# Phương pháp mới phát hiện URL lừa đảo sử dụng thuật toán học máy kết hợp

**Nguyễn Mạnh Thắng, Lê Quang Anh, Hứa Song Toàn, Nguyễn Quốc Trung**

***Tóm tắt*— Tấn công lừa đảo là một loại tấn công mạng nhắm vào sự tin tưởng của người dùng bằng cách che giấu ý đồ ác ý của cuộc tấn công dưới dạng thông tin của các nguồn có uy tín. Mục tiêu là lấy cắp dữ liệu nhạy cảm của nạn nhân (thông tin ngân hàng, nhận dạng xã hội, thông tin đăng nhập,...) với nhiều mục đích khác nhau (bán để kiếm lợi nhuận, thực hiện việc đánh cắp danh tính, sử dụng như một đòn bẩy cho cuộc tấn công leo thang). Vào năm 2022, số lượng cuộc tấn công lừa đảo được báo cáo đạt đến con số khổng lồ là 255 triệu trường hợp, tăng 61% so với năm 2021. Bên cạnh đó các phương pháp hiện có để phát hiện đường dẫn URL lừa đảo bộc lộ nhiều sự hạn chế. Bài báo đề xuất một phương pháp để tăng độ chính xác trong việc phát hiện các URL độc hại bằng cách sử dụng các phương pháp học máy Vector Hỗ trợ Tuyến tính và đa thức Naive Bayes kết hợp với kỹ thuật voting (bỏ phiếu).**

***Abstract*— The phishing attack is the type of cyberattack that targets people's trust by masking the malicious intent of the attack as communications from reputable sources. The goal is to steal sensitive data from the victims (banking information, social identification, credentials, etc.) for various purposes (selling for monetary gain, performing identity thief, using as a lever for escalation attack). In 2022, the number of reported phishing attacks reach a whopping 255 million cases, an increment of 61% compared to 2021. Existing methods of phishing URL detection have limitations. The article proposes a method to increase the accuracy of detecting malicious URL by using machine learning methods Linear Support Vector Classification and multinomial Naive Bayes with voting mechanisms.**

Bài báo được gửi báo cáo trước đó tại Hội thảo quốc gia VNICT 2023, sau đó gửi Tạp chí vào ngày 15/9/2023. Bài báo được nhận xét bởi phản biện thứ nhất vào ngày 02/10/2023 và được chấp nhận đăng vào ngày 04/10/2023. Bài báo được nhận xét bởi phản biện thứ hai vào ngày 01/10/2023 và được chấp nhận đăng vào ngày 04/10/2023.

***Từ khóa— URL, lừa đảo, SVM, Naive Bayes, học máy.***

***Keywords— URL, phishing, SVM, Naive Bayes, machine learning.***

## I. GIỚI THIỆU

Trong suốt lịch sử phát triển của xã hội con người hiện đại, lừa đảo luôn là một trong những mối đe dọa không ngừng đối với thông tin và quyền riêng tư của người dùng. Từ những ngày đầu tiên, tội phạm đã sử dụng điện thoại rộng rãi để tống tiền người dùng thông qua nhiều hình thức giả mạo. Ngày nay, với sự phát triển của Internet và sự gia tăng liên tục giá trị của thông tin cá nhân, tội phạm mạng tiếp tục thực hiện các cuộc tấn công lừa đảo để khai thác tài sản và thông tin cá nhân của người dùng, trong một số trường hợp là dữ liệu riêng tư quan trọng. Internet ảnh hưởng đến cuộc sống ngày nay của chúng ta và thực trạng các lỗ hổng xuất hiện ngày một nhiều hơn, cho phép tội phạm mạng thực hiện các hành vi tấn công và khai thác thông tin.

Yếu tố đầu tiên và quan trọng nhất cho người dùng truy cập thông tin trên Internet là URL (Uniform Resource Locator). URL tương tự như địa chỉ của chúng ta, nó đại diện cho vị trí của các trang web trên Internet. Tội phạm mạng sẽ tạo ra các URL độc hại có vẻ giống với các URL mục tiêu để lừa người dùng và từ đó đánh cắp thông tin đăng nhập hoặc dữ liệu cá nhân của họ.

Với sự bùng nổ mạng Internet, con người dành nhiều thời gian trên ứng dụng mạng nhiều hơn trước đây. Điều này bao gồm cả những người dùng thông thường, sinh viên và các doanh nghiệp nhỏ, đến các tập đoàn đa quốc gia và chính phủ. Bởi vì yếu tố này, thông tin thu thập từ Internet rất được chú ý và vô cùng quý giá đối với cả chính phủ và doanh nghiệp. Thông tin thu thập có thể được sử dụng cho nhiều mục đích, từ

việc tăng độ chính xác của mô hình học máy, rộng hơn là mô hình trí tuệ nhân tạo hoặc cung cấp sản phẩm phù hợp hơn, quảng cáo đến một nhóm người, tìm kiếm nguồn gốc của một chủ đề hoặc mối quan tâm phổ biến trong một cộng đồng và nhiều mục đích khác. Nhưng những dữ liệu này cũng có thể được sử dụng với ý định xấu nếu rơi vào tay của tội phạm mạng. Ví dụ, nếu dữ liệu GPS của một người nào đó nằm trong tay một nhóm tội phạm, tin tặc có thể sử dụng nó để phân tích lịch trình và vị trí của nạn nhân nhằm đột nhập vào nhà họ hoặc định vị vị trí và thực hiện các hành vi xấu khác. Cuộc tấn công lừa đảo cũng hoạt động theo cách tương tự.

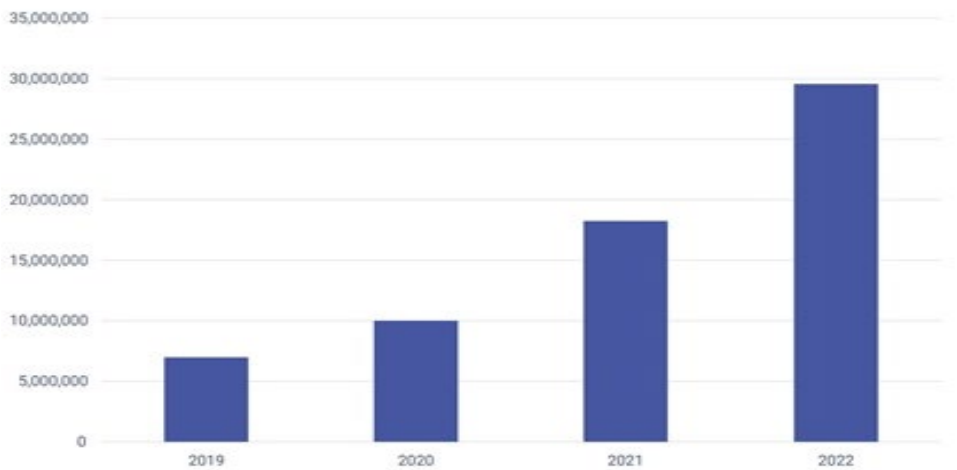
Có nhiều phương pháp khác nhau để phát hiện các URL lừa đảo: danh sách đen (blacklist), phương pháp heuristic và học máy. Phương pháp danh sách đen dễ thiết lập và sử dụng, ít xảy ra kết quả sai nhưng yêu cầu danh sách các URL lừa đảo cập nhật, cần phải được bổ sung liên tục. Ngoài ra, nó không hiệu quả đối với các cuộc tấn công lừa đảo mới xuất hiện (zero-days). Phương pháp heuristic cũng dễ dàng thiết lập và sử dụng, nhưng hiệu quả của chúng phụ thuộc vào tính chính xác của bộ quy tắc được áp dụng. Trong khi đó, phương pháp học máy rất hiệu quả trong việc phát hiện các mối đe dọa mới nhưng yêu cầu các tập dữ liệu chất lượng cao cùng tài nguyên tính toán đáng kể để huấn luyện và đánh giá mô hình học máy. Học máy đã đạt được độ chính xác đáng kể trong việc phát hiện các URL lừa đảo, với một số kết quả tồi nhất vẫn ở mức 0,75 và một số kết quả tốt nhất ở mức xung quanh 0,97.

