**ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH TRƯỜNG ĐẠI HỌC SÀI GÒN**

**Logo

Description automatically generated----------🙡🕮🙣----------**

**LUẬN VĂN NGHIÊN CỨU KHOA HỌC**

**PHƯƠNG PHÁP VÀ NGHIÊN CỨU KHOA HỌC TRONG CNTT**

**ĐỀ TÀI:**

**PHÂN TÍCH LUẬN VĂN NGHIÊN CỨU KHOA HỌC**

**SHORT-TERM ELECTRICITY LOAD FORECASTING WITH MACHINE LEARNING**

**MÃ HỌC PHẦN: 841324**

**NHÓM MÔN HỌC: 02**

**HỌC KỲ:** **02**

**NHÓM BÁO CÁO: 03**

**GVHD**: **ĐỖ NHƯ TÀI**

**THÀNH VIÊN**:

**3120410101 Đỗ Xuân Quý Dương**

**3121410515 Đăng Ngọc Đoan Trang**

**3123410306 Đình Đào Thái Sơn**

**3123410294 Nguyễn Anh Quân (Nhóm trưởng)**

**TP HỒ CHÍ MINH, THÁNG 4 NĂM 2025**

**ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**Logo

Description automatically generated----------🙡🕮🙣----------**

**LUẬN VĂN NGHIÊN CỨU KHOA HỌC**

**PHƯƠNG PHÁP VÀ NGHIÊN CỨU KHOA HỌC TRONG CNTT**

**ĐỀ TÀI:**

**PHÂN TÍCH LUẬN VĂN NGHIÊN CỨU KHOA HỌC**

**SHORT-TERM ELECTRICITY LOAD FORECASTING WITH MACHINE LEARNING**

**MÃ HỌC PHẦN: 841324**

**NHÓM MÔN HỌC: 02**

**HỌC KỲ:** **02**

**NHÓM BÁO CÁO: 03**

**GVHD**: **ĐỖ NHƯ TÀI**

**THÀNH VIÊN**:

**3120410101 Đỗ Xuân Quý Dương**

**3121410515 Đăng Ngọc Đoan Trang**

**3123410306 Đình Đào Thái Sơn**

**3123410294 Nguyễn Anh Quân (Nhóm trưởng)**

**TP HỒ CHÍ MINH, THÁNG 4 NĂM 2025**

**Mục lục**

[I. Giới thiệu 2](#_Toc195911415)

[1. Đặt vấn đề 2](#_Toc195911416)

[2. Phạm vi và mục đích nghiên cứu 3](#_Toc195911417)

[II. Cơ sở lý thuyết 4](#_Toc195911418)

[1. Mô hình học máy (Machine Learning Models) 4](#_Toc195911419)

[2. Các yếu tố ảnh hưởng đến phụ tải điện 5](#_Toc195911420)

[III. Phương pháp thu thập và xử lý dữ 8](#_Toc195911421)

[1. Thu thập dữ liệu 8](#_Toc195911422)

[2. Làm sạch và chuẩn hóa dữ liệu 9](#_Toc195911423)

[3. Tạo đặc trưng và lựa chọn đặc trưng 9](#_Toc195911424)

[4. Xử lý và phân chia dữ liệu 9](#_Toc195911425)

[5. Cross-validation và mô hình học máy 9](#_Toc195911426)

[6. Các mô hình học máy sử dụng 10](#_Toc195911427)

[7. Tìm kiếm và tối ưu hóa hyperamerter 10](#_Toc195911428)

[IV. Nhận xét và đánh giá 11](#_Toc195911429)

[1. Kết quả đạt được 11](#_Toc195911430)

[2. Hạn chế của công trình 13](#_Toc195911431)

[V. Tài liệu kham khảo 14](#_Toc195911432)

**LỜI CAM ĐOAN**

Nhóm xin cam đoan rằng bài báo cáo với tiêu đề **"Phân tích luận văn nghiên cứu khoa học: Short-Term Electricity Load Forecasting with Machine Learning"** là kết quả của quá trình làm việc nhóm nghiêm túc, được thực hiện bởi các thành viên trong nhóm. Tất cả nội dung trong báo cáo đều được xây dựng dựa trên việc đọc hiểu, phân tích bài báo khoa học gốc và các tài liệu có liên quan, hoàn toàn không sao chép từ bất kỳ nguồn nào khác mà không trích dẫn.

Trong suốt quá trình thực hiện, mọi người trong nhóm đã cùng nhau tìm hiểu, thảo luận và thống nhất nội dung trình bày nhằm đảm bảo tính khách quan và trung thực của báo cáo. Các ý kiến nhận xét, đánh giá trong bài đều xuất phát từ quá trình làm việc và tư duy phân tích của nhóm.

Nhóm sẽ chịu hoàn toàn trách nhiệm về tính trung thực, nguyên bản và tính chính xác của bài báo cáo này. Mọi tài liệu tham khảo đều đã được ghi rõ nguồn.

Nhóm xin chân thành cảm ơn giảng viên hướng dẫn đã tận tình hỗ trợ và tạo điều kiện để nhóm hoàn thành tốt bài báo cáo này

**TÓM TẮT LUẬN VĂN**

Bài luận văn này tập trung phân tích nội dung chính của công trình nghiên cứu **“Short-Term Electricity Load Forecasting with Machine Learning”**, được công bố trên tạp chí *Information* năm 2021 bởi hai tác giả **Ernesto Aguilar Madrid** và **Nuno António**.

