

Áp dụng một số phương pháp máy học để dự báo các vấn đề có thể phát sinh của bơm điện chìm trong quá trình sử dụng

Phạm Sơn Tùng¹, Trịnh Quang Khải^{1*}

¹ Khoa Kỹ thuật Địa chất và Dầu khí, Trường Đại học Bách Khoa – Đại học Quốc gia Thành phố Hồ Chí Minh

TỪ KHOẢ

Bơm điện chìm
Mạng nơron nhân tạo
Extreme gradient boosting

TÓM TẮT

Bơm điện chìm (ESP) từ lâu đã trở thành một trong những phương pháp nâng nhân tạo hiệu quả nhất góp phần duy trì tốc độ khai thác và kéo dài tuổi thọ của các giếng dầu. Bên cạnh những lợi ích to lớn, ESP cũng đặt ra vấn đề về việc giám sát, sửa chữa cho hệ thống ESP tránh những tình trạng hư hỏng. Đối mặt với những thiệt về kinh tế do hư hỏng ESP gây ra rất nhiều các phương pháp được đề ra để dự báo trước được những bất thường của hệ thống ESP, từ đó đưa ra các kế hoạch thay thế và sửa chữa. Nhiệm vụ của người kĩ sư dầu khí là lựa chọn được phương pháp hiệu quả và ít tốn kém thời gian cũng như chi phí nhất. Những năm gần đây AI – Artificial Intelligence (Trí tuệ nhân tạo) phát triển vô cùng mạnh mẽ trong đó ML – Machine Learning (Máy học) là một thành tựu nổi bật. Với ML việc dự báo lỗi hay xu hướng hoạt động của ESP không còn là công việc quá khó khăn như trước. ML sử dụng bộ số liệu hoạt động trong quá khứ của các ESP trước đó để dự báo những sự kiện có thể xảy ra trong tương lai một cách chính xác và dễ dàng. Từ đó giúp giảm chi phí và thời gian để sửa chữa thay thế hệ thống ESP. Trong nghiên cứu này sẽ đề cập đến hai phương pháp ML hiệu quả và phổ biến nhất trong thời điểm hiện nay là Extreme Gradient Boosting (XGboosting) và Mạng Nơron nhân tạo (Artificial Neural Network - ANN). Bài nghiên cứu này sẽ sử dụng hai phương pháp này để dự báo các lỗi có thể phát sinh trong quá trình hoạt động của bơm điện chìm. Kết quả của hai thuật toán sẽ được so sánh với nhau để tìm ra được mô hình tối ưu hơn trong công tác dự báo lỗi ESP. Ngoài ra bài nghiên cứu sẽ đánh giá mức độ ảnh hưởng của từng thông số đến lỗi của ESP. Từ đó giúp ta hiểu hơn về ESP và sớm đề ra kế hoạch nhằm ngăn ngừa các lỗi ESP tiềm ẩn.

KEYWORDS

Electrical submersible pump
Artificial Neural Network
xtreme gradient boosting

ABSTRACT

The electrical submersible pump (ESP) has long been one of the most effective artificial-lift methods to improve the production rates and extend the life of oil wells. Although ESP has many great advantages, the monitoring and repairing process of ESP poses some problems. In order to minimize the economic losses caused by ESP errors, many methods have been proposed to predict the abnormalities of ESP system before making replacement and repair plans for ESP. The petroleum engineer must choose the most effective, time-saving and economical method. In recent years, AI - Artificial Intelligence has been extremely strongly developing, in which ML - Machine Learning is a prominent achievement. With ML, the prediction of ESP's error and operating trend is no longer as difficult as before. ML uses the historical data set of previous ESPs to accurately and easily forecast possible future events. As the result, the cost and the time to repair and replace the ESP system may reduce. This research mentioned two most effective and popular ML methods: Extreme Gradient Boosting (XGboosting) and Artificial Neural Network (ANN). These two methods were used to predict errors that may arise during the ESP operation. The results of the two algorithms were compared together to find out the more optimal model in ESP error prediction. In addition, the research also evaluated the influence of each parameter on the error of ESP with the aim of preventing hidden ESP errors.

1. Giới thiệu

Thu hồi dầu tăng cường bằng các hệ thống nâng nhân tạo đang trở thành giai đoạn quan trọng và có tính quyết định đến ngành dầu khí. Hiện nay hệ thống hiệu quả và đang được sử dụng phổ biến nhất

là hệ thống ESP. ESP sở hữu nhiều tính năng như: làm việc ở độ sâu lớn, các giếng có quỹ đạo phức tạp, hoạt động hiệu quả với mọi loại chất lưu và chế độ dòng chảy... Nhược điểm của ESP là cấu tạo phức tạp và phải nâng toàn bộ hệ thống lên trong mỗi lần sửa chữa hay bảo trì thiết bị. Vì vậy khi hệ thống gặp sự cố, công đoạn sửa chữa sẽ vô

*Liên hệ tác giả: trinhquangkhaihn24051999@outlook.com.vn

Nhận ngày 21/05/2021, sửa xong ngày 12/06 /2021, chấp nhận đăng 04/11/2021

cùng mất thời gian và tốn chi phí lớn do phải ngừng sản xuất. Đối mặt với việc đó việc giám sát cũng như dự báo trước các sự cố liên quan đến ESP là vô cùng cần thiết và mang tính quyết định đến quá trình khai thác.

