Áp dụng thuật toán hồi quy Random Forest để xây dựng mô hình dự báo hệ số Poisson tầng Oligocene mỏ Bạch Hổ

Trương Văn Từ[1,*], Nguyễn Thế Vinh[1], Nguyễn Tiến Hùng[1], Nguyễn Khắc Long[1], Nguyễn Trọng Tài[2], Đào Hiệp[3]

[1] *Trường Đại học Mỏ - Địa chất*
[2] *Zarubezhneft E&P Vietnam*
[3] *Trường Cao đẳng Công Thương Miền Trung*

Tóm tắt

Nghiên cứu này đã sử dụng thuật toán Random Forest chạy trong môi trường phần mềm mã nguồn mở với ngôn ngữ lập trình python để xây dựng mô hình dự báo hệ số Poisson tầng oligocene mỏ Bạch Hổ. Đồng thời, nhóm tác giả sử dụng công cụ trí tuệ nhân tạo để thiết lập mối tương quan giữa hệ số Poisson và các thông số khoan (tải trọng lên choòng, lưu lượng nước rửa, mô-men xoắn, tốc độ cơ học khoan, tổng tổn hao áp suất) trong tầng Oligocene mỏ Bạch Hổ. Để xây dựng, huấn luyện, kiểm tra độ chính xác của mô hình trí tuệ nhân tạo, chúng tôi sử dụng dữ liệu đo được từ 02 giếng A và B trong địa tầng Oligocene, mỏ Bạch Hổ. Hai giếng này có nhiều điểm tương đồng về đặc điểm và thành phần thạch học. Kết quả nghiên cứu cho thấy, các mô hình trí tuệ nhân tạo có khả năng dự báo hệ số Poisson có độ chính xác cao với hệ số tương quan đối với mô hình Random forest - RF là 0,84 và với giá trị sai số tương đối trung bình (AAPE) là 3,06%.

Từ khóa: Random Forest, Mô hình dự báo, Hệ số Poisson.

1. Giới thiệu chung

Tầng dầu Oligocen được phát hiện đầu tiên tại giếng khoan BH-4 mỏ Bạch Hổ vào năm 1985. Đá chứa có đặc trưng chủ yếu dạng thấu kính, thấm chứa kém, bất đồng nhất cao. Cho đến nay, tầng dầu khí này đã và đang được khai thác tại nhiều khu vực, với sự đóng góp sản lượng chủ yếu từ khu vực Đông Bắc Bạch Hổ. Tuy nhiên, Oligocen có sản lượng dầu thấp hơn so với các đối tượng khai thác còn lại, do đó cần thiết tìm thân dầu mới nhằm bổ sung sản lượng cho việc duy trì khai thác ổn định.

Ngoài ra, theo kết quả nghiên cứu tại các giếng khoan khu vực ghi nhận áp suất vỉa đá chứa Oligocen suy giảm nhanh, hệ số bù khá thấp, trong khi hệ thống bơm ép có ảnh hưởng không rõ rệt. Điều này cho thấy đối tượng có mức độ bất đồng nhất cao, sự liên thông thủy động kém. Đối với các vỉa dầu Oligocen có khả năng chứa thấp do nhiều nguyên nhân khác nhau, cần thiết áp dụng giải pháp địa-kỹ thuật nhằm cải thiện lưu lượng dầu trong đó có nứt vỡ thủy lực.

Nứt vỡ thủy lực là một phương pháp rất tốt để tăng lưu lượng các giếng khai thác dầu khí và độ tiếp nhận của các giếng bơm ép. Bản chất của phương pháp này là bơm chất lỏng dưới áp suất cao để tạo và mở rộng thêm các khe nứt trong vỉa sản phẩm, sau đó bơm ép hỗn hợp chất lỏng-hạt chèn vào vỉa để chèn ép những khe nứt tạo thành với mục đích đảm bảo khả năng chảy thấm tốt và duy trì độ thấm sau này khi kết thúc quá trình nứt vỡ. Trong thiết kế NVTL, khi mô phỏng khe nứt có đòi hỏi một trong các thông số đầu vào là Module Young và hệ số Poisson liên quan đến tính chất cơ học của đất đá (Từ Thành Nghĩa và nnk, 2017).

