

Dự báo giá cổ phiếu XOM đóng cửa và các chiến lược giao dịch bằng phương pháp Bayesian Ridge Regression và thuật toán tối ưu hóa Grid Search

TRẦN BÁ THUẬN*
TRẦN THỊ DIỆU THUẦN**

Tóm tắt

Nghiên cứu này áp dụng mô hình Bayesian Ridge Regression (BRR) kết hợp với thuật toán Grid Search để dự báo giá đóng cửa cổ phiếu XOM (Exxon Mobil). Dữ liệu lịch sử được sử dụng để huấn luyện mô hình và Grid Search được áp dụng để tối ưu hóa các siêu tham số nhằm cải thiện độ chính xác dự báo. Kết quả đánh giá mô hình cho thấy, BRR tối ưu hóa đạt độ chính xác cao và vượt trội so với các mô hình học máy khác trong việc dự báo giá cổ phiếu. Tuy nhiên, để nâng cao hiệu suất, có thể bổ sung dữ liệu cảm xúc hoặc thử nghiệm các mô hình tiên tiến hơn như Bayesian Neural Networks và các kỹ thuật tối ưu hóa như Optuna. Việc điều chỉnh mô hình liên tục và theo dõi biến động thị trường cũng là yếu tố quan trọng để cải thiện tính hiệu quả của mô hình trong môi trường tài chính năng động.

Từ khóa: Hồi quy Ridge Bayesian, tìm kiếm lưới, học máy, dự báo cổ phiếu XOM

Summary

This study applies the Bayesian Ridge Regression (BRR) model combined with the Grid Search algorithm to forecast the closing price of XOM (Exxon Mobil) stock. Historical data is used to train the model, and Grid Search is applied to optimize hyperparameters to improve forecasting accuracy. The model evaluation results show that the optimized BRR achieves high accuracy and outperforms other machine learning models in forecasting stock prices. However, it is possible to add sentiment data or experiment with more advanced models to improve the performance, such as Bayesian Neural Networks, and optimization techniques, such as Optuna. Continuous model tuning and monitoring market fluctuations are crucial factors to improve the model's effectiveness in a dynamic financial environment.

Keywords: Bayesian Ridge regression, Grid Search, machine learning, XOM stock forecasting

ĐẶT VẤN ĐỀ

Trong bối cảnh thị trường tài chính toàn cầu biến động mạnh, dự báo giá cổ phiếu là yêu cầu cấp thiết, giúp nhà đầu tư đưa ra quyết định mua, bán hoặc giữ cổ phiếu, đồng thời quản lý rủi ro và tối ưu hóa lợi nhuận. Các phương pháp truyền thống còn hạn chế trong việc nắm bắt tính phi tuyến và bất ổn của thị trường, trong khi các kỹ thuật học máy (Machine Learning) đã chứng minh hiệu quả vượt trội. Bayesian Ridge Regression (BRR), một phương pháp dựa trên xác suất Bayes, được đánh giá cao nhờ khả năng giảm quá khớp và mô hình hóa các quan hệ phi tuyến giữa các yếu tố ảnh hưởng đến giá cổ phiếu. Kết hợp với thuật toán Grid Search để tối ưu siêu tham số, BRR có

thể cải thiện đáng kể hiệu suất dự báo. Nghiên cứu này áp dụng BRR kết hợp Grid Search để dự báo giá cổ phiếu XOM (ExxonMobil), một công ty dầu khí hàng đầu thế giới. XOM được chọn làm đối tượng nghiên cứu không chỉ vì tầm ảnh hưởng trên thị trường, mà còn do ngành dầu khí chịu nhiều tác động từ yếu tố kinh tế vĩ mô, giá dầu và chính sách.

Mục tiêu của nghiên cứu này là áp dụng mô hình Bayesian Ridge Regression kết hợp Grid Search để dự báo giá cổ phiếu XOM, so sánh hiệu quả dự báo giữa các phương pháp học máy khác nhau như: Linear Regression, Lasso Regression và Adaboost Regressor. Đồng thời, nghiên cứu cũng sẽ phát triển các chiến lược giao dịch tối ưu nhằm giúp nhà đầu tư ra quyết định đúng đắn trong bối cảnh biến động của thị trường tài chính.

