



Universidade Federal
de São João del-Rei

Trabalho Prático 2 - Inteligência Artificial

Arthur Antunes Santos Silva e Lucas Nojiri

Documentação do trabalho prático da disciplina de Inteligência Artificial, ministrada pela professora Erica da Costa Reis Carvalho, do curso de Ciência da Computação da Universidade Federal de São João del-Rei.

São João del Rei

Julho de 2025

Sumário

1	Introdução	2
2	Metodologia	2
3	Resultados	2
3.1	Testes com Comentários Positivos	3
3.2	Testes com Comentários Negativos	3
3.3	Testes com Comentários Ambíguos, Neutros e Sarcásticos	4
3.4	Análise das Limitações em Comentários Ambíguos	7
3.4.1	O Modelo Foi Treinado para uma Tarefa Binária	7
3.4.2	Dificuldade em Compreender Nuances Humanas	7
3.4.3	O Objetivo é a Polaridade, Não a Compreensão	7
4	Conclusão	8
5	Referências	8

1 Introdução

O presente projeto tem como objetivo o desenvolvimento de um sistema de Inteligência Artificial capaz de realizar a análise de sentimentos em textos de avaliações de produtos. Com o aumento expressivo do volume de reviews em plataformas online, a análise manual tornou-se uma tarefa impraticável. A automação desse processo oferece a empresas e consumidores insights valiosos e rápidos sobre a percepção de um produto, permitindo a identificação de pontos fortes e fracos a partir das opiniões dos usuários.

O público-alvo principal da aplicação são os setores de atendimento ao cliente e comércio, que podem se beneficiar diretamente da análise automatizada de feedback para aprimorar seus produtos e serviços.

2 Metodologia

Para o desenvolvimento do sistema, foi utilizada a plataforma Hugging Face e o modelo de linguagem DistilBERT, especificamente a versão **distilbert-base-uncased-finetuned-sst-2-english**, pré-treinada para análise de sentimentos. O conjunto de dados escolhido para treinar e avaliar o modelo foi o "Amazon Polarity", que contém avaliações de produtos da Amazon. Assim o modelo DistilBERT é treinado usando o Amazon Polarity. Em resumo código utiliza um modelo treinado no SST-2 e aplica esse conhecimento para analisar os textos do dataset Amazon Polarity. Prática comum chamada de "aprendizado por transferência"(transfer learning).

As principais funcionalidades desenvolvidas foram:

- **Classificação de Sentimentos:** O sistema recebe um texto de avaliação como entrada e o classifica como "Positivo"ou "Negativo".
- **Interface de Demonstração:** Foi criada uma interface web simples e interativa com a biblioteca Gradio, permitindo que o usuário insira um texto e visualize a classificação do sentimento em tempo real.
- **Análise de Acurácia:** O modelo foi testado para validar sua eficácia na classificação de sentimentos.

