Nghiên cứu phát hiện sự cố trên hệ thống tuabin gió dựa trên học máy

Nguyễn Quốc Minh[1*], Nguyễn Chất Phát[2], Nguyễn Trọng Khiêm[1],
Trần Văn Đại[1], Nguyễn Xuân Tùng[1]

[1]Trường Điện - Điện tử, Đại học Bách khoa Hà Nội, Số 1 Đại Cồ Việt, Hai Bà Trưng, Hà Nội, Việt Nam;
[2]Trung tâm Điều độ Hệ thống điện Quốc gia, Số 11 Cửa Bắc, Ba Đình, Hà Nội, Việt Nam.
[*]Email: minh.nguyenquoc@hust.edu.vn
Nhận bài: 14/11/2023; Hoàn thiện: 05/02/2024; Chấp nhận đăng: 08/4/2024; Xuất bản: 22/04/2024.
DOI: https://doi.org/10.54939/1859-1043.j.mst.94.2024.3-10

TÓM TẮT

Năng lượng tái tạo nói chung và năng lượng gió nói riêng đang ngày càng nhận được nhiều sự quan tâm với mục tiêu giảm phát thải khí nhà kính và sản xuất năng lượng sạch. Trong những năm gần đây, các nhà máy và trang trại gió tăng lên đáng kể, thúc đẩy năng lượng gió trở thành một nguồn năng lượng vô cùng tiềm năng. Tuy nhiên, với tính chất bất định của nguồn năng lượng gió thì việc đảm bảo hệ thống tuabin gió vận hành an toàn, giảm thiểu thời gian ngừng hoạt động do sự cố đóng vai quan trọng để tối ưu hóa chi phí sản xuất và nâng cao độ tin cậy của hệ thống điện. Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng các mô hình học máy để phát hiện sự cố xảy ra trên hệ thống tuabin gió. Các thông số vận hành đo được từ SCADA được sử dụng làm dữ liệu đầu vào cho các mô hình học máy. Kết quả cho thấy, các mô hình học máy có thể phát hiện sự cố trên hệ thống tuabin gió với độ chính xác lên đến hơn 99% với thời gian huấn luyện mô hình chỉ cỡ vài chục ms.

Từ khoá: Năng lượng tái tạo; Tuabin gió; Sự cố; SCADA; Học máy.

