

# NGHIÊN CỨU THUẬT TOÁN TRÍ TUỆ NHÂN TẠO GIẢI BÀI TOÁN DỰ BÁO NĂNG LƯỢNG TIÊU THỤ

Lê Quốc Khương<sup>1</sup>, Huỳnh Phát Triển<sup>1</sup>, Trần Trung Khánh<sup>1</sup>,  
Phan Huỳnh Minh Thư<sup>2</sup>, Huỳnh Quốc Anh<sup>2</sup> và Trần Thị Cẩm Tiên<sup>2</sup>

<sup>1</sup>Trường Đại học Kỹ thuật Công nghệ Cần Thơ

<sup>2</sup>Sinh viên Khoa Điện - Điện tử, Trường Đại học Kỹ thuật Công nghệ Cần Thơ

Email: lqkhuong@ctuet.edu.vn

## Thông tin chung

Ngày nhận bài:

22/5/2025

Ngày nhận bài sửa:

10/7/2025

Ngày duyệt đăng:

25/7/2025

**Từ khóa:** Dự báo năng lượng, Hồi quy tuyến tính, Hồi quy vector hỗ trợ, Rừng ngẫu nhiên, Quản lý năng lượng.

## TÓM TẮT

Ứng dụng công nghệ học máy đang ngày càng được quan tâm trong việc tối ưu hóa hệ thống quản lý năng lượng (EMS). Nghiên cứu tập trung vào ứng dụng và so sánh hiệu suất của ba thuật toán hồi quy: Hồi quy tuyến tính (LR), Hồi quy vector hỗ trợ (SVR) và Rừng ngẫu nhiên (RF), nhằm giải quyết bài toán dự báo năng lượng tiêu thụ (điện lưới và điện mặt trời) tại tòa nhà khu C, Trường Đại học Kỹ thuật - Công nghệ Cần Thơ. Kết quả đánh giá bằng các chỉ số MSE, RMSE,  $R^2$  và MAE cho thấy mô hình hồi quy tuyến tính thể hiện hiệu suất tốt nhất trong cả dự báo tiêu thụ điện lưới và điện mặt trời, đặc biệt về mặt sai số. Phương pháp Rừng ngẫu nhiên cũng cho thấy khả năng dự báo tốt, trong việc nắm bắt xu hướng dữ liệu điện mặt trời ( $R^2 = 1,00$ ), trong khi SVR có sai số cao hơn. Việc dự báo năng lượng tiêu thụ trong các công trình cung cấp thông tin giá trị để hỗ trợ các quyết định đầu tư sửa chữa và cải tạo công trình một cách hiệu quả.

## 1. ĐẶT VẤN ĐỀ

Trong bối cảnh toàn cầu hóa và đô thị hóa diễn ra mạnh mẽ, các tòa nhà cao tầng và khu phức hợp hiện đại ngày càng trở thành những trung tâm hoạt động kinh tế, văn hóa và xã hội. Tuy nhiên, sự gia tăng về số lượng và quy mô của các công trình này cũng đặt ra những thách thức không nhỏ trong việc quản lý vận hành, đặc biệt là vấn đề tiêu thụ năng lượng. Năng lượng tiêu thụ trong các tòa nhà chiếm khoảng 40% năng lượng tiêu thụ toàn cầu [1]. Song, việc quản lý năng lượng một cách thiếu hiệu quả không chỉ gây ra những lãng phí tài chính khổng lồ mà còn tác động tiêu cực đến môi trường, góp phần vào biến đổi khí hậu và cạn kiệt nguồn tài nguyên.

Hệ thống quản lý năng lượng (EMS) là một hệ thống điều khiển số chuyên dụng, được lập trình để tự động hóa và tối ưu hóa hoạt động của các hệ thống tiêu thụ năng lượng trong tòa nhà. Về bản chất, EMS là một

nền tảng phần cứng và phần mềm tích hợp, cho phép giám sát, thu thập dữ liệu và điều khiển các thiết bị [2]. Dữ liệu được lưu trữ, xử lý, phân tích xu hướng, xác định khu vực tiêu thụ cao và tiềm năng tiết kiệm. EMS cảnh báo sự cố, vượt ngưỡng và tạo báo cáo trực quan, giúp nhà quản lý đánh giá hiệu suất, điều chỉnh vận hành thông minh, xây dựng chiến lược tiết kiệm và hướng tới phát triển bền vững.

Nhận thức rõ tầm quan trọng cấp thiết của việc sử dụng năng lượng bền vững và hiệu quả trong lĩnh vực xây dựng [3], bài báo này tập trung nghiên cứu và đề xuất một giải pháp dựa trên các thuật toán phân tích dự đoán tiên tiến [4]. Mục tiêu chính là xây dựng một khung quản lý năng lượng thông minh nhằm dự báo năng lượng tiêu thụ trong công trình [5], hệ thống này còn được trang bị khả năng phân tích và dự đoán nhu cầu năng lượng trong tương lai, từ đó hỗ trợ các nhà quản lý

đưa ra những quyết định tối ưu hóa vận hành, giảm thiểu chi phí và nâng cao hiệu quả sử dụng năng lượng tổng thể [3]. Việc tích hợp giải pháp này vào các hệ thống lưới điện thông minh sẽ góp phần thúc đẩy quá trình chuyển đổi năng lượng trong tương lai.

