**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**KHOA CƠ KHÍ CHẾ TẠO MÁY**



**ĐỀ TÀI: ÁP DỤNG AI TRONG DỰ ĐOÁN GIÁ CỔ PHIẾU**

**GVHD: PGS.TS NGUYỄN TRƯỜNG THỊNH**

**MÃ MÔN HỌC: ARIN337629\_22\_2\_09**

**SINH VIÊN THỰC HIỆN: NGUYỄN ĐỨC QUYỀN**

**MSSV: 20146148**

*Hồ Chí Minh, tháng 5 năm 2023*

# **DANH MỤC TỪ VIẾT TẮT**

LSTM Long short-term memory

RNN Recurrent Neural Network

# **DANH MỤC HÌNH ẢNH**

[Hình 1. 1 Kiểm tra dữ liệu 3](#_Toc135220602)

[Hình 1. 2 Lấy dữ liệu "Ngày" và "Đóng cửa" 4](#_Toc135220603)

[Hình 1. 3 Mạng hồi quy RNN 4](#_Toc135220604)

[Hình 1. 4 Xây dựng mô hình RNN 5](#_Toc135220605)

[Hình 1. 5 Compile mô hình 6](#_Toc135220606)

[Hình 1. 6 Huấn luyện mô hình 7](#_Toc135220607)

[Hình 1. 7 Kết quả huấn luyện mô hình 7](#_Toc135220608)

[Hình 3. 1 Bảng kết quả dự đoán giá cổ phiếu trên tập kiểm tra 10](#_Toc135220656)

[Hình 3. 2 Biểu đồ kết quả dự đoán giá cổ phiếu trên tập kiểm tra 10](#_Toc135577050)

[Hình 3. 3 Đánh giá độ chính xác 11](#_Toc135577051)

[Hình 3. 4 Kết quả dự đoán ngày tiếp theo 11](#_Toc135577052)

**MỤC LỤC**

[DANH MỤC TỪ VIẾT TẮT i](#_Toc135577545)

[DANH MỤC HÌNH ẢNH ii](#_Toc135577546)

[GIỚI THIỆU 1](#_Toc135577547)

[1. Đặt vấn đề 1](#_Toc135577548)

[2. Mục tiêu nghiên cứu 1](#_Toc135577549)

[3. Đối tượng và phạm vi nghiên cứu 2](#_Toc135577550)

[CHƯƠNG 1. NỘI DUNG NGHIÊN CỨU 3](#_Toc135577551)

[1.1 Thu thập dữ liệu 3](#_Toc135577552)

[1.2 Tiền xử lý dữ liệu 3](#_Toc135577553)

[1.3 Sử dụng mô hình RNN cho bài toán dự đoán giá cổ phiếu 4](#_Toc135577554)

[1.4 Giải thích về kiến trúc mô hình RNN 5](#_Toc135577555)

[1.5 Trình bày về thuật toán huấn luyện mô hình 6](#_Toc135577556)

[CHƯƠNG 2: MÔ HÌNH VÀ THUẬT TOÁN 8](#_Toc135577557)

[2.1 Mô tả chi tiết về mô hình RNN 8](#_Toc135577558)

[2.2 Cấu trúc và chức năng của các lớp LSTM và lớp kết nối mạng 8](#_Toc135577559)

[CHƯƠNG 3: KẾT QUẢ NGHIÊN CỨU 10](#_Toc135577560)

[3.1 Kết quả của mô hình dự đoán giá cổ phiếu 10](#_Toc135577561)

[3.2 Thảo luận kết quả và giải thích 12](#_Toc135577562)

[KẾT LUẬN 13](#_Toc135577563)

[PHỤ LỤC 14](#_Toc135577564)

# **GIỚI THIỆU**

1. **Đặt vấn đề**

Bài toán dự đoán giá cổ phiếu luôn thu hút sự quan tâm của nhà đầu tư và các chuyên gia tài chính. Dự đoán chính xác về giá cổ phiếu mang lại lợi nhuận lớn và hỗ trợ người đầu tư đưa ra quyết định đúng đắn. Trong lĩnh vực này, sử dụng trí tuệ nhân tạo, đặc biệt là mô hình RNN, đã được áp dụng rộng rãi để dự đoán giá cổ phiếu. Dự đoán giá cổ phiếu hỗ trợ nhà đầu tư đưa ra quyết định đầu tư thông minh dựa trên thông tin dự đoán về giá cổ phiếu trong tương lai. Điều này giúp giảm rủi ro và tăng khả năng đạt được lợi nhuận. Ngoài ra, dự đoán giá cổ phiếu cung cấp thông tin quan trọng để tối ưu hóa quản lý danh mục đầu tư. Nhà đầu tư có thể cân nhắc việc phân bổ tài sản và điều chỉnh danh mục đầu tư dựa trên đánh giá tiềm năng tăng trưởng và rủi ro của các cổ phiếu khác nhau. Dự đoán giá cổ phiếu cũng hỗ trợ cho các chuyên gia tài chính và nhà phân tích thị trường trong việc phân tích thị trường và dự báo xu hướng. Qua việc phân tích dữ liệu lịch sử và các yếu tố tài chính khác, họ có thể hiểu được sự biến động của thị trường và đưa ra dự báo về hướng đi của giá cổ phiếu. Ngoài ra, dự đoán giá cổ phiếu cung cấp thông tin quan trọng cho nhà giao dịch, giúp họ đưa ra quyết định giao dịch và kinh doanh cổ phiếu một cách hiệu quả. Điều này giúp tận dụng các cơ hội giao dịch và giảm thiểu rủi ro. Cuối cùng, dự đoán giá cổ phiếu hỗ trợ trong xây dựng chiến lược đầu tư dài hạn. Nhờ có thông tin về xu hướng và tiềm năng tăng trưởng của cổ phiếu, nhà đầu tư có thể xác định được cổ phiếu tiềm năng và lựa chọn đầu tư dài hạn vào các công ty có triển vọng.

1. **Mục tiêu nghiên cứu**

Mục tiêu của bài toán là tạo ra một mô hình dự đoán giá cổ phiếu chính xác và tin cậy sử dụng mô hình RNN. Bằng cách sử dụng dữ liệu lịch sử về giá cổ phiếu và các chỉ số liên quan, mô hình RNN có khả năng học và nhớ thông tin từ quá khứ để đưa ra dự đoán về giá cổ phiếu trong tương lai.

Mục tiêu cụ thể của bài toán là xây dựng một mô hình RNN hiệu quả, đạt được độ chính xác cao và khả năng dự đoán tốt. Bằng cách tiền xử lý dữ liệu, xây dựng mô hình RNN với các lớp LSTM và kết hợp các lớp kết nối mạng, chúng ta hy vọng có thể đưa ra dự đoán chính xác về giá cổ phiếu trong tương lai gần.

1. **Đối tượng và phạm vi nghiên cứu**

Đối tượng nghiên cứu của đề tài là dữ liệu lịch sử về giá cổ phiếu và các chỉ số của Công ty Cổ phần Chứng khoán Tân Việt trong 5 năm gần đây.

Phạm vi nghiên cứu tập trung vào áp dụng trí tuệ nhân tạo, đặc biệt là mô hình RNN, trong việc dự đoán giá cổ phiếu. Nghiên cứu tập trung vào quá trình xây dựng mô hình RNN, tiền xử lý dữ liệu cổ phiếu, quy trình huấn luyện mô hình và đánh giá hiệu suất của mô hình dự đoán. Bao gồm việc xác định thời gian quan sát phù hợp trong 5 năm gần nhất, cũng như các chỉ số và yếu tố tài chính cần thiết để xây dựng một mô hình dự đoán giá cổ phiếu chính xác và đáng tin cậy sử dụng dữ liệu của công ty chứng khoán Tân Việt trong thời gian nghiên cứu.

Top of Form

# **CHƯƠNG 1. NỘI DUNG NGHIÊN CỨU**

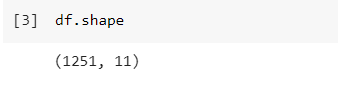
* 1. **Thu thập dữ liệu**

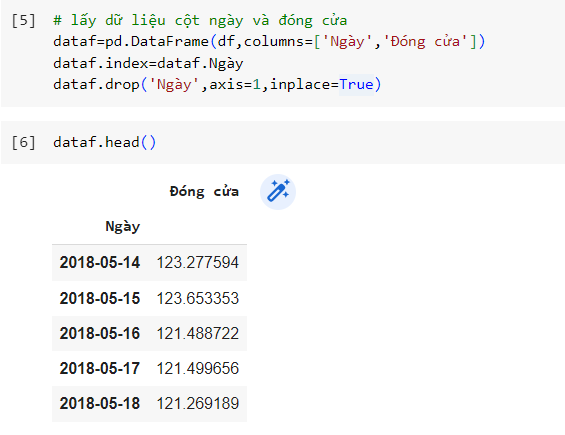
Trước khi đi sâu vào nghiên cứu, cần tiến hành thu thập dữ liệu lịch sử về giá cổ phiếu và các chỉ số liên quan trong 5 năm gần nhất của cổ phiếu của Công ty Cổ phần Chứng khoán Tân Việt. In thử 5 giá trình của bảng số liệu để kiểm tra lại dữ liệu.

Hình 1. 1 Kiểm tra dữ liệu

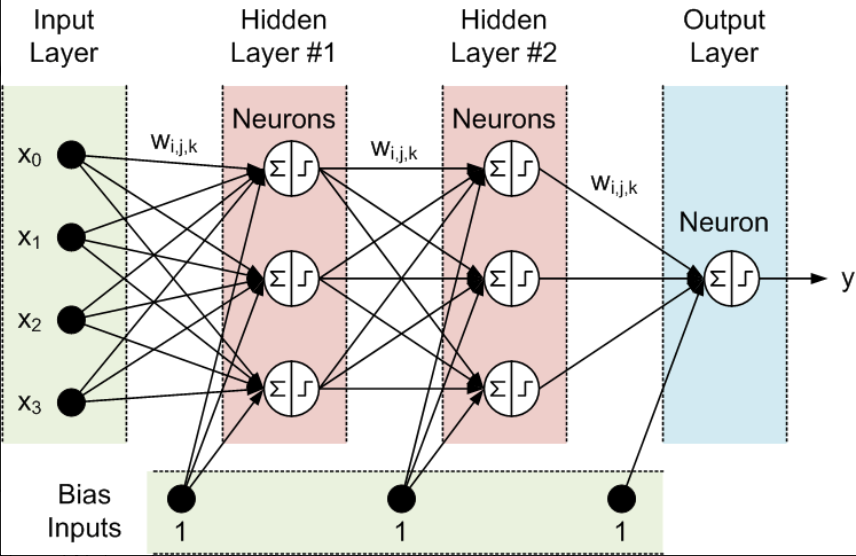
* 1. **Tiền xử lý dữ liệu**

Tập dữ liệu dùng trong nghiên cứu bao gồm số liệu của 1251 ngày trên sàn chứng khoán. Chúng ta sẽ lấy 2 thông tin là “Ngày” và “Đóng” cửa để dự đoán cho các ngày tiếp theo.



*Hình 1. 2 Lấy dữ liệu "Ngày" và "Đóng cửa"*

* 1. **Sử dụng mô hình RNN cho bài toán dự đoán giá cổ phiếu**

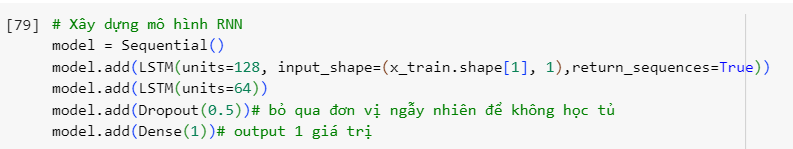
Mô hình RNN được chọn để dự đoán giá cổ phiếu vì khả năng xử lý dữ liệu chuỗi thời gian, khả năng học mô hình phức tạp và khả năng điều chỉnh độ dài cửa sổ thời gian. Qua đó, mô hình RNN có tiềm năng đưa ra các dự đoán chính xác và phản ánh các mô hình và xu hướng trong dữ liệu giá cổ phiếu.

Hình 1. 3 Mạng hồi quy RNN

**1.4 Giải thích về kiến trúc mô hình RNN**

Mô hình RNN là một kiến trúc mạng nơ-ron được thiết kế đặc biệt để xử lý dữ liệu chuỗi thời gian. Nó có khả năng lưu trữ thông tin từ quá khứ và sử dụng thông tin đó để dự đoán các giá trị trong tương lai. Kiến trúc RNN bao gồm các lớp LSTM và các lớp kết nối mạng.

* + 1. **Lớp LSTM**

LSTM là một loại lớp đặc biệt trong mô hình RNN, được thiết kế để giải quyết vấn đề mất mát thông tin trong quá trình lan truyền ngược gradient. Lớp LSTM có khả năng lưu trữ và quản lý thông tin từ các đầu vào trước đó thông qua các cơ chế cổng (gate) như cổng quên (forget gate), cổng đầu vào (input gate) và cổng đầu ra (output gate). Các cổng cho phép LSTM quyết định xem thông tin nào cần được lưu trữ và thông tin nào cần bị bỏ qua, giúp mô hình học và nhớ thông tin quan trọng từ quá khứ.

Hình 1. 4 Xây dựng mô hình RNN

* + 1. **Lớp kết nối mạng:**

Sau các lớp LSTM, thông thường có một hoặc nhiều lớp kết nối mạng (fully connected layers) để chuyển đổi và biểu diễn thông tin từ LSTM thành đầu ra dự đoán. Lớp kết nối mạng bao gồm các nơ-ron và ma trận trọng số, cho phép tính toán và kết hợp các đặc trưng đã học từ LSTM để tạo ra dự đoán cuối cùng về giá cổ phiếu. Kiến trúc mô hình RNN với các lớp LSTM và lớp kết nối mạng cho phép mô hình học và nhớ thông tin từ quá khứ thông qua cơ chế cổng của LSTM. Các đặc trưng đã học từ LSTM sau đó được kết hợp và biểu diễn trong các lớp kết nối mạng để tạo ra dự đoán về giá cổ phiếu trong tương lai. Kiến trúc này giúp mô hình RNN có khả năng xử lý dữ liệu chuỗi thời gian và dự đoán chính xác giá cổ phiếu.

## **Trình bày về thuật toán huấn luyện mô hình**

**1.5.1 Cấu hình tham số:**

Kích thước batch (batch size): Đây là số lượng các mẫu dữ liệu được sử dụng trong mỗi lần cập nhật gradient. Kích thước batch thường được chọn là một giá trị nhỏ để giảm thời gian tính toán và tăng tính ổn định của quá trình huấn luyện.

Số lượng epoch (epochs): Đây là số lần mô hình được huấn luyện trên toàn bộ tập huấn luyện. Một số lượng epoch thích hợp giúp mô hình học đủ thông tin từ dữ liệu.

* + 1. **Quá trình tối ưu hóa:**

Hàm mất mát (loss): Một hàm mất mát được chọn để đo lường sự khác biệt giữa dự đoán của mô hình và giá trị thực tế. Trong bài toán dự đoán giá cổ phiếu, có thể sử dụng hàm mất mát như sai số bình phương trung bình (Mean Squared Error) để đo lường chênh lệch giữa giá dự đoán và giá thực tế.

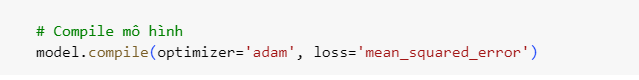
Thuật toán tối ưu (optimizer): Một thuật toán tối ưu được chọn để tối thiểu hóa hàm mất mát và điều chỉnh các trọng số của mô hình. Các thuật toán tối ưu phổ biến cho mô hình RNN là Stochastic Gradient Descent (SGD), Adam, hoặc RMSprop. Mỗi thuật toán có cách tối ưu và tốc độ hội tụ khác nhau.

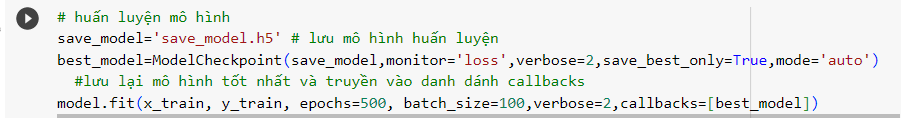
Tốc độ học (learning rate): Đây là một tham số quan trọng của thuật.

* + 1. **Lưu kết quả:**

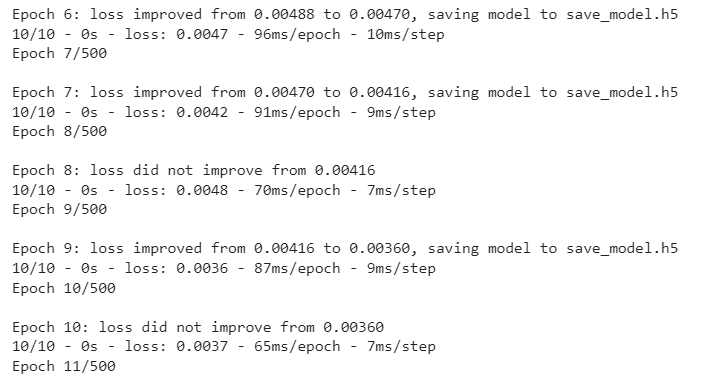
Tạo callback để lưu lại mô hình có hiệu suất tốt nhất(callback): Callback này sẽ giúp lưu mô hình vào tệp được chỉ định khi có sự cải thiện về giá trị hàm mất mát.

Lưu model (save\_model): Đặt đường dẫn và tên tệp để lưu mô hình huấn luyện. Câu lệnh này cho phép bạn chỉ định nơi và tên tệp để lưu mô hình sau khi hoàn tất quá trình huấn luyện.

Hình 1. 5 Compile mô hình



Hình 1. 6 Huấn luyện mô hình



Hình 1. 7 Kết quả huấn luyện mô hình

# **CHƯƠNG 2: MÔ HÌNH VÀ THUẬT TOÁN**

* 1. **Mô tả chi tiết về mô hình RNN**

Mô hình RNN được sử dụng trong bài toán dự đoán giá cổ phiếu có thể có một cấu trúc đơn giản như sau: Lớp đầu vào (Input Layer), Lớp LSTM (Long Short-Term Memory), Lớp kết nối mạng (Fully Connected Layer) và Lớp đầu ra (Output Layer).

* + 1. **Lớp đầu vào (Input Layer)**

Lớp này nhận các đặc trưng đầu vào liên quan đến giá cổ phiếu trong bài này là giá trị đóng cửa.

* + 1. **Lớp LSTM (Long Short-Term Memory)**

Mô hình RNN sử dụng các lớp LSTM để xử lý dữ liệu chuỗi thời gian và học các mô hình dự đoán. Mỗi lớp LSTM có một số lượng nơ-ron (hidden units) và các cổng (gates) như cổng quên (forget gate), cổng đầu vào (input gate) và cổng đầu ra (output gate) để điều chỉnh thông tin được lưu trữ và truyền qua. Các lớp LSTM nhận đầu vào từ lớp trước đó và truyền kết quả đến lớp tiếp theo trong chuỗi thời gian.

* + 1. **Lớp kết nối mạng (Fully Connected Layer)**

Sau các lớp LSTM, thông thường có một hoặc nhiều lớp kết nối mạng (fully connected layers) để kết hợp thông tin từ LSTM và tạo ra đầu ra dự đoán. Lớp kết nối mạng có thể sử dụng các hàm kích hoạt như ReLU (Rectified Linear Unit) hoặc sigmoid để tạo ra các giá trị đầu ra dự đoán.

* + 1. **Lớp đầu ra (Output Layer)**

Lớp cuối cùng trong mô hình RNN là lớp đầu ra, nơi dự đoán giá cổ phiếu được tạo ra. Đầu ra có thể là một giá trị dự đoán duy nhất hoặc một vector dự đoán cho các tham số khác nhau của giá cổ phiếu, chẳng hạn như giá mở cửa, giá cao nhất, giá thấp nhất, và giá đóng cửa.

* 1. **Cấu trúc và chức năng của các lớp LSTM và lớp kết nối mạng**
     1. **Lớp LSTM (Long Short-Term Memory)**

Cấu trúc: Mỗi lớp LSTM bao gồm một số lượng nơ-ron (hidden units) và các cổng (gates) như cổng quên (forget gate), cổng đầu vào (input gate) và cổng đầu ra (output gate).

Chức năng:

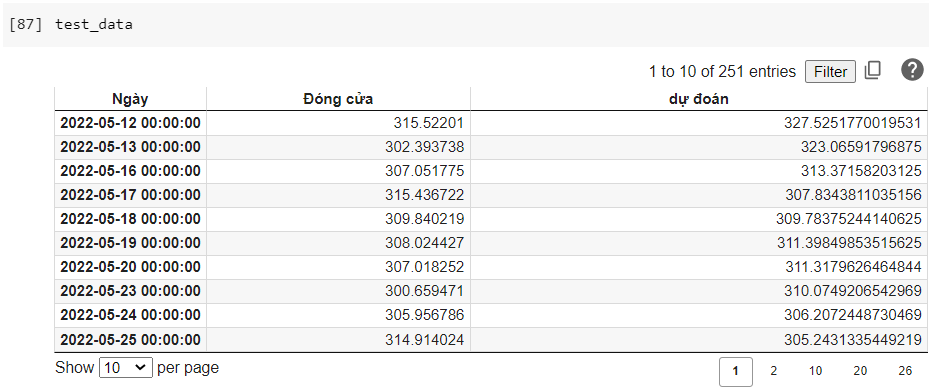
* Cổng quên (forget gate): Điều chỉnh thông tin trạng thái trước đó được lưu trữ trong LSTM. Nó xác định những thông tin nào cần bị loại bỏ hoặc quên đi.
* Cổng đầu vào (input gate): Xác định những thông tin mới nào sẽ được lưu trữ vào trạng thái hiện tại. Nó điều chỉnh thông tin mới từ đầu vào và thông tin trạng thái trước đó.
* Cổng đầu ra (output gate): Xác định những thông tin nào từ trạng thái hiện tại sẽ truyền đi để tạo ra đầu ra. Nó điều chỉnh đầu ra của LSTM dựa trên thông tin trạng thái hiện tại.
  + 1. **Lớp kết nối mạng (Fully Connected Layer)**

Cấu trúc: Lớp kết nối mạng là một lớp trong mô hình RNN, được đặt sau các lớp LSTM.

Chức năng: Lớp này thực hiện kết hợp thông tin từ các lớp LSTM và tạo ra đầu ra dự đoán. Mỗi nơ-ron trong lớp kết nối mạng nhận thông tin từ các nơ-ron trong lớp LSTM và áp dụng hàm kích hoạt (ví dụ: ReLU, sigmoid) để tính toán giá trị đầu ra. Các nơ-ron trong lớp kết nối mạng có thể được kết nối với tất cả các nơ-ron trong lớp LSTM hoặc chỉ một phần, tùy thuộc vào cấu trúc và mục tiêu của mô hình.

Tổng quát, lớp LSTM trong mô hình RNN có chức năng xử lý và lưu trữ thông tin chuỗi thời gian, trong khi lớp kết nối mạng thực hiện kết hợp thông tin từ LSTM và tạo ra đầu ra dự đoán. Các lớp này giúp mô hình RNN học và mô hình hóa quan hệ phức tạp trong dữ liệu chuỗi thời gian để dự đoán giá cổ phiếu.

# **CHƯƠNG 3: KẾT QUẢ NGHIÊN CỨU**

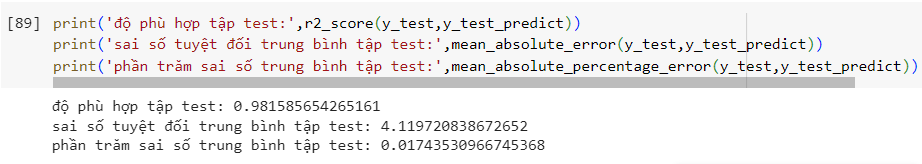
* 1. **Kết quả của mô hình dự đoán giá cổ phiếu**

Hình 3. 1 Bảng kết quả dự đoán giá cổ phiếu trên tập kiểm tra

Trong ví dụ trên, chúng ta có bảng kết quả hiển thị giá cổ phiếu thực tế và giá cổ phiếu dự đoán cho mỗi ngày trên tập kiểm tra. Ta có thể thấy sự khác biệt giữa giá thực tế và giá dự đoán.

Hình 3. 2Biểu đồ kết quả dự đoán giá cổ phiếu trên tập kiểm tra

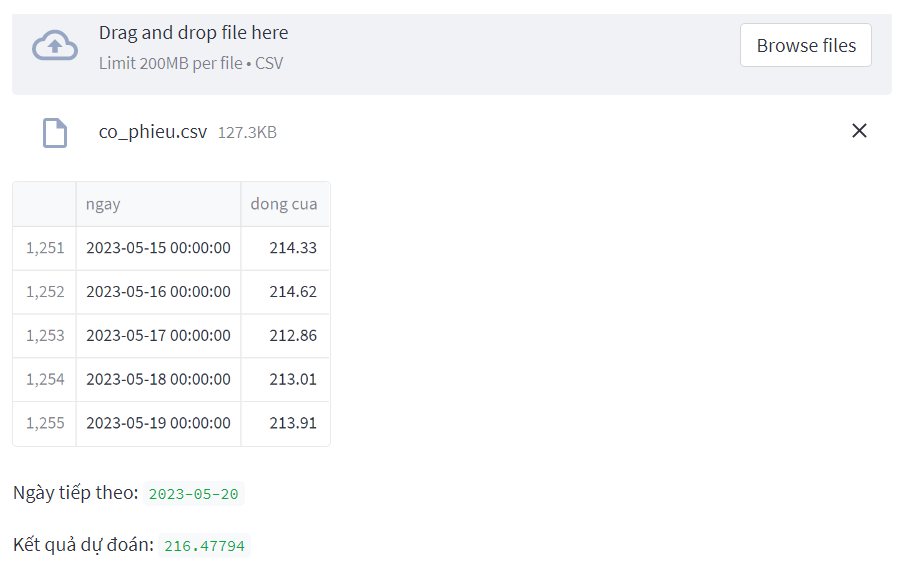
Trong biểu đồ trên, trục x đại diện cho thời gian (ngày, tuần, tháng) và trục y đại diện cho giá cổ phiếu. Đường màu red đại diện cho giá cổ phiếu thực tế, đường mài green đạo diện cho tập train và đường màu blue đại diện cho giá cổ phiếu dự đoán. Bằng cách so sánh hai đường red và blue, chúng ta có thể đánh giá độ chính xác của mô hình được in ra dưới đây.



Hình 3. 3 Đánh giá độ chính xác

So sánh dự đoán của mô hình với giá cổ phiếu thực tế và đánh giá độ chính xác và hiệu suất của mô hình. Mức độ phù hợp (accuracy) là một chỉ số quan trọng để đánh giá khả năng dự đoán của mô hình. Nó thể hiện tỷ lệ dự đoán đúng trong số tất cả các dự đoán. Với mức độ phù hợp trên 98%, mô hình đạt được độ chính xác cao trong việc dự đoán giá cổ phiếu. Điều này cho thấy mô hình có khả năng đưa ra dự đoán chính xác và tin cậy về giá cổ phiếu trong tương lai.

Xây dựng app dự đoán cổ phiếu của sàn chứng khoán việt tân để dự đoán giá cổ phiếu ngày tiếp theo dựa vào mô hình đã xây dựng và dữ liệu được cũng cấp bởi người dùng.



Hình 3. 4 Kết quả dự đoán ngày tiếp theo

## **Thảo luận kết quả và giải thích**

* 1. **.1 Thảo luận**

Sự khác biệt giữa dự đoán của mô hình và giá cổ phiếu thực tế có thể được giải thích bởi một số yếu tố. Thị trường tài chính và giá cổ phiếu thường trải qua sự biến động tự nhiên do nhiều yếu tố không thể dự đoán trước. Điều này có nghĩa là mô hình dự đoán cổ phiếu không thể hoàn toàn phản ánh được sự biến động này, dẫn đến sự khác biệt giữa dự đoán và giá thực tế. Mô hình RNN cũng có những giới hạn riêng, bao gồm khả năng xử lý dữ liệu lớn, khả năng mô hình hóa các mối quan hệ phức tạp và vấn đề overfitting. Những giới hạn này có thể ảnh hưởng đến khả năng dự đoán chính xác của mô hình và gây ra sự khác biệt với giá thực tế.

**3.2.2 Đánh giá**

Mô hình RNN có nhược điểm và hạn chế trong bài toán dự đoán giá cổ phiếu, bao gồm khả năng xử lý thông tin xa kém, vấn đề vanishing/exploding gradient, hạn chế đối với độ dài chuỗi dữ liệu, đòn bẩy thông tin lịch sử đồng nhất và cấu trúc đơn giản. Các biến thể như LSTM và GRU đã được phát triển để giảm nhược điểm này và nâng cao hiệu suất dự đoán giá cổ phiếu.

# **KẾT LUẬN**

Trong bài toán dự đoán giá cổ phiếu bằng mô hình RNN, chúng ta đã tiến hành nghiên cứu và áp dụng mô hình RNN để dự đoán giá cổ phiếu. Kết quả và những điểm quan trọng của bài toán này được tổng kết như sau:

Ý nghĩa và ứng dụng của dự đoán giá cổ phiếu: Dự đoán giá cổ phiếu có ý nghĩa quan trọng trong lĩnh vực tài chính và đầu tư. Nó giúp các nhà đầu tư và nhà quản lý tài sản có thể đưa ra quyết định dựa trên các thông tin dự báo và định hướng thị trường.

Phương pháp và quy trình sử dụng mô hình RNN: Chúng ta đã áp dụng mô hình RNN với kiến trúc LSTM và lớp kết nối mạng để dự đoán giá cổ phiếu. Quy trình bao gồm chuẩn bị dữ liệu, huấn luyện mô hình và dự đoán giá cổ phiếu trên tập kiểm tra hoặc dữ liệu mới.

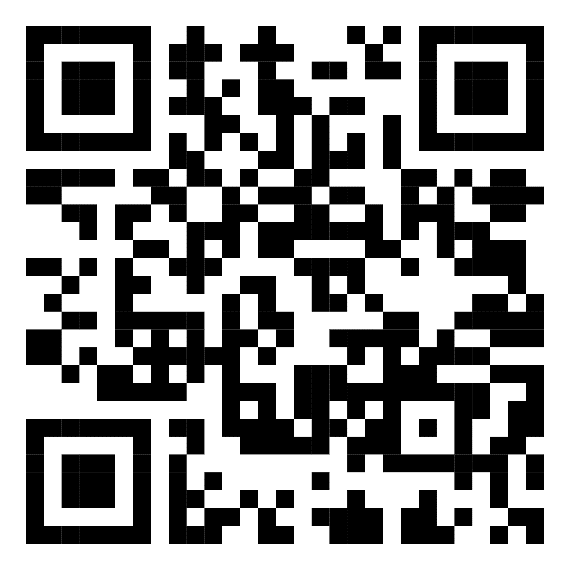
Kết quả dự đoán: Chúng ta đã hiển thị kết quả dự đoán của mô hình trên tập kiểm tra hoặc dữ liệu mới. So sánh giữa dự đoán và giá thực tế được thực hiện để đánh giá độ chính xác và hiệu suất của mô hình.

Thảo luận và giải thích: Chúng ta đã thảo luận về các kết quả và giải thích sự khác biệt giữa dự đoán và giá thực tế. Đánh giá độ chính xác, hiệu suất và những khía cạnh khác của mô hình đã được thảo luận để đưa ra nhận định tổng quan về hiệu quả của mô hình RNN trong bài toán dự đoán giá cổ phiếu.

Những điểm chính và hạn chế của mô hình RNN: Cuối cùng, chúng ta đã xem xét những điểm mạnh và yếu của mô hình RNN trong bài toán dự đoán giá cổ phiếu. Điều này giúp chúng ta có cái nhìn tổng quan về khả năng và giới hạn của mô hình trong việc dự đoán giá cổ phiếu.

Dựa trên các kết quả và thảo luận trên, chúng ta có thể kết luận rằng mô hình RNN có thể được sử dụng để dự đoán giá cổ phiếu với một mức độ chính xác và hiệu suất tương đối. Tuy nhiên, việc áp dụng mô hình cần được cân nhắc kỹ lưỡng và những giới hạn của mô hình cần được hiểu rõ để đưa ra quyết định và đánh giá chính xác trong lĩnh vực tài chính và đầu tư."

# **PHỤ LỤC**

1. **QR Github**
2. **Google drive**

[Link Dataset](https://bom.so/Nvroer)