iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71338

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

A Data Mining Approach to Food and Health Recommendations

Mrs. Priyanka S[1], Mr. Ajith D[2], Mr. Bharathiraja V[3], Mr. Boopathi P[4], Mr. Kamalesh V[5], Mr. Ajay S[6]
[1]Assistant professor, Department of Computer Science and Engineering, Muthayammal Engineering College Rasipuram, India.
[2,3,4,5,6]UG, Department of Computer Science and Engineering, Muthayammal Engineering College Rasipuram, India.

Abstract: The exchange between nourishment and wellbeing is imperative for infection anticipation and in general well-being. This extend presents a information mining approach to analyze dietary propensities and suggest personalized nourishment choices custom-made to person wellbeing conditions. By utilizing huge datasets containing dietary data, dietary designs, and wellbeing records, the framework utilizes progressed information mining methods such as classification, clustering, and affiliation run the show mining to reveal important experiences. The proposed framework distinguishes relationships between dietary lacks and maladies, advertising noteworthy dietary proposals. It too highlights unthinkable nourishments and appropriate choices based on particular wellbeing conditions, such as diabetes, weight, or cardiovascular maladies. Compared to conventional measurable strategies, the framework illustrates made strides precision and proficiency in analyzing complex datasets.

I. INTRODUCTION

The project, A Data Mining Approach To Food And Health Recommendations , leverages advanced data mining techniques to analyze the relationship between food, nutrition, and health. It aims to provide personalized dietary recommendations to individuals based on their health conditions and dietary habits, thereby enhancing their overall well-being and aiding in disease prevention.

The system operates by utilizing large datasets containing information on nutritional values, disease symptoms, and dietary patterns. It employs machine learning algorithms like Random Forest for classification and prediction tasks. By analyzing these datasets, the system identifies potential health risks, recommends suitable foods, and warns against taboo foods for specific medical conditions.

This project overcomes the inefficiencies of traditional statistical methods by offering higher accuracy and performance. Through preprocessing, noise removal, and feature extraction, the system ensures data quality and reliability. The architecture includes both an admin module for dataset management and recommendation generation, and a user module for accessing personalized dietary advice.

The application is designed to support healthcare providers, researchers, and individuals by enabling effective nutritional planning and early detection of diet-related diseases. Its user-friendly interface and high- performance algorithms make it a scalable healthcare challenges.

II. OBJECTIVES

The primary objective of this project is to analyze the dietary patterns and their relationship with various diseases to provide personalized food recommendations. By utilizing data mining techniques, the system aims to identify diseases linked to an individual's dietary habits and recommend suitable foods to prevent or mitigate disease progression.

The project focuses on improving healthcare decision-making by extracting meaningful insights from large and complex datasets, emphasizing accuracy, performance, and user-friendliness.

Additionally, the project aims to support healthcare professionals by providing valuable insights into the nutritional ingredients conducive to disease rehabilitation, ensuring that individuals receive scientifically guided dietary advice. This approach not only aids in disease prevention but also fosters long- term health management through data-driven solutions.

The system is designed to identify diseases that may arise due certain food consumption patterns and recommend appropriate dietary changes to prevent or manage these conditions effectively. By leveragin large datasets containing nutritional information and disease data, the project aims to deliver personalized food recommendations and highlight taboo foods for specific health issues.

This project also seeks to address the limitations of traditional healthcare data analysis by improving the accuracy, efficiency, and reliability of dietary recommendations. Additionally, it serves as a valuable tool for healthcare professionals and researchers to uncover positive nutritional factors that support disease rehabilitation. Ultimately, the system promotes better decision-making, enhanced health management, and long-term well-being for users through data- driven insights.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

III. PROPOSED SYSTEM

The proposed system leverages data mining techniques, particularly Random Forest classification, to offer accurate and personalized food and health recommendations. By analyzing extensive datasets of nutritional information, disease- related data, and individual health conditions, the system generates tailored dietary suggestions.

This solution uses machine learning to classify diseases based on dietary patterns and recommend preventive measures. The system uses preprocessing techniques to clean and standardize data, feature extraction to identify the most relevant attributes, and algorithms to ensure accurate predictions.

