Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd
Anatolian Science

ISSN,e-ISSN: 2548-1304
Volume:9, Issue:1, pp:1-17, 2024
https://doi.org/10.53070/bbd.1389078
Research Paper

DEĞİŞİMLİ EN KÜÇÜK KARELER VE KOSİNÜS BENZERLİK TEKNİKLERİ KULLANILARAK YEMEK TAVSİYE SİSTEMİ OLUŞTURMA

BUILDING FOOD RECOMMENDATION SYSTEMS USING ALTERNATING LEAST SQUARES AND COSINE SIMILARITY TECHNIQUES

Merve CENGIZ[*1], Tugba OZKAL YILDIZ[2]

[1]Data Science, Dokuz Eylul University, Izmir, Turkey

(merwes1@gmail.com, tugba.ozkal@deu.edu.tr)

Received:Nov.10, 2023 Accepted: Jan.08, 2024 Published:Jun.01, 2024

Özetçe— Bu çalışmada, Allrecipes.com web sitesindeki yemek tariflerine ve üyeler tarafından verilen oylara dayalı bir yemek tavsiye sistemi geliştirildi. Toplam 1840 yemek tarifi (Diyabetik - Glutensiz - Ketojenik - Düşük Sodyum - Düşük Kolesterol - Vejetaryen – Vegan) ve 66012 kullanıcı Allrecipes.com'dan web scraping yöntemi ile kazındı ve Python'da analiz edildi. En yüksek oy (5 puan) sayısı 63.568, en az oy (1 puan) sayısı ise 2.298 olarak bulundu. Veri setinde düşük kolesterollü yemek sayısının en fazla (324), ketojenik diyet yemek sayısının ise en az (210) olduğu belirlendi. Ortalama oya göre, vegan, vejetaryen ve ketojenik tariflerin en fazla tercih edilen; düşük kolesterol ve diyabetik tariflerin ise en az tercih edilen diyet tarifleri olduğu tespit edildi. Bunun yanı sıra, toplam oya göre en fazla ve en az tercih edilen diyet tariflerinin glutensiz ve ketojenik olduğu belirlendi. Tavsiye Sistemi, Değişimli En Küçük Kareler (DEKK) yöntemi kullanılarak oluşturuldu. Diyet Yemek Tavsiye Sistemi, kosinüs benzerlik yöntemi kullanılarak gerçekleştirildi. DEKK yönteminin büyük veri ile uygulaması bulut üzerinde gerçekleştirildi. Modelin hata kareler ortalamasının karekökü 0.495 olarak bulundu. Modelin önerdiği yemekler kullanıcı bazlı incelendi ve sonuçların tutarlı olduğu belirlendi. En çok tavsiye edilen yemekler incelendiğinde, vejetaryen tariflerin ilk sırada yer aldığı; toplamda ise ketojenik tariflerin yüksek sayıda önerildiği görüldü. Sonuç olarak, yemek tarifleri aracılığıyla yiyecekler hakkında fikir sahibi olmak ve diyetlerine göre yiyecek seçmek isteyen kullanıcılara öneriler üreten web tabanlı bir yemek öneri sistemi oluşturuldu.

Anahtar Kelimeler : *Tavsiye sistemi, değişimli en küçük kareler, kosinüs benzerlik, büyük veri*

Abstract— The aim of this study was to develop a food recommendation system based on the recipes on the Allrecipes.com website and the ratings given by the members. Total 1840 recipes (Diabetic - Gluten Free - Ketogenic - Low Sodium - Low Cholesterol - Vegetarian - Vegan) and 66,012 users were scraped from Allrecipes.com and analysed in Python. The highest rating (5 scores) count was 63,568 while the lowest rating (1 score) count was found to be 2,298. It was observed that there are mostly low cholesterol foods (324) in the dataset whereas there are the least number of ketogenic foods (210). Vegan, vegetarian and ketogenic recipes ranked in the top 3 based on the mean rating. According to the total number of ratings, the most and the least rated recipes were gluten free and ketogenic, respectively. Recommendation System was performed by using Alternating Least Square (ALS). Diet Food Recommendation System was performed by using cosine similarity technique. The application of ALS technique with big data was performed on the cloud. The root mean squared error of the model was found 0.495. The foods recommended by the model were examined on a user basis and it was determined that the results were consistent. When the most recommended foods were examined, vegetarian recipes were ranked first, and in total, a high number of ketogenic recipes were recommended. Consequently, we created a web-based food recommendation system that generates recommendations to users who want to have ideas about foods via recipes and choose food according to their diet.

Keywords : *Recommendation system, alternating least squares, cosine similarity, big data.*

1. Introduction

Nutrition is one of the factors that significantly affect an individual's health, such as physical exercise, sleep, and genetics, and it is the easiest factor to be changeable in our lives. That's why small changes can have a significant impact. Throughout human history, cooking and nutrition have been indicators of the development of society and civilization. During the Covid-19 pandemic quarantine period, individuals returned to their home kitchens to prepare meals. It was found that about half of American adults started cooking at home more often by a survey in April 2020. As a result, click-through rates of recipe websites such as Allrecipes and Tasty have increased. Consequently, in the light of global food, material shortages and price fluctuations, it has become even more important to help individuals cook healthy meals and improve their nutrition (Li and McAuley, 2020). However, it is not possible to examine all of the nutrition options that are constantly increasing and make a decision. Therefore, the only way to maximize healthy nutrition options and minimize unhealthy ones is to use recommendation systems by taking into account user preferences. Because the selection of meals with appropriate nutritional values depends on the individual's health conditions and food preferences. Therefore, it is important to provide personalized meal recommendations according to individual needs (Vairale and Shukla, 2021). For instance, it is a known fact that a personalized recommendation system would clearly provide benefits compared to a system provide general nutrition recommendation. (Oh et al., 2010).

There are various data filtering techniques that can be used to generate recommendations, including content-based and collaborative-based techniques that are often used in recommender systems. In this study, food recommendation systems based on both collaborative and content-based filtering methods are developed. Accurate recommendations are provided with the relationships established using Alternating Least Squares (ALS) which is one of the collaborative filtering techniques. In addition, with Cosine Similarity, the foods with the similar food content are listed as recommendations. With the ALS model, users were given the opportunity to evaluate the foods on a scale of 1 to 5 (1: worst, 5: best). Thus, the model was retrained with the new scores added, and the user was given recommendations for foods that he/she had not previously evaluated.

This study focuses on the development, evaluation and design of a food recommendation system. The system generates accurate recommendations using data such as users' ratings and food contents. The recommendation system that built with collaborative filtering addresses the data sparsity, scalability and cold-start problems of recommendation systems. In this study, the reason for preferring the ALS method is to address the sparsity and scalability problems of collaborative filtering. For the cold-start problem, we have used an approach to have the user's prior knowledge. Consequently, this study presents a web-based food recommendation system using collaborative and content-based filtering algorithms.

In the first section of the study, it was explained that why recommendation systems are needed. In the second section, the methods and platforms used to build recommender systems were explained. In the third section, the implementation and results were given. Finally, conclusions were presented in the fourth section.

2. Material and Methods

The flow chart of the Food Recommendation Systems application method is given below (Figure 1). Data scraping was performed using the BeautifulSoup library to retrieve recipes (Diabetic - Gluten Free - Ketogenic - Low Sodium - Low Cholesterol - Vegetarian - Vegan) and other features. In the data preprocessing step, tokenization, removing unnecessary characters such as punctuation marks - emoji - numbers - symbols from reviews, lemmatization were performed. MongoDB was used for data storage. Thus, data correction and management operations were performed effectively. Users and recipes were analysed on the basis of ratings with exploratory data analysis. ALS model was constructed. To create the Diet Recipes Recommendation System cosine similarities were created for each of diet types. Both recommendation systems were deployed by using Flask application.

