Dataset Anotado de Sentimentos a partir de comentários de Aplicativos Móveis

Vitor X. Siqueira[1], Ricardo Luiz Hentges Costa[1], Tales Schifelbein Soares[1], Gabriel M. Lunardi[2], Williamson Silva[1]

[1]Universidade Federal do Pampa (UNIPAMPA)
Av. Tiaraju, 810 – 97546-550 – Alegrete – RS – Brasil

[2]Universidade Federal de Santa Maria (UFSM)
Av. Roraima, 1000 – 97105-900 – Santa Maria – RS – Brasil

{vitorsiqueira, ricardohentges, talessoares}.aluno@unipampa.edu.br

williamsonsilva@unipampa.edu.br

gabriel.lunardi@ufsm.br

Abstract. *This work introduces a novel Brazilian Portuguese dataset for sentiment analysis, comprising 3,000 user reviews extracted from 10 popular Google Play Store applications. Reviews were manually classified into seven basic emotions through a collaborative annotation process and validated by multiple annotators. Analysis reveals a predominance of negative emotions, suggesting potential for future research on the relation between emotions and user satisfaction. This resource aims to fill the gap of Portuguese datasets and foster the development of tools and research in areas such as human-computer interaction, marketing, and software engineering.*

Resumo. *Este trabalho apresenta um novo dataset em português brasileiro para análise de sentimentos, composto por 3.000 avaliações de usuários extraídas de 10 aplicativos populares da Google Play Store. As avaliações foram manualmente classificadas em sete emoções básicas através de um processo de anotação colaborativa e validado por múltiplos avaliadores. A análise demonstra uma predominância de emoções negativas, indicando potencial para pesquisas sobre a relação entre emoções e satisfação dos usuários. Este trabalho visa suprir a lacuna de datasets em português e impulsionar o desenvolvimento de ferramentas e pesquisas em áreas como interação humano-computador, marketing e engenharia de software.*

1. Introdução

Os aplicativos móveis tornaram-se peças centrais na vida cotidiana, conectando usuários a serviços, informações e entre si [Quiroga e Bessa 2024]. A Google Play Store, como principal plataforma de distribuição de apps para o sistema Android, exemplifica essa relevância, hospedando milhões de aplicativos e bilhões de usuários ativos [Fabro 2021]. Nesse contexto, as avaliações dos usuários emergem como um recurso para entender a receptividade do público, a qualidade dos apps e as necessidades dos usuários. A partir da análise de sentimentos expressos nesses comentários, empresas podem identificar funcionalidades problemáticas, direcionar o desenvolvimento de novas *features* e aprimorar a experiência do usuário [Motger et al. 2024].

Considerando a importância das avaliações de usuários, os comentários podem auxiliar projetistas e desenvolvedores a analisar e compreender as opiniões sobre características específicas, bem como a obter novas perspectivas para possíveis requisitos futuros ou atualizações [Guzman e Maalej 2014]. Essas avaliações ajudam ainda a identificar falhas e erros, validar funcionalidades relevantes, compreender o mercado, fomentar a inovação e demonstrar o valor dos usuários, garantindo que o produto atenda às necessidades e expectativas de forma efetiva [Maalej e Nabil 2015, Nakamura et al. 2022]. Tendo isso em vista, um conjunto de dados consistente sobre avaliações de aplicativos móveis é de extrema importância para aprimoramentos contínuos e para a satisfação do usuário final [Barbosa et al. 2022].

Dada essa necessidade, a comunidade científica e profissional começou a desenvolver conjuntos de dados para apoiar os profissionais da área. No entanto, embora existam *datasets* — que são coleções estruturadas de dados utilizadas para análise, pesquisa ou treinamento de modelos de aprendizado de máquina — a disponibilidade desses conjuntos em português ainda é limitada. [Alcoforado et al. 2022, Sanches et al. 2022]. Optar pela tradução pode ser desvantajoso, uma vez que muitas vezes palavras e expressões em diferentes idiomas possuem nuances e significados específicos que não podem ser totalmente capturados em uma tradução direta [Oliver 2020, Moreira et al. 2023].

