Applications of Recommendation Systems in Food Industry

Devansh Sharma - *E19CSE385*, Dhruv Devasthale - *E19CSE167*, Yedla Jaswanth Naidu - *E19CSE071*

***Abstract*—For whatever we buy or travel to a new place, the recommendation system is inescapable. Restaurants also require recommendation systems in order to attract more customers from the management side and to allow consumers to sample preferred, well-known dishes in the restaurant. Finding preferred and well-known foods, especially in a new region, is a difficult endeavor. We offer a restaurant recommendation system based on rating distribution and other (ambience, service etc) rating distribution calculated using the matrix density in this article. In addition, we created a popularity-based recommender model for customers to use when making restaurant recommendations. Our technique firstly distributes the data, then uses TF-IDF Vectorization to analyse. Finally, the model's output includes 10 recommendations for the most popular restaurants and meal items offered by the relevant restaurant.**

Modern recommendation systems/paradigms for restaurants are not very personalized and human in nature and rather are highly dependent on data and statistics, rather than feedback from unbiased customers.

***Index Terms*— Machine learning, content-based filtering, collaborative filtering, user reviews, recommendation**

I. INTRODUCTION

A recommendation system is a set of algorithms that learns from the input and after processing, it can recommend the needed restaurants/cafes to the customers. In essence, it is a program that recommends items to users based on data such as search history, user similarity, and similar behaviours, as well as ratings. YouTube, Amazon, Facebook, and other real-time popular examples come to mind. The majority of their operating processes are based on historical data. The available things are rated based on the data, and the most relevant are delivered to the customers.

A. Problem Statement

India's foodservice sector was valued at INR 423,000 crores in the financial year 2018-19, according to the NRAI[1] study of 20 cities. Mumbai's market was projected to be worth $40,880 crore, while Delhi's was worth $31,132 crore and Bangalore's was worth $20,014 crore. The unorganized market in Mumbai is projected to be around $26,601 crore, while the organized sector is at $14,279 crore. Approximately 28,000 of the 87,000 outlets are organized, accounting for only 32%.

According to Zomato's current algorithm, a will get eateries that are either in their immediate proximity or those they have previously ordered from. Hundreds more unorganized eateries, "dhabas," and local enterprises are not visible to users trying to make an order during this process.

Additionally, travelers and individuals visiting cities for work trips frequently use their food ordering applications to seek fast recommendations. Instead of getting suggested authentic and traditional, they generally show fast-food restaurants such as McDonald’s, KFC, BurgerKing, and others.

B. Motivation

When it comes to eating a meal, we frequently hang out with our family, friends, and coworkers. People worry more about how we will like a restaurant as users of recommendation apps. People used to get restaurant recommendations from their pals. Although this approach is simplistic and user-friendly, it has several significant drawbacks. To begin with, suggestions from friends or other ordinary people are restricted to restaurants/eateries they have already visited. As a result, the user is unable to learn about destinations that their friends have not visited. Aside from that, there is a risk that consumers will dislike the location that their friends have recommended.

Bangalore is one of these urban communities, with north of 95,000 eateries (NRAI, 2019) giving cooking styles from everywhere the world. With the lastest cafés sending off consistently, the area hasn't yet arrived at immersion, and requests are developing continuously. Aside from the enormous interest, the most recent cafés are finding it progressively challenging to stand separated from the laid out foundations.
Most of them give similar dishes. Since they lack opportunity and willpower to set up, most individuals here depend intensely on eatery cooking. With such a popularity for diners, it's more indispensable than any other time in recent memory to investigate an area's socioeconomics. We'll use an informational index from Zomato that incorporates Bangalore's information.

C. Background knowledge

Recommender systems are content-based filtering systems that aid with information overload by filtering and segmenting huge volumes of dynamically produced data and producing pieces based on the user's choices, interests, or observed behavior concerning a certain item or things. Based on the user's profile and previous information, a recommender system may anticipate whether a specific user would like an item or not.

A recommendation system is said to accelerate the standard and speed of decision-making. Recommender systems increase marketing income in large e-commerce scenarios since they are an efficient way of selling more items. Recommender systems in academic libraries assist and enable users to go beyond conventional catalog searches. Like a conclusion, the necessity of implementing efficient and reliable recommendation algorithms inside a system that gives relevant and trustworthy ideas to customers cannot be overstated.

