



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

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

## **Estudo de Caso 5 Pré-Processamento de Dados Para E- Commerce Analytics**



## Formação Cientista de Dados 3.0

### Big Data Real-Time Analytics com Python e Spark

#### Estudo de Caso 5

#### Pré-Processamento de Dados Para E-Commerce Analytics

Este Estudo de Caso é uma continuação do trabalho feito nos capítulos anteriores. Agora aplicaremos técnicas de pré-processamento de variáveis categóricas (Label Encoding e One-Hot Encoding) e Feature Scaling de variáveis numéricas (Normalização e Padronização).

O objetivo é pré-processar os dados para uma etapa de modelagem preditiva.

Todo o Estudo de Caso é no contexto de um problema de negócio em E-Commerce Analytics e usaremos como dataset o arquivo gerado ao final do Estudo de Caso 4.