Este trabalho busca suprir essa lacuna ao apresentar um novo *dataset* em português brasileiro composto por 3.000 avaliações de usuários, extraídas de 10 aplicativos populares da Google Play Store. A seleção dos apps considerou critérios de relevância e popularidade, abrangendo diferentes categorias e perfis de usuários. As avaliações foram manualmente classificadas em sete emoções básicas (felicidade, surpresa, tristeza, medo, nojo, raiva e neutro) através de um processo de anotação colaborativa, garantindo a qualidade e confiabilidade dos dados. O *dataset* está disponível na plataforma Zenodo, a partir do link `https://zenodo.org/records/10823148`, e visa contribuir com o desenvolvimento de pesquisas e ferramentas de análise de sentimentos em português, com aplicações em áreas como interação humano-computador, marketing e aprimoramento de software.

Este artigo organiza-se da seguinte forma: na Seção 2, são elencados trabalhos relacionados. Na Seção 3, são dados detalhes sobre as etapas para composição do *dataset*, sua organização e propriedades quantitativas compreendidas a partir do conjunto resultante. Na Seção 4, são listadas as possibilidades de utilização, desafios e limitações encontrados, e melhorias previstas para trabalhos futuros. Por fim, na Seção 5, é apresentada uma contextualização final do escopo de trabalho explorado no presente artigo.

2. Trabalhos Relacionados

A análise de sentimentos em reviews de aplicativos e plataformas online tem ganhado crescente atenção, impulsionada pela riqueza de informações que esses dados fornecem sobre a experiência e as opiniões dos usuários. Diversos estudos exploram essa área, com diferentes focos, abordagens e idiomas.

Saif et al. (2013) abordam o desafio da identificação de emoções em tweets, um contexto similar ao de reviews de apps em termos de informalidade e concisão da linguagem. Eles analisaram oito conjuntos de dados públicos de tweets anotados manualmente com diferentes categorias de sentimentos e emoções e propuseram um novo dataset, o

STS-Gold, para superar as limitações dos datasets anteriores. Uma das lacunas identificadas por Saif et al. (2013) é a falta de distinção entre o sentimento geral do tweet e o sentimento direcionado a entidades específicas [Saif et al. 2013]. A presente pesquisa aborda essa questão de forma similar ao STS-Gold, permitindo a anotação de diferentes emoções para o review como um todo e para as entidades mencionadas, o que possibilita uma análise mais precisa dos sentimentos dos usuários em relação a diferentes aspectos do aplicativo.

Mora e Lavid-López (2018) investigam a expressão de sentimentos em reviews de apps em inglês e espanhol utilizando a teoria linguística de "Appraisal", que se concentra na análise da linguagem para identificar atitudes, julgamentos e emoções expressas nos textos. Os autores criaram um sistema de anotação para identificar diferentes categorias de "Appraisal" em reviews e realizaram experimentos com anotadores humanos para validar o sistema [Mora e Lavid-López 2018]. Essa pesquisa, embora aprofunde a análise linguística, limita-se ao inglês e espanhol e não explora a classificação granular de emoções, como a presente pesquisa, que se concentra no português brasileiro e utiliza sete categorias de emoções básicas.

Umer et al. (2021) investigam a discrepância entre as avaliações numéricas (estrelas) e o conteúdo textual dos reviews na Google Play Store, utilizando deep learning para prever a nota "real" com base no texto, revelando um viés de cerca de 25% para cima nas notas atribuídas pelos usuários [Sadiq et al. 2021]. Enquanto Umer et al. (2021) focam na detecção de vieses nas avaliações, a presente pesquisa busca analisar as emoções expressas nos reviews. A predominância de emoções negativas em nosso dataset corrobora a hipótese de Umer et al. (2021) de que as notas numéricas podem ser inflacionadas, indicando que os usuários podem expressar insatisfação de forma mais sutil no texto, mesmo quando atribuem notas altas.

Saputra et al. (2023) constroem um dataset de sentimentos e emoções em indonésio a partir de reviews de apps, evidenciando a necessidade de datasets em diferentes idiomas e contextos culturais. O dataset, com mais de 21 mil avaliações, foi anotado manualmente com três categorias de sentimentos e seis categorias de emoções. Essa pesquisa, assim como a nossa, ressalta a importância de datasets em idiomas além do inglês [Saputra et al. 2023]. No entanto, diferentemente do trabalho de Saputra et al. (2023), que foca em um idioma específico, a presente pesquisa aborda o português brasileiro, contribuindo para a área de análise de sentimentos em um idioma com grande representatividade online, mas ainda pouco explorado nesse contexto.

