

# A análise de sentimentos e emoções de mensagens de redes sociais envolvendo serviços de entrega de alimentos

João Pedro Rodrigues Alves  
Mario Olimpio de Menezes

Faculdade de Computação e Informática (FCI)  
Universidade Presbiteriana Mackenzie São Paulo, SP – Brasil

[<10382746@mackenzista.com.br>](mailto:10382746@mackenzista.com.br)  
[<mario.menezes@mackenzie.br>](mailto:mario.menezes@mackenzie.br)

2024

## Resumo

*Considerando a pandemia de COVID-19, houve uma expansão significativa nos serviços de entrega, especialmente os de alimentos, tornando-se relevante a investigação das opiniões dos clientes desses serviços. Com base nessa motivação e em trabalhos anteriores, este estudo teve como objetivo investigar técnicas de mineração de texto e Aprendizagem de Máquina (AM) na análise de sentimentos (classificação entre positivo, negativo ou neutro) e emoções (classificação entre felicidade, tristeza, raiva ou neutro) em mensagens publicadas em redes sociais relacionadas aos serviços dos aplicativos Rappi, iFood e Zé Delivery. Para alcançar esse objetivo, foram realizadas as seguintes etapas: coleta de dados em redes sociais como Twitter, Bluesky e Facebook, com rotulação manual dos textos coletados; pré-processamento dos dados utilizando técnicas de Processamento de Linguagem Natural (PLN), como remoção de stopwords, case folding e tokenization; teste da técnica SMOTE para balanceamento das classes e possível incremento na eficácia dos resultados; e experimentos com diferentes técnicas de AM. Da literatura, foram selecionados os algoritmos Support Vector Machine (SVM), Random Forest e Naive Bayes. Os resultados obtidos incluem: um conjunto de dados final composto por 1.728 textos rotulados; um estudo comparativo entre combinações de técnicas de PLN e AM; e um repositório publicado no GitHub, documentando as combinações mais assertivas envolvendo o SVM.*

**Palavras-chave:** Análise de Dados; Aprendizagem de Máquina; Processamento de Linguagem Natural (PLN).

## Abstract

*Considering the COVID-19 pandemic, there was significant growth in delivery services, particularly food delivery, making it relevant to investigate customer opinions regarding these services. Based on this context and previous studies, this research aimed to explore text mining and Machine Learning (ML) techniques in sentiment analysis (classification into positive, negative, or neutral) and emotion analysis (classification into happiness, sadness, anger, or neutral) of opinion messages published on social media about the services offered by the Rappi, iFood, and Zé Delivery applications. To achieve this goal, the following steps were undertaken: data collection from social media platforms such as Twitter, Bluesky, and Facebook, followed by manual labeling of the collected texts; data preprocessing using Natural Language Processing (NLP) techniques, including stopword removal, case folding, and tokenization; testing the SMOTE technique to balance classes and potentially enhance the effectiveness of the results; and conducting experiments with different ML techniques. From the literature, the selected ML algorithms were Support Vector Machine (SVM), Random Forest, and Naive Bayes. The results obtained include: a final dataset composed of 1,728 labeled texts; a comparative study of combinations of NLP and ML techniques; and a repository published on GitHub, documenting the most accurate combinations involving SVM.*

**Keywords:** Data Analysis; Machine Learning; Natural Language Processing (NLP).

## 1 Introdução

Os serviços de entrega (*delivery*) de alimentos já vinham ganhando relevância na última década; no entanto, a pandemia de COVID-19 acelerou significativamente sua popularização e importância. De acordo com a pesquisa de Brandão (2021), 80% dos brasileiros pretendiam continuar utilizando serviços de *delivery* no período pós-pandemia. Além disso, 58% dos entrevistados destacaram a valorização da qualidade na entrega (BRANDAO, 2021). Assim, para manter os clientes engajados, as empresas do setor devem priorizar a qualidade dos serviços oferecidos (SAYYIDAH et al., 2021). Nesse contexto, uma estratégia eficaz para permanecer competitivo no mercado e melhorar a qualidade dos serviços é ouvir e compreender a opinião dos clientes.

Segundo Rossi (2019), a exploração de dados urbanos é uma ferramenta crucial para a tomada de decisões e para o desenvolvimento de serviços inteligentes, um dos pilares fundamentais de uma Cidade Inteligente. Atualmente, diversos recursos digitais ou inteligentes são capazes de capturar aspectos sociais em ambientes urbanos. Em consonância com essa perspectiva, este projeto propõe investigar os sentimentos expressos nas opiniões dos usuários de serviços de *delivery* em suas publicações nas redes sociais, fornecendo subsídios para a tomada de decisões (YUGOSHI, 2018; SAYYIDAH et al., 2021). Visando compreender melhor as opiniões expressas nos textos analisados, este trabalho busca ampliar o escopo da análise de sentimentos, incorporando também a análise de emoções, possibilitando uma avaliação mais detalhada do impacto e desempenho desses serviços.

Este trabalho dá continuidade à pesquisa desenvolvida durante um projeto de Iniciação Científica, cujos resultados foram publicados no artigo A análise de sentimentos de mensagens do Twitter envolvendo serviços de entrega de alimentos. Naquela pesquisa, foram exploradas as bases para a análise de sentimentos em mensagens publicadas no Twitter relacionadas a aplicativos de *delivery*. O presente Trabalho de Conclusão de Curso aprofunda as propostas sugeridas como continuidade daquela pesquisa, com foco

em expandir a base de dados, incluindo outras redes sociais, e em ampliar a análise de sentimentos para contemplar também a análise de emoções.

Com base no exposto, o problema de pesquisa abordado neste projeto é formulado pela seguinte questão: Os algoritmos de classificação tradicionais são eficientes na tarefa de análise de sentimentos e emoções de mensagens extraídas de redes sociais? Para responder a essa questão, o projeto teve como objetivo realizar um estudo comparativo sobre o uso das diferentes técnicas e algoritmos previamente levantados, com o intuito de analisar seus resultados por meio de métricas de desempenho de modelos preditivos.