Việc tích hợp các thuật toán học máy tiên tiến vào hệ thống quản lý năng lượng (EMS) có tiềm năng lớn để nâng cao khả năng phân tích và dự đoán năng lượng tiêu thụ [5]. Nghiên cứu này tập trung vào ba thuật toán hồi quy phổ biến và hiệu quả: Hồi quy tuyến tính (LR), một phần mở rộng của hồi quy đơn giản để xác định mối quan hệ giữa các biến, đã được ứng dụng trong dự đoán tiêu thụ năng lượng [4], Hồi quy vector hỗ trợ (SVR), một kỹ thuật học máy có giám sát cân bằng giữa độ phức tạp của mô hình và sai số dự đoán, hiệu quả với dữ liệu đa chiều và lớn về năng lượng [5] và Rừng ngẫu nhiên (RF), một phương pháp học tập tổng hợp kết hợp nhiều cây quyết định để giảm thiểu phương sai mà không tăng độ lệch, được công nhận về hiệu quả và độ mạnh mẽ trong xử lý dữ liệu phức tạp, đa chiều, với các nghiên cứu cho thấy hiệu suất dự đoán năng lượng tốt hơn so với ANNs và SVR [5]. Mỗi thuật toán mang lại những ưu điểm riêng phù hợp với các dạng dữ liệu và mối quan hệ khác nhau ảnh hưởng đến tiêu thụ năng lượng.

Việc ứng dụng đồng thời ba thuật toán này trong hệ thống EMS mang lại nhiều lợi ích tiềm năng. Bằng cách so sánh hiệu suất dự đoán của từng thuật toán trên cùng một tập dữ liệu, các nhà quản lý có thể lựa chọn mô hình phù hợp nhất với đặc thù của tòa nhà và mục tiêu dự đoán cụ thể [5]. Hơn nữa, việc kết hợp kết quả dự đoán từ nhiều mô hình (Ensemble Forecasting) có thể giúp tăng cường độ chính xác và độ tin cậy của các dự báo năng lượng [4].

Bài báo này trình bày quy trình thu thập và tiền xử lý dữ liệu năng lượng tòa nhà, tương tự như các nghiên cứu khảo sát trước đó [3]. Sau đó, nhóm nghiên cứu mô tả quá trình xây dựng và huấn luyện các mô hình dự đoán

năng lượng dựa trên ba thuật toán hồi quy tuyến tính (Linear Regression - LR), hồi quy vector hỗ trợ (Support Vector Regression - SVR) và rừng ngẫu nhiên (Random Forest - RF), tương tự như các mô hình AI hiệu quả trong lĩnh vực xây dựng [5], được huấn luyện và kiểm tra bằng cách chia dữ liệu thành tập huấn luyện (80%) và tập kiểm tra (20%). Hiệu suất của các mô hình này được đánh giá bằng các chỉ số thống kê MSE (Mean Squared Error), RMSE (Root Mean Squared Error), R-squared ( $R^2$ ) và MAE (Mean Absolute Error) trên dữ liệu thực tế, tương tự như các nghiên cứu dự báo AI khác. Cuối cùng là phần thảo luận về kết quả và đề xuất các hướng nghiên cứu tiếp theo để tối ưu hóa quản lý năng lượng thông minh trong các tòa nhà hiện đại.

## 2. PHÁT TRIỂN CÁC MÔ HÌNH TRÍ TUỆ NHÂN TẠO (AI) TRONG DỰ BÁO NĂNG LƯỢNG TIÊU THỤ CỦA TÒA NHÀ

### 2.1. Mô hình dự báo hồi quy tuyến tính

Hồi quy tuyến tính là một thuật toán cơ bản nhưng mạnh mẽ trong việc mô hình hóa mối quan hệ tuyến tính giữa biến mục tiêu (tiêu thụ năng lượng) và các biến độc lập (ví dụ: thời gian, nhiệt độ môi trường, công suất hoạt động của các thiết bị chính) [4]. Với sự đơn giản và khả năng giải thích kết quả tốt, LR có thể cung cấp những hiểu biết ban đầu về các yếu tố có ảnh hưởng tuyến tính đến mức tiêu thụ năng lượng tòa nhà.

Mô hình hồi quy tuyến tính này được thể hiện bằng công thức [6]:

$$y = \beta_0 + \beta_1 x + \epsilon \quad (1)$$

Trong đó:  $\beta_0$  là hệ số chặn, thể hiện giá trị kỳ vọng của  $y$  khi  $x = 0$ .  $\beta_1$  là hệ số góc, biểu thị sự thay đổi của  $y$  tương ứng với một đơn vị thay đổi của  $x$ .  $\epsilon$  là thành phần sai số ngẫu nhiên, thể hiện sự chênh lệch giữa giá trị quan sát và đường hồi quy dự đoán.

Mô hình giả định rằng sai số  $\epsilon$  có trung bình bằng 0 và phương sai không đổi, nghĩa là phân phối của  $y$  tại mọi giá trị của  $x$  đều

giống nhau. Đường thẳng trong mô hình được gọi là đường hồi quy, biểu thị giá trị trung bình của  $y$  tại các giá trị khác nhau của  $x$  [6].

### 2.2. Mô hình dự báo hồi quy vector hỗ trợ

Hồi quy vector hỗ trợ là một thuật toán dựa trên học máy có giám sát để giải quyết các bài toán hồi quy, có khả năng mô hình hóa các mối quan hệ phi tuyến tính phức tạp trong dữ liệu [7]. SVR là một kỹ thuật mở rộng của thuật toán phân loại Máy vector hỗ trợ (SVM). SVR có thể ước tính các hàm với số chiều cao mà không nhất thiết đòi hỏi số lượng lớn các quan sát [7]. Bằng cách ánh xạ dữ liệu vào một không gian chiều cao hơn, SVR có thể tìm ra một hàm hồi quy tối ưu, ít nhạy cảm hơn với các giá trị ngoại lai và có khả năng khái quát hóa tốt trên dữ liệu mới [7]. Điều này đặc biệt hữu ích trong việc dự đoán năng lượng tiêu thụ [5], vốn có thể bị ảnh hưởng bởi nhiều yếu tố phi tuyến tính như thói quen sử dụng của người dùng hoặc hiệu suất thay đổi theo thời gian của các thiết bị [4].

