The Scientific Temper
SEPTEMBER 2026 Vol. 17 (9): 6971-6980
ISSN: 0976-8653, E-ISSN: 2231-6396
DOI: 10.58414/SCIENTIFICTEMPER.2026.17.9.2557
A WEB OF SCIENCE JOURNAL

RESEARCH ARTICLE

Nutrient-Aware Food Recommendation for Diabetes Using Item–Item Collaborative Filtering

S. BALASUBRAMANIAN[1]*, DR. K.N. ABDUL KADER NIHAL[2]

[1]*Research Scholar, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620020, India*

[2]*Assistant Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620020, India*

***Corresponding Author: lakshibala@gmail.com**

ABSTRACT

Diet management is an important part of diabetes care, but food recommendation is more demanding than ordinary item recommendation because a useful suggestion must satisfy both user preference and nutritional constraints. This study develops an item–item collaborative filtering (CF) model for diabetes-oriented food recommendation by combining rule-based nutrient filtering with cosine-similarity-based collaborative prediction. A nutrient composition dataset containing 12,044 food records and 29 raw attributes was cleaned and mapped to canonical nutrient fields. After removal of malformed records and duplicate food names, 1,781 unique foods remained; a sugar threshold of 12 g per reported unit reduced the candidate catalogue to 1,779 items. Because public data linking diabetic-user food preferences with detailed nutrient profiles were not available, 100 synthetic users and a controlled 1–5 rating matrix were generated from nutrient-weighted preference profiles with Gaussian noise. The resulting matrix contained 21,203 observed ratings at 88.08% sparsity. An 80:20 per-user train/test split was used, and item–item cosine similarity with an eight-neighbourhood was applied for prediction and Top-10 recommendation. Among 4,200 held-out ratings, predictions were available for 1,640 user–item pairs, producing RMSE = 0.9463 and MAE = 0.7101. Across 100 users, Precision@10 = 0.0091, Recall@10 = 0.0075, and HitRate@10 = 0.0909. The representative Top-10 list was dominated by low-recorded-sugar, protein- and fibre-rich foods. The findings support the feasibility of the proposed pipeline as a reproducible baseline, while the modest ranking performance, synthetic ratings, sparse matrix, mean-imputation strategy, and incomplete source fields limit clinical interpretation and motivate validation with real preference data.

Keywords: *collaborative filtering, item–item similarity, cosine similarity, food recommender system, diabetes diet management, nutrient-aware recommendation, synthetic rating data.*

INTRODUCTION

Diabetes mellitus is a prolonged metabolic condition categorized by elevated levels of glucose in blood. Dietary controlling is a process which must be carried out for a long period, but still has its lacuna in selecting the right food with needed calorie,

Received: 5th September, 2026 **Accepted: 7th September, 2026** **Published: 28th September, 2026**

and other essential minerals, vitamins etc. Recommender systems offer a novel plan which categorizes the broader array of available food items in to compatible, easy available, health oriented requirements. Choice of food or recommendations or availability differs. A food may be highly preferred but unsuitable for a diabetes-oriented dietary profile. For this reason, the present study uses a two-stage design: a transparent nutrient-based filter first restricts the candidate catalogue, and collaborative filtering then ranks the remaining foods according to learned preference relationships.

Collaborative filtering estimates preference for an unseen item from patterns observed in user–item interactions. In item–item CF, the model first estimates relationships among items from their rating patterns and then predicts a user's preference from the ratings of similar items. This approach is attractive for food recommendation because food-item relationships can remain relatively stable even when individual user tastes vary.

Two practical difficulties arise in diabetes-oriented CF. First, public datasets that contain both detailed nutrient composition and user-level food ratings are scarce. Second, conventional rating data are not necessarily restricted to nutritionally appropriate foods. The present study therefore combines a nutrient-annotated food dataset with a controlled synthetic-rating procedure so that the item–item CF pipeline can be evaluated in a reproducible setting.

The study addresses a gap in condition-specific food recommendation by explicitly applying a diabetes-oriented nutrient filter before collaborative ranking and by auditing the nutrient composition of the resulting. Top N list rather than reporting recommendation accuracy alone.

