# Ivan Huaman

## Performance_workers_University (1).pdf

My Files

My Files

Universidad Peruana Union

### Detalles del documento

Identificador de la entrega
**trn:oid:::29566:556622864**

Fecha de entrega
**13 feb 2026, 12:12 p.m. GMT-5**

Fecha de descarga
**13 feb 2026, 12:14 p.m. GMT-5**

Nombre del archivo
**Performance_workers_University (1).pdf**

Tamaño del archivo
**791.7 KB**

**15 páginas**

**7230 palabras**

**40.232 caracteres**

# 8% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

## Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Trabajos entregados
- Fuentes de Internet

## Grupos de coincidencias

**48** Sin cita o referencia 8%
Coincidencias sin una citación ni comillas en el texto

**0** Faltan citas 0%
Coincidencias que siguen siendo muy similar al material fuente

**0** Falta referencia 0%
Las coincidencias tienen comillas, pero no una citación correcta en el texto

**0** Con comillas y referencia 0%
Coincidencias de citación en el texto, pero sin comillas

## Fuentes principales

0% Fuentes de Internet
8% Publicaciones
0% Trabajos entregados (trabajos del estudiante)

## Marcas de integridad

**N.º de alerta de integridad para revisión**

**Caracteres reemplazados**
22 caracteres sospechosos en N.º de páginas
Las letras son intercambiadas por caracteres similares de otro alfabeto.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

## Grupos de coincidencias

**48** Sin cita o referencia 8%
Coincidencias sin una citación ni comillas en el texto

**0** Faltan citas 0%
Coincidencias que siguen siendo muy similar al material fuente

**0** Falta referencia 0%
Las coincidencias tienen comillas, pero no una citación correcta en el texto

**0** Con comillas y referencia 0%
Coincidencias de citación en el texto, pero sin comillas

## Fuentes principales

0% Fuentes de Internet
8% Publicaciones
0% Trabajos entregados (trabajos del estudiante)

## Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1 Publicación
Williams, Kane C.. "Metal Recovery and Recycling Using Machine Learning and Ed... <1%

2 Publicación
Benard O. Ogol, Evans Omondi, John Olukuru, Betsy Muriithi, Kennedy Senagi. "P... <1%

3 Publicación
Dr Sudip Chakraborty, Praval Pratap Singh, Jasmeen Kaur, Mohit Jakhar, Shihan V... <1%

4 Publicación
Ruoyao Wang, Yanyan Huang, Guoliang Zhang, Yi Yang, Qizhi Dong. "Optimizing ... <1%

5 Publicación
Kennedy C. Onyelowe, Shadi Hanandeh, Viroon Kamchoom, Ahmed M. Ebid et al. ... <1%

6 Publicación
Mariusz Ptak, Mariusz Sojka, Katarzyna Szyga-Pluta, Teerachai Amnuaylojaroen. "... <1%

7 Publicación
Thompson, Chelsie C.. "A Machine Learning Approach to Evaluate Ransomware a... <1%

8 Publicación
Jorge Sánchez-Garcés, Juan J. Soria, Josué E. Turpo-Chaparro, Himer Avila-George, ... <1%

9 Publicación
Konstantinos Bisiotis, Dimitris Christopoulos, George Tzougas. "Forecasting Carb... <1%

10 Publicación
Md Khalid Hossen, Yan-Tsung Peng, Asher Shao, Meng Chang Chen. "An ODE base... <1%

11 Publicación
Rungroad Suppakul, Duy Tan Tran, Suraparb Keawsawasvong, Peem Nuaklong, S... <1%

12 Publicación
Wilfredo Meza Cuba, Juan Carlos Huaman Alfaro, Hasnain Iftikhar, Javier Linkolk ... <1%

13 Publicación
"AI Technologies for Information Systems and Management Science", Springer Sc... <1%

14 Publicación
Godwin Ekedegwa, Anthony Inalegwu. "Performance Analysis of Two Hybrid Opti... <1%

15 Publicación
Chien-Chang Lin, Eddie S.J. Cheng, Anna Y.Q. Huang, Stephen J.H. Yang. "DNA of L... <1%

16 Publicación
V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challeng... <1%

17 Publicación
Dai, Hanjun. "Learning Neural Algorithms with Graph Structures.", Georgia Instit... <1%

18 Publicación
Yiyang Huang, Zhizhuo He, Yuchen Qin, Yichen Lu, Kaida Chen. "Optimizing office... <1%

19 Publicación
Hasnain Iftikhar, Atef F. Hashem, Moiz Qureshi, Paulo Canas Rodrigues. "Clinical ... <1%

20 Publicación
Ana Maria Căvescu, Nirvana Popescu. "Predictive Analytics in Human Resources ... <1%

21 Publicación
Md. Taufeeq Uddin, Md Monirul Islam. "Extremely randomized trees for Wi-Fi fing... <1%

22 Publicación
Beatriz Dias, Andre Gloria, Pedro Sebastiao. "Prediction of Link Quality for IoT Clo... <1%

23 Publicación
Chen, Zijie. "Physics-Informed and Data-Driven Modeling for Radiative Transport i... <1%

24 Publicación
Faridoon Khan, Hasnain Iftikhar, Imran Khan, Paulo Canas Rodrigues, Abdulmaje... <1%

| | | |
|---|---|---|
| 25 | Publicación | Jiayue Xu, Le Xuan, Cong Li, Mengxue Zhang, Xuhui Wang. "Exploring the Impact ... <1% |
| 26 | Publicación | Roy, Poulomee. "Analyzing Equity in Resource Allocation for Wildfire Containmen... <1% |
| 27 | Publicación | Junfeng Zhang, Yaxin Shen. "Model Modeling the Spatiotemporal Vitality of a Hist... <1% |
| 28 | Publicación | Paolo Ferro, Harinadh Vemanaboina, Chander Prakash. "Computational Techniqu... <1% |
| 29 | Publicación | Cenk Temizel, Salih Tutun, Celal Hakan Canbaz, Ekrem Alagoz et al. "Artificial Inte... <1% |
| 30 | Publicación | Ong, Yang. "Transparent and Equitable Pay Structures: Using Machine Learning a... <1% |
| 31 | Publicación | Paul D. McNicholas, Peter A. Tait. "Data Science with Julia", CRC Press, 2019 <1% |
| 32 | Publicación | S.J. Xavier Savarimuthu, Sivakannan Subramani, Alex Noel Joseph Raj. "Artificial I... <1% |

MDPI

*Article*

# Classification model for job performance supported by neural networks for University institutions.

12 **Iván Huamán Fernández [1], Esteban Tocto-Cano [1], Hasnain Iftikhar [1] and Javier Linkolk López-Gonzales[1,‡]**

[1] Escuela de Posgrado, Universidad Peruana Unión, Lima, Peru
* Correspondence: javierlinkolk@gmail.com

## 1. Introduction

A group of collaborators, composed of psychologists and administrative staff, after reviewing ways to measure job performance, decided to implement the methodology based on functions and general and specific competencies developed by Martha Alles [1]. Based on this methodology, a web-based software was implemented [2] to evaluate performance. The performance evaluation is conducted once a year and involves teaching, non-teaching, administrative, and management staff [3].