Netflix and other media conglomerates utilize a recommendation engine to provide movie and show recommendations to their consumers. Amazon ando ther such up and coming ecommerce platforms use its recommendation engine to provide product suggestions to users. While one utilizes the other for somewhat different objectives, they both have the same overall goal: to increase sales, increase customer engagement, and provide more tailored customer experiences.

Recommendations usually speed up searches and make it simpler for consumers to obtain the information they've always wanted to see, as well as surprising them with many offerings they wouldn't have looked for. Sending out personalized emails with links to fresh offers that match the recipients' likes, or ideas of films and TV shows that suit their specific profiles, allows businesses to attract new consumers.

Here are collaborative filtering and content-based filtering, the two types of recommendation methods:

A. Collaborative filtering

Collaborative filtering procedure for proposal frameworks is a technique in the event that one needs to deliver new suggestions that are absolutely founded on past connections between the client base and things that have been recorded by the association. Cooperative Filtering attempts to figure out what practically identical clients need and give ideas for those clients, as well as sort clients into clubs of comparable sorts and propose every client in light of their gatherings inclinations.

The basic principle that drives collaborative techniques is that by analyzing prior user-item interactions, it becomes sufficient to recognize comparable people or similar things in order to generate predictions based on these estimated facts/insights.

These memory-based methods work directly with the values of recorded interactions or data and are based on nearest neighbors search.

The developed model works by presuming that there is an underlying "generative" insight that describes user-item interactions and then attempting to uncover it in order to produce new predictions.

Furthermore, without depending on feature engineering manually, the embeddings may be learned automatically.

The characteristics of the objects do not need to be provided for the collaborative filtering approach to work. A feature vector describes each user and object.

The Nearest Neighborhood algorithm is the standard approach adopted by Collaborative Filtering. Filtering may be done in a variety of ways, including user-based and item-based Collaborative Filtering.

The drawbacks of this method is that the model's prediction for a particular user is the dot product of the tags that are related. As a result, if an item isn't encountered during training, the system won't be able to generate an embedding for it, and therefore won't be able to query the model with it. The cold-start problem is the name for this issue.

B. Content-based filtering

Extra data on individuals as well as items is utilized in the substance based approach. This kind of separating utilizes thing highlights to recommend extra things that are equivalent to what the client likes, as well as earlier activities or unequivocal information. Assuming we take the instance of a film recommender framework, the additional data could incorporate the client's age, orientation, occupation, or some other individual data, as well as the classification, significant entertainers, length, or different highlights of the motion pictures (for example the things).

The fundamental goal of this approach is to try to create a model that explains observed user-item interactions using the given "features." We can train a model in such a way that it can explain why this is occurring while still taking into account consumers and movies. By merely checking into a user's profile and finding appropriate movies to recommend based on their information, we can swiftly produce new recommendations for them.

For Content-Based Methods, we can apply a Utility Matrix. A Utility Matrix can be used to represent a user's preference for specific objects. We can establish a relationship between the products that the user likes and those that the user dislikes using the data obtained from the user; for this, the utility matrix can be used to its full potential.

Each user-item combination is given a certain value, which is referred to as the degree of preference, and a matrix of the user is drawn with the relevant objects to show their preference relationship.

Since new individuals or items might be characterized by their characteristics, for example the substance, content-based methods seem to experience the ill effects of the virus start issue impressively not exactly cooperative frameworks, appropriate ideas can be made for these new elements.

Just new clients or things with beforehand obscure highlights will hypothetically experience the ill effects of this imperfection, yet after the framework has been adequately prepared, this is probably not going to happen.

It fundamentally expects to be that assuming an individual was keen on something previously, they would be keen on it again later on. Comparable products are many times arranged together in view of their qualities. Client profiles are made by examining past associations or by getting some information about their inclinations. There are extra frameworks that utilize client individual and social information that aren't viewed as rigorously satisfied based.

II. Related Work

There are several approaches for developing a restaurant recommendation system. The following are some of the existing systems and how they work. A user's decision to visit a restaurant is influenced by a number of aspects, including the restaurant's cuisine, its location, its atmosphere, its pricing range, its popularity, and its ratings.

They produced recommendations based on the user's choices in this recommender system. It was inspired by the discovery that varied characteristics expressed in reviews influence a user's inclination against an item.

The authors of other researchers have described the recommendation system with ML algorithms. Finally, regression models were used to determine the user-restaurant association.

Many algorithms have been tried to find the one that gives the most precise result. In most of these recommendation systems, user based and Item based collaborative algorithms are used.

