

DỰ BÁO NĂNG LƯỢNG GIÓ HƯỚNG TỚI PHÁT TRIỂN BỀN VỮNG

WIND ENERGY FORECAST TOWARDS SUSTAINABLE DEVELOPMENT

Nguyễn Đức Huy¹, Vũ Xuân Cẩm Tú^{2*}, Phạm Thanh Sơn³

¹FPT Software, Quy Nhơn, Việt Nam

²Viện Nghiên cứu và Đào tạo Việt Anh - Đại học Đà Nẵng, Việt Nam

³FPT Greenwich, Đà Nẵng, Việt Nam

*Tác giả liên hệ / Corresponding author: tu.vu@vnuk.udn.vn

(Nhận bài / Received: 26/9/2024; Sửa bài / Revised: 22/10/2024; Chấp nhận đăng / Accepted: 12/11/2024)

Tóm tắt - Dự báo công suất điện gió đóng vai trò then chốt trong việc quản lý nguồn năng lượng tái tạo và góp phần thực hiện các mục tiêu bền vững. Tính ngẫu nhiên, không liên tục và khó dự đoán của tốc độ gió tạo ra những thách thức đáng kể trong việc dự đoán chính xác, ảnh hưởng trực tiếp đến các quyết định đầu tư vào năng lượng tái tạo. Nghiên cứu này so sánh và đánh giá các thuật toán dự báo năng lượng gió, bao gồm các phương pháp thống kê, học máy, học sâu và học tập kết hợp, với mục tiêu phát triển các mô hình có độ chính xác cao trong dự báo công suất điện gió. Nghiên cứu sẽ làm rõ một số mô hình mang lại kết quả dự báo chính xác hơn, giúp tối ưu hóa việc tích hợp điện gió vào lưới điện. Kết quả này không chỉ hỗ trợ cho việc lập kế hoạch và quản lý hiệu quả hệ thống điện, mà còn đóng góp quan trọng vào lĩnh vực tài chính bền vững bằng cách tăng cường khả năng dự báo nguồn năng lượng tái tạo.

Từ khóa - Dự báo năng lượng gió; mô hình học tập kết hợp; phát triển bền vững; năng lượng tái tạo

Abstract - Wind power forecasting plays a crucial role in managing renewable energy sources and contributes to achieving sustainable financial goals. The randomness, discontinuity, and unpredictability of wind speed create significant challenges in accurate forecasting, directly impacting investment decisions in renewable energy. This study compares and evaluates wind power forecasting algorithms, including traditional statistical methods, Machine Learning, Deep Learning, and Ensemble Learning, to develop highly accurate models for wind power forecasting. By analyzing the accuracy of each method, the study will clarify why certain models yield more accurate forecasts, helping to optimize the integration of wind power into the electrical grid. These results not only support effective planning and management of power systems but also make important contributions to sustainable development by enhancing the forecasting capability of renewable energy supply.

Key words - Wind forecasting; ensemble learning; sustainable development; renewables

1. Đặt vấn đề

Biến đổi khí hậu trong những năm qua đã tạo ra những tác động ngày càng nghiêm trọng, đặc biệt qua các hiện tượng thời tiết cực đoan và tình trạng nước biển dâng, ảnh hưởng trực tiếp đến an ninh lương thực, an ninh nguồn nước, và sự phát triển bền vững [1]. Hội nghị COP 26 nhấn mạnh rằng thế giới cần tiếp tục duy trì mục tiêu giữ mức tăng nhiệt độ trung bình toàn cầu dưới 1,5 độ C so với thời kỳ tiền công nghiệp [2]. Năng lượng tái tạo, đặc biệt là năng lượng gió, đóng vai trò then chốt trong việc giảm phụ thuộc vào nhiên liệu hóa thạch và giảm phát thải khí nhà kính, từ đó góp phần vào phát triển kinh tế bền vững và bảo vệ môi trường.

