

NGHIÊN CỨU ỨNG DỤNG LIGHTGBM VÀO KIỂM KÊ PHÁT THẢI CỦA TÀU CONTAINER TẠI CẢNG NAM ĐÌNH VŨ

APPLICATION OF LIGHTGBM IN THE EMISSION INVENTORY OF CONTAINER SHIPS AT NAM DINH VU PORT

ĐOÀN LÊ NAM KHÁNH¹, LÊ CÔNG DOANH¹, NGUYỄN THỊ HẰNG¹,
PHAN VĂN HÙNG^{2*}

¹Sinh viên Khoa Hàng hải, Trường Đại học Hàng hải Việt Nam

²Khoa Hàng hải, Trường Đại học Hàng hải Việt Nam

*Email liên hệ: phanvanhung@vamaru.edu.vn

DOI: <https://doi.org/10.65154/jmst.974>

Tóm tắt

Nghiên cứu này tập trung ứng dụng mô hình học máy Light Gradient Boosting Machine (LightGBM) nhằm nội suy hệ số phát thải của tàu container tại cảng biển Nam Đình Vũ, qua đó góp phần nâng cao độ chính xác trong công tác ước tính phát thải khí từ hoạt động tàu biển - một trong những nguồn phát thải chính tại khu vực cảng. Phạm vi nghiên cứu là 693 lượt tàu container ra vào cảng Nam Đình Vũ trong năm 2023. Kết quả đánh giá mô hình cho thấy LightGBM đạt độ chính xác rất cao đối với tất cả các chất ô nhiễm, với hệ số xác định R^2 luôn lớn hơn 0,97, sai số phần trăm MAPE nhỏ hơn 7% và sai số tuyệt đối MAE ổn định theo quy mô phát thải. Đặc biệt, các thành phần có vai trò quan trọng trong kiểm kê một số loại khí nhà kính như CO_2 , CH_4 , N_2O và các chất gây ô nhiễm như SO_x , HC và bụi mịn đều đạt mức dự báo chính xác vượt trội. Phân tích ma trận tương quan cho thấy các biến kỹ thuật và vận hành của tàu có mối quan hệ phi tuyến phức tạp với hệ số phát thải, đồng thời không tồn tại hiện tượng đa cộng tuyến nghiêm trọng, khẳng định sự phù hợp của mô hình LightGBM trong bài toán nội suy hệ số phát thải. Cách tiếp cận này mở ra hướng ứng dụng hiệu quả cho công tác quản lý môi trường cảng biển và có tiềm năng mở rộng cho các cảng biển khác tại Việt Nam.

Từ khóa: Mô hình học máy LightGBM, cảng Nam Đình Vũ, khí thải tàu container.

Abstract

This study focuses on applying the Light Gradient Boosting Machine (LightGBM) model to interpolate emission factors of container vessels at Nam Dinh

Vu Port, thereby improving the accuracy of emission estimation from maritime shipping activities, which represent one of the major sources of emissions in port areas. The research covers 693 container vessel calls at Nam Dinh Vu Port in 2023. Model evaluation results demonstrate that LightGBM achieves very high accuracy for all pollutants, with the coefficient of determination R^2 consistently exceeding 0.97, mean absolute percentage error (MAPE) remaining below 7%, and mean absolute error (MAE) remaining stable relative to emission magnitude. Notably, the components fundamental to the inventory of specific greenhouse gases including CO_2 , CH_4 and N_2O as well as pollutants such as SO_x , hydrocarbons (HC), and particulate matter (PM), have all achieved exceptional levels of predictive accuracy. Correlation matrix analysis indicates that vessel technical and operational variables exhibit complex nonlinear relationships with emission factors, while no serious multicollinearity is observed, further confirming the suitability of the LightGBM model for emission factor interpolation. This approach provides an effective pathway for port environmental management and demonstrates strong potential for extension to other seaports in Vietnam.

Keywords: LightGBM; Nam Dinh Vu Port; container vessels emissions.

1. Mở đầu

Trong bối cảnh hiện nay, vận tải biển toàn cầu ngày càng phát triển, dù vận tải biển cũng được xem làm một hình thức thân thiện với môi trường tuy nhiên sự gia tăng lưu lượng tại các cảng đã tác động làm gia tăng lượng phát thải khí, đối với ngành vận tải container đường biển toàn cầu trong năm 2024 đã đạt mức phát thải carbon cao nhất trong lịch sử với 240,6 triệu tấn tăng 14,3% so với năm 2023 [1], gây ra những ảnh hưởng xấu tới sức khỏe và chất lượng cuộc sống của người dân các khu vực xung quanh cảng [2,3] và đóng góp lớn vào biến đổi khí hậu.

Nhằm thực hiện mục tiêu giảm phát thải và phát triển bền vững IMO đã ban hành phụ lục VI MARPOL đưa ra những giới hạn dành cho SO_x, NO_x nhằm hạn chế mức phát thải [4], gần đây nhất là chiến dịch “Net Zero” phát thải ròng bằng 0 vào năm 2050, đạt được nhiều hưởng ứng của các quốc gia. Đối với Việt Nam đã tích cực tham gia công ước và thực hiện cam kết chung như tham gia phụ lục VI MARPOL 73/78 và

đặt mục tiêu “Net Zero” phát thải ròng bằng 0 vào năm 2050 tại COP 26 [5], [6].

Nhằm đánh giá và đưa ra biện pháp hiệu quả để giảm phát thải khí nhà kính, xây dựng công cụ kiểm kê với độ chính xác cao là yêu cầu bắt buộc. Đối với việc kiểm kê thường được sử dụng tại Việt Nam, các nghiên cứu trước đây tập trung vào tính toán theo khu vực bằng cách sử dụng các hệ số phát thải hiện có dựa trên nhiên liệu hay động cơ do các tổ chức như EPA, EEA, IMO,... đề xuất. Trên thực tế phát thải từ hoạt động tàu phụ thuộc chặt chẽ vào đặc trưng kỹ thuật của tàu và điều kiện vận hành của tàu. Mối quan hệ giữa công suất động cơ, tốc độ hành trình, thời gian neo đậu và trạng thái điều động của tàu với lượng phát thải CO₂ thường mang tính phi tuyến cao. Các phương pháp kiểm kê truyền thống, mặc dù đơn giản và dễ triển khai, chưa phản ánh đầy đủ sự khác biệt về đặc tính và chế độ hoạt động của từng tàu, dẫn đến sai số đáng kể trong việc ước tính. Các mô hình học máy là phương pháp tiềm năng giúp giải quyết vấn đề này.

