# A ANÁLISE DE SENTIMENTOS DE MENSAGENS DO TWITTER ENVOLVENDO SERVIÇOS DE ENTREGA DE ALIMENTOS

João Pedro Rodrigues Alves (IC), Ivan Carlos Alcantara de Oliveira (Orientador)

**Apoio:PIBIC Mackenzie**

## RESUMO

Considerando a pandemia de COVID 19, houve uma expansão significativa nos serviços de entrega, especialmente o de alimentos, o que tornou relevante a investigação da opinião de clientes desses serviços. Baseado nisso, este trabalho tem por objetivo investigar técnicas de mineração de texto e Aprendizado de Máquina (AM) na análise de sentimentos em mensagens de opiniões publicadas no Twitter sobre os serviços dos aplicativos Rappi, iFood e Zé Delivery. Para alcançar esse objetivo, foi realizada uma revisão da literatura; uma coleta de dados no Twitter, rotulados manualmente; feito um pré-processamento nos dados obtidos, com técnicas de PLN, na tentativa de equilibrar a base e incrementar a eficácia dos resultados; e realizados experimentos com as técnicas de AM selecionadas da literatura: *SVM*, *K-NN*, *Random Forest* e *Naive Bayes*. Como resultados, obteve-se: um *dataset* final rotulado com 958 dados; um estudo comparativo entre combinações de técnicas de PLN e AM; e um modelo publicado no GitHub com uso do Naive Bayes.

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

## ABSTRACT

*Considering the COVID-19 pandemic, there has been a significant expansion in delivery services, especially in the food sector, making it relevant to investigate customer opinions of these services. Based on this, this study aims to explore text mining and Machine Learning (ML) techniques in sentiment analysis of opinions posted on Twitter about the services of the Rappi, iFood, and Zé Delivery apps. To achieve this goal, a literature review was conducted, data was collected on Twitter and manually labeled, preprocessing was performed on the obtained data using NLP techniques in an attempt to balance the dataset and enhance result effectiveness, and experiments were conducted with selected ML techniques from the literature: SVM, K-NN, Random Forest, and Naive Bayes. Results include a final labeled dataset with 958 entries, a comparative study between combinations of NLP and ML techniques, and a model published on GitHub using Naive Bayes.*

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

## 1. INTRODUÇÃO

Os serviços de entrega (*delivery*) de alimentos estavam em crescente relevância na última década, porém a pandemia causada pela COVID-19 alavancou a popularização e a relevância desse tipo de serviço. Segundo a pesquisa de Brandão (2021), 80% dos brasileiros pretendiam continuar usando *delivery* no pós-pandemia, além disso 58% revelaram valorizar a qualidade de entrega. Com isso, para continuar a manter os clientes engajados, fornecer qualidade do serviço deve ser um objetivo para as empresas do ramo (SAYYIDAH, 2021). A partir disso, uma estratégia para se manter no mercado e melhorar a qualidade dos serviços é ouvir a opinião dos clientes.

Segundo Rossi (2019), a exploração de dados urbanos auxilia a análise para tomada de decisões e o estabelecimento de serviços inteligentes, que é um dos pré-requisitos de uma Cidade Inteligente. Hoje em dia uma grande variedade de recursos digitais, ou inteligentes, podem capturar aspectos sociais em cidades assim. Nessa linha, este projeto propõe a investigação a respeito do sentimento da opinião dos usuários dos serviços de *delivery* atrelada a sua publicação em redes sociais e, a partir disso, fornecer elementos para tomada de decisão (YUGOSH, 2018; CONOR, EOGHAN, KEVIN, 2019; SAYYIDAH, 2021). Uma rede social de alta popularidade e que geralmente é utilizada pelos clientes para publicar suas opiniões é o Twitter. Por esse fato, o Twitter foi selecionado para captura de opiniões publicadas e avaliação de alguns serviços de entrega.

Com base no exposto, o problema de pesquisa explorado neste projeto é pautado na seguinte pergunta: Qual abordagem em mineração de texto e aprendizado de máquina é mais assertiva para avaliação de sentimentos em mensagens de opiniões publicadas no Twitter de serviços de entrega de alimentos?

A seguir, este artigo irá discorrer sobre o referencial teórico, apresentando, de forma resumida, técnicas, métodos e algoritmos que foram usados na pesquisa, utilizando-se da literatura de apoio levantada pelo pesquisador, e que foram base para nortear este projeto. Em seguida, encontra-se a metodologia usada durante o desenvolvimento do projeto com métodos usados em sua ordem cronológica. Depois, são mostrados os resultados obtidos, enfatizando as suas contribuições. Enfim, na conclusão, é apresentada uma síntese dos resultados, uma breve discussão e apontadas propostas para trabalhos futuros.

## 2. REFERENCIAL TEÓRICO

Neste tópico, será abordada uma síntese do referencial teórico base deste projeto. Cada subtópico irá apresentar os conceitos e assuntos derivados relevantes que foram estudados para atingir os objetivos desta pesquisa.

### 2.1. MINERAÇÃO DE TEXTO

Segundo Peixoto (2021), textos podem ser considerados dados não-estruturados e representam a maioria dos dados disponíveis hoje em dia, isso devido ao grande avanço de tecnologias sociais que popularizaram os meios digitais e produzem por dia uma quantidade imensa de dados, sobretudo no formato de texto. Além disso, o autor aponta que a natureza não padronizada desse tipo de dado torna sua extração mais complexa do que outros tipos. Dessa maneira, cabe a Mineração de Dados ajudar a extrair informações úteis de dados textuais, consistindo em um conjunto de técnicas matemáticas, estatísticas e computacionais que juntas permitem a análise de seu conteúdo.

#### 2.1.1. MINERAÇÃO DE TEXTO NO TWITTER

O Twitter é uma rede social muito popular de postagens curtas na qual, até o momento da escrita deste texto, permite publicações de até 280 caracteres, em que uma diversidade imensa de pessoas posta sobre qualquer aspecto de suas vidas, entre eles opiniões, reclamações ou indignação sobre os serviços que usaram. Esses fatos colaboram para seu uso neste projeto, pois sua natureza permite a análise de opinião de seus usuários de forma fácil e sucinta (SHELAR, HUANG, 2018; ALSAEEDI, KHAN, 2019).