OBJECTIVES AND CONTRIBUTIONS

- ✓ Construct a diabetes-oriented food catalogue using transparent preprocessing and a sugar-content rule.
- ✓ Forecasting the top 10-recommendations by using item-item CF
- ✓ Assess the recommendation proposed by the devised model.

RELATED WORK

Food recommender systems have gradually moved beyond simply matching users with foods they already like. Recent systems increasingly combine user preferences with nutritional requirements, health conditions, and contextual information. The earlier work suggested that item the collaborative filtering is a wonderful method when user-item interactions happens. (Linden, 2003; Sarwar et al., 2001). Herlocker et al. (2004) But at the same time the evaluation of this method must be carried out carefully because sometime the actual predictions and reality differs in some cases. Koren et al. (2009) in later days the method that was used to find the preferences of items that are mapped in matrices was explored in an excellent way. The domain of food, Trattner and Elsweiler (2017) studied diverse types of food items recommender systems, which consists of various methods like content based, collaborative etc. Tran et al. (2018) studied that the recommendations can be done in a way that it must resemble the preferences of user and it must be balanced by means of nutrients. but this approach also has its hidden challenges like inadequate data that was given by the user, their frequent changes or transformation as for as food intake methods are concerned etc are the factors of research and this must be studied carefully in order to have an optimal solution.

The studies show that as for as the considerations of nutrition are concerned, the recommender system helps the user to make a choice based on the food choice by means of various filtering methods. But the comparison of these methods must be done with care because all these methods follows different data repository. (Amini et al., 2019; Bondevik et al., 2024). Chen et al. (2020) highlights the food recommendations on the basis of healthier recipes recommendations on the basis of NutRec. Li et al. (2023) explored the idea by utilizing the knowledge graphs.

The studies reveal that research area of food recommendation is changing by means of user preferences and as well as providing healthy meal plan. Silva et al. (2022) provided important evidence
through a collaborative-filtering study by exploring the extensive data set from ELSA-Brasil cohort which is designed only for dietary. The main motive of this study is to analyze 12,667 participants, and then compare the User and Item based CF and secured the good precision value 88%–91%. The prediction error is also seeming to me in a minimal number. User-based collaborative filtering also showed lower prediction error than item-based filtering. These results made a pavement for the researchers that CF can be used with real-world dietary data, only thing that how much data is directly deals with the interaction of User-Food.

Yera et al. (2023), study gives useful insights about the available data set and the lacuna, and where the research has to be done and why. The CF can be used to develop food recommendation for the diabetic patients but still the research has to

move further. Because so far that has been explored in this domain is limited due to non-availability of domain specific data which explains the direct link between User- Food preferences based on Diet. A diabetes management system has been designed by Sowah et al. (2020) which is new one and works by combining the identification of food with its diet value and adopted K-nearest neighbour. The model gave an excellent meal plan which is due to the wider range of its functionality.

Ahmed et al. (2025) studied that nowadays the research has been exploiting all latest computer science based methods and mechanism in the field of research and this approach is started to give results. One among them was to utilize the core features of Machine Learning algorithms to study the diabetes data set. By this novel idea it is identified that the data set must be widened so that the optimal results can be had. Prabakaran et al. (2025) analysed various food recommendation methods and suggested that the method or model must be limited and it must be based on nutrition and User preferences without compromising health. Incorporating the suggestions and intelligence of AI based systems also plays pivotal role in devising Food recommendation for diabetes patients. But this method has its own potential barrier. For example, combining User preferences which is Unique, a diversified data, studying the basic nature of data and practical use (Tsolakidis et al., 2024) are much more difficult. A chat based nutrition recommendation system ChatDiet show a notable contribution in this domain and explores how interactive models yields the results (Yang et al., 2024).

This study addresses the research lacuna by using a two-stage architecture. In the first stage a food catalogue has been developed which is based on nutrition requirements, filtering the sugar. In the second stage item–item CF explores these data set and ranks them on the basis of user preferences. This approach is reliable with the wider literature reviews which is a base of this work, and this which increasingly demands for recommendation systems that are aims to give Medicare on the basis of personalized, nutritionally apt, and appropriate for real-life use.

