Case Study: Analysis of Customer Support Interactions on Twitter

By
Hassan Javed

1. Introduction

Social media platforms have become an important medium for customers to voice their concerns, seek assistance, and engage with businesses. To maintain a favourable brand image and improve customer satisfaction, companies must address customer queries effectively.

This case study focuses on the analysis of the sentiment of customer support interactions on Twitter, using a publicly available dataset which consists of real interactions between customers and support accounts of some of the world's most popular brands.

2. Objectives

The objective of this study is to obtain valuable insights into customer interactions by utilizing Data Science principles and best practices through analyzing customer sentiments. The resulting analysis can highlight potential areas of improvement in products as well as inform whether customer support processes and strategies need a revamp.

3. Data Collection and Description

The data in this case study is sourced from a publicly available dataset on Kaggle. It includes around 3 million tweet interactions between customer support accounts and users between 2008-05-08 to 2017-12-03. Moreover, the dataset is made available under a creative commons license i.e. CC BY-NC-SA 4.0. Under this license users can freely share and adapt this dataset for non-commercial purposes but must provide attribution.

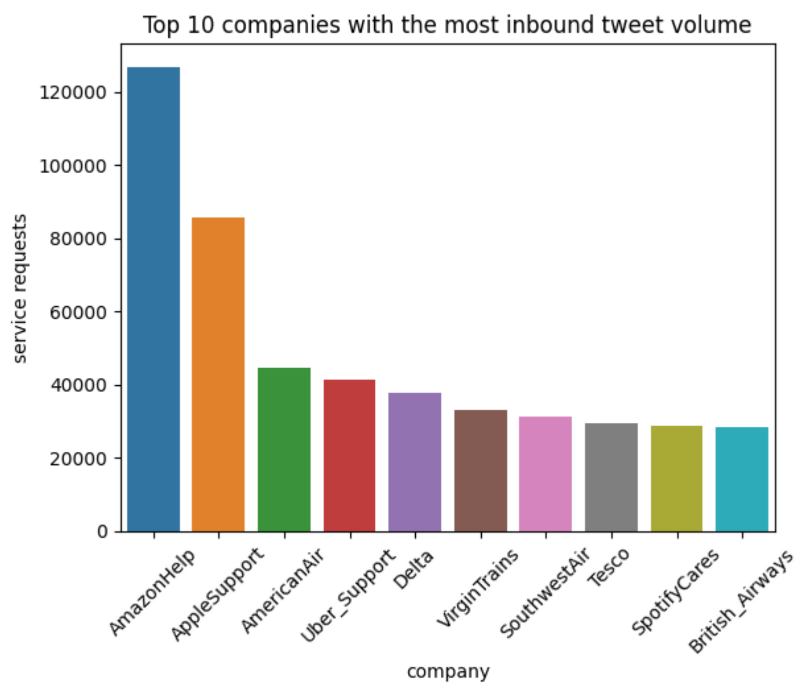
The dataset is in tabular format with 7 columns and their description is as follows:

- Tweet_id
  - The unique ID for this tweet
- Author_id
  - The unique ID for this tweet author (anonymized for non-company users)
- Inbound
  - Whether or not the tweet was sent (inbound) to a company
- Created_at
  - When the tweet was created
- Text
  - The text content of the tweet
- Response_tweet_id

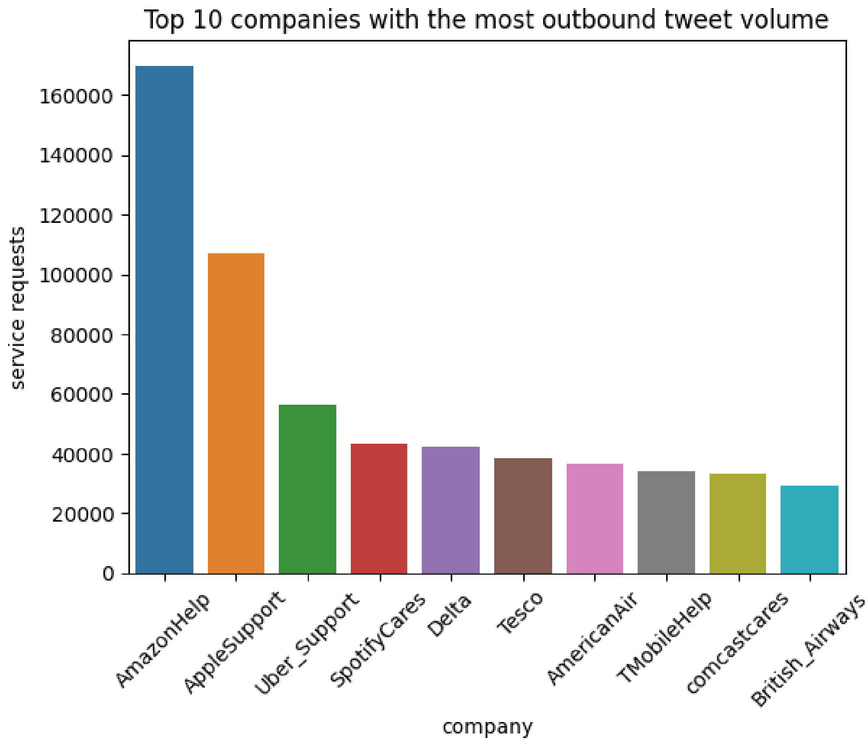
        ➢ The tweet that responded to this one, if any
- ❖ In_response_to_tweet_id
    - ➢ The tweet this tweet was in response to, if any

