December 13th, 2020

**Helio: Smartphone Sentiment Analysis**

**Consulting Firm:** Alert! Analytics

**Analysis Curated by:** Monika Spreitzer

**Project Overview:** Helio, a mobile app developer, has built a telehealth application to aid health care workers in 3rd world countries. Helio intends to bundle their app with a smartphone device and needs help deciding which one to choose. Helio is asking us, Alert! Analytics, to scrape the internet to find text sentiment regarding smartphones with the object to uncover which smartphone has the most positive sentiment.

**Findings:** Based on the predicted sentiment score, from randomized website scrapping, we can conclude that the majority of web pages analyzed have polarizing sentiments with 42% - 48% of web pages reporting negative **or** positive reviews of both Galaxy & iPhone. The neutral sentiment scores were less than 7% for both devices with the “somewhat positive” scores leading for iPhone.

**Sentiment Count Sentiment Average**

**Table

Description automatically generatedTable

Description automatically generated**

**Chart, bar chart

Description automatically generated**

**Confidence**: Our best model provided us with 85% accuracy of both the iPhone & Galaxy datasets. This accuracy score tells us that we can expect to project the sentiment score with only 15% chance of inaccuracies. We were able to achieve this accuracy score by reducing the sentiment scores into more succinct options. With this minor change, we saw an increase in accuracy by roughly 10%.

**Implications**: Our results exposed that the internet provides us polarizing articles about mobile devices. While our results did produce strong confidence, these results should only be used to accompany additional research. Helio should consider sending a survey to individuals in the medical profession to gather their perspectives on their preferred device to ensure the most informed decision can be made.

**Methodology:** This project is a traditional categorical (classification) problem where we analyzed a baseline dataset (training data) in order to identify patterns and predict uncategorized data (supervised learning). For categorical (classification) problems, we measure success using accuracy and kappa. I like to think of accuracy and kappa like studying for an exam. If we study really hard and hone in on the most important principals (aka tune the model), we will likely get a high grade/score (or higher accuracy). If we know the exam will have true/false or multiple-choice questions, there’s a chance we could get lucky and guess all the correct answers but that doesn’t prove we truly know the material. Kappa helps us understand the reliability of our projections and helps eliminate guessing the answer.