

Ajuste de unidades de medida

Durante a fase de **Preparação do Pipeline do Big Data**, procuramos investigar correlações entre variáveis, ou seja, tentar descobrir se ao mudar o conteúdo de uma variável, o conteúdo de outra variável apresenta um comportamento correlacionado. Por exemplo, imagine que estamos tentando avaliar automóveis. Se a variável "quilômetros rodados" for muito alta, provavelmente a variável "preço do veículo" tenderá a cair. Neste caso, dizemos que as variáveis estão correlacionadas de forma negativa (uma cresce, a outra diminui). Uma condição importante para a descoberta de correlações entre variáveis consiste em padronizar seus valores para que não existam discrepâncias muito acentuadas.

Pergunta-se: como se chama a operação que padroniza valores entre variáveis de um mesmo *dataset*?

Selecione uma alternativa

A Nulificação

B Fatoração

C Normalização

D Transformação