

TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT
BỘ MÔN THIẾT BỊ DẦU KHÍ VÀ CÔNG TRÌNH

BÁO CÁO HỌC THUẬT

Chuyên đề: *“Tìm hiểu mô hình mô phỏng nhiệt động lực học và/hoặc thủy động lực học một hệ thống công nghệ chính giàn xử lý khí condensate mỏ Hải Thạch - Mộc Tinh (HT-MT) theo thời gian thực”*

Thực hiện: Nguyễn Thanh Tuấn

Đơn vị công tác: Bộ môn Thiết bị Dầu khí và Công trình

Hà Nội, tháng 6 năm 2026

Giới thiệu chung

Cụm mỏ khí Hải Thạch - Mộc Tinh (HT-MT) bao gồm 1 giàn đầu giếng tại Lô 05-3 (mỏ Mộc Tinh) và 1 giàn đầu giếng cùng 1 giàn xử lý trung tâm tại Lô 05-2 (mỏ Hải Thạch) do Công ty Điều hành Dầu khí Biển Đông xây dựng và đưa vào vận hành khai thác từ năm 2012. Giàn được thiết kế với công suất cực đại đạt 10 triệu m³ khí/ ngày đêm và 25.000 thùng dầu/ ngày đêm.

Để tối ưu hóa hiệu quả khai thác của các mỏ, việc tối ưu hóa điều kiện vận hành bằng cách giảm áp suất đầu vào của hệ thống công nghệ là một trong những phương pháp được sử dụng. Tuy nhiên, thay đổi điều kiện vận hành đầu vào ảnh hưởng trực tiếp lên hệ thống công nghệ, đặc biệt quan trọng là hệ thống máy nén để xuất khí qua đường ống Nam Côn Sơn. Khi áp suất đầu vào giảm, nhiệt độ đầu ra của máy nén có thể tăng vượt quá ngưỡng làm việc an toàn dẫn đến hư hỏng các đệm làm kín, cũng như dầu nhớt bôi trơn sử dụng cho máy nén. Để tính toán sự ảnh hưởng của thay đổi điều kiện vận hành của máy nén lên nhiệt độ đầu ra của mỗi cấp, các phần mềm thương mại để tính toán mô phỏng nhiệt động học như Hysys, ProII thường được sử dụng để dự đoán ảnh hưởng của việc thay đổi điều kiện vận hành lên hệ thống công nghệ, từ đó người vận hành có thể mô phỏng và lựa chọn các điều kiện làm việc tối ưu, đảm bảo an toàn cho hệ thống công nghệ. Tuy nhiên, chi phí bản quyền phần mềm thương mại cao, hơn 30.000 USD/ năm.

Bên cạnh đó, một công cụ hiện đại và đã chứng minh có thể sử dụng để dự đoán các thông số vận hành dựa trên cơ sở dữ liệu vận hành trong quá khứ đó là ứng dụng các thuật toán học máy. Nhiều nghiên cứu đã tìm cách đưa ra những dự đoán chính xác cho hiệu suất hoạt động của máy nén, nhằm mục đích nâng cao hiệu quả của công tác tối ưu hóa vận hành và mô phỏng hệ thống công nghệ. Nhiều phương pháp đã được sử dụng như mạng thần kinh nhân tạo và các thuật toán dựa trên lý thuyết xác suất thống kê khác.

Thuật toán học máy đã chứng minh có nhiều lợi thế như thuật toán đơn giản, mô hình có thể vận hành liên tục, không cần hiệu chỉnh khi thay đổi điều kiện vận hành và các mô hình này được xem như là một giải pháp thay thế để so sánh với các phần mềm mô phỏng thương mại như Hysys.

1. Tổng quan về nhiệm vụ dự báo nhiệt độ đầu ra của cấp 1 và cấp 2 của máy nén khí EGC tại mỏ HT - MT

Máy nén khí ly tâm là thiết bị được sử dụng rộng rãi trong nhiều lĩnh vực kỹ thuật ứng dụng khác nhau. Việc dự đoán hiệu năng cũng như điều kiện làm việc của máy nén là một trong những công việc chính trong quá trình thiết kế, tối ưu hóa vận hành và mô phỏng hệ thống.

Để dự đoán các thông số vận hành của máy nén, người ta thường sử dụng các phương pháp tính toán truyền thống dựa trên các phương trình khí lý tưởng [1,2,3] $PV^k = \text{Hằng số}$, trong đó k là hệ số tỷ lệ giữa nhiệt dung riêng đẳng áp (C_p) và nhiệt dung riêng đẳng tích (C_v). Các công cụ tính toán thủ công thường sử dụng các phương trình nhiệt động dành cho khí lý tưởng bởi tính đơn giản của các công thức tính toán. Tuy nhiên, kết quả tính toán thường có độ sai lệch lớn so với số liệu thực tế.

Để tăng độ chính xác của các công thức tính toán, tác giả John M.Schultz [4] đã đưa ra các phương trình tính toán cho khí thực (real gas). Bên cạnh hệ số nén của khí Z , tác giả đã đưa thêm vào các hàm số X và Y . Các hàm số bổ sung này có thể được khái quát hóa và vẽ biểu đồ cho các tính chất của chất khí ở các trạng thái tương ứng. Tuy nhiên, các phương trình tính toán của tác giả phức tạp, thường được đưa vào cơ sở dữ liệu tính toán nhiệt động học cho các phần mềm mô phỏng thương mại hiện nay như Hysys, Pro II.

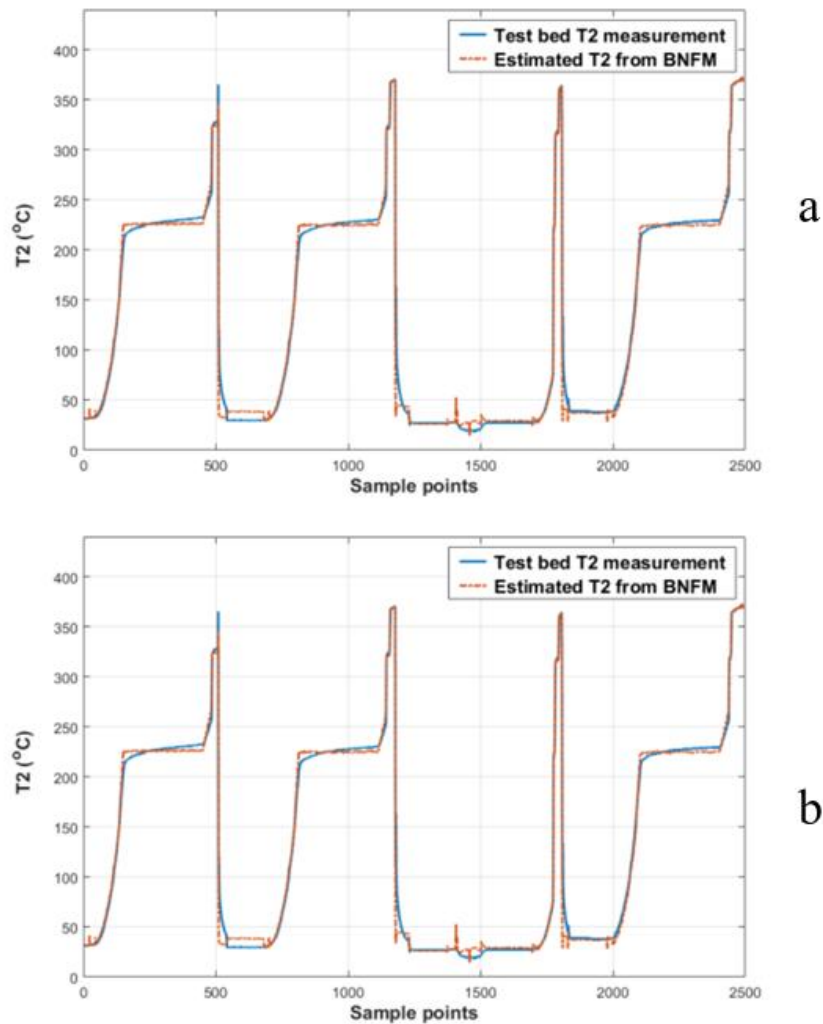
Tác giả Wei Chunzhi [5] dựa trên các kết quả nghiên cứu của tác giả John M.Schultz [4] để dự đoán nhiệt độ đầu ra của máy nén và kết quả cho thấy nhiệt độ đầu ra của máy nén tính theo các phương trình trạng thái của khí thực có độ chênh lệch khoảng 0,77%, trong khi giá trị nhiệt độ đầu ra của máy nén tính theo phương trình trạng thái khí lý tưởng có độ chênh lệch hơn 8% so với thông số thử nghiệm thực tế do nhà thầu (Vendor) cung cấp.

Việc tính toán các thông số vận hành của máy nén khí điều kiện vận hành thay đổi hiện nay tại công ty BDPOC đang sử dụng phần mềm Hysys để tính toán. Kết quả thực tế cho thấy kết quả tính toán có độ chính xác cao. Tuy nhiên, cần phải lấy mẫu định kỳ để cập nhật sự thay đổi tính chất của lưu chất và chi phí cho việc thuê phần mềm tương đối cao.

Trong những năm gần đây, nhiều nghiên cứu đã tìm cách đưa ra những dự đoán chính xác cho hiệu suất hoạt động của máy nén, nhằm mục đích nâng cao hiệu quả của công tác thiết kế, tối ưu hóa và mô phỏng hệ thống. Nhiều phương pháp đã được sử dụng như nội suy, điều chỉnh đường cong (curve fitting), mạng thần kinh (NN-Neural Network) và các thuật toán dựa trên lý thuyết xác suất thống kê khác. Tuy nhiên, các phương pháp này thường cần một lượng lớn dữ liệu và việc thu thập dữ liệu có thể tốn đáng kể tài nguyên máy tính hoặc tài nguyên thử nghiệm.

Tác giả Yu Zhang cùng cộng sự [6] đã nghiên cứu dự đoán nhiệt độ đầu ra của máy nén bằng cách sử dụng một mô hình hướng dữ liệu (data-driven model) và mô hình này được xây dựng dựa trên nền tảng cơ sở dữ liệu thử nghiệm (test bed) động cơ của nhà thầu. Tác giả đề xuất sử dụng mô hình hóa mờ thần kinh Bayes (Bayesian Neuro-Fuzzy Modelling - BNFM), kết hợp hệ thống suy luận mờ thần kinh thích ứng (Adaptive Neuro-Fuzzy Inference System - ANFIS) và kỹ thuật mô hình hỗn hợp Bayesian Gaussian (Variational Bayesian Gaussian Mixture Model - VBGMM). Mô hình máy nén dựa trên các cơ sở dữ liệu được xây dựng bằng ANFIS và VBGMM đã được áp dụng trong giai đoạn xây dựng thiết bị để lựa chọn một cách tự động các dữ liệu đầu vào trong hệ thống mờ (fuzzy system). Phương pháp tiếp cận BNFM được thiết lập thông qua các thử nghiệm thực tế đối với tuốc bin khí công suất dưới 15MW và kết quả từ mô hình được xây dựng dựa trên các cơ sở dữ liệu thử nghiệm được chứng minh là có độ tin cậy cao trong việc dự đoán nhiệt độ đầu ra của máy nén trên tuốc bin khí trong quá trình vận hành.

Hình 1.1a và Hình 1.1b mô tả kết quả dự đoán nhiệt độ đầu ra của máy nén (T2) so với nền tảng cơ sở dữ liệu thử nghiệm (test bed) và giá trị thực tế thực nghiệm tại hiện trường.



Hình 1.1a-b: a) Kết quả dự đoán nhiệt độ đầu ra của máy nén (T2) bằng BNFM so với cơ sở dữ liệu thử nghiệm của nhà thầu (test-bed).
b) Kết quả dự đoán nhiệt độ đầu ra của máy nén (T2) bằng BNFM so với số liệu thực tế tại hiện trường.

Kết quả cho thấy, giá trị ước tính của nhiệt độ đầu ra máy nén bằng phương pháp sử dụng BNFM đưa ra số liệu có độ tin cậy cao, do vậy, kết quả dự đoán từ mô hình này có thể được sử dụng để thay thế trong trường hợp các cảm biến đo đặc hoạt động thiếu chính xác hay bị hư hỏng không thể sử dụng

được. Bên cạnh đó, kết quả tính toán bằng mô hình cũng có thể được sử dụng như một cảm biến ảo để cung cấp thêm bằng chứng cho việc giám sát tình trạng của máy nén bên cạnh các cảm biến vật lý hiện có.

Tác giả Jingzhou Fei cùng cộng sự [7] đã sử dụng mạng thần kinh nhân tạo mới tích hợp mạng thần kinh truyền ngược nguồn cấp dữ liệu với chức năng chính của thuật toán Gaussian để dự đoán hiệu suất hoạt động của máy nén. Kết quả nghiên cứu chỉ ra rằng mạng thần kinh được đề xuất có hiệu suất dự đoán vượt trội so với mạng thần kinh truyền ngược nguồn cấp dữ liệu chuyển tiếp và máy vector hỗ trợ hiện có, đặc biệt là đối với phép ngoại suy với các mẫu nhỏ. Hơn nữa, nghiên cứu này có thể được sử dụng trong việc tinh chỉnh mô hình dựa trên hiệu suất hiện có để cải thiện phân tích mô phỏng, giám sát tình trạng và chẩn đoán lỗi của máy nén tuabin khí.

