

Online Smart Phone Sentiment Analysis

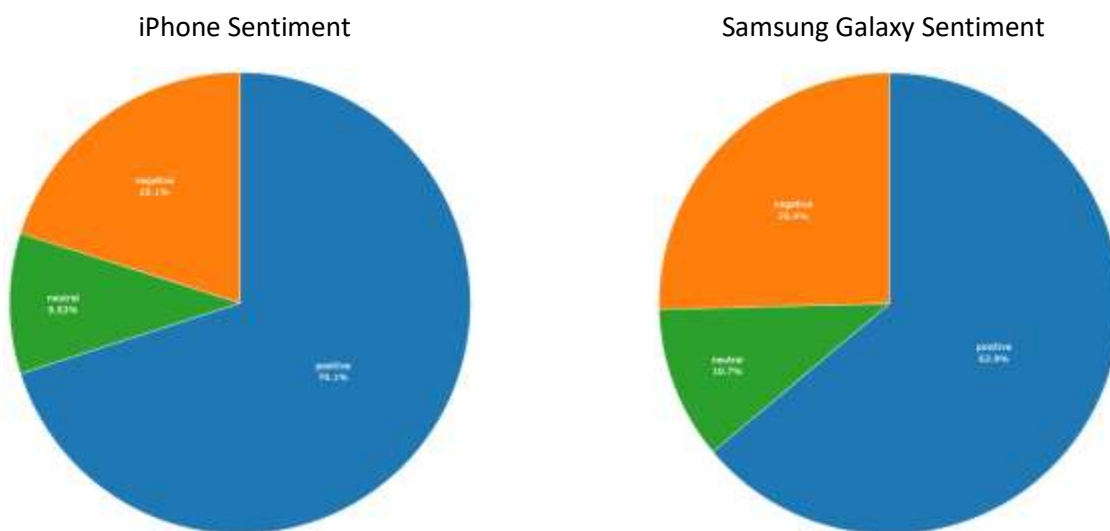
A Comparison of User Opinions of iPhone and Samsung Galaxy
for Helio Electronics

Compiled by
Alert! Analytics

Project Overview

Helio, a smart phone and tablet app developer, is working with a government health agency to create a suite of smart phone medical apps for use by aid workers in developing countries. This suite of apps will enable the aid workers to manage local health conditions by facilitating communication with medical professionals located elsewhere. The government agency requires that the app suite be bundled with one model of smart phone. Our task is to analyze online customer sentiment toward particular brands of smart phones, namely iPhone and Samsung Galaxy, to determine if one model is a better option to bundle the app suite with.

Findings



	Sentiment Unclear	Negative Sentiment	Neutral Sentiment	Positive Sentiment
iPhone	16385	1201	587	4183
Galaxy	16358	1524	642	3832

Figure 1: Comparison of iPhone and Galaxy Sentiment Scores

In analyzing over 22,000 websites scraped from the internet, we found that of those with a clear sentiment expressed regarding one or both of the phone models of interest, there was a measurable preference for iPhone over Samsung Galaxy. Not only was there a higher percentage of positive sentiment expressed for the iPhone, there was also a lower percentage of negative sentiment. Neutral sentiment for both phones was more or less comparable.

Methodology

Data was collected from the Common Crawl repository of web data stored on Amazon's Data Sets. We scanned for documents that expressed meaningful sentiment about five potential phone partners and retained only those with mentions of at least one of the phones or a relevant OS term and at least one of the following words: review; critique; looks at; in depth; analysis; evaluate; evaluation; assess. For each document, we collected information about the sentiment toward specific phone features. A set of words that indicated positive, neutral, and negative sentiments was compiled and used to create a Python mapper script that could comb through these websites to tally unique instances of these relevant words within a set distance from a mention of a phone. Initially, roughly 12,000 sites mentioning one or more of the phones were collected and sentiment word counts were tallied and stored in a data matrix.

