PlayReviewsPT: Um Dataset de 97 Milhões de Avaliações do Google Play em Português Brasileiro

Tiago de Melo[1]

[1]Universidade do Estado do Amazonas (UEA)
Manaus – AM – Brasil

tmelo@uea.edu.br

Abstract. *This paper presents PlayReviewsPT, a public dataset of 97.2 million Brazilian Portuguese reviews from 96 Google Play applications (2010–2025). The data are released in JSONL format, partitioned by application, with structured metadata and the collector source code. We describe the collection methodology, the dataset schema, and an initial quantitative characterization. The analyses reveal a long-tail distribution across applications, strong rating polarization, a predominance of short comments, and steady growth in review volume while rating proportions remain stable. PlayReviewsPT is a large-scale resource for natural language processing, app review mining, and empirical software engineering.*

Resumo. *Este trabalho apresenta o PlayReviewsPT, um dataset público com 97,2 milhões de avaliações em português brasileiro de 96 aplicativos do Google Play (2010–2025). Os dados são disponibilizados em formato JSONL, particionados por aplicativo, com metadados estruturados e o código do coletor. O artigo descreve a metodologia de coleta, o esquema dos dados e uma caracterização inicial. As análises mostram concentração em poucos aplicativos, forte polarização das notas, predominância de comentários curtos e crescimento expressivo do volume ao longo do tempo. O PlayReviewsPT é um recurso em larga escala para processamento de linguagem natural, mineração de avaliações de aplicativos e engenharia de software empírica.*

1. Introdução

Plataformas de distribuição de aplicativos móveis, como o Google Play, representam um dos maiores repositórios de *feedback* de usuários gerados organicamente em escala global [Harman et al. 2012]. Estes comentários constituem uma fonte relevante de dados para pesquisa em múltiplos domínios: processamento de linguagem natural (PLN), análise de satisfação de usuários, engenharia de software empírica e identificação de padrões comportamentais em comunidades digitais. A capacidade de processar, analisar e extrair conhecimento deste volume massivo de comentários é cada vez mais crítica para desenvolvedores de software, pesquisadores e formuladores de políticas em plataformas digitais.

Contudo, existe um desequilíbrio significativo na disponibilidade de *datasets* públicos para pesquisa em avaliações de aplicativos. Recursos de qualidade concentram-se predominantemente no idioma inglês, deixando pesquisadores e desenvolvedores de

contextos linguísticos minoritários com acesso limitado a dados comparáveis em seus idiomas nativos. O português brasileiro, falado por aproximadamente 220 milhões de pessoas, continua classificado como "*low-resource language*" pela comunidade de processamento de linguagem natural, apesar de sua importância econômica e demográfica [Pereira 2021]. Esta lacuna é particularmente crítica para: (1) desenvolvimento de ferramentas robustas de PLN em português; (2) validação de modelos multilíngues em contextos linguísticos específicos; (3) pesquisa sobre padrões de comportamento de usuários em ecossistemas digitais brasileiros; (4) capacitação de pesquisadores brasileiros em análise de grandes volumes de dados não-estruturados.

Este trabalho apresenta o PlayReviewsPT, um dataset contendo 97,2 milhões de avaliações originárias de 96 aplicativos populares coletados entre 2010 e 2025, mantido após a rigorosa validação de qualidade. O *dataset* abrange cinco categorias principais (Redes Sociais, Jogos, Utilitários, Streaming, Produtividade), capturando a diversidade do ecossistema de aplicativos móveis brasileiro. Os dados estão disponíveis publicamente no Hugging Face, com documentação técnica completa, formato normalizado (JSONL) e metadados estruturados, viabilizando reprodutibilidade e reutilização.

Neste artigo, o termo avaliação refere-se ao registro completo extraído do Google Play, incluindo texto, nota (*rating*), data e metadados associados. O termo comentário é utilizado especificamente para denotar o conteúdo textual escrito pelo usuário, enquanto nota (*rating*) se refere ao valor numérico de 1 a 5 estrelas atribuído à avaliação.

A Figura 1 ilustra a natureza e a estrutura dos dados contidos no PlayReviewsPT, mostrando um exemplo real de avaliação do app Amazon Shopping. O exemplo demonstra os metadados estruturados, como usuário, data, *rating* e *feedback* negativo específico sobre problemas funcionais e engajamento comunitário (27 pessoas consideraram útil). Este tipo de dado é representativo do corpus de análise.

