Análise de Sentimento
Reviews One Piece Live Action (IMDb)

Amadeo Neto - atcn
Arthur Henrique - ahac
Bruno Marques - bhsm
Caue Addae - casg

Centro de
Informática
UFPE

UNIVERSIDADE
FEDERAL
DE PERNAMBUCO

Objetivo

Explicar as predições de um modelo de análise de sentimento textual.

Metodologia

1. Revisão bibliográfica para descobrir as técnicas do estado da arte da área.
   - ➤ LIME [1] e SHAP [2] são técnicas de explicabilidade comuns para esse objetivo [3] [4]
2. Modelo a ser usado: **DistilBERT** [5] pré-treinado do Hugging Faces
3. Dataset de fine-tuning: **IMDB Reviews 50K** [6]
4. Técnica de avaliação para explicabilidade: **SHAP** [2] com dataset **One Piece Live Action IMDb** [7]

[1] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "" Why should i trust you?" Explaining the predictions of any classifier." Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016.

[2] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017).

[3] Kumar, Akshi, Shubham Dikshit, and Victor Hugo C. Albuquerque. "Explainable artificial intelligence for sarcasm detection in dialogues." *Wireless Communications and Mobile Computing* 2021.1 (2021): 2939334.

[4] Cirqueira, Douglas, et al. "Explainable sentiment analysis application for social media crisis management in retail." (2020).

[5] Sanh, Victor, et al. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." *arXiv preprint arXiv:1910.01108* (2019).

[6] Maas, Andrew, et al. "Learning word vectors for sentiment analysis." *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*. 2011.

[7] https://www.kaggle.com/datasets/itsnobita/one-piece-live-action-imdb-reviews. Accessed 13/07/2024

cin.ufpe.br

DistilBERT

O DistilBERT é um modelo de NLP baseado na arquitetura Transformers.

O modelo foi pré-treinado nas bases da Wikipédia e Book Corpus e encontra-se disponível no Hugging Face.

Ele funciona como um modelo base no qual um fine-tuning pode ser realizado para realizar uma tarefa específica.
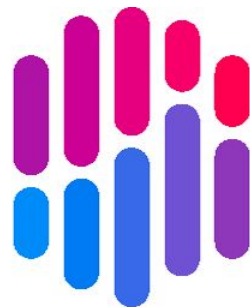
SHAP

O SHAP é uma técnica de explicação de modelos similar ao Lime, mas que usa os "Valores de Shapley" da Teoria dos Jogos. A técnica permite tanto explicação de texto quanto de dados categóricos e imagens.

O SHAP permite tanto explicações locais quanto globais.

Uma biblioteca está disponível para realizar os cálculos dos valores de Shapley e também realizar plots.

IMDb Reviews 50K

Dataset para classificação binária de sentimentos. Constitui-se de 50.000 avaliações de filmes classificadas em positivas ou negativas, contendo não mais que 30 análises por filme, e uma distribuição balanceada de forma que uma predição cega teria 50% de chance de acerto.

- 25.000 reviews para treinamento
- 25.000 reviews para teste

cin.ufpe.br

One Piece Live Action (OPLA) IMDb Reviews

Dataset incluindo análises e notas para a série live action de One Piece da Netflix. Foi utilizado em conjunto ao **SHAP** para realizar a avaliação de explicabilidade do modelo **DistilBERT**.

Como o dataset possui avaliações que variam de 0 a 10, utilizamos a técnica de definição de labels do artigo do dataset **IMDb Reviews 50K** que consiste em:

- nota ≤ 4 → análise negativa
- nota ≥ 7 → análise positiva

Estrutura do Dataset

Antes do tratamento

Title	Review	Date	Rating
String	String	Date	Integer

Estrutura do Dataset

Após o tratamento

Review	Word Count	Rating
String	Integer	String

Regras aplicadas:
- A review **negativa** tem o score ≤ 4
- A review **positiva** tem o score ≥ 7
- Reviews com score > 4 e < 7 foram descartadas
- Remoção da coluna **Title**
- Remoção da coluna **Date**
- **Remoção** das linhas com **Word count** maior que 300 (de 765 linhas para 680)
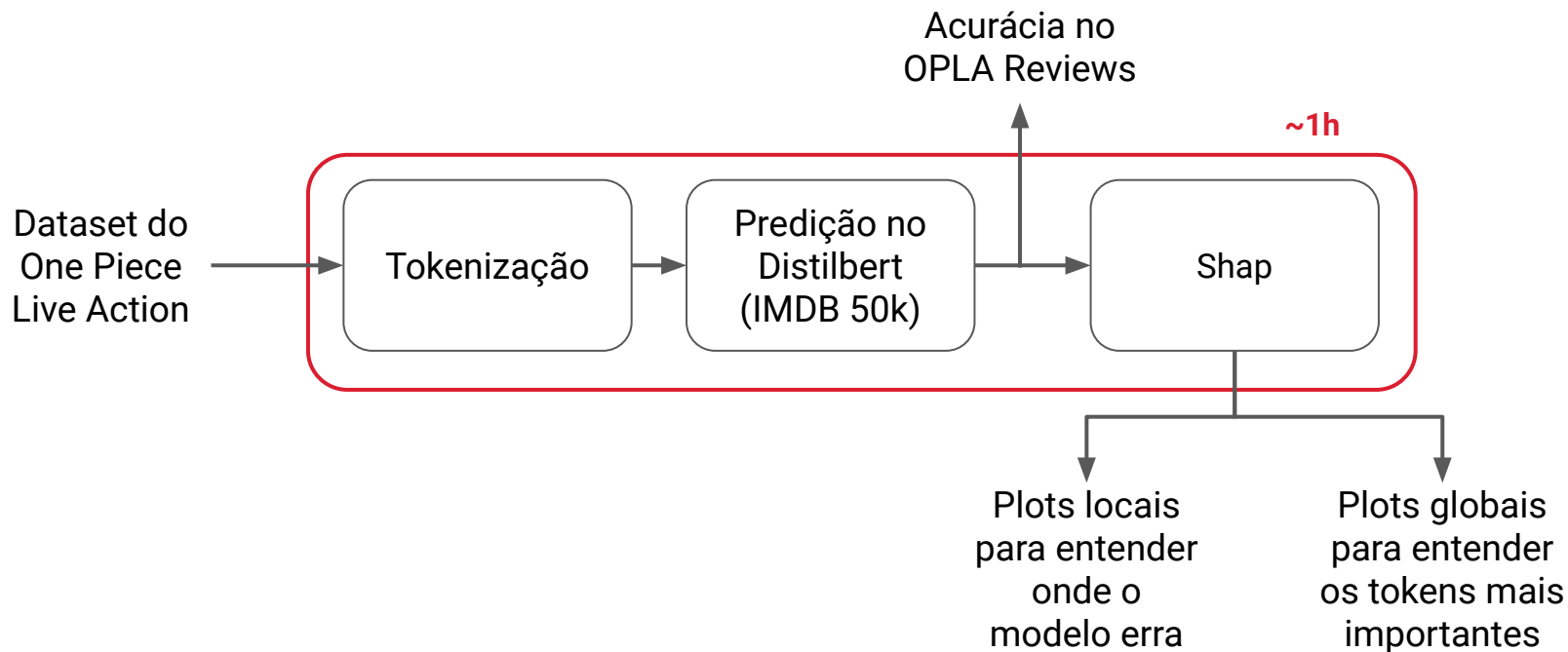
Pipeline do Treino

~30 min

Distilbert pré-treinado → Fine-Tuning (1 época) → Cálculo das métricas → Acurácia e F1-Score