Tác giả Hongsheng Jiang cùng cộng sự [8] tiến hành xây dựng mô hình tỷ lệ áp suất và hiệu suất đẳng entropi của máy nén bằng cách sử dụng một số lượng mẫu dữ liệu hạn chế trong khi vẫn đảm bảo được độ chính xác cao. Đầu tiên, dữ liệu mẫu được tạo ra bằng phương pháp mô phỏng Vista CCD. Tiếp đó, lưu lượng đã hiệu chỉnh và tốc độ quay đã hiệu chỉnh được sử dụng làm các biến độc lập và các biểu thức hồi quy với giá trị trung bình của tỷ lệ áp suất và hiệu suất đẳng entropi được rút ra và đơn giản hóa thông qua phân tích nhiệt động lực học và phân tích tổn thất của máy nén ly tâm. Các mô hình kriging dựa trên mô hình mạng nơ ron và đa thức bậc hai cũng đã được xây dựng. Kết quả cho thấy rằng, khi tính toán dự đoán bên trong phạm vi giới hạn của dữ liệu, mô hình dựa trên phân tích tổn thất và mô hình kriging tạo ra dự đoán có độ chính xác cao hơn ngay cả khi có một cơ sở dữ liệu nhỏ và kết quả dự đoán là ổn định, trong khi mô hình mạng nơ ron chỉ cung cấp kết quả tốt hơn trong một tập dữ liệu lớn hơn. Đối với dự đoán bên ngoài phạm vi giới hạn của dữ liệu, mô hình dựa trên phân tích tổn thất có thể cung cấp kết quả tương đối

chính xác. Bên cạnh đó, mất ít thời gian hơn để đào tạo và sử dụng mô hình dựa trên phân tích tổn thất (loss-analysis-based model) so với các mô hình khác.

Như vậy, việc áp dụng các công cụ hiện đại như trí tuệ nhân tạo và học máy đã cho thấy có nhiều lợi thế trong lĩnh vực xây dựng mô tiên đoán hiệu suất hoạt động của máy nén ly tâm, cũng như dự đoán sự thay đổi của các thông số vận hành một cách chính xác. Do vậy, việc ứng dụng các công cụ trí tuệ nhân tạo và học máy vào việc dự báo nhiệt độ đầu ra của cấp 1 và cấp 2 của máy nén khí EGC tại mỏ Mộc Tinh – Hải Thạch là phù hợp nhằm tối ưu hóa hoạt động sản xuất, cũng như tiết kiệm được các chi phí thuê mướn phần mềm chuyên dụng.

2. Các phương pháp đang được áp dụng để dự báo nhiệt độ đầu ra của cấp 1 và cấp 2 của máy nén khí EGC tại mỏ HT - MT

Khí khai thác từ giàn Hải Thạch – Mộc tinh, sau khi đã được xử lý đạt các tiêu chuẩn của khí thương phẩm, được xuất bán qua đường ống Nam Côn Sơn bằng hai dây chuyền máy nén ly tâm hai cấp.

Trong nghiên cứu này, nhóm tác giả chọn ra 3 trường hợp đại diện để thực hiện dự đoán nhiệt độ đầu ra của mỗi cấp, trong đó trường hợp 1 và 2 tương ứng với điều kiện vận hành với áp suất đầu vào của cấp 1 đạt giá trị cực tiểu và cực đại và trường hợp 3 là điểm làm việc mà công ty dự kiến sẽ vận hành trong thời gian tới. Những số liệu của trường hợp 3 là do nhà thầu (Vendor) máy nén cung cấp Bảng 2.1.

Bảng 2.1: Bảng tổng hợp các thông số vận hành thực tế của máy nén

Trường hợp	Đầu vào cấp 1		Đầu ra cấp 1		Đầu vào cấp 2		Đầu ra cấp 2	
	TT-	PT-	TT-	PT-	TT-	PT-	TT-	PT-
	2620 (C)	2620 (barg)	2626 (C)	2626 (barg)	2625 (C)	2625 (barg)	2622 (C)	2622 (barg)
1	27,1	46,4	59,5	67,6	31,2	67,1	87,7	126,2
2	26,7	54,8	53,7	75,6	30,6	74,8	76,8	127,9
3*	41,6	30,0	95,9	54,9	50,0	54,1	151,8	158,0

*: Số liệu do nhà thầu máy nén cung cấp.

2.1. Phương pháp tính toán truyền thống

Thông số nhiệt độ đầu ra của máy nén là một thông số vận hành quan trọng cần được theo dõi trong hoạt động của máy nén khí. Nhiệt độ đầu ra mong muốn của máy nén thông thường được duy trì từ mức 121°C đến 135°C để đảm bảo tuổi thọ của các lớp đệm cũng như tránh dầu nhớt bôi trơn bị hư hỏng do tiếp xúc với nhiệt độ cao cũng như đảm bảo hiệu suất hoạt động của máy nén [1]. Ở nhiệt độ cao vượt quá 149°C, ngoài việc dầu nhớt bôi trơn bị hư

hồng, nếu có sự hiện diện của oxi, hiện tượng dầu bôi trơn tự bốc cháy có thể xảy ra. Trong mọi trường hợp, nhiệt độ đầu ra của máy nén không được phép vượt quá 177°C [2].

Nhiệt độ đầu ra của máy nén có thể được giảm xuống bằng cách:

Giảm nhiệt độ khí đầu vào

Giảm tỷ số nén (P_d/P_s)

Để tính nhiệt độ đầu ra của máy nén, ta có thể sử dụng công thức (2.1) trong tài liệu Surface Production Operation [2]. Công thức tính nhiệt độ đầu ra của máy nén chi tiết như sau:

$$T_d = T_s \left(\frac{P_d}{P_s} \right)^{\frac{k-1}{k} \times \frac{1}{\eta}} \quad (2.1)$$

Trong đó:

T_d : Nhiệt độ đầu ra của mỗi cấp, °R Chuyển đổi °C=(R-491,67) x (5/9)

T_s : Nhiệt độ đầu vào của mỗi cấp, °R

P_d : Áp suất đầu ra của mỗi cấp, bara

P_s : Áp suất đầu vào của mỗi cấp, bara

k : Tỷ số nhiệt dung riêng, $\frac{C_p}{C_v}$, đối với khí tự nhiên, $k=1,3$

C_p : Nhiệt dung riêng đẳng áp

C_v : Nhiệt dung riêng đẳng tích.

η : Hiệu suất đẳng nhiệt, 0,8 cho máy nén ly tâm.

Kết quả so sánh nhiệt độ tính toán nhiệt độ đầu ra cấp 1 và cấp 2 theo công thức (2.1) với giá trị thực tế, được tính toán trong cùng trong cùng một điều kiện vận hành nhiệt độ áp, suất đầu vào và áp suất đầu ra của mỗi cấp, được tổng hợp trong bảng sau:

Bảng 2.2: So sánh nhiệt độ thực tế và nhiệt độ tính toán theo công thức (2.1) và nhiệt độ thực

Trường hợp	Nhiệt độ đầu ra Cấp 1 (C)			Nhiệt độ đầu ra Cấp 2 (C)		
	Thực tế	Tính toán	Sai số (%)	Thực tế	Tính toán	Sai số (%)
1	59,5	60,9	2,3	87,7	91,2	4,1
2	53,7	55,3	3,1	76,8	80,9	5,3
3	95,9	99,9	4,3	151,8	165,5	9,0

Kết quả so sánh tại Bảng 2.2 cho thấy nhiệt độ đầu ra của máy nén tính toán bằng công thức 2.1 có kết quả chính xác không cao, sai số kết quả tính toán so với giá trị thực tế từ 2% đến 9%. Nguyên nhân được xác định do công thức (2.1) được sử dụng dựa trên tính chất của khí lý tưởng và nhiệt độ đầu ra tại mỗi cấp của máy nén phụ thuộc vào độ chính xác của hệ số k , cũng như hiệu suất đẳng nhiệt η của máy nén. Do đó, nhiệt độ được tính toán từ phương pháp trên có thể khác biệt đáng kể so với điều kiện thực tế.

2.2. Phương pháp tính toán bằng các phần mềm mô phỏng

Bên cạnh đó, một phương pháp thuận tiện hơn là dùng các phần mềm mô phỏng như Hysys để dự đoán thông số vận hành của máy nén. Phần mềm Hysys sử dụng các mô hình nhiệt động có chính xác cao, dựa trên các phương trình tính toán đa biến với các vòng lặp tính toán. Các thông số cài đặt hoặc đường đặc tính thực tế của máy nén được cung cấp bởi nhà thầu được đưa vào mô hình tính toán nhằm tăng độ chính xác của các phương trình nhiệt động học. Phương trình tính toán hiệu suất (efficiency) và chiều cao cột áp (head) của máy nén

được sử dụng trong phần mềm mô phỏng Hysys được liệt kê chi tiết trong tài liệu kỹ thuật của phần mềm [9], và được tóm tắt như sau:

Các ký hiệu sau được sử dụng trong các công thức tính toán:

Ideal : Đẳng entropi (Hiệu suất = 100%)

Actual: Thực tế

H : Enthalpy khối lượng (kJ/kg)

Out : Dòng ra

In : Dòng vào

P : Áp suất (kPa)

MW : Phân tử khối (Molecular Weight – MW)

Z : Hệ số nén

ρ : Tỷ trọng (kg/m³)

f : Hệ số cột áp đa biến (Polytropic Head Factor)

n : Hệ số mũ đa biến (Polytropic Exponent)

k : Hệ số mũ đoạn nhiệt (Isentropic Exponent)

Hiệu suất đoạn nhiệt (adiabatic) và hiệu suất đa biến (polytropic) của máy nén được tính theo công thức sau:

$$AdiabaticEff = \frac{Work\ Required_{(ideal)}}{Work\ Required_{(actual)}} = \frac{(H_{out} - H_{in})_{ideal}}{(H_{out} - H_{in})_{actual}} \quad (2.2)$$

PolytropicEff

$$= \frac{\left[\left(\frac{P_{out}}{P_{in}} \right)^{\left(\frac{n-1}{n} \right)} - 1 \right] \times \left[\left(\frac{n}{(n-1)} \right) \times \left(\frac{k-1}{k} \right) \right]}{\left[\left(\frac{P_{out}}{P_{in}} \right)^{\left(\frac{k-1}{k} \right)} - 1 \right]} \times AdiabaticEff \quad (2.3)$$

Trong đó hệ số mũ đa biến (n) và hệ số mũ đoạn nhiệt (k) được tính theo các công thức:

$$n = \frac{\log(P_{out}/P_{in})}{\log(\rho_{out,actual}/\rho_{in})} \quad (2.4)$$

Và:

$$k = \frac{\log(P_{out}/P_{in})}{\log(\rho_{out,ideal}/\rho_{in})} \quad (2.5)$$

Chiều cao cột áp đoạn nhiệt (Adiabatic Head) của máy nén được tính toán theo công thức sau:

$$AdiabaticHead = \left[\frac{WorkRequired_{(actual)}}{MassFlowRate} \right] \times AdiabaticEff \times \frac{1}{(g/g_c)} \quad (2.6)$$

Chiều cao cột áp đa biến (Polytropic Head) được tính toán dựa trên phương pháp của ASME:

PolytropicHead

$$= f \times \left(\frac{n}{n-1} \right) \times \left[\left(\frac{P_{out}}{\rho_{out,actual}} \right) - \left(\frac{P_{in}}{\rho_{in}} \right) \right] \times \frac{1}{(g/g_c)} \quad (2.7)$$

Trong đó:

$$f = \frac{H_{out,ideal} - H_{in}}{\left(\frac{k}{k-1} \right) \times \left[\left(\frac{P_{out}}{\rho_{out,ideal}} \right) - \left(\frac{P_{in}}{\rho_{in}} \right) \right]} \quad (2.1)$$

$$n = \frac{\log(P_{out}/P_{in})}{\log(\rho_{out,actual}/\rho_{in})} \quad (2.9)$$

Để có được kết quả mô phỏng chính xác, trước hết cần phải có kết quả phân tích thành phần khí đầu vào cũng như tính chất vật lý của các phân đoạn có khối lượng phân tử lớn có trong khí. Thành phần cấu tử và tính chất vật lý của các phân đoạn nặng sử dụng cho mô hình mô phỏng được tính toán bằng giá trị trung bình của từng cấu tử được sản xuất trong tháng 6 năm 2022.

