

Level of Budget Execution According to the Professional Profile of Regional Governors Applying Machine Learning Models

José Luis Morales Rocha¹, Mario Aurelio Coyla Zela², Nakaday Irazema Vargas Torres³
Jarol Teófilo Ramos Rojas⁴, Daniel Quispe Mamani⁵, José Oscar Huanca Frias⁶

Universidad Nacional de Moquegua, Moquegua, Perú^{1,2,3,4}
Universidad Nacional de Juliaca, Juliaca, Perú^{5,6}

Abstract—Machine Learning is a discipline of artificial intelligence that implements computer systems capable of learning complex patterns automatically and predicting future behaviors. The objective was to implement a Machine Learning model that allows to identify, classify and predict the influence of the professional training of the governors in the execution of the public spending of the regional governments of Peru. Of the 14 indicators of academic training, professional experience and university studies were selected as significant indicators that contribute to the execution of public spending by the 25 governors of Peru. For the prediction of the execution of the public spending of the regional governors, a supervised learning algorithm was implemented. The mean square error for the Machine Learning regression model was 4.20 and the coefficient of determination was 0.726, which indicates that the execution of public spending by regional governments is explained with 72.6% by the professional experience and university studies of the governors. The regional governors of Peru with university studies and professional experience achieve better results in the execution of public spending in the regional governments of Peru.

Keywords—Machine learning; multiple regression; professional experience; university studies; public budget; governor; public spending

I. INTRODUCTION

The execution of public spending is carried out by the institutions of the public sector of a country over a year. It is carried out to acquire goods and services and for the provision of subsidies and transfers, in order to satisfy the needs of its inhabitants, public consumption and contribute to the redistribution of wealth [1]. In [2], the reduction in public expenditure causes the economic growth, and the deterioration of the population's living conditions. Currently, it is observed that public expenditure is insufficient to adequately meet the duties that the State has with its population. Educational training [3] is the key factor for the development of a country.

Through the observation technique, 14 indicators of academic training of the governors of Peru were collected through the electoral platform of the National Elections Jury, the execution of public spending through the Portal of Economic Transparency of the friendly consultation of the Ministry of Economy and Finance. With the multiple regression analysis, professional experience and university studies were selected as significant indicators that contribute to

the execution of public spending by the governors of the 25 regions of Peru. The determination coefficient (0.726) of the Machine Learning regression model indicates that the execution of public spending by regional governors is explained in 72.6% by professional experience and university studies.

The purpose of the research is to propose a machine learning model with Machine Learning techniques that allows to identify and predict the influence of the professional training of the governors in the execution of the public spending of the regional governments of Peru.

In this context, [4], [5] the professional profile of regional governors stands out, as they must achieve results with significant coefficients in executing the budget according to a schedule with the objective of reducing the gaps of economic inequality and unsatisfied social needs in an efficient, effective and transparent manner, for the welfare of citizens and thus achieving regional and national development.

This paper is organized as follows. Section II reviews some related works. Section III is made up of the theoretical background; Section IV is the presentation of the obtained results. Section V is titled discussion and contains the analysis of the results. Section VI contains the paper conclusion, followed by the last section, which is Section VII and presents suggestions for future research.

II. RELATED WORKS

This section presents the references of different investigations related to Machine Learning, academic training and public spending.

In [6] proposes a machine learning approach to detect and prevent cyberbullying using machine learning techniques. Evaluation of the proposed approach to the cyberbullying dataset shows that the neural network works best achieving 92.8% accuracy and support vector machines reaches 90.3.

In [7] he considers that many Machine Learning approaches are used to generate different models for prediction. However, she claims that the success of Machine Learning-based approaches depends on several factors. Likewise, it considers that no particular Machine Learning technique is effective in all its applications and that the success of the technique depends on the application in the problem to

be solved, so it is important to understand its behavior that guarantees to use the technique. It also uses conventional statistical techniques for bioclimatic modeling.

In [8] he tries to predict the Facebook profile using the Machine Learning technique, stating that the candidates can be chosen. They identified the characteristics that can be extracted from Facebook, through which the personality prediction is viable, the data has been extracted using the Facebook Graph API, which was carried out on a web page. To build the Machine Learning knowledge base, the personality test was implemented for students close to graduating from the University, in order to execute the training and categorization of the machine learning models through the use of the tool for the knowledge analysis, and thus check the degree of accuracy of the algorithms used in predicting the personality of the Facebook user.

