HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

KHOA VIỄN THÔNG 1

BÁO CÁO BÀI TẬP LỚN

MÔN: AN TOÀN MẠNG THÔNG TIN

ĐỀ TÀI: Ứng dụng AI/ML trong phát hiện xâm nhập mạng

Giảng viên hướng dẫn: Nguyễn Thanh Trà

NHÓM: 7

Trần Duy Lăng	B20DCVT224
Nguyễn Đăng Khải	B20DCVT212
Phạm Văn Hưng	B20DCVT198
Nguyễn Bá Quốc Khánh	B20DCVT214

Hà Nội

PHÂN CHIA CÔNG VIỆC

Họ và tên	Nhiệm vụ
Nguyễn Đăng Khải	Chương 1: phần 1+2
Nguyễn Bá Quốc Khánh	Chương 1: phần 3+4
Trần Duy Lăng	Chương 2: phần 1+2
Phạm Văn Hưng	Chương 2: phần 3+4

MỤC LỤC

LỜI MỞ ĐẦU 4
DANH MỤC HÌNH VẼ 5
DANH MỤC BẢNG BIỂU 5
THUẬT NGỮ VIẾT TẮT 6
Chương I: Kiến thức chung 7
1.Giới thiệu về đề tài 7
1.1 Lý do chọn đề tài 7
1.2 Mục tiêu nghiên cứu 8
1.3 Tầm quan trọng của đề tài 9
2.Nguy cơ xâm nhập mạng 10
2.1 Xâm nhập mạng là gì? 10
2.2 Những hậu quả của xâm nhập mạng trái phép 11
2.3 Những phương pháp phòng chống xâm nhập mạng 11
3.Giới thiệu về trí tuệ nhân tạo học máy 12
3.1 Trí tuệ nhân tạo (AI) và máy học (ML) 12
3.2 Bài toán phân loại trong ML 15
4.Nghiên cứu và xây dựng mô hình AI/ML 16
4.1 Thu thập và tiền xử lý dữ liệu 16
4.2 Sử dụng những mô hình đào tạo mô hình 17
4.3 Tối ưu và cải thiện mô hình 18
Chương II: Xây dựng mô hình phát hiện xâm nhập mạng dựa trên bộ dữ liệu NSL-KDD 20
1. Giới thiệu bộ dữ liệu NSL-KDD 20
1.1 Nguồn gốc bộ dữ liệu 20
1.2 Các trường dữ liệu trong NSL-KDD 20
1.3 Các kiểu tấn công có trong tập dữ liệu 22
2. Xây dựng mô hình máy học 25
2.1 Tiền xử lý dữ liệu 26
3. Xây dựng mô hình 28

3.1 Phân chia dữ liệu 28
3.2 Thử nghiệm với các mô hình 28
KẾT LUẬN 35
Tài liệu tham khảo 36

LỜI MỞ ĐẦU

Trong thời đại số hóa và thời nguyên internet, mạng máy tính đã trở thành một phần quan trọng của cuộc sống cá nhân và hoạt động của doanh nghiệp. Tuy nhiên, sự phổ biến và phức tạp của mạng cũng đã mang theo mối đe dọa mới và nguy cơ mà ngày càng trở nên nguy hiểm: xâm nhập mạng trái phép.

Xâm nhập mạng trái phép là xâm nhập vào bảo mật và dữ liệu riêng của mạng hệ thống và dữ liệu. Đây là một hoạt động bất hợp pháp mà kẻ xâm nhập thực hiện để truy cập, thay đổi hoặc đánh cắp thông tin quý giá. Điều này gây ra hậu quả nghiêm trọng cho cá nhân, tổ chức và doanh nghiệp, không chỉ về mặt tài chính mà còn về mặt danh tiếng và sự kiện riêng.

Trong bối cảnh này, sự phát triển nhanh chóng của Trí tuệ Nhân tạo (AI) và Học Máy (ML) đã mang lại những cơ hội mới để phát hiện và ngăn chặn trái phép xâm nhập mạng. Giải pháp dựa trên AI/ML có khả năng phân tích mạng dữ liệu hàng triệu một cách nhanh chóng và tự động, xác định các hoạt động bất thường và nguy cơ tiềm ẩn một cách hiệu quả hơn.

Đề xuất tập tài liệu này về cách khám phá mà AI và ML có thể được áp dụng trong trái phép phát hiện và ngăn chặn xâm nhập mạng. Chúng tôi sẽ xem xét các phương pháp, kỹ thuật và công cụ cụ thể mà Trí tuệ Nhân tạo và Học Máy có thể cung cấp để tăng cường bảo vệ mạng một cách hiệu quả hơn.

Đề tài này không chỉ mang tính nghiên cứu mà còn hướng dẫn ứng dụng thực tế, với hy vọng rằng những nghiên cứu và phát triển trong lĩnh vực này sẽ đóng góp vào việc xây dựng một môi trường mạng an toàn và đáng tin cậy hơn .

DANH MỤC HÌNH VẼ

Hình 1.1.Các kiểu tấn công trong NSL-KDD 23
Hình 1.2 Phân bổ các loại tấn công trong NSL-KDD 25
Hình 2.1 Phân bố thông tin trong trường “attack” 26
Hình 2.2 Kết quả quá trình học của SVM 29
Hình 2.3 Kết quả quá trình học của KNN 30
Hình 2.4 Kết quả quá trình học của Logistic Regresstion.....31

DANH MỤC BẢNG BIỂU

Bảng 1. Các trường dữ liệu trong NSL-KDD.....22

THUẬT NGỮ VIẾT TẮT

Tên viết tắt	Ý nghĩa
AI	Artificial Intelligence
ML	Machine Learning
DoS	Denial of Service
DDoS	Distributed Denial of Service
U2R	User to root
U2L	Remote to Local
SVM	Support Vector Machine
IMCP	Internet Control Message Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol

Chương I: Kiến thức chung

1.Giới thiệu về đề tài

1.1Lý do chọn đề tài

Trong thời đại hóa ngày càng phát triển, mạng thông tin trở thành một phần quan trọng trong cuộc sống và hoạt động của chúng ta. Tuy nhiên, với sự gia tăng đáng kể của các cuộc tấn công mạng, việc bảo vệ thông tin và dữ liệu cá nhân đã trở thành một công thức ngày càng lớn. Vì thế, nhóm chúng tôi đã chọn đề tài "Ứng dụng Trí tuệ Nhân tạo (AI) và Học Máy (ML) trong Phát hiện Xâm nhập Mạng" giúp nâng cao bảo mật mạng và bảo vệ thông tin cá nhân bởi một số lý do sau:

Một trong những lý do hàng đầu cho việc đơn giản chọn đề tài này là sự gia tăng đáng kể của các cuộc tấn công công mạng. Trong một thời đại mà mọi thứ đều kết nối mạng, các tổ chức và cá nhân trở nên ngày càng phụ thuộc vào mạng internet. Điều này đã tạo ra một môi trường có lợi cho các hành vi xâm nhập và tấn công mạng. Những cuộc tấn công này không chỉ gây tổn hại đến tài chính mà còn có thể đe dọa sự riêng tư và an toàn của thông tin quan trọng. Do đó, việc phát hiện và ngăn chặn xâm nhập mạng trở thành ưu tiên hàng đầu.

Khả năng phân loại dữ liệu lớn và phát triển trí tuệ nhân tạo và học máy là một lý do quan trọng khác. AI và ML đã được xử lý trong quá trình xử lý và phân tích dữ liệu lớn. Thuật toán và mô hình hóa có khả năng tự động học từ dữ liệu đã giúp tạo ra các kết quả giải pháp hiệu quả để phát hiện mạng xâm nhập xâm nhập. Chúng tôi có khả năng phát hiện các mẫu không bình thường và các hoạt động đáng ngạc nhiên trong mạng dữ liệu, ngay cả khi chúng thay đổi theo thời gian.

Hơn nữa, việc ứng dụng AI và ML trong lĩnh vực này hứa hẹn mang lại sự hoạt động và khả năng thích nghi với môi trường mạng thay đổi liên tục. Hệ thống phát hiện truyền tải xâm nhập thông thường cần được cập nhật thường xuyên để hỗ trợ cho các cuộc tấn công mới. Trong thời gian đó, các mô hình AI/ML có khả năng tự động học từ dữ liệu mới và bản cập nhật mà không cần đến khả năng nhận diện của con người. Điều này giúp giảm bớt thời gian và công sức cần thiết để duy trì và nâng cấp hệ thống.

Điểm mạnh của AI/ML cũng có khả năng giảm thiểu các dạng sai lệch. AI/ML có tiềm năng giúp giảm thiểu số lượng báo động sai sót, giúp người quản lý mạng quản trị tập trung vào các đợt tấn công quan trọng và thực sự đáng lo ngại.

Cuối cùng, lựa chọn chủ đề tài năng này được thúc đẩy bởi sự phát triển nhanh chóng của lĩnh vực AI/ML. Có sẵn nhiều công cụ và thư viện hỗ trợ để phát triển các mô hình phát hiện chiến dịch dựa trên AI/ML, giúp công việc nghiên cứu và phát triển trở nên thú vị và hiệu quả tiềm năng.

Như vậy, lý do chọn đề tài "Ứng dụng Trí tuệ Nhân tạo (AI) và Học Máy (ML) trong Phát triển Xâm nhập Mạng" không chỉ là về việc nghiên cứu và phát triển một giải pháp kỹ thuật kỹ thuật , mà còn đóng góp công việc bảo vệ thông tin và dữ liệu quan trọng của mọi người trong thế giới kỹ thuật số ngày nay.

1.2 Mục tiêu nghiên cứu

Trong cuộc cách mạng hóa đầy đủ công thức của thế giới ngày nay, việc bảo vệ mạng thông tin sẽ trở nên quan trọng hơn bao giờ hết. Sự phát triển nhanh chóng của công nghệ đã mở cánh cửa cho sự tiến bộ nhưng cũng tạo ra những cơ hội mới cho các tên cướp mạng tinh vi.

Để đối phó với những hành vi xâm nhập trái phép ảnh hưởng tới quyền lợi của mọi người, chúng tôi đã đưa ra một giải pháp hiệu quả cho việc phát hiện chiến lược xâm nhập, sử dụng sức mạnh của trí tuệ nhân tạo và học máy. Chúng tôi tin rằng việc kết hợp AI/ML với lĩnh vực bảo mật mạng có thể mang lại những đột phá đáng kể trong công việc bảo vệ thông tin và dữ liệu của cá nhân, tổ chức và doanh nghiệp.

Việc tích hợp các AI/ML mô hình vào thực tế mạng hệ thống cũng là mục tiêu quan trọng. Chúng tôi mong muốn tạo ra giải pháp có thể hoạt động hiệu quả trong môi trường mạng thực tế của các tổ chức và doanh nghiệp. Nghiên cứu này cũng tập trung vào việc nghiên cứu giảm thiểu số lượng những động báo sót lại, giúp người quản lý mạng quản trị tập trung vào các tấn công quan trọng hơn.

