

ỨNG DỤNG MÔ HÌNH RANDOM FOREST ĐỂ DỰ BÁO GIÁ QUẶNG SẮT THẾ GIỚI

● VŨ DIỆP ANH

TÓM TẮT:

Quặng sắt là một nguồn tài nguyên có giá trị đóng vai trò quan trọng đối với sự phát triển kinh tế, nhất là công nghiệp, của mọi quốc gia. Sự biến động giá quặng sắt có thể ảnh hưởng đến các ngành công nghiệp có liên quan như ngành thép, xây dựng, ô tô, tàu biển,... Do đó, việc dự báo chính xác giá quặng sắt thế giới đã thu hút được sự quan tâm của các doanh nghiệp, nhà đầu tư và các nhà khoa học trong các lĩnh vực liên quan đến sắt thép. Bài báo này ứng dụng mô hình trí tuệ nhân tạo Random Forest để dự báo giá quặng sắt thế giới trên cơ sở bộ dữ liệu theo tháng của các nhân tố ảnh hưởng tới giá quặng sắt bao gồm chỉ số giá sắt phế liệu, chỉ số sản xuất thép, xuất khẩu quặng sắt của Úc, cước phí vận tải đường biển, tỷ giá của đồng đôla Úc và của đồng nhân dân tệ, đại dịch Covid-19 và cơ chế thương lượng giá quặng sắt trong giai đoạn từ tháng 3/1990 đến tháng 2/2022. Kết quả nghiên cứu cho thấy mô hình Random Forest có thể dự báo tốt giá quặng sắt thế giới. Kết quả dự báo từ tập huấn luyện và tập kiểm tra bám sát giá thực tế và có sai số dự báo ở mức thấp, với sai số trung bình tuyệt đối lần lượt là $MAE = 1,329$ và $8,78$; và căn bậc hai trung bình bình phương sai số dự báo tương ứng là $RMSE = 2,848$ và $13,65$.

Từ khóa: quặng sắt, giá quặng sắt thế giới, dự báo, mô hình Random Forest.

1. Đặt vấn đề

Trí tuệ nhân tạo (artificial intelligence - AI) là một lĩnh vực của khoa học và công nghệ nhằm làm cho máy có những khả năng của trí tuệ và trí thông minh của con người, tiêu biểu như biết suy nghĩ và lập luận để giải quyết vấn đề, biết giao tiếp do hiểu ngôn ngữ và tiếng nói, biết học và tự thích nghi,... Trí tuệ nhân tạo ra đời từ năm 1956 tại hội nghị mùa hè tại trường Dartmouth, Mỹ và không ngừng phát triển trong hơn 60 năm qua. Ngày nay, trí tuệ nhân tạo được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau trong đời sống như giáo dục, y tế, sản xuất, tài chính, kinh doanh, các thiết bị như tivi, điện thoại,... Các mô hình trí tuệ nhân tạo được phát triển khá nhanh và được nhiều

nhà nghiên cứu đề xuất sử dụng trong việc dự báo các hiện tượng trong nhiều lĩnh vực khác nhau như sinh học, y học, sinh thái, thiên văn, tự động hóa, thời tiết, ngân hàng, chứng khoán, giá cả hàng hóa,... bởi các mô hình này có thể xử lý được lượng lớn các dữ liệu nhanh hơn nhiều và đưa ra các dự đoán chính xác hơn khả năng của con người. Các mô hình này có thể phân tích "hành vi" của dữ liệu mà không cần có những giả định liên quan đến phân bố thống kê của dữ liệu như trong các mô hình kinh tế lượng hay chuỗi dừng trong các mô hình tự hồi quy theo chuỗi thời gian. Các mô hình trí tuệ nhân tạo có ưu điểm hơn so với các mô hình kinh tế lượng do trí tuệ nhân tạo cho phép xử lý dữ liệu có xu hướng phức tạp, các mối quan hệ

phi tuyến tính. Hơn thế nữa, các mô hình trí tuệ nhân tạo cho các kết quả dự báo chính xác hơn. Kết quả nghiên cứu của Lasheras và cộng sự, Ramyar và Kianfar, Kristjanpoller và Hernandez, Dehghani, Alameer và cộng sự cho thấy các mô hình trí tuệ mạng nơ-ron nhân tạo cho kết quả dự báo chính xác hơn các mô hình kinh tế lượng truyền thống như mô hình chuỗi thời gian ARIMA, mô hình vectơ tự hồi quy (VAR), mô hình kinh tế lượng hợp phần GARCH [10, 14, 9, 4, 1].