A seguir, este artigo apresentará o referencial teórico, de forma resumida, expondo as técnicas, métodos e algoritmos utilizados na pesquisa, com base na literatura consultada pelo pesquisador, e que serviram como fundamentação para a condução deste projeto. Em sequência, será apresentada a metodologia adotada no desenvolvimento do projeto, com a descrição dos métodos utilizados, conforme sua ordem cronológica. Posteriormente, serão apresentados os resultados obtidos, destacando suas contribuições. Por fim, a conclusão sintetizará os resultados, realizará uma breve discussão e apontará sugestões para futuros trabalhos.

## 2 Referencial Teórico

Com base na metodologia apresentada no trabalho de [Alves e Oliveira \(2023\)](#), este estudo adota as técnicas e algoritmos levantados naquele para alcançar os objetivos específicos desta pesquisa. Aqueles que se mostraram ineficientes dentro do contexto, Doc2Vec e KNN, não foram considerados para o presente trabalho. Segue a base teórica revisada, como base para o presente projeto.

### 2.1 Mineração de texto

Segundo [Peixoto \(2021\)](#), textos podem ser classificados como dados não estruturados, os quais representam a maior parte dos dados disponíveis atualmente. Esse cenário decorre do avanço significativo das tecnologias sociais, que popularizaram os meios digitais e, diariamente, geram uma quantidade massiva de dados, especialmente em formato textual. Ainda de acordo com o autor, a natureza não padronizada desse tipo de dado torna o processo de extração de informações mais desafiador em comparação a outros formatos. Nesse contexto, a Mineração de Dados desempenha um papel fundamental ao possibilitar a extração de informações úteis a partir de dados textuais, utilizando um conjunto integrado de técnicas matemáticas, estatísticas e computacionais que permitem a análise eficiente de seu conteúdo.

### 2.2 Processamento de Linguagem Natural

A partir de dados textuais, técnicas de Processamento de Linguagem Natural (PLN) são empregadas para eliminar ruídos e informações desnecessárias, realizando o pré-processamento e preparando esses dados para análise. O PLN é o estudo estruturado da forma da linguagem humana, permitindo que os computadores a compreendam de maneira semelhante aos seres humanos, com foco em um processamento que extraia informações manipuláveis por algoritmos. [Sumathy e Chidambaram \(2013\)](#) destacam que o PLN integra o campo interdisciplinar da Mineração de Texto, podendo ser considerado um meio para alcançar a base de informações essencial para o início efetivo de uma análise.

Em pré-processamento de dados utilizando PLN existem técnicas que ajudam no tratamento de ruídos e outros aspectos nos dados textuais obtidos, como mostrado na Tabela 1

Tabela 1 – Técnicas de pré-processamento de linguagem natural

Técnica	Descrição
Tokenization	Segmenta todo o texto em palavras ou tokens, ou seja, as palavras em cada frase são separadas individualmente ( <i>token</i> ), e todos os caracteres especiais, incluindo <i>links</i> , pontuações e acentos, são retirados por serem desnecessários para a análise posterior.
Case Folding	Remove todas as distinções de letras maiúsculas e minúsculas, ajustando todas para minúsculas por padrão. Isso evita que palavras idênticas sejam consideradas diferentes umas das outras.
StopWords	Remove palavras ou pontuações irrelevantes para a análise dos textos em questão, melhorando a qualidade dos dados obtidos e impedindo que termos desnecessários para a análise de sentimentos ou emoções permaneçam neles.
Stemming	Simplifica o texto recuperando as palavras para a sua forma raiz, reduzindo flexões e derivações, facilitando o entendimento básico dos dados obtidos.

Fonte: (SUMATHY; CHIDAMBARAM, 2013; RAVI; RAVI, 2015; ROSSI, 2019; FACELI et al., 2021; NURDENI; BUDI; SANTOSO, 2021; PRIBADI et al., 2022)

Ainda na etapa de pré-processamento de dados, conforme Peixoto (2021), é nesse momento que os dados textuais são convertidos para um modelo de representação, transformando-os em representações numéricas. Entre os modelos de representação disponíveis, destaca-se o *Term Frequency-Inverse Document Frequency* (TF-IDF).

O TF-IDF é uma técnica matemática utilizada para calcular a relevância de um termo com base na sua frequência de ocorrência em um conjunto de documentos ou textos. Essa técnica converte textos em modelos vetoriais e é amplamente empregada como uma medida estatística, frequentemente utilizada como peso na Mineração de Dados (ROSSI, 2019; PEIXOTO, 2021).

## 2.3 Análise Descritiva

Antes de avançar para a fase de análise preditiva, com a aplicação de algoritmos de Aprendizagem de Máquina, é necessário extrair informações iniciais dos dados, que serão utilizadas na etapa subsequente. Uma análise exploratória ou estatística descritiva pode ser realizada para resumir as principais características dos dados coletados, empregando técnicas quantitativas para identificar tendências e possíveis variações atípicas. Essas informações numéricas recém-descobertas são cruciais para a seleção de modelos preditivos adequados e podem influenciar diretamente a aplicação de técnicas específicas (GALLAGHER; FUREY; CURRAN, 2019; VILLARROEL, 2020; SAYYIDAH et al., 2021).

Durante essa etapa, foi identificada uma desproporção entre as classes em relação ao total de dados, evidenciando um desbalanceamento na base. Esse desbalanceamento pode introduzir vieses nos modelos gerados. Uma solução potencial é a aplicação da técnica de *Oversampling*, que busca equilibrar a proporção entre a classe minoritária e a majoritária na base de dados (GHOSH et al., 2019; PRIBADI et al., 2022). Um dos algoritmos mais

utilizados para *Oversampling* é o *Synthetic Minority Over-sampling Technique* (SMOTE), que se baseia no algoritmo *K-Nearest Neighbors* (KNN) para gerar instâncias sintéticas da classe minoritária, equilibrando-a em relação à classe majoritária (OLUSEGUN et al., 2023; ANNISA; SETIAWAN, 2022; NYOTO; RULDEVIYANI, 2022).

