

Restaurant Recommendation System

Team Members

1. Bharath Reddy Sarabuddula - 70073991
2. Shashaank Reddy Dondapaati – 700740721
3. Jyothi Sai Rajesh Kunapareddy – 700742325

Abstract

Due to the obvious usage of recommendation algorithms, users get access to sets of services that are more suited to their individual needs. They are mainly intended to deliver recommendations or ideas (such as restaurants or destinations) that are in line with the user's priorities and may be applied in a number of settings because of this design. It is feasible to improve the overall quality of recommendation systems as well as the level of service they give and to correct any problems that may be linked with them by making use of a broad variety of effective tactics that are connected to data management. Users are making effective use of the hotel's services and offering feedback that can be considered beneficial. The objective of ingesting natural language processing (NLP) is to investigate and categorize all previous users' remarks (both positive and negative) for each hotel, then calculate the overall proportion of comments, and finally

save the outcome. Following the user's selection of the hotel's attributes according to his interests, the matching hotels are retrieved; following that, the user comments are analyzed to decide which hotel has the largest ranking; and last, the restaurant is suggested. There are a great number of restaurants in the city of Dhaka. There are times when the majority of us find ourselves in a situation in which we go to a restaurant, place an order for some food, and find that either the quality of the cuisine is subpar or the cost of the meal is exceptionally expensive in comparison to the prices at other restaurants. Apart from this, one of the most significant issues is that the restaurant is located in an inconvenient area. We are unable to locate the local establishments that best suit our interests for dining out. A personalized restaurant suggestion system that is based on the user's current location and is incorporated with mobile technology. It continuously analyzes, through the use of a Machine Learning algorithm, the user's pattern of behavior when visiting restaurants.

Keywords: Restaurant, recommendation, Machine Learning, Pattern, Behavior

Introduction

Which product, movie, vacation package, restaurant, or book should I choose? is a question most consumers have at some point. Recommendation Systems were created to provide an answer to this issue by providing the user with the most relevant and useful options. By factoring in the activities and preferences of the user's social circle, Recommendation Systems improve the user's decision-making process. In this research, we use Sentiment Analysis methods to the input data in an effort to improve the recommendation process. The field of Sentiment Analysis is concerned with determining the overall tone of a piece of data and determining whether it is positive, negative, or neutral. Findings from Sentiment Analysis projects may be used to establish social patterns, measure the popularity of products, and modify services in response to customers' wants and requirements. To determine the most helpful recommendations for a user, the suggested method integrates Sentiment Analysis and Recommendation Systems. Textual reviews of restaurants may be sorted into good and negative categories using Sentiment Analysis. To build a list of the top n restaurants, a recommendation system uses the Sentiment Analysis task's output and the collaborative filtering algorithm to make predictions about the rating of yet-to-be-visited establishments. When compared to the outcomes achieved when ignoring Sentiment Analysis during the recommendation process, this method fared better. One such city is Bengaluru, India, which is home to more than 12,000 eateries selling cuisine from across the world. The business isn't yet saturated, even though new eateries appear every day, and demand is rising steadily. Despite the growth in the

dining-out market, new establishments are finding it harder to compete with more established ones. The majority of them provide exactly the same menu. As India's technological hub, Bengaluru is an important economic and cultural center. Because of their busy schedules, most residents rely heavily on takeout and restaurant fare. Due to the high demand for eating establishments, research about the local population is now crucial. The city of Bangalore's data will come from a Zomato data collection that we'll use. The goal is to develop a content-based recommender system that, for example, when we type the name of a restaurant, the System will look at user evaluations of other restaurants to determine which ones are most like the one we typed and then rank them accordingly.

Motivation

We are continually impressed by the restaurants' ability to maintain their high standards in the face of such innovative rivalry. And what considerations there are to make while starting a restaurant. Does the population make a difference? Can the demographics of a neighborhood have any bearing on the type of eatery that opens there? Is it significant that the restaurant has a certain theme? Is one restaurant in the food chain category more likely to be successful than another? Exists a comparable area somewhere else? Does the similarity between two neighborhoods indicate that their residents share a common ancestor, that they belong to the same social group, or that they are in fact geographically close to one another? What kinds of dishes are the most well-liked among the locals. It seems like everyone in this area favors vegetarian cuisine. Is the neighborhood predominantly comprised of Jain, Marwari, or Gujarati