Nhóm thực hiện đã tổng hợp, hệ thống lại các nội dung cốt lõi của bài báo, bao gồm: mục tiêu nghiên cứu, các mô hình học máy được sử dụng, phương pháp thu thập và xử lý dữ liệu đầu vào, kết quả đạt được và các hạn chế còn tồn tại.

Thông qua việc phân tích công trình này, nhóm không chỉ tìm hiểu sâu hơn về khả năng ứng dụng của các mô hình học máy như **XGBoost, SVR, ANN, MLR hay Random Forest** trong bài toán dự báo phụ tải điện ngắn hạn, mà còn rèn luyện kỹ năng đọc hiểu tài liệu chuyên ngành, kỹ năng phân tích – tổng hợp và đánh giá các nghiên cứu khoa học.

Kết quả phân tích cho thấy nghiên cứu đã đưa ra được mô hình XGBoost có hiệu năng nổi bật trong điều kiện dữ liệu thực tế tại Panama, đồng thời mở ra nhiều hướng phát triển tiếp theo trong việc tích hợp dữ liệu thời gian thực hoặc ứng dụng học sâu. Bài báo cáo là một bước khởi đầu giúp nhóm sinh viên tiếp cận các phương pháp học máy ứng dụng trong ngành năng lượng, làm tiền đề cho các đề tài nghiên cứu sau này.

# Giới thiệu

Trong những năm gần đây, trí tuệ nhân tạo (AI) nói chung và học máy (Machine Learning) nói riêng đã trở thành xu hướng nghiên cứu nổi bật trong nhiều lĩnh vực, từ tài chính, y tế, giao thông đến năng lượng. Đặc biệt, với khả năng xử lý dữ liệu lớn và học hỏi từ dữ liệu lịch sử, các mô hình học máy đang được xem là công cụ hữu hiệu để giải quyết những bài toán dự báo – vốn mang tính phức tạp và có nhiều yếu tố ảnh hưởng.

Trong bối cảnh đó, việc ứng dụng Machine Learning vào dự báo phụ tải điện – một bài toán quan trọng trong ngành điện lực – không chỉ có giá trị về mặt học thuật mà còn mang ý nghĩa thiết thực đối với kinh tế và xã hội

## Đặt vấn đề

Trong thời đại công nghiệp hóa, hiện đại hóa và chuyển đổi số, năng lượng – đặc biệt là điện năng – đóng vai trò then chốt trong mọi hoạt động kinh tế, công nghiệp và đời sống xã hội. Việc quản lý và dự báo nhu cầu tiêu thụ điện chính xác trong từng khung giờ, từng ngày là yếu tố quyết định đến hiệu quả vận hành hệ thống điện quốc gia. Dự báo không chính xác có thể dẫn đến tình trạng thiếu điện gây gián đoạn sản xuất, hoặc dư thừa gây lãng phí tài nguyên và chi phí vận hành.

Bên cạnh đó, với sự phát triển mạnh mẽ của các nguồn năng lượng tái tạo như điện mặt trời, điện gió – vốn có tính không ổn định – thì nhu cầu có một hệ thống dự báo phụ tải thông minh và chính xác càng trở nên cấp thiết. Trong bối cảnh đó, các mô hình trí tuệ nhân tạo, đặc biệt là học máy (Machine Learning) đã chứng minh tiềm năng vượt trội trong việc phân tích dữ liệu lớn, rút trích quy luật và đưa ra các dự báo có độ chính xác cao.

Xuất phát từ thực tiễn trên, nhóm lựa chọn phân tích bài báo “Short-Term Electricity Load Forecasting with Machine Learning” – một nghiên cứu điển hình trong việc ứng dụng các thuật toán học máy vào lĩnh vực dự báo phụ tải điện ngắn hạn. Thông qua việc tìm hiểu và phân tích bài báo này, nhóm mong muốn nắm được cách tiếp cận khoa học hiện đại, hiểu rõ hơn về quy trình xây dựng và đánh giá mô hình dự báo, từ đó có thể vận dụng vào những nghiên cứu sau này hoặc áp dụng trong thực tiễn.

# Phạm vi và mục đích nghiên cứu

Bài báo cáo tập trung vào việc tìm hiểu, phân tích và đánh giá công trình nghiên cứu có tiêu đề **“Short-Term Electricity Load Forecasting with Machine Learning”** – một bài báo khoa học được công bố trên tạp chí Information vào năm 2021. Phạm vi thực hiện của nhóm chủ yếu xoay quanh việc làm rõ nội dung chính của bài báo, bao gồm: mục tiêu nghiên cứu, cơ sở lý thuyết được sử dụng, phương pháp thu thập và xử lý dữ liệu, kết quả mà nhóm tác giả đạt được, cũng như những hạn chế còn tồn tại trong quá trình nghiên cứu. Nhóm không đi sâu vào khía cạnh kỹ thuật hay toán học của từng thuật toán, mà chú trọng ở mức độ phân tích – tổng hợp và diễn giải nội dung theo hướng dễ hiểu, phù hợp với sinh viên chuyên ngành.

Mục tiêu của báo cáo là giúp nhóm sinh viên tiếp cận và hiểu rõ hơn về một hướng ứng dụng cụ thể của Machine Learning trong thực tế – đó là dự báo phụ tải điện ngắn hạn. Thông qua việc nghiên cứu bài báo, nhóm không chỉ nâng cao kiến thức chuyên môn mà còn rèn luyện kỹ năng đọc – hiểu tài liệu khoa học, kỹ năng làm việc nhóm và trình bày nội dung nghiên cứu một cách có hệ thống.