## 2.4 Aprendizagem de Máquina

A Aprendizagem de Máquina (AM) é uma subárea da Inteligência Artificial (IA) que engloba técnicas e algoritmos voltados para a identificação de padrões por meio da construção de sistemas que permitem às máquinas adquirirem conhecimento de forma automática. Esses sistemas acumulam experiências a partir de soluções ótimas e, com base nelas, tomam decisões. A AM pode ser preditiva, descritiva ou ambas, e tem como objetivo aprender e padronizar informações a partir de uma amostra de dados, visando atender ao propósito da análise em questão (YUGOSHI, 2018; PEIXOTO, 2021).

Conforme Yugoshi (2018), os algoritmos de *aprendizado supervisionado* utilizam exemplos rotulados, conhecidos como conjunto de treinamento, para criar uma função de mapeamento com base no tipo de atributo-alvo. Essa base rotulada orienta o aprendizado supervisionado e viabiliza a construção de modelos úteis para a classificação de sentimentos, que podem apresentar bom desempenho desde que disponham de uma base de dados adequada e da escolha correta de técnicas (ROSSI, 2019).

Para avaliar a qualidade dos modelos aplicados, são utilizadas métricas de desempenho que medem o nível de classificações equivocadas realizadas pelo algoritmo. Antes da aplicação dessas métricas, é construída uma Matriz de Confusão, conforme apresentado na Figura ?? (VILLARROEL, 2020; NURDENI; BUDI; SANTOSO, 2021; ANNISA; SETIAWAN, 2022; FAISAL et al., 2022).

		Predito	
		Positivo	Negativo
Original	Positivo	VP	FN
	Negativo	FP	VN

Fonte: Elaborado pelo autor.

Figura 1 – Matriz de Confusão genérica para duas classes

A partir da Matriz de Confusão, é possível calcular as métricas de desempenho dos modelos, como Acurácia, Precisão e Revocação. Essas métricas são fundamentais para avaliar o desempenho de modelos de classificação, e sua análise conjunta é essencial para selecionar o modelo preditivo mais eficiente (VILLARROEL, 2020; BINSAR; MAURITSIUS, 2020; NURDENI; BUDI; SANTOSO, 2021; ANNISA; SETIAWAN, 2022; FAISAL et al., 2022).

Neste trabalho, foi conduzido um estudo sobre técnicas de AM aplicadas à análise de sentimentos e emoções, com foco na classificação dos dados coletados. O objetivo foi extrair informações de mensagens dos clientes relacionadas à entrega de alimentos, identificando como eles se sentem em relação aos serviços prestados. A seguir, com base em trabalhos relacionados que serviram como referência para os experimentos deste projeto, apresenta-se uma breve descrição dos algoritmos de AM selecionados.

- *Naive Bayes* (NB): é um algoritmo probabilístico que calcula um conjunto de proba-

bilidades com base na frequência e combinação dos dados, utilizando o Teorema de Bayes como fundamento. Durante o treinamento, são calculadas as probabilidades de cada atributo em relação às classes, enquanto no teste, são determinadas as probabilidades de cada exemplo não visto, classificando aqueles com maior probabilidade de ocorrência com base no treinamento. O algoritmo é considerado "ingênuo" porque assume que todos os atributos são independentes e não exercem influência uns sobre os outros (YUGOSHI, 2018; ALSAEEDI; KHAN, 2019; ROSSI, 2019; LI, 2021).

- *Support Vector Machine* (SVM): amplamente reconhecido por seu bom desempenho em análises de sentimentos, é um classificador que identifica limites de decisão ao analisar e separar informações importantes em vetores. Cada dado é atribuído a uma classe, e o algoritmo determina a fronteira entre elas, minimizando ambiguidades. A melhor separação é obtida ao maximizar a margem, definida como a distância entre o hiperplano (linha que separa as classes) e os dados mais próximos (RAVI; RAVI, 2015; YUGOSHI, 2018; DHIR; RAJ, 2018; ROSSI, 2019).
- *Random Forest* (RF): amplamente aplicado em tarefas como a classificação de opiniões no Twitter, é considerado uma evolução do algoritmo *Decision Tree* (DT). Ele utiliza múltiplas árvores de decisão treinadas simultaneamente, caracterizando-se como um método baseado em *Ensemble Learning* (Aprendizado por Conjunto), que combina vários modelos para realizar a mesma tarefa. Cada árvore é treinada com um subconjunto dos dados de treinamento, superando limitações das árvores tradicionais, como o *overfitting*, que ocorre devido ao excesso de profundidade (FLORÊNCIO, 2016; ROSSI, 2019; DHIR; RAJ, 2018; RANE; KUMAR, 2018; BINSAR; MAURITSIUS, 2020; ZAHOR; ROHILLA, 2020).

Ao realizar o teste da aplicação de diferentes algoritmos de aprendizagem de máquina utilizando os dados da base e as métricas obtidas, torna-se necessário comparar o desempenho dos modelos gerados, levando-se em consideração as métricas mencionadas anteriormente.

## 2.5 Análise de Sentimentos e Emoções

Segundo Peixoto (2021), a satisfação do cliente está diretamente relacionada às suas emoções e sentimentos em relação ao que foi consumido, além da correspondência dessa experiência com suas expectativas. Dessa forma, para aumentar a satisfação do cliente, é essencial compreender o que ele espera dos serviços oferecidos. A análise desses sentimentos pode, portanto, fornecer dados valiosos para a tomada de decisões.

Yugoshi (2018) define a Análise de Sentimentos (AS), também conhecida como Mineração de Opiniões, como o estudo das emoções e opiniões das pessoas em relação ao serviço analisado. O autor também destaca o crescente volume de opiniões compartilhadas em redes sociais.