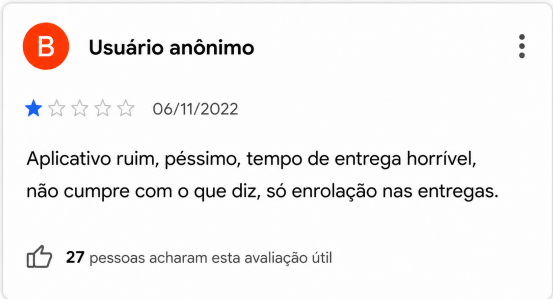


Figura 1. Exemplo autêntico de comentário do dataset PlayReviewsPT.

As contribuições principais deste trabalho são: (1) disponibilizar publicamente o PlayReviewsPT, um recurso de grande escala em português brasileiro, preenchendo uma lacuna significativa em repositórios de pesquisa multilíngues; (2) documentar padrões estruturais de *feedback* em aplicativos móveis (distribuição de cauda longa, polarização de ratings, características textuais); (3) analisar a dinâmica temporal, correlacionando o crescimento exponencial de volume com a notável estabilidade da satisfação dos usuários; e (4) estabelecer e documentar uma metodologia de coleta, normalização e organização dos dados, com disponibilização pública do código do coletor, viabilizando auditoria, reutilização e adaptação do processo por terceiros. O *dataset* viabiliza pesquisas futuras

em detecção de bugs através de mineração de comentários, extração de atributos, análise de sentimentos multilíngue, modelagem de tópicos em português e desenvolvimento de modelos de linguagem robustos para contextos específicos de domínio.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta limitações e considerações de uso do dataset. A Seção 3 discute possíveis usos do PlayReviewsPT. A Seção 4 descreve a metodologia de coleta e a caracterização estatística do dataset. A Seção 5 descreve a disponibilidade dos dados. Por fim, a Seção 6 traz as considerações finais.

2. Limitações e Considerações de Uso

O uso do PlayReviewsPT requer atenção a vieses inerentes ao dataset, como a concentração de avaliações em poucos aplicativos, a polarização das notas e a predominância de comentários curtos. Além disso, a seleção dos aplicativos e a coleta por *web scraping* podem introduzir limitações de cobertura. Estudos derivados devem considerar estratégias de amostragem, estratificação e pré-processamento adequadas, além das condições éticas e legais aplicáveis ao uso de conteúdo publicamente disponível.

3. Possíveis Usos do *Dataset*

3.1. Análise de Sentimentos em Plataformas de Apps

A análise de sentimentos em comentários de aplicativos é um domínio consolidado. Pagano e Maalej [Pagano and Maalej 2013] estabeleceram categorias fundamentais de *feedback* (*bug reports*, *feature requests*, UX, *ratings*). Estudos recentes [Samanmali and Rupasingha 2024] demonstram que redes LSTM superam SVM em classificação de sentimentos em avaliações do Google Play, com acurácia superior em análises de 33 mil avaliações de 15 aplicativos. Abordagens modernas de PLN viabilizam a análise automatizada em larga escala.

3.2. Mining de Reviews para Detecção de Bugs e Feedback

Iacob e Harrison [Iacob and Harrison 2013] validaram que as avaliações contêm informações valiosas sobre bugs e solicitações de funcionalidades. Harman *et al*. [Harman et al. 2012] argumentaram que a mineração de opinião em app stores é essencial na era do big data. Sorbo *et al*. [Di Sorbo et al. 2016] confirmaram a necessidade de plataformas escaláveis para a análise de feedback. Pesquisas recentes [Khan et al. 2024] demonstram que comentários de usuários frequentemente revelam *bugs* e problemas de performance antes de testes formais, com padrões comuns identificáveis por mineração.

3.3. Análise Temporal e Comparação de Aplicativos

Panichella *et al*. [Panichella et al. 2013] demonstraram a efetividade de modelagem de tópicos (*Latent Dirichlet Allocation* - LDA) para organizar feedback não-estruturado em categorias significativas, validando a aplicabilidade de aprendizado de máquina em larga escala. Análises comparativas entre aplicativos similares revelam diferenciais competitivos em satisfação dos usuários, conforme documentado em uma revisão sistemática de mineração de opinião em app stores [Genc-Nayebi and Abran 2017].

3.4. Datasets Multilíngues e Lacuna em Português Brasileiro

Enquanto datasets em inglês proliferam, recursos em português brasileiro são escassos. O português é classificado como "*low-resource language*" na comunidade de PLN, apesar de haver mais de milhões de falantes [Pereira 2021]. Desenvolvimentos recentes (2024-2025) incluem BERTimbau [Souza et al. 2020], PTT5-v2 [Piau et al. 2024] e GigaVerbo[1]. Contudo, até onde se sabe, não há um *dataset* público de avaliações do Google Play em português brasileiro com escala comparável à apresentada neste trabalho.