* Trường Đại học Kinh tế Huế | Email: tbthuan@hce.edu.vn

** Khoa Công nghệ Hóa học, Trường Đại học Công nghiệp TP. Hồ Chí Minh | Email: tranthidieuthuan@iuh.edu.vn

Ngày nhận bài: 05/12/2024; Ngày phản biện: 05/01/2025; Ngày duyệt đăng: 15/01/2025

TỔNG QUAN NGHIÊN CỨU

Các nghiên cứu gần đây đã cho thấy tiềm năng to lớn của phương pháp Bayes trong mô hình hóa, dự báo chi phí giao dịch và biến động thị trường tài chính. Briere và cộng sự (2020) sử dụng mạng Bayes để ước tính chi phí giao dịch trên thị trường chứng khoán Mỹ, đặc biệt khi sự mất cân đối lệnh giao dịch chỉ được quan sát một phần. Nghiên cứu này nhấn mạnh tầm quan trọng của kích thước lệnh và sự mất cân đối trong việc ảnh hưởng đến chi phí giao dịch.

Yao (2024) kết hợp mô hình Bayes và hồi quy đa biến để dự báo giá cổ phiếu S&P 500, cho thấy mô hình kết hợp vượt trội hơn so với từng phương pháp riêng lẻ. Tương tự, Martin và cộng sự (2023) đánh giá phương pháp dự báo Bayes trong kinh tế và tài chính, khẳng định khả năng xử lý bất định của mô hình này. Nguyễn và cộng sự (2020) áp dụng thống kê Bayes để dự báo giá cổ phiếu ngân hàng Việt Nam trong ngắn hạn, đạt độ chính xác 89.02%. Lê (2023) sử dụng Bayes để đánh giá tác động của giá dầu đến khả năng sinh lời của ngân hàng, chỉ ra mối quan hệ ngược chiều giữa giá dầu và ROA. Hortúa và Mora-Valencia (2024) áp dụng học sâu Bayes để dự báo chỉ số biến động VIX, trong đó mạng TCN đạt RMSE 0,189. Còn Ampomah và cộng sự (2021) đã khám phá tiềm năng của thuật toán Gaussian Naïve Bayes trong dự báo chuyển động giá cổ phiếu, khẳng định tính hiệu quả khi kết hợp với các kỹ thuật tiền xử lý dữ liệu.

Tổng quan nghiên cứu cho thấy, phương pháp Bayes ngày càng trở thành công cụ quan trọng trong dự báo và quản lý rủi ro tài chính.

“Dự báo giá cổ phiếu XOM đóng cửa và các chiến lược giao dịch sử dụng Bayesian Ridge Regression và thuật toán tối ưu hóa Grid Search” sẽ lấp đầy một khoảng trống trong nghiên cứu tài chính, đặc biệt là trong việc áp dụng phương pháp Bayesian Ridge Regression cho dự báo giá cổ phiếu cụ thể. Mặc dù các nghiên cứu trước đã sử dụng phương pháp Bayes trong dự báo tài chính, những việc tối ưu hóa mô hình với Grid Search, Random Search, BO, Optuna và kết hợp với chiến lược giao dịch cụ thể cho cổ phiếu XOM vẫn chưa được khai thác nhiều, mở ra cơ hội cải thiện độ chính xác trong dự báo và chiến lược giao dịch.

Tính mới của nghiên cứu “Dự báo giá cổ phiếu XOM và chiến lược giao dịch sử dụng Bayesian Ridge Regression với Grid Search” nằm ở việc áp dụng Bayesian Ridge Regression kết hợp với Grid Search để tối ưu hóa tham số, cải thiện độ chính xác trong dự báo giá cổ phiếu XOM. Nghiên cứu này so sánh hiệu quả của mô hình Bayesian với các phương pháp truyền thống, như: Linear Regression, Lasso Regression và Adaboost Regressor, cung cấp cái nhìn mới về khả năng dự báo trong thị trường tài chính, đồng thời tối ưu hóa các chiến lược giao dịch hiệu quả hơn.

PHƯƠNG PHÁP NGHIÊN CỨU

Dữ liệu nghiên cứu được thu thập từ Yahoo Finance, là giá cổ phiếu đóng cửa trong một khoảng thời gian 45 năm để đảm bảo nguồn dữ liệu đủ lớn (Big Data). Dữ liệu được xử lý kỹ lưỡng: loại bỏ giá trị thiếu, chuẩn hóa và chia thành các tập huấn luyện, kiểm tra. Ngoài ra, các ngoại lệ và tính đầy đủ của dữ liệu được kiểm tra cẩn thận nhằm đảm bảo chất lượng phân tích. Mô hình Bayesian Ridge Regression kết hợp Grid Search được sử dụng để tối ưu hóa siêu tham số, mang lại hiệu quả cao trong dự báo giá cổ phiếu.