IMDB 50k Reviews → Tokenização → Fine-Tuning (1 época)

cin.ufpe.br
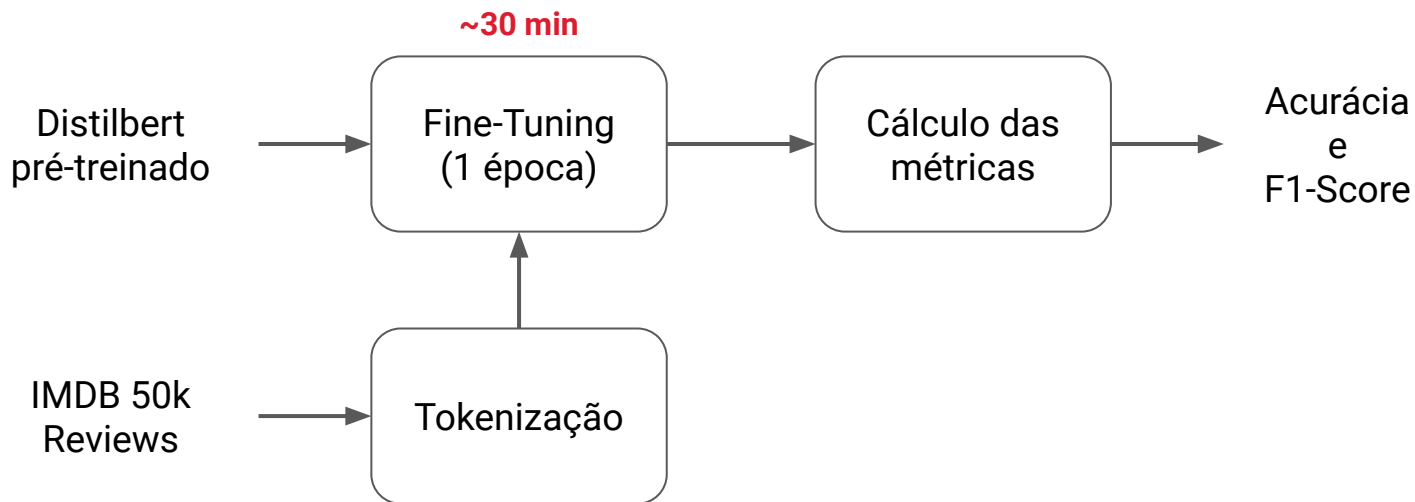
Resultados - DistilBERT no IMDB

Calculamos as métricas de acurácia e F1-Score do modelo no dataset

Métrica	Valor
Acurácia	0.925800
F1 - Score	0.926163

Pipeline do Teste

Resultados - DistilBERT no OPLA Reviews

Calculamos as métricas de acurácia geral e por classe

Métrica	Valor
Acurácia	0.9588
Acurácia POS	0.9642
Acurácia NEG	0.9062

Como o modelo performou bem nessa base, não foi necessário fazer um novo fine-tuning

Onde o modelo acerta?

