

Introdução à Análise de sentimento

Welton Antonio Gomes



Quem sou eu?



- ▶ Física
- ▶ Filosofia
- ▶ Cinema (I&S)
- ▶ Análise de Sistemas

- ▶ Data Science
- ▶ PLN
- ▶ Sistemas de diálogos
(aplicados a jogos)

Outline

- ▶ O que é análise de sentimento?
- ▶ Pra que serve?
- ▶ Por que é difícil?
- ▶ Como é feito?
- ▶ *Hands-on* (se houver tempo...)

O que é Análise de Sentimento?

“Análise de Sentimento (*Sentiment Analysis*), também chamada de Mineração de Opinião (*Opinion Mining*) é o campo de estudo que analisa a opinião, avaliações, atitudes e emoções de pessoas em relação a entidades tais como produtos, serviços e organizações” (LIU, 2012)

O que é Análise de Sentimento?

- ▶ Análise de Sentimento
 - ▶ *Texto subjetivo direcionado a uma entidade*
 - ▶ Texto
 - ▶ Subjetivo
 - ▶ Entidade

O que é Análise de Sentimento?

► Fato vs. Opinião

“Inteligência artificial derrota melhores jogadores profissionais de pôquer do mundo” (Gizmodo Brasil, 12/07/2019)

(trecho da matéria) *Cientistas da computação desenvolveram um robô que joga cartas, chamado Pluribus.*

“Claro encerra marca NET e incorpora TV por assinatura e internet fixa” (Gizmodo Brasil, 11/07/2019)

(comentário) *O atendimento telefônico ficou muito ruim!.*

O que é Análise de Sentimento?

► Fato vs. Opinião

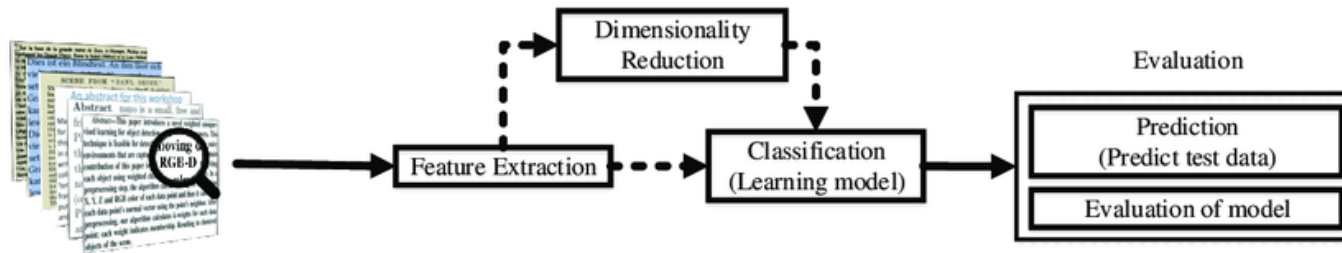
“Steve Wozniak, cofundador da Apple, acha que todo mundo deveria sair do Facebook” (Gizmodo Brasil, 08/07/2019)

(comentário) *Facebook é coisa de gente carente.*

O que é Análise de Sentimento? (Na prática...)

- ▶ Processamento de Linguagem Natural
- ▶ Aprendizado de Máquina
- ▶ Processamento de textos **opinativos** usando técnicas específicas:
 - ▶ (pré)Processamento de textos
 - ▶ Algoritmos de Classificação
 - ▶ Análise (e uso) dos resultados

O que é Análise de Sentimento? (Pipeline...)



- ▶ Pré-processamento
- ▶ Extração de Características
(Redução de Dimensionalidade)
- ▶ Classificação
- ▶ Predição/Avaliação
- ▶ [Modelo em Produção]

Aplicações da Análise de Sentimento

Pra quem interessa a opinião das pessoas?

