UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE CIÊNCIAS

DEPARTAMENTO DE COMPUTAÇÃO

CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DANIELE CARNAÚBA GONÇALVES

ANÁLISE DE SENTIMENTOS EM COMMITS DE PROJETOS DE CÓDIGO-FONTE ABERTO NO CONTEXTO DA PANDEMIA DE COVID-19

FORTALEZA

2023

DANIELE CARNAÚBA GONÇALVES

ANÁLISE DE SENTIMENTOS EM COMMITS DE PROJETOS DE CÓDIGO-FONTE ABERTO NO CONTEXTO DA PANDEMIA DE COVID-19

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Orientadora: Prof. Dra. Ticiana Linhares Coelho da Silva.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

G624a Gonçalves, Daniele Carnaúba.
Análise de sentimentos em commits de projetos de código-fonte aberto no contexto da pandemia de Covid-19 / Daniele Carnaúba Gonçalves. – 2023.
43 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Computação, Fortaleza, 2023.
Orientação: Profa. Dra. Ticiana Linhares Coelho da Silva.

1. Análise de sentimentos. 2. Commits. 3. Pandemia. 4. COVID-19. I. Título.

CDD 005

DANIELE CARNAÚBA GONÇALVES

ANÁLISE DE SENTIMENTOS EM COMMITS DE PROJETOS DE CÓDIGO-FONTE ABERTO NO CONTEXTO DA PANDEMIA DE COVID-19

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Aprovada em:

BANCA EXAMINADORA

Prof. Dra. Ticiana Linhares Coelho da Silva (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. José A. Fernandes de Macêdo
Universidade Federal do Ceará (UFC)

Prof. Me. Luís Gustavo Coutinho do Rego
Instituto Federal do Ceará - Campus Quixadá

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

AGRADECIMENTOS

Gostaria de expressar minha sincera gratidão à Prof. Dra. Ticiana, por toda a orientação, apoio e inspiração que proporcionou ao longo da elaboração desse trabalho. Sua dedicação, conhecimento e orientação foram fundamentais para o sucesso deste estudo. Ao longo deste processo, suas orientações valiosas não apenas refinaram o conteúdo do TCC, mas também contribuíram significativamente para o meu crescimento acadêmico e profissional. Seu comprometimento e paixão pelo ensino deixaram uma marca em minha jornada acadêmica.

Obrigada à minha família, que sempre esteve presente nos momentos em que precisei me dedicar ao ensino superior, mesmo estando fisicamente distante. Vocês sempre me apoiaram e aplaudiram meus avanços durante a graduação. Foram meu alicerce, minha fonte de força e motivação ao longo dessa jornada.

Aos professores, agradeço por compartilharem seu conhecimento, experiência e paixão pela aprendizagem. Suas orientações e dedicação contribuíram significativamente para o meu desenvolvimento acadêmico e pessoal. Aos queridos colegas de classe, agradeço pela parceria e amizade ao longo desses anos. Juntos, enfrentamos desafios acadêmicos, celebramos sucessos e construímos memórias que levo comigo para o futuro. O apoio mútuo e a colaboração foram essenciais para o nosso crescimento individual e coletivo. Agradeço à administração e a todos os membros da equipe acadêmica pela dedicação incansável em fornecer uma educação de qualidade e pelo compromisso em incentivar a busca pelo conhecimento.

Meus agradecimentos aos amigos Daniel, Cinthia, Anderson e Misael, companheiros de trabalhos e irmãos na amizade que fizeram parte da minha formação. Tenho a certeza de que essa conexão valiosa continuará a enriquecer minha vida no futuro.

“Se é a razão que faz o homem, é o sentimento que o conduz.” (Jean-Jacques Rousseau)

RESUMO

O estudo apresentado nesta pesquisa inicia-se com a análise dos sentimentos expressos em mensagens escritas por desenvolvedores de software por meio de *commits* em repositórios de código aberto, abrangendo os anos de 2019 a 2021. Esse período engloba o período anterior, durante e após as primeiras grandes ondas da pandemia de COVID-19. Foram examinados sete projetos de código aberto, totalizando 173.993 mensagens de *commit*. A análise foi conduzida utilizando dois modelos de Processamento de Linguagem Natural (PLN) baseados na arquitetura RoBERTa, os quais categorizam as mensagens de *commit* em três grupos: negativo, positivo e neutro. Posteriormente, foi realizada uma investigação da relação entre os sentimentos expressos nas mensagens de *commit* e os anos em que foram escritas, bem como a eficácia de cada modelo. Os resultados deste estudo sugerem que as mensagens tendem a ser mais negativas no início das ondas de COVID-19. Após a pandemia, esse número tende a diminuir. Por outro lado, observa-se um aumento nas mensagens positivas nesse período.

Palavras-chave: Análise de sentimentos, commits, pandemia, COVID-19

ABSTRACT

The study presented in this research begins with the analysis of feelings expressed in messages written by software developers through *commits* in open source repositories, covering the years 2019 to 2021. This period encompasses the previous period, during and after the first major waves of the COVID-19 pandemic. Seven open source projects were examined, totaling 173,993 *commit* messages. The analysis was conducted using two Natural Language Processing (NLP) models based on the RoBERTa architecture, which categorize *commit* messages into three groups: negative, positive and neutral. Subsequently, an investigation was carried out into the relationship between the feelings expressed in *commit* messages and the years in which they were written, as well as the effectiveness of each model. The results of this study suggest that messages tend to be more negative at the beginning of COVID-19 waves. After the pandemic, this number tends to decrease. On the other hand, there was an increase in positive messages during this period.

Keywords: Sentiment Analysis.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

SUMÁRIO

1 INTRODUÇÃO

O desenvolvimento de software é uma atividade altamente colaborativa na qual os participantes utilizam listas de discussão, fóruns, repositórios de códigos de software e ferramentas de rastreamento de problemas, entre outros, para gerenciar seu trabalho Guzman *et al.* 2014. Em meio a esse cenário dinâmico, os repositórios de código aberto emergiram como espaços cruciais de colaboração global, onde desenvolvedores de diferentes partes do mundo contribuem, compartilham e aprimoram projetos de software. O GitHub[1], como um dos maiores e mais influentes desses repositórios, não só se tornou uma plataforma central para o desenvolvimento colaborativo, mas também uma fonte rica de dados que captura as interações, discussões e emoções dos desenvolvedores durante o processo de desenvolvimento de software.

No contexto desse ambiente digital e interativo, os sentimentos que um desenvolvedor projeta durante o desenvolvimento são importantes porque podem ter impacto na produtividade Sinha *et al.* 2016. A análise de sentimentos tornou-se uma ferramenta poderosa para decifrar as nuances emocionais presentes nas mensagens de *commit*. As mensagens de *commit*, embora frequentemente consideradas como registros técnicos e objetivos, são, na realidade, veículos de comunicação que refletem as experiências, desafios e triunfos dos desenvolvedores. Entender o espectro emocional dessas mensagens pode proporcionar uma visão mais profunda não apenas do progresso técnico, mas também da dinâmica humana por trás do código.

Este trabalho visa explorar e analisar a expressão de sentimentos nas mensagens de *commit* do GitHub. Ao longo de um período significativo de três anos, de 2019 a 2021, o estudo se propõe a investigar como os sentimentos evoluíram em resposta a eventos externos, com foco especial nos períodos que antecederam, coincidiram e sucederam as primeiras ondas da pandemia do coronavirus (COVID-19)[2]. A escolha desse intervalo de tempo estratégico permite a análise das respostas emocionais dos desenvolvedores em um período de mudanças significativas e desafios globais.

Dentro desse escopo, as questões de pesquisa que guiam este trabalho são: **(Q1)** Como os sentimentos expressos nas mensagens de *commit* variam ao longo do tempo? **(Q2)** Há padrões distintos relacionados a eventos externos ao longo dos anos, como a pandemia de COVID-19?

Para alcançar esses objetivos, serão empregadas técnicas avançadas de Processamento

[1] https://github.com/about
[2] https://covid19.who.int/

de Linguagem Natural (PLN), utilizando modelos de análise de sentimentos treinados Hartmann *et al.* 2021 Camacho-collados *et al.* 2022 Loureiro *et al.* 2022. Essa abordagem permitirá uma investigação detalhada das flutuações emocionais ao longo do tempo, destacando padrões, tendências e mudanças significativas.

Ao conduzir essa análise de sentimentos, este trabalho almeja contribuir para a compreensão mais profunda da dinâmica emocional no desenvolvimento de software colaborativo. Além disso, ao trazer à tona os aspectos emocionais, este trabalho visa contribuir para a melhoria do ambiente de trabalho no contexto do desenvolvimento colaborativo. A conscientização sobre as emoções expressas nas mensagens de commit pode levar a práticas mais empáticas e inclusivas, promovendo uma cultura de colaboração saudável. Entender como as emoções impactam o processo de desenvolvimento de software é crucial para criar ambientes que incentivem a inovação, a criatividade e o engajamento duradouro dos desenvolvedores e ainda opodem servir como um guia para a implementação de estratégias de gestão de equipe

2 PRELIMINARES

Em cada repositório, os desenvolvedores podem contribuir com alterações no projeto em que estão trabalhando. Uma ou mais alterações podem ser enviadas para o projeto através do *commit*, cada *commit* tem várias informações sobre a alteração enviada. As informações enviadas no *commit* incluem a mensagem de *commit*, na qual o desenvolvedor pode informar detalhadamente quais alterações foram realizadas no código do projeto. Os *commits* são úteis para o gerenciamento do projeto, e as mensagens das alterações podem ser utilizadas para análises, como neste trabalho, onde será aplicada a análise de sentimentos das mensagens de *commit* em projetos que tem o código disponível ao público, sendo, por isso, chamados de projetos de código-fonte aberto.

A análise de sentimentos visa extrair informações emocionais de textos, classificando-os em categorias como positivo, negativo ou neutro. O objetivo é decifrar o tom emocional subjacente em mensagens, avaliações, comentários e outros tipos de texto, proporcionando uma compreensão mais profunda das atitudes e percepções dos autores. Estudos têm conectado fatores humanos, analisando aspectos como humor e emoções por meio de métodos de análise de sentimentos, a uma variedade de áreas, incluindo a resolução de problemas D Wang X 2014 e os resultados do processo de construção (build) da aplicação SOUZA R.; SILVA 2017.

Há diversas maneiras de detectar os sentimentos presentes em um texto. Uma delas é através da biblioteca Hugging Face Transformers[1], plataforma dedicada ao trabalho com modelos de Processamento de Linguagem Natural (PLN), especialmente modelos baseados em transformers. Essa biblioteca oferece uma ampla variedade de modelos pré-treinados incluindo BERT, GPT, RoBERTa, T5 e muitos outros. Nesse estudo foram utilizados dois modelos, o Hartmann *et al.* 2021 com 5.304 postagens anotadas manualmente nas redes sociais e acurácia 86,1 porcento e o Camacho-collados *et al.* 2022 Loureiro *et al.* 2022 pré-treinado com aproximadamente 124 milhões de tweets de janeiro de 2018 a dezembro de 2021. Os modelos estão disponíveis no Hugging Face, todos baseados na arquitetura RoBERTa Liu *et al.* 2019, que é uma variação melhorada do modelo BERT (Bidirectional Encoder Representations from Transformers) Batra *et al.* 2021 pois seu treinamento tem algumas vantagens sobre o BERT como por exemplo o treinamento bidirecionalmente ao longo de todo o texto que permite um maior entendimento de todo o texto enquanto no BERT o treinamento só vai até um determinado ponto do texto.

