

# CASO DE ESTUDIO DE RECURSOS HUMANOS

Piedrahita Allison; Ramírez Anyi; Vergara María

*Departamento de Ingeniería Industrial, Universidad de Antioquia.*

*Medellín, Colombia.*

**RESUMEN** ~ La rotación de personal representa un gran problema en las organizaciones, puesto que las empresas deben incurrir en gastos de hasta 12 veces el valor del salario del empleado que abandona el puesto. En este caso de estudio, se plantea una solución analítica para mejorar la tasa de retiros de una empresa de 4.000 empleados cuyo índice es de 15% anual. Para ello, se propone una solución analítica, se usan herramientas de machine learning y se recomiendan acciones que contribuyan a mejorar el problema de negocio que presenta la empresa actualmente.

## 1. INTRODUCCIÓN

Una empresa cuenta con alrededor de 4.000 empleados, cuya tasa de retiros es de aproximadamente 15% anual. Los directivos de la empresa consideran que este nivel de retiros es muy alto, ya que trae altos costos para la empresa asociados los procesos de reclutamiento, selección, contratación, capacitación y adaptación. Además, el retiro de las personas conlleva otras implicaciones negativas que se presentan a continuación:

- Proyectos atrasados que afectan satisfacción de clientes.
- Mayor cantidad de personas en reclutamiento.
- Pérdida de conocimiento.
- Empleados con mayores jornadas laborales.

La rotación del personal implica gastos para una empresa de hasta 12 veces el salario de la persona, debido a los procesos de selección, capacitación y adaptación que se deben llevar a cabo, lo cual implica tiempo y dinero irrecuperable. Se evidencia

que las personas abandonan su puesto de trabajo cuando no tienen proyección de crecimiento profesional, cuando no cuentan con ambientes sanos o cuando hay un desequilibrio de asignaciones [1]. Teniendo en cuenta lo anterior, se analizan bases de datos que contienen información de los empleados para establecer una solución analítica, que facilite la identificación y establecimiento de planes de mejora que permitan reducir la tasa de retiros, reduciendo los costos en los que incurre la empresa y las implicaciones negativas que actualmente se presentan en el desarrollo de las demás funciones.

## 2. METODOLOGÍA

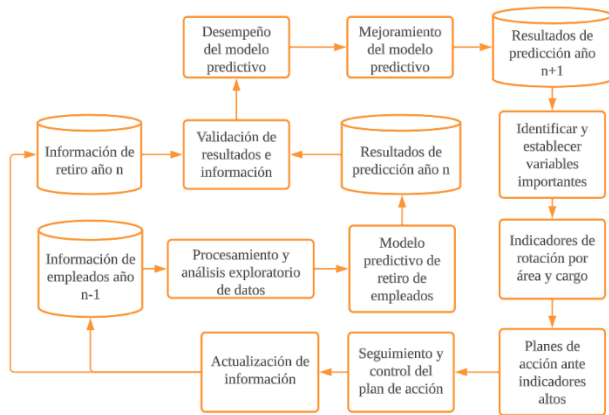
En las siguientes secciones se describe cada uno de los pasos desarrollados.

### 2.1.DISEÑO DE LA SOLUCIÓN

Se plantean los siguientes objetivos para desarrollar el caso de estudio:

- Predecir a tiempo si los empleados presentan posibilidades de abandonar su puesto de trabajo.
- Identificar las variables más significativas para establecer planes de acción preventivos específicos.
- Establecer indicadores para verificar y controlar la tasa de rotación.

Con base a lo anterior, se plantea la solución analítica de la Figura 1, en la cual se evidencian los objetivos anteriores y las condiciones de rezago de la información.



**Figura 1.** Diseño de la solución analítica.

## 2.2.LIMPIEZA Y TRANSFORMACIÓN

Se cuenta con seis bases de datos:

- “*general\_data*”: contiene información general de los empleados del año 2015.
- “*employee\_survey\_data*”: resultados de la encuesta realizada a los empleados en 2015 respecto a su nivel de satisfacción con su empleado actual.
- “*manager\_survey\_data*”: resultados de la evaluación de desempeño realizada en 2015.
- “*in\_time*”: hora de ingreso a la jornada laboral durante 2015. (Para establecer jornada laboral promedio por empleado)
- “*out\_time*”: hora de salida de la jornada laboral durante 2015. (Para establecer jornada laboral promedio por empleado)
- “*retirement\_info*”: información de los empleados retirados para el 2016.

En este paso, se establece una base de datos general que une la información de todas las bases de datos, se imputan valores faltantes y se hacen cambios de variables según el tipo de cada una.

