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



**KHOA CÔNG NGHỆ THÔNG TIN**



**KHÓA LUẬN TỐT NGHIỆP**

**PHÂN TÍCH VÀ DỰ ĐOÁN GIÁ CỔ PHIẾU DỰA TRÊN CHUỖI THỜI GIAN BẰNG CÁCH SỬ DỤNG MÔ HÌNH SARIMAX VÀ HỌC MÁY**

**Chuyên ngành: Khoa học dữ liệu**

**Châu Mỹ Uyên – 20087481**

**Đặng Quốc Toàn – 20051051**

**Giảng viên hướng dẫn 1: TS. Nguyễn Chí Kiên**

**Giảng viên hướng dẫn 2: TS. Vũ Đức Thịnh**

*TP. Hồ Chí Minh, tháng 12 năm 2024*

**INDUSTRIAL UNIVERSITY OF HO CHI MINH CITY**



**FACULTY OF INFORMATION TECHNOLOGY**

****

**GRADUATION THESIS**

**ANALYSIS AND FORECASTING OF STOCK PRICES BASED ON TIME SERIES USING SARIMAX AND MACHINE LEARNING MODELS**

**Major: Data Science**

**Chau My Uyen – 20087481**

**Dang Quoc Toan – 20051051**

**Instructor: PhD. Nguyen Chi Kien**

*Ho Chi Minh City, December 2024*

**ANALYSIS AND FORECASTING OF STOCK PRICES BASED ON TIME SERIES USING SARIMAX AND MACHINE LEARNING MODELS**

**Abstract**

With the continuous growth of the global economy, stock price prediction has garnered significant attention, becoming a focal point and widely discussed topic in the financial sector. This growing interest stems from the increasing demand among investors to leverage advanced tools for making effective decisions in highly volatile stock markets. In this study, we propose using the SARIMAX time series forecasting model, alongside popular machine learning models such as LSTM, Random Forest, and XGBoost, as well as a hybrid SARIMAX + XGBOOTS model, to predict stock prices. We then compare their performance and accuracy.

The research focuses on building a comprehensive dataset for predicting TESLA's stock prices by integrating three key factors: technical, fundamental, and sentiment analysis. Data from these three analytical approaches are combined with stock price data from Yahoo Finance to create a multidimensional dataset for predictive modeling. This dataset offers diverse insights, encompassing technical indicators, fundamental factors, and market sentiment, thereby enhancing the accuracy and efficiency of stock price predictions.

The results of the study underscore the superior performance of the SARIMAX model and the hybrid SARIMAX + XGBOOTS model in forecasting the next day’s closing prices of TESLA's stock. The SARIMAX model achieved RMSE, R², MAE, and MAPE values of (2.7048, 0.9989, 2.0341, 0.0049%), while the SARIMAX + XGBOOTS model achieved values of (3.3067, 0.9984, 2.6825, 0.0065%). Compared to other approaches such as LSTM, Random Forest, and XGBoost, SARIMAX and SARIMAX + XGBOOTS have proven to be more suitable for this task, delivering the highest levels of accuracy and reliability. The predictive accuracy of these two models provides investors with critical insights, helping them mitigate risks and develop effective investment strategies.

In conclusion, this study reaffirms the predictive power of the SARIMAX model and the hybrid SARIMAX + XGBOOTS model compared to widely used machine learning techniques for stock price forecasting. It makes a significant contribution to the financial field by equipping stock investors with intelligent decision-making tools, reducing investment risks, and achieving optimal returns in dynamic stock market environments.

**PHÂN TÍCH VÀ DỰ ĐOÁN GIÁ CỔ PHIẾU DỰA TRÊN CHUỖI THỜI GIAN BẰNG CÁCH SỬ DỤNG MÔ HÌNH SARIMAX VÀ HỌC MÁY**

**Tóm tắt**

Với sự phát triển không ngừng của nền kinh tế toàn cầu, đặc biệt tại các thị trường mới nổi như Việt Nam, việc dự đoán giá cổ phiếu đã thu hút sự chú ý đáng kể, trở thành tâm điểm và chủ đề thảo luận rộng rãi trong lĩnh vực tài chính. Sự quan tâm ngày càng tăng này xuất phát từ nhu cầu ngày càng lớn của các nhà đầu tư trong việc tận dụng các công cụ tiên tiến để đưa ra quyết định hiệu quả trên các thị trường chứng khoán có độ biến động cao. Trong nghiên cứu này, chúng tôi đề xuất sử dụng mô hình dự báo chuỗi thời gian SARIMAX, cùng với các mô hình học máy phổ biến như LSTM, Random Forest, XGBoost và mô hình kết hợp SARIMAX + XGBOOTS, để dự đoán giá cổ phiếu. Sau đó, chúng tôi so sánh hiệu suất và độ chính xác của chúng.

Nghiên cứu tập trung vào việc xây dựng một bộ dữ liệu toàn diện để dự đoán giá cổ phiếu TESLA bằng cách tích hợp ba yếu tố chính: phân tích kỹ thuật (technical), phân tích cơ bản (fundamental), và phân tích cảm xúc (sentiment). Dữ liệu từ ba phương pháp phân tích này được kết hợp với dữ liệu giá cổ phiếu từ Yahoo Finance để tạo nên một bộ dữ liệu đa chiều phục vụ cho mô hình dự đoán. Bộ dữ liệu này mang lại những góc nhìn đa dạng, từ chỉ số kỹ thuật, yếu tố cơ bản đến cảm xúc của thị trường, qua đó cải thiện độ chính xác và hiệu quả trong dự đoán giá cổ phiếu.

Kết quả nghiên cứu nhấn mạnh hiệu suất vượt trội của mô hình SARIMAX và mô hình kết hợp SARIMAX + XGBOOTS trong việc dự đoán giá đóng cửa của cổ phiếu TESLA vào ngày tiếp theo. Mô hình SARIMAX đạt các giá trị RMSE, R², MAE và MAPE lần lượt là (2.7048, 0.9989, 2.0341, 0.0049%), trong khi mô hình SARIMAX + XGBOOTS đạt các giá trị tương ứng là (2.7052, 0.9989, 2.0348, 0.0049%). So với các phương pháp khác như LSTM, Random Forest và XGBoost, hai mô hình này đã chứng minh tính hiệu quả vượt trội, đảm bảo độ chính xác và độ tin cậy cao hơn trong dự đoán. Kết quả này cung cấp những thông tin quan trọng cho các nhà đầu tư, hỗ trợ họ giảm thiểu rủi ro và phát triển chiến lược đầu tư một cách hiệu quả.

Tóm lại, nghiên cứu này khẳng định lại sức mạnh dự đoán của mô hình SARIMAX và mô hình kết hợp SARIMAX + XGBOOTS so với các kỹ thuật học máy được sử dụng rộng rãi để dự đoán giá cổ phiếu. Nghiên cứu đóng góp đáng kể vào lĩnh vực tài chính bằng cách trang bị cho các nhà đầu tư cổ phiếu những công cụ ra quyết định thông minh, giảm thiểu rủi ro đầu tư và đạt được lợi nhuận tối ưu trong môi trường thị trường chứng khoán năng động.

**LỜI CẢM ƠN**

Khóa luận tốt nghiệp là một cột mốc đánh dấu vô cùng quan trọng để kết thúc khoảng thời gian học tập và rèn luyện của mỗi sinh viên chúng em. Quá trình thực hiện luận văn tốt nghiệp này là quá trình để chúng em có thể nhớ lại và áp dụng những kiến thức mà thầy cô đã dạy, đồng thời lấy nó làm tiền đề để phát triển trong tương lai.

Chúng em xin gửi lời cảm ơn chân thành đến TS. Nguyễn Chí Kiên, . Chúng em biết thầy là một người vô cùng bận rộn với công việc của mình nhưng thầy đã luôn cố gắng hỗ trợ hết mình cho luận án của chúng em nói riêng và Khóa 16 chúng em nói chung. Nhờ có thầy mà chúng em có nhiều kinh nghiệm tích lũy và có thêm nhiều lòng tin để thực hiện luận án của mình.

Ngoài ra, chúng em muốn gửi lời cảm ơn đến hầy Trương Vĩnh Linh – một người thầy chủ nhiệm đã tiếp một niềm tin vô to lớn đến tất cả chúng em. Thầy luôn đồng hành với chúng em như một người bạn, thầy lo lắng, tận tụy, giúp đỡ chúng em rất nhiều trong học tập và cuộc sống, giúp chúng em có thể vững bước tiến về tương lai phía trước.

Đồng thời, chúng em xin cảm ơn tất cả Thầy, Cô Khoa Công nghệ thông tin và đặc biệt là các Thầy, Cô ngành Khoa học dữ liệu đã hỗ trợ, giúp đỡ chúng em trong suốt quá trình học tập để hoàn thành tốt các môn học và giúp chúng em học được nhiều kỹ năng quý báu.

Cuối cùng, chúng em xin cảm ơn những người bạn đã đồng hành với chúng em trong suốt quá trình học tập của mình, nhờ các bạn mà chúng em đã có rất nhiều niềm vui và kỷ niệm đáng nhớ trong suốt những năm đại học này. Một lần nữa, chúng em xin gửi lời cảm ơn chân thành và một niềm hi vọng về tương lai tốt đẹp đến với tất cả mọi người.

