

NLP Applications on Yelp Datasets

Chundi Guo Zhu Liang Xin Lu

April 2023

Team Background

- ▶ Economics department
- ▶ Basic level of programming skills in Python
- ▶ No experience in JavaScript

Table of Contents

1: Rating Prediction and Review Summary

2: Restaurant Recommender: An Interface by Python

Q&A

Project Objectives

Perform rating prediction and generate review summary based on customer reviews and rating.

- ▶ Text Preprocessing: `json`, `multiprocessing`, `pandas`, `nltk`, `pandas`, `re`, `string`
- ▶ Data Exploratory Analysis: `pandas`, `wordcloud`
- ▶ Model Training and Prediction: `scikit-learn`, `XGBoost`
- ▶ Sentimental Analysis: `matplotlib`, `pandas`, `scikit-learn`, `wordcloud`
- ▶ Summary Generator Interface: `IPython.display`, `ipywidgets`, `uszipcodes`

Main Contributors: Chundi Guo, Xin Lu

Results and Conclusion



(a) Most frequent words for 1 star



(b) Most frequent words for 5 star

Results and Conclusion

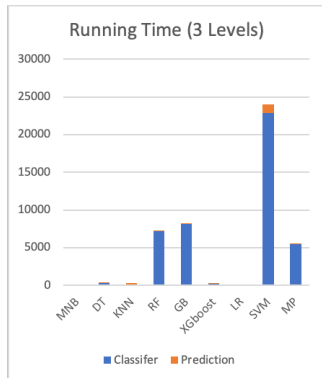


Table: Accuracy of 3-level star rating prediction

	Positive	Neutral	Negative	Weight Avg
MNB	0.92	0.41	0.73	0.83
DT	0.86	0.25	0.57	0.74
KNN	0.76	0.37	0.81	0.72
RF	0.81	0.67	0.83	0.8
GB	0.84	0.51	0.8	0.79
XGboost	0.87	0.51	0.79	0.81
LR	0.9	0.48	0.78	0.83
SVM	0.88	0.56	0.79	0.83
MP	0.91	0.39	0.74	0.82

Restaurant Review Summary Generator

Zipcode:

11790

Radius:

15

Category:

Seafood

Display:

☒ Positive Word Cloud

☐ Negative Word Cloud

☐ Bar Plot

Search

Oh! We have found 1 available choices for you!

The Chinese restaurant you are looking up now is:

China Szechwan,
1800 E Fort Lowell Rd,
Tucson,
AZ,
85719

The current rating for this restaurant is 4.0

The main categories that this restaurant belonging to are:

Restaurants, Chinese, Food, Szechuan



Table of Contents

1: Rating Prediction and Review Summary

2: Restaurant Recommender: An Interface by Python

Q&A

Project Objectives

Develop an interactive interface that recommends restaurants to users based on user-generated tip text.

1. Data processing, sentiment analysis, and database construction
2. Develop a reusable Python class for restaurant recommendations.
3. Build an interactive user interface (Jupyter Notebook and web page).

Main Contributor: Zhu Liang

Techniques and Tools

1. **Data processing, sentiment analysis, and database construction.**
 - ▶ pandas, NumPy, NLTK, re, string, TextBlob
2. **Develop a reusable Python class.**
 - ▶ uszipcode, urllib, random
 - ▶ Python Class and Methods, Standalone Script
3. **Build an interactive user interface.**
 - ▶ Notebook: IPython.display, ipywidgets
 - ▶ Web page: Flask
 - ▶ HTML, CSS, JavaScript

Results and Conclusion

- ▶ **Restaurant Recommender:** Users can get recommendations based on cuisine, zip code, and mile range.
- ▶ **Dine Dice:** Users can customize a pool of cuisines and receive a random recommendation.
- ▶ Seamless experience for users to discover new restaurants and make informed dining decisions.
- ▶ Flexibility and adaptability for diverse preferences and requirements.

Interface

Showcase: <http://127.0.0.1:5000/>

Restaurant Recommender

Cuisine: Zip Code: Mile Range: 10 miles

Search

[Not sure what to eat?](#)

[Fat Salmon \(★★★★☆\)](#)

719 Walnut St, Philadelphia, PA 19106

50 tips mentioned sushi

"Terrific. Best sushi in town. Great ambiance."

"Sushi wise - really fresh and tasty. Recommend."

"Fantastic food and a variety of sushi to choose from. Great atmosphere!"

[Bleu Sushi \(★★★★☆\)](#)

262 S 10th St, Philadelphia, PA 19107

25 tips mentioned sushi

"Sushi was okay for the price point. I would only recommend for byob+atmosphere"

"Really great prices for quality sushi"

"Great sushi and friendly service! A little too tight seating."

[Vic Sushi Bar \(★★★★\)](#)

2035 Sansom St, Philadelphia, PA 19103

22 tips mentioned sushi

"Pretty good sushi for a decent price. Small shop so it gets crowded best to go at a non high traffic time ie afternoon etc"

"Sushi is to die for. Amazing!"

"A nice walk to center city for a Saki and froyo lunch w my babe after an exam and abs class!!! What a beautiful day!"

Dine Dice

Step 1: Customize Your Pool

<input type="text" value="pizza"/>	<input type="text" value="sushi"/>	<input type="text" value="burger"/>	<input type="text" value="tacos"/>
<input type="text" value="ramen"/>	<input type="text" value="steak"/>	<input type="text" value="fried chicken"/>	<input type="text" value="pasta"/>
<input type="text" value="wings"/>	<input type="text" value="sashimi"/>	<input type="text" value="ribeye"/>	<input type="text" value="seafood"/>
<input type="text" value="hot pot; seafood"/>			

Create Pool

sushi, steak, fried chicken, hot pot, seafood

Step 2: Draw a Lot!

Zip Code: Mile Range: 10 miles

Draw

Let's try **steak**

[Pat's King of Steaks \(★★★\)](#)

1237 E Passyunk Ave, Philadelphia, PA 19147

83 tips mentioned steak

"See if there is an earlier flight I can fly standby? Or race into Philly to the birthplace of the Philly Cheesesteak for lunch? No brainer!"

"Good cheesesteak, classic Philly experience. I like Geno's better."

"Worth the drive. Now I know what a Philly Cheese Steak is suppose to be like."

[Go back to the main page](#)

Table of Contents

1: Rating Prediction and Review Summary

2: Restaurant Recommender: An Interface by Python

Q&A