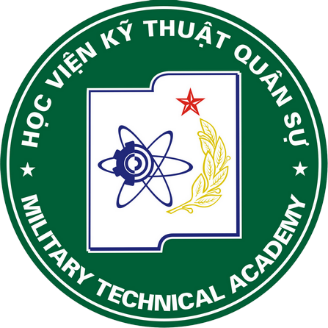
HỌC VIỆN KỸ THUẬT QUÂN SỰ  
**KHOA CÔNG NGHỆ THÔNG TIN  
——**🙞🙜🕮🙞🙜**——**

****

**BÁO CÁO BÀI TẬP LỚN**

**DỰ ĐOÁN GIÁ CHỨNG KHOÁN BẰNG MÔ HÌNH**

**LONG SHORT-TERM MEMORY**

**Giảng viên HD: Phạm Văn Việt**

**Học viên thực hiện: Hồ Huy Luật**

**Nguyễn Sĩ Hiệp**

***Vĩnh Phúc, năm 2023***

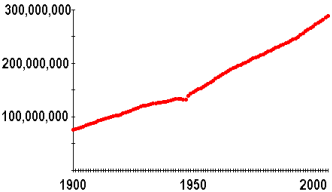
CHƯƠNG I – PHÁT BIỂU BÀI TOÁN

1. Phương pháp đề tài

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 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 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 1 : Biểu đồ Time Series của dân số Hoa Kỳ 1900 – 2000*

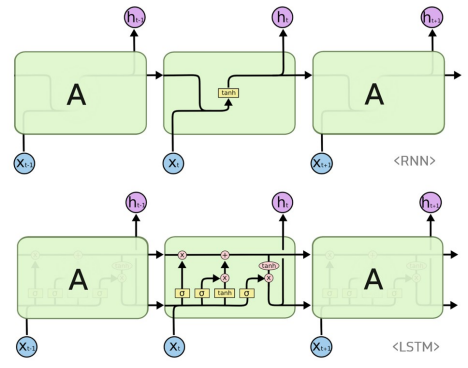
Đối với dạng bài toán có Time Series thì sẽ sử dụng mạng Neural mô hình Long Short-Term Memory ( LSTM ) để dự đoán giá chứng khoán. LSTM là một dạng đặt biệt của mạng nơ-ron hồi quy ( Recurrent Neural Network gọi là RNN ). Để nói Long Short Term Memory là một trong những mạng thần kinh nhân tạo được 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).

- Việc nhớ thông tin trong suốt thời gian dài là đặc tính mặc định của chúng, chứ không cần phải đào tạo để có thể nhớ được.

- Tức là ngay nội tại của nó đã có thể ghi nhớ được mà không cần bất cứ can thiệp nào.

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 đặt biệt.



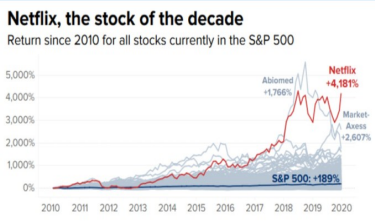
*Hình 2 : Mô hình sự khác nhau giữa RNN và LSTM*

2. Đối tượng và phạm vi dự đoán

2.1 Đối tượng dự đoán

Đối tượng ở đây muốn dự đoán là Netflix, Inc. Netflix, Inc là dịch vụ truyền dữ liệu video theo yêu cầu trên toàn cầu và cho thuê DVD trả phí tại Hoa Kỳ, nơi DVD và đĩa Blue-ray được gửi thông tin qua thư điện tử bởi Permiit Reply Mail. Công ty thành lập năm 1997 và có trụ sở tại Los Gatos, California. Năm 2002, Netflix ra mắt công chúng. Trong những năm 2003 – 2006, số lượng người dùng tích cực đã tăng từ hơn 1 triệu lên 5 triệu. Năm 2007, đánh dấu sự ra đời của tính năng phát trực tiếp cho người dùng Netflix xem ngay nội dung có sẵn và số lượng người dùng vượt 10 triệu trong năm 2007, 100 triệu vào năm 20171, 193 triệu vào năm 2020.

Cổ phiếu của Netflix ,Inc. ra mắt trên thị trường chứng khoán Mỹ vào ngày 23 tháng 5 năm 2002 với mức giá khởi đầu là 15 USD và sau đó thời điểm thị trường chứng khoán nổi lên, vốn hóa thị trường của công ty lên tới khoảng 300 triệu USD. Cùng đó sau khi ra mắt cổ phiếu Netflix bắt đầu tăng vọt cùng với số lượng người dùng nhiều. Netflix đã tăng từ mức vốn hóa ban đầu là 300 triệu USD lên hơn 240 tỷ USD và được xếp hạng trong số 40 công ty giao dịch công khai lớn nhất trên thế giới. Bên dưới là biểu đồ sự tăng trưởng của Netflix từ năm 2010 đến 2020.

