Sentiment Analysis on Airline Twitter Data

Mohit Ravi
Nishaat Farheen
Anuj Thacker
Lanyin Zhang
Gautam Gite

***Abstract* — Sentiment analysis over Twitter offer organisations a fast and effective way to monitor the publics' feelings towards their brand, business, directors, etc. In this project, we perform Sentiment analysis on tweets from five domestic US airlines. We then classify the tweets into two categories – positive and negative and carry out further analysis to compare the airlines.**

I. INTRODUCTION

The amount of user-generated content on the Internet has risen exponentially over the last decade, and such content is now always at our fingertips. As a result, nearly all our decision-making is social; before buying products (attending events, trying services, voting for candidates, visiting specialists), we see what our peers are saying about them. The fate of a new offering is often sealed by those evaluations. Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. It has become very useful for firms to monitor their brands. Sentiment Analysis is also increasing used by firms to improve their customer service. Through this businesses can quickly identify positive talking points around their brand to measure, inform and evaluate their digital strategy. Conversely, they can also identify negative conversational threads and emerging threats to their reputation. In this project, we have used sentiment analysis on Twitter data for five domestic US airlines. They are Jetblue, Southwest, United, Delta and American. The sentiments were classified as either positive or negative. We compared the user sentiments for two time periods – before thanksgiving and after thanksgiving and visualized the results.

II. MOTIVATION FOR THE PROJECT

Anyone who travels regularly recognizes that airlines struggle to deliver a consistent, positive customer experience. Through extensive interview and survey work, the American Customer Satisfaction Index (http://theacsi.org/) quantifies this impression. As a group, airlines falls at the bottom of their industry rankings, below the Post Office and insurance companies. Consequently, brand values of domestic US airlines are at an all-time low. Therefore, Sentiment analysis is crucial for airlines to understand customer grievances and in turn improve their service.

III. DATA PREPARATION

We used R programming language to carry out sentiment analysis. R has a package called 'TwitteR' which directly connects to the Twitter API. We used this package to load tweets into the R environment by using the search function in the package. For example, @united was used to load tweets for United Airlines and so on. The downloaded tweets for all airlines were stored in a Dataframe format. The drawback in this method is that Twitter only allows a maximum of 1500 tweets to be downloaded.

Before building the model, we need to clean the data. Some of tweets contains hashtags, retweets, URL's, mentions etc which are not necessary to quantify sentiments. So the data is cleaned by removing the above mentioned things.

To conduct sentiment analysis, we need a dictionary of words containing positive and negative words. In this project, we used Hu and Liu's (2004) lexicon to conduct the analysis. This lexicon includes around 6800 seed adjectives with known orientation of 2006 positive and 4783 negative words. So, now we are ready to build a model to classify the sentiments after cleaning the tweets and loading the opinion lexicon.

IV. SENTIMENT DETECTION

Determining sentiment polarity is done by comparing the tweets against a predefined corpus of subjective words from Hu and Liu's lexicon. In the algorithm, we match the words in the tweets against the positive and negative words in the dictionary. Then, we find out the total positive matches and negative matches for each tweet. Finally, the score for each tweet is calculated as:

Score = Sum of Positive matches – Sum of Negative matches. The entire procedure is summarized below.

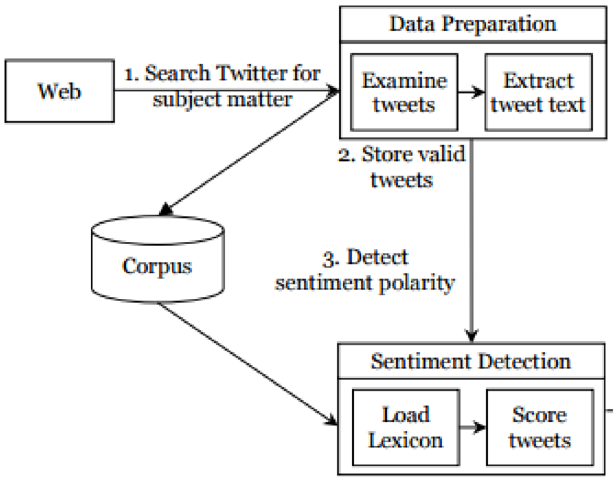


Figure 1 – Summary of the model.