Trong những năm gần đây thuật toán máy học đang phát triển vô cùng mạnh mẽ, thuật toán này đã được áp dụng cho nhiều ngành nghề và lĩnh vực trong xã hội và dầu khí cũng không nằm ngoài xu thế này. Hiện nay có nhiều 3 loại mô hình chính giám sát và dự báo hư hỏng cho ESP bao gồm:

Binary Classification models (phân loại nhị phân): mục tiêu của mô hình là xác định ESP có gặp sự cố hay không trong một khoảng thời gian hay một tương lai gần.

Multi-Class Classification models (phân loại dán nhãn): mục tiêu của mô hình là dự đoán sự cố ESP trong nhiều khung thời gian, xác suất gây ra sự cố của một bộ phận ESP hay một nhân tố môi trường nào đó. Mô hình còn có khả năng dự đoán nguyên nhân gây nên sự cố.

Regression models (mô hình hồi quy): khác với hai mô hình phân loại trên mô hình này được dùng để dự đoán tuổi thọ và thời gian hỏng hóc của bơm.

Trên cơ sở ba loại mô hình đó đã có nhiều phương pháp được chuyên gia dầu khí sử dụng như mô hình cảnh báo thời gian thực bằng thuật toán Principal Component Analysis (PCA) Model và Non-Linear Residual Models của Abhijeet, et al [1]. Mô hình phân loại lỗi ESP bằng thuật toán Decision Tree của Barrios, et al [2]. Mô hình xác định khoảng thời gian xuất hiện lỗi của ESP bằng phương pháp Time-series Shapelets của Om, et al [7]. Dự báo thời gian xảy ra lỗi của ESP cũng có thể thực hiện bằng phương pháp phân tích số liệu kết hợp với các thuật toán học máy tăng cường (Boost) của Rinat, et al [9].

Ngoài ra phương pháp PCA cũng được dùng để giám sát sự thay đổi của các tính năng động của ESP, từ đó tìm ra vùng vận hành nguy hiểm trong [3] và [5]. Trong sự phát triển của internet, hệ thống vận hành, giám sát ESP bằng IoT cũng được áp dụng bởi Nico, et al [6].

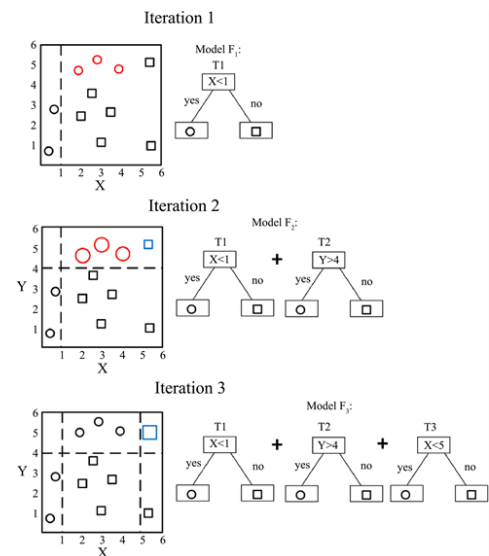
Nghiên cứu này sẽ trình bày một cách tiếp cận khác, đó là xây dựng một mô hình sử dụng 2 thuật toán khác nhau là Extreme Gradient Boosting và Mạng Nơron nhân tạo để dự đoán lỗi của ESP trong quá trình sử dụng thiết bị.

Khác với mô hình phân loại nhị phân cảnh báo lỗi của ESP như [1], [3], [5], [7] và [9] mô hình trong nghiên cứu sẽ dự đoán tình trạng của ESP và gắn đúng tên lỗi của ESP với từng bộ số liệu được cung cấp. Mô hình này có mục đích tương tự như [2], nhưng như ta đã biết Decision tree có một độ chính xác không cao, cũng như không được tối ưu như XGboosting và Mạng Nơron nhân tạo như đánh giá trong [8]. Bên cạnh đó mô hình sẽ đánh giá mức độ ảnh hưởng của các thông số đến lỗi ESP và sự thay đổi của dự đoán khi lược bỏ những thông số ít ảnh hưởng.