In [9] he considers that the promotion of intrusion detection methods in computer networks poses a challenge for researchers, because with the growth of computer networks, new content-based infiltrations constantly appear. The work constitutes different Machine Learning techniques applied to the data processing stages for detection. Taxonomies and connection attribute classification sketches are described. In the detection of anomalies from Machine Learning techniques, it is concluded that it is of great applicability for those who seek areas within the detection of intrusion in computer networks from Machine Learning techniques.

In [10] he mentions that Machine Learning or machine learning is based on the process of systems that learn from historical data to predict future data. Machine Learning in different environments requires a large amount of multivariate and multidimensional environment data, which need to be analyzed to diagnose through statistical procedures, which represents an area of opportunity for machine learning.

In [11], k-means is selected as the base clustering and provides an algorithm for clustering sets of multiple clusters of k-means based on a hypothesis. In addition, they study the extraction of credible local labels from a grouping of bases, the production of different groupings of bases, the construction of the grouping relationship, and the final assignment of each object. In [12] proposes a clustering method based on distance weighted and K-means. In [13] reviewed and applied two known and used clustering methods, k-means and hierarchical clumping, to air pollution studies.

In [14] he considers that applying the technique of documentary review and gathering information for the budget execution of the Portal of the Ministry of Economy and Finance and for the variable of professional training of the mayor of the National Elections Jury Portal. Results were obtained that show that through statistical tests, the influence of the mayor's professional training in the budget execution of the Puno Region is not significant, concluding that the labor practice in the public sector and the mayor's age, more district poverty has a significant influence on budget execution; However, the level of education and the specific profession of the mayor do not significantly influence the budget execution of the Puno Region, years 2015-2018.

As indicated [15] it analyzes public debt in Tamaulipas and compares it with the evolution of public spending, indicating which items and where spending has been directed in the period (2003-2013). First, the situations that affect the public indebtedness of the states are studied; low fiscal pressure, absence of fiscal sovereignty, high state public expenditures of the governing parties, restricted public financial transparency. The state's indebtedness is also analyzed in relation to federal transfers and the application of the resources transferred from the expenditure budgets of that period. It concludes that an assessment of the current situation of Tamaulipas that it presents in terms of indebtedness, verifying that the accelerated growth of debt and public spending in Tamaulipas in the analyzed period, which is reflected in comparable economic growth rates, losing positions in various rankings of competitiveness and safety.

As indicated [16] in their research *Influência dos public expenditures do not grow economico dos municípios da Região Sudeste do Brasil*. The public expenditures analyzed were related to health, education and culture in 2010, in this investigation it shows a multivariate nonlinear regression mathematical model, whose purpose is to analyze the relationship between economic growth and public spending. The results showed an estimation error of 14.98% on average for the municipalities analyzed. The explanatory power of the model was 97.7% with high reliability. The State of Sao Paulo showed the highest economic growth and the State of Rio de Janeiro among the smallest. The evidence found indicates that the southeast of Brazil there is a positive influence of public spending in relation to economic growth, with the highest spending on education and health. With the application of the model it is inferred that public spending drives the municipal and or state GDP. Concluding that public spending plays an important role for economic growth in the Southeast Region of Brazil.

As pointed out [17] in their article the effect of public spending on the chances of reelection of Spanish local governments during the period 2000-2007, using the logit methodology for panel data. The results of which indicate the increases in municipal public spending have a positive impact on the chances of reelection of local governments, and these increases when the pre-electoral period approaches. We can also observe that there are other variables that positively affect the probability of reelection, such as the volume of income from transfers, which are right-wing parties, or the fact that a municipality has obtained an absolute majority in the previous elections. But we can also observe that the number of years that a mayor is in office has a negative impact on the chances of reelection.

As indicated by [18] in their research *Institutional operational plan and the efficiency of public spending in regional governments* whose purpose was to determine if the Institutional Operational Plans (POI) affect the efficiency of public spending in the departments from Peru, being an applicative type research with a cross-sectional, analytical, observational and non-experimental design. The approach was quantitative and the level was explanatory correlation, whose purpose was to determine the degree of relationship of the variables, in a population of 25 institutional operational plans

of the Regional Governments of Peru, the sample was non-probabilistic (10 POI of the Regional Governments that had greater budget execution); the instrument that was used was the data record, structured in 8 sentences. Whose result was 0.078, that is, the institutional operational plans do not affect the efficiency of the public spending of the Regional Governments in fiscal year 2018 and this leads us to the following conclusion that the Institutional Operational Plan (POI) did not affect the efficiency of public spending of Regional Governments during 2018 in the health and education sector.

