

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

BỘ MÔN THỰC TẬP CƠ SỞ



BÁO CÁO ĐỀ CƯƠNG

ĐỀ TÀI MÔN THỰC TẬP CƠ SỞ

Giảng viên bộ môn: Kim Ngọc Bách

Sinh viên: Vũ Thị Đào

Mã sinh viên: B23DCCN123

Hà Nội – 2026

# 1. GIỚI THIỆU DỰ ÁN

## 1.1 Tên đề tài

Xây dựng hệ thống quản lý dinh dưỡng và dự đoán tiến trình cơ thể.

## 1.2 Lý do chọn đề tài

Trong những năm gần đây, tình trạng thừa cân, béo phì và rối loạn dinh dưỡng ngày càng gia tăng do chế độ ăn uống thiếu kiểm soát và lối sống ít vận động. Theo nguyên lý cân bằng năng lượng, sự thay đổi cân nặng phụ thuộc vào chênh lệch giữa năng lượng nạp vào và năng lượng tiêu hao (Hall et al., 2011). Tuy nhiên, đa số người dùng chưa có công cụ theo dõi và phân tích dữ liệu dinh dưỡng một cách khoa học.

Sự phát triển của Machine Learning cho phép xây dựng các mô hình dự đoán xu hướng thay đổi cân nặng dựa trên dữ liệu cá nhân. Do đó, đề tài “Hệ thống quản lý dinh dưỡng và dự đoán tiến trình cơ thể” được thực hiện nhằm kết hợp kiến thức dinh dưỡng với khoa học dữ liệu để giải quyết vấn đề thực tiễn.

## 1.3 Ý nghĩa của đề tài

- Hỗ trợ người dùng theo dõi lượng calo tiêu thụ hằng ngày.
- Tính toán nhu cầu năng lượng cá nhân (BMR, TDEE).
- Phân tích thặng dư/thâm hụt năng lượng.
- Dự đoán xu hướng thay đổi cân nặng dựa trên dữ liệu lịch sử.
- Đề tài góp phần nâng cao nhận thức về quản lý sức khỏe cá nhân dựa trên dữ liệu.

## 1.4 Tính ứng dụng

- Hệ thống có thể ứng dụng cho:
  - Cá nhân kiểm soát cân nặng
  - Người tập gym và thể hình
  - Huấn luyện viên cá nhân
  - Các nền tảng chăm sóc sức khỏe trực tuyến
- Hệ thống có thể mở rộng thành ứng dụng di động hoặc tích hợp thiết bị đeo thông minh.

## 1.5 Giá trị học thuật

- Ứng dụng bài toán Regression trong Machine Learning
- So sánh mô hình Linear Regression và Random Forest
- Đánh giá mô hình bằng MAE, MSE, R<sup>2</sup>
- Trực quan hóa dữ liệu theo chuỗi thời gian
- Kết hợp liên ngành: dinh dưỡng học và khoa học dữ liệu

## **2. CƠ SỞ LÝ THUYẾT VÀ CÔNG NGHỆ SỬ DỤNG**

### **2.1 Cơ sở lý thuyết về năng lượng**

#### **2.1.1 BMR (Basal Metabolic Rate)**

Tỷ lệ chuyển hóa cơ bản (BMR) là lượng năng lượng tối thiểu mà cơ thể cần để duy trì các hoạt động sinh lý cơ bản như hô hấp, tuần hoàn và duy trì thân nhiệt trong trạng thái nghỉ ngoại hoàn toàn. BMR chiếm tỷ lệ lớn trong tổng năng lượng tiêu hao hằng ngày của cơ thể[1].

Trong thực tế, một trong những công thức được sử dụng phổ biến để ước tính BMR là công thức Mifflin–St Jeor, được đánh giá có độ chính xác cao đối với người trưởng thành[2]. Công thức được xác định dựa trên các biến nhân trắc học gồm cân nặng, chiều cao và tuổi:

Đối với nam:

$$\text{BMR} = 10W + 6.25H - 5A + 5$$

Đối với nữ:

$$\text{BMR} = 10W + 6.25H - 5A - 161$$

Trong đó:

W là cân nặng (kg),

H là chiều cao (cm),

A là tuổi (năm).

