

Proposal: 食品營養成分數據分析與健康建議系統

1. Introduction (簡介)

- 背景知識

本專題的目標是開發一個基於食品營養成分數據的智能系統，專為現代消費者設計，提供精準的個性化飲食建議。隨著人們對健康生活方式的日益重視，飲食在疾病預防和健康管理中的角色變得越來越重要。雖然市面上已有飲食建議工具，例如 **MyFitnessPal**，但這些工具往往缺乏足夠的個性化和即時性，難以精確對應每個用戶的健康需求和生活習慣。通過對數據集中每種食品營養成分的深入分析，本系統將建立一個科學的模型，將用戶的日常飲食習慣與營養標準進行對比，幫助用戶全面了解飲食與身體健康之間的關聯，並做出更好的飲食選擇。

- 動機

隨著全球肥胖、心血管疾病、糖尿病等公共健康問題的日益嚴峻，飲食成為了健康管理中不可忽視的一環。根據世界衛生組織(WHO)2023年的數據，全球有超過20億成年人超重，其中6.5億人屬於肥胖。這些數據表明了全球飲食習慣與健康問題之間的緊密聯繫。然而，許多消費者在選擇食品時對其營養成分缺乏全面的了解，這導致了日常飲食中常見的營養失衡問題，如熱量攝入過高、缺乏必要的維生素等。為解決這一問題，本系統將應用現代數據分析技術和機器學習模型，為用戶生成個性化的飲食計劃，不僅涵蓋總卡路里和主要營養素的建議，還會考慮到用戶的個人健康需求和生活方式，幫助他們進行營養管理並達到健康目標。

- 專案目標/願景

本系統旨在通過分析用戶的健康數據、生活習慣與食品營養數據，為其提供量身定制的飲食建議。系統將根據用戶的具體需求，如減肥、增肌、控制血糖等，提供具體的建議，幫助用戶優化卡路里、蛋白質、脂肪、碳水化合物及維生素等營養素的攝取量。此外，通過結合可視化技術，系統將提供直觀的數據圖表，展示用戶的營養攝取趨勢和健康狀況變化，幫助用戶更加深入了解自己的飲食模式，從而有效調整飲食行為，提升整體健康水平。未來，系統可擴展至與其他健康應用整合，例如iPhone的健康應用，或者與健身追蹤器(如Apple Watch或Fitbit)聯動，進一步增強用戶的健康管理體驗。

2. Dataset (數據集)

- 目標數據集

為了支持系統的精準建議，本專案將使用來自 **Kaggle** 的食品營養成分數據集([Food Nutritional Facts](#))。該數據集涵蓋了1174種不同類型的食品的詳細營養成分，包括但不限於卡路里、蛋白質、脂肪、碳水化合物、纖維、糖、鈉、維生素和礦物質等信息。這些數據將作為系統建模和個性化飲食建議的關鍵基礎。

- 數據收集方式

首先，數據集將從 Kaggle 下載。下載後，我們將進行全面的數據清理和預處理，具體步驟包括：

- 去除重複項：確保數據集中的每一個項目都是獨立的，沒有冗餘數據。
- 處理缺失值：使用適當的方法填補缺失數據，避免對模型訓練造成負面影響。
- 數據正規化：將數據標準化處理，以確保不同的營養成分可以在同一尺度上進行分析和比較。

在預處理完成後，將根據模型的需求進行特徵工程，選擇和提取對飲食建議最為重要的營養成分指標。最終處理過的數據將被輸入機器學習模型，用於訓練和生成個性化的健康建議。此外，我們計劃定期更新數據集，確保系統能夠涵蓋最新的食品營養數據和健康建議趨勢，進一步提升系統的準確性和實用性。

3. Current Progress (當前進度)

- 初步方法 (Baseline Solution)