**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 11](#_Toc184843876)

[DANH MỤC BẢNG BIỂU 12](#_Toc184843877)

[DANH MỤC TỪ VIẾT TẮT 13](#_Toc184843878)

[CHƯƠNG 1. GIỚI THIỆU 1](#_Toc184843879)

[1.1. Tổng quan 1](#_Toc184843880)

[**1.1.1. Bối cảnh** 1](#_Toc184843881)

[**1.1.2. Lý do chọn đề tài** 2](#_Toc184843882)

[1.2. Mục tiêu nghiên cứu 2](#_Toc184843883)

[1.3. Phạm vi nghiên cứu 3](#_Toc184843884)

[1.4. Ý nghĩa khoa học và thực tiễn 3](#_Toc184843885)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 5](#_Toc184843886)

[2.1. Học máy 5](#_Toc184843887)

[2.2. Phân tích chuỗi thời gian 6](#_Toc184843888)

[2.3. Bài toán dự đoán giá cổ phiếu của TESLA bằng các mô hình chuỗi thời gian và học máy 7](#_Toc184843889)

[**2.3.1. Tổng quan** 7](#_Toc184843890)

[**2.3.2. Các nghiên cứu liên quan** 8](#_Toc184843891)

[CHƯƠNG 3: MÔ HÌNH ĐỀ XUẤT 11](#_Toc184843892)

[3.1. Tổng quan mô hình đề xuất 11](#_Toc184843893)

[3.2. Đặc trưng của mô hình đề xuất 12](#_Toc184843894)

[3.2.1. SARIMAX 12](#_Toc184843895)

[3.2.2. XGBOOST 14](#_Toc184843896)

[3.2.3. Random Forest 15](#_Toc184843897)

[3.3.4. LSTM 17](#_Toc184843898)

[3.3. Phương pháp đánh giá 18](#_Toc184843899)

[CHƯƠNG 4: THỰC NGHIỆM VÀ KẾT QUẢ 21](#_Toc184843900)

[4.1. Dữ liệu 21](#_Toc184843901)

[Bộ dữ liệu được sử dụng trong phân tích bao gồm ba nguồn thông tin chính, nhằm cung cấp góc nhìn toàn diện về hoạt động của Tesla từ ngày 01/01/2020 đến ngày 01/08/2024: 21](#_Toc184843902)

[Dữ liệu giá chứng khoán Tesla 21](#_Toc184843903)

[Dữ liệu này được thu thập từ Yahoo Finance (YF), cung cấp thông tin chi tiết về biến động giá cổ phiếu Tesla trong suốt giai đoạn nghiên cứu. Cụ thể, các biến trong bộ dữ liệu bao gồm: 21](#_Toc184843904)

[ Ngày giao dịch (Date): Xác định thời điểm diễn ra giao dịch. 21](#_Toc184843905)

[ Giá mở cửa (Open): Giá cổ phiếu khi bắt đầu phiên giao dịch. 21](#_Toc184843906)

[ Giá cao nhất (High) và giá thấp nhất (Low): Khoảng giá dao động trong ngày. 21](#_Toc184843907)

[ Giá đóng cửa (Close) và giá đóng cửa điều chỉnh (Adj Close): Giá cổ phiếu khi kết thúc phiên giao dịch, bao gồm phiên bản đã điều chỉnh cho các yếu tố như chia cổ tức hoặc tách cổ phiếu. 21](#_Toc184843908)

[ Khối lượng giao dịch (Volume): Tổng số cổ phiếu được giao dịch trong ngày. 21](#_Toc184843909)

[**4.1.1. Thu thập dữ liệu** 21](#_Toc184843910)

[**4.1.2. Tổng quan dữ liệu** 22](#_Toc184843911)

[**4.1.3. Khai phá và phân tích dữ liệu (EDA)** 22](#_Toc184843912)

[**4.1.3. Tiền xử lí dữ liệu** 22](#_Toc184843913)

[4.2. Thực nghiệm và đánh giá mô hình 22](#_Toc184843914)

[4.2.1. Thực nghiệm với các tham số mặc định 22](#_Toc184843915)

[4.2.2. Tinh chỉnh siêu tham số 22](#_Toc184843916)

[4.2.3. Tổng kết 22](#_Toc184843917)

[CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 23](#_Toc184843918)

[5.1. Kết luận 23](#_Toc184843919)

[5.2. Hạn chế 23](#_Toc184843920)

[5.3. Kiến thức và kỹ năng 23](#_Toc184843921)

[5.4. Hướng phát triển trong tương lai 23](#_Toc184843922)

[TÀI LIỆU THAM KHẢO 25](#_Toc184843923)

[KẾ HOẠCH THỰC HIỆN 27](#_Toc184843924)

[NHẬT KÝ LÀM VIỆC 28](#_Toc184843925)

# DANH MỤC HÌNH ẢNH

[Hình 1. Các nhiệm vụ trong lĩnh vực Computer Vision 4](#_Toc135438731)

c

# DANH MỤC BẢNG BIỂU

[Bảng 1. Ví dụ bảng 1 4](#_Toc179819583)

[Bảng 2. Ví dụ bảng 2 4](#_Toc179819584)

# DANH MỤC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **TỪ VIẾT TẮT** | **TỪ ĐẦY ĐỦ** | **Ý NGHĨA** |
| CNN | Convolutional Neural Network | Mạng nơ-ron tích chập |
| FCNN | Fully connected neural network | Mạng nơ-ron tích chập đầy đủ |
| AR | AutoRegressive | Tự hồi quy |
| MA | Moving Average | Trung bình động |
| I | Integrated | Tích hợp |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# CHƯƠNG 1. GIỚI THIỆU

## 1.1. Tổng quan

### **1.1.1. Bối cảnh**

Hiện nay, đầu tư chứng khoán ngày càng trở nên phổ biến và nhận được sự quan tâm rộng rãi, trở thành một kênh đầu tư hấp dẫn với tiềm năng sinh lời cao. Tuy nhiên, lợi nhuận lớn luôn đi kèm với những rủi ro tiềm ẩn. Các sự kiện toàn cầu như đại dịch COVID-19 và cuộc xung đột Nga - Ukraine đã gây ra những biến động nghiêm trọng, ảnh hưởng tiêu cực đến tính thanh khoản và sự ổn định của thị trường chứng khoán.

Nghiên cứu của Hatmanu và cộng sự (2021) [1] chỉ ra rằng, đại dịch COVID-19 đã đẩy thị trường chứng khoán toàn cầu vào tình trạng bất ổn lớn, gây ra các giai đoạn biến động mạnh mẽ. Baig và cộng sự (2021) [2] nhận định rằng, số ca nhiễm và tử vong do COVID-19, tâm lý tiêu cực từ các tin tức, cùng các quy định hạn chế di chuyển của chính phủ đã ảnh hưởng tiêu cực đến thanh khoản và mức độ biến động của thị trường, đặc biệt ở cấp độ cổ phiếu riêng lẻ.

Tương tự, vào ngày 24/02/2022, khi Nga phát động cuộc tấn công vào Ukraine, nghiên cứu của Boungou và cộng sự (2022) [3] đã chứng minh rằng sự kiện này gây tác động tiêu cực đến các chỉ số chứng khoán toàn cầu, và mức độ nghiêm trọng càng gia tăng sau cuộc xâm lược.

Ngoài các yếu tố trên, giá cổ phiếu còn chịu ảnh hưởng bởi nhiều yếu tố khác như tình hình kinh tế vĩ mô, hiệu quả hoạt động của doanh nghiệp, tâm lý nhà đầu tư và giá hàng hóa toàn cầu.

Trước những tác động này, việc dự đoán giá cổ phiếu ngày càng trở thành một chủ đề được quan tâm sâu sắc trong giới tài chính, đặc biệt với các nhà đầu tư. Những dự đoán chính xác không chỉ giúp nhà đầu tư hiểu rõ xu hướng thị trường mà còn hỗ trợ đưa ra các quyết định mua bán hợp lý, tối ưu hóa cơ hội sinh lời và giảm thiểu rủi ro hiệu quả.

### **1.1.2. Lý do chọn đề tài**

Thị trường chứng khoán đóng vai trò quan trọng trong nền kinh tế toàn cầu, và việc theo dõi kịp thời các xu hướng thị trường là vô cùng cần thiết. Đầu tư vào cổ phiếu không chỉ giúp phát triển doanh nghiệp trong nước mà còn thu hút vốn đầu tư nước ngoài, góp phần nâng cao ngân sách Nhà nước. Tuy nhiên, sự biến động khó lường của giá cổ phiếu là một thách thức lớn đối với các nhà đầu tư.

Vì vậy, nghiên cứu này sử dụng các mô hình phổ biến như LSTM, Random Forest, SARIMAX, XGBoost và kết hợp SARIMAX + XGBOOTS để dự đoán giá cổ phiếu. Những mô hình này được chọn vì khả năng nắm bắt mối quan hệ dài hạn, phi tuyến tính và yếu tố mùa vụ, đồng thời có thể xử lý các dữ liệu phức tạp. Mục tiêu của nghiên cứu là giúp giảm thiểu rủi ro và tối ưu hóa hiệu quả đầu tư. Với sự phát triển mạnh mẽ của công nghệ thông tin trong tài chính, nghiên cứu hy vọng sẽ phát triển một công cụ hỗ trợ các nhà đầu tư đưa ra quyết định thông minh, giảm thiểu rủi ro và tối đa hóa lợi nhuận.

## 1.2. Mục tiêu nghiên cứu

Nghiên cứu được thực hiện với mong muốn hỗ trợ các nhà đầu xác định thời điểm để mua, bán và giữ cổ phiếu một cách hợp lí, đưa ra quyết định dựa trên xu hướng giá cổ phiếu, từ đó tối ưu hóa lợi nhuận và giảm thiểu rủi ro đầu tư.

Đòng thời, nghiên cứu còn nhằm so sánh và đánh giá hiệu suất mô hình dự báo chuỗi thời gian SARIMAX với các mô hình học máy như LSTM, Random Forest, XGBoost và mô hình kết hợp SARIMAX + XGBOOTS trong việc dự đoán giá cổ phiếu, cụ thể là giá cổ phiếu của TESLA. Mục tiêu là đánh giá xem mô hình nào mang lại kết quả dự đoán chính xác và phù hợp nhất với dữ liệu thực tế.

