SDS 625 - Case Studies
Rachel Guan, Kelly Wang

Project Proposal

Project Description:

We aim to develop an intuitive application that tailors restaurant recommendations for users based on specific attributes they prioritize. In modern society where restaurants can be judged and assessed at the tip of a finger, customers often have distinct preferences that go beyond just the taste of the food; they might be focused on aspects such as hygiene, ambience, family-friendliness, vegan options, etc. By utilizing a sophisticated machine learning model, our application will filter through vast amounts of Yelp reviews, ensuring that users find restaurants aligned with their unique preferences. Not only will this enhance user satisfaction, but it will also provide a more tailored dining experience, saving users the time and effort typically expended in sifting through numerous reviews.

We will go through this by the use of machine learning algorithms as well as textual sentiment analysis derived from these comments by the users of Yelp. In doing so, we hope to generate something that not only tells us about the details of the restaurant that can't be directly seen from a surface level of the place, but we also hope to provide a more in depth understanding of how people feel and how their experiences reflect on the restaurant itself as well. We believe this will be of great utility to the public because it serves to provide a more detailed view of a wide scope of restaurants for the multifarious needs and preferences of everyday people. And on top of that, instead of anxiously scrolling through endless reviews to find the perfect place, we can utilize the application to briefly but effectively have an understanding of a certain restaurant.

Data Set and Source:

Our primary dataset will be sourced from the Yelp Open Dataset, available at the following URL: https://www.yelp.com/dataset. This dataset offers a wealth of information, including text reviews about restaurants, their corresponding ratings, and the number of "agreements" or endorsements from other users, signaling the perceived usefulness or accuracy of each review. Given the richness of this dataset, we can extract multifaceted insights that cater to a broad spectrum of user preferences.

Questions and Real-world Applicability:

With this application, we seek to answer the following questions:

1. How can we effectively match individual user preferences to restaurants that align with those specific attributes?
2. How can a machine learning model be fine-tuned to give weightage to certain attributes in Yelp reviews?
3. How do user endorsements (or "agreements") impact the credibility and relevance of a review?
4. How can we measure the sentiments and "positivity" or "negativity" of comments and how can we effectively display that to an audience?
5. What is a way people can easily obtain information and access about the credibility and appeal of restaurants? How can we personalize a large platform of restaurants to specific needs from individuals?

The real-world applicability of our project lies in its potential to revolutionize the way users interact with Yelp, making the restaurant selection process more personalized, efficient, and enjoyable. Given the rise of personalized content in various sectors, our application caters to an evident need in the restaurant and review environment.

Analysis and Data Usage:

We will employ a fine-tuned language model that ingests review text data about restaurants, their ratings, and the number of user endorsements. By transforming this input data, our model will output a multi-dimensional vector indicating the relevance and utility of the review based on the attributes chosen by the user. This vector, in turn, will drive our recommendation engine in the backend, presenting users with a curated list of restaurants that align with their preferences. We also hope to provide a refreshed rating of ratings based on the sentiment analysis of the comments in the Yelp database as well. This way, we can create an entirely new point of view of the restaurant ratings just from the comments as an addition.

By meticulously refining the data processing and model training stages, we aim to ensure that the datasets we employ fully support and enhance our analysis, leading to accurate and tailored restaurant recommendations for our users.