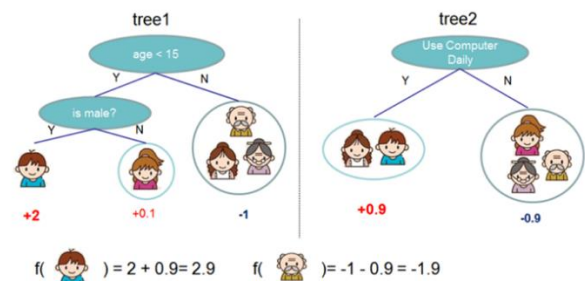
2. Cơ sở lý thuyết

2.1. Extreme Gradient Boosting (XGboosting)

XGboosting là một thuật toán học máy tăng cường là một phương pháp tổng hợp nhằm giảm phương sai và bias. Trong quá trình hoạt động phân loại, thuật toán sẽ tạo ra nhiều bộ phân loại yếu với độ chính xác vừa phải, sau đó sẽ gộp những bộ phân loại này tạo nên một bộ phân loại mạnh cuối cùng, quá trình này được biểu diễn ở Hình 1. Sau khi những bộ phân loại yếu được thêm vào dữ liệu của chúng sẽ được tái trọng số, trong đó những mẫu đã phân loại đúng sẽ giảm trọng số, những mẫu phân loại sai sẽ tăng trọng số. XGboosting được biết đến với độ chính xác và hiệu quả cao so với những thuật toán tương tự và tiết kiệm được tài nguyên [12]. Về độ phổ biến XGboosting là thuật toán được sử dụng trong nhiều lĩnh vực cũng như các cuộc thi về phân tích dữ liệu.



Hình 1. Trực quan quá trình hoạt động của XGboosting [10].



Hình 2. Phân loại đơn giản bằng XGboosting [11].

XGboosting đã được tạo ra bởi Tianqi Chen [12] như sau:

Cho một bộ dữ liệu có n mẫu và m đặc trưng:

$$D = \{(x_i, y_i)\} (|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R})$$

Mô hình học của XGboosting:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F$$

Trong đó: $F = \{f(x) = w_{q(x)}\} (q: \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T)$ là không gian của cây, T là số lá trong cây, w là trọng số của nút lá, q là cấu trúc của một cây ánh xạ mẫu dữ liệu vào nút lá tương ứng, f_k : cấu trúc các cây k độc lập của mô hình,

$\hat{y}_i^{(t)}$ là giá trị dự đoán của ví dụ thứ i và vòng lặp thứ t .

Hàm học của XGboosting được trình bày như sau:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

Trong đó: $\Omega = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$

l là hàm mất mát (loss function) thể hiện độ sai khác của giá trị thực và giá trị dự đoán. Ω là hàm thể hiện độ phức tạp của mô hình, được thêm vào để hạn chế hiện tượng overfitting. Tiến trình học được thể hiện thông qua các công thức bên dưới:

$$L^{(t)} = \sum_{i=1}^n l[y_i, \hat{y}_i^{(t-1)} + f_t(x_i)] + \Omega(f_t)$$

$$L^{(t)} \sim \sum_{i=1}^n l \left[y_i, \hat{y}_i^{(t-1)} + g_t f_t(x_i) + \frac{1}{2} h_t f_t^2(x_i) \right] + \Omega(f_t)$$

Trong đó: $g_t = \partial_{\hat{y}_i^{(t-1)}} l$ và $h_t = \partial_{\hat{y}_i^{(t-1)}}^2 l$

$$L^{(t)} = \sum_{i=1}^n l \left[g_t f(x_i) + \frac{1}{2} h_t f^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

$$L^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} (h_i + \gamma) w_j^2 \right] + \gamma T$$

2.2. Mạng nơ-ron nhân tạo

Một Mạng Nơ-ron nhân tạo thường được tổ chức các nơ-ron thành từng lớp và mỗi lớp chịu trách nhiệm cho một công việc cụ thể. Mạng Nơ-ron nhân tạo thường có 3 lớp: lớp nhập hay lớp đầu vào, lớp ẩn và lớp xuất [11]:

Lớp nhập (input layer) cung cấp cho mạng các số liệu cần thiết. Số lượng nơ-ron trong lớp nhập tương ứng với số lượng thông số đầu vào được cung cấp cho mạng và các thông số đầu vào này được giả thiết ở dạng véc tơ.

Lớp ẩn (hidden layer) chứa các nơ-ron ẩn giúp kết nối giá trị đầu vào đến giá trị đầu ra. Một mạng nơ-ron có thể có một hoặc nhiều lớp ẩn chịu trách nhiệm chính cho việc xử lý các nơ-ron của lớp nhập và đưa các thông tin đến nơ-ron của lớp xuất. Các nơ-ron này thích ứng với việc phân loại và nhận diện mối liên hệ giữa thông số đầu vào và thông số đầu ra.

