# SENTIMENT PREDICTIONS FOR APPLE IPHONE AND SAMSUNG GALAXY

October 15, 2020

Michael Ortiz
VP
Alert! Analytics

Re: Sentiment Predictions for Apple Iphone and Samsung Galaxy – Final Results

Dear Mr. Ortiz,

**Project background**

Alert!Analytics' client, Helio, is working with a government health agency to create a suite of smart phone medical apps for use by aid workers in developing countries. The government agency requires that the app suite be bundled with one model of smart phones. Helio is in the process of evaluating potential handset models to determine which one to bundle their software with. Helio has asked Alert!Analytics to assist with narrowing their list down to one device and examining the prevalence of positive and negative attitudes toward Apple Iphone and Samsung Galaxy on the web.

The project goals for Alert!Analytics were set to following:

1. Collect web-based information necessary for determining the overall sentiment toward Apple Iphone and Samsung Galaxy
2. Build appropriate predictive models
3. Generate sentiment predictions by category for Iphone and Samsung Galaxy based on the web-based information

**Data characteristics**

The data attributes mainly consist of counts of positive, negative, and unclear mentions regarding the operating system, camera, display, and the performance for Iphone, Samsung Galaxy, Sony Xperia, Nokia Lumina, and HTC Phone.

Training and testing of predictive models were performed on the data sets with manually determined overall sentiments toward Iphone and Samsung Galaxy.

The overall sentiment includes 6 categories: "0" – very negative, "1" – negative, "2" – somewhat negative, "3" – somewhat positive, "4" – positive, and "5" – very positive.

# SENTIMENT PREDICTIONS FOR APPLE IPHONE AND SAMSUNG GALAXY

Amazon Web Services (AWS), specifically Elastic Compute Cloud (EC2), Elastic MapReduce (EMR), and Simple Storage Service (S3), have been used to compile more than 20,000 records for the above mentioned attributes from Common Crawl, the open repository of web crawl data, for the month of August 2020. The interesting observation is that web pages dated closer to the end of August had twice as many mentions regarding our topic.

**Pre-processing**

Observations for 6 sentiment categories were condensed to 4 groups: "1" – negative, "2" – somewhat negative, "3" – somewhat positive, "4" – positive. To improve algorithms' Accuracy and Kappa (the possibility of the agreement occurring by chance) scores, the train/test sets for both Iphone and Samsung Galaxy have been balanced, and the automatic feature selection has been applied. Also, the selected feature values have been standardized.

**Model evaluation**

During testing the selected models showed the following performance:

| MODELS | PERFORMANCE METRICS | | | |
|---|---|---|---|---|
| | IPHONE | | SAMSUNG GALAXY | |
| | ACCURACY | KAPPA | ACCURACY | KAPPA |
| K- Nearest Neighbor | 0.7787 | 0.6099 | 0.8153 | 0.5923 |
| Support Vector Machine | 0.7485 | 0.5524 | 0.7436 | 0.3927 |
| Classification and Regression Tree | 0.7858 | 0.6221 | 0.8126 | 0.5856 |
| Random Forest | 0.8575 | 0.6956 | 0.8299 | 0.6264 |

Random Forest produced higher Accuracy score, but, most importantly, higher Kappa score for both Apple Iphone and Samsung Galaxy.

# SENTIMENT PREDICTIONS FOR APPLE IPHONE AND SAMSUNG GALAXY

**Final predictions**

Figure 1.1. demonstrates final sentiment predictions using Random Forest algorithm based on the sample data collected from different websites for August 2020.

