

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ
HỌC MÁY**

Người thực hiện: **PHẠM NHỰT MINH – 21021471**

NGUYỄN KHẮC LUẬT – 21099741

Lớp : 420300371101

Khoá : 17

Người hướng dẫn: **TS BÙI THANH HÙNG**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ
HỌC MÁY**

Người thực hiện: **PHẠM NHỰT MINH – 21021471**

NGUYỄN KHẮC LUẬT – 21099741

Lớp : 420300371101

Khoá : 17

Người hướng dẫn: **TS BÙI THANH HÙNG**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

LỜI CẢM ƠN

Đồ án được sự thực hiện bởi nhóm 13 gồm 2 thành viên (Phạm Nhật Minh) và Nguyễn Khắc Luật) lớp DHKHD17A dưới sự hướng dẫn của Thầy Bùi Thanh Hùng giảng viên môn Máy học bộ môn Khoa học dữ liệu. Thầy đã tạo điều kiện, giúp đỡ về kiến thức và hỗ trợ các thắc mắc về các vấn đề trong quá trình thực hiện để chúng tôi hoàn thành đồ án.

Chúng tôi muốn gửi lời cảm ơn thầy Bùi Thanh Hùng – giảng viên môn Máy học, đã đưa ra đề tài cuối kì hết sức hấp dẫn. Và cũng cảm ơn thầy đã cung cấp cho chúng tôi những kiến thức quý giá thông qua các tiết học, bài tập để hoàn thành đề tài này. Những gì thầy đã truyền tải giúp tôi hiểu sâu hơn về chủ đề và phát triển các kỹ năng cần thiết để giải quyết các vấn đề liên quan đến đồ án.

Chúng tôi cũng muốn gửi lời cảm ơn đến đồng đội của chúng em trong đề tài. Bạn đã hỗ trợ tôi trong việc tìm kiếm tài liệu, trao đổi ý kiến và đóng góp ý kiến quý báu để đưa ra những giải pháp tối ưu cho đề tài. Chúng tôi đã học hỏi được rất nhiều kinh nghiệm từ nhau và sẽ luôn nhớ đến những kỷ niệm đẹp trong quá trình làm việc cùng nhau.

Một lần nữa, xin chân thành cảm ơn thầy và đồng đội. Sự giúp đỡ và đóng góp của mọi người là cốt lõi để hoàn thành đề tài này.

ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH

Chúng em xin cam đoan đây là sản phẩm đồ án của riêng chúng em và được sự hướng dẫn của TS. Bùi Thanh Hùng. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính chúng em thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào chúng em xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình. Trường Đại học Công nghiệp TP Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do chúng em gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày 20 tháng 05 năm 2024

Tác giả

Phạm Nhựt Minh

Nguyễn Khắc Luật

PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

Tp. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

TÓM TẮT

Bài toán dự đoán thị trường chứng khoán Việt Nam đặt ra nhiều thách thức cho các nhà nghiên cứu và nhà đầu tư. Vấn đề chính là dự đoán biến động giá cả cổ phiếu trong tương lai dựa trên thông tin có sẵn và các yếu tố ảnh hưởng. Các nghiên cứu thường tập trung vào việc sử dụng các mô hình thống kê, học máy, hoặc trí tuệ nhân tạo để phân tích dữ liệu lịch sử và các chỉ số kinh tế tài chính, kỹ thuật giao dịch, tin tức và tâm lý thị trường để tạo ra các dự đoán có thể giúp nhà đầu tư đưa ra quyết định đầu tư thông minh.

Có một số hướng tiếp cận phổ biến trong dự đoán thị trường chứng khoán Việt Nam như:

- Cơ bản: Sử dụng các chỉ số tài chính, dữ liệu kinh tế và yếu tố kinh tế-xã hội để phân tích tỷ lệ giá cổ phiếu so với lợi nhuận, giá trị thị trường so với thu nhập, và các chỉ số tài chính khác.
- Kỹ thuật: Phân tích biểu đồ giá cổ phiếu và các chỉ báo kỹ thuật như MACD, RSI, Bollinger Bands để dự đoán xu hướng giá tương lai.
- Dựa trên tin tức: Sử dụng thông tin từ các nguồn tin tức tài chính, kinh tế và chính trị để đánh giá ảnh hưởng của các sự kiện và tin tức đối với thị trường chứng khoán, bao gồm phân tích cảm xúc và tâm lý thị trường.
- Học máy và trí tuệ nhân tạo: Sử dụng các mô hình học máy như mạng nơ-ron và SVM để phân tích dữ liệu lịch sử và tạo ra các mô hình dự đoán cho thị trường chứng khoán.

Chúng em dùng sử dụng phương pháp học máy để dự đoán thị trường chứng khoán theo các bước:

- Thu thập dữ liệu
- Tiền xử lý dữ liệu
- Xây dựng mô hình
- Đánh giá mô hình
- Tinh chỉnh mô hình
- Triển khai mô hình
- Theo dõi và cập nhật

Chúng em nhận được một số kết quả:

- Các độ đo thể hiện độ chính xác của mô hình cho thấy độ chính xác của dữ liệu dự đoán.
- Các kết quả dự đoán được từ mô hình, thấy được sự biến động của thị trường chứng khoán để ra quyết định đối với các nhà nghiên cứu và nhà đầu tư.
- Cung cấp thông tin chi tiết về mức độ ảnh hưởng của từng yếu tố đến giá cổ phiếu, giúp hiểu rõ hơn về các yếu tố tác động đến thị trường chứng khoán Việt Nam.

Những phát hiện cơ bản:

- Mô hình có hiệu suất ở mức trung bình, phản ánh khả năng dự đoán giá cổ phiếu với một mức độ chính xác chấp nhận được, nhưng vẫn còn sai số đáng kể. Các mô hình cần phải được cải thiện để tăng độ chính xác cho khả năng dự đoán.
- Tạo và lựa chọn các đặc trưng mới từ dữ liệu gốc, bao gồm các chỉ số kỹ thuật, các biến trễ, và các biến tương tác để cải thiện hiệu suất mô hình.

- Cho thấy tầm quan trọng của từng yếu tố đối với giá cổ phiếu. Một số yếu tố có ảnh hưởng lớn hơn, chẳng hạn như giá cổ phiếu ngày trước, khối lượng giao dịch, và các chỉ số tài chính quan trọng.

