




Detecção de Posicionamento

Aluno: Diogo Pedro, IF867: Introdução a Aprendizagem Profunda



Apresentação



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are highlighted with a double-circle outline. The lines are thin and gray, creating a mesh-like structure.

1. Contexto

A decorative network diagram in the top-left corner, consisting of a cluster of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric rings, suggesting different levels of connectivity or importance. The lines are thin and gray, connecting the nodes in a complex, web-like structure.

Contexto/Motivação

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It features a cluster of interconnected nodes and lines, with some nodes being larger and having concentric rings. The lines are thin and gray, forming a complex web.

Redes Sociais

81%

Percentual das pessoas brasileiras como usuárias de internet¹

40 mi

De “tweets” referentes às eleições brasileiras de 2014²

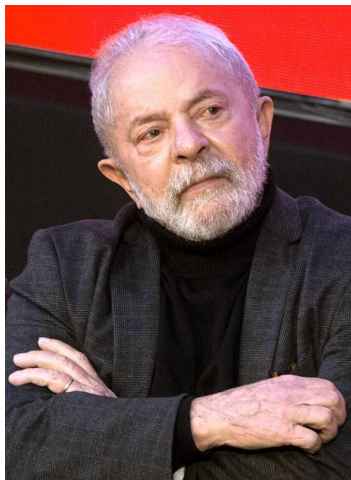


¹ TIC Domicílios 2020 (Edição COVID-19 - Metodologia Adaptada)

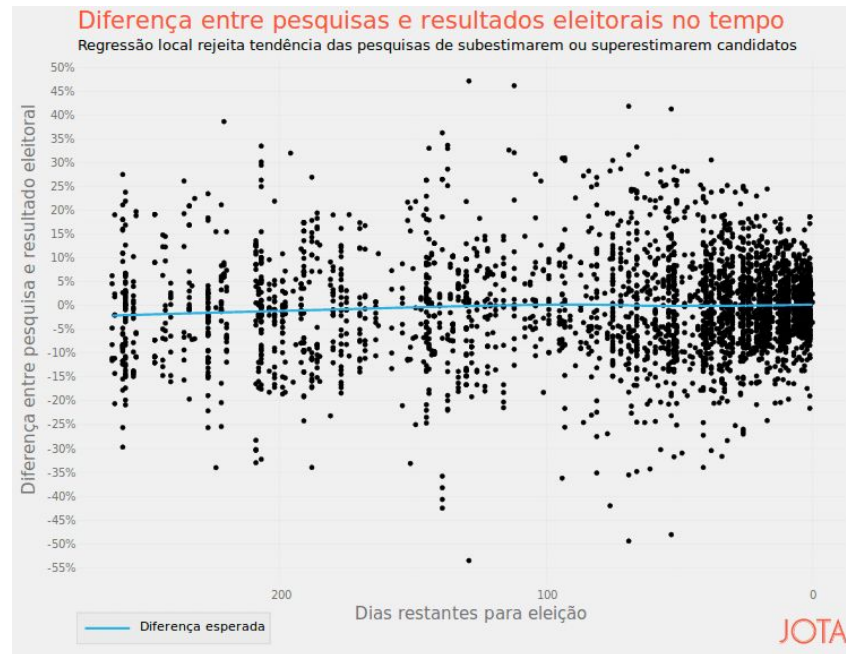
² Disponível em:

<http://g1.globo.com/politica/eleicoes/2014/noticia/2014/10/eleicoes-brasileiras-geraram-quase-40-milhoes-de-tuites-diz-twitter.html>

Detecção de posicionamento através dos "tweets" para as eleições presidenciais brasileiras de 2022 no segundo turno.



Justificativa



Objetivos do projeto

O objetivo deste trabalho é analisar os sentimentos de "tweets" da população brasileira a respeito dos posicionamentos políticos, isto é, julgamentos morais e comentários de cunho político-ideológico, dos candidatos à presidência do Brasil neste segundo turno.

Classificação do Usuário



Favor de Lula

Favor de Bolsonaro

Neutro

Trabalhos Relacionados

A Sentiment Analysis of Brazilian Elections Tweets

André L. Cristiani¹, Douglas D. Lieira^{2,3}, Heloisa A. Camargo¹

¹ Universidade Federal de São Carlos, UFSCar, Brazil
andre.cristiani@estudante.ufscar.br, heloisa@dc.ufscar.br

² Universidade Estadual Paulista, UNESP, Brazil
douglas.lieira@unesp.br

³ Instituto Federal de São Paulo, IFSP, Brazil

Análise de sentimentos das postagens
e comentários dos principais
candidatos à presidência durante a
corrida eleitoral de 2018

Rafael Francisco de Oliveira

TRABALHO DE
CONCLUSÃO DE CURSO

ORIENTAÇÃO:
Prof. Dr. Filipe Nunes Ribeiro

**Hate Towards the Political Opponent:
A Twitter Corpus Study of the 2020 US Elections on the
Basis of Offensive Speech and Stance Detection**

A stack of several books is shown, with the spines of the books visible. The books have various colored spines, including red, yellow, and blue. The text "Estado da Arte?" is overlaid in a large, bold, red font with a white outline, centered on the stack of books.

