JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8223

RANDOM FOREST AND K-MEANS FOOD RECOMMENDATIONS FOR TYPE-2 DIABETES MELLITUS BASED ON GLYCEMIC INDEX

Malik Aziz Ramadhan[1], Yohani Setiya Rafika Nur[1*], Dany Candra Febrianto[1]

Telkom University, Banyumas, Indonesia[1]
https://telkomuniversity.ac.id[1]
malikazizramadhan@student.telkomuniversity.ac.id, yohanin@telkomuniversity.ac.id*,
danycandra@telkomuniversity.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)



The creation is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License

Submitted: 06 February 2026, **Revised:** 27 July 2026, **Accepted:** 29 July 2026, **Published:** 12 November 2026.

Abstract— *Patients with Type 2 Diabetes Mellitus require appropriate dietary management to maintain stable blood glucose levels. However, selecting suitable foods remains challenging because information regarding glycemic index values and nutritional content is not always readily available or easily understood by patients. This study develops a glycemic index-based food recommendation system by combining the Random Forest and K-Means algorithms. In the proposed system, Random Forest is employed to predict the risk level of each food item using nutritional attributes, while K-Means groups foods with similar characteristics to provide safer alternative food recommendations. The dataset was constructed by integrating the International Tables of Glycemic Index and Glycemic Load Values 2021 with the Indonesian Food Composition Table (TKPI) and subsequently preprocessed before model development. The evaluation results indicate that the combination of Random Forest and SMOTE achieved the best classification performance, with an Accuracy of 93.10%, Precision of 95.86%, and F1-Score of 93.76%. For the clustering task, K-Means produced three clusters with a Silhouette Score of 0.43836. Finally, both models were integrated into a Streamlit-based prototype to support food risk classification and provide alternative food recommendations for patients with diabetes.*

Keywords: *Glycemic index, K-means, random forest, recommendation system, type-2 diabetes mellitus.*

INTRODUCTION

Health is an important aspect that affects the quality of human life. However, changes in modern lifestyles have led to an increase in chronic diseases, one of which is diabetes mellitus. According to the World Health Organization (WHO), diabetes is a major cause of death and a significant global economic burden [1]. Diabetes mellitus occurs when the body cannot produce insulin or use it effectively, causing blood sugar levels to rise [2]. Indonesia has the fifth-highest number of diabetes patients in the world and is predicted to reach 28.6 million people by 2050 [3]. Therefore, diet management is very important, especially for Type 2 diabetes patients, taking into account the Glycemic Index (GI), which measures the speed at which food increases blood sugar levels [4]. Consuming foods with a high GI poses a significant risk to diabetic patients [5].

Several studies have attempted to address dietary issues through technology by developing food classification and recommendation systems based on nutritional content and user preferences, showing that machine learning can support dietary decision making [6][7][8]. Recent studies have also explored personalized food recommendation systems specifically for diabetes management by integrating machine learning and nutritional information to generate individualized dietary recommendations [9]. However, most existing approaches remain general-purpose and do not specifically target diabetic patients or incorporate Glycemic Index parameters, which are essential for blood sugar control [6]. This research gap highlights the need for a system that specifically considers



glycemic impact in food recommendations. To overcome these limitations, this study uses a combined approach involving two machine learning methods, namely the Random Forest algorithm and the K-Means algorithm. The Random Forest algorithm was chosen to classify food risk levels because of its ability to handle high-dimensional data and provide reliable classification performance [10]. Meanwhile, K-Means was used as an unsupervised learning technique to group foods based on nutritional similarity in order to provide relevant food alternatives [8].

This study aims to design and implement a food recommendation system specifically for patients with Type 2 Diabetes Mellitus. This system not only classifies foods into safe, restricted, or risky categories, but also suggests alternative foods with similar nutritional characteristics. In this study, food categories were first assigned according to glycemic index thresholds defined by the ISO 26642:2010 standard, while the Random Forest classifier used only nutritional attributes (calories, protein, fat, carbohydrates, and fiber) as predictor variables. Glycemic index values were used solely to generate category labels and support the recommendation process.

