TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**PHÁT HIỆN XÂM NHẬP MẠNG SỬ DỤNG CÁC MÔ HÌNH HỌC MÁY**

**CHUYÊN ĐỀ NGHIÊN CỨU 1**

**KHOA HỌC MÁY TÍNH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**PHÁT HIỆN XÂM NHẬP MẠNG SỬ DỤNG CÁC MÔ HÌNH HỌC MÁY**

**CHUYÊN ĐỀ NGHIÊN CỨU 1**

**KHOA HỌC MÁY TÍNH**

Người hướng dẫn

**TS. Huỳnh Ngọc Tú**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**LỜI CẢM ƠN**

Chúng em xin chân thành cảm ơn Cô Huỳnh Ngọc Tú đã tận tình giúp đỡ chúng em hoàn thành Chuyên đề nghiên cứu này. Để có thể hoàn thành chuyên đề này, chúng em đã nỗ lực thực hiện và đồng thời cũng nhận được nhiều sự giúp đỡ từ cô và mọi người. Chúng em cũng xin gửi lời cảm ơn đến Trường đại học Tôn Đức Thắng đã tạo điều kiện tốt nhất để chúng em có thể học tập, trao đổi và nâng cao kiến thức của mình. Mặc dù chúng em đã rất cố gắng hoàn thành trong phạm vi và khả năng cho phép nhưng chắc chắn sẽ không tránh khỏi những thiếu sót. Em kính mong nhận được sự cảm thông và tận tình chỉ bảo của cô.

Chúng em xin chân thành cảm ơn!

*TP. Hồ Chí Minh, ngày 17 tháng 3 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Chúng em xin cam đoan đây là công trình nghiên cứu của riêng chúng tôi và được sự hướng dẫn khoa học của TS. Huỳnh Ngọc Tú. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Chuyên đề còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào chúng em xin hoàn toàn chịu trách nhiệm về nội dung Chuyên đề của mình**. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do chúng em gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 17 tháng 3 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

**PHÁT HIỆN XÂM NHẬP MẠNG**

**TÓM TẮT**

Chuyên đề nghiên cứu chúng em làm dưới đây liên quan đến vấn đề an ninh mạng, một vấn đề đang rất phổ biến trong thời gian gần đây. Trong bài nghiên cứu này, chúng ta sẽ được tìm hiểu về an ninh mạng, các thành phần của an ninh mạng, cách mà một cuộc tấn công sẽ diễn ra, các phương pháp để phòng chống một cuộc tấn công đó. Ngoài ra chúng ta còn tìm hiểu về mô hình học máy và học sâu, cách áp dụng các mô hình đó vào để phòng chống một cuộc tấn công an ninh mạng. Cuối cùng là triển khai thực hiện các phương pháp và mô hình học máy đó vào bộ dữ liệu để đưa ra kết quả, so sánh và đánh giá mô hình.

**NETWORK INTRUSION DETECTION**

**ABSTRACT**

The research topic we're working on is related to cybersecurity, which has become increasingly prevalent in recent times. In this research, we will delve into cybersecurity, its components, the anatomy of a cyber attack, and methods to mitigate such attacks. Additionally, we will explore machine learning and deep learning models, and how they can be applied to prevent cyber attacks. Finally, we'll implement these methods and machine learning models on datasets to produce results, compare, and evaluate the models.

**MỤC LỤC**

[**DANH MỤC HÌNH VẼ 7**](#_2u6wntf)

[**DANH MỤC CÁC CHỮ VIẾT TẮT 8**](#_19c6y18)

[**CHƯƠNG 1. MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI 1**](#_3tbugp1)

[1.1 Lý do chọn đề tài 1](#_28h4qwu)

[1.2 Mục tiêu thực hiện đề tài 2](#_nmf14n)

[**CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 3**](#_37m2jsg)

[2.1 Các mục tiêu chính của an ninh mạng 3](#_1mrcu09)

[2.2 Các giải pháp an ninh mạng 4](#_46r0co2)

[*2.2.1 Tường lửa (Firewalls): 4*](#_2lwamvv)

[2.3 Các phương thức phát hiện xâm nhập 6](#_1egqt2p)

[*2.3.1 IDS theo nền tảng được giám sát 6*](#_2dlolyb)

[*2.3.2 IDS dựa trên kỹ thuật được sử dụng trong quá trình phát hiện 9*](#_4bvk7pj)

[2.4 Cơ sở dữ liệu trong Phát hiện Xâm nhập 10](#_2r0uhxc)

[2.5 Mô hình học máy và Học sâu trong IDS 10](#_1664s55)

[*2.5.1 Học Máy trong IDS (Machine Learning in IDS) 10*](#_3q5sasy)

[*2.5.2 Học Sâu trong IDS (Deep Learning in IDS) 11*](#_23m2g9bouobf)

[2.6 Tập dữ liệu và đánh giá mô hình IDS 11](#_25b2l0r)

[2.7 Các thách thức trong đánh giá mô hình IDS 11](#_kgcv8k)

[**CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT 13**](#_34g0dwd)

[3.1 Dữ liệu sử dụng: 13](#_1jlao46)

[3.2 Tiền xử lý Dữ liệu: 13](#_43ky6rz)

[*3.2.1 Xử lý dữ liệu rỗng (null) và dữ liệu dạng chuỗi: 14*](#_2iq8gzs)

[*3.2.2 Chọn lọc đặc trưng đơn giản (Simple Feature Selection): 14*](#_xvir7l)

[3.3 Lựa chọn đặc trưng dữ liệu: 14](#_3hv69ve)

[3.4 Chọn và Huấn luyện Mô hình 15](#_1x0gk37)

[*3.4.1 Lựa Chọn Đặc Trưng với RFE 16*](#_4h042r0)

[*3.4.2 Huấn luyện Với Nhiều Mô hình với Tham Số Mặc Định 16*](#_2w5ecyt)

[3.5 Tuning Model: Điều Chỉnh Siêu Tham Số của Mô hình 22](#_3mzq4wv)

[3.6 Đánh giá mô hình 23](#_2250f4o)

[*3.6.1 Accuracy: Độ Chính xác 23*](#_haapch)

[*3.6.2 Precision: Độ Chính xác tích cực 23*](#_319y80a)

[*3.6.3 Recall: Độ Nhạy 23*](#_1gf8i83)

[*3.6.4 F1 Score: Điểm số F1 24*](#_40ew0vw)

[**CHƯƠNG 4. THỰC NGHIỆM 25**](#_2fk6b3p)

[4.1 Kết quả thực nghiệm 25](#_upglbi)

[*4.1.1 Trước khi Tuning Model 25*](#_3ep43zb)

[*4.1.2 Sau khi tiến hành tinh chỉnh tham số: 26*](#_1tuee74)

[**CHƯƠNG 5. KẾT LUẬN 27**](#_4du1wux)

[5.1 Kết luận 27](#_2szc72q)

[5.2 Hướng phát triển 28](#_184mhaj)

[*5.2.1 Tích hợp thêm yếu tố và nguồn dữ liệu đa dạng: 28*](#_3s49zyc)

[*5.2.2 Phát triển giao diện người dùng trực quan: 28*](#_279ka65)

[*5.2.3 Tối ưu hóa và cập nhật mô hình dựa trên phản hồi thực tế: 28*](#_meukdy)

[*5.2.4 Khám phá các kỹ thuật học máy mới và tiên tiến: 29*](#_36ei31r)

[**TÀI LIỆU THAM KHẢO 30**](#_1ljsd9k)

# **DANH MỤC HÌNH VẼ**

[HÌNH 2.1 Tường lửa (Firewall) trong hệ thống mạng 5](#_111kx3o)

[HÌNH 2.2 Tường lửa ở cổng Gateway 5](#_3l18frh)

[HÌNH 2.3 Tường lửa ở giữa nội bộ mạng và bộ định tuyến (router) 6](#_4k668n3)

[HÌNH 2.4 Các phương pháp phát hiện xâm nhập mạng 6](#_3ygebqi)

[HÌNH 2.5 Phát hiện xâm nhập dựa trên mạng (Network-based IDS) 7](#_sqyw64)

[HÌNH 2.6 Phát hiện xâm nhập dựa trên máy chủ (Host-based IDS) 8](#_3cqmetx)

[HÌNH 2.7 Phát hiện xâm nhập kết hợp (Hybrid IDS) 9](#_1rvwp1q)

[HÌNH 3.1 So sánh Linear Regression và Logistic Regression 16](#_3vac5uf)

[HÌNH 3.2 Thuật toán kNN 17](#_pkwqa1)

[HÌNH 3.3 Thuật toán Decision Trees 18](#_39kk8xu)

[HÌNH 3.4 Thuật toán Random Forest 19](#_1opuj5n)

[HÌNH 3.5 Thuật toán Gradient Boosting 20](#_48pi1tg)

[HÌNH 3.6 Thuật toán Support Vector Machine 21](#_2nusc19)

[HÌNH 3.7 Thuật toán Multilayer Perceptron 22](#_1302m92)

# **DANH MỤC CÁC CHỮ VIẾT TẮT**

| IDS  NIDS  TCP  IP  DMZ  LAN  DOS  DDOS  HTTP  HTTPS  SMB  SSH | Intrusion Detection System  Network Intrusion Detection System  Transmission Control Protocol  Internet Protocol  Demilitarized Zone  Local Area Network  Denial-of-service  Distributed Denial-of-service  Hypertext Transfer Protocol  Hypertext Transfer Protocol Security  Server Message Block  Secure Shell |
| --- | --- |

# **MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI**

## **Lý do chọn đề tài**

Trong thời đại kỷ nguyên kỹ thuật số mà chúng ta đang sống, việc đảm bảo an ninh mạng là một trong những ưu tiên hàng đầu không chỉ cho các doanh nghiệp mà cũng cho các tổ chức chính phủ và cá nhân trên khắp thế giới. Sự gia tăng về số lượng và tính chất nghiêm trọng của các cuộc tấn công mạng đòi hỏi phải có những phương pháp tiên tiến và tự động hóa để phát hiện và ngăn chặn chúng. Trong bối cảnh đó, việc phát hiện xâm nhập mạng sớm và chính xác trở thành một yếu tố quan trọng, góp phần vào việc bảo vệ dữ liệu quan trọng và duy trì hoạt động bình thường của hệ thống thông tin.

Các kỹ thuật Học Máy (Machine learning) [1] đã được chứng minh là công cụ mạnh mẽ trong việc giải quyết nhiều vấn đề phức tạp, từ phân tích dữ liệu lớn đến nhận dạng mẫu và dự đoán xu hướng. Trong lĩnh vực an ninh mạng, việc ứng dụng các kỹ thuật học máy vào phát hiện xâm nhập mạng [2] hứa hẹn sẽ mang lại khả năng nhận biết sớm các dấu hiệu của cuộc tấn công, từ đó giúp giảm thiểu thiệt hại và tăng cường độ tin cậy và an toàn của hệ thống thông tin.

Với tiềm năng lớn mà học máy mang lại, đề tài Phát hiện xâm nhập mạng sử dụng các kỹ thuật Học Máy (Network Intrusion Detection using Machine Learning) được chọn làm đề tài Chuyên đề nghiên cứu, nhằm mục tiêu nghiên cứu và ứng dụng các kỹ thuật học máy tiên tiến vào việc phát hiện và phòng chống xâm nhập mạng. Đề tài này không chỉ có ý nghĩa trong việc cải thiện an ninh mạng và bảo vệ thông tin mà còn đóng góp vào việc phát triển các phương pháp phòng ngừa hiệu quả, giảm thiểu rủi ro và tối ưu hóa quy trình quản lý an ninh thông tin, qua đó hỗ trợ cho việc ra quyết định và chiến lược an ninh mạng tốt hơn.

## **Mục tiêu thực hiện đề tài**

Mục tiêu của đề tài này bao gồm:

- Nắm vững cơ sở lý thuyết của các kỹ thuật học máy áp dụng trong phát hiện xâm nhập mạng, bao gồm nguyên lý hoạt động, phương pháp phân loại và nhận dạng hành vi mạng bất thường.

- Nghiên cứu và so sánh mô hình NIDS (Network Intrusion Detection System) hiệu quả sử dụng kỹ thuật học máy, nhằm đề xuất mô hình tối ưu có khả năng phân biệt giữa lưu lượng mạng bình thường và lưu lượng mạng có chứa dấu hiệu xâm nhập, bao gồm cả các cuộc tấn công DoS, DDoS [3] và các hình thức tấn công khác.