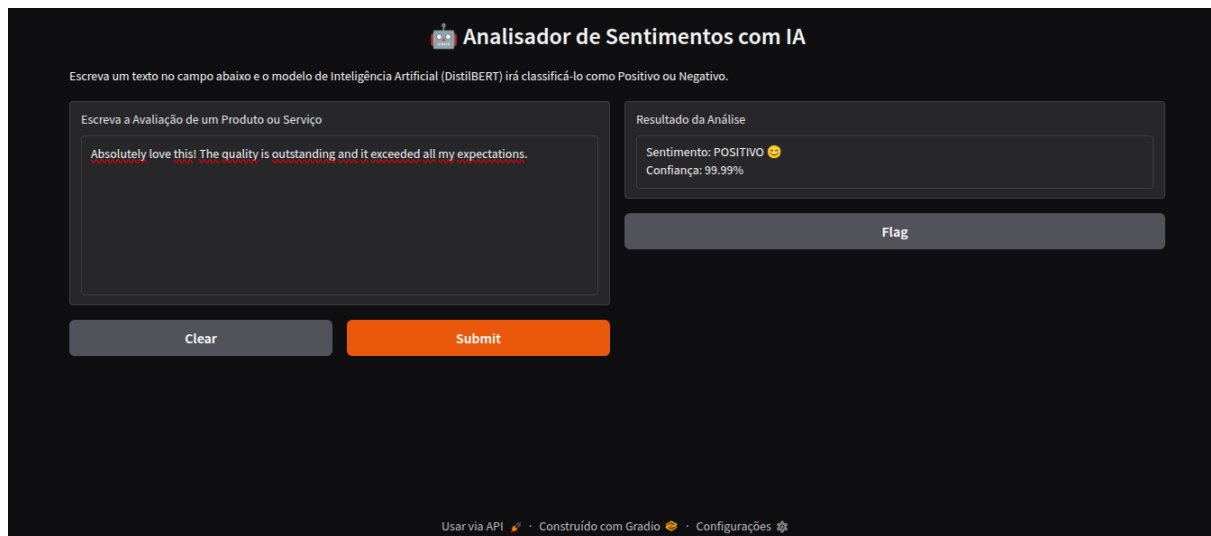
O processo de desenvolvimento envolveu o carregamento do modelo de análise de sentimentos, a implementação de uma função para analisar o texto inserido e a criação da interface com o usuário. O código em `interface.py` demonstra a integração do modelo com o Gradio, onde o texto é recebido, analisado pelo *pipeline* do **transformers**, e o resultado (positivo ou negativo, com a respectiva confiança) é exibido na interface.

3 Resultados

A seguir será apresentado resultados de diferentes tipos de comentários divididos em 3 categorias: positivos, negativos e neutros(com ambiguidade).

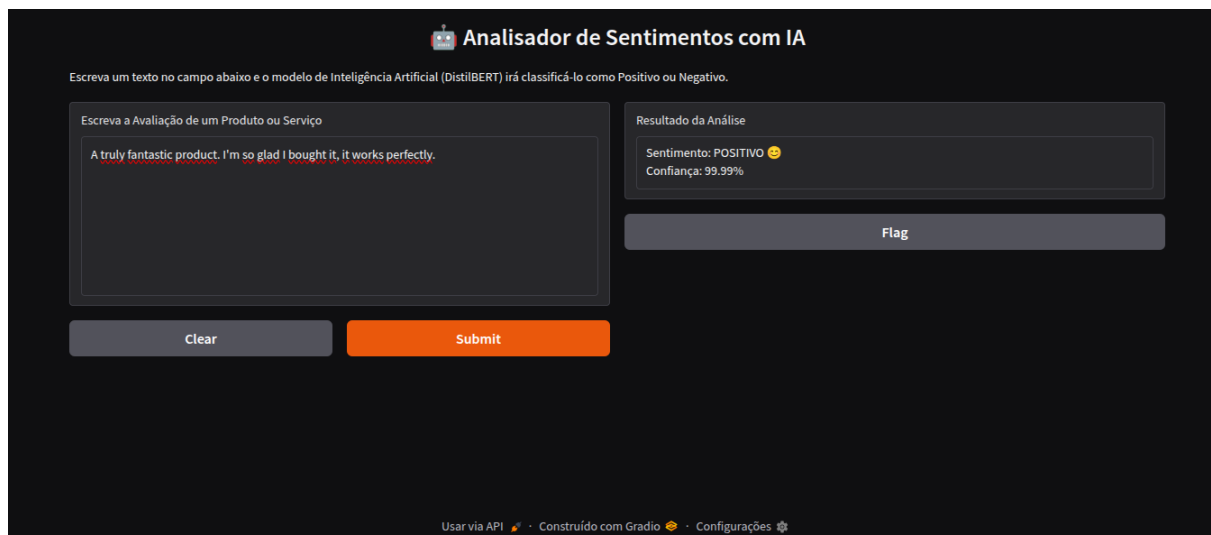
3.1 Testes com Comentários Positivos

Para textos com linguagem claramente positiva, o modelo demonstrou alta precisão e confiança, classificando-os corretamente como "POSITIVO".



