

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC ĐẠI NAM



BÀI TẬP LỚN

TÊN HỌC PHẦN: HỌC MÁY
ĐỀ TÀI: DỰ BÁO GIÁ VÀNG BẰNG HỒI
QUY TUYỂN TÍNH

Giáo viên hướng dẫn: Tạ Đăng Chí

Sinh viên thực hiện:

STT	Mã SV	Họ và tên	Lớp
1	1451020181	Phạm Duy Quang	CNTT1401
2	1451020262	Trần Anh Tú	CNTT1401
3	1451020273	Nguyễn Xuân Viễn	CNTT1401

Hà Nội, ngày 20 tháng 10 năm 2023

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC ĐẠI NAM**



BÀI TẬP LỚN

**TÊN HỌC PHẦN: HỌC MÁY
ĐỀ TÀI: DỰ BÁO GIÁ VÀNG BẰNG HỒI
QUY TUYẾN TÍNH**

STT	Mã SV	Họ và Tên	Ngày Sinh	Điểm	
				Bảng Số	Bảng Chữ
1	1451020181	Phạm Duy Quang	24/08/2002		
2	1451020262	Trần Anh Tú	28/05/2002		
3	1451020273	Nguyễn Xuân Viễn	05/11/2002		

CÁN BỘ CHẤM THI

Hà Nội, ngày 20 tháng 10 năm 2023

LỜI MỞ ĐẦU

Ngày nay với sự phát triển mạnh mẽ của nền công nghệ mới, đặc biệt sau cuộc cách mạng công nghệ 4.0 nền công nghệ đã có những bước tiến vượt bậc khiến cho cuộc sống của con người ngày càng tiên tiến hơn và hiện đại hơn. Bất kỳ trong lĩnh vực nào, sự góp mặt của trí tuệ nhân tạo (AI) sẽ giúp con người làm việc và hoàn thành tốt công việc hơn. Với sự giúp đỡ của trí tuệ nhân tạo con người đã giảm thiểu đi rất nhiều công việc dư thừa và tiết kiệm thời gian trong việc xử lý thông tin với độ chính xác cao. Và gần đây, một thuật ngữ “machine learning” rất được nhiều người quan tâm. Thay vì phải code phần mềm với cách thức thủ công theo một bộ hướng dẫn cụ thể nhằm hoàn thành một nhiệm vụ đề ra thì máy sẽ tự “học hỏi” bằng cách sử dụng một lượng lớn dữ liệu cùng những thuật toán cho phép nó thực hiện các tác vụ và hoàn thành nhiệm vụ của mình.

"Vàng" một trong những kim loại quý quý giá nhất trên thế giới, luôn là một đối tượng quan tâm của sự quan tâm, đầu tư và phân tích. Giá vàng không chỉ phản ánh sự biến động của thị trường tài chính mà còn được xem là một chỉ số quan trọng về tình hình kinh tế và tâm lý của nhà đầu tư. Việc dự đoán giá vàng tương lai là một thách thức đáng kể, nhưng có thể giúp nhà đầu tư, doanh nghiệp, người tiêu dùng đưa ra quyết định thông minh.

Chúng em xin trình bày báo cáo này với mục tiêu sử dụng mô hình hồi quy tuyến tính để dự báo giá vàng. Chúng em đã sử dụng dữ liệu lịch sử về giá vàng cùng với các biến độc lập như thời gian, tỷ giá ngoại tệ, chỉ số kinh tế và các yếu tố khác có thể ảnh hưởng đến giá vàng. Bằng việc áp dụng các nguyên tắc của Machine Learning, chúng em đã cố gắng xây dựng một mô hình dự đoán giá vàng một cách hiệu quả.

Chúng em hi vọng rằng báo cáo này sẽ cung cấp một cái nhìn tổng quan về quá trình nghiên cứu và phát triển mô hình dự đoán giá vàng. Chúng em đã đối mặt với nhiều thách thức trong quá trình này, từ việc thu thập dữ liệu đến xây dựng mô hình và đánh giá hiệu suất. Bằng cách chia sẻ kinh nghiệm và kết quả của chúng em, chúng em hy vọng rằng đây sẽ là một bước đầu trong việc hiểu sâu hơn về cách sử dụng Machine Learning để dự đoán giá vàng. Qua đó, cũng phần nào cũng cố được kiến thức mà chúng em đã học được trên lớp qua quá trình học tập và thực hành.

LỜI CẢM ƠN

Chúng em xin chân thành cảm ơn tới sự hỗ trợ của thầy Tạ Đăng Chí đã có những lời nhận xét và góp ý để chúng em có thể hoàn thành tốt nhất bài báo cáo này. Và hi vọng rằng báo cáo này sẽ có ích cho tất cả những ai quan tâm đến việc dự đoán giá vàng và sử dụng Machine Learning trong lĩnh vực này.

MỤC LỤC

LỜI MỞ ĐẦU	1
LỜI CẢM ƠN.....	4
CHƯƠNG 1: TỔNG QUAN VỀ MACHINE LEARNING	7
1.1. Khái niệm	7
1.2. Các bước cơ bản trong Machine Learning	8
1.3. Phân nhóm các thuật toán học máy	9
1.3.1. Học máy có giám sát (Supervised Learning)	9
1.3.2. Học máy không giám sát (Unsupervised Learning).....	10
1.3.3. Học tăng cường (Reinforce Learning).....	11
1.3.4. Học bán giám sát (Semi – Supervised Learning)	11
1.3.5. Học sâu (Deep Learning).....	12
1.4. Các bài toán cơ bản trong Machine Learning	12
1.4.1. Các bài toán cơ bản	12
1.4.2. Một số bài toán khác.....	12
1.5. Ứng dụng của Machine Learning	13
CHƯƠNG 2: GIỚI THIỆU VỀ ĐỀ TÀI.....	15
2.1. Lý do chọn đề tài.	15
2.2. Giới thiệu về bài toán dự đoán giá vàng.....	16
2.3. Giới thiệu về các phương pháp Dự đoán giá vàng	17
2.4. Ý nghĩa của việc xây dựng mô hình học máy Dự đoán giá vàng.....	18
2.5. Các bước cơ bản Dự đoán giá vàng.....	19
2.5.1. Thu thập dữ liệu.....	19
2.5.2. Tiền xử lý dữ liệu	19

2.5.3. Chọn biến độc lập (Features).....	19
2.5.4. Tạo mô hình hồi quy tuyến tính	19
2.5.5. Huấn luyện mô hình	19
2.5.6. Đánh giá mô hình	20
2.5.7. Dự đoán giá vàng.....	20
2.5.8. Cập nhật và giám sát.....	20
2.6. Ý nghĩa của bài toán	20
CHƯƠNG 3: THUẬT TOÁN HỒI QUY TRONG DỰ BÁO GIÁ VÀNG	21
3.1. Thu thập dữ liệu.....	21
3.2. Phân tích dữ liệu	22
3.3. Chọn các biến độc lập (Features)	23
3.4. Xây dựng mô hình hồi quy tuyến tính.....	24
3.5. Huấn luyện mô hình	26
3.6. Đánh giá mô hình	27
3.7. Dự báo giá vàng.....	28
3.8. Cập nhật và giám sát.....	29
CHƯƠNG 4: KẾT QUẢ ĐẠT ĐƯỢC VÀ HƯỚNG PHÁT TRIỂN.....	30
4.1. Kết quả đạt được.....	30
4.2. Phương hướng phát triển	31
KẾT LUẬN	33
TÀI LIỆU THAM KHẢO	34

CHƯƠNG 1: TỔNG QUAN VỀ MACHINE LEARNING

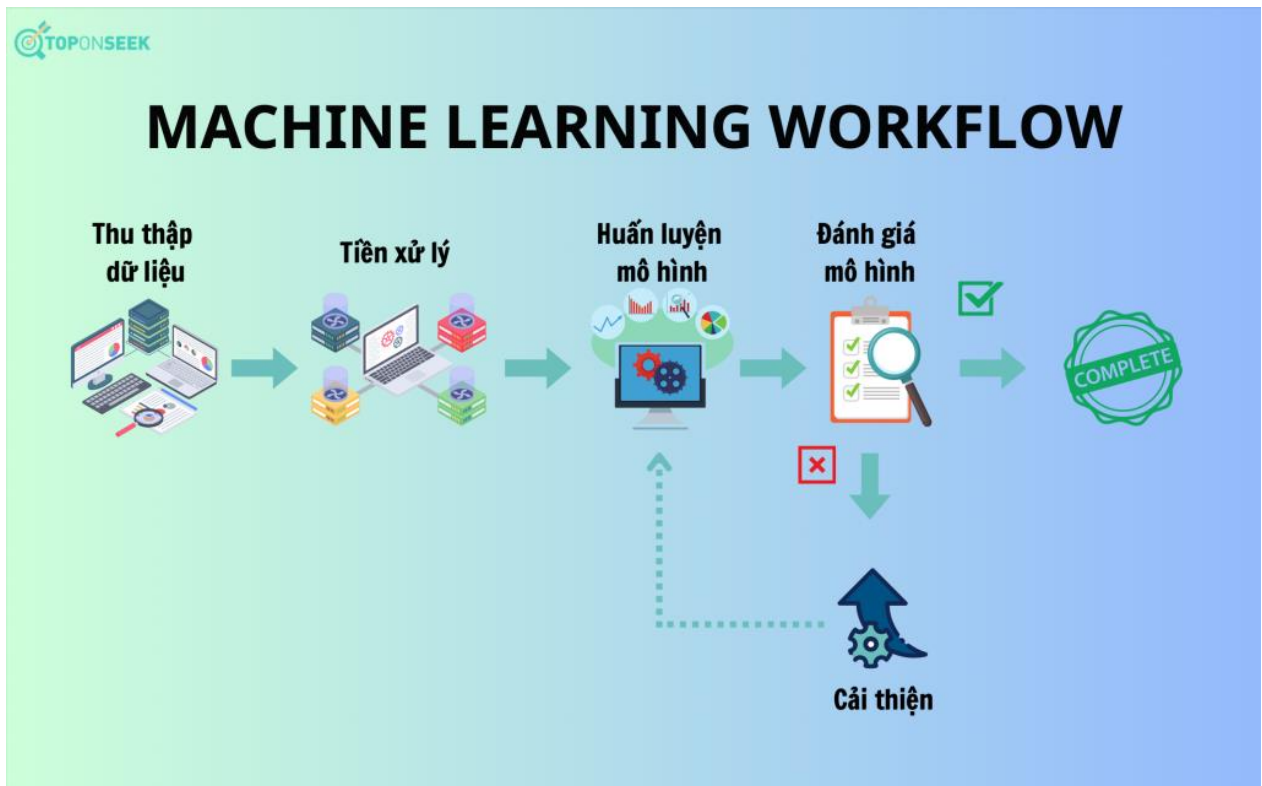
1.1. Khái niệm

Machine Learning (ML) hay học máy là một phần của trí tuệ nhân tạo AI (Artificial Intelligence), đây là lĩnh vực nghiên cứu cho phép máy tính có thể cải thiện bản thân của chúng dựa trên các dữ liệu mẫu là training data hay dựa vào kinh nghiệm là những gì đã được học trước đó. Machine Learning cũng có thể tự dự đoán cũng như đưa ra quyết định mà không cần lập trình hướng đi cụ thể, chi tiết.

