

Bài báo nghiên cứu

ỨNG DỤNG PHƯƠNG PHÁP EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) VÀ CÁC MÔ HÌNH HỌC MÁY TRUYỀN THỐNG XỬ LÝ BÀI TOÁN TIÊU THỤ NĂNG LƯỢNG TÁI TẠO

Nguyễn Thành Huy*, Đinh Tấn Lộc, Nguyễn Nhật Quang,
Trần Thị Kim Chi, Hoàng Quỳnh Anh, Lê Di Khanh

Trường Đại học Kinh tế Thành phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ: Nguyễn Thành Huy – Email: huynt@ueh.edu.vn.

Ngày nhận bài: 25-11-2024; Ngày nhận bài sửa: 17-4-2025; Ngày duyệt đăng: 20-5-2025

TÓM TẮT

Nghiên cứu này tập trung vào việc giải quyết thách thức dự báo sản lượng điện năng lượng mặt trời cho Công ty Enefit thông qua việc ứng dụng XAI kết hợp với các mô hình học máy. Mục tiêu là xác định những yếu tố quan trọng ảnh hưởng đến sản lượng điện và tiêu thụ, nhằm phục vụ cho quá trình tích hợp liền mạch các hệ thống quang điện (PV) vào lưới điện quốc gia. Nghiên cứu cũng đề xuất các phương pháp học máy khác nhau giúp nâng cao độ chính xác của dự báo, bằng cách kết hợp dữ liệu như: thời tiết lịch sử; giá khí đốt và giá điện, cùng với các yếu tố về dự báo thời tiết. Nhóm nghiên cứu đã phát triển nhiều mô hình như tuyến tính, phi tuyến tính và LightGBM (LGBM), tiếp tục tiến hành thí nghiệm để tìm ra mô hình hiệu quả nhất. Kết quả nghiên cứu đã giúp giảm thiểu sự mất cân bằng trong lưới điện, đồng thời nâng cao hiệu suất và độ tin cậy khi tích hợp năng lượng tái tạo. Những nỗ lực nghiên cứu tiếp theo sẽ tập trung vào việc tích hợp dữ liệu thời gian thực và áp dụng các kỹ thuật học sâu (deep learning) nhằm nâng cao độ chính xác và hiệu quả của dự báo.

Từ khóa: Photovoltaic system (PV); năng lượng tái tạo; mô hình truyền thống; XAI

1. Giới thiệu

Enefit Green, thuộc tập đoàn Eesti Energia, là một trong những nhà sản xuất năng lượng tái tạo lớn ở khu vực Baltic, hoạt động tại nhiều quốc gia như Estonia, Litva, Phần Lan, Ba Lan và Latvia. Công ti tập trung phát triển điện gió và điện mặt trời, với mục tiêu dẫn đầu khu vực vào năm 2026, đồng thời chuyển dần sang năng lượng mặt trời để giảm chi phí và tác động môi trường. Sự xuất hiện của các “Prosumers” – hộ gia đình sản xuất và tiêu thụ điện từ nguồn tái tạo – tạo ra thách thức mới khi họ không chỉ tiêu thụ mà còn bán điện dư thừa trở lại lưới điện, dẫn đến sự dao động giữa cung và cầu. Điều này gây quá tải lưới

Cite this article as: Nguyen, T. H., Dinh, T. L., Nguyen, N. Q., Tran, T. K. C., Hoang, Q. A., & Le, D. K. (2026). Applying explainable AI to traditional machine learning models for sustainable economic consumption. *Ho Chi Minh City University of Education Journal of Science*, 23(4), 955-965. [https://doi.org/10.54607/hcmue.js.23.4.4613\(2026\)](https://doi.org/10.54607/hcmue.js.23.4.4613(2026))

điện, tăng chi phí và phát sinh các vấn đề khác. Để giải quyết những thách thức này, nhóm nghiên cứu đã sử dụng dữ liệu từ Enefit Green (công bố trên Kaggle) với mục tiêu giảm hóa đơn tiền điện và duy trì nguồn cung thông qua việc sử dụng năng lượng hiệu quả. Ngoài ra, nghiên cứu phân tích các yếu tố chính ảnh hưởng đến tiêu thụ năng lượng, nhằm phát triển các chiến lược quản lý hiệu quả sự mất cân bằng cung-cầu, tăng cường sự ổn định của lưới điện, và thúc đẩy phát triển năng lượng bền vững.

Các nghiên cứu trước đây đã nêu rõ tác động tiêu cực của nguồn năng lượng không tái tạo, từ đó khẳng định sự cần thiết chuyển sang năng lượng tái tạo. Tolliver và cộng sự (2021), Sadorsky (2010) nhấn mạnh tầm quan trọng của việc tích hợp năng lượng tái tạo vào các lĩnh vực như khu dân cư, công nghiệp, dịch vụ, giao thông và phát triển bền vững. Peñasco và Anadón (2023) đã chỉ ra lợi ích của việc lắp đặt tấm pin mặt trời trong khu dân cư nhằm giảm phụ thuộc vào nhiên liệu hóa thạch (Peñasco et al., 2023).

Sự lệch pha giữa cung và cầu điện gây nhiều hệ lụy cho cả hệ thống điện lẫn môi trường. Khi thiếu điện vào giờ cao điểm dễ xảy ra mất điện, còn dư thừa lúc nhu cầu thấp lại làm tăng chi phí vận hành và phát thải. Tình trạng này càng khó kiểm soát khi nguồn năng lượng tái tạo như gió, mặt trời có tính biến động. Ngoài ra, mức tiêu thụ điện còn chịu ảnh hưởng bởi nhiều yếu tố như thu nhập, thời tiết, quy mô hộ gia đình và giá điện – trong đó hộ gia đình khá giả hoặc ở điều kiện khí hậu khắc nghiệt thường dùng nhiều điện hơn (Frederiks et al., 2015).