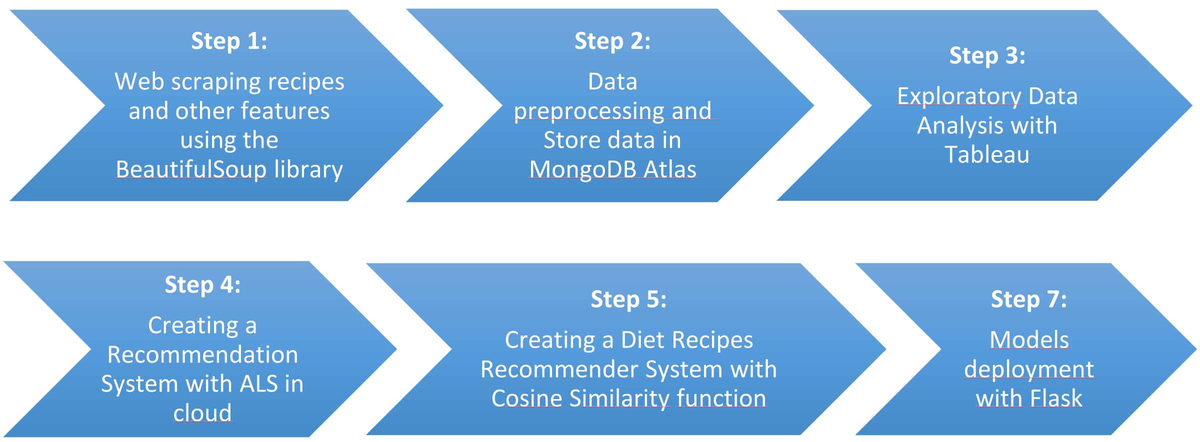


Figure 1. Food Recommendation Systems Application Method Flow Chart

The web scraping technique used in this study is summarized below. Recommendation systems which form the basis of the study are explained in detail in 2.2. section of the article.

2.1. Web Scraping:

Generally, data scraping is the process of extracting, obtaining, copying, scanning or collecting data from websites or internet-related sources. It is called web scraping, as well (Lawson, 2015).

Rapid developments in technology have made the information on the internet dynamic and transformed it into a real-time information source. We can now access information on the internet with a browser (Google Chrome, Safari, Mozilla Firefox, etc.) (Lawson, 2015). Mitchell (2018); "If the only way to access information on the Internet is through a browser, you are missing a wide range of possibilities". Browsers are better for viewing and reading data. But web scraping techniques are excellent at quickly collecting large amounts of data or databases without a series of adverts and suggestions of other popular sites (Mitchell, 2018). In this study, Beautifulsoup library in Python was used to scrape data.

2.2. Recommendation Systems:

Currently, consumers and businesses are faced with the problem of high-speed data growth (Li et al., 2018). This problem makes it difficult for users to find the product they are looking for on the internet. Recommendation systems, which emerged as a solution, increase the satisfaction of the user by making the recommendation in the best and shortest way, and besides increase the profit of businesses (Bozkurt and Acı, 2021).

Personalized recommendation systems are information filtering. It tries to find what is interesting for a user from a large number of items (food, film, music, book, website, news, etc.). The score or rating of an item for the user indicates the user's degree of interest. With the recommendation system, an item that the user has not used before can be predicted and the items with the highest ratings can be recommended to users (Xie et al., 2016).

For example, a user who wants to cook/eat is looking for a food that he/she will like, filters the foods he/she finds with various parameters such as ingredients, cooking time, etc., and may ignore a food that he/she may like very much in terms of content. Recommender systems can recommend a food that the user may like but has never tasted before (Bozkurt and Acı, 2021).

Recommender systems are generally divided into 3 types (Özcan and Çelik, 2018; Bozkurt and Acı, 2021):

- Content-based Filtering

- Collaborative Filtering

- Hybrid Filtering

In this study, we used content-based and collaborative filtering methods.

2.2.1. Content Based Filtering

In content based filtering, recommendations depend on users' previous preferences. The item description and the profile of the user (which item they rated for) play an important role. Content-based filtering algorithms try to recommend items based on the count of similarities (Thorat et al., 2015). For example, if film A, which a user preferred in the past, has common features (content, genre, etc.) with film B, film B is also recommended to that user (Özcan and Çelik, 2018).

In the study, we used cosine similarity technique which is one of content based filtering methods. Cosine similarity (CS) is a measure of the similarity between two vectors of an inner product space (Philip et al., 2014). Similarity is measured by the cosine of the angle between two vectors and determines whether two vectors point in roughly the same direction (Han et al., 2012).

When each element in the system is considered as a vector, the similarity between these vectors is given below;

$$Cos(\theta) = \frac{X.Y}{||X||\ ||Y||} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2}\ \sqrt{\sum_{i=1}^{n} Y_i^2}}$$

X_i: X user's rating for i^{th} food,

Y_i: Y user's rating for i^{th} food,

$||X||$ denotes the length of the vector (Philip et al., 2014; Han et al., 2012; Fathollahi and Razzazi, 2021).

A cosine value of 0 means that the two vectors are at 90 degrees to each other (orthogonal) and are not similar. When the cosine value is closer to 1 the smaller the angle and the greater the similarity between the vectors (Philip et al., 2014; Han et al., 2012; Fathollahi and Razzazi, 2021). CS is independent of vector size (Kaya, 2019). According to the degree of similarity found, a recommendation list is created by ranking the foods in descending order (Gündoğan and Kaya, 2021).

In the study, a Diet Food Recommendation System was established by creating cosine similarity functions for food recipes belonging to 7 diet types.

The details of the other filtering method used in the study are explained below.

2.2.2. Collaborative Filtering

Collaborative Filtering is the process of filtering and evaluating items based on peoples' views and opinions (Schafer et al., 2007). With the Collaborative Recommender System:

- Prediction for a specific item: Predicting the ratings of other items based on the ratings the user has given to items (Linden et al., 2003).

- Recommending an item: Ranking the predicted items to create a list of items that may be useful (Linden et al., 2003; Sarwar et al., 2001).

Although this filtering method has advantages, it has some problems. These problems are:

Data sparsity: As the number of products in the system increases, the proportion of products rated by the user decreases and accordingly, the relationship calculation becomes difficult.

Scalability: It is generally valid for large databases and can be defined as the prolonged running time and decreased performance of recommendation system algorithms due to the size of the dataset.

Cold Start/Early-rater: The system cannot draw any inferences for users or items about which it has not yet gathered sufficient information. In this case, the system cannot produce accurate and strong recommendations because it does not have prior knowledge of the user/item. The most commonly used solution is to present a certain number of random items to the user at the beginning and ask them to rate. In this way, it is aimed to create an initial profile for the user (Ünaldı, 2022).

In this study the reason for using the ALS technique is that it is a model-based method. With the model-based method, even though the data is sparse, the relationship can be calculated by estimating the ratings. In addition, since the ALS technique uses parallel calculation, the scalability problem is prevented. To solve the cold start problem, users were asked to rate on 5 different foods in the developed web application. The system is designed to recognize the user and recommend foods suitable for his/her profile. This approach is called Personalized Recommendation System. As a result, despite the disadvantages of collaborative filtering, the recommendation system was able to generate recommendations by using a model-based technique and adopting an approach to obtain user prior information.

2.2.2.1. Model Based Method: Alternating Least Squares

ALS is a technique for training data and finding similarities based on matrix factorization. Matrix Factorization based methods are frequently used due to time saving and scalability (Ünaldı, 2022). These methods aim to find latent user-item relationships (hidden factors) in user rating matrix (Jiang et al., 2020). Users or items with the same factors are assumed to be similar (Uluyağmur, 2012).

High similarity between item and user factors leads to a recommendation (Koren et al., 2009). The model-based method builds a learning model using the user's previous ratings to predict the ratings of items that the user has not rated before (Barakat, 2020).

A basic matrix factorization model can be explained as below:

Users and items to be represented in a common f dimensional hidden factor space,

y_i : Element vector containing each element (i) in the matrix $Y(n \times k)$, $y_i \in R^f$

x_u : User vector containing each user (u) in the matrix $X(m \times k)$, $x_u \in R^f$

The interaction between the user u and the item i, $\hat{r}_{ui} = x_u^T y_i$, gives the rating matrix that shows the user's interest in the item (Koren et al., 2009).