Học máy vẫn đòi hỏi sự đánh giá của con người trong việc tìm hiểu dữ liệu cơ sở và lựa chọn các kỹ thuật phù hợp để phân tích dữ liệu. Đồng thời, trước khi sử dụng, dữ liệu phải sạch, không có sai lệch và không có dữ liệu giả.

Các mô hình học máy yêu cầu lượng dữ liệu đủ lớn để "huấn luyện" và đánh giá mô hình. Trước đây, các thuật toán học máy thiếu quyền truy cập vào một lượng lớn dữ liệu cần thiết để mô hình hóa các mối quan hệ giữa các dữ liệu. Sự tăng trưởng trong dữ liệu lớn (big data) đã cung cấp các thuật toán học máy với đủ dữ liệu để cải thiện độ chính xác của mô hình và dự đoán.

1.2. Các bước cơ bản trong Machine Learning



Thu thập dữ liệu

Để máy tính có thể học và đưa ra các dự đoán, phân tích, lập trình viên cần cung cấp một bộ dữ liệu gọi là Dataset cho máy. Thông thường, bạn có thể thu thập các dữ liệu này hoặc sử dụng các Dataset có sẵn trên các nền tảng hỗ trợ học lập trình học máy.

Cần lưu ý lựa chọn những bộ dữ liệu từ những nguồn chính thống, như vậy máy tính mới có thể học được một cách chính xác và đưa ra những kết quả đúng đắn, có tỷ lệ hiệu quả cao hơn.

Tiền xử lý

Bước tiền xử lý trong Machine Learning dùng để chuẩn hóa các dữ liệu vừa thu thập được, giúp loại bỏ các thuộc tính không cần thiết, những dữ liệu bị hỏng, thiếu. Đồng thời bước này sẽ tiến hành gán nhãn, mã hóa các đặc trưng, trích xuất những đặc trưng và rút gọn bộ dữ liệu mà vẫn đảm bảo kết quả đầu ra.

Bước Preprocessing chiếm thời gian nhất trong toàn bộ workflow, tỷ lệ thuận với độ lớn, khối lượng dữ liệu mà bạn cung cấp. Từ đó, tổng thời gian thực hiện hai bước 1 và 2 chiếm tổng thời gian khoảng 70% toàn quá trình.

Huấn luyện mô hình

Bước huấn luyện mô hình này sử dụng để cho máy học trên dữ liệu mà bạn cung cấp và tiến hành xử lý ở hai bước đầu tiên.

Đánh giá mô hình

Sau khi đã tiến hành huấn luyện mô hình, bước tiếp theo trong Machine Learning đó là đánh giá mô hình vừa tạo ra. Tùy thuộc vào từng các loại độ đo khác nhau mà mô hình vừa huấn luyện được đánh giá là tốt hay không tốt khác nhau. Về cơ bản, độ chính xác của mô hình vừa huấn luyện đạt trên 80% được cho là đảm bảo hiệu quả.

Cải thiện

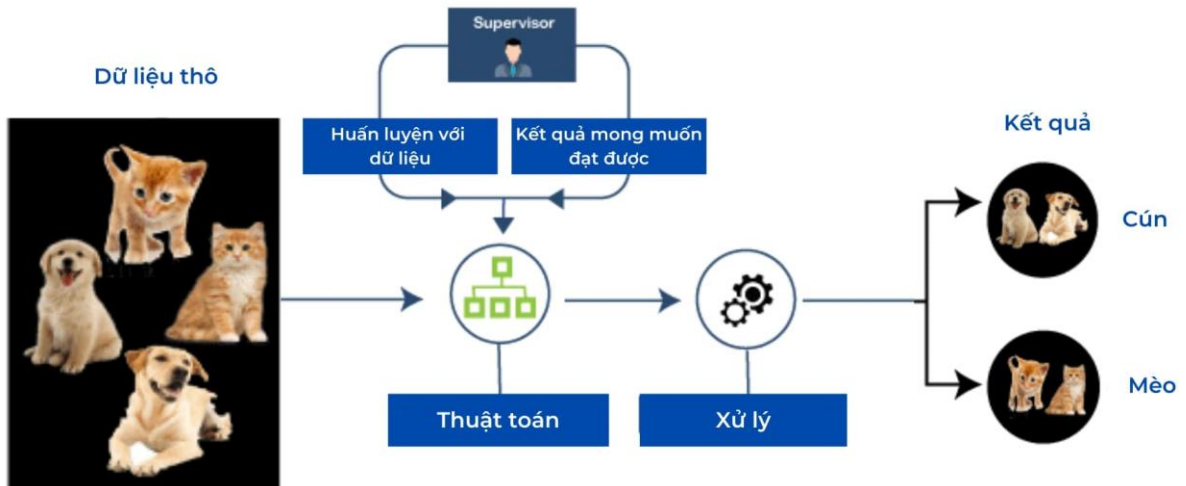
Trong bước cải thiện này, những mô hình sau khi đã được đánh giá nếu không đạt chuẩn thì sẽ được tiến hành lại bước thứ 3 cho đến khi độ chính xác đạt đúng kỳ vọng cần thiết. Ba bước cuối của Machine Learning Workflow là khoảng 30% tổng quá trình.

1.3. Phân nhóm các thuật toán học máy

1.3.1. Học máy có giám sát (Supervised Learning)

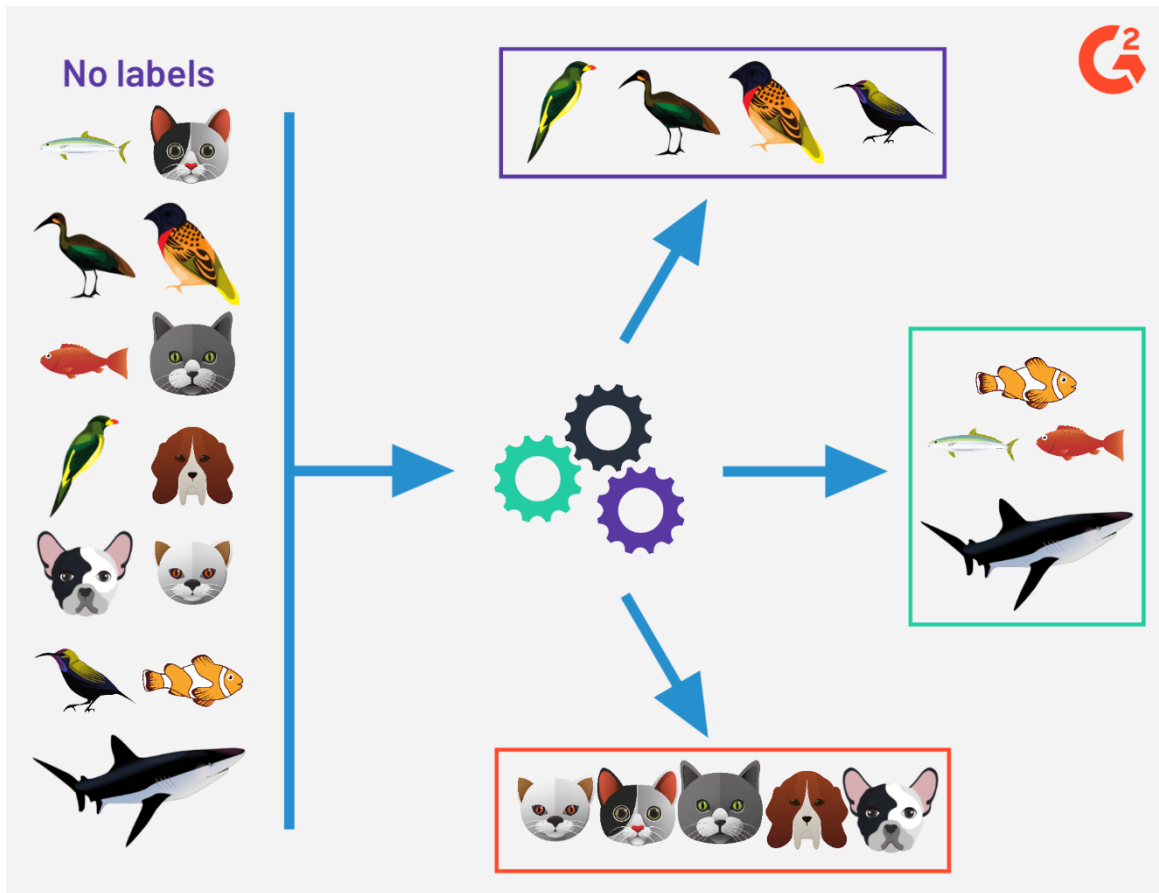
Học máy có giám sát hay Supervised learning là việc để cho máy tiến hành học trên dữ liệu đã được gán nhãn (label) trước đó, nói đơn giản đó là mỗi đầu vào A1 sẽ có đầu ra tương ứng là B1. Các dữ liệu này được gọi là dữ liệu huấn luyện và chúng là cặp các đầu vào-đầu ra. Học có giám sát sẽ xem xét các 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. Các mô hình như mạng Nơ-ron, SVM, CNN...

