

Giảng viên hướng dẫn:

**Thầy Nguyễn Thanh Bình**

**Thầy Nguyễn Hồ Duy Trí**

Sinh viên thực hiện:

**Nguyễn Tường Minh Quang 19522098**

**Lê Hữu Nghĩa 19521894**

**TP. Hồ Chí Minh, tháng 5 năm 2022**

**BÁO CÁO ĐỒ ÁN**

**HỆ HỖ TRỢ RA QUYẾT ĐỊNH**

**ĐỀ TÀI**

**DỰ ĐOÁN GIÁ VÀNG**

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

# **LỜI CẢM ƠN**

Trước tiên, nhóm chúng em dành lời cảm ơn sâu sắc đối với cuộc đời đã mang đến cho chúng em cuộc gặp định mệnh với môn học này. Cùng với đó, chúng em cảm ơn thầy Nguyễn Thanh Bình và thầy Nguyễn Hồ Duy Trí với tâm huyết dành cho môn học Hệ hỗ trợ ra quyết định. Hai thầy đã hướng dẫn, dẫn dắt chúng em tận tình trong suốt môn học. Chúng em cảm ơn cả hai thầy bằng lời cảm ơn sâu sắc nhất.

Thành phố Hồ Chí Minh, ngày 27 tháng 5 năm 2022

Nhóm sinh viên

# **NHẬN XÉT CỦA GIẢNG VIÊN**

# **MỤC LỤC**

[**LỜI CẢM ƠN** 1](#_Toc105443967)

[**NHẬN XÉT CỦA GIẢNG VIÊN** 2](#_Toc105443968)

[**MỤC LỤC** 3](#_Toc105443969)

[**DANH MỤC HÌNH ẢNH** 5](#_Toc105443970)

[**BÀI LÀM** 6](#_Toc105443971)

[**1. Tổng quan:** 6](#_Toc105443972)

[***1.1. Lý do chọn đề tài:*** 6](#_Toc105443973)

[***1.2. Mục tiêu:*** 6](#_Toc105443974)

[**2. Công cụ sử dụng:** 6](#_Toc105443975)

[**3. Ý nghĩa các cột trong dữ liệu giá vàng:** 7](#_Toc105443976)

[**4. Dữ liệu giá vàng:** 7](#_Toc105443977)

[**5. Bộ dữ liệu giá vàng** 8](#_Toc105443978)

[**6. Mô hình ARIMA** 8](#_Toc105443979)

[***6.1. Mô hình tự hồi quy AR – Auto Regression:*** 8](#_Toc105443980)

[***6.2. Mô hình trung bình trượt MA – Move Average:*** 9](#_Toc105443981)

[***6.3. Mô hình ARMA – Auto Regression Move Average*** 10](#_Toc105443982)

[***6.4. Mô hình ARIMA – Auto Regression Integrated Move Average*** 10](#_Toc105443983)

[***6.6. Phương pháp Box-Jenkins (BJ)*** 11](#_Toc105443984)

[**7. Dữ liệu code:** 13](#_Toc105443985)

[***7.1. Input – Output:*** 13](#_Toc105443986)

[***7.2. Chuẩn bị:*** 14](#_Toc105443987)

[***7.3. Nhận dạng p, d, q trong mô hình ARIMA:*** 20](#_Toc105443988)

[***7.4. Dự đoán tại thời điểm dữ liệu test:*** 22](#_Toc105443989)

[***7.5. Sai số:*** 25](#_Toc105443990)

[***7.6. Dự đoán 5 ngày tiếp theo:*** 25](#_Toc105443991)

[**8. Kết luận:** 27](#_Toc105443992)

[**9. Kiến thức nhận được:** 28](#_Toc105443993)

[**BẢNG PHÂN CÔNG CÔNG VIỆC** 29](#_Toc105443994)

[**THAM KHẢO** 30](#_Toc105443995)

# **DANH MỤC HÌNH ẢNH**

[Hình 1. Dữ liệu giá vàng tại website. 7](#_Toc105443942)

