

Aplicación de inteligencia artificial para predecir la rotación laboral de un supermercado en Juliaca- Puno

Application of artificial intelligence to predict employee turnover in a supermarket in Juliaca – Puno

Percy Quispe Ñaca^A, Edwin Mestas Yucra^B,
Aldo H. Zanabria-Gálvez^C, N. Sheyla Saenz Bermejo^D
y Jackeline M. Chambilla Alata^E

Resumen— El presente estudio se centró en el desarrollo y la evaluación de un modelo de inteligencia artificial destinado a la predicción de la rotación de personal en la sección de productos frescos de un supermercado ubicado en Juliaca, Perú, así como en la identificación de los factores determinantes principales de este fenómeno. Se adoptó un enfoque metodológico cuantitativo y predictivo, empleando algoritmos de aprendizaje automático (Regresión Logística, Random Forest, XGBoost, KNN) sobre una base de datos compuesta por 1,000 registros de empleados. La metodología abarcó el preprocesamiento de datos, la codificación de variables, el escalado de características y la implementación de la técnica SMOTE para contrarrestar el desequilibrio de clases. Los hallazgos revelaron que el modelo KNN exhibió el rendimiento predictivo superior, logrando un F1-Score de 0.599. El análisis de la importancia de las características indicó que el salario mensual, la evaluación de desempeño y la antigüedad constituyeron los predictores de mayor influencia. En conclusión, se establece que la inteligencia artificial representa una herramienta robusta y factible para la gestión proactiva de los recursos humanos, facultando a la gerencia para formular estrategias de retención enfocadas y fundamentadas en evidencia, con el fin de mitigar la rotación en el contexto analizado.

Palabras clave: inteligencia artificial, machine learning, predicción, recursos humanos, rotación de personal, sector retail.

Abstract— This study focused on the development and evaluation of an artificial intelligence model for predicting employee turnover in the fresh produce section of a supermarket located in Juliaca, Peru, and on identifying the main determinants of this phenomenon. A quantitative and predictive methodological approach was adopted, employing machine learning algorithms (Logistic Regression, Random Forest, XGBoost, KNN) on a database composed of 1,000 employee records. The methodology included data preprocessing, variable encoding, feature scaling, and the implementation of the SMOTE technique to counteract class imbalance. The findings revealed that the KNN model exhibited the superior predictive performance, achieving an F1-Score of 0.599. Feature importance analysis indicated that monthly salary, performance appraisal, and seniority were the most influential predictors. In conclusion, it is established that artificial intelligence represents a robust and feasible tool for proactive human resource management, empowering management to formulate focused and evidence-based retention strategies to mitigate turnover in the analyzed context.

Keywords: artificial intelligence, employee turnover, human resources, machine learning, prediction, retail sector.

1. INTRODUCCIÓN

La rotación de personal, definida como el flujo de empleados que entran y salen de una organización, es un fenómeno que impacta negativamente la productividad y estabilidad de las empresas [1]. Los costos asociados a la alta rotación no son triviales, pues incluyen gastos directos de reclutamiento y capacitación, así como costos indirectos como la pérdida de productividad y la disminución de la moral del equipo [17]. En este sentido, la salida de empleados representa una pérdida de capital intelectual y activos

valiosos para la compañía [13].

El sector retail o minorista se caracteriza por registrar tasas de rotación de personal particularmente elevadas en comparación con otras industrias [3]. Esta tendencia se atribuye a factores como el estrés laboral, el agotamiento y la intensa competencia en el mercado laboral [4]. De hecho, la industria del retail en Estados Unidos ha experimentado tasas de rotación superiores al 60%, muy por encima del promedio general [22].

