**TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TIỂU LUẬN KHAI THÁC DỮ LIỆU VÀ ỨNG DỤNG**

**Tên đề tài: Dự đoán biến động tỷ giá hối đoái giữa VND và USD**

**Giảng viên hướng dẫn : VÕ THỊ HỒNG THẮM**

**Sinh viên thực hiện : TRẦN ĐỨC ĐẠI**

**MSSV : 2200003458**

**Khoá : 2022**

**Ngành/ chuyên ngành : KHOA HỌC DỮ LIỆU**

TP. HCM, tháng 05 năm 2025

**TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TIỂU LUẬN KHAI THÁC DỮ LIỆU VÀ ỨNG DỤNG**

**Tên đề tài: Dự đoán biến động tỷ giá hối đoái giữa VND và USD**

**Giảng viên hướng dẫn : VÕ THỊ HỒNG THẮM**

**Sinh viên thực hiện : TRẦN ĐỨC ĐẠI**

**MSSV : 2200003458**

**Khoá : 2022**

**Ngành/ chuyên ngành : KHOA HỌC DỮ LIỆU**

TP. HCM, tháng 05 năm 2025

# LỜI CẢM ƠN

Em xin trân trọng gửi lời cảm ơn sâu sắc đến cô Võ Thị Hồng Thắm, người đã trực tiếp hướng dẫn và chỉ bảo em trong suốt quá trình thực hiện tiểu luận đồ án môn học. Những kiến thức chuyên sâu, tận tâm của cô trong việc hướng dẫn em, từ việc chọn đề tài cho đến việc hoàn thiện từng phần của đồ án, đã giúp em rất nhiều trong việc nâng cao khả năng nghiên cứu và ứng dụng lý thuyết vào thực tế.

Bằng sự nhiệt tình và trách nhiệm, cô đã không chỉ cung cấp những lời khuyên quý báu mà còn truyền cảm hứng cho em trong công việc hoàn thiện đồ án. Quá trình tinh chế của cô đã giúp em nhận được những điểm cần cải thiện và hoàn thiện.

Em cũng xin chân thành cảm ơn các bạn trong lớp đã cùng em đồng hành, hỗ trợ và chia sẻ kinh nghiệm trong quá trình thực hiện dự án. Sự học hỏi lẫn nhau, lời khuyên từ bạn bè đã tạo nên một môi trường học tập tuyệt vời

Cuối cùng, em xin cảm ơn các thầy cô trong khoa Công Nghệ Thông Tin đại học Nguyễn Tất Thành đã tạo điều kiện thuận lợi và cung cấp những tài liệu nghiên cứu quý giá, giúp em hoàn thiện bản đồ này. Nhờ sự hỗ trợ từ thầy cô và bạn bè, em đã có thể vượt qua những thử thách đó.

Em xin trân trọng cảm ơn!

# LỜI MỞ ĐẦU

Trong bối cảnh nền kinh tế toàn cầu phát triển mạnh mẽ, biến động tỷ giá giữa đồng Việt Nam (VND) và đồng đô la Mỹ (USD) ảnh hưởng lớn đến các quyết định kinh tế, tài chính và đầu tư. Tỷ giá không chỉ tác động đến xuất nhập khẩu mà còn ảnh hưởng đến chính sách tiền tệ, đầu tư nước ngoài và ổn định nền kinh tế quốc gia. Vì vậy, dự đoán chính xác biến động tỷ giá là yếu tố quan trọng trong việc quản lý rủi ro và tối ưu hóa quyết định tài chính.

Dự án "Dự đoán tỷ lệ biến động tỷ giá giữa VND và USD" nhằm phát triển mô hình dự đoán chính xác tỷ giá sử dụng phương pháp phân tích dữ liệu và học máy. Mô hình này giúp các tổ chức tài chính, doanh nghiệp và nhà đầu tư đưa ra chiến lược tài chính phù hợp, giảm thiểu rủi ro và tối ưu hóa lợi nhuận.

Thông qua ứng dụng các kỹ thuật phân tích dữ liệu tiên tiến, dự án cung cấp công cụ giảm thiểu rủi ro cho các tổ chức tài chính và cá nhân tham gia vào thị trường ngoại hối. Dự báo chính xác tỷ giá giúp doanh nghiệp điều chỉnh chiến lược tài chính, tối ưu hóa đầu tư và bảo vệ tài sản trong bối cảnh biến động thị trường.

Dự án sẽ sử dụng các yếu tố kinh tế vĩ mô và các chỉ số tài chính để dự đoán tỷ giá VND/USD, hỗ trợ các quyết định tài chính quan trọng. Các công cụ dự đoán này có thể tích hợp vào hệ thống quản lý tài chính của tổ chức, ngân hàng, doanh nghiệp xuất khẩu, giúp họ xây dựng chiến lược phòng ngừa rủi ro hiệu quả. Đồng thời, cá nhân và nhà đầu tư có thể sử dụng thông tin dự báo này để đưa ra quyết định mua bán ngoại tệ và đầu tư tài sản, từ đó giảm thiểu rủi ro và tối đa hóa lợi nhuận.

|  |  |
| --- | --- |
| TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH  **TRUNG TÂM KHẢO THÍ** | **KỲ THI KẾT THÚC HỌC PHẦN**  **HỌC KỲ II NĂM HỌC 2024 – 2025** |

**PHIẾU CHẤM THI TIỂU LUẬN/ĐỒ ÁN**

BM-ChT-11

Môn thi: Khai thác dữ liệu và ứng dụng Lớp học phần:22DTH4B

Nhóm sinh viên thực hiện :

1.Hồ Quốc Thái Tham gia đóng góp:33%

2.Nguyễn Duy Thái Tham gia đóng góp:33%

3. Trần Đức Đại Tham gia đóng góp:33%

4. Tham gia đóng góp:

5. Tham gia đóng góp:

6. Tham gia đóng góp:

7. Tham gia đóng góp:

8. Tham gia đóng góp:

Ngày thi:09/05/2025 Phòng thi:

Đề tài tiểu luận/báo cáo của sinh viên :Dự đoán biến động tỷ giá đối hoái giữa VND và USĐ

Phần đánh giá của giảng viên (căn cứ trên thang rubrics của môn học):

|  |  |  |  |
| --- | --- | --- | --- |
| **Tiêu chí (theo CĐR HP)** | **Đánh giá của GV** | **Điểm tối đa** | **Điểm đạt được** |
| Cấu trúc của báo cáo |  | 1 |  |
| Nội dung |  |  |  |
| * Các nội dung thành phần |  | 6 |  |
| * Lập luận |  | 2 |  |
| * Kết luận |  | 0.5 |  |
| Trình bày |  | 0.5 |  |
| **TỔNG ĐIỂM** |  | **10** |  |

|  |  |
| --- | --- |
|  | **Giảng viên chấm thi**  *(ký, ghi rõ họ tên)*  **TS.Võ Thị Hồng Thắm** |

# 

# MỤC LỤC

LỜI CẢM ƠN ii

LỜI MỞ ĐẦU iii

MỤC LỤC v

DANH MỤC HÌNH vi

CHƯƠNG 1 1

Cơ sở lý thuyết 1

I. Giới thiệu về tỷ giá hối đoái và các yếu tố tác động 1

II. Máy học trong dự báo tỷ giá hối đoái 3

III. LSTM mô hình Deep Learning 4

IV. MinMaxScaler và các kỹ thuật tiền xử lý dữ liệu 6

CHƯƠNG 2 10

Phân tích yêu cầu 10

I. Mục tiêu của dự án 10

II. Yêu cầu dữ liệu 12

III. Yêu cầu kỹ thuật 13

IV. Đối tượng và phạm vi ứng dụng 16

CHƯƠNG 3 18

Xây dựng mô hình 18

I. Tiền xử lý dữ liệu 18

II. Xây dựng mô hình LSTM 20

III. Tạo tập huấn luyện và kiểm tra 23

CHƯƠNG 4 24

Thực nghiệm mô hình 24

I. Dự báo tỷ giá hối đoái 24

II. Đánh giá mô hình 25

III. Dự báo tương lai 25

IV. Nhận xét về độ ứng dụng thực tế 26

CHƯƠNG 5 27

Kết luận và hướng phát triển 27

I. Kết quả đạt được 27

II. Hạn chế của đề tài 28

III. Hướng phát triển 28

TÀI LIỆU THAM KHẢO 30

# DANH MỤC HÌNH

Hình 1.1 Lịch sử tỷ giá trọng số GDP của một giỏ 6 đồng tiền quan trọng (Đô la Mỹ, Euro, Yên Nhật, Nhân dân tệ Trung Quốc, Franc Thụy Sỹ, Bảng Anh). 1

Hình 2.1 Sơ đồ xây dựng mô hình học có giám sát dự báo xu hướng tỷ giá ngoại hối. 4

Hình 3.1 Mô hình LSTM. 5

Hình 4.1 Kỹ thuật Data transformation & Feature Engineering với Python. 6

Hình 5.1 Tổng hợp về chuẩn hóa dữ liệu 9

Hình 6.2 Dự báo biến động tỷ giá giữa đồng VND và USD. 10

Hình 7.2 Mô hình học máy LSTM. 11

Hình 8.2 Logo Google Colab 13

Hình 9.2 Logo Visual Studio Code. 14

Hình 10.2 Logo PyCharm. 14

Hình 11.3 Code minh họa hàm pd.to\_datetime() 18

Hình 12.3 Code minh họa sử dụng MinMaxScaler. 19

Hình 13.3 Tạo chuỗi thời gian cho mô hình LSTM. 19

Hình 14.3 Biểu đồ so sánh giữa tỷ giá thực tế với tỷ giá dự đoán. 20

Hình 15.3 Code xây dựng mô hình LSTM. 20

Hình 16.3 Optimizer cho mô hình. 21

Hình 17.3 Huấn luyện model với các thông số. 21

Hình 18.3 Xây dựng mô hình LSTM trong Keras. 21

Hình 19.3 Code đánh giá mô hình LSTM. 22

Hình 20.3 Kết quả đánh giá mô hình. 22

Hình 21.3 Phân chia dữ liệu huấn luyện và kiểm tra. 23

Hình 22.4 Kết quả dự đoán. 24

Hình 23.4 Công thức tính MAE và MSE. 25

Hình 24.4 Phát triển mô hình LSTM 27

Hình 25.4 Overfitting trong Machine Learning. 28

Hình 26.4 Mô hình vĩ mô. 29

## CHƯƠNG 1

## Cơ sở lý thuyết

### Giới thiệu về tỷ giá hối đoái và các yếu tố tác động

#### Tỷ giá đối hoái, tỷ giá đối hoái giữa USD và VND

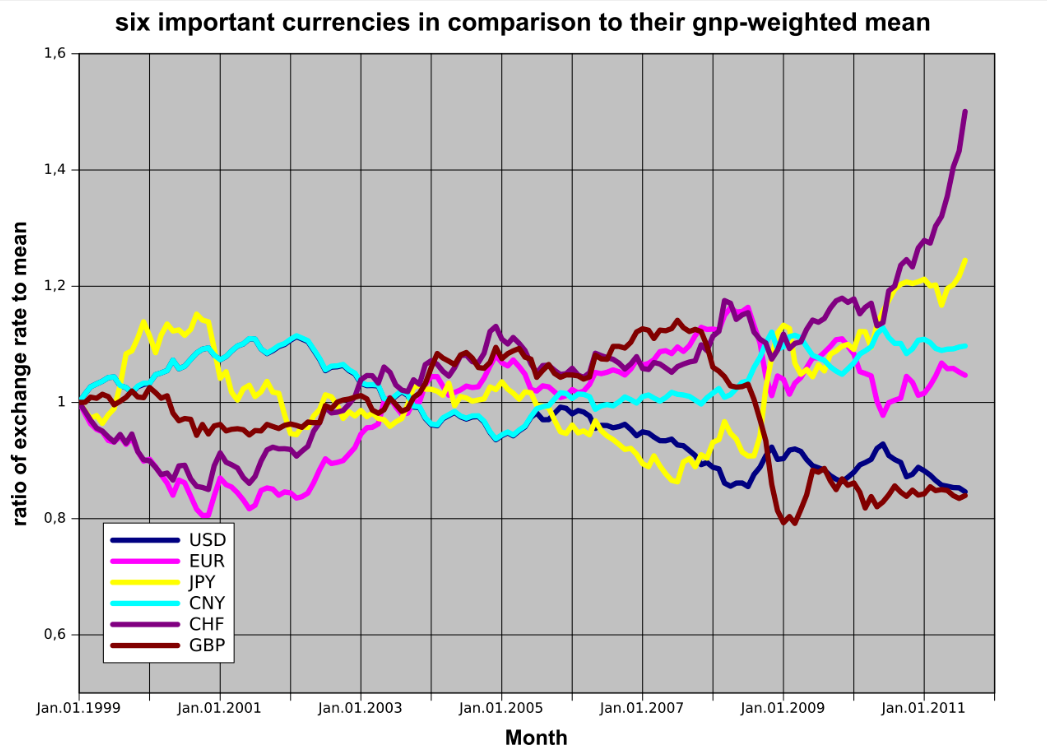
##### Tỷ giá đối hoái

Tỷ giá hối đoái (còn được gọi là tỷ giá trao đổi ngoại tệ, tỷ giá Forex, tỷ giá FX hoặc Agio) giữa hai tiền tệ là tỷ giá mà tại đó một đồng tiền này sẽ được trao đổi cho một đồng tiền khác. Nó cũng được coi là giá cả đồng tiền của một quốc gia được biểu hiện bởi một tiền tệ khác.

