PHÁT HIỆN GIAN LẬN GIAO DỊCH THƯƠNG MẠI ĐIỆN TỬ: TIẾP CẬN HỌC MÁY VÀ HÀM Ý QUẢN TRỊ

TS. Nguyễn Anh Tú*

Trong bối cảnh thương mại điện tử phát triển nhanh, phát hiện gian lận giao dịch trở thành thách thức quan trọng trong quản trị rủi ro số. Nghiên cứu này đánh giá hiệu quả của các thuật toán học máy trong nhận diện gian lận trên bộ dữ liệu 299.695 quan sát. Ba mô hình gồm Hồi quy Logistic, Rừng ngẫu nhiên và XGBoost được huấn luyện, kiểm định qua 30 vòng lặp nhằm bảo đảm độ tin cậy. Kết quả cho thấy các mô hình phi tuyến vượt trội, trong đó XGBoost đạt hiệu năng tốt nhất với AUC 0,97 và độ nhạy 88,8%, giúp phát hiện gian lận hiệu quả đồng thời kiểm soát báo động giả. Nghiên cứu cũng chỉ ra các biến ngữ cảnh và xác thực có giá trị dự báo cao hơn biến số tiền giao dịch, qua đó cung cấp cơ sở cho doanh nghiệp TMĐT tối ưu chiến lược khai thác dữ liệu và bảo vệ niềm tin khách hàng.

• Từ khóa: *dự đoán gian lận; học máy; phân loại; rừng ngẫu nhiên; thương mại điện tử; xgboost.*

In the context of rapidly growing e-commerce, transaction fraud detection has become a critical challenge in digital risk management. This study evaluates the performance of machine learning algorithms in identifying fraudulent transactions using a dataset of 299,695 observations. Three models, including Logistic Regression, Random Forest, and XGBoost, were trained and tested over 30 iterations to ensure reliability. The results indicate that nonlinear models outperform the linear model, with XGBoost achieving the best overall performance, with an AUC of 0.97 and the highest sensitivity of 88.8%, enabling effective fraud detection while controlling false alarms. The study also reveals that contextual and authentication variables have greater predictive value than conventional transaction amount variables. These findings provide practical implications for e-commerce businesses in optimizing data-driven fraud prevention strategies and maintaining customer trust.

• Key words: *e-commerce; fraud detection; machine learning; classification; random forest; xgboost.*

Ngày nhận bài: 23/3/2026
Ngày gửi phản biện: 27/4/2026
Ngày duyệt đăng: 13/5/2026
DOI: https://doi.org/10.71374/jfarv.v26.i313.48

1. Giới thiệu

Phát hiện gian lận trong giao dịch thương mại điện tử (TMĐT) đang trở thành chủ đề quan trọng trong quản trị rủi ro số do sự gia tăng mạnh của giao dịch trực tuyến kéo theo nguy cơ chiếm đoạt thông tin và lạm dụng tài khoản. Gian lận không chỉ gây thiệt hại tài chính mà còn làm suy giảm niềm tin của người dùng, vì vậy nhu cầu về các hệ thống phát hiện tự động, gần thời gian thực ngày càng cấp thiết (Zhang et al., 2025). Về phương pháp, các hệ thống truyền thống chủ yếu dựa vào luật thủ công và đặc trưng giao dịch như số tiền, thời điểm, thiết bị hay địa chỉ giao hàng. Tuy nhiên, các yếu tố này dễ bị giả mạo, làm giảm hiệu quả mô hình. Vì thế, xu hướng mới là khai thác dữ liệu hành vi người dùng như chuỗi thao tác, thời gian dừng trang và sinh trắc học hành vi nhằm tăng độ chính xác và khả năng thích nghi (Zhang et al., 2025). Nghiên cứu hiện nay phần lớn tập trung vào gian lận thanh toán trực tuyến, trong khi phát hiện theo thời gian thực còn hạn chế. Các mô hình học máy như mạng nơ-ron, K-láng giềng gần nhất và học kết hợp được sử dụng rộng rãi, nhưng bài toán mất cân bằng dữ liệu vẫn là thách thức lớn (Rodrigues et al., 2022). Việc kết hợp dữ liệu giao dịch với dữ liệu hành vi đa phương thức được xem là hướng đi hiệu quả, giúp nâng cao khả năng phát hiện với độ trễ thấp (Zeng et al., 2025). Từ góc độ học thuật, đây là bài toán phân loại bất cân bằng và biến động theo thời gian, đòi hỏi sự kết hợp giữa học máy, khai phá hành vi và thiết kế hệ thống, đồng thời có giá trị thực tiễn cao đối với an toàn giao dịch và quản trị rủi ro số (Carta et al., 2019).

