|  |  |
| --- | --- |
|  | BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM** |

**ĐỒ ÁN MÔN TRÍ TUỆ NHÂN TẠO**

**Dự đoán giá cổ phiếu bằng thuật toán Long-short term memory**

Ngành: **Công nghệ thông tin**

Chuyên ngành: **Công nghệ thông tin**

Giảng viên hướng dẫn : Nguyễn Thị Hải Bình

Sinh viên thực hiện :

1. MSSV: 2080600542 Họ tên: Võ Thương Trường Nhơn

2. MSSV: 2080601324 Họ tên: Phạm Đức Tài

3. MSSV: 1911060972 Họ tên: Nguyễn Hoàng Sơn

4. MSSV: 1911061398 Họ tên: Phạm Phú Quý

TP. Hồ Chí Minh, 2022

Mục lục

[I. Phát biểu bài toán: 2](#_Toc92054981)

[1. Giới thiệu bài toán 3](#_Toc92054982)

[2. Đầu vào, đầu ra của bài toán 3](#_Toc92054983)

[3. Những vấn đề khác của bài toán: 3](#_Toc92054984)

[4. Lựa chọn thuật toán LSTM 4](#_Toc92054985)

[II. Thuật toán sử dụng 4](#_Toc92054986)

[1. Mạng Nơ-ron hồi quy – Recurrent Neural Network 4](#_Toc92054987)

[a. Giới thiệu thuật toán RNN : 4](#_Toc92054988)

[b. Phân loại bài toán RNN 4](#_Toc92054989)

[c. Dữ liệu dạng Time series data: 5](#_Toc92054990)

[d. Ứng dụng RNN vào bài toán dự đoán giá cổ phiếu 6](#_Toc92054991)

[2. Long – short term memory(LSTM) 6](#_Toc92054992)

[a. Giới thiệu về LSTM 7](#_Toc92054993)

[b. Mô hình của LSTM 7](#_Toc92054994)

[3. Giải thích thuật toán LSTM 8](#_Toc92054995)

[a. Forget gate: 8](#_Toc92054996)

[b. Input gate: 9](#_Toc92054997)

[c. Output gate: 10](#_Toc92054998)

[4. Ứng dụng vào bài toán dự đoán cổ phiếu 11](#_Toc92054999)

[a. Thư viện sử dụng 11](#_Toc92055000)

[b. Lấy data 12](#_Toc92055001)

[c. Xử lý dữ liệu: 12](#_Toc92055002)

[d. Xây dựng mô hình nơ ron LSTM 13](#_Toc92055003)

[III. Kết quả thực nghiệm: 14](#_Toc92055004)

[1. Bộ dữ liệu sử dụng: 14](#_Toc92055005)

[2. Kết quả thực nghiệm ứng với các mã cổ phiếu: 15](#_Toc92055006)

[a. Các công thức sử dụng để đánh giá 15](#_Toc92055007)

[b. Phân tích độ chính xác của mô hình trên tập test: 16](#_Toc92055008)

[IV. Kết luận và nhận xét: 21](#_Toc92055009)

[V. Tài liệu tham khảo: 22](#_Toc92055010)

1. **Phát biểu bài toán:**
   1. Giới thiệu bài toán

Đầu tư chứng khoán là một trong những hình thức đầu tư tài chính(vàng, bất động sản, trái phiếu,...). Việc sử dụng những đồng tiền nhàn rỗi mua cổ phần của các doanh nghiệp góp phần sinh ra lợi nhuận không nhỏ. Giúp tạo ra một khoản thu nhập bị động và chống lại sự mất giá của đồng tiền.

Dẫu biết giá cổ phiếu phụ thuộc vào nhiều yếu tố dễ biến động, luôn thay đổi khó lường và phi tuyến tính. Dự đoán giá cổ phiếu chính xác là vô cùng khó khăn vì phụ thuộc vào nhiều yếu tố vĩ mô và vi mô, chẳng hạn như yếu tố chính trị, bối cảnh kinh tế trong nước và kinh tế toàn cầu, các cú sốc bất ngờ (Covid-19, thiên tai, mất mùa,...), tình hình hoạt động tài chính của công ty, ...

Tuy nhiên, tất cả những điều này cũng có nghĩa là có rất nhiều dữ liệu để sử dụng, khai phá giúp tìm ra giá cổ phiếu. Vì vậy, các chuyên gia phân tích tài chính, các nhà nghiên cứu và các nhà khoa học dữ liệu đang tiếp tục khám phá các kỹ thuật phân tích để phát hiện xu hướng và quy luật vận động của thị trường chứng khoán.

* 1. Đầu vào, đầu ra của bài toán

Ở bài toán dự đoán giá chứng khoán ta sẽ phân tích, sử dụng các giá trị trong quá khứ cụ thể đầu vào của bài toán này là lịch sử giá trong 10 năm gần đây và ta sẽ  dự đoán tương lai, giá và xu hướng của của cổ phiếu trong những ngày tiếp theo.

* 1. Những vấn đề khác của bài toán:

Để dự đoán giá của một mã cổ phiếu không chỉ dựa vào giá của một ngày, một tháng hay một năm trước đó mà phải dựa vào cả quá trình tăng giảm của giá cổ phiếu trong 10 năm, 20 năm.

Dữ liệu phải có dạng chuỗi, tính liên tục vì mô hình được huấn luyện dựa trên những dữ liệu trong quá khứ các quy luật trong quá khứ sẽ được lặp lại ở tương lai. Sau đó dịch chuyển các khung thời gian để dự báo tương lai.