Ω : All item and user pairs,

$(r_{ui} - x_u^T y_i)^2$: Lose function, λ : Regularization coefficient

$(|x_u|^2+|y_i|^2)$: Penalty term, the error function is calculated by assigning random values to the nulls in the rating matrix which is divided into two smaller matrices by matrix factorization. Thus, when learning the latent factors, it is necessary to find the lowest value of the error function (Awan, MJ. et al. 2021, Koren, Y. et al. 2009). Lambda is used to prevent the factors from taking too large values with the penalty term. In this way, the overfitting problem is overcome (Xie et al., 2016; Uluyağmur, 2012).

$$L(X,Y) = \sum_{u,i\in\Omega} (r_{ui} - x_u^T y_i)^2 + \lambda(|x_u|^2+|y_i|^2)$$

To solve the model, first when y_i is constant x_u is calculated and when x_u is constant y_i is calculated, hence the method is called as "Alternating".

$$L(X) = \sum_{i\in\Omega} (r_{ui} - x_u^T y_i)^2 + \lambda|x_u|^2$$

$$L(Y) = \sum_{i\in\Omega} (r_{ui} - x_u^T y_i)^2 + \lambda|y_i|^2$$

Now, the equation becomes a quadratic function. The process continues stepwise until the equation converges to a minimum (Chen J et al., 2017).

The partial derivative of x_u and y_i sets equal to 0. Consequently, x_u and y_i are found, where I is the unit matrix of rank k, r_u and r_i is the u^{th} and i^{th} row value of the R matrix, respectively (Chen J et al., 2017).

$$x_u = (Y^T Y + \lambda \mathrm{I})^{-1} Y^T r_u$$

$$y_i = (X^T X + \lambda \mathrm{I})^{-1} X^T r_i$$

The algorithm is iterated until it reaches the specified maximum cycle or error rate (Chen J et al. 2017). Finally, the matrices are multiplied again and the empty values in the rating matrix are predicted (Nguyen, 2021). As a result, $\hat{r}_{ui}$, rating matrix is predicted for recommendation to users. Thus, data sparsity is avoided (Xie et al., 2016).

3. Application and Results

3.1. Data Scraping from Allrecipes Website

Data scraping was performed from www.allrecipes.com website containing recipes. The variables obtained in the data scraping process are food name, total number of ratings, total number of reviews, total preparation and cooking time (minutes), energy value and nutrients, ingredients, recipe, user name, user rating and user reviews.

One of the features of the dataset is users. The feature refers the users name who reviewed about that food. Since rate, user id and food id are required in ALS analysis, we changed the size of the dataset by using user_id. In the new dataset, the feature shows how many times the food name is repeated. For instance, the food named Vegan Chocolate Cake was repeated as many times as the number of users who rate on it (Table 1). This structure was made for each food. The dataset was prepared for ALS analysis by assigning IDs to food names and users.

Table 1. The instance of the dataset for ALS

foodame	users	rating
Vegan Chocolate Cake	KADart05	5
Vegan Chocolate Cake	Abztrakt	3
Vegan Chocolate Cake	ginny1647	5
Vegan Chocolate Cake	Amber Waite	5
Vegan Chocolate Cake	flying chef	5
Vegan Chocolate Cake	JILLITH	5
Vegan Chocolate Cake	LAURIESUE	5

Vegan Chocolate Cake	Whrenwraith	5

3.2. Exploratory Data Analysis

The user's name of all users who comment on the website without being a member is automatically registered as "Allrecipes Member". This causes that the same user to be recorded as giving different rates for the same food. Therefore, Allrecipes Member users were deleted from the dataset.

In the study, there were seven diet types to create a Diet Meal Recipe Recommendation System using the CB technique, including diabetic, gluten-free, ketogenic, low-sodium, low-cholesterol, vegetarian, and vegan. Since a food can be categorised as more than one type of diet, the same food and the same user names were removed from the dataset. This implemented only in the ALS method. Thus, the problem of duplicate in the data set was prevented.

The range value of the rating score given by the users is observed as 1-5. The highest rating (5 scores) count is 63,568 while the lowest rating (1 score) count is found to be 2,298. (Figure 2).

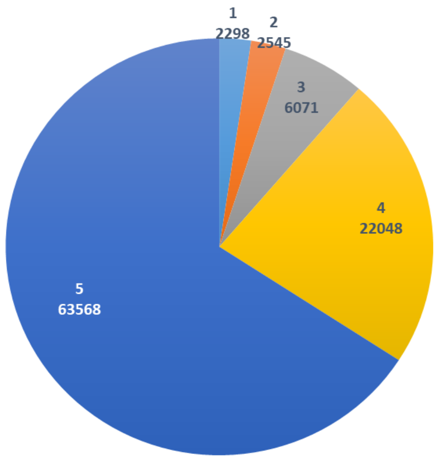


Figure 2. Distribution of ratings

In Figure 3, it is observed that there are mostly low cholesterol foods (324) in the dataset whereas there are the least number of ketogenic foods (210).

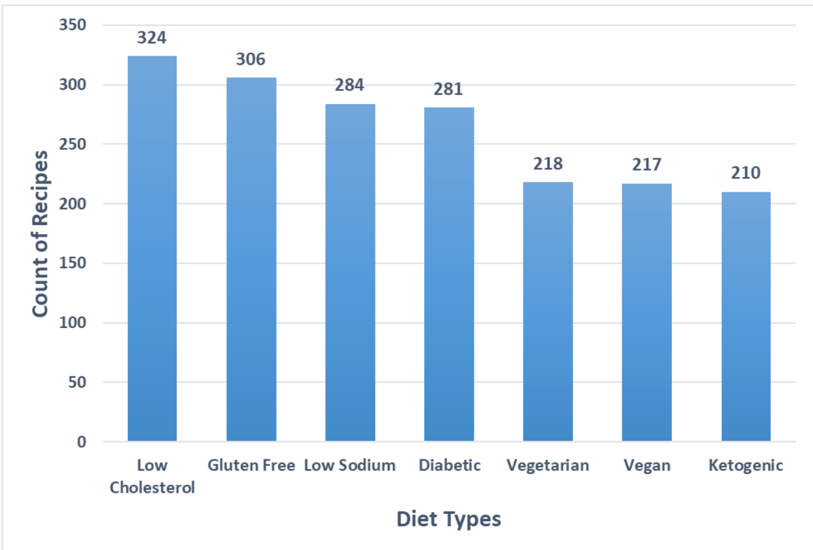


Figure 3. Count of recipes

According to the mean rate, vegan, vegetarian and ketogenic recipes are the most preferred while low cholesterol and diabetic recipes are the least preferred diet recipes. Low cholesterol has the highest number of recipes (Figure 3) but is the least preferred (Figure 4). This is explained by users giving low ratings to these recipes. Despite the low number of vegan, vegetarian, and ketogenic recipes, they rank in the top 3 based on the mean rating which is explained by high ratings for these recipes (Figure 4).

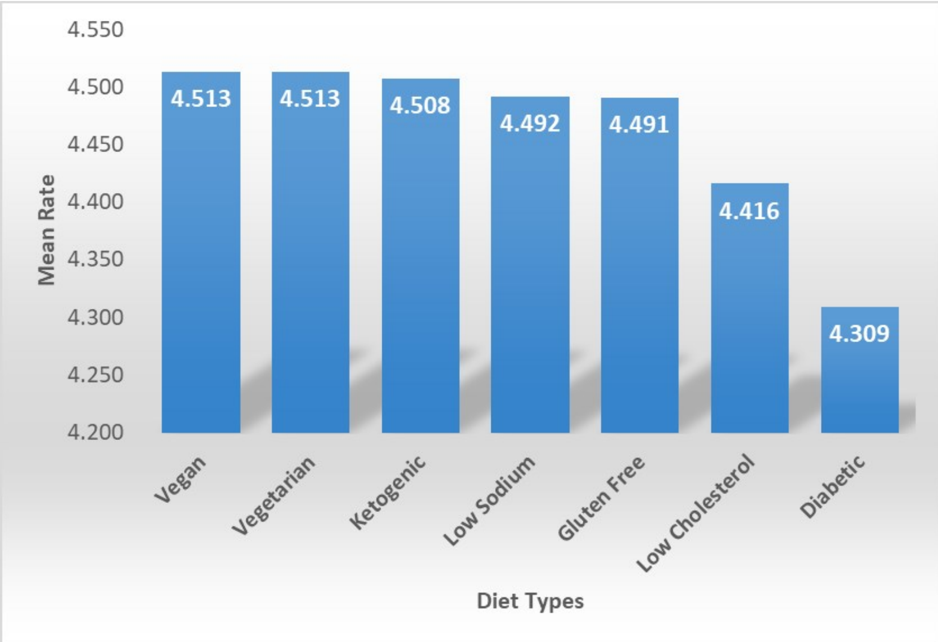


Figure 4. Diet types by mean rate

The names of the users who rated the recipes the most and the least are given in Table 2. In the table, out of total number of 66,012 users, only the 5 people who rated the recipes the most and the 5 people who rated the least are shown. When the values in the table are examined, for instance, the number 126 corresponding to the user Sarah indicates how many foods she rated for.