Để đạt được mục tiêu này, chúng tôi sẽ tiến hành nghiên cứu chuyên sâu về trí tuệ nhân tạo và học máy. Chúng tôi sẽ xem xét các thuật toán và mô hình học máy phù hợp nhất để tìm ra công việc phát hiện xâm nhập và tối ưu hóa chúng để đảm bảo hiệu quả tốt nhất.

Việc “ứng dụng Trí tuệ Nhân tạo (AI) và Học Máy (ML) trong Phát triển Xâm nhập Mạng" đánh dấu một bước tiến trình quan trọng trong việc bảo vệ mạng thông tin và dữ liệu của chúng khỏi các mạng cuộc tấn công mạng ngày càng tinh vi. Nghiên cứu này hứa hẹn mang lại nghiên cứu lợi ích cho cuộc sống số và bảo mật thông tin của chúng ta.

1.3 Tầm quan trọng của đề tài

Trong thời hiện đại nhanh chóng của mạng thông tin và kết nối mạng, việc đảm bảo bảo mật và an toàn cho dữ liệu trở thành một trong những công thức quan trọng nhất. Sự phát triển của công nghệ thông tin đã mang lại nhiều tiện ích và cơ hội, nhưng cũng đồng nghĩa với sự gia tăng của các cuộc tấn công mạng ngày càng phức tạp và phức tạp. Vậy nên "Ứng dụng Trí tuệ Nhân tạo (AI) và Học Máy (ML) trong Phát hiện Xâm nhập Mạng" trở nên vô cùng quan trọng và cần thiết.

Một trong những khía cạnh quan trọng nhất của vấn đề tài nghiên cứu này chính là công việc bảo vệ thông tin cá nhân và doanh nghiệp. Thông tin cá nhân của mọi người và dữ liệu doanh nghiệp là tài sản quý giá và đôi khi cực kỳ nhạy cảm. Sự xâm nhập vào mạng hệ thống có thể gây ra hậu quả nghiêm trọng, từ việc tiết lộ thông tin cá nhân đến việc thoát khỏi tài chính và danh tiếng thất bại. Đề tài này giúp bảo vệ những tài sản quý giá này khỏi những cuộc tấn công phá hoại.

Một khía cạnh khác của tầm quan trọng của tài liệu này là sự thay đổi liên tục của cách thức tấn công mạng. Hacker và kẻ tấn công luôn phát triển các chiến thuật mới để tránh bị phát hiện. Bằng cách sử dụng trí tuệ nhân tạo và máy học, chúng tôi có khả năng tạo ra các xâm nhập hệ thống có khả năng phát hiện các công thức tấn công mẫu mới và không rõ ràng. Điều này làm tăng khả năng chống lại các cuộc tấn công ngày càng phức tạp.

Đề tài này cũng quan trọng vì nó giúp giảm thiểu số lượng báo động sai sót. Hệ thống phát hiện xâm nhập xâm nhập thường xuyên tạo ra nhiều cảnh báo không cần thiết, mất thời gian và tạo ra hoang mang. Bằng cách sử dụng trí tuệ nhân tạo và học máy, chúng tôi có thể cải thiện độ chính xác của hệ thống này và giảm thiểu sai sót cảnh báo, giúp người quản lý mạng quản trị tập trung vào những trường hợp lý thực sự quan trọng .

Cuối cùng, đề tài này đánh dấu một bước tiến quan trọng trong quá trình phát triển trí tuệ nhân tạo và học máy. Việc sử dụng AI và ML trong việc phát hiện mạng xâm nhập không chỉ giúp tăng cường bảo mật thông tin mạng mà còn đóng góp cho việc phát triển các công nghệ tiên tiến. Điều này có tầm quan trọng lớn trong việc tạo ra một môi trường mạng thông tin an toàn và bảo mật hơn cho tương lai.

"Ứng dụng Trí tuệ Nhân tạo (AI) và Học Máy (ML) trong Phát triển Xâm nhập Mạng" không chỉ đơn thuần là một nghiên cứu mà còn là một sứ mệnh quan trọng để bảo vệ thông tin cá nhân nhân và doanh nghiệp từ các cuộc tấn công mạng ngày càng tinh vi. Nó đóng góp vào việc xây dựng một tương lai an toàn và bảo mật trong thế giới hóa học ngày nay.

2.Nguy cơ xâm nhập mạng

2.1 Xâm nhập mạng là gì?

Hiện nay, thời đại 4.0 công nghệ lên ngôi, mạng thông tin và kết nối mạng đóng vai trò quan trọng không chỉ trong công việc tiếp theo và truy cập thông tin mà còn trong hoạt động kinh doanh, giáo dục, y tế và nhiều khía cạnh khác của cuộc sống. Tuy nhiên, với những lợi ích của mạng thông tin này cũng đi kèm với những công thức và nguy cơ cơ bản, đặc biệt là liên quan đến vấn đề bảo mật và an toàn. Vì thế, chúng ta cần tìm hiểu xâm nhập mạng là gì?

Xâm nhập mạng (Network Intrusion) là hành động phi pháp của một cá nhân hoặc tổ chức nhằm vào một hệ thống máy tính, website, cơ sở dữ liệu, hạ tầng mạng, thiết bị của một cá nhân hoặc tổ chức thông qua mạng internet với những mục đích bất hợp pháp.

Mục tiêu của một cuộc xâm nhập mạng rất đa dạng, có thể là vi phạm dữ liệu (đánh cắp, thay đổi, mã hóa, phá hủy), cũng có thể nhắm tới sự toàn vẹn của hệ thống (gây gián đoạn, cản trở dịch vụ), hoặc lợi dụng tài nguyên của nạn nhân (hiển thị quảng cáo, mã độc đào tiền ảo). Ví dụ: Năm 2018, website của ngân hàng Vietcombank đã bị tấn công. Khi đó tuy chưa có tổn thất nhưng cũng phản ánh với các lỏng lẻo đối với phương thức bảo mật của một ngân hàng lớn. Khi Hacker để lại hai câu thơ chế "Trăm năm Kiều vẫn là Kiều/ Sinh viên thi lại là điều tất nhiên" trên chính website của ngân hàng này. Điều này mang đến các phản ánh đối với an ninh mạng chưa được đảm bảo.

Các hình thức xâm nhập mạng phổ biến hiện nay:

- Dùng các phần mềm độc hại (malware attack): là hình thức phổ biến nhất, tin tặc sẽ dùng các phần mềm độc hại để thông qua các lỗ hổng bảo mật, cũng có thể là dụ dỗ người dùng click vào một đường link hoặc email để phần mềm độc hại tự động cài đặt vào máy tính. Nó sẽ gây ra:
  - Ngăn cản người dùng truy cập vào một file hoặc folder quan trọng
  - Cài đặt thêm những phần mềm độc hại khác
  - Lén lút theo dõi người dùng và đánh cắp dữ liệu
  - Làm hư hại phần mềm, phần cứng, làm gián đoạn hệ thống.
- Tấn công giả mạo (phishing attack): là hình thức giả mạo thành một đơn vị/cá nhân uy tín để chiếm lòng tin của người dùng, thông thường qua email. Mục đích của tấn công Phishing thường là đánh cắp dữ liệu nhạy cảm như thông tin thẻ tín dụng, mật khẩu, đôi khi phishing là một hình thức để lừa người dùng cài đặt malware vào thiết bị (khi đó, phishing là một công đoạn trong cuộc tấn công malware).
- Tấn công từ chối dịch vụ (DoS và DdoS):
  - DoS (Denial of Service) là hình thức tấn công mà tin tặc "đánh sập tạm thời" một hệ thống, máy chủ, hoặc mạng nội bộ -> khiến cho hệ thống bị quá tải, từ đó người

dùng không thể truy cập vào dịch vụ trong khoảng thời gian mà cuộc tấn công DoS diễn ra.

- DDoS (Distributed Denial of Service): tin tặc sử dụng một mạng lưới các máy tính (botnet) để tấn công nạn nhân. Điều nguy hiểm là nạn nhân không hề hay biết bản thân đang bị lợi dụng để làm công cụ tấn công

➢ Tấn công trung gian (Man-in-the-middle attack): Loại hình này xảy ra khi:

- Nạn nhân truy cập vào một mạng Wifi công cộng không an toàn, kẻ tấn công có thể "chen vào giữa" thiết bị của nạn nhân và mạng Wifi đó. Vô tình, những thông tin nạn nhân gửi đi sẽ rơi vào tay kẻ tấn công.
- Khi phần mềm độc hại được cài đặt thành công vào thiết bị, một kẻ tấn công có thể dễ dàng xem và điều chỉnh dữ liệu của nạn nhân.

2.2 Những hậu quả của xâm nhập mạng trái phép

Các cuộc tấn công xâm nhập vào mạng không chỉ là mối đe dọa lớn nhất đối với cá nhân, tổ chức và doanh nghiệp trên khắp thế giới, mà còn vi phạm vào hệ thống và dữ liệu mà còn là một cuộc tấn công vào tính riêng tư và toàn vẹn của mỗi người.

Xâm nhập mạng trái phép gây ra nhiều hậu quả nghiêm trọng:

- Mất dữ liệu: chúng có thể đánh cắp hoặc xóa dữ liệu quý giá, bao gồm tài liệu công việc, thông tin khách hàng và thông tin quản lý.
- Thiệt hại về tài chính: Cuộc tấn công có thể dẫn đến mất tiền bạc thông qua trái phép giao dịch, lừa đảo tài khoản ngân hàng.
- Ảnh hưởng tới hoạt động kinh doanh: Tấn công từ chối dịch vụ (DDoS) có thể khiến dịch vụ mạng trở nên không hoạt động, gây rối loạn rối loạn trong việc cung cấp dịch vụ và mất khách hàng.
- Nguy cơ lây lan các phần mềm độc hại.
- Vi phạm pháp luật

Để ngăn chặn xâm nhập và giảm thiểu hậu quả, cần phát triển các biện pháp bảo mật mạng mạnh mẽ, theo dõi và đào tạo nhân viên về nguy cơ bảo mật và duy trì kế hoạch phục hồi dữ liệu.

2.3 Những phương pháp phòng chống xâm nhập mạng

Phòng chống xâm nhập mạng là một nhiệm vụ quan trọng đòi hỏi cần có cải tiến và phát triển các phương thức bảo mật hiện đại, tiên tiến để có thể chống lại những mối đe dọa ngày càng tinh vi và phức tạp.

Một số phương pháp bảo mật để phòng chống xâm nhập mạng:

- Mạng tường lửa : Sử dụng mạng tường lửa để kiểm soát việc lưu trữ mạng và quyết định cái nào được phép và cái nào bị từ chối. Tường lửa đóng vai trò là một biện pháp bảo vệ tường lửa giữa nội bộ mạng và Internet.