De acordo com Rossi (2019), quando um texto é considerado subjetivo, uma opinião apresenta cinco aspectos principais: o nome da entidade alvo do comentário, o aspecto específico da entidade (quando aplicável), a polaridade do sentimento, a fonte ou detentor do sentimento e o momento em que a opinião foi expressa. Ainda segundo Rossi (2019), a polaridade é definida como o grau da opinião expressa, podendo ser positiva, neutra ou negativa (classes do tipo ternário), representadas por intervalos numéricos ou categorizadas de forma qualitativa.

Villarroel (2020) estrutura a *AS* em três etapas principais: a identificação de opiniões (usando os cinco aspectos definidos por Rossi (2019)), a classificação da polaridade e a sumarização, que consiste na organização dos resultados de maneira concisa. Para realizar a análise proposta, existem diferentes abordagens, incluindo métodos estatísticos e técnicas de *Machine Learning* (AM).

Na abordagem estatística, é possível utilizar as informações obtidas na etapa de análise descritiva para a criação de nuvens de palavras, que representam a frequência e a intensidade de termos em um texto, e gráficos de diversos tipos (como de barras e pizza), que facilitam a visualização das informações de maneira prática. Por outro lado, a abordagem baseada em AM envolve a aplicação de uma ou mais técnicas semelhantes às descritas anteriormente (VILLARROEL, 2020).

### 3 Metodologia

A condução desta pesquisa foi realizada por meio de experimentos, considerando o problema de pesquisa e a sua questão. Isso permitiu testar hipóteses para estabelecer uma relação entre os elementos de estudo e os efeitos que eles produzem em um ambiente controlado. Dessa forma, foi realizado o seguinte conjunto de atividades durante o período deste projeto:

1. Coleta de Dados
2. Rotulação Manual dos Dados
3. Análise Descritiva
4. Pré-Processamento de Dados
5. Aprendizagem de Máquina

#### **Coleta de Dados:**

Foi utilizada a plataforma *Apify* para extração de dados das seguintes redes sociais: Bluesky, Twitter, Facebook, Instagram e Threads, além de agrupar esses dados com a base de dados coletada do Twitter do trabalho anterior a esse. A *Apify* se trata de uma plataforma de automação e *web scraping*, que permite fazer buscas de dados em diversos sites da web e obter os resultados em formato de relatórios no Excel. Dentro da plataforma existem aplicações específicas para *web scraping* em diferentes sites.

Para o Bluesky, rede similar ao Twitter que esteve em alta no Brasil em 2024, foi utilizada a aplicação *Bluesky Posts Scraper*, que coletou publicações que citavam diretamente o nome dos aplicativos analisados. Para o Facebook, foi utilizada a aplicação *Facebook Comments Scraper*, que buscou comentários das publicações mais recentes das páginas dos aplicativos. Já para o Threads, foi utilizada a aplicação *Threads Search Scraper*, que retornou uma quantidade ínfima de dados, sendo assim descartado da análise.

No momento da coleta dos dados para a pesquisa, o Twitter passava por um processo judicial no Brasil e estava fora do ar, dessa forma os dados utilizados dessa rede são exclusivamente aqueles que foram coletados pela própria API do Twitter, quando sua utilização era gratuita, no projeto anterior o qual essa pesquisa foi baseada.

#### **Rotulação Manual dos Dados:**



Após a coleta e definição de variáveis, foi realizada a etapa de rotulação dos dados. Esse processo foi feito manualmente em uma planilha do Excel, em que era lido e interpretado cada um dos *posts* coletados e classificado com uma opinião entre positiva, negativa ou neutra, e com uma emoção entre raiva, felicidade, medo, tristeza ou neutro. Foram analisados quase 2.500 *posts* das redes citadas, desde uma série de casos foram descartados por não terem nenhuma ligação com o objeto de estudo ou que não era possível definir nenhuma emoção, restando ao final uma base de dados com 1.728 *posts* classificados manualmente com sentimento e emoção. A distribuição e proporção dessas classes na base obtida podem ser observadas na Figura 2.

Sentimento			Emoção		
Classe	Quantidade	Porcentagem	Classe	Quantidade	Porcentagem
Negativo	698	40,39%	Raiva	508	29,40%
Positivo	585	33,85%	Felicidade	496	28,70%
Neutro	445	25,75%	Tristeza	446	25,81%
			Neutro	278	16,09%
Total	1728		Total	1728	

Figura 2 – Proporção da Base por Classe

### Análise Descritiva:

Elaboração de uma análise estatística descritiva dos dados brutos coletados, com a finalidade de encontrar informações relevantes a seu respeito, com tabelas, nuvem de palavras, gráficos de pizza e gráficos de barras. Para construção desses gráficos foram utilizadas as bibliotecas do Python: Seaborn, Mplotlib, Wordcloud e Numpy.

### Pré-Processamento de Dados:

Após isso, foi realizada a etapa de pré-processamento de dados com as diferentes técnicas de PLN. Para limpeza geral da base, foi criada uma função que, nessa ordem, remove: citações de perfis do Twitter (que começam com @), as *hashtags* (que começam com #), o marcador de *retweet* (a *string* RT isolada), links de sites, converte os textos completamente para letras minúsculas, remove pontuações, *emojis*, caracteres especiais, acentuação, números e espaços extras. Esse processo foi realizado utilizando-se das bibliotecas do Python: pandas, re, string e unicodedata.

Em seguida, foi aplicada a remoção de *stopwords*, utilizando-se do corpus em português da biblioteca do Python nltk. Por fim, foi feita a *tokenization* nos textos, separando cada palavra em um *token*, e aplicado *stemmization*, que pega cada *token* e reduz a raiz da palavra, diminuindo a variação de palavras de mesmo significado. Feito tudo isso, foi aplicada a técnica TF-IDF, com a biblioteca sklearn do Python, vetorizando os dados textuais limpos e pré-processados.

Já pensando no possível desequilíbrio que a base apresentaria, a técnica de *Over-sampling* foi levantada com o uso do método SMOTE, que gerou novas amostras sintéticas para equilibrar o *dataset*. Apesar de, diferente do trabalho anterior, as classes obtidas na base estabelecida no presente projeto não estão muito desequilibradas, foi-se avaliado que ainda há uma presença minoritária da classe neutra, podendo influenciar nos resultados.