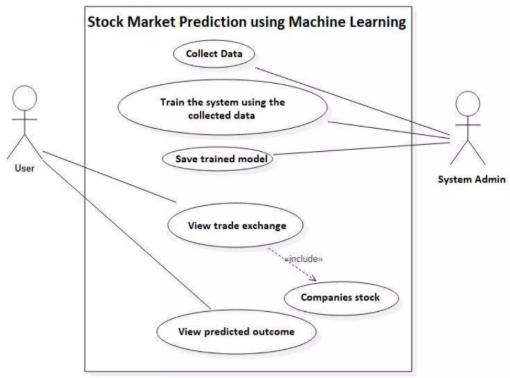


*Hình 3 : Giá chứng khoán Netflix từ năm 2010 - 2020*

2.2 Phạm vi dự đoán

Đối với một công ty có lịch sử tăng trưởng tốt như Netflix, Inc thì trong bài toán dự đoán giá chứng khoán này sẽ dự đoán giá mở cửa từ năm 2010 đến 2022. Dựa vào tất cả những giá trị giao dịch bao gồm các giá đóng cửa và giá điều chỉnh,…. Trong 12 năm để có thể dự đoán lại giá chứng khoán trong 12 năm đó. Cùng với đó sẽ dự đoán thêm giá cổ phiếu trong những ngày tương lai dựa vào mô hình LSTM đã dự đoán được giá cửa của những ngày trước đó. Từ đó có thể nhìn tổng quát được giá mở cửa dự đoán của mô hình LSTM và giá mở cửa thực tế .

3. Mô hình ca sử dụng



1. Dữ liệu ban đầu được thu thập từ các nguồn trực tuyến hoặc sàn giao dịch chứng khoán

2. Dữ liệu sau đó được sử dụng để huấn luyện hệ thống

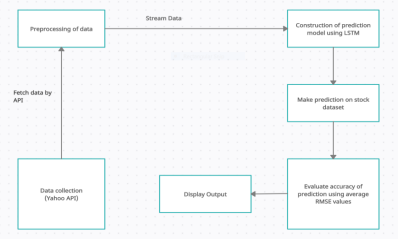
3. Mô hình đã huấn luyện được lưu

4. Người dùng xem trao đổi thương mại và cổ phiếu của một công ty

5. Sử dụng mô hình, giá đóng cửa được dự đoán.

CHƯƠNG II – THIẾT KẾ

1. Thiết kế kiến trúc



- Tiền xử lý dữ liệu: Dữ liệu giá cổ phiếu lịch sử được thu thập và xử lý trước để loại bỏ các giá trị còn thiếu, chuẩn hóa dữ liệu và chia thành các tập huấn luyện, xác thực và kiểm tra.

- Huấn luyện mô hình: Mô hình LSTM được huấn luyện bằng cách sử dụng tập dữ liệu huấn luyện và hàm mất mát được tối ưu hóa bằng cách sử dụng lan truyền ngược theo thời gian. Việc cắt giảm độ dốc có thể được sử dụng để ngăn chặn độ dốc bùng nổ trong quá trình đào tạo.

- Dự đoán: Sau khi mô hình được đào tạo, nó có thể được sử dụng để dự đoán giá cổ phiếu trong tương lai bằng cách sử dụng dữ liệu lịch sử mới nhất làm đầu vào.

- Đánh giá mô hình và xuất đầu ra: Hiệu suất của mô hình LSTM được đánh giá trên bộ xác thực để xác định các siêu tham số tốt nhất và ngăn chặn quá khớp. Bộ kiểm tra được sử dụng để đánh giá hiệu suất cuối cùng của mô hình.

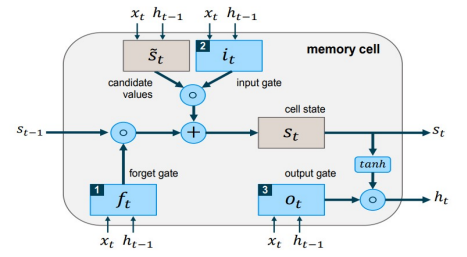
Nhìn chung, kiến trúc của mô hình LSTM để dự đoán giá cổ phiếu bao gồm sự kết hợp của quá trình tiền xử lý dữ liệu, các lớp LSTM, các lớp dày đặc cũng như các kỹ thuật đào tạo và đánh giá được thiết kế để tìm hiểu và dự đoán các mẫu phức tạp trong dữ liệu giá cổ phiếu lịch sử.

2. Mô tả bài toán hạt nhân và dữ liệu

Để giải quyết bài toán dự đoán giá chứng khoán của Netflix. Thì sẽ dùng mạng Neural mô hình LSTM để dự đoán giá trị cổ phiếu trong ngày tiếp theo, bộ dữ liệu được tải về từ Yahoo! Finance. Sử dụng giá mở cửa phiên giao dịch để dự đoán. Giá chứng khoán là một chuỗi các giá trị thay đổi theo thời gian N, gọi P là giá đóng cửa trong ngày i thì 0 < i < N. Sẽ tạo ra một cửa sổ trượt theo trục thời gian, kích thước của w là cố định và step mỗi lần lượt đúng bằng kích thước này để không xảy ra sự chồng chéo dữ liệu. Ý tưởng là dùng dữ liệu trong cửa sổ w(t) để dự đoán w(t + 1).

