

Data Science do ZERO

Capítulo 07

Mineração de Textos (NLP)

NLTK - Natural Language Toolkit



Porque usar NLTK?

- Toolkit para NLP
 - Processamento de textos.
 - Análise Exploratória de Textos.
 - Similaridade de palavras.
 - Análise de Palavras chave.
 - Suite de datasets de exemplos.
 - Integração com Recursos Léxicos.
 - Integração com algoritmos de Machine Learning (Sklearn)



NLTK - Processamento de textos

- Recursos para processamento de textos
 - Processamento de frequencia de termos
 - Similaridade de termos
 - Busca por conceitos semanticos
 - Integração como o Wordnet
 - Recursos para Tokenização



NLTK - Classificação de textos

- Classificação de textos
 - Algoritmos de Machine Learning:
 - Naive Bayes
 - Decision Tree
 - SVM
 - Logistic Regression
 - Métricas para avaliação de modelos



Hands on!