HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

NGUYỄN HỒNG SƠN

DỰ ĐOÁN TỶ GIÁ USD/VNĐ DÙNG HỌC MÁY

LUẬN VĂN THẠC SĨ KỸ THUẬT

HÀ NỘI – NĂM 2021

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

NGUYỄN HỒNG SƠN

DỰ ĐOÁN TỶ GIÁ USD/VNĐ DÙNG HỌC MÁY

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

LUẬN VĂN THẠC SĨ KỸ THUẬT

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. NGUYỄN VĂN THỦY

HÀ NỘI – NĂM 2021

LỜI CAM ĐOAN

Tôi cam đoan đề tài: ***"Dự đoán tỷ giá USD/VNĐ dùng học máy"*** là công trình nghiên cứu của riêng tôi dưới sự hướng dẫn của **TS. Nguyễn Văn Thủy.**

Những phân tích, kết luận, kết quả trong luận văn này đều là kết quả của tác giả, số liệu nêu ra là trung thực và chưa từng được công bố trong bất kỳ công trình nào khác.

Hà Nội, ngày tháng năm 2021

Tác giả

Nguyễn Hồng Sơn

LỜI CẢM ƠN

Lời đầu tiên cho tôi xin gửi lời cảm ơn chân thành đến các thầy, cô giáo của Học viện Công nghệ Bưu chính viễn thông đã tận tình chỉ bảo, hướng dẫn, giúp đỡ tôi trong suốt quá trình thực hiện luận văn này.

Tôi xin gửi lời cảm ơn chân thành đặc biệt tới thầy hướng dẫn khoa học **TS. Nguyễn Văn Thủy**, tận tình chỉ bảo và hướng dẫn, đưa ra định hướng đúng đắn giúp em hoàn thành được luận văn này.

Xin trân trọng gửi lời cảm ơn đến lãnh đạo cùng tập BIDV Chi nhánh Ba Đình và Viện Đào Tạo & Nghiên cứu BIDV, đã tạo điều kiện, cung cấp số liệu, thông tin chuyên ngành để hỗ trợ em hoàn thành luận văn.

Xin trân trọng cảm ơn các cảm ơn tập thể lớp Cao học hệ thống thông tin khoá 2019-2021 đã đồng hành, khích lệ và chia sẻ trong suốt quá trình học tập và làm luận văn.

Trong quá trình thực hiện luận văn, mặc dù bản thân đã cố gắng, chủ động sưu tầm dữ liệu, tài liệu, củng cố kiến thức… tuy nhiên không thể tránh khỏi những thiếu sót, hạn chế. Rất mong nhận được sự chỉ dạy, góp ý của các thầy, cô giáo và các bạn cùng lớp để luận văn được hoàn thiện hơn nữa và có tính ứng dụng cao hơn trong thực tiễn.

Xin trân trọng cảm ơn!

Hà Nội, ngày tháng năm 2021

Học viên

Nguyễn Hồng Sơn

MỤC LỤC

LỜI CẢM ƠN 2
DANH MỤC CÁC HÌNH 6
MỞ ĐẦU 1
1. Lý do chọn đề tài 1
2. Tổng quan về đề tài nghiên cứu 1
3. Mục tiêu nghiên cứu của đề tài 3
CHƯƠNG 1. BÀI TOÁN DỰ ĐOÁN TỶ GIÁ HỐI ĐOÁI 4
1.2. Tỷ giá hối đoái và các tác động ảnh hưởng đến tỷ giá hối đoái 6
1.2.1 Tỷ giá hối đoái 7
1.2.2 Cách thức phân loại tỷ giá hối đoái 7
1.2.3 Phương pháp xác định tỷ giá hối đoái 9
1.2.4 Các yếu tố ảnh hưởng đến tỷ giá hối đoái 9
1.3 Giới thiệu bài toán dự đoán tỷ giá hối đoái 11
1.3.1 Bài toán dự đoán tỷ giá 11
1.3.1.1 Khảo sát các nghiên cứu đã có 11
1.3.1.2 Xây dựng bài toán dự đoán tỷ giá USD-VND 14
1.3.2 Ứng dụng của bài toán 17
1.4 Kết luận chung chương 1 17
CHƯƠNG 2. ỨNG DỤNG CỦA HỌC MÁY CHO BÀI TOÁN DỰ ĐOÁN TỶ GIÁ 18
2.1 Tổng quan về học máy 18
2.2 Các công nghệ ứng dụng trong bài toán 24
2.2.1 Linear Regression 26
2.2.2 Random Forest 30
2.2.3 Neural Network 32
2.3 Kết luận chương 2 37
CHƯƠNG 3. THỬ NGHIỆM VÀ ĐÁNH GIÁ 38
3.1 Xây dựng bộ dữ liệu 38

3.1.1 Dữ liệu Tỷ giá USD/VNĐ 38
3.1.2 Dữ liệu giá vàng 39
3.1.3 Dữ liệu giá dầu 40
3.1.4 Dữ liệu chỉ số tiêu dùng 41
3.2 Cài đặt thuật toán học máy 42
3.2.1 Chuẩn hóa và import dữ liệu đầu vào 42
3.2.2 Tạo khung dữ liệu Data Frame 44
3.2.3 Xác định mục tiêu chu kỳ cần dự đoán 44
3.2.4 Thể hiện tính tương quan giữa các thuộc tính bằng biểu đổ cặp 44
3.2.5 Xây dựng Model (X,Y) 47
3.3 Thử nghiệm và đánh giá 49
3.3.1 Nội dung thử nghiệm 49
3.3.2 Kết quả thử nghiệm và đánh giá 50
3.4 Kết luận chương 3 56
KẾT LUẬN 57
IV. DANH MỤC TÀI LIỆU THAM KHẢO 58
BẢN CẢM ĐOAN 61

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng Việt
AI	Artificial Intelligence	Trí tuệ nhân tạo
BIDV	Bank for Investment and Development of Vietnam	Ngân hàng TMCP ĐT &PT Việt Nam
ERM	Exchange Rate Mechanism	Cơ chế tỷ giá
EU	European Union	Khối liên mình châu Âu
IMF	International Monetary Fund	Quỹ tiền tệ
LBFGS	Limited-memory BFGS	Bộ nhớ giới hạn Broyden Sĩ FletcherTHER GoldfarbTHER Shanno (BFGS)
ML	Machine Learning	Học máy
MLP	Multi-layer Perceptron	Perceptron nhiều lớp
USD	United States dollar	Đồng Đô la Mỹ
VND		Việt Nam Đồng
WB	World Bank	Ngân hàng thế giới

DANH MỤC CÁC HÌNH

Hình 1.1: Minh họa về dữ liệu nến trong 15 phút....12
Hình 1.2: Kết quả thực hiện thể hiện trên web 13
Hình 1.3: Minh họa mối tương quan giữa các đặc điểm....14
Hình 2.1: Mối liên hệ giữa AI, Machine Learning và Deep Learning 19
Hình 2.2: Cách thức hoạt động của Machine Learning 19
Hình 2.3: Ứng dụng Machine Learning 20
Hình 2.4: Nhận dạng đối tượng sử dụng học có giám sát 21
Hình 2.5: Phân loại nhóm đối tượng sử dụng học phi giám sát....22
Hình 2.6: Ví dụ học tăng cường 23
Hình 2.7: Dữ liệu đầu vào trong ví dụ sử dụng Linear Regression 29
Hình 2.8: Nghiệm của bài toán Linear Regression....29
Hình 2.9: Đồ thị mô tả dự đoán Linear Regression 29
Hình 2.10: Hình ảnh minh họa về Random Forest 31
Hình 2.11: Mạng netural network nhiều lớp ẩn....33
Hình 2.12: Ví dụ MLP với 2 hidden Layer....35
Hình 2.13: Các ký hiệu sử dụng trong MLP....36
Hình 2.14: Ví dụ MLPRegressor 37
Hình 3.1: Dữ liệu Tỷ giá USD/VNĐ 39
Hình 3.2: Dữ liệu giá vàng....40
Hình 3.3: Dữ liệu giá dầu....41
Hình 3.4: Dữ liệu chỉ số tiêu dùng CPI....41
Hình 3.5: Hàm chuẩn hóa dữ liệu 43
Hình 3.6: Cách thức gộp dữ liệu....43
Hình 3.7: Dữ liệu sau khi gộp 44
Hình 3.8: Tạo khung dữ liệu 44
Hình 3.9: Cặp bảng lưới của thuộc tính số....45
Hình 3.10: Tương quan giữa các thuộc tính....46

Hình 3.11: Phân chia dữ liệu X,Y trong Model 47
Hình 3.12: Dữ liệu tập Y 47
Hình 3.13: Dữ liệu tập X 48
Hình 3.14: Dữ liệu trong mô hình huấn luyện 48
Hình 3.15: Biểu đồ tần suất của Linear Regression 49
Hình 3.16: Biểu đồ tần suất của Netural Network 50
Hình 3.17: Biểu đồ thuộc tính quan trọng của Random Forest 50
Hình 3.18: Kết quả chạy của Linear Regression 52
Hình 3.19: Sơ đồ biểu diễn kết quả chạy của Linear Regression 52
Hình 3.20: Kết quả chạy của Random Forest 53
Hình 3.21: Sơ đồ biểu diễn kết quả chạy của Random Forest 53
Hình 3.22: Kết quả chạy của Netural Network 53
Hình 3.23: Sơ đồ biểu diễn kết quả chạy của Netural Network 54
Hình 3.24: Dữ liệu thực tế của tỷ giá USD/VNĐ 54
Hình 3.25: So sánh kết quả với thực tế trong 3 ngày 55
Hình 3.26: So sánh độ chênh lệch của kết quả đạt được với thực tế 55

MỞ ĐẦU

1. Lý do chọn đề tài

Hội nhập quốc tế đã và đang trở thành yêu cầu bức xúc, tất yếu đối với mỗi quốc gia trong điều kiện xu thế toàn cầu hóa hiện nay và Việt Nam đang vận hành nền kinh tế đi sâu vào hội nhập hóa quốc tế. Hiện nay, hoạt động thương mại quốc tế trong đó có hoạt động xuất nhập khẩu phát triển với một tốc độ chóng mặt. Với vai trò là nền huyết mạch kinh tế, hoạt động xuất nhập khẩu luôn được quốc gia quan tâm vì đây là con đường ngắn nhất góp phần tăng tích lũy của cải, giải quyết gánh nợ kinh tế. Hoạt động xuất nhập khẩu giữ vai trò vô cùng quan trọng và tỷ giá hối đoái được xem là công cụ hữu hiệu nhất để tối ưu hóa mục đích.

Tỷ giá hối đoái có ảnh hưởng sâu sắc và mạnh mẽ đến quan hệ kinh tế đối ngoại, tình trạng cán cân thanh toán, tăng trưởng kinh tế, lạm phát và thất nghiệp. Do vậy, việc dự đoán tỷ giá hối đoái mang lại giá trị to lớn cho các nhà quản lý trong nhiều lĩnh vực, đặc biệt là trong lĩnh vực tài chính ngân hàng.

Với mục đích đưa những tiến bộ công nghệ vào phục vụ cho lĩnh vực công việc chuyên môn, học viên xin chọn đề tài ***"Dự đoán tỷ giá USD/VNĐ dùng học máy"*** làm đề tài luận văn.

2. Tổng quan về đề tài nghiên cứu

Những năm gần đây, AI - Artificial Intelligence (Trí Tuệ Nhân Tạo), và cụ thể hơn là Machine Learning (Học Máy hoặc Máy Học) nổi lên như một bằng chứng của cuộc cách mạng công nghiệp lần thứ tư (1 - động cơ hơi nước, 2 - năng lượng điện, 3 - công nghệ thông tin). Trí Tuệ Nhân Tạo đã và đang len lỏi vào mọi lĩnh vực trong đời sống [10].

Trong lĩnh vực AI, một nhánh nghiên cứu về khả năng tự học của máy tính được gọi là học máy (*machine learning*). Hiện nay không có 1 định nghĩa chính thức nào về học máy nhưng có thể hiểu rằng nó là các kỹ thuật giúp cho máy tính có thể tự học mà không cần phải cài đặt các luật quyết định. Thông thường, một chương trình máy tính cần các quy tắc, luật lệ để có thể thực thi được một tác vụ nào đó như dán nhãn cho các email là thư rác nếu nội dung email có chứ từ khoá "quảng cáo".

Nhưng với học máy, các máy tính có thể tự động phân lại các thư rác thành mà không cần chỉ trước bất kỳ quy tắc nào cả. Có thể hiểu đơn giản là nó giúp cho máy tính có được cảm quan và suy nghĩ được như con người. Ở góc độ kỹ thuật thì học máy là phương pháp vẽ các đường thể hiện mối quan hệ của tập dữ liệu. Ví dụ như đường ngăn cách 2 loại dữ liệu cho nhãn khác nhau, đường thể hiện xu hướng của giá nhà phụ thuộc vào diện tích và vị trí hay các đường phân cụm dữ liệu.

Học sâu (tiếng Anh: deep learning) là một chi của ngành máy học dựa trên một tập hợp các thuật toán để 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 bằng cách khác bao gồm nhiều biến đổi phi tuyến.

Các giải thuật học máy được phân ra làm 2 loại chính là [7]:

- **Học có giám sát** (*Supervised Learning*): Là phương pháp sử dụng những dữ liệu đã được gán nhãn từ trước để suy luận ra quan hệ giữa đầu vào và đầu ra. Các dữ liệu này được gọi là dữ liệu huấn luyện và chúng là cặp các đầu vào-đầu ra. Học có giám sát sẽ xem xét các tập huấn luyện này, để từ đó có thể đưa ra dự đoán đầu ra cho 1 đầu vào mới chưa gặp bao giờ. Ví dụ dự đoán giá nhà, phân loại email.

Học phi giám sát (*Unsupervised Learning*): Khác với học có giám sát, học phi giám sát sử dụng những dữ liệu chưa được gán nhãn từ trước để suy luận. Phương pháp này thường được sử dụng để tìm cấu trúc của tập dữ liệu. Tuy vậy, không có phương pháp đánh giá được cấu trúc tìm ra được coi là đúng hay sai. Ví dụ như phân cụm dữ liệu, triết xuất thành phần chính của một chất nào đó.

Ngoài ra còn có 1 loại nữa là **học tăng cường** (*reinforcement learning*). Phương pháp học tăng cường tập trung vào việc làm sao để cho 1 tác tử Agent trong môi trường có thể hành động sao cho lấy được phần thưởng (reward) nhiều nhất có thể. Bản chất của việc học tăng cường là trial-and-error, nghĩa là thử đi thử lại và rút ra kinh nghiệm sau mỗi lần thử như vậy. Khác với học có giám sát, nó không có cặp dữ liệu gán nhãn trước làm đầu vào và cũng không có đánh giá các hành động là đúng hay sai. Ví dụ điển hình cho việc học tăng cường trong trò chơi như cờ vây, cờ vua,

cờ tướng hay StarCraft.

3. Mục tiêu nghiên cứu của đề tài

Mục tiêu nghiên cứu của luận văn là sử dụng một số thuật toán học máy để dự đoán tỷ giá ngoại tệ của đồng USD so với đồng VNĐ trong tương lai. Trong nghiên cứu này, mục tiêu dự báo tỷ giá ngoại tệ USD/VND theo ngày.

4. Đối tượng và phạm vi nghiên cứu

- **Đối tượng nghiên cứu:** Tỷ giá USD/VNĐ và phương pháp học máy

- **Phạm vi nghiên cứu:** Áp dụng cho tỷ giá USD/VNĐ trong lĩnh vực tài chính ngân hàng, sử dụng tỷ giá niêm yết (tỷ giá thị trường) phục vụ cho mục đích giao dịch mua bán ngoại tệ đối với khách hàng cá nhân.

5. Phương pháp nghiên cứu của đề tài

- ***Về mặt lý thuyết:*** Thu thập, khảo sát, phân tích các tài liệu liên quan đến bài toán dự báo tỷ giá hối đối. Nghiên cứu các thuật toán học máy để dự báo tỷ giá hối đoái trong tương lai

- ***Về mặt thực nghiệm:*** Thực nghiệm trên tập dữ liệu có sẵn, phân tích và đánh giá kết quả đạt được.

6. Bố cục luận văn

Luận văn được trình bày trong 3 chương:

- Chương 1 của luận văn sẽ trình bày tổng quan về bài toán dự đoán tỷ giá hối đoái.
- Chương 2 của luận văn tập trung nghiên cứu các thuật toán trong học máy ứng dụng vào bài toán dự đoán tỷ giá.
- Chương 3 của luận văn tập trung đưa ra cách thức xây dựng bộ dữ liệu, đồng thời đưa ra cài đặt thuật toán học máy cho việc dự đoán kết quả. Kết quả thử nghiệm sau đó được so sánh đối chiếu với giá trị thực tế để có nhận xét đánh giá độ phù hợp.

CHƯƠNG 1. BÀI TOÁN DỰ ĐOÁN TỶ GIÁ HỐI ĐOÁI

Chương 1 của luận văn sẽ trình bày tổng quan về bài toán dự đoán tỷ giá hối đoái, các khái niệm cơ bản, các thuật ngữ trong bài toán dự đoán tỷ giá hối đoái. Nội dung chính của chương 1 bao gồm:

1.1 Tìm hiểu về lịch sử hệ thống tiền tệ

Hệ thống tiền tệ quốc tế có lịch sử hơn 200 năm với những biến động cùng sự ra đời của các ngân hàng trung ương, thế chiến, cơ chế tỷ giá. Dưới đây là bài viết của Matthew Boesler của Business Insider sẽ giúp chúng ta có cái nhìn rõ hơn về lịch sử hệ thống tiền cũng như cơ chế tỷ giá hối đoái từ năm 1821 đến nay [24].

1.1.1 Bản vị vàng cổ điển

Từ năm 1821 đến 1914, hầu hết các tiền tệ trên thế giới đều được quy đổi sang vàng. Anh là quốc gia đầu tiên áp dụng cơ chế này vào năm 1821, tiếp đến là các nước khác vào những năm 1870. Hầu hết các nước thuộc địa sẽ quy đổi tiền tệ của mình theo vàng.

Kết quả là kinh tế toàn cầu được kết nối thông qua việc sử dụng chung cơ chế quy đổi vàng, tiền. Bảng Anh được coi là quan trọng nhất trong hệ thống này bởi Anh là siêu cường tại thời điểm đó, là nước đầu tiên áp dụng cơ chế quy đổi và cũng là nước kiên trì tuân thủ mức quy đổi 0,25 oz/bảng.

1.1.2 Thời kỳ thả nổi

Thế chiến thứ I giai đoạn 1915-1925 đã làm sụp đổ hệ thống bản vị vàng. Trong khi gần như cả thế giới ngừng quy đổi tiền tệ theo vàng thì Mỹ vẫn theo đuổi cơ chế này thêm vài năm nữa.

Chính điều này đã giúp nâng vị thế của đồng USD với vai trò là một đồng tiền dự trữ quốc tế. Giai đoạn thả nổi cơ chế tỷ giá đầu thế kỷ 20 – khi mà các chính phủ có thể tự do can thiệp thị trường – bị coi là giai đoạn hỗn loạn và bất ổn. Do đó, sau chiến tranh, các nước dự định khôi phục lại cơ chế trước Thế chiến thứ I.

1.1.3 Hệ thống bản vị vàng giữa 2 cuộc thế chiến

Sau Thế chiến I, việc khôi phục kinh tế trở thành nhu cầu cấp thiết cho các nước châu Âu, do đó các nước đã tiến tới thỏa thuận lập một trật tự mới trong quan hệ thương mại, tín dụng và tiền tệ quốc tế.

Tại hội nghị Genoa năm 1922, các nước thừa nhận vai trò của đồng bảng Anh là đồng tiền thanh toán và dự trữ quốc tế. Do đó thực tế, chế độ tiền tệ lúc này là chế độ bản vị bảng Anh.

1.1.4 Hệ thống thả nổi trước hiệp ước Bretton Woods

Các chính phủ châu Âu lần lượt phá vỡ cam kết tuân thủ hệ thống bản vị vàng khiến hệ thống tiền tệ quốc tế tiếp tục theo cơ chế thả nổi, duy chỉ có USD vẫn theo cơ chế neo tỷ giá với vàng. Tuy nhiên, hệ thống này suy yếu dần trong giai đoạn Đại suy thoái những năm 1930.

Đến năm 1934, tổng thống Mỹ lúc đó là Roosevelt đã ra sắc lệnh cấm dự trữ vàng và giảm tỷ lệ quy đổi còn khoảng từ 20 USD/oz -35 USD/oz Ngoài ra, sắc lệnh cũng yêu cầu người dân không xuất khẩu vàng, và chuyển vai trò dự trữ vàng sang cho Cục dự trữ liên bang Mỹ (Fed).