2. Áp dụng thuật toán vào bài toán

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



*Hình 4 : Biễu diễn kiến trúc bên trong của một Cell 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, Cell LSTM quyết định những thông tin nào cần được loại bỏ từ Cell Internal State ở bước thời gian trước đó S(t-1). Activation value Ft của Forget Gate tại bước thời gian t được dựa trên giá trị đầu vào hiện tại của Xt, giá trị đầu ra H(t-1) từ Cell LSTM ở bước trước đó và bias Bf của Forget Gate. Hàm sigmoid function biến đổi tất cả activation value về miền có giá trị trong khoảng từ 0 ( hoàn toàn quên ) và 1 ( hoàn toàn ghi nhớ ):



∙ Ở bước thứ hai, Cell LSTM quyết định những thông tin nào cần được thêm vào Cell Internal State St. Bước này bao gồm hai quá trình hoán đối S~t và Ft. Candidate value S~t biểu diễn những thông tin tiềm năng cần được thêm vào Cell Internal Sate được tính như sau :

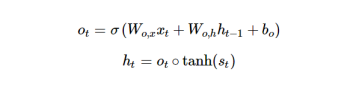


Activation value It của Input Gate theo đó cũng được tính như sau: 

∙ Ở bước thứ ba, giá trị mới của Cell Internal State St được tính dựa trên kết quả thu được từ các bước trước với phép nhân Hadamard theo từng phần tử ( Hadamard product ) được ký hiệu bằng o :

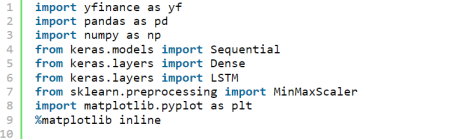
∙

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

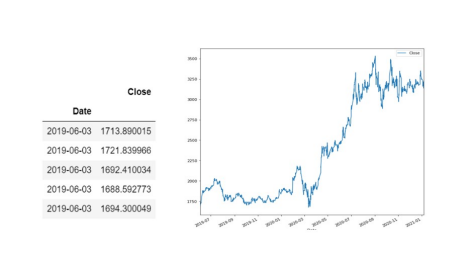


3. Thực nghiệm với thư viện Python

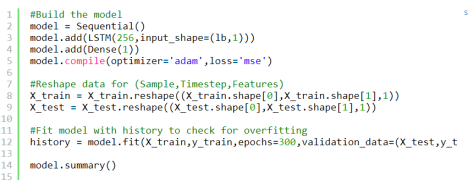
Ví dụ dự đoán giá chứng khoán của AMZN ( Amazon ) với mô hình LSTM. Giả sử để dự đoán giá đóng cửa trong khoản thời gian được chỉ định. Sử dụng thư viện Keras để xây dựng mô hình LSTM để dự đoán và một số thư viện cần thiết như Numpy, Pandas, Matplotlib.



Xây dựng một mô hình LSTM để dự đoán giá cổ phiếu hằng giờ. Giả dụ lấy giá đóng cửa của AMZN ( Amazon ) từ ‘2019-06-01’ đến ‘2021-01-07’ . In ra 5 dòng đầu tiên bằng thư viện Pandas và vẽ giá đóng cửa bằng thư viện Matplotlib.

*Hình 5 : Dữ liệu và biểu đồ giá đóng cửa của AMZN ( Amazon )*

Sau khi đã có dữ liệu thu thập thì ta sẽ biến đổi các giá trị trong các cột dữ liệu về định dạng mà mô hình LSTM có thể training và đưa ra dự đoán được bằng thư viện hỗ trợ xử lý dữ liệu như Numpy và bình thường hóa dữ liệu ( Normalize Data). Dùng thư viện Keras để tạo ra mô hình LSTM với nhiều lựa chọn mà thư viện cung cấp. Tạo LSTM bằng Keras rồi sau đó thêm các thông số cần cho một mô hình LSTM, sau khi tạo xong mô hình thì chỉ cần huấn luyện mô hình cho tập dữ liệu đã được chuẩn bị và xử lý ở bước trước từ đoạn mã bên dưới.



Bên dưới cho ra kết quả của mô hình LSTM sau khi được huấn luyện dựa trên giá đóng cửa và được trực quan hóa qua biểu đồ đường để ta có thể nhìn thấy được giá đóng cửa dự đoán của mô hình LSTM ( Predicted ) và giá đóng cửa thực tế ( Actual ) là rất gần giống nhau.



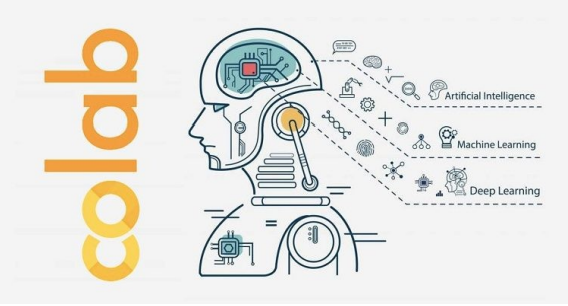
*Hình 6 : Biểu đồ giá đóng cửa dựa đoán và thực tế.*

CHƯƠNG III – XÂY DỰNG ỨNG DỤNG DỰ ĐOÁN GIÁ

CHỨNG KHOÁN NETFLIX BẰNG MÔ HÌNH LONG SHORT TERM MEMORY (LSTM)

1. Môi trường thực hiện và chuẩn bị dữ liệu

1.1. Google Colab

Google Colab là một sản phẩm từ Google Research, cho phép thực thi Python trên nền tảng đám mây. Google Colab có những lợi ích có sẵn cho Python và sử dụng miễn phí GPU cho các dự án AI, ML, DL.

*Hình 7 : Google Colab sử dụng để phát triển ứng dụng AI, ML, DL*

1.2 Chuẩn bị dữ liệu

Dữ liệu được lấy từ trang web Yahoo Finance (https://finance.yahoo.com/). Lấy dữ liệu chứng khoán của Netflix (NFLX) giá mở cửa, đóng cửa, cao nhất,… trên trang web và tải về dưới dạng CSV (Nflx.csv) Ngày bắt đầu dữ liệu từ 04/1/2010 đến 13/12/2022.