Trong những năm gần đây, năng lượng gió đã trở thành một trong những ngành năng lượng phát triển nhanh nhất thế giới do sở hữu nhiều ưu điểm, chẳng hạn như khả năng tái tạo, tiết kiệm chi phí, yêu cầu bảo trì thấp và chiếm tương đối ít không gian [3]. Tuy nhiên, một trong những thách thức lớn nhất của năng lượng gió là dự báo chính xác sản lượng, vốn phụ thuộc vào điều kiện khí hậu và các yếu tố theo mùa. Sự bất ổn này làm giảm độ chính xác và độ tin cậy của các phương pháp dự báo, ảnh hưởng trực tiếp đến việc hoạch định và đầu tư vào năng lượng tái tạo.

Việc nghiên cứu và phát triển các phương pháp dự báo năng lượng gió có độ chính xác và tin cậy cao không chỉ góp phần tối ưu hóa tích hợp năng lượng gió vào lưới điện mà còn là yếu tố quan trọng trong việc giảm rủi ro tài chính và môi trường. Đặc biệt trong bối cảnh các tiêu chuẩn phát triển bền vững đang trở thành thước đo quan trọng đối với các nhà đầu tư, việc cải thiện khả năng dự báo sẽ hỗ trợ các doanh nghiệp tuân thủ yêu cầu về môi trường, nâng cao hiệu quả tài chính và giảm thiểu rủi ro phát sinh từ biến động khí hậu [4].

2. Nội dung nghiên cứu

2.1. Các mô hình dự báo

Trong các nghiên cứu dự báo năng lượng tái tạo nói chung và năng lượng gió nói riêng, các nhà nghiên cứu thường áp dụng ba mô hình chính: mô hình vật lý, mô hình thống kê và mô hình học máy [5].

2.1.1. Mô hình vật lý

Mô hình vật lý ước lượng tốc độ gió hoặc sản lượng điện trong tương lai dựa trên dữ liệu khí tượng và các yếu tố vật lý. Những dự đoán này được rút ra từ việc giải các phương trình toán học liên quan đến động lực học nhiệt và thủy động lực học [5]. Các phương pháp này thường bao gồm việc phân tích dữ liệu lịch sử từ các trạm quan trắc khí

¹ FPT Software, Quy Nhơn, Vietnam (Nguyen Duc Huy)

² The University of Danang - VNUK Institute for Research and Executive Education, Vietnam (Vu Xuan Cam Tu)

³ FPT Greenwich Centre, FPT University, Danang, Vietnam (Pham Thanh Son)

tượng và sử dụng mô hình toán học để mô phỏng điều kiện gió trong tương lai. Tuy nhiên, mô hình vật lý gặp một số hạn chế trong việc nắm bắt các mối quan hệ phức tạp giữa các biến đầu vào và sản lượng điện, đặc biệt do tính không ổn định của nguồn gió.

2.1.2. Mô hình thống kê

Mô hình thống kê hồi quy được sử dụng để tìm ra mối quan hệ giữa các yếu tố đầu vào (như dữ liệu lịch sử về thời tiết) và kết quả đầu ra (dự báo sản lượng điện gió trong tương lai). Các mô hình này thường được áp dụng cho dự báo ngắn hạn, bao gồm phân tích chuỗi thời gian và bộ lọc Kalman. Phân tích chuỗi thời gian giúp phân tích và dự đoán các biến đổi theo thời gian, và nghiên cứu của Mahata [6] về dữ liệu sản xuất điện và thời tiết dựa trên báo cáo từ Cơ quan Điện lực Trung ương (CEA), Ấn Độ đã chỉ ra rằng mô hình ARIMA (Autoregressive Integrated Moving Average - Tự hồi quy tích hợp trung bình trượt) hiệu quả hơn các mô hình tiên tiến khác trong việc dự báo dữ liệu phi tuyến tính.

Bộ lọc Kalman, một kỹ thuật dùng để ước lượng trạng thái của hệ thống động theo thời gian, đặc biệt hữu ích khi dữ liệu bị nhiễu, giúp tăng cường độ tin cậy của các dự báo.