Table 2. Some users who rate the most and the least

Users	Total Rating for Foods
naples34102	171
Sarah	126
Jillian	122
Laura	115
Jennifer	114
Lisa Howard	1
‘Lil T	1
wb	1
_MJ	1
-steve-	1

In terms of the total rating, the most and least preferred foods and the total of ratings given to these foods are shown in Table 3. In the table, out of total number of recipes, only the 5 most and the 5 least rated foods are shown.

When the values in the table are reviewed, for instance, the number 942 corresponding to the food name Sauteed Apples shows how many ratings were given in total.

Table 3. Some foods rated the most and the least

Food Name	Total Rating
Sauteed Apples	942
Chef John's Shakshuka	940
Scott Ure's Clams And Garlic	940
Applesauce	940
Refried Beans Without the Refry	938
Keto Coconut Shrimp	2
Jicama Tortillas	2
Orange Surprise Pops	2
Basic Ketosher Omelette	1
Keto Flappers	1

The most and least preferred foods according to the mean rating are given in Figure 5. According to figure, 8 recipes are shown out of the total recipes. The first two most preferred foods (Colored Sugar - Mom's Sushi Rice) can be categorised as "foods that children like". The least preferred foods can be categorized as "sports meals" (Pepper Salad - Overnight Slow Cooker Apple Oatmeal).

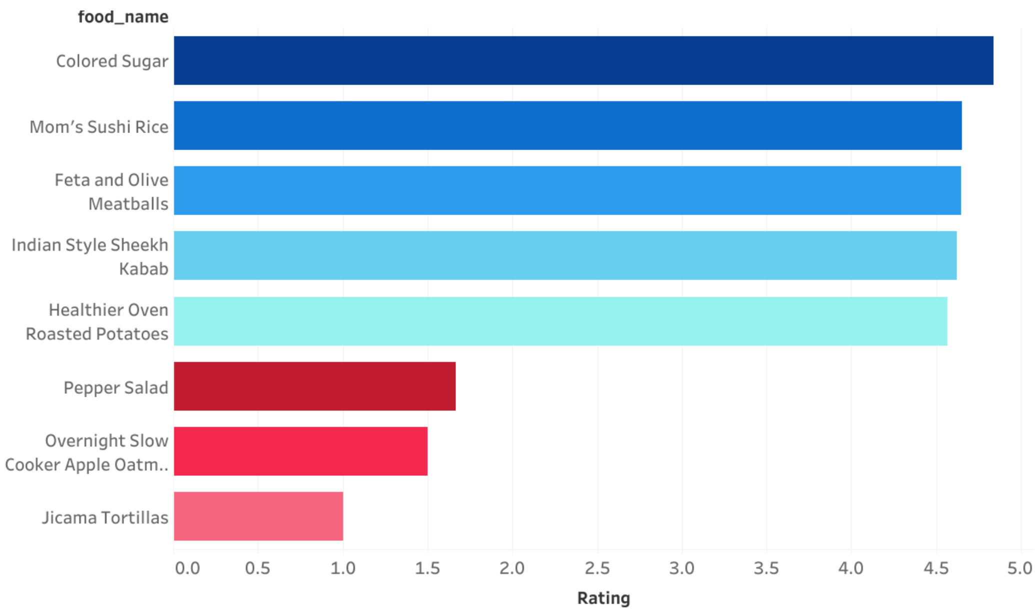


Figure 5. Some foods with the highest and lowest mean rating

Diet recipes according to the total number of ratings are given in Figure 6. The most and the least rated recipe is gluten free and ketogenic recipes, respectively. For instance, the count of ketogenic recipes and the total ratings given to these foods are low, but they are the one of the most preferred recipes according to the mean ratings shown in Figure 4. This is because the ratings given to the recipes are high.

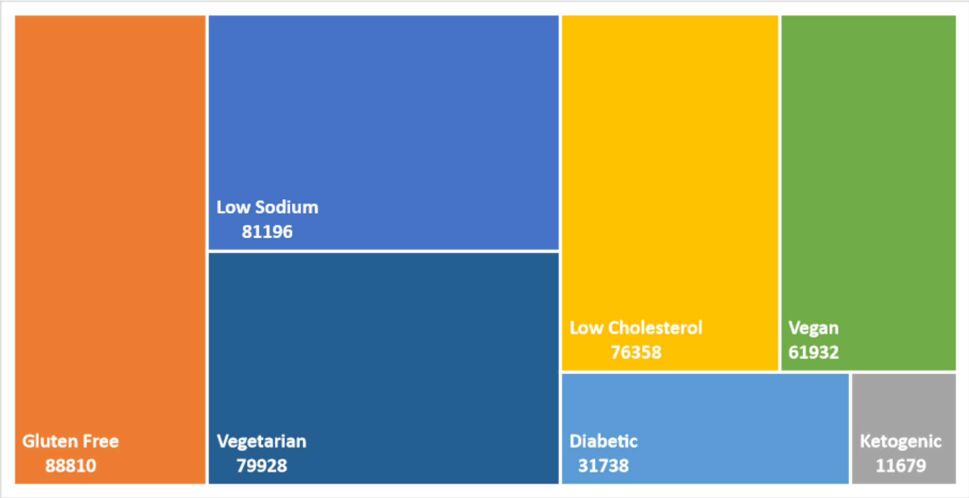


Figure 6. Total ratings of diet types

3.3. Results of Food Recommendation System Analysis Using ALS Technique

Before the ALS technique was applied, the dataset was divided into two parts as training and test set at a ratio of 0.8 and 0.2 using the randomSplit function. 10-fold cross validation was applied to the training set using the numFolds function. ParamGridBuilder function was used to achieve the best results for all hyperparameters. In order to find the optimal food recommendation system model, the root of the mean squared error (RMSE) was used as a model performance evaluation criterion. RMSE results according to the different hyperparameters are shown in Table 4.

Table 4. RMSE Results

Rank	maxIter	regParam	RMSE
25	75	0.05	1.257
5	10	0.3	1.160
5	10	0.5	1.154
25	10	0.05	0.571
50	100	0.03	0.542
25	20	0.03	0.524
25	20	0.05	0.523
50	50	0.03	0.519
50	20	0.05	0.510
50	20	0.03	0.499
50	10	0.03	0.495

Table 5. shows the optimum values of the hyperparameters for ALS technique.

Table 5. Optimum Hyperparameter Values for ALS Analysis

Rank (Number of hidden factors)	50
maxIter (Number of iterations)	10
regParam (Regularisation Parameter, λ)	0.03

Based on these findings, it is necessary to test the consistency of the foods recommended by the model with a group of users. For example, in Table 6 below, the model provides the top 5 food recommendations for a member with user id=7244 based on user ratings.

Table 6. Top 5 food recommendations for user id=7244

User id	Recommended foods	Recommended prediction
7244	Easiest Asparagus Recipe	3.893
7244	No-Churn Keto Ice Cream	3.874
7244	Crystallized or Candied Ginger	3.957
7244	Sarah's Ambrosia Fruit Salad	3.889
7244	One-Pan Keto Chicken Breast with Vegetable Ragout	4.016

To make a comparison, the average rating for each food in the table above was calculated. The results are shown in Table 7. Upon examination, it is observed that the average ratings for these foods are high. Therefore, the top 5 food recommendations with the highest ratings are accurate for this user.

Table 7. Mean rating of the recommended foods for user id=7244

Recommended foods	Mean rating
Easiest Asparagus Recipe	4.73
No-Churn Keto Ice Cream	5.0
Crystallized or Candied Ginger	4.58
Sarah's Ambrosia Fruit Salad	4.78
One-Pan Keto Chicken Breast with Vegetable Ragout	4.78

To perform a reverse check for the second user with user id=3918, first, the foods rated by this user are shown in Table 8 below.