The closest one to the existing recommendation system uses SVM to predict the location of the user. The authors of this research constructed an efficient recommendation engine for users in the form of a software application using a well-rounded, open source dataset supplied by Kaggle, which gives data not just of restaurant reviews, but also user-level information on their favoured restaurants.

In another paper, the authors had studied what and how the users use online reviews for a variety of reasons. Due to the large number of reviews it becomes difficult for the users to decide which one to adhere to.

As a result of past study, online review websites may be able to create a tailored review sorting system for each unique consumer.

One of the closest research to ours that the authors saw presented a content-filtering recommender system that assesses individual internet reviews and provides a numeric score to each review for each of the five consumer categories, based on five customer segments and ten restaurant features discussed in the review.

Individual consumers' online reviews can be sorted based on their restaurant preferences using the numeric scores.

With the growth in the total number of user comments and reviews on websites and social media, sentiment analysis has become a viable method for extracting their preferences.

Since clients express their needs and interests through remarks, acquiring their inclinations utilizing assessment mining and feeling examination could give more adaptable and dependable information than past techniques.

Subsequently, a few of the eatery's recommender frameworks dissect client remarks to decide their inclinations.

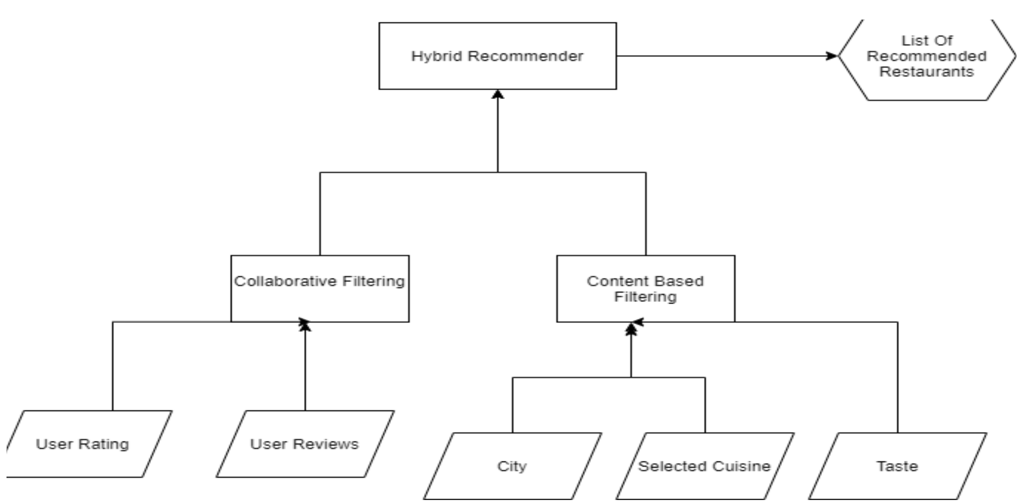
Analysis of the sort of environment that would fit a user, as well as other characteristics such as the type of food that can be provided and the quality of the cuisine, are critical requirements that may be examined and utilized to propose a restaurant to the user.

Meal plays an essential part in automatically rating restaurants based on various consumer reflections. To categorize the cuisine, the notion of document similarity metrics was used, as well as the user's perspective on the meal.

Swiggy, one of the most successful food ordering platforms, has been into developing Food Intelligence (FI) to have a better understanding of their catalog, which is relevant to their customers.

According to them, associating food with its distinctive characteristics enables personalization for someone with idiosyncratic and diverse taste preferences.

From the above reading, it is clear that numerous restaurant recommendation systems have been suggested, which assist users in selecting restaurants based on their interests. Restaurant recommendations are not adequate in today's environment. It is critical not only to define restaurants but also to determine the availability of dishes based on the mood.



Recommendations for items are by far, very different in the food domain when compared, in contrast to movie or book recommendations.

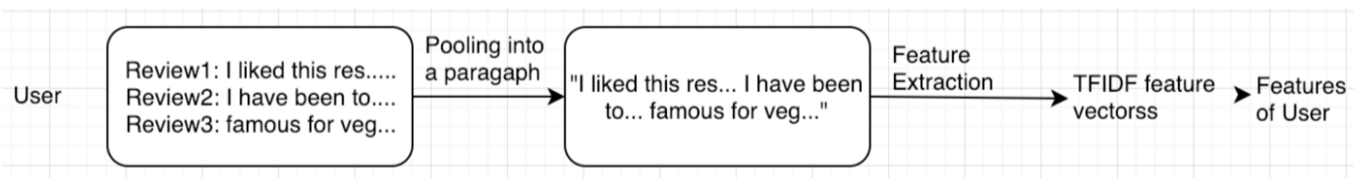
III. Proposed Methodology

The diagram above depicts the Model's entire operation. Both User Ratings and Reviews play a significant role. The Dataset is used to extract information about the city, cuisine, and taste.