#### 2.1.2. PROCESSAMENTO DE LINGUAGEM NATURAL

A partir de dados textuais, técnicas de Processamento de Linguagem Natural (PLN) são aplicadas para limpar os ruídos e informações desnecessárias, fazendo um pré-processamento e preparando esses dados para a análise. PLN é o estudo apropriado da forma da linguagem humana para que os computadores possam entender de maneira similar aos seres humanos, tendo como foco um processamento que obtenha informações possíveis de serem manipuladas por algoritmos. Sumathy e Chidambaram (2013) destacam que PLN faz parte do campo interdisciplinar da Mineração de Texto, podendo ser considerado um meio para alcançar a base de informações necessária para iniciar de fato uma análise.

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

**Quadro 1. Técnicas de pré-processamento de linguagem natural**

| | |
|---|---|
| *Tokenization* | Segmenta todo o texto em palavras ou *tokens*, ou seja, as palavras em cada frase são separadas individualmente (*token*) e todos os caracteres especiais, incluindo *links*, pontuações e acentos, são retirados por serem desnecessários para 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 permaneçam neles. |
| *Stemming* | Simplifica o texto recuperando palavras para a sua forma raiz, reduzindo flexões e derivações, facilitando o entendimento básico dos dados obtidos. |

Fonte: Elaborado pelo autor, adaptado de Sumathy e Chidambaram (2013), Ravi K. e Ravi V. (2015), Rossi (2019), Faceli (2021), Nurdeni, Budi e Santoso (2021) e Pribadi et al. (2022).

Ainda na etapa de pré-processamento de dados, segundo Peixoto (2021), os dados passarão para o modelo de representação, tornando dados textuais em representações numéricas. Entre os modelos de representação existentes pode ser citado o *Term Frequency-Inverse Document Frequency* (TF-IDF). O TF-IDF é uma técnica matemática usada para calcular a importância de um termo em relação a sua ocorrência em um conjunto de documentos/textos. Ela consiste em converter textos em modelos vetoriais, sendo uma medida estatística frequentemente usada como um peso na mineração de dados (ROSSI, 2019; PEIXOTO, 2021).

Outra técnica de PLN que pode ser usada no contexto de mineração de textos é o algoritmo de aprendizagem profundo (*Deep Learning*) Doc2Vec. Ele é usado para obter características relevantes, extraindo vetores de características numéricas de documentos de textos (FACELI, 2021).

## 2.2. ANÁLISE DESCRITIVA

Antes de entrar na fase de análise preditiva, aplicando-se os algoritmos de Aprendizagem de Máquina, é preciso extrair algumas informações iniciais dos dados para poder usá-las na próxima etapa. Uma análise exploratória ou estatística descritiva pode ser feita para resumir características principais dos dados obtidos, visando técnicas quantitativas, para descobrir tendencias e possíveis variações estranhas. Esses novos dados descobertos serão importantes para escolher modelos preditivos apropriados e podem influenciar diretamente a aplicação de técnicas especificas. (GALLAGHER, FUREY, CURRAN, 2019; VILLARROEL, 2020; SAYYIDAH, 2021)

Nessa etapa, foi descoberta uma desproporção de cada classe em relação ao total de dados, ou seja, foi observado um desbalanceamento da base de dados, o que pode causar vieses no modelo gerado. Uma possível solução é aplicar a técnica de *Oversampling*, que consiste em equilibrar a proporção da classe minoritária com a classe majoritária da base (GHOSH et al., 2019; PRIBADI et al., 2022). Um dos algoritmos mais comuns usados para *Oversampling* é o *Synthetic Minority Over-sampling Technique* (SMOTE), um método que se utiliza do algoritmo *K-Nearest neighbors* (KNN) para gerar instancias sintéticas do conjunto de dados da classe minoritária a fim de equilibrá-la com a classe majoritária. (OLUSEGUN et al., 2023; ANNISA, SETIAWAN, 2022; NYOTO, RULDEVIYANI, 2022)

## 2.3. APRENDIZAGEM DE MÁQUINA

Aprendizagem de Máquina (AM) é uma subárea da Inteligência Artificial (IA) que envolve técnicas e algoritmos para extração de padrões pela construção de sistemas que possibilitam as máquinas adquirirem conhecimento automaticamente. Ele acumula experiências de soluções ótimas e a partir disso toma decisões, podendo ser preditivo, descritivo ou ambos, tendo como objetivo aprender e padronizar, com base em experiencias extraídas de uma amostra de dados, a fim de alcançar o objetivo da análise em questão (YUGOSH, 2018; PEIXOTO, 2021).

Yugosh (2018) aponta que os algoritmos de aprendizado supervisionado aprendem utilizando exemplos rotulados, determinados como conjunto de treinamento, para fazer uma função de mapeamento dependendo do tipo de tributo alvo. Essa base rotulada guia a aprendizagem supervisionada e viabiliza a concepção de modelos úteis para classificação de sentimentos, apresentando bom desempenho se tiver uma boa base de dados e a escolha correta de técnicas (ROSSI, 2019).

Para definir a qualidade dos modelos aplicados são utilizadas métricas de desempenho que determinam o nível de classificações equivocadas inferidas pelo algoritmo. Antes de aplicar as métricas é construída uma Matriz de Confusão, conforme Quadro 2 (VILLARROEL, 2020; NURDENI, BUTI, SANTOSO, 2021; ANNISA, SETIAWAN, 2022; FAISAL et al., 2022).

**Quadro 2. Matriz de Confusão genérica para duas classes**

| | | Predito | |
|---|---|---|---|
| | | Positivo | Negativo |
| Original | Positivo | VP | FN |
| | Negativo | FP | VN |

Fonte: Elaborado pelo autor.