Lớp xuất (output layer) chứa các nơ-ron đầu ra nhằm chuyển thông tin đầu ra của các tính toán từ Mạng Nơ-ron nhân tạo đến người dùng. Một Mạng Nơ-ron nhân tạo có thể được xây dựng để có nhiều thông số đầu ra. Số nơ-ron của lớp nhập và lớp xuất sẽ do bài toán quyết định, số nơ-ron lớp ẩn và số lớp ẩn sẽ do người nhập quyết định. Tuy nhiên, việc chọn loại và số lượng của thông số đầu vào có ảnh hưởng lớn đến chất lượng của mạng.

Mô hình toán học của Mạng Nơ-ron nhân tạo được trình bày như sau:

$$y(x) = f \left(\sum_{i=1}^n w_i x_i \right)$$

Trong đó: y là giá trị đầu ra ứng với biến x , f là hàm kích hoạt hay hàm truyền, w là trọng số liên kết của nơ-ron, x là các giá trị đầu vào.

2.3 Ma trận nhầm lẫn (Confusion Matrix)

Là một phương pháp đánh giá kết quả của những bài toán phân loại với việc xem xét cả những chỉ số về độ chính xác và bao quát của các dự đoán cho từng lớp. Một ma trận nhầm lẫn gồm 4 chỉ số sau đối với mỗi lớp phân loại:

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Hình 3. Ma trận nhầm lẫn.

TP (True Positive): Số lượng dự đoán chính xác nhãn có giá trị Positive

TN (True Negative): Số lượng dự đoán chính xác nhãn có giá trị Negative

FP (False Positive): Số lượng dự đoán sai nhãn có giá trị Positive

FN (False Negative): Số lượng dự đoán sai nhãn có giá trị Negative

Từ 4 chỉ số này ta có 2 con số để đánh giá độ chính xác của mô hình:

Precision: trong tất cả các dự đoán Positive có bao nhiêu dự đoán chính xác:

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: trong tất cả các dự đoán Positive có bao nhiêu dự đoán là do mô hình đưa ra:

$$\text{Recall} = \frac{TP}{TP+FN}$$

Ngoài ra ta còn có một thông số khác là **Accuracy:** trong tất cả các dự đoán thì tỷ lệ dự đoán đúng là bao nhiêu

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

3. Thu thập số liệu đầu vào

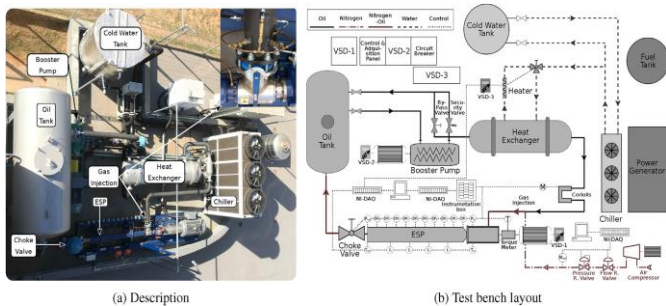
Bộ dữ liệu trong bài nghiên cứu lấy từ thí nghiệm của M. Barrios Castellanos, et al [2]. Mô hình thí nghiệm được mô tả trong hình 4 bao gồm: ESP, bồn chứa (tank), bơm tăng áp (booster pump) (hai trục vít), bộ trao đổi nhiệt, bộ truyền động, van tiết lưu (choke valve), hệ thống phun khí bao gồm máy phun khí, máy nén, van điều chỉnh và ống mềm.

Chất lưu được sử dụng là hỗn hợp hai phần khí và lỏng, phần lỏng là hỗn hợp dầu có API = 13-14^oAPI và dầu diesel S500, phần khí là nitơ.

Thí nghiệm được cài đặt để kiểm soát chính xác việc tăng áp suất, lưu lượng, nhiệt độ. Các biến được giám sát ở thí nghiệm này bao gồm: ΔP (bar) (chênh lệch giữa áp suất hút và áp suất xả của máy bơm), m_{liquid} (kg/s) (khối lượng lưu lượng của phần lỏng), T_{in} (°C) (nhiệt độ đầu vào của bơm), ω (rpm) (moment xoắn của bơm) và m_{gas} (kg/s) (khối lượng lưu lượng của phần khí).

Mục đích của thí nghiệm là thay đổi của các thông số của hệ thống thí nghiệm từ đó tìm ra điều kiện xảy ra các trạng thái của một ESP có thể gặp phải. Có hai trạng thái được đề cập ở đây là trạng thái bình thường (Normal) tại đó ESP hoạt động bình thường, trạng thái thứ hai là trạng thái lỗi (Fault) tại đó ESP gặp trục trặc, hư hỏng. Trạng thái lỗi bao gồm bốn trường hợp. GI (Gas Increasing) tăng bất thường của khí vào ESP gây ra hiện tượng surging or gas locking phenomena. VI (Viscosity Increasing) chất lưu tăng độ nhớt làm giảm hiệu suất của bơm và tăng ma sát trong đường ống. CV (Choke Valve closing) van tiết lưu đột ngột bị đóng. DIP (Decreasing of Input Pressure) giảm áp suất dầu vào.

