|  |  |
| --- | --- |
| Vũ Minh Đức | : 0950080093 |
| Phạm Thị Thảo Dương | : 0950080095 |
| Phạm Minh Tuấn | : 0950080115 |

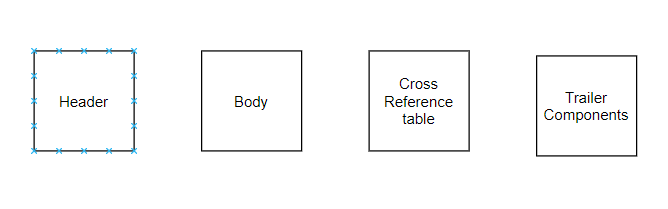
**PHÁT HIỆN PHẦN MỀM ĐỘC HẠI PDF DỰA TRÊN CÂY QUYẾT ĐỊNH CÓ THỂ TỐI ƯU HÓA**

**1. Giới thiệu**

Một đoạn mã độc hại có khả năng làm hỏng máy tính hoặc mạng được gọi là phần mềm độc hại. Khi các công nghệ phát hiện phần mềm độc hại dựa trên chữ ký thông thường trở nên vô dụng và không thể thực hiện được, những năm gần đây đã chứng kiến ​​sự gia tăng đáng kể của phần mềm độc hại. Các nhà phát triển phần mềm độc hại và tội phạm mạng đã áp dụng các kỹ thuật làm xáo trộn mã, làm giảm hiệu quả của các cơ chế phòng thủ chống lại phần mềm độc hại.

Phân loại và nhận dạng phần mềm độc hại vẫn là một thách thức trong thập kỷ này. Điều này phần lớn là do phần mềm độc hại tiên tiến phức tạp hơn và có khả năng tiên tiến trong việc ẩn giấu hoặc thay đổi mã hoặc hành vi của nó để hoạt động thông minh hơn. Kết quả là các phương pháp phát hiện và phân loại lỗi thời ngày nay ít hữu ích hơn. Do đó, trọng tâm đã chuyển sang học máy để nhận dạng và phân loại phần mềm độc hại tốt hơn .

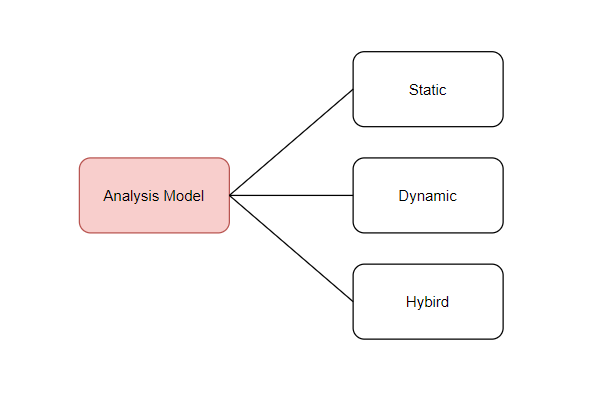
Phần mềm PDF độc hại là một trong những phương pháp hack phổ biến . Nghiên cứu pháp y bị cản trở do khó tách các tệp PDF có hại khỏi các tệp PDF lớn. Học máy đã phát triển đến mức giờ đây nó có thể được sử dụng để phát hiện các tài liệu PDF độc hại nhằm hỗ trợ các nhà điều tra pháp y hoặc bảo vệ hệ thống khỏi bị tấn công . Tuy nhiên, các kỹ thuật đối nghịch đã được phát triển để chống lại các bộ phân loại tài liệu độc hại. Các ví dụ thù địch dựa trên thao tác chính xác đã được chế tạo cẩn thận có thể bị phân loại sai. Điều này gây nguy hiểm cho nhiều máy dò dựa trên máy học . Đối với các cuộc tấn công cụ thể, nhiều phương pháp phân tích hoặc phát hiện khác nhau đã được cung cấp. Mối đe dọa từ các cuộc tấn công đối nghịch vẫn chưa được khắc phục hoàn toàn. Hình 1 mô tả các thành phần tiêu đề, nội dung, bảng tham chiếu chéo (xref) và đoạn giới thiệu của tài liệu PDF [9].



Hình 1. Structure of a PDF file

Phiên bản định dạng trình thông dịch sẽ được sử dụng được chỉ định trong tiêu đề. Phần thân của tệp PDF xác định nội dung của nó và bao gồm các khối văn bản, phông chữ, hình ảnh và siêu dữ liệu dành riêng cho tệp. Nội dung của tài liệu được chứa trong một nhóm các phần tử PDF. Những thứ này có thể thuộc một trong bốn loại: Booleans, chuỗi, luồng và số .

Công cụ phân tích hoặc phân tích có thể sử dụng các kỹ thuật phân tích phần mềm độc hại tĩnh, động hoặc kết hợp (Hình 2) . Kỹ thuật phân tích tĩnh kiểm tra mẫu mà không chạy mã và dựa vào các thuộc tính tệp, chẳng hạn như cấu trúc mã. Trong các phương pháp phân tích, tính năng động thực thi mã để quan sát hành vi của nó, chẳng hạn như các hoạt động của mạng chương trình .



Hình 2. Structure of a PDF file

Việc áp dụng các kỹ thuật ẩn nấp và che giấu nâng cao để che dấu hành vi nguy hiểm trong thời gian chạy khiến cho việc phân tích tĩnh dễ bị tổn thương. Chỉ thực hiện phân tích tĩnh trong môi trường bảo mật hiện tại là chưa đủ. Bất kỳ kẻ tấn công nào nghiêm túc về chiến dịch của họ sẽ làm xáo trộn và mã hóa mã của họ, thường không thể phát hiện được bằng phân tích tĩnh.

Mặt khác, các phương pháp động có khả năng chống lại sự xáo trộn mã tốt hơn, khiến chúng hiệu quả hơn trước các loại virus phức tạp. Để tránh bị tổn hại, các kỹ thuật động phải chạy vi-rút trong môi trường an toàn, có hộp cát. Cho dù nó có tin rằng phần mềm độc hại đang chạy trong hộp cát hay không, kẻ thù có thể thay đổi hành vi của vi-rút để cản trở quá trình phân tích phần mềm độc hại. Trong khi phân tích tĩnh thường nhanh chóng thì phân tích động thường chậm và khó. Phân tích lai đề cập đến sự kết hợp của hai phương pháp. Cách này chống lại phần mềm độc hại phức tạp hiệu quả hơn một trong hai cách, nhưng cũng tốn nhiều thời gian hơn và yêu cầu quá trình phân tích phức tạp hơn [16].

