Food and Restaurant Recommendation System Using Hybrid Filtering Mechanism

Amanuel Melese
Department of Information Engineering, Huzhou University

ABSTRACT
The application of machine learning and Artificial Intelligence system in the food industry is not more profound than the other sectors. Those available systems do not fully answer the expectations of customers, lacks user personalization’s. This paper presents the recommendation system for restaurants and food using a hybrid filtering mechanism. Multiple filtering mechanisms were applied on datasets to recommend food and restaurants for customers. The dataset was collected from recognized machine learning repositories of the U.S.A. government. Feature extraction and sampling are done on the datasets to test the performance of the system. This paper will try to answer user personalization preferences by applying Hybrid mechanisms. The recommendation was based on different customer preference like ratings, top sale, discount, weather condition etc. This paper combines content-based and collaborative based filtering mechanisms to provide the user with full functionalities of the recommender systems. It will adopt a hybrid system from the two mechanisms for effective implementation of the recommendation. For the development of this paper, data collection was done from known repositories, and feature extraction was done to filter out the unnecessary data. Sampling techniques were also applied to the dataset to distinguish between train and test data. 70% of the total dataset were allocated for training the system and 30% for test purpose. Sampling was also conducted in the model test stage by assigning 30% for evaluation purpose. To evaluate the performance of the proposed system, machine learning algorithms such as random forest, gradient boosting, decision tree, linear regression and K-Nearest neighbor were applied. The final performance of the model is so promising that it achieved an 83.5% success rate. Model loss and accuracy were also conducted, and the best fitting algorithms were selected. Based on the final result, the random forest algorithm shows significant performance with 0.859 accuracies and 0.1193 loss.

Keywords— RECOMMENDATION SYSTEM, COLLABORATIVE FILTERING, CONTENT BASED FILTERING, HYBRID FILTERING, CUSTOMERS, MACHINE LEARNING ALGORITHMS
Introduction

In recent years the lifestyle of human beings is almost changed from the traditional way of living into an easier technological advanced mode. A lot of technological innovations and systems are implemented which can assist daily activities and makes life easier. A recommendation system is the result of evolving technologies through times aimed to assist decision making. Pioneer to it, there are traditional programming which is a bit complex and difficult for decision making due to lack of intelligence and no deepest analysis of the data stored. The solution for such kind of problem is the emergence of a recommendation system. The recommendation system deeply analyzes the data, categorize and arrange them in a well-mannered form, apply some algorithms and make fast and precise decisions. As a result, it will save a great amount of time spent to make a decision, cost and energy lost on a specific course of actions.

According to F.O. Isinkaye et al, the rapid growth of digital information and visitors on the internet results in confusion of accessing personally engaging information. They also stated that search engines like Google, DevilFinder and Altavista have partially solved this problem, but still prioritization and personalization are not fulfilled [13]. Such kind problems is indicating that the demand for recommender systems is increasing more than ever before. The recommender system will investigate the big volume of user information and decide whether the user prefers the items or not decide based on user preferences, interest or behavior. Recommender system will greatly benefit both parties, users and service providers in a manner of saving time and costs. They reduce transaction costs of finding and selecting items on the internet and also improve decision making quality. In e-commerce recommending products based on customers search will enhance sale, in movies industries providing related movies for their search result will help to increase view, in libraries recommender systems helps readers by letting them discover further than catalogue searches.

The application of a recommendation system in the food industry assists customers to choose the restaurant, selecting the food based on their current mood and weather condition, even helping them decide what to eat occasionally in some specific restaurants. They will greatly help customers to select restaurant and recipe just by analyzing daily or monthly sales report or by considering the weather conditions, by age and gender or even more specific to customers need. There exist a large number of information in the food industry. Because of a large amount of data, the output of the web-based application is not easy and flexible to make a decision in the food industry unless we apply some techniques which enhance decision making. This is where A.I. driven systems are highly important and greatly reduce much effort to extract data for decision making.