Bảng 1. Tổng quan nghiên cứu liên quan

Nguồn	Bộ dữ liệu	Mô hình	Độ chính xác	Hạn chế
[7]	Công dữ liệu mở của Chính phủ Canada	LSTM, BiLSTM	R ² = 93,78%	Phụ thuộc dataset có sẵn, khó phân ánh dữ liệu thời gian thực và kịch bản phát thải động.
[8]	Dữ liệu thử nghiệm đường bộ bằng PEMS và OBDII	Gradient Boosting	R ² = 61,26%	Hiệu năng dự báo còn hạn chế (R ² = 0,61; MSE = 0,77).
[9]	GPS thực tế từ xe điện tại Nhật	LightGBM, XGBoost, ANN, LR	R ² = 98%	Chưa xét tải bổ sung, hành vi lái và giao thông.
[10]	Dữ liệu tài liệu về thiết bị khí hóa tầng sôi	Gradient Boosting	R ² = 84,68%	Có đề cập XAI nhưng chưa phân tích sâu tác động rộng hơn.
[11]	Hiệu suất xe hybrid với lớn hơn 20 tham số (cảm biến CO ₂ SprintIR-R 20)	UWS-LSTM	R ² = 97,50%	Chuẩn hóa dữ liệu chưa rõ, ảnh hưởng khả năng tái lập.
[12]	Dataset xe hạng nhẹ Chính phủ Canada (2017–2021)	Hồi quy đa thức một biến	R ² = 98,60%	Thiếu mô hình mạnh hơn và dataset nhiều đặc trưng.
[13]	Dataset WB & cơ quan giao thông Thổ Nhĩ Kỳ	MLP, XGBoost, SVM	R ² = 98,86%	Chỉ đánh giá 4 kịch bản tương quan đầu vào.
[14]	PEMS đo NO _x , CO ₂ ... từ xe diesel	Gradient Boosting Regression	R ² = 99%	Đầu vào ít biến dẫn tới giới hạn độ chính xác.
[15]	Dữ liệu lái xe thực tế từ các sự kiện phanh	CatBoost	R ² = 83%	Chỉ 600 sự kiện nên chưa đại diện đầy đủ.
[16]	Open Energy Data Initiative (OEDI)	ML Ensemble	R ² = 94,70%	Chi phí tính toán cao.

Mối quan hệ phi tuyến phức tạp được huấn luyện dựa vào của dữ liệu đầu vào và đầu ra, các mô hình học máy cho phép ước lượng hệ số phát thải mà không cần thực hiện các mô hình phi tuyến tốn thời gian. Một nghiên cứu đã tổng hợp có tới hơn 20 mô hình, thuật toán tối ưu trong việc ước tính CO₂ [17], cùng hàng loạt các mô hình đã được thử nghiệm tại các cảng biển trên thế giới như Explainable Artificial Intelligence (XAI) [18], PrE-PARE model [19],... Các mô hình học máy cho thấy hiệu quả đáng kể trong quá trình tính toán phát thải. Trong Bảng 1, nhóm tác giả đã tổng hợp 1 số nghiên cứu liên quan tới mô hình học máy và bộ dữ liệu được sử dụng trong các mô hình. Nhìn chung, các nghiên cứu được phân chia làm 2 loại, đa số các nghiên cứu sử dụng Machine Learning kết hợp với bộ dữ liệu được cung cấp thực địa, hai là sử dụng PEMS - đây là một công cụ đo đặc phát thải thực địa, Mặc dù hệ thống đo đặc phát thải di động (PEMS) cung cấp dữ liệu thực nghiệm có độ tin cậy cao ở cấp độ thời điểm, kết quả tổng hợp từ các nghiên cứu trước cho thấy hiệu năng dự báo dựa trên dữ liệu PEMS thường không ổn định khi mở rộng sang các kịch bản vận hành dài hạn hoặc quy mô lớn. Trong bảng tổng hợp, mô hình Gradient Boosting sử dụng dữ liệu thử nghiệm đường bộ từ PEMS chỉ đạt hệ số xác định $R^2=0,61$, thấp hơn đáng kể so với các nghiên cứu sử dụng bộ dữ liệu chuẩn hóa hoặc dữ liệu vận hành dài hạn với R^2 thường đạt hơn 0,9, tuy nhiên với nghiên cứu PEMS NOx, CO₂,... từ xe diesel với mô hình học máy Gradient Boosting Regression lại cho kết quả cao hơn với R^2 đạt 99%, tuy nhiên mô hình có biến vào ít biến dẫn tới giới hạn độ chính xác. Từ đó có thể nhận thấy rằng hiệu quả của các mô hình học máy không chỉ phụ thuộc vào việc sử dụng dữ liệu đo đặc trực tiếp như PEMS, mà còn phụ thuộc vào chất lượng chuẩn hóa dữ liệu, độ đa dạng biến đầu vào và khả năng học quan hệ phi tuyến của thuật toán. Trong bối cảnh đánh giá phát thải dài hạn cho tàu biển, cách tiếp cận kết hợp dữ liệu chuẩn hóa và mô hình học máy mang lại sự cân bằng tốt hơn giữa độ tin cậy, khả năng mở rộng và chi phí triển khai so với các phương pháp đo đặc thực nghiệm thuần túy. Nên trong bài nghiên cứu này, sẽ sử dụng mô hình học máy kết hợp với việc sử dụng bộ dữ liệu được báo cáo thực địa của từng tàu. Nhận thấy mô hình LightGBM nổi bật với khả năng xử lý các bài toán hồi quy phi tuyến phức tạp, tốc độ xử lý nhanh và cho ra độ chính xác cao.

