How Culture Shapes Customers: A Cross-Continent Analysis of Apps Reviews Using NLP Techniques

Maria Fernanda Azolin Kotsifas[1], **Ricardo Lüders**[1], **Thiago H. Silva**[1,2]

[1]Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, Brazil

[2]University of Toronto, Toronto, Canada

azolin@alunos.utfpr.edu.br, {luders,thiagoh}@utfpr.edu.br

Abstract. *Understanding customers' feedback is essential for businesses to improve products and adapt to different markets. This study analyzes 100,000 app reviews from Uber, Instagram, and WhatsApp across six countries (Brazil, U.S., Australia, India, U.K., and South Africa) to assess whether user concerns are global or culturally influenced. Using BERTopic for topic modeling, zero-shot classification for topic assignment, and Aspect-Based Sentiment Analysis, we identify twelve key topics in reviews. While some topics are shared globally, others are country-specific. For example, Uber's reliability was a major concern in South Africa and Australia, while Brazilian users discussed WhatsApp voice messages more frequently. These findings help businesses detect market-specific trends, benchmark competitors, and address regional needs strategically.*

1. Introduction

In today's digital age, mobile applications are central to daily life, with millions relying on them for communication, productivity, and entertainment. The number of global app downloads reached 257 billion in 2023, reflecting their widespread adoption [Statista 2025]. App marketplaces like the Apple App Store and Google Play Store allow users to leave public reviews, generating vast amounts of unstructured feedback. These reviews offer valuable insights into user experiences, making them a critical resource for businesses aiming to improve their products and understand diverse audiences.

However, manually analyzing such large volumes of reviews is challenging due to information overload [Aslam et al. 2020]. Machine learning (ML) and natural language processing (NLP) techniques, particularly topic modeling, have emerged as powerful tools to automate this process, uncovering patterns and themes in user feedback [Krishnan 2023]. These methods enable businesses to systematically identify trends, prioritize improvements, and tailor apps to global markets efficiently.

This study explores how user expectations and feedback vary across countries, investigating whether cultural differences influence app evaluations. We assume that feedback is any comment registered by a user as a review in an application, whereas expectation is any sentiment a user expresses as positive or negative regarding their comment. To extract these insights, we used BERTopic for topic modeling, zero-shot classification, and Aspect-Based Sentiment Analysis (ABSA). So, we address the research question: *Are the most common complaints and praised features of the same application consistent across countries, or are they culturally specific?*

2. Related Work

Prior research has employed NLP techniques to demonstrate that user reviews on products carry important messages, and the platform could influence them [Santos et al. 2020]. Specifically regarding mobile app reviews, there are several initiatives [Fatima et al. 2024, Ahn and Park 2023, Pranatawijaya et al. 2024], though often with limited scope. For instance, Ahn and Park (2023) [Ahn and Park 2023] used bag-of-words and sentiment analysis on fitness app reviews, finding emotional factors (e.g., affection) stronger predictors of satisfaction than functional aspects. Pranatawijaya et al. (2024) [Pranatawijaya et al. 2024] highlighted cultural influences on UX but focused on single-country datasets. Fischer et al. (2021) [Fischer et al. 2021] investigated how cultural dimensions influence the structure of reviews—such as review length, star rating distributions, and emoji usage—without exploring the semantic content of the reviews themselves. In contrast, Amirkhalili and Wong (2025) [Amirkhalili and Wong 2025] combined topic modeling and sentiment analysis to uncover functional and usability issues in banking app reviews, but their study was limited to a single app category and fewer countries.

While these works provide valuable insights, most are constrained by mono-domain, mono-platform, or mono-cultural settings. Our work advances this field by combining BERTopic, zero-shot classification, and ABSA to analyze 100K reviews from three distinct apps across six culturally diverse countries, uncovering both universal concerns and culturally specific feedback patterns in user experience.

3. Methodology

This section describes the methodology used for data collection and preprocessing, followed by three key data analyses: i) topic modeling using BERTopic; ii) zero-shot classification; iii) aspect-based sentiment analysis, according to Figure 1.