* 1. **Bố cục luận văn**

Để đảm bảo tính logic và thuận tiện trong việc trình bày nội dung, bài báo cáo được chia thành sáu phần chính. Phần đầu tiên là **Giới thiệu**, nhằm giới thiệu bối cảnh, lý do lựa chọn đề tài, phạm vi – mục tiêu thực hiện và bố cục tổng quan của bài viết.

Phần thứ hai là **Thông tin tổng quan bài báo**, cung cấp các thông tin cơ bản như tên đề tài, tác giả, nơi công bố và thời gian công bố. Tiếp theo là **Nội dung phân tích**, đây là phần trọng tâm của bài báo cáo, trình bày chi tiết các khía cạnh như mục tiêu nghiên cứu, cơ sở lý thuyết, phương pháp thu thập và xử lý thông tin, kết quả đạt được và các hạn chế của công trình.

Phần thứ tư là **Nhận xét và đánh giá**, trong đó nhóm sẽ đưa ra một số quan điểm cá nhân về chất lượng, mức độ ứng dụng và những điểm nổi bật của nghiên cứu. Phần thứ năm là **Kết luận**, nhằm tổng kết lại những gì đã tìm hiểu được, đồng thời nêu ra một vài định hướng nghiên cứu mở rộng.

Cuối cùng là **Tài liệu tham khảo**, nơi nhóm liệt kê các tài liệu đã sử dụng trong quá trình thực hiện báo cáo nhằm đảm bảo tính học thuật và minh bạch thông tin.

1. **Cơ sở lý thuyết**

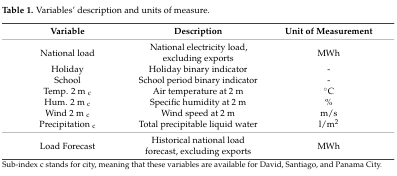
Dự báo phụ tải điện ngắn hạn (STLF) là quá trình ước tính nhu cầu tiêu thụ điện trong khoảng thời gian từ 1 đến 168 giờ. STLF đóng vai trò quan trọng trong việc lập kế hoạch và vận hành hệ thống điện, giúp giảm thiểu chi phí sản xuất và phân phối điện, đặc biệt khi có sự tham gia của các nguồn năng lượng tái tạo không ổn định như điện mặt trời và điện gió.

## Mô hình học máy (Machine Learning Models)

Trong bài báo, các tác giả đã áp dụng nhiều mô hình học máy khác nhau để dự báo phụ tải điện ngắn hạn. Các mô hình này đều có khả năng học từ dữ liệu quá khứ và đưa ra dự báo cho tương lai gần. Dưới đây là các mô hình học máy được sử dụng:

* **XGBoost (Extreme Gradient Boosting)**: Là một thuật toán học máy mạnh mẽ dựa trên phương pháp boosting, giúp tối ưu hóa kết quả dự báo. XGBoost đã được chứng minh là có hiệu suất cao trong việc xử lý dữ liệu với nhiều biến động và đặc biệt trong các bài toán dự báo chuỗi thời gian.
* **Artificial Neural Networks (ANN)**: Là mô hình học sâu (deep learning) mô phỏng hoạt động của mạng lưới thần kinh trong não người. ANN có khả năng tự học và tìm ra các mối quan hệ phức tạp trong dữ liệu, rất hữu ích trong các bài toán dự báo không tuyến tính.
* **Support Vector Regression (SVR)**: Là một thuật toán học máy dựa trên lý thuyết SVM (Support Vector Machine). SVR hiệu quả trong việc phân tích dữ liệu với các quan hệ phi tuyến và có thể dự báo với độ chính xác cao, đặc biệt khi dữ liệu có nhiễu.
* **Random Forest (RF)**: Là một mô hình ensemble (hội tụ nhiều mô hình con) sử dụng phương pháp cây quyết định (decision trees). RF có khả năng xử lý dữ liệu với nhiều đặc tính khác nhau và đưa ra các dự báo ổn định.
* **Multiple Linear Regression (MLR)**: Là mô hình hồi quy tuyến tính đa biến, được sử dụng để tìm mối quan hệ tuyến tính giữa các biến độc lập và biến phụ thuộc. MLR giúp dự đoán phụ tải điện dựa trên các yếu tố như nhiệt độ, độ ẩm, và các yếu tố thời gian.
  1. **Các yếu tố ảnh hưởng đến phụ tải điện**

Để xây dựng các mô hình học máy, nhóm tác giả đã sử dụng các biến liên quan đến phụ tải điện và các yếu tố ảnh hưởng. Dưới đây là các biến chính cùng với mô tả và đơn vị đo lường của chúng, được trình bày trong Table 1 của bài báo:



*Bảng các biến giá trị trong bài báo*

**National load**: Phụ tải điện quốc gia, không bao gồm xuất khẩu điện (Đơn vị: MWh)

**Holiday**: Biến chỉ thị về ngày lễ (Đơn vị: nhị phân: 1 nếu là ngày lễ, 0 nếu không phải ngày lễ)

**School**: Biến chỉ thị về kỳ học (Đơn vị: nhị phân: 1 nếu là kỳ học, 0 nếu không phải)

**Temp. 2 mc**: Nhiệt độ tại 2m (Đơn vị: °C)

**Hum. 2mc**: Độ ẩm tại 2m (Đơn vị: %)

**Wind 2mc**: Tốc độ gió tại 2m (Đơn vị: m/s)

**Precipitation c**: Lượng mưa (Đơn vị: l/m²)

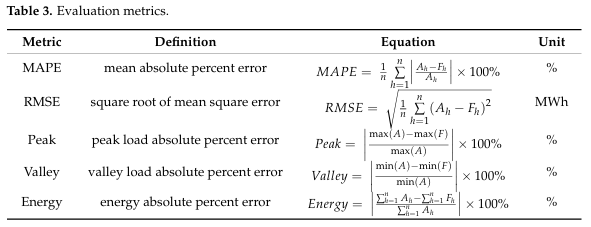
Các biến này đều có ảnh hưởng quan trọng đến mức tiêu thụ điện năng, vì nhu cầu sử dụng điện có thể thay đổi theo mùa, thời tiết, và các yếu tố xã hội (như kỳ học và ngày lễ).