Các phương pháp dự báo truyền thống như hồi quy hay chuỗi thời gian thường khó nắm bắt các mối quan hệ phức tạp (Zhao & Magoulès, 2012). Các mô hình như rừng ngẫu nhiên, máy tăng cường độ dốc và mạng nơ-ron sâu đã cải thiện độ chính xác nhờ khả năng xử lý dữ liệu lớn, nhưng vẫn gặp hạn chế về tài nguyên, dữ liệu và khả năng giải thích (Ahmad et al., 2018). Hơn nữa, khả năng giải thích của các mô hình này còn hạn chế, gây khó khăn trong việc hiểu các yếu tố chính dẫn đến dự báo (Molnar, 2024).

Nghiên cứu này sẽ phát triển một phương pháp XAI đa giai đoạn, có khả năng giải thích và tùy biến, được huấn luyện trên dữ liệu của Công ty Enefit. Mục tiêu là xác định các yếu tố chính ảnh hưởng đến tính bền vững và tiêu thụ năng lượng. Phương pháp này tận dụng khả năng linh hoạt của các khung XAI, đảm bảo tính bền vững và giảm phát thải ròng

2. Nội dung

2.1. Diễn giải (XAI) các mô hình học máy

Về dự báo năng lượng mặt trời, nhiều nghiên cứu và nỗ lực nghiên cứu đã đóng góp và mang lại cái nhìn sâu sắc và tiến bộ đáng kể (Allal et al., 2024). Các công cụ XAI như LIME, SHAP và ELI5 đã được sử dụng để hiểu và cải thiện hiệu suất của mô hình AI Random Forest trong việc dự báo sản lượng điện mặt trời PV (Kuzlu et al., 2020). Bên cạnh đó, ứng dụng XAI có thể cải thiện niềm tin và tính minh bạch của các mô hình học máy trong hệ thống năng lượng và điện lực bằng cách cung cấp các giải thích cục bộ và toàn cầu

tăng cường niềm tin và cho phép tối ưu hóa, đồng thời thảo luận về sự đánh đổi giữa các phương pháp XAI nội tại và hậu kỳ cho các mô hình mới và hiện có (Machlev et al., 2022).

Cụ thể, SHAP có nghĩa là “Shapley Additive exPlanations”, sử dụng các khái niệm lý thuyết trò chơi để đo lường ảnh hưởng của các đặc tính riêng lẻ đến dự đoán của một mô hình học máy (Van den Broeck et al., 2022). SHAP cung cấp các giải thích cục bộ chính xác hơn về dự đoán của mô hình so với các hệ số hồi quy tuyến tính toàn cục bằng cách khám phá cách thay đổi các đặc tính đầu vào riêng lẻ ảnh hưởng đến đầu ra cho mỗi điểm dữ liệu (Marcílio et al., 2020). SHAP được biểu diễn bằng công thức sau, đại diện cho tập hợp tất cả các biến đầu vào M:

$$\phi_i(f, \mathbf{x}) = \sum_{S \subset M \setminus i} \frac{|S|! (|M| - |S| - 1)!}{|M|!} [f(S \cup \{i\}) - f(S)]$$

trong đó, S biểu thị một tập con của M với đặc trưng thứ i bị loại trừ, và đóng góp của đặc trưng thứ i được cho bởi hàm $f(S \cup \{i\}) - f(S)$.

- $\phi_i(\mathbf{v})$ là đóng góp của đặc trưng thứ (i) vào hàm giá trị.
- (N) là tập hợp gồm (n) đặc trưng (từ 1, ..., (n)).
- (S) là một tập hợp con của (N) không chứa đặc trưng (i).
- |S| là số lượng phần tử trong tập hợp (S).

Mặt khác, LIME thường nhanh hơn SHAP vì không cần xét tất cả các tổ hợp đặc trưng, nên chi phí tính toán thấp hơn. Đây là một phương pháp giải thích cục bộ, giúp làm rõ dự đoán của mô hình bằng cách phân tích quanh từng điểm dữ liệu cụ thể. Cách hoạt động của LIME là tạo thêm các mẫu gần điểm cần giải thích, rồi huấn luyện một mô hình đơn giản để xấp xỉ hành vi của mô hình gốc. Từ đó, nó dùng các hệ số của mô hình này để giải thích kết quả. Tuy nhiên, chất lượng giải thích phụ thuộc nhiều vào cách lấy mẫu. Đây là công thức chính của LIME, nhằm tìm mô hình giải thích g tối ưu, bằng cách cân bằng giữa độ trung thành cục bộ (L) và độ phức tạp (Ω):

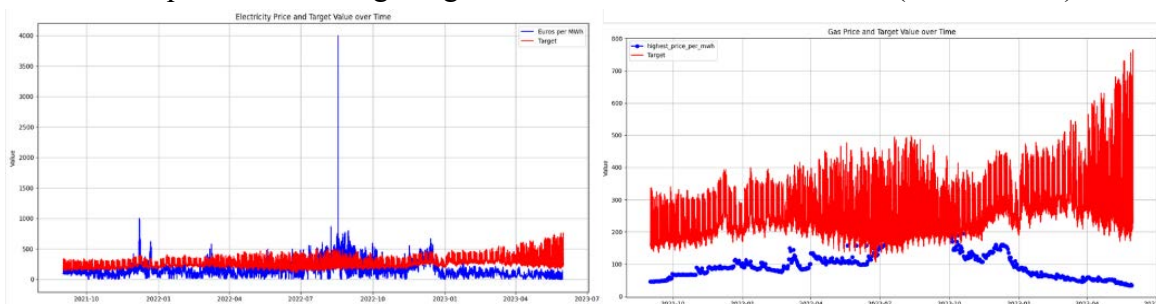
$$\xi(x) = \operatorname{argmin}_{g \in G} \mathcal{L}(f, g, \Pi_x) + \Omega(g)$$