Bộ số liệu gồm 806 mẫu trong đó có 392 mẫu Norma, 37 mẫu CV, 45 mẫu DIP, 42 mẫu VI, 290 mẫu GI được thể hiện rõ ở Hình 5.



Hình 4. Mô hình thí nghiệm ESP [2].

Fault	T _{in} (°C)	ω [rpm]	m _{gas} [kg/h]	Δp [bar]	m _{liquid} [kg/s]	# Experiments
Normal	50	3500	1-14	0.00-16.40	9-39	47
		3000	1-14	0.15-13.14	15-23	45
		2400	1-14	0.00-6.79	4-21	70
	45	1800	1-14	0.45-0.71	10-11	5
		3500	1-14	0.25-15.84	18-27	29
		3000	1-14	0.21-11.61	15-23	43
CV	50	2400	1-14	0.00-6.16	11-17	43
		3500	1-14	0.07-12.75	16-27	39
		3000	1-14	0.20-9.46	13-21	46
	40	2400	1-14	0.14-5.70	9-15	25
		3500	1-5	0.55-16.08	3-39	17
		2400	1-5	0.01-2.24	5-19	11
DIP	50	3500	1-5	0.17-9.06	11-25	9
		3500	2-5	0.61-16.43	3-39	25
		2400	2-5	0.00-2.36	4-21	12
	40	3500	2-5	1.67-8.90	13-21	8
		2400	4	0.00-1.02	11-15	42
		3500	0-14	1.23-19.63	17, 23, 28	26
VI	30-45	3000	0-14	0.37-14.80	14, 19, 23	34
		2400	0-14	0.18-8.97	11, 15, 18	49
		1800	0-14	0.43-0.69	10, 11	3
	45	3500	0-14	1.57-18.10	17, 22, 27	25
		3000	0-14	0.20-14.01	14, 19, 23	36
		2400	0-14	0.00-7.96	10, 14, 17	33
GI	40	3500	0-14	1.77-15.52	17, 22, 27	22
		3000	0-14	0.00-12.66	13, 17, 21	38
		2400	0-14	0.08-8.76	9, 12, 15	24

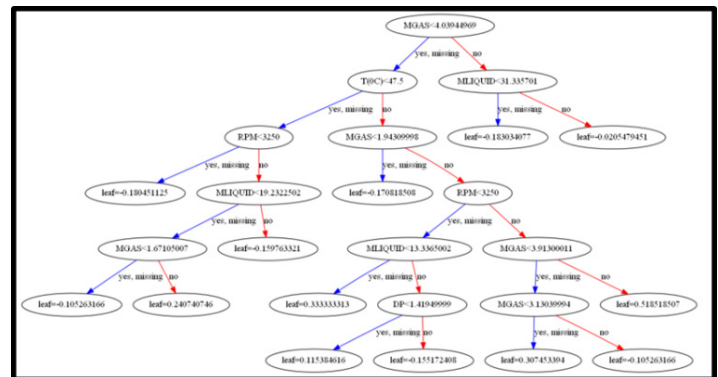
Hình 5. Bảng số liệu đầu vào của mô hình [2].

4. Xây dựng mô hình dự đoán lỗi của bơm điện chìm bằng hai phương pháp XGboosting và Mạng nơ-ron nhân tạo

Mục tiêu của việc xây dựng mô hình là từ những dữ liệu đầu vào ban đầu là các thông số, tham số ảnh hưởng tới hoạt động của ESP để dự đoán chính xác nhất thông số đầu ra là tình trạng của ESP.

Các mô hình sử dụng 2 phương pháp có học máy là XGboosting và Mạng Nơ-ron nhân tạo. Độ chính xác của mô hình được đánh giá bằng ma trận nhầm lẫn.

Mô hình XGboosting tiến hành trộn lẫn dữ liệu mẫu sau đó chia bộ dữ liệu ban đầu thành hai phần: bộ dữ liệu đào tạo chiếm 70% (564 mẫu) bộ dữ liệu ban đầu bộ, bộ dữ liệu kiểm tra chiếm 30% (242 mẫu) còn lại của bộ dữ liệu ban đầu. Mô hình XGboosting được thiết lập xây dựng 100 cây quyết định (e_estimator = 100). Độ sâu tối đa của cây là 6 (max_depth=6), giá trị khởi đầu là 0.5, cây đầu tiên của mô hình được biểu diễn ở hình, từ đó dự đoán tình trạng của hệ thống ESP.