Bayesian Ridge Regression (BRR) là một phương pháp hồi quy tiên tiến, kết hợp hồi quy tuyến tính với nguyên lý Bayes để ước lượng phân phối xác suất của các tham số. Không giống như các mô hình hồi quy tuyến tính truyền thống, BRR sử dụng phân phối xác suất hậu nghiệm thay vì một giá trị cụ thể cho các hệ số hồi quy. Điều này giúp xử lý tốt hơn tính bất định trong dữ liệu và giảm thiểu hiện tượng quá khớp (overfitting). Trong BRR, xác suất hậu nghiệm của các tham số w (hệ số hồi quy) được ước lượng dựa trên dữ liệu đầu vào X và đầu ra y , theo công thức:

$$p(w|X, y) = \frac{p(y|X, w).p(w)}{p(y|X)} \quad (1)$$

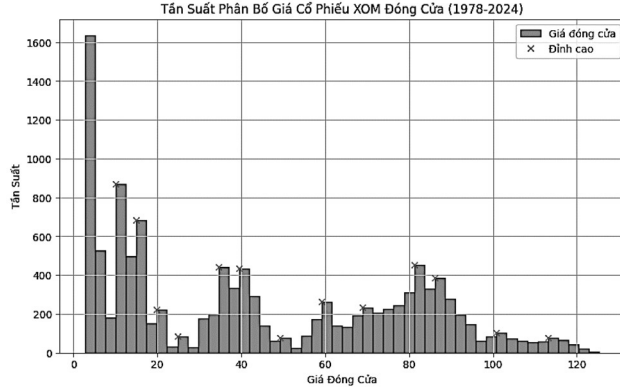
Trong đó: $p(w|X, y)$ là phân phối xác suất hậu nghiệm của tham số w khi biết dữ liệu X và nhãn y . $p(y|X, w)$ là hàm khả năng (likelihood) giả định phân phối Gaussian của đầu ra y với độ lệch chuẩn điều chỉnh thông qua các tham số. $p(w)$ phân phối tiên nghiệm của các tham số w , là Gaussian; $p(y|X)$ là hệ số chuẩn hóa (marginal likelihood).

BRR đặc biệt hiệu quả với các tập dữ liệu nhỏ hoặc khi các biến độc lập có tương quan cao, nhờ vào cơ chế điều chỉnh co giãn (shrinkage) của Ridge Regression và tính năng xác suất tiên nghiệm. Bằng cách này, BRR tạo ra các ước lượng ổn định và chính xác, giúp cải thiện hiệu suất mô hình trong các tình huống phức tạp.

Mô hình toán học của BRR thể hiện dạng phương trình hồi quy tuyến tính $y = \beta X + \varepsilon$ (2). Trong đó, y là vector đầu ra (phản hồi), X là ma trận các biến đầu vào (dữ liệu), β là vector các hệ số hồi quy và ε là sai số (giả sử có phân phối chuẩn với trung bình bằng 0 và phương sai là σ^2).

Mô hình BRR giả định rằng, $p(\beta|\alpha) \propto N(0, \alpha^{-1}I)$ (3), trong đó α là siêu tham số điều chỉnh mức độ co lại (prior). Với điều kiện y được biết: $p(y|X, \beta, \lambda) \propto N(X\beta, \lambda^{-1}I)$ (4), trong đó λ là phương sai của sai số ngẫu nhiên. Thông qua định lý Bayes, phân phối hậu nghiệm của β được tính như sau: $p(\beta|X, y, \alpha, \lambda) \propto p(y|X, \beta, \lambda)p(\beta|\alpha)$ (5). Mô hình BRR ước lượng phân phối hậu nghiệm của các hệ số hồi quy β dựa trên dữ liệu quan sát. Sau khi quan sát, ta có thể tính toán giá trị trung bình và độ lệch chuẩn của các hệ số. Ưu điểm của BRR là giảm thiểu hiện tượng quá khớp (overfitting) và khả năng điều chỉnh độ phức tạp mô

HÌNH 1: TẦN SUẤT PHÂN BỐ GIÁ CỔ PHIẾU XOM ĐÓNG CỬA



HÌNH 2: LỊCH SỬ GIÁ CỔ PHIẾU XOM ĐÓNG CỬA



Nguồn: Tổng hợp của nhóm tác giả

hình linh hoạt nhờ vào siêu tham số α điều chỉnh mức độ co lại. BRR đặc biệt hữu ích trong các bài toán dự đoán với dữ liệu hạn chế hoặc có đa cộng tuyến cao, giúp cải thiện độ tin cậy của các ước lượng.