After the decision was narrowed down to two options, iPhone or Galaxy, our analytics team manually assessed each collected site and labeled them with a sentiment rating, 0 through 5, for each phone with the following meanings: 0 = sentiment unclear; 1 = very negative; 2 = somewhat negative; 3 = neutral; 4 = somewhat positive; 5 = very positive. These data matrices were then used to train models that could predict user opinions based on the occurrence of the sentiment words.

Upon finalization of the models, they were applied to an unseen data set of over 22,000 sites collected using the same web scraping scripts as were used to collect the training set.

Confidence

Initially, four machine learning algorithms were used to construct out-of-the-box models to apply to the data sets. Performance was assessed based on accuracy and Kappa scores (a classification performance metric used to account for randomness). For both devices, Random Forest was the most reliable.

Model	Accuracy (Train)	Kappa (Train)	Accuracy (Test)	Kappa (Test)
C5.0	0.7700645	0.5523425	0.7753213	0.5619559
Random Forest	0.7733677	0.5597573	0.7760925	0.5654067
K-Nearest Neighbors	0.3648004	0.1793220	0.3609254	0.1817561
Support Vector Machines	0.7084131	0.4062715	0.7151671	0.4162251

Table 1: Comparison of Modeling Algorithm Performance (iPhone)

Model	Accuracy (Train)	Kappa (Train)	Accuracy (Test)	Kappa (Test)
C5.0	0.7651510	0.5281260	0.7659520	0.5303090
Random Forest	0.7679725	0.5341508	0.7698269	0.5375984
K-Nearest Neighbors	0.7564675	0.5145780	0.7563937	0.51410720
Support Vector Machines	0.7120026	0.3929717	0.7036941	0.3772725

Table 2: Comparison of Modeling Algorithm Performance (Galaxy)

Moving forward with Random Forest as our base model, feature selection and engineering was employed to tune the model and improve predictive ability. The three main issues encountered were noise from unnecessary features (irrelevant sentiment), ineffectiveness due to too many prediction classes (sentiment ratings), and a tendency to overpredict positive sentiment due to an overwhelming presence of positive scores in the training data.

Unnecessary features were identified and removed through Recursive Feature Elimination (RFE). For the iPhone model, this reduced the feature set from 58 to 25 and for the Galaxy model from 58 to 26. RFE alone had little effect on the performance of the models, but recoding the target variable (iPhone/Galaxy Sentiment Ratings) increased both accuracy and Kappa score roughly 7 percentage points. The number of classes was reduced from 6 to 4, where: 0 = sentiment unclear; 1 = negative; 2 = neutral; and 3 = positive.

While this was able to produce a model with nearly 85% accuracy, its flaw was that it over-predicted positive sentiment. This is caused by an imbalanced training set, in which the vast majority of reviews were positive. Due to this tendency, though the models accurately predicted 97-99% of positive reviews, they correctly identified only about 6% of negative reviews, 56-65% of neutral reviews, and 64-68% of unclear sentiments. As a result, though the overall accuracy looks impressive, it is not an effective model for predicting the less common classes.

To counter this, we employed a technique known as oversampling of the data, by which we padded the data with duplicate samples of the smaller classes to give the algorithm a balanced data set to train on. While we saw a significant drop in the overall accuracy score of the model, the ability to correctly predict anything other than positive sentiment was dramatically improved. There was still a tendency to over-predict positive, though it was not nearly as pronounced with oversampling. Ultimately the sacrifice of overall accuracy is justified by the increase in prediction accuracy of the individual classes.

Implications

Based on the findings from our analysis, both the iPhone and Samsung Galaxy seem to enjoy an overwhelmingly positive assessment from the general public and, with customer satisfaction as the primary metric for the decision, either would be a worthwhile choice for Helio to partner with. However, the data suggests that the iPhone is the more satisfactory device of the two by a significant enough margin that, barring other determining factors not considered here, it would be our recommendation to proceed with a partnership with iPhone for Helio's medical app suite.