2. Lược khảo các nghiên cứu liên quan

Các nghiên cứu gần đây cho thấy gian lận trong TMĐT không thể giải thích bằng một biến đơn lẻ mà là kết quả của sự kết hợp giữa động cơ, cơ hội và khả năng che giấu hành vi. Khung tam giác

* *Viện Nghiên cứu khoa học và Công nghệ ngân hàng, Trường Đại học Ngân hàng TP.HCM; email: tuna@hub.edu.vn*

gian lận với ba thành tố áp lực, cơ hội và hợp lý hóa vẫn là cơ sở phổ biến để lý giải gian lận, dù còn hạn chế trong môi trường số hóa (Homer, 2020). Trong TMĐT, các tín hiệu như uy tín người bán hay hình ảnh hồ sơ ảnh hưởng mạnh đến nhận thức rủi ro và quyết định của người mua, từ đó tạo điều kiện cho các hành vi gian lận dựa trên niềm tin (Dai et al., 2018). Đồng thời, các thông tin định danh như thiết bị, email hay địa chỉ có thể bị giả mạo, khiến tín hiệu tĩnh kém hiệu quả; vì vậy dữ liệu hành vi như cú nhấp chuột, chuỗi xem trang được xem là nguồn nhận diện đáng tin cậy hơn (Zhang et al., 2025).

Các phương pháp truyền thống trong phát hiện gian lận chủ yếu gồm quy tắc thủ công, kiểm tra chuyên gia và hồi quy logistic. Ali et al. (2022) cho rằng các cách tiếp cận này thường tốn kém, thiếu linh hoạt trước dữ liệu lớn và biến động liên tục. Hồi quy logistic có ưu điểm về khả năng giải thích và ước lượng xác suất gian lận, nhưng hạn chế ở giả định tuyến tính và khó nắm bắt tương tác phức tạp giữa hành vi người dùng, ngữ cảnh giao dịch và mạng lưới liên kết. Trong môi trường dữ liệu mất cân bằng cao, các mô hình này cũng dễ thiên lệch về lớp đa số (Ali et al., 2022; Zhang et al., 2025).

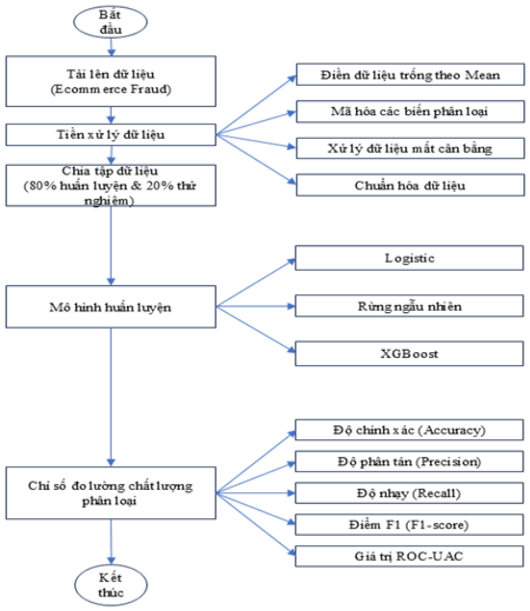
Sự phát triển của học máy đánh dấu chuyển dịch từ phát hiện dựa trên quy tắc sang dựa trên dữ liệu. Mutemi and Bacao (2024) cho thấy SVM và ANN là những thuật toán phổ biến, trong khi các thách thức chính vẫn là dữ liệu thực, mất cân bằng và khả năng khái quát hóa. Học máy có ưu thế nhờ khai thác đồng thời dữ liệu giao dịch, hành vi, văn bản, hình ảnh và quan hệ giữa các thực thể. Các nghiên cứu cũng chứng minh mô hình đa phương thức nâng cao hiệu quả phát hiện (Zhang et al., 2025), đặc trưng nội dung giúp nhận diện website TMĐT giả mạo (Sánchez-Paniagua, Fidalgo, Alegre, & Jáñez-Martino, 2025), còn mô hình tổ hợp cải thiện hiệu quả trong dữ liệu nhiễu và lệch lớp (Carta et al., 2019). Ngoài ra, các hệ thống như SCARFF cho thấy phát hiện gian lận cần gắn với xử lý dữ liệu luồng thời gian thực để triển khai quy mô lớn (Kodate, Chiba, Kimura, & Masuda, 2020). Vì vậy, học máy hiện được xem là hướng tiếp cận hiệu quả hơn kinh tế lượng truyền thống trong phát hiện gian lận TMĐT (Mutemi & Bacao, 2024).