4. Exploratory Data Analysis (EDA)

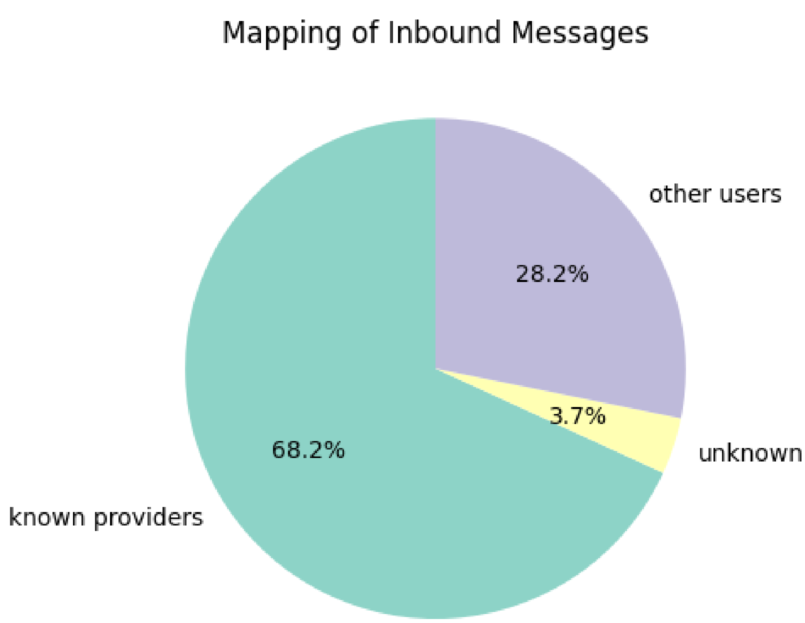
An exploration of the dataset reveals several interesting facts about service providers, customers and their interaction. The dataset consists of customer representatives from 108 companies. Here are the top 10 companies with the most inbound tweet volume.



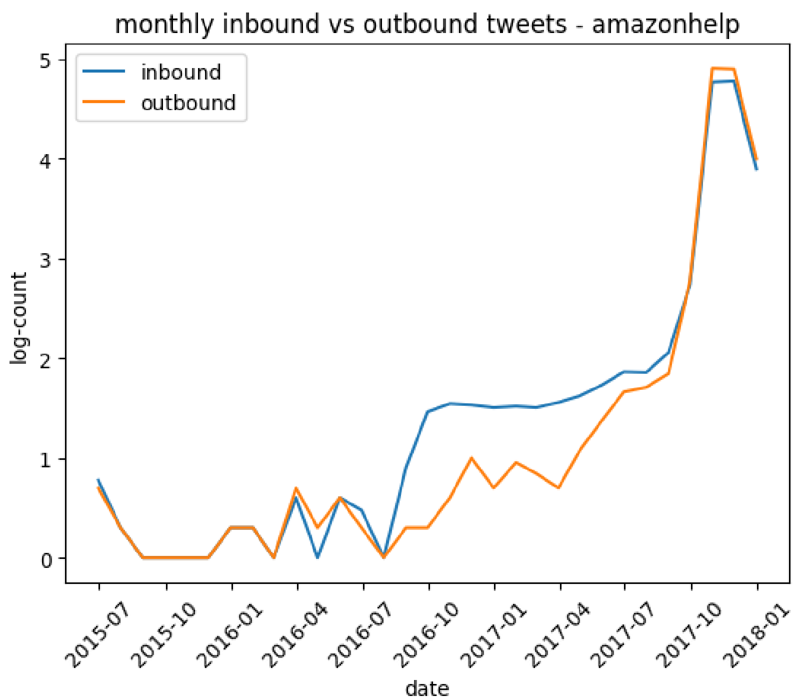
However, the top 10 companies with the most outbound tweet volume are slightly different. These are the companies with the most responses to customer requests.



