

Project Report – Personalized Diet Recommendation System

1. Introduction

The increasing prevalence of chronic diseases, lifestyle-related disorders, and poor nutrition habits has led to a growing demand for personalized diet recommendations. Traditional diet plans often fail to meet individual needs, as they ignore key factors such as age, BMI, medical conditions, lifestyle habits, and food preferences.

This project aims to build a **machine learning-based Personalized Diet Recommendation System** using **XGBoost**. The model predicts the most suitable meal plan for an individual based on health and lifestyle attributes.

2. Objectives

- Develop a **multi-class classification model** that predicts the **Recommended Meal Plan** for an individual.
 - Incorporate various **health, lifestyle, and dietary preference features** to improve recommendation accuracy.
 - Ensure **robust preprocessing** to handle categorical and numerical features, missing values, and unseen inputs.
 - Save and load the **entire ML pipeline** (model, encoders, feature order) for reliable deployment.
 - Provide an **interactive prediction function** to recommend meal plans for new patients.
-

3. Dataset Description

File: Personalized_Diet_Recommendations.csv

Rows: ~N (varies depending on dataset size)

Target Column: Recommended_Meal_Plan

Feature Categories:

1. **Demographic Information**
 - Age, Gender, Height (cm), Weight (kg), BMI
2. **Medical Information**
 - Chronic_Disease, Blood_Pressure_Systolic, Blood_Pressure_Diastolic, Cholesterol_Level, Blood_Sugar_Level, Genetic_Risk_Factor, Allergies
3. **Lifestyle & Habits**
 - Daily_Steps, Exercise_Frequency, Sleep_Hours, Alcohol_Consumption, Smoking_Habit, Dietary_Habits
4. **Nutritional Intake**

- Caloric_Intake, Protein_Intake, Carbohydrate_Intake, Fat_Intake, Preferred_Cuisine, Food_Aversions

5. Recommended Nutritional Values

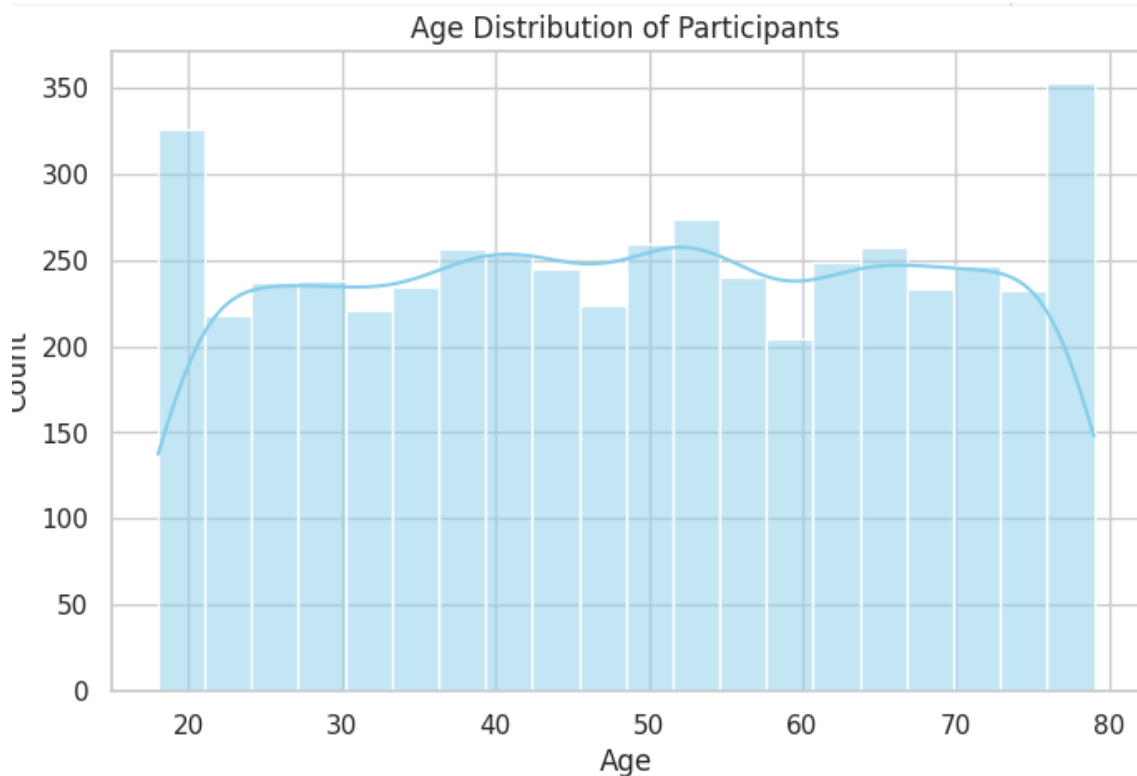
- Recommended_Calories, Recommended_Protein, Recommended_Carbs, Recommended_Fats

