

# **Análise de Sentimento em comentários de E-commerce Usando Processamento de Linguagem Natural (NPL) e Aprendizado de Máquina (ML)**

# Integrantes



Clayton dos Santos Lira



Lorena Vaz Cord



Lucas Vaz de Castro Oliveira



Tiago Clemente Rodrigues

# Introdução

## Contexto

O e-commerce cresce globalmente, gerando vastos volumes de dados, incluindo avaliações e comentários. A análise de sentimentos é usada para automatizar a interpretação desses dados textuais.

## Relevância no Brasil

O Brasil é um dos maiores mercados de e-commerce da América Latina. Compreender as opiniões dos clientes é essencial para manter competitividade e melhorar a experiência do cliente.

## Aplicação da Análise de Sentimentos

Utilização do processamento de linguagem natural (NLP) e algoritmos de aprendizado de máquina, como Random Forest.

# Introdução

## Análise de Sentimentos

- Utiliza técnicas como NLP, linguística computacional e mineração de dados.
- Automatiza a análise de grandes volumes de dados textuais.
- Avalia opiniões dos clientes e agiliza respostas a feedbacks negativos.

## Vantagens

- Reduz a necessidade de pesquisas de satisfação.
- Diminui custos com recursos humanos.
- Acelera o processo de tomada de decisão.

# Problema e Objetivos

## Problema

Empresas precisam entender e classificar sentimentos em avaliações de clientes. As informações podem ser usadas para melhorar produtos e serviços.

## Objetivo Geral

Construir um modelo de aprendizado de máquina para análise de sentimentos em comentários de e-commerce.

## Objetivos Específicos

Pré-processamento, representação de texto, treinamento do modelo e avaliação do modelo.

# Metodologia – Aquisição de dados

Origem dos Dados	Fornecido pela Olist, disponível no Kaggle.
Licença	Creative Commons CC BY-NC-SA 4.0 (uso não comercial).
Detalhes do Dataset	Contém 9 tabelas CSV, com foco no arquivo: <i>olist_order_reviews_dataset.csv</i> .
Registros (iniciais)	99.224
Atributos	review_id, order_id, <b>review_score</b> , <b>review_comment_title</b> , <b>review_comment_message</b> , review_creation_date, review_answer_timestamp.



# Pré-processamento e análise exploratória

1

## **Combinação de Textos**

Combinação de review\_comment\_title com review\_comment\_message.  
Criação do nova atributo combined\_text.

2

## **Remoção de Valores ausentes**

Total de registros válidos: 40.977.

3

## **Criação da Variável-Alvo (sentimento)**

Notas 1 e 2 → Negativo.

Nota 3 → Removido.

Notas 4 e 5 → Positivo.