1. MỞ ĐẦU

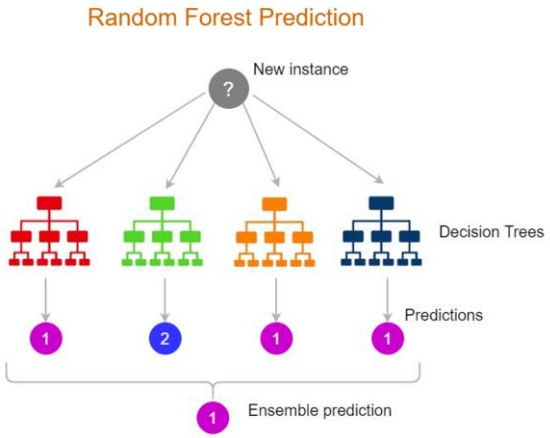
Năng lượng gió ngày càng được các nước quan tâm và phát triển ở trên thế giới nói chung, ở Việt Nam nói riêng. Nhiều chính sách đã được đưa ra để phát triển nguồn năng lượng sạch có tiềm năng lớn này. Theo quy hoạch điện 8 đã được chính phủ phê duyệt, đến năm 2030, công suất điện gió trên bờ đạt 22 GW, công suất điện gió ngoài khơi đạt 6 GW, định hướng đến năm 2050 công suất điện gió đạt khoảng 70 - 91.5 GW. Tại Việt Nam, nhiều dự án điện gió lớn đã được lắp đặt và đưa vào khai thác và sử dụng. Các tuabin gió được lắp đặt ở ngoài trời, ở trên bờ (onshore) hoặc ngoài biển khơi (offshore), những nơi có tốc gió lớn nên phải làm việc trong các điều kiện thời tiết tương đối khắc nghiệt. Việc thường xuyên phải làm việc trong các điều kiện như vậy dẫn đến các hỏng hóc và sự cố trên tuabin gió là điều không tránh khỏi. Bên cạnh đó, do có cấu tạo phức tạp (hình 1) nên sự cố trên tuabin gió tương đối đa dạng, có thể xuất hiện tại nhiều vị trí như sự cố trên rotor và cách quạt, sự cố ở hộp số, sự cố tại máy phát, sự cố tại vị trí ổ đỡ, sự cố nứt gãy trục tuabin, sự cố hỏng học hệ thống phanh, sự cố ở thiết bị cảm biến,… Điều này dẫn tới chi phí bảo dưỡng và vận hành của tuabin gió cũng khá cao, thường chiếm 15-20% chi phí đầu tư khi mới lắp đặt, và có thể tăng lên đến 20-35% trong suốt vòng đời của tuabin [1]. Việc phát triển các loại tuabin mới với chi phí bảo dưỡng và vận hành thấp đang được quan tâm bởi các nhà sản xuất lớn trong thời gian gần đây. Bên cạnh đó, việc chẩn đoán và phát hiện sớm các lỗi có thể xảy ra trong tương lai là biện pháp không tốn kém nhưng đem lại hiệu quả cao về mặt kinh tế. Phương pháp chẩn đoán dựa trên thử nghiệm không phá hủy khi xuất xưởng, và trong quá trình làm việc đã được đề xuất ở [2]. Một số phương pháp chẩn đoán và phát hiện lỗi dựa trên giám sát tình trạng làm việc và hiệu suất làm việc tại các bộ phận của tuabin gió được đề xuất ở [3-5]. Một số nghiên cứu đã sử dụng dữ liệu đo được từ hệ thống SCADA để phân tích nhằm chẩn đoán và cảnh báo sớm các sự cố xảy ra ở tuabin gió [6-8]. Gần đây, với sự phát triển mạnh mẽ của học máy, học sâu và trí tuệ nhân tạo đã mở ra hướng nghiên cứu mới trong lĩnh vực chẩn đoán, nhận dạng sự cố trong hệ thống điện [9, 10]. Ưu điểm của các mô hình này là có thể xử lý lượng dữ liệu lớn, có cấu trúc

phức tạp, và biến thiên liên tục. Trong bài báo này, nhóm nghiên cứu đề xuất sử dụng một số mô hình học máy trong lĩnh vực nhận dạng như Random Forest, Logistic Regression, Adaptive Boosting (AdaBoost), Multi-Layer Perceptron (MLP), Stochastic Gradient Descent (SGD) và Gradient Boost (GDB) để phát hiện lỗi của hệ thống tuabin gió dựa trên dữ liệu đầu vào đo được từ hệ thống SCADA. Dữ liệu được thu thập và ứng dụng trong mô hình bao gồm dữ liệu trạng thái và dữ liệu vận hành của tuabin gió.