The main issues discussed include the implementation of the Random Forest method using nutritional content for food classification and the K-Means method for similarity-based food recommendation, as well as the evaluation of system performance. The Random Forest model is evaluated using a confusion matrix and its derivative metrics. At the same time, the quality of K-Means clusters is measured using the Elbow Method, Silhouette Score, and Davies-Bouldin Index. The scope of this study covers a dataset consisting of 162 food items sourced from the International Tables of Glycemic Index 2021 and the Indonesian Food Composition Table (TKPI) 2020. This system was developed as a web-based prototype using the Streamlit framework for model testing and educational purposes, without replacing professional medical diagnosis. Through this research, it is hoped that a tool can be created to facilitate diabetes patients in managing their diet to prevent chronic complications.

MATERIALS AND METHODS

This study uses a quantitative experimental approach to develop a food recommendation system using secondary data. The main data sources were compiled from two key references, namely the Glycemic Index (GI) values sourced from the International Tables of Glycemic Index and Glycemic Load Values 2021 by Atkinson et al., and nutritional content data obtained from the 2020 Indonesian Food Composition Table (TKPI) published by the Ministry of Health of the Republic of Indonesia. Because the data came from different sources, a manual integration and validation process was carried out using a food-matching technique to align the food types in the TKPI with their GI references. From a total of 1,146 food types in the TKPI, 162 ready-to-eat foods were selected based on the relevance of their GI data. The final dataset consisted of 162 rows with attributes including food name, GI value, and nutritional content (calories, protein, fat, carbohydrates, and fiber), as presented in Table 1. In the classification stage, only nutritional attributes were used as input features for the Random Forest model, while GI values were used to assign category labels according to the ISO 26642:2010 standard and to support the recommendation process.

Table 1. Sample of Nutritional Content and Glycemic Index Dataset

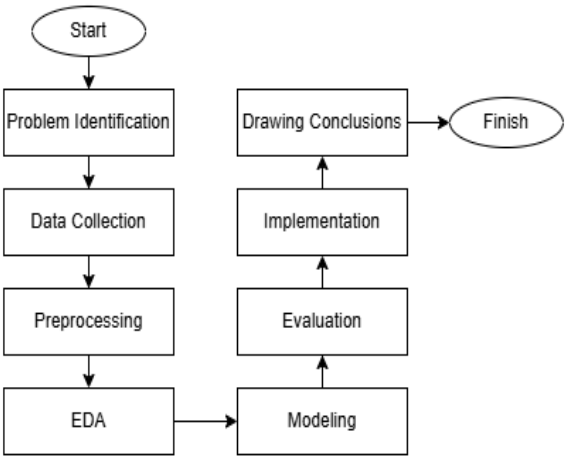
Food Name	Cal	Protein	Fat	Carbs	Fiber	GI	Category
Nasi	180	3	0.3	39.8	0.2	73	2
Ketan hitam	181	4	1.2	37.3	0.3	100	2
Bakwan	280	8.2	10.2	39	3.4	69	1
Mie ayam	102	6.2	3.9	10.5	1	58	1
Gado-gado	137	6.1	3.2	21	1.1	54	0
Apel, segar	58	0.3	0.4	14.9	2.6	44	0

Source: (Research Results, 2026)

The research process for this study, from data collection to evaluation, is shown in Figure 1. Prior to entering the modeling phase, data preprocessing was conducted to ensure data quality. This data preprocessing stage involved data cleaning to handle missing and duplicate values, and outlier detection using the interquartile range (IQR) method. The IQR is the difference between the third quartile (Q3) and the first quartile (Q1) [11]. Subsequently, GI values were used to assign food categories according to the ISO 26642:2010 standard. The resulting category labels served as target variables, while only nutritional attributes (calories, protein, fat, carbohydrates, and fiber) were used as predictor variables in the Random Forest classification model. This standard categorizes foods into three risk levels based on their glycemic index (GI). Foods with a GI ≤ 55 are classified as "safe," those with a GI between 56 and 69 as "restricted," and those with a GI ≥ 70 as "risky" [12].