Trên cơ sở đó, nghiên cứu này sẽ ứng dụng mô hình học máy LightGBM giúp nâng cao độ chính xác trong quá trình kiểm kê phát thải tàu biển tại cảng Nam Đình Vũ. Cách tiếp cận này mang tính mới tại

Việt Nam, khi lần đầu tiên mô hình LightGBM được áp dụng để cá thể hóa hệ số phát thải tàu biển ở quy mô cảng, thay vì sử dụng các hệ số phát thải trung bình theo tiêu chuẩn quốc tế. Phần sau của bài nghiên cứu sẽ trình bày các phần sau.

2. Tổng quan nghiên cứu

2.1. Thu thập và mô tả dữ liệu

Lượng tàu biển ra vào cảng Nam Đình Vũ tính từ ngày 1/1/2023 đến ngày 31/12/2023, với 693 lượt tàu ra vào cảng được thu thập theo báo cáo của cảng trong năm 2023. Các dữ liệu thu thập của tàu bao gồm 16 thông số như Bảng 2, trong đó:

Bảng 2. Thông tin của tàu được thu thập

TT	Dữ liệu	Thông số
1	Tên tàu (tên định danh)	CMA CGM FUJI
2	Trọng tải (GT)	25,165
3	Năm đóng	2015
4	Thời gian đến (giờ)	13/11/23 15:00:00
5	Thời gian đi (giờ)	17/11/23 17:00:00
6	Số giờ trong cảng (giờ)	98
7	Neo (giờ)	90,19
8	Hành trình (giờ)	0,24
9	Điều động (giờ)	7,57
10	Vận tốc - Hành trình (hải lý/ giờ)	19
11	Vận tốc - Điều động (hải lý/ giờ)	5
12	Vận tốc - Tối đa (hải lý/ giờ)	19
13	Công suất động cơ chính (kW)	13000
14	Công suất động cơ phụ (kW)	7200
15	Loại máy chính	MAN-B&W 6S60ME-C8
16	Tốc độ động cơ - RPM	105

Thời gian đến: Tính từ khi tàu tới điểm đón hoa tiêu;

Thời gian đi: Tính đến khi tàu rời điểm đón hoa tiêu;

Thời gian hành trình: Tàu hành trình từ điểm đón hoa tiêu tới phao số 0;

Thời gian điều động: Tàu điều động trong khu vực luồng và cập cảng;

Thời gian neo đậu: Khi tàu neo động để chờ làm hàng và thời gian tàu làm hàng.

Dữ liệu về hoạt động của tàu trong từng giai đoạn được thu thập bằng báo cáo tại cảng, theo dõi thời gian tàu hoạt động trong từng khu vực, việc này giúp hoạt động đánh giá phát thải của tàu được chính xác hơn, dữ liệu cần được thu thập đầy đủ trước khi được tính toán.

2.2. Xử lý dữ liệu và tiêu chí loại trừ ngoại lai

(a) Xử lý dữ liệu

Để phân loại dữ liệu, mô hình sử dụng kỹ thuật Label Encoding cho các biến như loại tàu và loại máy về dạng số học giúp mô hình xử lý được, trong trường hợp thiếu dữ liệu, dữ liệu được xử lý bằng cách điền giá trị trung bình hoặc trung vị của các tàu cùng loại.

Đối với các giai đoạn hoạt động của tàu được phân loại theo Bảng 3, tốc độ của tàu phụ thuộc vào lưu lượng tại cảng biển.

Bảng 3. Tốc độ hoạt động của tàu từng giai đoạn

Giai đoạn	Khoảng tốc độ điển hình (hải lý/ giờ)
Neo đậu	0
Điều động	3 – 6
Hành trình	8 – 11

(b) Tiêu chí loại trừ ngoại lai

Để đảm bảo độ tin cậy của mô hình và kết quả ước tính. Đối với các biến liên tục như tốc độ quay động cơ RPM và công suất máy, nghiên cứu này sử dụng phương pháp IQR (Interquartile Range), đây một phương pháp thống kê phi tham số, ít nhạy cảm với phân phối lệch và các giá trị cực trị. Cụ thể với mỗi biến liên tục X, các tứ phân vị thứ nhất Q1 và thứ ba Q3 được xác định, từ đó tính khoảng tứ phân vị:

$$IQR = Q3 - Q1 \quad (1)$$

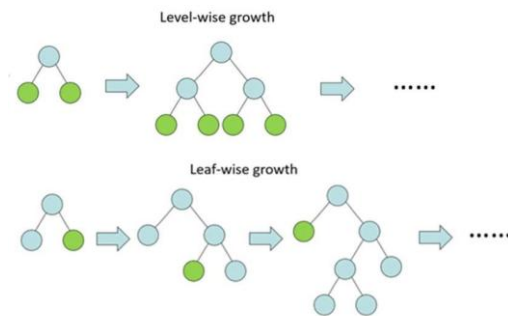
x_i được ngoại lai theo tiêu chí thống kê khi nằm trong khoảng $[Q1 - 1,5 \times IQR, Q3 + 1,5 \times IQR]$. Tuy nhiên trong ngành hàng hải, việc áp dụng máy móc các tiêu chí thống kê có thể loại bỏ các quan sát có ý nghĩa thực tiễn, đối với các tàu siêu trường, có thể là ngoại lai về mặt thống kê, nhưng phù hợp trong điều kiện vận hành thực tế. Do đó chỉ những giá trị

nào được xét vào cả tiêu chí ngoại lai của IQR và vi phạm những ràng buộc kỹ thuật cơ bản của hệ thống tàu biển (như công suất, tốc độ âm, công suất vượt quá giới hạn cho phép).

2.3. Mô hình LightGBM

LightGBM là một thuật toán học máy do nhóm Microsoft Research Asia phát triển, thuộc họ mô hình cây quyết định tăng cường viết tắt của Gradient Boosting Decision Tree - GBDT [20]. Thuật toán này được đánh giá cao nhờ khả năng huấn luyện nhanh, hiệu quả tính toán tốt và phù hợp với các bài toán dự đoán có quy mô dữ liệu lớn trong khi vẫn duy trì độ chính xác cao.

Thuật toán GBDT truyền thống có các đặc trưng đầu vào được sắp xếp trước nhằm xác định chính xác điểm chia tối ưu cho mỗi nút của cây quyết định. Cách tiếp cận này giúp nâng cao độ chính xác trong quá trình phân tách dữ liệu, nhưng đồng thời làm gia tăng đáng kể chi phí tính toán và yêu cầu bộ nhớ [20].