MỤC LỤC

LỜI CẢM ƠN	1
PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN	3
TÓM TẮT	4
MỤC LỤC.....	7
DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT	9
DANH MỤC CÁC HÌNH VẼ.....	10
DỰ ĐOÁN THỊ TRƯỜNG CHỨNG KHOÁN BẰNG CÁC PHƯƠNG PHÁP HỌC MÁY.....	11
1.1. Giới thiệu về bài toán	11
1.2. Cách thức tiến hành thu thập dữ liệu	11
1.2.1. Yêu cầu của bài toán	11
1.2.2. Các phương pháp giải quyết bài toán	12
1.2.3 Phương pháp đề xuất giải quyết bài toán.....	16
1.3 Phương pháp giải quyết bài toán.....	17
1.3.1 Mô hình tổng quát.....	17
1.3.2 Đặc trưng của mô hình đề xuất	20
1.4 Thực nghiệm	21
1.4.1 Dữ liệu.....	21
1.4.2 Xử lý dữ liệu	22
1.4.3 Công nghệ sử dụng	22
1.4.4 Cách đánh giá.....	23
1.5 Kết quả đạt được	23
1.6 Kết luận	25
1.6.1 Linear Regression	25
1.6.2 LSTM.....	28
LÀM VIỆC NHÓM	30

TÀI LIỆU THAM KHẢO.....	32
TỰ ĐÁNH GIÁ.....	33

DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT

CÁC KÍ HIỆU

$r^2 = r - \text{squared}$

CÁC CHỮ VIẾT TẮT

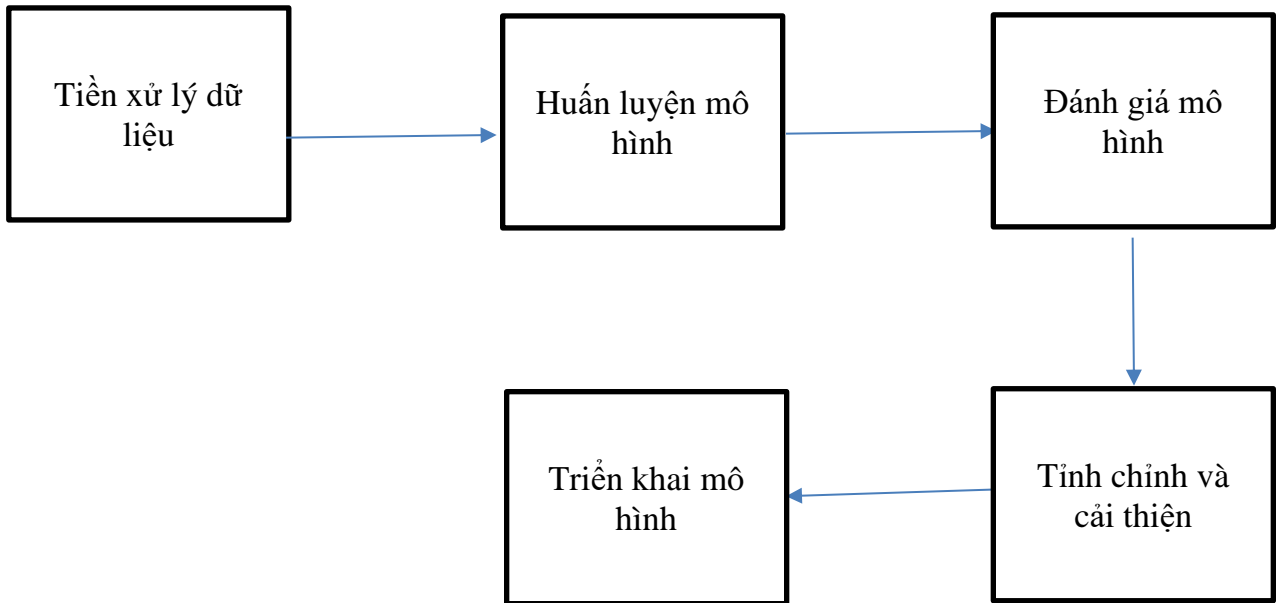
mse: Mean Squared Error

mae: Mean Absolute Error

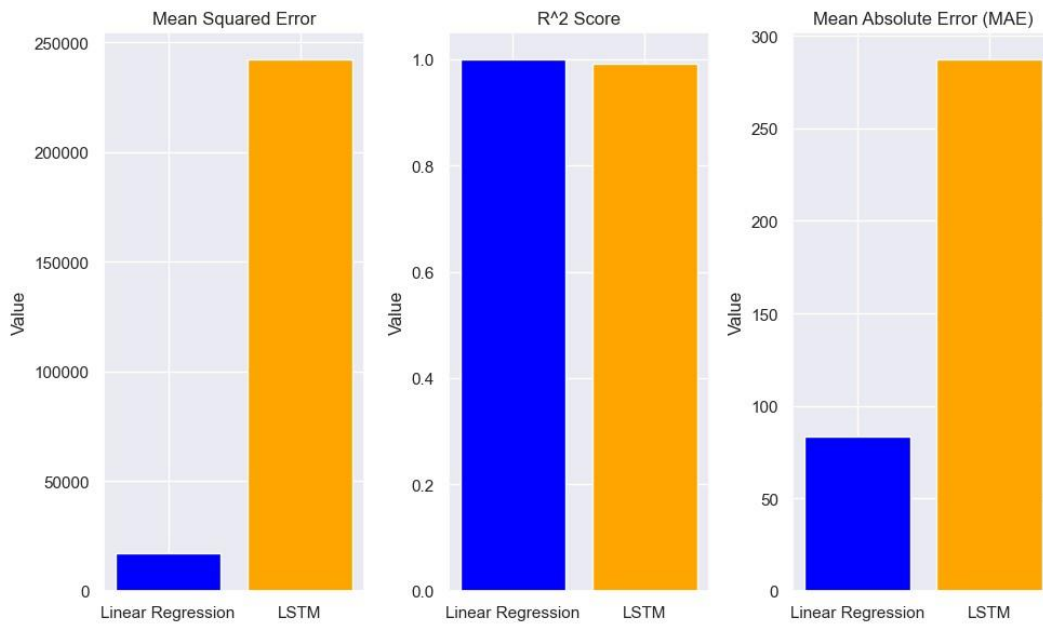
LSTM: Long Short Term Memory

DANH MỤC CÁC HÌNH VẼ

Hình 1.3.1. Mô hình tổng quát



Hình 1.5. Biểu đồ so sánh giữa các phương pháp theo các độ đo



DỰ ĐOÁN THỊ TRƯỜNG CHỨNG KHOÁN BẰNG CÁC PHƯƠNG PHÁP HỌC MÁY

1.1. Giới thiệu về bài toán

- Bài toán “Dự đoán thị trường chứng khoán bằng các phương pháp học máy” là một bài toán trong lĩnh vực dự đoán tài chính và thị trường chứng khoán. Bài toán này đặt ra mục tiêu sử dụng các phương pháp học máy để dự đoán giá cổ phiếu hoặc chỉ số thị trường trong tương lai dựa trên dữ liệu lịch sử và các yếu tố tác động. Mục tiêu cuối cùng là tạo ra một mô hình có khả năng dự đoán giá cổ phiếu với độ chính xác cao, tuy nhiên, thị trường chứng khoán là một môi trường rất phức tạp và không chắc chắn, do đó không có mô hình nào có thể dự đoán mọi biến động của thị trường một cách hoàn hảo.

