

PHƯƠNG PHÁP DỰ ĐOÁN GIÁ CỔ PHIẾU DỰA TRÊN THUẬT TOÁN HỒI QUY

Methods for Stock Price Prediction using Regression Algorithms

Cao Thành Nam, Nguyễn Hữu Phước

Khoa Công Nghệ Thông Tin

Trường Đại học Nông Lâm TP HCM

Tóm tắt – Thị trường chứng khoán được coi là hàn thử biểu của nền kinh tế bởi nó không những đóng vai trò huy động vốn cho nền kinh tế, cung cấp môi trường đầu tư cho công chúng mà còn đánh giá giá trị doanh nghiệp và tình hình của nền kinh tế, tạo môi trường giúp Chính phủ thực hiện chính sách kinh tế vĩ mô. Là một trong những kênh đầu tư hấp dẫn và có khả năng sinh lời cao kể cả trong ngắn hạn và dài hạn, nhưng kèm theo đó rủi ro cũng luôn kề bên trong quá trình đầu tư. Tính tới cuối tháng 5/2024 có gần 7.94 triệu tài khoản được mở. Về chủ trương chung, cuối năm 2023, Thủ tướng đã phê duyệt Chiến lược phát triển thị trường chứng khoán tầm nhìn đến năm 2030. Theo đó, đặt mục tiêu nâng tổng số lượng tài khoản trên thị trường chứng khoán lên 9 triệu tài khoản vào năm 2025 và 11 triệu tài khoản vào năm 2030. Do đó, việc dự đoán giá chứng khoán không chỉ mang tính nghiên cứu khoa học mà còn mang tính thực tiễn rất cao. Trong bài báo này, chúng tôi dự đoán giá cổ phiếu của công ty Apple và trình bày một số phương hồi quy áp dụng vào bài toán dự đoán giá chứng khoán như Multi-layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest và Linear Regression. Các thuật toán này sử dụng dữ liệu đầu vào do thư viện yfinance của Python cung cấp và thêm một số các thuộc tính MA10, MA20, MA50, MA100, MA200 và Tomorrow để có thể đưa ra giá cổ phiếu trong tương lai. Kết quả thực nghiệm với dữ liệu từ 01/01/1990 đến 01/06/2024 cho thấy hiệu suất của từng thuật toán như sau: ...

Abstract – The stock market is considered the barometer of the economy because it not only plays a role in mobilizing capital for the economy and providing an investment environment for the public but also in evaluating corporate value and the state of the economy, creating a platform for the Government to implement macroeconomic policies. As one of the most attractive investment channels with high potential returns both in the short and long term, it also entails significant risks during the investment process. As of the end of May 2024, nearly 7.94 million accounts have been opened. According to the general policy, by the end of 2023, the Prime Minister approved the Stock Market Development Strategy with a vision to 2030. Accordingly, the target is to increase the total number of accounts in the stock market to 9 million by 2025 and 11 million by 2030. Therefore, predicting stock prices is not only of scientific research interest but also of high practical significance. In this paper, we predict the stock price of Apple Inc. and present several regression models applied to the problem of stock price prediction, such as MLP, SVM, KNN, Random Forest, and Linear Regression. These algorithms use input data provided by the YFinance library of Python, along with additional attributes such as MA10, MA20, MA50, MA100, MA200, and Tomorrow to predict future stock prices. Experimental results with data from 01/01/1990 to 07/01/2024 demonstrate the performance of each algorithm as follows:...

1. GIỚI THIỆU

Hiện nay có nhiều kênh đầu tư khác nhau như gửi ngân hàng, đầu tư vàng, bất động sản... Nhưng mỗi kênh đầu tư lại đem đến lợi nhuận và rủi ro khác nhau cho các nhà đầu tư. Những nhà đầu tư thích sự an toàn thì gửi ngân hàng và mua vàng tích trữ là một lựa chọn phù hợp, đối với những nhà đầu tư thích sự an toàn thì kênh bất động sản và chứng khoán được ưu tiên hàng đầu, tất nhiên khái niệm an toàn hay rủi ro tùy thuộc vào cách sử dụng của mỗi người. Trong những năm gần đây khái niệm đầu tư

chứng khoán dần trở nên phổ biến hơn vì nền kinh tế ngày càng phát triển sẽ kích thích chứng khoán tăng cao đem đến cho những nhà đầu tư lợi nhuận khổng lồ nhưng cũng kèm theo những nguy cơ khôn lường về tài chính. Cùng với xu hướng đó, việc nghiên cứu một mô hình dự đoán giá cổ phiếu chính xác cao không chỉ mang tính nghiên cứu khoa học mà còn mang tính thực tiễn rất cao, đặc biệt là đối với những người mới đầu tư, vừa là công cụ hướng dẫn đầu tư mà còn là để quản trị rủi ro, giảm thiểu các nguy cơ về tài chính.