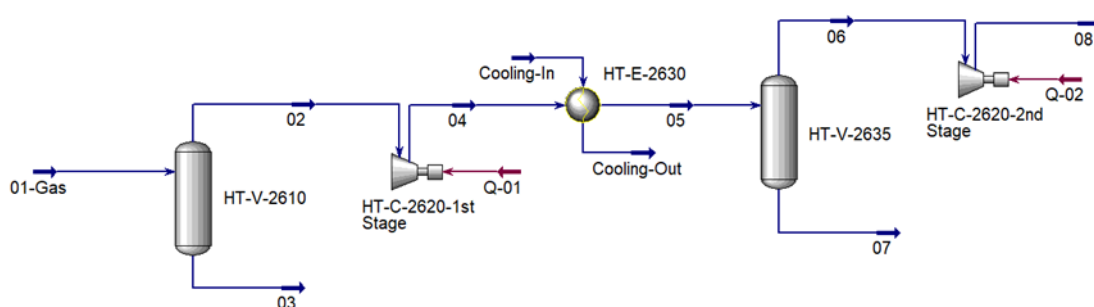
Bảng 2.3: Thành phần cấu tử

Cấu tử	% khối lượng
Methane	71,6898
Ethane	7,0413
Propane	4,9538
i-Butane	1,4133
n-Butane	1,8114
i-Pentane	0,7628
n-Pentane	0,6321
Nitrogen	0,0992
Carbon Dioxide	10,0762
Phân đoạn C6	0,5868
Phân đoạn C7	0,7830
Phân đoạn C8	0,1504
Phân đoạn C9+	0,0000

Bảng 2.4: Tính chất vật lý của phân đoạn nặng

STT	Phân đoạn	Nhiệt độ sôi [degC]	Phân tử khối
1	Phân đoạn C6	64,52	84,45
2	Phân đoạn C7	88,89	90,72
3	Phân đoạn C8	117,66	100,76
4	Phân đoạn C9+	150,82	128,26

Sơ đồ sau được xây dựng trên phần mềm Hysys để mô phỏng máy nén xuất khí EGC. Máy nén gồm có 2 cấp: cấp 1 và cấp 2. Giữa 2 cấp có thiết bị làm lạnh và tách lỏng trung gian. Sơ đồ mô phỏng máy nén như sau:



Hình 2.1: Sơ đồ mô phỏng công nghệ bằng phần mềm Hysys

Kết quả tính toán nhiệt độ đầu ra của máy nén cấp 1 và cấp 2 so với giá trị nhiệt độ thực tế vận hành, được tính toán trong cùng trong cùng một điều kiện vận hành nhiệt độ áp, suất đầu vào và áp suất đầu ra của mỗi cấp, được tổng hợp trong bảng sau:

Bảng 2.5: sánh nhiệt độ thực tế và nhiệt độ tính toán theo phần mềm mô phỏng Hysys

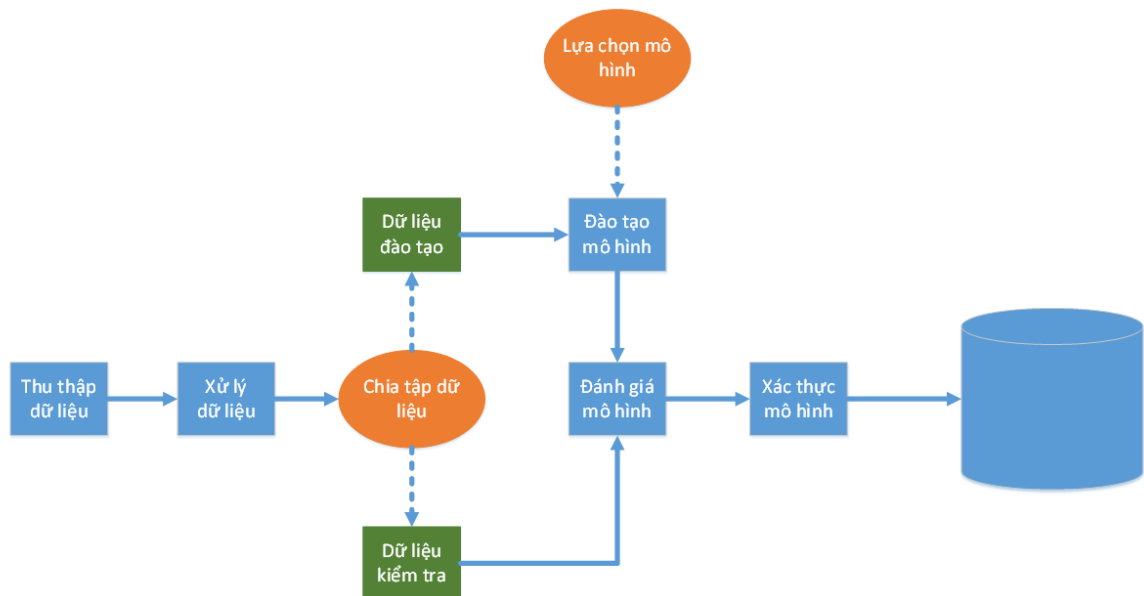
Trường hợp	Nhiệt độ đầu ra Cấp 1 (C)			Nhiệt độ đầu ra Cấp 2 (C)		
	Thực tế	Hysys	Sai số (%)	Thực tế	Hysys	Sai số (%)
1	59,5	60,4	1,5	87,7	88,5	0,9
2	53,7	55,2	2,8	76,8	78,5	2,2
3	95,9	97,0	1,2	151,8	154,5	1,8

Kết quả tổng hợp từ Bảng 2.5 cho thấy kết quả tính toán nhiệt độ đầu ra của máy nén bằng phần mềm mô phỏng Hysys có độ chính xác cao. Sai số của kết quả tính toán bằng phần mềm mô phỏng so với giá trị thực tế khoảng từ 0,9% đến 2,8%. Độ chính xác của kết quả tính toán bằng phần mềm Hysys phụ thuộc vào thành phần và tính chất của các cấu tử, do đó cần lấy mẫu định kỳ để phân tích kiểm tra sự thay đổi thành phần, cũng như tính chất của dòng lưu chất. Bên cạnh đó, chi phí thuê mướn phần mềm tương đối cao, khoảng 30 ngàn đô la mỹ một năm.

Từ những vấn đề nêu trên, nhóm tác giả đề xuất đề tài sử dụng thuật toán học máy (Machine learning) để dự đoán nhiệt độ đầu ra của máy nén cấp 1 và máy nén cấp 2 khi biết thông số nhiệt độ và áp suất đầu vào của máy nén cấp 1, cấp 2 và áp suất đầu ra của cấp 1 và cấp 2 dựa trên số liệu vận hành trong lịch sử để tăng độ chính xác cũng như giảm được chi phí thuê phần mềm.

3. Phương pháp nghiên cứu sử dụng AI dự báo nhiệt độ đầu ra tại mỗi cấp của máy nén

3.1. Các bước áp dụng bài toán ML



Hình 3.1: Các bước thực hiện bài toán ML

Tiến trình của ML được tuân theo các bước như sau:

- Thu thập dữ liệu: dữ liệu được thu thập từ nhiều nguồn khác nhau như cơ sở dữ liệu, cảm biến... cũng như thời gian thực hoặc từ lịch sử.
- Xử lý dữ liệu: Chất lượng của dữ liệu ảnh hưởng trực tiếp đến độ chính xác của hệ thống. Tuy nhiên, dữ liệu thường có điểm sai lệch, trùng lặp, thiếu sót không thể áp dụng trực tiếp vào hệ thống. Do đó, dữ liệu phải được xử lý bằng một hoặc nhiều phương pháp khác nhau như biến đổi, loại bỏ, chuẩn hóa, điều chỉnh trước khi cho mô hình học.
- Chọn mô hình: Đối với từng loại dữ liệu khác nhau cũng như mục tiêu đặt ra, ta phải lựa chọn các mô hình phù hợp để có hiệu quả tốt nhất. Ví dụ, đối với dữ liệu gắn nhãn và bài toán đặt ra là phân loại, thuật toán được sử dụng sẽ là thuật toán phân loại; đối với dữ liệu gắn nhãn và bài toán đặt ra là hồi quy, thuật toán được sử dụng sẽ là thuật toán hồi quy; đối với dữ liệu không được

gắn nhãn và bài toán đặt ra là phân loại, thuật toán được sử dụng sẽ là thuật toán phân cụm.

- Chia tập dữ liệu: Sau khi xử lý, dữ liệu thường được chia ra làm ba tập dữ liệu để đào tạo và đánh giá mô hình ML nhằm tăng hiệu quả của mô hình. Ba tập dữ liệu bao gồm: dữ liệu huấn luyện (Training Dataset) giúp mô hình hiểu các xử lý thông tin, mối quan hệ giữa dữ liệu và các kết quả liên quan của nó; dữ liệu xác thực (Validation Dataset) giúp đánh giá mô hình trong khi tinh chỉnh, tối ưu hóa nó, chẩn đoán xem mô hình có học tốt không; dữ liệu thử nghiệm (Test Dataset) giúp đánh giá mô hình cuối cùng một cách khách quan nhất. Sau đó mô hình sẽ được đào tạo, tinh chỉnh cũng như thử nghiệm dựa trên các tập dữ liệu đó.

- Đào tạo mô hình: Mô hình sẽ sử dụng tập dữ liệu huấn luyện để đào tạo nhằm giúp mô hình tìm hiểu về các tính năng và thông số. Trong giai đoạn này, thuật toán học tìm mối quan hệ giữa dữ liệu đầu vào và đầu ra và tạo ra một mô hình dựa vào.

- Đánh giá mô hình: Sau khi được đào tạo, mô hình sẽ được kiểm tra với dữ liệu thử nghiệm để đánh giá thông qua các thông số (evaluation metric) như sai số toàn phương trung bình (MSE), trung bình sai số trung bình tuyệt đối (MAE)... Trong giai đoạn này, nếu mô hình không hoạt động tốt, mô hình có thể xây dựng lại bằng cách sử dụng một hoặc nhiều tham số phức tạp được gọi là siêu tham số (hyperparameter). Để chọn một mô hình cuối cùng, chúng ta cần kiểm tra tác động của từng siêu thông số đối với hiệu suất của mô hình. Điều này xảy ra trong quá trình đào tạo mô hình.

- Xác thực mô hình: Sau khi hoàn thành đánh giá mô hình, ta sử dụng các dữ liệu thức tế để xác định mô hình có cần tiếp tục điều chỉnh hay có thể triển khai mô hình như một sản phẩm.

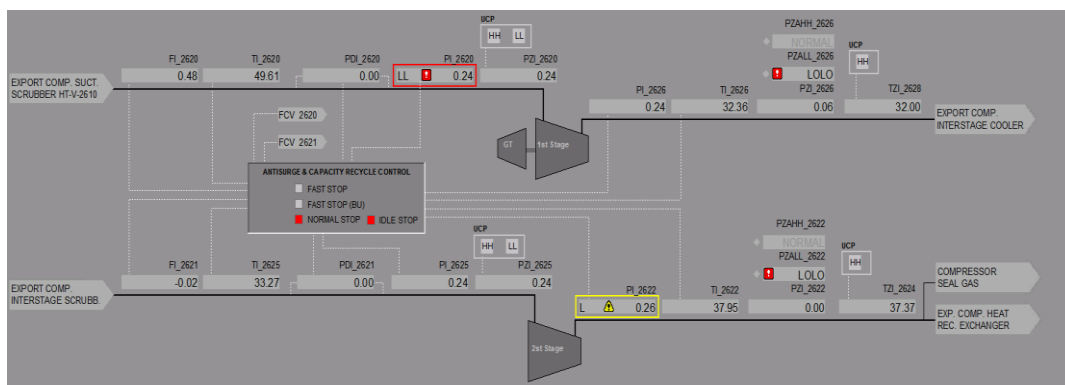
Đối với vấn đề được đặt ra trong chuyên đề này, nhóm tác giả đã áp dụng bài toán ML theo các bước sau:

- Dữ liệu sử dụng để đào tạo mô hình cho việc tính toán nhiệt độ đầu ra của máy nén được trích xuất từ phần mềm thu thập số liệu lịch sử (Historian) của máy nén khí cao áp export gas compressor (EGC).
- Dữ liệu được làm sạch và phân tích mối tương quan thông qua bản đồ nhiệt (heat map). Sau đó được phân chia thành 02 bộ dữ liệu: bộ dữ liệu đào tạo và bộ dữ liệu kiểm tra theo tỷ lệ lần lượt là 80% và 20%.
- Dựa trên thông tin phân tích dữ liệu, nhóm tác giả đã đánh giá và lựa chọn các mô hình ANN và LSTM để giải quyết vấn đề.
- Sau khi được đào tạo, các mô hình được đánh giá theo tiêu chí Căn Lỗi trung bình bình phương gốc bậc hai sai số toàn phương trung bình (RMSE) và hệ số R2 từ đó lựa chọn mô hình tốt nhất.

3.2. Thu thập dữ liệu

Hệ thống máy nén khí xuất bán khí qua đường ống Nam Côn Sơn bao gồm 2 dây chuyền máy nén khí hai cấp. Trong nghiên cứu này, nhóm tác giả sử dụng dữ liệu được trích xuất từ hệ thống thu thập số liệu lịch sử (Historian) của dây chuyền máy nén khí gas EGC số 1. Hệ thống này sử dụng máy nén khí hai cấp sử dụng động cơ turbine model TT130 của hãng Solar để dẫn động máy nén khí của hãng Dresser-rand với công suất thiết kế đạt 5 triệu m³ khí một ngày với áp suất nén hiện tại từ khoảng 50 bar lên áp suất 125 bar. Luồng khí đầu vào sau cấp nén thứ nhất từ 50 bar lên 70 bar, sau đó được làm lạnh qua bộ làm mát trung gian (interstage cooler) và tiếp tục vào máy nén cấp 2 để nén áp suất từ 70 bar tới 125 bar.