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8223



Source: (Research Results, 2026)
Figure 1. Research Methodology Flowchart

The dataset was then split into 80% training data and 20% test data using stratified sampling to maintain class proportions. The training process was additionally evaluated using 10-fold stratified cross-validation to assess the stability and consistency of the Random Forest model across different data partitions. Numerical attributes of the dataset are normalized using StandardScaler. Class imbalance in the training data is corrected by a data augmentation technique using synthetic oversampling of minority classes (SMOTE). SMOTE works by synthesizing new data among existing minority samples. The method is to calculate the difference between the feature vector and its closest neighbor, then multiply the difference by a random number between 0 and 1. The result of this multiplication is then added back to the initial feature vector, as formulated in Equation (1), where X_{syn} is the new synthetic data point, X_i is the original minority sample, and X_{knn} is one of the k nearest neighbors [13].

$$X_{syn} = X_i + (X_{knn} - X_i) \times \sigma \quad (1)$$

The system architecture integrates two machine learning algorithms to generate recommendations. The Random Forest algorithm is an ensemble classification algorithm that generates multiple decision trees from random subsets of data. The final decision is made by majority voting across all trees [10]. This algorithm is used to predict food risk levels based on nutritional characteristics. To optimize performance and prevent overfitting, the Random Forest model was set up using the following configuration: n_estimators = 200, max_depth = 10, class_weight = 'balanced'. The K-means algorithm was then applied to group foods according to the similarity of their nutritional characteristics. This algorithm works iteratively by initializing the centroid, assigning each data point to the nearest centroid using Euclidean distance, and updating the centroid until convergence is achieved [14]. In its recommendation logic, if a food is classified as "Risky," the system will search for other foods in the same cluster labeled "Safe" by calculating their similarity distance.

To assess the technical performance of the system, the evaluation was conducted in two parts. The Random Forest classification model was measured using the Confusion Matrix components, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Based on these components, Accuracy (Equation 2), Precision (Equation 3), Recall (Equation 4), and F1-Score (Equation 5) were calculated [11].

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

$$\text{F1} - \text{score} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision+Recall}} \quad (5)$$

Meanwhile, the quality of K-Means clustering is evaluated to determine the optimal number of clusters (K) using the Elbow Method to observe the decrease in Sum of Squared Errors (SSE). To measure how well cohesion and separation between groups are achieved, the Silhouette Score calculation is used according to Equation (6). In this formula, *a(i)* represents the average distance within the cluster itself, while *b(i)* is the average distance to the nearest neighboring cluster. In addition, the Davies-Bouldin Index (DBI) calculation is also performed with reference to Equation (7), where k is the total number of clusters and R_{ij} is the similarity level between cluster *i* and cluster *j*. A lower DBI value indicates better clustering performance [15][16].

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j}(R_{i,j}) \quad (7)$$

CC BY NC

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v12i2.8223

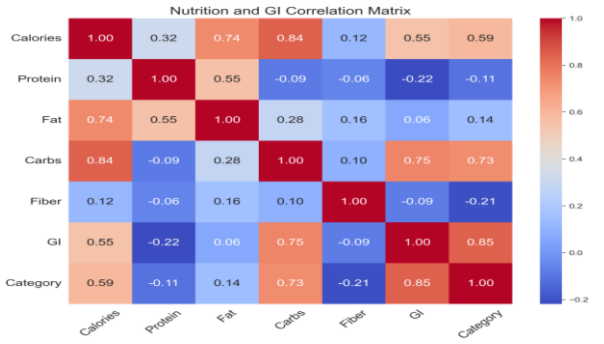
JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

RESULTS AND DISCUSSION

The initial stage of the research began with preprocessing 162 food data items to ensure the quality of input for the model. The data validation process did not detect any missing or duplicate values. However, the application of the outlier removal method with an IQR threshold of 3.0 identified 20 data points with extreme nutritional values, including sweet potato chips with relatively high fiber content and cashews with high fat content. Although these foods may represent valid nutritional variations, they were excluded in this study to reduce the influence of extreme values on model training and to improve model stability, resulting in a final dataset of 142 rows. The initial class distribution showed a significant imbalance, dominated by the "Safe" class with 86 items, followed by "At Risk" with 39 items, and "Restricted" with only 17 items. To overcome the bias towards the majority class, the SMOTE technique was applied to the training data, which successfully balanced the number of samples to 68 items per class (a total of 204 training data), while the test data was left according to its original distribution so that the validation remained objective.