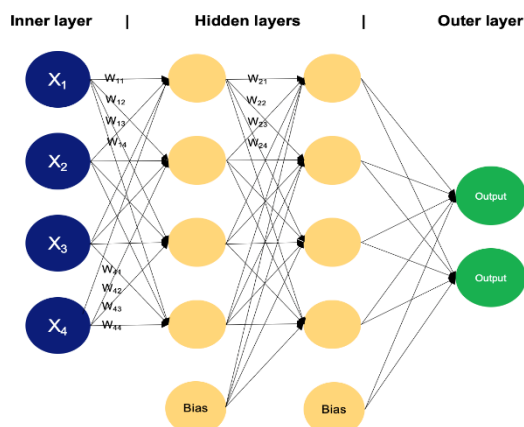
Thị trường chứng khoán là một thị trường nơi chứng khoán, chủ yếu là cổ phiếu, được mua và bán. Nó cung cấp một nền tảng cho các công ty huy động vốn bằng cách phát hành cổ phiếu và cho các nhà đầu tư giao dịch những cổ phiếu đó. Dự đoán giá chứng khoán là một trong những thách thức lớn nhất trong lĩnh vực tài chính và kinh tế. Khả năng dự đoán chính xác giá cổ phiếu không chỉ mang lại lợi ích lớn cho các nhà đầu tư mà còn đóng vai trò quan trọng trong việc ổn định và phát triển thị trường tài chính. Mặc dù có nhiều yếu tố tác động đến giá cổ phiếu, từ các yếu tố kinh tế vĩ mô như lãi suất và lạm phát đến các yếu tố vi mô như hiệu suất của công ty và tin tức thị trường, việc dự đoán chính xác giá cổ phiếu vẫn là một nhiệm vụ phức tạp và đầy thách thức. Nghiên cứu này nhằm khám phá các phương pháp và kỹ thuật khác nhau để dự đoán giá chứng khoán, với trọng tâm là các mô hình học máy và trí tuệ nhân tạo. Chúng tôi sẽ phân tích và so sánh hiệu quả của các mô hình như Linear Regression, Random Forest, Multi-layer Perceptron, K-Nearest Neighbors và Support Vector Machine trong việc dự đoán giá cổ phiếu. Bằng cách sử dụng dữ liệu lịch sử giá cổ phiếu và các yếu tố liên quan, chúng tôi kỳ vọng sẽ tìm ra mô hình dự đoán có độ chính xác cao nhất. Qua nghiên cứu này, chúng tôi mong muốn đóng góp vào kho tàng kiến thức hiện có và cung cấp những gợi ý hữu ích cho các nhà đầu tư và các nhà quản lý tài chính trong việc đưa ra quyết định đầu tư.

2. PHƯƠNG PHÁP

2.1. Thuật toán Multi-layer Perceptron (MLP)

MLP là một loại mạng neural nhân tạo dùng trong học sâu, thuộc loại Supervised Learning. Nó bao gồm một hoặc nhiều lớp ẩn (hidden layers) nằm giữa lớp đầu vào (input layer) và lớp đầu ra (output layer). Mỗi neural trong một lớp kết nối với mỗi neural trong lớp tiếp theo, và quá trình huấn luyện diễn ra thông qua thuật toán lan truyền ngược (backpropagation).

Quá trình học của MLP bao gồm hai bước chính: Lan truyền tiến (feedforward) và lan truyền ngược (backpropagation). Trong lan truyền tiến, dữ liệu được đưa vào mạng và đi qua các lớp để tạo ra đầu ra. Trong lan truyền ngược, lỗi giữa đầu ra dự đoán và đầu ra thực tế được tính toán và sử dụng để cập nhật trọng số của mạng.



Hình 1. Ví dụ thuật toán MLP

Ưu điểm:

- Có khả năng học các hàm phi tuyến tính phức tạp.
- Có thể học từ các tập dữ liệu lớn.