The screenshot shows a web application titled "Analizador de Sentimentos com IA". It has a dark theme. At the top, it says "Escreva um texto no campo abaixo e o modelo de Inteligência Artificial (DistilBERT) irá classificá-lo como Positivo ou Negativo." Below this is a text input area with the placeholder "Escreva a Avaliação de um Produto ou Serviço". The input field contains the text "Absolutely love this! The quality is outstanding and it exceeded all my expectations." with some words highlighted in red. To the right of the input field is a box labeled "Resultado da Análise" which displays "Sentimento: POSITIVO 😊" and "Confiança: 99.99%". Below the input field are two buttons: "Clear" and "Submit". At the bottom right of the "Resultado da Análise" box is a "Flag" button. At the very bottom of the page, there is a footer with links: "Usar via API", "Construído com Gradio", and "Configurações".

Figura 1: Teste com avaliação positiva. Entrada: "Absolutely love this! The quality is outstanding and it exceeded all my expectations.". Resultado: POSITIVO com 99.99% de confiança.

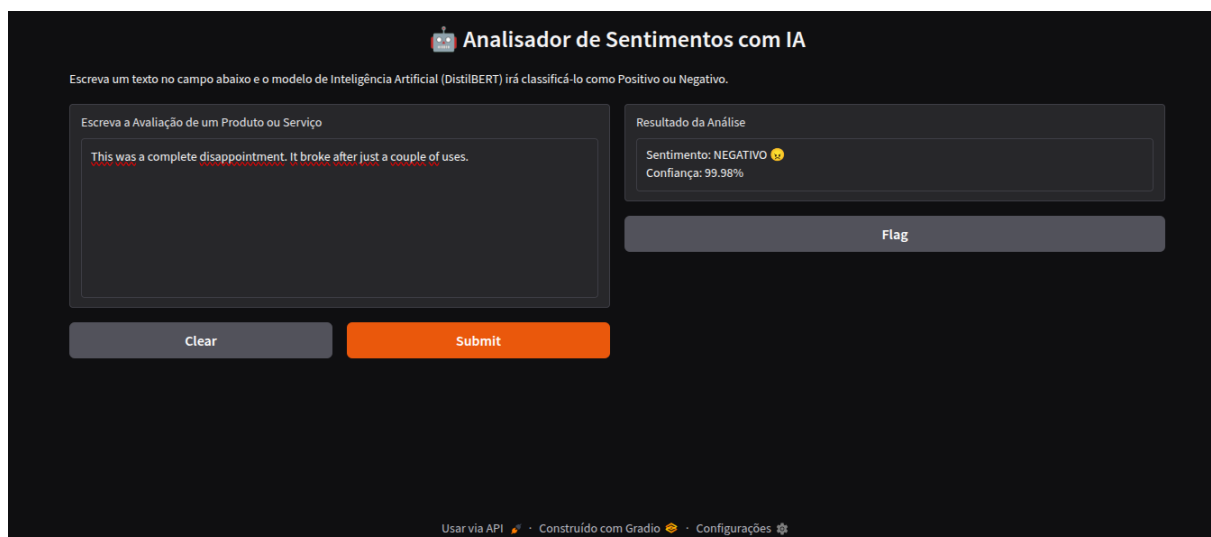


This screenshot is similar to the previous one, showing the same "Analizador de Sentimentos com IA" interface. The text input field now contains "A truly fantastic product. I'm so glad I bought it, it works perfectly." with some words highlighted in red. The "Resultado da Análise" box still shows "Sentimento: POSITIVO 😊" and "Confiança: 99.99%". The layout and buttons are identical to the previous figure.

Figura 2: Teste com avaliação positiva. Entrada: "A truly fantastic product. I'm so glad I bought it, it works perfectly.". Resultado: POSITIVO com 99.99% de confiança.