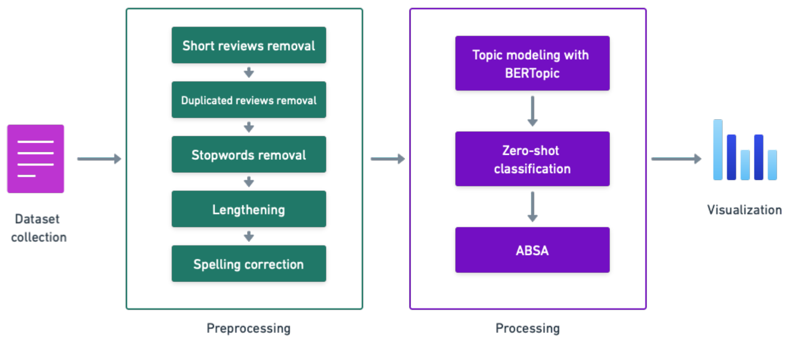


Figure 1. Methodological procedures.

3.1. Data Collection and Preprocessing

The datasets used in this study were collected using two Python libraries for web scraping: *app-store-scraper*[1] for the Apple App Store and *google-play-scraper*[2] for the Google Play Store. They contain reviews made between 2024 and 2025 and will be made available in a GitHub repository after publishing this article. We extracted more than 5,000

[1]https://github.com/facundoolano/app-store-scraper
[2]https://github.com/facundoolano/google-play-scraper

reviews from each of the three selected applications — Uber, Instagram, and WhatsApp — across six countries from different continents: Brazil, the United States, Australia, India, the United Kingdom, and South Africa, which resulted in a total of approximately 100,000 reviews. The data are compiled into a CSV file containing two key attributes: the numerical rating (1-5 stars) and the textual content of the review. To ensure a more comprehensive analysis, datasets from both app stores are merged, allowing for a broader representation of user feedback.

The collected data undergo a preprocessing pipeline. Language-specific preprocessing techniques were applied because the original reviews were written in English and Portuguese, according to the following steps: **Removal of very short reviews:** Reviews containing two words or fewer were discarded, as they provided little meaningful information for topic modeling and sentiment analysis. Short reviews often consisted of generic expressions that lacked context or depth regarding user experience. **Removal of duplicate reviews:** Identical reviews appearing multiple times in the dataset were removed. This step helped reduce redundancy and prevent over-representation of specific opinions, optimizing computational efficiency and ensuring that repeated comments did not bias the results. **Stopword removal:** Since the dataset contained reviews in both English and Portuguese, stopword lists for each language were applied selectively. **Lengthening:** Informal writing styles often include excessive letter repetition (e.g., ”loooove this app!!!”). To normalize text structure, repeated characters were reduced to their standard form while preserving meaning (”love this app!”). **Spell and Slang Correction:** Language-specific spelling and slang correction were applied using the *Enelvo*[3] Python library for Portuguese and the *SymSpell*[4] Python library for English.

3.2. Topic modeling

To identify meaningful themes in user reviews, we evaluate (not shown here due to space constraints) both traditional (LDA - Latent Dirichlet Allocation) and modern (BERTopic) topic modeling approaches. Then, we chose BERTopic because it demonstrated superior performance in topic coherence and interoperability, consistent with prior studies [Krishnan 2023]. This advantage stems from BERTopic's use of transformer-based embeddings and a class-based TF-IDF procedure, which better captures semantic relationships in multilingual text. We employed a pretrained multilingual BERTopic model [Grootendorst 2022] to handle our dataset's mix of English and Portuguese reviews.

3.3. Zero-shot classification

To categorize reviews into the topics identified by BERTopic, we employed the *facebook/bart-large-mnli*[5] zero-shot classifier. This model was selected after evaluating four classifiers (not shown here due to space constraints) against human-annotated samples, where it achieved the highest agreement rate. Its multilingual capability allowed consistent classification of both English and Portuguese reviews without language-specific fine-tuning. The zero-shot approach enabled efficient topic assignment at scale while avoiding the need for manually labeled training data. This proved critical for handling our large, multilingual dataset and ensured reproducible topic distributions across countries for comparative analysis.

