# **Task 7: Restaurant Hygiene Test**

Changhun Park

## 1. Title of Function

I implemented the function of topic modeling using LDA and the function of classification using SVM. The functions are used to predict if a restaurant would pass a hygiene inspection or not, based on such features as review text, zip code, review count, average rating, and provided cuisine list.

I used LDA for constructing topic vectors after tokenizing review texts, removing stop words and stemming. I used SVM for classification after constructing the feature table composed of topic vectors and additional data which are normalized or encoded as one-hot vectors.

## 2. Description of envisioned application scenario and intended users

Some users would like to know if this restaurant is clean enough to dine. The users might go through the review of a restaurant, but it is still hard to conclude. To solve this problem, I created an application system to decide if the restaurant has passed the hygiene test or not.

## 3. Description of interface design

The users can input the data of a restaurant they would like to test into 5 parts as shown in Figure 1.

First, the user inputs the restaurant name on the "Restaurant Name" field.

Next, the user inputs the 5-digit zip code on the "Zip Code" field; The user is allowed to input only the California zip code from 98101 to 98199 since the data trained in the SVM were those zip codes.

Next, the user inputs the rating of the restaurant from a scale of 0.0 to 5.0 on the "Rating" field.

Next, the user selects one or more types of provided cuisines from the list of 98 cuisines.

Next, the user inputs a review text of the restaurant on the "Review text" box.

Finally, the user presses the "Hygiene Test" button to see the results. The result will show the name of the restaurant and a message "Passed!" or "Failed!". For caution, if the user inputs the zip code in wrong, or inputs the rating without decimal point, the result will show simply "Failed!"

# Restaurant Hygiene Test Welcome to the restaurant hygiene test. You are able to test some restaurants in California(zipcode 98101–98199) and see whether the restaurant passed or failed the hygiene test Restaurant Name Zipcode(Make sure you are typing in 5 digit numbers) Rating (Add a number between 0–5 with decimal points) Culsines(Press ctrl to select multiple culsines) Aghan African African African (Rev) American (Traditional) Assen Fusion Barbeque B

Figure 1 interface of the UI

Figure 2 shows that this restaurant has passed the test.

### **Restaurant Hygiene Test**

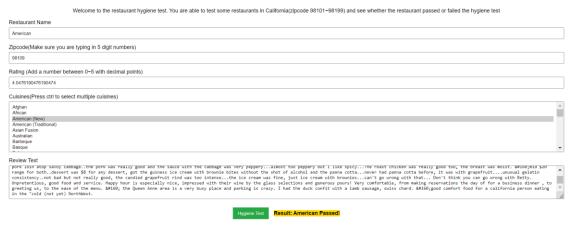


Figure 2 result passed

Figure 3 shows that this restaurant has failed the test.

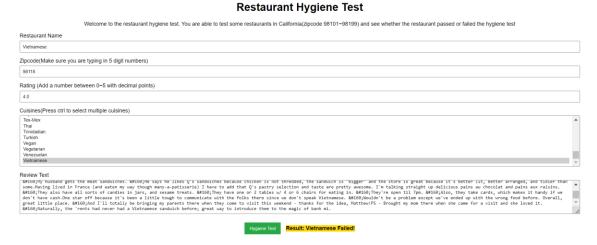


Figure 3 result failed

# 4. A working URL.

"http://ec2-43-203-173-84.ap-northeast-2.compute.amazonaws.com/"

I developed and deployed this system on a AWS EC2 instance. The public domain name of the instance is "ec2-43-203-173-84.ap-northeast-2.compute.amazonaws.com". The system can be remotely accessed from anywhere.

## 5. Novelty of the system

The novelty of this system is that it combined the numerical values and the textual data to construct the feature table for classification. It converted the textual data into a topic vector using topic modeling. It can be considered that using this kind of feature provides a better prediction than using numerical or text only.