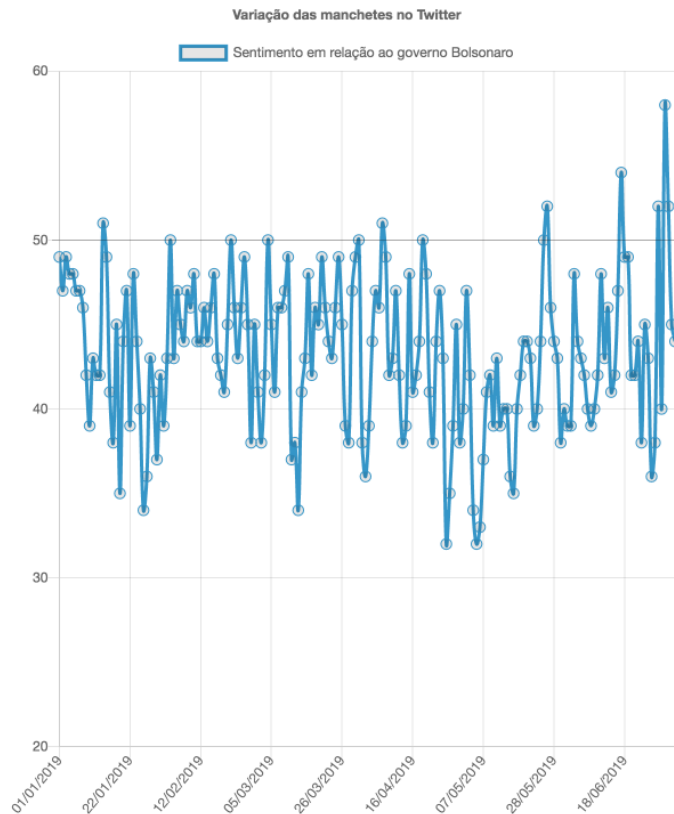
- ▶ Monitoramento de produtos/pessoas/marcas/etc;
- ▶ Pesquisas opinião;



Aplicações da Análise de Sentimento

Pra quem interessa a opinião das pessoas?

► “Termômetros”;



Aplicações da Análise de Sentimento

Pra quem interessa a opinião das pessoas?

- Análise de Sentimento “para o mal”



29/06/2014 13h45 - Atualizado em 29/06/2014 13h45

AFP

Em experimento secreto, Facebook manipula emoções de usuários

Algoritmo de 700 mil usuários foi manipulado por uma semana em 2012. Resultado mostrou que humor de usuários varia de acordo com conteúdo.

Da France Presse



Um estudo detalhando como o Facebook manipulou secretamente o feed de notícias de aproximadamente 700 mil usuários com o objetivo de avaliar o "contágio emocional" desencadeou revolta na rede social.

O que é Análise de Sentimento? (Porque é difícil?)

- ▶ Linguagem Natural
 - ▶ Dados não estruturados
 - ▶ Construções reais podem ser complexas
 - ▶ Sarcasmo, Ironia, etc.
- ▶ Sentimentos (opiniões) são complexos
- ▶ Identificar a Entidade pode não ser tão simples
- ▶ Acurácia de um ser humano?

O que é Análise de Sentimento? (Porque é difícil?)

“Claro encerra marca NET e incorpora TV por assinatura e internet fixa” (Gizmodo Brasil, 11/07/2019)

(comentários)

- > *Twitter caiu e o presidente da república ficou sem ter o que fazer por uns 30 minutos*
- > *tem vaga?*
- > *Ok mas eu vou poder continuar com o adesivo da Net que eu coloquei no meu Uno pra ele andar mais rápido?*

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

Hands On(?)

Toy-problem 1: Classificação de sentimento utilizando o DataSet de comentários do IMDb

- ▶ Dataset já estruturado (pouca necessidade de pré-processamento)
- ▶ Inglês
- ▶ *CountVectorizer (One-Hot Encoding)*
- ▶ Sentimentos binários (positivo e negativo)
- ▶ Foco nos algoritmos mais clássicos de Classificação
 - ▶ LR, KNN, SVM, *Random Forest*;
- ▶ Link:
<https://colab.research.google.com/drive/1Jd7ywToSPb8b2gVtBFk03HkZKSanH9Cu>