[Hình 2. Thêm thư viện cần thiết. 14](#_Toc105443943)

[Hình 3.. Lấy dữ liệu từ investing.com. 14](#_Toc105443944)

[Hình 4. Làm sạch dữ liệu. 15](#_Toc105443945)

[Hình 5. Kiểm tra dữ liệu. 16](#_Toc105443946)

[Hình 6. Thêm thư viện ARIMA. 16](#_Toc105443947)

[Hình 7. Thêm thư viện ADF và kiểm tra hợp lệ. 17](#_Toc105443948)

[Hình 8. Biểu đồ giá vàng. 17](#_Toc105443949)

[Hình 9. Kiểm tra mô hình thông qua 4 biểu đồ. 18](#_Toc105443950)

[Hình 10. Biểu đồ tính mùa vụ. 18](#_Toc105443951)

[Hình 11. Làm cho mô hình ARIMA có tính dừng. 19](#_Toc105443952)

[Hình 12. Kiểm tra tính dừng của mô hình. 20](#_Toc105443953)

[Hình 13. Mô hình ACF. 20](#_Toc105443954)

[Hình 14. Mô hình PACF. 21](#_Toc105443955)

[Hình 15.Mô hình ARIMA tốt nhất. 21](#_Toc105443956)

[Hình 16. Chia dữ liệu và huấn luyện. 22](#_Toc105443957)

[Hình 17. Dự đoán. 23](#_Toc105443958)

[Hình 18. Chỉnh lại cột index (datetime). 23](#_Toc105443959)

[Hình 19. So sánh dự đoán và thực tế. 23](#_Toc105443960)

[Hình 20. Dự báo giấ vàng trong khoảng thời gian của Test. 24](#_Toc105443961)

[Hình 21. Sai số MSE, RMSE. 25](#_Toc105443962)

[Hình 22. Tái huấn luyện và dự đoán 5 ngày kế tiếp. 25](#_Toc105443963)

[Hình 23. Tạo file datetime. 26](#_Toc105443964)

[Hình 24. Chỉnh lại index của file fcast. 26](#_Toc105443965)

[Hình 25. Biểu đồ dự đoán 5 ngày tiếp theo. 27](#_Toc105443966)

# **BÀI LÀM**

## **1. Tổng quan:**

### ***1.1. Lý do chọn đề tài:***

* Ban đầu chúng em đang có dự định làm dự đoán về giá của cổ phiếu vì thấy trong giai đoạn hiện nay cổ phiếu khá là thịnh hành. Sau khi chúng em chốt thì đã thấy có nhóm dự đoán về giá cổ phiếu nên cảm thấy nếu làm chung một dự đoán về giá trị nào đó thì sẽ khá là nhàm chán. Nên chúng em đã thảo luận tiếp và quyết định chọn một đề tài dự đoán khác là giá vàng.
* Bên cạnh đó, chúng em cũng muốn thử tìm hiểu xem thị trường vàng sẽ như thế nào và từ đó tạo ra những dự đoán mang tính nghiên cứu để xem xét.

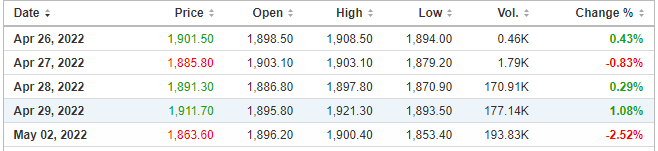
### ***1.2. Mục tiêu:***

* Dự báo giá vàng theo chuỗi thời gian.
* Tìm hiểu về chuỗi thời gian.
* Nghiên cứu và ứng dụng mô hình ARIMA trong dự báo giá vàng thế giới và kiểm tra sự chính xác của mô hình.

## **2. Công cụ sử dụng:**

* Ngôn ngữ: Python.
* Công cụ: Jupyter Notebook.

## **3. Ý nghĩa các cột trong dữ liệu giá vàng:**



Hình 1. Dữ liệu giá vàng tại website.

