TẠP CHÍ DẦU KHÍ
Số 9 - 2022, trang 16 - 23
ISSN 2615-9902

ỨNG DỤNG THUẬT TOÁN HỌC MÁY ĐỂ DỰ BÁO KHAI THÁC CHO ĐỐI TƯỢNG MÓNG NỨT NẺ, VÒM TRUNG TÂM, MỎ BẠCH HỔ

Trần Đăng Tú, Lê Thế Hùng, Trần Xuân Quý, Đoàn Huy Hiên, Phạm Trường Giang, Lưu Đình Tùng
Viện Dầu khí Việt Nam
Email: tutd@vpi.pvn.vn
https://doi.org/10.47800/PVJ.2022.09-03

Tóm tắt

Dự báo khai thác mỏ dầu là thách thức lớn trong ngành công nghiệp dầu khí. Mô hình và kết quả dự báo khai thác đặc biệt cần thiết cho công tác quản lý - điều hành mỏ. Các công cụ truyền thống đang được ứng dụng phổ biến để dự báo sản lượng hiện nay là mô hình mô phỏng thủy động lực học và phương pháp phân tích đường cong suy giảm...

Mô hình mô phỏng thủy động lực học cho thấy hiệu quả rõ rệt đối với các đối tượng trầm tích. Tuy nhiên, các kết quả dự báo khai thác sử dụng mô hình mô phỏng thủy động lực học cho đối tượng móng nứt nẻ đôi khi không đủ tin cậy do móng nứt nẻ là đối tượng địa chất phức tạp, khó dự báo các đặc điểm địa chất... Phương pháp phân tích đường cong suy giảm (DCA) sử dụng các hàm toán học ngoại suy đơn giản để dự báo sản lượng khai thác do vậy kết quả dự báo không phản ánh được các quá trình vận hành mỏ như đóng/mở tầng khai thác, thay đổi lưu lượng bơm ép nước...

Để khắc phục nhược điểm của các phương pháp dự báo khai thác truyền thống, Viện Dầu khí Việt Nam (VPI) đã nghiên cứu khả năng ứng dụng thuật toán học máy để dự báo khai thác cho đối tượng móng khu vực vòm Trung tâm, mỏ Bạch Hổ. Kết quả nghiên cứu cho thấy mô hình random forest (RF) cho kết quả dự báo có độ tin cậy cao với sai số tương đối trung bình 4%.

Từ khóa: Học máy, mô hình RF, dự báo khai thác, móng nứt nẻ, mỏ Bạch Hổ.

1. Giới thiệu

Học máy là phương pháp phân tích dữ liệu tự động hóa thông qua mô hình phân tích. Bằng cách sử dụng các thuật toán hiện đại để học từ dữ liệu, học máy cho phép máy tính tìm thấy những thông tin, giá trị ẩn sâu mà không thể lập trình rõ ràng. Cách lặp của học máy rất quan trọng để khi các mô hình mạng này được tiếp xúc với dữ liệu mới, có thể thích ứng một cách độc lập nhờ học từ các tính toán trước đó để đưa ra quyết định cũng như kết quả lặp lại đáng tin cậy.

Với sự phát triển mạnh mẽ của công nghệ như hiện nay, xu hướng phát triển các thuật toán học máy có khả năng tự động áp dụng các phép tính toán học phức tạp trên tập dữ liệu ngày càng lớn và nhanh hơn. Các thuật toán học máy phổ biến nhất gồm: Mô hình hồi quy tuyến tính, mô hình mạng trí tuệ nhân tạo (Artificial Neural Network - ANN), gradient tree boosting.

Ngày nhận bài: 2/9/2021. Ngày phản biện đánh giá và sửa chữa: 2 - 9/9/2022. Ngày bài báo được duyệt đăng: 12/9/2022.

Trong nghiên cứu này, mô hình hồi quy được sử dụng để thiết lập mối quan hệ giữa tổng sản lượng khai thác dầu với các giếng bơm ép được biểu thị theo công thức sau:

$$X = \{x_i\}_{i=1}^{n} \in R^{n \times d} \quad (1)$$

Trong đó:

n: Số lượng giếng bơm ép;

d: Số bước nhảy thời gian trong lịch sử bơm ép.

Giá trị dự báo sẽ được biểu thị dưới dạng vector như sau:

$$Y = \{y_i\}_{i=1}^{n} \in R^{n \times 1} \quad (2)$$

Sau đó, cần tìm 1 hàm gần đúng $\hat{f}(x, \theta) : X \rightarrow Y$ bằng cách tìm ra bộ số để độ lớn hàm mất mát (loss function) là nhỏ nhất (cực tiểu):

$$\sum_{i=1}^{n} L\left(\hat{f}(x, \theta), y_i\right) \rightarrow min\ \theta \quad (3)$$

Trong đó:

$\hat{f}(x, \theta)$: Mô hình hồi quy với thông số θ.