Estado da Arte?

Base de dados



Twitter API

Foi necessário criar uma nova base de dados para serem analisados, os motivos foram:

- Dados antigos não possuem grande grau de relevância.
- Dados focados apenas no segundo turno.
- Candidato à eleição presidencial no 2º turno é diferente se comparado com a última eleição presidencial.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, while others are smaller and solid. The lines are thin and gray, connecting the nodes in a non-linear fashion.

2. Metodologia

A stack of several books is shown, with the spines of the books visible. The books have various colored spines, including red, yellow, and blue. The text "Estado da Arte?" is overlaid in a large, bold, red font with a white outline. The background is a solid light gray.

Estado da Arte?

[illegible]

Staff research scientist at Google Brain



Thang Luong @lmthang · Oct 12, 2018

A new era of NLP has just begun a few days ago: large pretraining models (Transformer 24 layers, 1024 dim, 16 heads) + massive compute is all you need. BERT from @GoogleAI: SOTA results on everything arxiv.org/abs/1810.04805. Results on SQuAD are just mind-blowing. Fun time ahead!

Stanford University (Rajpurkar et al. '16)			
1 Oct 05, 2018	BERT (ensemble) Google A.I.	87.433	93.160
2 Oct 05, 2018	BERT (single model) Google A.I.	85.083	91.835
2 Sep 09, 2018	nlnet (ensemble) Microsoft Research Asia	85.356	91.202
2 Sep 26, 2018	nlnet (ensemble) Microsoft Research Asia	85.954	91.677
3 Jul 11, 2018	QANet (ensemble) Google Brain & CMU	84.454	90.490



13



510



1.1K



Bert - Artigo Original - 2018

BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on *eleven natural language processing tasks*, including pushing the GLUE score to 80.5% (7.7% point absolute improvement), MultiNLI accuracy to 86.7% (4.6% absolute improvement), SQuAD v1.1 question answering Test F1 to 93.2 (1.5 point absolute improvement) and SQuAD v2.0 Test F1 to 83.1 (5.1 point absolute improvement).

Bert

O Bert foi projetado para pré-treinar representações bidirecionais profundas de um texto não rotulado, trabalhando em contexto no lado esquerdo e direito em todas as camadas.

O modelo Bert pré-treinado pode ser ajustado com apenas uma camada de saída adicional para criar modelos de última geração para uma série de tarefas.

Bert

Novo paradigma:

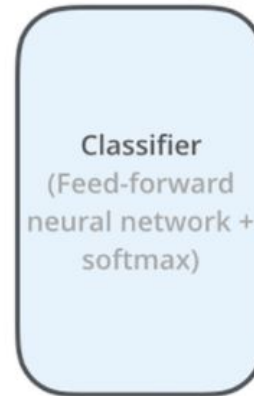
- Modelos pré-treinados
- Economiza Recursos: tempo, energia, processamento e.t.c
- Tarefas com alto desempenho de forma simplificada.

Bert

Input
Features

Output
Prediction

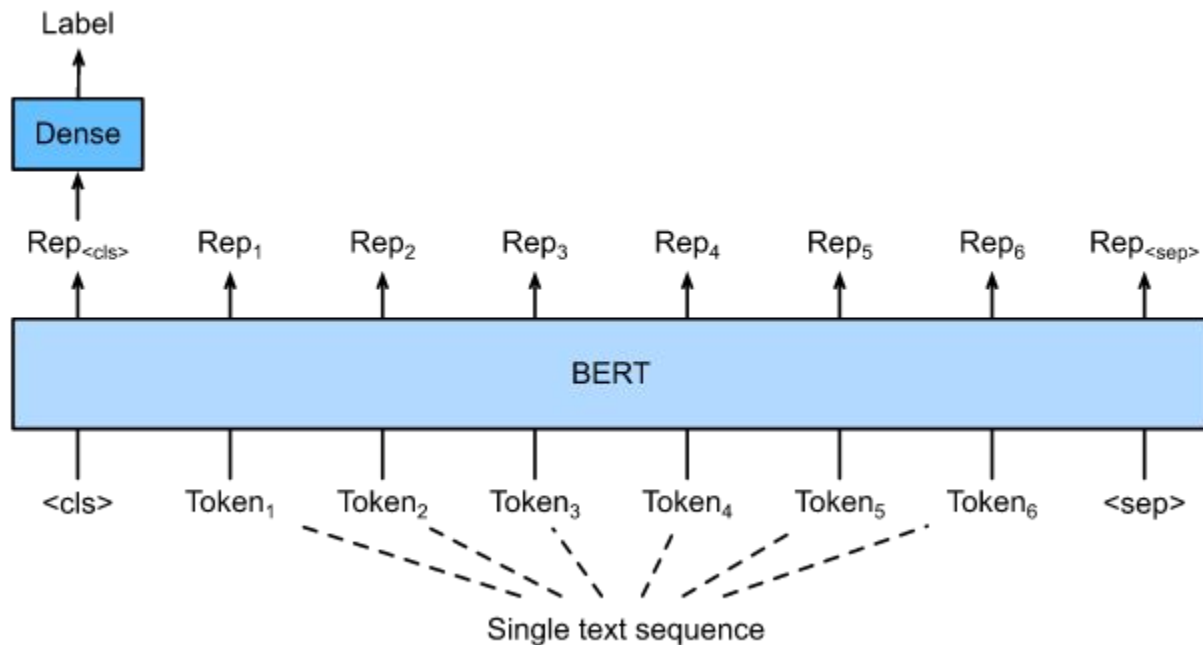
Help Prince Mayuko Transfer
Huge Inheritance