2.1.3. Mô hình máy học

Trong những năm gần đây, các kỹ thuật dự báo dựa trên máy học đã đạt được những bước tiến lớn nhờ khả năng tự động học từ các mẫu dữ liệu phức tạp [7]. Những phương pháp này ngày càng được coi là công cụ tin cậy và chính xác trong việc dự đoán sản lượng năng lượng gió [8]. Một ví dụ điển hình là nghiên cứu của Pathak [9], tiến hành phân tích toàn diện nhiều mô hình máy học như K-Nearest Neighbor, Random Forest, các biến thể của Gradient Boosting Machines, và Extreme Gradient Boosting Machine (XGBoost). Sử dụng dữ liệu từ cuộc thi dự báo năng lượng toàn cầu GEFCom2014, nghiên cứu này chỉ ra rằng mô hình hồi quy XGBoost đạt hiệu suất tốt nhất trong số các mô hình được thử nghiệm. Đồng thời, nghiên cứu cũng nhấn mạnh tầm quan trọng của việc tái sản xuất kết quả nghiên cứu thông qua việc cung cấp các siêu tham số tối ưu bằng cách điều chỉnh và tối ưu hóa mô hình một cách kỹ lưỡng.

Ngoài ra, các mô hình học sâu (deep learning) cũng đã được ứng dụng rộng rãi trong dự báo năng lượng gió. Đặc biệt, Taoussi [10] đã nghiên cứu mô hình dự báo năng lượng gió tại Tiểu bang El-Oued, Algeria nhằm tối ưu hóa việc sử dụng điện gió ngoài lưới điện và giảm thiểu rủi ro xói mòn do gió. Nghiên cứu này tập trung vào dự báo tốc độ gió ngắn hạn bằng cách kết hợp các phương pháp thống kê như ARIMA theo mùa và các kỹ thuật học sâu như LSTM (Long Short-term Memory - Bộ nhớ dài-ngắn hạn). Kết quả cho thấy mô hình LSTM đạt hiệu suất vượt trội hơn trên tập dữ liệu đã sử dụng, chứng minh tiềm năng lớn của học sâu trong lĩnh vực này.

2.1.4. Mô hình học tập kết hợp

Các mô hình dựa trên học tập kết hợp là sự tích hợp của nhiều mô hình dự đoán khác nhau, thường kết hợp với các công nghệ hiện có để đạt hiệu suất cao hơn so với các mô hình truyền thống. Cụ thể, Lang [11] đã đề xuất phương pháp dự báo năng lượng gió ngắn hạn bằng cách

kết hợp mô hình MGM (Minimal Gated Memory) và chiến lược điều chỉnh độ rộng khoảng cách cải tiến (Improved Interval Width Adaptive Adjustment Strategy) để nâng cao độ chính xác trong dự báo ngắn hạn về sản lượng điện gió.

Nhìn chung, các phương pháp dự báo năng lượng tái tạo, từ mô hình vật lý đến học máy và học sâu, đều có vai trò quan trọng trong việc hỗ trợ các quyết định tài chính bền vững. Bằng cách cải thiện độ chính xác trong dự báo, các nhà đầu tư và doanh nghiệp có thể giảm thiểu rủi ro tài chính liên quan đến sự biến động của năng lượng tái tạo, đồng thời đảm bảo rằng các hoạt động đầu tư góp phần phát triển kinh tế bền vững.

2.2. Các phương pháp dự báo

Dữ liệu năng lượng gió có tính phức tạp và biến động cao. Do đó, chúng tôi sử dụng các phương pháp khác nhau để khai thác các đặc điểm này:

(i) **Phương pháp tuyến tính (Hồi quy tuyến tính):** Giúp nhận biết xu hướng tuyến tính trong dữ liệu.