Tiền tệ thường là đơn vị tiền của mỗi quốc gia, nhưng cũng có thể là đơn vị tiền của các khu vực cấp dưới như Hồng Kông, hoặc là đơn vị tiền siêu quốc gia như euro.

Tỷ giá hối đoái còn biểu thị giá trị của tiền một nước so với tiền một nước khác. Ví dụ, tỷ giá hối đoái ngân hàng trung ương giữa 131 yên Nhật và Đô la Mỹ có nghĩa là ¥131 sẽ trao đổi lấy 1 đô la Mỹ, hoặc 1 đô la Mỹ sẽ trao đổi lấy ¥131. Trong ví dụ này, giá trị của một đô la so với yên là ¥131, hoặc tương đương với việc giá trị của một yên so với đô la là $1/131.

Mỗi quốc gia quyết định chế độ tỷ giá hối đoái áp dụng cho đơn vị tiền của mình. Ví dụ, tiền tệ có thể có chế độ tỷ giá nổi, tỷ giá cố định hoặc tỷ giá hỗn hợp. Chính phủ có thể áp đặt giới hạn và kiểm soát tỷ giá hối đoái. Ngoài ra, mỗi quốc gia có thể có đồng tiền mạnh hoặc yếu. Trong văn học kinh tế, không có sự đồng thuận về chính sách tỷ giá hối đoái quốc gia tối ưu (không giống như trong vấn đề thương mại, nơi thương mại tự do được coi là tối ưu). Thay vào đó, chế độ tỷ giá hối đoái quốc gia phản ánh yếu tố chính trị.



Hình 1.1 Lịch sử tỷ giá trọng số GDP của một giỏ 6 đồng tiền quan trọng (Đô la Mỹ, Euro, Yên Nhật, Nhân dân tệ Trung Quốc, Franc Thụy Sỹ, Bảng Anh).

##### Tỷ giá hối đoái giữa USD và VND

Tỷ giá hối đoái giữa đồng đô la Mỹ (USD) và đồng Việt Nam (VND) là tỷ lệ trao đổi giữa hai loại tiền tệ này, phản ánh giá trị của đồng VND trong quan hệ với đồng USD. Tỷ giá hối đoái có ảnh hưởng sâu rộng đến nền kinh tế quốc gia, đặc biệt trong các lĩnh vực như thương mại quốc tế, đầu tư nước ngoài, và chính sách tiền tệ.

Khái niệm về tỷ giá hối đoái USD/VND

Tỷ giá hối đoái USD/VND là tỷ lệ quy đổi giữa đồng đô la Mỹ và đồng Việt Nam. Ví dụ, tỷ giá hối đoái USD/VND ~ 25,000 có nghĩa là 1 đô la Mỹ có giá trị sấp sĩ là 25,000 đồng Việt Nam. Tỷ giá này thay đổi liên tục tùy thuộc vào các yếu tố kinh tế vĩ mô, các quyết định chính sách của chính phủ và ngân hàng trung ương, và biến động của thị trường tài chính quốc tế.

#### Các yếu tố kinh tế ảnh hưởng đến tỷ giá hối đoái

Thị trường mới nổi: Nghiên cứu về khu vực mục tiêu chủ yếu tập trung vào lợi ích của sự ổn định tỷ giá hối đoái đối với các nước công nghiệp, nhưng một số nghiên cứu cho rằng sự biến đổi không ổn định của tỷ giá hối đoái hai chiều giữa các nước công nghiệp một phần gây ra khủng hoảng tài chính ở các thị trường mới nổi. Theo quan điểm này, khả năng cạnh tranh của nền kinh tế thị trường mới nổi bị suy yếu vì nhiều đồng tiền được gắn liền với đô la Mỹ theo nhiều cách khác nhau, ngầm hoặc rõ ràng, do đó, biến đổi như việc đô la Mỹ tăng giá trị so với yen hoặc đồng Mark đã góp phần vào những sóng sốc gây mất ổn định. Hầu hết các quốc gia này là người nợ vốn ngoại, nợ của họ được định giá bằng một trong các đồng tiền G3.

Vào tháng 9 năm 2019, Argentina hạn chế khả năng mua đô la Mỹ . Mauricio Macri năm 2015 đã cam kết sẽ dỡ bỏ hạn chế do chính phủ cánh tả đặt ra, bao gồm kiểm soát vốn được sử dụng ở Argentina để ổn định kinh tế. Khi lạm phát tăng trên 20%, việc giao dịch được định giá bằng đô la Mỹ trở nên phổ biến khi người dân Argentina chuyển sang sử dụng đồng peso. Năm 2011, chính phủ Cristina Fernández de Kirchner hạn chế mua đô la, dẫn đến việc tăng mua đô la trên thị trường đen. Kiểm soát được giảm sau khi Macri nhậm chức và Argentina phát hành trái phiếu định giá bằng đô la Mỹ, nhưng khi nhiều yếu tố dẫn đến sự suy giảm giá trị của peso so với đô la, kiểm soát vốn đã được tái áp dụng để ngăn chặn sự suy giảm thêm trong bối cảnh sự bán ra đồng peso.

**Các yếu tố tác động đến biến đổi tỷ giá hối đoái**

* **Dịch vụ thanh toán:** Khi một quốc gia có thâm hụt thanh toán hoặc thương mại lớn, đồng tiền của nó suy giảm do nhu cầu về ngoại tệ vượt quá cung cấp.
* **Lãi suất:** Khi lãi suất tăng hoặc lãi suất nội địa cao hơn so với lãi suất ngoại hối, đồng tiền trong nước tăng giá trị do dòng vốn đổ vào.
* **Lạm phát:** Lạm phát cao làm giảm giá trị đồng tiền trong nước và tăng giá trị ngoại tệ. Quốc gia có lạm phát thấp thường có đồng tiền mạnh hơn.
* **Chính sách tài khóa và tiền tệ:** Chính sách tài khóa và tiền tệ lỏng lẻo, thâm hụt tài khóa gây suy giảm đồng tiền, còn chính sách thắt chặt làm tăng giá trị đồng tiền.
* **Phân tích:** Dự đoán tiền tệ sẽ tăng giá trị khiến nhà đầu tư mua nhiều, ngược lại khiến giá giảm. Phân tích ảnh hưởng đến biến động ngắn hạn của tỷ giá hối đoái.
* **Can thiệp thị trường của chính phủ:** Khi biến động tỷ giá hối đoái ảnh hưởng xấu đến nền kinh tế, thương mại của quốc gia hoặc chính phủ cần đạt được mục tiêu chính sách, cơ quan tiền tệ có thể tham gia giao dịch tiền tệ để thay đổi tỷ giá hối đoái.
* **Sức mạnh kinh tế của quốc gia:** Tăng trưởng kinh tế cao không tốt cho đồng tiền trong ngắn hạn, nhưng trong dài hạn, nó ủng hộ mạnh mẽ sức mạnh của đồng tiền quốc gia trên thị trường ngoại hối.

### Máy học trong dự báo tỷ giá hối đoái

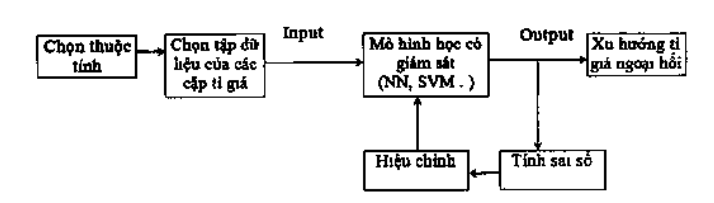
#### Học máy trong bài toán giao dịch ngoại hối

Mô hình học máy được xây dựng cho việc tìm kiếm các tri thức trong tập dữ liệu mà không cần sự tác động của con người. Hệ thống tư huấn luyện để sinh ra các tập luật, các chiến lược quyết định dựa trên dứ liêu đầu vào và đầu ra. Mô hình này có thể được xây dụng như là một nhà đầu tư thực sự trên thị trường giao dịch với đầy đủ các tri thức và kĩ năng để phân tích thị trường. Với những ưu điểm của mô hình mạng trí tuệ nhân tạo trong việc phân tích dữ liệu như: một khi mạng được huấn luyện thì hiệu quả của việc phân lớp không giảm nhiều nếu tập dữ liệu không đủ lớn. Nó không đòi hỏi người sử dụng nhập vào mức độ quan trọng của các thuộc tính, vì chính nó sẽ làm điều đó khi học. Hơn nữa, nó có thể điều khiến trong trường hợp chất lương dữ liệu kém (Carney và cộng sự, 1996). Trong khuôn khổ phạm vi bài viết này, một số mô hình học có giám sát sẽ được trình bày.

Piche (1995) sử dụng hình ảnh biểu đồ plot để mô tá xu hướng dịch chuyến các dao động của tỷ giá với ma trận xu hướng. Bằng cách thiết lập các tham số khác nhau của tỷ giá ngoại hối, các kết quả sẽ phản ánh các góc độ khác nhau của thị trường. Việc dự báo tỷ giá sử dụng mô hình dự báo chuỗi thời gian được biết đến với việc sử dụng phổ biến mô hình Regressive Integrated Moving Average (ARIMA) (Box và Jenkin, 1970). Tuy nhiên, ARIMA được phát triển dựa trên giả thiết rằng chuỗi thời gian dự báo là tuyến tính và không biến động. Thời gian gần đây, mô hình dự báo chứng khoán và ngoại hối được sử dụng phổ biến là GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) và các nghiên cứu chỉ ra rằng nó tốt hơn việc sử dung mô hình ARIMA (Wei, 2009, Kumar, 2006, Yu, 2002). Ngoài các mô hình mạng noron và SVM cũng được biết đến như là các mô hình được sử dụng để dự báo tỷ giá ngoại hối. Chan and Teong (1995) sử dụng một mang nơron đơn giản để cải tiến các kĩ thuật phân tích thường dùng. Mô hình này cho phép vào lệnh và thoát lệnh trước khi nhà đầu tư khác làm. Refenes và công sự (1992) chỉ ra rằng, bằng cách sử dụng mô hình mạng nơron với thuật toán lan truyền ngược với sự phản hồi theo giờ, hiệu quả của các giao dịch ngắn hạn có thể được cải thiện. Họ còn chỉ ra rằng dự báo nhiều bước thì mang lại hiệu quả hơn dự báo một bước và việc lựa trọn tập huấn luyện là rất quan trọng. Ở nghiên cứu của Kim (2003), mô hình SVM được sử dụng để dự báo cho giá chỉ số chứng khoán thông qua việc đề xuất tham số tối ưu của biên trên và hàm nhân Kernel. Nghiên cứu so sánh các kết quả này với các mô hình mạng noron, và chỉ ra kết quả tốt hơn.

#### Mô tả bài toán ngoại hối

Tỷ giá ngoại hối bị ảnh hưởng bởi một số các yếu tố như các sự kiện kinh tế hoặc chính trị, hay tâm lý của nhà đầu tư. Chính các điều này làm cho thị trường giao dịch không ổn định và phức tạp. Nhà đầu tư tham gia vào thị trưởng ngoại hối luôn tìm mọi cách để giải thích cho sự biến động của tỷ giá, và họ hi vọng việc dự báo của mình là chính xác. Dự báo chính xác tỷ giá và đúng thời điểm cho ra lợi nhuận tối đa. Tuy nhiên nếu giao dịch dựa trên việc tính toán sai dẫn đến nguy cơ mất là rất lớn, Bài toán giao dịch trên thị trường ngoại hối là bài toán dự báo xu hướng theo ngày của tỷ giá quy đổi của các cặp tiền tệ theo chuỗi thời gian biến động phức tạp.



Hình 2.1 Sơ đồ xây dựng mô hình học có giám sát dự báo xu hướng tỷ giá ngoại hối.

Đầu vào của bài toán là lịch sử tỷ giá trao đổi của các cặp tiền tệ, số liệu về các sự kiện kinh tế đã xảy ra.... Đầu ra của bài toán chính là xu hướng tăng hay giảm của các cặp tiền tệ cần xét. Một số lương lớn các thuộc tính từ dữ liệu chuỗi thời gian được sinh ra từ lịch sử tỷ giá các cặp tiền tệ nhằm mục đích chọn ra các tập dữ liệu phù hợp cho bài toán phân lớp.

### LSTM mô hình Deep Learning

#### Giới thiệu về Deep Learning

Deep learning là một nhánh của lĩnh vực học máy, dựa trên một tập hợp các thuật toán nhằm cố gắng mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý với cấu trúc phức tạp, hoặc bao gồm nhiều biến đổi phi tuyến.

Deep learning là một lớp của các thuật toán máy học:

* Sử dụng một tầng (cascade) nhiều lớp các đơn vị xử lý phi tuyến để trích trọn đặc trưng và chuyển đổi. Mỗi lớp kế tiếp dùng đầu ra từ lớp trước làm đầu vào. Thuật toán này có thể được giám sát hoặc không cần giám sát và được ứng dụng cho các mô hình phân tích (không có giám sát) và phân loại (giám sát).
* Dựa trên học (không có giám sát) của nhiều cấp các đặc trưng hoặc đại diện của dữ liệu. Các tính năng cao cấp bắt nguồn từ các tính năng thấp cấp hơn để tạo thành một đại diện thứ bậc.
* Là một phần của lĩnh vực máy học và rộng lớn hơn về việc học đại diện dữ liệu.
* Học nhiều cấp độ đại diện tương ứng với các mức độ trừu tượng khác nhau; các mức độ hình thành một hệ thống phân cấp của các khái niệm.

