**Abstract**

Founded in 2004, Yelp is a local-search service powered by crowd-sourced review forum. The company trains small businesses in how to respond to reviews, hosts social events for reviewers, and provides data about businesses. Often, a review describes various dimensions about a business and the experience of the user with respect to those dimensions.

The aims of the project are to analyze reviews, classify them and respond to the following questions:

* Do people tend to like a positive review or a bad review?
* What are the most common words in positive/negative reviews?
* Are reviews specific to one category of business?
* What are the most complaints about businesses? (Topic Modelling of the reviews) Can we classify them concerning the business categories?
* Are the categories of trending businesses different from top reviewed businesses?
* Are reviews influenced by user's friends? (relationship between users' friends and review patterns)

I will use a subset of the Yelp Challenge Round 12 dataset and apply text classification and sentiment analysis to respond to the questions above. All this will be done with R and Spark.