TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

**Sử dụng kỹ thuật máy học để phát hiện**

**lừa đảo trên trang web**

*Người hướng dẫn*: **TS. TRƯƠNG ĐÌNH TÚ**

*Người thực hiện*: **VOÒNG QUANG MẠNH – 52000573**

**NGÔ TƯỜNG VI – 52000430**

Lớp **: 20050401**

Khoá  **: 24**

**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**



**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

**Sử dụng kỹ thuật máy học để phát hiện**

**lừa đảo trên trang web**

Người hướng dẫn: **TS. TRƯƠNG ĐÌNH TÚ**

Người thực hiện: **VOÒNG QUANG MẠNH**

**NGÔ TƯỜNG VI**

Lớp **: 20050401**

Khoá  **: 24**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

# LỜI CẢM ƠN

Sau khoảng thời gian nghiên cứu và phát triển, chúng em đã hoàn thành được dự án công nghệ thông tin. Dự án này được hoàn thành tốt đẹp không chỉ nhờ vào công sức của chúng em mà còn nhờ vào sự hướng dẫn và chỉ bảo tận tình của TS.Trương Đình Tú. Chúng em xin gửi lời cảm ơn chân thành và sâu sắc đến thầy. Ngoài ra, chúng em cũng muốn gửi lời cảm ơn đến các thầy/cô trong Khoa công nghệ thông tin trường ĐH Tôn Đức Thắng đã trao dồi kiến thức nền tảng để chúng em triển khai cho dự án.

Chúng em xin chân thành cảm ơn.

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi / chúng tôi và được sự hướng dẫn của TS.Trương Đình 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 đồ án 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 tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 8 tháng 3 năm 2024*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Voòng Quang Mạnh*

*Ngô Tường Vi*

**PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN**

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

# TÓM TẮT

Ngày nay công nghệ phát triển ngày càng nhanh, song song với đó là nhiều cuộc tấn công lừa đảo chiếm đoạt thông tin người dùng thông qua cách thức giả mạo trang web chính thống. Đối tượng được hướng đến nhiều đó là những người không am hiểu về công nghệ. Vì thế để có thể góp phần cảnh báo đến người dùng những trang web giả mạo, chúng ta sẽ áp dụng các mô hình học máy (Machine Learning) như là: Support Vector Machines (SVM), Decision Trees, Random Forests và Neural Networks để phát hiện đâu là website chính thống và đâu là giả mạo.

Để thực hiện việc đào tạo mô hình , chúng ta sẽ phải thu thập và trích xuất các đặc trưng từ địa chỉ URL của website. Khi mô hình đào tạo hoàn thành, sẽ kết hợp với việc xây dựng một website để kiểm tra các đường dẫn URL.

**MỤC LỤC**

[LỜI CẢM ƠN i](#_Toc161518056)

[TÓM TẮT iv](#_Toc161518057)

[DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT 2](#_Toc161518058)

[DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ 3](#_Toc161518059)

[CHƯƠNG 1 – MỞ ĐẦU 4](#_Toc161518060)

[1.1 Lý do chọn đề tài 4](#_Toc161518061)

[1.2 Tổng quan về vấn đề nghiên cứu 5](#_Toc161518062)

[1.3 Mục đích và đối tượng nghiên cứu 5](#_Toc161518063)

[CHƯƠNG 2 – CƠ SỞ LÝ THUYẾT 5](#_Toc161518064)

[2.1 Tổng quan về tấn công lừa đảo(Phishing) 6](#_Toc161518065)

[2.2 Phương pháp chống và phát hiện Phishing trên môi trường mạng 8](#_Toc161518066)

[2.3 Tổng quan về Machine Learning 9](#_Toc161518067)

[2.4 Các thuật toán trong Machine Learning được áp dụng 10](#_Toc161518068)

[2.4.1 Decision Tree 10](#_Toc161518069)

[2.4.2 Random Forest 11](#_Toc161518070)

[2.4.3 Support Vector Machine 11](#_Toc161518071)

[2.4.4 Neural Network 12](#_Toc161518072)

[CHƯƠNG 3 – XẤY DỰNG MÔ HÌNH PHÁT HIỆN TẤN CÔNG PHISHING 14](#_Toc161518073)

[3.1 Thiết kế mô hình 14](#_Toc161518074)

[3.2 Bộ dữ liệu 14](#_Toc161518075)

[3.3 Phương pháp và đánh giá 19](#_Toc161518076)

[3.4 Hiện thực mô hình 19](#_Toc161518077)

[3.4.1 Trích xuất đặc trưng của các URL: 20](#_Toc161518078)

[3.4.2 Xây dựng mô hình Machine Learning: 22](#_Toc161518079)

[3.4.3 Xây dựng website kiểm tra URL: 24](#_Toc161518080)

[CHƯƠNG 4 – TỔNG KẾT 28](#_Toc161518081)

[4.1 Kết quả đạt được 28](#_Toc161518082)

[4.2 Hạn chế 28](#_Toc161518083)

[TÀI LIỆU THAM KHẢO 29](#_Toc161518084)

# DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

**CÁC KÝ HIỆU**

*Chữ cái Sigma dùng để tính tổng một chuỗi các số hoặc biểu thức*

**CÁC CHỮ VIẾT TẮT**

CEO “Chief Excutive Officer” – Giám đốc điều hành

CFO “Chief Financial Officer” – Giám đốc tài chính

URL “Uniform Resource Locator” – địa chỉ trên internet

HTTP “Hypertext Transfer Protocol” – giao thức truyền thông

HTTPS “Hypertext Transfer Protocol Secure” –biến thể HTTP và được mã hóa

2FA “Two-Factor Authentication” – xác thực hai yếu tố

RBF “Radial Basis Function” – hàm cơ sở xuyên tâm

# DANH MỤC CÁC BẢNG BIỂU, HÌNH VẼ, ĐỒ THỊ

**DANH MỤC HÌNH**

[Hình 1.1 Tổng quan bài toán phát hiện tấn công Phishing 6](#_1t3h5sf)

[Hình 2.1 Sơ đồ thuật toán Decision Tree[5] 11](#_1ksv4uv)

[Hình 2.2 So sánh giữa Decision Tree và Random Forest 12](#_2jxsxqh)

[Hình 2.3 Support Vector Machine [5] 13](#_3j2qqm3)

[Hình 2.4 Sơ đồ lớp mạng Neural [6] 14](#_2xcytpi)

[Hình 3.1 Tỉ lệ giá trị Phishing và Legitimate trong tập dữ liệu 16](#_3as4poj)

[Hình 3.2 Phân bố dữ liệu của một số đặc trưng 18](#_49x2ik5)

[Hình 3.3 Ma trận tương quan giữa các đặc trưng 19](#_2p2csry)

[Hình 3.4 Cách vận hành của website 28](#_1hmsyys)

**DANH MỤC BẢNG**

[Bảng 1 Bốn hàm nhân phổ biến tại SVM 13](#_1y810tw)

[Bảng 2 Hàm kích hoạt trong Neural Network 15](#_1ci93xb)

[Bảng 3 Các đặc trưng trích xuất từ URL 18](#_1pxezwc)

# CHƯƠNG 1 – MỞ ĐẦU

## Lý do chọn đề tài

Đối với sự phát triển công nghệ mạnh mẽ như ngày nay, thì thông tin cá nhân là nguồn dữ liệu bí mật và rất quan trọng với tất cả các cá nhân và doanh nghiệp. Hiện nay việc trao đổi thông tin và các giao dịch trên Internet diễn ra hằng ngày , hằng giờ với các cách thức vô cùng đơn giản và dễ sử dụng. Cũng chính vì thế mà tội phạm Internet ra đời và phát triển ngày càng nhiều. Chúng chuyên đi đánh cắp thông tin mật của người dùng là khách hàng của các ngân hàng nhằm mục đích chiếm đoạt tài sản dựa trên các tài khoản banking online. Ngoài ra các tên tội phạm này còn có mục đích phá hoại hệ thống máy tính của người dùng, doanh nghiệp bằng các mã độc. Đa phần chúng sẽ sử dụng cách tấn công Phishing để giả mạo trang web chính thống đánh lừa thị giác người dùng.

Tấn công lừa đảo (Phishing) [1] là hình thức giả mạo các thông tin của các website chính thống có uy tín nhằm tạo cái bẫy để người dùng tự cung cấp thông tin cá nhân. Các website thường bị mạo danh phổ biến là ngân hàng, thương mại điện tử, ví điện tử. Thông thường chúng sẽ giả mạo cả tên hiển thị của ngân hàng khi gửi tin nhắn SMS hoặc Email cho người dùng, kèm theo đó là các đường link để truy cập.