1.1.5 Hiệp định Bretton Woods về neo tỷ giá

Năm 1944, tại Bretton Woods, New Hampshire, 730 đại biểu đến từ 44 quốc gia đã gặp nhau để để xây dựng hệ thống tài chính thế giới sau chiến tranh, tránh nguy cơ tái diễn khủng hoảng kinh tế.

Một hệ thống tài chính được thành lập gọi là Bretton Woods - bao gồm Quỹ Tiền tệ Quốc tế (IMF), Ngân hàng Thế giới (WB) và chế độ tỷ giá hối đoái cố định được xây dựng quanh đồng USD gắn với vàng.

Tại thời điểm đó, hơn một nửa tiềm năng sản xuất của thế giới do Mỹ phụ trách và giữ gần như toàn bộ lượng vàng của thế giới. Bởi vậy, các nhà lãnh đạo quyết định gắn các đồng tiền thế giới với đồng USD với quy đổi theo vàng là 35 USD/oz.

1.1.6 Hiệp định Smithsonian

Tháng 12/1971, hiệp định Smithsonian ra đời, đòi hỏi USD giảm giá khoảng 8% so với các đồng tiền khác. Biên độ dao động giá trị các đồng tiền được nới rộng

đến 2,5% của tỷ giá ấn định. Tháng 3/1973, hiệp định này chấm dứt, kết thúc kỳ nguyên của Bretton Woods.

1.1.7 Thả nổi ở phương Tây và cơ chế neo tỷ giá linh hoạt ở các nước đang phát triển

Sau khi hiệp định Smithsonian chấm dứt, USD giảm mạnh so với vàng. Hàng loạt các nước đang phát triển bỏ cơ chế neo tỷ giá theo USD, trở về cơ chế thả nổi như đầu thế kỷ 20 và chỉ có một số ít nước, đặc biệt ở châu Á vẫn neo tỷ giá theo USD.

1.1.8 Cơ chế tỷ giá hối đoái châu Âu

Vài năm áp dụng cơ chế thả nổi, đến cuối những năm 1970, Cộng đồng châu Âu đã thiết lập một hệ thống hợp tác về tỷ giá hay còn gọi là ERM. Theo đó các ngân hàng trung ương thành viên có thể can thiệp thị trường để duy trì biên độ tỷ giá 2,25% giữa đồng tiền của nước họ với một đồng tiền khác. Đây có thể gọi là cơ chế bán neo tỷ giá (semi-peg).

Đến tận năm 1990, Anh mới tham gia vào hệ thống này, nhưng cũng rời hệ thống 2 năm sau đó khi chính phủ nước này không chấp nhận việc ấn định biên độ theo ERM. Cũng trong giai đoạn này, tỷ phú George Soros được cho là đã thu về 1 tỷ USD nhờ bán khống bảng Anh. Khoảng 10 năm sau đó, đồng euro ra đời và các tiền tệ khác được quy đổi theo euro khiến euro trở thành tiền tệ chính thức khi thị trường chứng khoán chính ở Italia, Đức và Pháp được định giá bằng đồng tiền này.

Trong khi đó, USD vẫn đóng vai trò là đồng tiền dự trữ toàn cầu, có tỷ giá thả nổi so với các đồng tiền chính và vàng. Hầu hết các nước Đông Nam Á vẫn duy trì cơ chế tỷ giá cố định hoặc linh hoạt với USD cho đến trước khủng hoảng tài chính khu vực năm 1997.

1.2. Tỷ giá hối đoái và các tác động ảnh hưởng đến tỷ giá hối đoái

Lịch sử của hệ thống tiền tệ đã cho ta cái nhìn khái quát hơn về cơ chế tỷ giá hối đoái. Tuy nhiên, chúng ta cũng cần làm rõ các vấn đề: thế nào là tỷ giá hối đoái, cách phân loại tỷ giá hối đoái, phương pháp xác định tỷ giá hối đoái, có những yếu tố nào ảnh hưởng đến tỷ giá hối đoái.... ?

1.2.1 Tỷ giá hối đoái

Tỷ giá hối đoái có cách gọi khác là tỷ giá trao đổi ngoại tệ. Được hiểu là tỷ giá của một đồng tiền này có thể được quy đổi cho một đồng tiền khác, tỷ giá giữa 2 loại tiền tệ, là số lượng đơn vị tiền tệ cần thiết để mua một đơn vị ngoại tệ. Theo Luật Ngân hàng Nhà nước Việt Nam (Số: 06/1997/QH10 ngày 12 tháng 12 năm 1997), tỷ giá hối đoái là tỷ lệ giá trị của đồng Việt Nam với giá trị đồng tiền nước ngoài. Tỷ giá này được hình thành dựa trên cơ sở cung cầu ngoại tệ, dưới sự điều tiết của Nhà Nước, do Ngân hàng Nhà nước Việt Nam xác định [14].

Trong ngành tài chính ngân hàng, tỷ giá hối đoái phản ánh mối quan hệ giá trị đồng tiền của hai nước với nhau. Ví dụ tỷ giá bán ra của Ngân hàng Ngoại thương Việt Nam ngày 21/11/2019 1 USD = 23.260VNĐ. Đây chính là tỷ giá hối đoái. Tỷ giá hối đoái được xem là một loại giá cả đặc biệt, là giá trị của tiền chứ không phải giá trị của hàng hóa. Cách đọc tỷ giá hối đoái: Đồng tiền đứng trước được hiểu là đồng tiền yết giá, đồng tiền đứng thứ hai gọi là đồng tiền định giá. Trong ví dụ về tỷ giá hối đoái trên thì USD là đồng tiền yết giá còn VNĐ là đồng tiền định giá.

Tỷ giá hối đoái còn được xem là quan hệ so sánh tiền tệ của các nước theo tiêu chuẩn nào đó. Trong chế độ bản vị vàng thì tiền tệ trong lưu thông hoạt động kinh doanh là tiền đúc bằng vàng và giấy và nó được đổi ra vàng căn cứ vào hàm lượng vàng. Vì thế, tỷ giá hối đoái có thể hiểu là mối quan hệ so sánh giữa tiền vàng của hai nước. Còn trong chế độ tiền giấy thì tiền đúc không còn được sử dụng nên ngang giá vàng không còn là cơ sở hình thành tỷ giá hối đoái. Theo đó thì việc so sánh các đồng tiền khác nhau được thực hiện bằng hình thức so sánh mức mua của hai tiền tệ với nhau.

1.2.2 Cách thức phân loại tỷ giá hối đoái [29]

Đối với thị trường hối đoái hiện nay, có rất nhiều loại tỷ giá khác nhau. Có một số cách phân chia tỷ giá hối đoái như sau:

a) Căn cứ dựa trên giá trị tỷ giá:

Dựa vào giá trị tỷ giá có thể chia thành 2 loại:

+) *Tỷ giá hối đoái thực*: Là tỷ giá mà trong đó có tác động của lạm phát và sức mua trong một cặp tiền tệ. Nó phản ánh giá cả hàng hóa tương quan có thể bán ra nước ngoài và hàng tiêu thụ trong nước. Tỷ giá này đại diện cho khả năng cạnh tranh trên trường quốc tế của nước đó.

+) *Tỷ giá hối đoái danh nghĩa*: Là tỷ giá của một loại tiền tệ theo giá hiện tại, không tính đến ảnh hưởng của lạm phát.

b) Căn cứ vào phương thức chuyển ngoại hối:

Dựa vào khái niệm Tỷ giá hối đoái là gì và căn cứ vào phương thức chuyển ngoại hối, chúng ta có thể chia làm 2 loại:

+) *Tỷ giá thư hối*: Là tỷ giá có hình thức chuyển ngoại hối bằng thư. Tỷ giá điện hối có xu hướng cao hơn tỷ giá thư hối.

+) *Tỷ giá điện hối*: Là tỷ giá có hình thức được niêm yết tại ngân hàng. Tỷ giá điện hối là tỷ giá chuyển ngoại hối bằng điện. Tỷ giá điện hối là tỷ giá cơ sở để xác định các loại tỷ giá khác.

c) Căn cứ vào thời điểm giao dịch ngoại hối:

Có thể chia ra thành 2 loại như sau:

+) *Tỷ giá mua:* Là tỷ giá mua ngoại hối vào của ngân hàng.

+) *Tỷ giá bán*: Là tỷ giá bán ngoại hối ra của ngân hàng.

d) Căn cứ vào kỳ hạn thanh toán:

Dựa trên kỳ hạn thanh toán, phân chia tỷ giá hối đoái thành:

+) *Tỷ giá giao dịch kỳ hạn (FORWARDS):* Là tỷ giá do tổ chức tín dụng tính toán và thỏa thuận với nhau nhưng phải đảm bảo trong biên độ quy định về tỷ giá kỳ hạn hiện hành của Ngân hàng Nhà nước tại thời điểm ký hợp đồng.

+) *Tỷ giá giao ngay (SPOT)*: Là tỷ giá do tổ chức tín dụng yết giá tại thời điểm giao dịch hoặc do hai bên thỏa thuận trong đó phải đảm bảo biểu độ do ngân hàng nhà nước quy định. Việc thanh toán giữa các bên phải được thực hiện trong vòng hai ngày làm việc tiếp theo, tính từ thời điểm sau ngày cam kết mua hoặc bán.

e) Căn cứ vào đối tượng xác định tỷ giá:

Dựa trên đối tượng xác định tỷ giá và những thông tin khái niệm "Tỷ giá hối đoái là gì" chúng ta có thể phân chia thành:

+) *Tỷ giá thị trường:* Tỷ giá được hình thành dựa trên quan hệ cung cầu của thị trường hối đoái.

+) *Tỷ giá chính thức*: Là tỷ giá do Ngân hàng trung ương của nước đó xác định. Các ngân hàng thương mại và các tổ chức tín dụng sẽ ấn định tỷ giá mua bán ngoại tệ giao ngay, có kỳ hạn, hoán đổi dựa trên tỷ giá chính thức.

1.2.3 Phương pháp xác định tỷ giá hối đoái [7]

Bản chất tỷ giá là giá cả của một đơn vị tiền tệ và phụ thuộc vào cung cầu về đồng tiền đó trên thị trường. Do vậy, tỷ giá sẽ thay đổi nếu cung cầu thay đổi. Có nhiều phương pháp xác định tỷ giá hối đoái khác nhau tuỳ thuộc vào mục đích kinh doanh, sự phát triển của thị trường tiền tệ và thị trường hàng hoá, dịch vụ trên thế giới. Việc xác định tỷ giá hối đoái giúp các nhà kinh doanh có thể xây dựng phương án kinh doanh sao cho có lợi nhất.

+) *Xác định tỷ giá hối đoái trên cơ sở ngang giá vàng (Gold parity)*: Đây là phương pháp so sánh hàm lượng vàng giữa hai đồng tiền với nhau.

+) *Xác định tỷ giá hối đoái trên cơ sở cân bằng sức mua (Purchasing Power Parity)*: Phương pháp này dựa trên cơ sở so sánh sức mua giữa hai đồng tiền, dùng để so sánh giá cả hàng hoá, dịch vụ, xây dựng phương án kinh doanh xuất nhập khẩu và thực hiện các nghiệp vụ hải quan.

1.2.4 Các yếu tố ảnh hưởng đến tỷ giá hối đoái [7]

Tỷ giá hối đoái có tính tương đối, và được thể hiện như sự so sánh giữa đồng tiền của hai quốc gia. Sau đây là một số yếu tố ảnh hướng đến tỷ giá hối đoái:

a) Yếu tố thương mại:

- Tình hình tăng trưởng kinh tế: Khi tốc độ tăng giá của sản phẩm xuất khẩu cao hơn tốc độ tăng giá sản phẩm nhập khẩu thì tỷ lệ trao đổi thương mại tăng kéo theo giá trị đồng nội tệ tăng và tỷ giá giảm. Ngược lại tốc độ tăng nhập khẩu cao hơn tốc độ tăng xuất khẩu thì cán cân thương mại giảm khiến cho tỷ giá hối đoái tăng.

- Cán cân thanh toán: Cán cân thanh toán quốc tế cao thì đồng ngoại tệ tăng và nội tệ giảm khiến tỷ giá hối đoái tăng.

b) Yếu tố lạm phát:

Vấn đề lạm phát trong nước là một trong những yếu tố quan trọng tác động đến hoạt động thương mại quốc tế và ảnh hưởng trực tiếp đến cung cầu ngoại tệ làm thay đổi tỷ giá. Đây cũng là yếu tố để trả lời cho câu hỏi yếu tố ảnh hưởng đến tỷ giá hối đoái là gì?

Nếu nội địa có tỷ lệ lạm phát cao hơn so với nước ngoài thì tỷ giá hối đoái sẽ tăng và giá trị nội tệ sẽ giảm. Ngược lại, với nội địa có tỷ lệ lạm phát thấp hơn so với nước ngoài thì tỷ giá hối đoái sẽ giảm và giá trị nội tệ sẽ tăng.

Ví dụ: Nếu tình hình trong nước (Ấn Độ) có tỷ lệ lạm phát cao hơn quốc gia nước ngoài (Mỹ). Khi đó, người tiêu dùng Ấn Độ sẽ có xu hướng chọn lựa hàng hoá Mỹ hơn do giá thành chi trả cho hàng hoá sẽ rẻ hơn và thị trường sẽ nhập khẩu hàng Mỹ tăng làm cầu đồng ngoại tệ (đô la Mỹ) tăng. Còn ở Mỹ, người dân sẽ hạn chế sử dụng hàng hoá từ Ấn Độ do giá cao và nhập khẩu giảm khiến cung ngoại tệ (đô la Mỹ) giảm

c) Yếu tố thu nhập:

Nếu đã biết tỷ giá hối đoái là gì thì có thể nói thu nhập của mỗi quốc giá cũng là yếu tố tác động đáng kể đến tỷ giá hối đoái.

- Tác động trực tiếp: nếu thu nhập của quốc gia đó tăng thì người dân sẽ có xu hướng muốn dùng hàng nhập khẩu nhiều hơn từ đó làm cầu ngoại tệ tăng làm tỷ giá tăng.

- Tác động gián tiếp: thu nhập cao thì người dân sẽ tăng mức chi tiêu trong nước làm cho tỷ lệ lạm phát cao làm tỷ giá tăng.

Ngược lại khi quốc gia có thu nhập giảm thì sẽ giảm cầu ngoại tệ dẫn đến giảm tỷ giá hối đoái

d) Yếu tố lãi suất:

Lãi suất có một phần ảnh hưởng đến các hoạt động đầu tư chứng khoán ở nước ngoài, từ đó ảnh hưởng trực tiếp đến tỷ giá hối đoái.

Ví dụ: Khi đất nước A có lãi suất thấp hơn so với các nước ngoài như Trung Quốc. Thì nhà đầu tư nước A sẽ có xu hướng đầu tư vào thị trường Trung Quốc hoặc gửi tiền tiết kiệm vào các ngân hàng nước ngoài đó. Như vậy sẽ giúp họ có thêm khoản lợi nhuận lớn hơn so với đầu tư vào thị trường trong nước. Khi đó, ngoại tệ Trung Quốc sẽ tăng lên và cung về ngoại tệ của nước A sẽ giảm.

Còn khi nội địa có lãi suất cao hơn nước ngoài thì tài chính nội địa hấp dẫn tỷ giá hối đoái giảm còn giá trị nội tệ sẽ tăng.

1.3 Giới thiệu bài toán dự đoán tỷ giá hối đoái

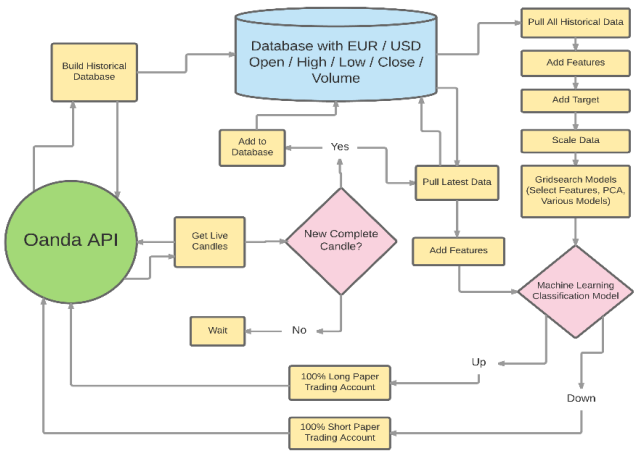
1.3.1 Bài toán dự đoán tỷ giá

Tỷ giá hối đoái có ý nghĩa quan trọng đối với doanh nghiệp cũng như đối với nhà nước. Dự đoán được kết quả sẽ mang lại những ý nghĩa thiết thực to lớn. Xây dựng bài toán dự đoán tỷ giá sẽ dựa trên tổng hợp thống kê của nhiều tổ hợp dữ liệu.

1.3.1.1 Khảo sát các nghiên cứu đã có

Hiện tại ở Việt Nam chưa có bài báo nghiên cứu về vấn đề sử dụng AI cho dự đoán tỷ giá hối đoái. Tuy nhiên, trên thế giới đã có một số tác giả nghiên cứu về vấn đề này. Ví dụ như tác giả Edeane có bài viết dự đoán tỷ giá EUR/USD trên Github [6] hay tác giả Robert Ritz có bài viết dự đoán tỷ giá hối đoái USD-MNT để phục vụ trong việc xuất nhập khẩu ở Mongolia [27].

Trong bài viết của tác giả Edeane, tác giả đã sử dụng sơ đồ dữ liệu như sau:



Sơ đồ luồng dữ liệu trên được thể hiện như sau: Từ cơ sở dữ liệu PostgreSQL của các giá (EUR/USD) gồm các nến thể hiện giá mở cửa, giá cao, giá thấp, giá đóng

cửa, khối lượng, tác giả dựa trên toàn bộ dữ liệu đấy để thêm các đặc điểm thuộc tính. Bước tiếp theo, tác giả thêm các hướng mục tiêu (target) trước khi thực hiện co giãn chuẩn hóa dữ liệu (Scaling Data). Mục đích của tác giả là để co giãn các đặc trưng về cùng một thước đo cụ thể, nhờ đó mà các hàm mục tiêu có thể hoạt động được đúng. Tiếp đến, tác giả sử dụng kỹ thuật phân tích Analytics Indicator để lựa chọn đặc điểm, tính toán tầm quan trọng của các tính năng, PCA để chính quy hóa tập mô hình dữ liệu. Sau đó, tác giả sử dụng mô hình học máy để phân tích, tính toán cho dự đoán thời gian dài hạn (long term), ngắn hạn (short term) của việc tăng/ giảm của tỷ giá Eur/USD. Thông qua API Oanda, tác giả lấy giá trị của nến ở thời điểm hiện tại, nếu đây là một nến mới hoàn chỉnh thì sẽ cập nhật dữ liệu vào Database, còn nếu không thì sẽ chờ đợi, tiếp tục 1 chu trình như vậy.

Về dữ liệu lịch sử:

Tác giả sử dụng API Oanda để tải các giá EUR / USD (ở đây là các nến) lịch sử xuống cơ sở dữ liệu PostgreSQL. Nến (candle) là giá mở, giá cao, giá thấp và giá đóng cửa trong một khoảng thời gian. Giá trung bình (giữa giá đặt mua / giá bán) đã được sử dụng. Khối lượng lượng cứ sau 5 giây, 10 giây, 15 giây, v.v. từ năm 2005 đến nay.

time	volume	open	high	low	close	complete
6:45:00	473	1.346250	1.348050	1.345950	1.348050	True
7:00:00	481	1.347950	1.348250	1.347350	1.348150	True
7:15:00	303	1.348150	1.348350	1.347300	1.347900	True
7:30:00	290	1.348000	1.350850	1.348000	1.350750	True
7:45:00	373	1.350650	1.353250	1.350250	1.352800	True
8:00:00	290	1.352800	1.354700	1.352500	1.352500	True
8:15:00	219	1.352400	1.353000	1.351250	1.351570	True

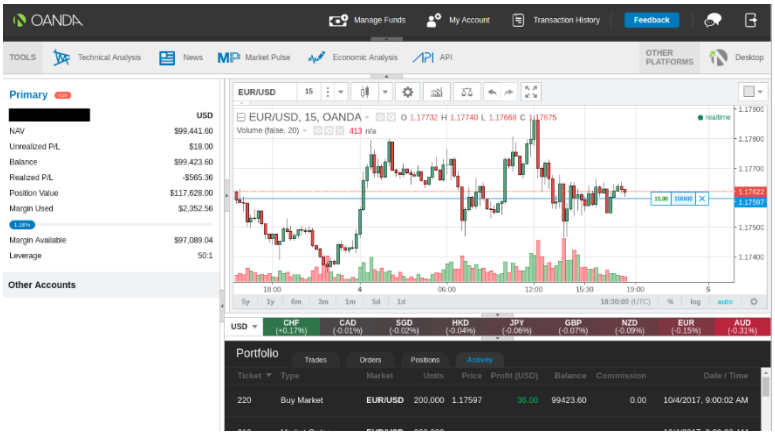
Hình 1.1: Minh họa về dữ liệu nến trong 15 phút

Về mô hình:

Tác giả sử dụng mô hình Logistic regression, boosted trees để thực hiện tính toán dự đoán. Ở đây, tác giả còn sử dụng máy chủ Amazon Web Service EC2 mạnh mẽ để tính toán song song việc tối ưu hóa tham số lưới tìm kiếm.