Freire et al. (2023) apresentam o AirBSet, um dataset com dados de imóveis do Airbnb em 18 cidades brasileiras, incluindo anúncios, avaliações e comentários, destacando a escassez de datasets sobre o Airbnb no Brasil e a necessidade de dados para pesquisas sobre impactos da plataforma [Freire et al. 2023]. Assim como Freire et al. (2023) identificaram uma lacuna de dados sobre o Airbnb no Brasil, a presente pesquisa visa suprir a carência de datasets de análise de sentimentos em português. Ambos os trabalhos demonstram a importância de se construir e disponibilizar datasets que reflitam o contexto brasileiro, impulsionando pesquisas e aplicações em áreas relevantes para o país.

3. Metodologia

Nesta seção, são apresentadas as etapas para a criação do *dataset* anotado de avaliações providas por usuários de aplicativos móveis. A Figura 1 apresenta essas etapas de forma geral. A primeira etapa consistiu na seleção da *Google Play Store* como a loja de aplicativos a ser analisada. Em seguida, foram escolhidos dez aplicativos da seção em destaque. Posteriormente, foram coletadas 300 avaliações de cada aplicativo selecionado, totalizando 3.000 avaliações. Após a coleta das avaliações, foram realizadas as classificações pelos anotadores. Em seguida, as etapas de processamento que foram executadas, culminando na consolidação do **dataset**.

Figura 1. Processo de desenvolvimento do *dataset*

3.1. Coleta de Avaliações da Play Store

A Google Play Store foi escolhida como plataforma de estudo por abrigar um público de usuários brasileiros consideravelmente expressivo e relevante no cenário global de desenvolvimento de apps. Com o sistema Android presente em quase 80% dos smartphones no Brasil [Statista 2023] em 2023, a Play Store se torna um espaço crucial para entender as opiniões, expectativas e reações dos usuários a aplicativos, orientando decisões estratégicas de desenvolvimento e aprimoramento. A análise de sentimentos nesse contexto, portanto, ganha ainda mais relevância para empresas que buscam atender às demandas específicas do mercado brasileiro.

Para garantir a representatividade do dataset dentro da plataforma, foram selecionados os 10 aplicativos mais baixados na categoria "Em Alta" em maio de 2023. Essa categoria reflete os apps com maior crescimento em número de downloads no período, representando tendências e preferências dos usuários brasileiros. Os apps selecionados abrangem diferentes categorias, como redes sociais, finanças, edição de fotos e utilitários: Shopee, SHEIN, TikTok Lite, Nubank, Instagram, Photo & File Detect, Whatsapp Messenger, Canva, CapCut e Gov.br.

A coleta das avaliações foi realizada por meio da linguagem de programação Python[1], utilizando a biblioteca *google-play-scraper*[2], sendo executado em uma maquina local. Essa biblioteca é responsável por fazer a captação das avaliações disponíveis na plataforma Google Play Store, possibilitando a obtenção de informações diretamente dos usuários. Por meio desta ferramenta, é possível analisar as percepções e respostas dos usuários, fornecendo informações importantes para a análise da experiência de usuário em relação ao aplicativo.

A Figura 2 mostra um exemplo de como a coleta de avaliações é feita. Na linha 1, é feita a importação de dois elementos da biblioteca *google-play-scraper*: *Sort* e *reviews*.

