

# Processamento de Linguagem Natural

- Luís Felipe Messias

Especialização  
*Deep Learning*



Centro de  
Informática  
UFPE



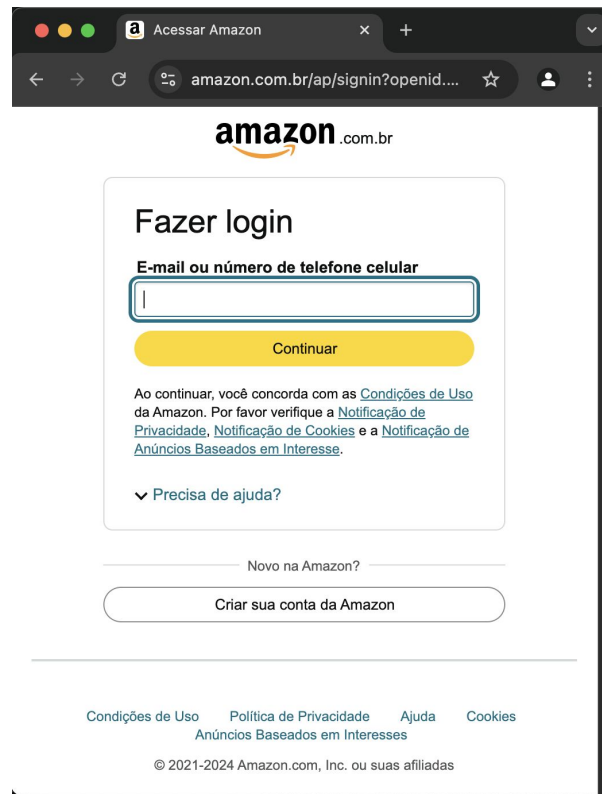
UNIVERSIDADE  
FEDERAL  
DE PERNAMBUCO

# Sumário

1. Dados
  - a. Web Scrapping Amazon.com
  - b. Dados Kaggle (Amazon Cell Phones Reviews)
2. Modelos
  - a. BOW + SVM
  - b. Embeddings + SVM
  - c. BERT
3. Resultados e conclusões

# Web Scrapping Amazon.com

# BeautifulSoup



# Web Scrapping Amazon.com

amazon.com.br Enviar para Luis Itabuna 45603295 Todos Pesquisar Amazon.com.br Olá, Luis Contas e Listas Devoluções e Pedidos Carrinho

Todos Venda na Amazon Comprar novamente luis Amazon.com.br Histórico de navegação Brinquedos e Jogos Atendimento ao Cliente Operação Natal - Assista agora

Apple iPhone 15 (128 GB) — Preto Avaliações de clientes

**Avaliações de clientes**  
★★★★★ 4,8 de 5  
3.342 classificações globais

Estilo: 128 GB Cor: Preto Alterar

5 estrelas 90%  
4 estrelas 6%  
3 estrelas 1%  
2 estrelas 1%  
1 estrela 3%

Escreva uma avaliação

Como as avaliações e classificações de clientes funcionam

Pesquisar Avaliações de clientes Pesquisar

CLASSIFICAR POR FILTRAR POR

Melhores avaliações Todos os avaliadores Somente 2 estrelas Todas as variantes Texto, imagem, vídeo

FILTRADO POR  
2 estrelas Limpar filtro  
19 classificações no total, 6 com avaliações

Veja todas as opções de compra

"Alexa, liga a luz."  
Monte sua casa inteligente  
Compre agora

# Web Scrapping Amazon.com



Apple iPhone 15 (128 GB) — Preto

Visite a loja Apple

4,8 ★★★★★ (3.342) | Pesquisar nesta página

Mais de 3 mil compras no mês passado

R\$ 5.100,53

Em até 10x R\$ 510,08 sem juros Ver parcelas disponíveis

Outros preços estão disponíveis em [mais opções de compra](#), com ofertas que podem não ser elegíveis para o Amazon Prime.