Module Young, hệ số Poisson và các thông số địa cơ của các tầng đất đá thường được xác định thông qua kết quả nghiên cứu nén mẫu lõi trong phòng thí nghiệm và minh giải dữ liệu địa vật lý giếng khoan. Các phương pháp này tồn tại những hạn chế: Đối với các giá trị được xác định thông qua số liệu đo địa vật lý giếng khoan được gọi là giá trị động "Dynamic", những giá trị này thường không thể dùng để tính toán độ ổn định thành giếng khoan. Để có được kết quả tính toán độ ổn định thành giếng tin cậy, cần phải chuyển đổi các giá trị "Dynamic" sang "Static" trong điều kiện địa tầng. Một số tài liệu cho rằng, giá trị hệ số Poisson động cao hơn giá trị tĩnh và giữa chúng không tồn tại mối quan hệ một cách rõ ràng. Thông thường các giá trị Dynamic lớn hơn giá trị Static và chúng phân dị, đặc biệt đối với các loại đất đá có độ đàn hồi thấp. Sự sai khác giữa những giá trị này được giải thích bằng sự ảnh hưởng của độ rỗng đất đá,

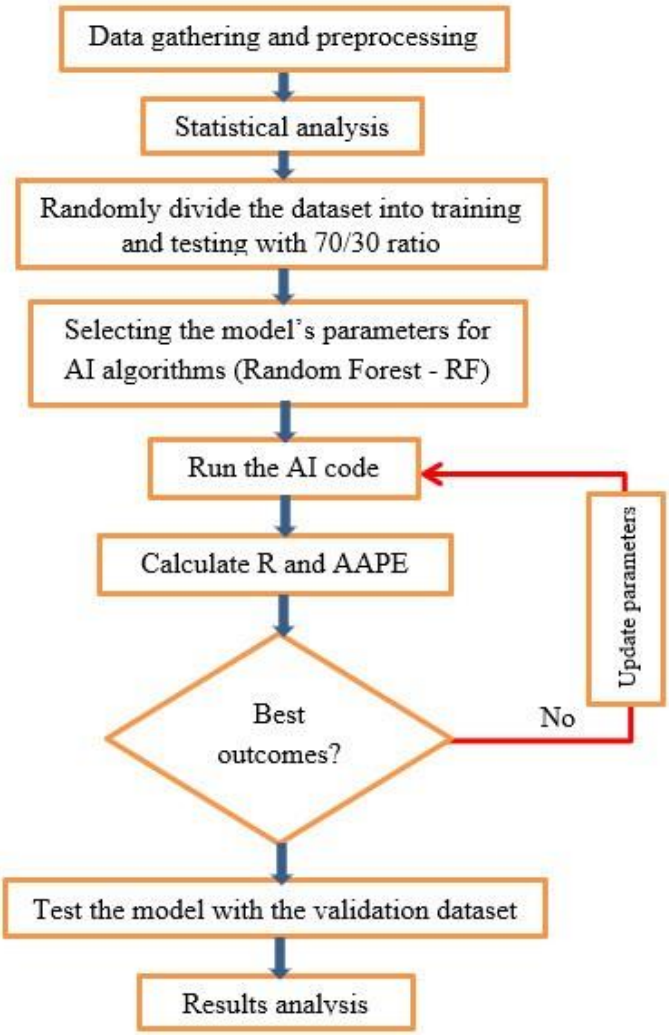
* *Tác giả liên hệ*
Email: truongvantu@humg.edu.vn

kích thước và hướng của các khe nứt hoặc thế nằm của đất đá. Để tìm ra các công thức chuyển đổi phù hợp đòi hỏi nhiều thời gian, chi phí và tương đối phức tạp (Abdallah và nnk, 2014; Lal, 1999); Phương pháp thí nghiệm nén mẫu lõi trong phòng có độ chính xác cao nhưng thường yêu cầu có sẵn mẫu thí nghiệm, đòi hỏi nhiều thiết bị đo phụ trợ và đôi khi yêu cầu kết quả đo mẫu lõi bổ sung tiêu tốn thời gian và chi phí lấy mẫu (Angelina Müller và nnk, 2019).

Vấn đề đặt ra cho các nhà nghiên cứu là tìm ra mối quan hệ biện chứng giữa hệ số Poisson với các thông khoan trong quá trình thi công giếng khoan. Một số tác giả trên thế giới cũng đã thực hiện nghiên ứng dụng trí tuệ nhân tạo (Artificial Intelligence-AI) dựa trên thông số đầu vào từ đường cong đo địa vật lý giếng khoan hoặc thông số khoan để có được giá trị các thông số địa cơ (modul đàn hồi dọc, hệ số Poisson, modul đàn hồi nén thể tích, modul đàn hồi ngang và ứng suất nhỏ nhất theo phương ngang) ảnh hưởng tới hướng phát triển khe nứt và hiệu quả nứt vỡ trong thiết kế nứt vỡ thuỷ lực đơn giản, nhanh chóng và tiết kiệm hơn (Elkatatny S., 2021; Mutalova R.F. và nkk, 2019; Siddig O. và nnk, 2021). Ứng dụng trí tuệ nhân tạo để thiết lập mối tương quan giữa hệ số Poisson và nhiều thông số ảnh hưởng đến nó bao gồm các thông số khoan thực tế hứa hẹn là giải pháp hữu hiệu để giải quyết những bất cập nêu trên mà các phương pháp truyền thống đang gặp phải.