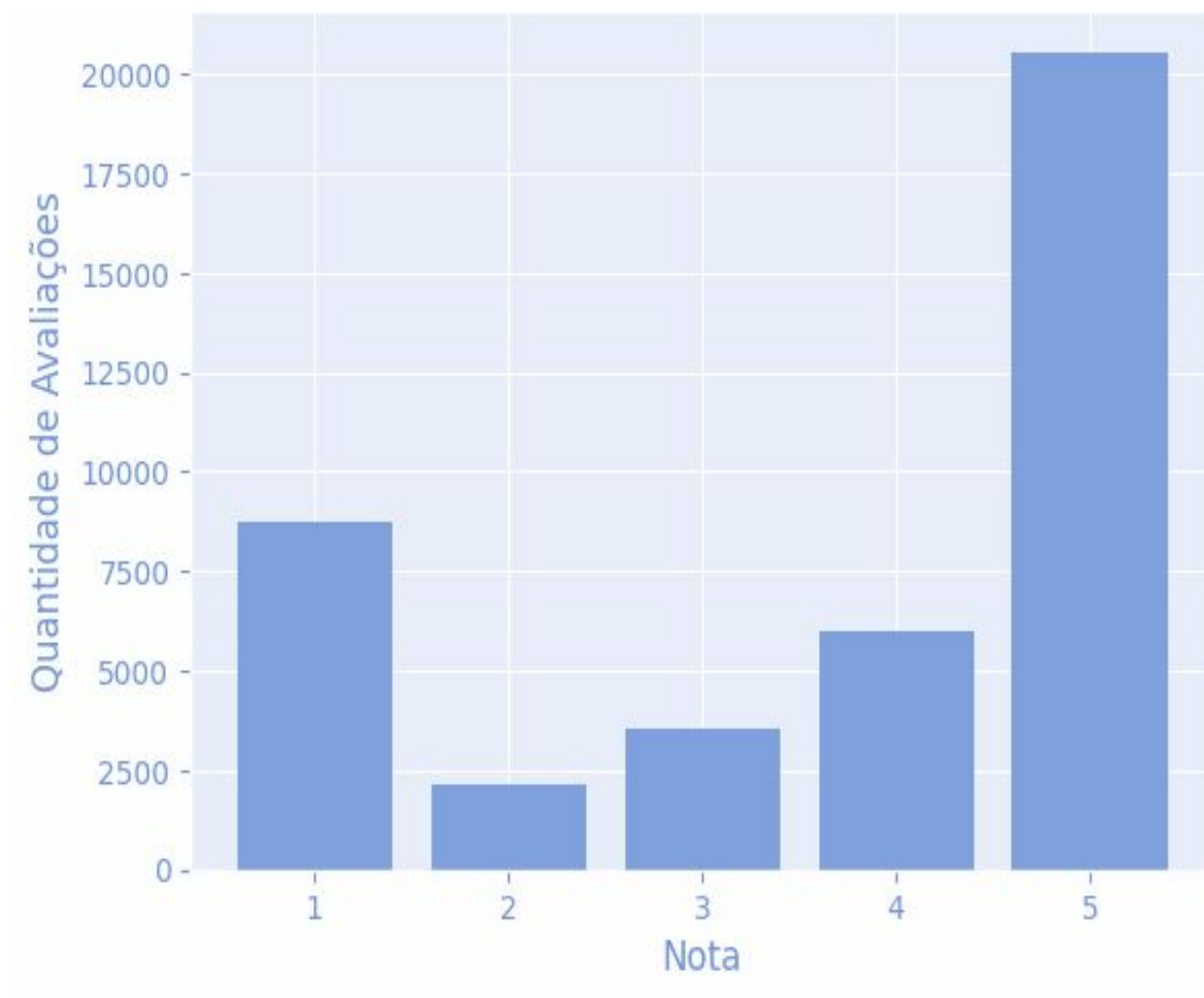
4

## **Resolução do Desequilíbrio de Classes**

Subamostragem para igualar positivos e negativos. Total após balanceamento: 21.780 registros (10.890 por classe).