A. Advantages

1) Improved Accuracy: By using advanced data mining algorithms, the proposed system provides more accurate disease predictions and dietary recommendations.
2) Personalized Recommendations: The system offers tailored advice based on individual health conditions and dietary habits, improving the quality of healthcare provided to users.
3) Scalability: The system can handle large and growing datasets, ensuring it remains effective as more health and food data are collected.
4) Efficiency: The automated data analysis and recommendation process saves time and reduces errors compared to manual methods.
5) Better Decision-Making: Healthcare professionals and individuals can make informed decisions about dietary changes and disease prevention based on data- driven insights.

IV. SYSTEM ARCHITECTURE OF FOOD RECOMMENDATION

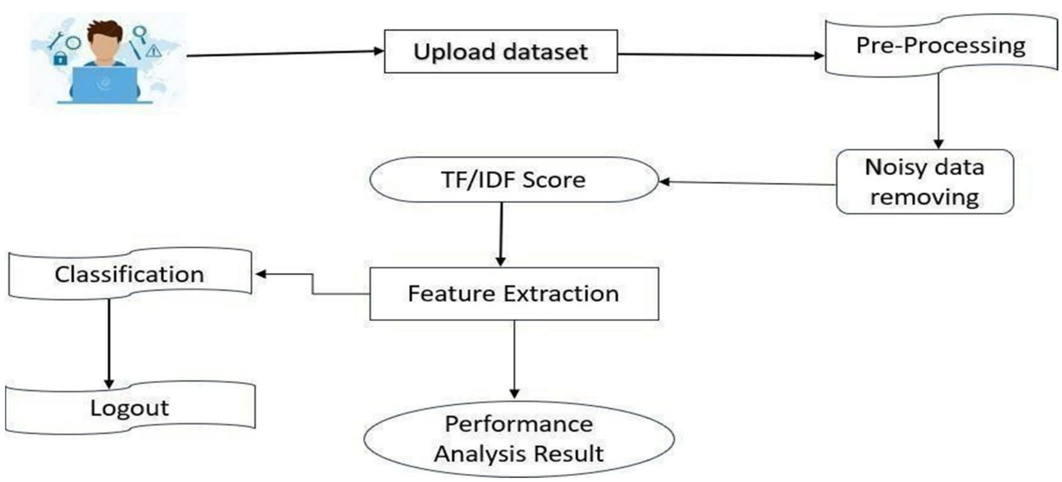


Figure 4.1 System Architecture of Food Recommendation

V. DATA SET

A. Food Data

This dataset would contain details about various foods, their nutritional information, and their relationship to health conditions.

Table 5.1 Food Nutritional And Health Data

FOOD ID	FOOD NAME	FOOD TYPE	CALORIES	FAT	ASSOCIATED DISEASES
1	APPLE	FRUIT	95	0.3	NONE
2	CARROT	VEGETABLE	41	0.2	DIABETES
3	CHICKEN	MEAT	335	30	HEART DISEASE
4	BROCCOLI	VEGETABLE	55	0.6	CANCER PREVENTION

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

B. Disease Data
This dataset will include information on different diseases, their severity, and potential dietary influences.

Table 5.2 Disease And Dietary Influence Data

DISEASE ID	DISEASE NAME	TYPE	SEVERITY	DISEASE FACTORS	RISK LEVEL
1	HYPERTENSION	CARDIOVASCULAR	HIGH	HIGH SODIUM INTAKE	HIGH
2	DIABETES	METABOLIC	MODERATE	HIGH SUGAR	MODERATE
3	CANCER	MALIGNANT	SEVERE	HIGH-FAT, LOW FIBER DIET	HIGH
4	OBESITY	ENDOCRINE	HIGH	OVEREATING	HIGH

C. Food Consumption
This will capture individual user inputs on food consumption, linking foods consumed to diseases or health conditions.

Table 5.3 Food Consumption Records

USER ID	FOOD ID	DATA CONSUMED	QUANTITY	MEAL TYPE
1001	1	2024-11-23	150	BREAKFAST
1002	2	2024-11-22	200	LUNCH
1003	3	2024-11-21	100	DINNER
1004	4	2024-11-20	124	SNACK

D. Health Data
User health data such as age, gender, existing conditions, and current health status.

Table 5.4 User Health Data

USER ID	AGE	GENDER	HIGHT	WEIGHT	EXISTING	HEALTH
1001	45	MALE	180	85	HYPERTENSION	STABLE
1002	35	FEMALE	160	70	NONE	HEALTHY
1003	23	MALE	175	95	OBESITY	HIGH RISK

E. Preprocessing
Preprocessing is a crucial step in any data analysis or machine learning project. It helps clean and transform raw data into a usable format for analysis and modeling. For your project analyzing health and dietary relationships.