[3]https://github.com/thalesbertaglia/enelvo
[4]https://github.com/wolfgarbe/SymSpell
[5]https://huggingface.co/facebook/bart-large-mnli

3.4. Aspect Based Sentiment Analysis (ABSA)

Aspect-Based Sentiment Analysis (ABSA) was applied to the dataset to gain deeper insights into user sentiment for each identified topic. For each topic detected by BERTopic, the ABSA model analyzed the sentiment distribution of the reviews, categorizing them as positive, negative or neutral. This enabled a structured understanding of user perception across different app features. E.g., if a topic related to ”Account Registration” had 70% negative reviews and 30% positive reviews, this would indicate that users frequently face issues in the registration process. ABSA was performed using *bert-base-multilingual-cased* [Devlin et al. 2018] model from the Transformers library, which supports multiple languages.

4. Results

Table 1 presents a sample of reviews from the Instagram dataset, showcasing the structure of the data before preprocessing and analysis.

Table 1. Sample reviews from Instagram dataset

Review	Rating	Topic
I can’t download this app	1	download
They really need to add the ”jump to message” thing when a text is searched (the one discord has) cuz it’s very frustrating to scroll and scroll.	3	search
It help full because when you have Instagram you don’t need WhatsApp or Facebook and also TikTok because you can watch reels or post’s and also text people and also you can post	5	tiktok

After preprocessing, topic modeling was performed using BERTopic. It generates clusters of thematically related reviews. The third column from Table 1 presents the topic associated with the review. Table 2 presents the main groups of topics for Instagram, Uber, and WhatsApp in six countries (Brazil, United States, Australia, India, United Kingdom, and South Africa).

Table 2. BERTopic generated groups of topics for Instagram, Uber, and WhatsApp

App	Main topics
Instagram	’update’, ’account’, ’filters’, ’like’, ’nazism’, ’palestin’, ’nickname’, ’tiktok’, ’download’, ’search’, ’battery’, ’ban’
Uber	’driver’, ’update’, ’polite’, ’friendly’, ’fast’, ’price’, ’features’, ’clean’, ’support’, ’reliable’, ’address’, ’waiting’
WhatsApp	’message’, ’color’, ’update’, ’status’, ’audio’, ’photo’, ’battery’, ’slow’, ’internet’

Figures 2, 3, and 4 present the distribution of classified topics in each country for Instagram, Uber, and WhatsApp, respectively. For Instagram, we see three main topics - *Like, Update* and *Account* - being discussed more often. However, for Brazil, the *Account* topic is not so relevant, whereas the *Nickname* one is. For Uber, we see that people often comment about the *Driver* and *Address* for most countries, and the *Reliable* topic in Brazil and the USA is not as common as in other countries. For WhatsApp, we see a clear distinction between the three main topics - *Message, Status*, and *Update* - and the others. Less predominant topics like *Audio* and *Slow* appear to have more meaning in Brazil.

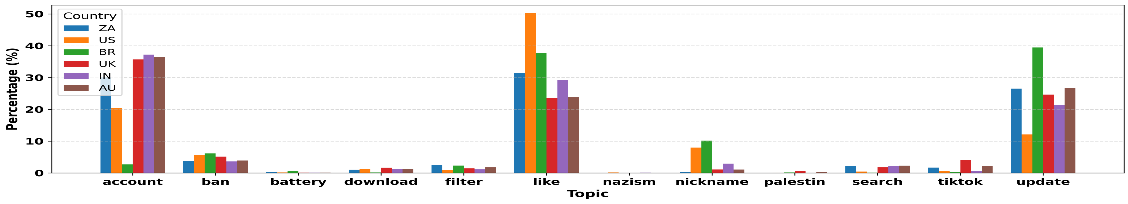


Figure 2. Distribution of topics in Brazil (BR), United States (US), Australia (AU), India (IN), United Kingdom (UK), and South Africa (ZA) for Instagram

