DIVIDERE IL DATASET

100% DEI DATI, PRENDO IL TRAINING SET (TRA IL 75% E IL 100% DEI DATI), ALLENO SU QUESTI DATI IL MIO ALGORITMO.

SUL RESTANTE 25% PROVO L’ALGORITMO

ENCODING

SCALAR: scalare i dati per la regressione logistica

Una volta trovato il modello che performa posso fare quella che viene definita:

FEATURE IMPORTANCE