********

**BAN CƠ YẾU CHÍNH PHỦ**

**HỌC VIỆN KỸ THUẬT MẬT MÃ**

---o0o---

**MÔN HỌC: HỌC MÁY TRONG AN TOÀN THÔNG TIN**

**TIỂU LUẬN**

**PHÁT HIỆN TẤN CÔNG BRUTE-FORCE BẰNG HỌC MÁY**

**GVHD: TS NGUYỄN AN KHƯƠNG**

**Lớp: CHAT3P**

**Nhóm học viên thực hiện:**

**CHAT3P02 - Tăng Quốc Cường**

**CHAT3P03 - Nguyễn Văn Tiến Dũng**

**CHAT3P16 - Nguyễn Duy Tùng**

***TP.Hồ Chí Minh, tháng 12 năm 2024***

**MỤC LỤC**

[CHƯƠNG 1 GIỚI THIỆU 3](#_Toc187412075)

[1.1 Đặt vấn đề 3](#_Toc187412076)

[1.2 Mục tiêu 4](#_Toc187412077)

[1.3 Thách thức 5](#_Toc187412078)

[CHƯƠNG 2: KIẾN THỨC NỀN TẢNG 6](#_Toc187412079)

[2.1 Tấn công brute-force 6](#_Toc187412080)

[2.2 Học máy 8](#_Toc187412081)

[2.3 Mô hình Random Forest (RF) 11](#_Toc187412082)

[2.4 Mô hình Support Vector Machine (SVM) 12](#_Toc187412083)

[CHƯƠNG 3: DỮ LIỆU VÀ THỰC NGHIỆM 16](#_Toc187412084)

[3.1 Phát biểu bài toán 16](#_Toc187412085)

[3.2 Dữ liệu thực nghiệm 16](#_Toc187412086)

[3.3 Xử lý dữ liệu 17](#_Toc187412087)

[3.4 Cài đặt mô hình 17](#_Toc187412088)

[CHƯƠNG 4: ĐÁNH GIÁ KẾT QUẢ 19](#_Toc187412089)

[4.1 Đánh giá kết quả 19](#_Toc187412090)

[4.2 Kết luận 19](#_Toc187412091)

CHƯƠNG 1 GIỚI THIỆU

1.1 Đặt vấn đề

Trong thời đại kỷ nguyên công nghệ số, mặc dù các phương thức xác thực tiên ra đời như: Xác thực sinh học, xác thực đa yếu tố, xác thực bằng khóa an toàn. Mặc dù các phương thức xác thực tiên tiến ra đời nhằm thay thế phương thức xác thực bằng mật khẩu. Tuy nhiên xác thực bằng mật khẩu vẫn là yếu tố then chốt trong hệ thống bảo mật kể từ khi Internet ra đời ban đầu, cho tới ngày nay phương thức xác thực vẫn là một phương thức xác thực phổ biến nhất. Rất nhiều ứng dụng và dịch vụ yêu cầu người dùng tạo mật khẩu bằng việc kết hợp giữa các chữ số, chữ cái và ký hiệu đặc biệt để giảm nguy cơ kẻ xấu đoán được mật khẩu Tuy nhiên, sự gia tăng các cuộc tấn công nhằm vào xác thực dựa trên mật khẩu cũng tạo ra những thách thức về bảo mật và khả năng sử dụng như người dùng khó nhớ mật khẩu phức tạp, người dùng có xu hướng lặp lại mật khẩu ở nhiều dịch vụ khác nhau,..

Tại hội thảo “An toàn thông tin trong chuyển đổi số” diễn ra tại Cần Thơ vào ngày 24 tháng 09 năm 2024 theo Cục An toàn thông tin cho biết: “Tình hình an toàn thông tin mạng nhiều rủi ro thiệt hại. Năm 2023, thế giới thiệt hại 8.000 tỷ USD, năm 2024 thiệt hại 9.500 tỷ USD. Có đến 353.027.892 người bị ảnh hưởng do vi phạm dữ liệu. 11 giây có 01 tổ chức bị tấn công mã độc tống tiền. Tấn công mạng quy mô lớn, chuyên nghiệp. Tấn công làm tê liệt sản xuất, kinh doanh. Gián điệp mạng, tình báo mạng, khủng bố mạng… Các sự cố tấn công mạng nghiêm trọng đã xảy ra như: Tháng 6/2017, tấn công hệ thống vận tải biển A.P. Moller Maersk của Đan Mạch, gây thiệt hại 300 triệu USD; Tháng 5/2021, sử dụng mã độc tống tiền vào nhà điều hành mạng lưới ống dẫn dầu Colonial Pipeline, gây thiệt hại 4,4 triệu USD; Tháng 4/2022, Băng đảng ransomware Conti đã phát động một cuộc tấn công kéo dài hàng tháng chống lại các tổ chức Chính phủ của Costa Rica, Băng đảng rò rỉ 672 GB dữ liệu và mất nhiều tháng mới phục hồi được hệ thống; Tháng 6/2024, tấn công mạng vào trung tâm dữ liệu quốc gia Indonesia gây ảnh hưởng đến 40 cơ quan của Chính phủ, đặc biệt làm gián đoạn dịch vụ nhập cư và ảnh hưởng đến hoạt động tại các sân bay lớn trong nhiều ngày…Tại Việt Nam, những sự cố tấn công mạng nghiêm trọng trong thời gian qua như: Năm 2022, 517.627 địa chỉ IP của Việt Nam nằm trong mạng botnet; Năm 2023, hơn 5,5 triệu tài khoản có tên miền .VN bị tấn công ransomware; Năm 2024, hàng loạt các vụ tấn công mã hóa dữ liệu đòi tiền chuộc: VNDirect, PVOil, VN Post gây thiệt hại hàng trăm tỷ VNĐ. Dự báo nguy cơ năm 2025, sẽ có 3.000 cuộc tấn công/giây; 12 mã độc/giây; 70 lỗ hổng/điểm yếu mỗi ngày. Đối tượng bị tấn công năm 2025 gấp 2,7 lần năm 2020. Năm 2030, gấp 7,5 lần năm 2020…” (https://baoxaydung.com.vn)

Trước bối cảnh an ninh mạng ngày càng phức tạp và nguy hiểm, việc nghiên cứu các phương thức phòng thủ mới như: Sử dụng trình quản lý mật khẩu, xác thực không cần bằng mật khẩu, tăng cường xác thực đa yếu tố trở nên cực kỳ quan trọng đối với việc đảm bảo an toàn thông tin. Trong đó, việc sử dụng các thuật toán trong học máy (Machine Learning) để phát hiện các hoạt động đáng ngờ được xem như là một phương pháp hiệu quả để phát hiện và ngăn chặn các cuộc tấn công nhờ khả năng phân tích dữ liệu lớn, nhận diện các mẫu hành vi bất thường và phản ứng kịp thời.