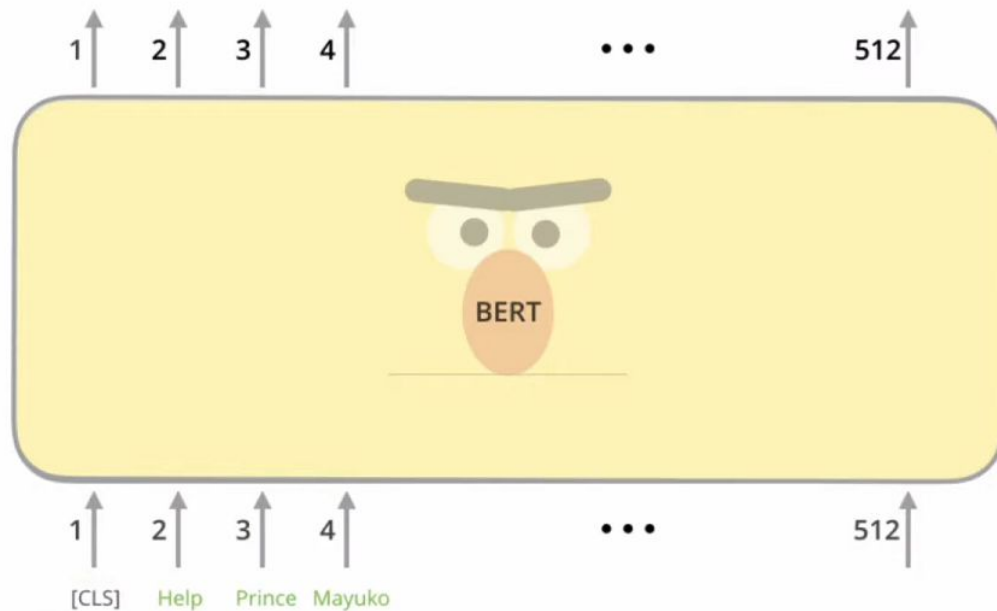
(Source: [jalammar.github.io](https://github.com/jalammar/jalammar.github.io))