Os dois modelos foram desenvolvidos para classificar o sentimento do texto em três

[1] https://huggingface.co/models

categorias: positivo, negativo e neutro. Durante a análise das mensagens, esses modelos atribuem probabilidades a cada classe, indicando em que medida a mensagem pode ser classificada como positiva, negativa ou neutra. Por exemplo, na frase "Os casos de Covid estão aumentando rapidamente!", o modelo 2 retorna pontuações nas classes, negativo: 0,724, neutro: 0,229 e positivo: 0,048, indicando que a frase é predominantemente negativa.

Para determinar o sentimento final de uma mensagem de *commit*, realizou-se a busca pela pontuação mais alta entre as classes. Desse modo, uma mensagem de *commit* pode ser categorizada como positiva (quando a maior pontuação é na classe positiva), negativa (quando a maior pontuação é na classe negativa) ou neutra (quando a maior pontuação é na classe neutra). O quadro 1 apresenta alguns exemplos de *commit*s e seus sentimentos correspondentes.

Quadro 1 – Exemplo de mensagens de *commit*

Mensagem	Sentimento
I am super happy to give back to Go on behalf of Hootsuite	Positiva
This test was failing when GOROOT was readonly	Negativa
This patch implements the algorithm from Ulf Adams	Neutro

Fonte: elaborada pelo autor.

3 TRABALHOS RELACIONADOS

Neste capítulo, são abordados trabalhos que associam a análise de sentimentos a elementos no campo do desenvolvimento de software, destacando suas convergências e divergências em relação ao que está sendo proposto.

O estudo Guzman *et al.* 2014 teve como objetivo analisar os comentários de *commits* em projetos hospedados no GitHub por meio da aplicação de técnicas de análise de sentimentos. Os pesquisadores coletaram dados para entender como as emoções expressas nos comentários dos *commits* influenciam a colaboração entre desenvolvedores. A análise incluiu a categorização das mensagens em positivas, negativas ou neutras, proporcionando insights sobre a dinâmica emocional nas comunidades de desenvolvimento de software. O estudo contribuiu para a compreensão das interações sociais e emocionais no contexto do desenvolvimento colaborativo de software, destacando a importância de considerar os aspectos emocionais no processo de colaboração e produção de código.

No SANTOS 2021, foram analisados os comentários presentes em pull requests com o objetivo de compreender se observações positivas podem influenciar na aceitação do pull request. Foram utilizados métodos para a extração desses dados, adotando abordagens de ponta para lidar com grandes volumes de dados (Big Data), e a análise foi realizada com a ferramenta SentiStrength.

O BOECHAT Gláucya; MOTA JR 2019 aborda a investigação das emoções presentes em discussões de issues reabertas no GitHub. Os pesquisadores empregam técnicas de análise de sentimentos para examinar o conteúdo dessas discussões, explorando possíveis padrões e correlações. O estudo visa fornecer insights sobre as dinâmicas emocionais envolvidas nas interações relacionadas a issues reabertas, contribuindo para uma compreensão mais aprofundada do ambiente colaborativo do GitHub.

O C White B 2022 realiza uma análise detalhada dos sentimentos expressos em dados de mídias sociais relacionados à vacina contra a COVID-19. A abordagem empregada envolve o ajuste fino de modelos de análise de sentimentos para lidar especificamente com o contexto das discussões sobre vacinas da COVID-19. O estudo conduz uma análise de sentimentos aplicada a conjuntos de dados provenientes de duas redes sociais populares, Reddit e Twitter. Para realizar essa análise, os pesquisadores empregaram um modelo personalizado que foi ajustado utilizando a linguagem de programação Python, juntamente com o pipeline de análise de sentimentos fornecido pela biblioteca Hugging Face.

Quadro 2 – Comparativo de trabalhos

Trabalho	**Qtd de sentenças analisadas**	**arquitetura *BERT-based***	**Modelo no Hugging Face**
GUZMAN et al. (2014)	60.425	Não	Não
SANTOS (2021)	2.637.650	Não	Não
BOECHAT GLáUCYA; MOTA JR (2019)	12.996	Não	Não
C WHITE B (2022)	9.570.000	Sim	Sim
Este trabalho	173.993	Sim	Sim

A abordagem usada para classificar os sentimentos em uma mensagem dos trabalhos Guzman *et al.* 2014, SANTOS 2021 e BOECHAT Gláucya; MOTA JR 2019 difere da proposta neste trabalho. Enquanto os estudos se baseam na análise das pontuações fornecidas pela ferramenta SentiStrength[1], este trabalho adota modelos baseados na arquitetura BERT como no trabalho C White B 2022. O Quadro 2 mostra a diferença entre os trabalhos relacionados a este trabalho.

[1] http://sentistrength.wlv.ac.uk/

4 METODOLOGIA

Este capítulo apresenta o processo de seleção dos projetos e modelos treinados, descreve o tratamento dos dados dos projetos escolhidos e detalha a análise de sentimentos realizada nas mensagens de *commits*.

4.1 Escolha dos projetos do Github

A seleção dos projetos foi guiada por 3 critérios:

a) Exigência de mensagens de *commit* redigidas em inglês;
b) Necessidade de apresentar um volume significativo de número de *commits* nos anos 2019, 2020 e 2021;
c) Consideração da popularidade dos projetos durante o período de análise.

No primeiro critério, é essencial que as mensagens de *commit* estejam em inglês, uma vez que os modelos foram treinados com mensagens nesse idioma, aprimorando a eficácia da análise em comparação a outros idiomas. O segundo critério foi estabelecido com o intuito de avaliar possíveis variações nos sentimentos das mensagens antes, durante e após o início da pandemia de COVID-19. Portanto, é crucial que o projeto tenha mantido atividade ao longo de todo o período selecionado. A escolha de projetos populares no GitHub reflete a intenção de que os projetos selecionados estejam alinhados com o contexto atual de desenvolvimento, proporcionando uma análise mais representativa do cenário.

4.2 Coleta e Pre-processamento dos dados dos projetos

Inicialmente, foram escolhidos sete projetos do GitHub: Pandas[1], Go[2], LLVM[3], React[4], Node[5], Vscode[6] e OpenBSD[7]. A coleta de dados desses projetos foi realizada a partir de um conjunto de dados[8] de mensagens de *commit* do gitHub disponibilizado no Kaggle, esses dados contém informações sobre mensagens de *commits* de vários projetos do GitHub.

Primeiramente, o conjunto de dados passou por uma etapa de filtragem dos dados,

[1] https://github.com/pandas-dev/pandas
[2] https://github.com/golang/go
[3] https://github.com/llvm/llvm-project
[4] https://github.com/facebook/react
[5] https://github.com/nodejs/node
[6] https://github.com/microsoft/vscode
[7] https://github.com/openbsd/src
[8] https://www.kaggle.com/datasets/dhruvildave/github-*commit*-messages-dataset

no qual foram extraídos os projetos selecionados, juntamente com seus dados, referentes ao período entre os anos de 2019 e 2021. Os projetos foram então separados em diferentes arquivos, considerando o ano em que as mensagens foram escritas.

A quantidade de *commits* em cada projeto está apresentada na Tabela 1. As mensagens de *commits* foram submetidas a alguns critérios durante o tratamento:

a) As mensagens não podem conter caracteres especiais e e-mails, pois esse tipo de dado não é interpretado pelo modelo como uma palavra que pode conter sentimento;
b) As mensagens não podem conter quebras de linha, uma vez que isso afeta a leitura das mensagens pelo modelo.

Tabela 1 – Quantidades de *commits* analisados em cada projeto

Projeto	Número de *commit*s
Pandas	7.967
Golang	8.634
llvm	77.910
Node	8.157
Openbsd	17.869
React	13.006
Vscode	35.664
Total	173.993

Fonte: elaborada pelo autor.

4.3 Análise de sentimentos

Na análise de sentimentos das mensagens de *commit*, foram utilizados dois modelos de Processamento de Linguagem Natural (NLP): o modelo 1 Hartmann *et al.* 2021, pré-treinado para tarefas de análise de sentimentos em inglês, e o modelo 2 Camacho-collados *et al.* 2022 Loureiro *et al.* 2022, pré-treinado com dados do Twitter para tarefas de análise de sentimentos também em inglês. Ambos os modelos foram construídos com base na arquitetura RoBERTa, uma variante da estrutura BERT. Esses modelos são reconhecidos pelo seu desempenho notável em diversas tarefas de PLN LECUN Y.; BENGIO 2015, incluindo a análise de sentimentos.

No processamento dos dados textuais, é necessário dividir cada mensagem de *commit* em diversas partes, denominadas tokens, por meio de um processo de tokenização. Durante esse procedimento, observou-se que nos projetos, devido ao tamanho das mensagens de *commit* apresentadas na Tabela 2, a quantidade de tokens gerada ultrapassa 512, que é a capacidade

máxima suportada pelos modelos utilizados. Consequentemente, foi necessário realizar o truncamento de cada mensagem de *commit* para a mensagem à extensão máxima permitida pelos modelos.

Tabela 2 – Quantidade de palavras por mensagem

Projeto	Ano	Maior Número de Palavras	Menor Número de Palavras	Média
Pandas	2019	224	2	9.02
Pandas	2020	1088	2	9.63
Pandas	2021	107	3	8.45
Golang	2019	846	7	72.15
Golang	2020	2235	7	75.86
Golang	2021	2171	10	78.86
LLVM	2019	4028	1	44.37
LLVM	2020	1031	0	42.91
LLVM	2021	1093	0	43.51
Node	2019	873	7	42.04
Node	2020	1021	7	40.67
Node	2021	742	8	34.95
OpenBSD	2019	403	1	25.61
OpenBSD	2020	475	1	25.61
OpenBSD	2021	416	1	26.33
React	2019	839	1	23.41
React	2020	749	1	40.26
React	2021	1093	0	43.51
VSCode	2019	370	1	7.22
VSCode	2020	269	0	7.49
VSCode	2021	545	1	8.16

Fonte: elaborada pelo autor.

Com a execução dos algoritmos, foram gerados gráficos para cada projeto, exibindo as informações de sentimento por ano. Além disso, foram criadas duas nuvens de palavras, uma destacando os termos mais frequentes nas mensagens classificadas como negativas e outra nas mensagens classificadas como positivas.

4.4 Interpretação e análise dos resultados

Para cada modelo, foi gerado um gráfico de barras com as informações gerais dos dados analisados sem levar em consideração cada projeto individualmente. Também foram gerados gráficos separando os dados por projeto. Os eixos dos gráficos estão divididos entre a quantidade de mensagens de *commits* e os anos em que as mensagens foram redigidas. Para cada ano, a quantidade de mensagens é segmentada em positivas, negativas e neutras, representadas por cores distintas.

