



MÁSTER EN INTELIGENCIA ARTIFICIAL

PRÁCTICA OBLIGATORIA CLASE 3

Prof. Alba M^a Vicente
alba.vfernandez@asociacionaepe.es

1. Objetivo

Desarrollo de un *Exploratory Data Analysis* (EDA) y preprocesamiento de datos previo a la generación de un modelo de Machine Learning.

1.1. Descargar datos de [Heart Attack Analysis & Prediction Dataset | Kaggle](#) y cargarlos en una estructura de datos apropiada, haciendo uso de la librería *pandas*.

1.2. Efectuar un EDA y documentarlo de manera detallada, indicando las conclusiones que se pueden extraer. Mostrar:

- Tamaño del dataset (nº de filas y de columnas, tamaño en memoria).
- Valores únicos de cada variable (si aplica).
- Estadísticos básicos de cada variable.
- Porcentaje de valores faltantes de cada variable.
- Box plot de las variables numéricas continuas. ¿Hay outliers?
- Matriz de correlación de las variables numéricas continuas.

1.3. Preprocesar los datos como sea necesario, en función de las conclusiones extraídas a partir del EDA. Para ello, considerar cuestiones como las siguientes:

- ¿Conviene escalar los datos?
- ¿Hay que codificar ciertas variables de alguna manera en concreto?
- ¿Se podrían reducir las dimensiones del dataset?

2. Formato

Agregar las respuestas de esta práctica al cuaderno Jupyter de la práctica 2, modificando el nombre del cuaderno de la siguiente manera: “**Apellidos, Nombre - Máster IA - Prácticas 2,3**”.

3. Plazo de entrega

2 semanas.

4. Recursos

- Diapositivas *Módulo I - Fundamentos de Inteligencia Artificial y Machine Learning*.
- Notebook [Heart Attack - EDA + Prediction \(90% accuracy\) | Kaggle](https://www.kaggle.com/namanmanchanda/heart-attack-eda-prediction-90-accuracy) (enlace completo:
<https://www.kaggle.com/namanmanchanda/heart-attack-eda-prediction-90-accuracy>)