

# APLICACIONES DE LA ANALÍTICA EN RECURSOS HUMANOS

Estudiantes: Laura Betancourt, Aura Molina, Gilberto Gil

## 1. Problema de negocio:

La empresa enfrenta una **alta rotación** de personal del 15% anual, lo que eleva los costos de contratación, provoca pérdida de conocimiento y retrasa proyectos clave, afectando la productividad y la satisfacción de los clientes. El objetivo es **reducir esta tasa** para optimizar costos, mejorar la eficiencia operativa y retener talento clave.

## 2. Problema analítico:

Se desarrolla un modelo de clasificación para **predecir la renuncia** de empleados en el próximo año. Se **identifican las variables clave** relacionadas con las renunciaciones para implementar acciones preventivas y reducir la tasa de rotación. El modelo se actualiza anualmente para mantener su precisión y relevancia.

## 3. Diseño de la solución:

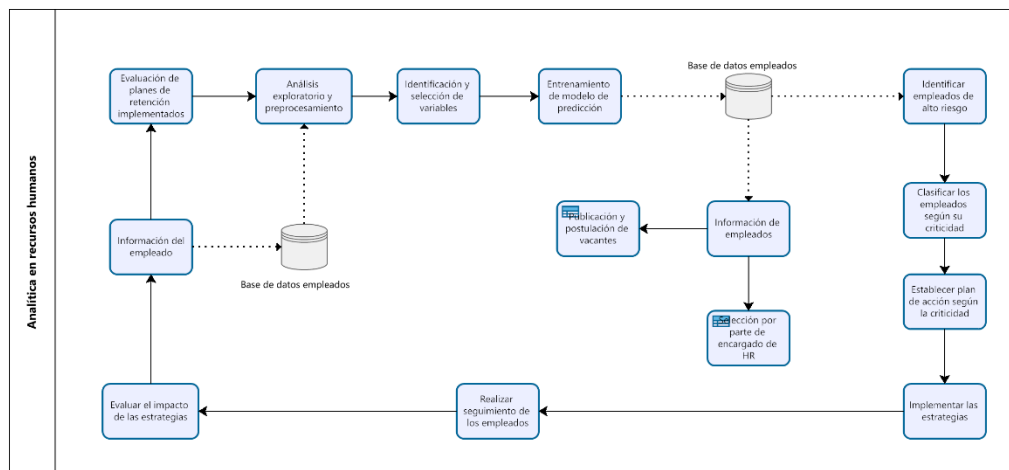


Figura 1. Diseño de la solución  
Fuente: Elaboración propia

## 4. Diseño Técnico:

Se va a entrenar un modelo de clasificación que predice las potenciales renunciaciones de empleados el siguiente año (2017), mediante la identificación de variables claves.

Los empleados se clasificarán según su probabilidad de abandonar en no crítico, crítico según las variables seleccionadas.

Las variables que se utilizarán para el modelo.

$$y_{pred} = \text{análisis de cada empleado } 2017 (t + 1)$$

$y_{real} = \text{desempeño de cada empleado 2016 } (t + 1)$

$x = \text{Info de cada empleado 2016 } (t)$

$x = \text{Info de cada empleado 2015 } (t)$

Los resultados obtenidos se entregarán en una base de datos a RRHH para la creación de estrategias de retención de los empleados. Además, el modelo se re-entrena cada año, después de realizada la encuesta de satisfacción a los empleados, para posteriormente realizar las predicciones.

## 5. Solución analítica:

Se implementará un modelo de Machine Learning para predecir la renuncia de empleados en el año 2017. Los resultados obtenidos serán utilizados para tomar medidas preventivas que reduzcan la deserción de empleados en la empresa. Se identificarán variables claves relacionadas con las renunciaciones, con el objetivo de disminuir la tasa de rotación que afecta a la organización. El modelo deberá ser actualizado cada año para mantener su precisión.