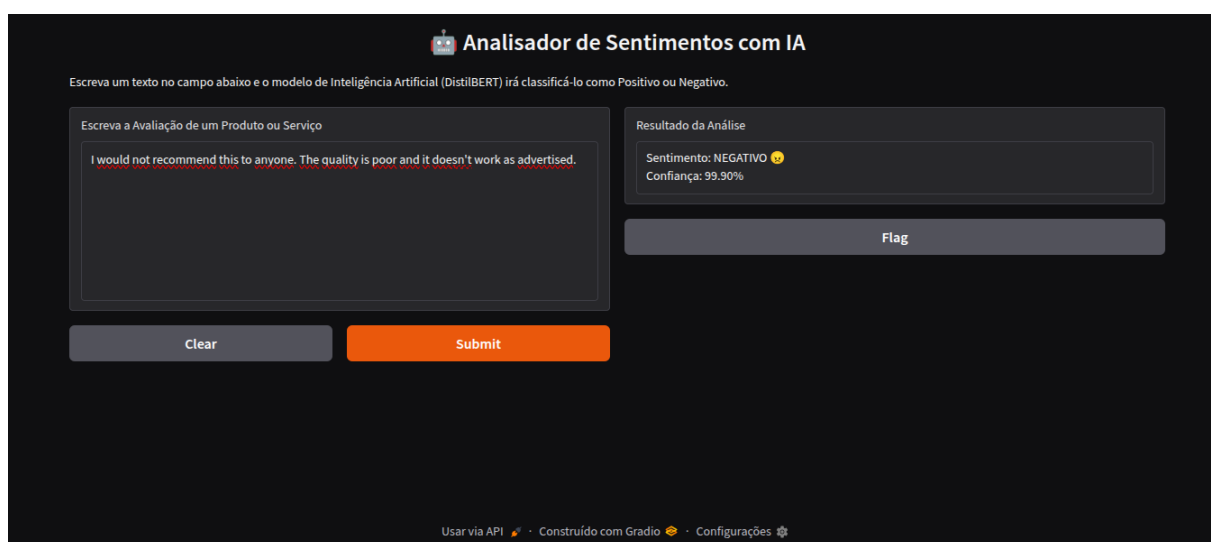
3.2 Testes com Comentários Negativos

De forma similar aos testes positivos, o modelo identificou com sucesso os comentários negativos, atribuindo a classificação "NEGATIVO" com alto grau de confiança.



The screenshot shows a web application titled "Analizador de Sentimentos com IA". It has a dark theme. At the top, it says "Escreva um texto no campo abaixo e o modelo de Inteligência Artificial (DistilBERT) irá classificá-lo como Positivo ou Negativo." Below this is a text input area with the placeholder "Escreva a Avaliação de um Produto ou Serviço". The input field contains the text: "This was a complete disappointment. It broke after just a couple of uses." Below the input field are two buttons: "Clear" and "Submit". To the right of the input field is a box labeled "Resultado da Análise" which displays "Sentimento: NEGATIVO 😞" and "Confiança: 99.98%". Below this box is a "Flag" button. At the bottom of the page, there is a footer with links: "Usar via API", "Construído com Gradio", and "Configurações".

Figura 3: Teste com avaliação negativa. Entrada: "This was a complete disappointment. It broke after just a couple of uses.". Resultado: NEGATIVO com 99.98% de confiança.