4. Descrição e Caracterização do Dataset

4.1. Metodologia de Coleta e Processamento

Os dados foram coletados através de web scraping automatizado da Google Play Store entre 2010 e 2025, abrangendo 96 aplicativos selecionados de diversas categorias. A escolha dos aplicativos deve-se à sua popularidade na loja do Google Play. A coleta foi implementada em Python com a biblioteca *google-play-scraper*, configurada com o parâmetro de localidade português do Brasil (`lang='pt', country='br'`), o que garante na origem que as avaliações recuperadas estejam em português brasileiro. As avaliações foram obtidas de forma paginada, das mais recentes para as mais antigas. Comentários brutos foram armazenados no formato JSONL, ocupando aproximadamente 37,0 GB de armazenamento em disco.

O código utilizado para a coleta dos dados também é disponibilizado publicamente junto ao *dataset*, no diretório coletor do repositório no Hugging Face[2]. Essa disponibilização tem como objetivo facilitar a auditoria do processo de coleta, a reprodutibilidade metodológica e a adaptação da ferramenta para futuras coletas em outros conjuntos de aplicativos. A coleta foi conduzida com mecanismos de limitação de requisições, incluindo intervalos aleatórios entre acessos, com o objetivo de reduzir o impacto sobre os servidores. O uso do *dataset* deve observar as condições legais e éticas aplicáveis à reutilização de conteúdo publicamente disponível. O dataset compreende 97.165.303 avaliações em português brasileiro. Os aplicativos foram categorizados em cinco grupos principais de acordo com sua funcionalidade. A Tabela 1 apresenta a distribuição das avaliações e o número de aplicativos por categoria.

4.2. Esquema e Dicionário de Dados

O PlayReviewsPT é armazenado em formato JSONL, no qual cada linha corresponde a uma avaliação individual de um aplicativo. Os arquivos são particionados por aplicativo, de modo que o aplicativo de origem pode ser identificado a partir do nome do arquivo. Por exemplo, o arquivo `facebook.jsonl` contém avaliações associadas ao aplicativo Facebook. Essa organização permite o processamento incremental dos dados e o download seletivo de subconjuntos específicos do dataset.

Cada registro contém o texto da avaliação, a nota atribuída pelo usuário, informações temporais e metadados relacionados à versão do aplicativo e à interação da comunidade com a avaliação. A Tabela 2 apresenta o dicionário de dados adotado no PlayReviewsPT.

[1]https://huggingface.co/datasets/TucanoBR/GigaVerbo-Text-Filter

[2]https://huggingface.co/datasets/UEA-LSI/googleplay-reviews-ptbr/tree/main/coletor

Tabela 1. Distribuição de avaliações por categoria de aplicativo

Categoria	Avaliações	Qtd. apps	Apps
Redes Sociais	41.848.527 (43,07%)	17	Bigo Live, Discord, Facebook, Instagram, Line, Messenger, Pinterest, Reddit, Snapchat, Telegram, Threads, Tiktok, Tinder, Twitter, Viber, Whatsapp, Whatsapp Business
Jogos	25.989.781 (26,75%)	23	Bowmasters, Candy Crush, Cats Crash Arena, Clash Of Clans, Clash Royale, Ea Sports Fc Mobile, Free Fire, Gardenscapes, Minecraft Pe, Mobile Legends, My Talking Tom, My Talking Tom Friends, Plants Vs Zombies, Pokemon Go, Pou, Roblox, Slither Io, Sonic Dash, Stumble Guys, Subway Surfers, Temple Run, Traffic Racer, Traffic Rider
Utilitários	19.894.257 (20,47%)	35	Amazon Shopping, Camscanner, Canva, Capcut, Cash App, Chase Mobile, Clean Master, Eight Ball Pool, Google App, Google Calculator, Google Chrome, Google Maps, Google Phone, Google Photos, Google Play Services, Life360, Meesho, Mercado Libre, Moto Camera3, Myjio, Newsbreak, Opera Mini, Picsart, Qr Barcode Scanner, Samsung Clock, Samsung Internet, Samsung My Files, Samsung One Ui Home, Samsung Pay, Shareit, Shazam, Uc Browser, Walmart, Waze, Wish
Streaming	6.623.520 (6,82%)	5	Iqiyi, Netflix, Spotify, Youtube, Youtube Music
Produtividade	2.809.218 (2,89%)	16	Adobe Reader, Chatgpt, Gmail, Google Calendar, Google Contacts, Google Drive, Google Meet, Google Messages, Microsoft Outlook, Microsoft Teams, Microsoft Word, Samsung Calendar, Samsung Email, Samsung Notes, Yahoo Mail, Zoom
Total	**97.165.303 (100,00%)**	**96**	–

4.3. Análise da Distribuição de Cauda Longa

O dataset apresenta uma distribuição de frequência característica de cauda longa, também conhecida como distribuição de Pareto. A Figura 2 ilustra este padrão, combinando um gráfico de barras com uma curva de frequência acumulada.

