

# Informe de Progreso:

## 1. Creación del DataSet

Desarrollamos un archivo en python para la generación y estructuración de un conjunto de datos ficticio con el fin de entrenar un modelo de aprendizaje automático. Este dataset contiene la siguiente información ,generada de forma aleatoria, sobre los empleados que consideramos conveniente: edad, antigüedad, departamento y la preferencia de notificación. El preprocesamiento de datos incluyó la conversión de variables categóricas a formato numérico mediante el objeto OneHotEncoder y LabelEncoder para su correcta interpretación por el modelo de machine learning.

## 2. Entrenamiento del Modelo de Árbol de Decisión

Utilizamos *Árboles de decisión* de la librería scikit-learn para construir un modelo de clasificación. Creemos que este algoritmo es óptimo para este proyecto, permite visualizar de forma clara el camino que toma los datos para la clasificación, y es más flexible al número de datos de entrada. Las acciones clave en esta fase fueron:

- División de los datos en conjunto de entrenamiento (70%) y prueba (30%).
- Aplicación de transformaciones a las variables categóricas y numéricas.
- Definición de parámetros, incluyendo la profundidad máxima del árbol
- Entrenamiento del modelo con los datos procesados.

## 3. Evaluación del Modelo

Para medir el desempeño del modelo, nos guiamos de las siguientes métricas:

- Precisión global
- Reporte de clasificación con métricas por categoría (precisión, recall, F1-score).
- Matriz de confusión para visualizar la relación entre predicciones y valores reales.
- Análisis de la profundidad del árbol y número de nodos.

La precisión del modelo fue el adecuado para el problema, por lo que el árbol de decisión es capaz de identificar correctamente las preferencias de notificaciones, esto ocurre en la mayoría de los casos.

## 4. Implementación de Predicciones

- Una vez entrenado el modelo, desarrollamos una función que permite:
- Generar nuevos datos de entrada con atributos específicos recibidos de un input.
- Aplicar el modelo para predecir la preferencia de notificación de un empleado.
- Visualizar la probabilidad de cada posible clasificación.

## 5. Visualización del Árbol de Decisión

Se incluyó la función *plot\_tree()* para graficar la estructura del árbol generado, lo que permite interpretar visualmente cómo el modelo toma sus decisiones.