1) Data Cleaning
Data cleaning involves removing errors, inconsistencies, and irrelevant information in your dataset.

a) Handling Missing Data:

Imputation: For numerical columns (e.g., calories, protein), you can replace missing values with the mean, median, or mode, depending on the distribution of the data.

Deletion: If missing values are too frequent (e.g., more than 30% of the data is missing in a column), you may decide to remove those rows or columns. Interpolation: For time-based data (e.g., health status changes over time), you can use interpolation to fill missing values based on nearby data points.

b) Removing Duplicates

Ensure there are no duplicate entries in your dataset, especially in the User Data and Food Consumption data. Remove duplicate rows using the drop_duplicates() function.

Standardize categorical values by converting them to lowercase or a consistent format.

c) Data Transformation

Data transformation involves converting data into a format that is suitable for analysis or modeling.

Normalization dataset includes variables with different scales (e.g., calories and protein), you might want to normalize or scale the data so that no single variable dominates the analysis.Min-Max Scaling: Scale numerical values to a [0, 1] range.

For machine learning algorithms, you need to convert categorical variables into numeric form.

2) Feature Engineering

This step involves creating new features or transforming existing ones to improve model performance.

Body Mass Index (BMI): If you have weight and height data, you can create a new feature for BMI, which can help in disease prediction.

3) Data Integration

If you're combining data from multiple sources (e.g., food data, disease data, and health data), you need to merge these datasets into a single unified dataset.

4) Data Sampling

If you have an imbalanced dataset (e.g., more data on healthy users than diseased users), you might want to perform oversampling or undersampling to balance the classes.Oversampling: Increase the number of examples in the minority class (e.g., using SMOTE).Undersampling: Reduce the number of examples in the majority class

5) Data Reduction

You may want to reduce the dimensionality of the dataset using techniques like Principal Component Analysis (PCA) if the number of features is very large. This can help reduce noise and improve model performance**.**

6) Data Split (Training and Testing)

Finally, split your dataset into training and testing sets for model evaluation. A typical split is 80% for training and 20% for testing.

VI. CONCLUSION

This project successfully demonstrates the use of data mining techniques to analyze the intricate relationship between dietary habits and health conditions. By leveraging datasets containing food details, disease information, and user health data, we developed a system that provides two key functionalities: predicting potential diseases based on dietary inputs and recommending suitable foods tailored to individual health profiles.

Key achievements of this project include

1) Comprehensive Data Integration and Analysis : By combining food, disease, and user data, the system highlights the influence of dietary choices on health outcomes. This integration enables more personalized insights into the role of nutrition in disease prevention and management.
2) Data Preprocessing for Reliable Analysis: Data preprocessing techniques such as noise removal, missing value handling, and feature extraction ensured the quality and reliability of the input data. This step significantly improved the accuracy of the system's predictions and recommendations.
3) Machine Learning for Disease Prediction and Food Recommendation The use of the Random Forest algorithm provided accurate disease predictions and relevant food recommendations.
4) User-Friendly Interface: A dual dashboard design, with separate views for administrators and end-users, makes the system accessible and practical. Admins can monitor datasets and insights, while users receive .

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

VII. FUTURE ENHANCEMENT

While the project achieves its primary objectives, there are areas for further Improvement and expansion: Incorporating Real-Time Data: Integrating real-time dietary tracking through wearable devices or mobile applications can enhance the system's accuracy and usability. Expanding the Dataset: Incorporating larger and more diverse datasets from global sources would improve the generalizability of the model. Advanced Recommendations: Adding support for personalized meal plans based on caloric needs, allergies, and cultural preferences could make the system even more user- centric.Exploring Other Algorithms: Testing other machine learning or deep learning models might yield better accuracy for specific tasks like disease prediction or pattern discovery.Admins can monitor datasets and insights, while users receive actionable recommendations based on their dietary inputs and health status.