A partir da Matriz de Confusão, pode-se então calcular as métricas de desempenho dos modelos, dentre elas pode-se citar: Acurácia, Precisão e Revocação. Elas são usadas para avaliar o desempenho de modelos de classificação, e sua interpretação conjunta é essencial para escolher o modelo preditivo mais eficiente (VILLARROEL 2020; BINSAR, MAURITSIUS, 2020; NURDENI, BUTI, SANTOSO, 2021; ANNISA, SETIAWAN, 2022; FAISAL et al., 2022).

Neste trabalho, foi realizado um estudo de técnicas de AM, envolvendo a análise de sentimentos, a classificação dos dados obtidos para extrair as informações de mensagens dos clientes relativos à entrega de alimentos e a identificação de como eles se sentem quanto aos serviços prestados. A seguir, tendo por base os trabalhos relacionados como mecanismos de experimentos para este projeto, é feita uma breve apresentação dos algoritmos de AM selecionados.

- ***Naive Bayes* (NB):** é um algoritmo probabilístico que calcula um conjunto de probabilidades a partir da frequência e combinações de dados, utilizando o Teorema de Bayes como base. Durante o treinamento, são calculadas as probabilidades de cada atributo em relação às classes, e durante o teste, são calculadas as probabilidades de cada exemplo não visto, considerando aqueles com maior probabilidade de ocorrência, com base no treinamento. Além disso, ele é considerado um classificador ingênuo, pois assume que todos os atributos são independentes e não exercem influência uns sobre os outros (YUGOSH, 2018; ALSAEEDI, KHAN, 2019; ROSSI, 2019; LI, 2021).
- ***Support Vector Machine* (SVM):** conhecido por executar bem em análise de sentimentos, é um classificador que analisa informações e caracteriza os limites de escolha, separando informações importantes em vetores. Cada dado é classificado em uma classe, e a máquina identifica a fronteira entre elas, reduzindo as escolhas ambíguas. A melhor separação é determinada pela comparação da distância do hiperplano, que é a linha que separa as classes, com a maior margem (RAVI K., RAVI V., 2017; DHIR, RAJ, 2018; ROSSI, 2019).
- ***K Nearest Neighbors* (KNN):** baseia-se na ideia de que a classificação de uma instância será semelhante àquelas próximas a ela no espaço vetorial, daí o termo "vizinho mais próximo". Ou seja, ele identifica o conjunto de registros que estão próximos em termos de alguma característica. Pode ser eficaz em classificação de sentimentos, pois os pesos dos conjuntos representam a relevância dos diferentes conjuntos de recursos, ao invés de atribuir a cada característica individualmente (DHIR, RAJ, 2018; GALLAGHER, FUREY, CURRAN, 2019; ROSSI, 2019).

- ***Random Forest* (RF):** dentre as muitas aplicações, vem sendo usado para classificação de opiniões no Twitter e é visto como uma evolução do algoritmo *Decision Tree* (DT), em que várias árvores são treinadas simultaneamente. Por isso, é considerado um método baseado em *Ensemble Learning* (Aprendizado por Conjunto), que combina vários modelos para treiná-los para a mesma tarefa. Cada árvore é treinada com um subconjunto dos dados de treinamento. Isso permite superar a limitação das árvores de decisão tradicionais, que podem sofrer *overfitting*, devido à tendência de crescerem muito em profundidade (FLORÊNCIO, 2016; ROSSI, 2019, DHIR, RAJ, 2018; RANE, KUMAR, 2018; BINSAR, MAURITSIUS, 2020; ZAHOOR, ROHILLA, 2020).

## 2.4. ANÁLISE DE SENTIMENTOS

Segundo Peixoto (2021), a satisfação do cliente está diretamente ligada ao seu sentimento sobre aquilo consumido e a relação dessa experiência com suas expectativas. Então, para aumentar essa satisfação é necessário descobrir o que o cliente espera de seus serviços. Por isso, analisar esse sentimento pode fornecer dados uteis para tomada de decisão.

Yugoshi (2018) define a Análise de Sentimentos (AS), ou Mineração de Opiniões, como um estudo de emoções das pessoas sobre o serviço ao qual está sendo analisado e aponta também uma crescente divulgação de opiniões em redes sociais como o Twitter.

De acordo com Rossi (2019), se um texto puder ser determinado como subjetivo, uma opinião detém cinco aspectos: nome da entidade alvo do comentário, aspecto da entidade (não obrigatório), polaridade do sentimento, detentor/fonte do respectivo sentimento e o instante em que a opinião foi exposta. Rossi (2019) ainda define polaridade como o grau da opinião expressada sendo positiva, neutra ou negativa (classes do tipo ternária), podendo ser representadas por intervalos numéricos ou simplesmente pela categorização de um dos três tipos possíveis.

Villarroel (2020) separa AS em três etapas: identificar opiniões (usando os cinco conceitos descritos por Rossi (2019) anteriormente), classificar a polaridade, e a sumarização, que consiste em organizar os resultados de forma sumarizada. Para atingir a análise pretendida, tem-se diferentes abordagens para análise de sentimentos, entre elas, a estatística e a por AM.

Na abordagem estatística pode-se utilizar das informações obtidas na etapa de análise descritiva para produção de nuvens de palavras, que representa a ocorrência e intensidade de palavras em determinado texto; e gráficos, de variados tipos (como

de barras e pizza), que possibilitam a visualização de informações de formas convenientes. Já a abordagem por AM envolve o uso de uma ou mais técnicas, similares as anteriormente apresentadas. (VILLARROEL, 2020).

## 2.5. TRABALHOS RELACIONADOS

Após a revisão de literatura, foram selecionados 23 trabalhos relacionados, que compõem o referencial teórico explicitado neste texto e que serviram de norte para o desenvolvimento deste projeto. O Quadro 3 sintetiza a quantidade de trabalhos que exploram algumas técnicas ou algoritmos encontrados, que foram selecionados como mecanismos de experimentos desta pesquisa.