Para cada projeto foram geradas nuvens de palavras com os termos que mais apare-

ciam no grupo de mensagens positivas e negativas.

A análise da relação entre as mensagens de *commits* (positivas e negativas) e o primeiro ano da pandemia de COVID-19 foi conduzida através da comparação do aumento ou redução na quantidade de mensagens no ano anterior e no ano subsequente.

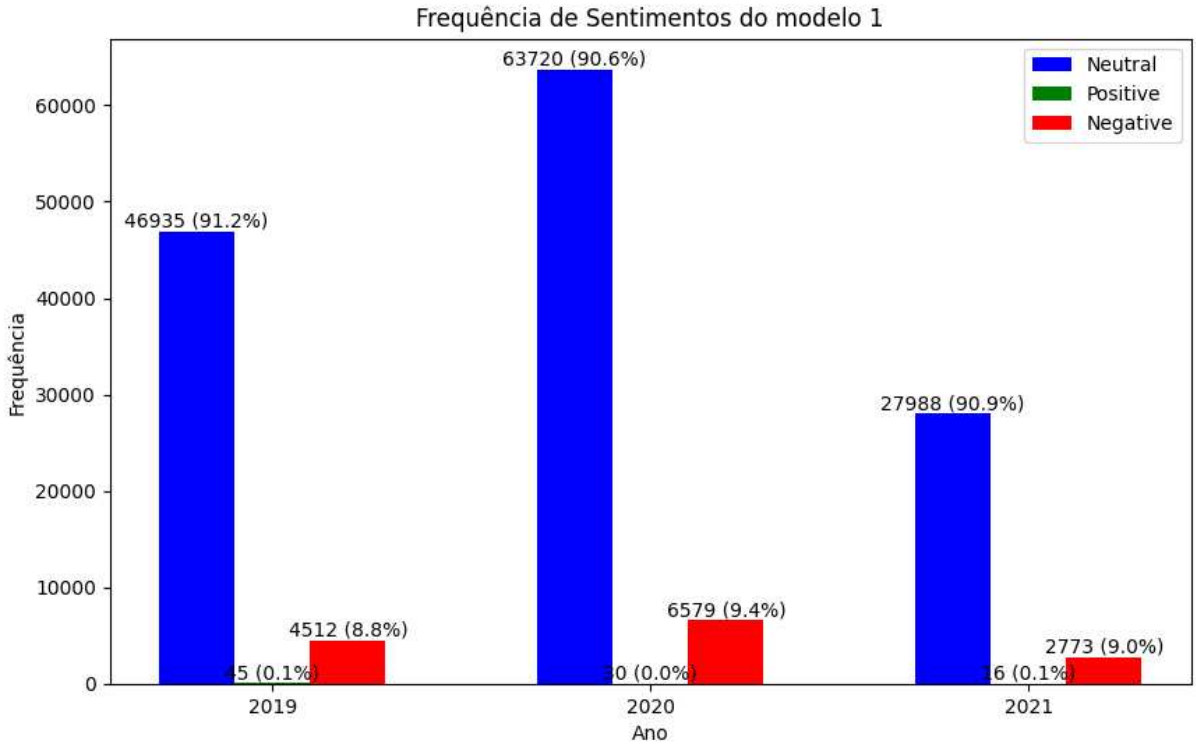
5 RESULTADOS

Neste capítulo, serão apresentados os resultados obtidos a partir da pesquisa que consistiu na coleta e análise de sentimento das mensagens de *commit* de sete projetos de código aberto no período de 2019 a 2021. A quantidade de mensagens analisadas por cada modelo está detalhada na Tabela 3 para o primeiro modelo e na Tabela 4 para o segundo modelo. Para abordar as questões Q1 e Q2 (Como os sentimentos expressos nas mensagens de *commit* variam ao longo do tempo?; Há padrões distintos relacionados a eventos externos, como a pandemia de COVID-19?), os sentimentos das mensagens de *commit* foram segmentados por ano e projeto, conforme ilustrado nas imagens seguintes.

5.1 Análise de sentimentos

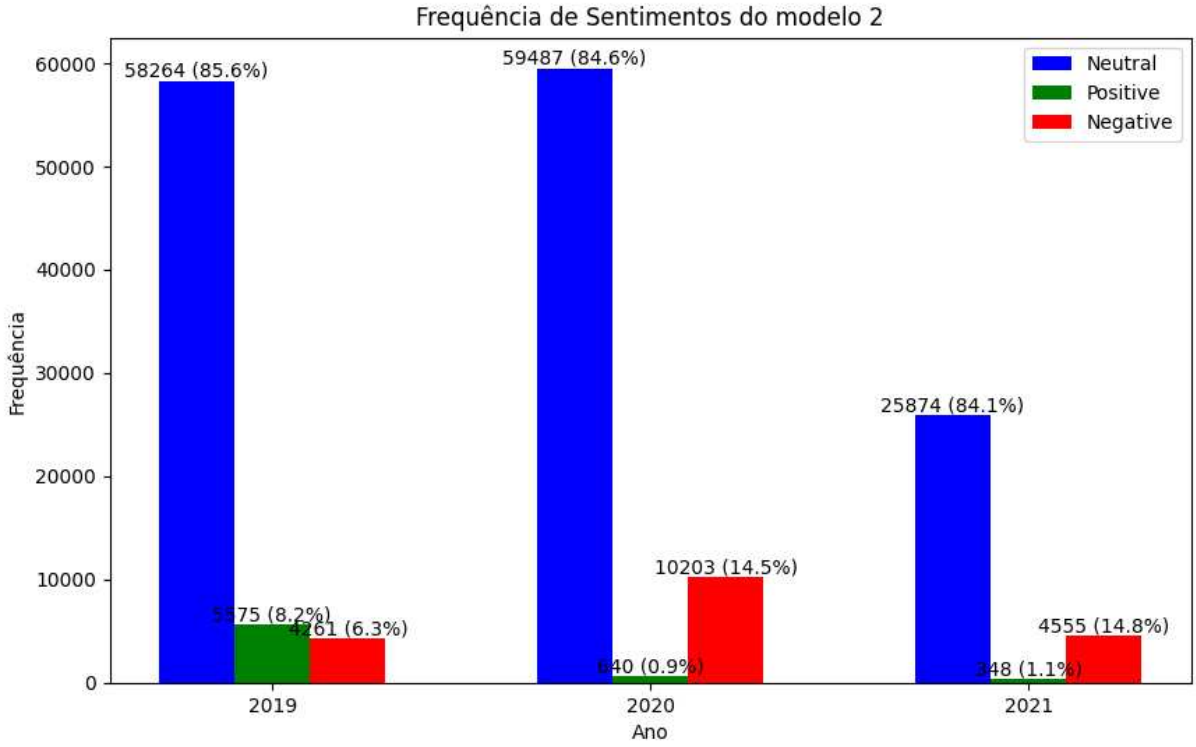
As Figuras 1 e 2 apresentam os resultados da análise de sentimentos pelos modelos 1 e 2. Em ambos os modelos, é perceptível um aumento nas mensagens negativas coincidindo com a chegada da COVID-19, seguido por uma diminuição no ano subsequente. Por outro lado, os sentimentos positivos nos modelos mostram uma tendência oposta: há uma queda durante a pandemia e um aumento significativo após o período pandêmico.

Figura 1 – Gráfico dos sentimentos identificados pelo modelo 1



Fonte: Elaborado pelo autor.

Figura 2 – Gráfico dos sentimentos identificados pelo modelo 2



Fonte: Elaborado pelo autor.

5.1.1 Resultados por projeto

Tabela 3 – Sentimentos das mensagens de *commit*s por projeto - Modelo 1

Projeto	Total de *commit*s	Negativo	Positivo	Neutro
Pandas	7.967	1.063	2	6.902
Golang	8.634	519	2	8.113
LLVM	61.302	5.772	10	55.520
Node	8.157	7.809	0	348
OpenBSD	17.869	15.858	23	1.988
React	11.481	11.852	15	1.139
VsCode	35.664	32.589	40	3.035
Total	151.074	75.462	92	77.045

Fonte: elaborada pelo autor.

Ao analisar o projeto Golang no modelo 1 nas Figuras 3 e 4, nota-se uma diminuição dos sentimentos negativos ao longo dos três anos. Além disso, observa-se que o modelo 1 não indentificou mensagens com sentimentos positivos em comparação com o modelo 2. No modelo 1, foram identificadas apenas duas mensagens com sentimentos positivos, são elas:

1. *All remove the nacl port part 1. You were a useful port and youve served your pur-*

Tabela 4 – Sentimentos das mensagens de *commit*s por projeto - Modelo 2

Projeto	Total de *commit*s	Negativo	Positivo	Neutro
Pandas	7.967	1.200	26	6.741
Golang	8.634	1.031	84	7.519
LLVM	77.910	13.173	886	63.851
Node	8.157	555	78	7.524
OpenBSD	17.869	2.746	284	14.839
React	13.006	1.858	165	10.790
VsCode	35.664	3.130	173	32.361
Total	173.993	28.722	1.646	143.625

Fonte: elaborada pelo autor.

pose. Thanks for all the play. A subsequent CL will remove amd64p32 including assembly files and toolchain bits and remaining bits The amd64p32 removal will be separated into its own CL in case we want to support the Linux x32 ABI in the future and want our old amd64p32 support as a starting point. Updates 30439. ChangeId Ia3a0c7d49804adc87bf52a4dea7e3d3007f2b1cd. RunTryBot Brad;

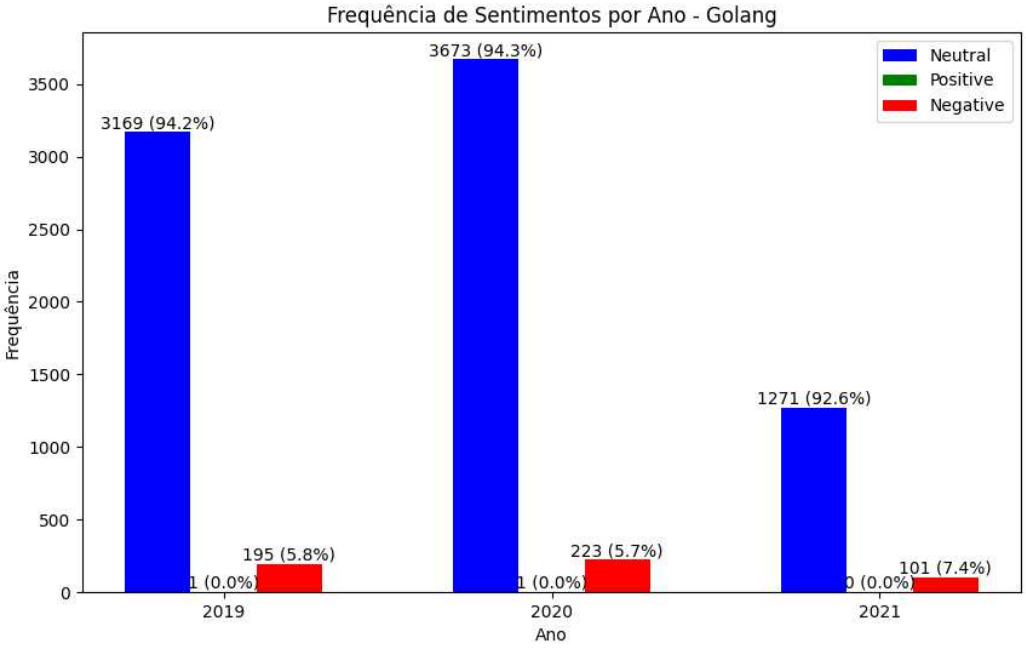
2. *AC add Kush Patel corporate CLA for Hootsuite Inc. Im from Hootsuite Were a Canadian tech company who provides products and services to businesses organizations and individuals to really help them succeed on social We have leveraged Go in our stack for the past 4 years I am super happy to give back to Go on behalf of Hootsuite through a small contribution to pkgsite with a few more in the works. We love this project and we love open source. Hopefully we can give back more*.