Review Positiva

outputs
NEGATIVE **POSITIVE**

base value $f_{POSITIVE}$(inputs)
0.586258   **0.958039**

0.2   0.4   0.586258   0.8   1.2

la   t   ov   a   pod   njoy   is   This

0.06
inputs
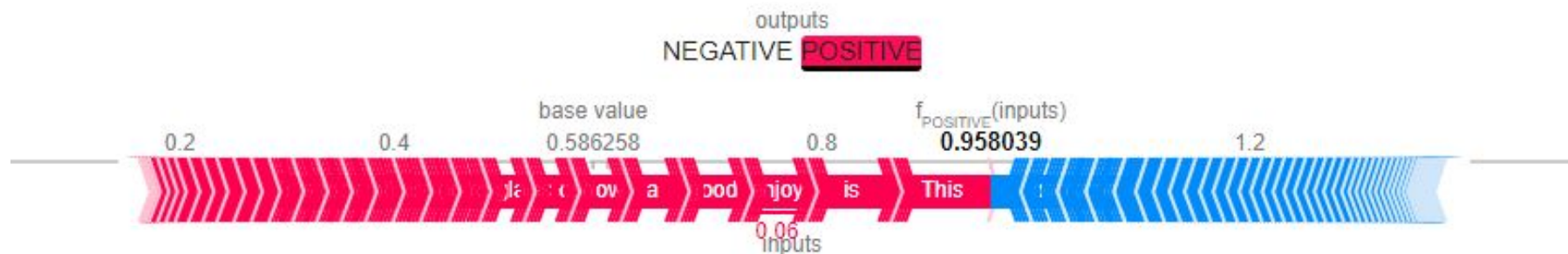
I have to admit that I've never watched any of the animated show One Piece or even heard about it for that matter. I'm really not a big fan of anime. Then after watching the trailers for this I thought it looked like a show for children. I was wrong. I kept reading about how good this show is and how it's #1 on Netflix so I thought I'd give it a chance. I'm glad I did because I couldn't have been more wrong. This is a show where people of all ages can enjoy. If you can just turn off your brain and watch I promise you'll have a good time. If you're looking for a new show to binge then give One Piece a try. You won't be disappointed!
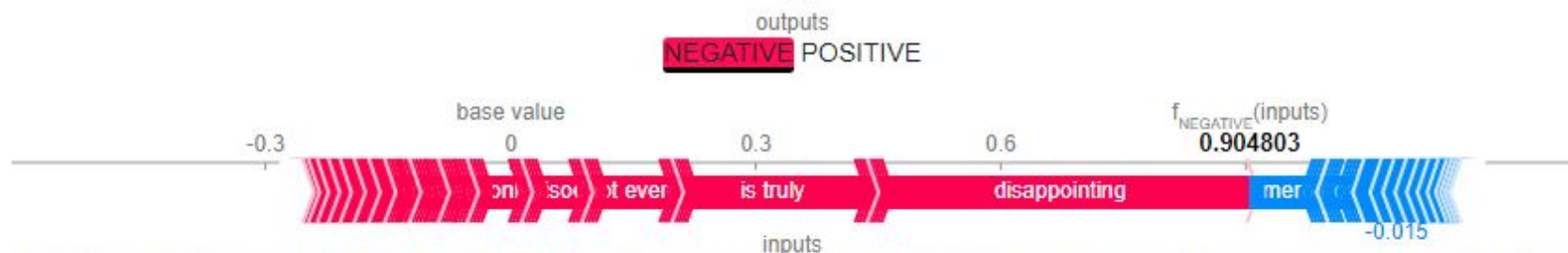
Onde o modelo acerta?

Review Negativa

base value

| 0 | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |

$f_{NEGATIVE}(inputs)$
0.996618

0.013
inputs
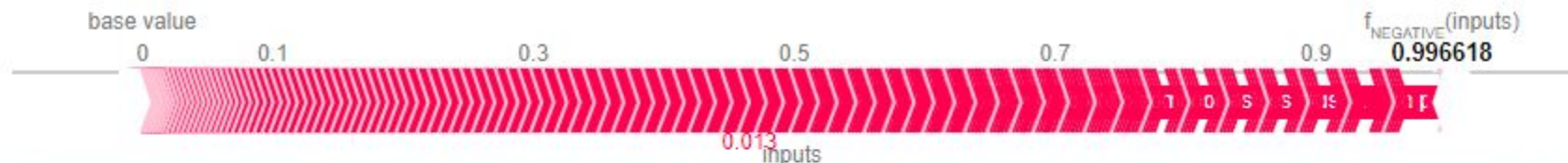
I wanted this to be more like Kung Fu Hustle or Hook, but it really didn't feel like the director had a clear vision for the final product. It feels more like a broadway play, with poor direction, cheesy dialogue and over-acting. Like many other manga to live action attempts, it really misses the mark and is riddled with plot holes as they desperately try to squeeze five episodes into one and take zero chances at veering away from the anime in fear of upsetting the fans. The cheesy music and dialogue mixed with violence was confusing, I couldn't tell if this was meant for adults or kids? If something is going to to be sooo unbearably tacky, they should have picked a lane and really ramped up the violence or put more effort into the script. Really stupid and definitely DOES NOT BREAK THE CURSE of all anime live actions being terrible. This one is absolute trash and not the exception.
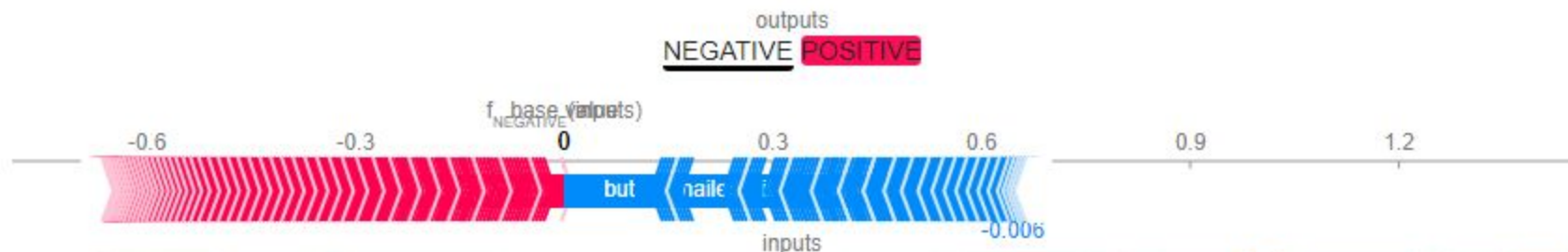
cin.ufpe.br