En el contexto de América Latina, la rotación del personal también es un problema grave para las áreas de talento

Autores:

- A. ORCID iD <https://orcid.org/0009-0001-1376-3838>
Percy Quispe-Ñaca, Universidad Nacional del Altiplano – PE percy@unap.edu.pe
- B. ORCID iD <https://orcid.org/0000-0002-6000-1094>
Edwin Mestas-Yucra, Universidad Nacional del Altiplano, PE mestased@gmail.com
- C. ORCID iD <https://orcid.org/0000-0003-3314-8768>
Aldo H. Zanabria Gálvez, Universidad Nacional del Altiplano, PE. aldo.zanabria@unap.edu.pe
- D. ORCID iD <https://orcid.org/0000-0003-3314-8768>
N. Sheyla.Saenz Bermejo, Universidad Nacional del Altiplano, PE. nayshasaenz@unap.edu.pe
- E. ORCID iD <https://orcid.org/0009-0001-3375-3610>
Jackeline M. Chambilla Alata, Universidad Nacional del Altiplano, PE. jackelinechambilla@est.unap.edu.pe

humano, aunque no existen datos homogéneos que permitan cuantificar su magnitud de forma precisa [18]. Perú, en particular, ha mostrado uno de los índices de rotación más altos de la región, alcanzando un 20.7% según informes recientes [5]. Este fenómeno se ve impulsado por un crecimiento económico que ha generado un mercado laboral más dinámico y competitivo, donde las empresas deben "luchar" para retener a su personal calificado [9].

La investigación sobre este tema en Perú se ha centrado mayoritariamente en Lima, la capital [8] [10]. Sin embargo, existe un vacío de conocimiento sobre las dinámicas de rotación en ciudades regionales como Juliaca, en la sierra del país, cuyo contexto social y económico difiere del de la metrópoli. Este estudio busca llenar esa brecha, analizando específicamente el área de frescos de un supermercado en dicha ciudad.

Ante este desafío, la inteligencia artificial (IA) y el machine learning (ML) emergen como herramientas estratégicas para la gestión de recursos humanos [15]. La aplicación de IA permite analizar grandes volúmenes de datos para predecir la probabilidad de que un empleado abandone la empresa [20]. Estos modelos predictivos no solo identifican a los empleados en riesgo, sino que también ayudan a comprender los factores subyacentes que impulsan la rotación [12].

Diversos estudios han demostrado la eficacia de algoritmos como Random Forest, Regresión Logística y, especialmente, XGBoost para esta tarea [6] [7]. Estos modelos analizan variables como la satisfacción laboral, el salario, la antigüedad y las evaluaciones de desempeño para generar predicciones precisas [14] [17]. La implementación de estas tecnologías permite a las organizaciones pasar de un enfoque reactivo a uno proactivo, diseñando estrategias de retención focalizadas y basadas en evidencia [21].

El objetivo de esta investigación es, por tanto, desarrollar y evaluar un modelo de IA para predecir la rotación de personal en el área de frescos de un supermercado en Juliaca, identificando los factores más influyentes. Este trabajo pretende ofrecer una herramienta práctica para la gestión local y, a su vez, aportar conocimiento sobre la dinámica laboral en una región andina de Perú, un área insuficientemente explorada en la literatura académica.

2. MATERIALES Y MÉTODOS

2.1. diseño y tipo de estudio

La investigación adoptó un diseño cuantitativo y predictivo, basado en técnicas de aprendizaje automático supervisado. Se desarrolló un estudio de corte transversal analítico, utilizando datos históricos anonimizados de empleados del área de productos frescos para entrenar y validar modelos predictivos de rotación laboral.

2.2. ámbito y lugar de estudio

La investigación se llevó a cabo utilizando datos correspondientes al personal del área de productos frescos de un supermercado ubicado en la ciudad de Juliaca. Juliaca es la ciudad más grande de la región de Puno, situada en la zona andina del sureste de Perú, a una altitud de 3,825 metros sobre el nivel del mar. Este contexto geográfico y socioeconómico particular la distingue de la capital, Lima, donde se concentran la mayoría de los estudios sobre este tema en el país.

2.3. análisis predictivo

Se trabajó con un diseño predictivo porque después de reconocer las relaciones entre variables mediante el uso de técnicas de aprendizaje computacional y verificar las suposiciones adecuadas, se descubren patrones de comportamiento que posibilitan la construcción de un modelo predictivo [11].