- Phân tích và đánh giá hiệu suất của mô hình NIDS được phát triển thông qua thử nghiệm trên bộ dữ liệu thực tế, để xác định khả năng phát hiện xâm nhập và độ chính xác trong các tình huống khác nhau.

- Khám phá và phân tích các ứng dụng thực tiễn của kỹ thuật học máy trong lĩnh vực an ninh mạng, như phân loại loại tấn công, phân tích hành vi người dùng và tối ưu hóa việc phát hiện xâm nhập.

Mục tiêu chính của đề tài này là không chỉ tìm ra một mô hình NIDS hiệu quả mà còn góp phần vào việc nâng cao nhận thức và kiến thức về an toàn thông tin. Qua đó, đề tài nhấn mạnh sự cần thiết của việc áp dụng các kỹ thuật học máy tiên tiến vào việc bảo vệ hệ thống thông tin trước những mối đe dọa an ninh mạng ngày càng tinh vi, vì thế (Network Intrusion Detection System - NIDS) hiệu quả là vô cùng quan trọng.

# **CƠ SỞ LÝ THUYẾT**

**An ninh mạng và hệ thống phát hiện xâm nhập (IDS)**

Trong bối cảnh hiện nay, an ninh mạng đã trở thành một phần không thể tách rời trong việc đảm bảo an toàn thông tin và hệ thống truyền thông của các tổ chức, doanh nghiệp. Nói cách khác, an ninh mạng không chỉ là hàng rào bảo vệ dữ liệu khỏi các mối đe dọa trực tuyến mà còn là nền tảng giúp duy trì sự toàn vẹn và khả năng sẵn có của hệ thống thông tin trong thời đại số. Để thực hiện nhiệm vụ này, các giải pháp an ninh mạng như phần mềm chống virus, tường lửa và các phương pháp kiểm soát truy cập đã được triển khai rộng rãi. Tuy nhiên, sự xuất hiện của Hệ thống Phát hiện Xâm nhập (Intrusion Detection System - IDS) [4] đánh dấu một bước tiến quan trọng, bổ sung cho những phương pháp truyền thống bằng cách sử dụng công nghệ tiên tiến để phát hiện sớm các dấu hiệu của các cuộc tấn công mạng, từ đó cảnh báo kịp thời và triển khai các biện pháp phòng ngừa.

## **Các mục tiêu chính của an ninh mạng**

Có năm thuộc tính bảo mật cốt lõi được sử dụng để đảm bảo tính bảo mật và độ tin cậy của hệ thống thông tin thường được viết tắt là: ‘CIA’ hoặc ‘CIAAN’ [5]:

* Bảo mật thông tin (Confidentiality): Thành phần này quyết định thông tin chỉ nên nằm trong tay những người hoặc hệ thống được ủy quyền. Trong thế giới kỹ thuật số, đây là cơ sở để xây dựng niềm tin, bảo vệ quyền riêng tư và giữ bí mật thông tin doanh nghiệp khỏi những ánh mắt tò mò.
* Tính toàn vẹn thông điệp (Integrity): Đảm bảo thông tin không bị thay đổi một cách trái phép hay không mong muốn, từ đó thông tin được duy trì một cách chính xác và đầy đủ. Tính toàn vẹn là yếu tố không thể thiếu để duy trì chất lượng dữ liệu và đảm bảo quyết định được đưa ra dựa trên thông tin chính xác.
* Xác thực (Authentication): Xác minh danh tính của người dùng hoặc thiết bị truy cập thông tin là yếu tố cơ bản để ngăn chặn sự truy nhập trái phép và giả mạo danh tính, từ đó duy trì một môi trường mạng an toàn và đáng tin cậy.
* Không phủ nhận (Non-repudiation): Đây là biện pháp ngăn chặn việc chối bỏ trách nhiệm về một hành động hoặc giao dịch đã thực hiện. Thông qua kỹ thuật chữ ký số, không phủ nhận góp phần tạo dựng sự tin cậy và khả năng kiểm chứng trong giao dịch điện tử.
* An ninh hoạt động (Operational Security): Bảo vệ thông tin quan trọng của tổ chức khỏi sự truy cập trái phép là một nhiệm vụ cốt yếu, nhằm ngăn chặn thông tin nội bộ rò rỉ hoặc bị sử dụng một cách xấu xa.

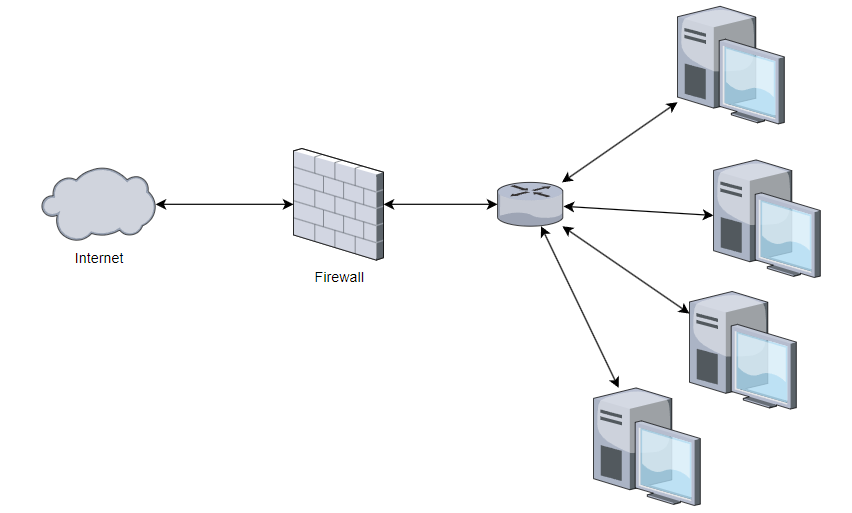
## **Các giải pháp an ninh mạng**

### ***Tường lửa (Firewalls):***

Tường lửa [6] là một thành phần quan trọng của các hệ thống an ninh mạng, hoạt động như một rào cản giữa một mạng nội bộ và các mạng bên ngoài, kiểm soát lưu lượng mạng dựa trên các quy tắc được xác định trước. Nó hoạt động như một bộ lọc, quản lý và kiểm soát luồng gói tin mạng vào và ra khỏi mạng, dựa trên các quy tắc được xác định trước để cho phép hoặc chặn giao tiếp dựa trên các tiêu chí như địa chỉ IP, cổng và loại giao thức.

Tường lửa có thể triển khai dưới dạng phần cứng hoặc phần mềm, hoặc sự kết hợp của cả hai. Chúng cung cấp một hàng rào phòng thủ đầu tiên cho mạng, giúp ngăn chặn các cuộc tấn công từ internet, bao gồm các cuộc tấn công từ xa như tấn công DDoS (từ chối dịch vụ), tấn công malware và các nỗ lực quét mạng.

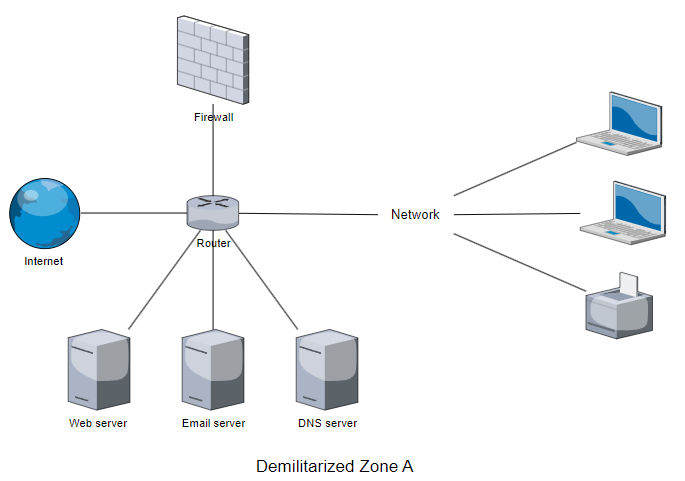
Tường lửa cũng có khả năng giám sát và nhật ký hoạt động mạng, cung cấp thông tin để phát hiện và đáp ứng với các mối đe dọa mạng. Đó là một công cụ thiết yếu trong việc tạo ra một môi trường mạng an toàn và bảo mật.



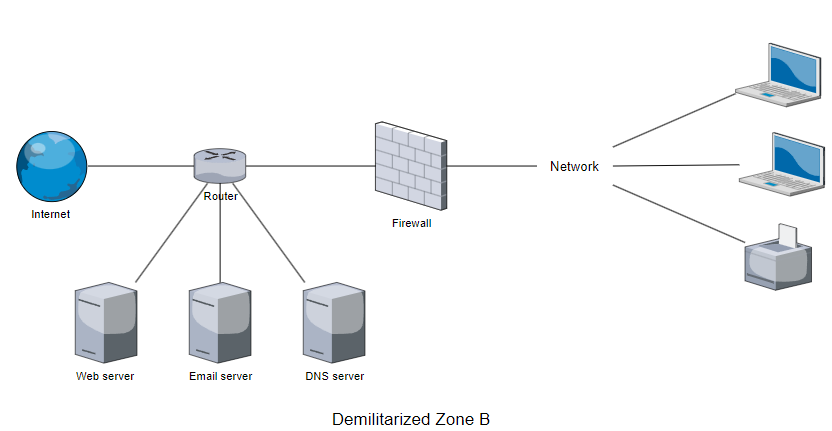
*HÌNH 2.1 Tường lửa (Firewall) trong hệ thống mạng*

**Vị trí của tường lửa [6]:**

Trong một hệ thống mạng, tường lửa có thể được đặt ở cổng gateway hoặc được đặt giữa nội bộ mạng và bộ định tuyến, đó cũng chính là 2 vị trí phổ biến nhất để đặt tường lửa.



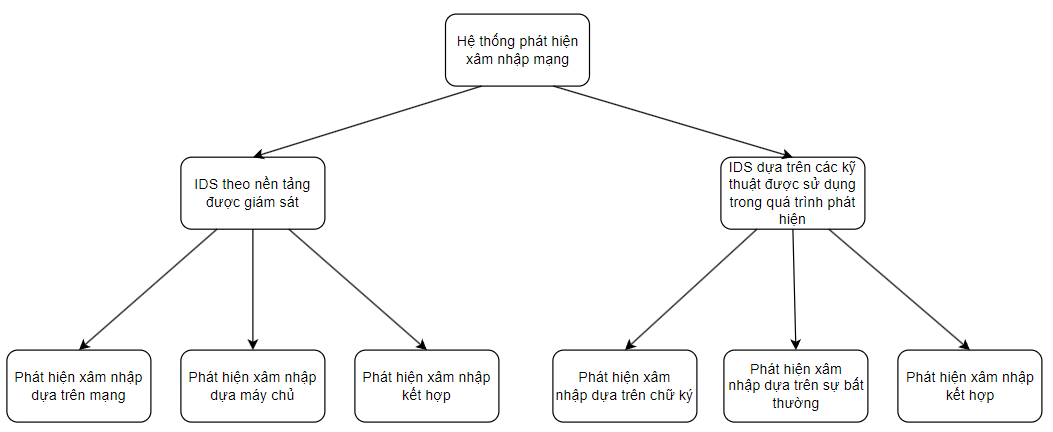
*HÌNH 2.2 Tường lửa ở cổng Gateway*



*HÌNH 2.3 Tường lửa ở giữa nội bộ mạng và bộ định tuyến (router)*

Tường lửa ở gateway có thể kiểm soát lưu lượng truy cập từ và đến internet, bảo vệ mạng khỏi các cuộc tấn công từ mạng ngoại bộ và kiểm soát quyền truy cập vào các tài nguyên nội bộ. Mặt khác tường lửa ở nội bộ mạng và router giúp kiểm soát lưu lượng mạng nội bộ, ngăn chặn các cuộc tấn công nội bộ và kiểm soát quyền truy cập giữa các phần khác nhau của mạng nội bộ.