## II. Đánh giá các nghiên cứu hiện đại về phát hiện URL lừa đảo

### *A. Phân loại tấn công lừa đảo và phương pháp phát hiện cuộc tấn công*

Các cuộc tấn công lừa đảo là việc sử dụng thông tin liên hệ giả mạo từ các nguồn đáng tin cậy như trang web ngân hàng, facebook,... để đánh lừa nạn nhân cung cấp thông tin nhạy cảm cho các tin tặc, điều này có thể đe dọa quyền riêng tư của người dùng. Các cuộc tấn công có thể tạo ra mối đe dọa về việc truy cập trái phép vào tài khoản trực tuyến của người dùng và dữ liệu cá nhân của họ, tin tặc có thể có khả năng sửa đổi và xâm phạm hệ thống liên quan. Đối với người dùng thông thường, điều này có thể gây ra một số rắc rối nhỏ hoặc mất mát về tài sản và dữ liệu, nhưng đối với những người có ảnh hưởng và tầm quan trọng đáng kể, đây có thể là một vấn đề nghiêm trọng.

Vào năm 2022, theo báo cáo của công ty an ninh mạng SlashNext [1], chứng kiến khoảng 255 triệu cuộc tấn công lừa đảo, tăng 61% so với năm 2021. 76% trong số các cuộc tấn công được phát hiện vào năm 2022 là về việc thu thập thông tin đăng nhập (credential harvesting), đó vẫn là nguyên nhân hàng đầu gây ra việc vi phạm nguyên tắc bảo mật. Ví dụ, đối với các công ty




Hình 1. Số lượng URL độc hại từ năm 2019 đến năm 2022

như Twilio, Cisco và Uber, các cuộc tấn công sẽ bắt đầu từ việc đánh cắp thông tin đăng nhập. Theo thống kê của hãng RedTeam Security [2], các lĩnh vực bị tấn công nhiều nhất là: Kinh doanh; Y tế/Y học; Ngân hàng/Tín dụng/Tài chính; Chính phủ/Quân đội; Giáo dục và Năng lượng/Dịch vụ tiện ích. Hình 1 cho thấy số lượng URL độc hại được báo cáo từ năm 2019 đến năm 2022, được thu thập bởi Liên minh quốc tế phòng chống lừa đảo APWG [3]. Từ biểu đồ, chúng ta có thể thấy sự gia tăng nhanh chóng số lượng URL độc hại suốt thời kỳ, tạo ra một xu hướng đáng lo ngại cho tương lai.

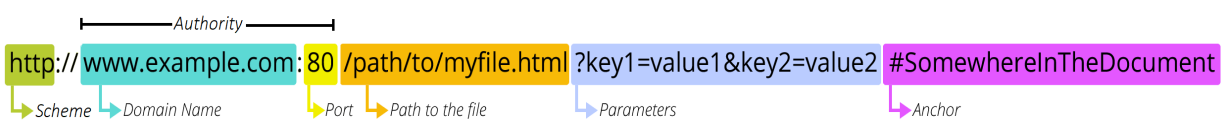
Để thực hiện cuộc tấn công lừa đảo, những kẻ tấn công cố gắng gửi các URL độc hại [4] đến nạn nhân, Hình 2 thể hiện cấu trúc thông thường của một URL. Mặc dù nhận thức về tấn công lừa đảo đã tăng lên trong nhiều năm qua, thế nhưng những kẻ tấn công vẫn đang phát triển các kỹ thuật của chúng để phá vỡ các tuyến phòng thủ của người dùng. Một số phương pháp cơ bản thường được phát triển bởi các tin tặc như việc kết hợp các liên kết hợp pháp với các liên kết độc hại. Vì nhiều ứng dụng như Gmail có khả năng lọc email có chứa liên kết độc hại, những kẻ tấn công sẽ cố gắng sử dụng ít nhất một liên kết hợp pháp cùng với các liên kết lừa đảo. Như vậy, email độc hại sẽ tránh được sự bảo vệ của ứng dụng Gmail.

*Lạm dụng chuyển hướng:* Một số URL lừa đảo có thể đánh lừa người dùng để cung cấp thông tin nhạy cảm. Khi thông tin này đã được nhập, người dùng sẽ được chuyển hướng đến trang web chính thống, khi đó người dùng có thể nghĩ rằng mình đã nhập chưa đúng thông tin nào đó và tiếp tục thực hiện thao tác lại trên trang web chính thức và sẽ không phát hiện ra.

*Ẩn mã độc bằng hình ảnh:* Một số bộ lọc quét văn bản trong email để tìm các yếu tố độc hại, vì vậy những kẻ tấn công thường muốn "làm mờ" người dùng bằng cách sử dụng hình ảnh và đồ họa để che giấu mã độc và có khả năng tránh các bộ lọc. Trong tất cả các loại cuộc tấn công lừa đảo thông thường, người dùng cần phải nhấp vào một liên kết độc hại, trừ trường hợp lừa đảo qua cuộc gọi thoại (voice phishing). Điều này có nghĩa là để các cuộc tấn công này thành công, người dùng cần truy cập vào các URL cụ thể đó. Vì vậy, nếu người quản trị có thể lọc URL từ đầu để xác định xem nó có phải là lừa đảo hay lành tính mà không cần sự nhận biết của người dùng, đó là cách tốt nhất để ngăn chặn tấn công.

Trong tình huống tốt nhất, chúng ta có thể giáo dục mọi người về lừa đảo và đảm bảo họ có đủ kiến thức về nó, nhưng việc này gần như không thể thực hiện được trong các tập đoàn lớn, vì nhiều người và mỗi người sẽ có các mức độ hiểu biết về công nghệ khác nhau. Hiện nay, cách phổ biến nhất để phát hiện các URL lừa đảo là sử dụng danh sách đen. Đây là phương pháp dễ dàng để thiết lập, triển khai và duy trì. Các công ty có thể mua một cơ sở dữ liệu về URL lừa đảo luôn được cập nhật và thiết lập một bộ lọc trên mạng của họ để kiểm tra với cơ sở dữ liệu này. Danh sách đen đơn giản là việc tạo ra một cơ sở dữ liệu các liên kết đã biết là độc hại hoặc liên kết không mong muốn. Khi người dùng cố gắng truy cập một URL, hệ thống sẽ so sánh URL đã nhập với dữ liệu trong cơ sở dữ liệu. Phương pháp này có thể đạt được kết quả phát hiện hoàn toàn chính xác, nhưng có thể xảy ra trường hợp phân loại sai khi dữ liệu không chính xác được nhập vào cơ sở dữ liệu.

Với heuristic, phương phán này sẽ phân tích, tìm các mẫu và đặc điểm của các URL lừa đảo, sau đó thiết lập một quy tắc để quyết định liệu một URL có phải là lừa đảo hay lành tính. Heuristic thường phát hiện nhanh chóng và dễ triển khai, nhưng có thể không hiệu quả trong các




Hình 2. Các thành phần của một URL

trường hợp quy tắc không bao gồm tất cả các biến thể có thể của một URL độc hại.

Phương pháp học máy là một thuật toán huấn luyện để nhận biết các mẫu của các URL, sau đó sử dụng để phân loại các URL mới là lừa đảo hoặc lành tính. Phương pháp này có thể rất hiệu quả trong việc phát hiện các mối đe dọa mới, nhưng có thể đòi hỏi lượng lớn dữ liệu cho quá trình đào tạo và tài nguyên tính toán để hoạt động một cách hiệu quả.

*B. Phát hiện lừa đảo sử dụng học máy*

Trong nghiên cứu [5] được công bố vào năm 2022, các tác giả đã đề xuất một mô hình học máy để phát hiện lừa đảo thông qua phân tích URL. Mô hình được huấn luyện trên một tập dữ liệu gồm 6000 dữ liệu được thu thập từ PhishTank và Alexa. Họ đã trích xuất 10 đặc trưng từ một URL:

- Độ dài của URL.
- Số lượng dấu '.' trong một tên miền.
- Sử dụng địa chỉ IP.
- Sự có mặt của kí hiệu '@' trong URL.
- Sự có mặt của kí hiệu '-' trong tên miền.
- Sự tồn tại của HTTPS trong một tên miền.
- Sự chuyển hướng của HTTP trong URL.
- Dịch vụ URL rút gọn.
- Số lượng trang thành phần.

