

UNIDAD TEMÁTICA 2: Procesamiento previo de los datos

PRACTICOS DOMICILIARIOS INDIVIDUALES #4

Ejercicio 1 – Análisis del dataset “Titanic” con python

En este ejercicio se desea construir un modelo de predicción para el dataset “Titanic”, utilizando Python y sklearn.

Utilizar el blog publicado en Kaggle para realizarlo, paso por paso:
<https://www.kaggle.com/samsonqian/titanic-beginner-s-guide-with-sklearn>

Documenta todos los pasos, análisis estadísticos y modelos, e integra todo en un apartado específico en tu portafolios – incluyendo todo el código desarrollado y ejecutando correctamente (será evaluado!!!)

PASOS OBLIGATORIOS EN ESTA UNIDAD:

1. Importar librerías y paquetes
2. Cargar el dataset y mostrarlo
3. Gestionar los valores faltantes
4. Graficar los datos
5. Feature Engineering

PASOS OPCIONALES EN ESTA UNIDAD (SERÁN OBLIGATORIOS EN UNIDADES SIGUIENTES):

1. Modelado y Predicción con sklearn
2. Evaluación del Rendimiento del Modelo

Ejercicio 2

Tomando como base el blog de la Universidad de Stanford “A Titanic Probability”, disponible en <http://web.stanford.edu/class/archive/cs/cs109/cs109.1166/problem12.html>

Y a partir del dataset “Titanic” que hemos estado utilizando, desarrolla un programa en Python que responda las siguientes consultas:

1. Calcular la probabilidad condicional de que una persona sobreviva, dados su sexo y clase de pasajero , aplicarlo a las siguientes preguntas concretas:
 1. $P(S = \text{true} \mid G = \text{female}, C = 1)$
 2. $P(S = \text{true} \mid G = \text{female}, C = 2)$
 3. $P(S = \text{true} \mid G = \text{female}, C = 3)$
 4. $P(S = \text{true} \mid G = \text{male}, C = 1)$
 5. $P(S = \text{true} \mid G = \text{male}, C = 2)$
 6. $P(S = \text{true} \mid G = \text{male}, C = 3)$
2. ¿Cuál es la probabilidad de que un niño que tenga 10 o menos años de edad y esté en 3era clase sobreviva?

$$P(S = \text{true} \mid A \leq 10, C = 3)$$