Capstone Project - Restaurant Classification

November 21, 2018

Business Question

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

Business Question

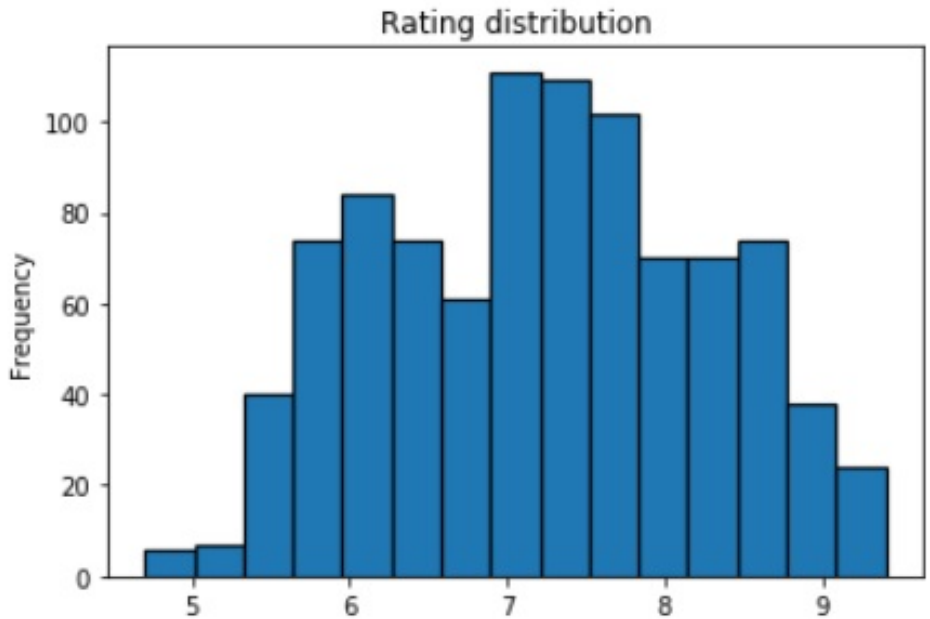
Can we classify whether a food place in Stockholm county is "satisfactory" or "not satisfactory" before ordering in it?

Which type of restaurant, in which price tier, and in which location maximizes the probability of being a "satisfactory" restaurant?

What is "satisfactory"?

Definition

We define a restaurant to be satisfactory if its customer rating is above average and "not satisfactory" otherwise.



Average rating: ~ 7.164

Input/Output of model

Inputs:

- **Venue Category**: One of 19 categories, one-hot encoded
- **Price Tier**: An integer 1-4
- **In Stockholm City**: A binary feature

Note: These inputs are easily observed by the user of the model.

Output:

- Probability p of restaurant being "satisfactory".

Analytical Model

- Binary classifier "satisfactory" is fuzzy so we use Bayesian approach.
- We therefore choose **logistic regression** as our analytical model.
- All data is standardized before model is fitted.
- Data is split 80/20 between training and test data.

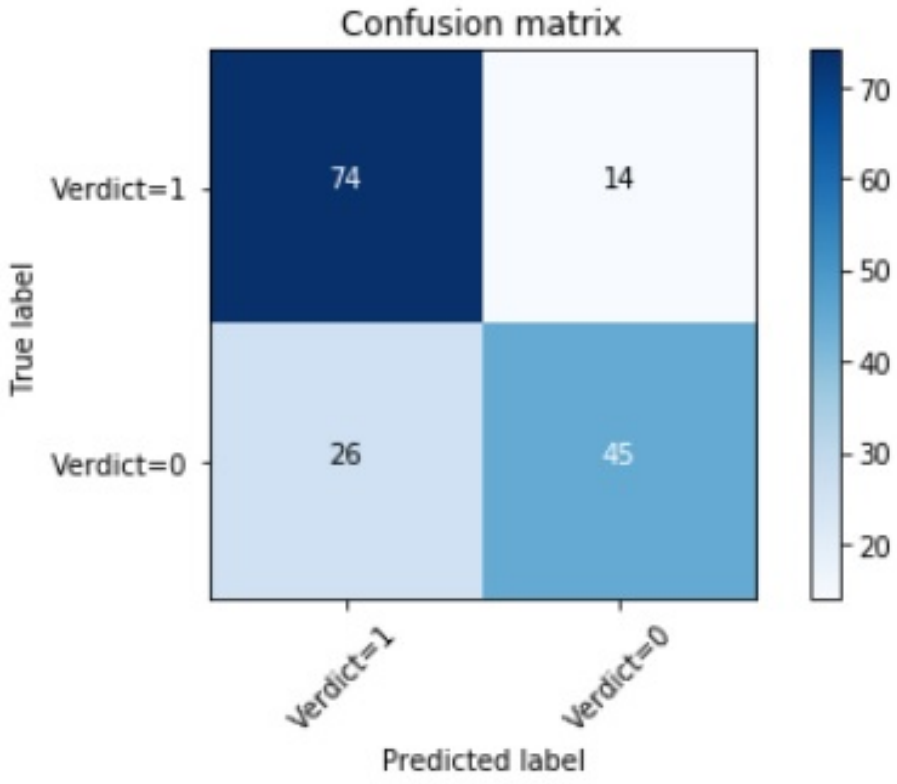
Modeling choices

- Using all city locations in Stockholm county and all venue categories leads to model with bad accuracy due to too many features.
- We limit venue categories to categories with at least 10 data points.
- Majority of data points belong to Stockholm city, so we reduce location data to a binary flag whether or not restaurant is in Stockholm city or not.
- Restaurants with rating but no price tier is filled in using the average price tier which is 2.

Data Sources

- Postcode and City information from SCB (Swedish government agency) spreadsheet.
- Geolocations via Here API
- Venue, Price Tier and Rating information via Foursquare API

Results



Verdict = 1 means "Satisfactory" and Verdict = 0 means "Not Satisfactory".

Precision = 75%, f1 score = 0.74, LogLoss = 0.6

Conclusion

An **expensive Scandinavian Restaurant in Stockholm City** has the maximum probability ($\sim 80\%$) of being a "satisfactory" restaurant according to customers.

A **cheap Fast Food Restaurant Restaurant not in Stockholm City** has the minimum probability ($\sim 50\%$) of being a "satisfactory" restaurant according to customers.