Received: 10 July 2023 | Revised: 8 August 2023 | Accepted: 11 September 2023 | Published online: 12 September 2023

RESEARCH ARTICLE

Archives of Advanced Engineering Science
2024, Vol. 2(1) 30–36
DOI: 10.47852/bonviewAAES32021326

BON VIEW PUBLISHING

# Application of Ordinary Least Squares Regression and Neural Networks in Predicting Employee Turnover in the Industry

**Bogart Yail Marquez[1,*], Arturo Realyvásquez-Vargas[2], Nelson Lopez-Esparza[1] and Carlos E. Ramos[1]**

[1]*Department of Systems and Computing, National Technological Institute of Mexico (Campus Tijuana), Mexico*

[2]*Department of Industrial Engineering, National Technological Institute of Mexico (Campus Tijuana), Mexico*

**Abstract:** Employee turnover, also known as labor turnover or employee attrition, refers to the flow of employees entering and leaving an organization within a specific time. It is an indicator used to measure the number of employees who leave a company and are replaced by new hires. This project aims to create and implement an artificial intelligence (AI) model using the Python programming language and the TensorFlow library. The focus is developing a dashboard to facilitate the model training process and enable predictions related to employee turnover in the business context. The goal is to enhance predictive capabilities and provide valuable strategic and human resources talent management and decision-making insights. By harnessing the power of AI, the project aims to identify patterns and factors that influence employee turnover. This, in turn, will enable the implementation of preventive measures and corrective actions to reduce turnover rates and maintain workforce stability in the company.

**Keywords:** artificial intelligence techniques, staff turnover, data mining, tensor flow, ordinary least squares regression

## 1. Introduction

In any company, employee turnover is a crucial indicator of the organization. It determines the difference between the number of employees who leave and rejoin in relation to the total number of employees within an organization. This indicator is a crucial variable as it incurs significant costs and losses for the company, resulting in decreased productivity due to constant personnel turnover. New employees require initial training to perform their assigned tasks, slowing the learning curve and significantly impacting the company's productivity (Appelbaum & Batt, 1994; Navarrete et al., 2024).

Employee turnover can be voluntary or involuntary. Voluntary turnover occurs when employees leave the company for various reasons, such as better job opportunities, job dissatisfaction, lack of professional development, conflicts with the organizational culture, and other personal motives (Mia et al., 2023). On the other hand, involuntary turnover happens when the company ends the employment relationship with an employee due to poor performance, restructuring, or layoffs due to economic reasons.

Employee turnover can have significant implications for a company. A high turnover rate can result in additional costs, such as recruitment and selection expenses for new employees, training costs, and decreased employee productivity and morale (Hidayat & Masdupi, 2023). Furthermore, constant turnover can impact business continuity and the quality of work performed. The present research study was conducted with the objective of analyzing and developing a predictive tool that enables the determination of indicators for turnover causes within a specific time for companies located in the city of Tijuana, Mexico. This tool aims to analyze and identify turnover indicators over a designated time frame.

Companies often monitor and analyze employee turnover rates to identify underlying causes and take measures to retain key employees. This may involve improving working conditions, offering growth and professional development opportunities, strengthening the organizational culture, and providing a proper work-life balance, among other strategies.

One of the problems that impacts not only this company but also companies in general is employee turnover, which opens up a research gap where innovative approaches are sought to address this issue (Ahmad et al., 2021; Yahia et al., 2021).

Creating an artificial intelligence (AI) model could prove valuable as it can help identify dates with a high risk of employee attrition in the near future (Davenport, 2018; Yadav et al., 2018). This knowledge could give the company a competitive advantage by enabling them to retain their most valuable and talented employees (Hadijah, 2023; Lawler, 2010). Additionally, it could facilitate better long-term hiring planning, allowing for more informed and strategic decision-making.

*Corresponding author: Bogart Yail Marquez, Department of Systems and Computing, National Technological Institute of Mexico (Campus Tijuana), Mexico. Email: bogart@tectijuana.edu.mx

According to Pedro Montejo Peterson, the President of Index Zona Costa BC, industrial parks in Baja, California reportedly began the year with a turnover rate of up to 120% in their workforce. He further elaborated that the industrial sector throughout the state started the year with over 30,000 job vacancies, with Tijuana having the highest demand for employment opportunities to date. Montejo Peterson stated, "There are industrial parks with very high turnover rates, averaging 84% annually at a general level. However, some industrial parks exceed 100%, and some even reach 120%" (Castañeda, 2023). A predictive tool is developed and implemented based on qualitative data processing and analysis. The master data collected from employee exit surveys covering 2019, 2020, 2021, and 2022 apply to companies.