- Ý nghĩa của bài toán này rất lớn đối với các nhà đầu tư và quản lý tài sản. Bằng cách sử dụng các mô hình học máy, họ có thể cố gắng dự đoán hướng diễn biến của thị trường chứng khoán, từ đó đưa ra quyết định đầu tư thông minh và tối ưu hóa lợi nhuận. Nếu có thể dự đoán được xu hướng của thị trường, các nhà đầu tư có thể đưa ra các quyết định mua, bán hoặc giữ cổ phiếu một cách hiệu quả hơn, giảm thiểu rủi ro và tối đa hóa sinh lợi. Tuy nhiên, việc dự đoán thị trường chứng khoán là một nhiệm vụ khó khăn do sự biến động và không chắc chắn của thị trường. Các dự đoán không thể hoàn toàn chính xác và luôn có yếu tố rủi ro.

1.2. Cách thức tiến hành thu thập dữ liệu

1.2.1. Yêu cầu của bài toán

Bài toán dự đoán thị trường chứng khoán thông qua các phương pháp học máy đặt ra có thể bao gồm các yêu cầu sau:

- Dự đoán hướng thị trường: Mục tiêu chính là dự đoán xem giá chứng khoán sẽ tăng, giảm hoặc duy trì ổn định trong một khoảng thời gian nhất định trong tương lai.

- Tính chính xác và đáng tin cậy của mô hình: Mô hình cần có khả năng dự đoán chính xác và đáng tin cậy, để nhà đầu tư có thể dựa vào dự đoán để đưa ra quyết định giao dịch.

- Xử lý dữ liệu: Yêu cầu sử dụng các kỹ thuật xử lý dữ liệu để loại bỏ nhiễu, xử lý dữ liệu thiếu, và chuẩn hóa dữ liệu để đảm bảo tính nhất quán và đồng nhất cho việc huấn luyện mô hình.

- Chọn và xây dựng mô hình phù hợp: Yêu cầu chọn lựa và xây dựng các mô hình học máy phù hợp hoặc các mô hình tổ hợp, dựa trên tính chất của dữ liệu và mục tiêu dự đoán.

- Tính ứng dụng thực tế: Mô hình cần phải có tính ứng dụng thực tế, có khả năng triển khai và sử dụng trong môi trường thực tế của thị trường chứng khoán.

1.2.2. Các phương pháp giải quyết bài toán

Có một số phương pháp phổ biến được các tác giả áp dụng để giải quyết bài toán này

Phương pháp 1: Dùng Máy vector hỗ trợ (SVM) và Phân tích thành phần độc lập (ICA)[1]:

Phương pháp giải quyết bài toán:

- Máy Vector Hỗ Trợ (SVM):

SVM là một phương pháp học máy phổ biến được sử dụng trong việc phân loại và dự đoán. SVM cố gắng tìm ra ranh giới quyết định tốt nhất giữa các điểm dữ liệu thuộc vào các lớp khác nhau.

Trong bài toán dự đoán thị trường chứng khoán, SVM có thể được áp dụng để dự đoán hướng diễn biến của giá cổ phiếu hoặc để phân loại các biến động của thị trường (tăng, giảm, ổn định).

- Phân Tích Thành Phần Độc Lập (ICA):

ICA là một kỹ thuật phân tích tín hiệu thống kê được sử dụng để phân tích một tín hiệu thành các thành phần độc lập, mỗi thành phần đại diện cho một tín hiệu độc lập.

Trong bài toán dự đoán thị trường chứng khoán, ICA có thể được sử dụng để phân tích và tách riêng các yếu tố tác động vào giá cổ phiếu từ dữ liệu phức tạp. Bằng cách này, ICA có thể giúp xác định và nắm bắt các mô hình ẩn đằng sau sự biến động của thị trường chứng khoán một cách hiệu quả.

Tên bài báo: Grigoryan, H. (2016). A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA). Database Systems Journal, 7(1).

Phương pháp giải quyết:

Phương pháp giải quyết được đề xuất trong bài báo này là kết hợp hai kỹ thuật chính: Máy Vector Hỗ Trợ (SVM) và Phân Tích Thành Phần Độc Lập (ICA).

- Trích xuất đặc trưng với ICA:

Sử dụng Phân Tích Thành Phần Độc Lập (ICA) để trích xuất các đặc trưng quan trọng từ dữ liệu nghiên cứu. ICA được áp dụng để chuyển đổi các tín hiệu gốc thành các thành phần độc lập giả định và phi Gaussian.

- Dự đoán với SVM:

Sử dụng Máy Vector Hỗ Trợ (SVM) để thực hiện dự đoán chuỗi thời gian dựa trên các đặc trưng được trích xuất từ ICA.

SVM được sử dụng để xây dựng mô hình dự đoán dựa trên các đặc trưng đã chọn và dữ liệu lịch sử.

- So Sánh và Đánh Giá:

Kết quả từ phương pháp SVM-ICA được so sánh với kết quả từ mô hình dựa trên SVM mà không sử dụng bất kỳ bước tiền xử lý nào.

Sử dụng độ đo sai số trung bình bình phương căn (RMSE) để đánh giá hiệu suất của các mô hình.

Dữ liệu thực nghiệm:

Bài báo sử dụng hai tập dữ liệu thực nghiệm từ thị trường chứng khoán khác nhau:

- Tập dữ liệu đầu tiên được lấy từ Sở giao dịch chứng khoán Bucharest và bao gồm các quan sát hàng tuần trong giai đoạn từ 3/9/2008 đến 11/30/2014, tổng cộng 350 quan sát. Dữ liệu này bao gồm 39 biến số, trong đó có 35 biến được lựa chọn từ phân tích kỹ thuật và 4 biến số kinh tế cơ bản được lấy từ phân tích cơ bản của cổ phiếu OMV Petrom.

- Tập dữ liệu thứ hai được lấy từ Sở giao dịch chứng khoán Baltic và bao gồm các quan sát hàng ngày trong giai đoạn từ 03/12/2012 đến 12/30/2014, tổng cộng 700 quan sát. Dữ liệu này chỉ bao gồm 35 biến số được lựa chọn từ phân tích kỹ thuật của cổ phiếu Tallink.

Kết quả đạt được:

- Kết quả của phương pháp SVM-ICA được đánh giá dựa trên độ đo sai số trung bình bình phương căn (RMSE).

- Kết quả thực nghiệm cho thấy rằng mô hình SVM-ICA vượt trội hơn so với mô hình SVM đơn giản trong việc dự đoán giá cổ phiếu trên cả hai tập dữ liệu.

- RMSE của mô hình SVM-ICA thấp hơn so với RMSE của mô hình SVM đơn giản, cho thấy hiệu suất dự đoán của phương pháp kết hợp SVM-ICA là tốt hơn.

Phương pháp 2: Dừng Random Forest [2]

- Random Forest là một thuật toán dựa trên mô hình cây quyết định (Decision Tree).

- Thuật toán này tạo ra một "rừng" (tập hợp) các cây quyết định dựa trên một tập dữ liệu huấn luyện bằng cách lấy mẫu ngẫu nhiên từ tập dữ liệu huấn luyện và từng cây trong rừng được huấn luyện độc lập.

