



**Data Science
Academy**

www.datascienceacademy.com.br

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

Mini-Projeto 2

Análise de Dados do Twitter em Tempo Real



O Twitter é uma das redes sociais mais dinâmicas disponíveis atualmente. É possível coletar em tempo real informações preciosas e o sentimento das pessoas sobre os mais variados temas. São milhões de tweets por minuto em todo mundo e muita informação preciosa está escondida em cada tweet.

O Streaming de dados gerado pelo Twitter pode alimentar aplicações analíticas, permitindo que empresas compreendam, em tempo real, o que clientes, parceiros e fornecedores estão pensando (e escrevendo) sobre você, sua marca, produto ou serviço.

Neste projeto, coletamos dados de Streaming do Twitter, aplicamos técnicas de análise em tempo real (à medida que os dados são gerados) e obtemos insights sobre um determinado assunto. O projeto pode ser facilmente reproduzido localmente em seu computador ou em um ambiente de cluster na nuvem. Tecnologias utilizadas:

Python/NLTK – construção do modelo de análise de sentimento

Spark Streaming – coleta do streaming de dados

API do Twitter – conexão a partir da nossa aplicação

Todo o conteúdo do Mini-Projeto você encontra no arquivo zip ao final deste capítulo.