2. Phương pháp nghiên cứu

Để dự báo hệ số Poisson từ các thông số khoan nhóm tác giả sử dụng sơ đồ thuật toán (hình 1). Dữ liệu cho các thông số khoan và hệ số Poisson thực tế theo tài liệu địa vật lý giếng khoan đã được thu thập cho hai giếng. Dữ liệu từ giếng A đã được sử dụng để xây dựng mô hình bằng một số kỹ thuật học máy. Tập dữ liệu từ giếng B được sử dụng để xác thực mô hình đã xây dựng.

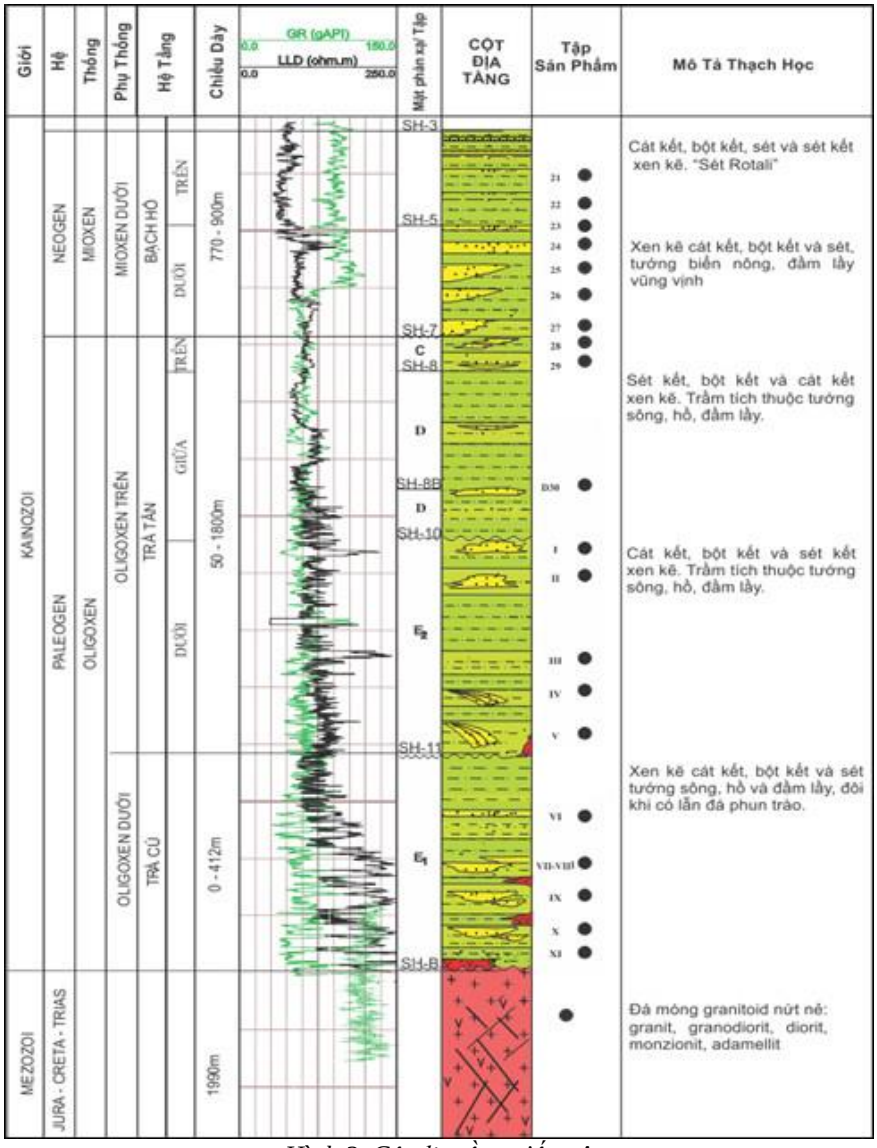


Hình 1. Sơ đồ thuật toán được sử dụng để dự đoán hệ số Poisson

2.1. Mô tả dữ liệu

Dữ liệu thu thập cho nghiên cứu này được thu thập từ các hoạt động giai đoạn khoan ở mỏ Bạch Hổ, Việt Nam. Dữ liệu bao gồm các thông số khoan và các giá trị hệ số Poisson có liên quan trong quá trình khoan công đoạn đường kính8,5". Thành phần thạch học tầng Oligocene trên và dưới công đoạn khoan này bao gồm: sét, sét kết và cát kết thể hiện trong hình 2. Giếng A có tổng cộng 714 điểm dữ liệu được sử dụng để xây dựng mô hình với 70% điểm dữ liệu dùng để huấn luyện và 30% dùng để kiểm tra mô hình.