- Mỗi cây quyết định trong rừng sẽ đưa ra một dự đoán riêng và kết quả cuối cùng là kết hợp của các dự đoán từ các cây, thông qua việc sử dụng phương pháp đa số hoặc trung bình.

Tên bài báo: Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. arXiv preprint arXiv:1605.00003.

Phương pháp giải quyết:

Bài báo trình bày một phương pháp giải quyết bài toán dự đoán xu hướng thị trường chứng khoán sử dụng một mô hình phân loại rừng ngẫu nhiên (random forest classifier). Dưới đây là phương pháp giải quyết được mô tả trong bài báo:

- Thu thập dữ liệu: Tác giả thu thập dữ liệu thị trường chứng khoán từ các nguồn đáng tin cậy, chẳng hạn như dữ liệu từ Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), và Samsung.

- Tiền xử lý dữ liệu: Trước khi sử dụng dữ liệu cho mô hình, tác giả thực hiện các bước tiền xử lý như làm mịn dữ liệu (exponential smoothing) để loại bỏ sự biến động ngẫu nhiên và làm cho quá trình học trở nên dễ dàng hơn.

- Xây dựng mô hình phân loại rừng ngẫu nhiên: Tác giả sử dụng một mô hình phân loại rừng ngẫu nhiên để dự đoán xu hướng thị trường chứng khoán. Mô hình này được xây dựng dựa trên việc sử dụng nhiều cây quyết định (decision trees) trong đó mỗi cây được huấn luyện trên một phần của dữ liệu.

- Đánh giá hiệu suất: Sau khi huấn luyện mô hình, tác giả đánh giá hiệu suất của mô hình bằng cách sử dụng các thước đo như độ chính xác (accuracy), độ nhớ (recall), độ chính xác (precision), và đặc trưng (specificity).

- So sánh với các nghiên cứu trước đó: Tác giả so sánh kết quả của mô hình đề xuất với các nghiên cứu trước đó trong lĩnh vực dự báo thị trường chứng khoán. Qua đó, họ chứng minh rằng mô hình của họ vượt trội hơn so với các mô hình đã được nghiên cứu trước đó.

Dữ liệu thực nghiệm:

Tác giả đã sử dụng dữ liệu từ các công ty lớn như Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), và Samsung để xây dựng và đánh giá mô hình dự đoán xu hướng thị trường chứng khoán.

Kết quả đạt được:

Mô hình phân loại rừng ngẫu nhiên đã đạt được độ chính xác cao, trong khoảng từ 85% đến 95% cho dự báo dài hạn.

Các phép đo hiệu suất khác như precision, recall và specificity cũng cho thấy mô hình hoạt động hiệu quả.

So sánh với các nghiên cứu trước đó cho thấy mô hình đề xuất vượt trội hơn với kết quả tốt hơn.

1.2.3 Phương pháp đề xuất giải quyết bài toán

Hướng giải quyết bài toán dự đoán thị trường chứng khoán trong đề tài này của chúng em sử dụng mô hình LSTM (Long Short-Term Memory) và mô hình Linear Regression.

- Lý do chọn LSTM:

+ Xử lý dữ liệu chuỗi: LSTM được thiết kế đặc biệt để xử lý dữ liệu chuỗi, như giá cổ phiếu theo thời gian. Cấu trúc nội tại của LSTM giúp nó nhớ được thông tin từ các quãng thời gian trước đó trong quá trình dự đoán, làm cho nó phù hợp cho việc dự đoán trong môi trường thị trường chứng khoán đầy biến động.

+ Khả năng học được mối quan hệ phức tạp: LSTM có khả năng học được mối quan hệ phi tuyến tính giữa các biến đầu vào và đầu ra. Trên thực tế, thị trường chứng khoán thường chứa nhiều yếu tố phức tạp và không tuyến tính, điều này làm cho LSTM trở thành một lựa chọn hợp lý để mô hình hóa các mối quan hệ này.

+ Khả năng xử lý dữ liệu lớn: LSTM có thể xử lý dữ liệu lớn một cách hiệu quả, bao gồm cả dữ liệu lịch sử lâu dài của thị trường chứng khoán. Điều này cho phép nó

học được các xu hướng dài hạn và ngắn hạn từ dữ liệu phong phú, giúp tăng tính khả quan và chính xác của dự đoán.

- Lý do chọn Linear Regression:

+ Đơn giản và dễ hiểu: Linear Regression là một trong những thuật toán đơn giản nhất trong học máy. Nó dễ hiểu và dễ triển khai, giúp người dùng có cái nhìn rõ ràng về mối quan hệ giữa các biến độc lập (các yếu tố ảnh hưởng đến giá cổ phiếu) và biến phụ thuộc (giá cổ phiếu).

+ Hiệu quả tính toán: Linear Regression không đòi hỏi nhiều tài nguyên tính toán như các mô hình phức tạp hơn. Điều này có lợi khi xử lý lượng dữ liệu lớn mà không cần đến hạ tầng phần cứng mạnh mẽ.

+ Khả năng diễn giải mô hình: Các hệ số trong mô hình hồi quy tuyến tính cung cấp thông tin về mức độ ảnh hưởng của từng yếu tố đến giá cổ phiếu. Điều này giúp các nhà phân tích dễ dàng diễn giải và đưa ra quyết định dựa trên mô hình.[5]

+ Phát hiện xu hướng: Linear Regression có thể giúp phát hiện các xu hướng tuyến tính trong dữ liệu, điều này có thể hữu ích trong việc dự đoán xu hướng giá cổ phiếu trong ngắn hạn.

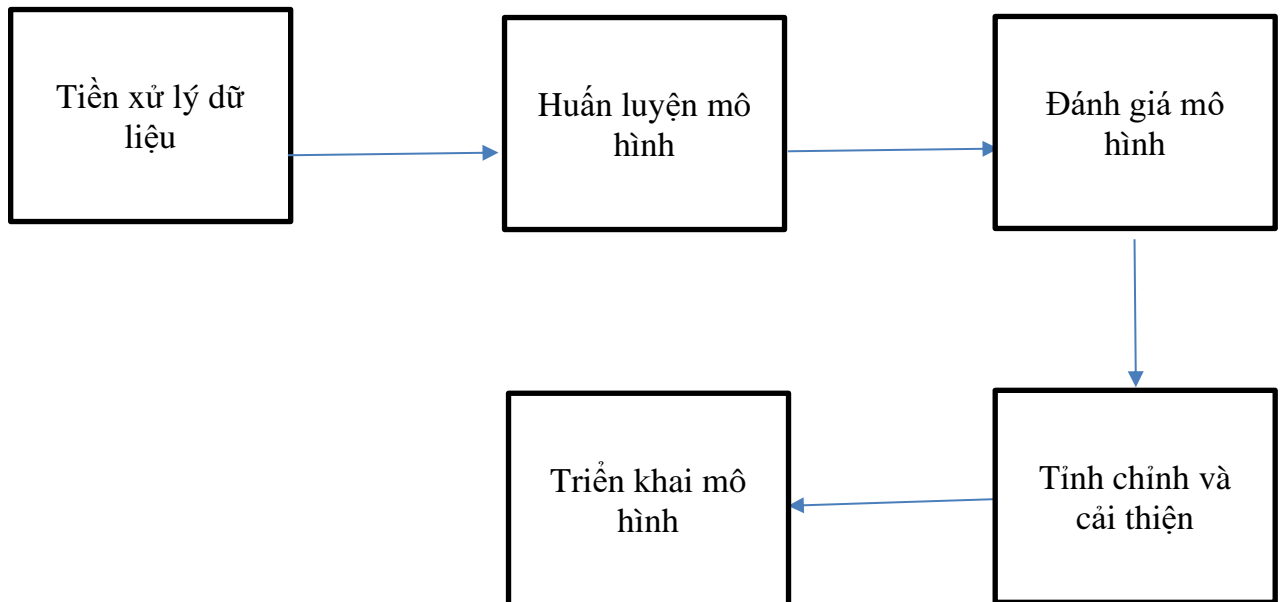
+ Khả năng mở rộng: Mô hình Linear Regression có thể dễ dàng được mở rộng và điều chỉnh bằng cách thêm các biến mới (như các chỉ số kinh tế, tin tức tài chính, etc.) hoặc bằng cách sử dụng các kỹ thuật nâng cao như Polynomial Regression nếu mối quan hệ giữa các biến không hoàn toàn tuyến tính.

