

Doble Grado en Ingeniería Informática y Administración  
de Empresas

2023-2024

*Aprendizaje Automático*

# Práctica 1: APRENDIZAJE SUPERVISADO

---

Cristina Ouro Alvite, [100454392@alumnos.uc3m.es](mailto:100454392@alumnos.uc3m.es)

Marina Gómez Lema, [100454255@alumnos.uc3m.es](mailto:100454255@alumnos.uc3m.es)

14/04/2024

# Introducción

En el siguiente documento se expone el uso realizado de la herramienta ChatGPT durante el desarrollo de la práctica I de la asignatura de Aprendizaje Automático.

La herramienta se ha usado, principalmente, para funciones de interpretación de errores, clarificación de conceptos teóricos, aclaraciones procedimentales sobre los pasos a dar o pequeñas propuestas de código para acciones específicas. En este caso, se han realizado prompts tanto en inglés como en castellano. Por ello, este informe presenta ejemplos en ambos idiomas.

## Usos de la herramienta CHATGPT

### A. Aclaraciones conceptuales

Se ha empleado ChatGPT para proporcionar ejemplos, explicaciones y aclaraciones de conceptos teóricos necesarios para desarrollar la práctica. Esto ha permitido adquirir una mayor comprensión de los conceptos y facilitado su manejo en el código implementado.

Ejemplo: se pide aclaración sobre el concepto de *index* y el uso de la función *set\_index*.

Q: what is an index in pandas DataFrame index? What does the set index method does?

A: In a pandas DataFrame, an index is a way to uniquely identify and label each row. It provides a way to access and manipulate the data in the DataFrame using these labels. The index can be thought of as the "row names" or "row labels" of the DataFrame. The set\_index method in pandas is used to set one or more columns as the index of the DataFrame. When you call `df.set_index(column_name)`, it returns a new DataFrame with the specified column (or columns) set as the index. The original DataFrame remains unchanged unless you use the `inplace=True` parameter, which modifies the DataFrame in place.

### B. Aclaraciones procedimentales

En otras ocasiones, se le pregunta a ChatGPT por formas de plantear las soluciones, obteniendo pasos a seguir que pudieran guiar el diseño de alguna sección específica de la práctica.

Ejemplo:

Q: Quiero seleccionar el mejor método (usando la evaluación inner), evaluarlo, construir modelo final y hacer predicciones. Cómo hago esto en python si tengo cuatro modelos (uno de KNN, otro de Regression Tree, otro de Linear Regression y otro de SVM) óptimos escogidos mediante gridsearch de cada uno tras haber hecho HPO.

A: Para seleccionar el mejor método utilizando la evaluación interna (inner evaluation), evaluarlo, construir el modelo final y hacer predicciones para la competición, puedes seguir estos pasos en Python:

1. Seleccionar el mejor método: Utiliza la evaluación interna (por ejemplo, mediante validación cruzada) para comparar el rendimiento de los cuatro modelos óptimos (KNN, árbol de regresión, regresión lineal y SVM) en un conjunto de datos de validación.

2. Evaluar el mejor método: Una vez que hayas seleccionado el mejor método, evalúalo en un conjunto de datos de prueba separado (si es posible) o mediante validación cruzada en el conjunto de datos completo.
3. Construir el modelo final: Entrena el modelo final utilizando todos los datos de entrenamiento disponibles.
4. Hacer predicciones para la competición: Utiliza el modelo final entrenado para hacer predicciones en los datos de la competición.

### C. Aclaraciones de código

En ciertas ocasiones se ha complementado la consulta de documentación oficial con prompts que proporcionaban aproximaciones más directas al uso de librerías, métodos o acciones específicas para el manejo de los objetos. Esto, principalmente, se ha empleado para comprender mejor el funcionamiento de la librería pandas y cómo operar con los DataSets para lograr efectos concretos.

Ejemplo:



You

Cómo itero sobre los valores de una columna en un dataset de andas



ChatGPT

Puedes iterar sobre los valores de una columna en un DataFrame de Pandas utilizando un bucle `for` o métodos específicos de Pandas como `iterrows()` o `itertuples()`. Aquí te muestro diferentes enfoques:

#### 1. Usando un bucle for:

```
python Copy code  
  
import pandas as pd  
  
# Suponiendo que 'df' es tu DataFrame y 'column_name' es el nombre de la columna  
for value in df['column_name']:  
    print(value)
```