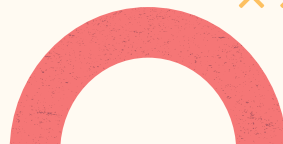
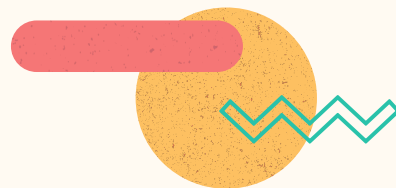


Análise de sentimentos dos usuários brasileiros no Twitter com relação ao COVID-19

Orientadora: Prof^a Dr^a Ana Régia de M. Neves
Aluno: Nathália Teixeira Guimarães



Agenda



01

Contextualização

- a) Problema
- b) Objetivo geral

03

Material e Métodos

- a) Base de dados
- b) Treinamento

02

Revisão Sistemática

- a) Questão norteadora
- b) Trabalhos Correlatos

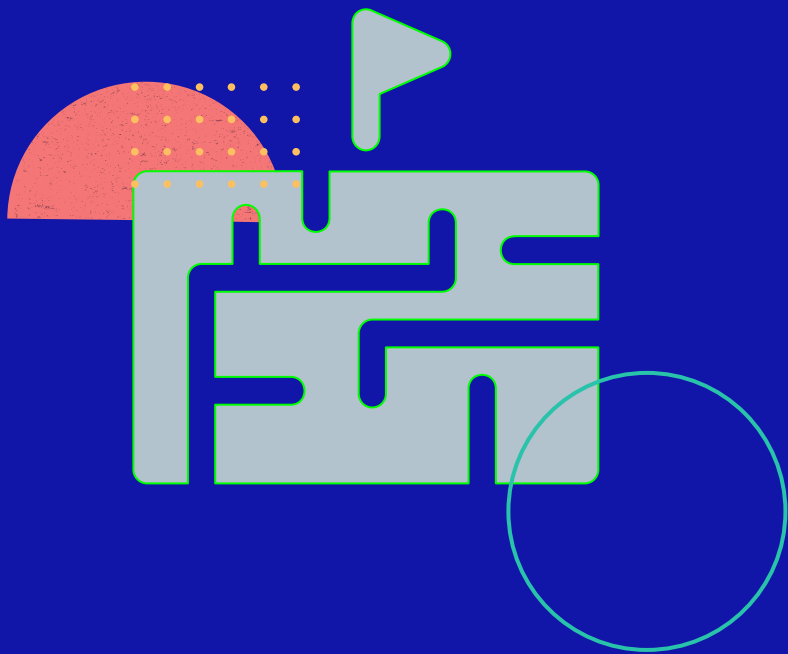
04

Resultados



05

Considerações Finais

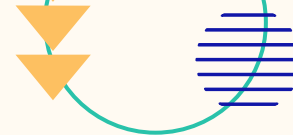


01

Contextualização



Problema



Pandemia da Covid-19

- Alastramento do vírus
- Medidas de contenção
- Aumento do uso das redes sociais¹
 - Twitter

¹<https://wearesocial.com/sg/blog/2021/07/digital-audiences-swell-but-there-may-be-trouble-ahead/>

Problema

Pandemia Covid-19

Impacto emocional¹

- Ansiedade
- Síndrome do pânico
- Medo da contaminação
- Solidão

¹<https://www.saude.gov.br/noticias/12609-pandemia-da-covid-19-provoca-impacto-negativo-na-saude-mental>



