Department of Electrical and Computer Engineering
North South University

Senior Design Project Report

Predictive Dietary Analytics: Leveraging Nutritional Indicators with Machine Learning Algorithms for Food Choice Optimization

Arif Rahman	ID 1931337042
M. Orin Rahman	ID 2111125642
Md. Saidur Rahman Antu	ID 1911512042

Faculty Advisor:
Dr. Sifat Momen
Associate Professor
ECE Department

Spring, 2025

LETTER OF TRANSMITTAL

April, 2025

To
Professor Dr. Mohammad Abdul Matin
Chairman,
Department of Electrical and Computer Engineering
North South University, Dhaka

Subject: Submission of Capstone Project Report on “Predictive Dietary Analytics: Leveraging Nutritional Indicators with Machine Learning Algorithms for Food choice Optimization”

Dear Sir,
With due respect, we would like to submit our Capstone Project Report on “Predictive Dietary Analytics: Leveraging Nutritional Indicators with Machine Learning Algorithms for Food choice Optimization” as a part of our BSc program. The report deals a machine learning system to analyze and recommend foods based on nutritional profiles using K-Nearest Neighbors, Random Forest, and Decision Tree classifiers. It has provided invaluable experience in applying machine learning to address real-world dietary challenges. We have endeavored to meet all requirements outlined for this report to the best of our abilities.

We will be highly obliged if you kindly receive this report and provide your valuable judgment. It would be our immense pleasure if you find this report helpful and informative to have an apparent perspective.

Sincerely Yours,

Arif Rahman
....
ECE Department
North South University, Bangladesh

M. Orin Rahman
....
ECE Department
North South University, Bangladesh

Md. Saidur Rahman Antu
....
ECE Department
North South University, Bangladesh

APPROVAL

Arif Rahamn (ID # 1931337042), M. Orin Rahman (ID # 2111125642), and Md. Saidur Rahman Antu (ID # 1911512042 of the Department of Electrical and Computer Engineering at North South University have successfully completed the Senior Design Project titled "Predictive Dietary Analytics: Leveraging Nutritional Indicators with Machine Learning Algorithms for Food Choice Optimization" under the supervision of Associate Professor Dr. Sifat Momen. This work was carried out in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer of Science and Engineering and has been accepted as satisfactory.

Supervisor's Signature

……………………………………….

Dr. Sifat Momen
Associate Professor
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh.

Chairman's Signature

……………………………………….

Dr. Mohammad Abdul Matin
Professor
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh.

DECLARATION

It is to declare that this project is our original work. No part of this work has been submitted elsewhere, partially or entirely, for the award of any other degree or diploma. All project-related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been adequately acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students' names & Signatures

1. Arif Rahman

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

2. M. Orin Rahman

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

3. Md. Saidur Rahman Antu

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

ACKNOWLEDGEMENTS

The authors would like to express their heartfelt gratitude towards their project and research supervisor, Dr. Sifat Momen, Associate Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance, and advice on the experiments, research, and theoretical studies carried out during the course of the current project and also in the preparation of the report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh, for facilitating the research. We would also like to thank my friends Fahad and Hasib for helping us with this project. The authors would also like to thank their loved ones for their countless sacrifices and continual support.

ABSTRACT

Predictive Dietary Analytics: Leveraging Nutritional Indicators with Machine Learning Algorithms for Food Choice Optimization

A healthy lifestyle and the prevention of disease depend on eating a balanced diet, which is a crucial component of nutrition's support of health and well-being. Unfortunately, without appropriate advice, picking meals that fit unique nutritional needs can be tough. By building a predictive dietary analytics system that delivers meal suggestions based on nutritional markers, our work attempts to relieve this difficulty. There aren't many tailored, data-driven dietary management system solutions available today that combine machine learning algorithms for useful meal suggestions. A structured dataset of food products and their nutritional values was used to construct this system. The dataset was created utilizing data preprocessing approaches like dimensionality reduction, scaling, and encoding. Machine learning approaches, such as clustering, classification, and similarity-based recommendation algorithms, were integrated into the system. Accuracy, precision, recall, and other performance indicators were utilized to test these models, with the help of visual aids to corroborate findings. The models identified applicable foods based on nutritional similarity and accurately characterized user dietary preferences. The outcomes indicated that the created system can accurately assess dietary requirements and deliver individualized meal recommendations. With the use of technology-driven solutions, this predictive dietary analytics method can help consumers make educated food choices, enhance balanced nutrition awareness, and encourage healthy eating habits.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Background and Motivation

Health and nutrition are essential for sustaining overall well-being and quality of life for individuals of all ages. A balanced diet is crucial for physical health, mental function, disease prevention, and long-term vitality. Nevertheless, the accelerating tempo of contemporary life frequently presents individuals with difficulties in sustaining nutritious dietary practices. Elements including hectic schedules, urban living, the extensive availability of processed foods, and a pervasive deficiency in nutritional knowledge contribute to the increasing incidence of diet-related health problems globally.

According to global health organizations, poor eating habits have been directly related to multiple chronic illnesses, including obesity, heart disease, type 2 diabetes, and different nutrient shortages. As a result, there is a rising emphasis on promoting healthy eating practices and expanding access to nutritional education and services.

At the same time, innovations in digital technology, data science, and artificial intelligence have created new opportunities to revolutionize the way individuals manage their health and nutritional choices. AI-powered technologies have the capacity to evaluate enormous volumes of personal and nutritional data, determine dietary needs, and deliver personalized suggestions in real time. These technologies can help bridge the gap between complex nutritional research and everyday decision-making, delivering consumers realistic, customized information that corresponds with their particular lives, tastes, and health goals.