Bộ dữ liệu là thông số vận hành của các cảm biến được lắp trên hệ thống máy nén như Hình 3.2 dưới đây.



Hình 3.2: Sơ đồ công nghệ của hệ thống máy nén khí gas trên giàn PQP

Các thiết bị đo được thu thập trong bộ dữ liệu được thể hiện trong Bảng 3.1.

Bảng 3.1: Danh sách thiết bị đo sử dụng cho bộ dữ liệu

Tên thiết bị	Mô tả
TT-2620	Nhiệt độ khí vào máy nén cấp 1
PT-2620	Áp suất khí vào máy nén cấp 1
TT-2626	Nhiệt độ khí ra máy nén cấp 1
PT_2626	Áp suất khí ra máy nén cấp 1
TT-2625	Nhiệt độ khí vào máy nén cấp 2
PT-2625	Áp suất khí vào máy nén cấp 2
TT-2622	Nhiệt độ khí ra máy nén cấp 2
PT-2622	Áp suất khí ra máy nén cấp 2

Bộ dữ liệu là file excel bao gồm 17521 dòng và 08 cột ghi lại dữ liệu vận hành của máy nén khí cao áp trong một năm. Mỗi dòng tương ứng với một mốc thời gian còn các cột là giá trị của những cảm biến trên. Hình 3.3 là những thông tin chi tiết về bộ dữ liệu.

	TT_2620	PT_2620	PT_2622	PT_2626	TT_2625	PT_2625	TT_2626	TT_2622
0	29.450321	50.097248	125.836075	72.357353	36.298080	71.639389	60.333336	85.512825
1	29.530235	50.097248	125.771980	72.357353	36.394230	71.629776	60.381416	85.572647
2	29.613785	50.637859	125.900177	72.849884	36.361115	72.176788	60.210472	85.132477
3	29.613785	50.767139	125.776253	72.982368	36.445515	72.261192	60.144234	85.000008
4	29.586540	50.622906	125.831802	72.886208	36.445515	72.116959	60.144234	85.128212
...	...	...	...	...	...	...	...	...
17516	32.434406	49.017097	126.284805	71.517593	37.426281	70.844498	64.759621	89.358971
17517	32.474358	48.898502	126.028397	71.395790	37.500000	70.719490	64.855774	89.423088
17518	32.556095	48.850426	125.771980	71.299637	37.503204	70.671417	64.903854	89.423088
17519	32.541565	48.847221	125.643768	71.229126	37.522438	70.552818	64.926285	89.487183
17520	32.583332	48.847221	125.643768	71.293228	37.548080	70.575256	64.907059	89.487183

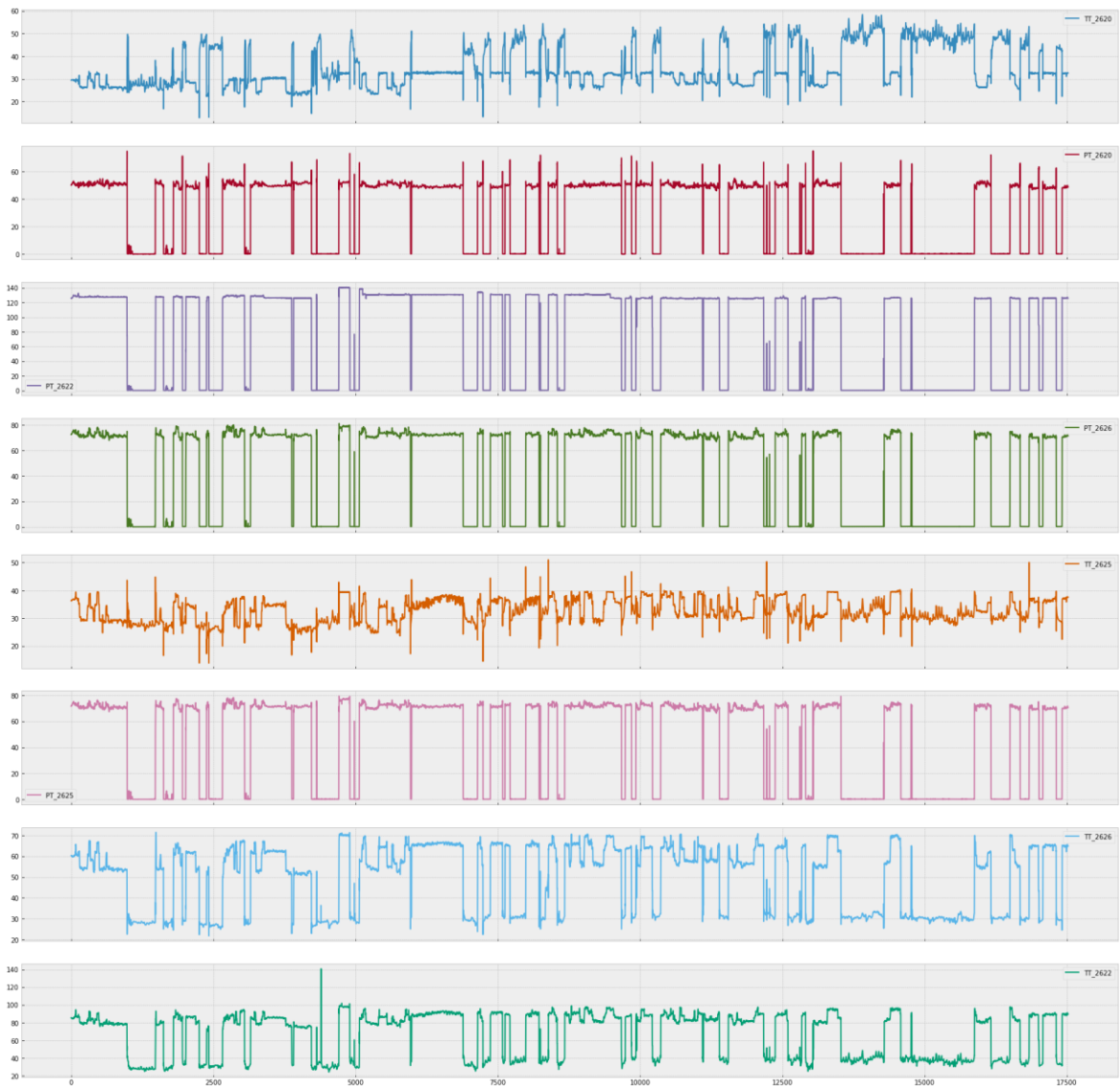
17521 rows × 8 columns

Hình 3.3: Bộ dữ liệu sử dụng nhằm dự báo nhiệt độ đầu ra của máy nén

3.3. Phân tích và tiền xử lý dữ liệu

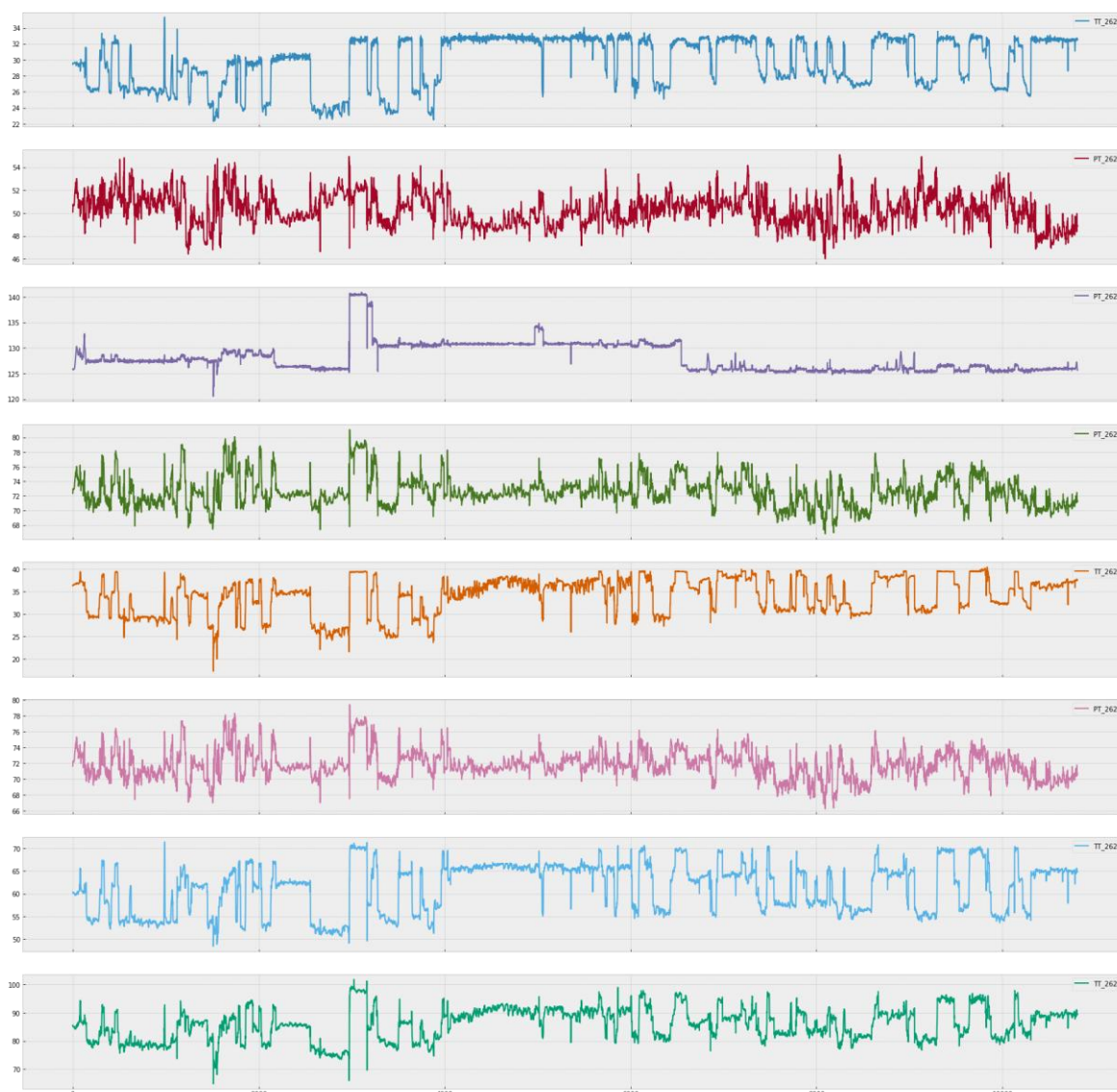
Làm sạch dữ liệu:

Biểu diễn bộ dữ liệu trên đồ thị Hình 3.4 quan sát thấy bộ dữ liệu bao gồm những đoạn đứt quãng có giá trị áp suất khí đầu ra cấp 2 của máy nén PT_2622 giảm từ giá trị vận hành xuống dưới 110 bar thể hiện khi đó máy từ chế độ vận hành chuyển về chế độ dừng, khí trong máy đã được xả hết.



Hình 3.4: Biểu diễn bộ dữ liệu trên đồ thị

Bằng việc loại bỏ những thời điểm máy dừng ra khỏi bộ dữ liệu dựa vào thời điểm áp suất khí đầu ra máy nén cấp 2 dưới 110 bar là áp suất tối thiểu để có thể xuất khí vào đường ống chuyển về bờ, bộ dữ liệu mới Hình 3.5 bao gồm 10808 mẫu chỉ bao gồm những thời điểm máy hoạt động bình thường.



Hình 3.5: Bộ dữ liệu sau khi được làm sạch

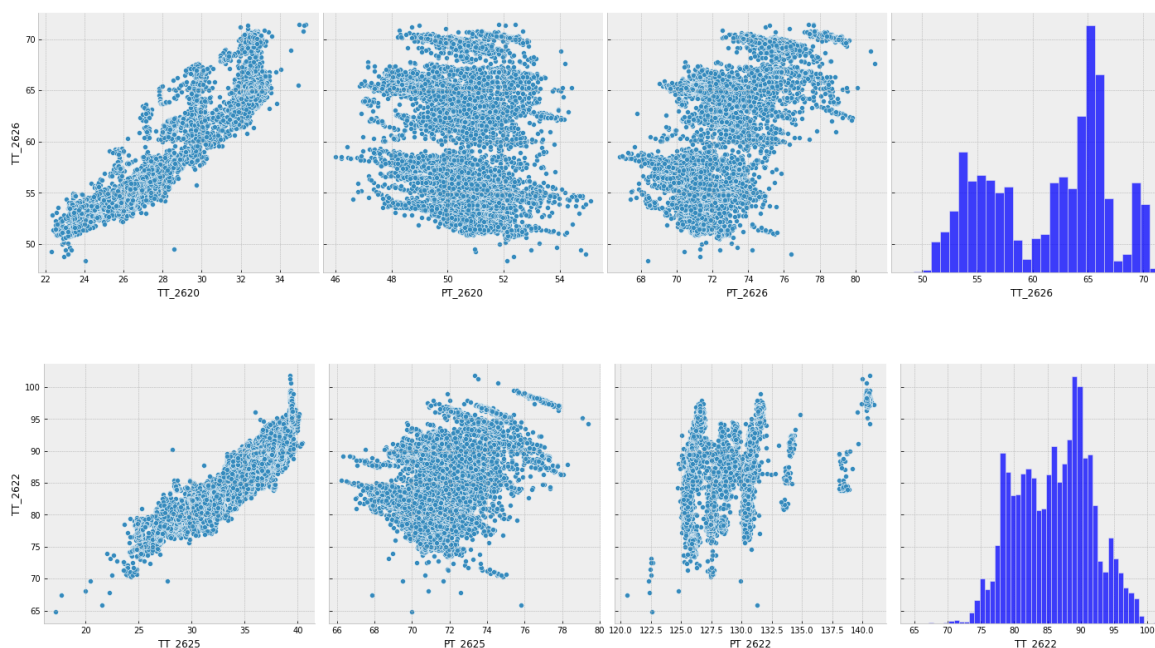
Phân tích thống kê dữ liệu:

Thống kê cho thấy điều kiện vận hành máy của bộ dữ liệu có áp suất khí đầu vào PT_2620 từ 46 tới 55 bar, áp suất khí đầu ra PT_2622 từ 120,5 tới 140,9 bar, nhiệt độ khí đầu ra cấp 1, TT_2626 từ 48,3°C tới 71,4°C, nhiệt độ khí đầu ra cấp 2, TT_2622 từ 64,8°C tới 101,8°C. Đây là khoảng làm việc mà mô hình sẽ dự đoán chính xác nhất.