If we look at inbound requests closely, we can identify that these are addressed to three different groups i.e. the majority are directed to known service providers, then there are tweets which are addressed to other users or their responses to a service request thread and finally there are some tweets marked as inbound but do not mention any service provider or other user.



In this study, we’ll focus on **amazonhelp** tweets as it has the most inbound and outbound requests as shown in the above charts. The following chart illustrates the no. of inbound requests to responses from amazonhelp over time. The y-axis is on log axis to provide better resolution. It shows an interesting pattern, that between 2016-07 to 2017-0 the number of inbound requests exceeded the responses from this account. In the rest of the chart, responses seem more aligned with the requests.



5. **Data Preparation and Feature Engineering**

As mentioned in the previous section, we have chosen a subset of data related to one of the largest accounts in this dataset i.e. **amazonhelp**. Specifically, we’ll be focusing on the inbound requests from users on this account as the customer sentiment is our primary focus in this study.

Furthermore, we’ll apply the following two transformations to the ‘text’ column of our dataset, which contains tweets from customers.

1. We’ll be removing emoticons to reduce the tweet text to natural language only.
2. We’ll be filtering to select only those tweets for which the language is English.

The above two transformations would clean the data and prepare it for sentiment analysis.

6. Modeling and Analysis

In order to model the dataset for sentiment analysis, we can take several different approaches including unsupervised and supervised models. Unsupervised models like VADER and SentiWordNet etc. assign sentiment scores to individual words based on lexical and grammatical rules. These methods can be computationally less expensive but they might be less sensitive to context and there is no ground truth to compare the results.

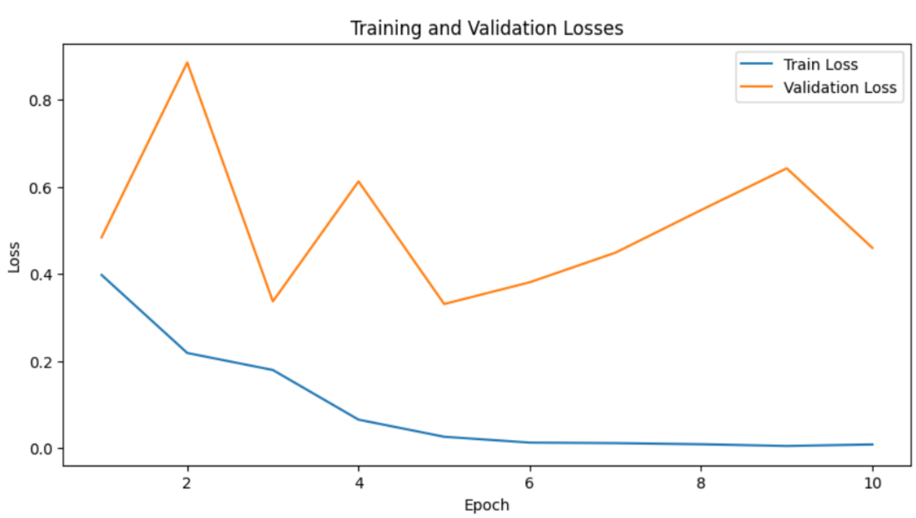
Supervised approaches portray the problem as a classification problem and require labelled datasets. While obtaining labels is time consuming and expensive, it has the advantage that we can evaluate our model more thoroughly against a set of ground truth.

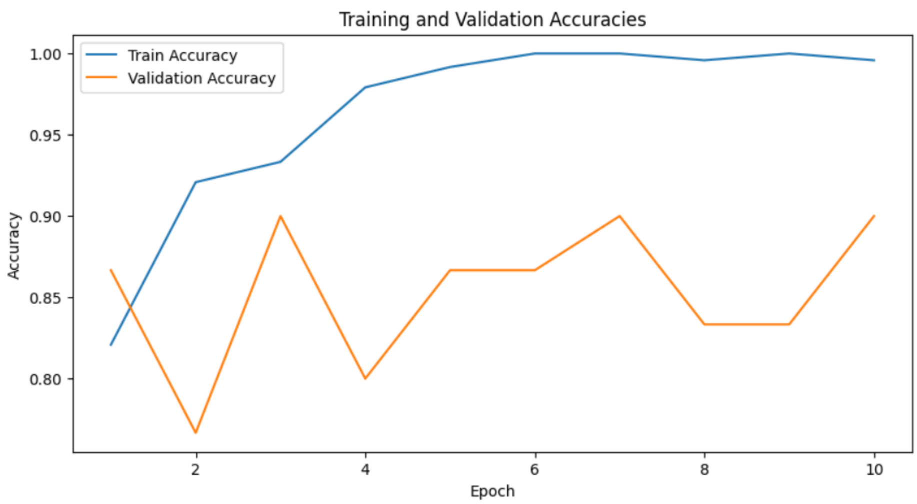
For this reason, in this study we pose sentiment analysis as a classification problem. Furthermore, we'll take a small sample of the amazonhelp dataset and label it into three sentiments i.e. positive, negative and neutral. Although customer requests are mostly biased towards the negative sentiments, we took a balanced annotated sample.

