

# Análise de sentimentos em avaliações de supermercados: entendendo a opinião dos consumidores.

Larissa Vitória Vasconcelos Sousa  
*Universidade Federal do Ceará*  
*Campus Jardins de Anita*  
Curso de Ciência de Dados  
Itapajé, CE-Brazil  
larissavvsousa@alu.ufc.br

Ruan Rodrigues Sousa  
*Universidade Federal do Ceará*  
*Campus Jardins de Anita*  
Curso de Ciência de Dados  
Itapajé, CE-Brazil  
ruanrodrigues@alu.ufc.br

## Resumo

Este estudo propõe uma investigação sobre os sentimentos expressos em avaliações online relacionadas à supermercados, utilizando técnicas de análise de sentimentos. A pesquisa abrange uma amostra de opiniões coletadas de 3 empresas para entender as percepções dos consumidores em relação à experiência de compra em supermercados. Por meio de métodos de processamento de linguagem natural, identificamos e categorizamos comentários positivos, negativos e neutros associados a diferentes aspectos, como atendimento ao cliente, qualidade dos produtos e eficiência operacional. Os resultados oferecem insights, revelando não apenas a satisfação dos consumidores, mas também áreas específicas que podem ser aprimoradas pelos estabelecimentos. Essa análise de sentimentos contribui para uma compreensão aprofundada da experiência do cliente, e também fornece informações cruciais para estratégias de gestão e melhoria contínua no setor de supermercados.

**Palavras-chave** — Análise de Sentimentos, NLTK, Supermercados.

## Abstract

This study proposes an investigation into the sentiments expressed in online reviews related to supermarkets, using sentiment analysis techniques. The research covers a sample of reviews collected from 3 companies to understand consumer perceptions of the supermarket shopping experience. Using natural language processing methods, we identified and categorized positive, negative and neutral comments associated with different aspects, such as customer service, product quality and operational efficiency. The results offer insights, revealing not only consumer satisfaction, but also specific areas that can be improved by establishments. This sentiment analysis contributes to an in-depth understanding of the customer experience, and also provides crucial information for management strategies and continuous improvement in the supermarket sector.

**Key words** — Sentiment Analysis, NLTK, Supermarkets.

## I. INTRODUÇÃO

A relevância da Análise de Sentimentos (AS) em comentários de supermercados consiste na capacidade de coletar a opinião dos consumidores. Segundo Martins *et al.* (2020, p. 252) a análise de sentimentos aborda o estudo computacional de opiniões, sentimentos e emoções expressos em texto. Essa abordagem faz uso de técnicas específicas para identificar e extrair informações subjetivas contidas em um determinado material textual [2].

Além disso, em um contexto onde a competitividade entre supermercados é acirrada, a reputação online desempenha um papel crucial na atração e retenção de clientes. Affonso *et al.* (2019, p.44) afirmam que "Clientes ou consumidores são extremamente importantes, pois, além de representarem a fonte de lucro do negócio, se satisfeitos, atuam na divulgação do negócio por meio do boca a boca, o que ajuda na formação da reputação e imagem do negócio." [1].

A análise de sentimentos não apenas fornece insights sobre a percepção do supermercado, mas também possibilita a identificação de áreas específicas que necessitam de aprimoramento, contribuindo assim para o fortalecimento da reputação e a fidelização do consumidor.

De acordo com o estudo realizado por Friedric e Tramontin (2021), a análise de sentimentos tem se destacado nos últimos anos, especialmente com a popularização das redes sociais, proporcionando uma vasta quantidade de informações textuais para extração de conhecimento estruturado em escalas ilimitadas [4].

Segundo a pesquisa de Friedrich *et. al.* (2022), a avaliação de opinião tem experimentado um crescimento significativo devido à popularização das mídias sociais e dos smartphones nas últimas décadas [3].

Nesse contexto, investigar as opiniões dos clientes torna-se essencial para promover uma experiência de compra mais satisfatória e alinhada às necessidades da sociedade. A análise de sentimentos presentes nos comentários não apenas fornece uma compreensão mais profunda das emoções dos consumidores, mas também permite às empresas ajustar suas estratégias de negócios de acordo com as expectativas.