[1]https://www.python.org
[2]https://pypi.org/project/google-play-scraper/

```python
from google_play_scraper import Sort, reviews
appId = 'com.shopee.br'
br_reviews = reviews(
    appId,
    lang='pt',
    country='br',
    sort=Sort.MOST_RELEVANT,
    count=300,
)
```

Figura 2. Coleta de Avaliações da Play Store

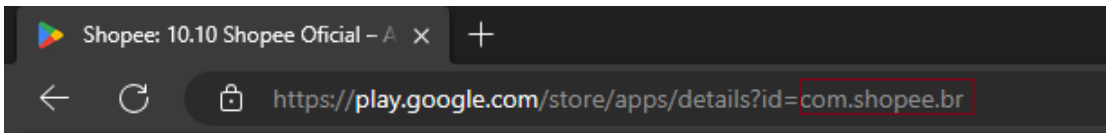

Figura 3. Id do Aplicativo

Sort é uma enumeração das formas de escolher as avaliações, e *reviews* é uma função que permite a coleta de avaliações na Google Play Store. Na linha 3, é chamada a função *review*, para qual são passados os seguintes parâmetros:

- **appId:** Primeiro parâmetro, no qual é colocado o ID encontrado na URL da Play Store. A Figura 3 demonstra de onde é realizada a coleta desse ID;
- ***lang:*** Segundo parâmetro, no qual é definido o idioma como "pt", correspondendo às avaliações em português;
- ***country:*** Terceiro parâmetro, no qual é definido o país como "br", correspondendo ao Brasil;
- ***sort:*** Quarto parâmetro, no qual é escolhida uma das opções de filtragem, podendo optar pelas mais relevantes (*Sort.MOST_RELEVANT*) ou pelas mais novas (*Sort.NEWEST*);
- ***count:*** Quinto parâmetro, definindo a quantidade de avaliações que serão coletadas.

A Figura 4 mostra o retorno da função *reviews*, a partir da qual é possível observar uma série de informações. Cada retorno é um dicionário com diversos campos, sendo estes: o identificador da avaliação (reviewId), imagem de usuário (userImage), conteúdo da avaliação (content), pontuação atribuída ao aplicativo (score), contagem de "polegares para cima" (thumbsUpCount), versão do aplicativo que o usuário estava utilizando (reviewCreatedVersion), data e hora em que a avaliação foi feita (at), entre outras. O campo *content* é particularmente interessante para este trabalho, pois contém o texto produzido pelo usuário, onde podemos encontrar suas experiências, opiniões e sugestões.

3.2. Organização dos Dados

A organização dos dados também é realizada utilizando a linguagem Python, com o auxílio da biblioteca Pandas[3]. Essa biblioteca oferece uma abordagem rápida, flexível e

[3]https://pandas.pydata.org/

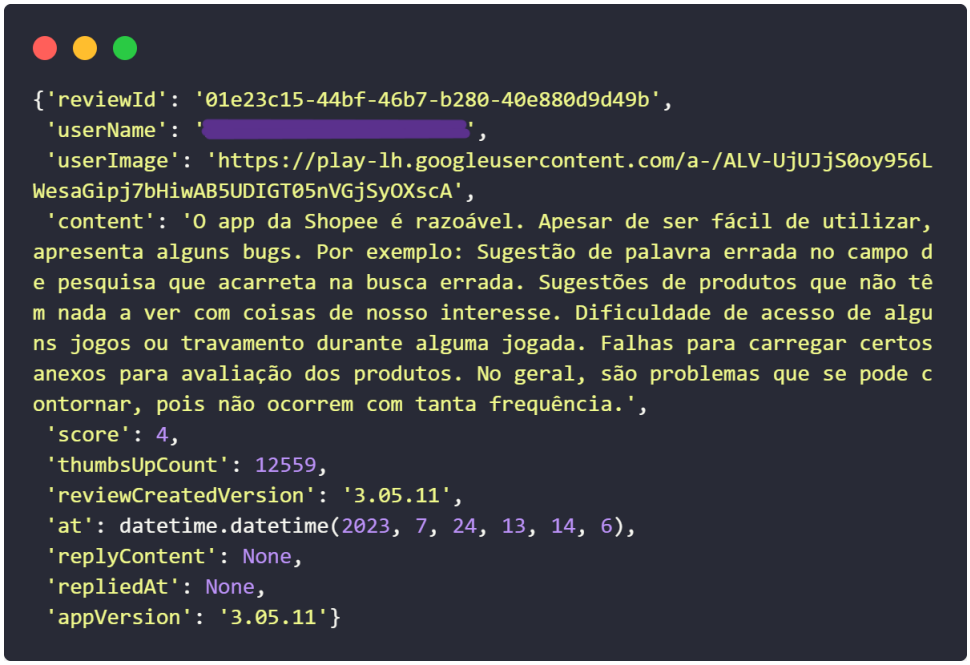

Figura 4. Retorno da função *reviews*

intuitiva para trabalhar com dados relacionais (rotulados), muito utilizada na área de Ciência de Dados[4]. O Pandas possui estruturas robustas que facilitam a manipulação e análise de dados, constituindo-se assim como uma ferramenta eficiente para tarefas como processamento e exploração de informações [Almeida 2023].