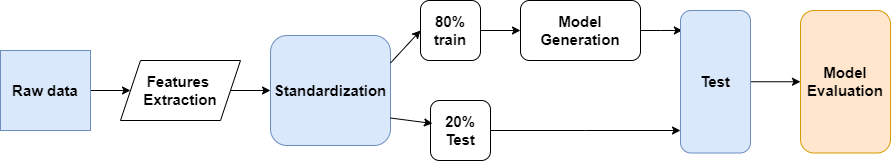
Vì lý do đó, tụi em đã quyết định chọn đề tài này nhằm hiểu hơn về cách thức hoạt động cũng như là tìm cách để phòng tránh những vấn nạn về phishing, đưa ra những lời khuyên cho người dùng và cộng đồng người dùng Internet. Song, đóng góp cho cộng động một công cụ để cảnh báo những trang web lừa đảo hiện đang tràn lan trên mạng.

Hiện nay, vấn nạn tấn công lừa đảo các nạn nhân không chuyên về bảo mật và công nghệ đang ngày càng nhiều và chúng sử dụng những biện pháp rất tinh vi, ví dụ điển hình là các đường link lừa đảo nhằm đánh cắp các thông tin của người dùng. Nhưng do không có nhiều cách thức để phát hiện ra tấn công lừa đảo (Phishing) hiện nay do những tội phạm mạng ngày càng phát triển với những kỹ thuật tinh vi hơn nên vấn nạn này vẫn còn đang gây nhức nhối trong cộng đồng mạng nói chung.

Vì vậy, tụi em quyết định chọn đề tài này nhằm nghiên cứu từ những địa chỉ giả mạo được lưu trong blacklist từ trước. Từ đó, phân tích ra những đặc trưng mà địa chỉ đó đang có sau đó sử dụng các kỹ thuật học máy để dự đoán đường link đó có an toàn hay không. Song, để dễ dàng áp dụng vào thực tế cho người dùng chúng em đã xây dụng một trang web áp dụng kỹ thuật học máy đó để đưa ra dự đoán cho người dùng. Những kĩ thuật học máy (Machine Learning) mà tụi em áp dụng bao gồm: Decision Tree, Random Forest, SVM, Neural Networks.

## Tổng quan về vấn đề nghiên cứu

Nghiên cứu cách thức tấn công giả mạo(Phishing): Dữ liệu được thu thập từ bộ dữ liệu, sau đó sẽ dựa vào địa chỉ url để phân tích các đặc trưng (features) , tiếp đến là chuẩn hóa dữ liệu. Và bắt đầu đưa vào mô hình đào tạo.



*Hình 1.1 Tổng quan bài toán phát hiện tấn công Phishing*

## Mục đích và đối tượng nghiên cứu

Mục đích : Xây dựng mô hình học máy (Machine Learning) áp dụng các thuật toán : Decision tree, Random forest, SVM,Neural networks để phát hiện tấn công lừa đảo(Phishing) . Nghiên cứu về cơ sở lý thuyết và các kỹ thuật áp dụng để ngăn chặn Phishing. Thu thập dữ liệu có liên quan để đưa vào mô hình dự đoán thông qua việc huấn luyện.

Đối tượng: + Tìm hiểu tổng quan về Machine Learning

+ Tìm hiểu về các mô hình được áp dụng

+ Dữ liệu trích xuất đặc trưng của địa chỉ URL từ các bộ dữ liệu

có sẵn ở các website như Mendeley ,Kaggle…

# CHƯƠNG 2 – CƠ SỞ LÝ THUYẾT

*Tổng quan*: Chương này thể hiện các cơ sở lý thuyết về Machine Learning và các thuật toán được áp dụng trong việc phát hiện tấn công phishing. Nhằm giải thích và phân tích quá trình thực hiện huấn luyện cho tập dữ liệu được đề cập ở Chương 3.

## 2.1 Tổng quan về tấn công lừa đảo(Phishing)

Tấn công lừa đảo(Phishing) là một dạng tấn công mạng phi kĩ thuật nhưng sức mạnh của nó đang nguy hại đến người dùng. Hình thức để triển khai cuộc tấn công này là các tên tội phạm sẽ giả dạng các website phổ biến mà mọi người hay sử dụng. Đặc biệt là website của các sàn giao dịch thương mại điện tử, online banking. Mục đích của chúng là chiếm đoạt thông tin cá nhân của người dùng khi họ nhập vào các website giả mạo này. Hành vi của chúng vô cùng tinh vi và khó phát hiện, chúng hướng đến các đối tượng thiếu kiến thức về bảo mật thông tin trên môi trường Internet. Những loại tài khoản như : Facebook, Gmail, tài khoản ngân hàng… sẽ là miếng mồi ngon mà chúng sẽ để ý đến.

Các tội phạm an ninh mạng được tiếp cận rất nhiều về các loại tấn công phishing. Một số cách tấn công phổ biến mà họ thường sử dụng [2] :

* **Email Phishing:** Phương thức dùng email để tấn công là cái phổ biến nhất. Chúng giả dạng các email của các doanh nghiệp lớn có uy tín để trao đổi với người dùng. Những email này thường bị sai về lỗi chính tả, chúng cố tình để sự nhầm lẫn này diễn ra vì đó là đón đánh vào thị giác người dùng. Chẳng hạn những chữ cái đứng gần nhau thì sẽ gây nhầm lẫn như : chữ “r” đứng cạnh “n” sẽ cho ra như này “rn” , nó tương tự như chữ “m”. Ngoài ra phần đứng sau ký tự “@” họ sẽ thêm vào đó là tên của các doanh nghiệp lớn uy tín.
* **Spear Phishing:** Đây là cách tấn công trực tiếp hướng đến bộ phận IT của doanh nghiệp, vì vai trò của bộ phận này có thể truy cập và quản lý các dự án quan trọng của doanh nghiệp.
* **Whaling:** Cuộc tấn công này hướng đến đối tượng có vai trò to lớn và quan trọng bậc nhất của doanh nghiệp như là CEO, CFO. Chúng thường mạo danh là đối tác có quyền hạn ngang bằng để trao đổi nhằm buộc bên còn lại cung cấp các thông tin nhạy cảm và có giá trị cao.
* **Vishing and Smishing:** Loại tấn công Smishing, chúng sẽ nhắm đến số điện thoại của người dùng, thay vì gửi email thì thay thế bằng tin nhắn SMS. Nội dung tin nhắn thường đi đôi với đường link vào trang web của chúng, hầu như trên link điều để tên tương tự của ngân hàng nào đó, khi click vào link đó và làm theo yêu cầu của họ thì thông tin sẽ bị chiếm đoạt. Vishing còn được gọi là Voice Vishing, chúng sẽ cử người gọi điện trao đổi trực tiếp với người dùng với danh xưng là một nhân viên ngân hàng hay là nhân viên báo trúng thưởng quà tặng để yêu cầu chúng ta cung cấp thông tin cá nhân như: tên, năm sinh, địa chỉ… để chúng đối chiếu với tài khoản ngân hàng của nạn nhân.
* **Pharming:** Loại tấn công này sử dụng kỹ thuật rất nhiều nên khó mà bị phát giác. Cách thức của chúng là sẽ chuyển hướng truy cập của một URL bằng cách đổi sang địa chỉ IP khác. Khi người dùng truy cập theo địa URL đó thì lập tức sẽ chuyển sang trang mà chúng đã thay đổi.
* **HTTPS Phishing:** Hầu hết người dùng sẽ nghỉ rằng protocol HTTPS đã được bảo mật an toàn hơn HTTP. Và cũng cho rằng đó là website chính thống nên nhấp vào. Nhưng hiện nay các tên tội phạm mạng đã tận dụng điều này nên website lửa đảo cũng sẽ được sử dụng protocol HTTPS để tăng mức độ tin tưởng.
* **Pop-up Phishing:** Hiện nay các browser sẽ có chức năng chặn cửa sổ hiển thị. Nhưng cửa sổ vẫn tự bật lên và đồng thời mang theo các nguy hại. Vì các tên tội phạm thường đặt mã độc trong đó và chúng sẽ kích hoạt khi ta lỡ nhấp vào nó.
* **Search Engine Phishing:** Cách tấn công này sẽ đánh vào sự tò mò và lòng tham của người dùng. Chúng sẽ làm ra các website ưu đãi, khuyến mãi lớn, nhận quà tặng miễn phí. Thường thì các web này có tên miền khá là dễ kiếm trên thanh tìm kiếm. Điều kiện của chúng là buộc người dùng nhập thông tin cá nhân để chúng ghi nhận và gửi quà tặng về tới nhà.