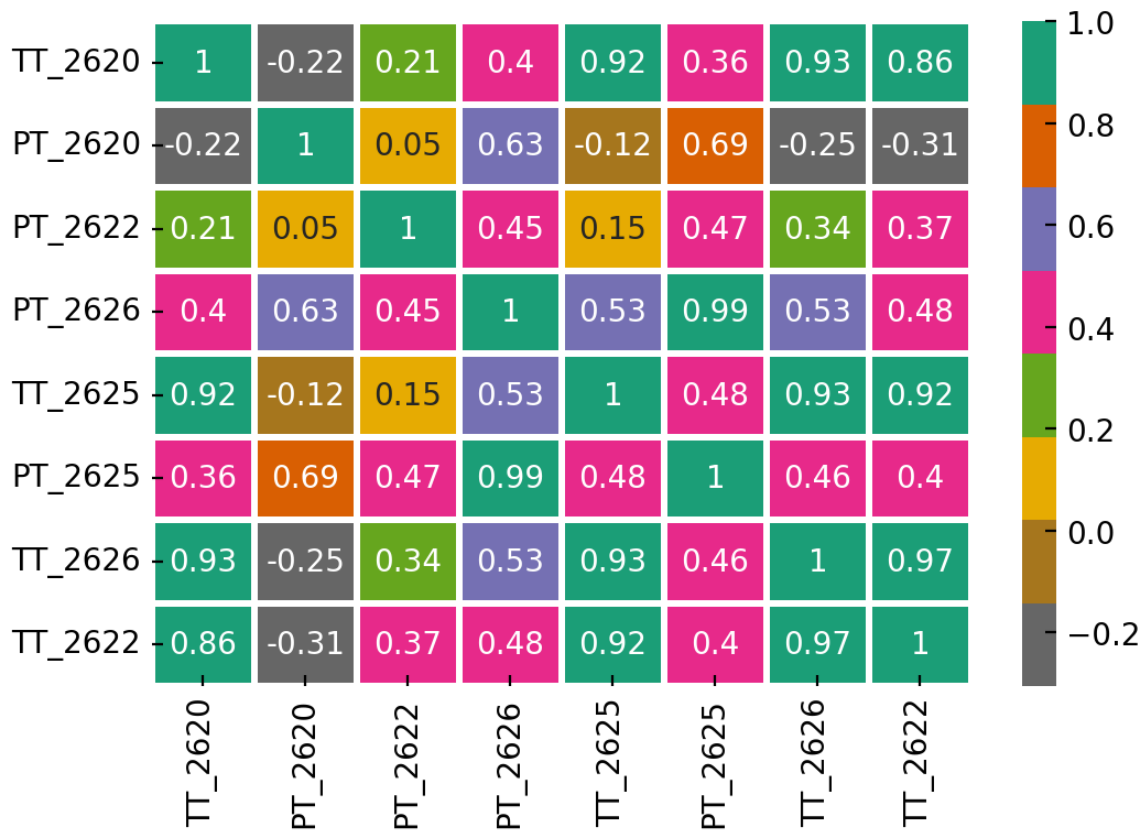
	TT_2620	PT_2620	PT_2622	PT_2626	TT_2625	PT_2625	TT_2626	TT_2622
count	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000	10808.000000
mean	29.802230	50.262590	128.124816	72.586570	34.044567	71.773163	61.381664	86.012226
std	3.084856	1.321416	2.867430	2.027851	4.119456	1.746670	5.398998	5.619714
min	22.308868	46.004211	120.532539	66.825180	17.201925	66.251442	48.371799	64.799156
25%	27.077860	49.284730	125.792988	71.299637	30.650640	70.674622	56.198715	81.474358
50%	30.766187	50.196609	127.366035	72.309273	35.000000	71.626572	62.884621	86.405983
75%	32.610577	51.148556	130.592590	73.466347	37.451923	72.709656	65.576920	90.000008
max	35.334938	55.094151	140.917603	81.062660	40.432693	79.379936	71.419876	101.790604

Hình 3.6: Thống kê bộ dữ liệu vận hành

Mối tương quan của nhiệt độ đầu ra tại hai cấp nén (cấp 1 và cấp 2) được thể hiện trong hình Hình 3.7 và Hình 3.8 cho thấy nhiệt độ đầu vào và đầu ra của máy nén có độ tương quan cao 0,97 là phù hợp vì nhiệt độ này có thể tính toán bằng công thức cũng như bằng phần mềm mô phỏng, phần sai số của tính toán đến từ áp suất làm việc, thành phần khí, điều kiện làm việc của bộ làm mát khí sau mỗi chu trình nén và việc áp dụng máy học có thể được áp dụng để tìm ra mối liên hệ đó và dự đoán nhiệt độ đầu ra của máy nén dựa trên điều kiện làm việc của hệ thống.



Hình 3.7: Phân tích tương quan giữa nhiệt độ đầu ra máy nén và các thông số khác



Hình 3.8: Phân tích mối tương quan của các thông số bằng bản đồ nhiệt

Scale dữ liệu:

Để giảm sai số cho quá trình huấn luyện và dự đoán, dữ liệu được tỷ lệ lại từ 0 đến 1 bằng hàm tỷ lên nhỏ nhất -lớn nhất (min-max scaler) được tính theo công thức:

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.9)$$

Trong đó:

X là giá trị được tỷ lệ

X_{max} là giá trị lớn nhất của bộ dữ liệu

X_{min} là giá trị nhỏ nhất của bộ dữ liệu

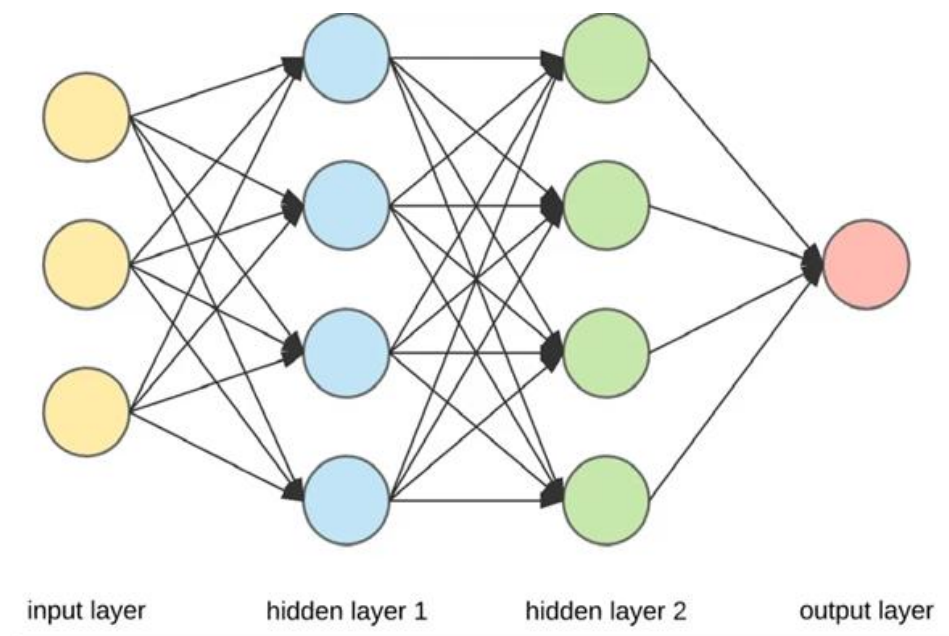
Chia dữ liệu thành tập đào tạo và tập kiểm tra:

Nhóm nghiên cứu đã trộn dữ liệu để tăng độ biến thiên của dữ liệu giúp cho việc training model nhanh đạt độ chính xác cao hơn, sau đó tập dữ liệu được chia thành tập dành cho đào tạo và tập dành cho kiểm tra với tỷ lệ 80/20 nghĩa là 80% các mẫu được sử dụng để đào tạo mô hình, 20% còn lại được sử dụng để kiểm tra.

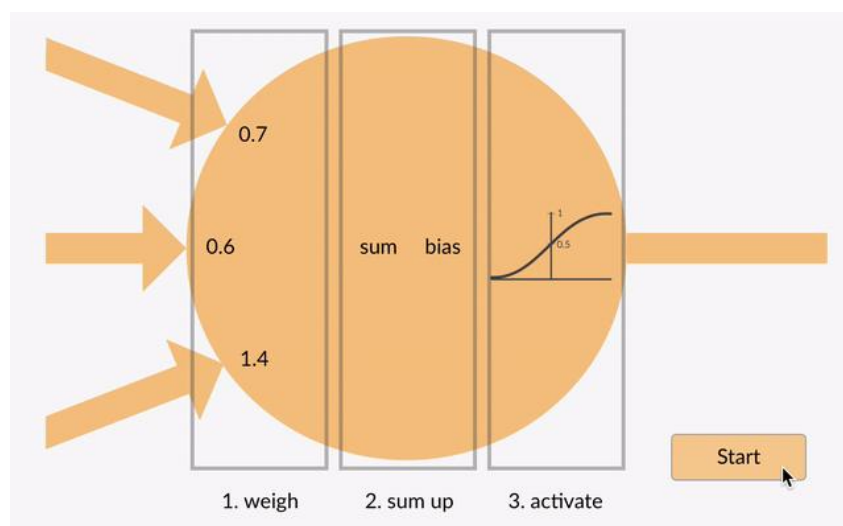
2.1.1. Nghiên cứu phát triển mô hình

Artificial Neural Network (ANN) hay mạng thần kinh nhân tạo là một hệ thống tính toán lấy cảm hứng từ sự hoạt động của các thần kinh trong hệ thần kinh. Chúng ta có thể coi ANN là dữ liệu thống kê phi tuyến. Điều này đồng nghĩa với một mối quan hệ phức tạp, được xác định giữa đầu vào và đầu ra. Kết quả là chúng tôi sẽ có nhiều mẫu khác nhau.

ANN gồm 3 thành phần chính: Input layer và output layer chỉ gồm 1 layer, hidden layer có thể có 1 hay nhiều layer tùy vào bài toán cụ thể. ANN hoạt động theo hướng mô tả lại cách hoạt động của hệ thần kinh với các neuron được kết nối với nhau.



Hình 3.9: Mô hình mạng thần kinh nhân tạo Artificial Neural Network (ANN)



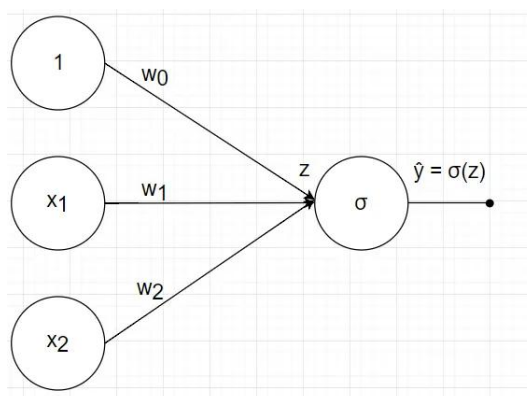
Hình 3.10: Nút mạng (node)

Mỗi node trong hidden layer và output layer:

- Liên kết với tất cả các node ở layer trước đó với các hệ số w riêng.
- Mỗi node có 1 hệ số bias b riêng.

- Diễn ra 2 bước: tính tổng linear và áp dụng hàm truyền (activation function).

Logistic regression là mô hình neural network đơn giản nhất chỉ với input layer và output layer: $\hat{y} = \sigma(w_0 + w_1 * x_1 + w_2 * x_2)$ với σ là hàm sigmoid được coi là hàm kích hoạt (activation function).



Hình 3.11: Mô hình Logistic regression

ANN được sử dụng để giải quyết các vấn đề liên quan đến dữ liệu dạng bảng, hình ảnh hay văn bản. ANN có khả năng học bất kỳ hàm phi tuyến nào, do đó, nó thường được biết đến cái tên “Phương pháp Xấp xỉ Hàm phổ quát” (Universal Function Approximators). Một trong những lý do chính đằng sau tính xấp xỉ phổ quát là hàm kích hoạt. Các hàm kích hoạt thể hiện các thuộc tính phi tuyến cho mạng. Điều này giúp mạng tìm hiểu bất kỳ mối quan hệ phức tạp nào giữa đầu vào và đầu ra.

