

Project Title

Preparing for Future: Help future businesses determine success in starting a venture

Team Info

Mohammad Kothawala (solo)

Objective

Determine how reviews, tips, locations of different businesses help prospective business owners to set up business.

Overview

Yelp was founded in 2004 to help people find great local businesses like dentists, hair stylists and mechanics. This project aims to help for individuals who want to set up a business and are encouraged to take corrective measures from the initial stages so as to have a lasting impression. This will be accomplished by the analyzing of sales patterns, reviews, tips, location, timings, star ratings and several other factors.

Data Mining Tasks to be employed:

Since yelp data set is a huge data set, partitioning a large set of objects into homogeneous clusters will be a fundamental operation in this project. While doing the cluster analysis, I will first partition the set of data into groups based on data similarity and then assign the label to the groups. Decision tree will be used to assist future owners to determine if their business would be successful or not. Map-reduce algorithm shall also be made use of

End Goals to be accomplished:

Through this project the end user (prospective business) will be able to view the success rate of a business based on the factors outlined above. All of the results obtained will be done graphically as well as visually on a website.

Challenges of this project:

Dealing with such a large dataset itself poses a big challenge. Using portioning algorithm and converting it to training data will also be a big challenge. Also only reviews and rating are not the only factors that determine the profitability of the business. Hence each attribute governing the profitability shall be taken in to account thus promising more complexity.

Addressing challenges

To overcome the challenges I plan to model the data in such a manner such that a similar pattern is rendered. Moreover various algorithms and models will be employed to solve the complexity of such a dataset

Evaluating Efficacy

The efficacy of this project shall be determined by testing manually whether the business in question is assigned a success label or not, on the basis of reviews, neighborhood etc.

Division of tasks

I intend to do the whole project on my own, hence no division of labor