

# Data Science do ZERO

## Capítulo 07

**Mineração de Textos (NLP)**

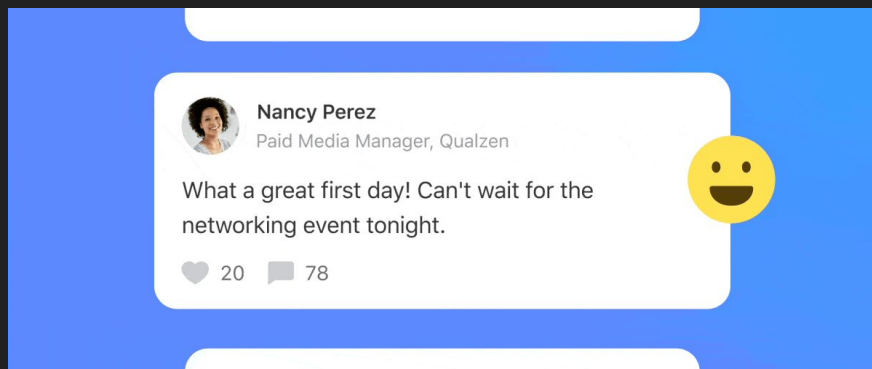
**Análise de Sentimentos**

# O que é Análise de Sentimentos?

*"Análise de Sentimentos ou Mineração de opiniões refere-se ao uso de NLP cujo objetivo é identificar, extrair e quantificar a polaridade expressada nos dados."*

# Análise de Sentimentos

- Análise de Sentimentos pode ser aplicada em dados de áudio ou ainda em dados de vídeo ou imagem.
- 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.

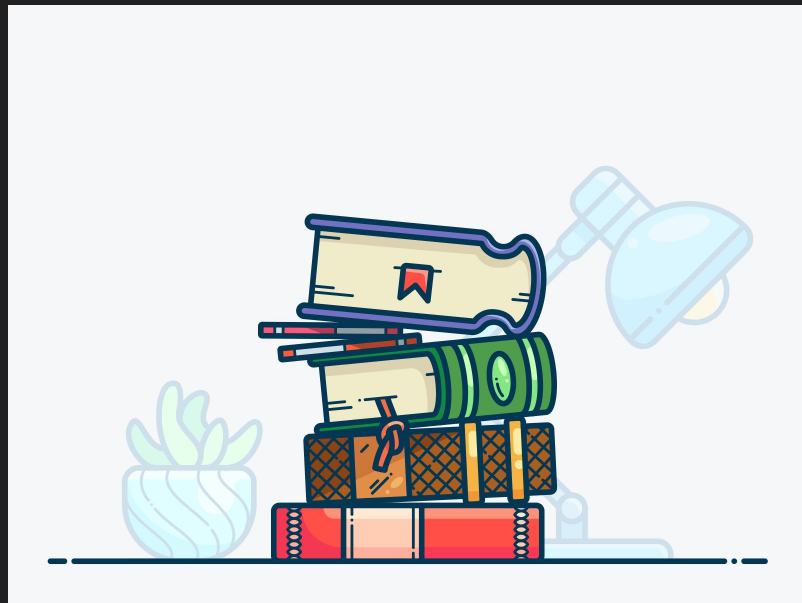


# Análise de Sentimentos

- Porque Análise de Sentimentos é difícil?
  - Negações: Uso de negações altera a polaridade de uma sentença ou uma palavra.
    - Exemplo: "I like this music" e "I don't like this music"
  - Metáforas, Ironias, Brincadeiras: As máquinas têm dificuldades de identificar esse tipo de colocação nos textos.
  - Muitos Sentimentos em uma sentença:
    - Exemplo: "*Esse smartphone é muito bonito, porém, a bateria deixa muito a desejar*".
  - Nesse exemplo temos dois sentimentos expressados, como agregar isso?