Việc áp dụng học máy không chỉ giúp giảm thiểu thiệt hại mà còn đóng góp vào việc xây dựng các hệ thống bảo mật tiên tiến, đáp ứng trước các thách thức an ninh mạng ngày càng tinh vi.

1.2 Mục tiêu

Mục tiêu của nghiên cứu này tập trung phân tích về tấn công brute-force, một trong những hình thức tấn công phổ biến nhưng nguy hiểm. Tiểu luận cũng nghiên cứu bản chất, các phương thức hoạt động, và những hệ quả mà tấn công brute-force có thể gây ra đối với hệ thống thông tin. Điều này giúp cung cấp nền tảng lý thuyết vững chắc để hiểu rõ về các điểm yếu mà các hệ thống bảo mật hiện tại có thể gặp phải.

Tiếp theo, nhóm nghiên cứu tập trung vào việc ứng dụng học máy (machine learning) trên một tập dữ liệu mới, được tạo ra từ dữ liệu lưu lượng mạng thu thập từ một mạng sản xuất thực tế. Việc phát hiện ở cấp độ mạng giúp tăng tính khả thi và bảo vệ các thiết bị không có cơ chế bảo vệ nội bộ. Ứng dụng học máy như một công cụ tiên tiến và hiệu quả trong việc phát hiện và ngăn chặn các cuộc tấn công brute-force. Thông qua việc phân tích các mẫu hành vi, nhận diện những bất thường, và dự đoán các hành động có thể xảy ra, các thuật toán học máy sẽ được tích hợp để tự động hóa quá trình phát hiện tấn công, từ đó nâng cao tốc độ và độ chính xác trong việc xử lý các mối đe dọa.

Cuối cùng, nghiên cứu đặt mục tiêu xây dựng một "tường lửa" thông minh, không chỉ có khả năng bảo vệ hệ thống thông tin mà còn có tính năng tự học và thích nghi với các hình thức tấn công ngày càng tinh vi. Tường lửa này sẽ được thiết kế để tối ưu hóa khả năng phòng thủ, giảm thiểu thiệt hại tiềm tàng và đảm bảo tính liên tục của các hoạt động trong hệ thống. Những mục tiêu này hướng tới việc tạo ra một giải pháp bảo mật toàn diện, hiện đại, và bền vững trước những thách thức an ninh mạng ngày càng phức tạp.

1.3 Thách thức

Thách thức lớn nhất trong việc phát hiện tấn công brute-force bằng học máy, đó là chất lượng dữ liệu. bộ dữ liệu có thể thiếu thông tin cần thiết hoặc chứa nhiều giá trị thiếu, điều này ảnh hưởng đến khả năng huấn luyện mô hình.

Thứ hai, dữ liệu không chính xác có thể gây khó khăn trong việc phân tích và phát hiện.

Thách thức thứ ba, đặc điểm của tấn công brute-force đó là kẻ tấn công thường thay đổi phương pháp để tránh bị phát hiện, sử dụng các khoảng thời gian dài giữa các lần thử mật khẩu, hoặc thực hiện các cuộc tấn công tinh vi hơn.

Thứ tư, thách thức đến từ việc lựa chọn mô hình, việc tìm ra mô hình học máy phù hợp với dữ liệu và yêu cầu cụ thể có thể khó khăn, đặc biệt khi một mô hình quá phức tạp có thể dẫn đến overfitting. Tối ưu hóa tham số cho từng mô hình cũng tốn thời gian và yêu cầu kiến thức chuyên sâu.

Thứ năm, đánh giá hiệu suất của mô hình là một thách thức khác, khi việc xác định các chỉ số đánh giá phù hợp và đảm bảo khả năng tổng quát trên dữ liệu chưa thấy là rất quan trọng. Hành vi người dùng cũng có thể thay đổi theo thời gian, ảnh hưởng đến hiệu quả của mô hình nếu không được cập nhật thường xuyên.

Thách thứ cuối cùng, yêu cầu về mặt cấu hình máy tính đủ mạnh để huấn luyện các mô hình phức tạp, cùng với việc tuân thủ bảo vệ dữ liệu nhạy cảm về các quy định về quyền riêng tư trong quá trình thu thập và xử lý dữ liệu.

Những thách thức này cần được xem xét và giải quyết để nâng cao hiệu quả của việc phát hiện tấn công brute-force bằng học máy.

CHƯƠNG 2: KIẾN THỨC NỀN TẢNG

2.1 Tấn công brute-force

Tấn công brute-force là một phương pháp tấn công mật khẩu các nhân hoặc hệ thống rất phổ biến. Các cuộc tấn công brute-force liên quan đến việc 'đoán' tên người dùng và mật khẩu để truy cập trái phép vào hệ thống, bằng cách thử các ký tự, từng từ hoặc từng số thông qua một danh sách các giá trị có thể tìm ra mật khẩu chính xác. Kẻ tấn công sẽ sử dụng một danh sách các từ điển mật khẩu phổ biến hoặc tạo ra các chuỗi ký tự ngẫu nhiên để thử mật khẩu. Việc sử dụng phương pháp này kẻ tấn công có thể thử hàng triệu hoặc thậm chí hàng tủy lần trong một khoảng thời gian rất ngắn.

Các cuộc tấn công brute-force thường được tự động hóa với sự hỗ trợ của các công cụ chuyên dụng. Những công cụ này cho phép kẻ tất công sử dụng danh sách các tên người dùng và mật khẩu phổ biến, thực hiện hàng loạt lần đăng nhập tự động với tốc độ cao để tăng khả năng thành công. Dù là một phương pháp đơn giản, brute-force có tỷ lệ thành công đáng kể, phụ thuộc vào độ dài và độ phức tạp của mật khẩu. Mật khẩu yếu có thể bị bẻ khóa chỉ trong vài giây, trong khi các mật khẩu mạnh hơn có thể yêu cầu nhiều thời gian và tài nguyên hơn để phá vỡ.

Đặc biệt, hình thức tấn công này thường khai thác các điểm yếu phổ biến như:

Người dùng đặt mật khẩu không an toàn, dễ đoán hoặc sử dụng rộng rãi.

Mật khẩu liên quan đến thông tin cá nhân, dễ dàng thu thập từ các mạng xã hội như tên, ngày sinh.

Máy chủ không thiết lập giới hạn số lần nhập sai, tạo điều kiện để hacker thực hiện hàng loạt lần thử mật khẩu.

Tấn công Brute Force thường được chưa làm các dạng sau đây:

**Simple Brute Force Attacks**: Hacker cố gắng đoán một cách hợp lý thông tin đăng nhập của bạn – hoàn toàn không được hỗ trợ từ các công cụ phần mềm hoặc các phương tiện khác. Chúng có thể tiết lộ mật khẩu và mã PIN đơn giản. Ví dụ: mật khẩu được đặt là “guest12345”.