It is possible to make machine learning-based decision just by feeding data and train the machine how to decide without much effort on programming. Such kind of decision making is greatly important for food companies to satisfy the customer, improve the experience and create loyal long term customers. There are a lot of online food supplies, and customers have a variety of choices to go for. The suppliers must make
customers remain a long time by providing a smart system which helps to decide a product to order based on seasons, age, customer mood, experience, top sale, discounts etc. The application of the recommendation system sounds loud here as a solution for this kind of problems.

**Literature review**

There is plenty of definition of recommender system. Many scholars define it from some point of view. For example, Faisal Rehman et al; defines recommender systems (R.S.) as a tool that provides the users with suggestions of information that may be useful to them [8], while Raciel Yera Toledo et al, [5] define recommender systems as any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options.

According to Adomavicius et al. [4] recommender system is rooted back in extensive work in cognitive science, approximation theory, information retrieval, forecasting theories, and also have links to management science and to consumer choice modelling in marketing and dated back to the mid-1990s. They stated that the recommendation problem is formulated by estimating ratings for the items that have not been seen by a user, which will be based on the ratings given by the users. They formally describe the formulation as follows

\[
\forall c \in C, S_{c1} = \arg\max_{s \in S} u(c, S)
\]

while C is the set of all users and S be the set of all possible items that can be recommended. They classified the recommendation system into three parts. Content-based filtering, collaborative filtering and hybrid methods. They also briefly explain the three categories.

On the other hand, Peter Forbes et al. [15] stated that the majority of recommender systems today, such as those used by Netflix and Amazon, apply either collaborative filtering or content-based filtering. They stated that Netflix adopted collaborative filtering based on the movies you like or not and find out users with similar taste. On the other hand, Pandora, a well-known music recommender, uses content-based filtering based on the song you like or dislike, whether you like fast, slow, soft music with female vocals or not. According to this paper, collaborative filtering doesn't require machine analysis content; it will recommend without understanding the item itself. Because the new items don't have much user feedback, this type of filtering suffer a cold start problem. However, this problem can be solved by exploiting any known content information.

Anusha Jayasimhan et al. [6] propose the development of a restaurant recommendation system that will help a user to decide which restaurant one should visit, and Elahi, M. et al. [7] proposed tags and latent factor technology used for android based food recommender system. The system recommends personalized recipe
to the user based on tags and ratings provided in user preferences. Latent feature vectors and matrix factorization algorithm are applied to the system. According to this paper, prediction accuracy is measured by using tags which closely match the recommendation of users preference. There are also systems proposed for menu recommendation system [9] and diet plan recommendations [10], and recipe recommendations [11].

The majority of recommendation systems uses users' preferences like users ratings [15], recipe choices [16], and browsing history [17]. For example, in [15], a recipe recommendation system is proposed using content-based matrix factorization, which will use users rating as users preference. In [16], they create three concepts for the experiments in recipe choice. First, they suggest a recipe to the user so that the application selects the recipes based on personal preferences; in the second concept, users receive their intake of different nutrients and can choose personally like fibres or vitamin A. Based on the selected recipes, the specified nutritional information is combined and visualized in the third concept giving advice to the users on how to improve their eating behavior based on past choices. In [17], a recipe recommendation system is based on recipe similarity measure.

Keiji Yanai et al. [1] proposed optical recognition technology for food recipe recommendation. They stated that real-time visual object recognition of food ingredients, and recommendation of cooking recipes using the built-in camera on a smartphone to food ingredients, a user can get to know a related cooking recipe instantly. To speed up object recognition for enabling the system to recommend cooking recipes in a real-time way, they propose colour-histogram-based bag-of-features extracted from multiple frames as an image representation and a linear kernel SVM as a classifier. On the other hand, Lee et al. [2] propose a system to learn an object template from a video stream and localize and track the corresponding object in live video. This system will use mobile object recognition technology on a smartphone, which recognized registered objects in a real-time way.

Rachel Yera Toledo et al. [5] stated that there is still an open and active research problem on menu generation for users. According to this research paper, the framework for menu generation is still lack personalization features. Basically, they focus on two research clusters—building complex information models as the basis for personalized services and nutritional information processing. For the first time, this research paper proposes the development of a food recommendation model that integrates both nutritional and user preferences-related information and integration of multi-criteria decision making (M.C.D.M) sorting processed together nutritional information-awareness within the food recommendation domain.