Já o modelo 2 identificou diversas mensagens positivas que não foram identificadas pelo modelo 1, como por exemplo:

1. *Runtime incorporate hbits advancement in scanobject into loop This makes it clearer that i and hbits advance together. As a bonus it generates slightly better code. ChangeId I24d51102535c39f962a59c1a4a7c5c894339aa18. Trust Josh Bleecher Snyder joshariangmailcom. RunTryBot Josh Bleecher Snyder joshariangmailcom. TryBotResult Go Bot gobotgolangorg. Reviewedby Austin Clements austingooglecom*;
2. *Time change genzabbrsgo to fetch windowsZonesxml file from GitHub. It seems that windowsZonesxml file has moved to Github I opened in my browser and it redirected me to. Very nice of them and we could see windowsZonesxml change history now We could even probably file issues against this file if we find problems. Anyway this CL adjusts genzabbrsgo to use new GitHub location I also run go generate command with updated genzabbrsgo to update*.

Na classificação dos sentimentos oferecida pelo modelo 2, observa-se que, ao entrar no ano em que o COVID-19 se iniciou, as mensagens positivas diminuíram em relação ao ano anterior e aumentaram após 2020, ao passo que os sentimentos negativos diminuíam.

Figura 3 – Gráfico do projeto Golang analisado pelo modelo 1



Fonte: Elaborado pelo autor.

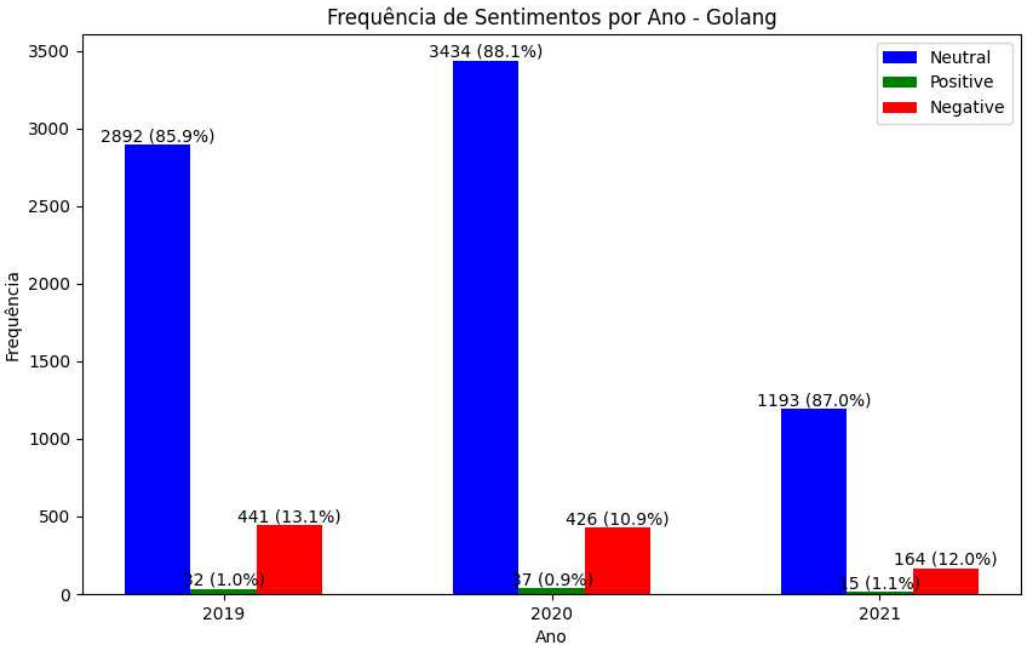
É perceptível, também, nos demais projetos, que o modelo 1 não foi tão eficaz na identificação de mensagens com sentimentos positivos em relação ao modelo 2. No caso do projeto LLVM nas Figuras 5 e 6, por exemplo, nos dois modelos é evidente que o número de mensagens negativas aumentou durante o ano do início da pandemia de COVID-19 e diminuiu no ano seguinte, acompanhado por um aumento no número de mensagens positivas.

A classificação de sentimentos no projeto Node reportado nas Figuras 7 e 8, é possível observar que ambos os modelos apresentaram um aumento nos sentimentos negativos durante o início da pandemia de COVID-19, seguido por uma diminuição desses sentimentos após 2020. Por outro lado, os sentimentos positivos apenas diminuíram ao longo dos anos.

A redução dos sentimentos negativos também é evidente no projeto Openbsd pós 2020. Comparando os resultados obtidos pelo modelo 1 (Figura 9) e pelo modelo 2 (Figura 10), sendo mais evidente essa diminuição nos resultados do modelo 2. De toda forma, em ambos os casos, foi observado que os sentimentos positivos não eram evidentes.

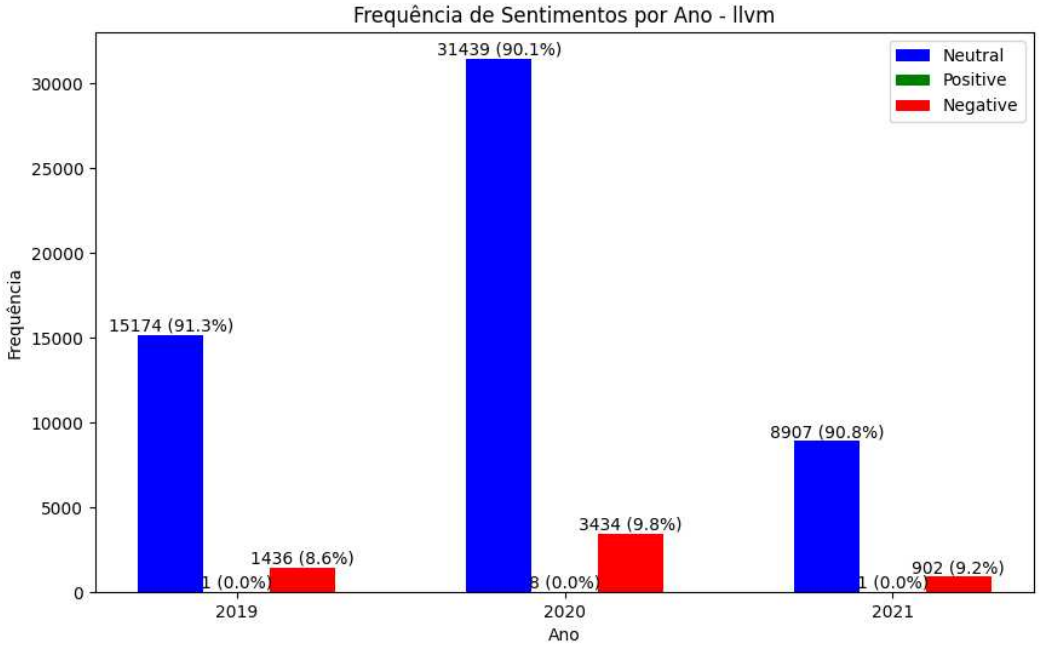
No projeto Pandas, os resultados da classificação dos sentimentos das mensagens

Figura 4 – Gráfico do projeto Golang analisado pelo modelo 2



Fonte: Elaborado pelo autor.

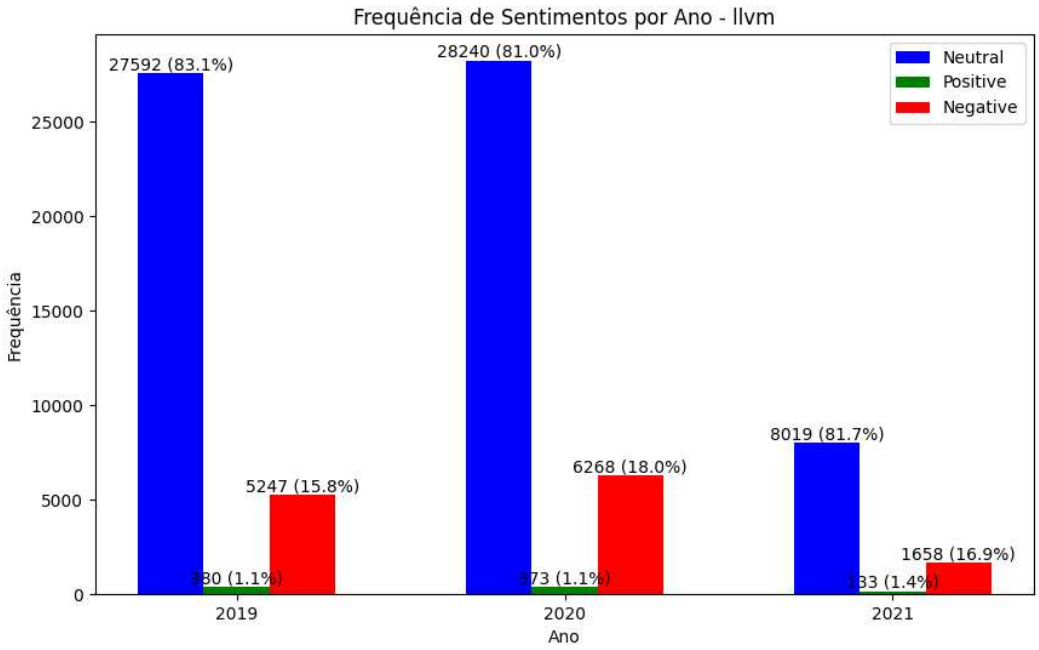
Figura 5 – Gráfico do projeto LLVM analisado pelo modelo 1



Fonte: Elaborado pelo autor.