Hình 1. Mô tả mô hình LightGBM

Để khắc phục nhược điểm này, LightGBM áp dụng phương pháp xây dựng cây dựa trên biểu đồ, kết hợp với chiến lược phát triển cây theo chiều lá.

Bên cạnh đó, LightGBM cho phép thiết lập giới hạn độ sâu tối đa của cây nhằm kiểm soát độ phức tạp mô hình, qua đó hạn chế nguy cơ khớp quá mức trong quá trình huấn luyện. Nhờ sự kết hợp giữa thuật toán dựa trên biểu đồ và chiến lược phát triển cây theo chiều lá, LightGBM đạt được sự cân bằng giữa độ chính xác, tốc độ tính toán và khả năng mở rộng, trở thành một trong những mô hình GBDT hiệu quả nhất hiện nay. Mô hình được thiết kế để xấp xỉ hàm ảnh xạ như sau:

$$EF = f(P, LF, RPM, V, Age, Type, Engine) \quad (2)$$

Trong đó:

EF: Là hệ số phát thải tàu biển (g/kW),

P: Là công suất máy (kW),

LF: Là hệ số tải của tàu biển (%),
RPM: Là tốc độ quay của động cơ diesel (vòng),
V: Là vận tốc (hải lý/ giờ),
Age: Là tuổi tàu (tính từ khi xuất xưởng),
Type: Là loại tàu (container, hàng rời,...),
Engine: Là loại động cơ (MAN, Thường, có van trượt hoặc không có van trượt).

2.3. Huấn luyện và đánh giá mô hình

2.3.1. Huấn luyện mô hình

Tập dữ liệu gồm 693 lượt tàu được chia thành hai tập độc lập: 80% cho huấn luyện và 20% cho kiểm tra, bảo đảm việc đánh giá mô hình phản ánh đúng khả năng tổng quát hóa đối với dữ liệu chưa từng xuất hiện trong quá trình học. Quá trình huấn luyện tuân theo cơ chế gradient boosting, trong đó mỗi cây quyết định mới được xây dựng nhằm hiệu chỉnh phần sai số còn lại của tổ hợp các cây trước đó, giúp mô hình dần hội tụ đến nghiệm tối ưu.

2.3.2. Đánh giá mô hình

Mô hình sử dụng các chỉ số đánh giá độ chính xác phổ biến bao gồm: R^2 , MAE và MAPE nhằm đánh giá độ chính xác của mô hình dự báo so với mô hình truyền thống.

a) Sai số R^2

R^2 hay còn gọi là Coefficient of Determination, nó đo tỷ lệ phần trăm biến thiên (phương sai) của biến mục y mà mô hình có thể giải thích dựa trên x . Giá trị R^2 càng tiến gần 1 cho thấy mô hình càng có khả năng mô tả tốt mối quan hệ giữa các biến giải thích và EF. Công thức tính R^2 và R^2 hiệu chỉnh như sau:

$$R^2 = 1 - \frac{ESS}{TSS} \quad (3)$$

$$R^2_{hc} = 1 - \left(\frac{\frac{ESS}{n}}{\frac{TSS}{n-1}} \right) = 1 - \frac{n-1}{n-k} \times (1 - R^2) \quad (4)$$

Trong đó:

R^2 : Là độ lệch bình phương giải thích từ hồi quy,

ESS : Là tổng các độ lệch bình phương phần dư,

TSS : Là tổng các độ lệch bình phương toàn bộ,

n : Là số lượng mẫu quan sát,

k : Là số tham số của mô hình, bằng số lượng biến độc lập cộng 1.

b) Sai số MAPE

MAPE hay còn được gọi là chỉ số phần trăm tuyệt đối trung bình được dùng để đánh giá tương đối giữa hệ số phát thải dự đoán bởi mô hình Light GBM

và hệ số phát thải tham chiếu theo tiêu chuẩn của EPA. MAPE phản ánh sai số dự đoán dưới dạng phần trăm, cho phép so sánh hiệu quả mô hình trong các trường hợp dữ liệu có sự khác biệt lớn về quy mô hệ số phát thải giữa các tàu.

Công thức tính MAPE được xác định như sau:

$$MAPE = \frac{1}{n} \times \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

Trong đó:

\hat{y}_i : Là hệ số phát thải EF dự đoán (g/kWh);

y_i : Là hệ số phát thải chuẩn theo EPA (g/kWh);

n : Là số lượng tàu trong tập dữ liệu.

Việc sử dụng EF theo tiêu chuẩn do EPA ban hành làm mốc tham chiếu cho phép đánh giá khách quan mức độ cải thiện của mô hình LightGBM so với phương pháp ước tính truyền thống.

c) Sai số MAE

$$MAE = \frac{\sum_{t=1}^n |e_t|}{n} = \frac{\sum_{t=1}^n |Y_i - \hat{Y}_i|}{n} \quad (6)$$

Trong đó:

Y_i : Là hệ số phát thải chuẩn theo EPA của tàu i (g/kW),

\hat{Y}_i : Là hệ số phát thải dự đoán bởi mô hình LightGBM của tàu thứ i (g/kW),

n : Là số lượng tàu trong tập dữ liệu.

Chỉ số MAE phản ánh mức độ sai lệch tuyệt đối trung bình của hệ số phát thải EF thông qua việc đo lường độ chênh lệch tuyệt đối giữa giá trị dự đoán và giá trị tham chiếu theo Bảng 4.

Bảng 4. Đánh giá sai số của mô hình

Hệ số phát thải	R^2	MAPE (%)	MAE (g)
CO	0.95915	6.7570	6.82×10^3
NO _x	0.98719	6.6697	9.31×10^4
HC	0.97455	4.8692	2.28×10^3
CH ₄	0.97455	4.8692	4.56×10^1
PM ₁₀	0.97326	4.7368	1.01×10^3
PM _{2.5}	0.97299	4.7028	9.01×10^2
CO ₂	0.98257	3.9001	3.64×10^6
SO _x	0.98276	3.8874	2.20×10^3
N ₂ O	0.98384	3.7802	1.53×10^2

(a) Đánh giá tổng thể bằng R^2

Kết quả cho thấy mô hình đạt độ phù hợp rất cao đối với tất cả các chất ô nhiễm, với R^2 luôn lớn hơn 0,97. Trong đó, các thành phần N₂O ($R^2 = 0,98384$), SO_x (0,98276), CO₂ (0,98257) và NO_x (0,98719) thể

hiện khả năng mô hình hóa vượt trội. Điều này khẳng định mô hình LightGBM đã nắm bắt tốt các quan hệ phi tuyến phức tạp giữa đặc trưng kỹ thuật - vận hành của tàu và hệ số phát thải.