**Quadro 3. Relação de uso de métodos dos trabalhos relacionados**

| Técnica/Algoritmo | Quantidade de trabalhos |
|---|---|
| Dados do Twitter | 14 |
| PLN | 16 |
| TF-IDF | 6 |
| Doc2vec | 4 |
| SMOTE | 5 |
| KNN | 7 |
| SVM | 16 |
| *Naive Bayes* | 14 |
| *Random Forest* | 9 |
| Análise de Sentimentos | 18 |
| Acurácia | 21 |
| Precisão | 18 |
| Revocação | 18 |
| Matriz de Confusão | 11 |

Fonte: elaborado pelo autor.

## 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:

- **Revisão da Literatura**: apuração dos textos previamente selecionados e de mais trabalhos pertinentes para auxiliar nesta pesquisa, obtidos nas bases de dados: Google Scholar, Universidade de São Paulo (USP), Universidade Federal de Pernambuco (UFP), *Science Direct* e IEEE *Xplore* (banco de dados digital de pesquisa do *Institute of Electrical and Electronics Engineers* - IEEE);
- **Investigação das Técnicas**: estudo e análise de mineração de texto, PLN e AM utilizando-se da literatura pesquisada e com enfoque no tema deste projeto;

- **Coleta de Dados**: elaboração de um software que coleta dados por meio de mineração de texto no Twitter, utilizando-se de sua própria API e bibliotecas específicas do Python. Nessa coleta, foram considerados os dados dos serviços de entrega: Rappi, iFood e Zé delivery, montado um *dataset* bruto que foi rotulado manualmente com os sentimentos: positivo, negativo ou neutro;
- **Análise Descritiva**: elaboração de uma análise estatística descritiva dos dados brutos coletados, com a finalidade de encontrar informações relevantes a seu respeito, com nuvem de palavras, tabelas e gráficos de pizza e barras;
- **Pré-Processamento de Dados**: aplicação de técnicas de PLN, eliminando redundância, ambiguidade, *outliers* e limpando irregularidades, a fim de facilitar o processo de análise, utilizando-se das técnicas de *StopWords*, *Tokenization*, *Stemization* e *Case Folding*, além de SMOTE e Doc2Vec, obtendo dessa forma um *dataset* preparado;
- **Aprendizagem de Máquina**: tendo por base as técnicas encontradas nos trabalhos relacionados, a saber: KNN, SVM, *Naive Bayes* e *Random Forest*, seus algoritmos foram experimentados no *dataset* preparado, sendo dividido em treino e teste considerando os percentuais: 60, 40; 65, 35; 70,30; 75, 25; e 80, 20; e os resultados obtidos foram documentados com as métricas de desempenho: acurácia, precisão, revocação e matriz de confusão;
- **Seleção do Modelo**: baseado nos resultados obtidos com os experimentos e suas métricas, comparações foram realizadas, a fim de obter aquele modelo preditivo com o melhor resultado.
- **Publicação no GitHub**: a documentação feita, o *dataset* bruto/final, o código fonte e os resultados encontrados foram publicados no GitHub no endereço: https://github.com/JoaoPedroAlves03/Analise_de_sentimentos_delivery.git. Com o código, é possível realizar mais experimentos e incluir dados no *dataset*, fazendo uso de um notebook do Google Colab, como usado neste projeto, ou do Jupyter, em Python e suas bibliotecas para análise de dados e predição.

## 4. RESULTADOS E DISCUSSÃO

Este projeto produziu como resultados e contribuições um *dataset* rotulado de *tweets* sobre serviços de *delivery* de alimentos, um estudo comparativo da aplicação de diferentes algoritmos e técnicas para predição nesse contexto e um modelo preditivo obtido com a técnica mais assertiva. Para a análise deste projeto, foram escolhidos três dos serviços de entrega de alimentos mais populares no Brasil: iFood, Rappi e Zé Delivery. Uma síntese desses resultados será apresentada nas subseções a seguir.

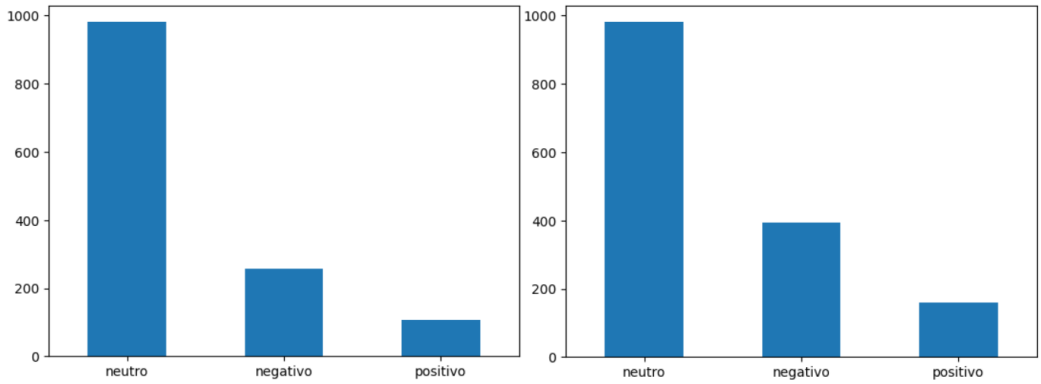
## 4.1. DATASET

Utilizando-se da API do Twitter, foram feitas coletas de dados da rede social, por meio da biblioteca Tweepy. Cada serviço teve uma busca específica com o *token* de citação direta do seu nome, dessa forma identificando publicações que estavam comentando sobre ele. Foram coletados como variáveis: o nome da conta que publicou, a data publicada, o conteúdo da publicação e o serviço que estava sendo referenciado. Para os fins da análise, foram mantidos apenas o *tweet* e o serviço referenciado no *dataset* para o desenvolvimento das análises desejadas.

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 *tweets* coletados e classificado como uma opinião positiva, negativa ou neutra. Como critério para classificação, foi estabelecido inicialmente que a classe positiva e negativa seriam os *tweets* que “expressavam alguma opinião aparente sobre o serviço prestado”, seja positiva ou negativa, e a classe neutro seriam os *tweets* que “não expressavam opinião aparente ou não tinham relação direta com o serviço”.