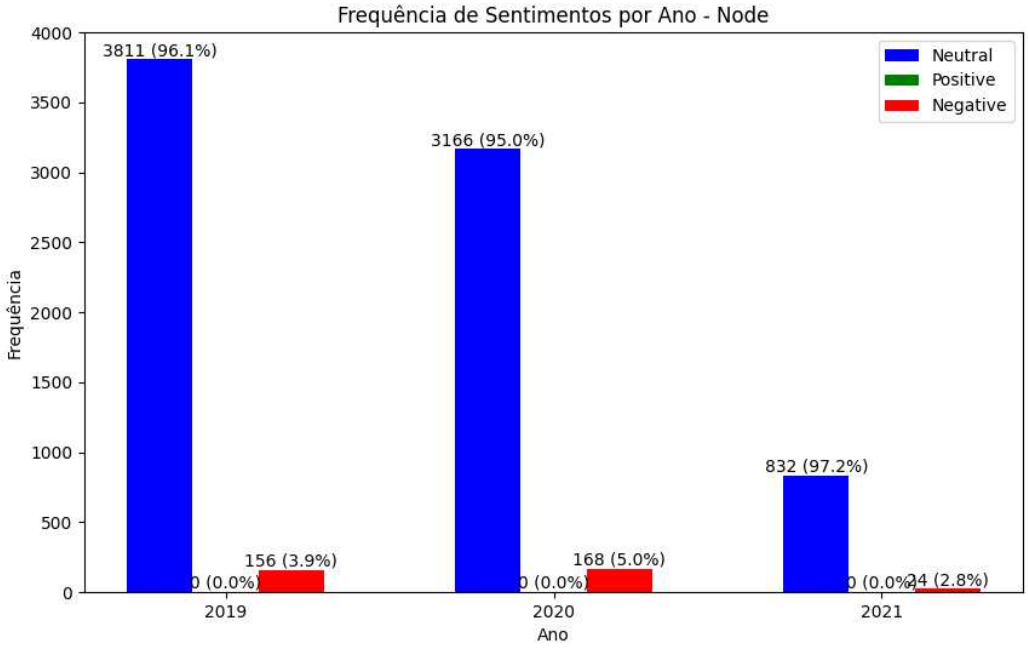
são reportados nas Figuras 11 e 12, respectivamente. Neste projeto, foi identificado um cenário diferente: os sentimentos negativos no modelo 1 aumentaram durante todos os anos de análise, enquanto no modelo 2 foi possível identificar uma queda desses sentimentos negativos durante

Figura 6 – Gráfico do projeto LLVM analisado pelo modelo 2



Fonte: Elaborado pelo autor.

Figura 7 – Gráfico do projeto Node analisado pelo modelo 1

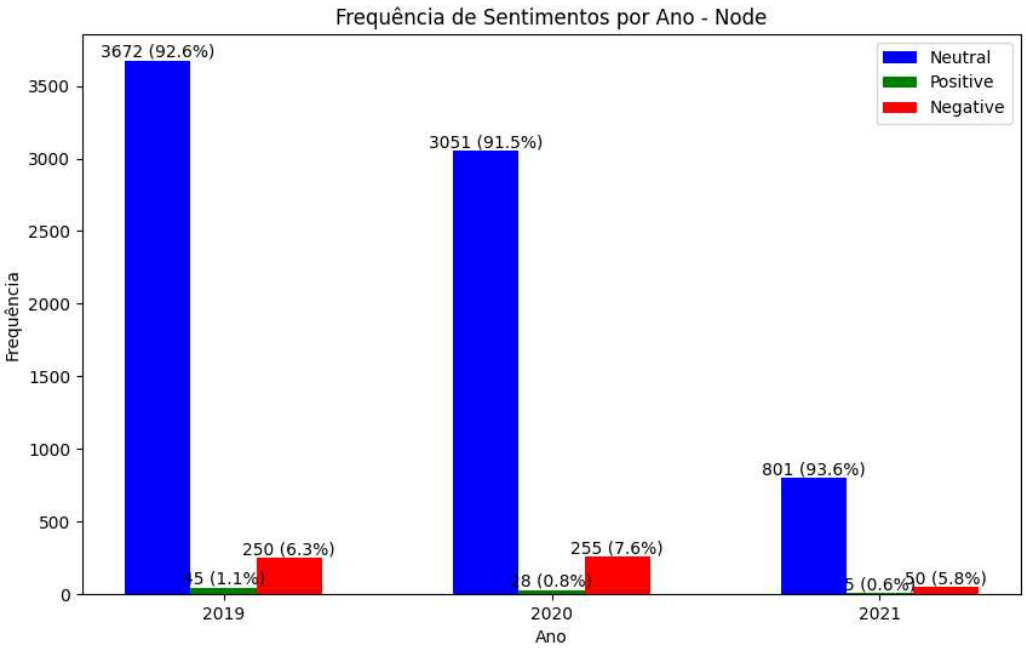


Fonte: Elaborado pelo autor.

2020, seguida por um aumento no ano subsequente.

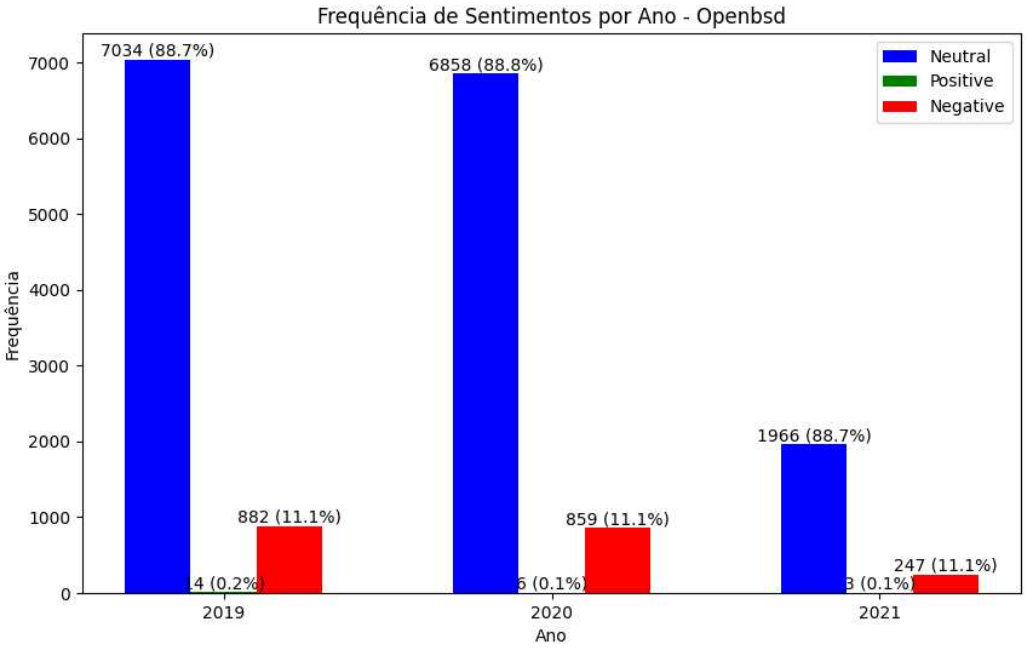
Já no projeto React, exposto o resultado das classificações de sentimentos usando os modelo 1 e modelo 2, respectivamente, nas Figuras 13 e 14, foi de aumento no número de

Figura 8 – Gráfico do projeto Node analisado pelo modelo 2



Fonte: Elaborado pelo autor.

Figura 9 – Gráfico do projeto Openbsd analisado pelo modelo 1

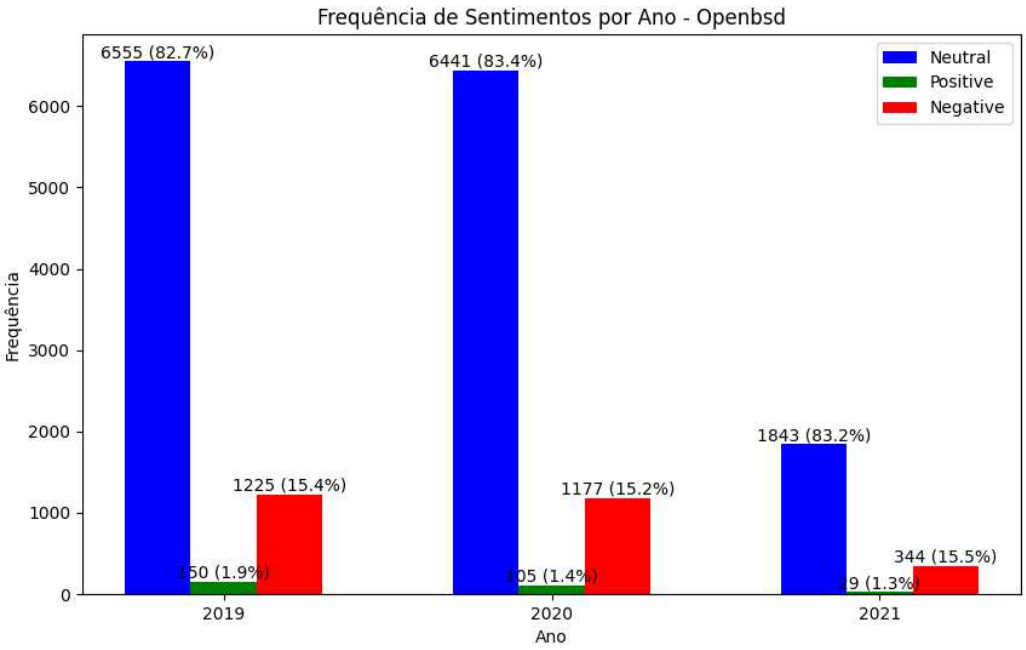


Fonte: Elaborado pelo autor.

mensagens de *commit* negativas, acompanhada por um aumento no número de sentimentos positivos após 2020.

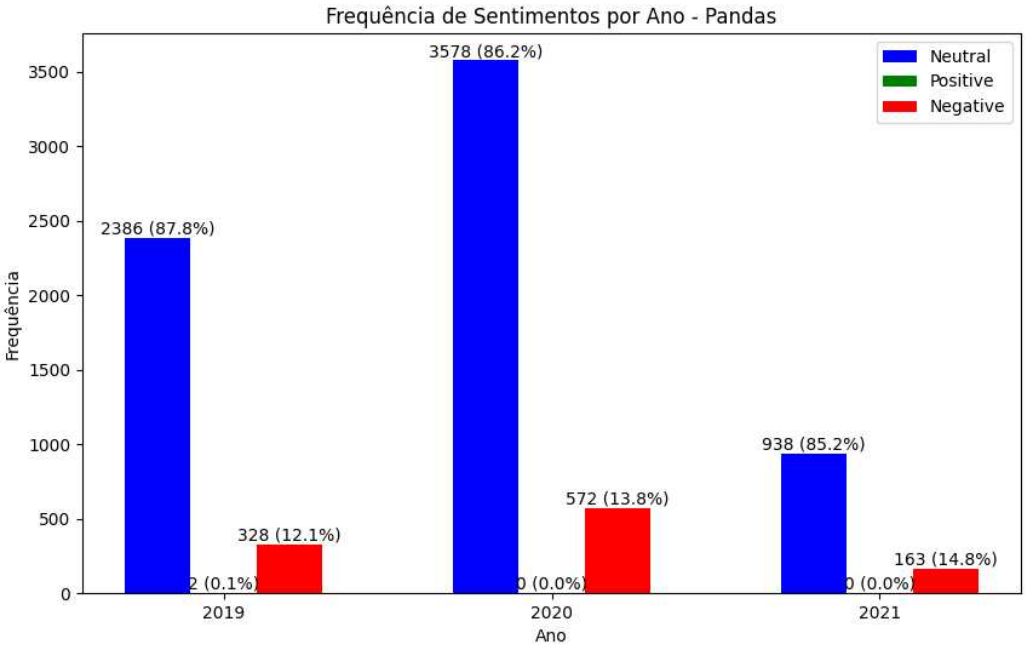
As Figuras 15 e 16 referentes a classificação de sentimentos nas mensagens de

Figura 10 – Gráfico do projeto Openbsd analisado pelo modelo 2



Fonte: Elaborado pelo autor.