Problema



**Impacto
emocional da
pandemia**

• — Estudos

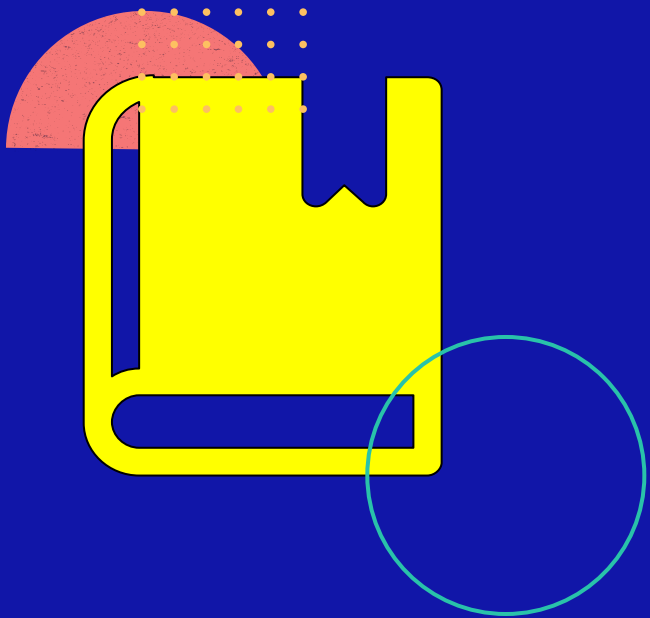
• — Comportamento
emocional

• — Redes sociais



Objetivo geral

Realizar uma análise de sentimentos dos *tweets* publicados pelos brasileiros sobre a Covid-19 durante o período de 4 de fevereiro a 7 de março de 2021 baseada nas seis emoções universais de Paul Ekman.



02

Revisão

Sistemática da
Literatura





Questão norteadora



Quais foram os principais sentimentos dos brasileiros no Twitter com relação à Covid-19 durante a pandemia?





Fluxograma linha de pesquisa

Linha 1) Análise de sentimento no *Twitter* sobre a Covid-19

Predominância: polaridade

Psicólogo Paul Ekman: raiva, nojo, medo, felicidade, tristeza e surpresa

Linha 2) Análise de sentimento no *Twitter* sobre a Covid-19
fazendo uso da Teoria dos Sentimento, de Paul Ekman

Critérios de Inclusão e Exclusão



Tabela 1. Critérios definidos para a revisão sistemática

Critérios de Inclusão	Critérios de Exclusão
Análise de sentimento no <i>Twitter</i>	Análise de sentimento fora do <i>Twitter</i>
Análise de sentimento para o idioma português no <i>Twitter</i>	Análise de sentimento para idiomas que não sejam o português no <i>Twitter</i>
Análise de sentimento no <i>Twitter</i> durante a pandemia da Covid-19	Análise de sentimento aplicada fora do <i>Twitter</i> durante o período da pandemia da Covid-19
Análise de sentimento fazendo uso dos sentimentos da teoria de Paul Ekman	Análise de sentimento sem o uso dos sentimentos da teoria de Paul Ekman



Fontes e String de Pesquisa



Tabela 2. Fontes e *Strings* de busca utilizadas

Base de dados	Palavra-chave	Resultados
IEEE XPLORE	(Covid19 OR Covid-19) AND (Twitter) AND (Sentimental Analysis OR Opinion Mining) AND (Paul Ekman) - Open Access	78
Google Scholar	(coronavirus)(covid19)(covid-19) (twitter) (sentimental analysis) (Paul Ekman) - Open Access - Artigos de Revisão	111
Google Scholar	(twitter) (sentimental analysis) (análise de sentimento) (análise de sentimentos) (português) - open access	546
Google Scholar	(twitter) (sentimental analysis) (análise de sentimento) (análise de sentimentos) (português)(Paul Ekman)	49
Periódicos Capes	(coronavirus OR covid19 OR covid-19) AND (twitter OR social network OR redes sociais)	698
Science Direct	(sentimental analysis OR opinion mining OR análise de sentimento) AND (twitter) AND (portuguese OR português) - open access	79



Seleção dos estudos

Identificação

Pesquisa

Incluído

Identificação de novos estudos em base de dados

Registros nas Bases de dados = 1.561

Registros Excluídos = 1.550

Estudos analisados = 11

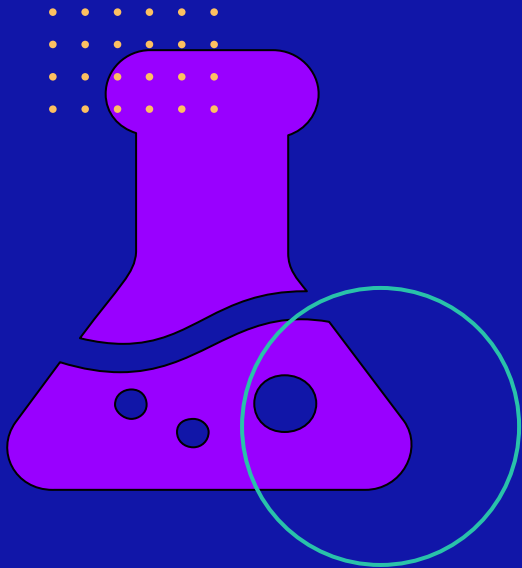
Registros excluídos:

- Duplicação
- Extrapolação objetivo
- Falta detalhamento
- Sem teoria de Paul Ekman

Métodos



- **Léxico:** lexiconPT
- **Aprendizado de máquina:** Naive Bayes

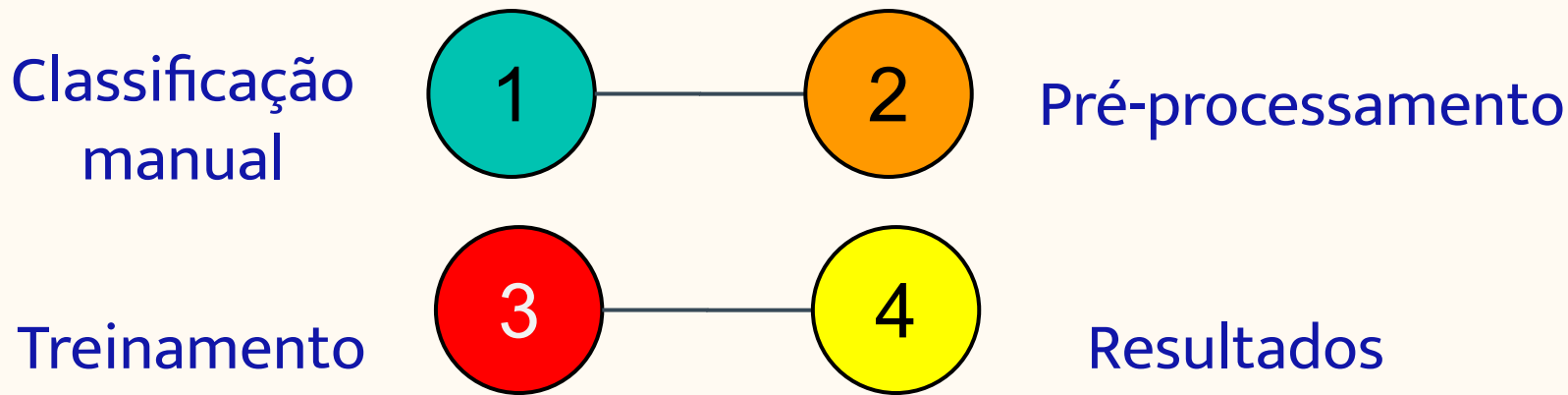


03

Material e Métodos

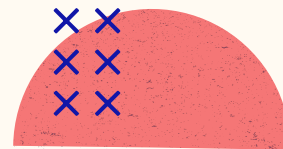


Fluxo Aprendizado Máquina¹



Base de dados utilizada

- ✗ Mineração de Texto no Twitter: uma ferramenta auxiliar na detecção de epidemias¹.
- ✗ Mineração no *Twitter*: entre 4 de fevereiro e 7 de março de 2021.
- ✗ Possui 7.179 mil tweets.



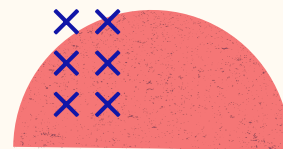
¹Simões and Neves 2021

Base de dados utilizada

✕ Google sheets

Polaridade; positivo, negativo, neutro

Paul Ekman: raiva, tristeza, nojo, felicidade, medo e surpresa.



