

EJERCICIO 3

Del dataset del punto 1 realice en PYTHON, cinco algoritmos de preprocesamiento; dos de ellos deben ser onehotencoder y otro escalado. Explique por cada uno la razón de su uso

Implementacion en github:

https://github.com/alvaro-tc/dat-245/tree/main/PrimerParcial/3_ejercicio/

One Hot Encoder Convierte datos nominales a binarios, es útil para otros algoritmos que estrictamente necesiten que todos sus datos sean numericos

Escalado Convierte los números a una escala entre 0 y 1, para evitar que una característica con valores grandes domine sobre otras con valores más pequeños. Esto es útil cuando las características tienen diferentes unidades o escalas, ya que ayuda a que todas las características tengan un impacto similar en el modelo.

Imputación de valores faltantes (Missing Value Imputation) Es el proceso de rellenar los valores faltantes en un conjunto de datos con valores estimados. Los valores faltantes pueden ocurrir por diversas razones, como errores en la recopilación de datos o pérdida de información. La imputación de valores faltantes es importante para garantizar que no se pierda información crucial y que los algoritmos de aprendizaje automático puedan trabajar con conjuntos de datos completos.

Normalización de características (Feature Normalization) Es el proceso de ajustar las características para que tengan una escala común. Esto se logra restando la media de la característica y dividiendo por la desviación estándar. La normalización de características es útil cuando las características tienen diferentes unidades o escalas, para que sea mas fácil emplear los algoritmos de aprendizaje. Al normalizar las características, se asegura que todas tengan una influencia similar en el modelo.

Selección de características (Feature Selection) Es el proceso de elegir un subconjunto relevante de características para utilizar en un modelo. Algunas características pueden no ser informativas o pueden estar altamente correlacionadas con otras, lo que puede afectar negativamente la precisión del modelo o aumentar el tiempo de entrenamiento. La selección de características ayuda a eliminar características redundantes o irrelevantes.