**Dictionary Attacks:** Các cuộc tấn công từ điển là công cụ cơ bản nhất trong các cuộc tấn công brute force. Mặc dù không nhất thiết phải là các cuộc tấn công Brute Force, nhưng chúng thường được sử dụng như một thành phần quan trọng để bẻ khóa mật khẩu. Một số tin tặc chạy qua các từ điển không kết hợp và bổ sung các từ bằng các ký tự và chữ số đặc biệt hoặc sử dụng các từ điển từ đặc biệt, nhưng kiểu tấn công tuần tự này rất phức tạp.

**Hybrid Brute Force Attacks:** Hacker lợi dụng các thông tin bên ngoài và của bạn để một cách logic của họ để cố gắng lấy thông tin đăng nhập. Một cuộc tấn công hỗn hợp thường kết hợp các cuộc tấn công từ điển và brute force. Các cuộc tấn công này được sử dụng để tìm ra mật khẩu kết hợp trộn các từ phổ biến với các ký tự ngẫu nhiên. Một ví dụ về cuộc tấn công vũ phu về bản chất này sẽ bao gồm các mật khẩu như NewYork1993 hoặc Spike1234.

**Reverse Brute Force Attacks:** Đây là một hình thức tấn công đảo ngược chiến lược tấn công bằng cách bắt đầu với một mật khẩu đã biết. Sau đó, hacker tìm kiếm hàng triệu tên người dùng cho đến khi họ tìm thấy một kết quả trùng khớp. Nhiều tên tội phạm trong số này bắt đầu với mật khẩu bị rò rỉ có sẵn trực tuyến từ các vi phạm dữ liệu hiện có.

**Credential Stuffing:** Nếu một hacker có tổ hợp tên người dùng và mật khẩu hoạt động cho một trang web, họ cũng sẽ thử nó cho rất nhiều trang web khác. Vì người dùng đã được biết là sử dụng lại thông tin đăng nhập trên nhiều trang web, họ là mục tiêu độc quyền của một cuộc tấn công như thế này.

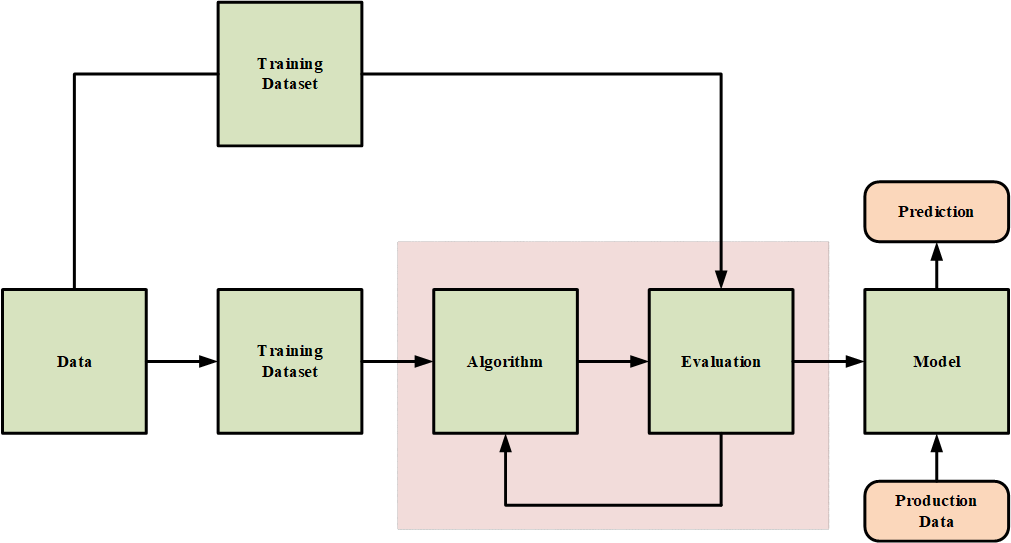
Mặc dù brute-force là hình thức tấn công dễ phòng chống thông qua các biện pháp bảo mật cơ bản, nhưng trên thực tế, rất nhiều hệ thống vẫn dễ dàng trở thành mục tiêu bởi sự chủ quan trong việc áp dụng các biện pháp bảo vệ cần thiết. Điều này nhấn mạnh tầm quan trọng của việc nâng cao nhận thức về bảo mật mật khẩu, triển khai giới hạn số lần nhập sai, và sử dụng các phương pháp bảo vệ nâng cao để ngăn chặn tấn công brute-force hiệu quả.

2.2 Học máy

“Lĩnh vực nghiên cứu mang lại cho máy tính khả năng học hỏi mà không cần thiết lập rõ ràng” – theo Arthur Samuel - người tiên phong trong lĩnh vực trí tuệ nhân tạo- đã định nghĩa về học máy (Machine learning).

Machine learning (ML) là một nhánh của trí tuệ nhân tạo, sử dụng các thuật toán được đào tạo trên các bộ dữ liệu để tạo ra các mô hình tự học có khả năng dự đoán kết quả và phân loại thông tin mà không cần sự can thiệp của con người. ML bắt đầu bằng việc cung cấp dữ liệu và sau đó đào tạo máy tính bằng cách xây dựng các mô hình học máy bằng cách sử dụng dữ liệu và các thuật toán khác nhau.

Ý tưởng cơ bản của mọi quy trình học máy là xây dựng mô hình dựa trên một số thuật toán để thực hiện một nhiệm vụ cụ thể như phân loại, phân lớp, hồi quy... Giai đoạn huấn luyện được thực hiện dựa trên dữ liệu đầu vào và mô hình được xây dựng để dự đoán đầu ra. Kết quả đầu ra phụ thuộc mục tiêu ban đầu và việc thực hiện. Chi tiết quy trình học máy gồm các bước như sau:



Hình 1. Quy trình học máy

Quy trình học máy cơ bản được chia làm các giai đoạn sau:

Thu thập dữ liệu (gathering data): Quá trình thu thập dữ liệu phụ thuộc vào loại dự án mà chúng ta mong muốn xây dựng, ví dụ nếu chúng ta muốn xây dựng dự án học máy mà sử dụng dữ liệu thực để chúng ta có thể xây dựng một hệ thống IoT từ các dữ liệu cảm biến khác nhau. Dữ liệu chúng ta có thể thu thập từ các nguồn dữ liệu khác nhau như một tập tin, cơ sở dữ liệu, cảm biến ...

Tiền xử lý dữ liệu (data pre-processing): Tiền xử lý dữ liệu là một trong những giai đoạn quan trọng trong học máy, nó giúp xây dựng mô hình học máy chính xác. Tiền xử lý dữ liệu là một quá trình làm sạch dữ liệu thô, dữ liệu được thu thập từ nhiều nguồn trong thế giới thực và được chuyển thành một tập dữ liệu sạch. Dữ liệu thô ban đầu có một số đặc điểm như dữ liệu bị thiếu sót, không nhất quán, nhiễu vì vậy dữ liệu này phải được xử lý trước khi đưa vào học máy.

