CS 491 – Introduction to Machine Learning

Project Proposal

**Team Members:**

* Asif Mansoor Amanullah (674784173)
* Dileep Kumar Irvichetty (654232610)
* Rajnish Kumar (675531042)

**Tentative title:** Predicting the number of useful votes a yelp review will get.

**Project Overview:**

The objective of our project is to find the number of useful votes that a yelp review will receive. This project was part of Kaggle’s[[1]](#footnote-2) competition. The dataset is also provided as a part of the competition.

Yelp is an online platform where users post reviews on the different businesses. These reviews are read by lot of users. The reviews are sometimes voted whether they are ‘useful’, ’cool’ or ‘funny’.

We propose to model a machine learning algorithm, which would learn from our training data of reviews and useful votes and will predict the useful votes on the test data.

**Dataset:**

The data for this project is got from Kaggle competition and the training set has 11537 business, 8282 checking, 43873 users and 229908 reviews datasets. We are planning to use the major part of this data for training and remaining part for testing.

**Machine learning tasks:**

We will do exploratory data analysis to understand the characteristics of this dataset.

* Data Pre-processing – Our data is in JSON format and it has to be pre-processed, for fetching the required data like user\_id, review\_id, review, useful\_votes count etc.,
* We plan to use following text preprocessing techniques like stemming, removing the stop words, parsing, tokenization etc.,
* Feature generation.
* Identify the machine learning model which best fits our data.
* We will evaluate our model based on the RMSLE value. (Root Mean Squared Logarithmic Error)

**Machine learning techniques planning to employ:**

1. We are looking to employ following supervised learning regression techniques to find number of useful votes.

* Multiple Linear Regression
* Support Vector Regression
* Decision Tree

1. Kaggle is an online platform for predictive modeling competitions [↑](#footnote-ref-2)