Mục tiêu của SVR là tìm một hàm  $f(x)$  sao cho độ lệch của nó so với dữ liệu thực tế  $y$  nhỏ hơn một giá trị  $\epsilon$  cho càng nhiều điểm dữ liệu huấn luyện càng tốt, đồng thời giữ cho độ phức tạp của hàm ở mức thấp nhất [4]. Cho một tập dữ liệu huấn luyện  $\{(x_i, y_i)\}_{i=1}^l$ , với  $x_i \in \mathbb{R}^n$  là vector đặc trưng và  $y_i \in \mathbb{R}$  là giá trị mục tiêu. SVR tìm kiếm một hàm  $f(x)$  có dạng [8]:

$$f(x) = w^T \phi(x) + b \quad (2)$$

trong đó,  $w$  là vector trọng số,  $\phi(x)$  là hàm ánh xạ dữ liệu đầu vào lên không gian đa chiều,  $\phi(x) = x$ ,  $b$  là hệ số thiên lệch. Mục tiêu tối ưu hóa của SVR có thể được biểu diễn như sau:

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (3)$$

với các ràng buộc:

$$y_i - (w^T \phi(x_i) + b) \leq \epsilon + \xi_i$$

$$\begin{aligned} (w^T \phi(x_i) + b) - y_i &\leq \epsilon + \xi_i^* \\ \xi_i &\geq 0, \xi_i^* \geq 0, i=1, \dots, l \end{aligned}$$

Trong đó,  $\frac{1}{2} \|w\|^2$  là thành phần chính tắc hóa, nhằm kiểm soát độ phức tạp của mô hình,  $C > 0$  là tham số chính tắc hóa, xác định sự đánh đổi giữa việc giảm thiểu sai số huấn luyện và việc giữ cho độ phức tạp của mô hình thấp,  $\epsilon \geq 0$  định nghĩa một biên dung sai xung quanh hàm dự đoán. Các điểm dữ liệu nằm trong ống này không gây ra bất kỳ mất mát nào,  $\xi_i$  và  $\xi_i^*$  là các biến bù cho phép một số điểm dữ liệu nằm ngoài biên dung sai  $\epsilon$ .

### 2.3. Mô hình dự báo rừng ngẫu nhiên

Rừng ngẫu nhiên là một tập hợp các cây dự đoán, trong đó mỗi cây phụ thuộc vào các giá trị của một vector ngẫu nhiên được lấy mẫu độc lập và có cùng phân phối cho tất cả các cây trong rừng [9]. RF tạo ra một số lượng lớn các cây quyết định từ số lượng  $N$  mẫu huấn luyện. Đối với mỗi cây trong rừng, lấy mẫu khởi động (bootstrap sampling) được thực hiện để tạo các tập huấn luyện mới, trong khi các mẫu không được chọn được gọi là các tập ngoại vi (out-of-bag sets) [10].

Cung cấp một RF là một tập hợp các cây  $C$   $T_1(X), T_2(X), \dots, T_C(X)$ , trong đó  $X = x_1, x_2, \dots, x_m$  là một vector đầu vào có  $m$  chiều. Kết quả tổng hợp tạo ra  $C$  đầu ra được định nghĩa [5]:

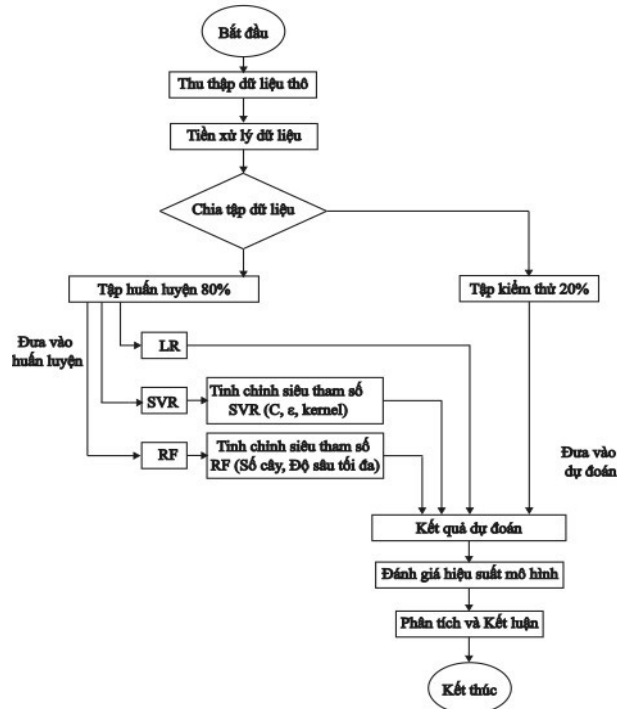
$$\begin{aligned} Y_{(pred\_1)} &= T_1(X), Y_{(pred\_2)} = \\ T_2(X), \dots, Y_{(pred\_C)} &= T_C(X) \end{aligned} \quad (4)$$

Trong đó,  $Y_{(pred\_C)}$  là giá trị dự đoán thu được từ số cây quyết định  $C$ .

Phương pháp lấy trung bình kết quả dự đoán từ tất cả các cây được tạo ngẫu nhiên này giúp giảm phương sai của mô hình và làm cho dự đoán ổn định hơn [10]. Đối với các bài toán phân loại, Random Forest thường sử dụng phương pháp bầu chọn đa số (majority voting). Mỗi cây sẽ bầu cho một lớp và lớp được nhiều cây bầu nhất sẽ là dự đoán cuối cùng [11].

Random Forest là một thuật toán mạnh mẽ vì nó kết hợp sức mạnh của nhiều cây quyết định [5] và giảm thiểu các vấn đề quá khớp mà các cây riêng lẻ có thể gặp phải [9]. Việc sử dụng các tập con ngẫu nhiên của dữ liệu (bootstrap sampling) và các thuộc tính ngẫu

nhiên (random feature selection) trong quá trình xây dựng cây giúp tạo ra sự đa dạng trong các cây, từ đó cải thiện khả năng khái quát hóa của mô hình [10]. Random Forest mang lại độ chính xác tốt và chạy hiệu quả trên các tập dữ liệu lớn [11].



