Data Collection and Preprocessing Phase

Date	05 February 2026
Team ID	LTVIP2026TMIDS65517
Project Title	Advancing Nutrition Science through GeminiAI – NutriGen
Maximum Marks	6 Marks

Preprocessing Template

In the **NutriGen** project, data preprocessing focuses on user-provided health and nutrition-related textual input rather than images or large external datasets. Since the application uses a pre-trained Gemini AI model, no traditional dataset collection or image preprocessing is required. Instead, preprocessing ensures clean, valid, health-relevant, and structured text input to generate accurate, personalized nutrition guidance.

Section	**Description**
Data Overview	The data consists of user-entered inputs such as age, gender, weight, height, dietary preferences, health conditions (e.g., diabetes, PCOS), fitness goals, allergies, and desired calorie intake. No external dataset is directly used.
Text Cleaning	User inputs are cleaned by removing unnecessary spaces, special characters (if any), and handling null or empty values.
Input Validation	Ensures required fields (age, weight, health goal) are provided and values fall within logical and medically reasonable ranges.
Token Handling	The formatted input is passed to the Gemini model, which internally performs tokenization and advanced language processing.

Prompt Formatting	User health data is structured into a scientifically formatted nutrition prompt before sending it to Gemini AI for analysis and personalized diet generation.
Error Handling	Handles invalid health inputs, unrealistic calorie values, or API-related errors meaningful user feedback.
Data Preprocessing Areas	
Loading Data	User health details are collected directly through input forms (e.g., Streamlit or web interface fields such as text boxes, dropdowns, and sliders).
Input Validation	The system ensures: • Age is within a valid range (e.g., 5–100 years) • Weight and height values are realistic • Health goals are selected • Medical conditions are properly specified
Prompt Creation	The validated data is converted into a structured nutrition analysis prompt such as: "Generate a scientifically balanced diet plan for a 25-year-old male weighing 75 kg, 170 cm tall, BMI 25.9, aiming for weight loss, vegetarian diet, with mild hypertension."
Model Invocation	The formatted health prompt is sent to the Gemini AI model for personalized meal planning, macro distribution, and nutritional advice generation.
Output Handling	The generated nutrition plan is: • Displayed in structured sections (Breakfast, Lunch, Dinner, Snacks)