

# 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)

Cuisines(Press ctrl to select multiple cuisines)  

Alghan  
African  
American (New)  
American (Traditional)  
Asian Fusion  
Australian  
Barbeque  
Basque

Review Text

Hygiene Test Result

Figure 1 interface of the UI

Figure 2 shows that this restaurant has passed the test.

**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)

Cuisines(Press ctrl to select multiple cuisines)

Afghan

African

American (New)

American (Traditional)

Asian Fusion

Australian

Barbeque

Basque

Review Text

pork loin atop saavy cabbage...the pork was really good and the sauce with the cabbage was very peppery...almost too peppery but I like spicy...the roast chicken was really good too, the breast was moist. \$160;his xlv range for both..dessert was \$8 for any dessert, got the guiness ice cream with brownie bites without the shot of alcohol and the panna cotta...never had panna cotta before, it was with grapefruit....unusual gelatin consistency...not bad but not really good, the candied grapefruit rind was too intense...the ice cream was fine, just ice cream with brownies...can't go wrong with that... Don't think you can go wrong with Betty. Unpretentious, good food and service. Happy hour is especially nice, impressed with their wine by the glass selections and generous pours! Very comfortable, from making reservations the day of for a business dinner, to greeting us, to the ease of the menu. \$160; the Queen Anne area is a very busy place and parking is crazy. I had the duck confit with a lamb sausage, swiss chard. \$160;good comfort food for a california person eating in the "cold (not yet) Northwest.

Hygiene Test    **Result: American Passed!**

Figure 2 result passed

Figure 3 shows that this restaurant has failed the test.

**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)

Cuisines(Press ctrl to select multiple cuisines)

Tex-Mex

Thai

Trinidadian

Turkish

Vegan

Vegetarian

Venezuelan

Vietnamese

Review Text

\$160;my husband gets the meat sandwiches. \$160;he says he likes Q's sandwiches because chicken is not shredded, the sandwich is "bigger" and the store is great because it's better lit, better arranged, and tidier than some.Having lived in France (and eaten my way though many-a-patisserie) I have to add that Q's pastry selection and taste are pretty awesome. I'm talking straight up delicious pains au chocolat and pains aux raisins. \$160;They also have all sorts of candies in jars, and sesame treats. \$160;They have one or 2 tables w/ 4 or 6 chairs for eating in. \$160;They're open til 7pm. \$160;Also, they take cards, which makes it handy if we don't have cash.One star off because it's been a little tough to communicate with the folks there since we don't speak Vietnamese. \$160;wouldn't be a problem except we've ended up with the wrong food before. Overall, great little place. \$160;And I'll totally be bringing my parents there when they come to visit this weekend - thanks for the idea, Matthew!PS - Brought my mom there when she came for a visit and she loved it. \$160;naturally, the 'rents had never had a Vietnamese sandwich before; great way to introduce them to the magic of banh mi.

Hygiene Test    **Result: Vietnamese Failed!**

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.