A análise revela que a concentração de avaliações é significativamente desigual entre os aplicativos. Os dois aplicativos com maior volume (Instagram e WhatsApp) juntos acumulam 24,26% de todas as avaliações do *dataset* (23.573.831 registros). Este padrão intensifica-se quando se considera um número maior de aplicativos. Os cinco principais concentram 42,10% das avaliações, enquanto os dez aplicativos com maior

Tabela 2. Dicionário de dados da versão pública do PlayReviewsPT.

Campo	Tipo	Descrição
`reviewId`	string	Identificador da avaliação, podendo ser anonimizado por hash.
`content`	string	Texto da avaliação escrita pelo usuário.
`score`	integer	Nota atribuída ao aplicativo, variando de 1 a 5 estrelas.
`thumbsUpCount`	integer	Número de usuários que marcaram a avaliação como útil.
`reviewCreatedVersion`	string/null	Versão do aplicativo no momento da avaliação, quando disponível.
`at`	datetime	Data e hora de publicação da avaliação.
`replyContent`	string/null	Resposta do desenvolvedor à avaliação, quando disponível.
`repliedAt`	datetime/null	Data e hora da resposta do desenvolvedor, quando disponível.
`appVersion`	string/null	Versão do aplicativo associada à avaliação, quando disponível.
`appId`	string	Identificador do aplicativo.
`category`	string	Categoria atribuída ao aplicativo no dataset.

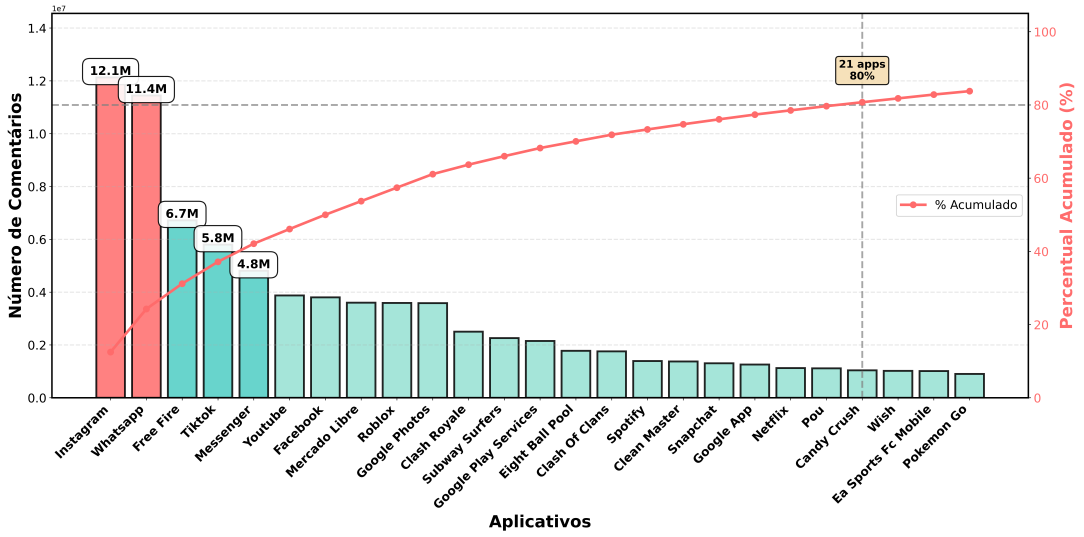


Figura 2. Distribuição de cauda longa das avaliações por aplicativo.

volume de avaliações representam 61,09% do total.

Observa-se ainda que aproximadamente 21 aplicativos (21,9% do total) são necessários para atingir 80% de todos os comentários, confirmando o princípio de Pareto clássico para este *dataset*. Inversamente, os 75 aplicativos restantes (78,1% da amostra) contribuem conjuntamente com apenas 20% do volume total de avaliações.