Để dự báo giá quặng sắt thế giới, các nhà nghiên cứu hiện nay chủ yếu sử dụng các phương pháp dự báo truyền thống. Zhu đã sử dụng phương pháp mô phỏng Monte Carlo để dự báo giá quặng sắt và kết luận rằng kết quả dự báo nhất quán với dữ liệu giá lịch sử và tốc độ tăng trưởng GDP thế giới là nhân tố quyết định tới việc dự báo giá quặng sắt [18]. Pustov và cộng sự sử dụng lý thuyết chi phí biên và giá khuyến khích để thiết lập mô hình nghiên cứu giá quặng sắt thực trong dài hạn. Theo đó, giá quặng sắt được dự báo nằm trong khoảng 85-125\$ và tăng lên 150-220\$ vào năm 2020 [13]. Sử dụng mô hình EGARCH để phân tích, Ma chỉ ra rằng cơ chế thương lượng giá mới làm giảm sự biến động của giá quặng sắt thế giới [11]. Với mô hình VECM và các dữ liệu theo tháng từ 1/2003-8/2012 và giai đoạn 1/2003 - 6/2017 của các biến độc lập GDP của Trung Quốc, chi phí vận tải và biến giá quặng sắt với độ trễ một năm, Warell chứng minh rằng việc thay đổi chính sách giá quặng sắt không có ảnh hưởng tới giá quặng sắt và tăng trưởng GDP của Trung Quốc có ảnh hưởng mạnh nhất tới giá quặng sắt trong ngắn hạn [16, 17]. Tương tự, Haque cũng sử dụng mô hình VECM để phân tích và chỉ ra rằng tỷ giá hối đoái AUD/USD không tác động tới giá quặng sắt [7]. Tuy nhiên, kết quả phân tích khi sử dụng phương pháp hồi quy SVAR lại cho thấy giá quặng sắt có phản ứng mạnh với các cú sốc trong tỷ giá AUD/USD.

Tổng quan nghiên cứu cho thấy tới nay ảnh hưởng của đại dịch Covid-19 và các nhân tố liên quan đến cung, cầu quặng sắt, vốn là những thông tin quan trọng ảnh hưởng đến giá quặng sắt, chưa được sử dụng trong các mô hình dự báo giá quặng sắt. Các công trình nghiên cứu chủ yếu dựa vào dữ

liệu theo năm của các biến số để dự báo giá quặng sắt và các dữ liệu này chưa được cập nhật đến năm 2022. Để bổ sung cho các nghiên cứu về giá quặng sắt thế giới, bài báo này sử dụng mô hình trí tuệ nhân tạo Random Forest đa biến, với các biến độc lập là chỉ số giá sắt phế liệu, chỉ số sản xuất thép, xuất khẩu quặng sắt của Úc, cước phí vận tải đường biển, tỷ giá của đồng đôla Úc và của đồng nhân dân tệ, đại dịch Covid-19 và cơ chế thương lượng giá quặng sắt, để dự báo giá quặng sắt dựa trên các số liệu được thu thập theo tháng từ tháng 3/1990 đến tháng 2/2022.

2. Phương pháp nghiên cứu

2.1. Mô hình Random Forest

Mô hình Random Forest, hay còn gọi là rừng ngẫu nhiên, là phương pháp phân lớp thuộc tính, một thuật toán trong phương pháp học máy, sử dụng nhiều cây phân loại hoặc hồi quy trong một nhóm được phát triển bởi Leo Breiman [2, 3]. Mô hình này sử dụng kỹ thuật đóng gói (bagging) cho phép lựa chọn một nhóm nhỏ các thuộc tính tại mỗi nút (node) của cây phân lớp để phân chia thành các mức tiếp theo. Cụ thể, đây là thuật toán phân lớp bao gồm một tập các phân lớp có cấu trúc cây quyết định với những vectơ độc lập, tương tự nhau được phân bố một cách ngẫu nhiên và mỗi cây sẽ “bỏ một phiếu bầu” cho lớp phổ biến nhất ở vectơ đầu vào, làm cơ sở ra quyết định cho thuật toán. Các phương pháp học nhóm kết hợp với các kết quả riêng lẻ của từng cây thường mang lại kết quả tốt hơn.