**Hình 1. Lưu đồ quá trình triển khai các mô hình học máy trong dự báo năng lượng**

Để làm rõ quy trình triển khai các mô hình học máy đã được giới thiệu, Hình 1 minh họa chi tiết các bước áp dụng từ giai đoạn thu thập dữ liệu thô đến khi đưa ra kết quả dự đoán và đánh giá hiệu suất. Quy trình bắt đầu bằng việc thu thập dữ liệu năng lượng tiêu thụ. Sau đó, dữ liệu được tiền xử lý để đảm bảo chất lượng, bao gồm làm sạch, chuẩn hóa và tạo đặc trưng. Tiếp theo, dữ liệu được chia thành tập huấn luyện (80%) và tập kiểm thử (20%). Tập huấn luyện được sử dụng để đào tạo ba mô hình hồi quy. Đối với SVR và RF, các siêu tham số quan trọng như  $C$ ,  $\epsilon$ , kernel (cho SVR) và số cây, độ sâu tối đa (cho RF) được tinh chỉnh để tối ưu hóa hiệu suất. Mô hình đã huấn luyện sau đó được dùng để dự đoán trên tập kiểm thử. Cuối cùng, hiệu suất của các mô hình được đánh giá bằng các chỉ số như MSE, RMSE,

$R^2$  và MAE, từ đó rút ra phân tích và kết luận về khả năng dự báo của từng thuật toán.

### 3. ĐÁNH GIÁ TÍNH HIỆU QUẢ VÀ ĐỘ CHÍNH XÁC CÁC MÔ HÌNH AI TRONG BÀI TOÁN DỰ BÁO NĂNG LƯỢNG TIÊU THỤ

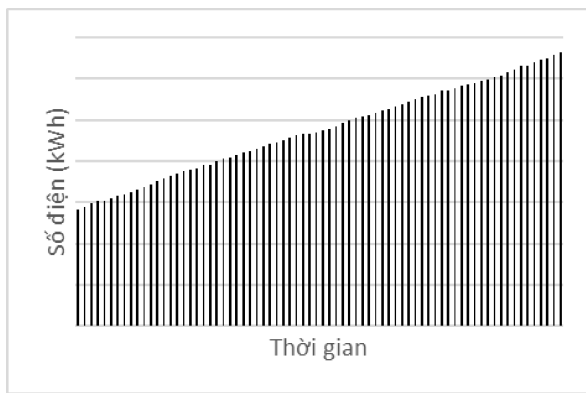
#### 3.1. Thu thập dữ liệu

Quá trình lấy mẫu dữ liệu tại tòa nhà khu C, Trường Đại học Kỹ thuật - Công nghệ Cần Thơ được thực hiện nhằm thu thập thông tin toàn diện về việc sử dụng năng lượng, từ đó hỗ trợ quản lý năng lượng hiệu quả. Dữ liệu được ghi nhận theo từng giờ, ngày và tháng trong khoảng thời gian từ 01/02/2025 đến 15/3/2025, với tổng cộng 74 mẫu, được thể hiện ở Hình 2. Mục tiêu chính là hiểu rõ các yếu tố ảnh hưởng đến mức tiêu thụ điện.

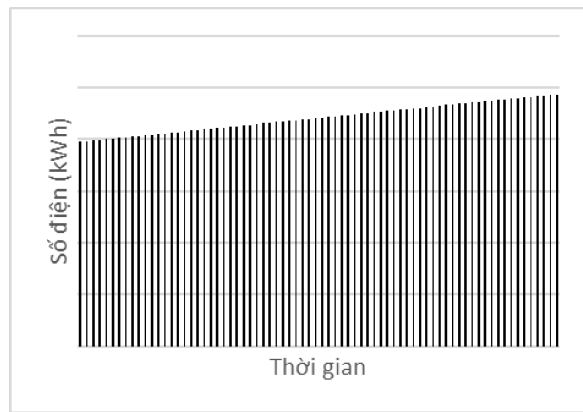
Để đánh giá chi tiết các nguồn tiêu thụ, dữ liệu được thu thập về số lượng và công suất của hệ thống chiếu sáng, làm mát, thiết bị văn phòng, cùng với tần suất sử dụng thang máy. Đặc biệt, dữ liệu tiêu thụ điện được thu thập khác nhau giữa Vùng 1 (điện lưới, chỉ số từ công tơ điện) và Vùng 2 (điện mặt trời, chỉ số từ inverter), cho phép so sánh hiệu quả sử dụng năng lượng giữa hai nguồn cung cấp.

Kết quả thu thập dữ liệu cho thấy sự khác biệt đáng chú ý về tiêu thụ điện giữa hai khu

vực. Vùng 1 (điện lưới) có mức tiêu thụ cao hơn và dao động mạnh hơn (87653-95248 kWh, trung bình 91486,60 kWh, độ lệch chuẩn 2178,06 kWh) so với Vùng 2 (điện mặt trời) (19785-24367 kWh, trung bình 22079,07 kWh, độ lệch chuẩn 1358,95 kWh). Mức tiêu thụ trung bình ở Vùng 1 cao gấp hơn bốn lần so với Vùng 2, cho thấy sự phụ thuộc lớn hơn vào lưới điện. Dù có sự dao động lớn, đường dự báo tuyến tính cho thấy xu hướng tiêu thụ điện lưới ở Vùng 1 là tương đối ổn định.



