

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO THỰC TẬP TỐT NGHIỆP
XÂY DỰNG MÔ HÌNH DỰ ĐOÁN TÀI CHÍNH BẰNG
LSTM

GIẢNG VIÊN HƯỚNG DẪN: TS. VÕ THỊ HỒNG THẨM

SINH VIÊN THỰC HIỆN : BÙI TIẾN SANG

MSSV : 2000004684

NIÊN KHÓA : 2020

ĐƠN VỊ THỰC TẬP : VIỆN CƠ HỌC VÀ TIN HỌC ỨNG DỤNG

CHUYÊN NGÀNH : KHOA HỌC DỮ LIỆU

Tp.HCM, tháng 01 năm 2024

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH
KHOA CÔNG NGHỆ THÔNG TIN**



BÁO CÁO THỰC TẬP TỐT NGHIỆP

**XÂY DỰNG MÔ HÌNH DỰ ĐOÁN TÀI CHÍNH BẰNG
LSTM**

GIẢNG VIÊN HƯỚNG DẪN: TS. VÕ THỊ HỒNG THẨM

SINH VIÊN THỰC HIỆN : BÙI TIẾN SANG

MSSV : 2000004684

NIÊN KHÓA : 2020

ĐƠN VỊ THỰC TẬP : VIỆN CƠ HỌC VÀ TIN HỌC ỨNG DỤNG

CHUYÊN NGÀNH : KHOA HỌC DỮ LIỆU

Tp.HCM, tháng 01 năm 2024

LỜI CẢM ƠN

Em xin bày tỏ lòng kính trọng và tri ân sâu sắc đến quý thầy cô giáo của khoa Công Nghệ Thông Tin, Đại học Nguyễn Tất Thành, những người đã truyền đạt những tri thức quý giá và chia sẻ kinh nghiệm thực tế trong suốt thời gian em theo học tại đây. Em cũng muốn cảm ơn cô TS. Võ Thị Hồng Thắm, một giáo viên tại trường, người đã không chỉ hướng dẫn mà còn mở ra cơ hội liên kết với các doanh nghiệp, giúp em tìm kiếm được môi trường thực tập lý tưởng. Cô đã thể hiện sự quan tâm và nhiệt huyết trong việc hỗ trợ em hoàn thiện hồ sơ thực tập và góp ý cho báo cáo của em.

Em cũng muốn gửi lời cảm ơn chân thành đến TS. Trương Nguyên Vũ, Viện trưởng Viện Cơ học và Tin học ứng dụng, cùng ban quản lý Viện đã tạo điều kiện thuận lợi nhất cho kỳ thực tập của em. Lời cảm ơn cũng được gửi đến thầy ThS. Hồ Khôi, Bộ phận Quản lý Hệ thống Thông tin, đã cung cấp sự hướng dẫn xuất sắc trong suốt quá trình em thực tập tại Viện, cùng với sự chia sẻ tài liệu bổ ích và sự động viên, hỗ trợ từ các đồng nghiệp tại Bộ phận.

Em xin gửi lời cảm ơn chân thành và sâu sắc.

BÙI TIẾN SANG

LỜI MỞ ĐẦU

Trong suốt thời gian theo học và làm việc tại viện, em đã tích lũy được một lượng lớn kiến thức chuyên ngành. Em đã học được nhiều điều về các phương pháp tiên tiến trong lĩnh vực Data Science, bao gồm Machine Learning và Deep Learning, giúp em hiểu sâu hơn về cách phân tích dữ liệu, phát triển mô hình dự đoán và xây dựng các ứng dụng.

Quá trình tự học của em tập trung vào việc áp dụng kiến thức lý thuyết vào thực tiễn. Em đã tham gia nhiều khóa học trực tuyến, đọc tài liệu liên quan và tự thực hiện các dự án nhỏ để cải thiện kỹ năng thực hành. Điều này giúp em nâng cao khả năng phân tích và phản biện.

Trong thời gian thực tập, em đã được trải nghiệm môi trường làm việc thực tế, áp dụng và mở rộng kiến thức của mình trong việc giải quyết các vấn đề cụ thể, dẫn đến việc phát triển một hệ thống học máy. Quá trình này không chỉ giúp em làm quen với công nghệ mới mà còn rèn luyện kỹ năng làm việc nhóm và hợp tác.

Tổng thể, những trải nghiệm học tập và làm việc này đã góp phần quan trọng vào sự phát triển chuyên môn và kỹ năng thực hành của em, từ việc hiểu biết về lý thuyết cho đến khả năng tham gia vào công việc thực tế một cách độc lập.

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

1. Hình thức (Bố cục, trình bày, lỗi, các mục, hình, bảng, công thức, phụ lục,)

2. Nội dung (mục tiêu, phương pháp, kết quả, sao chép, các chương, tài liệu,..)

3. Kết luận

TPHCM, Ngày tháng năm 2024

Giáo viên hướng dẫn

(Ký tên, đóng dấu)

MỤC LỤC

CHƯƠNG I : GIỚI THIỆU NƠI THỰC TẬP VÀ ĐỀ TÀI	1
1. Khái quát về Viện Cơ học và Tin học ứng dụng.....	1
1.1 Thông tin chung ^[1]	1
1.2 Chức năng và nhiệm vụ ^[1]	1
1.3 Quá trình hình thành và phát triển ^[1]	2
1.4 Cơ cấu tổ chức ^[1]	2
1.4.1 Lực lượng cán bộ	2
1.4.2 Các phòng chuyên môn	2
1.4.3 Các đơn vị quản lý nghiệp vụ	3
1.5 Các hoạt động thường xuyên của đơn vị ^[1]	3
1.6 Các thành tích nổi bật ^[1]	3
1.6.1 Lĩnh vực nghiên cứu cơ bản	3
1.7 Các lĩnh vực khoa học công nghệ ^[1]	4
1.8 Nghiên cứu ứng dụng triển khai công nghệ ^[1]	4
1.9 Thông tin về quá trình thực tập	5
1.9.1 Vị trí thực tập.....	5
1.9.2 Các yêu cầu thực tập.....	5
1.9.3 Yêu cầu từ đề tài	5
1.10 Giới thiệu đề tài thực tập.....	5
1.10.1 Phương pháp đề tài	6
1.10.2 Lý do chọn đề tài	7
1.10.3 Mục tiêu đề tài	8

1.11 Cấu trúc của báo cáo	10
CHƯƠNG II. CƠ SỞ LÝ LUẬN VỀ VẤN ĐỀ NGHIÊN CỨU	11
2.1 Định nghĩa và tầm quan trọng của dự báo trong tài chính	11
2.2 Các nguyên tắc cơ bản của dự báo tài chính và kinh doanh	11
2.3 Lý thuyết nền tảng về học máy và dự báo tài chính	11
2.3.1 Supervised Learning	12
2.3.2 Unsupervised Learning	12
2.3.3 Ứng dụng	12
2.4 Khám phá các mô hình dự báo truyền thống và hiện đại.....	13
2.5 Khung lý thuyết về chuỗi thời gian và phân tích dữ liệu	13
2.6 Tiếp cận nghiên cứu: phương pháp luận và kỹ thuật	14
2.7 Tổng kết	14
CHƯƠNG III. MÔ HÌNH LÝ THUYẾT	16
3.1 Lý thuyết về dự báo trong tài chính và kinh doanh.	16
3.1.1 Giới thiệu về dự báo trong tài chính và kinh doanh.	16
3.1.2 Khái niệm về dự báo.....	16
3.1.3 Các loại dự báo.	17
3.2 Dữ liệu chuỗi thời gian.....	18
3.3 Lý thuyết về Time Series Decomposition.....	19
3.3.1 Shifting and lags	19
3.4 Technical Indicators.....	20
3.4.1 Simple Moving Average.....	20
3.4.2 Exponential Moving Average.....	21
3.4.3 Relative Strength Index	22

3.4.4 Moving Average Convergence Divergence	23
3.5 Differencing	23
3.6 Mô hình LSTM.	23
2.5.1 Sơ đồ của mô hình LSTM.....	25
3.7 Đánh giá mô hình.	26
3.7.1 Thu thập dữ liệu	26
3.7.2 Xử lý dữ liệu	26
3.7.4 Dự đoán và đánh giá	27
CHƯƠNG IV : THỰC NGHIỆM	28
4.1 Mô tả tập dữ liệu	28
4.1.1 Thông tin cung cấp	28
4.1.2 Tại sao giá đóng cửa quan trọng.....	29
4.2 Trực quan hóa dữ liệu.	29
4.2.1 Giá đóng của cổ phiếu	30
4.2.2 Phân phối giá đóng cửa.....	31
4.2.3 Closing Price: Range Slider and Selectors	32
4.2.4 Lợi nhuận, thanh trượt và bộ chọn phạm vi.....	33
4.2.5 Biểu đồ nến	33
4.2.6 Shifting and lags	36
4.3 Technical Indicators	37
4.3.1 Simple Moving Average.....	37
4.3.2 Exponential Moving Average (EMA)	39
4.3.3 Relative Strength Index (RSI)	40
4.3.4 MACD	41

4.4 Time Series Data Analysis – Resampling.....	42
4.4.1 Seasonal Factors	43
4.4.1.1 Tần suất hằng ngày	43
4.4.1.2 Tần suất 365 ngày	44
4.4.1.3 Tần suất 366 ngày	45
4.4.2 Stationary Test / ADF Test	46
4.5 Differencing	46
4.5.1 First Difference	46
4.5.2 Seasonal Difference	47
4.5.3 Seasonal Daily Difference	48
4.6 Xây dựng mô hình LSTM.....	49
4.6.1 Xây dựng mô hình LSTM với 10 bước thời gian	51
4.6.1.1 Dự đoán 10 ngày tiếp theo.....	53
4.6.2 Xây dựng mô hình LSTM với 30 bước thời gian	54
4.6.2.1 Dự đoán 30 ngày tiếp theo.....	57
4.7 So sánh 2 bước chạy.....	59
CHƯƠNG V : KẾT LUẬN VÀ KIẾN NGHỊ	60
5.1 Kết luận	60
5.2 Các mục tiêu đưa ra đã hoàn thành	60
5.3 Kiến nghị.....	61
DANH MỤC TÀI LIỆU THAM KHẢO	62

DANH MỤC CÁC BẢNG BIỂU

Bảng 1: Bảng dữ liệu cổ phiếu APPLE	28
Bảng 2 : First Difference	47
Bảng 3 : Seasonal Difference	47
Bảng 4 : Seasonal Daily Difference	48
Bảng 5 : Thông số đánh giá mô hình qua 10 bước thời gian	52
Bảng 6 : Giá đóng cửa dự đoán 10 ngày tiếp theo	53
Bảng 7 : Thông số đánh giá mô hình qua 10 bước thời gian	56
Bảng 8 : Giá đóng cửa dự đoán 30 ngày tiếp theo	57
Bảng 9 : So sánh 10 steps và 30 steps	59

DANH MỤC CÁC BẢNG HÌNH

Hình 1 : Viện Cơ Học Và Tin Học Ứng Dụng	1
Hình 2 : Sơ đồ LSTM	26
Hình 3 : Hình ảnh biểu diễn giá đóng cửa giai đoạn 2018-2023	30
Hình 4 : Hình ảnh phân phối của giá đóng cửa của cổ phiếu giai đoạn 2018-2023	31
Hình 5 : Closing Price: Range Slider and Selectors	32
Hình 6 : Lợi nhuận, thanh trượt và bộ chọn phạm vi giai đoạn 2018-2023	33
Hình 7 : Mô tả sơ bộ về biểu đồ nến	34
Hình 8 : Biểu đồ nến thể hiện giá đóng cửa của cổ phiếu giai đoạn 2018-2023	35
Hình 9 : Giá đóng cửa hằng ngày và được dịch chuyển trong giai đoạn 2018-2023	36
Hình 10: Biểu đồ biểu diễn so sánh giữa giá trị đóng cửa với SMA5 và SMA15	37
Hình 11 : So sánh giữa giá trị đóng cửa SMA14 thông qua UPPER và LOWER14	38
Hình 12 : Biểu đồ so sánh giữa Close với Exponential Moving Average (EMA)5,15	39
Hình 13 : Biểu đồ biểu diễn giá trị Relative Strength Index (RSI) giai đoạn 2018-2023	40
Hình 14 : MACD và MACD Signal	41
Hình 15 : Time Series Data Analysis – Resampling giai đoạn 2018 - 2023	42
Hình 16: Biểu đồ phân tích mùa vụ (hằng ngày)	43
Hình 17 :Biểu đồ phân tích mùa vụ (365 ngày)	44
Hình 18 :Biểu đồ phân tích mùa vụ (366 ngày)	45
Hình 19 : Mô tả trung bình lăn với độ lệch chuẩn của dữ liệu giai đoạn 2018-2023	46
Hình 20 : Biểu đồ Differencing	49
Hình 21 : Hình ảnh dự đoán của LSTM với 10 bước thời gian	51
Hình 22 : Dự đoán giá đóng cửa trong vòng 10 ngày tới	54
Hình 23 : Dự đoán LSTM với 30 bước thời gian	55
Hình 24 : Dự đoán giá đóng cửa trong vòng 30 ngày tới	58

DANH MỤC CÁC TỪ VIẾT TẮT

CHỮ VIẾT TẮT	Ý NGHĨA
LSTM	Long Short-Term Memory
RSI	Relative Strength Index
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MSE	Mean Squared Error
MACD	Moving Average Convergence Divergence
DA	Data Analysis
RSS	Residual Sum of Squares

CHƯƠNG I : GIỚI THIỆU NƠI THỰC TẬP VÀ ĐỀ TÀI

1. Khái quát về Viện Cơ học và Tin học ứng dụng

1.1 Thông tin chung^[1]



Hình 1 : Viện Cơ Học Và Tin Học Ứng Dụng

Viện Cơ học và Tin học ứng dụng là một đơn vị nghiên cứu khoa học hàng đầu của Việt Nam, được thành lập từ năm 1984 bởi Chính phủ. Viện có nhiệm vụ nghiên cứu cơ bản và ứng dụng các lĩnh vực liên quan đến cơ học và tin học, đồng thời cung cấp các tư vấn khoa học cho các cơ quan nhà nước trong việc xây dựng chính sách, chiến lược và quy hoạch phát triển kinh tế, xã hội. Viện cũng đào tạo các nhà khoa học, kỹ sư và chuyên gia có trình độ cao trong các ngành cơ học và tin học. Viện được trang bị nhiều thiết bị hiện đại và tiên tiến, cho phép thực hiện các nghiên cứu sâu và rộng. Viện hiện có trụ sở chính tại số 291 Điện Biên Phủ, Phường 7, Quận 3, TP. Hồ Chí Minh, và một số chi nhánh khác tại các tỉnh thành trong cả nước.

1.2 Chức năng và nhiệm vụ^[1]

Nghiên cứu cơ bản trong khoa học và công nghệ đóng vai trò thiết yếu trong việc mở đường cho các phát kiến mới, bao gồm cả trong cơ học biển, dầu khí, và tự động hóa, cũng như trong lĩnh vực công nghệ thông tin. Các công trình này không chỉ tập trung vào chế tạo máy và vật liệu xây dựng mà còn hướng đến việc quản lý môi trường và phát triển công nghệ môi trường. Việc xác định các chỉ tiêu cơ - lý của vật liệu và điều kiện địa cơ học góp

phần vào thiết kế hiệu quả các công trình xây dựng. Ngoài ra, ứng dụng của những nghiên cứu này trong ngành công nghiệp dược, y tế, thực phẩm, và hàng tiêu dùng cũng rất rõ rệt. Hợp tác quốc tế và đào tạo chuyên sâu trong lĩnh vực này cũng là một phần quan trọng, góp phần thúc đẩy sự phát triển của khoa học và công nghệ hiện đại.

Những định hướng nghiên cứu - triển khai của Viện:

Viện tập trung vào các định hướng nghiên cứu và triển khai chủ chốt bao gồm: Cơ khí và Tự động hóa, Cơ học Tính toán cùng Cơ học Công trình và Chẩn đoán kỹ thuật, Cơ học Thủy và khí – tập trung vào Cơ học sông và biển cũng như Môi trường, và cuối cùng là lĩnh vực Công nghệ thông tin.

1.3 Quá trình hình thành và phát triển^[1]

Viện Cơ học và Tin học ứng dụng là một tổ chức quốc gia đa ngành, được thành lập bởi Chính phủ từ năm 1984. Qua nhiều năm, Viện đã không ngừng phát triển, đầu tư vào các thiết bị chuyên ngành hiện đại, nổi bật trong lĩnh vực của mình. Là một phần của Chính phủ, Viện chủ yếu tập trung vào nghiên cứu cơ bản trong khoa học tự nhiên và phát triển công nghệ.

Viện đảm nhận vai trò trọng yếu trong việc cung cấp cơ sở lý luận khoa học cho các quyết định quản lý về khoa học và công nghệ, góp phần vào việc hoạch định chính sách, chiến lược và quy hoạch phát triển kinh tế-xã hội. Bên cạnh đó, Viện còn có nhiệm vụ đào tạo nhân sự có chất lượng cao, đáp ứng các tiêu chuẩn chuyên môn theo quy định pháp luật.

1.4 Cơ cấu tổ chức^[1]

1.4.1 Lực lượng cán bộ

Tổng số cán bộ, viên chức (CBVC) tại tổ chức là 74 người, trong đó bao gồm 66 người ở vị trí biên chế và 08 người làm việc dưới hợp đồng. Về trình độ chuyên môn, có 04 Phó giáo sư, 07 Tiến sĩ, 17 Thạc sĩ, và 40 Kỹ sư/Cử nhân. Ngoài ra, có thêm 10 người thuộc nhóm "Khác".

1.4.2 Các phòng chuyên môn

Tại tổ chức này, có nhiều bộ phận chuyên môn, bao gồm: Bộ phận chuyên về kỹ thuật máy móc, bộ phận nghiên cứu địa cơ học, công trình dưới lòng đất và phân tích kỹ

thuật; bộ phận tập trung vào cơ học liên quan đến biển và vấn đề môi trường; bộ phận phân tích cơ học dựa trên tính toán và ứng dụng trong xây dựng; bộ phận phát triển tự động hóa và robot; bộ phận quản lý hệ thống thông tin; bộ phận quản lý và bảo trì mạng máy tính; bộ phận chuyên về công nghệ tính toán và khám phá tri thức; và cuối cùng là bộ phận nghiên cứu và triển khai hệ thống thông tin, chịu trách nhiệm phát triển các giải pháp thông tin hiện đại. Mỗi bộ phận này đóng góp một phần quan trọng trong việc thúc đẩy sự tiến bộ trong lĩnh vực chuyên môn của mình.

1.4.3 Các đơn vị quản lý nghiệp vụ

Bộ phận Quản lý hành chính tổng quát, kèm theo các Trung tâm chuyên ngành như Trung tâm Công nghệ, Trung tâm Phát triển và Ứng dụng Công nghệ Môi trường, Trung tâm Tư vấn và Ứng dụng Nghiên cứu Khoa học Môi trường, và Trung tâm Giám sát cùng Ứng dụng Công nghệ Môi trường. Ngoài ra, còn có nhiều xưởng chuyên biệt như Xưởng Cơ nhiệt và Thiết bị Môi trường, Xưởng Chế tạo Cơ khí Chính xác, Xưởng Sản xuất Cơ khí, Xưởng Vật liệu và Máy Công nghệ, Xưởng Sản xuất Thiết bị Vật liệu Xây dựng, Xưởng Sản xuất Thiết bị Tự động Công nghiệp, và Xưởng Sản xuất Phụ tùng cùng Công nghệ Polymer.

1.5 Các hoạt động thường xuyên của đơn vị^[1]

Tổ chức này thực hiện nhiều hoạt động chuyên môn, bao gồm thực hiện các nghiên cứu khoa học và dự án trên nhiều cấp độ. Họ tập trung vào việc áp dụng kết quả nghiên cứu vào thực tiễn, nghiên cứu ứng dụng và phát triển công nghệ. Trong lĩnh vực dịch vụ khoa học kỹ thuật, họ cung cấp các giải pháp thiết kế, chế tạo thiết bị cho các ngành công nghiệp, tư vấn và thi công các công trình môi trường; cung cấp dịch vụ thông tin, tư vấn, đo lường, kiểm định, hiệu chỉnh và chuyển giao công nghệ, cũng như đào tạo và nâng cao kỹ năng chuyên môn. Ngoài ra, họ còn tham gia vào công tác đào tạo sau đại học và thực hiện các nhiệm vụ hợp tác quốc tế.

1.6 Các thành tích nổi bật^[1]

1.6.1 Lĩnh vực nghiên cứu cơ bản

Nghiên cứu sâu rộng trên nhiều lĩnh vực, bao gồm phân tích sự ổn định và động lực học của các hệ thống phức tạp, từ hệ động lực phi tuyến và hỗn độn đến hệ thống ngẫu

nhiên. Các nghiên cứu tập trung vào ổn định của máy móc và các chi tiết cơ khí như ổ trục và vòng bi, cân bằng rotor. Họ cũng khám phá ổn định của các công trình ngầm trong đất yếu, và xây dựng các mô hình toán học và cơ học cho các vấn đề thủy động lực. Nghiên cứu tiếp tục với việc phát triển các phương pháp số để giải quyết các bài toán biên trong cơ học và kỹ thuật, sử dụng công nghệ thông tin. Công việc còn bao gồm lập trình và tính toán song song, phát triển hệ điều hành và ngôn ngữ lập trình. Cuối cùng, họ nghiên cứu các chỉ tiêu cơ lý của vật liệu và xác định tham số của các hệ thống cơ học, cũng như lý thuyết chẩn đoán kỹ thuật cho các công trình và thiết bị.

1.7 Các lĩnh vực khoa học công nghệ^[1]

Tổ chức này chủ yếu tập trung vào nghiên cứu khoa học liên quan đến tương tác giữa biển và các công trình biển, cung cấp nền tảng quan trọng cho việc xây dựng các tiêu chuẩn và quy phạm cho công trình biển. Họ cũng nghiên cứu về độ thấm trong các mỏ dầu để tối ưu hóa quy hoạch khai thác. Đồng thời, việc tính toán ổn định các đập thủy điện chịu tải trọng phức tạp là một phần quan trọng trong công việc của họ. Ngoài ra, họ khám phá các vấn đề như lún sụp mặt đất và sạt lở bờ sông, mái dốc. Họ cũng phát triển phần mềm để tính toán các bài toán động học sông, biển và lan truyền chất, ứng dụng trong dự báo ô nhiễm môi trường nước và không khí, thiết kế đê đập. Các phần mềm này còn ứng dụng trong lĩnh vực chẩn đoán kỹ thuật, kinh tế, y tế từ xa, và quản lý nhà nước trong các lĩnh vực khoa học-công nghệ, tài nguyên-môi trường, giao thông-vận tải.

Đặc biệt, họ còn tính toán lũ lụt cho ĐBSCL, dự báo ô nhiễm ở các hồ và cửa sông, và xây dựng cơ sở dữ liệu cho Biển Đông dựa trên ảnh vệ tinh nhỏ, hỗ trợ dự đoán lũ lụt, nước dâng, thảm thực vật, và biến đổi đường bờ sông biển.

1.8 Nghiên cứu ứng dụng triển khai công nghệ^[1]

Tổ chức này chuyên thiết kế, chế tạo và thử nghiệm các thiết bị công nghệ cao như dây chuyền sản xuất sữa, thuốc lá, giày da, và tấm lợp. Họ cũng phát triển các thiết bị tự động hóa như cân kỹ thuật số, thiết bị y tế, và các thiết bị đo tự động. Ngoài ra, họ sản xuất thiết bị chính xác cho ngành y tế, dược phẩm, và thực phẩm, cũng như các thiết bị áp lực và nghiên cứu cho ngành sản xuất.

1.9 Thông tin về quá trình thực tập

1.9.1 Vị trí thực tập

Chức vụ thực tập: Nhà phân tích dữ liệu

Khoảng thời gian thực tập: từ ngày 08/10/2023 đến 04/12/2023

Người hướng dẫn: Ths. Hồ Khôi

Phương thức làm việc: Trực tuyến

1.9.2 Các yêu cầu thực tập

Nhiệm vụ cần thực hiện bao gồm:

Phát triển kế hoạch cho toàn bộ giai đoạn thực tập

Tạo ra mô hình LSTM cho việc dự báo thị trường chứng khoán

Hoàn thành mã nguồn

Soạn thảo báo cáo thực tập sử dụng Microsoft Word

Tạo bản trình bày PowerPoint

Hoàn thiện báo cáo thực tập

1.9.3 Yêu cầu từ đề tài

Đề tài nằm trong lĩnh vực Nhà phân tích dữ liệu

Cần có tính ứng dụng thực tế

Sử dụng mô hình ML

Thu thập dữ liệu chứng khoán theo yêu cầu của nơi thực tập

Tiến hành xử lý bộ dữ liệu

Visualize dữ liệu

Phân tích các chỉ số kỹ thuật

Nghiên cứu và áp dụng mô hình LSTM

Thực hiện thí nghiệm với mô hình LSTM

1.10 Giới thiệu đề tài thực tập

Trong một thị trường tài chính đầy biến động và không dự đoán được, khả năng dự báo trở thành chìa khóa quyết định, mở ra cơ hội chiến lược để đạt đến thành công. Dự án "Xây dựng mô hình dự báo tài chính bằng LSTM" không chỉ là một nhiệm vụ nghiên cứu mà còn là một cuộc hành trình khám phá thế giới phức tạp của dự báo tài chính. Chúng ta

đang sống trong một thời đại không chắc chắn, với thị trường tài chính không ngừng biến động, phản ứng nhanh chóng đối với những yếu tố đa dạng và khó lường. Trong ngữ cảnh này, khả năng đọc hiểu và dự báo trở thành vũ khí quan trọng, giúp nhà đầu tư và doanh nghiệp duy trì sự ổn định và đưa ra quyết định linh hoạt.

Dự án này không chỉ là một dự án nghiên cứu mà còn là một hành trình sâu rộng vào thế giới phức tạp của dự báo tài chính. Tại đây, chúng ta sẽ đối mặt với những thách thức của việc hiểu biết và mô phỏng biến động thị trường, nơi dữ liệu thời gian và mô hình học máy LSTM gặp nhau, hứa hẹn mang lại cái nhìn sắc bén và chiến lược cho nhà đầu tư và doanh nghiệp.

Tâm điểm của dự án là LSTM, một dạng mô hình học máy đặc biệt, có khả năng "gặp nhớ" thông tin từ quá khứ và áp dụng hiểu biết đó vào việc dự báo tương lai. Điều này không chỉ là một công nghệ mới mẻ mà còn là một công cụ mạnh mẽ có thể giúp chúng ta vượt qua những thách thức của thị trường đầy biến động.

Mục tiêu của chúng ta là xây dựng một mô hình dự báo tài chính linh hoạt, giúp chúng ta đọc hiểu biến động thị trường và dự đoán xu hướng tương lai một cách chính xác. Điều này không chỉ là một thách thức nghiên cứu mà còn mang lại giá trị thực tế cho cộng đồng tài chính.

Hãy cùng nhau bắt đầu cuộc hành trình này, khám phá sức mạnh của dữ liệu thời gian và LSTM để tạo ra một công cụ dự báo tài chính mạnh mẽ và đột phá trong thế giới đầy thách thức của thị trường tài chính ngày nay. Chúng ta không chỉ đối mặt với thách thức, mà còn mang lại những giải pháp sáng tạo và chiến lược chiến thắng.

1.10.1 Phương pháp đề tài

Trong nghiên cứu này, thu thập dữ liệu tài chính từ các nguồn uy tín và đáng tin cậy, đảm bảo tính đa dạng và toàn diện. Quá trình tiền xử lý dữ liệu được thực hiện thông qua việc áp dụng các kỹ thuật hiện đại để loại bỏ nhiễu và thiếu sót, nhằm tăng cường độ chính xác và đồng nhất của dữ liệu. Một mô hình LSTM được xây dựng, phù hợp với đặc tính của dữ liệu thu thập, có thể bao gồm nhiều lớp LSTM và các lớp kết nối, và sử dụng các phương pháp tối ưu hóa như gradient descent để điều chỉnh các tham số.

Trong giai đoạn huấn luyện và đánh giá mô hình, chúng tôi sử dụng tập dữ liệu huấn luyện để mô hình có thể học hỏi và điều chỉnh tham số, đồng thời áp dụng các chỉ số như Mean Squared Error (MSE), Mean Absolute Error (MAE), và hệ số xác định để đánh giá hiệu suất. Dựa trên kết quả đánh giá, chúng tôi tiếp tục tối ưu hóa mô hình để cải thiện hiệu suất, tinh chỉnh các tham số như tốc độ học và số lượng epoch.

Mô hình được sử dụng để dự đoán giá tài chính trong tương lai, với việc kiểm thử tính ổn định và chính xác trên tập dữ liệu kiểm thử. Ngoài ra, đánh giá tính ứng dụng của mô hình trong thực tế trên thị trường tài chính, đề xuất chiến lược cho nhà đầu tư và doanh nghiệp dựa trên kết quả thu được. Báo cáo và trình bày kết quả thông qua biểu đồ, đồ thị giúp cung cấp cái nhìn rõ ràng và chi tiết về quá trình nghiên cứu và kết quả đạt được.

Cuối cùng, liên tục theo dõi và nghiên cứu các phương pháp mới để cải thiện hiệu suất của mô hình, tích hợp công nghệ mới và tiên tiến. Đồng thời, chúng tôi cũng thực hiện phản biện và đề xuất hướng nghiên cứu tiếp theo, nhìn nhận những hạn chế và tiềm năng phát triển của đề tài. Tổng kết những kiến thức mới, phát hiện quan trọng và bài học học được, và nếu thích hợp, chúng tôi sẽ triển khai mô hình vào môi trường thực tế để kiểm tra tính ứng dụng và hiệu suất thực tế.

1.10.2 Lý do chọn đề tài

Trong thời đại công nghệ phát triển như hiện nay, các thị trường tài chính đang liên tục đối mặt với sự biến động và các tình huống không dự đoán trước được, chịu ảnh hưởng từ nhiều yếu tố như biến động giá cả, thông tin thị trường, và tâm lý nhà đầu tư. Mô hình LSTM, với khả năng xử lý và nhớ thông tin lịch sử, cho thấy tiềm năng lớn trong việc mô phỏng những quan hệ phức tạp này. LSTM được thiết kế đặc biệt cho dữ liệu dạng chuỗi thời gian, làm cho nó trở nên quan trọng trong việc giải quyết những thách thức liên quan đến biến động giá cổ phiếu, lợi nhuận, và các chỉ số kinh tế khác. LSTM không chỉ giữ lại thông tin quan trọng từ quá khứ mà còn sử dụng nó để dự đoán tương lai, làm cho nó trở nên lý tưởng cho việc dự báo trong lĩnh vực tài chính. Nó cũng có khả năng học hỏi mối quan hệ cả tuyến tính và phi tuyến tính, giúp nó trở nên vô cùng hữu ích trong ngành tài chính, nơi mà các mối quan hệ thường không tuân theo quy luật tuyến tính.

Phát triển một mô hình LSTM trong dự báo tài chính không chỉ mang lại lợi ích đáng kể trong việc hỗ trợ các quyết định đầu tư mà còn tiếp tục thúc đẩy sự phát triển của trí tuệ nhân tạo trong ngành tài chính. Mô hình này, nếu có thể dự đoán chính xác các xu hướng thị trường và biến động giá cả, sẽ là công cụ quý giá cho các nhà đầu tư và quản lý rủi ro trong việc ra quyết định dựa trên thông tin và dữ liệu phân tích. Sự thành công của dự án này không chỉ chứng minh hiệu quả của nó mà còn góp phần vào sự tiến triển của trí tuệ nhân tạo trong lĩnh vực này, mở rộng hiểu biết của cộng đồng nghiên cứu và ngành công nghiệp về khả năng và hạn chế của những kỹ thuật máy học.

Do đó, việc chọn lựa đề tài này không chỉ dựa vào việc áp dụng một công nghệ tiên tiến như LSTM, mà còn vì giá trị ứng dụng thực tế cao của nó trong lĩnh vực tài chính, đòi hỏi sự am hiểu sâu sắc về cả kỹ thuật máy học và kiến thức tài chính.

1.10.3 Mục tiêu đề tài

Mục tiêu chính

Phát triển một mô hình dự báo trong lĩnh vực tài chính sử dụng công nghệ LSTM (Long Short-Term Memory), nhằm mục đích cải thiện năng lực dự báo và quản lý rủi ro trong ngành tài chính.

Mục tiêu nhỏ 1: Nghiên cứu và hiểu về LSTM

Trước những đặc thù phức tạp của dữ liệu tài chính, mục tiêu này tập trung vào việc nghiên cứu sâu về mô hình LSTM, cách thức hoạt động của nó trong việc dự đoán các chuỗi dữ liệu thời gian. Các hoạt động cụ thể gồm:

Nghiên cứu kỹ lưỡng về cấu trúc và nguyên lý hoạt động của LSTM, hiểu rõ cách mô hình này xử lý và duy trì thông tin trong dài hạn. Khảo sát các ứng dụng thực tế của LSTM trong ngành tài chính như dự đoán giá cổ phiếu, biến động thị trường và các chỉ số quan trọng khác. Phân tích ưu và nhược điểm của LSTM trong dự báo tài chính, từ đó phát triển phương pháp tiếp cận hiệu quả nhất cho từng loại dữ liệu.

Mục tiêu 2 : Chuẩn bị và xử lý dữ liệu

Một tập dữ liệu chất lượng cao và chính xác là yếu tố quan trọng cho sự thành công của mô hình LSTM. Các bước chính bao gồm:

Thu thập dữ liệu tài chính từ các nguồn đáng tin cậy, đảm bảo độ chính xác và đầy đủ. Tiến hành các bước tiền xử lý như chuẩn hóa, loại bỏ nhiễu và xử lý các vấn đề dữ liệu

cụ thể. Phân chia dữ liệu thành tập huấn luyện và kiểm tra, xác định chiến lược phù hợp để đảm bảo mô hình được huấn luyện và đánh giá một cách toàn diện.

Mục tiêu 3 :Xây dựng mô hình LSTM

Bước này tập trung vào quá trình xây dựng và đào tạo mô hình LSTM để dự báo tài chính. Các bước cụ thể bao gồm:

Thiết kế kiến trúc mô hình LSTM phù hợp với đặc điểm của dữ liệu tài chính, bao gồm số lớp, số nơ-ron, và các tham số quan trọng khác.

Xác định và tinh chỉnh các siêu tham số của mô hình, bao gồm tỷ lệ học, số vòng lặp, và kích thước batch.

Thực hiện quá trình đào tạo trên tập huấn luyện và theo dõi hiệu suất trên tập kiểm tra để đảm bảo mô hình hội tụ một cách hiệu quả. Đánh giá hiệu suất của mô hình trên tập kiểm tra, so sánh với các mô hình dự báo khác và đánh giá khả năng ứng dụng thực tế của nó. Bằng cách thực hiện các mục tiêu nhỏ này, mục tiêu lớn của việc xây dựng mô hình dự báo tài chính sẽ được đạt được một cách có hệ thống và hiệu quả.

Mục tiêu 4 : Tối ưu hóa và đánh giá mô hình

Giai đoạn này tập trung vào việc đánh giá và cải tiến mô hình LSTM: Đánh giá hiệu suất mô hình qua các chỉ số chính xác và độ tin cậy. Sử dụng các kỹ thuật tối ưu hóa tiên tiến để điều chỉnh tham số mô hình, nhằm đạt được hiệu quả cao nhất. Thử nghiệm độ bền và ổn định của mô hình trên các tập dữ liệu kiểm thử độc lập và đối mặt với biến động thị trường.

Phân tích khả năng quản lý rủi ro của mô hình, đánh giá rủi ro liên quan và điều chỉnh chiến lược đầu tư dựa trên dự đoán của mô hình.

Cuối cùng, phân tích khả năng quản lý rủi ro của mô hình là bước quan trọng trong việc xác định khả năng của mô hình đối với rủi ro tài chính. Bằng cách xác định và đánh giá rủi ro liên quan đến các dự đoán của mô hình, có thể cung cấp thông tin quan trọng cho quản lý rủi ro và điều chỉnh chiến lược đầu tư. Đồng thời, việc xác định các biến quan trọng và ảnh hưởng giúp tạo nền tảng cho việc liên tục cải thiện và tối ưu hóa mô hình trong các tình huống thực tế của thị trường tài chính đầy biến động.

1.11 Cấu trúc của báo cáo

Để hiểu rõ hơn về báo cáo có cấu trúc như thế nào thì đây là phần miêu tả

CHƯƠNG 1: GIỚI THIỆU NƠI THỰC TẬP VÀ ĐỀ TÀI : Cung cấp thông tin tổng quan về nơi thực tập, bao gồm lịch sử phát triển, cơ cấu tổ chức, hoạt động, và đề tài thực tập cụ thể.

CHƯƠNG II: CƠ SỞ LÝ LUẬN VỀ VẤN ĐỀ NGHIÊN CỨU : Trình bày lý thuyết cơ bản liên quan đến vấn đề nghiên cứu, bao gồm lý thuyết về dự báo trong tài chính và kinh doanh, cũng như công cụ và kỹ thuật sử dụng trong nghiên cứu.

CHƯƠNG III: MÔ HÌNH LÝ THUYẾT : Đưa ra mô hình lý thuyết được sử dụng trong nghiên cứu, kể cả lý thuyết về phân tích chuỗi thời gian, các chỉ số kỹ thuật, và mô hình LSTM.

CHƯƠNG IV: THỰC NGHIỆM : Mô tả quá trình thực nghiệm, bao gồm mô tả dữ liệu, phân tích, xây dựng mô hình LSTM, và đánh giá mô hình.

CHƯƠNG V: KẾT LUẬN VÀ KIẾN NGHỊ : Tổng kết kết quả nghiên cứu, đưa ra kết luận và các kiến nghị dựa trên phân tích và kết quả nghiên cứu.

DANH MỤC TÀI LIỆU THAM KHẢO : Liệt kê các nguồn tài liệu tham khảo sử dụng.

CHƯƠNG II. CƠ SỞ LÝ LUẬN VỀ VẤN ĐỀ NGHIÊN CỨU

2.1 Định nghĩa và tầm quan trọng của dự báo trong tài chính

Dự báo tài chính là việc sử dụng dữ liệu lịch sử và phân tích để dự đoán sự kiện tài chính trong tương lai. Nó chủ yếu hỗ trợ quyết định đầu tư, quản lý rủi ro và lập kế hoạch chiến lược.

Dự báo đóng vai trò quan trọng trong hình thành chiến lược kinh doanh, nhận diện cơ hội và thách thức, cũng như giúp quản lý tài chính hiệu quả. Đây là công cụ quan trọng trong môi trường kinh doanh đầy biến động.

2.2 Các nguyên tắc cơ bản của dự báo tài chính và kinh doanh

Trong lĩnh vực dự báo tài chính và kinh doanh, các nguyên tắc cơ bản đóng vai trò quan trọng trong việc xây dựng những mô hình dự báo hiệu quả. Tính liên tục của dữ liệu là một nguyên tắc quan trọng, yêu cầu sự theo dõi và thu thập thông tin liên tục để xác định xu hướng chính xác và đưa ra dự báo đáng tin cậy. Đồng thời, sự ổn định của mô hình dự báo theo thời gian là yếu tố quyết định để áp dụng mô hình cho các điều kiện thị trường thực tế.

Tính chất chu kỳ của thị trường cũng được nhấn mạnh, vì hiểu biết về những biến động này giúp dự báo mạnh mẽ hơn và nhận diện những cơ hội và thách thức. Điều này kết hợp với sự hiểu biết vững về môi trường kinh tế toàn cầu và các yếu tố tác động đến thị trường, bao gồm cả các yếu tố kinh tế, chính trị, và xã hội.

Mô hình dự báo cần phải linh hoạt và chính xác để đối mặt với những biến động bất ngờ trong môi trường kinh doanh. Việc đánh giá rủi ro và thiết lập biện pháp phòng ngừa cũng là một phần quan trọng, giúp giảm thiểu tác động tiêu cực và làm cho quá trình dự báo trở nên đáng tin cậy và hiệu quả hơn. Những nguyên tắc này không chỉ giúp tăng cường độ chính xác của dự báo mà còn hỗ trợ quyết định thông minh và linh hoạt trong môi trường kinh doanh đầy biến động.

2.3 Lý thuyết nền tảng về học máy và dự báo tài chính

Machine Learning đóng một vai trò quan trọng trong lĩnh vực dự báo tài chính, mang lại sự hiện đại và tính chính xác cao. Sự tích hợp của học máy trong dự báo tài chính

thường bao gồm việc sử dụng các mô hình dựa trên dữ liệu để phát hiện mẫu và xu hướng không thể dễ dàng nhận diện bằng các phương pháp truyền thống.

2.3.1 Supervised Learning

Học có giám sát giúp mô hình học từ dữ liệu huấn luyện có nhãn, tức là dữ liệu đã được gán nhãn trước đó. Trong dự báo tài chính, điều này có thể áp dụng cho việc dự đoán giá cổ phiếu dựa trên dữ liệu lịch sử.

2.3.2 Unsupervised Learning

Học không giám sát tập trung vào việc khám phá cấu trúc ẩn trong dữ liệu mà không cần nhãn. Trong ngữ cảnh dự báo tài chính, nó có thể giúp phân loại các nhóm đối tượng tài chính mà không cần biết trước về chúng.

2.3.3 Ứng dụng

Học máy có giám sát và học máy không giám sát đều đóng vai trò quan trọng trong lĩnh vực dự báo tài chính, mang lại những lợi ích đặc biệt và đồng thời đối mặt với những thách thức cụ thể.

Mô hình học có giám sát, dựa trên dữ liệu đã được gán nhãn, tăng cường khả năng dự báo thông qua việc học mối quan hệ giữa các đặc trưng và kết quả đã được nhãn. Điều này không chỉ giúp dự đoán giá cổ phiếu mà còn hỗ trợ quyết định đầu tư thông minh. Tuy nhiên, đối mặt với thách thức của việc yêu cầu lượng lớn dữ liệu đã được gán nhãn và nguy cơ overfitting khi mô hình quá mức học thuật từ dữ liệu đào tạo.

Ngược lại, mô hình học không giám sát không cần dữ liệu nhãn và tập trung vào việc khám phá cấu trúc ẩn trong dữ liệu. Nó có thể tự động phân loại và khám phá nhóm tài chính, đem lại khả năng tự động hóa quy trình phân loại và nhóm. Tuy nhiên, mô hình này thường tạo ra kết quả khó giải thích và đôi khi phụ thuộc nhiều vào lượng lớn dữ liệu.

Tổng cộng, cả hai loại học máy đều mang lại những cơ hội đáng kể trong lĩnh vực dự báo tài chính. Tuy nhiên, để tận dụng hết tiềm năng, chúng ta cần đối mặt với những thách thức của việc quản lý dữ liệu lớn, đảm bảo tính giải thích được của mô hình, và xử lý những rủi ro tiềm ẩn. Điều này sẽ giúp chúng ta phát triển những công cụ mạnh mẽ và linh hoạt hơn trong việc đối mặt với động lực phức tạp của thị trường tài chính ngày nay.

2.4 Khám phá các mô hình dự báo truyền thống và hiện đại

Trong lĩnh vực dự báo tài chính, sự so sánh giữa các mô hình truyền thống như ARIMA và các mô hình hiện đại như LSTM giúp ta hiểu rõ về tính linh hoạt và hiệu suất của từng phương pháp.

Mô hình ARIMA, dựa trên phân tích thống kê, thường được ưa chuộng khi làm việc với các chuỗi dữ liệu có xu hướng và chu kỳ rõ ràng. Với ưu điểm dễ hiểu và thực hiện, ARIMA là lựa chọn phù hợp cho các tình huống mà dữ liệu biểu hiện sự biến động theo các quy luật tuyến tính. Tuy nhiên, mô hình này có hạn chế đối với dữ liệu phi tuyến và có biến động phức tạp.

Ngược lại, LSTM, một mô hình hiện đại trong họ mạng nơ-ron hồi quy, nổi bật với khả năng xử lý chuỗi dữ liệu dài hạn và giữ lại thông tin lâu dài qua các chuỗi dữ liệu. Điều này làm cho LSTM trở nên hữu ích khi đối mặt với các dữ liệu tài chính động và phức tạp. Tuy nhiên, độ phức tạp của LSTM yêu cầu lượng lớn dữ liệu và tài nguyên tính toán cao.

Tổng cộng, sự lựa chọn giữa ARIMA và LSTM phụ thuộc vào đặc tính cụ thể của dữ liệu và yêu cầu của mô hình dự báo. ARIMA thích hợp cho dữ liệu tuyến tính, trong khi LSTM thích hợp cho dữ liệu phi tuyến và phức tạp. Quyết định giữa hai mô hình này cần xem xét kỹ lưỡng để đảm bảo phương pháp dự báo phù hợp nhất với tình huống tài chính cụ thể.

2.5 Khung lý thuyết về chuỗi thời gian và phân tích dữ liệu

Chuỗi thời gian là một dạng dữ liệu trong đó các quan sát được ghi lại theo thời gian tuần tự. Trong lĩnh vực dự báo tài chính, phân tích chuỗi thời gian đóng vai trò quan trọng để hiểu biến động và xu hướng của dữ liệu theo thời gian.

Phân rã chuỗi thời gian: Tách chuỗi thành các thành phần như xu hướng, chu kỳ và ngẫu nhiên để hiểu rõ cấu trúc của dữ liệu.

Phương pháp trung bình động: Tính trung bình của giá trị liên tiếp trong khoảng thời gian để làm mịn dữ liệu và giảm nhiễu, từ đó nhận biết xu hướng dài hạn.

Phân tích tần suất: Sử dụng biến đổi Fourier hoặc các phương pháp khác để xác định chu kỳ và tần suất trong chuỗi thời gian.

Tóm lại phân tích chuỗi thời gian không chỉ giúp hiểu biến động thị trường mà còn dự báo xu hướng tương lai. Áp dụng những phương pháp này tạo ra cái nhìn toàn diện và chính xác về thị trường tài chính, hỗ trợ quyết định đầu tư và rủi ro.

2.6 Tiếp cận nghiên cứu: phương pháp luận và kỹ thuật

Ở mục này tập trung vào việc mô tả chi tiết các phương pháp luận và kỹ thuật nghiên cứu được sử dụng. Bước đầu tiên trong quá trình nghiên cứu là thu thập dữ liệu, nơi chúng tôi lựa chọn các nguồn dữ liệu công khai hoặc cơ sở dữ liệu chuyên ngành để thu thập dữ liệu tài chính, chẳng hạn như giá cổ phiếu và chỉ số thị trường. Tiếp theo, quá trình tiền xử lý dữ liệu được thực hiện để đảm bảo tính chính xác và sạch sẽ của dữ liệu, bao gồm việc loại bỏ nhiễu, xử lý dữ liệu thiếu và chuẩn hóa.

Phân phân tích dữ liệu của nghiên cứu này bao gồm việc sử dụng cả phương pháp thống kê và học máy. Mục tiêu là khám phá dữ liệu, tìm kiếm mẫu và xu hướng, và áp dụng các phương pháp như học có giám sát và không giám sát để phân tích và dự báo. Khi xây dựng mô hình, chúng tôi chú trọng đến lựa chọn tham số và cấu trúc mô hình phù hợp, cũng như tối ưu hóa mô hình thông qua các kỹ thuật như cross-validation.

Cuối cùng, đánh giá mô hình là một bước quan trọng trong nghiên cứu. Áp dụng các chỉ số đánh giá như MSE, RMSE, và MAE để đánh giá độ chính xác và độ tin cậy của các dự đoán mô hình. Qua đó, chúng tôi có thể xác định được điểm mạnh và yếu của từng mô hình, cũng như cung cấp cơ sở cho việc so sánh chúng với nhau. Phương pháp luận này giúp đảm bảo tính khoa học và chính xác trong quá trình nghiên cứu, từ đó đóng góp vào việc cung cấp cái nhìn sâu sắc và hữu ích về việc dự báo trong lĩnh vực tài chính.

2.7 Tổng kết

Trong chương này đã trình bày và tóm tắt những khía cạnh quan trọng và nền tảng lý thuyết cần thiết cho nghiên cứu. Nghiên cứu đã bắt đầu bằng cách định nghĩa và nhấn mạnh tầm quan trọng của dự báo trong lĩnh vực tài chính, từ đó chuyển sang phân tích các nguyên tắc cơ bản của dự báo tài chính và kinh doanh. Sự tập trung vào lý thuyết nền tảng về học máy và ứng dụng của nó trong dự báo tài chính là bước quan trọng để hiểu rõ hơn về tiềm năng và thách thức của kỹ thuật này.

Nghiên cứu cũng đã khám phá và so sánh giữa các mô hình dự báo truyền thống và hiện đại, đề cập đến cách chúng được sử dụng để hiểu và dự đoán các xu hướng tài chính.

Một phần quan trọng khác của chương này là phân tích chuỗi thời gian, nơi chúng tôi đã xem xét cách thức phân tích dữ liệu chuỗi thời gian có thể hỗ trợ trong việc đưa ra dự đoán chính xác hơn.

Cuối cùng, đã trình bày phương pháp luận và kỹ thuật nghiên cứu, từ việc thu thập và tiền xử lý dữ liệu đến phương pháp phân tích và đánh giá mô hình. Phần này không chỉ giúp định hình cách tiếp cận nghiên cứu của chúng tôi mà còn góp phần vào việc xây dựng một nền móng vững chắc cho các phân tích và kết luận sắp tới.

Tóm lại, chương II không chỉ cung cấp một cái nhìn tổng quan về các khái niệm, lý thuyết, và phương pháp liên quan đến dự báo tài chính và học máy mà còn đặt nền tảng vững chắc cho các chương tiếp theo của khóa luận này, giúp đảm bảo một tiếp cận toàn diện và hệ thống trong việc nghiên cứu và phân tích.

CHƯƠNG III. MÔ HÌNH LÝ THUYẾT

3.1 Lý thuyết về dự báo trong tài chính và kinh doanh.

Lý thuyết dự báo luôn là quan trọng trong lĩnh vực kinh doanh mảng tài chính, đó là một bước đi vô cùng quan trọng và thông minh trong quá trình ra quyết định. Dự báo là quá trình áp dụng những thông tin từ quá khứ cho tới hiện tại để tạo ra các thông tin hỗ trợ công việc tính toán cho tương lai, hỗ trợ cho tổ chức doanh nghiệp và cá nhân trong thị trường tài chính và kinh doanh để hiểu và lập kế hoạch cho những tính huống xấu nhất có thể xảy ra trong tương lai.

3.1.1 Giới thiệu về dự báo trong tài chính và kinh doanh.

Dự báo trong lĩnh vực này đóng vai trò quan trọng trong quá trình đưa ra quyết định và lập kế hoạch kinh doanh. Quá trình này giúp cho biết về các sự kiện tương lai, bao gồm thông tin về lợi nhuận, doanh thu, và các yếu tố quan trọng khác. Nó mang lại cho người tham gia tham khảo quan trọng cho việc quản lý tài chính và hình thành chiến lược kinh doanh, đặc biệt là khi chúng ta muốn hiểu về lợi nhuận.

3.1.2 Khái niệm về dự báo

Luôn đóng vai trò rất quan trọng trong công việc dự đoán hay là đánh giá các vấn đề ở tương lai. Quá trình này chủ yếu dựa vào việc sử dụng dữ liệu hiện tại và quá khứ để tạo ra các ước tính về những điều có thể được sẽ xảy ra trong tương lai. Ở ngữ cảnh này, dự báo tập trung vào việc dự đoán lợi nhuận, doanh thu, và các yếu tố khác trong lĩnh vực tài chính.

Quy trình được thể hiện như sau:

Bước 1. Thu thập dữ liệu về giá đóng cửa cổ phiếu của công ty Apple từ các nguồn đáng tin cậy ví dụ như là yahoo finance.

Bước 2. Tiền xử lý dữ liệu để xử lý các giá trị thiếu, sử dụng các kỹ thuật như điền giá trị trung bình hoặc giá trị tối đa.

Bước 3. Phân chia dữ liệu thành hai phần tập huấn luyện và tập kiểm tra. Tập huấn luyện sẽ dùng cho công việc giúp cho mô hình học tập, trong khi tập kiểm tra dùng trong việc đưa ra đánh giá cho hiệu suất mô hình.

Bước 4. Sử dụng ARIMA với LSTM để huấn luyện cho việc dự báo lợi nhuận. cả hai được đào tạo để học từ dữ liệu từ tập huấn luyện.

Bước 5. Sử dụng tập kiểm tra để tiến hành đánh giá về hiệu suất của dự báo. Kết quả thu được dùng so sánh để xem xét xác định mô hình nào có hiệu suất tốt nhất.

Bước 6. Tinh chỉnh mô hình để cải thiện độ chính xác trong dự đoán và hiệu suất cả hai.

Bước 7. Sử dụng các mô hình đã được huấn luyện để dự báo lợi nhuận trong tương lai.

Bước 8. Đưa ra một đánh giá khách quan của dự báo để đảm bảo rằng chúng đủ chính xác và đáng để ta tin cậy và đưa ra các quyết định liên quan đến kinh doanh.

3.1.3 Các loại dự báo.

Dự báo tương lai đóng vai trò khá là quan trọng trong những quá trình đưa ra những quản lý và lập kế hoạch, nhằm đảm bảo sự phát triển được ổn định với bền vững cho tổ chức. Thông qua việc đưa ra nhiều phân loại để đưa vào dự báo theo thời gian, chúng ta có thể được hiểu biết rõ ràng hơn về mục tiêu hay là phạm vi của từng loại dự báo.

Dự báo dài hạn sẽ kéo dài từ 5 năm trở lên, mang đặc tính chiến lược giúp định hướng cho tương lai. Đây là công cụ mạnh mẽ để cho quản lý, chính trị gia, và các nhà quy hoạch xây dựng mục tiêu hoặc chiến lược cho quốc gia hoặc tổ chức. Ở kinh tế thị trường, dự báo dài hạn giúp xác định cơ hội và thách thức chiến lược, định hình chính sách kinh tế, và định hình hướng phát triển dài hạn. Nó cũng đóng vai trò quan trọng trong việc đảm bảo bền vững về môi trường, xã hội và kinh tế.

Trung hạn, với thời gian 3 tới 5 năm, thường được dùng xây dựng các kế hoạch chi tiết và chiến lược cho kỳ vọng trung hạn. Nó cung cấp cho chúng ta thông tin nếu cần thiết để quản lý giúp ta đẩy nhanh công việc thực hiện dự án, thực nghiệm và cải tiến quy trình. Nó vô cùng hữu ích để quy hoạch và tối ưu hóa tài nguyên, giúp quản lý những rủi ro, và chắc rằng các mục tiêu về trung hạn cho kinh tế, xã hội, khoa học và văn hoá có thể dễ dàng hình dung được.

Dự báo ngắn hạn, có thời gian dưới 3 năm, chủ yếu phục vụ quyết định kịp thời và thực hiện các kế hoạch trong tương lai gần. Rất quan trọng cho quản lý hàng ngày và đảm bảo rằng nó đạt được tính hiệu quả của các hoạt động cho hiện tại. Dự báo ngắn hạn cung cấp thông tin về tình hình ở hiện tại cho dự kiến đến tương lai gần, giúp quyết định về mục tiêu tài chính, nhân sự và các hoạt động được diễn ra ở thường ngày.

Sự kết hợp giữa cả ba dự báo này sẽ dễ dàng giúp quản lý hay các lãnh đạo hiểu rõ hơn về động lực và xu hướng dài hạn, xây dựng chiến lược trung hạn và đảm bảo hiệu quả trong việc quản lý chính chu hơn các hoạt động hàng ngày. Các dự báo này đóng vai trò cực kỳ vô cùng quan trọng trong việc tạo ra một tương lai bền rất bền vững và phát triển cho các quốc gia hoặc tổ chức.

3.2 Dữ liệu chuỗi thời gian.

Dữ liệu chuỗi thời gian luôn đóng một vai trò vô cùng quan trọng và phổ biến trong nhiều ngành nghề như thống kê, kinh tế học đặc biệt là khoa học dữ liệu. Loại dữ liệu này ghi lại thông tin qua các khoảng thời gian cố định, giúp chúng ta theo dõi các biến động và thay đổi của giá trị mà chúng ta quan tâm theo thời gian. Việc này mang lại cơ hội phân tích và dự đoán xu hướng, chu kỳ, và biến động trong tương lai, hỗ trợ quyết định một cách vô cùng là hiệu quả.

Trong lĩnh vực đang được nghiên cứu, dữ liệu của chuỗi thời gian đặc biệt quan trọng. Ví dụ, thông tin về giá cổ phiếu của Apple từ năm 2013 đến 2023 có ý nghĩa lớn đối với những nhà đầu tư và nhà phân tích. Điều này giúp xác định xu hướng dài hạn hay ngắn hạn của giá cổ phiếu, hiểu rõ sự ảnh hưởng của nó đến các sự kiện kinh tế và thay đổi trong ngành công nghiệp công nghệ. Những giá trị này hỗ trợ quyết định đầu tư và cung cấp cái nhìn về diễn biến thị trường.

Dữ liệu chuỗi thời gian cũng sử dụng để dự đoán tình hình trong tương lai cho một công ty. Các mô hình được dùng dự báo dữ liệu chuỗi thời gian như ARIMA hoặc mạng LSTM có thể áp dụng vào các thông tin dữ liệu từ lịch sử để dự đoán giá cổ phiếu, doanh thu, hoặc lợi nhuận. Điều này giúp chuẩn bị kế hoạch kinh doanh và đầu tư một cách hiệu quả.

Ngoài lĩnh vực tài chính, Nó còn được ứng dụng rộng rãi trong nhiều lĩnh vực khác. Ví dụ, trong dự báo thời tiết, về nhiệt độ, độ ẩm, áp suất, và gió được sử dụng để đưa ra thông tin dự báo thời tiết trong tương lai. Trong quản lý sản xuất, dữ liệu về sản lượng và hiệu suất sản xuất qua các giai đoạn thời gian giúp tối ưu hóa các quy trình trong sản xuất và dự đoán các nhu cầu từ thị trường. Dữ liệu mạng xã hội cũng được thu thập để theo dõi thay đổi trong lượng truy cập trang website hoặc tương tác trên mạng xã hội, và phân tích xu hướng truyền thông.

Tổng cộng, dữ liệu chuỗi thời gian vô cùng quan trọng nó giúp chúng ta hiểu và dự đoán biến động theo thời gian và đưa ra nhiều khía cạnh của cuộc sống và kinh tế. Bằng việc kết hợp với các phương pháp trong phân tích thống kê và kỹ thuật dự báo, nó có thể giúp ra những quyết định thông minh và tối ưu kế hoạch trong nhiều lĩnh vực khác nhau.

3.3 Lý thuyết về Time Series Decomposition

Phương pháp dùng để phân rã Timeseries Decomposition, là một kỹ thuật giúp tách một data từ chuỗi thời gian thành những thành phần rất cơ bản, nhằm hiểu rõ hơn về xu hướng, chu kỳ, và thành phần ngẫu nhiên. Thông thường, Việc phân rã ra ba thành phần chính: xu hướng dữ liệu với chu kỳ, và thành phần được cho là ngẫu nhiên [1].

Xu hướng thể hiện hướng chung của chuỗi thời gian, có thể là tăng, giảm, hoặc ổn định theo thời gian. Chu kỳ đại diện cho các biến động lặp lại trong chuỗi thời gian, ví dụ như sự tăng mạnh trong doanh số bán hàng vào mùa giáng sinh. Thành phần ngẫu nhiên là sự biến động không thể giải thích bằng xu hướng hoặc chu kỳ.

Có hai phương pháp chính để phân rã chuỗi thời gian: phân rã Additive và phân rã Multiplicative.

Phân rã Additive theo công thức :

$$Y_t = Trend_t + Seasonal_t + Residual_t \quad (1)$$

Phân rã Multiplicative thì nó nhân với nhau

$$Y_t = Trend_t * Seasonal_t * Residual_t \quad (2)$$

Ở đây Y_t là giá trị mua bán tại thời điểm t

$Trend_t$ cho là xu hướng mua bán tại t

$Seasonal_t$ thành phần của chu kỳ tại t

$Residual_t$ là thành phần là ngẫu nhiên tại t

Phương pháp phân rã chuỗi thời gian giúp ta nắm bắt cấu trúc của dữ liệu, từ đó hỗ trợ trong việc dự đoán và phân tích theo thời gian.

3.3.1 Shifting and lags

"Shifting" và "lags" quan trọng trong quá trình đưa ra thẩm định tập dữ liệu thời gian, để hiểu rõ về sự biến động, xu hướng của dữ liệu theo thời gian. Shifting đề cập đến việc dịch chuyển dữ liệu trong bước thời gian cố định. Khi chúng ta thực hiện "shifting,"

giá trị của dữ liệu tại mỗi điểm thời gian được dịch chuyển đến phía trước hoặc phía sau một số bước nhất định. Điều này giúp quan sát được sự biến đổi dữ liệu và tìm hiểu thêm về mối quan hệ các quan sát với nhau.

Lags là khái niệm liên quan, mô tả thời gian ở giữa hai sự kiện trong chuỗi thời gian. Lag dương đại diện cho sự trễ theo thời gian dương, trong khi lag âm đại diện cho sự trễ theo thời gian âm. Việc này giúp chúng ta đánh giá tương tác thời gian giữa các sự kiện và hiểu rõ hơn của sự biến động trong tập data.

Sự giúp đỡ linh hoạt giữa "shifting" với "lags" cung cấp cho người nhìn Một cái nhìn sâu rộng và chi tiết hơn về cấu trúc thời gian của dữ liệu, từ đó hỗ trợ trong việc dự đoán thông qua xu hướng và có được quyết định chính xác trong lĩnh vực phân tích dữ liệu chuỗi thời gian.

3.4 Technical Indicators.

Chỉ số kỹ thuật khá quan trọng với việc phân tích các kỹ thuật, hỗ trợ nhà đầu tư và người giao dịch hiểu rõ hơn về hướng đi và sức mạnh của những thị trường tài chính. Các chỉ số đa dạng, từ những công cụ theo dõi xu hướng trung bình động (SMA) đến các chỉ số đo lường động lượng như Relative Strength Index (RSI). Chiến lược theo dõi xu thế được cho là rất quan trọng trong giao dịch bằng những kỹ thuật, với việc áp dụng SMA và EMA để xác định rõ ràng hơn hướng chung của giá. Các chỉ số động lượng như RSI đưa ra thông tin về hai trạng thái quá mua hoặc quá bán của một hay nhiều tài sản, trong khi các chỉ số biến động như Bollinger Bands dùng để đo những mức độ tỉ lệ biến động của các giá trị.

Các chỉ số dao động như MACD đưa ra tín hiệu mua/bán thông qua sự cắt giao giữa những đường trung bình động. Ở việc cứu này sẽ tập trung vào các danh sách chỉ số báo kỹ thuật được các chuyên viên sử dụng rộng rãi cũng như những chỉ báo mà tôi tin là có lợi nhất trong giao dịch tự động. Danh sách các chỉ số như sau:

3.4.1 Simple Moving Average

SMA là công cụ khá là quan trọng trong công việc đưa ra các kỹ thuật, nó đo lường các xu hướng chung thị trường hoặc là tài sản cụ thể. Chức năng của SMA là làm mịn dữ liệu thông qua tính những giá trị mean của chuỗi khoảng Time Series nhất định, giúp loại bỏ những biến động ngắn hạn không quan trọng.

Người giao dịch lựa chọn khoảng thời gian hợp lý với các chiến lược đầu tư của họ, chẳng hạn như SMA 10 ngày, 50 ngày hoặc 200 ngày. SMA cung cấp tín hiệu quan trọng như "Golden Cross" (khi giá vượt lên trên đường SMA, tạo ra tín hiệu đó là mua) , "Death Cross" (khi giá xuống dưới SMA, tạo ra tín hiệu bán). Tuy nhiên, SMA cũng có điểm yếu khi không đưa ra được chính xác biến động của thị trường một cách vô cùng nhanh chóng luôn để có thể tạo ra tín hiệu trễ. Do đó, nhiều nhà đầu tư thường dùng SMA và những chỉ số và công cụ phân tích khác để có quan sát và đưa ra những con số đầu tư đúng đắn. Ở thực tế, SMA thường được dùng trong chiến lược dài hạn và luôn đóng vai trò của bộ công cụ phân tích của nhà đầu tư. Công thức của SMA được mô tả qua công thức (3):

$$SMA = \frac{P_1 + P_2 + P_3 + \dots + P_n}{n} \quad (3)$$

Trong đó SMA là giá trị trung bình và $P_1 + P_2 + \dots + P_n$ là giá đóng của tài sản trong các buổi giao dịch cụ thể n là số lượng các phiên được dùng để tính SMA. Nếu tính SMA cho giá đóng của cổ phiếu trong 10 ngày gần đây, bạn sẽ lấy giá đóng của trong 10 ngày gần đây và tính được trung bình cộng của chúng.

3.4.2 Exponential Moving Average

Trung bình động mở rộng (EMA) là một biến thể của trung bình động, nhưng với đặc điểm quan trọng là áp dụng trọng số sẽ cao hơn giúp những giá trị mà gần đây hơn so với những giá trị cũ hơn. Điều này giúp EMA nhanh chóng phản ánh sự biến động mạnh mẽ, làm cho nó trở nên khá linh hoạt, nhạy bén hơn so với SMA.

Cho P_t đóng cửa tại thời điểm t , EMA_t ở thời gian t và α là hệ trọng số qua đó ta có công thức được biểu diễn như sau :

$$EMA_t = \alpha \cdot P_t + (1 - \alpha) \cdot EMA_{t-1} \quad (4)$$

EMA_t thể hiện EMA tại thời điểm t

P_t giá thay đổi tại thời gian t

EMA_{t-1} là giá trị EMA thời điểm trước đó $t - 1$

α là trọng số, được tính theo công thức $\alpha = \frac{2}{n+1}$ n biểu thị cho số ngày được sử dụng cho chu kỳ EMA

3.4.3 Relative Strength Index

Chỉ số này đóng vai trò cực kỳ quan trọng trong giai đoạn phân tích, giúp cho nhà đầu tư và những người đo mức độ biến động của giá trị gần đây để đánh giá tình trạng việc mua bán quá mức của một giá trị tài chính.

Để tính toán RSI, chúng ta thực hiện bốn bước chính. Bước đầu tiên là tính toán sự tăng giá (Upward Price Change) và sự giảm giá (Downward Price Change) cho mỗi phiên giao dịch.

$$Gain_i = (Close_i - Close_{i-1}, 0) \quad (5)$$

$$Loss_i = (Close_i - Close_i, 0) \quad (6)$$

Trong đó $Close_i$ là giá đóng phiên giao dịch

$Close_{i-1}$ là giá đóng của phiên trước đó $i-1$

$Gain_i$ là sự tăng giá trong phiên bản i

$Loss_i$ là sự giảm giá trong phiên i

Sau đó tiến hành Average Gain và Average Loss cho n phiên bản giao dịch theo

$$Average\ Gain = \frac{\sum_{i=1}^n Gain_i}{n} \quad (7)$$

$$Average\ Loss = \frac{\sum_{i=1}^n Loss_i}{n} \quad (8)$$

$\sum_{i=1}^n Gain_i$ là tổng của giá trị $Gain_i$ của n phiên giao dịch.

$\sum_{i=1}^n Loss_i$ là tổng của giá trị $Loss_i$ của n phiên.

n được biết là các số lượng phiên giao dịch sẽ sử dụng để tính trung bình.

Relative Strength (RS)

$$RS = \frac{Average\ Gain}{Average\ Loss} \quad (9)$$

Relative Strength Index (RSI)

$$RSI = 100 - \frac{100}{1 + RS} \quad (10)$$

Các bước trên giúp đo lường mức độ mạnh của tăng so với giảm trong giai đoạn của n phiên của giao dịch và chuyển đổi nó thành một chỉ số RSI nằm trong một khoảng mà giá trị bắt đầu 0 đến 100. Công thức này mang lại cái nhìn quan trọng về sự thay đổi của giá trị cổ phiếu và là công cụ được cho là quyết định quan trọng trong giai đoạn này.

3.4.4 Moving Average Convergence Divergence

Bao gồm hai phần chính: đường MACD và đường kích thích (Signal Line). Nó là sự chênh lệch thường được tính bằng cách trừ đường bình quân động dài hạn từ đường bình quân động ngắn hạn. Đường kích thích là một bình quân động giúp làm mịn đường MACD và tạo ra tín hiệu để tiến hành giao dịch. Với chênh lệch giữa đường MACD và đường kích thích thường được áp dụng vào việc đánh giá sức mạnh của phân tích xu thế và dự đoán là những công việc thể hiện sự đảo chiều của thị trường.

$$MACD\ Line = EMA_{12} - EMA_{26} \quad (11)$$

Tính toán Signal Line (đường kích thích): Sử dụng dữ liệu của MACD Line để tính toán đường Signal Line, là EMA của MACD Line với kỳ hạn 9 ngày.

$$Signal\ Line = EMA_{MACD\ Line, 9} \quad (12)$$

3.5 Differencing

Differencing là một phương pháp trong thống kê và phân tích timeseries, được sử dụng để giảm biến động và làm cho chuỗi thời gian trở nên ổn định hơn. Quá trình differencing bao gồm việc tính hiệu của giá trị tại một thời điểm và giá trị tại một thời điểm trước đó. Y_t là giá trị của timeseries tại t thì Differencing thực hiện bằng $Y_t - Y_{t-1}$. Kết quả của quá trình là first difference hoặc hiệu số đầu tiên. Nếu muốn giảm biến động một lần nữa, có thể thực hiện differencing thêm một lần nữa, và kết quả được gọi là second difference, và có thể tiếp tục như vậy.

Kết quả của nó gọi là first difference hoặc hiệu số đầu tiên. Nếu muốn giảm biến động một lần nữa, có thể thực hiện differencing thêm một lần nữa, và kết quả được gọi là second difference, và có thể tiếp tục như vậy.

Differencing có thể giúp loại bỏ xu hướng và tính thời vụ từ chuỗi cũ, tạo ra một chuỗi mới có giá trị mean ổn định và dễ dàng phân tích. Nó thường áp dụng trong việc chuẩn bị data cho các mô hình dự đoán chuỗi thời gian

3.6 Mô hình LSTM.

LSTM là một kiến trúc mạng neural sâu được dùng trong NLP và Timeseries [1]. Đây là một cải tiến từ mô hình mạng RNN, có khả năng lưu trữ các thông tin thời gian dài và được tối ưu hóa giúp bắt các mẫu phi tuyến trong dữ liệu chuỗi của thời gian [13].

LSTM giải quyết thách thức tạo nên các mô hình học máy dự báo chính xác, khi đối mặt với dữ liệu có tính phi tuyến và phi định của chuỗi thời gian. Trong quá trình làm sạch các dữ liệu, khi chuỗi có nhiễu và giá trị thiếu, chúng ta thường áp dụng kỹ thuật làm mịn và thay thế bằng phương pháp mà chúng ta thấy thích hợp [14].

Việc chúng ta sử dụng LSTM vào data yêu cầu thời gian được trực tuyến giúp mô hình phản ứng tốt giữa input của dữ liệu và output, đồng thời có thể so sánh hiệu suất với Một loạt các phương pháp timeseries đã được áp dụng [15,16]. Trong khuôn khổ mô hình LSTM, trạng thái ở tại thời điểm t sẽ mô tả như sau.

LSTM, có một trạng thái ở thời gian t cụ thể được mô tả như sau:

- Đầu ra: c_t , được biết đến như tình trạng ô, h_t là tình trạng ẩn
- Đầu vào (Input): c_{t-1}, h_{t-1}, x_t , x_t là đầu vào là đầu vào tại thời điểm thứ t của mô hình. c_{t-1}, h_{t-1} là đầu ra của tầng trước, có vai trò tương tự với s RNN, c_t là trạng thái ô mới của LSTM.:

x_t : Giá đóng cửa tại thời điểm hiện tại (bước thời gian t).

c_{t-1} : Các biến kỹ thuật hoặc các thông tin khác từ bước thời gian của trước đó.

h_{t-1} : Trạng thái ẩn của LSTM từ bước thời gian của trước đó.

Các thời điểm trong chuỗi, ví dụ như 10 bước thời gian, sẽ cung cấp thông tin cho mô hình LSTM học cấu trúc chuỗi và tiên đoán giá đóng cửa cho thời điểm tiếp theo ($t+1$).

F_t, i_t, o_t , tương ứng với cổng quên, cổng vào và cổng ra

$$\text{- Cổng quên: } f_t = \sigma(U_f * x_t + W_f * h_{t-1} + b_f) \quad (13)$$

$$\text{- Cổng vào: } i_t = \sigma(U_i * x_t + W_i * h_{t-1} + b_i) \quad (14)$$

$$\text{- Cổng ra: } o_t = \sigma(U_o * x_t + W_o * h_{t-1} + b_o) \quad (15)$$

Trong đó: $0 < f_t, i_t, o_t < 1$; b_f, b_i, b_o là các con số từ bias; hệ số của W và U tương tự ở RNN

$\tilde{c} = \tanh \tanh (U_c * x_t + W_c * h_{t-1} + b_c)$ (16), giai đoạn này thì công thức (16) được tính như là s_t trong RNN

Cổng quên xác định mức độ quên từ trạng thái trước và cổng ra quyết định mức độ nạp thông tin mới từ đầu vào và trạng thái trước theo công thức (17) :

$$\tilde{c} = f_c \times c_{t-1} + i_t \times \tilde{c}_t \quad (17)$$

Cổng ra quyết định cách sử dụng trạng thái để tính toán trạng thái ẩn và đầu ra của trạng thái ẩn, ngoài ra h_t trong công thức (13) cũng được sử dụng tính output y_t cho trạng thái t .

$$h_t = o_t \times \tanh(c_t) \quad (18)$$

2.5.1 Sơ đồ của mô hình LSTM.

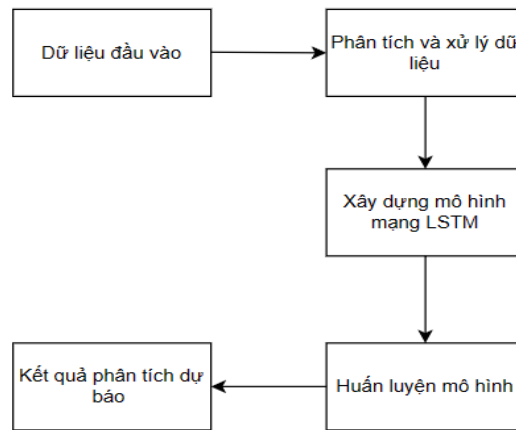
Bước 1: Đi thu thập dữ liệu thông qua các nguồn dữ liệu uy tín, có thông tin đến mục tiêu, giả sử ở đây là dữ liệu về giá đóng cửa.

Bước 2: Để chuẩn bị cho LSTM, thực hiện tiền xử lý các dữ liệu bằng cách làm sạch và xử lý giá trị bị thiếu. Bước này cũng bao gồm phân tích để hiểu các đặc điểm và quan hệ của chúng.

Bước 3: Thiết kế, xây dựng cho kiến trúc mạng LSTM, quyết định số lớp, số nơ-ron và các chỉ số khác của mô hình. Mô hình LSTM sẽ tiến hành học từ các dữ liệu để dự đoán hoặc phân tích các mẫu, xu hướng.

Bước 4: Sử dụng data đã được thông qua tiền xử lý để tiến hành train cho mô hình LSTM, cung cấp dữ liệu và cải thiện trọng số và tham số dựa trên dữ liệu được đào tạo.

Bước 5: Sau khi mô hình đã được học tập, thực hiện dự đoán hoặc phân tích tập dữ liệu. Kết quả áp dụng giúp đưa ra những các quyết định hoặc phân tích những xu hướng và dự đoán trong dữ liệu dựa trên khả năng học của mô hình LSTM.



Hình 2 : Sơ đồ LSTM

3.7 Đánh giá mô hình.

Quá trình thẩm định mô hình là bước không thể thiếu trong việc đánh giá hiệu quả và năng lực dự đoán của các mô hình học máy hoặc phân tích thống kê. Mục đích chính là để xác minh khả năng hoạt động của mô hình với dữ liệu mới, không được sử dụng trong giai đoạn huấn luyện của nó. Qua đó, mô hình được kiểm tra để đảm bảo nó không rơi vào tình trạng quá khớp (overfitting) hay thiếu khớp (underfitting) và có khả năng áp dụng rộng rãi một cách hiệu quả.

3.7.1 Thu thập dữ liệu

Để bắt đầu quá trình đánh giá, Công việc này làm để có được dữ liệu lịch sử về lĩnh vực tài chính, từ các nguồn đáng tin cậy. Với việc sử dụng Python, cùng với các thư viện như pandas, để tự động hóa công việc thu thập dữ liệu.

3.7.2 Xử lý dữ liệu

Khi đã có data , sau đó tiến hành các công việc với dữ liệu để cho quá trình đánh giá một mô hình. Những nhiệm vụ này bao gồm quá trình làm sạch dữ liệu để loại bỏ sự nhiễu hay xử lý giá trị bị thiếu, biến đổi tập dữ liệu để đưa vào định dạng hợp lý, và tạo ra các đặc trưng liên quan đến chuỗi thời gian. Trong tất cả những công đoạn này được thực hiện bằng Python, và chúng tôi sử dụng thư viện pandas thực hiện chúng và sau đó trực quan.

3.7.4 Dự đoán và đánh giá

Sau quá trình đào tạo cả hai ARIMA và LSTM, Việc áp dụng chúng để thực hiện dự đoán các giá trị kết quả trong tương lai. Thông qua Python, chúng tôi đánh giá dự đoán thông qua thực hiện quá trình dự đoán của cả hai. MSE, RMSE, MAE được sử dụng để đánh giá mức độ sai số giữa giá trị dự đoán và giá trị thực tế, thường ta thực hiện điều này bằng cách so sánh giữa các giá trị mà mô hình dự đoán với những giá trị thực tế tương ứng

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (19)$$

Trong việc thực hiện dự báo và ML, RMSE là giá trị đo phổ biến thường được sử dụng ở mục đích đánh giá sự chính xác của dự đoán. RMSE có một trong những ưu điểm là nó trả về kết quả có đơn vị đo giống với dữ liệu gốc, giúp dễ dàng đánh giá mức độ sai số dự đoán và hiểu được ý nghĩa thực tế của nó [19]. Công thức tính RMSE được mô tả như sau (20):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (20)$$

MAE là một chỉ số thống kê được nhiều người biết và được rất nhiều trong lĩnh vực tiên đoán và đánh giá. Nó đo lường được các mức độ cho sai lệch trung bình tuyệt đối giữa các giá trị dự đoán thông qua các giá trị hiện thực. Để tính toán MAE, chúng ta bắt đầu với một tập dữ liệu chứa các cặp giá trị, trong đó một giá trị dùng làm thông tin dự đoán (thường được ký hiệu là \hat{y}), và giá trị còn lại là thông tin của các giá trị thực tế (thường được ký hiệu là y). Để đo lường sai lệch giữa dự đoán, thực tế, chúng ta tính các sai tuyệt đối (absolute error) cho những cặp giá trị bằng cách lấy các giá trị tuyệt đối của hiệu giữa \hat{y} và giá trị thực tế y . Sau đó, chúng ta tính trung vị cộng của tất cả sai khác tuyệt đối trong một tập dữ liệu để có giá trị MAE. Công thức của MAE được thể hiện qua công thức (21).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (21)$$

CHƯƠNG IV : THỰC NGHIỆM

4.1 Mô tả tập dữ liệu

Thông tin về giá cổ phiếu của Apple (APPL) được ghi lại trong vòng hơn 5 năm, bao gồm dữ liệu từ ngày 1 tháng 1 năm 2018 đến ngày 5 tháng 11 năm 2023. Thu thập từ <https://finance.yahoo.com/> bằng cách crawl (khai thác dữ liệu), về giá cổ phiếu trong suốt khoảng thời gian này. Các cột trong bộ dữ liệu bao gồm:

1. Ngày (Date): Đây là ngày giao dịch tương ứng, được sắp xếp theo từng ngày, cho biết ngày cụ thể trong khoảng thời gian đã nêu.
2. Giá mở (Open): giá mở cửa sàn của giao dịch tại thời điểm của ngày hôm đó.
3. Giá đóng (Close): Nó đề cập đến giá của một cổ phiếu riêng lẻ tại thời gian đóng giao dịch trong ngày. Nó đại diện cho lệnh mua-bán cuối cùng được thực hiện giữa hai nhà giao dịch và đây chính là phần quan trọng.
4. Mức giá tối đa (High): Giá cao nhất hiển thị ở ngày giao dịch hôm đó.
5. Mức giá tối thiểu (Low): Nó biểu thị thấp nhất mà cổ phiếu đã đạt được trong ngày.
6. Giá đóng cửa điều chỉnh (Adj Close): Khi đã được thông qua chỉnh sửa thông các yếu tố như cổ tức hoặc chia tách hay là phát hành cổ phiếu mới. Nó giúp duy trì bao gồm được tính toán vẹn thông tin từ dữ liệu chứng khoán.
7. Lượng giao dịch (Volume): Biểu diễn khối lượng được giao dịch trong ngày.

Bảng 1: Bảng dữ liệu cổ phiếu APPLE

Date	Open	High	Low	Close	Adj Close	Volume
2018-01-02	42.54	43.075	42.315	43.065	40.7229	102223600
2018-01-03	43.1325	43.6375	42.99	43.0575	40.7158	118071600
2018-01-04	43.135	43.3675	43.02	43.2575	40.9049	89738400
-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----
11/2/2023	175.5200	177.779	175.46	177.57	177.3363	77334800
11/3/2023	174.240	176.820	173.350	176.649	176.417	79763700

4.1.1 Thông tin cung cấp

Bộ dữ liệu này cung cấp chi tiết về giá cổ phiếu hàng ngày, bao gồm giá khai mạc, giá kết thúc, mức giá cao nhất, mức giá thấp nhất trong ngày, cũng như giá đóng cửa đã

điều chỉnh và giá đóng cửa thực tế. Nó có vai trò quan trọng trong việc theo dõi và phân tích những biến động giá trị của cổ phiếu trong khoảng thời gian được nghiên cứu.

4.1.2 Tại sao giá đóng cửa quan trọng

Nó là giá cuối cùng mà cổ phiếu được giao dịch trong một ngày tại sàn chứng khoán, thường được xem là giá chốt cuộc trong ngày và là một thước đo quan trọng cho giá cả cổ phiếu tại thời điểm đó. Trong quá trình phân tích lịch sử giá cổ phiếu, các tổ chức tài chính và nhà đầu tư thường dựa vào giá đóng cửa như một tiêu chuẩn tham khảo chính vì một số nguyên nhân:

Phản ánh giá trị cuối biểu hiện của giá trị cuối cùng của cổ phiếu trong ngày và thường được xem xét như giá chốt tại thời điểm đó.

Tính toán khoảng thời gian dài: Áp dụng để tính giá trị cổ phiếu cho các khoảng thời gian sẽ dài hơn như tuần, tháng, quý và năm, cung cấp cái nhìn tổng quan về các hiệu suất của cổ phiếu của mình trong những giai đoạn đó.

Thước đo lợi nhuận và hiệu suất: Khoảng cách giữa giá khai mạc và kết thúc của cổ phiếu, được tính bằng tỷ lệ phần trăm, thường được sử dụng để đánh giá hiệu suất hoặc lợi tức của cổ phiếu trong một ngày cụ thể. Đây là một phương tiện quan trọng giúp nhà đầu tư xác định tiềm năng sinh lợi trong ngày đó.

Nói chung, sự chênh lệch này không chỉ phản ánh kỳ vọng cuối ngày của cổ phiếu mà còn là thước đo cho sự biến đổi giá trị trong suốt thời gian giao dịch. Điều này cung cấp cái nhìn chi tiết hơn về hiệu suất của cổ phiếu qua các chu kỳ thị trường khác nhau.

4.2 Trực quan hóa dữ liệu.

Để có cái nhìn chi tiết và nắm bắt xu hướng của tập dữ liệu, việc thông qua các biểu đồ và đồ thị là rất quan trọng. Trực quan hóa dữ liệu để tạo ra hình ảnh minh họa và phân tích rõ ràng hơn về đặc điểm của dữ liệu.

4.2.1 Giá đóng cửa của cổ phiếu

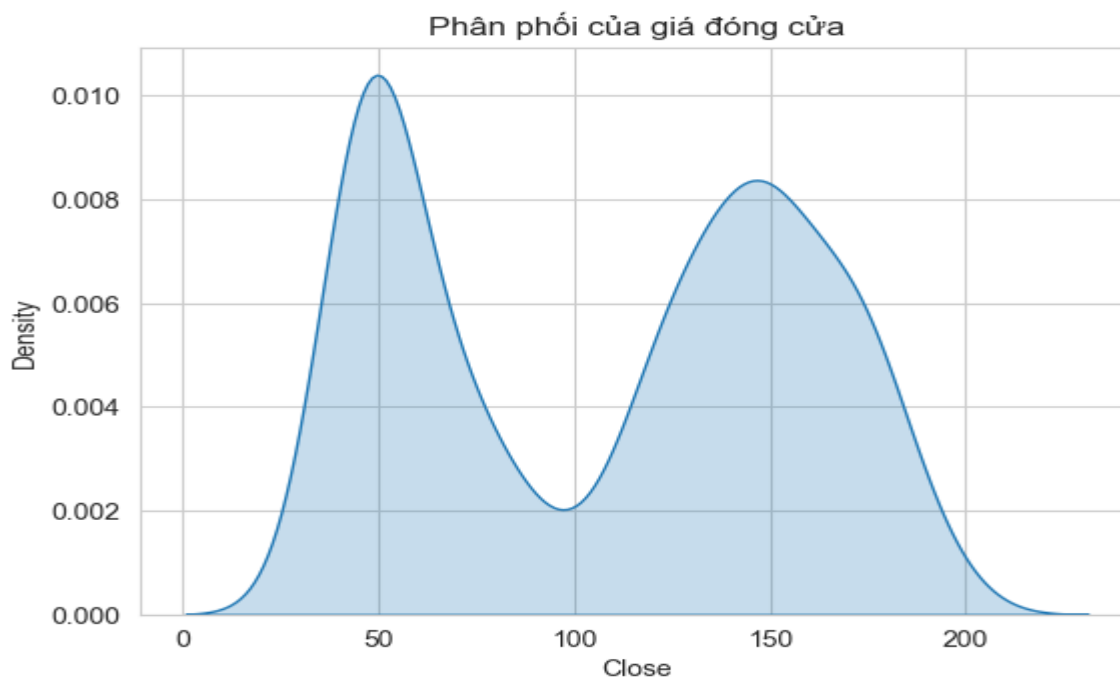


Hình 3 : Hình ảnh biểu diễn giá đóng cửa giai đoạn 2018-2023

Biểu đồ thể hiện giá đóng cửa của thị trường cổ phiếu Apple 2018 - 2023 có thể thấy trước mắt chúng ta đó một hành trình đầy biến động và thành công. Trên trục thời gian x, từ năm 2018 đến năm 2023 được diễn đạt một cách rõ ràng, trong khi trục giá trên y thể hiện giá cổ phiếu theo mức 50,100,150,200. Đường màu xanh lam nét đậm biểu diễn giá đóng cửa thực tế của cổ phiếu, truyền tải thông điệp về sự mạnh mẽ về sự tăng trưởng mạnh mẽ của Apple. Qua những năm, tốc độ tăng trưởng đều đặn của giá cổ phiếu Apple, tạo ra một cuộc hành trình vững chắc và đầy động lực. Các đỉnh và đáy trên đường xanh lam kể lên câu chuyện về những thời kỳ thách thức và những bước nhảy vọt thành công. Mỗi đường đi của thông tin trên biểu đồ hỗ trợ chúng ta xây dựng một cái nhìn về cơ hội và thách thức xuất hiện trong mỗi giai đoạn cụ thể.

Với sự biến động mạnh mẽ này, biểu đồ không chỉ thống kê mà còn thể hiện cho thấy một cửa sổ mở ra về sự phát triển của Apple trong ngữ cảnh thị trường toàn cầu. Những người quan tâm có thể tìm thấy trong đây những dấu hiệu, xu hướng và chiến lược kinh doanh hết sức thiết yếu, hỗ trợ họ trong việc ra những quyết định chính xác trong thế giới đa dạng và phức tạp của kinh doanh.

4.2.2 Phân phối giá đóng cửa



Hình 4 : Hình ảnh phân phối của giá đóng cửa của cổ phiếu giai đoạn 2018-2023

Biểu đồ phân phối giá đóng cửa của cổ phiếu Apple tạo ra một bức tranh động và sinh động về biến động của thị trường. Trục x, biểu diễn giá đóng cửa, đưa ta đến những con số chính xác, trong khi trục y thể hiện mật độ của giá đóng cửa, giúp chúng ta nhận diện sự tập trung của các mức giá.

Đường màu xanh nổi bật trên biểu đồ là biểu hiện của phân phối giá, chia thành nhiều vùng biến động khác nhau. Trong sự phức tạp này, có thể dễ dàng nhận ra hai đỉnh chính, một tại giá trị xấp xỉ 50 và một khác tại giá trị xấp xỉ 150. Điều này đề xuất rằng giá đóng cửa của cổ phiếu Apple thường có xu hướng tập trung quanh những mức giá này, tạo ra một hình ảnh về mức độ biến động và các điểm dừng phổ biến trên thị trường.

Nhìn vào biểu đồ phân phối, Những người đầu tư quan tâm có thể nhận ra mức giá chủ chốt và cơ hội giao dịch tiềm năng, đồng thời cung cấp khả năng đánh giá đúng sự phổ biến giá cụ thể. Đã làm nổi bật tính động và tích cực của thị trường cổ phiếu Apple, mang lại thông tin vô cùng quan trọng khi muốn đưa ra ý định đầu tư.

4.2.3 Closing Price: Range Slider and Selectors



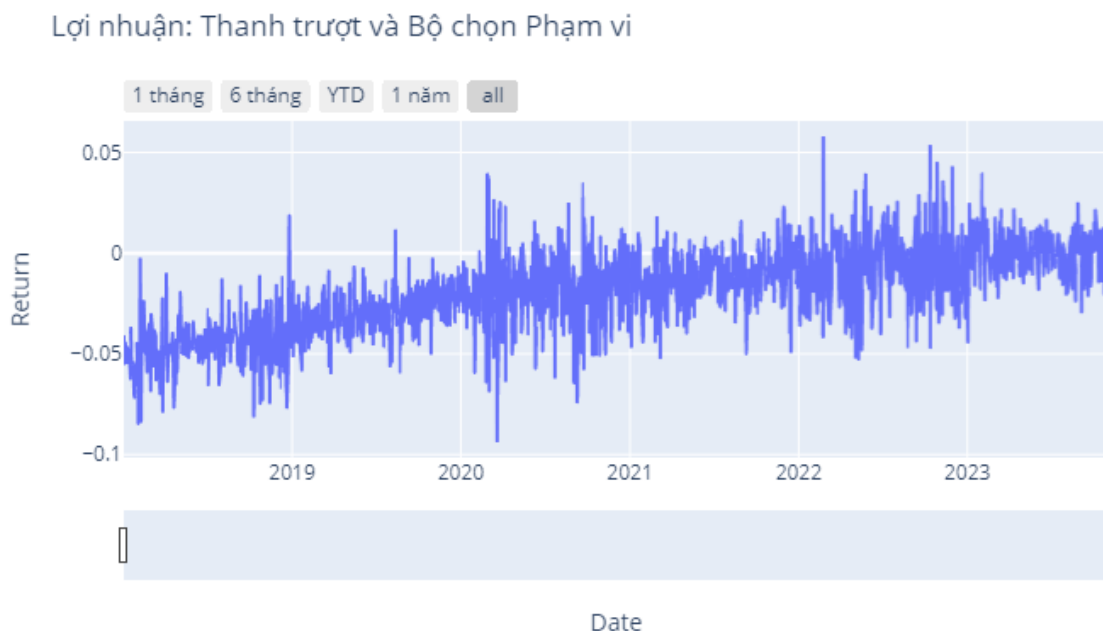
Hình 5 : Closing Price: Range Slider and Selectors

Biểu đồ trên là biểu đồ giá đóng và thanh trượt của cổ phiếu Apple, nơi mà thông tin về thời điểm và giá trị đóng được hiển thị rõ ràng là ở đây. Trục x biểu diễn thời gian, trong khi trục y thể hiện giá đóng. Đường màu xanh lam là hình ảnh chân thực của sự biến động giá, mang lại cái nhìn chi tiết vào sự diễn biến của thị trường.

Thanh trượt phạm vi dưới cùng là công cụ quan trọng, cho phép chọn thời gian cụ thể để xem xét. Để giúp tập trung vào các giai đoạn quan trọng hoặc thị trường có biến động lớn. Các tùy chọn nằm ở đỉnh cung cấp sự linh hoạt hơn, giúp bạn chọn thời gian cụ thể như 1 month, 6 month, YTD, 1 years, tất cả. Điều này giúp nhanh chóng ra xu hướng và biến động trong những thời gian khác nhau.

Người xem có thể nhìn nhận theo dõi và hiểu rõ hơn về các biến động giá trị dự kiến trên thị trường trong tương lai của Apple theo thời gian, đồng thời nắm bắt được các thời kỳ quan trọng và cơ hội giao dịch.

4.2.4 Lợi nhuận, thanh trượt và bộ chọn phạm vi



Hình 6 : Lợi nhuận, thanh trượt và bộ chọn phạm vi giai đoạn 2018-2023

Biểu đồ trên là biểu đồ lợi nhuận của cổ phiếu Apple, nơi thời gian và lợi nhuận được minh họa một cách rõ ràng. Trục x từ 2018 đến 2023, trục y biểu diễn lợi nhuận của cổ phiếu. Đường màu xanh lam là hình ảnh chân thực của sự biến động trong lợi nhuận, mang lại cái nhìn chi tiết về hiệu suất của cổ phiếu.

Thanh trượt phạm vi ở dưới cùng cũng có tính năng tương tự như phần ở trên đã được mô tả nên phần này không cần nhắc lại bạn có thể xem lại bên trên.

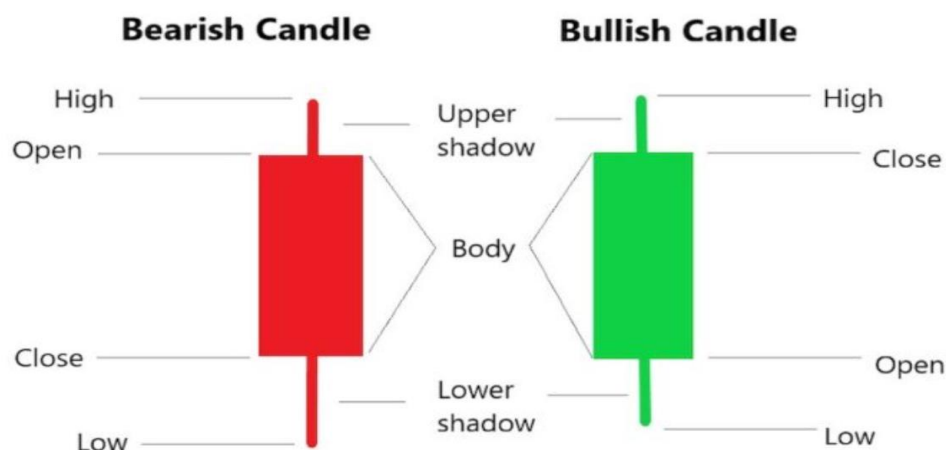
Hình ảnh giúp theo dõi và hiểu rõ hơn về diễn biến lợi nhuận của cổ phiếu Apple theo thời gian, đồng thời nắm bắt được các thời kỳ quan trọng và cơ hội đầu tư.

4.2.5 Biểu đồ nến

Mọi chuyên gia tài chính đều đồng thuận rằng biểu đồ nến là một thứ cực kỳ mạnh mẽ để đưa ra các diễn biến cổ phiếu. Đơn giản, hiệu quả, nó giúp ta dễ dàng nhận biết những thông tin chính về giá trị mà nó mang lại trong khoảng thời gian cụ thể.

Mỗi thanh thể hiện được lúc mở, cao, thấp, và đóng cửa của các giá trị trong những thời gian nhất định. Nến màu xanh biểu thị giai đoạn tăng giá, ở đây cho biết đóng cửa sẽ cao hơn mở. Ngược lại, nến đỏ đại diện cho giai đoạn giảm giá, khi giá mở thấp hơn đóng

Không chỉ là biểu đồ minh họa, Thậm chí đây là nguồn thông tin linh hoạt. Nhìn vào mô hình của nến, Các nhà đầu tư dễ dàng xem xét sự biến động, xu hướng, và áp lực mua/bán trên thị trường. Điều này sẽ giúp có được những quyết định hết sức sáng suốt và chính xác dựa trên các phương pháp kỹ thuật rất tinh tế.



Hình 7 : Mô tả sơ bộ về biểu đồ nến



Hình 8 : Biểu đồ nến thể hiện giá đóng cửa của cổ phiếu giai đoạn 2018-2023

Trên biểu đồ này, chúng ta chứng kiến hai đường biểu diễn sự biến động thị trường Apple. Đầu tiên, đường màu xanh đó là giai đoạn tăng giá, nơi mà giá đóng sẽ cao hơn giá mở. Điều này đồng nghĩa với sự tăng giá liên tục và ổn định từ năm 2018 đến 2023, một biểu hiện tích cực về hiệu suất cổ phiếu.

Ngược lại, đường màu đỏ biểu diễn giai đoạn giảm giá, trong đó giá đóng thấp hơn giá mở. Mặc dù thấy được sự giảm nhẹ từ 2018 cho tới 2023, nó không dẫn đến biến động đột ngột hoặc sự biến động lớn.

Tổng quan về biểu đồ cho ta nhận thức về sự liên tục trong sự gia tăng của giá cổ phiếu, đặc biệt là qua giai đoạn màu xanh lá cây. Hiện tại, ổn định với sự phát triển cực kỳ mạnh mẽ và liên tục của giá cổ phiếu tạo ra một dấu hiệu tích cực, mang lại hỗ trợ quan trọng cho nhà đầu tư và những người xem có thể theo dõi thị trường khi họ thực hiện quyết định về thông tin đầu tư và hiệu suất

4.2.6 Shifting and lags



Hình 9 : Giá đóng cửa hàng ngày và được dịch chuyển trong giai đoạn 2018-2023

Trên biểu đồ này, chúng ta có cơ hội nhìn được sự phát triển của cổ phiếu Apple qua thời gian thông qua hai đường biểu diễn xu hướng. Đầu tiên, đường màu xanh dương, được đặt tên là 'Close_M', thể hiện giá đóng hàng ngày. Bắt đầu từ mức khoảng 50 vào năm 2018, đường này liên tục được tăng lên và đã đạt đến mức khoảng 150 vào năm 2023. Với sự gia tăng này chính là bằng chứng cho một xu hướng tích cực trong giá cổ phiếu trong khoảng thời gian nói trên.

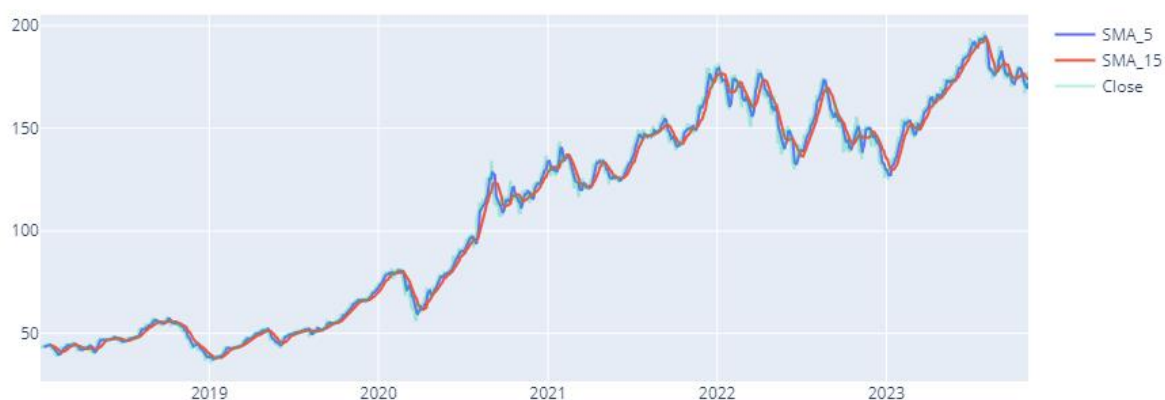
Mặt khác, đường màu đỏ, gọi là 'Lag_Close_M', biểu diễn giá đóng cửa được dịch chuyển lên với lags là 10. Từ mức khoảng 100 vào năm 2018, đường này tăng lên khoảng 200 vào năm 2023. Sự dịch chuyển này giúp ta nhìn rõ mối quan hệ giữa giá trị hiện tại với 10 ngày trước được thông qua bằng đường màu đỏ, không chỉ thể hiện sự tăng ổn định mà còn cung cấp thông tin quan trọng về xu hướng dài hạn của cổ phiếu.

So sánh giữa 'Close_M' và 'Lag_Close_M' cho phép phân tích sự biến đổi của giá trong quá khứ và Hỗ trợ trong việc dự đoán triển vọng tương lai khả năng đầu tư khác của Apple. Cung cấp cái nhìn chi tiết về thời gian của dữ liệu, giúp chúng ta sẽ hiểu rõ về biến động từ thị trường.

4.3 Technical Indicators

Technical Indicators là những công cụ và phương pháp sẽ áp dụng cho phân tích chính xác cho đo lường kèm luôn cả đánh giá hiệu suất của một cổ phiếu, thị trường, hoặc một tài sản hay là tài chính khác. Các chỉ số sẽ hỗ trợ những nhà đầu tư hình dung được tình hình của thị trường, từ đó hỗ trợ trong việc quyết định để đầu tư. Để được hiểu rõ hơn các chỉ số này thì việc tiến hành tính toán, trực quan hóa để hiểu một cách chân thực hơn từ các chỉ số.

4.3.1 Simple Moving Average



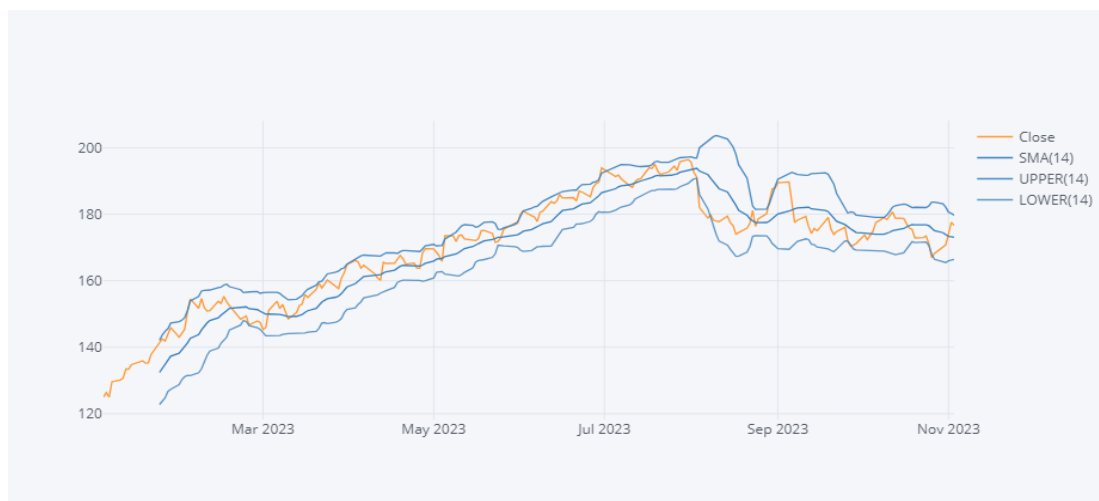
Hình 10: Biểu đồ biểu diễn so sánh giữa giá trị đóng cửa với SMA5 và SMA15

Biểu đồ giá cổ phiếu của Apple từ năm 2018 đến 2023 trình bày một hình ảnh vô cùng rõ ràng chi tiết về phát triển thông qua thị trường. Đường giá đóng cửa (màu xanh nhạt) thể hiện xu hướng tăng ổn định của giá cổ phiếu qua thời gian, đặc biệt là khi dùng so sánh với SMA.

SMA5 (màu xanh) là chỉ số ngắn hạn, cung cấp cái nhìn động của giá trong vòng 5 ngày. Sự tương quan giữa đường giá đóng cửa và SMA5 giúp phác họa xu hướng ngắn hạn, với sự đổi chiều nhanh chóng của giá trong khoảng thời gian ngắn.

Đường SMA15 (màu đỏ tập trung vào hướng biến của giá cổ phiếu có thể là giảm xuống hoặc tăng lên ở thời gian trung hạn, được tính trên 15 ngày. Nếu SMA15 vượt lên trên đường giá đóng cửa, điều này thường biểu thị một xu hướng tăng trung hạn của cổ phiếu.

Tổng thể, biểu đồ này chứng minh sự tích lũy và tăng giá ổn định của cổ phiếu Apple, với đường giá đóng cửa làm nổi bật xu hướng chính, trong khi SMA5 và SMA15 giúp hiểu rõ hơn về biến động ngắn và trung hạn của thị trường.



Hình 11 : So sánh giữa giá trị đóng cửa của SMA14 thông qua UPPER và LOWER14

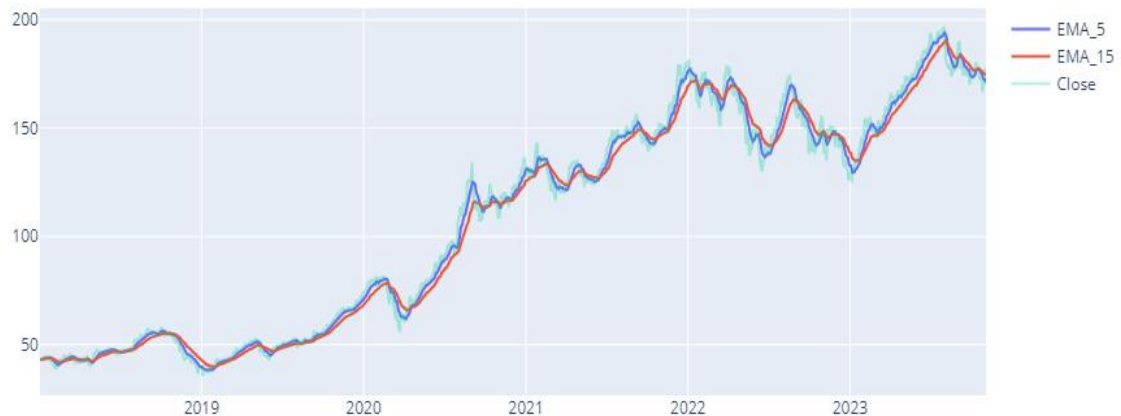
Trên biểu đồ này, chúng ta có ba góc quan trọng để làm nổi bật lên các khía cạnh vô cùng quan trọng của giá cổ phiếu. Đường màu cam, biểu diễn giá đóng cửa hàng ngày, là điểm xuất phát cho sự theo dõi của chúng ta. Chúng ta cũng có đường màu xanh dương, là SMA(14) hay trung bình động đơn giản trong 14 ngày, nó giúp chúng ta nhìn nhận xu hướng chung của giá trị trong những khoảng thời gian đó. Điều này giúp làm mịn đồ thị, làm nổi bật những biến động cơ bản của giá.

Hai đường màu xám và xanh, được gọi là LOWER(14) và UPPER(14) tương ứng, hình thành một khoảng tin cậy xung quanh giá. Nói cách khác, Có thể dựa vào những đường biến động này để có thể đánh giá và nhìn nhận giá có nằm trong phạm vi dự mà mình đưa vào dự đoán hay không. Với những người mà chỉ cần quan tâm đến đường xảy ra những biến động mạnh mẽ từ thị trường.

Việc giá trị đã vượt qua giới hạn bên trên (UPPER(14)) có thể là dấu hiệu của một xu hướng đang tăng đột ngột, trong lúc giảm xuống dưới giới hạn dưới (LOWER(14)) có thể là xu hướng đang bị giảm một cách đột ngột.

Qua cách này, biểu đồ cung cấp một tầm nhìn có thể nói là toàn diện và về sự biến động mạnh hoặc nhẹ của cổ phiếu, giúp đầu tư với người mua bán đưa ra quyết định thông tin và tiên đoán sự náo loạn mạnh mẽ một cách chính xác vô cùng đầy đủ.

4.3.2 Exponential Moving Average (EMA)



Hình 12 : Biểu đồ so sánh giữa Close với Exponential Moving Average (EMA) 5,15

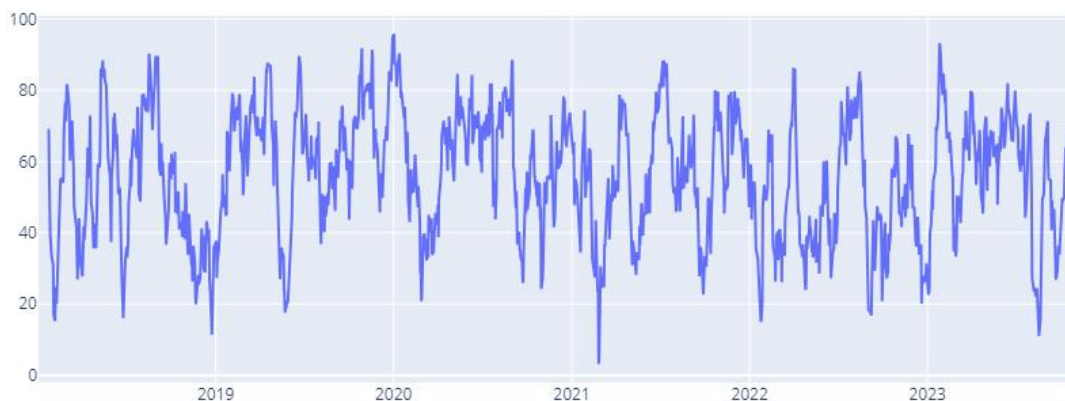
Trên biểu đồ này, chúng ta quan sát ba đường quan trọng đại diện cho những sự nhiễu loạn mạnh mẽ của các giá trị của cổ phiếu theo thời gian. Đầu tiên, đường màu xanh dương, gọi là 'EMA 5', là một chỉ báo nhạy bén theo dõi giá trung bình trong 5 ngày gần nhất. Việc trọng số cao đặt cho giá mới nhất giúp nó phản ánh nhanh chóng các biến động gần đây, tạo ra một hình ảnh động của giá.

Đường màu đỏ, 'EMA 15', mặc dù ít nhạy bén hơn, nhưng cũng cung cấp một tầm nhìn dài hạn hơn về xu hướng tăng hay giảm của giá. Trọng số cao hơn cho dữ liệu quá khứ giúp làm mịn đường đồ thị, giảm thiểu những biến động ngắn hạn.

Cuối cùng, đường màu xanh nhạt biểu diễn giá đóng cửa xảy ra hàng ngày, là chỉ số quan trọng cho những người xem diễn biến các sự thay đổi tổng thể các giá trị.

Tất cả ba đường đều thể hiện xu hướng tăng tổng thể của cổ phiếu theo thời gian. Về tầm quan trọng của nó điều này làm nên một sự mạnh mẽ, nổi bật và tích cực của xu hướng thị trường, Đưa ra kịp thời các kiến thức cực kỳ hữu ích cho nhà tài chính hoặc người mua bán, giao dịch trong việc có được các quyết định đúng đắn.

4.3.3 Relative Strength Index (RSI)

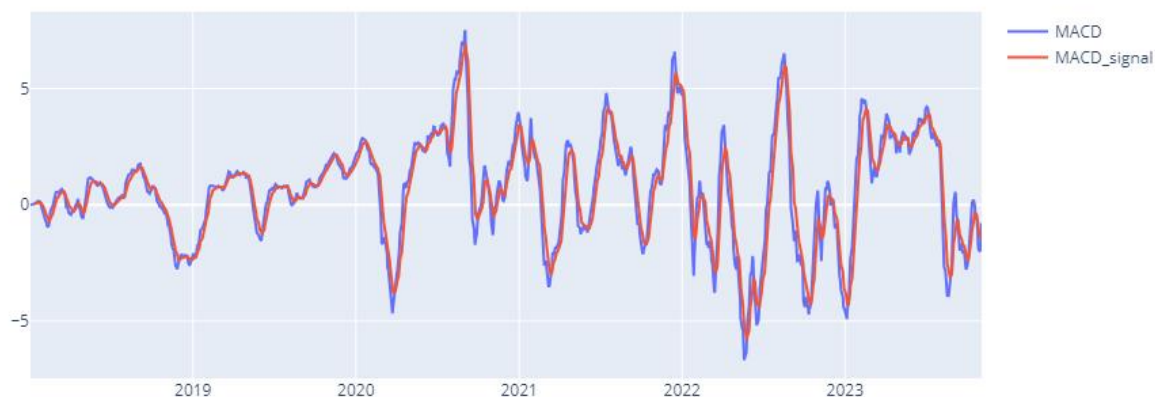


Hình 13 : Biểu đồ biểu diễn giá trị Relative Strength Index (RSI) giai đoạn 2018-2023

Biểu đồ chỉ số động lượng RSI rất quan trọng trong phân tích tài chính, đặc biệt trong giai đoạn từ năm 2018 đến 2023. RSI là một công cụ phân tích kỹ thuật thiết yếu, được sử dụng để đánh giá mức độ mua quá mức hoặc bán quá mức của cổ phiếu. Nó cung cấp cái nhìn sâu sắc về hiệu suất của cổ phiếu, giúp các nhà đầu tư nhận diện được thời điểm thích hợp để mua hoặc bán cổ phiếu. Chỉ số này đóng một vai trò không thể thiếu trong việc phân tích và quyết định đầu tư, giúp nhà đầu tư đưa ra những quyết định chính xác hơn dựa trên thông tin thị trường. Các đường được biểu diễn trên biểu đồ là biểu diễn thông tin của chỉ số RSI theo thời gian. Chỉ số RSI này dao động giữa hai giới hạn chính là 40 và 80, trong đó vùng từ 40 đến 80 thường được coi là khu vực an toàn, không thuộc về vùng quá mua (>70) hoặc quá bán (<30).

Khoảng từ năm 2018 đến 2023 không thấy xuất hiện dấu hiệu quá mua hoặc quá bán trên biểu đồ, đặc điểm này có thể được tính là sự ổn định trên thị trường cổ phiếu trong những thời gian được biểu diễn trên đó. Không có biến động nằm ngoài khu vực an toàn (40-80), cho thấy sự ổn định và cân bằng trong hành vi giá của cổ phiếu. Thông tin này quan trọng để hỗ trợ cho tuyệt vời đối với đầu tư và những người tham gia trong việc đưa ra các quyết định dựa trên tâm lý và tình hình thị trường.

4.3.4 MACD



Hình 14 : MACD và MACD Signal

Biểu đồ đưa ra thị trường chỉ thông qua chỉ số MACD , thực sự đây là một công cụ được biết đến với nhiệm vụ phân tích kỹ thuật một cách mạnh mẽ để xác định những xu thế thị trường và thường cung cấp cho người xem các tín hiệu mua/bán một rõ ràng và hợp lý.

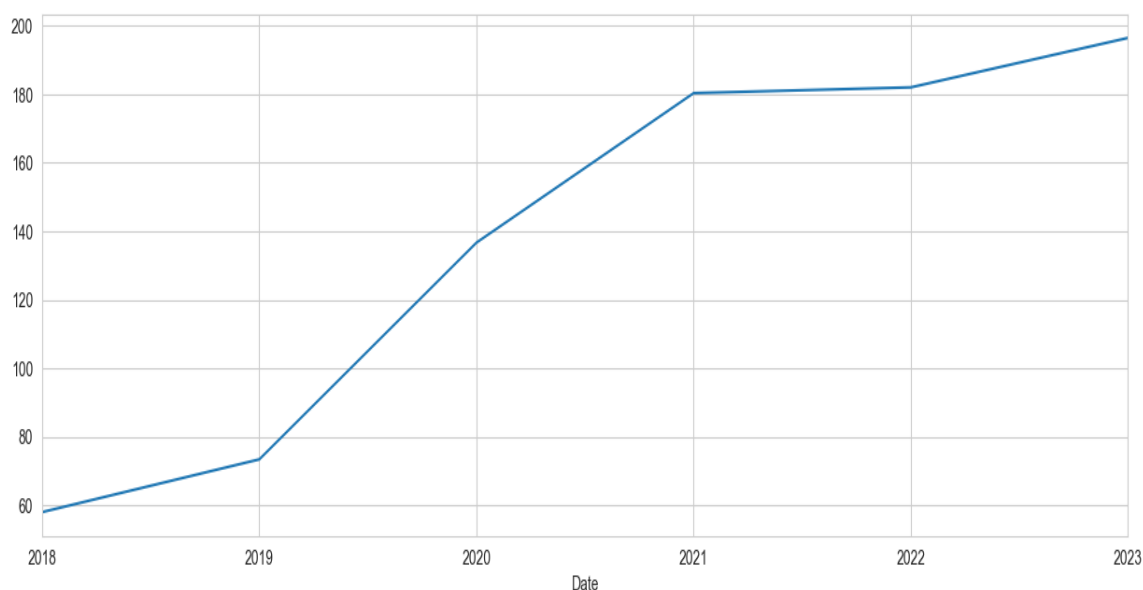
Đường màu xanh dương trên biểu đồ đại diện cho đường MACD, được tính bằng sự chênh lệch giữa đường EMA (Exponential Moving Average) 12 kỳ và EMA 26 kỳ. Đường này thường phản ánh được những sự biến động giữa hai đường của trung bình động và thể hiện sự hội tụ hoặc phân kỳ giữa chúng.

Đường màu đỏ trong biểu đồ là đường tín hiệu MACD, thường được tính thông qua cách lấy chỉ số của EMA 9 kỳ từ đường MACD. Đường này có vai trò là một tín hiệu mềm, giúp làm mịn đường MACD và tạo ra các điểm cắt quan trọng.

Tín hiệu cho phép mua được xuất hiện khi MACD chạy qua đường tín hiệu từ phía dưới lên trên, là chỉ báo cho thấy MACD đang chuyển lên trên đường chỉ báo tín hiệu. Ngược lại, tín hiệu được bán xuất hiện khi MACD cắt qua đường tín hiệu từ phía trên xuống.

MACD cung cấp thông tin chi tiết về sự dao động của giá cả, mở ra cơ hội cho các nhà đầu tư và trader trong việc đưa ra các quyết định thông minh dựa trên xu hướng diễn biến của thị trường tài chính.

4.4 Time Series Data Analysis – Resampling



Hình 15 : Time Series Data Analysis – Resampling giai đoạn 2018 - 2023

Biểu đồ mô tả được sự tăng trưởng của giá trị đóng cửa tối đa của cổ phiếu từ năm 2018 đến 2023. Trên trục hoành, chúng ta có các năm từ 2018 đến 2023, mỗi năm được đánh dấu để vẽ ra một bức tranh biến động qua những thời gian rõ ràng.

Trục tung, mặc dù không có nhãn, được chia thành các đơn vị từ 0 đến 200, với mỗi đơn vị có khoảng cách là 20. Điều này tạo ra cho chúng ta một tầm nhìn cực chi tiết về biên độ giao động của giá trị đóng cửa tối đa, đồng thời giúp theo dõi được các biến động mạnh mẽ của các giá trị tăng theo từng thời gian.

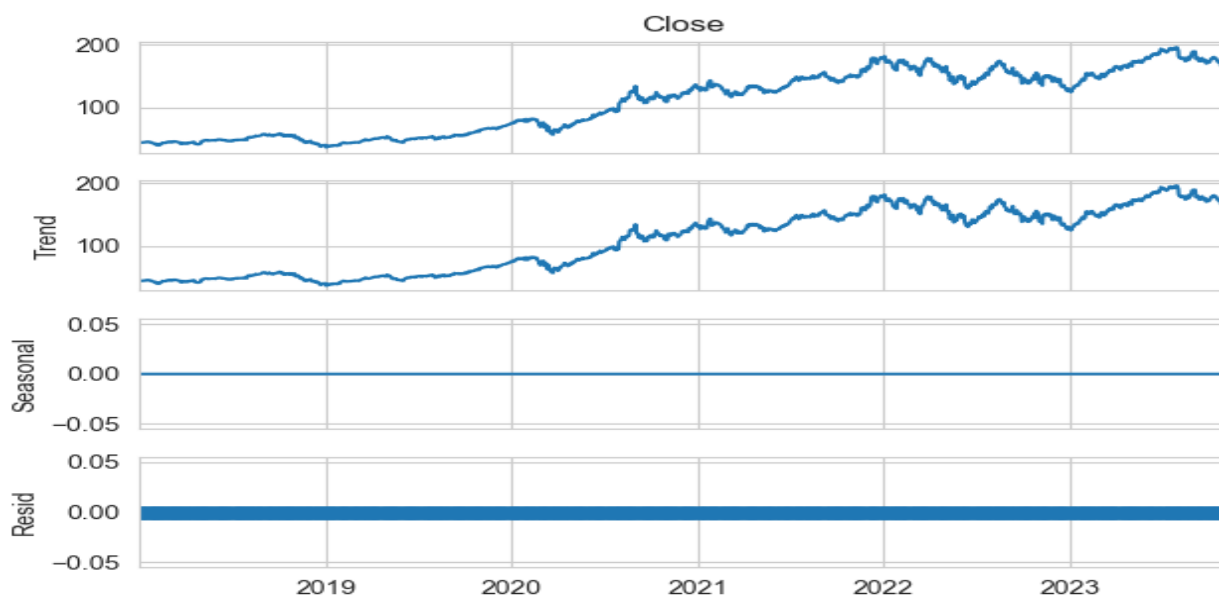
Từ việc nhìn nhận về nó ta nhận biết rằng giá trị đóng cửa tối đa bắt đầu từ mức khoảng 60 vào năm 2018 và trải qua một quá trình tăng dần đều đặn, đạt đến khoảng 180 vào năm 2023. Sự di chuyển liên tục của đường màu xanh lam thể hiện sự ổn định và tăng trưởng của cổ phiếu qua các giai đoạn.

Điều này không chỉ giúp nhìn thấy sự phát triển của giá trị đóng cửa tối đa mà còn tạo ra một cái nhìn xu hướng các biến động của cổ phiếu trong thời gian quan sát

4.4.1 Seasonal Factors

Đây là một yếu tố hết sức quan trọng trong việc phân tích dữ liệu và tiến hành dự báo, nhất là trong lĩnh vực kinh doanh và thị trường chứng khoán. Nó thường được áp dụng để hiểu rõ hơn và dự đoán những biến động dữ liệu theo mùa, như sự biến đổi theo mùa trong các ngành như nông nghiệp, bán lẻ, du lịch và nhiều lĩnh vực khác.

4.4.1.1 Tầng suất hằng ngày

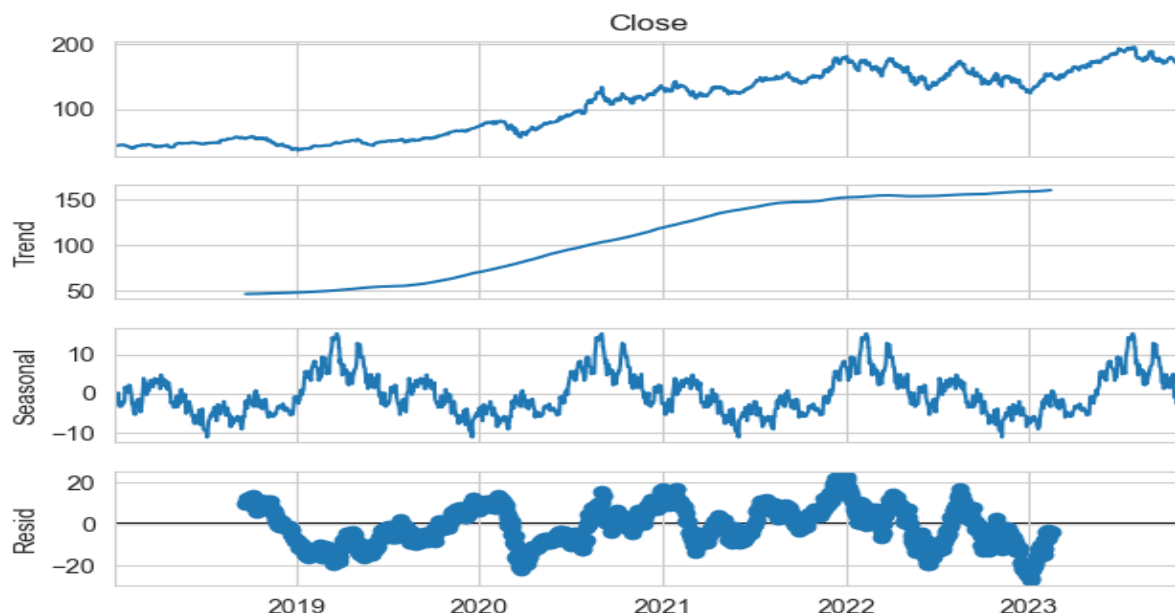


Hình 16: Biểu đồ phân tích mùa vụ (hằng ngày)

Biểu đồ dữ liệu chuỗi thời gian mang đến cái nhìn toàn diện về dữ liệu theo dõi hàng ngày. Mục 'Đóng cửa' phản ánh sự thay đổi hàng ngày trong giá cổ phiếu hoặc chỉ số kinh tế, hiện ra một khuynh hướng tăng dần qua các năm, mặc dù có những dao động ngắn hạn. Phần Xu hướng làm nổi bật khuynh hướng lâu dài của dữ liệu, loại bỏ tác động của những biến động ngắn hạn và yếu tố theo mùa. Biểu đồ này cho thấy một xu hướng tăng nhất quán theo thời gian, với một vài thời kỳ tăng giảm nhẹ. Mục Mùa vụ cho thấy sự dao động theo mùa, tuy nhiên không có sự biến động nổi bật. Điều này có thể do yếu tố mùa vụ không quan trọng hoặc dữ liệu đã được điều chỉnh để khử yếu tố này. Phần Dư thừa thể hiện phần còn lại sau khi loại bỏ xu hướng và yếu tố mùa vụ. Sự ổn định của dư thừa và giữ giá trị gần không cho thấy mô hình phân tích chuỗi thời gian hiệu quả trong việc lý giải dữ liệu, bỏ qua ít thông tin quan trọng và không chừa lại nhiều phần không giải thích được.

Tóm lại, biểu đồ này cung cấp cái nhìn toàn cảnh về khuynh hướng lâu dài, ảnh hưởng mùa vụ (nếu có), và độ ổn định của phần dư, làm rõ khả năng của mô hình phân tích chuỗi thời gian trong việc nắm bắt những yếu tố chính từ dữ liệu.

4.4.1.2 Tàng suất 365 ngày



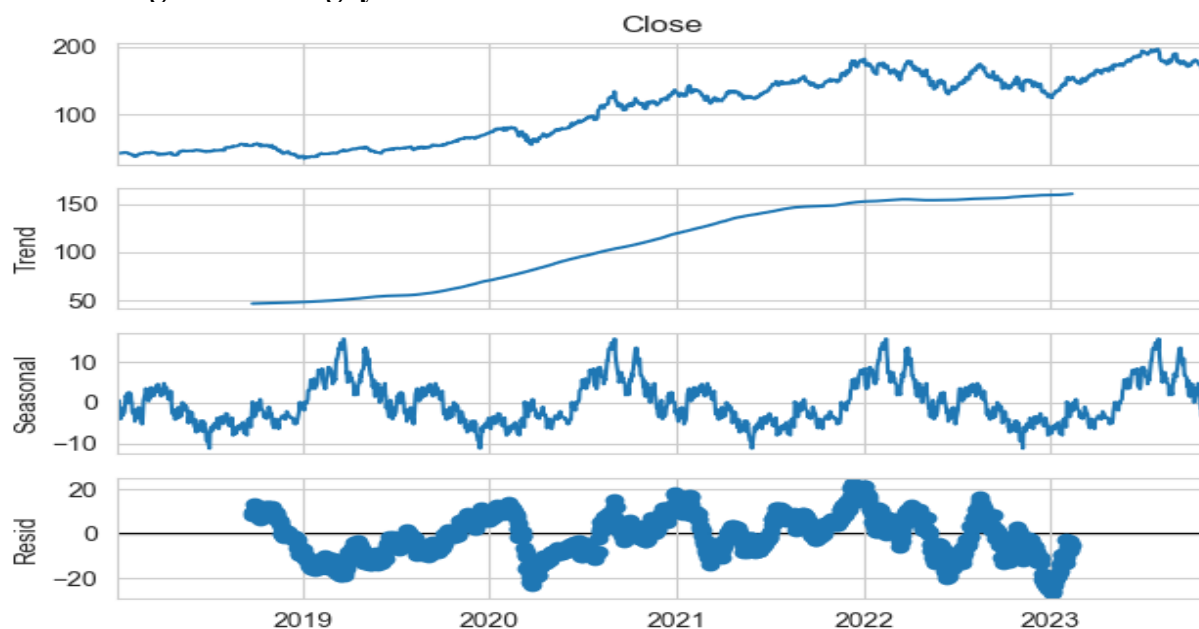
Hình 17 :Biểu đồ phân tích mùa vụ (365 ngày)

Đây là một phân tích tổng quan về dữ liệu được ghi nhận hàng ngày, với chu kỳ mùa vụ là 365 ngày. Các thành phần chính của biểu đồ đã được phân tích chi tiết: Close biểu diễn giá trị thực tế 2019 đến 2023. Xu hướng tăng nhẹ theo thời gian là rõ ràng trong dữ liệu thô này. Đường xu hướng chỉ ra sự tăng trưởng dài hạn của dữ liệu, loại bỏ biến động ngắn hạn và ảnh hưởng mùa vụ. Có xu hướng tăng qua các năm, tăng từ năm 2019 và đặc biệt đáng chú ý đến năm 2023. Biểu đồ mùa vụ cho thấy biến động định kỳ trong dữ liệu, có thể phản ánh các yếu tố mùa như quy luật mua sắm hoặc yếu tố thời tiết. Mô hình định kỳ hàng năm là rõ ràng, với các đỉnh và đáy xảy ra vào cùng thời điểm mỗi năm. Resid phần dư thể hiện sự biến động ngẫu nhiên hoặc không giải thích được sau khi loại bỏ xu hướng và mùa vụ. Điều này có thể là do các yếu tố bên ngoài hoặc biến động ngẫu nhiên trong dữ liệu.

Tóm lại, biểu đồ cung cấp cái nhìn sự thay đổi của dữ liệu với thời gian, bao gồm cả xu hướng dài hạn, biến động theo mùa và những yếu tố ngẫu nhiên. Thông tin này giúp

hiều rõ hơn về ảnh hưởng của những yếu tố khác nhau đối với chuỗi thời gian và hỗ trợ cho quyết định dựa trên thông tin này.

4.4.1.3 Tần suất 366 ngày



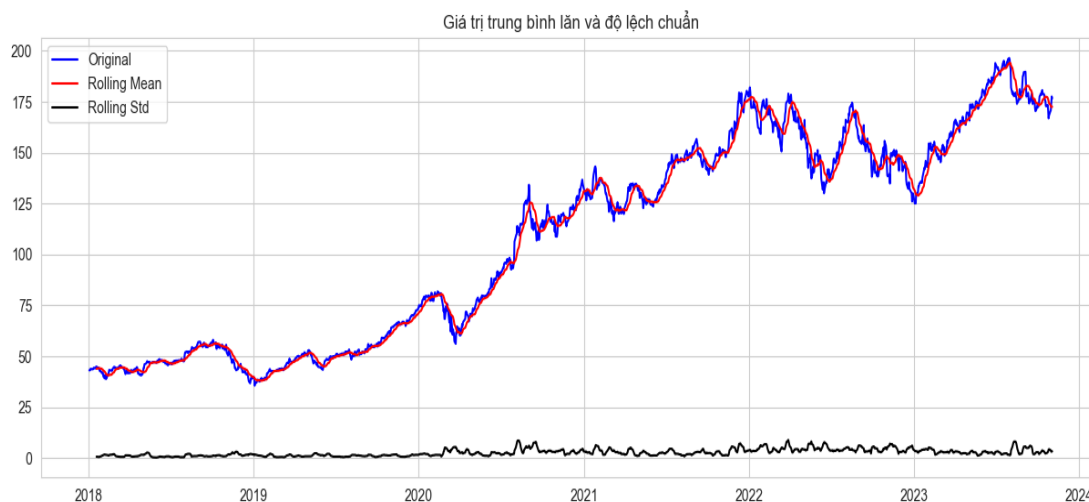
Hình 18 :Biểu đồ phân tích mùa vụ (366 ngày)

Tiếp tục phân tích dữ liệu hàng ngày, nhưng lần này với một tần suất mùa vụ là 366 ngày, ngụ ý rằng dữ liệu có thể bao gồm một năm nhuận. Close biểu thị giá trị hàng ngày của chuỗi thời gian từ 2019 đến 2023, có thể là giá cổ phiếu, chỉ số kinh tế hoặc dữ liệu tài chính tương tự. Xu hướng tăng dần qua thời gian là rõ ràng, kèm theo một số biến động ngắn hạn. Biểu đồ xu hướng thể hiện sự tăng trưởng dài hạn của dữ liệu. Đường cong mượt mà hơn so với dữ liệu thô, chỉ ra rằng, sau khi loại bỏ các yếu tố mùa vụ và ngẫu nhiên, xu hướng chung vẫn là tăng. Seasonal này thể hiện biến động mùa vụ với chu kỳ, mức độ biến động lớn hơn so với tần suất 365 ngày, có thể do bao gồm cả ngày nhuận. Biến động mùa vụ rõ ràng và chu kỳ, có thể được tạo ra bởi các yếu tố kinh tế, thời tiết hoặc các yếu tố khác có ảnh hưởng theo mùa. Resid phần dư sau khi loại bỏ xu hướng và yếu tố mùa vụ. Phần dư này biến động mạnh và không đều, cho thấy sự ảnh hưởng của những yếu tố khác không được mô hình hóa trong hai thành phần trước đó đối với dữ liệu.

Tổng cộng, việc phân tích với tần suất 366 ngày có thể chính xác hóa độ chi tiết của các yếu tố mùa vụ trong chuỗi thời gian, đặc biệt quan trọng đối với các chuỗi thời gian dài

kéo dài nhiều năm. Phần dư mang lại thông tin về sự biến động ngẫu nhiên hoặc các yếu tố không được mô hình hóa, từ đó hỗ trợ trong việc cải thiện mô hình phân tích trong tương lai.

4.4.2 Stationary Test / ADF Test



Hình 19 : Mô tả trung bình lăn với độ lệch chuẩn của dữ liệu giai đoạn 2018-2023

Thông qua biểu đồ được vẽ bên trên , Có thể dễ dàng nhận ra rằng giá trị trung bình hay là độ lệch chuẩn của dữ liệu chuỗi thời gian luôn có xu hướng tăng lên hoặc giảm nhẹ bằng việc thông qua điều này ta nhận ra rằng chuỗi của chúng ta là không có tính chất dừng. Khi xem xét giá trị p của giá trị kiểm định Dickey-Fuller, Giá trị của p lớn hơn 0.05 đây là ngưỡng thông thường để xác định tính dừng của chuỗi. Chính vì vậy, chúng ta sẽ không có đủ các bằng chứng để tiến hành bác bỏ giả thuyết null ,Nói chung có thể đưa ra kết luận rằng các thời gian của chúng ta không có tính chất là dừng.

Hơn nữa, giá trị của thống kê thông qua các giá trị được kiểm định có giá trị lớn hơn các giá trị quan trọng, làm cho chúng ta càng không thể bác bỏ giả thuyết null. Tất cả những điều này hỗ trợ quan điểm rằng dữ liệu không có tính chất dừng, Cần thực hiện các phương pháp hợp lý để chuyển đổi hoặc làm cho chuỗi dừng để sử dụng các mô hình thích hợp.

4.5 Differencing

4.5.1 First Difference

Thể hiện sự biến động giữa giá đóng cửa tại các thời điểm liên tiếp, giúp loại bỏ xu hướng và tính thời vụ.

Bảng 2 : First Difference

Test Statistics	-39.721215
p-value	0.000000
No. of lags used	0.000000
Number of observations used	1469.000000
dtype	float64

Kết quả này đưa ra các thông tin quan trọng để đánh giá tính dừng của chuỗi thời gian. Trong trường hợp này, giá trị p (p-value) rất nhỏ (bằng 0), cho thấy chúng ta có thể bác bỏ giả thuyết về tính dừng. Điều này có nghĩa là sau khi thực hiện differencing, chuỗi thời gian không chứa xu hướng hay tính thời vụ đáng kể.

4.5.2 Seasonal Difference

Đây là sự chênh lệch giá đóng và mở với một khoảng cách là 12 tháng.

Bảng 3 : Seasonal Difference

Test Statistics	-5.816200e+00
p-value	4.286788e-07
No. of lags used	2.400000e+01
Number of observations used	1.434000e+03
dtype	float64

Kết quả của kiểm định Augmented Dickey-Fuller (ADF) trên chuỗi thời gian gốc là đáng chú ý. Giá trị thống kê kiểm định là -5.816200, một giá trị âm lớn mạnh mẽ cho thấy sự không ổn định của chuỗi thời gian, có khả năng được coi là định tính. Xác suất p (4.286788e-07) rất thấp, dưới ngưỡng 0.05, mạnh mẽ bác bỏ giả thuyết về tính dừng, làm tăng độ tin cậy vào việc chuỗi thời gian có khả năng định tính.

Số lượng độ trễ sử dụng trong mô hình là 24, đảm bảo rằng kiểm định được thực hiện với độ chính xác cao. Số lượng quan sát sử dụng là 1434, tổng số điểm dữ liệu có sẵn để thực hiện kiểm định.

Tổng cộng, kết quả mạnh mẽ này cung cấp đủ bằng chứng cho quan điểm rằng chuỗi thời gian không có tính dừng và có thể được xem xét là định tính. Điều này là quan trọng

khi chuẩn bị dữ liệu cho mô hình phân tích chuỗi thời gian, đảm bảo tính chính xác và độ tin cậy của các kết quả dự báo.

4.5.3 Seasonal Daily Difference

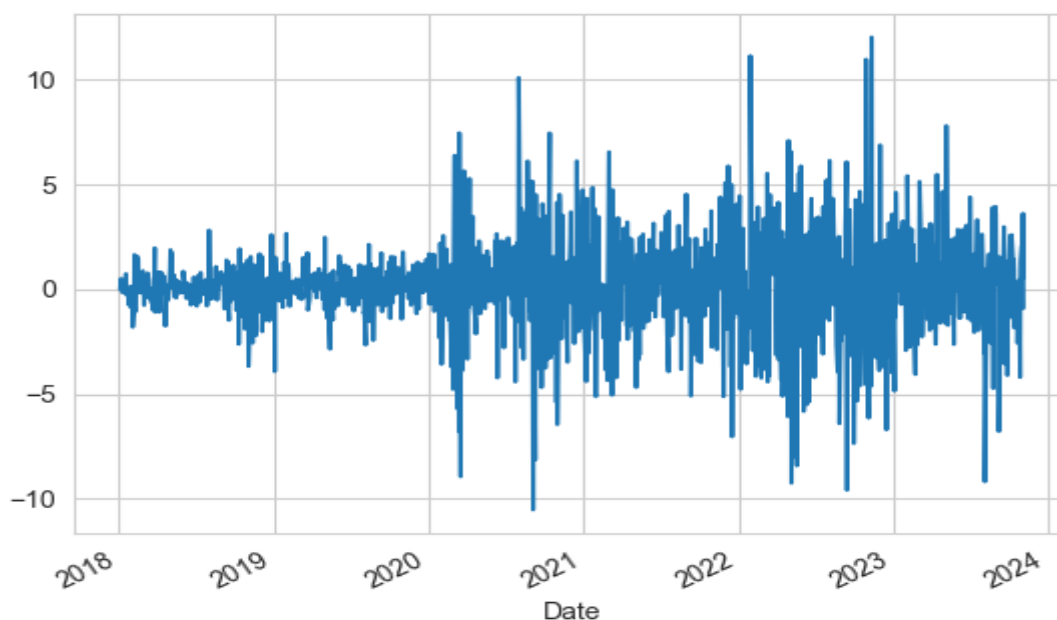
Bảng 4 : Seasonal Daily Difference

Test Statistics	-1.299284e+01
p-value	2.789172e-24
No. of lags used	2.300000e+01
Number of observations used	1.434000e+03
dtype	float64

Kết quả của kiểm định Augmented Dickey-Fuller (ADF) trên chuỗi thời gian hiện tại là rất đặc biệt. Giá trị thống kê kiểm định là -1.299284e+01, một giá trị âm lớn mạnh mẽ cho thấy sự không ổn định của chuỗi thời gian và có khả năng xem xét chuỗi này là định tính. Giá trị p (2.789172e-24) rất thấp, dưới ngưỡng 0.05, mạnh mẽ bác bỏ giả thuyết về tính dừng, cung cấp đủ bằng chứng cho tính định tính của chuỗi thời gian.

Số lượng độ trễ sử dụng trong mô hình là 2.3, cho thấy sự cân nhắc cẩn thận về số lượng quan trọng để đảm bảo kiểm định được thực hiện với độ chính xác cao. Số lượng quan sát sử dụng là 1.434, tổng số điểm dữ liệu có sẵn để thực hiện kiểm định.

Kết quả này không chỉ mạnh mẽ ủng hộ quan điểm rằng chuỗi thời gian không có tính dừng, mà còn cung cấp độ tin cậy cao trong việc xác định đặc điểm định tính của chuỗi. Điều này là quan trọng để chuẩn bị dữ liệu cho mô hình phân tích chuỗi thời gian, đảm bảo kết quả dự báo chính xác và đáng tin cậy.



Hình 20 : Biểu đồ Differencing

Biểu đồ này cung cấp cái nhìn về sự biến động giá đóng cửa hàng ngày của cổ phiếu sau khi thực hiện phép differencing, với dữ liệu kéo dài đến tháng 11 năm 2023. Kỹ thuật Differencing giúp làm nổi bật sự thay đổi giá giữa các ngày giao dịch liên tiếp, được biểu thị qua các giá trị trên trục tung, dao động từ -10 đến +10. Trong tháng 11 năm 2023, biểu đồ cho thấy một mức độ biến động ổn định mà không có bất kỳ sự kiện bất thường nào được phản ánh. Các điểm dữ liệu chủ yếu tập trung gần giá trị không, cho thấy các biến động ngày càng giảm và không có các đỉnh hoặc thung lũng đáng kể. Sự ổn định này có thể cho thấy rằng không có sự kiện lớn nào ảnh hưởng đến cổ phiếu trong khoảng thời gian này, hoặc nó phản ánh sự cân bằng của thị trường sau các biến động trước đó. Nhìn chung, mức độ biến động nhẹ trong tháng 11 năm 2023 có thể là dấu hiệu của sự ổn định trong cảm nhận của nhà đầu tư đối với công ty phát hành cổ phiếu này.

4.6 Xây dựng mô hình LSTM

Quá trình này bắt đầu với việc tải dữ liệu cổ phiếu, được lưu trữ trong một Data Frame có chỉ mục thời gian (datetime index) và một cột giá đóng cửa ('Close'). Dữ liệu này được lọc để chỉ bao gồm các bản ghi từ sau năm 2021, sau đó chia thành hai bộ phận: dữ liệu huấn luyện (trước năm 2023) và dữ liệu kiểm tra (từ năm 2023 trở đi).

Tiếp theo, dữ liệu giá đóng cửa được chuẩn hóa sử dụng MinMaxScaler để đưa giá trị về trong khoảng từ 0 đến 1. Việc này giúp cải thiện khả năng học của mô hình LSTM bằng cách đảm bảo rằng tất cả các giá trị đều ở cùng một quy mô.

Để chuẩn bị dữ liệu cho mô hình LSTM, cần phải tạo ra các chuỗi thời gian. Điều này được thực hiện thông qua hàm `create_sequences`, chuyển đổi dữ liệu đã được chuẩn hóa thành các cặp X (đầu vào) và y (nhãn). Mỗi cặp này chứa thông tin về giá trị giá đóng cửa trong quá khứ (X) và giá trị tiếp theo cần dự đoán (y). Mô hình LSTM được xây dựng sử dụng TensorFlow và Keras. Mô hình này gồm ba lớp LSTM, mỗi lớp có 50 đơn vị, cùng một lớp Dense cuối cùng để dự đoán giá trị giá đóng cửa. LSTM là một dạng mạng nơ-ron tái phát (RNN), đặc biệt hiệu quả trong việc xử lý dữ liệu chuỗi thời gian như dữ liệu cổ phiếu.

Mô hình được biên dịch với hàm mất mát là `'mean_squared_error'` và thuật toán tối ưu hóa `'adam'`. Sau đó, nó được huấn luyện trên tập dữ liệu huấn luyện với 50 kỳ lặp và kích thước là 32. Quá trình huấn luyện được kiểm soát chặt chẽ để tránh vấn đề quá khớp, đảm bảo mô hình có khả năng tổng quát hóa tốt trên dữ liệu mới.

Sau khi huấn luyện, mô hình được sử dụng để dự đoán giá trị trên tập dữ liệu kiểm tra. Các dự đoán này sau đó được so sánh với giá trị thực tế bằng cách tính toán các chỉ số đánh giá như RMSE (Root Mean Squared Error), MSE (Mean Squared Error) và MAE (Mean Absolute Error), cung cấp một cái nhìn sâu sắc về hiệu suất của mô hình.

Cuối cùng, kết quả được trực quan hóa sử dụng Plotly, một thư viện mạnh mẽ cho phép tạo ra các biểu đồ tương tác. Biểu đồ này không chỉ hiển thị dữ liệu cổ phiếu toàn bộ và dữ liệu được dự đoán bởi mô hình, mà còn so sánh chúng với dữ liệu huấn luyện và kiểm tra. Điều này cho phép nhìn thấy rõ ràng sự khác biệt giữa giá trị thực tế và dự đoán, cũng như hiệu suất của mô hình trên cả hai phần dữ liệu.

Quá trình này từ đầu đến cuối cho thấy sức mạnh của học sâu và LSTM trong việc mô phỏng và dự đoán các xu hướng phức tạp trong dữ liệu thị trường chứng khoán. Đây là một ví dụ điển hình về cách ứng dụng trí tuệ nhân tạo có thể mang lại cái nhìn sâu sắc và giá trị trong lĩnh vực tài chính và đầu tư.

4.6.1 Xây dựng mô hình LSTM với 10 bước thời gian



Hình 21 : Hình ảnh dự đoán của LSTM với 10 bước thời gian

Biểu đồ này bao gồm một đồ thị chuỗi thời gian với các đường khác nhau đại diện cho các bộ dữ liệu khác nhau: Đường màu xanh dương đại diện cho bộ dữ liệu đầy đủ về giá cổ phiếu của Apple theo thời gian. Nó cho thấy một xu hướng tăng chung từ năm 2019 trở đi với một số biến động. Điều này phản ánh tính chất thường thấy của dữ liệu giá cổ phiếu, thường biểu hiện sự kết hợp giữa xu hướng chung và biến động ngắn hạn. Đường màu xanh lá cây minh họa phần dữ liệu được sử dụng để huấn luyện mô hình LSTM. Nó sát theo đường màu xanh dương cho đến cuối năm 2022, đánh dấu điểm chia giữa dữ liệu huấn luyện và dữ liệu kiểm tra. Dữ liệu huấn luyện là những gì mô hình LSTM sử dụng để học các mô hình cơ bản trong biến động giá cổ phiếu của Apple. Nối tiếp đường màu xanh lá, đường màu đen đại diện cho dữ liệu kiểm tra, đó là dữ liệu từ năm 2023 trở đi mà mô hình chưa từng thấy trong quá trình huấn luyện. Bộ này được sử dụng để đánh giá hiệu suất của mô hình, để xem nó có thể dự đoán giá cổ phiếu dựa trên những gì nó đã học từ dữ liệu huấn luyện như thế nào. Cuối cùng, đường màu đỏ đại diện cho giá trị dự đoán của mô hình LSTM cho giai đoạn dữ liệu kiểm tra. Với $\text{time_steps_lstm} = 10$, mô hình đang thực hiện dự đoán dựa trên các mô hình nó đã học từ 10 ngày gần nhất của giá cổ phiếu. Đường màu đỏ có vẻ bắt đầu từ đầu năm 2023 và cố gắng theo dõi những biến động của đường màu đen, thể hiện nỗ lực của mô hình trong việc dự báo giá trong tương lai.

Biểu đồ cho thấy mô hình LSTM, sau khi được huấn luyện trên dữ liệu giá cổ phiếu lịch sử cho đến hết năm 2022, có khả năng tạo ra các dự đoán theo xu hướng chung của dữ liệu kiểm tra khá tốt. Tuy nhiên, có những điểm mà giá trị dự đoán chênh lệch so với giá trị thực tế, chỉ ra những khu vực mà dự đoán của mô hình không hoàn hảo, điều này thường gặp trong dự đoán giá cổ phiếu do tính không chắc chắn và biến động của thị trường chứng khoán.

Nhìn chung, biểu đồ mô tả khả năng sử dụng mô hình LSTM cho việc dự báo chuỗi thời gian trong tài chính. Nó cho thấy cách dữ liệu quá khứ có thể được sử dụng để huấn luyện một mô hình có khả năng dự đoán các chuyển động tương lai, mặc dù độ chính xác của những dự đoán như vậy có thể thay đổi tùy thuộc vào nhiều yếu tố, bao gồm biến động thị trường, điều kiện kinh tế, và các thông số được chọn của chính mô hình, như số lượng bước thời gian được sử dụng trong LSTM

Ngoài ra việc đánh giá còn thông qua các chỉ số qua bảng như sau :

Bảng 5 : Thông số đánh giá mô hình qua 10 bước thời gian

RMSE for LSTM: 0.05023
MSE for LSTM: 0.00252
MAE for LSTM: 0.0411

Kết quả đánh giá hiệu suất cho mô hình LSTM trong dự đoán giá cổ phiếu của Apple mang đến một cái nhìn tích cực và đầy hứa hẹn. Các độ đo hiệu suất đã được tính toán và phản ánh mức độ chính xác và hiệu quả của mô hình trong việc ước lượng giá cổ phiếu trong tương lai. Trong đó, Root Mean Squared Error (RMSE), một chỉ số quan trọng đo lường độ lệch trung bình giữa giá trị dự đoán và giá trị thực tế, đạt giá trị thấp là 0.0502. Điều này chỉ ra rằng mô hình có khả năng dự đoán giá cổ phiếu với độ chính xác cao, và sự chênh lệch giữa giá trị dự đoán và thực tế là rất nhỏ. MSE cũng thu được giá trị thấp là 0.0025, một biểu hiện khác về sự chính xác và hiệu quả của mô hình trong việc dự đoán giá cổ phiếu. Giá trị thấp của MSE cho thấy mức độ chênh lệch giữa dự đoán và thực tế được giảm thiểu. Ngoài ra, MAE với giá trị là 0.0411 thể hiện mức độ chính xác trung bình của

dự đoán. Mô hình có khả năng ước lượng giá cổ phiếu với độ chính xác cao, đặc biệt là khi so sánh với giá trị thực tế.

Tổng thể, sự tích cực của các độ đo này cùng với đồ thị biểu diễn cho thấy mô hình LSTM làm việc hiệu quả trong việc dự đoán giá cổ phiếu của Apple. Các kết quả này là một bước tiến quan trọng trong việc áp dụng mô hình chuỗi thời gian trong lĩnh vực tài chính, và chúng có thể cung cấp cơ sở để tối ưu hóa mô hình và nâng cao khả năng dự đoán trong tương lai.

4.6.1.1 Dự đoán 10 ngày tiếp theo.

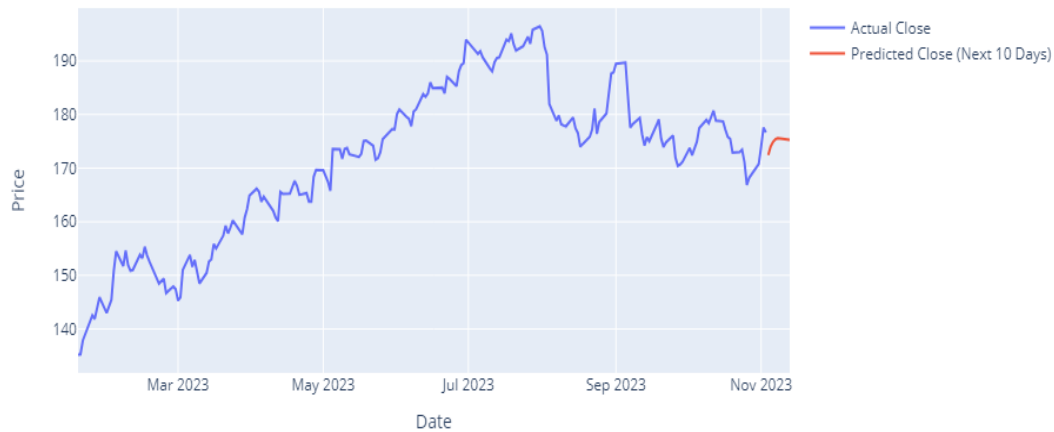
Giá đóng cửa được dự đoán cho 10 ngày tiếp theo như sau:

Bảng 6 : Giá đóng cửa dự đoán 10 ngày tiếp theo

2023-11-03: 172.4403
2023-11-04: 174.0241
2023-11-05: 174.8820
2023-11-06: 175.3798
2023-11-07: 175.5999
2023-11-08: 175.5446
2023-11-09: 175.4988
2023-11-10: 175.4629
2023-11-11: 175.3479
2023-11-12: 175.3011

Dựa trên những dự đoán về giá đóng cửa của cổ phiếu cho 10 ngày tiếp theo, xu hướng chung cho thấy sự ổn định và tăng trưởng nhẹ của giá cổ phiếu. Trong giai đoạn đầu, dự đoán cho thấy một mức giá khá ổn định, dao động từ 172.44 đến 175.60, cho thấy sự ổn định tại thị trường và khả năng duy trì giá ổn định trong khoảng này. Mặc dù có sự biến động nhỏ trong giá, nhưng nó duy trì ở mức cao, điều này có thể phản ánh sự tích lũy sức mạnh trong thị trường cổ phiếu.

Stock Price Prediction



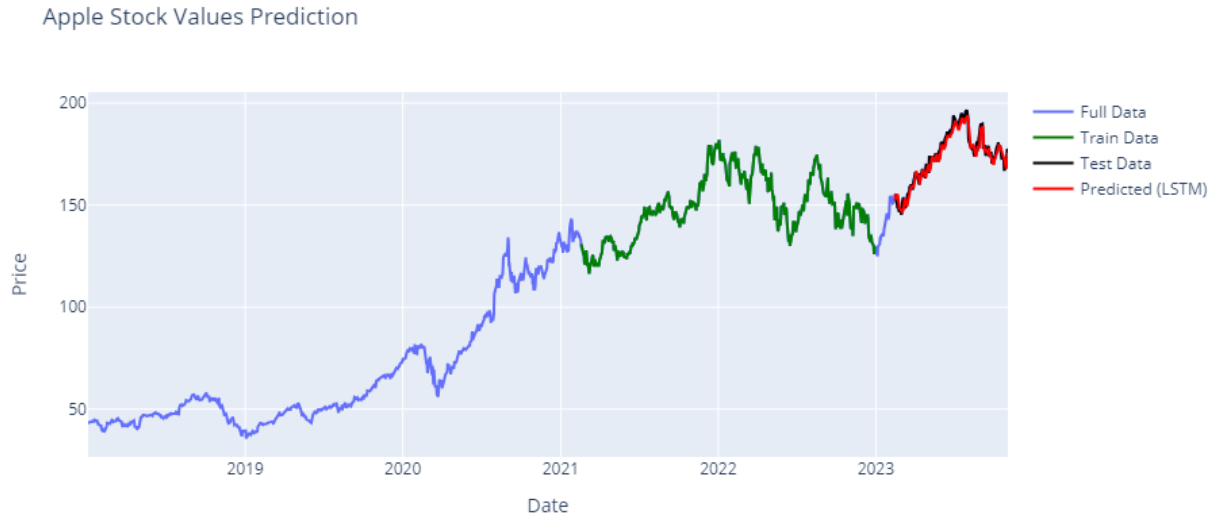
Hình 22 : Dự đoán giá đóng cửa trong vòng 10 ngày tới

Đường màu xanh đại diện cho giá đóng cửa thực tế của cổ phiếu, theo dõi diễn biến giá từ tháng Ba đến cuối năm 2023. Giá trị có vẻ biến động tăng dần từ khoảng 150 đến gần 190, sau đó giảm nhẹ về cuối biểu đồ, cho thấy sự biến động tự nhiên của thị trường. Đường màu đỏ thì đại diện cho giá đóng cửa dự đoán của cổ phiếu cho 10 ngày tiếp theo, được mô hình dự đoán tạo ra. Đường màu đỏ bắt đầu từ điểm cuối cùng của đường màu xanh dương và có xu hướng giảm. Điều này cho thấy mô hình dự đoán rằng giá cổ phiếu sẽ giảm trong ngắn hạn.