Tuy nhiên, ANN cũng có một số nhược điểm. ANN dễ xảy ra hiện tượng Vanishing Gradient và Exploding Gradient như một đặc điểm cố hữu của các mạng thần kinh (Neural Network), nhất là khi số lượng hidden layer lớn. Trong quá trình huấn luyện các mạng nơ ron, trong một số trường hợp thời gian training có thể kéo dài hay kết quả trả về có độ chính xác không như mong muốn. Một trong những nguyên nhân gây nên những hiện tượng trên có liên quan đến Gradients (tốc độ biến thiên của hàm số), hay cụ thể hơn là Vanishing / Exploding Gradients. Gradient của một hàm số là đạo hàm của hàm số đó

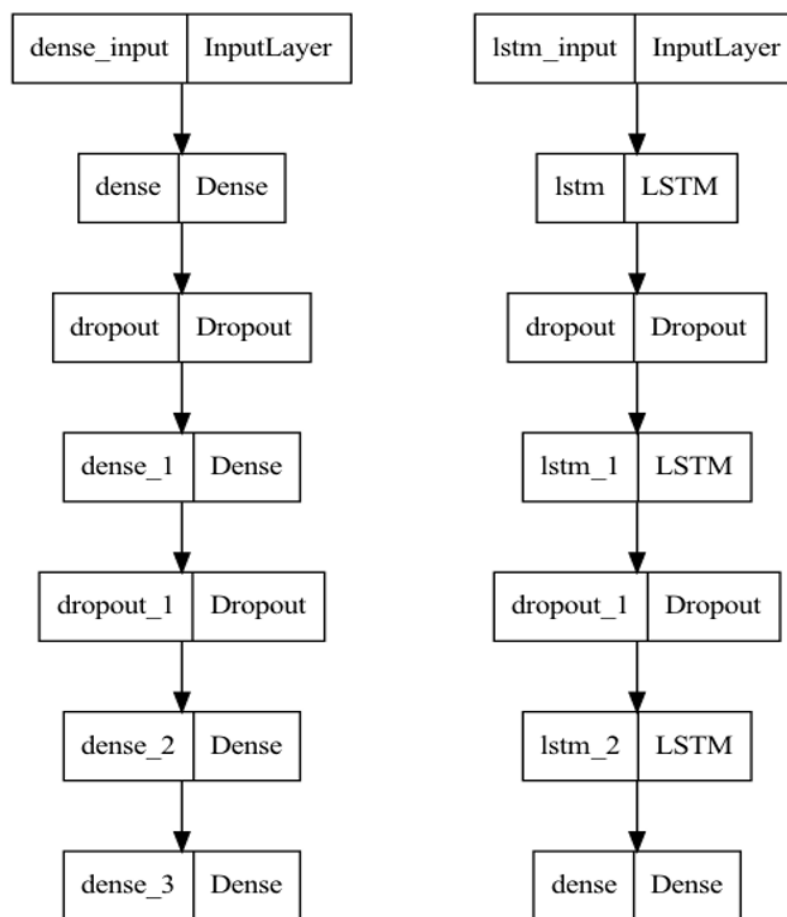
tương ứng với mỗi biến của hàm. Đối với hàm số đơn biến, chúng ta sẽ sử dụng khái niệm Derivative thay cho Gradient. Mỗi thành phần được gọi là đạo hàm riêng (đạo hàm từng phần hay Partial Derivative) tương ứng với một biến nào đó.

Tuy nhiên trên thực tế gradients thường sẽ có giá trị nhỏ dần khi đi xuống các lớp thấp hơn. Kết quả là các cập nhật thực hiện bởi Gradient Descent không làm thay đổi nhiều trọng số của các layer đó, khiến chúng không thể hội tụ và DNN sẽ không thu được kết quả tốt. Hiện tượng này được gọi là gradient biến mất (Vanishing gradients). Trong nhiều trường hợp khác, gradients có thể có giá trị lớn hơn trong quá trình lan truyền ngược, khiến một số layers có giá trị cập nhật cho trọng số quá lớn khiến chúng phân kỳ (phân rã), tất nhiên DNN cũng sẽ không có kết quả như mong muốn. Hiện tượng này được gọi là gradient bùng nổ (Exploding gradients), và thường gặp khi sử dụng Mạng nơ-ron hồi quy (RNN). Tóm lại, trong quá trình huấn luyện DNN chúng ta có thể gặp phải các vấn đề liên quan đến việc gradients không ổn định khiến cho tốc độ học (learning rate) của các layer khác nhau chênh lệch khá nhiều.

Hai hiện tượng trên là một trong những nguyên nhân khiến neural networks không nhận được sự quan tâm trong một thời gian khá dài. Tuy nhiên trong một nghiên cứu được thực hiện bởi Xavier Glorot và Yoshua Bengio năm 2010, các tác giả đã đưa ra một số nguyên nhân dẫn đến hiện tượng trên. Trong đó việc lựa chọn hàm kích hoạt (activation function) và kỹ thuật khởi tạo trọng số (weight initialization) là hai nguyên nhân chính.

Đầu tiên với dữ liệu dạng bảng (như hình ảnh), ta phải chuyển về vector 1 chiều trước khi huấn luyện mô hình, khiến kích thước của dữ liệu tăng lên cũng như mất các tính năng của dữ liệu. ANN cũng không nắm bắt được các thông tin dạng chuỗi (sequential/ series data) do luồng dữ liệu được đưa vào đồng thời.

Mạng thần kinh nhân tạo (ANN) và mạng bộ nhớ dài hạn, ngắn hạn được nhóm lựa chọn là hai mô hình để thực hiện dự đoán cho nhiệt độ đầu ra của máy nén, về cấu trúc cả hai mô hình nhóm sử dụng đều bao gồm 04 lớp mạng xếp chồng lên nhau, giữa các lớp mạng có sử dụng lớp Dropout để làm chậm quá trình đào tạo và tránh hiện tượng quá khớp (overfitting) **Error! Reference source not found.**



Hình 3.12: Cấu trúc mô hình ANN và LSTM

Các tham số cụ thể được nêu trong Bảng 3.2.

Bảng 3.2: Số tế bào thần kinh tại mô hình ANN và LSTM

Mô hình	Lớp	Số tế bào thần kinh
ANN	Dense	128
	Dense_1	64
	Dense_2	32
	Dense_3	2

LSTM	LSTM	128
	LSTM_1	64
	LSTM_2	32
	Dense	2

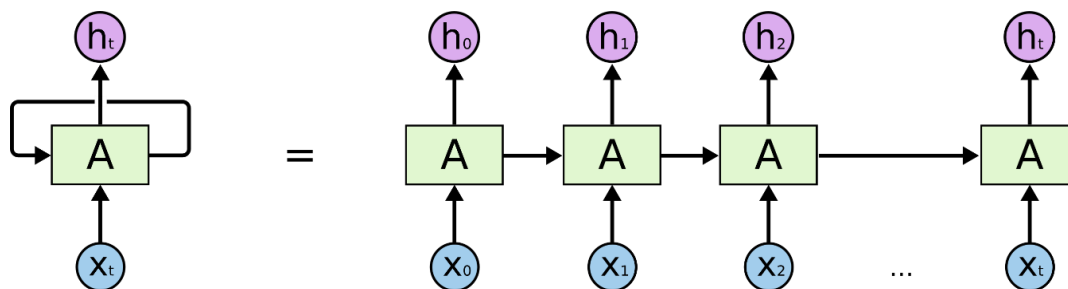
Quá trình đào tạo mô hình với các siêu tham số:

Bảng 3.3: Các siêu tham số sử dụng để đào tạo mô hình

Siêu tham số	Chức năng	Giá trị
Hàm kích hoạt	Chuyển đổi tổng trọng số của đầu vào thành đầu ra từ một hoặc nhiều nút trong lớp mạng	Đơn vị tuyến tính chỉnh lưu (ReLU)
Hàm mất mát	Tìm hiểu các hàm mất mát cung cấp hiệu suất tốt nhất theo chỉ số được xác định trước	MSE
Dropout	Một phần của các đơn vị đầu vào để giảm (hoặc trọng lượng bị đóng băng trong quá trình huấn luyện)	0,2
Kích thước lô và số lượng epoch	Số lượng vòng lặp cho quá trình đào tạo	500
Hàm tối ưu	Tối ưu trong quá trình đào tạo mô hình	Adam

Đối với Long short-term memory (LSTM) hay bộ nhớ dài - ngắn hạn, được phát triển từ mạng nơ-ron hồi quy (Recurrent Neural Network - RNN), Long short-term memory (LSTM) hay bộ nhớ dài - ngắn hạn cũng có các kết nối phản hồi. LSTM có thể xử lý không chỉ các điểm dữ liệu đơn lẻ (chẳng hạn như hình ảnh) mà còn dữ liệu dạng chuỗi (chẳng hạn như speech hoặc video).

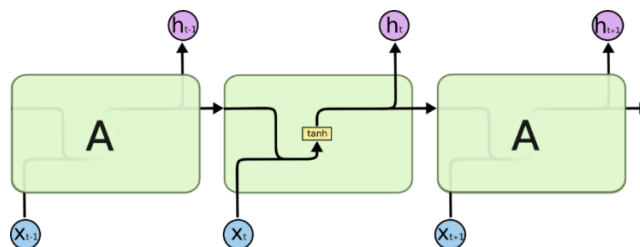
Error! Reference source not found. thể hiện mô hình cơ bản của RNN, ta có thể thấy RNN có thể mang thông tin từ các lớp (layer) trước đến các lớp sau:



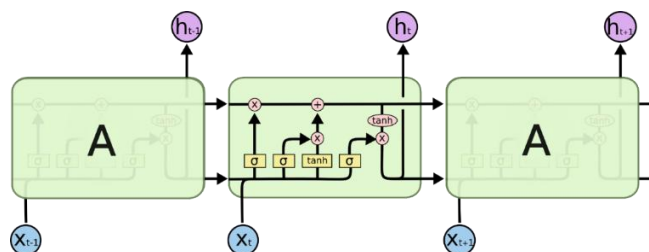
Hình 3.13: Mô hình cơ bản RNN

Tuy nhiên thực tế là thông tin chỉ mang được qua một số lượng trạng thái (state) nhất định, sau đó thì sẽ bị gặp hiện tượng gradient biến mất (Vanishing gradient), hay nói cách khác là model chỉ học được từ các trạng thái gần nó hay bộ nhớ ngắn hạn (Short term memory). Khi cần sử dụng được các thông tin từ trạng thái ở trước đó rất xa hay bộ nhớ dài hạn (Long term memory), LSTM được ra đời được phát triển để khắc phục một phần nhược điểm hiện tượng gradient biến mất trên RNN.

Một đơn vị LSTM chung bao gồm một tế bào (cell), một cổng nhập (input gate), một cổng xuất (output gate) và một cổng quên (forget gate). Tế bào ghi nhớ các giá trị trong khoảng thời gian tùy ý và ba cổng điều chỉnh luồng thông tin nhập và xuất. LSTM rất phù hợp để phân loại, phát triển, và dự đoán có khoảng thời gian không xác định.

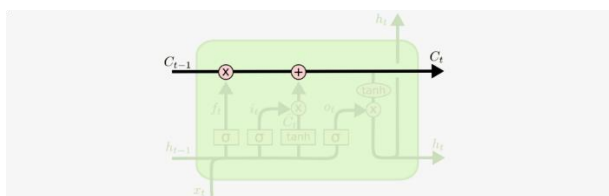


Hình 3.14: Mô hình của RNN chuẩn



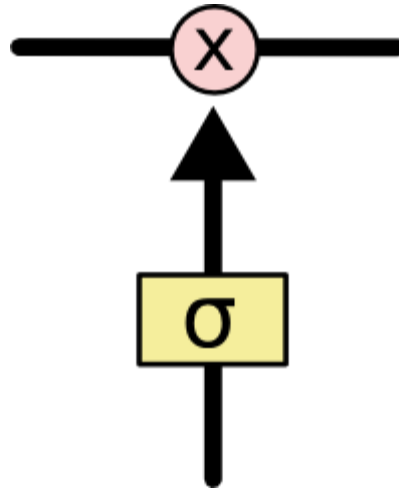
Hình 3.15: Mô hình của LSTM

Chìa khóa của LSTM là trạng thái tế bào (cell state) – chính đường chạy thông ngang phía trên của sơ đồ Hình 3.16:



Hình 3.16: LSTM trạng thái tế bào (cell state)

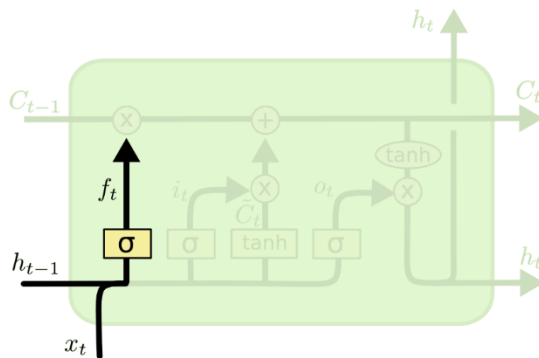
Trạng thái tế bào là một dạng giống như băng truyền. Nó chạy xuyên suốt tất cả các tế bào và chỉ tương tác tuyến tính đôi chút. Vì vậy mà các thông tin có thể dễ dàng truyền đi thông suốt mà không sợ bị thay đổi. LSTM có khả năng bỏ đi hoặc thêm vào các thông tin cần thiết cho trạng thái tế bào, chúng được điều chỉnh cẩn thận bởi các nhóm được gọi là cổng (gate). Các cổng là nơi sàng lọc thông tin đi qua nó, chúng được kết hợp bởi một tầng mạng sigmoid và một phép nhân.