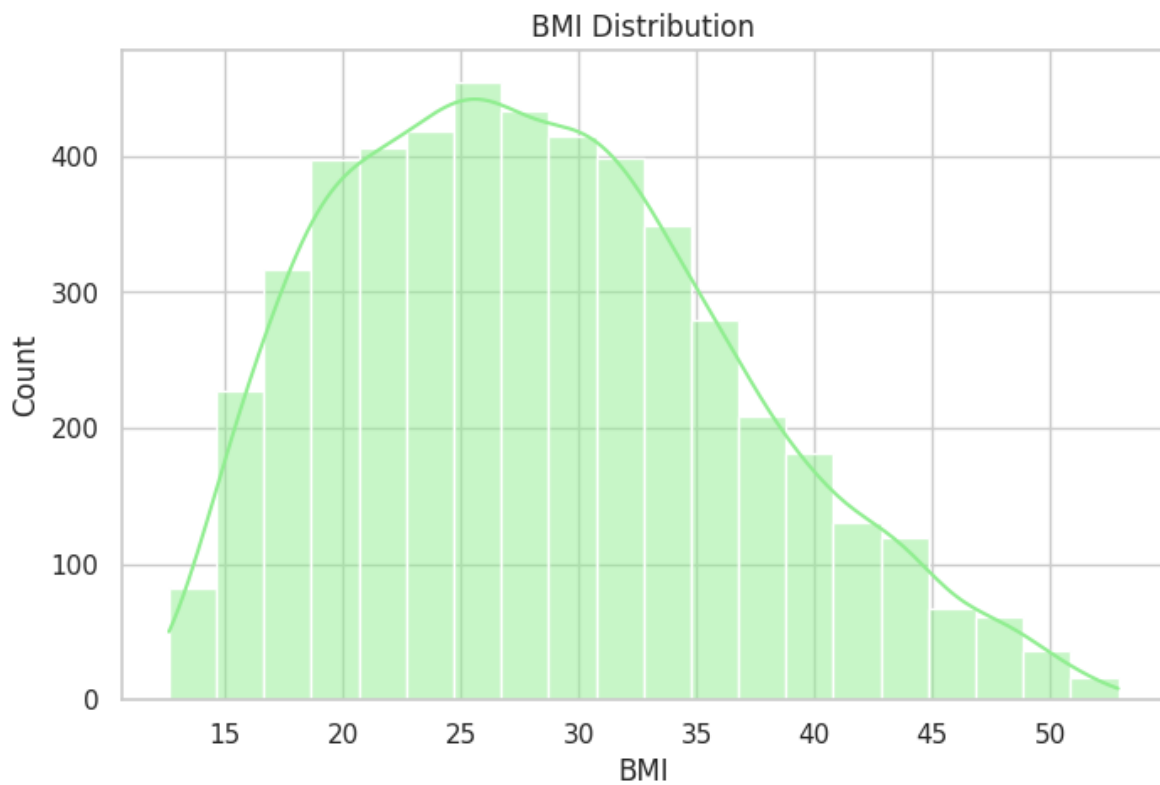