Xây dựng mô hình phù hợp cho loại dữ liệu (researching model): Mục tiêu chính của chúng ta là xây dựng mô hình thực hiện tốt nhất dựa trên một số thuật toán phân loại và phân lớp.

Huấn luyện và kiểm thử mô hình trên dữ liệu (training and testing model): để huấn luyện một mô hình, ban đầu chúng ta chia mô hình thành 03 giai đoạn bao gồm: dữ liệu huấn luyện (training data), dữ liệu xác nhận (validation data) và dữ liệu kiểm thử (testing data). Để huấn luyện bộ phân lớp ta sử dụng tập hợp dữ liệu huấn luyện (training data set), để tinh chỉnh các tham số ta sử dụng tập hợp xác nhận (validation set) và sau đó kiểm tra hiệu suất của bộ phân loại chưa biết sử dụng tập hợp dữ liệu kiểm thử (test data set). Một lưu ý quan trọng là trong quá trình huấn luyện bộ phân lớp là dữ liệu kiểm thử không được sử dụng để huấn luyện.

Đánh giá (evaluation): Đánh giá mô hình là một phần quan trọng trong quy trình phát triển mô hình, nó giúp tìm ra mô hình tốt nhất để đại diện cho dữ liệu của chúng ta và mô hình được chọn sẽ hoạt động tốt như thế nào trong tương lai.

Có 3 loại máy học được phân loại dựa theo mục tiêu và bộ dữ liệu, bao gồm học có giám sát (Supervised Learning), học không giám sát (Unsupervised Learning) và học tăng cường (Reinforcement Learning).

Học có giám sát (Supervised Learning): Học có giám sát là mô hình học máy có bộ dữ liệu đầu vào được gọi là dữ liệu huấn luyện (training data) được gắn nhãn hoặc kết quả đã biết làm đầu ra. Vì vậy, nó hoạt động theo nguyên tắc cặp đầu vào-đầu ra. Học có giám sát là loại mô hình dựa trên nhiệm vụ và tập dữ liệu phải được dán nhãn.

Học không giám sát (Unsupervised Learning): Các mô hình Học máy không giám sát thực hiện quy trình học ngược lại với học có giám sát, có nghĩa là nó cho phép mô hình học từ tập dữ liệu huấn luyện không được gắn nhãn. Nó được đào tạo để nhóm các đối tượng theo các đặc điểm chung.

Học tăng cường (Reinforcement Learning): Trong học tăng cường, thuật toán được tạo ra để tự đào tạo bằng nhiều thử nghiệm thử và sai. Thuật toán học các hành động đối với một tập hợp các trạng thái nhất định dẫn đến trạng thái mục tiêu. Đây là mô hình học tập dựa trên phản hồi, nhận tín hiệu phản hồi sau mỗi trạng thái hoặc hành động bằng cáchtương tác với môi trường. Học tăng cường xảy ra khi thuật toán tương tác liên tục với môi trường thay vìdựa vào dữ liệu huấn luyện. Một trong những ví dụ phổ biến nhất về học tăng cường là lái xe tự động.

Mô hình học máy là một tệp đã được đào tạo để nhận dạng các mẫu trong 1 tệp dữ liệu hoặc đưa ra dự đoán. Các mô hình học máy được tạo từ các thuật toán học máy, trải qua quá trình đào tạo bằng cách sử dụng các tệp dữ liệu mà chúng ta cung cấp cho nó. Sau khi huấn luyện mô hình, ta có thể sử dụng nó để suy luận về dữ liệu mà nó chưa từng thấy trước đó và đưa ra dự đoán về những dữ liệu đó.

Ví dụ, học máy có thể được sử dụng để phân tích các mẫu hành vi bất thường trong quá trình đăng nhập. Ví dụ, một hệ thống học máy có thể học được tần suất đăng nhập thông thường của người dùng và phát hiện các lần đăng nhập liên tục, tốc độ cao, hoặc sử dụng nhiều tổ hợp mật khẩu - dấu hiệu của tấn công brute-force.

Có hai loại mô hình học máy bao gồm mô hình phân loại (Classification) và mô hình hồi quy (Regression). Đôi khi, cùng một thuật toán có thể được sử dụng để tạo mô hình phân loại hoặc hồi quy, tùy thuộc vào cách nó được huấn luyện.

Mô hình phân loại (Classification)Mô hình phân loại được sử dụng khá rộng rãi trong ML để đưa ra kết luận từ các giá trị được quan sát ở dạng phân loại. Khi tệp dữ liệu đầu vào có các nhãn là 1 số hữu hạn các nhóm thì ta sẽ sử dụng classification.

Ví dụ: Một mô hình phân loại có thể được huấn luyện để phân biệt giữa tập tin "hợp lệ" và "mã độc" dựa trên các đặc trưng như kích thước tập tin, các hành động mà tập tin thực hiện khi được chạy, hoặc các thuộc tính tĩnh khác. Phân loại tập tin thành các nhóm như: malware (mã độc), adware (quảng cáo độc hại), hoặc benign (không độc hại). Một số mô hình học máy phổ biến: Decision Tree, Random Forest, hoặc Deep Neural Network.

Hay mô hình phân loại sử dụng các đặc trưng từ email như tiêu đề, nội dung, hoặc các liên kết trong email để xác định email đó là hợp lệ hay lừa đảo. Phân loại email thành các nhóm: Phishing Email (email lừa đảo) hoặc Legitimate Email (email hợp lệ). Các thuật toán thường dùng: Support Vector Machine (SVM), Gradient Boosting Machines.

Mô hình hồi quy (Regression).

Trong lĩnh vực an toàn thông tin, mô hình hồi quy (Regression) đóng vai trò quan trọng trong việc dự đoán các giá trị liên tục, khác với các bài toán phân loại (Classification) vốn tập trung vào các nhãn cố định. Hồi quy được sử dụng để xử lý các bài toán mà đầu ra là các giá trị theo khoảng, giúp dự đoán hoặc ước lượng các thông số an ninh quan trọng. Một số ứng dụng tiêu biểu bao gồm:

Ví dụ, sử dụng mô hình hồi quy để dự đoán số lượng yêu cầu bất thường trong một khoảng thời gian dựa trên các yếu tố như lưu lượng mạng, số lượng người dùng, hoặc tần suất truy cập Giúp phát hiện sự gia tăng bất thường của lưu lượng mạng, thường liên quan đến các cuộc tấn công từ chối dịch vụ (DDoS). Hay sử dụng các đặc trưng như độ phức tạp của mật khẩu, tốc độ thử nghiệm của hệ thống, và khả năng xử lý của hacker để dự đoán thời gian cần thiết để hoàn tất cuộc tấn công brute-force, giúp các hệ thống bảo mật đưa ra cảnh báo và tăng cường biện pháp phòng thủ kịp thời.

