Effects of Restaurant Neighborhood and Cuisine on Rating

Pradeep Gopinathan

July 12, 2020

1. Introduction

1.1 Background

Urban areas are some of the hardest areas for new restaurants to be built in, as the costs of doing business are high, and the competition is fierce. However, the results of success can be quite lucrative. One important factor to consider is the population of the neighborhood, and what kinds of cuisine they do or do not want. Would-be restauranteurs need to consider if the neighborhood they are scouting as a location has an actual demand for what they are selling. Social media sites like Foursquare can help us understand this demand by showing current restaurants in the neighborhood, and what visitors have rated them as.

1.2 Problem

Our goal for this analysis is to attempt to build a predictive model that can see what rating a restaurant might get depending on the neighborhood it is located in and its type of cuisine. For this preliminary research, we are limiting ourselves to just downtown Toronto and the surrounding boroughs. Foursquare ratings are given as rankings between 1-10, so we will treat the problem as a classification one.

1.3 Interest

Our hope is to give restauranteurs a better understanding of what the population of each neighborhood is looking for in a restaurant, and where the food they hope to serve will be most welcomed.

2. Data

2.1 Data Source

For this preliminary investigation, we will limit ourselves to data accessed by the Foursquare API. First, we will collect a list of restaurants within each neighborhood of Toronto. Afterwards, we will collect the rating and the number of tips for each restaurant through the Foursquare API.

Foursquare's API allows developers to search for venues a certain distance away from a specified latitude and longitude. We will collect our data based of looking for all venues .8km away from the latitude/longitude of each neighborhood center (which was collected prior to this analysis), and will filter for only venues whose Foursquare assigned category is Food or some sub-category of Food.

2.2 Data Processing

First, we will have to de-dup the data collected, to ensure venues that are close enough to multiple neighborhood centers are not double counted. Then, we will filter our all venues that Foursquare could not provide a rating for.

Rating data is often provided as an average of user ratings. To turn this value into a classification, we will convert each rating provided into an integer between 1-10 by rounding the number given.

Finally, we will use one-hot encoding to turn the neighborhood and category of each venue into a list of binary features usable to machine learning, and these encoded features will make up the feature set used by our model.