MATERIALS AND METHODS

Research Framework

The experimental workflow (Figure 1) includes five levels like data collection, cleaning them, then applying filtering due to nature of research, then user rating and generation, and the final one is item–item collaborative filtering.

The figure-1 explains the total available records before preprocessing. This Frame work explains the available data set after preprocessing, filtering level, cosine functionality, training and testing ration etc.

Figure 1. Research and Experimental Framework

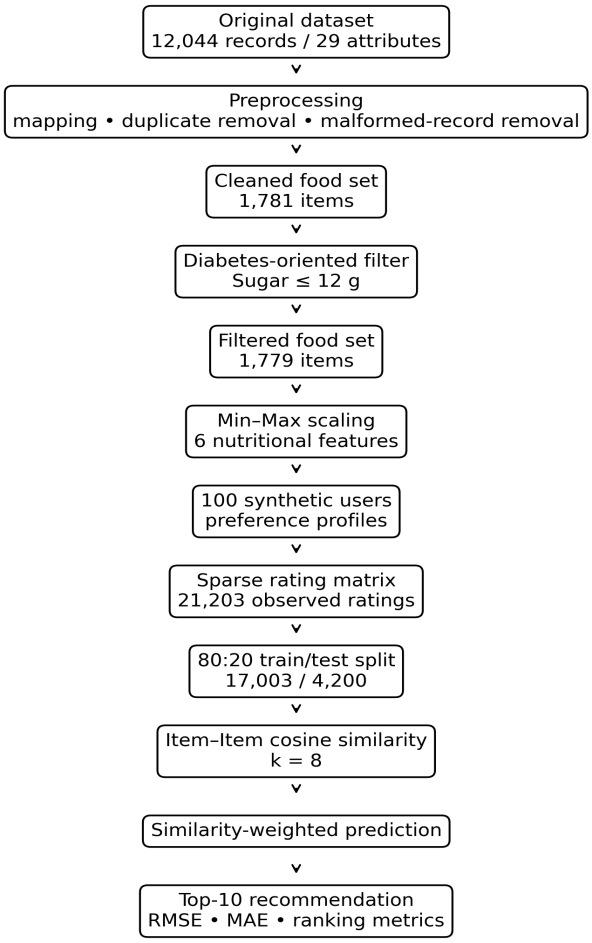


Figure 1. End-to-end research framework from the original food dataset to Top-10 recommendation and evaluation.

Dataset and Preprocessing

The source dataset selected for this study includes12,044 food records with 29 attributes like energy, protein, fat, carbohydrate, sugar-related fields, dietary fibre, and other variables. The data which is no longer needed for the study due to its nature and duplicates were removed. After this removal the source dataset is reduced to 1,781 food items.

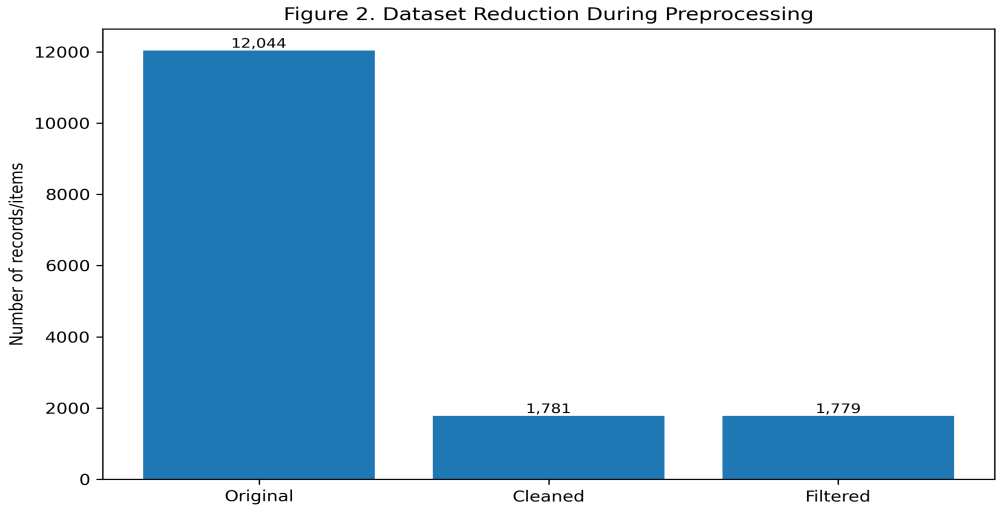


Figure 2. Reduction of the food catalogue during preprocessing and diabetes-oriented filtering.