Hạn chế:

- Đòi hỏi một lượng lớn dữ liệu để huấn luyện và thường rất tốn thời gian.
- Có thể dễ dàng bị overfitting nếu không có các kỹ thuật regularization thích hợp.

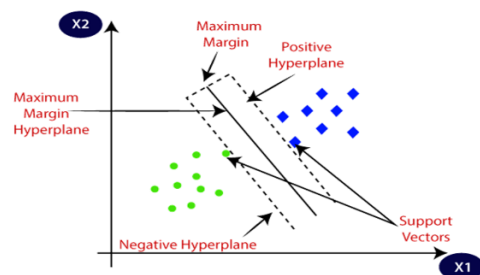
Ứng dụng:

- **Nhận diện chữ viết tay:** MLP có thể phân loại các ký tự viết- tay, như trong hệ thống nhận diện chữ viết tay của bưu điện.
- **Dự đoán tài chính:** Dùng để dự đoán giá cổ phiếu hoặc các chỉ số kinh tế.
- **Xử lý ngôn ngữ tự nhiên:** Dịch máy, phân loại văn bản, phân tích cảm xúc.

2.2. Thuật toán Support Vector Machine (SVM)

SVM là một thuật toán học máy dùng cho cả phân loại và hồi quy, thuộc loại Supervised Learning. Nó tìm cách tìm ra một siêu phẳng (hyperplane) tối ưu để phân chia dữ liệu thành các lớp khác nhau. Trong không gian nhiều chiều, siêu phẳng này có thể là một mặt phẳng hoặc một đường thẳng.

SVM sử dụng khái niệm "margin" để chọn siêu phẳng tối ưu, tức là khoảng cách lớn nhất giữa các lớp. Khi dữ liệu không tuyến tính, SVM có thể sử dụng kernel trick để biến đổi không gian dữ liệu và làm cho nó tuyến tính.



Hình 2. Ví dụ thuật toán SVM

Ưu điểm:

- Hiệu quả cao với các tập dữ liệu có kích thước nhỏ đến trung bình.
- Có khả năng hoạt động tốt với dữ liệu không tuyến tính nhờ vào kernel trick.

Hạn chế:

- Không hiệu quả với dữ liệu lớn do chi phí tính toán cao khi huấn luyện.
- Chọn thông số kernel và các tham số khác như C (regularization parameter) có thể phức tạp.

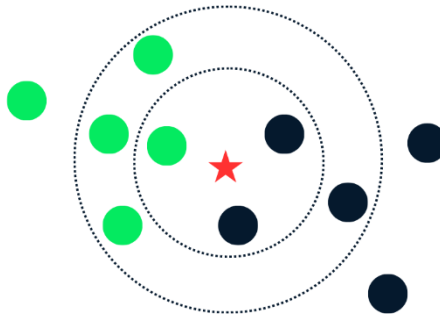
Ứng dụng:

- **Nhận dạng khuôn mặt:** Phân loại hình ảnh khuôn mặt thành các danh mục khác nhau.
- **Phân loại văn bản và thư rác:** SVM có thể được dùng để phân loại email thành thư rác hoặc không phải thư rác.
- **Phân tích gen:** SVM giúp phân loại các loại bệnh dựa trên dữ liệu gen.

2.3. Thuật toán KNN(K-Nearest Neighbors)

KNN là một thuật toán phân loại dựa trên việc tìm k điểm dữ liệu gần nhất trong

không gian đặc trưng (feature space), thuộc loại Supervised Learning. Dữ liệu mới sẽ được phân loại dựa trên đa số các nhãn của k điểm lân cận nhất. Thuật toán đơn giản nhưng hiệu quả, yêu cầu tính toán khoảng cách giữa điểm dữ liệu mới và tất cả các điểm trong tập huấn luyện.



Hình 3. Ví dụ thuật toán KNN

Ưu điểm:

- Dễ hiểu và dễ triển khai.
- Không cần huấn luyện phức tạp, chỉ cần lưu trữ dữ liệu huấn luyện.

Hạn chế:

- Chậm khi dữ liệu lớn vì phải tính toán khoảng cách cho mỗi điểm dữ liệu mới.
- Dễ bị ảnh hưởng bởi nhiễu và các điểm dữ liệu không chuẩn.