1.1. Mô hình hồi quy tuyến tính

Hồi quy tuyến tính là phương pháp thống kê để hồi quy dữ liệu với biến phụ thuộc có giá trị liên tục trong khi các biến độc lập có thể có 1 trong 2 giá trị, liên tục hoặc phân loại. Nói cách khác hồi quy tuyến tính là phương pháp để dự đoán biến phụ thuộc Y dựa trên giá trị của biến độc lập X, có thể được sử dụng cho các trường hợp dự đoán liên tục.

Mô hình hồi quy tuyến tính được biểu diễn như sau:

$$\hat{f}(x, w) = w^T x + w_0 \quad (4)$$

Quá trình huấn luyện mô hình này tìm ra bộ số để độ lớn hàm sai số giữa số liệu tính toán và số liệu thực tế (lost function) là nhỏ nhất và các trọng số $w^T \in R^{n\times 1}$, $w_0 \in R$ sẽ được cập nhật để tìm các thông số của mô hình tối ưu nhất theo công thức sau:

$$[w^T, w_0] = \underset{w^T, w_0}{argmin} \sum_{i=1}^{n} (w^T x_i + w_0 - y_i)^2 + R(w, \alpha) \quad (5)$$

Kỹ thuật chuẩn hóa (regularisation) R(w, α) được thêm vào hàm mất mát để xử lý hiện tượng quá khớp lịch sử, đồng thời có thể giúp giảm hệ số ảnh hưởng của các đầu vào ít có ảnh hưởng đến kết quả dự báo. Có một số loại chuẩn hóa:

$$Lasso\ R(w, \alpha) = \alpha \sum_{j=1}^{d} |w_j| \quad (L_1\ regularisation);$$

$$Ridge\ R(w, \alpha) = \alpha \sum_{j=1}^{d} w_j^2 \quad (L_2\ regularisation);$$

Siêu tham số α liên quan đến số hạng chuẩn hóa L_1/L_2 sẽ được tinh chỉnh bằng cách sử dụng thư viện Scikit-learn Lasso and Ridge cho tập dữ liệu nghiên cứu [1].

1.2. Mô hình mạng trí tuệ nhân tạo (ANN)

ANN là mô hình xử lý thông tin mô phỏng theo cách thức xử lý thông tin của các hệ neural sinh học. ANN được tạo từ số lượng lớn các phần tử (gọi là phần tử xử lý hay neural) kết nối với nhau thông qua các liên kết (gọi là trọng số liên kết) làm việc như 1 thể thống nhất để giải quyết vấn đề cụ thể nào đó.

ANN được cấu hình cho 1 ứng dụng cụ thể (nhận dạng mẫu, phân loại dữ liệu...) thông qua quá trình học từ tập các mẫu huấn luyện. Về bản chất, học chính là quá trình hiệu chỉnh trọng số liên kết giữa các neural.

Mô hình ANN gồm các nút (đơn vị xử lý, neural) được nối với nhau bởi liên kết neural. Mỗi liên kết kèm theo 1 trọng số đặc trưng cho đặc tính kích hoạt/ức chế giữa các neural. Có thể xem các trọng số là phương tiện để lưu thông tin dài hạn trong mạng neural và nhiệm vụ của quá trình đào tạo mạng là cập nhật các trọng số khi có thêm thông tin về các mẫu học [2].

Mô hình ANN có thể được biểu diễn như sau:

$$\hat{f}(x) = \delta_k(w_k \delta_{k-1}(\dots w_2 \delta_1(w_1 x + b_1) + b_2) \dots) + b_k) \quad (6)$$

Trong đó:

δ_i: Hàm kích hoạt;

k: Số lớp;

$w_i \in R^{out_i \times in_i}$): Ma trận trọng số;

b_i: Độ lệch cho lớp thứ i.

Cũng giống như mô hình hồi quy tuyến tính, việc tìm ra bộ số để độ lớn hàm sai số giữa số liệu tính toán và số liệu thực tế là nhỏ nhất sẽ được thực hiện để tái lặp lại lịch sử bằng cách sử dụng thuật toán giảm dần tốc độ (stochastic gradient descent) hoặc thuật toán Adam [3] như sau:

$$[w_1, b_1, \dots, w_k, b_k] = \underset{w_1, b_1, \dots, w_k, b_k}{argmin} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}(x_i) - y_i\right)^2 \quad (7)$$

1.3. Gradient tree boosting

Boosting được giới thiệu bởi Schapire [4] sử dụng thuật toán cây quyết định để tạo các mô hình mới. Boosting gán trọng số cho các mô hình dựa trên hiệu suất của chúng. Có nhiều biến thể của thuật toán boosting như LogitBoost (LB) và AdaBoost (AB). Schapire đã chứng minh rằng các kỹ thuật học yếu (tốt hơn không nhiều so với đoán ngẫu nhiên) có thể được kết hợp với mục đích tạo ra 1 cụm hoặc 1 nhóm các kỹ thuật máy học yếu từ đó tạo thành 1 mô hình mạnh duy nhất. Đối với bài toán phân loại, các hàm cơ bản là các hàm phân loại riêng lẻ $G_m(x) \in \{-1,1\}$. Phần mở rộng này có được biểu diễn trong phương trình sau cho bài toán hồi quy:

$$\hat{f}(x) = \sum_{m=1}^{M} \beta_m b(x; \gamma_m) \quad (8)$$

Trong đó:

β_m, m = 1,2,3...M: Các hệ số mở rộng;

$b(x; \gamma_m) \in R$: Các hàm đơn giản của các đối số đa biến x được đặc trưng bởi thông số γ;

Thông thường, các mô hình này khớp lịch sử bằng cách tìm ra bộ số để độ lớn hàm sai số giữa số liệu tính toán và số liệu thực tế là nhỏ nhất trên tập đào tạo:

$$[\beta_1, \gamma_1, \dots, \beta_M, \gamma_M] = \underset{\beta_1, \gamma_1, \dots, \beta_M, \gamma_M}{argmin} \frac{1}{n} \sum_{i=1}^{n} L\left(y_i, \sum_{m=1}^{M} \beta_m b(x; \gamma_m)\right) \quad (9)$$

Trong thư viện Skit-learn [5], có một số thuật toán gradient boosting gồm: AdaBoost, gradient tree boosting, ExtraTree, Radom forest được phân loại là học theo nhóm. Ngoài ra, XGBoost là 1 trong những thuật toán học máy phổ biến và được sử dụng rộng rãi nhất hiện nay vì giúp xử lý nhanh, tiết kiệm thời gian cũng như chi phí tính toán [6]. XGBoost tương tự như gradient boost nhưng có một số tính năng bổ sung mạnh hơn đáng kể gồm:

- Sử dụng sự thu nhỏ theo tỷ lệ của các nút lá (cắt tỉa) để cải thiện tính tổng quát của mô hình;

- Newton boosting tạo ra 1 tuyến đường trực tiếp đến cực tiểu, thay vì giảm độ dốc;

- Bổ sung tham số ngẫu nhiên giúp giảm mối tương quan giữa các cây để cải thiện sức mạnh của nhóm.

3. Dữ liệu khai thác đối tượng móng, vòm Trung tâm, mỏ Bạch Hổ

Thân dầu móng khu vực vòm Trung tâm bắt đầu đưa vào khai thác năm 1988 với áp suất vỉa ban đầu tại độ sâu tuyệt đối 3.650 mTVDSS đạt 417 at. Ở giai đoạn đầu khai thác áp suất vỉa suy giảm mạnh, cơ chế suy giảm năng lượng tự nhiên và đàn hồi ảnh hưởng chính đến thân dầu khai thác. Do vậy, giải pháp bơm ép nước được áp dụng khi áp suất vỉa trung bình đạt 280 at tại phần đáy của thân dầu nhằm duy trì năng lượng và gia tăng hiệu quả thu hồi. Từ năm 1995, sau 2 năm bơm ép, tốc độ suy giảm áp suất vỉa dần dần ổn định. Tại thời điểm cuối tháng 8/2022, quỹ giếng khai thác của thân dầu đá móng khu vực vòm Trung tâm có 48 giếng đang khai thác (13 giếng tự phun và 35 giếng gaslift) và 8 giếng bơm ép hoạt động (921, 817B, 7004, 424, 437, 8, 448, 8014). Sản lượng dầu cộng dồn đạt 174,413 triệu tấn, độ ngập nước 55%. Sản lượng bơm ép cộng dồn đạt 274,1 triệu m^3. Tổng sản lượng khai thác dầu và lưu lượng bơm ép của các giếng bơm ép được thể hiện trong Hình 1.

4. Tiền xử lý dữ liệu

Các giếng khai thác và bơm ép tại thân dầu đá móng nằm chủ yếu ở khối Trung tâm, với 48/57 giếng khai thác và 8/10 giếng bơm ép (Hình 1). Do giếng bơm ép 8014 mới hoạt động trở lại từ tháng 9/2021 với lưu lượng trung bình 154 m^3/ngày nên sẽ không được sử dụng làm đầu vào mô hình học máy. Như vậy, chỉ có 7 giếng bơm ép sẽ được sử dụng làm dữ liệu đầu vào mô hình học máy.

Hình 1. *Tổng sản lượng khai thác và lưu lượng bơm ép đối tượng móng khu vực vòm Trung tâm*

Để đánh giá mức độ tương quan tuyến tính giữa các cặp giếng bơm ép và giữa các giếng bơm ép với tổng sản lượng dầu, nhóm tác giả đã sử dụng hệ số tương quan Pearson. Hình 2 cho thấy có 4 giếng bơm ép có tương quan nghịch với tổng sản lượng dầu và 3 giếng có tương quan thuận với tổng sản lượng dầu. Về mặt thống kê, có thể giảm dần lưu lượng bơm ép từ 4 giếng có tương quan nghịch và tăng lưu lượng bơm ép từ 3 giếng có tương quan thuận để đánh giá hiệu quả kế hoạch bơm ép cũng như gia tăng tổng sản lượng dầu khai thác. Ngoài ra, hệ số tương quan giữa các giếng bơm ép này không quá cao, do đó nhóm tác giả lựa chọn lưu lượng bơm ép từ 7 giếng bơm ép này làm dữ liệu đầu vào để dự báo tổng sản lượng dầu.