Esta distribuição desequilibrada tem implicações diretas para análises estatísticas do dataset. Métodos de análise que pressupõem distribuições uniformes ou normais podem gerar conclusões enviesadas se não contabilizarem adequadamente esse desequilíbrio. Além disso, a presença de uma cauda longa pronunciada sugere que diferentes padrões comportamentais podem emergir entre aplicativos predominantes versus aplicativos de nicho, recomendando estratificação ou ponderação nas análises subsequentes.

A magnitude desta concentração também reflete a dinâmica real do mercado de aplicativos móveis, onde poucos aplicativos dominam o uso massivo (e, consequentemente, as avaliações), enquanto uma longa cauda de aplicativos especializados ou de menor alcance completa o ecossistema.

4.4. Distribuição dos *Ratings*

A Figura 3 mostra que a distribuição de *ratings* no *dataset* reflete comportamentos bem documentados em estudos sobre avaliações de aplicativos móveis. Especificamente, observa-se uma concentração pronunciada nos extremos da escala (1 e 5 estrelas), consistente com o modelo *"speak with your feet"*, onde usuários insatisfeitos desinstalam aplicativos e usuários muito satisfeitos deixam avaliações positivas de forma preferencial.

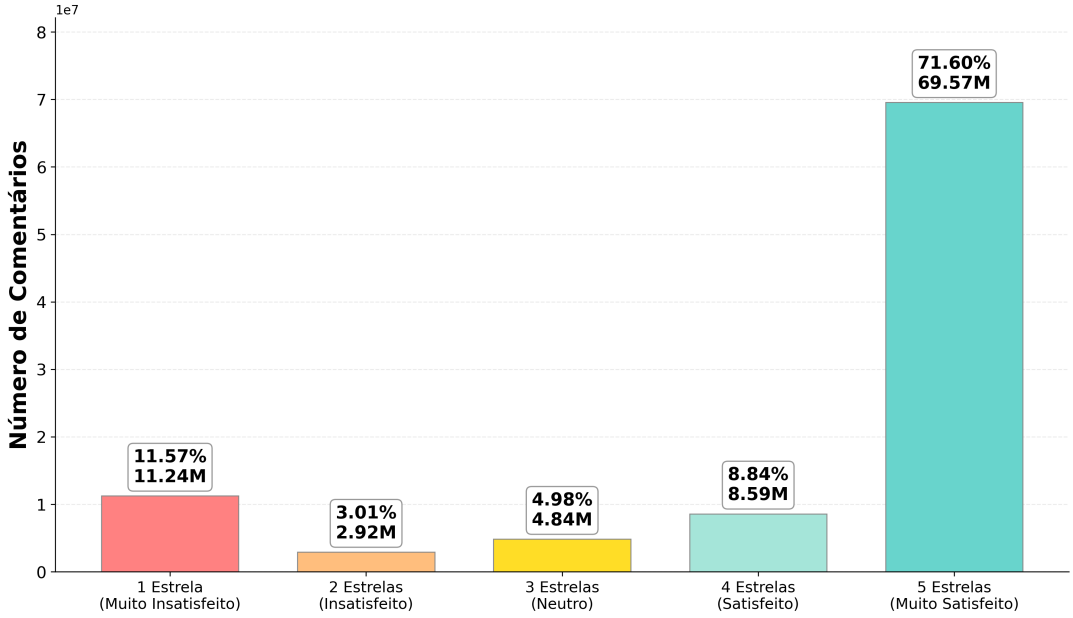


Figura 3. Distribuição de comentários por rating (1-5 estrelas).

A análise da distribuição revela que 80,44% dos comentários recebem ratings positivos (4-5 estrelas), enquanto 14,58% recebem ratings negativos (1-2 estrelas). A média ponderada de 4,26 estrelas mascara a forte polarização subjacente, indicando uma base de usuários com satisfação altamente estratificada.

Esta assimetria é particularmente acentuada em categorias como Redes Sociais e Jogos, onde a satisfação do usuário é mais polarizada, pois os usuários tendem a ser altamente satisfeitos ou migram para alternativas sem deixar avaliações intermediárias. A presença diminuta de avaliações neutras (3 estrelas), representando apenas 4,98% do total, sugere que a escala de 5 pontos funciona primariamente como um mecanismo binário de satisfação/insatisfação para esses usuários.

Consequentemente, análises que dependem exclusivamente de classificações numéricas agregadas podem perder informações relevantes presentes no conteúdo textual dos comentários, especialmente aquelas nuances críticas que não se alinham perfeitamente com a escala discreta de avaliação.

