NUTRIAI: Nutrition Understanding and Tracking Recommendations using Intelligent Algorithms

Jun Jie Lim 1, a), Hui-Ngo Goh 1, b), Choo-Yee Ting 1, c) and Albert Quek 1, d)

1Faculty of Computing and Informatics, Multimedia University, 63000 Cyberjaya, Selangor Malaysia

*b) Corresponding author: hngoh@mmu.edu.my  
a)1201103290@student.mmu.edu.my*

*c)cyting@mmu.edu.my*

*d)quek.albert@mmu.edu.my*

**Abstract.** Dieting is important to maintain a healthy lifestyle. However, the restrictive guidelines and lack of appealing food choices creates a challenging for dieting. Hence, this presents a need for a personalized food recommendation system that balances the healthiness of the food recommended and food preferences. The food recommendation system proposed recommends food through food rating classification model, and multiple nutritional guidelines to ensure healthiness of recipe for various health conditions. personalized recommendation is done by comparing the similarity of each recipe with foods consumed by user. The features for training recipe rating classification model were selected based on the strength of their correlation with recipe rating. Six derived features deriving from existing numerical features were created due to weak correlation of 13 numerical features with rating. The six derived features were found to have strong correlation with rating. This reduces the number of numerical features for model training from 13 to six. The experimental results showed that XGBoost model with default parameters is the best model for rating classification. Explainable Artificial Intelligence (XAI) was used to advise and explain the healthiness of recommended recipes in accordance with user health information. In extension of this study, time-awareness can be implemented such that the recommendation system can track nutrients consumed by user over time. Furthermore, more features such as images or user reviews can be considered for model training.

# INTRODUCTION

Dieting is important for a healthy lifestyle as it offers health benefits. However, adhering to a diet can often be challenging due to restrictive guidelines and the lack of appealing food choices. This often results in deviations of diet plans or cessation of the dietary regimen which increases the risk of health complications and diseases due to malnutrition. According to the World Health Organization (WHO)[1], millions of adults and children were diagnosed with obesity in 2022 due to malnutrition. This problem presents a need for a personalized food recommendation system that takes food preferences and nutritional guidelines into consideration.

Food recommendation systems were categorised into four categories. These categories are collaborative filtering (CF), content-based filtering (CBF), rule-based filtering (RBF), and hybrid filtering. CF suggests items suggests items based on the preference of other like-minded users while CBF suggests items based on the similarity of items to the items previously preferred by user [2]. RBF suggests items based on a set of predefined rules [3]. For example, a rule stating that all foods must have ingredients aligned with user preferences will only recommend foods that does not violate the defined rule. Hybrid filtering is a combination of the filtering methods mentioned previously [2].

Based on previous studies of food recommendation systems, the most common filtering methods used is CF [4-10] indicating the effectiveness of food recommendation though past behaviour of other users. Furthermore, CF tends to be used with CBF in hybrid filtering [4], [6], [10]. This is done to overcome the cold-start problem faced by CF due to the lack of user-item interaction data [2]. Furthermore, the combination of CF and CBF allow more personalized recommendations. Most studies focusing on recommending healthy food tend to use RBF [4,6,7,10]. Models for food recommendation consist of neural networks [4], [5], [7], [8] and Singular Value Decomposition (SVD) [6], [9], [10] were commonly used in food recommendation systems. The common gaps faced by most studies is the lack of features to train their model [4-6], [8-10]. Furthermore, studies utilizing neural networks [4], [5], [7], [8] lack the necessary volume of data needed to train their model.

This study aims to develop a dietary recommendation system utilizing machine learning algorithms to capture food preferences. To achieve this objective, the methods used in this study include data collection, data cleaning, data analysis, data pre-processing, feature selection, feature engineering, model evaluation, hyperparameter fine-tuning, and user interface development. The expected outcome after achieving the objective is a food recommendation system that recommends food based on food preferences and dietary needs. This study serves significance to the health of society by increasing adherence to dietary guidelines which can improve the overall health of individuals through personalized dietary recommendation of food that caters to individual food preferences.