Các đặc điểm này huấn luyện và kiểm tra trên 8 thuật toán học máy khác nhau (DT - Decision Tree, RF - Random Forest, LR - Logistic Regression, XGboost, SVM - Support Vector Machine, KNN - K Nearest Neighbors, A/B testing, MLP - Multi-layers Perceptron) và đã phát hiện ra rằng MLP đã tạo ra độ chính xác tốt nhất ở mức 85,41%.

Trong bài báo [6] được công bố vào năm 2022, các nhà nghiên cứu đã thực hiện phân tích về việc phát hiện các trang web lừa đảo dựa trên học máy. Họ đã so sánh 05 kỹ thuật học máy: DT, RF, KNN, Gaussian, NB và XGBoost. Cuộc thử nghiệm được thực hiện trên một tập dữ liệu gồm 11.430 URL (với tỷ lệ cân bằng giữa các URL lừa đảo và các URL lành tính) với 87 đặc trưng được trích xuất. Các đặc trưng này được chia thành ba loại:

- 56 đặc trưng dựa trên cấu trúc và cú pháp của các URL.
- 24 đặc trưng dựa trên nội dung của các trang tương ứng.
- 7 đặc trưng dựa trên việc truy vấn các dịch vụ khác.

Với các đặc trưng và thuật toán được lựa chọn, các nhà nghiên cứu đã tiến hành thử nghiệm. Họ đã thấy thuật toán RF đạt được độ chính xác cao nhất là 97%, vượt qua 4 thuật toán khác, trong đó XGBoost đứng ở vị trí thứ hai với 94,79%.

Trong bài báo [7] được công bố vào năm 2020, tác giả đã đề xuất một mô hình sử dụng RF, SVM và mạng neuron lan truyền. RF và SVM đã xuất hiện trong nhiều bài báo khác về vấn đề này được dùng để so sánh thực nghiệm. Việc sử dụng cách thức lan truyền ngược nhằm giảm lỗi trong kết quả cuối cùng khi lỗi được truyền ngược lại và trọng số được gán cho mỗi neuron trong các lớp ẩn thay đổi sau mỗi lần lặp. Tập dữ liệu trong bài báo này được lấy từ kho lưu trữ học máy UCI, chứa khoảng 11.000 URL với 6.157 URL lừa đảo và 4.898 URL lành tính. Để xử lý đầu vào, các nhà nghiên cứu sử dụng một tiện ích mở rộng trên Google Chrome để trích xuất 13 đặc trưng. Cuối cùng, các nhà nghiên cứu đã đạt được độ chính xác trên 3 thuật toán là: RF 97,37%, SVM 97,45% và NN Backpropagation 97,26%.

Trong bài báo [8] được công bố vào năm 2020 đã thảo luận về hiệu quả của các phương pháp khác nhau để phát hiện lừa đảo: danh sách đen, heuristic, dựa trên tương đồng hình ảnh, dựa trên máy tìm kiếm và các phương pháp học máy. Họ đã chỉ ra một số điểm mạnh và yếu của mỗi

phương pháp và những gì có thể được thực hiện để cải thiện hiệu quả trong việc phát hiện lừa đảo.

Bài báo [9] được công bố vào năm 2021, các tác giả đã đánh giá 11 thuật toán như: DT, KNN, GB - Gradient Boosting, LR, NB, RF, SVM, kNN, ET, AB testing và B. Tập dữ liệu mà họ sử dụng chứa 2.843 URL lành tính và 8.495 URL lừa đảo, tỷ lệ mất cân bằng rất cao. Họ đã trích xuất 9 đặc trưng từ URL để huấn luyện tất cả các mô hình. Kết quả thu thập từ tất cả các thuật toán dao động từ 0,75 đến 0,9, thuật toán thực hiện tốt nhất là Extra Tree và kết quả kém nhất là Neural Network. Một số điểm số thấp có thể là kết quả của việc sử dụng một tập dữ liệu mất cân bằng, điều này là cần chú ý trong nghiên cứu và thử nghiệm của chúng ta.

Trong bài báo [10] được công bố vào năm 2021, các nhà nghiên cứu đã đề xuất một phân loại theo ba giai đoạn của một URL lừa đảo - hợp lệ, không đủ dữ liệu, không hợp lệ. Họ đã tiến hành các thử nghiệm trên các tập dữ liệu do Đại học California Irvine cung cấp với 5 thuật toán học máy (Gradient Tree Boosting, RF, SVM, Extra RF, LR). Họ đã đạt được độ chính xác trong khoảng từ 74-82%, với LR có độ chính xác tốt nhất.

Trong bài báo [11] được công bố vào năm 2023, các nhà nghiên cứu đã đề xuất một mô hình học máy để phân loại các URL độc hại bao gồm các URL lừa đảo. Bài báo này sử dụng một tập dữ liệu gồm 651.191 trang web lấy từ kho lưu trữ Kaggle, bao gồm 428.103 URL lành tính, 96.457 URL bị thay đổi, 94.111 URL lừa đảo và 32.520 URL độc hại. Tỷ lệ của tập dữ liệu này rất mất cân bằng giữa các URL độc hại và lành tính. Các nhà nghiên cứu đã trích xuất 18 đặc trưng từ URL. Để huấn luyện mô hình, họ sử dụng 3 thuật toán (RF, LightGBM, XGBoost) với độ chính xác là 0,966; 0,932 và 0,956 theo thứ tự. Đây là kết quả đầy hứa hẹn từ một tập dữ liệu rất mất cân bằng.

Thông qua các bài báo nghiên cứu trong những năm gần đây, chúng ta có thể thấy học máy đã đạt được độ chính xác đáng kể trong việc phát hiện các URL lừa đảo, với một số kết quả tệ nhất vẫn ở mức 0,75 và một số kết quả tốt nhất ở mức xung quanh 0,97. Có một loạt các đặc điểm đặc biệt được trích xuất từ URL để huấn luyện mô hình học máy, nhưng một số đặc điểm ổn định bao gồm độ dài của URL, tên miền, tham số URL, địa chỉ IP và số lượng của mỗi ký tự đặc biệt.

Trong khảo sát các nghiên cứu trước đó, chúng ta thấy có đánh giá về nhiều phương pháp khác nhau để phát hiện các URL lừa đảo. Phương pháp danh sách đen dễ thiết lập và chạy, ít dễ bị sai lệch tích cực nhưng yêu cầu một cơ sở dữ liệu lớn về các URL lừa đảo và cần phải được cập nhật liên tục để hiệu quả và không hiệu quả đối với các cuộc tấn công phishing zero-day. Phương pháp heuristic cũng nhanh chóng và dễ dàng để thiết lập và chạy, nhưng nó chỉ hiệu quả khi được áp dụng đúng bộ quy tắc. Phương pháp học máy rất hiệu quả trong việc phát hiện các mối đe dọa mới, nhưng nó đòi hỏi một lượng lớn dữ liệu cho quá trình huấn luyện và một tập dữ liệu chất lượng. Không chỉ vậy, nó còn đòi hỏi sức mạnh tính toán đáng kể cho tất cả quá trình huấn luyện, phát hiện và đánh giá các mô hình học máy

## III. PHÁT TRIỂN PHƯƠNG PHÁP PHÁT HIỆN URL LỪA ĐẢO SỬ DỤNG HỌC MÁY

### *A. Nghiên cứu về các phương pháp hiện có*

phishing_site_urls.csv > find.py > ...

```
import csv
import time

URL = input("Enter URL: ")
start_time = time.time()

with open('D:\master 2nd year\prac\phishing_site_urls.csv\phishing_site_urls.csv','r',encoding='utf8') as f:
    reader = csv.reader(f, delimiter=',')
    for row in reader:
        if URL == row[0] :
            print("is in file")
            print(row[1])
            print("--- %s seconds ---" % (time.time() - start_time))
            break
    else:
        print("not in file")
        print("--- %s seconds ---" % (time.time() - start_time))
```

Hình 3. Tập lệnh Python để kiểm tra URL

*Danh sách đen:* Đầu tiên tiến hành khảo sát phương pháp danh sách đen. Nhóm tác giả sẽ viết một đoạn script nhỏ để kiểm tra các URL đã có trong tập dữ liệu thu thập được. Tập dữ liệu đã

được phân loại thành 2 lớp với các URL bình thường và các URL lừa đảo.