Nhìn vào bảng dữ liệu ta có các cột như :

⮚ Date : Ngày

⮚ Open : Giá mở cửa

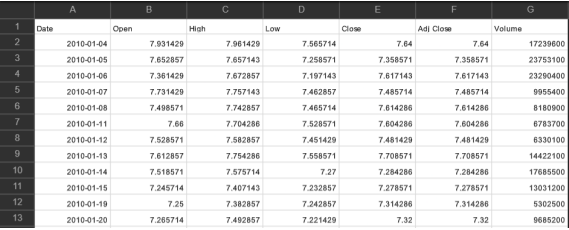
⮚ High : Giá cao nhất

⮚ Low : Giá thấp nhất

⮚ Close : Giá đóng cửa

⮚ Adj Close : Giá điều chỉnh

⮚ Volume : Tổng khối lượng giao dịch .



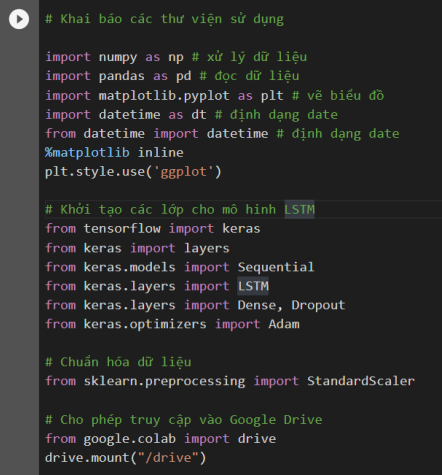
*Hình 8 : Chi tiết dữ liệu chứng khoán Netlix NFLX.CSV*

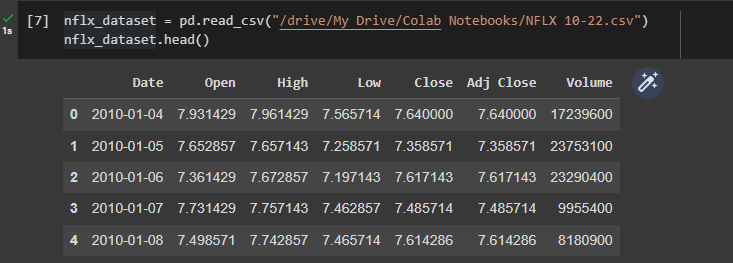
2. Xây dựng ứng dụng ( ngôn ngữ Python ).

Bước 1 : Đọc dữ liệu và tiền xử lý dữ liệu ( Read data and Data pre-processing )

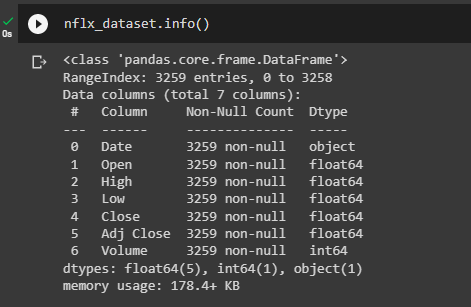
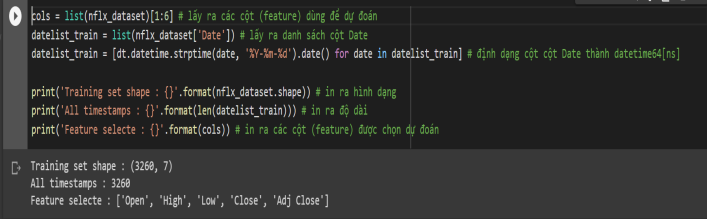
- Thêm các thư viện cần dùng trong việc xử lý dữ liệu, đọc dữ liệu và trực quan hóa dữ liệu .

- Thư viện Tensorflow, Keras cho phép gọi tạo mô hình LSTM.

- Thư viện Sklearn dùng để chuẩn hóa dữ liệu trước khi đưa vào mô hình LSTM.

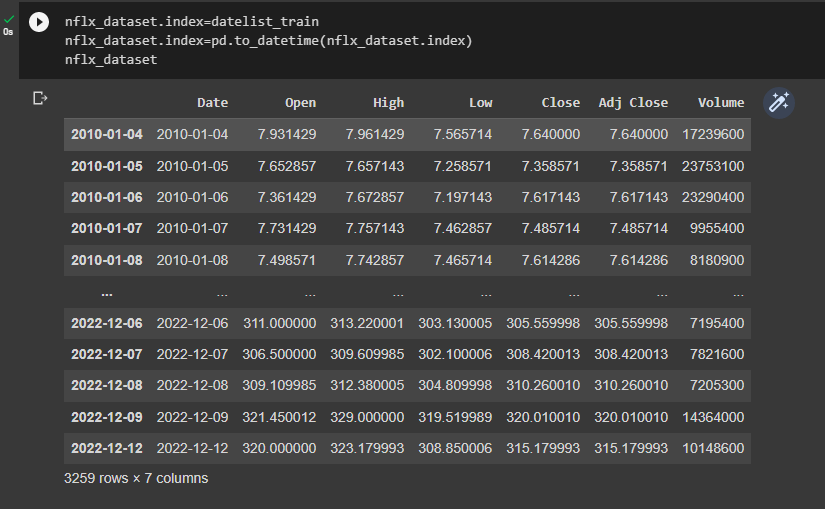
- Đọc file NFLX.CSV nhờ thư viện Pandas ( pd.read\_csv) và in ra 5 dòng đầu tiên của tập dữ liệu.