Học máy có giám sát



1.3.2. Học máy không giám sát (Unsupervised Learning)

Học không có giám sát hay Unsupervised learning là việc cho máy tiến hành học trên những dữ liệu cung cấp chưa được gán nhãn (label). Những thuật toán sử dụng trong Machine Learning sẽ tìm ra điểm tương quan dữ liệu, thực hiện mô hình hóa dữ liệu, nói cách khác là làm cho máy có kiến thức về dữ liệu. Từ đó, máy sẽ phân loại được dữ liệu thành các nhóm tương đồng nhau, trong đó chúng đã được học hay giảm số chiều của dữ liệu. Ví dụ như phân cụm dữ liệu, triết xuất thành phần chính của một chất nào đó. K mean Ứng dụng phổ biến nhất của học không giám sát là gom cụm (cluster).



Trong thuật toán này, chúng ta không biết được dữ liệu đầu ra hay nhãn mà chỉ có dữ liệu đầu vào. Thuật toán Học không giám sát dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó, ví dụ như phân nhóm hoặc giảm số chiều của dữ liệu để thuận tiện trong việc lưu trữ và tính toán.

1.3.3. Học tăng cường (Reinforce Learning)

Phương pháp học tăng cường tập trung vào việc làm sao để cho 1 tác tử trong môi trường có thể hành động sao cho lấy được phần thưởng nhiều nhất có thể. 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.

1.3.4. Học bán giám sát (Semi – Supervised Learning)

Các bài toán khi chúng ta có một lượng lớn dữ liệu X nhưng chỉ một phần trong chúng được gán nhãn được gọi là Semi-Supervised Learning. Những bài toán thuộc nhóm này nằm giữa hai nhóm được nêu bên trên. Một ví dụ điển hình của nhóm này là chỉ có một

phần ảnh hoặc văn bản được gán nhãn (ví dụ bức ảnh về người, động vật hoặc các văn bản khoa học, chính trị) và phần lớn các bức ảnh/văn bản khác chưa được gán nhãn được thu thập từ internet.

Thực tế cho thấy rất nhiều các bài toán Machine Learning thuộc vào nhóm này vì việc thu thập dữ liệu có nhãn tốn rất nhiều thời gian và có chi phí cao. Rất nhiều loại dữ liệu thậm chí cần phải có chuyên gia mới gán nhãn được (ảnh y học chẳng hạn). Ngược lại, dữ liệu chưa có nhãn có thể được thu thập với chi phí thấp từ internet.

1.3.5. Học sâu (Deep Learning)

Học sâu là một lĩnh vực quan trọng trong Machine Learning, nổi bật với khả năng tự động học biểu diễn từ dữ liệu thay vì yêu cầu biểu diễn được xác định trước. Điểm đặc biệt của học sâu là sử dụng các mạng nơ-ron sâu với nhiều tầng ẩn, cho phép chúng học biểu diễn cấp cao từ dữ liệu đầu vào.

Học sâu đã đạt được sự thành công ấn tượng trong nhiều lĩnh vực như thị giác máy tính, xử lý ngôn ngữ tự nhiên, và nhiều ứng dụng khác. Các mô hình nổi tiếng bao gồm Mạng nơ-ron tích chập (CNNs), Mạng nơ-ron hồi quy (RNNs), Mạng nơ-ron hồi quy dài hạn (LSTMs), và Mạng nơ-ron hồi quy áp dụng cho Dịch máy (Seq2Seq).

Học sâu đòi hỏi dữ liệu lớn và hiện đang là một trong những công cụ mạnh mẽ nhất trong việc xử lý và hiểu dữ liệu phức tạp.

1.4. Các bài toán cơ bản trong Machine Learning

Dưới đây là các dạng bài toán cơ bản mà chúng ta hay gặp và sử dụng trong machine learning. Mỗi bài toán đều có những đặc điểm và các chức năng riêng tùy thuộc vào độ phức tạp của bài toán mà chúng ta xử lý sẽ chọn phương pháp tối ưu nhất và sử dụng.

1.4.1. Các bài toán cơ bản

- Phân cụm (Clustering)
- Phân lớp (Classification)
- Hồi quy (Regression)
- Học tăng cường (Reinforce Learning)

1.4.2. Một số bài toán khác

- Trích rút luật (Rule Extraction)

- Ước lượng mật độ (Density estimation)
- Giảm chiều (Dimensionality reduction)
- Học để học (meta-learning)

1.5. Ứng dụng của Machine Learning

Nhiều hoạt động hàng ngày của chúng ta được trợ giúp bởi các thuật toán machine learning, bao gồm:

- *Trong y tế*: xác định bệnh lý của người bệnh mới dựa trên dữ liệu lịch sử của các bệnh nhân có cùng bệnh lý có cùng các đặc điểm đã được chữa khỏi trước đây, hay xác định loại thuốc phù hợp.
- *Trong lĩnh vực ngân hàng*: xác định khả năng khách hàng chậm trả các khoản vay hoặc rủi ro tín dụng do nợ xấu dựa trên phân tích Credit score; xác định xem liệu các giao dịch có hành vi phạm tội, lừa đảo hay không.
- *Trong giáo dục*: phân loại các học sinh theo hoàn cảnh, học lực để xem cần hỗ trợ gì cho những học sinh ví dụ như hoàn cảnh sống khó khăn nhưng học lực lại tốt.
- *Trong thương mại điện tử*: phân loại khách hàng theo sở thích cụ thể để hỗ trợ personalized marketing hay xây dựng hệ thống khuyến nghị, dựa trên dữ liệu từ website, social media.
- *Trong kinh tế nói chung*: giúp dự báo các sự kiện kinh tế trong tương lai, dự báo tình hình thời tiết trong nông nghiệp, xác định xu hướng thị trường chứng khoán để lên kế hoạch đầu tư thích hợp.

Một ví dụ cụ thể cho ứng dụng này như: Phân tích văn bản. Phân tích văn bản hay Text analysis sử dụng trong việc trích xuất, phân loại những thông tin từ văn bản (có thể là bài đăng Facebook, thư điện tử, tài liệu...) với các ứng dụng:

- *Lọc tin nhắn rác (Spam filtering)*: Đây là ứng dụng phổ biến nhất của Machine Learning trong phân tích văn bản. Bộ lọc sẽ tiến hành phân loại xem thư điện tử, tin nhắn có phải là spam không dựa trên nội dung mà tin nhắn gửi đến

- *Khai thác thông tin (Information Extraction)*: Từ một văn bản, máy có thể phân tích để trích xuất ra những thông tin quan trọng, cần thiết như tên người, các keyword...

CHƯƠNG 2: GIỚI THIỆU VỀ ĐỀ TÀI

2.1. Lý do chọn đề tài.

Ngày nay với sự phát triển mạnh mẽ của nền khoa học hiện đại, con người gần như đã bước qua kỷ nguyên mới với sức mạnh của nền CNTT đang trên đà phát triển mạnh mẽ. Cùng song hành với nó là sự phát triển mạnh mẽ của một số lĩnh vực như: Trí tuệ nhân tạo (AI), Internet Of Thing, Cloud Computing,... Trong các lĩnh vực đó chúng ta biết đến Học máy hay Máy học (Machine Learning) là một công nghệ phát triển từ lĩnh vực trí tuệ nhân tạo. Công nghệ máy học (Machine learning) là một nhánh quan trọng của trí tuệ nhân tạo (AI), giúp tự động hóa và tối ưu quy trình sản xuất, kiểm tra, quản lý... thúc đẩy sự phát triển chung của các thành phố thông minh, doanh nghiệp cũng như đời sống con người.

Để tiếp cận và góp phần đẩy mạnh sự phổ biến của nền công nghệ máy học ở Việt Nam. Qua đó, chúng em đã chọn đề tài "Xây dựng hệ thống học máy dự đoán giá vàng" cho bài tập lớn (BTL) môn học máy để tìm hiểu về phương hướng và các phương pháp hay cách thức để phát triển một bài toán dự báo.

- *Tính quan trọng của giá vàng:* Giá vàng là một trong những chỉ số tài chính quan trọng và có tầm quan trọng toàn cầu. Nó ảnh hưởng đến nhiều khía cạnh trong kinh tế và tài chính thế giới, bao gồm đầu tư, thương mại quốc tế, và chính trị. Khả năng dự đoán giá vàng có thể cung cấp thông tin quý báu cho các nhà đầu tư, ngân hàng, và chính phủ.
- *Thách thức trong việc dự đoán giá vàng:* Dự đoán giá vàng không phải lúc nào cũng dễ dàng do nó bị ảnh hưởng bởi nhiều yếu tố phức tạp như biến động thị trường, tình hình kinh tế thế giới, chính trị, và các sự kiện tự nhiên. Điều này tạo ra một thách thức thú vị cho việc áp dụng các kỹ thuật học máy để dự đoán giá vàng.
- *Tính ứng dụng rộng rãi:* Kỹ thuật học máy và dự đoán giá vàng có tiềm năng ứng dụng rộng rãi. Các công ty giao dịch vàng, ngân hàng, và các nhà đầu tư đều có thể tận dụng các dự đoán này để tối ưu hóa quyết định đầu tư, quản lý rủi ro, và dự báo xu hướng thị trường.

- *Khả năng sử dụng dữ liệu thời gian thực*: Dữ liệu hàng ngày về giá vàng từ nhiều nguồn khác nhau có sẵn, và nó thường được cập nhật liên tục. Điều này tạo cơ hội để xây dựng và kiểm tra các mô hình dự đoán thời gian thực, giúp chúng ta hiểu rõ hơn về biến động của giá vàng trong thời gian gần nhất.
- *Phát triển kỹ năng trong máy học*: Đề tài này cung cấp cơ hội phát triển kỹ năng trong lĩnh vực máy học, bao gồm xử lý dữ liệu lớn, lựa chọn và đánh giá mô hình, và tối ưu hóa. Nó sẽ giúp chúng em hiểu rõ hơn về quá trình phát triển mô hình dự đoán thời gian thực trong môi trường thực tế.