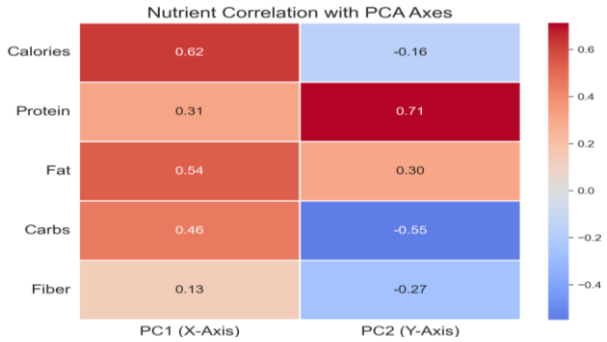
The relatively small dataset size was mainly caused by the limited availability of food items that could be consistently matched between the Indonesian Food Composition Table and published glycemic index references. Consequently, the 80:20 train-test split resulted in a relatively small test set, and the reported performance should therefore be interpreted within the scope of the available dataset rather than as evidence of broad generalizability. In addition, the current evaluation was limited to classification and clustering metrics and did not include external validation through user studies, nutrition expert assessments, or comparisons with existing dietary recommendation systems. Nevertheless, preprocessing, stratified data splitting, class balancing, and 10-fold stratified cross-validation were employed to reduce potential bias and support a more robust evaluation of the proposed model.

Feature correlation analysis, as visualized in the Correlation Heatmap Matrix (Figure 2), indicates that Carbohydrates have the highest correlation with the target variables of Category at 0.75 and Glycemic Index at 0.85. These results are consistent with the theory that carbohydrates are the main factor triggering the glycemic response. In addition, analysis using Principal Component Analysis (PCA) shows that the first principal component (PC1) represents the "Energy Density" dimension, which is dominated by Calories and Fat, while the second component (PC2) represents the "Macronutrient Profile" that distinguishes between protein and carbohydrate sources (Figure 3).



Source: (Research Results, 2026)

Figure 2. Feature Correlation Matrix using Heatmap



Source: (Research Results, 2026)

Figure 3. Principal Component Analysis (PCA) Visualization

Based on this understanding of the data, Random Forest model testing was conducted in two scenarios, namely without augmentation (Scenario 1) and with SMOTE augmentation (Scenario 2). The evaluation results showed improved performance in Scenario 2, where the accuracy increased from 86.21% to 93.10%, accompanied by higher precision (95.86%) and an F1-score of 93.76%. The most notable improvement was observed in the minority class ("Restricted"), where recall increased from 33.33% to 100%, as shown in Table 2. These results indicate that SMOTE helped alleviate class imbalance during model training. However, the observed performance improvement should be interpreted with caution because the evaluation was conducted on a relatively small dataset with a limited test set, and synthetic oversampling may not fully represent the variability of real-world food data. Therefore, additional validation using larger and more diverse datasets is required to further confirm the generalizability of the proposed model.

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8223

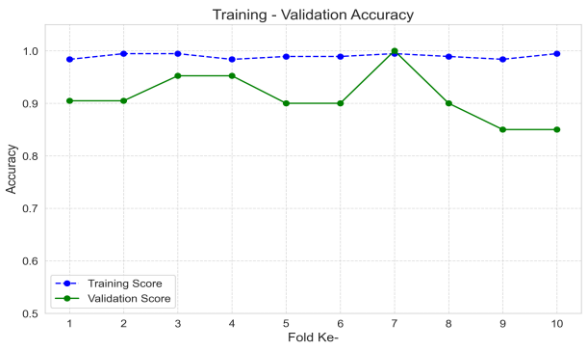
Table 2. Model Performance Comparison in Scenario 1 and Scenario 2