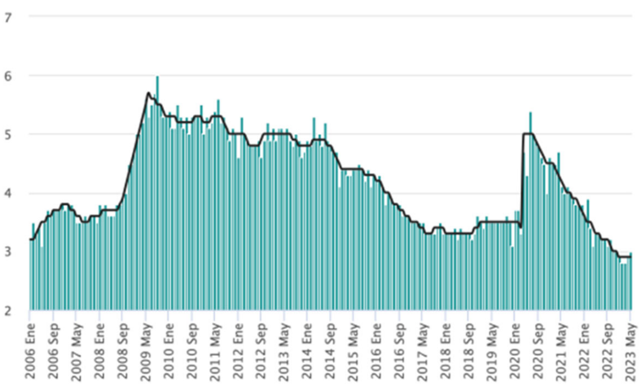
Therefore, we will discuss the problem statement, where important unemployment indicators are demonstrated. We will then present our methodology, implementing ordinary least squares regression (OLS) and neural networks. Finally, we will explain the obtained results.

## 2. Problem Statement

Employee turnover is a widely studied phenomenon that affects organizations in various ways. However, in the specific context of the company located in the city of Tijuana, a Mexican city experiencing constant economic growth and with a significant industrial presence, it is crucial to understand how this phenomenon impacts local businesses. The high employee turnover rate in Tijuana can have significant implications in terms of costs, work quality, and job stability. Therefore, it is essential to investigate and comprehend employee turnover's impact on Tijuana companies (INEGI, 2023).

Figure 1 shows the unemployment rate as a seasonally adjusted series and trend cycle. The unemployment rate refers to the percentage of working-age individuals who are unemployed and actively seeking employment to the economically active population. This measure is used to analyze unemployment in a specific economy or region. The seasonally adjusted series refers to removing seasonal effects that can influence the unemployment rate, such as expected variations in employment due to climatic or seasonal factors. A more accurate and comparable measure of the unemployment rate over time is obtained by removing these variations. The trend cycle refers to the long-term fluctuations in the unemployment rate that are not due to seasonal factors. This time series component shows the general direction of the unemployment rate over several periods and can reveal economic growth or contraction patterns.

**Figure 1**
**Unemployment rate**




The economically active population (PEA) refers to the group of individuals who are of working age and are available and willing to participate in economic activities, including paid employment or self-employment. This population includes employed individuals and those unemployed but actively seeking employment (Mayorga & Isabel, 2023). The PEA encompasses not only individuals currently employed but also those who are unemployed, actively seeking work, and available to start working. These individuals are considered economically active as they participate in the labor market and are part of the workforce of a country or region (INEGI, 2023).

The PEA is an essential indicator for analyzing the labor dynamics of an economy and assessing its ability to generate employment. It is also used to calculate the unemployment rate and other indicators related to the labor market. There are 60,089,308 economically active people in Mexico in this registered year (INEGI, 2023); Figure 2 shows how the economically active population increased.

**Figure 2**
**Economically active population**

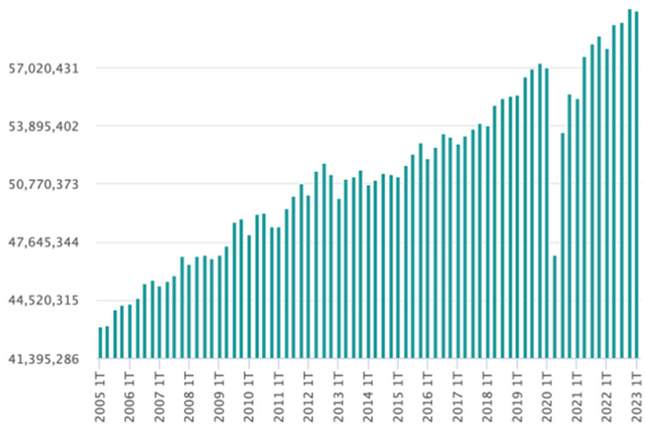



Figure 3 shows that "unemployment rates" refer to the proportion of the employed population working fewer hours than they would like or are willing to work more. These rates indicate the extent of underemployment in an economic (INEGI, 2023).