Dessa forma, primeiramente foi obtida uma base com 1346 dados. Porém, foi observado um grande desequilíbrio na proporção das classes, obtendo-se muito mais do que o dobro de neutro em comparação a soma de negativos e positivos. Então, foi realizada uma nova coleta, obtendo-se mais de 1500 dados. Na fase de rotulação desses novos dados, foram excluídos os neutros e incluído apenas os classificados como positivos ou negativos, obtendo-se um novo total de 1537 dados, ilustrado à direita na imagem 1.

**Imagem 1. Gráficos de comparação da distribuição de classes com a 1ª coleta (à esquerda) e com o complemento da 2ª coleta (à direita).**




Fonte: elaborado pelo autor.

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 *hastags* (que começam com #), o marcador de *retweet* (a *string* RT isolada), *links* de sites, converte os textos completamente para letras minusculas, 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 portugês 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. Outra técnica aplicada para testes foi o algoritmo Doc2Vec, em busca de suprir possiveis deficiencias que o TF-IDF pudesse apresentar. Já pensando no desequilibrio que a base apresentaria, a técnica de *Oversampling* foi levantada com o uso do método SMOTE, que gerou novas amostras sintéticas para equilibrar o *dataset*.

Alguns experimentos iniciais não foram satisfatórios com SVM e NB, indicando a necessidade de sua reavaliação. Então, foi ponderado o tamanho do desiquilibrio do *dataset* e a criação de dados com o SMOTE. Por esse motivo, uma revisão mais aprofundada na rotução do *dataset* foi realizada. Nessa investigação, foi observado que alguns *Tweets* não pertenciam ao contexto deste trabalho e foram eliminados. Exemplos deles podem ser observados no Quadro 4.

**Quadro 4. Exemplos de *tweets* que foram removidos da base**

| Exemplos de *Tweets* | Motivo da retirada |
|---|---|
| "A Agência Pública teve acesso a um contrato do iFood com terceirizada que prevê escala e turno de entregadores e até mesmo direitos trabalhistas não cumpridos. Nele, a contratada se compromete a “isentar” app de processos na Justiça. #Arquivo https://t.co/ITqe0yy3nT" | Notícia que cita o iFood |
| “@jp_barbosa João, Pedimos desculpas por isso. Poderia nos chamar na DM, por favor? Estamos à disposição para ajudar!” | Mensagem do próprio iFood respondendo um usuário |
| “Gente vcs tbm já notaram como o tamanho do Rappi 10 tá menor???” | Comentário que se refere a marca de alimentos “Rap10”, mas que confunde a escrita com o nome do aplicativo Rappi |
| “@newscolina Jogou o suficiente para nunca mais o Zé delivery usar a camisa do Vasco” | Comentário que se refere a um jogador de futebol apelidado de “Zé Delivery” |

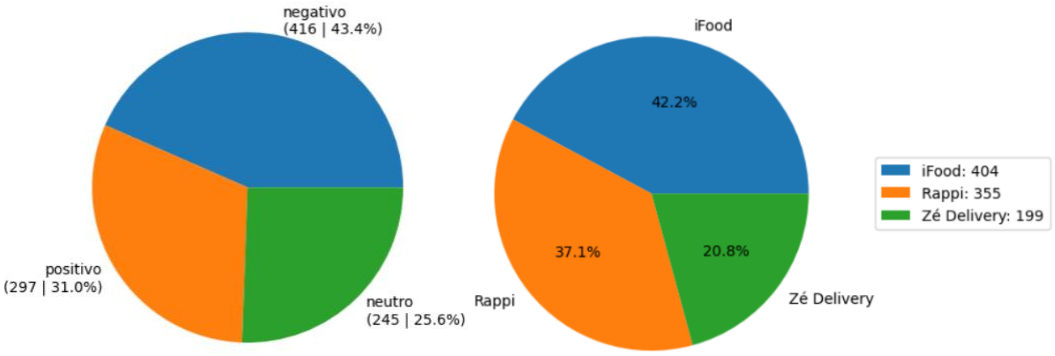
Fonte: elaborado pelo autor.

Todos os exemplos mostrados no Quadro 4 são *tweets* que foram capturados para a base de dados, mas que não apresentam um cliente relando sua experiência

com algum dos aplicativos de *delivery* selecionados. Por isso, eles e mais uma quantidade significativa foram removidos da base de dados. Além das remoções, uma reclassificação foi feita nessa revisão mais minuciosa de cada dado presente na base.

Ao final dessa revisão dos dados, o *dataset* resultante passou para um total de 958 dados, mas com uma quantidade um pouco mais proporcional de valores positivos, negativos e neutros (Imagem 2).

**Imagem 2. Parcela representativa de cada sentimento e de cada aplicativo**




Fonte: elaborado pelo autor.

Nessa imagem 2, observa-se que a classificação neutra virou minoria, e no geral temos uma distribuição com menor desbalanceamento do que antes. Também é interessante observar, no segundo gráfico (à direita), a porcentagem representativa de cada aplicativo na base obtida.

Uma alternativa considerada, como forma de aumentar os dados, foi uma nova coleta no Twitter. Porém, na metade desta pesquisa, o Twitter passou por mudanças internas, alterndo o modo de captura de dados com sua API, sendo necessário o pagamento de um valor mensal elevado, inviabilizando essa opção. Outra opção, seria o uso de *web scraping* nas páginas do Twitter, mas devido ao consumo de tempo para o seu desenvolvimento, foi desconsiderada.

A base completa, estava finalizada e pronta para ser aplicada nas técnicas de pré-processamento, que já tinham sido desenvolvidas. Desse modo, foi realizada uma análise descritiva do *dataset* produzido, para observar características gerais da base. A Imagem 3 apresenta uma nuvem de palavras que mostra as palavras mais usadas nos *tweets* capturados na base de dados atualizada.

Por fim, o Quadro 5 apresenta um comparativo das parcelas representativas das classificações de sentimento de cada serviço, mostrando o valor absoluto em número de dados e a porcentagem relativa de cada um deles individualmente, sendo possível comparar diretamente as quantidades do *dataset* final obtido.