4.5. Análise Textual dos Comentários

A análise textual dos comentários do PlayReviewsPT identifica padrões linguísticos no *feedback* de usuários de aplicativos móveis. Diante da natureza espontânea e não estruturada desses dados, investigam-se o comprimento dos textos e a distribuição lexical. Essa

análise é essencial para compreender limites e oportunidades no uso de técnicas de PLN em textos curtos, com vocabulário simplificado e alta variabilidade expressiva.

4.5.1. Características dos Textos

Os comentários do *dataset* apresentam características linguísticas típicas de avaliações produzidas em plataformas móveis, marcadas por informalidade, brevidade e baixa elaboração textual. A distribuição do comprimento dos comentários, apresentada na Figura 4, revela forte concentração em textos curtos, com mediana de 3 palavras e média de 7,2 palavras. A diferença entre esses dois valores indica uma distribuição assimétrica à direita, na qual a maior parte dos comentários é bastante breve, enquanto um conjunto menor de avaliações mais longas eleva a média.

Esse padrão sugere que muitos usuários recorrem a expressões rápidas para registrar sua percepção sobre os aplicativos, frequentemente utilizando poucas palavras. Essa característica deve ser considerada em tarefas de PLN, pois textos muito curtos podem limitar a disponibilidade de contexto semântico, dificultando análises mais refinadas, como a identificação de aspectos, a detecção de problemas específicos e a classificação de sentimentos baseada exclusivamente no conteúdo textual.

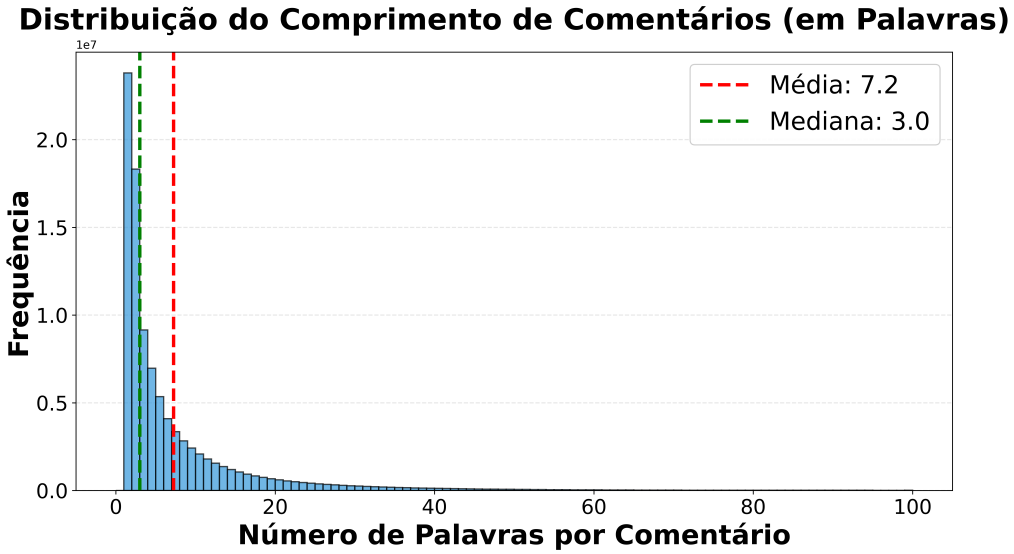


Figura 4. Distribuição do comprimento dos comentários em número de palavras.

4.5.2. Vocabulário e Termos Frequentes

A análise de frequência lexical (Figura 5) identifica os termos mais utilizados nos comentários. As palavras mais frequentes — “bom”, “muito”, “ótimo” e “legal” — indicam que a maioria dos comentários utiliza um vocabulário limitado e baseado em adjetivos de avaliação. Esta simplicidade vocabular é notável e sugere que os usuários confiam primariamente em termos genéricos para expressar satisfação ou insatisfação.

