

DỰ ĐOÁN TRẠNG THÁI GIAO HÀNG ĐA LỚP BẰNG TRÍ TUỆ NHÂN TẠO CÓ KHẢ NĂNG GIẢI THÍCH

LÙ THỊ HẰNG^{1*} - TỐNG KHÁNH LINH² - NGUYỄN ĐẶNG LAN PHƯƠNG³
- NGUYỄN BẢO NGÂN⁴ - BÙI ÁNH TUYẾT⁵

*Tác giả liên hệ - Email: luhang1311@gmail.com

Ngày nhận bài: 7/8/2026

Ngày sửa bài: 20/8/2026

Ngày duyệt đăng bài: 7/9/2026

Tóm tắt:

Giao hàng trễ làm suy giảm chất lượng dịch vụ và gia tăng chi phí xử lý trong chuỗi cung ứng, trong khi các mô hình dự báo phức tạp thường khó chuyển hóa thành quyết định quản trị. Thực nghiệm sử dụng bộ dữ liệu tổng hợp SynDelay công bố năm 2025 gồm 155.488 quan sát. Sau khi loại các biến định danh và biến chỉ xuất hiện sau thời điểm tiếp nhận đơn, 3 thuật toán Logistic Regression, Random Forest và XGBoost được đánh giá trên tập kiểm tra phân tầng 20%. Kết quả cho thấy, XGBoost đạt Macro-F1 cao nhất (0,4825), với precision 0,8753 và recall 0,5152 đối với lớp giao trễ. Phân tích SHAP xác định phương thức vận chuyển là yếu tố chi phối mạnh nhất, tiếp theo là giá trị đơn hàng và các đặc trưng địa lý. Kết quả minh họa tiềm năng kết hợp dự báo và giải thích để hỗ trợ sàng lọc đơn hàng rủi ro, đồng thời chỉ ra hiệu quả dự báo cần được diễn giải thận trọng do dữ liệu tổng hợp và sự mất cân bằng giữa các lớp.

Từ khóa: giao hàng trễ, trí tuệ nhân tạo có khả năng giải thích, XGBoost, SHAP, logistics

Explainable artificial intelligence for multi-class delivery status prediction

Abstract:

Late deliveries reduce service quality and increase exception-handling costs in supply chains, while complex predictive models are often difficult to translate into managerial decisions. This study uses the 2025 SynDelay synthetic dataset comprising 155,488 observations. After removing identifier variables and variables available only after order receipt, three algorithms: Logistic Regression, Random Forest, and XGBoost, are evaluated on a stratified 20% test set. The results show that XGBoost achieves the highest Macro-F1 score of 0.4825, with a precision of 0.8753 and a recall of 0.5152 for the delayed-delivery class. SHAP analysis identifies shipping mode as the most influential predictor, followed by order value and geographic features. The findings illustrate the potential of combining prediction and explainability to support shipment-risk screening, while also indicating that predictive performance should be interpreted cautiously because of the synthetic nature of the data and class imbalance.

Keywords: delivery delay, explainable artificial intelligence, XGBoost, SHAP, logistics

1. Đặt vấn đề

Tính đúng hạn là một chỉ tiêu trung tâm của chất lượng dịch vụ logistics. Khi một đơn hàng không đạt thời điểm cam kết, doanh nghiệp có thể phát sinh chi phí điều phối lại, chăm sóc khách hàng, bồi hoàn và suy giảm uy tín. Trước sự bùng nổ của công nghệ số, học máy mở ra khả năng chuyển dữ liệu tác nghiệp thành cảnh báo sớm, nhưng giá trị thực tế phụ thuộc không chỉ vào độ chính xác mà còn vào khả năng giải thích cho người ra quyết định.

Theo Baryannis, Dani và Antoniou (2019), dự báo rủi ro chuỗi cung ứng luôn tồn tại đánh đổi giữa hiệu năng và khả năng diễn giải. Một mô hình hộp đen có thể cho điểm dự báo tốt nhưng khó được chấp nhận nếu nhà quản trị không biết yếu tố nào làm tăng rủi ro. SHAP do Lundberg và Lee (2017) đề xuất cung cấp giá trị đóng góp của từng biến cho một dự báo, nhờ đó kết nối mô hình phức tạp với nhu cầu giải trình. Bài viết đặt 3 câu hỏi: (RQ1) mô hình nào cân bằng tốt nhất 3 trạng thái giao hàng; (RQ2) mô hình nhận diện lớp giao trễ ở mức nào; và (RQ3) những biến nào chi phối dự báo giao trễ? Đóng góp của bài viết là một quy trình thử nghiệm tránh dùng thông tin tương lai và lớp giải thích SHAP hướng tới ứng dụng vận hành.

