*International Journal of Latest Research in Humanities and Social Science (IJLRHSS)*
*Volume 06 - Issue 12, 2023*
*www.ijlrhss.com || PP. 70-79*

# Unemployment Rate Prediction Model Applying WEKA Technology Platform Peru Case and Latin America Comparative

Wilfredo Carranza[1], Francisca Vera[2], Jorge Lira[3]
[1]*Universidad Nacional Federico Villarreal, Facultad de Ingeniería Industrial y de Sistemas, Lima, Perú*
[2]*Universidad Nacional Federico Villarreal, Facultad de Ingeniería Industrial y de Sistemas, Lima, Peru*
[3]*Universidad Nacional Federico Villarreal, Facultad de Ingeniería Industrial y de Sistemas, Lima, Peru*

**Abstract:** Since it is of great interest to know the macroeconomic variables that affect employability in Peru, the aim of this research was to identify and learn about a basic econometric model that allows us to explain the relationships with the unemployment rate and, subsequently, to formulate a predictive model applying Machine Learning technology, so that the machine learns and provides us with estimated values with the highest degree of assertiveness, in the short term.

**Keywords:** unemployment, machine learning, WEKA, data mining, data mining

## 1. Introduction

In August 2023, the GDP contracted by 0.6% compared to the same month last year, according to figures from the National Institute of Statistics and Informatics (INEI) and, in October 2023, the government has acknowledged that we are entering a recession. Recession in Peru is an economic event that appears after more than 20 years and, today we are facing an uncertain economic situation, although the Ministry of Economy-MEF expressed optimism about an eventual recovery. However, the coastal El Niño climate event is expected to last until the autumn of 2024, according to the latest communiqué of the National Study of the El Niño Phenomenon (ENFEN). And, the consequences will be unfavourable, affecting the health of the most precarious populations due to the tremendous rains and floods, causing massive outbreaks of infections and the consequent economic outlays on health by the government, limiting public investment and reducing private business activity by limiting the hiring of personnel and leading to a rise in the level of unemployment.

Today, specialised opinions for the economy at the end of the year are divided: on the one hand, consultants anticipate a decrease of around 0.2% and, in contrast, international and governmental organisations predict an increase of 0.9%. This marked difference in the estimated figures creates the need to know the macroeconomic variables that influence the generation of unemployment as it is a social indicator of consumers' capacity to acquire goods or services.

In this context, we resort to the use of technological tools that facilitate the development of predictive models to identify and use correlated macroeconomic variables to forecast an unemployment rate. To this end, we have used the WEKA digital platform, which was created at the University of Waikato, New Zealand, for academic purposes.

## 2. State of the Art

[1] in his thesis detailed that the evolution of unemployment in Peru went from 9.4% in 2003 to 8.0% in 2005, and 5.9% by the end of 2018 showing a stable behaviour. The author used maximum likelihood regression using the Error Correction Model and found that the level of production, level of employment and income from work have an inverse effect on unemployment. The International Labour Organization (ILO) estimates unemployment in Latin America and the Caribbean at 28.8 million people by 2022, representing a decrease of 1.3 million unemployed compared to 2021, but 4.5 million more than in 2019.

[2] The issue of unemployment is one of the most prominent research topics related to the labour market and economic development in Peru. The purpose of this study was to evaluate the impact of macroeconomic variables on the Unemployment Rate during the period 2001-2019 in the Peruvian context. To this end, a multiple regression model using time series was employed to test the hypotheses. The results of various econometric analyses indicate that a decrease in macroeconomic variables is associated with a negative impact resulting in an increase in the unemployment rate. This supports the inverse relationship between the Unemployment Rate and Gross Domestic Product (GDP), Gross Fixed Investment and Public Expenditure.

[3] Machine learning, also known as Machine Learning, is a discipline within computer science that enables computers to learn and adjust using data, rather than requiring detailed programming for each new situation. Contrary to conventional programming methods, where a different solution is needed for each

problem, machine learning focuses on developing flexible algorithms that can detect and use patterns in diverse data sets. This area allows machines to optimise their performance and decision-making capabilities based on previous experience, eliminating the need for specific human guidance for each type of problem encountered. However,

[4] states that the machine learning algorithm consists of a set of mathematical methods and principles designed to enable a computer system to acquire knowledge from available data. These advanced predictive systems represent one of the key advantages of contemporary technology. They facilitate process automation, optimise decision-making and promote continuous learning based on the information collected. Moreover, these systems have the ability to self-improve over time, adapt to business development and are capable of adjusting to environments that are constantly evolving.