## 2.3.ANÁLISIS EXPLORATORIO

Después de realizar un análisis para comprender las características de los empleados, se observa que:

- Las variables numéricas presentan desviaciones significativas, además, las escalas son muy diferentes, por lo que se vuelve necesaria la

normalización de los datos para evitar sesgos en las predicciones.

- La empresa cuenta con empleados entre los 18 y 60 años, donde el 50% de ellos son mayores de 36 años.
- El 50% de los empleados debe desplazarse mínimo 7km para llegar al trabajo y el 25% de ellos, mínimo 14km.
- El 50% de los empleados ha trabajado en dos o menos compañías.
- Los empleados han recibido un incremento de salario en el 2015 de al menos 11%.
- El 50% de los empleados tienen ingresos de 49.190 dólares mensuales, sin embargo, el 25% de los empleados tienen ingresos de 29.110 dólares mensuales.
- Existen empleados que no fueron capacitados en el año 2015, y el 75% fueron capacitados al menos 2 veces al año.
- El 75% de los empleados lleva al menos 3 años en la empresa.
- Más del 25% de los empleados trabajan en promedio más de la jornada laboral de 8 horas.
- El área de investigación y desarrollo cuenta con cerca del 75% de los empleados. A su vez, alrededor del 75% de empleados rara vez viajan por cuestiones laborales.
- Cerca del 36% de los empleados reportan baja o mediana satisfacción con su trabajo y ambiente laboral.
- Se presentan correlaciones menores al 77% entre variables numéricas.

## 2.4.SELECCIÓN DE ALGORITMOS Y VARIABLES

En primer lugar, se eliminan las variables que no presentan variabilidad y, por lo tanto, información para la construcción del modelo, por ejemplo, el ID del empleado, la jornada laboral estándar, etc. Posteriormente, teniendo en cuenta la importancia de la interpretabilidad de la solución analítica, se plantea el uso de los siguientes algoritmos iniciales

para evaluar sus desempeños: regresión logística, árbol de decisión, bosque aleatorio y extreme gradient boosting (XGB). El desempeño se evalúa bajo tres condiciones de selección de variables: sin selección, con selección bajo la función “SelectFromModel” y con selección bajo el método integrado Lasso. Para evaluar el desempeño, se tienen en cuenta cuatro métricas en la validación cruzada: Accuracy, Precisión, Recall y F1-Score, dando prioridad al Recall por ser una medida que busca determinar los valores correctos de retiros predichos por el modelo y al Accuracy para medir el desempeño general del modelo. Teniendo en cuenta lo anterior, se obtienen los resultados de validación de la Tabla 1.

**Tabla 1.** Desempeño de los métodos y modelos iniciales.

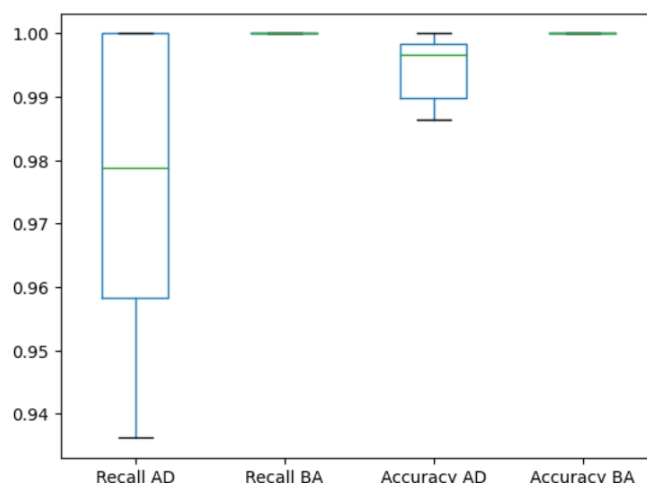
Método	Modelo	Recall	Accuracy	N° de Variables
Sin selección de variables	Regresión logística	76,9%	73,9%	45 variables
	Árbol de decisión	98,2%	99,4%	
	Bosque aleatorio	99,9%	100,0%	
	XGB	99,7%	100,0%	
SelectFormModel (2 veces la media)	Regresión logística	71,5%	70,4%	15 variables
	Árbol de decisión	97,6%	99,4%	
	Bosque aleatorio	100,0%	100,0%	
	XGB	99,3%	99,9%	
Lasso (Alpha de 0.01)	Regresión logística	73,6%	72,8%	12 variables
	Árbol de decisión	96,3%	99,0%	
	Bosque aleatorio	99,6%	99,9%	
	XGB	98,3%	99,7%	

En los resultados anteriores, se evidencia que a pesar de la reducción significativa de las variables en los métodos 2 y 3, se siguen obteniendo variables de desempeño cercanas al 100% en la mayoría de los modelos. Además, se observa que, la mayoría de los modelos presentan mejores métricas en el método 2. Por dichos motivos, se deciden emplear las 15 variables seleccionadas bajo el método SelectFromModel, las cuales son: edad, frecuencia de viajes, distancia de casa, área de trabajo, rol de

trabajo, estado marital, tiempo promedio de jornada, ingresos mensuales, campo de educación, total de años trabajados, años en la compañía, años desde la última promoción y años con el actual jefe.