Github link : <https://github.com/Rk-oo7/ML--Project.git>

vegetarians, for example? Since Bengaluru is one of these cities, with more than 12,000 restaurants serving dishes from all over the world, the primary goal of analyzing the Zomato dataset is to get a fair idea about the factors affecting the establishment of various types of restaurants at various locations in Bengaluru, as well as an aggregate rating of each restaurant. New restaurants are opening every day, so the market isn't yet saturated, and the demand keeps growing. However, new restaurants are having a harder time breaking into the market than ever before, despite the fact that there is a higher need for their services. The menus are largely similar at these establishments. Bengaluru is India's technological hub. Because of their busy schedules, most residents rely heavily on takeout and restaurant fare. Because of the high demand for eating establishments, it is now vital to investigate the demographics of a region.

Main Contributions and Objectives

1. Examine and investigate the dataset.
2. The representation of the datasets.
3. Rating prediction of the restaurant.
4. A sentiment analysis of the dataset's reviews
5. System for recommending restaurants using keywords

Related Work

Uzma Fasahteet.al, the hybrid recommendation method, which makes use of user evaluations in the form of text as well as rating data, is a strategy that is proposed in order to investigate the behavior of an individual user. The Recommendation System is now being employed by the hotel business. This system reveals the traveler's feelings with regard to the hotel by mining traveler's reviews, which further assists in assessing user's preferences.

Jun Zeng et.al, a recommendation is made to make use of the mobile environment for the restaurant recommender system. The system incorporates a user preference model by taking into account the number of restaurants that the user has been to in addition to the location details of both the user and the restaurant. As a result, the system is able to produce recommendation results in a manner that is both dynamic and personalized. The proposed restaurant recommender system made use of BMCS and BWCS, and the results of the case study demonstrated that the system was able to successfully take into account the user's preferences.

Ling Li et.al, contains three alterations that are suggested to be made to the conventional UCF algorithm. The accuracy of the UCF algorithms was quite poor due to the fact that the user's preference for a restaurant was impacted by a large number of characteristics. In conclusion, the real private information of registered internet users are being used in order to determine the degree of similarity connected to the

Github link : <https://github.com/Rk-oo7/ML--Project.git>

user's qualities. The results make it abundantly evident that the ACF-modified method improves the accuracy of the similarity computation, providing the customer with an exceptionally exact restaurant suggestion.

Nanthaphat Koetphrom et.al, a method for predicting the customer satisfaction rating that is based on actual information (connected to customer/restaurants traits) and similarities in the preferences of customers is presented as a potential solution. The three strategies for filtering information that are now under consideration are content-based, collaborative, and hybrid. The results indicate that the hybrid filtering method is superior to both the content-based filtering method and the collaborative filtering method. The hybrid filtering method makes use of regression, while the collaborative filtering method employs cluster-based methodology to control information among its peers.

Md. Ahsan Habib et.al, proposes a novel location, preference and time-based restaurant recommendation system by employing users' existing geospatial location. The technique analyses individual users' check-in accounts to explore his visiting patterns, food priorities, and restaurants popularity. For measuring recommendation scores four major factors are employed, these are: 1.) score of user's preference 2) restaurants distance 3) the time of a day, and 4) restaurant's popularity scores. Open dataset is being employed to exhibit the recommendation technique.

Github link : <https://github.com/Rk-oo7/ML--Project.git>

Proposed Framework

A recommendation system will use the results of the Sentiment Analysis job as well as the collaborative filtering algorithm to produce predictions about the ratings of restaurants that have not yet been visited. These predictions will then be used to compile a list of the top n restaurants. This strategy did better in comparison to the outcomes attained while neglecting Sentiment Analysis during the process of making recommendations, as compared to the outcomes. One of these cities is Bengaluru, which is located in India and is home to over 12,000 restaurants serving food from all over the world. Despite the fact that new restaurants open their doors each and every day and demand is consistently on the rise, the market is not yet considered to be saturated. In spite of the expansion of the market for eating out, new businesses are finding it increasingly difficult to compete with those that have been there for longer. The bulk of them provide precisely the same food selections on their menus. The majority of locals have such hectic schedules that they frequently rely on food that may be taken out or purchased from restaurants. The study on the local population is becoming increasingly important as a result of the growing demand for dining places. The data for the city of Bangalore will be obtained via a Zomato data collection, which will be utilized by us. The objective is to design a content-based recommender system in which, for instance, when we type the name of a restaurant, the System will look at user evaluations of other restaurants to determine which restaurants are most similar to the one we typed in and then rank them in accordance with that similarity. Ingesting natural language processing (NLP) has as its primary purpose the