Việc tính toán BMR giúp xác định mức năng lượng nền của mỗi cá nhân, làm cơ sở cho các bước phân tích tiếp theo[2].

#### **2.1.2 TDEE (Total Daily Energy Expenditure)**

Tổng năng lượng tiêu hao hằng ngày (TDEE) phản ánh lượng năng lượng cơ thể sử dụng trong một ngày, bao gồm năng lượng cho chuyển hóa cơ bản, hoạt động thể chất và hiệu ứng nhiệt của thực phẩm.

TDEE được tính bằng cách nhân BMR với hệ số vận động tương ứng với mức độ hoạt động của cá nhân:

$$\text{TDEE} = \text{BMR} \times \text{Hệ số vận động}$$

Hệ số vận động được phân loại theo các mức như ít vận động, vận động nhẹ, vận động trung bình và vận động cao[3]. Chỉ số TDEE đóng vai trò quan trọng trong việc xác định trạng thái cân bằng năng lượng.

### **2.1.3 Nguyên lý cân bằng năng lượng và thay đổi cân nặng**

Sự thay đổi khối lượng cơ thể phụ thuộc vào chênh lệch giữa năng lượng nạp vào và năng lượng tiêu hao. Khi năng lượng nạp vào lớn hơn TDEE, cơ thể sẽ tích lũy năng lượng dư thừa dưới dạng mỡ; ngược lại, khi năng lượng nạp vào thấp hơn TDEE, cơ thể sẽ sử dụng nguồn dự trữ để bù đắp, dẫn đến giảm cân[3].

Các nghiên cứu định lượng chỉ ra rằng trung bình khoảng 7.700 kcal chênh lệch năng lượng tương đương với khoảng 1 kg khối lượng mỡ cơ thể. Tuy nhiên, quá trình thay đổi cân nặng trên thực tế mang tính động và chịu ảnh hưởng bởi nhiều yếu tố sinh lý phức tạp[3]. Nguyên lý cân bằng năng lượng là nền tảng để xây dựng mô hình dự đoán xu hướng thay đổi cân nặng trong hệ thống.

## **2.2 Cơ sở lý thuyết Machine Learning**

### **2.2.1 Bài toán hồi quy trong dự đoán cân nặng**

Bài toán dự đoán tiến trình cơ thể được xem là một bài toán hồi quy (Regression), trong đó mục tiêu là ước lượng giá trị liên tục (ví dụ: cân nặng trong tương lai) dựa trên các biến đầu vào như lượng calo tiêu thụ, TDEE và thời gian[4].

Hồi quy tuyến tính (Linear Regression) là mô hình cơ bản, giả định mối quan hệ tuyến tính giữa biến độc lập và biến phụ thuộc. Mô hình này có ưu điểm đơn giản, dễ diễn giải và phù hợp khi dữ liệu có xu hướng tuyến tính rõ ràng[8].

### **2.2.2 Random Forest Regression**

Random Forest là phương pháp học máy dựa trên kỹ thuật tập hợp (ensemble learning), trong đó nhiều cây quyết định được huấn luyện song song trên các tập dữ liệu con khác nhau. Kết quả dự đoán cuối cùng được tổng hợp từ các cây thành phần[9].

Ưu điểm của Random Forest là giảm hiện tượng quá khớp (overfitting), tăng khả năng khai quát hóa và xử lý tốt các mối quan hệ phi tuyến giữa biến đầu vào và biến mục tiêu[9]. Điều này phù hợp với bài toán dự đoán cân nặng, vốn chịu ảnh hưởng của nhiều yếu tố phức tạp.