Các mô hình hồi quy, bao gồm Linear Regression, Random Forest Regression, và Neural Network Regression, giữ vai trò quan trọng trong việc nâng cao hiệu quả dự đoán và phân tích dữ liệu trong lĩnh vực an toàn thông tin. Nhờ khả năng xử lý các giá trị liên tục và đưa ra kết quả chính xác, các mô hình này không chỉ hỗ trợ phát hiện sớm các mối đe dọa mà còn giúp tối ưu hóa chiến lược phòng thủ, góp phần quan trọng trong việc đảm bảo an ninh mạng toàn diện.

2.3 Mô hình Random Forest (RF)

Random Forest được lấy ý tưởng cũng là một trong bài toán cây quyết định (decision tree), đây là thuật toán học có giám sát (supervised learning). Các bài toán phân lớp, hồi quy thường áp dụng Random Forest để xử lý, vì nó dễ sử dụng và linh hoạt. RF ý tưởng tạo ra nhiều cây quyết định từ các mẫu được chọn ngẫu nhiên và đưa ra đánh giá ở mỗi cây, kết quả lựa chọn bằng cách bỏ phiếu (vote) chọn ra các cây có kết quả tốt nhất, từ đó sẽ tìm ra các tính năng (feature) tốt nhất có yêu tố quyết định trong phân lớp hoặc hồi quy.

Random Forest cải thiện được những hạn chế mà k-NN gặp phải đó là k-NN coi các tính năng đều có giá trị như nhau dẫn đến mô hình học, học tất cả các tính năng từ tập dữ liệu ảnh hưởng lớn đến quá trình dự đoán. Còn Random forest lựa chọn tìm ra những tính năng tốt để học “không học đại trà, tràn lan” vì vậy kết dự đoán tốt hơn trong quá trình dự đoán một tập dữ liệu mới. Random Forest đưa ra lựa chọn dựa trên giá trị trung bình từ các dự đoán, nó sẽ tránh được các dự đoán có thiên hướng sai lệch nghiêng về nhãn có số lượng nhiều, điều này giúp cho thuật toán tránh trường hợp overfitting (mất cân bằng nhãn).

Thuật toán hoạt động gồm 4 bước:

Bước 1: Chọn k tính năng ngẫu nhiên trong tập dữ liệu gồm n tính năng (k<=n)

Bước 2: Thiết lập cây quyết định cho k tính năng được chọn

Bước 3: Bỏ phiếu cho cây quyết định theo từng tính năng được tạo theo giá trị trung bình của mỗi dự đoán.

Bước 4: Chọn cây quyết định có giá trị votting lớn nhất.

Random Forest là một thuật toán mạnh mẽ, đáng tin cậy, và dễ sử dụng trong các bài toán an toàn thông tin. Với khả năng xử lý dữ liệu phức tạp và đưa ra dự đoán chính xác, nó đã trở thành một công cụ không thể thiếu trong việc bảo vệ hệ thống và dữ liệu khỏi các mối đe dọa ngày càng tinh vi.

2.4 Mô hình Support Vector Machine (SVM)

SVM (support vector machine) là một khái niệm trong thống kê và khoa học máy tính cho một tập hợp các phương pháp học có giám sát liên quan đến nhau để phân loại và phân tích hồi quy. SVM dạng chuẩn nhận dữ liệu đầu vào và phân loại chúng vào hai lớp khác nhau. Do đó SVM là một thuật toán phân loại nhị phân. Với một bộ các ví dụ huấn luyện thuộc hai thể loại cho trước, thuật toán huấn luyện SVM xây dựng một mô hình SVM để phân loại các ví dụ khác vào hai thể loại đó. Một mô hình SVM là một cách biểu diễn các điểm trong không gian và lựa chọn ranh giới giữa hai thể loại sao cho khoảng cách từ các ví dụ huấn luyện tới ranh giới là xa nhất có thể. Các ví dụ mới cũng được biểu diễn trong cùng một không gian và được thuật toán dự đoán thuộc một trong hai thể loại tùy vào ví dụ đó nằm ở phía nào của ranh giới.

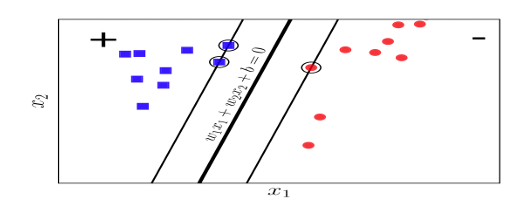
SVM xây dựng một siêu phẳng hoặc một tập hợp các siêu phẳng trong một không gian nhiều chiều hoặc vô hạn chiều, có thể được sử dụng cho phân loại, hồi quy, hoặc các nhiệm vụ khác. Một cách trực giác, để phân loại tốt nhất thì các siêu phẳng nằm ở càng xa các điểm dữ liệu của tất cả các lớp (gọi là hàm lề) càng tốt, vì nói chung lề càng lớn thì sai số tổng quát hóa của thuật toán phân loại càng bé.

Trong nhiều trường hợp, không thể phân chia các lớp dữ liệu một cách tuyến tính trong một không gian ban đầu được dùng để mô tả một vấn đề. Vì vậy, nhiều khi cần phải ánh xạ các điểm dữ liệu trong không gian ban đầu vào một không gian mới nhiều chiều hơn, để việc phân tách chúng trở nên dễ dàng hơn trong không gian mới. Để việc tính toán được hiệu quả, ánh xạ sử dụng trong thuật toán SVM chỉ đòi hỏi tích vô hướng của các vector dữ liệu trong không gian mới có thể được tính dễ dàng từ các tọa độ trong không gian cũ. Tích vô hướng này được xác định bằng một hàm hạt nhân K(x,y) phù hợp. [1] Một siêu phẳng trong không gian mới được định nghĩa là tập hợp các điểm có tích vô hướng với một vectơ cố định trong không gian đó là một hằng số. Vector xác định một siêu phẳng sử dụng trong SVM là một tổ hợp tuyến tính của các vector dữ liệu luyện tập trong không gian mới với các hệ số αi. Với siêu phẳng lựa chọn như trên, các điểm x trong không gian đặc trưng được ánh xạ vào một siêu mặt phẳng là các điểm thỏa mãn: Σi αi K(xi,x) = hằng số.

Ghi chú rằng nếu K(x,y) nhận giá trị ngày càng nhỏ khi y xa dần khỏi x thì mỗi số hạng của tổng trên được dùng để đo độ tương tự giữa x với điểm xi tương ứng trong dữ liệu luyện tập. Như vậy, tác dụng của tổng trên chính là so sánh khoảng cách giữa điểm cần dự đoán với các điểm dữ liệu đã biết. Lưu ý là tập hợp các điểm x được ánh xạ vào một siêu phẳng có thể có độ phức tạp tùy ý trong không gian ban đầu, nên có thể phân tách các tập hợp thậm chí không lồi trong không gian ban đầu.