(ii) **Phương pháp phi tuyến (K-nearest neighbors, Cây quyết định, Adaboost, Mạng nơ-ron):** Các phương pháp này xử lý tốt các mối quan hệ phi tuyến tính và tương tác phức tạp trong dữ liệu, phù hợp với bản chất phi tuyến của năng lượng gió.

(iii) **Phương pháp học tập kết hợp (ensemble learning)** để kết hợp nhiều mô hình dự báo khác nhau và xây dựng một hệ thống trọng số nhằm tạo ra dự đoán chính xác nhất.

Việc lựa chọn các phương pháp dự báo khác nhau giúp chúng tôi đạt được mục tiêu tăng tính linh hoạt và chính xác vì mỗi phương pháp đều khai thác những khía cạnh khác nhau của dữ liệu, giúp giảm sai số và tối ưu hóa dự báo. Bên cạnh đó, chúng tôi có thể so sánh và lựa chọn mô hình phù hợp nhất cho bài toán dự báo cụ thể. Trong phần này, chúng tôi sẽ trình bày tóm tắt về từng thuật toán dự báo được sử dụng, cùng với các tiêu chí đánh giá hiệu suất của từng thuật toán và quy trình thử nghiệm.

2.2.1. Hồi quy tuyến tính

Hồi quy tuyến tính là một trong những mô hình đơn giản nhất, trong đó biến phụ thuộc được mô hình hóa như một tổ hợp tuyến tính của các biến dự báo. Các hệ số của mô hình được ước tính bằng cách tối thiểu hóa tổng bình phương phần dư, thông qua phương pháp bình phương tối thiểu.

2.2.2. Thuật toán K láng giềng gần nhất

Thuật toán K láng giềng gần nhất (K-nearest neighbors) hoạt động bằng cách xác định k điểm dữ liệu gần nhất với điểm cần dự đoán trong không gian dữ liệu. Dựa trên các giá trị của những điểm lân cận này, thuật toán tính toán giá trị dự đoán cho điểm cần dự báo bằng cách lấy trung bình hoặc kết hợp các giá trị của các điểm lân cận. Thuật toán này không xây dựng mô hình dựa trên hình học cụ thể, mà dựa vào sự tương đồng trực tiếp giữa các điểm dữ liệu [12].

2.2.3. Thuật toán cây quyết định (Decision tree)

Cây quyết định là một mô hình dự đoán phổ biến, trong đó mỗi nút của cây biểu thị một điều kiện dựa trên giá trị

của các biến đầu vào. Dữ liệu được phân chia thành các nhánh theo các điều kiện này, dẫn đến các lá chứa giá trị dự đoán cuối cùng. Mô hình này đặc biệt hữu ích trong việc mô hình hóa các mối quan hệ phi tuyến giữa các biến đầu vào và đầu ra [13].

2.2.4. Thuật toán tăng cường thích ứng (AdaBoost)

AdaBoost là một thuật toán dựa trên phương pháp Boosting nhằm cải thiện hiệu suất của mô hình dự báo [14]. Bằng cách kết hợp nhiều mô hình yếu, thường là các cây quyết định nhỏ, AdaBoost tạo ra một mô hình mạnh mẽ hơn, giúp nâng cao độ chính xác tổng thể của dự báo.

2.2.5. Mạng nơ-ron (Artificial Neural Network)

Mạng nơ-ron nhân tạo (ANN) hoạt động bằng cách kết hợp các biến dự báo (features) thành các tổ hợp tuyến tính, sau đó áp dụng một hàm phi tuyến để mô hình hóa biến phụ thuộc [15]. Trong nghiên cứu này, chúng tôi sử dụng mạng nơ-ron truyền thẳng đa lớp (Multilayer Perceptron), một dạng phổ biến của mạng nơ-ron trong các bài toán dự báo.

