

Mô tả tổng quan hệ thống

Hệ thống là gì?

Social Media Well-being Predictor là một ứng dụng web ứng dụng **Machine Learning** nhằm:

- Phân tích hành vi sử dụng mạng xã hội (Instagram) và các yếu tố lối sống của người dùng.
- Dự đoán **mức độ căng thẳng (Stress)**, **mức độ hạnh phúc (Happiness)**.
- Phân loại người dùng vào các **nhóm hành vi (Persona)**.

Hệ thống **không sử dụng dữ liệu người dùng thật**. Toàn bộ mô hình được huấn luyện từ **dữ liệu synthetic** (giả lập) nhằm đảm bảo an toàn, riêng tư và phù hợp với mục đích học thuật.

Mục tiêu của hệ thống




Hệ thống được xây dựng để đạt các mục tiêu sau:

- Áp dụng Machine Learning vào một bài toán có ý nghĩa thực tế liên quan đến hành vi con người.
- Minh họa mối liên hệ giữa:
- Thời gian và cách thức sử dụng mạng xã hội
- Lối sống (ngủ, vận động, công việc)
- Trạng thái tinh thần (stress, happiness)
- Cung cấp một **demo web trực quan, dễ sử dụng**.
- Thể hiện quy trình **end-to-end**:

Train model → Đánh giá → Deploy → Tích hợp Backend & Frontend

Đối tượng sử dụng

Hệ thống hướng tới các đối tượng:

-  **Sinh viên / người học AI - Data Science** (mục đích học tập)
-  **Giảng viên** (đánh giá khả năng áp dụng ML)
-  **Người dùng demo** (nhập dữ liệu giả lập để xem kết quả)

 **Lưu ý:** Hệ thống **không dùng cho chẩn đoán y khoa**, chỉ mang tính phân tích và gợi ý học thuật.

Các chức năng chính

1) Nhập thông tin người dùng (Frontend)

Người dùng nhập các thông tin mô phỏng hành vi và lối sống, ví dụ:

- Tuổi, giới tính, quốc gia
- Thời gian ngủ, vận động

- Thời gian dùng Instagram
- Tần suất xem reels/stories, mức độ tương tác (like/comment)

Frontend **tự động sinh form** dựa trên danh sách feature lấy từ Backend (/features), nhờ đó:

- Không cần hard-code form
 - Dễ mở rộng khi thêm feature mới
-

2) Dự đoán Stress (Regression)

Hệ thống sử dụng mô hình **RandomForestRegressor** để:

- Phân tích các yếu tố hành vi và lối sống
- Dự đoán **Stress score** (giá trị số)

Ý nghĩa:

- Minh họa mức độ căng thẳng tương đối
 - Thấy được ảnh hưởng của việc sử dụng mạng xã hội quá mức
-

3) Dự đoán Happiness (Regression)

Song song với Stress, hệ thống dự đoán:

- **Happiness score** (mức độ hạnh phúc tự báo cáo)

Ý nghĩa:

- So sánh mối quan hệ giữa stress và happiness
 - Thể hiện rằng stress không nhất thiết luôn tỷ lệ nghịch tuyệt đối với happiness
-

4) Phân loại Persona (Clustering)

Hệ thống sử dụng **KMeans Clustering** để phân nhóm người dùng theo hành vi Instagram.

Ví dụ các persona:

- **Doom-Scroller**: dùng nhiều, xem reels liên tục, ít tương tác chất lượng
- **Engaged Socializer**: tương tác cao, like/comment nhiều
- **Light User**: dùng ít, thời gian ngắn
- **Casual Browser**: dùng vừa phải, không quá lệ thuộc

Persona giúp:

- Giải thích kết quả ML theo cách dễ hiểu với người không chuyên
 - Là nền tảng để mở rộng phần gợi ý hành vi trong tương lai
-

5) Trả kết quả & hiển thị trực quan

Sau khi dự đoán, Backend trả về cho Frontend:

- Stress score
- Happiness score
- Persona name

Frontend hiển thị:

- Kết quả dạng số
- Nhóm persona tương ứng

Có thể mở rộng trong tương lai:

- Màu sắc cảnh báo
- Gợi ý hành vi
- Dashboard theo thời gian

Kiến trúc hệ thống (khái niệm)

Hệ thống gồm 3 tầng:

◆ Frontend (Presentation Layer)

- Giao diện web
- Thu thập input
- Hiển thị kết quả

◆ Backend (Application Layer)

- Nhận request từ Frontend
- Validate dữ liệu theo `features.json`
- Gọi mô hình ML để predict
- Trả response JSON

◆ Machine Learning Layer

- Các pipeline ML đã train sẵn
- Không train lại khi chạy hệ thống
- Chỉ load và predict

Phạm vi & giới hạn

Hệ thống làm được

- Phân tích dữ liệu hành vi
- Dự đoán chỉ số mang tính tương đối
- Demo quy trình ML end-to-end

Hệ thống không làm

- Không chẩn đoán tâm lý/y khoa
 - Không thu thập dữ liệu thật
 - Không can thiệp trực tiếp vào hành vi người dùng
-

Ý nghĩa học thuật & thực tiễn

Hệ thống thể hiện khả năng:

- Feature engineering
- Regression & Clustering
- Model deployment
- Backend-Frontend integration

Hướng mở rộng:

- Hệ thống gợi ý hành vi lành mạnh
- Dashboard phân tích xu hướng
- A/B testing giả lập