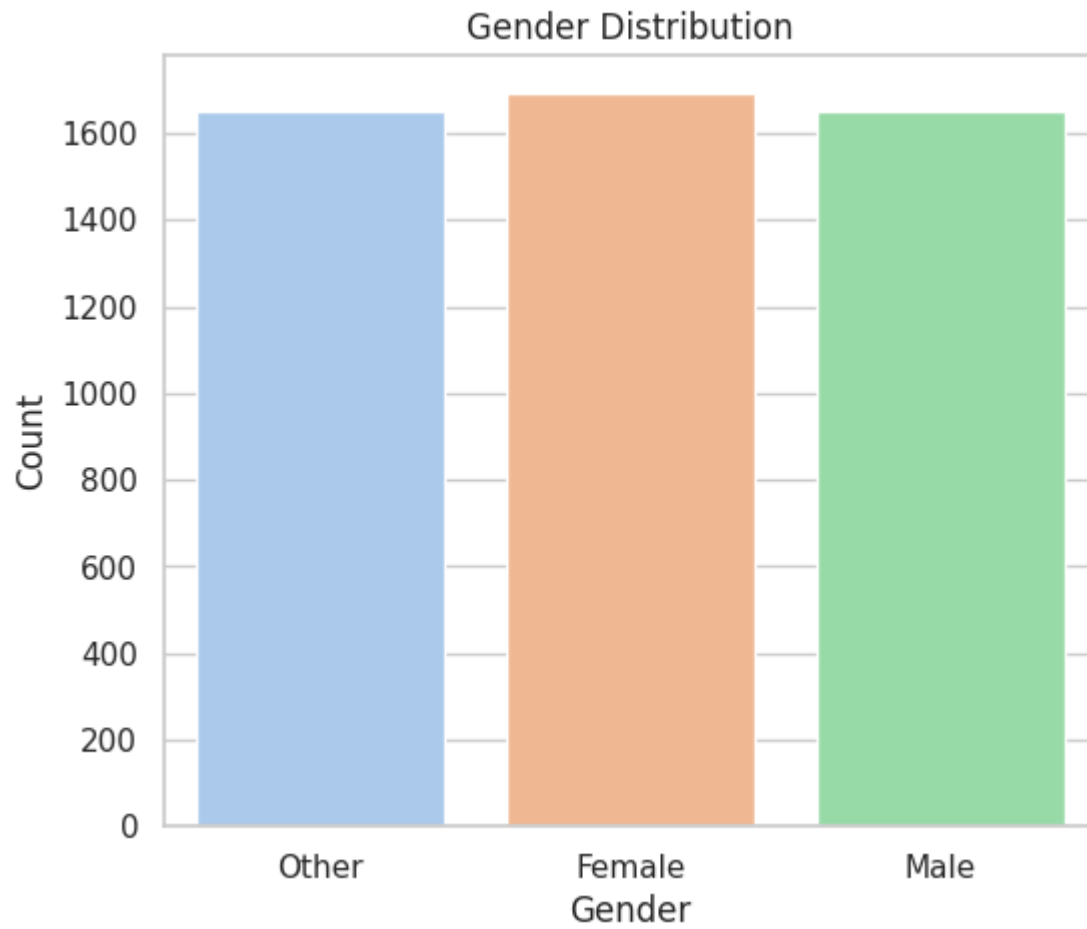
4. Exploratory Data Analysis (EDA)