Về kết quả:

Tác giả có 1 số kết quả thực hiện như sau:



Hình 1.2: Kết quả thực hiện thể hiện trên web

Không giống với Edeane, tác giả Robert Ritz viết về dự đoán tỷ giá USD-MNT với chu kỳ khác. Ở trong nghiên cứu này, tác giả dự đoán tỷ giá theo thời gian 3 tháng, 6 tháng và 12 tháng trong tương lai.

Về dữ liệu:

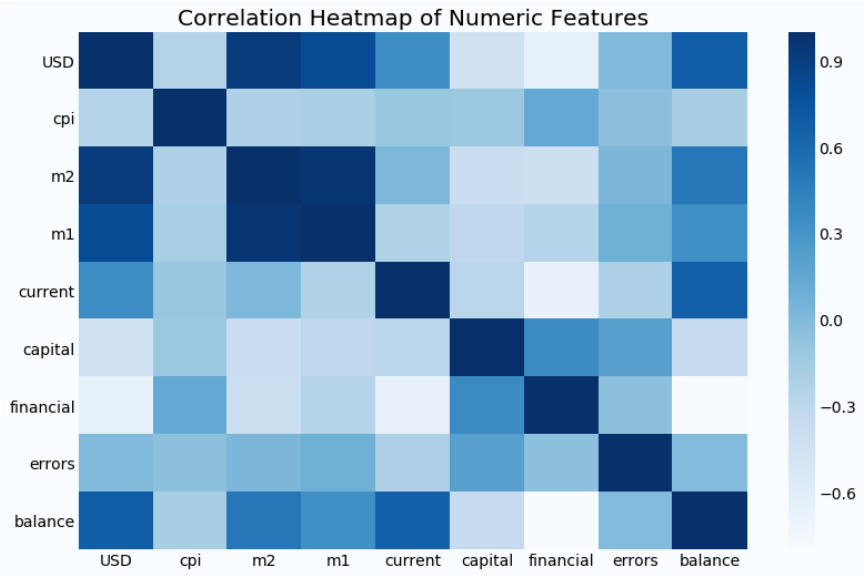
Tác giả sử dụng đặc điểm dữ liệu như sau:

- USD — USD-MNT Tỷ giá giữa đồng USD với đồng Mongolia
- CPI — Mức độ tiêu dùng. Đây là mức thay đổi CPI so với tháng trước.
- m2 cung tiền
- m1 cung tiền
- current: tài khoản vãng lai từ cán cân thanh toán
- capital: tài khoản vốn từ cán cân thanh toán
- financial: tài khoản tài chính từ cán cân thanh toán
- errors: lỗi và thiếu sót từ cán cân thanh toán
- balance: cán cân xuất khẩu và nhập khẩu ở Mông Cổ

Về mô hình:

Tác giả sử dụng nhiều mô hình để thực hiện và so sánh đối chiếu như Linear Regression, Random Forest và Extremely Random Trees.

Về kết quả thu được như sau:



Hình 1.3: Minh họa mối tương quan giữa các dặc điểm

Trong hình minh họa 1.3 thể hiện mối tương quan giữa các đặc điểm, trong đó màu sắc càng đậm thì sẽ thể hiện sự tương quan càng lớn, ngược lại với màu sắc càng nhạt thì độ tương quan ảnh hưởng càng ít. Dựa vào hình minh họa 1.3, ta thấy được quan hệ tương quan giữa các cặp được thể hiện khá rõ, ví dụ đối với USD, thì đặc điểm thuộc tính m2 là có quan tương quan tốt nhất (màu sắc đậm nhất), sau đó lần lượt là m1, balance, current, errors, cpi, capital và financial.

Có thể thấy mỗi tác giả đều có cách riêng của mình trong việc dự đoán tỷ giả tương lai. Tuy nhiên cả hai nghiên cứu này đều có đặc điểm chung đó là lựa chọn loại dữ liệu mà đặc điểm có thể có tương quan ảnh hưởng lẫn nhau, tiếp theo đó là kỳ tính toán và cuối cùng là cùng chọn 1 số loại mô hình AI để thực hiện. Dựa trên những điểm chung ở trên, học viên sẽ xây dựng bài toán dự đoán tỷ giá ở phần tiếp theo.

1.3.1.2 Xây dựng bài toán dự đoán tỷ giá USD-VND

Như đã nêu trên, có nhiều yếu tố ảnh hưởng đến tỷ giá hối đoái như yếu tố thương mại, yếu tố lạm phát, yếu tố thu nhập, yếu tố lãi suất. Dựa trên những bài

nghiên cứu của những người đi trước, đồng thời tìm hiểu và chọn lọc dữ liệu, học viên xác định có 1 số thuộc tính làm cơ sở để xây dựng bài toán. Các thuộc tính này đều có tính tương quan, ảnh hưởng đến thay đổi của dự đoán tỷ giá là :

+) Thuộc tính giá vàng

+) Thuộc tính giá dầu thô

+) Thuộc tính CPI

Trong đó thuộc tính giá vàng và giá dầu thô nằm trong yếu tố thương mại, còn thuộc tính CPI nằm trong yếu tố lạm phát. Dưới đây là một số dẫn chứng, chứng minh các thuộc tính giá vàng, giá dầu hay chỉ số tiêu dùng CPI có ảnh hưởng đến thị trường tiền tệ, cũng như ảnh hưởng đến tỷ giá.

Trong bài viết "**Phía sau việc giá vàng đắt chưa từng có**" trên tapchitaichinh.vn của tác giả Lan Hương/vtc.vn có viết "Giá vàng tăng dồn dập qua từng phiên và đang đắt chưa từng có, nhiều chuyên gia nhận định giá sẽ còn tăng tiếp, vậy phía sau đó, chuyện gì đang diễn ra?"[23]. Theo dự đoán của đồng tác giả TS. Nguyễn Trí Hiếu, với tình trạng giá vàng thế giới cao ở mức kỷ lục sẽ kéo theo tình trạng giá vàng trong nước gặp nhiều biến động. Giá vàng đạt mốc kỷ lục 55 triệu đồng/lượng sẽ có nhiều khả năng sẽ cán mốc, tuy nhiên trước khi đạt được mốc đó thì thị trường vàng cũng sẽ phải chứng kiến rất nhiều biến động, thậm chí có lúc xuống rất thấp [23].

Hay trong đoạn "Trong thời gian qua, bên cạnh yếu tố dịch bệnh, thế giới còn chứng kiến các vấn đề về chính trị như chiến tranh thương mại Mỹ - Trung, căng thẳng trên bán đảo Triều Tiên, Afghanistan, các nước Trung Đông. Ở châu Âu, Anh rút khỏi EU cũng gây ảnh hưởng không nhỏ đến nền kinh tế thế giới. Trong khi đó, nội bộ chính trị Mỹ hết sức bất ổn trong cuộc đua bầu cử Tổng thống giữa các Đảng phái. Hàng loạt vấn đề trên nhiều khả năng vẫn còn tiếp tục thêm thời gian nữa, thậm chí có thể diễn biến phức tạp hơn. Và khi mà những dự đoán về khủng hoảng, suy thoái, lạm phát… vẫn còn phủ bóng đen lớn trên nền kinh tế toàn cầu thì giá vàng vẫn còn nhận được sự hỗ trợ tăng giá. Phía sau đà tăng mạnh của giá vàng, giới chuyên gia dự đoán, kinh tế thế giới cần một lượng tiền khổng lồ để giải quyết các vấn đề bất

ổn và cái giá phải trả là lạm phát tăng vọt. Nhất là trong bối cảnh đồng tiền định giá vàng trên thị trường tài chính thế giới là USD liên tục suy yếu." Theo TS.Bùi Trinh cũng cho rằng "Nếu giá vàng tăng cao quá mức sẽ ảnh hưởng đến lạm phát, ảnh hưởng dây chuyền đến các mặt hàng khác*"* [23].

Để xác định giá dầu thô ảnh hưởng đến thị trường tiền tệ cũng như ảnh hưởng đến tỷ giá xuất nhập khẩu, bài viết "**Giá dầu giảm sâu tác động thế nào đến kinh tế Việt Nam?**" của TS. Cấn Văn Lực và Nhóm tác giả Viện Đào tạo và Nghiên cứu BIDV trên trang cafef.vn đã chứng minh giá dầu thô là một trong yếu tố thương mại.

Trong bài viết "Báo cáo cho biết, từ đầu năm 2020 đến nay (31/3), giá dầu thế giới đã giảm trên 60% đã có nhiều tác động đối với nền kinh tế toàn cầu cả tích cực lẫn tiêu cực. Đối với Việt Nam, giá dầu thế giới giảm góp phần giảm chi phí sản xuất cho doanh nghiệp và người tiêu dùng, qua đó kích thích đầu tư và tiêu dùng, đồng thời tiết kiệm được lượng ngoại tệ nhập khẩu xăng dầu, hỗ trợ kiểm soát lạm phát, ổn định kinh tế vĩ mô. Tuy nhiên, giá dầu giảm cũng ảnh hưởng đến nguồn thu ngân sách, hoạt động đầu tư, khai thác và lọc hóa dầu. Trong năm 2020, giá dầu được dự báo tiếp tục ở mức thấp sẽ có những tác động nhất định đối với kinh tế Việt Nam" [30].

Tương tự, trong bài viết "Chỉ số CPI và diễn biến thị trường tiền tệ: Mục tiêu kép cần bảo vệ", tác giả TS. Nguyễn Thị Kim Oanh có viết "Diễn biến chỉ số giá tiêu dùng trong 7 tháng đầu năm có xu hướng tăng nhẹ, đến tháng 7 CPI so với cùng kỳ chỉ tăng 2,39%, lạm phát cơ bản nhìn chung ổn định, đến tháng 7 ở mức 1,85% thấp hơn mức tăng của tháng 6 (1,88%).Giá USD so với cùng kỳ có xu hướng giảm mạnh, đến tháng 7 chỉ số giá USD chỉ tăng 2,21% so với cùng kỳ và giảm so với tháng 12/2015. Giá vàng đã có xu hướng giảm dần và đi vào thế ổn định" [31].

Đó chính là cơ sở để học viên lựa chọn các thuộc tính trên làm dữ liệu nghiên cứu trong việc xây dựng bài toán dự đoán tỷ giá USD-VND. Cụ thể hơn, với bài toán dự đoán tỷ giả USD/VNĐ, dữ liệu đầu vào sẽ gồm dữ liệu tỷ giá bán ra của ngân hàng đối với ngoại tệ USD. Dữ liệu ở đây lấy vào thời điểm cuối ngày, được chốt trước hết

phiên giao dịch trong ngân hàng. Dữ liệu sẽ nằm trong khoảng 04/05/2015 đến ngày 04/05/2020.

1.3.2 Ứng dụng của bài toán

- Kết quả dự đoán tỷ giá USD/VNĐ sẽ giúp nhà đầu tư, doanh nghiệp, bộ phận quản lý thị trường xuất nhập khẩu, ban quản lý ngoại tệ, khối ngân hàng tăng khả năng chính xác trong việc đưa ra quyết định đối với các vấn đề liên quan đến kiểm soát nền kinh tế như hoạt động xuất nhập khẩu, bình ổn tỷ lệ lạm phát hay thúc đẩy tăng trưởng kinh tế.

1.4 Kết luận chung chương 1

Trong chương 1, luận văn đã nghiên cứu tổng quan chung về tiền tệ, lịch sử hệ thống tiền tệ, tỷ giá hối đoái, cách phân loại tỷ giá hối đoái, phương pháp xác định tỷ giá hối đoái, và những yếu tố nào ảnh hưởng đến tỷ giá hối đoái và các nghiên cứu đã có. Qua đó, nghiên cứu cơ sở trên làm tiền đề để xây dựng các thông tin đầu vào cho bài toán dự đoán tỷ giá USD/VNĐ. Chương tiếp theo sẽ trình bày các thuật toán trong học máy và cách thức ứng dụng vào bài toán một cách hiệu quả.

CHƯƠNG 2. ỨNG DỤNG CỦA HỌC MÁY CHO BÀI TOÁN DỰ ĐOÁN TỶ GIÁ

Chương 2 của luận văn tập trung trình bày giới thiệu về học máy, các công nghệ ứng dụng trong bài toán dự đoán tỷ giá.

2.1 Tổng quan về học máy

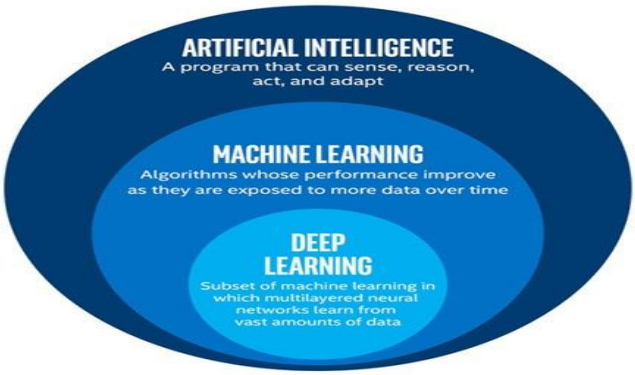
Cuộc cách mạng công nghiệp 4.0, các thuật ngữ như trí tuệ nhân tạo (AI), học máy (machine learning-ML) và học sâu (deep learning–DL) đang ngày càng phổ biến và trở thành những khái niệm mà các công dân thời kỳ kỷ nguyên này buộc phải nắm được.

Học máy là một nhánh nghiên cứu của AI về khả năng tự học của máy tính. Hiện tại, thế giới chưa có 1 định nghĩa chính thức nào về học máy nhưng chúng ta có thể hiểu rằng nó là các kỹ thuật giúp cho máy tính có thể tự học mà không cần phải cài đặt các luật quyết định. Thông thường, một chương trình máy tính cần các quy tắc, luật lệ để có thể thực thi được một tác vụ nào đó. Cuốn sách Machine learning của Tom Mitchell có một cách định nghĩa về Machine learning như sau: "Một chương trình máy tính được cho là học từ kinh nghiệm E, E liên quan đến một số loại nhiệm vụ T, nếu hiệu suất của nó ở trong T, được đo bằng P, và cải thiện theo thời gian" [2].

Dưới đây là một số thuật ngữ cần được nắm rõ:

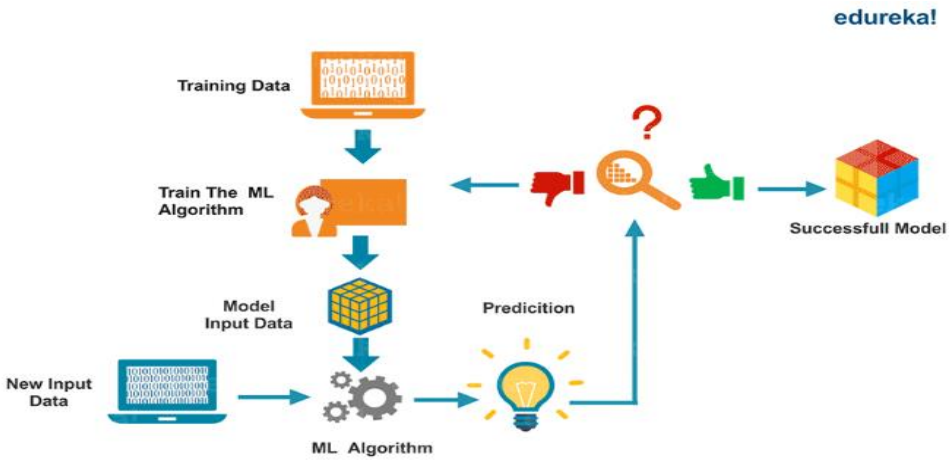
- Trí tuệ nhân tạo (AI) là ngành học tổng hợp bao gồm mọi thứ liên quan đến việc làm cho máy móc trở nên thông minh. Cho dù đó là rô bốt, tủ lạnh, ô tô hay ứng dụng phần mềm, nếu bạn đang biến chúng trở nên thông minh, thì đó chính là AI. Học máy (ML) thường được sử dụng cùng với AI nhưng chúng không giống nhau [13].
- Machine Learning (ML) còn được gọi học máy, là một tập hợp con của AI. ML đề cập đến các hệ thống có thể tự học. Các hệ thống ngày càng thông minh hơn theo thời gian mà không cần sự can thiệp của con người [20].
- Deep Learning (DL) là ML nhưng được áp dụng cho các tập dữ liệu lớn. Hầu hết công việc của AI hiện nay đều liên quan đến ML vì hành vi thông

minh đòi hỏi kiến thức đáng kể và học tập là cách dễ nhất để có được kiến thức đó. Hình ảnh 2-1 dưới đây ghi lại mối quan hệ giữa AI, ML và DL.



Hình 2.1: Mối liên hệ giữa AI, Machine Learning và Deep Learning [26]

Cách thức hoạt động của Machine Learning: Thuật toán Học máy được đào tạo bằng cách sử dụng tập dữ liệu đào tạo để tạo mô hình. Khi *dữ liệu đầu vào mới* được đưa vào thuật toán ML, nó sẽ đưa ra dự đoán trên cơ sở mô hình. Dự đoán được đánh giá về độ chính xác và nếu độ chính xác được chấp nhận, thuật toán Máy học được triển khai. Nếu độ chính xác không được chấp nhận, thuật toán Học máy sẽ được đào tạo lại nhiều lần với tập dữ liệu đào tạo tăng cường. Hình 2.2 mô tả cách thức hoạt động của Machine Learning.

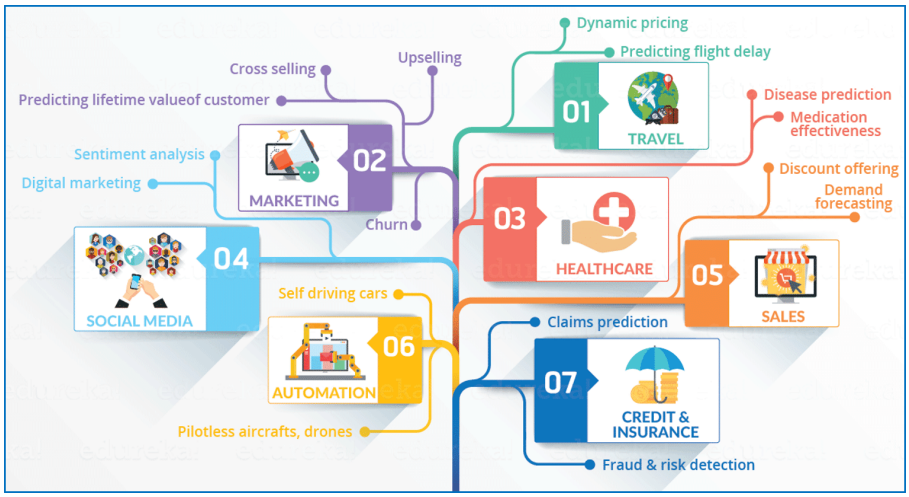


Hình 2.2: Cách thức hoạt động của Machine Learning [5]

Ứng dụng của Machine Learning được sử dụng rất nhiều trong cuộc sống. Nó được thể hiện qua một số lĩnh vực như:

- Trong lĩnh vực du lịch: định giá động (định giá 1 chuyến du lịch bao gồm từ vé máy bay, nhà nghỉ, đặt phòng, thời gian) hay Dự đoán độ trễ của chuyến bay.
- Trong Quảng cáo Marketing: Dự đoán giá trị vòng đời của khách hàng, hay trong bán chéo sản phẩm.
- Trong Chăm sóc sức khỏe: Dự đoán bệnh, các tác động hiệu quả của thuốc.
- Trong truyền thông xã hội: Phân tích tình cảm, số hóa trong tiếp thị.
- Trong tự động hóa: lái xe tự động, máy bay không người lái.

