

## Project Proposal

**Help new business to develop faster by analyzing trends, discovering good and bad in already well-established similar ventures using yelp data.**

Team members: Jvalin Dave, Mudit Pradhan

### **Objective and overview:**

Owners of the new ventures such as hotels, restaurants or home services need a comprehensive study on latest trends in the location. for e.g. if we take restaurant business, services such as home delivery, customer follow up and complementary offers can be trending in one location while restaurant types such as Asian, Lebanese or Mexican can be trending in another area based on density of people from one origin in that location. It is important to know the latest trends for which people really care about in a particular business, and this has to be overall and also location specific. This would help him establish business with the appropriate product and services. Apart from this, the new owners would also want to know good and bad qualities in already established similar ventures in the same location, this would help them develop faster and also lessen the risk of failure. for e.g. if a new branch of Mc Donald's needs to be opened some another country like U.K., they need to find out the what the local established burger restaurants are doing and also what people looking for in terms of services and product in the reviews. We need to mine the data very specifically for one particular kind of business.

### **Data mining tasks:**

1. Data Cleaning
2. Semantic Data analysis: trend analysis w.r.t location, over all trends, Reviews analysis to find out the best products, services, quality control offered.
3. Classification: location based, business based, product based, price range based, timings, services based, quality based, number of reviews, review based.
4. Clustering: product types, location trends, services, overall trends in a business.
5. Predictive tasks: prediction of trends, top practices and worst practices from established businesses.

### **Deliverable:**

Tool to help a new business setup to develop by identifying the top trends for a particular business, firstly specific to an area and also overall trend. A comprehensive study of good and bad practices by the well established businesses in that location represented as charts and graphs. This would help a new business owner to understand the market requirements, establish, develop faster and reduce the risk of failure. To demo our work we would prepare a paper explaining used techniques and demonstrate visualizations of findings on a website.

### **Challenges:**

- Clean the data to suit our needs.
- Understand the trends in chronological order.
- The model should be capable of understanding review sentiments and notion. Using n-grams would produce a meaningful description
- Trend analysis in a particular region.

**Efficacy evaluation:**

Create a local test data along with good reviews and negative reviews. Search for local/known data. Check it via different dimensions such as just service or combination of service and quality.

**Design and implementation:**

Cleaning the data would be the most challenging task in order to achieve accurate results. We would use classic data cleaning techniques based on location and time relevant data. To design and implement the solution we would use data mining techniques. The techniques specifically used would be tf-idf for the textual similarity, decision tree approach for classification and K Means for clustering. Use n-grams to get accurate result. To further our predictions we would consider using Naive Bayes for the outcome.

**Task Division-**

Mudit Pradhan:

- 1) Analyze reviews, determine test data.
- 2) Data cleaning.
- 3) Implementation of classification algorithm.
- 4) Test for efficacy.

Jvalin Dave:

- 1) Data cleaning.
- 2) Textual analysis implementation.
- 3) Clustering algorithms.
- 4) Visualization on website.