**Cellphone Predictive Sentiment**

The goal of this project was to gain better insights into cellphone customers’ perceptions of specific IPhone and Galaxy cell phone components such as: operating system, hardware(performance), camera, and display. By capturing the customer’s sentiments through a sample of websites, this report provides key observations regarding the overall customer sentiments toward the IPhone and Galaxy, specific areas of promise/concern within these phones, and the development of a predictive framework to address future customer needs.

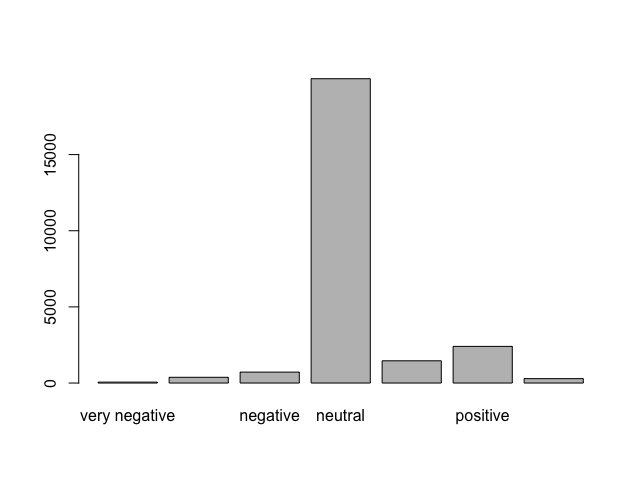
**Data Insights**

**Overall Sentiment**

The IPhone and Galaxy datasets both have 25,292 observations with 60 different variables. While there are several attributes to focus (which this reports highlights after this section), the main attribute of importance consists of looking at the overall sentiment of the specific phones. The IPhone overall sentiment variables ranges from – 794 (very negative) to 782 (very positive) with an average rating of 1.472. Here is the breakdown of the IPhone sentiment by customer perceptions:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| IPhone Sentiment | | | | | | |
| Very Negative | Somewhat Negative | Negative | Neutral | Somewhat Positive | Positive | Very Positive |
| 60 | 377 | 718 | 19982 | 1460 | 2406 | 289 |

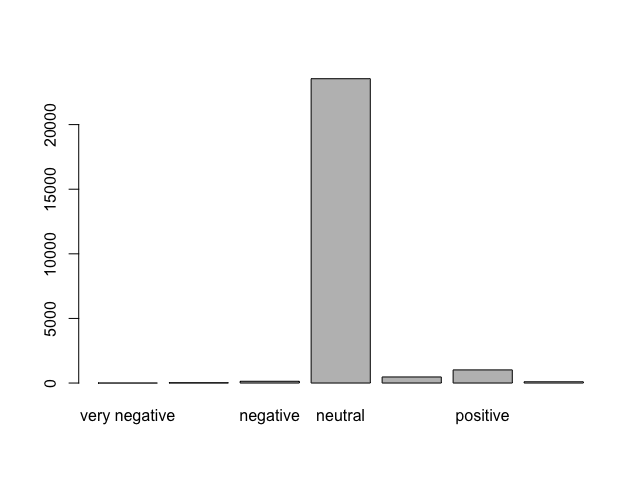
The highlighted areas indicate the largest distribution of IPhone sentiments numbers and lean toward a more overall positive sentiment of the IPhone based upon all categories. However, like the Galaxy results as well, one key limitation is the significant neutral observations within the results. The following chart illustrates the prevalence of neutral sentiments within the IPhone analysis:



The Galaxy overall sentiment variables range from -107 (very negative) to 407 (very positive) with an average rating of 1.472. Similar to the IPhone data, the Galaxy results illustrate a large percentage of the overall Galaxy sentiment in the neutral and Positive classification.

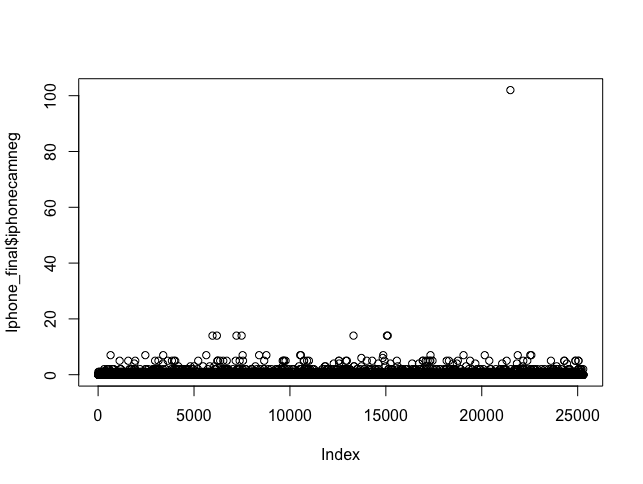
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Galaxy Sentiment | | | | | | |
| Very Negative | Somewhat Negative | Negative | Neutral | Somewhat Positive | Positive | Very Positive |
| 2 | 20 | 137 | 23551 | 470 | 1016 | 96 |

