Restaurant Recommendation System

Parth Shrivasatava
Department of CSE
Lovely Professional University

***Abstract*—** The collection of data has led to a new era of information. It is being used in many different and new technologies and efficient systems, and this is where Recommendation Systems also begins their influence. Recommendation Systems are a type of information filtering system as they improve the quality of search results and provide items that are more relevant to the search item /are related to the search history of the user or as requirement provided by user. They are active information filtering systems that personalize the information coming to a user based on their interests, relevance of the information, etc. They are widely used for recommending movies, articles, restaurants, places to visit, items to buy, etc. Recommendation systems are being enforced to offer a personalized set of services to users. They are basically built to produce recommendations or suggestions (like restaurants, places...) that comply with user concerns and can be applied to multiple fields. To enhance the quality and service of systems and resolve any issues related to them, various effective techniques linked to data management can be utilized. In the given paper, I've surveyed various recommendation systems and tried to find out the best way to combine and get the best results. The focus of the research work is to offer a list of recommended restaurants located in Bangalore and provide user-required reviews and cuisines. This will use NLP for sentiment analysis of reviews. The conclusion and results reveal quiet high accuracy using different techniques.

Keywords—Restaurant recommendation system, Sentimental analysis, Natural language processing, Machine learning, Amenities, Tagging, collaborative ¢lustering, electronic commerce, recommender

I. INTRODUCTION

Recommender systems were originally defined as systems in which "people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients" (Resnick & Varian, 1997). Today, however, the term has a broader meaning, describing any system that generates individualized recommendations or guides users in a personalized way to objects that are interesting or useful in a large array of options. Such systems are especially appealing in an environment where the volume of online information vastly surpasses an individual's ability to navigate it. Recommender systems are now integral to e-commerce platforms such as Amazon.com and CDNow (Schafer, Konstan & Riedl, 1999).

It is the criteria of "individualized" and "interesting and useful" that distinguish recommender systems from traditional information retrieval systems or search engines. While a search engine relies on "matching" by returning items ranked by the degree of query match, recommendation systems go further by incorporating relevance feedback and other methods to personalize results. Google, for instance, incorporates "authoritativeness" into its ranking (Brin & Page, 1998), which uses the recursive sum of authoritative links to a page to enhance the usefulness of results.

The research around recommender systems often emphasizes the importance of combining different recommendation techniques to optimize performance. Each known recommendation technique has specific strengths and weaknesses, leading researchers to explore hybrid systems that combine various methods. This paper examines different recommendation techniques by analyzing the data that supports them, as well as the algorithms that operate on this data, and evaluates various hybridization techniques. This analysis reveals several unexplored hybrid methods. A noteworthy improvement in performance was observed when combining collaborative filtering with a knowledge-based approach, which was demonstrated in the Entree system. Additionally, the semantic ratings available from the knowledge-based part of the system added a performance boost to the hybrid (Schafer, Konstan & Riedl, 1999).

Data mining, or Knowledge Discovery in Databases (KDD), is a developing discipline that identifies significant patterns within large datasets. This field integrates techniques from machine learning, artificial intelligence, databases, and statistics. The primary aim of data mining is to transform information from raw databases into a human-understandable form, aiding in decision-making. While the web is a vast source of valuable information, it also presents challenges, making the decision-making process more complex. Thus, filtering and personalizing information is essential, particularly in recommender systems, where information must be tailored to the user's preferences.

In this study, Data, including general restaurant details, comments, reviews, and ratings, is compiled within a database. Using Natural Language Processing (NLP), user comments are analyzed and tagged to highlight features and sentiments (positive or negative).

II.Literature Review

In the paper "Hotel Recommendation System," the authors propose a hybrid recommendation approach using users' reviews (in the form of text and rating data) to analyze user behavior. The recommender system aims to reveal traveler sentiments about hotels by mining reviews, which helps in examining user preferences【1】.

In their research, Jun Zeng et al. present a restaurant recommender system designed for mobile environments, incorporating user preference by analyzing restaurants visited by users along with location details. The system dynamically generates recommendations based on user preferences, employing techniques such as BMCS and BWCS, which were shown to effectively capture user preferences【2】.