## 3. Methodology

The methodology of this research was based on an explanatory-correlational approach, where the macroeconomic variables that are causal to the unemployment rate are analysed, maintaining a direct correlation with it in a socio-economic context with a tendency to reduce unemployment, but with contradictory expectations regarding economic growth aggravated by threats of nearby natural disasters.

The elements of the methodological approach developed in the research study are described below:

**a) Spatial and temporal scope of the study**

The research is national in scope and comprises historical data, from 2003 to 2022, with predictions going forward, with 2023 being used to validate the learning tests of the Machine Learning based predictive model.

**b) Study sample**

The valid sample for the study is obtained from official sources such as the World Bank, which presents the unemployment rate for the period 2003 to 2022.

**c) Unit of analysis**

The unit of analysis that constitutes the object of the research is the macroeconomic variable Unemployment at the national level, studied over the last 20 years.

**d) Data collection techniques**

Data collection techniques will focus on direct observation, documentary analysis and content analysis, considering that our source is provided by official bodies, who carry out direct surveys of the target public, for example, INEI's National Household Survey.

**e) Procedure**

The research was carried out following a procedure specially designed for the Supervised Machine Learning methodology, comprising the following activities, grouped by phases:

**A. Identification of incident variables**

1. Identify the macroeconomic variables that affect the Unemployment variable, application of the Causal Diagram.

**B. Data acquisition**

2. Search and selection of official documents to obtain the time series of macroeconomic variables.
3. Collection and tabular recording of data.

**C. Validation of key macroeconomic variables**

4. Determine the correlation of the incident variables with the target variable - Unemployment.

**D. Machine Learning (ML) Modelling**

5. Study of supervised Machine Learning (ML) algorithms.
6. Training of the members of the research group in the use of the WEKA technological platform.
7. Training and testing with WEKA modules, especially Classification.
8. Train a supervised learning model.

**E. Pre-processing**

9. WEKA parameter settings
10. Data entry and pre-processing (using WEKA platform).
11. Selection of the algorithms to be used.

**F. Use of the Classification module**

12. Application of the model (processing) with selected algorithms.

**G. Analysis of results**

13. Identification of the predictive model with the highest assertiveness.
14. Assessment of the quality of the model.

**H. Comparison of Unemployment Rates in Latin America**

15. Search and selection of official documents from Latin American countries.
16. Comparative analysis of the resulting macroeconomic variables on Unemployment in Peru and nearby Latin American countries.

## 4. Results and Conclusions

The **results** are presented according to the order of the **specific objectives**:

1) Identify the macroeconomic variables that affect the Unemployment variable, formulating a Cause-Effect Diagram.

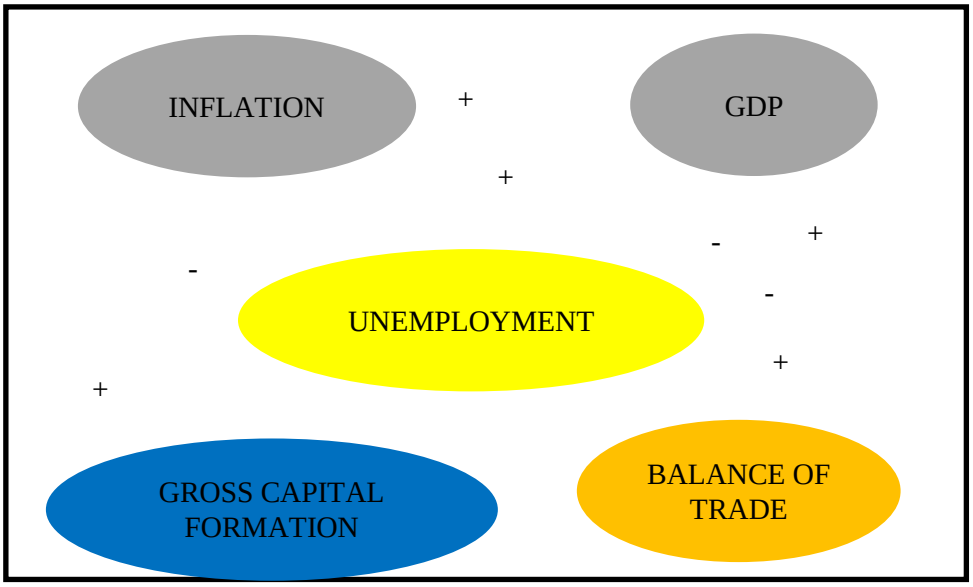



Figure 1 Cause-and-effect diagram of macroeconomic variables

2) Obtain time series of macroeconomic variables from official sources.