- $g \in G$ là mô hình giải thích, được chọn từ tập G - tập các mô hình có thể diễn giải được, như mô hình tuyến tính, cây quyết định, hoặc falling rule lists. Chúng ta sẽ tìm một mô hình g từ các mô hình có thể diễn giải được.
- $\Omega(g)$ là hàm đo độ phức tạp của mô hình giải thích g. Chúng ta muốn tìm một g có độ phức tạp đủ thấp để con người có thể hiểu được.
- f là mô hình gốc cần được giải thích. Trong phân loại, f(x) cho biết xác suất (hoặc chỉ số nhị phân) rằng x thuộc về một lớp nhất định.
- $\pi_x(\mathbf{z})$ là hàm đo khoảng cách (proximity) giữa z và x. Hàm này định nghĩa “tính cục bộ” xung quanh x cần được giải thích.
- $L(f, g, \pi_x)$ là hàm đo độ không trung thành (unfaithfulness) của mô hình giải thích g so với mô hình gốc f, trong phạm vi cục bộ xác định bởi π_x .

2.2. Dữ liệu và phân tích

Do đề tài tập trung vào dự đoán mức tiêu thụ điện năng từ năng lượng mặt trời, các yếu tố ảnh hưởng đến sản lượng điện được xem xét một cách cẩn trọng. Theo nghiên cứu của Bhuiyan và cộng sự (2022), bức xạ mặt trời, nhiệt độ và các yếu tố thời tiết khác có ảnh hưởng đáng kể đến sản lượng điện mặt trời. Phân tích dữ liệu cho thấy sự khác biệt rõ rệt về tần suất và mức độ tiêu thụ hoặc sản xuất điện giữa các khoảng thời gian trong ngày, đặc biệt là theo chu kỳ ngày (daily) và chu kỳ 12 giờ (hourly). Cụ thể, sản lượng điện mặt trời có xu hướng tăng cao hơn vào ban ngày do cường độ bức xạ mặt trời lớn, trong khi mức tiêu thụ điện thường tăng vào buổi chiều tối, phản ánh nhu cầu sinh hoạt của người dùng. Ngoài ra, trong các ngày mùa hè với điều kiện nắng nóng, người dân có xu hướng gia tăng sản xuất điện từ năng lượng mặt trời. Bên cạnh yếu tố bức xạ, các yếu tố gián tiếp như điều kiện thời tiết, nhiệt độ và độ ẩm cũng góp phần tạo nên sự biến động trong sản lượng và mức tiêu thụ điện.

Dựa trên bộ dữ liệu Enefit đã cung cấp, có thể thấy rằng các yếu tố kinh tế vĩ mô bao gồm giá điện, giá nhiên liệu và các biến số ảnh hưởng đến năng lượng mặt trời như điều kiện thời tiết và đặc điểm nhân khẩu học đóng vai trò quan trọng trong việc xác định giá trị dự đoán của chủ đề nghiên cứu hiện tại. Tuy nhiên, kết quả phân tích khám phá dữ liệu (Exploratory Data Analysis – EDA) đối với tập dữ liệu trong giai đoạn từ ngày 01/09/2021 đến ngày 29/05/2023 cho thấy mức độ biến động cao cùng với quy mô dữ liệu lớn. Điều này có thể cản trở khả năng của mô hình dự đoán trong việc giải thích đầy đủ mức độ biến động của dữ liệu, qua đó ảnh hưởng đáng kể đến độ chính xác của dự đoán (Enefit, 2023).



Hình 1. Hình ảnh so sánh sự phân bố các yếu tố giá điện, giá gas với lượng điện năng trời tiêu thụ

Do đó, nhóm nghiên cứu quyết định cải thiện đặc điểm của bộ dữ liệu ban đầu bằng cách tạo ra các đặc trưng mới (feature engineering). Cụ thể, các đặc trưng độ trễ theo chuỗi thời gian (time-lag features) của các biến liên quan được bổ sung nhằm làm phong phú thông tin đầu vào cho mô hình. Quá trình hợp nhất các bảng dữ liệu ban đầu được thực hiện chủ yếu dựa trên trường datetime, đóng vai trò là khóa liên kết chính trong việc xây dựng tập dữ liệu phục vụ huấn luyện. Đồng thời, biến datetime cũng được phân rã thành các thành phần riêng biệt như ngày, tháng, năm và các đặc trưng thời gian khác, nhằm đáp ứng yêu cầu của các mô hình học máy. Bên cạnh đó, đối với các biến có tính biến động theo thời gian, bao gồm: giá điện, giá nhiên liệu, dữ liệu thời tiết... đã được tạo ra để nắm bắt xu hướng và tính phụ thuộc theo các khoảng trễ từ 1 đến 7 ngày. Ngoài ra, nhóm nghiên cứu cũng nhận thấy

ràng hành vi tiêu thụ điện chịu ảnh hưởng đáng kể từ các giá trị tiêu thụ trong quá khứ. Do đó, biến mục tiêu cũng được mở rộng thông qua việc bổ sung các độ trễ tương ứng, nhằm nâng cao khả năng dự báo của mô hình.