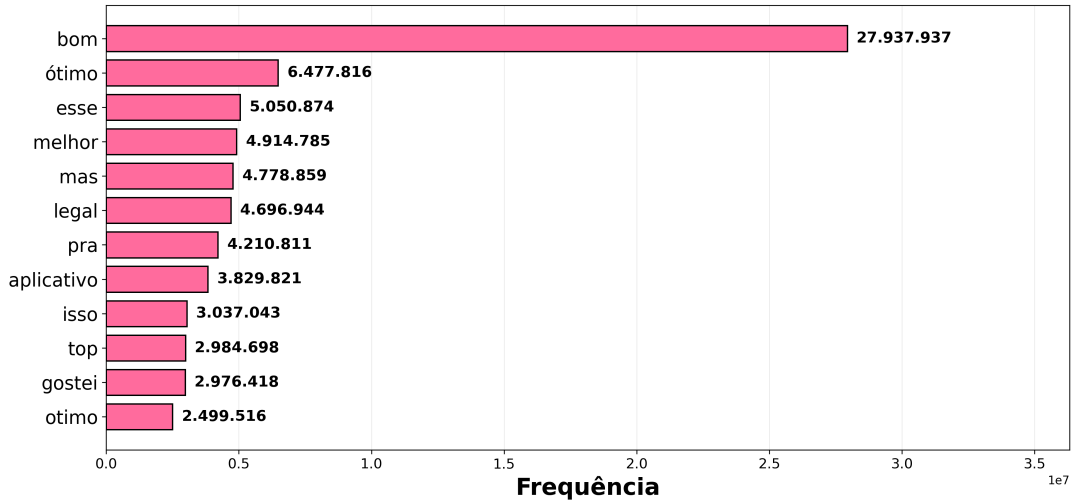
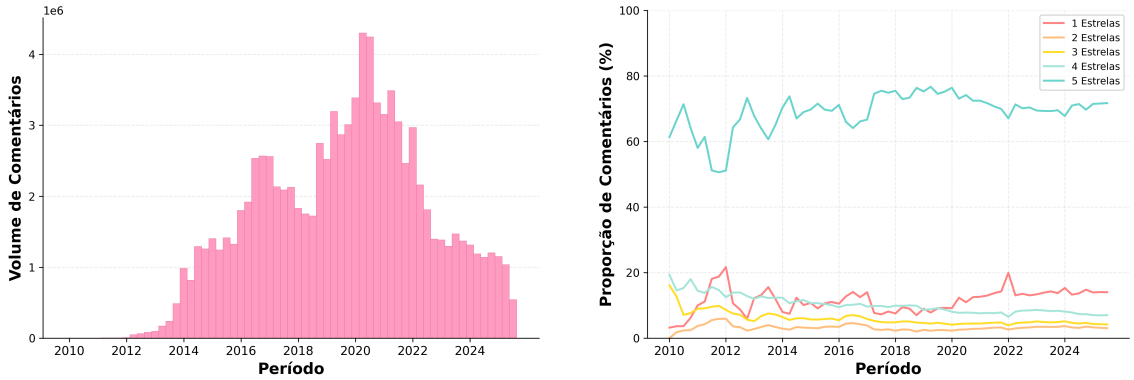


Figura 5. Palavras mais frequentes nos comentários

4.6. Análise Temporal das Avaliações

A evolução temporal do *dataset* revela padrões significativos de crescimento do volume de avaliações, contrastados com uma notável estabilidade na distribuição de ratings. O período analisado abrange 63 trimestres, de 2010 Q1 a 2025 Q3, durante o qual dinâmicas distintas de volume e satisfação emergiram.

O volume de comentários apresentado na Figura 6 (a) indica três fases distintas: (1) *crescimento inicial gradual* (2010-2017), onde o volume aumenta lentamente à medida que mais usuários adotam avaliações móveis; (2) *aceleração exponencial* (2017-2021), coincidindo com a pandemia de COVID-19 e o aumento massivo do uso de aplicativos móveis; (3) *estabilização relativa* (2021-2025), onde o volume continua crescendo, mas em ritmo moderado. O pico absoluto foi registrado em 2020, coincidindo temporalmente com o período de isolamento social global. Essa coincidência sugere uma possível relação entre eventos macrosociais e engajamento em plataformas móveis, embora essa hipótese demande investigação específica.



(a) Volume de comentários por trimestre. **(b)** Proporção dos *ratings*.

Figura 6. Evolução temporal do dataset.

O gráfico de proporções de *ratings* da Figura 6 (b) indica uma predominância

persistente das avaliações de 5 estrelas ao longo do período analisado. Considerando a média trimestral, as avaliações de 5 estrelas representam aproximadamente 69,0% dos comentários, enquanto as avaliações de 1 estrela permanecem em torno de 11,5%. As categorias intermediárias apresentam menor participação média, com 3,2%, 5,9% e 10,3% para 2, 3 e 4 estrelas, respectivamente. Embora sejam observadas oscilações pontuais ao longo da série temporal, especialmente nos anos iniciais, a hierarquia entre as categorias de *rating* permanece relativamente estável, sugerindo que o crescimento do volume de comentários não foi acompanhado por uma mudança proporcional expressiva na distribuição das notas.

A divergência entre volume (crescimento exponencial) e distribuição de *ratings* (estabilidade) indica que o crescimento do conjunto de dados não representa uma melhoria nem deterioração sistemática da qualidade percebida. Mais usuários deixando comentários não implica que seus níveis de satisfação mudaram. Esta observação sugere que a dinâmica de satisfação é determinada primariamente por características inerentes de cada aplicativo, sendo relativamente insensível a flutuações contextuais macroscópicas.