# METHODOLOGY

## Dataset Characteristics

The dataset used for this study is the Food.com Recipes and Interactions dataset [19]. This publicly available dataset is sourced from Kaggle. This dataset contains data related to user rating and food recipes. This dataset contains eight different files and a total of 49 features. The dataset has more than 180 thousand recipes and more than 700 reviews over the span of 18 years. The dataset was chosen due to the sufficient number of features and volume for this study. Although the dataset contains eight different files, only three files were selected for this study. These files are RAW\_recipes.csv, interactions\_train.csv, and interactions\_test.csv. The files contain features related to food recipes and food rating given by users such as ingredients, cooking steps, amount of nutrients, and rating.

## Framework

The food recommendation system needs to suggest foods based on user food preferences while ensuring the suggested foods align with dietary needs and restrictions of the user. The proposed framework for the recommendation can be seen in Figure 1.

|  |
| --- |
|  |

FIGURE 1. Proposed framework of dietary recommendation system

The framework consists of four sections (Section A, Section B, Section C, Section D) where each section serves a distinct function in the recommendation system. The system outputs a list of recommended recipes aligned with dietary guidelines based on user input.

Section A is a CF unit that generates the average predicted rating for recipes. It consists of a rating classification model trained on a user-recipe interaction training set to predict the recipe rating. After rating classification, average of predicted ratings from the classification model was calculated for each recipe. The output of this section is a list of recipes with average predicted rating.

Section B is an RBF unit that ensures all recipes align with dietary guidelines and diet restrictions. This ensures the safe consumption of recipes for users with specific health diseases. This section removes recipes that do not align with diet restrictions of users with high blood pressure, obesity, and Type 2 diabetes. Recipe removal is done by removing recipes with excessive or insufficient amount of nutrients based on diet restrictions for specific health diseases. The diet restrictions for the diseases mentioned were obtained from medical institutions or research papers by medical experts [11-17]. After the removal of recipes, daily nutritional intake was calculated based on nutritional

guideline by [18]. The daily nutritional intake was then used to calculate the nutritional deficiency score which represents the overall lack or excessiveness of nutrients in a recipe. The output of this section are recipes with nutritional deficiency score.

Section C is a CBF unit that compares the similarity between recipes, and recipes from the user-recipe interaction history. The user-recipe interaction history contains recipes with past user interactions. This allows more personalized

recommendations towards based on food preferences of a specific user. The output after comparisons are recipes with similarity score. If there are no recipes from the user-recipe interaction history, no comparison between recipes and user interacted recipes is done. In this case, the output are recipes without similarity score.

Section D combines the output of all pervious sections by calculating a recipe score for each recipe. The resulting recipe score of each recipe will be used for recommendation. All recipes are then sorted in a descending order based on their recipe score before recommending the top-k recipes to the user. Explainable Artificial Intelligence (XAI) was then used to explain the healthiness of recipes and provide health advice regarding consumption of recipes.

## Exploratory Data Analytics

Before proceeding with further data analysis, random stratified under-sampling was applied due to imbalanced classes in rating. Under-sampling reduced the number of rows from 698901 rows to 20046 rows. This is to ensure accurate analysis through unbiased distribution of rating. A correlation heatmap in Figure 2 shows the correlation between all numerical features.

|  |
| --- |
| A chart with red and blue squares  Description automatically generated |

FIGURE 2. Correlation heatmap of numerical features

The numerical features shown in the heatmap are related to rating, recipe publication date, recipe preparation time in minutes, number of cooking steps, number of ingredients required, and amount of nutrients. The results in the correlational heatmap indicates no correlation between all numerical features and rating. This means that numerical features do not capture user food preferences. Numerical features associated with amount of nutrients are strongly correlated with each other. This is due to the relationship between nutrients in a nutritional context.