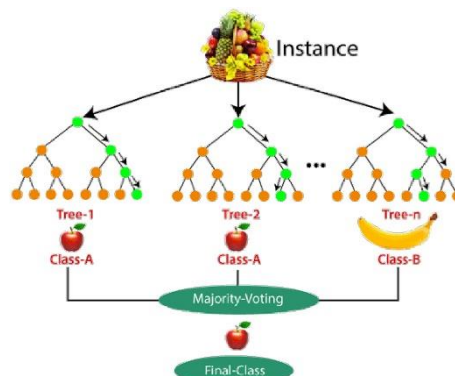
Ứng dụng:

- **Phân loại hình ảnh:** Phân loại hình ảnh thành các danh mục khác nhau dựa trên các đặc trưng của hình ảnh.
- **Dự đoán sở thích người dùng:** Dùng để đề xuất sản phẩm hoặc dịch vụ dựa trên sở thích của người dùng tương tự.
- **Phát hiện bất thường:** Giúp nhận diện các mẫu bất thường trong dữ liệu, như trong phát hiện gian lận.

2.4. Thuật toán Random Forest

Random Forest là một mô hình ensemble learning dựa trên nhiều cây quyết định (decision trees). Mỗi cây quyết định được huấn luyện trên một tập con ngẫu nhiên của dữ liệu, và quyết định cuối cùng được đưa ra dựa trên sự đồng thuận của tất cả các cây.

Random Forest sử dụng phương pháp bagging (bootstrap aggregating) để giảm thiểu sự đa dạng của các cây và tăng độ chính xác của mô hình.



Hình 4. Ví dụ thuật toán Random Forest

Ưu điểm:

- Giảm thiểu hiện tượng overfitting so với các cây quyết định đơn lẻ.
- Có khả năng xử lý các biến số có tính chất khác nhau và dữ liệu thiếu.

Hạn chế:

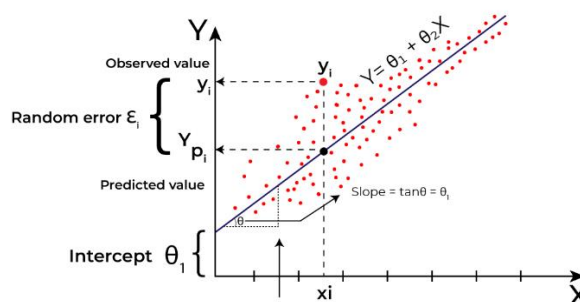
- Mô hình có thể trở nên rất phức tạp và khó hiểu.
- Tốn nhiều tài nguyên tính toán khi số lượng cây lớn.

Ứng dụng:

- **Chẩn đoán y khoa:** Dự đoán bệnh dựa trên dữ liệu y khoa của bệnh nhân.
- **Phân loại hình ảnh:** Dùng trong các bài toán nhận diện đối tượng và phân loại hình ảnh.
- **Phân tích thị trường:** Dự đoán xu hướng thị trường và phân tích tài chính.

2.5. Thuật toán Linear Regression

Hồi quy tuyến tính dự đoán mối quan hệ giữa hai biến bằng cách giả định rằng chúng có mối quan hệ tuyến tính. Nó tìm ra đường thẳng tốt nhất để giảm thiểu sự khác biệt giữa giá trị dự đoán và giá trị thực tế. Được sử dụng trong các lĩnh vực như kinh tế và tài chính, nó giúp phân tích và dự báo các xu hướng dữ liệu. Hồi quy tuyến tính cũng có thể liên quan đến nhiều biến (hồi quy tuyến tính đa biến) hoặc được điều chỉnh cho các câu hỏi có/không (hồi quy logistic).



Hình 5. Ví dụ thuật toán Linear Regression

Ưu điểm:

- Đơn giản và dễ hiểu.
- Yêu cầu tài nguyên tính toán thấp.

Hạn chế:

- Chỉ mô hình hóa mối quan hệ tuyến tính, không phù hợp với dữ liệu phi tuyến.
- Dễ bị đa cộng tuyến nếu các biến độc lập có mối quan hệ mạnh với nhau.

Ứng dụng thực tế:

- **Phân tích xu hướng:** Phân tích xu hướng trong dữ liệu thời gian như sự thay đổi của giá cổ phiếu theo thời gian.
- **Dự đoán tiêu dùng:** Dự đoán mức tiêu thụ hàng hóa dựa trên các yếu tố như thu nhập, giá cả, và quảng cáo.

