

Projeto Final de NLP

Curso: Especialização em Deep
Learning - UFPE

Disciplina: Processamento de
Linguagem Natural

Autor: Gabriel Souza Marques

Introdução

- **Objetivo do projeto:** Explorar e resolver um problema de NLP utilizando técnicas de Deep Learning.
- **Contexto:** Aplicar conceitos teóricos da disciplina em um problema prático.
- **Tecnologias utilizadas:** Python, bibliotecas como TensorFlow, PyTorch, scikit-learn, e pandas.

Metodologia

- Definição do Problema:

Coletar avaliações de um produto selecionado

Treinar classificadores de sentimentos:

- SVM + bow
- SVM + embeddings
- BERT

Métricas de performance:

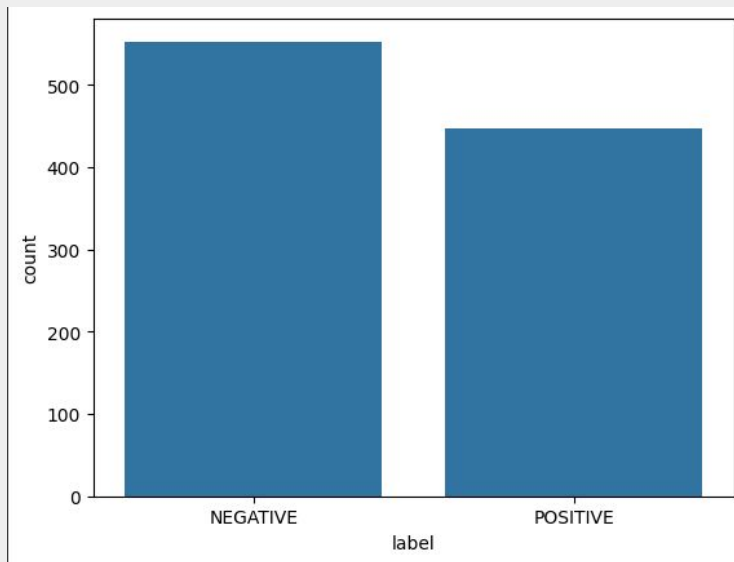
- Acurácia
- F1-score
- Precisão
- Recall

Dataset

| Review | label |
|---|----------|
| I don't know what happened between now and las... | NEGATIVE |
| AwesomeðŸˆŽ to use for hands free ðŸ™ˆ listeni... | POSITIVE |
| love spotify. but when I close all my tabs aft... | NEGATIVE |
| Buggy, stops all of the sudden, cannot pause, ... | NEGATIVE |
| Very great quality, only backfire is that ever... | POSITIVE |

Spotify User Reviews

51.000 avaliaões de usurios do aplicativo Spotify da Google Play Store



Pré- processamento

- **Limpeza:**
 - **Remoção de caracteres especiais**
 - **Remoção de mais de uma ocorrência de espaços em branco**
 - **Remoção de espaços no início e no fim**
 - **Conversão para letras minúsculas**
- **Tokenização**
- **Remoção de stopwords**
- **Lematização**
- **Stemming**

Extração de features

- **Bag of words**
 - **CountVectorizer**
- **Embeddings**
 - **Word2Vec**

Treinamento

- **80% de treino / 20% para testes**
- **SVM**
 - **Parâmetros padrão**
- **Bert**
 - **10 epochs**
 - **Batch size 32**
 - **Otimizador**
 - **AdamWeightDecay**
 - **Learning rate inicial: 2e-5**
 - **Beta 1 = 0.9**
 - **Beta 2 = 0.999**
 - **epsilon=1e-6**
 - **Loss**
 - **SparseCategoricalCrossentropy**

Resultados

- **SVM + BoW**
 - **Acurácia: 0.8350**
 - **F1-score: 0.8344**
 - **Precisão: 0.8372**
 - **Recall: 0.8350**
- **SVM + Embeddings**
 - **Acurácia: 0.7000**
 - **F1-score: 0.6977**
 - **Precisão: 0.7023**
 - **Recall: 0.7000**
- **BERT**
 - **Acurácia: 0.8**
 - **F1-score: 0.8**
 - **Precisão: 0.8**
 - **Recall: 0.8**

Obrigado!