

Restaurant Recommendation System based on Cuisine

Hemanth Pasupuleti
S20180010131
UG4 - CSE
IIIT Sri City
Andhra Pradesh, India
satyasaihemanth.p18@iiits.in

Suma Shreya T V
S20180010170
UG4 - CSE
IIIT Sri City
Andhra Pradesh, India
sumashreya.t18@iiits.in

Meda Rukmini
S20180010102
UG4 - CSE
IIIT Sri City
Andhra Pradesh, India
rukmini.m18@iiits.in

Abstract — Restaurant recommendation has been a widely experimented Natural Language Processing task. It has many commercial uses to customers as well as businesses to improve with the increase in digital information and usage. In this work, we create a system to recommend restaurants based on the user's preferred cuisine based on Yelp's restaurant reviews dataset. We pre-process the data by removing stopwords, punctuation and stemming by implementing Porter Stemmer. We predict ratings of Yelp reviews with a naive bayes classifier and obtain sentiment polarities. Top K (10) restaurants are then recommended based on the computed polarities. Instead of using direct restaurant ratings, the equipped sentiment analysis based sorting brings normalization among the different restaurants. An accuracy of 40% in predicting ratings for Yelp reviews through the Naive Bayes Classifier is observed and accurate top 10 restaurant recommendations are produced.

1. Introduction

Information Retrieval is one of the many interesting challenges in Natural language processing applications. One such application is the restaurant recommendation. With thousands of restaurants and several hundreds of thousands of reviews, recommending restaurants is an intimidating task. Data mining and more concretely text mining techniques allow us to explore a massive corpus. In this work, we intend to create an Information Retrieval model that recommends top restaurants for a specified cuisine.

1.1. Motivation

Today, customer reviews in social media have a large impact on the chances of success of any business. Yelp, a vast database of reviews, ratings, and general information provided by the community about businesses, provides consumers with a number of options and information even when searching for an especially specific service or goods niche. Yelp has many reviews of customers' experience regarding food quality, service, and ambience. Apart from that, it receives thousands of reviews each day from every corner of the world and summarizing or extracting specific pieces of information from such a big corpus is a challenging task. Reviews content is very diverse. They are a wealth of information and usually are more informative than a numeric rating. Reading all the reviews of a single business alone is time consuming and requires more effort than the average user is willing to expend. As a result, we believe users could greatly benefit from a recommendation system. Recommendation Systems have historically been created for various Machine Learning applications in numerous disciplines. So, we focus on creating a recommendation system for Yelp users in application to potential restaurant choices they could make. When customers look for a place to eat, they can ask the service for a list of nearby restaurants for a cuisine category.

2. State of the art

Several past works have focussed on restaurant reviews employing various kinds of natural language processing and data mining techniques. R. M. Gomathi et al.[1] proposed a machine learning algorithm to resolve the issue of personalized

restaurant selection relying upon tripadvisor.com search data. They proposed that users choose the hotel's features, the corresponding labels are fetched, user comments are examined and the highest rated hotel is recommended. Their evaluation results reveal that the proposed NLP algorithm improves the performance when compared to already existing algorithms. Elham Asani et al.[2] proposed a context-aware recommender system that extracts the food preferences of individuals from their comments and suggests restaurants in accordance with these preferences. Document and Sentence based opinion mining focus on overall priority of document and sentence respectively and do not describe the important aspects of each opinion which is more accurate. Hence, I.K.C.U Perera and H.A. Caldera[3] proposed an aspect based opinion mining algorithm on restaurant reviews. Gianni Barlacchi et al.[4] proposed an innovative approach for finding restaurants based on the dishes a user would like to taste rather than using the name of food facilities or their general categories. Yifan Gao et al.[5] proposed a method that first explores topic modelling to discover hidden aspects from review text, then create profiles for users and restaurants separately based on aspects discovered in their reviews and finally utilize regression models to detect the user - restaurant relationship. Jun Zeng et al.[6] proposed a restaurant recommender system in mobile environment by adopting a user preference model by using the features of user's visited restaurants and also utilizing the location information of user and restaurants to dynamically generate the recommendation results. Shreya Joshi et al.[7] presented a survey for a restaurant recommendation system by incorporating the narrowed down information of methods and data which are already being used for research in the domain. Various techniques and their advantages are also presented. Shoujin Wang et al.[8] provided a systematic review of Graph Learning Based Recommender Systems by discussing how they extract important knowledge from graph based representations to improve the accuracy, reliability and explainability of the recommendations. Shreya Joshi et al.[9] presented a k - means nearest neighbor and Naive Bayes classifier - based systems for recommendation of restaurants. Yancui Shi et al.[10] proposed the restaurant recommendation based on the

improved collaborative filtering method (ICFM) by analyzing the users' features.

