Pablo Arranz ropero y Juan Alberto camino Sáez



Si me lo como, ¿me muero?

El propósito de este proyecto es aplicar diferentes algoritmos de aprendizaje automático a un conjunto de datos de setas encontrado en la siguiente URL:

<https://www.kaggle.com/uciml/mushroom-classification>

El conjunto de datos presenta un total de 8124 instancias con 23 atributos cada una.

Cada uno de los datos representa lo explicado en el Excel adjunto.

sed -i '/^.\*?.\*$/d' mushrooms.csv

sed -i 's/^e/0/g' mushrooms.csv

sed -i 's/^p/1/g' mushrooms.csv

sed -i -E 's/^(.{1}\,)b/\11/g' mushrooms.csv 🡪 para sustituir los cap-shape

sed -i -E 's/^(.{1}\,.{1}\,)f/\11/g' mushrooms.csv 🡪 para sustituir los cap-surface

sed -i -E 's/^([0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,)t/\11/g' mushrooms.csv 🡪 cambiada la regex para que coja numeros de varios digitos

sed -i -E 's/^([0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,)s/\14/g' mushrooms.csv 🡪 regex para stalk-surface-below-ring

sed -i -E 's/^([0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,[0-9]+\,)g/\11/g' mushrooms.csv 🡪 ultima regex usada