

Data Science do ZERO

Capítulo 07

Mineração de Textos (NLP)

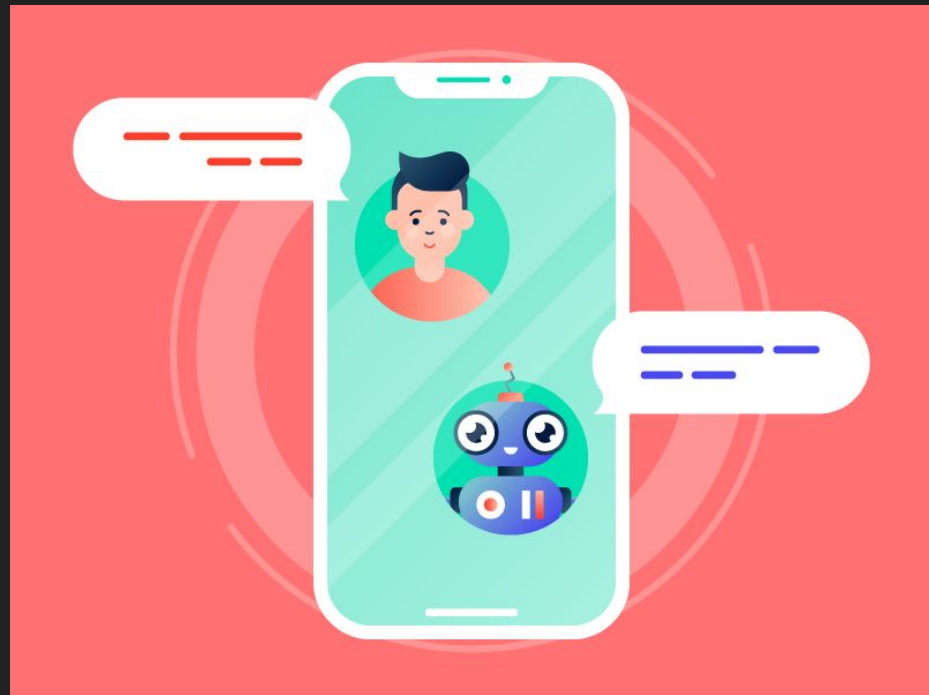
O que é Mineração de Textos?

- Mineração de Textos é uma técnica usada para tirar valiosas informações de bases de dados de textos.
- Também chamamos de **Natural Language Processing** que significa Processamento de linguagem Natural ou **NLP**.
- Essa área está em crescendo muito devido a quantidade de dados que temos hoje.
- Existem tarefas que podem resolver problemas importantes ou podem gerar bastante valor.



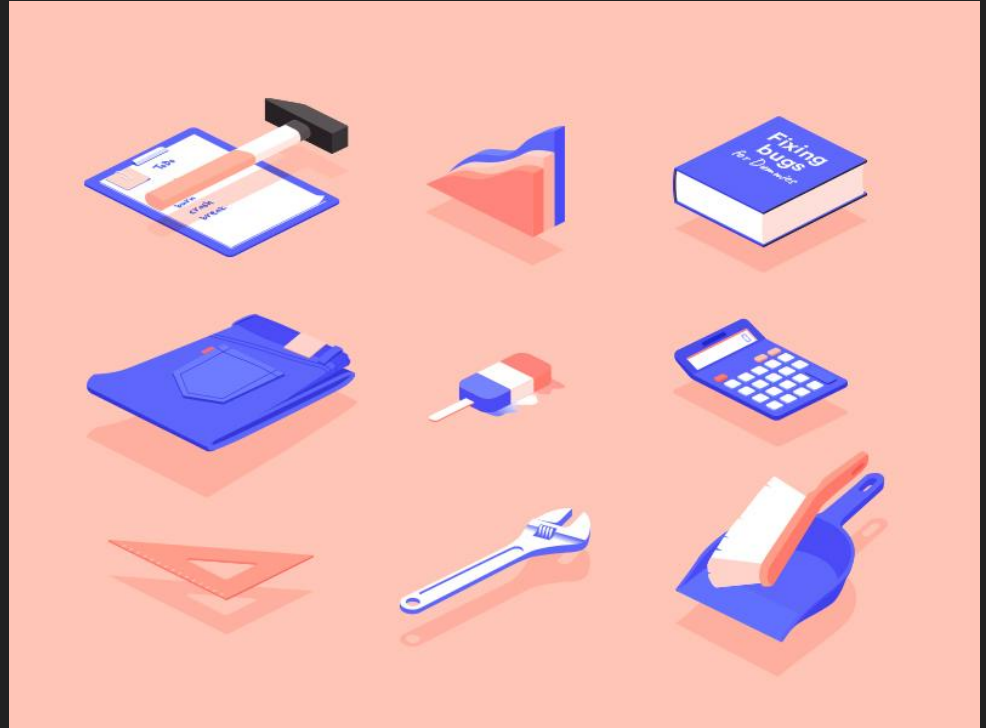
Mineração de Textos

- Por que Aprender Mineração de Textos?
 - Automatização da comunicação entre pessoas ou serviços.
 - Extração de conhecimento em bases de dados de texto.