2.6. Đánh giá mô hình hồi quy

2.6.1. Mean Absolute Error (MAE)

Công thức:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Trong đó:

- y_i : là giá trị thực tế.
- \hat{y}_i : là giá trị dự đoán.

Ý nghĩa: MAE đo lường giá trị trung bình của các sai số tuyệt đối giữa giá trị thực tế và giá trị dự đoán, cho biết mức độ chính xác của dự đoán.

2.6.2. Mean Squared Error (MSE)

Công thức:

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

Trong đó:

- y_i : là giá trị thực tế.
- \hat{y}_i : là giá trị dự đoán.

Ý nghĩa: MSE đo lường trung bình bình phương của các sai số, nhấn mạnh các sai số lớn hơn.

2.6.3. R-squared (R^2)

Công thức:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Trong đó:

- y_i : là giá trị thực tế.
- \hat{y}_i : là giá trị dự đoán.
- \bar{y} : là giá trị trung bình.

Ý nghĩa: R^2 đo lường tỷ lệ biến thiên của dữ liệu thực tế được giải thích bởi mô hình, nằm trong khoảng từ 0 đến 1, giá trị gần 1 chỉ ra mô hình phù hợp tốt với dữ liệu.

2.6.4. Mean Absolute Percentage Error (MAPE)

Công thức:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Trong đó:

- y_i : là giá trị thực tế.
- \hat{y}_i : là giá trị dự đoán.

Ý nghĩa: MAPE đo lường phần trăm sai số trung bình tuyệt đối, giúp dễ hiểu và so sánh các mô hình trên các tập dữ liệu khác nhau.

3. KẾT QUẢ VÀ THẢO LUẬN

Trong phần này, chúng tôi sẽ trình bày các bước tiền xử lý dữ liệu do thư viện yfinance cung cấp và các kết quả thực nghiệm với dữ liệu.

3.1. Dữ liệu

3.1.1. Tiền xử lý dữ liệu

Trong bài báo này, chúng tôi sử dụng dữ liệu giao dịch của công ty Apple do yfinance cung cấp. Cụ thể, dữ liệu bao gồm 10958 ngày giao dịch và 7 thuộc tính bao gồm: Open, High, Low, Close, Volume, Dividends và Stock Splits trong đó thuộc tính Close sẽ là nhãn.

Để áp dụng các thuật toán hồi quy chính xác hơn, chúng tôi sẽ thêm và loại bỏ các thuộc tính cần thiết và không cần thiết. Ngoài ra chúng tôi cũng sẽ chỉ lấy dữ liệu từ năm 01/01/1990 đến 01/06/2024 cho tập dữ liệu.

Chúng tôi sẽ bỏ các thuộc tính Dividends và Stock Splits. Vì 2 thuộc tính này đa số đều bằng 0 nên ít tác động đến quá trình hồi quy. Đồng thời thêm các thuộc tính MA10, MA20, MA50, MA100, MA200, Date và Tomorrow.

Trong đó:

- Date: là ngày giao dịch, chúng tôi sẽ chuyển sang giây. Trong chứng khoán sẽ có những chu kỳ kinh tế sẽ ảnh hưởng đến giá của chứng khoán. Ví dụ trong một năm, vào tháng 5 người ta thường có câu “Sell in may and go away” hoặc vào tháng 11, 12 và 1 thường được gọi là hiện tượng Santa Claus Rally, nghĩa là sẽ tăng giá liên tục trong những tuần cuối năm, tất nhiên đây cũng chỉ là xác suất. Ngoài ra còn có các chu kỳ dài hơn như 3, 5, 10, 20 năm.
- MA10: là trung bình cộng của giá trong 10 ngày gần nhất.
- MA20: là trung bình cộng của giá trong 20 ngày gần nhất.
- MA50: là trung bình cộng của giá trong 50 ngày gần nhất.
- MA100: là trung bình cộng của giá trong 100 ngày gần nhất.
- MA200: là trung bình cộng của giá trong 200 ngày gần nhất.
- Tomorrow: là giá đóng cửa của ngày tiếp theo. Cho nên dữ liệu cuối cùng sẽ bị trống thuộc tính này và đây cũng là giá trị mà cần dự đoán.

