

DỰ BÁO GIÁ CỔ PHIẾU CATL ĐÓNG CỬA VÀ CÁC CHIẾN LƯỢC GIAO DỊCH SỬ DỤNG XGBOOST VÀ THUẬT TOÁN TỐI ƯU HÓA OPTUNA

● TRẦN BÁ THUẤN - TRẦN THỊ DIỆU THUẦN

TÓM TẮT:

Trong nghiên cứu này, chúng tôi đã áp dụng thuật toán XGBoost kết hợp với tối ưu hóa tham số qua Optuna để dự báo giá cổ phiếu CATL và phát triển các chiến lược giao dịch hiệu quả. Kết quả đánh giá cho thấy mô hình XGBoost đạt chỉ số R^2 là 0.9201, MAE là 0.1982 và MSE là 0.0825, vượt trội hơn so với các mô hình khác như ElasticNet và Decision Tree. Đồng thời, chiến lược giao dịch sử dụng dự báo giá cổ phiếu đã đưa ra các khuyến nghị mua, bán, hoặc giữ trong 30 ngày tới, giúp nhà đầu tư tối ưu hóa lợi nhuận. Mô hình XGBoost thể hiện khả năng dự đoán chính xác và ổn định, đặc biệt khi được tối ưu hóa thông qua Optuna, giúp tăng cường hiệu quả trong việc đưa ra các quyết định giao dịch. Tuy nhiên, nghiên cứu cũng chỉ ra rằng, việc tích hợp các mô hình deep learning phức tạp hơn, hoặc các thuật toán tối ưu hóa khác có thể mở ra cơ hội cải thiện thêm độ chính xác và tính hiệu quả. Các hướng nghiên cứu tiếp theo có thể bao gồm ứng dụng các phương pháp deep learning, hoặc kết hợp thêm các yếu tố từ dữ liệu phi cấu trúc để cải thiện dự báo và chiến lược giao dịch.

Từ khóa: dự báo giá cổ phiếu CATL, mô hình XGBoost, pin xe điện, thuật toán tối ưu hóa Optuna, năng lượng bền vững.

1. Đặt vấn đề

Trong bối cảnh sự phát triển nhanh chóng của ngành công nghiệp xe điện và nhu cầu về năng lượng bền vững ngày càng tăng, các công ty sản xuất pin xe điện, đặc biệt là Contemporary Amperex Technology Co. Limited (CATL), đã trở thành những nhà đầu tư quan trọng trong thị trường toàn cầu. Với sự đóng góp quan trọng của CATL vào ngành công nghiệp xe điện, việc dự báo giá cổ phiếu CATL có thể giúp các nhà đầu tư đưa ra quyết định đầu tư chính xác hơn, tối ưu hóa chiến lược giao dịch và góp phần vào sự phát triển bền vững của ngành. Dự báo giá cổ phiếu trong thị

trường tài chính luôn là một bài toán phức tạp do sự biến động mạnh mẽ của các yếu tố ảnh hưởng, từ thông tin kinh tế đến sự thay đổi trong tâm lý thị trường. Tuy nhiên, với sự phát triển của các mô hình học máy, đặc biệt là XGBoost, cùng với các thuật toán tối ưu hóa như Optuna, bài toán này có thể được giải quyết hiệu quả hơn. XGBoost, với khả năng xử lý dữ liệu phi tuyến tính và đưa ra dự báo chính xác, đã trở thành một công cụ mạnh mẽ trong các ứng dụng tài chính. Hơn nữa, thuật toán tối ưu hóa Optuna giúp tinh chỉnh các siêu tham số của mô hình, từ đó cải thiện khả năng dự báo và tối ưu hóa hiệu quả giao dịch.

Mục tiêu chính của nghiên cứu này là xây dựng một mô hình dự báo giá cổ phiếu CATL đóng cửa sử dụng XGBoost và tối ưu hóa các siêu tham số của mô hình thông qua thuật toán Optuna. Bằng cách này, nghiên cứu sẽ không chỉ giúp dự báo chính xác hơn giá cổ phiếu CATL mà còn đề xuất các chiến lược giao dịch tối ưu cho các nhà đầu tư, góp phần thúc đẩy sự phát triển của ngành công nghiệp năng lượng bền vững và xe điện. Thông qua việc áp dụng mô hình này, các nhà đầu tư có thể nắm bắt các xu hướng thị trường và ra quyết định kịp thời, tối đa hóa lợi nhuận từ việc đầu tư vào CATL.

