|  |
| --- |
| Mobile device with apps |
| **Helio Smart Phone Sentiment Analysis**  Presented by Alert! Analytics |
| |  |  |  | | --- | --- | --- | |  | Jelly Edwards |  | |

**Overview**

Helio is working with a government health agency to create a suite of smart phone medical apps to be used by aid workers. This government agency requires that the app suite be bundled with one model of smart phone. Helio has reached out to Alert! Analytics to assist with pinpointing which smart phone device would best fit their needs. Alert! was requested to evaluate the prevalence of positive and negative sentiment toward these devices on the web. This report is a summary of the findings that Alert! collected during their analysis for the iPhone and Galaxy devices.

**Sentiment Analysis Results**

Alert! was requested to evaluate websites on the overall sentiment toward iPhone and Galaxy devices on the web. We placed these sentiments into six categories: Sentiment Unclear, Very Negative, Somewhat Negative, Neutral, Somewhat Positive, and Very Positive. After running our machine learning models and calculating the predictions, we recorded our sentiment results and created figures that display these results. From our results, it is clear that a large percentage of sites do not have a clear sentiment towards either device. This is especially clear in *Figure 1*, where the Sentiment Unclear category dominates the bar chart.

Chart, bar chart

Description automatically generated

***Figure 1.*** *Number of Predicted Ratings for iPhone and Galaxy*

To further explore this data, we decided to only examine the findings of the other sentiment categories and compare the results between iPhone and Galaxy. We created a set of pie charts to display the percentage of each sentiment in *Figure 2*. With this, we can tell that both have very similar sentiments. Most users view the devices in a positive light, with the results skewing mostly to Very Positive. The sentiment between the two varies very little – Galaxy surpasses iPhone by a minute 3.09 % in the Very Positive category.

*Chart, pie chart

Description automatically generated*

***Figure 2.*** *Percentage of Sentiment without Unclear Data Points*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Device | Sentiment Unclear | Very Negative | Somewhat Negative | Neutral | Somewhat  Positive | Very Positive |
| *iPhone* | 20911 | 0 | 1236 | 861 | 1 | 5349 |
| *Galaxy* | 20706 | 3 | 1198 | 872 | 0 | 5579 |

***Figure 3.*** *Raw Predicted Sentiment Findings for iPhone and Galaxy*

**Sentiment Analysis Prediction Confidence**

Alert’s analysis required the used of Machine Learning Models and Prediction algorithms. While we did experiment with several algorithms, the primary algorithm we settled on was RandomForest (RF). To examine the performance of this model, we checked the Accuracy and Kappa metrics. Accuracy is a metric that shows the percentage of correctly classified cases out of all possible instances. Kappa is a metric that compares predicted values accuracy with the expected values accuracy, and it informs the user of the overall performance of the model. The higher the Accuracy and Kappa scores, the better the model.

In *Figure 4,* it is clear that Accuracy is relatively high. While a high accuracy is a positive trait to have, we need to consider the models Kappa as well. The closer Kappa is to 1, the better the classification model’s performance. The Kappa scores for the models are substantial and above 0.50. While this means that model itself performed pretty well, it also wasn’t absolutely perfect. I think it can be improved by collecting more preliminary iPhone and Galaxy sentiment data to train the model with. Increasing the training data will give the model more examples to learn from, improving its likelihood to identify correctly.

|  |  |  |
| --- | --- | --- |
| Device | RF Accuracy | RF Kappa |
| iPhone | 0.81 | 0.65 |
| Galaxy | 0.80 | 0.60 |

***Figure 4.*** *Performance Metrics for RandomForest Model*

**Sentiment Analysis Implications**

From our prediction models, we can see that the sentiment to iPhone and Galaxy are quite similar to each other. Neither device has a strong preference over the other. Due to the data showing no preference, Helio will have to select their device based on other outside factors. One factor I would like to focus on is the mobile operating system available for each device. iPhone uses Apple’s IOS operating system, while Galaxy uses Google’s Android. IOS is an OS only available for Apple devices, limiting what devices can be used if Helio chooses this device for their application. Andriod, however, is an OS that is available to more brands and devices other than just Galaxy. If circumstances arise in the future that requires Helio to switch devices, it may be best to choose Galaxy due to the flexibility of moving their application to another Android device. In the end, it is up to Helio to choose what device they think is best for them based on our analysis and other outside factors.

**Methodology**

For our analysis, we received our data from Common Crawl - a non-profit organization that crawls and archives the entire readable internet once per month. It stores these archived files in Amazon Web Services (AWS). We used Python scripts to analyze these archived files and document if the page was review for that device. It would then document what was the overall sentiment of that device. After analyzing and collecting the appropriate data from AWS, we used R and RStudio to evaluate the data further. We used our machine learning algorithms on the collected sample data to train the models. Once trained, we applied these models to ~20,000 recorded instances that were pulled from Common Crawl/AWS. This predicted the sentiment for each device.