3. Phương pháp nghiên cứu

3.1. Quy trình nghiên cứu

Hình 1 mô tả quy trình nghiên cứu trong bài viết này. Trong đó, nghiên cứu nhấn mạnh tới việc xử lý tiền dữ liệu với việc kiểm tra bộ dữ liệu và làm sạch bộ dữ liệu trước khi đưa vào các mô hình huấn luyện. Sau đó, nghiên cứu áp dụng hai phương pháp học kết hợp là Rừng ngẫu nhiên (RF) và XGBoost để áp dụng cho bài toán phân loại học có giám sát. Cả 2 thuật toán này đều được so sánh cùng phương pháp cơ sở là hồi quy logistic để cho thấy tính tối ưu của 2 thuật toán đề xuất (Ali et al., 2022). Sau cùng, để đánh giá kết quả, nghiên cứu áp dụng các chỉ số đánh giá: Độ chính xác (Accuracy), Độ phân tán (Precision), Độ nhạy (Recall), Điểm số F1 (F1-score), và Giá trị AUC (AUC) nhằm so sánh các mô hình Logistic, RF và XGBoost về hiệu quả phân loại.

Hình 1. Quy trình nghiên cứu



3.2. Mô tả bộ dữ liệu

Nghiên cứu sử dụng bộ dữ liệu thứ cấp Ecommerce Fraud từ Kaggle với 299.695 quan sát, phản ánh đặc trưng giao dịch TMĐT và trạng thái gian lận thực tế tại nhiều quốc gia.

Biến phụ thuộc là is_fraud, cho biết giao dịch có gian lận hay không. Khoảng 2,2% giao dịch được gắn nhãn gian lận, cho thấy dữ liệu mất cân bằng lớp và cần xử lý để tránh thiên lệch dự báo.

Các biến giải thích gồm bốn nhóm chính: (i) định danh giao dịch/người dùng như *transaction_*

id, *user_id*; (ii) lịch sử và hành vi người dùng như *account_age_days*, *total_transactions_user*, *avg_amount_user*; (iii) đặc điểm giao dịch hiện tại như *amount*, *transaction_time*; và (iv) ngữ cảnh giao dịch như *country*, *bin_country*, *channel*, *merchant_category*, *cvv_result*. Những biến này giúp nhận diện các dấu hiệu bất thường về chi tiêu, thời gian, địa lý và xác thực thanh toán. Việc mô tả chi tiết 15 biến giải thích và 01 biến phụ thuộc được thể hiện ở dữ liệu gốc tại https://www.kaggle.com/datasets/umuttuygurr/e-commerce-fraud-detection-dataset.

3.3. Tiền xử lý dữ liệu

Các biến số như *amount*, *avg_amount_user*, *account_age_days*, *total_transactions_user* và *shipping_distance_km* được giữ nguyên giá trị gốc và chuẩn hóa. Các biến nhị phân như *channel*, *promo_used*, *avs_match*, *cvv_result* và *three_ds_flag* được giữ nguyên dạng ban đầu. Sau đó, toàn bộ dữ liệu được chuyển thành ma trận số để huấn luyện mô hình.

Để xử lý mất cân bằng dữ liệu, nghiên cứu áp dụng kỹ thuật **SMOTE** trên tập huấn luyện nhằm tăng đại diện cho lớp gian lận. Phương pháp này tạo mẫu tổng hợp từ các điểm lân cận, giúp mô hình xác định ranh giới phân loại rõ hơn và cải thiện khả năng nhận diện giao dịch gian lận (Seo & Kim, 2018). Dữ liệu được chia theo tỷ lệ 80% huấn luyện và 20% kiểm tra, với *random_state = 42* để đảm bảo khả năng tái lập.

3.4. Mô hình huấn luyện và chỉ số đo lường chất lượng phân loại

Nghiên cứu này sử dụng 3 thuật toán bao gồm: hồi quy logistic, rừng ngẫu nhiên (RF) và Extreme Gradient Boosting (XGBoost) nhằm so sánh chất lượng bài toán phân loại. Dữ liệu được chia theo tỷ lệ 80% cho tập huấn luyện và 20% cho tập kiểm tra và cài đặt "random_state=42" để đảm bảo khả năng tái lập. Nghiên cứu áp dụng các chỉ số đánh giá trên tập kiểm tra bao gồm Độ chính xác (Accuracy), Độ phân tán (Precision), Độ nhạy (Recall), Điểm số F1 (F1-score), và Giá trị AUC (AUC) nhằm so sánh chất lượng phân loại các mô hình Logistic, RF và XGBoost (Lu, Wang, & Nguyen, 2023; Nguyen, Lu, & Nguyen, 2021; Tu, Khoa, Nhan, & Le Kieu Oanh, 2024). Lưu ý rằng để các giá trị có giá trị thống kê, cả 3 thuật toán đều được đưa vào vòng lặp 30 lần và lấy giá trị trung bình của các chỉ số đánh giá. Toàn bộ các thuật toán đều được phát triển bằng mã Python chạy trên nền tảng Google Colab.

