



Módulo 4

Sesión N° 1



ACTIVIDAD:



Predicción de Diabetes con Ajuste de Hiperparámetros

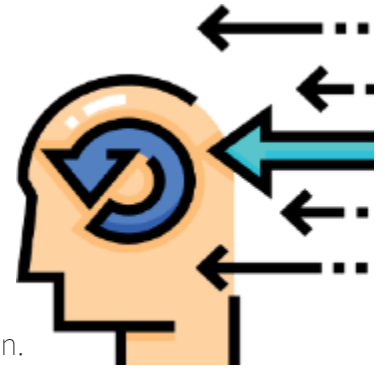
- Objetivo: Desarrollar un modelo de clasificación para predecir la presencia de diabetes tipo II utilizando el Pima Indians Diabetes Dataset, aplicando y comparando técnicas de ajuste de hiperparámetros: Grid Search, Random Search y Optimización Bayesiana.



Instrucciones:

- Cargar y explorar el dataset:
 - Utilizar el dataset Pima disponible públicamente (<https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.csv>).
 - Realizar una breve exploración del conjunto de datos: dimensiones, tipos de variables, valores faltantes o anómalos.
- Preprocesamiento:
 - Escalar las variables numéricas.
 - Dividir en conjunto de entrenamiento y prueba (por ejemplo, 80/20).
- Entrenamiento inicial del modelo base:
 - Usar un Random Forest sin ajuste de hiperparámetros para establecer una línea base.
 - Evaluar el modelo utilizando F1-Score, Precisión, Recall y AUC.
- Aplicar técnicas de optimización:
 - Implementar al menos dos de las siguientes técnicas:
 - Grid Search
 - Random Search
 - Optimización Bayesiana con Optuna
 - Registrar el mejor conjunto de hiperparámetros obtenido con cada técnica.





5. Comparar resultados:

- Comparar el rendimiento del modelo antes y después de la optimización.
- Analizar cuál técnica fue más efectiva en términos de resultados obtenidos y tiempo de ejecución.

6. Visualización y reflexión:

- Crear gráficos simples para mostrar el impacto de los hiperparámetros sobre el rendimiento.
- Reflexionar sobre la importancia del ajuste y cómo podría cambiar en contextos con mayor volumen o complejidad de datos.

Entrega:

- Formato ejecutable (Google Colab o Jupyter Notebook).
- Modalidad: en pareja.
- Tiempo: 100 min.

Anexo: Recursos recomendados

- [Scikit-learn: GridSearchCV](#)
- [Scikit-learn: RandomizedSearchCV](#)
- [Optuna – Optimización Bayesiana](#)