Deep learning còn là phương pháp nâng cao của mạng nơ-ron nhân tạo (Artificial Neural Networks) khai thác khả năng tính toán ngày càng rẻ từ các chip xử lý hiện đại. Phương pháp này nhắm tới việc xây dựng nhiều hơn các mạng nơ-ron phức tạp cũng như giải quyết bài toán semi-supervised do tập dữ liệu khổng lồ thường được gán nhãn không đầy đủ.

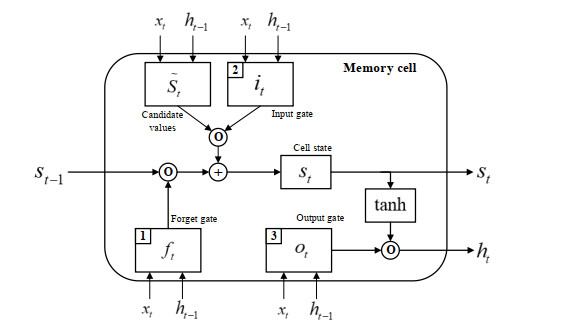
Deep learning có nhiều thuật toán khác nhau nên nó phát triển rất nhanh nhiều biến thể và thuật toán mới ra đời liên tục trong thời gian ngắn, có thể kể đến một số thuật toán như: DNN - Deep Nơ-ron Network, Deep Belief Network, RNN, LSTM,...

#### Mô hình LSTM

LSTM là một phiên bản mở rộng của mạng Recurrent Neural Network (RNN), nó được thiết kế để giải quyết các bài toán về phụ thuộc xa (long-term dependencies). RNN là mạng nơ-ron có chứa vòng lặp. Mạng này có khả năng lưu trữ thông tin, thông tin được truyền từ lớp này sang lớp khác. Đầu ra của lớp ẩn phụ thuộc vào thông tin của các lớp tại mọi thời điểm.

RNN đã được sử dụng phổ biến trong xử lý ngôn ngữ tự nhiên hay các bài toán có dữ liệu tuần tự. Tuy nhiên, do kiến trúc của RNN khá đơn giản nên khả năng lien kết các lớp có khoảng cách xa là không tốt. Nó cơ bản không có khả năng ghi nhớ thông tin từ các dữ liệu có khoảng cách xa, và do đó, những phần tử đầu tiên trong chuỗi đầu vào thường không có nhiều ảnh hưởng đến kết quả dự đoán phần tử cho chuỗi đầu ra các bước sau.

Nguyên nhân của việc này là do RNN chịu ảnh hưởng bởi việc đạo hàm bị thấp dần trong quá trình học – biến mất đạo hàm (vanishing gradient). Mạng LSTM được thiết kế để khắc phục vấn đề này. Cơ chế hoạt động của LSTM là chỉ ghi nhớ những thông tin liên quan, quan trọng cho việc dự đoán, còn các thông tin khác sẽ được bỏ đi.



Hình 3.1 Mô hình LSTM.

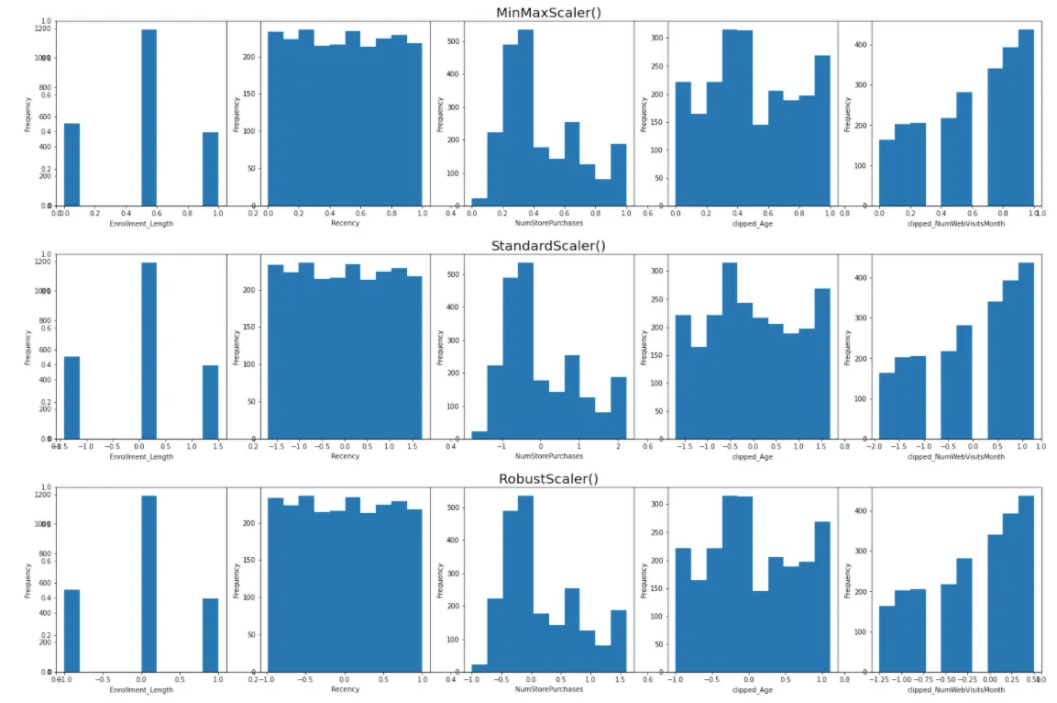
### MinMaxScaler và các kỹ thuật tiền xử lý dữ liệu

#### Giới thiệu về MinMaxScaler và các kỹ thuật chuẩn hóa dữ liệu

Trong các bài toán học máy, việc chuẩn hóa dữ liệu là một bước rất quan trọng để đảm bảo mô hình có thể học và dự đoán chính xác. Dữ liệu không chuẩn hóa có thể khiến các thuật toán học máy gặp khó khăn trong việc tối ưu hóa và học các đặc trưng (features). Một trong những phương pháp chuẩn hóa phổ biến là MinMaxScaler, cùng với các kỹ thuật chuẩn hóa khác, giúp cải thiện hiệu suất của mô hình học máy.

##### MinMaxScaler

MinMaxScaler là một phương pháp chuẩn hóa dữ liệu phổ biến, được sử dụng để chuyển đổi các giá trị của dữ liệu vào trong một phạm vi xác định, thường là từ 0 đến 1. Phương pháp này sử dụng công thức để chuẩn hóa mỗi giá trị dữ liệu.



Hình 4.1 Kỹ thuật Data transformation & Feature Engineering với Python.

***Ưu điểm của MinMaxScaler***:

- Phạm vi đồng nhất: Đảm bảo tất cả các đặc trưng của dữ liệu đều nằm trong một phạm vi cụ thể, giúp các thuật toán học máy, đặc biệt là các mô hình như mạng nơ-ron (neural networks), hoạt động hiệu quả hơn.

- Nhanh chóng và dễ sử dụng: Quá trình chuẩn hóa diễn ra nhanh chóng và dễ dàng cài đặt trong các thư viện Python như `sklearn`.

***Nhược điểm của MinMaxScaler:***

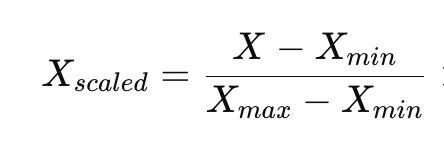
- Ảnh hưởng của ngoại lai: MinMaxScaler rất nhạy cảm với các giá trị ngoại lai (outliers). Nếu dữ liệu có các giá trị quá cao hoặc quá thấp, phạm vi dữ liệu sau chuẩn hóa có thể bị bóp méo, dẫn đến giảm chất lượng dự đoán của mô hình.

##### Các kỹ thuật chuẩn hóa dữ liệu

Khi làm việc với dữ liệu, việc chuẩn hóa giúp cải thiện hiệu quả và độ chính xác của các mô hình học máy. Có nhiều kỹ thuật chuẩn hóa dữ liệu khác nhau, tùy thuộc vào đặc điểm của dữ liệu và mục tiêu của bài toán. Dưới đây là một số kỹ thuật chuẩn hóa phổ biến:

* ***Chuẩn hóa Min-Max (Min-Max Scaling)***

- Công thức:

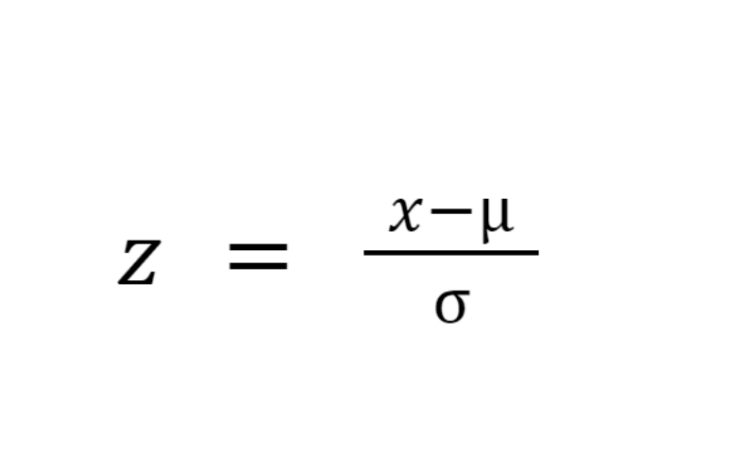


- Mô tả: Kỹ thuật này chuyển đổi dữ liệu sao cho giá trị của mỗi tính năng nằm trong khoảng [0, 1] hoặc bất kỳ khoảng nào bạn chỉ định.

- Ứng dụng: Thích hợp khi bạn muốn giữ tỷ lệ giữa các giá trị hoặc khi các thuật toán yêu cầu dữ liệu phải nằm trong một phạm vi cụ thể (ví dụ, mạng nơ-ron).

* ***Chuẩn hóa Z-score (Z-score Normalization)***

- Công thức:



- Trong đó: µ là giá trị trung bình và sigma là độ lệch chuẩn.

- Mô tả: Dữ liệu sẽ được chuyển đổi sao cho giá trị trung bình bằng 0 và độ lệch chuẩn bằng 1. Kỹ thuật này giúp dữ liệu có phân phối chuẩn.

- Ứng dụng: Thường dùng khi dữ liệu có phân phối không đều hoặc khi sử dụng các thuật toán yêu cầu phân phối chuẩn (ví dụ, hồi quy tuyến tính, phân tích thành phần chính PCA).

* ***Chuẩn hóa Robust (Robust Scaler)***

- Công thức:

A mathematical equation with black text

AI-generated content may be incorrect.

- Mô tả: Kỹ thuật này sử dụng các phần tử không bị ảnh hưởng bởi các giá trị ngoại lai (outliers) như phương pháp Min-Max hoặc Z-score.

- Ứng dụng: Thích hợp khi dữ liệu có nhiều giá trị ngoại lai.

* ***Chuẩn hóa theo log (Log Transformation)***

- Mô tả: Phương pháp này áp dụng hàm logarit cho dữ liệu. Thường được sử dụng khi dữ liệu có sự phân phối lệch mạnh hoặc có sự chênh lệch lớn giữa các giá trị.

- Ứng dụng: Thích hợp cho các dữ liệu có sự phân phối không đều và giá trị chênh lệch rất lớn (ví dụ, dữ liệu tài chính).

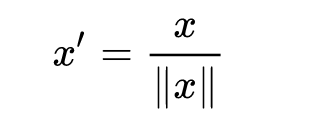
* ***Chuẩn hóa theo Root (Square Root Transformation)***

- Mô tả: Đây là một biến thể của chuẩn hóa log, nhưng thay vì sử dụng logarit, bạn sử dụng căn bậc hai của các giá trị.

- Ứng dụng: Thích hợp cho dữ liệu có phân phối không chuẩn, với các giá trị không quá lệch như log.

* ***Chuẩn hóa Unit Vector (Vector Scaling)***

- Công thức:



- Trong đó: |X| là độ dài chuẩn của vector.

- Mô tả: Kỹ thuật này chuẩn hóa dữ liệu sao cho tổng bình phương các giá trị của mỗi vector bằng 1 (chuẩn hóa vector).

- Ứng dụng: Thường được sử dụng trong các bài toán xử lý văn bản, phân tích dữ liệu, hay các bài toán liên quan đến không gian véc-tơ (ví dụ, trong học máy với các mô hình như Naive Bayes).

* ***Chuẩn hóa Quantile***

- Mô tả: Phương pháp này thay thế giá trị của mỗi điểm dữ liệu bằng phân vị tương ứng trong phân phối chuẩn.

- Ứng dụng: Sử dụng khi bạn muốn dữ liệu có phân phối đồng nhất, tức là phân phối chuẩn hoặc phân phối đồng đều.

* ***Binarization (Chuẩn hóa nhị phân)***

- Mô tả: Kỹ thuật này chuyển đổi các giá trị thành 0 hoặc 1 dựa trên ngưỡng cho trước. Các giá trị trên ngưỡng sẽ được gán là 1, và các giá trị dưới ngưỡng sẽ là 0.

- Ứng dụng: Thường được sử dụng trong các bài toán phân loại nhị phân.