4. Kết quả thực nghiệm

4.1. Thống kê mô tả các nhóm biến trong mô hình

Bảng 1 trình bày thống kê mô tả của 299.695 giao dịch TMĐT, trong đó tỷ lệ gian lận "*is_fraud*" khoảng 2,21%, cho thấy dữ liệu mất cân bằng rõ rệt giữa giao dịch gian lận và hợp lệ.

Về biến định lượng, *account_age_days* trung bình 973 ngày, cho thấy đa số tài khoản đã tồn tại khá lâu; *total_transactions_user* đạt 50,67 giao dịch, phản ánh mức độ hoạt động tương đối ổn định. Các biến *avg_amount_user* và *amount* có giá trị trung bình cao hơn trung vị, cho thấy phân phối lệch phải và tồn tại một số giao dịch giá trị lớn. Tương tự, *shipping_distance_km* có biến động mạnh với một số giá trị ngoại lai.

Bảng 1. Thống kê mô tả các biến trong mô hình

	Account_ age _days	Total_ transactions _user	Avg_ amount_ user	Amount	Promo	Avs_ match	Cvv_ result	Three_ ds_ flag	Distance	Fraud
Tổng	299695	299695	299695	299695	299695	299695	299695	299695	299695	299695
Trung bình	973	50.67	148	177	0.153	0.84	0.87	0.78	357	0.02
Độ lệch chuẩn	525	5.97	200	306	0.36	0.37	0.33	0.41	427	0.15
Nhỏ nhất	1	40	3.52	1	0	0	0	0	0	0
25%	516	46	46.19	42.1	0	1	1	1	136	0
50%	975	51	90.13	89.99	0	1	1	1	273	0
75%	1425	56	173	191	0	1	1	1	409	0
Lớn nhất	1890	60	4565	16994	1	1	1	1	3748	1

Nguồn: Tác giả tổng hợp

Đối với biến xác thực, *promo_used* đạt 15,36%, còn *cvv_result* là 78,46%, cho thấy phần lớn giao dịch vượt qua bước kiểm tra cơ bản nhưng vẫn có tỷ lệ đáng kể không khớp. Nhìn chung, rủi ro gian lận có xu hướng gắn với các giao dịch khác thường về giá trị, khoảng cách giao hàng và trạng thái xác thực.

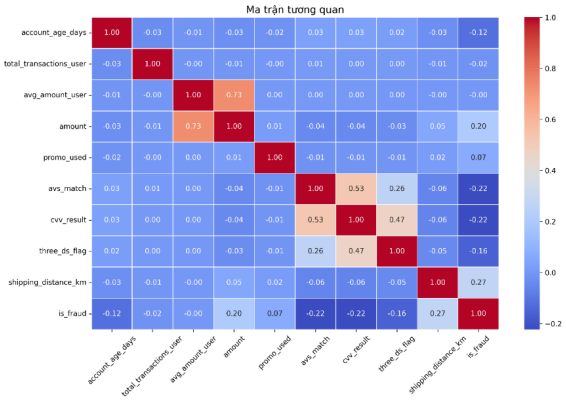
4.2. Phân tích tương quan giữa các biến

Hình 2 cho thấy hầu hết các cặp biến có hệ số tương quan thấp, không xuất hiện đa cộng tuyến nghiêm trọng khi tất cả hệ số đều nhỏ hơn 0,8 (Gujarati, 2004). Điều này cho thấy các biến phản ánh những khía cạnh khác nhau của hành vi giao dịch.

Các biến xác thực như *avs_match*, *cvv_result* và *three_ds_flag* có tương quan dương mức trung bình (r từ 0,26 đến 0,53), cho thấy các cơ chế này cùng phản ánh mức độ tin cậy của giao dịch và có tính bổ trợ lẫn nhau.