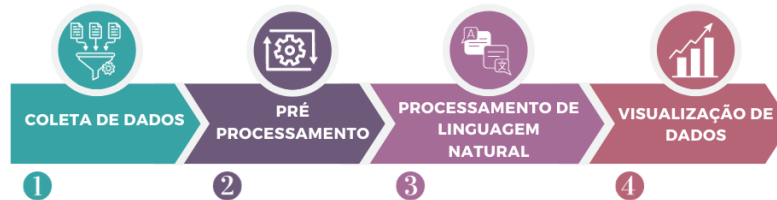
Diante disso, a proposta deste trabalho é realizar uma análise, utilizando técnicas de Processamento de Linguagem Natural (NLP), para avaliar o desempenho de três supermercados específicos com base nas avaliações dos consumidores. A ênfase será na compreensão da satisfação dos clientes e na utilização de abordagens de NLP para a interpretação de insights a partir dessas avaliações.

## II. MATERIAIS E MÉTODOS

Para a condução deste estudo, foram selecionados três supermercados bastante buscados em uma determinada cidade.

A abordagem metodológica adotada neste estudo foi baseada em quatro etapas, tal como exibido na Figura 1.

Fig. 1: Fluxograma



Na primeira etapa, a coleta de dados foi realizada a partir das avaliações disponíveis no *Google Reviews*, empregando técnicas de *web scraping*, que consiste na prática de coletar dados de forma automatizada em páginas da web por meio da ferramenta *Instant Data Scraper*. As informações extraídas incluíram o nome do cliente, o período em que o comentário foi realizado, o conteúdo do comentário, a quantidade de *likes* recebidos, a resposta do supermercado, o status de *local guide*, uma "medalha" concedida aos usuários que frequentemente contribuem com avaliações no *Google Reviews*, a quantidade total de comentários, quantidade total de imagens capturadas do consumidor no geral e a nota, que neste último, foi necessário utilizar JavaScript para coletar a informação.

A nota dos clientes no conjunto de textos é atribuída baseado na quantidade de estrelas que cada consumidor atribui ao estabelecimento. Neste trabalho, considerou-se um comentário negativo se for atribuída até duas estrelas, neutro se possuir 3 estrelas e positivo se possuir 4 ou 5 estrelas. Neste contexto, o corpus ficou constituído por 83 comentários negativos, 127 neutros e 1.134 positivos. Após a coleta de todos os dados, foi construído um corpus contendo um total de 1.344 textos, conforme a Tabela 1 que mostra três exemplos.

TABLE I: Amostra da base de dados

id	Comentário	Supermercado	Nota	Classificação	Feedback
4	"Ótimo atendimento."	A	5	Positivo	Não
352	"Boa"	B	2	Negativo	Sim
139	"Pouco espaço"	C	3	Neutro	Sim

A segunda etapa abrange o pré-processamento dos dados, onde foram realizadas ações como a atribuição de nomes às variáveis nos três conjuntos de dados, a fim de facilitar a compreensão, tratamento de valores ausentes, visando garantir uma melhor análise nos resultados obtidos. Para preservar a privacidade e o anonimato dos indivíduos mencionados nos comentários, procedeu-se à substituição dos nomes dos clientes por identificadores únicos (id), essa abordagem seguiu a regra de que, por exemplo, se João realizou comentários nos três supermercados, ele seria atribuído ao mesmo identificador nos três casos. Logo após, fez-se a junção das três bases de dados tratadas, como indicado na amostra da Tabela 1.

Na etapa três, deu-se início à uma análise exploratória junto de técnicas de Processamento de Linguagem Natural (NLP), efetuadas por meio da biblioteca NLTK, que, nesse caso, foi aplicada a remoção de "stopwords", palavras que podem ser consideradas irrelevantes para o conjunto de dados, sendo seis mais removidas ['e', 'a', 'de', 'um', 'com', 'sem']. A Figura 2 descreve um processo de pré-processamento de texto, com o objetivo de preparar os dados para uma análise. O processo começa com a remoção de stopwords, em seguida, pontuações e símbolos são eliminados do texto para simplificar a análise, finalmente, ocorre a tokenização, dividindo o texto em tokens.