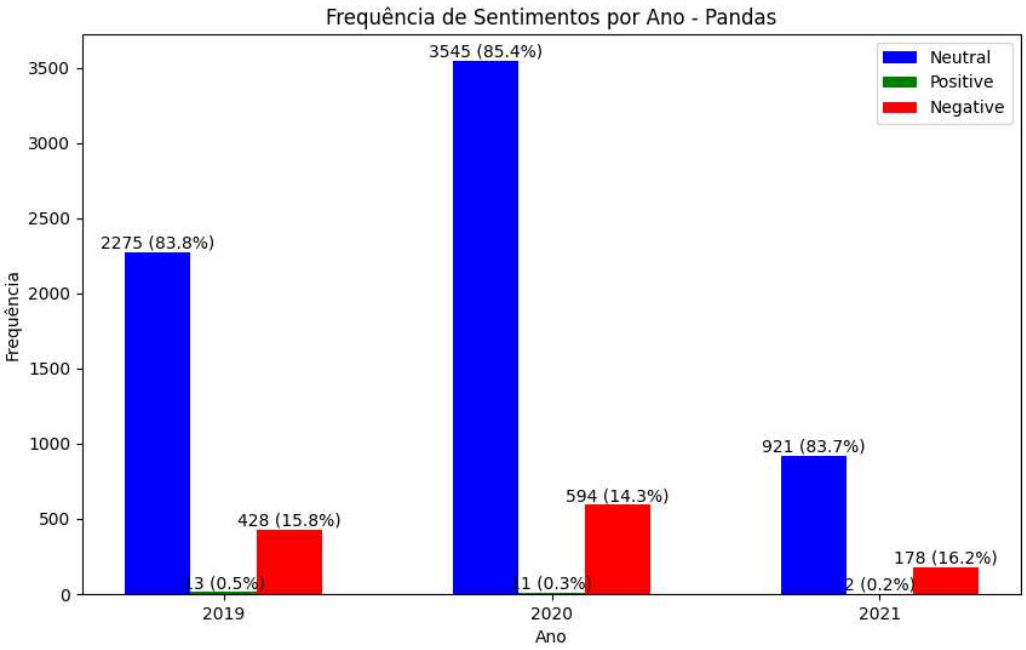
Figura 11 – Gráfico do projeto Pandas analisado pelo modelo 1



Fonte: Elaborado pelo autor.

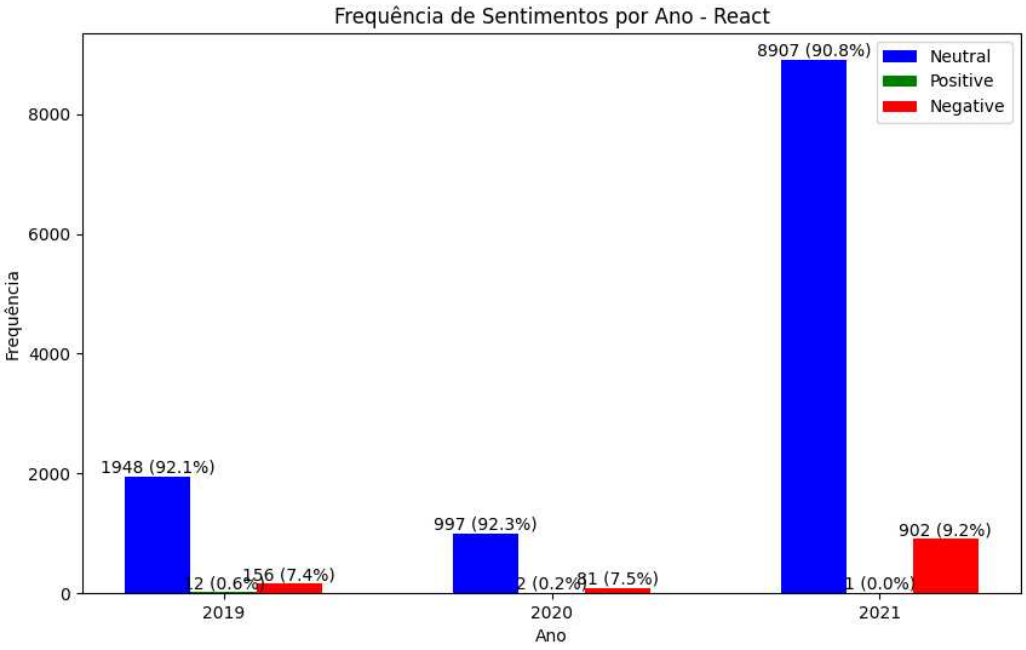
commit do projeto VSCode pelo modelo 1 e modelo 2, respectivamente. Observou-se que, nos dois modelos, houve um leve aumento da quantidade de sentimentos positivos ao longo dos anos analisados. Além disso, a quantidade de mensagens de *commit* com sentimentos negativos

Figura 12 – Gráfico do projeto Pandas analisado pelo modelo 2



Fonte: Elaborado pelo autor.

Figura 13 – Gráfico do projeto React analisado pelo modelo 1

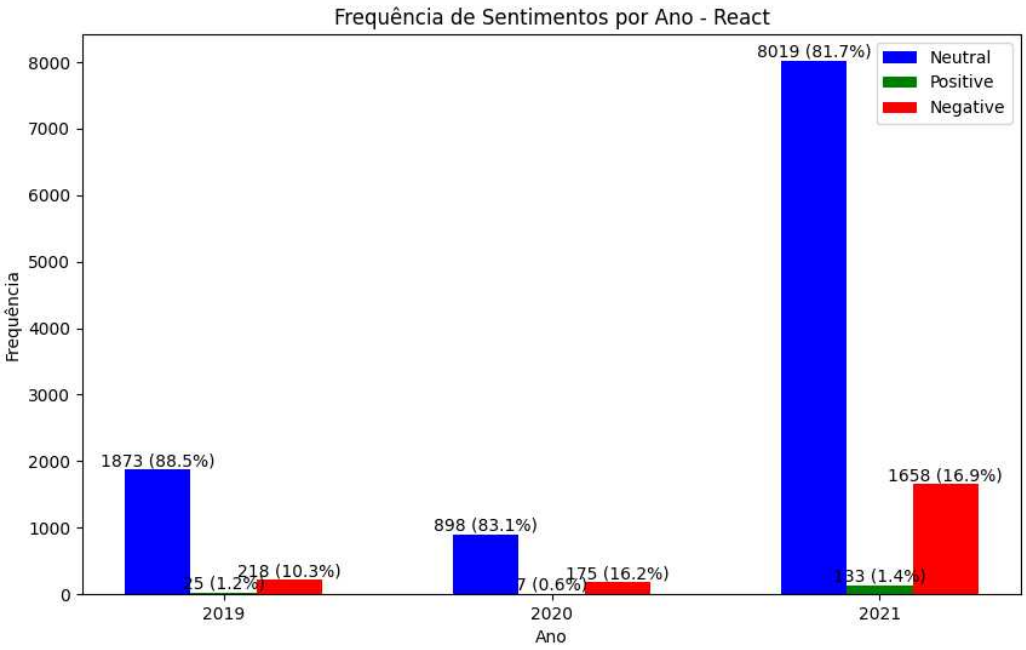


Fonte: Elaborado pelo autor.

diminuiu com o passar dos anos no modelo 1, enquanto no modelo 2 houve queda apenas nos dois primeiros anos.

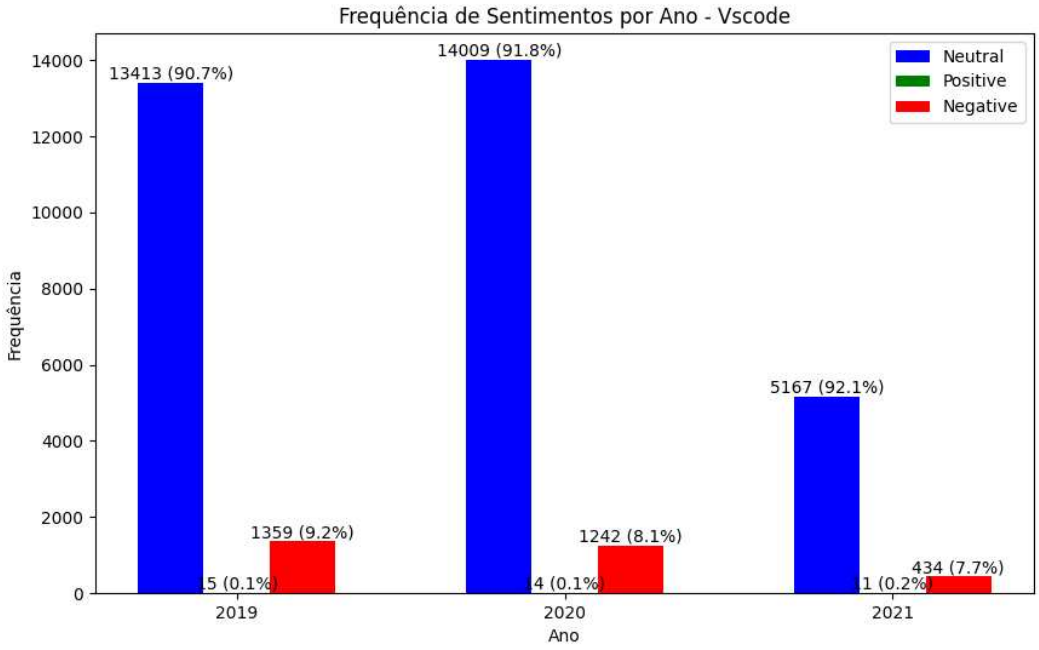
As Figuras 17-43 apresentam as palavras que mais aparecem nos *commits* positivos

Figura 14 – Gráfico do projeto React analisado pelo modelo 2



Fonte: Elaborado pelo autor.