For the execution, both collaborative filtering and content-based filtering approaches are used. The Hybrid Recommender that results is what generates the list of suggested eateries.

The recommendation system will also be context-aware. The user's preference is a crucial piece of contextual information that will be used in the recommendation system

We evaluate the input using the NLTK toolkit and then run the data through a Cosine Similarity test. The recommendations are based on the information that has been analyzed.



For the best results, Similar Ratings are combined with the default data of location, taste, and cuisine.

The interface of the entire application is simple. The user logs in. Basic preference about the cuisine type, taste, and the origin city of the user is entered.

With the help of responsive REST API, the data is relayed to the ML Model. The Model then works with the received data and provides the user with the restaurants similar to the users' preference.

The final output is extracted by the API and sent to the user for display.

A. Preprocessing

Before analysing and selecting our features for our algorithm, we performed a lot of pre-processing tasks on our acquired dataset. There were unwanted features such as restaurant URL, phone, dish_liked for which the columns were removed.

Then to ensure integrity, we removed duplicates and NaN and null values from the dataset. To make things simpler we renamed column names to proper snake_case. The feature 'cost' was transformed from double to string values and was cleansed of unwanted characters such as commas and hyphens.

The next step was to remove Stop words (articles, prepositions, punctuations, pronouns, conjunctions) from the dataset to give proper substance to our dataset.

B. *Distribution of Data*

After extensive research and analysis, the Zomato dataset for the city of Bangalore was chosen. The data has 12 distinct features and 51,717 entries.

The entire dataset is localized to the city of Bangalore, India. Before working on the data, it is important to extract all the necessary details, so as to understand the task and figure out the course of action.

Name, Location, Restaurant Type, Cuisines, Approximate Cost, and Reviews are primary features worth considering. All the features are self explanatory.

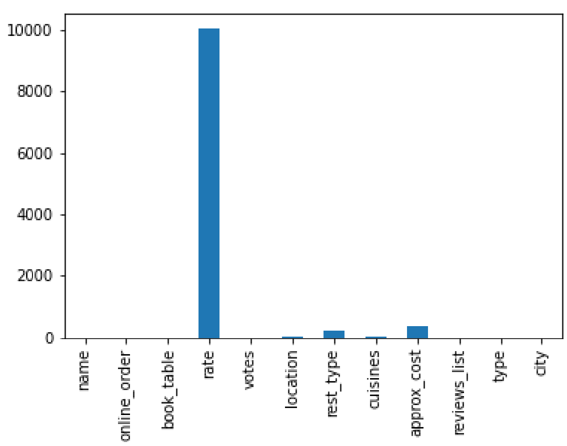
First step is to ensure the data is usable and clean. That is, to remove the redundant data and fill the missing values. Upon further analysis, 72 rows were eliminated due to redundancy.

```
df.drop_duplicates(inplace=True)
```

It was found that the features Rate, Location, Type, Cuisine and Cost had a few missing values. eliminating those columns is not feasible.

To tackle this problem, a different approach was utilized. The empty data points were replaced with either 'NaN 'or with the mean of all the data entries in that particular column. This aided in making the entire dataset systematic.

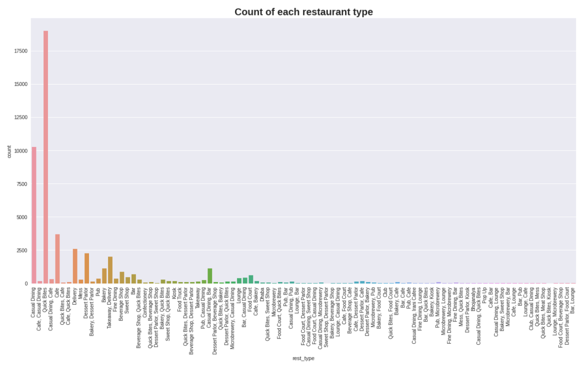
	rate	votes	Number_of_cuisines_offered
rate	1.00	0.43	0.19
votes	0.43	1.00	0.23
Number_of_cuisines_offered	0.19	0.23	1.00



The Rate feature had the maximum undefined values. The blanks were replaced with the mean of all the rates.

```
df['rate'].fillna(df['rate'].mean(),inplace=True)
```

After the cleaning, the data is prepared for the Exploratory Data Analysis. The data analysis revealed key information points.