Fig. 2: Técnicas de Processamento de Linguagem Natural



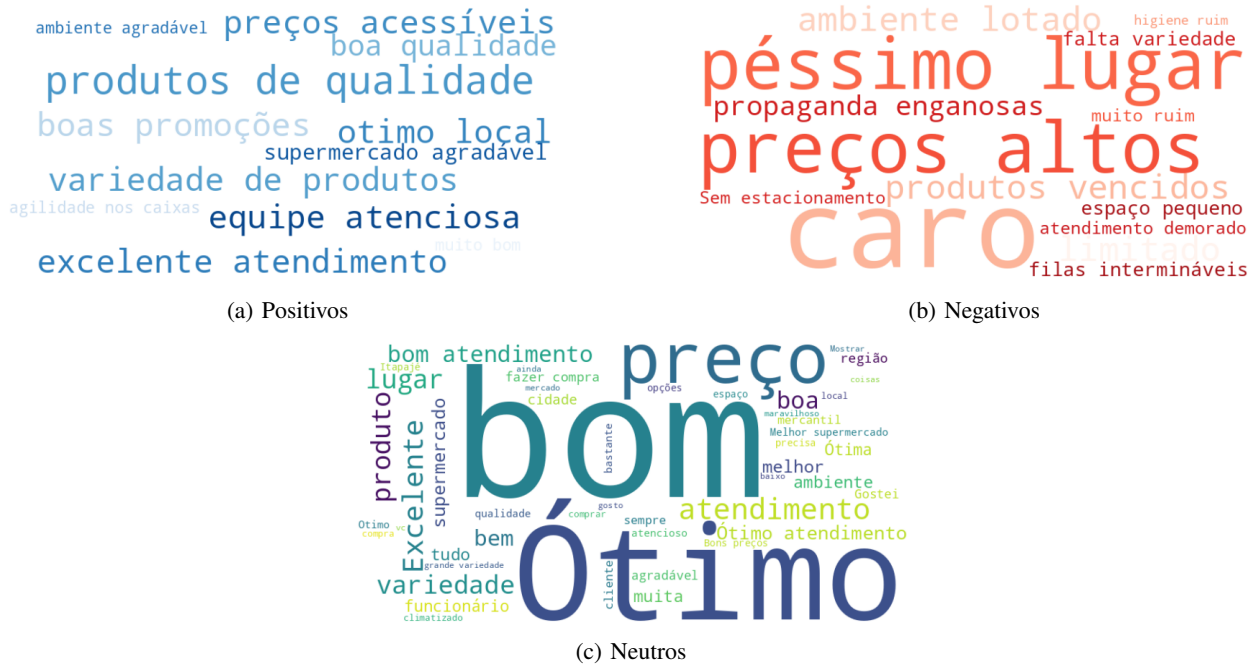
Na quarta e última etapa, procedeu-se à criação de uma nuvem de palavras com o propósito de visualizar graficamente os tokens mais frequentes nos comentários analisados. Além disso, realizou-se uma análise gráfica da distribuição das notas, visando investigar se existem padrões textuais específicos associados às avaliações mais positivas e negativas. Adicionalmente, foi conduzida uma investigação sobre os dez tokens mais prevalentes nos termos, buscando compreender melhor as tendências e temas recorrentes na linguagem utilizada pelos usuários.

Além disso, foi feita a avaliação dos supermercados que demonstram resposta proativa diante das críticas ou elogios dos clientes. Para complementar, o estudo incluiu a apresentação da média das notas, bem como a identificação do comentário mais curtido.

### III. RESULTADOS E DISCUSSÕES

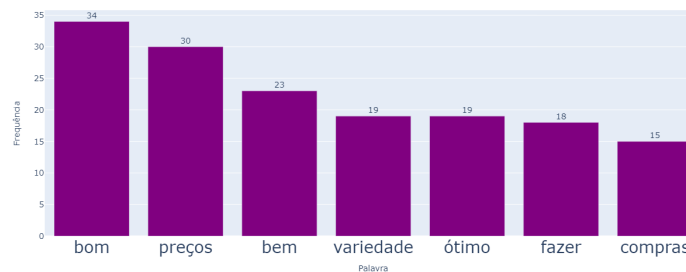
A análise dos resultados inclui a visualização por meio de nuvens de palavras, tanto para uma visão geral, como também baseada na classificação dos comentários em positivos, negativos e neutros, considerando as notas atribuídas. Veja a Figura 3 a seguir.