## 2.2 Phương pháp chống và phát hiện Phishing trên môi trường mạng

Hiện nay các cuộc tấn công Phishing rất tinh vi và khó phát hiện đề phòng tránh. Vì thế, bước cơ bản để chống Phishing là phải nâng cao nhận thức cho người dùng. Đối với doanh nghiệp thì cần khóa huấn luyện về cách nhận biết Phishing cho nhân viên. Dựa vào các dữ liệu đó , thì National Cyber Security Centre đã đưa ra một số bước để phòng tránh như sau [3] :

* **Phân quyền hạn cần thiết cho tài khoản nhân viên:** Để giảm thiểu các cuộc truy cập trái phép gây tổn thất kinh tế do tội phạm mạng, thì chỉ cấp quyền hạn đúng với vai trò của nhân viên. Còn đối với quản trị viên phải chắc rằng không dùng tài khoản này để tùy ý duyệt web và email. Vì chúng luôn thích tấn công vào vị trí này ở doanh nghiệp, quyền hạn của quản trị viên có thể thay đổi thông tin các tài khoản khác. Nâng cao bảo mật hơn thì nên thiết lập xác minh hai bước(2FA) ,mã số xác minh sẽ được gửi về email.
* **Tập huấn nhân viên:** Hướng dẫn cách nhận biết Phishing và cảnh báo về các kĩ thuật tấn công Phishing cơ bản được áp dụng. Nhắc nhở nhân viên đọc kĩ yêu cầu từ email được gửi đến, vì bình thường nếu là email giao dịch thì không cần phải cung cấp thông tin chi tiết hay mật khẩu.
* **Kiểm tra dấu hiệu của Phishing:** Cảnh giác với các email được gửi từ các khu vực ngoài nước. Thường thì những email này không được hoàn chỉnh về mặt chính tả hoặc ngữ pháp. Ngoài ra các tên tội phạm mạng chuyên nghiệp hơn thì những điều này sẽ không xảy ra, vì thế khi chưa có lệnh từ cấp trên thì không được nhấp vào bất kì link nào trong email đó. Phải đảm bảo tên và địa chỉ email phải trùng khớp với email mà cấp trên thường sử dụng.
* **Báo cáo lại các cuộc tấn công Phishing:** Nếu nhân viên bị tấn công thì trước hết thì nên nhắc nhở họ thay đổi mật khẩu. Sau đó bản thân họ nên báo cáo ngay đến quản trị viên hoặc cấp trên người mà có quyền quản lý tài khoản để ngăn chặn truy cập từ thiết bị khác.
* **Kiểm tra dấu vết thông tin:** Để cuộc tấn công của họ trở nên thuyết phục thì các tội phạm mạng sẽ đi thu thập dữ liệu từ tài khoản mạng xã hội liên quan đến nhân viên và doanh nghiệp. Vì thế phải hạn chế chia sẻ thông tin quan trọng có ảnh hưởng đến doanh nghiệp hoặc bản thân trên mạng xã hội. Các tổ chức, doanh nghiệp nên truyền đạt nhận biết được thông tin nào không được phép chia sẻ và phải bí mật giữ kín để tránh sự nhầm lẫn của nhân viên.

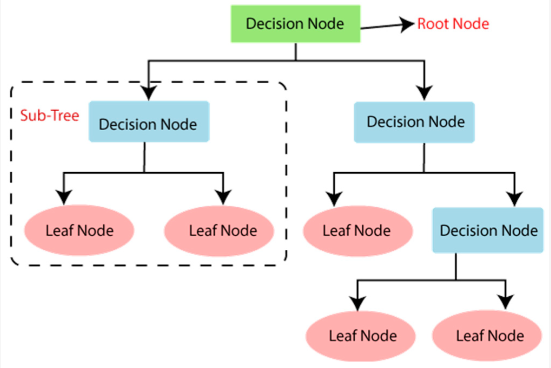
## 2.3 Tổng quan về Machine Learning

Machine learning [4] là một kỹ thuật thuộc mảng trí tuệ nhân tạo, các mô hình của học máy được huấn luyện nhờ vào các dữ liệu có sẵn. Đây được coi là phương thức hiệu quả trong việc phân tích dữ liệu. Nhờ đó mà học máy có ưu điểm chính là về việc phát hiện tấn công Phishing một cách linh hoạt dựa vào việc đào tạo mô hình. Vì muốn ngăn chặn tấn công lừa đảo thì vấn đề là phải phân loại Phishing hay Legitimate. Có thể nói học máy là một công cụ mạnh mẽ và hữu ích cho việc phân loại này. Ngoài ra với sự thích ứng nhanh với sự thay đổi để xác định ra các đặc trưng (feature) của các web lừa đảo. Và các phát hiện đó sẽ được đưa vào mô hình huấn luyện dựa trên việc học của mô hình. Một số thuật toán của học máy được áp dụng trong nghiên cứu này.

## 2.4 Các thuật toán trong Machine Learning được áp dụng

### 2.4.1 Decision Tree

Decision tree [5] thường được sử dụng để phân tích các mối quan hệ trong các tập dữ liệu lớn. Cấu trúc của decision tree là dạng sơ đồ cây vì chúng học cách phân vùng đệ quy. Tính năng này có thể giúp xử lý các tập dữ liệu khác nhau bao gồm dữ liệu số hay phân loại đều được xử lý với độ phân giải cao. Mục tiêu của decision tree là quan sát một quá trình. Hình 2, minh họa cho thuật toán này. Tỷ lệ thành công xác định từ vài khía cạnh. như là kích thước của bộ dữ liệu. Tuy nhiên, nhược điểm là lạm dụng dữ liệu và khó khăn trong việc cập nhật mẫu mới.

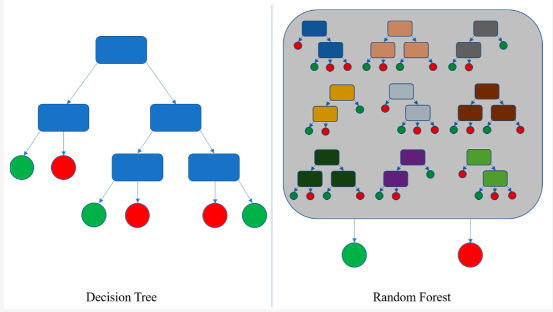


*Hình 2.2 Sơ đồ thuật toán Decision Tree[5]*

Nút gốc là nơi khởi đầu của decision tree(nút cha). Kế tiếp sẽ tách thành các nhóm đồng nhất(nút con). Bước cuối cùng là các nút lá, là nơi xuất ra cuối cùng không thể chia được nữa. Tách là hành động dựa trên điều kiện được chỉ định.

### 2.4.2 Random Forest

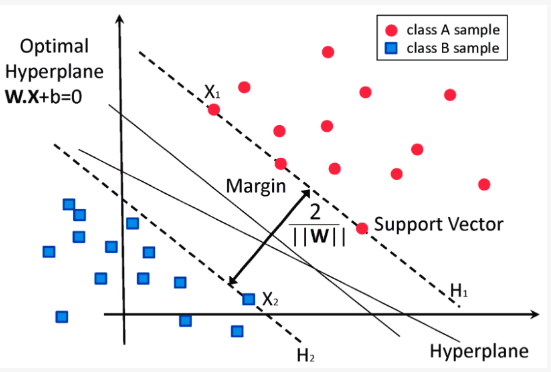
Random Forest [5] có thể phân loại và hồi quy các thuật toán có giám sát. Tốc độ xử lý và độ chính xác là điểm mấu chốt giúp làm nổi bật Random Forest. Nó tập hợp các decision tree để chọn lọc đầu ra tốt nhất dựa kết quả dự đoán của decision tree. Tập dữ liệu trong random forest được chia thành tập huấn luyện và tập kiểm tra. Tiếp theo, từ tập huấn luyện sẽ chọn ngẫu nhiên nhiều mẫu. Sau đó nó sẽ dựa vào xử lý của decision tree để đưa ra dự đoán. Nếu trong một random forest có chứa số lượng decision tree lớn thì hiệu suất và việc dự đoán sẽ đạt được sự ổn định cao hơn đồng thời gian xử lý tăng. Với việc cải thiện hiệu suất thì cần chú ý đến số lượng đặc trưng tối đa. Hình 3 là mô phỏng về trách nhiệm của mỗi cây trong random forest.