* 1. **Đo lường hiệu xuất mô hình**

Để đánh giá hiệu suất của các mô hình học máy trong việc dự báo phụ tải điện, tác giả đã sử dụng hai chỉ số chính:

**MAPE (Mean Absolute Percentage Error)**: Là sai số tuyệt đối trung bình theo tỷ lệ phần trăm, đo lường độ chính xác của mô hình dự báo. MAPE có thể giúp đánh giá mức độ sai lệch của dự báo so với giá trị thực tế, và giá trị càng thấp thì mô hình càng chính xác.

**RMSE (Root Mean Square Error)**: Là sai số bình phương trung bình căn, cũng là một chỉ số quan trọng để đánh giá độ chính xác của mô hình. RMSE đánh giá sự sai lệch giữa giá trị thực tế và giá trị dự báo, với giá trị càng thấp cho thấy mô hình có độ chính xác càng cao.



Bảng đo lường hiệu suất mô hình

**Giải thích các ký hiệu trong công thức**:

* Ah ​ là giá trị thực tế của phụ tải điện tại thời điểm h.
* Fh  là giá trị dự báo của phụ tải điện tại thời điểm h.
* n là số lượng điểm dữ liệu (hoặc số giờ cần dự báo).
* max(A) và min(A) là giá trị cực đại và cực tiểu của dữ liệu thực tế trong khoảng thời gian dự báo.
* max(F) và min(F) là giá trị cực đại và cực tiểu của dữ liệu dự báo trong khoảng thời gian dự báo.
* Chỉ số **Peak** đánh giá độ chính xác của mô hình trong việc dự báo giá trị cao điểm (peak) của phụ tải điện. Trong thực tế, phụ tải điện có thể đạt giá trị cực đại vào những thời điểm cụ thể (ví dụ, vào giờ cao điểm). Chỉ số này giúp xem xét mức độ sai lệch giữa giá trị dự báo và giá trị thực tế của phụ tải tại điểm cực đại.
* Chỉ số Valley đo lường độ chính xác của mô hình trong việc dự báo giá trị thấp điểm (valley) của phụ tải điện. Phụ tải điện có thể đạt giá trị cực tiểu vào những thời điểm nhất định (ví dụ, vào ban đêm hoặc khi có ít người sử dụng điện). Chỉ số này đánh giá mức độ sai lệch giữa giá trị dự báo và giá trị thực tế tại điểm thấp nhất.
* Chỉ số **Energy** đo lường độ chính xác tổng thể của mô hình trong việc dự báo tổng năng lượng tiêu thụ trong khoảng thời gian dự báo (tổng phụ tải điện trong một ngày hoặc nhiều giờ). Chỉ số này giúp đánh giá độ chính xác tổng thể của mô hình bằng cách tính tổng sai lệch giữa giá trị thực tế và dự báo của tất cả các giờ trong chuỗi thời gian.

