**Alert Analytics**

**iPhone vs Samsung Galaxy**

**Predictive Sentiment Analysis**

- Report –

[Introduction 2](#_Toc509858861)

[Analysis: iPhone vs Samsung Galaxy 3](#_Toc509858862)

[Summary of Findings 6](#_Toc509858863)

# Introduction

**Goal**

*Helio’s goal:* To determine which handset model from the list of five devices will be bundled with the suite of smart phone medical apps for use by aid workers in developing countries.

*Alert Analytics’ role*: Support the decision process through delivery of the sentiment analysis based on website reviews toward set of handset devices conducted through the use of machine learning methods.

**Data**

*Datasets:* Two product review datasets for iPhone and for Samsung Galaxy.

*Source of data:* Common Crawl, an open repository of webpages, stored on Amazon’s Public Data Sets.

*Structure of the data:* each dataset includes the count of words relevant to the analysis organized in 30291 rows (instances) and 60 columns (attributes). Each row represents a webpage containing words associated with sentiment toward a particular device. Columns refer to positive, negative and neutral reviews on smartphones and their key features: camera, display, performance and operating system.

The value in the sentiment column represents the weighted sentiment for each instance, enabling to examine the prevalence of positive and negative attitudes toward devices expressed on the webpages.

**Analytical tools used for data analysis and machine learning**

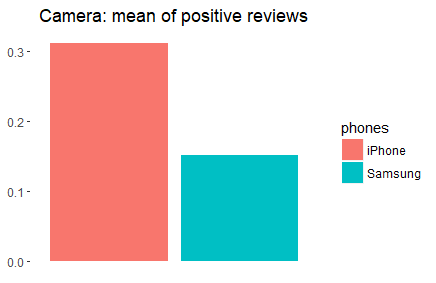
Data collection and the analysis was conducted through the use of the cloud computing platform provided by Amazon Web Services (AWS) and the R programming language.

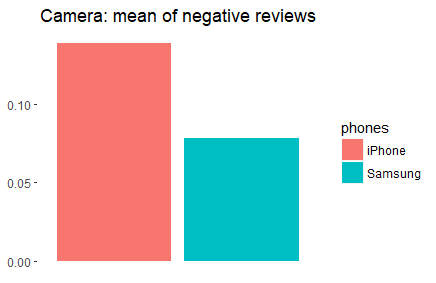
# Analysis: iPhone vs Samsung Galaxy

Twodevices with the highest ratings – iPhone and Samsung Galaxy – have been chosen for the final analysis.

Analysis, based on raw counts of words expressing sentiments, reveals strong trends.

* Dominant position of iPhone (mean 0.84) as well as IOS (0.39) compering to Samsung Galaxy (0.12) and Google Android (0.10).
* Dominant position of iPhone’s features in general – in positive as well as in negative reviews.
* The following charts illustrate differences regarding positive and negative reviews for cameras in both phones.

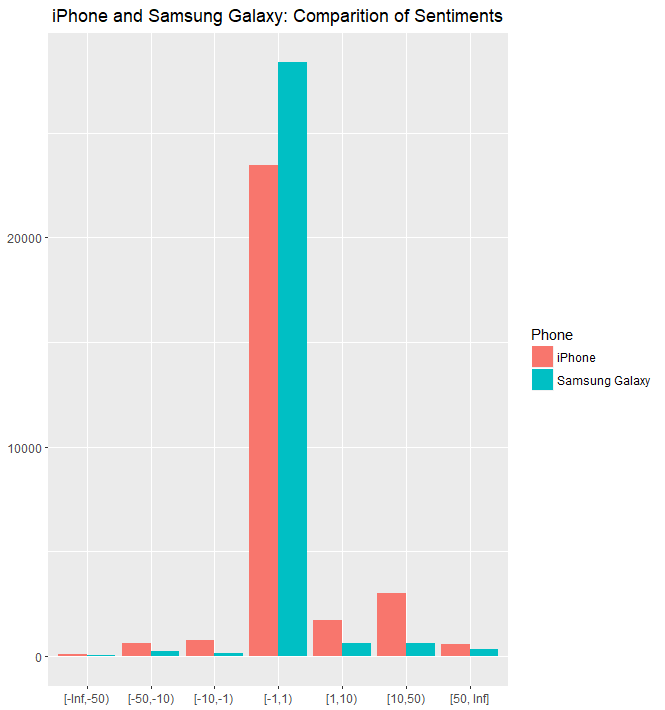




The same pattern is present in all other categories: like display (mean of positive reviews for iPhone – 0.46 and for Samsung – 0.18; mean of negative opinions for iPhone: 0.24, Samsung – 0.11) or performance, as well as IOS (mean of positive reviews – 0.14, negative – 0.09) vs Google Android (0.5 and mean of negative reviews – 0.04).

**Sentiment Analyses**

The following chart illustrates the number of instances in each level for iPhone and Samsung Galaxy.



In general, the highest – middle – bar represents neutral reviews [-1,1), first three groups of bars [-Inf, -1) – negative, and last three groups of bars [1, Inf] – positive reviews[[1]](#footnote-1).

iPhone Sentiment column, which represents the weighted sentiment for each instance, is ranging from -389 to 3607 where “-“ indicates negative sentiment toward handset. The Samsung Galaxy Sentiment column is ranging from -151 to 2829. Therefore sentiments towards iPhone are much more polarized than toward Samsung, which are more evenly distributed.

**Modeling**

Two optimized models to predict the overall sentiment toward iPhone and Samsung Galaxy handsets has been developed in order to conduct future predictions on new data.

For iPhone sentiment modeling, Random Forest was chosen with accuracy: 0.89 and kappa: 0.67.

Random Forest used for Samsung Galaxy sentiment modeling reached accuracy: 0.97 and kappa: 0.75.

# Summary of Findings

iPhone is the most popular product in all the categories, which is strongly supported by the data presenting very clear picture of comparison. There are more positive as well as negative reviews, because in general iPhone and its features are widely discussed on Internet fora, in contrast to Samsung Galaxy. At the same time opinions on iPhone are also more polarized while reviews regarding Samsung more neutral.

The analysis tells a lot about popularity of iPhone vs Samsung Galaxy, however can not be treated as a reliable source of objective knowledge about quality of particular features.

Sentiment analysis shows that there is one leader – iPhone, and the further analysis should be focused on technical aspects of the particular features and solutions. In this case it will be for example camera, as the high quality of pictures will be essential for the application enabling diagnosis of conditions by examining images and other patient data uploaded by aid workers.

1. [-Inf, -50) ***very negative***: iPhone: 72, Samsung: 30 instances

   [-50, -10) ***negative***: iPhone: 632, Samsung: 224 instances

   [-10, -1) ***somewhat negative***: iPhone: 754, Samsung: 138 instances

   [-1, 1) ***neutral***: iPhone: 23485, Samsung: 28375 instances

   [1, 10) ***somewhat positive***: iPhone: 1731, Samsung: 608 instances

   [10, 50) ***positive***: iPhone: 3027, Samsung: 602 instances

   [50, Inf] ***very positive***: iPhone: 590, Samsung: 314 instances [↑](#footnote-ref-1)