There are two main types of unemployment:

**Figure 3**
**Underemployment rate. Seasonally adjusted series**

Underemployment due to insufficient hours occurs when workers have a job but desire to work more than they currently have. It can result from limited full-time job opportunities or an involuntary reduction of working hours.

Overemployment refers to workers who have a full-time job but work more hours than they desire or are willing to work. Economic necessity, work-related pressures, or a lack of alternative employment options may drive it. Unemployment rates are essential indicators for understanding the quality of jobs and the extent of labor underutilization in an economy. They help identify needs and challenges related to the labor market, such as insufficient decent employment opportunities or exploiting workers. The unemployment rate, economically active population, and underemployment are related yet distinct concepts that can provide relevant information about employee turnover in Tijuana. These indicators can help understand the labor conditions and factors that may influence worker mobility in the labor market of Tijuana.

Two important lines of research are analysis of factors influencing employee turnover and trends related to employee turnover.

Analysis of Factors Influencing Employee Turnover: Employee turnover results from internal and external factors that affect an individual's behavior and attitudes. External factors range from job opportunities to personal economic aspects, while internal factors include growth opportunities and company salary policies (Kowssarie & Mansour, 2023). Organizational employee turnover is estimated to have an economic impact of 1.1 trillion pesos annually. The decline in productivity, along with the costs of recruitment and training when an employee leaves the company, results in significant losses in operations and finances. According to a survey by Indeed, 42% of workers in Mexico believe that the ideal time to stay in a company is less than 5 years (Forster & Kahan, 2023). Suppose employees do not experience what is promoted within the organization. In that case, the company may experience a sharp decline in job vacancies, requests from qualified candidates, and employee retention, as explained by Madalina Secareanu, Senior Manager of Corporate Communications for Indeed in Latin America (Forster & Kahan, 2023). The delay in implementing employee retention strategies is often due to the inability to recognize signals indicating dissatisfaction among the workforces. These symptoms often suggest that a worker is considering leaving the company. Providing a safe and healthy work environment, offering professional development opportunities, establishing a positive organizational culture, providing fair compensation and benefits, promoting work-life balance, and fostering employee engagement and involvement are some strategies that every company can use to enhance employee job satisfaction (Ballesteros et al., 2023). Hence, our AI model is employed to address the question of interest for the Human Resources personnel: "During which period does employee turnover have the greatest impact?". This question serves as the foundation for our investigation, guiding our analysis and helping us understand the critical factors influencing employee turnover within the specified time frames. By addressing this question, we aim to provide valuable insights into the patterns and trends of employee turnover, enabling organizations to make informed decisions and implement effective strategies to manage and reduce turnover rates effectively.

## 3. Methodology

The data provided serve as a guiding compass for our model, enabling us to discern its trajectory. By analyzing these data, we can define relevant categories or classes to which the observed data points belong. It is crucial to note that our model is based on regression analysis, necessitating quantifiable data utilization. These quantifiable attributes empower us to extract meaningful insights and make informed predictions based on statistical patterns and trends. This approach ensures a robust and data-driven foundation for our analysis, enhancing the accuracy and reliability of our model's predictions. As we delve deeper into the data, we uncover valuable information that enables us to understand the underlying dynamics and relationships within the dataset comprehensively.

Data Collection: The first step was acquiring data to feed the AI model. Data were collected through surveys administered to former employees who left the company. These surveys consisted of a letter that formalized the termination of their employment, from which the date and the employee's signature were collected. Additionally, an exit interview form was attached, capturing the employee's ID number, name, and a table to mark the reason for their departure, which included options such as relocating to another city, termination due to absenteeism, unjustified or poor performance, transportation issues, dissatisfaction with the job, pursuing further studies, finding another job with better salary, experiencing a family member's death, caring for children, changes in working hours, problems with immediate supervisors, health-related issues, termination due to low production or disciplinary reasons, starting a personal business, changing residence, pension, and lack of opportunities, among others.

The model adopted for this study is a linear regression, as it aims to predict the number of individuals leaving the company. Furthermore, the collected data can allow us to identify categories or classes to which the observed variables belong. In this case, the time-related variables are the ones under observation.