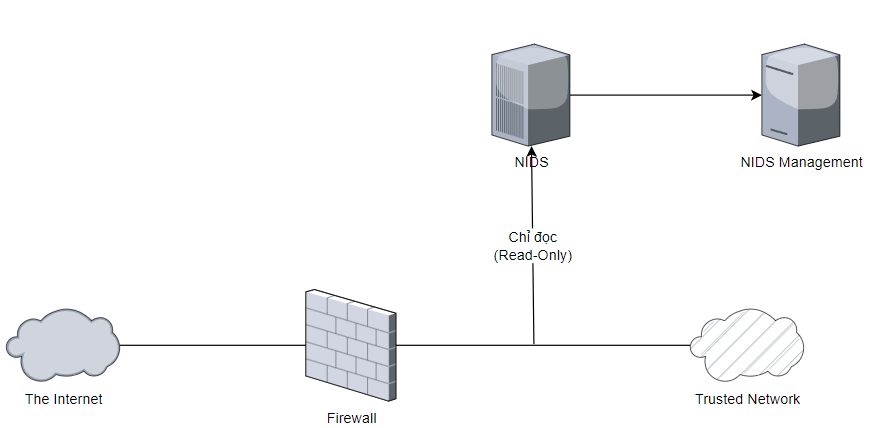
## **Các phương thức phát hiện xâm nhập**



*HÌNH 2.4 Các phương pháp phát hiện xâm nhập mạng*

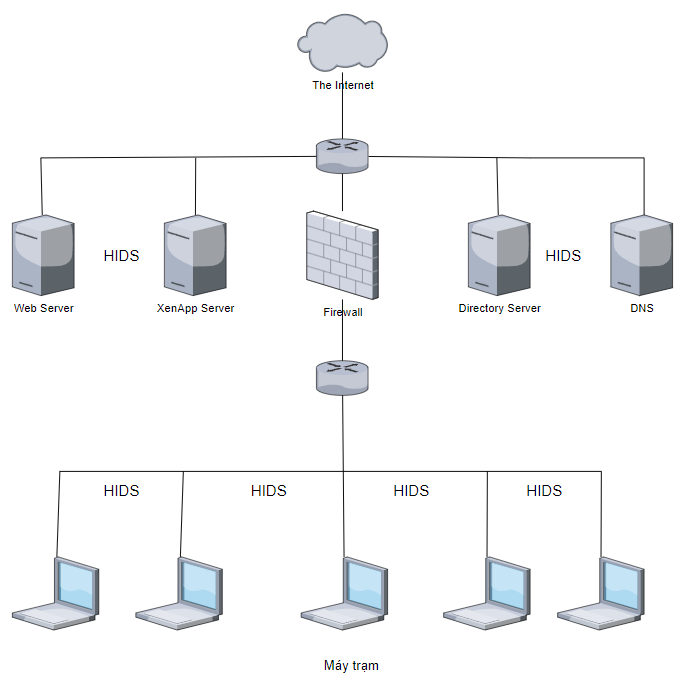
### ***IDS theo nền tảng được giám sát***

Network-based IDS (NIDS) (hệ thống phát hiện xâm nhập dựa trên mạng): Để bảo vệ tất cả các máy chủ, loại IDS này thường được cài đặt tại điểm vào của mạng, có thể là router biên của công ty, điểm vào của mạng nội bộ, hoặc khu vực DMZ (Vùng không quân). NIDS hoạt động bằng cách khám phá một loạt các điểm dữ liệu từ các nguồn mạng khác nhau. Để xác định xem có hoạt động độc hại hoặc bất thường nào đã xảy ra không, các tiêu đề gói tin, thống kê và luồng dữ liệu giao thức/ứng dụng được kiểm tra. Nó có thể được áp dụng để phát hiện các lỗ hổng bảo mật tiềm ẩn trên một hệ thống, chẳng hạn như các cuộc tấn công sniffer và các cuộc tấn công vào các dịch vụ HTTP/S, SMB, SSH và các dịch vụ khác. Tuy nhiên, chúng cũng có nhược điểm, chẳng hạn như khó khăn trong việc phân tích tất cả các gói tin từ một mạng lớn và quá tải.



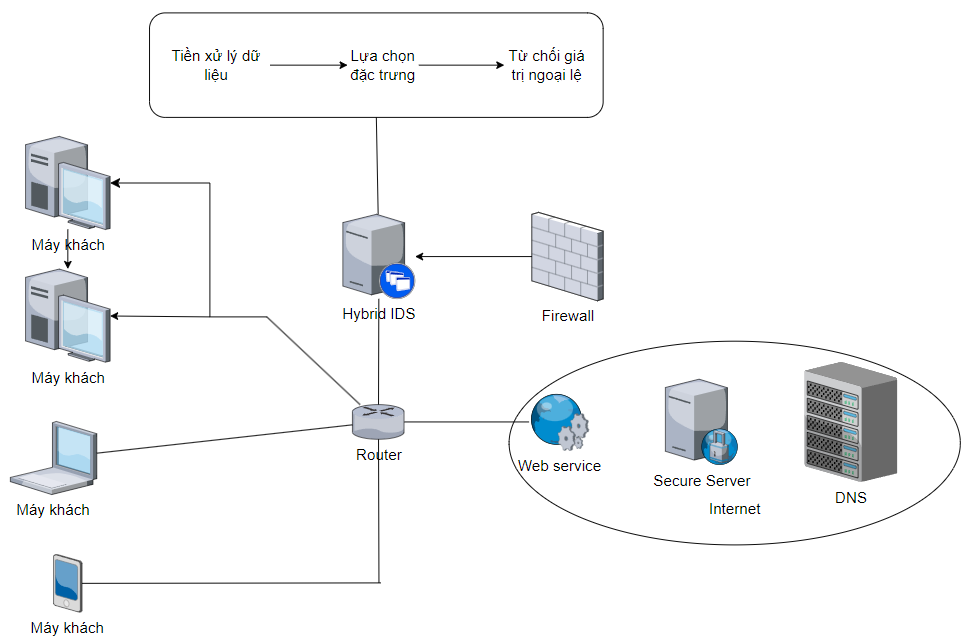
*HÌNH 2.5 Phát hiện xâm nhập dựa trên mạng (Network-based IDS)*

* Host-based IDS (HIDS) (hệ thống phát hiện xâm nhập dựa trên máy chủ): Các máy chủ đơn lẻ là trường hợp sử dụng phổ biến nhất cho HIDS. Chúng được cài đặt trên máy chủ như là phần mềm. Điểm khác biệt giữa HIDS và NIDS là HIDS có thể được cài đặt trên nhiều loại máy khác nhau như máy chủ, máy trạm hoặc máy tính xách tay. Loại IDS này đảm bảo an ninh cho máy chủ bằng cách giám sát chỉ lưu lượng gói tin vào và ra của máy chủ và cảnh báo người dùng hoặc quản trị viên nếu phát hiện hoạt động đáng ngờ. Khi lưu lượng được truyền đến máy chủ, nó sẽ được phân tích và chuyển qua máy chủ nếu hệ thống không phát hiện gói tin mang mã nguy hiểm bên trong. Nó bảo vệ khỏi các hệ thống phát hiện xâm nhập và tường lửa không thể phát hiện các cuộc tấn công không thể phát hiện được. HIDS có một lợi ích bổ sung so với NIDS là nó có thể phát hiện xem một kẻ tấn công có thành công hay không nhanh chóng.



*HÌNH 2.6 Phát hiện xâm nhập dựa trên máy chủ (Host-based IDS)*

* Hybrid IDS (hệ thống phát hiện xâm nhập kết hợp): Các Hệ thống Phát hiện Xâm nhập Kết hợp được cài đặt giữa một Hệ thống Phát hiện Xâm nhập Máy chủ và một Hệ thống Phát hiện Xâm nhập Mạng. Chúng được tạo ra bằng cách thu thập thông tin hoặc dữ liệu để phân tích từ mạng và máy chủ. Những hệ thống này sẽ rất thành công trong tương lai vì chúng giải quyết nhược điểm của cả HIDS và NIDS.



*HÌNH 2.7 Phát hiện xâm nhập kết hợp (Hybrid IDS)*

### ***IDS dựa trên kỹ thuật được sử dụng trong quá trình phát hiện***

* IDS dựa trên chữ ký (Signature-based IDS): Sử dụng các đặc trưng cố định, được biết đến từ trước của các cuộc tấn công, IDS dựa trên chữ ký là công cụ mạnh mẽ trong việc nhận diện và ngăn chặn các mối đe dọa đã được định nghĩa trước.
* IDS dựa trên bất thường (Anomaly-based IDS): Phương pháp này nhận diện các hoạt động bất thường bằng cách so sánh chúng với hành vi mạng được xác định là bình thường. IDS dựa trên bất thường có khả năng phát hiện những tấn công mới mà chưa từng được biết đến trước đây, nhưng cũng đối mặt với thách thức là tỷ lệ báo động giả cao do phải phân biệt giữa hoạt động bất thường và hoạt động hợp pháp không theo mô hình thông thường.
* IDS kết hợp (Hybrid IDS): Bằng cách tích hợp cả hai phương pháp trên, IDS kết hợp tận dụng ưu điểm của cả hai để tối ưu hóa khả năng phát hiện, đồng thời giảm thiểu số lượng báo động giả, đem lại một lớp bảo vệ thông tin tốt hơn.

## **Cơ sở dữ liệu trong Phát hiện Xâm nhập**

Trong việc phát triển và đánh giá các hệ thống IDS, dữ liệu đóng một vai trò không thể phủ nhận. Các bộ dữ liệu như NSL-KDD và các tập dữ liệu khác đã được sử dụng rộng rãi trong cộng đồng nghiên cứu để làm nền tảng cho việc kiểm tra, đánh giá và huấn luyện các thuật toán phát hiện xâm nhập dựa trên học máy và học sâu. Những cơ sở dữ liệu này cung cấp cho nhà nghiên cứu một khung tham chiếu về các loại tấn công, mô hình hoạt động mạng thông thường và bất thường, từ đó giúp cải thiện độ chính xác và khả năng tổng quát hóa của mô hình IDS.

## **Mô hình học máy và Học sâu trong IDS**

Học máy và học sâu đã cách mạng hóa cách chúng ta tiếp cận với phát hiện xâm nhập mạng, cung cấp khả năng tự động hóa phát hiện và phân loại các hành động đáng ngờ nghi trong mạng máy tính.

### ***Học Máy trong IDS (Machine Learning in IDS)***

Các thuật toán học máy như Support Vector Machines (SVM), Decision Trees, và Random Forests đã được áp dụng rộng rãi trong việc phân loại giao thông mạng và phát hiện các hoạt động bất thường. Chúng cho phép máy tính học từ dữ liệu quá khứ và phát triển khả năng dự đoán hoặc nhận dạng các mẫu xâm nhập mới. Tuy nhiên, một thách thức lớn là sự không cân đối dữ liệu khi mà giao thông bình thường chiếm đa số và giao thông xâm nhập là hiếm. Do đó, các phương pháp tái cấu trúc và tái cân bằng dữ liệu như SMOTE (Synthetic Minority Over-sampling Technique) và các biến thể của nó đã được áp dụng để cải thiện khả năng nhận dạng tấn công của mô hình.

### ***Học Sâu trong IDS (Deep Learning in IDS)***

Học sâu, một nhánh của học máy, sử dụng các mạng nơ-ron sâu để phân tích dữ liệu. Các mô hình như Convolutional Neural Networks (CNN) và Recurrent Neural Networks (RNN), đặc biệt là Long Short-Term Memory networks (LSTM), được sử dụng để phát hiện các mẫu phức tạp trong dữ liệu mạng. Các mô hình này đặc biệt hữu ích trong việc xử lý và học từ dữ liệu chuỗi thời gian, một đặc điểm chung của giao thông mạng. Ngoài ra, việc áp dụng các kiến trúc nâng cao như Gated Recurrent Units (GRU) hoặc Attention Mechanisms đã giúp tăng cường khả năng hiểu biết về các chuỗi sự kiện trong dữ liệu mạng.

## **Tập dữ liệu và đánh giá mô hình IDS**

NSL-KDD, CICIDS2017 và các tập dữ liệu mới hơn như CSE-CIC-IDS 2018 hay Bot-IoT là những nguồn dữ liệu có giá trị cho việc đánh giá IDS. Tuy nhiên, việc lựa chọn đánh giá mô hình không chỉ dừng lại ở độ chính xác phát hiện mà còn phải cân nhắc đến tỉ lệ báo động giả (false positive rate), thời gian phản hồi và khả năng tích hợp của mô hình vào hệ thống an ninh mạng hiện hữu. Ngoài ra, phải có một hệ thống so sánh trực tiếp hiệu suất giữa các mô hình khác nhau. Điều này đặt ra yêu cầu về sự phát triển của các chỉ số đánh giá toàn diện và các phương pháp thử nghiệm chuẩn mực, đồng thời cũng cần sự chú trọng đặc biệt đến khả năng áp dụng thực tế của các mô hình trong môi trường mạng đa dạng và liên tục thay đổi.