Và còn rất nhiều lĩnh vực khác. Hình 2.3 sẽ thể hiện một số vai trò tích cực khi ứng dụng Machine Learning



Hình 2.3: Ứng dụng Machine Learning [5]

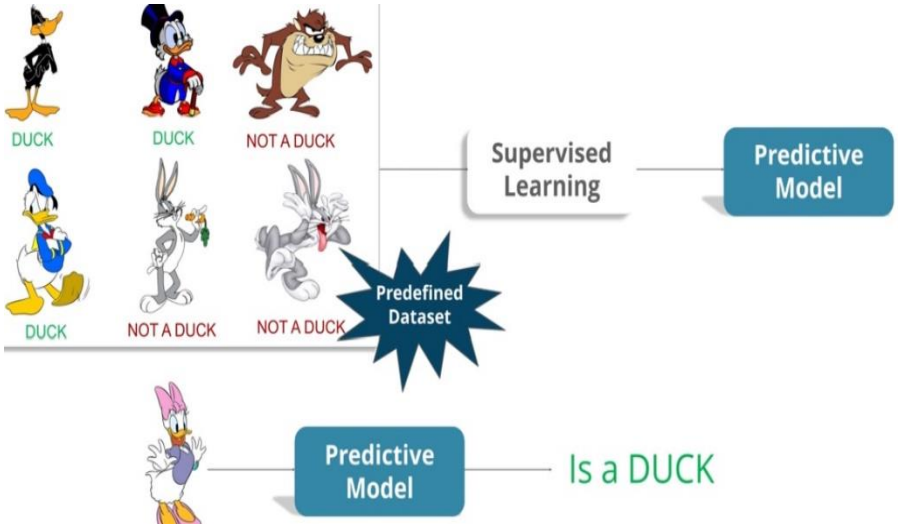
Phân loại thuật toán trong học máy:

Các giải thuật trong học máy gồm 3 loại:

- Học có giám sát (*Supervised Learning*).
- Học phi giám sát (*Unsupervised Learning*).
- Học tăng cường (reinforcement learning).

Học có giám sát (*Supervised Learning*): Học có giám sát về mặt kỹ thuật có nghĩa là học một hàm cung cấp đầu ra cho một đầu vào nhất định dựa trên một tập hợp các cặp đầu vào-đầu ra xác định. Nó thực hiện điều này với sự trợ giúp của 'dữ liệu huấn luyện' được gắn nhãn bao gồm một tập hợp các ví dụ huấn luyện [3].

Một ví dụ khác như trong Hình 2.4, hệ thống có rất nhiều nhân vật hoạt hình. Các nhân vật này sẽ được xác định (gán nhãn) là vịt hoặc không phải là vịt. Thông qua học có giám sát, hệ thống cho ra được mô hình dự đoán. Khi đối tượng mới được đưa vào, dựa trên mô hình dự đoán đoán đó để xác định đối tượng này có phải là vịt hay không. Hình 2.4 phản ánh đúng bản chất của học có giám sát.



Hình 2.4: Nhận dạng đối tượng sử dụng học có giám sát [5]

Tùy thuộc vào loại đầu ra mong muốn, ta chia nhỏ **học có giám sát** gồm:

• Phân loại (*Classification*): Khi đầu ra mong muốn là 1 tập hữu hạn và rời rạc. Chi tiết hơn của mục tiêu đầu ra là lớp nhãn hoặc lớp danh mục. Đối với trường hợp dự đoán giá trị liên tục, nhưng giá trị liên tục ở dạng xác suất đối với lớp thì bài toán đó cũng được coi là bài toán phân loại [20]. Các dạng bài toán trên sẽ được gọi là **bài toán phân loại**. Như tên cho thấy, các thuật toán phân loại đảm nhận công việc dự đoán một nhãn hoặc đưa biến vào một loại (phân loại).

Một số thuật toán trong phân loại của học máy là: Logistic Regression, K-Nearest Neighbours, Support Vector Machines, Kernel SVM, Naïve Bayes, Decision Tree Classification, Random Forest Classification… [19].

• Hồi quy(*Regression*): Khi đầu ra không xác định một "danh mục" của biến mà chỉ định một số lượng/ con số cho nó dựa trên dữ liệu lịch sử [3]. Các điểm dữ liệu trong hồi quy có giá trị liên tục. Thuật toán sử dụng mối quan hệ giữa một biến độc

lập và một biến phụ thuộc từ dữ liệu mối quan hệ lịch sử và dự đoán một số lượng. Các bài toán như vậy sẽ được gọi là **hồi quy.**

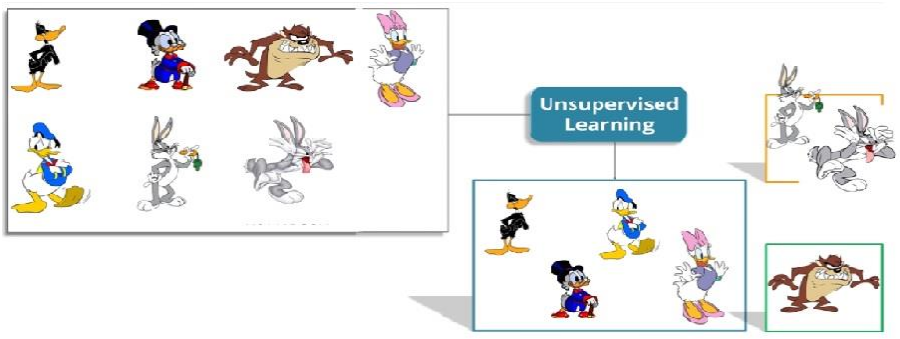
Một số thuật toán trong hồi quy của học máy là: Simple Linear Regression, Multiple Linear Regression, Polynomial Regression, Support Vector Regression, Decision Tree Regression, Random Forest Regression …[19].

Học phi giám sát / Học không giám sát (*Unsupervised Learning*): sử dụng những dữ liệu chưa được gán nhãn từ trước để suy luận. Phương pháp được sử dụng với mục đích tìm cấu trúc của tập dữ liệu. Tuy vậy, không có phương pháp đánh giá được cấu trúc tìm ra được là đúng hay sai.

Một cách thể hiện khác về học không giám sát như sau:

Học không giám sát là nơi bạn chỉ có dữ liệu đầu vào (X) và không có biến đầu ra tương ứng. Mục tiêu của việc học không giám sát là mô hình hóa cấu trúc hoặc phân phối cơ bản trong dữ liệu để tìm hiểu thêm về dữ liệu [21]. Đây được gọi là học không giám sát vì không giống như học có giám sát ở trên, không có câu trả lời chính xác và không có giáo viên.

Một ví dụ cho học phi giám sát, được thể hiện trong Hình 2.5. Một danh sách các nhân vật hoạt hình không được xác định. Thông qua học phi giám sát, hệ thống có thể phân loại ra nhóm các nhân vật là vịt, thỏ hay sói giống như ảnh bên dưới trong Hình 2.5.



Hình 2.5: Phân loại nhóm đối tượng sử dụng học phi giám sát [5]

Các vấn đề học tập không được giám sát có thể được nhóm lại thành các vấn đề phân cụm và liên kết. Trong đó:

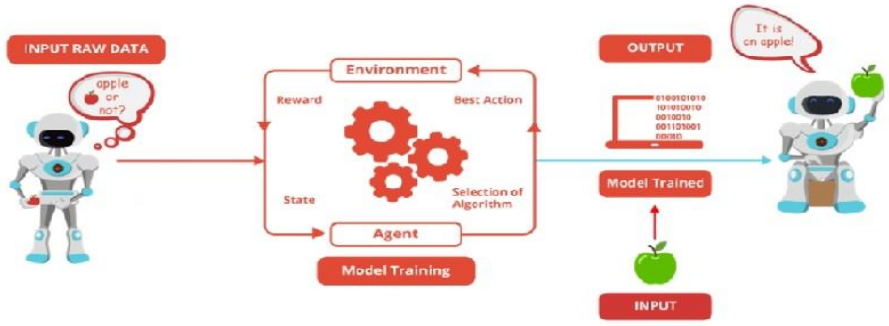
- Phân cụm: Vấn đề phân cụm là nơi bạn muốn khám phá các nhóm vốn có trong dữ liệu, chẳng hạn như nhóm khách hàng theo hành vi mua hàng.
- Liên kết: Một vấn đề học tập quy tắc liên kết là nơi bạn muốn khám phá các quy tắc mô tả phần lớn dữ liệu của bạn, chẳng hạn như những người mua X cũng có xu hướng mua Y.

Một số ví dụ phổ biến về thuật toán học không giám sát là: k-Means clustering, k-Medians, Expectation Maximization (EM), Apriori,Independent Component Analysis (ICA), Principal Component Analysis (PCA)…

Điểm khác biệt lớn nhất của thuật toán Supervised Learning với Unsupervised Learning, đó là cách chúng ta cung cấp tập dữ liệu huấn luyện cho mô hình. Cách thuật toán sử dụng dữ liệu và loại vấn đề phù hợp để chúng giải quyết.

Học tăng cường (*reinforcement learning*). Phương pháp học tăng cường tập trung vào việc làm sao để cho 1 tác tử Agent trong môi trường có thể hành động sao cho lấy được phần thưởng (reward) nhiều nhất có thể. Bản chất của việc học tăng cường là trial-and-error, nghĩa là thử đi thử lại và rút ra kinh nghiệm sau mỗi lần thử như vậy. Khác với học có giám sát, nó không có cặp dữ liệu gán nhãn trước làm đầu vào và cũng không có đánh giá các hành động là đúng hay sai.

Ví dụ: Nhận diện đồ vật của robot



Hình 2.6: Ví dụ học tăng cường [5]

Trong ví dụ Hình 2.6, dữ liệu thô ban đầu là một đối tượng mà hệ thống hoàn toàn không có định nghĩa về nó. Thông qua khả năng của một tác nhân (agent) để tương tác với môi trường và tìm ra kết quả tốt nhất là gì. Nó tuân theo khái niệm về phương pháp đánh và thử nghiệm (hit and trial). Tác nhân Agent sẽ được thưởng hoặc

phạt 1 điểm cho câu trả lời đúng hoặc sai. Trên cơ sở các điểm thưởng tích cực đạt được, mô hình tự đào tạo. Và một lần nữa, nó đã sẵn sang để dự đoán dữ liệu mới được trình bày cho nó. Trong ví dụ Hình 2.6, con robot lúc đầu không nhận biết được vật trên tay nó cầm là cái gì. Sau nhiều lần tự đào tạo theo khái niệm được cung cấp. Mô hình tự đào tạo đến ngưỡng nào đó, nó có thể nhận biết được được đối tượng đưa vào là gì.

Trong thực tế, học tăng cường đã được áp dụng thành công cho nhiều bài toán, trong đó có điều khiển robot, điều vận thang máy, viễn thông, các trò chơi backgammon và cờ vua.

2.2 Các công nghệ ứng dụng trong bài toán

Trước khi công nghệ AI phát triển và sự bùng nổ mạnh mẽ về dữ liệu, cách thức xác định dự đoán tỷ giá đa phần đều sử dụng phương thức xác suất thống kê với 3 cách làm thông thường là sức mua tương đương, sức mạnh kinh tế tương đối và các mô hình kinh tế lượng. Theo tác giả Joseph Nguyễn của bài viết "*3 cách phổ biến để dự báo tỷ giá hối đoái tiền tệ*" trên trang investopedia [22], sức mua tương đương - Purchasing Power Parity (PPP) là phương pháp phổ biến nhất do nó được phổ biến trong hầu hết các sách giáo khoa kinh tế. Phương pháp dự báo PPP dựa trên quy luật lý thuyết về một mức giá , trong đó nói rằng các hàng hóa giống nhau ở các quốc gia khác nhau nên có giá giống nhau. Phương pháp này tiếp cận sức mạnh kinh tế tương đối so sánh mức độ tăng trưởng kinh tế giữa các quốc gia để dự báo tỷ giá hối đoái và sử dụng các mô hình kinh tế lượng có thể xem xét một loạt các biến số khi cố gắng tìm hiểu xu hướng trên thị trường tiền tệ. Cách tiếp cận PPP dự báo rằng tỷ giá hối đoái sẽ thay đổi để bù đắp những thay đổi về giá do lạm phát dựa trên nguyên tắc cơ bản này. Còn về phương pháp sức mạnh kinh tế tương đối, phương pháp này xem xét sức mạnh tăng trưởng kinh tế ở các quốc gia khác nhau để dự báo hướng của tỷ giá hối đoái. Cơ sở lý luận đằng sau cách tiếp cận này dựa trên ý tưởng rằng một môi trường kinh tế mạnh mẽ và tiềm năng tăng trưởng cao có nhiều khả năng thu hút đầu tư từ các nhà đầu tư nước ngoài. Và để mua các khoản đầu tư ở quốc gia mong muốn, nhà đầu tư sẽ phải mua tiền tệ của quốc gia đó tạo ra nhu cầu tăng lên khiến đồng tiền

tăng giá. Cách tiếp cận này không chỉ xem xét sức mạnh kinh tế tương đối giữa các quốc gia mà còn cần một cái nhìn tổng thể hơn và xem xét tất cả các dòng vốn đầu tư. Ví dụ, một yếu tố khác có thể thu hút các nhà đầu tư đến một quốc gia nhất định là lãi suất. Lãi suất cao sẽ thu hút các nhà đầu tư tìm kiếm lợi tức cao nhất cho các khoản đầu tư của họ, khiến nhu cầu về tiền tệ tăng lên, điều này một lần nữa sẽ dẫn đến sự tăng giá của đồng tiền này. Phương pháp sức mạnh kinh tế tương đối không dự báo tỷ giá hối đoái sẽ là bao nhiêu. Không giống như phương pháp PPP, cách tiếp cận này cung cấp cho nhà đầu tư một cảm giác chung về việc liệu một loại tiền tệ sẽ tăng giá hay giảm giá và cảm nhận tổng thể về sức mạnh của chuyển động. Nó thường được sử dụng kết hợp với các phương pháp dự báo khác để tạo ra một kết quả hoàn chỉnh. Tiếp đến là phương pháp các mô hình kinh tế lượng của dự báo tỷ giá hối đoái. Đây là một phương pháp phổ biến khác được sử dụng để dự báo tỷ giá hối đoái liên quan đến việc thu thập các yếu tố có thể ảnh hưởng đến chuyển động tiền tệ và tạo ra một mô hình liên hệ các biến số này với tỷ giá hối đoái. Các yếu tố được sử dụng trong các mô hình kinh tế lượng thường dựa trên lý thuyết kinh tế, nhưng bất kỳ biến nào cũng có thể được thêm vào nếu nó được cho là có ảnh hưởng đáng kể đến tỷ giá hối đoái. Ví dụ dự báo tỷ giá hối đoái USD / CAD trong năm tới của một công ty. Họ tin rằng một mô hình kinh tế lượng sẽ là một phương pháp tốt để sử dụng và đã nghiên cứu các yếu tố mà họ cho là ảnh hưởng đến tỷ giá hối đoái. Từ nghiên cứu và phân tích của mình, họ kết luận các yếu tố có ảnh hưởng lớn nhất là: chênh lệch lãi suất giữa Mỹ và Canada (INT), chênh lệch tỷ lệ tăng trưởng GDP (GDP) và chênh lệch tỷ lệ tăng trưởng thu nhập (IGR) giữa hai Quốc gia. Mô hình kinh tế lượng mà họ đưa ra được hiển thị như sau: USD / Cad (1 - Năm) = z + a(INT) + b(GDP) + c(IGR). Trong đó, z=Tỷ giá hối đoái cơ bản không đổi. a ,b và c=Hệ số đại diện cho tương đối trọng lượng của từng yếu tố. INT=Sự khác biệt về lãi suất giữa Hoa Kỳ và Canada. GDP=Sự khác biệt về tốc độ tăng trưởng GDP. IGR=Sự khác biệt về tỷ lệ tăng thu nhập. Sau khi tạo mô hình, các biến INT, GDP và IGR có thể được cắm vào để tạo dự báo. Các hệ số a, b và c sẽ xác định mức độ ảnh hưởng của một yếu tố nào đó đến tỷ giá hối đoái và hướng của tác động (cho dù nó là tích cực hay tiêu cực). Phương

pháp này có lẽ là cách tiếp cận phức tạp và tốn thời gian nhất, nhưng một khi mô hình được xây dựng, dữ liệu mới có thể dễ dàng được thu thập và cắm vào để tạo ra các dự báo nhanh chóng. Tuy nhiên, cả 3 phương pháp trên đều gặp một số hạn chế đó là tỷ lệ hiệu suất thấp hơn, kém hiệu quả hơn và giao tiếp chậm hơn. Với những tiến bộ trong máy tính công nghệ, hệ thống Trí tuệ nhân tạo (AI) ngày nay, đặc biệt là học máy (ML) đã có thể đáp ứng, khắc phục các hạn chế trên. Bởi vậy, trong đề tài này, học viên xác định sử dụng học máy để thực hiện dự đoán tỷ giả hối đoái.

Mục tiêu nghiên cứu của đề tài là mong muốn xác định giá trị tương lai của bài toán dự đoán tỷ giá USD/VNĐ, chi tiết hơn là giá trị tương lai của ngày tiếp theo đó. Do vậy, học viên xác định đây là bài toán thuộc loại bài toán Hồi quy (Regression). Học máy (ML) có rất nhiều mô hình để giải quyết bài toán hồi quy (Regression). Trong đó, mô hình cơ bản nhất là Linear Regression, tiếp đến để tăng dần độ phức tạp, để giải quyết bài toán xử lý các giá trị bị thiếu, và duy trì sự chính xác của một tỷ lệ lớn dữ liệu, không thể không nhắc đến mô hình Random Forest. Và một trong những mô hình ngày nay đang được áp dụng có yêu cầu độ chính xác cao với lượng lớn dữ liệu phức tạp, đó là mô hình mạng trí tuệ nhân tạo (Neural Network). Đây cũng là 3 mô hình được học viên lựa chọn để kiểm nghiệm về độ chính xác của dự báo trong luận văn này.

2.2.1 Linear Regression

Linear Regression là một thuật toán trong học máy, thuộc loại giải thuật học có giám sát. Thuật toán này nằm giải quyết các bài toán hồi quy trong học có giám sát. Linear Regression còn được gọi là hồi quy tuyến tính. Trong thống kê, hồi quy tuyến tính là một cách tiếp cận tuyến tính để mô hình hóa mối quan hệ giữa một phản ứng vô hướng và một hoặc nhiều biến giải thích (còn được gọi là biến phụ thuộc và độc lập). Ở học máy, Linear Regression có các mối quan hệ được mô hình hóa bằng cách sử dụng các hàm dự báo tuyến tính mà các tham số mô hình chưa biết được ước tính từ dữ liệu. Các mô hình như vậy được gọi là mô hình tuyến tính [8].

Phân tích toán học:

a) Dạng của Linear Regression:

Cho tập dữ liệu $(\mathbf{x}^{(i)}, \boldsymbol{y}^{(i)})_{i=1}^{n}$ với n là đơn vị thống kê, với $\mathbf{x}=\{x_1, x_2, \dots, x_n\}$ là tập các thuộc tính và **y** là tập giá trị đầu ra. Có thể hiểu đơn giản (*x*,*y*)- một bản ghi đầy đủ của dữ liệu mà ta thu thập được. Mô hình Linear regression giả định mối quan hệ tuyến tính giữa **y** và **x**, sao cho giá trị $\hat{y} \approx \mathbf{y}$. Khi đó phương trình có dạng tổng quát như sau [11]:

$$\mathbf{y} \approx \hat{y}, \text{ với } \hat{y} = f(x)$$

$$f(x) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n \quad (2.1)$$

Trong đó, $w_0, w_1, w_{2,\dots}, w_n$ là các hằng số, w_0 còn được gọi là bias. Mối quan hệ $y = f(x)$ bên trên là một mối quan hệ tuyến tính (linear).

Nếu viết dưới dạng ma trận thì phương trình (2.1) có dạng như sau:

$$\hat{y} = \bar{\boldsymbol{x}}\boldsymbol{w} \quad (2.2)$$

Với $\bar{\boldsymbol{x}} = [1, x_1, x_{2,\dots}, x_n]$ là vector (hàng) dữ liệu đầu vào.

Và $\mathbf{w} = [w_0, w_1, w_{2,\dots}, w_n]^T$ là vector (cột) hệ số cần phải tối ưu.

b) Sai số dự đoán:

Chúng ta mong muốn rằng sự sai khác *e* giữa giá trị thực **y** và giá trị dự đoán $\hat{y}$ là nhỏ nhất. Nói cách khác, chúng ta muốn giá trị sau đây càng nhỏ càng tốt:

$$\frac{1}{2} e^2 = \frac{1}{2} (\boldsymbol{y} - \hat{y})^{\mathbf{2}} = \frac{1}{2} (\boldsymbol{y} - \bar{\boldsymbol{x}}\boldsymbol{w})^{\mathbf{2}} \quad (2.3)$$

c) Hàm mất mát:

Điều chúng ta muốn, tổng sai số là nhỏ nhất, tương đương với việc tìm **w** để hàm số sau đạt giá trị nhỏ nhất:

$$\mathcal{L}(\boldsymbol{w}) = \frac{1}{2} \sum_{i=1}^{n} (y_i - \bar{\boldsymbol{x}}_i \mathbf{w})^2 \text{ ta gọi đây là phương trình (2.4)}$$

Hàm số $\mathcal{L}(\boldsymbol{w})$ được gọi là **hàm mất mát** (loss function) của bài toán Linear Regression. Chúng ta luôn mong muốn rằng sự mất mát (sai số) là nhỏ nhất, điều đó đồng nghĩa với việc tìm vector hệ số **w** sao cho giá trị của hàm mất mát này càng nhỏ càng tốt. Giá trị của **w** làm cho hàm mất mát đạt giá trị nhỏ nhất được gọi là *điểm tối ưu* (optimal point), ký hiệu: $\boldsymbol{w}^* = \arg min\ \mathcal{L}(\boldsymbol{w})$

Hàm mất mát ở phương trình (2.4) có thể được viết đơn giản hơn dưới dạng như sau: $\mathcal{L}(\boldsymbol{w}) = \frac{1}{2} \sum_{i=1}^{n} (y_i - \bar{\boldsymbol{x}}_i \mathbf{w})^2 = \frac{1}{2} \|\boldsymbol{y} - \bar{\boldsymbol{X}}\boldsymbol{w}\|_2^2 \quad (2.5)$

Với $\mathbf{y} = [y_1, y_2, \ldots, y_n]$ là một vetor chứa tất cả các *output* của *training data*

$\overline{\boldsymbol{X}} = [\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \ldots, \overline{\boldsymbol{X}}_n]$ là ma trận dữ liệu đầu vào mà mỗi hàng của nó là một điểm dữ liệu.