2.2.6. Học tập siêu kết hợp (Super learners)

Thuật toán siêu học (Super Learners) sử dụng phương pháp học tập kết hợp (ensemble learning) để kết hợp nhiều mô hình dự báo khác nhau và xây dựng một hệ thống trọng số nhằm tạo ra dự đoán chính xác nhất. Mục tiêu của Super Learners là cung cấp các dự đoán tốt bằng hoặc tốt hơn bất kỳ mô hình đơn lẻ nào [16]. Thuật toán này là một ứng dụng của kỹ thuật tổng quát hóa xếp chồng (stacking), hay còn gọi là pha trộn (blending), kết hợp với phương pháp kiểm định chéo k lần. Trong quy trình này, tất cả các mô hình đều sử dụng cùng một tập chia k lần của dữ liệu, và một siêu mô hình (meta-model) được huấn luyện dựa trên các dự đoán ngoài mẫu của các mô hình cơ bản. Quy trình của thuật toán siêu học có thể được tóm tắt như sau:

(i) Triển khai kiểm định chéo k-fold trên tập dữ liệu huấn luyện.

(ii) Chọn m mô hình cơ bản (base-models).

(iii) Đối với mỗi mô hình cơ bản: a. Đánh giá mô hình bằng kiểm định chéo k lần. b. Lưu trữ tất cả các dự đoán. c. Huấn luyện mô hình trên toàn bộ tập dữ liệu huấn luyện và lưu trữ.

(iv) Huấn luyện một siêu mô hình (meta-model) trên các dự đoán ngoài mẫu.

(v) Đánh giá hiệu suất của siêu mô hình meta-model.

2.3. Dữ liệu nghiên cứu

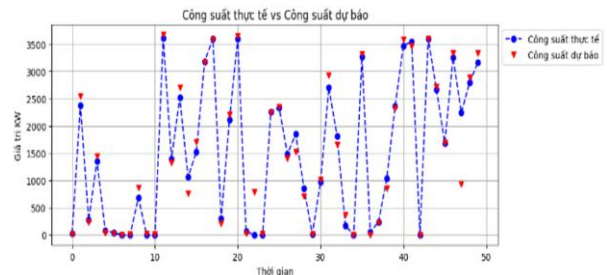
Nghiên cứu này sử dụng dữ liệu từ các tuabin gió ở Thổ Nhĩ Kỳ [17]. Bộ dữ liệu được lấy từ Kaggle, bao gồm các tham số liên quan đến gió và sản xuất điện, chẳng hạn như tốc độ gió (m/s), hướng gió tại độ cao trục tuabin ($^{\circ}$), công suất lý thuyết do nhà sản xuất tuabin chỉ định (kilowatt-giờ, kWh), và sản lượng điện thực tế theo thời gian (kilowatt, kW). Tổng cộng có 50.530 điểm dữ liệu được thu thập từ ngày 1/1 đến 31/12/2018.

Việc tiêu thụ và sản xuất năng lượng thường phụ thuộc đến các yếu tố thời gian, đặc biệt là theo quy mô hàng ngày và hàng tháng. Bằng cách trích xuất các đặc trưng có ý

nghĩa từ dữ liệu ngày/giờ, chúng tôi có thể giúp mô hình máy học khai thác tốt hơn những yếu tố này. Ví dụ, 13 giờ ngày 21 tháng 2 thì có tạo thành 3 đặc trưng lần lượt như giờ (13) ngày (21) và tháng (2).

Để làm sạch dữ liệu, chúng tôi áp dụng phương pháp khoảng tứ phân vị (Interquartile Range, IQR), một kỹ thuật thống kê phổ biến dùng để phát hiện và loại bỏ các giá trị ngoại lai. Cụ thể, các giá trị ngoại lai là các giá trị quan sát nằm dưới ngưỡng ($Q1 - 1.5 \text{ IQR}$) hoặc trên ngưỡng ($Q3 + 1.5 \text{ IQR}$). Trong đó, $Q1$ và $Q3$ lần lượt là tứ phân vị thứ nhất và thứ ba, và $\text{IQR} = Q3 - Q1$. Phương pháp này được áp dụng cho tất cả các biến sử dụng.