As stated, [19], [20] in their article The impact of public spending on private investment in Mexico (1980-2015). The objective of which was to investigate the relationship between private investment, public spending and public investment in Mexico, for the period 1980-2015. A time series analysis was carried out using an ADL model, which included the variables private investment, primary public spending and GDP. The results obtained that the total net effect of primary public spending and GDP on private investment is positive and considerable. Therefore, in Mexico, between 1981 and 2015, the fall in private investment as a proportion of GDP can be explained by the fall in different types of public spending as a proportion of GDP. Coming to the following conclusion.

Government investment was reduced to such an extent that it acted to the detriment of the country's total investment and by remaining at low levels it cannot explain the behavior of private investment. Regarding the relationship between private and government investment, government investment spending must be very small and that it positively influences private investment must exceed the minimum level. The results of the econometric analysis show that the total net effect of primary public spending on private investment is positive and of considerable magnitude.

III. THEORETICAL BACKGROUND

Machine learning is considered a discipline in the field of artificial intelligence, responsible for developing systems that are capable of automatically learning complex patterns with a huge amount of data to predict future behaviors [21].

A. Multiple Regression Model

Machine Learning [22] is considered by many researchers as Artificial Intelligence, one of the most important characteristics of Artificial Intelligence is the ability to learn. Machine learning was designed with the purpose of implementing computer systems that can adapt and benefit from its knowledge. The Machine Learning technique consists of learning the data inputs, evaluating the results of the model and optimizing the output [23].

Machine learning algorithms consist of the following main models:

- 1) *Supervised learning*: the algorithm is trained with the inputs and outputs, in order to predict the output of future inputs.
- 2) *Unsupervised learning*: the algorithm is presented with inputs without desired outputs.
- 3) *Reinforced learning*: the algorithm interacts with the environment and achieve a specific goal without training.

The machine learning algorithm [24] is composed of supervised, semi-supervised and unsupervised learning, it is applied in different areas of knowledge.

Machine learning [25] instead of feeding the data into the program, now uses the data and the output it has collected to derive its program (also known as a model). This model can be used to make predictions.

Machine learning is a collection of algorithms and techniques that are used to design systems that learn from data. Machine learning algorithms have a solid mathematical and statistical foundation, but do not take into account domain knowledge. Machine learning consists of the following disciplines [25]:

- 1) Scientific computing
- 2) Mathematics
- 3) Statistics

B. Types of Machine Learning Algorithms

There are the following types of machine learning algorithms:

1) *Supervised learning algorithms*: are those algorithms that are trained with labeled data. This means that they are data composed of examples of the desired responses. Most of the machine learning is supervised.

2) *Unsupervised learning algorithms*: are those algorithms that are used in data without labels, and the objective is to find relationships in the data.

C. Multiple Regression Model

The construction of the multiple regression model is an iterative process, this process of construction of the model is helped by graphs that help to visualize the relationships between the different variables in the data, generate associations between the variables of the problem under study and consider the importance of develop such relationships between variables. This model can be fit and inferred, performing fit diagnostics to verify the assumptions of the model [26]. It is essential to know how the academic training of the governors intervenes in the budget execution.

Estimation and inference in multiple linear regression. The multiple linear regression model is as follows:

$$E(Y|X_1 = x_1, \dots, x_p = x_p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (1)$$

So,

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + e_i \quad (2)$$

Where:

e_i = random fluctuation (or error) in Y_i such that $E(e_i|X) = 0$. In this case, the response variable Y_i is predicted from p predictor (or explanatory) variables X_1, X_2, \dots, X_p , and the relationship between Y_i and X_1, X_2, \dots, X_p , is linear in the parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_p$.

1) Least squares estimates: The least squares estimate of $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are the values of $b_0, b_1, b_2, \dots, b_p$ for the sum of the squared residuals.

$$RSS = \sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

$$RSS = \sum_{i=1}^n (y_i - b_0 - b_1x_{1i} - b_2x_{2i} - \dots - b_px_{pi})^2 \quad (4)$$

is a minimum. For RSS to be minimal with respect to $b_0, b_1, b_2, \dots, b_p$ we need

$$\frac{\partial RSS}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1x_{1i} - b_2x_{2i} - \dots - b_px_{pi}) = 0$$

$$\frac{\partial RSS}{\partial b_1} = -2 \sum_{i=1}^n x_{1i} (y_i - b_0 - b_1x_{1i} - b_2x_{2i} - \dots - b_px_{pi}) = 0$$

⋮

$$\frac{\partial RSS}{\partial b_p} = -2 \sum_{i=1}^n x_{pi} (y_i - b_0 - b_1x_{1i} - b_2x_{2i} - \dots - b_px_{pi}) = 0$$

This gives a system of $(p + 1)$ equations in $(p + 1)$ unknowns. In practice, a software package is needed to solve these equations and thus obtain the least squares estimates, $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$.