Extensive research was conducted on different libraries to develop the AI model. Firstly, TensorFlow was utilized (Pau et al., 2023), which provides a wide range of tools for creating AI models (Pattanayak, 2023). Alongside TensorFlow, the Keras internal library was employed to facilitate the construction of neural networks.

The Pandas library was also helpful for data reading and manipulation (Hodeghatta & Nayak, 2023). Specifically, it was employed to read CSV files that contain the data used to feed the AI model. The Matplotlib library was utilized for data visualization by creating graphs. It also provides tools for manipulating data structures such as vectors, matrices, and lists. To divide the data into training and testing sets, the Sklearn library was used, which is a versatile self-learning tool. Given the project's requirement for handling large amounts of data, the Numpy library was employed. This library is handy for efficiently managing numerical data and offers data structures like vectors and matrices suitable for storing and processing large volumes of information.

The support vector machines (SVMs) algorithm was initially used, which can only find hyperplanes in problems that allow linear separation. However, due to the nature of our dataset, finding linearly separable hyperplanes was not feasible. The rotation patterns exhibited a sigmoidal behavior with significant spikes in certain months, making it challenging to achieve the linearity the algorithm seeks.

To address the non-linearity of the data, the logistic regression algorithm was employed to measure the relationship between dependent variables, such as the number of rotations and the time in months. However, it encountered issues during model training and produced unreliable results, stagnating at specific points.

Ultimately, the OLS algorithm was the best fit for the data. OLS estimates coefficients for linear regression equations that describe the relationship between one or more quantitative independent variables and a dependent variable. This regression method minimizes the sum of squared errors, making it well-suited for this study.

OLS outperformed the other algorithms, displaying better behavior with the data. Unlike SVMs, it provided consistent predictions without stagnation issues, which overly maximized the margin in the presence of few data points, resulting in unrealistic rotation predictions. The OLS algorithm yielded more reliable predictions, ensuring better accuracy, and avoiding exponential fluctuations between months.

It is essential to highlight that various algorithms were employed to adjust the model. One of them was logistic regression, which was used due to the non-linear behavior of the data (Sonderegger, 2023). Specifically, the aim was to measure the relationship between the dependent variable, such as the turnover rate, and the time in months. Given a sample of input data $X = (X_1, X_2, \ldots, X_n)$ and its corresponding binary dependent variable $Y$ (0 or 1), the logistic regression model seeks to find the coefficients $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_n)$ that maximize the likelihood function.

The likelihood function is defined as:

$$L(\beta) = \Pi[P(Y = 1|X; \beta)\hat{}Y * (1 - P(Y = 1|X; \beta))\hat{}(1 - Y)] \quad (1)$$

$P(Y = 1|X; \beta)$ is the conditional probability of $Y$ being equal to 1 given $X$ and the coefficients $\beta$.

The likelihood function is maximized using optimization techniques, such as the gradient descent algorithm, to find the optimal values of the coefficients $\beta$. Once the coefficients are obtained, the logistic regression model can be used to predict the probability that a new sample X belongs to the positive class ($Y = 1$) using the following formula:

$$P(Y = 1|X) = 1/(1 + \exp(-\beta_0 - \beta_1 X_1 - \beta_2 X_2 - \ldots - \beta_n * X_n)) \quad (2)$$

where exp is the exponential function. This formula represents the logistic curve, which maps the input variables $X$ to the probability of belonging to the positive class. However, during the training of the model, it was observed that it did not converge to a satisfactory result.

Another tested algorithm is the SVM algorithm, a supervised machine learning method used for classification and regression. Its objective is to find a hyperplane in a feature space that optimally separates the data of different classes (Kurani et al., 2023). The optimization of SVM involves finding the optimal values of $w$ and $b$.

$$f(x) = w \cdot x + b \quad (3)$$

represents the equation of the hyperplane in SVM, where $x$ is a vector of input features. The variable $b$ is the bias term that affects the position of the hyperplane in space, $w$ and $b$ are parameters that need to be optimally adjusted during the SVM training process to achieve effective separation between classes.

The OLS algorithm was chosen as the most suitable by testing various algorithms and aiming for simplicity.

The OLS algorithm is a method used to find the best fitting line that minimizes the sum of squared differences between observed values and predicted values. It is primarily used in regression problems, where the goal is to establish a relationship between a dependent variable and one or more independent variables.

The algorithm works as follows:

Training data are collected, consisting of pairs of values $(X, y)$, where $X$ represents the independent variables and $y$ represents the dependent variable.