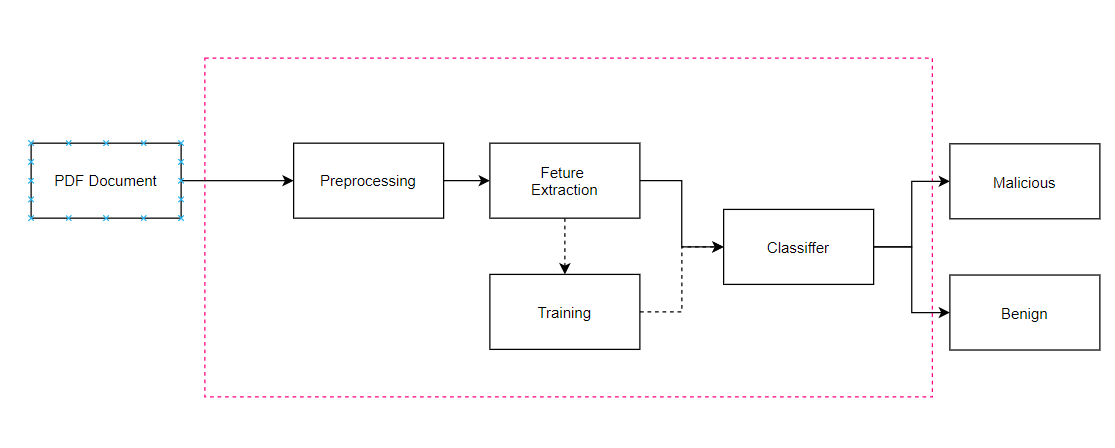
Trong bài báo này, chúng tôi trình bày một hệ thống phát hiện mới có thể phân tích tài liệu PDF để xác định tệp PDF lành tính khỏi tệp PDF chứa phần mềm độc hại. Hệ thống được đề xuất sử dụng cây quyết định AdaBoost với các siêu tham số tối ưu , được đào tạo và đánh giá trên bộ dữ liệu tổng hợp hiện đại, tức là. Lẩn tránh-PDFMal2022. Đánh giá điều tra cho thấy hệ thống phát hiện PDF nhẹ và chính xác, đạt độ chính xác dự đoán 98,84% với khoảng thời gian dự đoán ngắn là 2,174 µSec. Cuối cùng, mô hình đề xuất vượt trội hơn các mô hình tiên tiến khác trong cùng khu vực nghiên cứu. Do đó, hệ thống được đề xuất có thể phát hiện phần mềm độc hại PDF một cách hiệu quả với hiệu suất phát hiện cao và chi phí phát hiện thấp.

Phần còn lại của bài viết này được tổ chức như sau: Phần 2 trình bày tổng quan một cách có hệ thống và toàn diện về các bài viết liên quan gần đây trong cùng lĩnh vực nghiên cứu. Phần 3 cung cấp kiến trúc mô hình hóa cho hệ thống phát hiện phần mềm độc hại PDF. Phần 4 trình bày và thảo luận về kết quả thực hiện và đánh giá thực nghiệm. Cuối cùng, Phần 5 đưa ra các nhận xét kết luận.

**2. Tổng quan lý thuyết**

Các phương pháp học sâu, đặc biệt là Mạng lưới thần kinh sâu (DNN), đã trở nên phổ biến trong các lĩnh vực học thuật và công nghiệp . Ứng dụng của họ có thể được tìm thấy trong nhiều lĩnh vực khác nhau, bao gồm cả phân tích phần mềm độc hại. Đối với các tác vụ đòi hỏi nhiều tài nguyên như nhận dạng giọng nói, xử lý ngôn ngữ tự nhiên và nhận dạng hình ảnh, DNN hoạt động tốt. Tuy nhiên, người ta đã chứng minh rằng việc phân loại các hệ thống dựa trên máy học rất dễ bị ảnh hưởng bởi các cài đặt thù địch với các nỗ lực trốn tránh tiên tiến .

Để xác định phần mềm độc hại, học máy có giám sát thường được sử dụng. Một số trình phát hiện sử dụng công nghệ này đã được tạo riêng cho tệp PDF trong mười năm qua. Việc chọn xem bất kỳ tệp PDF không xác định nào sẽ được phân loại là có hại hay lành tính là mục tiêu chính của trình phát hiện máy học để nhận dạng tài liệu độc hại. Các hệ thống như vậy có thể hoạt động bằng cách kiểm tra dữ liệu được lấy từ nội dung hoặc cấu trúc của tài liệu. Luồng quy trình chung của chúng được mô tả trong Hình 3, bao gồm ba phần chính : tiền xử lý phân tích các tệp PDF và cung cấp quyền truy cập vào dữ liệu cần thiết để phát hiện.



Hình 3. Structure of a PDF file

Thông tin được chuyển thành vectơ chuẩn hóa như một phần của quá trình trích xuất đặc trưng. Để đảm bảo dự đoán chính xác, bộ phân loại chọn thuật toán học tốt nhất để huấn luyện và sửa đổi nhằm thu được các tham số được cải thiện. Bởi vì chất lượng của các đặc trưng có thể có tác động rõ rệt đến hiệu suất dự đoán nên việc trích xuất đặc trưng là rất quan trọng .

Một phương pháp tích hợp để phát hiện phần mềm độc hại sử dụng các tính năng tĩnh và động đã được giới thiệu bởi . Việc kết hợp các đặc điểm tĩnh và động đã cải thiện độ chính xác nhận dạng so với việc sử dụng các phương pháp tĩnh hoặc động riêng biệt. Theo kết quả nghiên cứu, phương pháp học máy vectơ hỗ trợ (SVM) là phương pháp hiệu quả nhất trong việc phân loại dữ liệu. Tuy nhiên, ngoài những cải thiện về tỷ lệ dương tính giả (FP) và âm tính giả (FN), rừng ngẫu nhiên (RF) cũng cải thiện độ chính xác . Kết quả phân loại cho thấy phân tích động vượt trội hơn so với các phương pháp tĩnh dựa trên mã. So với phương pháp tĩnh, phương pháp động có độ chính xác cao hơn. Chiến lược tích hợp cải thiện độ chính xác phát hiện, phù hợp với mục tiêu nghiên cứu.

Các tác giả đã đề xuất một hệ thống phát hiện phần mềm độc hại nhúng hoàn toàn mới dựa trên các kỹ thuật phát hiện bất thường về thống kê. Đây là phương pháp phát hiện phần mềm độc hại dựa trên sự bất thường đầu tiên để xác định vị trí lây nhiễm trong tệp bị nhiễm. Máy dò n-gram Markov được đề xuất vượt trội hơn các máy dò hiện có về tốc độ phát hiện. Ngoài ra, khi được sử dụng với phần mềm chống vi-rút thương mại sẵn có (COTS-AV) hiện tại, trình phát hiện được đề xuất có thể mang lại tỷ lệ dương tính giả rất thấp do khả năng xác định phần mềm độc hại được nhúng.

Một kỹ thuật không dựa trên chữ ký để kiểm tra nội dung tệp ở mức byte đã được đề xuất bởi [25]. Phương pháp như vậy mang lại khả năng chống lại các chiến lược che giấu thông thường, đặc biệt là các chiến lược sử dụng phần mềm độc hại được đóng gói lại để ẩn chữ ký. Nghiên cứu này đã phát hiện ra rằng các tệp bị nhiễm và tệp lành tính khác nhau về cơ bản, thậm chí ở cấp độ byte. Mười ba tính năng lý thuyết thông tin và thống kê độc đáo được tính toán trên 1, 2, 3 và 4 g của mỗi khối tệp được sử dụng trong phương pháp đề xuất, có bộ tính năng phong phú.

