Nghiên cứu bài toán phát hiện tin nhắn rác trên thiết bị di động và một số vấn đề gợi mở

Vũ Duy Hiến[1], Chu Văn Huy[2] , Bùi Thị Hồng Nhung[3], Lê Thị Hồng Nhung[4]
Học viện Ngân hàng, Việt Nam

Ngày nhận: 06/06/2024 Ngày nhận bản sửa: 20/06/2024 Ngày duyệt đăng: 01/07/2024

Tóm tắt: *Dịch vụ tin nhắn di động (Short Message Service-SMS) ngày càng trở nên phổ biến. Tuy nhiên, tội phạm công nghệ cao đã lợi dụng tin nhắn rác để gây ra các vụ lừa đảo chiếm đoạt tài sản với mức độ thiệt hại nghiêm trọng. Đến nay, cộng đồng nghiên cứu và các nhà phát triển ứng dụng đã đề xuất những giải pháp phát hiện tin nhắn rác có độ chính xác cao, nhưng trên thực tế số vụ việc và mức độ thiệt hại tài chính của người dùng di động do vấn nạn này gây ra có dấu hiệu gia tăng trong những năm gần đây. Sử dụng phương pháp nghiên cứu phân tích, tổng hợp những công trình nghiên cứu điển hình trong việc phát hiện tin nhắn rác trong năm năm trở lại đây ở Việt Nam và trên thế giới kết hợp cùng với kết quả quan sát và thu thập số liệu thực tế về vấn nạn tin nhắn rác từ các nguồn tin cậy, bài viết phân tích những vấn đề còn tồn tại trong bài toán phát hiện tin nhắn rác trên thiết bị di động và đề xuất một số khuyến nghị nhằm giảm thiểu tối đa những thiệt hại tài chính mà người dùng di động đang gặp phải.*

Từ khóa: *Tin nhắn rác, Học máy, Phân loại văn bản, Ứng dụng di động, Tài chính-ngân hàng, Lừa đảo*

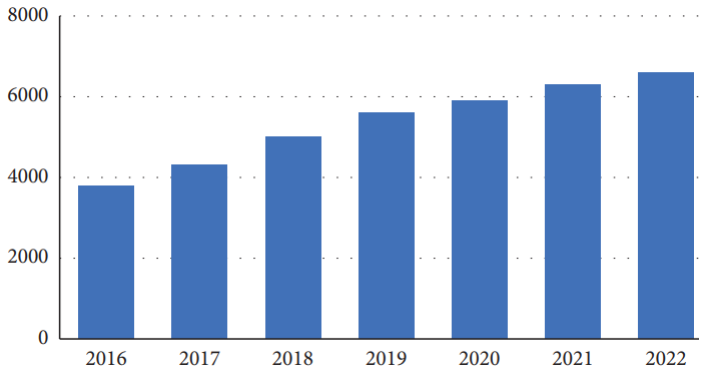
A study on the spam sms detection problem and some recommendations

Abstract: Short Message Service-SMS is a popular communication method for us. However, high-tech criminals have used spam SMS to cause scams with serious damage. Up to now, the research community and application developers have proposed solutions to detect spam SMS with high accuracy, but in fact, the number of spam SMSs and the amount of financial loss of mobile users caused by this problem have still increased in recent years. Using the research methods of analysis and synthesis on the typical related work, and the research methods of observation and collection actual data from truthful and confident sources, this paper points out the existing problems of spam SMSs detection and proposes some recommendations for creating more efficient and practical solutions.

Keywords: Spam SMS, Machine learning, Text classification, Mobile apps, Banking-finance, Fraud
DOI: 10.59276/JELB.2024.07CD.2763

Vu, Duy Hien[1], Chu, Van Huy[2] , Bui, Thi Hong Nhung[3] , Le, Thi Hong Nhung[4]
Email: hienvd@hvnh.edu.vn[1], huycv@hvnh.edu.vn[2], nhungbth@hvnh.edu.vn[3], nhunglth@hvnh.edu.vn[4]
Organization of all: Banking Academy of Vietnam

ISSN 3030 - 4199



(đơn vị: triệu người)

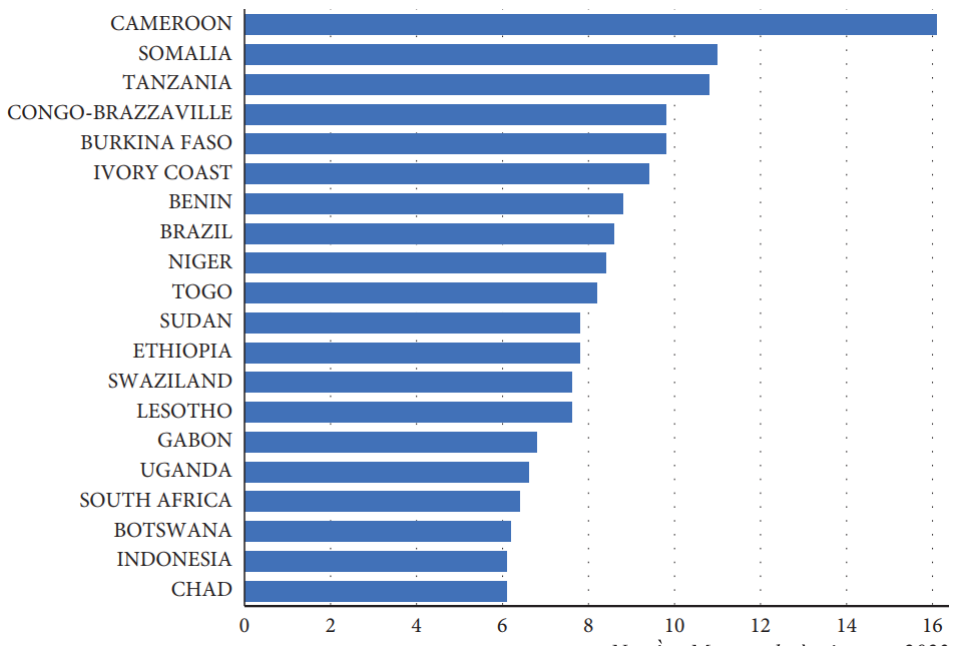
Nguồn: Degenhar, 2024; Maqsood và cộng sự, 2023