## 2.5.SELECCIÓN DEL MODELO

El modelo a emplear se selecciona teniendo en cuenta la interpretabilidad, la complejidad y las métricas de desempeño. Teniendo en cuenta lo anterior, se descarta la regresión logística por su bajo desempeño, asimismo, XGB se descarta debido a la pérdida de interpretabilidad que se origina de las bases teóricas de construcción del modelo. Luego, para seleccionar entre el árbol de decisión (AD) y el bosque aleatorio (BA) se realiza el diagrama de la Figura 2, construido a partir de la validación cruzada realizada para cada modelo.



**Figura 2.** Métricas de desempeño de los modelos finales.

En la figura se observa el posible sobreajuste que podría tener el bosque aleatorio respecto al árbol de decisión, además, este último presenta mayor interpretabilidad para definir la estructura de variables que permiten predecir el retiro de un empleado. Por tal motivo, se selecciona el árbol de decisión para desarrollar la solución analítica.

## 2.6.AFINAMIENTO DE HIPERPARÁMETROS

Para mejorar el modelo seleccionado, se hace un afinamiento de hiperparámetros por medio de

método de búsqueda aleatoria. La cuadrícula de hiperparámetros implementada para la búsqueda contiene los siguientes atributos:

- Profundidad máxima del árbol: con variaciones entre 16 y 30 niveles en pasos de 2.
- Máximo de nodos hoja: con variaciones entre 150 y 350 hojas en pasos de 50.
- Cantidad de características a considerar al buscar la mejor división.

Con dicha cuadrícula, se afinan los hiperparámetros teniendo en cuenta 20 iteraciones, la optimización de la métrica Recall y una división de validación cruzada de 20, obteniendo los siguientes resultados:

Tabla 2. Afinamiento de hiperparámetros.

Hiperparámetros	Árbol de decisión (15 variables)
max_leaf_nodes	350
max_features	auto
max_depth	30
Best Score	0,94

Con base a los resultados anteriores, se construye el árbol de decisión, teniendo en cuenta también un criterio de calidad de la decisión por el método Gini y método balanceado por la distribución de la variable objetivo.

2.7.ANÁLISIS DEL MODELO

Con el modelo construido, se evalúan las métricas por medio de la validación cruzada con un divisor CV de 100, y se obtienen los resultados de la Tabla 3 y la Figura 3.

Tabla 3. Resultados Validación Cruzada CV = 100

Métrica	Entrenamiento	Validación
F1 - Score	98,29%	96,34%
Precisión	96,73%	95,45%
Recall	99,94%	97,61%
Accuracy	99,43%	98,78%

En los resultados se evidencia un desempeño óptimo del modelo, con tasas de entrenamiento y validación cercanas a 100% y una distribución de las métricas acorde a los resultados esperados. No

se evidencia sobreajuste debido a que las métricas son similares en entrenamiento y validación.

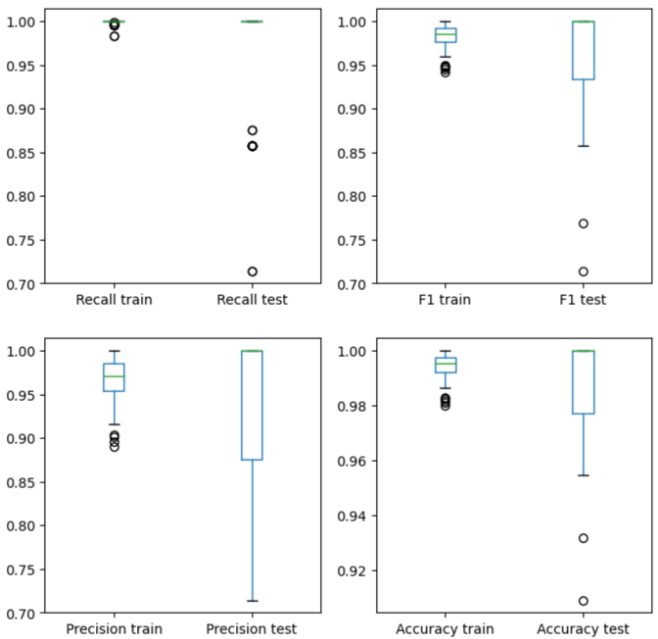


Figura 3. Distribución de las métricas del modelo.