Grid Search là kỹ thuật tối ưu hóa siêu tham số trong học máy, dùng để tìm bộ siêu tham số tối ưu cho mô hình. Phương pháp này thử nghiệm tất cả các kết hợp có thể của siêu tham số và chọn ra sự kết hợp mang lại kết quả tốt nhất dựa trên chỉ số đánh giá như MAE hoặc MSE. Các bước chính bao gồm: xác định các siêu tham số và giá trị thử nghiệm, tạo các kết hợp siêu tham số, huấn luyện mô hình với mỗi kết hợp và đánh giá hiệu suất. Kết quả là bộ siêu tham số tối ưu cho mô hình.

Linear Regression thích hợp với các mối quan hệ tuyến tính giữa biến độc lập và biến phụ thuộc. Tuy nhiên, nó gặp khó khăn khi dữ liệu có tính phi tuyến tính, nhiễu cao, hoặc hiện tượng đa cộng tuyến. Lasso Regression thêm thành phần L1 regularization, giúp thực hiện feature selection (lựa chọn đặc trưng) bằng cách đưa một số hệ số hồi quy về 0. Phù hợp khi có nhiều biến độc lập không quan trọng, nhưng có thể dẫn đến mất thông tin khi hệ số bị triệt tiêu hoàn toàn. Adaboost Regressor là mô hình kết hợp nhiều cây quyết định nhỏ (weak learners) để giảm thiểu sai số và cải thiện độ chính xác. Tuy nhiên, Adaboost dễ bị ảnh hưởng bởi nhiễu trong dữ liệu và đòi hỏi nhiều tài nguyên tính toán khi làm việc với tập dữ liệu lớn. Bayesian Ridge Regression (BRR) kết hợp kỹ thuật regularization (L2 regularization) với phương pháp Bayesian inference. BRR xử lý tốt dữ liệu nhiễu, giảm thiểu ảnh hưởng của đa cộng tuyến, và mang lại

độ ổn định cũng như tính chính xác cao hơn trong các mô hình hồi quy phức tạp.

Thang đo đánh giá mô hình BRR như sau:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad STD = \sqrt{\frac{1}{n} \sum_{i=1}^n (error_i - \bar{error})^2}$$

NGHIÊN CỨU THỰC NGHIỆM

Tiền xử lý dữ liệu: Dữ liệu giá cổ phiếu XOM được tải từ Yahoo Finance từ ngày 01/01/1978 đến ngày 16/11/2024, chỉ lấy cột giá cổ phiếu XOM đóng cửa và chuẩn hóa bằng StandardScaler. Cột dự đoán được tạo bằng cách dịch chuyển giá trị đóng cửa lên 30 ngày. Dữ liệu được chia thành X (đặc trưng) và y (giá trị dự đoán), loại bỏ giá trị không khớp. Cuối cùng, dữ liệu được tách thành tập huấn luyện và kiểm tra theo tỷ lệ 80/20 (Bài viết sử dụng cách viết số thập phân theo chuẩn quốc tế).

Các mức giá đỉnh cao trong biểu đồ giá cổ phiếu XOM đóng cửa (Hình 1), nơi các đỉnh đại diện cho những mức giá phổ biến, có thể là tín hiệu quan trọng để ra quyết định giao dịch mua hoặc bán. Trong suốt chiều dài lịch sử của công ty, các mức giá đỉnh cao này bao gồm: [10.07048186, 14.97684445, 19.88320704, 24.78956964, 34.60229482, 39.50865741, 49.32138259, 59.13410777, 68.94683295, 81.21273943, 86.11910202, 100.83818979, 113.10409627]. Kết quả này hỗ trợ tối ưu hóa chiến lược đầu tư.

