Scotia Data Science: From Reviews to Revolution

Team Name: Triangle of Statness:
Team Members: Tianyi Wu, Daisy Li, Jingyi Fang, Stone Luo

This brief write-up outlines the approach we used to generate valuable insights from user reviews leveraging LLM and NLP. By systematically addressing irrelevant input, cleaning the data, and leveraging advanced models, we prioritized topics and identified pain points, desired features, and recommendations.

Remove Easter Eggs

The first step was identifying Easter eggs since irrelevant input tends to produce inaccurate output. Any review with an obvious presence of banking-related terms was removed. Then LLM, namely ChatGPT, was used to label the completely irrelevant reviews as Easter eggs. By iteratively applying this divide and conquer strategy, the problem of finding Easter eggs was simplified into a series of binary decisions, significantly reducing the initial complexity.

Cleaning

The next step was cleaning the reviews. Cleaning the reviews was essentially removing words that do not contribute to future analysis, which were decided based on the length, syntax, and relevance of each word. Words that meet any of the following criteria were removed:

1. Length: contains two or fewer characters
2. Syntax: is emoji, punctuation, or special characters
3. Relevance: does not relate to any specific topic but frequently appears in the reviews, such as “best” and “worst”

Seed list

In addition to the cleaned reviews, the other critical input for the model is a list of seed words - keywords that capture the semantic meaning of the topics. The seed words are weighted higher in the model and used more frequently in the output representative words for each topic, thus increasing the accuracy of the model. The list of seed words was initially generated by tokenizing the given topic description, then modified manually to better capture the essence of each topic.

Modelling for Topic Ranking

A guided BERTopic model was trained with the cleaned reviews and the list of seed words. The model generated around 100 topics with counts and words that represent that topic. Based on the representative words, the model-generated topics were manually classified into the 20 given topics. Before the manual classification, we attempted using keyword matching for further classification but received zero counts for some topics, so was aborted.

Category	Sum of Count
2SV	2210
NC	1463
Cheque_Deposit	678
Errors	416
Login_and_Logout_Issues	374
Email_Money_Transfer	193
PP	120
Chat	117
Biometric_Login	107
Quick_Balance	83
Application Performance	78
Info_Alerts	75
Appointment_Booking	70
Fee	58
Accessibility	43
Save_and_Share_Statements	42
Investments	38
Request_New_Card	34
Credit_Score	32
International_Money_Movement	31
Budgeting	24
Rewards	24
Grand Total	**6310**

Filter Representative Reviews

Pain points and desired features fall between reviews that accurately represent users, which can be determined by the number of upvotes. By filtering reviews with many likes, we identified representative reviews for future analysis that further identified pain points, desired features, and recommendations.

Sentiment Analysis for Pain Points/Desired Features/Recommendations:

A sentiment analysis was performed using the RoBerta model. The model evaluated the sentiment of each review by generating scores for positive, negative, and neutral sentiments. Next, the extremely negative reviews – those with a significant negative score – were identified and fed into the Berttopic model to identify the pain points. A similar process was conducted to identify desired features and recommendations.

In summary, this comprehensive process not only allowed us to filter out noise and irrelevant information but also discovered valuable insights from user reviews. Moving forward, these insights can inform strategic decision-making and product improvements.