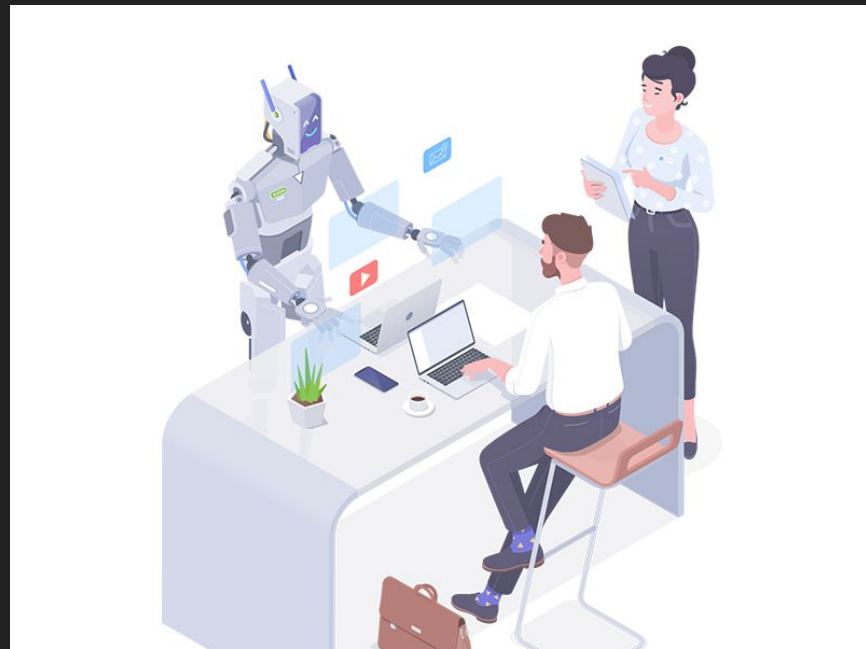
# Análise de Sentimentos

- Tipos de Abordagens:
  - **Dicionários Léxicos:**
    - Dicionários léxicos contém informação de sentimentos. Com isso o sentimento das palavras podem ser combinados para gerar uma classificação de uma sentença.
    - Alguns léxicos são criados levando em consideração uma área de conhecimento como: Psicologia, Direito etc.
    - A principal vantagem é a não necessidade de ter dados de treino.
    - A principal desvantagem é a dependência do idioma.



# Análise de Sentimentos

- Tipos de Abordagens:
  - **Machine Learning:**
    - Uma forma muito comum de fazer análise de sentimentos é utilizar machine learning para classificação de sentimentos.
    - Essa abordagem normalmente consiste em modelar os dados de texto no formato **Bag of Words**.
    - A principal vantagem é a não dependência de idiomas
    - A principal desvantagem é a necessidade de ter dados de treino.

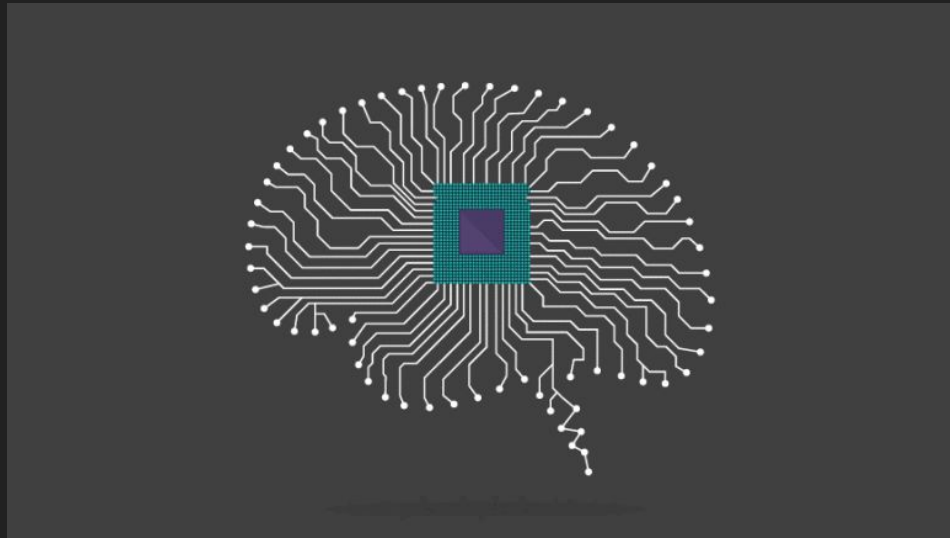


# Análise de Sentimentos

- Tipos de Abordagens:

- Deep Learning:

- Algoritmos de deep learning estão sendo muito utilizados para análise de sentimentos
    - Alguns algoritmos conseguem aprender padrões de escrita ou contexto na qual o texto inserido.
    - A principal vantagem é que muitos algoritmos desse tipo exigem menos trabalho de feature engineer
    - A principal desvantagem é a necessidade de muitos dados para treino.



Hands on!