

# Comparação de Modelos de Regressão Logística e Modelos de Linguagem de Grande Escala na Classificação de Sentimentos em Texto

Odemir Depieri Jr

**Resumo** – Este estudo compara a eficácia de um modelo de Regressão Logística e de um Modelo de Linguagem de Grande Escala (LLM) na classificação de sentimentos em textos. Utilizamos um conjunto de dados composto por textos rotulados em quatro classes de sentimento: neutro, positivo, imparcial e negativo. Inicialmente, um modelo de Regressão Logística foi treinado e testado, atingindo uma acurácia de 88%. Em seguida, uma LLM foi empregada para a mesma tarefa, resultando em uma acurácia significativamente inferior de 52%. O objetivo deste estudo é analisar as diferenças de desempenho entre as abordagens tradicionais e as de última geração, destacando suas vantagens e limitações.

## Introdução

A classificação de sentimentos em textos é uma tarefa fundamental na análise de dados textuais, com aplicações que vão desde a análise de redes sociais até o atendimento ao cliente. Tradicionalmente, modelos de Regressão Logística têm sido amplamente utilizados devido à sua simplicidade e eficácia. Recentemente, Modelos de Linguagem de Grande Escala (LLMs), como os desenvolvidos com arquiteturas de deep learning, têm demonstrado capacidades avançadas em várias tarefas de processamento de linguagem natural. No entanto, a eficácia dessas novas abordagens em comparação com métodos tradicionais na classificação de sentimentos ainda precisa ser explorada de forma mais detalhada. Este estudo investiga a performance de um modelo de

Regressão Logística e de uma LLM na classificação de sentimentos em textos, utilizando um conjunto de dados rotulado com quatro classes de sentimento: neutro, positivo, imparcial e negativo.

## Dados de estudo

Os dados selecionados para este estudo foram extraídos da Kaggle, a maior plataforma de competição de ciência de dados. Os dados foram coletados do Twitter e classificados como análise de sentimento na plataforma. Dada uma mensagem e uma entidade, a tarefa é julgar o sentimento da mensagem sobre a entidade em:

- a) Positivo
- b) Negativo
- c) Neutro
- d) Irrelevante

## Análise das Features

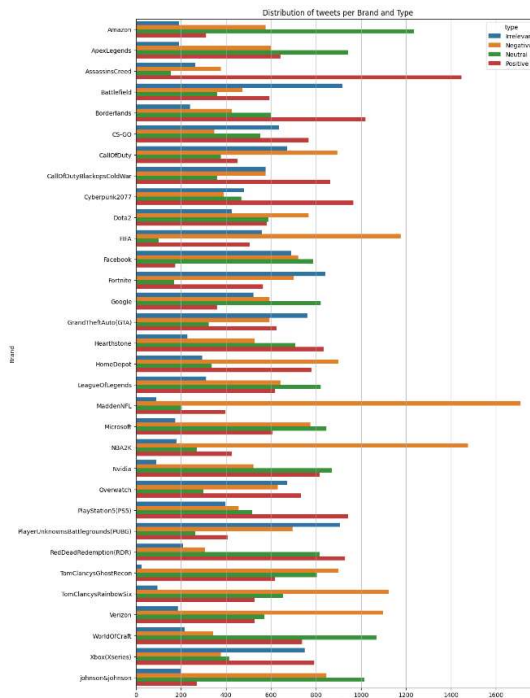
O *Dataset* possui 2 conjuntos de dados:

1. Dados de Treinamento: 74.642 registros
2. Dados de Validação: 1.000 registros

id	information	type	text
3364	Facebook	Irrelevant	I mentioned on Facebook that I was struggling ...
352	Amazon	Neutral	BBC News - Amazon boss Jeff Bezos rejects clai...
8312	Microsoft	Negative	@Microsoft Why do I pay for WORD when it funct...
4371	CS-GO	Negative	CSGO matchmaking is so full of closet hacking...
4433	Google	Neutral	Now the President is slapping Americans in the...
...	...	...	...
4891	GrandTheftAuto(GTA)	Irrelevant	★ Toronto is the arts and culture capital of ...
4359	CS-GO	Irrelevant	THIS IS ACTUALLY A GOOD MOVE TOT BRING MORE VI...
2652	Borderlands	Positive	Today sucked so it's time to drink wine n play...
8069	Microsoft	Positive	Bought a fraction of Microsoft today. Small wins.
6960	johnson&johnson	Neutral	Johnson & Johnson to stop selling talc baby po...