Trong khi đó, 196 điểm dữ liệu từ giếng B được sử dụng để xác thực mô hình đã xây dựng. Mỗi điểm dữ liệu chứa các tham số khoan được sử dụng làm đầu vào để dự báo hệ số Poisson. Các thông số khoan được liệt kê bên dưới và được lấy từ các phép đo tại hiện trường sử dụng để xây dựng mô hình dự báo: Tải trọng lên choòng WOB; Mô-men quay choòng TQR;Tổng tổn hao áp suất SPP;Tốc độ quay choòng RPM; Lưu lượng nước rửa FLOWIN; Tốc độ cơ học khoan ROP.



Hình 2. Cột địa tầng giếng A

2.2. Phân tích dữ liệu

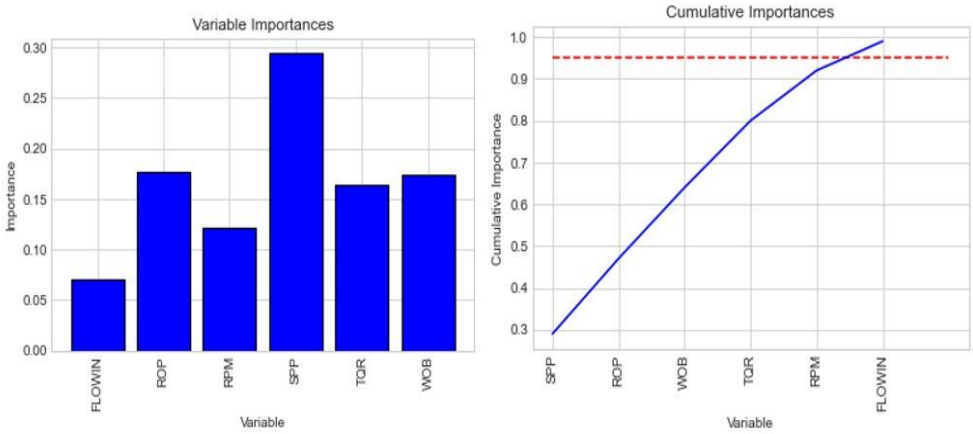
Trước khi chạy dữ liệu vào các thuật toán máy học, các bộ dữ liệu đã được lọc nhiễu và các giá trị ngoại lai bằng xác định các giá trị ngoại lai Z-score (Tripathy S. S. và nnk, 2013), phân tích dữ liệu dựa vào mối tương quan giữa 2 biến số. Phân tích thống kê của bộ dữ liệu được sử dụng để xây dựng mô hình được trình bày trong bảng 1.

Bảng 1. Thống kê dữ liệu xây dựng mô hình

	ROP	WOB	RPM	TQR	SPP	FLOWIN	POISSON
Count	714	714	714	714	714	714	714.000
Mean	17.44	8.27	117.47	1575.95	195.26	37.78	0.316
Std	9.99	1.90	20.96	204.29	24.18	9.30	0.029
Min	0.78	2.33	40.00	1037.60	143.32	22.06	0.200
25%	11.86	7.07	116.00	1534.15	184.28	34.77	0.301
50%	17.35	8.41	121.00	1593.30	202.85	38.05	0.320
75%	21.07	9.20	122.00	1669.88	212.06	38.10	0.337
Max	45.40	13.87	161.00	2185.30	224.00	54.64	0.392

Việc lựa chọn các thông số đầu vào để đào tạo và kiểm tra là một bước quan trọng quyết định tới độ chính xác của mô hình. Các hệ số tương quan giữa hệ số Poisson và các thông số khoan khác nhau được đưa ra trong hình 3. Qua hình 3a có thể nhận thấy, các hệ số tương quan của các thông số khoan và hệ số Poisson đều nhỏ hơn 0,3. Do đó, áp dụng các mô hình trí tuệ nhân tạo sẽ cho kết quả tốt hơn các phương pháp hồi quy tuyến tính vì có thể xấp xỉ các mối quan hệ phức tạp hơn.

Trong hình 3b cho thấy mối tương quan tương đối chặt chẽ giữa hệ số Poisson và một số thông số khoan như tổng tổn hao áp suất SPP, mô-men quay choòng TQR, tải trọng lên choòng WOB và tốc độ cơ học khoan ROP. Các hệ số tương quan thấp hơn đối với các tham số khác không nhất thiết ngụ ý sự vắng mặt của mối quan hệ giữa các đầu vào này và tỷ lệ Poisson, mà có nghĩa là phương trình tuyến tính không mô tả mối quan hệ giữa đầu vào và đầu ra.