The official documents for obtaining the time series of macroeconomic variables correspond to World Bank sources:

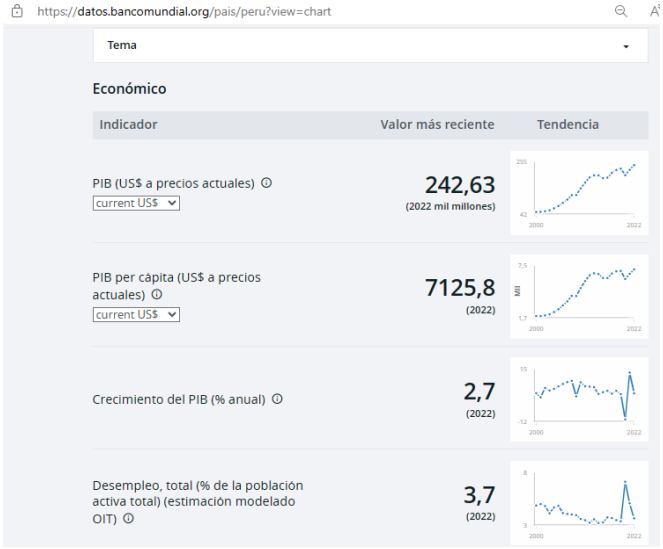



Figure 2 Statistical data from the World Bank

**3) Train a supervised machine learning model.**

**a. Validation of key macroeconomic variables**

The correlation coefficients with the target variable - Unemployment

| Inflation with Unemployment | GDP with Unemployment | Trade Balance with Unemployment | Gross Capital Formation with Unemployment |
|---|---|---|---|
| -0.15 | -0.43 | 0.24 | -0.57 |

It is observed that the variables with the highest incidence are GDP and GFCF.

**b. Machine Learning (ML) Modelling**

The WEKA platform was used, generating a predictive model with the annual time series data of the macroeconomic variables mentioned. The predictive model is based on the WEKA platform algorithms, which support Machine Learning applications.

**c. Pre-processing**

Previously, the Excel file is ready with the time series data of the incident macroeconomic variables and the values of the Unemployment variable, as shown in the following table.

| Year | Inflation | GDP | Balance of trade | FBC | Unemployment rate |
|---|---|---|---|---|---|
| 2003 | 2.26 | 4.17 | 0.40 | 17.51 | 4.15 |
| 2004 | 3.66 | 4.96 | 3.89 | 16.86 | 4.71 |
| 2005 | 1.62 | 6.29 | 6.34 | 17.28 | 4.87 |
| 2006 | 2.00 | 7.53 | 9.29 | 19.65 | 4.17 |
| 2007 | 1.78 | 8.52 | 7.36 | 22.03 | 4.08 |
| 2008 | 5.79 | 9.13 | 1.02 | 26.18 | 4.03 |
| 2009 | 2.94 | 1.10 | 4.77 | 19.98 | 3.96 |
| 2010 | 1.53 | 8.33 | 3.98 | 23.76 | 3.60 |
| 2011 | 3.37 | 6.33 | 5.02 | 24.20 | 3.48 |
| 2012 | 3.61 | 6.14 | 2.28 | 24.61 | 3.24 |
| 2013 | 2.77 | 5.85 | -0.17 | 25.57 | 3.57 |
| 2014 | 3.41 | 3.25 | -1.64 | 24.67 | 3.21 |
| 2015 | 3.40 | 3.95 | -2.59 | 24.31 | 3.27 |
| 2016 | 3.56 | 3.95 | -0.19 | 22.02 | 3.74 |
| 2017 | 2.99 | 2.52 | 1.93 | 20.71 | 3.69 |
| 2018 | 1.51 | 3.97 | 1.75 | 21.31 | 3.49 |
| 2019 | 2.25 | 2.24 | 1.13 | 20.83 | 3.38 |
| 2020 | 2.00 | -10.87 | 1.59 | 18.38 | 7.18 |
| 2021 | 4.27 | 13.42 | 3.02 | 21.88 | 5.10 |
| 2022 | 8.33 | 2.68 | 0.22 | 22.15 | 3.66 |

Table 1: Macroeconomic variables data

The file is converted to .csv format, which is one of the formats recognised by WEKA.

