



Módulo 4

Sesión N° 3



ACTIVIDAD:



Optimización de Modelos en Salud usando Técnicas Bayesianas

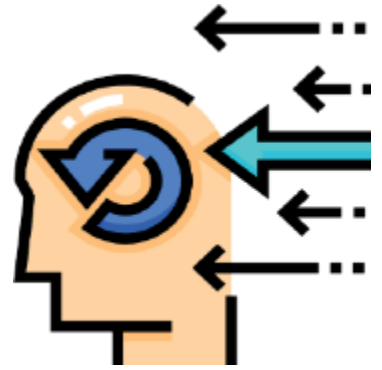
- Objetivo: Aplicar la Optimización Bayesiana para ajustar los hiperparámetros de un modelo de clasificación binaria (Random Forest) sobre un problema de salud pública, comparando dos enfoques populares: Scikit-Optimize (skopt) y Hyperopt, evaluando el rendimiento del modelo y la eficiencia de cada técnica.



Instrucciones:

1. Cargar y preparar los datos:
 - Utiliza el dataset de cáncer de mama (load_breast_cancer) de Scikit-learn.
 - Aplica StandardScaler para escalar las variables.
 - Realiza la división en conjunto de entrenamiento y prueba (70/30).
2. Entrenar modelo base:
 - Implementa un modelo RandomForestClassifier sin ajuste de hiperparámetros.
 - Evalúa usando classification_report y F1-Score.
3. Aplicar Optimización Bayesiana – Parte A (Scikit-Optimize):
 - Define un espacio de búsqueda para n_estimators, max_depth y min_samples_split.
 - Ejecuta BayesSearchCV con cv=3 y scoring='f1'.
 - Evalúa resultados en el test set.
4. Aplicar Optimización Bayesiana – Parte B (Hyperopt):
 - Define el mismo espacio de búsqueda con hp.quniform.
 - Utiliza fmin() con algoritmo tpe.suggest.
 - Entrena un modelo final con los mejores hiperparámetros encontrados.





5. Comparar y reflexionar:

- Contrasta F1-Score, tiempo de ejecución y claridad de resultados.
- Comenta qué técnica fue más efectiva y por qué.

6. Documentación y presentación:

- Estructura tu notebook con secciones comentadas.
- Incluye conclusiones sobre el uso de Optimización Bayesiana frente a técnicas tradicionales.

Entrega:

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

Anexo: Recursos recomendados

- [Scikit-Optimize – Documentación](#)
- [Hyperopt – Documentación oficial](#)
- Dataset: `from sklearn.datasets import load_breast_cancer`