## 6. Selección de algoritmos:

En un inicio, para abordar la predicción del retiro de empleados, se contempló la utilización de cuatro modelos de machine learning: Regresión Logística, Árbol de Decisión, Random Forest y Gradient Boosting Classifier. Estos modelos fueron seleccionados debido a sus características y su capacidad para resolver problemas de clasificación:

- **Regresión Logística:** elegido por su simplicidad y capacidad para ofrecer interpretaciones claras sobre cómo variables como el ingreso o la distancia al hogar influyen en la probabilidad de renuncia.
- **Decision Tree:** seleccionado por su habilidad para capturar relaciones no lineales, comunes en este contexto, y por su fácil interpretación. Permite reflejar interacciones complejas, como la relación entre la satisfacción laboral y los años desde la última promoción.
- **Random Forest:** contemplado como una mejora del Árbol de Decisión, reduce el sobreajuste, un desafío frecuente en conjuntos de datos complejos con múltiples predictores. Este modelo es especialmente relevante para manejar la diversidad de factores que pueden influir en la decisión de un empleado de dejar la empresa.
- **Gradient Boosting:** seleccionado por su capacidad para ofrecer un alto rendimiento en clasificación al combinar múltiples modelos débiles, lo que permite capturar patrones ocultos y complejos entre las variables, como las interacciones entre la satisfacción con el entorno y la frecuencia de viajes de trabajo. Esto lo hace adecuado para abordar el problema multifactorial del retiro de empleados.

## 7. Selección de algoritmos

Se realizó una selección de variables para reducir el conjunto de datos a las características más relevantes para la predicción, se llegó a un total de 13 variables, donde destacan la edad, la satisfacción con el entorno, el ingreso mensual, estado civil: soltero, años en la empresa, número de empresas en las que se ha trabajado, entre otras.

Se entrenaron los modelos y se calculó el F1-score, obteniendo los mejores resultados con el Decision Tree y el Gradient Boosting, con puntajes de 0.838651 y 0.790158 respectivamente. Posteriormente, se realizó una selección de variables clave, entrenando nuevamente los modelos, y se observó que ambos mantuvieron un rendimiento similar, con F1-scores de 0.836950 y 0.790158. Esto confirma que estos dos modelos son los más adecuados para predecir el retiro de empleados.

## 8. Prueba de sobre ajuste:

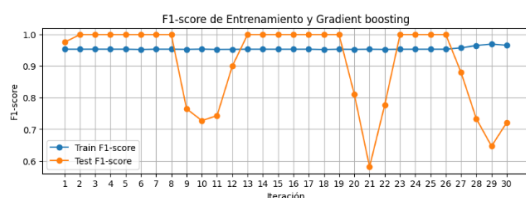


Figura 2. Prueba sobre ajuste Gradient Boosting

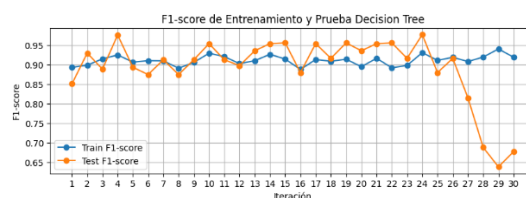


Figura 3. Prueba sobre ajuste Decision Tree

En el modelo de **Gradient Boosting**, el F1-score se mantiene constantemente en 1 para los datos de entrenamiento, lo que indica un claro **sobreajuste**. Sin embargo, en los datos de prueba, el F1-score muestra fluctuaciones significativas, con caídas notables en las iteraciones 9, 10, 19, 21 y 26, lo que sugiere que el modelo es muy sensible a pequeñas variaciones.

Por otro lado, en el **Decision Tree**, el F1-score es estable y alto en los datos de entrenamiento, y en los datos de prueba se mantiene relativamente constante, con algunos pequeños picos antes de una caída en la iteración 27. A pesar de esto, el modelo tiene un buen rendimiento general, lo que lo convierte en la mejor opción.