Nhóm tác giả chia dữ liệu thành 2 tập:

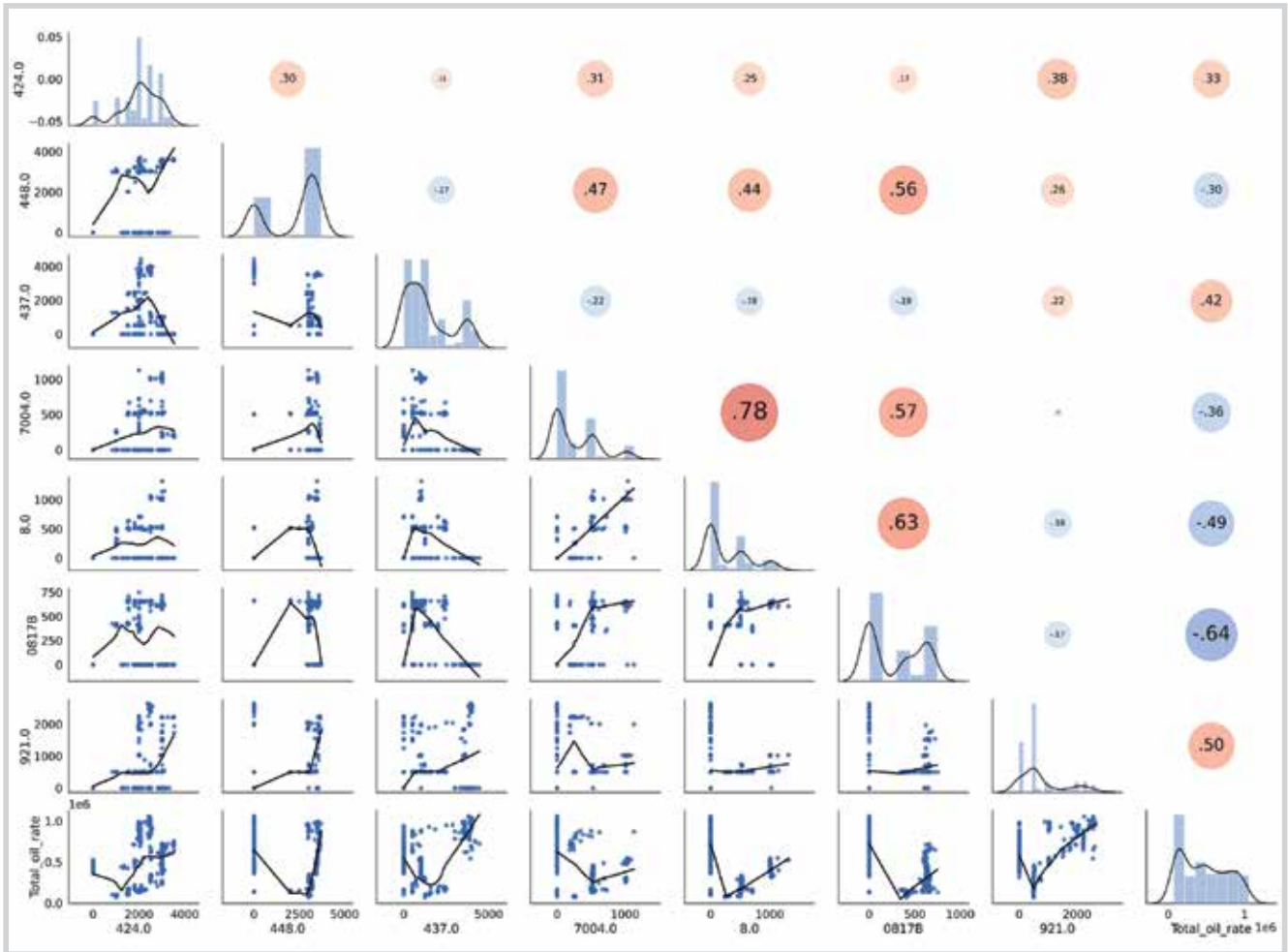
- Tập dữ liệu I sử dụng 339 tháng dữ liệu (từ tháng 6/1993 - 8/2021) được xáo trộn và chọn ngẫu nhiên để xây dựng mô hình cấu trúc và xác định thuật toán tối ưu;

- Tập dữ liệu II sử dụng 12 tháng dữ liệu lưu lượng bơm ép nước sẽ được lấy theo kế hoạch bơm ép của Liên doanh Việt - Nga "Vietsovpetro" (từ tháng 9/2021 - 8/2022) với) để dự báo sản lượng khai thác dầu.