* 1. Lựa chọn thuật toán LSTM

Mạng LSTM được công bố vào năm 1997 do hai nhà khoa học người Đức Hochreiter và Schmidhuber phát minh. Sau này thuật toán được nhiều nhà khoa học đi sau phát triển và ngày càng trở nên hoàn thiện. LSTM là một cải tiến của mạng hồi quy truyền thống vẫn có thể xử lý dữ liệu dạng chuỗi đồng thời giúp giải quyết tốt hiện vanishing gradient trong kĩ thuật, hiện tượng gây mất các phụ thuộc ở xa, làm mất thông tin giá ở xa.

Trong lĩnh vực tài chính thì mạng Long- short term memory (LSTM),  đã đạt được nhiều thành tựu và được sử dụng rộng rãi. Mạng LSTM có thể sử dụng mạng để duy trì bộ nhớ của các sự kiện gần đây nhờ cách xử lý dữ liệu dạng chuỗi hiệu quả  và xây dựng các kết nối giữa mỗi đơn vị của mạng và khả năng học được các phụ thuộc xa nên thuật toán phù hợp đối với bài toán dự đoán giá cổ phiếu nơi mà dữ liệu ở dạng chuỗi và có tính thứ tự.

1. **Thuật toán sử dụng**

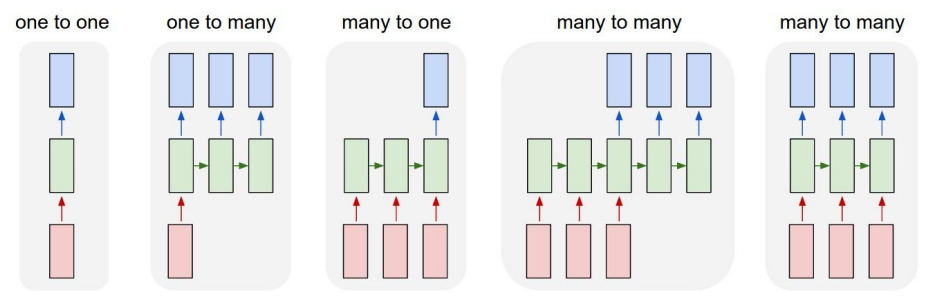
## Mạng Nơ-ron hồi quy – Recurrent Neural Network

### Giới thiệu thuật toán RNN :

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.

Với đầu vào là dữ liệu dạng chuỗi có thứ tự (time series data) một chuỗi các điểm dữ liệu, thường bao gồm các phép đo liên tiếp được thực hiện từ cùng một nguồn trong một khoảng thời gian (video, đoạn văn, giá cả thị trường,…).  Đầu ra ra là véc - tơ ŷ chứa các thông tin từ các state trước.

* + 1. Phân loại bài toán RNN



One to one: mẫu bài toán cho Neural Network (NN) và Convolutional Neural Network (CNN), 1 input và 1 output, ví dụ với bài toán phân loại ảnh MNIST input là ảnh và output là ảnh đấy là số nào.

One to many: bài toán có 1 input nhưng nhiều output, ví dụ với bài toán caption cho ảnh, input là 1 ảnh nhưng output là nhiều chữ mô tả cho ảnh đấy, dưới dạng một câu.

Many to one: bài toán có nhiều input nhưng chỉ có 1 output, ví dụ bài toán phân loại hành động trong video, input là nhiều ảnh (frame) tách ra từ video, ouptut là hành động trong video

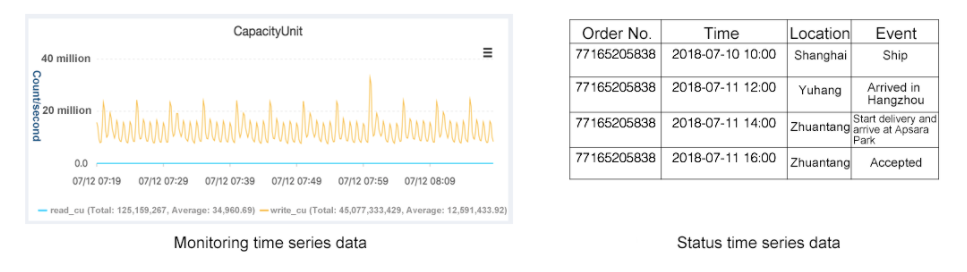
Many to many: bài toán có nhiều input và nhiều output, ví dụ bài toán dịch từ tiếng anh sang tiếng việt, input là 1 câu gồm nhiều chữ: "I love Vietnam" và output cũng là một câu gồm nhiều chữ "Tôi yêu Việt Nam". Để ý là độ dài sequence của input và output có thể khác nhau.

* + 1. Dữ liệu dạng Time series data:

Time-series Data: là một chuỗi các điểm dữ liệu, thường bao gồm các phép đo liên tiếp được thực hiện từ cùng một nguồn trong một khoảng thời gian. Phân tích chuỗi thời gian có mục đích nhận dạng và tập hợp lại các yếu tố, những biến đổi theo thời gian mà nó có ảnh hưởng đến giá trị của biến quan sát.

Trong Time-series Data, có hai loại chính:

* Chuỗi thời gian thông thường (regular time series) là số liệu.
* Chuỗi thời gian bất thường (events) là những sự kiện.



* + 1. Ứng dụng RNN vào bài toán dự đoán giá cổ phiếu

RNN có thể mang thông tin của giá cổ phiếu từ những ngày trước, cuối cùng là sự kết hợp của tất cả giá của cổ phiếu để dự đoán. Tuy nhiên RNN chỉ có thể học được những thông tin gần các thông tin ở xa thì càng bị vanishing gradient và các hệ số không được cập nhật.