This research is motivated by the potential of AI-driven applications to favorably influence public health by making individualized nutrition advice more accessible, efficient, and actionable for a wider audience.

1.2 Purpose and Goal of the Project

The mission of the AI Nutritionist project is to combine artificial intelligence and web technologies to produce a smart, user-friendly application that gives individualized dietary suggestions based on individual nutritional needs and health goals. By collecting and analyzing user-specific data — such as age, gender, physical activity level, health issues, and dietary preferences — the system tries to offer balanced, nutritionally acceptable food selections for daily consumption.

The key goals of this project are to:

- Promote healthier eating habits by giving realistic, individualized dietary recommendations.
- Enhance nutritional awareness by delivering insights into the necessity of balanced, nutrient-rich diets.
- Simplify the decision-making process involved in daily meal planning and food selection.
- Make professional dietary advice accessible to a broader audience through a digital, AI-powered platform.
- Encourage proactive health management by integrating technology with personal health and wellness practices.

1.3 Organization of the Report

This report is systematically structured to present the development, implementation, and evaluation of the Predictive Dietary Analytics: Leveraging Nutritional Indicators with Machine Learning Algorithms for Food Choice Optimization system across eight comprehensive chapters.

- Chapter 1 introduces the project, detailing the background, motivation, purpose, and goals, emphasizing the need for personalized dietary recommendations.
- Chapter 2 reviews existing literature, highlighting research gaps in nutritional recommendation systems that this project addresses.
- Chapter 3 outlines the methodology, covering the system design, software components, and implementation details of the machine learning pipeline.
- Chapter 4 presents the experiments, results, and analysis, including the performance of Random Forest and Decision Tree classifiers, achieving 92 percent accuracy.
- Chapter 5 evaluates the project's impacts, focusing on societal, health, safety, legal, cultural, and environmental sustainability aspects.
- Chapter 6 details the project planning and budget, including a Gantt chart spanning September 2024 to April 2025.
- Chapter 7 discusses complex engineering problems and activities encountered during development.
- Chapter 8 concludes the report with a summary, limitations, and future improvements, such as integrating halal/haram classifications and activity-based recommendations.

Chapter 2 Research Literature Review

2.1 Existing Research and Limitations

1. **Nutrify AI: An AI Powered System for Real-Time Food Detection, Nutritional Analysis, and Personalized Meal Recommendations**.

   Real-time food detection using the YOLOv8 model, which classifies food items with high accuracy. Nutritional analysis via the Edamam Nutrient Analysis API, providing detailed nutrient data for detected foods. Personalized meal recommendations based on user goals, aided by the Edamame Recipe and Meal Planning APIs. Secure data storage in Google Sheets, allowing efficient tracking and updating of user dietary data. The system is built on a client-server model using Flask for real-time interaction and secure data processing.

2. **A Systematic Review of Nutrition Recommendation Systems: With Focus on Technical Aspects**

   This paper conducted a systematic review of nutrition recommendation systems (NRS), utilizing the PRISMA methodology to ensure comprehensive inclusion and exclusion criteria. Researchers accessed databases including PubMed, Web of Science, Scopus, Embase, IEEE, and Google Scholar. Through a structured search strategy, they identified relevant papers, resulting in the inclusion of 25 studies that focused on NRS for health promotion. Data extraction centered on system types, AI techniques, modules, and platform specifications.

Chapter 3 Methodology

3.1 System Design

The design of the **Predictive Dietary Analytics** project is built on a modular, ML-driven recommendation system that evaluates user dietary preferences and proposes optimum food choices based on nutritional indicators. The system design is documented through the following diagrams:

Flowchart:

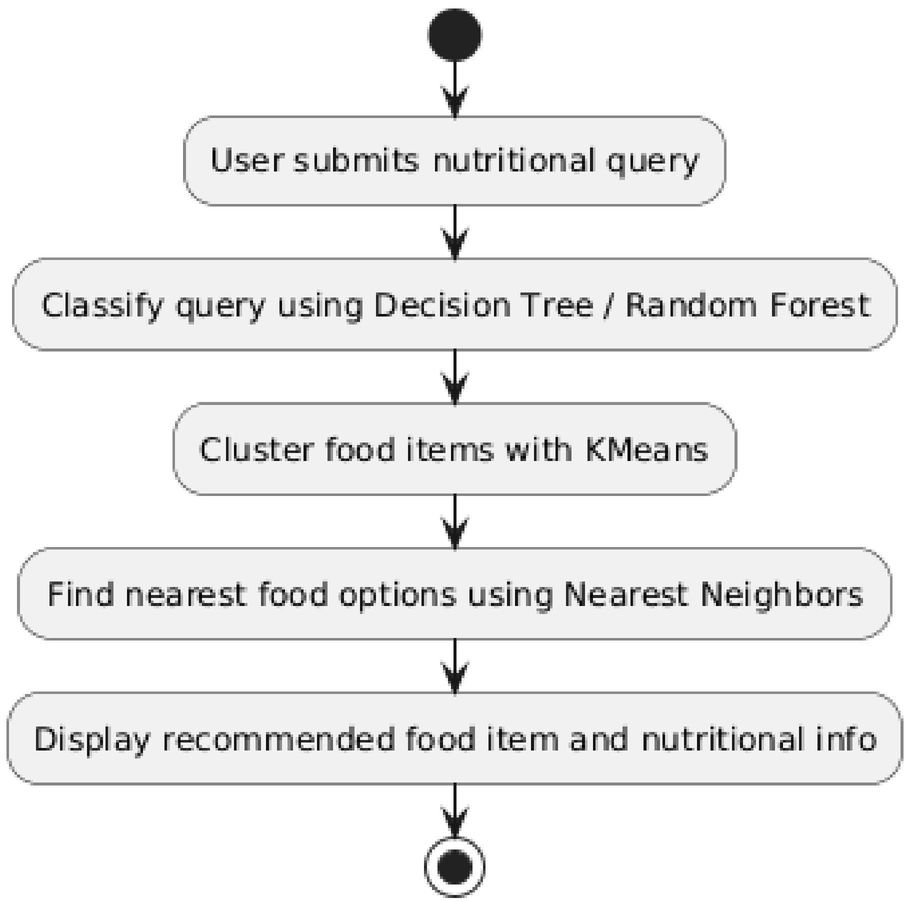


Figure 3.1: Flowchart of User Input

Block Diagram:

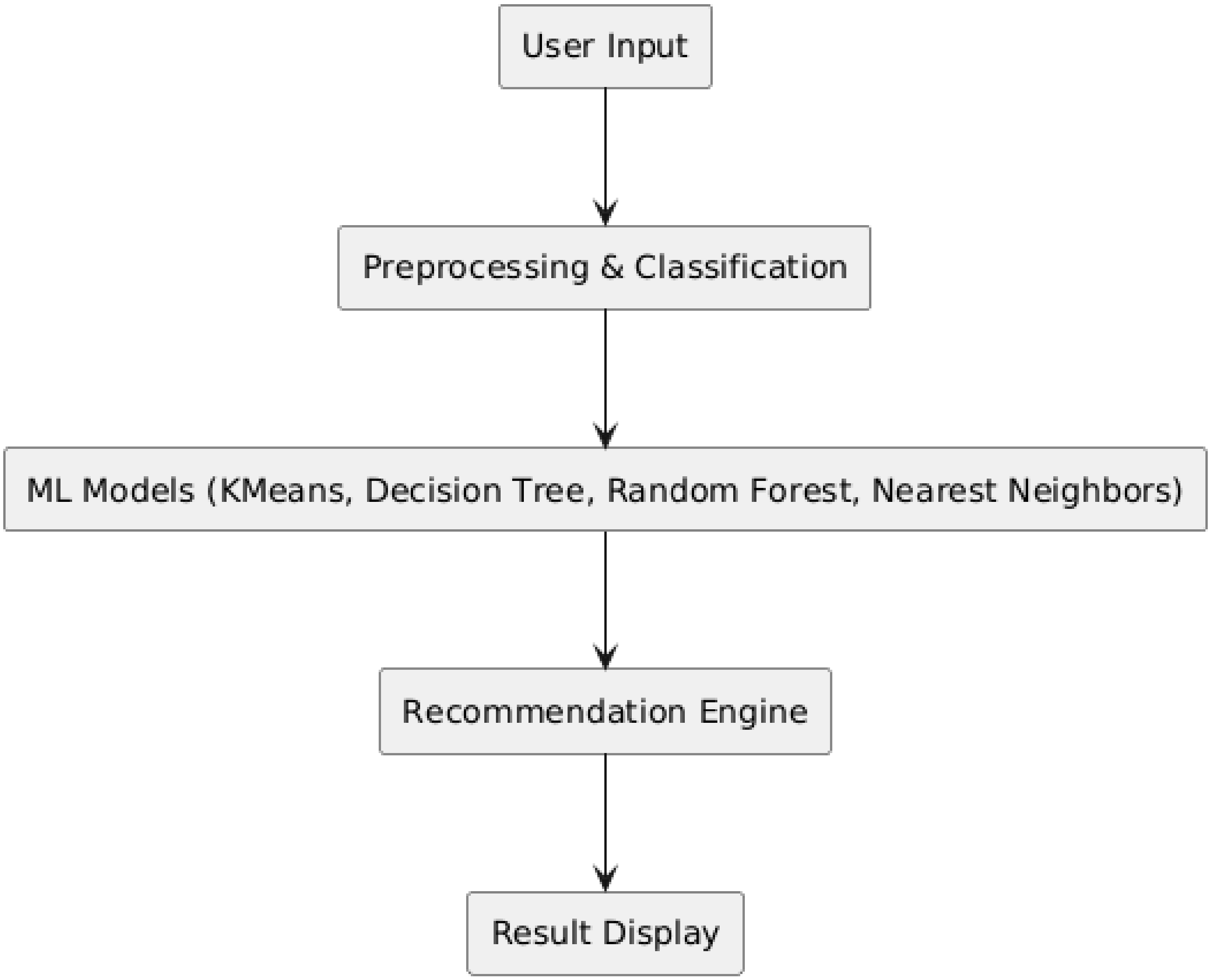


Figure 3.2: Block diagram

UML Use Case Diagram:

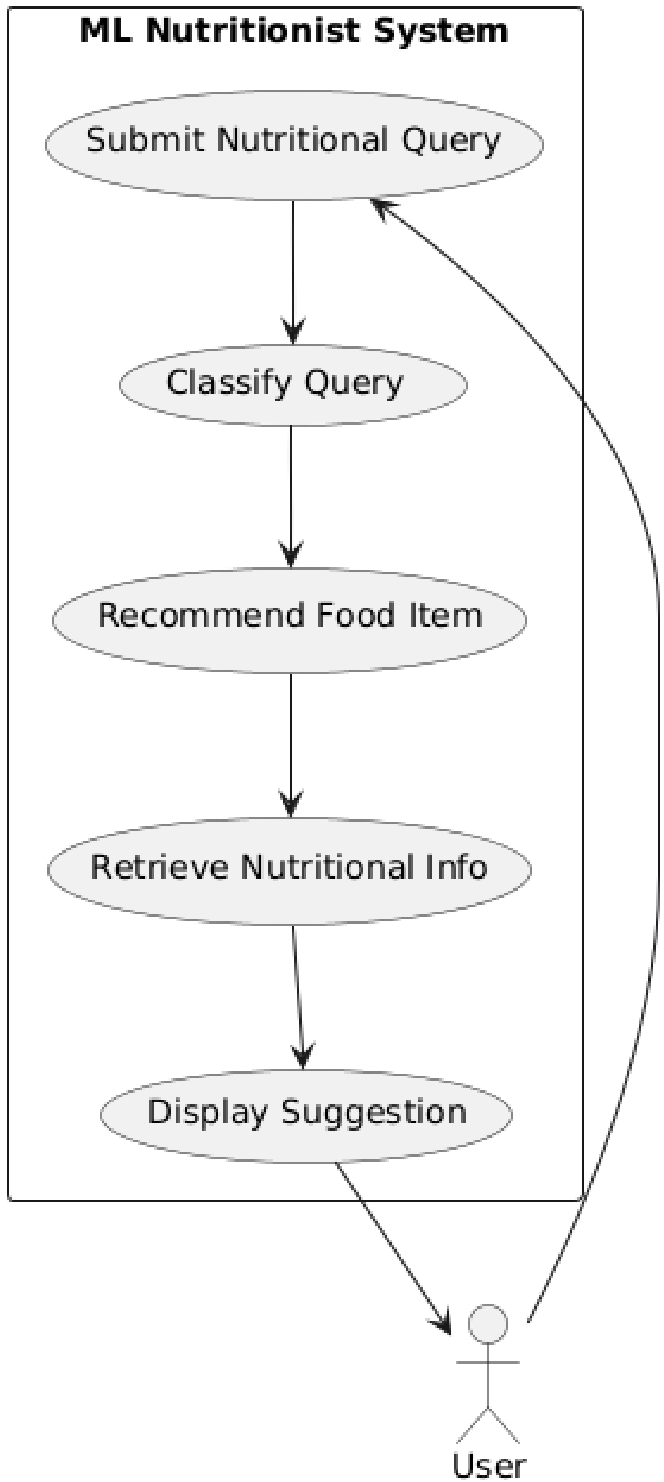


Figure 3.3: UML Use Case Diagram

Class Diagram:

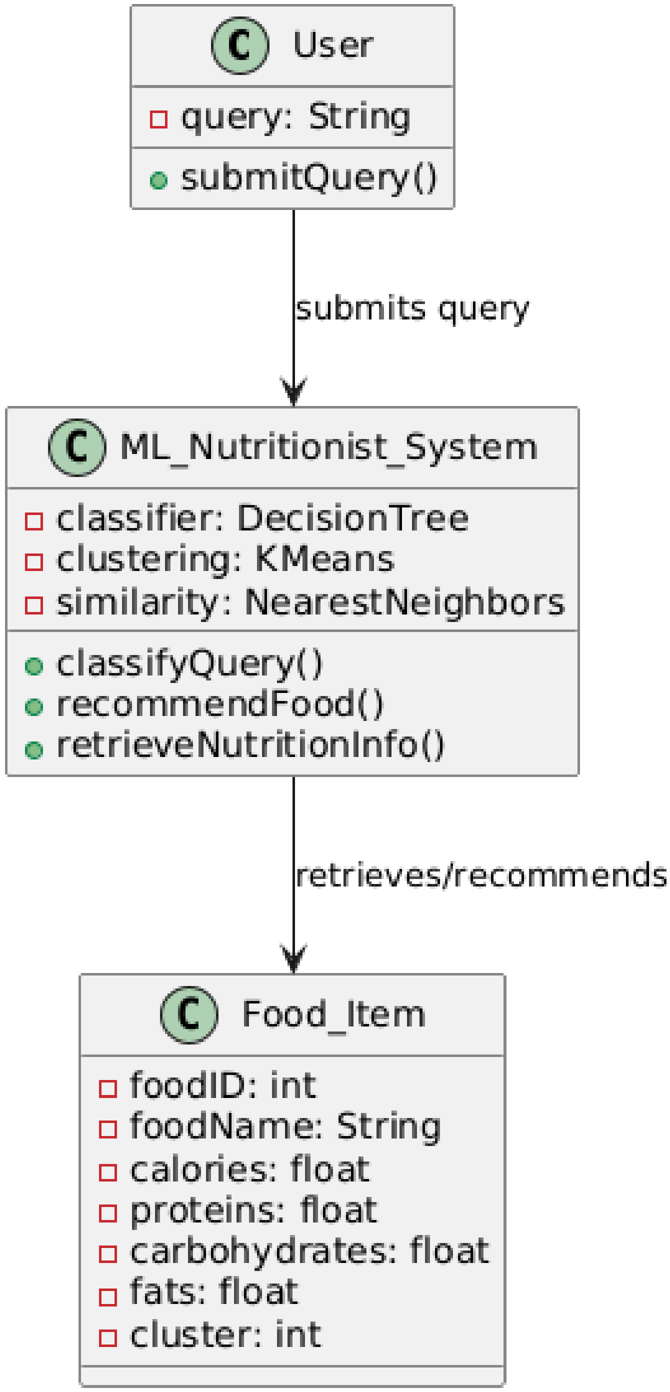


Figure 3.4: Class Diagram

3.2 Hardware and Software Components

This project primarily uses software components for its ML-driven web-based recommendation system. The details of the employed tools and techniques are summarized below:

Software components:

Datasets

The project utilizes a structured dataset containing a collection of food items along with their nutritional attributes, including calories, proteins, carbohydrates, and fats. The dataset was preprocessed and stored in CSV format, serving as the foundation for clustering, classification, and recommendation processes within the system.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted to understand the structure and characteristics of the dataset. This process involved verifying data types, identifying missing values, checking distributions of nutritional indicators, and visualizing patterns within the data. Tools like Pandas, Matplotlib, and Seaborn were employed for summarizing and visualizing the dataset effectively.