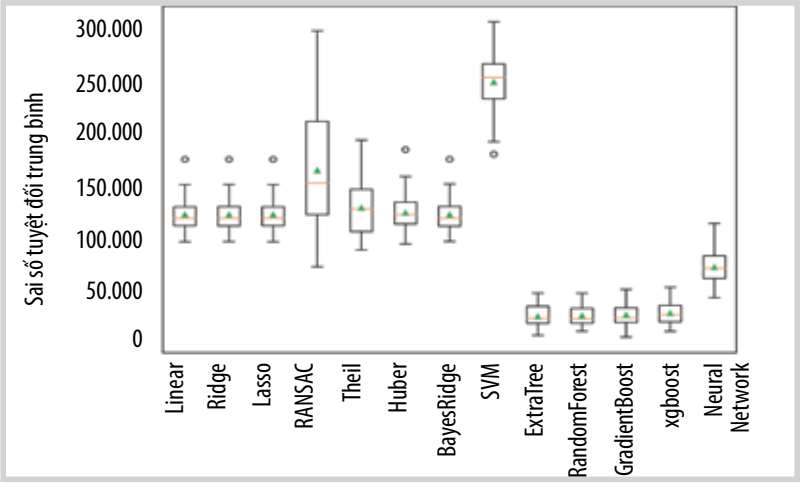
Để nâng cao chất lượng của mô hình học máy và tránh hiện tượng mô hình quá khớp hoặc không khớp lịch sử trên tập dữ liệu huấn luyện thì tập dữ liệu I được phân chia làm 2 giai đoạn: đào tạo và kiểm tra. Tập đào tạo sử dụng 271 tháng dữ liệu ngẫu nhiên (80%). Tập kiểm tra sử dụng 68 tháng dữ liệu (20%) để kiểm tra chất lượng mô hình trong quá trình huấn luyện.

5. Lựa chọn và tinh chỉnh mô hình

Để đánh giá lựa chọn mô hình tối ưu nhất trên tập dữ liệu nghiên cứu, nhóm tác giả thử nghiệm 10 thuật toán học máy gồm: Random forest, ExtraTree, XGBoost, GradientBoosting, Neural Network, BayesRidge, Ridge, Linear, Lasso, Huber. Sử dụng các thông số mặc định trên thư viện Scikit-learning [6] và chỉ số đánh giá sai số tuyệt đối trung bình (MAE) để đánh giá lựa chọn các thông số



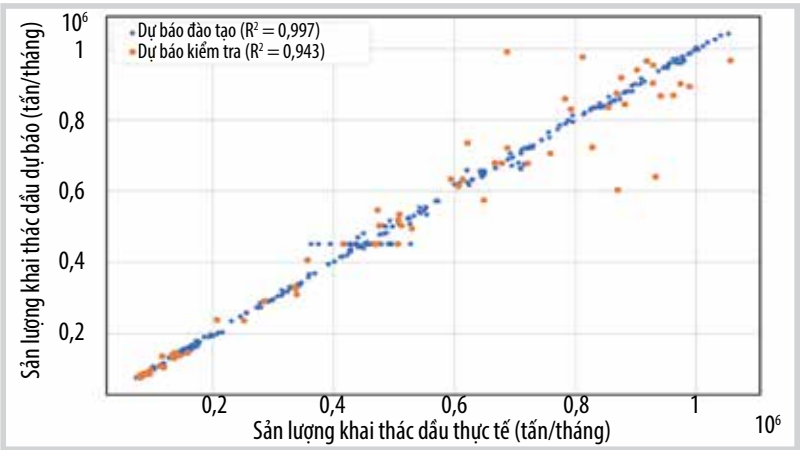
Hình 2. *Hệ số tương quan Pearson giữa các cặp sản lượng khai thác dầu và lưu lượng bơm ép nước đối tượng móng, khu vực vòm Trung tâm, mỏ Bạch Hổ.*



Hình 3. *Sai số tuyệt đối trung bình của các thuật toán học máy.*

Bảng 1. *Kết quả tối ưu các thông số của mô hình RF*

Các thông số	Khoảng giá trị	Kết quả tối ưu
Độ sâu tối đa (Max depth)	2 - 7	6
Số lượng mẫu tối thiểu (Min sample split)	2 - 10	7
Số lượng cây quyết định (Number of estimator)	10 - 120	107
Trình tạo ngẫu nhiên (Random state)	0 - 10	0



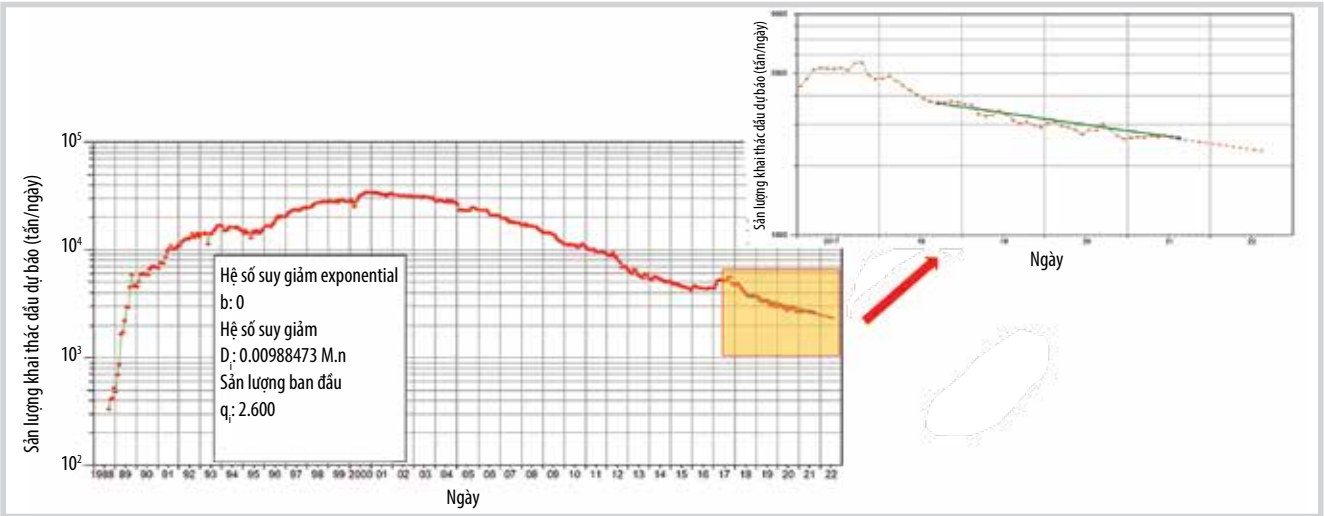
Hình 4. *Hệ số tương quan R^2 giữa kết quả dự báo sản lượng dầu khai thác và sản lượng dầu khai thác thực tế trên tập đào tạo và tập kiểm tra.*

