CAPSTONE PROJECT - RESTAURANT CLASSIFICATION

NIMA AMINI

INTRODUCTION

Being somewhat disappointed after trying out a new restaurant is quite a common experience in Stockholm county. It would therefore be beneficial for restaurant customers to get some guidance as to whether a restaurant in Stockholm county is worth trying based on a few immediately visible attributes which are easy to input by the customer. Thus our question is:

Can we classify whether a food place in Stockholm county is "good enough" or "not good enough" before ordering in it?

Of course we need to define what "good enough" means. A natural approach is to put a threshold based on the average rating given by customers over all restaurants in Stockholm county. We shall define a restaurant to be "good enough" if its rating is above average and "not good enough" otherwise. Since putting a hard threshold is likely to lead to marginal misclassifications we are also interested in the relative probabilities between "good enough" and "not good enough" to get a sense of how strong the classification is. Our aim is to see how much two fundamental attributes: *venue category* and *price tier* can tell us about the deviation of restaurant rating from the mean.

DATA

To build our model we first require the geocoordinates of all restaurants in Stockholm county. One way to collect this data is to look at a small (e.g 500 meter) radius around the geolocation of each postcode in Stockholm county. A list of all postcodes in Stockholm county can be retrieved via the postcode population spreadhsheet provided by SCB (the Swedish government agency responsible for producing official statistics regarding Sweden). To get the geolocation for each postcode we use the 'Here API'. Moreover to explore the venues around each postcode we use the 'Foursquare API'. We filter the results on venue category for words like 'Restaurant', 'Burger', 'Food' and 'Steak'. To each food place venue we retrieve rating (a float 1-9) and price tier (an integer 1-4) via another (premium) call to the 'Foursquare API'.

	Postcode	City	Latitude	Longitude	Venue	Venue Category	Venue ID
0	11115	STOCKHOLM	59.33913	18.06768	Surfers Stockholm	Szechuan Restaurant	5491ccf5498ee346e18b6a76
1	11115	STOCKHOLM	59.33913	18.06768	Vassa Eggen Restaurant	Steakhouse	4adcdaf0f964a520535b21e3
2	11115	STOCKHOLM	59.33913	18.06768	Doktor Mat	Modern European Restaurant	573eda6f498eff71f2749af4
3	11115	STOCKHOLM	59.33913	18.06768	Bar Central	Eastern European Restaurant	54f04f99498e311c661d5966
4	11115	STOCKHOLM	59.33913	18.06768	Ingers Kitchen	Asian Restaurant	53e7a0f3498e8e59724e4b9c

	Venue ID	Price Tier	Rating
0	5491ccf5498ee346e18b6a76	NaN	8.9
1	4adcdaf0f964a520535b21e3	4.0	8.3
2	573eda6f498eff71f2749af4	NaN	8.6
3	54f04f99498e311c661d5966	2.0	8.0
4	53e7a0f3498e8e59724e4b9c	2.0	7.9

Rating distribution