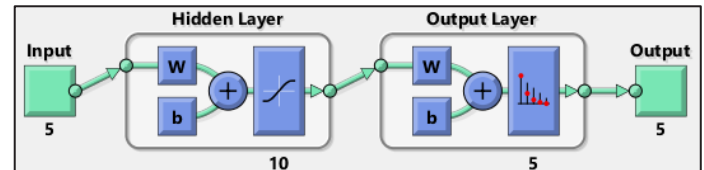


Hình 6. Cây quyết định đầu tiên của Xgboosting.

Mô hình Mạng Nơ-ron nhân tạo chia bộ dữ liệu ban đầu thành 3 phần bao gồm:

Bộ dữ liệu đào tạo (training data) chiếm 70% bộ dữ liệu ban đầu, bộ dữ liệu đối sánh chéo (validation data) chiếm 15% bộ dữ liệu ban đầu, bộ dữ liệu kiểm tra (test data) chiếm 15% còn lại của bộ dữ liệu ban đầu.

Mô hình Mạng Nơ-ron nhân tạo sử dụng mạng nơ-ron có 2 lớp truyền thẳng, lớp ẩn có 10 nơ-ron.



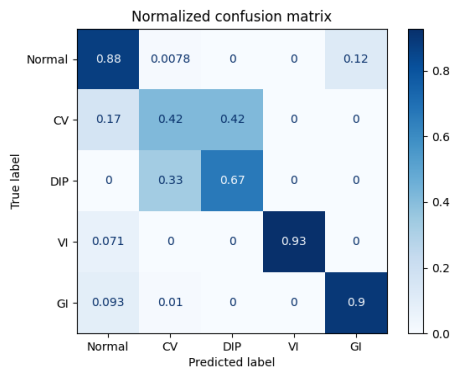
Hình 7. Mô hình mạng nơ-ron.

5. Kết quả và thảo luận

5.1. XGboosting

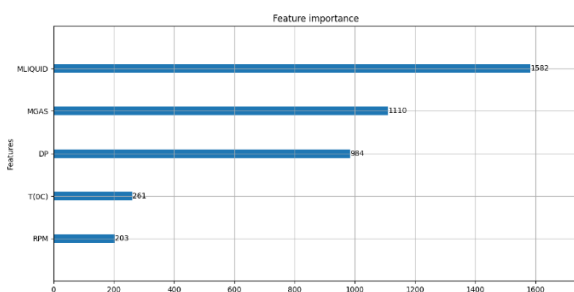
Mô hình XGboosting cho ta các dự giá trị dự đoán có độ chính xác cao ở các hiện tượng như Normal (88 %), VI (93 %), GI (90 %) và độ chính xác ở mức khá như DIP (67%) và trung bình như CV (42%). Độ chính xác tổng thể là Accuracy = 85.33834586466166 %. Nguyên nhân của việc dự đoán không chính xác hai lỗi CV và DIP là do hai

nguyên nhân sau. Thứ nhất là do số lượng mẫu xuất hiện hai hiện tượng này là khá ít với 37 mẫu ở CV và 45 mẫu ở DIP. Nguyên nhân thứ hai là do điều kiện xuất hiện của hai hiện tượng này là gần như tương đồng nhau nên gây ra hiện tượng nhầm lẫn, dự báo sai qua lại giữa hai hiện tượng.



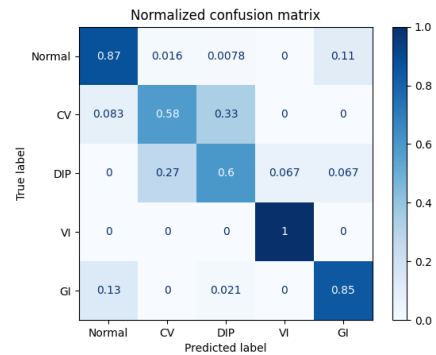
Hình 8. Ma trận nhầm lẫn đánh giá độ chính xác của Xgboosting.

XGboosting cung cấp đánh giá mối tương quan của các đặc tính đối với kết quả dự đoán của mô hình. Hình 9 thể hiện sự ảnh hưởng của các đặc tính từ thấp đến cao đối với mô hình. Ta có m_{liquid} có sự ảnh hưởng lớn nhất với 1582 điểm, kế đến là m_{gas} và ΔP có sức ảnh hưởng gần như nhau lần lượt là 1110 điểm và 984 điểm, hai đặc trưng còn lại là T và ω có ảnh hưởng khá nhỏ nếu so sánh với 2 đặc trưng kể trên. Nhờ vào bảng đánh giá này ta chạy lại mô hình với việc lược bỏ bớt đi hai thông số ít ảnh hưởng hơn là T và ω để xem có thể loại bỏ bớt nhiều cho mô hình hay không.



Hình 9. Tầm ảnh hưởng của các đặc trưng tác động đến mô hình.