+ Dễ dàng kiểm tra giả định và chẩn đoán lỗi: Linear Regression cho phép kiểm tra dễ dàng các giả định của mô hình (như tính tuyến tính, tính độc lập, tính bình thường của sai số, etc.). Điều này giúp phát hiện và khắc phục sớm các vấn đề có thể ảnh hưởng đến độ chính xác của mô hình.[4]

1.3 Phương pháp giải quyết bài toán

1.3.1 Mô hình tổng quát

Vẽ mô hình: *Hình 1.3.1*



Tiền xử lý dữ liệu → Huấn luyện mô hình → Đánh giá mô hình → Tinh chỉnh và cải thiện → Triển khai mô hình

Trình bày tổng quát về mô hình LSTM:

1. Tiền xử lý dữ liệu:

- Thu thập dữ liệu: Sử dụng thư viện vnstock để lấy dữ liệu lịch sử về giá cổ phiếu, khối lượng giao dịch và các chỉ số khác từ 4 ngân hàng của Việt Nam. Khoảng thời gian cần thu thập dữ liệu là từ tháng 5/2014 đến tháng 5/2024.

- Xử lý dữ liệu: Thống kê mô tả, loại bỏ dữ liệu thiếu, chia dữ liệu thành tập huấn luyện và tập kiểm tra.

2. Huấn luyện mô hình LSTM:

- Chia dữ liệu thành tập huấn luyện và tập kiểm tra.
- Xây dựng mô hình LSTM với các lớp LSTM, lớp Dropout để tránh overfitting.
- Huấn luyện mô hình trên tập huấn luyện.

3. Đánh giá mô hình:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)

4. Tinh chỉnh và cải thiện:

- Dựa trên kết quả đánh giá, tiến hành tinh chỉnh mô hình bằng cách điều chỉnh các siêu tham số như số lượng lớp LSTM, số lượng neuron trong mỗi lớp, tỷ lệ dropout, và tốc độ học.[7]

5. Triển khai và theo dõi:

- Triển khai mô hình: Sử dụng mô hình đã huấn luyện để dự đoán thị trường chứng khoán trên dữ liệu của vnstock.

- Theo dõi hiệu suất: Theo dõi hiệu suất của mô hình trên dữ liệu thực tế và điều chỉnh nếu cần thiết.

Trình bày tổng quát về mô hình Linear Regression:

1. Tiền xử lý dữ liệu:

- Thu thập dữ liệu: Sử dụng thư viện vnstock để lấy dữ liệu lịch sử về giá cổ phiếu, khối lượng giao dịch và các chỉ số khác từ 4 ngân hàng của Việt Nam. Khoảng thời gian cần thu thập dữ liệu là từ 2012 đến 2024.

- Xử lý dữ liệu: Thống kê mô tả, loại bỏ dữ liệu thiếu, chia dữ liệu thành tập huấn luyện và tập kiểm tra.

2. Huấn luyện mô hình:

- Chọn các biến 'open', 'high', 'low', 'volume' để làm biến feature (biến đặc trưng). Và biến 'close' sẽ là biến target (mục tiêu)

- Chọn tỉ lệ train/test là 8/2 để chia tập train/set và huấn luyện mô hình

3. Đánh giá mô hình:

- Sử dụng 3 độ đo mse, r^2 và mae để đánh giá mô hình.

4. Tinh chỉnh và cải thiện:

- Tạo thêm các đặc trưng mới để có thể dự đoán được giá một cách chính xác của dữ liệu.

5. Triển khai và theo dõi:

- Triển khai mô hình: Sử dụng mô hình đã huấn luyện để dự đoán thị trường chứng khoán trên dữ liệu của vnstock.

- Theo dõi hiệu suất: Theo dõi hiệu suất của mô hình trên dữ liệu thực tế và điều chỉnh nếu cần thiết.

1.3.2 Đặc trưng của mô hình đề xuất

Mô hình LSTM:

- Cấu trúc mạng nơ-ron:

- + LSTM có một cấu trúc lặp lại trong đó mỗi đơn vị LSTM được lặp lại qua thời gian. Điều này cho phép nó xử lý dữ liệu chuỗi một cách hiệu quả bằng cách lưu trữ và trích xuất thông tin từ các quãng thời gian trước đó.

- + Mỗi đơn vị LSTM bao gồm các cổng (gate) để kiểm soát luồng thông tin, bao gồm cổng quên (forget gate), cổng nhập (input gate), và cổng đầu ra (output gate), giúp điều chỉnh việc ghi và đọc thông tin từ bộ nhớ dài hạn và ngắn hạn.[6]

- Bộ nhớ dài hạn và ngắn hạn:

- + LSTM có khả năng lưu trữ thông tin trong bộ nhớ dài hạn (long-term memory) và bộ nhớ ngắn hạn (short-term memory). Bộ nhớ dài hạn giúp mô hình ghi nhớ các thông tin quan trọng trong dữ liệu chuỗi, trong khi bộ nhớ ngắn hạn giúp mô hình tạm thời lưu trữ các thông tin ngắn hạn.

Mô hình Linear Regression:

- Giả định tuyến tính: Giả định mối quan hệ tuyến tính giữa các biến độc lập và giá cổ phiếu.[3]

- Đơn giản và dễ hiểu: Dễ dàng diễn giải và hiểu cách từng yếu tố ảnh hưởng đến giá cổ phiếu.

- Hiệu quả tính toán: Yêu cầu ít tài nguyên và có thể huấn luyện nhanh chóng.

- Nhạy cảm với outliers: Dễ bị ảnh hưởng bởi các giá trị ngoại lai, làm giảm độ chính xác.