- Sử dụng một phần mềm diệt Virus uy tín: giúp tiêu diệt các phần mềm độc hại.
- Bảo vệ mật khẩu cá nhân bằng cách: đặt mật khẩu phức tạp, bật tính năng bảo mật 2 lớp – xác nhận qua điện thoại,…
- Tuyệt đối không tải các file hoặc nhấp vào đường link không rõ nguồn gốc vì rất dễ nguy cơ lây nhiễm các phần mềm độc hại.
- Luôn cập nhật phần mềm, firmware lên phiên bản mới nhất
- Sử dụng các dịch vụ đám mây uy tín cho mục đích lưu trữ
- Lựa chọn các phần mềm, đối tác một cách kỹ càng. Ưu tiên những bên có cam kết bảo mật và cam kết cập nhật bảo mật thường xuyên.
- Đánh giá bảo mật & Xây dựng một chiến lược an ninh mạng tổng thể cho doanh nghiệp, bao gồm các thành phần: bảo mật website, bảo mật hệ thống máy chủ, mạng nội bộ, hệ thống quan hệ khách hàng (CRM), bảo mật IoT, bảo mật hệ thống CNTT – vận hành…
- Tổ chức các buổi đào tạo, training kiến thức sử dụng internet an toàn cho nhân viên

3.Giới thiệu về trí tuệ nhân tạo học máy

3.1 Trí tuệ nhân tạo (AI) và máy học (ML)

Trí tuệ nhân tạo (AI)

Trí tuệ nhân tạo (AI) là một ngành thuộc lĩnh vực khoa học máy tính, chuyên nghiên cứu và phát triển các ứng dụng máy tính có thể thực hiện các chức năng thông minh như con người. AI được ứng dụng trong nhiều lĩnh vực của đời sống, từ sản xuất, kinh doanh, dịch vụ, đến y tế, giáo dục, giải trí,...

Trí tuệ nhân tạo có thể chia thành nhiều lĩnh vực nghiên cứu chính, bao gồm:

- Học máy (machine learning): là lĩnh vực nghiên cứu về các thuật toán cho phép máy tính tự học từ dữ liệu. Học máy là nền tảng cho nhiều ứng dụng AI hiện đại, như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên,...
- Xử lý ngôn ngữ tự nhiên (natural language processing): là lĩnh vực nghiên cứu về việc giao tiếp giữa con người và máy tính bằng ngôn ngữ tự nhiên. Xử lý ngôn ngữ tự nhiên được ứng dụng trong nhiều lĩnh vực, như dịch máy, nhận dạng giọng nói, chatbot,...
- Trí tuệ nhận thức (cognitive intelligence): là lĩnh vực nghiên cứu về các mô hình trí tuệ tương tự như trí tuệ của con người, bao gồm nhận thức, suy luận, giải quyết vấn đề,... Trí tuệ nhận thức được ứng dụng trong các hệ thống tự động hóa phức tạp, như xe tự lái, robot,...

- Trí tuệ tổng hợp (artificial general intelligence): là lĩnh vực nghiên cứu về các hệ thống AI có khả năng thực hiện nhiều nhiệm vụ thông minh khác nhau, giống như con người. Trí tuệ tổng hợp là mục tiêu cuối cùng của nghiên cứu AI.

Trong những năm gần đây, AI đã có những bước tiến vượt bậc, được ứng dụng rộng rãi trong nhiều lĩnh vực của đời sống. AI đang có tiềm năng tác động to lớn đến tương lai của nhân loại, từ cách chúng ta làm việc, học tập, giải trí, đến cách chúng ta tương tác với thế giới xung quanh.

Dưới đây là một số ví dụ về ứng dụng của AI trong thực tế:

- Trong sản xuất: AI được ứng dụng để tự động hóa các quy trình sản xuất, tối ưu hóa hiệu quả sản xuất, và nâng cao chất lượng sản phẩm.
- Trong kinh doanh: AI được ứng dụng để phân tích dữ liệu, tối ưu hóa quy trình kinh doanh, và cung cấp các trải nghiệm khách hàng cá nhân hóa.
- Trong dịch vụ: AI được ứng dụng để cung cấp dịch vụ khách hàng 24/7, tự động hóa các tác vụ hành chính, và cá nhân hóa trải nghiệm khách hàng.
- Trong y tế: AI được ứng dụng để chẩn đoán bệnh, điều trị y tế, và phát triển các loại thuốc mới.
- Trong giáo dục: AI được ứng dụng để cá nhân hóa việc học, cung cấp các bài học trực tuyến, và đánh giá học tập.
- Trong giải trí: AI được ứng dụng để tạo ra các hình ảnh, âm thanh, và video ảo.

AI là một công nghệ mang tính đột phá, có tiềm năng thay đổi thế giới theo nhiều cách. Tuy nhiên, AI cũng có những tiềm năng rủi ro, như mất việc làm, phân biệt đối xử, và sai sót hệ thống. Do đó, cần có sự phát triển AI một cách có trách nhiệm và bền vững, đảm bảo lợi ích của con người.

Máy Học (ML)

Máy học (ML) là một lĩnh vực của trí tuệ nhân tạo (AI), chuyên nghiên cứu và phát triển các thuật toán cho phép máy tính tự học từ dữ liệu. Máy học là nền tảng cho nhiều ứng dụng AI hiện đại, như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên,...

Máy học khác với lập trình truyền thống ở chỗ nó không yêu cầu lập trình cụ thể cho từng nhiệm vụ. Thay vào đó, máy học sử dụng các thuật toán để học từ dữ liệu và tự động tìm ra các quy tắc để thực hiện các nhiệm vụ.

Có hai loại chính của máy học:

- Học có giám sát: trong học có giám sát, máy tính được cung cấp một tập dữ liệu có nhãn, trong đó mỗi mẫu dữ liệu được gắn nhãn với một kết quả mong muốn. Máy tính sử dụng tập dữ liệu này để học các quy tắc để dự đoán kết quả cho các mẫu dữ liệu mới.
- Học không giám sát: trong học không giám sát, máy tính không được cung cấp tập dữ liệu có nhãn. Thay vào đó, máy tính phải tự tìm ra các quy tắc để phân nhóm các mẫu dữ liệu hoặc xác định các mẫu trong dữ liệu.

Máy học được ứng dụng trong nhiều lĩnh vực của đời sống, bao gồm:

- Sản xuất: máy học được sử dụng để tự động hóa các quy trình sản xuất, tối ưu hóa hiệu quả sản xuất, và nâng cao chất lượng sản phẩm.
- Kinh doanh: máy học được sử dụng để phân tích dữ liệu, tối ưu hóa quy trình kinh doanh, và cung cấp các trải nghiệm khách hàng cá nhân hóa.
- Dịch vụ: máy học được sử dụng để cung cấp dịch vụ khách hàng 24/7, tự động hóa các tác vụ hành chính, và cá nhân hóa trải nghiệm khách hàng.
- Y tế: máy học được sử dụng để chẩn đoán bệnh, điều trị y tế, và phát triển các loại thuốc mới.
- Giáo dục: máy học được sử dụng để cá nhân hóa việc học, cung cấp các bài học trực tuyến, và đánh giá học tập.
- Giải trí: máy học được sử dụng để tạo ra các hình ảnh, âm thanh, và video ảo.

Máy học là một công nghệ mang tính đột phá, có tiềm năng thay đổi thế giới theo nhiều cách. Tuy nhiên, máy học cũng có những tiềm năng rủi ro, như mất việc làm, phân biệt đối xử, và sai sót hệ thống. Do đó, cần có sự phát triển máy học một cách có trách nhiệm và bền vững, đảm bảo lợi ích của con người.

Máy học là một lĩnh vực đang phát triển nhanh chóng, với những ứng dụng ngày càng rộng rãi trong nhiều lĩnh vực của đời sống.

3.2 Bài toán phân loại trong ML

Bài toán phân loại trong ML là một loại bài toán học máy có giám sát, trong đó mục tiêu là phân loại dữ liệu đầu vào thành một trong hai hoặc nhiều lớp.

Ví dụ, một bài toán phân loại có thể là xác định xem một email là spam hay không, hoặc xác định xem một bệnh nhân mắc bệnh tiểu đường hay không.

Để giải quyết một bài toán phân loại, chúng ta cần một tập dữ liệu có nhãn, trong đó mỗi mẫu dữ liệu được gắn nhãn với một lớp. Tập dữ liệu này được sử dụng để đào tạo mô hình phân loại.

Có nhiều thuật toán phân loại khác nhau, chẳng hạn như:

- Hồi quy logistic: Đây là một thuật toán phân loại phổ biến, được sử dụng để phân loại dữ liệu nhị phân.
- Máy quyết định: Đây là một thuật toán phân loại dựa trên cây quyết định, có thể được sử dụng để phân loại dữ liệu với nhiều lớp.
- Nhóm hỗ trợ vector: Đây là một thuật toán phân loại hiệu quả, được sử dụng để phân loại dữ liệu với nhiều lớp.

Sau khi đào tạo mô hình phân loại, chúng ta có thể sử dụng mô hình này để dự đoán lớp của dữ liệu đầu vào mới.

Dưới đây là một số ví dụ về ứng dụng của bài toán phân loại trong ML:

- Trong kinh doanh: bài toán phân loại được sử dụng để xác định các khách hàng tiềm năng, phân loại các khách hàng theo nhóm, và phát hiện gian lận.
- Trong y tế: bài toán phân loại được sử dụng để chẩn đoán bệnh, phát hiện sớm bệnh, và nghiên cứu các bệnh mới.
- Trong an ninh: bài toán phân loại được sử dụng để phát hiện tội phạm, phát hiện khủng bố, và giám sát các khu vực công cộng.

Bài toán phân loại là một bài toán quan trọng trong ML, với nhiều ứng dụng trong thực tế.

4.Nghiên cứu và xây dựng mô hình AI/ML

4.1 Thu thập và tiền xử lý dữ liệu

Thu thập và tiền xử lý dữ liệu là hai bước quan trọng trong quá trình phát triển hệ thống AI/ML. Dữ liệu là nền tảng cho mọi hệ thống AI/ML, và chất lượng của dữ liệu sẽ quyết định chất lượng của hệ thống.

Dữ liệu có thể thu thập từ nhiều nguồn khác nhau, bao gồm:

- Nguồn nội bộ: Đây là dữ liệu được tạo ra bởi chính hệ thống AI/ML, chẳng hạn như dữ liệu từ các cảm biến, dữ liệu từ các hệ thống khác, hoặc dữ liệu từ người dùng.
- Nguồn bên ngoài: Đây là dữ liệu được thu thập từ các nguồn bên ngoài, chẳng hạn như dữ liệu từ các trang web, dữ liệu từ các cơ sở dữ liệu, hoặc dữ liệu từ các nhà cung cấp dữ liệu.