Hình 1. Thống kê số người dùng thiết bị di động giai đoạn 2016- 2022

1. Giới thiệu

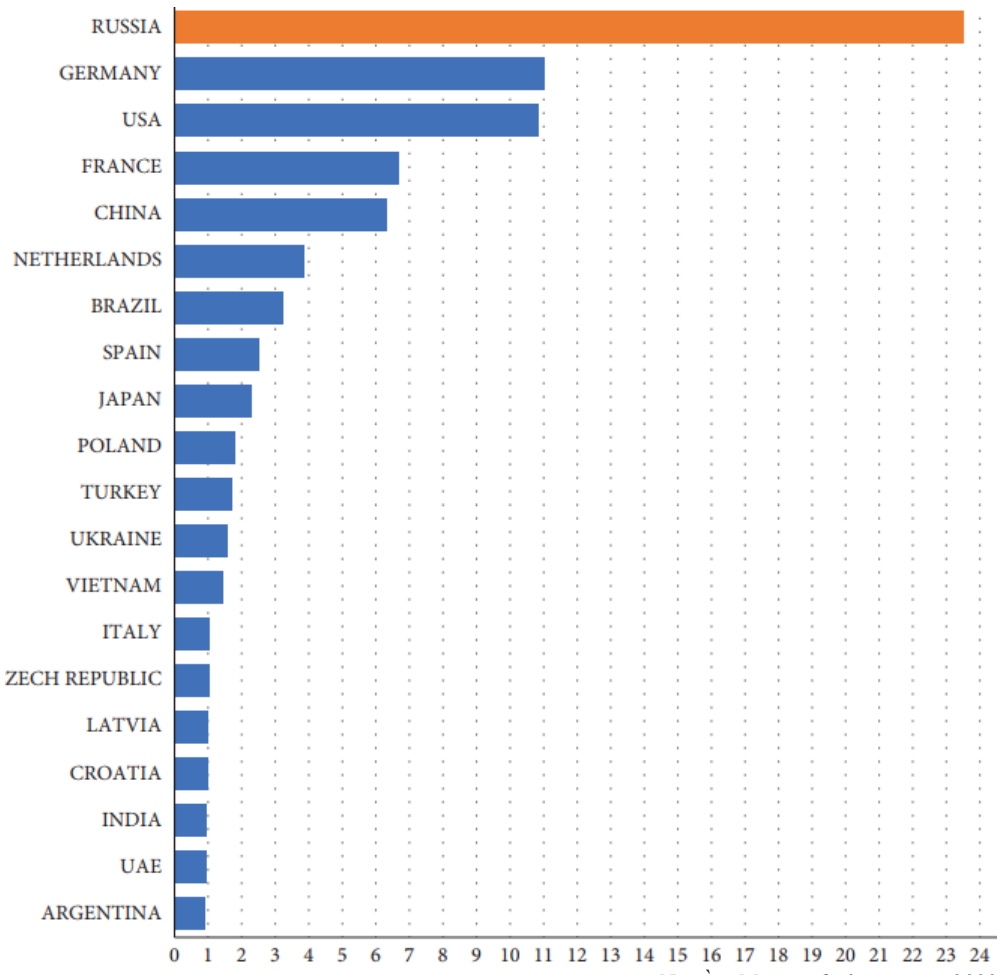
Trong cuộc sống hiện đại ngày nay, điện thoại di động đã trở thành thiết bị hữu ích hỗ trợ chúng ta kết nối, cập nhật thông tin, học tập, làm việc và giải trí. Theo thống kê của tổ chức uy tín Statista, trên toàn thế giới có đến hơn 6 tỷ người dùng di động tính đến năm 2022 (Degenhar, 2024; Maqsood và cộng sự, 2023).

Tuy nhiên, thực tế cho thấy rằng điện thoại di động cũng mang lại nhiều phiền toái cho người dùng. Nhiều dữ liệu nhạy cảm hoặc riêng tư lưu trữ trên thiết bị di động có thể bị rò rỉ nếu như người dùng không cẩn trọng, và mỗi thiết bị thông minh có thể bị điều khiển trở thành công cụ tấn công của tin tặc. Điển hình như vào năm 2017, trang lowyat.net đã công bố và báo cáo lên Ủy ban Truyền thông và đa phương tiện Malaysia về vụ việc 46 triệu người dùng di động của nước này đã bị rò rỉ những dữ liệu nhạy cảm và quan trọng như địa chỉ, số thẻ định danh, thông tin thẻ sim điện thoại và các thông tin cá nhân khác (Bình, 2017). Đầu năm 2024, Công ty viễn thông AT&T- nhà cung cấp dịch vụ không dây bán lẻ lớn thứ 3 của Mỹ đã thông báo rằng dữ liệu của công ty này (bao gồm các thông tin cá nhân như số an sinh xã hội…) liên quan tới 7,6 triệu chủ tài khoản hiện tại và 65,4 triệu chủ tài khoản trước đây đã bị rò rỉ và có khả năng sự việc này đã diễn ra từ trước năm 2019 (Chi, 2024). Tại Việt Nam,



Nguồn: Maqsood và cộng sự, 2023

Hình 2. Số tin nhắn rác trung bình tháng/người dùng di động của 20 quốc gia chịu ảnh hưởng lớn nhất từ vấn nạn tin nhắn rác năm 2022



Nguồn: Maqsood và cộng sự, 2023

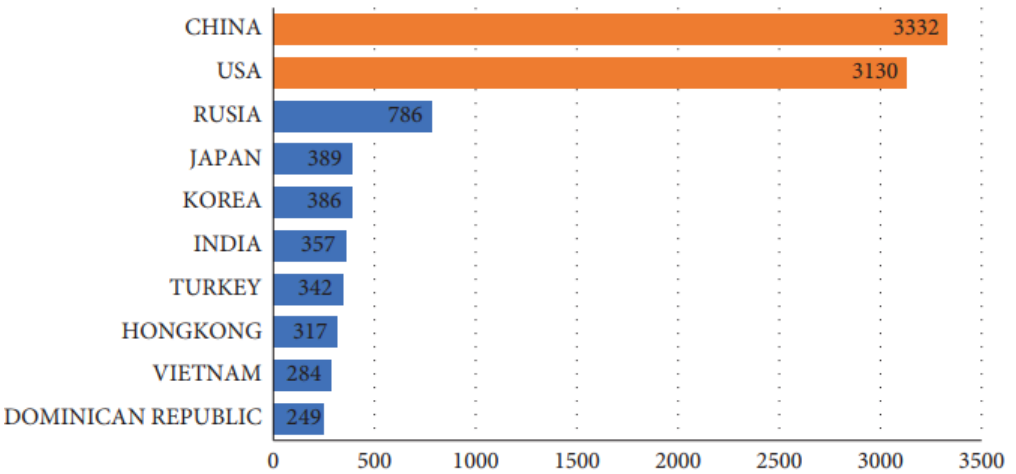
Hình 3. Các quốc gia có lượng tin nhắn rác gửi đi nhiều nhất thế giới năm 2022

trong những năm qua cũng đã xảy ra nhiều vụ việc lộ lọt dữ liệu di động của một số cá nhân nổi tiếng khiến cho danh tiếng, uy tín và đời sống riêng tư của họ bị ảnh hưởng nghiêm trọng.

Nghiên cứu này tập trung vào vấn đề rất phổ biến đối với người sử dụng thiết bị di động, đó chính là vấn nạn tin nhắn rác (hay còn gọi là tin nhắn spam SMS). Về bản chất, tin nhắn rác là những tin nhắn mà người dùng không yêu cầu hoặc không mong muốn nhận được thường chứa thông tin quảng cáo, các thông điệp không liên quan, hoặc thậm chí là lừa đảo hoặc mã độc cài cắm vào thiết bị người dùng (Nhi, 2024).

Theo hãng phát triển ứng dụng chặn cuộc gọi và tin nhắn rác hàng đầu Truecaller (https://www.truecaller.com) và (Maqsood và cộng sự, 2023), trung bình một tháng của năm 2022, một người dân ở Cameroon nhận được 16 tin nhắn rác, ở Nam Phi là khoảng 6 tin nhắn rác và dường như các nước châu Phi phải đối mặt với nạn tin nhắn rác nghiêm trọng hơn so với các khu vực khác trên thế giới (xem Hình 2).

Tổ chức thống kê nổi tiếng Statista (https://www.statista.com) cho biết nước Nga đứng đầu danh sách quốc gia phát tán thư/tin nhắn rác nhiều nhất chiếm 23,5%, Đức đứng thứ hai chiếm 11% và Mỹ thứ ba chiếm 10,85% trong tổng số thư/tin nhắn rác năm 2022 (Hình 3).