2) *Least squares matrix formulation*: A convenient way to study the properties of the least square estimates, $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ is to use matrix and vector notation. Define the vector $(n \times 1)$, Y , the matrix $n \times (p + 1)$, X , the vector $(p + 1) \times 1$, β of unknown regression parameters and the vector $(n \times 1)$, e of random errors by.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

We can write the multiple linear regression model in matrix notation as:

$$Y = X\beta + e \quad (5)$$

Also, let x_i denote the i th row of matrix X . Then:

$$x_i = (1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}) \quad (6)$$

is a row vector $1 \times (p + 1)$ that allows us to write in the following way:

$$E(Y|X = x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + e_i = x_i\beta \quad (7)$$

The residual sum of squares as a function of β can be written in matrix form as:

$$RSS(\beta) = (Y - X\beta)'(Y - X\beta) \quad (8)$$

D. K-Means Algorithms

In the k-means algorithm [13] k points are randomly selected from a data set to be considered as the initial central points for the grouping. The Euclidean distance is then

determined by using the distance to determine the distance between the data points and the centroids of the cluster, and the data set is grouped again according to the distance.

Finally, the average distance in each group is calculated, and the new center is adjusted in the data set of each group, finally the final result of the grouping is obtained through multiple iterations.

The k-means clustering algorithm [27] is a fundamental clustering technique in the field of machine learning. Clustering begins by randomly initializing a centroid for each of a total of k groups.

$$c_i = \operatorname{argmin} |x_i - \mu_j|^2 \quad (9)$$

Also, m and k are the number of data and groups, respectively, and c_i is the index of the group to which the i th center is now assigned.

Second, assuming that the group assigned to each x_i is constant, the new centroids are calculated using a new subset of points, where m y k are the number of data and groups, respectively, and c_i is the index of the group to which i th data that is now assigned.

$$\mu_j = \frac{\sum_{i=1}^m [c_i=j] X_i}{\sum_{i=1}^m [c_i=j]} \quad (10)$$

where X shows 1 if the condition is true and 0 otherwise. This two-step process is repeated iteratively until convergence is achieved, ultimately minimizing the error criterion.

$$\varepsilon = \sum_{j=1}^k \sum_{i=1}^m [c_i = j] \|x_i - \mu_j\|^2 \quad (11)$$

The algorithm will converge to a minimum value, so that the result is sensitive to the initial position of the centroids and will generally only converge to a minimum value.

Centroids at different locations are often initialized and error checked to determine the optimal value.

IV. RESULTS

This section shows the results obtained from applying supervised machine learning algorithms.

A. Research Variables

The indicators of academic training of the governors of the regions of Peru were obtained from the National Elections Jury (consultation of jurisdictional files), the execution of public spending of the regional governments from the Portal of Economic Transparency of the friendly consultation of the Ministry of Economy and Finance (MEF).

Of the 14 indicators of academic training, professional experience and university studies were selected as significant indicators that contribute to the execution of public spending by the governors of the 25 regions of Peru.

B. Statistic Analysis

The measures of central tendency on the quantitative characteristics of the research are shown.

In Table I, it is shown that the average age of the governors is 57 years, with 11 years of professional experience on

average, an average income of PEN 91298 and 86% on average of execution of public spending by regional government. Other descriptive measures such as the amount of data, standard deviation, maximum and minimum values and quartiles are observed for each of the quantitative characteristics of the regional governors of Peru.

In Fig 1, the correlations that exist between the quantitative variables are shown. We can observe that only between the execution of public spending and professional experience there is a moderate correlation ($r = 0.43$), which is not observed with the other variables.

In Fig. 2, a scatter diagram is shown, indicating that the greater the professional experience, the greater the execution of public spending in the regional governments of Peru. It is also observed that governors with university studies show a greater execution of public spending.

C. Regression Analysis

In Table II, the summary of the multiple regression model is shown.

The results show a coefficient of determination of 0.378, which indicates that the execution of public spending is explained in 37.8% by the professional experience and by the university studies of the regional governors of Peru.

