

Universidad de Lima  
Facultad de Ingeniería  
Carrera de Ingeniería Industrial



# **TECHNIQUES OF MACHINE LEARNING APPLIED TO REDUCE EMPLOYEE TURNOVER IN A COMPANY CLEANING AND DISINFECTION**

Tesis para optar el Título Profesional de Ingeniero Industrial

**Erika Noemi Romero Rojas**

**Código 20162488**

**Asesor**

Juan Carlos Quiroz Flores

Lima – Perú

Marzo de 2024



## Título

*Techniques of Machine Learning applied to reduce employee turnover in a Company cleaning and disinfection*

## Autor(es)

Erika Noemi Romero Rojas

[20162488@aloe.ulima.edu.pe](mailto:20162488@aloe.ulima.edu.pe)

Facultad de Ingeniería y Arquitectura, Universidad de Lima, Perú

Yvan Jesús López García

[ygarcia@ulima.edu.pe](mailto:ygarcia@ulima.edu.pe)

Facultad de Ingeniería y Arquitectura, Universidad de Lima, Perú

Juan Carlos Quiroz Flores

[jcquiroz@ulima.edu.pe](mailto:jcquiroz@ulima.edu.pe)

Facultad de Ingeniería y Arquitectura, Universidad de Lima, Perú

**Resumen:** La rotación de personal en las grandes industrias peruanas de manufactura ha venido incrementando en los últimos años. Si bien la rotación laboral es un efecto natural en las organizaciones, cuando es no deseada genera mayores costos de entrenamiento para el nuevo personal e impacta en el desempeño y clima laboral. Ante esta problemática nace la necesidad de poder identificar las posibles causas de rotación del personal operativo y predecir estos sucesos a través del análisis de datos en una etapa temprana para evitar y/o reducir su impacto en la compañía. El presente artículo de enfoque cuantitativo y alcance exploratorio-explicativo tiene como objetivo principal determinar los factores que influyen en la rotación de personal operativo de una empresa de manufactura del rubro de limpieza y desinfección a través de la recolección de datos empleando Machine Learning y fomentar propuestas que permitan dar soluciones ante la rotación de personal. Para el análisis de datos se empleó el software Orange, en donde los datos fueron entrenados con diferentes modelos de inteligencia como Random Forest, Logistic Regression, Decision Tree, y SCV, y Python para correr el modelo y obtener indicadores numéricos como el Área bajo la curva (AUC) y el análisis de la curva ROC. El estudio propuesto muestra que los modelos tienen un buen desempeño en la clasificación, con altas tasas de precisión y recall, 96% y 97% respectivamente, así como una exactitud general del 96%.

**Palabras Clave:** Machine Learning – Rotación de personal – Manufactura – Recursos Humanos – Personal operativo

**Abstract:** Staff turnover in large Peruvian manufacturing industries has been increasing in recent years. While job rotation is a natural effect in organizations, it generates higher training costs for new staff and impacts work performance and climate when unwanted. Given this problem arises the need to identify the possible causes of rotation of operational personnel and predict these events through data analysis at an early stage to avoid and reduce its impact on the company. This article of quantitative approach and exploratory scope-explanatory aims to identify the propensity of rotation of the operation of a company manufacturing cleaning and disinfection through a model of forecast by collecting data using Machine Learning and encourage proposals that enable solutions to be found to the factors influencing staff turnover. MS Excel and Orange software were used for data analysis, where the data were trained with different intelligence models such as Random Forest, Logistic Regression, Decision Tree, and SVM, and Python to run the model and get numerical indicators like the Area under the curve (AUC) and the analysis of the ROC curve. The proposed study shows that the models perform well in classification, with high accuracy and recall rates, 96% and 97%, respectively, and an overall accuracy of 96%.

