



**Data Science
Academy**

www.datascienceacademy.com.br

Big Data Real-Time Analytics com Python e Spark

Desafio
Pipeline com PCA e Regressão Logística



Você gosta de desafio?

Usando o que aprendeu neste capítulo, junto com um trabalho de pesquisa na documentação do Scikit-Learn, você deve criar um Pipeline, ou seja, um conjunto de modelos concatenados, especificamente o PCA (para redução de dimensionalidade) com Regressão Logística (para classificação). Com um Pipeline, realizamos duas operações em conjunto!

O script com os dados foi fornecido para você junto com os demais arquivos do capítulo. A solução será fornecida no próximo capítulo!

Bom trabalho!