Tóm lại, việc lựa chọn đề tài "Xây dựng hệ thống học máy dự đoán giá vàng" không chỉ mang tính thực tiễn mà còn thúc đẩy sự phát triển cá nhân và kiến thức trong lĩnh vực máy học và tài chính. Chúng em tin rằng việc nghiên cứu và xây dựng hệ thống này sẽ đem lại giá trị cho nhiều khía cạnh trong lĩnh vực này.

2.2. Giới thiệu về bài toán dự đoán giá vàng

Giá vàng, trong bối cảnh tài chính và kinh tế, luôn thu hút sự quan tâm và theo dõi rất sát sao từ cả những nhà đầu tư cá nhân đến các chuyên gia tài chính và cả chính phủ. Sự biến động của giá vàng có thể ảnh hưởng đến nhiều khía cạnh của cuộc sống, từ đầu tư và thương mại quốc tế đến sự ổn định kinh tế toàn cầu.

Bài toán Dự đoán Giá Vàng là một trong những thách thức phức tạp nhất trong lĩnh vực dự đoán tài chính. Điều này bởi vì giá vàng bị ảnh hưởng bởi nhiều yếu tố khác nhau và thường biến đổi theo thời gian thực. Các yếu tố này có thể bao gồm tình hình kinh tế thế giới, chính trị, sự kiện toàn cầu, biến động thị trường và thậm chí là những tình huống khẩn cấp như đợt khủng bố hoặc đợt bùng phát dịch bệnh.

Trong bối cảnh này, áp dụng học máy để dự đoán giá vàng trở thành một thách thức thú vị và quan trọng. Việc này đòi hỏi chúng ta phải xử lý một lượng lớn dữ liệu thời gian thực, phát triển các mô hình học máy phức tạp để hiểu và dự đoán các mô hình biến động của giá vàng và tương tác với các biến thể của nó.

Bài báo cáo này sẽ giới thiệu về việc xây dựng một hệ thống học máy dự đoán giá vàng, tập trung vào việc thu thập dữ liệu, tiền xử lý dữ liệu, lựa chọn mô hình, và đánh giá hiệu suất của mô hình. Chúng ta sẽ khám phá các phương pháp học máy phù hợp để dự

đoán giá vàng và những thách thức đặc biệt của bài toán này. Ngoài ra, chúng ta sẽ thảo luận về ứng dụng tiềm năng của dự đoán giá vàng trong thế giới thực, từ quản lý đầu tư đến dự báo rủi ro tài chính. Bài báo cáo này sẽ đưa ra một cái nhìn tổng quan về việc áp dụng học máy để giải quyết một trong những bài toán quan trọng và phức tạp nhất trong lĩnh vực tài chính.

2.3. Giới thiệu về các phương pháp Dự đoán giá vàng

Khi nghiên cứu về dự đoán giá vàng, có nhiều phương pháp khác nhau có thể được áp dụng từ lĩnh vực thống kê và học máy. Mỗi phương pháp có ưu điểm và hạn chế riêng, và lựa chọn phụ thuộc vào bản chất của dữ liệu và mục tiêu cụ thể của dự đoán. Dưới đây là một số phương pháp thường được sử dụng:

- *Hồi quy (Regression)*: Hồi quy là một phương pháp cổ điển trong dự đoán giá vàng. Nó tạo ra một mô hình dự đoán dựa trên mối quan hệ tuyến tính giữa các biến độc lập và biến mục tiêu (trong trường hợp này là giá vàng). Hồi quy có thể được điều chỉnh để xử lý dữ liệu thời gian thực và các biến phức tạp.
- *Mạng Nơ-ron (Neural Networks)*: Mạng nơ-ron sâu đã trở thành một phương pháp phổ biến cho dự đoán tài chính. Chúng có khả năng học các mô hình phi tuyến tính phức tạp và có thể xử lý dữ liệu chuỗi thời gian một cách hiệu quả.
- *Mô hình chuỗi thời gian (Time Series Models)*: Mô hình chuỗi thời gian như ARIMA (AutoRegressive Integrated Moving Average) và LSTM (Long Short-Term Memory) có thể sử dụng để dự đoán giá vàng dựa trên dữ liệu lịch sử. Chúng tập trung vào phân tích các mô hình chuỗi thời gian và mô hình hóa các biến động theo thời gian.
- *Phân tích cơ bản (Fundamental Analysis)*: Phương pháp này dựa trên việc phân tích các yếu tố cơ bản ảnh hưởng đến giá vàng như tình hình kinh tế toàn cầu, chính trị, và sự kiện toàn cầu. Nó sử dụng thông tin về cung cầu và các yếu tố tác động đến giá vàng để đưa ra dự đoán.

Tại sao chọn phương pháp hồi quy?

Trong bài toán dự đoán giá vàng, chúng em đã chọn sử dụng phương pháp Hồi quy vì những lý do sau:

- *Tính đơn giản và diễn giải:* Hồi quy tạo ra một mô hình tuyến tính, dễ hiểu và diễn giải, cho phép ta xác định được mối quan hệ giữa các biến độc lập và giá vàng.
- *Khả năng xử lý biến động thời gian thực:* Hồi quy có thể được điều chỉnh để xử lý dữ liệu thời gian thực bằng cách sử dụng các biến số thời gian như tháng, năm, hoặc ngày trong mô hình.
- *Khả năng đối mặt với nhiễu và biến động:* Hồi quy có thể được cải thiện bằng cách sử dụng các biến số bổ sung hoặc kỹ thuật chống nhiễu để ứng phó với sự biến động của giá vàng.

Tuy nhiên, điều quan trọng là việc lựa chọn phương pháp không phải lúc nào cũng là tùy chọn tốt nhất và cần phải xem xét cẩn thận dữ liệu cụ thể và mục tiêu của mô hình. Trong quá trình nghiên cứu, chúng ta cũng có thể kết hợp nhiều phương pháp để tạo ra một mô hình dự đoán mạnh mẽ hơn.

2.4. Ý nghĩa của việc xây dựng mô hình học máy Dự đoán giá vàng

Ý nghĩa thực tiễn:

- *Quản lý rủi ro tài chính:* Các nhà đầu tư và các tổ chức tài chính có thể sử dụng các mô hình dự đoán giá vàng để quản lý rủi ro và ra quyết định đầu tư thông minh.
- *Dự báo thị trường tài chính:* Dự đoán giá vàng có thể cung cấp thông tin quý báu về xu hướng thị trường tài chính tổng quan, giúp các chuyên gia tài chính và nhà quản lý thị trường hiểu rõ hơn về tình hình kinh tế và tài chính toàn cầu.
- *Nghiên cứu tài chính và kinh tế:* Việc xây dựng các mô hình dự đoán giá vàng có thể cung cấp những thông tin quan trọng để nghiên cứu và hiểu rõ hơn về mối quan hệ giữa giá vàng và các yếu tố tác động khác trong nền kinh tế và tài chính.
- *Cải thiện quản lý đầu tư cá nhân:* Cho phép nhà đầu tư cá nhân có kiến thức để đưa ra quyết định đầu tư dựa trên thông tin dự đoán giá vàng.

- *Ứng dụng trong ngân hàng và tài chính:* Các ngân hàng và tổ chức tài chính có thể sử dụng các mô hình dự đoán giá vàng để cải thiện dịch vụ của họ và tối ưu hóa quyết định tài chính.

Tóm lại, việc xây dựng một mô hình học máy dự đoán giá vàng mang ý nghĩa quan trọng trong lĩnh vực tài chính và kinh tế, giúp cải thiện quản lý rủi ro, dự báo thị trường, và cung cấp thông tin quan trọng cho nghiên cứu và ứng dụng thực tế. Tuy nhiên, nó cũng đặt ra nhiều thách thức do tính phức tạp và biến đổi của dữ liệu giá vàng.

2.5. Các bước cơ bản Dự đoán giá vàng

2.5.1. Thu thập dữ liệu

Bước đầu tiên là thu thập dữ liệu giá vàng. Dữ liệu này thường được cung cấp từ các nguồn tài chính như trang web tài chính, thị trường tài chính, hoặc các dịch vụ cung cấp dữ liệu tài chính.

2.5.2. Tiền xử lý dữ liệu

Dữ liệu giá vàng thường không hoàn hảo và có thể chứa nhiễu. Trong bước này, cần tiến hành tiền xử lý dữ liệu để loại bỏ dữ liệu bất thường hoặc thiếu sót và làm sạch dữ liệu.

2.5.3. Chọn biến độc lập (Features)

Chọn các biến độc lập (hoặc đặc trưng) mà bạn tin rằng có thể ảnh hưởng đến giá vàng. Điều này có thể bao gồm các yếu tố tài chính như tỷ giá ngoại tệ, lãi suất, chỉ số kinh tế, hay thậm chí là các sự kiện toàn cầu như chiến tranh hoặc khủng bố.

2.5.4. Tạo mô hình hồi quy tuyến tính

Bây giờ, bạn cần tạo một mô hình hồi quy tuyến tính. Mô hình này sẽ cố gắng xác định mối quan hệ tuyến tính giữa giá vàng và các biến độc lập bạn đã chọn. Công thức tổng quát cho mô hình hồi quy tuyến tính là:

$$\text{Giá vàng} = a * \text{Biến 1} + b * \text{Biến 2} + c * \text{Biến 3} + \dots + \text{Intercept}$$

Trong đó, a, b, c là các hệ số và Intercept là hệ số chặn.

2.5.5. Huấn luyện mô hình

Sử dụng dữ liệu lịch sử, bạn huấn luyện mô hình hồi quy tuyến tính để tìm các giá trị tối ưu cho các hệ số (a, b, c và Intercept). Quá trình này thường dựa trên phương pháp tối thiểu hoá sai số (least squares method) để đưa ra mô hình tốt nhất.

2.5.6. Đánh giá mô hình

Đánh giá hiệu suất của mô hình bằng cách sử dụng các thước đo như sai số bình phương trung bình (Mean Squared Error - MSE), sai số tuyệt đối trung bình (Mean Absolute Error - MAE), hoặc hệ số xác định (Coefficient of Determination - R-squared).