Ngoài ra, mục tiêu của bài nghiên cứu này là tìm hiểu và phân tích những yếu tố ngoại sinh làm ảnh hưởng đến kết quả dự đoán, đánh giá tác động của các yếu tố phân tích cơ bản, kỹ thuật và tâm lý đến giá cổ phiếu. từ đó cải thiện mô hình và đưa ra được kết quả dự đoán chính xác nhất.

## 1.3. Phạm vi nghiên cứu

Nghiên cứu này tập trung vào việc dự đoán giá cổ phiếu TESLA (TSLA) bằng cách sử dụng các mô hình chuỗi thời gian và học máy. Dữ liệu được thu thập và phân tích trong khoảng thời gian từ 01/01/2020 đến 01/08/2024.

Phạm vi nghiên cứu bao gồm:

* Dữ liệu Kỹ Thuật: Thu thập dữ liệu giá cổ phiếu TESLA từ Yahoo Finance, bao gồm giá mở cửa, giá đóng cửa và khối lượng giao dịch. Tính toán các chỉ số kỹ thuật như EMA, RSI và MACD,...
* Dữ liệu Cơ Bản: Thu thập báo cáo tài chính của TESLA để tính toán các chỉ số tài chính quan trọng như EPS, PE, và Net Income,...
* Dữ liệu Cảm Xúc: Sử dụng API của EODHD để thu thập bài báo liên quan đến TESLA và phân tích cảm xúc qua mô hình Meta-LLaMA.
* Mô Hình Dự Báo: Áp dụng các mô hình SARIMAX, LSTM, Random Forest và XGBoost để dự đoán giá cổ phiếu, so sánh hiệu quả của các mô hình, sau đó là sự kết hợp SARIMAX + XGBOOTS.

## 1.4. Ý nghĩa khoa học và thực tiễn

***Ý nghĩa khoa học:***

Nghiên cứu này đóng góp vào việc phát triển các phương pháp dự đoán giá cổ phiếu thông qua việc áp dụng các mô hình học máy tiên tiến như LSTM, Random Forest, XGBoost, SARIMAX và SARIMAX + XGBOOTS. Việc so sánh các mô hình này giúp nâng cao độ chính xác trong dự báo giá cổ phiếu và cung cấp cái nhìn sâu sắc về hiệu quả của các phương pháp phân tích kỹ thuật, cơ bản và tâm lý thị trường. Nghiên cứu cũng mở rộng khả năng tích hợp các yếu tố ngoại sinh và tâm lý thị trường vào mô hình dự báo, làm phong phú thêm nền tảng lý thuyết trong phân tích tài chính.

***Ý nghĩa thực tiễn:***

Việc dự đoán giá cổ phiếu có thể hỗ trợ và giúp ích nhiều trong thực tế, đặc biệt là các nhà đầu tư chứng khoán:

* Cung cấp cho nhà đầu tư các công cụ dự báo chính xác về biến động giá cổ phiếu, từ đó giúp họ đưa ra những quyết định đầu tư thông minh hơn. Các mô hình dự đoán có thể giúp nhà đầu tư nhận diện xu hướng giá cổ phiếu sắp tới, giảm thiểu rủi ro và tối ưu hóa lợi nhuận trong các giao dịch.
* Các mô hình dự báo giúp nhà đầu tư và tổ chức tài chính xây dựng chiến lược đầu tư hiệu quả, ra quyết định chính xác và ứng phó kịp thời với biến động của thị trường chứng khoán.
* Doanh nghiệp, đặc biệt là ngân hàng, có thể sử dụng các mô hình dự báo để tối ưu hóa quyết định về huy động vốn, tài chính và chiến lược mua bán cổ phiếu.
* Nghiên cứu này cũng có ý nghĩa quan trọng đối với các cơ quan quản lý nhà nước và tổ chức tài chính. Các kết quả nghiên cứu có thể giúp cơ quan quản lý hiểu rõ hơn về các yếu tố tác động đến giá cổ phiếu của các ngân hàng, từ đó xây dựng các chính sách ổn định thị trường chứng khoán và phát triển thị trường tài chính một cách bền vững.

# CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

## 2.1. Học máy

Học máy (Machine Learning) là một lĩnh vực thuộc trí tuệ nhân tạo cho phép máy tính học hỏi và cải thiện hiệu suất mà không cần được lập trình rõ ràng cho từng nhiệm vụ. Nó hoạt động bằng cách phân tích dữ liệu để nhận dạng các mẫu và từ đó đưa ra dự đoán hoặc quyết định.

Các loại kỹ thuật học máy:

* Học có giám sát (Supervised Learning): là phương pháp mà mô hình được huấn luyện bằng dữ liệu đã được gắn nhãn. Nghĩa là, mỗi điểm dữ liệu đầu vào (input) đều được gắn kèm với một đầu ra đúng, giúp phân loại dữ liệu hoặc dự đoán kết quả một cách chính xác.
* Học không giám sát (Unsupervised Learning): là phương pháp được sử dụng khi dữ liệu không có nhãn. Mục tiêu chính là tìm ra các mẫu, mối quan hệ, hoặc cấu trúc ẩn bên trong dữ liệu mà không cần sự can thiệp trực tiếp từ con người.
* Học bán giám sát (Semi-supervised Learning): là phương pháp kết hợp giữa học có giám sát và học không giám sát, được sử dụng khi có một lượng lớn dữ liệu không có nhãn và một lượng nhỏ dữ liệu có nhãn. Mục tiêu là tận dụng cả dữ liệu có nhãn và không có nhãn để cải thiện hiệu quả mô hình học máy.

Dự đoán giá cổ phiếu là một ứng dụng điển hình của học máy trong lĩnh vực tài chính. Hiện nay, có nhiều thuật toán học máy được ứng dụng phổ biến vào việc dự đoán giá cổ phiếu như: LSTM, XGBoost, Random Forest, SVM, KNN,...

Độ chính xác của các mô hình học máy phụ thuộc rất nhiều vào ngữ cảnh, chịu ảnh hưởng của các yếu tố như loại dữ liệu, thuật toán cụ thể được sử dụng và kích thước và chất lượng của tập dữ liệu.

Các yếu tố có thể ảnh hưởng đến độ chính xác của các mô hình học máy:

* Lặp lại dữ liệu : Tăng kích thước tập dữ liệu thông qua việc lặp lại thử nghiệm đã cải thiện đáng kể độ chính xác của các mô hình học máy, đặc biệt đối với các phương pháp tổng hợp như Random Forest, lên tới 23% [4].
* Lý luận nhân quả : Việc xây dựng lại chẩn đoán dưới dạng nhiệm vụ suy luận phản chứng đã cải thiện độ chính xác của các mô hình học máy, đưa chúng vào top 25% bác sĩ trong một bộ thử nghiệm lâm sàng [5].
* Bộ dữ liệu nhỏ : Việc kết hợp các ước tính thô về các thuộc tính trong không gian tính năng có thể tăng cường khả năng dự đoán của các mô hình học máy bằng cách sử dụng các bộ dữ liệu nhỏ, đạt được mức độ chính xác tiên tiến [6].

## 2.2. Phân tích chuỗi thời gian

Phân tích chuỗi thời gian (Time Series Analysis) là một lĩnh vực quan trọng trong khoa học dữ liệu, liên quan đến việc nghiên cứu các điểm dữ liệu được thu thập hoặc ghi lại theo các khoảng thời gian cụ thể. Phân tích này rất cần thiết để hiểu các cơ chế cơ bản của nhiều hiện tượng khác nhau và để đưa ra dự đoán về các giá trị trong tương lai.

Các thành phần của chuỗi thời gian:

* Xác định xu hướng (Trend): Biểu thị sự tiến triển dài hạn của chuỗi. Nó cho thấy hướng chung mà dữ liệu đang di chuyển trong một thời gian dài, có thể là tăng, giảm hoặc giữ nguyên [7].
* Mùa vụ (Seasonality): Đây là những mô hình hoặc chu kỳ lặp lại thường xuyên trong dữ liệu xảy ra trong một khoảng thời gian cụ thể, chẳng hạn như một năm, tháng hoặc tuần. Các biến động theo mùa thường bị ảnh hưởng bởi các điều kiện khí hậu, ngày lễ hoặc các sự kiện lặp lại khác [7].
* Bất thường (Irregular): Còn được gọi là thành phần dư hoặc nhiễu, nó nắm bắt các biến thể ngẫu nhiên, không thể đoán trước trong dữ liệu mà không thể quy cho các thành phần xu hướng, theo chu kỳ hoặc theo mùa. Những điều này thường là do các sự kiện hoặc bất thường không lường trước được [7].
* Chu kỳ (Cyclic): Đây là những biến động xung quanh đường xu hướng xảy ra theo các khoảng thời gian không đều, thường liên quan đến chu kỳ kinh tế hoặc kinh doanh. Chúng phản ánh các giai đoạn mở rộng và thu hẹp trong dữ liệu [7], [8].

Các mô hình phổ biến trong phân tích chuỗi thời gian:

* ARIMA (AutoRegressive Integrated Moving Average): ARIMA là một mô hình mạnh mẽ trong phân tích chuỗi thời gian, chuỗi dữ liệu không có tính mùa vụ. ARIMA kết hợp ba thành phần: tự hồi quy (AR), tích hợp (I), và trung bình động (MA).
* SARIMA (Seasonal ARIMA): SARIMA mở rộng ARIMA bằng cách bổ sung các yếu tố mùa vụ, có ích khi phân tích dữ liệu có tính chu kỳ rõ ràng.
* Exponential Smoothing (Hàm làm mượt số học): Phương pháp này sử dụng một chỉ số mượt hóa để dự đoán các giá trị trong tương lai, với các dạng như Holt-Winters phù hợp cho dữ liệu có xu hướng và tính mùa vụ.
* VAR (Vector Autoregression): VAR là một phương pháp mạnh mẽ khi phân tích các chuỗi thời gian liên quan lẫn nhau.