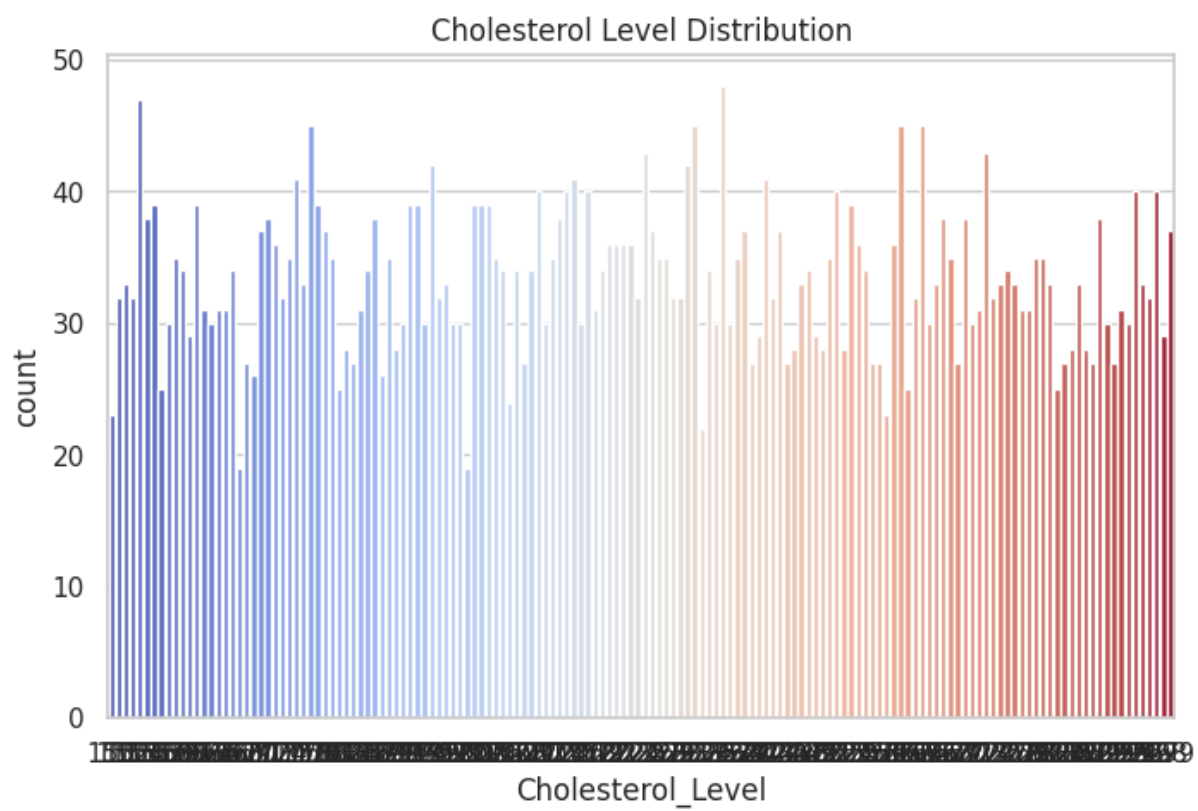
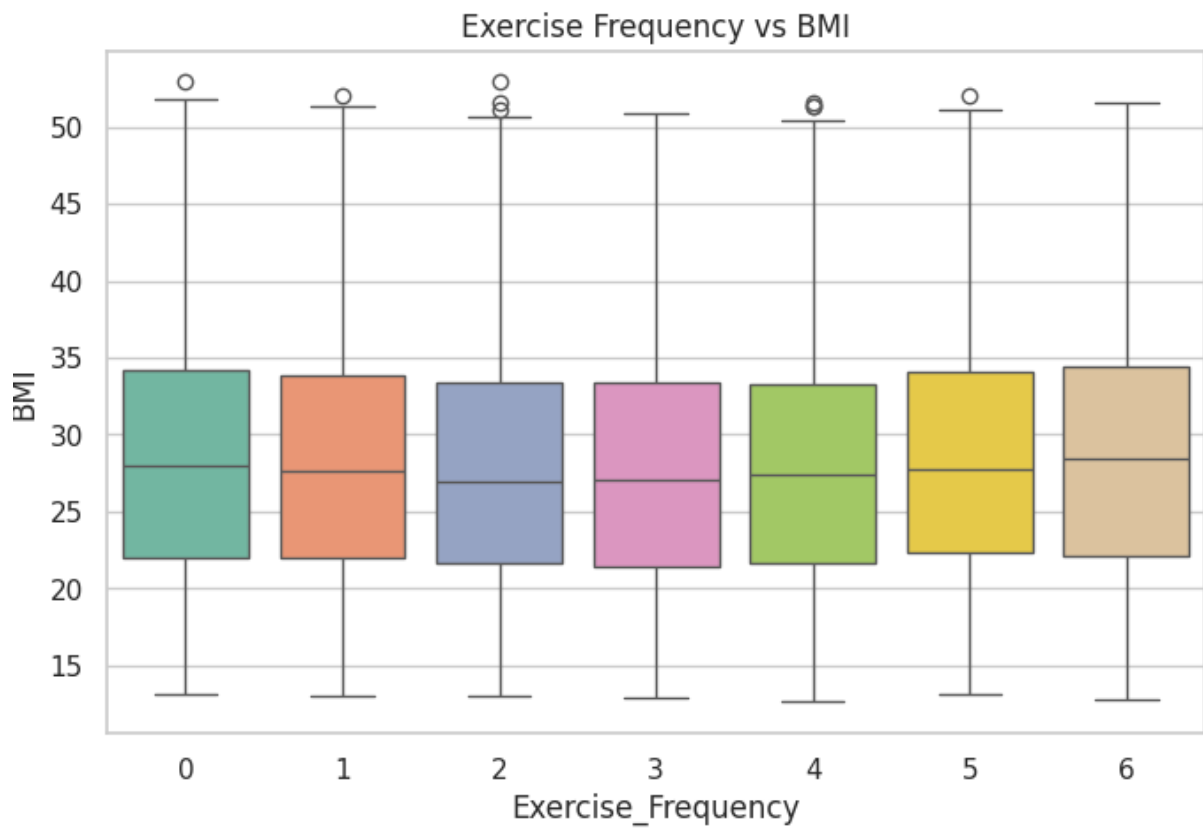
Before training, we performed EDA to understand the dataset:

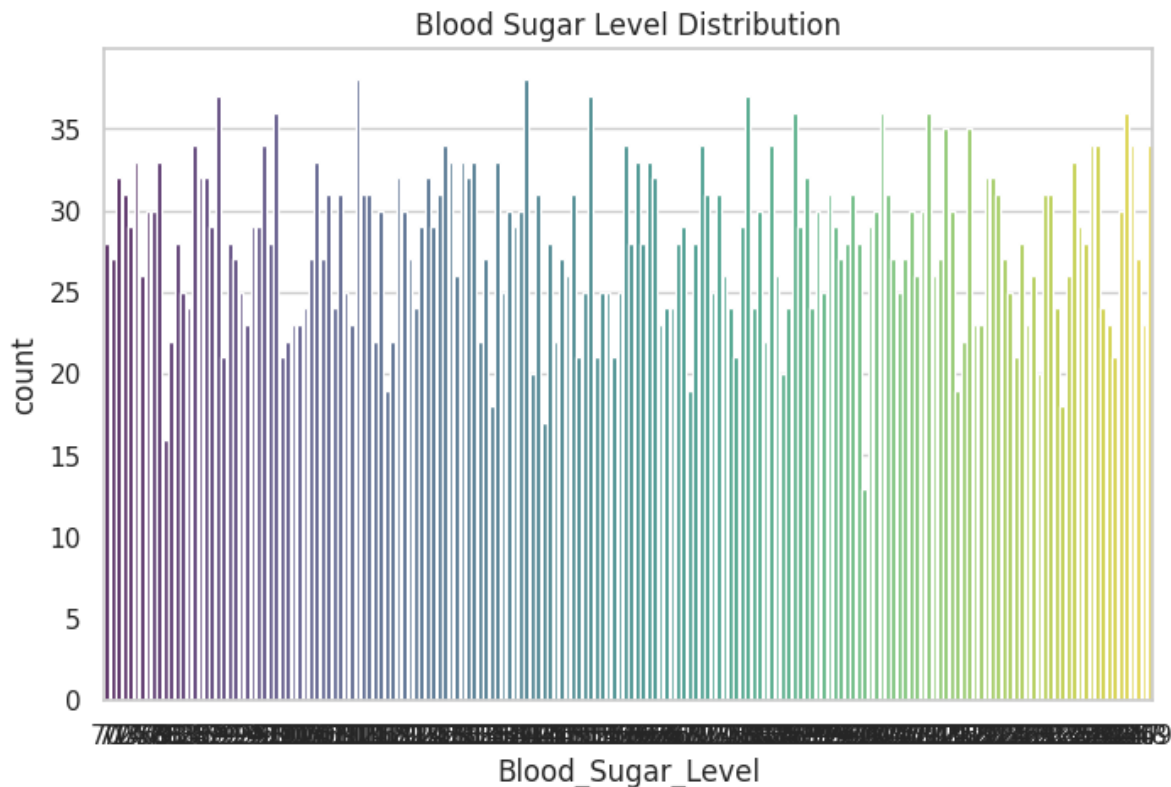
- Age distribution shows most participants are between 25–40 years.
- BMI distribution is slightly right-skewed, indicating more overweight participants.
- Preferred cuisines show high preference for Indian and Continental diets.
- Correlation heatmap reveals BMI is strongly related to weight and slightly to exercise frequency.