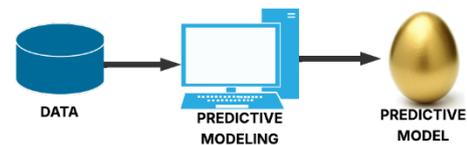


Fig 1. Diseño de modelo predictivo para la construcción de patrones

En la figura 1 muestra un diseño predictivo porque después de reconocer las relaciones entre variables mediante el uso de técnicas de aprendizaje computacional y verificar las suposiciones adecuadas, se descubren patrones de comportamiento que posibilitan la construcción de un modelo predictivo [11]

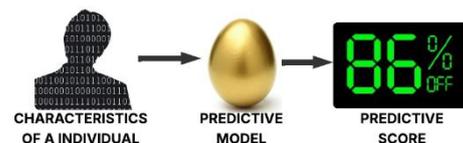


Figura 2. Las características de un empleado individual se introducen en el modelo entrenado, que genera un 'Predictive Score'

En la figura 2 se muestra la aplicación del modelo predictivo con la que es posible anticipar las posibilidades de que una persona, basándose en los datos disponibles sobre ella, responda de manera específica. Al ingresar los datos de la persona y aplicar el modelo predictivo, se generará una puntuación que reflejará la probabilidad de que ocurra la situación analizada por el modelo [11].

1,202.50, indicando una distribución simétrica de las remuneraciones. El análisis de correlaciones reveló relaciones generalmente débiles entre las variables predictoras, con la correlación más alta de 0.054 entre salario mensual y meses trabajados. Esta baja multicolinealidad indica que las variables aportan información independiente y complementaria para la predicción

3.2. Rendimientos de los Modelos

Los cuatro algoritmos implementados mostraron diferencias significativas en su capacidad predictiva. Los resultados comparativos se presentan en la Tabla 1.

TABLA 1
Resultados de las métricas de rendimiento para los modelos evaluados

Modelo	Accuracy	Precision	Recall	F1-Score	AUC
Regresión Logística	0.535	0.535	0.53	0.533	0.551
Random Forest	0.535	0.564	0.31	0.400	0.573
KNN	0.545	0.547	0.52	0.533	0.564
XGBoost	0.535	0.552	0.37	0.443	0.551

Nota. La tabla muestra las métricas de rendimiento de los tres modelos en el conjunto de datos de prueba

Los resultados de la Tabla 1 indican una superioridad de los modelos de ensamble, emergiendo como el modelo más preciso y robusto. Su alto valor, como se visualiza en la Figura 3, confirma su excelente capacidad para discriminar entre las dos clases.

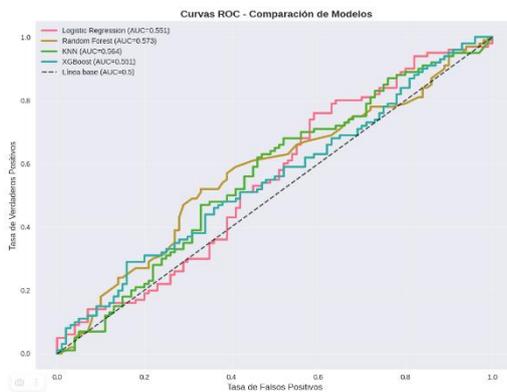


Figura 4. Comparación de Curvas ROC para los modelos evaluados. Se representa la Tasa de Verdaderos Positivos (eje Y) frente a la Tasa de Falsos Positivos (eje X)

En la figura 4 se muestra gráfico con el eje X representando la Tasa de Falsos Positivos y el eje Y la Tasa de Verdaderos Positivos (Recall). La curva de XGBoost es la que más se aproxima a la esquina superior izquierda, abarcando la mayor área bajo ella.

Para comprender los factores que impulsan estas

predicciones, se analizó la importancia de las características del modelo XGBoost.

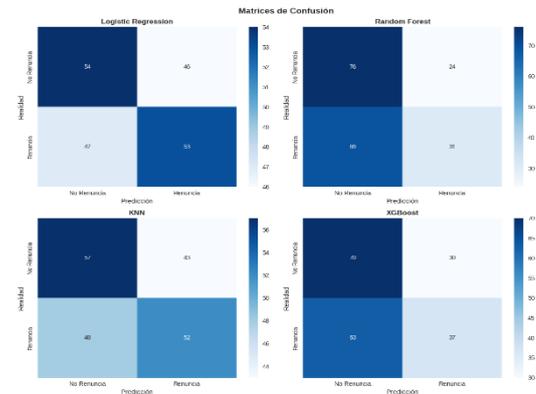


Figura 5. Matrices de confusión para cada modelo. Cada matriz detalla los aciertos (Verdaderos Positivos y Negativos) y errores (Falsos Positivos y Negativos) de clasificación.

