**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**KHÓA LUẬN TỐT NGHIỆP**

**PHÂN TÍCH VÀ DỰ ĐOÁN THỊ TRƯỜNG CỔ PHIẾU VIỆT NAM**

**Chuyên ngành: Khoa học dữ liệu**

**Trần Văn Hùng– 20029141**

**Vũ Văn Hưng – 20036941**

**Giảng viên hướng dẫn: TS. Nguyễn Chí Kiên**

*TP. Hồ Chí Minh, tháng 12 năm 2024*

**INDUSTRIAL UNIVERSITY OF HO CHI MINH CITY**

**FACULTY OF INFORMATION TECHNOLOGY**

****

**GRADUATION THESIS**

**VIETNAM STOCK MARKET ANALYSIS AND PREDICTION**

**Major: Data Science**

**Tran Van Hung – 20029141**

**Vu Van Hung – 20036941**

**Instructor: PhD. Nguyen Chi Kien**

*Ho Chi Minh City, December 2024*

**VIETNAM STOCK MARKET ANALYSIS AND PREDICTION**

**Abstract**

The Vietnamese stock market has recently been developing at an unprecedented rate, catching the eye of both domestic and foreign investors. However, this market is highly volatile and inefficient due to its dynamic nature as an emerging market, thus rendering the analysis and prediction of its stock price movements highly complicated. This paper aims at improving these challenges by integrating traditional financial theories and advanced machine learning techniques in analyzing and stock price forecasting in Vietnam.

The research consists of two most important parts: fundamental and technical analysis. The fundamental analysis deals with intrinsic values of the stock based on financial ratios and macroeconomic indicators, while the technical analysis provides a prediction about the trend based on historical price and volume information. Furthermore, it proposes state-of-the-art models for machine learning, including LSTM and Transformer, for more precise forecasts of the closing price in stocks.

The findings of this research highlight the effectiveness of traditional and modern approaches in analyzing the Vietnamese stock market. The LSTM and Transformer models were powerful in giving good predictions that are helpful for investors and policymakers. This study also demonstrates the need to access more reliable data and utilize advanced technologies for better investment decision-making in emerging markets like Vietnam.

**LỜI CẢM ƠN**

Đầu tiên chúng em xin gửi lời cảm ơn đến trường Đại học Công nghiệp TP.HCM, nơi đã trang bị cho chúng em các nền tảng kiến thức vững chắc và những kỹ năng cần thiết cũng như đã tạo điều kiện thuận lợi cho chúng em tham gia học tập, nghiên cứu trong khoảng thời gian vừa qua.

Chúng em xin gửi lời cảm ơn chân thành nhất tới thầy Nguyễn Chí Kiên và thầy Vũ Đức Thịnh - Hai giảng viên hướng dẫn cho đồ án tốt nghiệp này cũng như thầy Trương Vĩnh Linh - Giáo viên chủ nhiệm. Chúng em rất biết ơn các thầy vì đã luôn sẵn sàng hướng dẫn và động viên nhóm chúng em trong suốt quá trình thực hiện khoá luận. Nhờ có sự tận tâm của các thầy, chúng em đã có thể hoàn thành được đồ án với kết quả tốt đẹp. Chúc em rất trân trọng những lời khuyên và những lời góp ý của hai thầy. Chúng em hi vọng vẫn sẽ nhận được sự quan tâm và hỗ trợ của các thầy trong tương lai.

Xin chân thành cảm ơn các thầy!

**NHẬN XÉT VÀ ĐÁNH GIÁ CỦA GIÁO VIÊN HƯỚNG DẪN**

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

Tp. Hồ Chí Minh, ngày.......tháng 12 năm 2024

Giáo viên hướng dẫn

(Ký ghi rõ họ tên)

**NHẬN XÉT VÀ ĐÁNH GIÁ CỦA GIÁO VIÊN PHẢN BIỆN 1**

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

Tp. Hồ Chí Minh, ngày.......tháng 12 năm 2024

Giáo viên phản biện 1

(Ký ghi rõ họ tên)

**NHẬN XÉT VÀ ĐÁNH GIÁ CỦA GIÁO VIÊN PHẢN BIỆN 2**

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

..........................................................................................................................................

Tp. Hồ Chí Minh, ngày.......tháng 12 năm 2024

Giáo viên phản biện 2

(Ký ghi rõ họ tên)

**MỤC LỤC**

[DANH MỤC HÌNH ẢNH 8](#_Toc409152792)

[DANH MỤC BẢNG BIỂU 9](#_Toc1671557647)

[DANH MỤC TỪ VIẾT TẮT 10](#_Toc753733205)

[CHƯƠNG 1. GIỚI THIỆU 11](#_Toc173027676)

[1.1. Tổng quan 12](#_Toc2058336642)

[1.1.1. Bối cảnh 12](#_Toc1213685148)

[1.1.2. Lý do chọn đề tài 12](#_Toc1814545574)

[1.2. Mục tiêu nghiên cứu 13](#_Toc474739204)

[1.3. Phạm vi nghiên cứu 13](#_Toc1472592482)

[1.4. Ý nghĩa khoa học và thực tiễn 13](#_Toc1322908131)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 14](#_Toc2133338759)

[2.1.Tổng quan về thị trường chứng khoán và giá đóng cửa của cổ phiếu 15](#_Toc1236470372)

[2.1.1.Tổng quan về cổ phiếu 15](#_Toc1835904392)

[2.1.2. Các loại giá cổ phiếu và phương pháp định giá 16](#_Toc1115990645)

[2.1.3 Ý nghĩa giá đóng cửa của cổ phiếu 17](#_Toc1836958451)

[2.2. Các nhân tố ảnh hưởng dến giá cổ phiếu 18](#_Toc1948846759)

[2.2.1. Yếu tố nội tại doanh nghiệp 18](#_Toc1203182736)

[2.2.2. Yếu tố ngành và lĩnh vực 19](#_Toc1365078586)

[2.2.3. Yếu tố vĩ mô kinh tế 19](#_Toc330982047)

[2.2.4. Yếu tố tâm lý và hành vi thị trường 19](#_Toc1418890406)

[2.3. Phương pháp phân tích và dự đoán cổ phiếu 20](#_Toc460294442)

[2.3.1. Phân tích cơ bản (Fundamental Analysis) 20](#_Toc1816508349)

[2.3.2. Phân tích kỹ thuật (Technical Analysis) 21](#_Toc595473588)

[2.3.3. Phương pháp định lượng và thống kê 21](#_Toc444844613)

[2.3.4. Phương pháp học máy và học sâu (Machine Learning và Deep Learning) 22](#_Toc1490387866)

[2.4 Tổng quan các nghiên cứu liên quan 23](#_Toc814621254)

[CHƯƠNG 3: DỮ LIỆU VÀ MÔ HÌNH 24](#_Toc1556665218)

[3.1. Dữ liệu 25](#_Toc133189891)

[3.2. Phương pháp phân tích dữ liệu 25](#_Toc1256094452)

[3.2.1. Thu thập dữ liệu 25](#_Toc157775729)

[3.2.2. Xử lý dữ liệu 25](#_Toc1632938368)

[3.2.3. Mô tả 26](#_Toc1495602198)

[3.3 Các mô hình dự đoán 26](#_Toc1323351689)

[3.3.1 Đường trung bình động (Moving Average - MA) 26](#_Toc989032198)

[3.3.2 Long Short-Term Memory (LSTM) 28](#_Toc28009742)

[3.3.3 Hồi quy tuyến tính (Linear Regression) 30](#_Toc800186997)

[3.3.4. Transformer 31](#_Toc2037447864)

[CHƯƠNG 4: THỰC NGHIỆM VÀ KẾT QUẢ 35](#_Toc830591419)

[4.1. Thực nghiệm 35](#_Toc1924257718)

[4.1.1. Dữ liệu 35](#_Toc1563527707)

[4.1.2. Tham số khi train mô hình 35](#_Toc764832024)

[4.2. Thực nghiệm 35](#_Toc275127044)

[4.2.1. Đường trung bình động (Moving Average) 35](#_Toc1693670478)

[4.2.2 Stock prediction with Linear Regression: 37](#_Toc673785504)

[4.2.3 Long Short-Term memory (LSTM) 39](#_Toc755336910)

[4.2.4 Tranformer 42](#_Toc17758264)

[CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 45](#_Toc1589022961)

[5.1. Kết luận 46](#_Toc1205646479)

[5.2. Phân tích 46](#_Toc901032006)

[5.3. Hạn chế va hướng phát triển 47](#_Toc1481188991)

[TÀI LIỆU THAM KHẢO 47](#_Toc264911400)

[KẾ HOẠCH THỰC HIỆN 49](#_Toc851365466)

# DANH MỤC HÌNH ẢNH

Ảnh 3.1 : Mô hình kiến trúc của Transformer..........................................................22

Ảnh 4.1: Biểu đồ trực quan hóa kết quả MA..............................................................27

Ảnh 4.2: Biểu đồ trực quan hóa kết quả dự đoán Linear regression...........................29

Ảnh 4.3: Biểu đồ trực quan hóa kết quả dự đoán Long Short-Term memory............32

# DANH MỤC BẢNG BIỂU

Bảng 1 : Bảng chia dữ liệu ..........................................................................................25

Bảng 2 : Bảng thông số train mô hình.........................................................................25

Bảng 3: Bảng kết quả dự đoán mô hình......................................................................36

# DANH MỤC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **TỪ VIẾT TẮT** | **TỪ ĐẦY ĐỦ** | **Ý NGHĨA** |
| LSTM | Long short term memory | Mạng thần kinh hồi quy(RNN) |
| RSME | Root Mean Squared Error | một độ đo thước đo sai số giữa giá trị dự đoán và giá trị thực tế |
| MA | Moving Average | Trung bình động |

# CHƯƠNG 1. GIỚI THIỆU

## 1.1. Tổng quan

### **1.1.1. Bối cảnh**

Theo số liệu thống kê thị trường Việt Nam được đánh là là 1 trong những thị trường có sức chống chịu và phục hồi tốt nhất. Quy mô thị trường vượt cả mục tiêu do Chính phủ đề ra, và hiện tại vẫn duy trì tăng trưởng. Thanh khoản của thị trường Việt Nam tăng mạnh lên mức kỷ lục cho thấy sức hấp dẫn của thị trường chứng khoán Việt. Số lượng tài khoản được mở mới tăng kỷ lục vào năm 2020. Thị trường chứng khoán tiếp tục khẳng định là kênh huy động vốn quan trọng với tổng mức huy động tăng 13.6% so với cuối năm 2023.

Theo các số liệu thống kê, nhóm cổ phiếu thuộc ngành Ngân hàng tại Việt Nam hiện đang thu hút sự quan tâm đáng kể từ cả nhà đầu tư trong nước lẫn quốc tế. Đây là một trong những lĩnh vực được quản lý chặt chẽ bởi các quy định pháp luật, với mức độ cải cách và minh bạch ngày càng cao so với nhiều ngành khác. Các vấn đề như quản trị, quản lý rủi ro hay vận hành theo chuẩn quốc tế cũng được các ngân hàng chú trọng triển khai.

Ngoài ra, cổ phiếu ngành Ngân hàng thường có giá trị vốn hóa lớn, nằm trong nhóm dẫn đầu về lợi nhuận trên thị trường chứng khoán Việt Nam. Với vai trò quan trọng trong việc thúc đẩy sự phát triển của thị trường, đây là nhóm cổ phiếu được đánh giá có tiềm năng đầu tư cao, nhờ số lượng ngân hàng niêm yết khá lớn. Tuy vậy, các yếu tố cụ thể ảnh hưởng đến giá cổ phiếu của ngành này vẫn cần được nghiên cứu kỹ lưỡng để làm rõ hơn.

### **1.1.2. Lý do chọn đề tài**

Phân tích và định giá cổ phiếu đóng vai trò quan trọng trên thị trường chứng khoán, ngoài ra còn có ý nghĩa quan trọng đối với các cá nhân, doanh nhiệp tham gia thị trường. Doanh nghiệp, cá nhân sử dụng phân tích để dưa ra các quyết định về huy động nguồn vốn, cho vay nhằm mục đích kiếm lời. Phân tích tài chính đóng vai trò quan trọng là thế, nhưng hiện nay phân tích tài chính, cổ phiếu gặp nhiều khó khăn, trở ngại trong việc chọn mô hình định giá, trong ước lượng tham số cho việc định giá vay hoặc các phát sinh từ Chính phủ, các cơ sở nhà nước, các công ty chứng khoán, các nhà đầu tư.

Trong các cổ phiếu được định giá thì các cổ phiếu của doanh nghiệp có mức tăng trưởng cao, ổn định có tầm quan trọng rất lớn. Bản thân các doanh nghiệp này thường xuyên phải định giá lại cổ phiếu để đưa ra các quyết định quan trong trong việc huy động vốn, các ngân hàng thường cho các doanh nghiệp này vay dễ dàng hơn, các nhà đầu tư cũng ưa thích đầu tư vào các doanh nghiệp này, vì vậy công tác phân tích, định giá cổ phiểu đóng có vai trò quan trọng trong việc đầu tư.

Vì vậy em chọn đề tài “Phân tích và dự đoán giá cổ phiếu” làm đề tài nghiên cứu cho khóa luận của mình.

## 1.2. Mục tiêu nghiên cứu

Mục đích nghiên cứu của khóa luận nhằm mục đích tổng hợp các mô hình có hiệu quả tối ưu nhất trong việc dự đoán giá đóng cửa cổ phiếu, để từ đó lựa chọn ra mô hình phù hợp nhất. Đề tài tập trung áp dụng phân tích cơ bản để đánh giá giá trị nội tại của cổ phiếu thông qua các chỉ số tài chính kết hợp với phân tích kỹ thuật nhằm nhận diện xu hướng giá và các tín hiệu giao dịch. Đồng thời nêu lên những khó khăn trong phân tích và định giá giá đóng cửa trên thị trường để từ đó đưa ra những giải pháp, kiến nghị giúp cho việc phân tích, định giá chính xác và thuận lợi hơn.

## 1.3. Phạm vi nghiên cứu

Đối tượng của khóa luận là các cổ phiếu niêm yết trên thị trường chứng khoán Việt Nam. Phạm vi nghiên cứu của khóa luận tập chung vào giai đoạn từ năm 01/2023 đến 09/2024 để từ đó định giá cổ phiếu.

## 1.4. Ý nghĩa khoa học và thực tiễn

Phân tích và dự đoán thị trường cổ phiếu tại Việt Nam không chỉ mang ý nghĩa quan trọng trong lĩnh vực nghiên cứu tài chính mà còn có giá trị thực tiễn cao, đặc biệt trong bối cảnh nền kinh tế đang phát triển mạnh mẽ và hội nhập quốc tế.

*-****Ý nghĩa khoa học***

+ Phát triển mô hình phân tích tài chính: Việc nghiên cứu thị trường cổ phiếu giúp xây dựng và cải tiến các mô hình phân tích định lượng, dự đoán xu hướng giá cổ phiếu, và đánh giá rủi ro.

+ Củng cố lý thuyết thị trường hiệu quả: Qua phân tích thực tiễn, luận văn cung cấp thêm bằng chứng để kiểm định lý thuyết thị trường hiệu quả (Efficient Market Hypothesis - EMH) trong bối cảnh thị trường mới nổi như Việt Nam.

+ Ứng dụng công nghệ và phương pháp hiện đại: Sử dụng các công cụ như trí tuệ nhân tạo (AI), học máy (machine learning), và phân tích dữ liệu lớn (big data) để cải thiện độ chính xác của dự báo, đồng thời đề xuất phương pháp tối ưu trong nghiên cứu tài chính.

***- Ý nghĩa thực tiễn***

+ Hỗ trợ nhà đầu tư: Kết quả nghiên cứu cung cấp thông tin hữu ích để các nhà đầu tư, từ cá nhân đến tổ chức, có cơ sở ra quyết định đầu tư hiệu quả hơn.

+ Đóng góp cho cơ quan quản lý: Giúp cơ quan quản lý thị trường chứng khoán hiểu rõ hơn về xu hướng và yếu tố tác động, từ đó điều chỉnh chính sách phù hợp nhằm đảm bảo sự minh bạch và ổn định của thị trường.

+ Nâng cao nhận thức tài chính: Luận văn góp phần nâng cao kiến thức tài chính cho công chúng, giúp họ hiểu rõ hơn về cách vận hành và rủi ro của thị trường cổ phiếu.

# CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

## 2.1.Tổng quan về thị trường chứng khoán và giá đóng cửa của cổ phiếu

### **2.1.1.Tổng quan về cổ phiếu**

Cổ phiếu là một loại chứng khoán thể hiện quyền sở hữu một phần vốn điều lệ của công ty cổ phần. Người sở hữu cổ phiếu được gọi là cổ đông và có quyền tham gia vào các quyết định quan trọng của công ty thông qua các cuộc họp Đại hội đồng cổ đông. Đồng thời, cổ đông cũng có quyền hưởng cổ tức nếu công ty hoạt động có lãi hoặc chuyển nhượng cổ phiếu của mình trên thị trường chứng khoán. Cổ phiếu không chỉ là công cụ đầu tư phổ biến mà còn là phương tiện giúp doanh nghiệp huy động vốn hiệu quả từ công chúng, thúc đẩy sự phát triển của nền kinh tế.

- Phân loại cổ phiếu:

+ Cổ phiếu thường (Common Stock):

* Quyền lợi: Người sở hữu cổ phiếu thường có quyền biểu quyết tại Đại hội đồng cổ đông, được nhận cổ tức không cố định phụ thuộc vào kết quả kinh doanh của công ty.
* Rủi ro: Trong trường hợp công ty phá sản, cổ đông thường chỉ nhận lại tài sản sau khi công ty đã thanh toán nợ và nghĩa vụ cho các cổ đông ưu đãi.

+ Cổ phiếu ưu đãi (Preferred Stock):

* Quyền lợi: Người sở hữu cổ phiếu ưu đãi được nhận cổ tức cố định trước cổ đông thường và có thể nhận vốn góp trước trong trường hợp công ty giải thể.
* Hạn chế: Thường không có quyền biểu quyết hoặc quyền hạn chế so với cổ đông thường.

- Đặc điểm của cổ phiếu

+ Tính thanh khoản cao: Cổ phiếu có thể được mua bán trên thị trường chứng khoán một cách dễ dàng, giúp nhà đầu tư nhanh chóng chuyển đổi tài sản đầu tư thành tiền mặt.

+ Lợi nhuận cao đi kèm rủi ro lớn: Giá trị cổ phiếu có thể tăng mạnh, mang lại lợi nhuận cao cho nhà đầu tư. Tuy nhiên, cổ phiếu cũng chịu tác động lớn từ các biến động thị trường, gây ra rủi ro về giá.

+ Không có thời hạn hoàn trả:Không giống như trái phiếu, cổ phiếu không có thời hạn đáo hạn, đồng nghĩa với việc doanh nghiệp không phải trả lại vốn góp cho cổ đông.

+ Tính chia nhỏ:Cổ phiếu có thể được chia nhỏ để tăng khả năng giao dịch, đồng thời tạo cơ hội cho nhiều nhà đầu tư tham gia thị trường.

### **2.1.2. Các loại giá cổ phiếu và phương pháp định giá**

#### **2.1.2.1.Giá mở cửa**

Giá mở cửa là mức giá tại thời điểm bắt đầu giao dịch trong một phiên làm việc của thị trường chứng khoán. Đây là giá của giao dịch đầu tiên được thực hiện khi thị trường mở cửa sau một kỳ nghỉ hoặc phiên giao dịch trước đó. Giá mở cửa phản ánh sự kỳ vọng của nhà đầu tư và tâm lý thị trường vào thời điểm bắt đầu phiên giao dịch, đồng thời có thể bị ảnh hưởng bởi các yếu tố vĩ mô, tin tức kinh tế, hoặc các yếu tố bên ngoài khác xảy ra sau khi phiên giao dịch trước đó kết thúc.

#### **2.1.2.2 Giá cao nhất**

Giá cao nhất của cổ phiếu trong một phiên giao dịch là mức giá đạt được cao nhất trong suốt thời gian mở cửa cho đến khi kết thúc phiên giao dịch. Đây là chỉ báo quan trọng phản ánh sự quan tâm mạnh mẽ và nhu cầu mua cổ phiếu trong phiên, thường liên quan đến các yếu tố như sự tích cực của thị trường hoặc các tin tức hỗ trợ tích cực đối với công ty. Giá cao nhất có thể giúp đánh giá mức độ biến động và sức mạnh của cổ phiếu trong ngày giao dịch.

#### **2.1.2.3.Giá thấp nhất**

Giá thấp nhất của cổ phiếu trong một phiên giao dịch là mức giá thấp nhất mà cổ phiếu đạt được trong suốt thời gian thị trường mở cửa. Đây là chỉ số quan trọng phản ánh sự giảm giá trong suốt phiên giao dịch, có thể do tác động của các yếu tố tiêu cực hoặc sự thay đổi tâm lý của nhà đầu tư. Giá thấp nhất giúp cung cấp thông tin về mức độ suy giảm giá trị của cổ phiếu trong ngày và có thể là cơ sở cho các quyết định mua bán của nhà đầu tư.

#### **2.1.2.4 Giá đóng cửa**

Giá đóng cửa của cổ phiếu là mức giá của cổ phiếu trong giao dịch cuối cùng của phiên, khi thị trường kết thúc hoạt động trong một ngày giao dịch. Đây là mức giá thể hiện sự tổng hợp của tất cả các giao dịch trong phiên, phản ánh kết quả cuối cùng của thị trường trong ngày và được coi là chỉ báo quan trọng để đánh giá xu hướng giá cổ phiếu. Giá đóng cửa thường được sử dụng làm cơ sở để phân tích và dự đoán biến động giá trong các phiên giao dịch tiếp theo, vì nó cung cấp cái nhìn tổng quan về sự cân bằng giữa cung và cầu trong suốt thời gian giao dịch.

#### **2.1.2.5 Giá điều chỉnh**

Giá điều chỉnh của cổ phiếu là mức giá được điều chỉnh lại để phản ánh các sự kiện đặc biệt có ảnh hưởng đến giá trị cổ phiếu, chẳng hạn như chia cổ tức, chia tách cổ phiếu, hoặc sáp nhập công ty. Mục đích của việc điều chỉnh giá là để loại bỏ tác động của những sự kiện này, giúp việc phân tích và so sánh giá cổ phiếu qua thời gian trở nên chính xác hơn. Giá điều chỉnh cung cấp một cái nhìn rõ ràng hơn về xu hướng thực sự của cổ phiếu mà không bị ảnh hưởng bởi các yếu tố ngoại lai.

### **2.1.3 Ý nghĩa giá đóng cửa của cổ phiếu**

Giá đóng cửa của cổ phiếu đóng vai trò rất quan trọng trong việc đánh giá tình hình thị trường trong một phiên giao dịch và dự đoán xu hướng giá trong tương lai. Đây là mức giá cuối cùng khi thị trường kết thúc, phản ánh sự tổng hợp của tất cả các giao dịch diễn ra trong suốt phiên.

Dự báo xu hướng thị trường: Giá đóng cửa là chỉ báo chính được các nhà đầu tư và phân tích kỹ thuật sử dụng để xác định xu hướng tiếp theo của cổ phiếu. Mức giá này thường được xem là tín hiệu quan trọng cho các quyết định mua hoặc bán trong các phiên giao dịch tiếp theo.

Cung cấp thông tin chính xác về tâm lý thị trường: Giá đóng cửa phản ánh sự cân bằng giữa cung và cầu của cổ phiếu, là dấu hiệu cho thấy sức mua và sức bán trong suốt phiên giao dịch. Một giá đóng cửa cao so với mức giá mở cửa có thể cho thấy tâm lý tích cực từ nhà đầu tư, trong khi giá đóng cửa thấp có thể là dấu hiệu của sự suy giảm niềm tin vào cổ phiếu.

Công cụ cho các chỉ báo kỹ thuật: Giá đóng cửa là đầu vào chủ yếu cho hầu hết các chỉ báo kỹ thuật như đường trung bình động (MA), chỉ báo sức mạnh tương đối (RSI), và các mô hình giá. Nhờ vào giá đóng cửa, các nhà phân tích có thể xây dựng các mô hình và dự đoán chính xác hơn về diễn biến giá trong các ngày tiếp theo.

Đo lường hiệu quả hoạt động của cổ phiếu: Giá đóng cửa cũng là công cụ quan trọng để đo lường sự hiệu quả của một cổ phiếu trong một khoảng thời gian cụ thể, giúp các nhà đầu tư theo dõi sự thay đổi trong giá trị cổ phiếu và đưa ra quyết định đầu tư hợp lý.

## 2.2. Các nhân tố ảnh hưởng dến giá cổ phiếu

### **2.2.1. Yếu tố nội tại doanh nghiệp**

Giá cổ phiếu chịu tác động lớn từ tình hình hoạt động và quản trị của doanh nghiệp phát hành.

Hiệu quả hoạt động kinh doanh: Doanh thu, lợi nhuận ròng và biên lợi nhuận của công ty có mối liên hệ trực tiếp với giá cổ phiếu. Khi doanh nghiệp báo cáo kết quả kinh doanh tích cực, giá cổ phiếu thường tăng nhờ sự kỳ vọng của nhà đầu tư.

Quản trị doanh nghiệp: Một ban lãnh đạo có tầm nhìn chiến lược, khả năng quản trị rủi ro và thực hiện các quyết định đầu tư hiệu quả sẽ tăng niềm tin của nhà đầu tư, qua đó tác động tích cực đến giá cổ phiếu.

Tình hình tài chính:Các chỉ số tài chính như tỷ lệ nợ trên vốn chủ sở hữu (D/E), khả năng thanh toán và dòng tiền tự do đều là những yếu tố ảnh hưởng đến sự ổn định và tiềm năng tăng trưởng của cổ phiếu.

### **2.2.2. Yếu tố ngành và lĩnh vực**

- Cạnh tranh trong ngành:Mức độ cạnh tranh có thể ảnh hưởng đến thị phần và lợi nhuận của doanh nghiệp. Một ngành có sự cạnh tranh khốc liệt thường tạo áp lực giảm giá cổ phiếu.

- Xu hướng phát triển ngành: Những ngành đang tăng trưởng nhanh (như công nghệ, năng lượng tái tạo) thường thu hút dòng vốn đầu tư lớn, làm tăng giá cổ phiếu của các công ty trong lĩnh vực đó.

- Quy định pháp luật: Các chính sách, quy định và thuế áp dụng cho ngành sẽ ảnh hưởng đến hoạt động kinh doanh và giá cổ phiếu.

### **2.2.3. Yếu tố vĩ mô kinh tế**

Tăng trưởng kinh tế: Sự tăng trưởng GDP thường đi kèm với kỳ vọng tích cực cho doanh nghiệp, tạo động lực tăng giá cổ phiếu.

Lãi suất và chính sách tiền tệ: Lãi suất cao làm tăng chi phí vay vốn và giảm khả năng sinh lời, từ đó có thể làm giảm giá cổ phiếu. Ngược lại, lãi suất thấp thường thúc đẩy đầu tư và tăng giá trị cổ phiếu.

Lạm phát: Lạm phát vừa phải có thể thúc đẩy giá cổ phiếu do doanh nghiệp có khả năng điều chỉnh giá bán. Tuy nhiên, lạm phát quá cao làm tăng chi phí và giảm sức mua, gây áp lực giảm giá cổ phiếu.

Tỷ giá hối đoái: Những doanh nghiệp xuất khẩu lớn sẽ được hưởng lợi nếu đồng nội tệ giảm giá, làm tăng giá cổ phiếu. Ngược lại, các doanh nghiệp nhập khẩu sẽ chịu áp lực chi phí tăng cao.

### **2.2.4. Yếu tố tâm lý và hành vi thị trường**

Tin tức và sự kiện: Những tin tức tích cực như ký kết hợp đồng lớn, mở rộng thị trường hay thay đổi ban lãnh đạo hiệu quả có thể làm tăng giá cổ phiếu. Ngược lại, tin tức tiêu cực như bê bối tài chính hoặc phá sản sẽ làm giảm giá.

Tâm lý nhà đầu tư: Sự kỳ vọng hoặc lo sợ của nhà đầu tư thường dẫn đến những biến động giá không phản ánh chính xác giá trị thực của cổ phiếu.

Giao dịch của các tổ chức lớn: Hoạt động mua hoặc bán của các quỹ đầu tư, tổ chức tài chính có thể tạo áp lực tăng hoặc giảm giá cổ phiếu đáng kể, do họ thường nắm giữ khối lượng cổ phiếu lớn.

## 2.3. Phương pháp phân tích và dự đoán cổ phiếu

### **2.3.1. Phân tích cơ bản (Fundamental Analysis)**

Phân tích cơ bản tập trung vào việc đánh giá giá trị nội tại của cổ phiếu bằng cách xem xét các yếu tố tài chính và kinh tế ảnh hưởng đến doanh nghiệp

1.1 Các yếu tố tài chính:

- Báo cáo tài chính: Doanh thu, lợi nhuận ròng, dòng tiền.

- Các chỉ số tài chính:

* Tỷ lệ P/E (Price-to-Earnings): Đo lường mối quan hệ giữa giá cổ phiếu và lợi nhuận mỗi cổ phiếu.
* Tỷ lệ P/B (Price-to-Book): So sánh giá cổ phiếu với giá trị sổ sách của doanh nghiệp.
* ROE (Return on Equity): Phản ánh khả năng sinh lời trên vốn chủ sở hữu.

1.2 Các yếu tố kinh tế và ngành:

* Tăng trưởng GDP, tỷ lệ lạm phát, chính sách tiền tệ và thuế.
* Cạnh tranh trong ngành, xu hướng phát triển và thị phần của doanh nghiệp.

- Ưu điểm:

* Phù hợp để đánh giá giá trị dài hạn của cổ phiếu.
* Cung cấp cái nhìn tổng quan về tình hình tài chính và triển vọng doanh nghiệp.

- Hạn chế:

* Đòi hỏi dữ liệu chi tiết và thời gian nghiên cứu lâu dài.
* Không hiệu quả trong việc dự đoán biến động giá ngắn hạn.

### **2.3.2. Phân tích kỹ thuật (Technical Analysis)**

Phân tích kỹ thuật dựa trên dữ liệu giá và khối lượng giao dịch trong quá khứ để dự đoán xu hướng giá cổ phiếu trong tương lai.

2.1 Nguyên lý cơ bản:

* Giá thị trường phản ánh tất cả các yếu tố kinh tế và tâm lý.
* Lịch sử giá có xu hướng lặp lại do hành vi của nhà đầu tư.

2.2 Các công cụ phân tích:

* Đường trung bình động (Moving Average): Theo dõi xu hướng giá cổ phiếu trong một khoảng thời gian nhất định.
* Chỉ báo sức mạnh tương đối (RSI): Đo lường tốc độ và mức thay đổi của giá để xác định cổ phiếu quá mua hoặc quá bán.
* Mô hình nến Nhật: Phân tích tâm lý nhà đầu tư qua các mẫu hình giá.

- Ưu điểm:

* Hiệu quả trong phân tích xu hướng ngắn hạn.
* Dễ dàng áp dụng với các công cụ trực tuyến.

- Hạn chế:

* Dựa trên giả định quá khứ lặp lại, có thể không phù hợp trong các trường hợp bất thường.
* Không xem xét yếu tố cơ bản của doanh nghiệp.

### **2.3.3. Phương pháp định lượng và thống kê**

Các phương pháp này sử dụng mô hình toán học và dữ liệu lịch sử để dự đoán giá cổ phiếu.

3.1 Mô hình hồi quy:

* Mô hình tuyến tính hoặc phi tuyến tính dựa trên mối quan hệ giữa các yếu tố như lãi suất, GDP, hoặc chỉ số tài chính với giá cổ phiếu.

3.2 Mô hình chuỗi thời gian:

* ARIMA (AutoRegressive Integrated Moving Average): Dự đoán giá cổ phiếu dựa trên các giá trị trong quá khứ và xu hướng biến động.
* GARCH (Generalized Autoregressive Conditional Heteroskedasticity): Xác định sự biến động của giá theo thời gian.

- Ưu điểm:

* Đưa ra dự đoán dựa trên dữ liệu cụ thể và có căn cứ.
* Phù hợp với các nhà đầu tư tổ chức hoặc nghiên cứu khoa học.

- Hạn chế: Phức tạp, đòi hỏi kỹ năng xử lý dữ liệu và công cụ thống kê.

### **2.3.4. Phương pháp học máy và học sâu (Machine Learning và Deep Learning)**

Các phương pháp hiện đại sử dụng trí tuệ nhân tạo để dự đoán giá cổ phiếu dựa trên dữ liệu lớn và phi tuyến tính.

4.1 Học máy (Machine Learning):

* Các thuật toán phổ biến: Hồi quy tuyến tính, cây quyết định (Decision Trees), rừng ngẫu nhiên (Random Forest).

4.2 Học sâu (Deep Learning):

* Mạng nơ-ron nhân tạo (Artificial Neural Networks): Mô phỏng hoạt động của não bộ để học và dự đoán.
* Mạng LSTM (Long Short-Term Memory): Phân tích chuỗi thời gian để dự đoán các xu hướng phức tạp của giá cổ phiếu.

- Ưu điểm:

* Xử lý tốt các tập dữ liệu lớn và phi tuyến tính.
* Hiệu quả trong việc dự đoán giá cổ phiếu ngắn hạn và trung hạn.

- Hạn chế:

* Yêu cầu cao về kỹ thuật và công cụ phân tích.
* Phụ thuộc vào chất lượng và độ đầy đủ của dữ liệu.

## 2.4 Tổng quan các nghiên cứu liên quan

**Fama (1970)**: Đề xuất lý thuyết "Thị trường hiệu quả" (Efficient Market Hypothesis), cho rằng không thể dự đoán giá cổ phiếu dựa trên thông tin lịch sử vì thị trường đã phản ánh tất cả thông tin. Tuy nhiên, nghiên cứu sau này cho thấy các chỉ báo kỹ thuật có thể khai thác được các bất thường trong thị trường.

**Box & Jenkins (1970)**: Phát triển mô hình ARIMA, cho phép dự đoán chuỗi thời gian tài chính. Mô hình này đã được ứng dụng rộng rãi để dự đoán giá cổ phiếu trong các thị trường phát triển.

**Prophet (Facebook, 2017):** Một mô hình chuỗi thời gian được thiết kế để phân tích và dự đoán dữ liệu tài chính có tính chu kỳ và biến động mạnh.

==> Kết luận : Tổng kết lại, trong bài luận văn nghiên cứu của chúng tôi đã lựa chọn một cách tiếp cận sử dụng các phương pháp từ phân tích kỹ thuật cơ bản như đường Trung bình động (Moving Average) đến các phương pháp phức tạp hơn như học máy và học sâu Long Short-Term Memory (LSTM), Hồi quy tuyến tính (Linear Regression). Mỗi phương pháp đều có những ưu điểm và hạn chế riêng, do đó việc sử dụng chúng giúp tận dụng được những điểm mạnh và giảm thiểu các nhược điểm. Các phương pháp đơn giản như Moving Average và hồi quy tuyến tính có thể hoạt động hiệu quả trong ngắn hạn, trong khi các mô hình máy học và học sâu như Tranfomer và LSTM giúp cải thiện dự đoán khi xử lý dữ liệu phức tạp và dài hạn.

Trong tương lai, nghiên cứu có thể mở rộng bằng cách sử dụng các mô hình phức tạp hơn, cải tiến thuật toán học sâu để giảm thiểu tình trạng overfitting và thử nghiệm với các bộ dữ liệu lớn hơn để đánh giá tính hiệu quả trong các điều kiện thị trường thay đổi nhanh chóng. Việc áp dụng các mô hình học sâu và phân tích dữ liệu lớn trong thị trường chứng khoán có thể giúp nâng cao độ chính xác của các dự đoán, đồng thời cung cấp các chiến lược đầu tư thông minh hơn cho các nhà đầu tư và các chuyên gia tài chính.

# CHƯƠNG 3: DỮ LIỆU VÀ MÔ HÌNH

## 3.1. Dữ liệu

Dữ liệu sử dụng trong nghiên cứu này được thu thập từ thông tin giao dịch trên thị trường chứng khoán Việt Nam bằng cách sử dụng vnstock của python, dữ liệu được lấy trong giai đoạn từ tháng 01/2023 đến tháng 09/2024. Bộ dữ liệu chứa các trường thông tin cơ bản bao gồm:

* time: Ngày giao dịch.(object)
* open: Giá mở cửa của cổ phiếu trong phiên giao dịch.(int64)
* high: Giá cao nhất được ghi nhận trong phiên giao dịch. (int64)
* low: Giá thấp nhất được ghi nhận trong phiên giao dịch. (int64)
* close: Giá đóng cửa của cổ phiếu trong phiên giao dịch. (int64)
* volume: Khối lượng giao dịch (số lượng cổ phiếu được trao đổi). (int64)
* ticker: Mã cổ phiếu . (object)

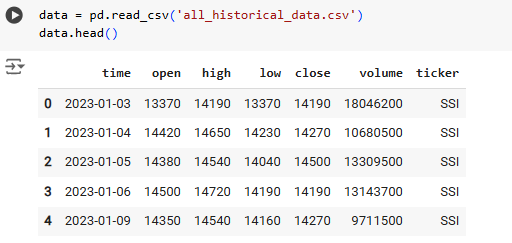
## 3.2. Phương pháp phân tích dữ liệu

Phần này trình bày chi tiết các phương pháp được sử dụng trong nghiên cứu để phân tích và dự đoán giá đóng cửa của cổ phiếu. Quy trình bao gồm việc xử lý dữ liệu, áp dụng các kỹ thuật phân tích kỹ thuật, sử dụng mô hình thống kê, máy học, và học sâu để tối ưu hóa kết quả dự đoán.

### **3.2.1. Thu thập dữ liệu**

Dữ liệu giao dịch của cổ phiếu được thu thập từ thư viện vnstock.

Các trường dữ liệu bao gồm: ngày giao dịch, giá mở cửa, giá cao nhất, giá thấp nhất, giá đóng cửa và khối lượng giao dịch.



### **3.2.2. Xử lý dữ liệu**

Kiểm tra dữ liệu thiếu: Loại bỏ hoặc thay thế các giá trị thiếu bằng phương pháp nội suy hoặc giá trị trung bình.

Loại bỏ ngoại lệ: Kiểm tra và điều chỉnh các giá trị bất thường để tránh ảnh hưởng tiêu cực đến mô hình.

Chuẩn hóa dữ liệu: Chuyển đổi các biến đầu vào về cùng thang đo để phù hợp với yêu cầu của các thuật toán.

### **3.2.3. Mô tả**

Tóm tắt các đặc điểm cơ bản của dữ liệu như trung bình, phương sai, và xu hướng.

Xác định các mối quan hệ giữa các biến như giá mở cửa, giá cao nhất và giá thấp nhất với giá đóng cửa.

Áp dụng các mô hình phân tích và dự đoán: Sử dụng các kỹ thuật phân tích kỹ thuật, mô hình thống kê, máy học và học sâu để xây dựng các mô hình dự đoán giá.

Sử dụng các tiêu chí đánh giá MAE, RMSE (Root Mean Squared Error) để đo lường độ chính xác của các mô hình

## 3.3 Các mô hình dự đoán

Phần này trình bày chi tiết về các mô hình được sử dụng trong nghiên cứu để dự đoán giá đóng cửa của cổ phiếu. Các mô hình bao gồm từ những phương pháp phân tích kỹ thuật cơ bản đến các mô hình máy học và học sâu hiện đại. Mỗi mô hình được lựa chọn đều có vai trò và đóng góp riêng trong việc nâng cao độ chính xác của dự đoán.

### **3.3.1 Đường trung bình động (Moving Average - MA)**

Đường trung bình động (Moving Average - MA) là một công cụ phân tích kỹ thuật cơ bản được sử dụng để làm mịn chuỗi dữ liệu giá cổ phiếu, giúp nhận diện xu hướng chính bằng cách loại bỏ các biến động ngẫu nhiên. MA được tính bằng cách lấy trung bình giá cổ phiếu trong một khoảng thời gian nhất định.

- Các loại đường trung bình động phổ biến bao gồm:

* MA-10: Trung bình giá trong 10 ngày gần nhất.
* MA-20: Trung bình giá trong 20 ngày gần nhất.
* MA-50: Trung bình giá trong 50 ngày gần nhất.

- Công thức : Công thức tính trung bình động đơn giản (Simple Moving Average - SMA) cho 𝑛 ngày là:

Trong đó :

: Trung bình động trong ngày.

: Giá đóng cửa của cổ phiếu vào ngày i.

: Số ngày được tính trung bình

- Ý nghĩa của đường trung bình động :

Đường trung bình động đóng vai trò quan trọng trong phân tích kỹ thuật, giúp nhà đầu tư nhận diện xu hướng giá cổ phiếu:

+ Xác định xu hướng chính:

* Khi giá cổ phiếu nằm trên đường MA, xu hướng tăng thường được xác nhận.
* Khi giá nằm dưới đường MA, xu hướng giảm có thể xảy ra.

+ Loại bỏ nhiễu: MA giúp giảm thiểu tác động của các biến động giá bất thường, cung cấp cái nhìn rõ ràng hơn về xu hướng tổng thể.

+ Tín hiệu giao dịch:

* Giao điểm giữa các đường MA ngắn hạn và dài hạn (crossover) thường được sử dụng làm tín hiệu mua hoặc bán.
* Golden Cross (giao cắt vàng): Đường MA ngắn hạn cắt lên đường MA dài hạn, báo hiệu xu hướng tăng.
* Death Cross (giao cắt tử thần): Đường MA ngắn hạn cắt xuống đường MA dài hạn, báo hiệu xu hướng giảm.

- Ứng dụng trong nghiên cứu: Trong nghiên cứu này, các đường trung bình động được tính toán dựa trên dữ liệu giá đóng cửa của cổ phiếu với các kỳ phổ biến: MA-10, MA-20, MA-50.

+ Quy trình thực hiện:

* Tính toán các đường MA: Sử dụng công thức trên để tính trung bình động của giá đóng cửa trong các khoảng thời gian 10, 20, và 50 ngày.
* Trực quan hóa: Vẽ các đường MA cùng với giá đóng cửa để phân tích xu hướng và tín hiệu giao dịch

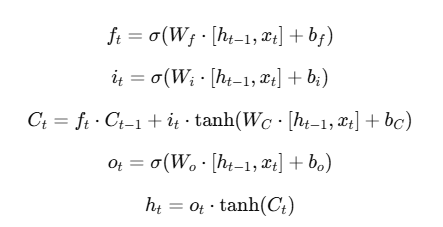
### **3.3.2 Long Short-Term Memory (LSTM)**

Long Short-Term Memory (LSTM) là một kiến trúc đặc biệt của mạng nơ-ron hồi tiếp (Recurrent Neural Network - RNN). Mạng LSTM được thiết kế để giải quyết vấn đề bóng mờ gradient trong RNN, giúp lưu trữ và xử lý thông tin dài hạn và ngắn hạn hiệu quả. Đây là lý do LSTM được sử dụng rộng rãi trong các bài toán chuỗi thời gian, như dự đoán giá cổ phiếu.

- Cấu trúc của LSTM bao gồm ba loại cổng:

* Cổng quên (Forget Gate): Quyết định loại thông tin nào cần loại bỏ khỏi trạng thái bộ nhớ.
* Cổng đầu vào (Input Gate): Xác định loại thông tin mới nào cần lưu trữ.
* Cổng đầu ra (Output Gate): Quyết định phần thông tin nào sẽ được sử dụng làm đầu ra.

- Công thức cốt lõi:



* Ở đây , , lần lượt là cổng quên, đầu vào, đầu ra, là trạng thái bộ nhớ.

- Đặc điểm nổi bật của LSTM :

* Giữ thông tin dài hạn: Nhờ cổng quên và ô nhớ , LSTM có khả năng lưu trữ và xử lý các thông tin dài hạn mà không bị mất mát.
* Giảm thiểu vấn đề vanishing gradient: Sử dụng cơ chế tính toán tuyến tính trong trạng thái ô nhớ , LSTM hạn chế được hiện tượng đạo hàm trở nên quá nhỏ, thường xảy ra trong RNN truyền thống.
* Linh hoạt trong xử lý chuỗi: LSTM không yêu cầu chiều dài chuỗi cố định, phù hợp với nhiều loại bài toán chuỗi thời gian.

- Ứng dụng:

* Xử lý dữ liệu chuỗi phức tạp, không tuyến tính.
* Kết hợp thông tin ngắn hạn và dài hạn, giúp cải thiện độ chính xác.
* Giảm thiểu độ nhiễu từ các yếu tố thị trường bất thường.

- Lợi ích khi sử dụng LSTM cho dự đoán giá cổ phiếu

* Hiệu quả với chuỗi dài hạn: LSTM khai thác tốt các mẫu trong dữ liệu chuỗi thời gian phức tạp, bao gồm cả xu hướng dài hạn và biến động ngắn hạn.
* Khả năng tổng quát hóa: Xử lý tốt các yếu tố phi tuyến trong dữ liệu cổ phiếu, vốn bị ảnh hưởng bởi nhiều yếu tố thị trường.
* Tích hợp dễ dàng: LSTM có thể được mở rộng với các kỹ thuật khác như các chỉ số phân tích kỹ thuật hoặc tích hợp vào các hệ thống dự đoán phức tạp.

- Hạn chế của LSTM :

* Yêu cầu tài nguyên lớn: Huấn luyện mô hình LSTM thường yêu cầu thời gian và tài nguyên tính toán cao.
* Độ phức tạp: Cần thiết kế cẩn thận để tránh overfitting hoặc underfitting khi áp dụng cho dữ liệu tài chính.
* Nhạy cảm với siêu tham số: Hiệu quả của mô hình phụ thuộc vào việc lựa chọn đúng số lớp, số nơ-ron, và các siêu tham số khác.

### **3.3.3 Hồi quy tuyến tính (Linear Regression)**

Linear Regression là một phương pháp thống kê và học máy cơ bản được sử dụng để mô hình hóa mối quan hệ giữa một biến phụ thuộc (biến đầu ra) và một hoặc nhiều biến độc lập (biến đầu vào). Mục tiêu của hồi quy tuyến tính là tìm ra một đường thẳng (hoặc siêu phẳng nếu có nhiều biến đầu vào) tối ưu nhất để biểu diễn mối quan hệ này.

Khái niệm và công thức cơ bản: Trong hồi quy tuyến tính đơn biến (Simple Linear Regression), mối quan hệ giữa biến đầu ra 𝑦 và biến đầu vào x được biểu diễn bởi phương trình tuyến tính:

: ệ số chặn (intercept), biểu diễn giá trị của y khi x = 0

: Hệ số góc (slope), biểu diễn mức độ thay đổi của *y* khi x thay đổi một đơn vị.

: Sai số ngẫu nhiên, thể hiện phần nhiễu không được mô hình giải thích.

* Trong trường hợp hồi quy tuyến tính đa biến (Multiple Linear Regression), công thức được mở rộng thành:

Các trọng số tương ứng với từng biến đầu vào.

: Các biến đầu vào.

Ứng dụng : Hồi quy tuyến tính được áp dụng trong phân tích và dự đoán giá cổ phiếu bằng cách mô hình hóa mối quan hệ giữa giá cổ phiếu (biến phụ thuộc) và các yếu tố khác (biến độc lập) như khối lượng giao dịch, chỉ số kinh tế, hoặc các chỉ báo kỹ thuật.

- Ưu điểm :

* Đơn giản, dễ hiểu và dễ triển khai.
* Hiệu quả với các bài toán có quan hệ tuyến tính rõ ràng giữa các biến.
* Tính toán nhanh chóng, phù hợp với dữ liệu nhỏ và ít biến đầu vào.

- Nhược điểm:

* Không hiệu quả nếu mối quan hệ giữa các biến không phải là tuyến tính.
* Dễ bị ảnh hưởng bởi các ngoại lệ (outliers).
* Không thể xử lý tốt các bài toán có dữ liệu phi tuyến hoặc tương quan phức tạp giữa các biến.

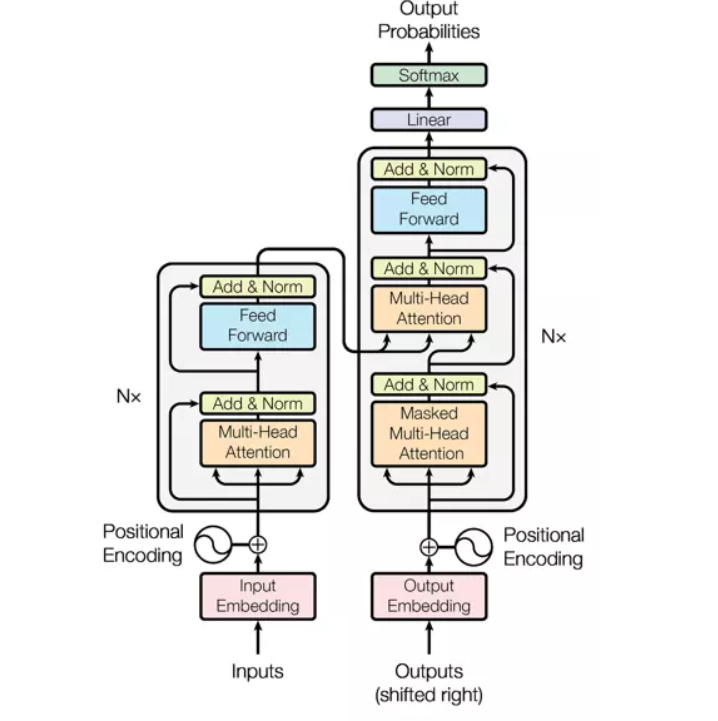
### **3.3.4. Transformer**

Transformer là một mô hình học sâu tiên tiến, được giới thiệu lần đầu tiên bởi Vaswani et al. (2017) trong bài báo nổi tiếng "Attention is All You Need". Đây là một kiến trúc đột phá trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP) và hiện nay được ứng dụng rộng rãi trong nhiều bài toán khác, bao gồm phân tích chuỗi thời gian, xử lý hình ảnh, và dự đoán tài chính. Khác với các mô hình truyền thống như RNN hay LSTM vốn gặp hạn chế trong việc xử lý chuỗi dài, Transformer sử dụng hoàn toàn cơ chế self-attention để học các mối quan hệ giữa các phần tử trong chuỗi, bất kể khoảng cách giữa chúng. Điều này giúp mô hình có khả năng xử lý hiệu quả các chuỗi dữ liệu lớn và phức tạp. Đồng thời, nhờ khả năng tính toán song song, Transformer cải thiện đáng kể tốc độ huấn luyện so với các kiến trúc tuần tự.

Kiến trúc Transformer:

- Encoder: Trích xuất thông tin từ chuỗi đầu vào, sử dụng các khối Multi-Head Attention và Feed-Forward để học các đặc trưng phức tạp.

- Decoder: Tạo ra chuỗi đầu ra dựa trên thông tin từ Encoder và cơ chế attention, phù hợp với các tác vụ sinh chuỗi như dịch ngôn ngữ.



*Ảnh 3.1. Mô hình kiến trúc của Transformer*

a. Input Embedding & Positional Encoding

Input Embedding: Mỗi dữ liệu đầu vào (giá đóng cửa, chỉ báo kỹ thuật, khối lượng giao dịch) được chuyển thành vector nhờ một embedding layer hoặc normalization.

Positional Encoding: Vì dữ liệu chuỗi thời gian có thứ tự, thông tin về vị trí thời gian được thêm vào embedding để mô hình hiểu được ngữ cảnh tuần tự của dữ liệu.

b. Encoder

Multi-Head Self-Attention: Thành phần này giúp Transformer tập trung vào các thời điểm quan trọng trong chuỗi dữ liệu. Ví dụ, khi dự đoán giá đóng cửa, mô hình có thể chú ý nhiều hơn đến các ngày trước đó khi có biến động lớn.

Feed Forward: Mạng nơ-ron fully connected xử lý các đặc trưng đã học từ cơ chế attention, giúp mô hình hiểu sâu hơn mối quan hệ phi tuyến trong dữ liệu.

Add & Norm: Chuẩn hóa giúp ổn định việc học và tăng tốc độ hội tụ của mô hình.

c. Decoder (không bắt buộc)

Nếu bạn muốn dự đoán từng bước liên tục (như nhiều ngày giá cổ phiếu trong tương lai), Decoder có thể được sử dụng để tạo ra chuỗi dự đoán. Tuy nhiên, trong trường hợp dự đoán giá đóng cửa tại một ngày cụ thể, phần Decoder có thể được đơn giản hóa hoặc bỏ qua.

d. Output Layer

Dữ liệu đầu ra là giá đóng cửa được dự đoán (một giá trị liên tục). Lớp đầu ra sử dụng activation linear để dự đoán giá trị thực.

- Cơ chế Attention trong dự đoán giá cổ phiếu:

* Self-Attention giúp Transformer tập trung vào những thời điểm quan trọng trong lịch sử giá cổ phiếu:
* Xác định mối quan hệ giữa các giá trị tại thời điểm hiện tại với các mức giá trong quá khứ.
* Phát hiện xu hướng hoặc mẫu biến động, chẳng hạn như khi giá cổ phiếu phá vỡ ngưỡng kháng cự hoặc hỗ trợ.

- Ưu điểm khi sử dụng Transformer:

* Hiệu quả với chuỗi dài: Transformer không gặp hiện tượng "quên dài hạn" như LSTM, giúp nó xử lý các chuỗi dữ liệu cổ phiếu dài.
* Tính toán song song: Tăng tốc độ huấn luyện và dự đoán.
* Khả năng học quan hệ phức tạp: Học được cả mối quan hệ phi tuyến và tuyến tính trong dữ liệu giá cổ phiếu.

# CHƯƠNG 4: THỰC NGHIỆM VÀ KẾT QUẢ

## 4.1. Thực nghiệm

### **4.1.1. Dữ liệu**

Chúng ta sẽ chia dữ liệu thành train và test để xác minh kết quả dự đoán. Vì đây là dữ liệu dạng Time series nên ta không thể chia ngẫu nhiên được. Vậy ta sẽ chia 80% đầu là train và 20% là còn lại là test.

|  |  |
| --- | --- |
| Tập dữ liệu | Số lượng |
| Train | 1910 |
| val | 430 |

Bảng 1. Bảng chia dữ liệu

### **4.1.2. Tham số khi train mô hình**

|  |  |
| --- | --- |
| Tham số | Giá trị |
| Số lượng epoc | 20 |
| Batch-size | 8 |

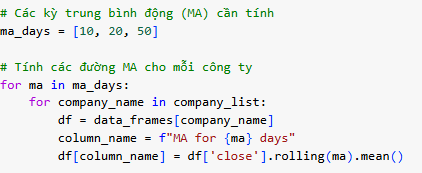
Bảng 2. Bảng thông số khi train mô hình

## 4.2. Thực nghiệm

### **4.2.1. Đường trung bình động (Moving Average)**

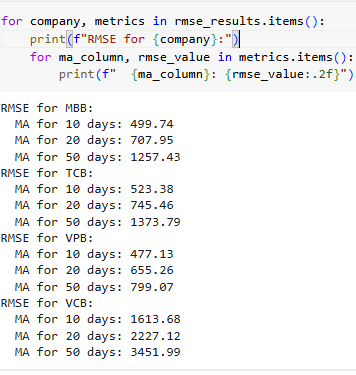
**Trích đặc trưng:**

Lấy dữ liệu giá cổ phiếu lịch sử và tính toán các chỉ báo Moving Averages (MA) cho các kỳ 10, 20, và 50 ngày cho từng công ty. Dữ liệu này là đầu vào cho mô hình dự đoán.



Tính toán các giá trị MA cho mỗi cổ phiếu, giúp tạo các đặc trưng (features) cho mô hình.

**Tính chỉ số RSME cho mô hình:**



Chỉ số RMSE theo trung bình động của giá đóng của trong 10, 20, 50 ngày của MBB, TCB, VPB, VCB.

**Trực quan hóa giá trị dữ liệu :**



*Ảnh 4.1. Biểu đồ trực quan kết quả dự đoán của đường trung bình động*

### **4.2.2 Stock prediction with Linear Regression:**

Mô hình này áp dụng cách tiếp cận tuyến tính để mô hình hóa mối quan hệ giữa biến phụ thuộc và (các) biến độc lập, đồng thời cũng là hình thức học máy đơn giản nhất.

**Trích chọn đặc trưng :**

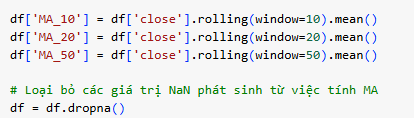
- Tính toán các chỉ số trung bình động (Moving Averages): Để cung cấp các đặc trưng có ý nghĩa cho mô hình, chúng tôi tính toán ba loại trung bình động của giá đóng cửa (Close Price), bao gồm:

MA\_10: Trung bình động của giá đóng cửa trong 10 ngày.

MA\_20: Trung bình động của giá đóng cửa trong 20 ngày.

MA\_50: Trung bình động của giá đóng cửa trong 50 ngày.

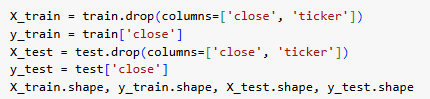
Các chỉ số này được sử dụng để phản ánh xu hướng ngắn hạn, trung hạn và dài hạn của giá cổ phiếu, từ đó giúp mô hình có cái nhìn tổng quan hơn về biến động giá.



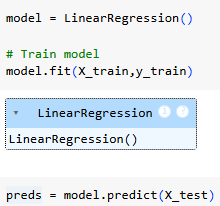
**Split data thành tập train và test** : chúng ta chọn 80%, 20% cho train và test của bộ dữ liệu:



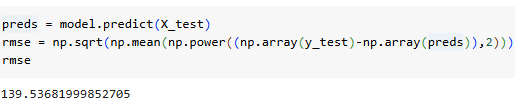
**Tạo train test mô hình hồi qui tuyến tính:**



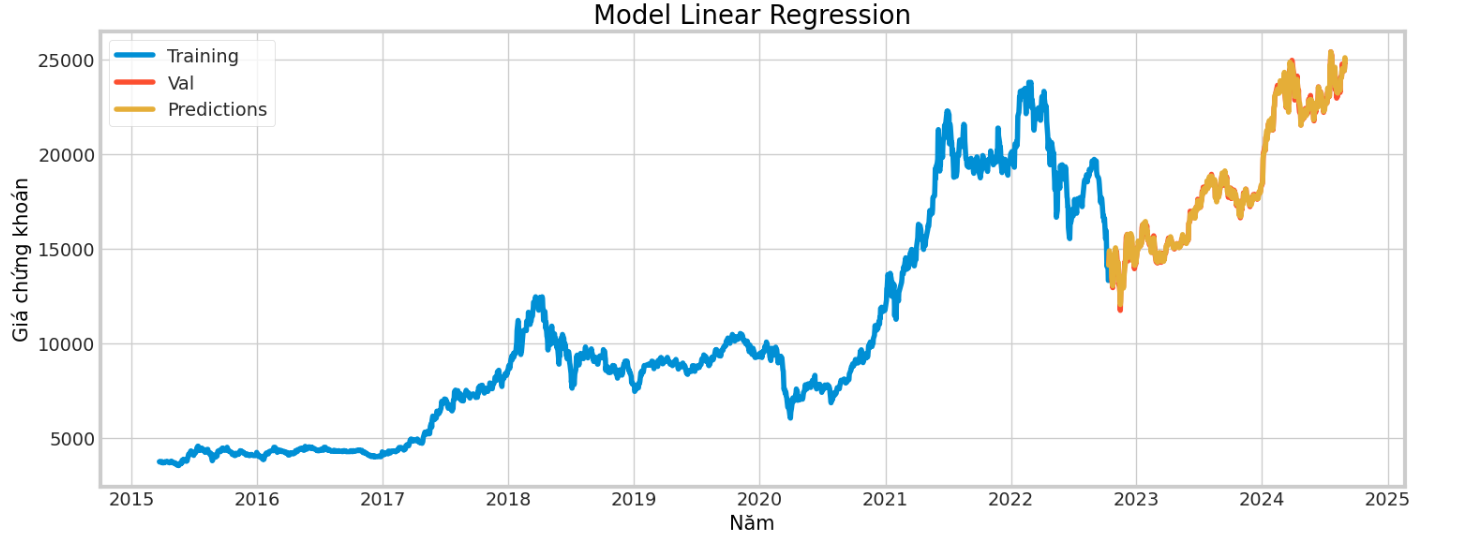
**Train model:**



**Tính chỉ số RSME cho mô hình:**



Chỉ số RSME của mô hình hồi qui tuyến tính là 139.53

**Trực quan hoá dữ liệu dự đoán và thực tế:** 

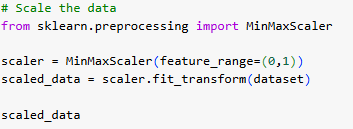
Ảnh 4.3. Ảnh trực quan kết quả dự đoán của mô hình Linear Regression

### **4.2.3 Long Short-Term memory (LSTM)**

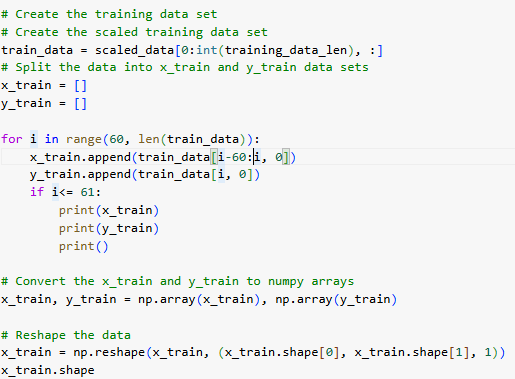
LSTM là một Recurrent Neural Network (tạm dịch: Mạng thần kinh tái phát) có thể lưu trữ những thông tin quan trọng trong quá khứ và quên đi những thông tin không quan trọng. Nó sử dụng các quy trình bộ nhớ ngắn hạn để tạo ra bộ nhớ dài hơn và giới thiệu khái niệm về cổng để kiểm soát luồng thông tin trong mạng lưới bằng cách có cơ chế cổng đầu vào, đầu ra và cổng quên.

**Split data thành tập train và test** : chúng ta chọn 80%, 20% cho train và test của bộ dữ liệu

**Feature scaling :** Dòng mã này tạo ra một đối tượng MinMaxScaler với phạm vi chuẩn hóa từ 0 đến 1, tức là tất cả giá trị trong dữ liệu sẽ được chuyển đổi sao cho nằm trong khoảng từ 0 đến 1.



**Tạo dữ liệu training :**



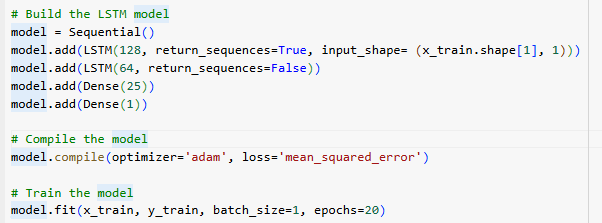
**Tạo và train model**

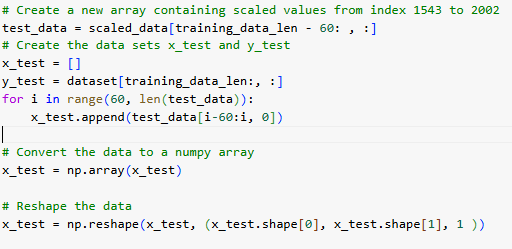
Khởi tạo mô hình: Sử dụng Sequential từ Keras, mô hình được xây dựng theo dạng tuyến tính, trong đó các lớp được thêm vào theo thứ tự.

Các lớp LSTM:

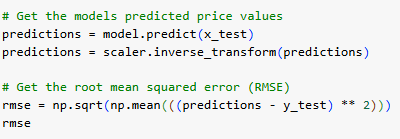
* Lớp LSTM đầu tiên có 128 đơn vị và trả về chuỗi (sequence) để chuyển tiếp vào lớp LSTM thứ hai. Lớp LSTM thứ hai có 64 đơn vị và trả về một giá trị duy nhất. Các lớp Dense:
* Lớp Dense thứ nhất có 25 đơn vị, học các mối quan hệ phức tạp. Lớp Dense cuối cùng có 1 đơn vị, dùng để dự đoán giá cổ phiếu. Biên dịch mô hình: Sử dụng Adam làm thuật toán tối ưu và mean\_squared\_error làm hàm mất mát.

Huấn luyện mô hình: Mô hình được huấn luyện với dữ liệu x\_train và y\_train, sử dụng batch size = 1 và 20 epochs.

**Tạo train và test**



**Tính chỉ số RSME**



Chỉ số RMSE của mô hình Long Short-Term Memory trả về giá trị là 390.17 (một con số thấp vượt trội cho thấy mức độ hiệu quả được cải thiện đáng kể của mô hình so với những mô hình dự đoán trước).

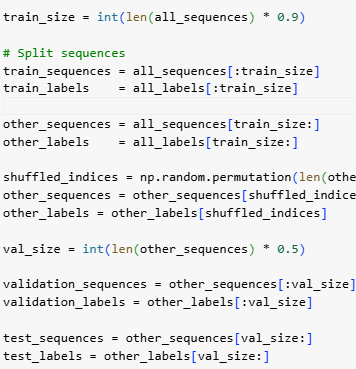
**Trực quan hóa dự đoán :**

*Ảnh 4.4. Ảnh trực quan dự liệu dự đoán của mô hình LSTM*

### **4.2.4 Tranformer**

Transformer là một kiến trúc mạng nơ-ron được thiết kế để xử lý chuỗi dữ liệu, nổi bật với cơ chế self-attention giúp mô hình học được mối quan hệ giữa các phần trong chuỗi, bất kể khoảng cách của chúng. Trong bài toán dự đoán giá đóng cửa (Close Price), Transformer có thể được áp dụng hiệu quả để phân tích chuỗi thời gian tài chính, nhờ khả năng nắm bắt cả các mối quan hệ ngắn hạn và dài hạn.

**Split data thành tập train và test** : chúng ta chọn 90%, 5%, 5% cho train và test, val của bộ dữ liệu



**Feature Engineering** : Đặc trưng của mô hình được tính toán thông qua các hàm:

1. **Bollinger Bands:**

* Độ rộng Bollinger phản ánh mức độ biến động (volatility) của giá trong một khoảng thời gian.
* Khi độ rộng lớn: Giá biến động mạnh (cơ hội giao dịch cao).
* Khi độ rộng nhỏ: Giá ít biến động (thị trường yên tĩnh).

1. **RSI**



* Đo lường sức mạnh tương đối của giá (momentum).
* RSI > 70: Quá mua (overbought) → Có thể giảm.
* RSI < 30: Quá bán (oversold) → Có thể tăng.

1. **ROC**



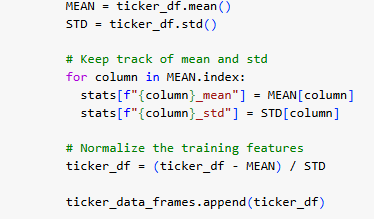
* Phản ánh tốc độ thay đổi giá trong khoảng thời gian cụ thể.
* ROC dương: Giá đang tăng.
* ROC âm: Giá đang giảm.
* Dùng để nhận diện xu hướng (trend) và đảo chiều (reversal).

1. **Volume, Diff, và Percent Change:**



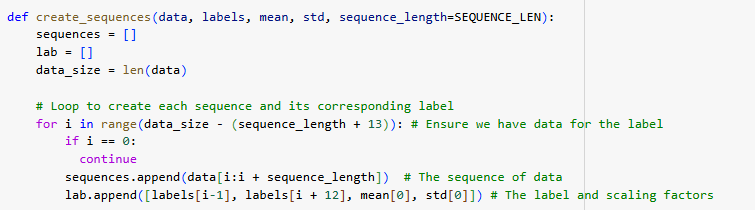
* Xác nhận xu hướng thị trường, đo lường sự tham gia của nhà đầu tư.
* Phân tích mức độ dao động giá ngắn hạn qua từng phiên.
* Chuẩn hóa biến động giá theo phần trăm, dễ so sánh giữa các cổ phiếu

**Xử lý dữ liệu :**



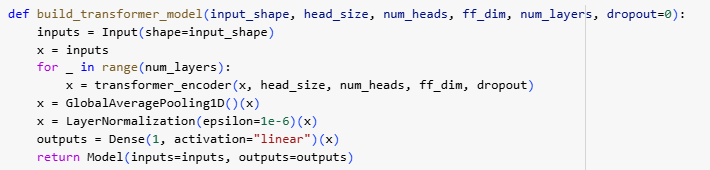
* Chuẩn hóa (normalize): Đưa dữ liệu về dạng có phân phối chuẩn để dễ học.
* Xử lý giá trị thiếu (dropna): Loại bỏ các giá trị không hợp lệ để tránh sai sót.

**Tạo dữ liệu training :**



* Tạo chuỗi dữ liệu sequence\_length (ví dụ: 24) từ các đặc trưng để mô hình học.

**Build tranfomer :**



* Hàm build\_transformer\_model xây dựng một mô hình Transformer cho các bài toán học sâu với dữ liệu chuỗi, như chuỗi thời gian. Mô hình này sử dụng các lớp Transformer Encoder với cơ chế multi-head attention để học các mối quan hệ trong dữ liệu, kết hợp với Global Average Pooling để tóm tắt thông tin và Layer Normalization để ổn định huấn luyện. Cuối cùng, mô hình sử dụng một lớp Dense với hàm kích hoạt linear để dự đoán giá trị liên tục, phù hợp với các bài toán hồi quy.

**Tính chỉ số RSME :**

Chỉ số RSME của mô hình tranfomer là 2626.68, điều đó cho thấy độ chính xác của mô hình tranfomer xác suất dự đoán giá đóng cửa đúng của cổ phiếu có tỉ lệ chính xác thâp hơn LSTM, Hồi qui tuyến tính.

# CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## 5.1. Kết luận

Bảng so sánh kết quả của các mô hình:

|  |  |
| --- | --- |
| Mô hình | RMSE |
| Dự báo trung bình động (Forecasting moving average) | 3544.87 |
| Hồi qui tuyến tính | 139.53 |
| Long short-term memory (LSTM) | 390.17 |
| Tranfomer | 2626.68 |

Bảng 3 : Bảng kết quả mô hình dự đoán

Trong nghiên cứu này, chúng tôi đã tiến hành xây dựng và so sánh hiệu quả của ba mô hình dự báo gồm: Hồi quy tuyến tính, Long Short-Term Memory (LSTM) và Transformer, dựa trên dữ liệu thời gian thực tế.

Kết quả thực nghiệm cho thấy, Đường trung bình động với chỉ số RMSE = 3544.87 ,mô hình Hồi quy tuyến tính đạt được hiệu suất tốt nhất với chỉ số RMSE = 139.54, tiếp theo là mô hình LSTM với RMSE = 390.17, và cuối cùng là mô hình Transformer có RMSE = 2626.68. Điều này cho thấy các mô hình phức tạp hơn như LSTM và Transformer không hoạt động hiệu quả trong bài toán này so với mô hình đơn giản hơn là Hồi quy tuyến tính.

## 5.2. Phân tích

Hiệu quả của Hồi quy tuyến tính: Mặc dù là mô hình đơn giản, Hồi quy tuyến tính cho thấy khả năng dự báo tốt nhất trong bài toán này, có thể do dữ liệu có mối quan hệ tuyến tính rõ ràng và không quá phức tạp. Điều này chứng minh rằng không phải lúc nào các mô hình phức tạp hơn cũng mang lại hiệu quả tốt hơn.

Hiệu quả của LSTM: LSTM, mặc dù được thiết kế để xử lý dữ liệu chuỗi thời gian và có khả năng nắm bắt thông tin về phụ thuộc dài hạn, vẫn không vượt qua được mô hình hồi quy tuyến tính. Nguyên nhân có thể do sự phức tạp của mô hình, hoặc do dữ liệu không có đủ đặc trưng phi tuyến tính để mô hình phát huy hiệu quả.

Hiệu quả của Transformer: Kết quả của Transformer với RMSE cao nhất cho thấy mô hình này không phù hợp với dữ liệu trong nghiên cứu. Transformer thường yêu cầu lượng dữ liệu lớn và phức tạp để đạt hiệu quả cao, nhưng trong trường hợp này, dữ liệu có thể chưa đáp ứng được yêu cầu của mô hình.

## 5.3. Hạn chế và hướng phát triển

Hạn chế :

* Dữ liệu huấn luyện có thể chưa đủ lớn hoặc chưa chứa đựng đủ thông tin phức tạp để các mô hình như LSTM hay Transformer thể hiện ưu thế.
* Việc lựa chọn siêu tham số và cấu trúc mô hình có thể chưa được tối ưu hóa triệt để, đặc biệt đối với các mô hình học sâu.

Hướng phát triển :

* Tăng cường dữ liệu: Thử nghiệm trên các bộ dữ liệu lớn hơn hoặc có nhiều đặc trưng hơn (ví dụ: thêm các yếu tố kinh tế vĩ mô hoặc dữ liệu liên ngành).
* Tối ưu hóa mô hình: Tinh chỉnh các siêu tham số hoặc thử nghiệm các biến thể của LSTM và Transformer, chẳng hạn như BiLSTM hoặc các Transformer cải tiến.
* Kết hợp mô hình: Xây dựng mô hình lai kết hợp giữa các phương pháp tuyến tính và phi tuyến tính để khai thác tối đa hiệu quả của từng loại mô hình.

==> Kết luận: mặc dù mô hình Hồi quy tuyến tính đạt hiệu quả cao nhất trong nghiên cứu này, tiềm năng cải thiện của các mô hình phức tạp như LSTM và Transformer vẫn rất lớn nếu được áp dụng với các dữ liệu và thiết lập phù hợp hơn trong tương lai.

# TÀI LIỆU THAM KHẢO

1. Security Analysis - Benjamin Graham và David Dodd (1934)
2. The Cross-Section of Expected Stock Returns - Fama và French (1992)
3. Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management - Ludwig B. Chincarini, Daehwan Kim (2006).
4. Machine Learning for Asset Managers - Marcos López de Prado (2020).
5. Technical Analysis of the Financial Markets - John J. Murphy (1999).
6. Advances in Financial Machine Learning - Marcos López de Prado (2018).
7. Phân tích đầu tư chứng khoán - Nguyễn Minh Kiều (2021).
8. A Machine Learning Approach for Stock Price Prediction Using Fundamental and Technical Analysis - J. Patel, S. Shah, P. Thakkar, K. Kotecha (2015).
9. Forecasting Stock Prices Using LSTM Neural Networks - S. Fischer et al. (2018).
10. Vietnam Stock Market: A Review of the Literature and Future Research Directions - Trần Đức Thảo (2022).
11. Applying Deep Learning Models for Predicting Vietnam Stock Market Indices - Lê Quang Huy (2021).
12. Economic Integration and Its Impact on Vietnam Stock Market Volatility - Phạm Thị Hồng (2020).

# KẾ HOẠCH THỰC HIỆN

|  |  |  |  |
| --- | --- | --- | --- |
| **Tuần** | **Từ** | **Đến** | **Nội dung công việc** |
| **1** | 09/09/2024 | 13/09/2024 | Tìm hiểu chủ đề và chuẩn bị môi trường làm việc. |
| **2** | 16/09/2024 | 20/09/2024 | Nghiên cứu đề tài và dữ liệu. |
| **3** | 23/09/2024 | 27/09/2024 | Thực hiện phân tích cơ bản (Fundamental Analysis). |
| **4** | 30/09/2024 | 04/10/2024 | Thực hiện phân tích kỹ thuật (Technical Analysis). |
| **5** | 07/10/2024 | 11/10/2024 | Xây dựng mô hình Long Short Term Memory (LSTM). |
| **6** | 14/10/2024 | 18/10/2024 | Đánh giá và tối ưu hóa mô hình Long Short Term Memory. |
| **7** | 21/10/2024 | 25/10/2024 | Xây dựng mô hình Linear Regression Transformer. |
| **8** | 28/10/2024 | 01/11/2024 | Đánh giá và tối ưu hóa mô hình Linear Regression và Transformer. |
| **9** | 04/11/2024 | 08/11/2024 | So sánh các mô hình và lựa chọn mô hình phù hợp. |
| **10** | 11/11/2024 | 15/11/2024 | Trực quan hoá kết quả dự đoán trên mô hình phù hợp nhất. |
| **11** | 18/11/2024 | 22/11/2024 | Viết báo cáo kết quả thực tập. |
| **12** | 25/11/2024 | 29/11/2024 | Hoàn thiện báo cáo và chuẩn bị cho buổi thuyết trình kết quả. |