Unidad	Unidad 1: Interpretabilidad de algoritmos
Entrega	Método de entrega de la actividad en el campus virtual

1. Enunciado

En esta actividad, deberéis seleccionar

a) Seleccionar un conjunto de datos real de su preferencia, por ejemplo de la plataforma Kaggle o UCI Machine Learning Repository (ver **datasets no admitidos**), y un modelo de aprendizaje automático que sea interpretable según lo estudiado en clase,
b) realizar el análisis exploratorio de datos (EDA) del dataset seleccionado y
c) explorar la interpretabilidad del modelo.

El objetivo principal es aprender a evaluar y mejorar la interpretabilidad del modelo utilizando diversas métricas y métodos de evaluación.

Además, la realización de esta actividad tiene los siguientes objetivos secundarios:

- Realizar un EDA detallado para entender las características y patrones del conjunto de datos seleccionado, identificar anomalías y preparar los datos para el entrenamiento del modelo.
- Entrenar y evaluar el modelo de aprendizaje automático interpretable seleccionado
- Aplicar métricas de rendimiento estándar y
- Evaluar la interpretabilidad del modelo y de sus predicciones, comprendiendo la importancia de las métricas de interpretabilidad.
- Documentar y comunicar de manera efectiva sus procesos y hallazgos, presentando visualizaciones, análisis y resultados detallados y
- Participar en discusiones constructivas para recibir y proporcionar retroalimentación.

2. Pasos para realizar la actividad

- Selección de un conjunto de datos admisible (ver **datasets no admitidos**) que tenga características diversas y sea relevante para el análisis.
- Selección del modelo de aprendizaje automático interpretable apropiado para el análisis del conjunto de datos.
- Realiza un Análisis exploratorio de datos (EDA) detallado del conjunto de datos seleccionado, incluyendo limpieza de datos, análisis de estadísticas descriptivas, visualización de distribuciones de características y detección de posibles anomalías.
- Documenta los resultados y prepara visualizaciones que ayuden a entender las características principales del conjunto de datos.
- Entrena el modelo de aprendizaje automático con el conjunto de datos preparado.
    - Evalúa el rendimiento inicial del modelo utilizando métricas estándar.

- Evaluación de la interpretabilidad aplicando, por ejemplo, gráficos de pesos y efectos, importancia general de las características, etc.
- Documenta todo el proceso, desde el EDA hasta la evaluación de la interpretabilidad. Incluye visualizaciones, resultados y análisis detallados.
- Prepara una presentación para compartir tus hallazgos con la clase. La presentación debe destacar las técnicas utilizadas, los resultados obtenidos y las posibles mejoras en la interpretabilidad del modelo.
- Discusión y retroalimentación: Presenta tu trabajo a la clase y participa en una discusión abierta sobre los desafíos y hallazgos del análisis de interpretabilidad.

3. Detalles de la entrega

Documentos a entregar a través de la plataforma:

- Informe sobre el análisis exploratorio de los datos llevado, detallando proveniencia del conjunto de datos (página web), motivación para su selección, gráficos y tablas obtenidos como parte del análisis acompañados del correspondiente análisis y conclusiones extraídas.
- Código del programa utilizado para el EDA en formato (.py, .ipynb, etc.) que facilite su ejecución mediante el IDE adecuado
- Informe sobre la selección del modelo de aprendizaje automatizado interpretable, explicando los diferentes pasos realizados para su entrenamiento y la evaluación de sus métricas de desempeño y de interpretabilidad.
- Código del programa utilizado para entrenar y evaluar el modelo en formato (.py, .ipynb, etc.) que facilite su ejecución mediante el IDE adecuado
- Presentación para compartir sus hallazgos con la clase, destacando las técnicas utilizadas, resultados obtenidos y posibles mejoras en la interpretabilidad del modelo

Subir de forma individual los documentos a la actividad 1 del campus virtual.

Recuerda

Al ser un ejercicio puntuable, se evaluarán los aspectos ya señalados. Además, se puntuarán muy firmemente los siguientes:

- La originalidad: es decir, las copias de Internet, de un compañero o de otra bibliografía no serán válidas.
- La simplicidad: enunciados claros y soluciones fáciles de seguir.
- La solución: no es suficiente con unos buenos ejercicios si estos están mal resueltos.
- La planificación: entregas a tiempo y consulta con el profesor la evolución de la tarea.

Es importante dedicarle el tiempo necesario a cada uno de los ejercicios, ya que esta tarea, en su totalidad, es puntuable para la nota final de la materia.

4. Anexo

Criterios para seleccionar los conjuntos de datos:

a) Deben ser conjuntos de datos provenientes de la realidad cotidiana (investigaciones, encuestas) tomados de páginas de los gobiernos con estadísticas, de competencias de Kaggle u otras similares, repositorios como UCI, etc.)
b) No se aceptarán conjuntos de datos obtenidos de forma artificial mediante código creado para generarlos como se acostumbra para hacer demostraciones en las páginas web y ejemplos de cursos.
c) Los conjuntos de datos deben tener más de 1000 muestras.
d) **No** se aceptarán para las tareas ninguno de los siguientes conjuntos de datos o versiones similares
   - scikit-learn
   - MNIST
   - California Housing
   - Iris flower
   - Breast cancer
   - Diabetes
   - cats vs dogs
   - Heart disease
   - Titanic (cualquier versión de este dataset)
     - https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv
     - https://www.kaggle.com/competitions/titanic/data
     - https://www.kaggle.com/datasets/yasserh/titanic-dataset

Bibliografía:

1. Molnar, C. (2025). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (3rd ed.), christophm.github.io/interpretable-ml-book/
2. Molnar, C. (2021) Aprendizaje automático interpretable: Una guía para hacer que los modelos de caja negra sean explicables. https://fedefliguer.github.io/AAI/