Khi thu thập dữ liệu, cần lưu ý các vấn đề sau:

- Số lượng dữ liệu: Số lượng dữ liệu càng lớn, hệ thống AI/ML sẽ càng chính xác.
- Độ đa dạng của dữ liệu: Dữ liệu cần đa dạng để phản ánh thực tế.
- Chất lượng của dữ liệu: Dữ liệu cần chính xác và không có lỗi.

Tiền xử lý dữ liệu

Tiền xử lý dữ liệu là quá trình chuẩn hóa dữ liệu để phù hợp với mô hình AI/ML. Các công việc tiền xử lý dữ liệu bao gồm:

- Loại bỏ dữ liệu lỗi: Xóa các mẫu dữ liệu có lỗi hoặc không phù hợp.
- Làm sạch dữ liệu: Xóa các lỗi ngớ ngẩn trong dữ liệu, chẳng hạn như các lỗi chính tả hoặc các lỗi về định dạng.
- Chuẩn hóa dữ liệu: Chuyển đổi dữ liệu thành định dạng phù hợp với mô hình AI/ML.
- Tính toán các đặc trưng: Tạo các đặc trưng mới từ dữ liệu đầu vào để giúp mô hình AI/ML học tập tốt hơn.

Nghiên cứu và xây dựng mô hình AI/ML

Sau khi thu thập và tiền xử lý dữ liệu, chúng ta có thể bắt đầu nghiên cứu và xây dựng mô hình AI/ML. Quá trình này bao gồm các bước sau:

- Chọn mô hình AI/ML: Có nhiều mô hình AI/ML khác nhau, mỗi mô hình có ưu và nhược điểm riêng. Cần chọn mô hình phù hợp với bài toán cụ thể.
- Chọn thuật toán tối ưu: Có nhiều thuật toán tối ưu khác nhau, mỗi thuật toán có hiệu quả khác nhau đối với các mô hình AI/ML khác nhau. Cần chọn thuật toán tối ưu phù hợp với mô hình AI/ML đã chọn.
- Huấn luyện mô hình: Huấn luyện mô hình AI/ML bằng cách sử dụng tập dữ liệu đã thu thập và tiền xử lý.
- Đánh giá mô hình: Đánh giá hiệu quả của mô hình AI/ML bằng cách sử dụng tập dữ liệu đánh giá.

Sau khi xây dựng mô hình AI/ML, chúng ta cần triển khai mô hình này để đưa vào sử dụng. Trong quá trình triển khai, cần lưu ý các vấn đề sau:

- Kiểm tra mô hình trên dữ liệu mới: Kiểm tra xem mô hình có hoạt động tốt trên dữ liệu mới hay không.
- Cập nhật mô hình: Theo dõi hiệu quả của mô hình và cập nhật mô hình khi cần thiết.

Thu thập và tiền xử lý dữ liệu, nghiên cứu và xây dựng mô hình AI/ML là những công việc quan trọng trong quá trình phát triển hệ thống AI/ML. Các công việc này đòi hỏi kiến thức và kỹ năng chuyên môn, và cần được thực hiện cẩn thận để đảm bảo chất lượng của hệ thống AI/ML.

4.2 Sử dụng những mô hình đào tạo mô hình

Các mô hình đào tạo mô hình (ML) được sử dụng trong nghiên cứu và xây dựng mô hình AI/ML để giúp các nhà nghiên cứu và nhà phát triển học hỏi từ dữ liệu và tạo ra các mô hình mới. Các mô hình này có thể được sử dụng để tự động hóa các tác vụ, chẳng hạn như thu thập và tiền xử lý dữ liệu, lựa chọn mô hình và thuật toán, và đánh giá hiệu quả của mô hình.

Có nhiều loại mô hình đào tạo mô hình khác nhau, mỗi loại có ưu và nhược điểm riêng. Một số mô hình phổ biến bao gồm:

- Mô hình huấn luyện trước: Mô hình này được đào tạo trên một tập dữ liệu lớn và có thể được sử dụng để học hỏi từ dữ liệu mới.
- Mô hình huấn luyện sau: Mô hình này được đào tạo trên một tập dữ liệu nhỏ và có thể được tùy chỉnh cho một nhiệm vụ cụ thể.

- Mô hình huấn luyện trên đám mây: Mô hình này được đào tạo trên các máy tính đám mây và có thể truy cập từ bất kỳ nơi nào.

Các mô hình đào tạo mô hình đang trở nên ngày càng phổ biến trong nghiên cứu và xây dựng mô hình AI/ML. Chúng giúp các nhà nghiên cứu và nhà phát triển tiết kiệm thời gian và công sức, đồng thời cho phép họ tạo ra các mô hình mới hiệu quả hơn.

Dưới đây là một số lợi ích của việc sử dụng các mô hình đào tạo mô hình trong nghiên cứu và xây dựng mô hình AI/ML:

- Tăng tốc độ phát triển mô hình: Các mô hình đào tạo mô hình có thể tự động hóa các tác vụ, chẳng hạn như thu thập và tiền xử lý dữ liệu, lựa chọn mô hình và thuật toán, và đánh giá hiệu quả của mô hình. Điều này giúp các nhà nghiên cứu và nhà phát triển tiết kiệm thời gian và công sức, đồng thời cho phép họ tạo ra các mô hình mới nhanh hơn.
- Tăng cường hiệu quả của mô hình: Các mô hình đào tạo mô hình có thể giúp các nhà nghiên cứu và nhà phát triển tạo ra các mô hình mới hiệu quả hơn bằng cách sử dụng các phương pháp học máy tiên tiến.
- Mở rộng khả năng tiếp cận: Các mô hình đào tạo mô hình có thể truy cập từ bất kỳ nơi nào, giúp các nhà nghiên cứu và nhà phát triển trên toàn thế giới chia sẻ kiến thức và công cụ của họ.

Các mô hình đào tạo mô hình là một công cụ mạnh mẽ có thể giúp các nhà nghiên cứu và nhà phát triển tạo ra các mô hình AI/ML mới hiệu quả hơn. Chúng đang trở nên ngày càng phổ biến và có khả năng tác động đáng kể đến cách chúng ta phát triển và sử dụng AI/ML.

4.3 Tối ưu và cải thiện mô hình

Tối ưu và cải thiện mô hình là một quá trình liên tục trong nghiên cứu và xây dựng mô hình AI/ML. Các nhà nghiên cứu và nhà phát triển cần liên tục kiểm tra và đánh giá hiệu quả của mô hình, và thực hiện các thay đổi cần thiết để cải thiện hiệu quả.

Có nhiều cách để tối ưu và cải thiện mô hình AI/ML, bao gồm:

- Chọn mô hình và thuật toán phù hợp: Mô hình và thuật toán phù hợp là yếu tố quan trọng nhất quyết định hiệu quả của mô hình. Các nhà nghiên cứu và nhà phát triển cần chọn mô hình và thuật toán phù hợp với bài toán cụ thể.

- Chuẩn hóa dữ liệu: Dữ liệu chuẩn hóa sẽ giúp mô hình học tập hiệu quả hơn. Các nhà nghiên cứu và nhà phát triển cần chuẩn hóa dữ liệu bằng cách loại bỏ các mẫu dữ liệu lỗi, làm sạch dữ liệu, và chuyển đổi dữ liệu thành định dạng phù hợp với mô hình.
- Huấn luyện mô hình với nhiều tập dữ liệu: Huấn luyện mô hình với nhiều tập dữ liệu sẽ giúp mô hình tổng quát tốt hơn. Các nhà nghiên cứu và nhà phát triển cần huấn luyện mô hình với nhiều tập dữ liệu, bao gồm cả tập dữ liệu huấn luyện, tập dữ liệu kiểm tra, và tập dữ liệu đánh giá.
- Chọn tham số tối ưu: Tham số tối ưu sẽ giúp mô hình đạt hiệu quả cao nhất. Các nhà nghiên cứu và nhà phát triển cần chọn tham số tối ưu cho mô hình bằng cách thử nghiệm các tham số khác nhau.
- Sử dụng các kỹ thuật tối ưu hóa: Các kỹ thuật tối ưu hóa có thể giúp mô hình học tập hiệu quả hơn. Các nhà nghiên cứu và nhà phát triển có thể sử dụng các kỹ thuật tối ưu hóa, chẳng hạn như khử bội, khử nhiễu, và tăng cường dữ liệu, để cải thiện hiệu quả của mô hình.

Dưới đây là một số mẹo để tối ưu và cải thiện mô hình AI/ML:

- Thử nghiệm các mô hình và thuật toán khác nhau: Không có một mô hình hoặc thuật toán nào phù hợp với tất cả các bài toán. Các nhà nghiên cứu và nhà phát triển cần thử nghiệm các mô hình và thuật toán khác nhau để tìm ra mô hình phù hợp nhất với bài toán cụ thể.
- Theo dõi hiệu quả của mô hình: Các nhà nghiên cứu và nhà phát triển cần theo dõi hiệu quả của mô hình trong quá trình phát triển. Điều này sẽ giúp họ xác định các lĩnh vực cần cải thiện.
- Sẵn sàng thay đổi: Các nhà nghiên cứu và nhà phát triển cần sẵn sàng thay đổi mô hình khi cần thiết. Điều này có thể bao gồm việc thay đổi mô hình, thuật toán, hoặc tập dữ liệu.

Tối ưu và cải thiện mô hình là một quá trình không dễ dàng, nhưng cần thiết để tạo ra các mô hình AI/ML hiệu quả. Các nhà nghiên cứu và nhà phát triển cần có kiến thức và kỹ năng chuyên môn, đồng thời kiên trì thử nghiệm và đánh giá để tìm ra các cách cải thiện hiệu quả của mô hình.

Chương II: Xây dựng mô hình phát hiện xâm nhập mạng dựa trên bộ dữ liệu NSL-KDD

1. Giới thiệu bộ dữ liệu NSL-KDD

Tập dữ liệu NSL-KDD được xây dựng để phục vụ cho việc thử nghiệm một số phương pháp phát hiện xâm nhập mạng. Tập dữ liệu cung cấp dữ liệu chuẩn và đáng tin cậy để thử nghiệm cũng như đánh giá các phương pháp phát hiện xâm nhập mạng máy tính. Bộ dữ liệu hiện đang được sử dụng rộng rãi trong các thuật toán và mô hình trí tuệ nhân tạo phát hiện xâm nhập mạng trái phép.

1.1 Nguồn gốc bộ dữ liệu

Bộ dữ liệu NSL-KDD là tập dữ liệu được công khai, phát triển từ bộ dữ liệu KDD'99 và giải quyết được một số vấn đề cố hữu cua tập dữ liệu KDD'99.