TABLE I. DESCRIPTIVE STATISTICS

Statistics	Age	Professional experience	Income	Public spending execution
count	25.00	25.00	25.00	25.00
mean	57.16	11.48	91298.16	86.07
std	10.92	11.39	110761.60	5.91
min	38.00	0.00	0.00	68.65
25%	51.00	4.00	21600.00	83.44
50%	57.00	7.00	63700.00	87.05
75%	67.00	14.00	106200.00	90.64
max	74.00	42.00	412938.00	94.65

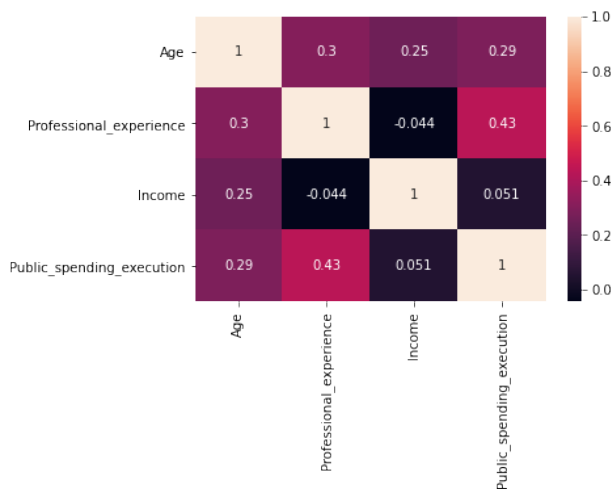


Fig. 1. Correlation Matrix.

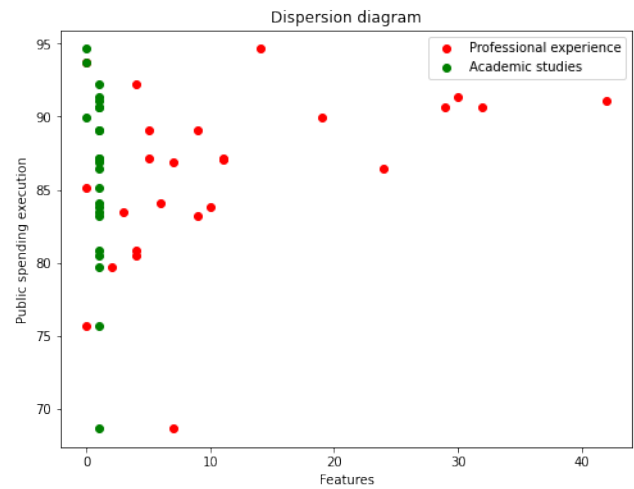


Fig. 2. Dispersion Diagram.

TABLE II. MODEL SUMMARY

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,615 ^a	,378	,321	4,86272

a. Predictors: (Constant), Professional experience, University studies

In Table III, the Analysis of Variance (ANOVA) 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 = 0.005) is less than 0.05 of significance, this means that the regression model is significant.

The regression model (Table IV) shows that the coefficients of the regression are significant, this indicates that professional experience and university studies are determining factors that contribute to the execution of public spending by the regional governors of Peru.

TABLE III. ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	315,726	2	157,863	6,676	,005 ^b
Residual	520,214	22	23,646		
Total	835,940	24			

a. Dependent Variable: Public spending execution

b. Predictors: (Constant), Professional experience, University studies

TABLE IV. COEFFICIENTS^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	90,293	2,967		30,437	,000
University studies	-7,761	2,993	-,436	-2,593	,017
Professional experience	,228	,087	,440	2,616	,016

a. Dependent Variable: Public spending execution

D. Supervised Machine Learning Algorithms

The Machine Learning linear regression model algorithm has the following characteristics:

- Independent variables: 02
- Dependent variables: 01
- Training set: 70%
- Test set: 30%

The coefficients of the multiple linear regression model applying Machine Learning techniques are shown in the table.

In Table V the regression model with the application of Machine Learning techniques is shown:

$$Pse = 90.25 + 0.2299 Prof.exp. - 8.0086 Univ.stud \quad (12)$$

The mean square error for the Machine Learning regression model is: MSE = 4.20303.

Table VI shows the prediction values with the Machine Learning model.

E. Evaluation and Validation of the Algorithm

Mean square error is the most widely used endpoint for regression supervised learning problems.

In Fig. 3, it is observed that the prediction values for cases, the results obtained are very similar to the test data set, the low value of the mean square error (4.20) indicates that there is a good fit of the regression model of Machine Learning.