- Giả định về phân phối sai số: Sai số giữa giá trị dự đoán và giá trị thực tế tuân theo phân phối chuẩn và có phương sai không đổi.

- Đa cộng tuyến: Bị ảnh hưởng khi các biến độc lập có mối quan hệ tuyến tính cao với nhau.

- Hạn chế với quan hệ phi tuyến: Không thể nắm bắt các mối quan hệ phi tuyến giữa các biến độc lập và giá cổ phiếu.

- Khả năng diễn giải mô hình: Hệ số hồi quy cho biết mức độ ảnh hưởng của từng biến độc lập đến giá cổ phiếu.

- Dễ dàng đánh giá mô hình: Sử dụng các chỉ số như R-squared, MSE, RMSE, và MAE.

- Tiện lợi trong kiểm tra giả định và chẩn đoán lỗi: Dễ dàng kiểm tra các giả định của mô hình và chẩn đoán lỗi.

1.4 Thực nghiệm

1.4.1 Dữ liệu

Sử dụng thư viện vnstock để lấy dữ liệu lịch sử về giá cổ phiếu, khối lượng giao dịch và các chỉ số khác.

Dữ liệu đã lấy về bao gồm các thông tin sau:

- time: Ngày giao dịch.
- open: Giá mở cửa của cổ phiếu trong ngày. Đây là giá mà cổ phiếu được giao dịch lần đầu tiên trong một ngày giao dịch.
- high: Giá cao nhất của cổ phiếu trong ngày.
- low: Giá thấp nhất của cổ phiếu trong ngày.
- close: Giá đóng cửa của cổ phiếu trong ngày. Đây là giá cuối cùng mà cổ phiếu được giao dịch trong một ngày giao dịch.
- volume: Khối lượng giao dịch của cổ phiếu trong ngày. Đây là số lượng cổ phiếu đã được giao dịch trong phiên giao dịch đó.
- ticker: mã ngân hàng của ngân hàng đó gồm 3 ký tự

1.4.2 Xử lý dữ liệu

Cần phải tiền xử lý dữ liệu, vì:

- Có thể dữ liệu bị thiếu.
- Tùy vào phương pháp học máy có thể cần phải chuẩn hóa dữ liệu

Xử lý dữ liệu như thế nào?

- Kiểm tra dữ liệu thiếu: Đầu tiên, ta cần kiểm tra xem có dữ liệu thiếu không trong các cột. Nếu có, có thể quyết định xóa các dòng hoặc điền giá trị thiếu bằng cách sử dụng phương pháp như điền giá trị trung bình, trung vị hoặc phương pháp khác.

- Tính toán các đặc trưng mới: Tính toán các đặc trưng mới từ dữ liệu hiện có như các chỉ số kỹ thuật, biến đổi giá trị (ví dụ: tỷ lệ biến đổi hàng ngày), hoặc các đặc trưng thời gian (ví dụ: tuần này so với tuần trước).

- Tạo tập dữ liệu huấn luyện và kiểm tra: Sau khi đã xử lý dữ liệu, cần chia dữ liệu thành tập dữ liệu huấn luyện và tập dữ liệu kiểm tra. Tập dữ liệu huấn luyện sẽ được sử dụng để huấn luyện mô hình, trong khi tập dữ liệu kiểm tra sẽ được sử dụng để đánh giá hiệu suất của mô hình.

- Khám phá dữ liệu (EDA): Trước khi bắt đầu xây dựng mô hình, thực hiện EDA để hiểu rõ hơn về dữ liệu, như phân phối của các biến, sự tương quan giữa các biến, và các xu hướng trong dữ liệu.

1.4.3 Công nghệ sử dụng

- Ngôn ngữ lập trình: Python.

- Thư viện cung cấp dữ liệu chứng khoán được sử dụng là vnstock (pip install vnstock).

- Các thư viện được sử dụng là các thư viện hỗ trợ trong ngôn ngữ lập trình Python dùng để xử lý bảng dữ liệu như pandas, các thư viện số học như numpy, các thư viện trực quan hóa dữ liệu như matplotlib và plotly, thư viện hỗ trợ cho quá trình huấn luyện học máy như Scikit-learn...

- Công cụ: Jupyter Notebook, Colab, Visual Studio Code.

1.4.4 Cách đánh giá

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

Trong đó:

- n là số lượng điểm dữ liệu.
- y_i là giá trị thực tế của điểm dữ liệu thứ i .
- \hat{y}_i là giá trị được dự đoán tương ứng với điểm dữ liệu thứ i

$$r^2 = \frac{RSS}{TSS} = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2}$$

Trong đó:

- y là giá trị thực tế của biến phụ thuộc.
- \hat{y} là giá trị dự đoán của biến phụ thuộc dựa trên các giá trị độc lập được sử dụng trong mô hình.
- \bar{y} là giá trị trung bình của biến phụ thuộc.

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

Trong đó:

- n là Số lượng mẫu trong tập dữ liệu
- y_i là giá trị dự đoán thứ i
- x_i Là giá trị thực tế tương ứng thứ i

1.5 Kết quả đạt được

- Linear Regression:
 - + test_size = 0.2
 - + random_state = 42

+ coefficients = [-5.77584099e-01 8.01844096e-01 7.72491946e-01
1.71738720e-05]

+ intercept = 6.828597615506624

LSTM:

+ epochs = 10

+ batch = 50

+ dropout rate = 0.5

– Các kết quả đạt được theo các độ đo:

Model	Mean Squared Error	R^2 – Squared	Mean Absolute Error
Linear Regression	17154.434379	0.999420	83.171267
LSTM	242172.330192	0.991836	287.273797

– Biểu đồ so sánh giữa các phương pháp theo các độ đo: *Hình 1.5*

– Giải thích về kết quả đạt được và sự khác biệt giữa các phương pháp:

Mean Squared Error (MSE):

- Linear Regression có MSE là 17154.434379.
- LSTM có MSE là 248819.631683.

Giải thích: MSE đo lường trung bình của bình phương của sai số giữa giá trị thực tế và giá trị dự đoán. MSE càng thấp thì mô hình dự đoán càng chính xác. Trong trường hợp này, Linear Regression có MSE thấp hơn so với LSTM, cho thấy mô hình Linear Regression dự đoán chính xác hơn so với mô hình LSTM.

R^2 – Score:

- Linear Regression: R^2 – squared là 0.999420.
- LSTM: R^2 – Score là 0.991612.

Giải thích: R^2 Score là một phép đo cho thấy phần trăm phương sai của biến phụ thuộc mà mô hình giải thích được. Giá trị càng gần 1, mô hình càng tốt. Linear Regression có R^2 Score gần 1 hơn so với LSTM, cho thấy Linear Regression giải thích được phần lớn sự biến thiên trong dữ liệu hơn so với LSTM.

Mean Absolute Error (MAE):

- Linear Regression: MAE là 83.171267.
- LSTM: MAE là 286.930596.