Technique	Pluses	Minuses
Collaborative filtering (CF)	A. Can identify cross-genre niches. B. Domain knowledge not needed. C. Adaptive: quality improves over time. D. Implicit feedback sufficient	I. New user ramp-up problem J. New item ramp-up problem K. 'Gray sheep' problem L. Quality dependent on large historical data set. M. Stability vs. plasticity problem
Content-based (CN)	B, C, D	I, L, M
Demographic (DM)	A, B, C	I, K, L, M N. Must gather demographic information
Utility-based (UT)	E. No ramp-up required F. Sensitive to changes of preference G. Can include non-product features	O. User must input utility function P. Suggestion ability static (does not learn)
Knowledge-based (KB)	E, F, G H. Can map from user needs to products	P Q. Knowledge engineering required.

The authors of "Collaborative Filtering based on User Attributes and User Ratings for Restaurant Recommendation" propose three improvements to the traditional User Collaborative Filtering (UCF) algorithm. The UCF algorithm's accuracy was found to be lower due to various parameters influencing user preferences for restaurants. By utilizing actual private details of registered users, the ACF-modified algorithm enhances similarity computation, offering users more precise recommendations【3】.

In "Comparing Filtering Techniques in Restaurant Recommendation System," the authors explore a hybrid technique based on customer and restaurant attributes, incorporating customer satisfaction ratings. The proposed hybrid filtering surpasses content-based filtering through regression techniques and collaborative filtering via a cluster-based approach that manages information within groups【4】.

Md. Ahsan Habib et al. propose a location- and time-based restaurant recommendation method that leverages users' geospatial data, check-in history, food preferences, and restaurant popularity. The model scores recommendations by considering user preferences, restaurant distance, time of day, and restaurant popularity. The approach demonstrates the system's effectiveness using an open dataset to validate its recommendation technique【5】.

In the paper "Restaurant Rating Based on Textual Feedback," Sanjukta Saha et al. focus on using textual feedback for restaurant ratings in Kolkata. The recommender system first evaluates the user's sentiment and computes sentiment scores for specific food items. Collaborative filtering is then applied, enhancing accuracy by basing recommendations on both textual sentiment analysis and rating data【6】.

In "Sentimental Analysis for Reviews of Restaurant in Myanmar Text," authors develop a Myanmar sentiment lexicon to analyze restaurant and food reviews. The system interprets data from 500 Myanmar language reviews with a high accuracy rate

of 96%, emphasizing the lexicon-based approach for precise recommendation outcomes 【7】.

Guolei Yang et al. introduce a recommender system that mines user-POI (Point of Interest) attributes by leveraging spatio-temporal trajectory data. The study demonstrates that attributes derived from spatio-temporal data can accurately predict POI popularity, supported by data from Foursquare 【8】.

The authors of "Context-Aware Hotel Recommendation System based on Hybrid Approach to Mitigate Cold-Start Problem" suggest a context-aware hybrid recommendation system to enhance accuracy in predicting hotel preferences. By integrating collaborative filtering with sentiment analysis, the system produces personalized hotel recommendations, improving user experience by adapting to their preferences and contexts 【11】.

Finally, Zhou Xing et al. propose a content-based recommendation model for podcasts using natural language processing (NLP) techniques. This model builds latent

embeddings for podcasts, capturing content similarity among items with text-based attributes. Cross-validation shows that this approach effectively predicts and classifies podcast content, providing users with accurate recommendations based on content similarity 【13】

III. DataSet

Dataset taken for this work is Zomato Bangalore Restaraunts from kaggle by ‘Himanshu poddar’. It’s main content is as follows:

“The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the establishment of different types of restaurants at various places in Bengaluru, and the aggregate rating of each restaurant. Bengaluru, being one such city, has more than 12,000 restaurants serving dishes from all over the world. With new restaurants opening every day, the industry hasn’t yet been saturated, and the demand is increasing day by day. Despite the rising demand, it has become difficult for new restaurants to compete with established ones, as many of them serve the same type of food. Bengaluru is known as the IT capital of India, and most people here rely on restaurant food, as they don’t have time to cook for themselves. Given the overwhelming demand for restaurants, it has therefore become essential to study the demographics of a location, such as what types of food are more popular in a locality. Does the entire locality prefer vegetarian food? If so, is that locality primarily populated by certain groups, such as Jains, Marwaris, or Gujaratis, who are mostly vegetarian? This type of analysis can be conducted using the data, by examining factors such as:

- Location of the restaurant
- Approximate price of food
- Whether the restaurant is theme-based or not
- Which locality in the city has the most restaurants serving a specific cuisine

- The needs of people seeking the best cuisine in the neighborhood
- Whether a particular neighborhood is known for a specific type of food

"Just so that you have a good meal the next time you step out."