Nguồn: Maqsood và cộng sự, 2023

Hình 4. Các quốc gia có số lượng tin nhắn rác trực tiếp nhiều nhất thế giới năm 2021

Tiếp theo, báo cáo năm 2021 (Gajić, 2022) được trình bày ở Hình 4 cho thấy rằng Trung Quốc đứng đầu, còn Việt Nam đứng thứ 9 trong danh sách các quốc gia có số lượng tin nhắn rác trực tiếp (tin nhắn mà tin tặc đang tương tác trực tiếp với người dùng để lừa đảo) cao nhất thế giới (xét tại thời điểm viết báo cáo).

Bằng việc sử dụng tin nhắn rác, tội phạm công nghệ cao có thể đánh cắp dữ liệu riêng tư & nhạy cảm (thông tin cá nhân, mật khẩu đăng nhập của các loại tài khoản, mã OTP xác thực giao dịch, ảnh, clip, tin nhắn…) và chiếm đoạt tiền trong tài khoản ngân hàng hoặc tài khoản di động của người dùng. Theo tổ chức nghiên cứu và tư vấn nổi tiếng Fact.MR (Fact.MR, 2021), 75% người dùng di động có cài đặt ứng dụng tài chính trên thiết bị của mình khiến cho họ phải đối mặt với nguy cơ cao bị thất thoát tài chính bởi những tin nhắn rác lừa đảo.

Như vậy có thể thấy rằng, nạn tin nhắn rác là một vấn đề nhức nhối đối với bất kỳ quốc gia, ở bất kỳ châu lục nào. Trong suốt nhiều năm qua, những tin nhắn rác không chỉ làm phiền người dùng mà chúng còn gây ra thiệt hại lớn về mặt tài chính. Do đó, việc nghiên cứu đề xuất các giải pháp công nghệ hiện đại nhằm phát hiện tin nhắn rác với độ chính xác cao và phát triển những ứng dụng ngăn chặn tin nhắn rác là rất quan trọng và có ý nghĩa trong việc đảm bảo an toàn nói chung và hạn chế rủi ro về mặt tài chính cho người dùng di động.

Đến nay, cộng đồng nghiên cứu và các nhà phát triển ứng dụng đã đề xuất những giải pháp phát hiện tin nhắn rác có độ chính xác cao, tuy nhiên thực tế lại cho thấy rằng số vụ việc và mức độ thiệt hại tài chính của người dùng di động do vấn nạn này gây ra có dấu hiệu gia tăng trong những năm gần đây. Vì vậy, bài viết này tập trung vào hai đóng góp chính như sau: (1) Đánh giá những công trình nghiên cứu điển hình cho vấn đề phát hiện tin nhắn rác dựa trên phương pháp nghiên cứu phân tích, tổng hợp từ đó xác định các tồn tại của vấn đề và (2) Đưa ra những khuyến nghị quan trọng đối với các vấn đề đang tồn tại của bài toán phát hiện tin nhắn rác.

Để đạt được hai đóng góp kể trên, nhóm tác giả sử dụng ba phương pháp nghiên cứu chính là: phương pháp phân tích, tổng hợp; phương pháp quan sát và thu thập số liệu thực tế; và phương pháp chuyên gia. Bên cạnh đó, dữ liệu phục vụ nghiên cứu này bao gồm: những công trình nghiên cứu điển hình cho vấn đề phát hiện tin nhắn rác

đã được công bố trên các tạp chí và hội nghị khoa học uy tín, và số liệu thực tế từ các báo cáo thiệt hại về vấn nạn này của những tổ chức tin cậy.

Ngoài Mục Giới thiệu và Kết luận, kết cấu nội dung chính của bài báo này bao gồm: Phần 2 khảo sát thực trạng nghiên cứu và phát triển giải pháp phát hiện tin nhắn rác và các vấn đề thực tế mà người dùng gặp phải bởi vấn nạn này, từ đó xác định khoảng trống nghiên cứu của bài toán phát hiện tin nhắn rác; Phần 3 đề xuất một số khuyến nghị nhằm khắc phục các vấn đề còn tồn tại của bài toán phát hiện tin nhắn rác trên thiết bị di động.

2. Tổng quan về bài toán phát hiện tin nhắn rác trên thiết bị di động

2.1. Tin nhắn rác và các phương pháp, kỹ thuật phát hiện tin nhắn rác

Như đã đề cập ở trên, tin nhắn rác (Nhi, 2024) là những tin nhắn mà người dùng không yêu cầu hoặc không mong muốn nhận được, thường chứa thông tin quảng cáo, tiếp thị thậm chí là lừa đảo hoặc phần mềm mã độc tấn công người dùng. Cho đến nay, hai cách tiếp cận phổ biến được sử dụng để đề xuất giải pháp cho bài toán phát hiện tin nhắn rác là phương pháp phát hiện dựa trên luật và phương pháp lọc dựa trên nội dung (Xia & Chen, 2021).

- *Phương pháp phát hiện dựa trên luật* được những tập đoàn công nghệ lớn như Google, Symantec, McAfee ưa thích ứng dụng để loại bỏ những tin nhắn, thư điện tử rác (M. Hameed & Hussein Ali, 2021). Phương pháp này khá hữu hiệu nhưng tập luật sẽ ngày một lớn dần lên để tăng cường độ chính xác và khiến cho thời gian rà soát mỗi tin nhắn ngày một tăng lên ảnh hưởng tới chất lượng dịch vụ.
- *Phương pháp lọc dựa trên nội dung* phân loại tin nhắn rác dựa trên các mô hình học máy được huấn luyện từ những bộ dữ liệu tin nhắn. Với phương thức tiếp cận này, các hệ thống lọc tin nhắn rác tự động với tốc độ nhanh và độ chính xác cao. Chính vì thế, phương pháp phân loại tin nhắn rác dựa trên lọc nội dung sử dụng học máy được nghiên cứu và phát triển nhiều hơn trong những năm trở lại đây (Hsu, 2020; Xia & Chen, 2021). Trong nghiên cứu này, nhóm tác giả sẽ tập trung phân tích, tổng hợp và đánh giá những giải pháp điển hình được đề xuất theo phương pháp này.

2.2. Thực trạng nghiên cứu đề xuất giải pháp phát hiện tin nhắn rác trên thiết bị di động

Bài nghiên cứu cung cấp kết quả khảo sát những bộ dữ liệu tin nhắn di động công khai được sử dụng phổ biến và những công trình nổi bật đề xuất giải pháp phát hiện tin nhắn rác dựa trên học máy trong thời gian gần đây.

2.2.1. Các bộ dữ liệu tin nhắn di động phổ biến

Hiện nay, bên cạnh các nghiên cứu chưa sẵn sàng công khai dữ liệu thì đã có một số bộ dữ liệu tin nhắn SMS được công bố phục vụ cho nghiên cứu về bài toán lọc tin nhắn rác. Bộ sưu tập tin nhắn SMS (Almeida và cộng sự, 2011) được coi là bộ dữ liệu công khai đầu tiên năm 2012 trên trang Kaggle và kho dữ liệu nghiên cứu của Đại học California Irvine (University of California Irvine, 2012), trong đó có tất cả 5.574 tin nhắn và 747 tin nhắn rác được trích xuất từ kho (Hidalgo và cộng sự, 2006), bộ dữ liệu từ luận án (Tagg, 2009), kho tin nhắn (Chen & Kan, 2012) và trang web diễn đàn Grumbletext của Anh (grumbletext.co.uk). Đại học Quốc Gia Singapore cũng đã công bố kho tin nhắn bao gồm hơn 67.000 tin