Khi đó, nhóm tác giả nhận được kết quả phù hợp kỳ vọng, với một URL đã biết thì việc phát hiện là 100% chính xác, nhưng nếu chúng ta thay đổi bất cứ điều gì ngay cả một ký tự, nó sẽ không thể phát hiện được. Nhóm tác giả đã chạy kịch bản trên phần cứng mạnh mẽ hiện đại (Ryzen 9 6900HS, CPU 8 nhân 16 luồng), nhưng vẫn thấy sự khác biệt lớn về thời gian chạy (gần 500 lần khác nhau) giữa các URL đầu vào đã biết và chưa biết. Điều này có nghĩa là với một tập dữ liệu lớn hơn, chúng ta có thể mong đợi thấy sự khác biệt về thời gian chạy còn cao hơn.

Mặt khác của việc sử dụng danh sách đen là danh sách trắng, phương pháp này cực kỳ hạn chế. Nó chỉ cho phép người dùng truy cập vào một danh sách cụ thể các trang web tương tự danh sách đen nhưng chứa các trang web đã được xác minh. Phương pháp này có sự tiện lợi của việc sử dụng Internet an toàn, nhưng số lượng các website quá nhiều nên không thể dự đoán được trang web nào nên được sử dụng hoặc không nên sử dụng, vì vậy phương pháp này không phải là một lựa chọn hợp lý.

*Phương pháp dựa trên heuristic:* Trong phương pháp heuristic, chúng ta có thể tạo ra một tập lệnh Python đơn giản và một bộ quy tắc, sau đó sử dụng tập dữ liệu đầu tiên làm đầu vào. Các quy tắc như sau:

- Từ khóa đáng ngờ: Kiểm tra xem URL có chứa bất kì từ khóa đáng ngờ nào sau đây không: “login”, “sign in”, “account”, “security”, “update”, “verify”, “validate”, “reset”, “support”, hoặc “service”.

- Phần tên miền phụ dài: Kiểm tra xem phần miền phụ của URL có dài hơn 15 kí tự hay không.

- Nhiều miền phụ: Kiểm tra xem URL có nhiều hơn một thành phần miền phụ hay không.

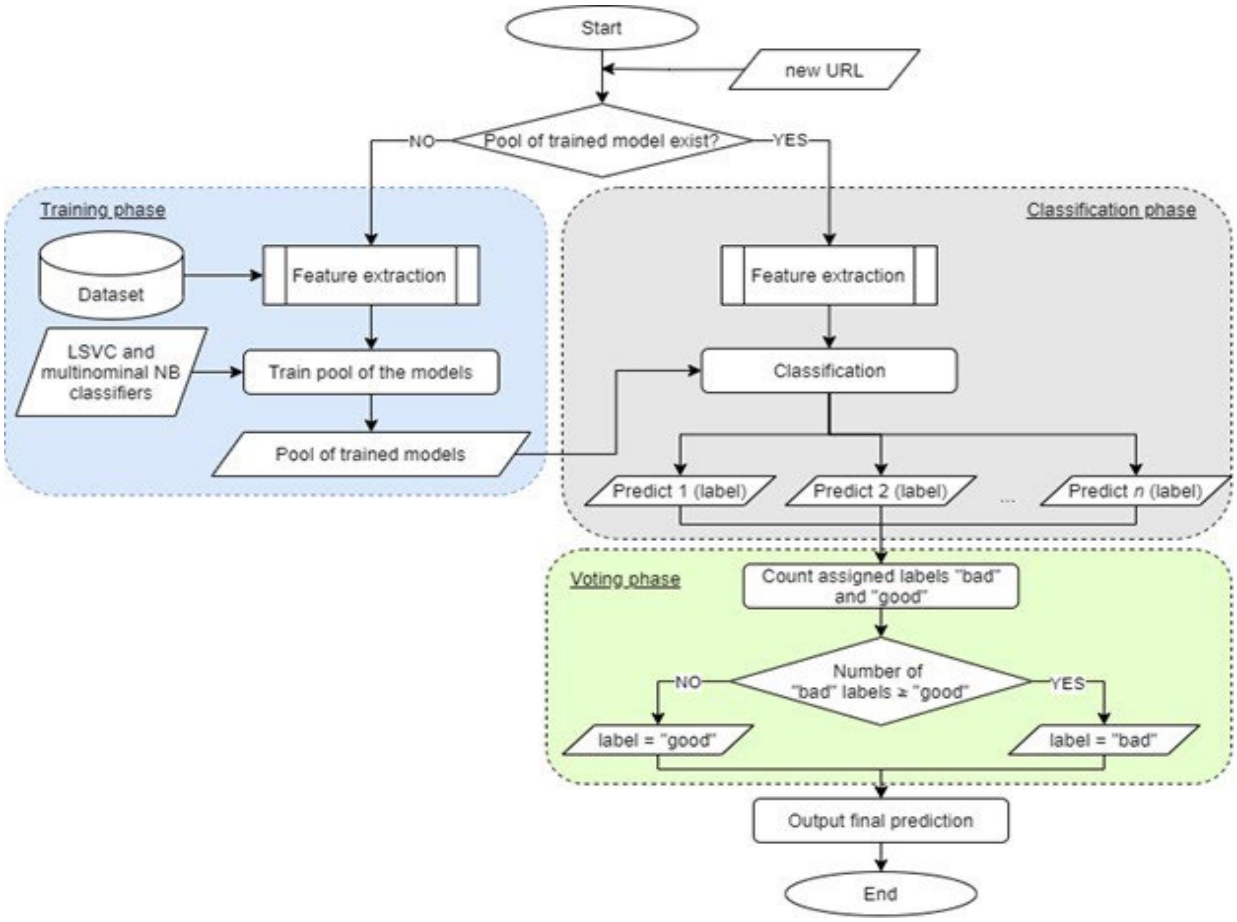
- Miền giả mạo: Kiểm tra xem phần mở rộng tên miền cấp cao (TLD) của URL có đáng ngờ hay không. Điều này được thực hiện bằng cách so sánh TLD với danh sách các TLD đã biết thường được sử dụng bởi các trang web chính thống.

Nếu bất kỳ URL nào không đáp ứng một trong bất kỳ những quy tắc thì URL sẽ được đánh dấu là lừa đảo, nếu ngược lại sẽ được đánh dấu là một URL lành tính. Với phương pháp này, chúng ta chỉ có thể đạt được độ chính xác khoảng 50%, tốt hơn so với phương pháp dựa trên danh sách đen, vì nó không yêu cầu một tập dữ liệu các URL và đạt được một số độ chính xác trên các URL không biết nhưng nó chỉ hiệu quả như tập luật được thiết lập bởi người phát triển.

*Phương pháp dựa trên đám đông:* Phương này hoạt động dựa trên sự bỏ phiếu của người dùng Internet để quyết định xem một trang web có phải là lừa đảo hay lành tính hay không. Một trang web nổi tiếng cung cấp dịch vụ này là PhishTank. Phương pháp dựa trên đám đông chậm trong việc đối phó với bất kỳ cuộc tấn công zero-day hoặc 1-day nào vì nó đòi hỏi sự bỏ phiếu đáng kể từ người dùng để phân loại một trang web. Không chỉ vậy, người dùng có thể bỏ phiếu cho các trang web cũng cần được xác minh hoặc tin tặc có thể sử dụng bot để bỏ phiếu hàng loạt cho các trang web lừa đảo của chúng. Phương pháp này cũng không phù hợp để áp dụng trong các tập đoàn vì nó đòi hỏi một lực lượng lao động lớn để thực hiện việc bỏ phiếu, nó chỉ phù hợp để xây dựng cơ sở dữ liệu cho các phương pháp dựa trên danh sách đen.

Để mô phỏng và xác định cách phương pháp này hoạt động, cuộc thử nghiệm đã được tiến hành bằng cách mời một nhóm người tham gia làm công cụ phân loại. Các URL chưa biết sẽ được gửi cho họ trong các khoảng thời gian khác nhau và với số lượng khác nhau. Một URL chỉ được xác định khi đa số người bỏ phiếu đưa ra cùng một ý kiến là lành tính hoặc lừa đảo.

