Scotiabank: Data Science Discovery Days

Ambar Chakraborty[1], Gurnika Kaur[1], Kshama Patel[1], and Sev Ladtchenko[1]

[1]Waterloo University, Waterloo, Canada

February 2, 2024

Exploratory Data Analysis

We started with a word cloud to find commonly occurring words. We discovered many reviews of hotels. These are Easter Eggs that have been injected into the dataset.

Easter Eggs

We went with a Mixture-Of-Experts approach by using several models: a top-performing open-source text embedding model, Instructor-xl, and a framework for few-shot fine-tuning of Sentence Transformers, SetFit.

First, we generated a small labeled dataset of hotel reviews using ChatGPT. Next, we fine-tuned the Sentence Transformer model on labeled examples for the Easter Egg detection task. Finally, we trained a classifier head on the generated embeddings and applied it to the real dataset to predict class labels.

Using t-SNE for visualization of the text embeddings (fig 1), we found most irrelevant reviews in a cluster. SetFit found additional Easter Eggs outside this cluster.

Popularity of 20 Topics

Next, we rank the 20 topics by occurrence. To train a classifier, we need labeled data. We supplied the topic descriptions from the data dictionary to ChatGPT to generate sample reviews for each topic.

As not all reviews fall under the 20 given topics, we introduced two additional topics for "short" reviews under 3 words, and "emotional" reviews that carry no information about the user experience.
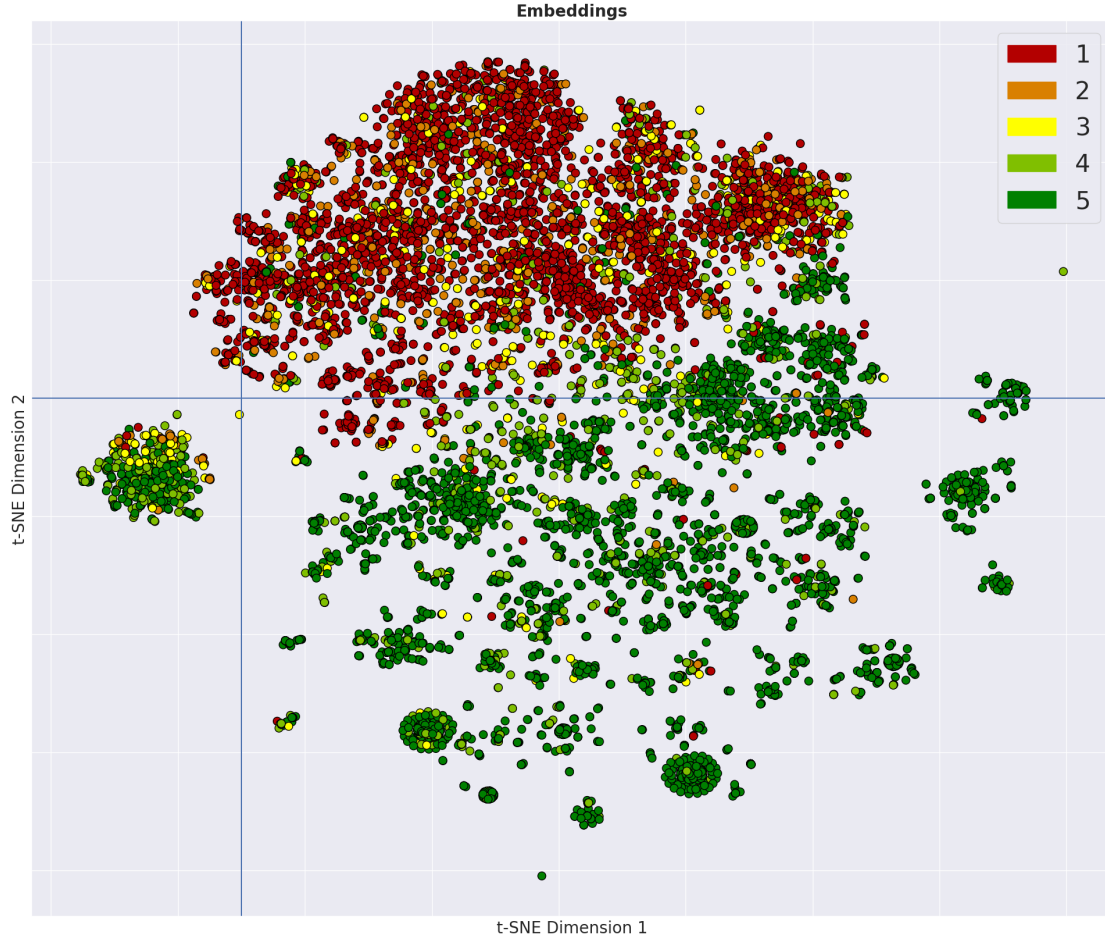
Figure 1: t-SNE projection of the review embeddings. Easter Eggs are in a cluster on the bottom left.