*Hình 2.3 So sánh giữa Decision Tree và Random Forest*

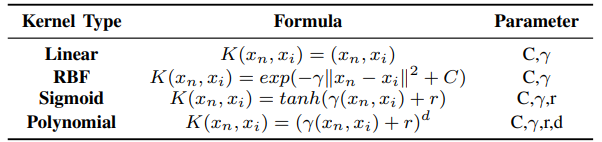
### 2.4.3 Support Vector Machine

Support Vector Machine (SVM) [5] là một thuật toán giám sát, có thể áp dụng cho việc phân loại hoặc đệ quy. Ý tưởng của SVM là dùng khoảng cách tối đa giữa các lớp để chọn điểm gần nhất giữa hai lớp. SVM dự đoán nhãn bằng cách tạo ra ranh giới quyết định. Các điểm dữ liệu và vectơ hỗ trợ được xử lý bởi siêu phẳng. Nhờ vào khoảng cách giữa các điểm để phân loại từng lớp một cách độc lập. Tuy nhiên điểm yếu của SVM là việc khi tiếp xúc với dữ liễu nhiễu sẽ dễ bị overfitting. Hình 4 mô tả về SVM dự đoán nhãn.



*Hình 2.4 Support Vector Machine [5]*

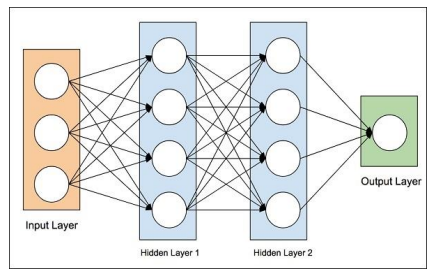
Bốn hàm nhân phổ biến tại [4] SVM là linear(tuyến tính), RBF (hàm cơ sở xuyên tâm), sigmoid và đa thức(polynomial), được nhắc đến trong Bảng 1. Mỗi hàm nhân có các tham số cụ thể phải được tối ưu hóa để thu được kết quả tốt nhất.



*Bảng 1 Bốn hàm nhân phổ biến tại SVM*

### 2.4.4 Neural Network

Neural Network [6] được hình thành từ các nút liên kết thành một mạng lưới thần kinh tương tự như các tế bào thần kinh trong não người. Cách di chuyển của các nút là chuyển tiếp, hướng di chuyển chỉ có một. Các lớp có trong một mạng neural gồm lớp đầu vào, lớp đầu ra và lớp trung gian. Nếu một mạng neural chứa nhiều lớp thì đó còn gọi là Deep Learning. Với số lượng lớp trong một mạng càng lớn thì việc xử lý các thông tin có phần phức tạp sẽ diễn ra đơn giản hơn. (Hình 5)



*Hình 2.5 Sơ đồ lớp mạng Neural [6]*

Trong mạng neural, các nút đều được gán một trọng số. Giá trị của thông tin được gán thường được thể hiện qua trọng số. Khi truyền thông tin từ nút này sang một nút khác, thì việc đầu tiên nó sẽ đi tính tổng trọng lượng hoặc giá trị của thông tin. Thông tin sẽ được chuyển qua lớp tiếp theo khi mà số lượng vượt quá mức tối đa được phép. Trọng số và ngưỡng của các neural được đặt ngẫu nhiên khi bắt đầu huấn luyện. Chúng sẽ được điều chỉnh bằng phương pháp giảm độ dốc để tối ưu và đem lại kết quả chính xác. Trong mạng neural có hàm kích hoạt phổ biến [7]:

|  |  |
| --- | --- |
| **Hàm kích hoạt** | **Công thức** |
| Linear |  |
| Binary step |  |
| Sigmoid |  |
| Tanh |  |
| ReLU |  |
| Softmax | ; j=1,…,k |

*Bảng 2 Hàm kích hoạt trong Neural Network*

# CHƯƠNG 3 – XẤY DỰNG MÔ HÌNH PHÁT HIỆN TẤN CÔNG PHISHING

## 3.1 Thiết kế mô hình

Để phát hiện tấn công Phishing thông qua địa chỉ URL , thì ở nghiên cứu này ta sẽ dùng features extraction để đưa ra các đặc trưng của các địa chỉ tên miền của website. Dữ liệu đầu vào cho mô hình train là 49 đặc trưng được trích xuất. Trong đó các dữ liệu sẽ có kiểu dữ liệu như sau:

+ Đối với các đặc trưng là đếm các số lượng thì sẽ có đơn vị là number

+ Đặc trưng thuộc dạng trả lời câu hỏi có hoặc không sẽ là 1,0 (1 là có, 0 là không có)

+ Ngoài ra lớp dùng để phân loại phishing hay legitimate sẽ có kiểu dữ liệu string

Với ý tưởng là huấn luyện mô hình để phân loại, khi đó dữ liệu đầu vào sẽ là địa chỉ URL, và kết quả trả ra là đó có là một cuộc tấn công phishing hay không.

## 3.2 Bộ dữ liệu

Để có tập dữ liệu để huấn luyện cho việc chống tấn công Phishing, bộ dữ liệu Web page phishing detection trên Mendeley Data [8] sẽ được sử dụng trong nghiên cứu này.

Thông tin dữ liệu được cung cấp bao gồm 11430 địa chỉ URL và 89 đặc trưng được trích xuất. Trong đó có 2 đặc trưng là địa chỉ url và nhãn để phân loại. Bộ dữ liệu này đã được thiết kế phù hợp cho việc sử dụng học máy để huấn luyện. Được phân chia các đặc trưng theo các lớp khác nhau:

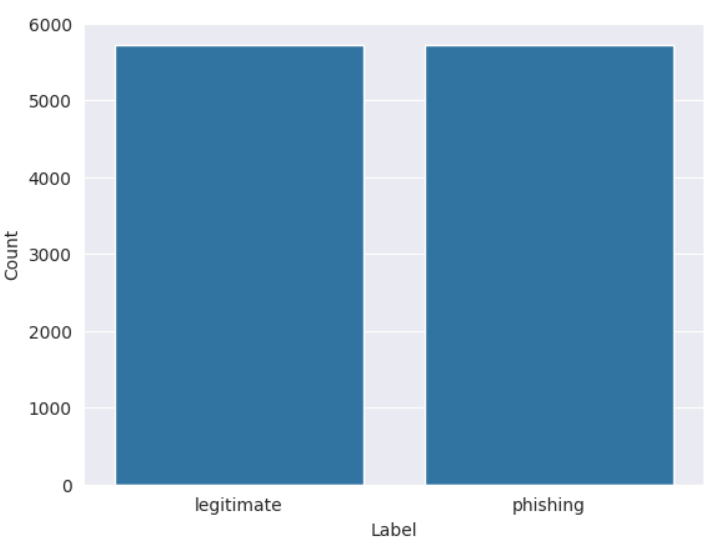
+ Các đặc trưng trích xuất dựa trên cấu trúc của địa URL thì có số lượng là 56.

+ 24 đặc trưng được trích xuất từ content của trang thông qua kiểm tra các thẻ

javascript,html,css.

+ 7 đặc trưng còn lại là dựa vào các dịch vụ uy tín ở bên ngoài.

Đây là bộ dữ liệu cân bằng gồm 50% lừa đảo (Phishing) và 50% chính thống (Legitimate). Trong đó “status” là nhãn chứa hai giá trị để phân loại “Phishing” và “Legitimate”. Tỉ lệ giữa hai giá trị này là bằng nhau (Hình 3.1).