Vùng 1

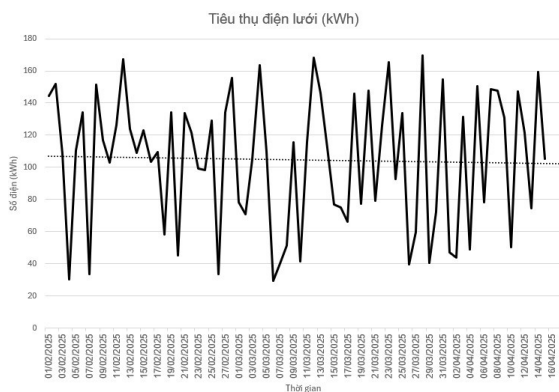


Vùng 2

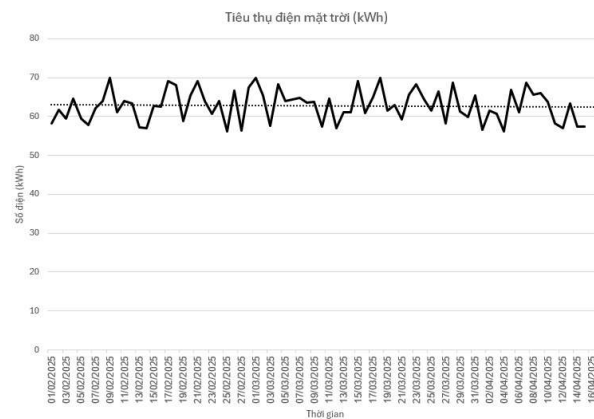
**Hình 2. Dữ liệu năng lượng tiêu thụ trong tòa nhà**

Độ lệch chuẩn cao ở Vùng 1 cho thấy nhu cầu điện lưới biến động đáng kể, do ảnh hưởng bởi hoạt động tòa nhà, thời gian và cường độ sử dụng thiết bị, số lượng người dùng, loại hình hoạt động, hiệu suất thiết bị, mùa vụ và thời tiết. Dữ liệu cho thấy sự dao động lớn hàng ngày (20 kWh - 180 kWh), phản ánh nhu cầu sử dụng điện không đồng

đều, dù đường dự báo tuyến tính cho thấy xu hướng tiêu thụ tổng thể tương đối ổn định quanh mức trung bình 100 kWh, nhưng không thể hiện hết các biến động chi tiết. Sự biến động hàng ngày lớn này gây ra thách thức đáng kể cho việc xây dựng các mô hình dự báo cho khu vực sử dụng điện lưới.



Vùng 1



Vùng 2

**Hình 3. Kết quả dự báo năng lượng tiêu thụ tại tòa nhà**

**Bảng 1. Thống kê năng lượng tiêu thụ trong tòa nhà**

Dữ liệu	Số liệu thống kê năng lượng tiêu thụ (kWh)			
	Giá trị lớn nhất	Giá trị nhỏ nhất	Giá trị trung bình	Độ lệch chuẩn
Vùng 1	95248	87653	91486,60	2178,06
Vùng 2	24367	19785	22079,07	1358,95

Dữ liệu tiêu thụ điện mặt trời ở Vùng 2 ít biến động hơn (60-70 kWh), cho thấy nhu cầu ổn định với xu hướng tuyến tính phẳng quanh mức trung bình 63 kWh. Mức tiêu thụ này thấp hơn đáng kể so với điện lưới (100 kWh), có thể do hệ thống mặt trời chỉ đáp ứng một phần nhu cầu.

Vùng 1 sử dụng điện lưới có mức tiêu thụ dao động mạnh theo ngày quanh mức trung bình 100 kWh, nhưng xu hướng tổng thể ổn định. Ngược lại, Vùng 2 dùng điện mặt trời ổn định hơn với mức tiêu thụ hàng ngày hẹp quanh 63 kWh và xu hướng tổng thể cũng ổn định. Sự ổn định của điện mặt trời được thể hiện qua độ lệch chuẩn thấp hơn, cho thấy tiềm năng làm nguồn cung cấp cơ bản, dù khả năng phát điện phụ thuộc vào các yếu tố bên ngoài.

### 3.2. Đánh giá hiệu suất mô hình bằng các chỉ số MSE, RMSE, $R^2$ và MAE

Để đánh giá hiệu quả của các mô hình dự báo AI, nghiên cứu này sử dụng bốn chỉ số thống kê phổ biến: sai số bình phương trung

bình, căn bậc hai của sai số bình phương trung bình, hệ số xác định và sai số tuyệt đối trung bình. Các chỉ số này được áp dụng để so sánh hiệu suất giữa ba thuật toán dự báo khác nhau: Hồi quy tuyến tính, Hồi quy vector hỗ trợ và Rừng ngẫu nhiên.

Chỉ số MSE đo lường bình quân của bình phương sai số giữa giá trị thực tế và giá trị dự báo, với giá trị càng nhỏ thể hiện độ chính xác cao hơn. RMSE, có đơn vị tương đương với dữ liệu gốc, được sử dụng để dễ diễn giải hơn và nhạy cảm với các sai số lớn.

MAE đo lường sai số trung bình tuyệt đối, không bị ảnh hưởng bởi giá trị cực đoan.  $R^2$  thể hiện tỷ lệ phương sai của dữ liệu đầu ra có thể được giải thích bởi mô hình, với giá trị gần 1 cho thấy sự phù hợp tốt. Kết quả đánh giá cho thấy LR hiệu quả với dữ liệu tuyến tính, SVR mạnh mẽ với quan hệ phi tuyến và RF có khả năng tổng quát hóa tốt. Việc kết hợp các chỉ số này giúp đưa ra đánh giá toàn diện về hiệu suất của từng mô hình.

