Yelp Data Analysis

Bharat Mukheja(bkmukhej)
Mitkumar Pandya (mhpandya)
Rachit Thirani (rthiran)

1 Introduction

This project utilizes different algorithms and techniques related to Business Intelligence to find answers to questions important to the business and the users.

2 Data Set

For this project is Yelp Data Set which is available on. `http://www.yelp.com/dataset_challenge` The data set consists of user details, business details, reviews given by the users for different businesses, business related tips given by the user and checkin details of different businesses for 7 days on hourly basis. This dataset includes information about local businesses in 11 cities across 4 countries. The dataset provided consists of json objects.

3 Business Problems

In this project we aim to find solution of 3 business related problems.

3.1 User review analysis

For the first problem we want to analyze what the users says about a business and what are the user's sentiments for the particular business. As a solution we want to analyze sentiment of the user reviews and further we would like to analyze trend and seasonality on positive and negative review sentiments using Timeseries analysis.

3.2 Checkin Analysis

In addition to reviews we want to analyze the Check-In data to understand the headcount patterns at particular business. In addition to providing the most favorable times and headcounts at those times for the business, it would allow us to supplement such data to the sentiment analysis to form a better recommendation system.

3.3 Anomaly Detection

For this problem we are trying to find out potential users whose reviews are not trust worthy.

4 Papers to read

[1] M. Joshi, D. Das, K. Gimpel, and N. A. Smith. (2010) Movie reviews and revenues: An experiment in text regression, *NAACL-HLT*

[2] Mladen Marovic , Marko Mihokovic , Mladen Miksa, Sinisa Pribil, & Alan Tus. (2011) Automatic movie ratings prediction using machine learning , *MIPRO*

[3] Sentiment analysis using product review data, Journal of Big Data, 2015, Volume 2, Xing Fang, Justin Zhan

[4] Extracting Strong Sentiment Trends from Twitter, Patrick Lai

5 Work Division

Bharat - Time Series Analysis on Checkins
Mitkumar - Sentiment analysis and timeseries analysis on sentiments
Rachit - Anomaly Detection