The data is accurate as of what was available on the Zomato website up to 15 March 2019. It was scraped from Zomato in two phases. After analyzing the website structure, it was noted that each neighborhood has 6-7 categories of restaurants, including Buffet, Cafes, Delivery, Desserts, Dine-out, Drinks & nightlife, and Pubs and bars."[14]

Phase I:

In Phase I of extraction, only the URL, name, and address of the restaurants visible on the front page were extracted. The URLs for each restaurant on Zomato were recorded in a CSV file, so that the data could later be extracted individually for each

url contains the url of the restaurant in the zomato website	address contains the address of the restaurant in Bengaluru	name contains the name of the restaurant	online_order whether online ordering is available in the restaurant or not	book_table table book option available or not	ra cont ratin out c
51717 unique values	**11495** unique values	**8792** unique values	true 0 0% false 0 0%	true 0 0% false 0 0%	[null NEW Othe
https://www.zomato.com/bangalore/jalsa-banashankari?context=eyJzZSI6eyJlIjpbNTg2OTQsIjE4Mzc1NDc0Iiwi...	942, 21st Main Road, 2nd Stage, Banashankari, Bangalore	Jalsa	Yes	Yes	4.1/
https://www.zomato.com/bangalore/spice-elephant-banashankari?context=eyJzZSI6eyJlIjpbIjU4Njk0IiwxODM..	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th Block, Kathriguppe, 3rd Stage, Banashankari, Bangalore	Spice Elephant	Yes	No	4.1/
https://www.zomato.c nato-bangalore-restaurants/data	1112, Next to KIMS Medical College,	San Churro Cafe	Yes	No	3.8/

restaurant. This made the extraction process easier and reduced the load on my machine. Data for each neighborhood and each category can be found here.[14]

Phase II:

In Phase II, the recorded data for each restaurant and each category was read, and data for each restaurant was scraped individually. Fifteen variables were extracted in this phase. For each neighborhood and category, the following details were scraped: online_order, book_table, rate, votes, phone, location, rest_type, dish_liked, cuisines, approx_cost (for two people), reviews_list, and menu_item.[14]