2. Tổng quan nghiên cứu

Trong bối cảnh phát triển nhanh chóng của các mô hình học máy, thuật toán XGBoost đã chứng tỏ hiệu suất vượt trội trong việc dự báo giá cổ phiếu và tối ưu hóa chiến lược giao dịch. Nghiên cứu của Ting Li (2024) đã xây dựng mô hình dự báo giá điện sử dụng cơ chế hình thành giá và tối ưu hóa tham số bằng thuật toán Optuna, cho thấy XGBoost có độ chính xác cao hơn so với các mô hình khác như Gradient Boosting Regression Tree và Random Forest. Tương tự, nghiên cứu của Kyung Keun Yun et al. (2021) kết hợp XGBoost với thuật toán di truyền (GA-XGBoost) và ba giai đoạn tiền xử lý đặc trưng, cải thiện đáng kể độ chính xác dự báo và tối ưu hóa tập đặc trưng hiệu quả. Farshid Balaneji et al. (2024) đã kết hợp phân tích cảm xúc từ tin tức và dữ liệu thị trường vào mô hình XGBoost, tối ưu hóa tham số bằng Bayesian Optimization, chứng minh sức mạnh của mô hình trong việc phản ánh sự biến động của thị trường qua tin tức. Trong lĩnh vực giao dịch tiền điện tử, Abdul Jabbar et al. (2024) đã so sánh hiệu suất của 41 mô hình học máy, trong đó XGBoost và Random Forest nổi bật với độ chính xác cao và khả năng quản lý rủi ro tốt. Arsalan Dezhkam et al. (2023) kết hợp Hilbert-Huang Transform (HHT) với XGBoost để dự báo xu hướng giá cổ phiếu, đạt hiệu suất cao hơn 99,8% so với dữ liệu tài chính thô. Trong khi đó, Jin Xiao et al. (2024) sử dụng mô hình OPT-LightGBM, mặc dù cải tiến đáng kể về độ chính xác và hiệu suất, vẫn để ngỏ khả năng ứng dụng XGBoost trong các tình huống phi tuyến tính và

phức tạp hơn. Nghiên cứu của Trương Thị Thùy Dương (2023) tập trung vào chỉ số chứng khoán VnIndex, sử dụng thuật toán XGBoost và chỉ báo kỹ thuật, đạt độ chính xác trên 80%, đồng thời chỉ ra vai trò quan trọng của khối lượng giao dịch trong dự báo xu hướng thị trường. Solar Energy Forecasting Study (2024) chứng minh rằng các mô hình cây quyết định như Random Forests, Bagging và Extremely Randomized Trees vượt trội hơn Logit Regression trong việc dự báo hướng giá cổ phiếu năng lượng mặt trời, nhưng hiệu quả chiến lược giao dịch vẫn phụ thuộc vào khả năng tối ưu hóa và điều chỉnh tham số.

Từ các kết quả nghiên cứu trên, có thể thấy, XGBoost là một công cụ mạnh mẽ trong dự báo giá cổ phiếu và tối ưu hóa chiến lược giao dịch. Tuy nhiên, vẫn tồn tại một khoảng trống nghiên cứu liên quan đến việc tối ưu hóa tham số XGBoost với Optuna trong bối cảnh dự báo giá cổ phiếu CATL, đặc biệt là trong việc xây dựng chiến lược giao dịch hiệu quả. Nghiên cứu này đề xuất mô hình “Dự báo giá cổ phiếu CATL đóng cửa và các chiến lược giao dịch sử dụng XGBoost và thuật toán tối ưu Optuna”, nhằm khắc phục các hạn chế hiện có, nâng cao độ chính xác dự báo, đồng thời tối ưu hóa chiến lược giao dịch thông qua phân tích sâu hơn về mối quan hệ giữa các chỉ báo kỹ thuật và biến động giá. Tính mới của nghiên cứu nằm ở sự kết hợp giữa XGBoost, kỹ thuật tối ưu tham số từ Optuna và ứng dụng thực tế trên dữ liệu cổ phiếu CATL, mở ra hướng đi mới trong lĩnh vực dự báo tài chính và chiến lược giao dịch cổ phiếu.