Bảng 1. Các bộ dữ liệu tin nhắn SMS phổ biến cho bài toán phát hiện tin nhắn rác

Bộ dữ liệu	Năm công bố	Tổng số tin nhắn	Số tin nhắn rác
Kho tin nhắn SMS của Almeida và cộng sự, 2011	2011	5.574	747
Kho tin nhắn SMS của Đại học Quốc Gia Singapore (Chen & Kan, 2012)	2015	67.063	Không phân loại
Kho tin nhắn trích xuất bởi SpamHunter (Tang và cộng sự, 2022)	2022	25.889	947

Nguồn: Nhóm nghiên cứu tổng hợp

nhắn (Chen & Kan, 2012) tính đến đợt cập nhật cuối cùng năm 2015. Thời gian gần đây, (Tang và cộng sự, 2022) đã đề xuất một công cụ mang tên *"SpamHunter"* nhằm trích xuất tin nhắn rác từ những hình ảnh được công khai trong giai đoạn 2018-2022 trên mạng xã hội Twitter (https://twitter.com) (nay được đổi tên là X). Bảng 1 tổng hợp các bộ dữ liệu tin nhắn SMS phổ biến được sử dụng cho bài toán phát hiện tin nhắn rác.

2.2.2. Những nghiên cứu đề xuất điển hình trong phát hiện tin nhắn rác trên thiết bị di động

Những giải pháp đề xuất cho phát hiện tin nhắn rác trên thiết bị di động dựa trên học máy chủ yếu sử dụng *những kỹ thuật phân lớp truyền thống* như Naïve Bayes, láng giềng gần nhất-K-Nearest Neighbors (KNN), Support vector machine (SVM), cây quyết định hoặc *các mô hình học sâu* như Long Short Term Memory (LSTM) và Convolutional Neural Networks (CNN) (HUSSEIN và cộng sự, 2023; LIU và cộng sự, 2021; SALMAN và cộng sự, 2024; Xia & Chen, 2021). Dưới đây là kết quả tổng hợp những đề xuất giải pháp lọc tin nhắn rác điển hình theo từng kỹ thuật học máy.

2.2.2.1. Các giải pháp phát hiện tin nhắn rác điển hình dựa trên những kỹ thuật học máy truyền thống

Trong Bảng 2 dưới đây, nhóm nghiên cứu tổng hợp lại các giải pháp phát hiện tin nhắn rác điển hình gần đây dựa trên những kỹ thuật học máy truyền thống theo ba phương diện: thuật toán sử dụng, bộ dữ liệu huấn luyện và thử nghiệm, và độ tính xác. Kết quả trong Bảng 2 cho thấy rằng, những mô hình dựa trên học máy truyền thống đã phát hiện ra tin nhắn rác với độ chính xác tương đối cao (đều trên 90%). Đặc biệt là, những giải pháp Ayaz và cộng sự, 2024; Ghourabi & Alohaly, 2023; Maqsood và cộng sự, 2023; Srinivasarao & Sharaff, 2023 đều có độ chính xác hơn 99%.

Bảng 2. Tổng hợp các giải pháp phát hiện tin nhắn rác điển hình dựa trên những kỹ thuật học máy truyền thống

Công trình	Thuật toán	Bộ dữ liệu sử dụng	Độ chính xác (%)
Arifin và cộng sự, 2016	Naïve Bayes	Almeida và cộng sự	98,5
Agarwal và cộng sự, 2016	Cây quyết định	Almeida và cộng sự với kho tin nhắn tiếng Ấn Độ	96,04
	SVM	Almeida và cộng sự với kho tin nhắn tiếng Ấn Độ	98,23
Sonowal & Kuppusamy, 2018	SVM	Almeida và cộng sự	94,20
	Cây quyết định	Almeida và cộng sự	94,20
	Naïve Bayes	Almeida và cộng sự	94,20

Công trình	Thuật toán	Bộ dữ liệu sử dụng	Độ chính xác (%)
Jain & Gupta, 2019	Cây quyết định	Almeida và cộng sự	94,20
	SVM	Almeida và cộng sự	94,20
Sjarif và cộng sự, 2019	Naïve Bayes	Kho tin nhắn SMS của Almeida và cộng sự	97,5
Mishra & Soni, 2020	Naïve Bayes	Almeida và cộng sự	96,29
Xia & Chen, 2020	Markov ẩn	Almeida và cộng sự	95,90
Sousa và cộng sự, 2021	KNN	Almeida và cộng sự	98,15
Xia & Chen, 2021	Markov ẩn	Almeida và cộng sự	96,90
Ghourabi & Alohaly, 2023	Kết hợp SVM, KNN, LightGBM, CNN	Almeida và cộng sự	99,91
	KNN	Almeida và cộng sự	99,91
Srinivasarao & Sharaff, 2023	SVM	Bộ dữ liệu riêng tư	99,82
	KNN	Bộ dữ liệu riêng tư	99,82
Maqsood và cộng sự, 2023	SVM	Almeida và cộng sự	99,6
Ayaz và cộng sự, 2024	Naïve Bayes	Kho tin nhắn ngôn ngữ Latinh (Romanized messages)	97,33
	SVM	Kho tin nhắn ngôn ngữ Latinh (Romanized messages)	99,42
	Cây quyết định	Kho tin nhắn ngôn ngữ Latinh (Romanized messages)	97,33

Nguồn: Nhóm nghiên cứu tổng hợp

2.2.2.2. Các giải pháp phát hiện tin nhắn rác điển hình dựa trên học sâu

Tương tự như với cách làm ở trên, các giải pháp phát hiện tin nhắn rác điển hình dựa trên những kỹ thuật học sâu được đề xuất trong bốn năm trở lại đây đã được nhóm nghiên cứu tổng hợp lại trong Bảng 3. Kết quả này cho thấy rằng, những giải pháp trên cũng đều có độ chính xác trên 90% trong việc phát hiện tin nhắn rác, trong đó phần lớn trên 95%, và khá nhiều giải pháp đạt được đến độ chính xác hơn 99% như của HUSSEIN và cộng sự, 2023; MAMBINA và cộng sự, 2024; Roy và cộng sự, 2020; Tuan và cộng sự, 2022.

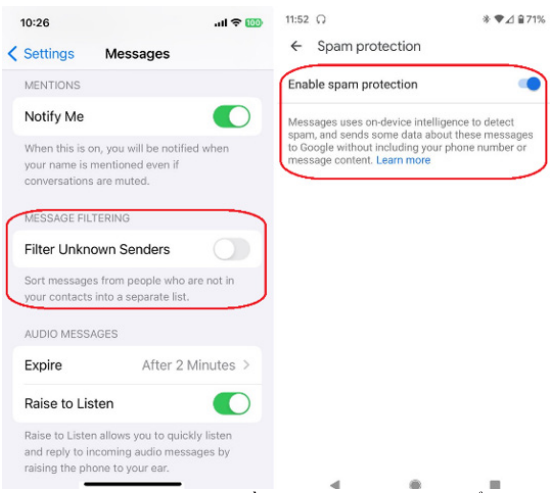
2.3. Thực trạng phát triển ứng dụng/công cụ phát hiện tin nhắn rác trên thiết bị di động

Bảng 3. Tổng hợp các giải pháp phát hiện tin nhắn rác điển hình dựa trên những kỹ thuật học sâu

Công trình	Thuật toán	Bộ dữ liệu sử dụng	Độ chính xác (%)
Roy và cộng sự, 2020	LSTM	Almeida và cộng sự	95,30
	CNN	Almeida và cộng sự	97,90
	LSTM kết hợp CNN	Almeida và cộng sự	99,44
Ghourabi và cộng sự, 2020	LSTM kết hợp CNN	Almeida và cộng sự	98,37
LIU và cộng sự, 2021	LSTM kết hợp CNN	Almeida và cộng sự	98,92
Osa & Elaigwu, 2021	Mô hình tuần tự	Almeida và cộng sự	98,30