### **2.2.3 Đánh giá mô hình dự đoán**

Để đánh giá hiệu quả của các mô hình hồi quy, các chỉ số thường được sử dụng gồm[8]:

- MAE (Mean Absolute Error): đo sai số trung bình tuyệt đối giữa giá trị dự đoán và giá trị thực tế.
- MSE (Mean Squared Error): đo sai số trung bình bình phương, nhấn mạnh các sai số lớn.
- R<sup>2</sup> Score: thể hiện mức độ giải thích phương sai của mô hình đối với dữ liệu.

Việc sử dụng nhiều chỉ số đánh giá giúp đảm bảo quá trình so sánh mô hình được thực hiện khách quan và toàn diện[8],[9].

### **2.3 Công nghệ sử dụng**

- Python – Ngôn ngữ lập trình chính để xây dựng hệ thống và xử lý dữ liệu.
- Scikit-learn – Xây dựng và huấn luyện các mô hình hồi quy (Linear Regression, Random Forest).
- NumPy, Pandas – Xử lý, phân tích và tiền xử lý dữ liệu.
- Flask – Phát triển API backend và xử lý logic hệ thống.
- SQLite – Lưu trữ dữ liệu người dùng và lịch sử dinh dưỡng.
- ReactJS – Xây dựng giao diện người dùng (frontend).
- Chart.js – Trực quan hóa dữ liệu và kết quả dự đoán.

## **3. PHÂN TÍCH YÊU CẦU DỰ ÁN**

### **3.1 Yêu cầu chức năng**

#### **3.1.1 Quản lý tài khoản người dùng**

Hệ thống cần cho phép:

- Đăng ký tài khoản mới
- Đăng nhập và đăng xuất
- Cập nhật thông tin cá nhân
- Lưu trữ dữ liệu riêng biệt cho từng người dùng

Thông tin cá nhân bao gồm: giới tính, tuổi, chiều cao, cân nặng hiện tại và mức độ vận động.

#### **3.2.2 Quản lý thông tin dinh dưỡng**

Hệ thống cần cung cấp chức năng:

- Nhập dữ liệu món ăn theo ngày
- Nhập khối lượng và lượng calo tương ứng

- Tính tổng lượng calo tiêu thụ trong ngày
- Lưu lịch sử ăn uống để theo dõi dài hạn

Dữ liệu này là cơ sở để phân tích trạng thái năng lượng và huấn luyện mô hình dự đoán.

### **3.2.3 Tính toán chỉ số năng lượng**

Hệ thống phải tự động:

- Tính BMR dựa trên thông tin cá nhân
- Tính TDEE theo hệ số vận động
- So sánh tổng calo tiêu thụ với TDEE
- Xác định trạng thái thặng dư hoặc thâm hụt năng lượng

Kết quả tính toán được hiển thị rõ ràng để người dùng dễ theo dõi.

### **3.2.4 Dự đoán tiến trình cân nặng**

Hệ thống cần:

- Thu thập dữ liệu lịch sử cân nặng và năng lượng
- Huấn luyện mô hình hồi quy
- Dự đoán cân nặng sau một khoảng thời gian (ví dụ: 7 ngày, 30 ngày)
- Hiển thị kết quả dự đoán dưới dạng số liệu và biểu đồ

Đây là chức năng cốt lõi tạo nên giá trị khác biệt của hệ thống.

## **3.2 Yêu cầu phi chức năng**

### **3.2.1 Yêu cầu về hiệu năng**

- Thời gian phản hồi cho các thao tác cơ bản không vượt quá 2 giây
- Thời gian dự đoán mô hình ở mức chấp nhận được (dưới 5 giây)

### **3.2.2 Yêu cầu về tính khả dụng**

- Giao diện trực quan, dễ sử dụng
- Bố cục rõ ràng, dễ thao tác
- Hiển thị biểu đồ trực quan

### **3.2.3 Yêu cầu về bảo mật**

- Mật khẩu được mã hóa khi lưu trữ
- Dữ liệu người dùng được phân tách riêng biệt
- Hạn chế truy cập trái phép