và mô hình tốt nhất bằng phương pháp GridSearch. Kết quả đánh giá sai số trung bình tuyệt đối cho mỗi thuật toán được thể hiện trong Hình 3, cho thấy mô hình vector máy hỗ trợ cho sai số cao nhất, sau đó lần lượt là nhóm hồi quy tuyến tính, nhóm ANN và nhóm học kết hợp. Trong nhóm học kết hợp, thuật toán RF có sai số tuyệt đối trung bình nhỏ nhất sẽ được chọn để tinh chỉnh thêm và tối ưu các thông số của mô hình.

Các thông số được tối ưu của thuật toán RF được tóm tắt trong Bảng 1. Hệ số tương quan R^2 trên tập đào tạo và tập kiểm tra tương đối cao lần lượt là 0,997 và 0,943 (Hình 4). Từ đó, cho thấy mô hình sử dụng thuật toán RF có độ tin cậy cao có thể sử dụng để dự báo sản lượng khai thác dầu trên tập dữ liệu II.

6. Dự báo sản lượng khai thác dầu sử dụng phương pháp phân tích đường cong suy giảm và mô hình RF

Nhóm tác giả sử dụng mô hình RF và phương pháp phân tích đường cong suy giảm (DCA) để dự báo sản lượng khai thác dầu cho 12 tháng (từ tháng 9/2021 - 8/2022) của thân dầu móng khu vực vòm Trung tâm, mỏ Bạch Hổ.



Hình 5. *Dự báo sản lượng khai thác dầu sử dụng phương pháp DCA.*

6.1. Kết quả dự báo khai thác sử dụng phương pháp DCA

Phương pháp phân tích đường cong suy giảm được sử dụng để dự báo sản lượng khai thác dầu cho thân dầu móng khu vực vòm Trung tâm, mỏ Bạch Hổ với lịch sử khai thác từ giữa năm 2018 đến hết tháng 8/2021. Kết quả dự báo cho thấy khu vực Trung tâm có hệ số suy giảm tương đối thấp (0,0098 M.n), kết quả dự báo được thể hiện trong Bảng 2 và Hình 5.

6.2. Kết quả dự báo khai thác sử dụng mô hình RF

Nhóm tác giả sử dụng mô hình RF để dự báo sản lượng khai thác dầu trên tập dữ liệu II từ tháng 9/2021 - 8/2022 thể hiện trong Hình 6 và Bảng 3.

7. Đánh giá, so sánh kết quả dự báo khai thác sử dụng phương pháp phân tích đường cong suy giảm và mô hình RF

7.1. Đánh giá, so sánh kết quả dự báo khai thác từ 9/2021 đến 8/2022

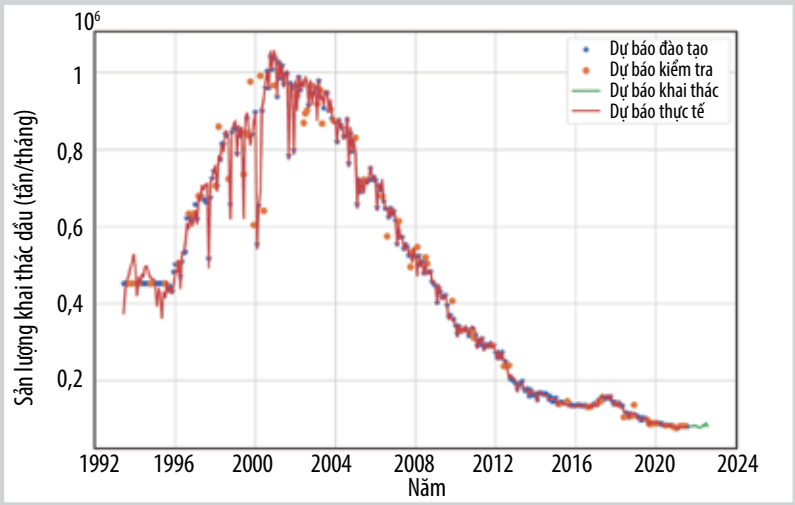
Kết quả dự báo khai thác sử dụng phương pháp DCA và mô hình RF sẽ được so sánh với dữ liệu khai thác thực tế được thể hiện trong Hình 6 và Bảng 4. Bảng 4 cho thấy các kết quả đánh giá sai số tuyệt đối (AE) và sai số tương đối trung bình (ARE) của mô hình RF lần lượt là 2,9 nghìn tấn và 4%. Các kết quả sai số này được đánh giá là tương đối thấp và dưới giới hạn cho phép. Điều này chứng tỏ kết quả dự báo khai thác đã phản ánh được ảnh hưởng của các thông số vận hành như sản lượng khai thác dầu và sản lượng bơm ép. Trong khi đó, dự báo khai thác sử dụng phương pháp DCA không phản ánh đúng xu hướng và cho sai số tương đối trung bình cao (ARE ~ 9%).