Pagamentos e Segurança Política de devolução

Estilo: 128 GB

128 GB	256 GB	512 GB
R\$ 5.100,53	R\$ 5.821,80	R\$ 7.321,11

Cor: Preto

Marca: Apple

Sistema operacional: iOS 16

Capacidade de armazenamento da memória: 128 GB

Tamanho da tela: 6,1 Polegadas

Nome do modelo: iPhone 15

```
>>> pd.read_csv("data/dataProcessed.csv")
Unnamed: 0      text  score
0      0  Sinceramente, foi-se o tempo que iPhones apres...      1
1      1                    Prezados,                        1
2      2  Demorei uns dias usando para fazer minha avali...      1
3      3  Nao gostei, comprei um iPhone achando que era ...      1
4      4  Celular zerado e já tenho que levar na Apple a...      1
..      ..      ...      ...
243    243  O aparelho veio bem embalado, muito bonito e c...      5
244    244                    Chegou no dia certinho, perfeito e lindo.      5
245    245  Produto excelente com entrega antes do prazo, ...      5
246    246                    Produto excelente e a entrega muito rapida.      5
247    247                    Ótimo produto. Superou minhas expectativas      5

[248 rows x 3 columns]
>>> 
```

# Dados Kaggle (Amazon Cell Phones Reviews)

- 720 produtos
- 67986 reviews

```
>>> pd.read_csv("Kaggle-dataset/raw/20191226-items.csv")
```

	asin	brand	title	...	totalReviews	price	originalPrice
0	B0000SX2UC	NaN	Dual-Band / Tri-Mode Sprint PCS Phone w/ Voice...	...	14	0.00	0.00
1	B0009N5L7K	Motorola	Motorola I265 phone	...	7	49.95	0.00
2	B000SKTZ0S	Motorola	MOTOROLA C168i AT&T CINGULAR PREPAID GOPHONE C...	...	22	99.99	0.00
3	B001A040UC	Motorola	Motorola i335 Cell Phone Boost Mobile	...	21	0.00	0.00
4	B001DCJAJG	Motorola	Motorola V365 no contract cellular phone AT&T	...	12	149.99	0.00
...	...	...	...	...	...	...	...
715	B07ZPKZSSC	Apple	Apple iPhone 11 Pro, 64GB, Fully Unlocked - Sp...	...	1	949.00	0.00
716	B07ZQSGP53	Xiaomi	Xiaomi Redmi Note 8, 32GB/3GB RAM 6.3" FHD+ Di...	...	3	150.96	0.00
717	B081H6STQ0	Sony	Sony Xperia 1 Unlocked Smartphone and WH1000XM...	...	70	948.00	0.00
718	B081TJFVCJ	Apple	Apple iPhone X, 64GB, Gray - Fully Unlocked (R...	...	1	478.97	0.00
719	B0825BB7SG	Samsung	Straight Talk Samsung Galaxy A10e Smartphone 5...	...	1	139.00	139.26

[720 rows x 10 columns]

URL: [amazon-cell-phones-reviews](https://www.kaggle.com/abhishek/amazon-cell-phones-reviews)

# Dados Kaggle (Amazon Cell Phones Reviews)

- 720 produtos
- 67986 reviews

```
>>> pd.read_csv("Kaggle-dataset/raw/20191226-reviews.csv")
```

	asin	name	...	body	helpfulVotes
0	B0000SX2UC	Janet	...	I had the Samsung A600 for awhile which is abs...	1.0
1	B0000SX2UC	Luke Wyatt	...	Due to a software issue between Nokia and Spri...	17.0
2	B0000SX2UC	Brooke	...	This is a great, reliable phone. I also purcha...	5.0
3	B0000SX2UC	amy m. teague	...	I love the phone and all, because I really did...	1.0
4	B0000SX2UC	tristazbimmer	...	The phone has been great for every purpose it ...	1.0
...	...	...	...	...	...
67981	B081H6STQQ	jande	...	I love the camera on this phone. The screen is...	1.0
67982	B081H6STQQ	2cool4u	...	I've been an Xperia user for several years and...	1.0
67983	B081H6STQQ	simon	...	buy one more for my cousin	NaN
67984	B081TJFVCJ	Tobiasz Jedrysiak	...	Product looks and works like new. Very much re...	NaN
67985	B0825BB7SG	Owen Gonzalez	...	I love the size and style of this phone. It is...	NaN

```
[67986 rows x 8 columns]
```

```
>>> █
```