VIII. RESULT

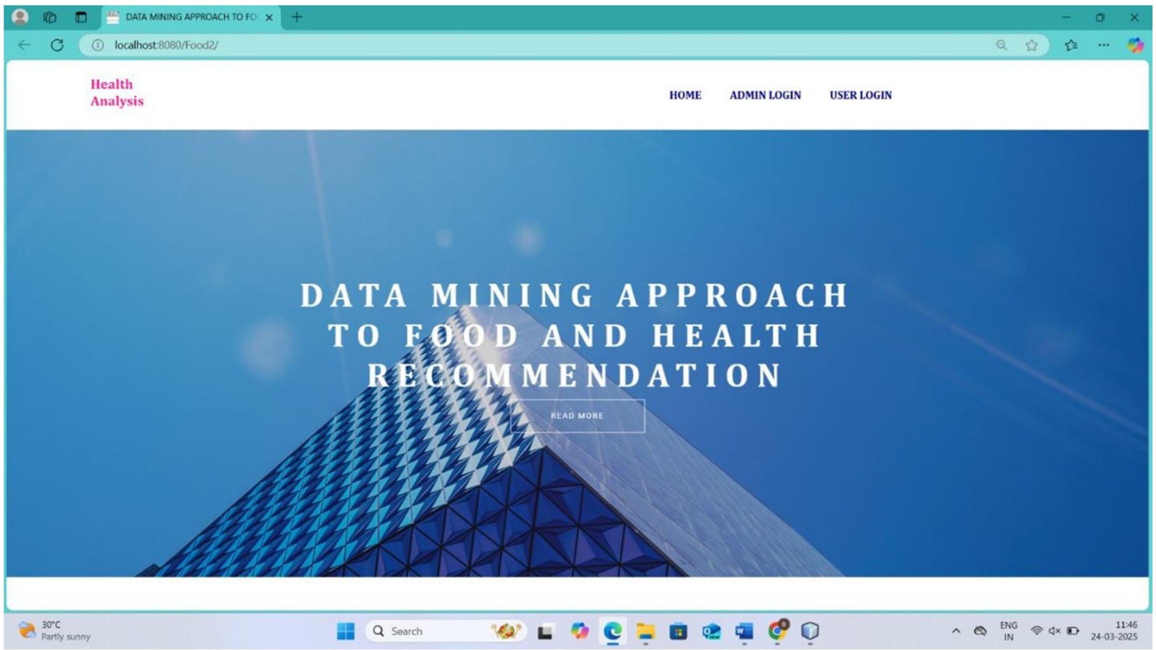


Fig 8.1 Home Page

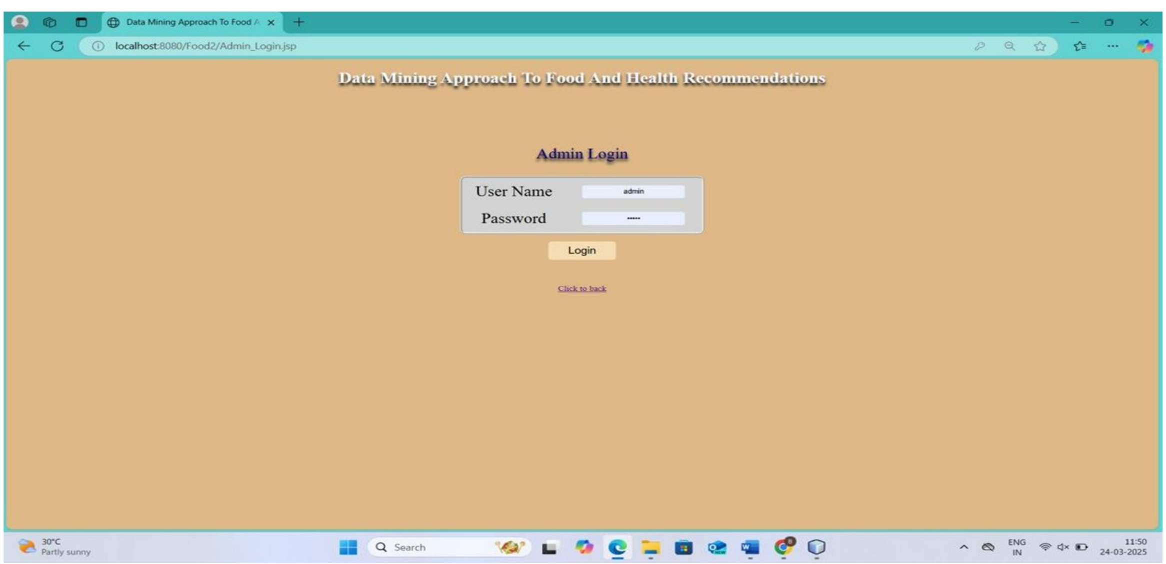