**Bảng 2. Kết quả đánh giá các mô hình khi dùng điện lưới**

Chỉ số đánh giá độ chính xác của mô hình	Hồi quy tuyến tính – LR	Hồi quy vector hỗ trợ – SVR	Rừng ngẫu nhiên – RF
MSE	88,65	6834,99	547,00
RMSE	9,42	82,67	23,39
$R^2$	1,00	0,93	0,99
MAE	7,90	65,21	18,68

Bảng 2 so sánh hiệu suất dự đoán điện lưới của LR, SVR và RF dựa trên MSE, RMSE,  $R^2$  và MAE. Hồi quy tuyến tính (LR) cho thấy hiệu suất tốt nhất với MSE (88,65), RMSE (9,42) và MAE (7,90) thấp nhất, cùng  $R^2$  cao nhất (1,00), cho thấy khả năng giải thích dữ liệu hoàn hảo. Rừng ngẫu nhiên (RF) cũng đạt kết quả ấn tượng với MSE (547,00), RMSE (23,39),  $R^2$  (0,99) và MAE (18,68), cho thấy tiềm năng xử lý dữ liệu phức tạp. Ngược lại, Hồi quy vector hỗ trợ (SVR) có

hiệu suất thấp nhất với MSE (6834,99), RMSE (82,67),  $R^2$  (0,93) và MAE (65,21), cho thấy chưa phù hợp với bộ dữ liệu này. Nhìn chung, LR là mô hình hiệu quả nhất cho dự đoán điện lưới, tiếp theo là RF, trong khi SVR cần tinh chỉnh thêm.

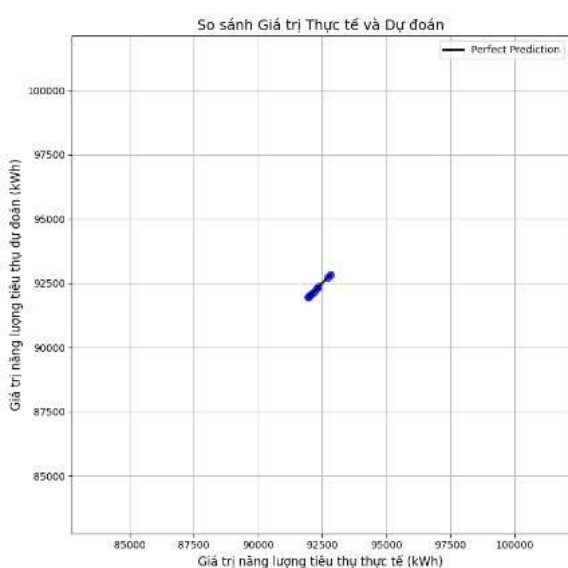
Bảng 3 so sánh hiệu suất dự đoán điện mặt trời của LR, SVR và RF dựa trên MSE, RMSE,  $R^2$  và MAE. Tương tự điện lưới, Hồi quy tuyến tính (LR) vẫn hiệu quả nhất với MSE (1992,78), RMSE (44,64) và MAE

(39,10) thấp nhất, cùng  $R^2$  hoàn hảo (1,00). Rừng ngẫu nhiên (RF) cũng đạt  $R^2$  là 1,00, cho thấy khả năng nắm bắt xu hướng tốt dù sai số cao hơn LR: MSE (8375,13), RMSE (91,52) và MAE (74,40). Hồi quy vector hỗ trợ (SVR) có sai số cao nhất: MSE (74303,04), RMSE (272,59) và MAE

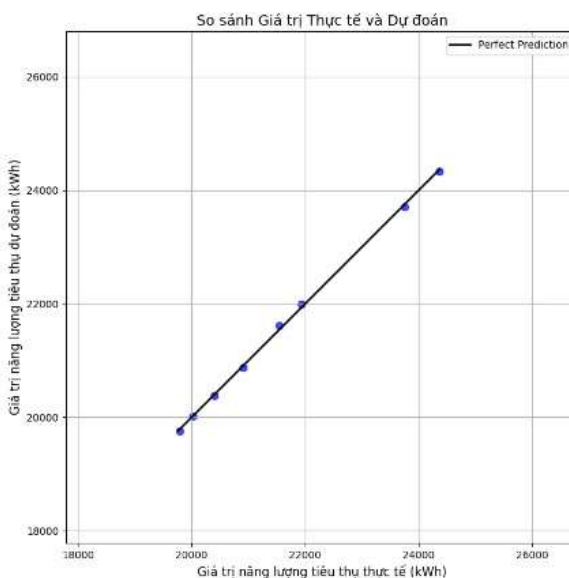
(186,54), nhưng  $R^2$  vẫn ở mức 0,97. Mô hình Hồi quy tuyến tính LR tiếp tục vượt trội trong dự đoán điện mặt trời, rừng ngẫu nhiên RF tốt trong nắm bắt xu hướng, còn hồi quy vector hỗ trợ SVR cần tối ưu thêm để cải thiện độ chính xác.

**Bảng 3. Kết quả đánh giá các mô hình khi dùng điện mặt trời**

Chỉ số đánh giá độ chính xác của mô hình	Hồi quy tuyến tính – LR	Hồi quy vector hỗ trợ – SVR	Rừng ngẫu nhiên – RF
MSE	1992,78	74303,04	8375,13
RMSE	44,64	272,59	91,52
$R^2$	1,00	0,97	1,00
MAE	39,10	186,54	74,40



Vùng 1



Vùng 2

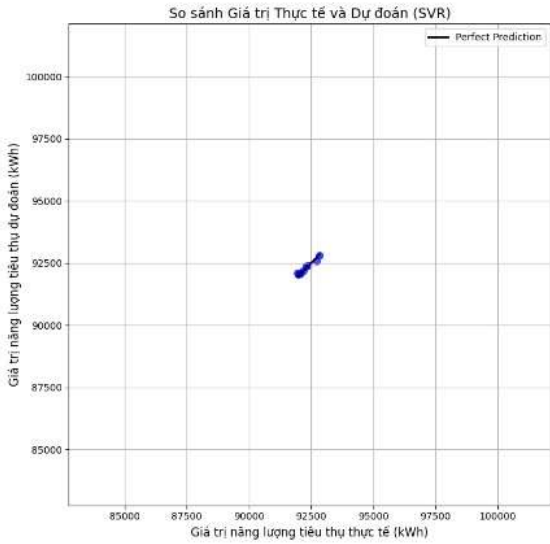
**Hình 4. Kết quả so sánh giá trị thực tế và dự đoán của mô hình LR**

Phân tích chỉ ra rằng các chỉ số sai số (MSE, RMSE, MAE) của tất cả các mô hình đều có xu hướng tăng khi dự đoán tiêu thụ điện từ nguồn điện mặt trời so với nguồn điện lưới. Sự khác biệt này gợi ý rằng việc mô hình hóa tiêu thụ năng lượng từ điện mặt trời có thể gặp nhiều thách thức hơn. Có khả năng là do tính chất biến động vốn có của quá trình phát điện mặt trời, vốn phụ thuộc vào các yếu tố khó kiểm soát như thời tiết, ánh sáng và hiệu suất của hệ thống pin. Trong khi đó, điện lưới có nguồn cung ổn định và dễ dự đoán hơn.