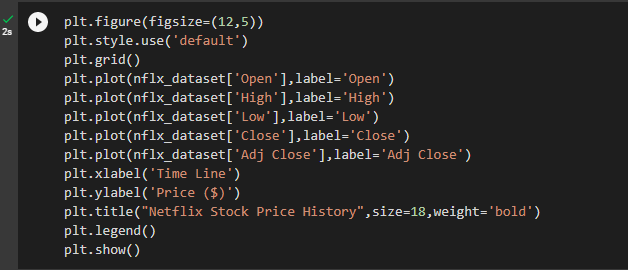
- In ra thông tin dữ liệu của mỗi cột trong file NFLX.CSV bao gồm 3259 dòng và 7 cột.



- Lấy ra 5 cột bao gồm : Open, High, Low, Close, Adj Close để dùng làm đầu vào cho quá trình dự đoán. Trích xuất cột Date dùng để định dạng từ kiểu dữ liệu Object sang kiểu dữ liệu Datetime64[ns].

- Dùng danh sách cột Date đã định dạng trước đó áp vào bộ dữ liệu làm chỉ mục (index) cho tập dữ liệu .

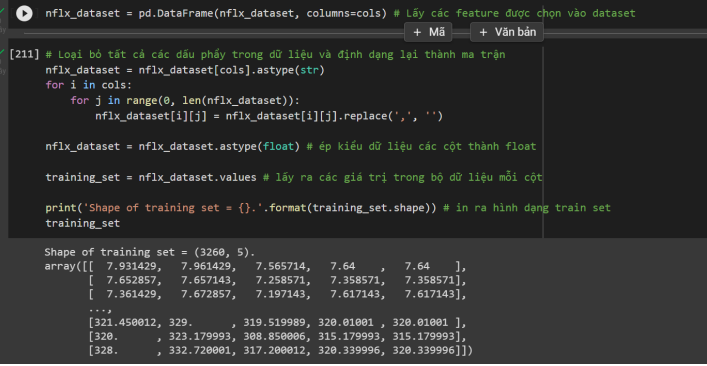
- Trực quan hóa dữ liệu lịch sử tăng giảm giá chứng khoán qua từng năm.



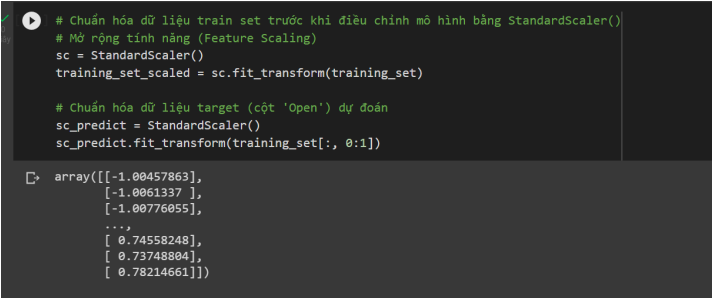


*Hình 9 : Biểu đồ giá chứng khoán của Netflix trong vòng 12 năm*

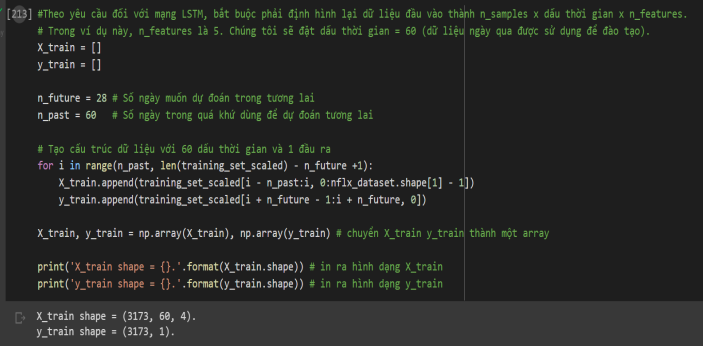
- Lấy 5 cột đã chọn cho dự đoán tạo thành một tập dữ liệu bao gồm 5 cột đã trích xuất : Open, High, Low, Close, Adj Close. Và chuyển tập dữ liệu sang dạng ma trận để có thể dùng cho mô hình LSTM.



- Vì mô hình LSTM dữ liệu trước khi đưa vào phải được chuẩn hóa để cho ra kết quả tốt nhất. sử dụng thư viện Sklearn dùng StandardScaler() để chuẩn hóa tập dữ liệu trainig\_set ( Feature Scaling ), và chuẩn hóa dữ liệu giá trị cột Open (Target).