A linear regression model is established, a mathematical function that estimates the value of $y$ based on the values of $X$. The model has the form:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n, \quad (4)$$

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are the coefficients that need to be determined. The OLS algorithm is used to find the optimal values of the coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ that minimize the sum of squared differences between observed values and predicted values. This is achieved through mathematical calculations involving differentiation and solving a system of linear equations. Once the optimal coefficients are obtained, the linear regression model predicts new data. The values of the independent variables are inserted into the model, and the predicted value for the dependent variable is calculated.

To implement the model, the following is carried out:

Gather historical data on staff turnover and predictor variables such as work experience, salary, and job satisfaction. Clean the data by removing outliers or missing values, and transform variables if needed (e.g., encode categorical variables as dummy variables).

Split the data into training and testing sets and divide the data into two sets, one for training the model (training set) and another for evaluating its performance (testing set).

Use the training set to fit the OLS model. This involves estimating coefficients for each predictor variable, minimizing the sum of squared differences between observed values and predicted values.

Use the testing set to assess the performance of the fitted model. Calculate evaluation metrics such as mean squared error or R-squared to determine how well the model fits the testing data (Borzì et al., 2023).

Utilize the fitted model to make predictions of staff turnover on new data. Apply the predictor variables to the model and obtain corresponding predictions (Margarat et al., 2023).

Analyze the estimated coefficients of the model to determine the relative influence of each predictor variable on staff turnover. Communicate the results clearly and coherently, presenting conclusions and potential implications based on the OLS model.

## 4. Results

In Figure 4, the system dashboard is presented, where the relationship of the total number of employees is displayed using a pie chart. This representation reveals a total of 308 employees, with an average of 3 rotations out of a total of 338 rotations. Additionally, a list of hired employees is provided alongside these charts.

Figure 5 shows the application of the OLS, showing a particular trend. Although it is a basic algorithm, it stands out for its ease of implementation and adaptability to the specific problem of the project.

To assign the columns of our data source, we first excluded the categorical variables (representing the reasons for employee turnover) as they are not involved in the current model development. Consequently, we retained only quantifiable variables, such as time in months. The input data will consist of the months we want to predict the number of rotations stored in the variable $X$. The dependent variable, denoted as y, holds the number of rotations that occurred each month. Given that this

**Figure 4**
**Dashboard showing employees and their staff turnover**

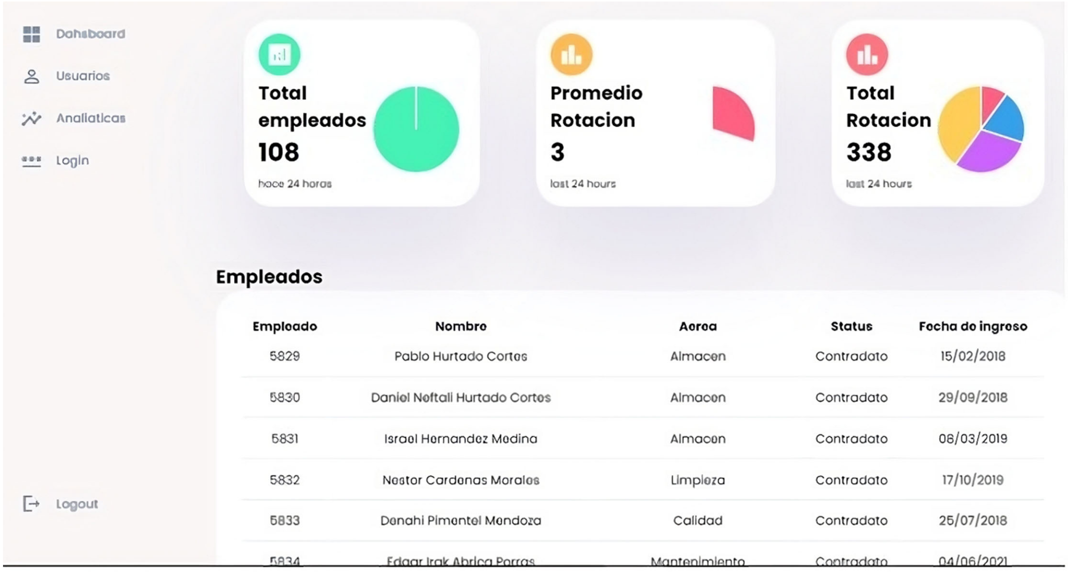
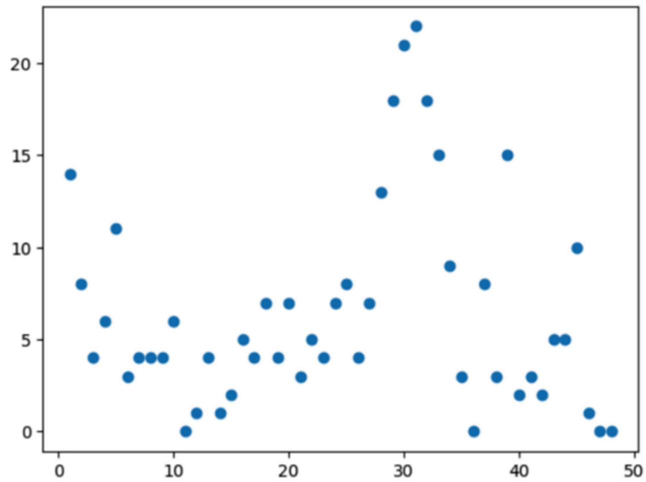



