

# Classificador de Avaliações de Produto

Gustavo Aires Tiago

Especialização  
*Deep Learning*



Centro de  
Informática  
UFPE



UNIVERSIDADE  
FEDERAL  
DE PERNAMBUCO

# Introdução

Problema

Desenvolver três classificadores de avaliação de produto

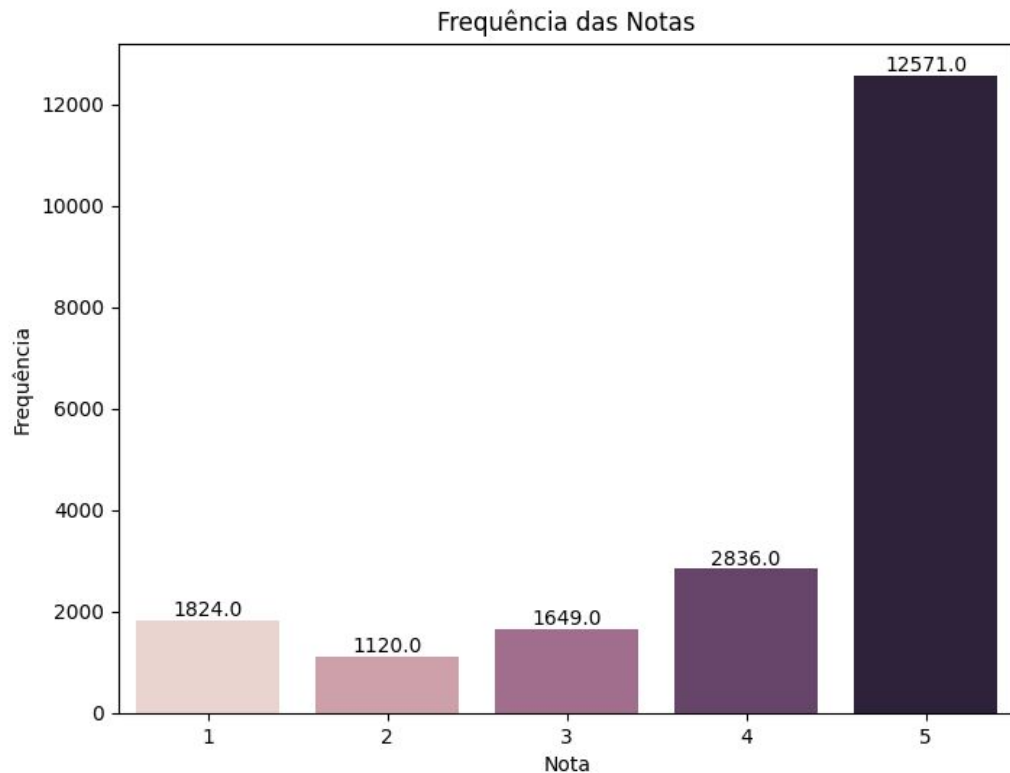
Dados

Avaliações retiradas da Amazon (20 mil avaliações)

Proposta

SVM + Bag of Words (BoW)  
SVM + Embeddings  
BERT (Bidirectional Encoder  
Representations from Transformers)