Hình 1. *Cấu tạo của tuabin gió.*

2. CÁC MÔ HÌNH HỌC MÁY ỨNG DỤNG TRONG BÀI TOÁN NHẬN DẠNG

2.1. Random Forest



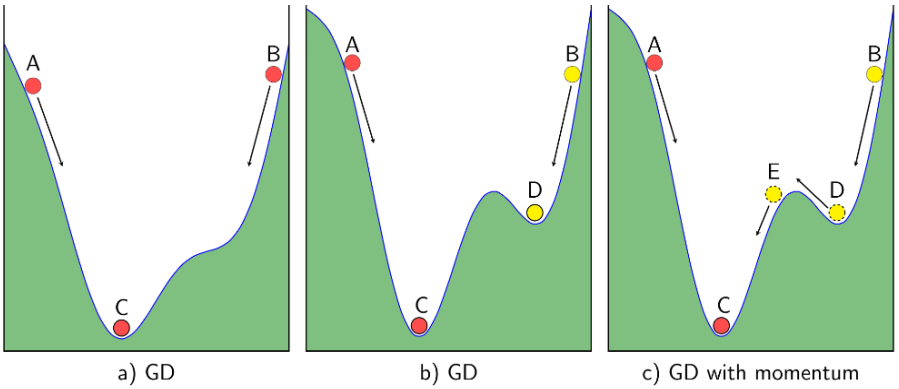
Hình 2. *Mô hình Random Forest.*

Random Forest là một phương pháp thống kê mô hình hóa bằng máy dùng để phục vụ các mục đích phân loại, tính hồi quy và các nhiệm vụ khác bằng cách xây dựng nhiều cây quyết định (hình 2). Một cây quyết định là một cách đơn giản để biểu diễn một giao thức. Nói cách khác, cây quyết định biểu diễn một kế hoạch, trả lời câu hỏi phải làm gì trong một hoàn cảnh nhất định. Mỗi nút của cây sẽ là các thuộc tính, và các nhánh là giá trị lựa chọn của thuộc tính đó. Bằng cách đi theo các giá trị thuộc tính trên cây, cây quyết định sẽ cho ta biết giá trị dự đoán. Nhóm thuật toán cây quyết định có một điểm mạnh đó là có thể sử dụng cho cả bài toán phân loại và hồi quy. Random Forest có khả năng tìm ra thuộc tính nào quan trọng hơn so với những thuộc tính khác. Trên thực tế, nó còn có thể chỉ ra rằng một số thuộc tính là không có tác dụng trong cây quyết định. Để các đặc tính trong mô hình được tương quan thì thuật toán lấy mẫu được áp dụng. Thuật toán này giúp nhận biết những đặc tính mạnh có thể ảnh hưởng đến quả đầu ra và từ đó nó được chọn nhiều hơn

trong mô hình này. Và từ đó Random Forest sẽ sắp xếp sự quan trọng của các đặc tính trong bài toán phân loại và hồi quy.

2.2. Stochastic Gradient Descent

Mô hình ban đầu của mô hình Stochastic Gradient Descent là Gradient Descent (GD). Việc xây dựng mô hình SGD là để tránh những điểm cực tiểu cục bộ nhưng không phải là cực tiểu toàn cục của hàm mục tiêu (hình 3).



Hình 3. *Mô hình SGD.*

Thuật toán Stochastic Gradient Descent bao gồm các bước như sau:

- Khởi tạo: Khởi tạo ngẫu nhiên các tham số của mô hình.
- Đặt tham số: Xác định số lần lặp và tốc độ học (alpha) để cập nhật tham số.
- Vòng lặp giảm dần ngẫu nhiên: Lặp lại các bước sau cho đến khi mô hình hội tụ hoặc đạt số lần lặp tối đa:
  + Xáo trộn tập dữ liệu huấn luyện để tạo ra tính ngẫu nhiên.
  + Lặp lại từng ví dụ huấn luyện theo thứ tự được xáo trộn.
  + Tính toán độ dốc của hàm chi phí đối với các tham số mô hình bằng cách sử dụng mẫu huấn luyện hiện tại.
  + Cập nhật các tham số mô hình bằng cách thực hiện một bước theo hướng độ dốc âm, được chia tỷ lệ theo tốc độ học.
  + Đánh giá các tiêu chí hội tụ, chẳng hạn như sự khác biệt trong hàm chi phí giữa các lần lặp của độ dốc.
- Trả về các tham số được tối ưu hóa: Sau khi đáp ứng các tiêu chí hội tụ hoặc đạt đến số lần lặp tối đa, trả về các tham số mô hình được tối ưu hóa.

2.3. AdaBoost

Thuật toán AdaBoost, viết tắt của Adaptive Boosting (tăng cường thích ứng), là một mô hình học máy có giám sát, được sử dụng trong lớp các bài toán phân loại. Nó được gọi là tăng cường thích ứng vì các trọng số của mô hình được cập nhật lại sau mỗi phiên bản, với trọng số cao hơn được gán cho các phiên bản được phân loại không chính xác. Việc này được thực hiện để giảm độ lệch và phương sai cho việc học có giám sát. Thuật toán hoạt động theo nguyên tắc phát triển tuần tự. Ngoại trừ lần đầu tiên, mỗi lần học tiếp theo đều được phát triển từ những lần học đã tính toán trước đó. Nói một cách đơn giản, AdaBoosting sử dụng mô hình phân loại có điều chỉnh lại trọng số để thu được mô hình phân loại với độ chính xác cao nhất.