Biểu diễn và lập chỉ mục dữ liệu chuỗi thời gian hiệu quả là rất quan trọng để xử lý các tập dữ liệu lớn. Các kỹ thuật trong lĩnh vực này giúp giảm chiều và cải thiện hiệu quả tính toán của phân tích chuỗi thời gian. Các phương pháp này được sử dụng để so sánh các chuỗi thời gian khác nhau và phân đoạn chúng thành các phần có ý nghĩa, điều này rất cần thiết cho việc nhận dạng mẫu và phát hiện bất thường [9]

## 2.3. Bài toán dự đoán giá cổ phiếu của TESLA bằng các mô hình chuỗi thời gian và học máy

### **2.3.1. Tổng quan**

Cổ phiếu là một loại chứng khoán xác nhận quyền sở hữu một phần vốn điều lệ của công ty cổ phần. Khi mua cổ phiếu, nhà đầu tư trở thành cổ đông của công ty và có các quyền lợi như nhận cổ tức, tham gia biểu quyết trong các cuộc họp cổ đông, và hưởng lợi từ sự tăng trưởng giá trị của công ty.

Giá cổ phiếu là một chỉ báo quan trọng về giá trị thị trường của một công ty và chịu ảnh hưởng của nhiều yếu tố bao gồm điều kiện thị trường, các yếu tố cơ bản và kỹ thuật, và các biến số kinh tế vĩ mô. Vậy nên, nếu hiểu được những ảnh hưởng này và sử dụng các mô hình dự đoán, nhà đầu tư có thể dễ dàng đưa ra quyết định đầu tư sáng suốt, giảm thiểu được rủi ro và tối ưu hóa lợi nhuận khi đầu tư vào thị trường chứng khoán,

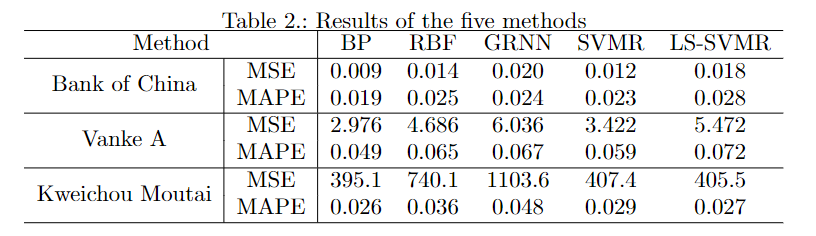
### **2.3.2. Các nghiên cứu liên quan**

Dự đoán giá cổ phiếu là một lĩnh vực đang nhận được rất nhiều sự quan tâm trong giới tài chính. Vì vậy, đã có rất nhiều nghiên cứu và phương pháp đề xuất được đưa ra để dự đoán vấn đề này. Basak & cộng sự (2019) [10] đã dự đoán hướng đi của giá cổ phiếu bằng cách sử dụng rừng ngẫu nhiên (Random Forest) và cây quyết định tăng cường gradient (XGBoost). Nghiên cứu này nhấn mạnh việc sử dụng các chỉ báo kỹ thuật làm tính năng và báo cáo độ chính xác cao cho các dự đoán trung hạn đến dài hạn. Trong các thí nghiệm của họ, dữ liệu chuỗi thời gian thu được trước tiên được làm mịn theo cấp số nhân. Sau đó, các chỉ báo kỹ thuật được trích xuất. Các chỉ báo kỹ thuật cung cấp thông tin chi tiết về hành vi giá cổ phiếu dự kiến ​​trong tương lai. Các chỉ báo kỹ thuật này được sử dụng làm các tính năng để đào tạo các bộ phân loại.

Ngoài ra, Yue-Gang Song & cộng sự (2018) [11] đã sử dụng mạng nơ-ron để dự đoán giá cổ phiếu, đánh giá hiệu quả của năm mô hình: Mạng nơ-ron truyền ngược (BP), Mạng hàm cơ sở xuyên tâm (RBF), Mạng hồi quy tổng quát (GRNN), Mạng hồi quy tổng quát (GRNN), Hồi quy máy vector hỗ trợ bình phương nhỏ nhất (LS-SVMR). Dữ liệu bao gồm giá cổ phiếu của ba công ty lớn tại Trung Quốc: Bank of China, Vanke A, Kweichou Moutai. Dữ liệu được lấy từ năm 2006 đến năm 2018.

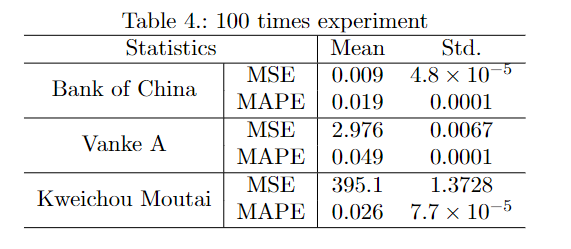
Một mặt, kernel tuyến tính là tốt nhất trong nhiệm vụ dự đoán này và luôn vượt trội so với ba kernel còn lại. Mặc dù kernel RBF là kernel mặc định trong nhiều gói phần mềm vì tính linh hoạt với nhiều nguồn dữ liệu khác nhau, nhưng nó không đủ tốt trong trường hợp này. Vì vậy, chúng ta nên thử các kernel khác để so sánh khi thực hiện các dự án dự đoán tương tự.

Mặt khác, BP vẫn vượt trội so với SVMR với kernel tuyến tính, mặc dù lợi thế này không rõ ràng. Hai phương pháp có lỗi dự đoán tương tự, có thể là nhờ vào việc cả hai đều sử dụng trung bình trọng số, giúp nắm bắt một số mối quan hệ tuyến tính trong mạng. Kết quả như hình dưới đây:



Hình 1: Kết quả dự đoán nghiên cứu của Yue-Gang Song & cộng sự [11]

Khi triển khai thuật toán BP, cần khởi tạo trọng số ngẫu nhiên, điều này có thể gây ra sự bất ổn trong kết quả. Để chứng minh rằng kết quả của BP là ổn định, mạng nơ-ron được huấn luyện 100 lần và tính toán giá trị trung bình và độ lệch chuẩn. Bảng bên dưới giúp giải quyết lo ngại này vì độ lệch chuẩn rất nhỏ so với giá trị trung bình tương ứng, chứng tỏ rằng kết quả của mỗi lần thử nghiệm là đáng tin cậy.



Hình 2: Kết quả 100 lần thử nghiệm của Yue-Gang Song & cộng sự [11]

# CHƯƠNG 3: MÔ HÌNH ĐỀ XUẤT

## 3.1. Tổng quan mô hình đề xuất

Time Series Forecasting

Step 2:

Data Cleaning, Data Processing And Feature Engineering

SARIMAX

LSTM

RANDOM FOREST

XGBOOTS

Gird Search/ auto\_arima

Mean squared Error, Cross validation

Fundamental

Dataset

Step 1:

Define the problem

Technical

Sentiment

Exploratory Data Analysis (EDA)

Step 6:

Expand: SARIMAX + XGBOOTS

* Root Mean Squared Error (RMSE)
* R-squared(R2),
* Mean Absolute Error (MAE)
* Mean Absolute Percentage Error(MAPE)

Step 5:

Evulation/Summary

END

Step 4:

Test the model result/performance

Step 3:

Hyperparameter Tunning & Build Model Training

Hình 3: Mô hình đề xuất

Các bước để xây dựng mô hình cho bài toán dự đoán giá cổ phiếu:

Bước 1: Xác định vấn đề (Define the problem):

* Sử dụng các phân tích kỹ thuật (Technical), phân loại cảm xúc (Sentiment), phân tích cơ bản (Fundamental) để hình thành tập dữ liệu.
* Xác định vấn đề là sử dụng phương pháp dự báo chuỗi thời gian ( Time Series Forecasting ) để giải quyết đề tài này.

Bước 2: Làm sạch dữ liệu, xử lý dữ liệu và kỹ thuật đặc trưng

* Thực hiện quy trình EDA (Exploratory Data Analysis) để khai phá dữ liệu, phân tích thống kê cơ bản, xử lý dữ liệu, xem xét mối quan hệ giữa các features và trực quan hóa dữ liệu
* Mục tiêu: chuẩn bị dữ liệu đầu vào để đảm bảo chất lượng và độ chính xác.

Bước 3: Điều chỉnh siêu tham số và huấn luyện mô hình

* Sử dụng các kỹ thuật như Grid Search hoặc auto\_arima.
* Xây dựng và huấn luyện các mô hình: SARIMAX, XGBoost, LSTM và Random Forest.

Bước 4: Kiểm tra hiệu suất mô hình

* Áp dụng các chỉ số như Mean Squared Error (MSE), Cross Validation.

Bước 5: Đánh giá và tóm tắt

* Sử dụng các chỉ số như Root Mean Squared Error (RMSE), R-squared (R²), Mean Absolute Error (MAE), và Mean Absolute Percentage Error (MAPE) để đánh giá độ chính xác của mô hình.

Bước 6: Mở rộng (Expand):

* Kết hợp các mô hình (như SARIMAX và XGBoost) để tạo ra dự đoán tốt hơn bằng cách tách riêng các thành phần (như Residual và Noise) và tổng hợp chúng.

## 3.2. Đặc trưng của mô hình đề xuất

## 3.2.1. SARIMAX

SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) là một mô hình mở rộng của SARIMA, thêm vào khả năng xử lý các biến ngoại sinh (exogenous variables). Mô hình này kết hợp tính thời vụ (seasonality) và các yếu tố bên ngoài để cung cấp dự báo chính xác hơn, đặc biệt khi chuỗi thời gian bị ảnh hưởng bởi yếu tố thời vụ và các biến ngoại sinh

Thành phân của SARIMAX:

* Thành phần theo mùa (S): Ghi lại các mô hình tuần hoàn trong dữ liệu, coi hạn như chu kỳ hàng tuần, hàng tháng hoặc hàng năm.
* Thành phần hồi quy tự động (AR): Biểu tượng mối quan hệ giữa giá trị hiện tại và các giá trị trước đó trong chuỗi thời gian.
* Thành phần tích hợp (I): Bao gồm các công việc phân tích để làm cho chuỗi thời gian trở nên tĩnh tại bằng cách loại bỏ hướng và tính theo mùa.
* Thành phần trung bình động (MA): Tính sự phụ thuộc của giá trị hiện tại vào các điều khoản lỗi trong quá khứ, được sử dụng để tính xu hướng.
* Hồi quy ngoại lệ (X): Cho phép đưa vào các biến bên ngoài có thể ảnh hưởng đến thời gian chuỗi [12].