After building the model, we need a sanity check to see if this model works. So, we decided to run this model against three sentences that were not part of any tweets we collected.

For the first sentence – "Had the best time of my life in New York! What a wonderful experience!", we received a score of 2. Since this is clearly a positive sentiment, as expected we receive a positive score.
For the second sentence - "The food was absolutely horrible! It was so bland!", we received a score of -2. Again this was as per our expectations since it was clearly a negative sentiment.
For the third sentence – "Wow!! You learnt how to make Coffee?? That's amazing! I'm so impressed", we received a score of 3. We know that this is a sarcastic comment and should be classified as a negative sentiment. But the model gives a positive score because of words like "amazing" and "impressed". This shows that model does not work for sarcastic comments.

After this, we run the model against the twitter data for all the five airlines. Below is the histogram of scores for the airlines.

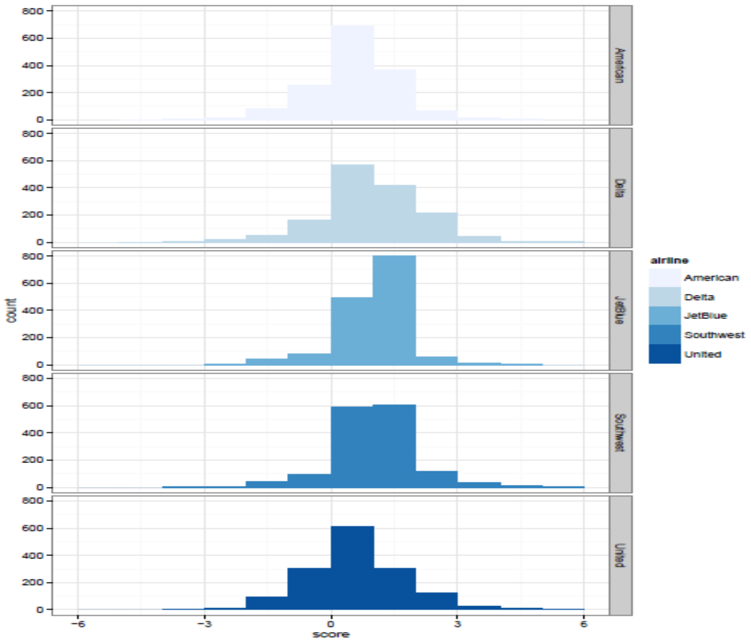


Figure 2: Histogram of scores calculated

We can see that majority of the tweets have a score of 0. To capture true positive or negative sentiments we focus on the extreme left and right of the distribution. All the scores greater than 2 are classified as positive and all the scores less than 2 are classified as negative.

V. Results

The results after quantifying the negative and positive sentiments are shown below.

Airline	Code	Positive count	Negative count	Total count
Delta	DL	272	77	349
Southwest	WN	158	54	212
Jetblue	B6	79	44	123
United	UA	164	115	279
American	AA	92	101	193

Figure 3: Total Sentiment count.

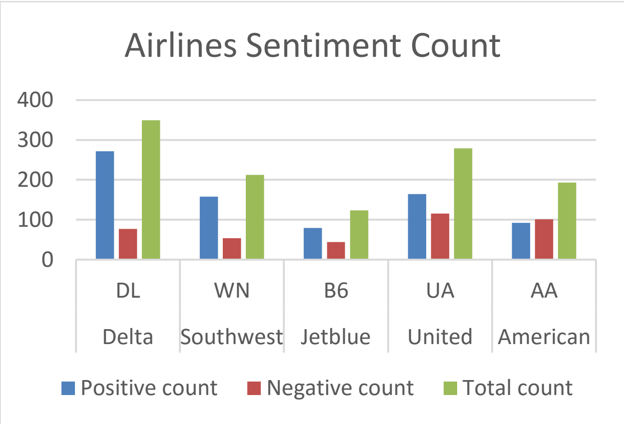


Figure 4: Sentiment count in bar graphs

As we can see Delta airlines has the highest number of positive sentiments followed by United and Southwest. United had the highest number of negative sentiments followed closely by American airlines.

Although, we cannot come to a definite conclusion that one airline is better than the other because our sample size is small.

