Listening to Users at Scale: A Topic Modeling Study of Google Play Store Reviews

Gaurisha Pandey

Digital Humanities / NLP Case Study

Abstract. User reviews on app stores contain rich information about user satisfaction, frustration, and unmet needs. However, the scale of such data makes manual analysis impractical. This study applies Natural Language Processing (NLP) and Latent Dirichlet Allocation (LDA) topic modeling to thousands of Google Play Store reviews of a productivity application to automatically discover dominant user concerns and positive feedback themes. The analysis reveals seven coherent topics, including subscription and pricing complaints, reminder reliability issues, synchronization problems, bug reports, feature requests, general frustration, and positive habit-building feedback. Topic frequency and rating analysis show that while positive feedback is the most common theme, the most emotionally negative and impactful issues are related to monetization and the reliability of core features. The study demonstrates how large-scale qualitative user feedback can be converted into actionable product insights for data-driven product and UX decision-making.

1. Introduction

Modern mobile applications receive thousands of user reviews every day. These reviews are not merely ratings; they are narratives describing real user experiences: frustration when an app crashes, anger when essential features are locked behind paywalls, or satisfaction when an application genuinely improves daily productivity. For product teams, such feedback is invaluable. However, the sheer volume of reviews makes manual reading and qualitative coding infeasible. This project addresses a practical and important question: can we teach a computer to read user feedback at scale and summarize what users truly care about? Using Natural Language Processing (NLP) and topic modeling, this study analyzes Google Play Store reviews of a productivity application to automatically discover major themes in user feedback and to translate unstructured text into structured, actionable insights.

2. Objectives

The objectives of this project are: To preprocess and clean large-scale user review text data. To apply topic modeling to discover latent themes in user feedback. To quantify the importance of each theme using topic distribution. To connect discovered topics with user ratings to assess dissatisfaction and satisfaction. To derive product and UX recommendations grounded in empirical evidence.

3. Dataset Description

The dataset consists of real Google Play Store user reviews collected using a Google Play scraping tool. Each row represents a single user review along with metadata such as rating, timestamp, and application identifier. The primary field used for analysis is the *content* column, which contains the natural language text written by users. After removing empty entries and invalid rows, the dataset contains several thousand reviews, providing a sufficiently large corpus for unsupervised text mining and topic modeling. Because each entry corresponds to a real user, the dataset offers a direct window into authentic user experience and sentiment.

4. Data Cleaning and Preprocessing

Raw user-generated text is noisy and inconsistent. To make the data suitable for modeling, the following preprocessing steps were applied: lowercasing all text, removal of URLs, numbers, emojis, and special characters, tokenization of text into words, removal of stopwords while explicitly retaining important negation words such as “not”, and lemmatization to reduce words to their base forms (e.g., “working” to “work”). Finally, empty or near-empty reviews were removed. This step is crucial because topic models are highly sensitive to noise; poor preprocessing would lead to incoherent and uninterpretable topics.

5. Exploratory Analysis

Before applying topic modeling, basic exploratory analysis was performed. This included analyzing review length distributions, star rating distributions, and the most frequent words in the corpus. The exploratory phase indicated that users frequently discuss reminders, pricing, synchronization, updates, and features, suggesting that the dataset is rich in both functional feedback and emotional responses.

6. Methodology: Topic Modeling with LDA

Latent Dirichlet Allocation (LDA) was used for topic modeling. The cleaned text corpus was converted into a document-term matrix using CountVectorizer. Several values for the number of topics were experimented with, and seven topics were chosen as the final model because they provided the most interpretable and semantically coherent themes. Each review was then assigned a dominant topic based on the highest topic probability.

7. Results

The model discovered seven major themes in user feedback: Positive Feedback and Habit Building Subscription and Pricing Complaints Reminder and Notification Problems App Bugs and Update Issues Feature Requests and UI Improvements Synchronization and Account Issues General User Frustration Representative review samples from each topic confirmed that these themes are semantically coherent and grounded in real user concerns.

8. Critical Analysis of Topic Distribution and Ratings

Topic distribution analysis shows that the most frequent topic is positive feedback, indicating that many users genuinely value the application. However, the most dominant negative topic is subscription and pricing complaints, followed closely by reminder and notification failures. Importantly, when topic membership is combined with star ratings, these same topics also exhibit the lowest average ratings. This reveals a critical insight: the issues that appear most often are also the ones that generate the strongest negative emotional responses. In other words, monetization strategy and reliability of core features are not minor annoyances; they are central to user dissatisfaction and directly threaten long-term user trust and retention.

9. Discussion

The findings suggest a classic product tension. On one hand, the application has strong product-market fit, as evidenced by the large volume of positive habit-building feedback. On the other hand, reliability and monetization issues undermine this goodwill. A productivity app whose reminders fail or whose synchronization breaks violates its core value proposition. Similarly, an overly aggressive monetization strategy can convert satisfied users into frustrated critics. From a product strategy perspective, this implies that engineering reliability and fair value communication should be prioritized over purely adding new features.

10. Product Recommendations

Based on the empirical evidence from the analysis, the following recommendations are proposed: Prioritize fixing reminder and notification reliability, as this is a core feature. Revisit monetization strategy to improve the free-tier experience and clarify premium value. Stabilize synchronization and account systems to prevent trust erosion. Strengthen testing pipelines before updates to reduce regression bugs. Leverage positive feedback in marketing and onboarding to reinforce perceived value.

11. Limitations

This study is limited to reviews from a specific application and time period. Topic modeling is also an unsupervised method, and while qualitative validation was performed, topic boundaries can still overlap. Future work could compare multiple applications, incorporate temporal analysis, or integrate sentiment models more deeply.

12. Conclusion

This project demonstrates how NLP and topic modeling can transform large-scale, unstructured user feedback into structured, actionable insights. By teaching machines to listen to users at scale, product teams can base decisions on real evidence rather than intuition. The approach shown here is directly applicable to modern, data-driven product and UX research.