SARIMA được biểu diễn dưới dạng: SARIMA(p, d, q) × (P, D, Q)ₛ

* (p, d, q): Các tham số của phần ARIMA (phi tuyến chính, tích hợp, trung bình trượt).
* p: Bậc của phần tự hồi quy (AR).
* d: Bậc sai phân (làm cho chuỗi thời gian trở nên dừng).
* q: Bậc của phần trung bình trượt (MA).
* (P, D, Q): Các tham số của phần mùa vụ (theo chu kỳ mùa vụ) tương ứng.
* P: Bậc tự hồi quy mùa vụ.
* D: Bậc sai phân mùa vụ.
* Q: Bậc trung bình trượt mùa vụ.
* s: Chu kỳ của mùa vụ.

Mô hình SARIMA tổng quát:

Trong đó:

*): Phần tự hồi quy (AR) của chuỗi không mùa vụ.*

*: Phần tự hồi quy mùa vụ (Seasonal AR).*

*: Sai phân bậc d để làm dừng chuỗi thời gian không mùa vụ.*

*​: Sai phân bậc D để làm dừng chuỗi thời gian theo mùa vụ.*

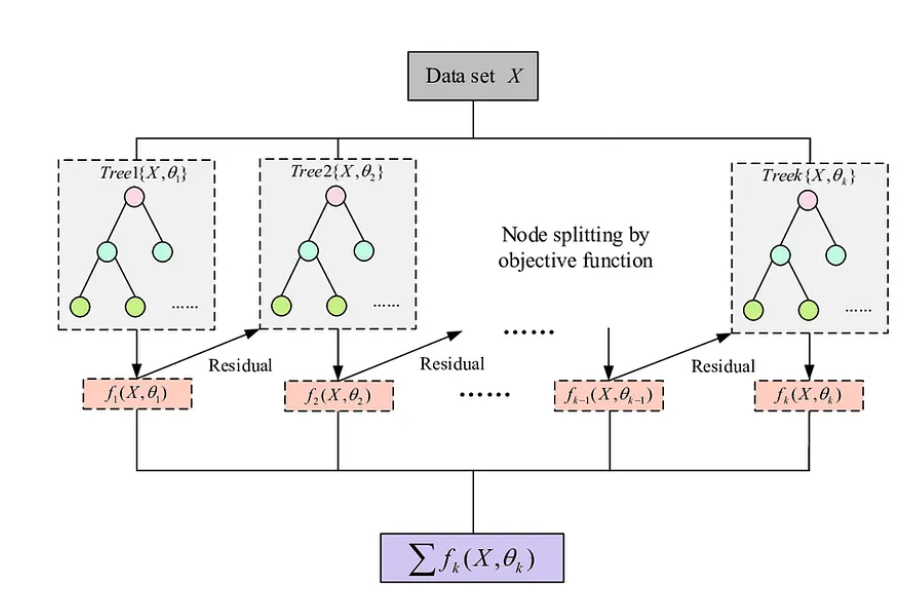
*: Phần trung bình trượt (MA) của chuỗi không mùa vụ.*

*: Phần trung bình trượt mùa vụ (Seasonal MA).*

*​: Nhiễu trắng (white noise) hoặc sai số ngẫu nhiên.).*

## 3.2.2. XGBOOST

XGBoost (Extreme Gradient Boosting) là một thuật toán học máy mạnh mẽ dựa trên kỹ thuật Gradient Boosting, được sử dụng để giải quyết các bài toán hồi quy, phân loại, và xếp hạng. XGBoost được phát triển để tối ưu hóa hiệu suất và tốc độ, cung cấp các giải pháp mạnh mẽ cho cả dữ liệu nhỏ và lớn.



Hình 4: Mô hình hoạt động của XGBoots [13]

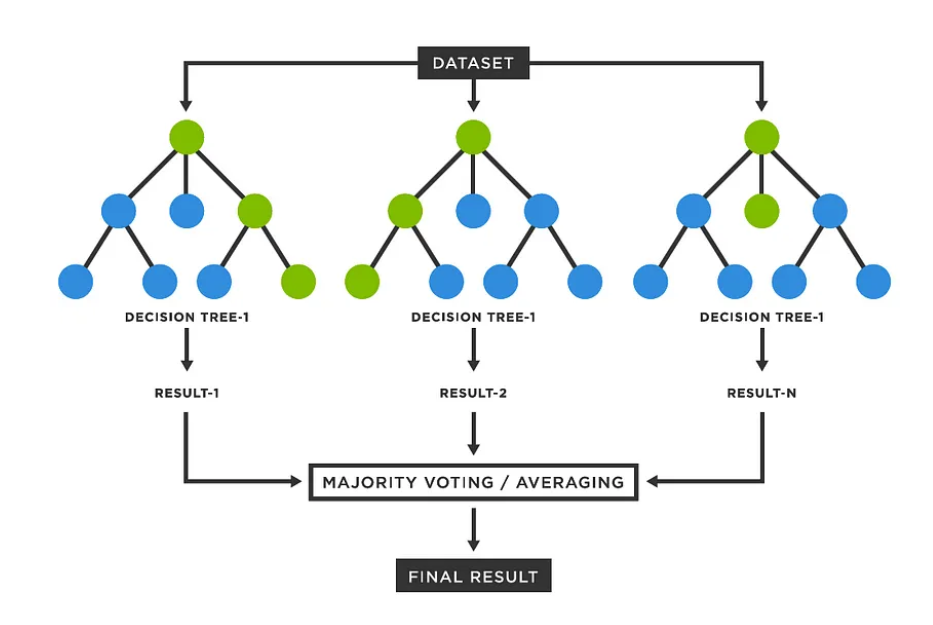
Cách thức hoạt động của XGBoost:

* Dự đoán ban đầu: XGBoost bắt đầu bằng cách đưa ra dự đoán đơn giản trên dữ liệu đào tạo, thường sử dụng giá trị trung bình của biến mục tiêu.
* Tính toán lỗi: Sau đó tính toán phần dư, là sự khác biệt giữa các giá trị dự đoán và các giá trị thực tế trong dữ liệu đào tạo. Về cơ bản, các phần dư này biểu thị các lỗi trong dự đoán ban đầu.
* Xây dựng cây quyết định đầu tiên: XGBoost xây dựng cây quyết định đầu tiên trong nhóm. Cây này tập trung vào việc học các giá trị dư này, nhằm mục đích giảm thiểu lỗi tổng thể. Để làm được điều này, thuật toán tìm các điểm phân chia tốt nhất trong các tính năng sẽ giảm lỗi nhiều nhất.
* Cây tiếp theo và Sửa lỗi: Đây là nơi phép thuật của việc tăng cường độ dốc diễn ra. XGBoost không loại bỏ cây trước đó. Thay vào đó, nó sử dụng lại các giá trị còn lại, nhưng lần này là cho các dự đoán được thực hiện bởi toàn bộ nhóm cho đến nay (bao gồm cả cây đầu tiên). Cây mới nhắm mục tiêu cụ thể vào các lỗi còn lại này, cải thiện thêm độ chính xác của mô hình.
* Giảm thiểu hàm mất mát: Trong suốt quá trình, XGBoost tối ưu hóa hàm mất mát. Hàm này đo lường về mặt toán học mức độ phù hợp giữa dự đoán của mô hình với các giá trị thực tế. Bằng cách giảm thiểu hàm mất mát, XGBoost đảm bảo rằng nhóm đang đi đúng hướng để đưa ra dự đoán chính xác.
* Tiêu chí dừng: XGBoost thêm cây cho đến khi đạt được tiêu chí dừng nhất định. Các tiêu chí này có thể là số lượng cây tối đa, cải thiện tối thiểu trong hàm mất mát hoặc đạt đến một mức độ chính xác nhất định [14].

## 3.2.3. Random Forest

Random Forest là một thuật toán học máy ensemble learning dựa trên các cây quyết định (Decision Trees). Thuật toán này sử dụng kỹ thuật Bagging (Bootstrap Aggregating) để kết hợp nhiều cây quyết định thành một mô hình mạnh mẽ hơn, giảm hiện tượng overfitting và cải thiện độ chính xác của dự đoán.

Nó được sử dụng cho cả bài toán phân loại và hồi quy, và đặc biệt hiệu quả khi làm việc với dữ liệu lớn, phức tạp hoặc nhiều chiều.