*Hình 3.6 Tỉ lệ giá trị Phishing và Legitimate trong tập dữ liệu*

Để cho việc nghiên cứu đơn giản thì trong file đính kèm tác giả cũng đã cung cấp folder scripts chứa code Python dùng để trích xuất đặc trưng. Nhưng với bài nghiên cứu này thì chúng em chỉ sử dụng 50 đặc trưng trong tổng 89 đặc trưng. Bảng 2 sẽ thể hiện rõ các đặc trưng được sử dụng.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Tên cột** | **Kiểu dữ liệu** | **STT** | **Tên cột** | **Kiểu dữ liệu** |
| 1 | url | Object | 26 | have\_prefixOrSuffix | Int |
| 2 | status | Object | 27 | dns\_expiration | Int |
| 3 | is\_ip | Int | 28 | Count\_redirect | Int |
| 4 | length\_url | Int | 29 | page\_rank | Int |
| 5 | length\_hostname | Int | 30 | domain\_age | Int |
| 6 | Tiny\_url | Int | 31 | domainRegLen | Int |
| 7 | CountAtSign | Int | 32 | RatioLinksTag | Float |
| 8 | CountQuestionMark | Int | 33 | RatioAnchorURL | Float |
| 9 | CountHyphen | Int | 34 | poppup\_window | Int |
| 10 | CountDot | Int | 35 | abnormal\_subdomain | Int |
| 11 | CountComma | Int | 36 | iframe | Int |
| 12 | CountSemicolon | Int | 37 | Count\_www\_path | Int |
| 13 | CountDollar | Int | 38 | Count\_com\_path | Int |
| 14 | CountSlash | Int | 39 | length\_word\_raw | Int |
| 15 | have\_redirect | Int | 40 | avg\_row\_words | Float |
| 16 | CountEqual | Int | 41 | avg\_row\_words\_host | Float |
| 17 | CountPercent | Int | 42 | avg\_row\_words\_path | Float |
| 18 | CountUnderScore | Int | 43 | longest\_words\_raw | Int |
| 19 | CountDotHostName | Int | 44 | longest\_word\_host | Int |
| 20 | CountColon | Int | 45 | longest\_word\_path | Int |
| 21 | CountStar | Int | 46 | shortest\_words\_raw | Int |
| 22 | CountHttp | Int | 47 | shortest\_word\_host | Int |
| 23 | checkHttps | Int | 48 | shortest\_word\_path | Int |
| 24 | RatioDigitHost | Float | 49 | webTraffic | Int |
| 25 | RatioDigitsURL | Float | 50 | whoisRegistered | Int |

*Bảng 3 Các đặc trưng trích xuất từ URL*

A graph of different sizes and numbers

Description automatically generated with medium confidence

*Hình 3.7 Phân bố dữ liệu của một số đặc trưng*

Hình 3.2 là 9 trên 50 đặc trưng được trích xuất trong tập dữ liệu. Từ hình ảnh trên ta thấy dữ liệu ở mỗi đặc trưng được phân bổ rải rác từ cao đến thấp, không đều và có độ chênh lệch cao. Lấy ví dụ ở đặc trưng domain\_age, có số lượng phân bố nhiều ở giá trị trong khoảng 0-500 theo trục y và phân bố ít nhất là khi giá trị trên 1000. Để thấy rõ về sự phân bố dữ liệu thì ta sẽ có biểu đồ ma trận tương quan.(Hình 3.3).

A red and white grid with white text

Description automatically generated

*Hình 3.8 Ma trận tương quan giữa các đặc trưng*

Nhìn vào Hình 3.3 là một ma trận tương quan của các đặc trưng trong tập dữ liệu. Dãy màu bên phải là thể hiện phạm vi tương quan tiêu chuẩn từ 1.0 đến -0.6, trong đó phạm vi khi phạm vi càng cao thì đặc trưng đó càng không cần thiết cho mô hình. Trong hình cho thấy rằng các đặc trưng đa số đều rất cần thiết cho mô hình dự đoán.

## 3.3 Phương pháp và đánh giá

Độ chính xác (Accuracy) [9] được sử dụng để đo hiệu suất trong mô hình học máy. Khi đánh giá các mô hình phân loại thì độ chính xác sẽ được áp dụng. Trong bài nghiên cứu này cũng sử dụng độ chính xác để phân loại phishing hay legitimate, nó sẽ đưa ra tỉ lệ phần trăm về độ chính xác. Giá trị đó thu được dựa trên tổng số dự đoán đúng so với tổng số dự đoán thực hiện. Nó được thể hiện như sau:

Trong trường hợp đó là bài toán phân loại nhị phân, thì mục tiêu là tách các trường hợp tích cực(Positive) và tiêu cực (Negative), sau đó kiểm tra độ chính xác để xem số trường hợp đúng. Công thức như sau:

\* TP : True Positive là mô hình phân loại đối tượng ở lớp Positive vào đúng lớp Positive ( dự đoán đúng)

\* TN : True Negative là mô hình phân loại đối tượng ở lớp Negative vào cùng lớp Negative (dự đoán đúng)

\* FP : False Positive là mô hình phân loại đối tượng lớp ở Negative vào lớp Positive (dự đoán sai)

\* FN ; False Negative là mô hình phân loại đối tượng lớp ở Positive vào lớp Negative(dự đoán sai)

## 3.4 Hiện thực mô hình

Để hiện thực mô hình, với nghiên cứu này thì sẽ sử dụng bộ dữ liệu Web page phishing detection [8] trên Mendely và Google Colaboratory, Python và Machine Learning. Ngoài ra, khi đã có mô hình hoàn chỉnh thì sẽ đưa mô hình vào Back-end là Django và Front-end là React Js để xây dựng một website cơ bản nhận vào địa chỉ URL và trả kết quả có phải là phishing hay không ra cho người dùng.

### 3.4.1 Trích xuất đặc trưng của các URL:

* Với bộ dataset có url và status thì dựa vào url đó trích xuất 48 đặc trưng như sau:
* is\_ip: kiểm tra xem trang url đó có sử dụng ip hay không
* length\_url: trả về độ dài của url
* length\_hostname: trả về độ dài của host
* tiny\_url: kiểm tra xem có sử dụng các đường link rút gọn hay không
* CountAtSign: đếm trên url có bao nhiêu dấu ‘@’
* CountQuestionMark: đếm trên url có bao nhiêu dấu ‘?’
* CountHyphen: đếm trên url có bao nhiêu dấu ‘-’
* CountDot: đếm trên url có bao nhiêu dấu ‘.’
* CountComma: đếm trên url có bao nhiêu dấu ‘,’
* CountSemicolon: đếm trên url có bao nhiêu dấu ‘;’
* CountDollar: đếm trên url có bao nhiêu dấu ‘$’
* CountSlash: đếm trên url có bao nhiêu dấu ‘/’
* have\_redirect: có điều hướng hay không ‘//’
* CountEqual: đếm trên url có bao nhiêu dấu ‘=’
* CountPercent: đếm trên url có bao nhiêu dấu ‘%’
* CountUnderScore: đếm trên url có bao nhiêu dấu ‘\_’
* CountDotHostName: đếm xem có bao nhiêu dấu ‘.’ trong hostname
* CountColon: đếm trên url có bao nhiêu dấu ‘:’
* CountStar: đếm trên url có bao nhiêu dấu ‘\*’
* CountHttp: đếm trên url có sử dụng bao nhiêu http trừ giao thức http://
* checkHttps: kiểm tra giao thức của url có là https hay không
* RatioDigitHost: tỉ lệ số xuất hiện ở hostname
* RatioDigitsURL: tỉ lệ số xuất hiện trong cả url
* have\_prefixOrSuffix: url có tiền tố hay hậu tố không
* dns\_expiration: kiểm tra dns đã hết hạn hay chưa
* Count\_redirect: đếm số lần url điều hướng
* page\_rank: kiểm tra ranking của page
* domain\_age: số tuổi tồn tại của domain
* domainRegLen: độ tuổi mà domain đã đăng ký
* RatioLinksTag: tỉ lệ phần trăm các liên kết có trong <Script>,<Link>
* RatioAnchorURL: tỉ lệ xuất hiện url tại các thẻ <a>
* poppup\_window: Kiểm tra tính năng cửa sổ có được mở
* abnormal\_subdomain: Kiểm tra cú pháp url có đủ hay không (thiếu http hay https, hoặc các kí tự đặc biệt không hợp lệ)
* iframe: Kiểm tra trang web có sử dụng iframe và frameBorder để nhúng các dữ liệu hoặc chuyển hướng khác hay không
* Count\_www\_path: đếm số lần xuất hiện của chuỗi “www” trên phần path của url
* Count\_com\_path: đếm số lần xuất hiện của “com” trên phần path của url
* length\_word\_raw: số lượng từ xuất hiện trên url sau khi lọc bỏ “/”,”https://”,”http://”
* avg\_row\_words: tính độ dài trung bình của các từ trong danh sách sau khi lọc bỏ “/”,”https://”,”http://”
* avg\_row\_words\_host: tính độ dài trung bình của các từ thuộc hostname sau khi lọc bỏ “/”,”https://”,”http://”
* avg\_row\_words\_path: tính độ dài trung bình của các từ thuộc phần path của url sau khi lọc bỏ “/”,”https://”,”http://”
* longest\_words\_raw: độ dài của từ dài nhất trong url
* longest\_word\_host: độ dài của từ dài nhất trong hostname
* longest\_word\_path: độ dài của từ dài nhất trong phần path
* shortest\_words\_raw: độ dài của từ ngắn nhất trong url
* shortest\_word\_host: độ dài của từ ngắn nhất trong hostname
* shortest\_word\_path: độ dài của từ ngắn nhất trong phần path
* webTraffic: mực độ lưu lượng truy cập web
* whoisRegistered: tên miền có được công nhận trong whois hay không