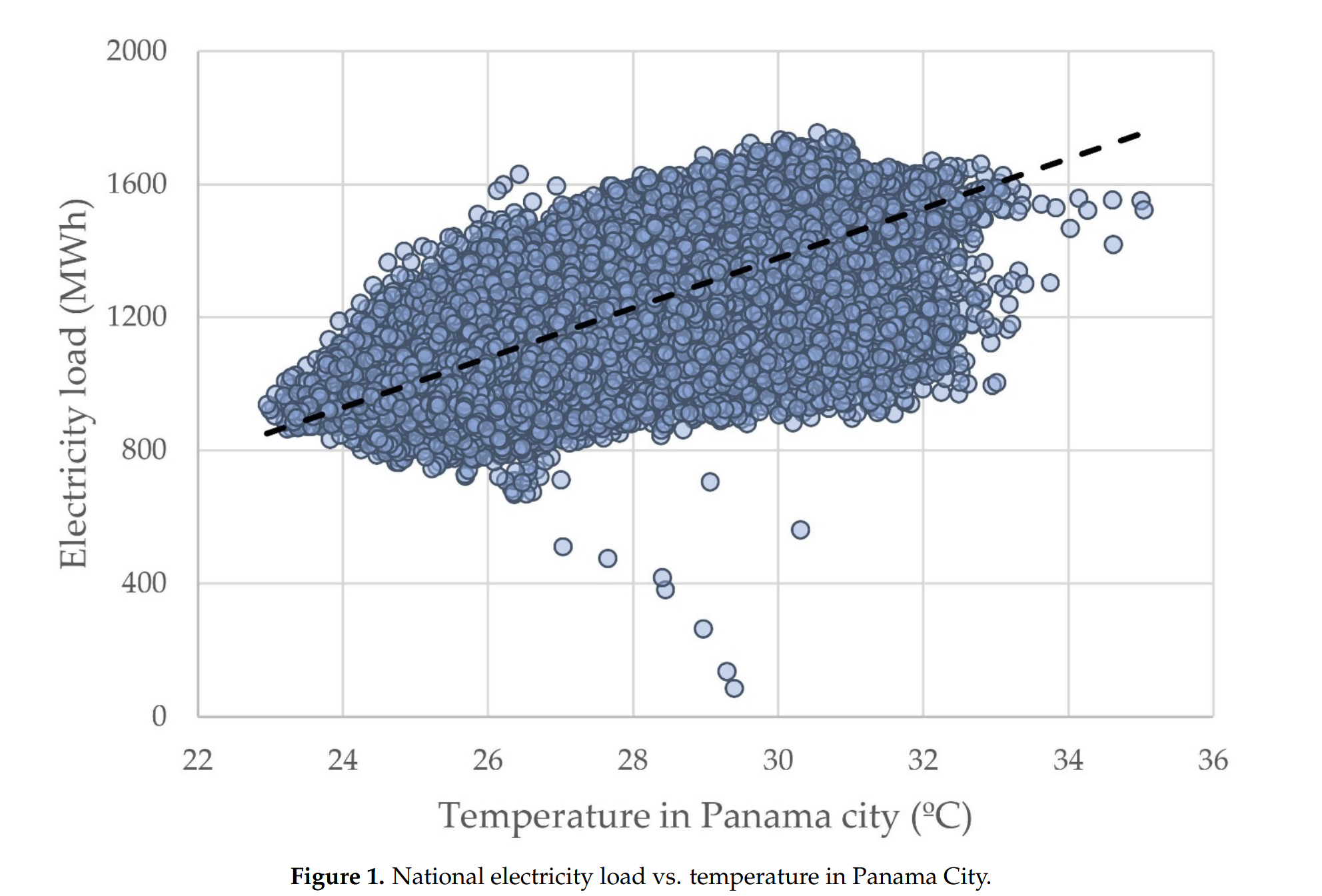
Có thể nói, **MAPE** và **RMSE** là các chỉ số tổng quát giúp đánh giá hiệu suất của mô hình, các chỉ số **Peak**, **Valley**, và **Energy** có thể **trừ sai số** trong các tình huống đặc thù và giúp đưa ra cái nhìn toàn diện hơn về hiệu quả của mô hình.

# Phương pháp thu thập và xử lý dữ

* 1. **Thu thập dữ liệu**

Dữ liệu được thu thập từ các nguồn công khai liên quan đến tải điện, nhiệt độ, độ ẩm, gió và mưa trong khoảng thời gian từ **năm 2015 đến 2020**. Dữ liệu này được lấy từ các cơ sở dữ liệu giám sát phụ tải điện và các trạm khí tượng địa phương.

Một ví dụ về mối liên hệ giữa **tải điện** và **nhiệt độ** tại Panama City được thể hiện trong **Figure 1** dưới đây, cho thấy sự biến động của tải điện theo sự thay đổi của nhiệt độ.



*Hình: Tải điện quốc gia so với nhiệt độ tại thành phố Panama*

1. **Làm sạch và chuẩn hóa dữ liệu**

Sau khi thu thập, dữ liệu sẽ được làm sạch để loại bỏ các giá trị thiếu hoặc ngoại lệ. Dữ liệu về **nhiệt độ**, **độ ẩm**, **gió**, **mưa** được chuẩn hóa về cùng một đơn vị đo để có thể so sánh trực tiếp.

1. **Tạo đặc trưng và lựa chọn đặc trưng**

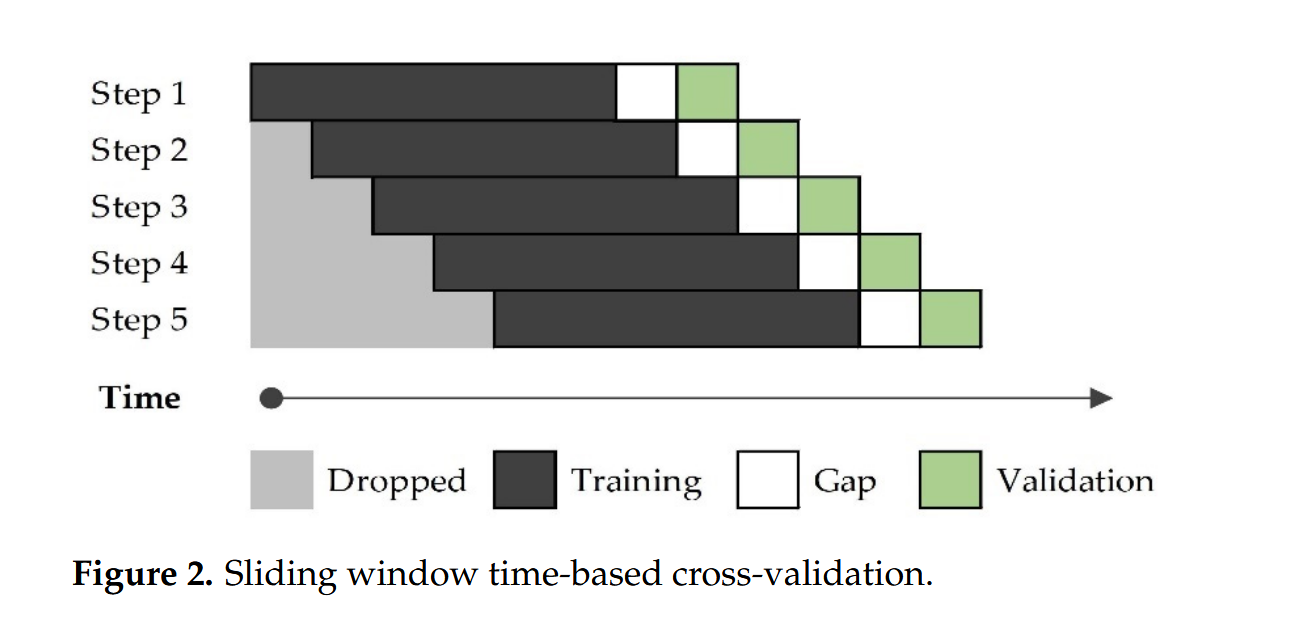
Một số đặc trưng quan trọng như **nhiệt độ cao nhất** và **lượng mưa trung bình** được tạo ra và lựa chọn để cải thiện hiệu quả dự báo. Các đặc trưng này giúp làm tăng tính dự đoán của mô hình.