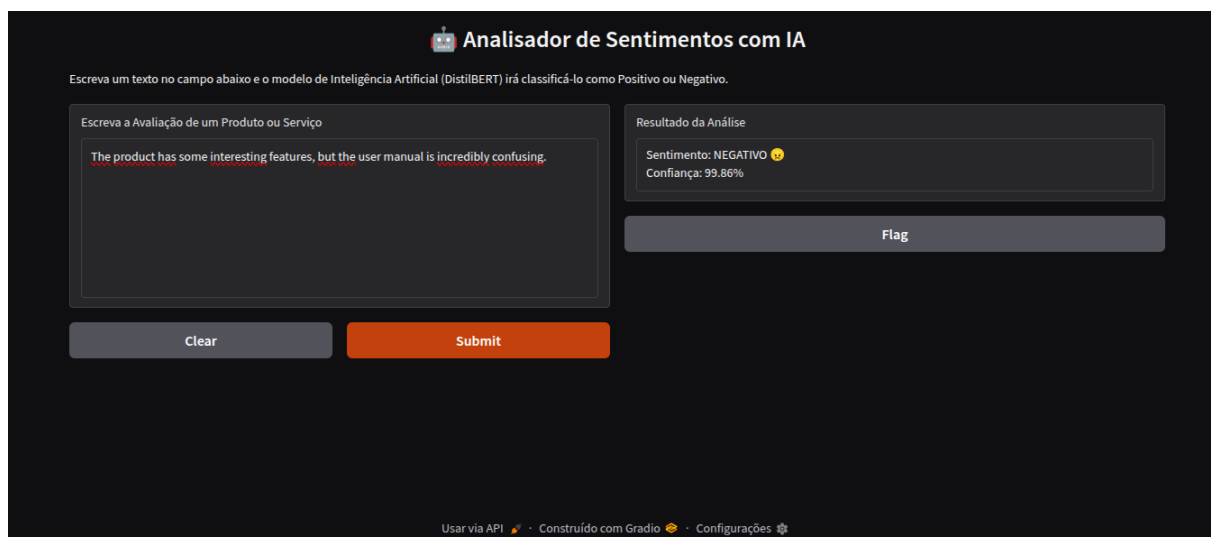


This screenshot shows the same web application as Figure 3, but with a different input. The text input field now contains: "I would not recommend this to anyone. The quality is poor and it doesn't work as advertised." The "Resultado da Análise" box displays "Sentimento: NEGATIVO 😞" and "Confiança: 99.90%". The "Flag" button is still present. The footer remains the same.

Figura 4: Teste com avaliação negativa. Entrada: "I would not recommend this to anyone. The quality is poor and it doesn't work as advertised.". Resultado: NEGATIVO com 99.90% de confiança.

3.3 Testes com Comentários Ambíguos, Neutros e Sarcásticos

Nesta categoria, observa-se a dificuldade do modelo em lidar com nuances da linguagem. Por ter sido treinado em um dataset binário (positivo/negativo), ele tende a forçar uma classificação, resultando em interpretações por vezes equivocadas.



Analizador de Sentimentos com IA

Escreva um texto no campo abaixo e o modelo de Inteligência Artificial (DistilBERT) irá classificá-lo como Positivo ou Negativo.

Escreva a Avaliação de um Produto ou Serviço

The product has some interesting features, but the user manual is incredibly confusing.

Resultado da Análise

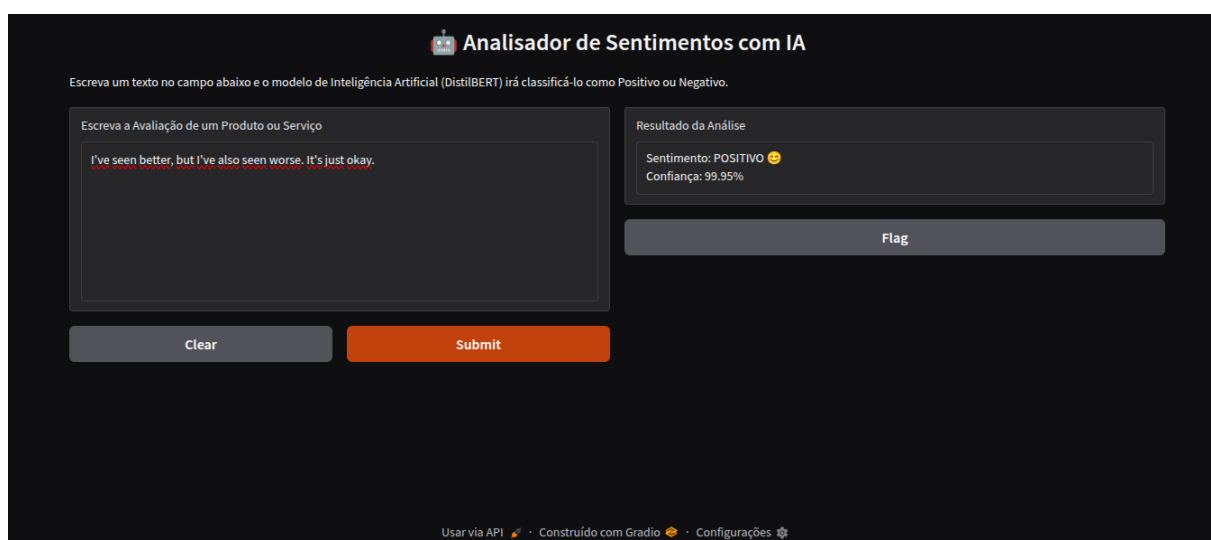
Sentimento: NEGATIVO 😞
Confiança: 99.86%