Hikmaturokhman và cộng sự, 2022	Mạng nơ-ron dày đặc	Kho tin nhắn SMS tiếng Indonesia	95,63
	LSTM	Kho tin nhắn SMS tiếng Indonesia	94,76
	Bi-LSTM	Kho tin nhắn SMS tiếng Indonesia	94,75
Tuan và cộng sự, 2022	Kết hợp mô hình DNN và PhoBERT	Kho tin nhắn SMS tiếng Việt	99,53
HUSSEIN và cộng sự, 2023	LSTM kết hợp CNN	Almeida và cộng sự	99,56
Giri và cộng sự, 2023	CNN	Bộ dữ liệu riêng tư	98,44
MAMBINA và cộng sự, 2024	Kết hợp CNN, LSTM, LSTM lai	Kho tin nhắn SMS từ các công ty viễn thông của Tanzania	99,98
	Kết hợp CNN và BiLSTM	Almeida và cộng sự	98,38

Nguồn: Nhóm nghiên cứu tổng hợp



Nguồn: Nhóm nghiên cứu tổng hợp

Hình 5. Tính năng lọc tin nhắn rác trên iOS và Android

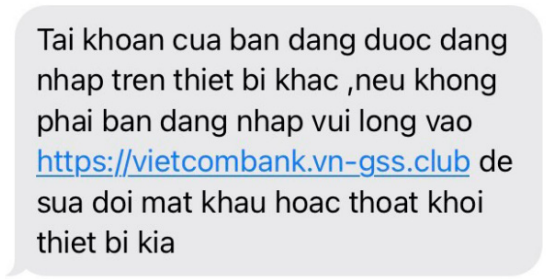
Để hỗ trợ cho người dùng di động, Google và Apple đã cung cấp tính năng lọc tin nhắn rác trên các nền tảng hệ điều hành Android và iOS.

Bên cạnh đó, các nhà phát triển phần mềm cung cấp ra thị trường một loạt ứng dụng lọc tin nhắn rác và cuộc gọi làm phiền gồm cả bản miễn phí và trả phí. Điển hình là những phần mềm TrueCaller, Robokiller, Key Messages với những tính năng nổi bật như chặn tin nhắn văn bản và cuộc gọi rác được báo cáo là spam bởi những người khác hoặc nhập bằng tay, được phát hiện từ danh sách liên lạc hoặc cuộc gọi lịch sử; cho phép xác minh số điện thoại nhờ việc tự động tìm kiếm các số chưa biết trong mạng lưới toàn cầu.

2.4. Một số thống kê thiệt hại của nạn tin nhắn rác

Trên thực tế, trong những năm gần đây vấn nạn tin nhắn rác và mức độ thiệt hại của người dùng di động không có dấu hiệu thuyên giảm mà còn tăng lên. Cụ thể như những số liệu dưới đây được nhóm nghiên cứu quan sát và thu thập từ những nguồn tin cậy.

Tại Mỹ, theo thống kê mới nhất của Robokiller (Robokiller, 2022)-ứng dụng chặn cuộc gọi và tin nhắn rác hàng đầu, người Mỹ đã nhận được 225 tỷ tin nhắn rác vào năm 2023 (tăng 157% so với năm trước) và 19.2 tỷ tin nhắn rác vào tháng 3 năm 2024. Theo Techreport (Laborde, 2024), trong số các vụ lừa đảo ở Mỹ thì có đến 21% liên quan đến tin nhắn rác. Nghiêm trọng hơn nữa, Truecaller và Ủy ban Thương mại Liên bang cho biết 68.4 triệu người Mỹ trở thành nạn nhân của các vụ lừa đảo qua điện thoại (Truecaller,



Nguồn: Nhóm nghiên cứu tổng hợp

Hình 6. Một tin nhắn rác lừa đảo

2022) với tổng thiệt hại hơn 330 triệu USD (Khristopher, 2023) trong năm 2022.

Ở Việt Nam, tổ chức bảo mật BKAV cho biết 75% người dùng từng nhận được tin nhắc/cuộc gọi rác lừa đảo (XM, 2023). Bằng việc sử dụng những trạm thu phát sóng di động BTS giả, tội phạm công nghệ cao có thể phát tán lên đến 100.000 tin nhắn rác/lừa đảo mỗi ngày (theo báo cáo của Bộ Thông tin và Truyền thông năm 2023 (An, 2024)). Hậu quả của những tin nhắn rác kể trên đã khiến cho nhiều người dùng di động mất tiền trong tài khoản ngân hàng. Điển hình là vào năm 2022 một khách hàng nữ đã mất 3 tỷ đồng chỉ trong vài phút sau khi làm theo những gì mà tin nhắn rác mạo danh ngân hàng yêu cầu thực hiện (Xuân, 2022). Một vụ việc khác tương tự đó là khách hàng đã mất sạch 38 triệu đồng trong tài khoản ngân hàng sau khi cố gắng đăng nhập để xác minh tài khoản theo yêu cầu của tội phạm mạo danh (Thuần & Hồng, 2021).

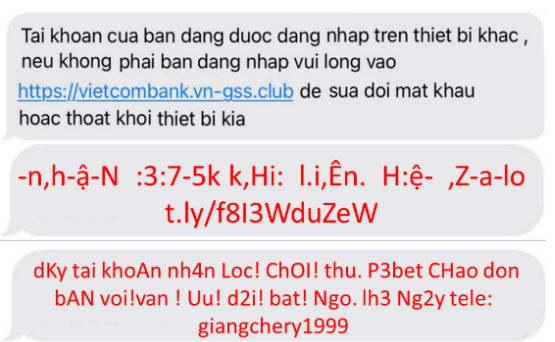
2.5. Những vấn đề còn tồn tại trong bài toán phát hiện tin nhắn rác

Qua kết quả nghiên cứu khảo sát được trình bày trong Bảng 2 và 3 ở trên, có thể dễ dàng thấy rằng hầu hết những giải pháp hiện có được đề xuất cho bài toán phát hiện tin nhắn rác đạt kết quả chính xác rất cao trên 90%, thậm chí những giải pháp điển hình được đề xuất trong các công trình (Ghourabi và cộng sự, 2020; HUSSEIN và cộng sự, 2023; MAMBINA và cộng sự, 2024; Maqsood và cộng sự, 2023; Roy và cộng sự, 2020; Srinivasarao & Sharaff, 2023; Tuan và cộng sự, 2022) còn có độ chính xác tiệm cận tới 100%. Người dùng di động cũng đã được cung cấp nhiều giải pháp khác nhau trong việc lọc tin nhắn rác. Tuy nhiên, những số liệu được quan sát và thu thập ở mục 2.4 lại phản ánh thực tế rằng vấn nạn tin nhắn rác và mức độ thiệt hại của người dùng di động không hề có dấu hiệu thuyên giảm. Như vậy, vấn đề phát hiện và ngăn chặn tin nhắn rác trong thực tế còn đối mặt với nhiều thách thức tiềm ẩn. Trong phần này, nhóm nghiên cứu sẽ chỉ ra những vấn đề còn tồn tại trong bài toán phát hiện tin nhắn rác.