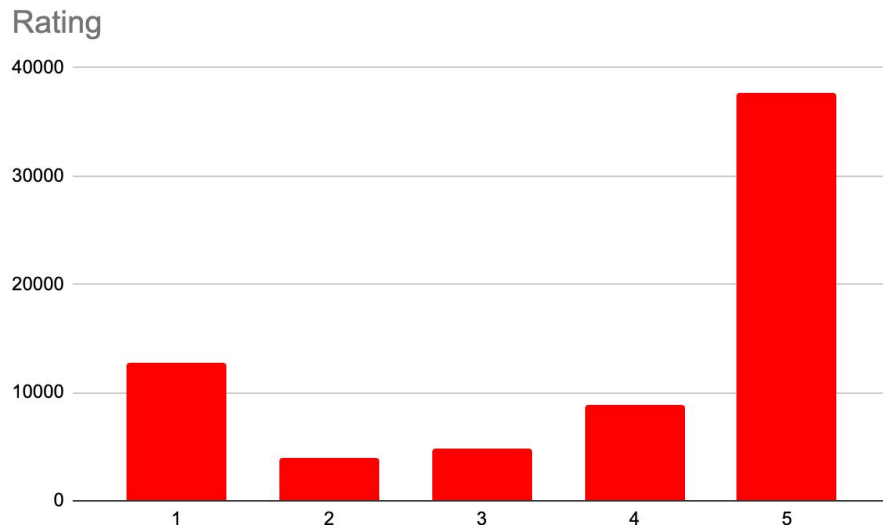
URL: [amazon-cell-phones-reviews](https://www.kaggle.com/datasets/alexisbcook/amazon-cell-phones-reviews)

# Dados Kaggle (Amazon Cell Phones Reviews)

- 720 produtos
- 67986 reviews

Split:

- 80% Treino
- 12% Validação
- 8% Teste



URL: [amazon-cell-phones-reviews](https://www.kaggle.com/abhishek/amazon-cell-phones-reviews)

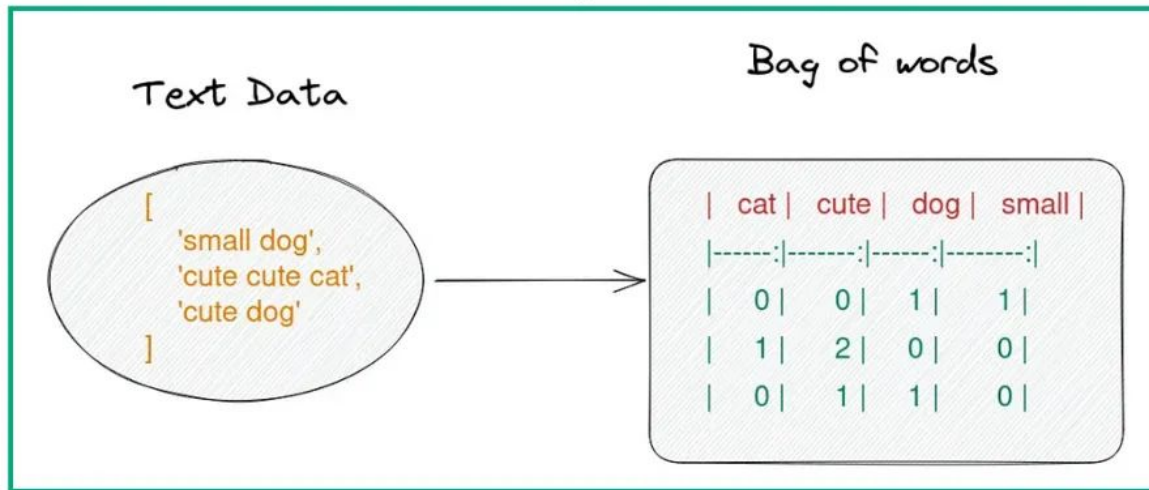


## Dados Kaggle (Amazon Cell Phones Reviews)



## BOW + SVM

- Acurácia: 0.745
- F1:
  - 0.749
  - 0.171
  - 0.295
  - 0.327
  - 0.859



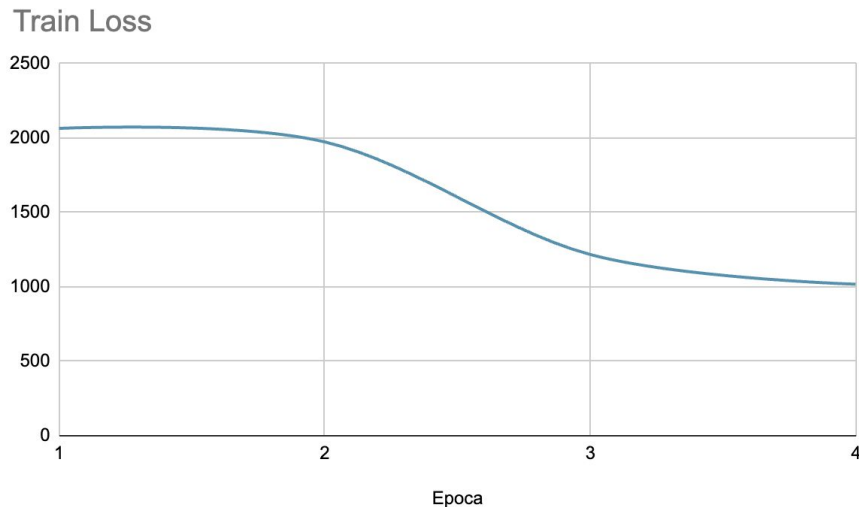
Fonte: <https://ayselaydin.medium.com/4-bag-of-words-model-in-nlp-434cb38cdd1b>

