Develop Models to Predicts Sentiment Towards Smart Phones

Client: Helio

From: Alert Analytics

By: Usman Iftikhar

# Overview

Alert Analytics’ client Helio, a smart phone and tablet app developer, has asked Alert Analytics to evaluate sentiment towards mobile phones. Helio is working with a government health agency to create a suite of smart phone apps for use by aid workers in developing countries. The apps will assist the aid workers to manage local health conditions by facilitating communication with medical professionals located elsewhere. For example, one of the apps enables specialists in communicable disease to diagnose conditions by examining images and other patient data uploaded by local aid workers. The government agency requires that the app suite be bundled with one model of smart phone. After initial investigation, Helio has created a short list of five devices that are all capable of executing the app suite’s function. The list of devices includes iPhone, Samsung Galaxy, Nokia Lumia, Sony Xperia, and HTC. Helio has tasked Alert Analytics to examine the prevalence of positive and negative sentiment towards these devices on the web. Alert Analytics is tasked with providing Helio an analysis of sentiment towards the target devices.

# Findings

Helio narrowed the list of smart phones down to, Apple and Samsung after initial discussion with the manufacturers. Therefore, the analysis focuses on sentiment towards Samsung Galaxy and Apple iPhone devices over other smart phones. In order to train models to predict sentiment towards these devices found on the web, a smaller set of data has been collected for both Galaxy and iPhone. The smaller data sets were manually examined and labeled by the Alert! Team to effectively train algorithms. The best model generated from the training data set is then applied to the larger data set collected from the web to predict the sentiment towards iPhone and Galaxy. The predicted sentiment is shown in Figure 1. The analysis shows the sentiment towards both devices is negative, and results are analogous. Negative sentiment towards both devices is more than double the positive sentiment.

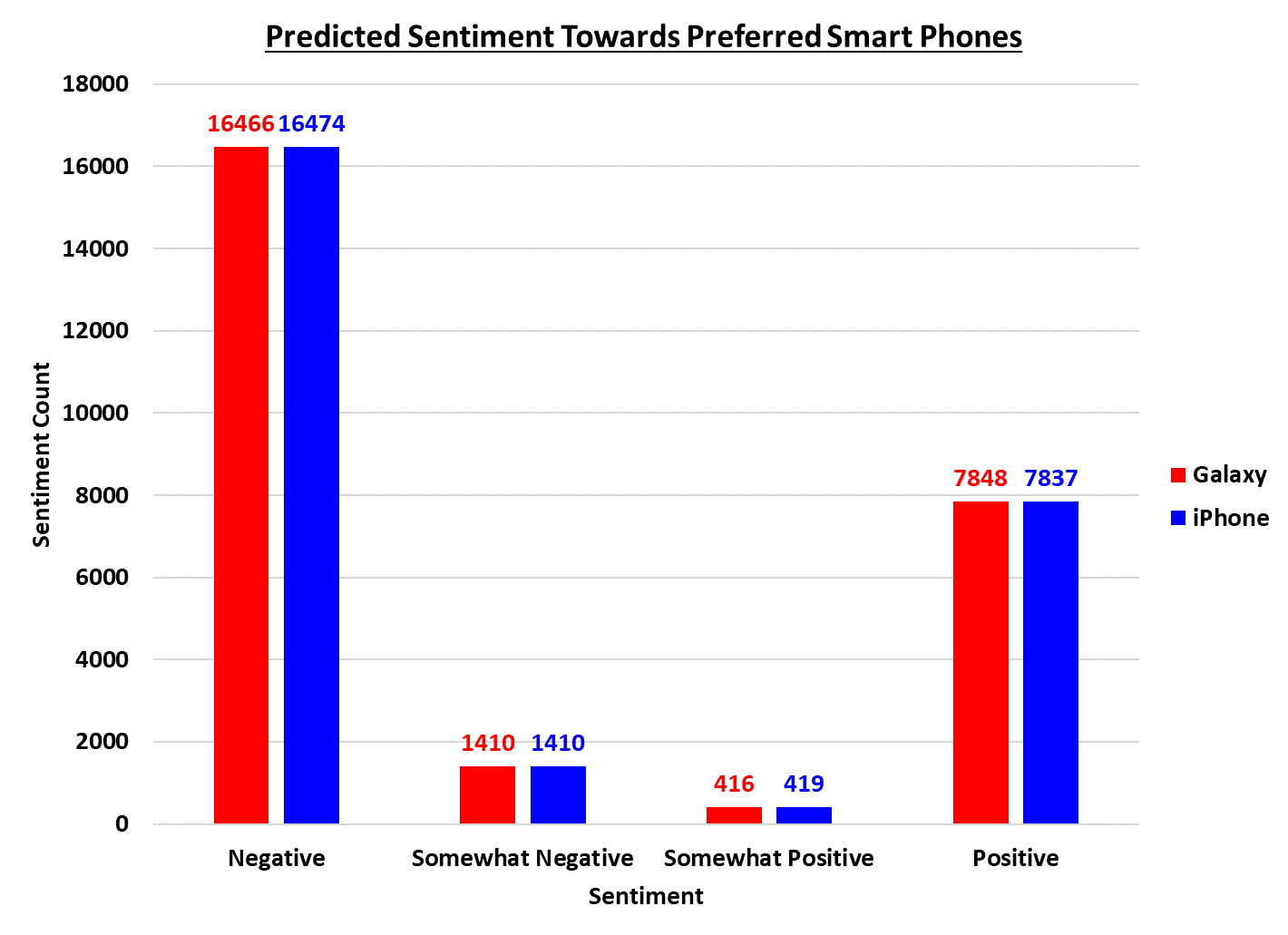


Figure . Sentiment towards the Samsung Galaxy and Apple iPhone from the provided data sets.

# Confidence

To measure the performance and accuracy of the best model, 30% of the smaller data sets (test/validation sets) were never exposed to the training algorithms. This was done to mimic this part of the data set to be the ‘real world’ data set. Each model generated from the algorithm was then used to predict the sentiment in the test data set. The predicted sentiment was then compared against the already known sentiment in the test data set, known as the ground truth. How often the predicted value matched the ground truth is the accuracy of the model. If the trends exhibited the data remain the same going forward, then the confidence in predictions is 85% with the 95% confidence interval ranging from 83.9% to 86.2%.

# Implications

Based on the above stated results, Helio should continue to consider other smart phones as the sentiment towards both iPhone and Galaxy is negative. Furthermore, data set should be expanded to include other attributes such as the price of smart phones. Since the smart phone devices will be used by medical professionals, perhaps sentiment from only the medical professionals may help. Data collection could also be done for other similar smart phone apps currently being used along with the sentiment of those apps on the desired smart phone devices.

# Methodology

In order to train models to predict sentiment towards these devices found on the web, a smaller set of data has been collected for both Galaxy and iPhone that is manually labeled with the sentiment. A larger data set is collected from the web that is used to predict the sentiment based on the trained data sets. In each case, data is first imported into R for exploratory analysis to gain an understanding of the attributes and determine if any of the attributes need to be removed to help the model make more accurate predictions. The prepared data set is then used to generate models. Models are then compared and selected based on highest accuracy. The selected model is used to predict the sentiment for the data collected from the web. Results are interpreted considering the business objective.