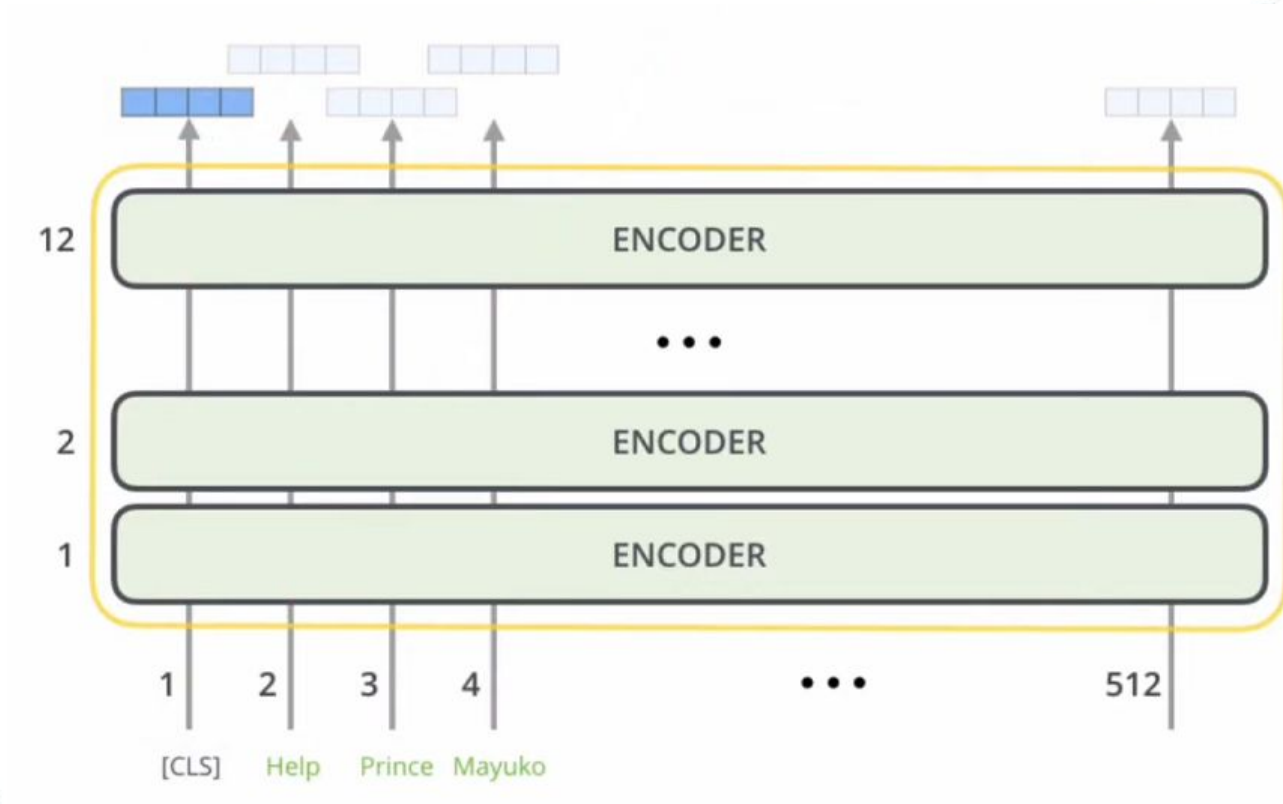
fine-tuning

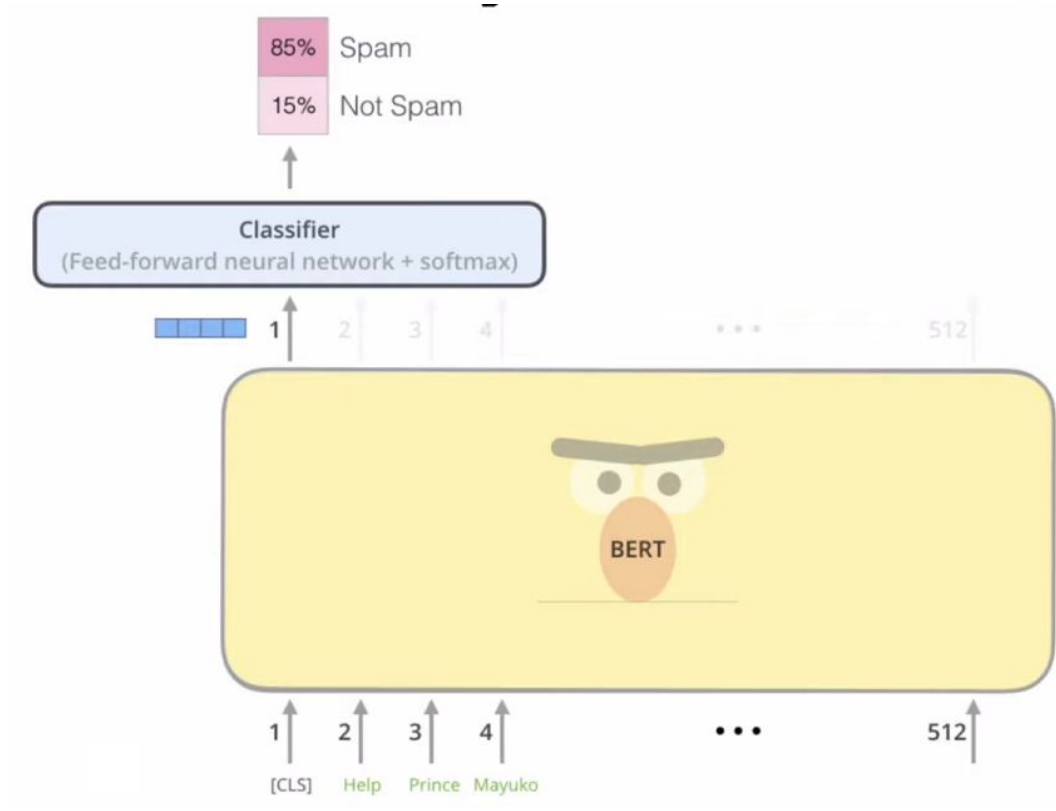
Bert - Entrada do Modelo



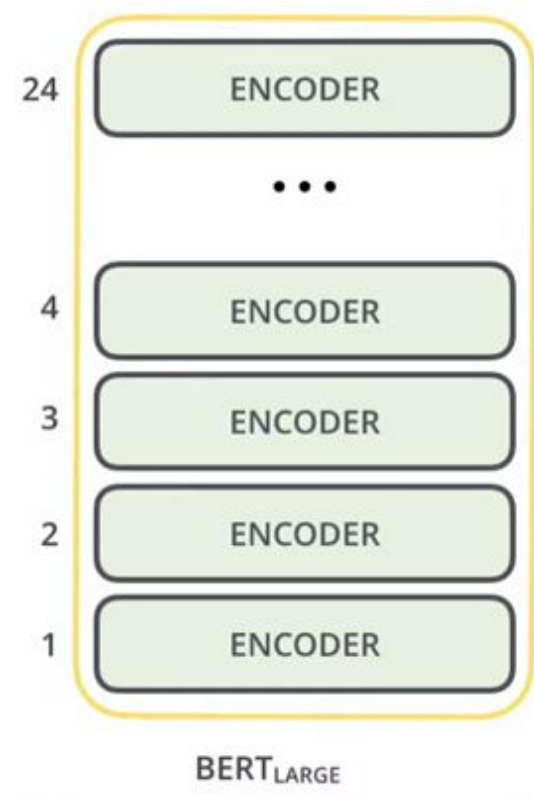
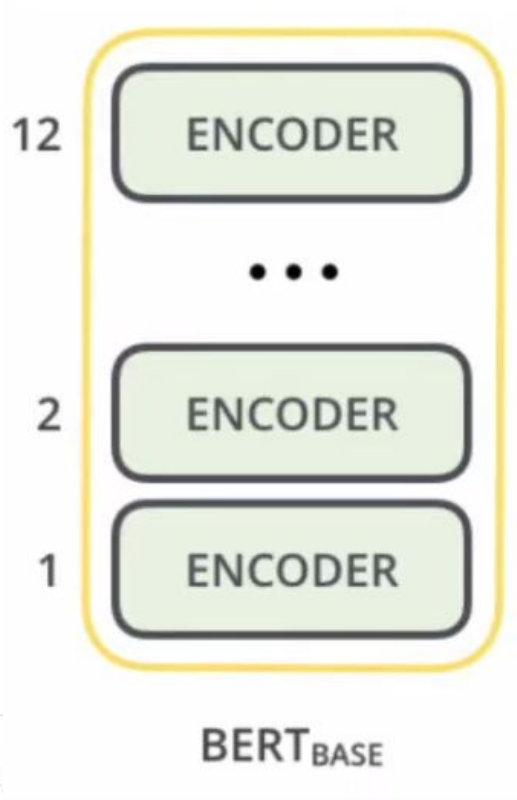
Entrada do modelo







Bert - Arquitetura



Exemplo:

```
[ ] PRE_TRAINED_MODEL_NAME = 'neuralmind/bert-base-portuguese-cased'  
  
    tokenizer = BertTokenizer.from_pretrained(PRE_TRAINED_MODEL_NAME)
```

```
[ ] sample_txt = 'Uma frase para a cadeira de aprendizagem profunda'
```

```
[ ] tokens = tokenizer.tokenize(sample_txt)  
    token_ids = tokenizer.convert_tokens_to_ids(tokens)  
  
    print(f' Sentence: {sample_txt}')    print(f'  Tokens: {tokens}')    print(f'Token IDs: {token_ids}')
```

Sentence: Uma frase para a cadeira de aprendizagem profunda

Tokens: ['Uma', 'frase', 'para', 'a', 'cadeira', 'de', 'aprendizagem', 'profunda']

Token IDs: [1431, 8422, 221, 123, 9104, 125, 11403, 6511]

Exemplo:

```
[ ] print(len(encoding['input_ids'][0]))  
    encoding['input_ids'][0]
```

```
16  
tensor([ 101, 1431, 8422, 221, 123, 9104, 125, 11403, 6511, 102,  
         0,    0,    0,    0,    0,    0])
```

```
[ ] print(tokenizer.convert_ids_to_tokens(encoding['input_ids'][0]))
```

```
['[CLS]', 'Uma', 'frase', 'para', 'a', 'cadeira', 'de', 'aprendizagem', 'profunda', '[SEP]', '[PAD]', '[PAD]', '[PAD]', '[PAD]', '[PAD]', '[PAD]']
```