Trong [26], các tác giả giới thiệu một khuôn khổ để phát hiện mạnh mẽ các tài liệu gian lận dựa trên máy học. Phương pháp được đề xuất dựa trên các thành phần được lấy từ cấu trúc và siêu dữ liệu của tài liệu. Nghiên cứu chứng minh sự phù hợp của một số thuộc tính tài liệu nhất định để nhận dạng phần mềm độc hại và khả năng phục hồi của các tính năng này trước các chủng vi-rút mới bằng cách sử dụng bộ dữ liệu trong thế giới thực. Giai đoạn phân tích cho thấy rằng Rừng ngẫu nhiên của bộ phân loại tổng hợp, chọn ngẫu nhiên các đặc điểm cho từng cây phân loại riêng biệt, tạo ra tỷ lệ phát hiện cao nhất.

Trong cuộc điều tra này, hai nguồn dữ liệu chính đã được sử dụng. Đầu tiên là bộ sưu tập dữ liệu Contagio được sử dụng rộng rãi [27], nhằm mục đích kiểm tra và nghiên cứu chữ ký. Nguồn dữ liệu này được chọn vì nó có nhiều tài liệu được phân loại là độc hại và lành tính, bao gồm một tỷ lệ khá lớn từ các cuộc tấn công có chủ đích. Nguồn này cung cấp một số bộ ocument. Bộ sưu tập thứ hai đến từ việc giám sát mạng của một khuôn viên trường đại học lớn. Các tệp này được lấy từ lưu lượng Giao thức truyền thư đơn giản (SMTP) và Giao thức truyền siêu văn bản (HTTP).

Các tác giả trong [28] đã nghĩ ra một phương pháp để xác định một tập hợp các tính năng được trích xuất bằng các công cụ hiện có và rút ra một bộ tính năng mới từ việc cải thiện khả năng phát hiện maldoc PDF và kéo dài tuổi thọ của các công nghệ phân tích và phát hiện hiện có. Các tính năng dẫn xuất được đánh giá bằng hàm bao bọc sử dụng ba thuật toán học có giám sát cơ bản [29] (Rừng ngẫu nhiên, Cây quyết định C5.0 và Máy vectơ hỗ trợ hai lớp) và mạng lưới thần kinh sâu chuyển tiếp nguồn cấp dữ liệu để xác định tầm quan trọng của tính năng là. Cuối cùng, một bộ phân loại mới được xây dựng bằng cách sử dụng các tính năng có tầm quan trọng cao nhất, cải thiện đáng kể hiệu suất phân loại với thời gian đào tạo ít hơn. Kết quả đã được xác nhận bằng cách sử dụng bộ dữ liệu khá lớn từ VirusTotal.

Các tác giả của [30] trình bày một kỹ thuật hoàn toàn mới để xác định các khó khăn về dữ liệu của một bộ phân loại tổng hợp. Khi có đủ số phiếu bầu của bộ phân loại riêng lẻ xung đột trong quá trình phát hiện, dự đoán của bộ phân loại tổng thể được chứng minh là không chính xác. Không cần thêm sự thật cơ bản bên ngoài, kỹ thuật được đề xuất, phân tích thỏa thuận chung của bộ phân loại tổng thể, cho phép phát hiện ra nhiều kiểu trốn tránh bộ phân loại.

Bằng cách sử dụng PDFrate, một trình phát hiện phần mềm độc hại PDF, các tác giả đã thử nghiệm chiến lược được đề xuất và chứng minh rằng phần lớn các dự đoán có thể được tạo ra với sự đồng thuận phân loại tổng thể cao bằng cách sử dụng dữ liệu từ toàn bộ mạng và phương pháp của chúng tôi. Chín tình huống bắt chước có chủ đích từ hai đợt nghiên cứu gần đây nằm trong số các nỗ lực trốn tránh của bộ phân loại thường mang lại kết quả không rõ ràng, cho thấy rằng bộ phân loại không thể đưa ra dự báo đáng tin cậy cho những dữ liệu này [31]. Để chứng minh khả năng ứng dụng rộng rãi của phương pháp này, tác giả đã thử nghiệm nó với trình phát hiện phần mềm độc hại Drebin Android, trong đó hầu hết các cuộc tấn công đặc biệt đều được dự đoán chính xác là không chắc chắn. Phương pháp đề xuất có thể được áp dụng rộng rãi hơn để giảm hiệu lực của các cuộc tấn công vào các máy vectơ hỗ trợ được thực hiện thông qua ước tính độ dốc và mật độ hạt nhân. Yếu tố quan trọng nhất để cho phép phát hiện sự trốn tránh dựa trên tính đa dạng của bộ phân loại tổng hợp là đóng gói tính năng.

Các tác giả trong [32] giới thiệu Lux0R, đôi khi được gọi là “Tham chiếu phân biệt Lux 0n”, một phương pháp độc đáo và di động để xác định mã JavaScript gian lận. Cách tiếp cận được đề xuất dựa trên việc mô tả đặc điểm mã JavaScript thông qua các tham chiếu đến API của nó, bao gồm các hàm, hằng, đối tượng, phương thức, từ khóa và thuộc tính được Giao diện lập trình ứng dụng JavaScript (API) nhận dạng nguyên bản. Phương pháp được đề xuất sử dụng công nghệ máy học để xác định một tập hợp con các tham chiếu API biểu thị mã nguy hiểm, sau đó sử dụng các tham chiếu đó để xác định phần mềm độc hại JavaScript. Người ta nói rằng cơ chế lựa chọn là “an toàn theo thiết kế” chống lại việc trốn tránh bằng cách sử dụng các cuộc tấn công bắt chước. Xác định mã JavaScript nguy hiểm trong tài liệu PDF là miền ứng dụng liên quan được tác giả tập trung vào trong công trình này.

Kỹ thuật này có thể đạt được độ chính xác phát hiện phần mềm độc hại vượt trội ngay cả trên các mẫu khai thác các lỗ hổng chưa được biết đến trước đây, tức là không có trường hợp nào trong dữ liệu đào tạo. Cuối cùng, đánh giá thử nghiệm về khả năng chống lại các cuộc tấn công bắt chước của Lux0R dựa trên việc tăng cường tính năng được thực hiện.

Công trình này [33] trình bày một cách tiếp cận mới kết hợp mô-đun trích xuất tính năng liên quan chặt chẽ đến cấu trúc của tệp PDF với trình phân loại mạnh mẽ. Kỹ thuật này đã cho thấy hiệu quả hơn hầu hết các chương trình chống vi-rút thương mại và các công cụ nghiên cứu tiên tiến khác để phát hiện các tệp PDF nguy hiểm. Hơn nữa, do khả năng thích ứng của nó, nó có thể được sử dụng để nâng cao hiệu quả của phần mềm chống vi-rút đã được cài đặt hoặc dưới dạng chương trình độc lập.