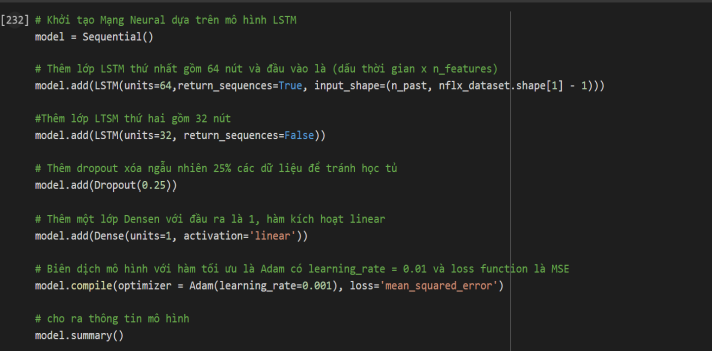


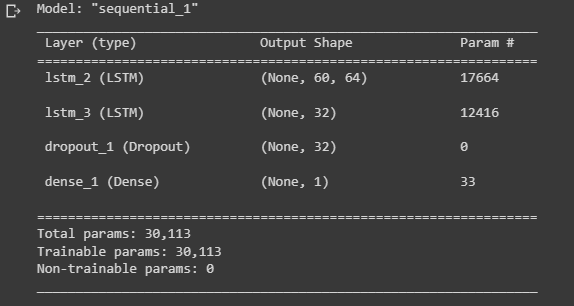
- Chia tập dữ liệu ra 2 phần bao gồm : X\_train, y\_train là dữ liệu đã được chuẩn hóa dạng ma trận đầy đủ theo yêu cầu của mạng LSTM gồm hình dạng ( n\_samples x dấu thời gian x n\_features - 1 ). Hình dạng của X\_train ( 3260, 60, 4 ), trong đó 60 là dấu thời gian với 60 ngày trước và 4 là số cột ta dự đoán ( 5 cột – 1 cột ( Adj Close ) ) ( Multi – Variable ).



Bước 2 : Xây dựng mạng Neural dựa trên LSTM và thực hiện huấn luyện mô hình

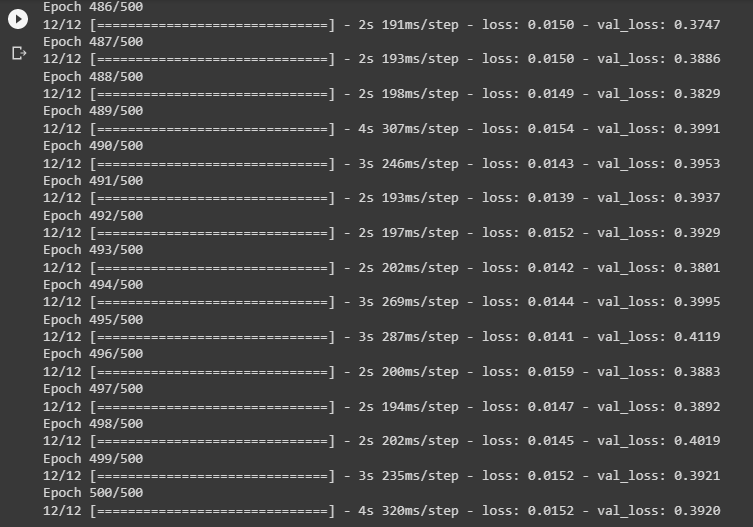
- Đầu tiên khởi tạo mạng Neural dựa trên LSTM bằng Senquence(). Thêm lớp LSTM Layer đầu tiên với 64 nút và return\_sequences = True xác định xem trả về đầu ra cuối cùng trong chuỗi đầu ra hay toàn bộ chuỗi input\_shape dưới dạng của tập huấn luyện (X\_train). Input\_shape sẽ có dạng là (dấu thời gian x n\_future – 1). Tiếp tục thêm lớp LSTM Layer thứ hai với 32 nút và return\_sequences = False không trả về đầu ra cuối cùng trong chuỗi. Thêm lớp Dropout = 0.25 tức là 25% số lớp sẽ bị loại bỏ để tránh mô hình mạng học tủ. Sau đó thêm lớp Dense với chỉ định đầu ra của 1 đơn vị và hàm kích hoạt là linear. Kế tiếp là biên dịch mô hình với trình tối ưu hóa là Adam với learning\_rate = 0.01 và Loss được tính là mean\_square\_error. Cuối cùng in ra thông tin mô hình. Bắt đầu huấn luyện mô hình với epochs = 500, batch\_size = 256.



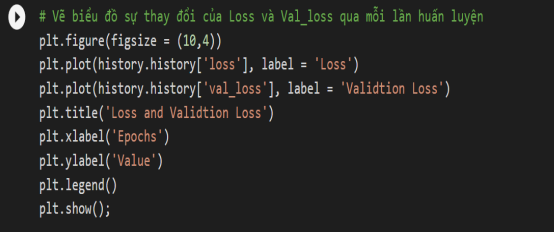


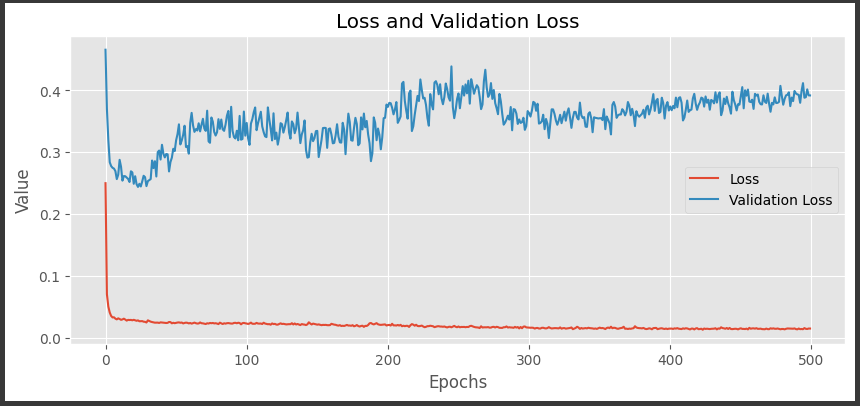
- Bắt đầu huấn luyện mô hình với epochs = 500, batch\_size = 256, valid\_split = 0.1, verbose = 1.