Nova era para embeddings

Nova era de embeddings

Palavras precisam ser representadas numericamente

Word embeddings capturam semântica e relacionamentos

Word2vec, glove embeddings estáticos

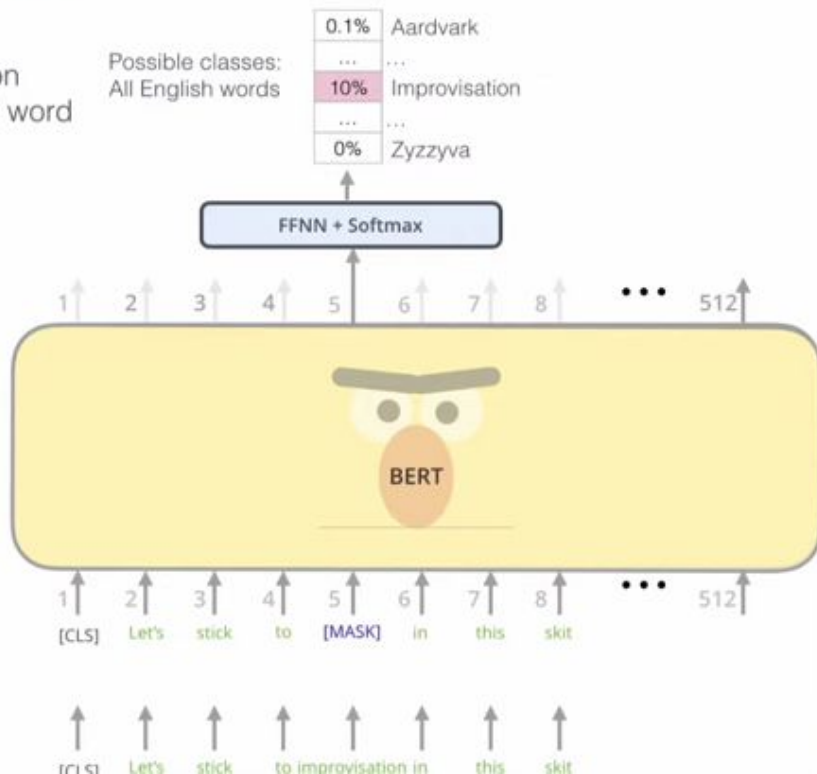
Ideia chave: embedding baseado no contexto atual

Embeddings contextualizados

Como era feito a esquerda, como Bert realiza a direita.

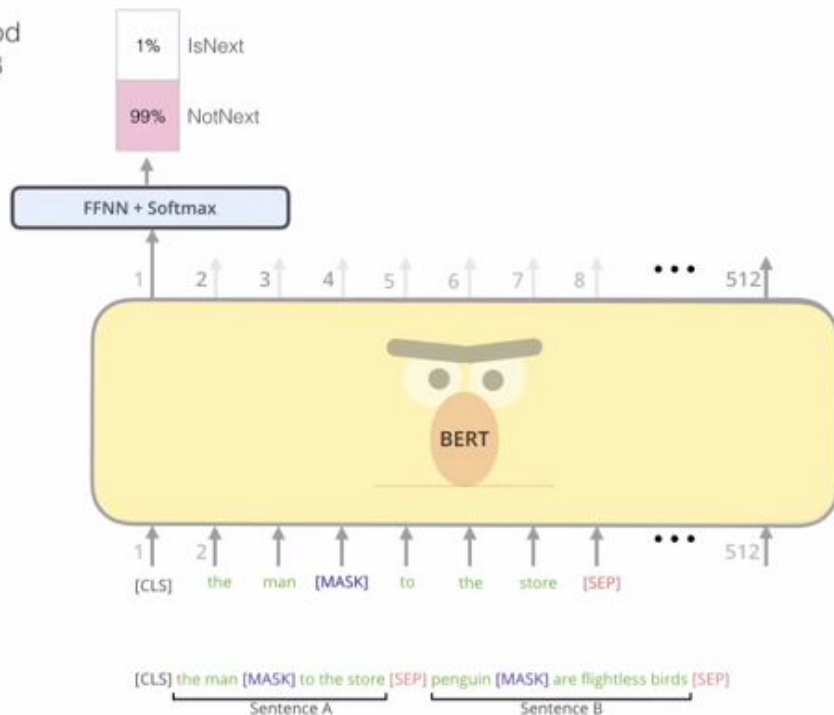
Modelo de Linguagem com máscaras

Use the output of the masked word's position to predict the masked word



Previendo relacionamento entre sentenças

Predict likelihood that sentence B belongs after sentence A



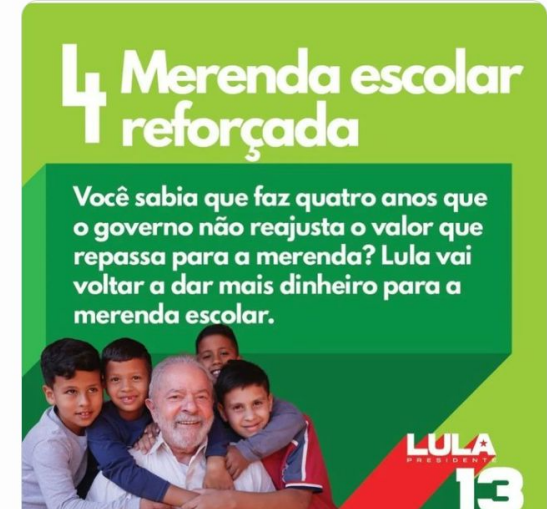
Coleta, Processamento e Rotulação dos dados

