**Project Title: Negative Reviews Extraction**

**Team Information:** Long Ly ([long.ly@mavs.uta.edu](mailto:long.ly@mavs.uta.edu)) 1000866434

**Objective and Overview:**

Location used to be the most important factor that determines the success of the business. Choosing the best location would definitely boost the chance of being success. However, with the fast growing of the Internet, reviews now are even more important than the business’ location. Reviews of a business do not only contain additional information about the business, but also provide customers’ feedback on various categories during their visit such as experiences, services, food, etc. The higher the rating of the business, the more popular that business would become. In addition, the review is not only important to the business itself, but also to the new customers. Many people look on the current reviews on Yelp before they pay visit to the business. Moreover, negative reviews are even more important than positive ones. Business owners would look into fixing what they are lacking off from negative reviews so that they can improve their business; new customers would also have more information about what they should expect from the business that they are planning to visit. Therefore, the objective of this project is to extract the information that has the most negative reviews for a particular business. Due to the large size of the dataset, the scope of this project is limited to restaurants only.

**Data Mining Tasks:**

Words in customer reviews will be extracted, stemmed and tokenized using NLTK. In addition, negative review does not necessarily mean low rating review; it can be any information that customers do not like about in any reviews including both low and high rating ones. In addition, business dataset will be classified into different categories and only restaurant category will be used for this project.

**Plan and Implementation:**

The deliverable would be a program that demonstrate the objective of this project by the end of the semester. Similar to Yelp’s website, users can type in keywords and choose the location that they want to search. A list of restaurant will be showed and users will have to choose the restaurant they want to have negative information.

**Challenges:**

The most challenge of this project is the size of the dataset. Due to the large dataset, the process of classification and extraction should be efficient in term of time and space.

**Evaluation:**

The efficacy of the program can be determined by manually looking at reviews from Yelp’s website. Since the dataset is not up to date, we should expect to have some inconsistency between the result from the program and the Yelp’s website.

**Task Partition:**

Since I’m the only member, I will be responsible for doing all the task.