Bushra Ramzan et al. [12] proposed intelligent data analysis for hotel recommendation using machine learning approaches. They use a novel collaborative recommendation approach in which opinion based sentiment analysis is used to achieve hotel feature matrix by polarity identification. They propose a true
recommendation system that will use a hotel feature utility matrix to recommend a suitable hotel to a user on
the basis of both quantitative (numerical) and qualitative (textual) features to achieve true recommendations.
This system will recommend the hotel based on users true preference such as room and food and cleanliness,
pool, spa and gym, Wi-Fi and computer.

**Problem statement**

Unlike health, military and commercial industry, A.I. is not widely and sufficiently applied to the food
industry. Due to this reason, there are not sufficient systems which help users and owners of the food
industry to have a smart assistant on daily consumption, even though there exists a wide range of system
which helps users to order online, receive foods where they are. Giant food suppliers like McDonald's,
K.F.C. and Starbucks have their own system which connects customers to their services wherever they are,
but few of them only have a rich and smart system that helps the user to personalize the order. Only a few
food suppliers have a smart assistant system that helps the company to decide how much amount of
ingredients to buy based on sales experience of previous days or how many ingredients to supply for the
coming day's sale. Most of them don't know which product customers order rarely or regularly to enhance
customer service.

The biggest food supply APP in China, called takeout, have a huge amount of local food suppliers, but most
of them doesn't have a recommendation system for their clients. Whenever customers are searching for
foods, the systems don't recommend best-selling food of the day in a specific restaurant or the food they have
to eat based on their specific mood, weather condition, calories etc. Most of the times, customers order the
type of food they frequently eat or try just one blindly, which might not satisfy them. Sometimes they might
even change the restaurants to look for better choice. Hence they will take a bit long time to search and order
food, and also addicted to only one kind of food they frequently order rather than tasting a variety of kinds of
foods.

It is also difficult to choose a restaurant based on a specific recipe in mind unless there is a recommendation
system in place that will assist in choosing where to go. A lot of customers spent more times searching the
best restaurants for the type of food they want to eat. It is true that most customers know limited restaurants
for enjoying the type of food they want. Whenever they feel to try a new recipe or eat their favorite one in
new environments, there must be a recommender system in place which will assist them in deciding which
restaurant to visit and what to eat there.

**Proposed solution**

This system is proposing the recommender system using hybrid filtering used to recommend products to
customers based on different criteria such as ratings, top sale of the day, seasons, discounts, age, gender and
mood of customers and browsing history etc. This system will recommend the food and restaurant based on
users preferences.

The customers $U = (u_1, u_2, \ldots, u_m)$ will rate the set of food in the restaurants $I = (i_1; i_2, \ldots, i_n)$. For these food a set of ratings has been provided from the customers, $R = (R_{ij})$. These ratings can be arranged in a $n \times m$ matrix $T$.

$$T = \begin{bmatrix} u_1 & \cdots & u_m \\ i_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots \\ \vdots & \ddots & \ddots \\ i_n & \cdots & u_m \end{bmatrix}$$

The ratings are provided on a separate scale by the user, $rij \in \{1; 2; 3; 4; 5\}$, the system uses these ratings for analysis and recommendation of restaurants and foods. The purpose of a recommender system here is to estimate a rating of food and restaurant $i$ for a customer $j$ where $rji \notin R$, i.e. $j$ has not rated the item before.

Whenever a customer wants to visit the restaurant for a specific type of food, such a system will guide them to the right place. It will assist the customers in selecting the restaurants based on different conditions like price, discount, services they provide and tastes. The proposed system will also save time and effort to find the best foods in specific restaurants. Such kind of system is vital for giant food suppliers as well as the middle and small business food industry to be competent. It is difficult to manage customers interest unless the company provide them flexible and fast food selection assistant system. The application of such a system will help to create loyal customers and make them stay with the company for a long time. As a result, the company will appear competent in the markets than the other competitors. Recommendation system will play a great role in increasing revenues.