When the file is opened in the WEKA platform, the interface shown in Fig. 7. is observed, where the attribute YEAR is removed as it is irrelevant, as it does not affect the Unemployment value.

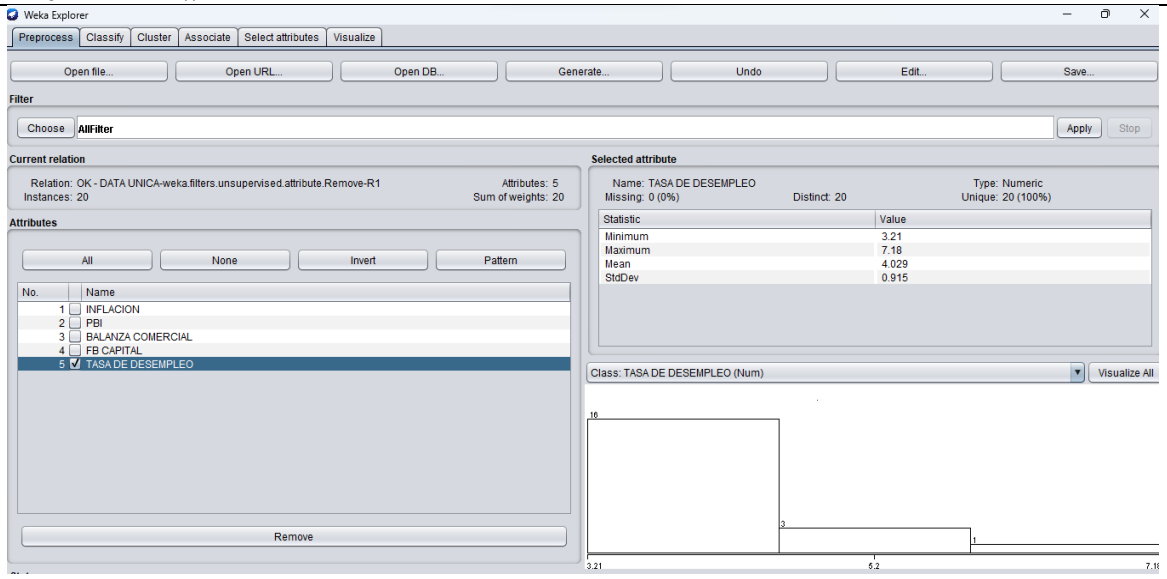



Figure 3. Pre-processing interface

**d. Use of the Classification module**

Model application (processing) with selected algorithms in Classification. Five envelopment strategy algorithms were used: Linear Regression (LR), Multilayer Perceptron (MLP), M5Rules (M5R) and, Zero R.

The algorithms that built the prediction model, formulating the Unemployment Rate, were:

- the Linear Regression model and,
- the M5Rules model.

Similarities were observed in the statistical values in Figures 4-A and 5-A.