In addition, the chart below highlights the significant neutrals sentiments in the analysis:



**Key Cellphone Components**

The attributes that illustrated the most significant range in perceptions included: IPhone Negative Camera Review (0-102), IPhone Positive Display (0-87), and IPhone Negative Performance (0-76). In addition, the smallest range of IPhone sentiments included HTC Positive Camera (0-9), HTC Negative Camera (0-8), and Sony Negative Performance (0-5). This extreme range can be the result of outlies as illustrated in the IPhone Negative Camera Review in the graph below:



**Model Performance & Confidence**

The first step of the analysis was to look at the attribute relationships to the Overall Sentiment and their relationship to each other. Based upon these relationships, several attributes were removed due to their correlation with the attribute, Overall Sentiment, and could distract from developing an accurate predictive framework. With 22 attributes removed from the IPhone dataset, the remaining 38 attributes for the iPhone analysis consisted of the following:

|  |
| --- |
| IPhone Final Attributes |
| “id" "iphone" "samsunggalaxy" "sonyxperia" "htcphone" "ios" "googleandroid" "iphonecampos" "samsungcampos" "htccampos" "iphonecamneg" "htccamneg" "iphonecamunc" "htccamunc" "iphonedispos" "samsungdispos" "sonydispos" "htcdispos" "iphonedisneg" "samsungdisneg"  "sonydisneg" "iphonedisunc" "htcdisunc" "iphoneperpos" "sonyperpos" "htcperpos" "iphoneperneg" "sonyperneg" "iphoneperunc" "samsungperunc" "sonyperunc" "nokiaperunc" "htcperunc" "iosperpos" "iosperneg" "iosperunc" "googleperunc" "iphoneSentiment" |

Based upon these 38 attributes, different models were implemented to determine accuracy so that Alert Analytics may address specific concepts of each phone. The following models were utilized, and it was determined to use Support Vector Machine (SVM) model because

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RMSE | R Squared | Accuracy | Kappa |
| KNN | 8.412233 | 0.8159849 | 0.9212850 | 0.9212850 |
| Random Forrest | 4.319590 | 0.9460792 | 0.9610944 | 0.8913057 |
| SVM | 4.452376 | 0.96672787 | 0.9960549 | 0.9960549 |
| C50 |  |  | 1.0 | 1.0 |

Although the C50 showed a perfect fit for the IPhone dataset and for the Galaxy dataset, it is suspected that there was over fit with this model and needs further analysis to determine parameters. Therefore, given that most models showed a high level of confidence in predicting overall sentiment, the selection of the SVM model provides strong confidence of predicting overall sentiment as all models applied to the data showed similar sentiment results.

In terms of the Galaxy data, a similar process was conducted as the Galaxy data as to analyze the relationships between the various attributes. Based upon these relationships, several attributes were removed due to their correlation with the attribute, Overall Sentiment, and could distract from developing an accurate predictive framework. With 23 attributes removed from the Galaxy dataset, the remaining 37 attributes for the Galaxy analysis consisted of the following

|  |
| --- |
| Galaxy Final Attributes |
| "id" "iphone" "samsunggalaxy" "sonyxperia" "htcphone"  "ios" "googleandroid" "iphonecampos" "samsungcampos" "htccampos" "iphonecamneg" "htccamneg" "iphonecamunc" "htccamunc" "iphonedispos" "sonydispos" "htcdispos" "iphonedisneg" "sonydisneg" "iphonedisunc" "samsungdisunc" "htcdisunc" "iphoneperpos" "sonyperpos" "htcperpos" "iphoneperneg" "sonyperneg" "iphoneperunc" "samsungperunc" "sonyperunc" "nokiaperunc" "htcperunc" "iosperpos" "iosperneg" "iosperunc" "googleperunc" "galaxySentiment" |

Based upon these 38 attributes, different models were implemented to determine accuracy so that Alert Analytics may address specific concepts of each phone. The following models were utilized, and it was determined to use Random Forrest (RF) model because of the highest Accuracy result.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RMSE | R Squared | Accuracy | Kappa |
| KNN (5) | 5.518470 | 0.6946736 | 0.9807988 | 0.8450510 |
| SVM | 4.851685 | 0.9998349 | 0.9811379 | 0.84779 |
| Random Forrest | 4.942965 | 0.7573926 | 0.9846908 | 0.9405622 |
| C50 |  |  | 1.0 | 1.0 |

**Insights & Recommendations**

There are several key insights that can be drawn from the predictive sentiment analysis that can benefit Alert Analytics:

* There is a general positive overall sentiment to both phones and it could be beneficial for future analysis to focus on each specific component of the cameras and research sentiments in more detail as there was a significant portion of neutral statements.
* Given the large number of “Neutral” sentiments it could also be very beneficial to further investigate these categories to determine how to more adequately address the search parameters for sentiment analysis.
* Outliers – There were several outliers to further investigate the language of the sentiment analysis.