3. Proposed System

3.1. Dataset

Yelp Open Dataset[11] is the primary dataset chosen. The Yelp dataset is a subset of Yelp's businesses, reviews, and user data that has been made publicly available for use for personal, educational, and academic purposes. The dataset contains 5,200,000+ reviews, 174,000+ businesses and 11 metropolitan areas. The dataset contains 6 csv files namely business.csv, business_hours.csv, reviews.csv, users.csv, tip.csv, checkin.csv. These files contain information about all business IDs, details about the weekly timings of businesses, user reviews for the businesses, list of all the users and their details, details about small compliments given by the users and mapping of business with date respectively in that order.

3.2. Pre-processing

There are a total of 1,74,567 businesses. Out of which, 1,46,702 businesses are open and 27,865 businesses are closed. We consider only the businesses that are open and consider the restaurants in Ontario (ON) state that have the highest number of food/restaurant related businesses..

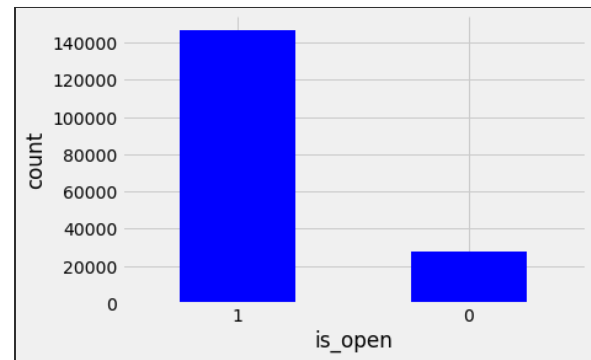


Fig1. No of open and closed businesses

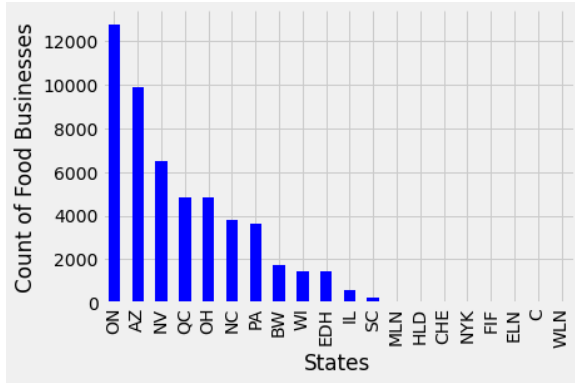


Fig 2. Count of food businesses or restaurants in each state

Out of over 100 different types of cuisines, we select the top 25 cuisines that are offered by more than 50 restaurants.

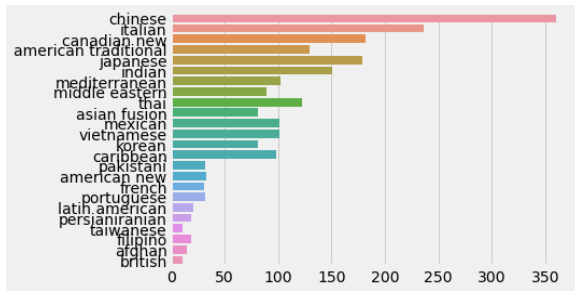


Fig 3. Count of each cuisine based businesses

The categories column in each business consists of a minimum of one category and a maximum of 10 categories. Thus, we have split the categories into category1, category2, etc for analysis.

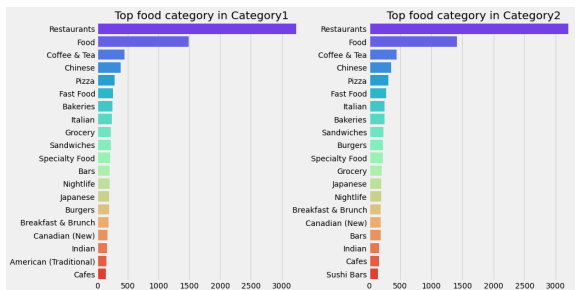


Fig 4. Count of different categories

Toronto city in Ontario has the highest number of food or restaurant related reviews.