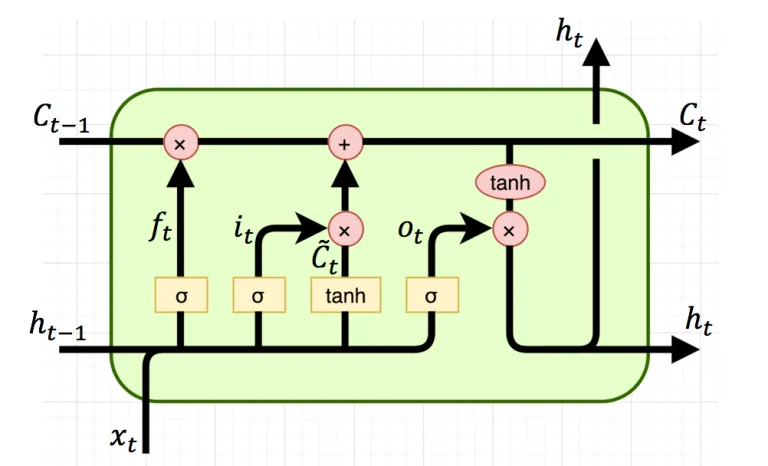
Vanishing gradient: Là vấn đề xảy ra khi huấn luyện các mạng nơ ron nhiều lớp. Khi huấn luyện, giá trị đạo hàm là thông tin phản hồi của quá trình lan truyền ngược. Giá trị này trở nên vô cùng nhỏ tại các lớp nơ ron đầu tiên khiến cho việc cập nhật trọng số mạng không thể xảy ra.

Vậy thực tế RNN chỉ mang được thông tin của một vài ngày gần đó sau đó thì sẽ bị vanishing gradient cụ thể RNN chỉ học được giá trong giá những ngày gần đây đối với những 9, 10 năm trước thì không được khiến model trở nên thiếu chính xác.

* 1. Long – short term memory(LSTM)
     1. Giới thiệu về LSTM

Long - Short Term Memory networks, thường được gọi là LSTM là một dạng đặc biệt của RNN, cũng xử lý tốt dữ liệu dạng chuỗi có thứ tự tuy nhiên LSTM có khả năng chống được hiện tượng vanishing gradient từ đó học được các phụ thuộc xa.

* + 1. Mô hình của LSTM



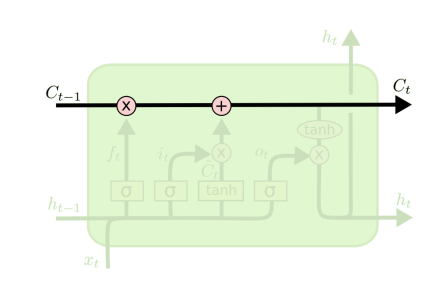
**Mô hình mạng LSTM**

* Input: Ct-1, ht-1, xt. Trong đó xt là input ở state thứ t của model. Ct-1, ht-1 là output của layer trước.
* Output: Ct, ht, ta gọi c là cell state, h là hidden state.

Ký hiệu: σ, tanh ý là bước đấy dùng sigmoid, tanh activation function. Phép nhân ở đây là element-wise multiplication, phép cộng là cộng ma trận.

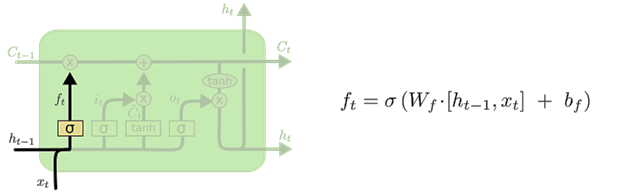
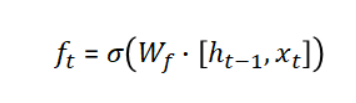
Trong đó: ft, it, Ot tương ứng với forget gate, input gate và output gate.

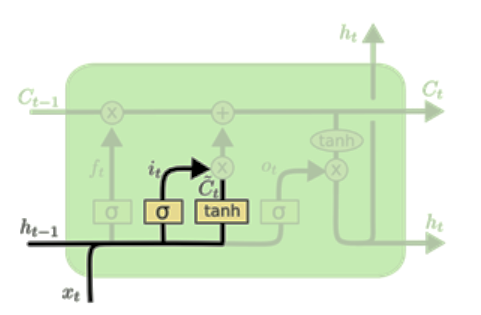
* Forget gate: ft = σ(Uf \* xf + Wf \* ht-1+ bf)
* Input gate: it = σ(Ui \* xt +Wi \* ht-1 + bi)
* Output gate: ot = σ (Uo \* xt + Wo\* ht-1+ bo)
  1. Giải thích thuật toán LSTM

Mạng LSTM là một cải tiến của mạng hồi quy truyền thống nên mô hình có những điểm mới sau:

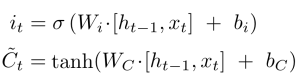
Cell state là đường ngang chạy qua đầu sơ đồ, giống như một băng chuyền, bộ nhớ của mạng LSTM. Nó chạy qua toàn bộ chuỗi, chỉ với một số tuyến tính nhỏ các tương tác LSTM có khả năng loại bỏ hoặc thêm thông tin vào cell state, được điều chỉnh cẩn thận bởi các cấu trúc được gọi là cổng.

Cổng là một cách tùy chọn để thông qua thông tin. Chúng sử dụng hàm kích hoạt sigmoid và tanh. Một LSTM có ba các cổng, để bảo vệ và kiểm soát trạng thái tế bào.