1. **Xử lý và phân chia dữ liệu**

Sau khi làm sạch và chuẩn hóa dữ liệu, nó sẽ được chia thành **dữ liệu huấn luyện** và **dữ liệu kiểm tra**. Tỷ lệ chia dữ liệu thường là 70% cho huấn luyện và 30% cho kiểm tra.

1. **Cross-validation và mô hình học máy**

Để đảm bảo tính chính xác của mô hình, bài báo áp dụng kỹ thuật **cross-validation theo cửa sổ trượt theo thời gian**. **Hình 2** dưới đây minh họa cách thức hoạt động của phương pháp này, giúp đánh giá mô hình qua các phân đoạn thời gian liên tiếp. Trong đó mỗi cửa sổ trượt thời gian sẽ giúp huấn luyện mô hình dựa trên các phân đoạn liên tiếp của dữ liệu lịch sử và đánh giá hiệu suất của mô hình trên các cửa sổ tiếp theo.

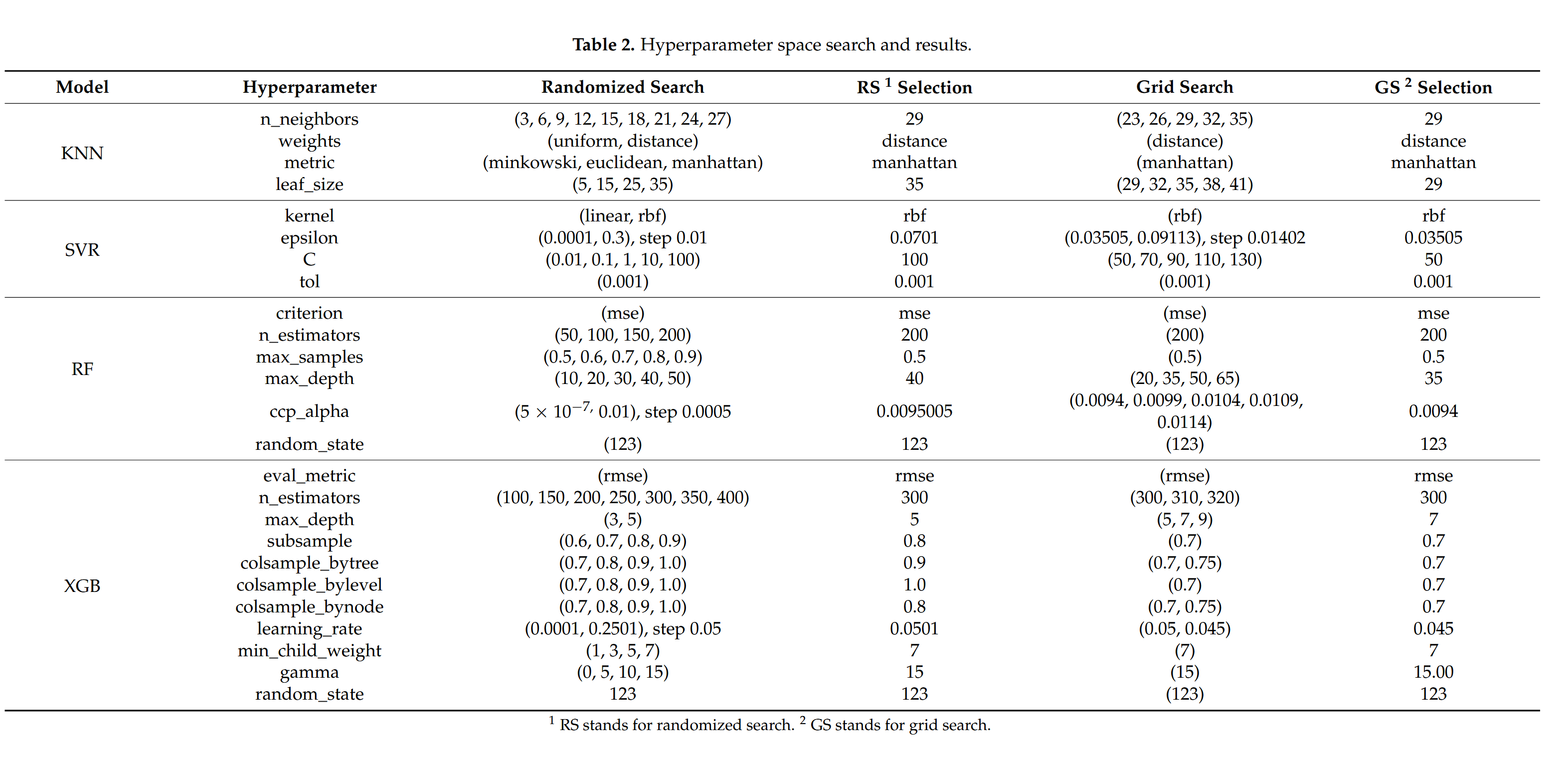


1. **Các mô hình học máy sử dụng**

Các mô hình học máy như **K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Random Forest (RF) và XGBoost (XGB)** đã được thử nghiệm trong nghiên cứu này. Các mô hình này sử dụng nhiều tham số khác nhau như số lượng láng giềng, kernel, n\_estimators, max\_depth, và learning\_rate để tìm ra mô hình tối ưu.