Diabetes-Oriented Filtering and Nutritional Features

The investigation starts from applying rule-based diabetes profile was implemented so that the food items failed to pass the criteria will be removed **(Figure-2).** The principal rule removed food items with sugar content above 12 g per reported unit and this rule removed two additional items, producing a final candidate catalogue of 1,779 foods.

Figure-2 gives an insight that the original records, cleaned after preprocessing and then filtered records etc**.**

As for as the core data attributes were concerned, six nutrient features like protein, fibre, sugar, fat, calories, and carbohydrates were taken for modelling and min–max normalized to the [0,1] interval. These structures were used to build synthetic user preferences and provide a nutrient depiction that can support future content-based or hybrid extensions.

Synthetic Users and Ratings

Due to non –availability of authenticated patient-level food ratings, 100 synthetic users were generated. Each user acknowledged nutrient preference weights on behalf of variation in dietary significances. Protein received a positive weight in [0.2, 1.0], fibre in [0, 1.0], sugar a mainly negative weight in [−1.0, 0.2], and fat a user-specific weight in [−0.5, 0.5]. A user bias in [2.0, 4.0] recognized the baseline tendency on the 1–5 rating scale. Calories and carbohydrates continued in the feature matrix but established zero weight in this rating-generation trial.

Table-1 explains The Original food records, cleaned unique foods, Filtered candidate foods, Synthetic users, Observed ratings, Actual sparsity and other relevant data set descriptions are present in **(Table-1).**

Table 1. Experimental configuration and dataset summary.

Experimental parameter	Value
Original food records	12,044
Cleaned unique foods	1,781
Filtered candidate foods	1,779
Synthetic users	100
Observed ratings	21,203
Actual sparsity	88.08%
Rating scale	1–5
Gaussian noise (σ)	0.25
Train/test split	80:20 per user
Item-similarity neighborhood (k)	8
Top-N size	10
Relevance threshold	4.0
Random seed	42

Train/Test Split

As the customary practice in model development, the total data set is divided into training and testing data. So 20% of the user's rated items were used as test data and the remaining ratings were used for training.

PROPOSED ITEM–ITEM COLLABORATIVE FILTERING MODEL

Item Representation and Similarity

To develop or conceive a model, each food item was symbolized by a vector of user ratings from the training matrix. Missing values within an item's rating vector were replaced with the item's training-set mean; any remaining missing entries were set to zero.

In order to strengthen the model, Cosine similarity was then calculated between pairs of item vectors:

$$\hat{r}_{u,i} = \frac{\sum_{j \in N(i)} w_{(ij)} r_{u,j}}{\sum_{j \in N(i)} w_{ij}}$$

Figure-3 explains the resulting 1,779 × 1,779 matrix was symmetric. Across off-diagonal pairs, similarity ranged from 0.0000 to 1.0000, with mean 0.9937, median 0.9955, and standard deviation 0.0335. The strong concentration near one is important for interpretation because it indicates limited discrimination between many items under the chosen sparse-data imputation scheme.