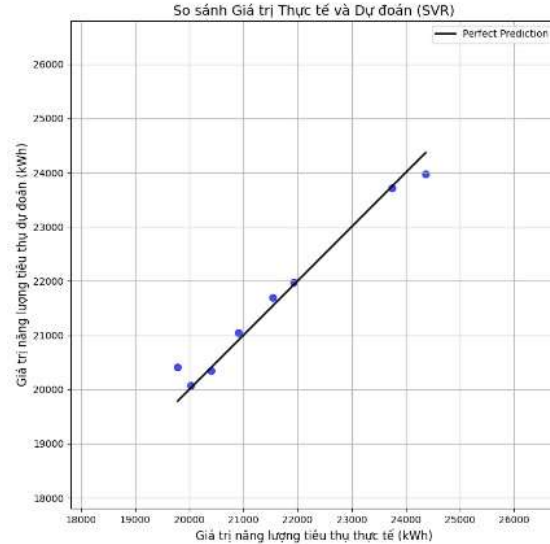
Tuy nhiên, điều đáng chú ý là giá trị  $R^2$  của cả mô hình LR và RF vẫn duy trì ở mức cao, thậm chí đạt 1,00 trong cả hai trường hợp. Điều này cho thấy rằng, bất kể nguồn năng lượng, các mô hình này vẫn hiệu quả trong việc giải thích phương sai của dữ liệu. Nói cách khác, các mô hình có khả năng nắm bắt được các mối quan hệ giữa các biến đầu vào và đầu ra, cho phép dự đoán chính xác xu hướng tiêu thụ năng lượng của hệ thống điện mặt trời, mặc dù có thể gặp khó khăn trong việc dự đoán giá trị tuyệt đối tại một thời điểm cụ thể. Hiệu suất cao của LR cho cả dữ

liệu điện lưới và điện mặt trời cho thấy sự phù hợp của nó, có thể do dữ liệu có tính tuyến tính đáng kể.  $R^2$  cao của RF cho thấy tiềm năng mô hình hóa các mối quan hệ phức tạp, với sai số cao hơn có thể là sự đánh đổi cho

khả năng tổng quát hóa. Hiệu suất thấp của SVR cho thấy cần tinh chỉnh thêm. Tóm lại, LR hoạt động tốt nhất, RF là lựa chọn tốt cho dữ liệu phức tạp, còn SVR cần tối ưu hóa. Dữ liệu điện mặt trời có vẻ khó dự đoán hơn.

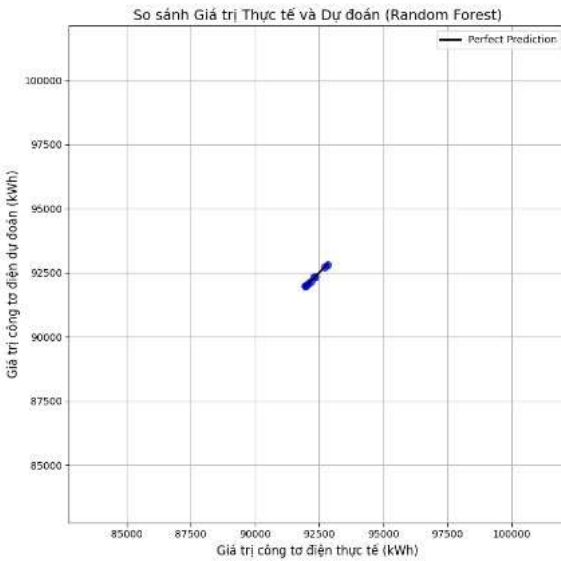


Vùng 1

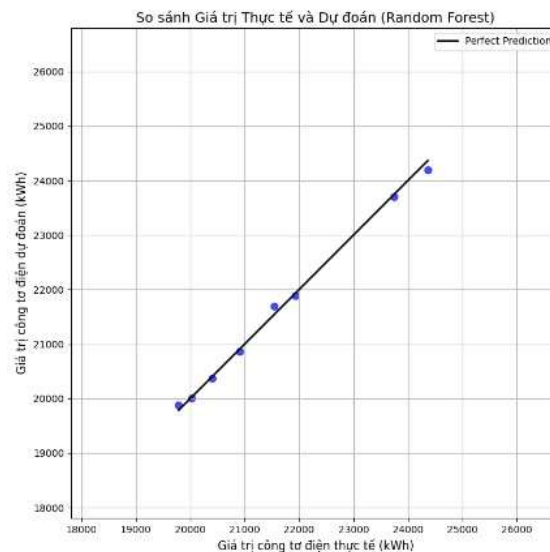


Vùng 2

Hình 5. Kết quả so sánh giá trị thực tế và dự đoán của mô hình SVR



Vùng 1



Vùng 2

Hình 6. Kết quả so sánh giá trị thực tế và dự đoán của mô hình RF

Do đó, để nâng cao độ chính xác dự báo cho điện mặt trời, các hướng nghiên cứu tiếp theo cần tập trung vào việc thu thập và tích hợp thêm các yếu tố đầu vào có ảnh hưởng trực tiếp

đến hiệu suất hệ thống pin mặt trời, chẳng hạn như cường độ bức xạ mặt trời, góc nghiêng của tấm pin, nhiệt độ môi trường xung quanh và tình trạng bụi bẩn trên bề mặt tấm pin.