Với $\|\boldsymbol{z}\|_2$ là Euclidean norm (chuẩn Euclid, hay khoảng cách Euclid), nói cách khác $\|\boldsymbol{z}\|_2^2$ là tổng của bình phương mỗi phần tử của vector

d) Nghiệm cho bài toán Linear Regression:

Cách phổ biến nhất để tìm nghiệm cho một bài toán tối ưu là giải phương trình đạo hàm bằng 0. Đạo hàm theo **w** của hàm mất mát là:

$$\frac{\partial \mathcal{L}(\boldsymbol{w})}{\partial \mathbf{w}} = \overline{\boldsymbol{X}}^T(\overline{\boldsymbol{X}}\boldsymbol{w} - \boldsymbol{y})$$

Phương trình đạo hàm bằng 0 :

$$\frac{\partial \mathcal{L}(\boldsymbol{w})}{\partial \mathbf{w}} = 0 \Leftrightarrow \overline{\boldsymbol{X}}^T(\overline{\boldsymbol{X}}\boldsymbol{w} - \boldsymbol{y}) = 0$$

$$\Leftrightarrow \overline{\boldsymbol{X}}^T\overline{\boldsymbol{X}}\boldsymbol{w} - \overline{\boldsymbol{X}}^T\boldsymbol{y} = 0$$

$$\Leftrightarrow \overline{\boldsymbol{X}}^T\overline{\boldsymbol{X}}\boldsymbol{w} = \overline{\boldsymbol{X}}^T\boldsymbol{y} \qquad (2.6)$$

Đặt $\overline{\boldsymbol{X}}^T\boldsymbol{y}$ bằng **b**

Nếu ma trân vuông $\mathbf{A} \triangleq \overline{\boldsymbol{X}}^T\overline{\boldsymbol{X}}$ khả nghịch, thì phương trình (2.6) có nghiệm duy nhất: $\mathbf{w} = \mathbf{A}^{-1} - \mathbf{b}$

Nếu ma trận **A** không khả nghịch (tức có định thức bằng 0), thì phương trình (2.6) vô nghiệm hoặc có vô số nghiệm.

Nếu ma trận **A** không vuông, ma trận không khả nghịch ta dùng khái niệm giả nghịch đảo. ➔ điểm tối ưu của bài toán Linear Regression có dạng:

$$\mathbf{w} = \mathbf{A}^+ \mathbf{b} = (\overline{\boldsymbol{X}}^T\overline{\boldsymbol{X}})^+ \overline{\boldsymbol{X}}^T\boldsymbol{y} \quad (2.7)$$

Ví dụ:

Dự đoán khả năng chịu lực của một cánh tay đòn. Biết rằng có thông tin của 15 trường hợp được coi là mẫu. Trong đó, mỗi mẫu sẽ cho biết thông tin chiều dài cánh tay đòn và khả năng chịu lực của cánh tay đòn đó. Hãy dự đoán khả năng chịu lực của cánh tay đòn có chiều dài 155 cm và 160 cm.

Dữ liệu đầu vào:

```
Chieu dai tay don:     Kha nang chiu luc:
[[147]                 [[68]
 [150]                  [67]
 [153]                  [66]
 [158]                  [64]
 [163]                  [63]
 [165]                  [62]
 [168]                  [60]
 [170]                  [59]
 [173]                  [58]
 [175]                  [54]
 [178]                  [51]
 [180]                  [50]
 [183]]                 [48]]
```

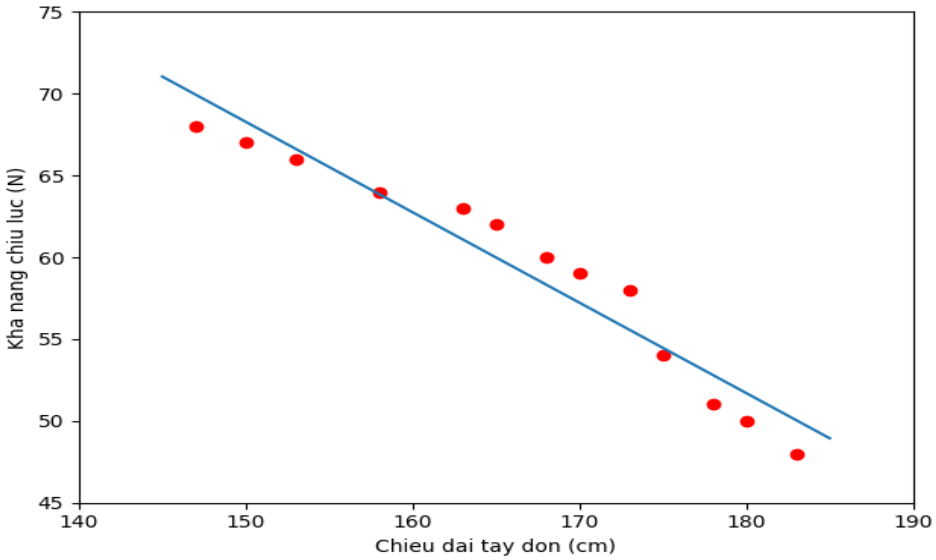
Hình 2.7: Dữ liệu đầu vào trong ví dụ sử dụng Linear Regression

Dữ liệu sau khi chạy được:

```
w =  [[151.22624754]
 [ -0.55290856]]
Du doan kha nang chiu luc cua tay don dai 155 cm la: 65.53 (N), số liệu thật: 65 (N)
Du doan kha nang chiu luc cua tay don dai 160 cm la:, 62.76 (N) số liệu thật: 62 (N)
Nghiệm tìm được bằng scikit-learn  :  [[151.22624753  -0.55290856]]
Nghiệm tìm được từ phương trình (2-6):  [[151.22624754  -0.55290856]]
```

Hình 2.8: Nghiệm của bài toán Linear Regression

Kết quả dự đoán:



Hình 2.9: Đồ thị mô tả dự đoán Linear Regression

2.2.2 Random Forest

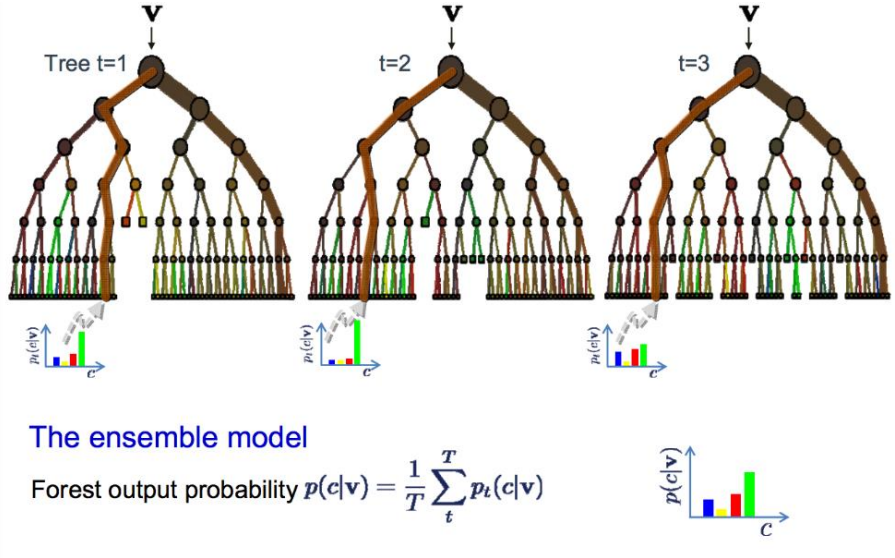
a) Tổng quan về RandomForest

Random Forest là một phương pháp Supervised Learning do vậy có thể xử lý được các bài toán về Classification (phân loại) và Regression (dự báo về các giá trị). Đúng như tên gọi của nó Random Forest - rừng ngẫu nhiên: đây là phương pháp xây dựng một tập hợp rất nhiều cây quyết định và sử dụng phương pháp bầu cử/ bỏ phiếu (voting) để đưa ra quyết định về biến target cần được dự báo. Có rất nhiều thông tin xung quanh làm cơ sơ tham khảo. Mỗi một ý kiến tương ứng với một Decision Tree trả lời các câu hỏi. Sau đó bạn sẽ có một rừng các câu trả lời để quyết định xem mình sẽ đi thực hiện thế nào. Random Forest hoạt động bằng cách đánh giá các Decision Tree sử dụng cách thức voting để đưa ra kết quả cuối cùng [28].

b) Đặc điểm của thuật toán

Random Forest (Rừng ngẫu nhiên) hay Random Descision forests (rừng quyết định ngẫu nhiên) là một phương pháp học tập tổng hợp để phân loại, hồi quy và các nhiệm vụ khác hoạt động bằng cách xây dựng vô số cây quyết định tại thời điểm đào tạo và xuất ra lớp là chế độ của các lớp (phân loại) hoặc dự đoán trung bình / trung bình (hồi quy) của các cây riêng lẻ [9].

Rừng quyết định ngẫu nhiên phù hợp với thói quen thích nghi quá mức của cây quyết định đối với tập huấn luyện của chúng [1]. Rừng ngẫu nhiên thường hoạt động tốt hơn cây quyết định, nhưng độ chính xác của chúng thấp hơn cây được tăng cường độ dốc. Tuy nhiên, đặc điểm dữ liệu có thể ảnh hưởng đến hiệu suất của chúng.



Hình 2.10: Hình ảnh minh họa về Random Forest

Thuật toán đào tạo cho các khu rừng ngẫu nhiên áp dụng kỹ thuật tổng hợp bootstrap hoặc đóng gói chung cho những người học cây. Cho một tập huấn luyện X = x $_1$,..., x $_n$ với các phản hồi Y = y $_1$,..., y $_n$, đóng gói lặp lại (B lần) chọn một mẫu ngẫu nhiên thay thế tập huấn luyện và lắp các cây vào mẫu:

Đối với b = 1,..., B :

1. Ví dụ huấn luyện mẫu, với thay thế, *n* từ *X* , *Y* ; gọi chúng là X_b , Y_b.
2. Huấn luyện cây phân loại hoặc hồi quy f_b trên X_b , Y_b.

Sau khi huấn luyện, các dự đoán cho các mẫu chưa nhìn thấy x' có thể được thực hiện bằng cách lấy trung bình các dự đoán từ tất cả các cây hồi quy riêng lẻ trên x' hoặc bằng cách lấy đa số phiếu trong trường hợp phân loại cây. :

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \quad (2.8)$$

Quy trình khởi động này dẫn đến hiệu suất mô hình tốt hơn vì nó làm giảm phương sai của mô hình mà không làm tăng độ chệch. Điều này có nghĩa là trong khi các dự đoán của một cây đơn lẻ rất nhạy cảm với nhiễu trong tập huấn luyện của nó, thì trung bình của nhiều cây thì không, miễn là các cây không tương quan. Chỉ cần huấn luyện nhiều cây trên một tập huấn luyện duy nhất sẽ cho các cây có tương quan chặt chẽ (hoặc thậm chí cùng một cây nhiều lần, nếu thuật toán huấn luyện là xác

định); lấy mẫu bootstrap là một cách khử tương quan giữa các cây bằng cách hiển thị cho chúng các tập huấn luyện khác nhau.

c) Một số ứng dụng của thuật toán Random Forest:

- Trong ngân hàng: Khu vực ngân hàng bao gồm rất nhiều người dùng. Có nhiều khách hàng trung thành và cũng có những khách hàng lừa đảo. Để xác định xem khách hàng là khách hàng trung thành hay gian lận, phân tích của Random Forest được đưa vào. Với sự trợ giúp của thuật toán Random Forest trong học máy, chúng ta có thể dễ dàng xác định xem khách hàng là gian lận hay trung thành. Một hệ thống sử dụng một tập hợp các thuật toán ngẫu nhiên để xác định các giao dịch gian lận bằng một loạt các mẫu [4].

- Trong y học lĩnh vực thuốc: Thuốc cần sự kết hợp phức tạp của các hóa chất cụ thể. Như vậy, để nhận biết sự kết hợp tuyệt vời trong các vị thuốc, có thể sử dụng Random Forest. Với sự trợ giúp của thuật toán học máy, việc phát hiện và dự đoán độ nhạy thuốc của một loại thuốc đã trở nên dễ dàng hơn. Ngoài ra, nó giúp xác định bệnh của bệnh nhân bằng cách phân tích bệnh án của bệnh nhân [4].

- Thị trường chứng khoán: Máy học cũng đóng vai trò trong việc phân tích thị trường chứng khoán. Khi bạn muốn biết hành vi/xu hướng của thị trường chứng khoán, với sự trợ giúp của thuật toán Random Forest, xu hướng của thị trường chứng khoán có thể được phân tích. Ngoài ra, nó có thể hiển thị lỗ hoặc lợi nhuận dự kiến có thể được tạo ra trong khi mua một cổ phiếu cụ thể [4].

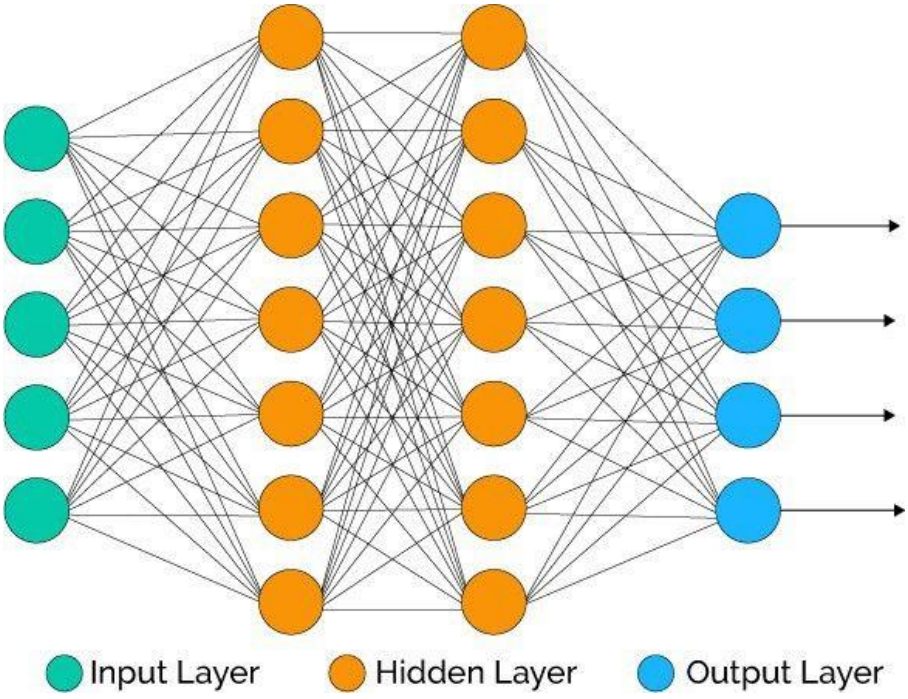
- Thương mại điện tử: Khi bạn cảm thấy khó khăn trong việc giới thiệu hoặc gợi ý loại sản phẩm mà khách hàng của bạn nên xem. Đây là nơi bạn có thể sử dụng một thuật toán Random Forest. Sử dụng hệ thống máy học, bạn có thể đề xuất các sản phẩm có nhiều khả năng được khách hàng quan tâm hơn. Sử dụng một mẫu nhất định và theo sở thích sản phẩm của khách hàng, bạn có thể đề xuất các sản phẩm tương tự cho khách hàng của mình [4].

2.2.3 Neural Network

a) Tổng quan về Neural Network

Mạng neural được xây dựng dựa trên mạng neural sinh học. Các neural (nút) nối với nhau và xử lý thông tin dựa trên cách truyền theo các kết nối và tính giá trị tại các nút [25]. Mạng neuron với mỗi nút sẽ có những dữ liệu đầu vào, biến đổi những dữ liệu đầu vào này bằng cách tính tổng các input với weight tương ứng trên các đầu vào, sau đó áp dụng một hàm biến đổi phi tuyến tính cho phép biến đổi này để tính toán trạng thái trung gian. 3 bước trên tạo thành 1 lớp và hàm biến đổi còn được gọi là activation funtion. Các output của layer này sẽ là input của layer phía sau. Thông qua việc lặp lại các bước trên, neural-network học thông qua nhiều layer và các nút phi tuyến tính rồi sau đó kết hợp lại ở layer cuối cùng để cho ra 1 dự đoán. Trong nerual network nếu mô hình có 1 lớp hidden hoặc nhiều lớp hideen được gọi Multi Layer Perceptron (MLP). Ví dụ như trong Hình 2.11. Trường hợp không có bất kỳ lớp hidden nào thì sẽ được gọi Single Layer Perceptron (SLP).

Neural-network học bằng cách tạo ra các tín hiệu lỗi đo lường sự khác biệt giữa các dự đoán của mạng và giá trị mong muốn, sau đó sử dụng tín hiệu lỗi này để cập nhật lại weight và bias trong activation function để việc dự đoán sau đó chính xác hơn



Hình 2.11: Mạng netural network nhiều lớp ẩn

Activation function là 1 thành phần rất quan trọng của neural-network. Activation có nhiệm vụ chuẩn hóa Output. Nó quyết định khi nào thì 1 neuron được kích hoạt hoặc không. Liệu thông tin mà neuron nhận được có liên quan đến thông tin được đưa ra hay nên bỏ qua.

*Y= Activation((weight * input) + bias)* (2.9)

Với Weight: là trọng số của đường nối

Activation function là 1 phép biến đổi phi tuyến tính mà chúng ta thực hiện đối với tín hiệu đầu vào. Đầu ra được chuyển đổi này sẽ được sử dụng làm đầu vào của neuron ở layer tiếp theo.

Nếu không có activation function thì weight và bias chỉ đơn giản như 1 hàm biến đổi tuyến tính. Giải 1 hàm tuyến tính sẽ đơn giản hơn nhiều nhưng sẽ khó có thể mô hình hóa và giải được những vấn đề phức tạp. Một mạng neuron nếu không có activation function thì cơ bản chỉ là 1 model hồi quy tuyến tính. Activation function thực hiện việc biến đổi phi tuyến tính với đầu vào làm việc học hỏi và thực hiện những nhiệm vụ phức tạp hơn như dịch ngôn ngữ hoặc phân loại ảnh là khả thi.

Activation function hỗ trợ back-propagation (tuyên truyền ngược) với việc cung cấp các lỗi để có thể cập nhật lại các weight và bias, việc này giúp mô hình có khả năng tự hoàn thiện. Dưới đây là các hàm Activation trong Nerual Network

- Binary step: $f(x) = 1, x \geq 0$
- Linear: $f(x) = ax$
- Sigmoid: $f(x) = \frac{1}{1 + e^{-x}}$
- Tanh: $tanh(x) = \frac{2}{1 + e^{-2x}} - 1$
- ReLU: $f(x) = \max(0, x)$
- Softmax: $a(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_j}}$ $với\ j = 1, \ldots, k$

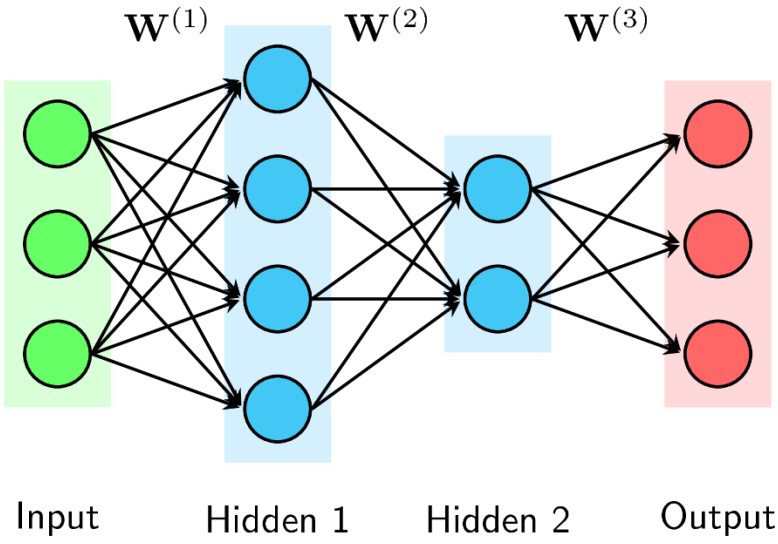
Cách lựa chọn activation function:

- Các hàm sigmoid và sự kết hợp của chúng thường phù hợp với những bài toán phân loại

- Sigmoid và tanh đôi khi nên tránh sử dụng đồng thời vì có thể khiến gradient biến mất
- ReLU là 1 activation function phổ biến và thường dùng nhất hiện nay Nếu gặp những trường hợp có tế bào neuron chết trong mạng thì leaky thì ReLU là 1 lựa chọn hoàn hảo
- ReLU function chỉ có thể được sử dụng trong những hidden layer

b) MLP Regressor

Với 1 layer Input và 1 layer Output , một Multi-layer Perceptron (MLP) có thể có nhiều Hidden layers ở giữa. Các *Hidden layers* theo thứ tự từ input layer đến output layer được đánh số thứ thự là *Hidden layer 1*, *Hidden layer 2*, … Hình 2.12 dưới đây là một ví dụ với 2 Hidden layers [12].