Thực quan hóa dữ liệu giúp nhận diện xu hướng giá, theo dõi và phân tích biến động một cách hiệu quả. Hình 2 cho thấy, mức giá cao nhất của cổ phiếu XOM là 125.37 USD vào ngày 07/10/2024, phản ánh sự tăng trưởng gần đây, trong khi mức giá thấp nhất là 2.71 USD vào ngày 26/01/1978, là thời điểm cổ phiếu XOM được niêm yết lần đầu. Sự biến động này cho thấy tiềm năng tăng trưởng và lịch sử phát triển của công ty.

Thiết lập thuật toán Grid Search: Thực hiện việc tối ưu hóa siêu tham số của mô hình BRR bằng phương pháp Grid Search. Thiết lập một lưới các giá trị siêu tham số như: `alpha_1`, `alpha_2`, `lambda_1`, `lambda_2` và các tham số khác như `fit_intercept`, `tol`, `n_iter`, `compute_score`. Grid Search sẽ thử tất cả các kết hợp của các giá trị này và sử dụng phương pháp cross-validation (`cv=5`) để đánh giá hiệu suất của mỗi kết hợp thông qua chỉ số `neg_mean_absolute_error`. Quá trình này giúp tìm ra các siêu tham số tối ưu cho mô hình, từ đó cải thiện độ chính xác dự báo. Bộ siêu tham số tìm thấy cho mô hình BRR với dữ liệu của XOM như Bảng 1.

Bộ siêu tham số tối ưu của BRR gồm: `alpha_1 = 1e-06` và `alpha_2 = 0.001`, giúp điều chỉnh phân rã trọng số để giảm quá khớp và cải thiện độ ổn định. `compute_score = True` tính toán log marginal

likelihood, hỗ trợ đánh giá mô hình. $\text{fit_intercept} = \text{False}$ phù hợp khi dữ liệu đã được chuẩn hóa, loại bỏ bias không cần thiết. $\text{lambda}_1 = 0.001$ và $\text{lambda}_2 = 1e-06$ điều chỉnh trọng số nhỏ và lớn, đảm bảo sự cân bằng giữa độ phức tạp và khả năng khái quát. $n_iter = 200$ xác định số vòng lặp tối đa để đảm bảo hội tụ, trong khi $\text{tol} = 0.001$ kiểm soát ngưỡng dừng của quá trình tối ưu hóa. Bộ tham số này phù hợp với dữ liệu tài chính phức tạp như cổ phiếu XOM, giúp tăng hiệu suất dự báo và giảm thiểu sai số.

Xây dựng mô hình BRR tối ưu: Khởi tạo mô hình BRR với các siêu tham số tối ưu đã được tìm ra từ quá trình tinh chỉnh siêu tham số (hyperparameter tuning).

Mô hình BRR tối ưu được huấn luyện trên tập dữ liệu (x_train, y_train) để tìm trọng số tối ưu, sau đó đánh giá qua các chỉ số hiệu suất. MAE (Mean Absolute Error) đo sai số tuyệt đối trung bình giữa dự đoán và thực tế, với giá trị 0.0682 chỉ ra mô hình dự đoán chính xác. MSE (Mean Squared Error) đo sai số bình phương trung bình, với giá trị 0.0126, biểu thị sai số rất nhỏ. RMSE (Root Mean Squared Error), căn bậc hai của MSE, có giá trị 0.1124, chứng tỏ sai số trung bình nhỏ và hiệu suất dự báo tốt. MAPE (Mean Absolute Percentage Error) cho thấy, sai số theo phần trăm so với giá trị thực tế, với giá trị 0.3779%, xác nhận độ chính xác cao trong dự báo. Các chỉ số này thể hiện rằng, mô hình BRR tối ưu có hiệu suất tốt, với sai số nhỏ và dự báo chính xác.

Kết quả tính toán của nhóm nghiên cứu cho biết R^2 (hệ số xác định) là 0.9873, cho thấy mô hình giải thích 98.73% sự biến động của dữ liệu, chứng tỏ hiệu quả rất cao. Độ lệch chuẩn (STD) là 0.1121, gần với RMSE, cho thấy sai số phân bố đều và không quá lớn. Các chỉ số này khẳng định mô hình BRR có độ chính xác cao, với R^2 gần 1 và MAE, RMSE, MAPE nhỏ, phù hợp với dữ liệu XOM. Mô hình BRR tối ưu dự báo chính xác giá cổ phiếu XOM đóng cửa trong 26 ngày tới.