Kết quả thu thập từ cuộc thử nghiệm này rất chính xác trong việc phân biệt URL lừa đảo và liên kết lành tính, nhưng thời gian cần thiết để




Hình 4. Quy trình làm việc của phương pháp được đề xuất

hoàn thành quá trình phân loại biến động mạnh tùy thuộc vào thời gian mọi người nhận được URL mới để xác định và thời gian mà URL mới được gửi đi.

*B. Nghiên cứu trên thuật toán học máy*

Có nhiều thuật toán trong học máy được sử dụng cho nhiệm vụ phân loại và tất cả đều có ưu điểm và nhược điểm riêng của chúng. Dưới đây là một số thuật toán học máy thường được đề cập và sử dụng trong nghiên cứu khoa học về việc phát hiện các URL lừa đảo.

- *LSVC:* Phương pháp học máy phân tách hoặc phân loại dữ liệu đầu vào thành một "mặt phẳng tốt nhất" để làm cho nó tương ứng với dữ liệu mà người dùng cung cấp. Sau đó, chúng ta có thể cung cấp một số đặc điểm cho bộ phân loại để lấy được lớp "dự đoán" sau khi có được mặt phẳng tốt nhất.

- *Bộ phân loại LR:* Một kỹ thuật phân loại học máy sử dụng một số biến phụ thuộc để tính toán xác suất của một lớp cụ thể. Mô hình LR kết hợp tất cả các đặc điểm của đầu vào và ước tính hợp lý của kết quả.

- *Multinomial Naiev Bayes:* Một chiến lược học Bayesian phổ biến trong xử lý ngôn ngữ tự nhiên (NLP). Định lý Bayes được sử dụng để xác định một thẻ dữ liệu. Nó đánh giá khả năng của mỗi thẻ cho một mẫu cụ thể và đầu ra với khả năng cao nhất.

- *Bộ phân loại DT:* Thuật toán cho cả các nhiệm vụ hồi quy và phân loại thường sử dụng cây quyết định. Nói một cách đơn giản, chúng nói đến một tập hợp có thứ tự các câu hỏi if/else dẫn đến một lựa chọn. Xác định chuỗi câu hỏi if/else dẫn chúng ta đến lời giải chính xác nhanh nhất là yếu tố quan trọng để hiểu cách sử dụng một cây. Thuật toán duyệt qua tất cả các thử

nghiệm có thể và chọn thử nghiệm chứa nhiều thông tin nhất về biến mục tiêu để tạo cây.

- *Bộ phân loại RF:* Rừng ngẫu nhiên RF hiện nằm trong số các kỹ thuật học máy phổ biến nhất cho phân loại và hồi quy. Theo cách cơ bản, RF là một nhóm các cây quyết định DT, trong đó mỗi cây được chọn ngẫu nhiên làm cây gốc.

Ý tưởng đằng sau RF là trong khi mỗi cây có thể đưa ra dự đoán khá chính xác, chúng gần như chắc chắn sẽ bị "khớp" overfitting trên một số phần của dữ liệu. Bằng cách trung bình hóa kết quả của nhiều cây, chúng ta có thể giảm bớt mức độ quá khớp. Chúng hiệu quả và thường hoạt động tốt mà không cần nhiều điều chỉnh tham số cũng như không yêu cầu chuẩn hóa dữ liệu.

### *C. Phát triển phương pháp kết hợp để phát hiện URL lừa đảo bằng cách sử dụng học máy*

Phương pháp lai ghép tổng hợp nhiều mô hình khác nhau và dự đoán kết quả đầu ra dựa trên mô hình có độ tin cậy cao nhất làm đầu ra. Nó đơn giản là tổng hợp kết quả của mỗi bộ phân loại được truyền vào Bộ bỏ phiếu và dự đoán lớp đầu ra dựa trên phần đông bằng phiếu bầu. Trong phương pháp lai ghép của nhóm tác giả sử dụng các thuật toán LSVC và đa thức NB. Ý tưởng là thay vì tạo ra các mô hình trên độc lập và tìm độ chính xác cho mỗi mô hình, nhóm tác giả tạo ra một mô hình duy nhất huấn luyện và dự đoán đầu ra dựa trên tổng hợp đa số phiếu bầu cho mỗi đầu ra. Ví dụ, nếu dự đoán cho một URL, cụ thể là mô hình 1 - "xấu", mô hình 2 - "xấu", mô hình 3 - "tốt". Cơ chế bỏ phiếu sẽ phân loại mẫu như "xấu" dựa trên nhãn lớp phần đông. Trong bước trích xuất đặc trưng, có 12 đặc trưng được xác định từ tập các URL:

- Độ dài của URL được cung cấp.

- Vị trí của tên miền lớp trên cùng.

- URL có tồn tại dưới dạng địa chỉ IP (is_ip).

- Số lượng dấu '–' có mặt trong tên miền (domain_hyphens).

- Số lượng gạch dưới '_' có mặt trong tên miền (domain_inderscocres).

- Số lượng dấu '–' có mặt trong đường dẫn của URL.

- Số lượng dấu '_' có mặt trong đường dẫn của URL.

- Số lượng dấu '/' có mặt trong URL.

- Số lượng dấu chấm '.' có mặt trong URL.

- Số lượng phần phụ tên miền có mặt trong URL (num_subdomains).

- Thông tin tên miền trong URL (domain_token).

- Thông tin đường dẫn trong URL (path_tokens).

Điểm mới về mặt khoa học của phương pháp đề xuất (Hình 4) nằm ở việc phân loại kết hợp các thuật toán học máy, phương pháp này được sử dụng để phát hiện các URL lừa đảo (sử dụng phương pháp học máy LSVC và NB đa thức với cơ chế bỏ phiếu).

## IV. ĐÁNH GIÁ HIỆU QUẢ CỦA CÁC GIẢI PHÁP SO SÁNH VỚI TRÌNH ĐỘ KHOA HỌC VÀ CÔNG NGHỆ HIỆN ĐẠI

### *A. Các chỉ số đánh giá chất lượng*

Accuracy (ACC) được định nghĩa trong Công thức (1) là phần trăm dự đoán đúng cho dữ liệu kiểm tra. Nó có thể được tính trực tiếp bằng cách chia số lượng dự đoán đúng cho số lượng tất cả các dự đoán:

$$ACC = \frac{TP+TN}{TP+TN+\mathrm{FP}+FN} \quad (1)$$

Precision được định nghĩa trong Công thức (2) là tỷ lệ các mẫu có liên quan (dương tính thật) trong số tất cả các mẫu được dự đoán thuộc về một lớp nhất định:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Giá trị Recall được định nghĩa trong Công thức (3) là tỷ lệ các mẫu được dự đoán thuộc về một lớp so với tất cả các mẫu thực sự thuộc về lớp đó:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Chỉ số phân loại:

- True Positive (TP): Tổng số trường hợp dự báo khớp Positive.

- True Negative (TN): Tổng số trường hợp dự báo khớp Negative.

- False Positive (FP): Tổng số trường hợp dự báo các quan sát thuộc nhãn Negative thành trường hợp Positive.

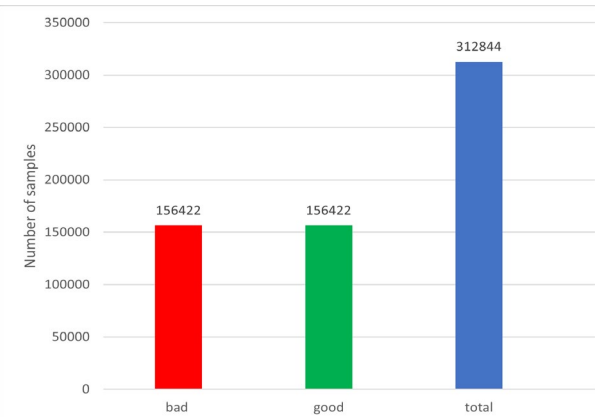
- False Negative (FN): Tổng số trường hợp dự báo các quan sát thuộc nhãn Positive thành trường hợp Negative.