Nó hoạt động tốt hơn đáng kể so với Wepawet, một công cụ mạnh mẽ được tạo ra bởi các học giả. Wepawet đã được tạo ra để phát hiện các mối đe dọa khác nhau, bao gồm các tệp PDF độc hại, nhưng chương trình được phát triển tập trung vào việc phát hiện các cuộc tấn công PDF.

Nó có thể được tăng cường hơn nữa bằng cách đánh giá khả năng phục hồi của hệ thống được đề xuất trước các lỗ hổng mới và tăng cường quy trình phân tích cú pháp. Công cụ được đề xuất cũng có thể là một thành phần của hệ thống đa phân loại, trong đó mỗi phân loại tập trung vào việc xác định các mối nguy hiểm cụ thể. Làm cho hệ thống bảo mật của chúng tôi mạnh mẽ hơn trước nhiều mối nguy hiểm hơn và cung cấp cho chúng khả năng dự đoán các mối đe dọa mới là một thách thức cho tương lai khi các chiến thuật của kẻ tấn công ngày càng phát triển

Các tác giả trong [34] đã phát hiện ra các lỗ hổng trong các trình trích xuất tính năng hiện có cho các tệp PDF bằng cách xem xét chúng và kiểm tra cách triển khai mẫu độc hại. Sau đó, các tác giả đã tạo ra một trình trích xuất tính năng mạnh mẽ có tên là FEPDF (trình trích xuất tính năng-PDF), có thể trích xuất các tính năng mà các trình trích xuất tính năng thông thường có thể bỏ qua và nắm bắt thông tin thực tế về các thành phần trong tệp PDF. Các tác giả đã tạo ra nhiều tệp PDF độc hại hoàn toàn mới làm mẫu để kiểm tra các công cụ chống vi-rút và trình trích xuất tính năng hiện tại. Các phát hiện chứng minh rằng một số công cụ chống vi-rút hiện tại không thể nhận dạng các tệp PDF có hại mới, nhưng FEPDF có thể trích xuất các yếu tố quan trọng để nâng cao khả năng phát hiện tệp PDF nguy hiểm.

Nghiên cứu này [35] chứng minh khả năng chống chịu yếu hơn của thuật toán phân loại K-gần nhất (KNN) điển hình trong môi trường đối nghịch bằng cách sử dụng phương pháp tấn công giảm dần độ dốc để thay đổi các mẫu độc hại trong bộ thử nghiệm nhằm tránh bị bộ phân loại phát hiện. Các tác giả cung cấp một phương pháp trong đó các mẫu đối thủ đã tạo sẽ được thêm vào tập huấn luyện, sau đó sử dụng tập huấn luyện để tạo bộ phân loại KNN mới và kiểm tra độ mạnh mẽ của chúng trước các cường độ tấn công khác nhau.

Cuối cùng, các thử nghiệm chứng minh rằng độ mạnh của bộ phân loại KNN có thể tăng lên đáng kể mà không ảnh hưởng đến hiệu suất tổng quát hóa của bộ phân loại KNN bằng cách đưa vào các mẫu đối nghịch được tạo ra bởi các cuộc tấn công giảm dần độ dốc trên tập huấn luyện.

Một cách tiếp cận dựa trên khai thác dữ liệu mới được cung cấp bởi các tác giả của [36], được giới thiệu để xác định các tệp PDF gian lận. Thuật toán đề xuất có hai giai đoạn: lựa chọn và phân loại đặc trưng. Bước lựa chọn tính năng được sử dụng để chọn số lượng tính năng lý tưởng được trích xuất từ tệp PDF nhằm đạt được tỷ lệ phát hiện cao và tỷ lệ dương tính giả thấp với chi phí tính toán ít. Theo dữ liệu thử nghiệm, thuật toán được đề xuất có thể đạt tỷ lệ phát hiện 99,77%, độ chính xác 99,84% và tỷ lệ dương tính giả 0,05%.

Nó có thể hoạt động tốt hơn bằng cách so sánh thuật toán được đề xuất với các chương trình chống vi-rút từ CalamAV, TrendMicro, MacAfee và Symantec. Thuật toán được đề xuất dựa trên các kỹ thuật khai thác dữ liệu, mang lại lợi thế cho phần mềm chống vi-rút trong việc phát hiện các tệp PDF có hại chưa từng thấy trước đây. Do đó, phương pháp được đề xuất có thể xác định tốt hơn các mối đe dọa liên tục nâng cao (APT).

Bằng cách sử dụng phương pháp giảm dần độ dốc (GD), SVM đơn giản được các tác giả sử dụng trong [37] đã dễ dàng bị chúng tôi đánh lừa. Các tác giả cũng nghĩ ra các biện pháp phòng vệ chống lại cuộc tấn công này bằng cách đặt ngưỡng cho từng đặc điểm được xem xét.

Điều này cho phép phương pháp được đề xuất thực tế có thể ngăn chặn tất cả các cuộc tấn công giảm dần độ dốc. Tiếp theo, ít tính năng hơn đã được chọn để có thể loại bỏ các tính năng được sử dụng trong cuộc tấn công giảm dần độ dốc. Điều này làm giảm khả năng tồn tại của cuộc tấn công hơn nữa, làm giảm độ chính xác của SVM [38]. Các tác giả cũng đề xuất sử dụng phương pháp học đối nghịch để huấn luyện SVM bằng cách sử dụng các tệp PDF giả mạo giảm dần độ dốc và lặp lại quy trình để giảm. Điều này cho phép phương pháp được đề xuất thực tế ngăn chặn tất cả các cuộc tấn công giảm dần độ dốc. Tiếp theo, ít tính năng hơn đã được chọn để có thể loại bỏ các tính năng được sử dụng trong cuộc tấn công giảm dần độ dốc. Điều này làm giảm khả năng tồn tại của cuộc tấn công hơn nữa, làm giảm độ chính xác của SVM [38]. Các tác giả cũng đề xuất sử dụng phương pháp học đối nghịch để huấn luyện SVM bằng cách sử dụng các tệp PDF giả mạo giảm dần độ dốc và lặp lại quy trình để giảm khả năng cuộc tấn công giảm dần độ dốc sẽ thành công. Chỉ sau ba chu kỳ, SVM đã thể hiện khả năng chống lại các cuộc tấn công bằng kỹ thuật giảm độ dốc.

Các tác giả trong [39] đưa ra những phân tích chuyên sâu về nội dung và cấu trúc JavaScript của tệp PDF. Họ đã tạo ra một bộ tính năng phong phú trong JavaScript bao gồm các tính năng nội dung như tên đối tượng, từ khóa và chuỗi có thể đọc được cũng như các tính năng về cấu trúc và siêu dữ liệu như kích thước tệp, phiên bản, phương pháp mã hóa và từ khóa. Việc tạo ra các ví dụ thù địch khi các tính năng rất đa dạng là một thách thức vì thuật toán học máy có khả năng chống lại những thay đổi nhỏ. Để giảm nguy cơ bị tấn công bất lợi, các nhà phân tích tạo ra các mô hình phát hiện sử dụng các loại mô hình hộp đen có đặc tính cấu trúc và nội dung. Các tác giả đã tạo ra cuộc tấn công đối nghịch để xác minh mô hình được đề xuất. Ngoài ra, chúng tôi thu thập các tài liệu lành mạnh với nhiều mã JavaScript khác nhau để làm nền tảng cho các mẫu thù địch.