Dự báo giá XOM từ ngày 17/11 đến 12/12/2024 (Bảng 3) cho thấy xu hướng biến động rõ rệt. Giai đoạn đầu (17-18/11), giá tăng nhẹ với khuyến nghị Mua, phản ánh kỳ vọng phục hồi ngắn hạn. Giai đoạn giữa (19-29/11), xu hướng giảm chiếm ưu thế với nhiều tín hiệu Bán, cho thấy áp lực bán mạnh và rủi ro giảm giá. Cuối kỳ (30/11-06/12), giá phục hồi nhẹ nhưng xen kẽ giữa các khuyến nghị Mua và Bán, phản ánh sự biến động và tâm lý thị trường thiếu chắc chắn. Chiến lược tối ưu tập trung vào việc Mua ở các mức giá thấp trong giai đoạn phục hồi (17-18/11, 20/11, 23-24/11, 01-06/12) và Bán khi giá giảm hoặc chạm đỉnh ngắn hạn (19/11, 21-29/11, 07-12/12). Nhà đầu tư nên tận dụng thời điểm Mua khi giá thấp và Bán khi xu hướng giảm chiếm ưu thế để tối đa hóa lợi nhuận và quản lý rủi ro. Giai đoạn cuối (07-12/12) cần thận trọng do thị trường thiếu xu hướng rõ ràng. Chiến lược này nhấn mạnh sự linh hoạt và nhạy bén trong việc ra quyết định giao dịch ngắn hạn.

BẢNG 1: BỘ SIÊU THAM SỐ TỐI ƯU CỦA BRR TÌM THẤY TỪ GRID SEARCH

alpha 1: 1e-06
alpha 2: 0.001
compute score: True
fit intercept: False
lambda 1: 0.001
lambda 2: 1e-06
n iter: 200
tol: 0.001

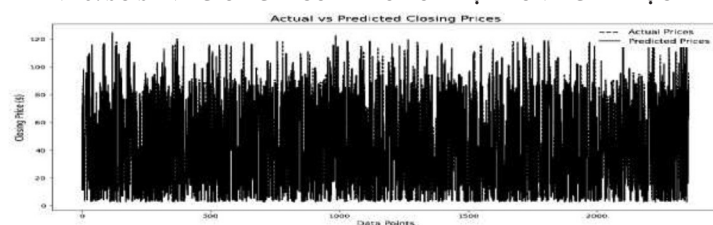
BẢNG 2: MÔ HÌNH BRR TỐI ƯU

```
# Bộ siêu tham số tối ưu của BRR
best_params = {'alpha_1': 1e-06, 'alpha_2': 0.001, 'compute_score':
True, 'fit_intercept': False, 'lambda_1': 0.001, 'lambda_2': 1e-06, 'n_
iter': 200, 'tol': 0.001}
# Mô hình BRR tối ưu
best_br_model = BayesianRidge(**best_params)
```

BẢNG 3: KẾT QUẢ DỰ BÁO GIÁ XOM ĐÓNG CỬA VÀ CHIẾN LƯỢC TỐI ƯU

Ngày	Giá-Chiến lược	Ngày	Giá-Chiến lược
17-11-2024	\$123.51-Mua	30-11-2024	\$116.60-Bán
18-11-2024	\$123.98-Mua	01-12-2024	\$116.69-Mua
19-11-2024	\$120.25-Bán	02-12-2024	\$114.86-Bán
20-11-2024	\$120.56-Mua	03-12-2024	\$118.52-Mua
21-11-2024	\$120.25-Bán	04-12-2024	\$118.86-Mua
22-11-2024	\$119.91-Bán	05-12-2024	\$120.90-Mua
23-11-2024	\$119.98-Mua	06-12-2024	\$121.05-Mua
24-11-2024	\$120.60-Mua	07-12-2024	\$121.01-Bán
25-11-2024	\$120.17-Bán	08-12-2024	\$120.37-Bán
26-11-2024	\$119.49-Bán	09-12-2024	\$120.25-Bán
27-11-2024	\$119.39-Bán	10-12-2024	\$121.37-Mua
28-11-2024	\$118.80-Bán	11-12-2024	\$120.46-Bán
29-11-2024	\$117.19-Bán	12-12-2024	\$119.21-Bán