### 3.4.2 Xây dựng mô hình Machine Learning:

* Đầu tiên dựa vào tập dữ liệu đầu vào thì khi trực quan hóa nó ra thì có thể thấy rằng một số cột đặc trưng có dữ liệu khá cao so với một số còn lại. Vì vậy, để tránh mô hình dự đoán sai khi có sự chênh lệch quá lớn giữa các đặc trưng thì trước khi tiến hành train mô hình chúng em sẽ thực hiện chuẩn hóa lại bộ dữ liệu của X (là tập gồm tất cả các đặc trưng trừ url và status) bằng cách sử dụng thư viện MinMaxScaler.
* Sau khi đã scale data lại thì có thể thấy rằng dữ liệu đã không còn chênh lệch nhau quá nhiều như ban đầu. Vì vậy tiến hành chia tập dữ liệu X và Y (là tập dữ liệu chỉ chứa status) ra thành 80% là tập dữ liệu để train và 20% là tập dữ liệu để test.
* Sử dụng các mô hình như sau:
* Sử dụng thư viện tensorflow để xây dựng và huấn luyện mô hình mạng neural, với 2 lớp ẩn lần lượt bao gồm 128 và 64 neural sử dụng hàm activation là relu, lớp đầu ra gồm 1 neural sử dụng hàm activation là sigmoid. Sau khi huấn luyện mô hình được chỉ số accuracy là 95% với chỉ số loss là ~0.187
* Tiếp theo sử dụng thư viện RandomForestClassifier để xây dựng mô hình thì được chỉ số accuracy là 96%
* Tiếp theo sử dụng thư viện DecisionTreeClassifier để xây dựng mô hình thì được chỉ số accuracy là 94%
* Tiếp theo sử dụng thư viện SVC để xây dựng mô hình thì được chỉ số accuracy là 93%

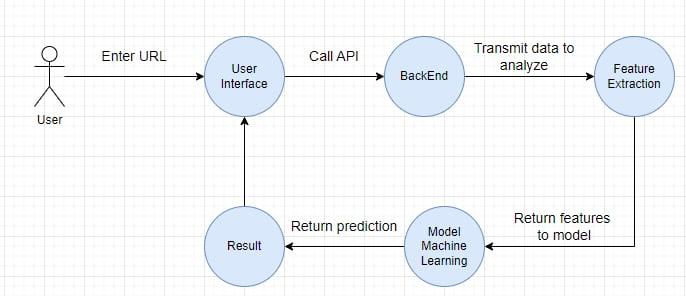
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **F1\_Score** | **Recall** | **Precision** |
| Random Forest | 0.964 | 0.964 | 0.971 | 0.958 |
| Neural Network | 0.952 | 0.954 | 0.948 | 0.960 |
| Decision Tree | 0.940 | 0.941 | 0.944 | 0.938 |
| Support Vector Machine | 0.938 | 0.939 | 0.950 | 0.927 |

* Sau khi so sánh các chỉ số thì có thể thấy rằng mô hình sử dụng thuật toán Random Forest cho ra kết quả chính xác cao nhất. Vì vậy, chúng em sẽ sử dụng mô hình này để dự đoán.
* Sau khi chọn được mô hình tốt nhất thì chúng em có thử thêm vào các url khác và kết quả dự đoán cho thấy rằng đúng 8/10 url được đưa vào.

|  |  |  |
| --- | --- | --- |
| **URL** | **Result** | **Expected** |
| https://www.google.com | 0 | 0 |
| http://205.174.165.80/CICDataset/ISCX-URL-2016/ | 1 | 1 |
| https://kotaku.com/the-witcher-3-wild-hunt-the-kotaku-review-1703766283 | 1 | 0 |
| http://docs.python.org:80/3/library/urllib.parse.html? | 1 | 0 |
| https://usps.address-shipwatchers.shop/ | 1 | 1 |
| https://grtakeprize.com/it/bp4f/index?c=2277&affId=BC6DC089&c1=48&c2=1opCaODZewbK&c3=1c1e607843434706847313349c7444a0&c4=&c7=&c5=&c8=&c6=&i1=&t1= | 1 | 1 |
| https://t.ly/EWZo9 | 1 | 1 |
| http://fis-lab.com/wp-includes/images/smilies/front/wwpp/ | 1 | 1 |
| https://www.kaggle.com/busrabetulcavusoglu/ | 0 | 0 |
| https://www.tdtu.edu.vn | 0 | 0 |

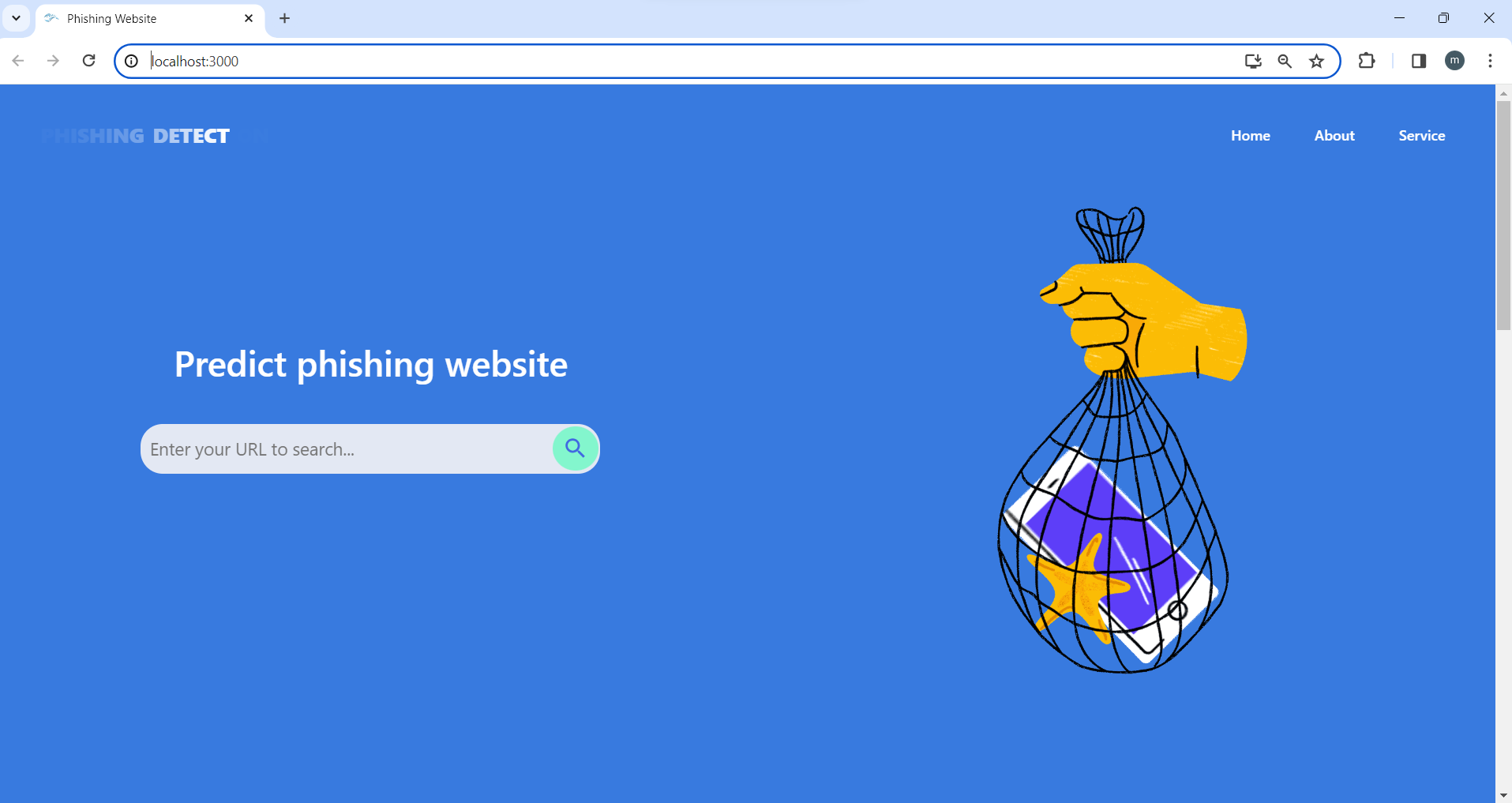
### 3.4.3 Xây dựng website kiểm tra URL:

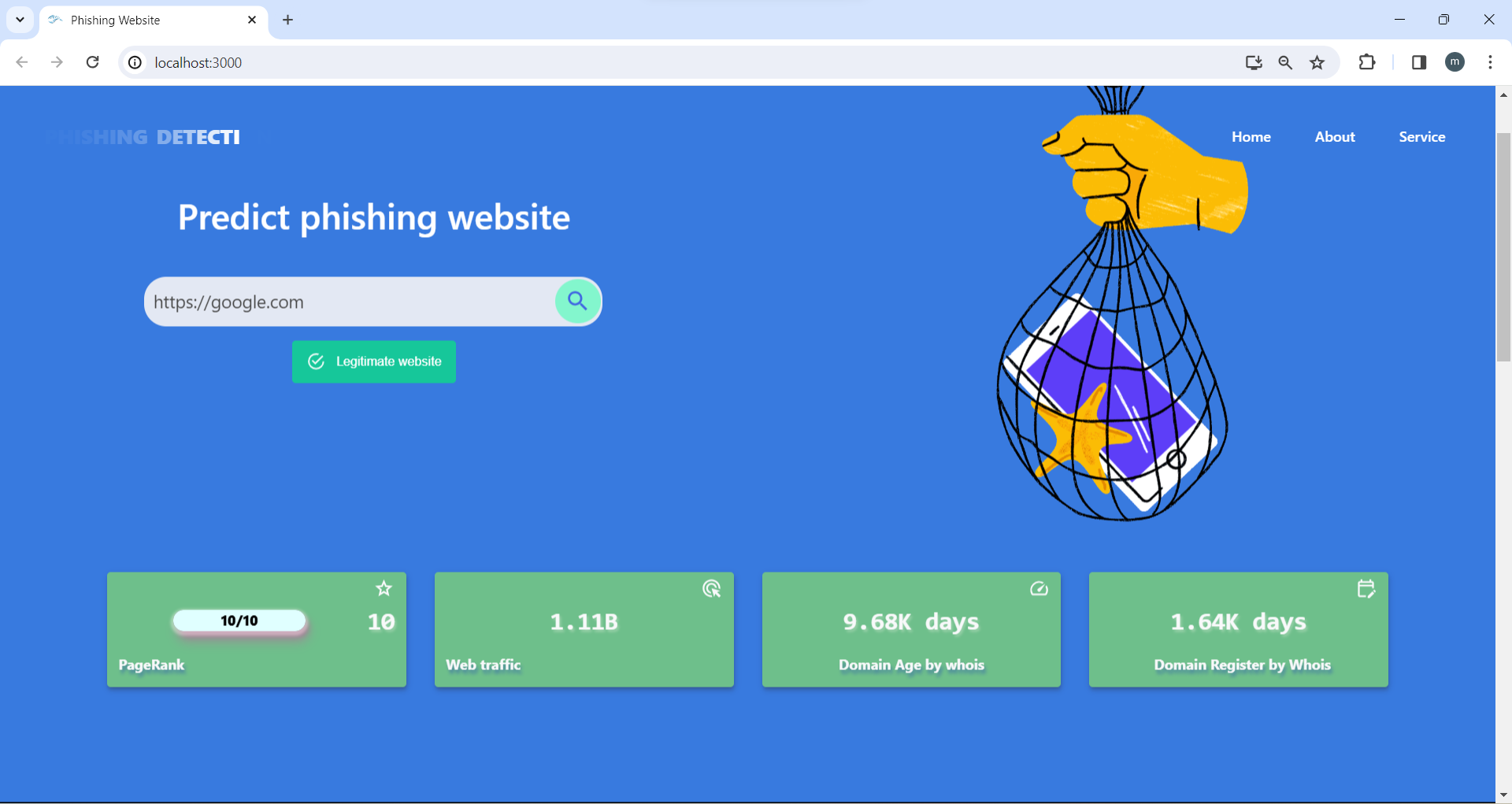
Sơ đồ hoạt động của trang web:



*Hình 3.9 Cách vận hành của website*

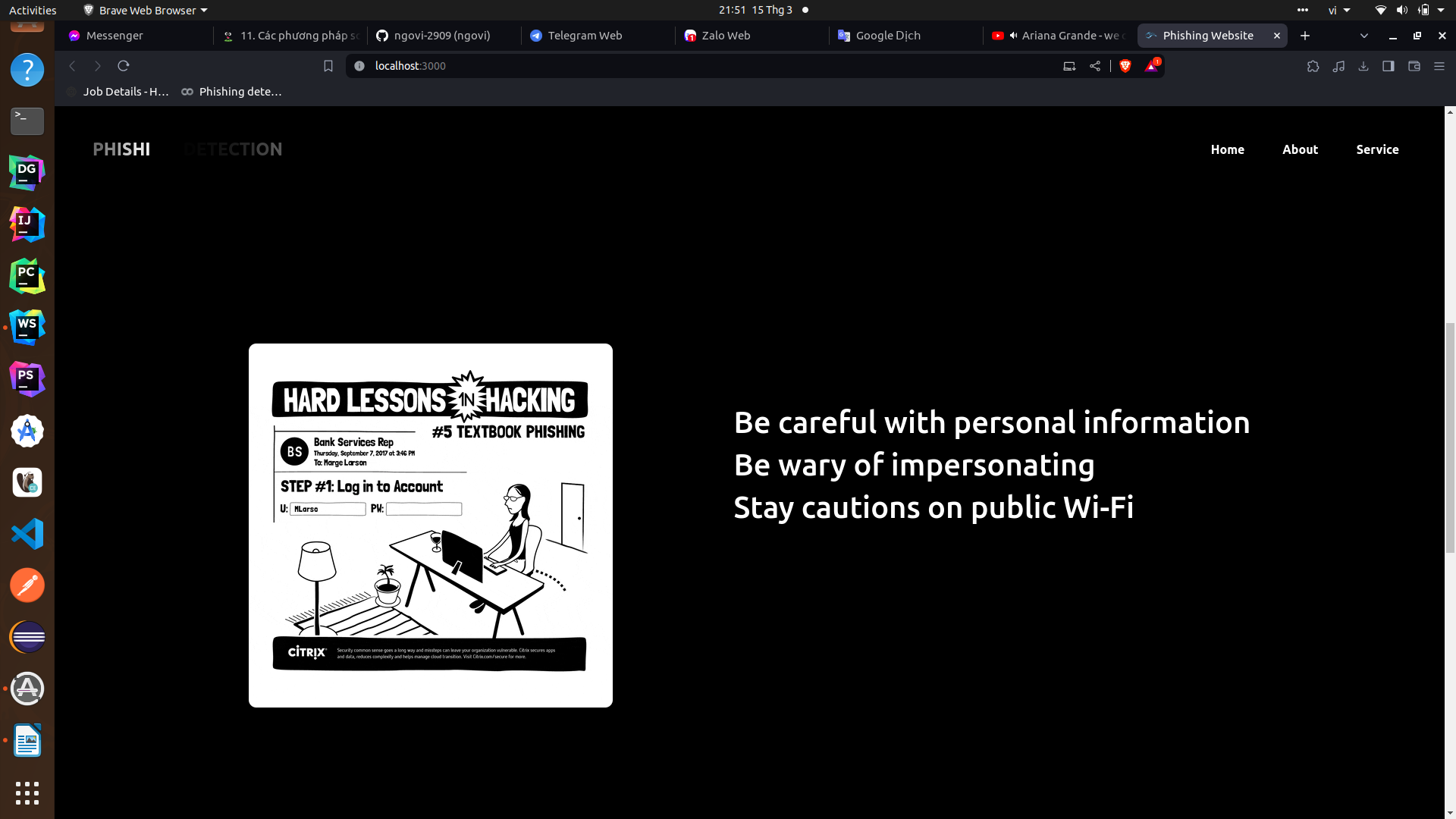
Giao diện của website:

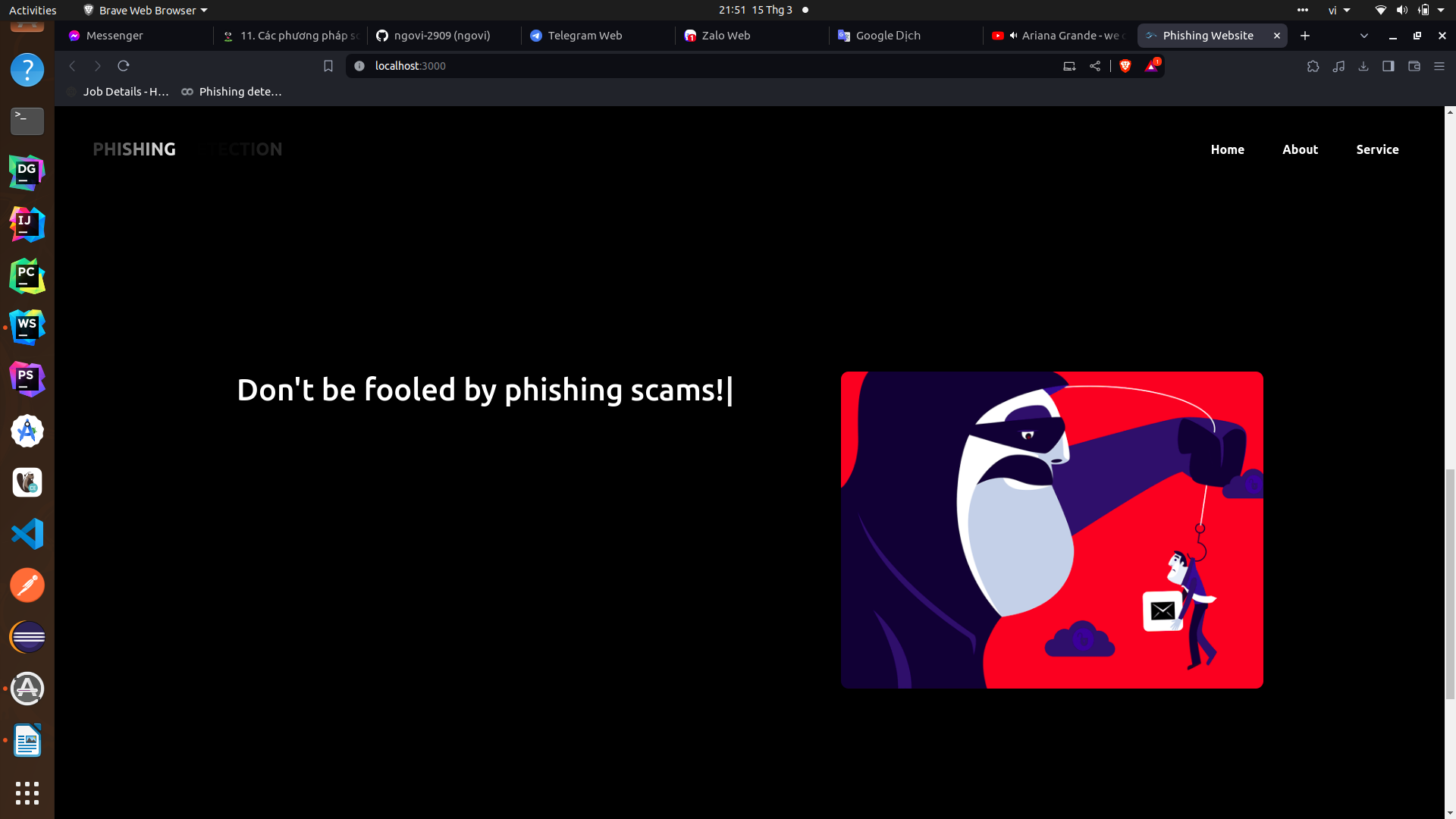




A screenshot of a computer

Description automatically generated





* Về phần giao diện của web tụi em sử dụng framework là ReactJS để xây dựng một landing page cơ bản cho người dùng nhập vào url mà họ muốn kiểm tra và tụi em sẽ show ra một số thông tin như pagerank của page, web traffic, domain age và domain register length được đăng ký trên whois. Bên cạnh đó tụi em cũng đưa ra một số thông điệp cho người dùng bên dưới để cảnh giác với các cuộc tấn công phishing.
* Về phần back-end thì tụi em sử dụng framework là Django Rest Framework để xây dựng API kiểm tra url. Khi người dùng nhập vào url bên phía front-end thì nó sẽ call đến api mà tụi em đã xây dựng. Khi đó bên phía back-end sẽ nhận được url và tiến hành kiểm tra bằng cách sử dụng mô hình học máy mà tụi em đã xây dựng trước đó để đưa ra dự đoán và trả về kết quả cho phía front-end dưới dạng JSON, bên front-end sẽ mapping kết quả đó vào để show ra cho người dùng.

# CHƯƠNG 4 – TỔNG KẾT

## 4.1 Kết quả đạt được

Tấn công lừa đảo (Phishing) là một trong những dạng tấn công đánh cắp thông tin mà các người dùng mạng và các tổ chức cung cấp dịch vụ phải đối mặt trong thời buổi công nghệ phát triển hiện nay. Với việc thực hiện nghiên cứu phát hiện tấn công phishing cũng là một đề tài vô cùng có ích cho việc bảo vệ an toàn thông tin. Nhờ sự phát triển không ngừng của các công nghệ mới cũng góp phần tăng cao hiệu suất phát hiện và ngăn chặn các cuộc tấn công này. Trong đó mô hình Machine Learning đã cho thấy điểm mạnh và sự tiện ích khi xử lý vấn đề này một cách khá hiệu quả. Trong bài nghiên cứu này, thì chúng em đã tìm hiểu về lý thuyết của các mô hình học máy và cách xây dựng mô hình huấn luyện, và tạo một trang web cơ bản có thể kiểm tra dựa trên các địa chỉ URL.

## 4.2 Hạn chế

Hiện tại, bài nghiên cứu này chỉ dừng lại ở việc phân tích 48 đặc trưng có thể xem là khá ít đối với việc phân tích cho một url phishing do đó nó chỉ có độ chính xác khoảng 96%.

Các url trong bộ dữ liệu này đã khá lâu vì vậy khi phân tích đặc trưng ra có thể sẽ có sự thay đổi hoặc sai sót so với các url phishing hiện tại khi các hacker đã cải tiến các kỹ thuật so với những năm về trước. Điều đó dẫn đến một số features của phishing là legitimate giống nhau dẫn đến sự sai sót trong dự đoán mô hình.

# TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | A. S. V.Bhavsar, "Study on Phishing Attacks," [Online]. Available: https://www.researchgate.net/profile/Shabnam-Sharma-2/publication/329716781\_Study\_on\_Phishing\_Attacks/links/5ef9867a92851c52d6069bf2/Study-on-Phishing-Attacks.pdf. |
| [2] | BlueVoyant, "8 Phishing Types and How to Prevent Them," [Online]. Available: https://www.bluevoyant.com/knowledge-center/8-phishing-types-and-how-to-prevent-them. |
| [3] | N. C. S. Centre, "Step 5 - Avoiding phishing attacks," 10 October 2017. [Online]. Available: https://www.ncsc.gov.uk/collection/small-business-guide/avoiding-phishing-attacks. |
| [4] | M. M. V.Shahrivari, "Phishing Detection Using Machine Learning," 20 Sep 2020. [Online]. Available: https://arxiv.org/pdf/2009.11116.pdf. |
| [5] | M. A. Shouq Alnemari, "Detecting Phishing Domains Using Machine Learning," 7 April 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/8/4649#B12-applsci-13-04649. |
| [6] | Wikipedia, "Neural Network (machine learning)," [Online]. Available: https://en.wikipedia.org/wiki/Neural\_network\_(machine\_learning). |
| [7] | T. T. Huynh, "ỨNG DỤNG REPRESENTATION LEARNING PHÁT HIỆN TẤN CÔNG PHISHING," 2023. [Online]. Available: https://ptithcm.edu.vn/wp-content/uploads/2023/03/2020\_HTTT\_TranHuynhTien\_LV.pdf. |
| [8] | S. Y. Abdelhakim Hannousse, "Web page phishing detection," 26 June 2021. [Online]. Available: https://data.mendeley.com/datasets/c2gw7fy2j4/3. |
| [9] | educative, "What is accuracy in machine learning?," [Online]. Available: https://www.educative.io/answers/what-is-accuracy-in-machine-learning. |