¿Cuál es la diferencia entre aprendizaje supervisado y no supervisado?

El aprendizaje supervisado consta de una matriz de caracteristicas y un vector objetivo, esto a grandes rasgos busca entrenar un modelo con sus respectivos resultados del vector objetivo para que asi pueda predecir valores nuevos con data nueva. El aprendizaje no supervisado solo consta de una matriz de caracteristicas y se utiliza comunmente para descubrir agrupaciones en los conjuntos de datos.

¿Cuál es la diferencia entre regresión y clasificación?

La regresion trata de un valor continuo, como un precio mientras que la clasificacion trata de un valor categorico.

¿Qué es un objetivo en aprendizaje automático supervisado?

Es la columna del conjunto de datos donde se quiere hacer las predicciones.

Explica el propósito de un train-test split.

El proposito es preparar un conjunto de datos para la correcta prediccion mediante una simulacion en la prediccion.

¿Qué es el aprendizaje automático?

Es un conjunto de reglas autoamtizadas y procesadas por la maquina para identificar patrones en un conjunto de datos.

¿Cuáles son algunas aplicaciones comunes del aprendizaje automático?

Predicciones de valores categoricos, agrupacion dentro de un conjunto de datos o la reduccion de la dimensionalidad de un conjunto de datos.

¿Cuáles son los beneficios de usar un pipeline?

La reduccion del codigo y la visibilidad o entendimiento de este.

¿Qué es la fuga de datos y por qué es un problema?

La fuga de datos son pistas o restos de data que quedan en el camino durante el proceso de entrenamiento los cuales se pueden mezclar con el conjunto de prueba y ocasionar una prediccion poco certera.

¿Qué se entiende por "ajustar" un modelo?

Se refiere a encontrar los mejores parametros o configuraciones para un algoritmo de aprendizaje automatico.

Describe los pasos principales del preprocesamiento de datos para el aprendizaje automático.

La recopilacion de datos, Limpieza, seleccion de caracteristicas, transformacion de datos, la division de datos y el balanceo de datos.