****

Hình 5: Mô hình Random Forest [15]

Quá trình hoạt động của Random Forest:

* Tạo ra các cây quyết định: Random Forest bắt đầu bằng việc tạo ra nhiều cây quyết định từ dữ liệu. Mỗi cây quyết định được xây dựng từ một tập con ngẫu nhiên của dữ. Tập con này được lấy mẫu với thay thế, nghĩa là một số mẫu có thể xuất hiện nhiều lần trong tập con.
* Chọn đặc trưng ngẫu nhiên: Trong quá trình xây dựng mỗi cây quyết định, Random Forest chọn một tập con ngẫu nhiên của các đặc trưng tại mỗi nút phân chia của cây. Điều này giúp giảm mối tương quan giữa các cây quyết định và làm cho mô hình trở nên mạnh mẽ hơn. Quá trình này còn giúp cây quyết định không bị phụ thuộc quá nhiều vào bất kỳ đặc trưng cụ thể nào, dẫn đến việc giảm thiểu overfitting.
* Huấn luyện các cây quyết định: Mỗi cây quyết định trong Random Forest được huấn luyện độc lập trên các tập con dữ liệu và tập con đặc trưng đã chọn. Nó thường tạo ra các cây có cấu trúc khác nhau và quyết định dự đoán khác nhau.
* Kết hợp kết quả: Sau khi các cây quyết định được huấn luyện, Random Forest kết hợp kết quả dự đoán từ tất cả các cây. Đối với bài toán phân loại, dự đoán cuối cùng là lớp mà có nhiều cây quyết định nhất chọn. Còn với bài toán hồi quy, dự đoán cuối cùng là giá trị trung bình của các dự đoán từ các cây quyết định.
* Đánh giá và cải thiện: Cuối cùng, kết quả của Random Forest có thể được đánh giá và so sánh với các mô hình khác để kiểm tra độ chính xác và hiệu suất. Trong trường hợp cần thiết, các tham số của thuật toán có thể được điều chỉnh để cải thiện hiệu suất [16].

## 3.3.4. LSTM

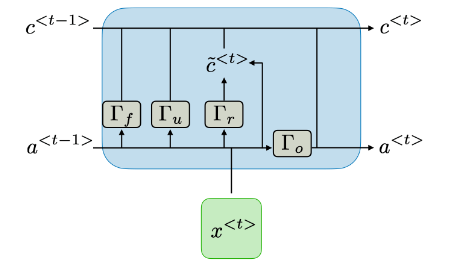
LSTM (Long Short-Term Memory) là một loại kiến trúc mạng Recurrent Neural Network (RNN), được thiết kế để giải quyết các vấn đề liên quan đến chuỗi dữ liệu tuần tự và phụ thuộc dài hạn (long-term dependencies). Nó được giới thiệu bởi Hochreiter và Schmidhuber vào năm 1997 nhằm khắc phục những hạn chế của RNN thông thường, đặc biệt là hiện tượng vanishing gradient khi học các chuỗi dài.

Kiến trúc LSTM liên quan đến ô nhớ được điều khiển bởi ba cổng: cổng vào, cổng quên và cổng ra. Cổng này quyết định thông tin nào sẽ được thêm vào, xóa và xuất ra khỏi bộ nhớ.

* Kiểm tra đầu vào Kiểm soát bất kỳ thông tin nào được thêm vào ô nhớ.
* Hãy quên kiểm tra thông tin nào đã bị xóa khỏi ô nhớ.
* Cổng điều khiển thông tin nào được đưa ra từ ô nhớ.

Điều này cho phép LSTM mạng giữ lại hoặc loại bỏ thông tin bằng cách chọn bộ lọc khi thông tin truyền qua mạng, cho phép chúng tìm hiểu các mối quan hệ phụ thuộc lâu dài.

LSTM duy trì trạng thái ẩn, hoạt động như thời hạn ngắn của mạng. Trạng thái ẩn được cập nhật dựa trên đầu vào, trạng thái ẩn trước đó và trạng thái hiện tại của ô nhớ [17].



Hình 6: Cấu trúc của một đơn vị trong mô hình LSTM [18]

Phương trình đặc trưng của LSTM:

c~<*t*> = tanh(*Wc*​[Γ*r*​⋆*a*<*t*−1>,*x*<*t*>]+*bc*​)

*c*<*t*> = Γ*u*​⋆*c*~<*t*>+Γ*f*​⋆*c*<*t*−1>

*a*<*t*> = Γ*u*​⋆*c*~<*t*>+Γ*f*​⋆*c*<*t*−1>

Trong đó:

c~<t>: Giá trị ứng cử cho trạng thái bộ nhớ tại thời điểm t (được tính qua hàm tanh).

c<t>: Trạng thái bộ nhớ cuối cùng tại thời điểm t, được tính từ trạng thái ứng cử c~<t> và trạng thái bộ nhớ trước c<t-1>.

a<t>: Trạng thái ẩn tại thời điểm t, được tính từ c~<t> và c<t-1>.

Γr: Trọng số cho phần tử đầu vào.

Γu: Trọng số cho phần tử cập nhật.

Γf: Trọng số cho phần tử quên.

x<t>: Đầu vào tại thời điểm t.

a<t-1>: Trạng thái ẩn tại thời điểm trước t.

bc: Hệ số điều chỉnh bổ sung cho c~<t>.

## 3.3. Phương pháp đánh giá

**3.3.1. Root Mean Squared Error (RMSE)**

Root Mean Squared Error (RMSE) là một chỉ số dùng để đo lường mức độ sai lệch giữa giá trị dự đoán và giá trị thực tế trong các mô hình học máy, đặc biệt trong các bài toán dự đoán (regression problems). RMSE là một trong những chỉ số quan trọng để đánh giá hiệu quả của mô hình.

Công thức tính RMSE như sau:

RMSE =

Trong đó:

*: số lượng dữ liệu*

*: Giá trị thực tế*

*: Giá trị dự đoán*

**3.3.2. R-squared (R2)**

R-squared (R²), còn được gọi là hệ số xác định, là một chỉ số thống kê dùng để đánh giá mức độ phù hợp của một mô hình hồi quy. Nó biểu thị tỷ lệ phần trăm của sự biến động trong biến phụ thuộc (biến đích) được giải thích bởi các biến độc lập (biến dự báo) trong mô hình.

R2 =

Trong đó:

: *Tổng bình* phương *sai số giữa giá trị thực tế và giá trị dự đoán*

: *Tổng bình phương sai số giữa giá trị thực tế và giá trị trung bình*

*: số lượng dữ liệu*

*: Giá trị thực tế*

*: Giá trị dự đoán*

*Giá trị trung bình của dữ liệu thực tế*

**3.3.3. Mean Absolute Error (MAE)**

Mean Absolute Error (MAE) là một chỉ số đánh giá hiệu suất được sử dụng phổ biến trong các bài toán dự đoán. Nó đo lường độ chính xác của mô hình bằng cách tính trung bình của các sai số tuyệt đối giữa giá trị dự đoán và giá trị thực tế.

MAE =

Trong đó:

*: số lượng dữ liệu*

*: Giá trị thực tế*

*: Giá trị dự đoán*

*: Sai số tuyệt đối giữa giá trị thực tế và giá trị dự đoán*

**3.3.4. Mean Absolute Percentage Error (MAPE)**

Mean Absolute Percentage Error (MAPE) là một chỉ số đánh giá hiệu suất của mô hình, đo lường mức độ sai lệch trung bình dưới dạng phần trăm giữa giá trị thực tế và giá trị dự đoán. Đây là một phương pháp phổ biến để biểu thị độ chính xác của mô hình dự đoán.

MAPE =

Trong đó:

*: số lượng dữ liệu*

*: Giá trị thực tế*

*: Giá trị dự đoán*

*: Sai số tương đối tuyệt đối giữa giá trị thực tế và giá trị dự đoán, tính bằng tỷ lệ phần trăm.*

# CHƯƠNG 4: THỰC NGHIỆM VÀ KẾT QUẢ

## 4.1. Dữ liệu

### **4.1.1. Thu thập dữ liệu**

Trong nghiên cứu này, dữ liệu được thu thập từ ba nguồn chính liên quan đến TESLA nhằm hỗ trợ xây dựng mô hình dự đoán giá cổ phiếu bằng các phương pháp học máy. Giai đoạn nghiên cứu kéo dài từ **01/01/2020** đến **01/08/2024**, bao quát các thông tin quan trọng từ biến động giá cổ phiếu, tác động của truyền thông, đến tình hình tài chính của công ty. Việc kết hợp ba bộ dữ liệu này giúp cung cấp cái nhìn toàn diện và đa chiều, đồng thời tạo ra một nền tảng mạnh mẽ để phát triển mô hình dự đoán chính xác hơn.

**Dữ liệu giá chứng khoán Tesla.**

Dữ liệu về giá cổ phiếu TESLA được lấy từ Yahoo Finance (YF) và cung cấp thông tin chi tiết về các phiên giao dịch hàng ngày. Các biến số chính bao gồm:

* Giá mở cửa (Open): Đại diện cho mức giá mà cổ phiếu được giao dịch khi thị trường bắt đầu phiên.
* Giá cao nhất (High) và giá thấp nhất (Low): Thể hiện biên độ dao động giá trong suốt ngày giao dịch, giúp đánh giá mức độ biến động.
* Giá đóng cửa (Close) và giá đóng cửa điều chỉnh (Adj Close): Là giá trị chính xác nhất phản ánh giá trị cổ phiếu khi thị trường kết thúc phiên, có tính đến các yếu tố như chia cổ tức hoặc tách cổ phiếu.
* Khối lượng giao dịch (Volume): Thể hiện số lượng cổ phiếu được mua bán, cung cấp thông tin về mức độ quan tâm của nhà đầu tư trong mỗi phiên.

Dữ liệu này là đầu vào quan trọng để phân tích xu hướng và biến động giá cổ phiếu, cung cấp nền tảng cho việc dự đoán.

**Dữ liệu bài báo theo ngày**

Truyền thông đóng vai trò quan trọng trong việc định hình tâm lý thị trường, và bộ dữ liệu bài báo liên quan đến TESLA được thu thập từ API của EODHD nhằm phân tích chi tiết các tác động này. Bộ dữ liệu bao gồm:

* Ngày đăng bài (Date): Xác định thời gian bài báo được phát hành, giúp liên kết thông tin truyền thông với các biến động thị trường cụ thể.
* Tiêu đề (Title) và nội dung bài viết (Content): Tóm tắt và mô tả chi tiết nội dung bài viết, cung cấp ngữ cảnh để phân tích.
* Nguồn phát hành (Source): Xác định đơn vị xuất bản bài báo, từ đó đánh giá độ tin cậy của thông tin.
* Danh mục bài báo (Category): Nhóm các bài báo theo chủ đề như công nghệ, tài chính, hoặc kinh tế.
* Cảm xúc bài báo (Sentiment): Phân loại cảm xúc (tích cực, tiêu cực, trung tính), đóng vai trò như một chỉ số đo lường tác động của truyền thông đến giá cổ phiếu.

Những bài báo này giúp làm rõ tác động của truyền thông đối với tâm lý thị trường và giá cổ phiếu, từ đó bổ sung yếu tố đầu vào cho mô hình học máy.

**Dữ liệu báo cáo tài chính Tesla**

Dữ liệu báo cáo tài chính được thu thập từ Yahoo Finance, cung cấp thông tin về tình hình kinh doanh của TESLA từ quý 1 năm 2020 đến quý 2 năm 2024. Các chỉ số chính bao gồm:

* **Tổng doanh thu (Total Revenue)** và **lợi nhuận gộp (Gross Profit)**: Đánh giá khả năng tạo doanh thu và hiệu quả sử dụng nguồn lực sản xuất.
* **Thu nhập từ hoạt động kinh doanh (Operating Income)** và **lợi nhuận ròng (Net Income)**: Cung cấp thông tin về khả năng sinh lời thực tế của công ty.

Báo cáo tài chính mang lại cái nhìn dài hạn về hiệu quả hoạt động, cung cấp thông tin định lượng để hỗ trợ dự đoán các xu hướng thị trường.

Mỗi bộ dữ liệu đóng vai trò bổ sung lẫn nhau, tạo nên một tập dữ liệu đa chiều. Ví dụ:

* Một sự kiện được đưa tin tiêu cực (dữ liệu bài báo) có thể dẫn đến sự sụt giảm giá cổ phiếu (dữ liệu giá chứng khoán) trong ngắn hạn.
* Sự cải thiện trong chỉ số tài chính (dữ liệu báo cáo tài chính) có thể làm tăng giá cổ phiếu trong trung hạn.

Khi kết hợp, các nguồn dữ liệu này cung cấp không chỉ ngữ cảnh mà còn là các đặc trưng quan trọng để mô hình học máy khai thác.

### **4.1.2. Tổng quan dữ liệu**

Dữ liệu giá cổ phiếu Tesla được thu thập từ Yahoo Finance, bao gồm các chỉ số cơ bản như giá mở cửa, giá đóng cửa, giá cao nhất, giá thấp nhất và khối lượng giao dịch. Tuy nhiên, để cải thiện chất lượng dự báo, bộ dữ liệu này đã được mở rộng thêm với các chỉ số phân tích kỹ thuật (Technical Analysis). Những chỉ số này giúp làm rõ xu hướng ngắn hạn và dài hạn của giá cổ phiếu, hỗ trợ mô hình học máy trong việc dự báo giá cổ phiếu Tesla.

Các yếu tố thời gian quan trọng, như năm, quý, tháng, ngày trong tuần, và một số chỉ báo kỹ thuật đã được tính toán và bổ sung vào bộ dữ liệu. Cụ thể, các biến sau đã được thêm vào:

* **year, quarter, month, day, dayofweek**: Những biến này giúp chia nhỏ dữ liệu theo các đơn vị thời gian khác nhau, từ đó giúp mô hình nhận diện các xu hướng và biến động theo mùa vụ hoặc các tuần trong năm.
* **is\_weekend**: Biến này cho biết liệu một ngày có phải là cuối tuần (thứ 7 hoặc chủ nhật), điều này quan trọng trong việc phân tích sự thay đổi giá cổ phiếu vào cuối tuần.
* **month\_end**: Biến này xác định ngày cuối tháng, giúp nhận diện sự biến động mạnh mẽ của giá cổ phiếu vào thời điểm này, khi các nhà đầu tư có xu hướng điều chỉnh danh mục đầu tư.

Ngoài các yếu tố thời gian, bộ dữ liệu cũng được tính toán các chỉ số phân tích kỹ thuật sau:

1. **EMA (Exponential Moving Average)**: Được tính cho 14 phiên giao dịch, chỉ báo này giúp theo dõi xu hướng ngắn hạn của giá cổ phiếu, với trọng số cao đối với các giá trị gần đây, giúp phản ánh chính xác sự thay đổi nhanh chóng của giá.
2. **CMA (Cumulative Moving Average)**: Trung bình động lũy kế theo chu kỳ 7 ngày giúp phân tích xu hướng dài hạn của cổ phiếu.
3. **Độ lệch chuẩn 7 ngày**: Đo lường sự biến động của giá cổ phiếu trong 7 ngày, giúp đánh giá mức độ rủi ro hoặc sự ổn định của cổ phiếu.
4. **Lợi nhuận hằng ngày (Daily Return)**: Được tính bằng phần trăm thay đổi giữa giá đóng cửa của ngày hiện tại và ngày trước đó, chỉ số này giúp theo dõi lợi nhuận ngắn hạn của cổ phiếu.
5. **Tỷ lệ chênh lệch giữa giá cao nhất và giá đóng cửa so với giá mở cửa (High-Close Ratio)**: Giúp đánh giá sức mạnh của xu hướng tăng giá trong một ngày giao dịch.
6. **Tỷ lệ chênh lệch giữa giá thấp nhất và giá mở cửa (Low-Open Ratio)**: Đánh giá xu hướng giảm giá trong ngày giao dịch.
7. **Cumulative Change (Thay đổi tích lũy)**: Tính toán tỷ lệ thay đổi tích lũy của giá đóng cửa so với tổng giá trị trong toàn bộ thời gian dữ liệu, từ đó đánh giá sự tăng trưởng của cổ phiếu.

Các chỉ số kỹ thuật bổ sung khác bao gồm:

* **Bollinger Bands**: Đo lường độ biến động của giá cổ phiếu và xác định các điểm mua bán tiềm năng.
* **SMA (Simple Moving Average)**: Trung bình động đơn giản giúp làm mượt dữ liệu và xác định xu hướng tổng thể.
* **RSI (Relative Strength Index)**: Chỉ báo xác định cổ phiếu đang ở trạng thái quá mua hay quá bán.
* **Stochastic Oscillator**: Đo lường động lượng và xác định các điều kiện mua quá mức hoặc bán quá mức.
* **Williams %R**: Đo lường tình trạng quá mua hoặc quá bán của cổ phiếu.
* **CCI (Commodity Channel Index)**: Giúp xác định các mức giá quá cao hoặc quá thấp, từ đó đưa ra tín hiệu mua bán.
* **PPO (Percentage Price Oscillator)**: Đánh giá sự thay đổi giá cổ phiếu theo tỷ lệ phần trăm.

Để làm phong phú thêm bộ dữ liệu và cải thiện khả năng dự báo giá cổ phiếu, chúng tôi cũng tích hợp thông tin từ các bài báo tài chính thông qua API EODHD. Những bài báo này cung cấp cái nhìn sâu sắc về các yếu tố tác động đến thị trường, đồng thời giúp tạo ra các tính năng bổ sung hỗ trợ cho mô hình dự báo giá cổ phiếu.

API EODHD cung cấp thông tin về các bài báo tài chính liên quan đến cổ phiếu Tesla. Vì số lượng bài báo được xuất bản hàng ngày là rất lớn, chúng tôi giả định mỗi bài báo sẽ đại diện cho một ngày giao dịch, qua đó liên kết thông tin bài báo với biến động thị trường trong ngày đó.

Dữ liệu bài báo bao gồm tiêu đề, nội dung bài báo và thời gian xuất bản. Những bài báo này đã được lấy thông qua API và sử dụng để tạo ra các tính năng bổ sung, như:

* **sentiment\_ref**: Cảm xúc tham khảo từ phân loại cảm xúc có sẵn trong API bài báo. Tuy nhiên, nhận thấy rằng phân loại này không hoàn toàn chính xác, chúng tôi đã cải tiến quy trình phân loại cảm xúc bằng cách sử dụng mô hình phân tích cảm xúc của mình.

Để nâng cao độ chính xác của phân loại cảm xúc, chúng tôi sử dụng mô hình phân tích cảm xúc của **DeepInfra**—một mô hình ngôn ngữ lớn (LLM) của Meta. Mô hình này có khả năng hiểu ngữ cảnh và phân tích cảm xúc của các bài báo tài chính, từ đó cung cấp kết quả phân loại cảm xúc chính xác hơn. Các cảm xúc này được phân thành ba loại: **positive, negative,** và **neutral**.

Từ kết quả phân loại cảm xúc, chúng tôi tạo ra thêm tính năng:

* **sentiment\_model**: Cảm xúc được phân loại lại bởi mô hình phân tích cảm xúc của DeepInfra. Đây là một cải tiến so với cảm xúc tham khảo ban đầu và sẽ được sử dụng như một biến độc lập trong mô hình học máy.

Cuối cùng, bộ dữ liệu bao gồm hai biến quan trọng về cảm xúc:

* **sentiment\_ref**: Cảm xúc tham khảo từ API EODHD.
* **sentiment\_model**: Cảm xúc từ mô hình phân tích cảm xúc của DeepInfra.

Các tính năng này không chỉ giúp làm phong phú bộ dữ liệu mà còn cung cấp cái nhìn sâu sắc về tâm lý thị trường, điều này có thể ảnh hưởng đến quyết định đầu tư và dự báo giá cổ phiếu.

Ngoài phân tích kỹ thuật và phân tích cảm xúc, chúng tôi cũng bổ sung các chỉ số tài chính cơ bản vào mô hình dự báo giá cổ phiếu. Các chỉ số quan trọng từ báo cáo tài chính của Tesla, như **EPS** (Earnings Per Share) và **P/E** (Price-to-Earnings Ratio), đã được tính toán và sử dụng để làm phong phú thêm mô hình.

