

Twitter Sentiment Analysis for Brands to Understand Customer Opinion – Progress Report

Srikanth Reddy Pullaihgari

srp10

The objective of this project is to glean insights through analysis of Twitter feeds for brands to understand their recent customer opinions and general sentiment. The key components of the project that I had outlined were Data Collection & Processing, Data Analysis and Data Visualization.

I have since broken these into a grouping of tasks including: researching the various ways of gathering Twitter feed(2 hours), collecting the data (5 hours), input/output design (3 hours), modelling/programming to analyse sentiment (10 hours), testing and final report (5 hours).

As a start, I began to research and understand the twitter feed gathering processing and the Tweepy API. I underestimated the process and as a result ran into a few issues. Firstly, I realized I needed a Twitter developer account. I applied for one and duly got rejected the very next day. I have since researched other ways to obtain Twitter feed without needing a Twitter developer account. I now intend to use the 'snsrape' library to scrape tweets as this allows fetching of tweets based on specific dates without needing to use Tweepy API. I then plan to use TextBlob to analyse the sentiment.

I still need to understand and design the interface for user input and the final output. I also need to figure out the data collection part along with the constraints on how much data can be analysed within a reasonable amount of time for the program to run without a long waiting time for the user. There are other problems I am likely to encounter as I test the feed and run the model. Given these problems, I am also considering limiting the number of brands through a drop-down model so user is only allowed to select one amongst a set of chosen brands for analysis.

Based on my work in the last couple of weeks, I have a better appreciation of the complexity of the work I have set out to do. I have realized I might not be able to complete everything I have originally set out to do. I am also in the process of simplifying the output this program would produce given the time constraints. While I had planned to produce some level of data visualization, I now plan to restrict the output to a document format which will display the sentiment analysis (negative/positive/neutral) for the brand selected and will include some example tweets outlining the sentiment.