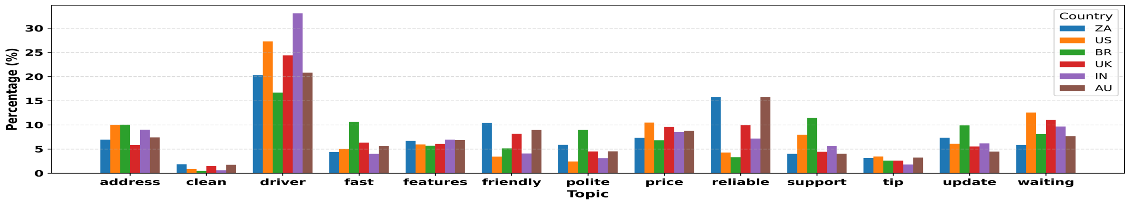


Figure 3. Distribution of topics in Brazil (BR), United States (US), Australia (AU), India (IN), United Kingdom (UK), and South Africa (ZA) for Uber

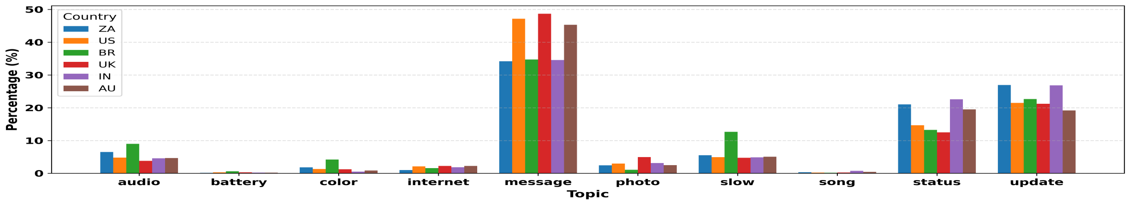
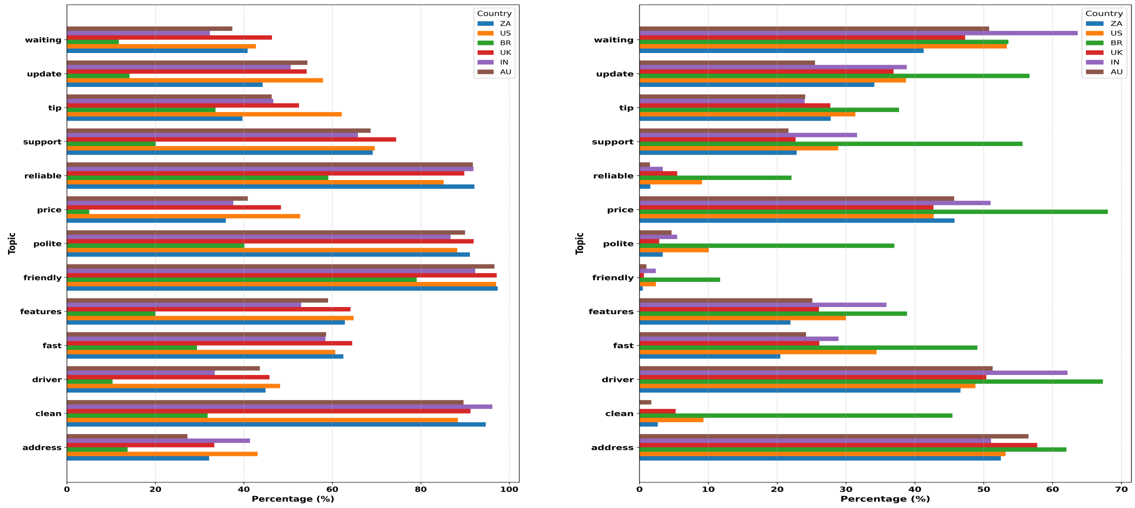


Figure 4. Distribution of topics in Brazil (BR), United States (US), Australia (AU), India (IN), United Kingdom (UK), and South Africa (ZA) for WhatsApp

Finally, Aspect-Based Sentiment Analysis (ABSA) was applied to evaluate the sentiment polarity for each identified topic. Figure 5(a) illustrates how much the reviews that are classified in each topic are positive, in each country for the Uber app. Figure 5(b) illustrates the negative percentages.