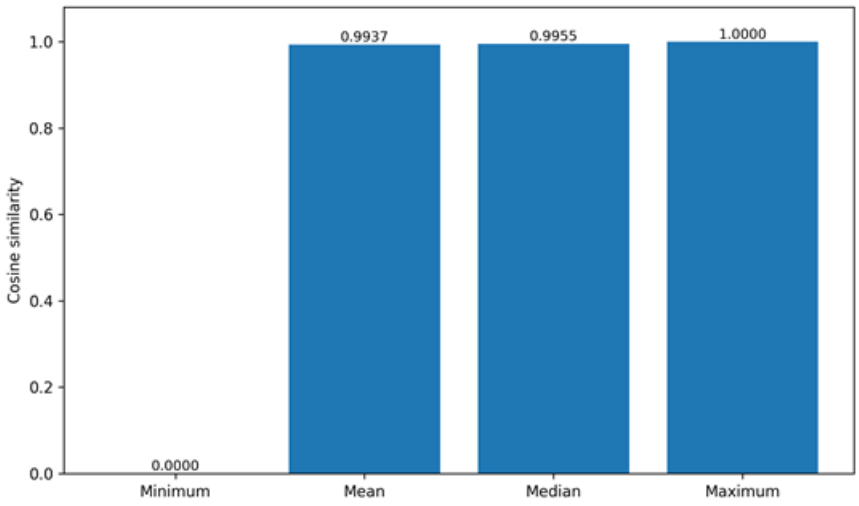


Figure 3. Summary of the observed item–item cosine similarity distribution.

Neighborhood and Rating Prediction

For each target item, the eight most similar items were selected as its neighbourhood. The predicted rating for user u and candidate item i was calculated as a similarity-weighted average of the user's known ratings on neighboring items.

If the user had no observed rating in the target item's neighbourhood, no prediction was returned. Predictions were clipped to [1,5]. For Top-10 recommendation, the model predicted ratings for unseen catalogue items, ranked them in descending order, and returned the ten highest-scoring items.

EXPERIMENTAL EVALUATION

All experiments were implemented in Python using pandas, NumPy, scikit-learn, and Matplotlib. A fixed random seed of 42 was used for synthetic-user generation, rating noise, sparsity masking, and train/test splitting.

Two groups of metrics were used. RMSE measures the square root of mean squared prediction error and therefore gives greater influence to larger errors. MAE measures the average absolute prediction error. For ranking, Precision@10 is the fraction of the ten recommendations that are relevant, Recall@10 is the fraction of relevant held-out items retrieved in the Top-10 list, and HitRate@10 measures the proportion of evaluated users for whom at least one relevant item is recommended. An item was treated as relevant when it's held-out rating was at least 4.0.

Although 4,200 ratings were assigned to the test partition, the model could generate predictions for only 1,640 users–item pairs because some test items lacked usable neighbourhood evidence.

RMSE and MAE were therefore calculated on those 1,640 predictable pairs, while the ranking evaluation covered all 100 users with the specified relevance criterion.

RESULTS

Rating Matrix and Similarity Structure

The final rating matrix contained 177,900 possible users–item interactions, of which 21,203 were observed and 156,697 were missing. The mean rating was 2.7310 with a standard deviation of 0.9068. The training and testing means were 2.7299 and 2.7353 respectively, showing that the random per-user split produced closely matched rating distributions and tabulated in Table-2.

Table-2 includes, Possible interactions, Observed ratings etc., which gives an useful insights about the work done.

Table 2. Composition of the synthetic user–food rating matrix.

Statistic	Value
Users	100
Food items	1,779
Possible interactions	177,900
Observed ratings	21,203
Missing ratings	156,697
Sparsity	88.08%
Training ratings	17,003
Testing ratings	4,200

Rating Prediction Performance

Figure -4 and Table-3**,** clearly explains **t**he larger RMSE compared with MAE suggests that a subset of predictions contained comparatively larger errors and it was depicted in the figure 4 and tabulated in Table-3. This shows the accuracy level has been reached, For the 1,640 held-out pairs for which a prediction could be generated, the item–item CF model achieved an RMSE of 0.946318 and an MAE of 0.710097. The MAE indicates an average absolute prediction error of about 0.71 rating points on the 1–5 scale.

.