## 9. Desempeño Decision Tree

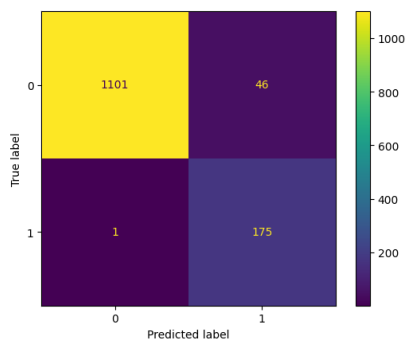


Figura 4. Desempeño Decision Tree

Accuracy test: 0.9644746787603931  
 Precisión: 0.9990925589836661  
 Recuperación (Sensibilidad): 0.959895379250218  
 F1-score: 0.8816120906801007  
 Especificidad: 0.9943181818181818

Figura 5. Resultados medida de desempeño

Para la empresa es prioridad reducir los falsos negativos debido a que estos generarían más costos a la empresa, pues predecir que un empleado no va a renunciar cuando en realidad si lo hará, conlleva que la empresa no tome medidas preventivas para la retención del empleado, lo que conduce a que se deba llevar procesos de contratación de un nuevo empleado. Aunque los falsos positivos también influyen en un costo para la empresa son menos riesgosos, pues no están ligados a rotación de personal.

El modelo muestra un sólido desempeño en la predicción de renunciaciones, con una precisión de 99.91% y un recall de 95.99%, lo que asegura que identifica correctamente la mayoría de los empleados que renunciarán y tiene muy pocos falsos positivos. La especificidad del 99.43% también indica que el modelo predice con precisión a los empleados que no renunciarán. Aunque el F1-score es un poco más bajo (88.16%), sigue reflejando un buen equilibrio entre precisión y recall, lo que lo convierte en una herramienta confiable para tomar medidas preventivas.

#### **10. Despliegue del modelo:**

Tras el ajuste de hiperparámetros del modelo, se realizó el despliegue del modelo para predecir los retiros de empleados en 2016. Los resultados indicaron que se prevé que 766 empleados se retiren de la empresa, lo que representa un aumento en la tasa de retiros al 17.4%, en comparación con el 15% anual previo, y umbral establecido. Este aumento resalta la necesidad de tomar medidas proactivas para abordar las razones detrás de estas decisiones.

Los pesos de las variables en el modelo revelan que el ingreso mensual (0.164) y la edad (0.124) son los factores más influyentes en la decisión de los empleados de dejar la empresa. La satisfacción laboral y del entorno de trabajo también son relevantes, aunque con menor peso (0.056 y 0.051, respectivamente). Esto demuestra que prestar atención a la compensación y al bienestar en el entorno laboral es crucial para reducir la tasa de retiros, y que las dinámicas demográficas deben ser áreas prioritarias de intervención. Además, la satisfacción laboral y del entorno de trabajo tienen un peso significativo, lo que sugiere que mejorar estos aspectos podría tener un impacto positivo en la retención de empleados.

Con base en los hallazgos anteriores, se proponen diversas estrategias enfocadas en las variables más influyentes del modelo. Primero, dado que el ingreso mensual se identifica como un factor crítico, es vital llevar a cabo un análisis exhaustivo de la estructura salarial y los beneficios ofrecidos. Asegurarse de que los salarios sean competitivos y que reflejen adecuadamente las contribuciones de los empleados puede ayudar a mitigar la insatisfacción y la intención de abandonar la empresa.

Además, dado que la edad también juega un papel importante en la retención, es esencial considerar las necesidades específicas de los empleados según su etapa de vida y carrera. Esto podría incluir programas de mentoría y desarrollo profesional que se adapten a diferentes grupos de edad, brindando así oportunidades de crecimiento y avance.

La satisfacción laboral y la satisfacción con el entorno de trabajo son igualmente relevantes, lo que implica que implementar encuestas periódicas para evaluar estos aspectos y realizar acciones concretas basadas en los resultados puede mejorar notablemente el ambiente laboral. Iniciativas como talleres de bienestar y actividades de integración pueden fomentar una cultura organizacional más positiva y colaborativa.

Finalmente, al considerar las dinámicas demográficas, se sugiere establecer políticas de trabajo flexible que se adapten a las diferentes situaciones personales de los empleados, lo que podría contribuir a un equilibrio más saludable entre la vida laboral y personal. Al abordar de manera integral estos factores, la empresa podrá crear un entorno más atractivo que fomente la retención y minimice la rotación de personal.