Após realizar a coleta das avaliações dos usuários, utilizando a função da Figura 2, os dados são organizados de forma mais compreensível com a ajuda da classe *DataFrame* do *pandas*.

O *DataFrame* é uma estrutura de dados tabular bidimensional que permite uma maneira eficiente de manipular e visualizar informações. A Figura 5 ilustra a organização dos dados.

	reviewId	content	appId	sentiment_polarity	sentiment
0	01e23c15-44bf-46b7-b280-40e880d9d49b	O app da Shopee é razoável. Apesar de ser fáci...	com.shopee.br		
1	df40ba7c-075d-48dc-b798-21466dfb41d6	O app é bem otimizado, fácil de visualizar os ...	com.shopee.br		
2	4669ecf6-4887-4505-98d7-cdddc759134b	Pontos positivos: Facilidade de compra, acessi...	com.shopee.br		

Figura 5. Avaliações de Usuários no *DataFrame*

3.3. Classificação das Avaliações

Após a estruturação dos dados, o próximo passo consiste na anotação das emoções identificadas nas avaliações, os quais serão classificados de acordo com as emoções básicas discutidas por Ekman e Cordaro (2011).

As avaliações foram classificadas conforme estão apresentadas na Tabela 1:

[4]A **Ciência de Dados** pode ser descrita como a extração de conhecimento a partir de dados utilizando análise estatística e aprendizado de máquina, [Amaral 2016].

Polaridade	**Positivo**			**Neutro**	**Negativo**		
Emoção	Felicidade	Surpresa	Tristeza	Neutro	Medo	Nojo	Raiva

Tabela 1. Tabela de emoções e polaridades

As avaliações foram classificadas com base na estrutura proposta por Britzolakis *et al.* (2021), que divide as emoções entre polaridades: as emoções positivas incluem felicidade, surpresa e tristeza; as negativas englobam medo, nojo e raiva; e, por fim, as neutras, conforme apresentado na Tabela 1.

Primeiramente, o autor deste trabalho fez a classificação das 3.000 avaliações. Cada uma das avaliações foi lida e classificada manualmente. As interpretações de como as avaliações foram feitas, dado o contexto deste trabalho, foram as seguintes:

- **Felicidade**: Representa uma experiência gratificante e positiva para o usuário ao usar o aplicativo;
- **Surpresa**: Indica uma reação inesperada e positiva do usuário durante a utilização do aplicativo;
- **Tristeza**: Reflete o sentimento de descontentamento ou a ausência de uma característica desejada no aplicativo, mas mantém uma apreciação geral pelo aplicativo;
- **Neutro**: Caracteriza a neutralidade na avaliação, com a ausência de manifestações emocionais;
- **Medo**: Sinaliza a presença de medo ou insegurança por parte do usuário relacionado ao uso do aplicativo;
- **Nojo**: Descreve uma sensação desagradável causada por uma característica do aplicativo, prejudicando a experiência do usuário;
- **Raiva**: Atribuída quando há um nível notável de agressividade na avaliação do usuário.

3.4. Validação das Classificações

Após as 3.000 classificações realizadas pelo autor deste trabalho, deu-se início ao processo de validação. Para isso, outras quatro pessoas da área de computação participaram do processo, sendo esses dois estudantes e dois professores orientadores, totalizando cinco envolvidos.

Primeiramente, foi estabelecido um material introdutório para os participantes, juntamente com uma reunião sobre o contexto do trabalho. Nesta reunião, foi apresentado o material, que continha as 3.000 avaliações a serem classificadas, acompanhadas por um documento que explicava as interpretações da Seção 3.3. Durante esta reunião, na qual os materiais foram distribuídos e as dúvidas foram esclarecidas, ficou acordado que os dois voluntários conduziriam seus trabalhos de forma independente, evitando qualquer comunicação entre eles sobre suas classificações para evitar possíveis vieses.