As observed in the visualizations, the most common topics tend to be consistent across different countries, reflecting universal user concerns (in the Instagram examples, topics related to likes and updates, or the driver topic for the Uber app). However, there are topics that appear moderately in some countries but are absent in others, indicating region-specific concerns. For instance, issues related to slowness and voice messages may be more prominent in some markets (Brazil) for WhatsApp than in others. Another notable finding is that the sentiment associated with certain topics can vary significantly across countries. For example, as seen in Figure 5, in the case of the Uber app, the *Clean* topic is predominantly negative in Brazil, often linked to complaints about vehicle hygiene or driver behavior. In contrast, the same topic appears with more positive sentiment in other countries, where users are more likely to praise cleanliness standards. The same occurs for Whatsapp, where the topic *Battery* is extremely negative for South Africa and more equilibrated for other countries. This suggests that while the topic itself may be globally relevant, user expectations around it are shaped by regional context.



(a) Positive (b) Negative

Figure 5. Sentiment per topic for Uber app.

5. Conclusion

This work presented results of collecting and analyzing a large volume of customers' feedback of Instagram, Uber, and WhatsApp in six countries (Brazil, the United States, Australia, India, the United Kingdom, and South Africa). It investigated whether the most common complaints and praised features are consistent across countries or if they are culturally specific. Using BERTopic for the identification of most frequently mentioned topics in users' reviews and ABSA for sentiment analysis, the results showed that while some aspects of the user experience are universally important, others may be influenced by local factors such as economic conditions, politics, and cultural expectations regarding service quality. However, even though country-specific concerns emerge in the analysis, the most frequently mentioned topics consistently dominate across different regions, indicating that certain aspects of an app's performance and usability tend to be global priorities for users. Some limitations of this work should be mentioned as well. It is restricted to US-based apps, which may not fully capture global diversity. The NLP models might miss cultural nuances, sarcasm, or evolving slang not present in their training data. Additionally, the lengthening step may have softened sentiment expressions. This research also provides valuable insights for both new and existing products, serving as a benchmarking tool to identify global patterns and localized pain points. It helps companies detect outliers in user behavior and adjust strategies accordingly. For instance, if a topic appears disproportionately in one country, it highlights a local issue or opportunity requiring attention. These findings support data-driven decisions in specific markets. Cross-cultural review analysis enables companies to align products with both global expectations and regional needs, supporting more effective and targeted product improvements. Future work could include more regions, apps, and culturally fine-tuned models.

Acknowledgments

This research was partially supported by the Fapesp SocialNet project (process 2023/00148-0), by CNPq (processes 314603/2023-9, 441444/2023-7, 409669/2024-5, and 444724/2024-9), and the INCT TILDIAR (process 408490/2024-1).

References

Ahn, H. and Park, E. (2023). Motivations for user satisfaction of mobile fitness applications: An analysis of user experience based on online review comments. *Humanities and Social Sciences Communications*, 10(1):1–7.

Amirkhalili, Y. and Wong, H. Y. (2025). Banking on feedback: Text analysis of mobile banking ios and google app reviews. *arXiv preprint arXiv:2503.11861*.

Aslam, N., Ramay, W. Y., Xia, K., and Sarwar, N. (2020). Convolutional neural network based classification of app reviews. *IEEE Access*, 8:185619–185628.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Fatima, E., Kanwal, H., Khan, J. A., and Khan, N. D. (2024). An exploratory and automated study of sarcasm detection and classification in app stores using fine-tuned deep learning classifiers. *Automated Software Engineering*, 31(2):69.

Fischer, R. A.-L., Walczuch, R., and Guzman, E. (2021). Does culture matter? impact of individualism and uncertainty avoidance on app reviews. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, pages 67–76. IEEE.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Krishnan, A. (2023). Exploring the power of topic modeling techniques in analyzing customer reviews: a comparative analysis. *arXiv preprint arXiv:2308.11520*.

Pranatawijaya, V. H., Sari, N. N. K., Rahman, R. A., Christian, E., and Geges, S. (2024). Unveiling user sentiment: Aspect-based analysis and topic modeling of ride-hailing and google play app reviews. *Journal of Information Systems Engineering and Business Intelligence*, 10(3):328–339.

Santos, G., Mota, V. F. S., Benevenuto, F., and Silva, T. H. (2020). Neutrality may matter: sentiment analysis in reviews of Airbnb, Booking, and Couchsurfing in Brazil and USA. *Social Network Analysis and Mining*, 10(1):45.

Statista (2025). Annual number of global mobile app downloads 2016-2023. https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/.