Đối với biến mục tiêu *is_fraud*, *distance_km* (r = 0,27) và *amount* (r = 0,20) tương quan dương, hàm ý giao dịch giá trị lớn hoặc khoảng cách giao hàng xa dễ liên quan đến gian lận hơn. Ngược lại, *avs_match* (r = -0,22), *cvv_result* (r = -0,22) và *three_ds_flag* (r = -0,16) có tương quan âm, cho thấy giao dịch không vượt qua xác minh có rủi ro cao hơn. Ngoài ra, *account_age_days* tương quan âm yếu (r = -0,12), gợi ý tài khoản mới có nguy cơ gian lận lớn hơn.

Hình 2. Biểu đồ tương quan giữa các biến trong mô hình



Nguồn: Tác giả tổng hợp

4.3. Kết quả mô hình dự báo và đánh giá hiệu năng

Bảng 3. So sánh hiệu năng của các mô hình đề xuất

Mô hình	Độ chính xác (Accuracy)	Độ phân tán (Precision)	Độ nhạy (Recall)	Điểm số F1 (F1-score)	Giá trị AUC
Logistic	87,93%	63,74%	84,72%	0,7275	0,93
RF	99,18%	92,79%	68,15%	0,7858	0,96
XGBoost	95,32%	80,63%	88,80%	0,8452	0,97

Nguồn: Tác giả tổng hợp

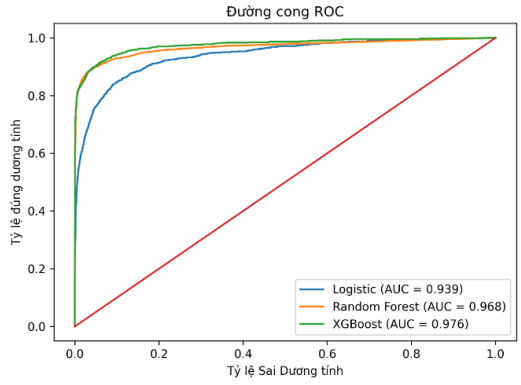
Hiệu năng của mô hình được đo lường trên tập kiểm tra (test dataset) nhằm đảm bảo tính khách quan và phản ánh đúng năng lực tổng quát hóa trong điều kiện dữ liệu thực tế. Các chỉ số đánh giá được sử dụng bao gồm Accuracy (độ chính xác), Precision (độ phân tán), Recall (độ nhạy), F1-score (chỉ số cân bằng giữa Precision và Recall) và AUC-ROC (khả năng phân biệt hai lớp trên toàn bộ các ngưỡng xác suất). Trong bối cảnh quản trị gian lận TMĐT, Recall mang ý nghĩa ưu tiên vì chi phí bỏ sót khách hàng có nguy cơ gian lận thường lớn hơn chi phí cảnh báo nhầm.

Từ Bảng 3 có thể rút ra một số nhận định quan trọng. Thứ nhất, xét riêng từng mô hình, hồi quy Logistic đạt Độ chính xác 87,93%, Độ phân tán 63,74%, Độ nhạy 84,72%, Điểm số F1 0,7275 và AUC 0,93. Kết quả này cho thấy hồi quy Logistic có khả năng bắt được tương đối nhiều giao dịch gian lận, thể hiện qua Độ nhạy khá tốt, nhưng Độ phân tán thấp hơn đáng kể so với hai mô hình còn lại. Điều đó hàm ý rằng mô hình tạo ra nhiều cảnh báo dương tính giả hơn, tức là dự đoán nhầm một số lượng giao dịch hợp lệ thành gian lận. Đây là đặc trưng thường gặp của mô hình tuyến tính khi phải xử lý dữ liệu có quan hệ phi tuyến và tương tác phức tạp. Đối với RF, mô hình cho Độ chính xác cao nhất trong ba mô hình (99,18%) và Độ phân tán cũng rất cao (92,79%), cho thấy khi RF dự đoán một giao dịch là gian lận thì xác suất đúng khá lớn. Tuy nhiên, Độ nhạy chỉ đạt 68,15%, thấp hơn đáng kể so với Logistic Regression và XGBoost. Điều này phản ánh mô hình có xu hướng thận trọng trong việc gắn nhãn gian lận, tức là ít báo động nhầm nhưng lại bỏ sót nhiều giao dịch gian lận thực sự. Vì vậy, nếu mục tiêu nghiên cứu là tối đa hóa khả năng phát hiện gian lận, RF chưa phải là thuật toán nên ưu tiên dù Độ chính xác rất cao. Trong khi đó, mô hình XGBoost có kết quả cân bằng nhất. Cụ thể, XGBoost đạt Độ chính xác 95,32%, Độ phân tán 80,63%, Độ nhạy 88,80%, Điểm số F1 0,8452 và AUC 0,97. Đây là mô hình có Điểm số F1 cao nhất và cũng có Độ nhạy cao nhất trong ba mô hình. Điều này cho thấy XGBoost không chỉ phát hiện được nhiều giao dịch gian lận hơn mà còn giữ được Độ phân tán tương đối tốt. Nói cách khác, mô hình này đạt được sự cân bằng tốt giữa việc phát hiện đúng gian lận và hạn chế báo động giả. Thứ hai, về AUC, cả ba mô hình đều có giá trị khá cao, trong đó XGBoost đạt 0,97, RF đạt 0,96, và hồi quy Logistic đạt 0,93. Điều này cho thấy cả ba mô hình đều có khả năng phân biệt tương đối tốt giữa giao dịch gian lận và giao dịch hợp lệ. Tuy nhiên, khi đặt trong bối cảnh thực tế của phát hiện gian lận TMĐT, Điểm số F1 và Độ nhạy thường có ý nghĩa thực tiễn hơn Độ chính