Mineração de Textos

- Recursos Disponíveis:
 - Funções Nativas da Linguagem Python
 - Biblioteca NLTK.
 - Bibliotecas para NLP
 - Dicionários Léxicos
 - Machine Learning e Deep Learning



Wordnet

- Dicionário Léxico:
 - Grande dicionário léxico de substantivos, verbos, adjetivos, advérbios que são agrupados por **conceitos**.
 - Utilizado no mundo todo para pesquisas.
 - Essa ferramenta está disponível gratuitamente.



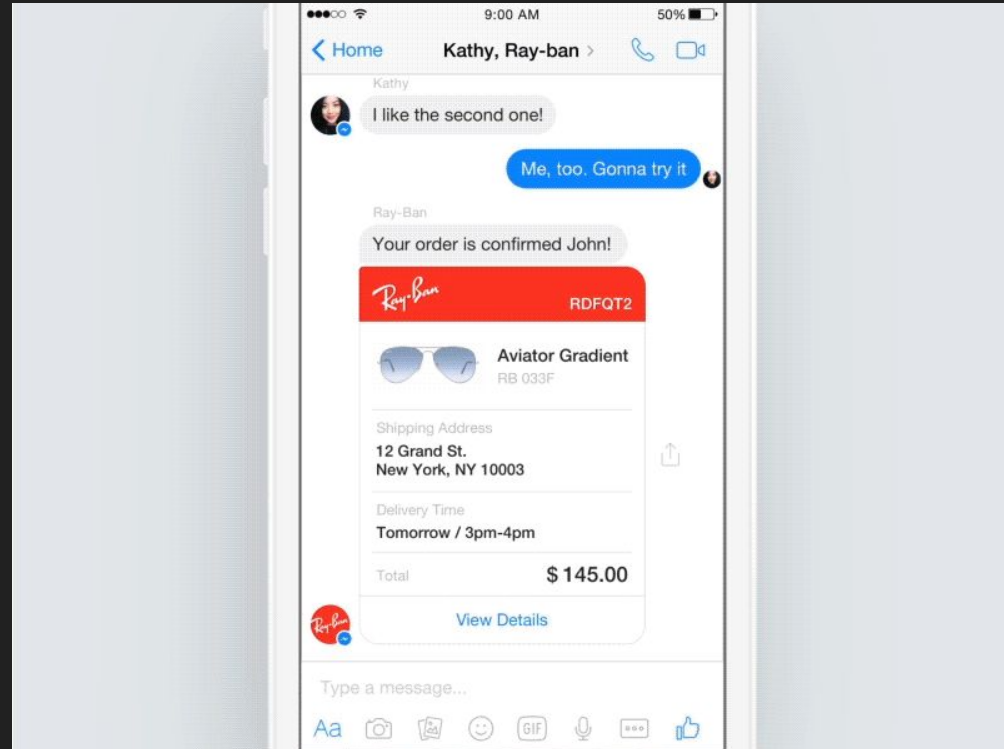
Part of Speech

- Partes da Fala
 - Consiste em classificar palavras pela função gramatical
 - Exemplo: “**Joãozinho** **comprou** **um doce**”
 - Nesse exemplo “**Joãozinho**” seria um substantivo, “**comprou**” seria um verbo e “**doce**” um substantivo.
 - Part of Speech é uma técnica difícil.



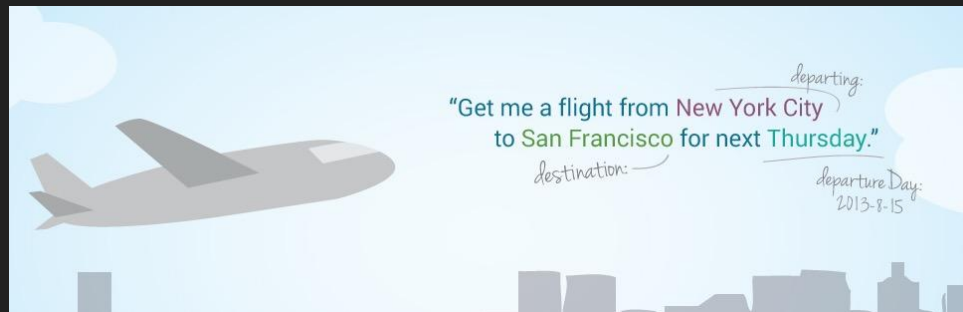
Part of Speech

- Onde utilizar PoS?
 - Existem inúmeras aplicações
 - Um exemplo seria um sistema de **chatbots**.
 - Identificar a função gramatical das palavras no texto é crucial nesse tipo de tarefa.
 - Essa tarefa é base para outras.