(a) (b)

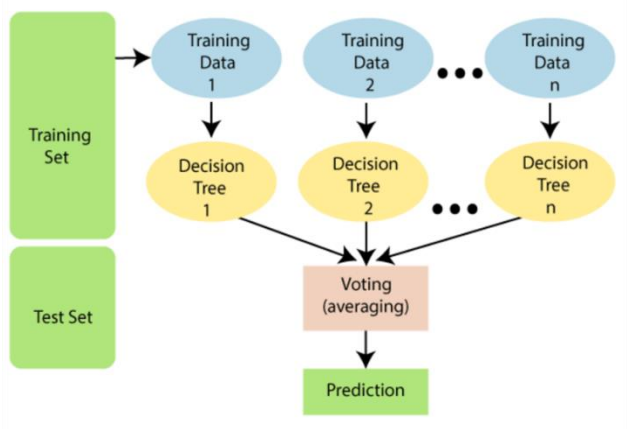
Hình 3. Mối tương quan giữa hệ số dự báo và các thông số dùng để dự báo

2.3. Xây dựng mô hình

Với mục đích xây dựng các mô hình mối quan hệ giữa hệ số Poisson và các thông số khoan, thuật toán máy học đã được sử dụng là Rừng ngẫu nhiên (Random forest - RF). Thuật toán này đều có hai khả năng là phân loại (classification) và hồi quy (regression). Trong khuân khổ ứng dụng của bài báo này, chỉ có hồi quy được sử dụng và trình bày.

2.3.1. Thuật toán Rừng ngẫu nhiên

Rừng ngẫu nhiên là một thuật toán học tập đồng bộ được đề xuất bởi Breiman năm 2001 (Breiman, L., 2001). Nó xây dựng một số lượng lớn các cây quyết định ngẫu nhiên trên các mẫu huấn luyện đã khởi động và tổng hợp các dự đoán của chúng bằng cách lấy trung bình các kết quả (James, G., 2015). Nó đã trở thành một công cụ khai thác dữ liệu chính cho cả vấn đề hồi quy và phân loại. Gần đây, tính nhất quán của Rừng ngẫu nhiên đã được chứng minh bởi Scornet năm 2015 (Scornet, E. và nnk, 2015). So với các thuật toán học máy khác như mạng thần kinh, Rừng ngẫu nhiên có thể đạt được hiệu suất dự đoán tương đối cao chỉ với một vài tham số để điều chỉnh (Genuer, R. và nnk, 2017).



Hình 4. Sơ đồ của thuật toán Rừng ngẫu nhiên

Thuật toán Rừng ngẫu nhiên:

- Từ tập dữ liệu đào tạo D, ta tạo dữ liệu ngẫu nhiên (mẫu bootstrap);

- Sử dụng các tập con dữ liệu lấy mẫu ngẫu nhiên D_1, D_2,…, Dn xây dựng nên các cây T_1, T_2, …, T_n;

- Kết hợp các cây quyết định: sử dụng chiến lược bình chọn theo số đông với bài toán phân loại hoặc lấy trung bình các giá trị dự báo từ các cây với bài toán hồi quy.

Có nhiều gói nguồn mở triển khai thuật toán Rừng ngẫu nhiên, một trong số đó là scikit-learn (F.Pedregosa và nnk, 2011; https://scikit-learn.org/) là thư viện được sử dụng phổ biến trong học máy được lựa chọn trong các nghiên cứu này với các bộ thông số được mô tả trong phần tiếp theo.

2.3.2. Lựa chọn bộ thông số của rừng ngẫu nhiên

Việc lựa chọn các bộ thông số của thuật toán Random forest được mô tả trong bước thứ 3 của sơ đồ thuật toán ở hình 1 (Selecting the model's parameters for AI algorithms (Random Forest - RF)). Các bộ thông số của cả hai thuật toán này này thể hiện lần lượt trong bảng 2 (https://scikit-learn.org/).

Bảng 2. Các bộ thông số được sử dụng cho mô hình Random forest

RANDOM FOREST					
n_estimators	max_features	max_depth	min_samples_split	min_samples_leaf	bootstrap
1	auto	None	1	1	True
5	sqrt	1	2	2	False
10		5	3	3	
100		10	4	4	
200		15	5	5	
300		20	10	10	
400		25			
500		30			
600					
700					
800					
900					
1000					

2.3.3. Đánh giá mô hình

Thuật toán Rừng ngẫu nhiên đã được sử dụng để xây dựng mô hình. Thuật toán này sử dụng 70% tập dữ liệu từ giếng A để xây dựng mô hình và 30% dữ liệu để kiểm tra nội bộ trong một số lần lặp lại và chọn ra mô hình phù hợp nhất. Sau khi có mô hình, dữ liệu từ giếng B được sử dụng làm bộ xác thực bên ngoài cho các mô hình. Để đánh giá tất cả các thử nghiệm của mô hình, hai tham số thống kê đã được sử dụng, hệ số tương quan (R) và sai số phần trăm tuyệt đối trung bình (AAPE). Hệ số tương quan (R) và sai số phần trăm tuyệt đối trung bình (AAPE) được tính bằng các phương trình (1) và phương trình (2):