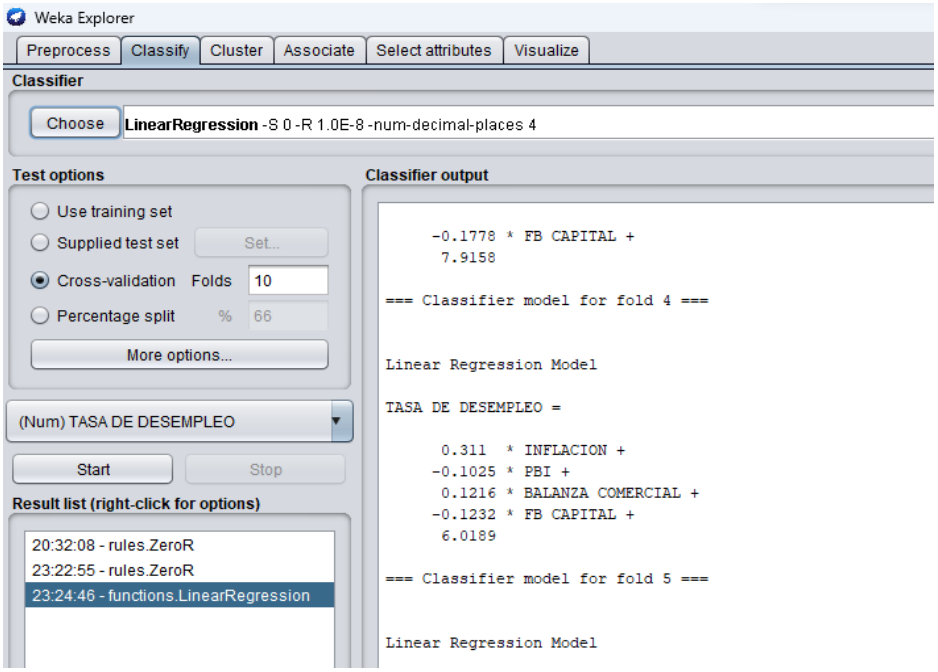



Figure 4: Linear Regression

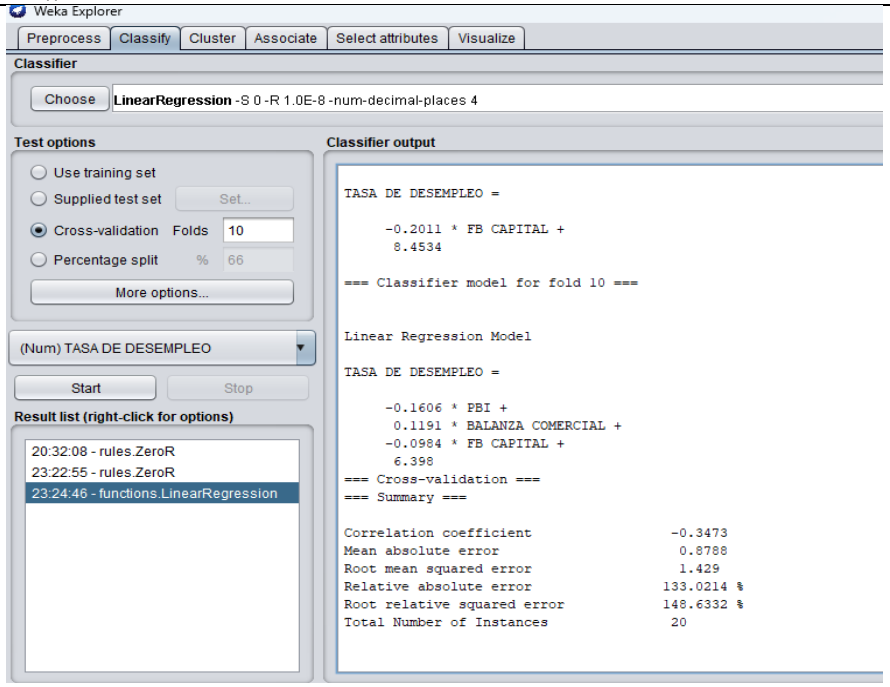



Figure 4-A: Linear Regression

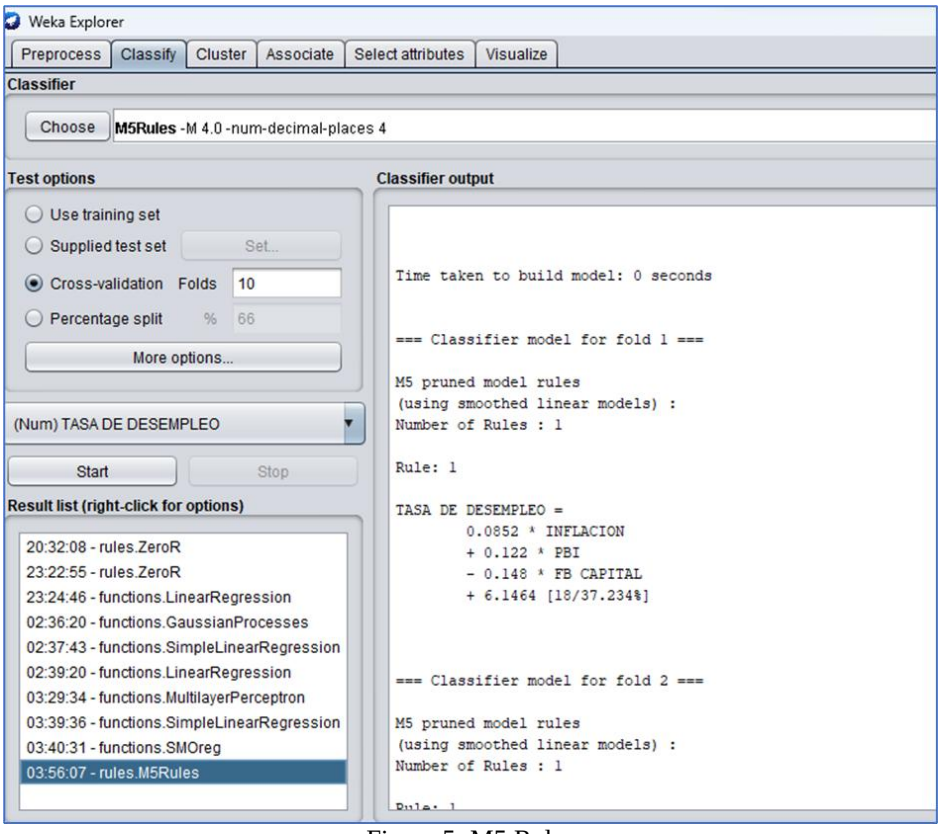



Figure 5: M5 Rules

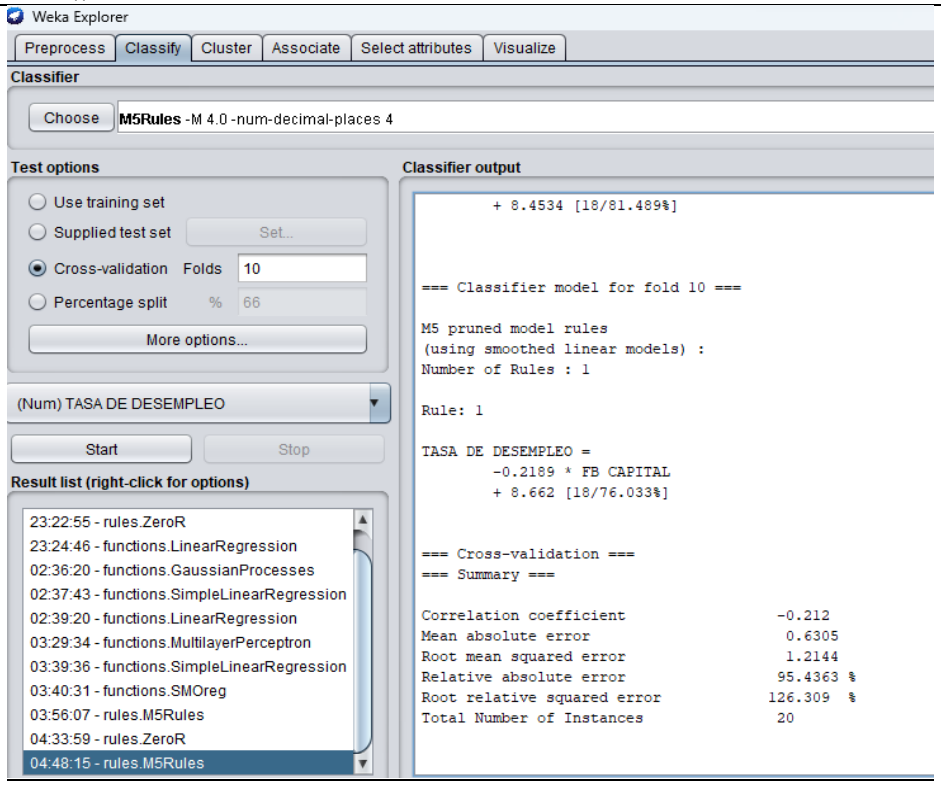



Figure 5-A: M5Rules

**4) Evaluate the quality of the predictive models (statistical data from the WEKA platform).**

The parameters and statistical indicators provided by the WEKA platform have been considered and are summarised in the following table.

Statistical values of the algorithms in WEKA, according to Table 4-A and 5-A

| Algoritmo | Coef. Correl. | Error Absoluto Medio | Error Absoluto Relativo |
|---|---|---|---|
| LinearRegression | -0.3473 | 0.8788 | 133.02% |
| M5Rules | -0.212 | 0.6305 | 95.43% |

Regarding the correlation coefficient:

- All of them vary between +1 and -1. With +1 being a perfect positive correlation and -1 being a perfect negative correlation.
- They are used as a measure of strength of association (effect size):
  - 0: no association.
  - 0.1: small partnership.
  - **0.3: median association.**
  - 0.5: moderate association.
  - 0.7: high association.
  - 0.9: very high association.