### Aprendizagem de Máquina:

A partir do *dataset* final, foi iniciada a aplicação dos algoritmos de Aprendizagem de Máquina com o auxílio da biblioteca do Python Scikit-learn. Como explicitado no referencial



teórico, foram selecionados três dos algoritmos mais citados na literatura levantada, sendo eles *Naive Bayes*, *Random Forest* e SVM. Todos esses foram aplicados e testados com diferentes combinações de técnicas de PLN, em busca dos melhores resultados.

Os testes foram conduzidos seguindo combinações de aplicação dos três algoritmos com aplicação ou não aplicação de remoção de *Stop Words*, o uso SMOTE para equilibrar o conjunto de treino e as seguintes proporções de treino e teste: 60%/40%, 70%/30% e 80%/20%. Além disso, os melhores modelos obtidos foram rodados com o *Search Grid* do Scikit-learn para testar diferentes configurações de parâmetros, com o objetivo de encontrar um modelo com melhores métricas de avaliação.

No SVM, além do *Kernel*, foram testados configurações do  $c$ , parâmetro de regularização que controla o quão generalizado ou minimização de erros o modelo será; e  $\gamma$ , que define modelos de mais complexos a mais simples. No NB foram testados configurações do  $\alpha$ , parâmetro de suavização de Laplace; e o `fit_prior`, que define se os *priors* das classes deve ser uniformes. Já no RF, foram testados: `n_estimators`, número de árvores na floresta; `max_depth`, profundidade máxima de cada árvore; `min_samples_split`, número mínimo de amostras para dividir um nó; `min_samples_leaf`, número mínimo de amostras que um nó folha deve conter; e `max_features`, número de recursos considerados ao procurar a melhor divisão.

Os resultados obtidos foram documentados com as métricas de desempenho descritas no referencial teórico: acurácia, precisão, revocação e matriz de confusão.

## 4 Resultados e discussão

Este projeto produziu como resultados e contribuições um *dataset* rotulado de mensagens sobre serviços de *delivery* de alimentos provindo de diversas redes sociais, um estudo comparativo da aplicação de diferentes algoritmos e técnicas para predição nesse contexto e uma análise dos dados com base no estudo obtido. Para a análise deste projeto, foram escolhidos três dos serviços de entrega de alimentos mais populares do Brasil: iFood, Rappi e Zé Delivery.

O projeto foi desenvolvido na linguagem de programação Python pelo ambiente controlado do Google Colab. Para manipulação do *dataset* foi usada a biblioteca Pandas, para construção dos gráficos da análise exploratória as bibliotecas Matplotlib e Seaborn, e para aplicação dos modelos de aprendizagem de máquina e suas métricas de desempenho foi usada a biblioteca Scikit-Learn.

A Figura 3 apresenta um gráfico mostrando as distribuições de emoções por sentimentos. Nele é possível notar, como já citado, a desproporção da classe neutra. Como esperado, a classe negativa predomina emoção de raiva e uma parcela considerável de tristeza. Já a na classe positiva, além da predominância da emoção felicidade, existe uma parcela de positivos tristes, isso pois alguns comentários apresentavam teor de tristeza mesmo com uma opinião positiva sobre o app, como por exemplo, casos em que o cliente se mostrava triste por não ter dinheiro para gastar com o aplicativo de *delivery*.

Em relação as redes sociais das quais os comentários foram retirados, a maioria dos dados foram retirados do Twitter, em projeto anterior, com revisão dos sentimentos rotulados e classificação das emoções. As outras redes que representam a base obtida foram Bluesky e Facebook, este último tendo uma distribuição pequena de comentários que puderam ser aproveitados na análise. A Figura 4 apresenta as proporções representativas

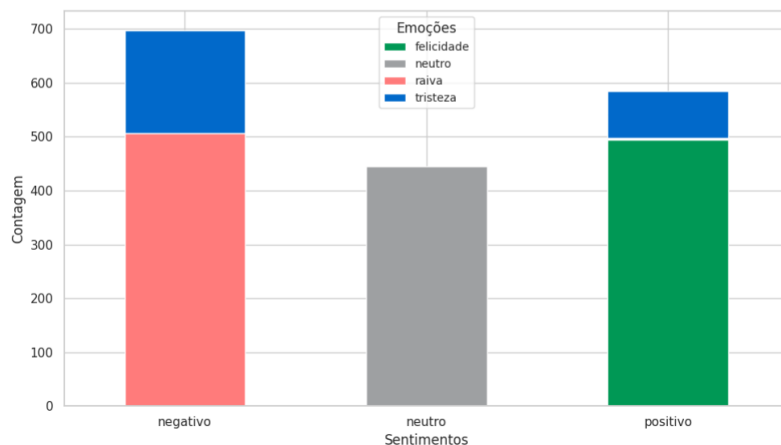


Figura 3 – Gráfico da distribuição das classes

de cada rede social na base.

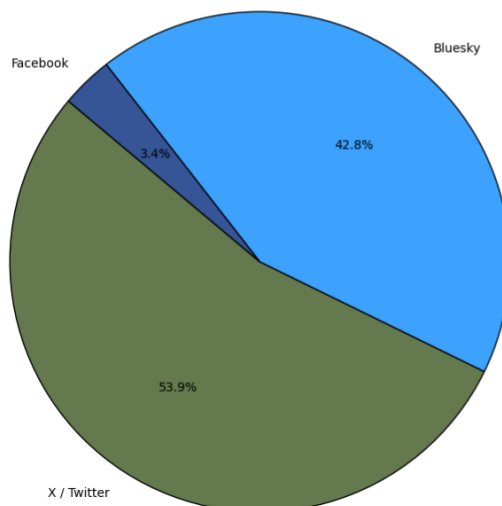


Figura 4 – Gráfico da distribuição das redes sociais

Dos aplicativos buscados, temos uma predominância abundante de dados sobre o iFood, algo esperado dada a popularidade do app no Brasil. Além disso, os novos dados obtidos neste projeto foram sobretudo retirados do Bluesky, que estava "substituindo" o X (antigo Twitter) durante o período em que esteve banido no país, este do qual apresentou poucas boas amostras para análise do Zé Delivery e nenhuma para o Rappi, aumentando consideravelmente a proporção do iFood, único que apresentou uma boa quantidade de dados pertinentes ao projeto. Na Figura 5 pode ser observada a distribuição exata dos apps citados.