Giải thích: MAE là trung bình của giá trị tuyệt đối của sai số giữa giá trị thực tế và giá trị dự đoán. Giá trị MAE càng thấp, mô hình càng chính xác. Trong trường hợp này, Linear Regression có MAE thấp hơn so với LSTM, cho thấy mô hình Linear Regression có sự sai số thấp hơn trong dự đoán so với mô hình LSTM.

Nguyên nhân của sự khác biệt giữa các phương pháp:

→ Linear Regression và LSTM là hai phương pháp khác nhau trong việc mô hình hóa dữ liệu. Linear Regression dựa vào các biến độc lập để tạo ra một mô hình tuyến tính, trong khi LSTM sử dụng kiến trúc mạng nơ-ron tái phát (recurrent neural network) để học các mẫu chuỗi thời gian trong dữ liệu. Sự phức tạp của mô hình LSTM có thể làm giảm hiệu suất dự đoán so với Linear Regression trong một số trường hợp, nhất là khi dữ liệu không có các mẫu chuỗi thời gian rõ ràng hoặc khi dữ liệu không đủ lớn để huấn luyện một mô hình phức tạp như LSTM.

1.6 Kết luận

1.6.1 Linear Regression

a. Tóm tắt kết quả đạt được

- Độ chính xác trung bình: Thị trường chứng khoán thường phức tạp và chứa nhiều yếu tố phi tuyến tính, do đó độ chính xác có thể không cao như mong đợi.
- Khả năng diễn giải: Các hệ số của mô hình cho biết mức độ ảnh hưởng của từng biến độc lập đến giá cổ phiếu, giúp các nhà phân tích tài chính hiểu rõ hơn về các yếu tố tác động đến thị trường.

- Hiệu quả tính toán: có hiệu quả tính toán cao, cho phép xử lý và huấn luyện mô hình nhanh chóng ngay cả với lượng dữ liệu lớn, hữu ích khi cần cập nhật mô hình thường xuyên với dữ liệu mới.
- Khả năng phát hiện xu hướng: phát hiện và dự đoán các xu hướng tuyến tính ngắn hạn trong dữ liệu chứng khoán. Với các xu hướng phi tuyến hoặc biến động mạnh, mô hình có thể không dự đoán chính xác.
- Độ nhạy với outliers và đa cộng tuyến: Linear Regression nhạy cảm với các giá trị ngoại lai (outliers) và đa cộng tuyến (multicollinearity) giữa các biến độc lập, có thể làm giảm độ chính xác của mô hình.
- Kết quả đánh giá mô hình: Thông qua các chỉ số đánh giá như Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), và R-squared, có thể thấy được hiệu suất của mô hình:
 - MSE/RMSE: Cho biết sai số trung bình giữa giá dự đoán và giá thực tế. Giá trị càng thấp càng tốt.
 - MAE: Đo lường sai số trung bình theo giá trị tuyệt đối, giúp dễ dàng so sánh.
 - R-squared: Cho biết tỷ lệ phương sai của biến phụ thuộc được giải thích bởi các biến độc lập. Giá trị gần 1 cho thấy mô hình giải thích tốt dữ liệu.
- Ứng dụng thực tiễn: Linear Regression có thể được sử dụng làm mô hình cơ bản hoặc bước đầu trong việc phân tích dữ liệu và dự đoán thị trường chứng khoán. Nó cung cấp một nền tảng để so sánh và cải thiện với các mô hình phức tạp hơn như các mô hình phi tuyến hoặc học sâu.

b. Hạn chế:

- Linear Regression không thể nắm bắt được các quan hệ phi tuyến tính và các biến động mạnh của thị trường chứng khoán, dẫn đến độ chính xác dự đoán không cao.

- Nhạy cảm với outliers và đa cộng tuyến: Mô hình nhạy cảm với các giá trị ngoại lai và đa cộng tuyến giữa các biến độc lập, có thể làm giảm độ chính xác của dự đoán.

- Giả định tuyến tính: Linear Regression giả định mối quan hệ giữa các biến là tuyến tính, điều này không luôn đúng với thị trường chứng khoán vốn phức tạp và chứa nhiều yếu tố phi tuyến.

- Không nắm bắt được sự phức tạp của dữ liệu: Mô hình không đủ mạnh để xử lý các tương tác phức tạp và các yếu tố ngoại lai ảnh hưởng đến thị trường chứng khoán.

c. Hướng phát triển trong tương lai

- Sử dụng các mô hình phi tuyến: Triển khai các mô hình phi tuyến như Polynomial Regression, Decision Trees, Random Forests, và Gradient Boosting để nắm bắt được các quan hệ phức tạp hơn trong dữ liệu.

- Áp dụng học sâu (Deep Learning): Sử dụng các mô hình học sâu như LSTM (Long Short-Term Memory) hoặc RNN (Recurrent Neural Networks) để dự đoán chuỗi thời gian và nắm bắt các xu hướng phức tạp trong dữ liệu thị trường chứng khoán.

- Kết hợp các phương pháp Ensemble: Kết hợp Linear Regression với các mô hình khác trong các phương pháp Ensemble (như Bagging, Boosting) để cải thiện độ chính xác và tính ổn định của dự đoán.

- Cải thiện tính năng (Feature Engineering): Tạo và lựa chọn các đặc trưng mới từ dữ liệu gốc, bao gồm các chỉ số kỹ thuật, các biến trễ, và các biến tương tác để cải thiện hiệu suất mô hình.

- Xử lý outliers và đa cộng tuyến: Sử dụng các kỹ thuật như Robust Regression, Ridge Regression, hoặc Lasso Regression để xử lý outliers và đa cộng tuyến, nâng cao độ chính xác của mô hình.

- Tăng cường dữ liệu: Kết hợp dữ liệu từ nhiều nguồn khác nhau (như dữ liệu kinh tế, dữ liệu xã hội, dữ liệu tin tức) để cung cấp một cái nhìn toàn diện và phong phú hơn về thị trường, từ đó cải thiện độ chính xác của dự đoán.

1.6.2 LSTM

a. Tóm tắt kết quả đạt được:

- Mô hình LSTM đã được huấn luyện trên dữ liệu thị trường chứng khoán với 10 epochs.
- Mất mát (loss) của mô hình đã giảm từ 0.0526 ở epoch đầu tiên xuống còn 0.0238 ở epoch cuối cùng.
- Dự đoán giá cổ phiếu trên tập huấn luyện có vẻ khá gần với giá thực tế, khi mức độ sai lệch không quá lớn.
- Kết quả đánh giá mô hình: Thông qua các chỉ số đánh giá như Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), và R-squared, có thể thấy được hiệu suất của mô hình:
 - MSE/RMSE: Cho biết sai số trung bình giữa giá dự đoán và giá thực tế. Giá trị càng thấp càng tốt.
 - MAE: Đo lường sai số trung bình theo giá trị tuyệt đối, giúp dễ dàng so sánh.
 - R-squared: Cho biết tỷ lệ phương sai của biến phụ thuộc được giải thích bởi các biến độc lập. Giá trị gần 1 cho thấy mô hình giải thích tốt dữ liệu.