**Keywords:** Machine Learning – Staff turnover – Manufacturing – Human Resources – Operational staff.

**Línea de investigación IDIC – ULIMA:** (5) - Productividad y Empleo

**Área y Sub-áreas de Investigación:** Gestión de Ingeniería

**Objetivo (s) de Desarrollo Sostenible (ODS) (8) - Trabajo Decente y Crecimiento Económico**

## PLANTEAMIENTO DEL PROBLEMA

En el Perú la rotación de personal en las grandes industrias ha venido incrementando en los últimos años. Si bien la rotación laboral es un efecto natural en las organizaciones, genera mayores costos de capacitación para el nuevo personal e impacta el desempeño y el clima laboral cuando no es deseado. Ante esta problemática surge la necesidad de identificar las posibles causas de la rotación del personal operativo y predecir estos eventos mediante el análisis de datos en una etapa temprana para evitar y reducir su impacto en la empresa. Por ello, se ha orientado el artículo hacia la aplicación de Técnicas de Machine Learning con la finalidad de identificar a tiempo las posibles causas de rotación.

Pregunta de investigación: ¿Cuáles son los factores que influyen en la rotación de personal operativo de una empresa de manufactura del rubro de limpieza y desinfección?

### OBJETIVOS

#### Objetivo general

Determinar los factores que influyen en la rotación de personal operativo de una empresa de manufactura del rubro de limpieza y desinfección a través de la recolección de datos empleando Machine Learning y fomentar propuestas que permitan dar soluciones ante la rotación de personal.

#### Objetivos específicos

- Analizar el impacto de la rotación de personal operativo de una empresa de manufactura del rubro de limpieza y desinfección.
- Identificar las estrategias de retención para el personal operativo en base a los factores que influyen en la rotación.

### JUSTIFICACIÓN

En los últimos años la presencia de herramientas de inteligencia artificial como el Machine Learning ha tomado gran presencia en las diversas áreas de las compañías. Sin embargo, aún existe una brecha de escepticismo por parte de los directores sobre el uso de la herramienta especialmente en el área de Recursos Humanos. Estudios previos muestran la eficiencia de la herramienta, la cual ha sido implementada en varios contextos, como el modelado de la rotación de personal y el compromiso de los empleados, con el objetivo de optimizar el rendimiento y producir un mejor retorno de la inversión para las organizaciones a través de la toma de decisiones basada en la recopilación de datos, métricas de recursos humanos y modelos predictivos. El Perú es uno de los países con un índice promedio de rotación laboral no deseada del 20.7% si bien todas las organizaciones se encuentran en constante cambio y desarrollo, lo que conlleva a la rotación laboral, cuando esta no es prevista o no se conocen las causas que motivan al suceso puede originar pérdida del control afectando en sus actividades e impactando económicamente en la organización. Ante ello, surge la propuesta de entrenar data del personal cesado entre los años 2018 - 2021 de una empresa del rubro de limpieza y desinfección a fin de poder conocer cuáles son los factores que más influyen en la rotación de personal y predecir a tiempo el evento, con el propósito de tomar acciones preventivas ante los posibles sucesos. Al realizar el entrenamiento y corrida de data se obtuvieron resultados positivos, logrando identificar las causas fuerza y disminuir el impacto a través de la marcha del plan de acción.

### HIPÓTESIS (Si aplica)

La aplicación de Técnicas de Machine Learning serán una buena alternativa para identificar cuáles son los factores que influyen en la rotación de personal operativo de una empresa de manufactura del rubro de limpieza y desinfección.

## DISEÑO METODOLÓGICO

**Tipo:** Investigación Practico – Experimental

**Enfoque:** Cuantitativo

**Alcance:** Exploratorio – Explicativo

### Técnicas e Instrumentos:

- Recolección de datos (Planillas de Operarios entre los años 2018 – 2021)
- Herramientas digitales: MS. Excel, Orange, Phyton.