# Frequência das notas



# Etapas

## Entrada dos dados:

- 20 mil avaliações com notas de 1 a 5

## Pré Processamento

- Limpeza de texto (remoção de stopwords, pontuações, etc.)
- Tokenização

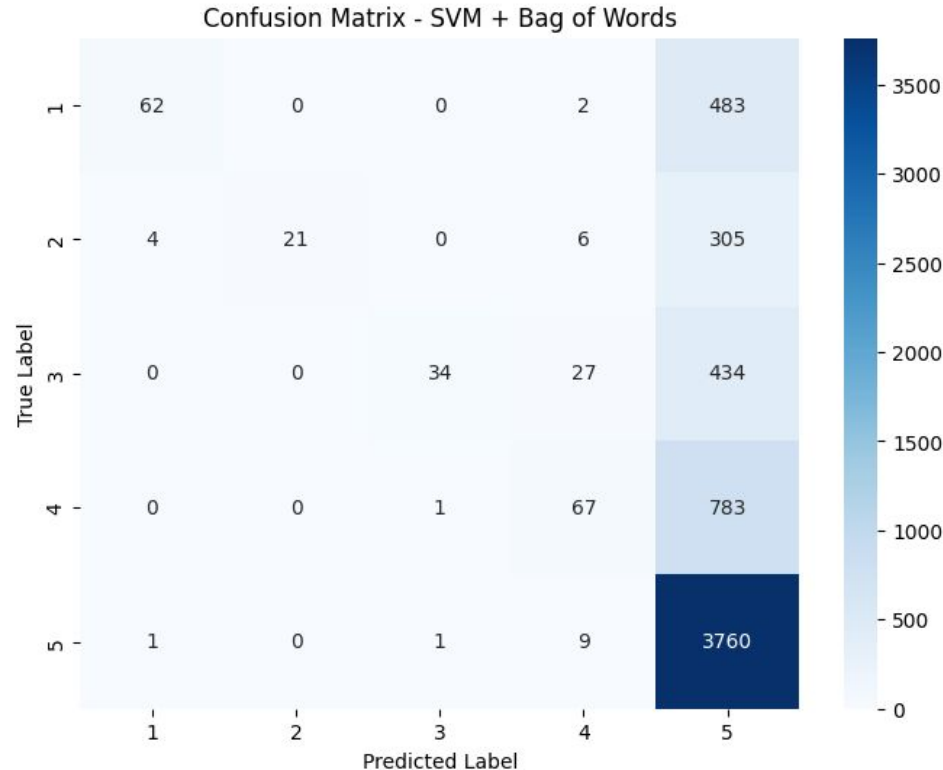
## Treinamento e Teste:

- Separação do conjunto de treinamento (70%) e teste (30%)
- Balanceamento das notas em cada conjunto

# SVM (Bag of words)

SVM + Bag of Words					
	precision	recall	f1-score	support	
1	0.93	0.11	0.20	547	
2	1.00	0.06	0.12	336	
3	0.94	0.07	0.13	495	
4	0.60	0.08	0.14	851	
5	0.65	1.00	0.79	3771	
accuracy			0.66	6000	
macro avg	0.83	0.26	0.28	6000	
weighted avg	0.71	0.66	0.55	6000	
Accuracy: 0.6573333333333333					
F1-score: 0.5509503211619117					

