Al Machine Learning - 3

Kevin Javier Castillo Hernández

Actividad 1: Construccion de un modelo de Machine Learning utilizando un algoritmo avanzado

**Informe: Sesgo Algorítmico en Modelos de IA Bancarios – El Caso Apple Card**

**1. Introducción**

En esta actividad se desarrolló un modelo de Machine Learning para la clasificación binaria del conjunto de datos **Breast Cancer Wisconsin**. El objetivo principal fue maximizar la precisión en la predicción de tumores malignos y benignos utilizando un algoritmo avanzado.

**2. Selección del Dataset**

Se utilizó el conjunto de datos **Breast Cancer Wisconsin**, disponible en la librería sklearn. Este dataset es ampliamente utilizado para tareas de clasificación binaria en el ámbito médico, ya que contiene características numéricas extraídas de imágenes de tumores, y un objetivo binario que indica si el tumor es maligno (1) o benigno (0).

**3. Eleccion del algoritmo**

Se eligió el algoritmo **XGBoost (Extreme Gradient Boosting)** por las siguientes razones:

* Es un método basado en árboles de decisión que ofrece alta precisión y eficiencia.
* Es robusto frente al sobreajuste gracias a técnicas de regularización incorporadas.
* Maneja bien datos tabulares y relaciones no lineales entre características.
* Tiene un buen soporte en Python y una amplia comunidad.

**4. Implementacion del modelo**

El pipeline de modelado consistió en los siguientes pasos:

1. **Carga de datos:** Se cargaron las características y etiquetas directamente desde sklearn.datasets.
2. **División de datos:** Se dividió el conjunto en entrenamiento (80%) y prueba (20%) con una semilla fija para reproducibilidad.
3. **Entrenamiento:** Se entrenó un clasificador XGBoost con parámetros por defecto y evaluación logloss.
4. **Evaluación:** Se calcularon las métricas de precisión (accuracy), matriz de confusión y reporte detallado (precisión, recall, F1-score).

**5. Resultados**

* **Accuracy obtenido:** Aproximadamente 0.96 (96%), lo que indica que el modelo clasifica correctamente la mayoría de los casos.
* **Matriz de confusión:** Muestra un bajo número de falsos positivos y falsos negativos, lo que es crucial para aplicaciones médicas.
* **Reporte de clasificación:** Proporciona métricas detalladas que confirman la buena performance del modelo tanto en la clase positiva como en la negativa.

Imagen que contiene Aplicación

El contenido generado por IA puede ser incorrecto.

**6. Conclusión**

El modelo XGBoost demostró ser altamente efectivo para la clasificación de tumores en el dataset Breast Cancer Wisconsin. Gracias a su capacidad para manejar características complejas y su robustez, se obtuvo una precisión cercana al 96%, un resultado adecuado para aplicaciones iniciales en diagnóstico médico.

Para trabajos futuros se podrían considerar:

* Ajustar hiperparámetros para mejorar aún más la precisión.
* Aplicar técnicas de validación cruzada.
* Evaluar otros algoritmos como SVM o redes neuronales para comparar desempeño.