2. Nền tảng và các nghiên cứu trước

2.1. Dự đoán giao hàng bằng học máy

Bài toán có thể được mô hình hóa dưới dạng phân loại đa lớp, trong đó hàm dự báo ánh xạ thông tin khả dụng khi tiếp nhận đơn sang một trong ba trạng thái. Logistic Regression tạo đường cơ sở tuyến tính và dễ kiểm tra. Random Forest kết hợp nhiều cây quyết định trên các mẫu và tập biến ngẫu nhiên, nhờ đó có khả năng mô hình hóa quan hệ phi tuyến và tương tác giữa biến (Breiman, L., 2001). XGBoost xây dựng tuần tự các cây tăng cường để sửa sai số của mô hình trước, đồng thời có cơ chế điều chuẩn và lấy mẫu giúp làm việc hiệu quả với dữ liệu bảng (Chen, T., & Guestrin, C., 2016).

2.2. Trí tuệ nhân tạo có khả năng giải thích

SHAP dựa trên giá trị Shapley để phân bổ chênh lệch giữa dự báo của một quan sát và giá trị nền cho từng biến đầu vào (Lundberg, S. M., & Lee, S.-I., 2017). Cách tiếp cận này không chứng minh quan hệ nhân quả, nhưng giúp kiểm tra tính hợp lý, phát hiện việc vô tình dùng thông tin tương lai và xác định biến cần ưu tiên trong vận hành. Tuy nhiên, cách giải thích cần phù hợp với người sử dụng. Nhân viên vận hành thường cần biết vì sao một đơn cụ thể bị cảnh báo; người quản lý lại quan tâm yếu tố nào thường xuyên ảnh hưởng đến toàn hệ thống.

2.3. Vai trò và giới hạn của dữ liệu tổng hợp

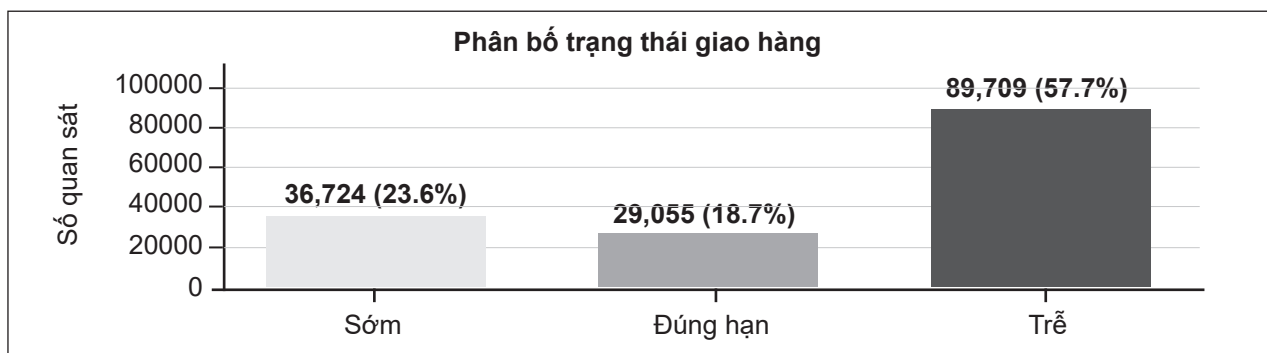
Dữ liệu logistics thực thường khó công bố vì chứa thông tin khách hàng, giá bán, tuyến vận chuyển và năng lực của đối tác. Dữ liệu tổng hợp học các mẫu thống kê của dữ liệu gốc rồi tạo ra các bản ghi mới, nhờ đó hỗ trợ chia sẻ và thử nghiệm mà không công bố trực tiếp từng giao dịch. Tổng quan của Fonseca và Bacao cho thấy, dữ liệu bảng tổng hợp có thể phục vụ bảo vệ riêng tư, tăng dữ liệu và huấn luyện mô hình; chất lượng vẫn phải được xem xét trên cả tính hữu ích, mức giống dữ liệu gốc và khả năng áp dụng cho bài toán cụ thể (Fonseca, J., & Bacao, F., 2023). Vì vậy, SynDelay phù hợp để so sánh phương pháp và kiểm tra quy trình, nhưng kết quả không phải là bằng chứng trực tiếp về tỷ lệ giao trễ của một doanh nghiệp hay quốc gia. Các nghiên cứu gần đây còn mở rộng dự báo chậm giao bằng dữ liệu theo thời gian và thông tin kinh tế bên ngoài. Chẳng hạn Gabellini và cộng sự sử dụng mô hình học sâu cùng các chỉ báo kinh tế để dự báo số ngày giao sớm hoặc trễ trong ngành công nghiệp ô tô (Gabellini, M., Civalani, L., Calabrese, F., & Bortolini, M., 2024).

