

## Project Progress Report - CS410 - NVK

Nikhil Khandekar (nikhilk5) (captain)

Vineet Chinthakindi (vineetc2)

Kyr Nastahunin (knasta2)

### 1) Which tasks have been completed?

So far we have been able to research text extraction algorithms and decided on implementing the RAKE keyword extraction algorithm for our project. We were able to successfully implement the algorithm and it has been producing good keyword results for the reviews we have tested it with so far. In addition, we have implemented the UI/backend to retrieve the customer reviews from Google Places and run our keyword extraction algorithm on them. This allows us to not hardcode anything, and instead, we can already dynamically fetch keywords for any business.

### 2) Which tasks are pending?

We are currently trying to improve our text extraction algorithm to filter out some keywords that are primarily stop words or words that don't have much meaning by themselves. In addition, we still need to start building a review sentiment classifier to be able to classify the keywords that we have identified in the reviews. We plan on training and evaluating the sentiment classification models next week. We need to identify proper training sets for those. We will also compare how our model does compared to the already existing VADER sentiment library in order to determine which sentiment model we will choose for our final project

### 3) Are you facing any challenges?

We had a problem with getting YELP reviews since the Yelp API returns only a small limited number of symbols in each review, which doesn't fit the idea of our project. Therefore, we have successfully switched to using the Google Places API.