Dữ liệu bài toán ban đầu sẽ gồm có 2 nhãn phân loại (chia dữ liệu thành 2 nhóm) y={1,-1}. Dựa vào mô hình phân loại ban đầu ta sẽ có được những dự báo $\hat{f}(x) \in \{-1,1\}$ cho các dữ kiện đầu vào x_i.

Thuật toán AdaBoost bao gồm các bước như sau:

Khởi tạo trọng số quan sát.

$$\omega_i = \frac{1}{N}, \forall i = \underline{1, N} \quad (1)$$

Lặp lại quá trình huấn luyện chuỗi mô hình ở mỗi bước b,b=1,2,…B.

Khớp mô hình $\hat{f}^b$ cho tập huấn luyện sử dụng trọng số w_i cho mỗi quan sát (x_i,y_i).

Tính sai số huấn luyện:

$$r_b = \frac{\sum_{i=1}^{N} \omega_i 1\left(y_i \neq \hat{f}^b(x_i)\right)}{\sum_{i=1}^{N} \omega_i} \quad (2)$$

Tính trọng số quyết định cho từng mô hình:

$$\alpha_b = log\ (\frac{1-r_b}{r_b}) \quad (3)$$

Cập nhật trọng số cho từng quan sát:

$$\omega_i = \frac{\omega_i e^{\alpha_b 1\left(y_i \neq \hat{f}^b(x_i)\right)}}{\sum_{i=1}^{N} \omega_i} \quad (4)$$

Cập nhật dự báo cuối cùng:

$$\hat{f}(x) = sign(\sum_{i=1}^{p} \alpha_i \widehat{f^i}(x)) \quad (5)$$

2.4. Gradient Boosting

Mô hình Gradient Boosting cũng có ý tưởng tương tự như AdaBoosting đó là huấn luyện liên tiếp các mô hình. Tuy nhiên, phương pháp này không sử dụng sai số của mô hình để tính toán trọng số cho dữ liệu huấn luyện mà sử dụng phần dư. Xuất phát từ mô hình hiện tại, thuật toán sẽ xây dựng một cây quyết định để khớp phần dư từ mô hình liền trước. Điểm đặc biệt của mô hình này là thay vì cố gắng khớp giá trị hàm mục tiêu thì mô hình sẽ tìm cách khớp giá trị sai số của mô hình trước đó. Sau đó, thuật toán sẽ đưa thêm mô hình huấn luyện vào hàm dự báo để cập nhật dần phần dư. Mỗi một cây quyết định trong chuỗi mô hình có kích thước rất nhỏ với chỉ một vài nút được xác định bởi tham số độ sâu trong mô hình.

Trong mô hình này, hệ số co λ được sử dụng. Đây là một hệ số dương với trị số nhỏ, gần giống như learning rate (tốc độ học), có tác dụng kiểm soát tỷ lệ mà phương pháp Gradient Boosting cập nhật số dư. Các giá trị của hệ số co thường là 0.01 hoặc 0.001 v.v. tuỳ thuộc vào từng bài toán và từng bộ dữ liệu cụ thể. Thông thường khi λ rất nhỏ có thể cần sử dụng một giá trị rất lớn của số lượng cây để đạt được hiệu suất tốt. Một tham số khác cũng được sử dụng là độ sâu d của cây quyết định, nó đại diện cho số lần phân chia tối đa trong mỗi cây. Dưới đây là các bước của thuật toán Gradient Boosting:

Thiết lập hàm dự báo $\hat{f}(x)$=0 và số dư r_0=y cho toàn bộ quan sát trong tập huấn luyện.

Lặp lại quá trình huấn luyện cây quyết định theo chuỗi tương ứng với b=1,2,…,B. Với một lượt huấn luyện gồm các bước con sau đây:

+ Khớp một cây quyết định $\hat{f}^b$ có độ sâu d là trên tập huấn luyện (X,rb).

