Bradley Mensah
Springboard Data Science, Nov 2nd Cohort

Capstone 3 Project Report

Applying Sentiment Analysis to Improve Customer Satisfaction

1. Context

Zendesk's 2020 Customer Experience Trends Report explains that 80% of customers who have had two or more bad experiences with a brand will consider switching to a different brand. The report also details that customer service was the second most important factor in brand loyalty for surveyed customers (the first being price). Therefore, it is imperative to determine how satisfied customers are with customer service interactions in order to increase retention.

As air travel is looking to rebound post-covid19, it is imperative now more than ever that airlines perform review monitoring to address and ease the concerns of returning customers.

From this context, I developed the following problem statement:

What opportunities exist for airlines to increase customer retention by 5% within one quarter through developing loyalty programs, offering more destinations, or improving brand reputation?

In order to answer this question, I sourced labelled tweets in order to train a model to distinguish between a positive and negative tweet. I also sourced airline customers' tweets and the airlines' responses. Both datasets were found on Kaggle. The labelled data was a 1.6 million tweet dataset (800,000 positive tweets, 800,000 negative tweets). The airline tweets were a part of a larger dataset of customer service tweets for over 100 companies.

Data source citations:

Labelled Sentiment Tweets:

Citation: Go, A., Bhayani, R. and Huang, L., 2009. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(2009), p.12.

Source: https://www.kaggle.com/kazanova/sentiment140

Customer Service Tweets:

Citation: Thought Vector and Axelbrooke, Stuart (December 2017). Customer Service on Twitter, Version 10. Retrieved 4 May 2021 from: https://www.kaggle.com/thoughtvector/customer-support-on-twitter.

License: CC BY-NC-SA 4.0

Source: https://www.kaggle.com/thoughtvector/customer-support-on-twitter

2. Data Wrangling and Exploratory Data Analysis

Among the over 100 companies in the customer service tweets dataset, 8 were airlines: Delta, American Air, British Airways, Southwest Air, Air Asia, Jet Blue, Alaska Air, and Virgin Atlantic.

As my goal was to compare tweets before and after a customer service interaction, I created two dataframes for each company: one with the customers' tweets that prompted a customer service Twitter account to respond, and another with the customers' tweets after the interaction.

I defined a series of functions to clean and transform each tweet into a series of tokens. I removed twitter handles, punctuation, non-alphabetic characters, stop words, and set all text to lowercase.

Next, I defined a vocabulary by using the Python Counter class to create a dictionary with each unique word as keys and their frequency in the entire corpus of tweets as the values. With this dictionary, I was able to identify and remove the least common words, as they would be unlikely to help the model predict sentiment.

Once the data was cleaned and tokenized and the vocabulary was trimmed, I performed exploratory data analysis. I found that Delta has the most tweets, with 42,253, and VirginAtlantic the least with 4,318 tweets.

AirAsia took the longest to respond to customer tweets, taking an average of 12 hours, while all of the other airlines were able to respond in less than 4 hours. AlaskaAir had the fastest response time, addressing customer tweets in only 9 minutes on average.

The average number of words per tweet was pretty consistent across airlines, with AmericanAir being the most curt with an average of 15 words per tweet, and Southwest being the most verbose, with 22 words per tweet on average.

Finally, I created word clouds for customer tweets and for customer service interventions. I found that before being addressed by a customer service agent, customers wrote about their flight, seating, gate, time, bookings, delays, and upgrades. Customer service agents replied politely, often writing please and sorry, while directing concerned customers to share their booking reference, confirmation number, and/or to send the airline a direct message to address the issue privately.

3. Training and Modeling

When developing my model, I tested four different encoding schemes: frequency within each document (freq), frequency over all documents (tfidf), binary/onehot encoding, and word count. I discovered that encoding text using the frequency within each document gave the most accurate results, with 76% accuracy on the test set.

I used my trained model to predict the sentiment of customer tweets before and after a customer service intervention by an airline. I discovered that for each airline in my dataset, interacting with customers on Twitter led to an improvement in sentiment. The best performing airline was JetBlue, with the percentage of positive tweets increasing by 26%, and AirAsia performed the worst, with the percentage of positive tweets rising by only 8%.

However, AirAsia also had the highest percentage of positive customer tweets before a customer service interaction, with 65% of customers sending tweets that were labelled as positive before interacting with a customer service agent.