7.2. Đánh giá, so sánh kết quả dự báo khai thác theo giai đoạn

Nhóm tác giả chia kết quả dự báo khai thác thành 2 giai đoạn để đánh giá chi tiết hiệu suất dự báo sản lượng khai thác của các mô hình theo chỉ tiêu đánh giá sai số

Bảng 2. *Dự báo sản lượng khai thác dầu sử dụng phương pháp DCA*

Tháng	9/21	10/21	11/21	12/21	1/22	2/22	3/22	4/22	5/22	6/22	7/22	8/22
Sản lượng dầu (nghìn tấn)	74	76	73	75	74	67	73	70	72	69	71	70



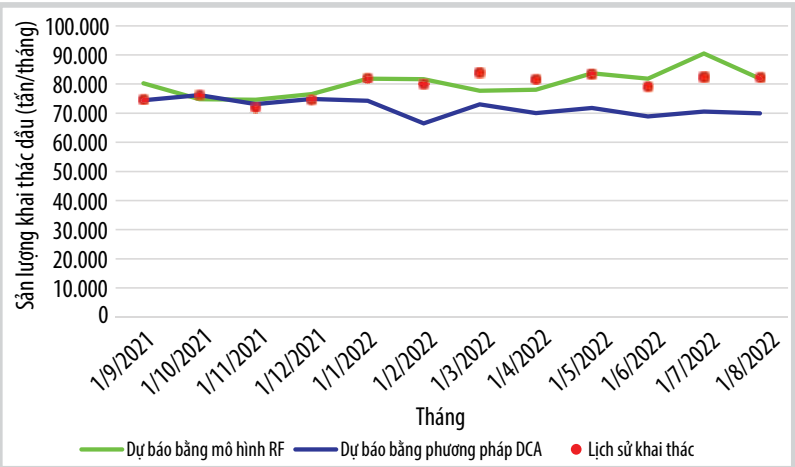
Hình 6. *Kết quả dự báo sản lượng dầu khai thác sử dụng mô hình RF.*

Bảng 3. *Kết quả dự báo sản lượng dầu khai thác sử dụng mô hình RF*

Tháng	9/21	10/21	11/21	12/21	1/22	2/22	3/22	4/22	5/22	6/22	7/22	8/22
Sản lượng dầu (nghìn tấn)	80	75	75	77	82	82	78	78	84	82	90	82

Bảng 4. *Bảng thống kê đánh giá sai số của các phương pháp dự báo khai thác*

Dự báo khai thác	AE_RF (tấn)	ARE_RF (%)	AE_DCA (tấn)	ARE_DCA (%)
Trung bình	2.903	4	7.620	9
Độ lệch chuẩn	2.502	3	5.455	7
Nhỏ nhất	101	0	26	0
Lớn nhất	8.043	10	13.443	17



Hình 7. *Kết quả dự báo sản lượng khai thác dầu.*

Bảng 5. *Bảng thống kê đánh giá sai số của các phương pháp dự báo khai thác trong giai đoạn I*

Giai đoạn I	AE_RF (tấn)	ARE_RF (%)	AE_DCA (tấn)	ARE_DCA (%)
Trung bình	2.875	4	480	1
Độ lệch chuẩn	1.814	2	489,54	1
Nhỏ nhất	1.398	2	26	0
Lớn nhất	5.488	7	1.175	2

Bảng 6. *Bảng thống kê đánh giá sai số của các phương pháp dự báo khai thác trong giai đoạn II*

Giai đoạn II	AE_RF (tấn)	ARE_RF (%)	AE_DCA (tấn)	ARE_DCA (%)
Trung bình	2.916	4	11.191	14
Độ lệch chuẩn	2.903	3	1.717	2
Nhỏ nhất	101	0	7.697	9
Lớn nhất	8.043	10	13.443	17

tuyệt đối (AE) và sai số tương đối trung bình (ARE) được thể hiện trong Bảng 5 và Bảng 6.