Flag

Clear Submit

Usar via API · Construído com Gradio · Configurações

Figura 5: Teste com sentimento misto. O modelo focou na parte negativa ("incredibly confusing"). Entrada: "The product has some interesting features, but the user manual is incredibly confusing.". Resultado: NEGATIVO com 99.86% de confiança.



Analizador de Sentimentos com IA

Escreva um texto no campo abaixo e o modelo de Inteligência Artificial (DistilBERT) irá classificá-lo como Positivo ou Negativo.

Escreva a Avaliação de um Produto ou Serviço

I've seen better, but I've also seen worse. It's just okay.

Resultado da Análise

Sentimento: POSITIVO 😊
Confiança: 99.95%

Flag

Clear Submit

Usar via API · Construído com Gradio · Configurações

Figura 6: Teste com comentário neutro. O modelo classificou como positivo, provavelmente devido a palavras como "better" e "okay". Entrada: "I've seen better, but I've also seen worse. It's just okay.". Resultado: POSITIVO com 99.95% de confiança.

The screenshot shows a web application titled "Analizador de Sentimentos com IA". It has a dark theme. At the top, it says "Escreva um texto no campo abaixo e o modelo de Inteligência Artificial (DistilBERT) irá classificá-lo como Positivo ou Negativo." Below this is a text input area with the placeholder "Escreva a Avaliação de um Produto ou Serviço". The input field contains the text "Well, that was certainly an experience." with "was" and "an" highlighted in red. To the right of the input field is a "Resultado da Análise" box showing "Sentimento: POSITIVO 😊" and "Confiança: 99.61%". Below the input field are "Clear" and "Submit" buttons. At the bottom right of the "Resultado da Análise" box is a "Flag" button. At the very bottom of the page, there is a footer with links: "Usar via API", "Construído com Gradio", and "Configurações".

Figura 7: Teste com sarcasmo. O modelo não detectou a ironia e classificou o comentário como positivo. Entrada: "Well, that was certainly an experience.". Resultado: POSITIVO com 99.61% de confiança.

The screenshot shows the same web application as Figure 7. The text input field now contains the text "The raw power of this machine is undeniable, though it lacks finesse." with "raw", "undeniable", and "lacks" highlighted in red. The "Resultado da Análise" box now shows "Sentimento: NEGATIVO 😞" and "Confiança: 76.85%". The "Flag" button is still present. The footer remains the same.

Figura 8: Teste com avaliação ambígua. A presença da palavra "lacks"(falta) levou o modelo a uma classificação negativa, apesar da confiança mais baixa (76.65%). Entrada: "The raw power of this machine is undeniable, though it lacks finesse.". Resultado: NEGATIVO com 76.65% de confiança.

3.4 Análise das Limitações em Comentários Ambíguos

3.4.1 O Modelo Foi Treinado para uma Tarefa Binária

O coração do problema está no fato de que o modelo `distilbert-base-uncased-finetuned-sst-2` foi especificamente treinado para uma tarefa de classificação binária. Como descrito, o objetivo é classificar o sentimento como "Positivo" ou "Negativo".

- **Conjunto de Dados:** O projeto utiliza o dataset `amazon_polarity`, que, como o nome sugere, contém avaliações rotuladas apenas com polaridade: 1 para positivo e 0 para negativo.
- **Ausência de Categoria "Neutra":** O modelo nunca "viu" exemplos de comentários neutros ou ambíguos durante seu treinamento. Ele aprendeu a reconhecer padrões de palavras e frases que estão fortemente associados a opiniões positivas ou negativas.