The text features ingredients and cooking steps were analysed using word cloud. Stop words removal and lemmatization were applied on text features to remove stop words and group inflected form of texts together. This allows the analysis to be focused on meaningful words with distinct meaning. The word clouds in Figure 3 shows the most common ingredients and cooking steps across ratings. The size of word corresponds with the frequency of an ingredient or a cooking step.

There are some distinctions in common ingredients across ratings. For instance, “water” is an ingredient more common in recipes with rating of zero and one but rarer in recipes with rating greater than one. Although there are similarities in ingredients across ratings, significant differences between them suggests the association between ingredients and rating. There are more similarities in common cooking steps across ratings with slight differences. For example, “minute” is most common across rating. However, “cook” is more common from rating zero to three and “add” is most common in rating one. This suggests cooking step still has a significant association with rating.

|  |  |
| --- | --- |
| A collage of different colored words  Description automatically generated  A collage of different colored words  Description automatically generated | A collage of words  Description automatically generated  A collage of words  Description automatically generated |
| (a) | (b) |

FIGURE 3. Word cloud of ingredients (a) and cooking steps (b) across ratings

## Feature Engineering

Due to the lack of numerical features correlating to rating, new features are created based on existing numerical features. The features created are as listed from Equation (1) to (6).

* Method Satisfaction

The satisfaction of user towards the number of steps needed to prepare a recipe.

|  |  |
| --- | --- |
|  | (1) |

* Ingredient Satisfaction

The satisfaction of user towards total number of ingredients in a recipe. Total ingredients refer to the total number of ingredients required to prepare the recipe.

|  |  |
| --- | --- |
|  | (2) |

* Time Satisfaction

The satisfaction of user towards amount of time needed to prepare a recipe. Minutes to prepare refers to the time in minutes required to prepare the recipe.

|  |  |
| --- | --- |
|  | (3) |

* Saturated Fat Satisfaction

The satisfaction of user towards the density of saturated fat in a recipe which derives from the ratio of saturated fat to total fat in a recipe.

|  |  |
| --- | --- |
|  | (4) |

* Carb Density Satisfaction

The satisfaction of user towards the density of carbohydrates in a recipe.

|  |  |
| --- | --- |
|  | (5) |

* Fat Density Satisfaction

The satisfaction of user towards the density of fat in a recipe.

|  |  |
| --- | --- |
|  | (6) |

The correlation between derived features and rating were analysed using correlation heatmap contained in Figure 4.

|  |
| --- |
| A screenshot of a graph  Description automatically generated |

FIGURE 4. Correlation heatmap of derived numerical features with rating

All derived features have strong correlation with rating. There are strong correlations between derived features due to the similarity in their derived features. This indicates the derived features strongly captured food preferences. However, most correlation between the derived features are weaker than their correlation to rating. Although there were 13 numerical features, the six derived features will be used for model evaluation. This is because the 13 numerical features have weak correlation with rating.

# RESULTS AND DISUCSSION

## Model Evaluation

Although there are 49 features in the dataset, eight features were used for model training. These features were six derived features along with ingredients and cooking steps. Synthetic Minority Over-sampling Technique (SMOTE) was used to balance classes of rating for training set and testing set. All numerical features were normalized for model training. All text features undergo stop words removal and lemmatization for model training. Text features were also vectorized using Bidirectional Encoder Representations from Transformers (BERT). Singular Value Decomposition (SVD) was used to reduce the number of dimensions generated by BERT with cumulative explained variance of 0.9. This ensures the number of SVD components capture 90% of textual information. The 768 dimensions generated by BERT was reduced to 182 dimensions. Numerical features were then concatenated with BERT word embeddings for model training.

The training set was split into 80% for training and 20% for validation while 20% of testing set was used for testing. Take note that both training set and testing set contains the same number of rows. Since there is multicollinearity in the training set, tree-based algorithms were chosen for training due to their robustness against multicollinearity [20]. The evaluation results on the tree-based algorithms can be seen in Table 1.