2.3. Phương pháp đề xuất

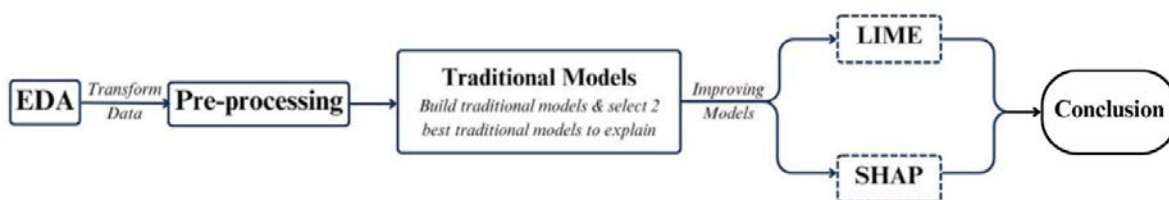
Phương pháp đề xuất của nhóm nghiên cứu là tổng hợp các đặc trưng quan trọng và cần thiết nhất cho các vấn đề liên quan bằng cách giải thích mô hình trong quá trình huấn luyện thông qua kỹ thuật LIME và SHAP trong lĩnh vực XAI. Từ đó có thể tóm tắt và trình bày quy trình xử lý dữ liệu cụ thể trong lĩnh vực nghiên cứu này. Ngoài các đặc trưng do bộ dữ liệu Enefit cung cấp, chúng tôi còn tạo ra các đặc trưng mới dựa trên độ trễ thời gian (time-lag), nhằm mở rộng thông tin và làm phong phú dữ liệu. Chúng tôi tin rằng, bên cạnh các đặc trưng cơ bản có sẵn trong bộ dữ liệu, các đặc trưng được tạo ra từ độ trễ thời gian cũng đóng vai trò quan trọng và ảnh hưởng đáng kể đến kết quả dự báo của mô hình.

Sau khi trải qua bước khám phá dữ liệu, và tạo ra được những đặc trưng mới dựa vào những phân tích từ bộ dữ liệu, nhóm nghiên cứu quyết định sẽ tiền xử lý dữ liệu theo nhiều cách khác nhau. Vì bộ dữ liệu này có số lượng quan sát rất lớn, nhóm quyết định chỉ chọn những quan sát từ năm 2022 trở đi nhằm tăng độ chính xác và ít lỗi trong quá trình dự đoán. Bên cạnh đó, để không làm mất đi sự phân phối vốn có của các đặc trưng, chúng tôi sẽ loại bỏ đi những quan sát có giá trị khuyết thiếu để không gây ảnh hưởng đến mô hình dự đoán.

Nhóm sử dụng hai phương pháp XAI chính để bổ trợ lẫn nhau: SHAP cho phân tích toàn cục dựa trên giá trị Shapley, giúp đánh giá mức độ đóng góp trung bình của các đặc trưng trên toàn bộ dữ liệu; trong khi đó, LIME được áp dụng cho việc phân tích cục bộ, tạo mô hình đơn giản xung quanh 3 “bản ghi đặc trưng” – 3 trung tâm của 3 cụm được sinh ra từ thuật toán gom cụm K-Means, từ đó giải thích chi tiết cách các thuộc tính ảnh hưởng đến từng dự đoán riêng lẻ. Cuối cùng, kết hợp cả 2 phương pháp nhằm có cái nhìn tổng quan về tầm quan trọng của từng biến cũng như hiểu cách mô hình hoạt động trên từng nhóm dữ liệu cụ thể, đảm bảo tính minh bạch, chính xác và hiệu quả trong việc giải thích mô hình.

2.4. Thiết lập thực nghiệm

Để đạt được kết quả chính xác nhất trong việc giải thích các mô hình đã nêu, nhóm nghiên cứu quyết định thực hiện các thí nghiệm toàn diện trên nhiều mô hình học máy khác nhau và giải thích chúng bằng các kỹ thuật XAI phổ biến. Quy trình mà chúng tôi tiến hành thực nghiệm được trình bày rõ trong Hình 2 dưới đây:



Hình 2. Tổng quan các bước thực hiện quy trình thực nghiệm

Tiền xử lý dữ liệu: Dữ liệu được chia tỉ lệ phổ biến là 66.7/33,3. Chúng tôi sẽ tiến hành huấn luyện mô hình dựa trên tập training và dùng tập validation để so sánh và đánh giá giữa các mô hình với nhau. Chúng tôi đã loại bỏ các quan sát thiếu dữ liệu vì nhận thấy việc này không làm thay đổi phân phối tổng thể của dữ liệu, sau đó chuẩn hóa để mô hình học hiệu quả hơn.

Đánh giá mô hình: Dựa vào những nghiên cứu trước, nhóm lựa chọn được những mô hình học máy được cho là đạt được hiệu suất cao với tham số được tinh chỉnh cơ bản sao cho phù hợp mà vẫn đảm bảo đạt được kết quả tốt trên tập kiểm định. Tiếp đến nhóm tiến hành đánh giá bằng kỹ thuật kiểm chứng chéo trên chỉ số R^2 và MAE và trực quan hóa nhằm tổng quát khả năng dự đoán cùng với độ phù hợp của các mô hình.

Xây dựng mô hình và điều chỉnh tham số: Các mô hình được xây dựng với tham số tối ưu, đồng thời nhóm cũng áp dụng các kỹ thuật tinh chỉnh để phân tích đặc trưng ảnh hưởng đến bài toán năng lượng đã được đề xuất trước đó.