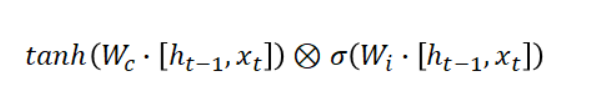
* + 1. Forget gate:
* Forget gate: thông qua lớp sigmod đưa ra quyết định thông tin có được vào cell state. Giá trị ht-1 và xt đi qua lớp sigmod cho ra một giá trị từ 0 đến 1 cho mỗi cell state.
* Lớp sigmoid xuất ra các số từ 0 đến 1, mô tả mức độ thông qua của từng thành phần. Giá trị bằng 0 có nghĩa là "không có gì thông qua", trong khi giá trị bằng một có nghĩa là "để mọi thứ thông qua!"
  + 1. Input gate:



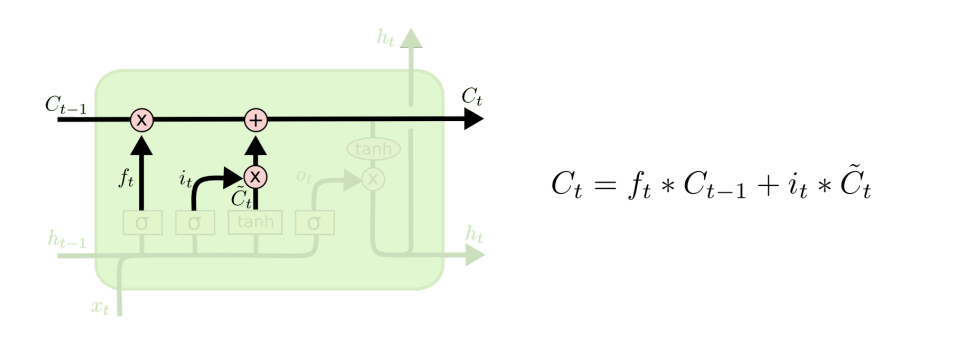
* Input gate: quyết định thông tin mới được lưu trữ ở cell state. Gồm hai phần: Lớp sigmod quyết định giá trị nào được cập nhật và một lớp tanh chứa những giá trị mới những giá trị có thể được thêm vào cell state .



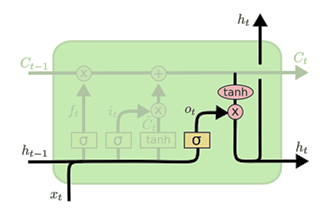
* Cuối cùng kết hợp hai thứ trên để tạo nên giá trị mới để cập nhật cho cell state.



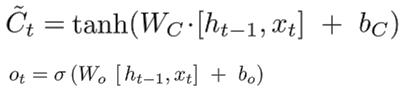
* Tiếp theo ta cập nhật cell state cũ Ct-1 bằng Ct . Nhân ft quên đi những thông tin cần quên và cộng thêm những giá trị mới



* + 1. Output gate:



* Output gate: quyết định những thông tin gì sẽ được thông qua. Đầu tiên, ta chạy một lớp sigmoid, lớp này quyết định phần nào của trạng thái tế bào mà ta sẽ xuất ra. Sau đó chúng ta đặt cell state thông qua hàm tanh (đẩy giá trị về khoảng từ -1 đến 1).



* Cuối cùng nhân với đầu ra của cổng sigmoid lấy những thông tin cần thiết.

ht = ot

* 1. Ứng dụng vào bài toán dự đoán cổ phiếu
     1. Thư viện sử dụng

Numpy (Numeric Python) == 1.21.5: là một thư viện toán học phổ biến và mạnh mẽ của Python. Cho phép làm việc hiệu quả với ma trận và mảng, đặc biệt là dữ liệu ma trận và mảng lớn với tốc độ xử lý nhanh hơn nhiều lần

Pandas == 1.3.5: là một thư viện mã nguồn mở, hỗ trợ đắc lực trong thao tác dữ liệu. Đây cũng là bộ công cụ phân tích và xử lý dữ liệu mạnh mẽ của ngôn ngữ lập trình python. Thư viện này sử dụng một cấu trúc dữ liệu riêng là Dataframe.

Scikit-learn (Sklearn) == 1.0 : là thư viện mạnh mẽ nhất dành cho các thuật toán học máy được viết trên ngôn ngữ Python. Thư viện cung cấp một tập các công cụ xử lý các bài toán machine learning và statistical modeling gồm: classification, regression, clustering, và dimensionality reduction.

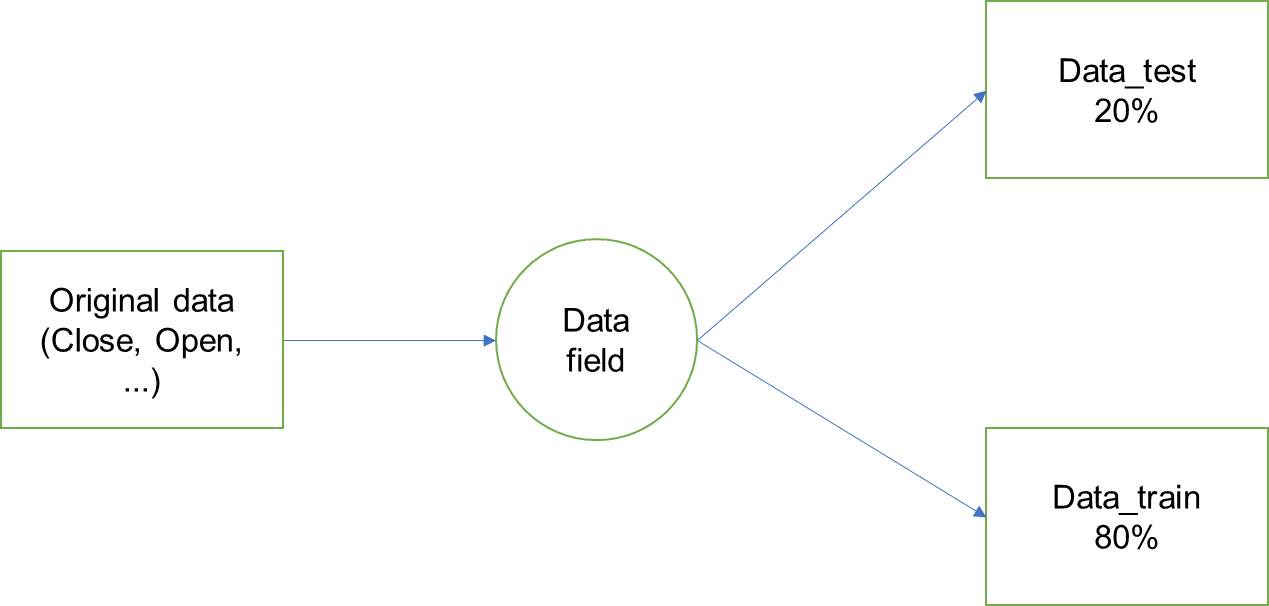
Tensorflow == 2.7.0 là thư viện mã nguồn mở hỗ trợ việc huấn luyện và suy luận trong học sâu.

