POSTWORK
SESIÓN 05

Objetivo:
Construir un algoritmo de Machine Learning de árboles de decisión para clasificación.

Si ya tienes un proyecto:

- Evalúa si tu proyecto puede resolverse con un árbol de decisión o un random forest: Tu problema puede resolverse con un solo árbol de decisión si puedes pensar en varios if / else que podrían clasificar tu problema en una de las dos clases. Este es un clasificador muy sencillo, y puedes intentar con un solo árbol de decisión.

- Si tu problema pareciera ser más difícil de clasificar que esto, puedes probar con múltiples árboles de decisión en un random forest para ver si la clasificación mejora. Una buena clasificación es aquella que tiene más del 90% de precisión en el dataset de prueba. Una excelente clasificación es aquella que tiene más del 95% de precisión.

- Un problema de clasificación requiere forzosamente una bitácora de pruebas. Aunque muchos no la usan, te recomiendo encarecidamente que lleves una, porque es fácil que te frustres y repitas experimentos innecesariamente.

- Te recomiendo que empieces con pocos árboles en tu random forest, si llegas a necesitarlos. Y también te recomiendo que utilices valores impares para la cantidad de árboles. Puede ser un caso realmente extraño, pero ¿qué pasa si en una clasificación binaria, la mitad de árboles se va para positivo y la mitad para negativo? para evitar eso, siempre intenta con al menos N + 1 árboles, donde N es el número de clases.

Si no tienes un proyecto y deseas crear uno:

- Existe un dataset que trata el análisis químico de vinos que han crecido en una misma región en Italia, por tres diferentes vinicultores (los llamaremos vinicultor 0, vinicultor 1 y vinicultor 2).

- Las características que se tomaron del vino fueron las siguientes:

  - Cantidad de alcohol
  - Cantidad de ácido málico
  - Cantidad de ceniza
  - Cantidad de la alcalinidad de la ceniza
  - Cantidad de magnesio
  - Total de fenoles
  - Total de flavonoides
  - Total de fenoles no-flavonoicos
  - Total de Protocianinos
  - Intensidad del color
  - Tonalidad
  - OD280 / OD315 de vinos diluídos
  - Cantidad de Prolina

- En total tienes 178 muestras, 50 en cada una de las tres clases. Propón un random forest para clasificar las muestras del vino, por medio de un dataset de pruebas.

- Para cargar y utilizar el dataset, utiliza el siguiente código:

```
from sklearn.datasets import load_wine

dataset = load_wine()
x, y = dataset.data, dataset.target
```