Mô tả thuật toán Random Forest:

- Chọn T là số lượng các cây thành phần sẽ được xây dựng.

- Chọn m là số lượng các thuộc tính được dùng để phân chia tại mỗi nút (node) của cây. p là tổng số các thuộc tính. m thường nhỏ hơn p rất nhiều. Giá trị m được giữ không đổi trong suốt quá trình xây dựng cây.

- Dựng T cây quyết định. Trong đó mỗi cây quyết định được hình thành như sau: (a) xây dựng tập mẫu khởi động (bootstrap) với n mẫu, hình thành từ việc hoán vị tập các mẫu ban đầu. Mỗi cây sẽ được dựng từ tập khởi động này; (b) Khi xây dựng cây quyết định, tại mỗi nút (node) sẽ

chọn ra m thuộc tính và sử dụng m thuộc tính này để tìm ra cách phân chia tốt nhất; (c) Mỗi cây quyết định được phát triển lớn nhất có thể và không bị cắt xén.

- Sau khi xây dựng được rừng ngẫu nhiên (Random Forest), để phân lớp cho đối tượng T, thu thập kết quả phân lớp đối tượng này trên tất cả các cây quyết định và tính bình quân giá trị dự báo của các cây quyết định để làm kết quả cuối cùng của thuật toán. Tỷ lệ lỗi của cây phụ thuộc vào độ mạnh của từng cây quyết định thành phần và mối quan hệ qua lại giữa các cây đó.

2.2. Các biến sử dụng trong mô hình

Theo quan điểm của các nhà kinh tế học hiện đại, giá cả hàng hóa chịu ảnh hưởng của nhiều nhân tố khác nhau như cung, cầu về hàng hóa đó, tác động của các chính sách kinh tế có liên quan, tỷ giá hối đoái, cước phí vận tải, điều kiện tự nhiên, dịch bệnh,... [12]. Dựa trên cơ sở lý luận về sự hình thành giá cả, các nhân tố ảnh hưởng tới giá cả, tổng quan nghiên cứu về các nhân tố ảnh hưởng tới giá cả và căn cứ vào khả năng thu thập dữ liệu cần thiết phục vụ cho nghiên cứu định lượng, bài báo sử dụng mô hình Random Forest để dự báo giá quặng sắt thế giới trên cơ sở các nhân tố ảnh hưởng tới giá quặng sắt bao gồm chỉ số giá sắt phế liệu, chỉ số sản xuất thép, xuất khẩu quặng sắt của Úc, cước phí vận tải đường biển, tỷ giá của đồng đôla Úc và của đồng nhân dân tệ, đại dịch Covid-19 và cơ chế thương lượng giá quặng sắt.

Trong đó:

- Biến phụ thuộc là giá quặng sắt thế giới (IO).

- Biến độc lập:

+ Chỉ số giá sắt phế liệu (SI) được coi là nhân tố ảnh hưởng tới cầu về quặng sắt do sắt phế liệu là hàng hóa thay thế cho quặng sắt trong các ngành công nghiệp có sử dụng quặng sắt. Khi chỉ số giá sắt phế liệu tăng lên, cầu về sắt phế liệu sẽ giảm xuống và cầu về quặng sắt sẽ tăng lên, kéo theo sự gia tăng của giá quặng sắt.

+ Chỉ số sản xuất thép (SPI) được coi là nhân tố ảnh hưởng tới cầu về quặng sắt do 98% quặng sắt khai thác ra được sử dụng để sản xuất thép. Nếu chỉ số sản xuất thép giảm xuống thì cầu đối với quặng sắt cũng giảm theo, làm giảm giá quặng sắt.