Về Xgboost (eXtreme Gradient Boosting): là thuật toán học máy mạnh mẽ, được thiết kế để giải quyết các bài toán hồi quy và phân loại với hiệu suất cao và khả năng tổng quát hóa tốt. Khác với các mô hình truyền thống, XGBoost tối ưu hóa quá trình huấn luyện bằng cách hỗ trợ tính toán song song, xử lý hiệu quả dữ liệu lớn và giảm thiểu hiện tượng quá khớp thông qua các tham số chuẩn hóa L1 và L2. Ngoài ra, thuật toán này có thể tự động xử lý giá trị khuyết thiếu và hỗ trợ kỹ thuật Early Stopping, giúp dừng huấn luyện sớm khi mô hình không còn cải thiện. Trong dự báo giá cổ phiếu CATL, XGBoost cho phép xử lý dữ liệu chuỗi thời

gian phức tạp, nắm bắt các xu hướng phi tuyến tính và biến động thất thường. Khi kết hợp với thuật toán tối ưu hóa Optuna, mô hình có thể tìm kiếm cấu hình tham số tối ưu như `learning_rate`, `max_depth` và `n_estimators`, từ đó cải thiện độ chính xác dự báo. Với những ưu điểm nổi bật này, XGBoost không chỉ nâng cao hiệu suất mô hình, mà còn đóng vai trò quan trọng trong việc xây dựng chiến lược giao dịch hiệu quả cho thị trường cổ phiếu đầy biến động như CATL.

Thuật toán Optuna: là một thư viện tối ưu siêu tham số tiên tiến, nổi bật với khả năng tìm kiếm thích ứng (Adaptive Search) dựa trên Bayesian Optimization và kỹ thuật Tree-structured Parzen Estimator (TPE). Khác với Grid Search và Random Search, Optuna không dò tìm ngẫu nhiên hay cố định, mà học hỏi từ kết quả trước để điều chỉnh tham số hiệu quả. So với Hyperband, Optuna kết hợp Early Stopping (Pruning) thông minh để giảm thiểu tài nguyên tính toán. Ngoài ra, Optuna hỗ trợ đa mục tiêu (Multi-Objective Optimization), giúp tối ưu hóa hiệu suất mô hình XGBoost nhanh chóng và chính xác.

Dữ liệu nghiên cứu: trong đề tài “dự báo giá cổ phiếu CATL đóng cửa và các chiến lược giao dịch sử dụng XGBoost và thuật toán tối ưu hóa Optuna”, dữ liệu nghiên cứu được tập trung hoàn toàn vào giá đóng cửa (Close) của cổ phiếu CATL (mã 300750.SZ). Dữ liệu được thu thập từ nguồn Yahoo Finance trong khoảng thời gian từ ngày 01/01/2010 đến 5/4/2024. Việc chỉ sử dụng giá đóng cửa làm đầu vào và đầu ra giúp giảm thiểu nhiễu từ các biến số khác, tạo điều kiện thuận lợi cho mô hình XGBoost tập trung vào dự báo xu hướng giá chính xác. Dữ liệu được xử lý và tiền xử lý để đảm bảo tính toàn vẹn, loại bỏ các giá trị khuyết thiếu và điều chỉnh định dạng thời gian. Sự lựa chọn này phù hợp với mục tiêu dự báo giá trị cụ thể và xây dựng chiến lược giao dịch dựa trên mô hình tối ưu.

Thang đo đánh giá mô hình Xgboost:

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

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

$$\text{và } STD = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{error}_i - \overline{\text{error}})^2}$$

3. Phân tích thực nghiệm

3.1. Thu thập dữ liệu

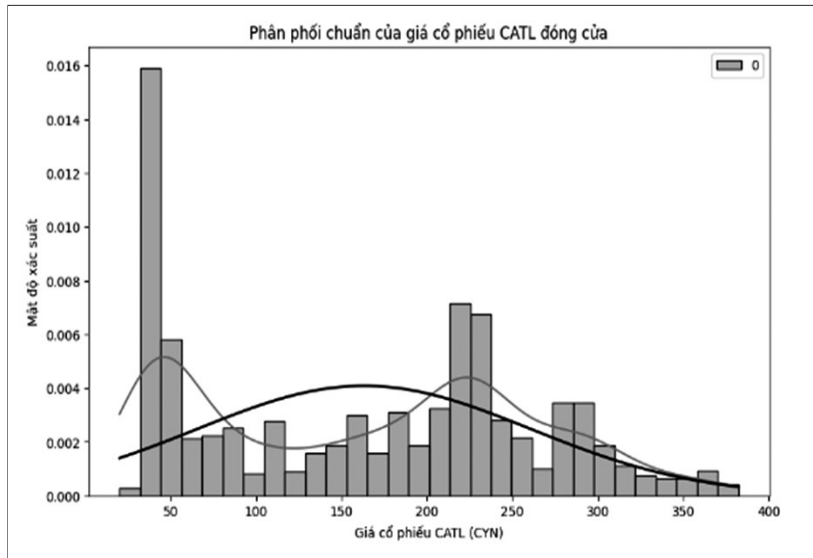
Tiền xử lý dữ liệu: giá cổ phiếu đóng cửa từ DataFrame được biểu diễn phân phối tần suất, tính toán giá trị trung bình và độ lệch chuẩn của giá cổ phiếu, từ đó tạo ra một phân phối chuẩn dựa trên các tham số này. Phân phối chuẩn được vẽ lên đồ thị để so sánh với histogram, giúp trực quan hóa độ phù hợp của dữ liệu với phân phối chuẩn.