# Metodologia - Pré-processamento

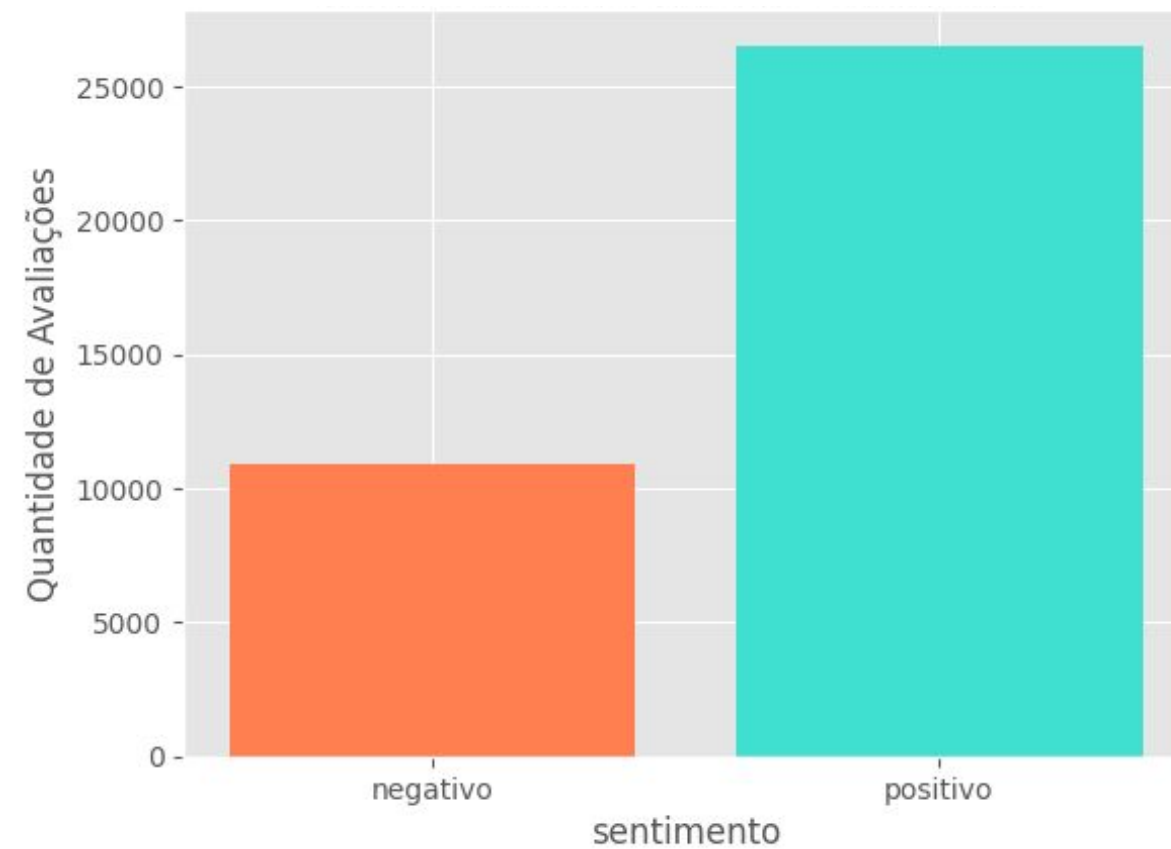
Figura 1 - Distribuição do atributo review\_score