xác vì bài toán này ưu tiên phát hiện càng nhiều gian lận càng tốt, đồng thời vẫn phải kiểm soát số lượng cảnh báo sai ở mức chấp nhận được. Do đó, Bảng 3 ngụ ý rằng XGBoost được xem là mô hình ưu tiên vì nó cho hiệu năng toàn diện hơn, đặc biệt ở những chỉ số phản ánh đúng bản chất của bài toán mất cân bằng.

Hình 3 cho thấy ROC của cả ba mô hình đều có khả năng phân biệt khá tốt giữa giao dịch gian lận và hợp lệ, vì các đường cong đều nằm xa đường chéo ngẫu nhiên và tiến khá sát góc trên bên trái của đồ thị. Điều này hàm ý rằng, trong nhiều ngưỡng phân loại khác nhau, cả ba mô hình đều duy trì được mức Tỷ lệ đúng dương tính cao trong khi vẫn kiểm soát được Tỷ lệ sai dương tính ở mức tương đối thấp. Xét theo giá trị AUC, XGBoost đạt kết quả tốt nhất với AUC = 0,976, tiếp theo là RF với AUC = 0,968, và hồi quy Logistic đạt AUC = 0,939. Sự chênh lệch này cho thấy XGBoost có năng lực phân tách lớp tốt nhất trong ba mô hình, đặc biệt trong việc nhận diện các giao dịch gian lận khi thay đổi ngưỡng quyết định. RF cũng thể hiện hiệu năng rất mạnh và chỉ thấp hơn XGBoost một khoảng nhỏ, cho thấy đây là một mô hình cạnh tranh và khá ổn định. Hồi quy Logistic vẫn cho kết quả tốt nhưng đường cong của nó thấp hơn hai mô hình còn lại ở hầu hết vùng Tỷ lệ sai dương tính thấp và trung bình, phản ánh hạn chế của mô hình tuyến tính khi phải nắm bắt các quan hệ phi tuyến và tương tác phức tạp giữa các biến.

Hình 3. Đường cong ROC của ba mô hình đề xuất



Nguồn: Tác giả tổng hợp

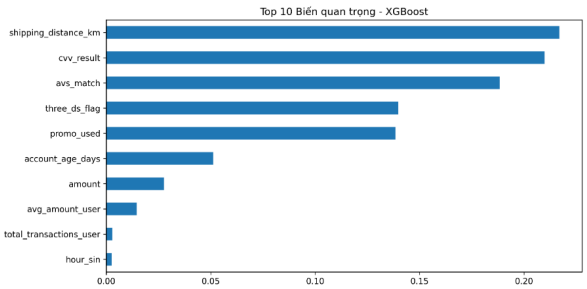
Một điểm đáng chú ý là ở vùng tỷ lệ sai dương tính thấp, khu vực quan trọng trong phát hiện gian lận, XGBoost và RF đạt độ nhạy cao hơn hồi quy Logistic dù số cảnh báo sai rất thấp. Đây là điều đặc biệt có giá trị trong thực tế, khi doanh nghiệp muốn hạn chế chặn nhầm giao dịch hợp lệ nhưng vẫn phát hiện hiệu quả giao dịch bất thường.

Kết quả cho thấy XGBoost là mô hình tốt nhất, tiếp theo là RF và Logistic. Điều này hàm ý các mô hình phi tuyến và nhiều cây quyết định khai thác tốt hơn dữ liệu gian lận TMĐT vốn phức tạp, mất cân bằng và có nhiều tương tác đặc trưng khó nắm bắt bằng mô hình tuyến tính. Phát hiện này phù hợp với Mutemi & Bacao (2024) và các nghiên cứu gần đây của Talukder et al. (2024).