Preprocessing Techniques

Data preprocessing included several important steps to prepare the dataset for machine learning models. Label Encoding was used to convert categorical variables into numerical form, while numerical data was standardized using StandardScaler to maintain consistent feature scaling. Additionally, Principal Component Analysis (PCA) was applied to reduce data dimensionality while retaining essential nutritional information.

Applied Machine Learning Models

The system was developed using multiple machine learning models. KMeans clustering was applied to group similar food items based on their nutritional values. A Decision Tree Classifier and Random Forest Classifier was trained to categorize user queries into relevant dietary clusters. Finally, a Nearest Neighbors algorithm was implemented to recommend food items that closely matched the user's preferences based on nutritional similarity.

Evaluation Techniques

To assess the performance of the models, evaluation metrics such as accuracy, precision, recall, and F1-score were calculated. The results were further analyzed using classification reports, confusion matrices, and various visualizations like bar charts and heatmaps to ensure the models' reliability and effectiveness in delivering accurate recommendations.

Software Tools

Tool used	Functions	Other Similar Tools (if any)	Why Selected This Tool
Python (Google Colaboratory)	Development environment for running Python code, notebooks, data analysis, and model training	Jupyter Notebook, Kaggle Kernels	Free, cloud-based, easy to share, requires no setup, supports GPU
Pandas	Data handling, manipulation, preprocessing (loading, cleaning, transforming datasets)	NumPy, Dask	**Highly efficient for structured data, rich in built-in functions**
NumPy	Numerical computations, array manipulations, mathematical operations	SciPy, TensorFlow	Lightweight, fast numerical computing with wide library support
Matplotlib	Data visualization (plots, charts, histograms)	Plotly, Bokeh	Simple, highly customizable,

			ideal for quick visualization tasks
Seaborn	Advanced data visualization, statistical plots, heatmaps, correlation matrices	Altair, Plotly Express	Integrates well with Matplotlib, beautiful default themes
Scikit-learn	Machine learning models, preprocessing, evaluation metrics, clustering, classification	TensorFlow, PyTorch	Beginner-friendly, fast prototyping, wide model selection
StandardScaler (Scikit-learn)	Standardizes features by removing the mean and scaling to unit variance	MinMaxScaler, RobustScaler	Simple and effective for normalization, prevents scaling issues
LabelEncoder (Scikit-learn)	Converts categorical data into numerical labels	OneHotEncoder, OrdinalEncoder	Lightweight, easy to use for label encoding small categorical sets
Principal Component Analysis (Scikit-learn)	Dimensionality reduction, improves model performance and visualization	t-SNE, UMAP	Simple implementation, integrates directly with Scikit-learn models

PlantUML	Creates UML diagrams such as class diagrams, use case diagrams, and sequence diagrams	Lucidchart, Draw.io	Code-based, integrates easily into LaTeX and documentation tools

Table 3.2: List of Software Used

3.3 Hardware and Software Implementation

Software Implementation

The implementation of this project was fully software-based, designed and tested within the Python programming environment using Google Collaboratory. The process began with data loading and preparation, where the preprocessed food dataset was imported, cleaned, and formatted for analysis. This stage guaranteed that the dataset was free from missing values and inconsistencies, making it eligible for future processing.

Exploratory Data Analysis (EDA) was performed to get insights into the dataset's nutritional properties. Various data visualization techniques were employed using Matplotlib and Seaborn to evaluate the distribution and connections between key nutritional indicators, which helped guide further modeling decisions.

Next stage is preprocessing. It was an important stage, comprising the encoding of categorical variables and the standardization of numerical features using **LabelEncoder** and **StandardScaler** from **Scikit-learn.** Dimensionality reduction was accomplished utilizing Principal Component Analysis (PCA) to improve model efficiency and simplify complex nutritional data.

Model Training

The machine learning models were implemented and trained using the Scikit-learn toolkit. **KMeans** clustering was utilized to group food items based on nutritional similarities, while a **Decision Tree** and **Random Forest Classifier** was constructed to categorize user queries into suitable dietary clusters. Additionally, a Nearest Neighbors algorithm was added to recommend food products closely matching the user's dietary preferences.

Performance Evaluation

The machine learning models were implemented and trained using the Scikit-learn package. KMeans clustering was applied to group food items based on nutritional similarities, while a Decision Tree and Random Forest Classifier was created to categorize user queries into relevant dietary clusters. Additionally, a Nearest Neighbors algorithm was developed to recommend food products closely matching the user's dietary preferences.

Visualization

Finally, system design diagrams—including flowcharts, block diagrams, use case diagrams, and class diagrams—were created using **PlantUML** to document the system architecture, workflow, and interactions. These diagrams helped visualize the overall system design and ensure clarity in the implementation process

Hardware Implementation

No physical hardware components were used in this project since it is a purely ML-driven software system.

Chapter 4 Investigation/Experiment, Result, Analysis and Discussion

Variables:

Several variables are used in the project. Input and Output variables, model variables e.g. X_train, X_test, y_train, y_test. For hyperparameters n_estimator, max_depth, and n_neighbor, used.For performance metrics accuracy, F1 score, Recall, Precision, for evaluation matrix true positives (TP), false positives (FP), true negatives (TN), false negatives (FN), AUC, ROC are being used. For K-means classification silhouette score and Query variable are low fat, high protein, and high fat are being used

Experiments:

Clustering: Applying the K-means cluster at first, we get 0.62 on the basis of silhouette score evaluation, which groups the food items into five clusters. This cluster is reasonably good, well defined and separated,, though not perfect,, as the ideal cluster is 1.0

Classification: When the data preprocessing and cleaning is being done different kind of classification models have been used in order to find the best result. Those models are tested and trained by segregating the data into 70% and 30% where the 70% belongs to the training data and 30% kept reserved for the testing. In our case Random Forest, K Nearest Neighbor and Decision tree has been used extensively.