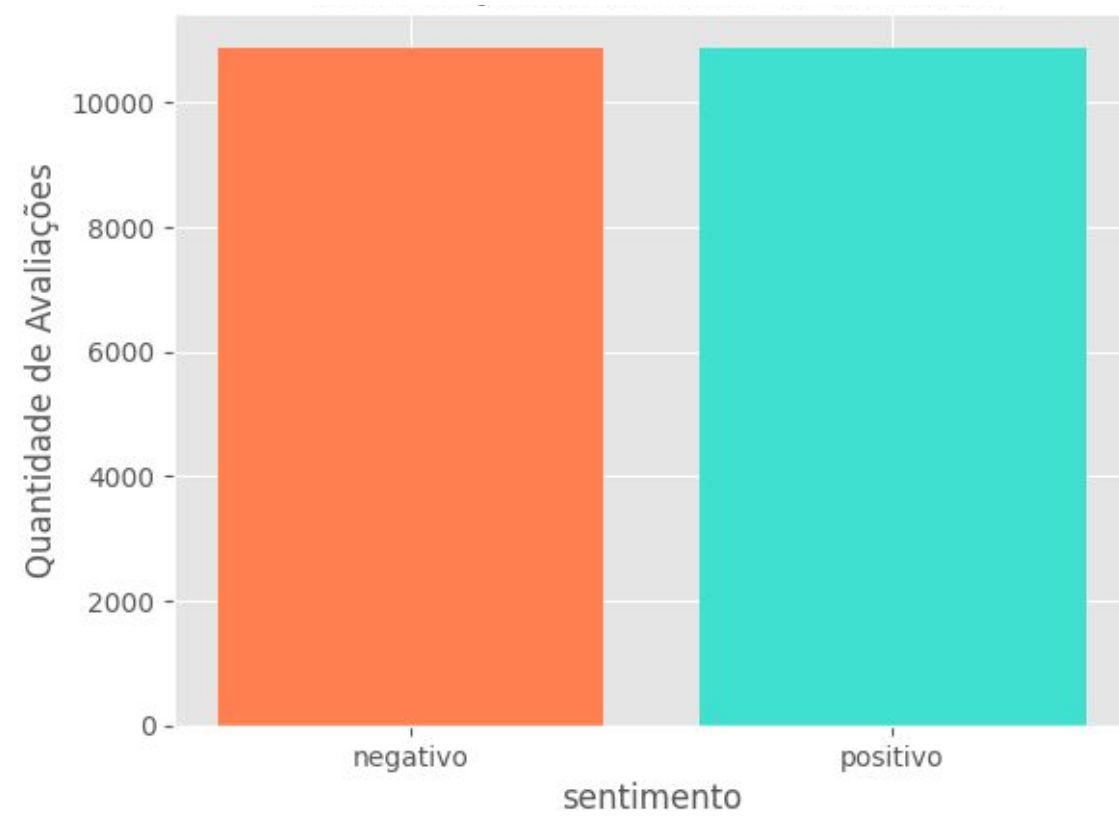
# Metodologia - Pré-processamento

**Figura 2 - Distribuição do atributo sentimento rotulado de acordo com as classes do review\_score**



# Metodologia - Pré-processamento

**Figura 3 - distribuição das classes positivo e negativo após subamostragem**



# Processamento de Texto - (NLP)



## Tokenização

Divisão do texto em palavras/tokens.



## Lematização

Reduzir palavras à sua forma base.



## Remoção de Stopwords

Eliminar palavras sem relevância (ex.: artigos).



## POS-Tagging

Atribuir categorias gramaticais a cada token.

# Representação vetorial - TF-IDF



## Term frequency - Inverse Document Frequency

Conversão do texto em vetores  
numéricos.

# Modelo e Treinamento

1

## Modelo Escolhido

Algoritmo Random Forest.

2

## Parâmetros Utilizados

`max_features='log2'` e `class_weight='balanced'`.

3

## Divisão dos Dados

Conjuntos: Treinamento e teste (usando `train_test_split`).

# Resultados

Acurácia Geral	92%
Classe Negativa	Precisão: 89% Recall: 95% F1-score: 0,92
Classe Positiva	Precisão: 95% Recall: 89% F1-score: 0,92

# Conclusão

## **Benefícios da Análise de Sentimentos**

Identificar áreas de insatisfação, respostas proativas, melhorar produtos e serviços.

## **Impacto**

Modelo com 92% de acurácia fornece insights confiáveis. Estratégia ajuda empresas a se destacarem no e-commerce brasileiro.



# Melhorias e perspectivas futuras

## Inclusão da classificação 3 - Neutro

A ausência do rótulo pode limitar a aplicabilidade do modelo. Manter as três classes representaria melhor a diversidade das opiniões

## Fine-tuning com GridsearchCV

Melhoria do desempenho do modelo através da seleção de hiperparâmetros mais adequados

# Referências

LIU, Yanqiu; YE, Fuming. Sentiment Analysis of Online Catering User Comments Based on Random Forest Feature Extraction. In: 2022 IEEE 10th International Conference on Information, Communication and Networks (ICICN). Zhangye, China: IEEE, 2022, p. 667–670.

SINGH, Shailendra Narayan; SARRAF, Twinkle. Sentiment Analysis of a Product based on User Reviews using Random Forests Algorithm. In: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). Noida, India: IEEE, 2020, p. 112–116.

KAUR, Manpreet. An Approach for Sentiment Analysis Using Gini Index with Random Forest Classification. In: SMYS, S.; TAVARES, João Manuel R. S.; BALAS, Valentina Emilia; et al (Orgs.). Computational Vision and Bio-Inspired Computing. Cham: Springer International Publishing, 2020, v. 1108, p. 541–554.

# Referências

SUMATHI, B.; SIJI GEORGE C.; et al. Classification of Sentiment on Business Data for Decision Making using Supervised Machine Learning Methods. *International Journal of Engineering and Advanced Technology*, v. 9, n. 3, p. 3595–3600, 2020.

ALLAHYARI, Mehdi; POURIYEH, Seyedamin; ASSEFI, Mehdi; et al. Text Summarization Techniques: A Brief Survey. *International Journal of Advanced Computer Science and Applications*, v. 8, n. 10, 2017.

JAKHOTIYA, Aachal; JAIN, Harshada; JAIN, Bhavik; CHANIYARA, Charmi. Text Pre-Processing Techniques in Natural Language Processing: A Review. *International Research Journal of Engineering and Technology (IRJET)*, v. 9, n. 2, p. 878–880, Feb. 2022.

**Obrigado!**