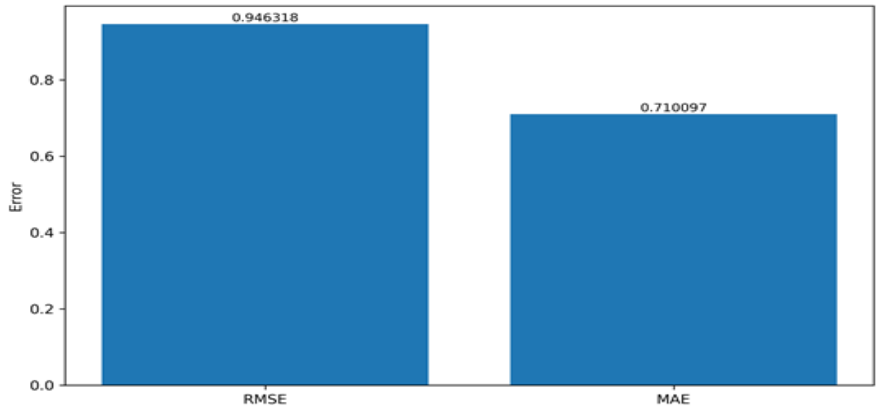


Figure 4. Rating-prediction error for item–item CF.

Table 3. Rating-prediction performance.

Metric	Result
RMSE	0.946318
MAE	0.710097
Predictable held-out pairs	1,640

Top-10 Ranking Performance

Across the 100 evaluated users, Precision@10 was 0.009091, Recall@10 was 0.007490, and HitRate@10 was 0.090909. Thus, the model rarely placed more than one relevant held-out item in a user's Top-10 list, (was depicted in Figure -5.

All the above figurative details were tabulated in Table -4, and approximately one in eleven users received at least one relevant item. Figure-5 highlights the precision value, recall. the values have been derived with reference to 100 evaluated users.

Table 4, gives an idea about the actual results with its overall performance metrics.

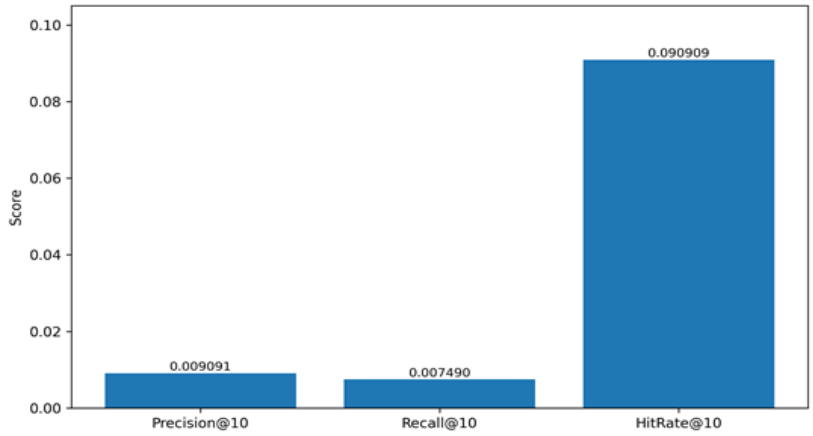


Figure 5. Top-10 ranking performance using Precision@10, Recall@10, and HitRate@10.

Table 4. Top-10 ranking performance.

Metric	Value
Precision@10	0.009091
Recall@10	0.007490
HitRate@10	0.090909
Users evaluated	100

Representative Recommendations and Nutritional Profile

For the representative user user_0, the Top-10 list was dominated by dry legumes and lean proteins. (Table 5) and Mean nutritional profile of the representative Top-10 recommendations was explained in Figure-6.

Table-5 gives the food names, its predicted ratings, its protein, fat, fiber values. This shows the effectiveness of the method followed**.**

Figure-6 explains that the protein plays a significant role then fat and lastly the fiber.

Table 5. Representative Top-10 recommendations for user_0. Food names are retained from the source composition records.

Rank	Food	Predicted rating	Protein (g)	Fat (g)	Fibre (g)
1	Beans, dry, pink, 732	3.5338	23.1	1.39	4.0
2	Beans, dry, pinto, 11f-8126	3.5338	24.2	1.05	4.6
3	Beans, dry, pinto, 497	3.5338	25.4	1.29	4.0
4	Beans, dry, pinto, 718	3.5338	23.9	1.22	4.0
5	Beef, eye of round, lean, raw	3.5338	23.8	2.07	0.0
6	Beans, dry, black, 227	3.5338	22.0	1.60	4.1
7	Beans, dry, great northern, 514	3.5338	23.8	1.11	4.4
8	Salsa, pace chunky, medium	3.5338	0.0	0.20	0.0
9	Beans, dry, black, 11f-8112	3.4181	21.9	1.33	4.4
10	Cheese, Swiss, slices	3.4181	0.0	29.90	0.0