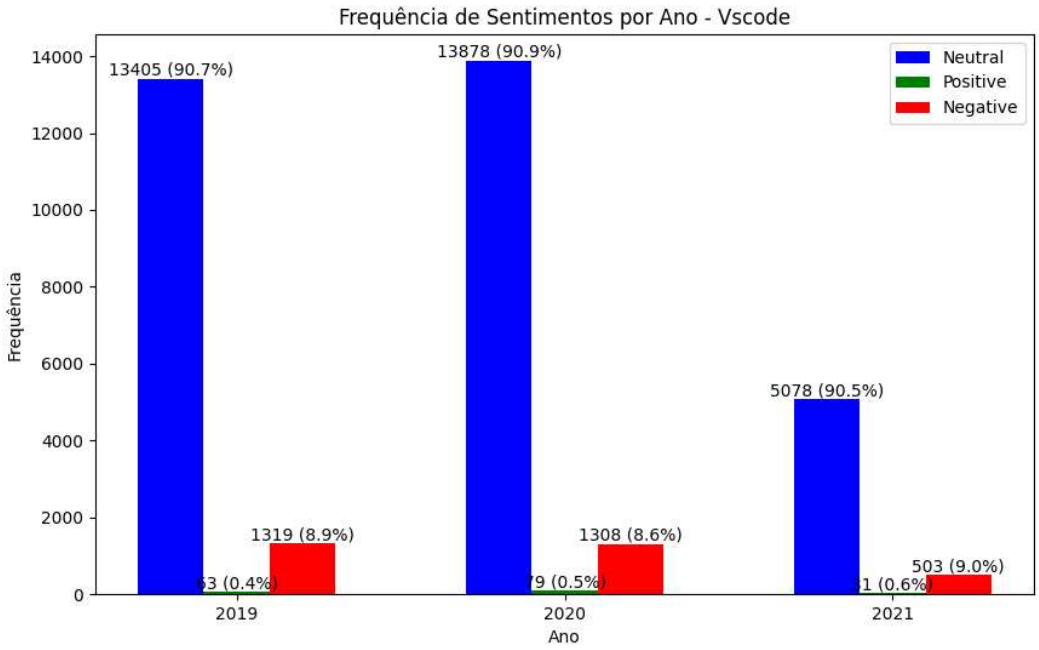
Figura 15 – Gráfico do projeto Vscode analisado pelo modelo 1



Fonte: Elaborado pelo autor.

e negativos de cada um dos projetos analisados. Como não foram eliminadas as *stopwords*, tais como artigos e preposições, entre as apresentadas *the, this, and, in, etc*. Tais palavras frequentemente ocorrem nos textos, boa parte das nuvens de palavras apresentadas apresentam

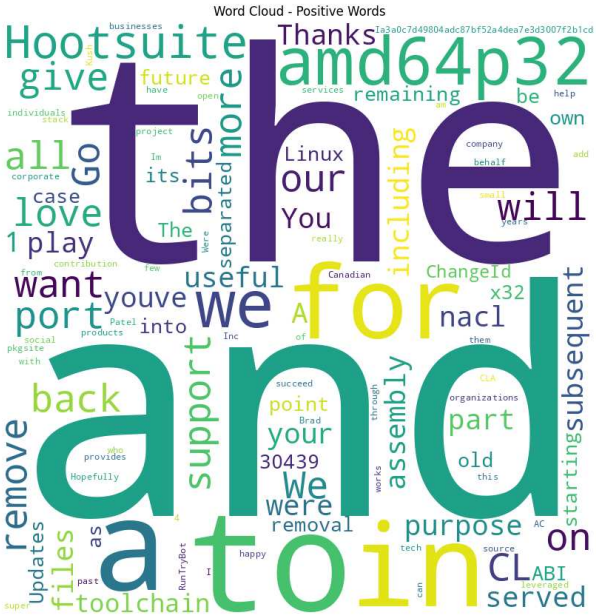
Figura 16 – Gráfico do projeto Vscode analisado pelo modelo 2



Fonte: Elaborado pelo autor.

Figura 17 – Palavras que mais aparecem em mensagens negativas do projeto Golang.

Figura 18 – Palavras que mais aparecem em mensagens positivas do projeto Golang.



tais palavras como as mais frequentes. Como neste trabalho foi utilizado modelos de linguagem *BERT-based* não há necessidade de eliminação dessas palavras dos textos para realização da análise de sentimentos. Palavras como *not, remove, unused, bug* remetem a sentimentos negativos, já palavras como *good, ok, fun*, realmente remetem a sentimentos positivos.

Figura 19 – Palavras que mais aparecem em mensagens negativas do projeto Golang.

Figura 20 – Palavras que mais aparecem em mensagens positivas do projeto Golang.

Figura 21 – Palavras que mais aparecem em mensagens negativas do projeto LLVM.

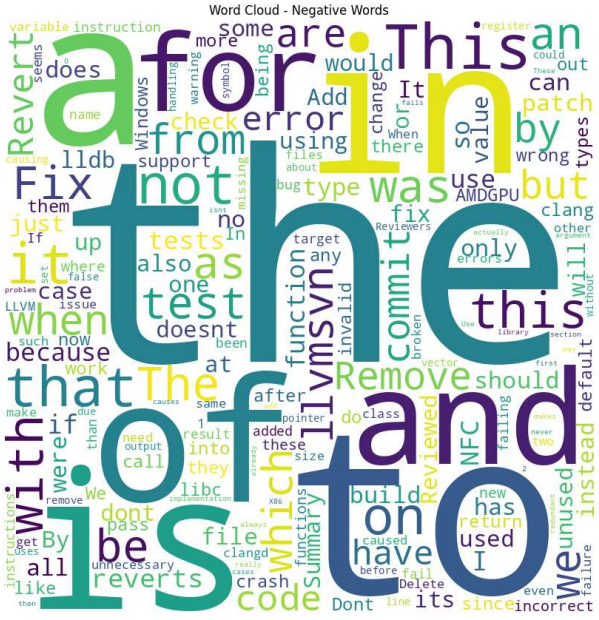


Figura 22 – Palavras que mais aparecem em mensagens positivas do projeto LLVM.

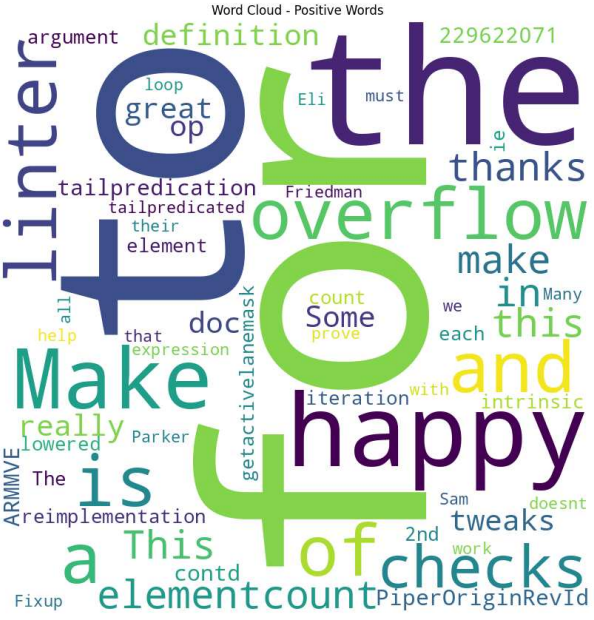


Figura 23 – Palavras que mais aparecem em mensagens negativas do projeto LLVM.

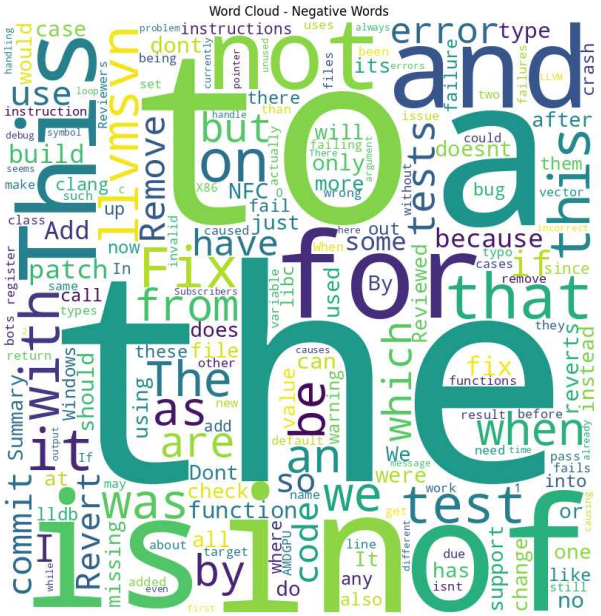


Figura 24 – Palavras que mais aparecem em mensagens positivas do projeto LLVM.

Figura 25 – Palavras que mais aparecem em mensagens negativas do projeto Node.

Figura 26 – Palavras que mais aparecem em mensagens negativas do projeto Node.

Figura 27 – Palavras que mais aparecem em mensagens positivas do projeto Node.

Figura 28 – Palavras que mais aparecem em mensagens negativas do projeto Openbsd.

Figura 29 – Palavras que mais aparecem em mensagens positivas do projeto Openbsd.

Figura 30 – Palavras que mais aparecem em mensagens negativas do projeto Openbsd.

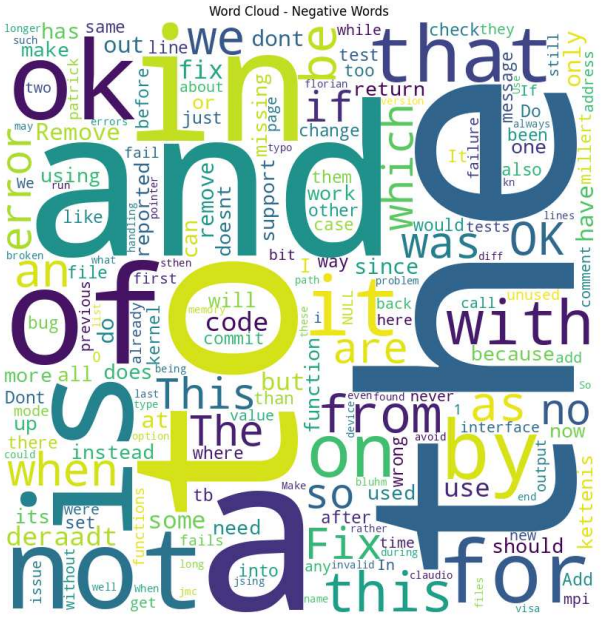


Figura 31 – Palavras que mais aparecem em mensagens positivas do projeto Openbsd.

Figura 32 – Palavras que mais aparecem em mensagens negativas do projeto Pandas.

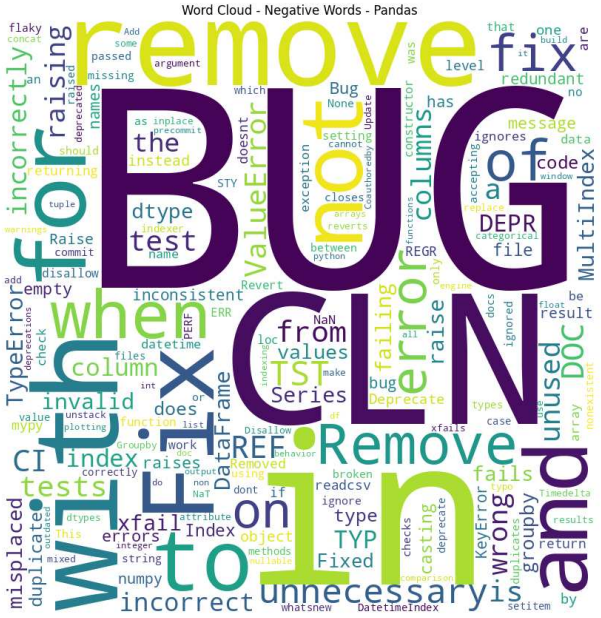


Figura 33 – Palavras que mais aparecem em mensagens positivas do projeto Pandas.