Hybrid filtering is basically a combination of content-based filtering, and collaborative filtering mechanism even includes knowledge-based filtering too. This paper uses the above mentioned two filtering mechanisms, collaborative and context-based filtering techniques and algorithms interchangeably for the purpose of recommending food and restaurant to customers.

**Content based food and restaurant recommendation system**

Content-based filtering is the earlier types of filtering mechanism which recommends items similar to the ones which are rated or liked previously by the user. It is composed of user and item profile which consists of attributes and features used for recommendation. Attributes such as user I.D., ratings, place I.D., cuisine, item description, actual price and others are used. For example, if users previously like Mexican food, another type of Mexican food are recommended to them, which they don't like before. The system will extract food attribute to generate food profile and user profile, then compare user profile with food profile and finally recommend the food which is not yet liked by this specific user. The Term frequency/inverse document frequency is used to decide the important attributes. The frequency of food attributes $wx$ in
dataset dy is

\[ \text{TF}_{xy} = \frac{f_{xy}}{\max f_{ij}} \quad \text{and} \]
\[ \text{IDF}_x = \log \frac{X}{X_i} \]

where \( f_{xy} \) is the frequency of food attributes \( w_x \) in datasets \( y \) and \( f_{ij} \) is the maximum of frequencies of all important food keywords in datasets. \( X \) is the total number of datasets and \( X_i \) is the number of datasets containing food attributes \( i \). The weight of food attributes \( w_x \) in dataset \( y \) is \( w_{x,y} \) defined in terms of TF and IDF as

\[ W_{x,y} = \text{TF}_{xy} \times \text{IDF}_x \]

Each dataset is represented as a vector of weights of each keyword,

\[ d_y = (w_{1,y}, w_{2,y}, \ldots, w_{n,y}) \]

**Collaborative based food and restaurant recommendation system**

A collaborative recommendation is dealing with the customers currently using the system. It is used to recommend food and restaurants for active users. This recommendation is considering other customers behavior and choice, which is similar to the active one for recommendation purpose. For example, the similarities between two users are finding out by considering different options such as the ratings of similar foods, the price preference of the two users, the similarity between the choice of foods etc. The system will detect any similarities between previous users and active ones and suggest food and restaurant for active users.

For a set of users \( U=\{u_1 ,u_2,u_3\ldots u_n\} \) and set of foods \( I=\{i_1 ,i_2,i_3\ldots i_n\} \), rating of a user \( u \) on food \( i \) is \( R_{u,i} \).

The similarity measurement is also applied to items profiles and user profile discussed above. Pearson coefficient similarity measurements are used to determine the food and user similarities, as illustrated below.

For active user the vector representation of similarity measurement of Pearson coefficient is Eq. 5.

\[ r_{ui} = \frac{\sum_{i \in I} (X_{ui} - \bar{X}) (Y_{uj} - \bar{Y})}{\sqrt{\sum_{i \in I} (X_{ui} - \bar{X})^2 \sum_{j \in I} (Y_{uj} - \bar{Y})^2}} \]

Where \( i \in I \) is the summation both the users have rated for each items. In item-based algorithm, the set of users denoted by \( a \in A \) who rated both food \( i \) and \( j \), then the Pearson Correlation is given in Eq.6.

\[ p_{ai} = \frac{\sum_{i \in A} (X_{ui} - \bar{X}) (Y_{uj} - \bar{Y})}{\sqrt{\sum_{i \in A} (X_{ui} - \bar{X})^2 \sum_{j \in A} (Y_{uj} - \bar{Y})^2}} \]

Where \( r_{ui} \) is the rating of user \( u \) on item \( i \).
Hybrid filtering based food and restaurant recommendation system

The combination of content-based and collaborative filtering can be used for recommending food and restaurants to customers. Content-based filtering will use TF/IDF mechanism to filter out the relevant food and restaurant words for specific customers. It will evaluate and compute the current customer's food or restaurant search result and find out the relevant one by comparing the search result and the datasets. If the type of search results which the current customers didn't like before was found, it would recommend it to them. For example, the system will be going to see the current users mood, age, food preference and other information and look into the dataset to find out similar data with customers profile and search result. If the system finds one, it will recommend it to the customers. The system will use different customers, food and restaurant information to make a recommendation.