* Date: Ngày giao dịch.
* Price (Close): Giá trị vàng tại phiên gần nhất của ngày giao dịch (hoặc giá trị vàng tại phiên đóng nếu là ngày đã qua).
* Open: Giá trị vàng tại phiên mở.
* High: Giá trị vàng cao nhất trong ngày.
* Low: Giá trị vàng thấp nhất trong ngày.
* Vol (Volume): Số lượng giao dịch xảy ra trong ngày hôm đó.
* Change %: % thay đổi theo giá trị.
* Currency: đơn vị tiền tệ.

## **4. Dữ liệu giá vàng:**

* Vị trí khai thác: investing.com/commodities/gold ứng với dùng thư viện của Python Investpy.
* Dữ liệu do công ty Fusion Media thuộc quyền sỡ hữu.
* Fusion Media và các Nhà cung cấp Bên Thứ ba sẽ không chịu trách nhiệm dưới bất kỳ hình thức nào đối với việc chấm dứt, gián đoạn, chậm trễ hoặc không chính xác của bất kỳ thông tin thị trường nào.
* Dữ liệu của trang web có thể không được cung cấp bởi bất kì sàn giao dịch hay thị trường nào, nhưng có thể cung cấp bởi nhà tạo lập thị trường (công ty hoặc cá nhân nào đó tự định giá).
* Nếu muốn dùng nền tảng này để kinh doanh hãy đọc kỹ những điều khoản của công ty. Hầu hết ít có lợi cho khách hàng.

## **5. Bộ dữ liệu giá vàng**

* Dữ liệu giá vàng sẽ được lấy từ ngày 01/08/2020 -> ngày hiện tại (06/06/2022)
* Số dòng dữ liệu 488 dòng.
* Số cột: 5 (Date, Open, High, Low, Close).
* Cột đã loại bỏ: Volume, Currency.
* Dữ liệu đã được kiểm tra bằng hàm trong Python, những ô trống và những ô có giá trị 0 sẽ bị xóa.
* Dữ liệu nhận được sẽ chia ra làm hai phần: dữ liệu train (<30/05/2022), và dữ liệu test (>=30/05/2022).
* Dữ liệu Date (Index của dataset) lưu theo định dạng datetime và các dữ liệu khác sẽ lưu theo định dạng float. Cả bảng sẽ lưu trữ theo định dạng DataFrame của thư viện Pandas trong Python.

## **6. Mô hình ARIMA**

### ***6.1. Mô hình tự hồi quy AR – Auto Regression:***

* Là một mô hình cơ bản trong xác suất- thống kê.
* Mô hình này dự báo giá trị của chuỗi thời gian dựa trên một hoặc nhiều **giá trị** của chuỗi thời gian trước đó cộng với một giá trị ngẫu nhiên gọi là nhiễu trắng (white noise).
* Mô hình hồi quy bậc p như sau: AR(p)



*Trong đó:*

* Yt: là giá trị của chuỗi thời gian tại thời điểm t.
* : là giá trị trung bình của chuỗi thời gian.
* 1*,* 2 *,..*p: là các tham số của mô hình.
* ut là nhiễu trắng tại thời điểm t. Đây là phân số ngẫu nhiên có phân phối độc lập với giá trị trung bình là 0 và phương sai là

### ***6.2. Mô hình trung bình trượt MA – Move Average:***

* Mô hình này dự báo giá trị của chuỗi thời gian dựa trên một hoặc nhiều **giá trị ngẫu nhiên** nhiễu trắng của chuỗi thời gian trước đó.
* Mô hình trung bình trượt bậc q như sau: MA(q)

Yt - = ut - 1 *t-1 -* 2 *t-2 -… -* q *t-q*

*Trong đó:*

* Yt: là giá trị của chuỗi thời gian tại thời điểm t.
* : là giá trị trung bình của chuỗi thời gian.
* 1*,* 2 *,..*p: là các tham số của mô hình.
* ut là nhiễu trắng tại thời điểm t. Đây là phân số ngẫu nhiên có phân phối độc lập với giá trị trung bình là 0 và phương sai là

### ***6.3. Mô hình ARMA – Auto Regression Move Average***

* Là mô hình kết hợp cả hai mô hình tự hồi quy AR và mô hình trung bình trượt MA. Dó đó mô hình ARMA mang đặc tính của hai mô hình hồi quy và trung bình trượt.