### 3.2.4 Yêu cầu về khả năng mở rộng

- Có thể tích hợp thêm mô hình dự đoán nâng cao
- Có thể mở rộng sang ứng dụng di động
- Có thể tích hợp với thiết bị theo dõi sức khỏe

## 4. KẾ HOẠCH THỰC HIỆN DỰ ÁN

### Tuần 1: Phân tích đề tài và khảo sát

- Xác định mục tiêu hệ thống
- Thu thập yêu cầu người dùng
- Nghiên cứu các hệ thống tương tự
- Hoàn thành tài liệu mô tả bài toán

### Tuần 2: Thiết kế hệ thống

- Thiết kế kiến trúc tổng thể
- Thiết kế Use Case Diagram
- Thiết kế Database (SQLite)
- Thiết kế sơ bộ giao diện

### Tuần 3: Xây dựng cơ sở dữ liệu

- Tạo database SQLite
- Thiết kế bảng: User, Meal, Nutrition, BodyMetrics, Prediction
- Viết các hàm CRUD cơ bản

### Tuần 4: Xây dựng chức năng quản lý người dùng

- Đăng ký / Đăng nhập
- Cập nhật thông tin cá nhân
- Nhập thông số cơ thể (cân nặng, chiều cao...)

## **Tuần 5: Xây dựng chức năng quản lý dinh dưỡng**

- Nhập bữa ăn
- Tính tổng calo trong ngày
- Tính TDEE/BMR
- Hiển thị lịch sử dinh dưỡng

## **Tuần 6: Xử lý dữ liệu và tiền xử lý**

- Chuẩn hóa dữ liệu
- Tính thặng dư / thâm hụt năng lượng
- Tạo tập dữ liệu huấn luyện

## **Tuần 7: Xây dựng mô hình dự đoán**

- Huấn luyện Linear Regression
- So sánh với Random Forest
- Đánh giá độ chính xác (MAE, MSE)
- Lưu mô hình

## **Tuần 8: Tích hợp hệ thống**

- Kết nối mô hình vào hệ thống
- Hiển thị kết quả dự đoán
- Kiểm thử chức năng tổng thể

## **Tuần 9: Kiểm thử và tối ưu**

- Test toàn bộ hệ thống
- Sửa lỗi
- Tối ưu giao diện
- Cải thiện tốc độ xử lý

## **Tuần 10: Hoàn thiện báo cáo và chuẩn bị báo cáo**

- Viết báo cáo hoàn chỉnh
- Chuẩn bị slide
- Chạy thử demo

- Chuẩn bị câu hỏi

## TÀI LIỆU THAM KHẢO

- [1] J. A. Harris and F. G. Benedict, “A biometric study of basal metabolism in man,” Proceedings of the National Academy of Sciences of the United States of America, vol. 4, no. 12, pp. 370–373, 1918, doi: 10.1073/pnas.4.12.370.
- [2] M. D. Mifflin, S. T. St Jeor, L. A. Hill, B. J. Scott, S. A. Daugherty, and Y. O. Koh, “A new predictive equation for resting energy expenditure in healthy individuals,” The American Journal of Clinical Nutrition, vol. 51, no. 2, pp. 241–247, Feb. 1990, doi: 10.1093/ajcn/51.2.241.
- [3] World Health Organization, Human energy requirements: Report of a joint FAO/WHO/UNU expert consultation, FAO Food and Nutrition Technical Report Series No. 1, Rome, Italy: FAO, 2004.
- [4] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
- [5] J. D. Hunter, “Matplotlib: A 2D graphics environment,” Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, May–Jun. 2007, doi: 10.1109/MCSE.2007.55.
- [6] Python Software Foundation, “Python language reference, version 3.x,” 2023. [Online]. Available: <https://www.python.org/>
- [7] SQLite Consortium, “SQLite documentation,” 2023. [Online]. Available: <https://www.sqlite.org/docs.html>
- [8] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2012.
- [9] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.