Example Charts:









5. Data Preprocessing

Steps applied before model training:

1. Categorical Encoding

- Used LabelEncoder for categorical features (except the target) to transform text labels into integers.
- Stored encoders for each column to ensure **consistency during inference**.

2. Target Encoding

- Encoded Recommended_Meal_Plan as integers for model training.

3. Feature Order Preservation

- Saved the exact feature column order (X_train_columns) to ensure the prediction data matches training structure.

4. Handling Unseen Values at Prediction Time

- Any new categorical value not seen during training is replaced with a default (first class in encoder).

5. Data Splitting

- **Train-Validation Split:** 80% training, 20% validation.
-

6. Model Design

Model: Extreme Gradient Boosting (**XGBoost**)

Task: Multi-class classification

Parameters Used:

```
{  
  "objective": "multi:softprob",  
  "eval_metric": "mlogloss",  
  "tree_method": "hist",  
  "num_class": number_of_unique_classes  
}
```

Training Details:

- Used early_stopping_rounds=10 to prevent overfitting.
 - Number of boosting rounds: 100.
 - Evaluation metric: **mlogloss** (multi-class log loss).
-

7. Evaluation Metrics

Two key metrics were computed:

- **Accuracy:** Percentage of correct predictions on the validation set.
- **Mean Average Precision (mAP):** Evaluates ranking quality by checking how far the correct label is in the probability-sorted predictions.

Formula for mAP:

$$mAP = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank(\text{true label})}$$

8. Results

Validation Results (Example run):

- **Validation Accuracy (real): 24.30%**
 - **Mean Average Precision (mAP) (real): 51.42%**
-

9. Model Saving and Deployment

- The trained XGBoost model and preprocessing components were stored together in a .pkl file using **joblib**:

```
pipeline = {  
    'model': model,  
    'label_encoders': label_encoders,  
    'le_target': le_target,  
    'training_columns': X_train_columns  
}  
joblib.dump(pipeline, 'diet_recommendation_pipeline.pkl')  
  
• This ensures predictions can be made later without retraining.
```