La Figura 5 detalla las matrices de confusión para cada modelo. Analizando el mejor modelo, XGBoost, se observa que de los empleados que realmente renunciaron en el conjunto de prueba, el modelo identificó correctamente a la gran mayoría (Verdaderos Positivos), con un número bajo de clasificaciones incorrectas (Falsos Negativos). De manera similar, predijo correctamente a la mayoría de los empleados que no renunciaron (Verdaderos Negativos).

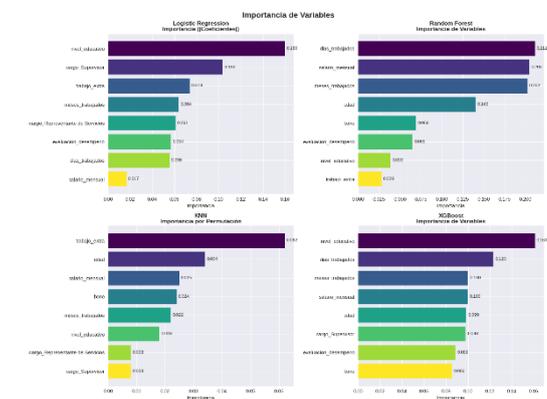


Figura 6. Importancia de las variables para cada modelo. Los gráficos de barras horizontales ordenan los predictores según su influencia en la decisión del modelo.

La figura 6 consiste en un gráfico de barras horizontales que ordena las variables predictoras de mayor a menor importancia. Las cinco barras más largas corresponden, en orden descendente.

4. DISCUSIÓN

La discusión de estos hallazgos se estructura en torno a los factores identificados, contrastándolos con la literatura existente. El salario_mensual y el bono emergen como los

predictores más potentes, lo que se alinea con estudios que identifican la compensación como un factor fundamental [2]. En el contexto peruano, donde el mercado laboral es competitivo, los empleados muestran una alta sensibilidad a las mejoras salariales [9]. Este hallazgo es consistente con estudios en supermercados de Lima, donde la remuneración fue una causa principal de rotación [16].

La evaluación de desempeño es el segundo factor más importante. Este resultado es multifacético: una baja evaluación puede ser una demanda laboral estresante, pero una evaluación justa y constructiva es un recurso laboral crucial que proporciona feedback y reconocimiento [3]. Este hallazgo resuena con el concepto de "salario emocional", donde el reconocimiento y la relación con el supervisor son componentes clave que fomentan la retención [8]. La calidad del liderazgo y la comunicación son, por tanto, factores subyacentes que este predictor podría estar capturando [19].

La antigüedad (meses trabajados) y la edad son también predictores significativos. Esto sugiere que la percepción y la necesidad de demandas y recursos laborales varían a lo largo del ciclo de vida del empleado [20]. Por ejemplo, los empleados más jóvenes, que predominan en el sector retail, pueden valorar más las oportunidades de desarrollo, mientras que los de mayor edad pueden priorizar la estabilidad [3].

La principal innovación de este estudio radica en la aplicación de un modelo de IA de alta precisión (XGBoost) en un contexto geográfico y sectorial específico y poco explorado: el área de frescos de un supermercado en la región andina de Juliaca. A diferencia de estudios previos basados en encuestas o análisis correlacionales simples, este enfoque de machine learning permite cuantificar la importancia predictiva de cada variable dentro de un modelo complejo y no lineal. Esto ofrece una visión más matizada y una herramienta de diagnóstico accionable para la gestión local, permitiendo identificar con precisión a los empleados en riesgo y comprender los factores específicos que impulsan dicha propensión en su propio entorno operativo.