Nhìn chung, biểu đồ cho thấy mô hình dự đoán đang cố gắng mô phỏng xu hướng giảm gần đây của thị trường. Tuy nhiên, do tính không chắc chắn và biến động cao của thị trường cổ phiếu, việc dự đoán chính xác giá cổ phiếu trong tương lai là rất khó khăn. Những dự đoán như vậy có thể được sử dụng như một công cụ để hỗ trợ quyết định đầu tư, nhưng chúng không nên được dùng làm cơ sở duy nhất cho các quyết định đó. Độ lệch giữa dữ liệu thực tế và dự đoán cũng cung cấp thông tin về khả năng và giới hạn của mô hình dự đoán được sử dụng.

4.6.2 Xây dựng mô hình LSTM với 30 bước thời gian

LSTM với 30 bước thời gian cũng tương tự so với 10 chỉ khác là nó học từ chuỗi 30 thay vì 10 mà thôi



Hình 23 : Dự đoán LSTM với 30 bước thời gian

Hình ảnh cung cấp một cái nhìn trực quan về dữ liệu giá cổ phiếu Apple từ năm 2019 đến năm 2023 và bao gồm cả dự đoán dựa trên mô hình LSTM. Đường màu xanh dương là dữ liệu giá đóng cửa thực tế của cổ phiếu Apple, cho thấy xu hướng tăng giá từ năm 2019 đến năm 2021, sau đó có biến động mạnh mẽ hơn từ năm 2021 đến 2023. Phần tiếp theo của dữ liệu, được hiển thị bằng màu xanh lá, đã được sử dụng để "dạy" cho mô hình LSTM. Đây là dữ liệu mà mô hình đã học và đã cố gắng tìm ra các mô hình và xu hướng trong quá khứ.

Đường màu đen là phần dữ liệu mà mô hình chưa từng thấy trong quá trình huấn luyện và được sử dụng để kiểm tra xem mô hình có thể dự đoán chính xác đến mức nào. Trong biểu đồ, đường màu đen bắt đầu từ năm 2023, cho phép chúng ta đánh giá hiệu quả của mô hình khi được áp dụng vào dữ liệu thực tế mà nó không được huấn luyện trước. Cuối cùng, đường màu đỏ biểu diễn giá trị dự đoán của mô hình LSTM.

Với việc sử dụng 30 bước chuỗi thời gian, mô hình đã xem xét 30 ngày gần nhất của dữ liệu giá đóng cửa để dự đoán giá cho ngày tiếp theo. Đường đỏ cho thấy các giá trị dự đoán sau khi mô hình đã được huấn luyện, và chúng ta có thể thấy mô hình đã nắm bắt được phần nào xu hướng giảm giá vào cuối biểu đồ.

Độ chính xác được thể hiện thông qua bảng sau :

Bảng 7 : Thông số đánh giá mô hình qua 10 bước thời gian

RMSE for LSTM : 0.0423
MSE for LSTM : 0.0017
MAE for LSTM : 0.0334

Dựa vào những đánh giá hiệu suất mới nhất cho mô hình LSTM, chúng ta có thể rõ ràng nhận thấy những bước tiến quan trọng trong khả năng dự đoán giá cổ phiếu. Root Mean Squared Error (RMSE) hiện đang ở mức rất thấp với giá trị là 0.0424, điều này chứng tỏ mô hình có khả năng dự đoán với độ chính xác cao đối với các giá trị giá cổ phiếu. Mean Squared Error (MSE) tiếp tục giảm xuống 0.0018, là dấu hiệu rõ ràng của sự chính xác ngày càng tăng lên và đồng thời giảm thiểu độ lớn của sai số. Trong khi đó, Mean Absolute Error (MAE) đạt giá trị là 0.0335, thể hiện mức độ chính xác trung bình của mô hình trong việc dự đoán giá cổ phiếu. Những con số ấn tượng này là kết quả của sự nỗ lực và cải tiến liên tục trong quá trình huấn luyện mô hình LSTM. Sự giảm nhỏ của các độ đo đánh giá này không chỉ phản ánh tính hiệu quả của mô hình trên dữ liệu đã biết mà còn thể hiện sự tự tin trong khả năng dự đoán trên dữ liệu mới. Tuy mô hình đã thể hiện khả năng dự đoán tích cực trong giai đoạn kiểm tra, nhưng vẫn tồn tại những khoảng cách đáng kể giữa dữ liệu thực tế và dự đoán. Điều này làm nổi bật thách thức của việc dự đoán chính xác giá cổ phiếu, một thách thức đầy phức tạp. Đường dự đoán của mô hình tương đối phản ánh được xu hướng chung của thị trường, nhưng các biến động và yếu tố không dự đoán được vẫn tạo nên những độ chênh lệch nhất định.

Tổng thể, mặc dù đã có sự tiến bộ rõ ràng, nhưng việc kiểm tra và điều chỉnh mô hình là hết sức quan trọng để nâng cao khả năng dự đoán và đối mặt với những thách thức phức tạp của thị trường tài chính.

4.6.2.1 Dự đoán 30 ngày tiếp theo

Sau khi tiến hành dự đoán 30 ngày tiếp theo thì giá đóng được dự đoán được thể hiện thông qua các con số sau :

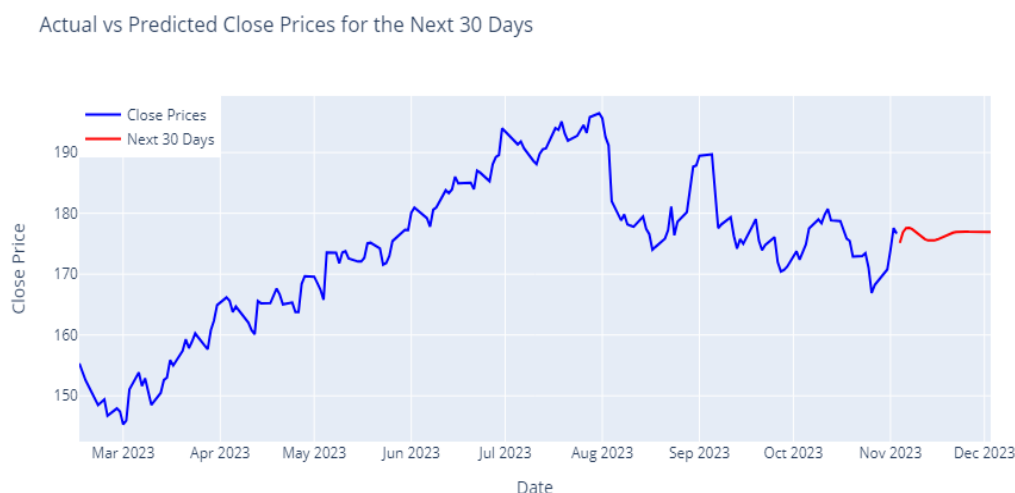
Bảng 8 : Giá đóng của dự đoán 30 ngày tiếp theo

2023-11-03: 175.6431	2023-11-16: 172.9875
2023-11-04: 176.7940	2023-11-17: 172.8435
2023-11-05: 176.5513	2023-11-18: 172.6574
2023-11-06: 175.7576	2023-11-19: 172.4459
2023-11-07: 174.8893	2023-11-20: 172.2234
2023-11-08: 174.1549	2023-11-21: 172.0040
2023-11-09: 173.6274	2023-11-22: 171.7967
2023-11-10: 173.3074	2023-11-23: 171.6063
2023-11-11: 173.1575	2023-11-24: 171.4355
2023-11-12: 173.1158	2023-11-25: 171.2814
2023-11-13: 173.1201	2023-11-26: 171.1407
2023-11-14: 173.1187	2023-11-27: 171.0114
2023-11-15: 173.0790	2023-11-28: 170.8913

Trong khoảng thời gian từ ngày 3/11 đến 2/12, dãy số dự đoán giá cổ phiếu mang đến một bức tranh tương đối tích cực. Dù có sự giảm dần nhẹ trong giá, nhưng mức giữ vững ở mức cao cho thấy sự ổn định và niềm tin từ phía nhà đầu tư. Bắt đầu từ mức cao nhất là 175.6431 vào ngày 3/11, giá cổ phiếu đã trải qua một sự giảm dần nhẹ, đạt đến mức thấp nhất là 170.4283 vào ngày 2/12. Sự biến động này không lớn, và thậm chí có thể thể hiện một đợt điều chỉnh tự nhiên sau một giai đoạn tăng trưởng. Điều đáng chú ý là giữa các ngày, sự giảm giá được duy trì ổn định và không có những độ dao động lớn. Điều này là một dấu hiệu tích cực, chỉ ra sự kiểm soát và tích lũy từ phía nhà đầu tư, thậm chí khi thị trường đang chịu áp lực. Tổng thể, dù có sự giảm giá nhẹ, dự đoán vẫn phản ánh một giai đoạn ổn định và tích lũy sức mạnh trong thị trường cổ phiếu. Tuy nhiên, như mọi dự đoán

khác, nhà đầu tư cần theo dõi những yếu tố bên ngoài và thay đổi thị trường để điều chỉnh chiến lược đầu tư một cách linh hoạt và hiệu quả.

Tóm lại, dãy số dự đoán này cung cấp một cái nhìn tích cực về tính ổn định và khả năng dự đoán của mô hình LSTM, đặc biệt khi áp dụng thông tin từ 30 bước thời gian trước đó. Tuy nhiên, để đánh giá toàn diện hơn, việc so sánh với giá thực tế, xem xét biểu đồ và thực hiện phân tích thống kê sẽ cung cấp cái nhìn chi tiết và chính xác hơn về hiệu suất của mô hình trong môi trường thị trường cụ thể.



Hình 24 : Dự đoán giá đóng cửa trong vòng 30 ngày tới

Biểu đồ "Actual vs Predicted Close Prices for the Next 30 Days" thể hiện giá đóng cửa thực tế so với giá đóng cửa dự đoán trong khoảng thời gian từ tháng Ba đến tháng Mười Hai năm 2023. Đường màu xanh dương này biểu thị giá đóng cửa thực tế của cổ phiếu theo từng ngày. Chúng ta thấy giá cổ phiếu có xu hướng tăng từ tháng Ba đến giữa tháng Sáu, sau đó biến động với sự tăng giảm giá, nhưng tổng thể vẫn giữ được xu hướng tăng lên. Đường màu đỏ chỉ ra dự đoán của mô hình cho giá đóng cửa trong 30 ngày tiếp theo. Dự đoán này bắt đầu từ điểm cuối cùng của dữ liệu thực tế và tiếp tục cho thấy một xu hướng giảm nhẹ. Điều này có thể phản ánh sự đánh giá của mô hình về các yếu tố thị trường gần đây hoặc một sự điều chỉnh sau một giai đoạn tăng giá.

Biểu đồ cung cấp một cái nhìn về cách mà mô hình dự đoán có thể được sử dụng để dự báo giá cổ phiếu trong ngắn hạn. Dự đoán không chính xác hoàn toàn so với giá thực tế nhưng lại cung cấp cái nhìn về xu hướng mà mô hình đã học được. Sự chênh lệch giữa dự

đoán và thực tế có thể do nhiều nguyên nhân, bao gồm sự không chắc chắn của thị trường và giới hạn của mô hình.

Đối với nhà đầu tư, thông tin từ biểu đồ như thế này có thể hữu ích trong việc xác định chiến lược đầu tư, dù rằng nó không nên được dùng làm cơ sở duy nhất cho các quyết định đầu tư do tính không chắc chắn và biến động của thị trường chứng khoán.

4.7 So sánh 2 bước chạy

Bảng 9 : So sánh 10 steps và 30 steps

10 steps	30 steps
RMSE: 0.0502	RMSE: 0.0423
MSE: 0.0025	MSE: 0.0017
LSTM: 0.0411	LSTM: 0.0334

Các chỉ số đánh giá của mô hình LSTM với 10 và 30 bước thời gian không chỉ là những con số khô khan mà còn là bức tranh sống động về khả năng dự đoán của chúng đối với giá cổ phiếu. Với mô hình 10 bước thời gian, chúng ta bắt gặp RMSE là 0.05023, MSE là 0.00252 và MAE là 0.0411. Đối lập với đó, mô hình với 30 bước thời gian thể hiện sự cải thiện đáng kể với RMSE giảm xuống 0.0423, MSE giảm xuống 0.0017 và MAE giảm xuống 0.0334. Trong quá trình so sánh, rõ ràng mô hình với 30 bước thời gian nổi bật với hiệu suất tốt hơn. Sự giảm mạnh của các chỉ số RMSE, MSE và MAE đồng nghĩa với khả năng dự đoán chính xác và ổn định hơn của mô hình khi đối mặt với một khung thời gian dài hơn. Sự cải thiện này có thể được giải thích bằng khả năng của mô hình học được thông tin dài hạn hơn, làm tăng khả năng của nó trong việc nắm bắt xu hướng và biến động của giá cổ phiếu.

Nhìn chung, mô hình LSTM với 30 bước thời gian trở nên là lựa chọn ưu việt, không chỉ về mặt chính xác mà còn về tính ổn định, khi so sánh với phiên bản 10 bước thời gian. Điều này là một dấu hiệu tích cực cho những người quan tâm đến việc sử dụng mô hình này trong dự đoán giá cổ phiếu trong tương lai.

CHƯƠNG V : KẾT LUẬN VÀ KIẾN NGHỊ

5.1 Kết luận

Trong suốt 10 tuần thực tập, đã thu thập được một lượng lớn kiến thức và kỹ năng quý báu trong ngành khoa học dữ liệu. Ban đầu, tập trung vào việc thu thập và xử lý dữ liệu, nâng cao độ chính xác và hiệu suất công việc. Kỹ năng trực quan hóa dữ liệu cũng đã phát triển, cho phép thể hiện dữ liệu phức tạp một cách rõ ràng và trực quan thông qua đồ họa và hình ảnh. Sự hiểu biết về các tiêu chí đánh giá trong thử nghiệm cũng đã mở rộng, cải thiện khả năng phân tích dữ liệu một cách hệ thống. Việc áp dụng các mô hình học sâu vào các dự án thực tế đã không chỉ củng cố kiến thức lý thuyết mà còn mở ra cơ hội thực hành.

Các kỹ năng mềm như làm việc nhóm, giao tiếp và giải quyết vấn đề cũng đã được cải thiện đáng kể. Học cách thích ứng và làm việc hiệu quả trong một môi trường đa văn hóa, nâng cao khả năng giao tiếp và tương tác với đồng nghiệp cũng như các bên liên quan khác. Bên cạnh đó, kỹ năng quản lý dự án cũng đã phát triển, từ việc xác định mục tiêu đến lập kế hoạch và theo dõi tiến độ công việc, đảm bảo rằng mọi dự án được thực hiện một cách trơn tru và hiệu quả.

Tóm lại, quãng thời gian thực tập đã không chỉ mang lại cơ hội học hỏi mà còn là bước đệm vững chắc cho sự phát triển nghề nghiệp của tôi trong lĩnh vực khoa học dữ liệu, mở ra nhiều hướng phát triển và cơ hội mới trong tương lai.

5.2 Các mục tiêu đưa ra đã hoàn thành

Dự án đã gặt hái được những thành công ấn tượng và đã hoàn thành nhiều mục tiêu đề ra. Quản lý dự án được thực hiện một cách chặt chẽ và hiệu quả, đảm bảo rằng mọi hoạt động luôn diễn ra đúng kế hoạch. Công đoạn thu thập và xử lý dữ liệu ban đầu cùng với việc trực quan hóa thông tin đã tạo dựng nền móng vững chắc cho các giai đoạn sau của dự án.

Trong quá trình nghiên cứu và phát triển các chỉ số kỹ thuật, chúng tôi đã mở rộng kiến thức sâu rộng về ngành tài chính. Các lý thuyết nền tảng đã được củng cố và tích hợp vào quá trình phát triển mô hình học máy. Phương pháp phân tích dữ liệu thời gian như

Resampling và Difference đã chứng minh sự hiệu quả trong việc xử lý dữ liệu không đều, tạo nên dữ liệu và biểu đồ mạch lạc, dễ dàng tích hợp vào mô hình.

Mô hình LSTM được xây dựng và huấn luyện một cách kỹ lưỡng với các cấu hình thời gian khác nhau, mang lại dự đoán chính xác và ứng dụng thực tế cao. Kết quả từ mô hình đã được đánh giá một cách chi tiết và nhận được phản hồi tích cực từ phía doanh nghiệp, chứng tỏ sự hiệu quả và giá trị thực tế của dự án. Việc tích hợp các chỉ số kỹ thuật như SMA, EMA, RSI, và MACD vào mô hình đã làm tăng độ chính xác và phức tạp của dự đoán, nâng cao hiệu quả dự án.

Cuối cùng, việc đề xuất các hướng cải tiến và mở rộng cho dự án là bước quan trọng giúp dự án tiếp tục phát triển và cải thiện hiệu suất. Dự án không chỉ là một thành tựu độc lập mà còn là một phần quan trọng trong quá trình phát triển và áp dụng công nghệ dự báo tài chính. Những cải tiến này sẽ giúp dự án duy trì tính linh hoạt và hiệu suất cao, phản ứng linh hoạt với sự biến động của thị trường.

5.3 Kiến nghị

Dựa trên những kết quả tích cực từ việc áp dụng mô hình LSTM trong nghiên cứu, đề xuất mở rộng việc sử dụng mô hình này với chuỗi thời gian dài hơn trong các ứng dụng thực tiễn. Đã rõ ràng rằng mô hình LSTM, khi sử dụng dữ liệu từ nhiều bước thời gian trước, cho thấy khả năng dự đoán xuất sắc, mở ra một phương pháp mới trong việc phân tích và dự báo xu hướng thị trường tài chính.

Xác định một số hướng nghiên cứu và phát triển tiềm năng cho dự án. Các hướng này bao gồm cải tiến mô hình hiện tại, mở rộng phạm vi nghiên cứu, kết hợp mô hình với các phương pháp phân tích khác, và khám phá việc tích hợp công nghệ mới vào quy trình làm việc. Ngoài ra, tăng cường ứng dụng thực tế của dự án và chia sẻ kiến thức thông qua việc xuất bản bài báo khoa học, thực hiện nghiên cứu học thuật và phát triển các đề tài luận văn tốt nghiệp. Những hướng phát triển này không chỉ giúp củng cố và hoàn thiện dự án, mà còn đóng vai trò quan trọng trong việc mở rộng ứng dụng của nó trong lĩnh vực dự báo tài chính và học máy.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] Hochreiter S. & Schmidhuber, J., "Long short-term memory", Neural computation, 9(8), 1735-1780, 1997".
- [2] Priya C.B. & Arulanand N., "Univariate and multivariate models for Short-term wind speed forecasting", Mater. Today Proc, 2021.
- [3] Shi H., Xu, M. & Li R., "Deep learning for household load forecasting-A novel pooling deep RNN", IEEE Trans. Smart Grid, 9,5271–5280, 2017.
- [4] Weng B., Martinez W., Tsai Y.T., Li C., Lu L., Barth J.R. & Megahed F.M., "Macroeconomic indicators alone can predict the monthly closing price of major US indices: Insights from artificial intelligence, time-series analysis and hybrid models", Appl. SoftComput., 71, 685–697, 2018.
- [5] 21. Qiao W., Wang Y., Zhang J., Tian W., Tian Y. & Yang, Q., "An innovative coupled model given in view of wavelet transform for predicting short-term", PM10 concentration. J. Environ. Manag., 289, 112438, 2021.
- [6] Box G. E. P., Jenkins G. M., Reinsel G. C., & Ljung G. M., "Time Series Analysis: Forecasting and Control" John Wiley & Sons, 2015.
- [7] Karimnezhad A. & Moradi F. Bayes, "E-Bayes and robust Bayes prediction of a future observation under precautionary prediction loss functions with applications", Appl. Math. Model, 40, 7051–7061, 2016.
- [8] Moon M.A., "Demand and Supply Integration: The Key to World-Class Demand Forecasting", Walter de Gruyter GmbH & Co KG: Berlin, Germany, 2018.
- [9] Bruzda J., "Demand forecasting under fill rate constraints-The case of re-order points", Int. J. Forecast., 36, 1342-1361, 2020.
- [10] Ali Abadi & Kouhikamali R., "CFD-aided mathematical modeling of thermal vapor compressors in multiple effects distillationunits", Appl. Math. Model., 40, 6850-6868, 2016.
- [11] Nia A.R., Awasthi A., Bhuiyan N. "Industry 4.0 and demand forecasting of the energy supply chain", A. Comput. Ind. Eng.,154, 107128, 2021.

- [12] Hu M., Qiu R.T., Wu D.C. & Song H., "Hierarchical pattern recognition for tourism demand forecasting", *Tour. Manag.*, 84,104263, 2021.
- [13] Hyndman R. J. & Athanasopoulos G., "Forecasting: Principles and Practice", OTexts, 2018.
- [14] Siامي-Namini S., Tavakoli N. & Namin A.S., "A comparison of ARIMA and LSTM in forecasting time series", In *Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, pp. 1394–1401, 2018.
- [15] Fattah J., Ezzine L., Aman Z., El Moussami H. & Lachhab A. "Forecasting of demand using ARIMA model", *Int. J. Eng. Bus. Manag*, 10, 1847979018808673, 2018.
- [16] Salman, A.G. & Kanigoro B. "Visibility forecasting using autoregressive integrated moving average (ARIMA) models", *ProcediaComput. Sci.*, 179, 252–259, 2021.
- [17] Hochreiter S. & Schmidhuber, J., "Long short-term memory", *Neural computation*, 9(8), 1735-1780, 1997.
- [18] Priya C.B. & Arulanand N., "Univariate and multivariate models for Short-term wind speed forecasting", *Mater. Today Proc*, 2021.
- [19] Shi H., Xu, M. & Li R., "Deep learning for household load forecasting-A novel pooling deep RNN", *IEEE Trans. Smart Grid*, 9,5271–5280, 2017.
- [20] Lewis C.D. "International and Business Forecasting Methods", Butterworths: London, UK, 1982.
- [21] Weng B., Martinez W., Tsai Y.T., Li C., Lu L., Barth J.R. & Megahed F.M., "Macroeconomic indicators alone can predict the monthly closing price of major US indices: Insights from artificial intelligence, time-series analysis and hybrid models", *Appl. SoftComput.*, 71, 685–697, 2018."