Dessa forma, deu-se início o processo de validação das classificações pelos dois participantes. O processo de validação adotou a abordagem da maioria de votos, em que uma avaliação precisava ser classificada por pelo menos dois dos três classificadores para receber uma emoção específica.

Reuniões semanais foram agendadas para discutir as avaliações que apresentaram discordância entre os três classificadores. Quando uma avaliação recebia classificações

diferentes, os classificadores discutiam suas interpretações e razões para a classificação, realizando uma segunda tentativa de classificação. Em casos de persistência na discordância, as classificações eram levadas aos professores orientadores, que forneciam suas próprias percepções sobre as classificações discordantes. Em seguida, chegava-se a um consenso. Esse processo de discussão de avaliação demorou em torno de 4 semanas, e ao todo foram realizadas 12 reuniões.

3.5. Análise de Concordância entre Anotadores

Após os anotadores concluírem suas classificações, foram obtidas três versões do *dataset*, cada uma representando a interpretação de seu respectivo anotador. Para avaliar a concordância entre eles, adotou-se o método Fleiss' kappa, que utiliza métodos estatísticos para medir o nível de acordo entre diferentes avaliadores, [Fleiss 1971]. A Tabela 2 mostra os possíveis níveis de concordância entre os anotadores de acordo com o método Fleiss' kappa.

Kappa	Nível de Concordância
> 0,8	Quase perfeito
> 0,6	Substancial
> 0,4	Moderado
> 0,2	Razoável
> 0	Pouco
< 0	Sem concordância

Tabela 2. Níveis de concordância de Fleiss Kappa [Fleiss 1971].

A abordagem adotada neste trabalho consistiu em calcular o Kappa de Fleiss para as classificações realizadas pelos anotadores. Primeiramente, as classificações dos anotadores foram organizadas em planilhas no Google Drive[5], onde cada linha representava uma avaliação e as colunas representam as classificações dos anotadores, conforme mostra a Figura 6. Em seguida, o Kappa de Fleiss foi calculado para avaliar a concordância na polaridade das emoções. Posteriormente, foi realizado o cálculo para analisar a concordância das emoções atribuídas. Por fim, foi calculada a média e a mediana dos valores obtidos para os 10 aplicativos avaliados.

3.6. Classificações

A Tabela 3 resume a classificação das emoções encontradas nas avaliações deste *dataset*, organizando-as conforme sua polaridade. Os números listados na tabela indicam o total de ocorrências de cada emoção e polaridade.

Polaridade	Negativo			Neutro	Positivo		
Emoções	Raiva	Nojo	Medo		Tristeza	Surpresa	Felicidade
Total de emoções	743	952	47	82	864	4	319
Total de polaridades	1742			82	1187		

Tabela 3. Classificações de polaridades e emoções

A maioria das classificações de polaridade e emoção se concentra em categorias negativas e positivas, com uma proporção menor de classificações neutras. Entre as emoções negativas, **raiva** e **nojo** foram as mais frequentemente identificadas, com **743** e **952**

[5]https://drive.google.com/drive/home

1	Anotador 1	Anotador 2	Anotador 3
2	negative	positive	neutral
3	positive	positive	positive
4	positive	negative	positive
5	negative	positive	positive
6	negative	negative	negative
7	positive	positive	positive
8	positive	positive	positive
9	positive	positive	positive
10	positive	positive	positive
11	negative	negative	negative
12	positive	positive	positive
13	positive	positive	positive
14	positive	negative	positive
15	positive	positive	positive
16	positive	positive	positive
17	positive	positive	positive
18	positive	negative	positive
19	positive	positive	positive
20	negative	negative	negative

Figura 6. Classificações das polaridades

ocorrências, respectivamente, enquanto **medo** foi mencionado em **47** classificações. Em contraste, **tristeza** e **felicidade** se destacaram entre as emoções positivas, com **864** e **319** classificações, respectivamente, enquanto **surpresa** foi identificada em apenas **4** casos. Além disso, o total de 3.011 classificações reflete comentários nos quais houve uma distinta mudança temporal e de opinião. Usuários retornaram as avaliações anteriores para expressar suas opiniões, frequentemente contradizendo pontos anteriormente feitos. Depois de anonimizado, o *dataset* foi publicado na plataforma Zenodo, conforme link disponibilizado na introdução deste documento [Siqueira et al. 2024].