TABLE 1. Evaluation Results of Tree-based Algorithms in Validation Set

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Validation Set | | | |  | Testing Set | | | |
| Algorithm | **Accuracy** | **Precision** | **Recall** | **F1-Score** |  | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| LightGBM | 0.97 | 0.97 | 0.97 | 0.97 |  | 0.73 | 0.74 | 0.73 | 0.74 |
| Random Forest | 0.77 | 0.76 | 0.77 | 0.76 |  | 0.73 | 0.73 | 0.73 | 0.73 |
| Decision Tree | 0.80 | 0.80 | 0.80 | 0.80 |  | 0.68 | 0.69 | 0.67 | 0.68 |
| XGBoost | 0.97 | 0.97 | 0.97 | 0.97 |  | **0.89** | **0.89** | **0.89** | **0.89** |
| HistGradientBoostingClassifier | **0.98** | **0.98** | **0.98** | **0.98** |  | 0.73 | 0.73 | 0.73 | 0.73 |

HistGradientBoostingClassifier has the best performance in the validation set among other algorithms. However, its performance across all metrics dropped by 25% in the testing set. This is due to poor generalization which signals overfitting. Despite Decision Tree achieving high scores across the metrics in validation set, its performance is the lowest among other algorithms in both validation set and testing set. Although both XGBoost and LightGBM performed equally well in validation set, XGBoost has better performance than LightGBM in testing set. This shows that XGBoost can generalize better than LightGBM. Although Random Forest generalizes the best among other algorithms, its overall performance across metrics is worse than XGBoost. Although there are signs of slight overfitting, XGBoost has better overall performance than other algorithms due to its high and consistent scoring across all metrics in both validation set and testing set.

## Health Evaluation Through Explainable Artificial Intelligence

ChatGPT was chosen to be used as XAI in evaluating the healthiness of recommended recipes. ChatGPT was given information about ingredients and nutritional value of recipe along with user health profile. Then, ChatGPT was prompted to state the healthiness of the recipe and the contribution of each ingredient towards each nutrient. In a list of recommended recipes, the top recipe for obesity, Type 2 diabetes, and high blood pressure were selected for health evaluation. Top recommended recipe for healthy individual was selected as well for comparison. The health evaluation by ChatGPT for the selected recipes can be seen in Figure 5.

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |
|  |  |
| (c) | (d) |

FIGURE 5. Health evaluation by ChatGPT on top recipe recommended for healthy individual (a), individual with obesity (b), individual with Type 2 diabetes (c), and individual with high blood pressure (d).

ChatGPT acknowledges the health condition and offer advice for healthy diet. Furthermore, ChatGPT pointed out the contribution of ingredients towards nutrients for each recipe with relevance to each health condition. Although ChatGPT raised concerns about high sodium content in the recipe for high blood pressure, the sodium content was still under the healthy limit for individuals with high blood pressure [11]. ChatGPT’s response to top recommended recipe for healthy person focuses on maintaining overall health. In contrast, its response to top recommended recipes for diseased individuals focuses on ingredients and nutritional aspects that may affect their health condition. Based on ChatGPT response, the top recommended recipes were considered healthy for individuals with obesity, high blood pressure, and Type 2 diabetes.

# CONCLUSION

In conclusion, the satisfaction of user towards amount of nutrients in food, and time needed to prepare recipe along with cooking steps, and ingredients have strong association with rating. This is because they can capture food preferences. Furthermore, XGBoost model is the best model for rating classification. This is due to the model’s high and consistent performance in rating classification. In addition, recipes recommended for different health conditions are healthy as evaluated using ChatGPT. Finally, ChatGPT was able to acknowledge health conditions when evaluating healthiness of food while listing contributions of ingredients towards nutrients in recipes. In extension to this study, more health conditions can be included for the food recommendation system. Furthermore, the recommendation system can be made time-aware by tracking nutrients consumed by user over time. Additionally, more features such as images or user reviews can be considered for model training.