Mỗi kỹ thuật chuẩn hóa có ưu điểm và nhược điểm riêng, và việc chọn phương pháp nào phụ thuộc vào dữ liệu và thuật toán sử dụng.

#### Tầm quan trọng của chuẩn hóa dữ liệu trong sử dụng mô hình máy học:

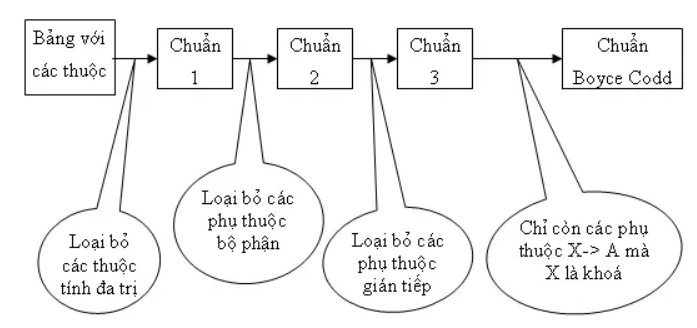
Chuẩn hóa dữ liệu là một bước quan trọng trong tiền xử lý khi xây dựng mô hình học máy, giúp đảm bảo các thuật toán hoạt động hiệu quả và chính xác. Các đặc trưng trong dữ liệu thường có phạm vi giá trị khác nhau, ví dụ như một đặc trưng có thể có giá trị từ 1 đến 1000, trong khi một đặc trưng khác chỉ có giá trị từ 0 đến 1. Khi các đặc trưng có phạm vi khác nhau, mô hình học máy có thể bị thiên lệch và không thể học được mối quan hệ giữa các đặc trưng một cách chính xác.

Chuẩn hóa giúp đưa các đặc trưng về một phạm vi giá trị đồng nhất, thường là từ 0 đến 1 hoặc có trung bình bằng 0 và độ lệch chuẩn bằng 1, giúp các mô hình học máy có thể học và tối ưu hóa các tham số hiệu quả hơn. Ngoài ra, chuẩn hóa cũng giúp tăng tốc quá trình huấn luyện mô hình, đặc biệt với các thuật toán sử dụng gradient descent như hồi quy tuyến tính hoặc mạng nơ-ron, vì nó giúp đảm bảo các gradient đều đặn và mô hình hội tụ nhanh hơn. Chuẩn hóa cũng giảm tác động của các giá trị ngoại lai, giúp mô hình học máy không bị sai lệch bởi những giá trị bất thường trong dữ liệu.

Bên cạnh đó, chuẩn hóa giúp các thuật toán học máy dựa trên khoảng cách như

K-means, KNN, hoặc SVM hoạt động hiệu quả hơn, vì nó đảm bảo rằng tất cả các đặc trưng đóng góp đồng đều vào việc tính toán khoảng cách giữa các điểm dữ liệu. Đặc biệt trong các mô hình mạng nơ-ron, chuẩn hóa dữ liệu giúp tránh hiện tượng vanishing gradients hoặc exploding gradients, giúp quá trình huấn luyện trở nên ổn định và chính xác hơn.

Tóm lại, chuẩn hóa dữ liệu là một yếu tố quan trọng giúp các mô hình học máy hoạt động hiệu quả, giảm thiểu rủi ro và cải thiện độ chính xác của kết quả dự đoán.



Hình 5.1 Tổng hợp về chuẩn hóa dữ liệu

# CHƯƠNG 2

# Phân tích yêu cầu

### Mục tiêu của dự án

#### Dự báo tỷ giá hối đoái giữa VND và USD với độ chính xác cao

Tỷ giá hối đoái giữa VND và USD là một yếu tố cực kỳ quan trọng trong nền kinh tế Việt Nam, ảnh hưởng trực tiếp đến các quyết định kinh doanh và đầu tư của các tổ chức tài chính, ngân hàng, và các doanh nghiệp xuất nhập khẩu. Biến động của tỷ giá không chỉ phụ thuộc vào các yếu tố kinh tế vĩ mô trong nước mà còn chịu tác động từ các yếu tố quốc tế như chính sách tiền tệ của Mỹ, lãi suất của các ngân hàng trung ương, tình hình chính trị toàn cầu, và các biến động tài chính quốc tế. Do đó, việc dự báo tỷ giá hối đoái không chỉ là việc dự đoán một con số đơn thuần mà còn phải hiểu và phân tích các yếu tố tác động đến nó một cách sâu sắc và toàn diện.



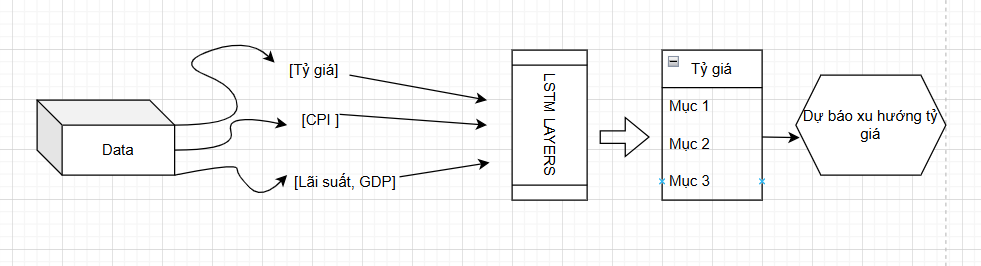
Hình 6.2 Dự báo biến động tỷ giá giữa đồng VND và USD.

Mục tiêu của dự án là xây dựng một mô hình học máy có thể dự báo chính xác các xu hướng biến động tỷ giá VND/USD trong các khoảng thời gian ngắn hạn và dài hạn, dựa trên dữ liệu lịch sử tỷ giá và các chỉ số kinh tế liên quan. Bằng cách áp dụng các kỹ thuật phân tích dữ liệu hiện đại và học máy, dự án sẽ cung cấp một công cụ dự báo không chỉ giúp các doanh nghiệp tối ưu hóa chiến lược tài chính mà còn hỗ trợ các nhà đầu tư trong việc đưa ra quyết định mua bán ngoại tệ, đầu tư tài sản và bảo vệ giá trị tài sản của mình.

#### Xây dựng mô hình học máy sử dụng LSTM để dự đoán xu hướng biến động tỷ giá trong tương lai

Để đạt được độ chính xác cao trong dự báo tỷ giá, dự án sử dụng **LSTM (Long Short-Term Memory)**, một loại mạng nơ-ron hồi tiếp sâu (recurrent neural network - RNN), đặc biệt hiệu quả trong việc xử lý và dự đoán các chuỗi dữ liệu thời gian (time series). LSTM có khả năng học và ghi nhớ các mối quan hệ dài hạn trong dữ liệu, điều này cực kỳ quan trọng trong việc dự báo tỷ giá hối đoái, bởi tỷ giá là một chuỗi dữ liệu có tính chất liên tục và bị ảnh hưởng bởi các yếu tố kinh tế vĩ mô thay đổi theo thời gian.

Các mô hình LSTM sẽ được huấn luyện trên một tập dữ liệu lớn bao gồm tỷ giá hối đoái lịch sử VND/USD và các yếu tố ảnh hưởng đến tỷ giá như chỉ số giá tiêu dùng (CPI), lãi suất, tăng trưởng GDP, và các yếu tố vĩ mô khác. Qua quá trình huấn luyện, mô hình sẽ học được cách nhận diện các mô hình biến động tỷ giá trong quá khứ và dự đoán xu hướng trong tương lai. Một điểm mạnh của LSTM là khả năng nhận diện các tác động của sự kiện bất ngờ và các xu hướng dài hạn mà không bị quên mất thông tin quan trọng trong quá khứ, điều này làm cho nó trở thành một công cụ mạnh mẽ trong việc dự báo tỷ giá hối đoái, nơi mà dữ liệu lịch sử có thể kéo dài hàng tháng hoặc thậm chí hàng năm.



Hình 7.2 Mô hình học máy LSTM.

Hơn nữa, LSTM sẽ không chỉ dự đoán tỷ giá trong một thời điểm cụ thể mà còn có thể dự đoán xu hướng biến động tỷ giá trong tương lai, giúp người dùng hiểu được mức độ rủi ro tiềm ẩn và đưa ra các quyết định chiến lược tài chính chính xác hơn. Các dự báo này sẽ có giá trị trong việc giúp các doanh nghiệp xuất nhập khẩu lập kế hoạch tài chính, đảm bảo tỷ giá ổn định cho các giao dịch quốc tế, đồng thời hỗ trợ các tổ chức tài chính trong việc phòng ngừa rủi ro tỷ giá (hedging).

#### Tạo ra giải pháp dự báo chủ động và thích ứng với biến động thị trường

Một trong những mục tiêu quan trọng của dự án là cung cấp một hệ thống dự báo chủ động, có thể linh hoạt thích ứng với các biến động thị trường và các yếu tố bất ngờ. Thị trường ngoại hối luôn có sự biến động mạnh mẽ và ảnh hưởng của các sự kiện toàn cầu có thể thay đổi nhanh chóng, do đó, một mô hình dự báo tỷ giá cần phải có khả năng điều chỉnh linh hoạt và đưa ra các dự báo kịp thời. Bằng việc ứng dụng LSTM, mô hình có thể không chỉ dự đoán tỷ giá trong tương lai gần mà còn cung cấp các dự báo dài hạn, từ đó giúp các tổ chức tài chính và doanh nghiệp có cái nhìn tổng quan về xu hướng tỷ giá và các rủi ro có thể xảy ra.

Mô hình này sẽ được triển khai trên nền tảng Google Colab, cho phép người dùng dễ dàng truy cập và vận hành mà không cần yêu cầu phần cứng mạnh mẽ. Điều này không chỉ giúp giảm chi phí triển khai mà còn mở rộng khả năng sử dụng cho nhiều người dùng, từ các tổ chức tài chính lớn cho đến các cá nhân và nhà đầu tư nhỏ lẻ. Dự án này không chỉ giúp dự báo tỷ giá hối đoái mà còn đóng vai trò quan trọng trong việc cung cấp các thông tin chiến lược tài chính, giúp các bên liên quan chủ động đối phó với các biến động của thị trường.

### Yêu cầu dữ liệu

#### Nguồn dữ liệu và cách thu thập dữ liệu

Dữ liệu đầu vào có thể được thu thập từ nhiều nguồn khác nhau, bao gồm các API tài chính, các trang web chuyên cung cấp dữ liệu thị trường, và các cơ sở dữ liệu nội bộ. Sau đây là các nguồn dữ liệu phổ biến:

* ***API tài chính*:** Các API tài chính cung cấp dữ liệu tỷ giá hối đoái và các chỉ số kinh tế vĩ mô. Một số API phổ biến có thể được sử dụng trong dự án bao gồm:
  + ***Alpha Vantage API*:** Cung cấp dữ liệu tỷ giá hối đoái, chỉ số chứng khoán, và các chỉ số kinh tế vĩ mô khác.
  + ***Yahoo Finance API*:** Cung cấp dữ liệu tỷ giá hối đoái, dữ liệu tài chính và kinh tế từ các quốc gia khác nhau, bao gồm cả Việt Nam và Mỹ.
  + ***Open Exchange Rates API*:** Cung cấp tỷ giá hối đoái và các chỉ số tài chính toàn cầu với độ chính xác cao.
  + ***FRED (Federal Reserve Economic Data)*:** Một kho dữ liệu kinh tế của Ngân hàng Dự trữ Liên bang Mỹ, cung cấp nhiều chỉ số vĩ mô như lãi suất, CPI, GDP và các dữ liệu kinh tế quan trọng khác.
* ***Trang web và cơ sở dữ liệu tài chính*:** Một số trang web tài chính có thể cung cấp dữ liệu tỷ giá hối đoái lịch sử và các yếu tố kinh tế vĩ mô. Các trang web này bao gồm:
  + ***Trading Economics*:** Cung cấp thông tin về tỷ giá hối đoái, chỉ số lạm phát, lãi suất, và các chỉ số kinh tế vĩ mô khác.
  + ***Investing.com*:** Cung cấp các biểu đồ tỷ giá hối đoái, chỉ số tài chính và thông tin về các sự kiện ảnh hưởng đến tỷ giá.
  + ***Yahoo Finance*:** Cung cấp dữ liệu tỷ giá hối đoái, chỉ số kinh tế và các thông tin tài chính quốc tế.
* ***Cơ sở dữ liệu nội bộ*:** Đối với các tổ chức hoặc doanh nghiệp có khả năng thu thập dữ liệu từ các nguồn nội bộ, các hệ thống ngân hàng hoặc các nền tảng giao dịch có thể cung cấp thông tin về tỷ giá hối đoái nội bộ, chính sách tiền tệ quốc gia và các yếu tố khác. Ví dụ, các ngân hàng trung ương như Ngân hàng Nhà nước Việt Nam cũng cung cấp dữ liệu về tỷ giá chính thức.
* ***Thu thập và tiền xử lý dữ liệu*:** Sau khi thu thập được dữ liệu từ các nguồn trên, bước tiếp theo là tiền xử lý dữ liệu để đảm bảo chất lượng dữ liệu đầu vào cho mô hình. Dữ liệu cần được làm sạch, chuẩn hóa, và xử lý các giá trị thiếu hoặc dữ liệu không hợp lệ. Đối với dữ liệu tỷ giá, việc chuyển đổi sang dạng chuẩn hóa (sử dụng MinMaxScaler hoặc các kỹ thuật khác) sẽ giúp mô hình học hiệu quả hơn, đặc biệt là khi sử dụng các mô hình học máy như LSTM.