MA10 và MA20 phản ánh xu hướng ngắn hạn.

MA50, MA100, MA200 phản ánh xu hướng trung và dài hạn.

Sau khi đã có đầy đủ các thuộc tính, theo như chúng tôi quan sát về dữ liệu có những nhận xét sau đây:

1. Sự chênh lệch giữa dữ liệu của các thuộc tính là quá lớn. Đặc biệt là Volume và Date so với các thuộc tính còn lại.
2. Sự chênh lệch giữa 2 điểm dữ liệu giữa hai ngày gần kề thì quá nhỏ. Còn so với dữ liệu tổng thể từ năm 1990 đến 2024 lại rất lớn (do sự tăng giá của cổ phiếu).

Để giải quyết vấn đề này chúng tôi sử dụng MinMaxScaler để chuẩn hóa dữ liệu:

MinMaxScaler là một phương pháp chuẩn hóa dữ liệu trong đó các giá trị của một tập dữ liệu được biến đổi để nằm trong một phạm vi xác định, thường là từ 0 đến 1.

Công thức:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Trong đó:

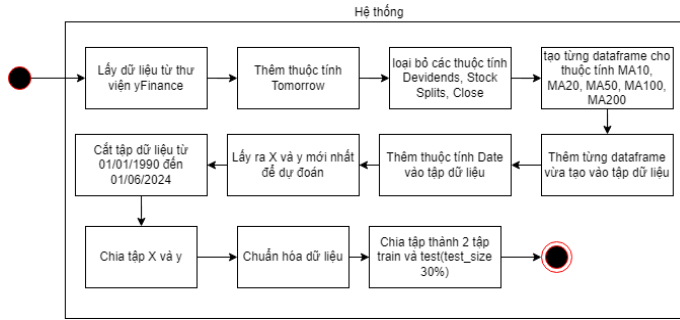
X là giá trị gốc của dữ liệu.

X_{\min} là giá trị nhỏ nhất của tập dữ liệu.

X_{\max} là giá trị lớn nhất của tập dữ liệu.

X_{scaled} là giá trị đã được chuẩn hóa.

Sau đây là quá trình tiền xử lý dữ liệu của chúng tôi:



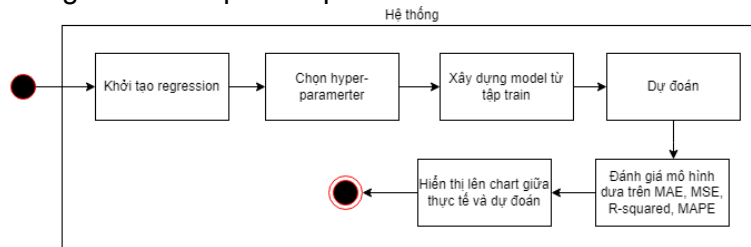
Hình 6. Mô tả từng bước tiền xử lý dữ liệu

3.1.2. Dữ liệu sau khi tiền xử lý

Sau khi áp dụng các bước trên để tiền xử lý dữ liệu, dữ liệu của chúng tôi bao gồm 8670 ngày giao dịch và 11 thuộc tính trong đó thuộc tính Tomorrow sẽ là nhãn.

3.2. Kết quả

Quy trình đánh giá mô hình phân lớp:



Hình 7. Mô tả từng bước quá trình đánh giá mô hình

Kết quả so sánh:

Thuật toán	MSE	MAE	R^2	MAPE
MLP	0	0.006	0.998	2.131
Linear Regression	0	0.002	1	0.035
Random Forest	0	0.002	1	0.035
SVM	0.005	0.071	0.909	62.597
KNN	0	0.002	0.999	0.105

Bảng 1. Hiệu suất của từng mô hình

MSE = 0 (do làm tròn), thực tế MSE bằng 0 là không khả thi. MSE = 0 cho thấy rằng mô hình đã khớp hoàn toàn với dữ liệu.

MAE của rất thấp thể hiện giá trị trung bình của các sai số tuyệt đối là rất nhỏ. SVM cao nhất.