2.8.DESPLIEGUE DEL MODELO

Para la implementación del modelo construido en la gestión de talento humano, se tienen en cuenta los siguientes aspectos:

- El modelo se entrenará de manera anual, con la información de los empleados del año anterior y la información de retiros del año actual, con el fin de hacer predicciones para el año siguiente con la información de empleados del año actual.

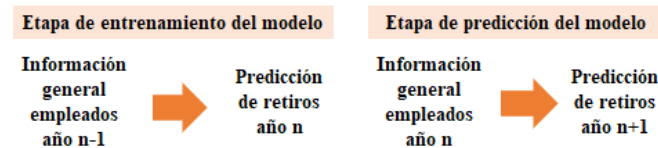


Figura 4. Flujo de información de la solución analítica.

- El modelo entregará una base de datos con las predicciones de los empleados que tienen posibilidad de retirarse en el año siguiente.
- Se entregará un reporte anual con las variables que más impactan en la decisión de retiro de un empleado, de manera que se construyan planes de acciones preventivos enfocados a dichas

variables y se planteen políticas efectivas de retención.

- Se entregará un reporte anual del porcentaje de retiros por área y cargo, y se entregan recomendación sobre acciones pertinentes.

### 3. RESULTADOS

Para el periodo de evaluación del modelo, se obtienen el siguiente nivel de importancia de las variables y los indicadores por área y cargo.

**Tabla 4.** Nivel de importancia de las variables.

Variable	Importancia
monthlyincome	17,73%
mean_time	17,23%
age	12,43%
yearsatcompany	11,31%
distancefromhome	10,99%
totalworkingyears	8,81%
yearssincelastpromotion	6,24%
yearswithcurrmanager	4,19%
maritalstatus_Single	2,98%
jobrole_Research Director	2,00%
businesstravel_Travel_Frequently	1,97%
businesstravel_Non-Travel	1,82%
jobrole_Manufacturing Director	1,10%
department_Human Resources	1,00%
educationfield_Human Resources	0,19%

**Tabla 5.** Indicadores de rotación por área.

Área	Rotación real	Rotación predicción
Human Resources (HR)	30,16%	29,63%
Research&Development (RD)	15,71%	15,75%
Sales (S)	15,02%	15,40%

**Tabla 6.** Indicadores de rotación por cargo.

Cargo	Rotación real	Rotación predicción
Healthcare Representative	14,50%	14,50%
Human Resources	13,46%	13,46%
Laboratory Technician	16,22%	15,44%
Manager	13,73%	13,73%
Manufacturing Director	11,03%	11,03%
Research Director	23,75%	23,75%
Research Scientist	18,15%	18,38%
Sales Executive	16,87%	17,48%
Sales Representative	14,46%	15,66%

### 4. CONCLUSIONES

De los resultados obtenidos, se evidencia que:

- Es necesario establecer políticas de desconexión laboral, puesto que desde la etapa de exploración se evidencia que el 25% de los empleados trabaja más de 8 horas diarias.
- Los ingresos mensuales es un factor importante de rotación, se evidencia que el 25% de los empleados presentan ingresos inferiores a la mediana y a la media de toda la población, por ende, se deben evaluar políticas de incentivos monetarios o extras para aumentar la motivación.
- La distancia recorrida de la casa al trabajo puede ser un factor decisivo, el 25% de los empleados debe desplazarse mínimo 14km, por lo tanto, se debería evaluar políticas de home office o beneficios como rutas propias para el transporte de los empleados.
- La edad es fundamental, se deben evaluar políticas de acuerdo a la edad de los empleados, enfocados a la formación, integración familiar, incentivos por años laborados en la empresa, acompañamiento psicosocial, etc.
- Se debe evaluar el ambiente laboral de manera frecuente, y realizar actividades de integración, pausas activas, mejoramiento de relaciones interpersonales, etc.
- Actualmente, se presentan altos índices de rotación en el área de recursos humanos, probablemente por el aumento de trabajo, es necesario establecer límites de tiempo y asignación para esta área.
- Los directores de investigación, científicos de investigación, ejecutivos de ventas y técnicos de laboratorio presentan las mayores tasas, es necesario evaluar sus funciones y satisfacción con el rol que desempeñan en su cargo.

### 5. REFERENCIAS

[1]. A. Ramírez Peña, “El costo de la alta rotación de personal para las empresas”, *Portafolio*, 18 de diciembre de 2016. Accedido el 7 de septiembre de 2023. [En línea]. Disponible: <https://www.portafolio.co/economia/empleo/costos-de-la-alta-rotacion-de-personal-en-las-empresas-502333>