

CONSTRUYENDO UN ÁRBOL DE CLASIFICACIÓN EN PYTHON

ABORDANDO UN PROBLEMA DE CLASIFICACIÓN

EN ESTA CLASE VAMOS A VER

- Exploración de un conjunto de datos en Python
- Preprocesado de datos para construcción de un modelo
- Construcción y validación de un modelo en Python
- Representación gráfica de un árbol de clasificación

TITANIC, UN PROBLEMA CON DATOS REALES

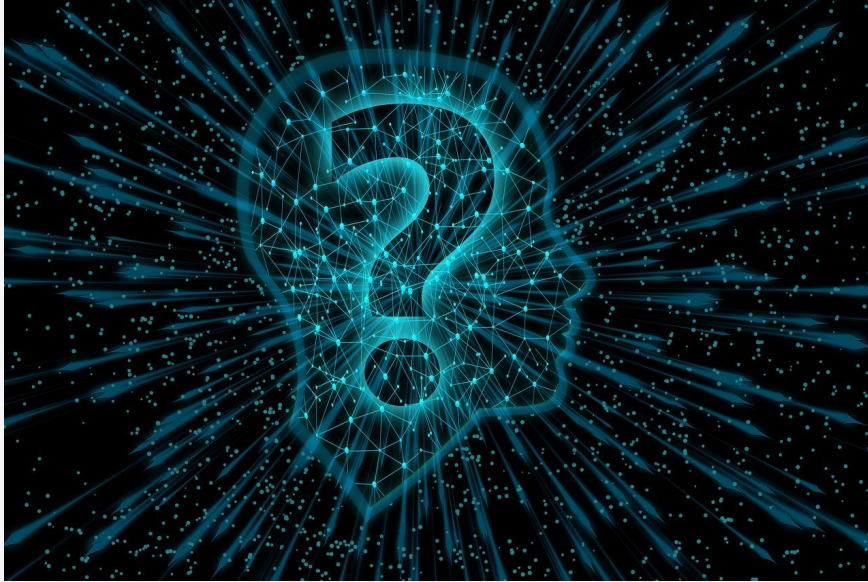
- En esta clase vamos a trabajar con un conjunto de datos reales sobre el naufragio del Titanic
- El conjunto de datos almacena información real de 861 pasajeros del Titanic
- Entre esta información conoceremos la clase en la que viajaban, su género, si viajaban con familia o solos...
- Para cada pasajero tendremos un indicador que señala si sobrevivió o no a la tragedia



TITANIC, UN PROBLEMA DE CLASIFICACIÓN

- El objetivo de nuestro modelo será predecir si una persona sobrevivirá o no al hundimiento
- Aunque evidentemente el problema no tiene sentido “a toro pasado” resulta interesante conocer los factores que más influencia tenían en la probabilidad de supervivencia de los pasajeros
- Nuestra variable objetivo solo toma dos valores: sobrevive o no, por lo que nos enfrentamos a un **problema de clasificación binaria**

¡PASEMOS A LA PRÁCTICA!



EN ESTA CLASE HEMOS VISTO

- Cómo cargar y explorar un conjunto de datos vía Python
- Cómo estudiar y visualizar las distribuciones de datos
- Cómo preprocesar un dataset haciendo cosas cómo generar “*variables dummy*” o imputar valores en los nulos
- Cómo construir y entrenar un modelo en unas pocas líneas de código
- Cómo evaluar con distintas métricas la calidad del modelo construido
- Cómo visualizar un árbol de decisión