Trực quan hóa dữ liệu: trực quan hóa giá cổ phiếu CATL đóng cửa giúp người dùng dễ dàng nhận diện xu hướng và biến động giá theo thời gian. Biểu đồ đường thể hiện sự thay đổi liên tục của giá, hỗ trợ phân tích các điểm mạnh yếu trong thị trường. (Hình 1)

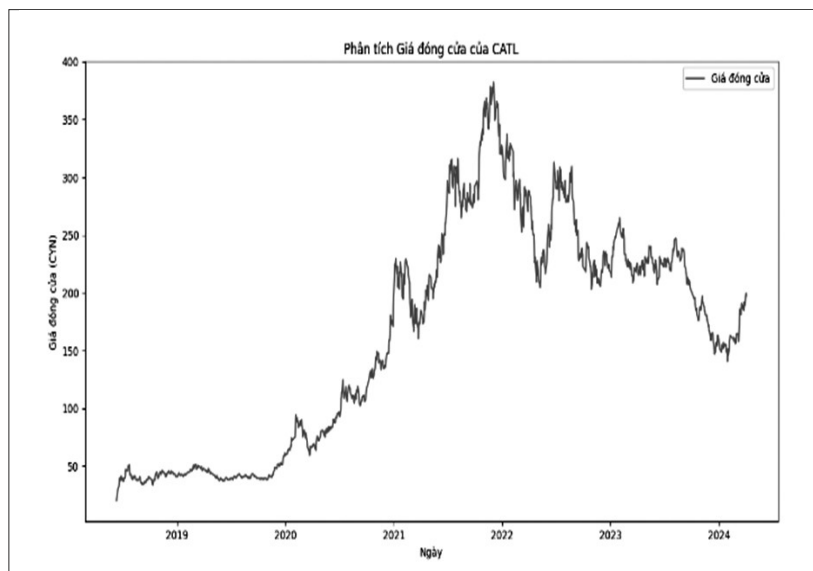
Thực hiện các bước chuẩn hóa và chuẩn bị dữ liệu để huấn luyện mô hình dự báo giá cổ phiếu CATL. Đầu tiên, giá cổ phiếu được chuẩn hóa bằng cách sử dụng StandardScaler để đảm bảo các giá trị có phân phối chuẩn và có thể so sánh trực tiếp. Sau đó, một cột dự đoán mới được tạo ra bằng cách dịch chuyển dữ liệu giá đóng cửa về phía trước 30 ngày (`shift(-upcoming_days)`), tạo ra một tập dữ liệu mục tiêu cho việc dự báo giá trong tương lai. Tiếp theo, dữ liệu được tách thành các đặc trưng (X) và nhãn (y), với X bao gồm tất cả các cột trừ ‘Prediction’ và y là giá trị dự đoán của cột ‘Prediction’. Cuối cùng, dữ liệu được chia thành tập huấn luyện và kiểm tra, với 80% dữ liệu dùng để huấn luyện mô hình và 20% còn lại để kiểm tra độ chính xác của mô hình sau khi huấn luyện. (Hình 2)

3.2. Xây dựng mô hình Xgboost tối ưu

Cài đặt thuật toán Optuna: sử dụng Optuna để tối ưu hóa các tham số của mô hình XGBoost. Hàm objective xác định các tham số cần tối ưu hóa, bao gồm số lượng cây (`n_estimators`), độ sâu tối đa của cây (`max_depth`), tỷ lệ học (`learning_rate`), tỷ lệ mẫu dùng trong mỗi cây (`subsample`) và tỷ lệ mẫu

