Alert Analytics

Shape

Description automatically generated with low confidenceHelio

Sentiment of smartphone devices

Icon

Description automatically generated

Contents

[Overview 3](#_Toc113306281)

[Findings 3](#_Toc113306282)

[Confidence 4](#_Toc113306283)

[Implications 4](#_Toc113306284)

[Methodology 5](#_Toc113306285)

# Overview

Helio is a smartphone and tablet app developer that currently works with a government health agency to create a suite of smartphone 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 smartphone. Helio asked Alert Analytics to narrow the list of devices to one by examining the prevalence of positive and negative sentiments towards the devices. In this case, we analyze the Galaxy and iPhone smartphones.

# Findings

The following sentiment categories were used:

0 <- unclear

1 <- negative

2 <- somewhat negative

3 <- somewhat positive

4 <- positive

There were 22,596 observations. Most of the sentiments towards both devices are unclear. Galaxy has a 20.1% positive sentiment and iPhone has a 19.4% positive sentiment. Therefore, the Galaxy smartphone has the most positive sentiment.

Chart

Description automatically generatedChart

Description automatically generated

Chart

Description automatically generatedChart

Description automatically generated

However, if we exclude the unclear sentiment from both devices then the results are different, and the observations are reduced to 5,946 and 6,307 for iPhone and Galaxy respectively. Galaxy has a 72% positive sentiment and iPhone has a 73.7% sentiment. Therefore, the iPhone smartphone has the most positive sentiment percentagewise. Nonetheless, Galaxy still has more positive sentiment counts than iPhone. However, the differences are minimal.

Chart

Description automatically generatedChart, waterfall chart

Description automatically generated

# Confidence

The accuracy of the modeling results of iPhone and Galaxy is 85% and 84% respectively. The kappa of the modeling results of iPhone and Galaxy is 0.62 and 0.59 respectively. Therefore, the modeling results are accurate around 85 times out of 100 and there is at least a moderate agreement between raters. The attributes selected don’t capture the whole sentiment with just counting how many times a certain word appears in an article, but it could improve if some adjectives related to a positive or negative attitude were included. For example, better, improved and lacking.

# Implications

Overall, the suite of apps could be done for either model of smartphone because the differences are minimal, and they could disappear in a larger or different sample of data.

# Methodology

Parallel processing was setup to speed up the process of analysis. Then, the small matrices were explored to better understand the data. After the exploratory data analysis, the small matrices were put under three different feature selection methods. The correlation, near zero variance, and recursive feature elimination methods were used to create three different matrices on each device. A total of 6 matrices. Then, each matrix was split 70% for training and 30% for testing.

The response variable of galaxy sentiment and iPhone sentiment were converted to factors as part of the pre-process. After the pre-process was ready, a total of 36 models were trained. The trained models were C5.0, Random Forest, Support Vector Machine with Linear Kernel, SVM with Polynomial Kernel, SVM with Radial Basis Function Kernel, and KKNN. Each model was run with the three different matrices that resulted from the feature selection. The models were repeated for each of the two smartphones. In other words, 2 smartphones, 3 feature selection matrices, and 6 models.

After the 36 models finished running, the results of each model in terms of accuracy and kappa were compared. For both devices, the combination of the RFE and the C5.0 model brought the best results. However, the accuracy of the models was at least 76%, so feature engineering was done. Changing the labeling of the response variable from 0,1,2,3,4,5 to 0,1,2,3,4 increased the accuracy to at least 84%. PCA was used separately but the accuracy almost didn’t rise.

Now that the model was improved with feature engineering, the predictions of the sentiments of each smartphone were made on the large matrix. Finally, the results were analyzed, and some visualizations were made to communicate the results.