Accuracy Matrix	Scenario 1	Scenario 2
Accuracy	0.8621	0.9310
Precision	0.8566	0.9586
Recall	0.8621	0.9310
F1-Score	0.8541	0.9376

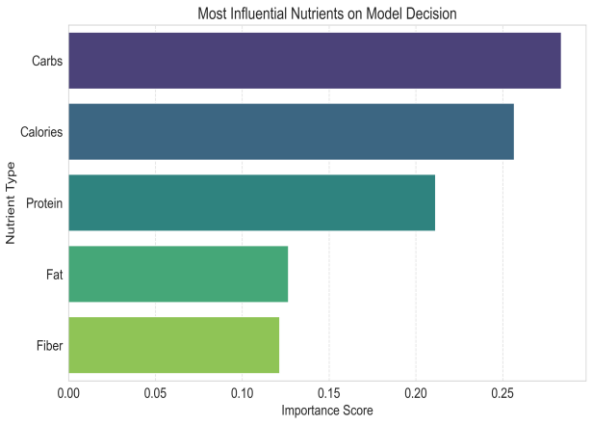
Source: (Research Results, 2026)

Model consistency validation using 10-Fold Stratified Cross-Validation resulted in an average accuracy of 91.14% with a standard deviation of 4.38%, as shown in Figure 4. Although there is a difference between the training accuracy (98.97%) and the validation accuracy, the stable performance above 90% indicates that the model has good generalization capabilities and does not experience severe overfitting. The relatively low standard deviation also suggests that the model performance remains consistent across different folds and is not heavily dependent on a particular data partition. However, these results should be interpreted within the scope of the filtered dataset, since foods with extreme nutritional characteristics may represent valid nutritional variations rather than erroneous observations.

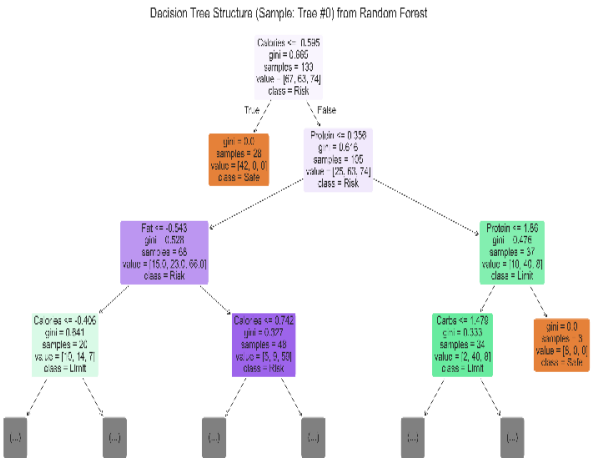
Although their exclusion helped improve model stability, it may reduce the practical coverage of the recommendation system for highly distinctive food profiles. Feature Importance Analysis (Figure 5) places Carbohydrates as the most dominant feature with a score of 0.2838, followed by Calories with a value of 0.2567 and Protein with a value of 0.2113. These findings are consistent with the nutritional characteristics of glycemic responses, where carbohydrate intake is one of the major determinants affecting postprandial blood glucose levels. As shown in the decision tree structure (Figure 6), the model places Calories as an initial filter to detect "Safe" foods, while more complex risk class separation is performed through a combination of other macronutrients.



Source: (Research Results, 2026)
Figure 4. 10-Fold Cross-Validation Performance Graph



Source: (Research Results, 2026)
Figure 5. Feature Importance in Classification Model



Source: (Research Results, 2026)
Figure 6. Visualization of Decision Tree Structure Sample (Tree-0)

On the other hand, to meet the need for alternative recommendation features, the optimal number of clusters was determined by testing a range of K values from 2 to 10. Based on the comparison of metrics in Table 3, K = 3 was selected because it produced the highest Silhouette Score of 0.43836, supported by a slowdown in the decline in inertia in the Elbow Method graph. Although this Silhouette Score indicates moderate cluster separation, it was considered sufficient for grouping foods with relatively similar nutritional characteristics to support computational recommendations. Therefore, the resulting clusters should be interpreted as nutritional similarity groups rather than clinically validated dietary substitution groups.

