2021

Helio Handset Sentiment Analysis

ALERT! ANALYTICS

BRIAN MATTIS (C5T3)

Overview

Alert! Analytics is conducting a broad web sentiment analysis on behalf of Helio. Helio is developing a novel mobile app for use by international aid workers in developing countries. To simplify development, support, and app rollout (including training), Helio will be choosing a single handset for implementation.

The primary goal of Alert! Analytics' sentiment analysis is to gain insight in the attitudes towards the Apple iPhone and the Samsung Galaxy. With this information, the best handset for the product can be selected.

Findings

Alert! Analytics gathered 30,000 relevant web pages across the internet, and assigned the sentiment of the page into one of four categories:

- Negative
- Somewhat Negative
- Somewhat Positive
- Positive

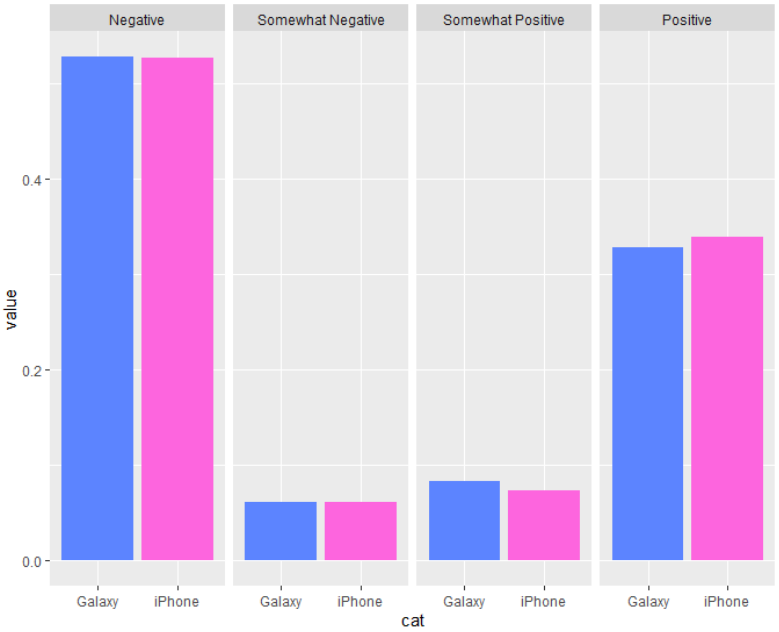

Figure 1: Modeled sentiment, normalized

Sentiment	Negative	Somewhat Negative	Somewhat Positive	Positive
Galaxy Sentiment	17341	2024	2751	10768
iPhone Sentiment	17307	2029	2413	11135

Figure 2: Predicted sentiment values

Upon compiled analysis of these pages we found that the sentiments of the iPhone and Galaxy were extraordinarily closely matched. As shown in Figure 1, the handsets were nearly matched in all categories, with slight offsets between *Somewhat Positive* and *Positive.* It also should be noted that the strong majority of all sentiments were *Negative* for both handsets. The raw sentiment counts of Figure 2 further illustrate the overall negative sentiment and the minor sentiment differences between handsets.

During our model generation, we used a different set of data where the sentiments were manually determined by reading each entry. To see if this data showed similar equality, the manual and modeled results were combined and normalized, as seen in Figure 3. While the overall sentiment differences between the models remain minor, the manual data set showed higher positive sentiments overall. To validate this oddity, the sentiment was remodeled with a different algorithm, and a similar trend occurred within the modeled data set. Therefore, we believe this sentiment shift between the manual and modeled data sets to be genuine.