Pré-processamento

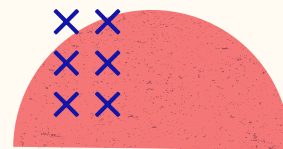
✗ Remoção de *tweets*:

- ✗ Não-relacionados à Covid
- ✗ Neutros na classificação de Paul Ekman

✗ Limpeza da base

- ✗ Caracteres especiais
- ✗ Palavras com menos de 3 caracteres

RESTARAM: 1.651 mil tweets

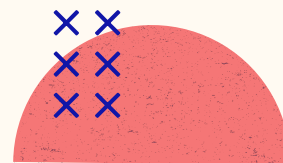


Pré-processamento

Tweets classificados - 1.651 mil

Polaridade

Negativo: **938** Neutro: **628** Positivo: **85**



Paul Ekman

Raiva: **602** Tristeza: **545** Medo: **241**
Felicidade: **144** Surpresa: **108** Nojo: **11**



RESTARTAM: 1.651 mil tweets

Treinamento

Plataforma: Google colab

Linguagem: Python

xx
xx
xx

Bibliotecas e pacotes

- **Manipulação:** Pandas, Numpy, NLTK, Matplotlib, Seaborn.
- **Treinamento:** Scikit-learn - train_test_split, CountVectorizer, Naive Bayes, MultinomialNB, cross_val_predict, confusion_matrix, accuracy_score, metrics, wordcloud