Streamlit == 1.0.0: là thư viện cung cấp các công cụ giúp chúng ta có thể tạo một trang web cho Machine learning và Data Science.

Yfinance == 0.1.64 là thư viện giúp chúng ta dễ dàng tải bộ dữ liệu về từ trang yahoo.finance.com.

Matplotlib == 3.3.4 là thư viện giúp chúng ta có thể trực quan hóa dữ liệu bằng nhiều loại biểu đồ khác nhau.

* + 1. Lấy data
* Dữ liệu được tải từ trang yahoo finance
* Từ bộ dữ liệu tải về ta tách lấy trường dữ liệu mà ta sẽ sử dụng để train
* Sau khi lấy được dữ liệu cần thiết, ta chia dữ liệu thành 2 bộ dữ liệu: Data\_train và Data\_test
  + Data\_train sẽ được chúng ta sử dụng trong việc training để tạo ra Model
  + Data\_test sẽ được dùng trong việc đánh giá Model



* + 1. Xử lý dữ liệu:
* Đến đây ta sử dụng hàm MinMaxScaler của thư viện scikit learn và chia tỷ lệ tập dữ liệu thành các số trong khoảng (0, 1) để đưa vào mạng nơ ron

Ví dụ: với một tập dữ liệu bất kỳ, chúng ta xác định được giá trị lớn nhất của 1 feature là 30, giá trị nhỏ nhất là -10. Như vậy, với 1 giá trị bất kỳ là 18.8:

y = (x - min) / (Max - min)

y = (18.8 - (-10)) / (30 - (-10)) = 28.8 / 40

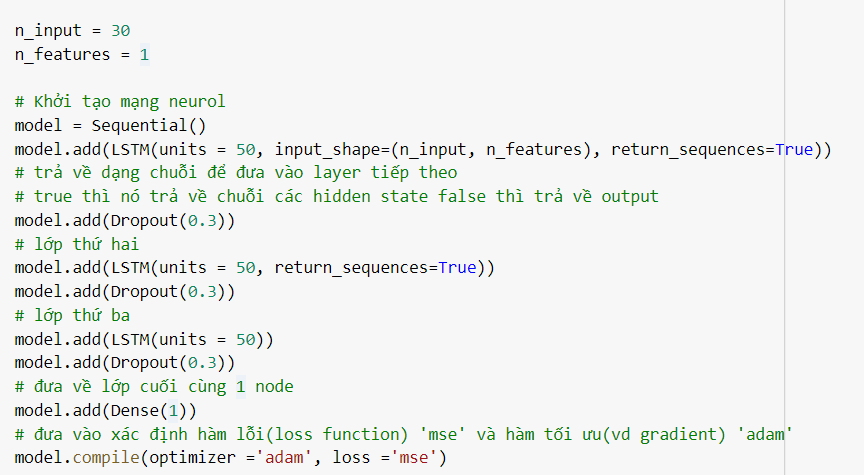
y = 0.72

### Xây dựng mô hình nơ ron LSTM

Bước đầu tiên, chúng ta cần khởi tạo lớp Sequential. Sequential là mô hình mà các lớp(layer) được xếp chồng tuyến tính

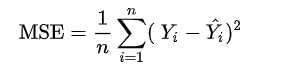
Lớp mô hình của bài toán gồm có các lớp LSTM, Dropout và Dense .

Ở trên ta thêm vào 3 lớp LSTM liên tiếp, và cứ qua 1 lớp là có 1 dropout 0.3 . Cuối cùng ta cho qua một tầng Dense với đầu ra là 1 chiều.

****

**Hàm loss :** một phép ước lượng là trung bình của bình phương các sai số. Giúp đánh giá độ chính xác của mô hình, mse càng nhỏ thì mô hình càng tốt. Trong bài toán này ta sử dụng hàm Mean squared error (MSE).

Công thức



Trong đó Yi là giá trị thực tế, Y^i là giá trị dự đoán

**Optimizer (**thuật toán tối ưu): là cơ sở để xây dựng mô hình neural network với mục đích "học " được các features ( hay pattern) của dữ liệu đầu vào, thuật toán cải thiện weight và bias theo từng bước có thể tìm 1 cặp weights và bias phù hợp để tối ưu hóa model. Ở đây bài toán sử dụng hàm “Adam”.