**Figure 5**
**Assignment to variables the columns time and staff turnover**




dependent variable is time dependent, we selected these two variables for model training as they are essential in capturing the temporal relationship and predicting rotation patterns accurately. Employee turnover can be voluntary or involuntary. Voluntary turnover occurs when employees leave the company for various reasons, such as better job opportunities, job dissatisfaction, lack of professional development, conflicts with the organizational culture, and other personal motives. On the other hand, involuntary turnover happens when the company ends the employment relationship with an employee due to poor performance, restructuring, or layoffs due to economic reasons.

Up to this point, several predictions were generated by adjusting biases and weights, which the TensorFlow library internally performs during the model training in the previous steps.

In the initial tests of the first experimental models, the following results were obtained:

As shown in Figure 6, for month 50, the results obtained were not realistic compared to the behavior of our data. Therefore, we discarded that option as the final model. To calibrate the model for better results, we decided to change the number of neurons used, so we opted to use 64 nodes. The predictions also improved as the model improved, as it exhibited similar behavior to the data feeding into the model. However, as the month moved further from the data, the prediction dropped to zero. Therefore, this option was also discarded.

**Figure 6**
**Prediction month 90 first model**

```
month = 90
months = model.predict([month])
print('Prediccion de rotacion = ', months)

1/1 [==============================] - 0s 69ms/step
Prediccion de rotacion =  [[9.180946]]
```

For the final model, it was chosen to have one input layer, one hidden layer, and one output layer, each consisting of 20 densely connected nodes with a Rectified Linear Unit (RELU) activation function.

The RELU function takes an input value $x$ and returns the maximum value between x and zero. Mathematically, it can be defined as $f(x) = \max(0, x)$. The RELU function returns the same input value when the input is positive. For example, if $x = 2$, then $f(x) = 2$. This means that there is no change for positive values.

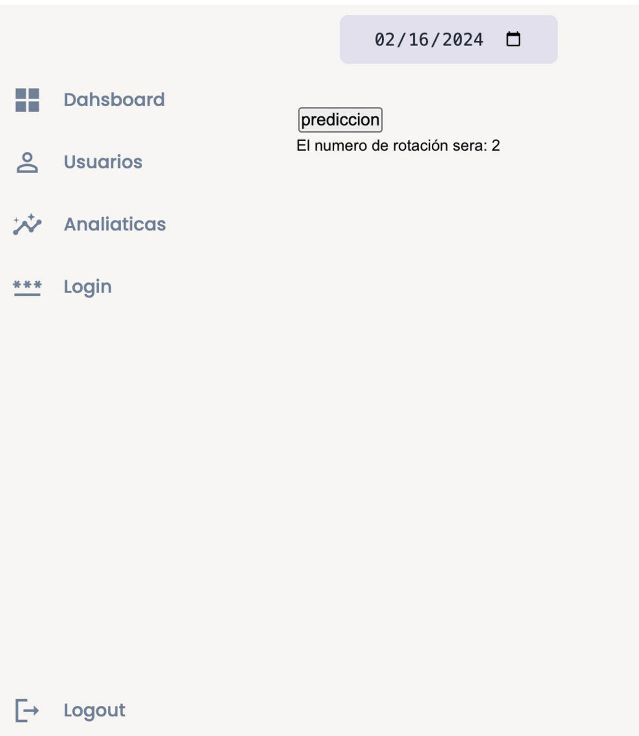
When the input is negative, the RELU function returns zero. For example, if $x = -2$, then $f(x) = 0$. This means that any negative value is turned into zero. The RELU function is known for its ability to introduce non-linearities in neural networks. Applying RELU as an activation function in a neural network allows the network to learn non-linear relationships and capture complex patterns in the data. In addition to its simplicity and computational efficiency, RELU addresses the issue of vanishing gradients that can occur in other activation functions like the sigmoid function. This means that RELU helps prevent training problems in deep neural networks by facilitating the flow of gradients during the backpropagation process.