Treinamento

Classificador: MultinomialNB do Naive Bayes

Polaridade: 75% treinamento e 25% teste

Max features: 1000

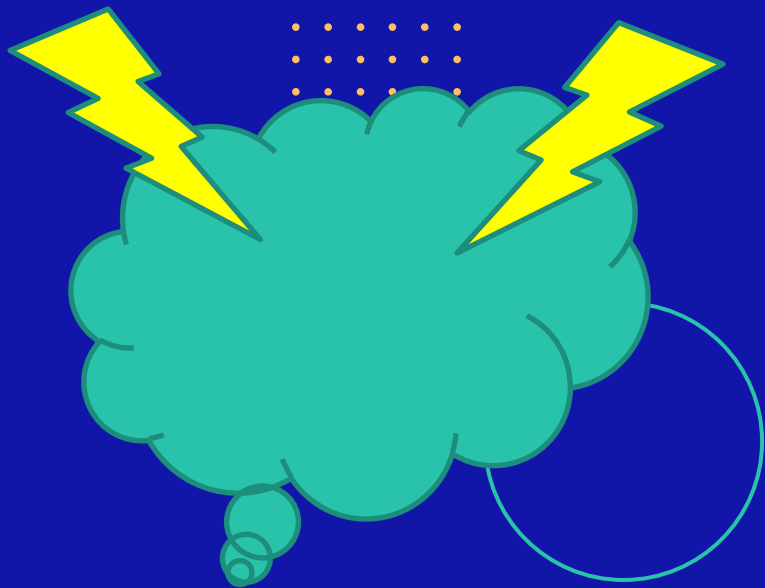
Validação cruzada: 15

x x
x x
x x

Paul Ekman: 70% treinamento e 30% teste

Max features: 1000

Validação cruzada: 15



04

**Resultados e
Discussão**

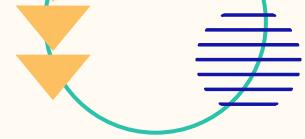




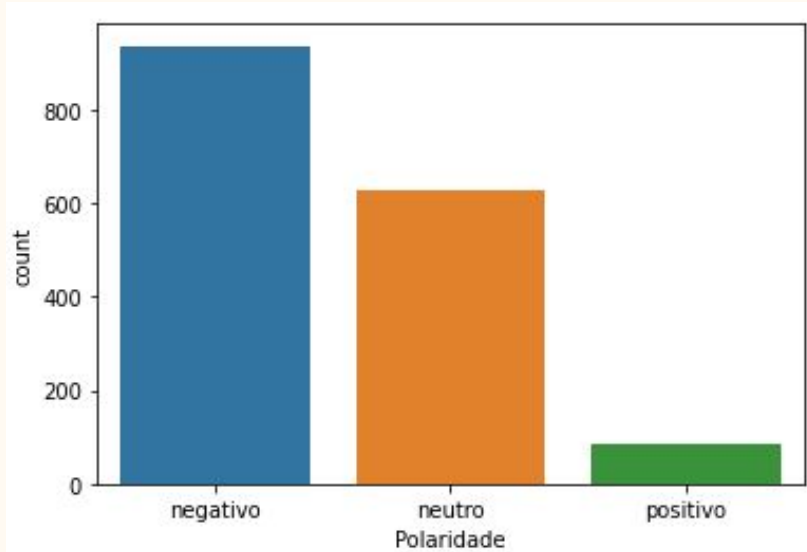


Negativo

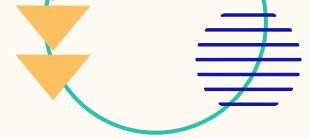




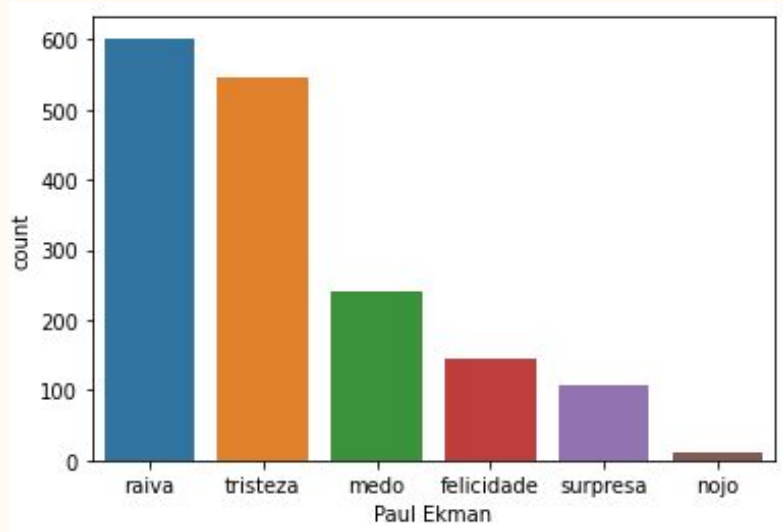
Métricas - Polaridade



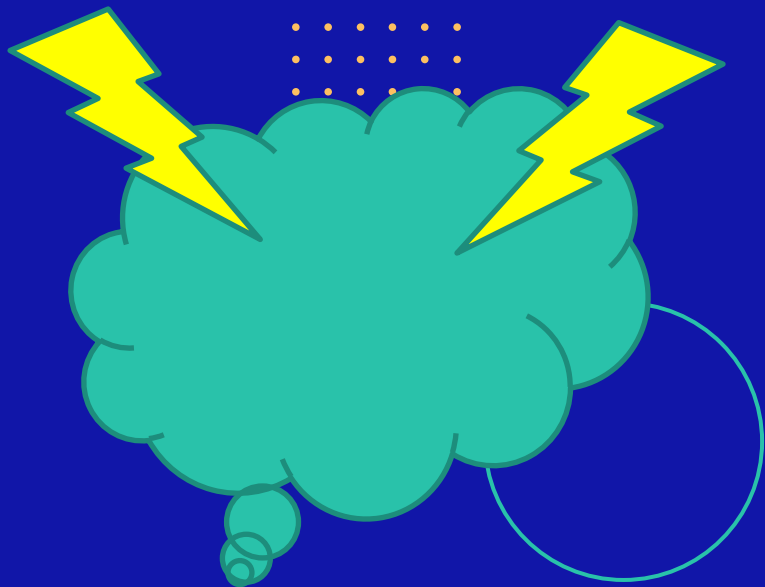
Modelo	<i>F1-score</i>	<i>Suporte</i>
<i>Negativo</i>	0.69	938
<i>Neutro</i>	0.47	628
<i>Positivo</i>	0.17	85



Métricas - Paul Ekman



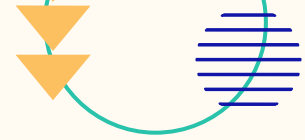
Modelo	<i>F1-score</i>	<i>Suporte</i>
<i>Felicidade</i>	0.29	144
<i>Medo</i>	0.39	241
<i>Nojo</i>	0.00	11
<i>Raiva</i>	0.67	602
<i>Surpresa</i>	0.07	108
<i>Tristeza</i>	0.58	545



05

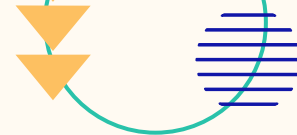
Considerações Finais





Considerações

- ✗ Observou-se que o sentimento predominante entre os brasileiros foi o de raiva, para a classificação sentimental de Paul Ekman, e o negativo, para a classificação de Polaridade. Essa predominância foi tão superior à das demais classes, que ocasionou um desbalanço da base, de modo a afetar a precisão da classificação dos modelos.
- ✗ Os modelos foram mais precisos nas classes que tinham maior volume amostral
- ✗ O sentimento de raiva esteve, principalmente, associado ao incômodo das pessoas por estarem sentindo algum sintoma da Covid-19.