Đưa dữ liệu vào mạng nơ – ron tiến hành huấn luyện model.fit()

= > Model cần tìm

1. **Kết quả thực nghiệm:**

## Bộ dữ liệu sử dụng:



Biểu đồ giá close và volume của cổ phiếu Netflix

Bộ dữ liệu được lấy từ package finance.yahoo.com. Các thông tin về dịch chứng khoán từ ngày 08/03/2010 đến ngày 31/10/2021 bao gồm 2795 dòng và 7 cột. Các trường dữ liệu:

* Date: thời gian giao dịch
* Open: giá mở cửa (giá thực hiện tại lần khớp lệnh đầu tiên trong ngày giao dịch chứng khoán)
* High: giá cao nhất trong ngày
* Low: giá thấp nhất trong ngày
* Close: giá đóng cửa (giá thị trường của các cổ phiếu vào thời điểm đóng cửa một phiên giao dịch) (target\_variable)
* Adj Close: là giá đóng cửa của cổ phiếu được điều chỉnh để phản ánh chính xác giá trị của cổ phiếu đó sau khi hạch toán.
* Volume: khối lượng cổ phiếu giao dịch trong ngày (target variable)
  1. Kết quả thực nghiệm ứng với các mã cổ phiếu:
     1. Các công thức sử dụng để đánh giá

MSE(mean square error): trung bình tổng bình phương chênh lệch giữa kết quả thực tế và dự đoán.

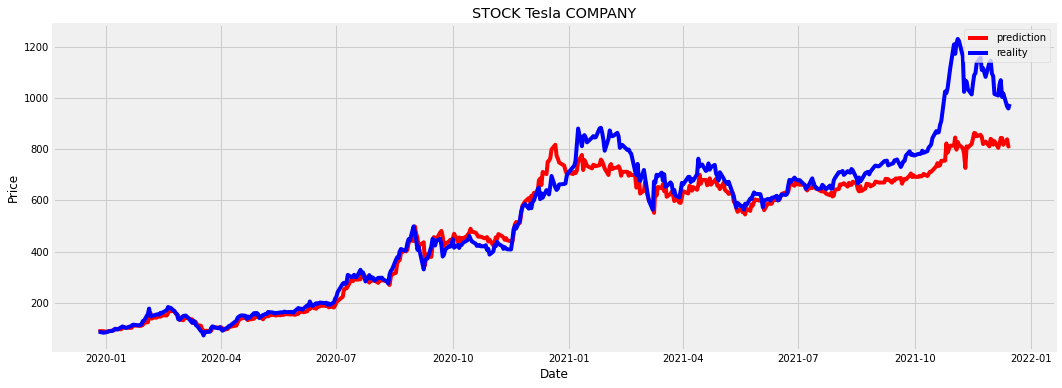
MAE(mean absolute error): trung bình tổng chênh lệch giữa kết quả thực tế và dự đoán.

Max: chênh lệch lớn nhất giữa kết quả thực tế và dự đoán.

Min: giá chênh lệch nhỏ nhất giữa kết quả thực tế và dự đoán.

Chú thích:

* là kết quả thực tế
* là kết quả dự đoán
* n số lượng kết quả thực tế (dự đoán)  
  + 1. Phân tích độ chính xác của mô hình trên tập test:
* **Độ chính xác của mô hình trên cổ phiếu Tesla:**

**** 

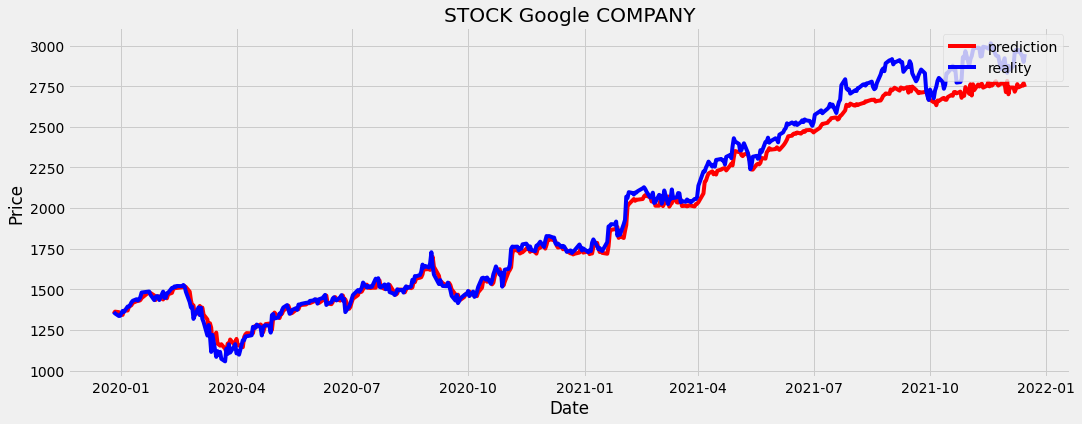
**Biểu đồ thể hiện giá dự đoán và giá thực tế của cổ phiếu Tesla giai đoạn 2020 - 2021**

* MSE = 4310.110725180327
* MAE= 43.00307399728602
* Max = 318.21807861328125
* Min = 0.0299072265625

Giá trị của phiếu Tesla có xu hướng tăng đều và tăng mạnh vào cuối năm, đường dự đoán và đường thực tế còn chưa khớp từ tháng 7/2022 đến 01/2022 dự đoán chưa chính xác cụ thể giá trị trung bình tổng bình phương chênh lệch cao khoảng 4310.11, chêch lệch giá trung bình đạt 43 $, chênh lệch giá cao nhất xấp xỉ 318.22 $, thấp nhất 0.03$