*B. Dữ liệu chuẩn bị*

Đầu tiên, cần thu thập dữ liệu cho bộ dữ liệu của chúng ta. Nhóm tác giả đã tìm thấy một số bộ dữ liệu công khai có sẵn trên Kaggle. Có nhiều cơ sở dữ liệu nổi tiếng hơn như PhishTank và OpenPhish, nhưng chúng yêu cầu một số gói đăng ký hoặc chỉ dành cho các nhà phát triển mới có thể truy cập vào cơ sở dữ liệu của họ. Nhóm tác giả sẽ sử dụng 2 bộ dữ liệu [12, 13] từ Kaggle, chứa lần lượt 549.346 bản ghi và 38.800 bản ghi riêng biệt của URL lừa đảo và URL lành tính đã được gắn nhãn là tốt (good) và xấu (bad). Bộ dữ liệu 1 lớn hơn đáng kể so với bộ dữ liệu 2, vì vậy nhóm tác giả sẽ sử dụng để huấn luyện và đánh giá ban đầu phương pháp của mình và bộ dữ liệu 2 sẽ được sử dụng làm dữ liệu kiểm tra. Bộ dữ liệu 1 có sự mất cân bằng lớn giữa các URL lừa đảo và tốt, có thể sử dụng chúng như vậy, nhưng cũng có thể loại bỏ một số dữ liệu để làm cho nó cân bằng hơn. Bộ dữ liệu 2 nhỏ hơn, nhưng cân bằng giữa hai loại.

Vì bộ dữ liệu 1 rất không cân bằng nên cần cân bằng lại chúng. Sau khi đã cân bằng lại bộ dữ liệu đầu tiên, thu được kết quả như Hình 5.

Trong cơ sở dữ liệu ban đầu chỉ nhận được hai cột URL và nhãn (tốt/xấu). Nhóm tác giả cần trích xuất dữ liệu để lấy thêm dữ liệu từ URL thô. Khi đó, nhóm sử dụng thư viện urllib.Parse




Hình 5. Tỷ lệ của bộ dữ liệu đầu tiên sau khi được cân bằng

[14] python3 đã xử lý dữ liệu trước khi trích xuất tính năng. Các thuộc tính dữ liệu trước khi trích xuất đặc trưng của bộ dữ liệu 1 với 6 loại thông tin, bao gồm: Giao thức: 108 mục; Tên miền: 312.840 mục; Đường dẫn: 308.407 mục; Tham số: 131 mục; Truy vấn: 58.645 mục; Đoạn: 317 mục.

Các thuộc tính dữ liệu trước khi trích xuất đặc trưng của bộ dữ liệu 2, bao gồm: Giao thức: 38.800 mục; Tên miền: 38.800 mục; Đường dẫn: 38.789 mục; Tham số: 93 mục; Truy vấn: 7.702 mục; Đoạn: 215 mục.

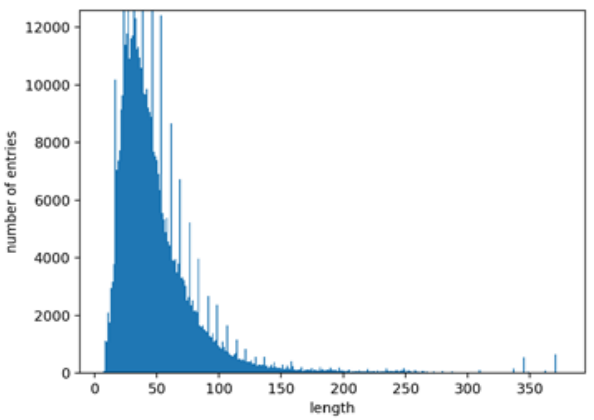
Từ việc trích xuất dữ liệu ban đầu, chúng ta có thể thấy rằng bộ dữ liệu 1 thiếu thông tin về giao thức của URL so với bộ dữ liệu 2. Cuối cùng, từ đây chúng ta có thể bắt đầu quá trình trích xuất đặc trưng và có thể thấy sự khác biệt rất lớn giữa các loại thông tin khác nhau, đây chỉ là một số URL có đoạn (fragment), tham số (params) và truy vấn (query) ít hơn đáng kể so với đường dẫn (path) và tên miền. Số lượng dữ liệu về giao thức trong các bộ dữ liệu 1 và thứ 2 hoàn toàn đối lập. Trong bộ dữ liệu 1, URL dường như thiếu dữ liệu này, có hai loại giao thức (HTTP và HTTPS), thông thường, HTTPS cho biết kết nối giữa máy khách và máy chủ là an toàn và thường liên quan đến một trang web hợp pháp, nhưng tin tặc đã nhận thấy điều này và có rất nhiều trang web lừa đảo với giao thức HTTPS.

Phân tích phân bổ của từng đặc trưng trong bộ dữ liệu 1. Hình 6 thể hiện phân bổ từ khóa

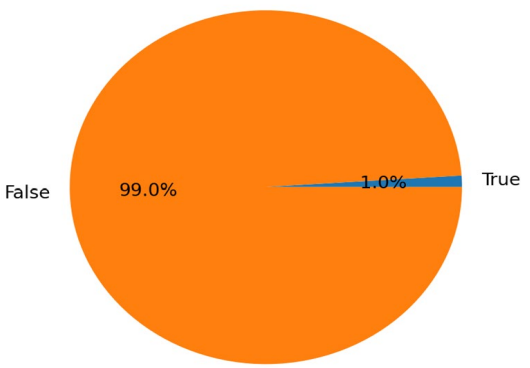

Hình 6. Phân phối các đặc trưng TLD trong bộ dữ liệu 1

trong tập hợp tên miền TLD của bộ dữ liệu 1 và cho thấy các TLD phổ biến nhất là "org", "net", "ca", "cu", "uk" và "org".

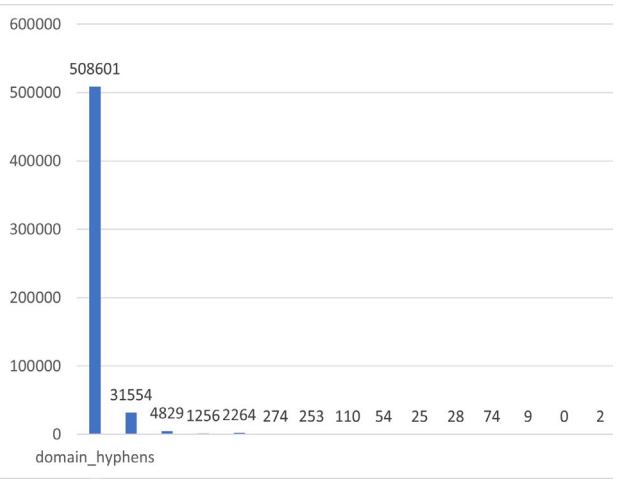
Hình 7 thể hiện phân phối các đặc trưng về độ dài trong bộ dữ liệu 1, cho thấy hầu hết dữ liệu dài dưới 50 từ.




Hình 7. Phân phối các đặc trưng về độ dài trong bộ dữ liệu 1




Hình 8. Phân phối về đặc trưng là IP của bộ dữ liệu 1




Hình 9. Phân phối tính năng dấu gạch nối tên miền trong tập dữ liệu 1

Hình 8 thể hiện phân phối số lượng địa chỉ IP có trong bộ dữ liệu, chiếm khoảng 1% của tổng số bộ dữ liệu.

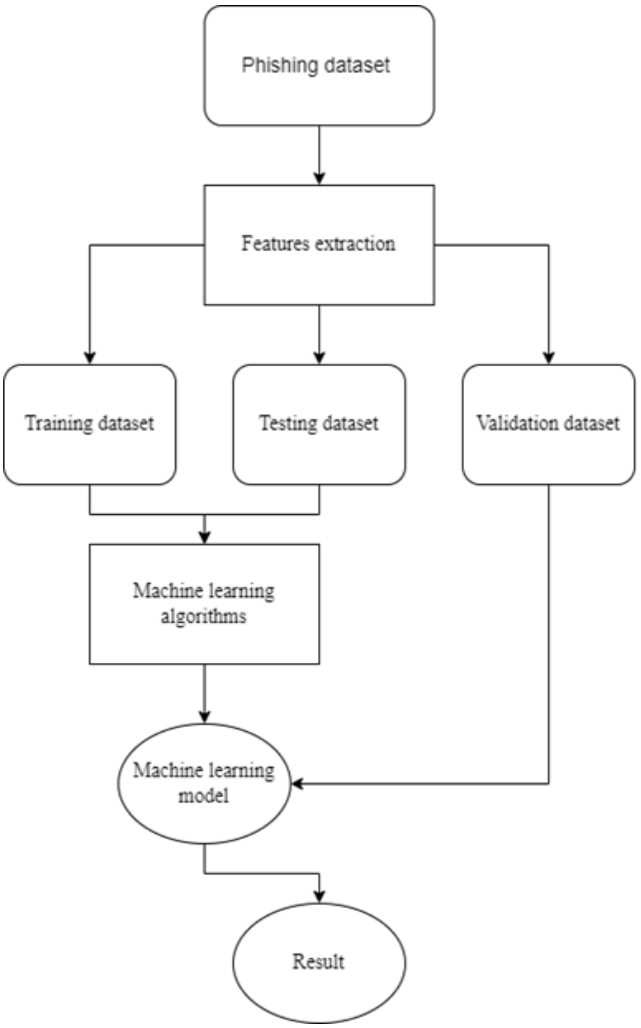
Hình 9 cho thấy hầu hết các URL chứa 1 dấu gạch ngang trong tên miền. Nhóm tác giả cũng thực hiện phân tích tương tự cho các đặc trưng còn lại và bộ dữ liệu 2.