- Trong giai đoạn I (từ tháng 9 - 12/2021), kết quả dự báo khai thác bằng phương pháp DCA phản ánh đúng xu hướng và sai số tương đối trung bình thấp (1%), trong khi đó kết quả dự báo khai thác sử dụng mô hình RF sai số tương đối trung bình 4% (Bảng 5). Điều này chứng tỏ, khi khu vực nghiên cứu có chế độ khai thác ổn định, không thực hiện các giải pháp địa kỹ thuật hay can thiệp giếng thì sử dụng phương pháp DCA để dự báo khai thác có độ tin cậy cao.

- Trong giai đoạn II (từ tháng 12/2021 - 8/2022), kết quả dự báo khai thác sử dụng mô hình RF phản ánh đúng xu hướng và sai số tương đối trung bình thấp (4%), trong khi đó kết quả dự báo khai thác bằng phương pháp DCA cho xu hướng không tốt, sai số tương đối trung bình cao (14%) (Bảng 6). Trong giai đoạn này, Vietsovpetro có thực hiện các giải pháp địa kỹ thuật (GTM) và thêm giếng mới giúp sản lượng dầu gia tăng 8 nghìn tấn từ tháng 12/2021 - 1/2022. Điều đó dẫn đến kết quả dự báo khai thác bằng phương pháp DCA trong giai đoạn này sẽ không phản ánh đúng xu hướng, đây cũng là hạn chế lớn nhất của phương pháp DCA.

8. Kết luận

Nghiên cứu cung cấp phương pháp dự báo khai thác mới trên tập dữ liệu lịch sử khai thác và cho thấy khả năng tổng quát hóa bài toán dự báo trên mô hình học máy trở thành công cụ hữu hiệu để có thể giải quyết hiệu quả nhiều bài toán khác nhau trong kỹ thuật khai thác mỏ.

Kết quả nghiên cứu cho thấy khả năng dự báo khai thác sử dụng mô hình RF cho độ chính xác cao với sai số 4% so với phương pháp dự báo truyền thống. Bên cạnh đó, dự báo bằng mô hình RF không phụ thuộc vào kinh nghiệm chủ quan của chuyên gia dự báo và cho kết quả có độ tin cậy cao ngay cả khi khu vực nghiên cứu trong tương lai có thực hiện các phương pháp GTM hay thêm giếng mới để gia tăng sản lượng khai thác dầu.

Lời cảm ơn

Nhóm tác giả trân trọng cảm ơn Viện Dầu khí Việt Nam (VPI) đã hỗ trợ nguồn lực và tài trợ kinh phí thực hiện nghiên cứu theo Quyết định giao nhiệm vụ số 5186/QĐ-VDKVN ngày 30/9/2021.

Tài liệu tham khảo

[1] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay, "Scikit-learn: Machine learning in Python", Journal of Machine Learning Research, Vol. 12, pp. 2825 - 2830, 2011.

[2] Tran Dang Tu, Nguyen The Duc, Le Quang Duyen, Pham Truong Giang, Le Vu Quan, Le Quoc Trung, Tran Xuan Quy, and Pham Chi Duc "An applied machine learning approach to production forecast for basement formation - Bach Ho field", *Petrovietnam Journal*, Vol. 6, pp. 48 - 57, 2019.

[3] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization", *3rd International Conference for Learning Representations, San Diego, USA, 7 - 9 May 2015*. DOI: 10.48550/arXiv.1412.6980.

[4] Robert E Schapire, "The strength of weak learnability", *Machine Learning*, Vol. 5, pp. 197 - 227, 1990. DOI: 10.1007/BF00116037.

[5] Oliver Kramer, "Scikit-learn", *Machine learning for evolution strategies*. Springer, 2016, pp. 45 - 53. DOI: 10.1007/978-3-319-33383-0.

[6] Tianqi Chen and Carlos Guestrin, "XGBoost: A scalable tree boosting system", *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 13 - 17 August 2016*. DOI: 10.1145/2939672.2939785.

APPLICATION OF MACHINE LEARNING ALGORITHM TO FORECAST PRODUCTION FOR FRACTURE BASEMENT FORMATION, CENTRAL ARCH, BACH HO FIELD

Tran Dang Tu, Le The Hung, Tran Xuan Quy, Doan Huy Hien, Pham Truong Giang, Luu Dinh Tung
Vietnam Petroleum Institute
Email: tutd@vpi.pvn.vn

Summary

Oil production forecast is a big challenge in the oil and gas industry. Simulation model and prediction results play an important role in field operation and management. Currently, dynamic simulation model, decline curve analysis are popular tools applied to forecast production. The dynamic simulation model shows a remarkable effect for sedimentary objects. However, production forecasting by this method for fracture basement formation sometimes gives unreliable results because the fracture basement formation is a complex of geological objects, which causes difficulties in predicting the geological characteristics. The decline curve analysis (DCA) method uses simple extrapolated mathematical functions to forecast oil production, therefore the results do not reflect the production operations such as opening/closing production interval.

To avoid the disadvantages of these traditional methods, Vietnam Petroleum Institute (VPI) has studied the applicability of machine learning to forecast oil production for fracture basement formation of Bach Ho field. The study results show that the random forest model has improved the production forecast with low relative error (4%).

Key words: Machine learning, random forest model, production forecast, fracture basement, Bach Ho field.