A Figura 6 apresenta uma nuvem de palavras que mostra as palavras mais usadas nos *posts* capturados para a análise do projeto.

Quando observa-se individualmente cada classe, é interessante comparar as nuvens de cada emoção, representadas pelas Figuras 7, 8 e 9.

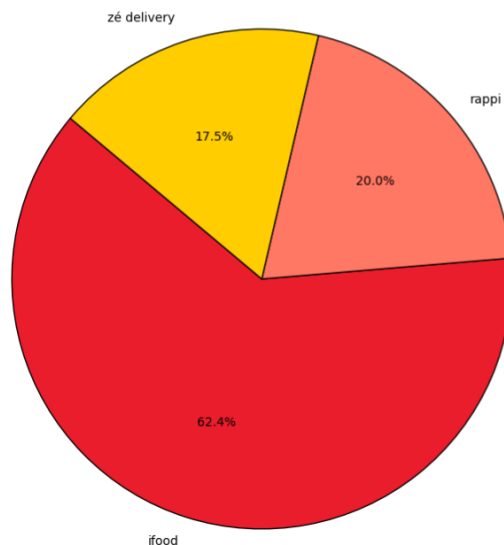


Figura 5 – Gráfico da distribuição dos aplicativos de delivery



Figura 6 – Nuvem de palavras da base geral

Na nuvem de felicidade é possível observar a relevância das palavras não, pedir, promoção, reais, cupom, amo, hoje e casa. Esses *tokens* indicam que a felicidade dos clientes está muito atrelada a compras que valem a pena, seja por comodidade ou por economia, quando se fala em felicidade dentro dessa base o valor e o desconto aplicado a compra são destaques nesse quesito.

Na nuvem de tristeza as palavras não, pedir e cupom ainda predominam, porém pode-se observar relevância em palavras como queria, vontade, comprar e entrega. Isso revela que essa classe se destaca pelos clientes que desejam comprar, mas por algum motivo não conseguem, seja por dificuldades com a plataforma, sendo um sentimento negativo em relação ao app, ou por não ter condições financeiras ou capacidade de escolha para pedir um *delivery*, esses casos inclinados para o sentimento positivo em relação ao app.

Já na nuvem de raiva, além do não, as palavras suporte, nada, hora, entregador, entrega, reembolso, ainda e odeio se mostram relevantes dentro da classe. Isso aponta que as reclamações que fazem os clientes sentirem raiva estão atreladas ao atraso no prazo de entrega e problemas com o suporte do aplicativo quando acontece algum problema com a



Figura 7 – Nuvem de palavras da emoção felicidade



Figura 8 – Nuvem de palavras da emoção tristeza

entrega.

As Figuras 10 e 11 mostram as quantidades e porcentagens de cada app por classificação de sentimento e emoção, respectivamente. Nelas é possível analisar a situação de cada aplicativo frente ao conjunto de dados obtido. Destaca-se que se espera a predominância do sentimento negativo, tendo em vista que a maioria dos clientes que se dispõem para comentar sobre sua experiência com um aplicativo nas redes sociais são motivados por problemas enfrentados.

O iFood mesmo sendo o aplicativo que mais aparece, consegue manter uma boa distribuição entre suas classes, mesmo com o negativo em primeiro a classe não se distanciar muito das outras, enquanto a emoção de raiva está bem próxima da de felicidade. Isso pode indicar que o domínio de mercado da plataforma seja resultado de medidas efetivas para manter o aplicativo estável em sua operação.

O Rappi já se destaca no lado negativo, tanto o sentimento negativo quanto a emoção de raiva predominam mais da metade de suas ocorrências na base. Esse resultado



- SVM:  $C = 1$ ,  $\gamma = 0.01$ ,  $\text{kernel} = \text{linear}$ ;
- RF:  $\text{max\_depth} = \text{None}$ ,  $\text{max\_features} = \log 2$ ,  $\text{min\_samples\_split} = 10$ ,  $\text{n\_estimators} = 500$ ,  $\text{min\_samples\_leaf} = 1$ ;
- NB:  $\alpha = 0.1$ ,  $\text{fit\_prior} = \text{True}$ .

As configurações encontradas com o *Search Grid* para emoções foram:

- SVM:  $C = 10$ ,  $\gamma = 1$ ,  $\text{kernel} = \text{rbf}$  (radial);
- RF:  $\text{max\_depth} = \text{None}$ ,  $\text{max\_features} = \log 2$ ,  $\text{min\_samples\_split} = 1$ ,  $\text{n\_estimators} = 200$ ,  $\text{min\_samples\_leaf} = 2$ ;
- NB:  $\alpha = 1$ ,  $\text{fit\_prior} = \text{False}$ .