The prediction of month 90 with the result 9.180946 corresponds to the results of the first model where the SVMs algorithm was generated, one of the discarded algorithms. When generating the new model (OLS) with 50 months, the prediction was 2.82727253, and using 90 months, the result was 5.8385496. As we can see, it now closely resembles the behavior in our input data, leading to the decision to adjust. Additionally, the use of the RELU activation function significantly improved the accuracy of the data prediction.

Lastly, the model is saved under "staff_turnover.h5". The Keras model is also converted to a TensorFlow format, typically saved as JSON. It is also instructed to create a folder named "output_folder".

Finally, in Figure 7, the analytics dashboard, the predictions for the model can be generated by simply clicking on the calendar icon, selecting a specific date, and then clicking on the "Prediction" button. This action will give us the predicted number of rotations for the selected date. The system utilizes the data and information from the trained model to perform these predictions accurately. This feature allows the user to conveniently access and use the predictive capabilities of the AI model for workforce rotation forecasting. It simplifies obtaining valuable insights about future employee turnover, empowering human resources personnel and decision-makers with helpful information to address workforce management and talent retention strategies. Moreover, the interface is designed to be user-friendly, ensuring that users with varying technical expertise can effortlessly use this predictive functionality, enhancing the system's overall usability and practicality.

**Figure 7**
**Analytics dashboard**




## 5. Conclusion

The turnover of personnel in the industry can be influenced by various factors beyond trends in the unemployment rate, the economically active population, and underemployment.

As the labor market strengthens and more jobs are created, workers may have more options and seek opportunities that better align with their needs and aspirations. This can result in higher turnover rates. Another possible cause is that workers may seek jobs that offer higher wages or better benefits, leading to personnel turnover in search of improved working conditions. Company culture and working conditions can also influence personnel turnover. Workers may switch jobs if companies do not provide a favorable work environment, development opportunities, or adequate recognition. Technological advancements and industry changes may require new skills and knowledge. If workers are not adequately trained or companies do not offer training opportunities, there may be higher personnel turnover as individuals seek jobs requiring in-demand skills.

These factors interact in complex ways and can vary in each industrial and geographic context. Therefore, conducting a comprehensive analysis to fully understand the specific causes of personnel turnover in the industry is essential.

The issue of employment and employee turnover is a relevant concern in Mexico and many other countries. High employee turnover can significantly impact organizations, affecting their productivity, efficiency, and financial outcomes. It is possible to develop a system for predicting employee turnover rates using data analysis and machine learning techniques. By identifying and analyzing significant variables, it is feasible to build a predictive model that estimates the risk of employee turnover at different times. Taking a data-driven and analytical approach, companies can make more informed decisions regarding addressing employee turnover challenges. By identifying variables that influence turnover and anticipating potential trends, organizations can implement effective talent retention strategies and improve the work environment, thus reducing turnover rates and enhancing employee stability.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## References

Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. *Journal of Cleaner Production*, *289*. https://doi.org/10.1016/j.jclepro.2021.125834

Appelbaum, E., & Batt, R. L. (1994). *The new American workplace: Transforming work systems in the United States*. USA: Cornell University Press.

Ballesteros, R., Castro, C. G., Charles-Leija, H., Guerrero, I., & Toledo, M. (2023). *Factor Wellbeing: El bienestar como ventaja competitiva*. (in Spanish). Mexico: Lid Editorial Mexicana Sa De Cv.