Các tệp PDF được sử dụng trong nghiên cứu này bao gồm 9000 tệp tài liệu lành tính và 11.097 tệp tài liệu độc hại được thu thập bởi phần mềm độc hại Contagio từ tháng 11 năm 2009 đến tháng 6 năm 2018 [39]. Các mẫu phần mềm độc hại được cung cấp thông qua trang web chứa phần mềm độc hại Contagio. Từ trang web, các nhà nghiên cứu có thể lấy mẫu phần mềm độc hại. Các mẫu bao gồm một lượng lớn thời gian. Các tác giả đã thu thập 115 tệp sạch với các tệp JavaScript riêng biệt để phát triển một cuộc tấn công đối nghịch nhằm xác thực. Các tác giả, ngoại trừ các tệp bị mã hóa, đã cấy thành công phần mềm độc hại vào 101 tệp sạch.

Về phương pháp học máy, các tác giả phát hiện ra rằng trong khi hầu hết các thuật toán học máy thông thường hoạt động hiệu quả trong việc phát hiện phần mềm độc hại, chúng hoạt động kém hơn đối với các mẫu đối nghịch, ngoại trừ thuật toán rừng ngẫu nhiên. Do khả năng chuyển đổi này, thuật toán rừng ngẫu nhiên có thể hoạt động tốt. Về phương pháp học máy, các tác giả phát hiện ra rằng trong khi hầu hết các thuật toán học máy thông thường hoạt động hiệu quả trong việc phát hiện phần mềm độc hại, chúng hoạt động kém hơn đối với các mẫu đối nghịch, ngoại trừ thuật toán rừng ngẫu nhiên. Do khả năng chuyển đổi này, thuật toán rừng ngẫu nhiên có thể hoạt động tốt.

Tác giả trong [40] đã đưa ra một cách bài bản một số khái niệm hướng dẫn lựa chọn đặc điểm nhằm giảm khả năng thoát hiểm mà vẫn giữ được độ chính xác cao. Những hướng dẫn này được tuân theo để trích xuất các tính năng và huấn luyện bộ phân loại hai giai đoạn. Các phát hiện thử nghiệm chứng minh rằng mô hình của chúng tôi hoạt động tuyệt vời về độ chính xác, khả năng khái quát hóa và độ tin cậy. Nó cũng có thể phân biệt giữa lỗ hổng được sử dụng trong các tệp độc hại.

Tác giả đã giới thiệu chiến lược xác định phần mềm đã tạo tệp PDF [41] dựa trên kiểu mã hóa: các mẫu cụ thể mà các nhà sản xuất PDF cụ thể chỉ tạo ra. Ngoài ra, họ còn xem xét hoạt động mã hóa của 900 tệp PDF được tạo bởi 11 nhà sản xuất PDF riêng biệt trên ba hệ điều hành khác nhau. Một bộ gồm 192 quy tắc có thể được sử dụng để xác định 11 nhà sản xuất PDF đã được các tác giả tiếp thu. Chúng tôi đã sử dụng 508.836 tệp PDF từ các nguồn in sẵn khoa học để kiểm tra phương pháp nhận dạng của mình. Công cụ được sử dụng có tỷ lệ chính xác 00% để xác định các nhà sản xuất cụ thể. Nhìn chung, nó vẫn phát hiện tốt (74%). Để hiểu cách hoạt động của các dịch vụ PDF trực tuyến và phát hiện sự không nhất quán, hãy sử dụng công cụ được cung cấp. Cuối cùng, Bảng 1 tóm tắt các nghiên cứu quan trọng có liên quan đã được xem xét.

Trong [42], các tác giả đã cung cấp một bản tóm tắt kỹ lưỡng về các phương pháp hiện đang được sử dụng để nhận dạng tài liệu độc hại. Các công cụ cơ bản thường được sử dụng trong các phương pháp phát hiện đã được đề cập. Các phương pháp khác nhau được phân loại dựa trên các đặc điểm đã chọn và phân tích tĩnh/động.

Một thước đo khoảng cách mới cho các đặc tính độ bền giới hạn đã được đề xuất trong [43]. Mô hình đề xuất duy trì độ chính xác 99,74% và tỷ lệ dương tính giả 0,56% trong khi đạt được độ chính xác mạnh mẽ (VRA) được xác minh lần lượt là 99,68% và 85,28% cho các thuộc tính chèn và xóa. Các phát hiện đã chứng minh rằng việc đào tạo các bộ phân loại bảo mật với các thuộc tính mạnh mẽ đã được xác minh là một hướng đầy hứa hẹn để nâng cao tiêu chuẩn cho những kẻ tấn công không bị kiềm chế.

Trong [44], tầm quan trọng của các mẫu nghi ngờ đã được xác nhận và mô hình phát hiện dựa trên học tập tích cực được sử dụng với các mẫu không chắc chắn đó. Một số lượng nhỏ các mẫu tập kiểm tra giàu thông tin, hay còn gọi là các mẫu nghi ngờ, được chọn để bổ sung cho tập huấn luyện trong mỗi giai đoạn đánh giá, dần dần nâng cao hiệu suất của bộ phân loại. So với các phương pháp đào tạo lại thông thường, các tác giả đã cắt giảm đáng kể thời gian đào tạo cần thiết. Nghiên cứu này sử dụng phương pháp học tập tích cực dựa trên sự thỏa thuận chung. Các tác giả đã sử dụng mô hình Hidost với khả năng học tập tích cực và phân tích thỏa thuận chung làm tiêu chí để chọn các mẫu nghi ngờ, cho phép chúng tôi sử dụng tập huấn luyện nhỏ hơn và cải thiện hiệu suất phân loại.

Các tác giả trong [45] đã đề xuất một mô hình mới gọi là OPEM. Trình phát hiện phần mềm độc hại kết hợp này kết hợp tần suất xuất hiện của mã hoạt động (được thu thập tĩnh) với các chi tiết về dấu vết thực thi có thể thực thi (được thu thập động). Các tác giả đã chứng minh rằng, khi được sử dụng độc lập, chiến lược kết hợp này sẽ cải thiện hiệu suất của cả hai phương pháp.

Các tác giả của [46] đã đề xuất một phương pháp dựa trên học tập mang tính cách mạng để phát hiện phần mềm độc hại PDF thông qua việc xử lý và hiển thị hình ảnh. Đồ họa thang độ xám được tạo từ các tệp PDF (sơ đồ byte và Markov). Sau đó, chất lượng hình ảnh khác nhau của hình ảnh được truy xuất bằng cách sử dụng các tính năng hình ảnh như bộ mô tả Keypoint và tính năng kết cấu.