* **EPS**: Chỉ số lợi nhuận trên mỗi cổ phiếu, được tính bằng cách chia lợi nhuận ròng cho số lượng cổ phiếu đang lưu hành. Chỉ số này phản ánh khả năng sinh lời của công ty đối với mỗi cổ phiếu.
* **P/E**: Tỷ lệ giá trên lợi nhuận, được tính bằng cách chia giá cổ phiếu cho EPS. Chỉ số này giúp đánh giá giá trị cổ phiếu trong bối cảnh thị trường và xác định mức độ đánh giá của thị trường đối với cổ phiếu của công ty.

Dù việc tính toán các chỉ số này theo ngày bằng cách chia lợi nhuận ròng cho 365 ngày và chia giá cổ phiếu cho EPS có thể không chính xác tuyệt đối, nhưng chúng vẫn cung cấp một cái nhìn sơ bộ về các yếu tố cơ bản ảnh hưởng đến giá cổ phiếu trong ngắn hạn. Những chỉ số này sẽ bổ sung vào mô hình của chúng tôi, kết hợp với các chỉ số kỹ thuật và cảm xúc từ bài báo, tạo nên một mô hình dự báo toàn diện hơn cho giá cổ phiếu Tesla.

### **4.1.3. Khai phá và phân tích dữ liệu (EDA)**

Trước khi tiến hành với mô hình học máy, tôi sẽ thực hiện một số phân tích khám phá dữ liệu (EDA) để hiểu rõ hơn về bộ dữ liệu. EDA là bước quan trọng trong quá trình phân tích, giúp phát hiện các mẫu dữ liệu, kiểm tra các giả định và phát hiện những điểm dữ liệu bất thường, như các giá trị ngoại lai. Trong nghiên cứu này, tôi sẽ không hiển thị tất cả biểu đồ và số liệu thống kê của từng tính năng, mà chỉ tập trung vào những yếu tố quan trọng và hữu ích để hỗ trợ cho việc dự đoán giá cổ phiếu Tesla.

### **4.1.3. Tiền xử lí dữ liệu**

## 4.2. Thực nghiệm và đánh giá mô hình

## 4.2.1. Thực nghiệm với các tham số mặc định

## 4.2.2. Tinh chỉnh siêu tham số

## 4.2.3. Tổng kết

# CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## 5.1. Kết luận

Bài luận này đã nghiên cứu và áp dụng các phương pháp dự đoán giá cổ phiếu thông qua việc sử dụng các mô hình học máy và chuỗi thời gian, bao gồm SARIMAX, LSTM, SARIMAX + XGBOOTS, Random Forest và XGBoost. Dữ liệu được sử dụng kết hợp từ ba nguồn chính: dữ liệu kỹ thuật (technical), dữ liệu cơ bản (fundamental) và dữ liệu cảm xúc (sentiment).

Kết quả thực nghiệm đã cho thấy mô hình SARIMAX và SARIMAX + XGBOOTS đạt hiệu suất tốt nhất trong việc dự đoán giá cổ phiếu TESLA, chứng tỏ khả năng mô hình này trong việc nắm bắt xu hướng và đặc điểm dữ liệu tuyến tính.

## 5.2. Hạn chế

Bộ dữ liệu cảm xúc chủ yếu dựa vào các bài báo, chưa bao gồm dữ liệu từ mạng xã hội hoặc diễn đàn, nơi phản ánh nhanh và trực tiếp tâm lý nhà đầu tư.

Còn nhiều yếu tố ngoại sinh chưa được tích hợp.

## 5.3. Kiến thức và kỹ năng

Qua quá trình thực hiện nghiên cứu và phân tích dữ liệu dự đoán giá cổ phiếu, chúng tôi đã hiếu thêm cách thu thập và xử lý dữ liệu tài chính từ các nguồn như Yahoo Finance và API tài chính, đồng thời nắm được cách sử dụng các chỉ số kỹ thuật (EMA, RSI, MACD) và các yếu tố tài chính như EPS hay PE. Bên cạnh đó, trang bị được thêm những kiến thức về các mô hình học máy (Random Forest, XGBoost, LSTM) và mô hình thống kê (SARIMAX), đồng thời nắm vững cách đánh giá hiệu suất mô hình thông qua các chỉ số RMSE, MAE, MAPE, và R². Ngoài ra, tôi còn hiểu thêm về dữ liệu phi tuyến tính và cách các mô hình học sâu như LSTM xử lý hiệu quả loại dữ liệu này.

## 5.4. Hướng phát triển trong tương lai

Trong tương lai, chúng tôi có mong muốn mở rộng phạm vi dữ liệu bằng cách thu thập thêm dữ liệu từ nhiều nguồn khác nhau nhằm tăng tính đa dạng và độ chính xác của dự đoán. Đồng thời, bổ sung các yếu tố kinh tế vĩ mô như lãi suất, tỷ giá hối đoái, và chỉ số thị trường cũng sẽ giúp đánh giá ảnh hưởng toàn diện đến giá cổ phiếu.

Ngoài ra, chúng tôi muốn tích hợp các mô hình tiên tiến hơn, chẳng hạn như các mô hình hiện đại như Transformers hoặc mô hình kết hợp giữa học sâu và mô hình thống kê để cải thiện hiệu suất xử lý dữ liệu phi tuyến tính. Bên cạnh đó, việc ứng dụng các phương pháp tối ưu hóa siêu tham số, như Bayesian Optimization, có thể giúp nâng cao hiệu quả dự đoán của các mô hình.

# TÀI LIỆU THAM KHẢO

[1] M. Hatmanu and C. Cautisanu, “The Impact of COVID-19 Pandemic on Stock Market: Evidence from Romania,” *Int J Environ Res Public Health*, vol. 18, no. 17, p. 9315, Sep. 2021, doi: 10.3390/ijerph18179315.

[2] A. S. Baig, H. A. Butt, O. Haroon, and S. A. R. Rizvi, “Deaths, panic, lockdowns and US equity markets: The case of COVID-19 pandemic,” *Financ Res Lett*, vol. 38, p. 101701, Jan. 2021, doi: 10.1016/j.frl.2020.101701.

[3] W. Boungou and A. Yatié, “The impact of the Ukraine–Russia war on world stock market returns,” *Econ Lett*, vol. 215, p. 110516, Jun. 2022, doi: 10.1016/J.ECONLET.2022.110516.

[4] A. Bustillo, R. Reis, A. R. Machado, and D. Yu. Pimenov, “Improving the accuracy of machine-learning models with data from machine test repetitions,” *J Intell Manuf*, vol. 33, no. 1, pp. 203–221, Jan. 2022, doi: 10.1007/s10845-020-01661-3.

[5] J. G. Richens, C. M. Lee, and S. Johri, “Improving the accuracy of medical diagnosis with causal machine learning,” *Nat Commun*, vol. 11, no. 1, p. 3923, Aug. 2020, doi: 10.1038/s41467-020-17419-7.

[6] R. Hudson and A. Urquhart, “War and stock markets: The effect of World War Two on the British stock market,” *International Review of Financial Analysis*, vol. 40, pp. 166–177, Jul. 2015, doi: 10.1016/J.IRFA.2015.05.015.

[7] E. Bee Dagum and S. Bianconcini, “Time Series Components,” 2016, pp. 29–57. doi: 10.1007/978-3-319-31822-6\_2.

[8] S. Beveridge and C. R. Nelson, “A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle,’” *J Monet Econ*, vol. 7, no. 2, pp. 151–174, Jan. 1981, doi: 10.1016/0304-3932(81)90040-4.

[9] T. Fu, “A review on time series data mining,” *Eng Appl Artif Intell*, vol. 24, no. 1, pp. 164–181, Feb. 2011, doi: 10.1016/j.engappai.2010.09.007.

[10] S. Basak, S. Kar, S. Saha, L. Khaidem, and S. R. Dey, “Predicting the direction of stock market prices using tree-based classifiers,” *The North American Journal of Economics and Finance*, vol. 47, pp. 552–567, Jan. 2019, doi: 10.1016/j.najef.2018.06.013.

[11] Y.-G. Song, Y.-L. Zhou, and R.-J. Han, “Neural networks for stock price prediction,” May 2018, [Online]. Available: http://arxiv.org/abs/1805.11317

[12] “Complete Guide To SARIMAX in Python.”

[13] Fraidoon Omarzai, “XGBoost Regression In Depth.”

[14] Fraidoon Omarzai, “XGBoost Regression In Depth.”

[15] Deniz Gunay, “Random Forest Deniz Gunay Deniz Gunay  · Follow.”

[16] Bizfly Cloud, “Random Forest là gì? Ứng dụng của thuật toán trong Machine Learning.”

[17] “What is LSTM – Long Short Term Memory?”

[18] Shervine Amidi and Afshine Amidi, “Recurrent Neural Networks cheatsheet.”

# KẾ HOẠCH THỰC HIỆN

|  |  |  |  |
| --- | --- | --- | --- |
| **Tuần** | **Từ** | **Đến** | **Nội dung công việc** |
| **1** |  |  |  |
| **2** |  |  |  |
| **3** |  |  |  |
| **4** |  |  |  |
| **5** |  |  |  |
| **6** |  |  |  |
| **7** |  |  |  |
| **8** |  |  |  |
| **9** |  |  |  |
| **10** |  |  |  |
| **11** |  |  |  |
| **12** |  |  |  |
| **13** |  |  |  |
| **14** |  |  |  |
| **15** |  |  |  |

# NHẬT KÝ LÀM VIỆC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Nội dung công việc** | **Dự kiến** | **Thực tế** | **Mức độ hoàn thành** |
| **1** |  |  |  |  |
| **2** |  |  |  |  |
| **3** |  |  |  |  |
| **4** |  |  |  |  |
| **5** |  |  |  |  |
| **6** |  |  |  |  |
| **7** |  |  |  |  |
| **8** |  |  |  |  |
| **9** |  |  |  |  |
| **10** |  |  |  |  |
| **11** |  |  |  |  |
| **12** |  |  |  |  |
| **13** |  |  |  |  |
| **14** |  |  |  |  |
| **15** |  |  |  |  |