# Embeddings + SVM

- Embeddings:
  - Spacy: en\_core\_web\_sm
  - URL: [https://spacy.io/models/en#en\\_core\\_web\\_sm](https://spacy.io/models/en#en_core_web_sm)
- Acurácia: 0.659
- F1:
  - 0.594
  - 0.021
  - 0.037
  - 0.156
  - 0.789

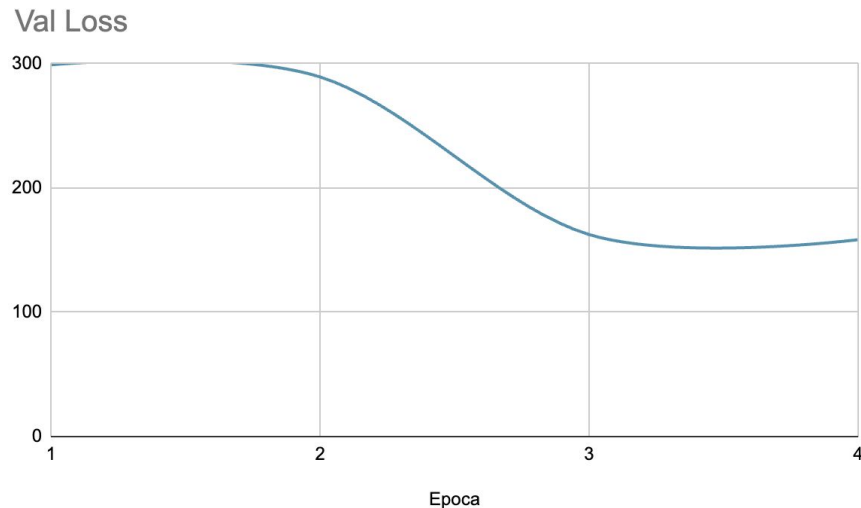
# BERT

- BERT: [url](#)
- 2 épocas de treinamento apenas para o classificador com  $lr=1e-4$
- 2 épocas de treinamento para todo o modelo com  $lr=1e-5$
- Acurácia: 0.767
- F1:
  - 0.777
  - 0.234
  - 0.409
  - 0.431
  - 0.897



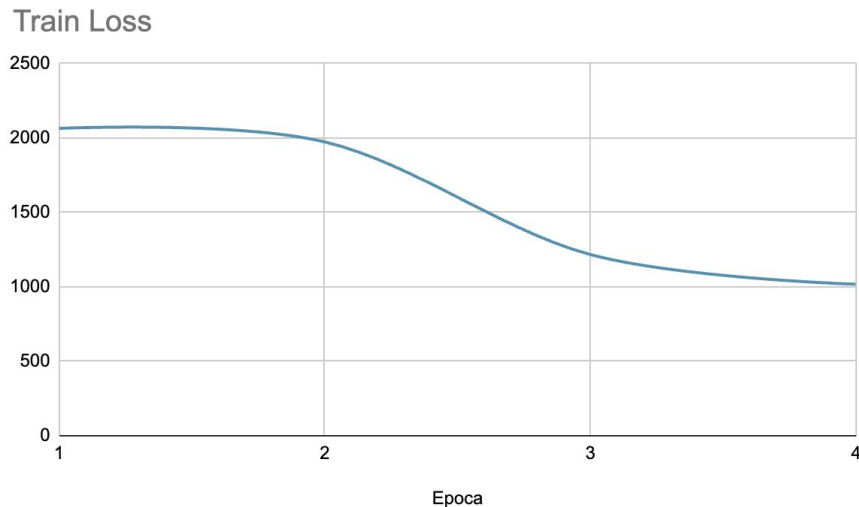
# BERT

- BERT: [url](#)
- 2 épocas de treinamento apenas para o classificador com  $lr=1e-4$
- 2 épocas de treinamento para todo o modelo com  $lr=1e-5$
- Acurácia: 0.767
- F1:
  - 0.777
  - 0.234
  - 0.409
  - 0.431
  - 0.897