1. **Tìm kiếm và tối ưu hóa hyperamerter**

Để đảm bảo tính chính xác của mô hình, bài báo áp dụng kỹ thuật **cross-validation theo cửa sổ trượt theo thời gian**. **Bảng 2** dưới đây minh họa cách thức hoạt động của phương pháp này, giúp đánh giá mô hình qua các phân đoạn thời gian liên tiếp.



*Tìm kiếm không gian hyperparameter và kết quả*

Bảng 2 tổng hợp kết quả của quá trình tìm kiếm hyperparameter cho các mô hình học máy khác nhau (KNN, SVR, RF, XGB) sử dụng các kỹ thuật **Randomized Search** và **Grid Search** để tối ưu hóa các tham số quan trọng. Các giá trị tham số tối ưu cho từng mô hình được ghi chú rõ ràng trong bảng

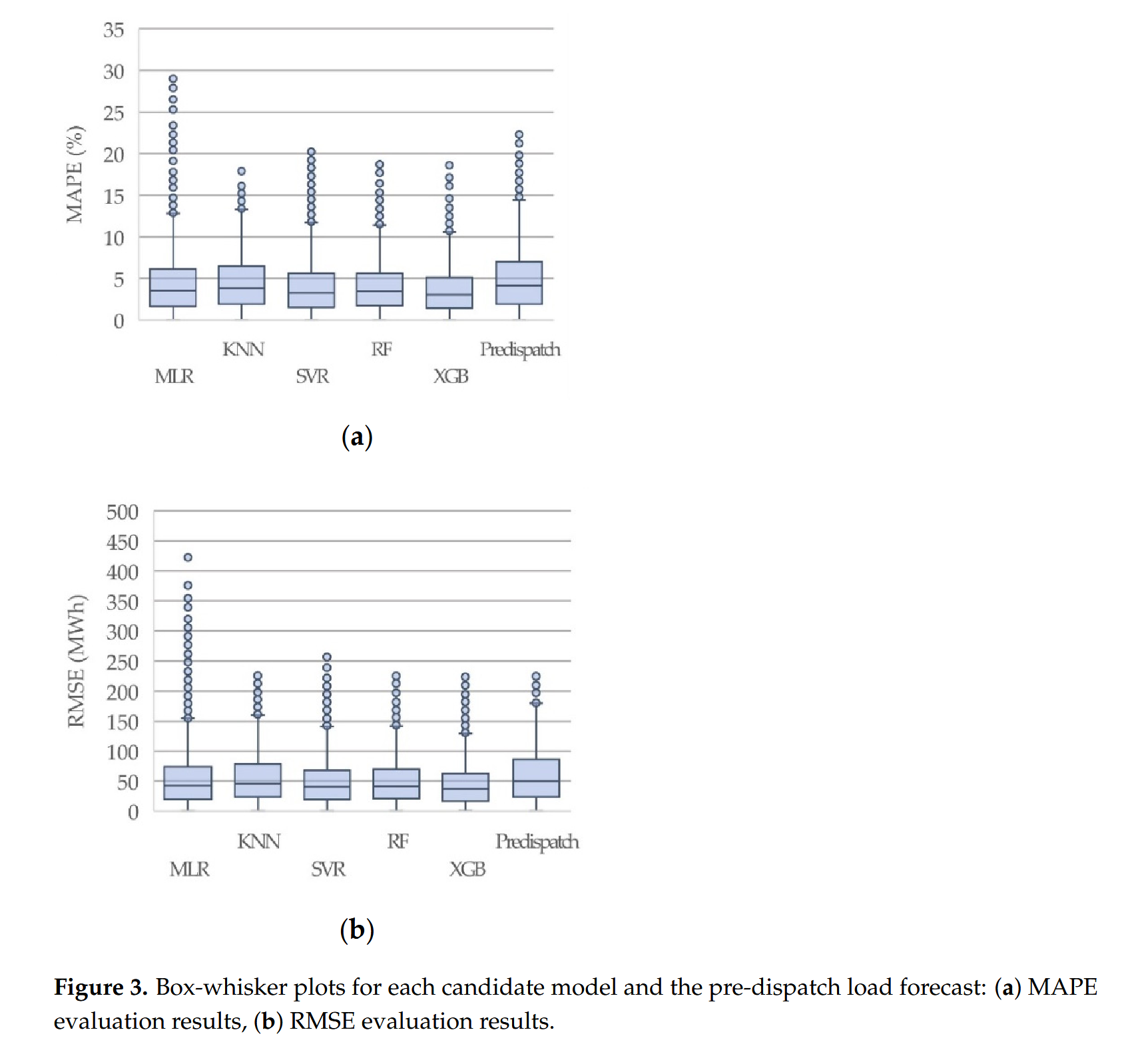
# Nhận xét và đánh giá

## 1. Kết quả đạt được

Sau quá trình xử lý và chuẩn hóa dữ liệu, nhóm tác giả đã tiến hành xây dựng các mô hình: **XGBoost (XGB)**, **Random Forest (RF)**, **Support Vector Regression (SVR)** và **Multiple Linear Regression (MLR)**. Để đánh giá hiệu suất của từng mô hình, tác giả sử dụng hai chỉ số phổ biến là:

* **MAPE (Mean Absolute Percentage Error)** – chỉ số sai số phần trăm tuyệt đối trung bình.
* **RMSE (Root Mean Squared Error)** – căn bậc hai của sai số trung bình bình phương.