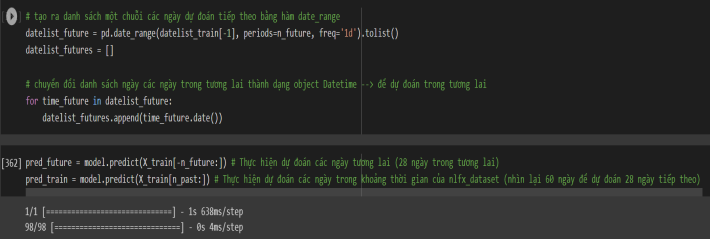
- Trực quan hóa quá trình huấn luyện và giá trị của Loss và Validition Los



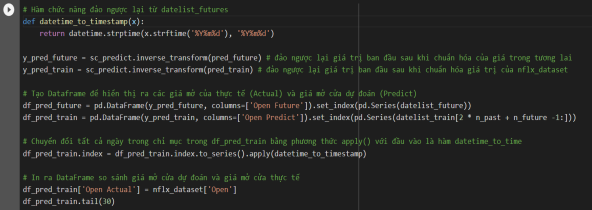


Bước 3 : Dự đoán giá mở cửa của tập huấn luyện và dự đoán giá mở cửa các ngày trong tương lai ( 28 ngày ) từ ngày 14 / 12 / 2022 đến 09 / 01 / 2023 với ngày cuối cùng của tập huấn luyện là từ ngày 13 / 12 / 2022.

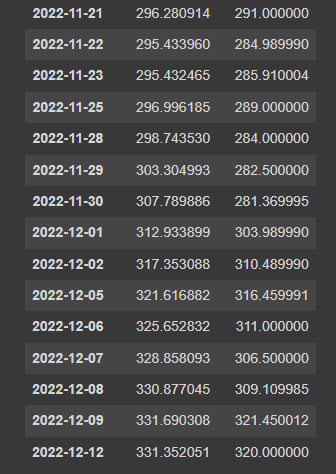
- Tạo ra một danh sách các ngày trong tương 28 ngày tiếp theo kể từ ngày cuối trong tập huấn luyện và định hình lại dữ liệu Datetime. Kế đến là dự đoán giá mở cửa của các ngày trong tập huấn luyện và dự đoán giá mở cửa của 28 ngày trong tương lai .



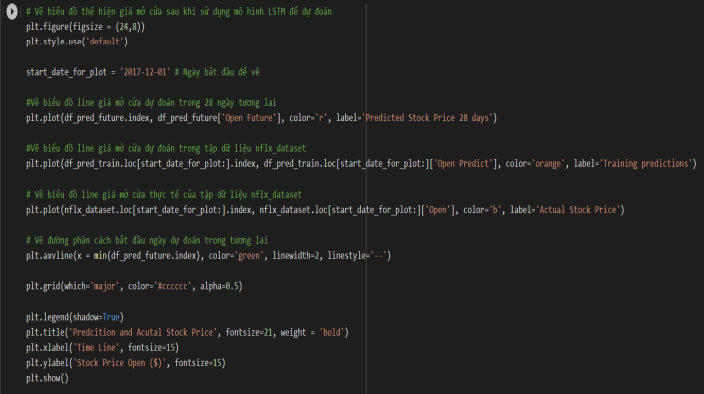
- Sau khi dự đoán thì ta cần đảo ngược lại giá mở cửa ban đầu của tập huấn luyện và giá mở cửa của 28 ngày trong tương lai sau khi ta sử dụng Sklearn để bình thường hóa dữ liệu ( Normalize Data ). Thực hiện đưa danh sách dữ liệu thành DataFrame để so sánh chênh lệch giữa giá mở cửa thực tế và giá mở cửa dự đoán của tập huấn luyện từ ngày 01 / 11 / 2022 đến ngày 13 / 12 / 2022.



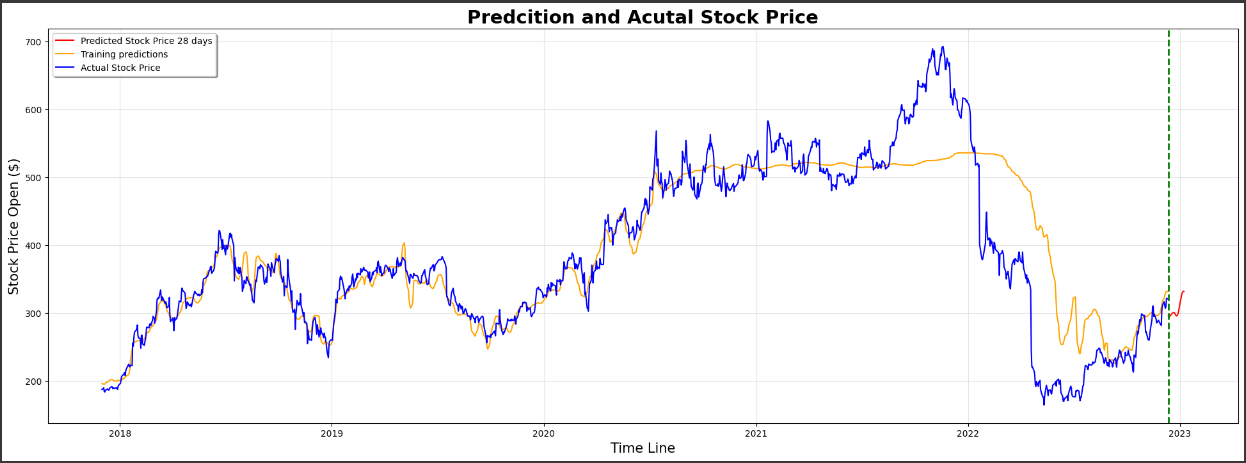
- In ra DataFrame giá mở cửa dự đoán của 28 ngày trong tương lai.