Effect of Thanksgiving

The sentiment scoring in the section above was done for tweets collected before thanksgiving. We also ran the same model for tweets collected after thanksgiving period. Thanksgiving period is one of the busiest days for air flights. So, if an airline performs well on one of the busiest days of the year, it bodes well for its brand.

To compare the effect of Thanksgiving on the sentiments, the percentage of positive sentiments was compared for the airlines as shown. The percentage of positive sentiments increased drastically for Southwest airlines whereas for other airlines, it decreased. This could imply that Southwest performed really well during Thanksgiving and the other airlines were not able to manage the rush.

	PercentPos_B4	PercentPos_AF
Delta	77.9%	61.0%
Southwest	74.5%	91.7%
Jetblue	64.2%	64.7%
United	58.8%	46.3%
American	47.7%	42.3%

Figure 5: Percentage change of Positive comments.

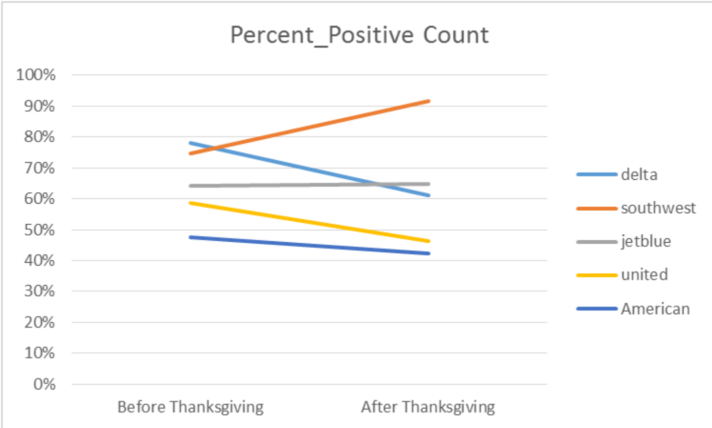


Figure 6: Change in Positive comments before and after Thanksgiving.

Also noteworthy result from the above graph is that Delta and United airlines had a decrease in their positive comments. Whereas, American and Jetblue airlines were more or less steady throughout.

To see if the difference in positive comments is actually due to thanksgiving, we conducted a Wilcoxon's Signed Rank test. This test checks if there is any difference between the sets of paired samples taken from two different populations. In this case the two different populations would be before thanksgiving and after thanksgiving.

```
> wilcox.test(comparison$PercentPos_B4, comparison$PercentPos_AF, paired=TRUE)

	Wilcoxon signed rank test

data:  comparison$PercentPos_B4 and comparison$PercentPos_AF
V = 9, p-value = 0.8125
alternative hypothesis: true location shift is not equal to 0
```

Figure 7: Wilcoxon's Signed Rank test results.

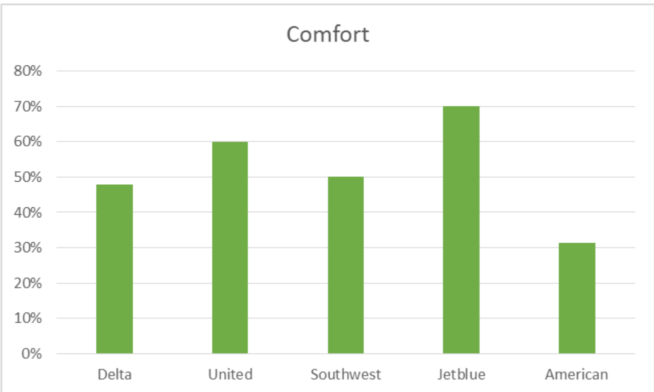
After conducting the test, we find that the p-value for this test is 0.8125 which implies that we cannot reject the null hypothesis that the two populations are the same. So, this implies that thanksgiving did not have an impact on the population of tweets.