4. Limitações

O dataset, composto por 3.000 avaliações de 10 aplicativos da Google Play Store, pode não capturar toda a diversidade de aplicativos e experiências dos usuários, limitando a generalização dos resultados. Além disso, a coleta direta das avaliações pode introduzir um viés de seleção, onde usuários insatisfeitos tendem a ser mais propensos a deixar avaliações do que usuários satisfeitos. A disponibilidade variável e as mudanças na Google Play Store ao longo do tempo podem também comprometer a replicabilidade de construção do dataset. Adicionalmente, os resultados podem não ser diretamente aplicáveis a outras plataformas de distribuição de aplicativos devido a diferenças nas políticas de avaliação e base de usuários. Outro ponto importante é que a anotação das emoções foi realizada por cinco pessoas da área de computação. A inclusão de especialistas em Psicologia forneceria uma perspectiva mais ampla e interdisciplinar, o que poderia enriquecer a análise emocional e comportamental [Lunardi et al. 2018, Lunardi 2019].

5. Considerações Finais

Este trabalho apresentou a construção de um dataset em português brasileiro para análise de sentimentos, composto por 3.000 avaliações de usuários coletadas de 10 aplicativos populares da Google Play Store. A metodologia empregada, desde a coleta e organização dos dados até a classificação manual das emoções em sete categorias básicas e a validação por múltiplos anotadores, demonstra o rigor e a confiabilidade do processo de construção do *dataset*. A predominância de emoções negativas, observada na análise preliminar dos dados, abre um leque de oportunidades para pesquisas futuras que investiguem a relação entre as emoções expressas pelos usuários e a sua satisfação com os aplicativos.

Indo além da categorização de sentimentos, este *dataset* pode ser aplicado em pesquisas que explorem a relação entre emoções, linguagem e comportamento do usuário em plataformas digitais nas mais diversas áreas como interação humano-computador, marketing direcionado e aprimoramento de software. Por exemplo, pode-se investigar como a expressão de diferentes emoções se correlaciona com a nota atribuída ao aplicativo, a extensão do texto da avaliação, ou até mesmo a probabilidade de um usuário recomendar o app para outros. A partir desses dados, novas ferramentas de análise de sentimentos poderão ser desenvolvidas, com o potencial de auxiliar desenvolvedores a entenderem a fundo as necessidades de seus usuários e a tomarem decisões mais estratégicas para o aprimoramento de seus produtos.

Como perspectivas futuras, pretende-se expandir o *dataset* com a inclusão de novas categorias de apps e períodos de coleta, visando abranger um espectro mais amplo de comportamentos dos usuários. Além disso, a exploração de técnicas de aprendizado de máquina para automatizar parte do processo de anotação poderá aumentar a escalabilidade da metodologia, enriquecendo ainda mais este recurso e impulsionando novas descobertas na área de análise de sentimentos em português.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Williamson Silva agradece pelo apoio financeiro da FAPERGS (Projeto ARD/ARC – processo 22/2551-0000606-0). Os autores gostariam de agradecer a todos os participantes voluntários do estudo pelo apoio e também ao apoio financeiro Edital nº 122/2022 (PROBIC/FAPERGS). Gabriel Machado Lunardi agradece pelo apoio financeiro da FAPERGS (Projeto ARD/ARC – processo 24/2551-0000645-1) e do CNPq (Projetos Universais 405973/2021-7 e 402086/2023-6).

Referências

Alcoforado, A., Ferraz, T. P., Gerber, R., Bustos, E., Oliveira, A. S., Veloso, B. M., Siqueira, F. L., e Costa, A. H. R. (2022). Zeroberto: Leveraging zero-shot text classification by topic modeling. Em *International Conference on Computational Processing of the Portuguese Language*, páginas 125–136. Springer.

Almeida, M. (2023). Pandas python: o que é, para que serve e como instalar.