3. Dữ liệu và phương pháp nghiên cứu

3.1. Bộ dữ liệu và biến mục tiêu

Nghiên cứu sử dụng tệp syndelay_v1.csv từ kho dữ liệu chính thức của SynDelay (Xu, L., Long, Y., & Brintrup, A., 2025). Bộ dữ liệu được tạo bằng mô hình sinh dữ liệu bảng học từ dữ liệu giao hàng thực, nhằm duy trì các quan hệ thống kê, đồng thời giảm rủi ro riêng tư. Tệp thực nghiệm có 155.488 quan sát, 41 biến đầu vào và biến label. Theo quy ước của bộ dữ liệu, label 0, 1 và 2 lần lượt biểu thị giao sớm, đúng hạn và trễ. Trong chính tệp v1 được sử dụng, 3 lớp có lần lượt 36.724, 29.055 và 89.709 quan sát; lớp trễ chiếm 57,7%.

Hình 1. Phân bố trạng thái giao hàng trong SynDelay



3.2. Chuẩn bị dữ liệu và tránh dùng thông tin tương lai

Thời điểm dự báo được xác định là khi doanh nghiệp tiếp nhận đơn. Vì vậy, nghiên cứu loại shipping_date, order_status, các biến định danh đơn hàng/khách hàng/sản phẩm và các biến lợi nhuận có thể chỉ hoàn thiện sau xử lý. 22 biến ban đầu được giữ lại; từ order_date sinh thêm tháng, thứ trong tuần và giờ, tạo thành 24 đặc trưng sau biến đổi. Không có ô thiếu trong tập biến được chọn. Biến số được điền trung vị và chuẩn hóa; biến phân loại được điền bằng giá trị xuất hiện nhiều nhất và chuyển thành mã số, trong đó giá trị chưa gặp ở tập huấn luyện nhận mã -1. Cách chọn này ngăn mô hình nhìn thấy kết quả tương lai rồi ghi nhớ đáp án.

Bảng 1. Nhóm đặc trưng sử dụng trong mô hình

Nhóm	Biến đại diện
Đơn hàng	Giá trị mua, giảm giá, số lượng, giá sản phẩm, loại thanh toán
Khách hàng	Phân khúc, quốc gia, bang, mã vùng, tọa độ
Tuyến giao	Thị trường, thành phố/quốc gia/vùng/bang nhận hàng
Thời gian	Tháng, thứ trong tuần và giờ đặt hàng
Vận chuyển	Phương thức vận chuyển

3.3. Cách huấn luyện và đánh giá

Dữ liệu được chia ngẫu nhiên có phân tầng thành 80% huấn luyện (124.390 quan sát) và 20% kiểm tra (31.098 quan sát), với số khởi tạo ngẫu nhiên 42. Trọng số lớp cân bằng được áp dụng trong quá trình huấn luyện. Logistic Regression, Random Forest 250 cây và XGBoost 500 cây được so sánh trên cùng tập kiểm tra. Các chỉ số gồm accuracy, Macro-F1, Weighted-F1 và precision, recall, F1 riêng cho lớp trễ. Sau đó TreeSHAP giải thích mô hình XGBoost trên mẫu ngẫu nhiên 1.200 quan sát kiểm tra. Cấu hình và số khởi tạo được cố định để người khác có thể chạy lại và kiểm tra kết quả.