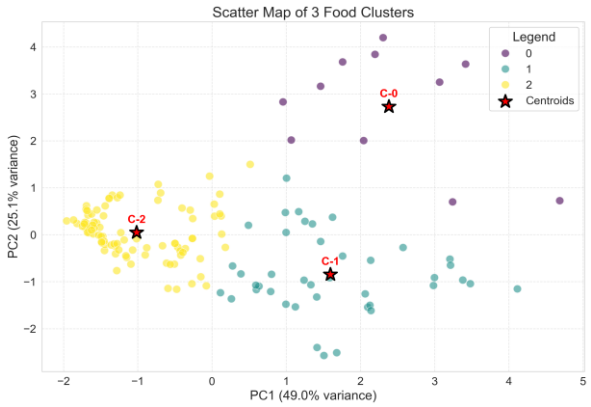
The PCA visualization (Figure 7) shows three separate food groups with distinctive nutritional characteristics, consisting of Cluster 0 (11 items) containing high-protein and high-fat foods

CC BY NC
Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v12i2.8223

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

representing animal products, Cluster 1 (41 items) containing high-calorie and carbohydrate foods representing staple foods, and Cluster 2 (90 items) containing low-calorie and low-fat foods representing fruits and vegetables. The largest proportion in Cluster 2, at 63.4%, indicates that the dataset contains a relatively wide variety of foods that can be considered as potential recommendation candidates.

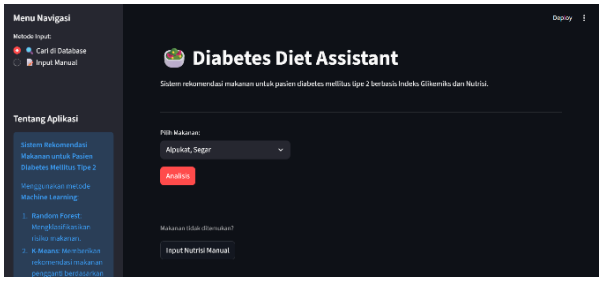


Source: (Research Results, 2026)

Figure 7. Cluster Distribution at $K = 3$

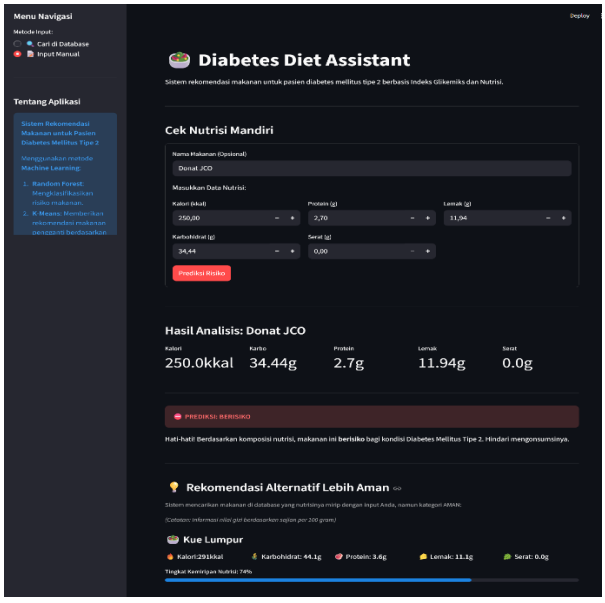
The integration of the two models was then tested through a web-based interface (Figure 8) that provides manual input and database search features. When users enter food items predicted as "At Risk" by Random Forest, the system automatically displays a warning and presents a list of alternative foods from the "Safe" category taken from the same K-Means cluster (Figure 9). This mechanism demonstrates that combining risk classification and similarity-based alternative searches can be used to generate nutritionally similar food recommendations. However, the identified alternatives should be interpreted as computational recommendations based on nutritional similarity and not as clinically equivalent dietary substitutions. Further validation involving nutrition experts and diabetic patients is necessary to assess the practical suitability of these recommendations in real-world dietary management. Future studies should also compare the proposed approach with existing dietary recommendation systems to better assess its practical effectiveness and clinical applicability. In addition, the current recommendation mechanism does not yet consider other important factors in diabetes dietary planning, such as glycemic load, portion size, patient profiles, medication use, comorbidities, allergies, and individualized nutritional requirements. Therefore, the proposed system should be considered as an assistive recommendation tool rather than a substitute for professional dietary assessment.