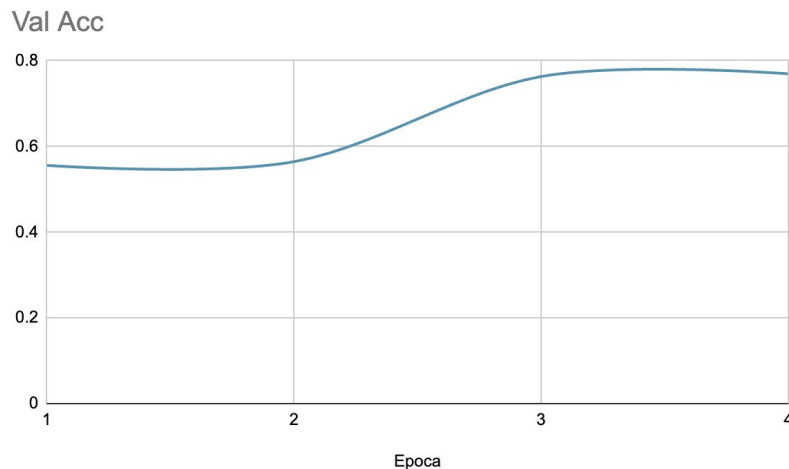
# BERT

- BERT: [url](#)
- 2 épocas de treinamento apenas para o classificador com  $lr=1e-4$
- 2 épocas de treinamento para todo o modelo com  $lr=1e-5$
- Acurácia: 0.767
- F1:
  - 0.777
  - 0.234
  - 0.409
  - 0.431
  - 0.897



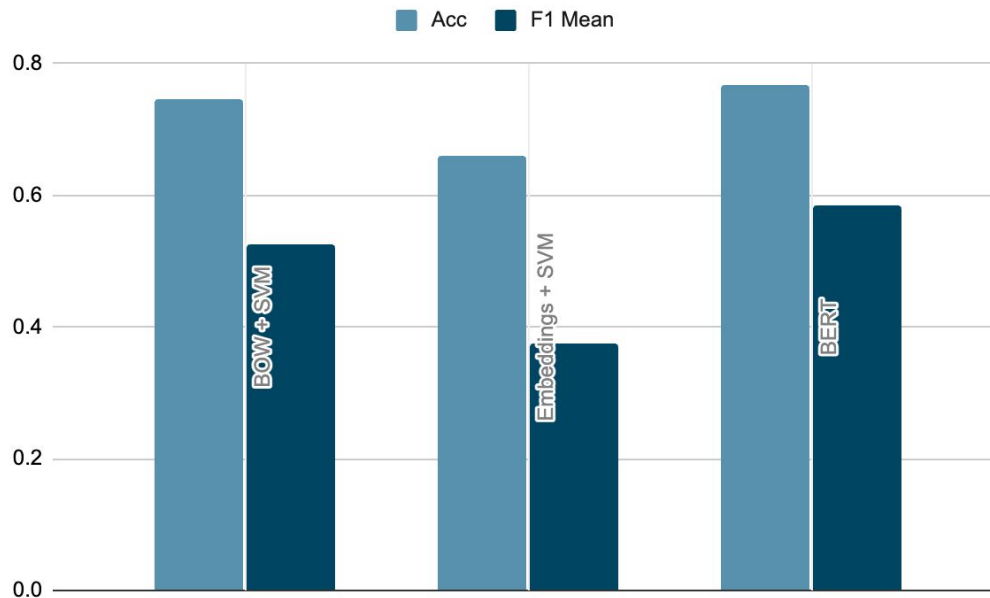
# BERT

- BERT: [url](#)
- 2 épocas de treinamento apenas para o classificador com  $lr=1e-4$
- 2 épocas de treinamento para todo o modelo com  $lr=1e-5$
- Acurácia: 0.767
- F1:
  - 0.777
  - 0.234
  - 0.409
  - 0.431
  - 0.897



# Resultados e conclusões

Acurácia e F1 médio para o conjunto de teste





# Obrigado

Especialização  
*Deep Learning*



Centro de  
Informática  
UFPE



UNIVERSIDADE  
FEDERAL  
DE PERNAMBUCO