+ Xuất khẩu quặng sắt của Úc (EX) là biến đại diện cho cung quặng sắt thế giới. Với vai trò là nước xuất khẩu quặng sắt lớn nhất trên thế giới, chiếm 53,6% tổng kim ngạch xuất khẩu quặng sắt trong năm 2021, sự biến động trong xuất khẩu quặng sắt của Úc sẽ dẫn đến sự thay đổi cung quặng sắt thế giới. Theo lý thuyết kinh tế, khi xuất khẩu quặng sắt của Úc giảm xuống, giá quặng sắt thế giới sẽ tăng lên.

+ Cước phí vận tải đường biển (FR): Giá quặng sắt chịu ảnh hưởng của cước phí vận tải đường biển vì quặng sắt được xuất nhập khẩu bằng đường biển. Cước phí vận tải rẻ hơn sẽ kéo giá quặng sắt giảm xuống.

+ Tỷ giá của đồng đôla Úc (ERa) và tỷ giá hối đoái của đồng nhân dân tệ (ERc): Tỷ giá này được hiểu là 1 đồng đôla Úc hoặc 1 đồng nhân dân tệ có thể đổi được bao nhiêu đôla Mỹ. Là nước xuất khẩu quặng sắt nhiều nhất trên thế giới nên sự biến động về giá trị của đồng đôla Úc có thể kéo theo sự thay đổi giá quặng sắt thế giới. Tỷ giá giảm xuống đồng nghĩa với việc đồng đôla Úc yếu đi, khi đó giá quặng sắt xuất khẩu của Úc cũng như giá quặng sắt thế giới sẽ trở nên rẻ hơn và ngược lại. Là nước nhập khẩu quặng sắt lớn nhất thế giới, chiếm 70,1% tổng kim ngạch nhập khẩu quặng sắt toàn cầu vào năm 2020, sự biến động của đồng nhân dân tệ Trung Quốc sẽ có những ảnh hưởng đến giá quặng sắt thế giới khi đồng tiền này được sử dụng trong các hợp đồng nhập khẩu quặng sắt.

+ Đại dịch Covid-19 (Covid) bùng phát toàn cầu trong 2 năm qua cùng các biện pháp giãn cách xã hội trong từng quốc gia và giữa các quốc gia trên thế giới đã ảnh hưởng đến hoạt động sản xuất kinh doanh của các doanh nghiệp trong mọi lĩnh vực của nền kinh tế thế giới, trong đó có các doanh nghiệp liên quan đến quặng sắt. Cùng với sự gián đoạn trong hoạt động vận tải quốc tế, có thể nói đại dịch Covid-19 đã ảnh hưởng đến cung, cầu đối với quặng sắt, và do đó, có thể dẫn đến sự biến động về giá quặng sắt thế giới.

+ Cơ chế thương lượng giá quặng sắt (PN): Trong suốt 40 năm kể từ trước tháng 11/2008, giá quặng sắt thế giới trước được xác định chủ yếu bằng cơ chế thương lượng giữa các nhà sản xuất

thép lớn nhất thế giới và các nhà cung cấp quặng sắt của họ. Theo đó, hàng năm họ sẽ gặp nhau để thương lượng, xác định giá quặng sắt cho cả 1 năm. Tuy nhiên, cơ chế thương lượng giá này đã bị phá bỏ do giá quặng sắt có biến động mạnh kể từ cuối năm 2008. Khi giá quặng sắt trên thị trường giao ngay thấp hơn mức giá thương lượng, một số nhà sản xuất thép lớn không tuân thủ thỏa thuận đã được thương lượng trước đó, yêu cầu được mua quặng sắt với mức giá thấp hơn hoặc chuyển sang mua quặng sắt trên thị trường giao ngay từ một số nhà cung cấp khác và kinh doanh kiếm lời nhờ chênh lệch giá. Đồng thời, các nhà cung cấp quặng sắt hàng đầu thế giới cũng muốn bán quặng sắt trên thị trường giao ngay nhất là khi thị trường quặng sắt và giá quặng sắt liên tục gia tăng trong giai đoạn tiếp đó. Chính vì vậy, việc áp dụng cơ chế thương lượng giá quặng sắt hay không có ảnh hưởng tới giá quặng sắt thế giới.