## **Các thách thức trong đánh giá mô hình IDS**

Trong quá trình đánh giá mô hình IDS, chúng ta gặp phải một loạt các thách thức phải được giải quyết. Sự không cân đối giữa lớp dữ liệu (data imbalance) là một vấn đề, vì trong thực tế, số lượng giao dịch hợp lệ thường nhiều hơn rất nhiều so với giao dịch tấn công, làm giảm khả năng nhận diện của mô hình. Sự biến đổi của các dạng tấn công (evolving attack patterns) làm cho việc phát hiện trở nên khó khăn hơn, đòi hỏi mô hình phải liên tục được cập nhật và tu chỉnh để theo kịp với các mối đe dọa mới. Bên cạnh đó, việc thiếu một phương pháp đánh giá thống nhất khiến việc so sánh hiệu quả giữa các mô hình khác nhau trở nên phức tạp và thiếu chính xác. Để giải quyết các vấn đề này, cần có một hệ thống đánh giá tiêu chuẩn và cập nhật, cùng với việc áp dụng các kỹ thuật cải thiện mô hình như học không cân đối (imbalance learning) và học tăng cường (reinforcement learning) để tối ưu hóa hiệu suất phát hiện của các mô hình IDS.

# **MÔ HÌNH ĐỀ XUẤT**

giới thiệu muốn làm gì

trước khi vào 3.1 thì mik giới thiệu muốn làm gì, mô tả sơ về bộ dữ liệu có gì (thay đổi kết hợp phía trên này và thay đổi dưới 3.1)

Trong chương này, chúng tôi sẽ giới thiệu quy trình xây dựng mô hình phát hiện xâm nhập mạng dựa trên việc lựa chọn đặc trưng thông minh và áp dụng các mô hình phân loại hiệu quả. Sử dụng tập dữ liệu từ một môi trường mạng quân sự mô phỏng, bao gồm dữ liệu TCP/IP với các hoạt động bình thường và tấn công, chúng tôi nhấn mạnh vào việc tinh giản không gian đặc trưng thông qua phương pháp Recursive Feature Elimination (RFE) kết hợp với mô hình RandomForest. Quy trình này giúp loại bỏ thông tin không cần thiết, giảm thiểu nguy cơ overfitting, và cải thiện đáng kể khả năng tổng quát hóa của mô hình. Phần tiếp theo sẽ chi tiết hóa việc tiền xử lý dữ liệu, lựa chọn đặc trưng, và cuối cùng là lựa chọn và huấn luyện mô hình, với mục tiêu nâng cao hiệu suất phát hiện xâm nhập trong môi trường mạng ngày càng phức tạp.

## **Dữ liệu sử dụng:**

Tập dữ liệu được sử dụng để xây dựng và đánh giá các mô hình phát hiện xâm nhập mạng trong nghiên cứu này được cung cấp thông qua một môi trường mạng quân sự mô phỏng đa dạng các loại xâm nhập. Dữ liệu TCP/IP thô được thu thập từ việc mô phỏng một mạng LAN điển hình của Không quân Hoa Kỳ, được thiết lập như một môi trường thực tế và chịu đựng nhiều cuộc tấn công. Một kết nối được định nghĩa là chuỗi các gói tin TCP bắt đầu và kết thúc trong một khoảng thời gian nhất định, trong đó dữ liệu được truyền từ một địa chỉ IP nguồn đến địa chỉ IP đích theo một giao thức rõ ràng. Mỗi kết nối được gắn nhãn là bình thường hoặc là một cuộc tấn công với một loại tấn công cụ thể.

Mỗi bản ghi kết nối chứa khoảng 100 byte thông tin được thu thập từ dữ liệu bình thường và tấn công, bao gồm 41 đặc trưng định lượng và định tính (3 đặc trưng định tính và 38 đặc trưng định lượng). Biến phân loại có hai danh mục:

- Bình Thường (Normal): Các kết nối không chứa hoạt động đáng ngờ hoặc tấn công.

- Bất Thường (Anomalous): Các kết nối chứa hoạt động xâm nhập hoặc tấn công, được gắn nhãn cụ thể với một loại hình tấn công nhất định.

Tập dữ liệu này cung cấp một cơ sở thực tế và chi tiết để phát triển và kiểm tra các thuật toán phát hiện xâm nhập, từ đó tăng cường khả năng phòng thủ của mạng trước các nguy cơ an ninh mạng hiện đại và phức tạp.

Tiếp theo, chúng tôi tiến hành phân tích khám phá dữ liệu (Data Exploratory Analysis) và trực quan hóa dữ liệu (Data Visualization), và tiến hành quá trình xử lý dữ liệu (Data Processing). Chúng tôi tiến hành một loạt các bước để hiểu rõ hơn về tập dữ liệu và những đặc điểm quan trọng của nó.

Đầu tiên, chúng tôi khám phá tổng quan về dữ liệu để xác định mọi thiếu sót hoặc dạng dữ liệu không đồng nhất, đồng thời phát hiện các mẫu hoặc xu hướng đặc biệt có thể hỗ trợ trong việc phát hiện xâm nhập.

Một bước quan trọng trong phân tích khám phá là đánh giá phân phối của các đặc trưng, bao gồm cả định lượng và định tính. Điều này giúp chúng tôi xác định được những đặc trưng có tiềm năng cao trong việc phân biệt giữa các kết nối bình thường và bất thường. Chúng tôi cũng tiến hành kiểm tra các giá trị null và loại bỏ hoặc điền các giá trị thiểu để đảm bảo dữ liệu đầu vào cho mô hình là hoàn chỉnh và chính xác nhất.

Tiếp theo, việc mã hóa các đặc trưng định tính thành định lượng được thực hiện thông qua phương pháp LabelEncoder và OneHotEncoder, cho phép các thuật toán học máy có thể hiểu và xử lý dữ liệu một cách hiệu quả. Quá trình này đặc biệt quan trọng vì nó chuyển đổi dữ liệu từ dạng có thể đọc được bởi con người sang dạng máy có thể xử lý được, mở ra khả năng áp dụng các kỹ thuật học máy tiên tiến.

Sau khi mã hóa, chúng tôi áp dụng phương pháp RFE để chọn lọc đặc trưng. Điều này giúp xác định 10 đặc trưng quan trọng nhất, đồng thời loại bỏ bớt sự phức tạp và tối ưu hóa quá trình huấn luyện. Những đặc trưng được lựa chọn không chỉ phản ánh mức độ quan trọng trong việc phân loại mà còn giúp giảm thiểu thời gian huấn luyện và cải thiện khả năng tổng quát hóa của mô hình.

Cuối cùng, phân tích khám phá dữ liệu cung cấp cái nhìn sâu sắc và toàn diện về tập dữ liệu, là nền tảng vững chắc cho các bước tiếp theo trong quy trình xây dựng mô hình phát hiện xâm nhập mạng. Qua đó, chúng tôi không chỉ hiểu rõ về dữ liệu mà còn có thể lựa chọn và áp dụng các phương pháp tiếp cận phù hợp nhất để đạt được hiệu suất cao nhất trong nhiệm vụ phát hiện xâm nhập.

## **Tiền xử lý Dữ liệu:**

Tiền xử lý dữ liệu là bước quan trọng để chuẩn bị dữ liệu trước khi đưa vào thuật toán học máy. Mục tiêu của việc chuẩn bị dữ liệu là loại bỏ sự mơ hồ và cung cấp cho Hệ thống Phát hiện Xâm nhập (IDS) dữ liệu chính xác. Quy trình này kết hợp việc chọn lọc và chuẩn hóa đặc trưng.

Trong nghiên cứu này, quá trình tiền xử lý dữ liệu được thực hiện như sau:

### ***Xử lý dữ liệu rỗng (null) và dữ liệu dạng chuỗi:***

a. Tất cả các giá trị null được kiểm tra và xử lý thích hợp để đảm bảo không có thông tin trống nào gây ảnh hưởng đến mô hình.

b. Các đặc trưng dạng chuỗi như `protocol\_type`, `service`, và `flag` được mã hóa thành dữ liệu số để thuật toán có thể xử lý. Điều này được thực hiện thông qua các phương pháp sau:

- Sử dụng `pd.to\_numeric` để chuyển đổi các cột dạng chuỗi thành số, đặt tham số `errors='coerce'` để chuyển đổi các giá trị không thể mã hóa thành NaN.

- `LabelEncoder` áp dụng cho các cột có danh mục, gán mỗi giá trị danh mục một số nguyên duy nhất.

- `OneHotEncoder` được sử dụng để chuyển các giá trị danh mục thành vector nhị phân, loại bỏ sự thiên vị mà số nguyên có thể tạo ra, mỗi danh mục được biểu diễn bởi một vector với một giá trị '1' ở vị trí tương ứng và '0' ở các vị trí khác.

Trong nghiên cứu này đã sử dụng phương pháp mã hóa LabelEncoder cho 3 cột nêu trên.

### ***Chọn lọc đặc trưng đơn giản (Simple Feature Selection):***

- Kiểm tra các cột có giá trị giống hệt nhau hoặc không đủ đa dạng, có thể không mang lại thông tin hữu ích cho mô hình, và xóa bỏ chúng nếu cần.

Việc tiền xử lý này giúp tăng cường chất lượng dữ liệu, làm cho việc học máy trở nên hiệu quả hơn và tạo điều kiện cho việc phát hiện xâm nhập diễn ra chính xác hơn. Đặc biệt, việc chuyển đổi dữ liệu từ dạng danh mục sang dạng số là bước thiết yếu để thuật toán có thể làm việc với dữ liệu một cách hiệu quả, nhất là trong các bài toán phân loại đa lớp và nhị phân, nơi mà dữ liệu được chia thành các lớp tấn công và lớp bình thường.

## **Lựa chọn đặc trưng dữ liệu:**

Trong khuôn khổ của việc xây dựng mô hình phát hiện xâm nhập mạng, việc lựa chọn đặc trưng (feature selection) giữa vai trò cực kỳ quan trọng, không chỉ góp phần nâng cao độ chính xác của mô hình mà còn tối ưu hóa quá trình huấn luyện. Phương pháp Recursive Feature Elimination (RFE) kết hợp với mô hình phân loại RandomForest được áp dụng nhằm mục đích tinh giản không gian đặc trưng, tập trung vào những thông tin có giá trị dự báo cao nhất.

Quá trình RFE bắt đầu bằng việc huấn luyện một mô hình phân loại trên toàn bộ tập hợp đặc trưng và đánh giá mức độ quan trọng của từng đặc trưng. Trong bối cảnh sử dụng Random Forest, mỗi đặc trưng được đánh giá dựa trên mức độ ảnh hưởng của nó đến độ chính xác của mô hình. Đặc trưng được xem là ít quan trọng nhất sẽ được loại bỏ, và quy trình này được lặp lại cho đến khi chỉ còn lại số lượng đặc trưng mong muốn. Quy trình này không chỉ giúp loại bỏ nhiễu và thông tin dư thừa mà còn giảm thiểu nguy cơ overfitting, qua đó tăng cường khả năng tổng quát hóa của mô hình.

Trong nghiên cứu phát hiện xâm nhập mạng của chúng tôi, RFE được sử dụng để xác định 10 đặc trưng quan trọng nhất giúp phân biệt giữa hoạt động bình thường và các mẫu xâm nhập. Việc này không chỉ giúp cải thiện độ chính xác và hiệu suất của mô hình mà còn giảm đáng kể thời gian huấn luyện và yêu cầu về tài nguyên tính toán. Hơn nữa, bằng cách tinh giản không gian đặc trưng, mô hình của chúng tôi còn có thể thích ứng tốt hơn với sự thay đổi trong môi trường dữ liệu, đảm bảo tính hiệu quả và độ tin cậy cao trong việc phát hiện xâm nhập mạng.

