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

Project Objectives

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

- Text Preprocessing: `json`, `multiprocesspandas`, `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

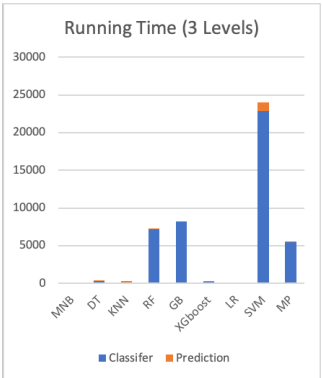


Table: Accuracy of 3-level star rating prediction

	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

Results and Conclusion

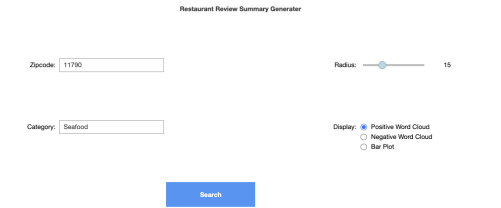


Table of Contents

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

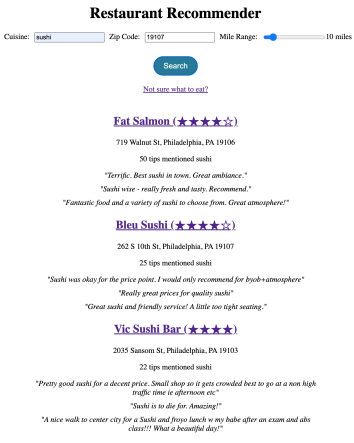
1. **Data processing, sentiment analysis, and database construction.**
   - `pandas`, `NumPy`, `NLTK`, `re`, `string`, `TextBlob`
2. **Develop a reusable Python class.**
   - `uszipcode`, `urllib`, `random`
   - Python Class and Methods, Standalone Script
3. **Build an interactive user interface.**
   - 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/`



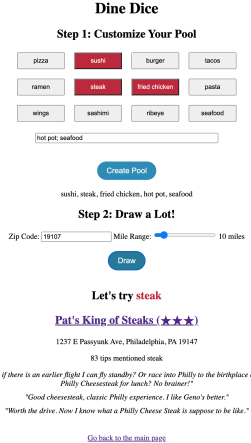


Table of Contents