HÌNH 3: SO SÁNH GIỮA GIÁ CỔ PHIẾU XOM DỰ BÁO VÀ GIÁ THỰC TẾ



BẢNG 3: KẾT QUẢ DỰ BÁO GIÁ XOM ĐÓNG CỬA VÀ CHIẾN LƯỢC TỐI ƯU

Ngày	Giá-Chiến lược	Ngày	Giá-Chiến lược
17-11-2024	\$123.51-Mua	30-11-2024	\$116.60-Bán
18-11-2024	\$123.98-Mua	01-12-2024	\$116.69-Mua
19-11-2024	\$120.25-Bán	02-12-2024	\$114.86-Bán
20-11-2024	\$120.56-Mua	03-12-2024	\$118.52-Mua
21-11-2024	\$120.25-Bán	04-12-2024	\$118.86-Mua
22-11-2024	\$119.91-Bán	05-12-2024	\$120.90-Mua
23-11-2024	\$119.98-Mua	06-12-2024	\$121.05-Mua
24-11-2024	\$120.60-Mua	07-12-2024	\$121.01-Bán
25-11-2024	\$120.17-Bán	08-12-2024	\$120.37-Bán
26-11-2024	\$119.49-Bán	09-12-2024	\$120.25-Bán
27-11-2024	\$119.39-Bán	10-12-2024	\$121.37-Mua
28-11-2024	\$118.80-Bán	11-12-2024	\$120.46-Bán
29-11-2024	\$117.19-Bán	12-12-2024	\$119.21-Bán

Nguồn: Tính toán của nhóm tác giả

KẾT QUẢ VÀ THẢO LUẬN

Mô hình dự đoán giá cổ phiếu XOM trong 26 ngày sử dụng các chỉ số hiệu suất tốt, thể hiện khả năng dự báo chính xác và tin cậy. Các chỉ số như MAE =

BẢNG 4: SO SÁNH HIỆU SUẤT CÁC MÔ HÌNH VỚI BRR

Mô hình	MAE	MSE	RMSE	MAPE	R ²	STD
Linear Regression	0.0682	0.0126	0.1124	0.3779	0.9873	1.1121
Lasso Regression	0.1170	0.0220	0.1482	0.3914	0.9779	0.1480
Bayesian Ridge	0.0682	0.0126	0.1124	0.3779	0.9873	0.1121
Adaboost Regressor	0.0673	0.0120	0.1095	0.7668	0.9880	0.1095

BẢNG 5: SO SÁNH HIỆU SUẤT BRR THÔNG QUA BO, RS, OPT VÀ GS

Metric	Bayesian Optimization	Random Search	Optuna	Grid Search
MAE	0.0682	0.0682	0.0683	0.0682
MSE	0.0126	0.0126	0.0126	0.0126
RMSE	0.1124	0.1124	0.1121	0.1124
MAPE	0.3779	0.3779	0.2991	0.3779
R ² Score	0.9873	0.9873	0.9874	0.9873
STD	0.1121	0.1121	0.1121	0.1121

Nguồn: Tính toán của nhóm nghiên cứu

0.0682, MSE = 0.0126 và RMSE = 0.1124 cho thấy, sai số dự đoán thấp, chứng minh mô hình có độ chính xác cao. Chỉ số MAPE = 0.3779% phản ánh sai số dự báo theo tỷ lệ phần trăm, trong khi R² = 0.9873, gần giá trị lý tưởng là 1, cho thấy mô hình giải thích rất tốt sự biến động của giá cổ phiếu. Ngoài ra, độ lệch chuẩn STD = 0.1121 cho thấy mức độ ổn định của sai số dự báo. Chiến lược giao dịch được đề xuất dựa trên kết quả dự báo: Bán trong các giai đoạn giá suy giảm để giảm rủi ro và Mua khi giá có xu hướng tăng trở lại nhằm tận dụng lợi nhuận tiềm năng. Giai đoạn hồi phục đáng chú ý cho thấy khả năng thị trường có thể phục hồi sau những đợt giảm giá. Hiệu suất mô hình BRR tối ưu với R² = 0.9873 đánh giá sự thành công trong việc giải thích biến động của dữ liệu, vượt trội so với các nghiên cứu trước đây. Mô hình này cho thấy khả năng tổng quát và học hỏi đặc trưng quan trọng từ dữ liệu, nhờ vào việc tối ưu hóa qua Grid Search, giúp cải thiện độ chính xác dự báo, đem lại sự tin cậy cao trong chiến lược giao dịch.