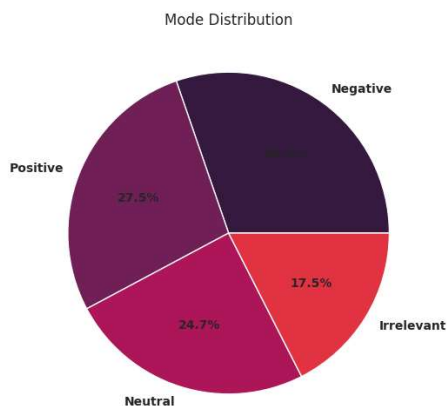
*Amostra dos dados de validação.*

## Visualização: Agrupamento por Informação e Tipo do Sentimento de Classificação



Agrupamento por informação e tipo do sentimento.

Havia uma distribuição na classe previsora para o treinamento do modelo de Regressão Logística.



Distribuição % da classe previsora.

## Treinamento da Regressão Logística

O modelo de Machine Learning “tradicional” escolhido foi o de Regressão Logística, usado no treinamento com os dados de treino. A validação foi feita com os dados de validação, obtendo uma acurácia de 88%.

```
validacao_amostra = model.predict(x_amostra)
print("Accuracy: ", accuracy_score(y_amostra, validacao_amostra) * 100)
Accuracy: 88.0
```

Print do código para apurar a acurácia do modelo de regressão logística.

## Treinamento da GEN AI

Para o modelo de GEN AI, utilizamos a arquitetura Gemini, empregando a API da arquitetura para classificar os comentários na base de validação. Utilizamos uma técnica de engenharia de prompt chamado “encadeamento de prompt”.

## Encadeamento de Prompt

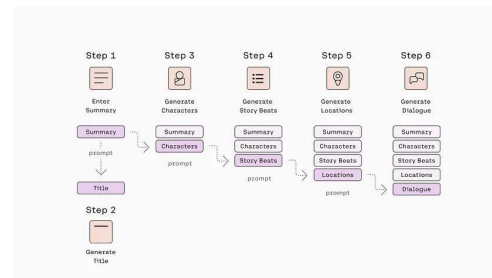


Imagem mostra o processo da técnica de encadeamento de prompt

O encadeamento de prompt é particularmente útil para tarefas que exigem um formato de resposta específico ou quando você precisa orientar o modelo passo a passo para garantir a precisão e a relevância da resposta.

A validação foi feita com os dados de validação. Executamos a API da LLM para cada comentário, pedindo que classifiesse a frase em: Negativo, Positivo, Neutro ou Irrelevante. A acurácia resultante foi de 52%.

```
print("Accuracy: ", accuracy_score(y_amostra, Acumulado['sentimento-IA']) * 100)
Accuracy: 52.0
```

Print do código para apurar a acurácia da LLM.

## Conclusão

Este estudo comparou a eficácia de um modelo de Regressão Logística e de um Modelo de Linguagem de Grande Escala (LLM) na tarefa de classificação de sentimentos em textos. Utilizando um conjunto de dados extraído do Twitter e rotulado em quatro classes de sentimento (neutro, positivo, imparcial e negativo), treinamos e testamos ambos os modelos. O modelo de Regressão Logística alcançou uma acurácia de 88%, enquanto a LLM apresentou uma acurácia significativamente inferior de 52%.

Os resultados sugerem que, apesar das capacidades avançadas das LLMs em várias tarefas de processamento de linguagem natural, sua aplicação direta na classificação de sentimentos pode não ser tão eficaz quanto os métodos tradicionais como a Regressão Logística, pelo menos nos contextos e com os dados utilizados neste estudo. A superioridade da Regressão Logística pode ser atribuída à sua simplicidade e à adequação dos dados de treinamento, que parecem ter favorecido este modelo.

No entanto, é importante considerar que as LLMs possuem potencial não totalmente explorado neste estudo, especialmente em cenários onde grandes volumes de dados e recursos de computação mais robustos estão disponíveis. A utilização de técnicas de fine-tuning e de melhor engenharia de prompts pode potencialmente melhorar o desempenho das LLMs.

Além disso, a integração de abordagens tradicionais e de última geração pode ser uma estratégia promissora. Combinações híbridas que aproveitem a robustez dos modelos tradicionais e as capacidades avançadas das LLMs podem resultar em sistemas de classificação de sentimentos mais precisos e eficientes.

Em resumo, a escolha do modelo para a classificação de sentimentos deve considerar o contexto específico, as características dos dados e os recursos disponíveis. Este estudo contribui para a compreensão das vantagens e limitações das abordagens tradicionais e de última geração na análise de sentimentos, fornecendo uma base para futuras pesquisas e aplicações práticas.

## Referências

- [1] Bird, Steven, Edward Loper, and Ewan Klein. "Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit." O'Reilly Media, Inc., 2009.
- [2] Kaggle. Disponível em: <https://www.kaggle.com/>
- [3] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, Carrie J Cai. "PromptChainer: Chaining Large Language Model Prompts through Visual Programming", 2022
- [4] Dietrich Trautmann. "Large Language Model Prompt Chaining for Long Legal Document Classification", 2023
- [5] Gemini Team Google. "Gemini: A Family of Highly Capable Multimodal Models", 2023