Sau khi dữ liệu được làm sạch, chúng tôi phân chia bộ dữ liệu thành hai tập con: tập huấn luyện (80%) và tập kiểm tra (20%) theo phương pháp ngẫu nhiên (random split). Chia ngẫu nhiên giúp đảm bảo rằng cả hai tập huấn luyện và tập kiểm tra đều có tính đại diện cao cho toàn bộ dữ liệu. Điều này giúp mô hình học được các đặc điểm chung mà không bị thiên lệch bởi các yếu tố thời gian cụ thể. Đây là cách phân chia tiêu chuẩn thường được sử dụng trong các nghiên cứu học thuật để đánh giá hiệu suất của các thuật toán.



Hình 1. Công suất thực tế và dự báo

Trong nghiên cứu này, các mô hình đều sử dụng các biến đầu vào chính bao gồm: hướng gió tại độ cao trục tuabin ($^{\circ}$), công suất lý thuyết, tốc độ gió, ngày, tháng, giờ. Đầu ra của tất cả các mô hình là công suất gió dự báo. Khi sử dụng đầy đủ các biến như hướng gió, tốc độ gió, và thời gian, mô hình có khả năng nắm bắt được mối quan hệ phức tạp giữa các yếu tố này, từ đó đưa ra dự báo chính xác hơn.

3. Kết quả và đánh giá

Các mô hình dự đoán trong nghiên cứu này được sử dụng dựa trên thư viện scikit-learn, một thư viện học máy dành cho Python. Các thí nghiệm mô phỏng đều được chạy trên Kaggle Notebook.

3.1. Kết quả

Kết quả của mô hình dự đoán sử dụng thuật toán học tập siêu kết hợp (Super Learner) được thể hiện trong Hình 1. Qua hình này, có thể thấy vẫn tồn tại sự chênh lệch giữa kết quả thực tế và dự đoán.

Để đánh giá chính xác hiệu suất của các thuật toán, các thước đo sau đã được sử dụng:

(i) RMSE (Sai số trung bình căn bậc hai): Đo lường mức sai số trung bình giữa giá trị dự đoán và giá trị thực tế. RMSE càng thấp, hiệu suất dự đoán càng tốt.

(ii) MAE (Sai số tuyệt đối trung bình): Trung bình của

các sai số tuyệt đối giữa dự đoán và thực tế, cung cấp một thước đo trực tiếp về độ chính xác.

(iii) MAPE (Sai số phần trăm tuyệt đối trung bình): Thể hiện sai số phần trăm trung bình giữa dự đoán và thực tế, giúp đánh giá tương quan tương đối của các sai số.

Kết quả so sánh hiệu suất giữa các thuật toán được tóm tắt trong Bảng 1. Từ đó có thể thấy rằng Super Learner có tỷ lệ sai số tương đối thấp hơn so với các phương pháp khác. Super Learner kết hợp nhiều mô hình "yếu" để tạo thành một mô hình "mạnh", mang lại độ chính xác cao hơn, cụ thể điểm mạnh của mô hình Super Learner được thể hiện ở các điểm sau:

(i) Tính linh hoạt cao: Super Learner không giới hạn ở một mô hình cụ thể mà có thể kết hợp nhiều mô hình với các thuật toán và tham số khác nhau. Điều này tạo ra một mô hình mạnh mẽ và linh hoạt hơn so với việc chỉ sử dụng một mô hình đơn lẻ.

(ii) Giảm phương sai và thiên lệch: Mỗi mô hình đơn lẻ có thể có xu hướng thiên lệch hoặc phương sai lớn, tức là có thể dự báo sai trong những trường hợp nhất định. Bằng cách kết hợp nhiều mô hình, Super Learner có thể giảm thiểu phương sai của một mô hình "yếu" và tận dụng các mô hình khác để bù đắp. Nhờ vậy, sai số tổng thể được giảm xuống so với các mô hình đơn lẻ.