Palavras chaves: #Lula13, #Bolsonaro22, #BrasilDaEsperança, #Eleição2022, #Nulo

	index	Nome	Twitter	Favor do candidato Lula	Favor do candidato Bolsonaro	Neutro
0	0	helo tavares	wilsonprattis eu tenho fe. eu acredito.lula'nd...	1	0	0
1	1	marco siqueira	sen_cristovam bora deixar essa discussao pra d...	1	0	0
2	2	leonardo lula silva	cordeironereida bom dia. estou te seguindo, me...	1	0	0
3	3	ana	as mulheres;'na liberdade religiosa;'no lgbtqi...	1	0	0
4	4	claudio ou nene (meu apelido e nene mesmo)	mbendixens vote lula13 p manicure, empreendedo...	1	0	0
...
616	622	jaissonc	choquei se der um zoom na conta bancaria da ch...	0	0	1
617	623	cil77733322	cristinagraemi , inumeros canais estao replica...	0	1	0
618	624	silas fiorotti	"filha de bolsonaro [laurinha] entrara em cole...	0	0	1
619	625	fiorentino	priscillalleite aqui nao tem espaco para ladao...	0	1	0
620	626	murakami baldwin	lula e bolsonaro sao dois mentirosos pensam qu...	0	0	1
...



Beatriz Haddad Lula @beatriz13L · 3m
#Lula13 ❤️



Coleta, Processamento e Rotulação dos dados

heytauat 1 turno/todos os votos

57.259.505

51.072.345

4.915.423

3.399.287

600.955

559.708

81.129

53.519

45.620

25.625

16.604

(provavelmente) os votos da simo e ciro vao para lula, quase tds da sora tabm como bozo ira ganhar?

fora bolsonaro forabolsonaro lula13

lulapresidente13

hoje pr - tf - vote

e dia 22 -- 22

as 17:00

jairbolsonaro tarcisiogdf neymarjr

#bolsonaro22 #bolsonaroreeleitoem2022

presidente otario = sem aumento de salario

presidente legal = salario com aumento real

lula lulapresidente eleicao2022

o cagao, digo, ladrao, arregou e deixou o caminho livre!
eleicao2022 debatesbt

wilmarpetista yaiasobreira vidaloukanomar augustotomanov victoriosales jacck13 botvaniasavio nena51344002
isabelfeijmach1 ladcasannova jsxv josealoi seabahia rogriafranca1 inesnasciment12 azevedo4134 liz6macielgmai1
amandakamber mariasolangece1 tweetando12 salaminucci robsonjaraui1 silvioosaresvi1 valdivialiette borgescm
de_borgez ramoseneildes rocha_oni carlosm17244219 annas_gdottir talentospt13 oalagoas analiseagora alvaroluizcunha
elisangelacavad dijoaquarent13 frass_2 cludia15687211 marquesks botelhovaniac antonio43643750 cissomartir
anja191083 medinaolorado1 joao2608sanches che_zl neiva73551106 joseaze91686404 vniacarvalhobo1 cmsclaudio
lulaoficial lula13

sobre "pintou um clima", e preciso destacar q o assunto e muito serio e nojento, pois um idoso se convidou p entrar na casa de meninas menores de idade e depois saiu falando q eram garotas de programal! cade as notas de repudio??

venezuelanas

eleicao2022 lulapresidente13

Coleta, Processamento e Rotulação dos dados

Favor de Lula

239 Tweets

Favor de Bolsonaro

241 Tweets

Neutro

148 Tweets

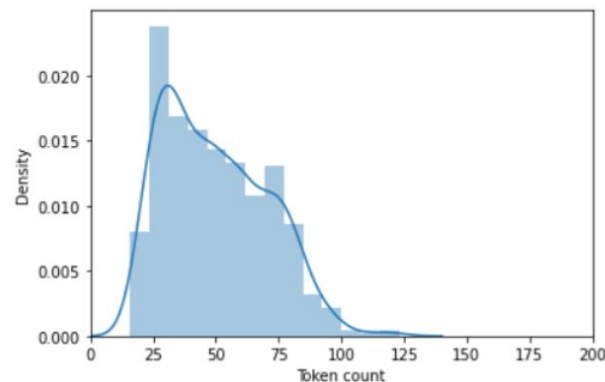
A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels or types of nodes. The lines are thin and gray, connecting the nodes in a non-linear fashion.