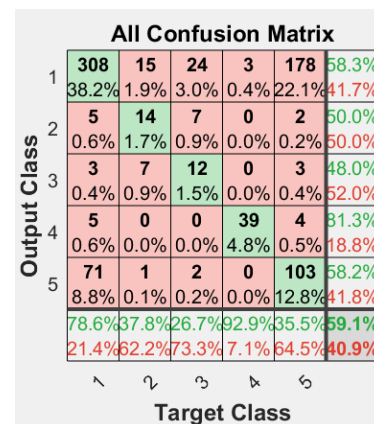
Việc chạy lại mô hình với hai thông số được lược bỏ làm tăng tính chính xác của việc dự báo như trong hình 10 CV tăng từ 42 đến 58 %, VI tăng từ 93 % đến 100 %, nhưng những dự báo về 3 hiện tượng còn lại bị giảm xuống từ đó làm cho độ chính xác tổng thể giảm còn 83.83458646616542 %. Những dự báo về Normal, GI, DIP tuy có giảm nhưng lại giảm một lượng không đáng kể so với việc dự báo CV tăng từ mức dưới trung bình đến mức trung bình. Kết luận để tăng khả năng dự đoán đúng ở từng nhãn có thể ta phải hy sinh đi độ chính xác tổng thể của mô hình.



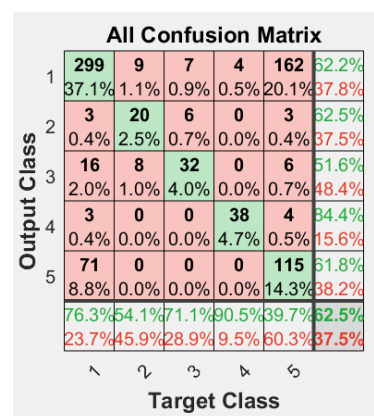
Hình 10. Ma trận nhầm lẫn sau khi đã lược bỏ hai đặc trưng T và ω .

5.2. Mạng Nơron nhân tạo

Mô hình Mạng Nơron nhân tạo cho ta các dự báo ở mức trung bình với nhãn Normal 58,3 %, CV 50 %, DIP 48 %, GI 58,2 % và đặc biệt khá chính xác ở nhãn VI 81,3 %. Vẫn giống với mô hình XGboosting với CV và DIP là hai nhãn có tỷ lệ dự báo đúng thấp nhất. Độ chính xác tổng thể của mô hình nằm ở mức 59,1 %.



Hình 11. Ma trận nhầm lẫn khi lớp ẩn có 10 nơron.



Hình 12. Ma trận nhầm lẫn khi lớp ẩn có 20 nơron.

Với 10 nơron lớp ẩn tỷ lệ dự báo không được như kì vọng nên ta tăng số nơron trong lớp ẩn lên 20 nơron và nhận được kết quả bất

là tỷ lệ dự đoán chính xác đều tăng lên nhân Normal (58,3 → 62,2 %), CV(50 → 62,5 %), DIP (48 → 51,6 %), VI (81,3 → 84,4 %), GI (58,2 → 61,2 %). Từ đó làm tăng độ chính xác tổng thể lên thành 62,5 %.

Vấn giữ số nơron lớp ẩn là 20 nơron và thử áp dụng cách làm tương tự XGboosting là lược bỏ 2 đặc trưng là T và ω ta sẽ được độ chính xác toàn phần là 56,9 % thấp hơn Mạng Nơron nhân tạo với 10

nơron lớp ẩn. Rút ra kết luận việc loại bỏ 2 đặc trưng trên không có hiệu trong mô hình Mạng Nơron nhân tạo.

5.3. So sánh độ chính xác của hai mô hình XGboosting và Mạng Nơron nhân tạo

Ta so sánh hai kết quả tốt nhất của 2 mô hình trong Bảng 1.

Bảng 1. So sánh độ chính xác của hai mô hình XGboosting và Mạng Nơron nhân tạo.

Model	Normal	CV	DIP	VI	GI	Accuracy
XGboosting	87 %	58 %	60 %	100 %	85 %	83,83 %
Mạng Nơron nhân tạo	62,2 %	62,5 %	51,6 %	84,4 %	61,8 %	62,5 %

Từ bảng trên ta thấy sự vượt trội của XGboosting trong bài toán phân loại lỗi của hệ thống ESP. Nguyên nhân của sự vượt trội này là do Mạng Nơron nhân tạo thích hợp hơn với những bài toán hồi quy hơn là những bài toán phân loại. XGboosting sử dụng các cây quyết định để xác định xác suất xuất hiện của các nhân, trong khi XGboosting hoàn toàn dùng đến các hàm tính toán do đó dễ bị nhiễu hơn.