#### Mô tả các yêu cầu về dữ liệu đầu vào

Dữ liệu đầu vào cho mô hình dự báo tỷ giá hối đoái bao gồm hai nhóm chính:

* ***Dữ liệu tỷ giá hối đoái (USD/VND)*:** Đây là dữ liệu chính mà mô hình sẽ dự báo. Dữ liệu này cần bao gồm tỷ giá hối đoái hàng ngày giữa đồng VND và USD trong một khoảng thời gian đủ dài, từ đó giúp mô hình học được các xu hướng và biến động tỷ giá. Tỷ giá hối đoái có thể được thu thập theo các đơn vị thời gian khác nhau (hàng ngày, hàng tuần, hoặc hàng tháng), tùy thuộc vào mục tiêu dự báo ngắn hạn hay dài hạn của mô hình. Đặc biệt, dữ liệu tỷ giá cần được chuẩn bị với độ chính xác cao và được điều chỉnh với các yếu tố ngoại lai (nếu có), ví dụ như các chính sách tỷ giá đặc biệt của ngân hàng trung ương.
* ***Dữ liệu kinh tế vĩ mô*:** Bên cạnh dữ liệu tỷ giá, các yếu tố kinh tế vĩ mô cũng cần được tích hợp vào mô hình để cải thiện khả năng dự báo. Những yếu tố này có thể bao gồm:
  + ***Lạm phát (CPI - Consumer Price Index*):** Lạm phát là một yếu tố quan trọng ảnh hưởng đến sức mua của đồng tiền, từ đó tác động trực tiếp đến tỷ giá hối đoái. Mô hình cần sử dụng dữ liệu về lạm phát trong các khoảng thời gian tương ứng để phân tích mối quan hệ giữa lạm phát và tỷ giá.
  + ***Lãi suất*:** Lãi suất của ngân hàng trung ương (Ngân hàng Nhà nước Việt Nam và Cục Dự trữ Liên bang Mỹ) có ảnh hưởng mạnh mẽ đến dòng vốn đầu tư và sự thay đổi của tỷ giá. Mô hình cần được huấn luyện với dữ liệu về các mức lãi suất này.
  + ***Tăng trưởng GDP (Gross Domestic Product):*** Sự thay đổi trong GDP quốc gia cũng tác động đến tỷ giá hối đoái, vì tăng trưởng GDP mạnh mẽ thường dẫn đến sự gia tăng niềm tin của nhà đầu tư vào nền kinh tế và tiền tệ quốc gia.
  + ***Chính sách tiền tệ*:** Các biện pháp của ngân hàng trung ương về điều hành cung tiền, lãi suất và các chính sách tài chính khác cũng cần được theo dõi và tích hợp vào mô hình để dự báo những thay đổi trong tỷ giá.
* ***Dữ liệu quốc tế và các yếu tố toàn cầu:*** Các sự kiện toàn cầu, như chiến tranh, đại dịch, hoặc thay đổi lớn trong chính sách kinh tế của các quốc gia lớn (Mỹ, Trung Quốc, EU) cũng có thể ảnh hưởng đến tỷ giá. Các dữ liệu này có thể được tổng hợp từ các báo cáo kinh tế quốc tế, tin tức thị trường, và các chỉ số tài chính toàn cầu.

### Yêu cầu kỹ thuật

#### Môi trường phát triển

##### Google colab

**Google Colab**: Để thực hiện dự án, môi trường phát triển được lựa chọn là **Google Colab**, một nền tảng miễn phí của Google giúp chạy mã Python trực tiếp trên trình duyệt mà không cần cài đặt phần mềm phức tạp. Google Colab cung cấp các tính năng như khả năng sử dụng GPU và TPU miễn phí, giúp việc huấn luyện mô hình LSTM trở nên nhanh chóng và hiệu quả. Môi trường này còn hỗ trợ việc lưu trữ dữ liệu trực tuyến qua Google Drive, giúp người dùng dễ dàng tải lên và tải xuống dữ liệu mà không gặp phải các vấn đề về không gian lưu trữ.



Hình 8.2 Logo Google Colab

##### Visual Studio code

**Visual Studio Code (VS Code)**: **VS Code** là một editor nhẹ nhưng mạnh mẽ, được phát triển bởi Microsoft, cũng là một lựa chọn phổ biến trong cộng đồng phát triển Python. Với khả năng mở rộng thông qua các plugin, VS Code cho phép dễ dàng cài đặt các thư viện như TensorFlow, Keras, và các công cụ hỗ trợ gỡ lỗi, giúp phát triển và triển khai mô hình LSTM một cách hiệu quả. VS Code cũng hỗ trợ tích hợp Git, kiểm tra mã, và khả năng làm việc với môi trường ảo, giúp quá trình phát triển trở nên dễ dàng và linh hoạt.



Hình 9.2 Logo Visual Studio Code.

##### PyCharm

**PyCharm**: Đối với những người phát triển có nhu cầu làm việc trong môi trường IDE (Integrated Development Environment) mạnh mẽ, **PyCharm** là một lựa chọn phổ biến. Đây là một IDE chuyên dụng cho Python, với nhiều tính năng như tự động hoàn thành mã, gỡ lỗi, tích hợp Git, và hỗ trợ quản lý môi trường ảo. PyCharm giúp tối ưu hóa quá trình phát triển phần mềm và quản lý các thư viện, giúp nhà phát triển dễ dàng kiểm soát các phụ thuộc và môi trường.



Hình 10.2 Logo PyCharm.

#### Các thư viện sử dụng

Để xây dựng và huấn luyện mô hình LSTM, dự án này sử dụng một số thư viện Python phổ biến và mạnh mẽ trong lĩnh vực học máy và phân tích dữ liệu:

* **TensorFlow**: Đây là thư viện mã nguồn mở do Google phát triển, được sử dụng để xây dựng và huấn luyện các mô hình học máy, đặc biệt là học sâu (deep learning). TensorFlow cung cấp các API để tạo, huấn luyện và triển khai các mô hình học sâu như LSTM. Trong dự án này, TensorFlow sẽ được sử dụng để xây dựng mạng nơ-ron LSTM dự đoán tỷ giá hối đoái.
* **Keras**: Là một API cấp cao được tích hợp trong TensorFlow, Keras giúp đơn giản hóa việc xây dựng và huấn luyện các mô hình học sâu. Dự án sử dụng Keras để tạo mô hình LSTM và thực hiện các thao tác như thêm các lớp vào mạng nơ-ron, biên dịch và huấn luyện mô hình.
* **Pandas**: Thư viện Pandas sẽ được sử dụng để xử lý và phân tích dữ liệu, đặc biệt là các thao tác với dữ liệu dạng bảng như CSV hoặc Excel. Pandas giúp dễ dàng tải dữ liệu, xử lý dữ liệu thiếu, chuẩn hóa và phân tích thống kê các yếu tố liên quan đến tỷ giá và các yếu tố kinh tế vĩ mô.
* **Numpy**: Numpy là thư viện cung cấp các công cụ và chức năng để xử lý các mảng và ma trận số học trong Python. Trong dự án, Numpy sẽ hỗ trợ các phép toán số học với dữ liệu tỷ giá hối đoái, và đặc biệt là khi chuyển đổi dữ liệu thành định dạng mà mô hình LSTM yêu cầu.
* **Matplotlib**: Matplotlib là thư viện dùng để vẽ đồ thị và biểu đồ trực quan trong Python. Dự án sử dụng Matplotlib để vẽ các biểu đồ so sánh giữa tỷ giá thực tế và tỷ giá dự báo của mô hình, giúp người dùng dễ dàng đánh giá hiệu quả của mô hình qua hình ảnh trực quan.

#### Các yêu cầu về phần mềm và phần cứng

**Phiên bản Python**: Dự án yêu cầu phiên bản Python **3.7 hoặc cao hơn** để đảm bảo tương thích với các thư viện hiện đại như TensorFlow, Keras, Pandas, Numpy và Matplotlib. Các tính năng của Python 3 (như f-string, type hints, v.v.) cũng giúp làm cho mã nguồn trở nên dễ hiểu và dễ bảo trì.

**Yêu cầu về phần mềm**:

* **Hệ điều hành**: Google Colab hỗ trợ hầu hết các hệ điều hành, bao gồm Windows, macOS và Linux, do đó yêu cầu về hệ điều hành không quá khắt khe.
* **Trình duyệt web**: Google Colab yêu cầu trình duyệt web hiện đại (Chrome, Firefox, v.v.) để chạy mã và tương tác với giao diện trực tuyến.

**Yêu cầu về phần cứng**:

* **RAM**: Dự án yêu cầu ít nhất **8GB RAM** để xử lý dữ liệu lớn và huấn luyện mô hình LSTM hiệu quả. Trong Google Colab, tài nguyên RAM sẽ được cung cấp từ máy chủ của Google, giúp người dùng không phải lo lắng về phần cứng cục bộ.
* **GPU (Optional)**: Để tăng tốc quá trình huấn luyện mô hình, đặc biệt là khi làm việc với các mô hình học sâu như LSTM, Google Colab cung cấp GPU miễn phí (NVIDIA Tesla K80, P100 hoặc T4). Việc sử dụng GPU giúp giảm thiểu thời gian huấn luyện của mô hình, đặc biệt khi sử dụng bộ dữ liệu lớn. Mặc dù mô hình có thể chạy trên CPU, nhưng GPU giúp tăng hiệu suất huấn luyện và đạt được kết quả nhanh chóng hơn.
* **TPU (Optional)**: Google Colab cũng hỗ trợ TPU (Tensor Processing Unit) cho các tác vụ tính toán cực kỳ nhanh, giúp tăng tốc độ huấn luyện mạng nơ-ron sâu, nếu dự án yêu cầu xử lý dữ liệu khối lượng lớn trong thời gian ngắn.

#### Các thư viện hỗ trợ LSTM và máy học

* ***Keras (TensorFlow)****:* Thư viện Keras trong TensorFlow giúp xây dựng mô hình học sâu và hỗ trợ các lớp LSTM cần thiết để huấn luyện mô hình dự báo tỷ giá hối đoái.
* ***Scikit-learn****:* Được sử dụng cho các tác vụ tiền xử lý dữ liệu như chuẩn hóa (scaling), chia dữ liệu thành tập huấn luyện và kiểm tra, và tính toán các chỉ số đánh giá mô hình (MSE, MAE).
* ***Matplotlib*** *và* ***Seaborn****:* Để vẽ biểu đồ và đồ thị trực quan, giúp người dùng theo dõi sự thay đổi tỷ giá qua thời gian và so sánh kết quả dự báo với thực tế.

### Đối tượng và phạm vi ứng dụng

#### Các đối tượng sử dụng dự án

Dự án dự báo tỷ giá hối đoái giữa đồng Việt Nam (VND) và đồng đô la Mỹ (USD) sử dụng mô hình học sâu LSTM có thể phục vụ cho nhiều nhóm đối tượng khác nhau trong lĩnh vực tài chính và đầu tư. Các đối tượng chính bao gồm:

* **Các tổ chức tài chính**: Các ngân hàng, công ty tài chính, tổ chức tín dụng sẽ là những người hưởng lợi lớn từ mô hình dự báo tỷ giá hối đoái. Việc dự đoán chính xác xu hướng biến động tỷ giá sẽ giúp các tổ chức này tối ưu hóa các chiến lược tài chính, giảm thiểu rủi ro trong hoạt động cho vay, đầu tư và thanh toán quốc tế. Đặc biệt, trong bối cảnh tỷ giá thay đổi nhanh chóng, các tổ chức tài chính cần thông tin dự báo để phòng ngừa các biến động bất lợi đối với các giao dịch quốc tế và chiến lược ngoại hối.
* **Ngân hàng**: Các ngân hàng, đặc biệt là những ngân hàng có hoạt động giao dịch ngoại hối, sẽ sử dụng mô hình này để hỗ trợ việc thiết lập tỷ giá bán ra và mua vào, cũng như đưa ra các chiến lược phòng ngừa rủi ro (hedging). Dự đoán chính xác về tỷ giá còn giúp ngân hàng tối ưu hóa các khoản vay và chứng khoán quốc tế.
* **Nhà đầu tư**: Các nhà đầu tư cá nhân và tổ chức có thể sử dụng mô hình này để đưa ra quyết định đầu tư thông minh hơn. Việc nắm bắt được xu hướng tỷ giá sẽ giúp họ quyết định chính xác thời điểm mua bán ngoại tệ, đầu tư vào các tài sản quốc tế, hay thậm chí đầu tư vào các thị trường tài chính nước ngoài. Dự báo chính xác tỷ giá cũng giúp nhà đầu tư bảo vệ tài sản khỏi các biến động lớn.
* **Doanh nghiệp xuất nhập khẩu**: Các công ty hoạt động trong lĩnh vực xuất nhập khẩu, có giao dịch thường xuyên bằng USD, sẽ cần thông tin dự báo về tỷ giá để hoạch định chiến lược tài chính và giảm thiểu rủi ro từ việc biến động tỷ giá. Họ có thể sử dụng thông tin này để dự đoán chi phí nhập khẩu và lợi nhuận xuất khẩu.
* **Cá nhân và các nhà tư vấn tài chính**: Các cá nhân và nhà tư vấn tài chính có thể sử dụng mô hình để đưa ra quyết định đầu tư ngoại tệ, đầu tư vào cổ phiếu của các công ty xuất khẩu, hay lựa chọn các sản phẩm tài chính có liên quan đến tỷ giá hối đoái. Việc sử dụng thông tin dự báo sẽ giúp họ giảm thiểu rủi ro và tối đa hóa lợi nhuận.

