C5\_T3\_Lessons Learned

Findings

Comparing the iPhone and Galaxy

What were the sentiment categories?

If there were any extremes, what were they?

Include visualizations

Confidence

Accuracies of various models

Implications

Discussion on the impact of analysis on client goals

Discussion on next steps, recommendations, what about similar projects in the future

Overview

The two major companies in this project are Helio and Alert Analytics. Helio is a smart phone and tablet developer, which is working with a government health agency to develop smart phone medical apps for aid workers in a developing country. As for Alert Analytics, which is a data analytics consulting form, it is responsible for conducting sentiment analysis for the sake of supporting Helio’s work.

The scope of this project is to employ sentiment analysis on web data related to smart phone devices. This analysis would be used to ascertain which smart phone devices would be appropriate to focus on for the sake of developing medical apps.

In other words, by ascertaining which smart phone devices are generally favored, this could help inform decisions regarding purchase costs and training. That is, by having a uniform suite of smart phone devices, this could help to reduce costs as phone-related costs are standardized. As for training, this uniformity could help streamline training processes for the medical apps as every aid worker would be using the same smart phone devices.

A necessary component of this project is to obtain web data so that is might be analyzed. This technically-involved aspect of the project was conducted through Amazon Web Services (AWS).

Methodologies

Without going into too much technical depth, the broad process for this project were as followed:

1. Carefully read through any relevant documentation and client requirements in order to understand the project scope and prepare a technical plan for the sentiment analysis.
2. Obtain data about smart phones via webscraping. Webscraping is the process of gathering information from websites. This was done using AWS.
3. Conduct some preliminary statistical analysis on different versions of the smart phone data, such as the distribution of iphone and galaxy sentiment.
4. Test models on different versions of the smart phone data.
5. Evaluate the accuracy performance of different models, with different versions of the smart phone data, in order to ascertain the best performing model.
6. Apply the best performing model to the complete dataset, which refers to step 1.
7. Summarize key findings in the form of a final document.

Findings

It appears that