$$R = \frac{\left[N\sum_{i=1}^{N}(\mu_{given\ i} \times \mu_{Predicted\ i})\right] - \left[\sum_{i=1}^{N}(\mu_{given\ i} \times \mu_{Predicted\ i})\right]}{\sqrt{\left[N\sum_{i=1}^{N}(\mu_{given\ i})^2 - \left(\sum_{i=1}^{N}\mu_{given\ i}\right)^2\right]\left[N\sum_{i=1}^{N}(\mu_{Predicted\ i})^2 - \left(\sum_{i=1}^{N}\mu_{Predicted\ i}\right)^2\right]}} \quad (1)$$

$$AAPE = \frac{\sum_{i=1}^{N}\frac{\mu_{given\ i} - \mu_{Predicted\ i}}{\mu_{given\ i}} \times 100\%}{N} \quad (2)$$

trong đó:

μ_{given} : Hệ số Poisson thực tế;

$\mu_{Predicted}$: Hệ số Poisson dự báo;

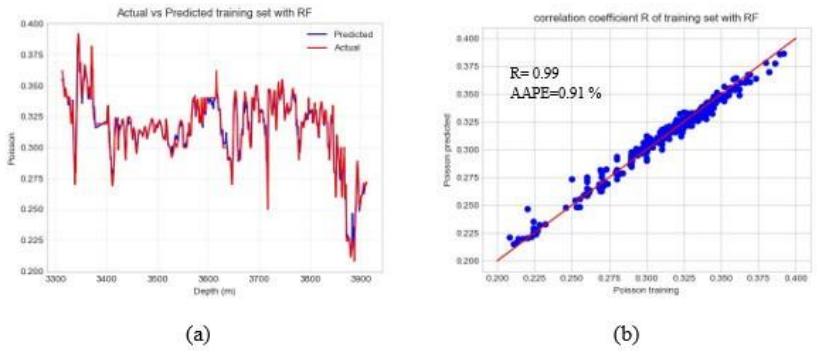
N: Tổng số điểm dữ liệu.

Sau khi thực hiện các vòng lặp của sơ đồ thuật toán như trong hình 1 và sử dụng các bộ thông số ở bảng 2, đã lựa chọn ra được bộ thông số tốt nhất đối với mô hình sử dụng thuật toán Random forest là 'n_estimators': 400, 'min_samples_split': 3, 'min_samples_leaf': 2, 'max_features': 'sqrt', 'max_depth': 20, 'bootstrap': False.

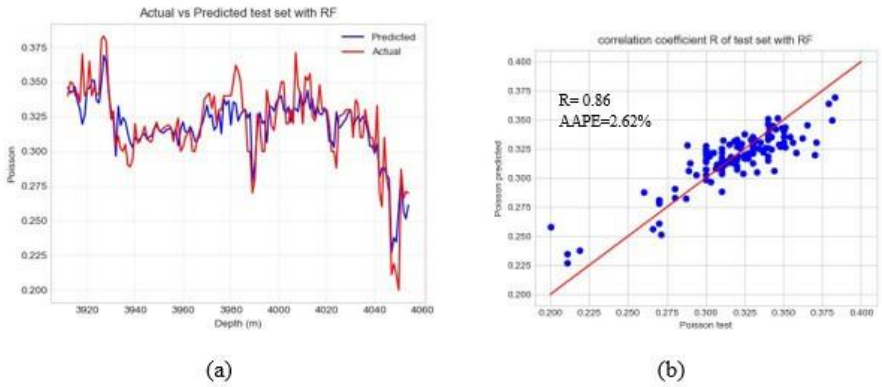
3. Kết quả và thảo luận

3.1. Đào tạo và kiểm tra mô hình xây dựng bằng thuật toán Rừng ngẫu nhiên

Mô hình Rừng ngẫu nhiên đã học cách dự đoán hệ số Poisson dựa trên: tải trọng lên choòng (WOB), mô-men quay choòng (TQR), tổng tổn hao áp suất (SPP), tốc độ quay choòng (RPM), tốc độ cơ học khoan (ROP), lưu lượng nước rửa (FLOWIN). Mô hình đã được đào tạo và kiểm tra trên 714 điểm dữ liệu từ giếng A. Hình 5a và 6a cho thấy hệ số Poisson thực tế và dự báo theo độ sâu rất sát nhau với tập đào tạo và kiểm tra. Số tương đối trung bình (AAPE) của các tập này tương ứng là 0,91% và 2,62 %. Điều này cũng được thể hiện rõ ở hình 5b và 6b là hình thể hiện sự tương thích của hệ số Poisson dự báo và thực tế, với các hệ số tương quan (R) lần lượt là 0,99 và 0,86.