2.5. Kết quả thực nghiệm

Kết quả trong Bảng 1 dưới đây cho thấy LightGBM là mô hình có hiệu suất tốt nhất trong tất cả các tập dữ liệu:

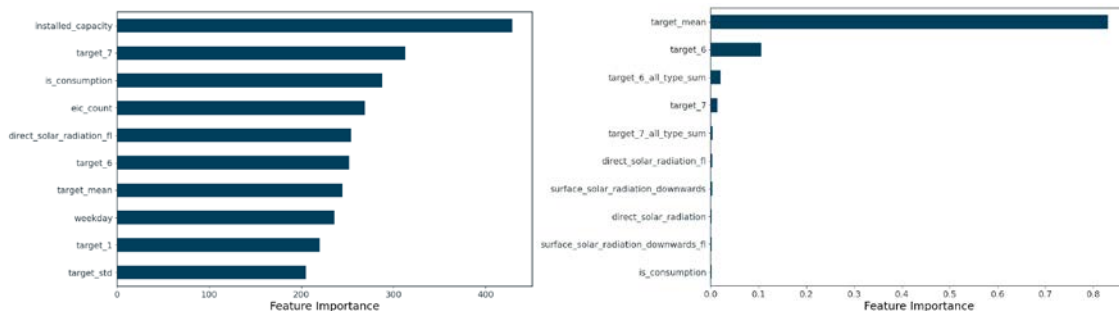
Bảng 1. So sánh hiệu suất các mô hình hồi quy trên các tập dữ liệu huấn luyện, xác thực và kiểm tra

	Train		Validation		Test	
	MAE	R^2	MAE	R^2	MAE	R^2
LightGBM	31.18	0.99	58.48	0.95	85.75	0.91
XGBoost	154.75	0.81	173.55	0.81	192.70	0.77
Decision Tree	44.36	0.97	72.50	0.93	94.48	0.89

Điều này chứng tỏ LightGBM có khả năng học tốt từ dữ liệu huấn luyện đồng thời vẫn giữ được khả năng khái quát hóa khi áp dụng lên dữ liệu chưa từng thấy. Trong khi đó, Decision Tree mặc dù thể hiện tương đối ổn định với R^2 dao động từ 0.97 đến 0.89, song MAE của nó trên tập Test (94.48) vẫn cao hơn đáng kể so với LightGBM (85.75), cho thấy nguy cơ hiện tượng overfitting nhẹ hơn và khả năng tổng quát kém hơn. Ngược lại, XGBoost thể hiện hiệu suất kém nhất với MAE lần lượt 154.75, 173.55 và 192.70 cùng R^2 chỉ khoảng 0.77–0.81. Sự chênh lệch lớn giữa MAE trên tập Train và Test của mô hình XGBoost cho thấy quá trình tối ưu hóa tham số có thể chưa đạt được sự cân bằng giữa độ phức tạp của mô hình và khả năng khái quát, dẫn đến hiệu quả dự đoán chưa cao như mong đợi. Ngoài ra, hiệu suất chưa tốt của XGBoost trong trường hợp này có thể bắt nguồn từ đặc thù của dữ liệu, chẳng hạn như mức độ nhiễu cao, tính phi tuyến mạnh hoặc sự mất cân bằng trong phân bố đặc trưng là những yếu tố mà LightGBM có vẻ xử lý tốt hơn, giúp tối ưu cả tốc độ học và độ chính xác trong dự đoán (Ke et al., 2017). Do đó, việc lựa chọn LightGBM và Decision Tree cho các phân tích tiếp theo là hợp lý, bởi chúng đại diện cho hai mức hiệu suất và đặc

tính mô hình khác biệt, giúp chúng ta so sánh sâu hơn về ưu nhược điểm và cơ chế vận hành của từng thuật toán.

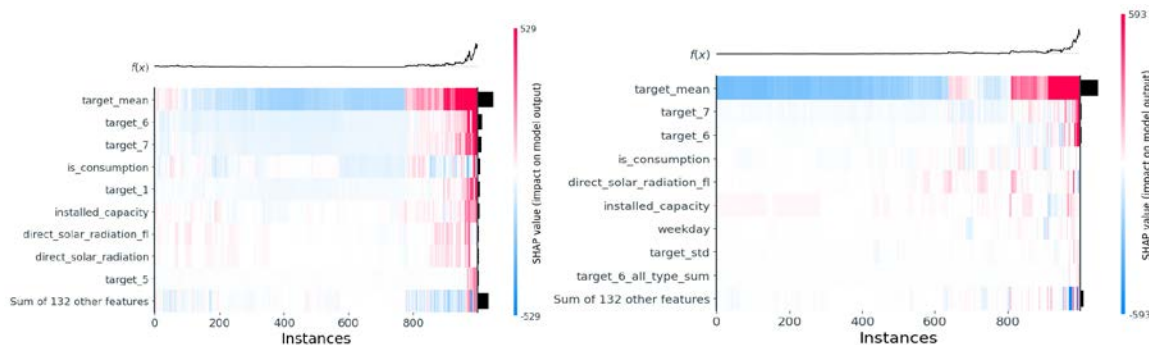
Dựa vào Hình 3, ta có thể thấy cách các mô hình xác định đặc trưng để đưa ra dự đoán cuối cùng cho bài toán là đa dạng với cách thức để tạo ra mô hình là dựa vào chỉ số Gini Impurity. Nhìn chung có thể thấy các yếu tố liên quan đến độ trễ của ngày cũng như các yếu tố liên quan đến thời gian và năng lượng bức xạ ảnh hưởng mạnh nhất đến kết quả của hai mô hình.



Hình 3. Top 10 đặc trưng quan trọng giữa 2 mô hình được chọn

Bên trái thể hiện cho mô hình LGBM. Bên phải thể hiện cho mô hình Decision Tree

2.6. Kết quả thực nghiệm với phương pháp SHAP



Hình 4. Top 10 đặc trưng quan trọng giữa 2 mô hình được chọn (theo SHAP).