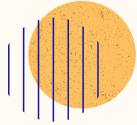
Considerações

O melhor F1-score obtido foi o de 0.67, na classificação de Paul Ekman e de 0.69 na classificação de Polaridade

Dificuldade na busca por trabalho que fazem uso de dicionário léxico para o português, o que facilitou a escolha do Aprendizado de Máquina no desenvolvimento desse projeto.

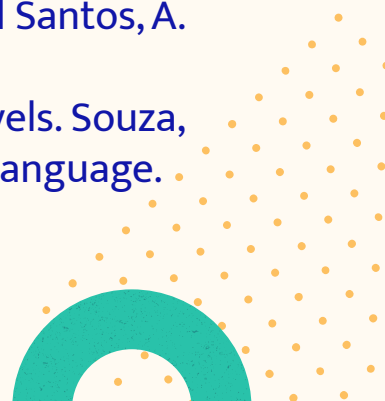
Dificuldade na limpeza da base de dados para uma melhor plotagem da nuvem de palavras e da vetorização dos termos na fase de treinamento dos modelos.

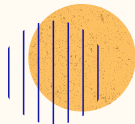
Usar base de dados com maior volume de tweets. Modificação das técnicas e parâmetros do Naive Bayes para obter um melhor resultado dos modelos.



Referências

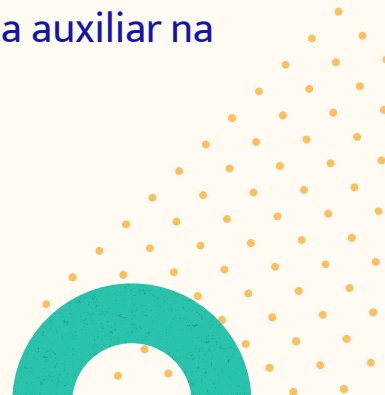
- × Araújo, M., Gonçalves, P., and Benevenuto, F. (2013).
- × Métodos para análise de sentimentos no twitter. Drus, Z. and Khalid, H. (2019).
- × Análise de sentimento em mídias sociais e sua aplicação: sistemática. Giachanou, A. and Crestani, F. (2016).
- × Like it or not: a survey of twitter sentiment analysis methods. Joaquim, C. E. L., Barbosa, C. H. M., and Ishikawa, E. (2021).
- × Análise de sentimentos da população brasileira durante a pandemia de covid-19 como ferramenta de exploração da expressão psicossocial no espaço cibernético. Longo, C. T. Y., Canovas, D. P. S., Fontes, A. S., Bessegato, L. F., and Santos, A. A. M. (2020).
- × Classification of futuristic technologies described in speculative fiction novels. Souza, M. and Vieira, R. (2012). Sentiment analysis on twitter data for portuguese language.
- ×

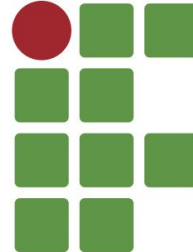




Referências

- × Desafios e repercussões da educação à distância no brasil em tempos de pandemia: um olhar à luz da análise de sentimentos no twitter. Pessanha, G. R. G., Fidelis, T. O., Freire, C. D., and Soares, E. A. (2020).
- × #fiqueemcasa: análise de sentimento dos usuários do twitter em relação ao covid-19. Silva, H., Andrade, E., Araújo, D., and Dantas, J. (2022).
- × Sentiment analysis of tweets related to sus before and during covid-19 pandemic. Silva, M. J., Carvalho, P., Costa, C., and Sarmento, L. (2010).
- × Automatic expansion of a social judgment lexicon for sentimento analysis. Simões, M. D. and Neves, A. R. M. (2021). Mineração de texto no twitter: uma ferramenta auxiliar na detecção de epidemias. Sittser, K. (2020).





**INSTITUTO
FEDERAL**
Brasília

Obrigada