Onde o modelo erra?

Review Positiva classificada como Negativa

outputs
NEGATIVE POSITIVE

base value
-0.3   0   0.3   0.6 $f_{NEGATIVE}$(inputs)
0.904803

oni iso ot ever | is truly | disappointing | mer
inputs
-0.015

Having watched a good portion of One Piece anime this show is exactly what you'd hope for! The outrageousness of the anime is truly on display in this live action and you wouldn't want it any other way. My sold issue with this show and it's not even the shows fault but the limited amount of episodes for season 1 is truly disappointing. So many of this new seasons are going to 8 episodes or so which is the companies fault. There are numerous shows that would have benefited so much if they could've had a few more episodes and for a anime show like this it DESERVES to have had more than just 8 to get started. Netflix scored on a perfection live action but burned us on limiting season 1 to only 8 episodes.
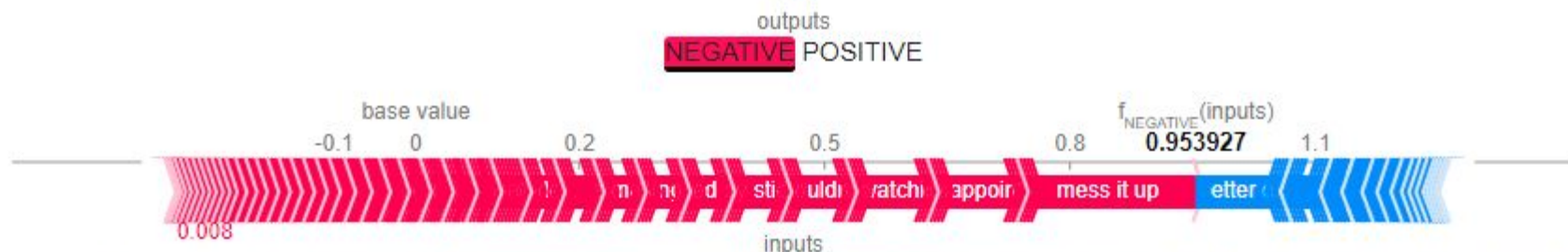
Onde o modelo erra?