Quá trình lựa chọn đặc trưng diễn ra qua các bước sau: Đầu tiên, mô hình RandomForestClassifier được khởi tạo làm cơ sở cho việc đánh giá mức độ quan trọng của các đặc trưng. Tiếp theo, RFE được huấn luyện dựa trên tập dữ liệu, với thông số n\_features\_to\_select được đặt là 10 để chỉ định số lượng đặc trưng cần lựa chọn. Các đặc trưng được RFE lựa chọn cuối cùng là kết quả của quá trình đánh giá và so sánh liên tục để tìm ra bộ đặc trưng tối ưu.

Những đặc trưng cuối cùng được chọn bao gồm: `protocol\_type`, `flag`, `src\_bytes`, `dst\_bytes`, `count`, `same\_srv\_rate`, `diff\_srv\_rate`, `dst\_host\_srv\_count`, `dst\_host\_same\_srv\_rate`, và `dst\_host\_same\_src\_port\_rate`. Sự lựa chọn này phản ánh một cách tiếp cận cẩn trọng, đảm bảo rằng mỗi đặc trưng được giữ lại đều làm tăng khả năng dự đoán của mô hình và có ý nghĩa thực tiễn trong việc phân tích và phát hiện các hoạt động xâm nhập mạng.

1. `protocol\_type`: Loại giao thức mạng (ví dụ: TCP, UDP, ICMP). Mỗi loại giao thức có đặc điểm riêng và thường được sử dụng trong các loại tấn công cụ thể, giúp nhận diện phương thức tấn công.

2. `flag`: Trạng thái của kết nối (ví dụ: S0, S1, REJ). Một số trạng thái cụ thể của kết nối có thể chỉ ra hoạt động đáng ngờ hoặc không bình thường.

3. `src\_bytes`: Lượng byte gửi từ địa chỉ IP nguồn đến địa chỉ IP đích. Lượng dữ liệu lớn không thường xuyên có thể chỉ ra một cuộc tấn công.

4. `dst\_bytes`: Lượng byte nhận được bởi địa chỉ IP đích từ địa chỉ IP nguồn. Giống như `src\_bytes`, lượng dữ liệu lớn không thường xuyên có thể là dấu hiệu của tấn công.

5. `count`: Số lượng kết nối đến cùng một host trong một khoảng thời gian nhất định. Sự tăng đột biến trong số lượng có thể báo hiệu một cuộc tấn công.

6. `same\_srv\_rate`: Tỷ lệ các kết nối đến cùng một dịch vụ. Một tỷ lệ thấp có thể chỉ ra sự scan dịch vụ hoặc tấn công từ chối dịch vụ.

7. `diff\_srv\_rate`: Tỷ lệ các kết nối đến các dịch vụ khác nhau. Một tỷ lệ cao có thể báo hiệu một hoạt động scan dịch vụ không bình thường.

8. `dst\_host\_srv\_count`: Số lượng kết nối từ địa chỉ IP nguồn đến địa chỉ IP đích với cùng một dịch vụ. Giá trị thấp có thể là dấu hiệu của việc scan cổng hoặc tấn công từ chối dịch vụ.

9. `dst\_host\_same\_srv\_rate`: Tỷ lệ kết nối đến cùng một dịch vụ từ host đích. Giá trị thấp có thể chỉ ra một hành vi scan hoặc tấn công khác nhau.

10. `dst\_host\_same\_src\_port\_rate`: Tỷ lệ kết nối đến host đích thông qua cùng một cổng nguồn. Tỷ lệ cao có thể báo hiệu một cuộc tấn công như tấn công từ chối dịch vụ hoặc scan cổng.

## **Chọn và Huấn luyện Mô hình**

Để giải quyết bài toán phân loại, tôi bắt đầu với việc thử nghiệm nhiều mô hình khác nhau, từ những mô hình cơ bản như Logistic Regression đến các mô hình phức tạp hơn như Random Forest và Gradient Boosting. Việc lựa chọn mô hình phù hợp là một bước quan trọng, giúp đảm bảo mô hình có khả năng học được những mẫu dữ liệu phức tạp từ tập dữ liệu.

tới phần đó, code đó có đặc điểm j, có j cho thấy mạng của mình bị xâm nhập

nói về các đặc trưng: dựa yếu tố nào để phát hiện bị xâm nhập -> nên mới lọc ra các đặc trưng đó

những cái đặc trưng để phát hiện ra cuộc tấn công có thể diễn ra -> chính vì vậy, những thứ đó là lý do nên để ý đến nó (PHÂN TÍCH DỮ LIỆU)

vì sao lại làm bước đó, tại sao lại có bước kia => (trong code có ghi rồi)

vì sao lại select được những futures đó => (dùng mô hình RFE)

### ***Huấn luyện Với Nhiều Mô hình với Tham Số Mặc Định***

Trong bối cảnh của việc giải quyết vấn đề phát hiện xâm nhập mạng (Network Intrusion Detection - NID), việc huấn luyện thử nghiệm và so sánh nhiều mô hình học máy khác nhau trở nên cực kỳ quan trọng và thiết yếu. Điều này không chỉ liên quan đến việc cải thiện hiệu suất và độ chính xác của hệ thống phát hiện xâm nhập mà còn đáp ứng nhiều yêu cầu cụ thể của lĩnh vực bảo mật mạng:

- Phát hiện tối ưu: Tìm ra mô hình có khả năng phát hiện các loại tấn công mạng một cách hiệu quả nhất.

- Độ chính xác và độ tin cậy: Đảm bảo hệ thống giảm thiểu được báo động giả và bỏ sót các mối đe dọa thực sự.

- Khả năng thích ứng: Lựa chọn mô hình có thể thích ứng tốt với sự thay đổi của dữ liệu và môi trường mạng.

- Hiểu biết sâu sắc về dữ liệu: Phân tích sâu để hiểu rõ các đặc điểm và mẫu trong dữ liệu, giúp tối ưu hóa mô hình.

- Tối ưu hóa tài nguyên: Chọn mô hình sử dụng hiệu quả tài nguyên hệ thống.

- So sánh và đánh giá: Cung cấp cơ sở để so sánh trực tiếp hiệu suất giữa các mô hình khác nhau.

- Sẵn sàng cho sự đổi mới: Chuẩn bị hệ thống cho việc tích hợp các giải pháp và thuật toán mới nâng cao khả năng phát hiện.

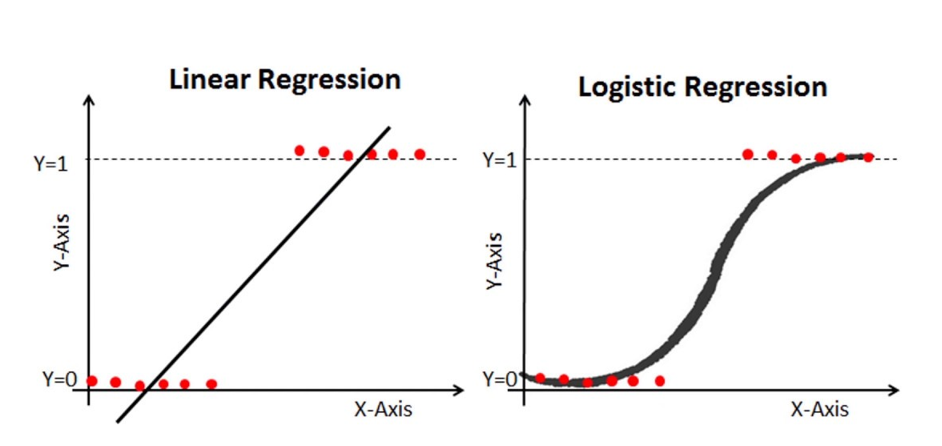
Những mục tiêu này giúp đảm bảo rằng hệ thống phát hiện xâm nhập mạng được tối ưu hóa để đối phó hiệu quả với các mối đe dọa hiện tại và tương lai, đồng thời sử dụng tài nguyên một cách hiệu quả nhất.

1. tại sao lại sử dụng nhiều mô hình -
2. tại sao lại chọn thuật toán này, nó có đặc điểm gì…mỗi mô hình sẽ có đặc điểm riêng, đặc trưng riêng j liên quan tới bài …nó có phục vụ cho mục tiêu của đề tài hay không,
3. tại sao lại phải điều chỉnh tham số (vì sao phải tuning)

#### Logistic Regression (Hồi quy Logistic):

Hồi quy Logistic sử dụng trong phát hiện xâm nhập mạng do khả năng phân loại hiệu quả giữa hoạt động bình thường và bất thường trên mạng. Sử dụng hàm logistic, mô hình này biến đổi đầu ra tuyến tính thành xác suất, cung cấp cơ sở định lượng để phân biệt hoạt động xâm nhập. Đặc biệt, hàm mất mát cross-entropy cho phép đánh giá chính xác sự khác biệt giữa dự đoán và thực tế, làm cơ sở để tối ưu hóa mô hình phát hiện xâm nhập mạng.

Hồi quy Logistic sử dụng hàm logistic để mô phỏng xác suất mà một đầu vào cụ thể thuộc về một lớp nào đó, thường là dạng nhị phân. biến đổi đầu ra của mô hình tuyến tính thành xác suất. Hàm mất mát được sử dụng là cross-entropy nơi mà `y` là nhãn thực và `p` là xác suất dự đoán của mô hình.



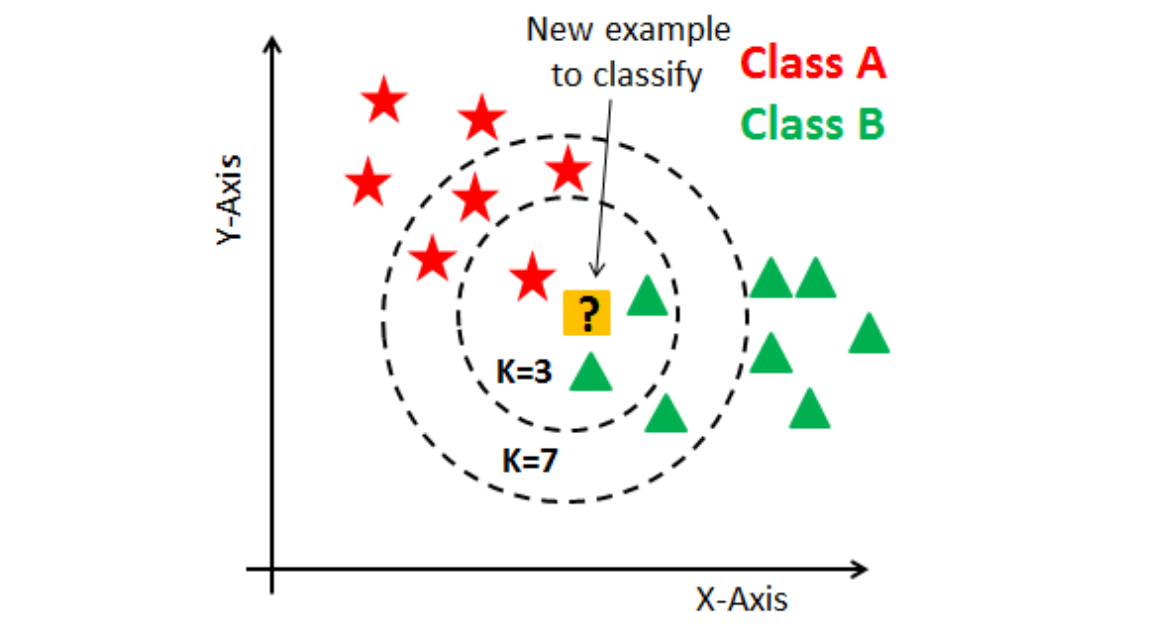
*HÌNH 3.1 So sánh Linear Regression và Logistic Regression*

(Nguồn: [Understanding Logistic Regression in Python Tutorial](https://www.datacamp.com/tutorial/understanding-logistic-regression-python), truy cập 14/02/2024)

#### K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) được áp dụng cho Network Intrusion Detection nhờ vào tính năng phân loại dựa trên sự tương đồng của dữ liệu. Cách tiếp cận này không yêu cầu một mô hình cụ thể, thích hợp cho việc phát hiện các loại tấn công mới mà chưa được định nghĩa rõ ràng trước đó. KNN xem xét 'k' điểm gần nhất và áp dụng bình chọn để xác định nhãn, giúp phát hiện các hoạt động đáng ngờ dựa trên mô hình hoạt động bình thường của mạng.

KNN là một thuật toán học không giám sát, phân loại một mẫu dữ liệu mới dựa trên 'k' điểm gần nhất trong không gian đặc trưng. Không sử dụng một hàm mất mát cụ thể, mô hình KNN đơn giản áp dụng phương pháp bỏ phiếu đa số từ 'k' lân cận gần nhất để quyết định nhãn của điểm dữ liệu mới.