Table 8. The foods rated by user id=3918

User id	Rated Food	Rating
3918	How to Make Pico de Gallo	2

In Table 9, the list of foods recommended to this user by the model is provided.

Table 9. Top 5 food recommendations user id=3918

User id	Recommended foods	Recommended prediction
3918	Yummy Korean Glass Noodles (Jap Chae)	1.961
3918	Keto Chicken-Broccoli Casserole	1.957
3918	Air Fryer French Fries	1.951
3918	Roasted Garlic-Parmesan Fingerling Potatoes	1.938
3918	Keto Margarita	1.264

To check the results, the average rating for each recommended food was calculated. When the results in Table 10 are examined, it is explained that the recommendation of these foods with low predicted values is due to the high average ratings for these meals. Therefore, the top 5 food recommendations with the highest ratings are accurate for this user.

Table 10. Mean rating of the recommended foods for user id=3918

Recommended foods	Mean rating
Yummy Korean Glass Noodles (Jap Chae)	4.92
Keto Chicken-Broccoli Casserole	4.73
Air Fryer French Fries	4.90
Roasted Garlic-Parmesan Fingerling Potatoes	4.89
Keto Margarita	4.93

When examining the most recommended food recipes and the diet types of these foods, the following results are obtained (Table 11).

Table 11. Most recommended foods and diet types

Most recommended foods	Diet type
Roasted Garlic-Parmesan Fingerling Potatoes	Vegetarian
Keto Mushroom-Stuffed Chicken Breasts	Ketogenic
Air Fryer French Fries	Low cholesterol
Keto Margarita	Ketogenic
Keto Lemon-Garlic Chicken Thighs in the Air Fryer	Ketogenic
Keto Chicken-Broccoli Casserole	Ketogenic
Easy Gluten-Free Carrot Cake	Gluten-free
Yummy Korean Glass Noodles (Jap Chae)	Vegan
Steel-Cut Oats and Quinoa Breakfast	Diabetic
No-Churn Keto Ice Cream	Ketogenic
Manhattan Clam Chowder	Diabetic
Cheesy-Crust Skillet Pizza	Gluten-free
Rebekah's Keto Egg Casserole	Ketogenic
Air Fryer Baked Potatoes	Vegetarian
Keto Cinnamon Granola	Ketogenic
Easter Grain Pie	Vegetarian
Maple Syrup Taffy	Low cholesterol
One-Pan Keto Chicken Breast with Vegetable Ragout	Ketogenic
Meyer Lemon Avocado Toast	Vegan
Keto Shrimp Scampi with Broccoli Noodles	Ketogenic

When the results are examined, it is observed that vegetarian recipes rank first, and in total, a high number of ketogenic recipes are recommended. According to the graph provided in Figure 4, both the ketogenic and vegetarian diet types were found to have high average ratings. Therefore, the recommendation system model recommending these recipes also supports exploratory data analysis results. It is believed that the high compatibility of ketogenic diet foods with many diet types has led users to choose this diet type to a large extent, which in turn affects the recommendation system.

3.4. Deployment Recommendation System Model

The recommendation system based on the ALS Model and the Cosine similarity has been deployed. The following steps were followed in the user-side deployment of the model created using the ALS technique (Table 12).

Table 12. Steps Followed in Deployment ALS Model

Step 1. Rating 5 foods from the current food list on a 5 scale
Step 2. Adding the rates of the selected foods to the data set
Step 3. Retraining the model with new rates added
Step 4. recommendForAllUsers() function to recommend foods for new users that they have not rated for before
Step 5. Displaying the recommended food list by pressing the *Recommend me!* button.

The following steps were followed in the user-side deployment of the Diet Recipes System created using the CS technique (Table 13).

Table 13. Steps in the Diet Recipes System Deployment

Step 1. The user is asked what he/she wants to eat and a one-word answer is expected from the user.
Step 2. The user is asked to choose one of the 7 diet types.
Step 3. Displaying the recommended food list by pressing the *Recommend me!* button.

3.5. Big Data Technologies and Flask

The application of ALS technique with big data was performed on the Databrick Community Edition platform. Mongo Database Atlas was used to store the data. Flask was used to develop a web application for Food Recommendation Systems. Python was used in the backend part of the application development and Javascript was used in the frontend part. Table 14 gives steps of web application in Flask.

Table 14. Web Application Steps in Flask

Step 1. Loading 7 datasets containing diet types and 7 cosine similarity functions into the Flask
Step 2. Loading the entire data set for the ALS model into the Flask
Step 3. Loading the names of the foods to be listed on the interface into the Flask
Step 4. Creating the app.py file that will contain all Python code
Step 5. Importing 7 cosine similarity functions with other libraries to be used
Step 6. Creating a style.css page in the static folder for easier visual styling of web applications
Step 7. Creating the home.html page in the templates folder to create the home page of the ALS model
Step 8. Creation of the ALS model
Step 9. Create the output.html page in the templates folder to create the page where the ALS model food recommendation list will be displayed and route the ALS model to this page
Step 10. Creation of the form_index.html page in the templates folder to create the home page of the recommendation system (Diet Recipes) created using cosine similarity
Step 11. Create the result_rec.html page in the templates folder and route the system to this page to create the food recommendation list to be obtained by cosine similarity
Step 12. Creation of back.jpg and backpic.jpg images in the static folder to be used in HTML pages to place background images

4. Conclusions

The goal of this study is to develop a food recommendation system based on the recipes and rates are given by the members of the Allrecipes.com. We used CS function for diet recommendation system and ALS for model-based recommendation analysis on this data. We aimed to discover hidden user-item relationships in matrix by using ALS model and the food which users would like were recommended. Besides, we recommend foods based on the food content similarities by using CS.

In this study, ALS method solves the sparsity and scalability problems of collaborative filtering that's why we prefer the method. At the same time, we solved the cold-start problem of collaborative filtering with the approach of rated on the foods to the users. This gave us the opportunity to get to know the user.

In the data, food name, total number of ratings, total number of reviews, total preparation and cooking time, energy value and nutrients, ingredients, recipe, user name, number of user rating and user reviews were assessed. According to the average rating, the first two most preferred foods (Colored Sugar and Mom's Sushi Rice) were categorised as "foods that children like". The least preferred foods (Pepper Salad and Overnight Slow Cooker Apple Oatmeal) were categorised as "sports meals". Considering the diet types, low-cholesterol recipes, while being the most numerous, are one of the least preferred diet types by users based on average ratings. This can be explained by the low ratings given to these foods. In terms of the total number of ratings, gluten-free recipes were received the most rating, while ketogenic recipes were received the fewest. However, it was observed that ketogenic recipes are one of the most preferred recipes based on the average rating. This is because these recipes received high ratings.

In this study, we used Big Data analytics techniques for recommendation. For this purpose, Mongo DB was utilized to store the data and ALS analysis was performed with SPARK on the Databrick Community Edition which is a cloud based big data platform. The ALS model was created with hyperparameters are found 5 for rank, 10 for maximum iteration and 0.03 for regularisation parameter. These hyperparameters are optimum for the ALS model since we used parameter grid builder to find optimum results for each hyperparameter. When the recommended foods by the model were examined on a user basis, it was found that the results were consistent. When the most recommended meals were examined, vegetarian recipes ranked first, and in total, a high number of ketogenic recipes were recommended. The fact that the recommendation system model suggests these recipes also supports the exploratory data analysis results. It is believed that the high compatibility of ketogenic diet meals with many diet types has led users to choose this diet type to a large extent, which in turn affects the recommendation system. Based on the obtained results, a web-based application was developed using Flask. The diet food recommendation application can be accessed at https://dietrec-3fdabc09c990.herokuapp.com

Consequently, we created a web-based food recommendation system that generates accurate recommendations to users who want to have ideas about foods via recipes and choose foods according to their diet.

Appendix A: The Github links

Appendix B: Web Applications Interface

References

Awan M J, Khan R A, Nobanee H, Yasin A, Anwar S , Naseem U, Singh V P (2021). A recommendation engine for predicting movie ratings using a big data approach. *Electronics*, 10(10): 1215. https://doi.org/10.3390/electronics10101215

Barakat M O S (2020) Pubmed Article Recommendation System Based On Collaborative Filtering, Master's thesis, Dokuz Eylul University, Izmir.