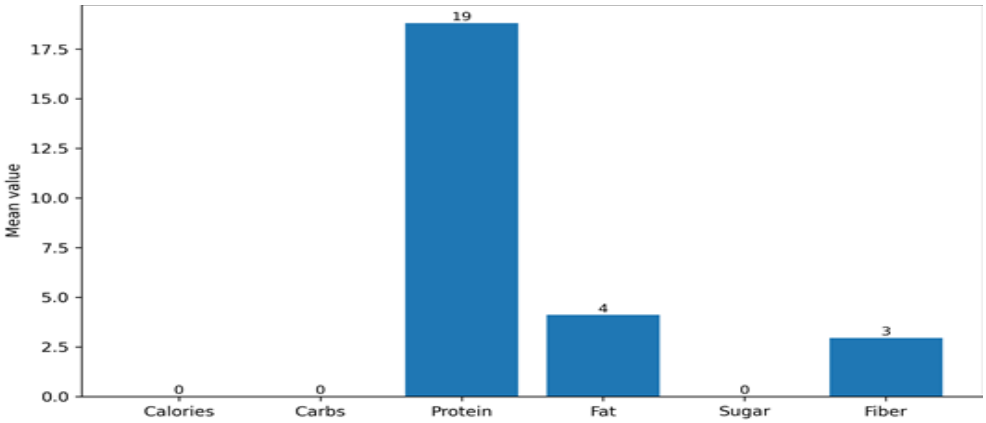


Figure 6. Mean nutritional profile of the representative Top-10 recommendations.

DISCUSSION

The experiment proves that an item–item CF can be placed after a nutrient filter to produce commendations that are dependable with the diabetes-oriented profile. This architecture has a practical advantage that any item which violates the rule particularly sugar rule cannot enter into the model. At the same time, the ranking results are designate that the model has limited ability to identify the user's most applicable foods from a catalogue.

The high similarity is plausibly related to the use of mean imputation before cosine calculation. With an 88.08% sparse matrix, replacing missing entries with item means can make rating vectors look artificially alike. This reduces the contrast needed by a neighborhood method to distinguish genuinely informative neighbors from broadly similar items.

The synthetic nature of the ratings creates a second interpretive issue. The same nutrient variables used to construct user preferences also influence the recommendation environment. Consequently, the observation that recommendations are low in recorded sugar and relatively rich in protein and fibre is partly expected from the experimental design rather than an independently discovered property of real users.

The catalogue size is also large relative to the number of users. With only 100 users and approximately 212 observed ratings per user on average, the model has limited co-rating evidence for distinguishing 1,779 candidate foods. The result is therefore better understood as a controlled baseline for the proposed architecture than as evidence of clinical or real-world recommendation effectiveness.

Limitations and Future Work

Four limitations are particularly important. First, the user ratings must be improved so that that can be incorporated in the viable model development. some source nutrient fields are unfinished, which affects the interpretation of zero-valued calorie and carbohydrate entries. Fourth, the study evaluates only item–item CF and does not provide an end-to-end experimental baseline.

CONCLUSION

The model food recommendation for diabetes patients constructively yielded a good result. An item–item collaborative filtering for diabetes-oriented food recommendation was framed with 12,044-record nutrient composition dataset was reduced to 1,779 through preprocessing. A rule based filtering like sugar-content rule, and 100 synthetic users generated 21,203 observed ratings at 88.08% sparsity. The model accuracy and its precision values are tabulated and verified, shows good improvement in the CF as for as present data set is concerned.

