

ESERCITAZIONE REGRESSIONE LINEARE

Corso di Data Analytics

20 Ottobre 2021

1. Caricare in un DataFrame Pandas il dataset contenuto del file "quality.csv". Stampare a video il numero di righe e di colonne.
2. Assicurarsi che il dataset NON contenga valori mancanti (NaN). Nel caso, rimuovere le righe associate ai valori mancanti. Stampare il numero di righe dopo l'operazione.
3. Applicare un modello di regressione lineare semplice tra le variabili "TEMP" (input) e "PM2.5" (output). Stampare il valore del MSE e dell' R^2 .
4. Visualizzare il dataset mediante scatterplot e la retta di regressione calcolata.
5. Ripetere l'esercizio considerando come variabili di input tutte quelle presenti nel dataset (tranne "No") e come variabile di output "PM2.5". Occorre effettuare l'encoding delle variabili categoriche presenti del dataset.
6. Applicare un modello di regressione logistica per riconoscere il nome della stazione meteo a partire dalle altre variabili presenti nel dataset (tranne "No").