Fig. 3: Nuvem de tokens



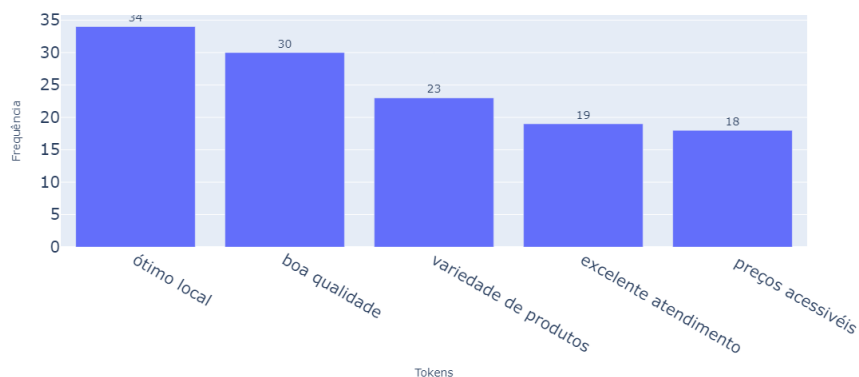
Além da análise das nuvens de palavras, foi conduzido um ranking dos termos mais frequentes nos comentários classificados como positivos, tal como mostra a Figura 4. A presença destes termos sugere áreas específicas que são percebidas de maneira positiva, como o atendimento, qualidade dos produtos, eficiência no serviço e satisfação geral.

Fig. 4: Ranking de termos



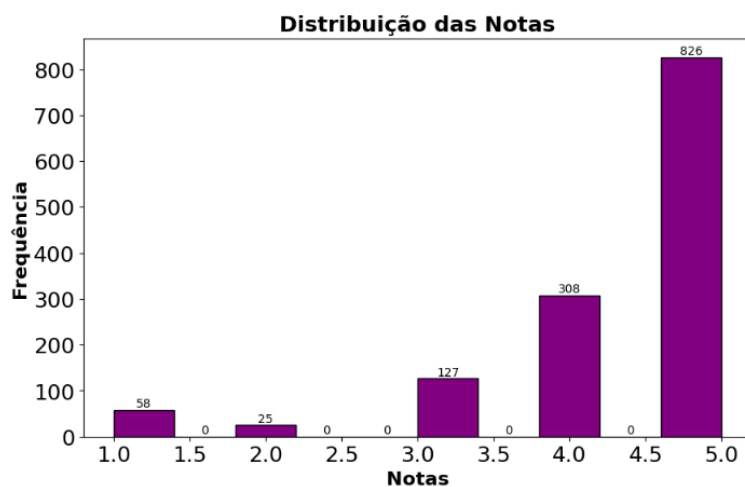
Na Figura 5, foram identificados termos específicos que se repetem nas avaliações positivas, oferecendo uma compreensão mais detalhada das experiências elogiadas pelos clientes. Esses termos indicam áreas específicas, como a qualidade do ambiente, preços atrativos, atendimento de qualidade e uma ampla variedade de produtos, que contribuem para avaliações favoráveis.