Although the results were promising, this study still has several limitations. The dataset is relatively small, the system does not yet consider individual user profiles, and the implementation is still limited to a web-based prototype. Based on these limitations, future studies can improve the system by predicting GI values from nutrition data or by adding more types of traditional local foods. Personalization can also be improved by including user information such as weight, age, and medical history. In addition, Deep Learning algorithms can be considered as a long-term approach for handling more complex datasets.



Source: (Research Results, 2026)

Figure 8. Recommendation System Main Interface



Source: (Research Results, 2026)

Figure 9. Prediction Result Display with Alternative Recommendations

CONCLUSION

This study successfully designed and implemented a food recommendation system for type 2 diabetes mellitus patients by combining the Random Forest method for food risk level

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.

VOL. 12. NO. 2 NOVEMBER 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i2.8223

classification and K-Means for providing safer food alternatives. The evaluation results show that the application of the SMOTE data augmentation technique significantly improves the performance of the classification model, with an accuracy of 93.10%, precision of 95.86%, and F1-Score of 93.76%. In addition, stability testing using 10-Fold Stratified Cross-Validation showed model consistency with an average accuracy of 91.14%.

The carbohydrate feature was identified as the most dominant determinant with a contribution of 28.38% to risk determination, which is in line with the glycemic index principle. Meanwhile, the K-Means algorithm was able to group 142 types of food with a Silhouette Score of 0.43836 into three clusters, namely high protein-fat, high carbohydrate-calorie, and low calorie. Visual validation of the data distribution for each cluster through PCA showed clear cluster separation representing 74.1% of data variance. These findings indicate that the combination of both methods can be implemented as a web-based decision support tool for food recommendation in patients with Type 2 Diabetes Mellitus. However, the proposed system does not yet incorporate important clinical factors such as glycemic load, portion size, patient profiles, medication use, comorbidities, allergies, and individualized nutritional requirements. Therefore, the system should be regarded as an assistive recommendation tool and requires further clinical validation before practical implementation.

DECLARATION OF AI USE

The authors declare that generative artificial intelligence (AI) tools were used in the preparation of this manuscript. An AI tool was used to assist with translating the text from Indonesian to English. The authors reviewed and revised all AI-assisted content and take full responsibility for the accuracy, originality, and final content of the manuscript.

AUTHOR CONTRIBUTIONS

This journal uses the Contributor Roles Taxonomy (CRediT) to identify individual author contributions to the research and manuscript preparation.

Author	Contribution Roles
Malik Aziz Ramadhan	Investigation, Formal Analysis, Writing – Original Draft.
Yohani Setiya Rafika Nur	Methodology, Supervision, Validation, Review & Editing.
Dany Candra Febrianto	Supervision, Validation, Review & Editing.

ACKNOWLEDGMENTS

The authors would like to thank Telkom University for providing the facilities and environment to support this research.

FUNDING

This research was funded by Telkom University.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest related to the research, writing, or publication of this article.

REFERENCE

[1] World Health Organization, "Diabetes," WHO. Accessed: Apr. 12, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes

[2] International Federation of Diabetes, "What Is Diabetes? International Federation of Diabetes," IDF. Accessed: Apr. 12, 2025. [Online]. Available: https://idf.org/about-diabetes/what-is-diabetes/

[3] International Federation of Diabetes, "IDF Diabetes Atlas Global Diabetes Data & Statistics," IDF. Accessed: Apr. 12, 2025. [Online]. Available: https://diabetesatlas.org/

[4] M. Peres, H. S. Costa, M. A. Silva, and T. G. Albuquerque, "The Health Effects of Low Glycemic Index and Low Glycemic Load Interventions on Prediabetes and Type 2 Diabetes Mellitus: A Literature Review of RCTs," *Nutrients*, vol. 15, no. 24, p. 5060, 2023, doi: 10.3390/nu15245060.