*C. Xây dựng mô hình, so sánh với các mô hình học máy khác.*

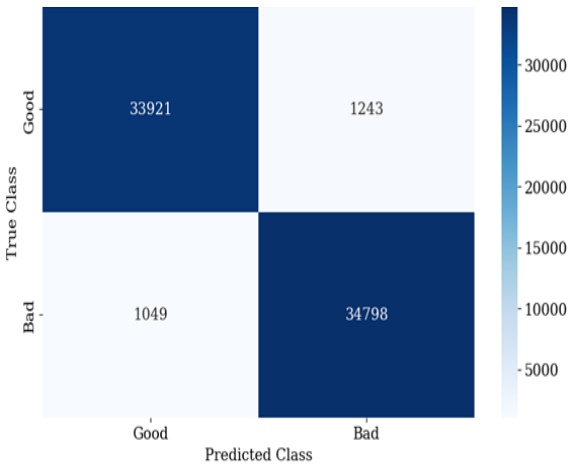
Để xây dựng mô hình, nhóm tác giả đã sử dụng thư viện Scikit-learn để xây dựng mô hình và sử dụng LSVC [15], LR [16], Multinomial NB [17], DT classifier [18], RF classifier [19] và Voting classifier [20]. Để huấn luyện các mô hình, nhóm tác giả thực hiện chia tách bộ dữ liệu thành 80% cho việc huấn luyện và 20% cho việc phát hiện. Hình 10 thể hiện quy trình làm việc của toàn bộ quá trình thử nghiệm.

Nhóm tác giả sử dụng Python để viết mã vì nó có nhiều thư viện hữu ích cho học máy như Scikit-learn, matplotlib, pandas và NumPy để huấn luyện mô hình, xử lý bộ dữ liệu và trực quan hóa dữ liệu. Từ kết quả thu thập được sẽ phát triển phương pháp mới của mình bằng cách sử dụng thư viện Voting classifier [20].

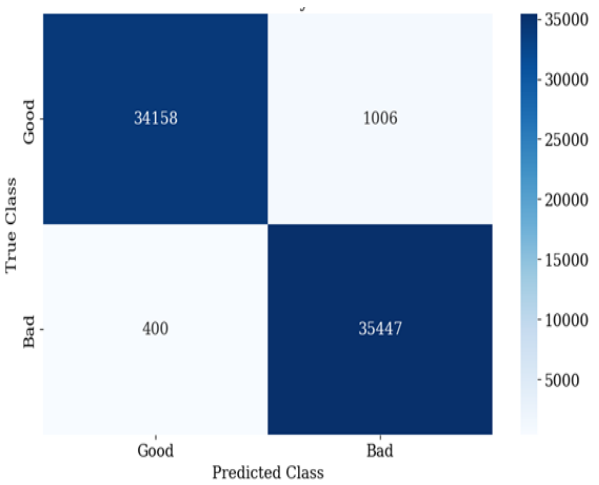
Sau khi huấn luyện và kiểm thử tất cả các thuật toán trên bộ dữ liệu đầu tiên, kết quả cho thấy LSVC đạt được độ chính xác tốt nhất, vì vậy ban đầu, nhóm đã kết hợp nó với các thuật toán




Hình 10. Quy trình làm việc của quá trình thử nghiệm




Hình 11. Ma trận nhầm lẫn của LSVC




Hình 12. Ma trận nhầm lẫn của phương pháp Hybrid

khác để xây dựng phương pháp của mình. Phương pháp của nhóm tác giả đã cho một kết quả có độ chính xác là 0,981, chỉ thấp hơn kết quả hàng đầu của LSVC (0,982) một chút với một biên độ nhỏ là 0,001.

Hình 11 thể hiện biểu đồ nhầm lẫn từ dữ liệu của mô hình LSVC chứa các giá trị của TP, TN, FP và FN. Điểm LSVC: 0,967.

Hình 12 thể hiện biểu đồ nhầm lẫn từ dữ liệu của mô hình Hybrid chứa các giá trị của TP, TN, FP và FN. Điểm số của phương pháp Hybrid là 0,980.

Bảng 1 thể hiện kết quả của phương pháp Hybrid trên các bộ dữ liệu của nhóm tác giả. Sau khi đánh giá, có thể thấy thuật toán Hybrid hoạt động tốt nhất trong số tất cả các thuật toán đã thử nghiệm. Nhóm tác giả đang thử nghiệm tất cả các thuật toán lại trên dữ liệu hoàn toàn mới. Trên bộ dữ liệu này, tất cả các kết quả thuật toán giảm độ chính xác. LSVC giảm từ 0,967 xuống 0,838; Multinomial NB phân loại giảm từ 0,968 xuống 0,803; LR classification giảm từ 0,961 xuống 0,904; DT classification giảm từ 0,880 xuống 0,669; RF classification giảm từ 0,826 xuống 0,715; phương pháp được đề xuất (phân loại Hybrid) giảm từ 0,980 xuống 0,929. Sau khi thử nghiệm với các bộ dữ liệu khác nhau, có thể

BẢNG 1. CÁC KẾT QUẢ CỦA PHƯƠNG PHÁP HYBRID

| Label | Precision | Recall |
|---|---|---|
| Bad | 0,99 | 0,97 |
| Good | 0,97 | 0,99 |

thấy phương pháp Hybrid vẫn hoạt động tốt hơn so với các thuật toán khác và cũng thấy sự giảm độ chính xác thấp nhất.

Trong những năm gần đây, đã có nhiều bài nghiên cứu tìm hiểu về vấn đề sử dụng học máy để phát hiện các URL lừa dảo. Nhiều mô hình đã được đề xuất với các thuật toán khác nhau được sử dụng và các đặc điểm khác nhau được trích xuất từ URL gốc [21], thậm chí có những mô hình vượt ra ngoài phạm vi của phân tích URL và kết hợp nhiều đặc điểm từ trang web chính như biểu tượng trang web (favicon), mã nguồn HTML, độ phổ biến của URL,…

Với kết quả thu thập được, chúng ta có thể thấy rằng phương pháp mới được phát triển bằng cách sử dụng học máy hoạt động tốt hơn so với tất cả 5 thuật toán khác: LSVC, LR, đa thức NB, DT, RF và thậm chí cả một số bài nghiên cứu khoa học đã được công bố. Với điều này, chúng ta cụ thể hóa được những ưu điểm của việc sử dụng máy học để phát hiện hơn so với các phương pháp hiện có khác và cũng đã phát triển một mô hình mới có thể phát hiện các URL không xác định với độ chính xác cao hơn nhiều so với các mô hình đã được đề xuất trong một số bài nghiên cứu khoa học. Ý nghĩa thực tế nằm ở việc phương pháp đề xuất cải thiện độ chính xác lên đến 0,98 (cao hơn 0,06 so với các phương pháp hiện có).