Borzì, G. R., Bonanno, E., Cavalli, N., Pace, M., Zirone, L., D'Anna, A., . . . , & Marino, C. (2023). A mathematical model to identify the output variations of TrueBeam linear accelerators using machine performance check evaluations. *Journal of Instrumentation*, *18*(07). https://doi.org/10.1088/1748-0221/18/07/T07002

Castañeda, E. (2023). *Rotación de personal estaría llegando hasta el 120% en industrias bajacalifornianas, Tijuana.* Retrieved from: https://www.elimparcial.com/tijuana/tijuana/Rotacion-de-personal-estaria-llegando-hasta-el-120-en-industrias-bajacalifornianas-20230205-0009.html

Davenport, T. H. (2018). *The AI advantage: How to put the artificial intelligence revolution to work*. USA: MIT Press.

Forster, P., & Kahan, R. (2023). *Indeed*. Retrieved from: https://mx.indeed.com/.

Hadijah, H. S. (2023). Implementation of talent management as a strategy for achieving company competitive advantage. *International Journal of Artificial Intelligence Research*, *6*(1). http://ijair.id/index.php/ijair/article/view/379

Hidayat, T., & Masdupi, E. (2023). The influence of transformational leadership, psychological capital and organizational support for work engagement. In *Ninth Padang International Conference on Economics Education, Economics, Business and Management, Accounting and Entrepreneurship*, 487–502. https://doi.org/10.2991/978-94-6463-158-6_44

Hodeghatta, U. R., & Nayak, U. (2023). Python programming for analytics. In U. R. Hodeghatta & U. Nayak (Eds.), *Practical business analytics using R and Python: Solve business problems using a data-driven approach* (pp. 635–672). Apress.

INEGI. (2023). *Encuesta Nacional de Ocupación y Empleo (ENOE), población de 15 años y más de edad*. (in Spanish). Retrieved from: https://www.inegi.org.mx/programas/enoe/

Kowssarie, S., & Mansour, S. (2023). *Employee turnover in warehouses: Exploring the role of internal marketing strategies in consulting firms*. Bachelor Thesis. Jönköping University.

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, *10*(1), 183–208. https://doi.org/10.1007/s40745-021-00344-x

Lawler, III, E. E. (2010). *Talent: Making people your competitive advantage*. USA: John Wiley & Sons.

Margarat, G. S., Kumar, S., Rajan, S., & Raj Mohan, B. (2023). Forecasting wind energy production using machine learning techniques. In *E3S Web of Conferences, 387*. https://doi.org/10.1051/e3sconf/202338701007

Mayorga, A., & Isabel, M. (2023). *Los aspectos socioeconómicos en el desempleo de la población económicamente activa del Ecuador*. (in Spanish). Bachelor's thesis, Universidad Técnica de Ambato.

Mia, M. A., Jibir, A., & Omeke, M. (2023). Employee turnover and the social outreach (breadth): International evidence from the microfinance industry. *Social Responsibility Journal*. https://doi.org/10.1108/SRJ-09-2022-0363

Navarrete, R. A., Cedillo, E. R., & Pérez, C. R. G. (2024). Estructura económica y empleo en México: Propuesta de análisis para políticas públicas. *Contaduría y Administración*, *69*(1).

Pattanayak, S. (2023). *Pro deep learning with TensorFlow 2.0: A mathematical approach to advanced artificial intelligence in Python*. Canada: Apress.

Pau, D. P., Naramo, T. A., & Dimbiniaina, M. (2023). Coding Mel spectrogram using Keras and Tensorflow for home appliances tiny classification. In *2023 IEEE International Conference on Consumer Electronics*, 1–5. https://doi.org/10.1109/ICCE56470.2023.10043378

Sonderegger, M. (2023). *Regression modeling for linguistic data*. USA: MIT Press.

Yadav, S., Jain, A., & Singh, D. (2018). Early prediction of employee attrition using data mining techniques. In *2018 IEEE 8th International Advance Computing Conference,* 349–354. https://doi.org/10.1109/IADCC.2018.8692137

Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. *IEEE Access*, *9,* 60447–60458. https://doi.org/10.1109/ACCESS.2021.3074559

**How to Cite:** Marquez, B. Y., Realyvásquez-Vargas, A., Lopez-Esparza, N., & Ramos, C. E. (2024). Application of Ordinary Least Squares Regression and Neural Networks in Predicting Employee Turnover in the Industry. *Archives of Advanced Engineering Science,* 2(1), 30–36, https://doi.org/10.47852/bonviewAAES32021326