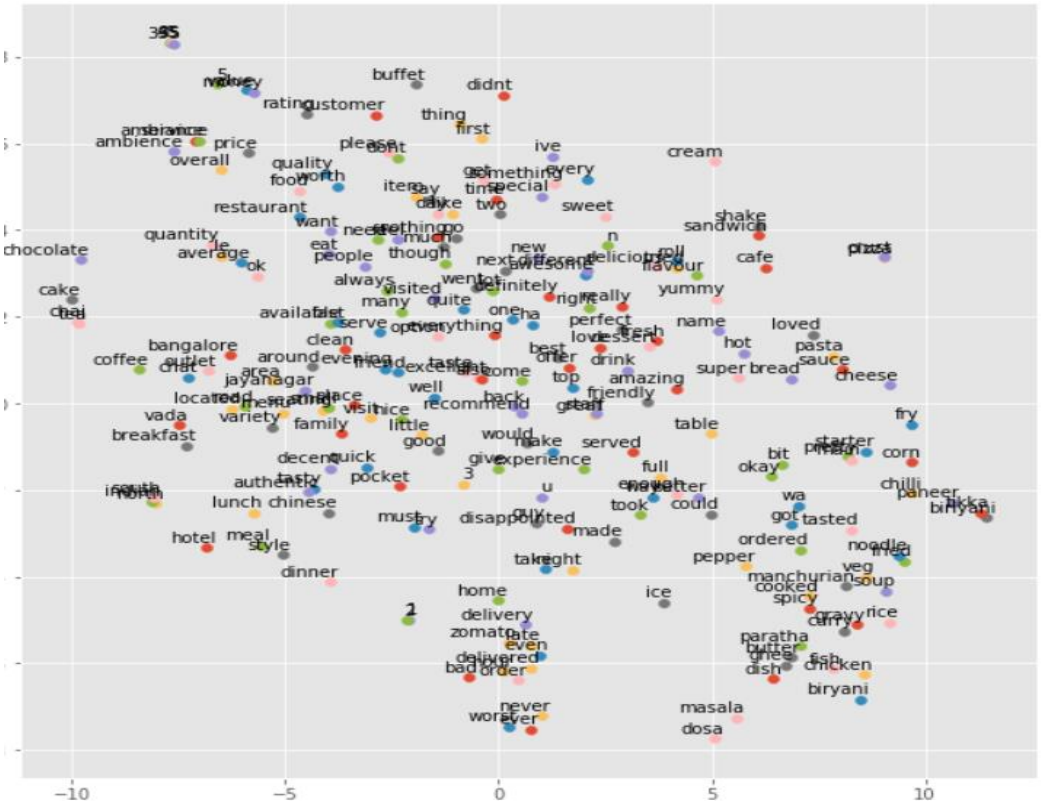
Acknowledgements

The data scraped was entirely for educational purposes only. Note that I don't claim any copyright for the data. All copyrights for the data is owned by Zomato Media Pvt. Ltd.