Một kỹ thuật lẩn tránh độc đáo dựa trên mạng đối thủ tạo vectơ đặc trưng (fvGAN) đã được đề xuất trong [47] để nhắm mục tiêu vào bộ phân loại phần mềm độc hại dựa trên học tập. Công việc này tạo ra các vectơ đặc trưng đối nghịch trong không gian đặc trưng bằng cách sử dụng fvGAN, sau đó chuyển đổi các vectơ đặc trưng đó thành các mẫu phần mềm độc hại đối địch thực tế. Các phát hiện chứng minh rằng mô hình fvGAN có tỷ lệ trốn tránh cao trong một thời gian hạn chế. Chiến lược được đề xuất cũng trái ngược với hai thuật toán tấn công hiện đang được sử dụng là Bắt chước và GD-KDE. Các phát hiện cho thấy kỹ thuật được đề xuất hoạt động tốt hơn cả về tỷ lệ trốn tránh và chi phí thực hiện.

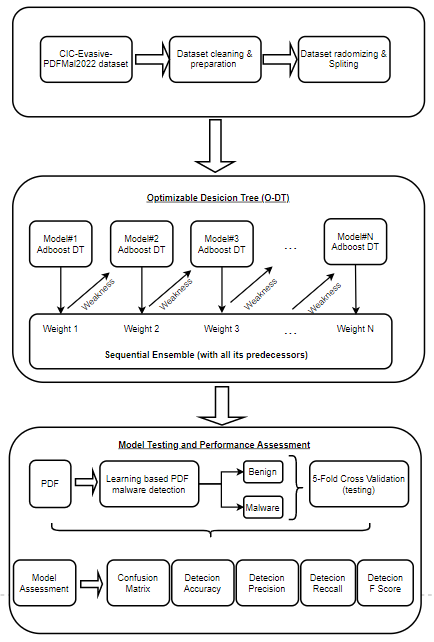
Trong [48], các tác giả đã nghiên cứu ba cuộc tấn công nổi tiếng—Bắt chước, Bắt chước+ và Bắt chước ngược—để so sánh mức độ chúng trốn tránh các bộ phân loại trong Hidost và Mimicus. Các phát hiện chứng minh rằng Bắt chước và Bắt chước+ thành công trong việc tránh các mô hình ở Mimicus nhưng không thành công ở Hidost, trong khi Bắt chước ngược thành công trong việc tránh cả hai mô hình Mimicus và Hidost.

Do việc sử dụng rộng rãi các kênh tấn công như tài liệu, phần mềm độc hại tiếp tục gây nguy hiểm cho an ninh mạng. Các vectơ lây nhiễm này che giấu mã độc hại khỏi người dùng nạn nhân, khiến việc lây nhiễm vào máy tính của họ dễ dàng hơn thông qua các phương pháp kỹ thuật xã hội [49].

3. Hệ thống phân loại đề xuất

Tệp định dạng Di động (PDF) là một trong những loại tệp được sử dụng phổ biến nhất trên toàn cầu. Giống như các loại tệp khác như tệp dot-com, PNG và Bitcoin, hacker có thể tìm cách sử dụng những tệp PDF thông thường vô hại này để tạo ra mối đe doạ bảo mật thông qua mã độc hại trong các tệp PDF. Điều này dẫn đến PDF Malware và đòi hỏi các kỹ thuật để nhận diện các tệp vô hại so với các tệp độc hại. Tài liệu PDF đã bị chiếm đóng và lợi dụng như một vector cho các hoạt động độc hại. Nhiều trình đọc và phần mềm PDF bị ảnh hưởng liên tục, như CVE-2018-14442, CVE-2017-10994 trong Foxit Reader và CVE-2018-8350 trong Thư viện PDF của Microsoft Windows [42]. Các hệ thống phát hiện thông minh gần đây được phát triển thông qua các kỹ thuật học máy/học sâu [54,55] và các kỹ thuật blockchain/tiền điện tử [56].

Trong phần này, chúng tôi giới thiệu hệ thống phát hiện đề xuất được sử dụng để phân tích các tệp PDF để cung cấp cái nhìn sâu sắc vào mô hình phát hiện, phân loại các tệp PDF thành vô hại hoặc có mã độc hại. Hình 4 cung cấp kiến trúc toàn diện cho hệ thống phát hiện đề xuất từ giai đoạn đầu vào đến giai đoạn đầu ra.

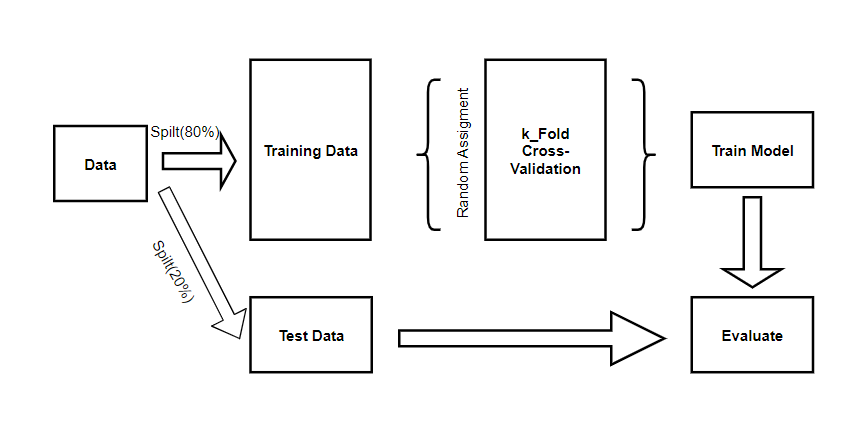


Hình 4. Structure of a PDF file

**3.1. Thu thập và Tiền xử lý Dữ liệu**

Do tính di động, tiện ích và độ tin cậy, tệp định dạng Di động (PDF) là định dạng tài liệu được sử dụng phổ biến nhất cho nhiều dịch vụ và ứng dụng. Tuy nhiên, danh tiếng và những tính năng của PDF này đã thu hút những hacker đen để khai thác chúng qua nhiều cách khác nhau. Thực sự, nhiều tính năng PDF quan trọng có thể được tận dụng bởi kẻ tấn công để tạo ra một tải trọng độc hại. Trong bài báo này, chúng tôi sử dụng một bộ dữ liệu PDF mới và toàn diện, tức là Evasive-PDFMal2022, bao gồm 10.025 bản ghi được phân phối thành 4.468 bản ghi vô hại và 5.557 bản ghi độc hại. Ngoài ra, Evasive-PDFMal2022 bao gồm 37 tính năng tĩnh quan trọng, bao gồm 12 tính năng chung và 25 tính năng cấu trúc được trích xuất từ mỗi tệp PDF [57]. Các ví dụ về tính năng bao gồm kích thước PDF, ký tự tiêu đề, mã hóa, kích thước siêu dữ liệu, số trang, đầu trang, số hình ảnh, văn bản, số đối tượng, đối tượng font, số tệp nhúng, và kích thước trung bình của tất cả các phương tiện nhúng.