Fig 8.2 Admin Page

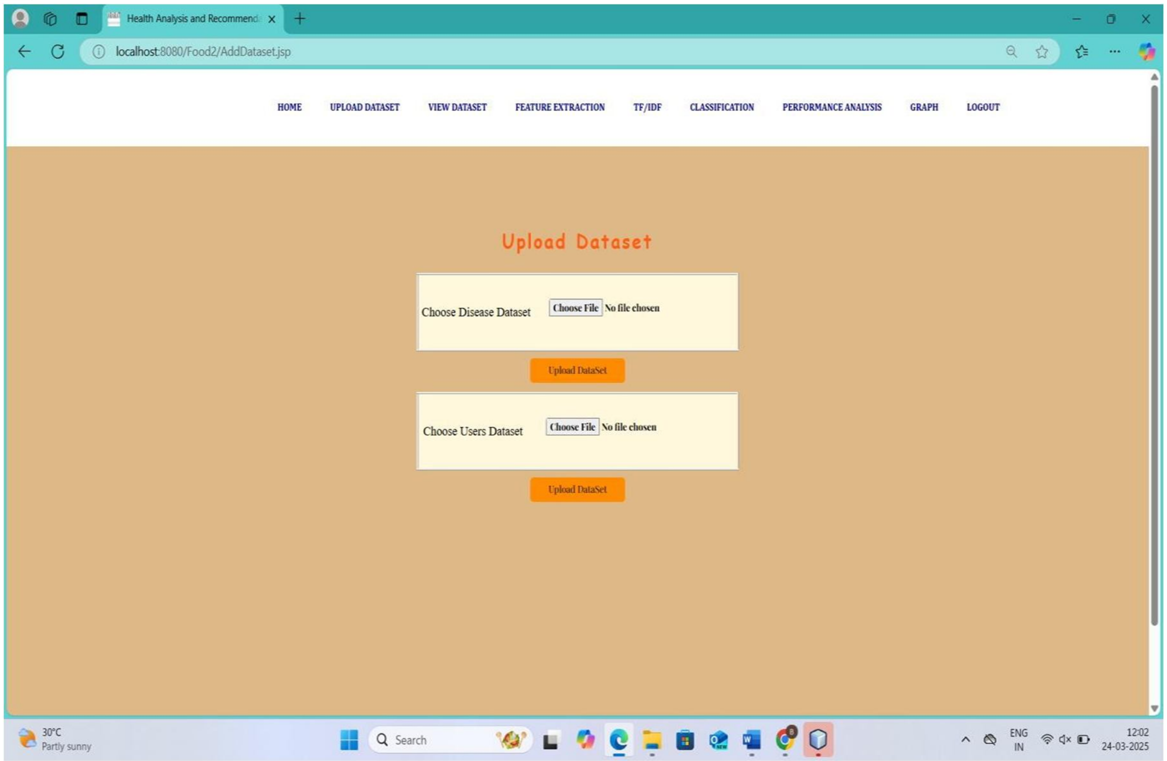
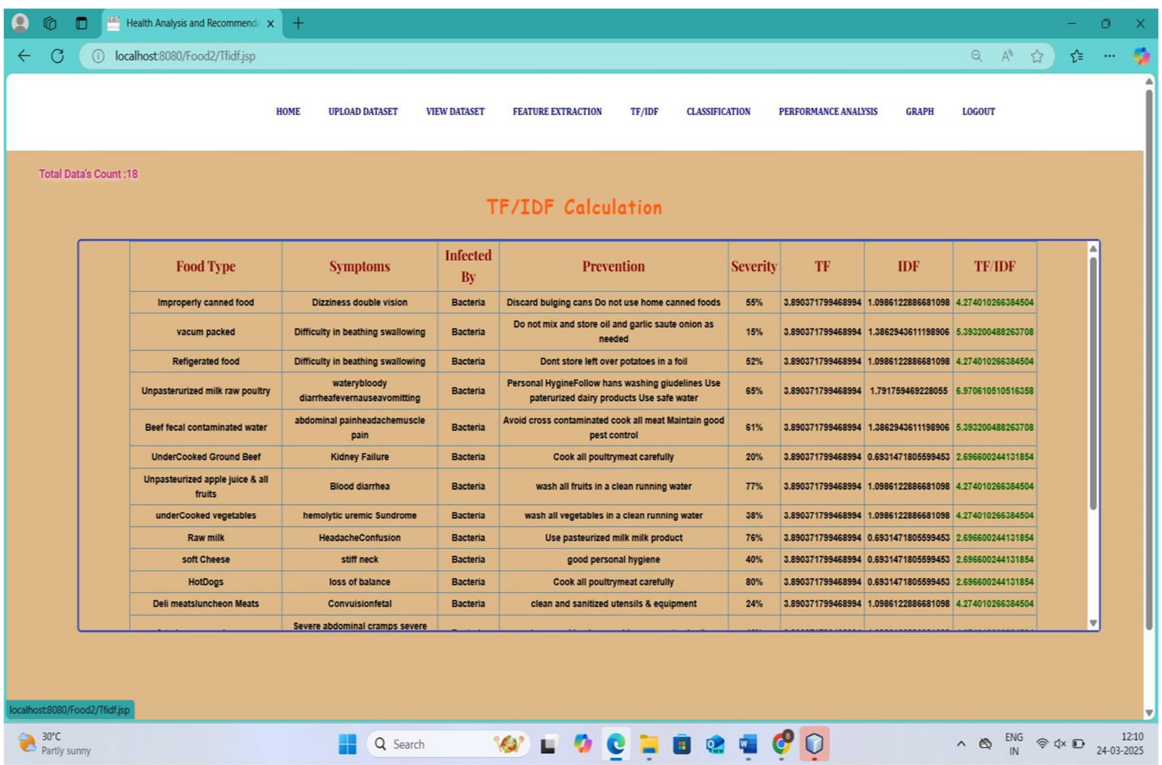