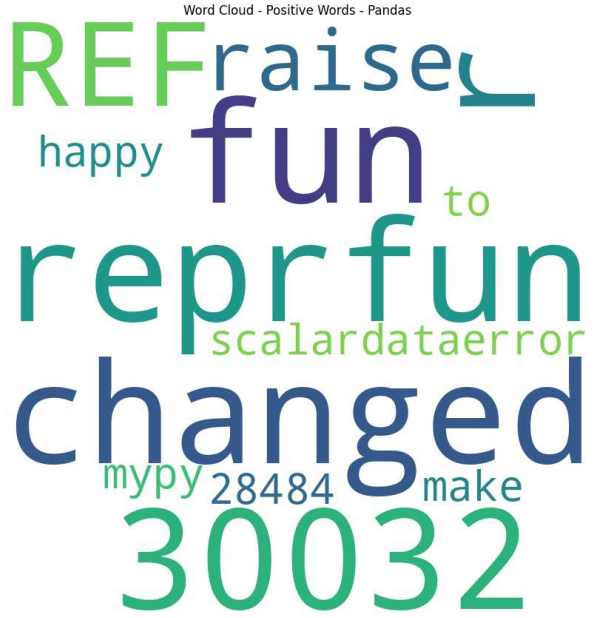


Figura 34 – Palavras que mais aparecem em mensagens negativas do projeto Pandas.

Figura 35 – Palavras que mais aparecem em mensagens positivas do projeto Pandas.

Figura 36 – Palavras que mais aparecem em mensagens negativas do projeto React.

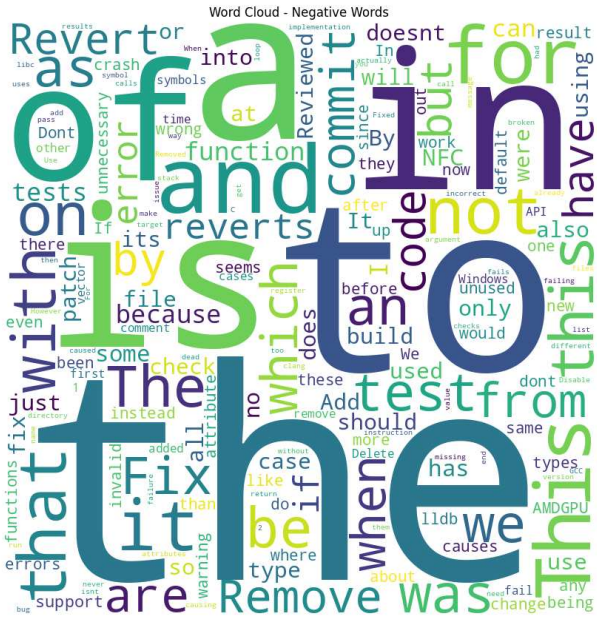


Figura 37 – Palavras que mais aparecem em mensagens positivas do projeto Pandas.

Figura 38 – Palavras que mais aparecem em mensagens negativas do projeto React.

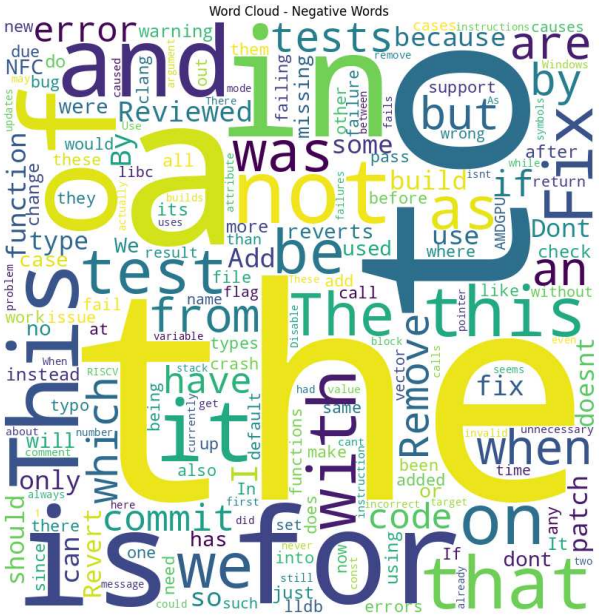


Figura 39 – Palavras que mais aparecem em mensagens positivas do projeto React.

Figura 40 – Palavras que mais aparecem em mensagens negativas do projeto Vscode.

Figura 41 – Palavras que mais aparecem em mensagens positivas do projeto Vscode.

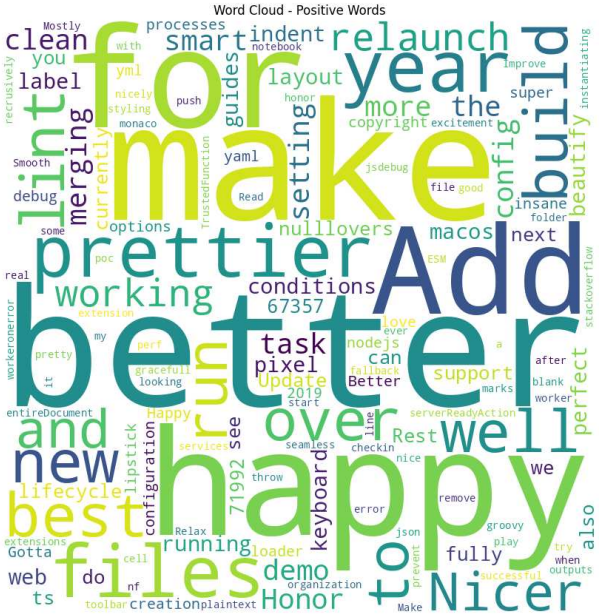


Figura 42 – Palavras que mais aparecem em mensagens negativas do projeto Vscode.

Figura 43 – Palavras que mais aparecem em mensagens positivas do projeto Vscode.

6 CONSIDERAÇÕES FINAIS

Este estudo analisou as mensagens de *commits* de sete projetos de código aberto disponíveis no GitHub, abrangendo o período de 2019 a 2021, englobando os momentos anteriores, durante e após as primeiras ondas da pandemia de COVID-19. Os sentimentos identificados foram organizados por ano, possibilitando uma análise temporal desses dados. Evidências revelaram um aumento significativo de mensagens com sentimentos negativos durante o período da pandemia, seguido por uma queda nas mensagens com sentimentos positivos. Após o término da pandemia, observou-se uma mudança nesse cenário, com o aumento das mensagens com sentimentos positivos e a redução das mensagens com sentimentos negativos. Isso sugere que os desenvolvedores foram impactados negativamente durante o período pandêmico, com uma posterior recuperação emocional.

Como trabalhos futuros, pretende-se:

a) Fazer esta mesma análise para um conjunto maior de projetos visando obter uma compreensão mais abrangente dos padrões de sentimentos em mensagens de *commits*;
b) Melhorar os gráficos, removendo o número de mensagens neutras;
c) Remover as *stop words* das nuvens de palavras para trazer uma maior visibilidade para as palavras significativas nas nuvens de sentimentos;
d) Investigar possiveis erros dos modelos nas mensagens definidas como neutras;
e) Fazer uma relação entre as mensagens de commis e os desenvolvedores

Essas extensões da pesquisa podem contribuir para uma compreensão mais completa do ambiente de trabalho na área de desenvolvimento de software.

REFERÊNCIAS

BATRA, H.; PUNN, N. S.; SONBHADRA, S. K.; AGARWAL, S. Bert-based sentiment analysis: A software engineering perspective. In: STRAUSS, C.; KOTSIS, G.; TJOA, A. M.; KHALIL, I. (Ed.). **Database and Expert Systems Applications**. Cham: Springer International Publishing, 2021. p. 138–148. ISBN 978-3-030-86472-9.

BOECHAT GLáUCYA; MOTA JR, J. M. I. M. M. **Análise de Sentimentos em Discussões de Issues Reabertas do Github.** Paraíba, Brasil: Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2019. 1-8 p. Disponível em: https://doi.org/10.5753/vem.2019.7579.

C WHITE B, D. R. B. R. S.-N. A. M. Fine-tuned sentiment analysis of covid-19 vaccine–related social media data: Comparative study. In: . J Med Internet Res, 2022. v. 24, n. 10:e40408, p. 1–8. Disponível em: https://www.jmir.org/2022/10/e40408.

CAMACHO-COLLADOS, J.; REZAEE, K.; RIAHI, T.; USHIO, A.; LOUREIRO, D.; ANTYPAS, D.; BOISSON, J.; ANKE, L. E.; LIU, F.; CÁMARA, E. M. *et al.* TweetNLP: Cutting-edge natural language processing for social media. In: **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**. Abu Dhabi, UAE: Association for Computational Linguistics, 2022. p. 38–49. Disponível em: https://aclanthology.org/2022.emnlp-demos.5.

D WANG X, A. P. G. Desenvolvedores de software felizes resolvem melhor os problemas: medições psicológicas na engenharia de software empírica. In: . PeerJ, 2014. v. 2, n. peerj.289. Disponível em: https://doi.org/10.7717/peerj.289.

GUZMAN, E.; AZóCAR, D.; LI, Y. **Sentiment Analysis of Commit Comments in GitHub: An Empirical Study**. New York, NY, USA: Association for Computing Machinery, 2014. 352–355 p. (MSR 2014). Disponível em: https://doi.org/10.1145/2597073.2597118.

HARTMANN, J.; HEITMANN, M.; SCHAMP, C.; NETZER, O. The power of brand selfies. **Journal of Marketing Research**, 2021.

LECUN Y.; BENGIO, Y. H. G. Deep learning. **Nature**, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.

LIU, Y.; OTT, M.; GOYAL, N.; DU, J.; JOSHI, M.; CHEN, D.; LEVY, O.; LEWIS, M.; ZETTLEMOYER, L.; STOYANOV, V. **RoBERTa: A Robustly Optimized BERT Pretraining Approach**. 2019.

LOUREIRO, D.; BARBIERI, F.; NEVES, L.; ANKE, L. E.; CAMACHO-COLLADOS, J. TimeLMs: Diachronic language models from Twitter. In: **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations**. Dublin, Ireland: Association for Computational Linguistics, 2022. p. 251–260. Disponível em: https://aclanthology.org/2022.acl-demo.25.

SANTOS, R. O. Análise de sentimentos em repositórios do github. (Trabalho de Conclusão de Curso - Artigo) – Curso de Bacharelado em Ciência da Computação, Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande, Paraíba, Brasil, 2021. Disponível em: http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/19697.

SINHA, V.; LAZAR, A.; SHARIF, B. Analyzing developer sentiment in commit logs. In: **Proceedings of the 13th International Conference on Mining Software Repositories**. New York, NY, USA: Association for Computing Machinery, 2016. (MSR '16), p. 520–523. ISBN 9781450341868. Disponível em: https://doi.org/10.1145/2901739.2903501.

SOUZA R.; SILVA, B. Sentiment analysis of travis ci builds. In: **2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)**. [*S. l.*]: [S.l.: s.n.], 2017. p. 459–462.