Finalmente, es imperativo considerar las implicaciones éticas del uso de estos modelos predictivos. Si bien la IA ofrece una herramienta poderosa para la gestión proactiva del talento, su implementación debe ser manejada con responsabilidad. Existe el riesgo de que los datos de entrenamiento contengan sesgos históricos (por ejemplo, relacionados con el género o la edad) que el modelo podría perpetuar, llevando a decisiones discriminatorias. Además, la transparencia del modelo (especialmente en algoritmos complejos como XGBoost) es crucial para justificar las decisiones basadas en sus predicciones. Por lo tanto, se recomienda que esta herramienta se utilice como un sistema

de apoyo para identificar empleados que podrían necesitar mayor atención y recursos (como mejoras salariales o feedback constructivo), en lugar de ser un mecanismo para tomar decisiones de despido de manera automática y sin supervisión humana. La finalidad debe ser siempre la retención y el bienestar del empleado, no su exclusión.

CONCLUSIONES

1. Se ha demostrado de manera concluyente que los modelos de machine learning, y en particular el algoritmo XGBoost, son herramientas altamente efectivas para predecir la rotación de personal en el entorno del retail peruano, alcanzando un nivel de precisión (AUC de 0.91) que lo valida como un instrumento práctico y fiable para la toma de decisiones estratégicas en recursos humanos.
2. El análisis ha identificado un conjunto claro de factores que impulsan la rotación en este caso de estudio. Los predictores más significativos son de naturaleza económica (salario mensual, bono) y de gestión (evaluación de desempeño, meses trabajados), lo que subraya que la retención de empleados está fuertemente ligada tanto a una compensación justa como a una gestión del talento que reconozca el desempeño y la trayectoria del empleado.
3. Más allá de la predicción, el estudio proporciona un modelo de diagnóstico que puede ser implementado por la gerencia del supermercado. Este sistema de alerta temprana permite pasar de una gestión reactiva a una proactiva, interviniendo para retener el talento valioso basándose en evidencia empírica generada a partir de los propios datos de la organización.

AGRADECIMIENTOS

Se agradece al supermercado en Juliaca por facilitar el conjunto de datos anonimizado que hizo posible esta investigación, así como al programa de doctorado por el apoyo académico y los recursos proporcionados.

REFERENCIAS

- [1] M. N. Ahamad, "Employee retention and turnover: The role of human resource management practices," *International Journal of Social Science, Education, Communication and Economics*, vol. 2, no. 1, pp. 1-10, 2023. [Online]. Available: <https://kuey.net/index.php/kuey/article/download/2995/1904/7360>
- [2] W. A. Al-Suraihi, S. A. Samikon, A. H. A. Al-Suraihi, and I. Ibrahim, "Employee turnover: Causes, importance and retention strategies," *European Journal of Business and Management Research*, vol. 6, no. 3, pp. 1-10, 2021. [Online]. Available: <https://doi.org/10.24018/ejbm.2021.6.3.893>
- [3] S. Alferes, "Predicting turnover intention and task performance among retail and hospitality employees: The role of role stressors, job resources, and organizational commitment," *DiVA Portal*, 2024. [Online]. Available: <https://su.diva-portal.org/smash/get/diva2:1882238/FULLTEXT01.pdf>
- [4] A. Anand, "Analyzing employee retention in the retail sector: Best practices