+ Cập nhật $\hat{f}(x)$ bằng cách cộng thêm vào giá trị dự báo của một cây quyết định, giá trị này được nhân với hệ số co λ:

$$\hat{f}(x) = \hat{f}(x) + \lambda \hat{f}^b(x) \quad (6)$$

+ Cập nhật phần dư cho mô hình:

$$r_{b+1} = r_b - \lambda \hat{f}^b(x) \quad (7)$$

Kết quả dự báo từ chuỗi mô hình sẽ là kết hợp của các mô hình con:

$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x) \quad (8)$$

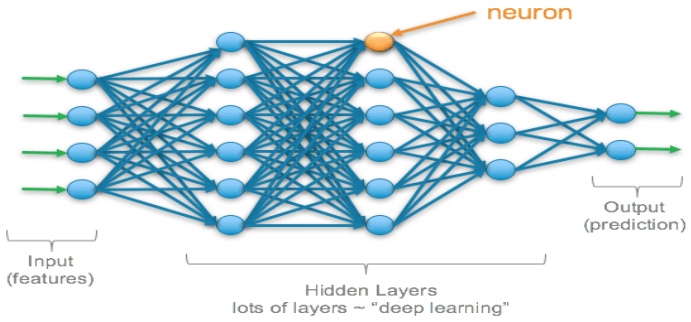
2.5. Mạng nơ ron nhân tạo nhiều lớp (Multi-Layer Perceptron)

Mạng nơ-ron nhân tạo nhiều lớp là một mô hình học máy dựa trên cấu trúc và hoạt động của bộ não con người. Nó bao gồm một mạng lưới các nơ-ron nhân tạo được kết nối với nhau (hình 4). Mỗi nơ-ron nhận các đầu vào từ các nơ-ron khác, tính tổng trọng số của các đầu vào đó và sau đó áp dụng hàm kích hoạt để tạo ra đầu ra. Các nơ-ron của một lớp liên kết với các nơ-ron của lớp liền kề thông qua các hàm kích hoạt (activation function) có trọng số. Các hàm kích hoạt này là các hàm phi tuyến, đặc trưng cho mối quan hệ phức tạp của dữ liệu.

Có hai hàm kích hoạt được sử dụng phổ biến trong mạng nơ ron nhân tạo là hàm Sigmoid và hàm ReLU

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (9)$$

$$R(x) = max\ (0, x\ ) \quad (10)$$



Hình 4. *Mô hình Multi-Layer Perceptron.*

Mô hình Logistic Regression cũng chính là một mạng rơ non nhân tạo với một lớp.

3. KẾT QUẢ

3.1. Dữ liệu đầu vào

Bảng 1. *Ví dụ về dữ liệu vận hành (SCADA data) thu thập được từ tua bin.*

DateTime	WEC: ava. windspeed	WEC: max. windspeed	WEC: min. windspeed	WEC: max. Rotation
5/1/2014 0:00	6.90	9.40	2.90	0.02
5/1/2014 0:09	5.30	8.90	1.60	0.01
5/1/2014 0:20	5.00	9.50	1.40	0.04
5/1/2014 0:30	4.40	8.30	1.30	0.08
5/1/2014 0:39	5.70	9.70	1.20	0.05
5/1/2014 0:49	8.40	9.90	5.40	0.04
5/1/2014 1:00	6.70	10.50	1.90	0.02
5/1/2014 1:09	3.70	9.10	1.10	0.04

Trong nghiên cứu này, tác giả đã tham khảo bộ dữ liệu vận hành (bảng 1) và dữ liệu trạng thái của tuabin gió (bảng 2) [11]. Dữ liệu vận hành được thu thập từ tuabin gió với công suất 3 MW,

cung cấp điện cho một cơ sở sản xuất thiết bị y sinh nằm gần bờ biển Ireland. Có 2 tệp dữ liệu được thu thập từ hệ thống tuabin là thông số vận hành và tình trạng của tuabin trong các khoản thời gian khác nhau. Trong đó, tình trạng hoạt động được thu thập trong khoảng thời gian dài nhất từ tháng 1 năm 2014 đến tháng 12 năm 2015 trong khi bộ dữ liệu ngắn nhất là thông số vận hành SCADA từ tháng 4 năm 2014 đến tháng 4 năm 2015. Do đó, khi thu thập dữ liệu SCADA, chúng ta có thể tham chiếu tình trạng hoạt động cũng như lỗi để đánh giá tình trạng làm việc các tuabin trong trong thời gian trên.