Hình 1: Phân phối chuẩn giá cổ phiếu CATL đóng cửa

Nguồn: Tác giả, 2024

Hình 2: Lịch sử giá cổ phiếu CATL đóng cửa

Nguồn: Tác giả, 2024

Bảng 1. Mô hình Xgboost tối ưu

Kết quả tham số tối ưu từ Optuna:

```
best_params = {'n_estimators': 148, 'max_depth': 4, 'learning_rate': 0.032307685000935296, 'subsample': 0.9005210044609593, 'colsample_bytree': 0.8328443096397664}
```

Mô hình Xgboost với tham số tối ưu:

```
model = XGBRegressor(**best_params, random_state=42)
```

cột trong mỗi cây (colsample_bytree). Mô hình XGBoost được huấn luyện trên dữ liệu huấn luyện và dự đoán giá trị trên tập kiểm tra, sau đó tính toán lỗi tuyệt đối trung bình (MAE) để đánh giá chất lượng mô hình. Quá trình tối ưu hóa được thực hiện với 50 lần thử nghiệm (n_trials=50) để tìm giá trị MAE tối thiểu. (Bảng 1)

Các siêu tham số tối ưu từ Optuna bao gồm: n_estimators (148) xác định số lượng cây trong mô hình, max_depth (4) giới hạn chiều sâu cây, learning_rate (0.032) điều chỉnh tốc độ học, subsample (0.90) là tỷ lệ mẫu ngẫu nhiên, và colsample_bytree (0.83) xác định tỷ lệ cột sử dụng trong mỗi cây.

3.3. Huấn luyện và đánh giá mô hình Xgboost tối ưu

Huấn luyện mô hình XGBoost: model.fit(x_train, y_train) trên dữ liệu huấn luyện x_train và nhãn mục tiêu y_train. Quá trình này tối ưu hóa các tham số của mô hình để học mối quan hệ giữa các đặc trưng và giá trị dự đoán. Sau khi huấn luyện, mô hình Xgboost tối ưu sẵn sàng để dự đoán trên dữ liệu mới.

Đánh giá mô hình Xgboost tối ưu trên tập dữ liệu kiểm tra cho

Nguồn: Tác giả, 2024

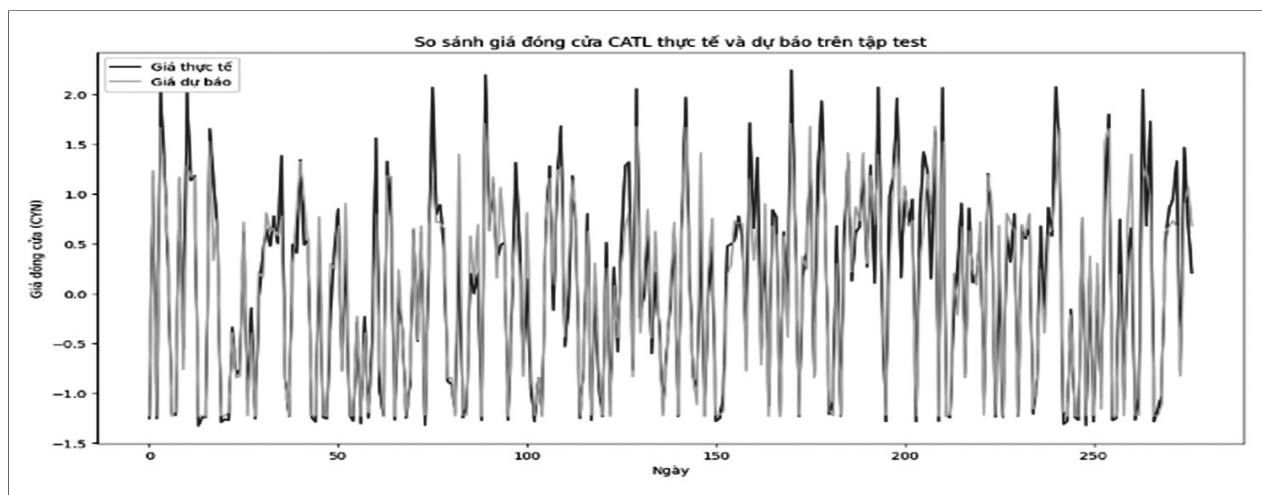
thấy mô hình có hiệu suất tốt với các chỉ số chính như sau: MAE (Mean Absolute Error) đạt 0.1982, chỉ ra độ lệch trung bình tuyệt đối của dự báo so với giá trị thực tế. MSE (Mean Squared Error) là 0.0825, phản ánh sai số bình phương trung bình. RMSE (Root Mean Squared Error) bằng 0.2872, cho biết độ lệch chuẩn của dự đoán. MAPE (Mean Absolute Percentage Error) là 1.165%, chỉ ra độ lệch phần trăm trung bình giữa dự đoán và giá trị thực tế. (Hình 3)