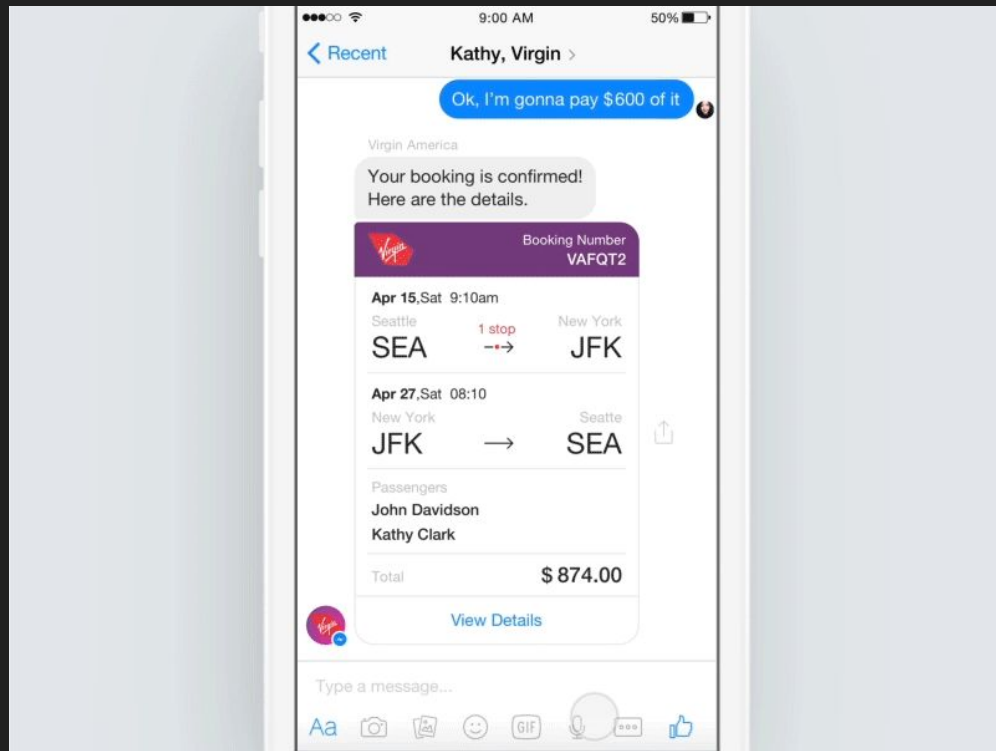
Reconhecimento de Entidades

- Reconhecimento de Entidades:
 - Reconhecimento de entidades no texto.
 - As entidades são padrões definidos pelo usuário.
 - O modelo aprende a identificar padrões.
 - Utiliza métodos de Machine Learning ou não.
 - Part of Speech pode ter uma função essencial.



Reconhecimento de Entidades

- Onde utilizar Reconhecimento de Entidades:
 - As entidades são padrões definidos pelo usuário.
 - O modelo aprende a identificar padrões.
 - Sistema pode tomar decisões conforme identifica as entidades.



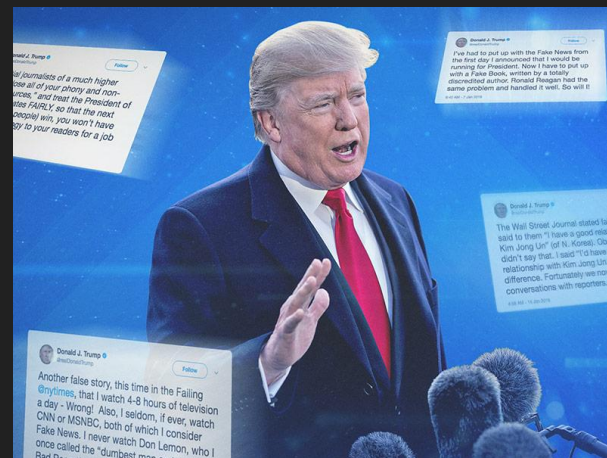
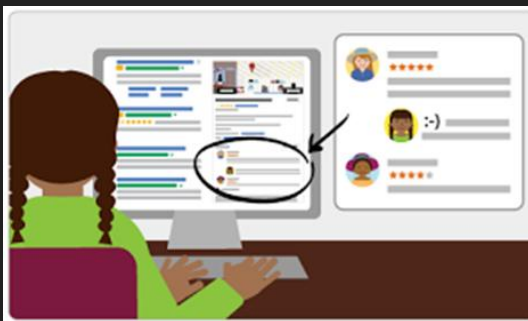
Análise de Sentimentos

- Análise de sentimentos é uma tarefa muito demandada quando se trabalha com textos.
- Pode ser utilizar diversos métodos como léxicos ou Machine Learning.
- Essa tarefa também pode ser base para outras aplicações.



Análise de Sentimentos

- Onde utilizar?
 - Análise de sentimentos de Marcas, produtos ou serviços.
 - Análise de sentimentos em posts de pessoas públicas.
 - Análise de emoções empregadas em textos ou áudio.



Hands on!