b. Hạn chế:

- Mặc dù mô hình đã đạt được một số kết quả khá tích cực, nhưng có thể còn sự sai lệch đáng kể giữa dự đoán và giá thực tế trên tập kiểm tra.
- Việc giảm mất mát trên tập huấn luyện không nhất thiết có nghĩa là mô hình đã học được các mẫu tổng quát, và có thể tồn tại hiện tượng overfitting.

c. Hướng phát triển trong tương lai:

- Nâng cao chất lượng dự đoán bằng cách sử dụng các kỹ thuật tinh chỉnh mô hình như thay đổi kiến trúc mạng (ví dụ: thêm lớp LSTM, điều chỉnh số lượng neuron), tăng cường dữ liệu huấn luyện, và tối ưu hóa siêu tham số.
- Điều chỉnh mô hình để giảm overfitting, có thể sử dụng kỹ thuật dropout hoặc regularization.
- Kết hợp mô hình LSTM với các phương pháp khác như mạng nơ-ron tích chập (CNN) để kết hợp thông tin không gian và thời gian từ dữ liệu thị trường chứng khoán.
- Nghiên cứu và áp dụng các kỹ thuật mới trong lĩnh vực học sâu (deep learning) để cải thiện hiệu suất dự đoán, chẳng hạn như mạng nơ-ron tái phát (autoencoder), mạng nơ-ron biến đổi (transformer), hoặc mô hình sinh (generative model).

LÀM VIỆC NHÓM

- Nhóm chúng em được hình thành với mục tiêu chung là cùng nhau trao đổi, chia sẻ, thực hiện và hoàn thành bài đồ án cuối kỳ môn Máy học một cách chính chu và tốt nhất.

- Sau khi nhận được đề tài, chúng em thống nhất sẽ sử dụng 2 mô hình Máy học Random Forest và Linear Regression. Mỗi thành viên sẽ chọn mô hình phù hợp với kiến thức đã học để nghiên cứu và triển khai mô hình để có thể dễ dàng so sánh các kết quả của mỗi mô hình với dữ liệu chứng khoán được dự đoán. Từ đó so sánh và rút ra kết luận tổng quan mô hình nào tốt hơn, mô hình nào có độ chính xác cao hơn...

- Trong quá trình làm việc với nhau chúng em hỗ trợ lẫn nhau về kiến thức kỹ thuật mà thành viên còn lại bị thiếu sót để đưa ra kết quả chính chu và hoàn thiện nhất đối với đồ án.

Tổng số lần gặp nhau (tính theo buổi): 10 buổi

Tổng thời gian gặp nhau (tính theo giờ): hơn 30 giờ

- Công việc được phân chia bình đẳng, cụ thể và phù hợp với năng lực mỗi người cụ thể như sau:

+ Buổi 1: nhận đề tài và bàn luận về đề tài được giao, tìm hiểu các phương pháp máy học đã được thực hiện với đề tài này trên Internet để tham khảo phương pháp.

+ Buổi 2: Chúng em họp, đưa ra phương pháp có thể áp dụng và đưa vào đề tài, thống nhất lấy dữ liệu từ đâu và sử dụng nó như thế nào.

+ Buổi 3, 4, 5, 6, 7: Triển khai và xem cách hoạt động để cải tiến mô hình của mỗi thành viên, đưa ra các độ đo của thành viên tham khảo và thực hiện với mô hình của mình để tính toán.

+ Buổi 8: Review mã (code) của mỗi thành viên và tinh chỉnh.

+ Buổi 9: Gộp mã (code) của mỗi thành viên vào thành 1 file notebook (ipynb) hoàn chỉnh bằng Colab và cùng review mã (code).

+ Buổi 10: Hoàn thành bảng báo cáo và review 1 lần nữa file notebook đã gộp và nộp bài lên lms.

- Phương thức họp, gặp mặt: online (trực tuyến) qua phần mềm Zavi (Zalo) và Discord.

TÀI LIỆU THAM KHẢO

1. Grigoryan, H. (2016). A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA). *Database Systems Journal*, 7(1).
2. Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*.
3. Altay, E., & Satman, M. H. (2005). Stock market forecasting: artificial neural network and linear regression comparison in an emerging market. *Journal of Financial Management & Analysis*, 18(2), 18.
4. Roy, S. S., Mittal, D., Basu, A., & Abraham, A. (2015). Stock market forecasting using LASSO linear regression model. In *Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014* (pp. 371-381). Springer International Publishing.
5. Qi, M., & Maddala, G. S. (1999). Economic factors and the stock market: a new perspective. *Journal of forecasting*, 18(3), 151-166.
6. Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Systems with Applications*, 113, 457-480.
7. Sunny, M. A. I., Maswood, M. M. S., & Alharbi, A. G. (2020, October). Deep learning-based stock price prediction using LSTM and bi-directional LSTM model. In *2020 2nd novel intelligent and leading emerging sciences conference (NILES)* (pp. 87-92). IEEE.

TỰ ĐÁNH GIÁ

(Bài nhóm)

STT	Nội dung	Điểm chuẩn	Tự chấm	Ghi chú
1 (8.5đ)	1.1 Giới thiệu về bài toán	0.5	0.5	
	1.2 Phân tích yêu cầu của bài toán	1	1	
	1.3 Phương pháp giải quyết bài toán	1.5	1.5	
	1.4 Thực nghiệm	4	3.75	
	1.5 Kết quả đạt được	1	0.5	
	1.6 Kết luận	0.5	0.5	
2 (1đ)	Báo cáo (chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng)	1đ	1	
3 (0.5đ)	Điểm nhóm (chú ý trả lời các câu hỏi trong mục làm việc nhóm)	0.5đ	0.5	
Tổng điểm			9.25	

TỰ ĐÁNH GIÁ (Nguyễn Khắc Luật)

STT	Nội dung	Điểm chuẩn	Tự chấm	Ghi chú
1 (9đ)	1.1 Giới thiệu về bài toán	0.5	0.5	
	1.2 Phân tích yêu cầu của bài toán	1	1	
	1.3 Phương pháp giải quyết bài toán	1.5	1.5	
	1.4 Thực nghiệm	4.5	4.25	
	1.5 Kết quả đạt được	1	0.5	
	1.6 Kết luận	0.5	0.5	
2 (1đ)	Báo cáo (chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng)	1đ	1đ	
Tổng điểm			9.25	

TỰ ĐÁNH GIÁ (Phạm Nhựt Minh)

STT	Nội dung	Điểm chuẩn	Tự chấm	Ghi chú
1 (9đ)	1.1 Giới thiệu về bài toán	0.5	0.5	
	1.2 Phân tích yêu cầu của bài toán	1	1	
	1.3 Phương pháp giải quyết bài toán	1.5	1.5	
	1.4 Thực nghiệm	4.5	4.25	
	1.5 Kết quả đạt được	1	0.5	
	1.6 Kết luận	0.5	0.5	
2 (1đ)	Báo cáo (chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng)	1đ	1đ	
Tổng điểm			9.25	