F. Cluster Analysis with K-Means

There are really many alternatives for Cluster formation, but for the present study k-means algorithms were used.

In Fig. 4, a characterization map of the regional governments of Peru is observed, according to some characteristics of the regional governors such as age, academic studies, professional experience and execution of public spending in the regional governments of Peru.

TABLE V. COEFICIENTES DEL MODELO

Model	B
(Constant)	90.2501657
University studies	0.22998494
Professional experience	-8.00861339

TABLE VI. PREDCCIÓN DEL MODELO

Y test	Y predicted
90.64	88.91111546
86.88	83.85144686
91.13	91.90091964
80.81	83.16149205
85.16	82.24155231
86.43	87.76119078
91.33	89.14110040
83.44	82.93150712

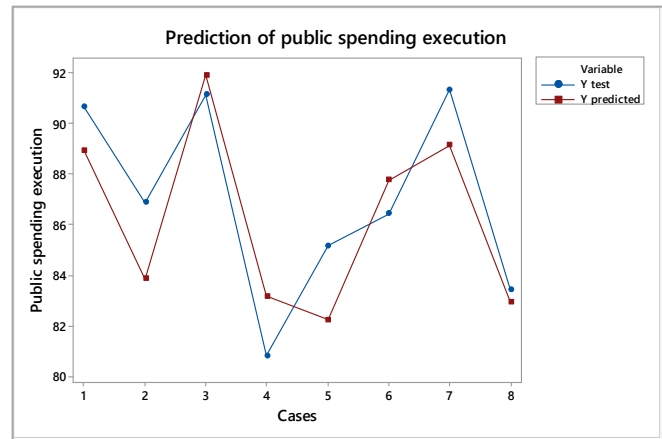


Fig. 3. Prediction of Public Spending Execution.

In Cluster 1, there are 7 departments, Ancash, Arequipa, Cusco, Junín, Pasco, Piura and Tacna.

In Cluster 2, we find 6 departments, Amazonas, La Libertad, Madre de Dios, Tumbes, Lima and the Constitutional Province of Callao.

Cluster 4 considers 7 departments, Apurímac, Ayacucho, Cajamarca, Huancavelica, Huánuco, Puno and Ucayali.

In Cluster 1, there are 5 departments, Ica, Lambayeque, Loreto, Moquegua and San Martín.

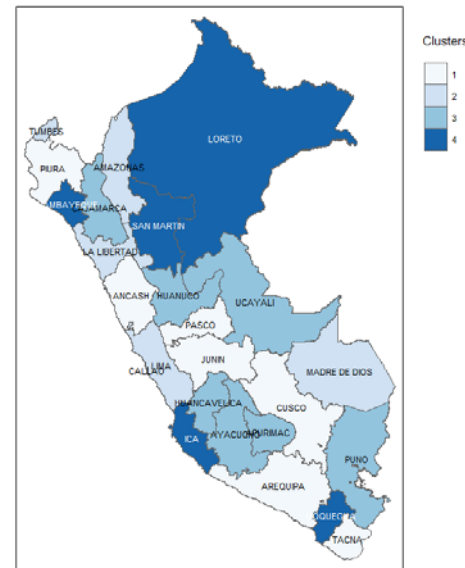


Fig. 4. Mapa De Caracterización De Gobiernos Regionales Del Perú.

TABLE VII. CLASIFICACIÓN DE LOS GOBIERNOS REGIONALES DEL PERÚ

Cluster	N	Age	Academic studies	Professional experience	Public spending execution
1	7	44	University	5	82%
2	6	68	University	6	83%
3	7	56	No university	9	90%
4	5	64	University	31	90%

According to the results of Table VII, Cluster 1 shows regional governors with an average age of 44 years, with university studies, with an average of 5 years of professional experience and are those who execute public spending on average in a 82%. In Cluster 2, there are those governors who, on average, are 68 years old, with university studies, on average 6 years of professional experience and execute 83% of public spending. In Cluster 3, there are governors with an average age of 56 years, they do not have university studies, but they have an average of 9 years of professional experience and they execute 90% of public spending. Finally, in Cluster 4, there are governors with 64 years of age on average, with university studies and with approximately 31 years of professional experience and show an execution of public spending of 90%.