**Imagem 3. Nuvem de palavras da base de *tweets***


Fonte: elaborado pelo autor

**Quadro 5. Parcela representativa da classificação de cada aplicativo**

| | | iFood | Rappi | Zé Delivery |
|---|---|---|---|---|
| Positivo | Porcentagem | 26% | 24,5% | 52,8% |
| | Valor absoluto | 105 | 87 | 105 |
| Negativo | Porcentagem | 37,1% | 61,4% | 24,1% |
| | Valor absoluto | 150 | 218 | 48 |
| Neutro | Porcentagem | 36,9% | 14,1% | 23,1% |
| | Valor absoluto | 149 | 50 | 46 |

Fonte: elaborado pelo autor

Observando-se os dados obtidos com essa análise, pode-se refletir um pouco sobre a situação de cada serviço. O iFood é o que possui maior equilíbrio de classes, o que tem maior representação e o que apresenta uma ordem de classificação esperada, visto que os negativos foram maioria. Acredita-se que a maioria das pessoas que se dispõem a comentar nas redes sociais sobre experiencia com algum produto é porque deseja fazer uma reclamação, o que também explica a minoria positiva. Pode-se deduzir que o iFood é o mais popular dentre os três, não somente por ter sido encontrado mais dados, mas também pelo equilíbrio, que pode significar uma estabilidade boa de funcionamento dos serviços da empresa.

Em relação ao Rappi, observa-se um elevado número de classificações negativas, o que leva a supor que embora ainda bastante usado, apresenta problemas, impactando nas reações mais infelizes no Twitter. Já o Zé Delivery, é o único que apresenta o sentimento positivo como maioria, algo bem curioso, que pode representar uma boa satisfação de seus clientes em relação aos seus serviços. Ainda sobre este último, pode-se inferir que o fato de entrega de bebidas, em especial bebidas alcoólicas, deixa as pessoas felizes, o que as incentiva a compartilhar dessa felicidade nas redes sociais, e isso realmente pôde ser observado nos seus *tweets*.

As suposições levantadas acima são apenas algumas observações que o *dataset* obtido pode indicar, porém a quantidade de dados é insuficiente para conseguir afirmar alguma delas, apenas perceber algumas tendencias encontradas no período de captura dos dados.

Por fim, cabe destacar que, um dos objetivos na proposta do projeto era automatizar a coleta e a análise de sentimento, possibilitando esse processo em tempo real, porém, com as mudanças no Twitter, isso se tornou inviável.

## 4.2. MODELO PREDITIVO

A partir do *dataset* final, foi iniciada a aplicação dos algoritmos de AM com o auxílio da biblioteca do Python Scikit-learn. Como explicitado no referencial teórico, foram selecionados quatro dos algoritmos mais citados na literatura levantada, sendo eles KNN, *Naive Bayes*, *Random Forest* e SVM. Todos esses foram aplicados e testados com diferentes combinações de técnicas de PLN, em busca dos melhores resultados. Inicialmente a divisão dos conjuntos de treino e teste foi definida como 60% treino e 40% teste, ficando 574 dados para treino e 384 dados para teste.

Os testes foram conduzidos seguindo combinações de aplicação dos quatro algoritmos com aplicação ou não aplicação de: remoção de Stop Words, remoção de números, SMOTE para equilibrar o conjunto de treino, e Doc2Vec. Após testes com a proporção 60%/40% para treino e teste, respectivamente, foram conduzidos testes com outras proporções destes dois conjuntos, pois como o projeto obteve uma quantidade de dados limitada também se tornou relevante conduzir experimentos com conjuntos de treino maiores, possibilitando que o modelo aprenda melhor.

Foram testadas mais de 100 combinações diferentes, usando-se as proporções 60%/40%, 65%/35%, 70%/30%, 75%/25%, 80%/20% e 85%/15% de treino e teste respectivamente, sendo documentado cada resultado com as respectivas matrizes de confusão e valores de Acurácia, Precisão e Revocação. Não é pertinente apresentar todos esses resultados neste artigo, ao invés disso serão apresentados os resultados mais relevantes obtidos após análise minuciosa das informações obtidas com os experimentos de cada um dos quatro algoritmos usados. O Quadro 6 mostra o melhor resultado obtido para cada um dos algoritmos.

Os melhores resultados não incluíram a aplicação do Doc2Vec, demonstrando-se ineficiente para o caso de estudo específico do projeto. Já o SMOTE performou bem e foi aplicado na maioria dos melhores resultados, apontando sua importância quanto a equilibrar bases com classes desiquilibradas. Observando-se a remoção de StopWords, ficou evidenciado que neste caso é melhor mantê-las, sem removê-las os

resultados melhoraram significativamente, diferente da remoção dos números, que a influência variou dependendo da combinação de técnicas usadas.

**Quadro 6. Melhores resultados obtidos**

| Algorit. | Prop. | remoção SW | remoção Nº | SMOTE | Acur. | Prec. | Rev. | Matriz de confusão |
|---|---|---|---|---|---|---|---|---|
| KNN (k = 15) | 60/40 | não | não | não | 0,63 | 0,63 | 0,63 | 133 8 22<br>40 44 23<br>33 17 64 |
| RF | 70/30 | não | não | sim | 0,63 | 0,63 | 0,63 | 95 12 15<br>21 40 14<br>29 15 47 |
| NB | 80/20 | não | sim | sim | 0,67 | 0,67 | 0,67 | 69 8 8<br>14 27 9<br>10 14 33 |
| SVM (linear) | 60/40 | não | sim | sim | 0,65 | 0,65 | 0,65 | 126 17 20<br>24 52 31<br>23 19 72 |

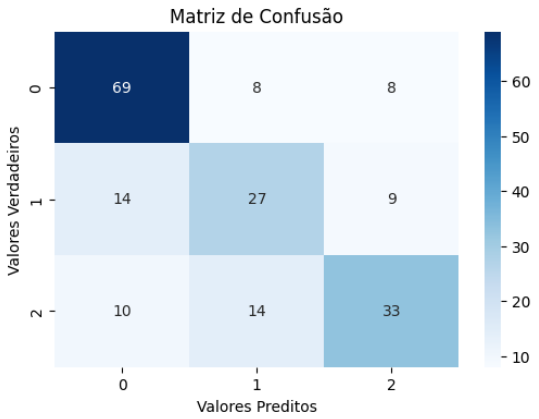
Fonte: elaborado pelo autor.

