ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

# ỨNG DỤNG HỌC MÁY TRONG PHÁT HIỆN MÃ ĐỘC CHO HỆ THỐNG CNTT BỆNH VIỆN

Học viên: Võ Anh Quân – 250101055
Ngành: Khoa học máy tính
Lớp: CS2205.SEP2025
Github:
YouTube:

# Vấn đề an toàn thông tin

### Tấn công gia tăng

Mã độc nhắm vào lĩnh vực y tế ngày càng nhiều.

### Các loại mã độc

Ransomware, trojan, spyware đe dọa bệnh viện.

### Nguy cơ gián đoạn

Ảnh hưởng nghiêm trọng đến khám chữa bệnh.

# Nhu cầu thực tiễn

### Phát hiện sớm

Tìm ra mã độc chưa biết một cách nhanh chóng.

### Ổn định hệ thống

Đảm bảo hoạt động liên tục, không gián đoạn.

### Phù hợp hạ tầng

Giải pháp triển khai được ở bệnh viện tuyến tỉnh.

# Mục tiêu nghiên cứu

1 **Nghiên cứu Malware**

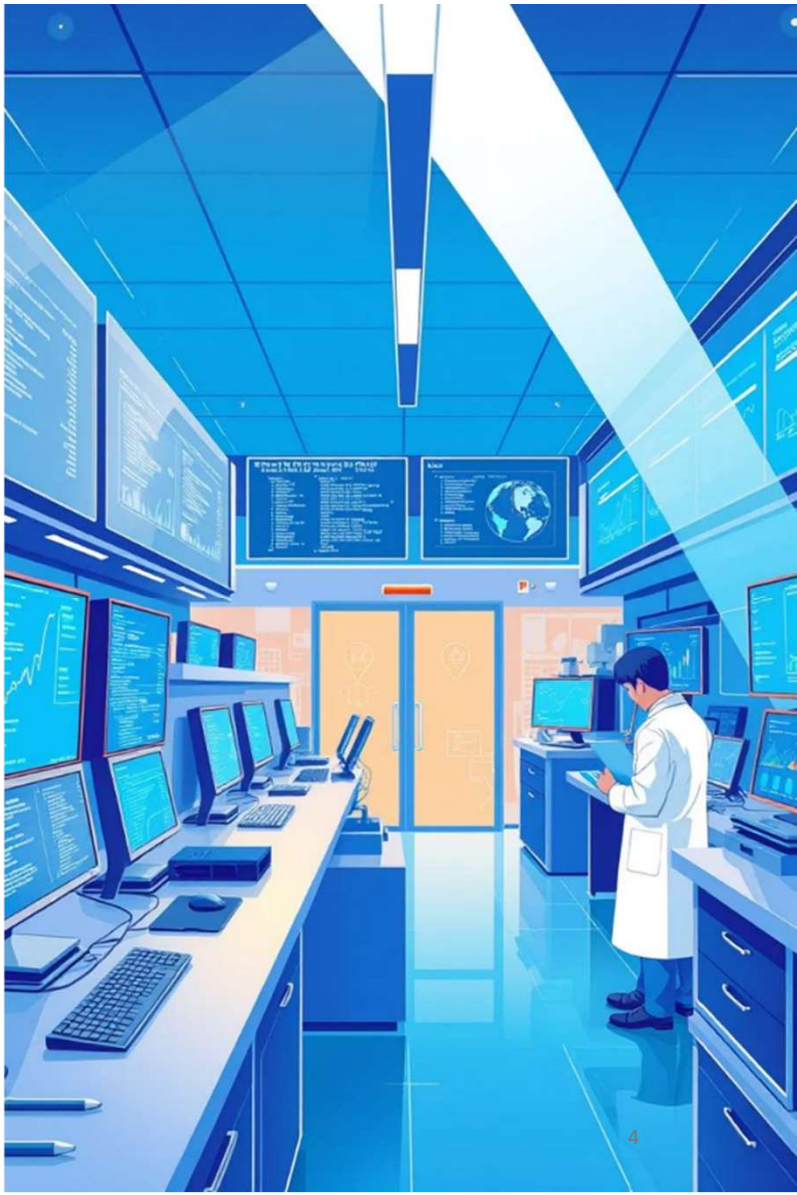
Tìm hiểu sâu về mã độc và kỹ thuật phân tích.

