

Term Project Proposal

Project Title:

Estimating the ratings of user reviews based on processed text review data.

Team Members: Anshul Vyas

Guidance By: Dr. Hui Yang

Associate Professor,
Computer Science Department, SF State

Project Description:

- About the company and the dataset:

Yelp! is a multi-national corporation which develops, hosts and markets Yelp.com and the Yelp mobile app. These publish crowd-sourced reviews about local businesses. The company also hosts social events for reviewers, and provides basic data about businesses, such as hours of operation.

Yelp! provides datasets which includes basic information about businesses, user reviews, ratings, as well as tips and also user data.

- The project will include processing of the text reviews provided by the user for local businesses across USA. It will address the problem of how well can we predict the reviews based on its text alone. Also, to identify various positive and negative words used in reviews. Maybe, also identify the correlation between user reviews and tips.
- I am particularly interested in this problem because by processing text data reviews will refine the rating system furthermore, making it more reliable. Moreover, using Yelp! Datasets gives me the opportunity to enter Yelp! Dataset Challenge, which will assist me to advance into progressively higher responsibilities while providing me the satisfaction I earn by successfully completing the challenge.
- The data has been made available by Yelp! to whoever takes part in the Dataset Challenge. The data is available in JSON format and it contains data of businesses, reviews, user and tips.

Learning Goals:

- To learn how to analyze huge amounts of data and learn more about Natural Language Processing (NLP) and other necessary concepts of knowledge discovery from data (KDD).
- To be able to detect new patterns and anomalies in the data to be processed and learn more about pattern mining and cluster analysis.
- To learn more about the “How to”s of challenges and competitions by competing in Yelp! Dataset Challenge and successfully completing the challenge.
- To expand my knowledge of Data Mining to aspects of this project I find particularly interesting.