Amaral, F. (2016). *Introdução à ciência de dados: mineração de dados e big data*. Alta Books Editora.

Barbosa, M., Nakamura, W., Valle, P., Guerino, G., Finger, A., Lunardi, G. M., e Silva, W. (2022). Ux of chatbots: An exploratory study on acceptance of user experience evaluation methods. Em *ICEIS*, volume 2, páginas 355–363.

Fabro, C. (2021). Google Play Store: conheça seis curiosidades sobre a loja de aplicativos. `https://tinyurl.com/y86b9w53`. Acessado em: 27/08/2024.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Freire, J., Costa, L. H., Dorneles, C., e Brandão, M. (2023). Airbset: Um conjunto de dados com imóveis brasileiros do airbnb e respectivas avaliações. Em *Anais do V Dataset Showcase Workshop*, páginas 79–86, Porto Alegre, RS, Brasil. SBC.

Guzman, E. e Maalej, W. (2014). How do users like this feature? a fine grained sentiment analysis of app reviews. Em *2014 IEEE 22nd international requirements engineering conference (RE)*, páginas 153–162. Ieee.

Lunardi, G. M. (2019). Representing the filter bubble: Towards a model to diversification in news. Em *International Conference on Conceptual Modeling*, páginas 239–246. Springer, Cham.

Lunardi, G. M., Machado, G. M., Al Machot, F., Maran, V., Machado, A., Mayr, H. C., Shekhovtsov, V. A., e de Oliveira, J. P. M. (2018). Probabilistic ontology reasoning in ambient assistance: Predicting human actions. Em *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, páginas 593–600.

Maalej, W. e Nabil, H. (2015). Bug report, feature request, or simply praise? on automatically classifying app reviews. Em *2015 IEEE 23rd international requirements engineering conference (RE)*, páginas 116–125. IEEE.

Mora, N. e Lavid-López, J. (2018). Building an annotated dataset of app store reviews with appraisal features in english and spanish. Em *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, páginas 16–24.

Moreira, L. S., Lunardi, G. M., de Oliveira Ribeiro, M., Silva, W., e Basso, F. P. (2023). A study of algorithm-based detection of fake news in brazilian election: Is bert the best. *IEEE Latin America Transactions*, 21(8):897–903.

Motger, Q., Franch, X., Gervasi, V., e Marco, J. (2024). Unveiling competition dynamics in mobile app markets through user reviews. Em *International Working Conference on Requirements Engineering: Foundation for Software Quality*, páginas 251–266. Springer.

Nakamura, W. T., de Oliveira, E. C., de Oliveira, E. H., Redmiles, D., e Conte, T. (2022). What factors affect the ux in mobile apps? a systematic mapping study on the analysis of app store reviews. *Journal of Systems and Software*, 193:111462.

Oliver, A. (2020). Human translation and machine translation: Specificities, uses, advantages and disadvantages.

Quiroga, F. L. e Bessa, R. d. (2024). A educação em tempos de smartphones e redes sociais: por uma crítica permanente no enfrentamento da dessubjetivação e monitoramento. *Texto Livre*, 17:e51341.

Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., e Nappi, M. (2021). Discrepancy detection between actual user reviews and numeric ratings of google app store using deep learning. *Expert Systems with Applications*, 181:115111.

Saif, H., Fernandez, M., He, Y., e Alani, H. (2013). Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold.

Sanches, M., de Sá, J., Foerste, H., Souza, R., Reis, J. D., e Villas, L. (2022). Textual datasets for portuguese-brazilian language models. Em *Anais do IV Dataset Showcase Workshop*, páginas 1–12, Porto Alegre, RS, Brasil. SBC.

Saputra, K. E. et al. (2023). Multilabel multiclass sentiment and emotion dataset from indonesian mobile application review. *Data in Brief*, 50:109576.

Siqueira, V., M. Lunardi, G., e Silva, W. (2024). A dataset of polarities and emotions from brazilian portuguese play store reviews.

Statista (2023). Market share of mobile operating systems in brazil from january 2019 to may 2023. `https://www.statista.com/statistics/262167/market-share-held-by-mobile-operating-systems-in-brazil/`. Acessado em: 27/08/2024.