Fig 8.3 Upload Dataset



Food Type	Symptoms	Infected By	Prevention	Severity	TF	IDF	TF/IDF
Improperly canned food	Dizziness double vision	Bacteria	Discard bulging cans Do not use home canned foods	55%	3.890371799468994	1.0986122886681098	4.274010266384504
vacum packed	Difficulty in beathing swallowing	Bacteria	Do not mix and store oil and garlic saute onion as needed	15%	3.890371799468994	1.3862943611198906	5.393200488263708
Refigerated food	Difficulty in beathing swallowing	Bacteria	Dont store left over potatoes in a foil	52%	3.890371799468994	1.0986122886681098	4.274010266384504
Unpasterurized milk raw poultry	waterybloody diarrheafevernauseavomitting	Bacteria	Personal HygineFollow hans washing giudelines Use paterurized dairy products Use safe water	65%	3.890371799468994	1.791759469228055	6.970610510516358
Beef fecal contaminated water	abdominal painheadachemuscle pain	Bacteria	Avoid cross contaminated cook all meat Maintain good pest control	61%	3.890371799468994	1.3862943611198906	5.393200488263708
UnderCooked Ground Beef	Kidney Failure	Bacteria	Cook all poultrymeat carefully	20%	3.890371799468994	0.6931471805599453	2.696600244131854
Unpasteurized apple juice & all fruits	Blood diarrhea	Bacteria	wash all fruits in a clean running water	77%	3.890371799468994	1.0986122886681098	4.274010266384504
underCooked vegetables	hemolytic uremic Sundrome	Bacteria	wash all vegetables in a clean running water	38%	3.890371799468994	1.0986122886681098	4.274010266384504
Raw milk	HeadacheConfusion	Bacteria	Use pasteurized milk milk product	76%	3.890371799468994	0.6931471805599453	2.696600244131854
soft Cheese	stiff neck	Bacteria	good personal hygiene	40%	3.890371799468994	0.6931471805599453	2.696600244131854
HotDogs	loss of balance	Bacteria	Cook all poultrymeat carefully	80%	3.890371799468994	0.6931471805599453	2.696600244131854
Deli meatsluncheon Meats	Convuisionfetal	Bacteria	clean and sanitized utensils & equipment	24%	3.890371799468994	1.0986122886681098	4.274010266384504
	Severe abdominal cramps severe						