In both cases, a correlation close to **0.3: median association** is observed. In further tests, there is the expectation of improving the correlation by introducing other macroeconomic variables. The M5Rules algorithm presents lower values for the Absolute Error, both mean and relative.

**5) Comparative Analysis of Unemployment with other Latin American Countries**

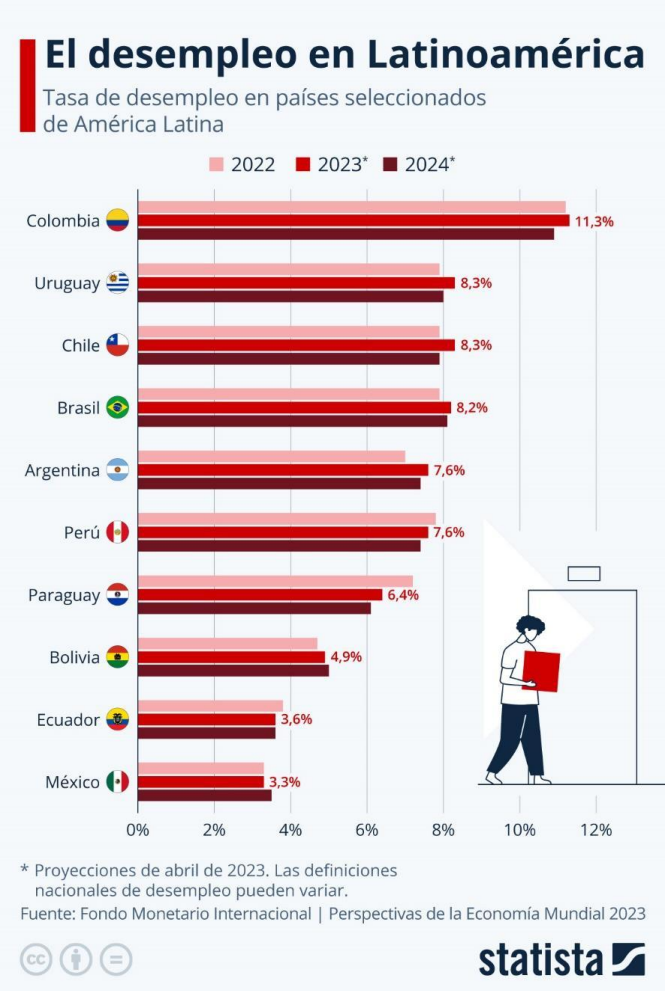
Information was compiled from the Economic Commission for Latin America and the Caribbean (ECLAC), an agency of the United Nations (UN), which presents the unemployment rate by country in 2020, according to the following **fact sheet:**

Latin America and the Caribbean: country performance rate in 2020
Source Economic Commission for Latin America and the Caribbean
Carried out by : Economic Commission for Latin America and the Caribbean
Period : 2020
Region : LAC

| **Unemployment rate by country (2020) in per cent** | |
|---|---|
| Costa Rica | 19.60% |
| Panama | 18.50% |
| Saint Lucia | 16.90% |
| Barbados | 15.60% |
| Colombia | 15.10% |
| Belize | 13.70% |
| Brazil | 13.50% |
| Argentina | 11.50% |
| Honduras | 10.90% |
| Chile | 10.80% |
| Latin America and the Caribbean | 10.50% |
| Uruguay | 10.10% |
| Bolivia | 8.30% |
| Paraguay | 7.70% |
| **Peru** | 7.60% |
| El Salvador | 6.90% |
| Jamaica | 6.40% |
| Ecuador | 5.90% |
| Dominican Republic | 5.80% |
| Nicaragua | 5% |
| Trinidad and Tobago | 4.70% |
| Mexico | 4.40% |
| Cuba | 1.40% |

Table 2: Comparative data, according to ECLAC

The table above shows that Peru, with 7.6 %, is below the LATAM average (10.50 %). And nearby countries, such as Bolivia, Chile, Argentina, Brazil and Colombia, have a higher unemployment rate, with the exception of Ecuador (5.9 %).