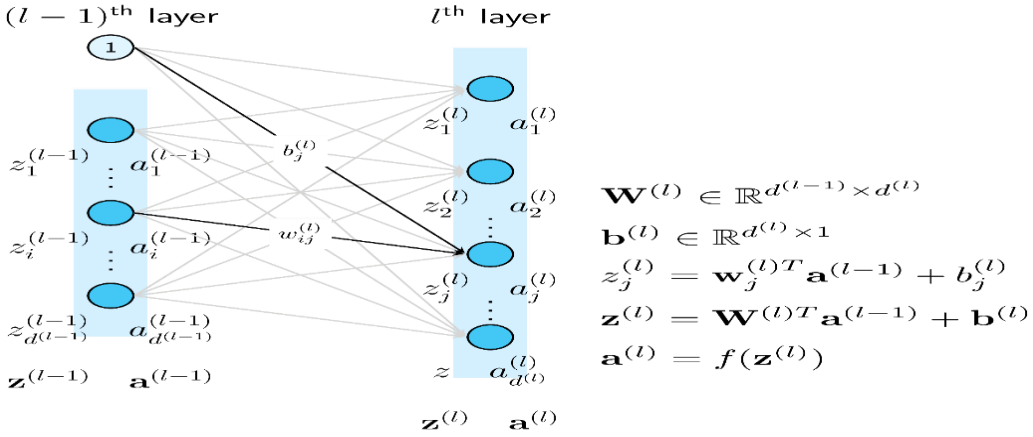


Hình 2.12: Ví dụ MLP với 2 hidden Layer

Trong đó 1 node hình trong trong 1 layer được gọi là 1 unit, hoặc có thể gọi là 1 cell. Unit ở lớp nào thì sẽ được gọi theo cấu trúc tên lớp + unit. Ví dụ Unit ở các input layer, hidden layers, và output layer được lần lượt gọi là input unit, hidden unit, và output unit.

Các ký hiệu ***z,*** **a**, **b, d**, **W** đều được thể hiện trong Hình 2.13. Trong đó, các hidden layer có đầu vào được ký hiệu bởi ***z*** và đầu ra được ký hiệu là ***a*** (thể hiện *activation*, tức giá trị của mỗi unit sau khi ta áp dụng activation function lên ***z***). Unit

thứ i trong layer thứ l có đầu ra được ký hiệu là $a_i^{(l)}$. Giả sử thêm rằng số unit trong layer thứ l (không tính bias) là $d^{(l)}$. Vector biểu diễn output của layer thứ l được ký hiệu là $\mathbf{a}^{(l)} \in \mathbf{R}^{d^{(l)}}$.



Hình 2.13: Các ký hiệu sử dụng trong MLP

Tập hợp các weights và biases lần lượt được ký hiệu là **W** và **b.** Với L ma trận trọng số cho một MLP có L layers. Các ma trận này được ký hiệu là $\mathbf{W}^{(l)} \in \mathbf{R}^{d^{(l-1)} \text{ x } d^{(l)}}$ với l=1,2,…, L trong đó $\mathbf{W}^{(l)}$ thể hiện các *kết nối* từ layer thứ l−1 tới layer thứ l (nếu ta coi input layer là layer thứ 0). Làm rõ hơn, phần tử $w_{ij}^{(l)}$ thể hiện kết nối từ node thứ i của layer thứ (l−1) tới node từ j của layer thứ (l). Các biases của layer thứ (l) được ký hiệu là $\mathbf{b}^{(l)} \in \mathbf{R}^{d^{(l)}}$.

Mỗi output của một unit (trừ các units ở lớp input layer) được tính dựa vào công thức: $$a_i^{(l)} = f(w_i^{(l)T} a^{(l-1)} + b_i^{(l)}) \quad (2.10)$$

Trong đó $f(.)$ là một (nonlinear) activation function. Ở dạng vector, biểu thức bên trên được viết là: $$\mathbf{a}^{(l)} = f(W^{(l)T}\, \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}) \quad (2.11)$$

Khi activation function $f(.)$ được áp dụng cho một ma trận (hoặc vector), có nghĩa nó được áp dụng cho *từng thành phần của ma trận đó*. Các thành phần này sẽ được sắp xếp lại đúng theo thứ tự để được một ma trận có kích thước bằng với ma trận đầu vào input.

Ví dụ về MLPRegressor

```
### Cai dat thuat toan

#####  Chay voi thuat toan Neural Network Regression
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Create MLPRegressor object
mlp = MLPRegressor()

# Train the model using the training sets
mlp.fit(X_train, y_train)

# Score the model
neural_network_regression_score = mlp.score(X_test, y_test)

neural_network_regression_score

# Make predictions using the testing set
nnr_pred = mlp.predict(X_test)
```

Hình 2.14: Ví dụ MLPRegressor

c) Ứng dụng của Nerual network:

Không thể phủ nhận được những thành công của Nerual network, đặc biệt là trong Deep Learning. Một số ứng dụng sử dụng Nerual network:

- Self-driving cars
- Voice Search & Voice-Activated Assistants
- Automatic Machine Translation
- Automatic Text Generation
- Nhận dạng ảnh (Image Recognition)

2.3 Kết luận chương 2

Trong chương 2, luận văn đã trình bày tổng quan về học máy, phân loại các giải thuật trong học máy, và một số thuật toán hay được sử dụng trong học máy. Dựa trên cách phân loại giải thuật cùng với các tiêu chí mong muốn của đề bài, học viên lựa chọn thuật toán phù hợp để giải quyết yêu cầu của bài toán. Trong chương 3, luận văn sẽ ứng dụng lý thuyết ở chương 2, chạy thử nghiệm và đánh giá kết quả đầu ra của bài toán.

CHƯƠNG 3. THỬ NGHIỆM VÀ ĐÁNH GIÁ

Chương 3 của luận văn sẽ nghiên cứu đưa ra cách thức xây dựng bộ dữ liệu, đồng thời đưa cài đặt thuật toán học máy cho dự đoán kết quả. Kết quả thử nghiệm được so sánh đối chiếu với giá trị thực tế để có nhận xét đánh giá độ phù hợp.

3.1 Xây dựng bộ dữ liệu

Trên thực tế, nhiều yếu tố ảnh hưởng, tác động đến tỷ giá hối đoái. Trong phạm vi lĩnh vực tài chính ngân hàng, các yếu tố chính ảnh hưởng đến tỷ giá USD/VNĐ là giá vàng thế giới, giá dầu thô và chỉ số tiêu dùng CPI. Đây là lý do học viên lựa chọn các biến của dữ liệu thô cho luận văn này. Với các biến đã được lựa chọn, chi phí và chất lượng của dữ liệu cần được xem xét trong quá trình thu thập dữ liệu. Bốn vấn đề cần được xem xét trong quá trình lựa chọn dữ liệu đó là (1) phương pháp tính toán, (2) dữ liệu không thể sửa đổi trở về trước, (3) sự chậm trễ thích hợp của dữ liệu và (4) đảm bảo rằng nguồn sẽ tiếp tục cung cấp dữ liệu trong tương lai. Trên cơ sở đó, học viên xác định sẽ lấy bộ dữ liệu trong khoảng 04/05/2015 đến ngày 04/05/2020. Đây sẽ là yếu tố ảnh hưởng đến độ dốc của dữ liệu và thời gian tính toán của các mô hình.

Đặc điểm mô tả của tập dữ liệu như sau, tập dữ liệu gồm tỷ giá USD/VND, giá vàng thế giới, giá dầu thô, chỉ số tiêu dùng CPI và ngày giao dịch trong khoảng thời gian 5 năm từ 04/05/2015 đến 04/05/2020. Trong đó, tỷ giá USD/VNĐ là dữ liệu tỷ giá bán ra của ngân hàng đối với ngoại tệ USD. Dữ liệu ở đây lấy vào thời điểm cuối ngày, được chốt trước hết phiên giao dịch trong ngân hàng. Các dữ liệu giá vàng thế giới, giá dầu thô, chỉ số tiêu dùng CPI được coi là 03 thuộc tính có tính tương quan, đi cùng trong quá trình dự đoán tỷ giá USD/VNĐ

3.1.1 Dữ liệu Tỷ giá USD/VNĐ

Nguồn dữ liệu được lấy ở ngân hàng TMCP ĐT& PT BIDV [15].

```
Du lieu dau vao USD / VND:
Du lieu 5 dong dau tien:
        Date     Data
0   5/4/2020  23446.0
1  4/29/2020  23429.0
2  4/28/2020  23445.0
3  4/27/2020  23456.5
4  4/24/2020  23510.0
.....
Du lieu 3 dong cuoi cung:
          Date     Data
1300  5/6/2015  21645.0
1301  5/5/2015  21645.0
1302  5/4/2015  21630.0
Tong so ban ghi : 1303
```

Hình 3.1: Dữ liệu Tỷ giá USD/VNĐ

Thông tin dữ liệu được đặt trong file USD_VND_Original_v1.csv của thư mục: ForexAI\DATA_RAW\

Trong đó dữ liệu Date: được thiết kế theo định dạng mm/dd/yyyy được mô tả như trong Hình 3.1.

Data: là số liệu tỷ giá USD/VNĐ ở thời điểm cuối ngày trước khi chốt phiên làm việc của ngày tương ứng

3.1.2 Dữ liệu giá vàng

Nguồn dữ liệu được lấy ở thị trường tài chính [17].

```
Du lieu dau vao ty gia VANG:
Du lieu 5 dong dau tien:
        Date     Data
0   5/4/2020  1701.69
1   5/1/2020  1700.41
2  4/30/2020  1685.05
3  4/29/2020  1712.39
4  4/28/2020  1708.64
.....
Du lieu 3 dong cuoi cung:
          Date     Data
1300  5/6/2015  1191.80
1301  5/5/2015  1193.33
1302  5/4/2015  1188.25
Tong so ban ghi : 1303
```

Hình 3.2: Dữ liệu giá vàng

Thông tin dữ liệu được đặt trong file XAU_USD_Original_v1.csv của thư mục: ForexAI\DATA_RAW\

Trong đó dữ liệu Date: được thiết kế theo định dạng mm/dd/yyyy yyyy được mô tả như trong Hình 3.2.

Data: là số liệu giá vàng của ngày tương ứng. Luận văn sử dụng dữ liệu của vàng thế giới, nên đây là tỷ giá của Vàng / USD

3.1.3 Dữ liệu giá dầu

Nguồn dữ liệu được lấy ở thị trường tài chính [18]

Thông tin dữ liệu được đặt trong file WTI_USD_Original_v1.csv của thư mục: ForexAI\DATA_RAW\

Trong đó dữ liệu Date: được thiết kế theo định dạng mm/dd/yyyy được mô tả trong Hình 3.3.

Data: là số liệu giá dầu thô của ngày tương ứng. Luận văn sử dụng dữ liệu dầu thế giới, nên đây là tỷ giá của Dầu/USD

```
Du lieu dau vao ty gia dau tho:
Du lieu 5 dong dau tien:
        Date   Data
0   5/4/2020  20.39
1   5/1/2020  19.78
2  4/30/2020  18.84
3  4/29/2020  15.06
4  4/28/2020  12.34
.....
Du lieu 3 dong cuoi cung:
          Date   Data
1317  5/6/2015  60.93
1318  5/5/2015  60.40
1319  5/4/2015  58.93
Tong so ban ghi : 1320
```

Hình 3.3: Dữ liệu giá dầu

3.1.4 Dữ liệu chỉ số tiêu dùng

Nguồn dữ liệu được lấy ở Tổng cục thống kê [16].

```
Du lieu dau vao Chi so tieu dung:
Du lieu 5 dong dau tien:
        Date    Data
0   5/4/2020  0.0000
1   5/1/2020  0.0000
2  4/30/2020 -0.0154
3  4/29/2020 -0.0154
4  4/28/2020 -0.0154
.....
Du lieu 3 dong cuoi cung:
          Date    Data
1324  5/6/2015  0.0016
1325  5/5/2015  0.0016
1326  5/4/2015  0.0016
Tong so ban ghi : 1327
```

Hình 3.4: Dữ liệu chỉ số tiêu dùng CPI

Thông tin dữ liệu được đặt trong file CPI_Original_v1.csv của thư mục: ForexAI\DATA_RAW\

Trong đó dữ liệu Date: được thiết kế theo định dạng mm/dd/yyyy được mô tả trong Hình 3.4.

Data: là số liệu CPI

3.2 Cài đặt thuật toán học máy

Học viên lựa chọn phần mềm Pycharm để viết chương trình với ngôn ngữ sử dụng python cùng các thư viện hỗ trợ như pandas, matplotlib.pyplot, numpy, seaborn và các mô hình học máy của Sklearn.

Các bước thực hiện trong cài đặt thuật toán học máy như sau:

- Chuẩn hóa và import dữ liệu đầu vào
- Tạo khung dữ liệu Data Frame
- Xác định mục tiêu chu kỳ cần dự đoán
- Thể hiện tính tương quan giữa các thuộc tính bằng biểu đồ cặp
- Xây dựng Model (X,Y) xác định tỷ lệ huấn luyện và kiểm thử
- Tiền xử lý dữ liệu mục đích để giảm thiểu độ nhiễu của dữ liệu thô
- Áp dụng mô hình (trình bày ở phần 3.3)
- Kết quả sau khi áp dụng (trình bày ở phần 3.3)
- Đánh giá kết quả mô hình (trình bày ở phần 3.3)

3.2.1 Chuẩn hóa và import dữ liệu đầu vào

- File đầu vào (có dạng đuôi file.csv)

- Dữ liệu chi tiết gồm 2 trường Date và Data

- Kiểu dữ liệu: trường Date có formate (mm/dd/yyyy) , trường Data là kiểu số

- Thực hiện chuẩn hóa: Dữ liệu Date trong các file thuộc tính là không giống nhau về tổng số bản ghi. Do vậy, ta tạo 1 danh sách chứa toàn bộ bản ghi của các ngày. Các ngày dữ liệu không có sự trùng lặp. Dữ liệu Data của các trường được bổ sung bằng cách lấy dữ liệu ngày gần nhất của trường đó. Cách thức thực hiện chuẩn hóa dữ liệu được biểu diễn như trong Hình 3.5.

- Thực hiện Import toàn bộ dữ liệu của các vào 1 file. Chú ý: trong file chỉ có 1 trường Date duy nhất như trong Hình 3.6.

- Kết quả thực sau khi thực hiện gộp toàn bộ dữ liệu sẽ được thể hiện trong Hình 3.7.

Ví dụ: thuộc tính vàng, có dữ liệu ban đầu : ở ngày 05/20/2019 là 1650.24 và ngày 05/24/2019 là 1675.49, khi so sánh với danh sách chứa toàn bộ các ngày, thì cần bổ sung thêm ngày 05/22/2019 và 05/23/2019. Do vậy, dữ liệu của ngày 05/22/2019 là 1650.24 (bằng dữ liệu của ngày 05/20/2019), tương tự với 05/23/2019 là 1650.24

```
def convert_StandardValue(dfSource1, dfSource2):
    listUsd = []

    if (dfSource1.shape[0] < dfSource2.shape[0]):
        return listUsd
    # elif (dfSource1.shape[0] == dfSource2.shape[0]):
    #     return dfSource2
    else:
        # lay totalRow cua Source 1 va Source 2
        i = dfSource1.shape[0] - 1
        k = dfSource2.shape[0] - 1

        while (i >= 0 and k >= 0):
            if (dfSource1["Date"].values[i] == dfSource2["Date"].values[k]):
                listUsd.append(dfSource2["Data"].values[k])
                i = i -1
                k = k -1
            else:
                listUsd.append(dfSource2["Data"].values[k])
                i = i -1
        listUsd.reverse()
    return listUsd
```

Hình 3.5: Hàm chuẩn hóa dữ liệu

```
## Tao. DataFrame va merge cac bang du lieu
#  Sau do tao ra 1 file csv dc cau' hinh tu nhieu bang du lieu
df = df_usd.merge(df_wti, how ='left',left_index=True, right_index=True)
df = df.merge(df_xau, how ='left',left_index=True, right_index=True)
df = df.merge(df_cpi, how ='left',left_index=True, right_index=True)
df.head()
df.tail()
df.to_csv('DATA_COMBINE/data.csv')
```

Hình 3.6: Cách thức gộp dữ liệu

```
Du lieu sau khi gop:
Du lieu 5 dong dau tien:
         Date  Data_USD  Data_WTI  Data_XAU  Data_CPI
0  2020-05-04   23446.0     20.39   1701.69    0.0000
1  2020-05-01   23446.0     19.78   1700.41    0.0000
2  2020-04-30   23446.0     18.84   1685.05   -0.0154
3  2020-04-29   23429.0     15.06   1712.39   -0.0154
4  2020-04-28   23445.0     12.34   1708.64   -0.0154
.....
Du lieu 3 dong cuoi cung:
            Date  Data_USD  Data_WTI  Data_XAU  Data_CPI
1324  2015-05-06   21645.0     60.93   1191.80    0.0016
1325  2015-05-05   21645.0     60.40   1193.33    0.0016
1326  2015-05-04   21630.0     58.93   1188.25    0.0016
Tong so ban ghi : 1327
```

Hình 3.7: Dữ liệu sau khi gộp

3.2.2 Tạo khung dữ liệu Data Frame

Dựa trên dữ liệu sau khi gộp, ta sẽ tạo ra khung dữ liệu DataFrame

```
plt.style.use('mds.mplstyle')
# df = pd.read_csv('DATA_COMBINE/data.csv', thousands=',', parse_dates=['Date'], dayfirst=True)
df = pd.read_csv('DATA_COMBINE/data.csv', thousands=',')
df = df.set_index('Date')
df.tail()
```

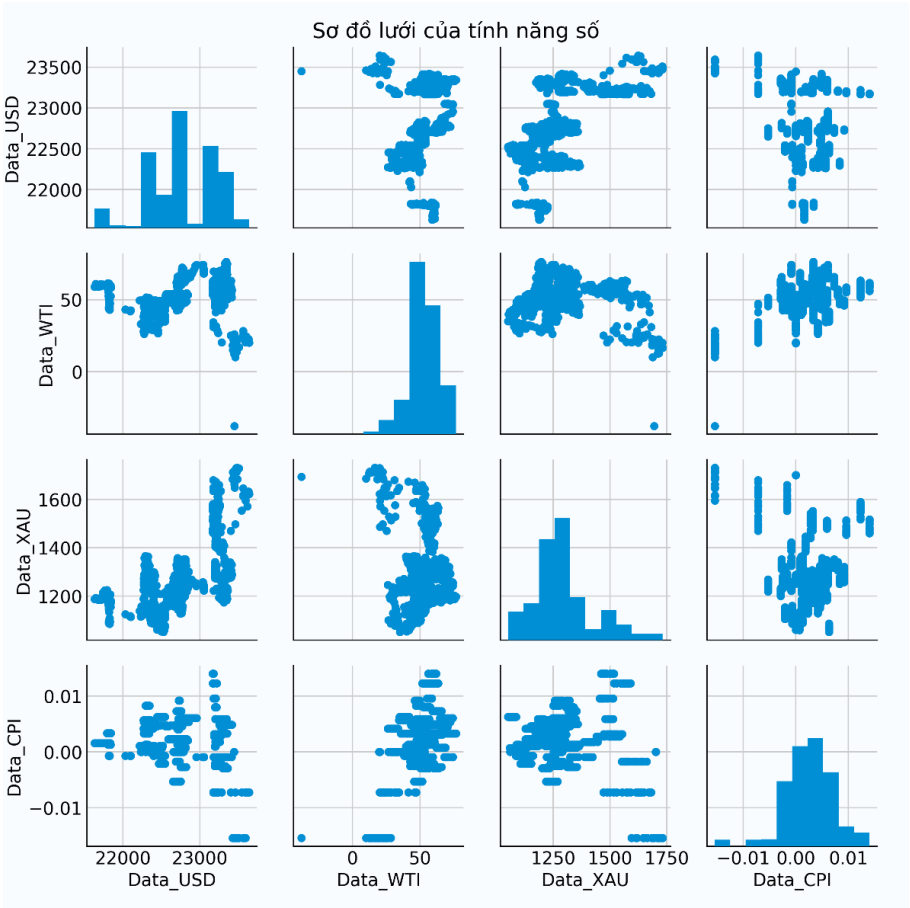
Hình 3.8: Tạo khung dữ liệu

3.2.3 Xác định mục tiêu chu kỳ cần dự đoán

Mục tiêu của bài toán là dự đoán theo ngày. Trong luận văn này, ta dự đoán chu kỳ là 1 ngày.