### Etapas del desarrollo de la Investigación:

- Recopilación de la información:  
Se recopiló la información sobre el personal operativo de la empresa que habría cesado entre los años 2018 - 2021
- Identificación del problema:  
Se procede a definir el problema y delimitar el tema de investigación
- Revisión bibliográfica  
A través de la revisión bibliográfica, se logra profundizar el tema e identificar los artículos base para la investigación.
- Definir objetivos  
Se establecen los objetivos generales y específicos.
- Metodología  
Se prepara la data de forma ordenada en Excel para entrenarla en Orange y posteriormente ingresarla al Phyton y hacer correr el modelo.
- Resultados  
Interpretación de las gráficas obtenidas como resultados del modelo.
- Discusión  
Brindar comentarios en base a la investigación, tomando en consideración los puntos de dificultad de la investigación.
- Comentarios finales  
Comentarios y recomendaciones finales en base a los resultados obtenidos de la investigación.

Diagrama de bloques del proceso:



## NOTAS (AGRADECIMIENTOS)

Al culminar esta linda etapa universitaria, agradezco a Dios, ya que sin él no sería posible alcanzar cada uno de los objetivos propuestos. Agradezco también a mis padres Luis y Maribel, por su confianza, motivación y amor incondicional durante todos estos años, a mis abuelos por cada uno de sus consejos y el apoyo incondicional que me brindaron desde el primer día, a Valeria, Naomi, y Renato, por acompañarme durante todo este camino y hacerlo más divertido y ameno.

De igual forma, extender el agradecimiento a mi asesor de tesis, por todo el apoyo y seguimiento continuo, ya que fueron claves para concluir con el trabajo en mención.

## REFERENCIAS

- Amazon, Disponible: <https://aws.amazon.com/es/what-is/logistic-regression/>, consultado el 14 de junio de 2023.
- Aracena, C., Villena, F., Arias, F. y Dunstan, J., Aplicaciones de aprendizaje automático en salud. *Revista Médica Clínica Las Condes*, vol. 33, núm. 6, págs. 568 - 575, 2022.
- Blanchard, W., Clark, R., Hui, F., Tian, R. y Woods, H., Tratando de la separación completa y la separación casi completa en la regresión logística para datos lingüísticos. *El Sevier*, vol. 2 , núm. 1, págs. 1-10, 2023.
- Computrabajo, Aviable: <https://recursos-empresa.computrabajo.com/el-coste-de-la-rotacion/>, Consultado el 13 de agosto de 2023.
- Editorial Ecoprensa, *El índice promedio de rotación laboral llega a 20,7% en las empresas peruanas, 16 de junio de 2019*, Aviable: <https://www.economistaamerica.pe/empresas-eAm-peru/noticias/10257100/12/19 /El-indice-promedio-de-rotacion-laboral-llega-a-207-en-las-empresas-peruanas.html>, Consultado el 21 de julio de 2023.
- Ekawati, AD, Análisis predictivo de la rotación de empleados: una revisión sistemática de la literatura. *Revista de Ciencias de la Decisión y la Información de Gestión*, vol. 22 , núm. 4, págs. 387-397, 2019.
- Evaluar, Aviable: <https://blogs.evaluar.com/por-que-la-rotacion-de-personal-cuesta-tanto>, Consultado el 13 de agosto de 2023.
- García Armendariz, J., Huertas Zúñiga, S., Lizárraga Portugal, CA, Quiroz Flores, JC y García López, YJ, Mejora de la previsión de la demanda mediante la implementación del aprendizaje automático en una empresa de producción avícola. *Repositorio de la Universidad de Lima*, vol. 8 , págs. 39 - 45, 2023.
- García López, YJ, Panduro, J. y Pumayari, S., Reducción de Backorders en el Proceso de Venta Cross Docking para el Servicio de Pedidos de Homecenter. *Repositorio de la Universidad de Lima*, vol. 12 , núm. 7, págs.11, 2022.
- Holwerda, J., ¿Grandes datos? Gran cosa: búsqueda de big data. *El Sevier*, vol. 64, núm. 4, págs. 391-399, 2021.
- King, KG, Análisis de datos en recursos humanos: estudio de caso y revisión crítica. *Revisión del desarrollo de recursos humanos*, vol. 15 , núm. 4, págs. 487-495, 2016.
- Mahesh, B., Algoritmos de aprendizaje automático: una revisión. *Revista Internacional de Ciencia e Investigación (IJSR)*, vol. 9 , núm. 1, págs. 381 - 386, 2020.
- Nocker, M. y Sena, V., Big Data y gestión de recursos humanos: el auge del análisis del talento. *Ciencias Sociales* , vol. 8, núm. 10, págs.273, 2019.
- Omondiaqbe, D., Veeramam, S. y Sidhu, A., Técnicas de clasificación de aprendizaje automático para el diagnóstico del cáncer de mama. *Serie de conferencias IOP: Ciencia e ingeniería de materiales* , vol. 495, págs. 012033, 2019.
- Panduro, JP y García, YJ, Uso de un Modelo de Aprendizaje Automático para la Reducción de Backorders en el Proceso de Ventas Cross Docking para el Servicio de Pedidos Homecenter. *Actas de la Primera Conferencia Internacional Australiana sobre Ingeniería Industrial y Gestión de Operaciones, Sydney* , vol. 12, núm. 7, págs. 2035 - 2045, 2022.
- Russo, C., Ramón, H., Alonso, N., Cicerchia, B., Esnaola, L., and Tessore, JP, Tratamiento masivo de datos utilizando técnicas de Machine Learning. *Repositorio Digital UNNOBA* , págs.131-134, 2016.
- Siddharth, M., Hao, L. y He, J., Caracterización de fracturas no invasivas basada en la clasificación de los tiempos de viaje de las ondas sónicas. *El Sevier* , vol. 4, págs. 243 - 287, 2020.
- Sujeet N., Dev R. y Yogesh P., Análisis predictivo de recursos humanos (HRPA) para la gestión de recursos humanos en las organizaciones. *Revista internacional de investigación científica y tecnológica*, vol. 5 , núm. 5, págs. 33-35, 2016.