Một là, vấn đề về bộ dữ liệu tin nhắn

Có thể dễ dàng thấy rằng, hầu hết những nghiên cứu điển hình chỉ sử dụng lặp đi lặp lại một số ít các bộ dữ liệu tin nhắn công khai (chủ yếu là tin nhắn tiếng Anh) và gần như không có những đóng góp mới về dữ liệu nghiên cứu trong vài năm trở lại đây.

Phân tích các bộ dữ liệu công khai được dùng trong những nghiên cứu điển hình ở trên, hầu hết các bộ dữ liệu tin nhắn này đều bị hạn chế về mặt số lượng tin nhắn (SALMAN và cộng sự, 2024) và thiếu cập nhật, trong khi đó tin tặc lại thường xuyên



Nguồn: Nhóm nghiên cứu tổng hợp

Hình 7. Một số thủ đoạn của tin tặc để tránh bộ lọc tin nhắn rác phát hiện

sử dụng nhiều thủ đoạn khác nhau nhằm thay đổi nhiều mẫu tin nhắn rác để tránh bị phát hiện.

Hai là, vấn đề về bài toán phát hiện tin nhắn rác và mô hình học máy

Hầu hết người dùng chưa ý thức được mức độ nguy hại của nạn tin nhắn rác và thực tế cho thấy 35% người dân Mỹ vẫn tò mò về tin nhắn rác được gửi tới từ số điện thoại lạ (theo kết quả khảo sát của Truecaller năm 2022 (Truecaller, 2022)). Chính vì điều này nên những người dùng di động ở Mỹ đã mất 330 triệu USD vì tin nhắn lừa đảo vào năm 2022, tăng từ 131 triệu USD vào năm 2021 và 86 triệu USD vào năm 2020. Do đó, những công cụ/ứng dụng chỉ mang tính chất cảnh báo tin nhắn *"thường"* hay *"rác"* là chưa đủ.

Ba là, vấn đề phát hành ứng dụng phát hiện tin nhắn rác và hoạt động tuyên truyền cộng đồng

Như đã đề cập ở trên, một số nhà mạng di động và hãng sản xuất điện thoại thông minh cũng đã triển khai công cụ cảnh báo tin nhắn rác cho người dùng, tuy nhiên những giải pháp này chủ yếu dựa trên luật và dừng lại ở mức độ cơ bản. Người dùng di động muốn được cảnh báo và bảo vệ trước nạn tin nhắn rác cần phải mua những phần mềm bản quyền thương mại như Robokiller, Truecaller. Đây có thể là một trong số những rào cản khiến cho công cuộc ngăn chặn vấn nạn tin nhắn rác di động chưa đạt được kết quả như kỳ vọng.

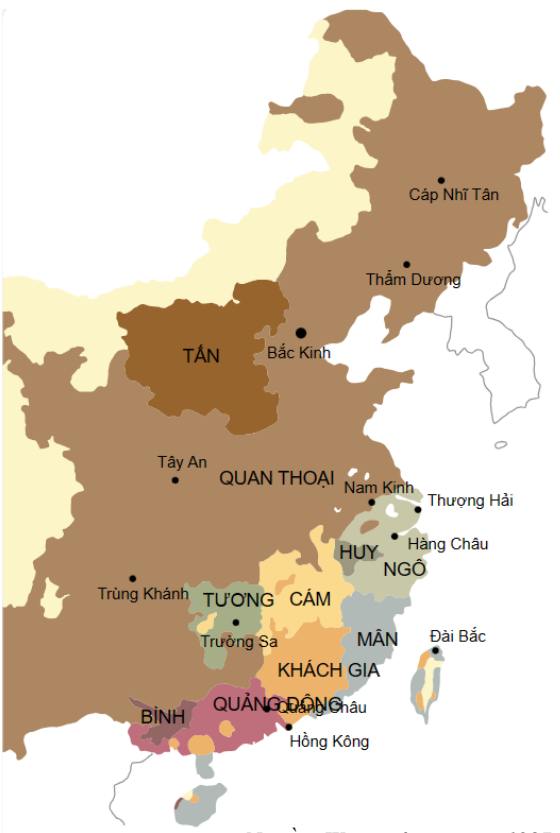
Bên cạnh đó, hoạt động tuyên truyền và hướng dẫn người dùng ứng phó với nạn tin nhắn/cuộc gọi rác ở các quốc gia chưa thực sự hiệu quả, bởi lẽ tại một đất nước phát triển như Mỹ mà vẫn có tới 35% dân số vẫn loay hoay hoặc tò mò nhấn vào những tin nhắn rác họ nhận được từ số điện thoại lạ (Truecaller, 2022).

3. Một số khuyến nghị cho các vấn đề còn tồn tại của bài toán phát hiện tin nhắn rác trên thiết bị di động

Trong phần này, bài báo đóng góp những khuyến nghị quan trọng giúp mang những giải pháp lý thuyết gần hơn với thực tế để có thể giảm thiểu tối đa những thiệt hại tài chính mà người dùng di động đang gặp phải.

3.1. Khuyến nghị đối với bộ dữ liệu tin nhắn

Đối với lĩnh vực trí tuệ nhân tạo và học máy nói chung, bộ dữ liệu đóng vai trò trọng yếu trong việc huấn luyện ra các mô hình thông minh. Theo Philip Russom-Ggiám đốc của tổ chức TDWI chuyên đào tạo và nghiên cứu chuyên sâu về quản



Nguồn: Wurm và cộng sự, 1987

Hình 8.
Phân bố của các phân chi tiếng Trung

trị & phân tích dữ liệu và trí tuệ nhân tạo (Russom, 2018), các yêu cầu đối với bộ dữ liệu dùng cho học máy và trí tuệ nhân tạo là: *đủ lớn về số lượng, đa dạng về nội dung, và luôn được cập nhật.* Vì thế, để xây dựng ra những mô hình lọc tin nhắn rác mạnh mẽ, các quốc gia cần đẩy mạnh việc đóng góp những bộ dữ liệu tin nhắn công khai và được cập nhật. Tất nhiên, họ cũng cần chú trọng để tránh vô tình làm lộ, lọt thông tin riêng tư, nhạy cảm có trong tin nhắn của người dùng.

Đối với những ngôn ngữ được sử dụng bởi nhiều đối tượng khác nhau, ví dụ như tiếng Trung Quốc, ngoài tiếng phổ thông là ngôn ngữ tiêu chuẩn của Quốc gia tỷ dân này và được sử dụng phổ biến ở các thành phố lớn thì tiếng Trung còn nhiều biến thể khác nhau như tiếng Khách Gia, tiếng Quảng Đông, tiếng Tứ Xuyên… (chi tiết xem Hình 8). Do đó, những bộ dữ liệu tin nhắn di động được thu thập mang tính cá nhân tự phát thì rất khó có thể mang tính khái quát và đa dạng vùng miền được.

Rõ ràng, việc thu thập bộ dữ liệu tin nhắn SMS đủ lớn, đa dạng biến thể cho từng ngôn ngữ để xây dựng được mô hình phổ quát phát hiện tin nhắn rác là rất cần thiết. Điều này chỉ có thể thực hiện được dưới sự hỗ trợ của cơ quan quản lý nhà nước và các nhà mạng cung cấp dịch vụ di động. Tuy nhiên, chúng tôi cũng cần nhấn mạnh lại rằng quá trình thu thập và xử lý dữ liệu cần phải tuân thủ pháp luật của Quốc gia đó và ràng buộc quyền riêng tư của người dùng.