Những ưu điểm của NSL-KDD so với KDD'99:

- Loại bỏ được những bản ghi thừa trong tập Train. Do đó các mô hình phân loại không bị lệch và hướng những bản tin xuất hiện thường xuyên.
- Số lượng bản tin được chọn từ mỗi nhóm mức độ khó tỷ lệ nghịch với tỷ lệ phần trăm bản ghi trong tập dữ liệu KDD'99. Khi đó, tỷ lệ phân loại của các phương pháp học máy khác nhau sẽ cao hơn, giúp đánh giá chính xác các kỹ thuật học khác nhau trở nên hiệu quả hơn
- Số lương bản tin trong tập Train và Test là hợp lý, lần lượt là 125,973 bản tin và 22,544 bản tin. Điều này giúp chạy thử nghiệm trên tập hợp hoàn chỉnh mà không cần phải chọn ngẫu nhiên một phần nhỏ. Nhờ đó, kết quả đánh giá của các công trình nghiên cứu khác nhau sẽ có tính nhất quán.
- Loại bỏ được những bản tin trùng lặp trong tập Test, do đó, hiệu quả của mô hình sẽ không bị sai lệch bởi các phương pháp có tỷ lệ phát hiện tốt hơn trên các bản tin thường xuyên.

1.2 Các trường dữ liệu trong NSL-KDD

Khi thu thập dữ liệu cung cấp cho thiết kế mô hình phát hiện xâm nhập mạng, các bản tin luôn rất lớn chứa cả những thông tin không hiệu quả, thông tin cần thiết và thông tin vector đa chiều. Những thông tin này gây ra những vấn đề như gây tổn thất tài nguyên, giảm hiệu quả và tốc độ của mô hình cũng như nguy cơ gây "overfitting" khi huấn luyện.

Việc lọc ra những thông tin dư thừa giúp giảm đáng kể số lượng tài nguyên máy tính, bộ nhớ và thời gian CPU cần thiết để phát hiện xâm nhập trái phép. [1]

Trong tập dữ liệu NSL-KDD, các thông tin dư thừa đã được loại bỏ thông qua nhiều kỹ thuật rút gọn đặc trưng. Theo đó, trong tập dữ liệu chỉ giữ lại 41 trường dữ liệu như sau:

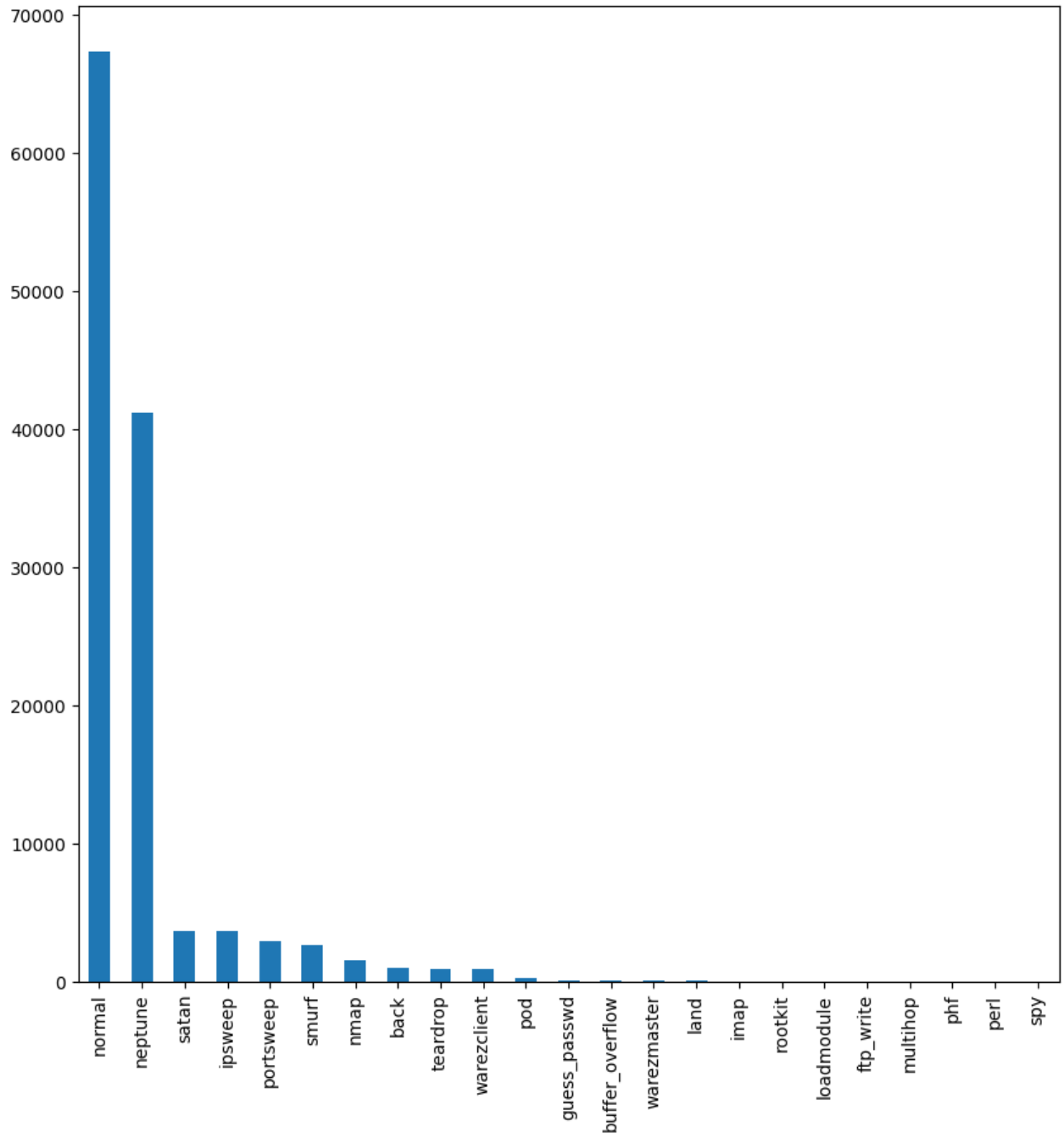
STT	Tên trường	Ý nghĩa	Dạng dữ liệu
1	duration	Thời gian bản tin duy trì	numerical
2	protocol_type	Giao thức kết nối	category
3	service	Dịch vụ	category
4	flag	Cờ trạng thái	category
5	src_bytes	Số byte gửi từ nguồn đến đích	numerical
6	dst_bytes	Số byte gửi từ đích đến nguồn	numerical
7	land	xác định thiết bị có đang gửi tin cho chính nó hay không	category
8	wrong_fragment	xác định số lượng byte bị hỏng	numerical
9	urgent	xác định độ ưu tiên của gói tin	category
10	hot		numerical
11	num_failed_logins	số lần đăng nhập thất bại	numerical
12	logged_in	đăng nhập thành công	category
13	num_compromised	số lượng máy tính bị xâm phạm	numerical
14	root_shell	kiểm tra kết nối root	category
15	su_attempted		
16	num_root	số lần truy cập root	numerical
17	num_file_creations	số lần tạo tệp tin mới	numerical
18	num_shells	số lượng phiên làm việc	numerical
19	num_access_files	số tệp tin đã truy cập	numerical
20	bum_outbound_cmds	số lượng lệnh hoặc gói tin đi ra được thực hiện	numerical
21	is_host_login	xác định đăng nhập máy chủ	category
22	is_guest_login	xác định đăng nhập máy chủ dưới quyền guest không	category
23	count	số lần các gói tin được thực hiện	numerical
24	srv_count	số lần các yêu cầu dịch vụ được thực hiện	numerical
25	serror_rate	tỷ lệ các gói tin bị lỗi	numerical
26	srv_serror_rate	tỷ lệ các yêu cầu dịch vụ mạng srv gây ra lỗi serror	numerical
27	rerror_rate	tỷ lệ các yêu cầu đến gây ra lỗi	numerical
28	srv_rerror_rate	tỷ lệ các srv gây lỗi rerror	numerical
29	same_srv_rate	tỷ lệ srv cùng một loại dịch vụ so với tổng số srv	numerical

30	diff_srv_rate	tỷ lệ srv các loại dịch vụ khác nhau so với tổng srv	numerical
31	srv_dif_host_rate	tỷ lệ srv thực hiện trên host khác nhau so với tổng srv	numerical
32	dst_host_count	số lần máy chủ đích được kết nối	numerical
33	dst_host_srv_count	số lần srv truy cập trên máy chủ đích	numerical
34	dst_host_same_srv_rate	tỷ lệ srv thuộc một loại dịch vụ trên máy chủ đích so với tổng srv	numerical
35	dst_host_diff_srv_rate	tỷ lệ srv các loại dịch vụ trên máy chủ đích so với tổng srv	numerical
36	dst_host_same_src_port_rate	tỷ lệ số lượng các kết nối mạng có cùng cổng nguồn trên máy khách đến cùng một cổng đích trên máy chủ so với tổng kết nối mạng	numerical
37	dst_host_srv_diff_host_rate	tỷ lệ số lần srv được truy cập trên máy chủ đích khác nhauso với tổng srv truy cập trên máy chủ đích	numerical
38	dst_host_serror_rate	tỷ lệ lỗi do srv gây ra trên máy chủ đích so với tổng srv thực hiện trên máy chủ đích	numerical
39	dst_host_srv_serror_rate	tỷ lệ lỗi serror do srv gây ra trên máy chủ đích so với tổng srv thực hiện trên máy chủ đích	numerical
40	dst_host_rerror_rate	tỷ lệ lỗi rerror do srv gây ra trên máy chủ đích so với tổng srv thực hiện trên máy chủ đích	numerical
41	dst_host_srv_rerror_rate	tỷ lệ lỗi rerror do srv gây ra trên máy chủ đích so với tổng srv thực hiện trên máy chủ đích	numerical

Bảng 1: Các trường dữ liệu trong NSL-KDD

1.3 Các kiểu tấn công có trong tập dữ liệu

Trong tập dữ liệu huấn luyện NSL-KDD, có 23 loại xâm nhập mạng được ghi nhận. Trong đó, xâm nhập "Normal" được coi là xâm nhập hợp pháp, các kiểu xâm nhập còn lại đều là những xâm nhập không hợp pháp.