3.2.4 Thể hiện tính tương quan giữa các thuộc tính bằng biểu đổ cặp

Biểu đồ cặp thể hiện tương quan giữa các thuộc tính được thể hiển ở Hình 3.9



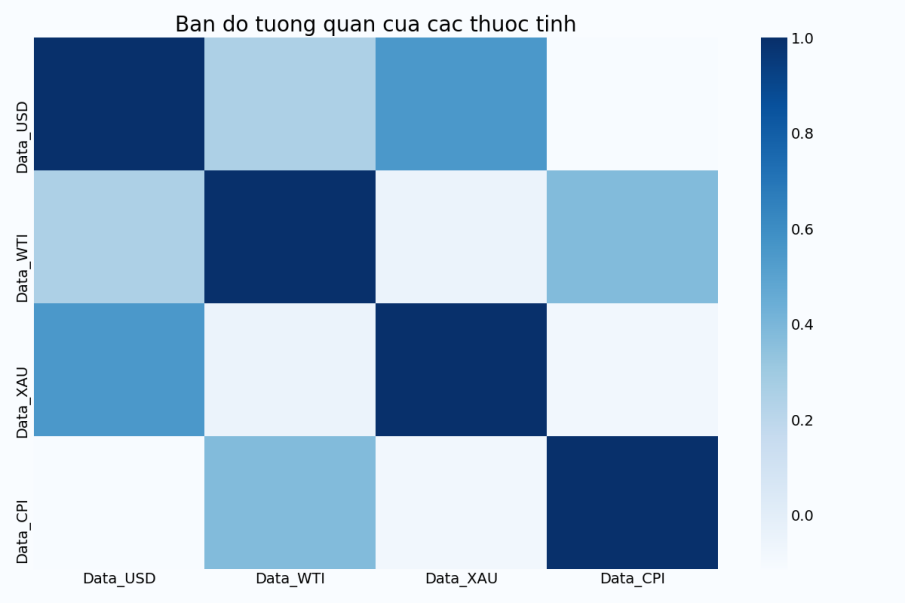
Hình 3.9: Cặp bảng lưới của thuộc tính số

Trong sơ đồ lưới tính năng số ở Hình 3.9, cho ta góc nhìn mối quan hệ giữa các thuộc tính. Cụ thể như sau:

- Xét cặp USD và CPI, ta nhận thấy khi dữ liệu USD trong khoảng từ 22000 đến 23000, thì chỉ số CPI tập trung phần nhiều ở ngưỡng 0 và 0.01. Trong dải USD từ trên 23000, thì CPI tập trung phần nhiều ở mức quanh 0 và lớn hơn 0 đôi chút (vẫn nhỏ 0.01)
- Xét cặp USD và XAU, ta nhận thấy dữ liệu USD trong khoảng 22000 đến 2300 thì dữ liệu của XAU rơi tập trung chủ yếu từ 1100 đến gần 1400. Còn khi USD ở mức trên 2300 chút xíu, thì giá trị XAU là dải chạy liên tục từ 1200 đến 1600.
- Xét cặp USD và WTI, ta nhận thấy dữ liệu USD trong khoảng 22000 đến 2300 thì dữ liệu của WTI rơi tập trung chủ yếu từ 23 đến gần 60. Còn khi USD ở mức trên 2300 chút xíu, thì giá trị WTI là dải chạy liên tục từ 40

đến 600. Khi USD trên 2300 một khoảng lớn, thì giá WTI lại có xu hướng giảm từ 23 về gần 0.

- Xét cặp USD và USD, đây chính là biểu đồ chỉ số của USD theo dang cột
- Các cặp khác cũng được xét tương tự.



Hình 3.10: Tương quan giữa các thuộc tính

Hình 3.10 thể hiện tương quan giữa các thuộc tính. Các vùng (ô) rất sáng và rất tối cho thấy mối tương quan tích cực hoặc tiêu cực giữa các thuộc tính. Ô mầu càng đậm thì tương quan càng tích cực, ô màu nhạt thì tương quan ngược lại. Cụ thể hơn:

-Trong cặp USD và CPI, ta nhận thấy màu ô giá tri rất sáng, thể hiện mối tương quan tiêu cực giữa hai đại lượng này. Mối quan hệ giữa USD và CPI là lỏng lẻo nhất.

- Trong cặp USD và XAU, ta thấy vùng khá đậm, cho thấy có sự ràn buộc giữa hai đại lượng này khá tích cực.

- Trong cặp USD và WTI, ta nhận thấy vùng màu đậm ít hơn so với vùng màu của USD và XAU, chứng tỏ quan hệ giữa đại USD và WTI không tích cực bằng USD và XAU.

- Cặp USD và USD, thì đơn giản là mối quan hệ với chính bản thân nó, dó vậy ở đây là vùng màu tối nhất (đạt ở ngưỡng 1).

- Xét các cặp khác cũng tương tự. Chú ý rằng, đây chỉ là mối tương quan chứ không phải quan hệ nhân quả. Tương quan ở đây được số hóa trong khoảng từ 0 đến

1, tức là từ ít tương quan đến tương quan nhiều, hoặc có thể hiểu là từ tường quan tiêu cực đến tương quan tích cực.

3.2.5 Xây dựng Model (X,Y)

Với mục tiêu dự đoán chu kỳ theo ngày trong tương lai, ta tạo thêm độ trễ d1, d2, d3 tương ứng với lùi 1 ngày, lùi 2 ngày, lùi 3 ngày. Ta cũng bổ sung đỗ trễ tương ứng với các thuộc tính.

Model sẽ gồm 2 tập X và Y. Trong đó, Y là tập các dữ liệu chứa cột d1,d2, d3. Tập Y được thể hiện như trong Hình 3.12.

Tập X là toàn bộ dữ liệu các dòng, nhưng không chứa các cột d1,d2,d3 được thể hiện như trong Hình 3.13.

```
# phan chia du lieu trong X va Y
### DataFrame cua Y , gom cac dong , va bao gom cac cot. d1, d2, d3
y = df[['d1', 'd2', 'd3']]
print('Du lieu tap Y= ')
print(y)

### DataFrame cua X , gom cac dong , va khong  bao gom cac cot. d1, d2, d3
# xoa cot khong lam bien giai thich
X = df.drop(['d1', 'd2', 'd3'], axis=1)
print('Du lieu tap X=')
print(X)
```

Hình 3.11: Phân chia dữ liệu X,Y trong Model

```
Du lieu tap Y=
                d1       d2       d3
Date
2020-04-28  23456.5  23510.0  23486.0
2020-04-27  23510.0  23486.0  23480.0
2020-04-24  23486.0  23480.0  23461.0
2020-04-23  23480.0  23461.0  23451.0
2020-04-22  23461.0  23451.0  23425.0
...             ...      ...      ...
2015-05-13  21710.0  21710.0  21680.0
2015-05-12  21710.0  21680.0  21680.0
2015-05-11  21680.0  21680.0  21645.0
2015-05-08  21680.0  21645.0  21645.0
2015-05-07  21645.0  21645.0  21630.0

[1320 rows x 3 columns]
```

Hình 3.12: Dữ liệu tập Y

```
Du lieu tap X=
            Data_USD  Data_WTI  ...  Data_USD-lag2-diff  Data_USD-lag3-diff
Date                            ...
2020-04-28   23445.0     12.34  ...               -17.0                 0.0
2020-04-27   23456.5     12.78  ...                16.0               -17.0
2020-04-24   23510.0     16.94  ...                11.5                16.0
2020-04-23   23486.0     16.50  ...                53.5                11.5
2020-04-22   23480.0     13.78  ...               -24.0                53.5
...              ...       ...  ...                 ...                 ...
2015-05-13   21740.0     60.50  ...                30.0               -27.5
2015-05-12   21710.0     60.75  ...               -55.0                30.0
2015-05-11   21710.0     59.25  ...               -30.0               -55.0
2015-05-08   21680.0     59.39  ...                 0.0               -30.0
2015-05-07   21680.0     58.94  ...               -30.0                 0.0

[1320 rows x 26 columns]
```

Hình 3.13: Dữ liệu tập X

Toàn bộ dữ liệu sẽ được huấn luyện, theo tỷ lệ 80/20. Trong đó 80% dữ liệu sẽ được huấn luyện trainning và 20 % dữ liệu sẽ được kiểm thử. Đây cũng là cách thường dùng của các nhà nghiên cứu áp dụng trong việc lựa chọn mô hình với hiệu suất tốt nhất.

```
### (X,Y) la 1 tap cac bo Xi --> Yi , dua. tren do ta chay cac thuat toan de tim gia tri tuong lai


from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test= train_test_split(X, y, test_size=0.20, random_state=420)


##### chu y: test_size=0.20  , the hien ty le cua tap du lieu dua vao phan tach' thu? nghiem. (o day 20% , 80%)
#####  random_state=420  : kiem soat viec xao tron duoc ap dung cho du lieu truoc khi ap dung. phan` tach' , cho phep tai' tao.
#####  random_state=420 , chon. 420 vi tong so ban ghi trong file csv ( file dau vao) khoang 1300 , do do' lay 1/3 lam gia tri
```

Hình 3.14-a: Dữ liệu trong mô hình huấn luyện

Tiền xử lý dữ liệu là quá trình xử lý các mẫu dữ liệu đầu vào và đầu ra. Các biến đầu vào và đầu ra hiếm khi được đưa vào mạng ở dạng thô. Ở bước này, học viên thực hiện chuyển đổi các biến đầu vào và và đầu ra để giảm thiểu độ nhiễu, tiếng ồn, mục đích làm nổi bật các mối quan hệ quan trong, đồng thời làm phẳng sự phân bố biến đổi và để phát hiện các xu hướng. Trong đề tài này, dữ liệu thô được chia tỷ lệ -1 và 1. Dưới đây là tiền dữ liệu xử lý trong Linear Regression, các mô hình khác cũng được học viên áp dụng tương tự

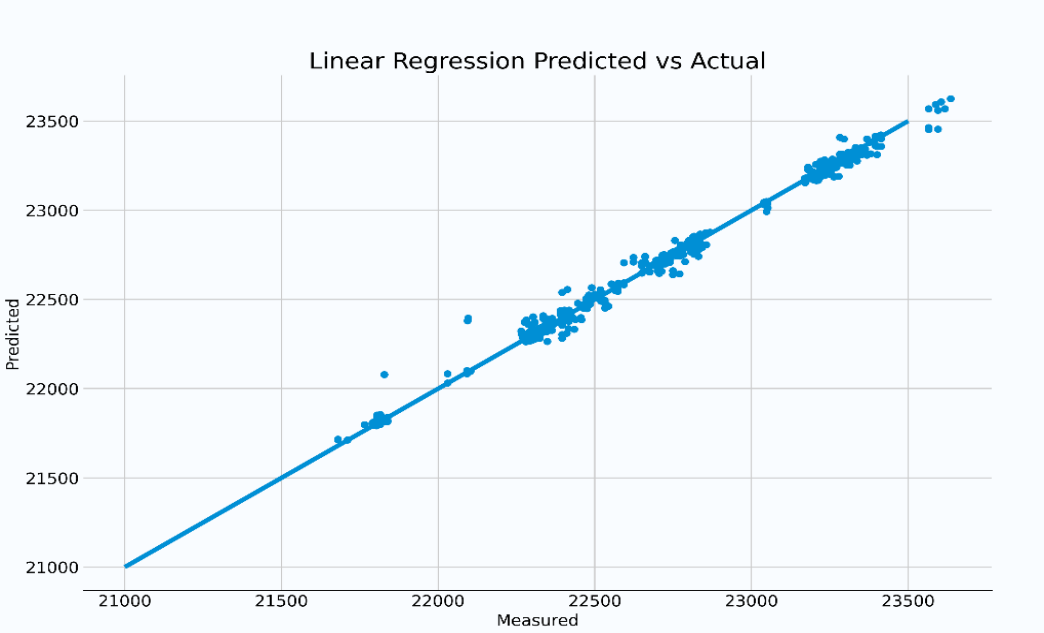
```
# Create linear regression object
regr = LinearRegression()

# Train the model using the training sets
regr.fit(X_train, y_train)

# Make predictions using the testing set
lin_pred = regr.predict(X_test)

linear_regression_score = regr.score(X_test, y_test)
linear_regression_score
```

Hình 3.14-b: Tiền xử lý dữ liệu trong Linear Regression

3.3 Thử nghiệm và đánh giá

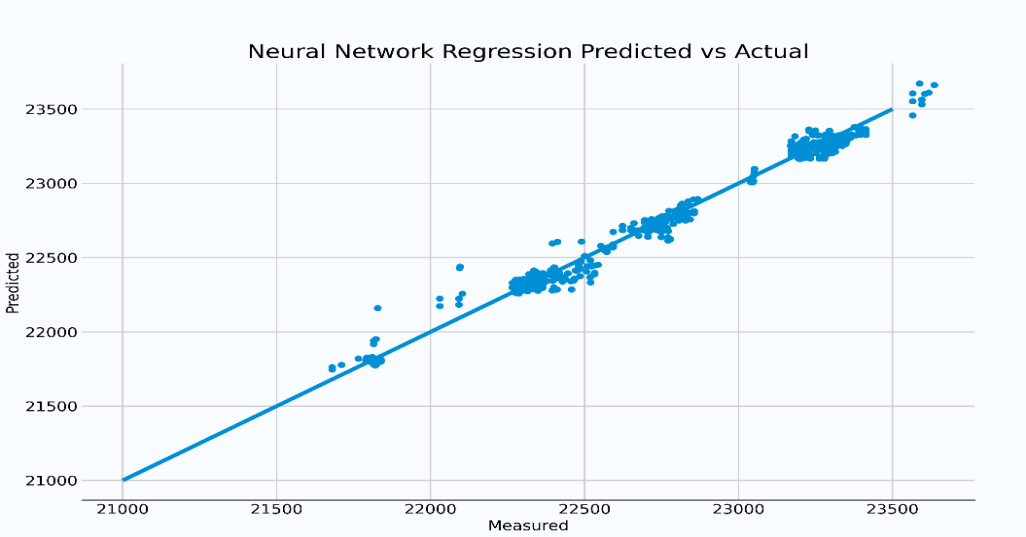
3.3.1 Nội dung thử nghiệm

Thuật toán Linear Regression (Hồi quy tuyến tính) được biểu diễn trong biểu đồ tần suất ở Hình 3.15.



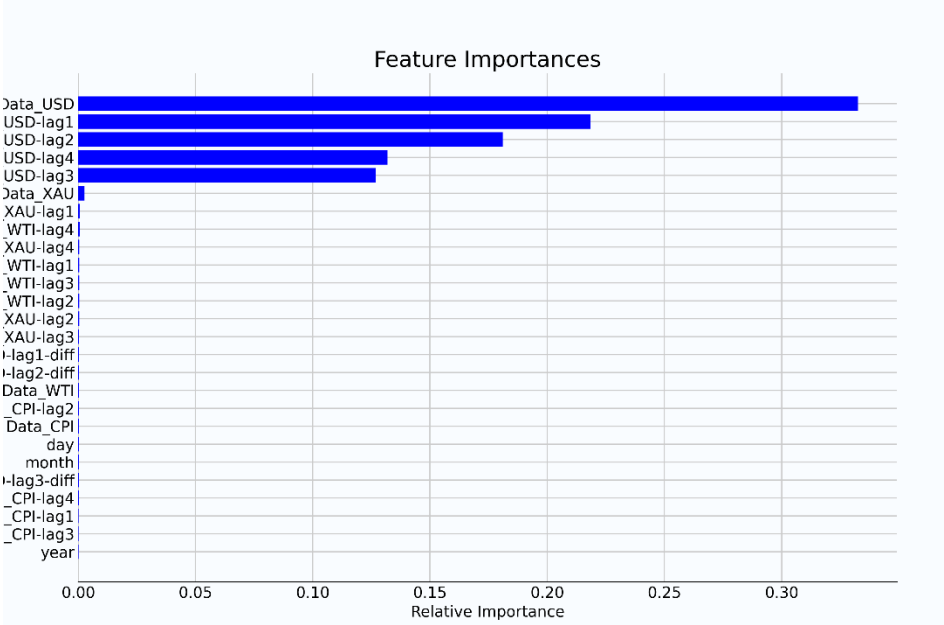
Hình 3.15: Biểu đồ tần suất của Linear Regression

Thuật toán Netural Network Regression có biểu đồ tần suất được thể hiện trong Hình 3.16



Hình 3.16: Biểu đồ tần suất của Netural Network

Ở thuật toán Random Forest, các thuộc tính quan trọng được biểu diễn như trong Hình 3.17



Hình 3.17: Biểu đồ thuộc tính quan trọng của Random Forest

3.3.2 Kết quả thử nghiệm và đánh giá

Một bước quan trọng việc xác định mô hình có phù hợp hay không trước khi đưa vào so sánh thực tế là xem xét, đánh giá các mức độ phù hợp thông qua một số các chỉ số. Ví dụ chỉ số MSE (Mean Square Error) có nghĩa là tính trung bình của bình phương sai số giữa giá trị thực tế và giá trị dự đoán. Một chỉ số khác cũng được

nhắc đến trong học máy khi đánh giá mô hình là MAE (Mean Absolute Error) là tính trung bình giá trị tuyệt đối sai số giữa giá trị thực tế và giá trị dự đoán. MAE được biết đến là mạnh mẽ hơn đối với các yếu tố ngoại lai (outliers) so với MSE. Một hệ số quan trọng nữa cần lưu ý là R^2 hay còn được biết đến với nhiều tên gọi như: R squared / R bình phương / coefficient of determination / hệ số xác định. Đây là một thước đo sự phù hợp của mô hình tuyến tính. Hệ số R square là hàm không giảm theo số biến độc lập được đưa vào mô hình, nếu chúng ta đưa thêm biến độc lập vào mô hình thì R^2 càng tăng. Tuy nhiên điều này cũng được chứng minh rằng không phải phương trình càng có nhiều biến thì càng tốt hơn. Một chỉ số rất quan trọng, hiện nay đang được sử dụng để đánh giá độ tin cậy của mô hình là RMSE (Root mean squared error). Lỗi trung bình bình phương (RMSE) là độ lệch chuẩn của phần dư (lỗi dự đoán). Phần dư là thước đo khoảng cách từ các điểm dữ liệu đường hồi quy. RMSE là thước đo mức độ lan truyền của những phần dư này. Nói cách khác, nó cho bạn biết mức độ tập trung của dữ liệu xung quanh dòng phù hợp nhất. Lỗi bình phương trung bình thường được sử dụng trong dự báo và phân tích hồi quy để xác minh kết quả thí nghiệm. Lỗi trung bình bình phương gốc (RMSE) là thước đo mức độ hiệu quả của mô hình của bạn. Nó thực hiện điều này bằng cách đo sự khác biệt giữa các giá trị dự đoán và giá trị thực tế. R-MSE càng nhỏ tức là sai số càng bé thì mức độ ước lượng cho thấy độ tin cậy của mô hình có thể đạt cao nhất.

Sau khi áp dụng, ta có kết quả đánh giá độ tin cậy của các mô hình như sau:

```
****** Neural Network Regression ******
Root mean squared error: 37.20
Mean absolute error: 21.22
R-squared: 0.99
****** Linear Regression ******
Root mean squared error: 33.81
Mean absolute error: 18.78
R-squared: 0.99
****** Random Forest Regression ******
```

Root mean squared error: 29.49

Mean absolute error: 16.82

R-squared: 1.00

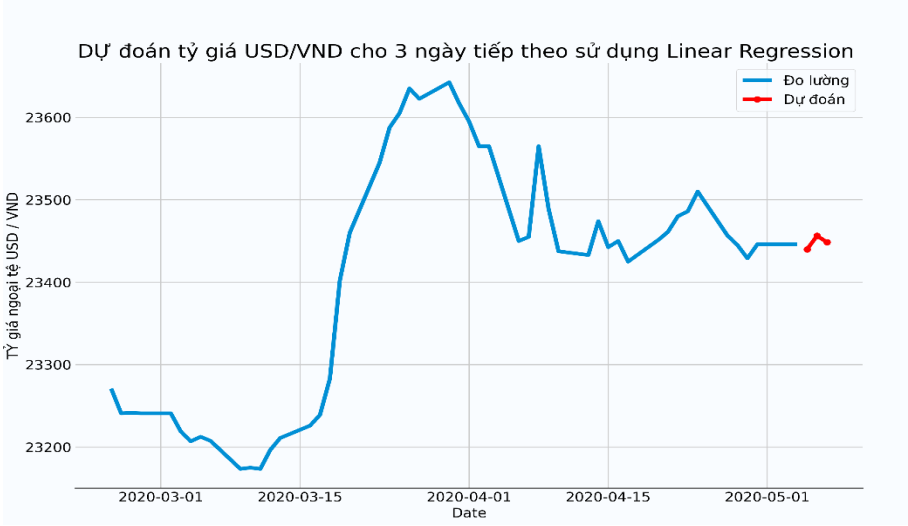
Theo kết quả trên, độ tin cậy phù hợp của các mô hình sẽ lần lượt là Random Forest với RMSE: 29.49, tiếp đến là Linear Regression với RMSE: 33.81 và cuối cùng là Neural Network Regression với RMSE: 37.20.