V. DISCUSSION

According to the objective, to identify the factors that influence the professional training of governors in the execution of public spending of the regional governments of Peru, the results of Table IV show that professional experience and university studies present significant coefficients, this indicates that are determining factors that contribute to the execution of public spending of the regional governors of Peru, results that when compared with what was found in [28] who indicate that the variables included in the model are significant, this means that significance is necessary for The variable is considered as a predictor variable. These results affirm that professional experience and university studies are more significant factors that contribute to the execution of public spending by the regional governors of Peru. In the regression [26] the information is collected, the dependent and the independent variables are selected, the model is built and finally it is validated. It means that once, at least one of the regressors has been determined, it is important to answer the question Which one(s) is(are) useful? special attention must be taken when including regressors, because only the significant one(s) that have value to explain the answer must be considered.

Regarding the objective: to implement a Machine Learning model that allows making predictions in the execution of public spending based on the professional experience and university studies of the governors of Peru, the results indicate that the governors of the regional governments of Peru with studies University students and professional experience achieve better results in the execution of public spending in the regional governments of Peru. [14] concludes stating that work experience in the public sector has a significant influence on budget execution; However, the level of education and the specific profession of the mayor, do not significantly influence the budget execution of the Puno Region, it also concludes that most of the mayors have a level of professional education, which constitutes 32.11% , of the total and that the level of education and specific professional training of the mayors of the Puno Region are indistinct and varied with respect to the budget execution. These results also affirm that professional experience and higher education influence the execution of public spending by the governors of Peru.

VI. CONCLUSIONS

Professional experience and university studies are the most significant factors that contribute to the execution of public spending by the regional governors of Peru.

The governors of the regional governments of Peru with university studies and more professional experience achieve better results in the execution of public spending in the regional governments of Peru.

With the application of k-means algorithms, the regional governors of Peru were classified according to characteristics such as age, academic studies, professional experience and execution of public spending in 4 groups: Cluster 1, there are 7 governors of the departments from Ancash, Arequipa, Cusco, Junín, Pasco, Piura and Tacna, in Cluster 2 the governors of the departments of Amazonas, La Libertad, Madre de Dios, Tumbes, Lima and the Constitutional Province of Callao, in Cluster 3, It considers the governors of the departments of Apurímac, Ayacucho, Cajamarca, Huancavelica, Huánuco, Puno and Ucayali and Cluster 4 includes the governors of the departments of Ica, Lambayeque, Loreto, Moquegua and San Martín. The governors with 64 years of age on average, with university studies and with approximately 31 years of professional experience, show a greater execution of public spending.

VII. FUTURE WORK

Future work will focus on proposing Machine Learning models to analyze:

The theory of public choice and orientation of expenditures executed by the Peruvian municipalities.

Incidence of the modalities of labor connection in the financial effectiveness of the Peruvian municipalities

REFERENCES

- [1] R. Durán, A. Barreix, L. Corrales, and A. Rasteletti, "Reengineering the Expenditure Budget of the Federation in Mexico The experience of a modern zero-based budget." Felipe Herrera Library of the Inter-American Development Bank, 2018, doi: <http://dx.doi.org/10.18235/0001225>.
- [2] C. Tello, "Austerity, public expenditure and economic growth with social justice," *Jornal Econ. Lit.*, vol. 16, pp. 54–60, 2019.
- [3] O. Holguín and J. R. Dos Santos, "Economy and Education: An Approach from the Perspective of Development and Economic Growth," *Comp. Res. Public Policies*, pp. 63–86, 2019.
- [4] M. Marcel and M. Guzman, "Budgets for development in Latin America." Felipe Herrera Library of the Inter-American Development Bank, 2014.
- [5] E. Cavallo and A. Powell, "Building opportunities to grow in a challenging world Latin America and the Caribbean Macroeconomic Report 2019." Felipe Herrera Library of the Inter-American Development Bank, 2019.
- [6] J. Hani, M. Nashaat, M. Ahmed, Z. Emad, E. Amer, and A. Mohammed, "Social media cyberbullying detection using machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 703–707, 2019, doi: 10.14569/ijacsa.2019.0100587.
- [7] M. Bhattacharya, "Machine Learning for Bioclimatic Modelling," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 2, pp. 1–8, 2013, doi: 10.14569/ijacsa.2013.040201.
- [8] E. R. Sarchi Albuja, "Predicting Personality from Facebook Profiles Using Machine Learning to Support the Selection of Human Talent Personnel," University of the Armed Forces, 2018.