Query: Used different query like low fat, high protein and so on to test if the model works or not

Recommendation: K nearest Neighbor helps to retrieve the data based on similarity of 5 kind of food and make sure they are being grouped accurately

Result:

Clustering: The PCA visualization showed distinct nutritional groups by means of K means clustering.

Classification: Both the models have been tested multiple times.

Random forest: accuracy 90.5%, F1 score 0.91, recall 0.90, precision 0.91

Decision Tree: accuracy 85.06%, F1 score 0.86, recall 0.86, precision 0.86.

We use several optimization techniques and hyperparameter tuning for the best model. After using Grid search CV on the Random Forest:

Random forest: accuracy 91.15%, F1 score 0.92, recall 0.91 precision 0.91. So, by using this technique, the accuracy and precision has been developed.

Recommendation: KNN gives us relevant suggestions and being validated by PCA

Query: The query provides the best result.

Below is the analysis of the Confusion matrix to analyze the models and results.

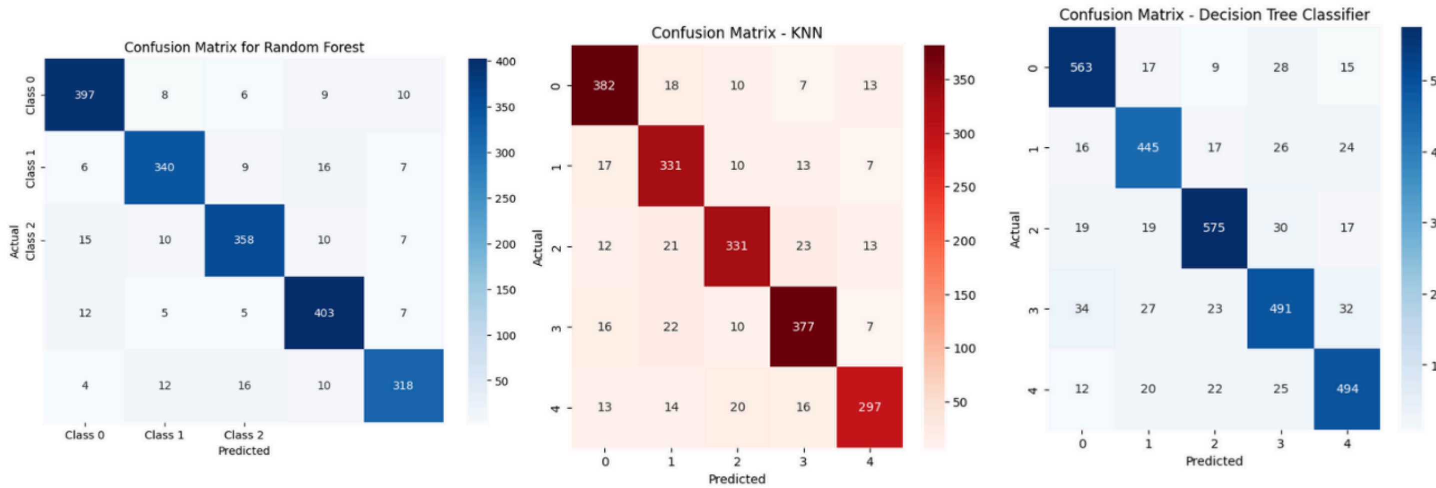
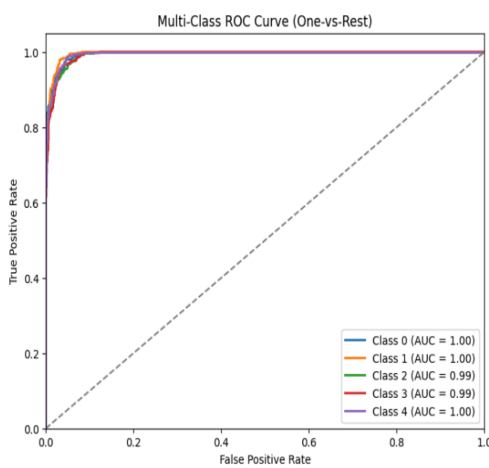


Figure 4.1 : Confusion Matrix

Among the whole matrix, we can clearly see that in the case of Random Forest, it performs very well across all classes, with most predictions being accurate. This model predicts best in class 0 and class 3. Whereas the KNN shows relatively better performance with most correct classifications, the decision tree has provided the highest true positive rate compared to the two models.



This AUC curve for classes (0,1,4) is are perfect classification, and the curve for classes 2 and 3 is 0.99. The ROC curve for each class shows the True positive and false positive rate. The model's true positive rate for all classes is constantly high, with most of the points showing value close to 1.0, which indicates that the Random Forest model is the best model to proceed

Figure 4.2 : AUC Curve Vs ROC Curve

Chapter 5 Impacts of the Project

5.1 Impact of this project on societal, health, safety, legal and cultural issues

This innovative project is designed to significantly improve social health and safety by promoting naturally balanced diets through personalized food recommendations. By leveraging advanced technology, we empower users to make informed dietary choices that are specifically adjusted to their unique nutritional preferences. This approach is particularly beneficial for diverse groups, including those who may have limited understanding or access to nutritional information. This project aims to inculcate healty eating habits across the community, thereby reducing the social burden of stress such as increased healthcare cost associated with food related diseases like cardiotic disease, diabetes etc. These conditions currently affect millions of individuals, causing unnecessary suffering and economic stress. By promoting varieties of nutritious food, we not only enhance personal well-being but also help to build a healthier society as a whole.