Tuy độ tin cậy của các mô hình đều đạt được kết quả tốt, nhưng ta vẫn cần phải xem xét độ hiệu quả trong thực tế, nhằm tránh rơi vào tình trạng overfitting (phù hợp quá mức) tức là kết quả huấn luyện và kiểm thử đều tốt, nhưng kết quả dự đoán có thể không phù hợp với mong đợi ở giá trị thực tế. Sau đây là kết quả sau khi chạy thử nghiệm với 3 thuật toán như sau:

a) Đối với Linear Regression

```
Du lieu du doan su dung Linear Regression:
        Date          Data
0 2020-05-05  23439.976540
1 2020-05-06  23456.221919
2 2020-05-07  23448.650078
```

Hình 3.18: Kết quả chạy của Linear Regression

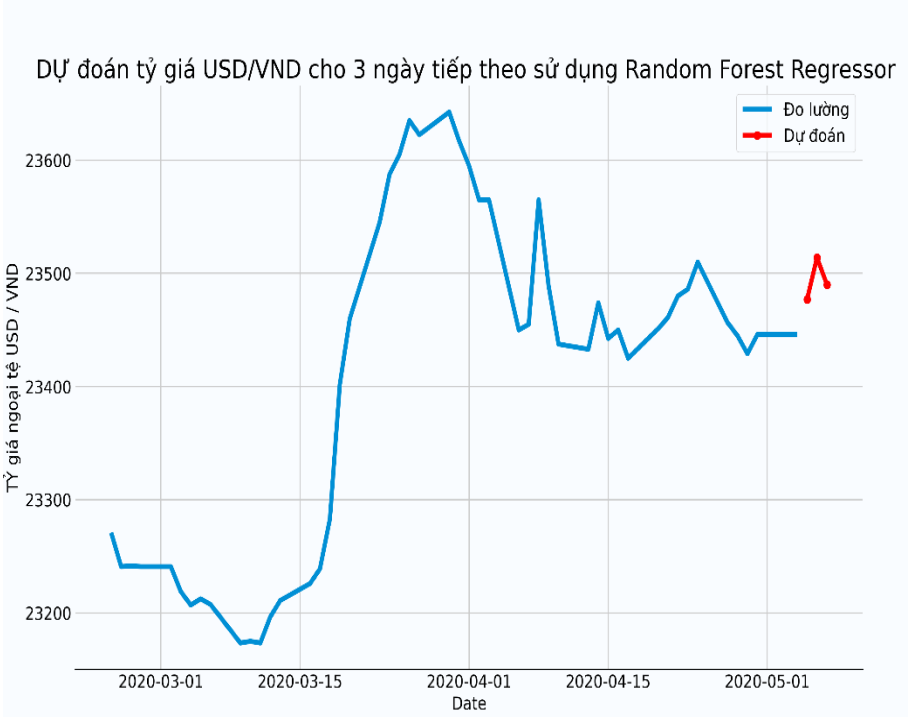


Hình 3.19: Sơ đồ biểu diễn kết quả chạy của Linear Regression

b) Đối với Random Forest

```
Du lieu du doan su dung Random Forest Regressor:
        Date        Data
0 2020-05-05  23477.0625
1 2020-05-06  23513.9425
2 2020-05-07  23490.0025
```

Hình 3.20: Kết quả chạy của Random Forest

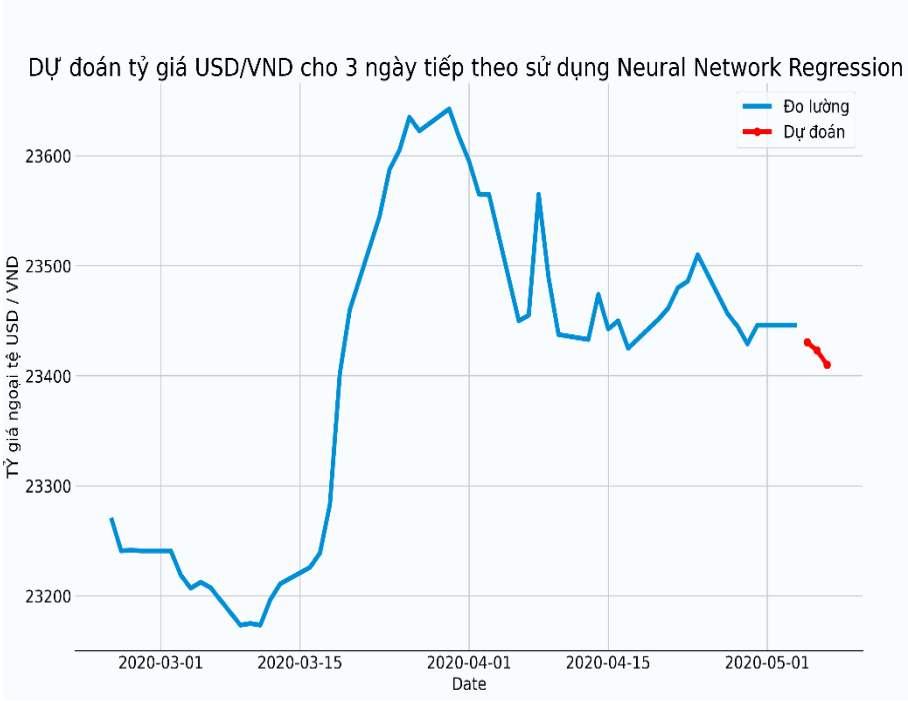


Hình 3.21: Sơ đồ biểu diễn kết quả chạy của Random Forest

c) Đối với Netural Network

```
Du lieu du doan su dung Neural Network Regression:
         Date          Data
0  2020-05-05  23430.497768
1  2020-05-06  23423.003220
2  2020-05-07  23410.174663
```

Hình 3.22: Kết quả chạy của Netural Network



Hình 3.23: Sơ đồ biểu diễn kết quả chạy của Netural Network

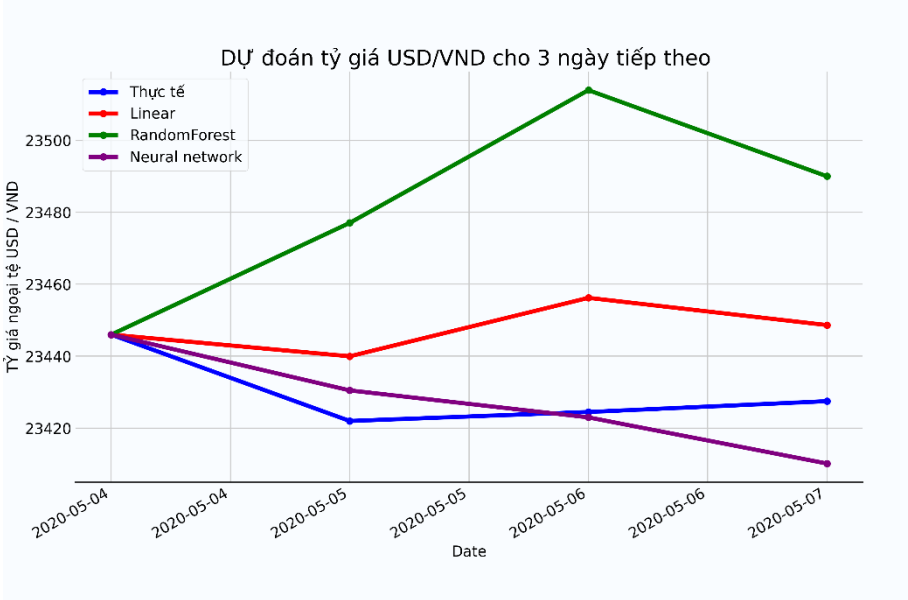
Đánh giá:

- So sánh kết quả đạt được với các giá trị thực tế :

Du lieu du thuc te:

	Date	Data
0	2020-05-04	23446.0
1	2020-05-05	23422.0
2	2020-05-06	23424.5
3	2020-05-07	23427.5

Hình 3.24: Dữ liệu thực tế của tỷ giá USD/VNĐ



Hình 3.25: So sánh kết quả với thực tế trong 3 ngày

- Trong Hình 3.24, kết quả xu hướng trong 01 ngày tiếp (ngày 05/05/2020) theo của Linear và Neural Network đang khá sát với thực tế đều có xu thế giảm với độ lệch lần lượt là 17 của Linear Regression và 8 của Neural Network. Riêng Random Forest thì lại có độ lệch lớn nhất là 55 đơn vị. Qua đó, ta nhận thấy khả năng dự đoán của Neural network có độ chính xác là cao nhất. Khi xem xét mở rộng hơn ở ngày tiếp theo 06/05/2020, dự đoán của Neural network vẫn đảm bảo được độ tin cậy với độ lệch là 1 đơn vị.
- Mỗi thuật toán đều đã cho ra kết quả. Khi so sánh kết quả dự báo với kết quả thực tế, kết quả dự báo có độ lệch khác khá xa (đặc biệt là ở Random Forest). Qua đây, ta cũng thấy được dù độ tin cậy của mô hình có tốt, nhưng vẫn phải bám sát thực tế để tránh rơi vào tình trạng overfitting.

Date	Dữ liệu thực tế	Dự đoán của LinearRegression	Dự đoán của RandomForest	Dự đoán của NeuralNetwork	Độ lệch LN	Độ lệch RF	Độ lệch NN
5/5/2020	23422	23439.97654	23477.0625	23430.49777	-17.97653979	-55.0625	**-8.497767784**
5/6/2020	23424.5	23456.22192	23513.9425	23423.00322	-31.72191861	-89.4425	1.496779833
5/7/2020	23427.5	23448.65008	23490.0025	23410.17466	-21.15007762	-62.5025	17.32533726

Hình 3.26: So sánh độ chênh lệch của kết quả đạt được với thực tế

Với kết quả như trên, chúng ta cần xem xét lại một số nguyên nhân có thể gây nên ảnh hưởng:

- Tính nhạy cảm của giá trị đầu vào (vì nếu xét dữ liệu trong 1 khoảng thời gian nhỏ, thì mật độ dữ liệu tương đối đều nhau. Tuy nhiên, khi xem xét trong 1 khoảng thời gian dài thì có độ chênh lệch khá lớn giữa giá trị lớn nhất và giá trị nhỏ nhất).
- Cần xem xét đánh giá lại độ lớn của dữ liệu, với khoảng thời gian như vậy liệu đã đủ phù hợp cho khả năng dự đoán.
- Cần xem xét lại các thuộc tính đưa vào, có thể thêm hoặc bớt để so sánh đánh giá tính phù hợp và ổn định của mô hình đem lại.

3.4 Kết luận chương 3

Chương 3 của luận văn đã xây dựng bộ dữ liệu, cài đặt và chạy chương trình cho ra được kết quả theo các thuật toán trong học máy. Dựa trên kết quả đạt được, ta thấy thuật toán neural network có độ chính xác cao nhất. Tuy nhiên, giá trị dự đoán vẫn còn có sự chênh lệch khác lớn ở kết quả đầu ra thực tế.

KẾT LUẬN

Kết quả dự kiến đạt được của luận văn:

Với mục tiêu nghiên cứu, áp dụng thuật toán trong học máy vào bài toán dự đoán tỷ giá USD/VNĐ, luận văn đã đạt được một số kết quả sau đây:

- Tổng quan về hệ thống học máy.
- Hiểu và áp dụng các thuật toán trong việc dự đoán tỷ giá USD/VNĐ.
- Kết quả của chương trình sẽ làm cơ sở xem xét, nâng cao hỗ trợ ra quyết định trong phán đoán xu hướng tăng giảm của thị trường ngoại tệ.

Hướng phát triển tiếp theo:

Học viên sẽ tiếp tục nghiên cứu, hoàn thiện, thử nghiệm với các tập dữ liệu và mô hình khác để tìm được giải pháp tối ưu hơn để có thể đưa kết quả gần sát với thực tế.

IV. DANH MỤC TÀI LIỆU THAM KHẢO

Tài liệu ngước ngoài

[1] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008) – *The Elements of Statistical Learning (2nd ed.)* – page 587 - 588

[2] Tom Mitchell - *Machine learning – page 2*

Tài liệu từ Internet:

[3] Anukrati Mehta, "An Ultimate Guide to Understanding Supervised Learning", trên trang: https://www.digitalvidya.com/blog/supervised-learning/ -Truy cập ngày:22/06/2020

[4] Anurag, "Random Forest Analysis in ML and when to use it", trên trang: https://www.newgenapps.com/blog/random-forest-analysis-in-ml-and-when-to-use-it/ -Truy cập ngày:08/10/2020

[5] Atul, "What is Machine Learning? Machine Learning For Beginners", trên trang: https://www.edureka.co/blog/what-is-machine-learning/ -Truy cập ngày:25/11/2020

[6] Edeane. trên trang: https://github.com/edeane/forex -Truy cập ngày:09/04/2020

[7] https://dominhhai.github.io/vi/2017/12/ml-intro/ -Truy cập ngày:07/05/2020

[8] https://en.wikipedia.org/wiki/Linear_regression -Truy cập ngày:08/05/2020

[9] https://en.wikipedia.org/wiki/Random_forest -Truy cập ngày:08/05/2020

[10] https://machinelearningcoban.com/2016/12/26/introduce/ -Truy cập ngày:09/05/2020

[11] https://machinelearningcoban.com/2016/12/28/linearregression/ -Truy cập ngày:10/05/2020

[12] https://machinelearningcoban.com/2017/02/24/mlp/ -Truy cập ngày:10/05/2020

[13] https://sonix.ai/articles/difference-between-artificial-intelligence-machine-learning-and-natural-language-processing -Truy cập ngày:15/08/2020

[14] https://thuvienphapluat.vn/van-ban/tien-te-ngan-hang/Luat-Ngan-hang-Nha-nuoc-1997-06-1997-QH10-41101.aspx -Truy cập ngày:03/10/2020

[15] https://www.bidv.com.vn/ -Truy cập ngày:15/11/2020

[16] https://www.gso.gov.vn/default.aspx?tabid=628 -Truy cập ngày:15/11/2020

[17] https://www.investing.com/currencies/xau-usd-historical-data -Truy cập ngày:15/11/2020

[18] https://www.investing.com/equities/w-t-offshore-inc-historical-data -Truy cập ngày:15/11/2020

[19] https://www.javatpoint.com/regression-vs-classification-in-machine-learning -Truy cập ngày:08/10/2020

[20] Jason Brownlee, "Difference Between Classification and Regression in Machine Learning", trên trang: https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/ -Truy cập ngày:15/05/2020

[21] Jason Brownlee, "Supervised and Unsupervised Machine Learning Algorithms", trên trang: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ -Truy cập ngày:15/05/2020

[22] Joseph Nguyễn, "3 Common Ways to Forecast Currency Exchange Rates", trên trang: https://www.investopedia.com/articles/forex/11/4-ways-to-forecast-exchange-rates.asp -Truy cập ngày:02/08/2021

[23] Lan Hương, "Phía sau việc giá vàng đắt chưa từng có", trên trang: http://tapchitaichinh.vn/ngan-hang/phia-sau-viec-gia-vang-dat-chua-tung-co-325433.html -Truy cập ngày:27/11/2020

[24] Matthew Boesler, "The Evolution Of The World's Currencies Since 1821 [Infographic]", trên trang: https://www.businessinsider.com/world-

currency-system-1821-infographic-2012-8#ixzz251EPIzIV -Truy cập ngày:10/06/2020

[25] Nguyễn Xuân Việt Cường, "Mạng Neural Network", trên trang: https://viblo.asia/p/mang-neural-network-WAyK84zpKxX -Truy cập ngày:20/05/2020

[26] Phạm Hải, "Machine learning là gì? Deep learning là gì? Sự khác biệt giữa AI, machine learning và deep learning", trên trang: https://quantrimang.com/su-khac-biet-giua-ai-hoc-may-va-hoc-sau-157948 -Truy cập ngày:11/09/2020

[27] Robert Ritz, "Forecasting USD-MNT Exchange Rate — Part 2: Machine Learning", trên trang: https://medium.com/mongolian-data-stories/forecasting-usd-mnt-exchange-rate-part-2-machine-learning-be00a765a741 -Truy cập ngày:09/04/2020

[28] Thanh Leo, "Random Forest và ứng dụng", trên trang: https://medium.com/@thanhleo92/random-forest-v%C3%A0-%E1%BB%A9ng-d%E1%BB%A5ng-b6965c1f0634 -Truy cập ngày:10/07/2020

[29] Tô Linh, "Tỷ giá hối đoái là gì? Những điều cơ bản bạn cần biết về ngoại tệ", trên trang: https://marketingai.admicro.vn/ty-gia-hoi-doai-la-gi/ -Truy cập ngày:08/06/2020

[30] TS. Cấn Văn Lực và Nhóm tác giả Viện Đào tạo và Nghiên cứu BIDV, "Giá dầu giảm sâu tác động thế nào đến kinh tế Việt Nam?", trên trang: https://cafef.vn/gia-dau-giam-sau-tac-dong-the-nao-den-kinh-te-viet-nam-20200331165853096.chn -Truy cập ngày:27/11/2020

[31] TS. Nguyễn Thị Kim Thanh, "Chỉ số CPI và diễn biến thị trường tiền tệ: Mục tiêu kép cần bảo vệ", trên trang: http://tapchitaichinh.vn/nghien-cuu-trao-doi/chi-so-cpi-va-dien-bien-thi-truong-tien-te-muc-tieu-kep-can-bao-ve-110128.html -Truy cập ngày:27/11/2020

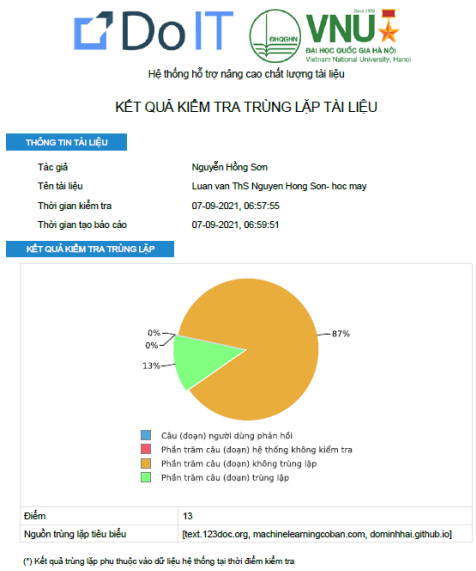
BẢN CẢM ĐOAN

Tôi cam đoan đã thực hiện việc kiểm tra mức độ tương đồng nội dung luận văn qua phần mềm DoIT một cách trung thực và đạt kết quả mức tương đồng % toàn bộ nội dung luận văn. Bản luận văn kiểm tra qua phần mềm là bản cứng của luận văn đã nộp để bảo vệ trước hội đồng. Nếu sai tôi xin chịu các hình thức kỷ luật hiện hành của Học viện.

Hà Nội, ngày tháng năm 2021
HỌC VIÊN

Nguyễn Hồng Sơn

Hình ảnh minh chứng kiểm tra kết quả trùng lặp dữ liệu



Hệ thống hỗ trợ nâng cao chất lượng tài liệu

KẾT QUẢ KIỂM TRA TRÙNG LẶP TÀI LIỆU

THÔNG TIN TÀI LIỆU

Tác giả	Nguyễn Hồng Sơn
Tên tài liệu	Luan van ThS Nguyen Hong Son- hoc may
Thời gian kiểm tra	07-09-2021, 06:57:55
Thời gian tạo báo cáo	07-09-2021, 06:59:51

KẾT QUẢ KIỂM TRA TRÙNG LẶP

Điểm	13
Nguồn trùng lặp tiêu biểu	[text.123doc.org, machinelearningcoban.com, dominhhai.github.io]

(*) Kết quả trùng lặp phụ thuộc vào dữ liệu hệ thống tại thời điểm kiểm tra

HỌC VIÊN **NGƯỜI HƯỚNG DẪN KHOA HỌC**

Nguyễn Hồng Sơn **TS. NGUYỄN VĂN THỦY**