- Trực quan hóa dữ liệu giá mở cửa dựa đoán ( đường vẽ màu Vàng ) và giá mở cửa tương lai ( đường vẽ màu Đỏ ) đối với giá mở cửa thực tế ( đường vẽ màu Xanh ). Với ngày bắt đầu trực quan của biểu đồ là từ 01 / 12 / 2017 cho 09 / 01 / 2023.



- Cuối cùng kết quả trực quan hóa với việc dự đoán giá mở cửa đối với mô hình LSTM.



*Hình 10 : Biểu đồ dự đoán giá mở cửa của Netflix từ năm 2017 - 2023.*

KẾT LUẬN

Qua quá trình phân tích dữ liệu và xử lý chuyển hóa dữ liệu cho phù hợp để đưa vào mô hình LSTM thì kết quả đạt được có thể nói là không tốt như mong đợi nhưng ta có thể nhìn vào biểu đồ trực quan ( Hình 10 ) . Và có thể thấy và nhận xét rằng mô hình LSTM đã dự đoán rất tốt giá mở cửa qua mỗi ngày mỗi tháng và mỗi năm dựa theo những ngày tháng năm trước đó. Từng thời điểm thì ta có thể thấy sự tăng giảm tương đồng giữa giá mở cửa thực tế và giá mở cửa dự đoán của mô hình. Nhìn từ năm 2020 đến 2022 thì giá chứng khoán của Netflix tăng cao hơn dự kiến mà mô hình không dự đoán được là vì trong 2 năm đó là những năm đại dịch COVID – 19 bùng phát nên hầu như mọi người đều ở nhà và nhờ vào đó một công ty chuyên cung cấp các phim trực tuyến trên trang web như Netflix được nhiều người sử dụng vào những năm đại dịch đó và từ đó trong 2 năm đó giá chứng khoán của Netflix được lên rất cao nên khó có thể nào mô hình của ta đoán được nó sẽ cao đến mức như thế chỉ trong 2 năm.

HẠN CHẾ VÀ HƯỚNG PHÁT TRIỂN

1. Hạn chế

Mô hình LSTM có thể không hoạt động tốt trên dữ liệu có số giá trị bị thiếu hoặc không liên tục, vì nó được thiết kế để học từ dữ liệu thời gian liên tiếp. Nếu dữ liệu có số giá trị bị thiếu hoặc không liên tục, mô hình có thể không hiểu được mối quan hệ giữa các giá trị và không thể dự đoán chính xác giá chứng khoán trong tương lai. Ngoài ra, mô hình LSTM cũng có thể không hoạt động tốt nếu dữ liệu không đủ lớn hoặc không đạt chất lượng tốt. Để huấn luyện mô hình LSTM hiệu quả, cần có một lượng dữ liệu lớn và đạt chất lượng tốt, bao gồm cả dữ liệu quan trọng và không quan trọng. Nếu không có đủ dữ liệu hoặc dữ liệu không đạt chất lượng tốt, mô hình có thể không hoạt động tốt và không thể dự đoán chính xác giá chứng khoán trong tương lai.

2. Hướng phát triển

Mô hình dự đoán giá chứng khoán LSTM đang trở thành một trong những công cụ phân tích thị trường chứng khoán phổ biến nhất hiện nay.Những nghiên cứu và phát triển mô hình LSTM đang tìm cách cải thiện hiệu quả của mô hình và giảm thiểu các hạn chế của nó. Một số có thể kể đến như sau :

1. Sử dụng mô hình khác cùng với LSTM : có thể sử dụng mô hình LSTM cùng với các mô hình khác có thể tăng hiệu quả dự đoán giá chứng khoán.

2. Sử dụng dữ liệu đa dạng hơn : để cải thiện hiệu quả của mô hình LSTM, có thể sử dụng dữ liệu đa dạng hơn để huấn luyện mô hình.

3. Sử dụng các kỹ thuật xử lý dữ liệu khác : để tăng độ hiệu quả của mô hình LSTM, có thể sử dụng các kỹ thuật xử lý dữ liệu khác như chuẩn hóa dữ liệu, phân tích biểu đồ ( chart analysis ) hay giảm chiều của dữ liệu.

4. Sử dụng phương pháp ước lượng khác : mô hình LSTM có thể không hoạt động tốt trong một số tình huống, vì vậy có thể sử dụng các phương pháp ước lượng khác như phương pháp ước lượng theo quy luật ( rule-based estimation ) hoặc phương pháp ước lượng cơ bản ( basic estimaton ) để tăng hiệu quả dự đoán giá chứng khoán.

5. Sử dụng các phương pháp huấn luyện khác : để cải thiện độ hiệu quả mô hình LSTM, có thể sử dụng các phương pháp huấn luyện khác như phương pháp huấn luyện theo từng lớp ( layer-wise training ) hoặc phương pháp huấn luyện theo từng điểm dữ liệu ( instance-wise training ).

6. Sử dụng các kỹ thuật tăng cường mô hình : để tăng khả năng dự đoán chính xác của mô hình LSTM, có thể sử dụng kỹ thuật tăng cường mô hình như phân tích đa chiều ( multivariate analysis ), sử dụng bộ lọc ( filter ) hoặc sử dụng các phương pháp điều chỉnh trọng số ( weight adjustment ).

SOURCE CODE DỰ ÁN

https://colab.research.google.com/drive/1CY\_JFtq2AV5DeaDEETNJUZzbhUVcOqbv?usp=sharing27