**Smart Phone Sentiment Analysis**

**Helio**

Alert! Analytics

Overview

Helio has partnered with a government health agency to create medical apps for smart phones. These apps are designed to help aid workers manage health conditions in developing countries through communication with medical professionals elsewhere. The government agency requires that the app suite be bundled with only one smart phone. The goal of this analysis is to determine which smart phone they should use based off of sentiment toward the target devices.

The data was found by counting words associated with sentiment toward the smart phones within relevant documents on the web. Amazon Web Services was used to conduct the analysis using data sets from Common Crawl.

Findings

The Sentiment categories were:

0: Sentiment Unclear

1: Very negative

2: Somewhat negative

3: Neutral

4: Somewhat positive

5: Very positive

The final results were very close. iPhone and the Samsung Galaxy both had 38 very positive votes. Galaxy had 6 neutral and 55 unclear while iPhone had 8 neutral and 53 unclear.



Confidence

This chart shows the accuracy of the Random Forest models for iPhone and Samsung.

Chart

Description automatically generated

Implications

Based on the data predictions, the client couldn’t go wrong with either device. Going off of the raw data, Samsung Galaxy has a slightly higher very positive sentiment count so that is the suggested phone to use.

Methodology

Background pattern

Description automatically generatedBackground pattern

Description automatically generatedCorrelations were explored first and had similar results, shown below:

Text, letter

Description automatically generated with medium confidenceNext, feature variance was explored. There were many variables that showed near zero variance.

These were the variables from both the iPhone and Galaxy matrices that showed near zero variance and were ultimately removed for modeling. In order for those models to work, the sentiment variable needed to be changed to factor so iphonesentiment and galaxysentiment were processed as factors.

The caret package was used to help run the prediction models. For this project, the data was split into 70% training and 30% testing. Four classification models were used to test: C5, Random Forest, SVM, and kknn. The random forest model resulted in the highest accuracy for both Galaxy and iPhone.

After deciding on the Random Forest model, predictions were run against the large data matrix.