目前我們已經設計了基礎的數據分析方法，並以描述性統計來分析每種食品的主要營養成分，這包括卡路里、蛋白質、脂肪、碳水化合物、纖維等多種營養素。我們的初步方案是設計簡單的規則，例如根據不同食品的卡路里攝入量或營養成分的攝取量，生成初步的飲食建議，這些建議將根據使用者的日常飲食輸入，提供即時的飲食反饋。

- 目前進度

- 數據集選定：

已選定 Kaggle 的 Food Nutritional Facts 數據集作為模型訓練的基礎資料。

- 數據清理與預處理：

- 去除重複的數據項目，確保數據集的唯一性。
- 填補了少量缺失的營養成分值，避免在模型訓練中出現數據空缺問題。
- 使用標準化技術對不同的營養成分數據進行正規化處理，確保各種營養成分在同一尺度下進行比較。

- 規則設計：

目前已設計了簡單的規則，根據用戶輸入的飲食數據，依照卡路里或營養成分的攝取量生成初步建議，這些建議將根據每日推薦攝取量(DRI)的標準進行簡單的飲食調整建議。

4. Plan/Schedule for the Following Weeks (未來計劃/進度表)

第1週：需求分析與系統規劃

- 確定系統功能：包括數據輸入、營養分析、機器學習模型、數據可視化。
- 製作系統架構圖
- 確認數據來源

第2、3週：前端基本框架搭建

- 開發前端 UI，設計基本界面，包括用戶資料輸入、飲食記錄輸入等頁面。
- 確定營養數據可視化方式，並進行初步設計。

第4週：後端架構設計與數據庫搭建

- 建立後端系統，實現資料處理，與前端進行連接。
- 設計資料庫結構，實現用戶基本資料和飲食記錄的存取。

第5週: 營養分析

- 從用戶輸入的食品中提取營養成分數據。
- 根據每日營養建議標準, 進行用戶實際攝取量與建議量的比較。
- 開發基礎的飲食建議算法, 根據分析結果給出簡單的建議。

第6週: 機器學習模型訓練

- 研究並選擇適合的機器學習模型, 用於飲食建議。
- 訓練模型, 讓系統能夠根據用戶歷史飲食數據和行為模式, 生成個人化的飲食建議。

第7週: 模型優化與系統整合

- 將機器學習模型與後端服務進行整合, 實現自動化的建議生成。
- 實現長期數據跟蹤與趨勢分析功能, 提供用戶的飲食趨勢圖。

第8週: 測試與調整

- 修復出現的問題, 優化系統性能和用戶體驗。
- 蒐集測試數據, 分析模型預測結果的準確性, 進行相應的調整。

第9週: 專題報告準備

- 完成系統的最終調整和優化, 確保所有功能正常運行。
- 準備專題報告資料, 系統功能展示、技術架構說明等等。

5. What to Accomplish by the End of the Semester (學期末預期完成的內容)

開發一個對食品營養成分進行數據分析的健康建議系統, 主要目的是透過使用者輸入的日常飲食內容, 精確分析其攝取的各類營養素, 包括卡路里、蛋白質、脂肪、碳水化合物以及維生素的攝取量。系統會自動提取這些食品的營養成分數據, 並與用戶的每日營養建議攝取量進行比較。

系統會根據使用者的個人基本資料(如年齡、性別、體重、身高以及活動水平)計算其每日所需的營養標準。當用戶輸入其飲食記錄後, 機器學習模型會對這些數據進行分析, 提取出卡路里、蛋白質、脂肪、碳水化合物和維生素等營養成分的具體數值, 並與用戶的日常需求進行比對。此外, 隨著時間的推移, 系統會收集並學習使用者的飲食行為, 根據累積的數據改進建議。接著, 系統根據分析結果, 提供個性化的飲食建議, 幫助用戶在攝取營養均衡的同時達到健康目標, 像是控制體重、提高蛋白質攝取或減少糖分和脂肪的攝取等。該系統也能通過長期跟蹤用戶的飲食數據, 提供更加精確的建議, 並用圖表和報告展示飲食趨勢, 讓使用者更容易理解和調整自己的飲食習慣, 從而實現更好的健康管理。