2.5.7. Dự đoán giá vàng

Khi mô hình đã được đào tạo và đánh giá, bạn có thể sử dụng nó để dự đoán giá vàng cho các điểm dữ liệu mới. Đưa vào các giá trị của các biến độc lập, mô hình sẽ đưa ra dự đoán về giá vàng tương ứng.

2.5.8. Cập nhật và giám sát

Việc dự đoán giá vàng cần được theo dõi và cập nhật liên tục, bởi vì thị trường và các yếu tố tác động đến giá vàng thay đổi liên tục. Mô hình cần được cập nhật với dữ liệu mới để duy trì hiệu suất.

2.6. Ý nghĩa của bài toán

Tóm lại, việc xây dựng một mô hình học máy dự đoán giá vàng mang ý nghĩa quan trọng trong lĩnh vực tài chính và kinh tế, giúp cải thiện quản lý rủi ro, dự báo thị trường, và cung cấp thông tin quan trọng cho nghiên cứu và ứng dụng thực tế. Tuy nhiên, nó cũng đặt ra nhiều thách thức do tính phức tạp và biến đổi của dữ liệu giá vàng.

CHƯƠNG 3: THUẬT TOÁN HỒI QUY TRONG DỰ BÁO GIÁ VÀNG

3.1. Thu thập dữ liệu

Thu thập dữ liệu là bước quan trọng nhất trong quá trình Machine Learning, bất kỳ dự án nào cũng bắt đầu từ đây. Dữ liệu chất lượng và đại diện cho bài toán là yếu tố quyết định đối với thành công của mô hình. Trong dự án dự đoán giá vàng của chúng em, quá trình thu thập dữ liệu được thực hiện như sau:

- *Xác định Nguồn Dữ Liệu:* Đầu tiên, chúng em đã xác định các nguồn dữ liệu liên quan đến giá vàng. Điều này có thể bao gồm các trang web tài chính, dữ liệu lịch sử từ các sàn giao dịch và các nguồn khác chứa thông tin về giá vàng. (Được lấy trong Yahoo finance)
- *Thu Thập Dữ Liệu Lịch Sử:* Chúng em đã thu thập dữ liệu lịch sử về giá vàng trong khoảng thời gian cần thiết. Điều này có thể bao gồm giá mở cửa, giá đóng cửa, giá cao nhất, giá thấp nhất và khối lượng giao dịch hàng ngày.
- *Thu Thập Dữ Liệu Biến Độc Lập:* Ngoài giá vàng, chúng em cũng thu thập dữ liệu về các biến độc lập có thể ảnh hưởng đến giá vàng. Điều này có thể bao gồm tỷ giá ngoại tệ, chỉ số kinh tế, tin tức toàn cầu và các yếu tố khác có thể có tác động lên giá vàng.
- *Kiểm Tra và Tiền Xử Lý Dữ Liệu:* Sau khi thu thập dữ liệu, chúng em kiểm tra tính nhất quán và độ chính xác của dữ liệu. Nếu có dữ liệu bị thiếu hoặc nhiễu, chúng em tiến hành tiền xử lý dữ liệu để làm sạch và chuẩn hóa nó.
- *Lưu Trữ Dữ Liệu:* Dữ liệu sau khi được thu thập và tiền xử lý được lưu trữ một cách an toàn và dễ quản lý để sử dụng cho quá trình huấn luyện và đánh giá mô hình.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Open	High	Low	Close*	Adj Close**	Volume	interest_rate	cpi	pmi	unemployment	usd_price
2	5-Oct-23	1,826.30	1,826.30	1,815.60	1,816.60	1,816.60	393	5.5	3.7	48	3.8	23975
3	4-Oct-23	1,821.80	1,828.20	1,816.40	1,818.50	1,818.50	357	5.5	3.7	48	3.8	23900
4	3-Oct-23	1,825.00	1,830.90	1,816.80	1,824.60	1,824.60	310	5.5	3.7	48	3.8	23900
5	2-Oct-23	1,844.10	1,845.90	1,826.70	1,830.00	1,830.00	2,309	5.5	3.7	47.6	3.8	23850
6	29-Sep-23	1,864.30	1,872.30	1,846.00	1,848.10	1,848.10	1,605	5.5	3.7	47.6	3.8	23850
7	28-Sep-23	1,875.50	1,878.20	1,856.70	1,860.40	1,860.40	4,654	5.5	3.7	47.6	3.8	23850
8	27-Sep-23	1,871.60	1,871.60	1,871.60	1,871.60	1,871.60	12,251	5.5	3.7	47.6	3.8	23850
9	26-Sep-23	1,905.00	1,905.10	1,900.40	1,900.40	1,900.40	161	5.5	3.7	47.6	3.8	23935
10	25-Sep-23	1,916.60	1,916.60	1,916.60	1,916.60	1,916.60	114	5.5	3.7	47.6	3.8	23965
11	22-Sep-23	1,925.40	1,925.40	1,925.40	1,925.40	1,925.40	1	5.5	3.7	47.6	3.8	23815
12	21-Sep-23	1,924.00	1,924.00	1,919.20	1,919.20	1,919.20	9	5.5	3.7	47.6	3.8	23750
13	20-Sep-23	1,945.60	1,945.60	1,929.20	1,945.60	1,945.60	119	5.5	3.7	47.6	3.8	23740
14	19-Sep-23	1,934.00	1,934.00	1,932.00	1,932.00	1,932.00	6	5.5	3.7	47.6	3.8	23735
15	18-Sep-23	1,928.20	1,931.50	1,926.40	1,931.50	1,931.50	19	5.5	3.7	47.6	3.8	23730
16	15-Sep-23	1,923.50	1,927.30	1,923.50	1,923.70	1,923.70	8	5.5	3.7	47.6	3.8	23720
17	14-Sep-23	1,903.00	1,910.00	1,903.00	1,910.00	1,910.00	169	5.5	3.7	47.6	3.8	23725
18	13-Sep-23	1,910.30	1,914.60	1,908.90	1,909.10	1,909.10	15	5.5	3.7	47.6	3.8	23750
19	12-Sep-23	1,920.80	1,920.80	1,911.30	1,911.30	1,911.30	365	5.5	3.7	47.6	3.8	23735
20	11-Sep-23	1,923.00	1,928.30	1,921.80	1,923.30	1,923.30	50	5.5	3.7	47.6	3.8	23680
21	8-Sep-23	1,924.00	1,924.70	1,918.40	1,918.40	1,918.40	105	5.5	3.7	47.6	3.8	23680
22	7-Sep-23	1,917.60	1,921.50	1,917.50	1,917.50	1,917.50	41	5.5	3.7	47.6	3.8	23685
23	6-Sep-23	1,924.10	1,925.30	1,916.30	1,918.10	1,918.10	24	5.5	3.7	47.6	3.8	23661
24	5-Sep-23	1,934.40	1,934.40	1,924.70	1,926.20	1,926.20	113	5.5	3.7	47.6	3.8	23665
25	1-Sep-23	1,950.00	1,950.00	1,936.60	1,939.80	1,939.80	167	5.5	3.7	47.6	3.8	23655
26	31-Aug-23	1,944.40	1,944.40	1,938.20	1,938.20	1,938.20	491	5.5	3.7	47.6	3.8	23665
27	30-Aug-23	1,936.00	1,946.40	1,934.60	1,944.30	1,944.30	396	5.5	3.7	47.6	3.8	23656
28	29-Aug-23	1,917.90	1,936.50	1,916.90	1,936.50	1,936.50	1,674	5.5	3.2	46.4	3.5	23647
29	28-Aug-23	1,915.50	1,921.50	1,915.50	1,917.90	1,917.90	99	5.5	3.2	46.4	3.5	23630
30	25-Aug-23	1,919.80	1,919.80	1,911.10	1,911.10	1,911.10	195	5.5	3.2	46.4	3.5	23635
31	24-Aug-23	1,920.00	1,920.80	1,918.20	1,918.20	1,918.20	337	5.5	3.2	46.4	3.5	23630
32	23-Aug-23	1,909.60	1,918.50	1,909.60	1,918.50	1,918.50	547	5.25	3.2	46.4	3.5	23630
33	22-Aug-23	1,894.60	1,896.80	1,892.80	1,896.40	1,896.40	46	5.25	3.2	46.4	3.5	23665

Quá trình thu thập dữ liệu đòi hỏi sự cẩn trọng và tính tổ chức. Dữ liệu chất lượng và đủ lớn rất quan trọng để đảm bảo mô hình học tốt và dự đoán chính xác.

3.2. Phân tích dữ liệu

Bước tiếp theo sau thu thập dữ liệu là phân tích dữ liệu. Phân tích dữ liệu giúp bạn hiểu sâu hơn về thông tin mà dữ liệu cung cấp và chuẩn bị cho việc xây dựng mô hình Machine Learning. Trong dự án dự đoán giá vàng, chúng em đã tiến hành các hoạt động sau để phân tích dữ liệu:

- *Khai phá dữ liệu (Data mining)*: Trong giai đoạn này, chúng em đã khai phá dữ liệu để tìm hiểu về sự phân phối của giá vàng, các biến độc lập và mối quan hệ giữa chúng. Các kỹ thuật data mining đã giúp chúng em tìm ra các mẫu, xu hướng và sự biến đổi trong dữ liệu.
- *Xử lý dữ liệu, phân tích dữ liệu*: Sau khi thu thập dữ liệu, chúng em đã thực hiện xử lý dữ liệu để làm sạch và chuẩn hóa nó. Điều này bao gồm xử lý giá trị ngoại lai, điền dữ liệu bị thiếu, và chuẩn hóa các biến. Chúng em đã sử dụng các phương pháp thống kê và machine learning để phân tích dữ liệu. Chúng em đã xem xét

mối quan hệ giữa giá vàng và các biến độc lập bằng cách tính toán hệ số tương quan và tạo biểu đồ tương ứng.