investigation and classification of all previous users' remarks (both good and bad) for each hotel, followed by the calculation of the total proportion of comments, and lastly the saving of the outcome. After the user has selected the qualities of the hotel in accordance with his interests, the matching hotels are obtained; next, the user comments are examined to determine which hotel has the highest ranking; and last, the restaurant is suggested. The city of Dhaka is home to a wide selection of different dining establishments. There are times when the vast majority of us find ourselves in a situation in which we go to a restaurant, place an order for some food, and then discover that either the quality of the cuisine is below average or the price of the meal is exceptionally expensive in comparison to the prices at other restaurants. Aside from this, one of the most major problems is that the restaurant is situated in a location that is inconvenient to its customers. We have been unable to discover any restaurants in the area that are a good fit for our preferences in terms of eating out.

Dataset

Primarily, researchers interested in the Zomato dataset hope to gain a better understanding of the factors that have contributed to the proliferation of eateries in different parts of Bengaluru, as well as an aggregate rating for each establishment. A wide variety of restaurants serving cuisines from across the world can be found in Bengaluru, which is home to more than 12,000. Even though new eateries are opening every day, demand is still growing, thus the hospitality industry is not yet at its saturation point. New eateries face stiff competition from those with longer track records despite the increased demand.

Github link : <https://github.com/Rk-oo7/ML--Project.git>

Common fare is offered at many of these establishments. Bengaluru is India's technological hub. Most people in this city don't bother cooking for themselves since they're too busy eating out. There's a lot of people out there that need to eat, so it's important to learn about the local population to see how you can best serve them.

Url: The URL of the restaurant

Address: The address of the restaurant

Online order: “Yes” if restaurant accepts online order else “No”

Book Table: Yes, if restaurant allows people to book the tables else “No”

Rate: Rating of the restaurant on the scale of 5.

Votes: No of Votes that the restaurant got.

Phone: Phone Number of the restaurant

Location: location of the restaurant

Restaurant Type: Type of restaurant like Café, Casual Dining, Quick Bites, Delivery, Mess, Pub, Bakery, Dessert Parlor, Food Truck, Sweet Shop, Beverage Shop, Lounge, Food Court, Microbrewery, Kiosk, Bar, Takeaway, Fine Dine.

Github link : <https://github.com/Rk-oo7/ML--Project.git>

Dish Liked: List of most liked dishes.

Cuisine: list of Cuisines served at restaurant

Review List: List of reviews.

Menu Items: List of menu items

Listed in Type: Type of restaurant listed in Zomato.

Listed in City: City of the restaurant listed in Zomato.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 17 columns):
url                    51717 non-null object
address               51717 non-null object
name                  51717 non-null object
online_order          51717 non-null object
book_table            51717 non-null object
rate                  43942 non-null object
votes                 51717 non-null int64
phone                 50509 non-null object
location              51696 non-null object
rest_type             51490 non-null object
dish_liked            23639 non-null object
cuisines              51672 non-null object
approx_cost(for two people) 51371 non-null object
reviews_list          51717 non-null object
menu_item             51717 non-null object
listed_in(type)       51717 non-null object
listed_in(city)       51717 non-null object
dtypes: int64(1), object(16)
memory usage: 6.7+ MB

```

Cleaning Data:

Dropping columns URL, dished liked and phone from the dataset.

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 43499 entries, 0 to 51716
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   address                               43499 non-null  object
1   name                                  43499 non-null  object
2   online_order                          43499 non-null  object
3   book_table                            43499 non-null  object
4   rate                                  43499 non-null  object
5   votes                                 43499 non-null  int64
6   location                              43499 non-null  object
7   rest_type                             43499 non-null  object
8   cuisines                              43499 non-null  object
9   approx_cost(for two people)           43499 non-null  object
10  reviews_list                          43499 non-null  object
11  menu_item                             43499 non-null  object
12  listed_in(type)                       43499 non-null  object
13  listed_in(city)                       43499 non-null  object
dtypes: int64(1), object(13)
memory usage: 5.0+ MB

```

Renaming the columns “approx cost for two people” as cost, “listed in type” as type and “listed in city” as “city”.

```

3         'listed_in(city)': 'city' })
4 zomato.columns

Index(['address', 'name', 'online_order', 'book_table', 'rate', 'votes',
      'location', 'rest_type', 'cuisines', 'cost', 'reviews_list',
      'menu_item', 'type', 'city'],
      dtype='object')

```

Changing cost from “object” and “Float64”.