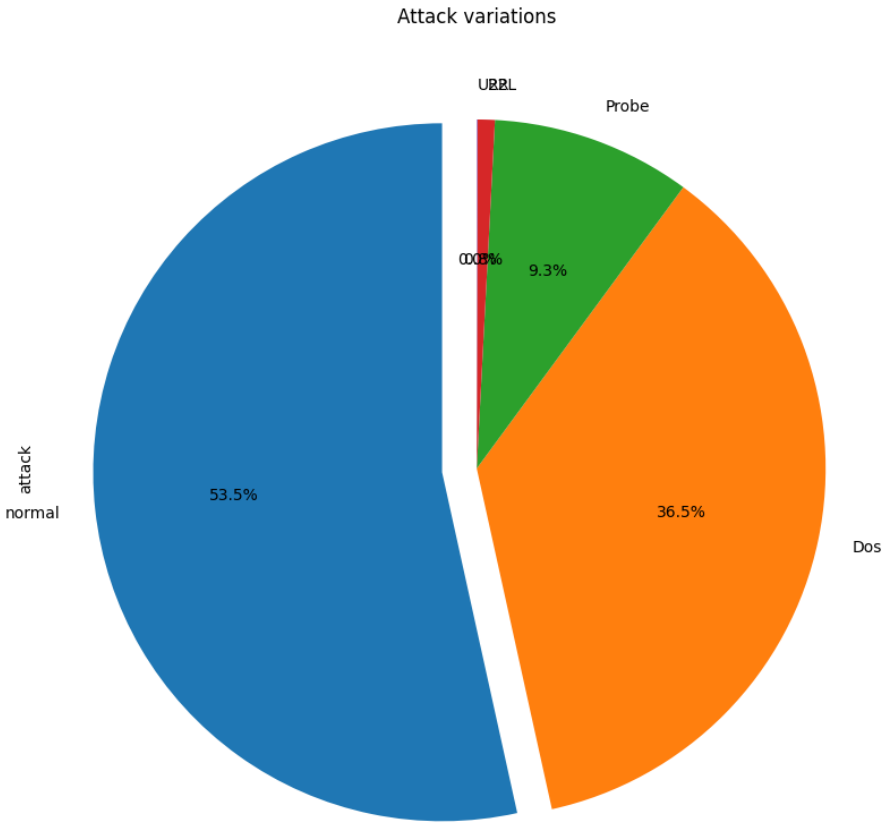
Hình 1.1.Các kiểu tấn công trong NSL-KDD

Những kiểu tấn công này thường được chia vào các loại tấn công mạng theo các dạng tấn công thường gặp [2]. Cụ thể là:

Attack Class	Attack Type
DoS	apache2, back, land, Neptune, pod, processtable, smurf, teardrop, udpstorm, worm
Probes	Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint

R2L	Guess_Password, Ftp_write, Imap, phf, Multihop, warezmaster, warezclient, spy, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, Named
U2R	Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps

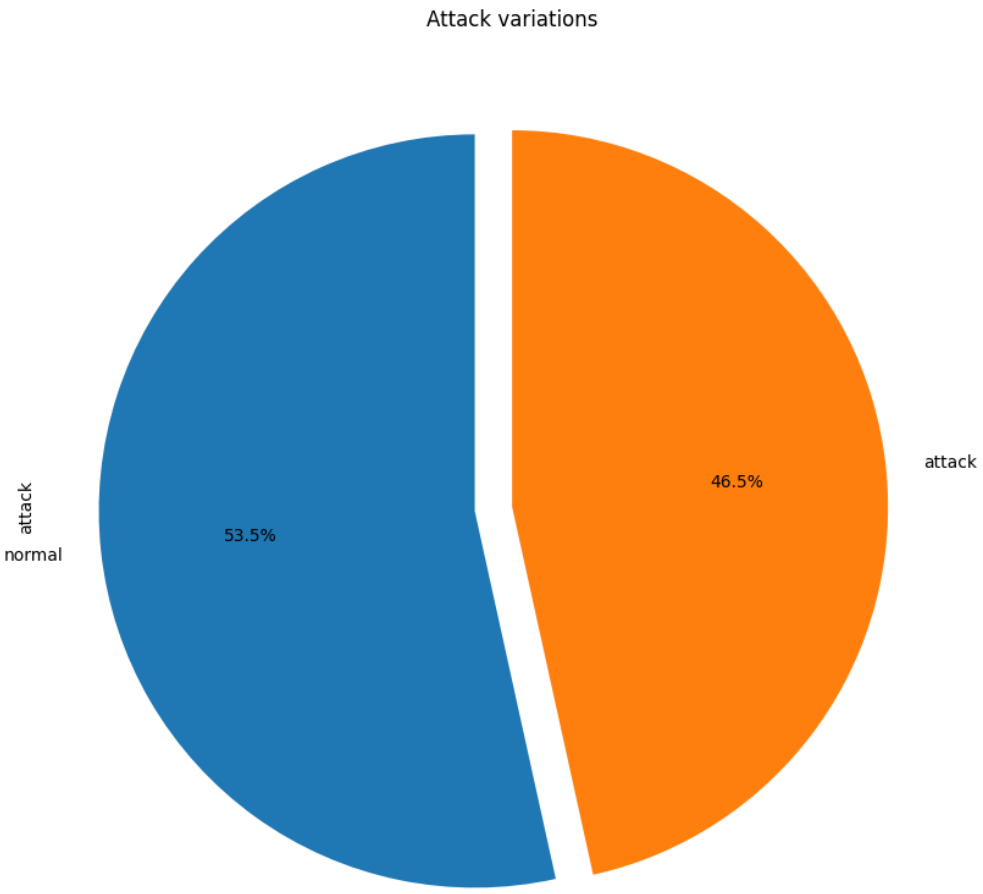
- DOS: Từ chối dịch vụ là một loại tấn công, làm cạn kiệt tài nguyên của nạn nhân, khiến nó không thể xử lý các yêu cầu hợp pháp – ví dụ như Syn Flood. Các tính năng liên quan: "byte nguồn" và "tỷ lệ phần trăm gói có lỗi"
- Thăm dò: Mục tiêu của cuộc tấn công giám sát và thăm dò khác là thu thập thông tin về nạn nhân từ xa, ví dụ như quét cổng. Các tính năng liên quan: "thời lượng kết nối" và "byte nguồn"
- U2R: truy cập trái phép vào đặc quyền siêu người dùng (root) cục bộ là một kiểu tấn công, trong đó kẻ tấn công sử dụng tài khoản bình thường để đăng nhập vào hệ thống nạn nhân và cố gắng giành được đặc quyền root/quản trị viên bằng cách khai thác một số lỗ hổng trong nạn nhân, ví dụ như bộ đệm các cuộc tấn công tràn. Các tính năng liên quan: "số lần tạo tệp" và "số lần nhắc nhở shell được gọi"
- R2L: truy cập trái phép từ máy từ xa, kẻ tấn công xâm nhập vào máy từ xa và giành quyền truy cập cục bộ vào máy nạn nhân. Ví dụ: đoán mật khẩu Các tính năng liên quan: Tính năng cấp độ mạng – "thời lượng kết nối" và "dịch vụ được yêu cầu" và các tính năng cấp độ máy chủ - "số lần đăng nhập thất bại"



Hình 1.2 Phân bổ các loại tấn công trong NSL-KDD

2. Xây dựng mô hình máy học

Trong bài báo cáo này, chúng ta sẽ chỉ dừng lại ở việc xây dựng mô hình học máy để xác nhận có hay không việc xâm nhập mạng trái phép. Do đó, trường thông tin “attack” trong tập dữ liệu NSL-KDD chỉ có 2 thông tin là “attack” và “normal”.



Hình 2.1 Phân bố thông tin trong trường "attack"

2.1 Tiền xử lý dữ liệu

Biến đổi dữ liệu dạng category thành dạng continuous

Các dữ liệu có dạng category trong NSL-KDD bào gồm các trường: protocol_type, service, flag, land, urgent, logged_in, is_host_login, is_guest_login. Những trường dữ liệu này cần được chuyển sang dạng continuous, nhằm loại bỏ sự thứ tự và mối quan hệ giữa các giá trị phân loại, ngăn chặn được những giả định sai lầm.

Trong mô hình này, chúng ta sẽ sử dụng kỹ thuật One-hot Encoder để biến đổi những trường category thành numerical. Cụ thể:

```
#One-Hot Encoder catergory attribute
objects = ["protocol_type", "service", "flag","land","urgent", "logged_in", "is_host_login","is_guest_login"]
for obj in objects:
  dummies = pd.get_dummies(X[obj])
  X = pd.concat([X,dummies], axis=1)
  X.drop(obj, axis = 1, inplace=True)
```

Đối với trường dữ liệu "attack" là nhãn, chúng ta sử dụng phương pháp Label Encoder để biến đổi.

```
# encode label
Lb = LabelEncoder()
Y = Lb.fit_transform(Y)
Y
```

```
array([1, 0, 1, ..., 1, 0, 1])
```

khi đó, trường dữ liệu nhãn Y sẽ chuyển các giá trị "attack" và "normal" thành dạng 0 và 1 với 0 là "attack" và 1 là "normal".

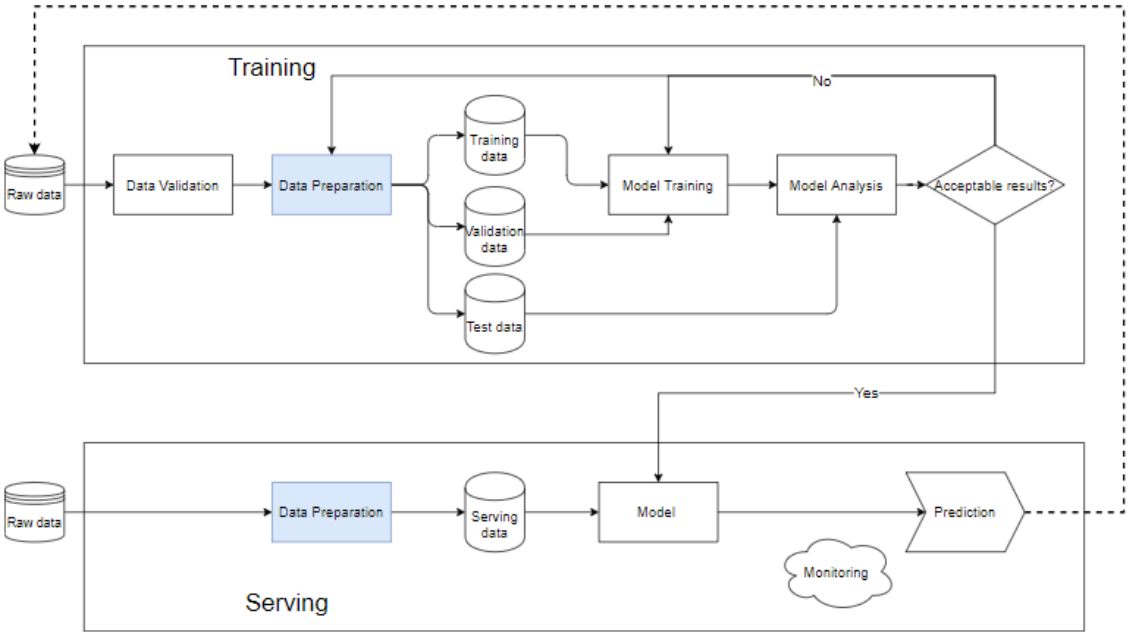
Chuẩn hóa dữ liệu

Đối với những mô hình máy học, việc chuẩn hóa dữ liệu có ảnh hưởng lớn đến hiệu suất. Chuẩn hóa giúp giảm ảnh hưởng của các giá trị ngoại lai, cải thiện sự hội tụ của thuật toán tối ưu hóa, loại bỏ sự biến đổi tỷ lệ…