4. Kết quả nghiên cứu

4.1. Khả năng dự đoán

Bảng 2. Kết quả trên tập kiểm tra

Mô hình	Acc.	Macro-F1	W-F1	P-late	R-late	F1-late
XGBoost	0.5496	0.4825	0.5508	0.8753	0.5152	0.6486
Random Forest	0.5733	0.4736	0.5568	0.8750	0.5495	0.6750
Logistic Regression	0.4959	0.4376	0.5119	0.8162	0.5020	0.6217

XGBoost đạt Macro-F1 cao nhất, 0,4825, nên được chọn theo tiêu chí chính. Random Forest đạt accuracy 0,5733 và Weighted-F1 0,5568, cao hơn XGBoost, nhưng phân bổ hiệu năng giữa 3 lớp kém cân bằng hơn. Logistic Regression có Macro-F1 0,4376, cho thấy quan hệ giữa đặc trưng và trạng thái giao hàng không hoàn toàn tuyến tính. Chênh lệch giữa accuracy và Macro-F1 cũng khẳng định việc chỉ báo cáo accuracy sẽ che khuất khó khăn ở các lớp ít quan sát.

Đối với lớp giao trễ, XGBoost đạt precision 0,8753 nhưng recall 0,5152. Nói cách khác, cảnh báo trễ của mô hình có độ tin cậy tương đối cao, song gần một nửa số đơn trễ chưa được phát hiện. Tùy chi phí bỏ sót, doanh nghiệp có thể hạ ngưỡng cảnh báo để tăng recall, đổi lại số cảnh báo sai tăng. Đây là quyết định nghiệp vụ không thể thay thế chỉ bằng tối ưu một chỉ số kỹ thuật.

Ma trận nhầm lẫn cho thấy, mô hình nhận diện 90,9% lớp giao sớm nhưng chỉ 20,2% lớp đúng hạn; phần lớn lỗi của lớp đúng hạn bị chuyển sang giao sớm hoặc giao trễ. Với lớp trễ, mô hình nhận diện đúng 9.243 trên 17.942 quan sát kiểm tra. Kết quả phản ánh ranh giới giữa 3 trạng thái khó hơn bài toán nhị phân và cho thấy, cần thêm dữ liệu về khoảng cách thực, tải mạng lưới, thời tiết hoặc SLA để phân biệt đúng hạn tốt hơn.

4.2. Giải thích mô hình bằng SHAP

Phương thức vận chuyển có Mean |SHAP| 1,2886, vượt xa các biến còn lại. Điều này cho thấy, lựa chọn mức dịch vụ vận chuyển là tín hiệu trung tâm mà XGBoost sử dụng để phân biệt rủi ro trễ. Sales per customer và tổng giá trị mặt hàng đúng tiếp theo, lần lượt đạt 0,0845 và 0,0798. Các biến kinh độ, mã vùng, thành phố giao, vĩ độ và bang giao hàng cùng xuất hiện trong nhóm đầu, phản ánh vai trò tổng hợp của không gian địa lý.

5. Ý nghĩa của kết quả và cách áp dụng

Thứ nhất, kết quả trả lời RQ1 rằng XGBoost cung cấp sự cân bằng tốt nhất giữa 3 lớp theo Macro-F1, dù Random Forest có độ chính xác cao hơn. Với các đơn trễ, độ chính xác cao phù hợp cho quy trình chỉ có nguồn lực xử lý một danh sách cảnh báo ngắn; ngược lại, mục tiêu bảo vệ SLA nghiêm ngặt cần tối ưu recall (Baryannis, G., Dani, S., & Antoniou, G., 2019).

Thứ hai, phương thức vận chuyển là tín hiệu có giá trị vận hành rõ nhất. Doanh nghiệp có thể xây dựng bảng theo dõi rủi ro theo phương thức vận chuyển, kết hợp tuyến địa lý và giá trị đơn để ưu tiên kiểm tra năng lực nhà vận chuyển, điều chỉnh mức dịch vụ hoặc chủ động thông báo khách hàng. Hệ thống nên hiển thị cả xác suất dự báo và 3 đến 5 yếu tố SHAP lớn nhất của từng đơn, thay vì chỉ trả về nhãn trễ/không trễ.