Việc áp dụng các kỹ thuật học máy nâng cao hơn, đặc biệt là các mô hình học sâu có khả năng xử lý dữ liệu chuỗi thời gian phức tạp và biến động mạnh như mạng nơ-ron hồi quy (RNN) hay mạng nơ-ron tích chập (CNN) kết hợp với các lớp LSTM (Long Short-Term Memory), có thể giúp nắm bắt tốt hơn các mối quan hệ phi tuyến tính và sự phụ thuộc thời gian trong dữ liệu điện mặt trời. Ngoài ra, việc nghiên cứu các phương pháp kết hợp mô hình (ensemble methods) có thể khai thác điểm mạnh của từng thuật toán, từ đó cải thiện độ ổn định và chính xác tổng thể của hệ thống dự báo năng lượng.

#### 4. KẾT LUẬN

Nghiên cứu này đã ứng dụng và so sánh hiệu suất của LR, SVR và RF trong dự báo tiêu thụ điện lưới và điện mặt trời. Kết quả cho thấy LR vượt trội về độ chính xác và sai số cho cả hai nguồn, gợi ý sự phù hợp của mô hình tuyến tính với dữ liệu nghiên cứu. RF cũng thể hiện khả năng dự báo tốt, đặc biệt trong việc nắm bắt xu hướng dữ liệu điện mặt trời với  $R^2$  là 1,00. Ngược lại, SVR có hiệu suất kém hơn về sai số, cho thấy cần điều chỉnh siêu tham số kỹ lưỡng hơn.

Phân tích xu hướng tiêu thụ năng lượng cho thấy điện lưới có mức tiêu thụ cao và biến động hơn so với điện mặt trời, được thể hiện rõ qua đường dự báo tuyến tính. Việc dự báo chính xác các xu hướng này quan trọng để tối ưu hóa phân bổ và sử dụng năng lượng, giảm chi phí và nâng cao hiệu quả vận hành tòa nhà.

Các hướng nghiên cứu tiềm năng để cải thiện dự báo năng lượng và hướng tới hệ thống quản lý năng lượng thông minh bao gồm mở rộng dữ liệu, nghiên cứu các thuật toán khác, kết hợp mô hình, lựa chọn đặc trưng phù hợp, tinh chỉnh siêu tham số (đặc biệt cho SVR) và xem xét các mô hình học sâu.

#### Tài liệu tham khảo

- [1] Yang L, Yan H, Lam J.C. Thermal comfort and building energy consumption implications - A review. *Appl Energy*. 2014; 115: 164 -173.
- [2] Ellis P.G, Torcellini P.A, Crawley D.B. Simulation of Energy Management Systems in EnergyPlus. In: *Proceedings of the Building Simulation 2007*, 3-6 September, Beijing, China. Golden, CO: National Renewable Energy Laboratory; 2007.
- [3] Nguyễn Thành Trung, Phạm Văn Tới. Nghiên cứu, đề xuất suất tiêu thụ điện năng cho văn phòng làm việc: Áp dụng tính toán cho văn phòng làm việc tại Hà Nội. *Tạp chí Khoa học và Công nghệ Xây dựng (STCE) - NUCE*. 2018; 12(2): 59 - 64.
- [4] Trần Đức Học, Lê Tân Tài. Tối ưu hóa ước tính mức tiêu thụ năng lượng trong các tòa nhà dựa trên các thuật toán trí tuệ nhân tạo. *Tạp chí Khoa học Công nghệ Xây dựng (KHCVXD) - ĐHXD*. 2020; 14(1V): 35 - 45.
- [5] Ngô Ngọc Tri, Phan Anh Đức, Mai Anh Đức. Phát triển mô hình trí tuệ nhân tạo nhằm dự báo năng lượng tiêu thụ của tòa nhà tại thành phố Đà Nẵng. *Tạp chí Khoa học và Công nghệ-Đại học Đà Nẵng*. 2024; 22(11A): 20 - 24.
- [6] Montgomery D.C, Peck E.A, Vining G.G. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley; 2012.
- [7] Zhang F, O'Donnell L.J. Support vector regression. In: *Machine Learning: Methods and Applications to Brain Disorders*. Elsevier. 2020; 123 - 140.
- [8] Qiu X, Zhang L, Suganthan P.N et al. Oblique random forest ensemble via least square estimation for time series forecasting. *Inf Sci (N Y)*. 2017; 420: 249 - 262.

- [9] Breiman L. Random forests. *Mach Learn.* 2001; 45: 5 - 32.
- [10] Pham A.D, Ngo N.T, Ha Truong T.T et al. Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *J Clean Prod.* 2020; 260: 121082.
- [11] Awad M, Khanna R. Support vector regression. In: *Efficient learning machines: Theories, concepts, and applications for engineers and system designers.* Apress. 2015; 67 - 80.

## RESEARCH ON ARTIFICIAL INTELLIGENCE ALGORITHMS FOR BUILDING ENERGY CONSUMPTION PREDICTION

### ABSTRACT

*The application of machine learning technologies is increasingly gaining attention for the optimization of Energy Management Systems (EMS). This research focuses on the application and performance comparison of three regression algorithms: Linear Regression (LR), Support Vector Regression (SVR), and Random Forest (RF), to address the problem of forecasting energy consumption (grid power and solar power) at Building C, Can Tho University of Technology. The evaluation results, using the metrics of MSE, RMSE,  $R^2$ , and MAE, indicate that the Linear Regression model exhibits the best performance in forecasting both grid power and solar power consumption, particularly in terms of error. Random Forest also demonstrates good forecasting capabilities, effectively capturing the trend of solar power data ( $R^2 = 1.00$ ), while SVR shows higher error rates. Predicting energy consumption in buildings provides valuable information to support efficient investment decisions for building repair and renovation.*

**Keywords:** *Building energy management, energy consumption forecasting, linear regression, support vector regression, random forest*