TWITTER SENTIMENT ANALYSIS

Shannon F. Hunley

November 4, 2025

OVERVIEW

- Our company, Apple, would like to evaluate public sentiment regarding our products, so that we can make business decisions with this information in mind, particularly as compared to our competitor, Google.

- Our goal with this project is to build a model which can rate the sentiment of a Tweet based on its content so that we may evaluate public sentiment in the most cost-effective manner.

DATA AND METHODS

- Our dataset is sourced from CrowdFlower via data.world. It contains over 9,000 Tweets which were rated by humans as positive, negative, or neither.

- We will use predictive ML modeling because this allows us to rate the sentiment of a tweet using our model rather than human raters, which are more costly.

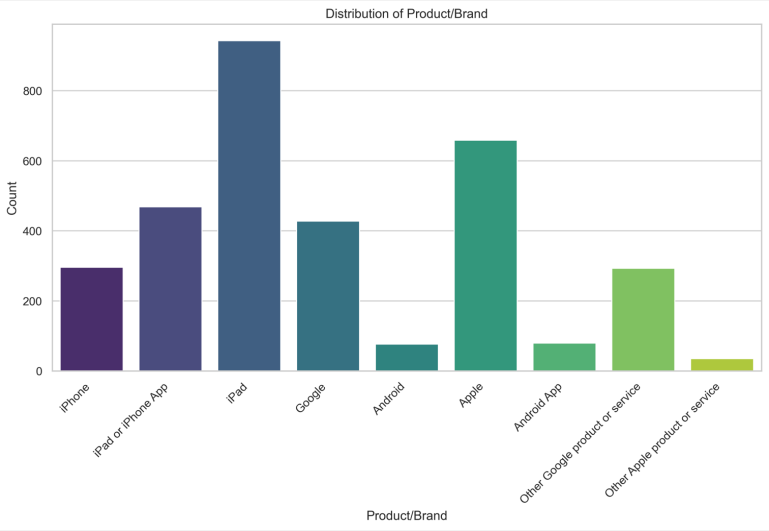
LIMITATIONS

- The limited amount of data is not ideal for modeling and additional data collection, particularly for the minority class, is warranted to improve the model.
- Selection bias - Twitter data only may not be representative of overall sentiment since a small portion of the consumer market actually tweets. Data collection from other social media platforms and/or consumer surveys might be more informative.

TWEETS BY PRODUCT/BRAND

The data contains a lot of Tweets about our brand and products compared to Google's.

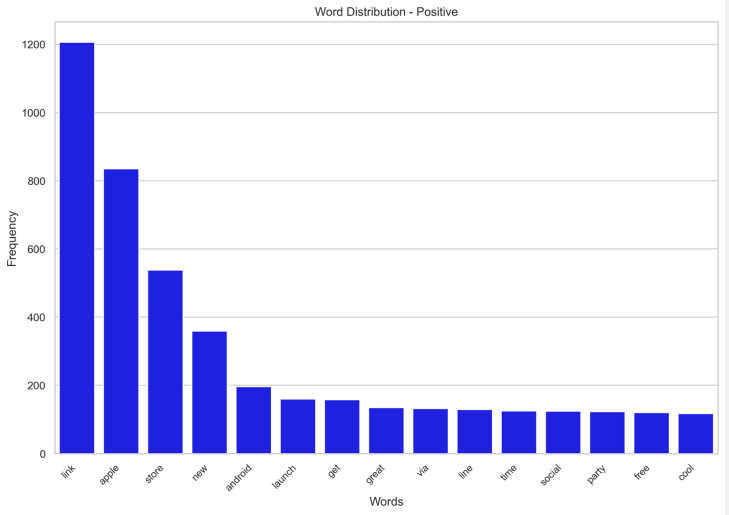
"iPad" and "Apple are the top two in the distribution.



MOST COMMON WORDS

We find that "Android" ranks in the top 5 words within Tweets considered positive sentiments.

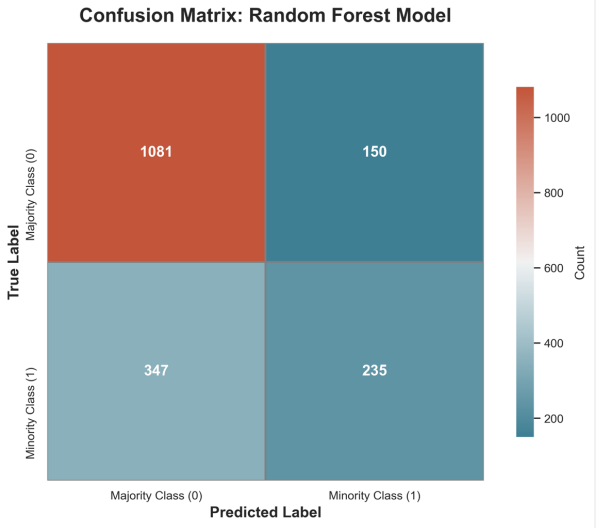
This is the case even though our Apple products and brands were mentioned many more times, and our competing product in not in the top 15 most words.



MODELING

- Our best model was a Random Forest binary classifier, which currently predicted sentiment with a precision of .71.

- We used precision because we want to prioritize preventing false positives… and catch the most possible true negative sentiment.



CONCLUSIONS

- Implement predictive advanced ML modeling to rate the sentiment of Tweets based on their content.
- Do this in place of paying human raters for a more cost-effective approach.
- Evaluate the possible reasons for the lack of chatter about iPhone and/or iOS versus Google's Android

NEXT STEPS

- Additional data collection from multiple sources to address selection bias (other social media platforms, customer survey for public opinion, etc.)
- Additional data collection to address class imbalance (more real data will improve our model)

Thank you!

Questions are welcome.

Email: Shannon.f.hunley@gmail.com

Github @ sfp13VA

LinkedIn: linkedin.com/in/Shannon-hunley