## ANEXO. Datos del artículo publicado

- **Nombre del artículo:** Techniques of Machine learning applied to reduce employee turnover in a company cleaning and disinfection
- **Autores:** Erika Noemi Romero Rojas
- **Co autor(es):** Yvan Jesús García López, Juan Carlos Quiroz Flores

**Publicación en revista**

- **Nombre de la revista:** 3rd Indian International Conference on Industrial Engineering and Operations Management
- **Organizador:** IEOM Society International
- **Sede:** Nueva Delhi, India
- **Año:** 2023
- **Pp:** 12 paginas
- **Enlace web donde se encuentra publicado el artículo (identificador DOI, ISBN, ISSN o equivalentes):**
- **ISSN:** 2169-8767
- **ISBN:** 979-8-3507-0545-4

**Presentación en congreso**

- **Nombre del congreso:** 3rd IEOM India Conference, November 2-4, 2023
- **Organizador:** Aligarh Muslim University and Delhi Technological University
- **Sede:** New Delhi, India
- **Año:** 2023

# Paper Erika Romero

---

## INFORME DE ORIGINALIDAD

---

5%

INDICE DE SIMILITUD

5%

FUENTES DE INTERNET

3%

PUBLICACIONES

2%

TRABAJOS DEL  
ESTUDIANTE

---

## ENCONTRAR COINCIDENCIAS CON TODAS LAS FUENTES (SOLO SE IMPRIMIRÁ LA FUENTE SELECCIONADA)

---

3%

★ [repositorio.uwiener.edu.pe](https://repositorio.uwiener.edu.pe)

Fuente de Internet

---

Excluir citas

Apagado

Exclude assignment  
template

Activo

Excluir bibliografía

Activo

Excluir coincidencias

< 15 words