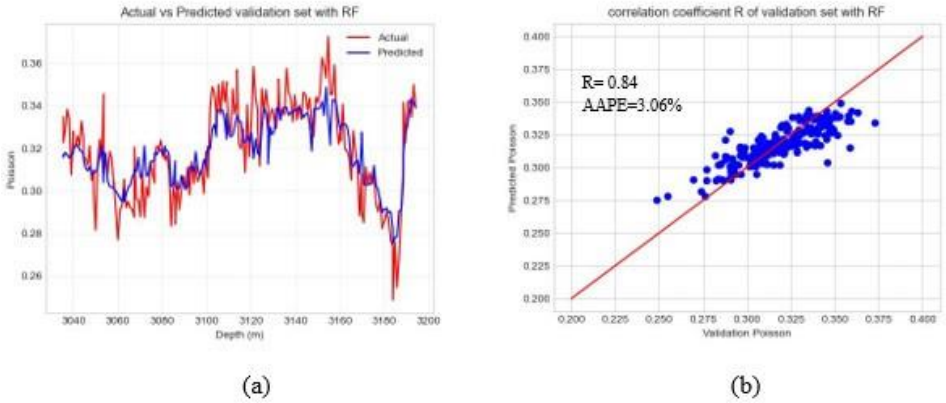
Hình 5. Hệ số Poisson thực tế và RF dự báo cho đào tạo



Hình 6. Hệ số Poisson thực tế và RF dự báo cho kiểm tra

3.2. Xác thực mô hình Rừng ngẫu nhiên

Nhóm tác giả xác thực mô hình Rừng ngẫu nhiên đã xây dựng trên 196 điểm dữ liệu từ giếng B. Hình 7a, 8b cho thấy độ chính xác của mô hình dự báo với giá trị sai số tương đối trung bình (AAPE) rất thấp chỉ 3,06% và hệ số tương quan (R) khá cao 0,84. Kết quả này khẳng định khả năng sử dụng tương quan thực nghiệm được phát triển trong nghiên cứu này dựa trên các thông số khoan để cho phép dự báo hệ số Poisson.



Hình 7. Hệ số Poisson thực tế và RF dự báo cho xác thực

Mô hình dự báo đối với tập đào tạo, kiểm tra và tập xác thực có hệ số tương quan cao và sai số thấp là do rừng ngẫu nhiên tạo ra một tập hợp các cây quyết định không cắt nhánh, mỗi cây được xây dựng trên tập mẫu bootstrap (lấy mẫu ngẫu nhiên có hoàn lại), tại mỗi nút phân hoạch tốt nhất được thực hiện từ việc chọn ngẫu nhiên một tập con các thuộc tính. Lỗi tổng quát của rừng phụ thuộc vào độ chính xác của từng cây thành viên trong rừng và sự phụ thuộc lẫn nhau giữa các cây thành viên. Giải thuật rừng ngẫu nhiên cho độ chính xác cao khi so sánh với các thuật toán học có giám sát hiện nay, giữ cho bias (lỗi của mô hình học không liên quan đến dữ liệu học) thấp và dùng tính ngẫu nhiên để điều khiển tính tương quan thấp giữa các cây trong rừng.

4. Kết luận

Thông thường, hệ số Poisson được xác định bằng hai phương pháp truyền thống: phương pháp sóng âm và phương pháp thí nghiệm nén mẫu lõi trong phòng. Tuy nhiên, các dữ liệu này không phải lúc nào cũng có sẵn. Dự báo hệ số Poisson thông qua các thông số khoan đã được đề cập trong báo cáo bằng việc ứng dụng trí tuệ nhân tạo. Dựa trên các kết quả được trình bày, nhóm tác giả có một số kết luận như sau:

- So với các phương pháp khác được sử dụng để xác định hệ số Poisson, thì sử dụng trí tuệ nhân tạo dự báo hệ số Poisson từ dữ liệu khoan sẽ có khả năng thiết lập mối tương quan giữa hệ số Poisson và các thông số liên quan chặt chẽ hơn, mang lại hiệu quả dự báo để phục vụ công tác nứt vỡ thủy lực trong vực khai thác và các lĩnh vực khác nhau trong công nghiệp dầu khí. Qua đó, tiết kiệm chi phí, thời gian, khắc phục thiếu các dữ liệu đo, thiết bị đo phụ trợ và kết quả đo mẫu lõi bổ sung. Vì vậy, việc dự đoán hệ số Poisson từ số liệu khoan sẽ mang lại nhiều lợi ích thiết thực.
- Hệ số tương quan giữa giá trị thực tế và giá trị dự đoán dao động trong khoảng 0,84 đến 0,86 sai số trung bình đều nhỏ hơn 5% khi sử dụng mô hình Rừng ngẫu nhiên (Random forest - RF). Mô hình tốt nhất được lựa chọn để dự báo hệ số Poisson với các bộ dữ liệu khác.
- Kết quả được trình bày trong bài báo này cho thấy, khả năng dự báo hệ số Poisson từ dữ liệu khoan là rất khả quan, tuy nhiên cũng nên nghiên cứu các phương pháp máy học khác. Ngoài ra, việc sử dụng dữ liệu khoan trong dự đoán các đặc tính địa cơ học khác có thể được nghiên cứu bằng cách sử dụng phương pháp tương tự. Cũng cần lưu ý rằng dữ liệu được sử dụng trong nghiên cứu này là từ một khu vực của mỏ Bạch Hổ tại Việt Nam, do đó để tạo ra mô hình có khả năng dự đoán đa dạng, cần đào tạo và kiểm tra với các bộ dữ liệu chung có thể được kết hợp từ các khu vực khác nhau của mỏ.

