CS 410 Group ZZZRQ
Term Project Proposal

Team Members Information:

Name: Zetian Li (captain) (NetID:zli123)
Name: Xinhe Xu (NetID:xinhexu2)
Name: Rick Wang (NetID:haoran14)
Name: Qiwei He (NetID: qiweihe2)
Name: Zhichao Fan (NetID: zhichao8)

Github Repo: https://github.com/zli123/CS_410_Group_ZZZRQ

Project Topic:

Sentiment analysis of Yelp reviews

Project Description:

Through retrieval and analysis of text data published by Yelp, come up with a predictive model that gives a score on the sentiment of customers towards a particular review, with 0 being the most negative and 10 being the most positive. Using this model, analyze a set of Yelp reviews dated from October 2019 to June 2022, and acquire the average sentiment score per month, which will be plotted against the time to visualize the average sentiment changes over the period of the pandemic. Results will be further de-aggregated according to restaurant type, time of the day when review was made and geographic location. This is to help restaurant owners better understand the correlation of sentiment of reviewers with the progression of time through the pandemic along with their sensitivity with several other factors at stake. The results will be evaluated through comparing the score with the associated yelp review score; they should be positively correlated with a high confidence level.

Tech Stack for project:

Python(Numpy, Pandas)
Microsoft Azure Text Analytics REST API, NLTK, whoosh
Dataset: Yelp's published dataset, sentiment dataset from various sources (will be cited)

Project tasks: (Estimate hour: 110 hr)

1. Data collection from Yelp published dataset and sentiment dataset from various sources (5 hr)
2. Data cleaning and tag the data for model training (30 hr)
3. Model Training (25 hr)

4. Data Analysis (30 hr)
5. Data Visualization (20 hr)

Schedule:

Data collection and Data cleaning (Week 10 - Week 12)
Model Training and Progress Report (Week 12 - Week 13)
Data Visualization and Data Analysis(Week 13 - Week 15)