Phân loại thống kê là một nhiệm vụ phổ biến trong học máy. Trong mô hình học có giám sát, thuật toán được cho trước một số điểm dữ liệu cùng với nhãn của chúng thuộc một trong hai lớp cho trước. Mục tiêu của thuật toán là 23 xác định xem một điểm dữ liệu mới sẽ được thuộc về lớp nào. Mỗi điểm dữ liệu được biểu diễn dưới dạng một vector p chiều và ta muốn biết liệu có thể chia tách hai lớp dữ liệu bằng một siêu phẳng p − 1 chiều, đây gọi là phân loại tuyến tính. Có nhiều siêu phẳng có thể phân loại được dữ liệu. Một lựa chọn hợp lý trong chúng là siêu phẳng có lề lớn nhất giữa hai lớp.

Xây dựng bài toán SVM Giả sử rằng các cặp dữ liệu của training set là (𝒙𝟏,𝒚𝟏), (𝒙𝟐,𝒚𝟐),.., (𝒙𝟑,𝒚𝟑) với vector 𝒙𝒊 𝞊 𝑹𝒅 thể hiện đầu vào của một điểm dữ liệu và 𝒚𝒊 là nhãn của điểm dữ liệu đó, d là số chiều của dữ liệu và N là số điểm dữ liệu. Giả sử rằng nhãn của mỗi điểm dữ liệu được xác định bởi 𝒚𝒊 = 1 (lớp 1) hoặc 𝒚𝒊 = -1 (lớp 2). Để dễ hình dung, chúng ta cùng xét trường hợp trong không gian hai chiều dưới đây. Không gian hai chiều để dễ hình dung, các phép toán hoàn toàn có thể được tổng quát lên không gian nhiều chiều.



Hình 2. 1 Phân tích bài toán SVM

Giả sử rằng các điểm vuông xanh thuộc lớp 1, các điểm tròn đỏ thuộc lớp -1 và mặt (𝑤𝑇𝑥 + 𝑏) = 𝑤1𝑥1 +𝑤2𝑥2 + 𝑏 là mặt phân chia giữa hai lớp. Hơn nữa, lớp 1 nằm về phía dương, lớp -1 nằm về phía âm của mặt phân chia. Nếu ngược lại, ta chỉ cần đổi dấu của w và b. Chú ý rằng chúng ta cần đi tìm các hệ số w và b. Ta quan sát thấy một điểm quan trọng sau đây: với cặp dữ liệu (𝑥𝑛,𝑦𝑛) bất kỳ, khoảng cách từ điểm đó tới mặt phân chia là:

Điều này có thể dễ nhận thấy vì theo giả sử ở trên, yn luôn cùng dấu với phía của 𝑥𝑛. Từ đó suy ra yn cùng dấu với (𝑤𝑇xn + b) và tử số luôn là 1 số không âm. Với mặt phần chia như trên, giới hạn (margin) được tính là khoảng cách gần nhất từ 1 điểm tới mặt đó (bất kể điểm nào trong hai lớp):

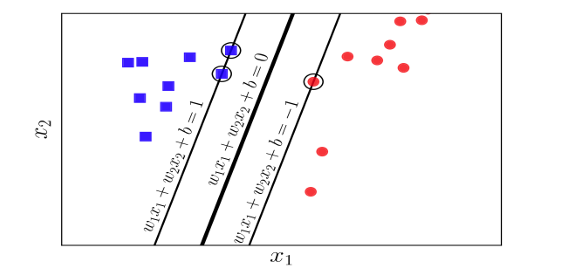
Bài toán tối ưu trong SVM chính là bài toán tìm w và b sao cho margin này đạt giá trị lớn nhất:

(w,b) = arg max(1)

Nhận xét quan trọng nhất là nếu ta thay vector hệ số w bởi kw và b bởi kb trong đó k là một hằng số dương thì mặt phân chia không thay đổi, tức khoảng cách từ từng điểm đến mặt phân chia không đổi, tức margin không đổi. Dựa trên tính chất này, ta có thể giả sử:

**Y**𝒏(**W**𝑻𝒙𝒏 + b) = 1

Với những điểm nằm gần mặt phân chia nhất như hình:



Hình 2. 2 Các điểm gần mặt phân cách nhất của hai class được khoanh tròn

Như vậy, với mọi n, ta có:

**Y**𝒏(**W**𝑻𝒙𝒏 + b) = 1

Vậy bài toán tối ưu (1) có thể đưa về bài toán tối ưu có ràng buộc sau đây:

(w,b) = arg

Subject to: yn (wTxn + b) 1, (2)

Bằng một biến đổi đơn giản, ta có thể đưa bài toán này về bài toán dưới đây:

(w,b) = arg

Subject to: 1-yn (wTxn + b) 0, (3)

Ở đây, chúng ta đã lấy nghịch đảo hàm mục tiêu, bình phương nó để được một hàm khả vi và nhân với 1/2 để biểu thức đạo hàm đẹp hơn. Xác định lớp (class) cho một điểm dữ liệu mới: sau khi tìm được mặt phân cách 𝑾𝑻𝒙+ 𝒃 =𝟎 , class của bất kỳ một điểm nào sẽ được xác định đơn giản bằng cách:

class(x) = sgn(𝑤𝑻x + b)

trong đó hàm sgn là hàm xác định dấu, nhận giá trị 1 nếu đối số là không âm và -1 nếu ngược lại.

CHƯƠNG 3: DỮ LIỆU VÀ THỰC NGHIỆM

3.1 Phát biểu bài toán

Trong bối cảnh an ninh mạng hiện nay, việc phát hiện và ngăn chặn các cuộc tấn công mạng ngày càng trở nên cấp thiết. Một trong những hình thức tấn công phổ biến là tấn công brute force, đặc biệt là những cuộc tấn công nhắm vào giao thức SSH và FTP. Để đối phó với thách thức này, cần xây dựng một hệ thống phát hiện tấn công tự động sử dụng machine learning.

Bài toán đặt ra yêu cầu phát triển một mô hình machine learning có khả năng phát hiện chính xác các cuộc tấn công brute force trên hai giao thức SSH và FTP. Mô hình cần được xây dựng và đánh giá dựa trên bộ dữ liệu chuẩn CSE-CIC-IDS2018, bao gồm hơn 80 đặc trưng lưu lượng mạng được thu thập từ môi trường AWS LAN. Dữ liệu sẽ được tiền xử lý để loại bỏ nhiễu và chuẩn hóa, sau đó phân chia thành các tập huấn luyện (60%), validation (20%) và kiểm thử (20%).

Nghiên cứu sẽ thực hiện đánh giá và so sánh hiệu suất của nhiều thuật toán machine learning khác nhau như: Random Forest, Support Vector Machine.