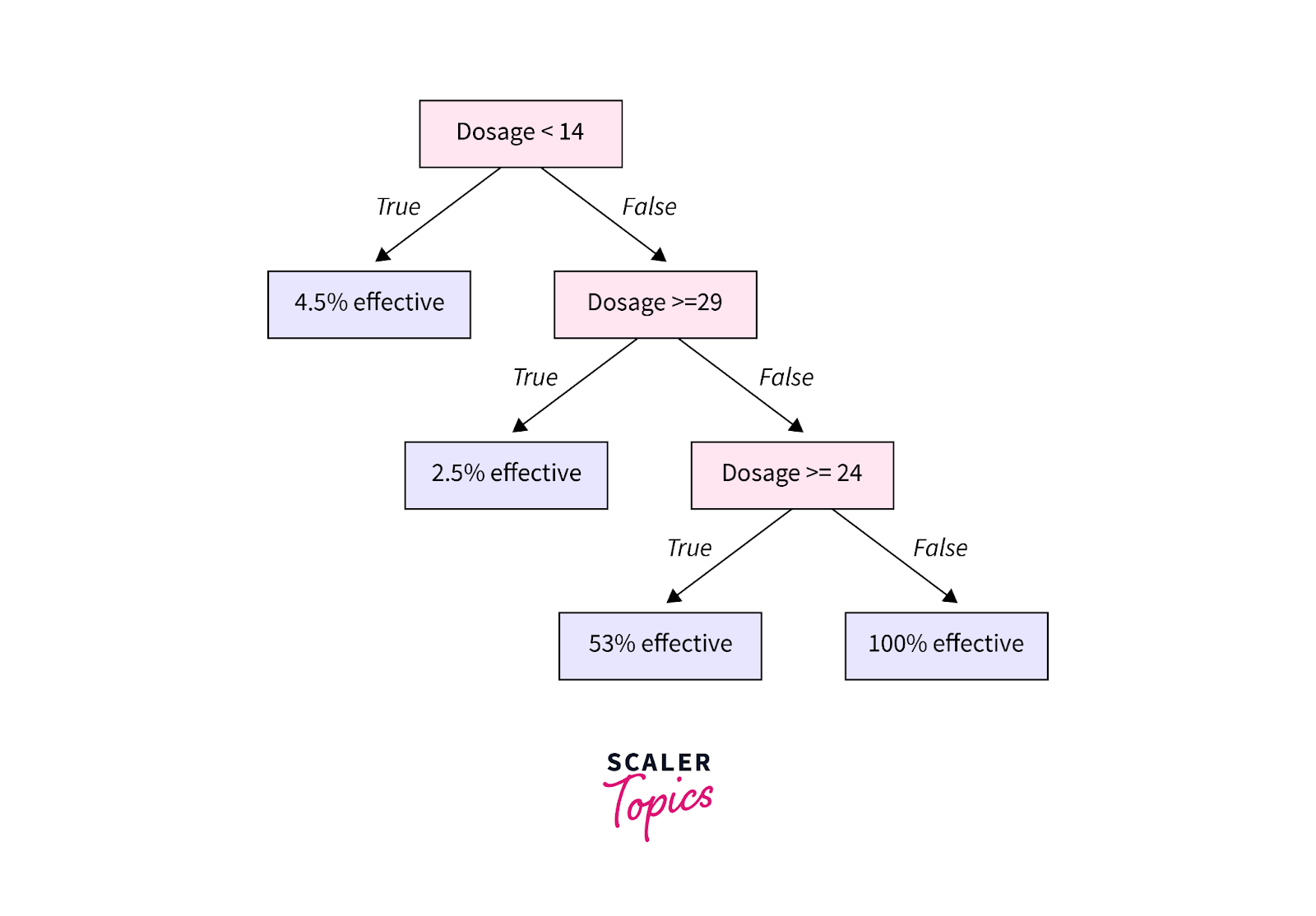
*HÌNH 3.2 Thuật toán kNN*

(Nguồn: [Introduction to k-Nearest Neighbors (kNN) Algorithm](https://ai.plainenglish.io/introduction-to-k-nearest-neighbors-knn-algorithm-e8617a448fa8), truy cập 14/2/2024)

#### Decision Tree (Cây quyết định):

Cây quyết định là công cụ quan trọng trong Network Intrusion Detection vì khả năng mô phỏng quyết định dễ hiểu, cung cấp cái nhìn trực quan về cách thức phân loại dữ liệu. Các quyết định tại mỗi nút dựa trên các thuộc tính của dữ liệu, cho phép phân biệt giữa traffic bình thường và xâm nhập. Sử dụng Gini impurity hoặc entropy để đánh giá độ "tinh khiết" của dữ liệu tại mỗi nút, giúp xác định cách thức tốt nhất để phân chia dữ liệu, tối ưu hóa khả năng phát hiện.

Cây quyết định là mô hình học có giám sát mô phỏng quá trình ra quyết định dạng cây. Mỗi nút đại diện cho một thuộc tính và mỗi nhánh đại diện cho một quyết định, dẫn tới kết quả ở các nút lá. Hàm mất mát thường liên quan đến độ tinh khiết của các nút, thường được đo bằng



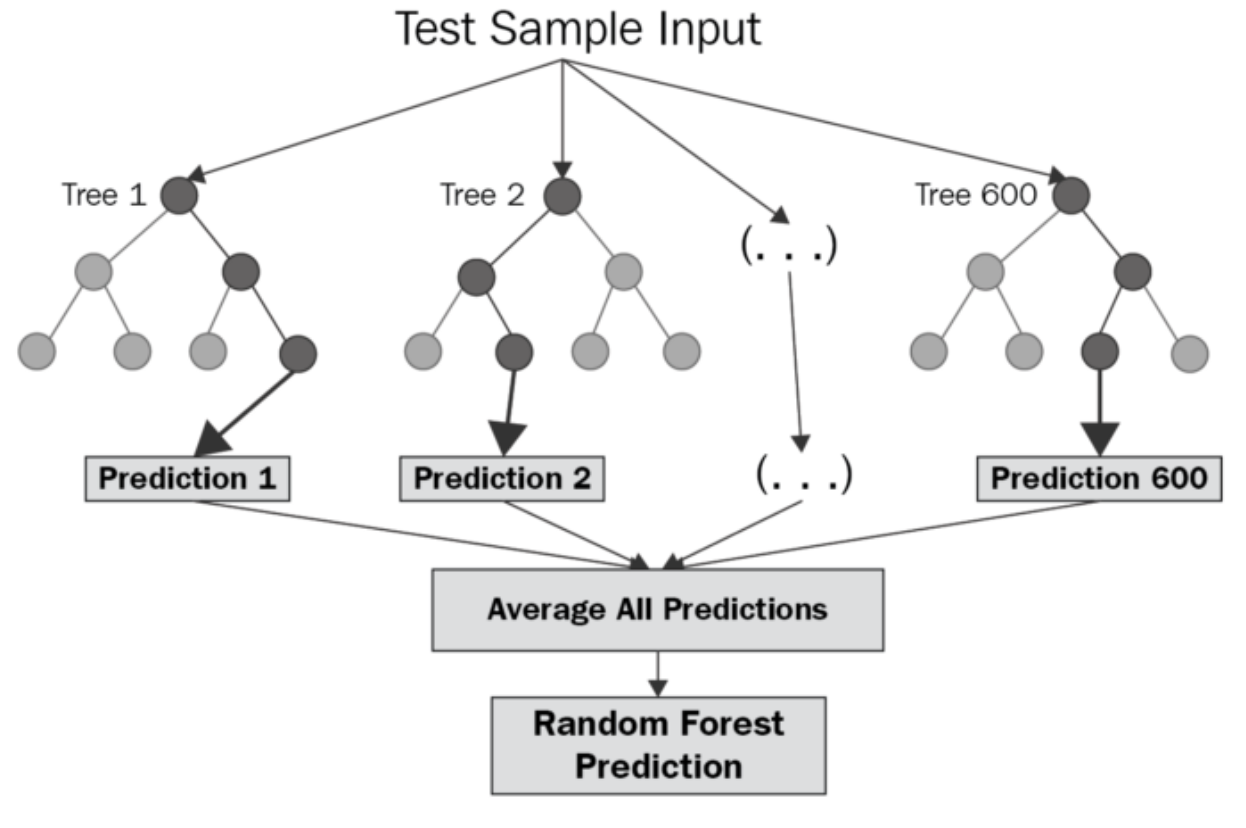
*HÌNH 3.3 Thuật toán Decision Trees*

(Nguồn: [What is Decision Trees in Machine Learning?](https://www.scaler.com/topics/machine-learning/what-is-decision-trees-in-machine-learning/), truy cập 14/02/2024)

#### Random Forest (Rừng ngẫu nhiên):

Rừng ngẫu nhiên, với sự kết hợp của nhiều cây quyết định, phù hợp với Network Intrusion Detection nhờ vào khả năng cải thiện độ chính xác và giảm overfitting. Mỗi cây làm việc với một tập con ngẫu nhiên của dữ liệu, giúp tăng cường độ đa dạng trong quyết định phân loại và tăng khả năng phát hiện các loại tấn công phức tạp, từ đó nâng cao chất lượng phát hiện xâm nhập.

Random Forest kết hợp nhiều cây quyết định, mỗi cây làm việc với một tập con ngẫu nhiên của dữ liệu và đặc trưng. Mô hình cuối cùng là kết quả tổng hợp của các cây. Hàm mất mát ở đây không chỉ giống như cây quyết định đơn lẻ mà còn bao gồm việc giảm sự phụ thuộc lẫn nhau giữa các cây để giảm overfitting.



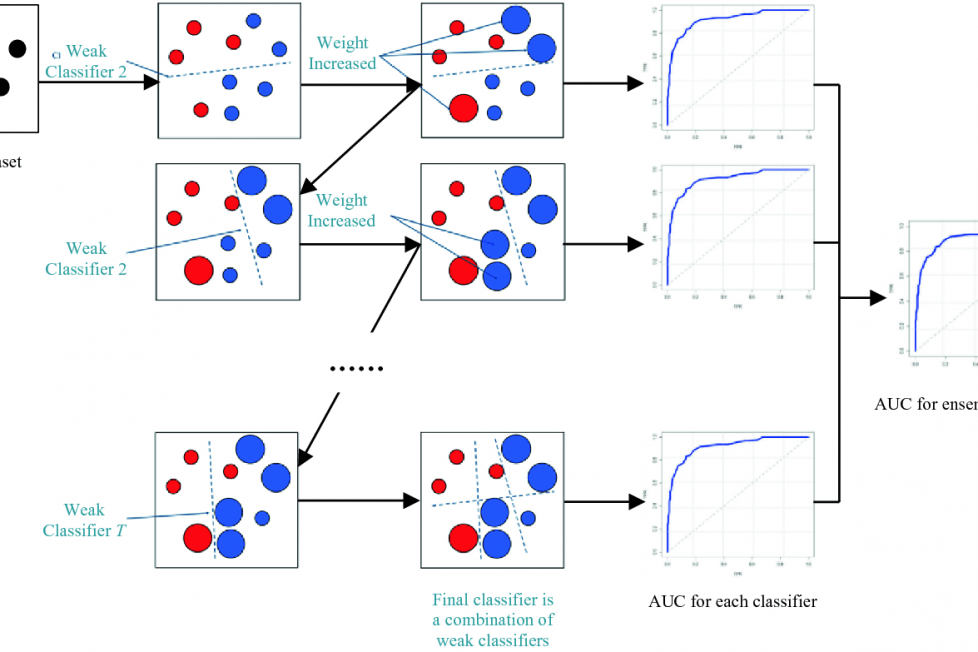
*HÌNH 3.4 Thuật toán Random Forest*

(Nguồn: [Diving into the Deep learning : Random Forest Algorithm](about:blank), truy cập 14/02/2024)

#### Gradient-Boosted Trees (Cây tăng cường gradient):

Cây tăng cường gradient tối ưu hóa hiệu suất phát hiện xâm nhập bằng cách tuần tự cải thiện những sai lầm của các mô hình trước, làm cho nó đặc biệt hữu ích trong việc đối phó với các vấn đề phân loại phức tạp. Phương pháp này tập trung vào việc giảm thiểu sai số dự đoán, qua đó cải thiện đáng kể khả năng phát hiện các dạng tấn công mới và phức tạp trong mạng. Việc áp dụng các cây tăng cường gradient cho phép mô hình tự động điều chỉnh dựa trên dữ liệu, giúp tối ưu hóa quá trình phân loại và nhận diện các mẫu tấn công cụ thể một cách chính xác hơn.