2 **Ứng dụng Học máy**

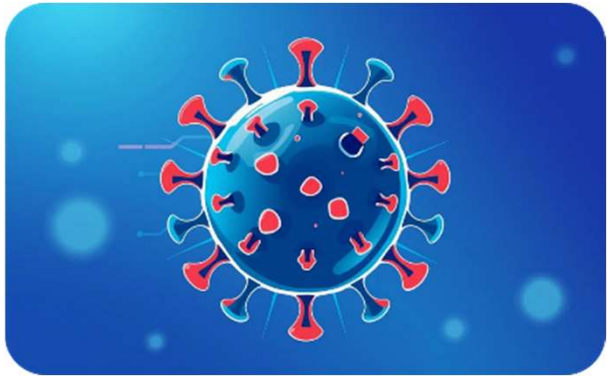
Phát triển giải pháp phát hiện mã độc bằng ML.

3 **Khả năng triển khai**

Hướng tới giải pháp có thể áp dụng thực tế.



# Tổng quan về mã độc



## Virus, Worm, Trojan

Các loại mã độc truyền thống với đặc điểm lây lan khác nhau.

## Ransomware, Spyware

Mã độc tống tiền và phần mềm gián điệp.

## Đặc điểm & Hành vi

Hiểu rõ cách chúng hoạt động để phòng chống hiệu quả.

# Các phương pháp phân tích mã độc

1

## Phân tích tĩnh

Kiểm tra mã mà không thực thi.

2

## Phân tích động

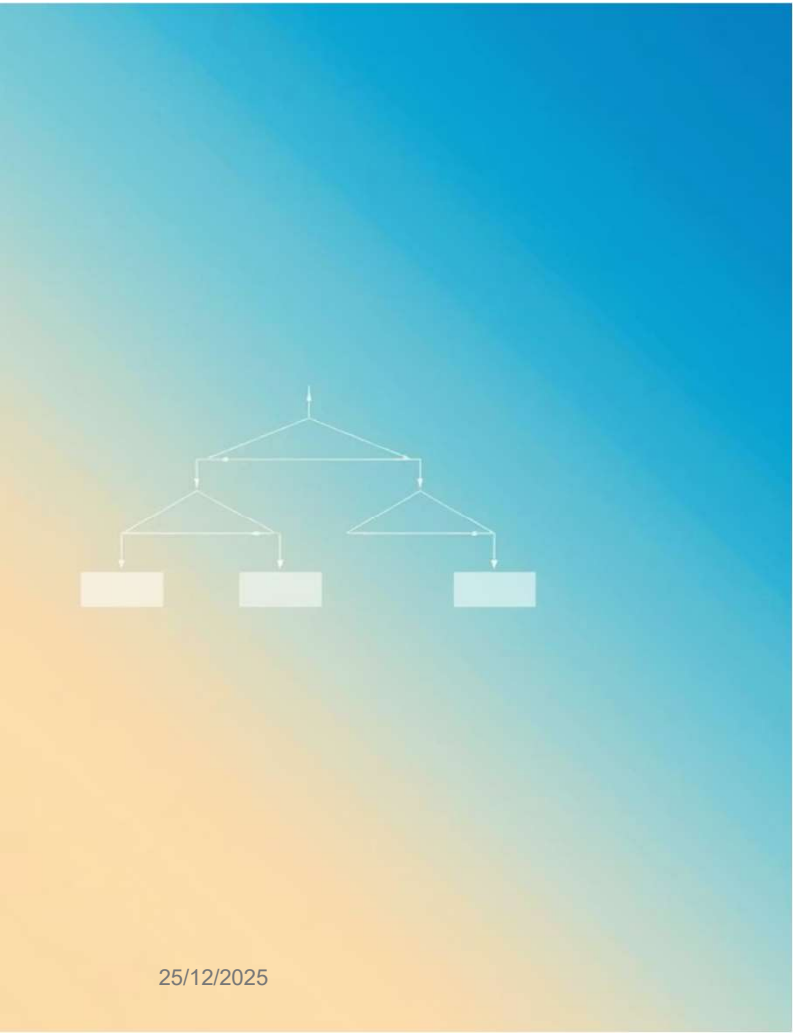
Quan sát hành vi khi mã chạy trong môi trường an toàn.

3

## Phân tích lai

Kết hợp cả hai phương pháp trên.

👉 Luận văn tập trung vào **phân tích tĩnh**.



# Thuật toán sử dụng

1

## Cây quyết định (Decision Tree)

Phân tích dữ liệu để đưa ra các quyết định phân loại.