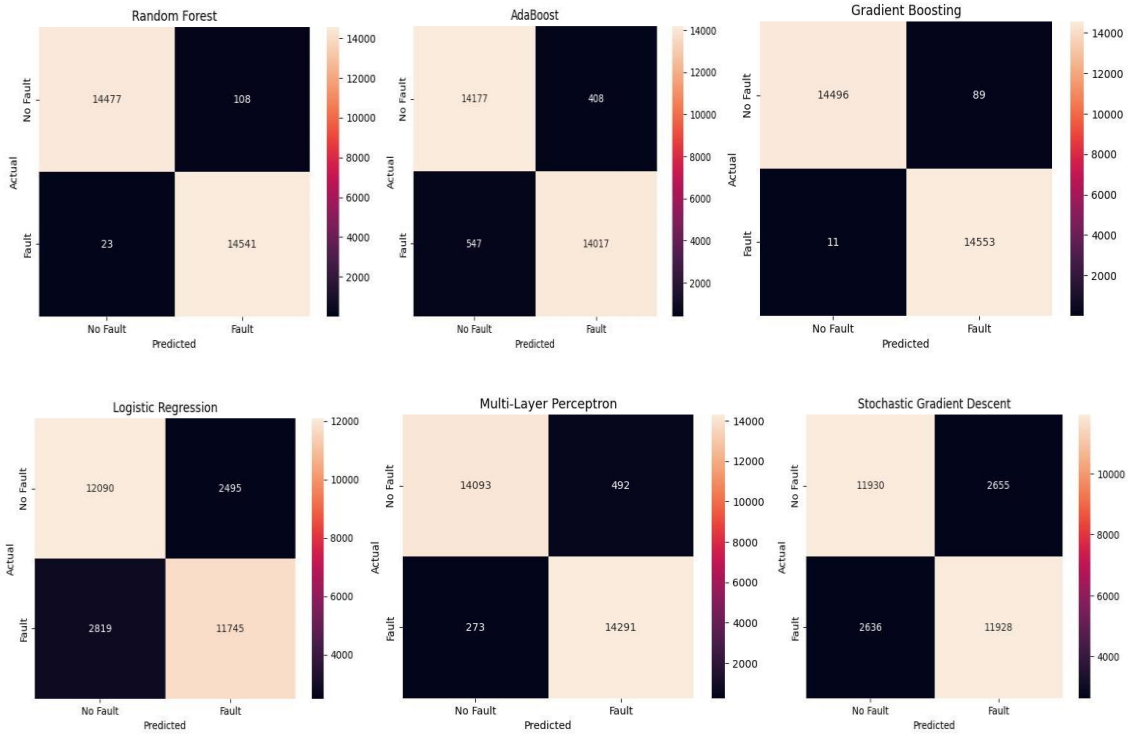
Bộ dữ liệu gồm 49134 dữ liệu các thông số vận hành của turbin gió và thông số trạng thái của nó. Bộ dữ liệu có đặc điểm là rất chênh lệch (hơn 90% trạng thái hoạt động là bình thường, còn lại là trạng thái sự cố), vì vậy, phương pháp tính toán hiệu suất của một mô hình phân loại theo các ngưỡng phân loại khác nhau (ROC-AUC) trở nên quan trọng, đặc biệt khi kết hợp với các kỹ thuật cân bằng dữ liệu như SMOTE và tinh chỉnh tham số bằng GridSearchCV.