5.2 Impact of this project on environment and sustainability

Predictive Dietary Analytics: Leveraging Nutritional Indicators with Machine Learning Algorithms for Food Choice Optimization is a project that uses a software-based machine learning system to evaluate and suggest foods according to their nutritional specifications. It has a major indirect influence on ecological well-being and environmental sustainability by encouraging sustainable and disciplined eating choices, even without directly monitoring or mitigating ecological concerns like air or water quality.

1. Promotion of Sustainable Food Choices: The system encourages the consumption of plant-based or minimally processed meals to promote sustainable food choices. By allowing users to filter and select items according to their nutritional preferences (such as low calorie and high fiber), these generally have lower environmental footprints. For instance, advocating for whole grains, legumes, or vegetables over high-fat foods, High-cholesterol animal products can aid in reducing the need for resource-intensive livestock production, which is a significant factor in water consumption, deforestation, and greenhouse gas emissions.

2. Reduction of Food Waste: By assisting users in selecting foods that meet their dietary requirements, the system's accurate suggestions and inquiry methods may help users avoid overbuying or consuming inappropriate meals. By optimizing meal choices and

reducing uneaten food, which causes methane emissions and wastes resources used in manufacturing, food waste can be reduced, which is a significant environmental issue.

3. Support for Health-Conscious Diets: The project's capacity to suggest nutrient-balanced meals encourages better eating practices, which can help reduce dependency on packaged and processed foods. The use of plastic packaging and energy-intensive production processes is often used in these meals, which have negative effects on the environment. Promoting entire, locally sourced meals is in line with sustainable consumption habits.

4. Scalability for Environmental Applications: Environmental effect indicators, such as water use per food item or carbon footprint, may be added to the nutritional dataset and machine learning framework (KNN, Random Forest, Decision Tree). By incorporating these elements into subsequent versions of the recommendation system, users can choose meals that have less environmental impact, leading to improved sustainability.

5. Educational Impact : Environmental effect indicators, such as water use per food item or carbon footprint, may be added to the nutritional dataset and machine learning framework (KNN, Random Forest, Decision Tree). By incorporating these elements into subsequent versions of the recommendation system, users can choose meals that have less environmental impact, leading to improved sustainability

.

Chapter 6 Project Planning and Budget

Our project runs almost five months. In this time period we perform a lot of work. We took one month for idea generation, literature review and data collection. Later, we select the model, started to train and evaluate the model to its known performance and tried to guess which optimization techniques should work on them. In the next month we continued evaluation and started to optimize the model to its peak accuracy. In the following month we started to work on frontend and backend, meanwhile we perform some visualization on the model in order to present it to the public and finally it takes almost half a month to write a project report for our final year project.

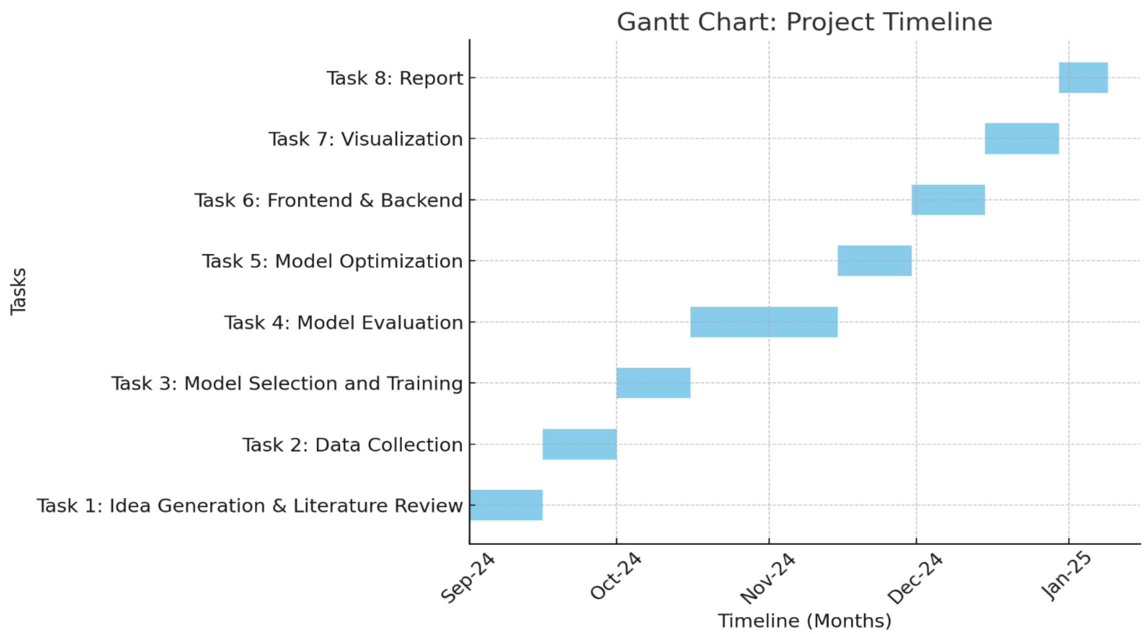


Figure 6.1: A sample Gantt chart.