Bozkurt M, Acı Ç İ (2021) Öneri algoritmalarının film önerme problemi üzerinde karşılaştırılması: Movielens örneği. *Bilgisayar Bilimleri ve Teknolojileri Dergisi*, 2(2): 36-42.

Chen J, Fang J, Liu W, Tang T, Chen X, Yang C. (2017) Efficient And Portable Als Matrix Factorization For Recommender Systems. 2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp.409-418.

Gündoğan, E, Kaya, M (2021) Bilimsel dergi tavsiyesi için içerik tabanlı bir yaklaşım. *Computer Science*, (Special): 41-47.

Han, J., Kamber, M., & Pei, J. (2012, January). Getting to know your data. In Data mining (Vol. 3, pp. 39-82). Boston, MA: Morgan Kaufmann. A chapter in the book, Data Mining (Third Edition) The Morgan Kaufmann Series in Data Management Systems https://doi.org/10.1016/B978-0-12-381479-1.00002-2.

Jiang J, Li W, Dong A, Gou Q, Luo X (2020) A fast deep autoencoder for high-dimensional and sparse matrices in recommender systems. *Neurocomputing*, 412: 381-391.

Kaya TS (2019) Veri Madenciliği Algoritmaları İle Kredi Kartı Kullanım Alışkanlıklarının Incelenmesi Ve Kişiye Özgü Kampanya Teklifi, Master's Thesis. İstanbul Üniversitesi.

Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30-37.

Lawson R. (2015) Web Scraping With Python, Packt Publishing Ltd.

Li JB, Lin SY, Hsu YH, Huang YC. (2018) Implementation Of An Alternating Least Square Model Based Collaborative Filtering Movie Recommendation System On Hadoop And Spark Platforms. International Conference on Broadband and Wireless Computing, Communication and Applications, pp.237-249.

Li S, McAuley J (2020) Recipes for Success: Data Science in the Home Kitchen. Harvard Data Science Review, 2. https://assets.pubpub.org/nzhfriaw/ca2af84f-38f9-48c4-8b38-5cfe915a7b7e.pdf. Accessed 02 Nov 2022

Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet computing*, 7(1): 76-80.

Mitchell R. (2018) Web Scraping With Python: Collecting More Data From The Modern Web, O'Reilly Media, Inc.

Nguyen W. (2021) A Literature Review of Collaborative Filtering Recommendation System using Matrix Factorization algorithms. In Proceedings of ACM Conference (Conference'17). ACM.

Oh Y, Choi A, Woo W (2010) u-BabSang: a context-aware food recommendation system. *The Journal of Supercomputing*, 54: 61-81.

Özcan İ, Çelik M. (2018) Developing Recommendation System Using Genetic Algorithm Based Alternative Least Squares. 2018 International Conference on Artificial Intelligence and Data Processing, pp.1-5.

Philip S, Shola P, Ovye A (2014) Application of content-based approach in research paper recommendation system for a digital library. *International Journal of Advanced Computer Science and Applications*, 5(10): 37-40.

Sarwar B, Karypis G, Konstan J, Riedl J (2001) Item-Based Collaborative Filtering Recommendation Algorithms. *In Proceedings of the 10th international conference on World Wide Web*: 285-295.

Schafer J B, Frankowski D, Herlocker J, Sen S. (2007) Collaborative Filtering Recommender Systems. In The Adaptive Web, Springer, Berlin, Heidelberg.

Fathollahi M S, Razzazi F (2021) Music similarity measurement and recommendation system using convolutional neural networks. *International Journal of Multimedia Information Retrieval*, 10(1): 43-53.

Thorat P B, Goudar R M and Barve S (2015) Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 110(4): 31-36.

Uluyağmur M (2012) Hibrit Film Öneri Sistemi, Master's thesis, Istanbul Teknik Üniversitesi.

Ünaldı, S S (2022) Veri Seyrekliği ve Ölçeklenebilirlik Problemlerini Gidermek İçin Derin Otomatik Kodlayıcı Tabanlı Yeni Bir Tavsiye Sistemi Modeli, Master's thesis, Necmettin Erbakan Üniversitesi.

Vairale VS, Shukla S. (2021) Recommendation Of Food Items For Thyroid Patients Using Content-Based Knn Method. Data Science and Security: Proceedings of IDSCS 2020, pp.71-77.

Xie L, Zhou W, Li Y (2016) Application of improved recommendation system based on spark platform in big data analysis. *Cybernetics and Information Technologies*, 16(6): 245-255.

Appendix A:

In this study, the codes were written in Python Programming Language. In Flask, we used the syntax to create web applications of recommendation analyses. For the purposes of the study, first part of the syntax is for Diet Food Recommendation System performed by cosine similarity technique. Second part of the syntax is for Food Recommendation System implemented by using ALS technique. This part is a slightly confusing but it's extremely readable.

You can access full codes and details by visiting our GitHub account: https://github.com/aarare/dietreconly

and https://github.com/aarare/recfoodonly

Moreover, you can visit our Tableau account of the recommendation analysis dashboard: https://public.tableau.com/app/profile/merve.cengiz/viz/DashboardofFoodRecommendationAnalysis/Dashboard1

Appendix B:

For our purposes on applications, we mainly interested in the design of interfaces of both recommendation analyses. The interface of the ALS Model is shown in the figure below (Figure 7). In the Food Recommendation System, users can select and rate desired foods from the drop-down menu, and when they click the "Recommend me!" button, a list of recommended foods is displayed (Figure 8). This application runs in the localhost. But for diet food recommendation system, we created a web application for everyone to access.