Conflict of Interest: None

Source of support: None

ACKNOWLEDGEMENT

We acknowledge the support of the PG & Research Department of Computer Science, Jamal Mohamed College (Autonomous), Trichy for providing computational resources. This research was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Abhari, S., Safdari, R., Azadbakht, L., Lankarani, K. B., Niakan Kalhori, S. R., Honarvar, B., Abhari, K., Ayyoubzadeh, S. M., Karbasi, Z., Zakerabasali, S., & Jalilpiran, Y. (2019). A systematic review of nutrition recommendation systems:

With focus on technical aspects. Journal of Biomedical Physics & Engineering, 9(6), 591–602. https://doi.org/10.31661/jbpe.v0i0.1248

Ahmed, E., Oumer, M., & Hassan, M. (2025). Diabetes-focused food recommender system (DFRS) to enabling digital health. PLOS Digital Health, 4(2), e0000530. https://doi.org/10.1371/journal.pdig.0000530

Bondevik, J. N., Bennin, K. E., Babur, Ö., & Ersch, C. (2024). A systematic review on food recommender systems. Expert Systems with Applications, 238, 122166. https://doi.org/10.1016/j.eswa.2023.122166

Chen, M., Jia, X., Gorbonos, E., Hoang, C. T., Yu, X., & Liu, Y. (2020). Eating healthier: Exploring nutrition information for healthier recipe recommendation. Information Processing & Management, 57(6), 102051. https://doi.org/10.1016/j.ipm.2019.05.012

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems, 22(1), 5–53. https://doi.org/10.1145/963770.963772

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30–37. https://doi.org/10.1109/MC.2009.263

Li, D., Zaki, M. J., & Chen, C.-H. (2023). Health-guided recipe recommendation over knowledge graphs. Journal of Web Semantics, 75, 100743. https://doi.org/10.1016/j.websem.2022.100743

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 76–80. https://doi.org/10.1109/MIC.2003.1167344

Prabakaran, N., Rajan, N., Chagalasetty, S., & Kakani, V. (2025). A comprehensive review of food recommendation systems in the context of nutritional therapy for diabetes mellitus. Current Nutrition & Food Science, 21(1), 14–34. https://doi.org/10.2174/0115734013297002240522063241

Sarwar, B., Karypis, G., Konstan, J. A., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web, WWW 2001 (pp. 285–295). Association for Computing Machinery. https://doi.org/10.1145/371920.372071

Silva, V. C., Gorgulho, B., Marchioni, D. M., Alvim, S. M., Giatti, L., de Araujo, T. A., Alonso, A. C., Santos, I. d. S., Lotufo, P. A., & Benseñor, I. M. (2022). Recommender system based on collaborative filtering for personalized dietary advice: A cross-sectional analysis of the ELSA-Brasil Study. International Journal of Environmental Research and Public Health, 19(22), 14934. https://doi.org/10.3390/ijerph192214934

Sowah, R. A., Bampoe-Addo, A. A., Armoo, S. K., Saalia, F. K., Gatsi, F., & Sarkodie-Mensah, B. (2020). Design and development of diabetes management system using machine learning. International Journal of Telemedicine and Applications, 2020, Article 8870141. https://doi.org/10.1155/2020/8870141

Tsolakidis, D., Gymnopoulos, L. P., & Dimitropoulos, K. (2024). Artificial intelligence and machine learning technologies for personalized nutrition: A review. Informatics, 11(3), 62. https://doi.org/10.3390/informatics11030062

Tran, T. N. T., Atas, M., Felfernig, A., & Stettinger, M. (2018). An overview of recommender systems in the healthy food domain. Journal of Intelligent Information Systems, 50, 501–526. https://doi.org/10.1007/s10844-017-0469-0

Trattner, C., & Elsweiler, D. (2017). Food recommender systems: Important contributions, challenges and future research directions. arXiv. https://arxiv.org/abs/1711.02760

Yang, Z., Khatibi, E., Nagesh, N., Abbasian, M., Azimi, I., Jain, R., & Rahmani, A. M. (2024). ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. Smart Health, 32, 100465. https://doi.org/10.1016/j.smhl.2024.100465

Yera, R., Alzahrani, A. A., Martínez, L., & Rodríguez, R. M. (2023). A systematic review on food recommender systems for diabetic patients. International Journal of Environmental Research and Public Health, 20(5), 4248. https://doi.org/10.3390/ijerph20054248