



# Módulo 4

## Sesión N° 2



### ACTIVIDAD:



## Ajustes de Hiperparámetros en Clasificación Médica

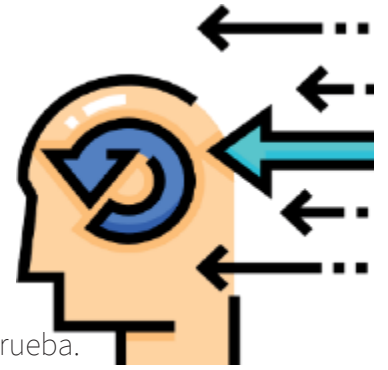
- Objetivo: Construir un modelo de clasificación para predecir la probabilidad de diabetes en pacientes usando el conjunto de datos Pima Indians Diabetes Dataset, aplicando Grid Search y Random Search para optimizar el rendimiento del modelo y comparar sus resultados en términos de precisión, F1-score y eficiencia computacional.



### Instrucciones:

1. Carga y exploración de datos:
  - Descargar y explorar el dataset desde [este enlace](#).
  - Describir brevemente las variables y su relación con la predicción de diabetes.
2. Preprocesamiento:
  - Escalar variables numéricas con StandardScaler.
  - Dividir en conjunto de entrenamiento y prueba (ej. 70/30).
3. Modelo base:
  - Entrenar un RandomForestClassifier sin optimización.
  - Evaluar el modelo con métricas como accuracy, recall, F1-Score, y AUC.
4. Aplicar Grid Search:
  - Definir un grid razonable de 2-3 hiperparámetros (n\_estimators, max\_depth, min\_samples\_split).
  - Evaluar el rendimiento y registrar la mejor combinación.
5. Aplicar Random Search:
  - Definir un rango más amplio para los mismos hiperparámetros.
  - Ejecutar al menos 10 iteraciones y guardar los mejores resultados.





6. Comparación y análisis:

- Evaluar ambos modelos (Grid y Random) sobre el mismo conjunto de prueba.
- Comparar métricas y tiempos de entrenamiento.

7. Reflexión final:

- ¿Cuál técnica fue más eficiente?
- ¿Cuál encontró el mejor modelo?
- ¿Qué hubieras hecho diferente?

Entrega:

- Formato ejecutable (Google Colab o Jupyter Notebook).
- Modalidad: Individual.
- Tiempo: 120 min.

Anexo: Recursos recomendados

- [Dataset original](#)
- [Documentación GridSearchCV](#)
- [Documentación RandomizedSearchCV](#)