Sentimento							
Algoritmo	Treino/ Teste	StopWord	SMOTE	Acurácia	Precisão	Revocação	Matriz de Confusão
SVM (linear)	60/40	Não	Não	0,649	0,642	0,649	[ 227 24 40 ] [ 50 69 62 ] [ 39 28 153 ]
SVM (linear) c/ Search Grid*	60/40	Não	Não	0,640	0,634	0,640	[ 222 27 42 ] [ 53 68 67 ] [ 33 27 153 ]
SVM (radial)	60/40	Não	Sim	0,639	0,641	0,639	[ 231 11 49 ] [ 56 58 67 ] [ 46 21 153 ]
RF com Search Grid*	70/30	Sim	Sim	0,636	0,634	0,636	[ 163 19 34 ] [ 35 62 49 ] [ 22 30 105 ]
RF	70/30	Sim	Sim	0,611	0,609	0,611	[ 153 26 33 ] [ 31 65 47 ] [ 29 36 99 ]
NB	60/40	Não	Sim	0,630	0,623	0,630	[ 222 31 38 ] [ 53 77 51 ] [ 35 48 137 ]
NB com Search Grid*	60/40	Não	Sim	0,616	0,609	0,616	[ 211 37 43 ] [ 58 74 56 ] [ 31 41 141 ]

Figura 12 – Resultados mais relevantes para análise de sentimentos

O uso das técnicas StopWord e SMOTE influenciaram nos resultados dependendo do algoritmo aplicado. Para sentimentos, o SMOTE não influenciou bem com SVM, porém para RF e NB ele ajudou a encontrar melhores índices, enquanto a remoção de *Stop Words* só ajudou com o RF. Para emoções, o uso do SMOTE variou mais, sendo que com NB os resultados foram melhores sem ele, enquanto para a remoção de *Stop Words*, NB e RF tiveram melhores resultados com sua aplicação.

Os resultados obtidos com os experimentos realizados não são muito representativos, talvez ainda por carência de uma base de dados maior e a dificuldade que os algoritmos tradicionais têm de lidar com a linguagem das redes sociais. Dentro da base obtida, é comum encontrar frases que usavam palavras com teor negativo para representar um sentimento positivo, e vice versa, o que pode confundir e induzir os algoritmos a classificarem esses dados com imprecisão, sobretudo quando falamos de emoções, pois temos muitas outras variações possíveis envolvidas.

Emoção							
Algoritmo	Treino/ Teste	StopWord	SMOTE	Acurácia	Precisão	Revocação	Matriz de Confusão
SVM (linear)	70/30	Não	Não	0,590	0,580	0,590	[ 96 30 10 6] [ 36 73 24 10] [ 7 26 117 3] [ 20 17 24 20]
SVM (sigmoid)	70/30	Não	Não	0,586	0,576	0,586	[ 94 31 12 5] [ 38 73 22 10] [ 8 24 118 3] [ 21 16 25 19]
SVM (radial) c/ Search Grid*	70/30	Não	Sim	0,551	0,536	0,551	[[ 85 31 14 7] [ 37 71 27 11] [ 9 23 112 8] [ 27 15 24 18]]
RF com Search Grid*	70/30	Sim	Sim	0,559	0,549	0,559	[[ 88 35 7 7] [ 34 80 17 15] [ 13 27 104 8] [ 22 22 22 18]]
RF	70/30	Sim	Sim	0,543	0,533	0,543	[ 81 33 16 12] [ 33 77 22 11] [ 12 29 105 7] [ 22 19 21 19]
NB	60/40	Não	Não	0,552	0,613	0,552	[ 148 21 23 0] [ 72 45 64 0] [ 18 9 188 0] [ 33 12 58 1]
NB com Search Grid*	60/40	Sim	Não	0,556	0,530	0,556	[[126 34 16 13] [ 51 69 54 14] [ 15 14 173 7] [ 32 14 43 17]]
NB	60/40	Sim	Não	0,552	0,585	0,552	[ 143 28 21 0] [ 66 49 65 1] [ 21 10 184 0] [ 38 14 46 6]

Figura 13 – Resultados mais relevantes para análise de emoções

Dessa forma, o presente estudo confirma que os algoritmos tradicionais de classificação, apesar de robustos, não devem ser o meio mais eficiente para classificação de mensagens de redes sociais, principalmente quando a base usada para treino é pequena.

Considerando as limitações do projeto, o estudo encontrou tanto para análise de sentimentos quanto para análise de emoções um modelo usando SVM, ambos usando *kernel* linear, em que para sentimentos não se removeu *Stop Words* da base, não se aplicou SMOTE e sua proporção de treino e teste foi de 60%/40%, com uma acurácia obtida de 0,65, enquanto para emoções usou-se a mesma configuração de técnicas com exceção da proporção de treino e teste de 70%/30%, com uma acurácia de 0,59.

Os resultados dessa pesquisa, incluindo a tabela de todos os experimentos realizados durante seu desenvolvimento, a base de dados rotulada e os modelos mais assertivos em notebooks do Google Colab, podem ser acessados no GitHub através do link: <[https://github.com/JoaoPedroAlves03/Analise\\_de\\_sentimentos\\_delivery.git](https://github.com/JoaoPedroAlves03/Analise_de_sentimentos_delivery.git)>.

## 5 Conclusão

A análise e classificação das opiniões de clientes de aplicativos de entrega de alimentos, como Rappi, iFood e Zé Delivery, são consideradas de grande relevância para as empresas do setor, que buscam se manter competitivas no mercado. Decisões estratégicas podem ser mais eficazes quando direcionadas pelas opiniões dos próprios clientes que utilizam o serviço.

Com base nas técnicas de PLN e AM aplicadas a textos de redes sociais ao longo do



desenvolvimento deste projeto, foi construída uma base de dados composta por 1.728 textos rotulados manualmente. Essa base incluiu conteúdos provenientes de diversas redes sociais e foi classificada suas emoções, além de sentimentos, possibilitando um estudo detalhado das técnicas levantadas no escopo dos dados obtidos. Embora os resultados obtidos não tenham sido altamente significativos em termos de métricas, foi possível analisar o comportamento de algoritmos clássicos na manipulação desse tipo de dado.

O modelo que apresentou melhor desempenho para a análise de sentimentos foi o *SVM*, configurado sem a remoção de *StopWords*, sem o uso de *SMOTE*, e com um conjunto de treinamento de 60% e teste de 40%. Esse modelo alcançou uma acurácia de 0,65. Para a análise de emoções, o melhor desempenho foi obtido também com o *SVM*, igualmente sem a remoção de *StopWords* e sem o uso de *SMOTE*, mas com uma divisão de 70% para o treinamento e 30% para o teste, alcançando uma acurácia de 0,59.