Review Negativa classificada como Positiva

outputs
NEGATIVE POSITIVE

f base value (inputs)
NEGATIVE

-0.6   -0.3   0   0.3   0.6   0.9   1.2

but   nailed

-0.006

inputs

There's so many reviews all the same 10/10 that go along the lines of I thought it was going to disappoint but I surprised by how good it is Live action adaption they nailed it The reality is. I'd say this show is the usual Netflix standard, like 5 out of 10. Max. Has all the usual issues of a Netflix show, actors don't seem to be cast based on talent, no meaningful story arcs, not many characters that you would like. Childish dialogue- nobody talks like these people I read a review that said they spent $17,000,000 an episode, I reckon Adam Goldberg could have made a more involving story for $17 an episode.
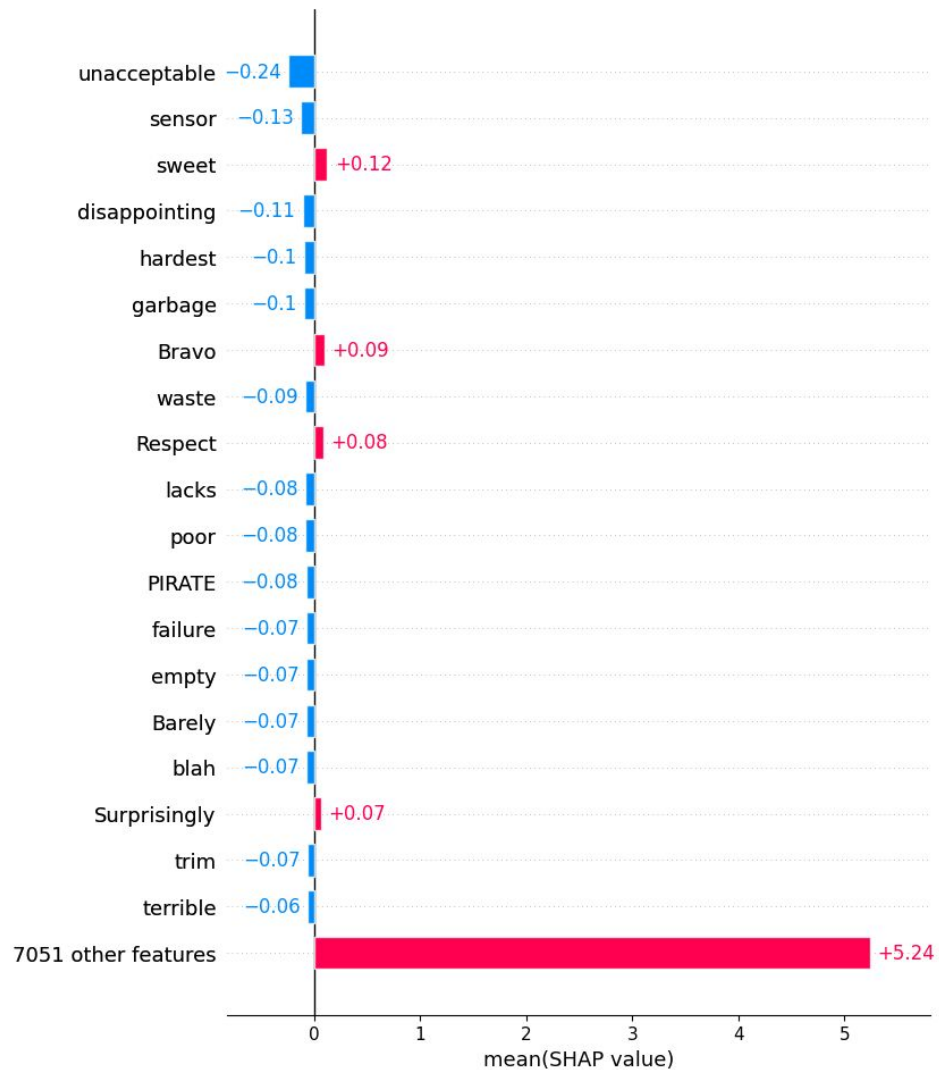
Onde o modelo erra?