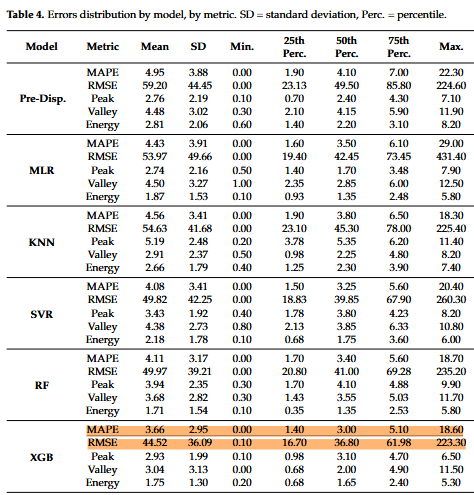
Kết quả đánh giá được trình bày trực quan bằng biểu đồ box-whisker, thể hiện phân bố sai số dự báo của từng mô hình:



*Biểu đồ box-wisker chỉ số MAPE và RMSE*

Các biểu đồ cho thấy mô hình **XGBoost** có sai số thấp và ổn định hơn so với các mô hình còn lại. Cụ thể, bảng tổng hợp sau (Bảng 4 từ bài báo) cho thấy các chỉ số cụ thể:

👉 **Bảng 4**: Sai số MAPE và RMSE của các mô hình



Từ kết quả này, có thể thấy **XGBoost là mô hình có hiệu quả dự báo cao nhất**, cả về độ chính xác (MAPE) và sai số tuyệt đối (RMSE), trong khi các mô hình khác như MLR hay SVR cho kết quả kém hơn đáng kể. Mô hình XGB cũng cho phân phối sai số hẹp và ít giá trị ngoại lai hơn, thể hiện độ ổn định trong khả năng dự báo.

### ****Hạn chế của công trình****

Mặc dù bài báo đạt được nhiều kết quả tích cực trong việc so sánh và lựa chọn mô hình dự báo tối ưu, vẫn tồn tại một số hạn chế cần lưu ý:

**Chỉ sử dụng dữ liệu từ một vùng địa lý** và trong một khoảng thời gian cụ thể (2015–2020), điều này có thể hạn chế khả năng tổng quát của mô hình nếu áp dụng ở khu vực hoặc thời điểm khác.

**Chưa đánh giá tác động của biến động bất thường** (ví dụ: thiên tai, sự kiện đột xuất) đến độ chính xác của mô hình. Các mô hình học máy có thể gặp khó khăn khi gặp dữ liệu ngoài phạm vi huấn luyện.

**Không xem xét yếu tố vận hành hệ thống điện**, như giới hạn công suất, lịch bảo trì,… điều này làm cho kết quả chỉ phản ánh tốt xu hướng phụ tải mà chưa tích hợp yếu tố vận hành thực tế.

Cuối cùng, **mặc dù mô hình XGB cho kết quả tốt nhất**, nhưng bài báo **chưa phân tích rõ ràng về độ phức tạp tính toán, thời gian huấn luyện**, hay khả năng triển khai thực tế của các mô hình này trong môi trường sản xuất.

# Tài liệu kham khảo

[[1](https://www.kaggle.com/datasets/saurabhshahane/electricity-load-forecasting)] Shahane, S. (n.d.). “Electricity Load Forcasting [Data set]. Kaggle. <https://www.kaggle.com/datasets/saurabhshahane/electricity-load-forecasting>

[[2](https://www.mdpi.com/1996-1073/15/10/3566)] Aisyah, S., & Simaremare, A. (2022). Exploratory weather data analysis for electricity load forecasting using SVM and GRNN, case study in Bali, Indonesia. \*Energies\*, 15(20), 7648. <https://www.mdpi.com/1996-1073/15/10/3566>

[[3](https://orca.cardiff.ac.uk/id/eprint/103793/1/Manuscript.pdf)] Kuster, C., Rezgui, Y., & Mourshed, M. (2022). Electrical load forecasting models: A critical systematic review. \*Sustainable Cities and Society\*, 85, 104047. <https://orca.cardiff.ac.uk/id/eprint/103793/1/Manuscript.pdf>

[[4](https://ieeexplore.ieee.org/document/9604919/metrics#metrics)] IEEE Transactions on Power Systems. (2021). Anomaly detection in power consumption. \*IEEE Transactions on Power Systems\*, 36(4), 3456-3465. <https://ieeexplore.ieee.org/document/9604919/metrics#metrics>

[[5](https://www.researchgate.net/publication/348690914_Short-Term_Electricity_Load_Forecasting_with_Machine_Learning)] IEEE Transactions on Power System. (2022). Short-term load forecasting using LSTM. \**IEEE Transactions on Power Systems*\*, 35(3), 2345-1354. <https://www.researchgate.net/publication/348690914_Short-Term_Electricity_Load_Forecasting_with_Machine_Learning>

[[6](https://www.researchgate.net/publication/348690914_Short-Term_Electricity_Load_Forecasting_with_Machine_Learning)] <https://www.researchgate.net/publication/348690914_Short-Term_Electricity_Load_Forecasting_with_Machine_Learning>