



Big Data Real-Time Analytics Com Python e Spark 3.0

Big Data Real-Time Analytics Com Python e Spark Versão 3.0

O Que é Label Encoding?

Label Encoding (codificação de rótulos) é uma técnica de codificação para lidar com variáveis categóricas. Nesta técnica, a cada rótulo é atribuído um número inteiro exclusivo com base na ordem alfabética.

Na codificação de rótulos em Python, substituímos o valor categórico por um valor numérico entre 0 e o número de classes menos 1. Se o valor da variável categórica contiver 5 classes distintas usamos (0, 1, 2, 3 e 4). Observe que os dados são modificados, mas sem perder a informação que eles representam.

Isso pode ser feito de maneira manual (usando dicionários em Python) ou com o algoritmo LabelEncoder do pacote Scikit-Learn. Veremos os dois métodos na sequência.