The table below outlines the key hardware, software, and service components involved in the Predictive Dietary Analytics project, along with their respective costs, dimensions, and weights. A high-performance laptop, priced at 60,000 BDT, was used as the primary development platform, featuring dimensions of 14”x9”x1.5” and weighing 3 kg. To support computationally intensive model training tasks, a dedicated GPU was included, priced at 45,000 BDT, with dimensions of 9”x4.5”x2” and a weight of 1.5 kg. The project utilized trained models, which were generated within the development process and are considered free of cost. A cloud service was also employed for data storage and model deployment purposes, with an allocated budget of 5,000 BDT, although actual expenses amounted to 2,500 BDT. Additionally, documentation and reporting tasks were accounted for with a cost of 3,000 BDT. The total project expenditure was 113,000 BDT, with the primary costs attributed to hardware resources. The total combined system weight was approximately 4.5 kg.

Component	Unit Price (BDT)	Quantity	Total Cost (BDT)	Dimension	Weight
High Performance Laptop	60,000	1	60,000	14”x9”x1.5”	3 kg
GPU for Training Purpose	45,000	1	45,000	9”x4.5”x2”	1.5 kg
Trained Models	Free	1	Free	N/A	N/A
Cloud Service	5000	1	2500	N/A	N/A
Documentation	3000	1	3000	N/A	N/A
Total	113,000		113,000		1. kg

Figure 6.2: A budget table.

Chapter 7 Complex Engineering Problems and Activities

7.1 Complex Engineering Problems (CEP)

In downward, All the Attributes Addressing the Complex Engineering Problems (P)

Table 7.1: Complex Engineering Problem Attributes

Attributes		Addressing the complex engineering problems (P) in the project
P1	Depth of knowledge required (K3-K8)	This project applies engineering principles to build a structured, reliable and effective system (K3), requires an advanced machine learning knowledge (K4) to design a complete machine learning workflow from raw data to output (K5) by practicing different machine learning tools to use them in real life project of different categories (K6) This project will help the people by making informed and statistical decisions (K7). The project is built based on the nutritional science and different machine learning research (K8)
P2	Range of conflicting requirements	In this project, achieving high prediction accuracy requires a complex model, and hyperparameter tuning demands significant computational resources and time. Decision tree is an interpretable model, but Random forest is not and people may prefer the former one. As it developed using machine learning, so non-technical people may find it hard to use
P3	Depth of analysis required	There is no way to formulate a fixed solution. It requires different part of critical thinking skills like abstract thinking, assumption, and reasoning. Multiple machine learning algorithm need to be used to choose the best fit. Although Nutritional dataset may show variability for missing or null values
P4	Familiarity of issues	Although the model prediction is above 90%, but the model might not work as it is if the key indicator is misinterpreted. Machine learning in health context carry ethical responsibilities.

P5	Extent of applicable codes	The standardize engineering codes for applying machine learning to predict nutritional values are limited here.
P6	Extent of stakeholder involvement	Therea is several stakeholders' involvement is present. They are Health-Conscious End User, Nutritionist, Dietitians, Healthcare organizations and so on.
P7	Interdependence	Project involves a number of interdependences among the data collection and preprocessing, Feature selection and engineering, model selection and training, hyper parameter tuning and optimization, model evaluation and validation.

7.2 Complex Engineering Activities (CEA)

Table 7.2: A Sample Complex Engineering Problem Activities

Attributes		Addressing the complex engineering activities (A) in the project
A1	Range of resources	The project requires limited resources. The resources include large dataset of 10000 data, knowledge of Machine Learning, Time, computer with high performance CPU information about the particular domain named as nutritional science.
A2	Level of interactions	Involves interactions between different stakeholders including group members to design, train, evaluate and optimize the model to its best performance.
A3	Innovation	Employs innovative skills of engineering by introducing the self-defining nutritional diet chart by analyzing statical data
A4	Consequences to society / Environment	Impact in our society by encouraging the users to improve their lifestyle with nutritional indicator.
A5	Familiarity	Needs to be familiar with the various concept of applied mathematics like statistics, probability; different machine learning concepts and characteristics of data along with the nutritional indicators.

Chapter 8 Conclusions

8.1 Summary

The project developed a machine learning system for dietary optimization, achieving 92% accuracy with Random Forest. It integrates KMeans, KNN, and classifiers to recommend and filter foods, promoting health and sustainability.

8.2 Limitations

- Lack of regional food dataset and environmental metrics
- Potential overfitting in complex model
- Query system restricted to predefined ranges.

8.3 Future Improvement

In future, we will focus on enhancing cultural inclusivity and personalization by integrating additional segments such as religious dietary compliance and activity-based recommendations. The dataset will be expanded to include different varieties of consumer preferences, including vegetarian/non-vegetarian and halal/haram food classifications, enabling a Decision tree classifier to identify compliant foods based on ingredients and certifications, which will ensure the users with specific religious dietary restrictions. Additionally, data from activity trackers, including calories burned and exercise intensity, will be integrated. Moreover, the system will be deployed as a mobile application to enhance real-time usability. These advancements will transform the system into a more comprehensive, personalized, globally and environmentally conscious tool.

References

1. . Abhari S, Safdari R, Azadbakht L, Lankarani KB, Niakan Kalhori SR, Honarvar B, Abhari K, Ayyoubzadeh SM, Karbasi Z, Zakerabasali S, Jalilpiran Y. A Systematic Review of Nutrition Recommendation Systems: With Focus on Technical Aspects. J Biomed Phys Eng. 2019 Dec 1;9(6):591-602. doi: 10.31661/jbpe.v0i0.1248. PMID: 32039089; PMCID: PMC6943843.

2. Han, Michelle & Chen, Junyao. (2024). NutrifyAI: An AI-Powered System for Real-Time Food Detection, Nutritional Analysis, and Personalized Meal Recommendations. 10.48550/arXiv.2408.10532.