2.3. Dữ liệu nghiên cứu

Để dự báo giá quặng sắt thế giới, tác giả tiến hành thu thập số liệu theo tháng về các biến trong mô hình cho giai đoạn từ tháng 3/1990 đến tháng 2/2022. Giá quặng sắt thế giới được tham khảo từ Index Mundi [8], chỉ số giá sắt phế liệu, chỉ số sản xuất thép được lấy từ Fred Economic Data [6], xuất khẩu quặng sắt của Úc và cước phí vận tải đường biển được thu thập từ Trading Economics [15], tỷ giá của đồng đôla Úc và của đồng nhân dân tệ được tham khảo từ Currency Converter [19]. Biến giả Covid-19 nhận giá trị bằng 0 đối với những tháng trước tháng 3/2020 và nhận giá trị bằng 1 đối với những tháng sau tháng 3/2020. Biến giả cơ chế thương lượng giá quặng sắt thế giới nhận giá trị bằng 1 đối với những tháng trước tháng 11/2008 và nhận giá trị bằng 0 đối với những tháng sau tháng 11/2008. Mẫu nghiên cứu bao gồm 384 quan sát.

3. Kết quả phân tích

Bài báo sử dụng mô hình Random Forest để dự báo giá quặng sắt thế giới. Kết quả phân tích được thể hiện trong Bảng 1, Hình 1 và Hình 2.

• Hệ số tương quan giữa các biến

Số liệu trong Bảng 1 cho thấy hệ số tương quan giữa các biến độc lập đều nhỏ hơn 0,8. Kết quả này chứng minh giữa các biến độc lập trong mô hình

không có mối quan hệ đa cộng tuyến. Hệ số tương quan giữa các biến độc lập và giá quặng sắt dao động trong khoảng 0,22-0,89, trong đó chỉ số giá sắt phế liệu (SI), xuất khẩu quặng sắt của Úc (EX), tỷ giá của đồng đôla Úc (ERa) và cơ chế thương lượng giá quặng sắt (PN) có quan hệ tương quan mạnh với giá quặng sắt do giá trị tuyệt đối của các hệ số tương quan >0,6.

• Kết quả dự báo giá quặng sắt thế giới theo mô hình Random Forest

Để dự báo giá quặng sắt thế giới theo mô hình Random Forest, bộ dữ liệu thu thập được về các biến trong mô hình được chia thành 2 tập là tập huấn luyện (training dataset), chiếm 80%, được dùng để điều chỉnh các tham số của mô hình; và tập kiểm tra (testing dataset), chiếm 20%, được dùng để đánh giá mô hình. Tập mẫu được rút ra từ tập huấn luyện để xây dựng các cây quyết định. Mô hình được huấn luyện dựa trên tập huấn luyện (training dataset). Sau khi kết thúc quá trình huấn luyện, mô hình sẽ được đánh giá hiệu quả trên cơ sở tập kiểm tra (testing dataset) bằng cách so sánh giá trị dự báo và giá trị thực tế của biến phụ thuộc và các chỉ tiêu đánh giá sai số dự báo.

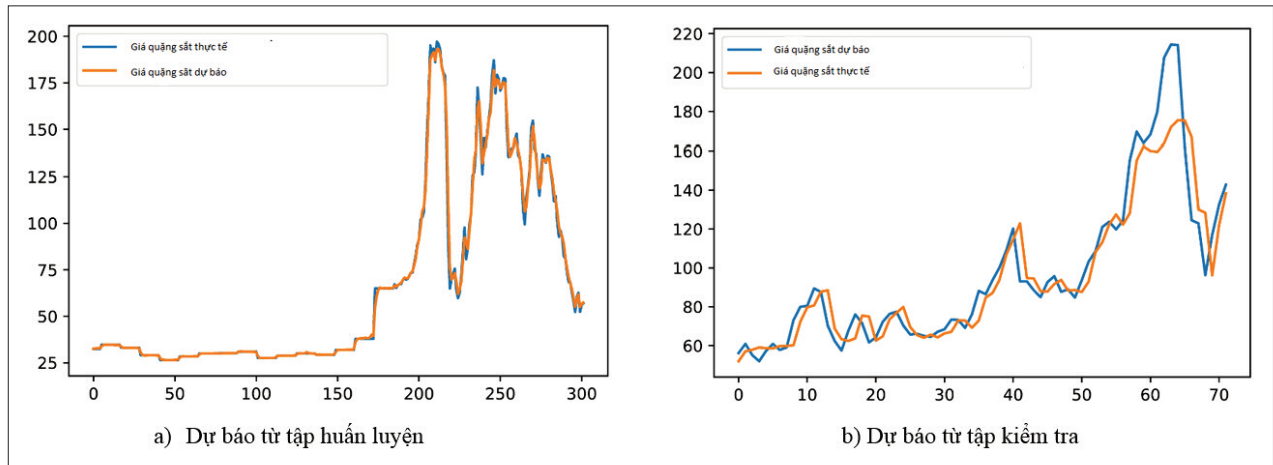
Kết quả dự báo theo mô hình Random Forest được thể hiện trong Hình 1. Theo đó, giá quặng sắt được dự báo theo tập huấn luyện bám rất sát giá quặng sắt thực tế trong giai đoạn nghiên cứu. Mô hình dự báo được kiểm tra lại trên cơ sở tập kiểm tra. Tuy độ chính xác của các dự báo không bằng so với kết quả thu được từ tập huấn luyện, nhưng kết quả dự báo giá quặng sắt từ tập kiểm tra cũng rất gần với giá quặng sắt thực tế.

