



Data Science Academy

www.datascienceacademy.com.br

Big Data Analytics com R e Microsoft Azure Machine Learning

Projeto 1
Análise de Redes Sociais
Data Mining, Web Scraping e Análise de
Sentimentos

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

Vamos coletar dados do Twitter, a partir do R, usando técnicas de web scraping. Então iremos limpar os dados, organizá-los, gerar visualizações e por fim realizar uma análise de sentimentos nos dados coletados, através de técnicas de processamento de linguagem natural e também usando um classificador Naive Bayes.

Mais importante que encontrar uma solução técnica (e existe mais de uma), procure praticar suas habilidades analíticas, identificando os dados que devem ser coletados, os pacotes R que devem ser usados, a forma como os dados devem ser remodelados e os possíveis gráficos gerados.

Para este projeto, você vai precisar de uma conta no Twitter e acesso como desenvolvedor. Após criar sua conta (você pode usar sua conta pessoal ou criar uma conta para o projeto), acesse o site: <https://apps.twitter.com> para habilitar seu acesso como desenvolvedor.

Aqui você encontra a documentação oficial do R sobre Text Mining:

<https://cran.r-project.org/web/views/NaturalLanguageProcessing.html>

Aqui você encontra a lista de stopwords em vários idiomas, incluindo português:

<http://www.ranks.nl/stopwords/brazilian>

Aqui um artigo explicando o que é Análise de Sentimento:

<http://www.cienciaedados.com/analise-de-sentimentos-e-machine-learning/>

Utilizaremos os pacotes twitter, plyr, stringr e tm, além dos pacotes sentiment e Rstem. Instale e estude o help sobre estes pacotes no RStudio. Compreenda como coletar os dados, realizar procedimentos de limpeza e organização. Monte um plano de ação sobre quais seriam as etapas necessárias para realizar este projeto.

A solução será apresentada com todos os arquivos do projeto.

Equipe Data Science Academy