**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.

Our main research question will focus on: What are the characteristics of leaders among users?

To do so the aims of the project will be to analyze reviews, define criteria for leaders and respond to the following questions related to those leaders:

* 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.

Research questions

discover how the characteristics of reviews related to ratings of those reviews

what characteristics make these user-generated reviews credible to other consumers

find the contextual commonalities the reviews have under each category of useful, cool or funny as classified by review readers

The purpose is to identify the trends that define the type of language and symbols used in each of the three categories that review readers used in their evaluations

helps researchers to better explore the ways that peer-driven communications work in a digital arena such as a consumer-driven website

This research closely examined the specific approach the users employed in the Yelp speech community to communicate their message to other people seeking information from fellow consumers like themselves

**characteristics of leaders among users**