Due to the lack of extensive labeled data for this problem, the chosen modelling methodology is based on fine-tuning a variant of RoBERTa (Robustly Optimized BERT approach) from HuggingFace called 'twitter-roberta-base-sentiment'. This variant is a sentiment classifier trained on thousands of general tweets data. We have fine-tuned this model specifically on amazonhelp customer support data with augmentation. The augmentation methodology utilizes contextual word embeddings from 'bert-base-uncased'. The model is trained on 10 epochs with AdamOptimizer and learning rate fixed to 0.1.

Evaluation and Validation:

To evaluate the model, the data set was trained on a 60-10-10 stratified split with equal classes on train-val-test sets.



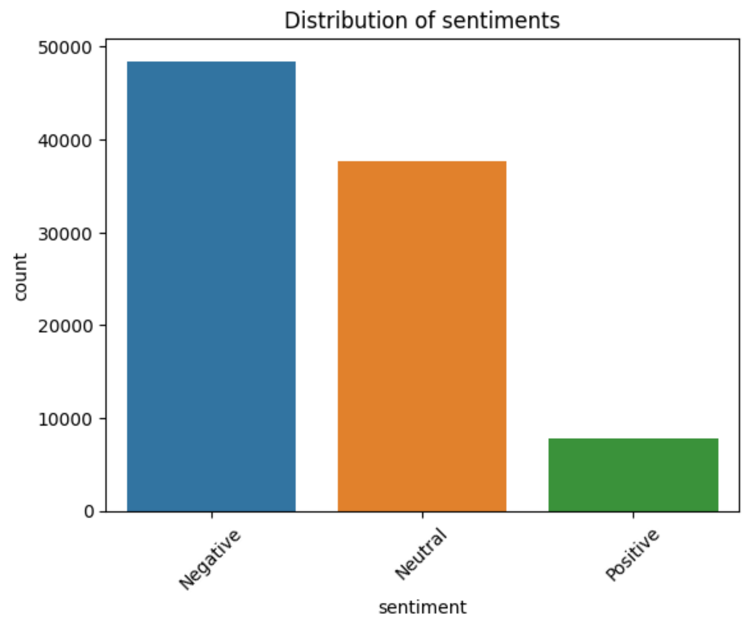


The train/validation curves start smoothing out around epoch no 10, where training curves have flattened. At the end of the 10th epoch, we achieve a train accuracy of 99% and a validation accuracy of 90%. The model achieves a 96% accuracy on the test set.

Here are the results in tabular format:

	train_loss	val_loss	train_acc	val_acc
0	0.397832	0.484213	0.820833	0.866667
1	0.218767	0.885393	0.920833	0.766667
2	0.179619	0.337100	0.933333	0.900000
3	0.065655	0.612610	0.979167	0.800000
4	0.026402	0.331064	0.991667	0.866667
5	0.012949	0.381095	1.000000	0.866667
6	0.011834	0.449022	1.000000	0.900000
7	0.009178	0.546781	0.995833	0.833333
8	0.005256	0.642590	1.000000	0.833333
9	0.008708	0.459832	0.995833	0.900000

Following is a distribution of sentiments across the whole dataset. As expected most of the sentiments are negative and positive sentiments are expressed in the least no. of tweets.



7. Limitations of the analysis

Following are some limitations of the analysis:

- The current analysis is based on one of the service providers out of 108. This can be extended to more customer support accounts. The model is expected to perform well on other customer support accounts but is not tested.
- The model was built on a small sample of labelled data. We can experiment with more customized models in the presence of a large no. of labeled examples.
- The current analysis only classifies the sentiment of individual tweets initiated by users towards the service accounts. These tweets in general have a negative sentiment. In an improved approach, we can transform tweets related to one user in a given time window into conversations and instead assign sentiment scores to conversations as the sentiment of an overall conversation might be more useful.

8. Conclusion

In this study, we analyzed the twitter customer support dataset and chose the amazonhelp subset to build a sentiment classification model. For this purpose, a small no. of tweets were hand-labelled and a general RoBerta based twitter sentiment model was fine-tuned on this customer support tweets data to provide a decent accuracy score on the test set.