Bảng 1. Kết quả phân tích tương quan

	IO	SI	SPI	EX	FR	ERa	ERc	Covid	PN
IO	1,00								
SI	0,89	1,00							
SPI	0,22	0,28	1,00						
EX	0,70	0,74	-0,19	1,00					
FR	0,35	0,25	0,37	-0,09	1,00				
ERa	0,69	0,71	0,34	0,30	0,17	1,00			
ERc	0,25	0,25	-0,22	0,28	-0,19	0,40	1,00		
Covid	0,36	0,33	-0,36	0,67	0,02	-0,06	0,10	1,00	
PN	-0,60	-0,72	0,24	-0,81	0,25	-0,47	-0,37	-0,30	1,00

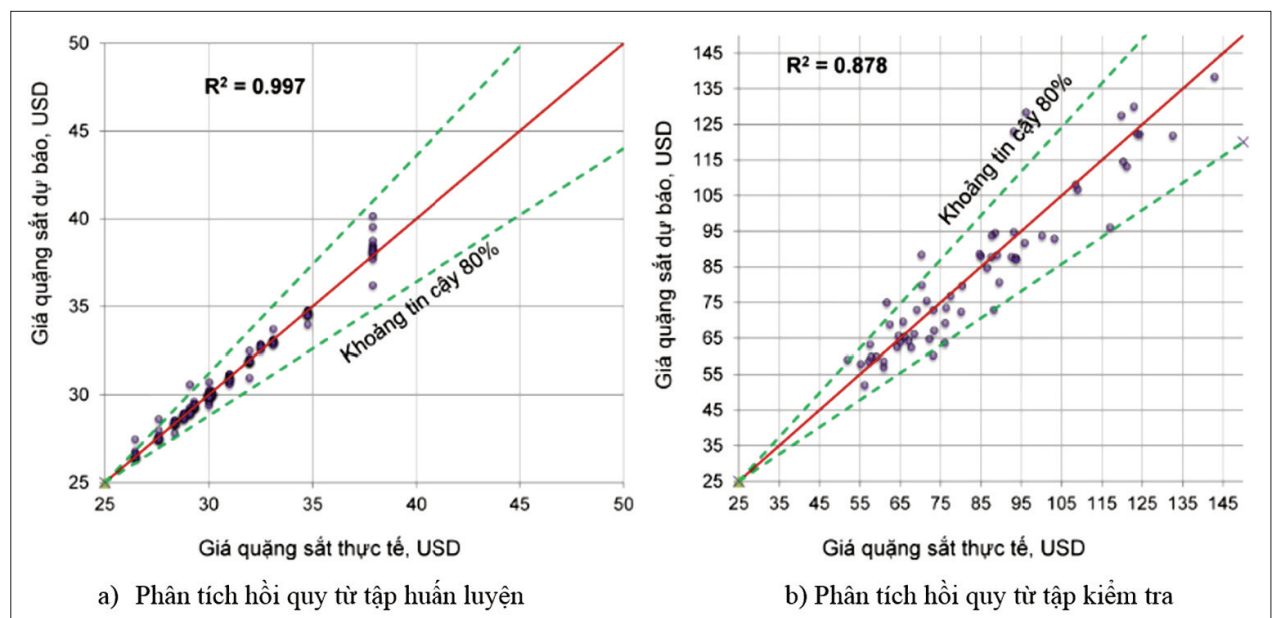
Nguồn: Kết quả phân tích của tác giả

Hình 1: Dự báo giá quặng sắt thế giới (USD/MT) theo mô hình Random Forest



Nguồn: Kết quả phân tích của tác giả

Hình 2: Độ phù hợp của hàm hồi quy



Nguồn: Kết quả phân tích của tác giả

Để đánh giá hiệu suất của mô hình Random Forest trong việc dự báo giá quặng sắt, tác giả sử dụng các chỉ tiêu đánh giá sai số dự báo là sai số trung bình tuyệt đối (MAE) và căn bậc hai trung bình bình phương sai số dự báo (RMSE). Các sai số dự báo từ tập huấn luyện, MAE = 1,329 và RMSE = 2,848. Trong khi đó, dự báo từ tập kiểm tra có MAE = 8,78 và RMSE = 13,65, cao hơn so với các sai số dự báo tương ứng từ tập huấn luyện, nhưng vẫn ở mức thấp.

• Độ phù hợp của hàm hồi quy