## V. Kết luận

Một cuộc đánh giá phân tích về tài liệu hiện đại về đề tài nghiên cứu đã cho thấy rằng học máy đã đạt được độ chính xác đáng kể trong việc phát hiện các URL lừa đảo, với một số kết quả tệ nhất vẫn đạt 0,75 và một số kết quả tốt nhất ở mức xấp xỉ 0,97. Có một loạt các đặc trưng độc đáo được trích xuất từ URL để huấn luyện mô hình học máy, nhưng một số đặc trưng ổn định bao gồm chiều dài của URL, miền cấp độ cao nhất, tham số URL, địa chỉ IP và số lượng mỗi ký tự đặc trưng. Cũng có nhiều thuật toán khác nhau được sử dụng, nhưng nổi bật phải kể đến như LR, SVM, RF và DT. Nghiên cứu đã khám phá rằng LSVC và đa thức NB là những phương pháp có triển vọng nhất cùng với phương pháp được phát triển.

Thông qua các thực nghiệm, một mô hình học máy mới với phương pháp được đề xuất đã được phát triển với độ chính xác là 0,98 (cao hơn 0,06 so với các phương pháp hiện tại). Các nghiên cứu thực nghiệm đã được tiến hành để đánh giá sự cải thiện về độ chính xác trong phát hiện. Phân tích kết quả cho thấy rằng phương pháp cho phép chúng ta tự tin phát hiện cả các URL lừa đảo đã biết và chưa biết. Sự so sánh với kết quả thu được và mức độ kỹ thuật hiện tại đã cho thấy rằng giải pháp đề xuất có khả năng đạt được độ chính xác cao hơn. Sự độc đáo trong lĩnh vực khoa học của phương pháp đề xuất nằm ở việc lần đầu tiên sử dụng phương pháp phân loại kết hợp để phát hiện các URL lừa đảo (sử dụng các phương pháp học máy LSVC và NB với cơ chế bỏ phiếu). Ý nghĩa thực tế của phương pháp đề xuất nằm ở việc phương pháp này cải thiện độ chính xác lên đến 0,98 (cao hơn 0,06 so với các phương pháp hiện tại).

Trong các nghiên cứu tương lai, nhóm tác giả sẽ thử nghiệm các thuật toán và tập hợp đặc trưng khác trên một tập dữ liệu lớn hơn và có thể thử nghiệm việc kết hợp một mô hình học máy tập trung vào các khía cạnh khác của các trang web (ví dụ, các tệp HTML, biểu tượng và nội dung của trang web) để tạo ra một mô hình thậm chí còn mạnh mẽ hơn.

TÀI LIỆU THAM KHẢO


[1]. The State of Phishing [Digital resource].– URL://www.slashnext.com/wp-content/uploads/2022/10/SlashNext-The-State-of-Phishing-2022.pdf (access date: 15.12.2022).

[2]. The Top 6 Industries At Risk For Cyber Attacks [Digital resource].– URL: https://redteamsecurity.com/blog/the-top-6-industries-at-risk-for-cyber-attacks (access date: 15.12.2022).

[3]. Phishing activity trends report 4th Quarter 2022 [Digital resource].– URL: https://docs.apwg.org/reports/apwg_trends_report_q4_2022.pdf?_gl=1*yoi676*_ga*NzA3MTgwODg0LjE2OTc3MjQ2NzU.*_ga_55RF0RHXSR*MTY5NzcyNDY3NS4xLjAuMTY5NzcyNDY3NS4wLjAuMA (access date: 15.12.2022).

[4]. What is URL phishing [Digital resource].– URL: https://surfshark.com/blog/what-is-url-phishing (access date: 15.12.2022).

[5]. Charan A. N. S., Chen Y. H., Chen J. L. Phishing Websites Detection using Machine Learning with URL Analysis /2022 IEEE World Conference on Applied Intelligence and Computing (AIC).– IEEE, 2022.– P. 808-812.

[6]. Uddin M. M. et al. A Comparative Analysis of Machine Learning-Based Website Phishing Detection Using URL Information //2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI).– IEEE, 2022.– P. 220-224.

[7]. Sindhu S. et al. Phishing detection using random forest, SVM and neural network with backpropagation //2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE).– IEEE, 2020, – P. 391-394.

[8]. Athulya A. A., Praveen K. Towards the detection of phishing attacks //2020 4th international conference on trends in electronics and informatics (ICOEI)(48184).– IEEE, 2020, – P. 337-343.

[9]. Bouijij H., Berqia A. Machine learning algorithms evaluation for phishing URL classification //2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT).– IEEE, 2021.– P. 01-05.

[10]. Amen K., Zohdy M., Mahmoud M. Machine Learning for Multiple Stage Phishing URL Prediction //2021 International Conference on Computational Science and Computational Intelligence (CSCI).– IEEE, 2021.– P. 794-800.

[11]. Dr U. S., Patil A., Mohana M. Malicious URL Detection and Classification Analysis using Machine Learning Models //2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT).– IEEE, 2023.– P. 470-476.

[12]. Phising and Benign Websites – URL: https://www.kaggle.com/datasets/peyamowar/phishing-and-benign- website (access date: 15.12.2022.).

[13]. Phising Site URL [Digital resource].– URL: https://www.kaggle.com/datasets/taruntiwarihp/phishing-site-URL (access date 15.12.2022).

[14]. Urlib.parse library [Digital resource].– URL: https://docs.python.org/3/library/urllib.parse.html (access date: 15.12.2022).

[15]. Linear support vector classifier [Digital resource].–URL: https://scikitlearn.org/stable/modules/generated/sklearn.svm.LinearSVC.html (access date: 15.12.2022).

[16]. Logistic regression [Digital resource].– URL: https://scikitlearn.org/stable/modules/generated/sklearn.linear model.LogisticRegression.html (access date: 15.12.2022).

[17]. Multinomial naive Bayes [Digital resource].– URL: https://scikitlearn.org/stable/modules/generated/sklearn.naive bayes.MultinomialNB.html (access date: 15.12.2022).

[18]. Decision tree classifier [Digital resource].– URL: https://scikitlearn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html (access date: 15.12.2022).

[19]. Random forest classifier [Digital resource].– URL: https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (access date: 15.12.2022).

[20]. Voting classifier [Digital resource].– URL: https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html (access date: 15.12.2022).

[21]. Thang, N. M., & Luong, T. T. (2022). Algorithm for detecting attacks on Web applications based on machine learning methods and attributes queries. Journal of Science and Technology on Information Security, 2(14), 26-34.

SƠ LƯỢC VỀ TÁC GIẢ

**Nguyễn Mạnh Thắng**
Đơn vị công tác: Học viện Kỹ thuật mật mã.
Email: chieumatxcova@gmail.com
2005-2007: Học kỹ sư Điều khiển tự động tại Học viện Kỹ thuật quân sự; 2007-2013: Tốt nghiệp ngành Toán ứng dụng và Tin học tại Đại học Sư phạm bang Lipetsk, Liên bang Nga; 2017-2020: Nghiên cứu sinh chuyên ngành An toàn thông tin tại Học viện FSO, Liên bang Nga và nhận bằng Tiến sĩ năm 2020.
Hướng nghiên cứu: Mạng máy tính; an ninh mạng; học máy; khai thác dữ liệu.

**Lê Quang Anh**
Đơn vị công tác: Tổng công ty giải pháp doanh nghiệp Viettel.
Email: lequanganh97@gmail.com
Quá trình đào tạo:
2015-2016: Học kỹ sư An toàn thông tin tại Học viện Kỹ thuật mật mã; 2017-2021: Nhận bằng cử nhân An toàn thông tin tại Đại học ITMO, Liên bang Nga; 2022-2023: Nhận bằng Thạc sĩ An toàn thông tin tại Đại học ITMO - St. Petersburg, Liên bang Nga.
Hướng nghiên cứu: Học máy; giám sát an toàn thông tin; an ninh mạng.

**Hứa Song Toàn**
Đơn vị công tác: Sở thông tin và Truyền thông Hải Phòng.
Email: huasongtoan@haiphong.gov.vn
Quá trình đào tạo: Nhận bằng cử nhân Công nghệ thông tin tại Học viện An ninh nhân dân năm 2017; Thạc sĩ An toàn thông tin tại Học viện Kỹ thuật mật mã năm 2019.
Hướng nghiên cứu: Học máy; an toàn thông tin.

**Nguyễn Quốc Trung**
Đơn vị công tác: Trung tâm phát triển Ngân hàng số BIDV.
Email: trung35118554@gmail.com
Quá trình đào tạo: Tốt nghiệp Đại học tại Trường Điện - Điện tử, Đại học Bách khoa Hà Nội năm 2023.
Hướng nghiên cứu: An toàn phần mềm.