Trong mô hình này, chúng ta sẽ sử dụng phương pháp Normalizer để chuẩn hóa dữ liệu, đưa dữ liệu về dạng các số trong đoạn từ 0 đến 1.

```
X.columns = X.columns.astype(str)
scaler = Normalizer(norm = 'l2').fit(X)
trainX = scaler.transform(X)
```

3. Xây dựng mô hình



3.1 Phân chia dữ liệu

Chúng ta sẽ không huấn luyện mô hình bằng toàn bộ dữ liệu trong tập huấn luyện của NSL-KDD, thay vì thế, chúng ta sẽ chia tập dữ liệu thành 2 tập dữ liệu nhỏ hơn. Điều này sẽ tránh lỗi "overfitting" cho mô hình

Thông thường, ta sẽ phân chia theo tỷ lệ 8:2 với 8 phần cho tập huấn luyện và 2 phần cho tập thử nghiệm.

```
from sklearn.model_selection import train_test_split
X_train, X_valid, y_train, y_valid = train_test_split(trainX,Y,train_size = 0.8, random_state = 42)
```

3.2 Thử nghiệm với các mô hình

Để có được mô hình tốt nhất, cần phải thử nghiệm với nhiều mô hình khác nhau. Trong bài luận này, chúng ta sẽ thử nghiệm 3 thuật toán phân loại trong học máy, là SVM, KNN và Logistic Regresstion. Những mô hình này đều đã được tích hợp sẵn trong thư viện sklearn của ngôn ngữ python.

Mô hình SVM

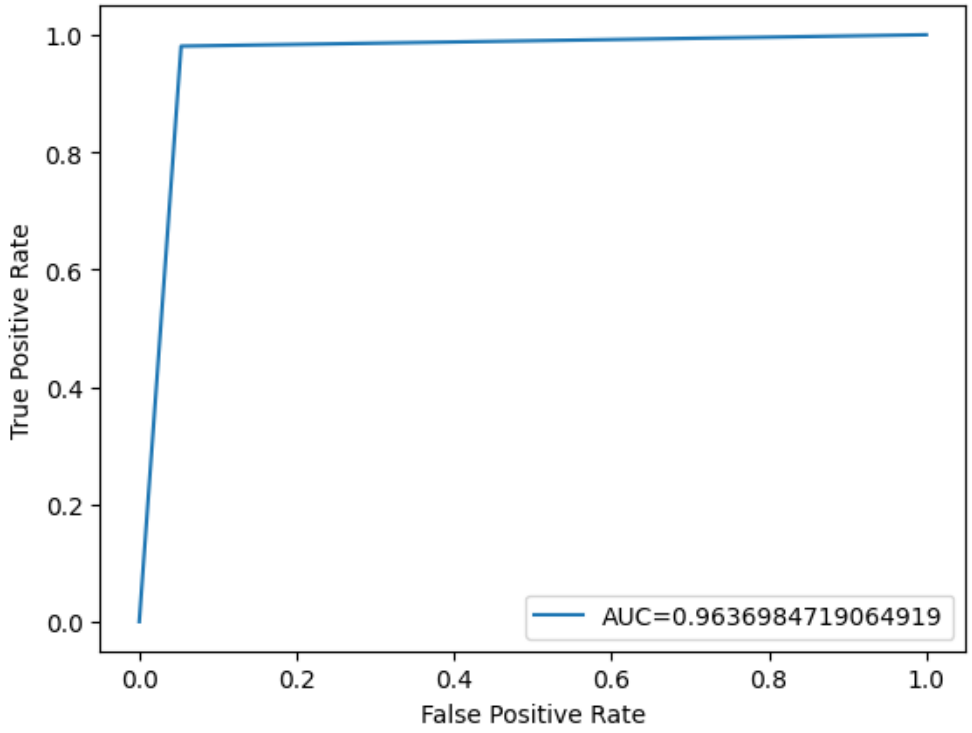
```
from sklearn.svm import SVC
model = SVC()
model.fit(X_train,y_train)
y_pred_svc = model.predict(X_valid)
accuracy = accuracy_score(y_valid, y_pred_svc)
```

Chỉ số đánh giá:

```
Accuracy: 0.964755
Precision: 0.954341
Recall: 0.980577
F1 score: 0.967281
              precision    recall  f1-score   support

           0       0.98      0.95      0.96     11809
           1       0.95      0.98      0.97     13386

    accuracy                           0.96     25195
   macro avg       0.97      0.96      0.96     25195
weighted avg       0.97      0.96      0.96     25195
```



Hình 2.2 Kết quả quá trình học của SVM

Mô hình KNN

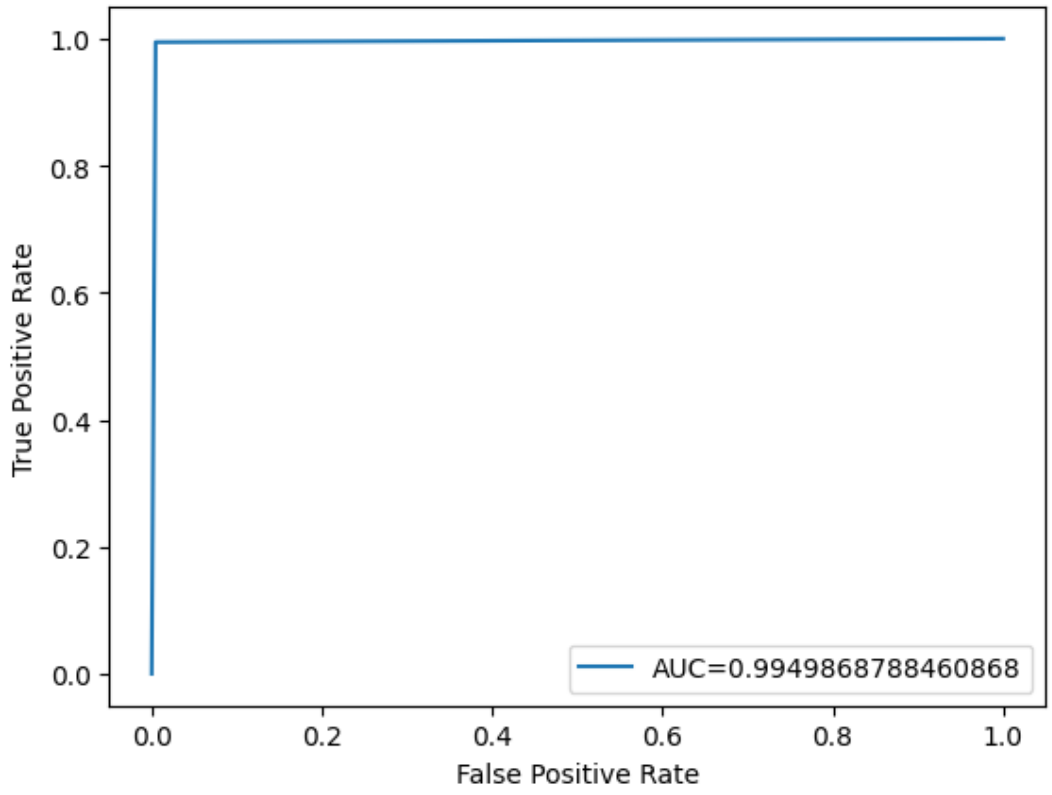
```
model_KNN = KNeighborsClassifier()
model_KNN.fit(X_train,y_train)
y_pred_KNN = model_KNN.predict(X_valid)
accuracy = accuracy_score(y_valid, y_pred_KNN)
```

Tham số đánh giá:

```
Accuracy: 0.994959
Precision: 0.995960
Recall: 0.994547
F1 score: 0.995253
              precision    recall  f1-score   support

           0       0.99      1.00      0.99     11809
           1       1.00      0.99      1.00     13386

    accuracy                           0.99     25195
   macro avg       0.99      0.99      0.99     25195
weighted avg       0.99      0.99      0.99     25195
```



Hình 2.3 Kết quả quá trình học của KNN

Mô hình Logistic Regresstion

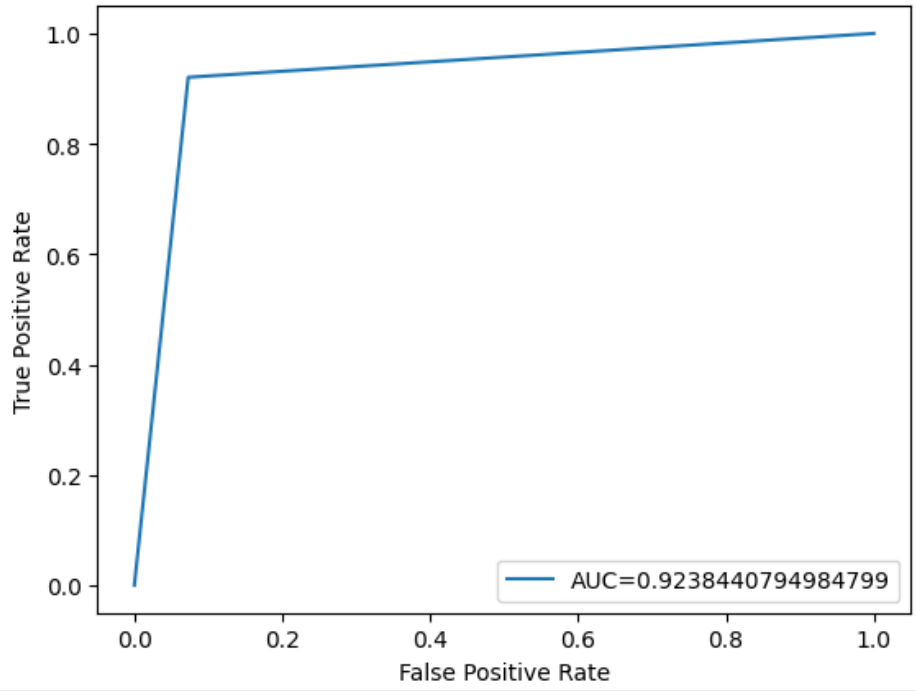
```
model_lc = LogisticRegression()
model_lc.fit(X_train,y_train)
y_pred_lr = model_lc.predict(X_valid)
accuracy = accuracy_score(y_valid, y_pred_lr)
```

Tham số đánh giá:

```
Accuracy: 0.923636
Precision: 0.934760
Recall: 0.920514
F1 score: 0.927582
              precision    recall  f1-score   support

           0       0.91      0.93      0.92     11809
           1       0.93      0.92      0.93     13386

    accuracy                           0.92     25195
   macro avg       0.92      0.92      0.92     25195
weighted avg       0.92      0.92      0.92     25195
```



Hình 2.4 Kết quả quá trình học của Logistic Regresstion

3. Kiểm thử và đánh giá mô hình

Các độ đo **accuracy, precision, recall, f1 score** tổng hợp từ ba thuật toán trên tập dữ liệu NSL-KDD

	Accuracy	**Precision**	**Recall**	**f1 score**
KNN	0.994959	0.995960	O.994547	0.994253
KVM	0.964755	0.954341	0.980577	0.967281

Logistic Regresstion	0.923636	0.934760	0.920514	0.927582

Ta có thể thấy kết quả tổng hợp của ba thuật toán KNN, KVM và Logistic Regresstion. Kết quả tổng hợp được lấy kết quả tốt nhất mỗi lần chạy của từng thuật toán. Dựa vào bảng tổng hợp kết quả, có thể thấy được KNN ghi nhận các thông số đầu ra vượt trội hơn so với hai thuật toán còn lại về độ chính xác, cụ thể là hơn KVM 3.02% và Logistic Regresstion 7.13%. Dựa các kết quả này, mô hình đề xuất sẽ có thể triển khai và áp dụng vào việc lựa chọn mô hình cho các bộ dữ liệu sau.

