

Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	SWTID1720259116
Project Title	Nutrition App Using Gemini Pro : Your Comprehensive Guide to Healthy Eating and Well-being"
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Data Exploration and Preprocessing Report: Nutrition App with Gemini Pro

1. Introduction

This report outlines the data exploration and preprocessing steps planned for the development of a mobile nutrition app using Gemini Pro.

2. Data Sources

The primary data sources for this project will be:

- **Food Image Dataset:** A large and diverse dataset of food images with corresponding nutritional information (calories, macronutrients, vitamins, minerals). This can be obtained from public datasets or licensed sources.
- **Nutritional Information Database:** A comprehensive database of food items with detailed nutritional information. This can be sourced from government agencies, nutritional institutes, or food science databases.
- **User Data (Optional):** If the app allows user accounts, data on user profiles, goals, and tracked meals can be collected (with proper consent).

3. Data Exploration

Food Image Dataset:

- **Exploration:** We will explore the dataset size, image format variations, image quality, food category distribution, and presence of duplicates.
- **Preprocessing:** Cleaning steps might include removing low-quality images, addressing duplicate entries, and ensuring image formats are compatible with the app.

Nutritional Information Database:

- **Exploration:** We will explore the database size, food item coverage, completeness of nutritional information, and consistency of data formats.
- **Preprocessing:** This might involve standardizing units (calories, grams), handling missing values, and ensuring consistency in data structure.

User Data (Optional):

- **Exploration:** User profile data might reveal user demographics, dietary goals, and food preferences. Meal tracking data can uncover eating habits and patterns.
- **Preprocessing:** This might involve anonymizing user data, handling outliers in tracked meals, and addressing any inconsistencies in data formats.

4. Data Preprocessing Techniques

- **Data Cleaning:** Remove duplicates, address missing values, correct errors in food labels or nutritional information.
- **Data Transformation:** Standardize units, convert data types if necessary (e.g., text to numerical), format dates and times consistently.
- **Data Augmentation (Optional):** Techniques like image flipping or rotation might be used to increase the food image dataset size.

5. Tools and Technologies

- Python libraries like Pandas, NumPy, and Scikit-learn for data exploration and manipulation.
- Image processing libraries like OpenCV or PIL for handling food images (if applicable).

6. Expected Outcomes

- Clean, consistent, and well-formatted data for training the food recognition model with Gemini Pro.
- User-friendly data structure for storing and analyzing nutritional information and diet tracking data.
- Insights from user data that can inform future app features and functionalities.