Bên trái thể hiện cho mô hình LGBM. Bên phải thể hiện cho mô hình Decision Tree

Kết quả thực nghiệm cho thấy kết quả hồi quy của cả LightGBM (LGBM) và Decision Tree đều rất phụ thuộc vào các biến trễ của biến mục tiêu. Trong đó:

Biến “target_mean” là biến đóng góp nhiều nhất vào kết quả dự đoán của mô hình. Giá trị của biến càng cao kéo giá trị SHAP của mô hình lên cao (màu đỏ), giá trị biến càng thấp càng kéo giá trị SHAP của mô hình xuống (màu xanh). Các biến “target_6”, “target_7” cũng cho thấy hành vi tương tự như biến “target_mean” nhưng với mức độ ảnh hưởng đến đầu ra của mô hình nhỏ hơn đáng kể.

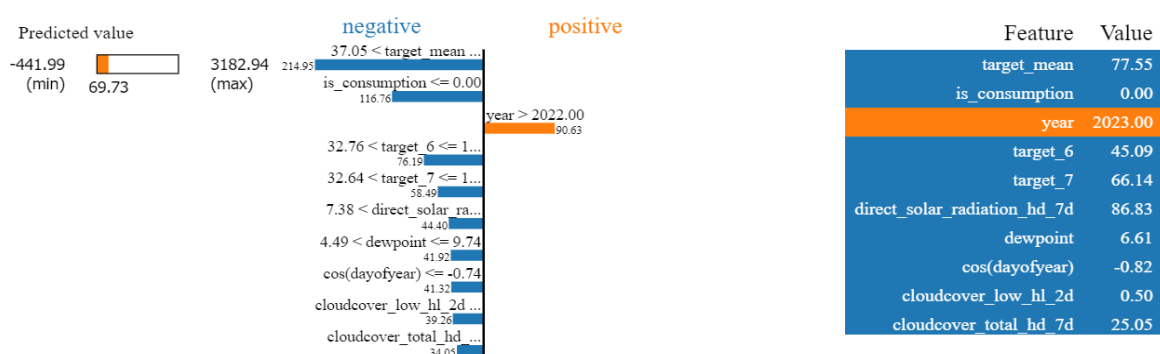
Một số biến liên quan đến yếu tố bức xạ mặt trời như “direct_solar_radiation_fl”, “installed_capacity” và biến liên quan đến loại hình doanh nghiệp như “is_consumption” cũng nằm trong nhóm 10 biến ảnh hưởng nhiều nhất đến kết quả dự đoán của mô hình. Tuy

nhiên, mức độ đóng góp của các biến này chủ yếu chỉ mang tính điều chỉnh nhẹ so với nhóm các biến trễ.

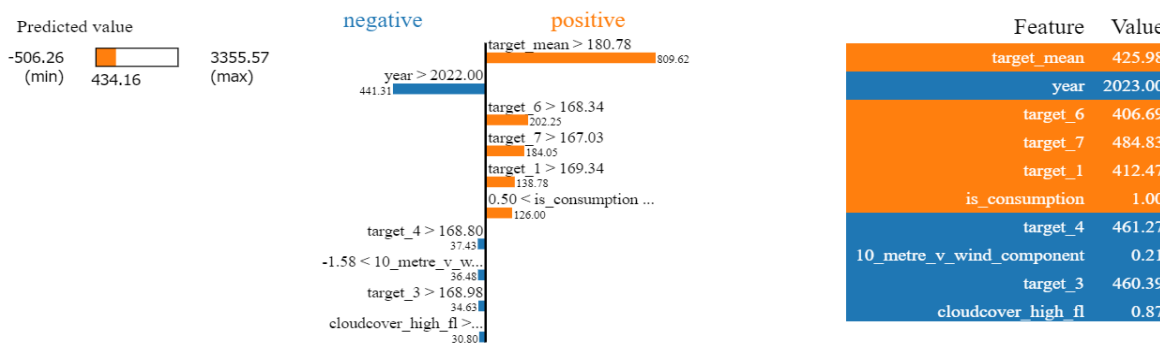
Kết luận lại, so sánh giữa hai mô hình LGBM và Decision Tree cho thấy rõ sự khác biệt. Với Decision Tree, SHAP values tập trung chủ yếu vào biến “target_mean”, phản ánh đặc điểm đặc trưng của mô hình, cho thấy những cây đơn thường ưu tiên một thuộc tính “đặc biệt” để chia nhánh. Ngược lại, mô hình LGBM cho thấy sự phân phối ảnh hưởng của các biến đa dạng hơn. Dù “target_mean” vẫn là yếu tố quan trọng nhất nhưng nhiều biến phụ trợ khác cũng đóng góp đáng kể vào quá trình dự đoán, giúp mô hình ít bị “quá phụ thuộc” vào một biến duy nhất.

2.7. Kết quả thực nghiệm với phương pháp LIME

Explainer for exp_Cluster 0 using LGBM:



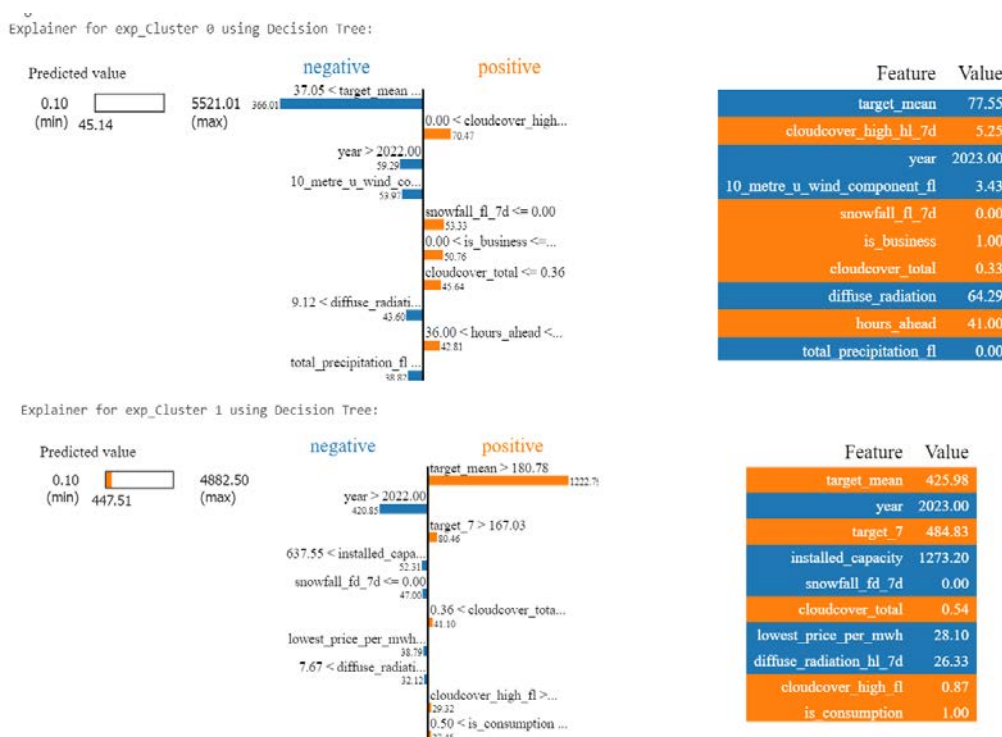
Explainer for exp_Cluster 1 using LGBM:



Hình 5. Phân tích LIME trên 3 bản ghi đại diện sử dụng LGBM

Trong cụm 0 (các bản ghi có giá trị dự đoán thấp), khi độ trễ của biến mục tiêu (ví dụ biến “target_mean”), các biến thời tiết (ví dụ biến “direct_solar_radiation_fl”, “dewpoint”) hoặc biến loại hình doanh nghiệp (is_consumption) có giá trị thấp, LIME cho thấy chúng đều có xu hướng kéo giá trị dự đoán xuống.

Thực nghiệm cho thấy kết quả từ việc ứng dụng phương pháp LIME cho các bản ghi cục bộ hoàn toàn khớp với kết quả từ việc ứng dụng phương pháp SHAP cho toàn bộ dữ liệu. Trong đó, các biến liên quan đến độ trễ luôn là quan trọng nhất, đặc biệt là biến “target_mean”, trong khi đó các biến còn lại chỉ đóng vai trò tinh chỉnh cho dự đoán của mô hình.



Hình 6. Phân tích LIME trên 3 bản ghi đại diện sử dụng Decision Tree

Sử dụng phương pháp LIME cho mô hình Decision Tree cũng cho ra kết quả không quá khác so với những gì quan sát được từ mô hình LGBM, cụ thể hơn:

Trong cụm 0 và cụm 1, một giá trị “target_mean” thấp sẽ kéo hoàn toàn kết quả đầu ra của mô hình xuống trong khi một giá trị “target_mean” cao sẽ kéo giá trị dự đoán của mô hình lên cao, và tương tự như LGBM, các biến còn lại chỉ mang ý nghĩa tinh chỉnh kết quả dự đoán. Tuy có sự thống trị của một biến duy nhất, song các biến còn lại tuy đóng góp ít vào kết quả mô hình hơn nhưng lại đa dạng hơn là LGBM.

Thực nghiệm với phương pháp LIME cho mô hình Decision Tree cho thấy có sự thống nhất về tính quan trọng, trong đó các biến “target_mean”, “target_6”, “target_7”, “installed_capacity”, “is_consumption”, “direct_solar_radiation_fl” vẫn là nhóm ảnh hưởng nhất. Bên cạnh đó, mô hình vẫn giữ được hướng tác động nhất quán như những gì quan sát được khi sử dụng phương pháp SHAP (giá trị của đặc trưng càng cao thì giá trị dự đoán của mô hình càng cao). Nhìn chung, Decision Tree linh hoạt hơn LGBM ở điểm có thể kết hợp nhiều đặc trưng, tuy nhiên giá trị dự đoán của mô hình này lại quá phụ thuộc vào một biến “target_mean” – điều này đồng thời củng cố lại quan sát của chúng ta ở Hình 3.

2.8. Nguyên nhân tiềm ẩn

Kết quả thí nghiệm cho thấy lịch sử tiêu thụ điện, dung lượng pin năng lượng mặt trời, và thời tiết là những yếu tố quan trọng trong dự đoán tiêu thụ và sản xuất điện của khách hàng Enefit. Các giá trị lịch sử như trung bình tiêu thụ (“target_mean”) và giá trị tại độ trễ 6-7 ngày (“target_6”, “target_7”) đóng vai trò lớn trong dự đoán, giúp khách hàng điều chỉnh hành vi sử dụng điện phù hợp với thực tế. Đặc biệt, “target_mean” có ảnh hưởng mạnh nhất trong các mô hình thử nghiệm, khẳng định tầm quan trọng của dữ liệu lịch sử.

Giả định rút ra là nếu khách hàng cân bằng tiêu thụ hoặc sản xuất điện trong 7 tuần, họ sẽ duy trì hành vi ổn định, với mức tiêu thụ cao đầu tuần thường được cân bằng vào cuối tuần. Ở bước phân tích dữ liệu khám phá và giải thích mô hình đã xác nhận độ tin cậy của giả định này.