For each topic, we used Instructor-xl to embed the sample reviews and averaged them to find the centroid for this topic. We compute the cosine similarity between Scotiabank mobile app reviews and the topic centroids. Each review will then have 22 similarity scores, the highest of which suggests the most relevant topic (fig 2).

Sentiment Analysis

Now that we generated categories for each review, we partition the reviews by category. For each category, we sort reviews by likes to see the most pressing issues and customer sentiment.

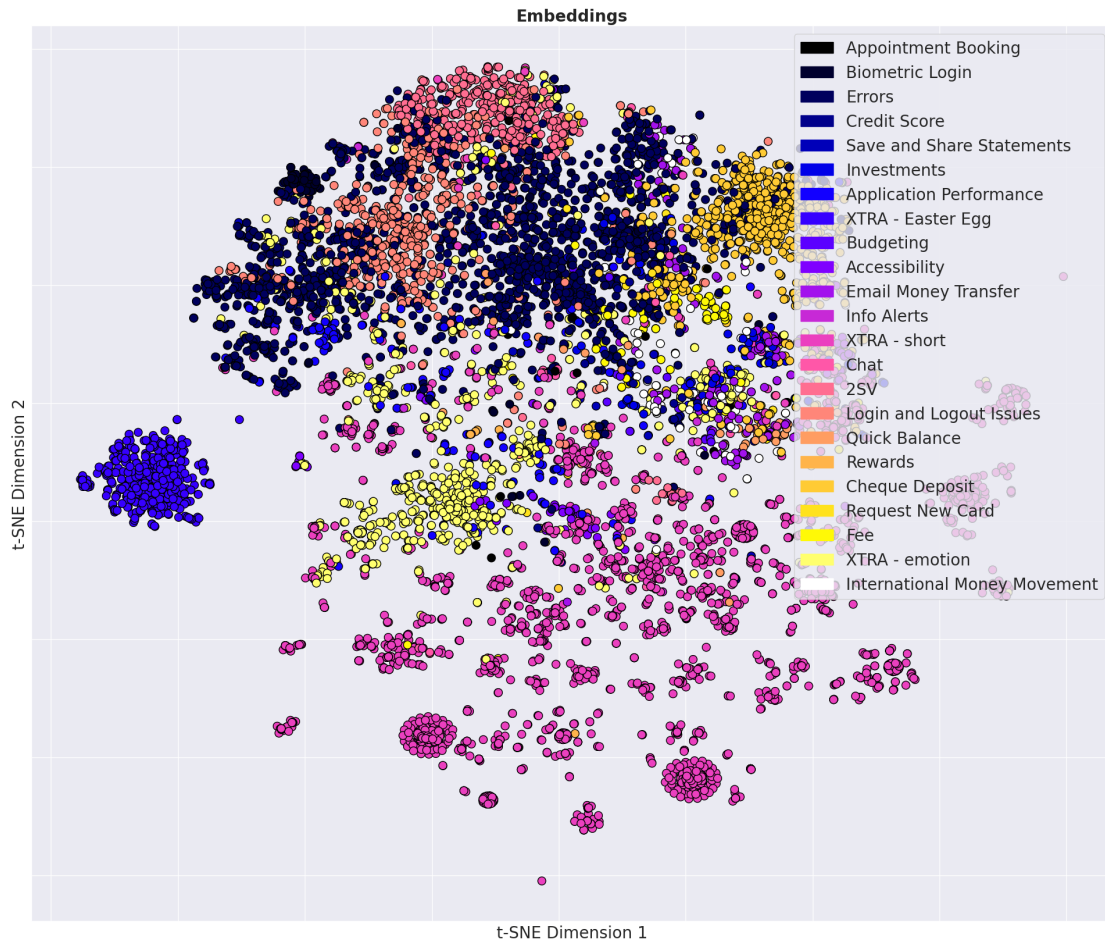We fed the reviews through Mixtral-8x7B. It is a top-performing open-source LLM

Figure 2: Embeddings colored by their classified topic.

that has been quantized to run on the free tier of Google Colab with a 16 GB RAM GPU.