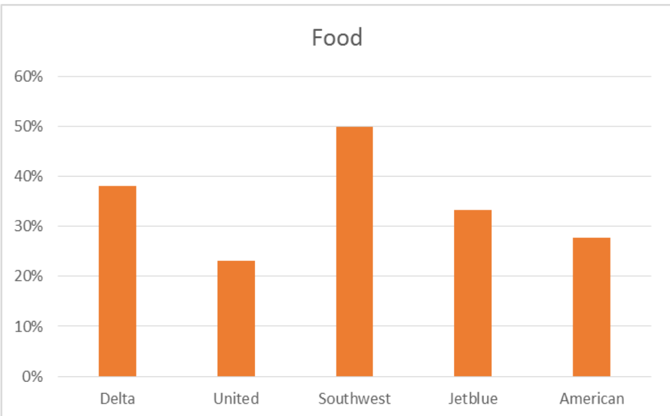
The sample means of all the airlines were compared using the Welch two sample t-test. The order of sample means obtained using the t-test is as follows: **Southwest > Jetblue > Delta > United > American** This is the same as the absolute value ordering, which means that it is statistically significant.

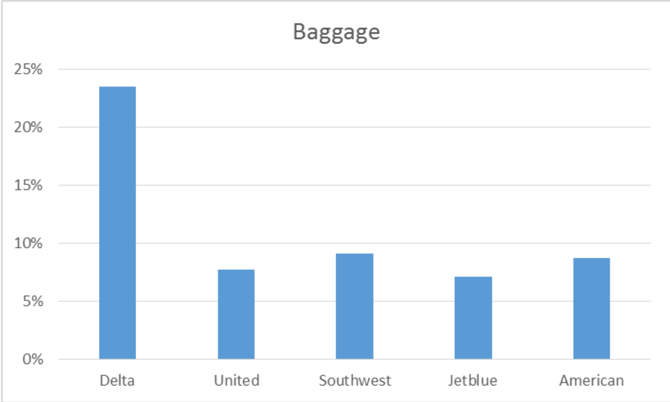
Subtweets Analysis

Besides computing the positive or negative sentiments, it is also very useful to understand the context behind these sentiments. If a customer had a good experience with an airline, he/she will express positive sentiments about the food served, In-flight entertainment, Baggage handling etc. Conversely, if a customer had a bad experience with an airline, he/she will vent their frustrations on missing baggage's, delayed flights, horrible food etc.

We decided to run Sentiment analysis on three keywords from the tweets that were collected. They were food, comfort and baggage. The results of the positive sentiments for the three airlines are shown below.







Figures 8: Comparison of Subtweets

As shown in the above graphs, Jetblue scored the highest in the comfort category followed closely by United airlines. Southwest scored the highest in the food category followed by Delta. For the baggage category, most of the airlines had only negative sentiments from customers but Delta was the only airline which received significantly more number of positive sentiments regarding baggage.

VI. Social Network Analysis

For the conclusion of this project, we performed a social network analysis of the tweets using Tableau. The interaction of millions of individuals through social networks generates vast amounts of data. The value here is in determining what forms public opinion and why viewers might feel positively or negatively about a certain airline. Gauging public opinion is not an easy target, but Twitter and the people using it, give a real-time view into the collective mind.

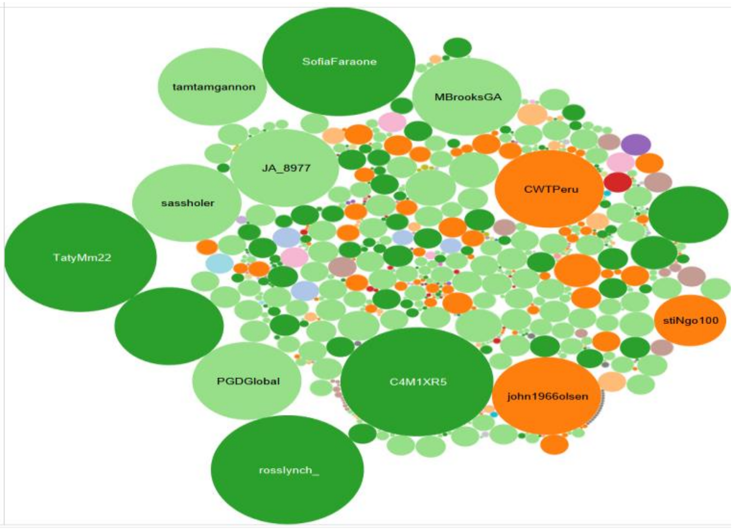


Figure 9: Influential Users in an airlines twitter's Network

The above graph tells us the most influential users contributing towards an airline's score. Colour represents the device usage. Size is the number of retweets for every single user. Identifying the influential nodes beforehand can help an airline reach out to bigger parts of the network faster and wider and spread their promotions in an efficient way. Combing the above with a user's network data like the number of followers, we can calculate the betweenness centrality to evaluate the force of an individual on a network.