R^2 Score đạt 0.9201, cho thấy mô hình giải thích 92.01% biến động của dữ liệu. Cuối cùng, độ lệch chuẩn (STD) là 0.2863, phản ánh sự phân tán của sai số dự đoán. Hiệu suất cao chứng tỏ mô hình

ngày 04/05/2024 cho thấy xu hướng biến động của giá cổ phiếu và các quyết định giao dịch phù hợp. Trong giai đoạn đầu, từ ngày 05/04 đến ngày 15/04, giá cổ phiếu CATL có xu hướng giảm, với các khuyến nghị “Bán” vào ngày 06/04 (CYN177.94) và tiếp tục giảm vào ngày 07/04 (CYN171.67), cho thấy chiến lược phòng ngừa rủi ro. Tuy nhiên, từ ngày 10/04 đến 13/04, giá cổ phiếu có sự phục hồi nhẹ, đi kèm với các khuyến nghị “Mua”, với mức giá tăng dần (CYN169.61 - CYN170.13).

Sang đến ngày 16/04 và tiếp theo, xu hướng tăng trưởng mạnh mẽ được ghi nhận với các khuyến nghị “Mua”, đặc biệt là trong khoảng thời gian từ ngày 18/04 đến ngày 02/05 (CYN183.14 -

Hình 3: Giá CATL dự báo và giá thực tế



Nguồn: Tác giả, 2024

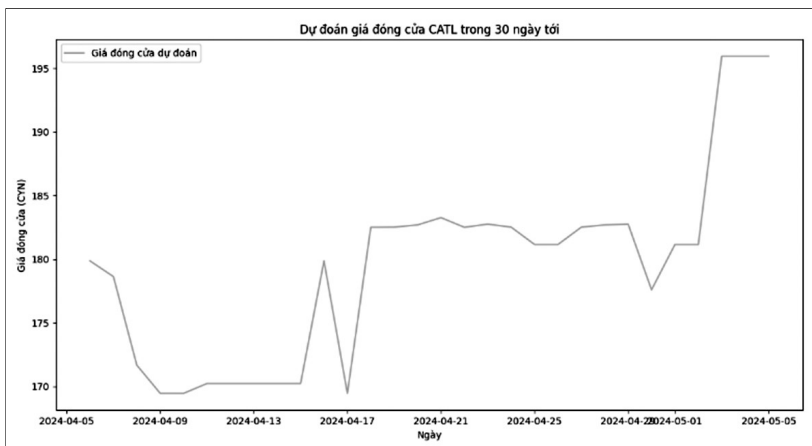
XGBoost mang lại độ chính xác cao và khả năng dự báo hiệu quả giá cổ phiếu.

3.4. Ứng dụng của Xgboost tối ưu

Tạo đồ thị dự đoán giá cổ phiếu CATL trong 30 ngày tới, sử dụng các ngày tương lai và hiển thị giá đóng cửa dự đoán. Nó cung cấp cái nhìn trực quan về xu hướng biến động giá cổ phiếu trong thời gian tới. (Hình 4)

Kết quả chiến lược giao dịch dự báo giá cổ phiếu CATL trong 30 ngày từ ngày 05/04/2024 đến

Hình 4: Dự báo giá cổ phiếu CATL đóng cửa



Nguồn: Tác giả, 2024

CYN198.00), cho thấy tiềm năng lạc quan trong các ngày này. Đặc biệt, các ngày cuối (03/05 - 04/05), các khuyến nghị “Giữ” (CYN198.00) chứng tỏ sự ổn định giá trị cổ phiếu. Tổng thể, chiến lược giao dịch này khuyến nghị duy trì sự linh hoạt trong việc điều chỉnh hành động dựa trên diễn biến của giá cổ phiếu, vừa phòng ngừa rủi ro trong giai đoạn giảm giá, vừa tận dụng cơ hội tăng trưởng mạnh mẽ khi thị trường phục hồi. (Bảng 2)

4. Kết quả và thảo luận

Trong việc so sánh hiệu suất của các mô hình ElasticNet, Decision Tree và Xgboost, mô hình Xgboost vượt trội nhất với MAE thấp nhất (0.1982), MSE (0.0825), RMSE (0.2872), và MAPE (1.1650%), cho thấy khả năng dự đoán chính xác và ổn định. Xgboost cũng có R2 Score cao nhất (0.9201), chứng tỏ mô hình này giải thích tốt sự biến động của dữ liệu. Decision Tree, mặc dù dễ triển khai và giải thích, nhưng hiệu suất không bằng Xgboost với MAE cao hơn và dễ bị overfitting.