(b) Phân tích sai số tương đối bằng MAPE

Sai số phần trăm MAPE của các thành phần đều nhỏ hơn 7%, phản ánh độ ổn định cao của mô hình trên toàn bộ các chất ô nhiễm. Đặc biệt, các khí N₂O (3,78%), SO_x (3,89%) và CO₂ (3,90%) đạt mức sai lệch rất thấp, cho thấy mô hình có khả năng dự báo cực kỳ chính xác đối với các khí có vai trò quan trọng trong kiểm kê phát thải khí nhà kính.

Trong khi đó, các thành phần CO (6,76%) và NO_x (6,67%) có MAPE cao hơn, phù hợp với đặc điểm biến động mạnh của các khí này trong quá trình điều động và vận hành tàu tại cảng.

(c) Phân tích sai số tuyệt đối bằng MAE

Giá trị MAE phản ánh quy mô sai số tuyệt đối theo đơn vị vật lý của từng hệ số phát thải. Khí CO₂ có MAE lớn nhất (3,64 × 10⁶ g), tuy nhiên điều này là hệ quả trực tiếp của quy mô phát thải lớn, không phản ánh sự suy giảm độ chính xác của mô hình mà điều này được xác nhận bởi MAPE thấp và R² cao trong ứng. Ngược lại, các thành phần như CH₄ (4,56 × 10¹ g) và N₂O (1,53 × 10² g) có MAE nhỏ, mô hình dự báo rất ổn định ngay cả đối với các chất có nồng độ phát thải thấp.

(d) Nhận định khoa học tổng hợp

Việc kết hợp đồng thời R² cao, MAPE thấp và MAE hợp lý theo quy mô phát thải chứng minh mô hình LightGBM đạt độ chính xác thống kê cao và đảm bảo tính tin cậy vật lý trong bài toán nội suy hệ số phát thải tàu biển.

(e) Kiểm định chéo

Thay vì chỉ dựa vào một lần chia dữ liệu huấn luyện-kiểm tra theo tỷ lệ 80-20, phương pháp kiểm định chéo K-fold (với K = 5) được xem xét nhằm đánh giá tính ổn định và khả năng khái quát hóa của mô hình. Cách tiếp cận này giúp giảm nguy cơ học vẹt (overfitting) bằng cách đánh giá hiệu năng mô hình trên nhiều cách phân chia dữ liệu khác nhau, đồng thời đảm bảo rằng hệ số xác định cao (R₂ ≈ 0,975) không phụ thuộc vào một lần chia dữ liệu ngẫu nhiên cụ thể mà vẫn duy trì ổn định trên các tập dữ liệu khác nhau.

(f) Phân tích độ nhạy

Bên cạnh đó, phân tích độ nhạy được thực hiện thông qua việc thay đổi các siêu tham số chính của mô

hình, bao gồm learning rate và max depth. Kết quả cho thấy hiệu năng của mô hình tương đối ổn định, với giá trị dao động trong khoảng 0,96 đến 0,975 giữa các cấu hình khác nhau. Mức biến thiên nhỏ này cho thấy mô hình có tính vững cao (robust) và không quá nhạy cảm với những thay đổi nhỏ của siêu tham số.

2.4. Kiểm kê phát thải tàu biển theo EF hiệu chỉnh bởi LightGBM

Trên cơ sở các hệ số phát thải được nội suy và cá thể hóa bởi mô hình LightGBM, nghiên cứu tiến hành xây dựng quy trình kiểm kê phát thải tàu biển tại cảng Nam Đình Vũ theo hướng dẫn của Bộ Theo hướng dẫn của Bộ Tài Nguyên và Môi trường về hướng dẫn kỹ thuật kiểm kê phát thải đối với các nguồn từ tàu thủy [21] và Cơ quan Bảo vệ Môi trường Hoa Kỳ (EPA) [22], công thức tính toán lượng khí thải được xác định như sau:

$$E = EF_{LightGBM} \times LF \times A \times P \quad (7)$$

Trong đó:

E là lượng phát thải của tàu biển (g);

P là công suất máy (kW);

LF là hệ số tải của động cơ (%);

A là thời gian hoạt động của tàu (giờ);

EF_{LightGBM} là hệ số phát thải đã được hiệu chỉnh bởi mô hình LightGBM (g/kWh).

+ Đối với máy chính thì [22]:

$$LF = (\text{Tốc độ thực tế} / \text{Tốc độ tối đa})^3 \quad (8)$$

+ Đối với các máy phụ thì tuân thủ hướng dẫn của [21] theo Bảng 5.

Bảng 5. Hệ số tải của động cơ phụ [12]

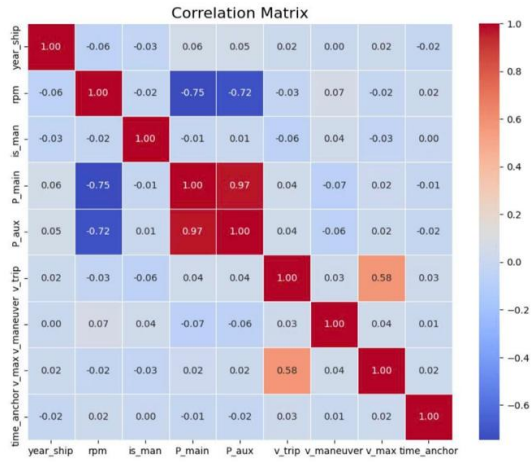
Loại tàu	Hành trình	Điều động	Dỡ hàng
Tàu hàng rời	0.17	0.45	0.10
Tàu container	0.13	0.48	0.19
Tàu du lịch	0.80	0.80	0.64
Tàu hàng tổng hợp	0.17	0.45	0.22
Tàu khác	0.17	0.45	0.22
Tàu RORO	0.15	0.45	0.26
Tàu lạnh	0.20	0.67	0.32

3. Kết quả tính toán và thảo luận

3.1. Phân tích mối quan hệ giữa các biến đầu vào trong mô hình LightGBM

Hình 2 thể hiện ma trận tương quan Pearson giữa các biến đầu vào được sử dụng trong quá trình huấn

luyện mô hình LightGBM, bao gồm các thông số kỹ thuật của tàu và điều kiện vận hành.