Github link : <https://github.com/Rk-oo7/ML--Project.git>

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 43499 entries, 0 to 51716
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   address         43499 non-null  object
 1   name            43499 non-null  object
 2   online_order    43499 non-null  object
 3   book_table      43499 non-null  object
 4   rate            43499 non-null  object
 5   votes           43499 non-null  int64
 6   location        43499 non-null  object
 7   rest_type       43499 non-null  object
 8   cuisines        43499 non-null  object
 9   cost            43499 non-null  float64
10  reviews_list    43499 non-null  object
11  menu_item       43499 non-null  object
12  type            43499 non-null  object
13  city            43499 non-null  object
dtypes: float64(1), int64(1), object(12)
memory usage: 5.0+ MB

```

Analysis

Computing Mean of rating for each restaurant

Github link : <https://github.com/Rk-oo7/ML--Project.git>

	address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	type	city	Mean Rating
0	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.1	775	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.0	[(('Rated 4.0', 'RATED')\n A beautiful place to ...	[]	Buffet	Banashankari	4.118182
1	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.1	787	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.0	[(('Rated 4.0', 'RATED')\n Had been here for din...	[]	Buffet	Banashankari	4.100000
2	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.8	918	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.0	[(('Rated 3.0', 'RATED')\n Ambience is not that ...	[]	Buffet	Banashankari	3.800000
3	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	False	False	3.7	88	Banashankari	Quick Bites	South Indian, North Indian	300.0	[(('Rated 4.0', 'RATED')\n Great food and proper...	[]	Buffet	Banashankari	3.700000
4	10, 3rd Floor, Lakshmi Associales, Gandhi Baza...	Grand Village	False	False	3.8	166	Basavanagudi	Casual Dining	North Indian, Rajasthani	600.0	[(('Rated 4.0', 'RATED')\n Very good restaurant ...	[]	Buffet	Banashankari	3.800000