4.4. Các nhân tố ảnh hưởng tới gian lận thương mại điện tử

Kết quả xếp hạng biến quan trọng của XGBoost cho thấy khả năng phân loại chủ yếu đến từ nhóm biến ngữ cảnh và xác thực giao dịch, nổi bật gồm "shipping_distance_km", "cvv_result", "avs_match", "three_ds_flag" và "promo_used", cao hơn các biến lịch sử như "amount" hay "avg_amount_user". Điều này phù hợp với nghiên cứu gần đây khi dữ liệu hành vi, tài khoản và bối cảnh giao dịch có giá trị phát hiện gian lận cao hơn tín hiệu giao dịch bề mặt (Zeng et al., 2025). Đồng thời, các cơ chế xác thực như AVS/CVV và 3-D Secure vẫn giữ vai trò quan trọng trong an toàn thanh toán trực tuyến, nên có giá trị dự báo cao trong mô hình (Plateaux et al., 2018). Kết quả cho thấy gian lận TMĐT thường được nhận diện qua sự bất nhất giữa xác thực, vận chuyển và hành vi giao dịch.

Hình 4. Phân tích các biến quan trọng ảnh hưởng tới khả năng gian lận TMĐT



Nguồn: Tác giả tổng hợp

5. Hàm ý quản trị

Kết quả nghiên cứu về ứng dụng học máy trong phát hiện gian lận TMĐT mang lại những hàm ý quan trọng cho các nhà quản trị rủi ro và các doanh

nghiệp trong hệ sinh thái kinh tế TMĐT. Thứ nhất, chuyển dịch từ hệ thống quản trị dựa trên quy tắc sang quản trị dựa trên dữ liệu. Các nhà quản lý cần nhận thức rõ rằng các hệ thống dựa trên luật thủ công (rule-based) và các đặc trưng tĩnh truyền thống hiện không còn đủ khả năng đối phó với các chiến thuật gian lận tinh vi và thích nghi. Nghiên cứu cho thấy XGBoost đạt hiệu năng vượt trội với giá trị AUC lên tới 0,97 và độ nhạy cao (88,8%). Do đó, doanh nghiệp nên ưu tiên đầu tư vào các mô hình học máy có khả năng nắm bắt các tương tác phi tuyến phức tạp thay vì chỉ dựa vào các mô hình kinh tế lượng tuyến tính truyền thống. Thứ hai, tối ưu hóa chiến lược thu thập và khai thác dữ liệu hành vi đa phương thức. Quản trị rủi ro hiện đại không chỉ tập trung vào các thông tin giao dịch bề mặt như số tiền, số giao dịch. Thay vào đó, nhà quản lý cần chú trọng khai thác các dấu vết hành vi khó làm giả hơn như khoảng cách giao hàng, kết quả xác thực CVV, và sự tương thích giữa quốc gia người mua với quốc gia phát hành thẻ. Việc kết hợp các biến ngữ cảnh này giúp tăng độ bền vững cho hệ thống phát hiện trước các hành vi giả mạo thông tin định danh. Thứ ba, cân bằng giữa trải nghiệm khách hàng và an ninh giao dịch. Một hàm ý quản trị quan trọng là việc lựa chọn ngưỡng phân loại dựa trên mục tiêu kinh doanh. Mô hình XGBoost được khuyến nghị vì nó giúp đạt được sự cân bằng tối ưu giữa việc phát hiện tối đa gian lận (giảm thiểu tổn thất tài chính) và hạn chế tối đa việc chặn nhầm các giao dịch hợp lệ (duy trì trải nghiệm khách hàng).