Bên cạnh đó, một khâu rất quan trọng nữa mang yếu tố quyết định đến chất lượng phân loại dữ liệu văn bản nói chung và phát hiện tin nhắn rác nói riêng đó là *"tiền xử lý dữ liệu"*. Về bản chất, để thực hiện các kỹ thuật phân loại văn bản, trước tiên nội dung văn bản thường được chuyển đổi bằng cách vector hóa sử dụng phương pháp tiếp cận dựa trên tần số hoặc dựa trên mô hình. Các công trình (Ghourabi & Alohaly, 2023; LIU và cộng sự, 2021) đã chứng minh rằng những phương pháp tiền xử lý dữ liệu đem lại kết quả phân lớp khác nhau, và đặc biệt là bằng cách tiền xử lý dữ liệu hợp lý có thể giúp cho kỹ thuật phân lớp cơ bản như KNN đạt được độ chính xác lên đến gần 100%. Do đó, các nhà nghiên cứu và phát triển giải pháp cần thử nghiệm những phương pháp tiền xử lý dữ liệu khác nhau để lựa chọn ra phương án tối ưu nhất.

3.2. Vấn đề về bài toán lọc tin nhắn rác và kỹ thuật học máy áp dụng

Như đã phân tích, việc cảnh báo/lọc tin nhắn *"thường"* hay *"rác"* chưa phát huy được tác dụng trong thực tế. Do đó, cần phải phát triển những mô hình phát hiện tin nhắn rác nâng cao có thể cảnh báo người dùng di động ở nhiều cấp độ khác nhau, đặc biệt là có khả năng nhận diện những tin nhắn tiềm ẩn rủi ro lớn về lĩnh vực tài chính-ngân hàng hoặc lừa đảo.

Đối với vấn đề lựa chọn kỹ thuật học máy, qua những kết quả nghiên cứu khảo sát của bài báo này và những công trình liên quan (Maqsood và cộng sự, 2023; SALMAN và cộng sự, 2024), có thể thấy rằng đã có rất nhiều kỹ thuật học máy như Naïve Bayes, KNN, SVM, cây quyết định, mạng Markov ẩn, CNN, LSTM được ứng dụng và không có kỹ thuật nào thể hiện sự hiệu quả vượt trội hơn hẳn khi được huấn luyện trên các bộ dữ liệu tin nhắn SMS khác nhau. Tuy nhiên, kết quả nghiên cứu của những công trình liên quan cũng đã cho thấy nhóm kỹ thuật học sâu thường đem lại độ chính xác cao hơn so với những kỹ thuật học máy còn lại. Vì vậy, khi xây dựng mô hình phát hiện tin nhắn rác, các nhà nghiên cứu cần thử nghiệm với nhiều kỹ thuật khác nhau và áp dụng nhiều phương pháp tiền xử lý dữ liệu

khác nhau (như đã nói đến ở mục 3.1) để đạt được kết quả khả dĩ nhất.

3.3. Vấn đề phát hành ứng dụng chặn tin nhắn rác tự động miễn phí và hoạt động tuyên truyền cộng đồng

Để bảo vệ tốt nhất cho người dùng di động, cộng đồng rất cần tới những phần mềm phát hiện tin nhắn rác tự động, mạnh mẽ và miễn phí (có sự kết hợp của công nghệ tự động hoá quy trình bằng Robot (RPA), học máy (ML), xử lý ngôn ngữ tự nhiên (NLP), v.v…) được phát hành bởi các nhà phát triển ứng dụng đáng tin cậy, cơ sở giáo dục đại học uy tín.

Bên cạnh đó, mỗi quốc gia cũng cần đẩy mạnh tuyên truyền cộng đồng nhằm nâng cao nhận thức của người dân về sự nguy hại và những rủi ro có thể gặp phải của tin nhắn rác.

4. Kết luận

Bài báo này đã tiến hành khảo sát, tổng hợp thực trạng nghiên cứu phát triển các giải pháp cho bài toán phát hiện tin nhắn rác trên thiết bị di động và phân tích những vấn đề còn tồn tại với bài toán này. Dựa trên những kết quả khảo sát, phân tích, tổng hợp trên, bài báo đã đưa ra những khuyến nghị quan trọng hỗ trợ cho cộng đồng nghiên cứu và các nhà phát triển có thể tạo ra những giải pháp phát hiện tin nhắn rác di động mạnh mẽ, hiệu quả và giúp cho người dùng di động có thể giảm thiểu tối đa những thiệt hại mà vấn nạn tin nhắn rác gây ra. Bài viết đề xuất một số hướng nghiên cứu có ý nghĩa nên thực hiện tiếp trong thời gian tới như phát triển mô hình nâng cao có khả năng nhận diện những tin nhắn rác tiềm ẩn rủi ro lớn, tăng cường tự động hoá nhận diện tin nhắn rác dựa trên một số công nghệ tiên tiến thông qua áp dụng RPA, mở rộng lĩnh vực áp dụng giải pháp, mở rộng ngôn ngữ áp dụng giải pháp (đặc biệt ngôn ngữ tiếng Anh và tiếng Việt). ■

Tài liệu tham khảo

Agarwal, S., Kaur, S., & Garhwal, S. (2016). SMS spam detection for Indian messages. International Conference on Next Generation Computing Technologies (NGCT). https://doi.org/10.1109/NGCT.2015.7375198

Almeida, T. A., J.M.G, H., & A, Y. (2011). Contributions to the study of SMS spam fltering: New collection and results. Proceedings of the 11th ACM symposium on Document engineering, 259–262. https://doi.org/10.1145/2034691.2034742

An, N. (2024). Tội phạm phát tán cả 100.000 tin nhắn rác mỗi ngày để lừa đảo, kích động bạo loạn. https://tuoitre.vn/toi-pham-phat-tan-ca-100-000-tin-nhan-rac-moi-ngay-de-lua-dao-kich-dong-bao-loan-20240117101737068.htm

Arifin, D. D., Shaufiah, & Bijaksana, Moch. A. (2016). Enhancing spam detection on mobile phone Short Message Service (SMS) performance using FP-growth and Naive Bayes Classifier. 2016 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob). https://doi.org/10.1109/APWiMob.2016.7811442

Ayaz, M., Nizamani, S., Chandio, A. A., & Kumar Luhana, K. (2024). Detection of Roman Urdu fraud/spam SMS in Pakistan Using Machine Learning. International Journal of Computing and Digital Systems, 15(1), 1053–1061. http://dx.doi.org/10.12785/ijcds/150174

Bình, A. (2017). Xung quanh vụ hơn 46 triệu thuê bao di động Malaysia rò rỉ dữ liệu. Báo Điện tử Chính phủ. https://baochinhphu.vn/xung-quanh-vu-hon-46-trieu-thue-bao-di-dong-malaysia-ro-ri-du-lieu-102229529.htm

Chen, T., & Kan, M.-Y. (2012). Creating a live, public short message service corpus: The NUS SMS corpus. Language Resources and Evaluation, 47, 299–335. https://doi.org/10.1007/s10579-012-9197-9

Chi, Q. (2024). Rò rỉ dữ liệu cá nhân của 73 triệu tài khoản di động ở Mỹ. VTV online. https://vtv.vn/the-gioi/ro-ri-du-lieu-ca-nhan-cua-73-trieu-tai-khoan-di-dong-o-my-20240401112854258.htm

Degenhar, J. (2024, Tháng Ba 28). Number of smartphone users worldwide from 2014 to 2029. Number of smartphone users worldwide from 2014 to 2029. https://www.statista.com/forecasts/1143723/smartphone-users-in-the-world