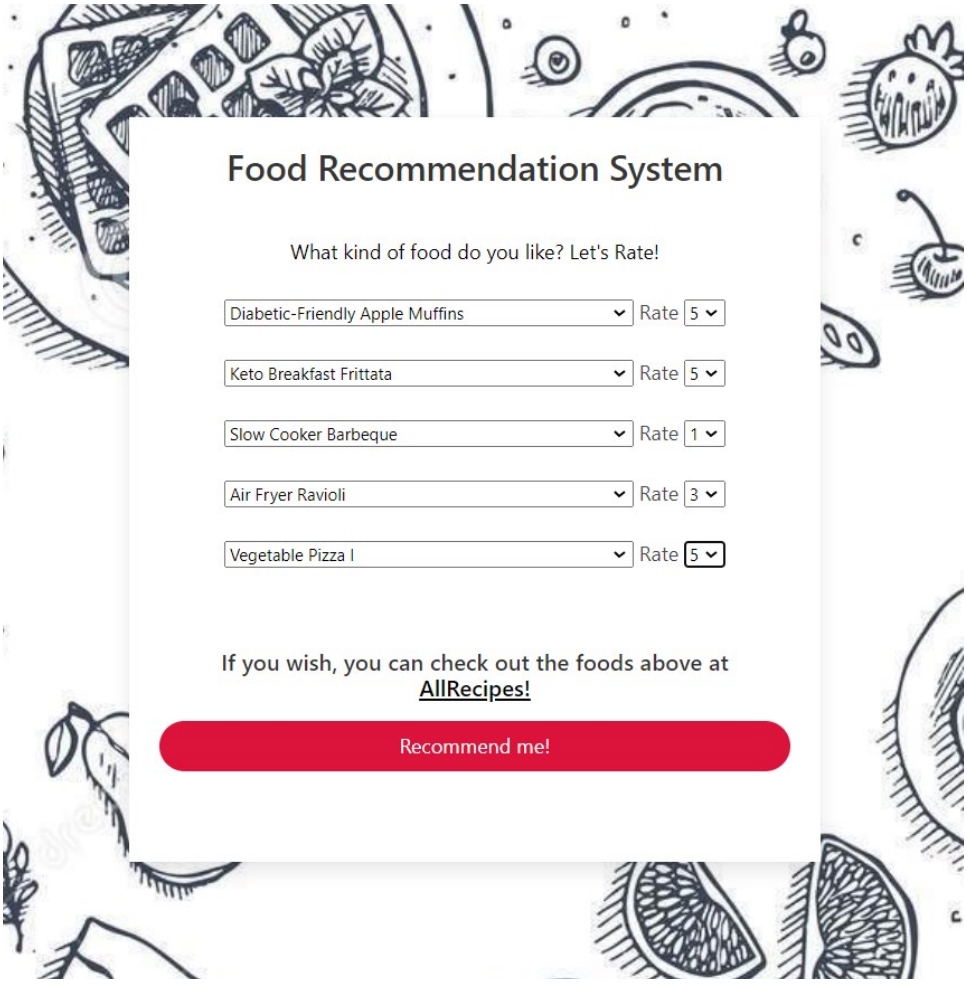
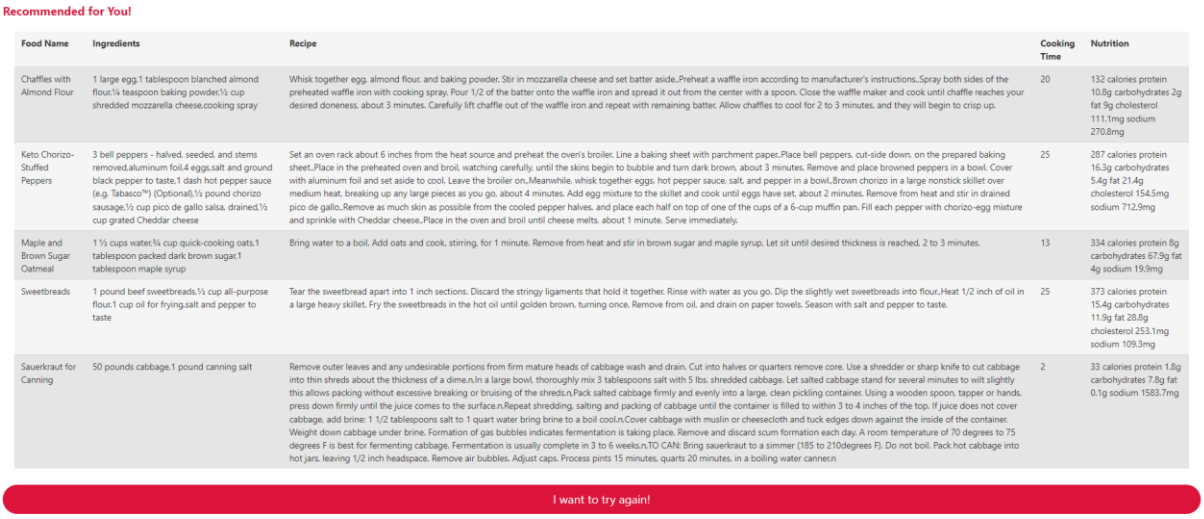


Figure 7. Food Recommendation System Interface



Recommended for You!

Food Name	Ingredients	Recipe	Cooking Time	Nutrition
Chaffles with Almond Flour	1 large egg,1 tablespoon blanched almond flour,¼ teaspoon baking powder,½ cup shredded mozzarella cheese,cooking spray	Whisk together egg, almond flour, and baking powder. Stir in mozzarella cheese and set batter aside.,Preheat a waffle iron according to manufacturer's instructions.,Spray both sides of the preheated waffle iron with cooking spray. Pour 1/2 of the batter onto the waffle iron and spread it out from the center with a spoon. Close the waffle maker and cook until chaffle reaches your desired doneness, about 3 minutes. Carefully lift chaffle out of the waffle iron and repeat with remaining batter. Allow chaffles to cool for 2 to 3 minutes, and they will begin to crisp up.	20	132 calories protein 10.8g carbohydrates 2g fat 9g cholesterol 111.1mg sodium 270.8mg
Keto Chorizo-Stuffed Peppers	3 bell peppers - halved, seeded, and stems removed,aluminum foil,4 eggs,salt and ground black pepper to taste,1 dash hot pepper sauce (e.g. Tabasco™) (Optional),½ pound chorizo sausage,½ cup pico de gallo salsa, drained,½ cup grated Cheddar cheese	Set an oven rack about 6 inches from the heat source and preheat the oven's broiler. Line a baking sheet with parchment paper.,Place bell peppers, cut-side down, on the prepared baking sheet.,Place in the preheated oven and broil, watching carefully, until the skins begin to bubble and turn dark brown, about 3 minutes. Remove and place browned peppers in a bowl. Cover with aluminum foil and set aside to cool. Leave the broiler on.,Meanwhile, whisk together eggs, hot pepper sauce, salt, and pepper in a bowl.,Brown chorizo in a large nonstick skillet over medium heat, breaking up any large pieces as you go, about 4 minutes. Add egg mixture to the skillet and cook until eggs have set, about 2 minutes. Remove from heat and stir in drained pico de gallo.,Remove as much skin as possible from the cooled pepper halves, and place each half on top of one of the cups of a 6-cup muffin pan. Fill each pepper with chorizo-egg mixture and sprinkle with Cheddar cheese.,Place in the oven and broil until cheese melts, about 1 minute. Serve immediately.	25	287 calories protein 16.3g carbohydrates 5.4g fat 21.4g cholesterol 154.5mg sodium 712.9mg
Maple and Brown Sugar Oatmeal	1 ½ cups water,¾ cup quick-cooking oats,1 tablespoon packed dark brown sugar,1 tablespoon maple syrup	Bring water to a boil. Add oats and cook, stirring, for 1 minute. Remove from heat and stir in brown sugar and maple syrup. Let sit until desired thickness is reached, 2 to 3 minutes.	13	334 calories protein 8g carbohydrates 67.9g fat 4g sodium 19.9mg
Sweetbreads	1 pound beef sweetbreads,½ cup all-purpose flour,1 cup oil for frying,salt and pepper to taste	Tear the sweetbread apart into 1 inch sections. Discard the stringy ligaments that hold it together. Rinse with water as you go. Dip the slightly wet sweetbreads into flour.,Heat 1/2 inch of oil in a large heavy skillet. Fry the sweetbreads in the hot oil until golden brown, turning once. Remove from oil, and drain on paper towels. Season with salt and pepper to taste.	25	373 calories protein 15.4g carbohydrates 11.9g fat 28.8g cholesterol 253.1mg sodium 109.3mg
Sauerkraut for Canning	50 pounds cabbage,1 pound canning salt	Remove outer leaves and any undesirable portions from firm mature heads of cabbage wash and drain. Cut into halves or quarters remove core. Use a shredder or sharp knife to cut cabbage into thin shreds about the thickness of a dime.n,In a large bowl, thoroughly mix 3 tablespoons salt with 5 lbs. shredded cabbage. Let salted cabbage stand for several minutes to wilt slightly this allows packing without excessive breaking or bruising of the shreds.n,Pack salted cabbage firmly and evenly into a large, clean pickling container. Using a wooden spoon, tapper or hands, press down firmly until the juice comes to the surface.n,Repeat shredding, salting and packing of cabbage until the container is filled to within 3 to 4 inches of the top. If juice does not cover cabbage, add brine: 1 1/2 tablespoons salt to 1 quart water bring brine to a boil cool.n,Cover cabbage with muslin or cheesecloth and tuck edges down against the inside of the container. Weight down cabbage under brine. Formation of gas bubbles indicates fermentation is taking place. Remove and discard scum formation each day. A room temperature of 70 degrees to 75 degrees F is best for fermenting cabbage. Fermentation is usually complete in 3 to 6 weeks.n,TO CAN: Bring sauerkraut to a simmer (185 to 210degrees F). Do not boil. Pack hot cabbage into hot jars, leaving 1/2 inch headspace. Remove air bubbles. Adjust caps. Process pints 15 minutes, quarts 20 minutes, in a boiling water canner.n	2	33 calories protein 1.8g carbohydrates 7.8g fat 0.1g sodium 1583.7mg

I want to try again!