- *Trực quan hóa dữ liệu:* Trực quan hóa dữ liệu là một phần quan trọng của phân tích dữ liệu. Chúng em đã tạo biểu đồ dạng đường để thể hiện sự biến đổi của giá vàng theo thời gian và sử dụng biểu đồ tương quan để hiển thị mối quan hệ giữa các biến.
- *Dự báo dự đoán với Azure Machine Learning:* Để dự đoán giá vàng, chúng em đã sử dụng Azure Machine Learning, một nền tảng mạnh mẽ cho việc xây dựng và huấn luyện các mô hình Machine Learning. Chúng em đã tạo mô hình hồi quy tuyến tính và dự đoán giá vàng dựa trên các biến độc lập đã chọn.
- *Lưu trữ:* Dữ liệu và kết quả của phân tích dữ liệu và dự đoán đã được lưu trữ một cách an toàn và dễ quản lý để sử dụng cho các bước tiếp theo trong quá trình dự báo giá vàng.

Những hoạt động này đóng vai trò quan trọng trong việc chuyển dữ liệu thành thông tin hữu ích và làm nền tảng cho việc xây dựng mô hình dự đoán giá vàng. Quá trình phân tích dữ liệu giúp chúng em hiểu rõ hơn về đặc điểm và xu hướng của dữ liệu, từ đó giúp cho việc xây dựng mô hình dự đoán giá vàng trở nên hiệu quả hơn.

3.3. Chọn các biến độc lập (Features)

Việc chọn biến độc lập là một bước quan trọng trong quá trình phân tích dữ liệu và xây dựng mô hình dự đoán giá vàng. Chúng em đã tiến hành các hoạt động sau để lựa chọn các biến độc lập quan trọng:

- *Phân Loại Biến Độc Lập:* Đầu tiên, chúng em đã phân loại các biến độc lập thành các loại khác nhau, bao gồm biến số, biến hạng mục và biến thời gian. Điều này giúp chúng em xác định cách xử lý và chọn biến cho từng loại.
- *Ưu Tiên Các Biến Có Ảnh Hưởng Dự Đoán Lớn:* Chúng em đã ưu tiên chọn các biến mà chúng em tin rằng có ảnh hưởng lớn đến giá vàng. Điều này dựa trên kiến thức lĩnh vực và phân tích sơ bộ của dữ liệu.

- *Phân Tích Tương Quan*: Chúng em đã sử dụng các phương pháp thống kê để phân tích tương quan giữa giá vàng và các biến độc lập. Các biến có mối quan hệ mạnh với giá vàng đã được ưu tiên.
- *Loại Bỏ Biến Tương Quan Cao*: Nếu hai hoặc nhiều biến có mối tương quan cao với nhau, chúng em chỉ chọn một trong số chúng để tránh hiện tượng đa cộng tuyến (multicollinearity) trong mô hình hồi quy.
- *Lựa Chọn Từ Kết Quả Phân Tích Dữ Liệu*: Kết quả phân tích dữ liệu, bao gồm tương quan và biểu đồ tương quan, đã hướng dẫn chúng em trong quá trình chọn biến độc lập. Những biến quan trọng đã được đưa vào mô hình.
- *Xác Định Biến Thời Gian Quan Trọng*: Trong trường hợp biến thời gian được sử dụng, chúng em đã xác định các biến thời gian quan trọng, bao gồm ngày, tháng, năm và các yếu tố khác có thể ảnh hưởng đến giá vàng.
- *Tiền Xử Lý Biến Hạng Mục*: Nếu có biến hạng mục, chúng em đã tiến hành tiền xử lý như mã hóa one-hot hoặc sử dụng các phương pháp phù hợp để biến chúng thành dạng số.

Chọn biến độc lập một cách cẩn thận giúp giảm chi phí tính toán, cải thiện hiệu suất của mô hình và giúp hiểu rõ hơn về tác động của các yếu tố lên giá vàng. Sau khi thực hiện quá trình này chúng em đã tiến hành chọn được ra các biến độc lập là: thời gian, giá mở cửa đóng cửa và khối lượng giao dịch, tỷ giá ngoại tệ...

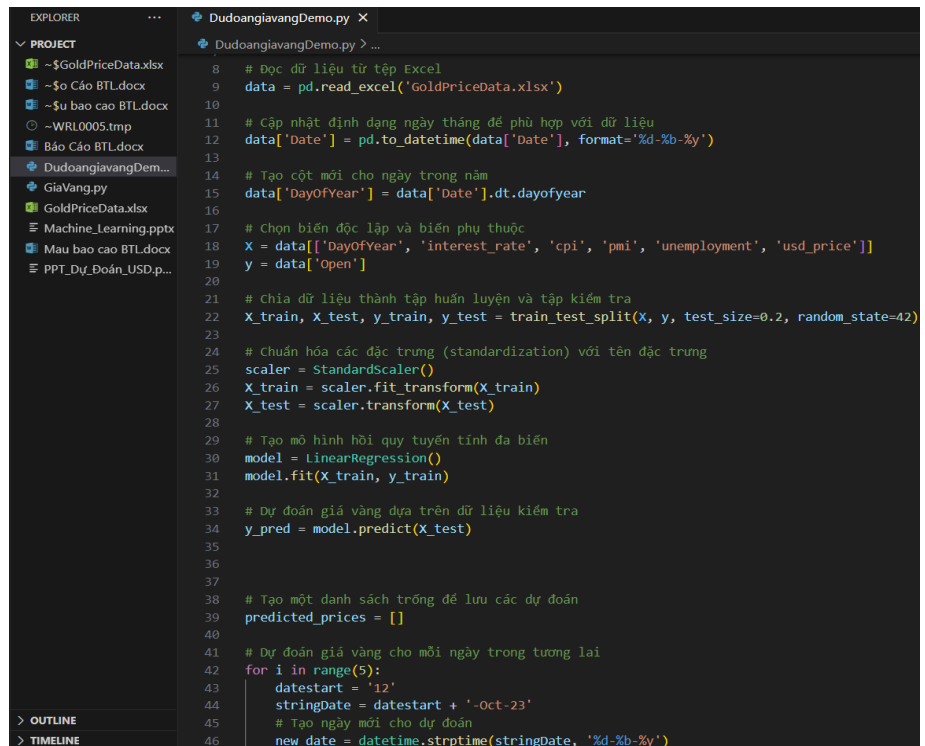
3.4. Xây dựng mô hình hồi quy tuyến tính

Sau khi đã thu thập và phân tích dữ liệu cùng việc chọn các biến độc lập quan trọng, bước tiếp theo trong quá trình dự đoán giá vàng là xây dựng mô hình hồi quy tuyến tính. Chúng em đã tiến hành các hoạt động sau để xây dựng mô hình:

- *Chọn loại mô hình*: Dựa trên tính chất của bài toán dự đoán giá vàng, chúng em đã quyết định sử dụng mô hình hồi quy tuyến tính đa biến (multiple linear regression). Mô hình hồi quy tuyến tính được chọn vì nó phù hợp với dự đoán mối quan hệ tuyến tính giữa giá vàng và các biến độc lập. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$

- *Chuẩn bị dữ liệu đầu vào:* Chúng em đã chuẩn bị dữ liệu đầu vào cho mô hình, bao gồm biến độc lập đã được chọn và biến phản hồi (giá vàng). Dữ liệu đã được chia thành tập huấn luyện và tập kiểm tra.
- *Xây dựng mô hình:* Chúng em đã sử dụng thư viện hoặc công cụ phù hợp để xây dựng mô hình hồi quy tuyến tính. Điều này bao gồm việc thiết lập mô hình, cài đặt siêu tham số, và huấn luyện mô hình trên tập dữ liệu huấn luyện.
- *Đánh giá mô hình:* Sau khi mô hình được huấn luyện, chúng em đã sử dụng tập kiểm tra để đánh giá hiệu suất của mô hình. Điều này bao gồm việc tính toán các thước đo như sai số trung bình bình phương (Mean Squared Error) và hệ số xác định R-squared trên tập kiểm tra.
- *Điều chỉnh mô hình:* Nếu mô hình không đạt hiệu suất mong muốn, chúng em đã điều chỉnh các siêu tham số hoặc thử nghiệm các biến khác để cải thiện mô hình.
- *Lưu mô hình:* Sau khi mô hình hoàn thành và đã đạt hiệu suất tốt trên tập kiểm tra, chúng em đã lưu mô hình để sử dụng cho việc dự đoán giá vàng trong tương lai.

Xây dựng mô hình là một phần quan trọng của quá trình dự đoán giá vàng và đòi hỏi sự cân nhắc cẩn thận về lựa chọn mô hình và điều chỉnh. Mô hình hồi quy tuyến tính được chọn dựa trên sự hiểu biết về bài toán, nhưng còn có thể sử dụng các mô hình khác nếu thích hợp.



```

EXPLORER
PROJECT
~$GoldPriceData.xlsx
~$o Báo BTL.docx
~$u báo cáo BTL.docx
~WRL0005.tmp
Báo Báo BTL.docx
DudoangiangDemo...
GiaVang.py
GoldPriceData.xlsx
Machine_Learning.pptx
Mau báo cáo BTL.docx
PPT_Dự_Đoán_USD.p...

DudoangiangDemo.py x
DudoangiangDemo.py > ...
8 # Đọc dữ liệu từ tập Excel
9 data = pd.read_excel('GoldPriceData.xlsx')
10
11 # Cập nhật định dạng ngày tháng để phù hợp với dữ liệu
12 data['Date'] = pd.to_datetime(data['Date'], format='%d-%b-%y')
13
14 # Tạo cột mới cho ngày trong năm
15 data['DayOfYear'] = data['Date'].dt.dayofyear
16
17 # Chọn biến độc lập và biến phụ thuộc
18 X = data[['DayOfYear', 'interest_rate', 'cpi', 'pmi', 'unemployment', 'usd_price']]
19 y = data['Open']
20
21 # Chia dữ liệu thành tập huấn luyện và tập kiểm tra
22 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
23
24 # Chuẩn hóa các đặc trưng (standardization) với tên đặc trưng
25 scaler = StandardScaler()
26 X_train = scaler.fit_transform(X_train)
27 X_test = scaler.transform(X_test)
28
29 # Tạo mô hình hồi quy tuyến tính đa biến
30 model = LinearRegression()
31 model.fit(X_train, y_train)
32
33 # Dự đoán giá vàng dựa trên dữ liệu kiểm tra
34 y_pred = model.predict(X_test)
35
36
37
38 # Tạo một danh sách trống để lưu các dự đoán
39 predicted_prices = []
40
41 # Dự đoán giá vàng cho mỗi ngày trong tương lai
42 for i in range(5):
43     datestart = '12'
44     stringDate = datestart + '-Oct-23'
45     # Tạo ngày mới cho dự đoán
46     new_date = datetime.strptime(stringDate, '%d-%b-%y')