Kết luận

Nghiên cứu đã đánh giá và so sánh hiệu quả các thuật toán học máy trong phát hiện gian lận giao dịch TMĐT trên bộ dữ liệu gần 300.000 quan sát. Kết quả cho thấy các mô hình phi tuyến như XGBoost và Random Forest vượt trội hơn hồi quy Logistic trong xử lý dữ liệu phức tạp. Đặc biệt, XGBoost đạt hiệu năng cao nhất với AUC 0,97 và độ nhạy 88,8%, giúp phát hiện tốt giao dịch gian lận đồng thời kiểm soát báo động giả. Nghiên cứu cũng khẳng định xử lý mất cân bằng dữ liệu là yếu tố quan trọng. Việc áp dụng SMOTE đã giảm thiên lệch do giao dịch hợp lệ chiếm đa số, giúp mô hình nhận diện gian lận chính xác hơn. Đồng thời, các biến ngữ cảnh và xác thực như khoảng cách giao hàng, kết quả CVV có giá trị dự báo cao hơn các thông tin định danh thông thường, cho thấy xu hướng chuyển từ quản trị dựa trên quy tắc tĩnh sang quản trị dựa trên hành vi và ngữ cảnh. Về thực tiễn, nghiên cứu nhấn mạnh nhu cầu xây dựng hệ thống phát hiện gian lận theo thời gian thực thông qua tích hợp học máy với hạ tầng xử lý dữ liệu luồng. Tuy nhiên, nghiên cứu còn hạn chế do phụ thuộc dữ liệu thứ cấp. Các nghiên cứu tiếp theo có thể mở rộng sang dữ liệu phi cấu trúc như văn bản, hình ảnh hoặc ứng dụng học sâu để nâng cao hiệu quả trong môi trường TMĐT biến động liên tục.

Tài liệu tham khảo:

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., . . . Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. Applied Sciences, 12(19), 9637.

Carta, S., Fenu, G., Reforgiato Recupero, D., & Saia, R. (2019). Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model. Journal of Information Security and Applications, 46, 13–22.

Dai, Y., Viken, G., Joo, E., & Bente, G. (2018). Risk assessment in e-commerce: How sellers' photos, reputation scores, and the stake of a transaction influence buyers' purchase behavior and information processing. Computers in Human Behavior, 84, 342–351.

Gujarati, D. N. (2004). Basic econometrics-Damodar N. Gujarati: McGraw-Hill.

Homer, E. M. (2020). Testing the fraud triangle: a systematic review. Journal of Financial Crime, 27(1), 172–187.

Kodate, S., Chiba, R., Kimura, S., & Masuda, N. (2020). Detecting problematic transactions in a consumer-to-consumer e-commerce network. Applied Network Science, 5(1), 90.

Lu, S.-H., Wang, H.-J., & Nguyen, A. T. (2023). Machine learning applications on box-office revenue forecasting: The Taiwanese film market case study. In Optimal Transport Statistics for Economics and Related Topics (pp. 384–402): Springer.

Mutemi, A., & Bacao, F. (2024). E-commerce fraud detection based on machine learning techniques: Systematic literature review. Big Data Mining and Analytics, 7(2), 419–444.

Nguyen, A.-T., Lu, S.-H., & Nguyen, P. T. T. (2021). Validating and forecasting carbon emissions in the framework of the environmental Kuznets curve: the case of Vietnam. Energies, 14(11), 3144.

Plateaux, A., Lacharme, P., Vernois, S., Coquet, V., & Rosenberger, C. (2018). A comparative study of card-not-present e-commerce architectures with card schemes: What about privacy? Journal of Information Security and Applications, 40, 103–110.

Rodrigues, V. F., Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., da Costa, C. A., Barbosa, J. L. V., . . . Arcot, T. (2022). Fraud detection and prevention in e-commerce: A systematic literature review. Electronic Commerce Research and Applications, 56, 101207.

Sánchez-Paniagua, M., Fidalgo, E., Alegre, E., & Jáñez-Martino, F. (2025). Fraud detection in e-commerce: a comparative analysis of features to enhance machine learning models. Electronic Commerce Research, 1–36.

Seo, J.-H., & Kim, Y.-H. (2018). Machine-learning approach to optimize smote ratio in class imbalance dataset for intrusion detection. Computational intelligence and neuroscience, 2018(1), 9704672.

Talukder, M. A., Khalid, M., & Uddin, M. A. (2024). An integrated multistage ensemble machine learning model for fraudulent transaction detection. Journal of Big Data, 11(1), 168.

Tu, N. A., Khoa, D. D., Nhan, N. C. T., & Le Kieu Oanh, D. (2024). Hybrid ARIMA and machine learning approach for the VND/USD exchange rate prediction in Vietnam: A comparative analysis. In Partial Identification in Econometrics and Related Topics (pp. 587–604): Springer.

Zeng, Q., Lin, L., Jiang, R., Huang, W., & Lin, D. (2025). NNEnsLeG: A novel approach for e-commerce payment fraud detection using ensemble learning and neural networks. Information Processing & Management, 62(1), 103916.

Zhang, Z., Yin, H., Rao, S. X., Yan, X., Wang, Z., Liang, W., . . . Jiang, J. (2025). Identifying E-Commerce Fraud Through User Behavior Data: Observations and Insights. Data Science and Engineering, 10(1), 24–39.