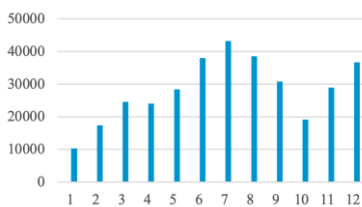
Hình 2. Ma trận tương quan giữa các đặc trưng kỹ thuật - vận hành của tàu container tại cảng Nam Đình Vũ

Kết quả cho thấy tồn tại mối tương quan rất mạnh giữa công suất máy chính và công suất máy phụ ($r \approx 0,97$). Điều này phản ánh đặc điểm thiết kế thực tế của tàu container, trong đó các tàu có công suất máy chính lớn thường được trang bị hệ thống máy phụ tương ứng để đáp ứng nhu cầu năng lượng trong quá trình neo đậu và khai thác tại cảng.

Bên cạnh đó, RPM có mối tương quan âm rõ rệt với công suất máy chính và máy phụ ($r \approx -0,72$ đến $-0,75$). Kết quả này phù hợp với quy luật kỹ thuật động cơ diesel tàu thủy, khi các động cơ công suất lớn thường vận hành ở dải tốc độ quay thấp hơn nhằm tối ưu hiệu suất và độ bền cơ học.

3.2. Kết quả kiểm kê phát thải khí nhà kính

Dựa trên các hệ số phát thải, ta tính toán được CO_{2e} tương đương, nghiên cứu tiến hành kiểm kê tổng lượng phát thải nhà kính của 693 lượt tàu container ra vào cảng Nam Đình Vũ trong năm 2023 theo từng tháng. Kết quả được tổng hợp và trình bày trong Hình 3.



(Đơn vị: Kg)

Hình 3. Lượng phát thải CO_{2e} theo tháng của tàu container tại cảng Nam Đình Vũ năm 2023

Biểu đồ cho thấy lượng phát thải CO_{2e} có sự biến động rõ rệt theo thời gian, phản ánh trực tiếp cường độ hoạt động khai thác tàu container tại cảng. Giá trị phát thải thấp nhất được ghi nhận vào tháng 1, đạt khoảng $1,03 \times 10^7$ kg, tương ứng với giai đoạn đầu năm khi lưu lượng tàu và nhu cầu vận chuyển hàng hóa còn thấp.

Từ tháng 3 đến tháng 6, lượng phát thải CO_{2e} tăng nhanh, đạt khoảng $3,8 \times 10^7$ kg vào tháng 6. Đỉnh phát thải xuất hiện vào tháng 7 với giá trị xấp xỉ $4,32 \times 10^7$ kg, cao gấp hơn 4 lần so với tháng thấp nhất trong năm. Giai đoạn này trùng với mùa cao điểm xuất nhập khẩu, khi số lượt tàu tăng mạnh và thời gian điều động, neo đậu tại cảng kéo dài.

Sau tháng 7, lượng phát thải có xu hướng giảm dần trong các tháng 9 và 10, trước khi tăng trở lại vào cuối năm. Tháng 12 ghi nhận mức phát thải tương đối cao, khoảng $3,67 \times 10^7$ kg, cho thấy sự phục hồi mạnh mẽ của hoạt động khai thác tàu container phục vụ nhu cầu logistics cuối năm.

3.3. Thảo luận

Kết hợp phân tích ma trận tương quan (Hình 2) và kết quả kiểm kê theo tháng (Hình 3) cho thấy các biến liên quan đến công suất máy chính và máy phụ có mức độ liên hệ cao, phản ánh đúng đặc trưng vận hành của tàu container trong điều kiện khai thác thực tế. Mối tương quan mạnh giữa công suất máy chính và máy phụ và cho thấy khi tải động cơ chính tăng, nhu cầu năng lượng phụ trợ cũng gia tăng, đặc biệt trong các giai đoạn chuyển trạng thái vận hành của tàu. Điều này phù hợp với các nguyên lý vận hành của hệ thống động lực tàu và giải thích vì sao các mô hình học máy dựa trên cây quyết định như LightGBM có thể khai thác hiệu quả các mối quan hệ phi tuyến tiềm ẩn.

Ngoài ra, sự biến động phát thải theo tháng phản ánh trực tiếp sự thay đổi về mật độ tàu và thời gian neo đậu tại cảng. Các giai đoạn có lưu lượng tàu cao cho thấy phát thải tăng chủ yếu do hoạt động của máy phụ trong quá trình cấp điện cho tàu khi neo bến. Điều này gợi ý rằng các giải pháp kỹ thuật như cấp điện bờ hoặc tối ưu hóa lịch trình tàu có thể mang lại hiệu quả giảm phát thải đáng kể trong bối cảnh cảng container.

Những kết quả đạt được không chỉ cung cấp bức tranh định lượng chi tiết về phát thải CO₂ tại cảng Nam Đình Vũ trong năm 2023, mà còn khẳng định tiềm năng ứng dụng của các mô hình học máy, đặc biệt là LightGBM, trong công tác kiểm kê và quản lý phát thải khí nhà kính tại các cảng biển ở Việt Nam.

4. Kiến nghị và giải pháp

4.1. Kiến nghị

Thứ nhất, chuyển đổi phương pháp kiểm kê phát thải từ hệ số trung bình sang hệ số cá thể hóa theo tàu. Kết quả nghiên cứu cho thấy mô hình LightGBM giúp cải thiện độ chính xác dự báo hệ số phát thải đạt độ chính xác rất cao đối với tất cả các chất ô nhiễm, với hệ số xác định R^2 luôn lớn hơn 0,97, sai số phần trăm MAPE nhỏ hơn 7% và sai số tuyệt đối MAE ổn định theo quy mô phát thải so với phương pháp sử dụng hệ số phát thải trung bình theo tiêu chuẩn quốc tế, tương tự hướng nghiên cứu trong các nghiên cứu tại các cảng container lớn trên thế giới [23], [24].