=> Mô hình dự đoán chưa tốt

**- Độ chính xác của mô hình trên cổ phiếu google:**

****



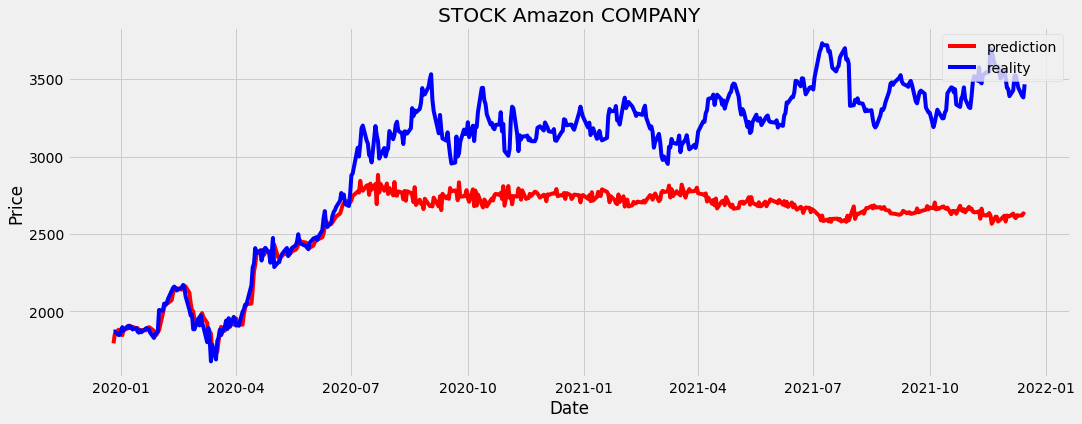
**Biểu đồ thể hiện giá dự đoán và giá thực tế của cổ phiếu Google giai đoạn 2020 – 2021**

* MSE = 3688.7373961612902
* MSA= 39.51614175458844
* Max = 291.381103515625
* Min = 0.021484375

Giá trị của phiếu Google có xu hướng tăng đều, đường dự đoán và đường thực tế còn khớp từ tháng 6/2021 đến 01/ 2022. Tuy nhiên từ tháng 6/2021 đến 01/ 2022 mô hình dự đoán chưa chính xác cụ thể giá trị trung bình tổng bình phương cao khoảng 3688.74, chêch lệch giá trung bình cao đạt 39.52 $, chênh lệch giá cao nhất xấp xỉ 291.38 $, thấp nhất 0.02$

=> Mô hình dự đoán tốt phần đầu từ 6/2021 mô hình dự đoán chưa tốt.

**- Độ chính xác của mô hình trên cổ phiếu Amazon:**

****



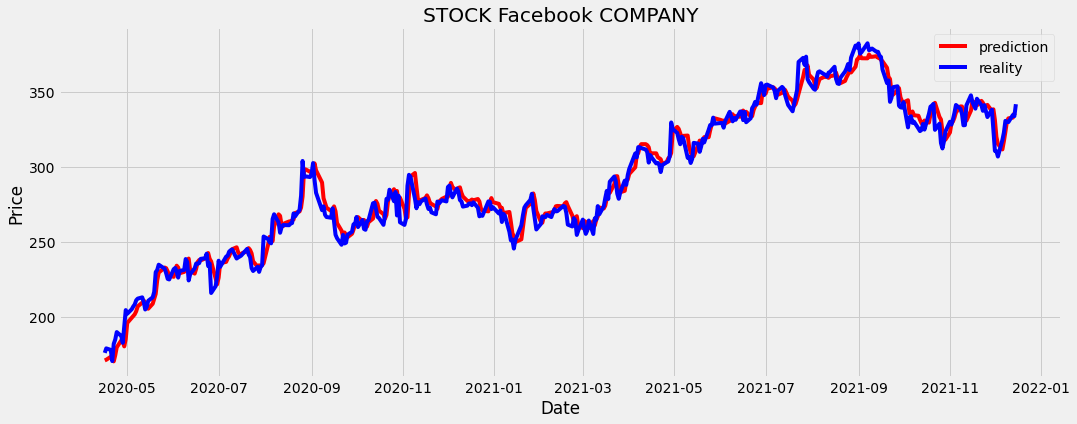
**Biểu đồ thể hiện giá dự đoán và giá thực tế của cổ phiếu Amazon giai đoạn 2020 - 2021**

* MSE = 245735.50692009853
* MAE= 303.2560640559828
* Max = 1386.48291015625
* Min = 0.12939453125

Giá trị của phiếu Amazon có xu hướng tăng, từ tháng 01/2020 đến 08/2020 giá cổ phiếu Amazon tăng mạnh và từ 08/2020 đến 01/2022 giá cổ phiếu biến động mạnh tăng giảm liên tục, đường dự đoán và đường thực tế còn khớp từ tháng 01/2020 đến 08/2020. Tuy nhiên từ tháng 08/2020 đến 01/2022 mô hình dự đoán rất thiếu chính xác cụ thể giá trị trung bình tổng bình phương rất cao khoảng 245735.51, chêch lệch giá trung bình cao đạt 303.26 $, chênh lệch giá cao nhất xấp xỉ 1386.48 $, thấp nhất 0.13$

=> Mô hình dự đoán rất tệ, khi giá dự đoán biến thiên liên tục mà mô hình dựa vào lịch sử giá làm mô hình dự đoán sai

**- Độ chính xác của mô hình trên cổ phiếu Facebook:**

****



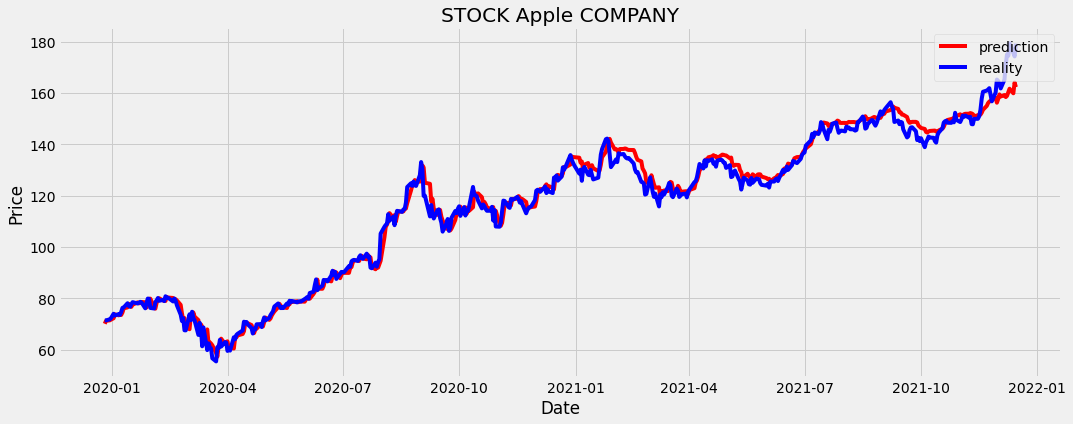
**Biểu đồ thể hiện giá dự đoán và giá thực tế của cổ phiếu Facebook giai đoạn 2020 – 2021**

