



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 Definição do Problema e Fonte de Dados



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

Uma empresa internacional de comércio eletrônico (E-commerce) que vende produtos eletrônicos deseja descobrir informações importantes de seu banco de dados de clientes.

Os produtos ficam armazenados em um armazém na sede da empresa. Após concluir a compra no web site da empresa, o cliente recebe o produto em casa, em qualquer parte do mundo. Os produtos são enviados de Navio, Avião ou Caminhão, dependendo da região de entrega.

Em cada compra o cliente pode receber um desconto dependendo do peso do produto comprado. Cada cliente pode fazer chamadas ao suporte da empresa no caso de dúvidas ou problemas e após receber o produto o cliente pode deixar uma avaliação sobre a experiência de compra. O único dado pessoal sobre o cliente que está disponível é o gênero.

O trabalho de Engenharia de Atributos foi realizado e novas variáveis estão disponíveis no conjunto de dados.

Nosso trabalho neste Estudo de Caso é pré-processar os dados e deixá-los prontos para a etapa de modelagem preditiva (que não será abordada). Nosso foco será nas técnicas de pré-processamento. Usaremos o dataset gerado no final do capítulo anterior.

Leia os manuais em pdf com diversas definições ao longo do capítulo.