It was discovered that the most famous restaurant chain was Cafe Coffee Day. The number of bookings available or not are also taken into account.

The restaurant recommendation is an amalgamation of all these factors. The Feedback, along with the Rating, and the Cuisine type are taken into account for the final recommendation. Top 10 restaurants are generated based on these factors.

```
df_new = pd.DataFrame(columns=['cuisines', 'Mean Rating', 'cost'])
```

C. *Approach and Algorithm*

Zomato Roulette used TF-IDF Vectorization to analyse and recommend restaurants. The algorithm creates a matrix where all columns have a specific word in the vocabulary. It's a process to procure the meaning of a set of characters in a given document/dataset.

With TF-IDF we used cosine similarities to find patterns and features of the same nature between two documents, in our case, two restaurants.

Mathematically, the division between the scalar product of numbers having direction and the product of the Euclidean norms or magnitude of each vector is viewed as the cosine similarity.

The generated TF-IDF matrices were used to compute the linear kernel, which is a unique case of the polynomial kernel with degree 1 and coeff 0, hence being homogeneous:

If x and y are column vectors, their linear kernel is:

$$k(x, y) = x^\top y$$

Initially, we created a user-item matrix that was obtained from ratings and reviews of the restaurants. We then used this matrix to calculate similarities.

Similarity calculations were carried out in 2 levels, one, calculating user-similarity, and two, user-feature similarity. Pearson correlation formula was helpful to obtain these calculations:

$$pearson(a,u) = \frac{\sum(r_{ai} - \bar{r_a})(r_{ui} - \bar{r_u})}{\sqrt{\sum(r_{ai} - \bar{r_a})^2 \sum(r_{ui} - \bar{r_u})^2}}$$

The last step was to extract restaurants having cosine-sim values similar to a restaurant with features based on the preferences selected by the user.

D. Recommendation

At this stage the application will provide the names of ten restaurants based on the results of the recommendation system.

IV. Result

The overall execution of the model results in recommending top 10 restaurants similar to the restaurant entered by the user. The recommendation is based on the cuisine, rating and the type of restaurant.

TOP 10 RESTAURANTS LIKE Pai Vihar WITH SIMILAR REVIEWS:

	cuisines	Mean Rating	cost
Burma Burma	Asian, Burmese	4.74	1.5
Nando'S	Portuguese, Wraps, Burger, Salad	4.13	1.2
Cravy Wings	American, Burger, Fast Food	4.11	400.0
Andhra Ruchulu	Andhra, North Indian	3.72	800.0
Cafe Aladdin	Cafe, Chinese	3.71	500.0
Green Pepper	Seafood, South Indian, Chinese, Kerala	3.65	600.0
3 Leafs	North Indian, South Indian, Chinese	3.45	600.0
Cafe @ Elanza	Chinese, North Indian, Cafe	3.45	1.0
Nys Kitchen	North Indian, Chinese	3.39	500.0
Chicken Corner	Biryani, North Indian, Fast Food, Chinese	3.35	400.0

The above results show that the restaurants have a similar rating and the cost as well. feedback of the users were also considered and the final outcome was generated.

V. Comparative Analysis

Previously existing sentiment analysis-based recommendation systems assess customers' views based on restaurant criteria such as service quality.

The authors have noticed and believe that the primary criteria of picking out a restaurant is based on its food[1]. For the extraction of names of the dishes, earlier recommendation applications[2] used generic approaches such as Bag of Word or the Term-Frequency (TF).

Repeating trends show the user’s preferences, however these applications also show all the elements that the consumer is uninterested in which is caused only due to repetition.

As a result, these systems' suggestion accuracy is poor. In this study, we employ content-based filtering to improve the extraction of user preferences.

Recommender systems attempt to determine a user's preferences in order to recommend a collection of products that best meet their priorities. Exact searches[3], star ratings[4], and user opinion analysis are all methods for extracting user preferences[5].

The client's inclinations are recovered utilizing inquiry based approaches toward the beginning of the login cycle by responding to specific inquiries as a survey. In such manner, when a client joins an eatery recommender framework, they are incited to choose their favored valuing and dinner type from a rundown of options.