Cây tăng cường gradient là một phương pháp ensemble mà trong đó các cây được thêm vào một cách tuần tự để sửa chữa những sai lầm của mô hình trước đó. Hàm mất mát cụ thể có thể khác nhau, nhưng nói chung mục tiêu là giảm dần sai số dự đoán thông qua việc tối ưu hóa gradient, thường là sử dụng để đo lường sai số bình phương.



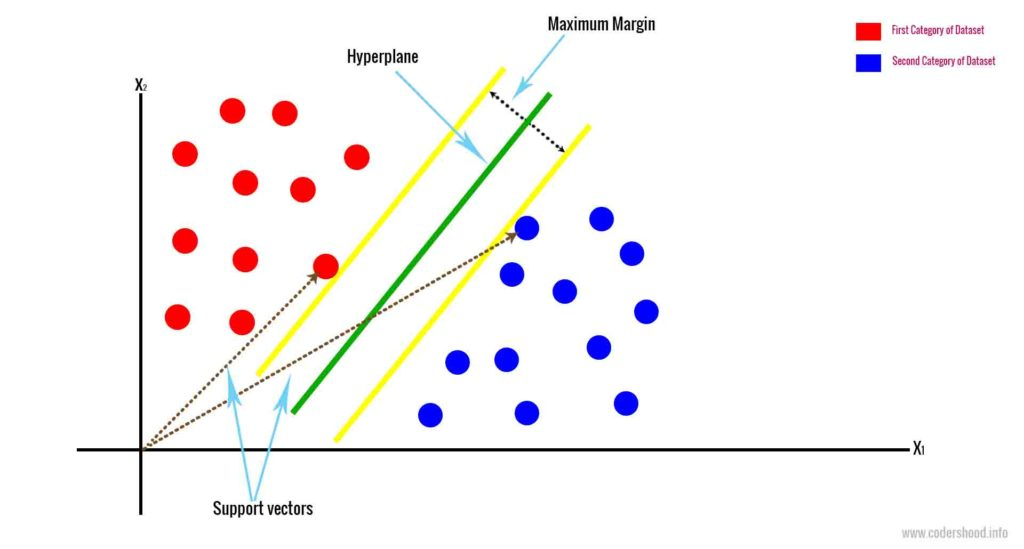
*HÌNH 3.5 Thuật toán Gradient Boosting*

(Nguồn: <https://datascience.eu/machine-learning/gradient-boosting-what-you-need-to-know/>, truy cập 14/02/2024)

#### Support Vector Machine (SVM):

Support Vector Machine (SVM) thích hợp cho Network Intrusion Detection do khả năng tạo ra mặt phẳng quyết định tối ưu để phân tách dữ liệu thành các lớp khác nhau. Bằng việc tối ưu hóa biên phân cách và giảm thiểu phân loại sai, SVM hiệu quả trong việc xử lý dữ liệu nhiều chiều, điều thường thấy trong phân tích traffic mạng. Cơ chế này đặc biệt hữu ích cho việc phân loại dữ liệu phức tạp, giúp nhận diện các hoạt động bất thường với độ chính xác cao.

SVM tìm ra mặt phẳng quyết định tốt nhất để phân chia dữ liệu thành các lớp khác nhau. Nó sử dụng hàm mất mát hinge, để tối ưu hóa biên phân cách và giảm thiểu các phân lớp sai. SVM có khả năng làm việc với không gian đặc trưng nhiều chiều và hiệu quả với dữ liệu có chiều cao mà không cần lớn về số lượng.



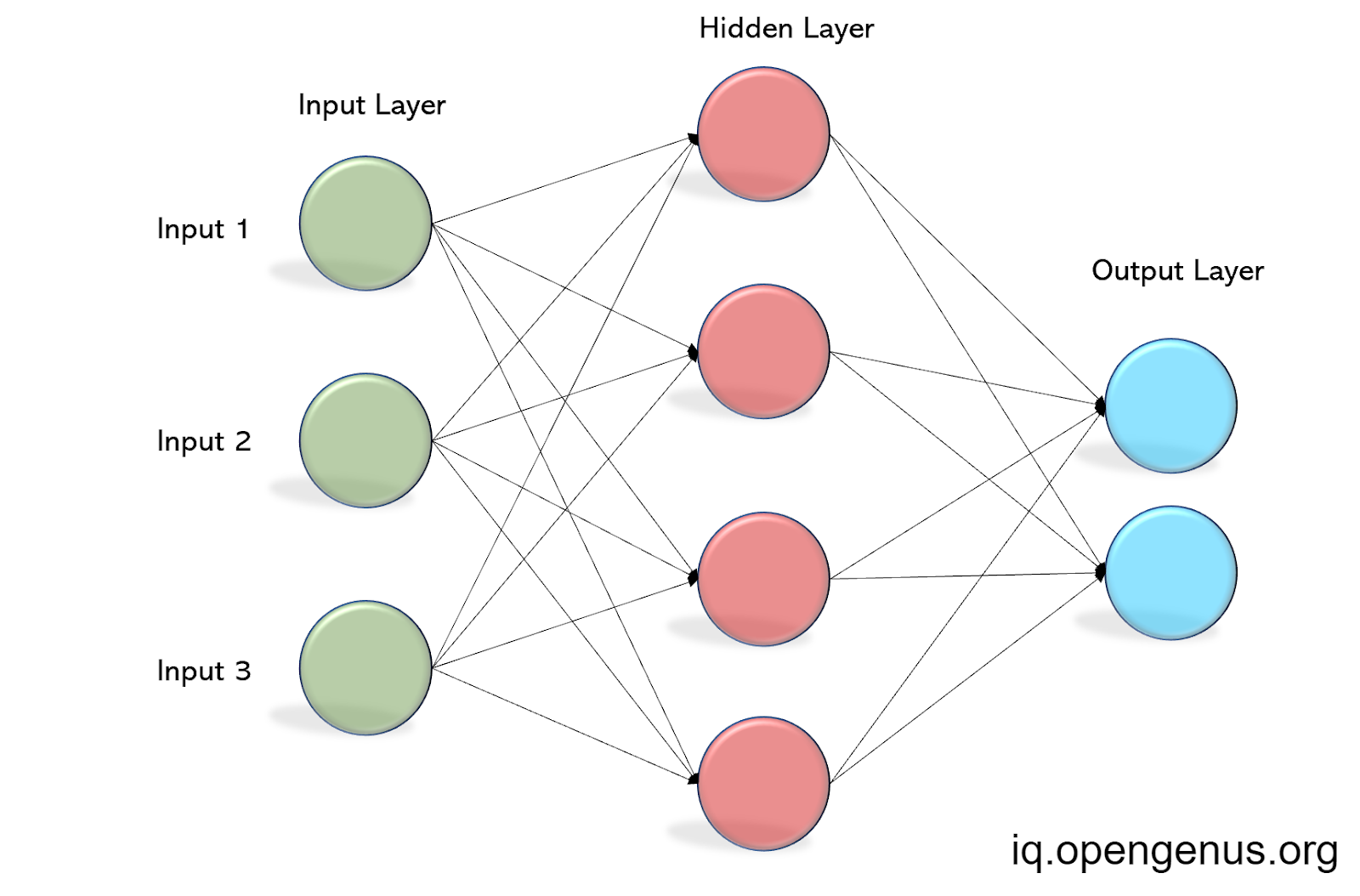
*HÌNH 3.6 Thuật toán Support Vector Machine*

(Nguồn: [Support Vector Machine Machine learning algorithm](about:blank), truy cập 14/02/2024)

#### Multilayer Perceptron (MLP):

Multilayer Perceptron (MLP), một dạng mạng neural nhân tạo, cung cấp khả năng mô hình hóa mối quan hệ phức tạp giữa các đặc trưng thông qua một hoặc nhiều lớp ẩn. Trong Network Intrusion Detection, MLP giúp phân tích các mẫu dữ liệu phức tạp, cho phép phát hiện các loại tấn công mà mô hình truyền thống khó nhận diện. Việc sử dụng backpropagation trong quá trình huấn luyện giúp tối ưu hóa hàm mất mát, nâng cao khả năng phân loại và nhận diện các hoạt động đáng ngờ trong mạng.

MLP là một dạng của mạng neural nhân tạo với một hoặc nhiều lớp ẩn, cho phép nó mô hình hóa các mối quan hệ phức tạp. Trong huấn luyện, nó thường sử dụng backpropagation để tối ưu hóa hàm mất mát, thường là hàm mất mát cross-entropy cho phân loại và mean squared error cho hồi quy.



*HÌNH 3.7 Thuật toán Multilayer Perceptron*

(Nguồn: [Multilayer Perceptrons algorithm](https://iq.opengenus.org/multilayer-perceptrons-vs-cnn/), truy cập 14/02/2024)

Mỗi mô hình trên đều có những ưu và nhược điểm riêng, cũng như các cách tiếp cận khác nhau đối với dữ liệu. Việc lựa chọn mô hình phù hợp phụ thuộc vào đặc thù của bài toán, kích thước và tính chất của dữ liệu.

## **Tuning Model: Điều Chỉnh Siêu Tham Số của Mô hình**

Nhằm tối ưu hóa hiệu suất của model, giảm thiểu báo động giả, thích ứng tốt với dữ liệu cụ thể, tối ưu hóa việc sử dụng tài nguyên, nâng cao độ chính xác và độ tin cậy, đồng thời chuẩn bị cho việc áp dụng các giải pháp và thuật toán mới, chúng ta cần điều chỉnh tham số (tuning) các model trong bài toán Network Intrusion Detection. Quá trình này không chỉ giúp mô hình phát hiện chính xác hơn các loại tấn công mạng trên một tập dữ liệu cho trước và cụ thể mà còn đảm bảo rằng mô hình có khả năng thích ứng và phản ứng hiệu quả với sự thay đổi của môi trường mạng và dữ liệu.

Chúng em áp dụng cả hai phương pháp điều chỉnh siêu tham số là thủ công và sử dụng thư viện Optuna. Optuna được chọn vì khả năng tối ưu hóa hiệu quả, giúp tìm ra bộ siêu tham số tốt nhất cho mô hình một cách nhanh chóng.

Chúng em bắt đầu với việc điều chỉnh siêu tham số cho mô hình Logistic Regression, mô hình đơn giản nhất trong số các mô hình được chọn. Sau đó, tiếp tục sử dụng Optuna để điều chỉnh siêu tham số cho ba mô hình tốt nhất: Random Forest, Decision Tree, và Gradient Boosting, cũng như Logistic Regression. Quá trình này giúp chúng em tìm ra bộ siêu tham số tốt nhất cho từng mô hình, từ đó nâng cao độ chính xác và hiệu suất của mô hình.

Sau khi hoàn thành quá trình điều chỉnh siêu tham số, chúng em tiến hành huấn luyện lại các mô hình với bộ siêu tham số tốt nhất đã được tìm ra. Quá trình này giúp đảm bảo mô hình đạt được hi

## **Đánh giá mô hình**

Đánh giá hiệu suất của mô hình phân loại là một phần không thể thiếu trong quá trình phát triển mô hình. Có nhiều tiêu chí đánh giá khác nhau được sử dụng để đánh giá khả năng của mô hình trong việc phân loại chính xác các mẫu dữ liệu. Trong nghiên cứu này, tôi sử dụng bốn tiêu chí đánh giá chính là:

* Accuracy: Độ Chính xác
* Precision: Độ Chính xác tích cực
* Recall: Độ Nhạy
* F1 Score: Điểm số F1

### ***Accuracy: Độ Chính xác***

Là tỉ lệ của tổng số mẫu được phân loại đúng so với tổng số mẫu. Độ chính xác không chỉ cho biết số lượng mẫu dự đoán chính xác mà còn phản ánh cả hai loại lỗi phân loại, cả dương tính giả (FP) và âm tính giả (FN).