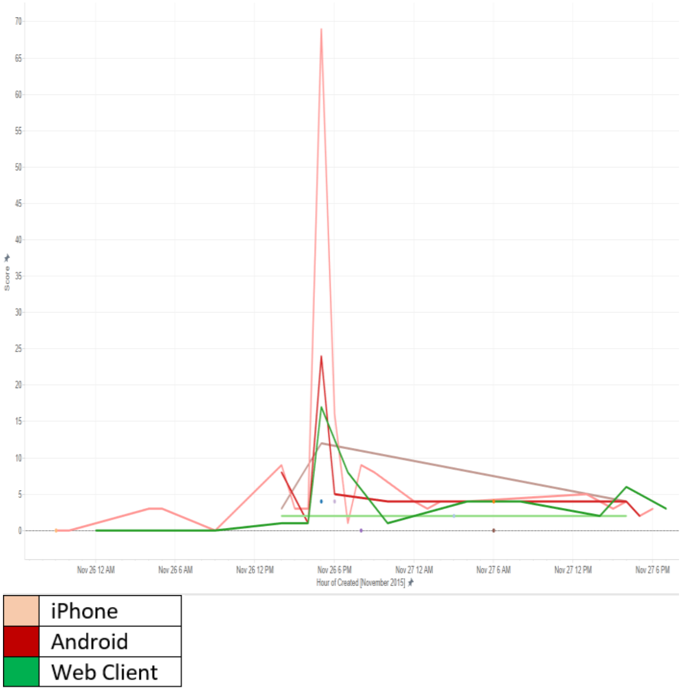


Figure 10: Score Vs Device Usage on Thanksgiving day for Delta Airlines.

The graph above shows cumulative score for Delta on the Thanksgiving Day broken down by hour of tweets created. The peak being between the hours of 4 to 6 PM. When comparing for more than one airline the number of devices and scores can be normalized to give a better perspective.
This interprets that for Delta, their highest scores are provided by iPhone users. This might prompt the company to release app only offers on the day of Thanksgiving Day. They can either boost marketing on one device or try to bring up the weaker ones to make a competitive market.

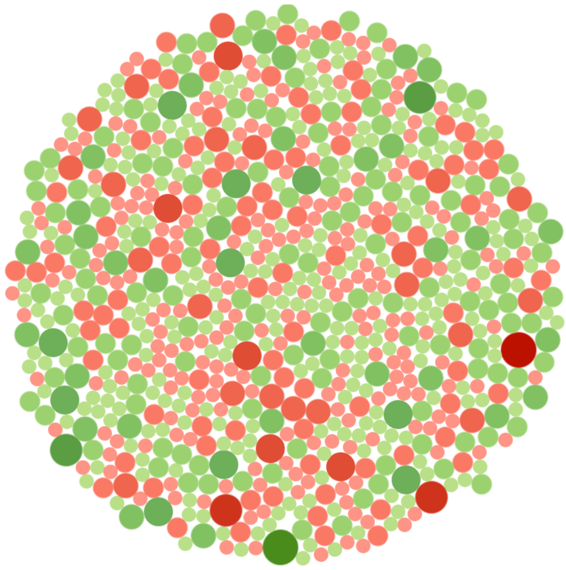


Figure 11: Sentiment Bubbles

An overall greener trend shows in general a positive inclination of sentiment towards an airline whereas an overall red trend shows a negative sentiment.
If monitored real time we can observe the colours and size change, which may provide an overview of the mood swings of people and which way their sentiment is holding.
Most analysis right now focus only on post-hoc interpretation of tweets, but if extracted real time, we can not only have sight over the present but may also have an insight into the future.

VII. References

1) http://www.inside-r.org/howto/mining-twitter-airline-consumer-sentiment

2) https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

3) http://mashable.com/2010/04/19/sentiment-analysis/#Lm8pYnWjMiqZ