# Future-Oriented Benchmarking through Social Media Analysis

**Team Members:** Claudino Alves, alvesc2014@my.fit.edu
Harshil Haumeer, hhaumeer2014@my.fit.edu

**Faculty Sponsor:** Dr Philip Chan, pkc@fit.edu

**Client:** Dr. Richard Griffith, griffith@fit.edu - *Department of Psychology*

**Sponsor meetings:** October 30th 2017

**Client meetings:** October 11th 2017
October 25th 2017

| Task | Completion | Claudino | Harshil | To do |
|---|---|---|---|---|
| Implement, test and demo extraction of reviews from websites (Web Scrapping) | 100% | 0% | 100% | None |
| Implement, test and demo classification of reviews into nodes of hierarchy | 100% | 20% | 80% | None |
| Implement, test and demo Graphical Visualizations | 100% | 0% | 100% | None |

**Discussion of each accomplished task for current milestone:**

**Task 1:** A web-scrapper was successfully created and implemented. To get the data(reviews) from social media platforms, a reporting platform to retrieve the SEO data was used. It was challenging to get the accurate and required data since there was a lot of errors on Orlando International Airport's NAP data. The NAP data had to be modified and then the API calls to the reporting platform made to get the data(reviews). Data were successfully obtained from several social media platforms including Facebook, Yelp, Foursquare, Yellow Pages among others.

**Task 2:** A text classifier was successfully built, which puts reviews into the different categories (7 factors). This was done using Bayes Naïve algorithm. The program uses sample data in from a csv file which has 2 columns, with column one containing the reviews and column 2 containing the categories to train the system. The system was then trained several times to improve the accuracy.

**Task 3:** A Graphical representation of the average sentiment score of all the categories (7 Factors) was successfully implemented. A python library, matplotlib, was used for that which displays the categories on the x-axis and the sentiment score on the y-axis.

**Discussion of Contribution for Each Task:**

**Claudino:** Claudino attempted to write a text classifier. He used the NLTK library and Bayes Naïve algorithm.

**Harshil:** Harshil created and implements a web-scrapper. He created an account on a reporting platform to obtain SEO data and an API key which was very important access data from social media platforms. The reporting platform were returning the wrong data since the NAP data were wrong and outdated. He had to fix the NAP data and generate a new report with the updated NAP data which improved the NAP score from 23% to 66% returning the correct data. Harshil also built and implemented a text classifier which used Bayes Naïve algorithms. He trained the system using sample input text which had assigned categories saved in a csv file. He then implemented a sentiment analysis using a python library. Harshil also implemented a graphical visualization to display the average sentiment score for each category.

**Plan for the Next Milestone:**

| Task | Claudino | Harshil |
|---|---|---|
| Implement, test and demo sentiment measurement of reviews (Sentiment analysis) | 30% | 70% |
| Implement, test and demo visualization of sentiment in different categories of hierarchy | 70% | 30% |
| Implement, test and demo benchmarking analysis tool | 50% | 50% |
| Set up Website | 50% | 50% |

**Discussion of each planned task for the next Milestone**

**Task 1:** We plan on implementing a sentiment analysis function which will calculate the sentiment score for each individual category and the average sentiment score for all the 7 categories.

**Task 2:** We will attempt to implement a graphical illustration of the sentiment score for the different categories and the overall sentiment score for the 7 categories

**Task 3:** We will attempt to implement and demo a successful benchmarking tool which will be used to compare Orlando International Airport's performance in its different categories as opposed to that of other airports.

**Task 4:** We will attempt to setup a website with at least the main page and hopefully the different GUIs described in the Design document.

**Sponsor Feedback:**

- parking lot review is not in output.
- 250 might not be sufficient (one month)
- ~~text~~ only text is relevant for classification.
- double check you're using Naive Bayes for text properly
- examples for more than 2 classes.
- train corpus: each review should have only one category.

**Sponsor Signature:** _____ **Date:** 10/30

**Sponsor Evaluation**

- Sponsor: detach and return this page to Dr. Chan (HC 322)
- Score (0-10) for each member: circle a score (or circle two adjacent scores for .25 or write down a real number between 0 and 10)

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Claudino | 0 | 1 | 2 | 3 | 4 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |
| Harshil | 0 | 1 | 2 | 3 | 4 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |

Sponsor Signature: _____ Date: _____