### ***Precision: Độ Chính xác tích cực***

Đo lường tỷ lệ các mẫu dự đoán là tích cực thực sự chính xác so với tổng số mẫu dự đoán là tích cực. Điều này quan trọng khi chi phí của việc dự đoán sai một mẫu là âm tính cao hơn nhiều.

### ***Recall: Độ Nhạy***

Còn được gọi là độ nhạy cảm hoặc tỉ lệ dương tính thực. Nó đo lường tỷ lệ các mẫu thực sự tích cực được phân loại đúng so với tổng số mẫu thực sự tích cực. Điều này quan trọng trong các tình huống mà việc bỏ sót một mẫu tích cực có thể dẫn đến hậu quả nghiêm trọng.

### ***F1 Score: Điểm số F1***

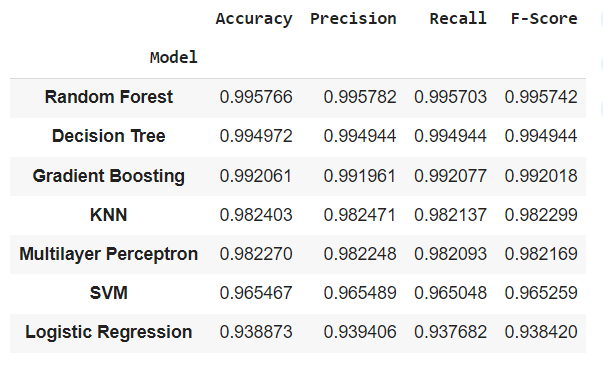
Là trung bình điều hòa của Precision và Recall, nó cung cấp một số đo đơn nhất của hiệu suất mô hình ở cả hai tiêu chí này. Điểm số F1 đặc biệt hữu ích khi ta cần cân bằng giữa Precision và Recall và khi có sự không cân đối giữa các lớp phân loại.

Sử dụng các tiêu chí này giúp tôi đánh giá khả năng tổng quát của mô hình trong việc phân loại cũng như phát hiện các điểm mạnh và điểm yếu cụ thể trong dự đoán của mô hình.

# **THỰC NGHIỆM**

## **Kết quả thực nghiệm**

### ***Trước khi Tuning Model***

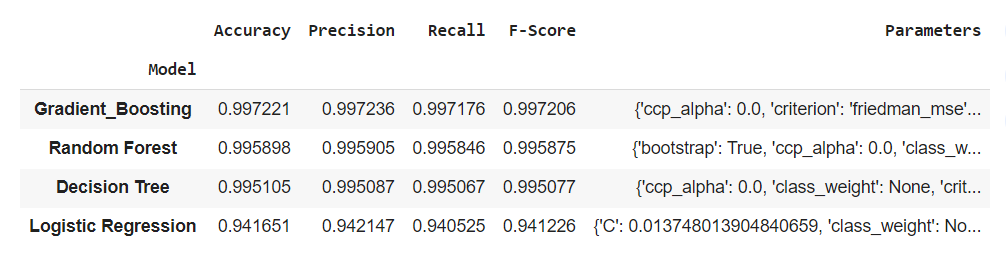


*HÌNH 4.1 Trước khi Tuning Model*

Đánh giá hiệu suất của các mô hình trước khi tinh chỉnh tham số, chúng ta có thể thấy rằng mô hình Random Forest, Decision Tree và Gradient Boosting đã thể hiện sự ưu việt so với các mô hình khác. Điều này có thể được giải thích thông qua cơ sở lý thuyết cũng như cách thức mà chúng xử lý và học hỏi từ dữ liệu.

* Random Forest: hiệu quả cao nhờ vào kỹ thuật ensemble, tức là kết hợp các cây quyết định thông qua bagging để giảm thiểu rủi ro overfitting và tăng khả năng tổng quát hóa. Sự đa dạng trong việc lựa chọn các tập con ngẫu nhiên của dữ liệu cho mỗi cây quyết định cũng góp phần làm tăng độ chính xác trong dự đoán.
* Decision Tree: lại ghi điểm với khả năng phân loại dựa trên các quy tắc rõ ràng và cấu trúc dữ liệu không tuyến tính, điều này giúp mô hình dễ dàng điều chỉnh để phù hợp với đặc thù của dữ liệu cụ thể.
* Gradient Boosting: phát huy sức mạnh từ việc cải thiện dự đoán bằng cách tuần tự sửa lỗi từ các mô hình trước, mỗi lần tinh chỉnh tăng cường mô hình tổng thể một cách có hệ thống và hiệu quả.

### ***Sau khi tiến hành tinh chỉnh tham số:***



*HÌNH 4.2 Sau khi Tuning Model*

* Gradient Boosting: không chỉ cải thiện về mặt độ chính xác mà còn cho thấy sự tăng cường độ nhạy và độ chính xác tích cực, phản ánh qua sự tăng của các chỉ số sau tuning. Sự điều chỉnh cẩn thận các tham số như tỷ lệ học và số lượng cây quyết định đã giúp mô hình này không chỉ cải thiện được khả năng học từ dữ liệu mà còn cải thiện đáng kể khả năng dự đoán chính xác.
* Random Forest sau tuning đã thể hiện được khả năng tổng quát hóa mạnh mẽ, giữ vững hiệu suất cao cả về độ chính xác lẫn độ nhạy, nhờ khả năng kết hợp thông tin từ một lượng lớn cây quyết định và giảm thiểu sự phụ thuộc vào từng đặc trưng cụ thể.
* Decision Tree, mặc dù có cấu trúc đơn giản hơn so với hai mô hình kia nhưng sau khi được tinh chỉnh đã cho thấy khả năng phân loại mạnh mẽ, đặc biệt là khi xử lý dữ liệu phức tạp hoặc không tuyến tính.

Kết quả từ quá trình tinh chỉnh tham số cho thấy sự cần thiết của việc lựa chọn đúng tham số để tối ưu hóa hiệu suất mô hình, đồng thời nâng cao khả năng phát hiện mẫu dữ liệu bất thường, điều rất quan trọng trong việc xây dựng hệ thống phát hiện xâm nhập mạng hiệu quả. Các phát hiện này cung cấp một cái nhìn toàn diện về tiềm năng cũng như thách thức trong việc triển khai các mô hình học máy và học sâu trong lĩnh vực an ninh mạng và phát hiện xâm nhập.

# **KẾT LUẬN**

## **Kết luận**

Trong nghiên cứu này, chúng em đã triển khai và đánh giá một hệ thống Phát hiện Xâm nhập Mạng sử dụng các kỹ thuật học máy tiên tiến. Qua việc thử nghiệm nhiều mô hình học máy khác nhau trên tập dữ liệu NSL-KDD, tôi nhận thấy mô hình Random Forest, Decision Tree và Gradient Boosting đã thể hiện hiệu suất vượt trội so với các mô hình khác. Lý thuyết cũng như kết quả thực nghiệm đã chứng minh khả năng hiệu quả của chúng trong việc phát hiện các gói tin xâm nhập mạng, đặc biệt là sau quá trình tinh chỉnh tham số.

Mô hình Random Forest với kỹ thuật ensemble đã cho thấy khả năng tổng quát hóa mạnh mẽ và giảm rủi ro overfitting. Decision Tree đã thể hiện sức mạnh với cấu trúc dữ liệu không tuyến tính và khả năng điều chỉnh để phù hợp với đặc thù của dữ liệu. Gradient Boosting cải thiện hiệu suất dự đoán bằng cách liên tục sửa lỗi từ các mô hình trước, và sau tinh chỉnh đã nâng cao đáng kể độ chính xác và độ nhạy cảm.

Các kết quả sau tinh chỉnh tham số cho thấy rõ sự cần thiết và hiệu quả của việc lựa chọn đúng tham số để tối ưu hóa hiệu suất mô hình. Sự cải thiện này không chỉ giúp mô hình học sâu hơn từ dữ liệu mà còn nâng cao khả năng dự đoán chính xác và phát hiện mẫu dữ liệu bất thường. Điều này đặc biệt quan trọng trong việc xây dựng hệ thống phát hiện xâm nhập mạng hiệu quả và đáng tin cậy.

Trong tương lai, việc tích hợp thêm dữ liệu từ môi trường mạng thực tế, cũng như tiếp tục cải tiến mô hình thông qua việc áp dụng các phương pháp học sâu và học củng cố, sẽ giúp tôi phát triển hệ thống phát hiện xâm nhập mạng ngày càng thông minh và nhạy bén hơn. Điều này góp phần vào mục tiêu rộng lớn hơn là tạo ra một nền tảng an ninh mạng vững chắc, có thể đáp ứng và thích ứng linh hoạt với những thách thức ngày càng phức tạp và biến đổi không ngừng của lĩnh vực an ninh mạng hiện đại.

.

## **Hướng phát triển**

Nhằm tiếp tục phát triển hệ thống phát hiện xâm nhập mạng (IDS) hiệu quả và chính xác hơn, các hướng nghiên cứu và phát triển trong tương lai bao gồm:

### ***Tích hợp thêm yếu tố và nguồn dữ liệu đa dạng:***

Phát triển hệ thống IDS của chúng em không chỉ dừng lại ở việc sử dụng dữ liệu từ tập NSL-KDD, mà còn hướng đến việc tích hợp và phân tích một loạt dữ liệu đa dạng hơn từ môi trường mạng thực tế. Bao gồm nhưng không giới hạn ở:

- Thu thập dữ liệu phong phú: Nguồn dữ liệu mới bao gồm thông tin từ các hệ thống, ứng dụng, và thiết bị IoT để có được cái nhìn đa chiều hơn về môi trường mạng.

- Phân tích tương quan: Xác định mối liên kết giữa các sự kiện mạng và các biến số khác như thời gian, sự kiện lớn (như các cuộc tấn công DDoS), và các yếu tố ngoại vi khác có thể ảnh hưởng đến an ninh mạng.

### ***Phát triển giao diện người dùng trực quan:***

Với sự phát triển của công nghệ, việc cung cấp một giao diện người dùng trực quan và thân thiện là yếu tố quan trọng để thu hút người dùng:

- Phát triển giao diện người dùng: Thiết kế giao diện đơn giản, dễ sử dụng nhưng đủ mạnh mẽ để thể hiện thông tin phức tạp một cách rõ ràng.

- Tối ưu hóa trải nghiệm người dùng: Kết hợp các tính năng thông minh như dự đoán và cảnh báo tự động, cung cấp cái nhìn toàn cảnh về tình trạng an ninh mạng.

### ***Tối ưu hóa và cập nhật mô hình dựa trên phản hồi thực tế:***

Các mô hình học máy, đặc biệt là các mô hình học sâu, cần được cập nhật và tối ưu hóa liên tục dựa trên phản hồi từ việc triển khai thực tế. Điều này bao gồm:

- Cập nhật mô hình: Thu thập và tích hợp dữ liệu mới để mô hình tiếp tục học hỏi và thích nghi với các mẫu tấn công mới và phức tạp hơn.

- Tối ưu hóa tham số: Dựa trên phân tích hiệu suất thực tế để điều chỉnh các tham số mô hình, cải thiện khả năng phát hiện và giảm báo động giả.

### ***Khám phá các kỹ thuật học máy mới và tiên tiến:***

Khám phá và thử nghiệm với các phương pháp học máy và học sâu tiên tiến, như học tăng cường và mô hình học chuyển giao, để phát hiện xâm nhập mạng chính xác và hiệu quả hơn.

Mục tiêu của tôi là tạo ra một hệ thống IDS không chỉ có khả năng thích ứng với các tình huống hiện tại, mà còn có khả năng dự đoán và phản ứng trước các mối đe dọa trong tương lai. Sự cải tiến không ngừng này sẽ góp phần xây dựng một nền tảng an ninh mạng mạnh mẽ, giúp bảo vệ cơ sở hạ tầng thông tin quan trọng trong kỷ nguyên kỹ thuật số.

# **TÀI LIỆU THAM KHẢO**

Tiếng Việt

…

Tiếng Anh

[1] IBM Cloud Education. Supervised Learning. Available online: <https://www.ibm.com/cloud/learn/supervised-learning> (accessed on 27 Dec 2023).

[2] <https://www.mdpi.com/1999-5903/15/7/243#B2-futureinternet-15-00243>

[3] <https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/>

[4] <https://www.ibm.com/topics/intrusion-detection-system>

[5] <https://securopia.wordpress.com/2011/08/25/security-models-cia-and-ciaan/>

[6] <https://www.techtarget.com/searchsecurity/definition/firewall>