Xét bộ dữ liệu NSL-KDD, trường dữ liệu protocol_type có tổng cộng ba loại giao thức là TCP, UDP và IMCP. Trong đó, số lượng của từng loại giao thức được mô tả cụ thể như sau: TCP, UDP, IMCP có số lượng lần lượt là 18880, 2621 và 1043. Theo kết quả có được, với dữ liệu có số lượng giao thức TCP lớn hơn số lượng các loại giao thức còn lại thì KNN sẽ là mô hình thích hợp cho bộ dữ liệu. Tương tự như vậy, đối với dữ liệu có số lượng UDP lớn hơn số lượng các giao thức còn lại thì SVM sẽ là mô hình phù hợp. Và cuối cùng là ứng với dữ liệu có số lượng IMCP lớn hơn số lượng các giao thức còn lại thì Logistic Regresstion sẽ là mô hình phù hợp.

4. So sánh với những phương pháp truyền thống

Sử dụng Máy học và Trí tuệ Nhân tạo (AL/ML) trong việc xâm nhập mạng có thể thay đổi cách chúng ta đối phó với việc bảo vệ mạng máy tính so với các phương pháp truyền thống. Dưới đây là một số sự so sánh giữa sử dụng AL/ML và phương pháp truyền thống trong việc xâm nhập mạng:

- Phát hiện sự xâm nhập:
  - Truyền thống: Các phương pháp truyền thống thường dựa vào việc so sánh dữ liệu mạng với các quy tắc cố định để xác định các hoạt động bất thường.
  - AL/ML: Mô hình AL/ML có khả năng học từ dữ liệu thời gian thực và phát hiện các mô hình phức tạp của xâm nhập mạng, bao gồm cả những tấn công mà chưa từng được biết đến trước đó.
- Tính năng thích nghi:
  - Truyền thống: Các giải pháp truyền thống cần cập nhật thường xuyên để đối phó với các mối đe dọa mới.
  - AL/ML: Mô hình AL/ML có thể tự động học và điều chỉnh mô hình của mình theo thời gian, từ đó cải thiện tính khả năng thích nghi với các hình thức tấn công mới.
- Phát hiện sớm:
  - Truyền thống: Các phương pháp truyền thống có thể bỏ lỡ các mô hình xâm nhập mới và chưa từng thấy trước đó.

- AL/ML: Mô hình AL/ML có thể phát hiện sớm các dấu hiệu tiềm năng của xâm nhập mạng, giúp ngăn chặn tấn công trước khi gây hậu quả nghiêm trọng.

➢ Sự phân loại:
- Truyền thống: Phương pháp truyền thống thường dựa vào các quy tắc cố định để phân loại các sự kiện mạng.
- AL/ML: Mô hình AL/ML có khả năng phân loại các sự kiện mạng dựa trên các đặc điểm thống kê và học tập từ dữ liệu, giúp nâng cao độ chính xác trong việc phát hiện xâm nhập.

➢ Phản ứng nhanh:
- Truyền thống: Phản ứng trong thời gian thực đòi hỏi sự can thiệp của người làm việc.
- AL/ML: Mô hình AL/ML có thể tự động thực hiện các biện pháp phản ứng nhanh, giúp giảm thiểu thời gian giữa việc phát hiện và phản ứng.

Sử dụng AL/ML trong việc xâm nhập mạng có thể cung cấp tính hiệu quả và tính khả năng thích nghi cao hơn so với phương pháp truyền thống. Tuy nhiên, việc triển khai và duy trì mô hình AL/ML cũng đòi hỏi kiến thức chuyên sâu và sự cập nhật liên tục để đảm bảo tính hiệu quả trong việc bảo vệ mạng máy tính.

Sử dụng AL/ML có những lợi ích sau:

- Tính tự động hóa: AL/ML có thể thực hiện quyết định và phản ứng tự động, giảm sự phụ thuộc vào can thiệp con người trong việc xử lý các tình huống xâm nhập.
- Phát hiện mẫu phức tạp: AL/ML có khả năng phát hiện các mô hình xâm nhập phức tạp và không dựa vào các quy tắc cố định, giúp phát hiện các tấn công mà phương pháp truyền thống có thể bỏ lỡ.
- Phản ứng nhanh: Mô hình AL/ML có thể phản ứng nhanh hơn để ngăn chặn các tấn công ngay khi chúng được phát hiện.
- Khả năng học và thích nghi: AL/ML có khả năng học từ dữ liệu mới và thích nghi với các mối đe dọa mới mà không cần cập nhật thủ công.
- Phân loại chính xác hơn: Mô hình AL/ML có thể cung cấp độ chính xác cao hơn trong việc phân loại các sự kiện mạng, giúp giảm thiểu các báo giả.

Bên cạnh những lợi ích đó thì vẫn có một vài hạn chế của sử dụng AL/ML:

- Sự phức tạp: Triển khai và duy trì các mô hình AL/ML có thể đòi hỏi kiến thức chuyên sâu và tài nguyên lớn.
- Độ tin cậy: Mô hình AL/ML có thể bị lừa dối bởi các tấn công tạo mô hình giả mạo nếu không được cấu hình chính xác.
- Nhận diện lỗi: AL/ML có thể tạo ra các báo động sai lệch nếu không được đào tạo hoặc cấu hình đúng cách.

- Bảo mật của mô hình: Mô hình AL/ML cũng có thể trở thành mục tiêu của các tấn công nếu không được bảo vệ cẩn thận.
- Chi phí cao: Sử dụng AL/ML có thể đòi hỏi đầu tư lớn về tài nguyên tính toán và con người.

Tóm lại, việc sử dụng AL/ML trong việc xâm nhập mạng có thể mang lại nhiều lợi ích về hiệu quả và khả năng phát hiện so với phương pháp truyền thống, nhưng cũng đòi hỏi sự đầu tư và quản lý kỹ lưỡng để đảm bảo tính an toàn và hiệu quả của mô hình. Tùy thuộc vào môi trường và mục tiêu cụ thể, có thể làm việc với một sự kết hợp của AL/ML và phương pháp truyền thống để đảm bảo tính an toàn của mạng máy tính. Việc sử dụng cả hai có thể tận dụng lợi ích của cả hai phương pháp và giảm bớt hạn chế. Ngoài ra, việc duy trì kiến thức về các mối đe dọa mới và cập nhật thường xuyên là quan trọng đối với cả hai phương pháp để duy trì bảo mật mạng hiệu quả.

KẾT LUẬN

Trong tương lai kỷ nguyên số hóa và mạng thông tin, việc phát hiện và ngăn chặn xâm nhập trái phép đã trở thành một chu kỳ ngày càng quan trọng và phức tạp. Mạng Xâm nhập không chỉ mục đích kiểm tra tính bảo mật và dữ liệu riêng tư, mà còn có thể gây ra hậu quả tài chính và danh tiếng đáng kể. Trong bối cảnh bối rối này, ứng dụng Trí tuệ Nhân tạo và Học Máy đã nổi lên như một giải pháp tiềm năng để tăng cường bảo vệ mạng và dữ liệu.

Qua bài tiểu luận này, chúng tôi đã tìm hiểu về cách AI và ML có thể được sử dụng để phát hiện các hoạt động bất thường trên mạng, xác định các mối đe dọa tiềm ẩn và sớm đưa ra các cảnh báo. Chúng tôi đã nhận thấy rằng các giải pháp này có khả năng cải thiện khả năng phát hiện và giảm thiểu thời gian phản ứng trong các xâm nhập xâm nhập trường hợp.

Tuy nhiên, không có giải pháp bảo mật nào là hoàn hảo. AI và ML có thể giúp cải thiện khả năng phát hiện, nhưng chúng cũng là đối tượng cho các công thức như độ tin cậy của dữ liệu đào tạo và đảm bảo tính riêng tư của thông tin cá nhân. Vì vậy, việc kết hợp giải pháp AI/ML với các giải pháp bảo mật khác là cần thiết để tạo ra một mạng hệ thống an toàn và đáng tin cậy.

Trong tương lai, chúng ta cần tiếp tục nghiên cứu và phát triển trong lĩnh vực này để hỗ trợ các mối đe dọa ngày càng tiên tiến. Sự hợp tác giữa cộng đồng nghiên cứu, các ngành công nghiệp và cơ sở chính phủ cũng là quan trọng để xây dựng một môi trường mạng an toàn và đáng tin cậy hơn. Chúng tôi hy vọng rằng những nỗ lực này sẽ giúp giảm thiểu nguy cơ xâm nhập mạng và bảo vệ thông tin quý giá của chúng ta trong tương lai số hóa.

Tài liệu tham khảo

[1] M. H. Hany, S. D. Abeer, A. Z. Afaf and A. G. Mohamed, "Selecting Optimal Subset of Features for Intrusion," pp. 180-181, January 2011.

[2] M. H. Hany, S. D. Abeer, A. Z. Afaf and A. G. Mohamed, "Selecting Optimal Subset of Features for Intrusion," pp. 180-181, January 2011.

[3] Statistics Department, Faculty of Science and Data Analytics, Sepuluh Nopember Institute of Technology (ITS), "Analysis of NSL-KDD Dataset for Classification of Attacks Based on," p. 508, 2022.

[4] https://web.archive.org/web/20150205070216/http://nsl.cs.unb.ca/NSL-KDD/

[5] https://cystack.net/blog/tan-cong-mang-cyber-attack