Độ phù hợp của hàm hồi quy được thể hiện thông qua hệ số xác định R^2 . Nếu $R^2 > 0,7$ thì có thể coi là hàm hồi quy xây dựng được là phù hợp. Kết quả phân tích ở Hình 2 cho thấy hệ số xác định của hàm hồi quy được đề xuất trong bài báo này lần lượt là $R^2 = 0,997$ và $R^2 = 0,878$ đối với dữ liệu phân tích từ tập huấn luyện và tập kiểm tra. Điều đó cho thấy các biến độc lập trong mô hình giúp giải thích lần lượt 99,7% và 87,89% sự biến động

của giá quặng sắt thế giới cho tập huấn luyện và tập kiểm tra. Hàm hồi quy xây dựng được có độ phù hợp cao.

4. Kết luận

Bài báo này vận dụng mô hình trí tuệ nhân tạo Random Forest để dự báo giá quặng sắt thế giới dựa trên 8 nhân tố ảnh hưởng tới giá quặng sắt, bao gồm chỉ số giá sắt phế liệu, chỉ số sản xuất thép, xuất khẩu quặng sắt của Úc, cước phí vận tải đường biển, tỷ giá của đồng đôla Úc và của đồng nhân dân

tệ, đại dịch Covid-19 và cơ chế thương lượng giá quặng sắt thế giới. Trên cơ sở mẫu nghiên cứu với 384 quan sát về các biến được thu thập theo tháng từ tháng 3/1990 đến tháng 2/2022, kết quả nghiên cứu cho thấy mô hình Random Forest được đề xuất trong nghiên cứu này có thể dự báo tốt giá quặng sắt thế giới. Kết quả dự báo từ tập huấn luyện và tập kiểm tra đều bám sát giá quặng sắt thực tế trong giai đoạn nghiên cứu với sai số dự báo (MAE và RMSE) thấp ■

Lời cảm ơn:

Bài báo được hoàn thành với sự hỗ trợ từ đề tài nghiên cứu cấp cơ sở mã số T20-34 của Trường Đại học Mở - Địa chất.

TÀI LIỆU THAM KHẢO:

- Alameer, Z., Elaziz, M.A., Ewees, A.A. et al (2019). Forecasting copper price using hybrid adaptive neuro-fuzzy inference system and genetic algorithms. *Natural Resources Research*, 28, 1385-1401. <https://doi.org/10.1007/s11053-019-09473-w>.
- Breiman, L. (1999). *Random forests*. UC Berkeley TR567.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <http://dx.doi.org/10.1023/A:1010933404324>.
- Dehghani, H. (2018). Forecasting copper price using gene expression programming. *Journal of Mining and Environment*, 9(2), 349-360. <https://dx.doi.org/10.22044/jme.2017.6195.1435>.
- Dehghani, H., Bogdanovic, D. (2018). Copper price estimation using BAT algorithm. *Resource Policy*, 55(C), 55-61. DOI: 10.1016/j.resourpol.2017.10.015.
- Fred Economic Data, <<https://fred.stlouisfed.org/series/IPN3311A2RS>>.
- Haque, Md.A., Topal, E., Lilford, E. (2015). Iron ore prices and the value of Australian Dollar. *Mining Technology*, 124(2), 107-120. <https://doi.org/10.1179/1743286315Y.0000000008>.
- Index Mundi, <<https://www.indexmundi.com/>>
- Kristjanpoller, W. và Hernandez, E. (2017). Volatility of main metal forecasted by a hybrid ANN-GARCH model with regressors. *Expert Systems with Application*, 84. <https://doi.org/10.1016/j.eswa.2017.05.024>.
- Lasheras, F.S., Cos Juez, F.J., Sanchez, A.S. et al (2015). Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. *Resource Policy*. 45:37-43. <https://doi.org/10.1016/j.resourpol.2015.03.004>.
- Ma, Y. (2013). Iron ore spot price volatility and change in forward pricing mechanism. *Resource Policy*, 38(4), 621-627. DOI: 10.1016/j.resourpol.2013.10.002.
- Nguyễn Văn Dân và Trần Xuân Hải, (2012). *Giáo trình Cơ sở hình thành giá cả*. Nhà xuất bản Tài chính.
- Pustov, A., Malanichev, A., Khabotilov, I. (2013). Long-term iron ore price modeling: Marginal costs vs. incentive price. *Resource Policy*, 38(4), 558-567. DOI:10.1016/j.resourpol.2013.09.003.
- Ramyar, S. và Kianfar, F. (2017). Forecasting crude oil prices: a comparison between artificial neural networks and vector autoregressive models. *Computational Economics*, 53, 743-761. DOI:10.1007/s10614-017-9764-7.
- Trading Economics, <<https://tradingeconomics.com/>>.

16. Warell, L. (2014). The effect of change in pricing regime on iron ore prices. *Resource Policy*, 41, 16-22. <https://doi.org/10.1016/j.resourpol.2014.02.002>.
17. Warell, L. (2018). An analysis of iron ore prices during the latest commodity boom. *Mineral Economics*, 31, 203-216. <https://doi.org/10.1007/s13563-018-0150-2>.
18. Zhu, Z.R. (2012). *Identifying supply and demand elasticities of iron ore*. PhD Thesis, Duke University, Durham.
19. Currency converter, <<https://fxtop.com/>>

Ngày nhận bài: 9/3/2022

Ngày phản biện đánh giá và sửa chữa: 7/4/2022

Ngày chấp nhận đăng bài: 17/4/2022

Thông tin tác giả:

TS. VŨ DIỆP ANH

Khoa Kinh tế và Quản trị kinh doanh

Trường Đại học Mở - Địa chất

USING RANDOM FOREST MODEL TO FORECAST GLOBAL IRON ORE PRICE

● **PhD. VU DIEP ANH**

Faculty of Economics and Business Administration
Hanoi University of Mining and Geology

ABSTRACT:

Iron ore is a valuable resource and it plays an important role in the global economic development. Fluctuations in iron ore prices may affect related industries like construction, steel making, car manufacturing, and ship building. Therefore, accurate prediction of global iron ore prices has attracted the attention of enterprises, investors, and scholars. In this study, an artificial intelligence model, namely Random Forest, is used to forecast global iron ore price and this model uses monthly data from March 1990 to February 2022 about the factors affecting iron ore price including the scrap price index, the steel production index, the Australian iron ore exports, the international ocean freight rate, the exchange rate of the Australian dollar, the exchange rate of the Chinese yuan, the Covid-19 pandemic and the iron ore price negotiation mechanism. The analysis results show that the proposed Random Forest model can predict the global iron ore prices well. The forecasted values are close to the actual prices for the training and testing datasets with low prediction errors, MAE = 1.329 and 8.78, RMSE = 2.848 and 13.65 for the training and testing datasets, respectively.

Keywords: iron ore, global iron ore price, forecast, Random Forest model.