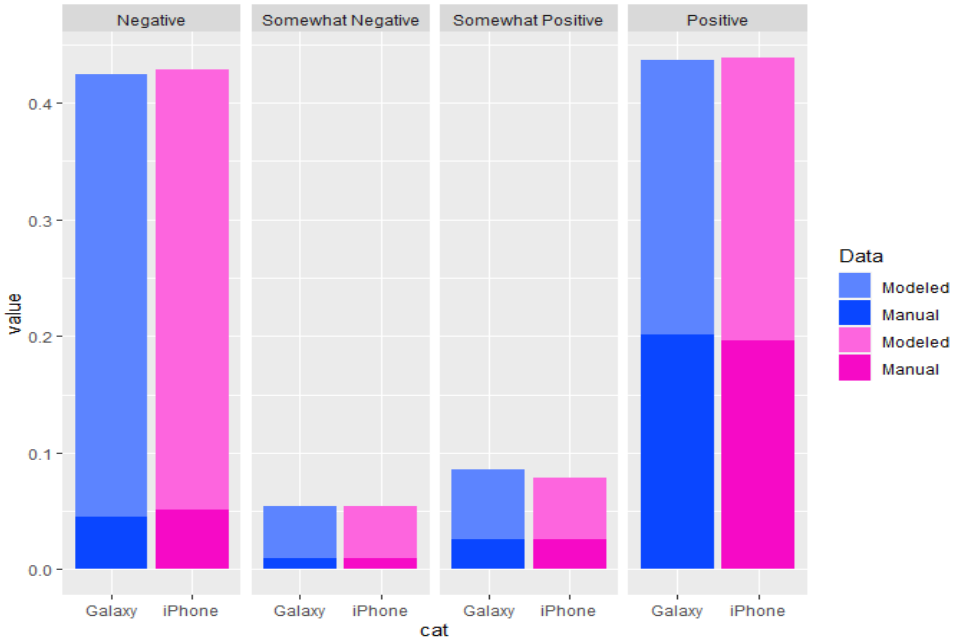

Figure 3: Combined sentiments for the manual and modeled data sets

Confidence

We are able to have high confidence in the accuracy of our models in these predictions. Our models demonstrated above 84% accuracy for identifying the proper webpage sentiment for each phone model. However, given the very minor differences in sentiment values in Figure 2, we believe these differences to be within the noise of our modelling and data collection. It is very possible that a subsequent web crawl and modelling could result in the Galaxy having a slightly higher *Positive* sentiment rating. As such, per this analysis, the models should be assumed to have effectively equal sentiments at the time of this writing.

Implications

While both of these name-brand devices would be functionally acceptable for implementation and have broad adoption in the commercial environment, the overall *Negative* sentiment is off-putting. Helio's app could potentially see more negative feedback simply because it is hosted on an undesirable device, and the user frustrations associated with it. It could be worthwhile for Helio and Alert! Analytics to extend this sentiment analysis to the other phones originally under consideration to see if a more positive device sentiment can be found.

Methodologies

Using the Common Crawl framework through Amazon Web Services, thousands of web pages were scanned looking for keywords associated with phone types, features (camera, display, performance, etc), and sentiment.

To understand how these identified keywords are associated with overall phone sentiment, over 12,000 web pages (per phone) were read by our team and assigned a sentiment score. This scoring was used to create predictive models to apply to a larger data set of over 32,000 web pages.

Alert! Analytics developed multiple machine learning models to predict the sentiment on the larger set of web articles. In building the most predictive model, we explored multiple data science models and multiple ways to organize the collected data.

Through model selection, feature selection, and feature engineering we were able to achieve prediction accuracies of 84% for iPhone sentiment and 84% for Galaxy sentiment. We also evaluated our models on their ability to predict simply overall positive or overall negative sentiment. This was performed because, for example, it's not as impactful for an article sentiment to be wrongly predicted to be "Somewhat positive" when it is actually "Positive", versus when it is wrongly predicted to be "Negative". In predicting the correct sentiment tone, our model accuracy increases to 87.6% for the iPhones and 89.2% for the Galaxy. With these high accuracies, we can have reasonable confidence in the ability of these models to classify the sentiment of these web pages.