* MSE = 47.55213519067378
* MAE = 5.282921711782391
* Max = 23.76751708984375
* Min = 0.061798095703125

Giá trị của phiếu Facebook có xu hướng tăng, từ biểu đồ trên ta thấy đường dự đoán khớp với đường thực tế cụ thể giá trị trung bình tổng bình phương khoảng 47.55, chêch lệch giá trung bình thấp đạt 5.28 $, chênh lệch giá cao nhất xấp xỉ 23.77 $, thấp nhất 0.06 $

=> Mô hình dự đoán tốt.

**- Độ chính xác của mô hình trên cổ phiếu Apple:**

****



**Biểu đồ thể hiện giá dự đoán và giá thực tế của cổ phiếu Apple giai đoạn 2020 - 2021**

* MSE = 11.935520102164654
* MAE = 1.9732292899250115
* Max = 17.104339599609375
* Min = 0.000919342041015625

Giá trị của phiếu Apple có xu hướng tăng, từ biểu đồ trên ta thấy đường dự đoán khớp với đường thực tế cụ thể giá trị trung bình tổng bình phương xấp xỉ 11.94, chêch lệch giá trung bình thấp đạt 1.97 $, chênh lệch giá cao nhất xấp xỉ 17.1 $, thấp nhất 0.0009 $

=> Mô hình dự đoán tốt.

1. **Kết luận và nhận xét:**

Sau khi chạy thực nghiệm ứng với mỗi mã cổ phiếu thì mô hình cho ra những kết quả với độ chính xác khác nhau. Chi tiết đối với mã cồ phiếu Amazon, Google, Tesla: giá thực tế và giá đự đoán lệch nhau nhiều mô hình dự đoán còn chưa tốt làm ảnh hưởng đến chất lượng của cuộc giao dịch tuy nhiên có thể dự đoán xu hướng tăng hay giảm của mã cổ phiếu. Còn đối với mã Facebook và Apple thì giá đự đoán khá khớp với giá thực tế.   
  
Tuy nhiên trên thực tế thị trường chứng khoán không chỉ phụ thuộc vào những con số mà còn dựa vào yếu tố chính trị, bối cảnh kinh tế trong nước và kinh tế toàn cầu, các cú sốc bất ngờ (Covid-19, thiên tai, mất mùa,...), tình hình hoạt động tài chính của công ty, …

Bên cạnh đó thì việc dự đoán còn thiếu đi sự chính xác do mạng LSTM vẫn có nhiều nhược điểm như: Thông tin phải được xử lý tuần tự, chỉ học được thông tin từ các state phía trước, không thể học được các thông tin ở xa do vanishing gradient.

Cách cải thiện và hướng phát triển: thêm dữ liệu, thêm feature(Lợi nhuận theo quý, Thu nhập từ hoạt động dịch vụ, Chi phí hoạt động khác, Chi phí quản lý doanh nghiệp, ...). Nghiên cứu sử dụng thuật toán khắc phục được những nhược điểm của mạng LSTM.

1. **Tài liệu tham khảo:**

1.https://vi.wikipedia.org/wiki/Sai\_số\_toàn\_phương\_trung\_bình

Truy cập ngày 20/12/2021 Wikipedia

2.<https://ndquy.github.io/posts/cac-phuong-phap-scaling/>

truy cập ngày 28/12/2021 ndquy blog

3. <https://medium.com/analytics-vidhya/long-short-term-memory-networks-23119598b66b>

truy cập ngày 28/12/2021 Author Vinithavn – medium.com

4. <https://medium.com/@asmello/introduction-to-model-evaluation-part-1-regression-and-classification-metrics-e75179d01db>

truy cập ngày 28/12/2021 Author André Mello – medium.com

5. <https://nttuan8.com/bai-14-long-short-term-memory-lstm/>

truy cập ngày 1/12/2021 Author Nguyễn Thanh Tuấn

6. https://nttuan8.com/bai-13-recurrent-neural-network/

truy cập ngày 1/12/2021 Author Nguyễn Thanh Tuấn

7. <https://www.kaggle.com/towarddatascience/sample-code?fbclid=IwAR3St8P2IhS6r18Oso18_PYLObLjH03lIUYoFusasFMR0tOKHM4_i0xv2As>

truy cập ngày 27/12/2021 – Kaggle

8. <https://viblo.asia/p/optimizer-hieu-sau-ve-cac-thuat-toan-toi-uu-gdsgdadam-Qbq5QQ9E5D8>

truy cập ngày 27/12/2021 – Author Trần Trung Trực

9. <https://streamlit.io>

truy cập ngày 15/12/2021 – Streamlit

10. <https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714>

truy cập ngày 15/12/2021 – Author Shi Yan

11. <https://stanford.edu/~shervine/l/vi/teaching/cs-230/cheatsheet-recurrent-neural-networks>

truy cập ngày 15/12/2021 - Author Shervine Amidi

12. <https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577>

truy cập ngày 15/12/2021 - Author Nir Arbel

13. <https://dominhhai.github.io/vi/2018/04/nn-bp>

truy cập ngày 28/11/2021- Hai's Blog