Dữ liệu được thu thập của Evasive-PDFMal2022 được nhập qua MATLAB 2021 để hỗ trợ được dừng lại và chuẩn bị để sử dụng với các thuật toán học có giám sát. Sau khi được nhập khẩu, tập dữ liệu ban đầu có sẵn dưới dạng định dạng tệp giá trị được phân tách bằng dấu phẩy (.CSV) được lưu trữ bởi MATLAB dưới dạng bảng chứa các bản ghi và đặc điểm. Sau đó, các bảng dữ liệu trải qua một số quy trình làm sạch dữ liệu, chẳng hạn như sửa các bản ghi không chính xác/không đầy đủ và loại bỏ các bản ghi trùng lặp/có lỗi. Cuối cùng, dữ liệu được chia thành đào tạo (70%), xác nhận (10%), và thử nghiệm (20%) bằng cách sử dụng xác thực chéo k-Fold (với k = 5), như minh họa trong Hình 5 bên dưới. Theo hình vẽ, 20% tập dữ liệu được chia ra để xác thực mô hình lần cuối. Để so sánh, 80% tập dữ liệu được sử dụng để huấn luyện và xác thực mô hình cho nhiều lần. Ở mỗi lần gấp (chẳng hạn như năm lần), việc chia ngẫu nhiên mới cho 80% là 70% cho việc huấn luyện và 10%để xác thực, dẫn đến năm tập huấn luyện và xác nhận khác nhau. Đối với mỗi lần gấp, mô hình được đánh giá và kết quả hiệu suất tổng thể cuối cùng là mức trung bình của các kết quả đạt được ở tất cả các nếp gấp. Tóm lại, đối với tập dữ liệu của chúng tôi, ~ 7000 mẫu được sử dụng để đào tạo (70%),~1000 mẫu được sử dụng để xác nhận (10%) và ~2000 mẫu được sử dụng để thử nghiệm (20%)



Hình 5. Structure of a PDF file

**3.2. Mô Hình Cây Quyết Định Có Thể Tối Ưu Hóa (O-DT)**

Một thuật toán cây quyết định (DT) là một phương pháp học có giám sát không tham số được sử dụng để thực hiện các nhiệm vụ phân loại và hồi quy. DT chủ yếu sử dụng một cây xác suất giúp quá trình ra quyết định của một quy trình cụ thể và có thể dự đoán giá trị của một biến mục tiêu. Ví dụ, việc quyết định giữa hai ý tưởng đầu tư dự án có thể được thực hiện thông qua cây quyết định. Ý tưởng chính của DT là xây dựng một mô hình để học các quy tắc quyết định được suy ra từ các đặc trưng dữ liệu, có thể được sử dụng sau này để đưa ra quyết định và dự đoán. Một cây quyết định có thể tối ưu hóa (O-DT) là một cây quyết định sử dụng các tham số và siêu tham số tối ưu để xây dựng một hệ thống phát hiện bằng cách thử nghiệm một không gian tìm kiếm được định trước cho các siêu tham số khác nhau.

Chúng tôi đã sử dụng thuật toán AdaBoost trong công việc này để xây dựng mô hình cây quyết định của chúng tôi với các tùy chọn siêu tham số khác nhau. AdaBoost (phương pháp học máy tăng cường tương thích) là một phương pháp tăng cường trong đó trọng số được phân phối lại cho mỗi ví dụ, với trọng số cao được phân phối cho các ví dụ được phân loại sai, giúp giảm thiểu độ chệch và phương sai trong quá trình học. Hình 4 cho thấy cách tăng cường được sử dụng trong AdaBoost DT bằng cách sử dụng một số người học mở rộng tuần tự. Ngoại trừ người học đầu tiên, mỗi người học kế tiếp được tạo ra từ những người học đã được tạo ra trước đó (người học yếu được biến đổi thành người học mạnh). Tóm lại, Bảng 2 trình bày các siêu tham số được tối ưu hóa cuối cùng để phát triển O-DT.

A close-up of a document

Description automatically generated

Bảng 2

**3.3. Kiểm Thử và Đánh Giá Mô Hình**

Quá trình kiểm thử và đánh giá mô hình là một bước quan trọng để hiểu về hiệu suất của một mô hình học máy và có thêm thông tin về ưu điểm và nhược điểm của mô hình. Trong nghiên cứu này, chúng tôi đã kiểm thử mô hình bằng cách sử dụng phương pháp kiểm thử chéo 5 lần và tập dữ liệu kiểm thử (~2000 mẫu), sau đó đánh giá hiệu suất của nó tương ứng. Các chỉ số đánh giá tiêu chuẩn đã được sử dụng để đánh giá hiệu quả của mô hình phát hiện trong các giai đoạn đào tạo, xác nhận và kiểm thử. Hình 6 tóm tắt các chỉ số đánh giá hiệu suất tiêu chuẩn được sử dụng trong công việc này.

Accuracy (CA): CA= (TP +TN)/(TP + FP + FN + TN)

Precision (PR): PR= TP/(TP + FP)

Recall (RE): RE= TP/(TP + FN)

F1 Score (F1): F1= 2TP/( 2TP + FP + FN)

A screenshot of a computer

Description automatically generated

**4. Kết Quả và Phân Tích**

Phần này trình bày kết quả đánh giá hiệu suất cho hệ thống phát hiện mã độc PDF được đề xuất dựa trên các chỉ số khác nhau. Ngoài ra, một so sánh với các mô hình hiện đại được thực hiện. Để bắt đầu, Hình 7 theo dõi quỹ đạo của sai số phân loại tối thiểu (MCE) trong suốt các vòng lặp đào tạo của cây quyết định có thể tối ưu hóa. Theo hình vẽ, MCE được ghi nhận ban đầu sau vòng lặp đầu tiên là 3.4%, ghi lại độ chính xác phân loại tối đa là 96.6%. Sau đó, MCE giảm đột ngột đến các siêu tham số MCE tối thiểu chỉ sau ba vòng lặp học, ghi lại MCE là 1.3% và độ chính xác phân loại là 98.7%. Sau đó, quỹ đạo của MCE tiếp tục giảm nhẹ đến các siêu tham số điểm tốt nhất, nơi nó ổn định sau vòng lặp thứ 13, ghi lại 1.16% MCE với 98.84% độ chính xác phân loại, và duy trì ổn định đến cuối quá trình học (30 vòng lặp).