[5] N. A. Elsayed *et al.*, "5. Facilitating Positive Health Behaviors and Well-being to Improve Health Outcomes: Standards of Care in Diabetes—2023," *Diabetes Care*, vol. 47, no. 1, pp. S68–S96, 2023, doi: 10.2337/dc23-S005.

[6] J. N. Bondevik, K. E. Bennin, Ö. Babur, and C. Ersch, "A systematic review on food recommender systems," *Expert Syst. Appl.*, vol. 238, no. 20, p. 4248, 2023, doi: 10.1016/j.eswa.2023.122166.

[7] A. Sosa-Holwerda, O. H. Park, K. Albracht-

Schulte, S. Niraula, L. Thompson, and W. Oldewage-Theron, "The Role of Artificial Intelligence in Nutrition Research: A Scoping Review," *Nutr.* , vol. 16, no. 13, pp. 1–22, 2024, doi: 10.3390/nu16132066.

[8] M. Ahmad, A. U. Khan, and M. Sajid, "A Diet Recommendation System for Persons with Special Dietary Requirements," *J. Comput. Biomed. Informatics*, vol. 5, no. 1, pp. 153–164, 2023.

[9] E. Ahmed, M. Oumer, and M. Hassan, "Diabetes-focused food recommender system (DFRS) to enabling digital health," *PLOS Digit. Heal.*, vol. 4, no. 2, pp. 1–22, 2025, doi: 10.1371/journal.pdig.0000530.

[10] Q. Tan, Y. Qin, R. Tang, S. Wu, and J. Cao, "A Multi-Layer Classifier Model XR-KS of Human Activity Recognition for the Problem of Similar Human Activity," *Sensors*, vol. 23, no. 23, p. 9613, 2023, doi: 10.3390/s23239613.

[11] W. Wunnasri, P. Musikawan, and C. So-In, "A Two-Phase Ensemble-Based Method for Predicting Learners' Grade in MOOCs," *Appl. Sci.*, vol. 13, no. 3, p. 1492, 2023, doi: 10.3390/app13031492.

[12] T. E. Eyinla, R. A. Sanusi, and B. Maziya-Dixon, "Evaluation of in vitro and in vivo Glycemic Index of common staples made from varieties of White Yam (Dioscorea rotundata)," *Front. Nutr.*, vol. 9, p. 983212, 2022, doi: 10.3389/fnut.2022.983212.

[13] H. A. Raihana, P. H. Gunawan, and N. Aquarini, "Breaking Class Imbalance: Machine Learning Solutions for Stunting Detection," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 16, no. 2, p. 03, 2025, [Online]. Available: https://ejournal3.unud.ac.id/index.php/lontar/article/view/762

[14] S. Sarmini, W. Ma'arifah, and I. Tahyudin, "Sistem Pendukung Keputusan Berbasis K-Means untuk Evaluasi Keberhasilan Bisnis dan Nilai Perusahaan," *J. Sist. Inf. Bisnis*, vol. 14, no. 4, pp. 363–374, 2024, doi: 10.21456/vol14iss4pp363-374.

[15] William and M. E. Johan, "Implementation of Customer Segmentation Model using K-Means and DBSCAN for Fashion Industry Product Transaction," *Int. J. Informatics Vis.*, vol. 9, no. 6, pp. 2559–2568, 2025, doi: 10.62527/joiv.9.6.2978.

[16] Firdaus *et al.*, "Text clustering for analyzing scientific article using pre-trained language model and k-means algorithm," *Bull. Electr. Eng. Informatics*, vol. 14, no. 5, pp. 3728–3740, 2025, doi: 10.11591/eei.v14i5.9670.

Accredited Rank 2 (SINTA 2) based on Decree No. 225/E/KPT/2022 issued by Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, dated December 7, 2022. Published by LPPM Universitas Nusa Mandiri.