O aumento de volume pode ser explicado por: (1) expansão do mercado de aplicativos móveis; (2) maior penetração de smartphones; (3) maior familiaridade dos usuários com sistemas de avaliação. Simultaneamente, a estabilidade na distribuição de *ratings* indica um patamar estrutural de satisfação que se mantém apesar de mudanças externas, sugerindo que fatores fundamentais de qualidade do software prevalecem sobre variações conjunturais na satisfação do usuário.

5. Disponibilidade e Acesso

O dataset é disponibilizado publicamente via Hugging Face[3]. Os dados estão em formato JSONL, particionados por aplicativo, o que permite o download seletivo de subconjuntos específicos. Além dos arquivos de dados, o repositório inclui o código do coletor utilizado no processo de obtenção das avaliações, disponível no diretório `coletor`. A disponibilização do código permite que terceiros inspecionem a metodologia de coleta, reproduzam parte do processo e adaptem a ferramenta para novas coletas, respeitadas as condições legais e éticas aplicáveis ao uso de conteúdo publicamente disponível.

6. Considerações Finais

Este *dataset* e suas análises são um recurso para pesquisa em ciência de dados aplicada, engenharia de software empírica e mercados digitais. As técnicas de processamento, normalização e análise são generalizáveis a outros contextos de avaliação de usuários, ampliando seu impacto científico. Espera-se que este trabalho estimule novas investigações sobre o comportamento de usuários em ecossistemas digitais e aprofunde a compreensão das dinâmicas de satisfação em contextos reais de grande escala.

Agradecimentos

Este trabalho foi parcialmente financiado pelo Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional, Tratamento e Disseminação de Informação (INCT-TILD-IAR), financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo nº 408490/2024-1.

[3]https://huggingface.co/datasets/UEA-LSI/googleplay-reviews-ptbr

Uso de Ferramentas de Inteligência Artificial

Este trabalho utilizou IA em etapas específicas. Claude (Anthropic) gerou e refinou código Python para processamento de dados JSONL, agregação temporal, análises estatísticas e criação de gráficos em matplotlib, com verificação manual prévia à execução. ChatGPT (OpenAI) revisou o texto (gramática, coerência e terminologia em português), com todas as sugestões criticamente avaliadas pelos autores.

Referências

Di Sorbo, A., Panichella, S., Alexandru, C. V., Shimagaki, J., Visaggio, C. A., Canfora, G., and Gall, H. C. (2016). What would users change in my app? summarizing app reviews for recommending software changes. In *Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering*, pages 499–510.

Genc-Nayebi, N. and Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, 125:207–219.

Harman, M., Jia, Y., and Zhang, Y. (2012). App store mining and analysis: Msr for app stores. In *2012 9th IEEE working conference on mining software repositories (MSR)*, pages 108–111. IEEE.

Iacob, C. and Harrison, R. (2013). Retrieving and analyzing mobile apps feature requests from online reviews. In *2013 10th working conference on mining software repositories (MSR)*, pages 41–44. IEEE.

Khan, N. D., Khan, J. A., Li, J., Ullah, T., and Zhao, Q. (2024). Mining software insights: uncovering the frequently occurring issues in low-rating software applications. *PeerJ Computer Science*, 10:e2115.

Pagano, D. and Maalej, W. (2013). User feedback in the appstore: An empirical study. In *2013 21st IEEE international requirements engineering conference (RE)*, pages 125–134. IEEE.

Panichella, A., Dit, B., Oliveto, R., Di Penta, M., Poshynanyk, D., and De Lucia, A. (2013). How to effectively use topic models for software engineering tasks? an approach based on genetic algorithms. In *2013 35th International conference on software engineering (ICSE)*, pages 522–531. IEEE.

Pereira, D. A. (2021). A survey of sentiment analysis in the portuguese language. *Artificial Intelligence Review*, 54(2):1087–1115.

Piau, M., Lotufo, R., and Nogueira, R. (2024). ptt5-v2: A closer look at continued pretraining of t5 models for the portuguese language. In *Brazilian Conference on Intelligent Systems*, pages 324–338. Springer.

Samanmali, P. and Rupasingha, R. A. (2024). Sentiment analysis on google play store app users' reviews based on deep learning approach. *Multimedia Tools and Applications*, 83(36):84425–84453.

Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: pretrained bert models for brazilian portuguese. In *Brazilian conference on intelligent systems*, pages 403–417. Springer.