```

3.5. Huấn luyện mô hình

Huấn luyện mô hình là bước quan trọng trong quá trình xây dựng mô hình dự đoán giá vàng. Chúng em đã thực hiện các hoạt động sau để huấn luyện mô hình hồi quy tuyến tính:

- *Chuẩn bị dữ liệu huấn luyện:* Trước khi bắt đầu quá trình huấn luyện, chúng em đã chuẩn bị dữ liệu huấn luyện bằng cách chọn các biến độc lập và biến mục tiêu (giá vàng) từ tập dữ liệu đã phân chia.
- *Lựa chọn mô hình:* Chúng em đã lựa chọn mô hình hồi quy tuyến tính dựa trên quá trình phân tích dữ liệu và phù hợp với tính chất của bài toán dự đoán giá vàng.
- *Xác định siêu tham số:* Siêu tham số của mô hình, như hệ số góc và hệ số chặn, đã được xác định trước hoặc thông qua quá trình tối ưu hóa. Chúng em đã sử dụng phương pháp gradient descent để tối ưu hóa các tham số này.
- *Chuẩn bị dữ liệu đầu vào:* Dữ liệu huấn luyện đã được chuẩn bị và biến đổi để phù hợp với mô hình. Điều này bao gồm việc tiền xử lý dữ liệu như chuẩn hóa, mã hóa biến hạng mục, và xử lý giá trị ngoại lai.
- *Huấn luyện mô hình:* Chúng em đã huấn luyện mô hình trên tập dữ liệu huấn luyện sử dụng phương pháp hồi quy tuyến tính. Quá trình này bao gồm tính toán hệ số hồi quy và chặn của mô hình sao cho chúng tối ưu hóa việc dự đoán giá vàng.
- *Đánh giá hiệu suất:* Sau khi mô hình đã được huấn luyện, chúng em đã đánh giá hiệu suất của nó bằng cách sử dụng tập kiểm tra. Chúng em đã sử dụng các thước đo như sai số trung bình bình phương (MSE) và hệ số xác định R-squared để đánh giá mức độ chính xác của mô hình.
- *Kiểm tra overfitting:* Chúng em đã kiểm tra xem mô hình có dấu hiệu overfitting hay không. Nếu có, chúng em đã thực hiện điều chỉnh mô hình để giảm overfitting.

- *Lưu mô hình*: Sau khi hoàn thành quá trình huấn luyện và đảm bảo rằng mô hình đạt hiệu suất tốt, chúng em đã lưu mô hình để sử dụng cho dự đoán giá vàng trong tương lai.

Quá trình huấn luyện mô hình đòi hỏi kiên nhẫn và sự cân nhắc cẩn thận để đảm bảo rằng mô hình có khả năng dự đoán giá vàng một cách chính xác.

3.6. Đánh giá mô hình

Bước cuối cùng trong quá trình xây dựng mô hình dự đoán giá vàng là đánh giá hiệu suất của mô hình. Chúng em đã thực hiện các hoạt động sau để đánh giá mô hình:

- *Sai số trung bình bình phương (mse)*: Chúng em đã tính toán sai số trung bình bình phương (MSE), một thước đo phổ biến để đánh giá hiệu suất của mô hình. MSE tính toán sự khác biệt giữa giá vàng thực tế và giá vàng được dự đoán, và sau đó lấy trung bình bình phương của tất cả các sai số. Một MSE thấp hơn cho thấy mô hình dự đoán gần với giá vàng thực tế.
- *Hệ số xác định R-squared (R^2)*: Chúng em cũng tính toán hệ số xác định R-squared, một thước đo khác để đo lường mức độ biến đổi trong giá vàng mà mô hình có thể giải thích. Giá trị R-squared càng gần 1, mô hình càng tốt trong việc dự đoán giá vàng.
- *Biểu đồ tương quan*: Chúng em đã vẽ biểu đồ tương quan giữa giá vàng thực tế và giá vàng được dự đoán bằng mô hình. Điều này giúp chúng em xem xét mối quan hệ giữa các dự đoán và giá trị thực tế.
- *Kiểm tra overfitting*: Chúng em đã kiểm tra xem mô hình có dấu hiệu overfitting hay không bằng cách so sánh hiệu suất trên tập huấn luyện và tập kiểm tra. Nếu mô hình hoạt động rất tốt trên tập huấn luyện nhưng kém trên tập kiểm tra, đây có thể là dấu hiệu overfitting.
- *So sánh với mô hình trước đó*: Nếu có sẵn mô hình hoặc phương pháp dự đoán giá vàng trước đó, chúng em đã so sánh hiệu suất của mô hình hiện tại với mô hình trước đó để đảm bảo rằng mô hình mới là sự cải thiện.

Dựa trên các kết quả đánh giá này, chúng em đã đưa ra nhận xét về hiệu suất của mô hình và quyết định xem liệu nó có thể được sử dụng để dự đoán giá vàng trong tương lai hay không. Nếu mô hình đạt hiệu suất tốt, nó có thể trở thành một công cụ hữu ích trong việc đưa ra dự đoán và quyết định đầu tư trong tương lai.

3.7. Dự báo giá vàng

Sau khi chúng em đã xây dựng và đánh giá mô hình dự đoán giá vàng, bước tiếp theo là sử dụng mô hình để thực hiện dự đoán giá vàng trong tương lai. Chúng em đã thực hiện các hoạt động sau để thực hiện dự đoán:

- *Chuẩn bị dữ liệu dự đoán:* Để thực hiện dự đoán, chúng em đã chuẩn bị dữ liệu đầu vào cho mô hình. Điều này bao gồm biến độc lập như thời gian, tỷ giá ngoại tệ, chỉ số kinh tế, và bất kỳ yếu tố khác nào có thể ảnh hưởng đến giá vàng trong tương lai.
- *Áp dụng mô hình dự đoán:* Chúng em đã sử dụng mô hình hồi quy tuyến tính mà chúng em đã xây dựng để thực hiện dự đoán. Bằng cách cung cấp các giá trị của các biến độc lập trong tương lai cho mô hình, chúng em đã nhận được dự đoán về giá vàng dự kiến.
- *Theo dõi kết quả dự đoán:* Chúng em đã theo dõi và ghi nhận kết quả của các dự đoán trong tương lai. Điều này giúp chúng em đánh giá sự chính xác của mô hình và điều chỉnh nếu cần thiết.
- *Cập nhật dữ liệu:* Để cải thiện mô hình dự đoán theo thời gian, chúng em đã cập nhật dữ liệu liên quan đến giá vàng và các biến độc lập. Dữ liệu mới giúp mô hình học hỏi từ thông tin mới nhất và cải thiện khả năng dự đoán.

```
Giá vàng dự đoán cho ngày 2023-10-14 00:00:00: 1904.4343153176137
C:\Users\Laptop24h\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names, but Standard
Scaler was fitted with feature names
  warnings.warn(
Giá vàng dự đoán cho ngày 2023-10-15 00:00:00: 1904.4756392234292
C:\Users\Laptop24h\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names, but Standard
Scaler was fitted with feature names
  warnings.warn(
Giá vàng dự đoán cho ngày 2023-10-16 00:00:00: 1904.5169631292447
```

Dự đoán giá vàng là một phần quan trọng trong quá trình này, và nó có thể hỗ trợ người đầu tư, doanh nghiệp và người tiêu dùng trong việc đưa ra quyết định dựa trên thông tin dự báo.

3.8. Cập nhật và giám sát

Sau khi chúng em đã xây dựng mô hình dự đoán giá vàng và thực hiện các dự đoán, quá trình cập nhật và giám sát là một phần quan trọng để duy trì và cải thiện mô hình theo thời gian. Chúng em đã thực hiện các hoạt động sau:

- *Cập nhật dữ liệu:* Chúng em đã thường xuyên cập nhật dữ liệu liên quan đến giá vàng và các biến độc lập. Dữ liệu mới giúp mô hình học hỏi từ thông tin mới nhất và điều chỉnh dự đoán.
- *Điều chỉnh mô hình:* Nếu dữ liệu mới hoặc kết quả dự đoán không khớp với thực tế, chúng em đã điều chỉnh mô hình. Điều này có thể bao gồm việc thay đổi siêu tham số của mô hình hoặc thay đổi biến độc lập.
- *Giám sát hiệu suất:* Chúng em đã theo dõi và giám sát hiệu suất của mô hình trên dự đoán thực tế. Điều này giúp chúng em đảm bảo rằng mô hình đang hoạt động đúng và đáng tin cậy.
- *Xây dựng mô hình mới:* Nếu cần, chúng em đã xây dựng mô hình mới hoặc cải thiện mô hình hiện có để đảm bảo rằng nó đáp ứng được sự biến đổi trong thị trường và dữ liệu.
- *Hợp nhất kiến thức lĩnh vực:* Chúng em đã hợp nhất kiến thức lĩnh vực và thông tin mới để cải thiện mô hình. Điều này có thể bao gồm việc thêm vào các biến độc lập mới hoặc điều chỉnh cách mô hình xử lý dữ liệu.
- *Tạo cơ hội tự động hoá:* Để đảm bảo tính liên tục và hiệu quả của quá trình cập nhật và giám sát, chúng em đã tạo cơ hội tự động hoá một phần của quá trình. Điều này giúp tiết kiệm thời gian và nguồn lực.