Preprocessing and analyzing of Review List:

There are different phases that are being done on the review list for sentiment analysis

Phase 1: Converting letters into lower case

Phase 2: Removing Punctuations

Phase 3: Removing stop words. The stop words are taken from the corpuses of natural language tool kit

the stop words are 'couldn't', 'were', 'y', 'you're', 'don't', 'of', 'mightn', 'their', 'have', 'you'd'.

Phase 4: Removing any urls if it has.

Before preprocessing text:

Github link : <https://github.com/Rk-oo7/ML--Project.git>

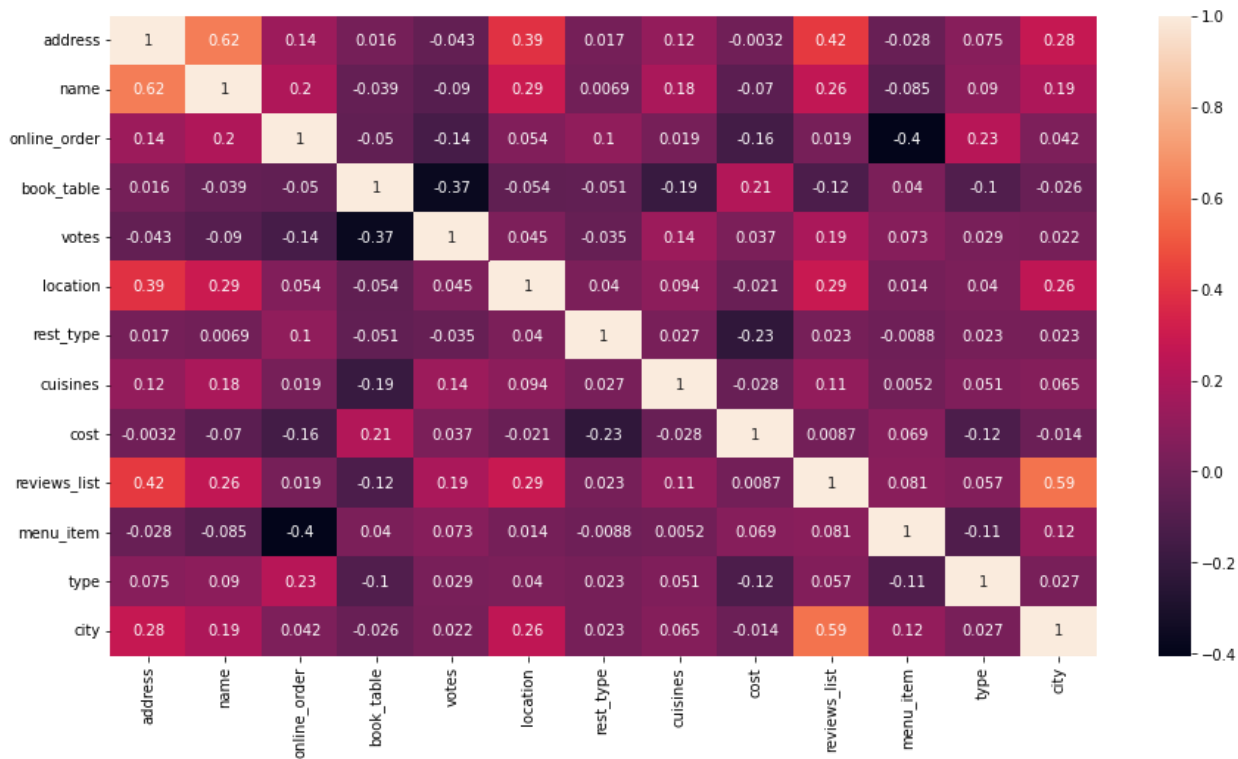
	reviews_list	cuisines
37033	['Rated 4.0', 'RATED\n Ordered anjal fish ta...	North Indian, Chinese, Seafood
16902	['Rated 4.0', 'RATED\n Initially I didn't wa...	Konkan, Mangalorean, North Indian, Seafood
2362	['Rated 5.0', 'RATED\n From the dark Caves, ...	North Indian, Afghani, Mughlai
6066	['Rated 4.0', 'RATED\n Sabudana khichdi is j...	North Indian, Beverages, Fast Food, Street Food
17945	['Rated 5.0', 'RATED\n For the first for a c...	Fast Food, Beverages

After preprocessing text:

	reviews_list	cuisines
34989	rated 40 ratedn 1q1 had been on my must visit...	Asian, Japanese, Thai, Malaysian, Vietnamese, ...
26513	rated 40 ratedn i have gone to this place mul...	North Indian, Chinese
22023	rated 50 ratedn good tasty food and suitable ...	Biryani, South Indian, Fast Food, Rolls, Kebab...
40267	rated 40 ratedn good quality shawarma the sha...	Arabian, North Indian, Biryani
34510	rated 40 ratedn had been here for lunch with ...	Chinese, Momos

Analyzing Correlation Between Different Variables:

Github link : <https://github.com/Rk-oo7/ML--Project.git>



Rating Prediction:

Using Extra Tree Regressor:

We used Extra Tree Regressor from sci kit-learn ensemble with estimators equal to 100

The r2 score for the Extra Tree Regressor for prediction of rating is 94%

Github link : <https://github.com/Rk-oo7/ML--Project.git>