Review Positiva classificada como Negativa

outputs
NEGATIVE POSITIVE

base value
-0.1   0   0.2   0.5   0.8 $f_{NEGATIVE}$(inputs)
0.953927   1.1

0.008

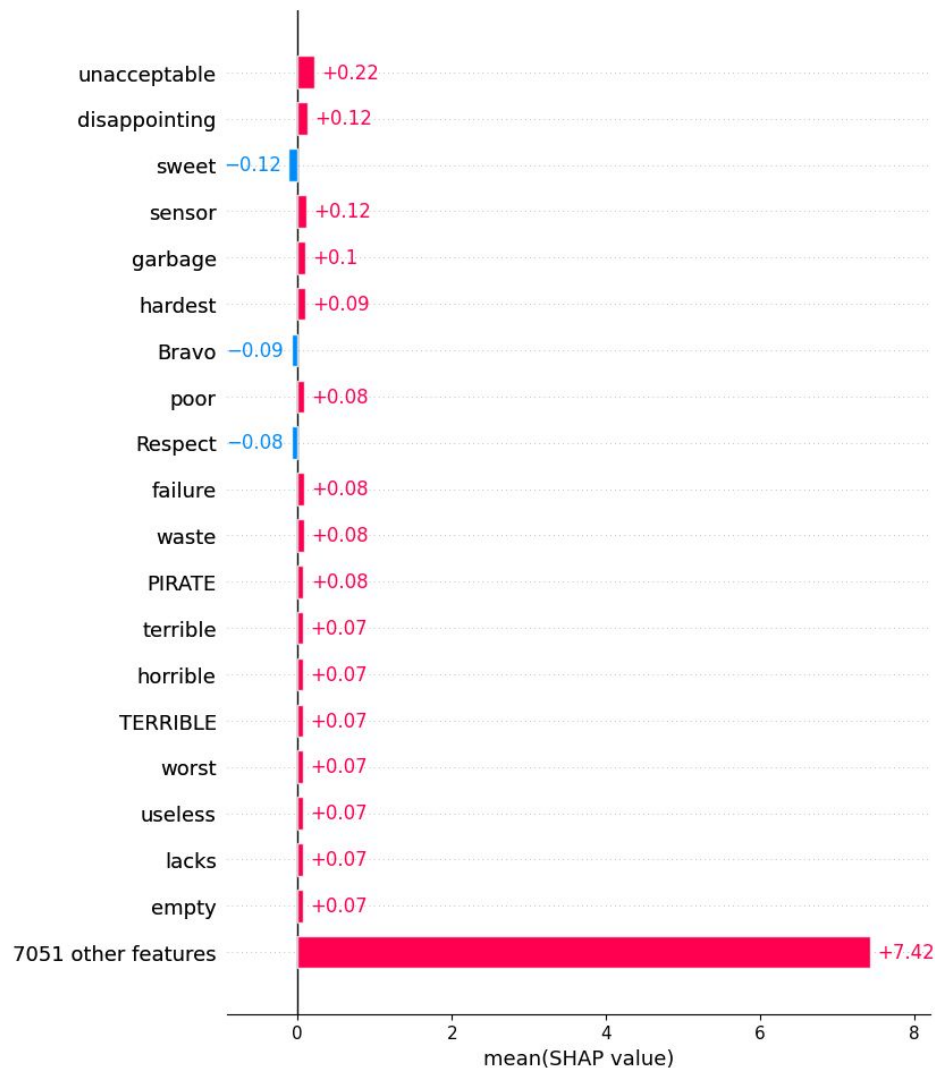n n d sti uld ratch appoir mess it up etter

inputs

I wasn't expecting this.. i mean i had only watched 125 something episodes of one piece and even after watching the same story, it really felt like One Piece. The crew had their personalities done really well. It was like watching a live action anime. Couldn't stop watching it. Binged watched it all in one go. Sooo looking forward to the next season and hopefully it will be an even better experience. Ot has so much potential. If they stick to the source material and keep the cgi the same. It will be remembered far better than witcher (let's not mention that disappointment). It will be interesting to see how Netflix does it and hopefully they won't mess it up.

cin.ufpe.br

Contribuição Global dos Tokens Mais Importantes

Reviews Positivas

Token	mean(SHAP value)
unacceptable	−0.24
sensor	−0.13
sweet	+0.12
disappointing	−0.11
hardest	−0.1
garbage	−0.1
Bravo	+0.09
waste	−0.09
Respect	+0.08
lacks	−0.08
poor	−0.08
PIRATE	−0.08
failure	−0.07
empty	−0.07
Barely	−0.07
blah	−0.07
Surprisingly	+0.07
trim	−0.07
terrible	−0.06
7051 other features	+5.24

Contribuição Global dos Tokens Mais Importantes

Reviews Negativas

[Link do Collab](#)

Obrigado!