6. Kết luận

Bài nghiên cứu đã trình bày một phương pháp tiếp cận chủ động để dự đoán lỗi của bơm điện chìm. Mô hình trong bài đã sử dụng hai phương pháp máy học thông dụng nhất hiện nay là XGboosting và Mạng Nơron Nhân Tạo. Kết luận được đưa ra là phương pháp XGboosting thích hợp hơn trong vấn đề phân loại lỗi bơm điện chìm. Kết quả của phương pháp này không chỉ có độ chính xác cao mà còn cho ta thấy sự ảnh hưởng của các tham số đối với quá trình phân loại. Thông qua kết quả của các mô hình ta thấy được hiệu quả của việc loại bớt những tham số có ảnh hưởng ít đến kết quả đầu ra trong bộ số liệu ban đầu. Việc loại bỏ những tham số nhiễu trong bộ dữ liệu làm quá trình dự đoán chính xác hơn, tiết kiệm tài nguyên cho mô hình.

Lời cảm ơn

Nghiên cứu được tài trợ bởi Trường Đại học Bách khoa – ĐHQG-HCM trong khuôn khổ Đề tài mã số To-ĐCDK-2021-07.

Tài liệu tham khảo

[1]. Abhijeet Sandeep Bhardwaj, Rahul Saraf, Geetha Gopakumar Nair, and Sridharan Vallabhaneni, Halliburton (2019), “SPE-197911-MS-Real-Time Monitoring and Predictive Failure Identification for Electrical Submersible Pumps”, Abu Dhabi International Petroleum Exhibition & Conference held in Abu Dhabi, UAE, 11-14 November 2019.

[2]. Barrios Castellanos Mauricio, Alberto Luiz Serpaa, Jorge Luiz Biazussib, Monte Verde William, do Socorro Dias Arrifano Sassim Natacheb (2020), “Fault identification using a chain of decision trees in an electrical submersible pump operating in a liquid-gas flow”.

[3]. Long Peng, Guoqing Han, Xianfu Sui, Arnold Landjobo Pagou, Liying Zhu, and Jin Shu (2021),” *Predictive Approach to Perform Fault Detection in Electrical Submersible Pump Systems*”.

[4]. Mike Pannel, Jeffrey Hsiung and V.B.Putcha, OspreyData Inc, (2018), “SPE-190090-MS Detecting Failures and Optimizing Performance in Artificial Lift Using Machine Learning Models”, SPE Western Regional Meeting held in Garden Grove, California, USA, 22-27 April 2018.

[5]. Mohanand Abdelaziz, Rafael Lastra and J.J.Xiao, Saudi Aramco (2017), SPE-1885130-MS “ESP Data Analytics Predicting Failures for Improved Production Performance”, Abu Dhabi International Petroleum Exhibitions and Conference held in Abu Dhabi, UAE, 13-16 November 2017.

[6]. Nico Jansen Van Rensburg, Lisa Kamin, and Skip (Robert) Davis, Siemens AG, (2019),” *Using Machine Learning-Based Predictive Models to Enable Preventative Maintenance and Prevent ESP Downtime*”, Abu Dhabi International Petroleum Exhibition & Conference held in Abu Dhabi, UAE, 11-14 November 2019.

[7]. Om Prasad Patri, Anand V. Panangadan, Charalampos Chelmiss, University of Southern California, Randall G.McKee, Chevron U.S.A. Inc., and Viktor K. Prasanna, University of Southern California (2014), “ *SPE-170680-MS Predicting Failures from Oilfield Sensor Data using Time Series Shapelets*”, Amsterdam, The Netherlands, 27–29 October 2014.

[8]. Ping Lu, Haitao Liu, Christopher Serratella, and Xiaozhi Wang, American Bureau of Shipping (2017), “*Assessment of Data-Driven, Machine Learning Techniques for Machinery Prognostics of Offshore Assets*”, Offshore Technology Conference held in Houston, Texas, USA, 1–4 May 2017.

[9]. Rinat Alfredovich Khabibullin and Arturas Rimo Shabonas, Gubkin Oil and Gas University Moscow; Nikolay Sergeevich Gurbatov, Deeplight; Alexey Vasilievich Timonov, Expert (2020), “SPE-201881-MS Prediction of ESPs Failure Using ML at Western Siberia Oilfields with Large Number of Wells”, SPE Russian Petroleum Technology Conference originally scheduled to be held in Moscow, Russia, 12-14 October 2020.

[10]. Pham Son Tung, Vo Phi Son and Nguyen Dac Nhat (2021) “*Effective Electrical Submersible Pump Management Using Machine Learning*. *Open Journal of Civil Engineering*”, 11, 70-80.

[11]. Ta Quoc Dung, Le The Ha, Pham Duy Khang (2019), “*Ứng dụng mạng nơron nhân tạo (Artificial Neural network) trong dự báo độ*”, Tạp chí Dầu Khí Số 7 - 2019, trang 18 – 27, ISSN-0866-854X

[12]. Tianqi Chen, Carlos Guestrin (University of Washington) (2016), “*XGBoost: A Scalable Tree Boosting System*”.