# Matriz de Confusão - SVM (Bag of words)

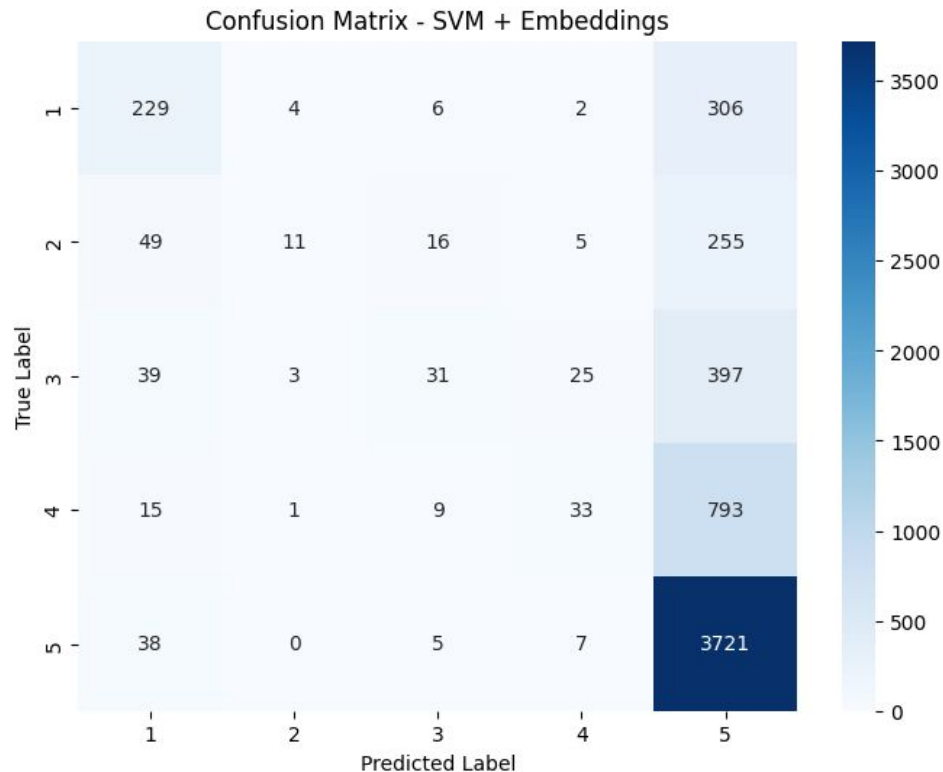


# SVM (Embeddings)

- Download do modelo “en\_core\_web\_md”
- Acurácia de 67% e F1 de 57%. (Maior do que com Bag of Words)
- F1 superior nas notas extremas

SVM + Embeddings				
	precision	recall	f1-score	support
1	0.62	0.42	0.50	547
2	0.58	0.03	0.06	336
3	0.46	0.06	0.11	495
4	0.46	0.04	0.07	851
5	0.68	0.99	0.81	3771
accuracy			0.67	6000
macro avg	0.56	0.31	0.31	6000
weighted avg	0.62	0.67	0.57	6000
Accuracy: 0.6708333333333333				
F1-score: 0.5742840750572394				

# Matriz de Confusão - SVM (Embeddings)





# BERT

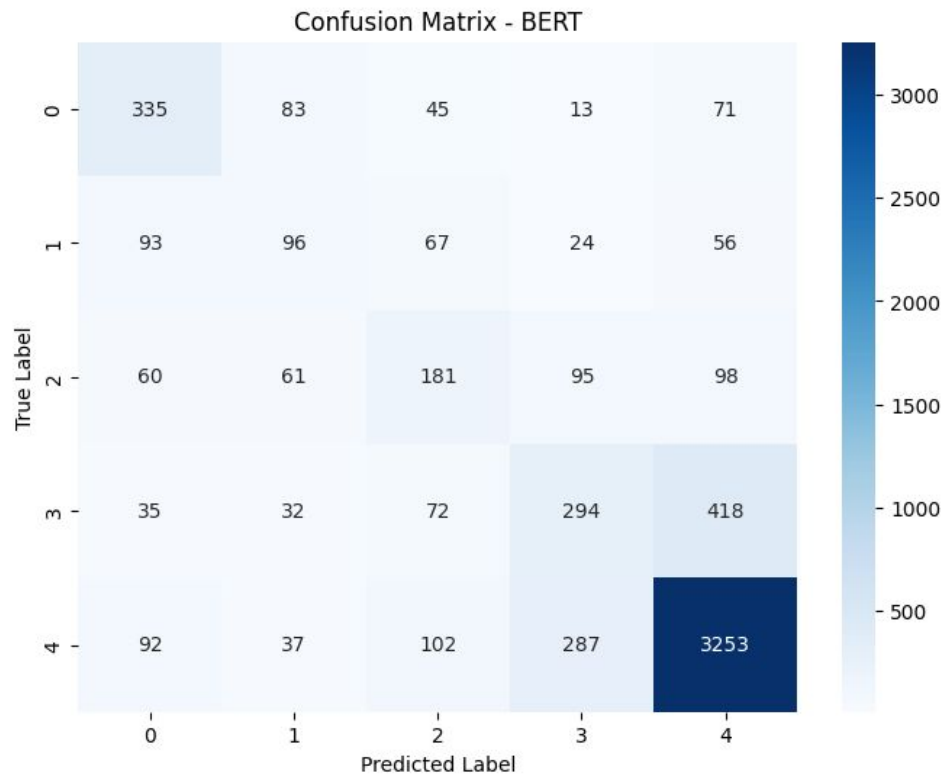
- Modelo pré-treinado utilizado: distilbert-base-uncased
- Treinamento com 10 épocas
- Performance superior aos dois últimos métodos utilizados
- A matriz de confusão apresenta erros mais próximos a nota mais positiva e mais negativa. Ou seja, a performance acaba sendo superior, já que notas mais altas estão sendo classificadas como “positivas” e mais baixas como “negativas”. Mesmo com o erro da nota