ElasticNet có hiệu suất thấp nhất, với MAE cao nhất (0.2542) và không thể vượt qua các mô hình phức tạp hơn như Xgboost. Do đó, Xgboost là lựa chọn tối ưu cho bài toán này.

Mặc dù Xgboost cho kết quả tốt, nhưng cần một số điều chỉnh và tối ưu hóa tham số để tránh overfitting như L1 và L2 regularization có thể được sử dụng để điều chỉnh mô hình theo cách giúp cải thiện khả năng tổng quát hóa của nó. Nghiên cứu thêm về việc kết hợp các mô hình học máy với deep learning, như mạng nơ-ron hồi tiếp (RNN, LSTM) để cải thiện hiệu suất dự báo trong các bài toán với chuỗi thời gian. (Bảng 3)

5. Kết luận

Trong nghiên cứu này, chúng tôi đã phát triển mô hình dự báo giá cổ phiếu CATL sử dụng thuật toán XGBoost kết hợp với tối ưu hóa tham số qua Optuna. Kết quả cho thấy mô hình XGBoost với các tham số tối ưu hóa mang lại hiệu suất cao, với chỉ số R² Score đạt 0.9201 và MAE là 0.1982,

Bảng 2. Kết quả dự báo giá CATL và chiến lược giao dịch 30 ngày

Ngày-tháng-năm	Giá-Giao dịch	Ngày-tháng-năm	Giá-Giao dịch
05/04/2024	CYN179.15	20/04/2024	CYN183.49-Giữ
06/04/2024	CYN177.94-Bán	21/04/2024	CYN183.14-Bán
07/04/2024	CYN171.67-Bán	22/04/2024	CYN183.15-Mua
08/04/2024	CYN169.00-Bán	23/04/2024	CYN184.26-Mua
09/04/2024	CYN169.00-Giữ	24/04/2024	CYN182.38-Bán
10/04/2024	CYN169.61-Mua	25/04/2024	CYN182.38-Giữ
11/04/2024	CYN170.13-Mua	26/04/2024	CYN184.26-Mua
12/04/2024	CYN169.61-Bán	27/04/2024	CYN183.49-Bán
13/04/2024	CYN170.13-Mua	28/04/2024	CYN183.15-Bán
14/04/2024	CYN169.61-Bán	29/04/2024	CYN178.76-Bán
15/04/2024	CYN179.15-Mua	30/04/2024	CYN182.38-Mua
16/04/2024	CYN169.00-Bán	01/05/2024	CYN182.38-Giữ
17/04/2024	CYN183.14-Mua	02/05/2024	CYN198.00-Mua
18/04/2024	CYN184.26-Mua	03/05/2024	CYN198.00-Giữ
19/04/2024	CYN183.49-Bán	04/05/2024	CYN198.00-Giữ

Nguồn: Tác giả, 2024

Bảng 3. So sánh hiệu suất của Xgboost với các mô hình

Thang đo\Mô hình	ElasticNet	Decision Tree	Xgboost
MAE	0.2542	0.2045	0.1982
MSE	0.1138	0.0889	0.0825
RMSE	0.3373	0.2981	0.2872
MAPE	0.9057%	1.3178%	1.1650%
R² Score	0.8843	0.9139	0.9201
STD	0.3341	0.2971	0.2863

Nguồn: Tác giả, 2024

chứng tỏ khả năng dự đoán chính xác giá cổ phiếu trong tương lai. Đồng thời, các chiến lược giao dịch dựa trên dự báo này cũng giúp đưa ra các quyết định mua, bán, hoặc giữ hợp lý, tối ưu hóa lợi nhuận cho nhà đầu tư. Mô hình XGBoost vượt trội so với các phương pháp khác như ElasticNet và Decision

Tree về độ chính xác và khả năng khái quát. Nghiên cứu này đóng góp vào việc ứng dụng học máy trong phân tích và dự báo thị trường chứng khoán, mở ra hướng nghiên cứu tiếp theo về các phương pháp tối ưu hóa kết hợp và các mô hình deep learning phức tạp hơn ■

TÀI LIỆU THAM KHẢO:

1. Dương, T. T. T. (2023). Dự báo chiều biến động của chỉ số chứng khoán bằng thuật toán tăng cường. Tạp chí Khoa học và Đào tạo Ngân hàng, 252, 2537.