Hình 3.17: Cổng LSTM (gate)

Một LSTM gồm có 3 cổng như vậy để duy trì và điều hành trạng thái của tế bào:

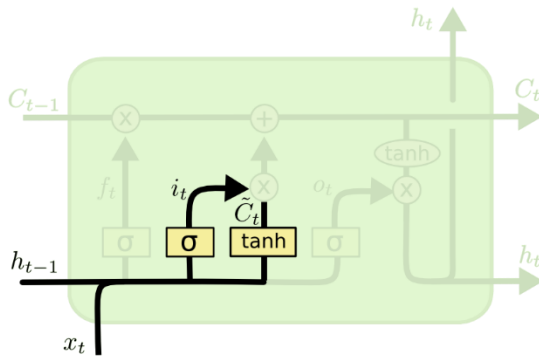
Bước đầu tiên của LSTM là quyết định xem thông tin nào cần bỏ đi từ trạng thái tế bào. Quyết định này được đưa ra bởi tầng sigmoid – gọi là tầng cổng quên (forget gate layer)



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Hình 3.18: Tầng cổng quên (forget gate layer)

Tầng cổng nhập (input gate layer) giúp quyết định xem thông tin mới nào ta sẽ lưu vào trạng thái tế bào

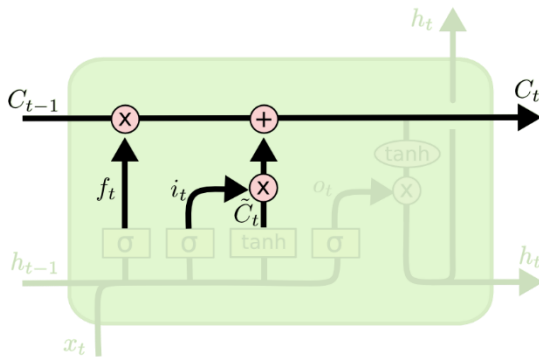


$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Hình 3.19: Tầng cổng nhập (input gate layer)

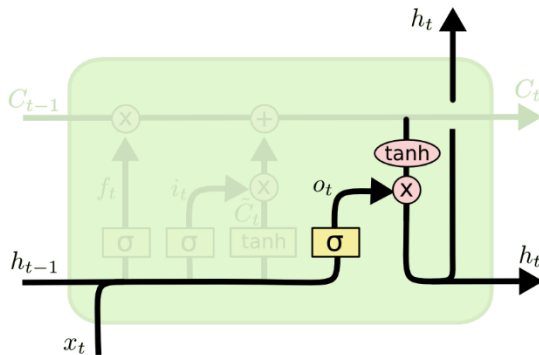
Việc tiếp theo là cập nhật trạng thái tế bào cũ thành trạng thái mới.



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Hình 3.20: Cập nhật trạng thái tế bào

Cuối cùng, ta cần quyết định xem ta muốn đầu ra là gì. Giá trị đầu ra sẽ dựa vào trạng thái tế bào, nhưng sẽ được tiếp tục sàng lọc. Quyết định này được gọi là tầng cổng xuất (output gate).



$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

Hình 3.21: Tầng cổng xuất – (output gate)

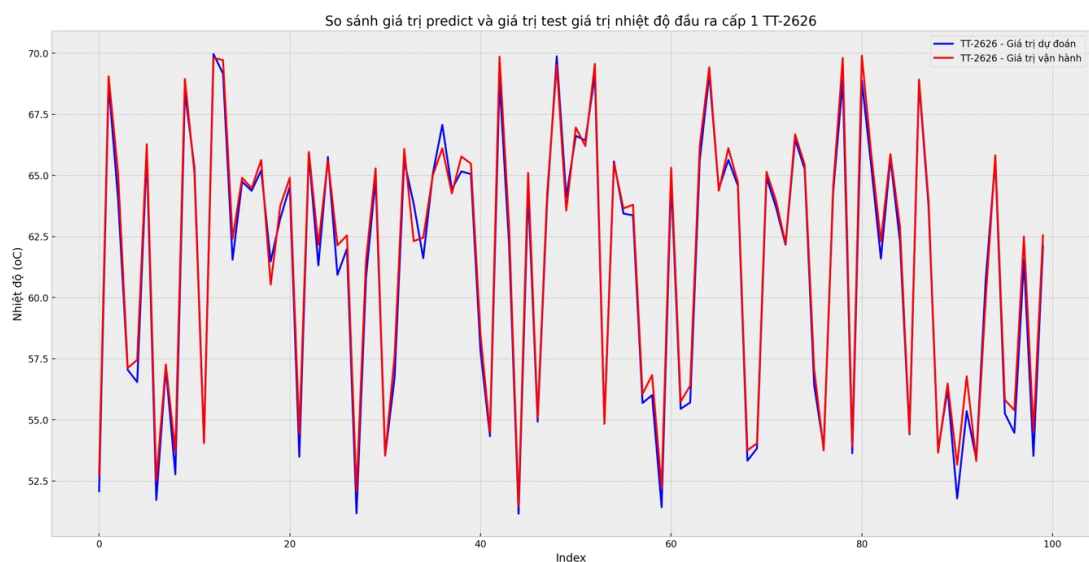
Vì vậy, với những cải tiến so với RNN thuần, LSTM đã và đang được sử dụng phổ biến, đồng thời rất phù hợp với bài toán và bộ dữ liệu của chuyên đề

này. Tuy nhiên, nó cũng tăng nhược điểm cố hữu của RNN là tính toán chậm khiến LSTM nay còn chậm hơn.

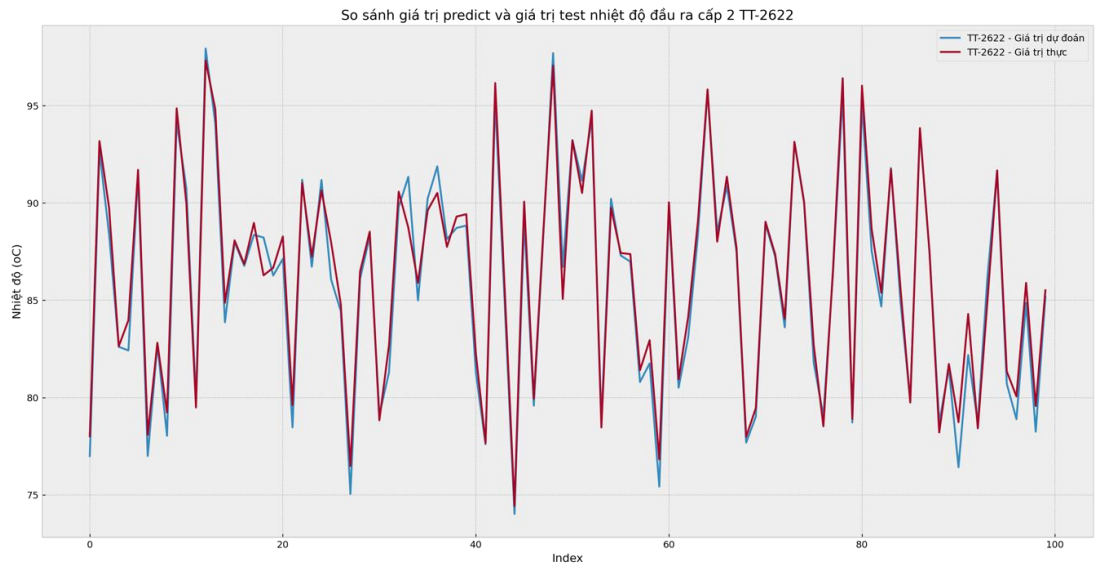
4. Kết quả và thảo luận

4.1. Kết quả chi tiết trong quá trình xây dựng, huấn luyện và tối ưu hoá mô hình

Cả hai mô hình đã dự đoán được giá trị nhiệt độ đầu ra của máy nén khí cao áp khi áp dụng tập dữ liệu kiểm tra, cụ thể như dưới đây:

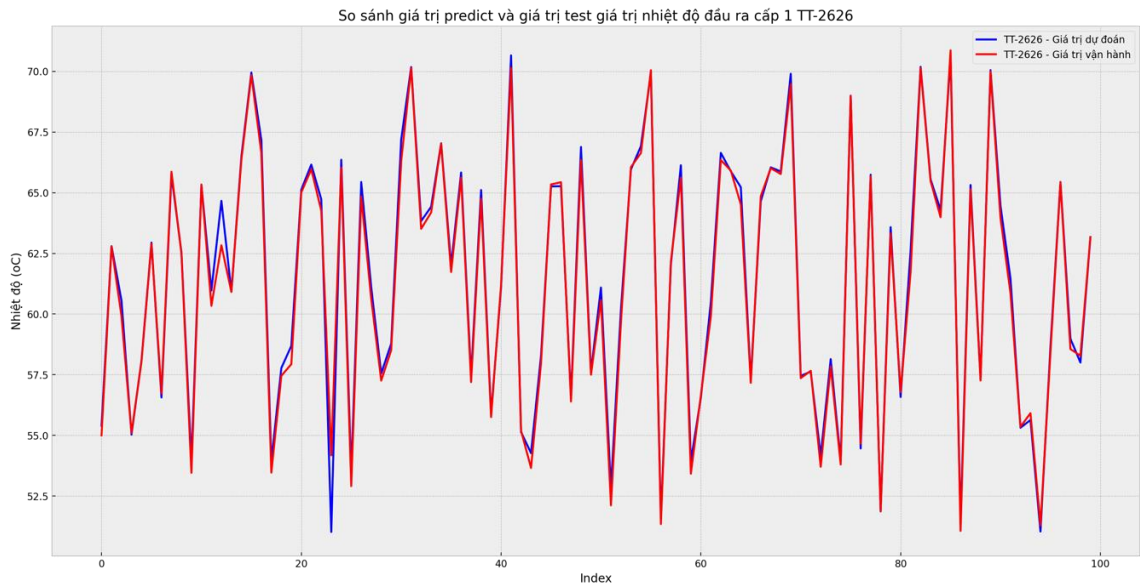


Hình 4.1: So sánh giá trị dự đoán và giá trị thực nhiệt độ đầu ra cấp 1 máy nén bằng mô hình ANN

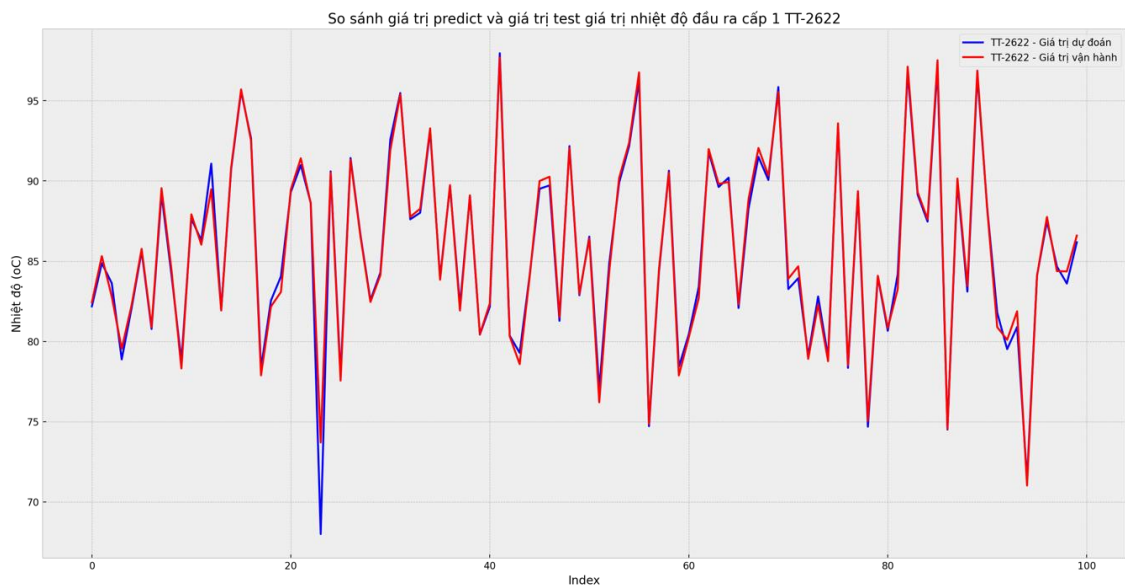


Hình 4.2: So sánh giá trị dự đoán và giá trị thực nhiệt độ đầu ra cấp 2 máy nén bằng mô hình ANN

Hình 4.1 và Hình 4.2 là kết quả dự báo của mô hình ANN cho nhiệt độ đầu ra cấp 1 và nhiệt độ đầu ra cấp 2 của máy nén khí cao áp cho 100 điểm đầu tiên trên bộ dữ liệu kiểm tra. Đường dự đoán là đường màu xanh trong khi đường màu đỏ là giá trị vận hành. Vẫn có những sai số đáng kể trong quá trình dự đoán $RMSE = 0,739$ và nhóm nghiên cứu vẫn đang tìm cách để nâng cao sai số dự đoán cho mô hình này.