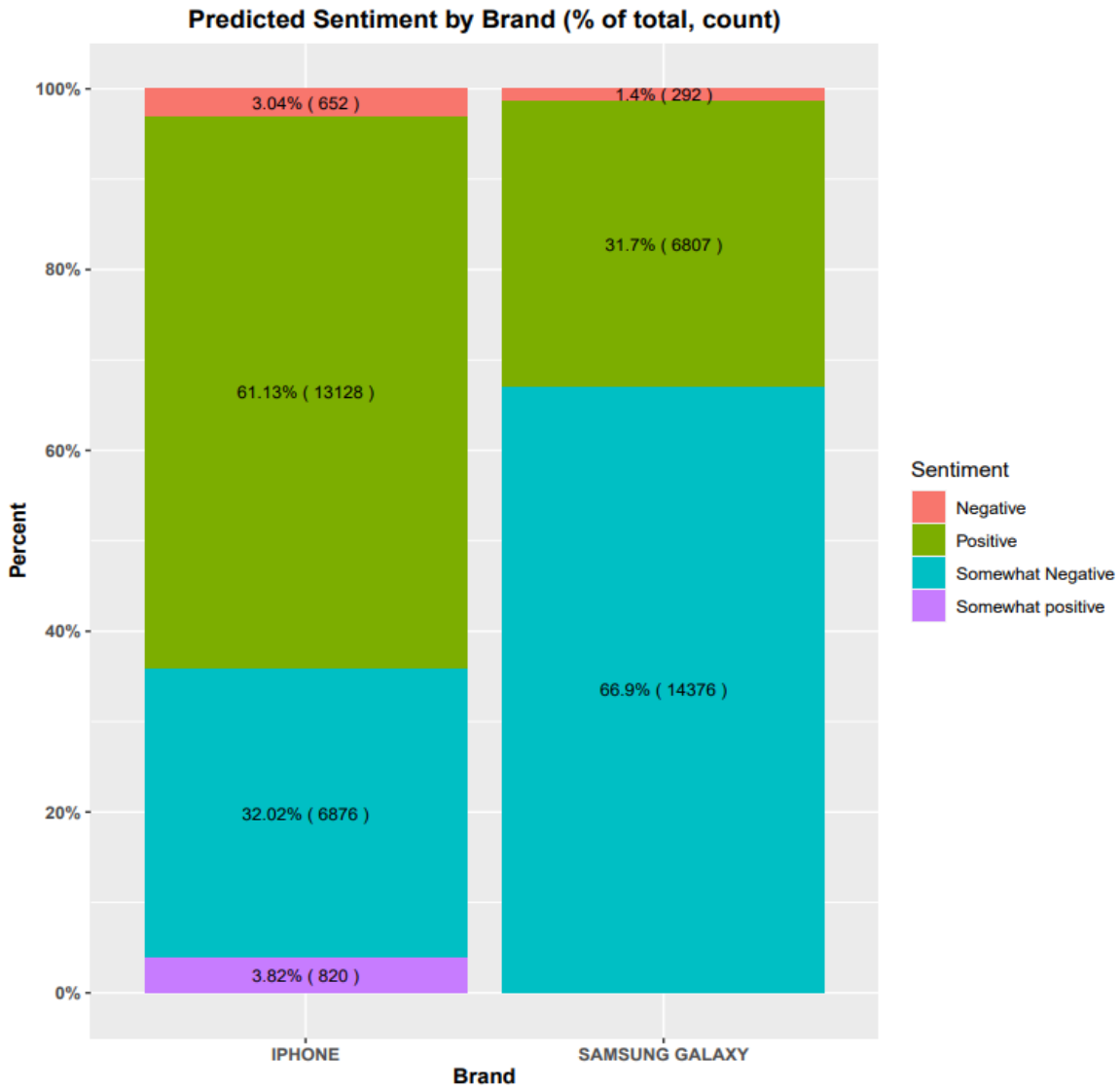


**Figure 1.1.**

It appears that Iphone is more preferred than Samsung Galaxy. A caution should be exercised while making final decisions based on these predictions. The sample used to make predictions represents the small fraction of information available online. Samples from multiple time periods/seasons (Christmas, Black Friday, new release events, etc.) are necessary to see the overall trend. While both Apple and

# SENTIMENT PREDICTIONS FOR APPLE IPHONE AND SAMSUNG GALAXY

Samsung have relatively stable pools of customers, consumer preferences might be impacted by new releases. Model specific comments for each handset (Iphone 12, Samsung Galaxy S20/S21, etc.) need to be incorporated.

**Summary**

While Helio can obtain the overall sense of customer preferences towards Iphone and Samsung Galaxy based on our predictions, the following is recommended:

1. Generate multiple samples from different time frames, new release events, holiday sales, etc. Predictions based on those samples will help us to confirm the trend.
2. Expand the search criteria/vocabulary in our algorithms to catch the handset model specific comments as well as collect more observations to build models with higher accuracy.
3. Research country specific consumer preferences as the app suite will be used in phones of aid workers in developing countries.

Please let me know if you have further questions.