A graph with green and orange lines

Description automatically generated

Hình 7

Ngoài ra, Hình 8 thể hiện kết quả ma trận nhầm lẫn nhị phân cho hệ thống phát hiện mã độc PDF được đề xuất sử dụng cây quyết định có thể tối ưu hóa. Ma trận được trình bày gồm các khối: (i) khối trên bên trái, biểu diễn True Positive (TP), thể hiện số lượng mẫu thực tế là mẫu vô hại và mô hình ML dự đoán là mẫu vô hại. Số TP trong ma trận này là 4412. (ii) Khối trên bên phải, biểu diễn False Positive (FP), thể hiện số lượng mẫu thực tế là mẫu vô hại và mô hình ML dự đoán là mẫu độc hại. Số FP trong ma trận này là 56. (iii) Khối dưới bên trái, biểu diễn False Negatives (FN), thể hiện số lượng mẫu thực tế là mẫu độc hại và mô hình ML dự đoán là mẫu vô hại. Số FN trong ma trận này là 60. (iv) Khối dưới bên phải, biểu diễn True Negatives (TN), thể hiện số lượng mẫu thực tế là mẫu độc hại và mô hình ML dự đoán là mẫu độc hại. Số TN trong ma trận này là 5497.

A blue squares with white text

Description automatically generated

Hình8

Hơn nữa, Bảng 3 cung cấp một tóm tắt về kết quả chỉ số hiệu suất thu được cho hệ thống phát hiện mã độc PDF được đềxuất dưới các chỉ số khác nhau, bao gồm Số mẫu Dự đoán Sai (IPS), Số mẫu Dự đoán Đúng (CPS), Tổng số Mẫu (TNS), Tốc độ Dự đoán (PS), Thời gian Dự đoán (PT), Thời gian Đào tạo (TT), Độ chính xác Dự đoán (CA), Diện tích dưới đường cong Dự đoán (AUC), Độ nhạy/Recall Dự đoán (RE), Độ chính xác Dự đoán (PR), Trung bình hòa các dự đoán/F-Score (F1), và Độ chính xác Cân bằng (BCA). Kết quả đạt được thể hiện hiệu suất và khả năng phát hiện cao cho hệ thống được đề xuất, đạt các yếu tố hiệu suất cao vượt quá 98,80% cho độ chính xác hệ thống, độ nhạy và độ chính xác.

A table with numbers and letters

Description automatically generated with medium confidence

Bảng 3

Cuối cùng, Bảng 4 so sánh hiệu suất của mô hình được đề xuất của chúng tôi với một số mô hình hiện tại khác trong cùng lĩnh vực nghiên cứu. Bảng so sánh tám hệ thống phát hiện mã độc PDF sử dụng các mô hình học khác nhau, bao gồm Zhang et al. [58] sử dụng mạng nơ-ron perceptron đa tầng (MLP-NN), Jiang et al. [34] sử dụng thuật toán học bán giám sát (Semi-SL), Li et al. [59] sử dụng công cụ thông minh được biết đến là JSUNPACK, Nissim et al. [60] sử dụng kỹ thuật máy vector hỗ trợ (SVM), Mohammed et al. [61] sử dụng mạng nơ-ron tích chập ResNet-50 (ResNet-50 CNN), Nataraj et al. [62] sử dụng bộ phân loại rừng ngẫu nhiên (RFC), Lakshmanan et al. [63] sử dụng bộ phân loại đa nguyên (VEC) sử dụng bộ phân loại rừng ngẫu nhiên (RFC) và k-nearest neighbor (kNN), và Cohen et al. [64] sử dụng kỹ thuật máy vector hỗ trợ (SVM). Ngoài yếu tố mô hình học, chúng tôi đã xem xét bốn yếu tố hiệu suất để so sánh với các mô hình hiện tại: độ chính xác, độ chính xác, độ nhạy và F-Score. Tổng thể, hệ thống được đề xuất vượt trội trong tất cả các yếu tố đánh giá, với hiệu suất đáng chú ý so với các mô hình khác dựa trên kỹ thuật SVM.

A table with numbers and text

Description automatically generated

Bảng 4

**5. Kết Luận và Ghi Chú**

Do xu hướng toàn cầu hướng về sự biến đổi số và công việc từ xa, nhu cầu về tài liệu kỹ thuật số đã tăng đáng kể. Sự tăng này trong việc sử dụng tài liệu số đã đi kèm một sự gia tăng đối ngược trong việc phát triển mã độc có thể đe dọa tệp tin và máy người dùng. Tệp PDF là một trong những loại tệp số phổ biến nhất trên toàn cầu, điều này khiến chúng trở nên rất dễ bị tổn thương bởi nhiều loại đe dọa và mã độc hại. Các vector nhiễm bệnh như vậy (được phát triển bởi các hacker) ẩn mã độc hại được nhúng trong tài liệu PDF để nhiễm bệnh máy của nạn nhân. Điều này dẫn đến mã độc PDF và đòi hỏi các kỹ thuật để phân biệt giữa các tệp đúng và tệp độc hại. Do đó, một hệ thống thông minh mới cho việc phát hiện mã độc PDF được đề xuất, phát triển và đánh giá trong bài báo này. Hệ thống được đề xuất sử dụng một mô hình học máy hiệu suất cao sử dụng cây quyết định có thể tối ưu hóa với thuật toán AdaBoost. Hệ thống được đào tạo và đánh giá trên một bộ dữ liệu mới toàn diện cho tài liệu PDF được biết đến với tên gọi Evasive-PDFMal2022. Kết quả mô phỏng đã chứng minh sự ưu việt của hệ thống đề xuất về độ chính xác phát hiện, độ chính xác, độ nhạy, F-Score và chi phí phát hiện. Cuối cùng, mô hình đề xuất vượt trội so với các mô hình hiện đại khác trong cùng lĩnh vực nghiên cứu. Mô hình đề xuất có thể được tổng quát hóa và áp dụng để cung cấp nhiều dịch vụ phát hiện trong nhiều lĩnh vực khác nhau [65,66]. Tóm lại, đóng góp của công việc này có thể được tóm tắt như sau:

* Một mô hình dựa trên học máy toàn diện để phân tích tài liệu PDF để nhận diện tệp PDF độc hại từ tệp vô hại.
* Mô hình đề xuất sử dụng cây quyết định có thể tối ưu hóa với thuật toán AdaBoost và siêu tham số tối ưu.
* Mô hình đề xuất dựa trên việc sử dụng một tập dữ liệu mới (Evasive-PDFMal2022) gồm 10.025 bản ghi phân phối và 37 đặc trưng tĩnh quan trọng (đặc trưng tổng quát và cấu trúc) được trích xuất từ mỗi tệp PDF.
* Kết quả thử nghiệm chứng minh hiệu suất của hệ thống phát hiện PDF đề xuất, đạt đến độ chính xác dự đoán 98.84% với khoảng thời gian dự đoán ngắn là 2.174 µSec.
* Thảo luận chỉ ra một số hạn chế trong các phương pháp hiện đại hiện tại có thể cung cấp hướng đi cho nghiên cứu tương lai.

Trong tương lai, một số tài liệu quan trọng và thông dụng khác có thể được xem xét trong quá trình phát hiện độc hại như tài liệu văn phòng (.docx, .xlxs, pptx., vv.). Ngoài ra, có thể sử dụng các mạng nơ-ron sâu hơn trong trường hợp có sự thay đổi chính xác trong mã độc cần các thuật toán học sâu như mạng nơ-ron tích chập (CNN) và bộ nhớ ngắn hạn dài (LSTM).