Bảng 2. *Ví dụ về dữ liệu trạng thái vận hành của tuabin gió.*

DateTime	Main Status	Sub Status	Full Status	Status Text
24-04-14 12:37	0	0	0:00	Turbine in operation
25-04-14 19:27	71	104	71 : 104	Insulation monitoring : Insulation fault Phase U2
26-04-14 9:30	8	0	8:00	Maintenance
26-04-14 10:05	8	0	8:00	Maintenance
26-04-14 10:05	8	0	8:00	Maintenance

3.2. Kết quả

Kết quả so sánh về độ chính xác và thời gian nhận dạng sự cố của các thuật toán đề xuất được thể hiện ở bảng 3, ma trận hợp nhất (confusion matrix) của các thuật toán được thể hiện ở hình 5.



Hình 5. *Kết quả Confusion matrix của các mô hình học máy.*

Bảng 3. *Kết quả so sánh giữa các mô hình học máy đề xuất.*

Thuật toán	**Random Forest**	**Logistic Regression**	**AdaBoost**	**MLP**	**SGD**	**GDB**
Độ chính xác	99.55	81.77	96.72	97.38	81.85	99.66
Thời gian nhận dạng trung bình	0.0197	0.0002	0.0342	0.0003	0.0002	0.0005
Thời gian nhận dạng lớn nhất	0.0340	0.0014	0.0420	0.0015	0.0015	0.0015

Bộ tham số tối ưu của các mô hình được thực hiện bởi thuật toán GridSearchCV với kết quả như sau:

Random Forest Best parameters: {'max_depth': 20, 'max_features': 'log2', 'min_samples_split': 10, 'n_estimators': 300}

Logistic Regression Best parameters: {'C': 0.1, 'penalty': 'l2'}

AdaBoost Best parameters: {'learning_rate': 1, 'n_estimators': 300}

Multi-Layer Perceptron Best parameters: {'activation': 'relu', 'alpha': 0.05, 'hidden_layer_sizes': (50, 100, 50), 'learning_rate': 'adaptive', 'solver': 'adam'}

Stochastic Gradient Descent Best parameters: {'alpha': 0.01, 'eta0': 1, 'learning_rate': 'adaptive', 'loss': 'hinge', 'penalty': 'l1', 'tol': 0.0001}

GradientBoosting Best parameters: {'learning_rate': 0.1, 'max_depth': 10, 'n_estimators': 200}

Kết quả cho thấy, các thuật toán học máy đề xuất đều cho kết quả độ chính xác phát hiện sự cố trên 80%, trong đó, thuật toán GDB cho độ chính xác cao nhất là 99.66%. Các thuật toán khác có độ chính xác dao động trong khoảng từ 81-99%. Thuật toán GDB đã thể hiện được ưu điểm rõ rệt trong các bộ dữ liệu lớn có cấu trúc dạng bảng, do cơ chế khớp trên những cây quyết định có kích thước rất nhỏ trên những phần dư từ đó cả thiện hàm dự báo. Bên cạnh độ chính xác thì thời gian nhận dạng cũng là một yếu tố quan trọng để đánh giá hiệu quả của thuật toán. Các thuật toán được sử dụng đều cho thời gian nhận dạng tương đối nhanh. Thuật toán Logistic Regression và Stochastic Gradient Descent có cấu trúc đơn giản nên thời gian nhận dạng trung bình nhanh nhất, tuy nhiên, cũng do cấu trúc đơn giản nên hai thuật toán này có độ chính xác chỉ đạt khoảng 81%.

4. KẾT LUẬN

Trong nghiên cứu này, nhóm đã đề xuất sử dụng các thuật toán học máy nhằm phát hiện sự cố trên tuabin gió dựa trên dữ liệu đo được từ hệ thống SCADA. Hai thông số quan trọng để đánh giá các mô hình nhận dạng là độ chính xác và thời gian tính toán. Kết quả cho thấy, thuật toán GDB có độ chính xác cao nhất khi nhận dạng sự cố trong thời gian ngắn. Một số thuật toán khác cho độ chính xác tương đối cao, tuy nhiên, thời gian nhận dạng lâu hơn. Đối với việc ứng dụng các thuật toán học máy vào phát hiện lỗi, dữ liệu SCADA thu thập là hết sức quan trọng. Dữ liệu được thu thập và ứng dụng trong mô hình bao gồm dữ liệu trạng thái của tuabin gió cũng như dữ liệu vận hành. Việc kết hợp hai loại dữ liệu này có thể phát hiện sự cố cũng như chuẩn đoán chúng một cách chính xác hơn. Với đặc điểm dữ liệu đo được từ SCADA là loại dữ liệu có cấu trúc, các thuật toán học máy nhìn chung tỏ ra có ưu điểm về độ chính xác cũng như thời gian tính toán. Việc phát hiện sớm và chính xác các sự cố có thể xảy ra trên tuabin gió có thể giúp giảm thiểu chi phí bảo dưỡng, vận hành, cũng như thời gian ngừng cung cấp điện do hỏng hóc, nâng cao độ tin cậy của hệ thống điện. Trong các nghiên cứu tiếp theo, nhóm sẽ tập trung vào việc cải thiện các mô hình hiện có nhằm nâng cao độ chính xác của bài toán nhận dạng sự cố.