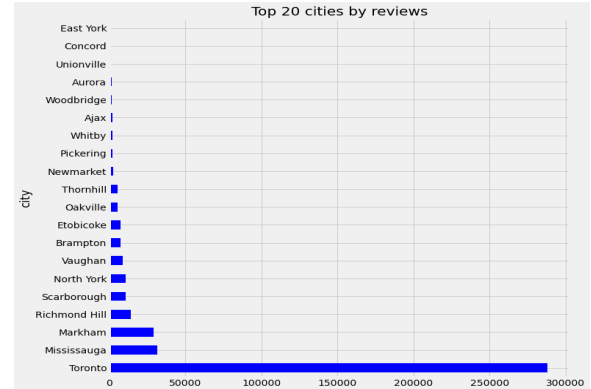


Fig 5. Count of user reviews in top 20 cities of ON (Ontario)

We convert all the text in the categories, reviews and cities to lowercase. Also, we pre-process the reviews by performing case normalization, removing punctuations, stopwords and returning cleaned text as a list of words.

```
y=preprocess('My hubby likes the apple pies there; crust is delicious.')
y
'hubby likes apple pies crust delicious'
```

Fig 6. Processed text output for a given string

3.3 Objective

The three main objectives that we have included in our project are Predict Ratings, Recommend Restaurants and Graphical User Interface. Firstly, the ratings of the restaurants are predicted based on the user reviews. Then, the restaurants are recommended based on the predicted ratings. To make the recommendation process more interactive, a user interface is built.

3.3.1 Ratings prediction

To predict the ratings of the reviews, we load the text of each review and stars rating into two different lists maintaining their corresponding index. We vectorize the complete data using CountVectorizer. We split it into train and test data. We use a Naive Bayes Classifier to classify the reviews into the five star ratings.

3.3.2 Restaurant recommendation

For each of the processed reviews, sentiment polarity values were computed. These were grouped by business_id and the mean of it was appended as a

new field in the dataframe. The polarity values range between -1.0 to 1.0. Then, the updated data frame is joined to the restaurant's data on business_id. Stars more than 3.5 are all distributed between the range 0 and 0.5 of sentiment polarity values. Consider only those business_ids that have values greater than 3.5 stars and polarity values greater than 0. This data can be used to recommend the restaurants to the users. A term to document index was maintained. Here, cuisine to a list of business_ids were stored. All the business_ids were sorted in descending order of their sentiment polarity values. Thus, given a cuisine (term), related restaurants were retrieved efficiently.

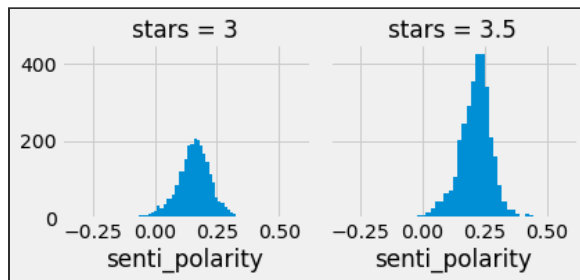


Fig7. Sentiment Polarity Vs Ratings

4. Results

The naive bayes classifier produced an accuracy of 40% over the data for predicting ratings of Yelp reviews.

	precision	recall	f1-score	support
1	1.00	0.44	0.62	9
2	0.00	0.00	0.00	6
3	0.24	0.18	0.21	22
4	0.41	0.74	0.53	62
5	0.40	0.12	0.18	52
accuracy			0.40	151
macro avg	0.41	0.30	0.31	151
weighted avg	0.40	0.40	0.35	151

fig8. Evaluation scores of Naive Bayes Classifier

Irrespective of the cuisine, the restaurant with maximum positive reviews (highest sentiment polarity values) is SunnySide Grill



Fig9. Highest Rated Restaurant

For Indian cuisine, Roti Palace is the best recommendation.

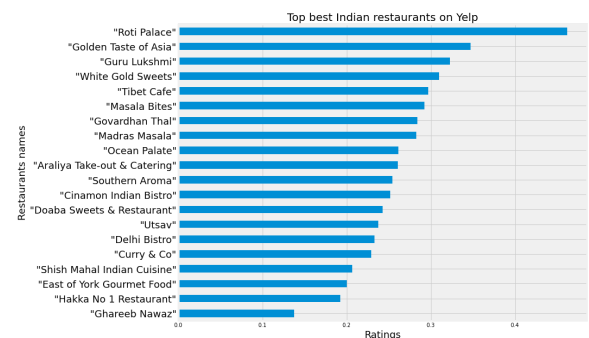


Fig10. Best restaurants for Indian cuisine

For Italian cuisine, Terra Cotta Inn is the best recommendation.