Figure 8. A list of recommended foods for new user

In Diet Food Recommendation System (Figure 9), for interactive application for users you can directly type one word then choose a diet type of your choice below. There must be a one word in the field, for instance please try the followings; banana, bread, chocolate etc. When you click the "Recommend me!" button, a list of recommended foods is displayed (Figure 10). For now, just be aware that since the size of the data is large the response time will be long. The diet food recommendation application can be accessed at https://dietrec-3fdabc09c990.herokuapp.com

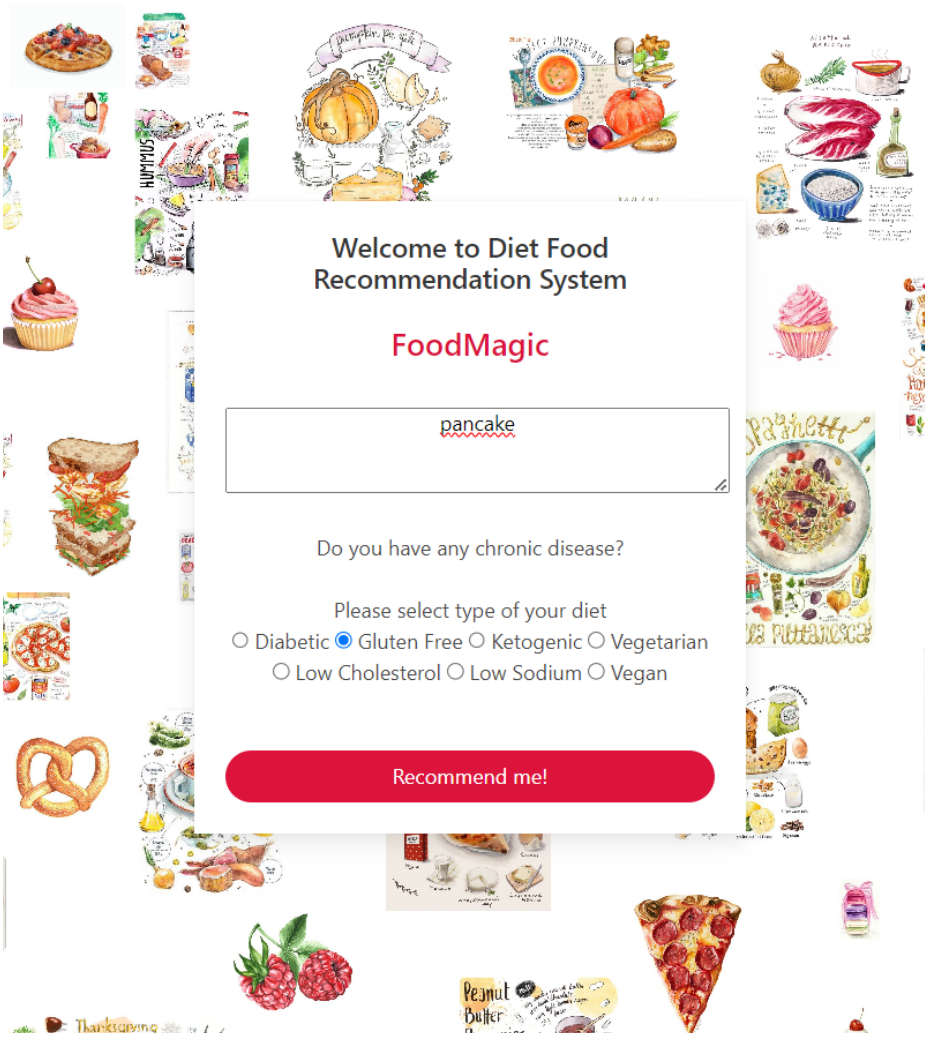
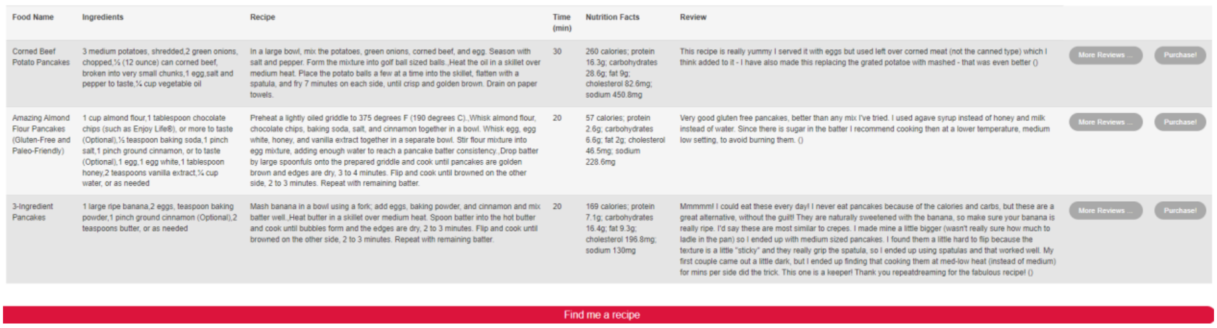


Figure 9. Diet Food Recommendation System Interface



Food Name	Ingredients	Recipe	Time (min)	Nutrition Facts	Review		
Corned Beef Potato Pancakes	3 medium potatoes, shredded,2 green onions, chopped,½ (12 ounce) can corned beef, broken into very small chunks,1 egg,salt and pepper to taste,¼ cup vegetable oil	In a large bowl, mix the potatoes, green onions, corned beef, and egg. Season with salt and pepper. Form the mixture into golf ball sized balls.,Heat the oil in a skillet over medium heat. Place the potato balls a few at a time into the skillet, flatten with a spatula, and fry 7 minutes on each side, until crisp and golden brown. Drain on paper towels.	30	260 calories; protein 16.3g; carbohydrates 28.6g; fat 9g; cholesterol 82.6mg; sodium 450.8mg	This recipe is really yummy I served it with eggs but used left over corned meat (not the canned type) which I think added to it - I have also made this replacing the grated potatoe with mashed - that was even better ()	More Reviews ...	Purchase!
Amazing Almond Flour Pancakes (Gluten-Free and Paleo-Friendly)	1 cup almond flour,1 tablespoon chocolate chips (such as Enjoy Life®), or more to taste (Optional),½ teaspoon baking soda,1 pinch salt,1 pinch ground cinnamon, or to taste (Optional),1 egg,1 egg white,1 tablespoon honey,2 teaspoons vanilla extract,¼ cup water, or as needed	Preheat a lightly oiled griddle to 375 degrees F (190 degrees C).,Whisk almond flour, chocolate chips, baking soda, salt, and cinnamon together in a bowl. Whisk egg, egg white, honey, and vanilla extract together in a separate bowl. Stir flour mixture into egg mixture, adding enough water to reach a pancake batter consistency.,Drop batter by large spoonfuls onto the prepared griddle and cook until pancakes are golden brown and edges are dry, 3 to 4 minutes. Flip and cook until browned on the other side, 2 to 3 minutes. Repeat with remaining batter.	20	57 calories; protein 2.6g; carbohydrates 6.6g; fat 2g; cholesterol 46.5mg; sodium 228.6mg	Very good gluten free pancakes, better than any mix I've tried. I used agave syrup instead of honey and milk instead of water. Since there is sugar in the batter I recommend cooking then at a lower temperature, medium low setting, to avoid burning them. ()	More Reviews ...	Purchase!
3-Ingredient Pancakes	1 large ripe banana,2 eggs, teaspoon baking powder,1 pinch ground cinnamon (Optional),2 teaspoons butter, or as needed	Mash banana in a bowl using a fork; add eggs, baking powder, and cinnamon and mix batter well.,Heat butter in a skillet over medium heat. Spoon batter into the hot butter and cook until bubbles form and the edges are dry, 2 to 3 minutes. Flip and cook until browned on the other side, 2 to 3 minutes. Repeat with remaining batter.	20	169 calories; protein 7.1g; carbohydrates 16.4g; fat 9.3g; cholesterol 196.8mg; sodium 130mg	Mmmmm! I could eat these every day! I never eat pancakes because of the calories and carbs, but these are a great alternative, without the guilt! They are naturally sweetened with the banana, so make sure your banana is really ripe. I'd say these are most similar to crepes. I made mine a little bigger (wasn't really sure how much to ladle in the pan) so I ended up with medium sized pancakes. I found them a little hard to flip because the texture is a little "sticky" and they really grip the spatula, so I ended up using spatulas and that worked well. My first couple came out a little dark, but I ended up finding that cooking them at med-low heat (instead of medium) for mins per side did the trick. This one is a keeper! Thank you repeatdreaming for the fabulous recipe! ()	More Reviews ...	Purchase!

Find me a recipe

Figure 10. A list of recommended foods