Tài liệu tham khảo

Abdallah Dauod Younis Elsadig, Hassan Osman Ali Elfaki et al, 2014. *Calibration of Wire-Line Mechanical Properties Using Core Measurements Results for Heglig Oilfield - Case Study*. Submitted to College of Petroleum Engineering & Technology for a partial fulfillment of the requirement for B.sc Degree.

Angelina Müller , Matthias C. Wapler and Ulrike Wallrabe, 2019. A quick and accurate method to

determine the Poisson's ratio and the coefficient of thermal expansion of PDMS. *Royal society of chemistry*.

Breiman, L., 2001. Random Forests. *Machine learning* 45 (1), 5-32.

Elkatatny S., 2021.Real-Time Prediction of the Dynamic Young's Modulus from the Drilling Parameters Using the Artificial Neural Networks. *Arab J Sci Eng*, 47(9),10933–10942.(5)

F. Pedregosa et al., 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, vol. 12, pp. 2825-2830.

Genuer, R., Poggi, J., Tuleau-Malot, C., Villa-Vialaneix, N., 2017. Random forests for big data. *Big Data Research* 9, 28-46.

G. James et al., 2021. *An Introduction to Statistical Learning*. Springer, NY, USA

James, G., Witten, D., Hastie, T., Tibshirani, R., 2015. *An introduction to statistical learning*. Springer, NY, USA

Lal M., 1999. Shale stability: drilling fluid interaction and shale strength. SPE Latin American and Caribbean Petrol Eng Conf held in Caracas, Venezuela. *SPE* 54356. Page 1-10.

Mutalova R.F. et al, 2019. Machine learning on field data for hydraulic fracturing design optimization. *Journal of Petroleum Science & Engineering.*

Scornet, E., Biau, G., Vert, J., 2015. Consistency of random forests. *The Annals of Statistics* 43 (4), 1716-1741

Siddig O., Gamal H., Elkatatny S., et al., 2021. Real-time prediction of Poisson's ratio from drilling parameters using machine learning tools. *Sci Rep*, 11(1), 12611.

Tripathy S. S., Saxena R. K., Gupta P. K., 2013. Comparison of statistical methods for outlier detection in proficiency testing data on analysis of lead in aqueous solution. *American Journal of Theoretical and Applied Statistics.* Volume 2, Issue 6, Pages: 233-242

https://scikit-learn.org/

Từ Thành Nghĩa, Nguyễn Thúc Kháng, và Nguyễn Quốc Dũng, 2016. *Công nghệ xử lý vùng cận đáy giếng các mỏ dầu khí ở thềm lục địa Nam Việt Nam*, Nhà xuất bản Đại học Quốc gia Tp. Hồ Chí Minh.

ABSTRACT

Application of Random Forest regression algorithm in building a predictive model for Poisson ratio in Oligocene of Bach Ho filed

Truong Van Tu[1], Nguyen The Vinh[1], Nguyen Tien Hung[1], Nguyen Khac Long[1], Nguyen Trong Tai[2], Dao Hiep [3]

[1] *Hanoi University of Mining and Geology*
[2] *Zarubezhneft E&P Vietnam*
[3] *Mientrung Industry and Trade College*

Study utilized the Random Forest algorithm implemented in the open-source Python programming environment to build a predictive model for the Poisson ratio in Oligocene of Bach Ho filed. Additionally, the authors employed artificial intelligence tools to establish correlations between the Poisson coefficient and drilling parameters (such as hook load, flow rates, torque, ROP, friction pressure) in the Oligocene formation at Bach Ho field. In order to build, train, and validate the accuracy of the artificial intelligence model, the offset data from two wells A and B which penetrated the Oligocene formation were utilized. These wells share many similarities in lithological characteristics and composition. The research results indicated that the artificial intelligence models are capable of accurately predicting the Poisson coefficient, with a high correlation coefficient for the Random Forest model (RF) at 0.84 and an average absolute percentage error (AAPE) of 3.06%.

Keywords: Random Forest, Predictive Model, Poisson ratio.