Quando o modelo se depara com um texto que não se encaixa claramente em nenhuma das duas caixas, ele é forçado a tomar uma decisão. O resultado é, muitas vezes, uma classificação com um baixo índice de confiança (score), pois o modelo está "em dúvida" sobre qual categoria é a mais apropriada.

3.4.2 Dificuldade em Compreender Nuances Humanas

Comentários ambíguos, sarcásticos ou com sentimentos mistos são extremamente complexos e dependem de um entendimento profundo do contexto, algo que ainda é um grande desafio para a IA.

- **Sarcasmo:** Um comentário como "Well, that was certainly an experience" pode ser positivo ou extremamente negativo dependendo do contexto. O modelo, que analisa principalmente as palavras usadas, não consegue captar a ironia.
- **Sentimentos Mistos:** Uma avaliação como "The product has some interesting features, but the user manual is incredibly confusing" contém elementos positivos e negativos. O modelo pode acabar se concentrando no trecho com a linguagem mais forte, classificando o todo de forma imprecisa.
- **Falta de Contexto Geral:** Um modelo de análise de sentimentos não "sabe" sobre o mundo. Ele não entende o que é um "manual de usuário" ou a frustração de não conseguir usar um produto. Ele apenas reconhece que "interessante" é geralmente positivo e "confuso" é geralmente negativo.

3.4.3 O Objetivo é a Polaridade, Não a Compreensão

É crucial entender que o objetivo do modelo não é compreender o texto da mesma forma que um humano, mas sim identificar a polaridade do sentimento expresso. O sistema foi projetado para responder à pergunta: "A opinião geral aqui é mais positiva ou mais negativa?". Ele não foi projetado para responder "O que essa pessoa está realmente querendo dizer?".

Em resumo, a aplicação tem dificuldades com comentários ambíguos porque o modelo subjacente foi treinado em um universo de dados estritamente dividido entre "positivo" e "negativo", e ele não possui a capacidade de interpretar as complexas nuances da linguagem humana que caracterizam a ambiguidade e a neutralidade.

4 Conclusão

O projeto alcançou com êxito seu objetivo de desenvolver um sistema funcional para análise de sentimentos. A utilização do modelo DistilBERT provou ser uma abordagem eficaz para a tarefa de PLN proposta. A implementação da interface com o Gradio tornou a solução acessível e fácil de usar, cumprindo todos os requisitos estabelecidos. A aplicação do modelo sobre o dataset amazon polarity demonstrou a capacidade notável da transferência de conhecimento, onde o modelo generalizou seu aprendizado para o domínio específico de avaliações de produtos, mesmo sem treinamento direto neste conjunto de dados. Os testes confirmam a capacidade do modelo em distinguir entre sentimentos positivos e negativos, mas também expõem suas limitações diante de textos com sarcasmo, neutralidade ou sentimentos mistos, o que abre portas para trabalhos futuros focados em modelos mais robustos para a detecção de comentários ambíguos.

5 Referências

- Vaswani, A., et al. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, 30. Disponível em: <https://arxiv.org/abs/1706.03762>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1. Disponível em: <https://arxiv.org/abs/1810.04805>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *5th Workshop on Energy Efficient Machine Learning and Cognitive Computing - NeurIPS 2019*. Disponível em: <https://arxiv.org/abs/1910.01108>
- Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level Convolutional Networks for Text Classification. *Advances in Neural Information Processing Systems*, 28. Disponível em: <https://arxiv.org/abs/1509.01626>
- Wolf, T., et al. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Disponível em: <https://www.aclweb.org/anthology/2020.emnlp-demos.6>
- Abid, A., Abdalla, A., Das, A. (2019). Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. *ICML 2019 Workshop on Reproducibility in Machine Learning*. Disponível em: <https://arxiv.org/abs/1906.02569>
-

Hugging Face. (2024). *Hugging Face Website*. Disponível em: [*https://huggingface.co/*](https://huggingface.co/)