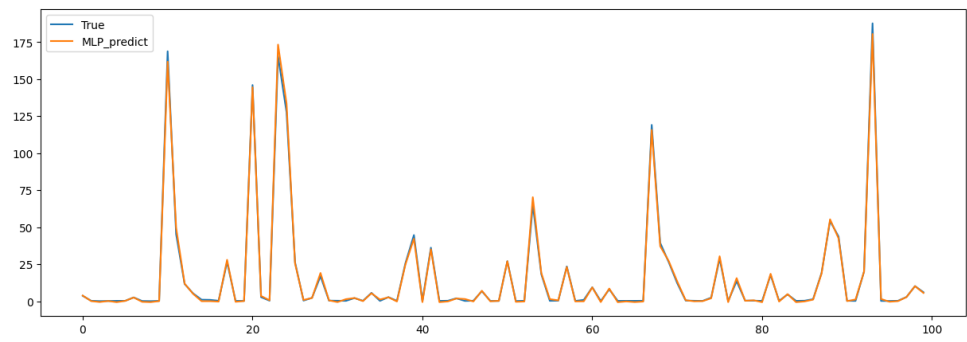
R^2 gần bằng 1 cho thấy mô hình thể hiện hoàn toàn sự biến thiên của dữ liệu

MAPE thấp cũng thể hiện mô hình có độ chính xác cao. SVM cao nhất.

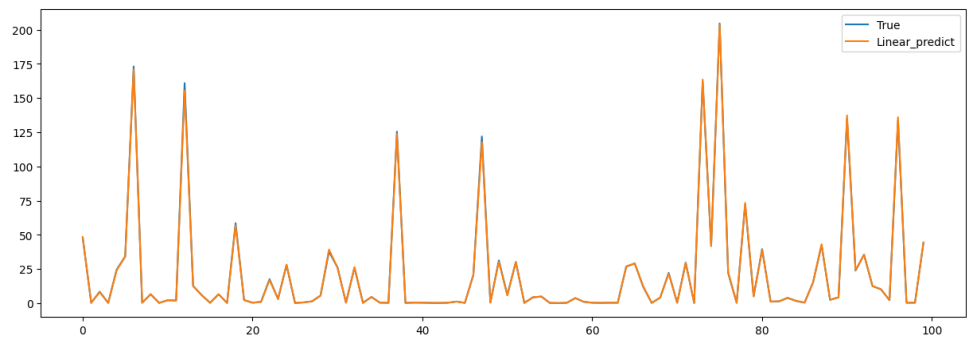
Theo kết luận của các chỉ số trên, chúng ta có thể thấy rằng Linear Regression và Random Forest là mô hình tốt nhất theo sau đó là KNN, MLP và SVM.

Do đó trong sản phẩm ứng dụng của đề tài chúng tôi đã sử dụng thuật toán Linear Regression hoặc Random Forest để dự đoán giá của cổ phiếu.

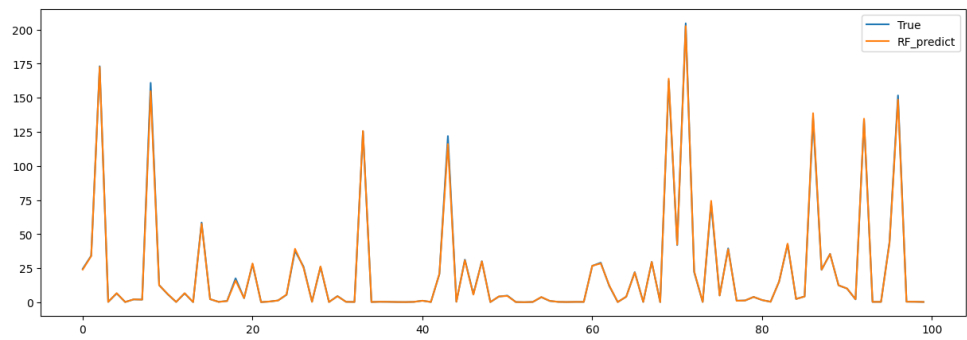
Để có thể nhìn rõ hiệu suất của mô hình chúng tôi đã ... hóa giữa giá trị thực tế và giá trị dự đoán như sau (chúng tôi lấy 100 mẫu bất kỳ vì do tập test lớn nếu plot lên sẽ không thể nhìn được):