Cảng Nam Đình Vũ phối hợp với Cục Hàng hải Việt Nam xem xét thí điểm áp dụng mô hình học máy (LightGBM) trong kiểm kê phát thải định kỳ, trước mắt tập trung vào khí CO_2 và NO_x - hai thành phần có ảnh hưởng lớn đến biến đổi khí hậu và chất lượng không khí khu vực cảng.

Thứ hai, chuẩn hóa và số hóa dữ liệu vận hành tàu phục vụ công tác kiểm kê phát thải. Phân tích ma trận tương quan và kết quả kiểm kê cho thấy các thông số vận hành như thời gian neo đậu, thời gian điều động, RPM và vận tốc có ảnh hưởng đáng kể đến lượng phát thải, song hiện nay các dữ liệu này chủ yếu được lưu trữ phân tán và chưa được khai thác hiệu quả cho mục tiêu môi trường.

Xây dựng hệ thống cơ sở dữ liệu số về hoạt động tàu tại cảng Nam Đình Vũ, làm tiền đề cho việc áp dụng các mô hình dự báo phát thải thông minh và phục vụ báo cáo môi trường theo yêu cầu trong nước và quốc tế.

4.2. Đề xuất giải pháp

Thứ nhất, tối ưu hóa hoạt động điều động và neo đậu tàu trong giai đoạn cao điểm. Kết quả kiểm kê cho thấy phát thải CO_2 đạt đỉnh trong các tháng cao điểm logistics (tháng 5-8), đặc biệt là tháng 7. Điều này cho thấy phát thải không chỉ phụ thuộc vào số lượt tàu mà còn chịu ảnh hưởng lớn từ thời gian chờ cầu bến và điều động kéo dài. Các nghiên cứu tại châu Âu và châu Á đã chứng minh rằng tối ưu hóa lịch tàu và điều phối khai thác có thể giúp giảm đáng kể phát thải khí tại cảng [25], [26].

Trên cơ sở đó, nghiên cứu đề xuất cảng Nam Đình Vũ cần tăng cường công tác điều phối lịch tàu nhằm hạn chế tình trạng tàu phải neo chờ kéo dài trong điều kiện tải thấp. Đồng thời, trong các khung giờ khai thác cao điểm, nên ưu tiên bố trí các tàu có công suất động cơ lớn vào những khoảng thời gian khai thác tối ưu,

qua đó rút ngắn tổng thời gian vận hành máy phụ và giảm phát thải phát sinh trong giai đoạn neo đậu và điều động.

Thứ hai, kết quả cho thấy thời gian neo đậu pha thải cao do máy phụ vẫn phải hoạt động liên tục trong bước triển khai hệ thống cấp điện bờ (Onshore Power Supply - OPS) cho tàu container trong thời gian neo đậu tại cảng. Phân tích dữ liệu vận hành cho thấy thời gian neo đậu chiếm tỷ trọng lớn trong tổng thời gian tàu lưu trú tại cảng Nam Đình Vũ, đồng nghĩa với việc các máy phụ phải hoạt động liên tục để cung cấp điện cho các hệ thống trên tàu, dẫn đến phát thải đáng kể CO_2 cũng như các chất ô nhiễm khác. Nhiều nghiên cứu quốc tế đã chứng minh rằng việc sử dụng OPS có thể giúp giảm từ 30-60% lượng phát thải CO_2 trong giai đoạn neo đậu, đồng thời giảm mạnh phát thải NO_x , SO_x và tiếng ồn tại khu vực cảng [27], [28], [29].

Từ cơ sở phân tích, đánh giá mô hình, nhóm nghiên cứu đề xuất tích hợp mô hình LightGBM vào hệ thống quản lý môi trường của cảng Nam Đình Vũ. Việc ứng dụng mô hình học máy trong hỗ trợ ra quyết định giúp nâng cao hiệu quả quản lý phát thải và tạo nền tảng khoa học cho công tác quy hoạch, điều hành khai thác và thực hiện các mục tiêu giảm phát thải bền vững tại cảng trong tương lai.

5. Kết luận

Nghiên cứu này đã chứng minh rõ ràng tính hiệu quả và tính khả thi của việc ứng dụng mô hình học máy LightGBM trong nội suy hệ số phát thải và kiểm kê phát thải CO_2 từ hoạt động tàu container tại cảng Nam Đình Vũ. Lượng phát thải của tàu biển trong cảng Nam Đình Vũ trong năm 2023 đạt 34.001 tấn CO_2e . Trên cơ sở bộ dữ liệu gồm 693 lượt tàu trong năm 2023, mô hình LightGBM cho thấy khả năng vượt trội trong việc nắm bắt các mối quan hệ phi tuyến phức tạp giữa đặc trưng kỹ thuật, điều kiện vận hành của tàu và hệ số phát thải, với mức cải thiện độ chính xác lên tới cao và sai số nhỏ với hệ số xác định R^2 luôn lớn hơn 0,97, sai số phần trăm MAPE nhỏ hơn 7%. Kết quả kiểm kê phản ánh rõ xu thế biến động phát thải theo thời gian và làm nổi bật vai trò của các yếu tố vận hành và hệ số tải trong tổng lượng phát thải ở quy mô cảng.

Việc ứng dụng LightGBM giúp nâng cao hiệu quả giảm phát thải trong công tác quản lý môi trường cảng biển. Cách tiếp cận dựa trên dữ liệu và học máy giúp nâng cao độ tin cậy và tính khoa học của kết quả kiểm kê phát thải, cung cấp cơ sở định lượng vững chắc cho việc xây dựng các giải pháp giảm phát thải có mục tiêu và hiệu quả. Đồng thời, nghiên cứu khẳng định

tiềm năng mở rộng và nhân rộng mô hình LightGBM cho các cảng biển khác tại Việt Nam, góp phần hỗ trợ quá trình thực hiện cam kết phát thải ròng bằng “0” vào năm 2050 theo định hướng phát triển bền vững của ngành hàng hải.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: **SV25-26.03**.