(iii) Khả năng khái quát hóa tốt hơn: Việc kết hợp các mô hình với nhiều quan điểm dự báo khác nhau cho phép Super Learner có khả năng khái quát hóa tốt hơn trên các tập dữ liệu chưa thấy trước, từ đó giúp cải thiện độ chính xác dự báo.

Đây là ưu điểm nổi bật của phương pháp, đặc biệt hữu ích trong các ứng dụng thực tế, nơi mà độ chính xác dự đoán đóng vai trò quan trọng. Tuy nhiên, cần cân nhắc thêm các yếu tố như độ phức tạp của mô hình và khả năng tổng quát hóa khi áp dụng trên các tập dữ liệu khác.

Bảng 1. Kết quả so sánh các thuật toán

Thuật toán	RMSE	MAPE	MAE
Linear Regression	394,928	4,354	165,137
Decision Tree	279,160	0,226	102,635
KNeighbors	336,720	0,636	132,128
AdaBoost	453,036	2,537	307,032
MLP	344,910	1,041	138,592
Super Learner	249,193	0,215	102,269

3.2. Đánh giá

Trong nghiên cứu này, chúng tôi nhận thấy rằng thuật toán học tập kết hợp mang lại độ chính xác vượt trội so với các phương pháp dự báo riêng lẻ. Mỗi mô hình dự báo được sử dụng trong nghiên cứu đều có những điểm mạnh và hạn chế riêng. Mô hình tuyến tính có hạn chế là bị giới hạn trong khả năng mô hình hóa các mối quan hệ phi tuyến giữa các yếu tố, có thể dẫn đến thiếu sót khi dự báo dữ liệu năng lượng gió phức tạp. Đối với các mô hình phi tuyến (Decision Tree, Adaboost, Mạng nơ-ron) có khả năng nắm bắt các tương tác phi tuyến, nhưng có thể dễ bị overfitting, đặc biệt khi dữ liệu có độ nhiễu cao hoặc thiếu tính ổn định. Đối với Super Learner, dù có thể kết hợp nhiều mô hình để cải thiện kết quả dự báo, nhưng quá

trình lựa chọn mô hình con và trọng số kết hợp có thể yêu cầu nhiều tính toán phức tạp.

Việc lựa chọn siêu tham số cho các mô hình học máy, đặc biệt là các mô hình phi tuyến như Random Forest, Gradient Boosting, và Mạng nơ-ron, là một thách thức lớn. Quy trình này phụ thuộc nhiều vào thử nghiệm và điều chỉnh, có thể dẫn đến sự thiếu ổn định trong kết quả nếu không được tối ưu hóa đúng cách. Chúng tôi đã sử dụng một số phương pháp tối ưu hóa siêu tham số (như Grid Search), nhưng có thể cần phát triển thêm những phương pháp tự động hóa, như Bayesian Optimization, để cải thiện hiệu quả.

Kết quả dự báo khá khả quan, tuy nhiên vẫn cần thử nghiệm thêm với các tập dữ liệu năng lượng gió khác nhau để xác minh tính ổn định của mô hình. Nghiên cứu đã phân tích một số yếu tố ảnh hưởng đến độ chính xác của các phương pháp dự đoán năng lượng gió, nhưng cần có một nghiên cứu tổng quát hơn để làm sáng tỏ lý do tại sao một số phương pháp lại hoạt động tốt hơn so với các phương pháp khác.

4. Kết luận

Trong bài báo này, chúng tôi đã tập trung vào các phương pháp dự đoán năng lượng gió bằng cách sử dụng các phương pháp dự báo khác nhau như thống kê truyền thống, học máy, học tập kết hợp. Chúng tôi hy vọng rằng kết quả của nghiên cứu sẽ đóng góp vào việc phát triển các mô hình dự đoán năng lượng gió đáng tin cậy và hiệu quả hơn trong tương lai. Các phương pháp dự báo không chỉ có giá trị trong lĩnh vực năng lượng mà còn đóng góp vào các mục tiêu phát triển bền vững toàn cầu.