TÀI LIỆU THAM KHẢO

[1]. McMillan, D.; Ault, G. *"Quantification of condition monitoring benefit for offshore wind turbines,"* Wind Energy, vol. 31, pp. 267-285, (2007).
[2]. Drewry, M.; Georgiou, G. *"A review of NDT techniques for wind turbines,"* Insight, vol. 49, pp. 137-141, (2007).
[3]. Aval, S.; Ahadi, A. *"Wind turbine fault diagnosis techniques and related algorithms,"* Int. J. Renew. Energy Res., vol. 6, pp. 80-89, (2016).
[4]. Hameed, Z.; Hong, Y.; Cho, Y.; Ahn, S.; Song, C. *"Condition monitoring and fault detection of wind turbines and related algorithms: A review,"* Renew. Sustain. Energy Rev., vol. 13, pp. 1-39, (2009).
[5]. Lu, B.; Li, Y.; Wu, X.; Yang, Z. *"A review of recent advances in wind turbine condition monitoring and fault diagnosis,"* In Proceedings of the IEEE Power Electronics and Machines in Wind Applications, Lincoln, NE, USA, pp. 1-7, (2009).
[6]. Y. Shi, Y. Liu and X. Gao, *"Study of Wind Turbine Fault Diagnosis and Early Warning Based on SCADA Data,"* in IEEE Access, vol. 9, pp. 124600-124615, (2021).
[7]. María Peco Chacón and F. Pedro García Márquez, *"SCADA data analytics for fault detection and diagnosis of wind turbines,"* 7th International Conference on Control, Instrumentation and Automation (ICCIA), Tabriz, Iran, pp. 1-6, (2021).
[8]. Karadayi, Y. Kuvvetli and S. Ural, *"Fault-related Alarm Detection of a Wind Turbine SCADA System,"* 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, pp. 1-5, (2021).
[9]. S. Zhang, Y. Wang, M. Liu and Z. Bao, *"Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM,"* in IEEE Access, vol. 6, pp. 7675-7686, (2018).
[10].C. Ren, H. Yuan, Q. Li, R. Zhang and Y. Xu, *"Pre-Fault Dynamic Security Assessment of Power Systems for Multiple Different Faults via Multi-Label Learning,"* in IEEE Transactions on Power Systems, vol. 38, no. 6, pp. 5501-5511, (2023).
[11].https://github.com/gauravjgogoi/Data-Analytics-and-Classification-Model-for-Failure-Detection-of-Wind-Turbine

ABSTRACT

Fault detection in wind turbine based on machine learning

Renewable energy, in general, and wind energy, in particular, are receiving increasing attention with the goal of reducing greenhouse gas emissions and producing clean energy. In recent years, wind farms and plants have significantly increased, driving wind energy to become an immensely potential energy source. However, due to the unpredictable nature of wind energy, ensuring the safe operation of wind turbine systems, minimizing downtime due to malfunctions play a crucial role in optimizing production costs and enhancing the system's reliability. In this study, the authors propose using machine learning models to detect issues occurring in wind turbine systems. Operational parameters measured from SCADA are used as input data for the machine learning models. The results indicate that machine learning models can detect issues in the wind turbine system with an accuracy of over 99%, with model training taking only tens of milliseconds.

Keywords: Renewable energy; Wind turbine; Fault; SCADA; Machine learning.