

Họ và tên:

MSSV:

(Nếu là HS ghi tên trường THPT)

Tên mentor:



**ML IoT**  
HCMUT EE MACHINE LEARNING & IOT LAB

HCMUT EE  
ML & IoT Lab

## BÀI KIỂM TRA HỌC PHẦN 2 MACHINE LEARNING

Ngày kiểm tra: 9/8/2025

Người ra đề: TA Trần Huy

Thời gian: 25 phút

Hình thức: Tự luận

### CHÚ Ý QUAN TRỌNG

- Không bắt buộc viết code chi tiết, tập trung vào ý tưởng và phương pháp
- Giải thích rõ ràng logic đằng sau mỗi quyết định
- Có thể sử dụng bất kỳ thuật toán/kỹ thuật nào phù hợp
- Chú ý đến tính thực tế và khả năng triển khai của giải pháp

### ĐỀ BÀI (ĐỀ GỒM CÓ 3 TRANG)

Một công ty công nghệ muốn phát triển một hệ thống AI để **tối ưu hóa quản lý năng lượng và an ninh trong nhà ở thông minh**. Hệ thống này nhằm:

- Dự đoán mức tiêu thụ năng lượng để tối ưu hóa sử dụng thiết bị
- Phát hiện các hành vi bất thường (ví dụ: xâm nhập, hỏng thiết bị) để cảnh báo kịp thời
- Hỗ trợ người dùng lập kế hoạch sử dụng năng lượng hiệu quả và tăng cường an ninh

**DỮ LIỆU CÓ SẴN:** Tập dữ liệu `smart_home_data.csv` gồm thông tin của 30,000 ngôi nhà, thu thập trong 5 năm, chia thành:

• **Bảng 1: Thông tin cơ bản về ngôi nhà**

Trường	Mô tả	Kiểu dữ liệu	Ví dụ
home_id	Mã định danh ngôi nhà	Chuỗi	H001, H002
location	Vị trí địa lý	Categorical	Urban, Suburban, Rural
home_size	Diện tích (m <sup>2</sup> )	Số thực	120.5, 250.8
num_residents	Số người cư trú	Số nguyên	2, 4, 6
home_age	Tuổi của ngôi nhà (năm)	Số nguyên	5, 20, 50

• **Bảng 2: Dữ liệu cảm biến năng lượng**

Trường	Mô tả	Kiểu dữ liệu	Ví dụ
power_usage	Tiêu thụ điện (kWh)	Số thực	10.5, 25.3
appliance_count	Số thiết bị điện	Số nguyên	5, 12, 20
hvac_usage	Tiêu thụ HVAC (kWh)	Số thực	5.2, 15.8
lighting_usage	Tiêu thụ ánh sáng (kWh)	Số thực	2.1, 8.4
time_of_day	Thời điểm trong ngày	Categorical	Morning, Afternoon, Night

• **Bảng 3: Dữ liệu an ninh**

Trường	Mô tả	Kiểu dữ liệu	Ví dụ
motion_sensor	Phát hiện chuyển động	Binary	0, 1
door_status	Trạng thái cửa	Binary	Open, Closed
window_status	Trạng thái cửa sổ	Binary	Open, Closed
camera_activity	Hoạt động camera	Categorical	None, Motion, Object
alarm_trigger	Kích hoạt cảnh báo	Binary	0, 1

• **Bảng 4: Dữ liệu môi trường**

Trường	Mô tả	Kiểu dữ liệu	Ví dụ
temperature	Nhiệt độ trong nhà (°C)	Số thực	22.5, 28.3
humidity	Độ ẩm (%)	Số thực	45.2, 70.1
air_quality	Chất lượng không khí	Số thực	50.0, 120.5
weather_condition	Thời tiết ngoài trời	Categorical	Sunny, Rainy, Cloudy

• **Bảng 5: Kết quả và chỉ số mục tiêu**

Trường	Mô tả	Kiểu dữ liệu	Ví dụ
energy_efficiency	Hiệu quả năng lượng	Số thực	0.75, 0.92
anomaly_detected	Phát hiện bất thường	Binary	0, 1
follow_up_days	Thời gian theo dõi (ngày)	Số nguyên	30, 90, 180

## MẪU DỮ LIỆU (Phần 1): Thông tin cơ bản và tiêu thụ điện

ID	Home	Location	Size	Residents	HomeAge	Power	Appliances	HVAC	Lighting
H001		Urban	150.2	4	10	12.5	8	6.2	3.1
H002		Rural	200.5	6	25	18.3	15	10.5	5.2
H003		Suburban	180.0	3	15	15.7	10	8.4	4.0
H004		Urban	120.8	2	5	10.2	6	4.8	2.5
H005		Rural	250.3	5	30	22.1	18	12.3	6.8

## MẪU DỮ LIỆU (Phần 2): Cảm biến, điều kiện môi trường và bất thường

ID	Home	Time	Motion	Door	Window	Camera	Alarm	Temp	Efficiency	Anomaly
H001		Morning	1	Closed	Closed	Motion	0	23.5	0.85	0
H002		Night	0	Closed	Open	None	0	20.1	0.78	0
H003		Afternoon	1	Open	Closed	Object	1	25.3	0.82	1
H004		Morning	0	Closed	Closed	None	0	22.0	0.90	0
H005		Night	1	Open	Open	Motion	1	19.8	0.70	1

Chú thích: Power=Tiêu thụ điện, Appliances=Số thiết bị điện, HVAC=Tiêu thụ HVAC, Lighting=Tiêu thụ ánh sáng, Time=Thời điểm trong ngày, Motion=Phát hiện chuyển động, Door=Trạng thái cửa, Window=Trạng thái cửa sổ, Camera=Hoạt động camera, Alarm=Kích hoạt cảnh báo, Temp=Nhiệt độ, Efficiency=Hiệu quả năng lượng, Anomaly=Phát hiện bất thường

## CÂU HỎI

### Câu 1: Phân tích bài toán (4 điểm)

- Xác định loại bài toán Machine Learning và giải thích tại sao lựa chọn này phù hợp
- Phân tích biến target và đề xuất cách xử lý (nếu cần)
- Đánh giá tính khả thi và thách thức khi triển khai trong môi trường nhà ở thực tế

### Câu 2: Khám phá và tiền xử lý dữ liệu (6 điểm)

- Mô tả các bước khám phá dữ liệu (EDA) mà bạn sẽ thực hiện
- Xác định và đề xuất cách xử lý dữ liệu

### Câu 3: Lựa chọn và đánh giá mô hình (6 điểm)

- Đề xuất ít nhất 1 thuật toán Machine Learning (có thể là mô hình truyền thống, hoặc kết hợp các kỹ thuật), giải thích lý do chọn từng thuật toán
- Đề xuất các metrics đánh giá phù hợp với bối cảnh nhà ở thông minh, giải thích tại sao chọn các metrics này

### Câu 4: Trình bày cách giải thích kết quả dự đoán cho người dùng và kỹ thuật viên (4 điểm)

—HẾT ĐỀ—  
Chúc các bạn làm bài tốt!