Fig. 5: Cinco tokens mais repetidos



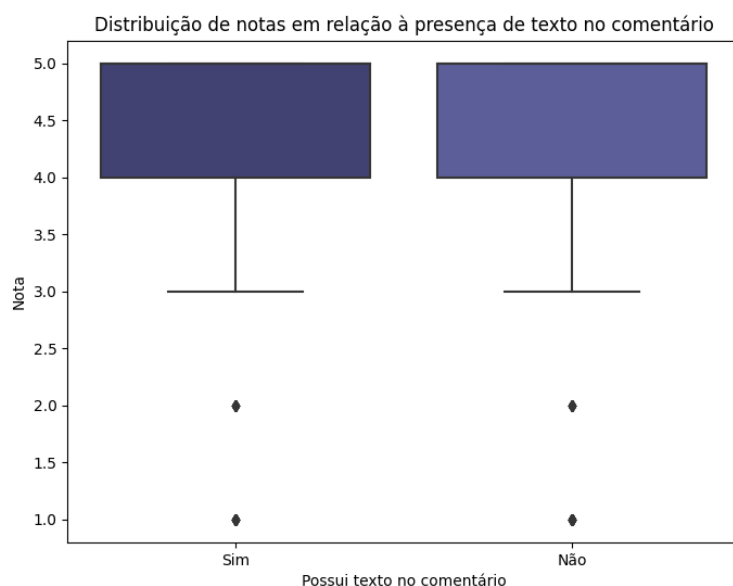
A análise da distribuição das notas revelou uma variedade significativa nas avaliações atribuídas pelos clientes aos supermercados. Nota-se uma predominância de avaliações positivas, com a maioria dos clientes atribuindo notas 4 e 5, conforme evidenciado na Figura 6 e, também, pela média de notas de 4.35. Essa tendência sugere uma satisfação geral dos clientes em relação aos serviços e experiências proporcionados pelos supermercados. No entanto, é importante analisar as avaliações mais baixas (notas 1 e 2) para entender as áreas específicas, como ilustrado anteriormente na nuvem de palavras para comentários negativos, Figura 3.b.

Fig. 6: Distribuição das notas



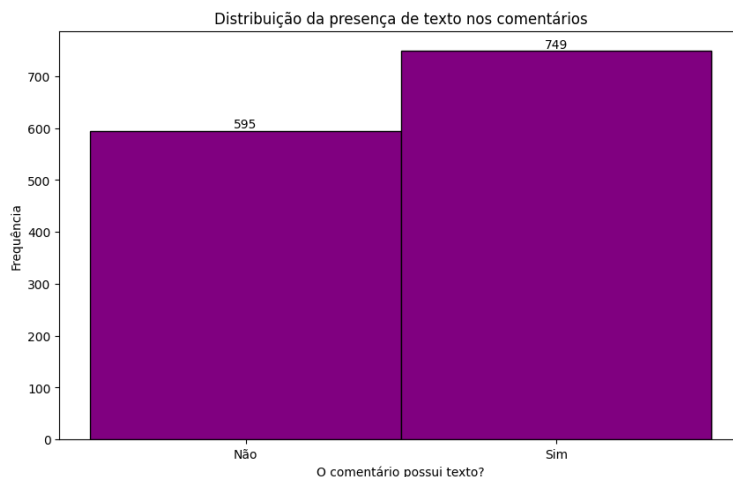
A análise realizada a seguir, na Figura 7, buscou investigar se havia uma relação entre a presença de texto nos comentários e a distribuição das notas atribuídas pelos clientes. Ao criar uma nova variável, que indica se o comentário possui ou não texto (considerando "Sem comentário" como ausência de texto), a intenção era visualizar se havia diferenças significativas nas notas em relação à presença ou ausência de texto. Dado que o gráfico mostrou resultados semelhantes de distribuição de notas tanto para comentários sem texto como para comentários com texto, isso sugere que, na amostra analisada, a presença ou ausência de texto nos comentários não influenciou nas avaliações atribuídas pelos clientes.

Fig. 7: As melhores/piores avaliações acompanham texto?



Na Figura 8, mostra-se a distribuição da presença de texto nos comentários, dado que, alguns clientes apenas colocam a nota (estrela), mas não comentam. É possível ver que, há uma diferença de apenas 154 comentários, em relação aos que comentaram e os que só atribuíram nota.