Observa-se também que os algoritmos mais relevantes encontrados na literatura corresponderam aos dois melhores resultados, SVM – que apareceu em 16 dos 23 trabalhos, e o NB – que apareceu em 14 dos 23 trabalhos. Apesar disso, é importante ressaltar que o *NB* apresentou variações baixas em geral comparando-se com o SVM e o *Random Forest*, que apresentaram resultados mais estáveis. Por outro lado, o KNN foi o que, em geral, se demonstrou pior, tendo a maioria das suas acurácias com valores abaixo de 0,57.

Reparando somente nas acurácias, o NB conseguiu melhor performance. Uma análise minuciosa na sua matriz de confusão foi feita, observando os valores de predição verdadeiros para as três classes e em especial o equilíbrio desses valores, pois não adianta selecionar um modelo que tem uma acurácia alta, mas está, por exemplo, aprendendo muito mais as classes positivo e negativo do que a classe neutra.

Apesar disso, os valores produzidos pelo NB se demonstraram melhores e mais equilibrados. Por essa razão, o modelo selecionado por este projeto aplica o algoritmo NB. Seu melhor resultado, sem remover Stop Words, removendo números, aplicando SMOTE e não aplicando Doc2Vec, com proporção de 80% para treino e 20% para teste, corresponde a melhor combinação testada nos experimentos produzidos por este projeto. O gráfico da matriz de confusão desse modelo é exposto na Imagem 4.

**Imagem 4. Matriz de confusão do modelo com NB melhor performático**




Fonte: elaborado pelo autor.

## 4.3. DISCUSSÃO DOS RESULTADOS

Apesar dos problemas e das mudanças repentinas forçadas por fatores externos, o projeto tem resultados satisfatórios. Mesmo não atingindo as expectativas iniciais, muitos experimentos relevantes foram feitos e documentados. Todos os testes apresentaram acurácias baixas, mesmo o modelo selecionado com 0,67 apenas.

O motivo provável do resultado obtido foi a quantidade insuficiente de dados. Para boas análises de dados, além de dados de qualidade, a quantidade também é muito importante, e 958 dados parece ter sido pouco para os algoritmos conseguirem aprender bem. O ideal era coletar mais dados, mas, mudanças na política de acesso pelos desenvolvedores ao Twitter no meio desta pesquisa inviabilizou essa opção.

Um outro motivo, não muito provável, pode ter relação com o uso de técnicas inadequadas. Todas as técnicas de PLN e de AM usados neste projeto foram baseados em artigos relacionados encontrados na literatura. Porém, outros experimentos poderiam ser realizados com outras técnicas ou combinações delas.

Mesmo assim, este estudo é relevante, produzindo também evidências sobre a performance das técnicas usadas e um *dataset*. O Doc2Vec não se mostrou eficiente em sua versão base, o que indica que é talvez melhor utilizar variações dele. Já o SMOTE apresentou melhoras significativas na eficiência dos modelos, provando que o equilíbrio das classes é essencial para encontrar resultados melhores de acurácia.

A experiencia com remoção ou não de StopWords e números mostrou que eles podem influenciar na performance do modelo, no caso, com a base testada, remover números apresentou melhores resultados. O KNN não se mostrou uma boa opção e o *Naive Bayes* se revelou uma boa alternativa mesmo não sendo possível aplicar o

Doc2Vec, pois o NB produz valores negativos e o Doc2Vec não consegue lidar com essas amostras.

No final deste projeto, era de interesse produzir uma aplicação analítica simples que coletasse os dados do Twitter automaticamente com o melhor modelo preditivo e, em tempo real, indicaria o sentimento. Porém, como citado anteriormente, devido as mudanças no Twitter, essa opção foi descartada.

Destaca-se que os resultados parciais deste trabalho foram apresentados no Workshop de Tendencias Tecnológicas (WTT), da Faculdade de Computação e Informática (FCI) da Universidade Presbiteriana Mackenzie (UPM).

## 5. CONSIDERAÇÕES FINAIS

A relação e classificação dos dados de opiniões de clientes de aplicativos de serviços de entrega de alimentos, como Rappi, iFood e Zé Delivery, é considerada relevante para observação das empresas do ramo, que buscam se manter no mercado. Tomadas de decisão estratégicas podem ser mais eficazes para o seu bem-estar se forem voltadas diretamente a opinião do próprio cliente que está utilizando o serviço.

Com base nas técnicas de pré-processamento, PLN e de AM aplicadas em textos do Twitter ao longo do desenvolvimento deste projeto, foi possível trazer resultados não muito altos na rotulagem dos sentimentos, podendo-se explicar pela quantidade insuficiente de dados.

Apesar disso, o projeto conseguiu fornecer resultados satisfatórios, como o *dataset* final obtido, mesmo que pequeno, e os experimentos realizados com as técnicas KNN, SVM, *Naive Bayes* e *Random Forest*. Baseado no *dataset* final, os experimentos e a questão de pesquisa, a técnica que obteve o melhor resultado foi o NB, com a proporção 80% dos dados para treino e 20% para teste, sem o uso de remoção de StopWords, com remoção de números, balanceamento dos dados com o SMOTE, apresentando acurácia, precisão e revocação iguais a 67%.

Considerando a continuidade e o aperfeiçoamento deste projeto, é proposto:

- Realizar novas coletas de dados no Twitter, talvez por *web scraping*;
- Utilizar outros algoritmos e/ou técnicas para verificar possíveis melhoras;
- Expandir a análise de sentimentos para uma análise de emoções, o que iria aprimorar a relevância dos dados obtidos com o modelo;
- 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.

