

Twitter Sentiment Analysis

Completed for the Marketing Team at Resale Electronics

By Rachel Fein

Summary

This analysis was completed for Resale Electronics to properly identify their target audience for their upcoming Twitter Ad Campaign. The goal is to use machine learning to find those tweets and include those users in their campaign. The final logistics regression model showed:

- The final model had an overall accuracy of 69% at correctly identifying the emotion in unseen tweets.
- The model is best at identifying neutral emotion tweets, with a precision of 73% on unseen data.
- As a result I recommend that more data is collected to improve the model.



Outline

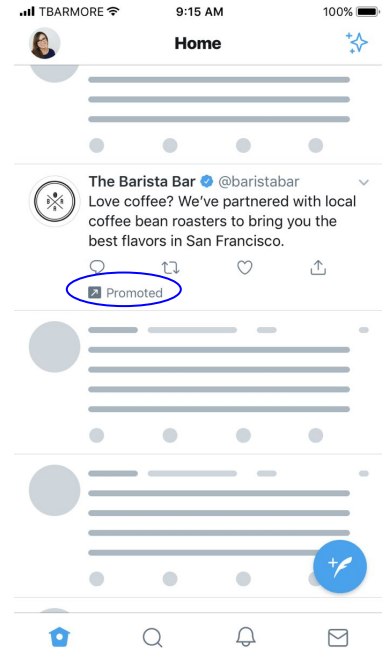
- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

This analysis was complete for the Marketing team at Resale Electronics. In preparation for your Twitter Ad Campaign the team is in need of a model that can correctly identify your target audience.



Your team's goal is to only have people who have either tweeted positively or neutrally about Apple or Google products included as recipients of their Ad Campaign.



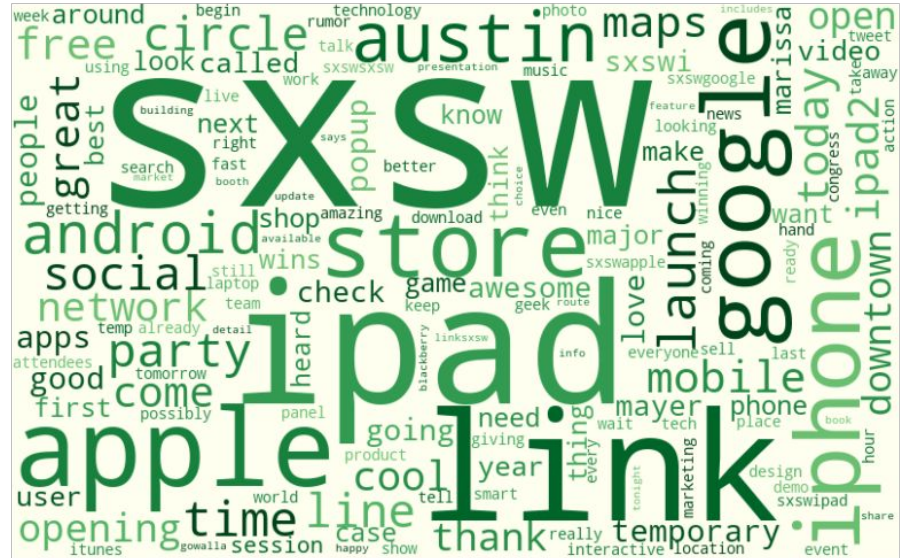
Data

This analysis used 9,065 tweets that were ranked by humans as having a positive, negative, or neutral sentiment towards Apple or Google. The variables used in modeling were the tweet text and emotion.

Most Common Words in Neutral Tweets



Most Common Words in Positive Tweets



Methods

Words that cause distraction for the model were removed, such as overly common & simple words. As well, words with the same root were reduced to their simplest form so the model would recognize them as the same word.

Multiple different models were used and tune to get the best performing model. The final model was a logistic regression model.


Methods were applied to attempt to help with the imbalance of the dataset.



Results

The business problem was to make a model that would identify people who are tweeting positive or neutral tweets about Google or Apple products. The goal in identifying these people is to narrow down the recipients of your twitter ad campaign & not spend money on people who won't buy your products.

The analysis does create a model that could aid in narrowing down potential targets for Resale Electronics' ad campaign, however sponsor tweets are expensive! The current accuracy of 69% of correctly identifying your target is not high enough.



Cost of a Promoted Tweet

Average **\$1.35** each time someone clicks, replies or retweets your tweet.

WebFX

Conclusions

In summary this analysis showed:

- The final model had an overall accuracy of 69% at correctly identifying the emotion in unseen tweets.
- The model is best at identifying neutral emotion tweets, with a precision of 73% on unseen data.
- As a result I recommend that more data is collected to improve the model.

Next steps:

- Attain more data. More data will improve the model's performance.
- Using more complex Deep Natural Language Processing methods such as neural networks.

Thank You!

Email: rachellfein@gmail.com

GitHub: @rachelfein

LinkedIn: [linkedin.com/in/rachellfein/](https://www.linkedin.com/in/rachellfein/)