- [9] J. L. Rivero, "Machine learning techniques for intrusion detection in computer networks," *Cuba. J. Comput. Sci.*, vol. 8, no. 4, pp. 52–73, 2014. [Online]. Available: <http://repositorio.espe.edu.ec/xmlui/handle/21000/14165>.
- [10] J. C. González-Islas, "Machine Learning in Physiotherapeutic Applications," *Pädi Sci. Bull. Basic Sci. Eng. ICBI*, vol. 7, no. Especial, pp. 104–110, 2019, doi: 10.29057/icbi.v7iespecial.4473.
- [11] L. Bai, J. Liang, and F. Cao, "A multiple k-means clustering ensemble algorithm to find nonlinearly separable clusters," *Inf. Fusion*, vol. 61, no. January, pp. 36–47, 2020, doi: 10.1016/j.inffus.2020.03.009.
- [12] W. Yang, H. Long, L. Ma, and H. Sun, "Research on clustering method based on weighted distance density and k-means," *Procedia Comput. Sci.*, vol. 166, pp. 507–511, 2020, doi: 10.1016/j.procs.2020.02.056.
- [13] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, 2020, doi: 10.1016/j.apr.2019.09.009.
- [14] Y. M. Poma, "Influence of the mayor's professional training in the budget execution of the Puno region, years 2015–2018," National University of the Altiplano, 2019.
- [15] F. García Fernández, R. A. Vaquera Salazar, and J. A. Serna Hinojosa, "Tamaulipas: indebtedness and public spending (2003–2013)," *Inf. Econ.*, vol. 403, pp. 70–90, 2017, doi: 10.1016/j.ecin.2017.05.005.
- [16] L. Degenhart, M. Vogt, and V. C. da S. Zonatto, "Influência dos gastos públicos no crescimento econômico dos municípios da Região Sudeste do Brasil," *REGE - Rev. Gestão*, vol. 23, no. 3, pp. 233–245, 2016, doi: 10.1016/j.rege.2016.06.005.
- [17] M. T. Balaguer-Coll and M. I. Brun-Martos, "The effect of public expenditure on the probability of mayors' reelection," *Account. Mag. Spanish Account. Rev.*, vol. 16, no. 1, pp. 74–80, 2013, doi: 10.1016/S1138-4891(13)70008-6.
- [18] E. S. Chuquiayauri and G. G. Robles, "The institutional operational plan and the efficiency of public spending on regional governments," no. 2018, pp. 6–16, 2020.
- [19] F. J. Santos, M. de la O. Barroso, and C. Guzmán, "The global economy and social enterprises," *World Econ. Mag.*, vol. 35, pp. 177–196, 2013, [Online]. Available: <https://www.redalyc.org/pdf/866/86629567010.pdf>.
- [20] F. S. Gutiérrez Cruz, "The impact of public spending on private investment in Mexico (1980–2015)," *Econ. UNAM*, vol. 14, no. 42, pp. 136–149, 2017, doi: 10.1016/j.eunam.2017.09.006.
- [21] A. D. J. Bedoya Carrillo, "Development of a web tool for classifying bone health in schoolchildren according to age and sex using machine learning techniques," Universidad Católica de Santa María, 2019.
- [22] A. Raghda Essam, E.-K. Hatem, L. Soha Safwat, and S. Yasmine Ibrahim, "Prediction of employee performance using machine learning techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 8, pp. 80–88, 2019, doi: 10.14569/IJACSA.2019.0100812.
- [23] F. K. Alsheref and W. H. Gomaa, "Blood diseases detection using classical machine learning algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 77–81, 2019, doi: 10.14569/ijacsa.2019.0100712.
- [24] N. Nadar and R. Kamatchi, "A novel student risk identification model using machine learning approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 11, pp. 305–309, 2018, doi: 10.14569/ijacsa.2018.091142.
- [25] L. Wei-Meng, *Python Machine Learning*, vol. 53, no. 9. Indianapolis: John Wiley & Sons, Inc, 2019.
- [26] D. J. Olive, *Linear regression*. Switzerland: Springer Nature, 2017.
- [27] T. A. Sipkens and S. N. Rogak, "Technical note : Using k -means to identify soot aggregates in transmission electron microscopy images," *J. Aerosol Sci.*, no. September, p. 105699, 2020, doi: 10.1016/j.jaerosci.2020.105699.
- [28] R. Vilà, M. Torrado, and M. Reguant, "Multiple Linear Regression Analysis with SPSS," *REIRE Rev. d 'Innovació i Recer. en Educ.*, vol. 12 (2), pp. 1–10, 2019, doi: 10.1344/reire2019.12.222704.