As mentioned previously, AirAsia had the slowest response time. The airline took 12 hours to respond to tweets on average. However, since customer inbound tweets tended to be scored as positive, I suspected that AirAsia's lack of activity on Twitter led many customers to deal with customer service complaints in other ways, such as by phone or email.

While AlaskaAir's rapid response time did not yield the best improvement in customer sentiment (three other airlines scored higher) it does appear to be essential for an airline to respond within at least 2 hours.

I explored four different factors that were correlated with the change in the percentage of positive tweets:

1. **the percentage of customers who follow up after getting a response from a customer service agent**
    - There is a positive correlation between the percentage of customers who follow up after getting a response from a customer service agent and the increase in positive sentiment. This correlation could mean that unsatisfied customers are not following up after a customer service intervention. However, this correlation is not conclusive. Some conversations end with airlines requesting that customers send them private direct messages on Twitter or send links to enter a private chat on a different website.
2. **response time**
    - There is a negative correlation between response time and increase in customer satisfaction. Airlines that take longer to respond tend to have a much smaller increase in the percentage of positive tweets.
3. **number of words per tweet**
    - There is a slight relationship between the average number of words per tweet and the increase in satisfaction. This could suggest that customers are happier when customer service agents provide detailed responses.
4. **total number of tweets**
    - The total number of tweets by an airline is negatively correlated with the increase in the percent of positive tweets. I speculate that this could be for one (or all) of the following reasons:
        - The more tweets it takes for a customers' issue to be resolved, the less likely they are to be content after the interaction. Airlines should aim to resolve issues within as few tweets as possible.
        - The airlines in the dataset with more data are just more active on Twitter, and therefore customers are more likely to view Twitter as a venue to voice complaints. Customers of airlines that are less active on Twitter might be more likely to make a phone call or write an email in order to reach a customer service agent.

- The airlines with more data are simply larger airlines, and therefore serve more customers. The more customers you serve, the more diverse the range of issues and the number of issues to resolve is, which may lead to a smaller percentage of customers having their issues resolved in a satisfactory way.

Finally, I used Latent Dirichlet Allocation (LDA) to model topics in the customer tweets before a customer service intervention and in the responses from airline accounts. I focused my analysis on the most successful airline, JetBlue.

I found that JetBlue customers are mostly concerned about four key areas: flight delays, flight changes, flight experience, and issues with their accounts. This narrow range of concerns suggests that there is potential for the customer service responses to be automated, which would allow human agents to focus on more ambiguous concerns.

4. Conclusions and Recommendations

From comparing the improvement in sentiment to the various differences in the customer service agents' responses, I conclude that there are some specific actions airlines, or any business, can take in order to better resolve customer issues on Twitter:

- Customers prefer faster responses. In order to have happy customers, airlines should respond as quickly as possible. Responding within at least 2 hours is essential.
- Customers prefer a small number of detailed tweets over multiple follow ups. Airlines should try to address concerns with as much detail as possible in as few tweets as possible, so that customers are not obliged to keep asking for clarification.

There are a few caveats that should be considered in order to properly understand these results:

- I based my analysis on comparing tweets before and after a customer service interaction, and satisfied customers may be more likely to follow up with a customer service rep after their issues have been resolved. This particular project does not address whether or not those customers who did not respond did so simply because they forgot, felt it wasn't necessary, or were genuinely frustrated and gave up on trying to get help on Twitter.
- Twitter only represents a small subset of customers, and therefore these results are not necessarily indicative of general trends for an enterprise. In order to ensure that the stated goal of increasing customer retention by 5% is achieved, airlines should conduct review

monitoring on surveys, emails, phone calls, and any other ways customers communicate their concern or appreciation.

I recommend the following continuations for this project:

1. Named Entity Recognition (NER) could be used to identify if a customer issues are associated with a particular airport.
2. The neural network was 76% accurate on the test set. I only used 5% of the available data (40,000 positive tweets, and 40,000 negative tweets) in order to be able to train and test the network locally. Better accuracy could be achieved by deploying the model on distributed servers over a cloud environment in order to process the full 1.6 million tweet dataset.
3. Time series analysis could be done to identify if there are seasonal trends around certain issues so that airlines can preemptively address concerns customers are likely to have in a given season.
4. While each customer is unique, customer concerns can generally be aggregated into distinct categories. I have shown that JetBlue customer issues can be grouped into 4 major groups (flight delays, flight changes, flight experience, and issues with customer accounts). Chatbots could be developed to answer some of the issues that are raised repeatedly. This would allow customer service agents to address more nuanced questions, leading to faster response times and higher customer satisfaction.