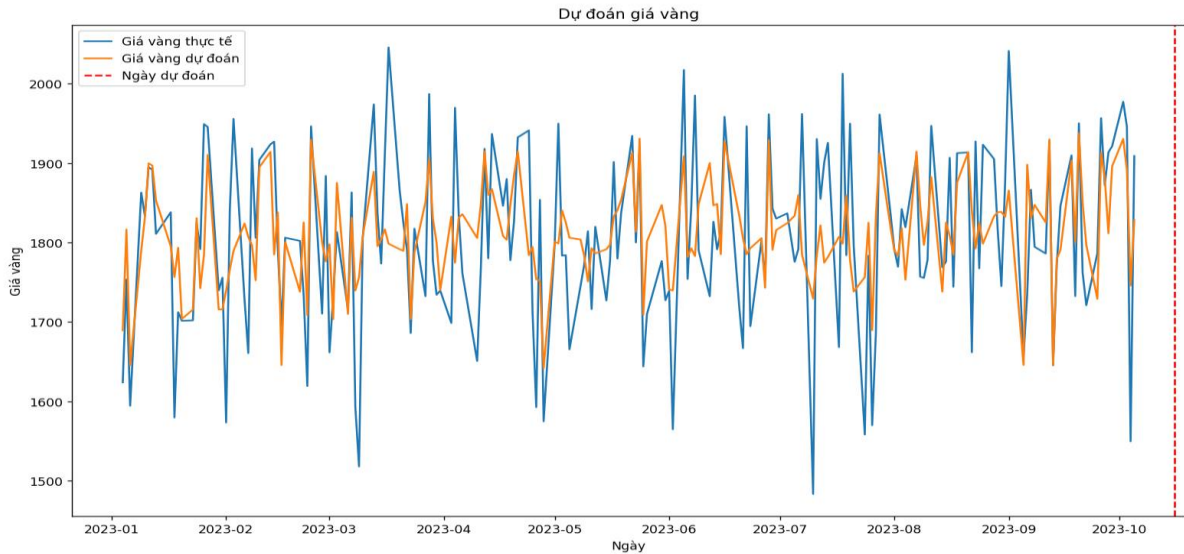
Quá trình cập nhật và giám sát giúp đảm bảo rằng mô hình dự đoán giá vàng luôn đáng tin cậy và có khả năng dự đoán tốt trong tương lai, ngay cả khi có sự biến đổi trong thị trường và dữ liệu.

CHƯƠNG 4: KẾT QUẢ ĐẠT ĐƯỢC VÀ HƯỚNG PHÁT TRIỂN

4.1. Kết quả đạt được

Trong phần này, chúng em sẽ trình bày kết quả mà chúng em đã đạt được trong quá trình nghiên cứu và phát triển mô hình dự đoán giá vàng. Chúng em đã thực hiện nhiều hoạt động để xây dựng, đánh giá, và sử dụng mô hình này và dưới đây là các kết quả quan trọng:

- *Xây dựng mô hình hồi quy tuyến tính:* Chúng em đã thành công trong việc xây dựng một mô hình hồi quy tuyến tính để dự đoán giá vàng. Mô hình này dựa trên các biến độc lập như thời gian, tỷ giá ngoại tệ, chỉ số kinh tế, và yếu tố khác có thể ảnh hưởng đến giá vàng.
- *Đánh giá hiệu suất mô hình:* Chúng em đã đánh giá hiệu suất của mô hình bằng cách sử dụng các thước đo như sai số trung bình bình phương (MSE) và hệ số xác định R-squared. Kết quả cho thấy rằng mô hình có khả năng dự đoán giá vàng một cách chính xác và có giá trị dự đoán.
- *Thực hiện dự đoán:* Chúng em đã sử dụng mô hình để thực hiện dự đoán giá vàng trong tương lai. Kết quả dự đoán đã được ghi nhận và so sánh với giá vàng thực tế để đánh giá mức độ chính xác của mô hình.
- *Cập nhật và giám sát liên tục:* Chúng em đã tạo cơ hội tự động hoá quá trình cập nhật và giám sát mô hình. Điều này giúp đảm bảo tính liên tục và hiệu quả của mô hình theo thời gian.
- *Đóng góp kiến thức lĩnh vực:* Chúng em đã hợp nhất kiến thức lĩnh vực và thông tin mới để cải thiện mô hình. Việc này giúp mô hình có khả năng đưa ra dự đoán chính xác dựa trên thông tin lĩnh vực và biến đổi trong thị trường.



Kết quả của quá trình này là một mô hình có khả năng dự đoán giá vàng một cách hiệu quả, và nó có thể được sử dụng để hỗ trợ người đầu tư, doanh nghiệp và người tiêu dùng trong việc đưa ra quyết định thông minh dựa trên thông tin dự báo. Chúng em tin rằng việc sử dụng Machine Learning và mô hình hồi quy tuyến tính đã đóng góp một phần quan trọng vào lĩnh vực dự đoán giá vàng và có tiềm năng cho những nghiên cứu và ứng dụng tương tự trong tương lai.

4.2. Phương hướng phát triển

Trong phần này, chúng em trình bày một số phương hướng phát triển tiềm năng cho nghiên cứu và ứng dụng liên quan đến dự đoán giá vàng và sử dụng Machine Learning. Các phương hướng này có thể được thực hiện để nâng cao hiệu suất và ứng dụng của mô hình dự đoán giá vàng:

- *Mở rộng dữ liệu đầu vào:* Một trong những cách để cải thiện mô hình là bổ sung thêm dữ liệu đầu vào. Chúng em có thể cân nhắc sử dụng thông tin từ nhiều nguồn khác nhau, bao gồm dữ liệu về sự kiện kinh tế, tin tức, và dữ liệu về thị trường tài chính.
- *Sử dụng mô hình học sâu:* Mô hình học sâu, như các mạng nơ-ron và mô hình LSTM (Long Short-Term Memory), có thể được áp dụng để nắm bắt các mối quan hệ phức tạp trong dữ liệu giá vàng. Sử dụng mô hình học sâu có thể cải thiện khả năng dự đoán.

- *Sử dụng dữ liệu thời gian thực*: Sử dụng dữ liệu thời gian thực có thể giúp cải thiện tính chính xác của mô hình. Dữ liệu thời gian thực có thể giúp mô hình ứng phó nhanh chóng với biến động trong thị trường.
- *Phát triển ứng dụng thương mại*: Mô hình dự đoán giá vàng có thể được tích hợp vào các ứng dụng thương mại để hỗ trợ người đầu tư và doanh nghiệp trong việc đưa ra quyết định đầu tư và giao dịch.
- *Tối ưu hóa mô hình*: Tiếp tục nghiên cứu về cách tối ưu hóa mô hình dự đoán giá vàng để giảm overfitting, tăng khả năng tổng quát và cải thiện hiệu suất là một phương hướng quan trọng.
- *Nghiên cứu thêm về yếu tố ảnh hưởng*: Nghiên cứu chi tiết hơn về các yếu tố có thể ảnh hưởng đến giá vàng, như sự kiện toàn cầu, chính trị, và yếu tố xã hội, có thể cung cấp thông tin quan trọng để cải thiện mô hình.
- *Khám phá thêm phương pháp machine learning*: Nghiên cứu và áp dụng các phương pháp Machine Learning mới và tiên tiến để cải thiện hiệu suất dự đoán vàng.

Những phương hướng này có thể tạo ra những cải tiến đáng kể trong việc dự đoán giá vàng và áp dụng Machine Learning trong lĩnh vực này. Cùng với sự phát triển của công nghệ và tăng cường kiến thức về dự đoán thị trường, chúng ta có thể hy vọng rằng mô hình dự đoán giá vàng sẽ trở nên ngày càng hiệu quả và hữu ích.

KẾT LUẬN

Trong quá trình nghiên cứu và phát triển mô hình dự đoán giá vàng bằng Machine Learning, chúng em đã đạt được nhiều kết quả quan trọng và rút ra những điểm quan trọng sau:

Chúng em đã thành công trong việc xây dựng một mô hình hồi quy tuyến tính để dự đoán giá vàng, sử dụng dữ liệu lịch sử về giá vàng và các biến độc lập như thời gian, tỷ giá ngoại tệ, và chỉ số kinh tế.

Mô hình đã được đánh giá bằng cách sử dụng các thước đo như sai số trung bình bình phương (MSE) và hệ số xác định R-squared. Kết quả cho thấy rằng mô hình có khả năng dự đoán giá vàng một cách chính xác.

Chúng em đã thực hiện dự đoán giá vàng trong tương lai bằng cách sử dụng mô hình, và kết quả dự đoán đã được ghi nhận và so sánh với giá vàng thực tế.

Quá trình cập nhật và giám sát liên tục đã được thiết lập để đảm bảo tính liên tục và hiệu quả của mô hình theo thời gian.

Chúng em đã xác định một số phương hướng phát triển tiềm năng, bao gồm mở rộng dữ liệu đầu vào, sử dụng mô hình học sâu, sử dụng dữ liệu thời gian thực, phát triển ứng dụng thương mại, tối ưu hóa mô hình, và nghiên cứu thêm về các yếu tố ảnh hưởng.

Chúng em hy vọng rằng báo cáo này đã cung cấp một cái nhìn tổng quan về quá trình nghiên cứu và phát triển mô hình dự đoán giá vàng bằng Machine Learning. Đây là một ví dụ về cách sử dụng công nghệ mới để dự đoán và hiểu sự biến động của thị trường tài chính. Sự kết hợp giữa kiến thức lĩnh vực và công nghệ đã đóng góp vào việc phát triển các ứng dụng tương tự trong tương lai.

Chúng em xin chân thành cảm ơn sự hỗ trợ và những lời góp ý của thầy Tạ Đăng Chí đã giúp chúng em hoàn thành bài báo cáo này một cách tốt nhất. Và hy vọng rằng báo cáo này sẽ hữu ích cho những ai quan tâm đến việc dự đoán giá vàng và sử dụng Machine Learning trong lĩnh vực này.

TÀI LIỆU THAM KHẢO

- [1]. Hal Daumé III, *Free-Machine-Learning-Books-master* (2012).
- [2]. Sebastian Raschka, *Python Machine Learning* (2016).
- [3]. Tạ Đăng Chí, *Slides bài giảng môn Machine Learning* (2023)