A synthesis of the main findings describes the following:

The prediction model of the unemployment rate using the WEKA Machine Learning platform confirms the results of the estimates of the official agencies - BCR, INEI, World Bank, achieving values with the highest degree of assertiveness, after testing with several algorithms it was determined that the Linear Regresion and M5Rules algorithms have a greater acceptance.

## 5. Conclusions

It is shown that the macroeconomic variables mentioned: Inflation, GDP, Trade Balance and Gross Capital Formation, affect the unemployment rate in an inverse relationship, i.e. if one increases, the other decreases.

The WEKA platform contains application software that has facilitated the testing of macroeconomic variables with various WEKA-specific algorithmic functions.

Also, an analysis of the unemployment rates of Latin American and Caribbean countries (provided by ECLAC) shows that Peru has maintained a lower unemployment rate in 2022 than the average of the other countries.

## 6. Recommendations

To make a Machine Learning application available to the university community using the WEKA platform, so that the multiple configuration options and the hundreds of models and algorithms contained in this academic platform, which was created precisely for this purpose, can be further investigated.

Conduct further research with other macroeconomic variables that may affect the unemployment rate in order to improve and refine the basic predictive model that has been developed in this research.

## References

[1] FLORES MAMANI, Adan Percy. Macroeconomic factors that determine unemployment in Peru, period 2001-2018.
[2] CALIXTO CORNEJO, Grecia Milagros; GOMEZ CONTRERAS, Maria Alejandra. Determinants of the variations of the unemployment rate as a function of macroeconomic variables for the period 2001-2019 in Peru. 2021.
[3] BOBADILLA, Jesús. Machine learning and deep learning: using Python, Scikit and Keras. Ediciones de la U, 2021.
[4] PINEDA, Javier Mora. Predictive models in health based on machine learning. Revista Médica Clínica Las Condes, 2022, vol. 33, no 6, p. 583-590.

## Author Profile

**CARRANZA BARRENA WILFREDO EDUARDO.** Systems Engineer, with CIP 89989; graduated from the National University of Engineering UNI-Peru; MBA in Business Management, graduated from the University of Tarapacá- UTA-Chile. Experience in project management of Information Technology and Communications ICTs - applying PMP and in projects of Continuous Process Improvement applying the CMMI Quality Model for the evaluation and improvement of processes; Business Process Outsourcing (BPO); business process re-engineering and implementation of integrated ERP systems; Strategic IT Planning, Business Process Modelling based on BPMN Trained in methodologies and ISO standards for process and software product quality (NTP ISO 12207-Software Cycle, ISO 15504-Quality Assessment, ISO 17799-Security, ISO 9126 and ISO14598 Software Product Assessment, ISO 9001:2015-Quality Management), 5S Method. Knowledge of ITIL and COBIT 5.0. GMD Internal Auditor - ISO 9001-2000 revision in the areas of Technology Solutions Division. Certified in BAAN IV-Manufacturing (ERP technology) Trained in ERP BAAN Enterprise Tools, Logistics, Distribution, Purchasing, Sales. Experience in multiplatforms: IBM Main Frame, Minicomputers, LAN and WAN Networks; Client/Server Architecture and Transactional Systems in Web environment. Lecturer at the Faculty of Industrial and Systems Engineering (FIIS) - Undergraduate, Universidad Nacional Federico Villarreal- UNFV. Lecturer at the Faculty FIIS- UNI in the Undergraduate and Postgraduate Diploma in Software Engineering and Quality Management. Member of ICACIT- Participation as Expert Evaluator (Program Evaluator) for Accreditation of University Vocational Training Programs and Educational Quality. Participation in Applied Research Projects at the Research Institute of the Faculty of Industrial and Systems Engineering of the UNI.

**SONIA FRANCISCA VERA TITO,** Systems Engineer, with a Master's Degree in Business Administration, with Doctorate studies in Systems Engineering, leadership with eminently humanistic, dynamic, scientific and technological training that I participate individually, collectively or forming multidisciplinary teams and with a systemic point of view with experience in General Management in Management and Educational Quality, I have experience in personnel management, administrative work, designing training programmes for teaching and administrative staff, experience in working with adolescents, young people and families with different problems, conducting workshops with leadership studies and the ability to develop and communicate in different cultural and denominational contexts. Experience and solid knowledge in the IT and administrative fields.

**JORGE LIRA CAMARGO,** Charismatic, proactive and responsible, focused on strategic planning, process management, organisational design and digital transformation. Eleven years of professional experience in public and private entities as analyst, specialist and consultant.