**C5T3 – Michelle Giniewicz – Findings Report**

**Overview of Project**

Our client is Helio, a smart phone and tablet app developer who is working with a government health agency to create mobile apps in developing countries. Since the government agency will provide tech support, they want to limit support to a single model of smart phone and operating system. To help determine which mobile device Helio should support for their mobile app, I performed data analysis to predict sentiment about both iPhone and Galaxy devices.

**Findings**

I performed analysis to predict sentiment for both iPhone and Galaxy devices using the following scale: 1) Negative, 2) Somewhat negative, 3) Somewhat positive, and 4) Positive. Below is what I found for both devices, based on sentiment of 20,319 websites:

Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated

As you can see from the above pie charts, most sentiment data is either at the negative or positive end of the scale, with very few users in the middle. Based on this, it seems that most reviewers either hate or love the iPhone or Galaxy – few reviewers are neutral. These visuals show that the frequency and distribution of the ratings of both devices are very similar – there are not many differences in the proportion of positive or negative sentiment in iPhones versus Galaxy devices.

**Confidence in Findings**

I analyzed multiple machine learning models using several versions of our training datasets for iPhone and Galaxy sentiment. The model that I used for the iPhone sentiment was Random Forest with the recoded sentiment (reducing the sentiment values from six [0-5] to only four values [1-4]) – this model predicts iPhone sentiment with 85% accuracy and a Kappa Score of 0.62. The model that I used for Galaxy sentiment was C5.0 with the recoded sentiment – this model predicts Galaxy sentiment with 84% accuracy and a Kappa Score of 0.59.

**Implications**

I think this analysis proved that iPhone and Galaxy are both good contenders for Helio to support for their mobile app. They both have around 33% positive sentiment and 57% negative sentiment, which I think is expected since most people seem to either really like or really hate iPhones/Galaxy devices. However, it is also difficult for machine learning to always categorize sentiment correctly – for example, if a user is being sarcastic in a review, or if a review is not completely positive or negative. Finally, I think additional analysis may need to be performed prior to selecting a device – for example, gather data about specific app support on iPhone vs. Galaxy – for example, does one platform have more support issues than the other?

**Methodology**

I began my analysis in R by preprocessing the data and performing feature selection to came up with four different versions of the iPhone data that I then ran through four different models. For each of these datasets, I split them into a 70/30 split for training/test data. I then used predictive analytics to determine which combination of dataset and model was the most accurate. After completing my analysis, my best model was Random Forest with recoded sentiment at 85% accuracy. I then performed the same machine learning methodology for the Galaxy datasets and ended up with a model with 84% accuracy (Random Forest with the reduced sentiment values).