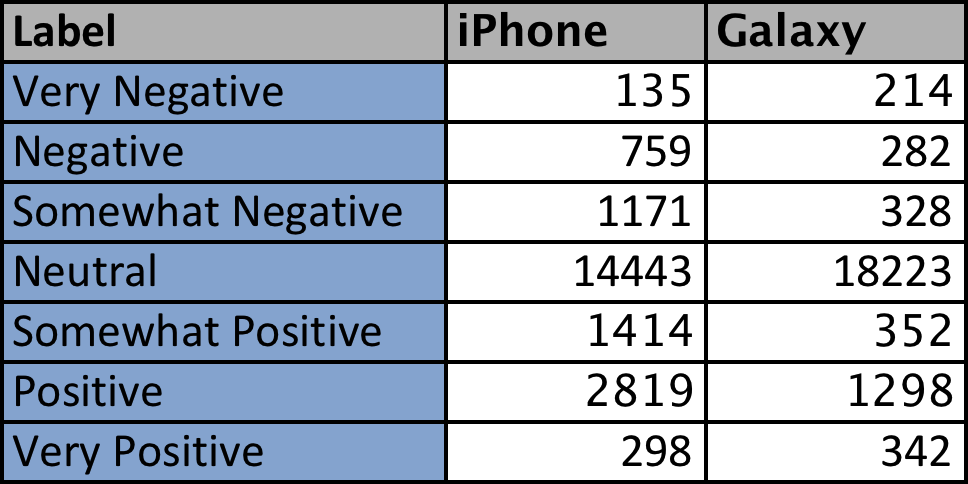
**Project: IPhone & Galaxy Sentiment Analysis**

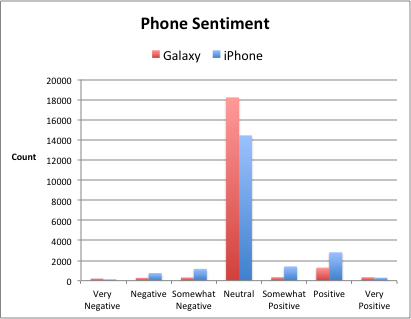
**Client: Helios**

**Nicholas Lynch, Data Analyst, Alert Analytics**

**Overview**: Helio is a smart phone and tablet app company for medical applications and they want to determine which phone model to focus their app development on. Helio wants us to determine the overall sentiment towards the iPhone and Galaxy phones ranging from [very negative, negative, somewhat negative, neutral (neither negative nor positive), somewhat positive, positive, very positive]. Data is based on 21,039 webpages obtained from <http://commoncrawl.org/connect/blog/>. Alert Analytics, applied machine-learning techniques using python, AWS servers, and R to perform sentiment analysis and predictive modeling for future web data. From our research, we recommend that Helio develop their medical app for iPhones and apply our RandomForest predictive model to future web data.

**Sentiment Findings**:





Both phones had more positive results than negative. The iPhone dataset had stronger sentiments over the entire range and had more positive sentiments. Therefore, it is recommended that Helio develop their medical application for iPhone platform. In addition, Apple’s operating system is similar across their phone models, which makes development for iPhone easier than Galaxy phones. The different Galaxy phones use a wide range of android operating systems versions and this makes application development more complicated.

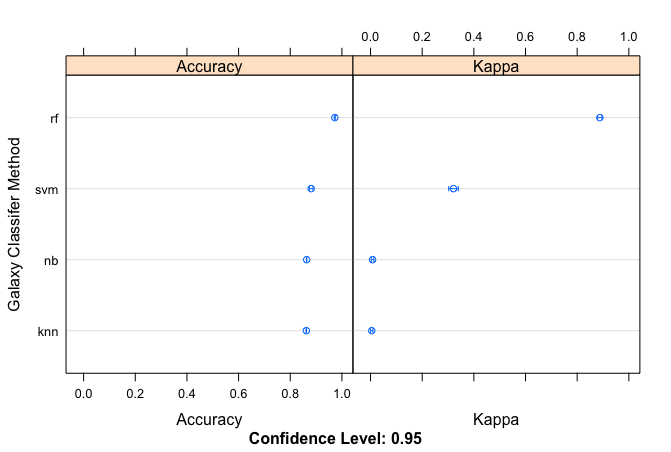
The majority of the webpages had a neutral sentiment because only a small fraction of webpages review phones on Internet. The datasets had a normal distribution. There were no abnormalities noted.

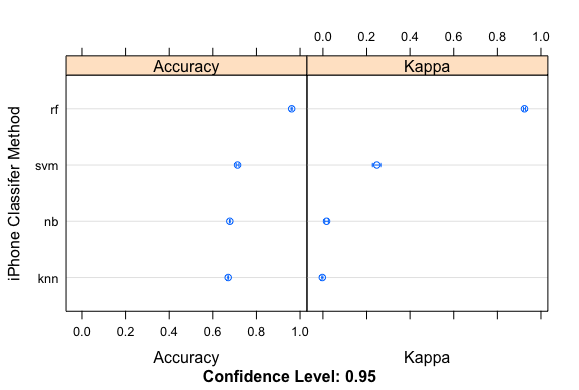
For future projects, Alert Analytics will determine if it is possible to get webpage data just for technology from common crawl organization. As a result, future findings will have more meaningful results because unnecessary data will be removed.

**Predictive model methods & results**: Alert Analytics selected 4000 samples out of 21,039 observations. To create a predictive model for future web data, the following machine learning were run on the samples: K Nearest Neighbor (“knn”), RandomForest (“rf”), Support Vector Machines (“svm”), and Naïve Bayes (“nb”). 2804 sample were used for model training and 1196 for testing our models. The optimal model was rf for the iPhone and Galaxy datasets with approximately 90% accuracy and kappa with 95% confidence interval. The optimal model accuracy was confirmed by applying the rf model to the testing data. No discrepancies noted.

The Alert Analytics rf model can be easily applied to future common web data to predict sentiment analytics in a timely manner.

**Galaxy predictive model training results**:



**iPhone predictive model training results**:

**RF testing data results**:

|  |  |  |
| --- | --- | --- |
| **Model** | **Accuracy** | **Kappa** |
| rf\_Galaxy\_testing\_results | 98.41% | 93.27% |
| rf\_iPhone\_testing\_results | 97.41% | 94.89% |