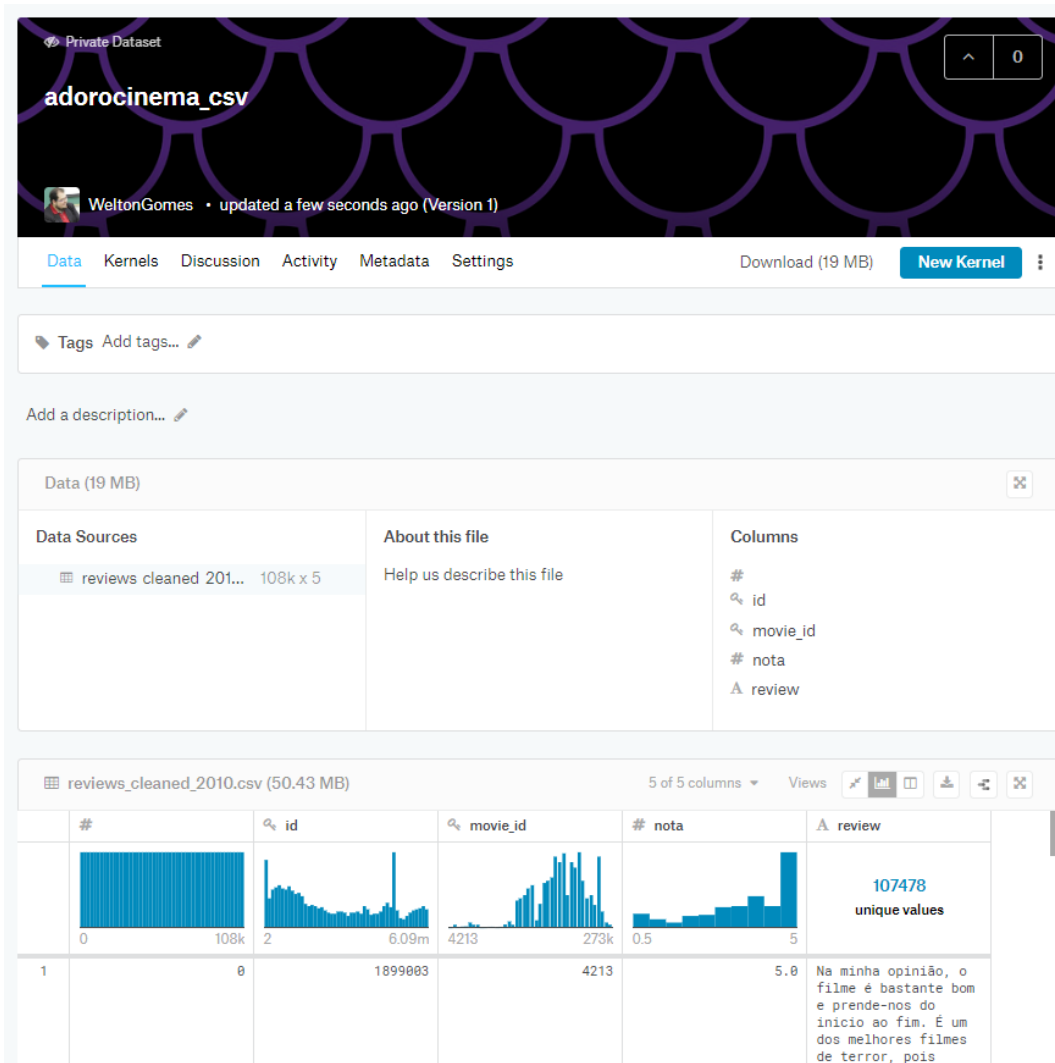
Toy-problem 1b: Classificação de sentimento utilizando o DataSet de comentários do IMDb (dataset traduzido)

- ▶ Dataset já estruturado (pouca necessidade de pré-processamento)
- ▶ Inglês / Português
- ▶ *CountVectorizer (One-Hot Encoding)*
- ▶ Sentimentos binários (positivo e negativo)
- ▶ Foco nos algoritmos mais clássicos de Classificação
 - ▶ LR, KNN, SVM, *Random Forest*;
- ▶ Link:
<https://colab.research.google.com/drive/1T3Z39rN8fBNsocfs6t8oqXMLDV5c1rFS>

Toy-problem 2: Captura de um dataset com comentários de filmes em PT-BR e classificação

- ▶ Dataset (Site Adoro Cinema)
- ▶ *Count-Vectorizer* vs. *TFIDF-Vectorizer*
- ▶ Sentimentos em “estrelas” (1-5)
- ▶ Classificação
- ▶ Link:
<https://colab.research.google.com/drive/1Jd7ywToSPb8b2gVtBFk03HkZKSanH9Cu>

Obtenção de Dataset



<https://www.kaggle.com/clorofila/adorocinema-csv>

Referências

AGUIAR, E. J. et al. **Análise de Sentimento em Redes Sociais para a Língua Portuguesa Utilizando Algoritmos de Classificação**. In: Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos. Porto Alegre: Sociedade Brasileira de Computação, 2018.

BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python**. O'Reilly, United States of America, 2009.

DENG L.; LIU Y. **Deep Learning in Natural Language Processing**. Springer Nature Singapore, 2018.

KOWSARI, K. *et al.* Text Classification Algorithms: A Survey. **Information**, v.10, n. 4, 150, 2019. doi:10.3390/info10040150.

LIU, B. **Sentiment Analysis and Opinion Mining**. Morgan & Claypool Publishers, 2012.

POZZI, F. A. *et al.* **Sentiment Analysis in Social Networks**. Elsevier, 2017.

SILVA N. F. F. **Análise de sentimentos em textos curtos provenientes de redes sociais**. 2016. 112 f. Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2016.

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

Duvidas???

Contato

[linkedin.com/clorophila](https://www.linkedin.com/company/clorophila)

github.com/clorophila

[kaggle.com/clorophila](https://www.kaggle.com/clorophila)