### ***6.4. Mô hình ARIMA – Auto Regression Integrated Move Average***

* Được xem là mô hình phù hợp với dữ liệu tuyến tính.
* Mô hình ARMA là mô hình dự báo dữ liệu chuỗi thời gian với giả thiết chuối thời gian có tính dừng. Nhưng trong thực tế thì không, tức phải kết hợp với Integrated.
* Do đó trước khi được dự báo, chuỗi thời gian không dừng phải biến về thành chuỗi dừng. Đây là ý tưởng của mô hình ARIMA.
* Mô hình ARIMA với các tham số p, d, q được biểu diễn dưới dạng ARIMA (p, d, q).

*Trong đó:*

* p là bậc hồi quy AR.
* q là bậc trung bình động MA
* d là số lần chuỗi thời gian phải được tính sai phân cho đến khi chuỗi thời gian có tính dừng.
* Một quá trình ARIMA (p, 0, 0) có nghĩa là quá trình có tính dừng AR(p) thuần túy; một quá trình ARIMA (0, 0, q) có nghĩa là quá trình có tính dừng MA(q) thuần túy.

### ***6.6. Phương pháp Box-Jenkins (BJ)***

Phương pháp Box-Jenkins được đặt theo tên của hai nhà toán học, người đề xuất ra phương pháp này, đó là ***George Box*** và ***Gwilym Jenkins***. Phương pháp Box-Jenkins xác định các bước cần thực hiện để xây dựng và dự báo dữ liệu chuỗi thời gian bằng mô hình ARIMA. Phương pháp này gồm bốn bước:

* ***Bước 1. Nhận dạng*.** Tức là, tìm các giá trị thích hợp của p, d và q.
* ***Bước 2. Ước lượng*.** Sau khi đã nhận dạng các giá trị thích hợp của p và q, bước tiếp theo là ước lượng các thông số của các số hạng tự hồi quy và trung bình trượt trong mô hình. Đôi khi phép tính này có thể được thực hiện bằng phương pháp bình phương tối thiểu nhưng đôi khi ta phải sử dụng các phương pháp ước lượng phi tuyến (thông số phi tuyến). Do bây giờ công việc này có thể được thực hiện tự động bằng một số phần mềm thống kê, ta không cần phải lo lắng về trình tự toán học của phép ước lượng này.
* ***Bước 3. Kiểm tra chẩn đoán*.** Sau khi đã lựa chọn mô hình ARIMA cụ thể và ước lượng các tham số của nó, ta tìm hiểu xem mô hình lựa chọn có phù hợp với dữ liệu ở mức chấp nhận hay không bởi vì có thể một mô hình ARIMA khác cũng phù hợp với dữ liệu. Một kiểm định đơn giản về mô hình lựa chọn là xem xem các phần dư ước lượng từ mô hình này có tính ngẫu nhiên thuần túy hay không; nếu có, ta có thể chấp nhận sự phù hợp này của mô hình; nếu không, ta phải lặp lại từ đầu.
* ***Bước 4. Dự báo*.** Một trong số các lý do về tính phổ biến của phương pháp lập mô hình ARIMA là thành công của nó trong dự báo. Trong nhiều trường hợp, các dự báo thu được từ phương pháp này tin cậy hơn so với các dự báo tính từ phương pháp lập mô hình kinh tế lượng truyền thống, đặc biệt là đối với dự báo ngắn hạn. Tất nhiên, từng trường hợp phải được kiểm tra cụ thể.

