

## Tarea 3

### Problema

Considere al conjunto de datos Adult<sup>1</sup>, un conjunto de datos con 48842 instancias, 14 atributos de entrada y uno de salida.

Se pide:

- a) Describa las principales características del conjunto de datos, incluyendo, por ejemplo, valores posibles para cada atributo, valores faltantes, distribución de la clase objetivo. Esto tiene como objeto entender el dataset con el que se está trabajando.
- b) Describa cómo dividirá el conjunto para el aprendizaje, sabiendo que deberá evaluar sobre un dataset de evaluación, y que la validación debe hacerse utilizando validación cruzada. Indique si es relevante utilizar estratificación o no, justificando.
- c) Implemente el algoritmo KNN (vecinos más cercanos) visto en el teórico. Evalúe sus resultados sobre el conjunto Adult.
- d) Implemente el algoritmo Naive Bayes visto en teórico. Evalúe sus resultados sobre el conjunto Adult.
- e) Aplique las implementaciones de Scikit-learn para (d) y (e) sobre Adult y compare los resultados con los de sus implementaciones.
- f) Presente los resultados de evaluación, incluyendo medidas de accuracy, precision, recall y medida-F.
- g) Comente brevemente los resultados obtenidos. Compare con otros resultados reportados en el mismo conjunto de datos.

Observaciones:

- Está permitido (y recomendado) utilizar bibliotecas Python para toda tarea auxiliar, no incluyendo, por supuesto, la implementación de los algoritmos en c) y d)
- Al implementar los algoritmos, recomendamos utilizar la biblioteca NumPy para trabajar con operaciones vectoriales y mejorar su rendimiento computacional.
- No es necesario, aunque tampoco prohibido, realizar selección de atributos

### Entregables

- Informe con las pruebas realizadas y los resultados obtenidos.
- El informe a entregar debe ser un Jupyter Notebook.
- Código escrito para resolver el problema.

### Fecha límite de entrega

Lunes 11 de mayo (inclusive)

---

<sup>1</sup> <https://archive.ics.uci.edu/ml/datasets/Adult>