Bảng 4 cho thấy, mô hình Bayesian Ridge Regression là sự lựa chọn tốt nhất trong số các mô

hình được đánh giá. Mô hình này đạt được sự cân bằng tốt giữa các chỉ số sai số thấp, R² cao và sự ổn định vượt trội (STD thấp). AdaBoost Regressor có hiệu suất dự đoán tốt nhất (thấp nhất về MAE, MSE, RMSE), nhưng MAPE cao chỉ ra rằng mô hình này không phù hợp với bài toán có dữ liệu biến đổi mạnh hoặc đòi hỏi giải thích tỷ lệ phần trăm sai số. Linear Regression có hiệu suất tương tự Bayesian Ridge về sai số và độ chính xác, nhưng độ phân tán sai số cao hơn (STD cao), dẫn đến kém ổn định hơn. Lasso Regression hoạt động kém hơn so với Bayesian Ridge và AdaBoost ở tất cả các tiêu chí, cho thấy không phù hợp với bài toán này.

KẾT LUẬN

Mô hình BRR đã chứng minh hiệu quả trong việc dự đoán giá cổ phiếu, với các chỉ số, như MAE = 0.0682, MSE = 0.0126, RMSE = 0.1124, MAPE = 0.3779%, R² = 0.9873 và độ lệch chuẩn STD = 0.1121, cho thấy độ chính xác cao. Đặc biệt, giá trị R² gần 1 khẳng định khả năng giải thích tốt sự biến động giá. Chiến lược giao dịch “Mua” và “Bán” được đề xuất mang lại thông tin hữu ích, giúp nhà đầu tư tối ưu hóa lợi nhuận. Phân tích xu hướng giá tăng, giảm từ dự đoán hỗ trợ xây dựng chiến lược chủ động. Mô hình kết hợp regularization (Ridge) và tính xác suất (Bayes), phù hợp dữ liệu nhiễu. Tuy nhiên, mô hình vẫn đối mặt với hạn chế, như khó lường trước biến động bất thường và tác động của yếu tố vĩ mô. Việc bổ sung dữ liệu cảm xúc và thử nghiệm các mô hình hiện đại hơn là Bayesian Neural Networks tối ưu hóa bởi thuật toán Optuna là hướng nghiên cứu tương lai để cải thiện hiệu suất. Mô hình BRR là một bước tiến thực tiễn, nhưng cần theo dõi và điều chỉnh để phù hợp với thị trường năng lượng và tài chính liên tục biến động.□

TÀI LIỆU THAM KHẢO

1. Ampomah, E. K., Nyame, G., Qin, Z., Addo, P. C., Gyamfi, E. O., Gyan, M. (2021), Stock market prediction with Gaussian naïve Bayes machine learning algorithm, *Informatica*, 45(2), 243-256, <https://doi.org/10.31449/inf.v45i2.3407>.

2. Brière, M., Lehalle, C. A., Nefedova, T., Raboun, A. (2020), *Modeling transaction costs when trades may be crowded: A Bayesian network using partially observable orders imbalance*, Wiley, <https://doi.org/10.1002/9781119751182.ch11>.

3. Hortúa, H. J., Mora-Valencia, A. (2024), Forecasting VIX using Bayesian deep learning, *International Journal of Data Science and Analytics*, <https://doi.org/10.1007/s41060-024-00562-5>.

4. Lê, T. T. (2023), Tác động của giá dầu đến khả năng sinh lời của các ngân hàng thương mại Việt Nam, *Tạp chí Khoa học Đại học Cần Thơ*, 59(5D), 207-215. <https://doi.org/10.22144/ctujos.2023.182>.

5. Martin, G. M., Frazier, D. T., Maneesoonthorn, W., Loaiza-Maya, R., Huber, F., Koop, G., Maheu, J., Nibbering, D., Panagiotelis, A. (2023), *Bayesian forecasting in economics and finance: A modern review*, ArXiv.

6. Nguyễn, P. Đ., Lê, T. H., Phạm, T. B. (2020), Thông kê Bayes và ứng dụng trong dự báo giá chứng khoán của các ngân hàng và công ty tài chính ở Việt Nam, *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 31(4), 40-63.

7. Yao, R. (2024), *Hybrid Bayesian and multiple regression models for stock market prediction: A comparative analysis of predictive accuracy and risk management*, Proceedings of the 3rd International Conference on Financial Technology and Business Analysis, <https://doi.org/10.54254/2754-1169/137/2024.18682>.