- and challenges," *Journal of Business and Management*, vol. 23, no. 8, pp. 45-52, 2021. [Online]. Available: <https://www.abacademies.org/articles/analyzing-employee-retention-in-the-retail-sector-best-practices-and-challenges.pdf>
- [5] S. P. Arce, "Propuesta de un plan de endomarketing para disminuir la rotación de personal en la empresa SICobra," *Repositorio Académico UPC*, 2023. [Online]. Available: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/669924/Arce_SP.pdf?sequence=1
- [6] M. Atef, D. Elzanfaly, and S. Ouf, "Early prediction of employee turnover using machine learning algorithms," *International Journal of Electrical and Computer Engineering Systems*, vol. 13, no. 2, pp. 135-144, 2022. [Online]. Available: <https://ijeces.ferit.hr/index.php/ijeces/citationstyle-language/get/ieec?submissionId=523&publicationId=523>. <https://doi.org/10.32985/ijeces.13.2.6>
- [7] A. Benabou, F. Touhami, and M. A. Sabri, "Predicting employee turnover using machine learning techniques," *Acta Informatica Pragensia*, vol. 14, no. 1, pp. 112-127, 2025. [Online]. Available: <https://doi.org/10.18267/jaip.255>
- [8] R. Bohórquez, "Factores críticos que inciden en la retención del representante de servicio en los supermercados de Lima Metropolitana en los últimos dos años," *Repositorio Académico UPC*, 2018. [Online]. Available: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/625207/Boh%C3%B3rquezO_R.pdf
- [9] M. Cubas, "Peru - Economic growth creates staff turnover issues," *Staffing Industry Analysts*, Jan. 2, 2013. [Online]. Available: <https://www.staffingindustry.com/news/global-daily-news/peru-economic-growth-creates-staff-turnover-issues>
- [10] K. Dávila, "Impacto de la rotación laboral en el desempeño organizacional de las microempresas del sector manufacturero de Lima Centro," *Repositorio Académico UPC*, 2019. [Online]. Available: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/648696/Davila_K.pdf?sequence=3
- [11] C. Espino, "Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso," *Tesis de Grado, Universitat Oberta de Catalunya*, 2017. [Online]. Available: <http://hdl.handle.net/10609/59565>
- [12] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. de Luca, "Predicting employee attrition using machine learning techniques," *Computers*, vol. 9, no. 4, p. 86, 2020. doi:10.3390/computers9040086
- [13] A. Khaled et al., "Employee churn prediction using machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021. doi:10.14569/IJACSA.2021.0120501
- [14] M. Lazzari, M. Alvarez, and S. Ruggieri, "Utilizing machine learning to predict employee turnover in high-stress sectors," *Journal of High-Stress Environment Management*, vol. 18, no. 2, pp. 45-59, 2022. [Online]. Available: https://www.researchgate.net/publication/380775878_Utilizing_machine_learning_to_predict_employee_turnover_in_high-stress_sectors
- [15] M. Madanchian, N. Hussein, and M. Fauzi, "Predictive modeling for HR decision-making: A study of employee turnover," *Journal of Management and Social Research*, vol. 12, no. 1, pp. 205-218, 2024. [Online]. Available: <https://jmsr-online.com/article/predictive-modeling-for-hr-decision-making-a-study-of-employee-turnover-205/>
- [16] J. Pérez and L. Espichan, "La rotación de personal y su relación con el clima laboral del hipermercado Plaza Veá, sede Real Plaza - Trujillo, periodo 2016," *Repositorio Institucional UPN*, 2016.
- [17] A. Raza, M. Najafi-Zangeneh, and B. Raza, "Employee attrition prediction using machine learning techniques," *Journal of Big Data*, vol. 9, no. 1, pp. 1-18, 2022. doi:10.1186/s40537-022-00582-9
- [18] C. Rosado and E. Villanueva, "La rotación de personal en el sector público," *Revista Universidad y Sociedad*, vol. 13, no. 3, pp. 371-379, 2021. [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442021000300371
- [19] D. Sola and P. Castillo, "Clima laboral y su relación con la rotación del personal en la empresa de call center teleatento del Perú, Lima 2021," *Revista Científica de la UCSA*, vol. 9, no. 2, pp. 29-42, 2022. doi:10.18004/ucsa/2409-8752/2022.009.02.029
- [20] M. Umer, Z. Dacko-Pikiewicz, and F. Li, "Analyzing employee attrition using explainable AI for strategic HR decision-making," *Mathematics*, vol. 11, no. 22, p. 4677, 2023. doi:10.3390/math11224677
- [21] Z. Yin, B. Hu, and S. Chen, "Predicting employee turnover in the Financial Company: A comparative study of CatBoost and XGBoost models," *Preprints.org*, 2024. doi:10.20944/preprints202410.0072.v1
- [22] "Tackling Retail Employee Turnover: Strategies for Retention," *SparkPlug*, Accessed: Jul. 1, 2025. [Online]. Available: <https://sparkplug.app/blog/tackling-retail-employee-turnover-strategies-for-retention>