



Data Science Academy

www.datascienceacademy.com.br

Big Data Analytics com R e Microsoft Azure Machine Learning

Mini-Projeto 3 Prevendo Despesas Hospitalares

O objetivo deste mini-projeto é praticar o conhecimento adquirido ao longo dos capítulos anteriores.

Neste mini-projeto, seu trabalho é prever as despesas hospitalares a partir de vários indicadores de diferentes pacientes. Para esta análise, vamos usar um conjunto de dados simulando despesas médicas hipotéticas para um conjunto de pacientes espalhados por 4 regiões do Brasil. Esse dataset possui 1.338 observações e 7 variáveis.

Uma das suas principais atividades antes de construir o modelo é analisar a correlação entre as variáveis. De acordo com a nossa análise inicial do dataset, nenhuma das correlações na matriz são consideradas fortes, mas existem algumas associações interessantes. Por exemplo, a idade e o bmi (IMC) parecem ter uma correlação positiva fraca, o que significa que com o aumento da idade, a massa corporal tende a aumentar. Há também uma correlação positiva moderada entre a idade e os gastos, além do número de filhos e os gastos. Estas associações implicam que, à medida que idade, massa corporal e número de filhos aumenta, o custo esperado do seguro saúde sobe. Seu trabalho é empregar técnicas de análise e confirmar ou negar essas afirmações. Na sequência, crie um modelo que faça as previsões quando receber novos conjuntos de dados.

Uma proposta de solução será apresentada com todos os arquivos do projeto.

Equipe Data Science Academy