Considerando a continuidade e o aperfeiçoamento deste projeto, sugere-se:

- Obter-se uma maior quantidade de dados rotulados, podendo estender para mais redes, ou maior proporção de outras se não X e Bluesky;
- Utilização de LLMs para avaliação e aperfeiçoamento de análise de sentimentos e emoções;
- Testas novas técnicas e/ou algoritmos para esse tipo de análise;
- Explorar diferentes aplicativos de entrega de alimentos ou ampliar o escopo para analisar a opinião de clientes em relação a variedade de apps de entrega.

## Referências

ALSAEEDI, A.; KHAN, M. Z. A study on sentiment analysis techniques of twitter data. *International Journal of Advanced Computer Science and Applications*, The Science and Information Organization, v. 10, n. 2, p. 361–374, 2019.

ALVES, J. P.; OLIVEIRA, I. C. de. *A ANÁLISE DE SENTIMENTOS DE MENSAGENS DO TWITTER ENVOLVENDO SERVIÇOS DE ENTREGA DE ALIMENTOS*. 2023. Disponível em: <<http://eventoscopq.mackenzie.br/index.php/jornada/xix/paper/view/3761>>.

ANNISA, R. A.; SETIAWAN, E. B. Aspect based sentiment analysis on twitter using word2vec feature expansion method and gradient boosting decision tree classification method. In: IEEE. *2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT)*. [S.l.], 2022. p. 273–278.

BINSAR, F.; MAURITSIUS, T. Mining of social media on covid-19 big data infodemic in indonesia. *J. Comput. Sci*, v. 16, n. 11, p. 1598–1609, 2020.

BRANDAO, R. *Praticidade deve seguir impulsionando delivery no pós-pandemia, diz pesquisa*. 2021. Disponível em: <<https://valor.globo.com/empresas/noticia/2021/07/06/praticidade-deve-seguir-impulsionando-delivery-no-pos-pandemia-diz-pesquisa.ghtml>>.

- DHIR, R.; RAJ, A. Movie success prediction using machine learning algorithms and their comparison. In: *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. IEEE, 2018. Disponível em: <<https://doi.org/10.1109/icsccc.2018.8703320>>.
- FACELI, K. et al. *Inteligência artificial: uma abordagem de aprendizado de máquina*. [S.l.]: LTC, 2021.
- FAISAL, M. S. et al. Prediction of movie quality via adaptive voting classifier. *IEEE Access*, Institute of Electrical and Electronics Engineers (IEEE), v. 10, p. 81581–81596, 2022. Disponível em: <<https://doi.org/10.1109/access.2022.3195228>>.
- FLORÊNCIO, J. C. P. *Análise e previsão de bilheterias de filmes*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2016.
- GALLAGHER, C.; FUREY, E.; CURRAN, K. The application of sentiment analysis and text analytics to customer experience reviews to understand what customers are really saying. *International Journal of Data Warehousing and Mining (IJDWM)*, IGI Global, v. 15, n. 4, p. 21–47, 2019.
- GHOSH, K. et al. Imbalanced twitter sentiment analysis using minority oversampling. In: IEEE. *2019 IEEE 10th international conference on awareness science and technology (iCAST)*. [S.l.], 2019. p. 1–5.
- LI, H. Using machine learning forecasts movie revenue. In: *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*. IEEE, 2021. Disponível em: <<https://doi.org/10.1109/icaice54393.2021.00094>>.
- NURDENI, D. A.; BUDI, I.; SANTOSO, A. B. Sentiment analysis on covid19 vaccines in indonesia: from the perspective of sinovac and pfizer. In: IEEE. *2021 3rd East Indonesia conference on computer and information technology (EIConCIT)*. [S.l.], 2021. p. 122–127.
- NYOTO, R. L. V.; RULDEVIYANI, Y. Infiltration wells program in jakarta: Twitter sentiment analysis. In: IEEE. *2022 1st International Conference on Information System & Information Technology (ICISIT)*. [S.l.], 2022. p. 352–357.
- OLUSEGUN, R. et al. Text mining and emotion classification on monkeypox twitter dataset: A deep learning-natural language processing (nlp) approach. *IEEE Access*, IEEE, v. 11, p. 49882–49894, 2023.
- PEIXOTO, L. H. R. *Aprendizado de Máquina Aplicado no Atendimento de Reclamações de Clientes*. Tese (Doutorado) — Universidade de São Paulo, 2021.
- PRIBADI, M. R. et al. Improving the accuracy of text classification using the over sampling technique in the case of sinovac vaccine. In: IEEE. *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. [S.l.], 2022. p. 106–110.
- RANE, A.; KUMAR, A. Sentiment classification system of twitter data for us airline service analysis. In: IEEE. *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*. [S.l.], 2018. v. 1, p. 769–773.
- RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, Elsevier, v. 89, p. 14–46, 2015.

ROSSI, R. H. P. S. *Análise de sentimentos para o auxílio na gestão das cidades inteligentes*. Tese (Doutorado), 2019. Disponível em: <<https://doi.org/10.11606/t.3.2019.tde-16092019-110422>>.

SAYYIDAH, A. et al. Effect of j&t express lead time and delivery price on customer satisfaction during the covid-19 pandemic. *Advances in Transportation and Logistics Research*, v. 4, p. 179–186, 2021.

SUMATHY, K.; CHIDAMBARAM, M. Text mining: concepts, applications, tools and issues-an overview. *International Journal of Computer Applications*, Citeseer, v. 80, n. 4, 2013.

VILLARROEL, R. G. *O uso da análise de sentimentos como ferramenta de apoio à gestão acadêmica*. Tese (Doutorado) — Universidade de São Paulo, 2020.

YUGOSHI, I. P. M. *Mineração de opiniões baseada em aspectos para revisões de produtos e serviços*. Tese (Doutorado), 2018. Disponível em: <<https://doi.org/10.11606/t.55.2018.tde-17102018-112458>>.

ZAHOOR, S.; ROHILLA, R. Twitter sentiment analysis using machine learning algorithms: a case study. In: IEEE. *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*. [S.l.], 2020. p. 194–199.