Thứ ba, độ chính xác vừa phải không phải thất bại của mô hình mà phản ánh giới hạn thông tin tại thời điểm đặt hàng. Một hệ thống triển khai có thể cập nhật dự báo theo 2 giai đoạn: lần đầu khi tiếp nhận đơn và lần hai khi có thông tin lấy hàng, tải mạng lưới hoặc tiến độ vận chuyển. Cấu trúc này vừa duy trì cảnh báo sớm vừa tận dụng dữ liệu mới, đồng thời phải lưu dấu phiên bản để bảo đảm khả năng giải trình.

Một cách triển khai đơn giản là thêm cột điểm rủi ro vào màn hình quản lý đơn hàng. Mỗi đầu ca, hệ thống xếp các đơn từ rủi ro cao xuống thấp và hiển thị lý do chính, chẳng hạn phương thức vận chuyển, tuyến địa lý hoặc giá trị đơn. Phản hồi của nhân viên cần được lưu để đối chiếu với kết quả giao thực tế. Sau một thời

gian, doanh nghiệp có thể đo số đơn trễ tránh được, chi phí can thiệp và số cảnh báo không cần thiết thay vì chỉ theo dõi điểm F1 trong môi trường thử nghiệm.

6. Kết luận

Nghiên cứu đã xây dựng và đánh giá quy trình XAI dự đoán 3 trạng thái giao hàng trên 155.488 quan sát SynDelay. Sau khi tránh dùng thông tin tương lai, XGBoost đạt Macro-F1 0,4825 và precision 0,8753 cho lớp trễ. SHAP cho thấy, phương thức vận chuyển, giá trị đơn hàng và địa lý là các nhóm tín hiệu chính. Kết quả cho thấy, việc kết hợp dự báo với giải thích có thể hỗ trợ doanh nghiệp sàng lọc đơn rủi ro và phân bổ nguồn lực can thiệp có căn cứ hơn. Nghiên cứu tiếp theo nên thu thập dữ liệu đơn hàng Việt Nam, bổ sung thời tiết - giao thông - khoảng cách, thực hiện đánh giá ngoài mẫu theo thời gian và hiệu chỉnh ngưỡng dựa trên chi phí nghiệp vụ ■

TÀI LIỆU THAM KHẢO:

- Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. *Future Generation Computer Systems*, 101, 993-1004. <https://doi.org/10.1016/j.future.2019.07.059>.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD*, 785-794. <https://doi.org/10.1145/2939672.2939785>.
- Fonseca, J., & Bacao, F. (2023). Tabular and latent space synthetic data generation: A literature review. *Journal of Big Data*, 10, 115. <https://doi.org/10.1186/s40537-023-00792-7>.
- Gabellini, M., Civolani, L., Calabrese, F., & Bortolini, M. (2024). A deep learning approach to predict supply chain delivery delay risk based on macroeconomic indicators: A case study in the automotive sector. *Applied Sciences*, 14(11), 4688. <https://doi.org/10.3390/app14114688>.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Xu, L., Long, Y., & Brintrup, A. (2025). SynDelay: A Synthetic Dataset for Delivery Delay Prediction. *arXiv:2509.05325*.

Thông tin tác giả:

LÙ THỊ HẰNG^{1*} - TỐNG KHÁNH LINH²

- NGUYỄN ĐẶNG LAN PHƯƠNG³ - NGUYỄN BẢO NGÂN⁴ - BÙI ÁNH TUYẾT⁵

¹Trường Đại học Mở Hà Nội
Hanoi Open University
*E-mail: luhang1311@gmail.com

²Trường Đại học Ngoại thương
Foreign Trade University
E-mail: tongkhanhlinh102003@gmail.com

³Trường Đại học Văn Lang
Van Lang University
E-mail: yolynghuyen1119@gmail.com

⁴Trường THPT Chuyên Bắc Ninh
Bac Ninh High School for the Gifted
E-mail: baonganamy@gmail.com

⁵Trường THPT Nguyễn Trãi
Nguyen Trai High School
E-mail: batuyet2009@gmail.com