Fig 8.4 TF & IDF Calculation

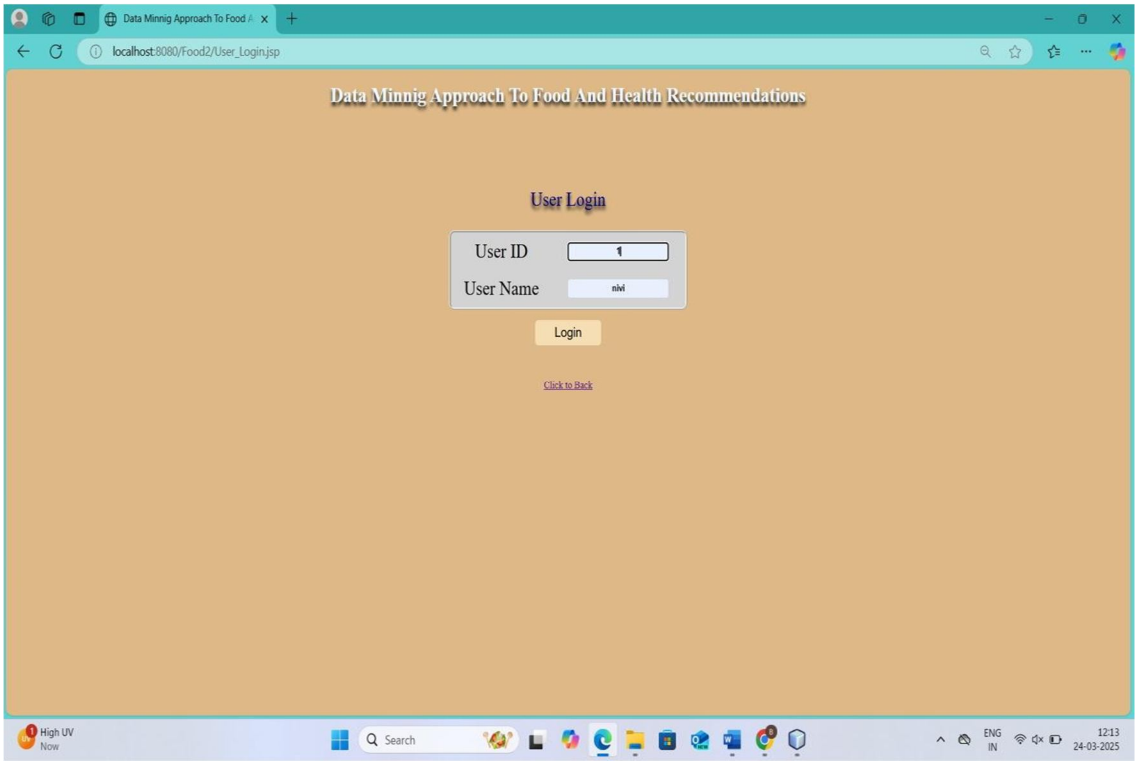
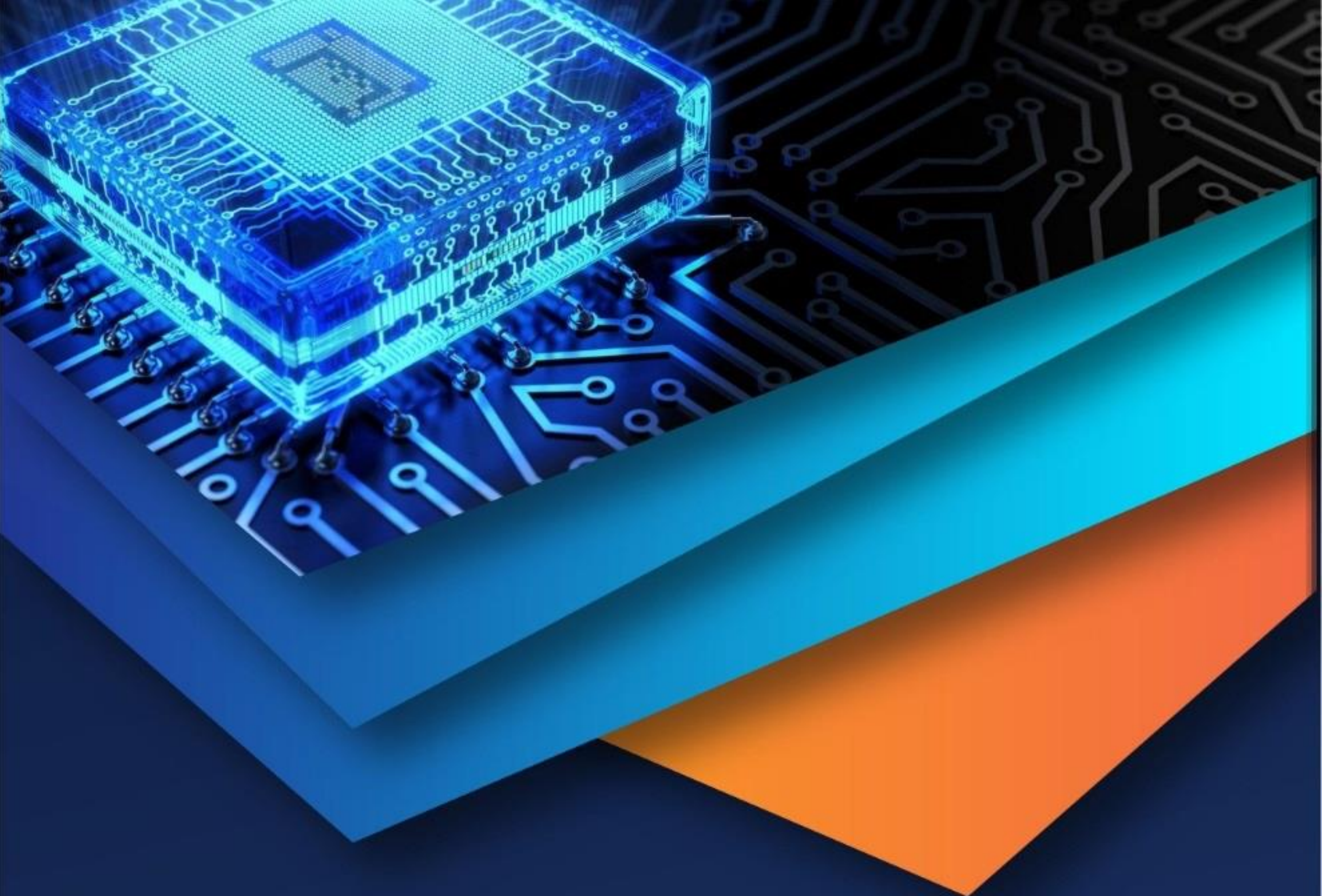