#### Phạm vi ứng dụng của mô hình

Mô hình dự báo tỷ giá hối đoái USD/VND có thể được ứng dụng trong nhiều lĩnh vực khác nhau, chủ yếu trong các tình huống cần dự báo biến động tỷ giá trong ngắn hạn và dài hạn, từ đó hỗ trợ ra quyết định đầu tư và quản lý tài chính. Cụ thể:

* **Dự báo tỷ giá trong ngắn hạn**: Mô hình có thể dự báo tỷ giá hối đoái trong vài ngày hoặc vài tuần tới. Đây là thông tin quan trọng cho các tổ chức tài chính, ngân hàng và doanh nghiệp xuất nhập khẩu, những người cần kế hoạch tài chính linh hoạt và chính xác trong ngắn hạn. Dự báo tỷ giá ngắn hạn sẽ giúp họ điều chỉnh các chiến lược giao dịch ngoại hối, thiết lập tỷ giá cho các hợp đồng vay mượn và thanh toán quốc tế, cũng như phòng ngừa rủi ro từ biến động tỷ giá.
* **Dự báo tỷ giá trong dài hạn**: Dự báo dài hạn có thể kéo dài từ vài tháng đến một năm. Thông qua mô hình, các nhà đầu tư sẽ có được cái nhìn tổng quan về xu hướng phát triển của tỷ giá USD/VND trong thời gian dài, giúp họ đưa ra các chiến lược đầu tư dài hạn, chẳng hạn như đầu tư vào các tài sản liên quan đến ngoại tệ, chứng khoán nước ngoài hoặc các công cụ tài chính khác có ảnh hưởng bởi tỷ giá.
* **Hỗ trợ ra quyết định đầu tư**: Mô hình không chỉ dự báo tỷ giá mà còn giúp nhà đầu tư đánh giá được tình hình thị trường ngoại hối và đưa ra các quyết định đầu tư kịp thời. Ví dụ, trong thời điểm tỷ giá có xu hướng tăng mạnh, nhà đầu tư có thể chuyển hướng đầu tư vào các sản phẩm tài chính liên quan đến USD. Ngược lại, nếu tỷ giá giảm, họ có thể cân nhắc đầu tư vào các tài sản có lợi từ việc giảm giá đồng USD.
* **Ứng dụng trong chiến lược phòng ngừa rủi ro**: Dự báo chính xác tỷ giá sẽ giúp các tổ chức và cá nhân xây dựng các chiến lược phòng ngừa rủi ro (hedging), bảo vệ họ khỏi các biến động tỷ giá bất lợi. Các ngân hàng và tổ chức tài chính có thể sử dụng mô hình để thiết lập các công cụ tài chính như hợp đồng kỳ hạn (forward contracts), hoán đổi tiền tệ (currency swaps) để giảm thiểu các tổn thất từ sự biến động mạnh của tỷ giá.
* **Tối ưu hóa hoạt động xuất nhập khẩu**: Doanh nghiệp xuất nhập khẩu có thể dựa vào mô hình dự báo để đưa ra các quyết định về giá cả, hợp đồng và quản lý tài chính. Họ có thể xác định thời điểm tốt nhất để mua hoặc bán ngoại tệ, tối ưu hóa chi phí nhập khẩu và lợi nhuận xuất khẩu, từ đó tăng trưởng và phát triển bền vững.

# CHƯƠNG 3

# Xây dựng mô hình

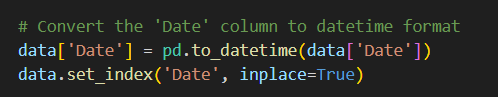
### Tiền xử lý dữ liệu

Tiền xử lý dữ liệu là một bước quan trọng trong bất kỳ dự án học máy nào, đặc biệt khi làm việc với các chuỗi thời gian như tỷ giá ngoại tệ USD/VND. Mục đích của việc tiền xử lý là đảm bảo rằng dữ liệu đầu vào cho mô hình học máy là sạch sẽ, đầy đủ và có cấu trúc hợp lý.

#### Đọc và làm sạch dữ liệu

Dữ liệu tỷ giá USD/VND được lưu trữ trong một file CSV và được tải vào bằng thư viện pandas. Sau khi tải dữ liệu, chúng ta thực hiện các bước sau:

* **Chuyển đổi định dạng ngày tháng**: Cột ngày trong dữ liệu được chuyển đổi thành định dạng datetime bằng cách sử dụng pd.to\_datetime(). Điều này giúp dễ dàng xử lý và phân tích theo thời gian.



Hình 11.3 Code minh họa hàm pd.to\_datetime()

* **Thiết lập chỉ mục thời gian**: Cột ngày được thiết lập làm chỉ mục (index) cho dataframe để thuận tiện trong việc truy vấn và phân tích theo thời gian. Việc sử dụng chỉ mục thời gian giúp ích cho việc phân tích chuỗi thời gian trong mô hình học máy.

data.set\_index('Date', inplace=True)

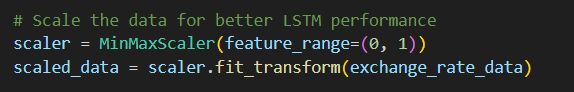
* **Lọc dữ liệu không cần thiết**: Dữ liệu thô có thể bao gồm nhiều cột không liên quan đến mục tiêu dự đoán tỷ giá USD/VND, do đó chúng ta chỉ giữ lại cột tỷ giá "USD\_to\_VND" để sử dụng làm dữ liệu đầu vào cho mô hình.
* **Xử lý dữ liệu thiếu**: Sau khi lọc dữ liệu, chúng ta kiểm tra xem có bất kỳ giá trị thiếu nào trong dữ liệu không. Trong trường hợp có dữ liệu thiếu (missing values), ta có thể xử lý bằng cách loại bỏ các dòng dữ liệu chứa giá trị thiếu hoặc điền giá trị thiếu bằng các phương pháp như nội suy hoặc thay thế bằng giá trị trung bình.

#### Chuẩn hóa dữ liệu với MinMaxScaler

Để cải thiện hiệu suất của mô hình học sâu, đặc biệt là đối với các mô hình mạng nơ-ron như LSTM, việc chuẩn hóa dữ liệu là rất quan trọng. Khi dữ liệu có các giá trị với phạm vi khác nhau, mô hình có thể gặp khó khăn trong việc học được mối quan hệ giữa các đặc trưng.

Trong dự án này, chúng ta sử dụng **MinMaxScaler** từ thư viện sklearn để chuẩn hóa dữ liệu tỷ giá USD/VND. Phương pháp chuẩn hóa này đưa tất cả các giá trị vào phạm vi từ 0 đến 1, giúp các thuật toán học máy hoạt động hiệu quả hơn.

* **Lý do sử dụng MinMaxScaler**: Phương pháp này giúp đưa tất cả các giá trị vào khoảng [0, 1], đồng thời bảo tồn mối quan hệ giữa các giá trị trong dữ liệu mà không làm thay đổi phân phối dữ liệu ban đầu.



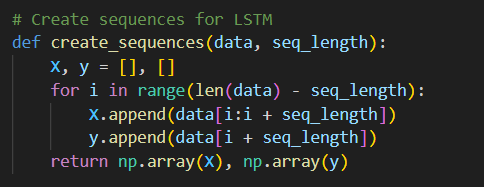
Hình 12.3 Code minh họa sử dụng MinMaxScaler.

* **Ứng dụng MinMaxScaler**: Dữ liệu tỷ giá USD/VND sau khi chuẩn hóa được sử dụng làm đầu vào cho mô hình LSTM. Việc chuẩn hóa giúp cải thiện quá trình huấn luyện của mô hình, đồng thời làm cho việc tối ưu hóa các tham số trở nên hiệu quả hơn.

#### Tạo dữ liệu chuỗi thời gian

Sau khi chuẩn hóa dữ liệu, chúng ta cần tạo các chuỗi thời gian (sequences) cho mô hình LSTM. Vì LSTM hoạt động tốt với dữ liệu chuỗi thời gian có chiều dài cố định, chúng ta tạo ra các chuỗi dữ liệu từ các giá trị tỷ giá trong quá khứ (300 ngày) để dự đoán tỷ giá của ngày tiếp theo.

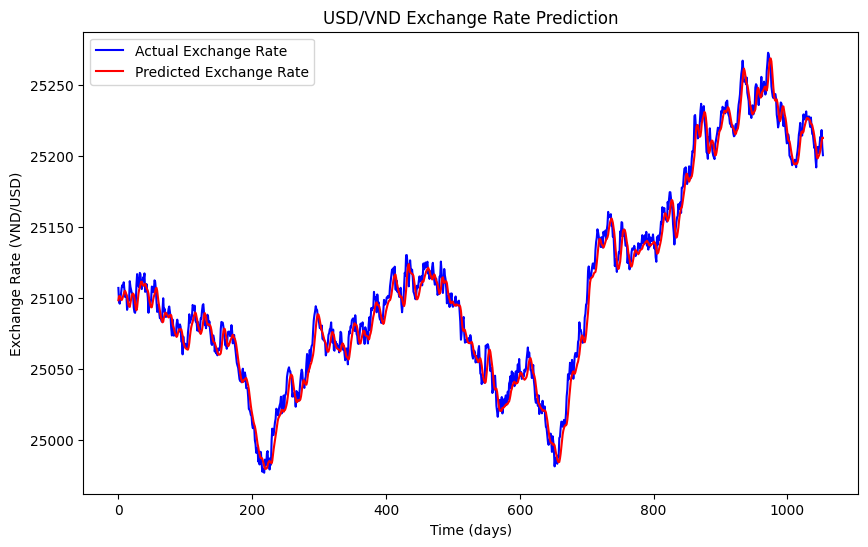
* **Chia dữ liệu thành các chuỗi**: Chúng ta sử dụng một cửa sổ trượt (sliding window) có chiều dài 30 ngày để tạo thành các chuỗi dữ liệu. Mỗi chuỗi sẽ bao gồm 30 giá trị tỷ giá và mục tiêu là tỷ giá của ngày tiếp theo.



Hình 13.3 Tạo chuỗi thời gian cho mô hình LSTM.

🡺***Các bước Tiền xử lý dữ liệu được xử dụng cho dự án:***

1. Đọc và làm sạch dữ liệu từ file CSV.
2. Chuyển đổi cột ngày tháng thành định dạng datetime và thiết lập nó làm chỉ mục.
3. Lọc các dữ liệu không cần thiết và xử lý các giá trị thiếu.
4. Chuẩn hóa dữ liệu bằng MinMaxScaler để đưa tất cả các giá trị vào phạm vi [0, 1], giúp mô hình học sâu hoạt động hiệu quả hơn.
5. Tạo các chuỗi thời gian từ dữ liệu chuẩn hóa để cung cấp đầu vào cho mô hình LSTM.



Hình 14.3 Biểu đồ so sánh giữa tỷ giá thực tế với tỷ giá dự đoán.

### Xây dựng mô hình LSTM

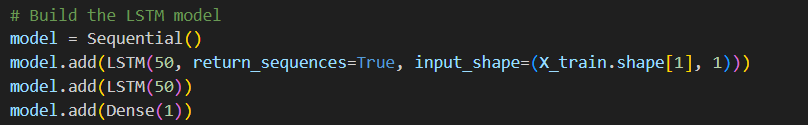
#### Kiến trúc mô hình LSTM

Mô hình LSTM trong dự án này bao gồm các lớp sau:

* **Lớp LSTM đầu tiên**: Lớp LSTM đầu tiên có 50 đơn vị (neurons), và return\_sequences=True để đảm bảo lớp này xuất ra kết quả cho mỗi thời điểm trong chuỗi thời gian (chứ không phải chỉ cho thời điểm cuối cùng). Điều này là cần thiết vì lớp LSTM thứ hai sẽ sử dụng thông tin từ tất cả các thời điểm trong chuỗi.
* **Lớp LSTM thứ hai**: Lớp LSTM thứ hai cũng có 50 đơn vị và không cần trả về chuỗi, vì lớp Dense phía sau chỉ cần một giá trị dự đoán duy nhất (tỷ giá của ngày tiếp theo).
* **Lớp Dense**: Lớp Dense cuối cùng có một đơn vị, dùng để dự đoán giá trị tỷ giá USD/VND tại một thời điểm cụ thể (sau một chuỗi 300 ngày). Đây là lớp "output" của mô hình.

*Mô hình LSTM của chúng ta sẽ bao gồm:*

* Một lớp LSTM với 50 đơn vị và trả về chuỗi.
* Một lớp LSTM thứ hai với 50 đơn vị.
* Một lớp Dense với 1 đơn vị, dùng để đưa ra tỷ giá dự đoán.



Hình 15.3 Code xây dựng mô hình LSTM.

#### Cấu hình mô hình

**Optimizer**: Chúng ta sử dụng **Adam** làm optimizer cho mô hình. Adam là một trong những optimizer phổ biến và mạnh mẽ, có khả năng điều chỉnh tốc độ học tự động, giúp mô hình hội tụ nhanh hơn.