**AGRADECIMENTOS**

Os autores agradecem ao apoio concedido com a bolsa PIBIC/Mackenzie - CFP no. 287/2022 durante a realização deste projeto.

**REFERÊNCIAS**

ALSAEEDI, Abdullah; KHAN, Mohammad Zubair. A study on sentiment analysis techniques of Twitter data. **International Journal of Advanced Computer Science and Applications**, New York, v. 10, n. 2, p. 361-374, 2019.

ANNISA, Rimdani Alya; SETIAWAN, Erwin Budi. Aspect Based Sentiment Analysis on Twitter Using Word2Vec Feature Expansion Method and Gradient Boosting Decision Tree Classification Method. **2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT)**. IEEE, 2022. p. 273-278.

BINSAR, Faisal; MAURITSIUS, Tuga. Mining of Social Media on Covid-19 Big Data Infodemic in Indonesia. **J. Comput. Sci**, v. 16, n. 11, p. 1598-1609, 2020.

DHIR, Rijul; RAJ, Anand. Movie success prediction using machine learning algorithms and their comparison. **first international conference on secure cyber computing and communication (ICSCCC)**. IEEE, 2018. p. 385-390.

BRANDÃO, Raquel. Praticidade deve seguir impulsionando delivery no pós-pandemia, diz pesquisa. **Valor Econômico**, Globo. São Paulo, 6 de agosto, 2021. Disponível em: https://valor.globo.com/empresas/noticia/2021/07/06/praticidade-deve-seguir-impulsionando-delivery-no-pos-pandemia-diz-pesquisa.ghtml. Acesso em: 27 mar. 2022.

FACELI, Katti et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. . Rio de Janeiro: LTC, 2021. Disponível em: https://repositorio.usp.br/directbitstream/ff933d41-4c3d-4b57-80c2-b4f1c805b1dc/3128493.pdf. Acesso em: 27 mar. 2022.

FAISAL, Muhammad Shahzad et al. Prediction of Movie Quality via Adaptive Voting Classifier. **IEEE Access**, v. 10, p. 81581-81596, ago. 2022. DOI: 10.1109/ACCESS.2022.3195228. Disponível em: https://ieeexplore.ieee.org/document/9845402. Acesso em: 30 maio 2023.

FLORÊNCIO, João Carlos Procópio. **Análise e predição de bilheterias de filmes**. 2016. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Pernambuco, 2016. Disponível em: https://repositorio.ufpe.br/handle/123456789/17639. Acesso em: 20 mar. 2022.

GALLAGHER, Conor; FUREY, Eoghan; CURRAN, Kevin. 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)**, v. 15, n. 4, p. 21-47, 2019.

GHOSH, Kushankur et al. Imbalanced twitter sentiment analysis using minority oversampling. **2019 IEEE 10th international conference on awareness science and technology (iCAST)**. IEEE, 2019. p. 1-5.

LI, Haibo. Using machine learning forecasts movie revenue. **2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)**. IEEE, 2021. p. 455-460.

NURDENI, Deden Ade; BUDI, Indra; SANTOSO, Aris Budi. Sentiment analysis on Covid19 vaccines in Indonesia: from the perspective of Sinovac and Pfizer. **2021 3rd East Indonesia conference on computer and information technology (EIConCIT)**. IEEE, 2021. p. 122-127.

NYOTO, Rebecca La Volla; RULDEVIYANI, Yova. Infiltration Wells Program in Jakarta: Twitter Sentiment Analysis. **2022 1st International Conference on Information System & Information Technology (ICISIT)**. IEEE, 2022. p. 352-357.

OLUSEGUN, Ruth et al. Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach. **IEEE Access**, v. 11, p. 49882-49894, maio 2023. DOI: 10.1109/ACCESS.2023.3277868. Disponível em: https://ieeexplore.ieee.org/document/10129946. Acesso em: 22 mar. 2022.

PEIXOTO, Luiz Henrique Rowan. **Aprendizado de Máquina Aplicado no Atendimento de Reclamações de Clientes**. 2021. Dissertação (Mestrado em Matemática, Estatística e Computação) - Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos, 2021. doi:10.11606/D.55.2021.tde-04022022-172002. Acesso em: 19 mar. 2022.

PRIBADI, Muhammad Rizky et al. Improving the accuracy of text classification using the over sampling technique in the case of sinovac vaccine. **2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)**. IEEE, 2022. p. 106-110.

RANE, Ankita; KUMAR, Anand. Sentiment classification system of twitter data for US airline service analysis. **2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)**. IEEE, 2018. p. 769-773.

RAVI, Kumar; RAVI, Vadlamani. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. **Knowledge-Based Systems**, Netherlands, v. 89, p. 14-46, nov. 2015.

ROSSI, Rosa Helena Peccinini Silva. **Análise de sentimentos para o auxílio na gestão das cidades inteligentes**. 2019. Tese (Doutorado em Sistemas Digitais) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2019. DOI: 10.11606/T.3.2019.tde-16092019-110422. Acesso em: 19 mar. 2022.

SAYYIDAH, Azhar 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.

SHELAR, Amrita; HUANG, Ching-Yu. Sentiment analysis of twitter data. **2018 International Conference on Computational Science and Computational Intelligence (CSCI)**. IEEE, 2018. p. 1301-1302.

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

VILLARROEL, Rosivaldo Gabriel. **O uso da análise de sentimentos como ferramenta de apoio à gestão acadêmica**. 2020. Tese de Mestrado. Universidade de São Paulo.

YUGOSHI, Ivone Penque Matsuno. **Mineração de opiniões baseada em aspectos para revisões de produtos e serviços**. 2018. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos, 2018. doi:10.11606/T.55.2018.tde-17102018-112458. Acesso em: 20 mar 2022.

ZAHOOR, Sheresh; ROHILLA, Rajesh. Twitter sentiment analysis using machine learning algorithms: a case study. **2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)**. IEEE, 2020. p. 194-199.

**Contatos:** 42083605@mackenzista.com.br e ivan.oliveira@mackenzie.br