Fig. 8: Quantidade de comentários



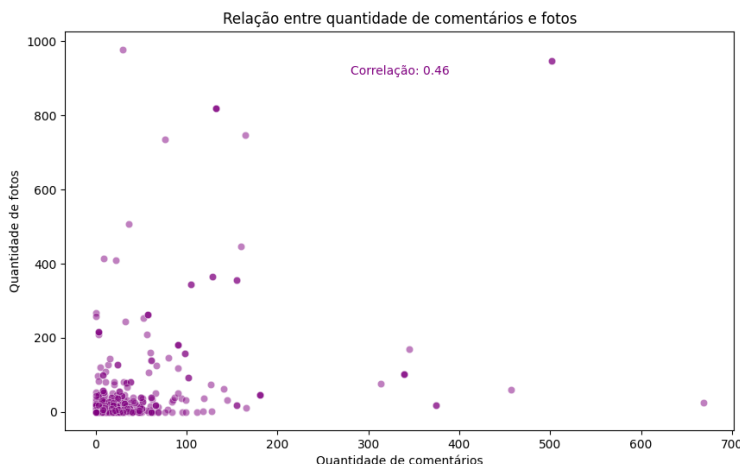
A Tabela 2 irá apresentar o comentário mais curtido entre os dados analisados, identificado pelo número de identificação único (id). O conteúdo do comentário destaca a experiência do cliente em relação ao atendimento do supermercado "A", concedendo uma nota 4. O comentário menciona a excelência do atendimento, mas ressalta demoras ocasionais na entrega, obteve significativa aprovação, recebendo 3 curtidas. Essa informação destaca a importância desse *feedback* específico na percepção dos demais clientes.

TABLE II: Comentário mais curtido

id	Comentário	Supermercado	Nota	Quantidade de likes
74	"O atendimento é ótimo! Mas em se tratando de entrega, às vezes, demora bastante."	A	4	3

A análise da relação entre a quantidade de comentários e a quantidade de fotos, revelou uma correlação positiva de 0.46. Isso indica que há uma associação moderadamente positiva entre essas duas variáveis. Esse resultado sugere que, em geral, à medida que a quantidade de comentários aumenta, há uma tendência de aumento na quantidade de fotos postadas. No entanto, é importante observar que uma correlação de 0.46 não indica uma relação linear perfeita, mas sim uma tendência geral na direção positiva, como mostra a Figura 9.

Fig. 9: Relação entre quantidade de comentários e fotos



A Tabela 3 apresenta insights significativos sobre a interação entre os consumidores e os supermercados, destacando o engajamento *online* e a capacidade de resposta de cada estabelecimento, que podem informar estratégias de gestão, melhorar a qualidade dos serviços e fortalecer o relacionamento com os clientes.

TABLE III: Feedbacks dos supermercados

Supermercado	Total de comentários	Comentários respondidos	Feedbacks (em %)
A	795	367	46,1%
B	364	78	21,4%
C	185	158	85,4%

O supermercado "A" lidera em termos de participação ativa, com 795 comentários, indicando um alto volume de interações dos clientes. Esse número elevado pode refletir a popularidade ou visibilidade do supermercado. Já sobre o quesito capacidade de resposta (*feedback*), o supermercado "C" se destaca, com 85,4% dos comentários sendo respondidos. Isso demonstra um compromisso notável com a atenção às preocupações dos clientes, o que pode contribuir positivamente para a fidelidade do cliente.

## REFERENCES

- [1] AFFONSO, Ligia M F.; RUWER, Léia M E.; GIACOMELLI, Giancarlo. Empreendedorismo. Rio Grande do Sul, Sagah: Grupo A, 2019. E-book. ISBN 9788595028326. Disponível em: <https://app.minhabiblioteca.com.br/books/9788595028326/>. Acesso em: 10 jan. 2024.
- [2] MARTINS, Júlio S.; LENZ, Maikon L.; SILVA, Michel Bernardo Fernandes Da; et al. Processamentos de Linguagem Natural. Rio Grande do Sul, Sagah: Grupo A, 2020. E-book. ISBN 9786556900575. Disponível em: <https://app.minhabiblioteca.com.br/books/9786556900575/>. Acesso em: 10 jan. 2024.
- [3] KUWAKI, V. T. F., Ladeira, M. N., Benitez, M. G. G., & Tramontin Junior, R. J. (2022). *Construção de um Corpus a partir de Avaliações de Supermercados em Português para Análise de Sentimento em Nível de Documento*. UDESC - Santa Catarina State University Joinville. <https://periodicos.univali.br/index.php/acotb/article/view/18741>
- [4] KUWAKI, V. T. F., & Tramontin Júnior, R. J. (2021). *Análise de Sentimentos em Resenhas de Supermercados: Construção de um Corpus Linguístico para o Português e Testes Usando Aprendizado de Máquina*. UDESC - Santa Catarina State University Joinville.
- [5] NLTK. DOCUMENTATION, 2023. Disponível em: org. Acesso em: 12 jan. 2023.