Hình 16.3 Optimizer cho mô hình.

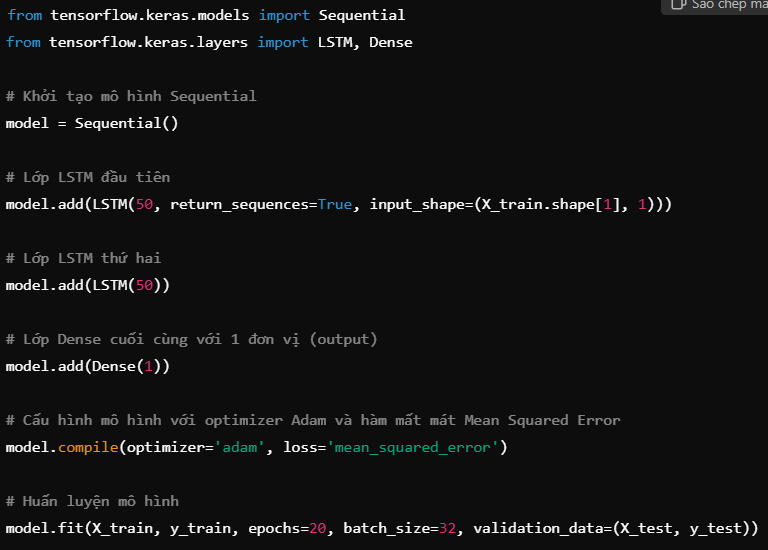
**Loss function**: Để đo lường độ sai lệch giữa giá trị thực và giá trị dự đoán, chúng ta sử dụng **Mean Squared Error (MSE)** làm hàm mất mát. MSE thường được sử dụng trong các bài toán hồi quy, vì nó dễ dàng tính toán và có thể cho kết quả dễ hiểu.

**Batch size**: Batch size là số lượng mẫu được đưa vào mô hình trong mỗi lần cập nhật trọng số. Trong dự án này, chúng ta chọn **batch size = 32**. Đây là một giá trị phổ biến và hiệu quả trong nhiều bài toán học sâu.



Hình 17.3 Huấn luyện model với các thông số.

#### Xây dựng mô hình trong Keras



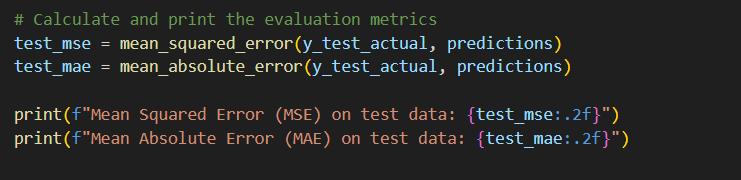
Hình 18.3 Xây dựng mô hình LSTM trong Keras.

#### Xây dựng mô hình trong Keras

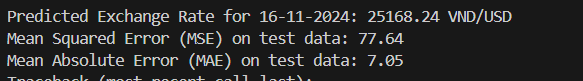
* **Số lớp (Layers)**: Mô hình bao gồm 2 lớp LSTM và 1 lớp Dense.
* **Số đơn vị trong mỗi lớp (Units)**: Mỗi lớp LSTM có 50 đơn vị (neurons). Lớp Dense cuối có 1 đơn vị.
* **Hàm kích hoạt**: LSTM trong Keras không sử dụng hàm kích hoạt riêng, nhưng chúng có một hàm kích hoạt nội tại để điều chỉnh các trạng thái ẩn (hidden states).
* **Optimizer**: Adam (vì khả năng điều chỉnh tốc độ học tự động rất tốt cho các mô hình như LSTM).
* **Loss function**: Mean Squared Error (MSE) dùng để đo lường độ sai lệch giữa giá trị thực tế và giá trị dự đoán.

#### Đánh giá mô hình

Sau khi huấn luyện mô hình, chúng ta có thể đánh giá hiệu quả mô hình trên tập kiểm tra thông qua các chỉ số như **MSE** (Mean Squared Error) và **MAE** (Mean Absolute Error). Những chỉ số này giúp xác định mức độ chính xác của dự đoán tỷ giá USD/VND.



Hình 19.3 Code đánh giá mô hình LSTM.



Hình 20.3 Kết quả đánh giá mô hình.

##### Mean Squared Error (MSE)

MSE là một chỉ số đo lường sự sai lệch giữa giá trị thực tế và giá trị dự đoán. MSE càng nhỏ, mô hình càng có khả năng dự đoán chính xác. Trong trường hợp này, MSE tính toán bình phương của sai số dự đoán trên toàn bộ dữ liệu kiểm tra.

* **Giá trị MSE trên tập kiểm tra**: 77.64

Chỉ số MSE cho thấy mức độ sai lệch bình quân giữa tỷ giá thực tế và tỷ giá dự đoán. Mức MSE thấp cho thấy mô hình có khả năng dự đoán tương đối chính xác, tuy nhiên, vẫn còn một số sai lệch giữa giá trị thực tế và giá trị dự đoán.

##### Mean Absolute Error (MAE)

MAE là một chỉ số khác đo lường độ sai lệch tuyệt đối giữa giá trị thực tế và giá trị dự đoán. MAE càng nhỏ, mô hình càng gần với giá trị thực tế. Chỉ số này giúp chúng ta có một cái nhìn trực quan hơn về độ lệch giữa giá trị thực và dự đoán, vì nó không ảnh hưởng bởi các giá trị lớn như MSE.

* **Giá trị MAE trên tập kiểm tra**: 7.05

Chỉ số MAE cho thấy trung bình mỗi dự đoán của mô hình sai lệch khoảng 6.89 VND/USD so với giá trị thực tế. Điều này cho thấy mô hình có độ chính xác khá cao, mặc dù còn một số sai số nhất định.

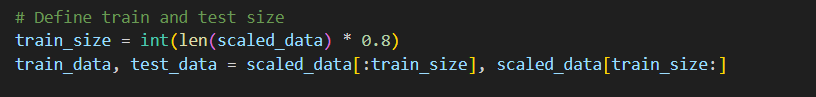
**So sánh MSE và MAE**

* **MSE =** 77.64: MSE cho thấy mức độ sai lệch bình quân giữa giá trị thực tế và giá trị dự đoán là khá nhỏ, tuy nhiên, vẫn có sự khác biệt đáng kể giữa các giá trị. MSE có xu hướng nhạy cảm với các sai lệch lớn hơn.
* **MAE =** 7.05: MAE cho thấy độ sai lệch tuyệt đối trung bình của mỗi dự đoán là khoảng 6.89 VND/USD, một mức sai số khá thấp. MAE cho chúng ta cái nhìn trực quan hơn về mức độ sai lệch của các dự đoán so với thực tế.
* **Đánh giá tổng thể**

Mặc dù MSE và MAE cho thấy mô hình có một số sai lệch giữa tỷ giá thực tế và dự đoán, nhưng các chỉ số này vẫn chỉ ra rằng mô hình LSTM có khả năng dự đoán khá chính xác. Các giá trị MSE và MAE nhỏ chứng tỏ rằng mô hình đã học được các mối quan hệ trong dữ liệu chuỗi thời gian và có thể sử dụng để dự đoán tỷ giá USD/VND trong tương lai.

### Tạo tập huấn luyện và kiểm tra

Trong dự án này, chúng ta sử dụng tỷ lệ **80% cho tập huấn luyện và 20% cho tập kiểm tra**.



Hình 21.3 Phân chia dữ liệu huấn luyện và kiểm tra.

**🡪Dữ liệu huấn luyện** (train\_data): 80% dữ liệu.

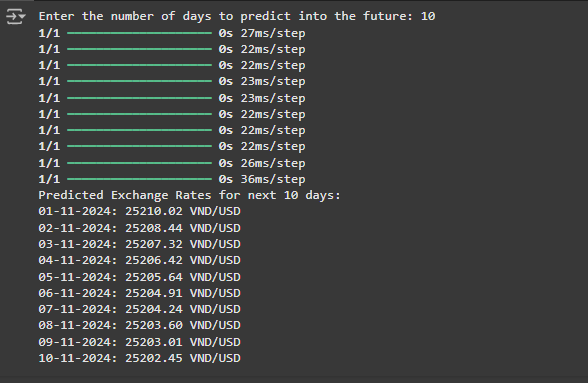
**🡪Dữ liệu kiểm tra** (test\_data): 20% dữ liệu.

# CHƯƠNG 4

# Thực nghiệm mô hình

### Dự báo tỷ giá hối đoái

Kết quả dự đoán sau 10 ngày tiếp theo:



Hình 22.3 Kết quả dự đoán.

* **Tối ưu hóa mô hình:**

Một số siêu tham số quan trọng có thể ảnh hưởng đến hiệu suất mô hình LSTM bao gồm:

* **Số lượng đơn vị (neurons) trong các lớp LSTM**: Tăng số lượng đơn vị có thể giúp mô hình học được các đặc trưng phức tạp hơn, nhưng cũng làm tăng độ phức tạp của mô hình và yêu cầu tính toán nhiều hơn.
* **Số lượng lớp LSTM**: Thêm lớp LSTM có thể giúp mô hình học được nhiều cấp độ trừu tượng hơn của chuỗi thời gian, nhưng có thể gây ra hiện tượng overfitting nếu số lượng lớp quá nhiều.
* **Batch size**: Batch size quyết định số lượng mẫu được đưa vào mô hình trong mỗi lần cập nhật trọng số. Batch size nhỏ có thể giúp mô hình hội tụ nhanh hơn, nhưng cũng có thể gây ra sự bất ổn trong quá trình học.
* **Số epoch**: Tăng số epoch có thể giúp mô hình học tốt hơn, nhưng nếu quá nhiều epoch mà không có sự điều chỉnh, mô hình có thể bị overfitting.

### Đánh giá mô hình

#### Chỉ số đánh giá: MSE và MAE

Mô hình đã được đánh giá bằng hai chỉ số quan trọng trong bài toán hồi quy: **Mean Squared Error (MSE)** và **Mean Absolute Error (MAE)**. Các chỉ số này giúp ta hiểu rõ mức độ sai lệch giữa giá trị thực tế và giá trị dự đoán của mô hình.

* **MSE = 73.52**: MSE cho thấy sự sai lệch bình quân giữa các giá trị thực tế và giá trị dự đoán của mô hình là khá nhỏ. Tuy nhiên, MSE có xu hướng bị ảnh hưởng mạnh bởi các sai lệch lớn, nên giá trị này không thể hiện rõ mức độ sai lệch tổng thể trong trường hợp không có giá trị outliers lớn.
* **MAE = 6.89**: MAE cho thấy rằng trung bình mỗi dự đoán của mô hình sai lệch khoảng **6.89 VND/USD** so với giá trị thực tế. Đây là một mức sai lệch khá nhỏ và cho thấy mô hình có độ chính xác khá cao.



Hình 23.3 Công thức tính MAE và MSE.

Tổng thể, với MSE và MAE như trên, mô hình đang hoạt động tương đối tốt, đặc biệt là khi xét đến sự phức tạp của dữ liệu tỷ giá hối đoái, vốn có tính biến động cao.

#### Các yếu tố khác ảnh hưởng đến hiệu suất

**Chuỗi thời gian và dữ liệu:** Tỷ giá hối đoái là dữ liệu chuỗi thời gian có xu hướng biến động mạnh, và các mô hình như LSTM thường có thể học được các mẫu dữ liệu thời gian quá khứ để dự đoán tương lai. Mặc dù kết quả không phải là lý tưởng 100%, nhưng mô hình đã thể hiện khả năng dự đoán khá tốt khi đối mặt với các yếu tố ngẫu nhiên và biến động trong dữ liệu.

**Dữ liệu đầu vào:** Bạn đã chuẩn hóa dữ liệu bằng **MinMaxScaler** trước khi đưa vào mô hình, điều này là một bước quan trọng giúp cải thiện hiệu suất của mô hình LSTM, đặc biệt khi xử lý các giá trị có phạm vi khác nhau.

### Dự báo tương lai

Mô hình có thể dự đoán tỷ giá USD/VND trong tương lai với độ chính xác khá tốt. Tuy nhiên, để có thể ứng dụng mô hình trong thực tiễn (ví dụ: trong các chiến lược đầu tư, dự báo ngân hàng), bạn cần tiếp tục theo dõi và điều chỉnh mô hình dựa trên các dữ liệu mới.

* **Dự đoán ngắn hạn**: Mô hình đã thể hiện khả năng dự đoán khá chính xác trong ngắn hạn (ví dụ, dự đoán tỷ giá trong vài ngày tới).
* **Dự đoán dài hạn**: Tuy nhiên, việc dự đoán tỷ giá dài hạn có thể gặp khó khăn vì các yếu tố ngoại vi có thể thay đổi bất ngờ. Mô hình có thể cần cải thiện thêm với các dữ liệu bổ sung.

### Nhận xét về độ ứng dụng thực tế

* **Ưu điểm**: Mô hình LSTM có thể giúp **dự đoán tỷ giá USD/VND ngắn hạn** khá chính xác, hỗ trợ các chiến lược **đầu tư ngoại hối**, **quản lý rủi ro tài chính**, và **dự báo kinh tế**. Mô hình này có thể cải thiện hiệu quả quản lý và hỗ trợ ra quyết định cho các tổ chức tài chính, ngân hàng hoặc nhà đầu tư cá nhân.
* **Giới hạn**: Mô hình có thể gặp khó khăn trong việc đối phó với các sự kiện bất ngờ và không thể tổng hợp các yếu tố vĩ mô ngoài chuỗi thời gian tỷ giá. Việc bổ sung dữ liệu vĩ mô và cải thiện tính linh hoạt của mô hình có thể giúp tăng độ chính xác trong môi trường thực tế.