2. Balaneji, F., Maringer, D., & Spasić, I. (2024). The power of words: Predicting stock market returns with fine-grained sentiment analysis and XGBoost. In Intelligent Systems and Applications (pp. 577-596). Springer. <https://doi.org/10.1007/978-3-030-69722-7-30>.

3. Dezhkam, A., & Manzuri, M. T. (2023). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert-Huang transform. Engineering Applications of Artificial Intelligence, 112, 105626. <https://doi.org/10.1016/j.engappai.2022.105626>.

4. El Zaar, A., Benaya, N., Bakir, T., Mansouri, A., & El Allati, A. (2023). Prediction of US 30-years-treasury-bonds movement and trading entry point using the robust 1DCNN-BiLSTM-XGBoost algorithm. Expert Systems, 40(3), e13459.

5. Iwai, K., & Hamagami, T. (2022). A new XGBoost inference with boundary conditions in real estate price prediction. IEEE Transactions on Engineering Management, 69(4), 1282-1290.

6. Jabbar, A., & Jalil, S. Q. (2024). A comprehensive analysis of machine learning models for algorithmic trading of Bitcoin. arXiv:2407.18334v1 [q-fin.TR]. [Online] Available at <https://arxiv.org/abs/2407.18334v1>.

7. Li, T. (2024). Research on electricity price forecast model based on electricity price formation mechanism and XGBoost and highly automated Optuna algorithm. In Proceedings of the 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (pp. 1-6). IEEE Xplore.

8. Sadorsky, P. (2022). Forecasting solar stock prices using tree-based machine learning classification: How important are silver prices? The North American Journal of Economics and Finance, 58, 101378.

9. Xiao, J., Wen, Z., Jiang, X., Yu, L., & Wang, S. (2024). Three-stage research framework to assess and predict the financial risk of SMEs based on hybrid method. Decision Support Systems, 177, 114090.

10. Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. Expert Systems with Applications, 168, 115716.

Ngày nhận bài: 20/4/2024

Ngày phản biện đánh giá và sửa chữa: 13/5/2024

Ngày chấp nhận đăng bài: 27/5/2024

Thông tin tác giả:

1. TRẦN BÁ THUẤN

Khoa Hệ thống Thông tin Kinh tế, Trường Đại học Kinh tế, Đại học Huế

2. TRẦN THỊ DIỆU THUẦN

Khoa Công nghệ Hóa học, Trường Đại học Công nghiệp Thành phố Hồ Chí Minh

FORECASTING CATL CLOSING STOCK PRICES AND OPTIMIZING TRADING STRATEGIES WITH XGBOOST AND OPTUNA

● TRAN BA THUAN¹

● TRAN THI DIEU THUAN²

¹Faculty of Economic Information Systems,
Hue University of Economics

²Faculty of Chemical Engineering,
Industrial University of Ho Chi Minh City

ABSTRACT:

This study employs the XGBoost algorithm, enhanced through hyperparameter optimization with Optuna, to forecast the closing stock prices of CATL and develop effective trading strategies. The XGBoost model achieved superior performance, with an R^2 score of 0.9201, a Mean Absolute Error (MAE) of 0.1982, and a Mean Squared Error (MSE) of 0.0825, outperforming benchmark models such as ElasticNet and Decision Tree. The resulting trading strategy generated actionable buy, sell, and hold recommendations over a 30-day horizon, aiding investors in profit optimization. The findings highlight the high prediction accuracy and stability of the XGBoost model, particularly when optimized with Optuna, thereby enhancing decision-making efficiency in stock trading. Nonetheless, the study underscores the potential for further improvement through the integration of advanced deep learning models, alternative optimization techniques, or ensemble learning approaches. Future research could also explore incorporating unstructured data to refine forecasting accuracy and expand the applicability of trading strategies.

Keywords: Forecasting CATL stock prices, XGBoost model, electric vehicle batteries, Optuna Optimization Algorithm, sustainable energy.