IV. **PROCEDURE AND METHODS**

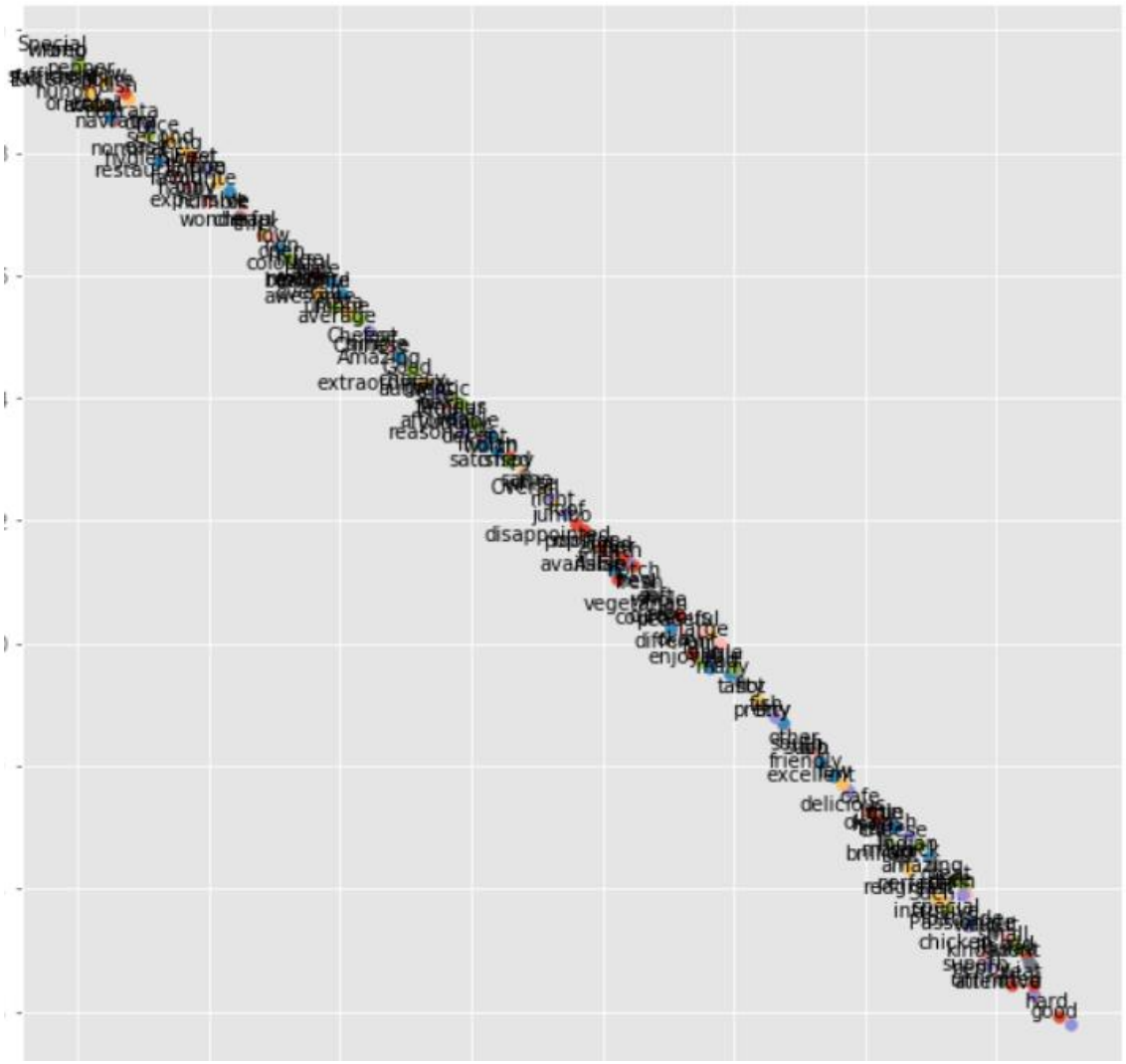
Sentimental Analysis

Sentiment can be termed as an attitude, outlook or by a user concerning to a specific entity. The amenities of the restaurant contain opinions these opinions are examined to discover the point of view of the user about the restaurant. In sentiment analysis user's preferences or opinions are being fetched, examined and inspected related to the subjective text. The analysis basically aims to parse the text and determine its polarity (whether neutral, positive or negative). It's also known as opinion mining as the users opinion is extracted which varies from person to person. Finally the reviews are categorized as positive and negative. The comments which say good thing about the restaurant considered as positive and bad comments as negative Sentiment analysis aids in comprehending users' reviews based on his perspective, this proves to be very significant in restaurant/hotel development approach.



Sentimental Analysis Using NLP

Sentiment Analysis is considered as a NLP-(natural language processing) approach that focuses on examining user's sentiments and preferences. The NLP processes are First tokenize the sentence into words.. Then use the POS tagging methods to tag each word of sentence of corpus as noun, verb, conjunction, number tag etc and then parsing is done to find the grammatical elements related to each other words and co reference resolution for pronoun identification of the sentence finally entity recognition is to understand which word refers to name, person etc. The tagged words are used to find the various aspects of the restaurant by Using NLP the free text is interpreted and made examinable. Abundant information is available in free text files.NLP analyzes user's feedback (whether neutral, positive or negative) thereby determining its polarity. Thereafter reviews are tagged by POS for identifying the restaurant's/hotel's positive and negative traits which are then stockpiled in the database. Sentiment analysis concerns with users sentiments, attitude, outlook and opinion regarding the restaurant.



Unsupervised- KMeans Clustering

KMeans clustering is unsupervised method of ML. Here I took 3 clusters , result provided by elbow method,and divided them on the basis of Location.

After doing all thse algorithms and getting their individual results they ewere combined to give final output.

V. **RESULT AND CONCLUSION**

The silhouette Score and Davies-bouldin score for the KMeans is 0.7442 and 0.3108 ,respectively. So the Clustering have given pretty good outputs .

In conclusion , The restaurant domain provides ample opportunity for suggesting dining choices to users based on their preferences, including historical data. Online travel site Tripadvisor.com holds substantial information linked to reviews and restaurant ratings. For some restaurants, users may leave negative reviews or low ratings, potentially damaging the restaurant's reputation, and the system's accuracy can suffer, producing unreliable recommendations. A machine learning algorithm is proposed to address the issue of personalized restaurant selection. In the proposed recommendation system, features provided by the restaurant and user reviews are analyzed and tagged with the help of Natural Language Processing (NLP). The technique yields effective results, improving functionality and providing precise outputs related to user preferences.

All existing recommender systems employ one or more of several fundamental techniques: content-based, collaborative, demographic, utility-based, and knowledge-based. A survey of these techniques shows that each has unique advantages and disadvantages. This reality has spurred research into hybrid recommender systems that combine methods for better performance. Significant recent research has focused on various hybrids, including six hybridization techniques discussed in this paper: weighted, mixed, switching, feature combination, feature augmentation, and meta-level. A survey of prominent research in the field shows that fewer than half of the 41 possible hybrid recommender combinations have been explored. Hence, This field have huge potential of exploration and can be used in various more fields than it is being used now.

Code : https://github.com/Parth-Shrivastava/ML_II-Project-Recommendation

VI. Citations and Refrences

[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 Zufle, “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 recommendation system 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 Chris Pouliot, “Content-based Recommendation for Podcast Audioitems using Natural Language Processing Techniques”, IEEE, pp. 2378-2382, 2016.

[14] https://www.kaggle.com/datasets/himanshupoddar/zomato-bangalore-restaurants/data