Mục tiêu cuối cùng là xây dựng được một hệ thống phát hiện tấn công brute force tự động với độ chính xác cao, tỷ lệ cảnh báo sai thấp và có khả năng hoạt động hiệu quả trong thời gian thực. Hệ thống cần đủ tin cậy để có thể triển khai trong môi trường mạng thực tế, góp phần bảo vệ các hệ thống khỏi các cuộc tấn công brute-force nhắm vào giao thức SSH và FTP.

3.2 Dữ liệu thực nghiệm

Để thực hiện nghiên tiểu luận phát hiện tấn công brute-force bằng học máy, nhóm đã sử dụng bộ dữ liệu CSE-CIC-IDS2018. Bộ dữ liệu này được tạo ra thông qua một dự án hợp tác giữa Communications Security Establishment (CSE) và Canadian Institute for Cybersecurity (CIC), nhằm phát triển một hệ thống bảo mật mạng hiệu quả.

Dữ liệu này bao gồm 7 kịch bản tấn công (Brute-force, Heartbleed, Botnet, DoS, DDoS, tấn công web, và xâm nhập từ bên trong), với 50 máy trong cơ sở hạ tầng tấn công và 420 máy tính cùng 30 máy chủ của tổ chức nạn nhân. Bộ dữ liệu chứa lưu lượng mạng và nhật ký từ các máy, cùng 80 tính năng lưu lượng được trích xuất bằng CICFlowMeter-V3.

Link tải: <https://registry.opendata.aws/cse-cic-ids2018/>

Ưu điểm của bộ dữ liệu này đó là mô phỏng môi trường mạng thực tế, có nhiều loại tấn công khác nhau, bộ dữ liệu này cung cấp nguồn dữ liệu chi tiết và đa chiều. Ngoài ra bộ dữ liệu này được xây dựng bởi các chuyên gia an ninh mạng và được công khai để các nhà nghiên cứu có thể áp dụng vào nghiên cứu các giải pháp an ninh mạng.

Bộ dữ liệu CSE-CIC-IDS2018 là một bộ dữ liệu quan trọng cho việc phát triển các giải pháp phát hiện tấn công sử dụng học máy, cho phép nhóm mô phỏng và phân tích các kịch bản tấn công phức tạp trong môi trường mạng.

Việc lựa chọn bộ dữ liệu này giúp cho nhóm có được bối cảnh thực tế và toàn diện, tạo điều kiện để phát triển các mô hình học máy có khả năng phát hiện và ứng phó với các loại tấn công khác nhau.

3.3 Xử lý dữ liệu

Quá trình xử lý dữ liệu bao gồm các bước chính:

* Làm sạch dữ liệu
* Trích xuất và lựa chọn đặc trưng
* Chuẩn hóa dữ liệu
* Xử lý mất cân bằng lớp

3.4 Cài đặt mô hình

Để cài đặt mô hình phát hiện tấn công Brute Force vào giao thức SSH và FTP được đề xuất, thực hiện các bước sau:

**3.4.1. Chuẩn bị môi trường:**

Cài đặt Python: Mô hình được phát triển bằng Python.

Cài đặt các thư viện cần thiết: Bao gồm Keras, TensorFlow, Scikit-learn, Pandas, NumPy. Bạn có thể sử dụng pip để cài đặt:

**3.4.2. Tải xuống tập dữ liệu CSE-CIC-IDS2018:**

Tập dữ liệu CSE-CIC-IDS2018.

Link tải: <https://registry.opendata.aws/cse-cic-ids2018/>

Chọn tập dữ liệu liên quan đến các cuộc tấn công Brute Force vào SSH và FTP.

**3.4.3. Tiền xử lý dữ liệu:**

Làm sạch dữ liệu: Xử lý các giá trị thiếu, ngoại lai và các lỗi khác trong tập dữ liệu.

Chuẩn hóa dữ liệu: Chuyển đổi các giá trị đặc trưng số về một thang đo chung, chẳng hạn như khoảng1.

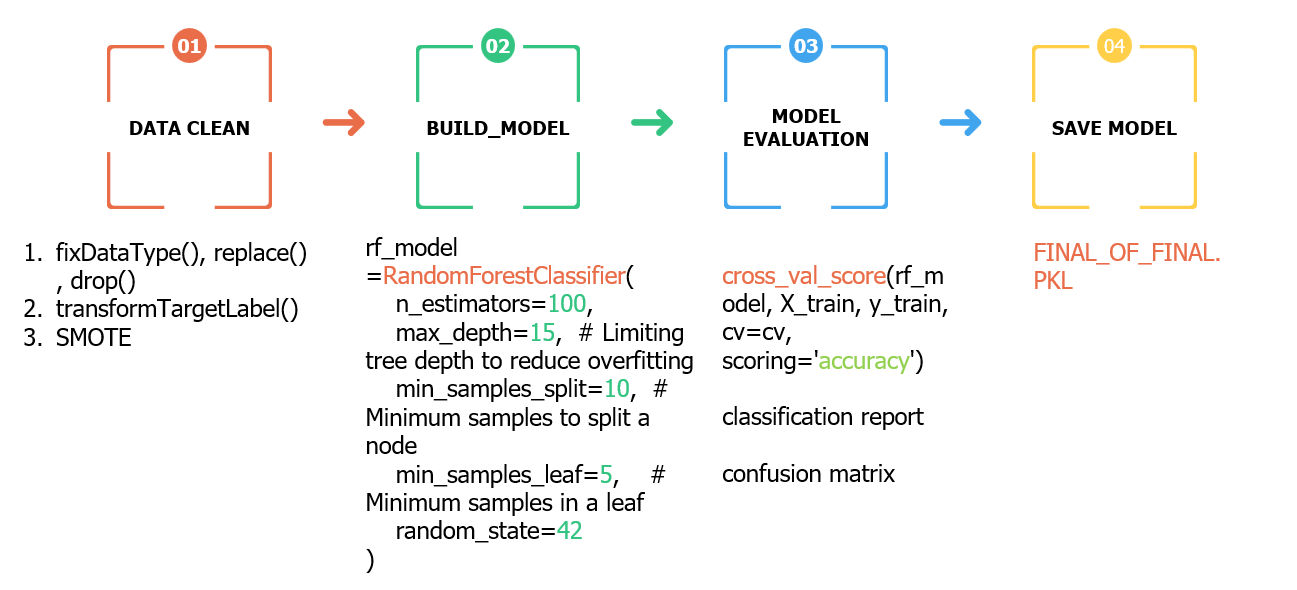
Chia dữ liệu: Chia tập dữ liệu thành ba phần: huấn luyện (60%), xác thực (20%) và kiểm tra (20%).

**3.4.4. Xây dựng mô hình Random Forest:**

Sử dụng thư viện Scikit-learn để tạo một mô hình Random Forest.

Điều chỉnh siêu tham số: Tối ưu hóa các siêu tham số của mô hình, chẳng hạn như số lượng cây quyết định, độ sâu tối đa của cây, v.v. để đạt được hiệu suất tốt nhất. Bạn có thể sử dụng kỹ thuật tìm kiếm lưới hoặc tìm kiếm ngẫu nhiên.