```
1 from sklearn.ensemble import ExtraTreesRegressor
2 ETree=ExtraTreesRegressor(n_estimators = 100)
3 ETree.fit(x_train,y_train)
4 y_predict=ETree.predict(x_test)
5 from sklearn.metrics import r2_score
6 r2_score(y_test,y_predict)]
```

0.9393181826949868

Sentiment analysis

The sentiment analysis is done by using TfidfVectorizer, all the reviews are converted to lower case first and removing of stop words. The related restaurants will be found using cosine similarities.

1 df_percent[df_percent.index == 'Vihar'].head()

	name	address	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	type	city	Mean Rating
Pai Vihar	16/A, Ground Floor, Kkmp Building, Vasanth Nag...		False	False	2.8	56	Vasanth Nagar	Quick Bites	South Indian, Street Food, Chinese, Fast Food	400.0	rated 30 ratedn 12 rate hereâx83ax83				

Github link : <https://github.com/Rk-oo7/ML--Project.git>

References

- [1] Uzma Fasahte, Deeksha Gambhir, Mrunal Merulingkar, Aditi Monde, Prof. Amruta Pokhare, “Hotel Recommendation System”, Imperial Journal of Interdisciplinary Research (IJIR), Vol. 3, Issue. 11, pp. 318–324, 2017.
- [2] Jun Zeng, Feng Li, Haiyang Liu, Junhao Wen, Sachio Hirokawa, “A Restaurant Recommender System Based on User Preference and Location in Mobile Environment”, 5th IIAI International Congress on Advanced Applied Informatics, IEEE, pp. 55-60, 2016.
- [3] Ling Li, Ya, Zhou, Han Xiong, Cailin Hu, Xiafei Wei, “Collaborative Filtering based on User Attributes and User Ratings for Restaurant Recommendation”, IEEE, pp. 2592-2596, 2017.
- [4] Nanthaphat Koetphrom, Panachai Charusangvittaya, Daricha Sutivong, “Comparing Filtering Techniques in Restaurant Recommendation System”, IEEE, pp. 46-51, 2018.
- [5] Ahsan Habib, Abdur Rakib, and Muhammad Abul Hasan, “Location, Time, and Preference Aware Restaurant Recommendation Method”, IEEE, 19th International Conference on Computer and Information Technology, pp. 315-319, 2016.
- [6] Sanjukta Saha, A. K. Santra, “Restaurant Rating Based on Textual Feedback”, IEEE, 2017.

- [7] Yu Mon Aye, Sint Sint Aung, “Sentimental Analysis for Reviews of Restaurant in Mynamar Text”, IEEE, pp. 321-325, 2017.
- [8] Guolei Yang, Andreas Zuffe, “Spatio-Temporal Site Recommendation” IEEE, 16th International Conference on Data Mining Workshops, pp.1173-1178, 2016.
- [9] Sonali R. Gandhi, Jaydeep Gheewala, “A Survey on recommendationsystem with collaborative filtering using big data”, IEEE, International Conference on Innovative Mechanisms for Industry Applications, pp.457- 460, 2017.
- [10] Ditdit Nugeraha Utama, Luqman Isyraqi Lazuardi, Hersy Ayu Qadrya, Bella Marisela Caroline, Tris Renanda, Atthiya Prima Sari, “Worth Eat: an Intelligent Application for Restaurant Recommendation based on Customer Preference” IEEE, Fifth International Conference on Information and Communication Technology, 2017.
- [11] Khushbu Jalan, Kiran Gawande, “Context-Aware Hotel Recommendation System based on Hybrid Approach to Mitigate ColdStart-Problem”, IEEE, Communication, Data Analytics and Soft Computing, pp. 2364-2369, 2017.
- [12] F.M Takbir Hossain, Ismail Hossain, Samia Nawshin, “Machine Learning Based Class Level Prediction of Restaurant Reviews”, IEEE, pp. 420-423, 2017.

- [13] Zhou Xing, Marzieh Parandehgheibi, Fei Xiao, Nilesh Kulkarni and uliot, “Content-based Recommendation for Podcast Audioitems using Natural Language Processing Techniques”, IEEE, pp.2378-2382, 2016.
- [14] Ding Xizojun, Iijima Junichi and Ho Sho. Unique Features of Mobile Commerce. In 4th Asian eBusiness Workshop, Journal of Electronic Science and Technology of China, University of Electronic Science and Technology of China, Vol 2, No 3, pp 205 - 210, Aug. 2004.)
- [15] Christos K. Georgiadis and Athanasios Manitsaris. Personalized Recommendations In Mobile Commerce. In 7th Hellenic European Conference on Computer Mathematics & its Applications, HERCMA 2005, Athens, Greece, September 2005.
- [16] Daniele Querciaxy, Neal Lathiaz, Francesco Calabresey, Giusy Di Lorenzoy and Jon Crowcroftx. Recommending Social Events fromMobile Phone Location Data. In Proceedings of ICDM. 2010, 971-976.
- [17] Raymond J. Mooney and Lorie Roy. Content-Based Book Recommending Using Learning for Text Categorization. In Proceedings of the SIGIR-99 Workshop on Recommender Systems: Algorithms and Evaluation, Berkeley, CA, August 1999.

- [18] Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, and John Zimmerman. I'm the Mayor of My House: Examining Why People Use foursquare - a Social-Driven Location Sharing Application. In CHI: Location Sharing, Vancouver, BC, Canada, 2011.
- [19] Mustansar Ali Ghazanfar and Adam Prugel-Bennett. An Improved Switching Hybrid Recommender System Using Naive Bayes Classifier and Collaborative Filtering. In IAENG International Conference on Data Mining and Applications, 17-19 March, 2010.
- [20] Wan-Shiou Yang, Hung-Chi Cheng and Jia-Ben Dia. A locationaware recommender system for mobile shopping environments. In Expert Systems with Applications 34, 2008.
- [21] Mu-Hsing Kuo, Liang-Chu Chen and Chien-Wen Liang. Building and evaluating a location-based service recommendation system.