We used the LLM to extract and identify customer needs, desired features, and pain points, which only required modifying the prompt. We combined the results together and again fed them to Mixtral-8x7B to summarize its own output recursively until we had one final summary. Processing through the LLM corrects spelling and incoherent comments, while summarizing a batch of comments.

```
User: In one sentence, what is the customer's main pain point in this
review? here is the review: THIS IS TERABLE

Mixtral: The primary pain point in this review is the customer's clear
```

dissatisfaction with the product or service, as indicated by their use
of the word "terrible."

Customer Needs, Desired Features, Pain Points

We used SetFit to find reviews that mention customers' needs and desires, sorted by likes. We combined this result with the LLM output. To identify pain points we used the sentiment analysis output to rank the topics by negative sentiment.

Our findings suggest that the customers want improvements in the check deposit feature, reduced bank account fees, and faster performance without lag (fig 3).
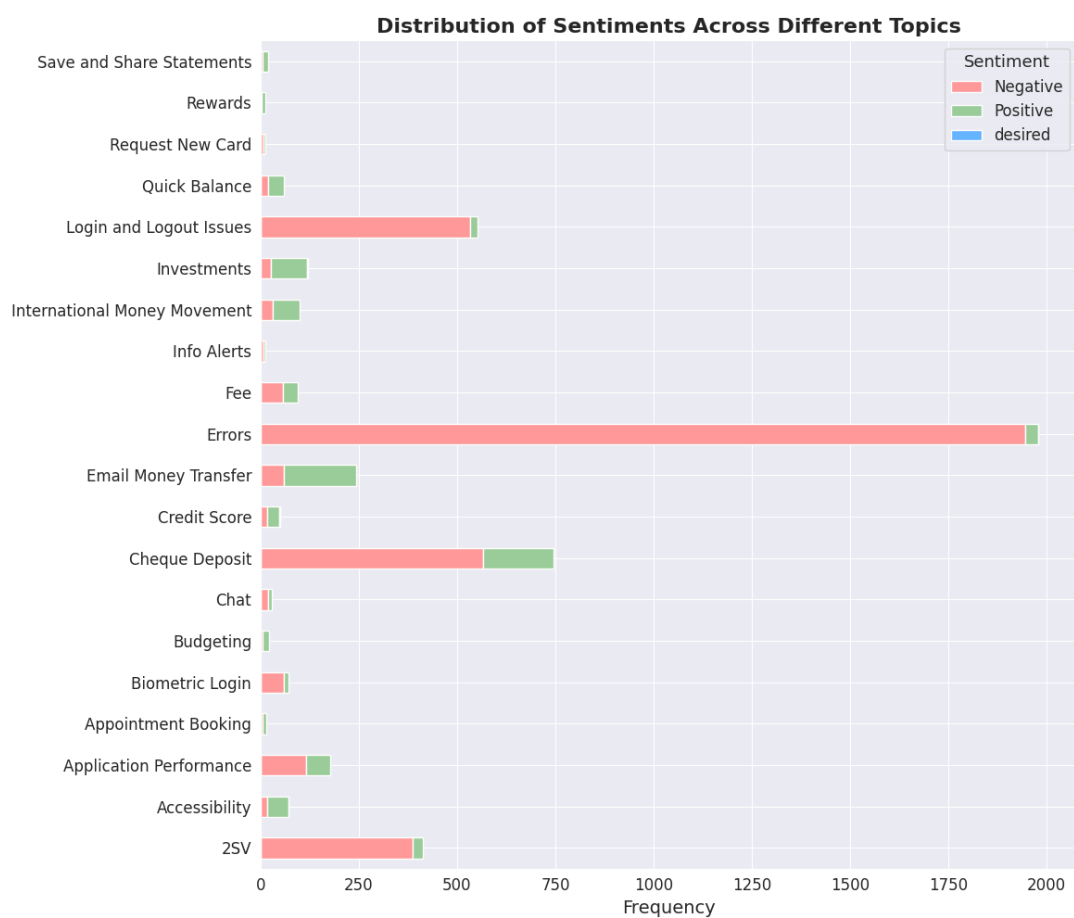

Figure 3: Sentiment distribution by topic.

Build/Fix One Feature of the App

We suggest fixing the check deposit and 2 Factor Authentication features, as they have received the largest number of negative comments.

Long Term Suggestions to Improve Customer Experience

We suggest Voice Assistant Integration to allow customers to perform banking tasks using voice commands, as many reviews have expressed difficulty in navigating the app. The voice assistant should be able to check account balances, transfer funds, and provide answers to frequently asked questions.