Hình 4.3: So sánh giá trị dự đoán và giá trị thực nhiệt độ đầu ra cấp 1 máy nén bằng mô hình LSTM



Hình 4.4: So sánh giá trị dự đoán và giá trị thực nhiệt độ đầu ra cấp 2 máy nén bằng mô hình LSTM

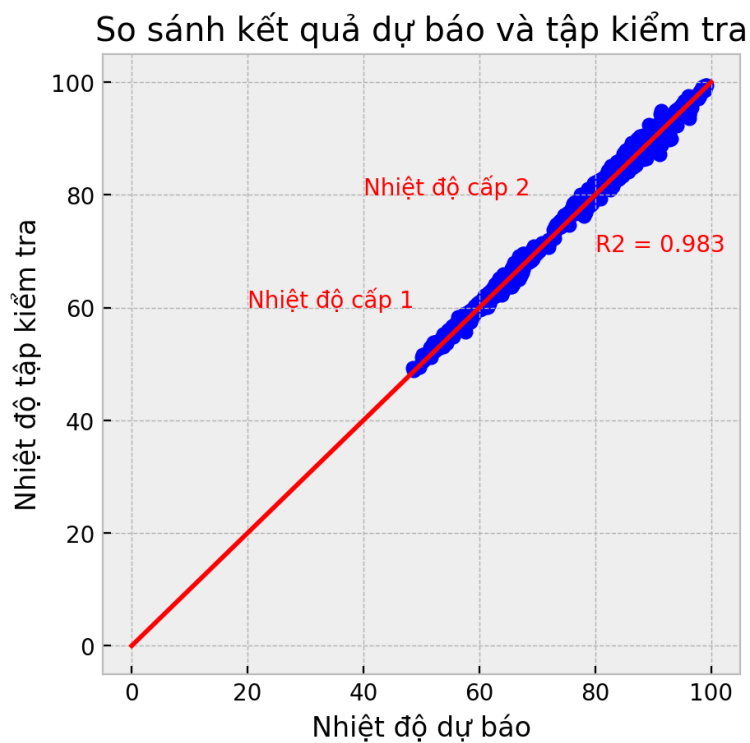
Hình 4.3 và Hình 4.4 thể hiện kết quả dự báo của mô hình LSTM, giá trị dự báo đã khớp hơn với giá trị tập kiểm tra, sai số cũng nhỏ hơn $RMSE = 0,424$.

4.2. Đánh giá độ tin cậy và tính khả thi của mô hình dự đoán

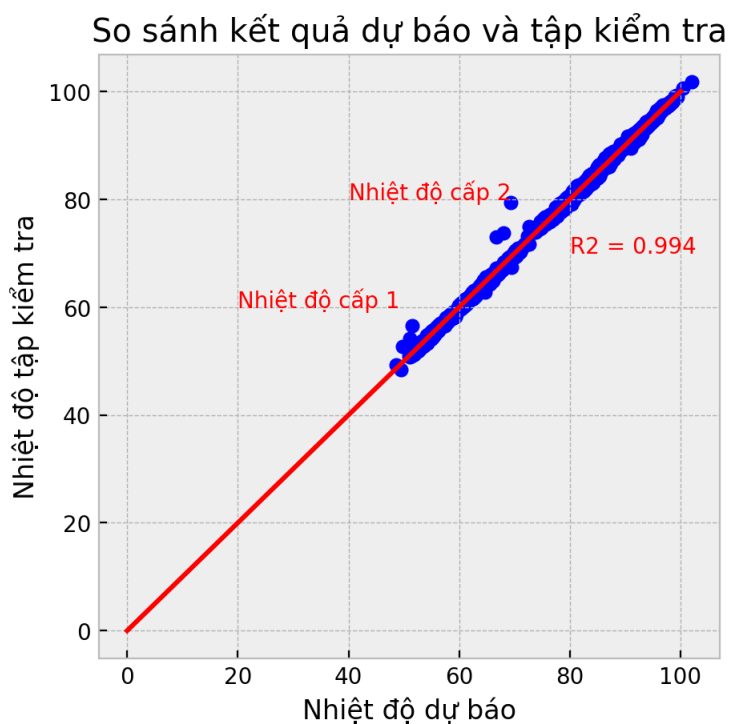
Các tiêu chí đánh giá cho cả hai mô hình được thể hiện trong Bảng 4.1 cho thấy mô hình sử dụng mạng LSTM có kết quả tốt hơn so với mô hình sử dụng mạng ANN cụ thể ở các tiêu chí về độ tương quan R^2 đạt kết quả cao hơn (0,994 so với 0,983) và sai số RMSE nhỏ hơn (0,424 so với 0,739) thể hiện mô hình LSTM có tính ưu việt hơn mô hình ANN đối với bộ dữ liệu này. Nhóm nghiên cứu sẽ tiếp tục cải thiện mô hình để đạt được kết quả dự báo chính xác hơn.

Bảng 4.1: So sánh hai mô hình dự đoán

Metric	Mô hình ANN	Mô hình LSTM
Val_loss	0,560	0,140
R^2	0,983	0,994
RMSE	0,739	0,4240



Hình 4.5: So sánh kết quả dự báo và kết quả của tập kiểm tra bằng mô hình ANN



Hình 4.6: So sánh kết quả dự báo và kết quả của tập kiểm tra bằng mô hình LSTM

4.3. So sánh với các phương pháp truyền thống sử dụng công thức vật lý và mô hình HYSYS

Bảng sau tổng hợp so sánh kết quả dự đoán nhiệt độ đầu ra của cấp 1 và cấp 2 của máy nén bằng thuật toán trí tuệ nhân tạo so với số liệu thực tế hoặc số liệu do nhà thầu cung cấp và kết quả tính toán bằng phần mềm mô phỏng Hysys trong cùng trong cùng một điều kiện vận hành nhiệt độ áp, suất đầu vào và áp suất đầu ra của mỗi cấp.

Bảng 4.2: So sánh kết quả dự đoán nhiệt độ đầu ra sử dụng các phương pháp dùng phần mềm mô phỏng và ML

Trường hợp	Nhiệt độ đầu ra máy nén cấp 1 (°C)				Nhiệt độ đầu ra máy nén cấp 2 (°C)			
	Thực tế	Hysys	ANN	LSTM	Thực tế	Hysys	ANN	LSTM
1	59,5	60,4	57,8	60,1	87,7	88,5	84,6	84,0
2	53,7	55,2	54,4	53,3	76,8	78,5	78,3	75,7
3*	95,9	97,0	88,8	97,3	151,8	154,5	129,6	131,7

**: Số liệu do Nhà thầu cung cấp và đây là điểm làm việc nằm ngoài bộ dữ liệu cho học máy.*

Kết quả cho thấy, đối với những điểm làm việc nằm trong khoảng thông số vận hành của bộ dữ liệu sử dụng để đào tạo cho các thuật toán học máy, cả hai mô hình sử dụng thuật toán mạng thần kinh nhân tạo ANN và mô hình sử dụng thuật toán bộ nhớ dài - ngắn hạn LSTM đều cho kết quả dự đoán nhiệt độ đầu ra tại mỗi cấp của máy nén có độ chính xác cao, tương đương với kết quả mô phỏng bằng phần mềm Hysys. Kết quả có độ sai lệch so với số liệu thực tế từ 0,8% đến 4,2%.

Đối với trường hợp điểm làm việc nằm ngoài khoảng thông số vận hành của bộ dữ liệu sử dụng để đào tạo cho các học máy (trường hợp 3), cả hai thuật toán đều cho kết quả có độ sai lệch cao, lên đến 10% so với số liệu thực tế.

Từ những kết quả trên cho thấy mô hình dựa trên các thuật toán máy học có thể được sử dụng để dự đoán một cách chính xác điều kiện vận hành của hệ thống công nghệ với điều kiện điểm làm việc nằm trong khoảng giới hạn của bộ dữ liệu sử dụng cho học máy. Ưu điểm của phương pháp sử dụng các thuật toán trí tuệ nhân tạo là có độ linh động cao, và có thể cập nhật cơ sở dữ liệu liên tục theo thời gian thực trong quá trình vận hành để tăng cường độ chính xác cho mô hình dự đoán thay vì phải lấy mẫu định kỳ của các mẫu khí để cập nhật sự thay đổi của thành phần và tính chất của các cấu tử và sau đó sử dụng các phần mềm chuyên dụng để tiến hành mô phỏng khi có sự thay đổi điều kiện vận hành với chi phí đắt đỏ. Như vậy, việc sử dụng mô hình dựa trên các thuật toán trí tuệ nhân tạo giúp tiết kiệm được thời gian, chi phí nhân lực, cũng như các chi phí để phân tích các mẫu và thuê phần mềm chuyên dụng.

Trong trường hợp điểm làm việc nằm ngoài khoảng thông số vận hành của bộ dữ liệu sử dụng cho học máy, các mô hình cần được tinh chỉnh để cho kết quả dự báo được chính xác hơn.

4.4. Đề xuất chương trình triển khai áp dụng tương tự cho 02 hệ thống công nghệ khác trên giàn

Kết quả sử dụng thuật toán máy học vào xây dựng mô hình dự đoán điều kiện vận hành của các thiết bị cho thấy có nhiều lợi thế so với các công cụ tính toán truyền thống khác. Hiện nay trên giàn PQP Hải Thạch còn có thêm các hệ thống công nghệ sử dụng máy nén khí khác như hệ thống thu hồi khí thấp áp (LP/IP FGC) và hệ thống thu hồi khí trung áp (MP FGC). Tương tự như kết quả nghiên cứu trên, nhóm tác giả đề xuất tiếp tục thực hiện nghiên cứu áp dụng các mô hình sử dụng thuật toán học máy vào công việc dự đoán nhiệt đầu ra cho hệ thống máy nén thu hồi thấp áp và trung áp.

KẾT LUẬN

Ý nghĩa khoa học

Việc dự đoán nhiệt độ đầu ra tại mỗi cấp của máy nén khí điều kiện vận hành thay đổi có ý nghĩa quan trọng trong công tác vận hành và tối ưu hóa hoạt động sản xuất hàng ngày tại giàn khai thác dầu khí PQP – Hải Thạch của Công ty Điều hành Dầu khí Biển Đông. Kết quả của việc ứng dụng các thuật toán học máy vào công việc dự đoán nhiệt độ đầu ra tại mỗi cấp của máy nén trong các điều kiện vận hành thay đổi đã giải quyết được các vấn đề khó khăn về thời gian cũng như sai số trong việc tính toán và dự đoán bằng các công thức tính toán.

Bên cạnh các phần mềm mô phỏng nhiệt động học thương mại, các thuật toán học máy cho thấy có nhiều ưu điểm như xây dựng mô hình đơn giản, tiết kiệm được thời gian tính toán cũng như tiết kiệm được chi phí thuê phần mềm và kết quả thu được có độ chính xác cao khi điểm làm việc nằm trong khoảng giá trị của bộ dữ liệu sử dụng để đào tạo của các công cụ học máy. Tuy nhiên, các mô hình này cần được tinh chỉnh cũng như gia tăng bộ dữ liệu sử dụng cho công việc đào tạo học máy để cải thiện độ chính xác cho việc dự đoán.

Ý nghĩa thực tiễn sản xuất

Công công tác vận hành hàng ngày, việc thay đổi điều kiện vận hành của các thiết bị thường được thực hiện để đáp ứng các nhu cầu khác nhau của sản xuất cũng như tối ưu hóa khả năng thu hồi sản phẩm từ các mỏ. Việc ứng dụng thuật toán học máy vào xây dựng mô hình dự đoán nhiệt độ đầu ra tại mỗi cấp của máy nén giúp cho người vận hành có thể dự đoán được các sự thay đổi về điều kiện vận hành có thể xảy ra từ đó có những chính sách, chiến lược vận hành phù hợp nhằm giảm thiểu các sự cố không mong muốn. Với tính linh động cao, thời gian thực hiện các dự đoán nhanh và chính xác, các mô hình sử dụng các thuật toán học máy là lựa chọn phù hợp để hỗ trợ công tác sản xuất hàng ngày.

Đề xuất chương trình triển khai áp dụng tương tự cho 02 hệ thống công nghệ khác trên giàn PQP-HT

Kết quả sử dụng thuật toán máy học vào xây dựng mô hình dự đoán điều kiện vận hành của các thiết bị cho thấy có nhiều lợi thế so với các công cụ tính toán truyền thống khác. Hiện nay trên giàn PQP – Hải Thạch còn có thêm các hệ thống công nghệ sử dụng máy nén khí khác như hệ thống thu hồi khí thấp áp (LP/IP FGC) và hệ thống thu hồi khí trung áp (MP FGC). Tương tự như kết quả nghiên cứu trên, nhóm tác giả đề xuất tiếp tục thực hiện nghiên cứu áp dụng các mô hình sử dụng thuật toán học máy vào công việc dự đoán nhiệt đầu ra cho hệ thống máy nén thu hồi thấp áp và trung áp.