Việc cải thiện độ chính xác không chỉ mang lại lợi ích kỹ thuật cho hệ thống điện mà còn tạo điều kiện thuận lợi cho các nhà đầu tư giảm thiểu rủi ro tài chính và môi trường. Các mô hình dự báo này đóng vai trò quan trọng trong việc thúc đẩy tài chính bền vững và phát triển năng lượng tái tạo, giúp doanh nghiệp và nhà đầu tư đạt được mục tiêu phát triển bền vững.

TÀI LIỆU THAM KHẢO

- [1] Meetings Coverage, "Devastating Impacts of Climate Change Threatening Farm Outputs, Increasing Global Hunger, Delegates Say as Second Committee Takes up Agriculture, Food Security", *United Nations*, 2018. [Online] Available: <https://press.un.org/en/2018/gaef3499.doc.htm> [Access September 01, 2024].
- [2] Conference Report, "COP26: Together for our planet", *United Nations*, 2023. [Online] Available: <https://www.un.org/en/climatechange/cop26> [Access September 01, 2024].
- [3] P. Sadorsky, "Wind energy for sustainable development: Driving factors and future outlook", *Journal of Cleaner Production*, vol. 289, no. 125779, 2021.
- [4] G. Friede, T. Busch, and A. Bassen, "The Role of ESG in Corporate Financial Performance: A Review of the Literature", *Journal of Sustainable Finance & Investment*, vol. 5, no. 4, pp. 210-233, 2015.
- [5] M. Lei, C. Shiyang, J. Fang, and Z. Hongling, "A review on the forecasting of wind speed and generated power", *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 915-920, 2009.
- [6] S. Mahata and R. Pathak, "Comparative study of time-series forecasting models for wind power generation in Gujarat, India", *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 8, no. 100511, 2024.

- [7] T. Alazemi, M. Darwish, and M. Radi, "Renewable energy sources integration via machine learning modelling: A systematic literature review", *Heliyon*, vol. 10, no. 4, 2024.
- [8] D. Bin Abu Sofian, H. R. Lim, H. S. M. Heli, Z. Ma, K. W. Chew, and P. L. Show, "Machine learning and the renewable energy revolution: Exploring solar and wind energy solutions for a sustainable future including innovations in energy storage", *Sustainable Development*, vol. 32, no. 4, pp. 3953-3978, 2024.
- [9] R. Pathak, A. K. Sharma, and A. Singh, "Comparative assessment of regression techniques for wind power forecasting", *IETE Journal of Research*, vol. 69, issue. 3, pp. 1393-1402, 2023.
- [10] Taoussi, M. Bouhorma, and A. Larhrib, "Seasonal ARIMA and LSTM Models for Wind Speed Prediction at El-Oued Region, Algeria", in *2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, El Oued, Algeria: IEEE, 2024, pp. 1-8.
- [11] J. A. Lang, "A novel two-stage interval prediction method based on minimal gated memory network for clustered wind power forecasting", *Wind Energy*, vol. 24, pp. 450-464, 2021.
- [12] E. Fix and J. L. Hodges, "Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties", USAF School of Aviation Medicine, University of Iowa, Randolph Field, Texas, Project No. 21-49-004, Report No. 4, 1951.
- [13] L. Breiman, *Classification and Regression Trees*, 1st edition. New York: Chapman and Hall/CRC, 2017.
- [14] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.
- [15] X. Glorot, A. Bordes, and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy: 2010, pp. 249-256.
- [16] E. C. Polley, M. J. van der Laan, and A. E. Hubbard, "Super Learner in Prediction", Division of Biostatistics, University of California, Berkeley, Working Paper 266, May 2010.
- [17] Erisen, "Wind turbine SCADA dataset", *Kaggle*, 2018. [Online]. Available: <https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset> [Accessed September 01, 2024]