TÀI LIỆU THAM KHẢO

- [1] Xeneta & Marine Benchmark (2025), *2024 emissions report: Record-breaking carbon emissions in ocean container shipping*.
- [2] Jia Shi, et al (2025), *The impact of low-sulfur marine fuel policy on air pollution in global coastal cities*, Sustainable Horizons, Vol.14, 100130.
- [3] César Ducruet, et al (2024), *Ports and their influence on local air pollution and public health: A global analysis*, Science of the Total Environment, Vol.915, 170099.
- [4] United States Environmental Protection Agency (2023), *MARPOL Annex VI and the Act to Prevent Pollution from Ships (APPS)*, United States Environmental Protection Agency Website.
- [5] Nguyễn Văn Trường, Vũ Thị Phương Oanh, Phạm Thị Thanh Niên (2024), *Đánh giá thực thi Phụ lục VI Công ước quốc tế về ngăn ngừa ô nhiễm từ tàu thuyền tại Việt Nam*, Tạp chí Khoa học Công nghệ Hàng hải, Số 78 (04/2024), tr.107-113.
- [6] Trần Thực (2021), *COP 26 và đóng góp của Việt Nam cho mục tiêu toàn cầu*, Tạp chí Khoa học & Công nghệ Việt Nam, Số 12, tr.18-20.
- [7] Al-Nefaie, A. H. & Aldhyani, T. H. H. (2023), *Predicting CO₂ emissions from traffic vehicles for sustainable and smart environment using a deep learning model*, Sustainability, Vol.15(9), p. 1765. Doi: 10.3390/su15097615.
- [8] Li, M. et al (2023), *Explaining a machine-learning lane change model with maximum entropy shapley values*, IEEE Transactions on Intelligent Vehicles.
- [9] Ullah, I. et al. (2021), *A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability*, Energy & Environment.
- [10] Pandey, D. S., Raza, H. & Bhattacharyya (2023), *S. Development of explainable ai-based predictive models for bubbling fluidised bed gasification process*, Fuel.
- [11] Tena-Gago, D., Golkarnarenji, G., Martinez-Alpiste, I., Wang, Q. & Calero, J. M. A. (2023), *Machine-learning-based carbon dioxide concentration prediction for hybrid vehicles*, Sensors, Vol.23(3), p.1350. Doi: 10.3390/s23031350
- [12] Hien, N. L. H. & Kor, A.-L. (2022), *Analysis and prediction model of fuel consumption and carbon dioxide emissions of light-duty vehicles*. Applied Sciences, Vol.12(2), p.803. Doi: 10.3390/app12020803
- [13] Cinarer, G., YESILYURT, M. K., Agbulut, U., Yilbasi, Z. & Kılıc, K. (2024), *Application of various machine learning algorithms in view of predicting the CO₂ emissions in the transportation sector*, Sustainable Transportation and Energy Transition.
- [14] Wen, H.-T., Lu, J.-H., Lu, J. & Jhang, D.-S. (2021), *Features importance analysis of diesel vehicles' nox and CO₂ emission predictions in real road driving based on gradient boosting regression model*, International Journal of Environmental Research and Public Health.
- [15] Wei, N. et al. (2022), *Applying machine learning to construct braking emission model for real-world road driving*, SSRN Electronic Journal.
- [16] Mohammad, F., Saleem, K. & Al-Muhtadi, J. (2023), *Ensemble-learning-based decision support system for energy-theft detection in smart-grid environment*, Energies.
- [17] Hala Salem Al Nuaimi, et al (2025), *Machine learning applications for carbon emission estimation*, Resources, Conservation & Recycling Advances, Vol.27, p. 200263.
- [18] Juhyang Lee, et al (2024), *The Development of a Machine Learning-Based Carbon Emission Prediction Method for a Multi-Fuel-Propelled Smart Ship by Using Onboard Measurement Data*, Sustainability, Vol.16(6), 2381.
- [19] Bojić Filip, Gudelj Anita, Bošnjak Rino (2025), *A Comprehensive Model for Quantifying, Predicting, and Evaluating Ship Emissions in Port Areas Using Novel Metrics and Machine Learning Methods*, Journal of Marine Science and Engineering, p. 1162.

- [20] Guolin Ke, et al (2017), *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*, Advances in Neural Information Processing Systems (NIPS), Vol.30, pp.3146-3154.
- [21] Bộ Tài nguyên và Môi trường, Công văn số 1074/BTNMT-KSONMT ngày 21/02/2024 về hướng dẫn kỹ thuật kiểm kê phát thải bụi và khí thải từ nguồn thải điểm, nguồn điện và nguồn di động.
- [22] U.S. Environmental Protection Agency (2009), *Current Methodologies in Preparing Mobile Source Port-Related Emission Inventories Final Report*, Virginia.
- [23] Lei Yanga, et al (2021), *An AIS-based emission inventory and the impact on air quality in Tianjin port based on localized emission factors*, Science of the Total Environment, Vol.783, p. 146896.
- [24] Chen, Z. S., et al (2024), *Prediction of harbour vessel emissions based on machine learning approach*, Transportation Research Part D: Transport and Environment, Vol.131, p. 104214.
- [25] Wen, X., et al (2024), *Multi-objective optimization for ship scheduling with port congestion and environmental considerations*, Journal of Marine Science and Engineering, Vol.12(114).
- [26] Port of Rotterdam (2024), *Rotterdam port introduces Geofence for Just-in-Time sailing*, Port of Rotterdam. Retrieved October 29, 2024.
- [27] Stolz, B., et al (2021), *The CO₂ reduction potential of shore-side electricity in Europe*, Applied Energy, Vol.285, p.116425.
- [28] Thalys Zis., et al (2014), *Evaluation of cold ironing and speed reduction policies to reduce ship emissions near and at ports*, Vol.16, pp.371-398.
- [29] Nguyen, D.-H., et al. (2021), *Reduction of NO_x and SO₂ emissions by shore power adoption*, Aerosol and Air Quality Research, Vol.21(7), p. 200100.

Ngày nhận bài:	17/01/2026
Ngày nhận bản sửa:	13/02/2026
Ngày duyệt đăng:	17/02/2026