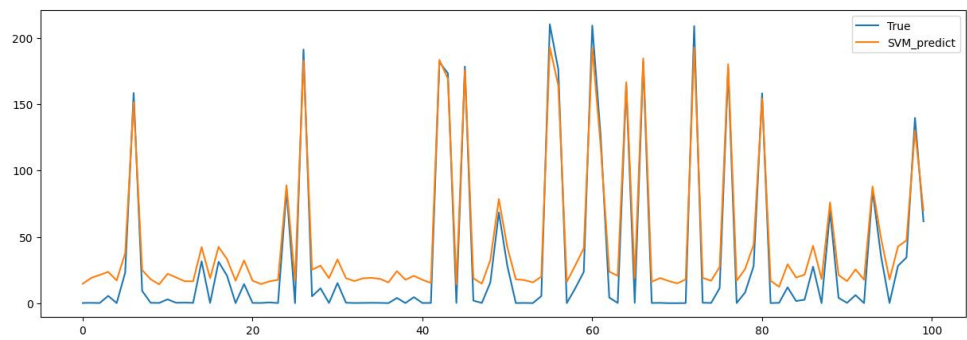
Hình 8: Thuật toán MLP



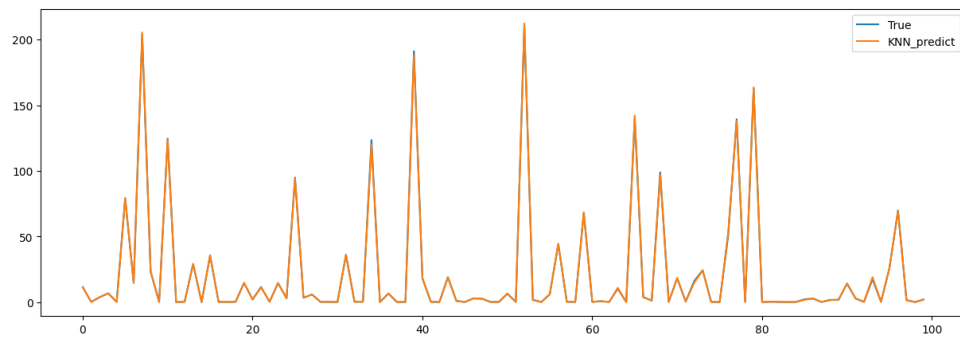
Hình 9: Thuật toán Linear Regression



Hình 10: Thuật toán Random Forest



Hình 11: Thuật toán SVM



Hình 8: Thuật toán KNN

4. KẾT LUẬN

Bài báo này của chúng tôi đã trình bày một số kỹ thuật hồi quy để ứng dụng vào việc dự đoán giá cổ phiếu của công ty Apple. Dữ liệu của chúng tôi lấy từ thư viện yFinance, sau đó dữ liệu sẽ được xử lý trước khi đưa vào các mô hình. Các thuật toán được sử dụng là MLP, Linear Regression, Random Forest, SVM và KNN. Kết quả thực nghiệm cho thấy tính khả thi của việc áp dụng mô hình hồi quy vào bài toán dự đoán giá cổ phiếu. Kết quả cho thấy hai thuật toán Linear Regression và Random Forest có kết quả tốt nhất và SVM có kết quả tệ nhất.

TÀI LIỆU THAM KHẢO

<https://maybank-kimeng.com.vn/kimengportal/news/printNewToPdf.do?id=182125#:~:text=Ch%E1%BB%A9ng%20kho%C3%A1n%20C4%91%C6%B0%E1%BB%A3c%20coi%20l%C3%A0,s%C3%A1ch%20kinh%20t%E1%BA%BF%20v%C4%A9%20m%C3%B4>

<https://vtv.vn/kinh-te/hon-132000-tai-khoan-chung-khoan-moi-duoc-mo-trong-thang-5-2024060917583746.htm>

<https://tapchitaichinh.vn/co-hon-7-9-trieu-tai-khoan-chung-khoan-duoc-mo.html>

<https://chatgpt.com/share/f5c1f2a6-075e-4018-a2d2-c964dfddab32>

<https://www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning>

<https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/>

<https://www.datacamp.com/tutorial/k-nearest-neighbors-knn-classification-with-r-tutorial>

<https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>

<https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/>

<https://www.geeksforgeeks.org/ml-linear-regression/>

<https://chatgpt.com/share/f2f62669-f6a8-4fd4-b7d4-552cadcecbcc>

<https://thuvienphapluat.vn/lao-dong-tien-luong/thi-truong-chung-khoan-la-gi-thi-truong-chung-khoan-co-tac-dong-gi-den-nguoi-lao-dong-10751.html>