[ECE 180] - Team 11 Project Proposal
Purpose

Ratings and reviews play a vital role in determining the performance and success of a business. Websites like TripAdvisor and Yelp enable users and consumers to rate services and products of different businesses on a scale of 1 to 5 stars, as well as provide a free-form text review accompanying the respective rating. The purpose of this project is to predict the ratings of businesses from different attributes, like location, price range and other such features. This would give us an understanding of how these attributes affect the rating and businesses can improve their rating by changing few of these attributes accordingly.

Data Acquisition and Analysis

We will be using the dataset from Yelp dataset challenge. It is available at yelp.com/dataset/challenge. This set includes information about local businesses in 11 metropolitan areas across 4 countries. Given the size of dataset, we would be concentrating only on specific cities in United States. The provided dataset is in JSON format and we intend to use the json library in Python to parse the data. For data mining and data analysis and scientific computation, we will be using libraries like Scikit-learn and NumPy respectively.

Github repo: https://github.com/tarekgouda/ECE180Team11

Expected Timeline

Task	Expected Time	Assigned Members
Data Pre-Processing & Formatting	February 9 - February 13	Tarek & Ibrahim & Chenhui
Data Analysis	February 14 - February 21	Ibrahim & Mayu & Kavya
Prediction Models	February 22 - March 3	Kavya & Mayu & Binh & Hao
Conclusion and Report	March 4 - March 7	All team members
Presentation	March 8 - March 15	All team members