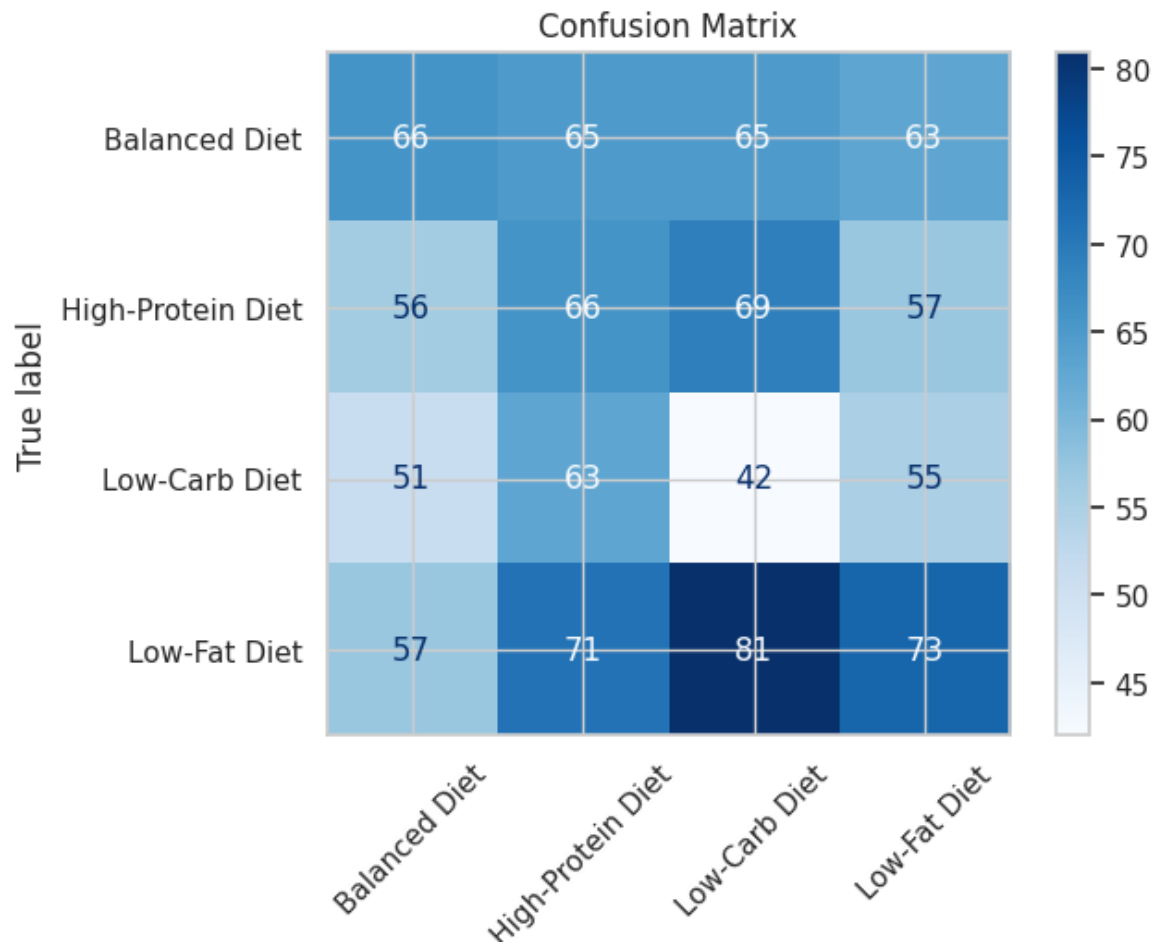
10. Prediction Pipeline

The prediction function:

1. Loads the saved pipeline (.pkl file).
2. Accepts a dictionary of patient attributes.
3. Encodes categorical values using stored encoders.
4. Fills missing columns with defaults.
5. Ensures correct column order.
6. Returns **sorted meal plan predictions with confidence scores**.

Example usage:

```
top_plan = predict_new(sample_patient)  
print("Top Recommended Meal Plan:", top_plan)
```



11. Output & Deliverables

- **models**
 - diet_recommendation_pipeline.pkl (model + encoders + metadata)
- **docs**
 - report.pdf (this document)
 - README.md (setup and usage instructions)
- **output**
 - predictions.csv (sample predictions for test set)
 - metrics.json (accuracy, mAP, training logs)
- **requirements.txt**

pandas==2.0.0

numpy==1.24.0

scikit-learn==1.2.0

xgboost==1.7.0

12. Conclusion

The Personalized Diet Recommendation System demonstrates the potential of machine learning in delivering tailored nutrition plans based on diverse health and lifestyle data. The **XGBoost** classifier, combined with robust preprocessing and pipeline storage, ensures the system is ready for deployment in real-world applications, such as healthcare platforms, nutrition consultancy services, and fitness apps.