Competency-based job performance is a modern method of human talent management that not only measures a worker's achievements, but also the way in which he or she achieves them, i.e., the skills he or she uses to reach his or her goals. Martha Alles, a leader in human resources management in Latin America, proposes a structured approach that directly links the skills model to performance evaluation, accounting for both technical and behavioral skills.

According to Alles, competencies are a set of knowledge, skills, and attitudes that are observable and affect job performance. They can be measured with objective tests. His approach uses tools such as the assessment center, 360°, 180°, and 90° assessment, competency interviews, and behavioral scales to ensure a comprehensive examination of individual and group performance [4].

According to Martha Alles' methodology, performance is also defined as the observed display of abilities, behaviors, and knowledge aligned with the role and strategy. The model consists of a list of competencies, sets of competencies, a job profile with clearly defined functions, and individual development plans. [5]

Numerous international organizations have adopted Martha Alles's methodology. In Colombia, a competency-based microbusiness management model has been implemented using Alles' methodology to assess and develop employee competencies [6]. In the academic field, 360-degree competency-based evaluation has also been used to improve student performance in educational institutions [7].

In different countries, Martha Alles is not explicitly mentioned, but performance evaluations based on competencies aligned with the methodology she proposes have been implemented.

La National Distance Education University UNED (Spain) and the Catholic University of Santiago de Guayaquil (Ecuador) they have designed a competency-based training model for teachers and education professionals with the aim of evaluating competencies and their great importance in the teaching process the results show that teachers between the ages of 25 and 35 prefer to value planning, communication, evaluation, methodology, digital, and tutoring skills, while those over 55 emphasize the importance of digital and innovation skills. [8]

In Peru, studies were conducted using a non-experimental, cross-sectional design with a descriptive-propositional approach. The results showed deficiencies in specific processes related to personnel selection and hiring, as well as a lack of incentives affecting staff development and motivation. When the model was implemented in human resources, it proved effective in improving work performance. [9]

1 **Citation:** Lastname, F.; Lastname, F.; Lastname, F. Classification model for
1 job performance. *Algorithms* **2022**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

8 **Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## 2. Related work

Performance is a key factor in a company's success, and research on artificial intelligence has been conducted in various countries. This is illustrated in the following study, which uses traditional methods such as KPIs and 360° evaluations, although these methods have limitations in capturing the complex aspects of job performance. To overcome these limitations, machine learning-based techniques allow for the processing of large amounts of multidimensional data, improving predictive capabilities. BP neural network models have demonstrated the capability to detect nonlinear trends in labor variables, leading to outcomes that are both more precise and scalable [10,11].

In any part of the world, without a doubt, a company's success lies in the good performance of its employees. Thanks to the good performance of the employees, productivity is improved. Identifying underperforming employees is a strategic step toward implementing improvements. In this regard, they use artificial intelligence to perform a comparative analysis of different algorithms to classify performance based on various factors. [12]

In matters related to employee behavior, he made predictions about staff turnover using machine learning algorithms, among which Logistic Regression, Decision Tree, Random Forest, Neural Networks, and Support Vector Machine stood out for their effectiveness in predicting employee turnover, for which employee data was used. It is particularly noteworthy that the Random Forest algorithm is the most widely used technique because it achieved the highest index and accuracy. [13]

Artificial intelligence is playing an important role in the management of human talent in companies, from recruitment processes and employee retention to improving worker performance. Human talent management is a strategic area in companies that want to achieve better performance in the key factor: the worker. In this research, predictive analysis is performed in human resource management, examining machine learning algorithms including XGBoost, SVM, random forest, and linear regression, with the result that using this technology can automate human resource processes. However, it does not ignore important concerns regarding data privacy, resistance to change, and biases in algorithms, for which it is important to adopt theoretical frameworks regarding the use of AI [14].

## 3. Methodological Framework

In this study, the soft systems methodology (SSM) is selected because recent literature highlights that SSM can function as a primary methodological framework, not just as a complement [15], as it has a clear structure, applicable stages, and analytical traceability, which makes it an organizational learning methodology that allows for comparing, analyzing, and negotiating changes[16].

Soft Systems Methodology (SSM) facilitates organizational learning, integration of perspectives, and identification of feasible and socially acceptable changes (see Figure 1). It is particularly effective in educational, organizational, and public policy contexts that are characterized by high complexity and uncertainty[17].

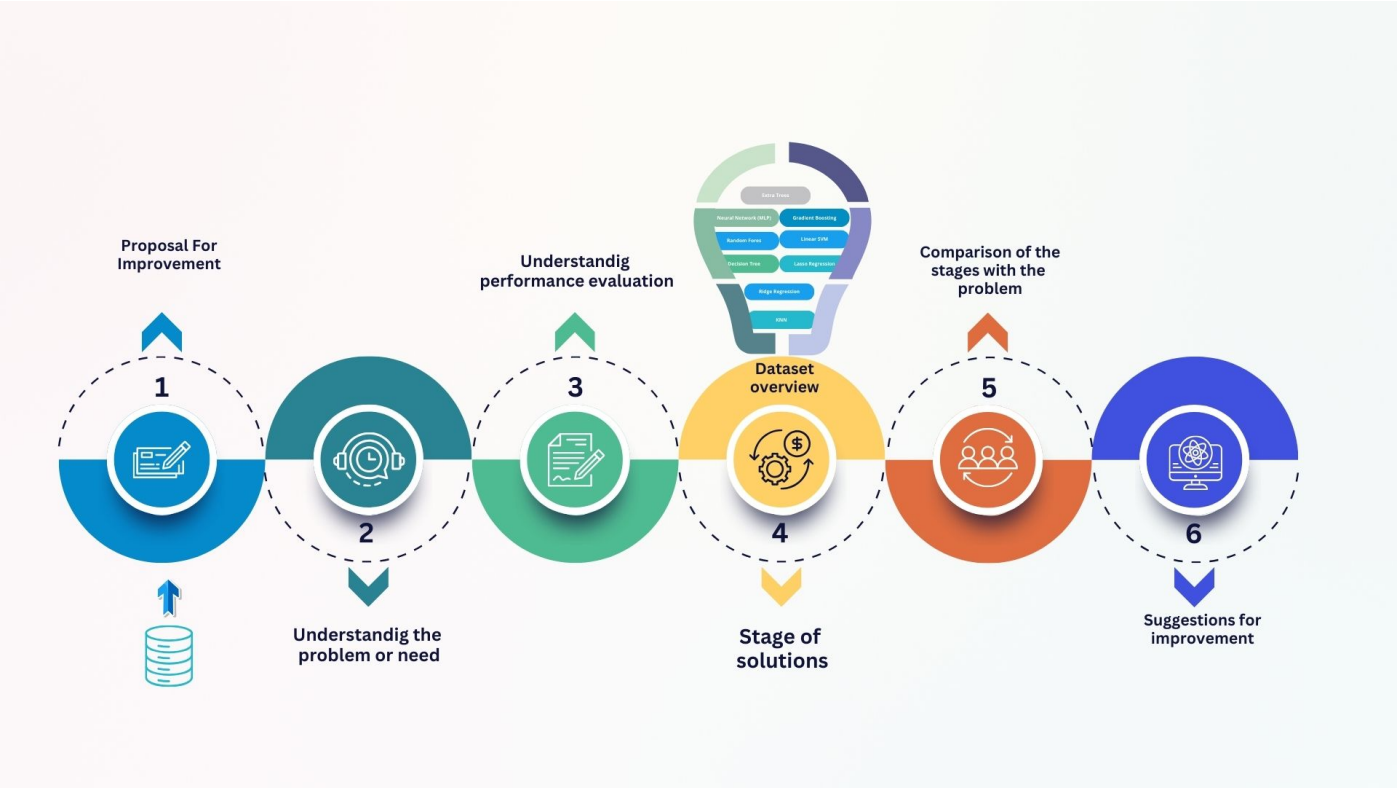



**Figure 1.** Soft System Methodology

### *3.1. Proposal for improvement*

This research proposes to evaluate performance in order to detect the behavior of employees in tasks that are relevant to the organization. However, although this process generates a wealth of data on university talent, very few subsequent analyses are produced based on the data collected. Taking advantage of the results of the performance evaluation process, giving value to the data obtained, can bring great benefits to the institution and, consequently, to the country in general. One of the clear results of the subsequent analysis of this data is the obtaining of relevant information to focus actions that tend to improve the performance of workers, establishing effective training plans, whose results can be quantified over time. Another important result that could be obtained from a subsequent analysis of performance evaluations has to do with university staff hiring and termination policies. Ultimately, these are also judged through employee performance, year after year. Another important result, now from an organizational point of view, is that a university human talent management policy, using instruments that are as objective as possible, guarantees an appropriate working environment, as employees clearly know how they will be judged for their work, without finding that their efforts are not appropriately rewarded when they are meritorious, but also, conversely, understanding that if the expected standards are not met, there will surely be undesirable consequences [18] [19]

### *3.2. Understanding the problem or need*

The performance management process is an integrated cycle that includes performance evaluation planning, supervision, communication, application of the evaluation tool, processing of the results obtained, and actions that can be taken to that effect. It is a complex process that includes reviewing the duties and responsibilities of each position, setting expectations, assessing performance, and determining objectives and goals. It also includes advising employees on performance evaluation and implementing an action plan.

At the University of Peru, the performance evaluation process is based on the text entitled “Competency-based performance: 360° evaluation.” [4]

The Peruvian government, through the Ministry of Education, seeks to ensure that students and educational institutions achieve quality learning, to provide higher education that fosters an environment conducive to national development and competitiveness. To this end, efforts are being made at the university level to strengthen the capacity of faculty and administrative staff. Two entities have been established to contribute to these objectives.

### *3.3. Understanding performance evaluation*

We use CATWOE to clarify, analyze, and define stakeholder perspectives and structure solutions. The analysis consists of breaking down the following. [20] [21] [22]

Customers:

- The Institution.
- University students.
- Society.

Actors:

- Teachers.
- Administrative staff.

Transformation:

- Improve the performance evaluation system, supported by machine learning solutions.

Worldview:

- The best teaching performance is an indispensable factor in an institution that guaranties high-quality education.

Owner:

- University council.
- Senior institutional management

Environment/Constraints:

- Educational regulations.
- Budget
- Organizational culture

### *3.4. Stages of solutions*

The dataset used in this study comprises a total of 18,947 samples and 14 features, occupying 9.63 MB of memory. It includes both categorical (8 columns) and numerical (6 columns) variables. The dataset is well-structured and complete, with no missing values across any feature, resulting in an overall data completeness of 100%. Only a negligible fraction of duplicate records (22 rows, 0.12%) was identified, indicating excellent data quality.

The primary outcome variable for regression tasks is the `Average score`, which ranges from 16.00 to 100.00, with a mean of 87.78, a standard deviation of 13.39, and a median of 92.00. The distribution is left-skewed (skewness = –1.486) and heavy-tailed (kurtosis = 2.486), with 2.53% of values flagged as potential outliers. To support classification tasks, a categorical performance indicator (`Performance_Class`) was derived from the average score using three thresholds: Low ($\leq$85.00), Medium ($\leq$97.00), and High (>97.00). The resulting class distribution is relatively balanced, with 6,426, 6,766, and 5,755 samples in the Low, Medium, and High categories, respectively.

The dataset covers the years 2012–2024 and includes demographic, functional, and organizational attributes. Key categorical variables are `LOCAL` (three locations, dominated by Lime with 13,431 cases), `Gender` (male: 10,371; female: 8,576), `Department` (96 unique values), `Position` (1,105 unique roles), and `Evaluator type` (four groups, with immediate supervisors representing the most significant share at 8,604 evaluations). Several columns

exhibit high cardinality, such as `Document number`, `Department`, `Birthdate`, and `Position`, which were further processed through feature engineering.

The numerical predictors are diverse in scale and content. For instance, `Age` spans from 18 to 87 years (mean = 43.72), while performance-related components such as `By function`, `Functional Comp.`, and `Organizational Comp.` show mean values of 87.48, 86.59, and 90.92, respectively, each distributed within the 6–100 range. The `Year` variable (2012–2024) was leveraged to construct tenure-related features.

Extensive preprocessing was conducted to ensure data readiness. Missing values were handled systematically (median imputation for numerical variables and mode imputation for categorical ones), and categorical features were encoded using `LabelEncoder`. Feature engineering steps included deriving `Age_Group` from birthdates, creating a combined `Dept_Position` feature with 1,625 unique department–position pairs, and computing a `Composite_Score` that synthesizes multiple performance indicators. Additionally, an `Experience_Proxy` variable was developed to approximate tenure based on the year of evaluation. To prepare the features for downstream modeling, scaling was applied using `StandardScaler`, ensuring consistent representation across regression and classification models.

Given the dataset's size, a random subset of 10,000 samples was used for training computationally intensive algorithms such as neural networks and support vector machines. For evaluation, both 80/20 and 50/50 train–test splits were used to assess model robustness across different data partitions [23].

Thus, the dataset is comprehensive, well-curated, and high-quality, providing a strong foundation for both regression and classification analyzes. The balanced target distributions, engineered features, and preprocessing steps enhance its suitability for applying a wide range of machine learning models[24].

### *3.5. Preprocessing Recommendations*

For categorical encoding, a mixed strategy is advised. Low-cardinality variables such as `LOCAL`, `Gender`, `Scale`, and `Evaluator type` can be efficiently processed using either label encoding or one-hot encoding. In contrast, high-cardinality features such as `Document number`, `Department`, `Birthdate`, and `Position` are better suited for target encoding to reduce dimensionality while retaining predictive information. All numerical features were standardized using the `StandardScaler` to ensure consistency across models.

Given the dataset size of 18,947 samples, a wide range of machine learning algorithms can be considered appropriate. In particular, regression models such as Ridge and Lasso regression provide robust baselines. Tree-based models, including Decision Tree, Random Forest, Extra Trees, and Gradient Boosting, are expected to capture nonlinear relationships effectively. Instance-based and kernel-based approaches, such as K-nearest neighbors (KNN) and support vector machines with radial basis function (RBF) kernels, are also suitable, particularly when applied with a 50/50 train–test evaluation strategy. Finally, neural networks (multi-layer perceptrons) are viable candidates for capturing complex patterns, though training efficiency considerations warrant the use of random sampling for computationally intensive runs.

### *3.6. Predictive Modeling*

#### Ridge Regression

Ridge Regression is a linear regression method with $L_2$ regularization that penalizes large coefficients to reduce overfitting[25]. Given training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, the Ridge optimization problem is:

$$\hat{\beta}_{\text{ridge}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \beta)^2 + \lambda \|\beta\|_2^2 \right\},$$

where $\lambda \geq 0$ controls the strength of regularization.

The closed-form solution is:

$$\hat{\beta}_{\text{ridge}} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}.$$

Lasso Regression

Lasso Regression introduces an $L_1$ penalty, promoting sparsity in the coefficient vector:

$$\hat{\beta}_{\text{lasso}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \beta)^2 + \lambda \|\beta\|_1 \right\}.$$

The $L_1$ penalty encourages some coefficients to become exactly zero, effectively performing feature selection[25].

Linear Support Vector Regression (Linear SVR)

Support Vector Regression builds a linear function with $\epsilon$-insensitive loss [26]:

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b.$$

The optimization is:

$$\min_{\mathbf{w}, b, \xi, \xi^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

subject to:

$$\begin{aligned} y_i - (\mathbf{w}^\top \mathbf{x}_i + b) &\leq \epsilon + \xi_i, \\ (\mathbf{w}^\top \mathbf{x}_i + b) - y_i &\leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* &\geq 0. \end{aligned}$$

Gradient Boosting

Gradient Boosting builds an ensemble of weak learners (usually decision trees) by sequentially minimizing a loss function. Starting with an initial model:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma),$$

each boosting stage adds a new tree $h_m(x)$:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x),$$

where $\eta$ is the learning rate and:

$$h_m(x_i) \approx -\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}.$$

Extra Trees

Extra Trees (Extremely Randomized Trees) are an ensemble of decision trees [27] where both:

- features are selected randomly,
- split thresholds are chosen at random.

For a regression tree, the prediction at a leaf is the mean of target values:

$$\hat{y}_{\text{leaf}} = \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} y_i.$$

The ensemble prediction is:

$$\hat{y} = \frac{1}{M} \sum_{m=1}^{M} T_m(x),$$

where $T_m$ is the $m$-th randomized tree.

Neural Network (MLP)

A Multilayer Perceptron with one hidden layer computes:

$$\hat{y} = f(\mathbf{x}) = \mathbf{w}_2^\top \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + b_2,$$

where:

- $\mathbf{W}_1$ and $\mathbf{w}_2$ are weight matrices,
- $\sigma(\cdot)$ is an activation function (e.g., ReLU, tanh),
- $\mathbf{b}_1$, $b_2$ are bias terms.

The network is trained by minimizing:

$$\min_{\Theta} \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i; \Theta))^2 + \lambda \Omega(\Theta),$$

where $\Theta$ is the set of all parameters.

Random Forest

A Random Forest averages predictions from many independently built decision trees using bootstrap samples. The model prediction is:

$$\hat{y}(x) = \frac{1}{M} \sum_{m=1}^{M} T_m(x),$$

where each tree $T_m$ is trained on a bootstrap subset, and feature subsets are randomly selected at each split.

Each tree is built by minimizing the node impurity:

$$\text{MSE} = \frac{1}{|S|} \sum_{i \in S} (y_i - \bar{y}_S)^2.$$

Decision Tree

A regression tree recursively partitions the feature space to minimize the mean squared error at each split. For a node $S$, the optimal split $(j, t)$ is:

$$(j^*, t^*) = \arg\min_{j,t} \left( \frac{|S_L|}{|S|} \text{MSE}(S_L) + \frac{|S_R|}{|S|} \text{MSE}(S_R) \right),$$

where $S_L$ and $S_R$ are left and right partitions.

The output at a leaf is:

$$\hat{y} = \bar{y}_S.$$

k-Nearest Neighbors (KNN)

KNN predicts values by averaging the $k$ nearest neighbors of a query point $x$ in feature space.

Define the neighbor set:

$$\mathcal{N}_k(x) = \{ i \mid x_i \text{ is among the } k \text{ closest points to } x \}.$$

The prediction is:

$$\hat{y}(x) = \frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} y_i.$$

Distance is typically measured using Euclidean distance:

$$d(x, x_i) = \|\mathbf{x} - \mathbf{x}_i\|_2.$$

### *3.7. Evaluation*

In this paper, model evaluation metrics quantify how well a predictive model approximates the relationship between input features and the target variable. The most commonly used metrics include the coefficient of determination ($R^2$), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Each metric captures different aspects of prediction quality and provides complementary information about model accuracy and error distribution.

#### 3.7.1. Coefficient of Determination

The coefficient of determination, denoted by $R^2$, measures the proportion of variance in the dependent variable that is explained by the model. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

where $y_i$ represents the true values, $\hat{y}_i$ the predicted values, and $\bar{y}$ the mean of the observed values.

The numerator,

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

is the sum of squared errors, while the denominator,

$$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2,$$

is the total sum of squares.

An $R^2$ value of 1 indicates perfect prediction, whereas a value of 0 implies that the model performs no better than predicting the sample mean. Negative values may occur when the model performs worse than this baseline.

#### 3.7.2. Mean Absolute Error (MAE)

Mean Absolute Error calculates the average magnitude of the errors without considering their direction. It is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|.$$

MAE provides a straightforward, interpretable measure of prediction accuracy in the same units as the target variable. Because it uses absolute differences, MAE treats all deviations linearly and is less sensitive to outliers compared to squared-error-based metrics.

#### 3.7.3. Root Mean Squared Error (RMSE)

Root Mean Squared Error measures the square root of the average squared differences between actual and predicted values:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

RMSE penalizes larger errors more heavily due to squaring, making it particularly sensitive to outliers or large deviations. Like MAE, RMSE is expressed in the original units of the target variable, making it interpretable and easy to compare across models.

*Interpretation and Comparison of Metrics*

- $R^2$**:** Measures explained variance. Higher values indicate better fit, but it does not capture error magnitude directly.
- **MAE:** Represents the average prediction error. It is robust to outliers and easy to interpret.
- **RMSE:** Similar to MAE but penalizes large errors more. A lower RMSE indicates better predictive accuracy and more stable predictions.

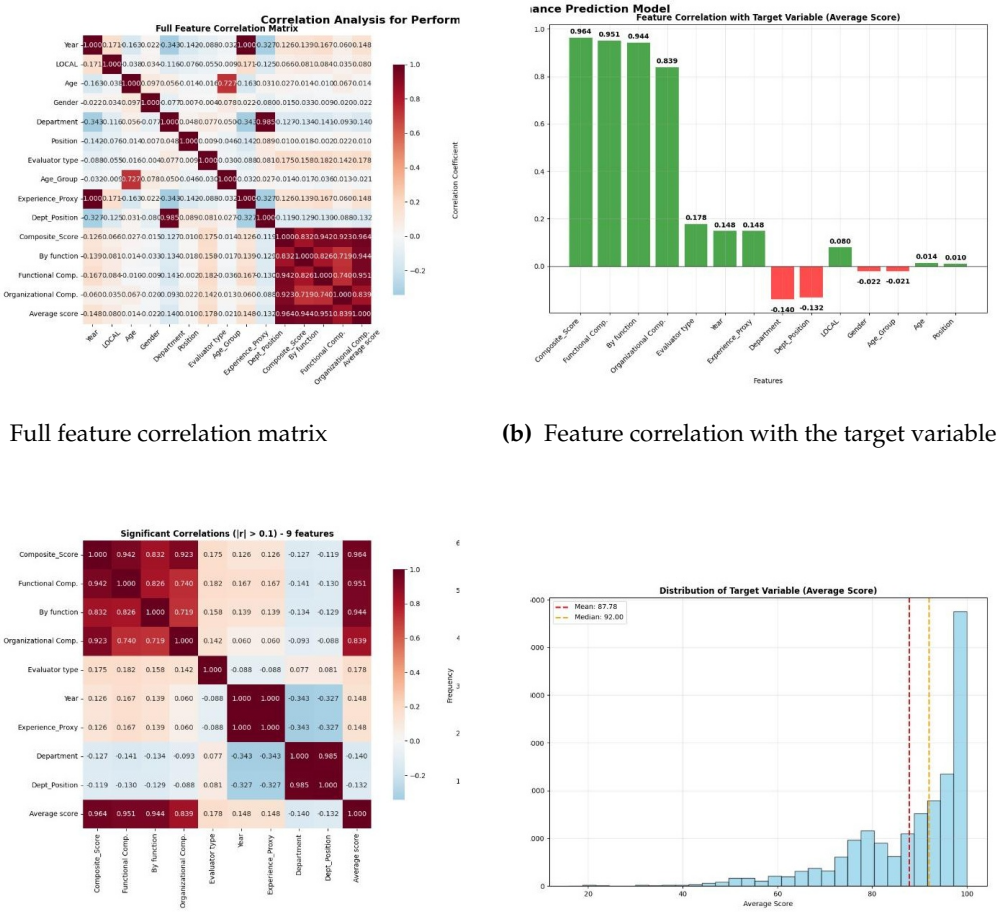
Together, these metrics provide a comprehensive evaluation of regression models, balancing goodness-of-fit ($R^2$) with direct measurements of prediction error (MAE and RMSE).

## 4. Results

Figure 2 presents a comprehensive correlation-based diagnostic of the performance prediction framework, offering insights into both feature interrelationships and their explanatory power with respect to the target variable (Average Score). However, the full feature correlation matrix (Figure 2(a)) reveals a clear structural pattern in the data. Performance-related constructs—namely Composite Score, Functional Component, By-Function Component, and Organizational Component—exhibit very strong positive correlations with one another, indicating that these dimensions capture closely related aspects of employee performance evaluation. In contrast, most demographic and administrative variables (e.g., Gender, Age, Position, and Local) show weak or near-zero correlations with both performance metrics and the target variable, suggesting limited direct influence on the final evaluation outcomes. Notably, Department and Department-Position display moderate negative correlations with performance-related measures, implying possible structural or organizational heterogeneity across units. On the other hand, the feature-target correlation analysis (Figure 2(b)) further clarifies the relative importance of predictors. Composite Score demonstrates the strongest association with the Average Score, followed closely by Functional Component and By-Function Component, all with correlation coefficients exceeding 0.94. This confirms that the Average Score is largely driven by aggregated and function-based performance measures. The Organizational Component also shows a strong positive relationship, though slightly weaker than the other core performance indicators. In contrast, evaluator characteristics (Evaluator Type), temporal factors (Year), and experience-related proxies contribute only marginally. Departmental variables show small but negative correlations, suggesting that institutional placement may influence scoring patterns rather than individual performance.

In the same way, Figure 2(c) focuses on statistically meaningful correlations ($|r| > 0.1$), highlighting a reduced set of influential variables. The heatmap underscores the dominance of performance composites, which maintain high internal consistency and strong links to the Average Score. The near-perfect correlation between Department and Department-Position indicates redundancy and suggests potential multicollinearity that should be addressed in predictive modeling. Meanwhile, the modest correlations of Year and Experience Proxy with the target suggest gradual temporal or tenure-related effects, but not primary drivers of performance outcomes. Finally, the distribution of the target variable (Figure 2(d))

illustrates a pronounced left-skewed pattern, with scores heavily concentrated at the upper end of the scale. The mean ($\approx$ 87.8) is below the median ($\approx$ 92.0), indicating a ceiling effect in which a large proportion of individuals receive high performance ratings. This distributional characteristic has important modeling implications, as it may reduce variance in the dependent variable and bias predictive accuracy if not properly handled. Therefore, the figures collectively demonstrate that composite and functional evaluation metrics overwhelmingly influences the performance prediction model, while demographic and organizational variables play a secondary, contextual role.




**(a)** Full feature correlation matrix

**(b)** Feature correlation with the target variable

**(c)** Significant correlations ($|r| > 0.1$)

**(d)** Distribution of the average score

**Figure 2.** Correlation analysis for the performance prediction model. (a) Full feature correlation matrix. (b) Correlation between individual features and the target variable (average score). (c) Heatmap of statistically significant correlations ($|r| > 0.1$). (d) Distribution of the target variable (average score).

The results obtained from Block 1, which used a 5-fold cross-validation scheme with an 80/20 train-test split, demonstrate consistently strong performance across all evaluated regression models. Among these, Ridge Regression emerged as the top-performing model, achieving the highest Test $R^2$ of 99.96% and an equally strong cross-validation mean of 99.95% with a very small standard deviation of $\pm 0.02$, highlighting both accuracy and exceptional stability. Lasso Regression and the Linear SVM also performed nearly identically to Ridge Regression, with Test $R^2$ scores of 99.95% and similarly low cross-validation variability, indicating that linear relationships dominate the dataset and that $L_1/L_2$ regularization or margin-based linear approaches are well-suited for this task.

Nonlinear ensemble models such as Gradient Boosting and Extra Trees showed slightly lower but still excellent performance, with Test $R^2$ scores of 99.93% and cross-validation results of 99.91% ($\pm$0.04) and 99.85% ($\pm$0.14), respectively. The slightly higher standard deviations for these models indicate increased sensitivity to the specific training subsets used during cross-validation, reflecting the inherent variance in tree-based approaches.

The Neural Network (MLP) achieved a Test $R^2$ of 99.79% with a perfect cross-validation standard deviation of $\pm$0.00, suggesting consistent convergence across folds. However, its RMSE and MAE were higher than those of the linear models, suggesting that linear functions more accurately represent the underlying data structure. Random Forests also performed well (Test $R^2$ = 99.49%), though with a higher cross-validation variance of $\pm$0.38, confirming that larger ensembles do not necessarily outperform simpler linear methods in this context.

The Decision Tree and KNN regressors exhibited notably lower performance, with Test $R^2$ values of 97.00% and 97.44%, respectively. Their higher cross-validation variances (Decision Tree: $\pm$1.04; KNN: $\pm$1.66) and larger error metrics indicate that they are more sensitive to sampling variation and less capable of capturing the global structure of the data.

**Table 1.** Block 1: Regression Results Using 5-Fold Cross-Validation with an 80/20 Train–Test Split.

| Model | Test $R^2$ (%) | CV5 $R^2$ (%) | CV5 Std | RMSE | MAE |
|---|---|---|---|---|---|
| Ridge Regression | 99.96 | 99.95 | $\pm$0.02 | 0.283 | 0.213 |
| Lasso Regression | 99.95 | 99.95 | $\pm$0.02 | 0.301 | 0.241 |
| Linear SVM | 99.95 | 99.92 | $\pm$0.03 | 0.293 | 0.237 |
| Gradient Boosting | 99.93 | 99.91 | $\pm$0.04 | 0.359 | 0.224 |
| Extra Trees | 99.93 | 99.85 | $\pm$0.14 | 0.365 | 0.182 |
| Neural Network (MLP) | 99.79 | 99.79 | $\pm$0.00 | 0.598 | 0.460 |
| Random Forest | 99.49 | 99.26 | $\pm$0.38 | 0.964 | 0.490 |
| Decision Tree | 97.00 | 96.13 | $\pm$1.04 | 2.328 | 1.215 |
| KNN | 97.44 | 90.77 | $\pm$1.66 | 2.153 | 1.536 |

Block 2 extends the analysis by employing 10-fold cross-validation, offering more robust and finely resolved evaluations of model stability and generalization. Across nearly all models, the results are consistent with those from Block 1, reaffirming the dominance of linear methods for this dataset. Ridge Regression once again delivers the best overall performance, achieving a Test $R^2$ of 99.96% and a 10-fold cross-validation mean of 99.95% with an extremely small standard deviation of $\pm$0.02. Lasso Regression and Linear SVM match this performance, further reinforcing the conclusion that the target variable is governed by highly linear relationships.

Tree-based and boosting models again follow closely behind. Gradient Boosting retains strong performance (Test $R^2$ = 99.93%, CV = 99.91% $\pm$0.04), while Extra Trees shows slightly higher instability (Test $R^2$ = 99.93%, CV = 99.87% $\pm$0.15), consistent with its randomized splitting mechanism. The Neural Network achieves a Test $R^2$ of 99.82% and a perfectly stable cross-validation spread ($\pm$0.00), confirming consistent training behavior across folds, though its higher RMSE and MAE again reveal that nonlinear architectures do not offer meaningful advantages over simpler linear methods.

Random Forest continues to trail the higher-performing models, with a Test $R^2$ of 99.49% and a noticeably larger cross-validation standard deviation of $\pm$0.46, suggesting greater sensitivity to variations in the training data. The lower-tier models, Decision Tree and KNN, again show limited performance. Decision Tree exhibits a Test $R^2$ of 97.00% with considerable variability ($\pm$1.95), while KNN achieves 97.44% Test $R^2$ but only 91.76% mean cross-validation performance, indicating substantial instability and the effects of high-dimensional distance limitations.

Overall, both experimental blocks reveal a consistent hierarchy of model performance. Ridge Regression, Lasso Regression, and Linear SVM stand out as the most accurate

and stable models, confirming the dominance of linear patterns in the underlying data. Gradient Boosting and Extra Trees provide strong but slightly less stable alternatives, while the Neural Network shows reliable but not superior performance. Random Forest, Decision Tree, and KNN exhibit lower accuracy and stability, making them less suitable for this highly deterministic regression task.

**Table 2.** Block 2: Regression Results Using 10-Fold Cross-Validation with an 80/20 Train–Test Split.

| Model | Test $R^2$ (%) | CV10 $R^2$ (%) | CV10 Std | RMSE | MAE |
|---|---|---|---|---|---|
| Ridge Regression | 99.96 | 99.95 | ±0.02 | 0.283 | 0.213 |
| Lasso Regression | 99.95 | 99.95 | ±0.02 | 0.301 | 0.241 |
| Linear SVM | 99.95 | 99.95 | ±0.02 | 0.293 | 0.237 |
| Gradient Boosting | 99.93 | 99.91 | ±0.04 | 0.359 | 0.224 |
| Extra Trees | 99.93 | 99.87 | ±0.15 | 0.365 | 0.182 |
| Neural Network (MLP) | 99.82 | 99.82 | ±0.00 | 0.556 | 0.430 |
| Random Forest | 99.49 | 99.29 | ±0.46 | 0.964 | 0.490 |
| Decision Tree | 97.00 | 96.64 | ±1.95 | 2.328 | 1.215 |
| KNN | 97.44 | 91.76 | ±1.40 | 2.153 | 1.536 |

## 5. Discussion

The results from both the 5-fold and 10-fold cross-validation experiments provide strong evidence that the dataset exhibits highly linear relationships between the input features and the target variable `Average_Score`. This is evident from the consistently superior performance of linear models such as Ridge Regression, Lasso Regression, and Linear SVM, all of which achieve near-perfect predictive accuracy with extremely low error metrics and minimal cross-validation variance. Among them, Ridge Regression emerges as the best-performing model, achieving a Test $R^2$ of 99.96% and cross-validation scores of 99.95% with a remarkably small standard deviation of only ±0.02. This suggests that the $L_2$ regularization used in Ridge effectively stabilizes coefficient estimates without sacrificing bias, making it highly suitable for this dataset.

The similar performance of Lasso Regression and Linear SVM further supports the notion that the underlying relationships in the data are predominantly linear. Lasso's ability to perform feature selection did not yield improvements over Ridge, indicating that most input variables contribute meaningful information and that sparsity is not essential for optimal performance. Likewise, the strong results from Linear SVM imply that the error distribution is well-contained within a linear margin, making nonlinear kernels unnecessary.

Nonlinear and ensemble-based models such as Gradient Boosting, Extra Trees, and Random Forest also demonstrate strong predictive capability but do not surpass the linear approaches. Gradient Boosting, for instance, reaches a Test $R^2$ of 99.93% and CV10 $R^2$ of 99.91%, showing that while it can capture nonlinear interactions, these interactions add only marginal predictive value beyond what linear models already capture. Extra Trees and Random Forest, although powerful for modeling complex feature interactions, exhibit higher variance across folds. This increased variability likely arises from their reliance on random splits and sensitivity to local data variations. The comparatively lower performance of Random Forest (CV10 $R^2 = 99.29\%$) suggests that aggregating multiple trees does not sufficiently address these instabilities when relationships in the data are dominantly linear.

The Neural Network (MLP) model presents an interesting case. While its Test $R^2$ (99.82%) is competitive, it does not reach the performance of the linear models or the best-performing ensemble methods. Its RMSE and MAE values are also relatively higher. This indicates that introducing nonlinear activation functions and deep-learning complexity does not yield significant benefits, and may even introduce unnecessary noise or overfitting. However, the Neural Network demonstrates perfect cross-validation stability (CV Std =

0.00), suggesting that the model converges consistently under the given architecture and hyperparameters. This stability may be attributed to well-tuned optimization settings and the model's capacity to generalize smoothly across training folds.

In contrast, the Decision Tree and KNN models perform substantially worse than all other candidates. The Decision Tree shows pronounced overfitting, reflected in its lower Test $R^2$ (97.00%) and its extremely high variance across folds (CV Std up to $\pm$1.95). Its hierarchical structure, while capable of modeling nonlinearity, is prone to instability when small perturbations in the data alter split choices. Similarly, KNN suffers due to the high dimensionality and scale sensitivity of the dataset. Its CV10 $R^2$ of 91.76% and relatively high RMSE/MAE values indicate that local averaging is insufficient for capturing global patterns, especially in a dataset where linear trends dominate.

Comparing the 5-fold and 10-fold CV results reveals a high degree of consistency across experimental settings. The performance rankings remain unchanged, and deviations between the two CV schemes are minimal. This further validates the robustness of the findings and indicates that the dataset offers sufficient sample diversity across folds.

Overall, the results highlight the strong suitability of linear modeling techniques for this prediction task. The exceptional performance of Ridge Regression, along with the sustained reliability of Lasso and Linear SVM, underscores the linearity and low-noise structure of the dataset. Ensemble and nonlinear models, despite their greater flexibility, do not offer significant advantages and may introduce additional complexity without corresponding improvements in predictive accuracy.

The observed trends also suggest several avenues for future work. Investigating the coefficients of the linear models could provide interpretable insights into which features contribute most strongly to `Average_Score`. Additionally, dimensionality reduction methods such as PCA or feature clustering could be explored to determine whether the removal of redundant information affects model stability or performance. Finally, evaluating the models on external or more heterogeneous datasets would help assess generalizability and ensure that the superior performance of linear models is not an artifact of dataset-specific properties.

## 6. Conclusion

The experimental findings demonstrate that linear models, particularly Ridge Regression, consistently deliver the highest predictive performance for estimating the `Average_Score` target variable across all evaluation settings. In both the 5-fold and 10-fold cross-validation blocks, Ridge Regression achieves the highest $R^2$ scores (Test $R^2 = 99.96\%$; CV = 99.95% $\pm$0.02), paired with the lowest RMSE and MAE values, indicating exceptional accuracy and robustness. Lasso Regression and Linear SVM follow closely, exhibiting similarly strong performance with minimal variance across folds, reinforcing the presence of dominant linear relationships within the dataset.

Ensemble-based methods such as Gradient Boosting and Extra Trees also perform at a high level (CV10 $R^2$: 99.91% and 99.87%, respectively), though they exhibit slightly higher variance, reflecting their sensitivity to complex feature interactions and potential overfitting. Tree-based models like Random Forest and Decision Tree show comparatively lower performance and substantially higher variability (e.g., Decision Tree CV Std up to $\pm$1.95), illustrating their limitations in modeling the smooth linear patterns present in the data. The KNN model performs the weakest (CV10 $R^2 = 91.76\%$), likely due to sensitivity to feature scaling, noise, and the curse of dimensionality.

The Neural Network (MLP) shows excellent predictive power (Test $R^2 \approx 99.8\%$), but its higher RMSE and MAE compared with linear models highlight that the additional nonlinearity does not provide meaningful advantages given the inherently linear structure of the dataset. Interestingly, the MLP demonstrates perfect fold-to-fold consistency (CV Std = 0.00), indicating stable optimization despite slightly inferior absolute performance.

Overall, the results confirm that the dataset is dominated by strong linear correlations, enabling linear models to outperform more complex nonlinear approaches. Features such as

`Composite_Score` and other aggregate academic indicators provide substantial explanatory power, reducing the need for complex architectures. Furthermore, the comparison between the 5-fold and 10-fold CV schemes shows consistent performance trends, reinforcing the reliability and stability of the models across different data partitions.

In future work, a deeper analysis of feature importance within linear models may help reveal the extent to which each predictor contributes to the near-perfect accuracy observed. Additionally, dimensionality reduction techniques (e.g., PCA) could be explored to evaluate whether model performance remains stable when reducing redundancy among correlated features. Investigating model generalization on external or more diverse datasets could further validate the robustness of these findings.

## References


1. Morales, S.F.G.; Zander, V.A.D. Evaluación del desempeño al personal de la empresa SOBOC Grafic ubicada en la ciudad de Quito. *Journal DL* **2015**.
2. Alzubaidi, L.; Zhang, J.; Humaidi, A.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data* **2021**, *8*, 1–74.
3. Gonzales, W.; Cordero, Z.; Abanto-Ramírez, C.D.; Torres Armas, E.A.; Iftikhar, H.; Lopez-Gonzales, J.L. A hybrid AI approach for predicting academic performance in RBE students. *Frontiers in Artificial Intelligence* **2025**, *8*, 1651100.
4. Alles, M.A. *Diccionario de Competencias*, 2 ed.; Granica: Buenos Aires, Argentina, 2006.
5. Alles, M. *Dirección Estratégica de Recursos Humanos: Gestión por Competencias*; Granica: Buenos Aires, 2006.
6. Nieto Rosado, M.O. Análisis del desempeño por competencias basado en la evaluación de 360° de la Gerencia de Administración y Finanzas del Gobierno Regional de Lambayeque bajo el modelo de Martha Alles. Master's thesis, Universidad Católica Santo Toribio de Mogrovejo, 2020.
7. Pullas, K.; Jiménez, V. La gestión del conocimiento y su relación con el desempeño laboral de los docentes en las instituciones educativas fiscales de la zona 8 del Ecuador, 2019. *Revista Atlante* **2019**.
8. Vargas, H.; et al. Standardizing Course Assessment in Competency-Based Settings. *Frontiers in Education* **2025**. https://doi.org/10.3389/feduc.2025.1579124.
9. Vela, I.; Contreras Julián, R.; Delgado Bardales, J. Effects of a human resource management model on job performance: an empirical study in Peruvian educational management. *Sapienza: International Journal of Interdisciplinary Studies* **2025**, *6*, e25046. https://doi.org/10.51798/sijis.v6i3.1067.
10. Song, H.; Li, Q. Employee Performance Prediction Model Based on BP Neural Networks. In Proceedings of the Proceedings of the 2025 International Conference on Digital Management and Information Technology (DMIT); , 2025.
11. Hussain, I.; Qureshi, M.; Ismail, M.; Iftikhar, H.; Zywiołek, J.; López-Gonzales, J.L. Optimal features selection in the high dimensional data based on robust technique: Application to different health database. *Heliyon* **2024**, *10*.
12. Patel, K.; Sheth, K.; Mehta, D.; Tanwar, S.; Florea, B.C.; Taralunga, D.D.; Altameem, A.; Altameem, T.; Sharma, R. RanKer: An AI-Based Employee-Performance Classification Scheme to Rank and Identify Low Performers. *Mathematics* **2022**, *10*.
13. Talebi, H.; Khatibi Bardsiri, A.; Khatibi Bardsiri, V. Machine Learning Approaches for Predicting Employee Turnover: A Systematic Review. *Engineering Reports* **2025**, *7*, e70298. https://doi.org/10.1002/eng2.70298.
14. Căvescu, A.M.; Popescu, N. Predictive Analytics in Human Resources Management: Evaluating AIHR's Role in Talent Retention. *AppliedMath* **2025**, *5*. https://doi.org/10.3390/appliedmath5030099.
15. Marnewick, C.; Romero-Torres, A.; Delisle, J. Rich pictures as a research method in project management: A way to engage practitioners. *Project Leadership and Society* **2024**, *5*, 100127. https://doi.org/10.1016/j.plas.2024.100127.
16. Goto, Y.; Miura, H. Using the Soft Systems Methodology to Link Healthcare and Long-Term Care Delivery Systems: A Case Study of Community Policy Coordinator Activities in Japan. *International Journal of Environmental Research and Public Health* **2022**, *19*, 8462. https://doi.org/10.3390/ijerph19148462.
17. Goto, Y.; Miura, H. Using the Soft Systems Methodology to Link Healthcare and Long-Term Care Delivery Systems: A Case Study of Community Policy Coordinator Activities in Japan. *International Journal of Environmental Research and Public Health* **2022**, *19*, 8462. https://doi.org/10.3390/ijerph19148462.
18. Almeida, R.; Chen, Y.; Robertson, I. Performance Appraisal Effectiveness in Modern Organisations: A Systematic Review of Outcomes, Moderators, and Contextual Factors. *Human Resource Management Review* **2025**, *35*, 100–118. https://doi.org/10.1016/j.hrmr.2024.100118.
19. Okechukwu, E.; Adewale, T. Impact of Performance Appraisal Mechanisms on Employee Productivity. *International Journal of Productivity and Performance Management* **2024**, *73*, 1214–1232. https://doi.org/10.1108/IJPPM-04-2024-0026.
20. Checkland, P.; Winter, M. Soft Systems Methodology: A Thirty-Year Retrospective. *Systems Research and Behavioral Science* **2022**, *39*, 567–582. https://doi.org/10.1002/sres.2856.
21. White, L.; Taket, A. Participatory methods in soft systems thinking. *Systems* **2023**, *11*, 85. https://doi.org/10.3390/systems11020085.

22. Zhu, Z.; Zhang, W.; Li, H. Soft systems thinking for organizational learning in higher education. *Higher Education Research & Development* **2024**, *43*, 45–60. https://doi.org/10.1080/07294360.2023.2254187.
23. Iftikhar, H.; Hashem, A.F.; Qureshi, M.; Rodrigues, P.C. Clinical application of machine learning models for early-stage chronic kidney disease detection. *Diagnostics* **2025**, *15*, 2610.
24. Saleem, A.; Shah, S.; Iftikhar, H.; Zywiołek, J.; Albalawi, O. A comprehensive systematic survey of iot protocols: Implications for data quality and performance. *IEEE Access* **2024**.
25. Khan, F.; Iftikhar, H.; Khan, I.; Rodrigues, P.C.; Alharbi, A.A.; Allohibi, J. A Hybrid Vector Autoregressive Model for Accurate Macroeconomic Forecasting: An Application to the US Economy. *Mathematics* **2025**, *13*, 1706.
26. Iftikhar, H.; Qureshi, M.; Zywiołek, J.; López-Gonzales, J.L.; Albalawi, O. Short-term PM 2.5 forecasting using a unique ensemble technique for proactive environmental management initiatives. *Frontiers in Environmental Science* **2024**, *12*, 1442644.
27. Iftikhar, H.; Qureshi, M.; Rodrigues, P.C.; Iftikhar, M.U.; López-Gonzales, J.L. Daily Crude Oil Prices Forecasting Using A Novel Hybrid Time Series Technique. *IEEE Access* **2025**.