Fact.MR. (2021). People Finance Mobile App Market. https://www.factmr.com/report/personal-finance-mobile-app-

market

Gajić, A. (2022). Spam Statistics. Spam Statistics. https://99firms.com/blog/spam-statistics/#gref

Ghourabi, A., A. Mahmood, M., & M. Alzubi, Q. (2020). A Hybrid CNN-LSTM Model for SMS Spam Detection in Arabic and English Messages. Future Internet, 12(156). https://doi.org/10.3390/fi12090156

Ghourabi, A., & Alohaly, M. (2023). Enhancing Spam Message Classification and Detection Using Transformer-Based Embedding and Ensemble Learning. Sensors, 23. https://doi.org/10.3390/ s23083861

Giri, S., Das, S., Das, S. B., & Banerjee, S. (2023). SMS Spam Classification–Simple Deep Learning Models With Higher Accuracy Using BUNOW And GloVe Word Embedding. Journal of Applied Science and Engineering, 26(10). https://doi.org/10.6180/jase.202310_26(10).0015

Hidalgo, J. M. G., Bringas, G. C., & Sánz, E. P. (2006). Content based SMS spam filtering. Proceedings of the 2006 ACM symposium on Document engineering, 107–114. https://doi.org/10.1145/1166160.1166191

Hikmaturokhman, A., Nafi'ah, H., Larasati, S., Wahyudin, A., Ariprawira, G., & Pramono, S. (2022). Deep Learning Algorithm Models for Spam Identification on Cellular Short Message Service. Journal of Communications, 17(9), 769–776. https://doi.org/10.12720/jcm.17.9.769-776

Hsu, B.-M. (2020). Comparison of Supervised Classification Models on Textual Data. Mathematics, 8(5). https://doi.org/10.3390/math8050851

Husein, A. A.-K., Mohammad-Reza, F.-D., & Saeid, P. (2023). Multi-Type Feature Extraction and Early Fusion Framework for SMS Spam Detection. IEEE Access, 11, 123756–123765. https://doi.org/10.1109/ACCESS.2023.3327897

Jain, A. K., & Gupta, B. B. (2019). Feature Based Approach for Detection of Smishing Messages in the Mobile Environment. Journal of Information Technology Research, 12(2), 17–35. https://doi.org/10.4018/JITR.2019040102

Khristopher, J. B. (2023). This is America's most common text-messaging scam, FTC says. https://www.cbsnews.com/news/text-message-scam-impersonating-bank-ftc/

Laborde, S. (2024). 60+ Smishing Statistics in 2024 (SMS Phishing Attacks). https://techreport.com/statistics/cybersecurity/smishing-statistics/

Liu, X., Lu, H., & Nayak, A. (2021). A Spam Transformer Model for SMS Spam Detection. IEEE Access, 9, 80253–80263. https://doi.org/10.1109/ACCESS.2021.3081479

M. Hameed, S., & Hussein Ali, Z. (2021). SMS Spam Detection Based on Fuzzy Rules and Binary Particle Swarm Optimization. International Journal of Intelligent Engineering and Systems, 14(2), 314–322. https://doi.org/10.22266/ijies2021.0430.28

MAMBINA, I. S., NDIBWILE, J. D., UWIMPUHWE, D., & MICHAEL, K. F. (2024). Uncovering SMS Spam in Swahili Text Using Deep Learning Approaches. IEEE Access, 12, 25164–25175. https://doi.org/10.1109/ACCESS.2024.3365193

Maqsood, U., Ur Rehman, S., Ali, T., Mahmood, K., Alsaedi, T., & Kundi, M. (2023). An Intelligent Framework Based on Deep Learning for SMS and e-mail Spam Detection. Applied Computational Intelligence and Soft Computing, 2023, 1–16. https://doi.org/10.1155/2023/6648970

Mishra, S., & Soni, D. (2020). Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis. Future Generation Computer Systems, 108, 803–815. https://doi.org/10.1016/j.future.2020.03.021

Osa, E., & Elaigwu, V. O. (2021). Modelling of a Deep Learning Based SMS Spam Detection Application. NIPES Journal of Science and Technology Research, 3(4), 163–173. https://doi.org/10.37933/nipes/3.4.2021.17

Robokiller. (2022). The Robokiller phone scam report 2022 insights & analysis. https://www.robokiller.com/robokiller-2022-phone-scam-report

Roy, P. K., Singh, J. P., & Banerjee, S. (2020). Deep learning to filter SMS Spam. Future Generation Computer Systems, 102, 524–533. https://doi.org/10.1016/j.future.2019.09.001

Russom, P. (2018). The Automation and Optimization of Advanced Analytics Based on Machine Learning. https://www.qubole.com/wp-content/uploads/2021/03/TDWI-Checklist-The-Automation-and-Optimization-of-Advanced-Analytics-Based-on-ML.pdf

Salman, M., Ikram, M., & Ali Kaafar, M. (2024). Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models. IEEE Access, 12, 24306–24324. https://doi.org/10.1109/ACCESS.2024.3364671

Sjarif, N. N. A., Azmi, N. F. M., Chuprat, S., Sarkan, H. M., Yahya, Y., & Sam, S. M. (2019). SMS Spam Message Detection using Term Frequency-Inverse Document Frequency and Random Forest Algorithm. Procedia Computer Science, 161, 509–515. https://doi.org/10.1016/j.procs.2019.11.150

Sonowal, G., & Kuppusamy, K. S. (2018). SmiDCA: An Anti-Smishing Model with Machine Learning Approach. The Computer Journal, 1143–1157. https://doi.org/10.1093/comjnl/bxy039

Sousa, G. J. de, Pedronette, D. C. G. ´ aes, Papa, J. P., & Guilherme, I. R. (2021). SMS Spam Detection Through Skip-gram Embeddings and Shallow Networks. 4193–4201. https://doi.org/10.18653/v1/2021.findings-acl.367

Srinivasarao, U., & Sharaff, A. (2023). Machine intelligence based hybrid classifier for spam detection and sentiment analysis of SMS messages. Multimedia Tools and Applications, 82, 31069–31099. https://doi.org/10.1007/s11042-023-14641-5

Tagg, C. (2009). A corpus linguistics study of SMS text messaging [University of Birmingham]. https://etheses.bham.ac.uk/id/eprint/253/1/Tagg09PhD.pdf

Tang, S., Mi, X., Li, Y., Wang, X., & Chen, K. (2022). Clues in tweets: Twitterguided discovery and analysis of SMS spam. Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, 2751–2764. https://doi.org/10.1145/3548606.3559351

Truecaller. (2022). Truecaller Insights 2022 U.S. Spam & Scam Report. https://www.truecaller.com/blog/insights/truecaller-insights-2022-us-spam-scam-report

Tuan, V. M., Thang, N. X., & Anh, T. Q. (2022). Vietnamese SMS spam detection with deep learning and pretrained language model. Journal of Science and Technology on Information and Communications, 1(2). https://jstic.ptit.edu.vn/jstic-ptit/index.php/jstic/article/view/484

University of California Irvine. (2012). SMS Spam Collection Dataset [dataset]. https://archive.ics.uci.edu/dataset/228/sms+spam+collection

Wurm, S. A., Li, R., Baumann, T., & Lee, M. W. (1987). Language Atlas of China. Longman.

Xia, T., & Chen, X. (2020). A Discrete Hidden Markov Model for SMS Spam Detection. Applied sciences, 10(5011). https://doi.org/10.3390/app10145011

Xia, T., & Chen, X. (2021). A weighted feature enhanced Hidden Markov Model for spam SMS filtering. Neurocomputing, 444, 48–58. https://doi.org/10.1016/j.neucom.2021.02.075