2

## Máy vector hỗ trợ (SVM)

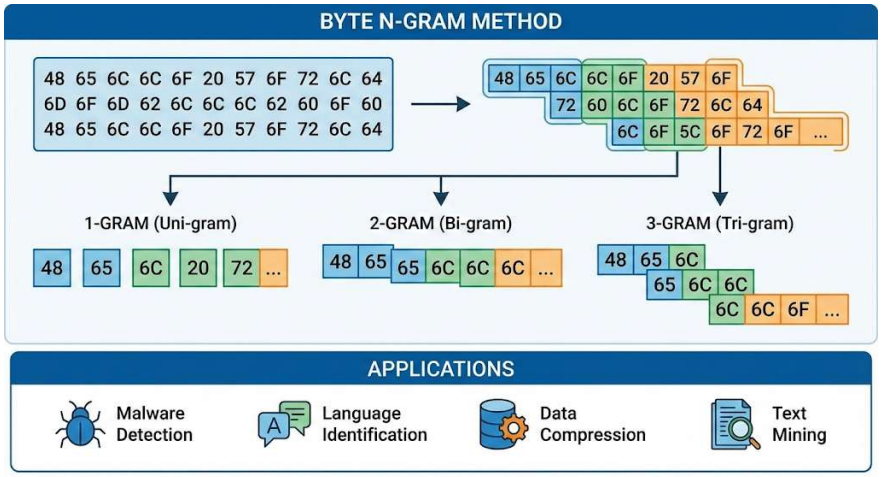
Tìm siêu phẳng tối ưu để phân tách các lớp dữ liệu.

3

## Học máy có giám sát

Sử dụng dữ liệu đã được gán nhãn để huấn luyện mô hình.

# Trích chọn đặc trưng




N

## Kỹ thuật n-gram byte

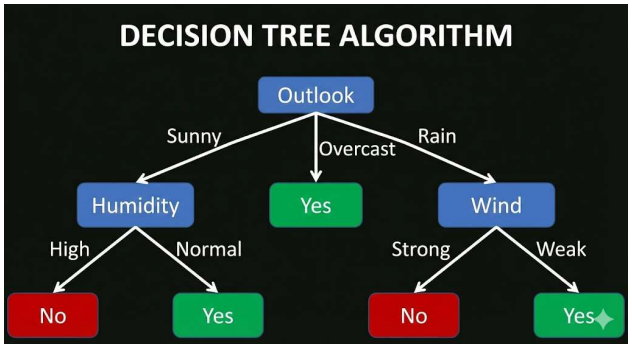
Phân tích các chuỗi byte liên tiếp trong mã độc.

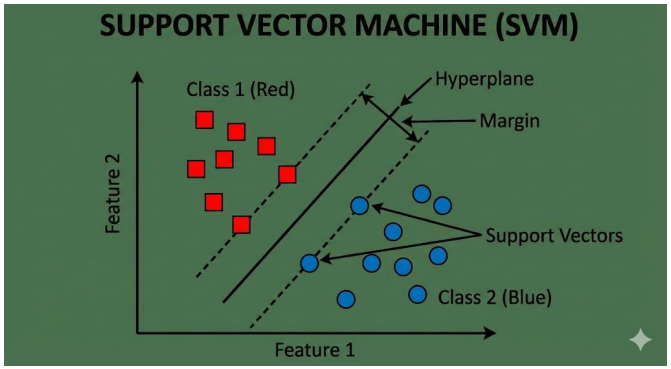
## Tính tần suất xuất hiện

Xác định mức độ phổ biến của các n-gram.

## Giảm chiều dữ liệu

Tối ưu hóa dữ liệu để tăng hiệu quả xử lý.




# Xây dựng mô hình

01

## Huấn luyện mô hình

Sử dụng dữ liệu mẫu để đào tạo thuật toán.

02

## Kiểm thử

Đánh giá hiệu suất trên tập dữ liệu mới, chưa từng thấy.

03

## Đánh giá độ chính xác

Đo lường khả năng nhận diện mã độc chính xác của mô hình.

# Lợi ích mang lại

### Nâng cao an toàn thông tin y tế

Bảo vệ dữ liệu bệnh nhân và hệ thống y tế khỏi các cuộc tấn công.

### Giảm phụ thuộc vào chữ ký mã độc

Phát hiện các mối đe dọa mới mà không cần cập nhật chữ ký liên tục.

### Phù hợp môi trường vận hành thực tế

Giải pháp linh hoạt, dễ dàng tích hợp vào hạ tầng hiện có.

# Tài liệu tham khảo

1. Shabtai A. et al., *Detecting Unknown Malicious Code Using Machine Learning*, 2012.
2. Moskovitch R., Elovici Y., *Static Malware Detection Survey*, 2009.
3. Hsu C.W., Chang C.C., Lin C.J., *A Practical Guide to Support Vector Classification*, 2016.
4. Oktavianto D., Muhardianto I., *Cuckoo Malware Analysis,* Packt Publishing*,* 2013.