# REFERENCES

1. World Health Organization: WHO, “Fact sheets - Malnutrition,” (2024).
2. Ryngksai and L. Chameikho, “Recommender Systems: Types of filtering techniques,” International Journal of Engineering Research **3**(11), 251–254 (2014).
3. B. Apolloni, S. Bassis, M. Mesiti, S. Valtolina, and F. Epifania, “A Rule Based Recommender System,” in *Advances in Neural Networks*, edited by S. Bassis, A. Esposito, F. C. Morabito, and E. Pasero (Springer International Publishing, Cham, 2016), pp. 87–96.
4. S. Hamdollahi Oskouei and M. Hashemzadeh, “FoodRecNet: a comprehensively personalized food recommender system using deep neural networks,” Knowledge and Information Systems **65**(9), 3753–3775 (2023).
5. J. Zhang, Z. Wang, W. Liu, X. Liu, and Q. Zheng, “A unified approach to designing sequence-based personalized food recommendation systems: tackling dynamic user behaviors,” International Journal of Machine Learning and Cybernetics **14**(9), 2903–2912 (2023).
6. P. Chavan, B. Thoms, and J. Isaacs, “A Recommender System for Healthy Food Choices: Building a Hybrid Model for Recipe Recommendations using Big Data Sets,” (2021).
7. A. Morales-Garzón, K. Gutiérrez-Batista, and M. J. Martin-Bautista, “Link prediction in food heterogeneous graphs for personalised recipe recommendation based on user interactions and dietary restrictions,” Computing **106**(7), 2133–2155 (2024).
8. A. A. Al-Asadi and M. N. Jasim, “Dietary Behavior Based Food Recommender System Using Deep Learning and Clustering Techniques,” World Journal of Computer and Management Sciences **2**(2), 1–8 (2023).
9. M. B. S. Siddik and A. T. Wibowo, “Collaborative Filtering Based Food Recommendation System Using Matrix Factorization,” Mib **7**(3), 1041 (2023).
10. A. Starke, C. Trattner, H. Bakken, M. Johannessen, and V. Solberg, “The cholesterol factor: Balancing accuracy and health in recipe recommendation through a nutrient-specific metric,” edited by H. Abdollahpouri, M. Elahi, and M. Mansoury (CEUR-WS.org, Amsterdam, The Netherlands, 2021).
11. American Heart Association, “Why should I limit sodium?,” (2024).
12. UF Health, “High Blood Pressure and Diet,” (n.d.).
13. J. Y. Kim, “Optimal Diet Strategies for Weight Loss and Weight Loss Maintenance,” Journal of Obesity & Metabolic Syndrome **30**(1), 20–31 (2021).
14. B. L. Strom, A. L. Yaktine, and M. Oria, editors, *Sodium Intake in Populations: Assessment of Evidence, Online-Ausg* (The National Academies Press, Washington, District of Columbia, 2013).
15. A. P. Campbell and T. M. Rains, “Dietary Protein Is Important in the Practical Management of Prediabetes and Type 2 Diabetes,” The Journal of Nutrition **145**(1), 164S–169S (2015).
16. American Diabetes Association, “Diabetes Meal Patterns: Science-Based Nutrition Plans for Weight Loss & Management,” (n.d.).
17. World Health Organization: WHO, “WHO guideline: sugar consumption recommendation,” (n.d.).
18. J. J. Otten, J. P. Hellwig, and L. Meyers, *Dietary Reference Intakes: The Essential Guide to Nutrient Requirements* (National Academies Press, Washington, 2006).
19. B. P. Majumder, S. Li, J. Ni, and J. McAuley, “Generating Personalized Recipes from Historical User Preferences,” (2019).
20. S. Chowdhury, Y. Lin, B. Liaw, and L. Kerby, “Evaluation of Tree Based Regression over Multiple Linear Regression for Non-normally Distributed Data in Battery Performance,” in *2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)* (IEEE, San Antonio, TX, USA, 2022), pp. 17–25.