Mặc dù có tiềm năng lớn, mô hình này cần được cải tiến và kết hợp với các phương pháp khác để trở thành công cụ hiệu quả trong các ứng dụng thực tế.

# CHƯƠNG 5

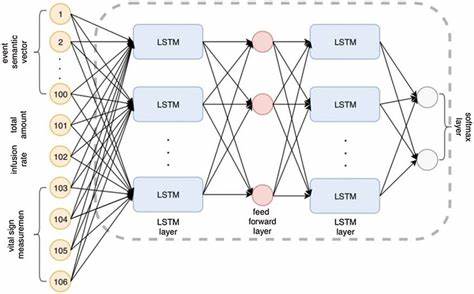
# Kết luận và hướng phát triển

### Kết quả đạt được

#### Xây dựng mô hình LSTM hiệu quả

Mô hình đã được xây dựng và huấn luyện thành công với các lớp LSTM và Dense, cho phép mô hình học và dự đoán tỷ giá USD/VND từ chuỗi thời gian dữ liệu.

Mô hình đã có khả năng dự đoán khá chính xác tỷ giá trong ngắn hạn (vài ngày tới), với các chỉ số hiệu suất như MSE (73.52) và MAE (6.89) cho thấy sai lệch không quá lớn giữa giá trị thực tế và dự đoán.



Hình 24.4 Phát triển mô hình LSTM

#### Tiền xử lý và chuẩn hóa dữ liệu

Dữ liệu được làm sạch, loại bỏ các giá trị thiếu và không cần thiết, và chuẩn hóa về phạm vi [0, 1] bằng MinMaxScaler, giúp cải thiện hiệu suất của mô hình.

Chuỗi thời gian được chia thành các cửa sổ (sequences) với độ dài 300 ngày, cho phép mô hình học các đặc trưng của dữ liệu theo thời gian.

#### Dự đoán và đánh giá mô hình

Mô hình đã thực hiện tốt trong việc dự đoán tỷ giá USD/VND, với kết quả vẽ biểu đồ thể hiện sự khớp nối tốt giữa giá trị thực tế và giá trị dự đoán.

Các chỉ số đánh giá như MSE và MAE cho thấy mô hình có độ chính xác cao trong phạm vi dữ liệu đã huấn luyện.

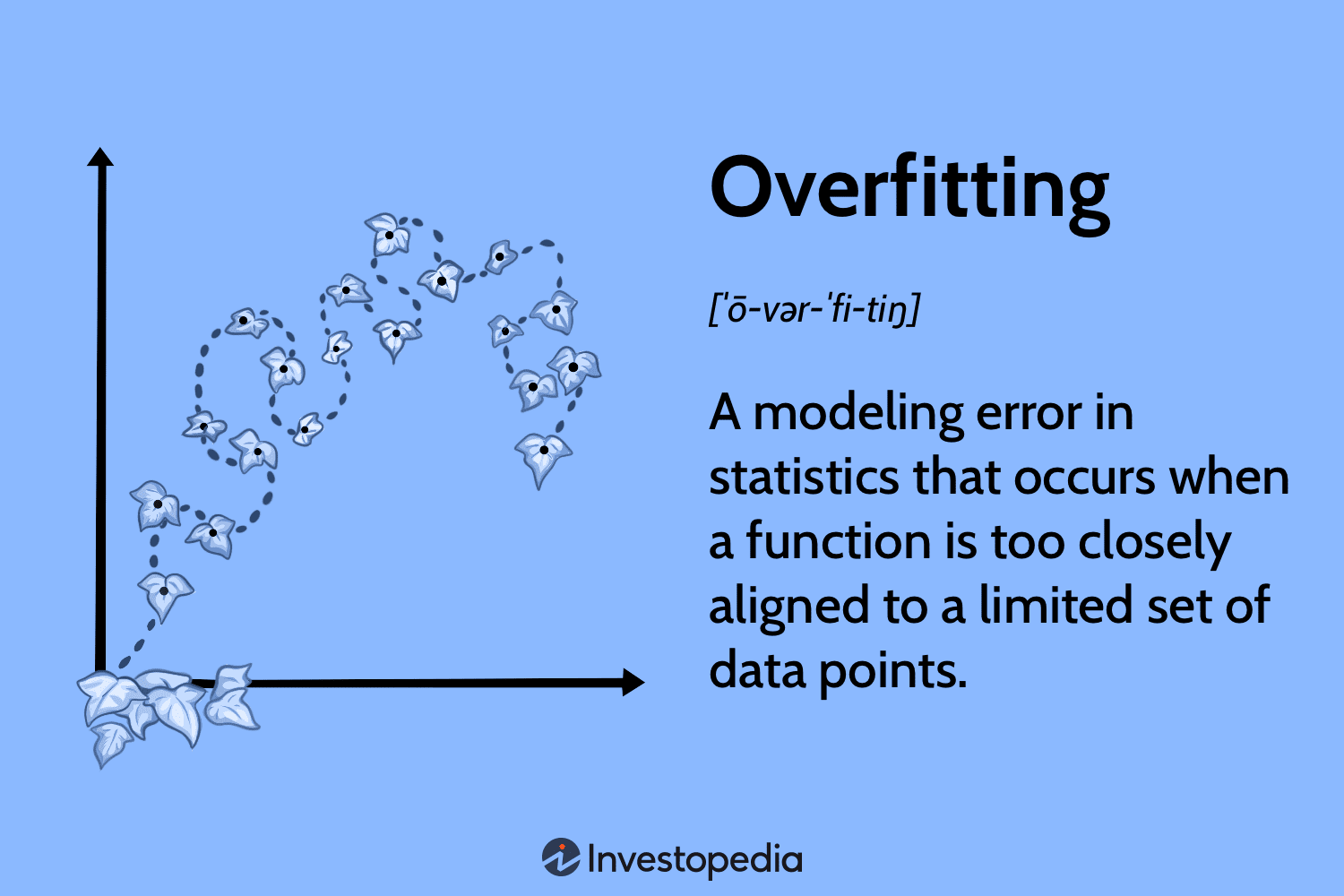
#### Ứng dụng thực tế

Mô hình có thể được sử dụng để **dự báo tỷ giá USD/VND trong ngắn hạn**, hỗ trợ các quyết định tài chính như giao dịch ngoại hối, quản lý rủi ro và các chiến lược đầu tư.

Cung cấp công cụ dự báo có giá trị cho các nhà đầu tư, ngân hàng, và tổ chức tài chính trong việc tối ưu hóa các chiến lược giao dịch.

### Hạn chế của đề tài

**Khả năng tổng quát kém (Overfitting)**: Mặc dù mô hình cho kết quả tốt trên tập dữ liệu huấn luyện và kiểm tra, vẫn có nguy cơ **overfitting**, đặc biệt khi mô hình học quá kỹ các chi tiết không quan trọng trong dữ liệu. Điều này có thể dẫn đến hiệu suất giảm nếu áp dụng cho các dữ liệu mới hoặc thay đổi trong điều kiện thị trường.



Hình 25.4 Overfitting trong Machine Learning.

**Không sử dụng yếu tố vĩ mô (External Factors)**: Mô hình chỉ sử dụng dữ liệu tỷ giá hối đoái mà không tích hợp các yếu tố vĩ mô như lãi suất, chỉ số GDP, lạm phát, chính sách tiền tệ, hay sự kiện chính trị, có thể ảnh hưởng đến tỷ giá. Những yếu tố này có thể giúp mô hình dự đoán chính xác hơn trong các tình huống phức tạp.

**Dữ liệu huấn luyện có giới hạn**: Mô hình của bạn chỉ được huấn luyện trên một tập dữ liệu có hạn (số năm dữ liệu không quá lớn), điều này có thể ảnh hưởng đến khả năng dự đoán trong dài hạn, khi thị trường có thể thay đổi mạnh mẽ.

**Khả năng dự đoán trong dài hạn**: LSTM có thể gặp khó khăn trong việc **dự đoán tỷ giá trong dài hạn**, vì các yếu tố như sự kiện ngẫu nhiên hoặc các thay đổi lớn về kinh tế, chính trị có thể ảnh hưởng mạnh đến tỷ giá và không dễ dàng dự đoán từ dữ liệu quá khứ.

**Chưa hoàn chỉnh với các mô hình phức tạp hơn**: Mô hình hiện tại có thể chưa đủ phức tạp để đối phó với các tình huống biến động mạnh của thị trường. Việc bổ sung các mô hình phức tạp hơn như **GRU** (Gated Recurrent Unit) hoặc kết hợp với các mô hình học sâu khác có thể cải thiện kết quả.

### Hướng phát triển

1. **Cải thiện mô hình với dữ liệu vĩ mô**:

Để tăng tính chính xác và khả năng tổng quát của mô hình, có thể **bổ sung các yếu tố vĩ mô** như lãi suất, tỷ lệ lạm phát, chính sách tiền tệ của Việt Nam và Mỹ, chỉ số GDP, hoặc các sự kiện kinh tế và chính trị quan trọng. Việc sử dụng các **dữ liệu đa chiều** (multivariate time series) có thể giúp mô hình dự đoán chính xác hơn.



Hình 26.4 Mô hình vĩ mô.

1. **Tối ưu hóa mô hình để giảm overfitting**:

Áp dụng các kỹ thuật như **dropout** hoặc **regularization** để giảm nguy cơ overfitting. Đồng thời, thử nghiệm với các chiến lược như **early stopping** để dừng huấn luyện khi mô hình không cải thiện hiệu suất trên tập kiểm tra.

1. **Tăng cường dữ liệu huấn luyện**:

Mở rộng tập dữ liệu huấn luyện bằng cách sử dụng **dữ liệu lịch sử dài hơn** để giúp mô hình học được các mẫu dài hạn và tăng độ chính xác. Việc sử dụng dữ liệu từ nhiều nguồn cũng có thể giúp cải thiện kết quả.

1. **Áp dụng các mô hình kết hợp**:

Thử nghiệm kết hợp **LSTM với các mô hình học máy khác** như **Random Forests** hoặc **XGBoost** để tăng cường độ chính xác dự đoán. Sự kết hợp này có thể tận dụng các đặc trưng mạnh mẽ của các mô hình học máy khác trong việc xử lý các biến động không thể dự đoán.

1. **Dự đoán tỷ giá dài hạn và mô phỏng các kịch bản**:

Cải thiện mô hình để có thể **dự đoán tỷ giá USD/VND trong dài hạn** bằng cách kết hợp với các phương pháp mô phỏng kịch bản hoặc phân tích độ nhạy, giúp đánh giá các thay đổi tiềm ẩn của thị trường ngoại hối.

1. **Cập nhật và bảo trì mô hình thường xuyên**:

Để mô hình luôn duy trì hiệu quả trong bối cảnh thay đổi không ngừng của thị trường ngoại hối, cần thiết phải thực hiện **cập nhật và bảo trì mô hình** định kỳ bằng cách huấn luyện lại mô hình với dữ liệu mới hoặc tinh chỉnh các siêu tham số theo yêu cầu của thị trường.

1. **Mở rộng ứng dụng cho các loại tài sản khác**:

Sau khi đạt được kết quả khả quan với tỷ giá USD/VND, mô hình có thể được mở rộng và ứng dụng cho các **tỷ giá hối đoái khác**, hoặc thậm chí **dự đoán giá cổ phiếu, chỉ số chứng khoán** dựa trên các mô hình chuỗi thời gian tương tự.

# TÀI LIỆU THAM KHẢO

1. Nguyễn Minh Tân (2021), Phương pháp dự đoán tỷ giá hối đoái trong kinh tế vĩ mô, Nhà xuất bản Kinh tế, Hà Nội.
2. Lê Hữu Phước (2020), Mô hình học sâu và ứng dụng trong phân tích dữ liệu tài chính, NXB Đại học Quốc gia Hà Nội.
3. Trần Thanh Bình, Lê Minh Tuấn (2019), Hệ thống thông tin tài chính trong phân tích dữ liệu thị trường, Nhà xuất bản Khoa học và Kỹ thuật, TP. Hồ Chí Minh.
4. Nguyễn Văn A (2018), Dự đoán thị trường ngoại hối bằng học máy, NXB Tài chính, Hà Nội.
5. Zhang, Y., & Wang, J. (2020), Deep Learning in Foreign Exchange Forecasting: A Review, Springer, New York.
6. Kingma, D. P., & Ba, J. (2015), Adam: A Method for Stochastic Optimization, International Conference on Learning Representations (ICLR).
7. Huang, C., & Lee, J. (2021), Application of LSTM for Stock Price Prediction, Journal of Machine Learning Research, 12(3), 243-256.
8. TensorFlow Documentation (2023), TensorFlow: Deep Learning Framework, https://www.tensorflow.org/ (truy cập ngày 12 tháng 10, 2024).
9. Brownlee, J. (2021), Machine Learning Mastery with Python, Machine Learning Mastery, https://machinelearningmastery.com (truy cập ngày 05 tháng 11, 2024).
10. Yahoo Finance API (2023), Yahoo Finance Historical Data, https://www.yahoofinanceapi.com/ (truy cập ngày 10 tháng 11, 2024).