**Perda no teste (Loss): 1.6080**

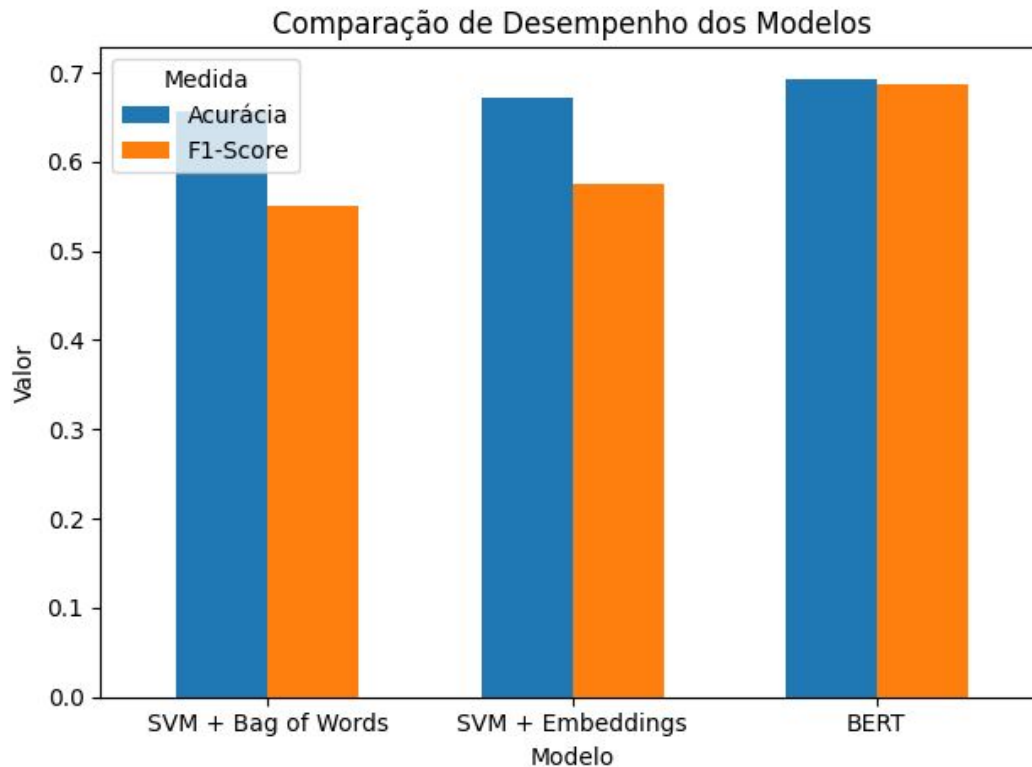
**Acurácia: 0.6932**

**F1 Score (ponderado): 0.6869**

# Matriz de Confusão - BERT



# Comparação dos modelos



# Conclusão

- **BERT superior aos outros modelos:** O modelo BERT obteve a maior acurácia e F1-Score, indicando que ele aprendeu representações mais robustas e discriminativas para a tarefa de classificação.
- **Embeddings melhoram o desempenho:** A utilização de embeddings em vez de Bag of Words resultou em uma melhoria no desempenho do modelo SVM, sugerindo que as embeddings capturam informações semânticas mais relevantes.