This method of obtaining preferences has several downsides, as dietary preferences may conflict with the items on the static questionnaire.

Moreover, some extra recommender systems collect users' preferences and change recommendations based on the points that users have rated previously. Customers are classified according to the similarity of ratings they have already completed in certain restaurant studies[6].

Similarly, restaurants are categorized based on how well they match customer evaluations for quality, service, cuisine, and pricing. Depending on the restaurant group's resemblance to the user group, a variety of eateries are recommended to them.

Extracting users' preferences through content-based filtering has become possible as the number of user comments and reviews on websites and social networks has increased.

Since clients express their needs and interests through remarks, gathering their inclinations utilizing assessment mining and feeling examination could give more adaptable and solid information than prior strategies.

As a result, several of the restaurant's recommender systems use user comments to determine their preferences[7]. If the goal of the system is to discover preferences based on user feedback, sentiment analysis techniques might be used to assess the polarity of the feedback, or whether it is positive or negative.

Dictionary-based techniques and machine learning models are commonly employed to achieve this goal.

To our knowledge, no previous restaurant recommender systems have fully exploited the power of opinion mining and content-based filtering to derive user meal preferences.

A couple of studies use content-based sifting to find through how purchasers feel about the nature of suppers, as indicated by a new overview in the field of food science in light of text mining[8].

Furthermore, one article released in 2020 [9] utilizes group correlations and customer preferences on restaurant comments to determine the amount of consumer satisfaction with meal quality. Both of these investigations look into distinct aspects of the present paper.

The extraction of food preferences is done in this study by analyzing consumers' text reviews. The restaurant menu, on the other hand, is utilized to extract the characteristics of each restaurant.

Moreover, prior user comments on restaurants are used as background information, and establishments with a sentiment analysis score for quality features and services below a threshold are eliminated from the list of suggestions.

Zomato Roulette explains and visualises the restaurant recommendation system with various ML techniques. They attempted numerous collaborative filtering strategies to anticipate evaluations between restaurants and users in order to create a solid machine-learning model.

Slope One, multiclass SVM classification and the k-Nearest Neighbors algorithm are the approaches they used. The TF-IDF Vectorization approach surpasses the other methods in our tests.

VI. Conclusion

The main objective of the study is to develop the restaurant recommendation system using machine learning with the web interface that can act as an application for the customers.

Zomato Roulette is a user-friendly application with a powerful and efficient recommendation engine at its core. The software responds quickly and makes it easy for users to identify restaurants in their area that they would appreciate.

Using collaborative and content-based filtering, the project was successfully completed. The dataset was scraped from Zomato for the city Bengaluru, and the basis characteristics were then preprocessed further. Finally, this data was utilised to create a model for the application.

For result sizes of up to 1000 items, the successful execution of was accomplished by developing a web application which shows the user top 10 restaurants based on their requisites.

References

[1] Anderson C., "A survey of food recommenders", Cornell University, Ithaca, NY, USA (2018)

[2] Hedge S.G., Satyapanvar S., Sety S., "Sentiment based food classification for restaurant business" in International conference on advances in computing, communications, and informatics

[3] J. Zeng, F. Li, H. Liu, and S. Hirokawa, "A Restaurant Recommender System Based on User Preference Location in Mobile Environment," in 5th International Congress on Advanced Applied Informatics, Kumamoto, Japan, 2016.

[4] Liu H., Wang Y., Peng Q., Wu F., Gan L., Pan L., Jiao P., "Hybrid neural recommendation with joint deep representation learning of ratings and reviews" Neurocomputing, 374 (1) (2020), pp. 77-85

[5] Abbasi-Moud Z., Vahdat-Nejad H., Sadri J., "Tourism recommendation system based on semantic clustering" in Expert Systems with Applications (2021)

[6] Mahadi M.I., Zainuddin N., Shah N.B., Naziron N.A., Rum S.F., "E-halal restaurant recommender system using collaborative filtering algorithm

[7] Ashok, M., Rajanna, S., Joshi, P. V., & Kamath, S. (2016). A personalized recommender system using Machine Learning based Sentiment Analysis over social data.

[8] Tao D., Yang P., Feng H., "Utilization of text mining as a big data analysis tool for food science and nutrition", Comprehensive Reviews in Food Science and Food Safety, 19 (2) (2020), pp. 875-894

[9] Zhang C., Zhang H., Wang J., "Personalized restaurant recommendation method combining group correlations and customer preferences", Information Sciences, 454 (1) (2018), pp. 128-143