On the other hand, collaborative filtering will deal with active users and compute the similarity between the other users and current users preferences to decide which food and restaurant to suggest for the active one. The information of previous users such as ratings, age, gender, types of food they like, restaurants they rate, the price preferences, mood, seasons and others available information's are used towards the suggestion of food and restaurant for the active one. While doing this, the system will use Pearson similarity computation to decide the similarity between the active and previous one.

The combination of these two filtering mechanisms will be interchangeably used in the system to provide relevant information to customers. The user preference problem will be solved using a hybrid recommendation system in such a way that the hybrid will consider the recommendation in aspects of multi-information about the customers, food and restaurants. Content and the collaborative engine will work together and share information about the decision they have to make for specific customers, as a result, relevant and cost-effective recommendation was made.
In order to provide relevant food and restaurant information to users, the system is composed of customer interacting with the system interface and the recommender system suggest food and restaurant information to customers. Users are interacting with the system through the system interface. The system will import datasets from Apps or websites and start data preprocessing and feature extraction to remove unnecessary fields from the datasets which have no significant values for decision making. The already extracted data is sent to the recommender system to decide which food and which restaurant. In a hybrid recommender system, both content and collaborative filtering mechanisms are applied to datasets interchangeably to provide relevant information for users. Relevant food and restaurant information are in respect to users are processed using TF/IDF computation in content-based, and similarity calculation was also done to check the relationships between customer, restaurant and foods. Based on this all computation, the system is going to recommend for the users.

**Methodology**

**Data analysis:** - Initial analysis process are performed on data so as to discover patterns, spot irregularities, test hypothesis and check assumptions with the help of summary statistics and graphical representations. Multiple regression multivariate analysis on data fields were performed to see if there is a significant statistical relationship between multiple variables. For each independent variables, multiple regression \( \{x_1,x_2,x_3..x_n\} \) were used to input to see the relationships between each of them. Simple regression: \( Y = d_0 + d_1 t \) as well as multiple regression: \( Y = d_0 + d_1 t_1 + d_0 + d_1 t_2...d_0...d_1 t \) were performed on datasets.
Based on the analysis, there are fields which has nothing to do with decision making and also errors in dataset entries that need to be removed for better performance of the model. While performing analysis, CourseID, CourseName, Sales ID, Sales Item and Quantity fields are observed as playing an important role in recommending the restaurant and food.

![Graph 1: Top selling restaurants](image)

**Sampling techniques:** - sampling data for training and test purpose, this research divide the whole data set into test and training as well as validation data. Of the total n dataset n(30/100) is allocated for test data, and the remaining n(70/100) are allocated for training data while n is the total number of datasets. Sampling was done on datasets processed after multivariate data analysis. For decision making purpose, the datasets are even divided into validation and test data to see the performance of the system.

**Application of machine learning algorithms:** - Machine learning algorithms are applied to a system to figure out how the system behaves and how accurate the decision-making process made. Linear regression, K.N.N., Decision trees, Gaussian NB, Random forests Classifier and Gradient Boosting Classifier, were applied.

The dataset used for this research was collected from the U.C.I. repository (https://archive.ics.uci.edu/ml/index.php) and the home of the U.S. government’s open data (https://catalog.data.gov/dataset/parks-golf-sales-detail). The first dataset is composed of different tables, which consists of many information about users, restaurants and foods. For the purpose of providing the best recommendation, the system will merge some tables like cuisine, users and ratings together based on their unique I.D. There are ratings of foods and restaurants in this datasets and are used accordingly to recommend to the users. Data preprocessing and feature extraction was done on this dataset to filter out only relevant attributes for recommendations so that unnecessary field is removed from the datasets. The second dataset consists of 329824 entries. The attributes are CourseID, CourseName, Sale ID, SaleDate, Report Date, Item ID, Item Description, Actual Price, Quantity and L.M.P. fields. In order for the system to facilitate the decision making capability of the system, I remove the attribute labelled as not important for decision making. L.M.P. field and Report date have nothing to do with recommending the restaurant or food. There are also entries that describe other activities such as managerial activities and has nothing to do with a
recommendation. This kind of entries is also eliminated from the dataset. Finally, after applying feature extraction, 12032 entries left out of 329824.