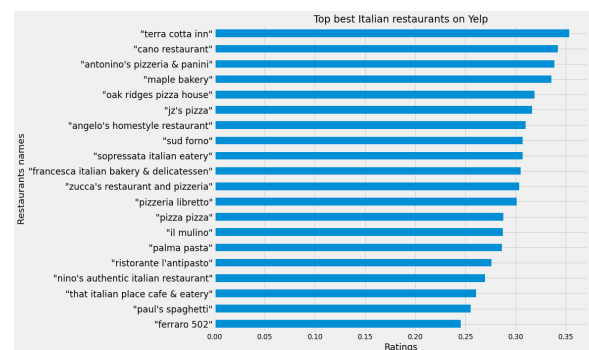


Fig11. Best restaurants for Italian cuisine

When a user enters a cuisine (Thai, here) top 10 restaurants are displayed in descending order of their

ratings.

	name	address	city	state	senti_polarity
100	"bangkok saigon noodle"	"136 holland street e, unit 1"	bradford	on	0.550298
4635	"thai wok n roll"	"30 broadleaf ave"	whitby	on	0.480556
3881	"thai cuisine experts"	"6970 lisgar drive, unit 2a"	mississauga	on	0.416845
3902	"bolton thai cuisine"	"18 king street e, unit 15"	caledon	on	0.396744
1887	"feng wok 'n' roll"	"28 dundas street e, unit 4"	mississauga	on	0.383282
4549	"thai express"	"178-200 windflower gate"	vaughan	on	0.378685
2766	"silk restaurant"	"446 parliament street"	toronto	on	0.359119
1530	"pho tai loi"	"1550 s gateway road, suite 1b"	mississauga	on	0.359035
538	"river tai restaurant"	"92 harbord street"	toronto	on	0.358654
4234	"bach yen"	"738 gerrard street e"	toronto	on	0.343499

Fig12. Top 10 restaurants for Thai cuisine

5. Conclusion & Future Work

We have developed a cuisine focused restaurant retrieval system. Ranking of the restaurants was done based on the pre-computed sentiment polarities on the reviews of that restaurant. Instead of using direct restaurant ratings, the equipped sentiment analysis based sorting brings normalization among the different restaurants. However, the model only focuses on cuisines and doesn't scale well over very large restaurant corpus. In future work, opinion mining can be employed targeting only specific reviews to enhance the performance of the model. Furthermore, graph based approaches can be experimented to scale over large corpus.

6. References

[1] R. M. Gomathi, P. Ajitha, G. H. S. Krishna and I. H. Pranay, "Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862048.

[2] Elham Asani, Hamed Vahdat-Nejad, Javad Sadri, Restaurant recommender system based on sentiment analysis, Machine Learning with Applications, Volume 6, 2021, 100114, ISSN 2666-8270, <https://doi.org/10.1016/j.mlwa.2021.100114>. (<https://www.sciencedirect.com/science/article/pii/S2666827021000578>)

[3] I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," 2017 2nd IEEE International Conference on Computational

Intelligence and Applications (ICCIA), 2017, pp. 542-546, doi: 10.1109/CIAPP.2017.8167276.

[4] Gao Y., Yu W., Chao P., Zhang R., Zhou A., Yang X. (2015) A Restaurant Recommendation System by Analyzing Ratings and Aspects in Reviews. In: Renz M., Shahabi C., Zhou X., Cheema M. (eds) Database Systems for Advanced Applications. DASFAA 2015. Lecture Notes in Computer Science, vol 9050. Springer, Cham. https://doi.org/10.1007/978-3-319-18123-3_33

[5] Barlacchi, Gianni et al. "Appetitoso: A Search Engine for Restaurant Retrieval based on Dishes." *CLIC-it/EVALITA* (2016).

[6] J. Zeng, F. Li, H. Liu, J. Wen and S. Hirokawa, "A Restaurant Recommender System Based on User Preference and Location in Mobile Environment," 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 2016, pp. 55-60, doi: 10.1109/IIAI-AAI.2016.126.

[7] <http://www.ijctm.in/Admin/upload/IJCTM-07-10-2018-1201872-sherya.pdf>

[8] Wang, Shoujin & Hu, Liang & Wang, Yan & He, Xiangnan & Sheng, Quan & Orgun, Mehmet & Cao, Longbing & Ricci, Francesco & Yu, Philip. (2021). Graph Learning based Recommender Systems: A Review. 10.24963/ijcai.2021/630.

[9] Joshi, Shreya and Jigyasu Dubey. "Restaurant Recommendation System Based on Novel Approach Using K-Means and Naïve Bayes Classifiers." (2020).

[10] Y. Shi, Q. Zhao, Y. Wang and J. Cao, "An Improved Restaurant Recommendation Algorithm Based on User's Multiple Features," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), 2018, pp. 191-194, doi: 10.1109/ICSESS.2018.8663850.

[11] <https://www.yelp.com/dataset>