Huấn luyện mô hình: Sử dụng tập dữ liệu huấn luyện để huấn luyện mô hình.



Hình 3.4.4 Sơ đồ quá trình Training theo mô hình Random Forest

**3.4.5. Xây dựng mô hình SVM:**

Sử dụng thư viện Scikit-learn để tạo một mô hình SVM.

Nhập các thư viện cần thiết như NumPy, Pandas, và các thành phần từ Scikit-learn để xử lý dữ liệu và xây dựng mô hình.

Tiếp theo, bạn tải dữ liệu và chia nó thành đặc trưng (X) và nhãn (y). Sau đó, sử dụng phương thức train\_test\_split để chia dữ liệu thành tập huấn luyện và tập kiểm tra. Để tối ưu hóa các siêu tham số của mô hình SVM, bạn định nghĩa một lưới các tham số cần điều chỉnh, bao gồm tham số C, loại kernel và tham số gamma.

Sử dụng GridSearchCV, bạn thực hiện tìm kiếm lưới để tìm ra các siêu tham số tốt nhất cho mô hình.

Cuối cùng, đánh giá hiệu suất của mô hình trên tập kiểm tra bằng cách dự đoán nhãn và sử dụng các chỉ số như báo cáo phân loại và ma trận nhầm lẫn để phân tích kết quả.

A diagram of a flowchart

Description automatically generated

Hình 3.4.5 Quy trình xử lý dữ liệu theo mô hình SVM

**3.4.5. Đánh giá mô hình:**

Sử dụng tập dữ liệu kiểm tra để đánh giá hiệu suất của mô hình.

Tính toán các chỉ số đánh giá: Bao gồm độ chính xác, độ chính xác, khả năng thu hồi và điểm F1.

Phân tích ma trận nhầm lẫn: Để hiểu rõ hơn về các lỗi phân loại của mô hình.

**3.4.6**. **Triển khai mô hình:**

Sau khi mô hình được huấn luyện và đánh giá, bạn có thể triển khai mô hình trong môi trường thực tế.

Tích hợp mô hình vào hệ thống phát hiện xâm nhập (IDS) hoặc hệ thống bảo mật khác.

CHƯƠNG 4: ĐÁNH GIÁ KẾT QUẢ

4.1 Đánh giá kết quả

Trong bối cảnh công nghệ thông tin ngày càng phát triển, vấn đề an ninh mạng trở thành một trong những ưu tiên hàng đầu của các tổ chức và doanh nghiệp. Các cuộc tấn công Brute-Force, với mục tiêu chiếm đoạt thông tin đăng nhập, ngày càng trở nên phổ biến và tinh vi. Bài tiểu luận nghiên cứu phát hiện tấn công Brute-Force bằng học máy nhằm tìm hiểu cũng như phát triển một số mô hình máy học để phát hiện và ngăn chặn những cuộc tấn công này. Nội dung tiểu luận sẽ đánh giá kết quả đạt được trong dự án, từ độ chính xác của mô hình cho đến khả năng ứng dụng thực tế.

Trong phần nghiên cứu, nhóm đã sử dụng hai mô hình chính: Random Forest và SVM. Cả hai mô hình này đã được huấn luyện và tối ưu hóa với dữ liệu thực tế, cho thấy chúng có khả năng phân loại chính xác các mẫu dữ liệu liên quan đến cuộc tấn công Brute Force. Đặc biệt, mô hình Random Forest đã đạt được độ chính xác cao, điều này chứng tỏ khả năng phát hiện hiệu quả của nó trong các tình huống thực tế.



Hình 4. 1 Kết quả thực nghiệm trên mô hình Random Forest

Bài tiểu luận của nhóm không chỉ dừng lại ở việc phát triển một mô hình duy nhất mà còn áp dụng nhiều thuật toán khác nhau. Việc sử dụng cả Random Forest và SVM cho thấy một chiến lược tiếp cận đa dạng, cho phép so sánh hiệu quả giữa các phương pháp khác nhau. Điều này không chỉ giúp người dùng lựa chọn mô hình phù hợp với yêu cầu cụ thể mà còn mở ra cơ hội cho việc cải tiến và phát triển các phương pháp mới trong tương lai.

Các mô hình mà nhóm nghiên cứu áp dụng có thể được triển khai trong các hệ thống giám sát an ninh mạng, giúp phát hiện sớm và ứng phó kịp thời với các cuộc tấn công Brute Force. Điều này không chỉ nâng cao tính bảo mật cho các hệ thống thông tin hiện đại mà còn giúp giảm thiểu thiệt hại do các cuộc tấn công gây ra. Khi người dùng có thể áp dụng những công nghệ tiên tiến này vào thực tiễn, sự an toàn của thông tin sẽ được đảm bảo hơn.

4.2 Kết luận

Bài tiểu luận nghiên cứu phát hiện tấn công Brute-Force bằng học máy đã thể hiện rõ tiềm năng của công nghệ máy học trong việc nâng cao an ninh mạng, đặc biệt là trong việc phát hiện và ngăn chặn các cuộc tấn công Brute Force. Qua việc phát triển và tối ưu hóa các mô hình như Random Forest và SVM, dự án không chỉ mang đến các giải pháp hiệu quả mà còn tạo ra một nền tảng vững chắc cho các nghiên cứu và ứng dụng tiếp theo trong lĩnh vực này.

Độ chính xác cao của các mô hình cho thấy khả năng nhận diện các mối đe dọa một cách kịp thời, giúp bảo vệ thông tin quan trọng của tổ chức và cá nhân. Sự đa dạng trong phương pháp tiếp cận cho phép người dùng linh hoạt lựa chọn giải pháp phù hợp nhất với nhu cầu cụ thể của họ, đồng thời mở rộng khả năng nghiên cứu thêm nhiều thuật toán mới trong tương lai.

Hơn nữa, tính dễ dàng trong việc triển khai và sử dụng các mô hình đã được huấn luyện giúp cho ngay cả những người không chuyên môn cũng có thể áp dụng công nghệ hiện đại này vào việc bảo mật hệ thống của mình. Điều này không chỉ tạo điều kiện thuận lợi cho việc bảo vệ an ninh thông tin mà còn góp phần xây dựng một môi trường mạng an toàn hơn cho tất cả mọi người.

Cuối cùng, những kết quả đạt được từ tìm hiểu và nghiên cứu khẳng định rằng việc ứng dụng máy học trong an ninh mạng không chỉ là xu hướng mà còn là một bước đi tất yếu trong việc bảo vệ dữ liệu trong kỷ nguyên số. Với những giải pháp như vậy, chúng ta đang tiến gần hơn đến một tương lai an toàn hơn, nơi mà công nghệ có thể giúp chúng ta đối phó hiệu quả với những thách thức ngày càng phức tạp trong lĩnh vực bảo mật thông tin.