Fig 8.5 User Login Page

Fig 8.6 User Dashboard

REFERENCES

[1] Al-Dhabyani, W., Gomaa, M., & Farag, A.(2021). Food classification and its impact on health using Random Forest models.Computers in Biology and Medicine, 137, 104788.[DOI: 10.1016/j.compbiomed.2021.104788].
[2] Chen, M., Wang, L., & Yang, H.(2020).Nutritional status prediction using machine learning techniques.Journal of Medical Systems, 44(2),36.[DOI: 10.1007/s10916-020-1536-8].
[3] Kumar, A., et al.(2022) .AI-powered Food Recommendation Engine for Healthy Aging [DOI: 10.3389/frai.2022.816531].
[4] Li, X., Zhang, Y., & Wang, Y.(2019).Data mining techniques for personalized dietary recommendations based on health profiles.IEEE Access, 7, 129594-129605.[DOI:10.1109/ACCESS.2019.2940050].
[5] Liu, K., et al.(2020) Dietary Pattern Analysis: Investigating Food Patterns and Chronic Disease Risk using Data Mining Techniques" (American Journal of Epidemiology)[DOI: 10.1093/aje/kwaa002].
[6] Mohan, A., & Singh, D.(2019).Data mining approaches for food-disease relationships.ACM Computing Surveys, 51(6), 1-26.[DOI:10.1145/3309632].
[7] Pal, S., & Prakash, J.(2017) A Data mining approach to Predict Health Risks Based on Dietary Habits .IEEE Access,12, 321-338.[DOI:10.1007/98-030- 46993-2_14].
[8] Patel, S., et al.(2021).Personalized Nutrition Advice through Data Analytics and Food Frequency Questionnaires, [DOI:10.1093/jn/nxab034].
[9] Singh, et al.(2022). Food-Health Analytics: A Review of Current Research and Future Directions,[DOI:10.1111/1750-3841.16151].
[10] Tseng, C., & Wang, R. (2018).Integrating data mining and machine learning techniques for dietary behavior prediction.Computers in Biology and Medicine, 95, 135-142.[DOI: 10.1016/j.compbiomed.2018.02.002].



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY