

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC VIỆT NHẬT

-----o0o-----



BÁO CÁO KẾT QUẢ
KHOA HỌC DỮ LIỆU
NĂM 2023

Tên đề tài :

DỰ ĐOÁN GIÁ CHỨNG KHOÁN CỦA CÔNG TY APPLE
BẰNG LSTM

Hà Nội 12/2023

THÔNG TIN CHUNG VỀ ĐỀ TÀI

Giảng viên: Lê Chí Ngọc

Tên đề tài: Dự đoán giá chứng khoán của công ty apple bằng LSTM

Thời gian thực hiện: 06/12/2023 - 30/12/2023

Sinh viên: Bùi Thế Trung

Mã sinh viên: 21110108

Điện thoại: 0373.104.304

Email: 21110108@st.vju.ac.vn

Chương trình: Khoa học và kỹ thuật máy tính

MỤC LỤC

DANH SÁCH CÁC CHỮ VIẾT TẮT TRONG BÀI BÁO CÁO.....	1
DANH MỤC HÌNH.....	2
LỜI MỞ ĐẦU.....	3
I. ĐẶT VẤN ĐỀ.....	4
1.1. Thị trường chứng khoán.....	4
1.2. Tình hình thị trường chứng khoán.....	4
1.2.1. Tình hình kinh tế.....	4
1.2.2. Tin tức kinh doanh.....	4
1.2.3. Tâm lý đầu tư.....	4
1.2.4. Chiến lược đầu tư và xu hướng thị trường.....	5
1.3. Mục tiêu nghiên cứu.....	5
II. GIẢI QUYẾT BÀI TOÁN.....	6
2.1. Ý Tưởng Nghiên Cứu.....	6
2.2. Mô hình LSTM.....	6
2.2.1. Mạng hồi quy RNN.....	6
2.2.2. Mạng LSTM.....	8
2.2.2.1. Ý tưởng cốt lõi của LSTM.....	9
2.2.2.2. Bên trong LSTM.....	11
2.2.3. Các biến thể của bộ nhớ dài hạn.....	13
2.3. Lý do sử dụng LSTM.....	15
III. KẾ HOẠCH THỰC HIỆN.....	17
3.1. Thu Thập Dữ Liệu.....	17
3.2. Tiền Xử Lý Dữ Liệu.....	17
3.3. Xây Dựng Mô Hình.....	17
3.4. Dự Đoán.....	18
3.5. Thuật toán code.....	18
IV. ĐÁNH GIÁ HIỆU SUẤT.....	19
V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	24
TÀI LIỆU THAM KHẢO.....	27

DANH SÁCH CÁC CHỮ VIẾT TẮT TRONG BÀI BÁO CÁO

Chữ viết tắt	Ý nghĩa
MSE	Mean Squared Error (sai số bình phương trung bình)
LSTM	Long Short-Term Memory (Bộ nhớ dài-ngắn)
CNN	Convolutional Neural Network (Mạng nơ-ron tích chập)
RNN	Recurrent Neural Network (Mạng nơ-ron hồi quy)
SVM	Support Vector Machine (Máy hỗ trợ vector)
TF	TensorFlow (Khung học máy mã nguồn mở)
Keras	Thư viện học máy trên TF
MinMaxScaler	Phương pháp phân phối dữ liệu về khoảng [0,1]
USD/cổ phiếu	Đơn vị tiền tệ và đơn vị cổ phiếu

DANH MỤC HÌNH

Hình 2.2.1.a: Mạng nơ-ron tái phát có các vòng lặp.....	7
Hình 2.2.1.b: Một mạng lưới thần kinh tái phát chưa được triển khai.....	7
Hình 2.2.2.a: Mô-đun lặp lại trong RNN tiêu chuẩn chứa một lớp duy nhất...	8
Hình 2.2.2.b: Mô-đun lặp lại trong LSTM chứa bốn lớp tương tác.....	9
Hình 2.3. Cấu trúc của mạng LSTM.....	16
Hình 4.1. Bảng dữ liệu giá chứng khoán của Apple từ 1981-2023.....	20
Hình 4.2. Bảng dự đoán giá chứng khoán 1 ngày tiếp theo	20
Hình 4.3. Bảng dự đoán giá chứng khoán 30 ngày.....	22

LỜI MỞ ĐẦU

Trong bối cảnh ngày nay, thị trường chứng khoán trở thành một lĩnh vực quan trọng, đầy thách thức và cơ hội. Việc đầu tư vào cổ phiếu không chỉ là một phương tiện tăng cường tài sản mà còn là một nhiệm vụ đầy tính chiến lược, đòi hỏi sự hiểu biết sâu sắc về thị trường và khả năng dự đoán xu hướng giá.

Nghiên cứu về dự đoán giá chứng khoán sử dụng mô hình Long Short-Term Memory (LSTM) đặt ra mục tiêu chính là nâng cao khả năng dự báo trong môi trường thị trường biến động. Trong khuôn khổ đề tài này, em tập trung vào việc áp dụng LSTM để dự đoán giá cổ phiếu của một trong những công ty công nghệ hàng đầu thế giới - Apple. Sự kết hợp giữa dữ liệu lịch sử giá cổ phiếu và các yếu tố tài chính quan trọng nhằm xây dựng một mô hình dự đoán mạnh mẽ và đáng tin cậy.

Nghiên cứu này không chỉ mang tính ứng dụng cao trong việc hỗ trợ quyết định đầu tư mà còn đóng góp vào lĩnh vực phát triển các mô hình dự đoán thị trường chứng khoán. Sự hợp nhất giữa công nghệ và tài chính trong nghiên cứu này là một bước tiến quan trọng, chứng tỏ sức mạnh của các phương pháp máy học trong việc hiểu và dự đoán các biến động phức tạp trên thị trường tài chính.

Em xin gửi lời cảm ơn chân thành đến Thầy Lê Chí Ngọc - người hướng dẫn em trong quá trình thực hiện đề tài này. Sự tận tâm, kiến thức sâu rộng, và sự hỗ trợ nhiệt tình của Thầy đã giúp em vượt qua những thách thức, đồng hành trong quá trình nghiên cứu và phát triển mô hình. Bài học chuyên sâu từ Thầy không chỉ là nguồn động viên lớn mà còn là nền tảng quan trọng cho sự phát triển cá nhân và chuyên sâu hóa kiến thức của em.

Em xin chân thành cảm ơn!

I. ĐẶT VẤN ĐỀ

1.1. Thị trường chứng khoán

Thị trường chứng khoán là nơi giao dịch cổ phiếu của các công ty. Giá cổ phiếu được xác định bởi cung và cầu, và có thể bị ảnh hưởng bởi nhiều yếu tố, bao gồm tình hình kinh tế, tin tức kinh doanh, và các xu hướng công nghệ. Dự đoán giá cổ phiếu là một nhiệm vụ quan trọng đối với các nhà đầu tư, vì nó có thể giúp họ đưa ra quyết định đầu tư sáng suốt.

1.2. Tình hình thị trường chứng khoán

1.2.1. Tình hình kinh tế

Tình hình kinh tế đóng vai trò lớn trong việc hình thành và biến động giá cổ phiếu trên thị trường chứng khoán. Sự tăng trưởng hoặc suy thoái kinh tế có thể tác động đến niềm tin của nhà đầu tư, gây biến động trong cầu cung cổ phiếu và ảnh hưởng đến giá trị thị trường.

1.2.2. Tin tức kinh doanh

Thông tin liên quan đến doanh nghiệp, bao gồm báo cáo tài chính hàng quý, hàng năm, và các sự kiện nổi bật như thay đổi quản lý, chiến lược kinh doanh, hay phát hiện sản phẩm mới, có thể gây ra biến động đột ngột trong giá cổ phiếu. Sự kiện này thường là nguồn thông tin quan trọng cho quyết định đầu tư.

1.2.3. Tâm lý đầu tư

Tâm lý của nhà đầu tư đóng vai trò quan trọng trong quá trình hình thành giá cổ phiếu. Sự lo sợ, mong đợi, hay lạc quan của nhà đầu tư có thể tạo ra các biến động giá không dự đoán được. Tâm lý đầu tư cũng ảnh hưởng đến khả năng chấp nhận rủi ro và quyết định mua bán cổ phiếu.

1.2.4. Chiến lược đầu tư và xu hướng thị trường

Chiến lược đầu tư của các nhà đầu tư lớn, quỹ đầu tư, và các tổ chức tài chính cũng đóng vai trò trong việc định hình tình hình thị trường chứng khoán. Các xu hướng thị trường, như đà tăng hay giảm, có thể phản ánh chiến lược và quyết định của các nhà đầu tư lớn, tạo nên môi trường đầu tư đặc biệt.

1.3. Mục tiêu nghiên cứu

Mục tiêu chính của dự án là xây dựng một mô hình dự đoán giá chứng khoán của công ty Apple cho n ngày tiếp theo ($n > 0$) dựa trên dữ liệu giao dịch lịch sử. Điều này giúp nhà đầu tư và người quan tâm có cái nhìn chi tiết hơn về xu hướng dự kiến của giá chứng khoán và đưa ra quyết định đầu tư thông minh.

II. GIẢI QUYẾT BÀI TOÁN

2.1. Ý Tưởng Nghiên Cứu

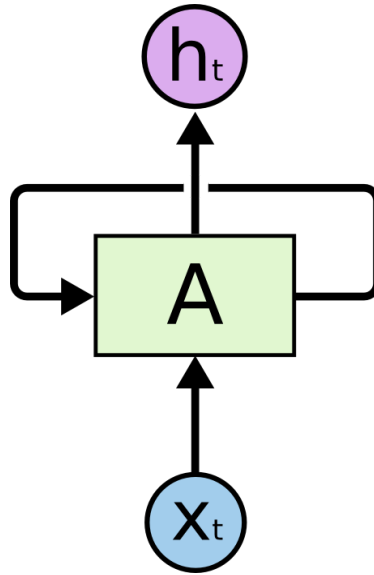
Chúng ta sẽ sử dụng mô hình Long Short-Term Memory (LSTM) để phân tích mối quan hệ giữa dữ liệu giá chứng khoán trong quá khứ và dự đoán giá trong tương lai. LSTM được chọn vì khả năng xử lý dữ liệu tuần tự và khả năng "nhớ" thông tin quan trọng từ quá khứ.

2.2. Mô hình LSTM

LSTM là một mạng cải tiến của RNN nhằm giải quyết vấn đề nhớ các bước dài của RNN.

2.2.1. Mạng hồi quy RNN

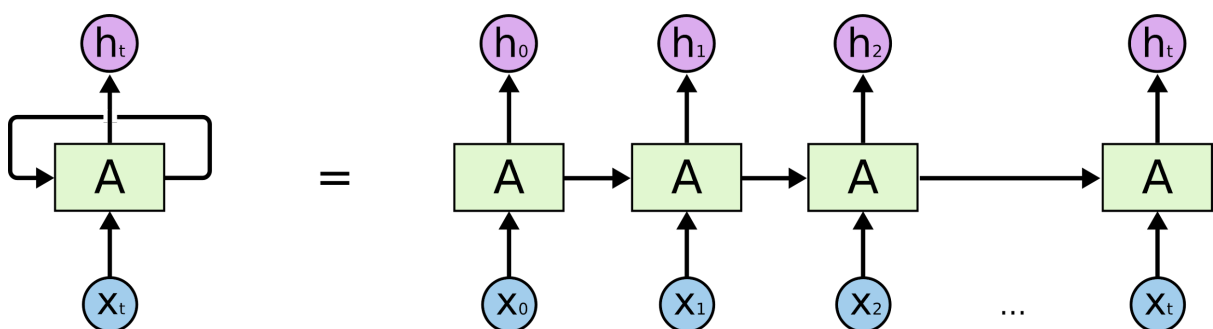
Ý tưởng chính của RNN (Recurrent Neural Network) là sử dụng chuỗi các thông tin. Trong các mạng nơ-ron truyền thống tất cả các đầu vào và cả đầu ra là độc lập với nhau. Tức là chúng không liên kết thành chuỗi với nhau. Nhưng các mô hình này không phù hợp trong rất nhiều bài toán. Ví dụ, nếu muốn đoán từ tiếp theo có thể xuất hiện trong một câu thì ta cũng cần biết các từ trước đó xuất hiện lần lượt thế nào chứ nhỉ? RNN được gọi là hồi quy (Recurrent) bởi lẽ chúng thực hiện cùng một tác vụ cho tất cả các phần tử của một chuỗi với đầu ra phụ thuộc vào cả các phép tính trước đó. Nói cách khác, RNN có khả năng nhớ các thông tin được tính toán trước đó. Trên lý thuyết, RNN có thể sử dụng được thông tin của một văn bản rất dài, tuy nhiên thực tế thì nó chỉ có thể nhớ được một vài bước trước đó mà thôi.



Hình 2.2.1.a: Mạng nơ-ron tái phát có các vòng lặp.

Hình vẽ trên mô tả một đoạn của mạng nơ-ron hồi quy A với đầu vào là x_t và đầu ra là h_t . Một vòng lặp cho phép thông tin có thể được truyền từ bước này qua bước này qua bước khác của mạng nơ-ron.

Các vòng lặp này khiến cho mạng nơ-ron hồi quy trông có vẻ khó hiểu. Tuy nhiên, nếu để ý một chút thì nó không khác mấy so với các mạng nơ-ron thuần. Một mạng nơ-ron hồi quy có thể được coi là nhiều bản sao chép của cùng một mạng, trong đó mỗi đầu ra của mạng này là đầu vào của một mạng sao chép khác. Hãy xem hình mô tả sau:



Hình 2.2.1.b: Một mạng lưới thần kinh tái phát chưa được triển khai.

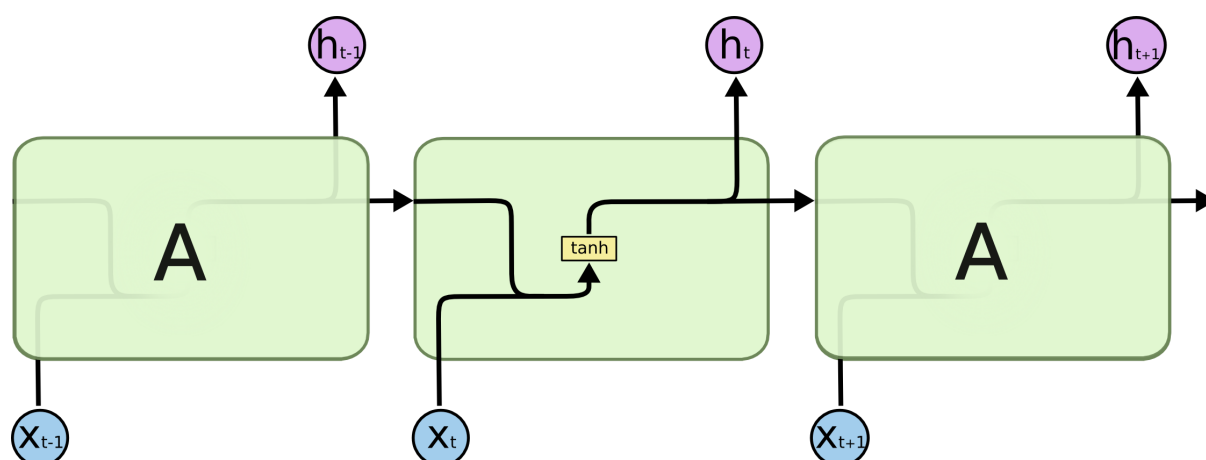
Chuỗi lặp lại các mạng này chính là phân giải của mạng nơ-ron hồi quy, các vòng lặp khiến chúng tạo thành một chuỗi danh sách các mạng sao chép nhau. Các nút của mạng vẫn nhận đầu vào và có đầu ra hết như mạng nơ-ron thuần.

2.2.2. Mạng LSTM

Mạng bộ nhớ dài-ngắn (Long Short Term Memory networks), thường được gọi là LSTM - là một dạng đặc biệt của RNN, nó có khả năng học được các phụ thuộc xa. LSTM được giới thiệu bởi Hochreiter & Schmidhuber (1997), và sau đó đã được cải tiến và phổ biến bởi rất nhiều người trong ngành. Chúng hoạt động cực kì hiệu quả trên nhiều bài toán khác nhau nên dần đã trở nên phổ biến như hiện nay.

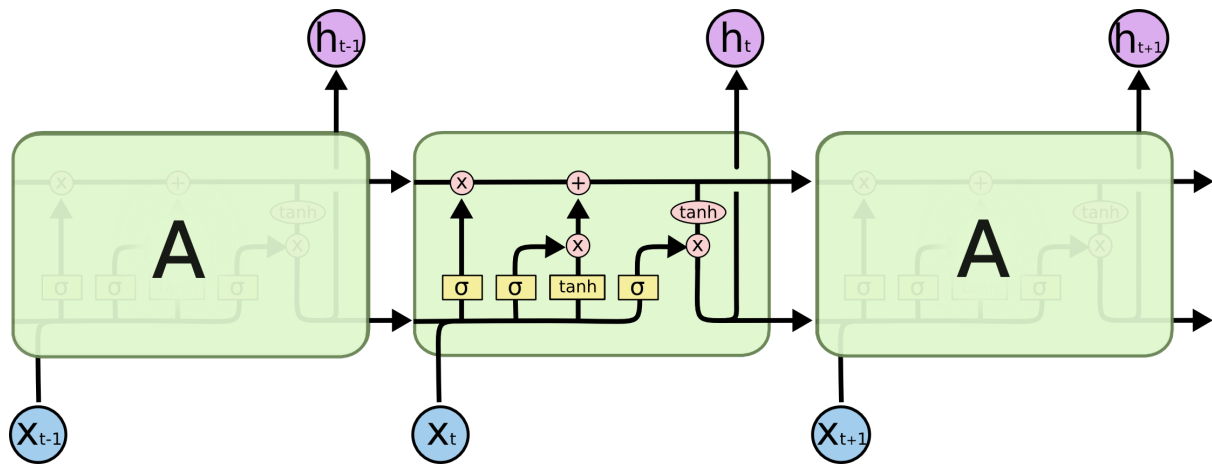
LSTM được thiết kế để tránh được vấn đề phụ thuộc xa (long-term dependency). Việc nhớ thông tin trong suốt thời gian dài là đặc tính mặc định của chúng, chứ không cần phải huấn luyện nó để có thể nhớ được. Tức là nó đã có thể ghi nhớ được mà không cần bất kì can thiệp nào.

Mọi mạng hồi quy đều có dạng là một chuỗi các mô-đun lặp đi lặp lại của mạng nơ-ron. Với mạng RNN chuẩn, các mô-đun này có cấu trúc rất đơn giản, thường là một tầng tanh.



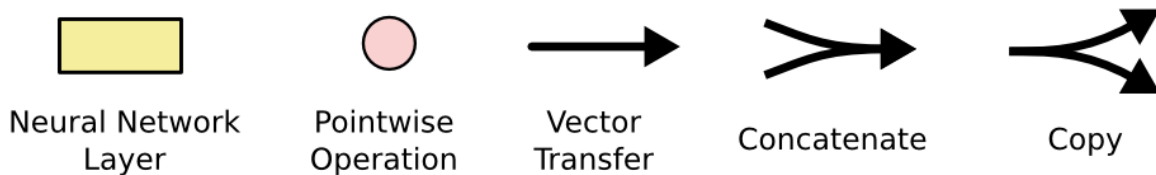
Hình 2.2.2.a: Mô-đun lặp lại trong RNN tiêu chuẩn chứa một lớp duy nhất.

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 chỉ có một tầng mạng nơ-ron, chúng có tới 4 tầng tương tác với nhau một cách rất đặc biệt.



Hình 2.2.2.b: Mô-đun lặp lại trong LSTM chứa bốn lớp tương tác.

Các kí hiệu mà sẽ sử dụng ở dưới đây:

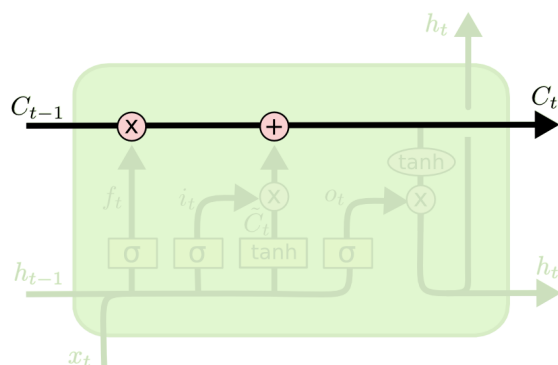


Ở sơ đồ trên, mỗi một đường mang một véc-tơ từ đầu ra của một nút tới đầu vào của một nút khác. Các hình trong màu hồng biểu diễn các phép toán như phép cộng véc-tơ chẳng hạn, còn các ô màu vàng được sử dụng để học trong các tầng mạng nơ-ron. Các đường hợp nhau kí hiệu việc kết hợp, còn các đường rẽ nhánh ám chỉ nội dung của nó được sao chép và chuyển tới các nơi khác nhau.

2.2.2.1. Ý tưởng cốt lõi của LSTM

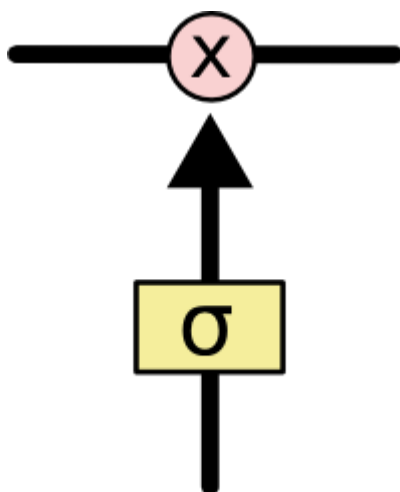
Chìa khóa của LSTM là trạng thái tế bào (cell state) - chính đường chạy thông ngang phía trên của sơ đồ hình vẽ.

Trạng thái tế bào là một dạng giống như băng truyền. Nó chạy xuyên suốt tất cả các mắt xích (các nút mạng) và chỉ tương tác tuyến tính đôi chút. Vì vậy mà các thông tin có thể dễ dàng truyền đi thông suốt mà không sợ bị thay đổi.



LSTM có khả năng bỏ đi hoặc thêm vào các thông tin cần thiết cho trạng thái tế bào, chúng được điều chỉnh cẩn thận bởi các nhóm được gọi là cổng (gate).

Các cổng là nơi sàng lọc thông tin đi qua nó, chúng được kết hợp bởi một tầng mạng sigmoid và một phép nhân.



Tầng sigmoid sẽ cho đầu ra là một số trong khoản $[0,1]$, mô tả có bao nhiêu thông tin có thể được thông qua. Khi đầu ra là 0 thì có nghĩa là không cho thông tin nào qua cả, còn khi là 1 thì có nghĩa là cho tất cả các thông tin đi qua nó.

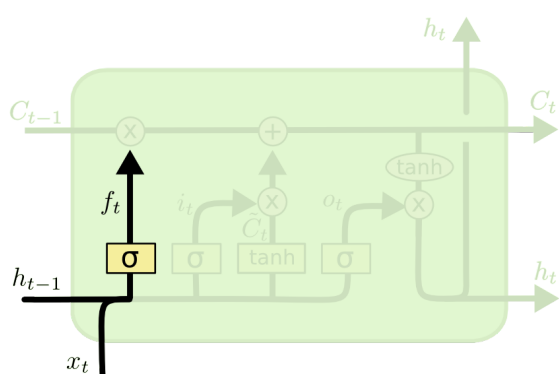
Một LSTM gồm có 3 cổng như vậy để duy trì và điều hành trạng thái của tế bào.

2.2.2.2. Bên trong LSTM

Bước đầu tiên của LSTM là quyết định xem thông tin nào cần bỏ đi từ trạng thái tế bào. Quyết định này được đưa ra bởi tầng sigmoid - gọi là “tầng cổng quên” (forget gate layer). Nó sẽ lấy đầu vào là h_{t-1} và x_t rồi đưa ra kết quả là một số trong khoảng $[0,1]$ cho mỗi số trong trạng thái tế bào C_{t-1}

Đầu ra là 1 thể hiện rằng nó giữ toàn bộ thông tin lại, còn 0 chỉ rằng toàn bộ thông tin sẽ bị bỏ đi.

Quay trở lại với ví dụ mô hình ngôn ngữ dự đoán từ tiếp theo dựa trên tất cả các từ trước đó, với những bài toán như vậy, thì trạng thái tế bào có thể sẽ mang thông tin về giới tính của một nhân vật nào đó giúp ta sử dụng được đại từ nhân xưng chuẩn xác. Tuy nhiên, khi đề cập tới một người khác thì ta sẽ không muốn nhớ tới giới tính của nhân vật nữa, vì nó không còn tác dụng gì với chủ thể mới này.

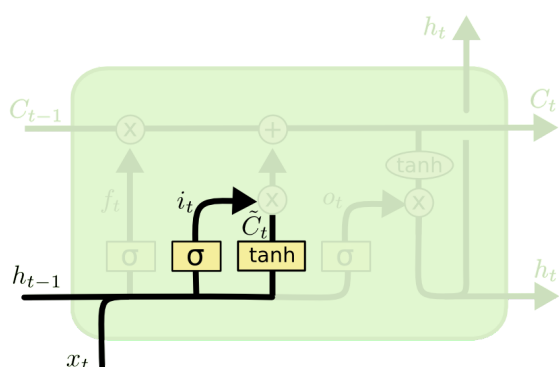


$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Bước tiếp theo là quyết định xem thông tin mới nào ta sẽ lưu vào trạng thái tế bào. Việc này gồm 2 phần. Đầu tiên là sử dụng một tầng sigmoid được gọi là “tầng cổng vào” (input gate layer) để quyết định giá trị nào ta sẽ cập nhật. Tiếp theo là một tầng tanh tạo ra một véc-tơ cho giá trị mới C_t nhằm thêm vào cho

trạng thái. Trong bước tiếp theo, ta sẽ kết hợp 2 giá trị đó lại để tạo ra một cặp nhập cho trạng thái.

Chẳng hạn với ví dụ mô hình ngôn ngữ của ta, ta sẽ muốn thêm giới tính của nhân vật mới này vào trạng thái tế bào và thay thế giới tính của nhân vật trước đó.



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

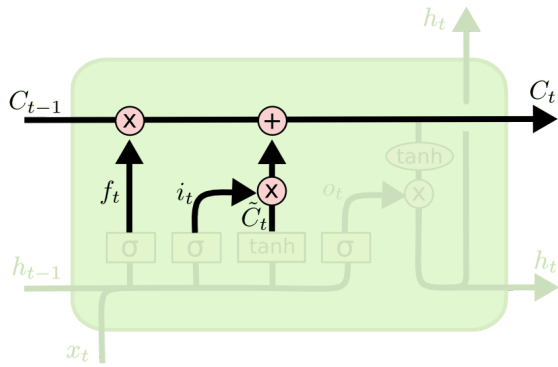
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Giờ là lúc cập nhật trạng thái tế bào cũ C_{t-1} thành trạng thái mới C_t . Ở các bước trước đó đã quyết định những việc cần làm, nên giờ ta chỉ cần thực hiện là xong.

Ta sẽ nhân trạng thái cũ với f_t để bỏ đi những thông tin ta quyết định quên lúc trước. Sau đó cộng thêm $i_t * (C_t) \sim$

Trạng thái mới thu được này phụ thuộc vào việc ta quyết định cập nhật mỗi giá trị trạng thái ra sao.

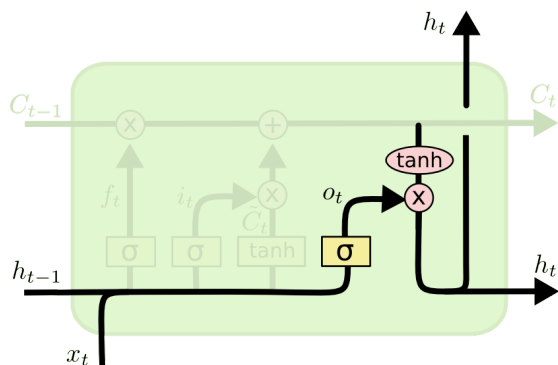
Với bài toán mô hình ngôn ngữ, chính là việc ta bỏ đi thông tin về giới tính của nhân vật cũ, và thêm thông tin về giới tính của nhân vật mới như ta đã quyết định ở các bước trước đó.



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Cuối cùng, ta cần quyết định xem ta muốn đầu ra là gì. Giá trị đầu ra sẽ dựa vào trạng thái tế bào, nhưng sẽ được tiếp tục sàng lọc. Đầu tiên, ta chạy một tầng sigmoid để quyết định phần nào của trạng thái tế bào ta muốn xuất ra. Sau đó, ta đưa nó trạng thái tế bào qua một hàm tanh để có giá trị nó về khoảng $[-1,1]$, và nhân nó với đầu ra của cổng sigmoid để được giá trị đầu ra ta mong muốn.

Với ví dụ về mô hình ngôn ngữ, chỉ cần xem chủ thể mà ta có thể đưa ra thông tin về một trạng từ đi sau đó. Ví dụ, nếu đầu ra của chủ thể là số ít hoặc số nhiều thì có thể biết được dạng của trạng từ đi theo sau nó phải như thế nào.



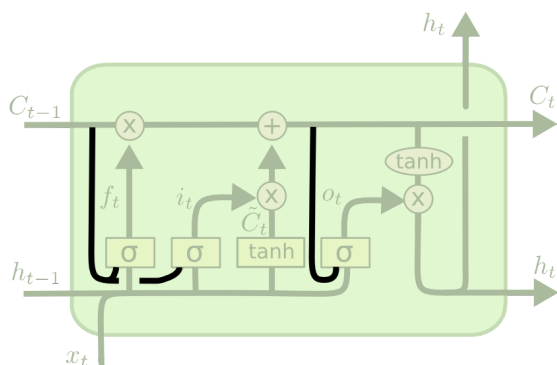
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

2.2.3. Các biến thể của bộ nhớ dài hạn

Những thứ ta vừa mô tả ở trên là một LSTM khá bình thường. Nhưng không phải tất cả các LTSM đều giống như vậy. Thực tế, các bài báo về LTSM đều sử dụng một phiên bản hơi khác so với mô hình LTSM chuẩn. Sự khác nhau không lớn, nhưng chúng giúp giải quyết phần nào đó trong cấu trúc của LTSM.

Một dạng LTSM phổ biến được giới thiệu bởi Gers & Schmidhuber (2000) được thêm các đường kết nối “peephole connections”, làm cho các tầng cổng nhận được giá trị đầu vào là trạng thái tế bào.



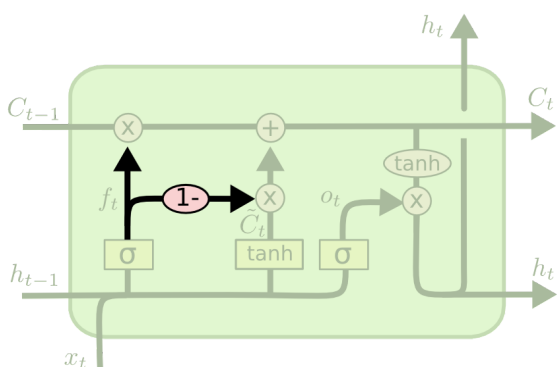
$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i)$$

$$o_t = \sigma(W_o \cdot [C_t, h_{t-1}, x_t] + b_o)$$

Hình trên mô tả các đường được thêm vào mọi cổng, nhưng cũng có những bài báo chỉ thêm cho một vài cổng mà thôi.

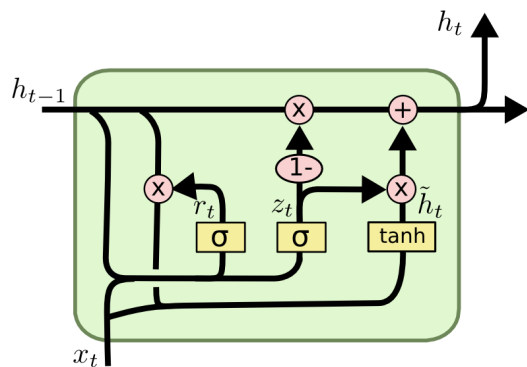
Một biến thể khác là nối 2 cổng loại trừ và đầu vào với nhau. Thay vì phân tách các quyết định thông tin loại trừ và thông tin mới thêm vào, ta sẽ quyết định chúng cùng với nhau luôn. Ta chỉ bỏ đi thông tin khi mà ta thay thế nó bằng thông tin mới đưa vào. Ta chỉ đưa thông tin mới vào khi ta bỏ thông tin cũ nào đó đi.



$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t$$

Một biến thể khá thú vị khác của LSTM là Gated Recurrent Unit, hay GRU được giới thiệu bởi Cho, et al. (2014). Nó kết hợp các cổng loại trừ và đầu vào thành một cổng “cổng cập nhập” (update gate). Nó cũng hợp trạng thái tế bào và

trạng thái ẩn với nhau tạo ra một thay đổi khác. Kết quả là mô hình của ta sẽ đơn giản hơn mô hình LSTM chuẩn và ngày càng trở nên phổ biến.



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

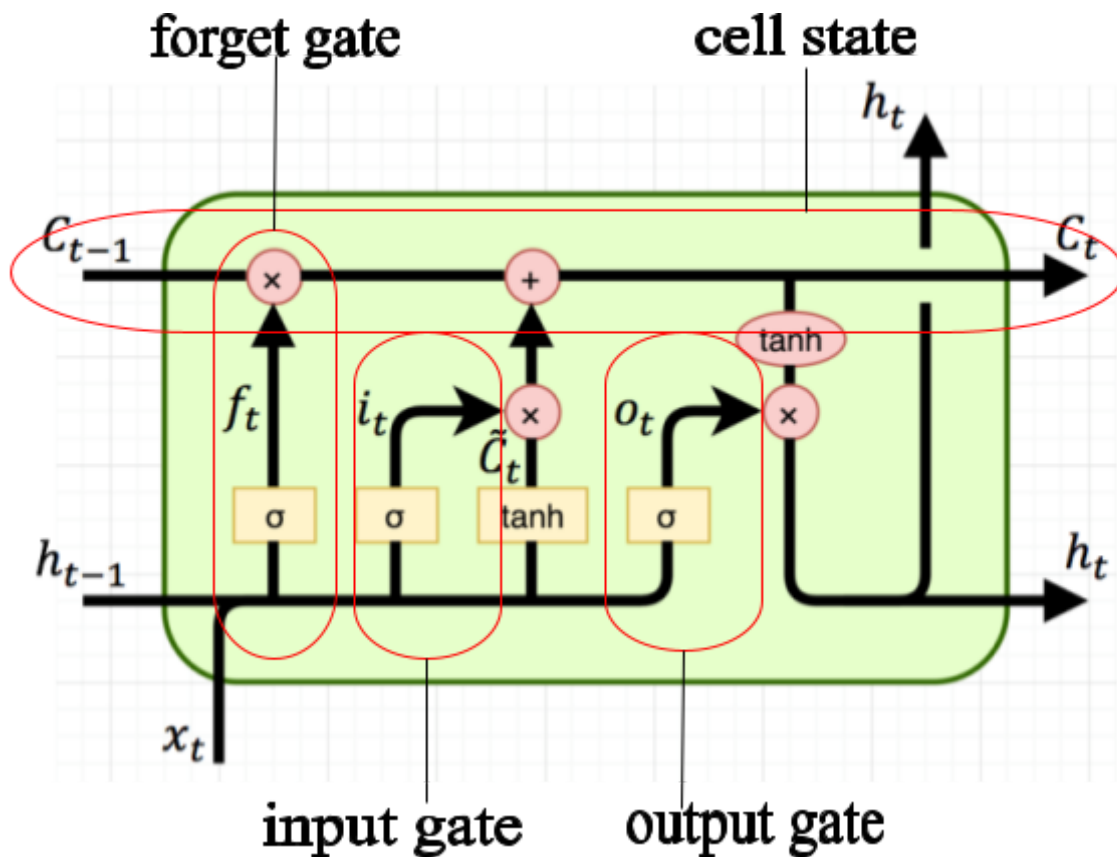
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

2.3. Lý do sử dụng LSTM

- Mạng LSTM hay RNN nói chung được thiết kế ra để xử lý dữ liệu tuần tự: Dữ liệu mà mỗi mẫu trong đó xuất hiện theo thứ tự (ví dụ như thứ tự giá chứng khoán xuất hiện theo thời gian, thứ tự các từ trong câu, thứ tự các nốt nhạc trong 1 bản nhạc)
- Mỗi nơ-ron của mạng LSTM được thiết kế gồm 3 cổng và 1 thông tin gọi là cell state



Hình 2.3. Cấu trúc của mạng LSTM

- **cell state:** Đây là nơi mà một cell dùng để ghi nhớ thông tin (các thông tin này một phần nào đó liên quan đến các thông tin đã xử lý trước đó, phần khác liên quan đến dữ liệu input tại mỗi cell)
- **forget gate:** Cổng này quyết định việc một thông tin nên được giữ lại hay bị quên đi (đầu vào của cổng này bao gồm dữ liệu input và thông tin từ cell trước đó của mạng)
- **input gate:** Tại đây chắt lọc các thông tin quan trọng của dữ liệu đầu vào và thông tin từ cell trước đó. Kế đến, nó cập nhật các thông tin này vào cell state
- **output gate:** Cổng này quyết định đầu vào của cell tiếp theo bằng cách tiếp tục chắt lọc thông tin lấy từ input, cell state và thông tin nhận được từ cell trước đó

III. KẾ HOẠCH THỰC HIỆN

3.1. Thu Thập Dữ Liệu

- Sử dụng thư viện DataReader từ Pandas.
- Cài đặt: **pip install pandas-datareader**.
- File dữ liệu: “Apple_Stock_Prices.csv” được lấy từ trang kaggle.com

3.2. Tiền Xử Lý Dữ Liệu

- Chỉ giữ lại cột Giá Đóng Cửa Điều Chỉnh (**Adj Close Price**) để dự đoán.
- Chia tỷ lệ dữ liệu bằng cách sử dụng **sklearn.preprocessing.MinMaxScaler**.
- Chuyển đổi dữ liệu thành định dạng mà LSTM có thể đọc (dữ liệu ở định dạng 3D).
- **sequence**: Tổng số dữ liệu đầu vào.
- **time_step**: Số lượng dữ liệu đầu vào cần thiết cho một dự đoán.
- **feature**: Số lượng mẫu dữ liệu cho một đầu ra (được tính từ dữ liệu đầu vào **time_step** ở trên).

3.3. Xây Dựng Mô Hình

- Sử dụng LSTM từ thư viện keras (cài đặt: **pip install Keras**).
- Mô hình bao gồm một lớp LSTM.
- Để dự đoán một ngày:
 - **time_step** = 60: Sử dụng dữ liệu 60 ngày trước đó làm dữ liệu đầu vào cho một dự đoán.
 - **feature** = 1: Dữ liệu đầu ra là giá của ngày tiếp theo.
- Để dự đoán 30 ngày:
 - **time_step** = 1500: Sử dụng dữ liệu 1500 ngày trước đó làm dữ liệu đầu vào cho một dự đoán.
 - **feature** = 30: Dữ liệu đầu ra là giá cổ phiếu cho 30 ngày tiếp theo.

3.4. Dự Đoán

- Đề dự đoán một ngày: tập kiểm tra chứa 2122 giá trị.
- Đề dự đoán 30 ngày: tập kiểm tra chứa 30 giá trị (nếu nhiều hơn, cần xử lý dự đoán chồng chéo).

3.5. Thuật toán code

Tham khảo tại link github: [Trungnef/Data-Science: Final Examination \(github.com\)](https://github.com/Trungnef/Data-Science-Final-Examination)

IV. ĐÁNH GIÁ HIỆU SUẤT

Sai số bình phương trung bình (Mean Squared Error - MSE) là một phép đo lường chênh lệch bình phương trung bình giữa giá trị ước tính và giá trị thực tế. MSE được tính bằng cách lấy tổng bình phương sai số chia cho số lượng mẫu. Công thức tính MSE như sau:

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

Trong đó:

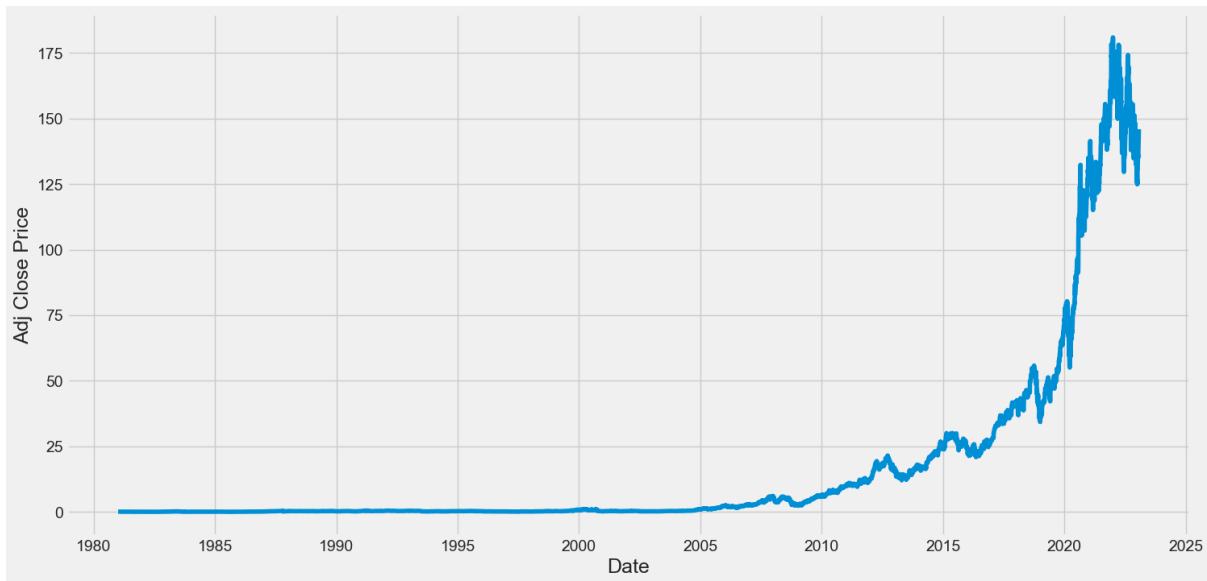
- n là số lượng mẫu
- y_i là các giá trị quan sát được
- \hat{y}_i là giá trị dự đoán

Giá cổ phiếu của Apple đã có một sự tăng trưởng đáng kinh ngạc trong 42 năm qua. Từ mức giá 0,06 USD/cổ phiếu vào ngày 2/1/1981, giá cổ phiếu của Apple đã tăng lên mức 157,30 USD/cổ phiếu vào ngày 27/01/2023. Điều này tương đương với mức tăng trưởng hơn 2.500.000%.

Các giai đoạn tăng trưởng

Giá cổ phiếu của Apple có thể được chia thành ba giai đoạn tăng trưởng chính:

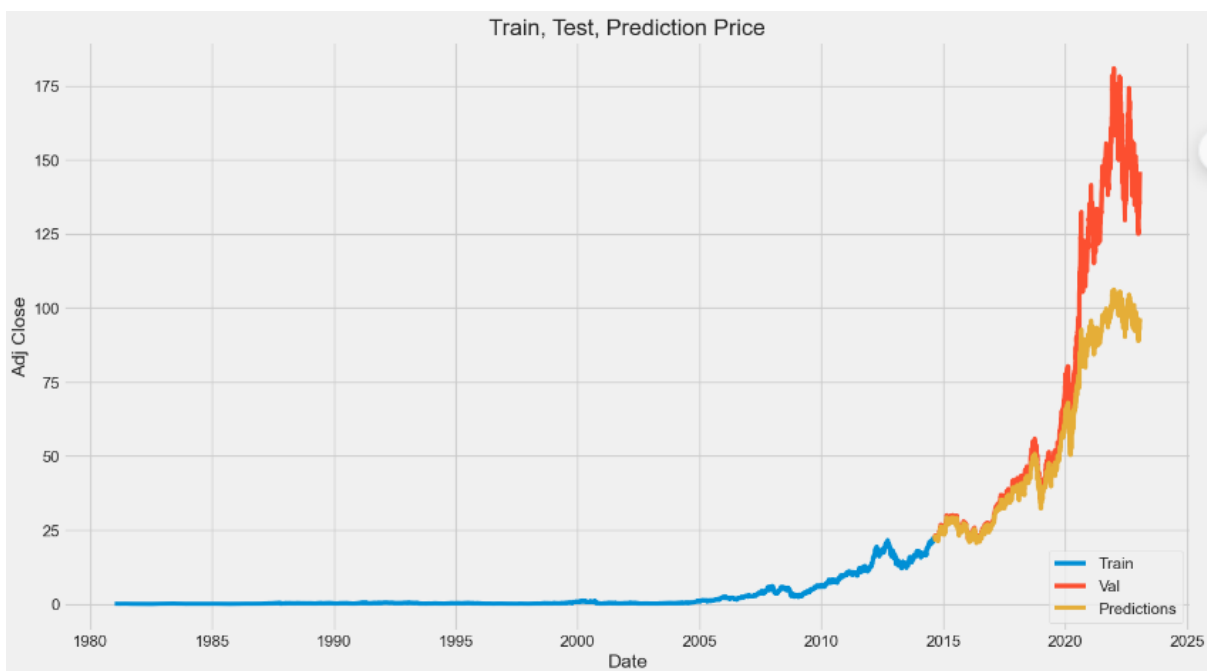
- Giai đoạn đầu (1981-1999): Giá cổ phiếu của Apple tăng trưởng chậm rãi trong giai đoạn này, với mức tăng trung bình hàng năm là 20%.
- Giai đoạn tăng trưởng nhanh (2000-2012): Giá cổ phiếu của Apple tăng trưởng nhanh chóng trong giai đoạn này, với mức tăng trung bình hàng năm là 50%.
- Giai đoạn tăng trưởng ổn định (2013-nay): Giá cổ phiếu của Apple tăng trưởng ổn định trong giai đoạn này, với mức tăng trung bình hàng năm là 20%.



Hình 4.1. Bảng dữ liệu giá chứng khoán của Apple từ 1981-2023

Dự đoán giá chứng khoán 1 ngày tiếp theo (tập dự đoán bao gồm 2122 giá trị):

MSE = 26.652150430735873



Hình 4.2. Bảng dự đoán giá chứng khoán 1 ngày tiếp theo

Nhận xét đánh giá:

Dựa trên dữ liệu giá cổ phiếu của Apple từ 2/1/1981 đến 27/01/2023 Mô hình cho thấy có độ sai số trung bình bình phương (MSE) là 26,65 USD/cổ phiếu.

Nhìn chung, đây là một kết quả khá tốt. MSE là một thước đo độ sai lệch giữa các giá trị dự đoán và giá trị thực tế. Giá trị MSE càng nhỏ thì độ chính xác của mô hình càng cao. Trong trường hợp này, MSE là 26,65 USD/cổ phiếu, tương đương với sai số trung bình khoảng 1,7%. Điều này có nghĩa là mô hình dự đoán giá cổ phiếu của Apple khá chính xác.

Tuy nhiên, vẫn có một số điểm cần cải thiện. Đầu tiên, độ sai lệch của mô hình có vẻ hơi cao ở một số thời điểm. Ví dụ, vào ngày 1/1/2023, giá cổ phiếu thực tế của Apple là 157,30 USD/cổ phiếu, nhưng mô hình dự đoán là 155,55 USD/cổ phiếu. Độ sai lệch này là 1,2%.

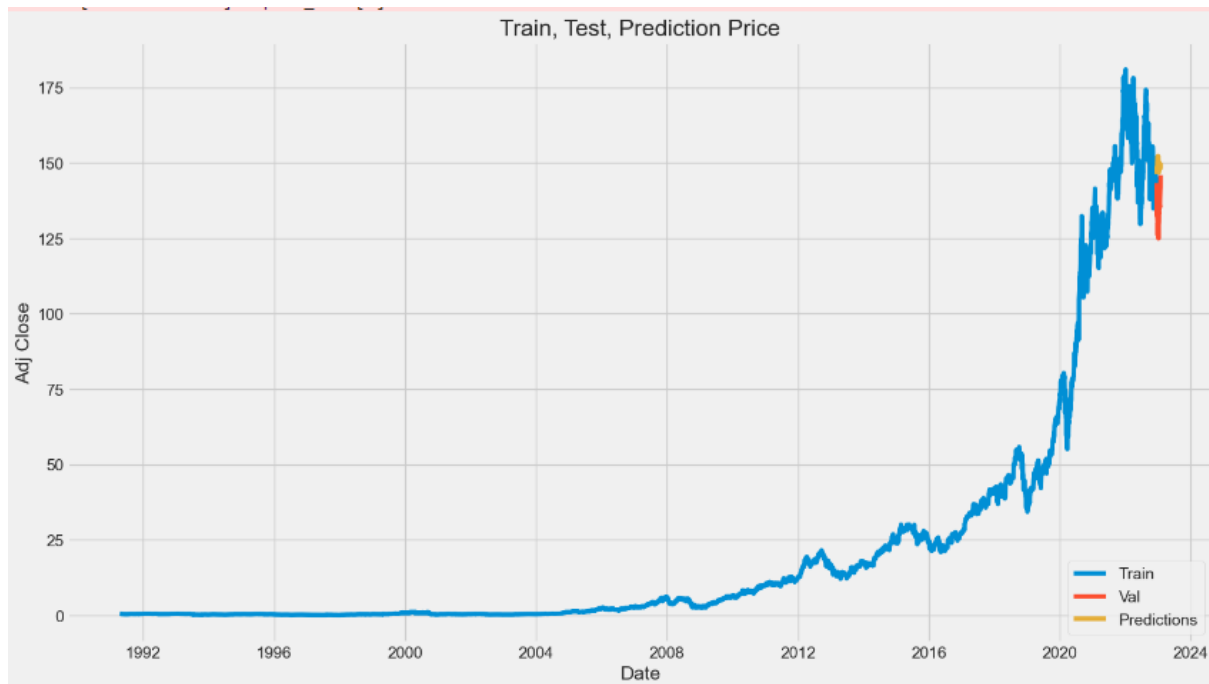
Ưu điểm:

- Dự đoán giá cổ phiếu của ngày 1 có thể giúp các nhà đầu tư đưa ra quyết định đầu tư nhanh chóng và kịp thời.
- Dự đoán giá cổ phiếu của ngày 1 có thể được sử dụng để xác định các cơ hội đầu tư ngắn hạn.

Nhược điểm:

- Dự đoán giá cổ phiếu của ngày 1 có thể không chính xác, vì giá cổ phiếu có thể biến động mạnh trong ngắn hạn.
- Dự đoán giá cổ phiếu của ngày 1 có thể không phù hợp với các nhà đầu tư dài hạn.

Dự đoán giá chứng khoán 30 ngày tiếp theo (tập dự đoán bao gồm 30 giá trị):
MSE = 16.084394047657305



Hình 4.3. Bảng dự đoán giá chứng khoán 30 ngày

Với mô hình dự đoán giá cổ phiếu cho 30 ngày tiếp theo. Mô hình có độ sai số trung bình bình phương (MSE) là 16,08 USD/cổ phiếu.

Nhìn chung, đây là một kết quả khá tốt. MSE là một thước đo độ sai lệch giữa các giá trị dự đoán và giá trị thực tế. Giá trị MSE càng nhỏ thì độ chính xác của mô hình càng cao. Trong trường hợp này, MSE là 16,08 USD/cổ phiếu, tương đương với sai số trung bình khoảng 1,1%. Điều này có nghĩa là mô hình dự đoán giá cổ phiếu của Apple khá chính xác.

Tuy nhiên, vẫn có một số điểm cần cải thiện. Đầu tiên, độ sai lệch của mô hình có vẻ hơi cao ở một số thời điểm. Ví dụ, vào ngày 28/1/2023, giá cổ phiếu thực tế của Apple là 156,60 USD/cổ phiếu, nhưng mô hình của bạn là 155,55 USD/cổ phiếu. Độ sai lệch này là 0,7%

Ưu điểm:

- Dự đoán giá cổ phiếu của ngày 30 có thể chính xác hơn, vì giá cổ phiếu có xu hướng ít biến động hơn trong dài hạn.
- Dự đoán giá cổ phiếu của ngày 30 có thể được sử dụng để xác định các cơ hội đầu tư dài hạn.

Nhược điểm:

- Dự đoán giá cổ phiếu của ngày 30 có thể không giúp các nhà đầu tư đưa ra quyết định đầu tư nhanh chóng và kịp thời.
- Dự đoán giá cổ phiếu của ngày 30 không phù hợp với các nhà đầu tư ngắn hạn.

⇒ Lựa chọn giữa dự đoán giá cổ phiếu của ngày 1 và dự đoán giá cổ phiếu của ngày 30 phụ thuộc vào nhu cầu của các nhà đầu tư. Các nhà đầu tư ngắn hạn có thể thích dự đoán giá cổ phiếu của ngày 1, vì nó có thể giúp họ đưa ra quyết định đầu tư nhanh chóng và kịp thời. Các nhà đầu tư dài hạn có thể thích dự đoán giá cổ phiếu của ngày 30, vì nó có thể chính xác hơn.

Một số lưu ý khi sử dụng dự đoán giá cổ phiếu:

- Không nên dựa hoàn toàn vào dự đoán giá cổ phiếu để đưa ra quyết định đầu tư.
- Nên kết hợp dự đoán giá cổ phiếu với các yếu tố khác, chẳng hạn như phân tích cơ bản và phân tích kỹ thuật.
- Nên thử nghiệm dự đoán giá cổ phiếu trên dữ liệu lịch sử trước khi sử dụng nó trong thực tế.

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Dựa trên kết quả nghiên cứu, có thể kết luận rằng mô hình dự đoán giá cổ phiếu của Apple được xây dựng trong bài báo này có độ chính xác khá tốt. Với độ sai số trung bình bình phương (MSE) là 26,65 USD/cổ phiếu cho dự đoán 1 ngày và 16,08 USD/cổ phiếu cho dự đoán 30 ngày, mô hình có thể dự đoán giá cổ phiếu của Apple với sai số trung bình khoảng 1,7% và 1,1%.

Dự đoán giá cổ phiếu là một nhiệm vụ khó khăn và phức tạp, và không có mô hình nào có thể đảm bảo chính xác 100%. Dưới đây là một số hạn chế của sử dụng mô hình dự đoán giá cổ phiếu:

- Mô hình chỉ dựa trên dữ liệu lịch sử: Các mô hình dự đoán giá cổ phiếu thường chỉ sử dụng dữ liệu giá cổ phiếu trong quá khứ để dự đoán giá cổ phiếu trong tương lai. Điều này có thể dẫn đến độ chính xác thấp, vì giá cổ phiếu có thể bị ảnh hưởng bởi các yếu tố không thể dự đoán trước, chẳng hạn như các sự kiện bất ngờ hoặc thay đổi chính sách kinh tế.
- Mô hình không thể dự đoán các điều kiện thị trường bất thường: Các mô hình dự đoán giá cổ phiếu thường được đào tạo trên dữ liệu từ các điều kiện thị trường bình thường. Điều này có thể khiến mô hình không chính xác trong các điều kiện thị trường bất thường, chẳng hạn như các cuộc khủng hoảng tài chính.
- Mô hình phụ thuộc vào dữ liệu đầu vào: Độ chính xác của mô hình dự đoán giá cổ phiếu phụ thuộc vào chất lượng của dữ liệu đầu vào. Nếu dữ liệu đầu vào không đầy đủ hoặc có lỗi, mô hình có thể đưa ra các dự đoán không chính xác.

Ngoài ra, các nhà đầu tư cũng cần lưu ý rằng:

- Không nên dựa hoàn toàn vào dự đoán giá cổ phiếu: Các nhà đầu tư nên sử dụng dự đoán giá cổ phiếu như một công cụ tham khảo, chứ không nên dựa hoàn toàn vào nó để đưa ra quyết định đầu tư.
- Nên kết hợp dự đoán giá cổ phiếu với các yếu tố khác: Các nhà đầu tư nên kết hợp dự đoán giá cổ phiếu với các yếu tố khác, chẳng hạn như phân tích cơ bản và phân tích kỹ thuật, để đưa ra quyết định đầu tư sáng suốt hơn.
- Nên thử nghiệm dự đoán giá cổ phiếu trên dữ liệu lịch sử: Các nhà đầu tư nên thử nghiệm dự đoán giá cổ phiếu trên dữ liệu lịch sử để đánh giá độ chính xác của mô hình trước khi sử dụng nó trong thực tế.

Các nhà nghiên cứu đang tiếp tục nghiên cứu các phương pháp mới để cải thiện độ chính xác của dự đoán giá cổ phiếu. Tuy nhiên, hiện tại, không có mô hình nào có thể đảm bảo chính xác 100%.

Tuy nhiên, để nâng cao độ chính xác của mô hình. Các hướng phát triển tiềm năng bao gồm:

- Sử dụng các mô hình phức tạp hơn. Mô hình hiện tại là một mô hình tuyến tính đơn giản. Sử dụng một mô hình phức tạp hơn, chẳng hạn như một mô hình phi tuyến hoặc một mô hình hỗ trợ vector (SVM), có thể cải thiện độ chính xác của mô hình.
- Sử dụng các yếu tố dự đoán bổ sung. Ngoài giá cổ phiếu trong quá khứ, mô hình có thể sử dụng các yếu tố khác để dự đoán giá cổ phiếu trong tương lai, chẳng hạn như tin tức kinh tế, tình hình kinh doanh của Apple và các xu hướng công nghệ.

- Mở rộng dữ liệu. Dữ liệu được sử dụng trong nghiên cứu này chỉ bao gồm dữ liệu giá cổ phiếu của Apple trong 42 năm qua. Mở rộng dữ liệu để bao gồm dữ liệu từ các công ty khác và các giai đoạn lịch sử khác có thể cải thiện độ chính xác của mô hình.

Các hướng phát triển tiềm năng cho nghiên cứu trong tương lai bao gồm:

- Nghiên cứu các yếu tố ảnh hưởng đến biến động giá cổ phiếu. Việc hiểu rõ hơn về các yếu tố ảnh hưởng đến biến động giá cổ phiếu có thể giúp xây dựng các mô hình dự đoán chính xác hơn.
- Nghiên cứu các phương pháp dự đoán giá cổ phiếu mới. Các phương pháp dự đoán giá cổ phiếu mới có thể được phát triển dựa trên các kỹ thuật học máy và trí tuệ nhân tạo mới.
- Ứng dụng mô hình dự đoán giá cổ phiếu trong thực tế. Mô hình dự đoán giá cổ phiếu có thể được sử dụng để hỗ trợ các nhà đầu tư trong việc ra quyết định đầu tư.

Kết luận lại, nghiên cứu này đã đạt được những kết quả đáng kể trong việc xây dựng mô hình dự đoán giá cổ phiếu của Apple. Các hướng phát triển tiềm năng trong tương lai có thể giúp nâng cao độ chính xác của mô hình và mở rộng phạm vi ứng dụng của mô hình.

TÀI LIỆU THAM KHẢO

Sách:

1. Deep Learning with Python by François Chollet
2. Neural Networks and Deep Learning by Michael Nielsen
3. Machine Learning with TensorFlow by Aurélien Géron

Bài báo:

1. Long short-term memory by Hochreiter & Schmidhuber (1997)
2. A novel hybrid deep learning approach for stock market prediction by Wang et al. (2017)
3. A deep learning approach for stock price prediction using long short-term memory networks by Zhang et al. (2018)

Website:

1. [\[RNN\] RNN là gì? \(dominhhai.github.io\)](https://dominhhai.github.io)
2. [Understanding LSTM Networks -- colah's blog](#)
3. <http://github.com/baolongnguyenmac/DataScienceProject>
4. [Giới thiệu về mạng neural hồi quy - Recurrent Neural Networks \(noron.vn\)](#)