SHAP và LIME đều là các phương pháp giải thích mô hình phổ biến nhưng có đặc điểm khác nhau. SHAP dựa trên lý thuyết trò chơi nên đảm bảo tính nhất quán và có thể giải thích cả ở mức toàn cục lẫn cục bộ, tuy nhiên chi phí tính toán cao và dễ gặp vấn đề khi các đặc trưng có tương quan. Trong khi đó, LIME xây dựng mô hình đơn giản quanh từng dự đoán để giải thích nên nhanh và dễ hiểu hơn, nhưng độ ổn định thấp do phụ thuộc vào lấy mẫu và chỉ cung cấp góc nhìn cục bộ, không phản ánh toàn bộ hành vi của mô hình.

3. Kết luận

Bài nghiên cứu tập trung vào việc giải quyết các vấn đề liên quan đến dự đoán lượng điện sản xuất và tiêu thụ của năng lượng mặt trời thông qua việc áp dụng XAI kết hợp với các mô hình học máy truyền thống. Bằng cách cải thiện dữ liệu thông qua việc tích hợp các yếu tố mới và độ trễ trong chuỗi thời gian, nghiên cứu đã nâng cao độ chính xác của các dự đoán, đồng thời đảm bảo tính phù hợp với các nguyên tắc kinh tế thực tiễn. Kết quả thực nghiệm cho thấy rằng việc bổ sung các yếu tố từ dữ liệu ban đầu đã giúp giảm thiểu đáng kể sai số, nâng cao hiệu quả và giảm thiểu tác động đến môi trường trong quản lý năng lượng. Tuy nhiên, nhóm nghiên cứu nhận thấy vẫn còn tồn tại những thách thức liên quan đến tài nguyên tính toán và khả năng thích ứng với sự biến động của dữ liệu thời gian thực.

❖ **Tuyên bố về quyền lợi:** Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.

TÀI LIỆU THAM KHẢO

- Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy and Buildings*, 165, 301–320.
- Allal, Z., Noura, H. N., Salman, O., & Chahine, K. (2024). Leveraging the power of machine learning and data balancing techniques to evaluate stability in smart grids. *Engineering Applications of Artificial Intelligence*, 133, 108304.
- Bhuiyan, M. A., Zhang, Q., Khare, V., Mikhaylov, A., Pinter, G., & Huang, X. (2022). Renewable Energy Consumption and Economic Growth Nexus—A Systematic Literature Review. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.878394>
- Frederiks, E., Stenner, K., & Hobman, E. (2015). The Socio-Demographic and Psychological Predictors of Residential Energy Consumption: A Comprehensive Review. *Energies*, 8(1), 573–609.
- Enefit. (2023). *Enefit – Predict energy behavior of prosumers*. Kaggle. <https://www.kaggle.com/competitions/predict-energy-behavior-of-prosumers>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: a highly efficient gradient boosting decision tree. *In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*.

- Kuzlu, M., Cali, U., Sharma, V., & Güler, Ö. (2020). Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools. *Ieee Access*, 8, 187814-187823. <https://doi.org/10.1109/ACCESS.2020.3031477>
- Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI*, 9, 100169. <https://doi.org/10.1016/j.egyai.2022.100169>
- Marcílio, W. E., & Eler, D. M. (2020, November). From explanations to feature selection: assessing SHAP values as feature selection mechanism. In *2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)* (pp. 340-347). Ieee.
- Molnar, C. (2024, July 31). *Interpretable Machine learning*.
- Peñasco, C., & Anadón, L. D. (2023). Assessing the effectiveness of energy efficiency measures in the residential sector gas consumption through dynamic treatment effects: Evidence from England and Wales. *Energy Economics*, 117, 106435.
- Sadorsky, P. (2010). The impact of financial development on energy consumption in emerging economies. *Energy policy*, 38(5), 2528-2535.
- Tolliver, C., Fujii, H., Keeley, A. R., & Managi, S. (2021). Green Innovation and Finance in Asia. *Asian Economic Policy Review*, 16(1), 67–87. <https://doi.org/10.1111/aepr.12320>
- Van den Broeck, G., Lykov, A., Schleich, M., & Suci, D. (2022). On the tractability of SHAP explanations. *Journal of Artificial Intelligence Research*, 74, 851-886.
- Zhao, H., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6), 3586–3592.

APPLYING EXPLAINABLE AI TO TRADITIONAL MACHINE LEARNING MODELS FOR SUSTAINABLE ECONOMIC CONSUMPTION

*Nguyen Thanh Huy**, *Dinh Tan Loc*, *Nguyen Nhat Quang*,
Tran Thi Kim Chi, *Hoang Quynh Anh*, *Le Di Khanh*

University of Economics, Ho Chi Minh City, Vietnam

**Corresponding author: Nguyen Thanh Huy – Email: huynt@ueh.edu.vn*

Received: November 25, 2024; Revised: April 17, 2025; Accepted: May 20, 2025

ABSTRACT

This study addresses the challenge of predicting solar power generation for Enefit by utilizing Explainable AI (XAI) in conjunction with traditional machine learning models to identify key factors. Accurate prediction is essential for the seamless integration of photovoltaic (PV) systems into the national grid. The research also proposes various machine learning methodologies to improve forecasting accuracy. By combining historical weather data, gas and electricity prices, and projected weather attributes, we developed a range of traditional models, such as linear, nonlinear, and LightGBM (LGBM) models. Experiments were carried out to determine the most effective models, allowing us to elucidate and investigate the factors influencing the issue across different models. The findings contribute to reducing power grid imbalance and improving the efficiency and reliability of renewable energy integration. Future research should focus on integrating real-time data and deep learning techniques to further refine the accuracy and effectiveness of predictions.

Keywords: Photovoltaic system (PV); renewable energy; traditional method; XAI