Selecting the appropriate programming language for the implementation is also done. Python was chosen as a platform to develop the system because of wide and rich libraries crucial for this project, such as sklearn, pandas, matplotlib, keras, tensorflow and others. Anaconda Navigator includes all these libraries into one package. Specific environments were created for the implementation and installed all the necessary packages listed above. Jupyter Notebook is used to write scripts and simulation of the results in browsers.

To evaluate how well a classifier is performing, we have to test the model on invisible data. For the purpose of this research, before building a model, the data is split into two parts: a training set and a test set. The training set is used to train and evaluate the model during the development stage. You then use the trained model to make predictions on the unseen test set. This approach gives you a sense of the model's performance and robustness. Luckily, sklearn have a function called train_test_split() which divides data into sets. The function randomly splits the data using the test_size parameter. For the purpose of this research, the test size represents 30% of the original dataset. The remaining make up the training data.

The below table describes the list and performance comparison of different algorithms applied to the system. The value of performance evaluation for Accuracy, Loss and model evaluation was mentioned between 0 and 1. For Accuracy measurement, the more the value is closer to 1, the more acceptable and accurate the system is. For the loss measurement, the more the value approaches 0, the more accurate the system.

<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>Accuracy</th>
<th>Loss</th>
<th>Model evaluation result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear regression</td>
<td>0.7407</td>
<td>0.4547</td>
<td>0.7407</td>
</tr>
<tr>
<td>2</td>
<td>K-Nearest Neighbor</td>
<td>0.7037</td>
<td>0.5390</td>
<td>0.666</td>
</tr>
<tr>
<td>3</td>
<td>Decision trees</td>
<td>0.7778</td>
<td>0.5193</td>
<td>0.7407</td>
</tr>
<tr>
<td>4</td>
<td>Gaussian NB</td>
<td>0.7778</td>
<td>0.5884</td>
<td>0.555</td>
</tr>
<tr>
<td>5</td>
<td>Random Forest</td>
<td>0.8519</td>
<td>0.1193</td>
<td>0.7777</td>
</tr>
<tr>
<td>6</td>
<td>Gradient Boosting</td>
<td>0.8148</td>
<td>0.5127</td>
<td>0.7037</td>
</tr>
</tbody>
</table>

According to the above analysis of algorithms random forest has highest accuracy and lowest loss followed by gradient boosting and decision tree algorithms. Linear regression and decision tree has the highest model evaluation following random forest. The graph below shows the loss and accuracy result of random forest.
Conclusion

A hybrid recommendation system will improve the recommendation system by considering the user's preference and personalization problems. This paper implements the content and collaborative filtering on different stages of datasets and users information. For the purpose of this paper, datasets were collected, and analysis was performed on data sets. The datasets were split into test and train data for performance evaluation, and in-depth analysis was conducted using different machine learning algorithms. The system successfully recommends the restaurants and the food based on customers preference. The recommendation was done based on price, quantity, ratings, geographical places and other attributes. Last but not least, researchers have to focus on improvement of the performance and also the realization of personalized
recommendation system in food and restaurant industries.

**Acknowledgment**

I thank Almighty God for instructing, teaching and leading me throughout my life and specifically in this research work. Next, I would like to thank my advisor, professor Lou Jungang, for his assistance and follow up throughout my work.

**References**


[2] Taehee Lee, Stefano Soatto, Learning and Matching Multiscale Template Descriptors for Real-Time Detection, Localization and Tracking, Computer Science Department University of California, Los Angeles, CA 90095, page 1457


[5] RACIEL YERA TOLEDO, AHMAD A. ALZAHRANI, AND LUIS MARTíNEZ, (Member, IEEE), A Food Recommender System Considering Nutritional Information and User Preferences, Knowledge Management Center, pages 96695 – 96709, 2019