3. Resultados

```
[ ] PRE_TRAINED_MODEL_NAME = 'neuralmind/bert-base-portuguese-cased'

tokenizer = BertTokenizer.from_pretrained(PRE_TRAINED_MODEL_NAME)
```

```
df = pd.read_csv("Dados_processadoPrincipal - Dados_processadoV3.csv")
df = df.dropna()
df = df.drop('Unnamed: 0', axis=1)
df = df.reset_index()
```



	precision	recall	f1-score	support
Favor de Lula	0.95	0.91	0.93	22
Neutro	0.78	0.93	0.85	15
Favor de Bolsonaro	0.88	0.81	0.84	26
accuracy			0.87	63
macro avg	0.87	0.88	0.87	63
weighted avg	0.88	0.87	0.87	63

neymar ganhou pontos comigo, se mostrou um patriota, um brasileiro,
que se preocupa com os valores cristãos e com a família, seguimos
juntos pra reeleger nosso presidente bolsonaro22

True sentiment: Favor de Bolsonaro

Review text: Lula para presidente

Favor de Lula 0.990285

Neutro 0.008259

Favor de Bolsonaro 0.001456

Name: 0, dtype: float64

=====

Review text: vote 22 pelo bem do brasil!

Favor de Lula 0.000369

Neutro 0.000303

Favor de Bolsonaro 0.999328

Name: 0, dtype: float64

=====

Review text: Na moral, todos os candidatos são péssimos.

Favor de Lula 0.000983

Neutro 0.998557

Favor de Bolsonaro 0.000460

Name: 0, dtype: float64

=====

Review text: Lula é tão competente quanto bolsonaro

Favor de Lula 0.000503

Neutro 0.000888

Favor de Bolsonaro 0.998609

Name: 0, dtype: float64

=====

Review text: Precisamos do fim da esquerda

Favor de Lula 0.994512

Neutro 0.004675

Favor de Bolsonaro 0.000813

Name: 0, dtype: float64

=====

Review text: Bolsonaro é péssimo

Favor de Lula 0.734492

Neutro 0.263123

Favor de Bolsonaro 0.002386

Name: 0, dtype: float64

=====

Review text: A esquerda destruiu o Brasil, a salvação é #bolsonaro22

Favor de Lula 0.000156

Neutro 0.000309

Favor de Bolsonaro 0.999535

Name: 0, dtype: float64

=====

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting a hierarchical or central structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.

4. Conclusão

Conclusão

Nesse projeto foi observado que em tópicos importantes como as eleições presidenciais possuem artigos com métodos que envelheceram mal e nos possibilitam o uso de tecnologias mais recentes e produzir modelos com maior acurácia e precisão. O modelo Bertimbau tem a capacidade de identificar bem a análise de sentimento, contudo o modelo apresentado no projeto não conseguiu generalizar bem, isso ocorreu por conta da base de dados ser pequena e com uma representativa questionável, a ideia de criar uma documentação de rotulação de dados e definir regras é a única forma de garantir uma extração efetiva de dados com uma representatividade real, já que parte dos dados extraídos são confusos para serem entendidos, até mesmo por nós humanos.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, while others are smaller and solid. The lines are thin and gray, connecting the nodes in a non-linear fashion.

5. Referências

5. Referências

- ◎ [1] Cristiani, André; Lieira, Douglas; Camargo, Heloisa. A Sentiment Analysis of Brazilian Elections Tweets. In: SYMPOSIUM ON KNOWLEDGE DISCOVERY, MINING AND LEARNING (KDMILE), 8, 2020, Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 153-160. ISSN 2763-8944.
- ◎ [2] Akilandeswari J, Jothi G, Sentiment Classification of Tweets with Non-Language Features, Procedia Computer Science, Volume 143, 2018, Pages 426-433, ISSN 1877-0509
- ◎ [3] Souza, F., Nogueira, R., Lotufo, R. (2020). BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In: Cerri, R., Prati, R.C. (eds) Intelligent Systems. BRACIS 2020. Lecture Notes in Computer Science(), vol 12319. Springer, Cham.

5. Referências

- ◎ [4] Ema Kušen, Mark Strembeck. Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections. Online Social Networks and Media, Volume 5, 2018, Pages 37-50. ISSN 2468-6964.
- ◎ [5] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In Proceedings of the ACL 2012 System Demonstrations (ACL '12). Association for Computational Linguistics, USA, 115–120.
- ◎ [6] Stieglitz, Stefan & Dang-Xuan, Linh. (2013). Emotions and Information Diffusion in Social Media — Sentiment of Microblogs and Sharing Behavior. Journal of Management Information Systems. 29. 217-248. 10.2753/MIS0742-1222290408.

5. Referências

- ◎ [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- ◎ [8] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pages 353–355, Brussels, Belgium. Association for Computational Linguistics.

5. Referências

- ◎ [9] André L. Cristiani , Douglas D. Lieira, Heloisa A. Camargo. 2018. A Sentiment Analysis of Brazilian Elections Tweets.
- ◎ [10] Lara Grimminger, Roman Klinger. 2018. 2020. Hate Towards the Political Opponent: A Twitter Corpus Study of the 2020 US Elections on the Basis of Offensive Speech and Stance Detection. BlackboxNLP:
- ◎ [11] Mickel Hoang. 2021. Aspect-Based Sentiment Analysis Using BERT.