**Đại học quốc gia hà nội  
Trường đại học việt nhật**

\*\*\*



**BÀI TẬP LỚN**

**dự đoán giá cổ phiếu bằng phân tích chuỗi thời gian**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn:** | TS. Giang Thành Trung |
| **Nhóm sinh viên thực hiện:** | 1. Nguyễn Hải Linh – 23110316  2. Nguyễn Doãn Nam Anh – 23110346  3. Lê Văn Thành Long – 23110190 |

*Hà Nội, Tháng 12 – 2024*

HÀ NỘI, NĂM 2024

**MỤC LỤC**

[**A. GIỚI THIỆU CHUNG 3**](#_Toc186390136)

[**I. Giới thiệu đề tài 3**](#_Toc186390137)

[**1. Lý do chọn đề tài 3**](#_Toc186390138)

[**2. Mục tiêu nghiên cứu 3**](#_Toc186390139)

[**3. Phạm vi nghiên cứu 4**](#_Toc186390140)

[**II. TỔNG QUAN VỀ TẬP ĐOÀN HOÀ PHÁT VÀ TÔN HOA SEN 5**](#_Toc186390141)

[**1.Giới thiệu 5**](#_Toc186390142)

[**2. Giới thiệu về Tôn Hoa Sen 6**](#_Toc186390143)

[**B. PHẦN NỘI DUNG 7**](#_Toc186390144)

[**I. Nguồn lấy dữ liệu 7**](#_Toc186390145)

[**II. Tiền xử lí dữ liệu 7**](#_Toc186390146)

[**1. Tạo bảng mối quan hệ chỉ số 7**](#_Toc186390147)

[**2. Kết hợp bảng mối quan hệ 8**](#_Toc186390148)

[**3. Chuẩn bị dữ liệu Bảng CĐKT, KQKD 9**](#_Toc186390149)

[**4. Chuẩn bị dữ liệu cổ phần 11**](#_Toc186390150)

[**5. Chuẩn bị thông tin cơ bản về cổ phiếu 12**](#_Toc186390151)

[**6. Tổng hợp các thông tin vừa xử lí vào một file excel 12**](#_Toc186390152)

[**III. Phân tích kết quả tổng quan và chọn ra cổ phiếu tiềm năng 14**](#_Toc186390153)

[**1. Một số thông số cần lưu ý khi đọc biểu đồ 14**](#_Toc186390154)

[**2. TỔNG QUAN 5 MÃ CỔ PHIẾU 19**](#_Toc186390155)

[**3. Phân tích và lựa chọn cổ phiếu tiềm năng 21**](#_Toc186390156)

[**IV. Áp dụng mô hình học máy LSTM 23**](#_Toc186390157)

[**1. Tổng quan về Dự báo thời gian và LSTM 23**](#_Toc186390158)

[**2. Xây dựng mô hình LSTM 25**](#_Toc186390159)

[**C. KẾT LUẬN 35**](#_Toc186390160)

[**I. Tổng kết bài nghiên cứu 35**](#_Toc186390161)

[**II. Hướng phát triển trong tương lai 35**](#_Toc186390162)

[**1. Hạn chế 35**](#_Toc186390163)

[**2. Hướng phát triển 36**](#_Toc186390164)

[**D. NGUỒN THAM KHẢO 37**](#_Toc186390165)

**MỤC LỤC HÌNH ẢNH**

[**Hình 1. Hướng dẫn lấy dữ liệu về cổ phiếu 7**](#_Toc186396402)

[**Hình 2. Bảng mối quan hệ giữa các thuộc tính kĩ thuật (chưa phân cấp) 8**](#_Toc186396403)

[**Hình 3. Bảng phân cấp giữa các chỉ số (đã phân cấp) 8**](#_Toc186396404)

[**Hình 4. Hàm tự định nghĩa chuyển đổi quý sang ngày 10**](#_Toc186396405)

[**Hình 5. Bảng cân đối kế toán sau xử lí 10**](#_Toc186396406)

[**Hình 6. Hàm tự định nghĩa lấy ra giá cổ phiếu 11**](#_Toc186396407)

[**Hình 7. Bảng giá đóng/ mở cửa cổ phiếu 11**](#_Toc186396408)

[**Hình 8. Hàm lấy các chỉ số kĩ thuật 12**](#_Toc186396409)

[**Hình 9. Các thông tin kĩ thuật được lấy về 12**](#_Toc186396410)

[**Hình 10. Bảng cân đối kế toán (BS) 13**](#_Toc186396411)

[**Hình 11. Bảng thông tin cơ bản của cổ phiếu (Stock\_info) 13**](#_Toc186396412)

[**Hình 12. Bảng giá của 5 mã cổ phiếu (Stock\_Price) 14**](#_Toc186396413)

[**Hình 13. Tổng quan về Cổ phiếu HPG 20**](#_Toc186396414)

[**Hình 14. Tổng quan về cổ phiếu HSG 20**](#_Toc186396415)

[**Hình 15. Tổng quan về cổ phiếu NKG 21**](#_Toc186396416)

[**Hình 16. Tổng quan về cổ phiếu TLN 21**](#_Toc186396417)

[**Hình 17. Tổng quan về cổ phiếu TLH 22**](#_Toc186396418)

[**Hình 18. Bảng so sánh các chỉ số kĩ thuật giữa 5 mã cổ phiếu 22**](#_Toc186396419)

[**Hình 19. Tăng trưởng kinh tế Hoa Kì (1900-2000) 25**](#_Toc186396420)

[**Hình 20 Biễu diễn kiến trúc bên trong của một Cell LSTM. 26**](#_Toc186396421)

[**Hình 21. Mô hình sự khác nhau giữa RNN và LSTM 27**](#_Toc186396422)

[**Hình 22. Đọc dữ liệu giá cổ phiếu 28**](#_Toc186396423)

[**Hình 23. Xử lý dữ liệu sai thiếu 29**](#_Toc186396424)

[**Hình 24. Chuẩn hoá MinMaxScaler 29**](#_Toc186396425)

[**Hình 25. Chia tập dữ liệu huấn luyện 29**](#_Toc186396426)

[**Hình 26. Dùng dữ liệu 60 ngày để huấn luyện 30**](#_Toc186396427)

[**Hình 27. Phân lớp dữ liệu 30**](#_Toc186396428)

[**Hình 28. Thiết lập huấn luyện mô hình 30**](#_Toc186396429)

[**Hình 29. Chuẩn bị dữ liệu kiểm tra 31**](#_Toc186396430)

[**Hình 30. Vẽ biểu đồ giá cổ phiếu 32**](#_Toc186396431)

[**Hình 31. Giá cổ phiếu HPG từ 01-07-2020 đến nay 32**](#_Toc186396432)

[**Hình 32. Giá cổ phiếu HPG từ đầu 2024 tới năm 32**](#_Toc186396433)

[**Hình 33. Giá cổ phiếu HSG từ 01-07-2020 đến nay 33**](#_Toc186396434)

[**Hình 34. Giá cổ phiếu HSG từ đầu 2024 đến nay 33**](#_Toc186396435)

[**Hình 35. Dự báo 15 ngày tiếp theo 34**](#_Toc186396436)

[**Hình 36. Vẽ đồ thị dự báo 15 ngày tiếp theo 34**](#_Toc186396437)

[**Hình 37. Kết quả dự báo giá cổ phiếu của HPG 15 ngày 35**](#_Toc186396438)

[**Hình 38. Kết quả dự báo giá cổ phiếu của HSG 15 ngày 35**](#_Toc186396439)

[**Hình 39. Đánh giá mô hình bằng MSE và RMSE 35**](#_Toc186396440)

**A. GIỚI THIỆU CHUNG**

## **I. Giới thiệu đề tài**

### **1. Lý do chọn đề tài**

Thị trường chứng khoán Việt Nam trong những năm gần đây đang chứng kiến sự phát triển mạnh mẽ và là một trong những điểm đến hấp dẫn đối với các nhà đầu tư trong và ngoài nước. Tuy nhiên, thị trường này cũng rất biến động, chịu ảnh hưởng từ các yếu tố như tình hình kinh tế, chính trị trong nước và quốc tế, cũng như tâm lý của nhà đầu tư. Chính vì vậy, việc dự đoán xu hướng giá cổ phiếu và chọn lựa những mã cổ phiếu tiềm năng là một nhiệm vụ quan trọng nhưng đầy thách thức.

Trong số các mã cổ phiếu niêm yết, nhóm nghiên cứu chọn phân tích 5 mã cổ phiếu tiêu biểu là ***HSG***, ***HPG***, ***TLN***, ***KNG*** và ***TVN***. Đây đều là những doanh nghiệp lớn trong ngành thép và sản xuất vật liệu xây dựng, có tác động mạnh mẽ đến nền kinh tế và thị trường chứng khoán Việt Nam. Tuy nhiên, không phải tất cả các mã cổ phiếu này đều có triển vọng phát triển mạnh mẽ trong tương lai.

Với mục tiêu lựa chọn những mã cổ phiếu tiềm năng nhất để đầu tư, nhóm nghiên cứu quyết định phân tích các yếu tố như biến động giá cổ phiếu, khối lượng giao dịch và các chỉ số tài chính của từng công ty. **Kết quả phân tích cho thấy, hai mã cổ phiếu *HSG* và *HPG* có tiềm năng tăng trưởng ổn định và mạnh mẽ nhất.** Do đó, chúng được chọn làm đối tượng nghiên cứu tiếp theo trong quá trình huấn luyện mô hình dự đoán giá cổ phiếu bằng phương pháp học máy, cụ thể là mô hình LSTM (Long Short-Term Memory).

Việc áp dụng mô hình LSTM vào dự đoán giá cổ phiếu sẽ giúp nhà đầu tư đưa ra những quyết định đầu tư thông minh hơn, dự báo các xu hướng giá trong tương lai và tối ưu hóa lợi nhuận. Mô hình này có khả năng học từ dữ liệu lịch sử để đưa ra dự đoán chính xác, hỗ trợ các nhà đầu tư trong việc ra quyết định trong một thị trường có nhiều biến động.

Vì vậy, nhóm nghiên cứu quyết định thực hiện đề tài "*Phân tích và dự đoán giá cổ phiếu của các công ty* ***HSG*** *và* ***HPG*** *bằng phương pháp LSTM*", với mục tiêu tìm ra các cổ phiếu tiềm năng để ứng dụng mô hình học máy vào việc dự đoán giá trị trong tương lai.

### **2. Mục tiêu nghiên cứu**

Mục tiêu nghiên cứu của đề tài là xử dụng **Power BI** để trực quan hoá dữ liệu kĩ thuật của 5 mã cổ phiếu được chọn nghiên cứu. Sau đó, xây dựng một mô hình dự đoán giá cổ phiếu của Công ty Cổ phần Tập đoàn Hoà Phát (***HPG***) sử dụng phương pháp học máy (machine learning), cụ thể là mô hình **LSTM (Long Short-Term Memory)**. Cụ thể, các mục tiêu của nghiên cứu này bao gồm:

***Phân tích yếu tố ảnh hưởng đến giá cổ phiếu****:*

* Phân tích các yếu tố chính ảnh hưởng đến sự biến động giá cổ phiếu, bao gồm yếu tố vĩ mô như tình hình kinh tế, chính trị, và các yếu tố đặc thù của công ty như kết quả kinh doanh, tình hình sản xuất và xuất khẩu thép.
* Tìm hiểu mối quan hệ giữa các yếu tố này và giá cổ phiếu để có thể đưa ra các khuyến nghị cho nhà đầu tư.

***Xây dựng mô hình dự đoán giá cổ phiếu****:*

* Áp dụng các kỹ thuật học máy để dự đoán xu hướng giá cổ phiếu ***HPG*** dựa trên các yếu tố lịch sử như giá mở cửa, giá đóng cửa, giá cao nhất, giá thấp nhất, và khối lượng giao dịch.
* Tối ưu hóa mô hình LSTM để nâng cao độ chính xác trong việc dự đoán giá cổ phiếu của Tập đoàn Hoà Phát và Tôn Hoa Sen.

### **3. Phạm vi nghiên cứu**

Nghiên cứu này tập trung vào việc **dự đoán giá cổ phiếu của Tập đoàn Hoà Phát (HPG)** dựa trên dữ liệu lịch sử giá cổ phiếu và các yếu tố tác động từ bên ngoài như chỉ số kinh tế vĩ mô, tâm lý thị trường, và các yếu tố ngành. Mô hình dự đoán sẽ được xây dựng bằng cách sử dụng các phương pháp học máy (machine learning) hiện đại, bao gồm các kỹ thuật như **hồi quy tuyến tính**. Cụ thể, phạm vi nghiên cứu sẽ bao gồm các nội dung chính sau:

***Dự đoán giá cổ phiếu Tập đoàn Hoà Phát****:*

* Thu thập và sử dụng dữ liệu lịch sử giá cổ phiếu của Hoà Phát (gồm các chỉ số như giá mở cửa, giá đóng cửa, giá cao nhất, giá thấp nhất và khối lượng giao dịch) trong một khoảng thời gian nhất định.
* Áp dụng các mô hình học máy (machine learning) để dự đoán giá cổ phiếu trong tương lai, dựa trên các yếu tố tác động bên ngoài như biến động thị trường, các chỉ số kinh tế vĩ mô và tin tức kinh tế.

***Xây dựng mô hình dự đoán bằng phương pháp học máy****:*

* Sử dụng phương pháp học máy mô hình **LSTM (Long Short-Term Memory)** cho dự đoán giá cổ phiếu theo chuỗi thời gian.
* Tiền xử lý dữ liệu bao gồm chuẩn hóa và xử lý các giá trị thiếu, sau đó chia dữ liệu thành các tập huấn luyện và kiểm tra.

## **II. TỔNG QUAN VỀ TẬP ĐOÀN HOÀ PHÁT VÀ TÔN HOA SEN**

### **1.Giới thiệu**

1. ***Tổng quan***

Tập đoàn Hòa Phát là một trong những tập đoàn sản xuất công nghiệp hàng đầu Việt Nam. Khởi đầu từ một công ty chuyên buôn bán các loại máy xây dựng từ tháng 8/1992, Hòa Phát lần lượt mở rộng sang các lĩnh vực khác như Nội thất, ống thép, thép xây dựng, điện lạnh, bất động sản và nông nghiệp. Ngày 15/11/2007, Hòa Phát chính thức niêm yết cổ phiếu trên thị trường chứng khoán Việt Nam với mã chứng khoán HPG.

Hiện nay, Tập đoàn Hòa Phát hoạt động trong 05 lĩnh vực: Gang thép (thép xây dựng, thép cuộn cán nóng) - Sản phẩm thép (gồm Ống thép, tôn mạ, thép rút dây, thép dự ứng lực) - Nông nghiệp - Bất động sản – Điện máy gia dụng. Sản xuất thép là lĩnh vực cốt lõi, chiếm tỷ trọng 90% doanh thu và lợi nhuận toàn Tập đoàn. Với công suất 8.5 triệu tấn thép thô/năm, Hòa Phát là doanh nghiệp sản xuất thép lớn nhất khu vực Đông Nam Á.

Tập đoàn Hòa Phát giữ thị phần số 1 Việt Nam về thép xây dựng, ống thép; Top 5 về tôn mạ. Hiện nay, Hòa Phát nằm trong Top 5 doanh nghiệp tư nhân lớn nhất Việt Nam, Top 50 doanh nghiệp niêm yết hiệu quả nhất, Top 30 doanh nghiệp nộp ngân sách Nhà nước lớn nhất Việt Nam, Top 3 doanh nghiệp có vốn điều lệ lớn nhất thị trường chứng khoán, Top 10 cổ phiếu có vốn hóa lớn nhất thị trường chứng khoán Việt Nam.

Tập đoàn Hòa Phát hiện nay do Chủ tịch Hội đồng Quản trị Trần Đình Long dẫn dắt. Với sự lãnh đạo của ông Trần Đình Long, Hòa Phát đã trở thành một thương hiệu vững mạnh, góp phần lớn vào sự phát triển của nền kinh tế Việt Nam.

Với triết lý kinh doanh "Hòa hợp cùng phát triển", Hòa Phát dành ngân sách hàng trăm tỷ đồng mỗi năm để thực hiện trách nhiệm xã hội của doanh nghiệp với cộng đồng.

1. ***Đặc điểm biến động cổ phiếu***

Cổ phiếu Hoà Phát (***HPG***) không hoàn toàn ổn định mà có sự biến động đáng kể, chủ yếu do ảnh hưởng của các yếu tố kinh tế vĩ mô và diễn biến thị trường thép. Mặc dù Hoà Phát là công ty sản xuất thép lớn nhất khu vực Đông Nam Á, với nền tảng tài chính mạnh mẽ và thị phần lớn trong ngành thép, giá cổ phiếu của Tập đoàn vẫn thường xuyên biến động theo sự thay đổi của giá thép thế giới, giá nguyên liệu đầu vào, và các yếu tố nhu cầu thị trường.

Ngoài ra, các chính sách của Nhà nước đối với ngành thép, như thuế xuất nhập khẩu hay hỗ trợ ngành thép, cũng có thể tác động mạnh đến giá cổ phiếu. Tuy nhiên, trong dài hạn, cổ phiếu ***HPG*** có xu hướng ổn định hơn nhờ vào chiến lược đa dạng hóa ngành nghề, bao gồm việc mở rộng sang các lĩnh vực như bất động sản và nông nghiệp, giúp giảm thiểu phần nào sự phụ thuộc vào ngành thép và giảm rủi ro từ biến động giá thép.

Tóm lại, cổ phiếu Hoà Phát có thể mang lại cơ hội sinh lời cao nhưng cũng đi kèm với biến động lớn, đặc biệt trong ngắn hạn. Nhà đầu tư cần theo dõi sát sao các yếu tố tác động đến thị trường thép và tình hình kinh tế để đưa ra quyết định đầu tư hợp lý.

### **2. Giới thiệu về Tôn Hoa Sen**

***a) Tổng quan***

Tập đoàn Hoa Sen (***HSG***), thành lập năm 2001, là doanh nghiệp hàng đầu Việt Nam và Đông Nam Á trong lĩnh vực sản xuất, kinh doanh tôn mạ, thép lá mạ, và ống thép. Với mạng lưới phân phối rộng khắp và sản phẩm xuất khẩu đến hơn 85 quốc gia, Hoa Sen giữ thị phần số 1 về tôn mạ tại Việt Nam.

Cổ phiếu ***HSG*** niêm yết trên ***HOSE*** từ năm 2008, khẳng định vị thế thương hiệu với chiến lược "Trung thực - Cộng đồng - Phát triển bền vững". Dưới sự lãnh đạo của Chủ tịch Lê Phước Vũ, Tập đoàn đã không ngừng đầu tư công nghệ hiện đại và mở rộng quy mô.

***b) Đặc điểm biến động cổ phiếu***

Cổ phiếu ***HSG*** biến động theo giá nguyên liệu đầu vào (***HRC***), nhu cầu thị trường, và chính sách ngành thép. Tuy có sự dao động lớn trong ngắn hạn, ***HSG*** được đánh giá tiềm năng nhờ vị thế dẫn đầu và chiến lược xuất khẩu mạnh mẽ. Nhà đầu tư cần theo dõi sát diễn biến thị trường thép và chiến lược kinh doanh của Tập đoàn.

# **B. PHẦN NỘI DUNG**

## **I. Nguồn lấy dữ liệu**

Truy cập [***Trang cổ phiếu SSI***](https://iboard.ssi.com.vn/analysis/fundamental-analysis) để dowload các báo cáo các mã cổ phiếu

Hình 1. Hướng dẫn lấy dữ liệu về cổ phiếu

**No table of figures entries found.**

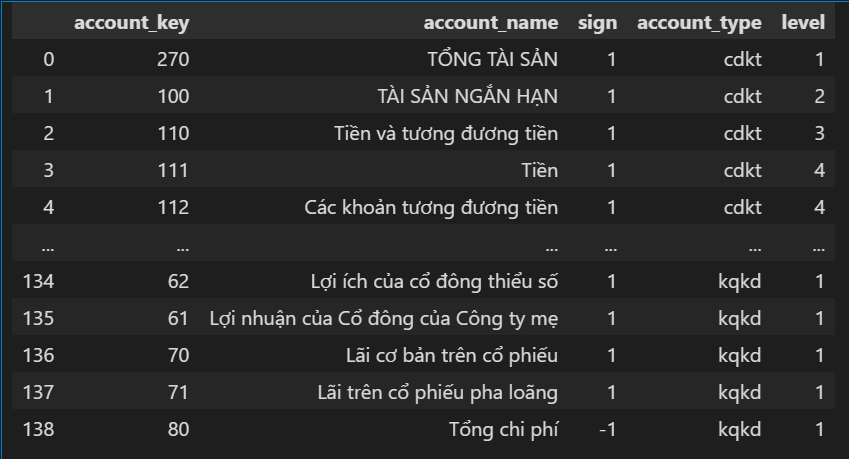
Dowload thành công và lưu tại Folder ‘**Raw Data’**

## II. Tiền xử lí dữ liệu

### **1. Tạo bảng mối quan hệ chỉ số**

Việc tạo một bảng mối quan hệ các chỉ số trong **DIM ACCOUNT** giúp tổ chức các tài khoản tài chính theo các nhóm, mối quan hệ và phân loại rõ ràng. Điều này giúp các nhà phân tích và người dùng báo cáo dễ dàng tìm và hiểu các chỉ số tài chính trong một cấu trúc có logic, như:

* **Account Key (Mã tài khoản):** Mã số duy nhất để nhận diện tài khoản.
* **Account Name (Tên tài khoản):** Tên mô tả của tài khoản.
* **Account Type (Loại tài khoản):** Phân loại tài khoản theo nhóm như "Tài sản", "Nợ phải trả", "Vốn chủ sở hữu", v.v.
* **Account Level (Cấp bậc tài khoản):** Chỉ ra mức độ chi tiết của tài khoản (level 1: tổng tài sản, level 2: tài sản ngắn hạn, level 3: tiền mặt, v.v.).
* **parent**: Đại diện cho tài khoản cha (hoặc tài khoản cấp cao), thường là các tài khoản tổng hợp, bao gồm nhiều tài khoản con.
* **child**: Đại diện cho tài khoản con, là các tài khoản chi tiết thuộc về một tài khoản cha. Các tài khoản con này có thể phản ánh các khoản mục nhỏ hơn trong báo cáo tài chính.



Hình 2. Bảng mối quan hệ giữa các thuộc tính kĩ thuật (chưa phân cấp)

### **2. Kết hợp bảng mối quan hệ**

Xây dựng một bảng thể hiện mối quan hệ phân cấp giữa các chỉ số **‘cân đối kế toán’** và **‘kết quả kinh doanh’.**



Hình 3. Bảng phân cấp giữa các chỉ số (đã phân cấp)

Class (Level 1):

- Tổng tài sản:

* Tài sản ngắn hạn
* Tài sản dài hạn

- Tổng cộng nguồn vốn:

* Nợ phải trả
* Vốn chủ sở hữu

**Sub class (level 2):**

- Tài sản ngắn hạn:

* Tiền và tương đương tiền
* Giá trị thuần đầu tư ngắn hạn
* Các khoản phải tìm
* Hàng tồn kho, ròng
* Tài sản lưu động khác

- Tài sản dài hạn:

* Khoản thu dài hạn
* Tài sản cố định
* Bất động sản đầu tư
* Tài sản dang dở dài hạn
* Đầu tư dài hạn
* Tài sản dài hạn khác

- Nợ phải trả:

* Nợ ngắn hạn
* Nợ dài hạn

- Vốn chủ sở hữu:

* Vốn và các quỹ
* Vốn ngân sách nhà nước và quỹ khác

### **3. Chuẩn bị dữ liệu Bảng CĐKT, KQKD**

Chuyển đổi từ Quý sang ngày đầu tháng: Tự định nghĩa hàm ‘**convert\_quarter\_to\_date’**

Cụ thể:

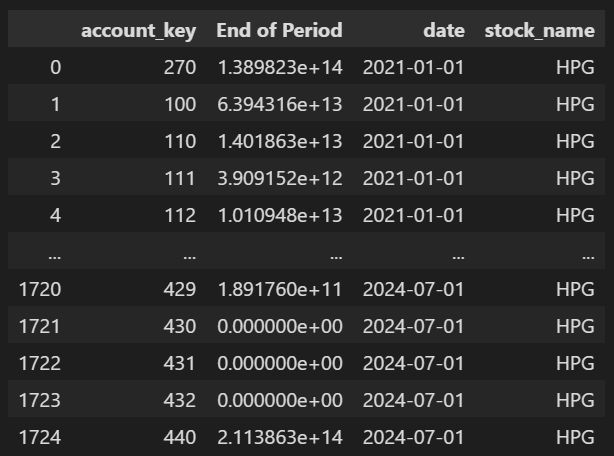
Hình 4. Hàm tự định nghĩa chuyển đổi quý sang ngày

* Đổi quý sang định dạng ngày, sau đó xoá bỏ cột **‘*quarter’***

Chuyển giá trị trong cột quarter (ví dụ: Q1 2024) thành định dạng ngày bắt đầu quý (Ví dụ: 2024-01-01) bằng cách sử dụng hàm ‘**convert\_quarter\_to\_date’**.

Lưu kết quả trong cột mới ‘**date’.**

* Loại bỏ các cột không dùng đến nữa: ‘***CHỈ TIÊU***’, ‘***ACCOUNT\_NAME***’, ‘***SIGN***’, ‘***ACCOUNT\_TYPE’***, ‘***LEVEL***’.
* Thêm cột **stock\_name** ghi tên các mã cổ phiếu.

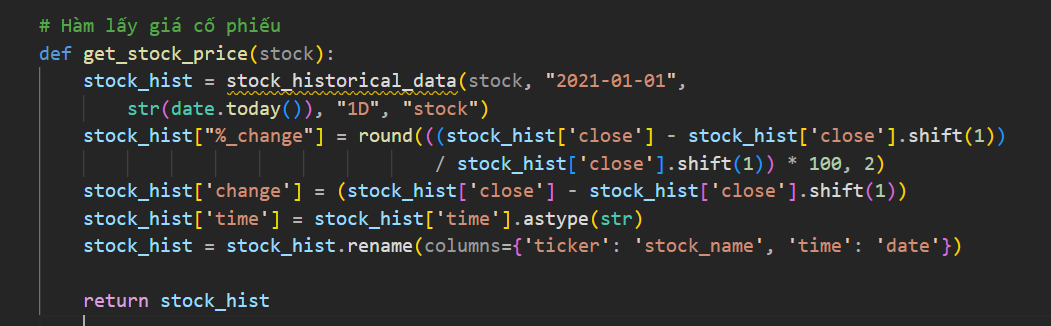


Hình 5. Bảng cân đối kế toán sau xử lí

### **4. Chuẩn bị dữ liệu cổ phần**

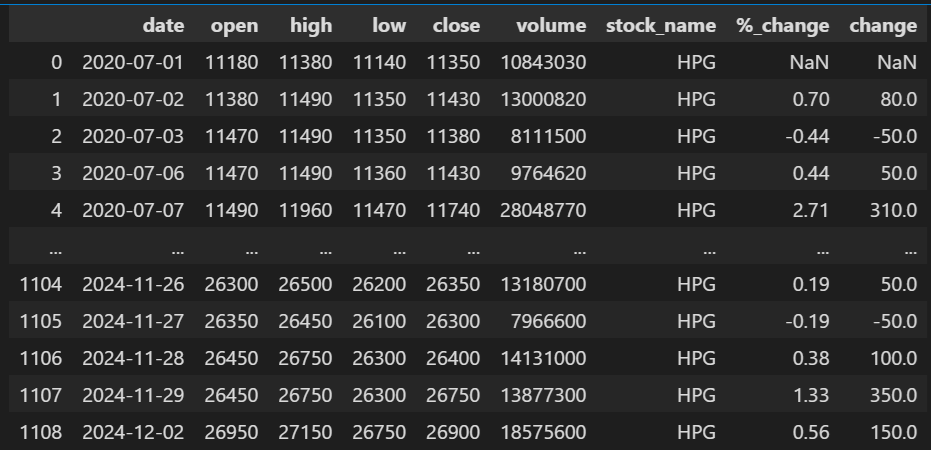
Sử dụng thư viện VNSTOCK để lấy dữ liệu

Tạo hàm **get\_stock\_price:** lấy ra các thông tin cơ bản và chỉ số tài chính của cổ phiếu



Hình 6. Hàm tự định nghĩa lấy ra giá cổ phiếu

Hàm này giúp chuẩn bị dữ liệu cổ phiếu với các chỉ số thay đổi giá, phục vụ cho việc phân tích xu hướng hoặc đánh giá biến động giá.



Hình 7. Bảng giá đóng/ mở cửa cổ phiếu

* ***Open****: giá cổ phiếu khi mở cửa phiên giao dịch*
* ***Close****: Giá cổ phiếu khi đống cửa phiên giao dịch*
* ***Low****: giá thấp nhất trong phiên giao dịch*
* ***High****: giá cao nhất phiên giao dịch*
* ***Volumn****: khối lượng cổ phiếu giao dịch trong phiên*
* ***Change****: mức chênh lệch giá khi mở và đóng cửa*
* ***% change****: chênh lệch giá mở và đóng cửa theo %*
* ***Date****: ngày ghi nhận giá cổ phiếu*

Đây là các thông tin cần thiết cho việc tạo ra Biểu đồ biến động giá cổ phiếu.

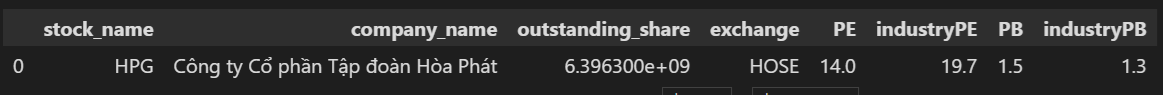
### **5. Chuẩn bị thông tin cơ bản về cổ phiếu**

Tạo hàm ***get\_stock\_fundamental\_info*:**

Sử dụng các hàm có sẵn trong thư viện **VNSTOCK**: ***company\_fundamental\_ratio*** *,* ***company\_profile, company\_overview, stock\_evaluation*.**



Hình 8. Hàm lấy các chỉ số kĩ thuật



Hình 9. Các thông tin kĩ thuật được lấy về

Lấy ra các thông tin cơ bản của các mã cổ phiếu:

* ***company\_fundamental\_ratio****: Lấy các chỉ số tài chính cơ bản của công ty (như P/E, P/B, ROE,...).*
* ***company\_profile****: Lấy thông tin hồ sơ cơ bản của công ty (tên, ngành nghề, địa chỉ,...).*
* ***company\_overview****: Lấy thông tin tổng quan về công ty (cổ phiếu lưu hành, sàn giao dịch,...).*
* ***stock\_evaluation****: Đánh giá cổ phiếu của công ty so với ngành (industryPE).*

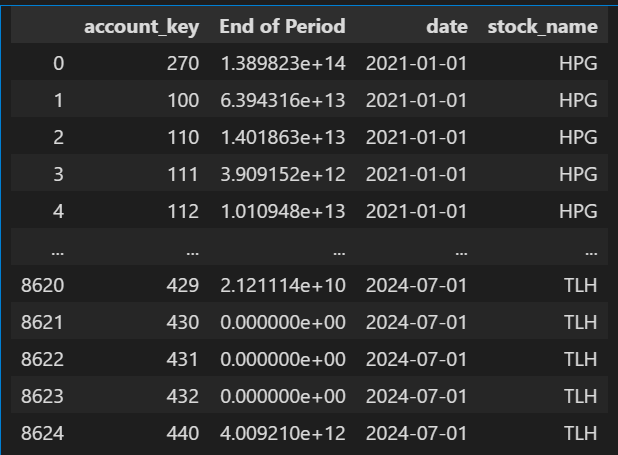
### **6. Tổng hợp các thông tin vừa xử lí vào một file excel**

Thực hiện việc tổng hợp và kết hợp dữ liệu từ nhiều cổ phiếu khác nhau:

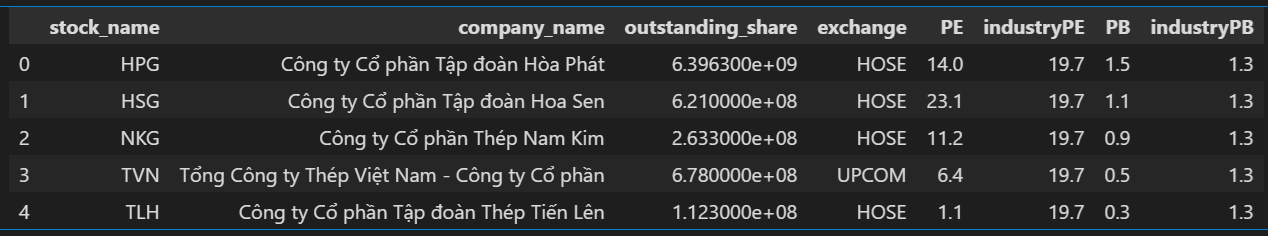
* **balance\_sheet:** Tạo báo cáo cân đối kế toán (Balance Sheet) cho cổ phiếu.
* **income\_statement:** Tạo báo cáo thu nhập (Income Statement) cho cổ phiếu.
* **stock\_price:** Lấy giá cổ phiếu theo thời gian (dữ liệu lịch sử).
* **stock\_info:** Lấy thông tin cơ bản về công ty, cùng với các chỉ số tài chính như P/E, P/B.

Sau đó, dữ liệu từ các báo cáo tài chính, giá cổ phiếu và thông tin cơ bản của từng cổ phiếu sẽ được lưu vào các danh sách (**bs\_list, is\_list, stock\_price\_list, stock\_info\_list)**.

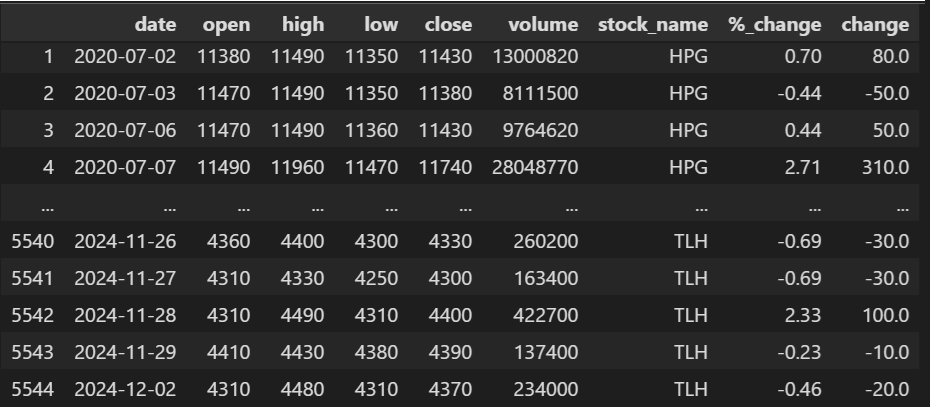
Cuối cùng, các danh sách này sẽ được kết hợp lại thành các DataFrame thông pd.concat, tạo ra các DataFrame tổng hợp cho tất cả các cổ phiếu.



Hình 10. Bảng cân đối kế toán (BS)



Hình 11. Bảng thông tin cơ bản của cổ phiếu (Stock\_info)



Hình 12. Bảng giá của 5 mã cổ phiếu (Stock\_Price)

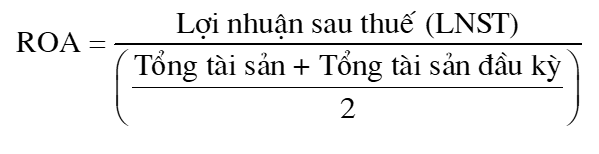
Cuối cùng sau khi xử lí xong ta cho các bảng tính vào thành một file excel ‘***data’*** tại folder “***Processed Data”.***

## **III. Phân tích kết quả tổng quan và chọn ra cổ phiếu tiềm năng**

### **1. Một số thông số cần lưu ý khi đọc biểu đồ**

*Chúng ta quan tâm tới một số chỉ số để đánh giá xem các mã cổ phiếu có triển vọng hay không.*Các chỉ số đó lần lượt là:

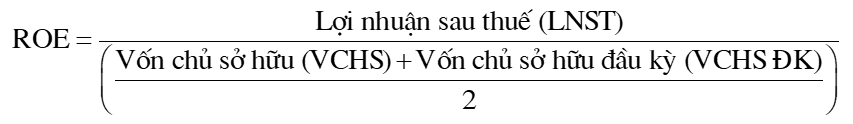
#### **1.1. ROA (Return on Assets) - Tỷ suất lợi nhuận trên tài sản**



Trong đó:

* **Lợi nhuận sau thuế (LNST)**: Là số tiền mà công ty thu được sau khi trừ đi tất cả các chi phí hoạt động, thuế và các khoản chi phí khác.
* **Tổng tài sản**: Là tổng giá trị tài sản của công ty tại cuối kỳ (hoặc cuối năm).
* **Tổng tài sản đầu kỳ**: Là tổng giá trị tài sản của công ty tại đầu kỳ (hoặc đầu năm).

#### **1.2. ROE (Return on Equity) - Tỷ suất lợi nhuận trên vốn chủ sở hữu**



Trong đó:

* **Lợi nhuận sau thuế (LNST)**: Là số tiền mà công ty thu được sau khi trừ đi tất cả các chi phí hoạt động, thuế và các khoản chi phí khác.
* **Vốn chủ sở hữu (VCHS)**: Là số tiền mà các cổ đông đầu tư vào công ty, thường được xác định trong báo cáo tài chính.
* **Vốn chủ sở hữu đầu kỳ (VCHS ĐK)**: Là số vốn chủ sở hữu của công ty tại đầu kỳ (hoặc đầu năm).

#### **1.3. PE và PB**

**PE thấp** và **PB thấp** có thể cho thấy cổ phiếu đang bị định giá thấp và có thể là cơ hội đầu tư tốt (mặc dù có thể có rủi ro).

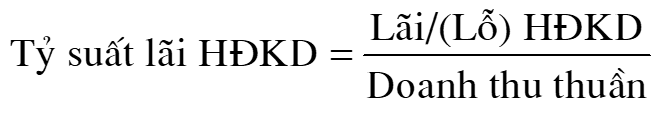
**PE cao** và **PB cao** có thể cho thấy cổ phiếu đang bị đánh giá cao, nhưng nếu công ty có tiềm năng tăng trưởng mạnh mẽ trong tương lai, thì đây có thể là một lựa chọn đầu tư hấp dẫn.

#### **1.4. Tỷ suất EBIT**

Trong đó:

* **LNTT** (Lợi nhuận trước thuế) là lợi nhuận của công ty trước khi tính thuế.
* **CPLV** (Chi phí lãi vay) là chi phí mà công ty phải trả cho các khoản vay hoặc nợ.
* **Doanh thu thuần** là tổng doanh thu mà công ty thu được sau khi trừ đi các khoản giảm trừ như chiết khấu, trả lại hàng, hoặc thuế doanh thu.

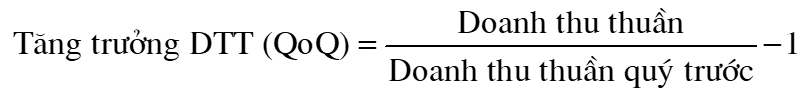
#### **1.5. Tỷ suất lãi HĐKD**



Trong đó:

* **Lãi/(Lỗ) HĐKD**: Là lợi nhuận hoặc lỗ từ hoạt động kinh doanh chính, tức là lợi nhuận trước thuế và lãi vay nhưng chỉ tính từ hoạt động kinh doanh (không tính đến thu nhập ngoài hoạt động kinh doanh như lãi từ đầu tư tài chính).
* **Doanh thu thuần**: Là tổng doanh thu mà công ty thu được từ bán hàng và cung cấp dịch vụ sau khi trừ các khoản giảm trừ như chiết khấu, trả lại hàng, hoặc thuế doanh thu.

#### **1.6. Tăng trưởng DTT**



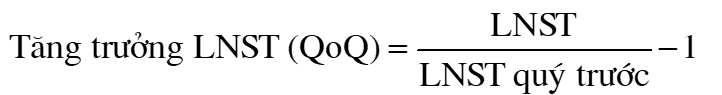
Trong đó:

* **Doanh thu thuần**: Là tổng doanh thu mà công ty thu được từ bán hàng và cung cấp dịch vụ sau khi trừ các khoản giảm trừ như chiết khấu, trả lại hàng, hoặc thuế doanh thu trong quý hiện tại.
* **Doanh thu thuần quý trước (Doanh thu thuần PQ)**: Là tổng doanh thu thuần của công ty trong quý liền kề trước đó (quý trước).

**Ý nghĩa:**

* **Tăng trưởng DTT QoQ > 0** (tăng trưởng dương): Cho thấy doanh thu của công ty trong quý hiện tại cao hơn quý trước, tức là công ty đang tăng trưởng về doanh thu.
* **Tăng trưởng DTT QoQ < 0** (giảm trưởng âm): Cho thấy doanh thu của công ty trong quý hiện tại thấp hơn quý trước, tức là công ty đang gặp khó khăn trong việc duy trì hoặc tăng trưởng doanh thu.
* **Tăng trưởng DTT QoQ = 0** (không thay đổi): Cho thấy doanh thu của công ty trong quý hiện tại không thay đổi so với quý trước.

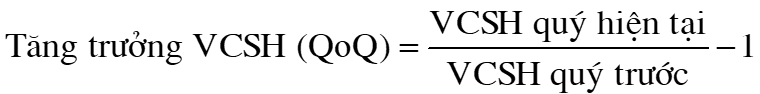
#### **1.7. Tăng trưởng LNST**



Trong đó:

* **LNST**: Lợi nhuận sau thuế của công ty trong quý hiện tại.
* **LNST quý trước (LNST PQ)**: Lợi nhuận sau thuế của công ty trong quý liền kề trước đó (quý trước).

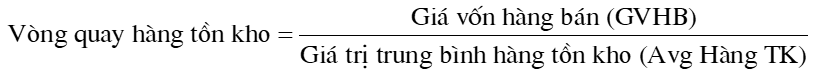
#### **1.8. Tăng trưởng VCSH**



Trong đó:

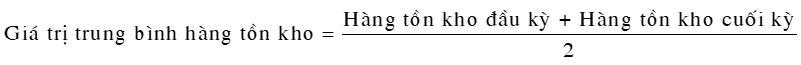
* **VCSH (Vốn chủ sở hữu)**: Là tổng giá trị tài sản thuộc sở hữu của cổ đông, bao gồm vốn điều lệ, lợi nhuận chưa phân phối, và các khoản đóng góp từ cổ đông.
* **VCSH quý hiện tại (VCHS)**: Là giá trị vốn chủ sở hữu của công ty trong quý hiện tại.
* **VCSH quý trước (VCHS ĐK)**: Là giá trị vốn chủ sở hữu của công ty trong quý trước (hoặc tại thời điểm đầu kỳ của quý hiện tại).

#### **1.9. Vòng quay hàng tồn kho**



Trong đó:

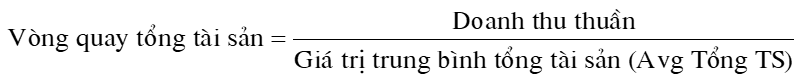
* **Giá vốn hàng bán (GVHB)**: Là chi phí của hàng hóa đã bán trong kỳ (thường là trong năm hoặc quý). Đây là chi phí sản xuất hoặc mua sắm các hàng hóa mà công ty đã bán cho khách hàng.
* **Giá trị trung bình hàng tồn kho (Avg Hàng TK)**: Là giá trị trung bình của hàng tồn kho trong kỳ. Giá trị này có thể được tính bằng cách lấy giá trị tồn kho đầu kỳ và cuối kỳ rồi chia cho 2:



**Ý nghĩa của chỉ số Vòng quay hàng tồn kho:**

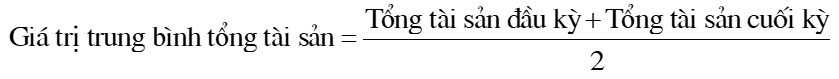
* **Vòng quay hàng tồn kho cao**: Thể hiện rằng công ty có khả năng bán nhanh hàng tồn kho, tức là công ty đang sử dụng nguồn lực tài chính một cách hiệu quả để duy trì mức tồn kho thấp và nhanh chóng chuyển thành doanh thu.
* **Vòng quay hàng tồn kho thấp**: Có thể chỉ ra rằng công ty gặp khó khăn trong việc bán hàng, dẫn đến lượng hàng tồn kho lớn và có thể gây áp lực tài chính cho công ty (chi phí lưu kho, giảm giá trị hàng tồn kho do lỗi mốt hoặc hư hỏng).

#### **1.10. Vòng quay tổng tài sản**



Trong đó:

* **Doanh thu thuần**: Là doanh thu từ hoạt động bán hàng của công ty, sau khi trừ đi các khoản giảm trừ (chiết khấu, hoàn trả hàng, giảm giá, v.v.). Đây là doanh thu thực tế mà công ty thu về từ các hoạt động kinh doanh chính.
* **Giá trị trung bình tổng tài sản (Avg Tổng TS)**: Là giá trị trung bình của tổng tài sản trong kỳ. Giá trị này có thể được tính bằng cách lấy tổng tài sản đầu kỳ và cuối kỳ rồi chia cho 2:



**Ý nghĩa của chỉ số Vòng quay tổng tài sản:**

* **Vòng quay tổng tài sản cao**: Cho thấy công ty sử dụng tài sản của mình một cách hiệu quả để tạo ra doanh thu. Điều này có thể chỉ ra rằng công ty có khả năng sinh lời cao từ tài sản hiện có.
* **Vòng quay tổng tài sản thấp**: Có thể chỉ ra rằng công ty không tận dụng hết tiềm năng tài sản của mình để tạo ra doanh thu, hoặc tài sản của công ty đang bị lãng phí.

#### **1.11. Vòng quay VCSH**



Trong đó:

* **Doanh thu thuần**: Là doanh thu thực tế mà công ty thu về từ các hoạt động kinh doanh chính, sau khi trừ đi các khoản giảm trừ như chiết khấu, hoàn trả hàng, giảm giá, v.v.
* **Giá trị trung bình VCSH (Avg VCSH)**: Là giá trị trung bình của vốn chủ sở hữu trong kỳ. Giá trị này được tính bằng cách lấy tổng vốn chủ sở hữu đầu kỳ và cuối kỳ rồi chia cho 2:

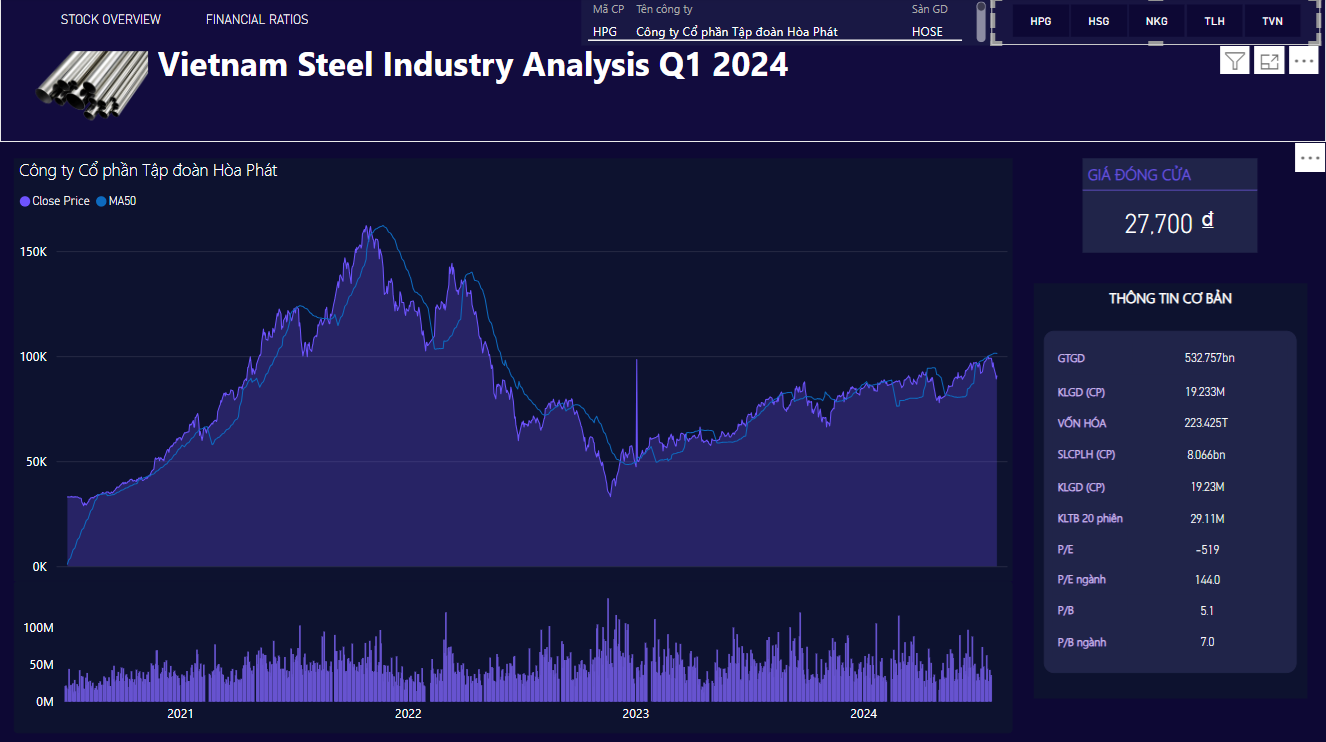


**Ý nghĩa của chỉ số Vòng quay VCSH:**

* **Vòng quay VCSH cao**: Chỉ ra rằng công ty sử dụng vốn chủ sở hữu hiệu quả để tạo ra doanh thu. Công ty có thể tạo ra nhiều doanh thu từ mỗi đồng vốn chủ sở hữu.
* **Vòng quay VCSH thấp**: Cho thấy công ty có thể không sử dụng vốn chủ sở hữu của mình hiệu quả, có thể dẫn đến việc không tối ưu hóa tài sản và vốn trong hoạt động kinh doanh.

### **2. TỔNG QUAN 5 MÃ CỔ PHIẾU**

#### **2.1. Công ty Cổ phần Tập đoàn Hòa Phát (HPG):**



Hình 13. Tổng quan về Cổ phiếu HPG

#### **2.2. Công ty Cổ phần Tập đoàn Hoa Sen (HSG):**



Hình 14. Tổng quan về cổ phiếu HSG

#### **2.3. Công ty Cổ phần Thép Nam Kim (NKG):**



Hình 15. Tổng quan về cổ phiếu NKG

#### **2.4. Tổng Công ty Thép Việt Nam - Công ty Cổ phần (TVN)**



Hình 16. Tổng quan về cổ phiếu TLN

#### **2.5. Công ty Cổ phần Tập đoàn Thép Tiến Lên (TLH):**



Hình 17. Tổng quan về cổ phiếu TLH

### **3. Phân tích và lựa chọn cổ phiếu tiềm năng**



Hình 18. Bảng so sánh các chỉ số kĩ thuật giữa 5 mã cổ phiếu

#### **3.1. Tóm tắt xu hướng kỹ thuật**

Dựa trên biểu đồ giá ***HPG, HSG*** và ***NKG*** vẫn tiếp tục duy trì trong xu hướng tăng dài hạn. Các đợt giảm giá gần đây đi kèm với thanh khoản thấp, cho thấy lực bán không mạnh và tiềm năng phục hồi. Đây có thể là cơ hội tốt để các nhà đầu tư mua thăm dò khi giá cổ phiếu tiếp cận các mức hỗ trợ trong kênh tăng giá dài hạn. Trong khi đó, cổ phiếu ***TLH*** vẫn trong giai đoạn đi ngang (sideways), dù có những phiên tăng giá mạnh với khối lượng giao dịch lớn nhưng chưa vượt qua được nền giá. Cổ phiếu ***TVN*** sau khi vượt nền và tăng 50% đã quay đầu giảm, điều này có thể gợi ý rằng mức giá đã đạt đến ngưỡng kháng cự mạnh, cần theo dõi thêm để đánh giá tiềm năng tăng trưởng tiếp theo.

***TVN*** có chỉ số vòng quay hàng tồn kho cao nhất trong ngành, đạt 1.52, cho thấy khả năng tiêu thụ hàng tồn kho nhanh. ***HSG*** có vòng quay tổng tài sản cao nhất với 0.48, chỉ ra hiệu quả sử dụng tài sản của công ty. ***NKG*** dẫn đầu về vòng quay vốn chủ sở hữu với 0.96, tiếp theo là HSG với 0.84, cho thấy khả năng tạo ra lợi nhuận trên vốn chủ sở hữu của các công ty này.

#### **3.2. Chỉ tiêu định giá**

PE trung bình của ngành thép là 28.8, với ***HPG*** và ***HSG*** lần lượt có PE là 19 và 16, cho thấy định giá của ***HPG*** và ***HSG*** thấp hơn so với trung bình ngành, có thể là cơ hội đầu tư hấp dẫn. PB của ***HPG*** là 1.7, cao nhất trong ngành, so với mức trung bình là 1.4, cho thấy sự đánh giá cao của thị trường đối với giá trị sổ sách của ***HPG***.

#### **3.3. Khả năng quản lý**

ROE của ***HSG*** và ***HPG*** lần lượt đạt 2.9% và 2.75%, cho thấy khả năng sinh lợi trên vốn chủ sở hữu của các công ty này là tốt nhất trong ngành. ROA của ***HPG*** đạt 1.75%, cao nhất trong ngành, tiếp theo là ***HSG*** với 1.67%.

#### **3.4. Tốc độ tăng trưởng**

***NKG*** dẫn đầu về tăng trưởng lợi nhuận sau thuế với 568.65%, cho thấy sự cải thiện mạnh mẽ trong hiệu suất kinh doanh. ***HSG*** cũng có mức tăng trưởng ấn tượng với 208.85%, trong khi ***TVN*** có mức giảm -510.45%.

#### **3.5. Khả năng sinh lợi**

Tỷ suất EBIT của ***HPG*** là 13%, cao nhất trong ngành, cho thấy khả năng sinh lợi từ hoạt động kinh doanh chính của công ty. ***NKG*** và ***HSG*** đều có tỷ suất EBIT là 4%, một mức tích cực. Tỷ suất lãi hoạt động kinh doanh của ***HPG*** đạt 9%, cao hơn hẳn so với ***NKG*** và ***HSG*** (đều là 4% và 3%).

#### **3.6. Khả năng thanh toán**

***TLH*** và ***HSG*** có chỉ số thanh toán hiện hành cao nhất (~1.5), thể hiện mức độ an toàn tài chính ổn của 2 DN. ***HPG*** có chỉ số thanh toán nhanh cao nhất (0.65), cho thấy khả năng thanh toán nợ ngắn hạn mà không cần bán hàng tồn kho. Chỉ số thanh toán tức thời nhanh của ***HPG*** cũng dẫn đầu với 0.16, cho thấy tình trạng tài chính lành mạnh.

#### **3.7. Kết luận:**

Dựa trên các chỉ số kỹ thuật và tài chính, ***HPG*** và ***HSG*** là những cổ phiếu tiềm năng cho quý 1/2024. ***HPG*** có tỷ suất EBIT và tỷ suất lãi hoạt động kinh doanh cao nhất trong ngành, cùng với khả năng quản lý tài chính hiệu quả và khả năng thanh khoản tốt. HSG nổi bật với khả năng quay vòng tài sản và vốn chủ sở hữu, cũng như tốc độ tăng trưởng lợi nhuận mạnh mẽ. Các yếu tố này làm cho ***HPG*** và ***HSG*** có tiềm năng trở thành các lựa chọn đầu tư hấp dẫn trong bối cảnh ngành thép hiện tại.

## **IV. Áp dụng mô hình học máy LSTM**

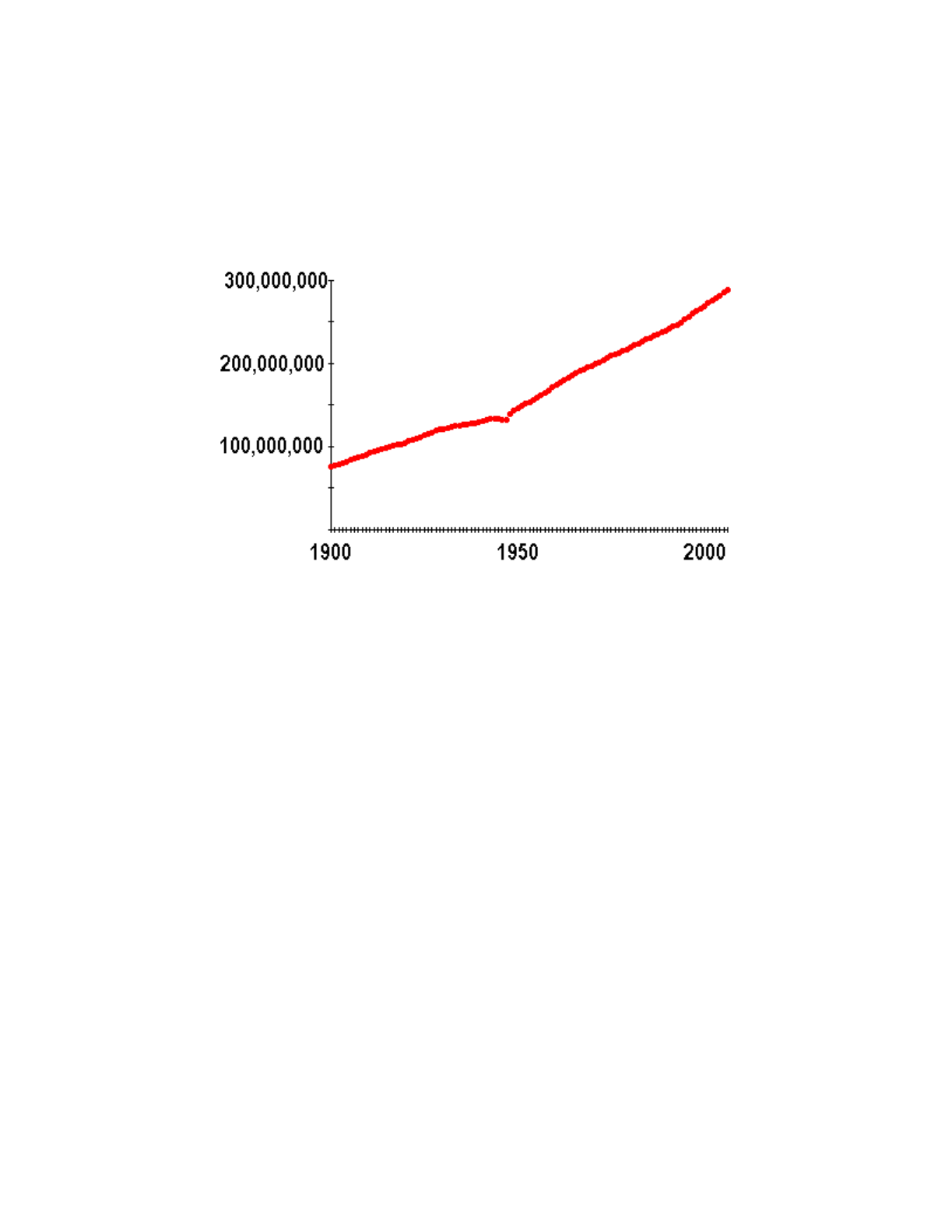
### **1. Tổng quan về Dự báo thời gian và LSTM**

#### **1.1. Phương pháp dự báo theo thời gian**

Để muốn phân tích một thứ gì đó hay dự đoán một điều gì như chứng khoán, giá vàng, thời tiết, bệnh tim mạch, giá xe, giá nhà thì chúng ta đều cần xác định kiểu tập dữ liệu mà chúng ta muốn dự đoán. Vì thế mỗi tập dữ liệu chúng ta muốn dự đoán và phân tích đều có mỗi dạng khác nhau, nên phương pháp dự đoán và phân tích đều có cách riêng biệt cho mỗi kiểu dữ liệu chúng ta thực hiện với chúng. Chẳng hạn, ta muốn dự đoán giá của một căn nhà dựa vào dữ liệu ta có là : phòng tắm, phòng ngủ, diện tích, vị trí,… Và cũng từ những giá trị đó ta có giá nhà cụ thể là cao hay thấp, giả sử ta có tập dữ liệu về nhiều giá nhà dựa trên các giá trị đó thì nhìn chung kiểu dữ liệu mà ta muốn dự đoán là một tập dữ liệu gồm nhiều giá trị với nhau và từ đó cho ra giá nhà cụ thể, đối với tập dữ liệu như trên thì thông thường phương pháp là sẽ sử dụng Linear Regresstion ( Hồi quy tuyến tính ) để dự đoán nhà mà ta muốn.

Đối tượng trong lần này là dự đoán xu hướng giá chứng khoán. Chúng ta có thể hiểu chứng khoán là một kiểu dữ liệu Time Series. Time Series là một chuỗi các điểm dữ liệu xảy ra theo thứ tự liên tiếp trong một khoảng thời gian. Trong đầu tư, một chuỗi thời gian theo dõi chuyển động của các điểm dữ liệu đã chọn, chẳng hạn như chứng khoán, trong một khoản thời gian xác định với các điểm dữ liệu được ghi lại theo các khoảng thời gian xác định với các điểm dữ liệu được ghi lại theo các khoảng thời gian đều đặn. Không có thời gian tối thiểu hoặc tối đa phải được đưa vào, cho phép dữ liệu được thu thập theo cách nhà cung cấp thông tin mà nhà đầu tư hoặc nhà phân tích tìm kiếm khi kiểm tra hoạt động. Điều này có thể được theo dõi trong suốt thời gian ngắn, chẳng hạn như giá của một chứng khoán vào giờ trong suốt một ngày làm việc hoặc thời gian dài và giá của một chứng khoán vào lúc đóng cửa vào ngày cuối cùng của mỗi tháng trong quá trình 2 năm.

Phân tích Time Series có thể hữu ích để xem là một tài sản, chứng khoán hoặc biến số kinh tế nhất định thay đổi như thế nào theo thời gian. Nó cũng có thể được sử dụng để kiểm tra xem những thay đổi liên quan đến điểm dữ liệu đã chọn so với những thay đổi của các biến khác nhau như thế nào trong một khoảng thời gian. Time Series được sử dụng trong một số bối cảnh phi tài chính, chẳng hạn như đo lường sự thay đổi dân số theo thời gian. Hình dưới đây mô tả một chuỗi thời gian như vậy với sự tăng trưởng nền kinh tế Hoa Kỳ dân số trong suốt những năm 1900 đến năm 2000.



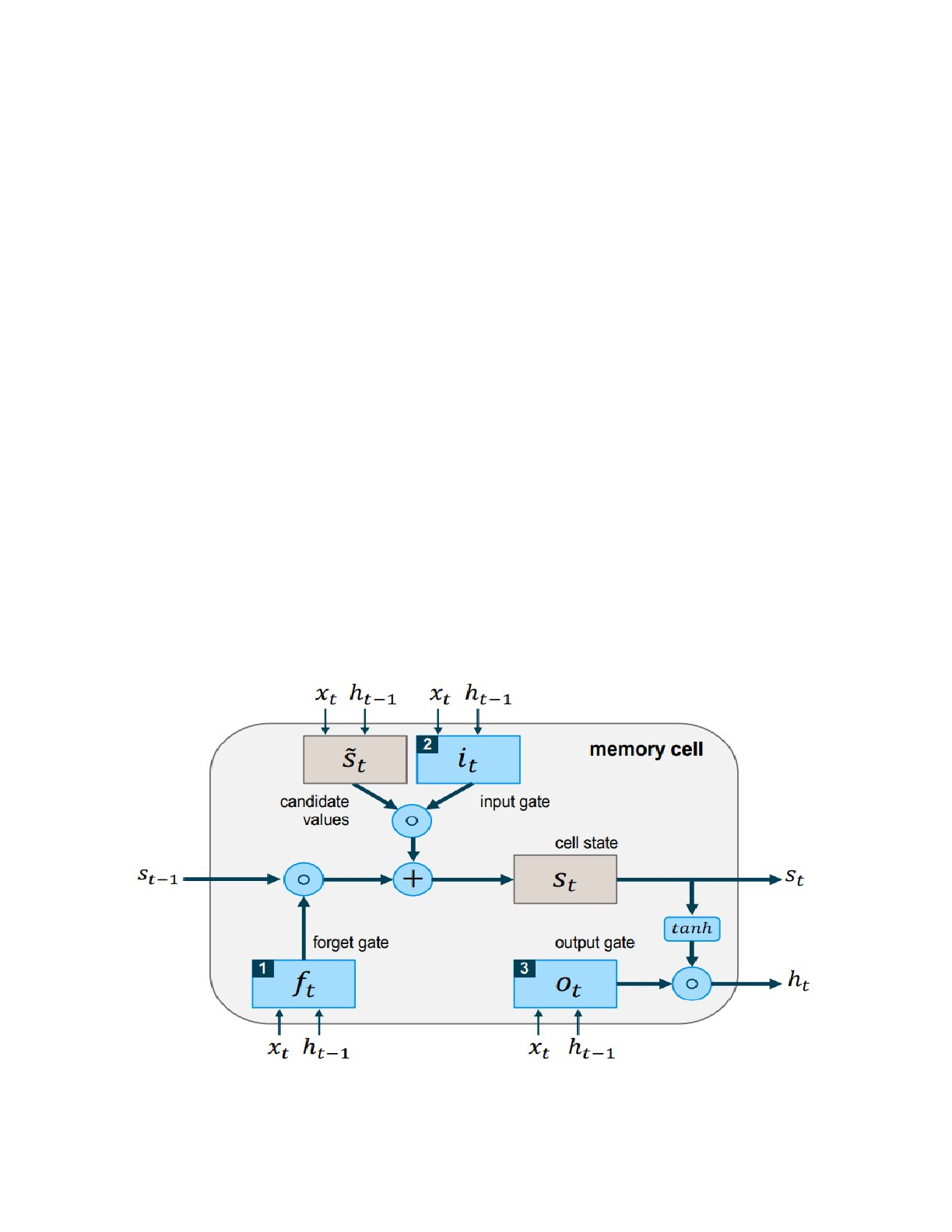
Hình 19. Tăng trưởng kinh tế Hoa Kì (1900-2000)

#### **1.2. Giới thiệu về LSTM**

Đối với dạng bài toán có **Time Series** thì sẽ sử dụng mạng Neural mô hình **Long Short-Term Memory** ( **LSTM** ) để dự đoán giá chứng khoán. **LSTM** là một dạng đặt biệt của mạng nơ-ron hồi quy **RNN** ( **Recurrent Neural Network** ). Để nói Long Short- Term Memory là một trong những mạng thần kinh nhân tạo được sử dụng phổ biến trong phân tích dữ liệu chuỗi thời gian ( Time Series ) thì có những đặc điểm sau :

* **LSTM** được thiết kế để tránh vấn đề phụ thuộc xa ( **Long-Term Dependency** ).
* **LSTM** cho phép lưu giữ thông tin qua nhiều bước thời gian mà không bị mất đi do các vấn đề như vanishinggradient thường xảy ra trong **RNN**.Điều này được thực hiện nhờ vào cách các trạng thái bộ nhớ và các cổng được huấn luyện để quyết định thông tin nào cần nhớ, quên hoặc cập nhật.

Trong kiến trúc mạng RNN chuẩn, các mô-dun này có cấu trúc rất đơn giản, thường là một tầng hàm tanh. LSTM cũng có kiến trúc dạng chuỗi như vậy, nhưng các mô-đun trong nó có cấu trúc khác với mạng RNN chuẩn. Thay vì một tầng mạng nơ-ron, chúng có tới 4 tầng tương tác với nhau một cách đặc biệt.



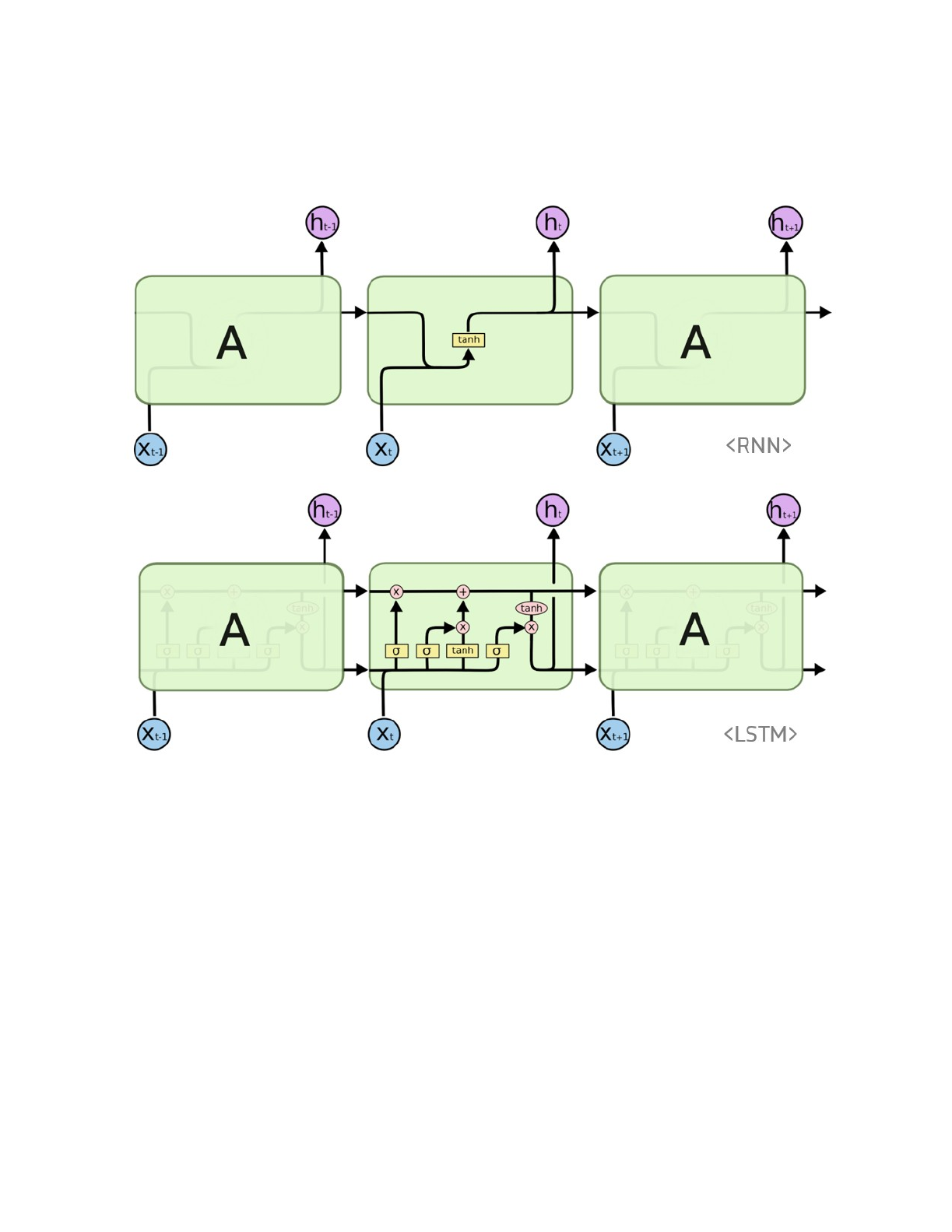
Hình 20 Biễu diễn kiến trúc bên trong của một Cell LSTM.

Để giải quyết bài toán dự đoán giá chứng khoán. Thì sẽ dùng mạng Neural mô hình LSTM để dự đoán giá trị cổ phiếu trong ngày tiếp theo . Sử dụng giá đóng cửa phiên giao dịch để dự đoán. Giá cổ phiếu là một chuỗi giá trị thay đổi theo thời gian, với P(i) là giá đóng cửa của ngày i trong chuỗi thời gian. Ý tưởng là sử dụng dữ liệu của cửa sổ thời gian cố định w ngày để dự đoán giá đóng cửa trong ngày w+1. Cửa sổ trượt theo trục thời gian với mỗi bước tiến đúng bằng kích thước w, không để xảy ra sự chồng chéo dữ liệu. Mô hình LSTM sẽ học từ các cửa sổ trượt này và dự đoán giá cổ phiếu của ngày tiếp theo. Dữ liệu từ P(i−w) đến P(i−1) được dùng để dự đoán P(i). Mô hình LSTM học các mối quan hệ phụ thuộc theo thời gian để tối ưu hóa việc dự đoán giá trong tương lai.

### **2. Xây dựng mô hình LSTM**

#### **2.1. Kiến trúc mô hình**

Để hiểu rõ thuật toán của mô hình LSTM xem hình bên dưới cho thấy cách hoạt động chi tiết của một Cell LSTM.



Hình 21. Mô hình sự khác nhau giữa RNN và LSTM

***Mạng LSTM có 3 cổng quan trọng:***

- Forget gate : Có nhiệm vụ loại bỏ những thông tin không cần thiết nhận được khỏi Cell Internal Sate.

- Input gate : Có nhiệm vụ chọn lọc những thông tin cần thiết khi nào được thêm vào Cell Intertal sate.

- Output gate : Có nhiệm vụ xác định những thông tin nào từ Cell Internal Sate được sử dụng như đầu ra.

***Các bước của thuật toán :***

- Ở bước đầu tiên, LSTM Cell quyết định những thông tin nào cần được loại bỏ từ Cell Internal State của bước thời gian trước đó S(*t-1)*. Activation value Ft của ForgetGate tại bước thời gian *t* được tính dựa trên dữ liệu đầu vào hiện tại X*t* , đầu ra của LSTM ở bước trước đó H(*t-1)*, và bias của Forget Gate. Hàm sigmoid được sử dụng để biến đổi giá trị kích hoạt này về một miền trong khoảng từ 0 (hoàn toàn loại bỏ thông tin) đến 1 (hoàn toàn giữ lại thông tin).



- Ở bước thứ hai, Cell LSTM quyết định những thông tin nào sẽ được thêm vào Cell Internal State S*t* ​. Bước này bao gồm hai quá trình: tính toán giá trị  và Forget gate F*t* . Giá trị  đại diện cho các thông tin tiềm năng cần được thêm vào Cell Internal Sate và được tính như sau:



Giá trị kích hoạt I*t* của Input Gate được tính như sau:



- Ở bước thứ ba, giá trị mới của trạng thái bên trong S*t* ​ được tính dựa trên kết quả từ các bước trước, sử dụng phép nhân Hadamard theo từng phần tử (Hadamard product), được ký hiệu là o :



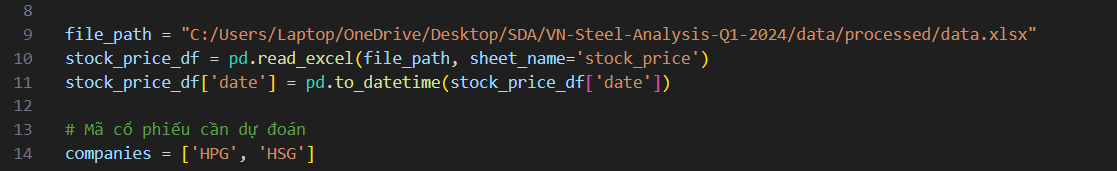
- Ở bước cuối cùng , giá trị đầu ra h*t* của Cell LSTM được tính toán dựa theo hai phương trình sau :





#### **2.2. Xây dựng và huấn luyện mô hình**

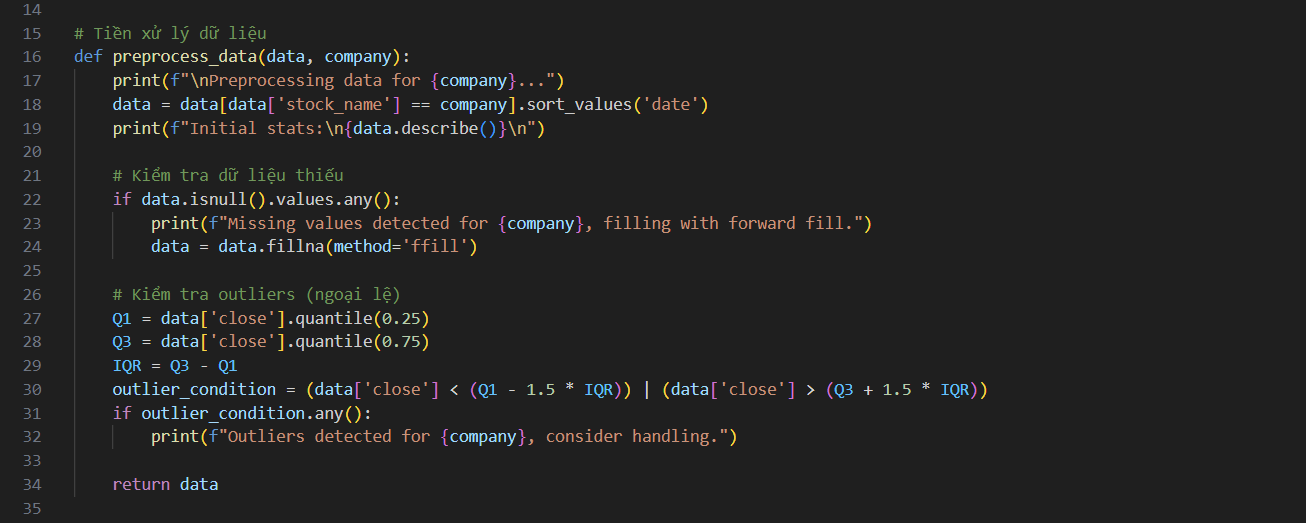
1. **Đọc, xử lý dữ liệu và lựa chọn mã cổ phiếu cần dự đoán**



Hình 22. Đọc dữ liệu giá cổ phiếu

* Đọc file ‘***data.xlsx’*** nhờ thư viện **Pandas** ( pd.read\_excel) và đặt lại định dạng thời gian của tập dữ liệu. Định dạng ngày tháng giúp sắp xếp và xử lý dữ liệu theo thứ tự thời gian.
* **Chọn mã cổ phiếu cần dự đoán gồm : “HPG” và “HSG”.**

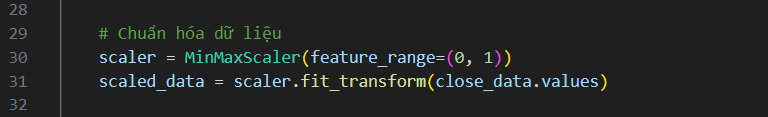
1. **Tiền xử lý dữ liệu**



Hình 23. Xử lý dữ liệu sai thiếu

* Các bước thực hiện:
  + Lọc dữ liệu theo mã cổ phiếu.
  + Kiểm tra và xử lý giá trị thiếu: Sử dụng phương pháp forward fill (fillna(method='fill')).
  + Phát hiện outliers (ngoại lệ): Sử dụng Interquartile Range (IQR). Nếu phát hiện ngoại lệ, hiện thông báo.
* Ưu điểm:
  + Bảo đảm dữ liệu đầy đủ và sạch.
  + Giảm thiểu ảnh hưởng của các giá trị bất thường đến mô hình.

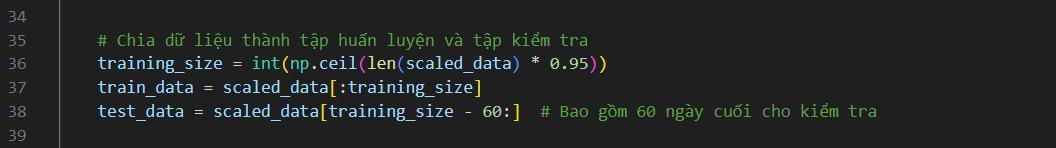
1. **Chuẩn hóa dữ liệu**



Hình 24. Chuẩn hoá MinMaxScaler

* Chuẩn hóa: Sử dụng MinMaxScaler để chuẩn hóa dữ liệu giá đóng cửa về phạm vi từ (0,1). Giúp các thuật toán học máy dễ dàng hơn và tăng hiệu quả tính toán.

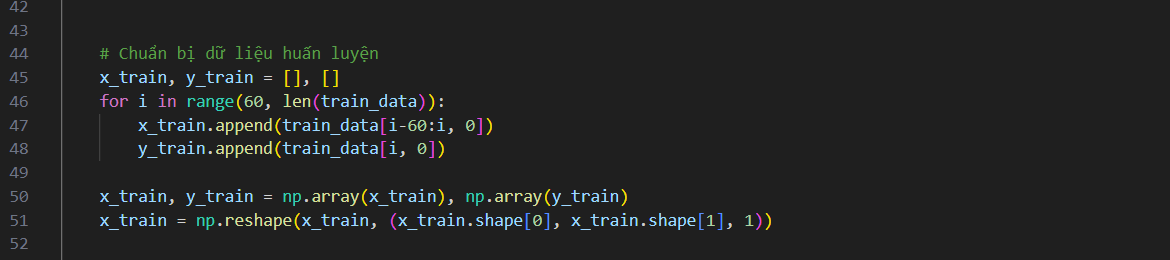
1. **Chia dữ liệu thành tập huấn luyện và kiểm tra**



Hình 25. Chia tập dữ liệu huấn luyện

* Tập huấn luyện: Chiếm 95% tổng số dữ liệu, còn lại 5% dùng để kiểm tra mô hình.
* Tập kiểm tra: Tập kiểm tra có kích thước 60 ngày cuối từ dữ liệu sau khi chia.

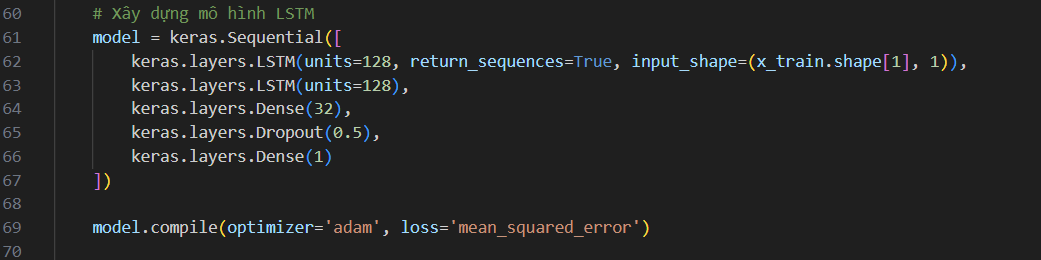
1. **Chuẩn bị dữ liệu huấn luyện**



Hình 26. Dùng dữ liệu 60 ngày để huấn luyện

* Tạo cửa sổ trượt: Sử dụng 60 ngày giá cổ phiếu trước đó để dự đoán giá đóng cửa của ngày tiếp theo.
* Dữ liệu huấn luyện: Mỗi cửa sổ dữ liệu có 60 giá trị đóng cửa (được dùng làm đầu vào x\_train), và giá đóng cửa của ngày tiếp theo là mục tiêu y\_train.

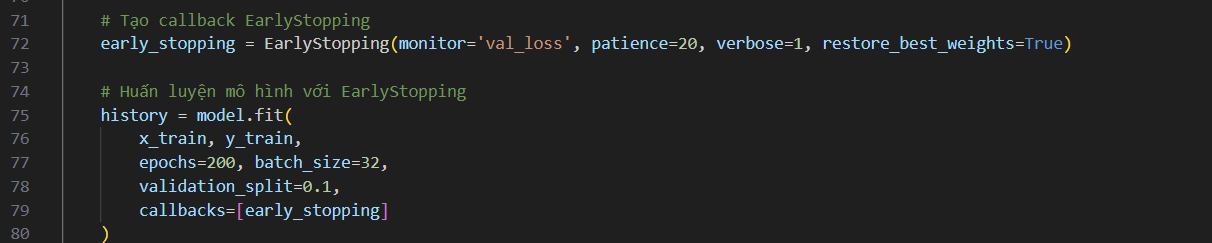
1. ***Xây dựng mô hình LSTM***



Hình 27. Phân lớp dữ liệu

* Mô hình LSTM:
* Hai lớp LSTM với 128 units giúp mô hình học các mối quan hệ dài hạn trong chuỗi thời gian.
* Một lớp Dense với 32 nút, sau đó là Dropout để giảm overfitting.
* Lớp Dense cuối cùng có 1 đơn vị để dự đoán giá cổ phiếu trong ngày tiếp theo.
* Biên dịch mô hình: Sử dụng bộ tối ưu hóa Adam và hàm mất mát mean\_squared\_error.

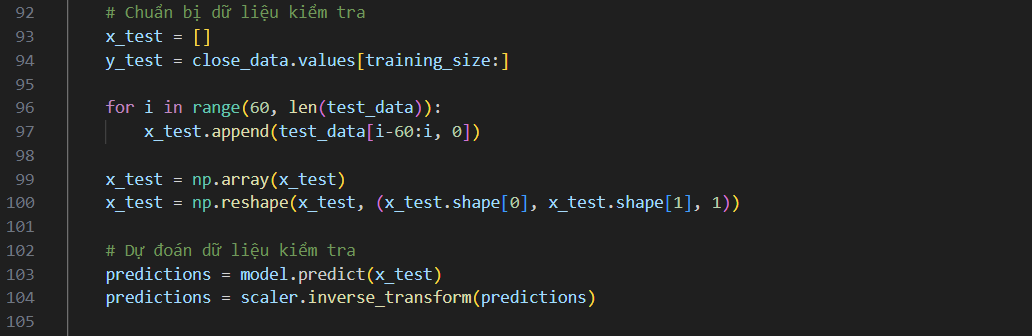
1. Huấn luyện mô hình



Hình 28. Thiết lập huấn luyện mô hình

* EarlyStopping: Nếu mô hình không cải thiện sau 20 epochs, quá trình huấn luyện sẽ dừng lại sớm để tránh overfitting.
* Huấn luyện: Mô hình được huấn luyện trong 200 epochs, với batch size là 16 và sử dụng 10% dữ liệu cho tập validation

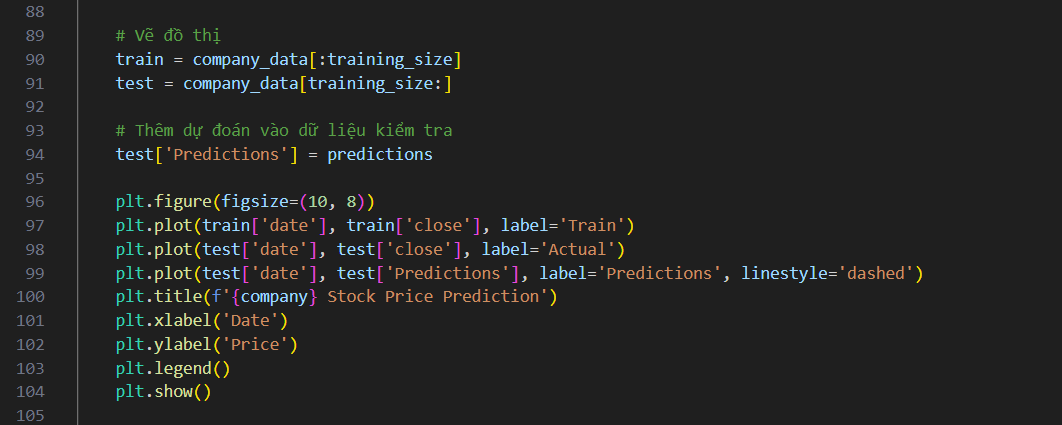
1. **Chuẩn bị và dự đoán dữ liệu kiểm tra**



Hình 29. Chuẩn bị dữ liệu kiểm tra

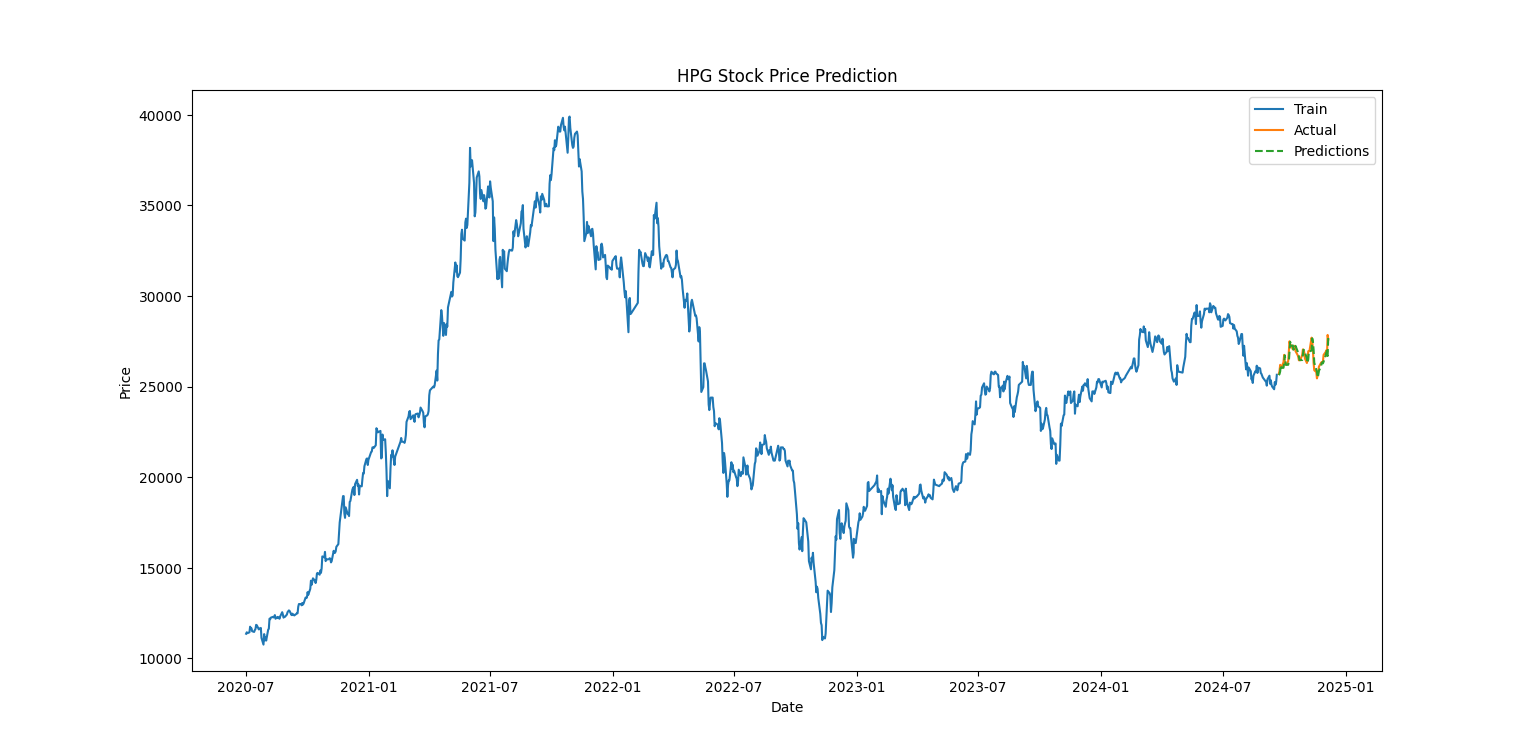
* Dữ liệu kiểm tra: Dữ liệu kiểm tra được chuẩn bị theo cách tương tự như dữ liệu huấn luyện, với 60 ngày trước đó làm đầu vào.
* Dự đoán: Mô hình dự đoán giá đóng cửa cho dữ liệu kiểm tra và giá trị dự đoán được đảo ngược chuẩn hóa (từ phạm vi [0,1] trở lại giá trị thực).

#### **2.3. Kết quả**

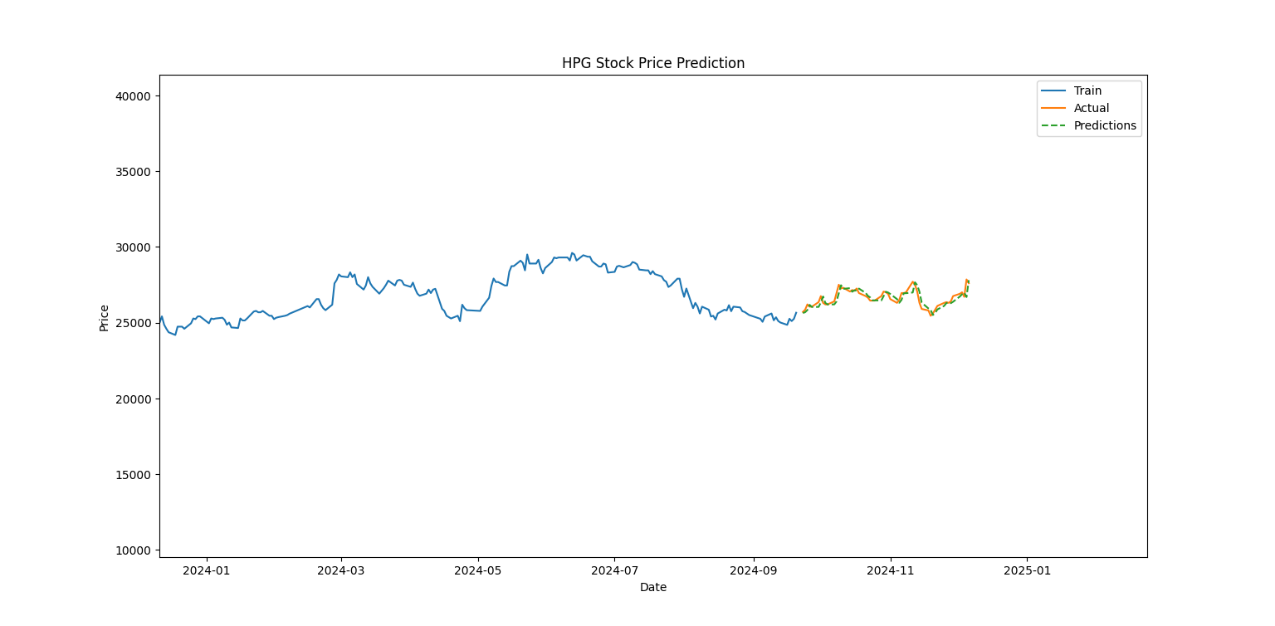


Hình 30. Vẽ biểu đồ giá cổ phiếu

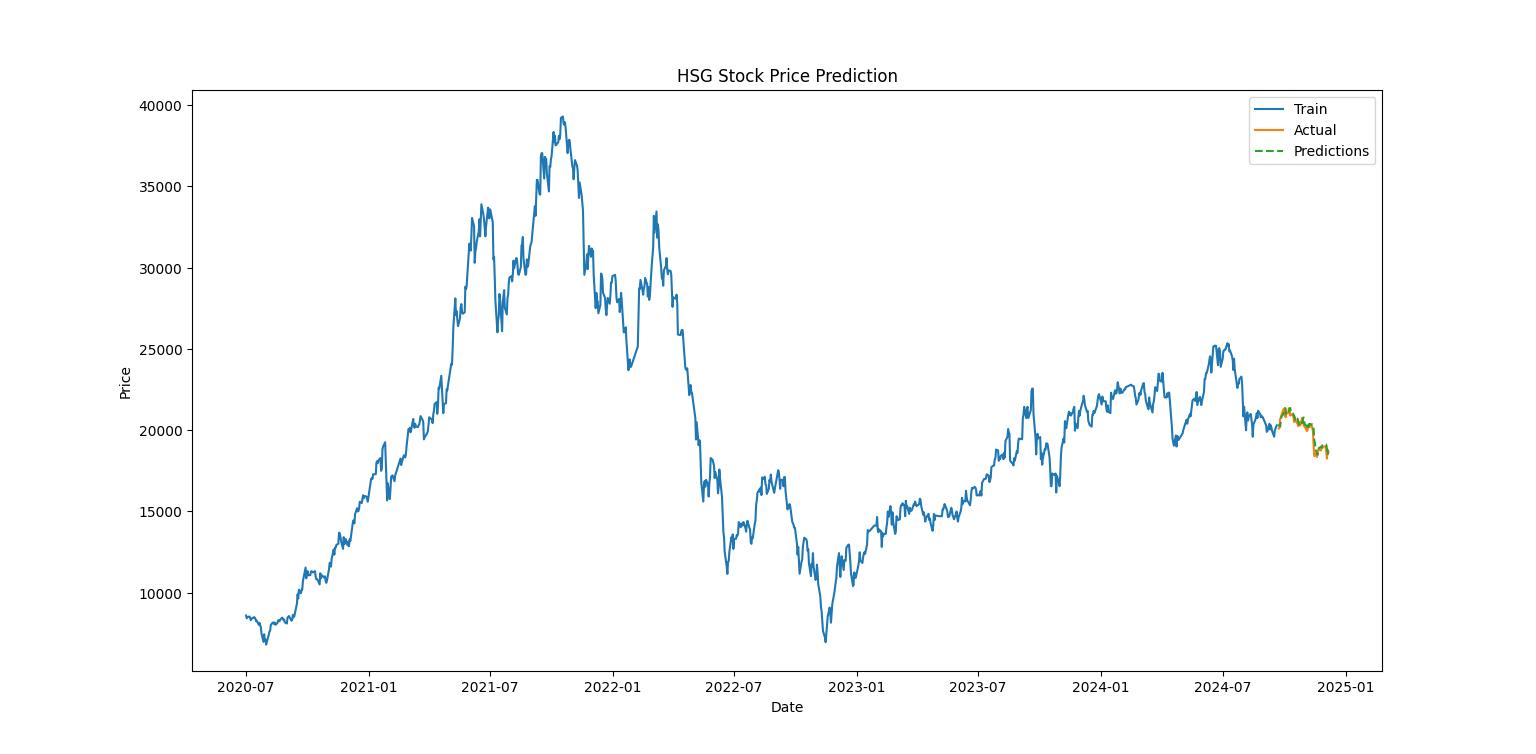
Vẽ đồ thị thể hiện giá cổ phiếu thực tế và dự đoán. Dữ liệu huấn luyện được vẽ với màu sắc khác, giá trị thực tế và dự đoán được vẽ để so sánh.



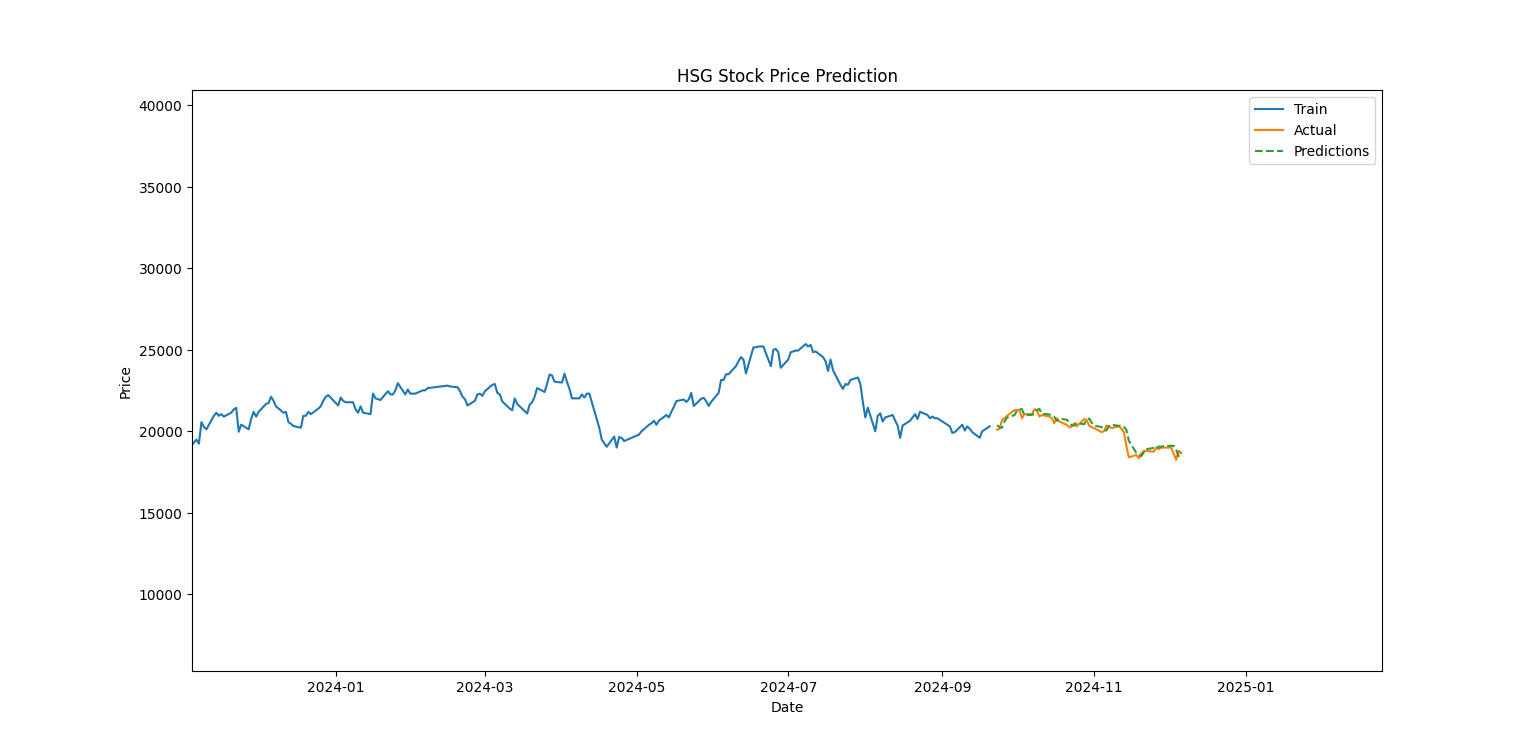
Hình 31. Giá cổ phiếu HPG từ 01-07-2020 đến nay



Hình 32. Giá cổ phiếu HPG từ đầu 2024 tới năm

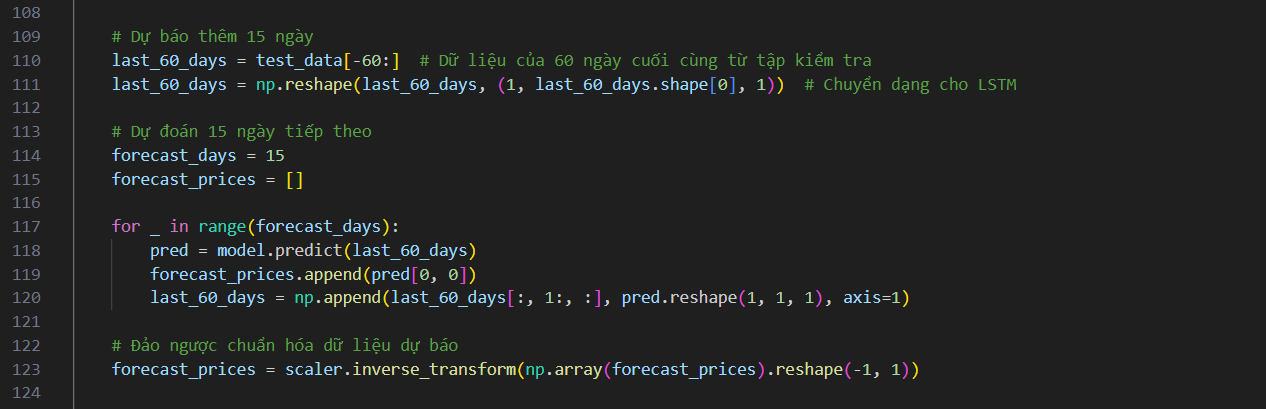


Hình 33. Giá cổ phiếu HSG từ 01-07-2020 đến nay



Hình 34. Giá cổ phiếu HSG từ đầu 2024 đến nay

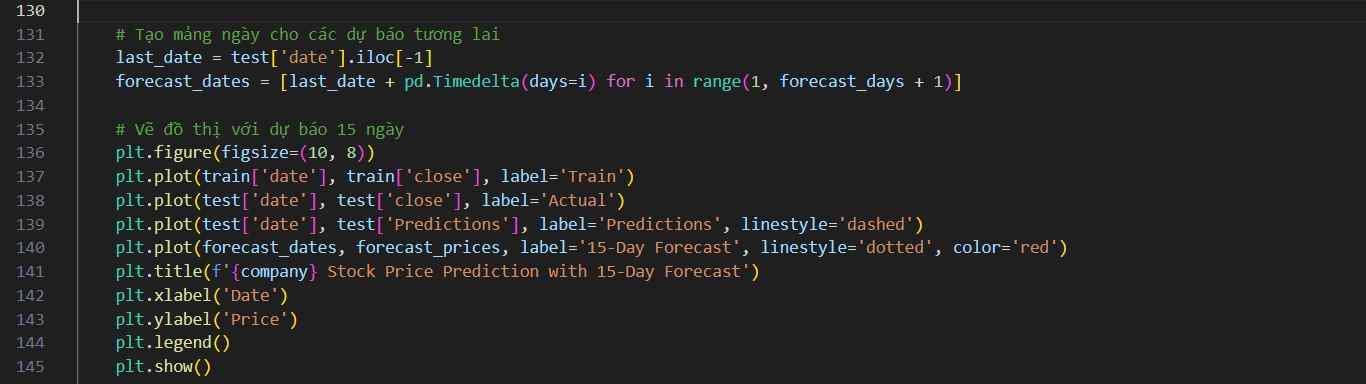
***Dự báo thêm 15 ngày***



Hình 35. Dự báo 15 ngày tiếp theo

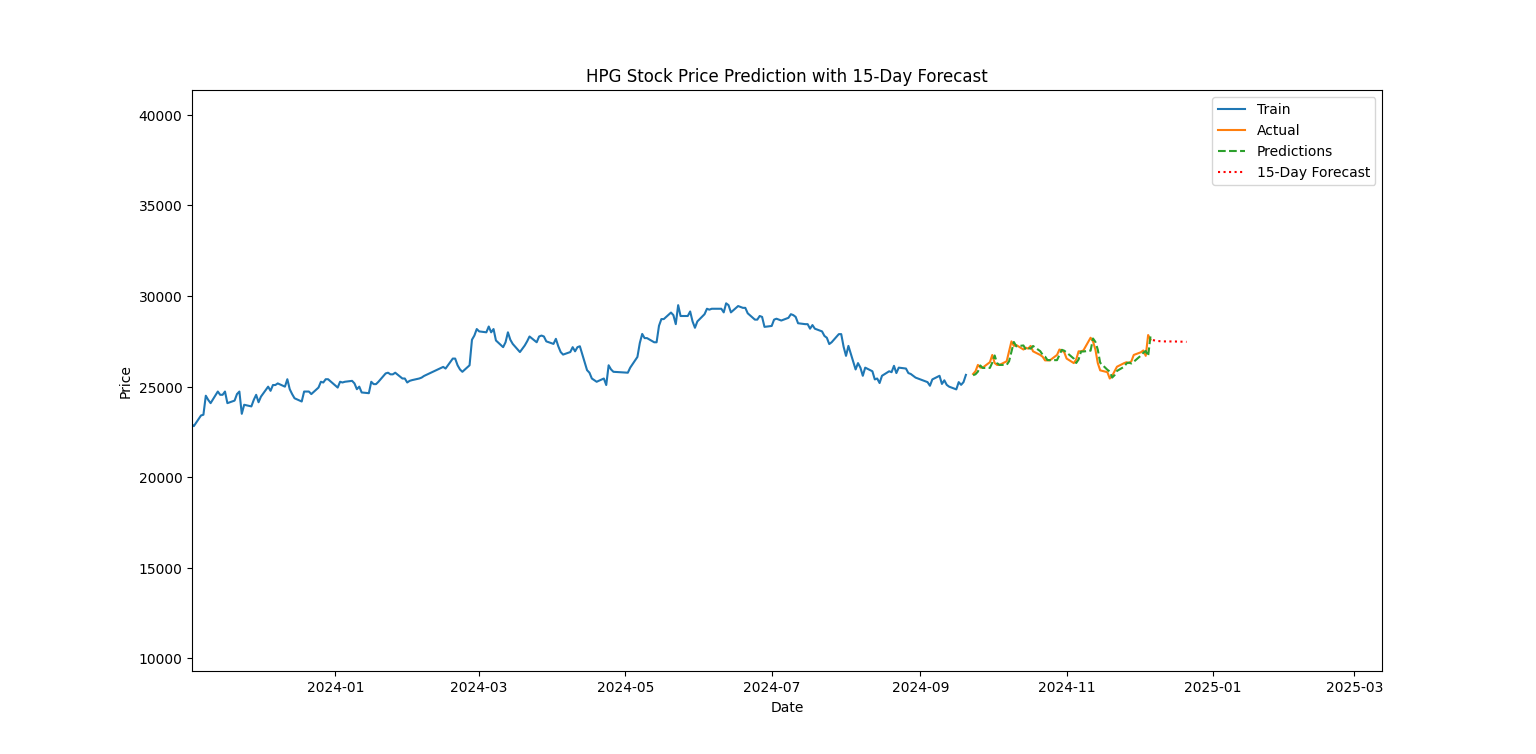
Dự báo tương lai: Sử dụng 60 ngày dữ liệu cuối cùng từ tập kiểm tra để dự đoán giá cổ phiếu trong 15 ngày tiếp theo.

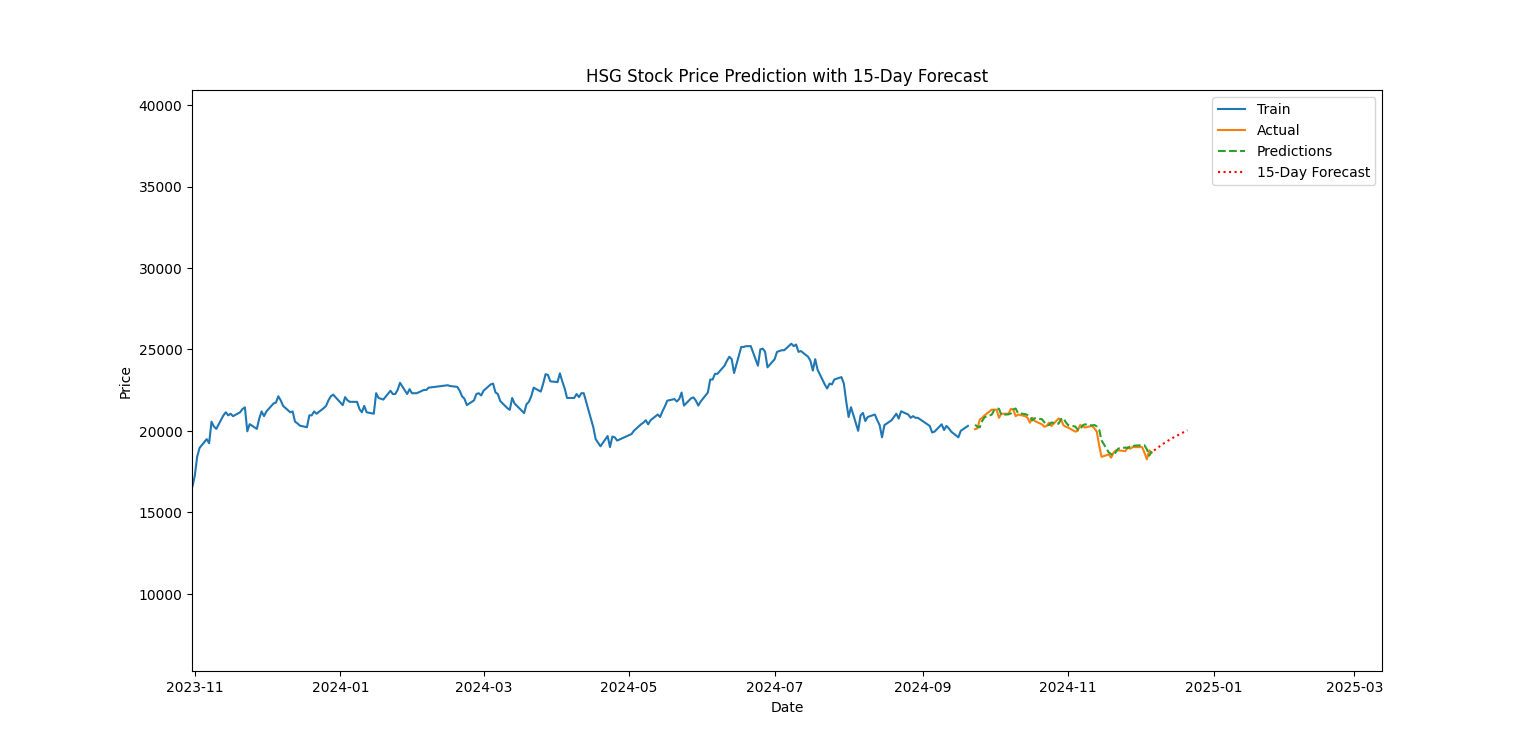
***Vẽ đồ thị dự báo 15 ngày***



Hình 36. Vẽ đồ thị dự báo 15 ngày tiếp theo

Dự báo 15 ngày: Vẽ đồ thị với dự báo giá cổ phiếu trong 15 ngày tiếp theo sau khi huấn luyện mô hình.



Hình 37. Kết quả dự báo giá cổ phiếu của HPG 15 ngày

Hình 38. Kết quả dự báo giá cổ phiếu của HSG 15 ngày

#### **2.4. Đánh giá mô hình**

Hình 39. Đánh giá mô hình bằng MSE và RMSE

* MSE (Mean Squared Error): Nó cho thấy sai số bình phương trung bình giữa các giá trị được dự đoán và thực tế. MSE là thước đo chất lượng của một công cụ ước tính - nó luôn không âm và các giá trị càng gần 0 càng tốt. Trong phân tích hồi quy, vẽ biểu đồ là một cách tự nhiên hơn để xem xu hướng chung của toàn bộ dữ liệu. Đơn giản MSE cho bạn biết mức độ gần của đường hồi quy với một tập hợp các điểm. Nó thực hiện điều này bằng cách lấy khoảng cách từ các điểm đến đường hồi quy (những khoảng cách này là “sai số”) và bình phương chúng. Bình phương là rất quan trọng để giảm độ phức tạp với các dấu hiệu tiêu cực. Nó cũng tạo ra nhiều trọng lượng hơn cho sự khác biệt lớn hơn.

***MSE = 112400***

* RMSE (Root Mean Squared Error): căn bậc hai mức trung bình của các sai số bình phương. RMSE là độ lệch chuẩn của các phần dư (sai số dự đoán). Phần dư là thước đo khoảng cách từ các điểm dữ liệu đường hồi quy; RMSE là thước đo mức độ dàn trải của những phần dư này, nói cách khác, nó cho bạn biết mức độ tập trung của dữ liệu xung quanh đường phù hợp nhất. Ảnh hưởng của mỗi lỗi đối với RMSE tỷ lệ với kích thước của lỗi bình phương; do đó các sai số lớn hơn có ảnh hưởng lớn đến RMSE một cách không cân xứng. Do đó, RMSE nhạy cảm với các yếu tố ngoại lai. Khi các quan sát và dự báo chuẩn hóa được sử dụng làm đầu vào RMSE, có mối quan hệ trực tiếp với hệ số tương quan . Ví dụ, nếu hệ số tương quan là 1, RMSE sẽ bằng 0, bởi vì tất cả các điểm nằm trên đường hồi quy (và do đó không có sai số).

***RMSE = 320***

# **C. KẾT LUẬN**

## **I. Tổng kết bài nghiên cứu**

Qua quá trình nghiên cứu và phát triển mô hình dự đoán giá cổ phiếu bằng cách sử dụng LSTM (Long Short-Term Memory), bài nghiên cứu đã tập trung vào việc khai thác dữ liệu lịch sử giá cổ phiếu để dự báo xu hướng giá trong ngắn hạn. Mặc dù mô hình LSTM là một công cụ mạnh mẽ trong dự đoán chuỗi thời gian, kết quả thu được từ mô hình chưa hoàn toàn đạt được kỳ vọng về độ chính xác, với các chỉ số **MSE = 112400** và **RMSE = 320.**

Tuy nhiên, qua việc phân tích biểu đồ trực quan, có thể thấy rằng mô hình LSTM vẫn phản ánh khá tốt sự tăng giảm xu hướng giữa giá thực tế và giá dự đoán, mặc dù vẫn tồn tại sai lệch đáng kể ở một số thời điểm. Điều này chỉ ra rằng mô hình có khả năng theo dõi xu hướng ngắn hạn của thị trường, nhưng vẫn còn thiếu độ chính xác trong việc dự đoán các biến động mạnh mẽ.

Mô hình LSTM có thể hoạt động hiệu quả khi dữ liệu đầu vào phản ánh đúng các yếu tố vĩ mô và các yếu tố tác động đến thị trường. Tuy nhiên, độ chính xác của mô hình bị ảnh hưởng bởi nhiều yếu tố ngoài dữ liệu quá khứ, như các thay đổi đột ngột trong thị trường tài chính, yếu tố chính trị, tình hình kinh tế vĩ mô, các thông tin bất ngờ và hành vi của các nhà đầu tư. Các sự kiện này khó có thể được mô hình LSTM dự đoán chỉ từ dữ liệu lịch sử.

Đặc biệt, trong giai đoạn nghiên cứu, những biến động mạnh mẽ của thị trường do tác động của đại dịch Covid-19 và các yếu tố chính trị, xã hội đã làm giảm khả năng dự đoán chính xác của mô hình. Các yếu tố này vượt ngoài phạm vi mà mô hình có thể học từ dữ liệu quá khứ, cho thấy sự cần thiết phải kết hợp mô hình LSTM với các phương pháp phân tích bổ sung để có thể nâng cao độ chính xác trong dự báo.

Từ kết quả nghiên cứu, có thể kết luận rằng mô hình LSTM có tiềm năng lớn trong việc dự đoán xu hướng giá cổ phiếu trong ngắn hạn, nhưng để đạt được độ chính xác cao hơn, cần cải thiện chất lượng dữ liệu đầu vào và kết hợp thêm các yếu tố ngoại cảnh và các phương pháp phân tích khác.

**II. Hướng phát triển trong tương lai**

**1. Hạn chế**

* **Dự báo liên tiếp và sai số tích lũy**: Mô hình dự báo liên tiếp có thể dẫn đến sai số tích lũy, đặc biệt là khi dự đoán xa trong tương lai. Dự đoán có thể mất độ chính xác do sự ảnh hưởng của các sai số nhỏ từ các bước trước. Điều này khiến cho các dự báo dài hạn có thể kém chính xác hơn so với các dự báo ngắn hạn.
* **Chỉ sử dụng giá đóng cửa**: Mô hình hiện tại chỉ sử dụng **giá đóng cửa** làm dữ liệu đầu vào, trong khi thực tế thị trường chứng khoán còn chịu ảnh hưởng của nhiều yếu tố khác như **khối lượng giao dịch, tin tức, các chỉ số kỹ thuật**, v.v. Việc bổ sung thêm các yếu tố này có thể giúp cải thiện đáng kể độ chính xác của mô hình.
* **Cửa sổ thời gian cố định (60 ngày)**: Việc sử dụng một cửa sổ cố định 60 ngày có thể không tối ưu đối với tất cả các loại cổ phiếu hoặc thị trường, vì mỗi cổ phiếu có thể có các xu hướng biến động giá khác nhau trong các giai đoạn khác nhau. Cửa sổ này có thể được điều chỉnh theo đặc điểm của từng mã cổ phiếu.

**2. Hướng phát triển**

* **Tinh chỉnh mô hình**: Các tham số như số lượng lớp LSTM, đơn vị LSTM, tỷ lệ dropout, và số lượng epoch có thể được điều chỉnh để cải thiện độ chính xác và hiệu quả của mô hình. Sử dụng các phương pháp và kỹ thuật như: Grid Search, Random Search, Bayesian Optimization, Gradien-Based Hyperparameter Optimization, Evolutionary Algorithms, v.v. Việc thử nghiệm kết hợp với các cấu trúc mạng khác (như GRU hoặc Transformer) cũng có thể là một lựa chọn để nâng cao hiệu quả.
* **Bổ sung dữ liệu**: Thêm các yếu tố khác vào mô hình như **chỉ số kỹ thuật (RSI, MACD, Moving Average)**, **khối lượng giao dịch**, hoặc **thông tin từ các tin tức tài chính** có thể cải thiện khả năng dự đoán của mô hình.
* **Dự đoán đa chiều**: Việc mở rộng mô hình để dự đoán các yếu tố khác như giá mở cửa, giá cao nhất/thấp nhất trong ngày cũng có thể giúp mô hình đưa ra những dự báo chi tiết hơn về diễn biến của giá cổ phiếu.

# **D. NGUỒN THAM KHẢO**

DataPot. (n.d.). *Tự học Power BI chuẩn chứng chỉ PL-300 của Microsoft*. YouTube. Retrieved December 9, 2024, from [https://www.youtube.com/watch?v=[VideoID](https://www.youtube.com/watch?v=%5BVideoID)]

Tuan, N. (n.d.). *Bài 14: Long Short-Term Memory (LSTM)*. Nttuan8.com. Truy cập ngày 9 tháng 12 năm 2024, từ <https://nttuan8.com/bai-14-long-short-term-memory-lstm/>

DataCamp. (n.d.). *LSTM in Python for Stock Market Prediction*. DataCamp. Retrieved December 9, 2024, from <https://www.datacamp.com/tutorial/lstm-python-stock-market>

Gülmez, B. (2023). Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. *Expert Systems With Applications, 227*, 120346. <https://doi.org/10.1016/j.eswa.2023.120346>

Nguyen, T. (n.d.). *Sử dụng mạng LSTM (Long Short-Term Memory) để dự đoán cổ phiếu*. Viblo. Truy cập ngày 9 tháng 12 năm 2024, từ <https://viblo.asia/s/su-dung-mang-lstm-long-short-term-memory-de-du-doan-co-phieu-24lJDz06KPM>

Hamoudi, H., & Elseifi, M. A. (2021). *Stock market prediction using CNN and LSTM*. Stanford University. Retrieved December 9, 2024, from <https://cs230.stanford.edu/projects_winter_2021/reports/70667451.pdf>

Sayah, F. (2021). *Stock market analysis & prediction using LSTM*. Kaggle. Retrieved December 9, 2024, from <https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm>

ProtonX. (2021, November 5). *Giải thích Mạng LSTM dễ hiểu và lập trình với Pytorch và Tensorflow*. YouTube. Retrieved December 9, 2024, from <https://www.youtube.com/watch?v=YcRPPy3EiJs&t=846s>