#### 6.5.1. Nhận dạng

* Xác định tham số d: d đại diện cho số lần tính sai phân của chuỗi thời gian cho đến khi chuỗi thời gian đó dừng.
* Xác định tham số p, q: p, q lần lượt là tham số dùng để xác định số lượng các số hạng hồi quy quy và các số hạng trung bình động. Chọn giá trị p tại các độ trễ mà tại đó giá trị của hàm PACF khác 0 về mặt thống kê. Tương tự chọn các giá trị của q tại các độ trễ mà tại đó giá trị của hàm ACF khác 0 về mặt thống kê. Các
* ACF: Hàm tự tương quan và PACF là hàm tự tương quan riêng phần.

#### 6.5.2. Ước lượng

* Phương pháp bình phương tối thiểu là phương pháp thường được sử dụng để ước lượng các tham số của phương trình hồi quy.
* Ý tưởng chính là xác định các tham số của phương trình hồi quy sao cho khoảng cách giữa đồ thị của phương trình hồi quy đang cần xấp xỉ và đồ thị của phương trình hồi quy được tạo bởi các tham số là nhỏ nhất có thể.
* Khoảng cách này thường được tính bằng bình phương hiệu giữa hai giá trị là giá trị của phương trình hồi quy cần xấp xỉ và giá trị của phương trình hồi quy được tạo bởi các tham số.
* Chính vì vậy mà phương pháp này thường được gọi là phương pháp bình phương tối thiểu.
* Giá trị của các tham số hồi quy sao cho phương trình hồi quy đó xấp xỉ tốt nhất cho phương trình hồi quy đang xét, khi đó các giá trị này được xem như là tham số của phương trình hồi quy.

#### 6.5.3. Kiểm định

* Sau khi có các tham số thì kiểm tra mô hình được chọn có phù hợp với dữ liệu chuỗi thời gian dùng để xây dựng mô hình bằng ADF (p-value < 0.05 thì đạt).
* Quan sát đồ thị hàm tự tương quan ACF và tự tương quan riêng phần PACF để xem xét ngẫu nhiên thuần túy.

#### 6.5.4. Dự báo

* Trong nhiều trường hợp, các dự báo thu được từ phương pháp này tin cậy hơn so với các dự báo tính từ phương pháp lập mô hình kinh tế lượng truyền thống, đặc biệt là đối với dự báo ngắn hạn.
* Tuy nhiên dữ liệu luôn biến động nên mô hình ARIMA cần được cập nhật thường xuyên để đảm bảo tính chính xác của kết quả dự báo.

## **7. Dữ liệu code:**

### ***7.1. Input – Output:***

* Đầu vào: Bộ dữ liệu giá vàng.
* Đầu ra: Dự đoán giá vàng 5 ngày tiếp theo.

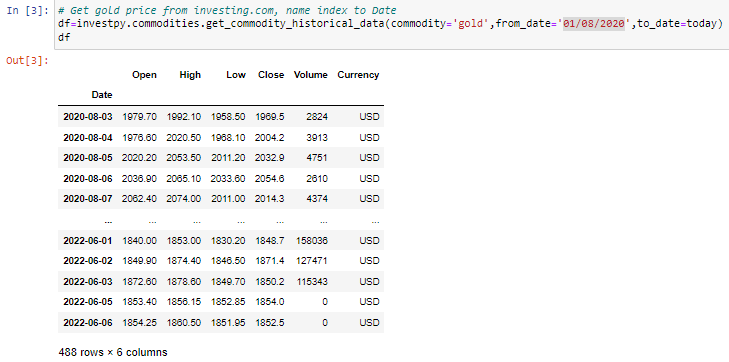
### ***7.2. Chuẩn bị:***

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 2. Thêm thư viện cần thiết.

* Thêm thư viện investpy được cung cấp bởi python để truy xuất dữ liệu trong investing.com.
* Bên cạnh đó lấy dữ liệu thời gian hiện tại lúc đang chạy code để xác nhận thời gian lấy dữ liệu tại trang investing.com.



Hình 3.. Lấy dữ liệu từ investing.com.

* Lấy dữ liệu giá vàng từ ngày 01/08/2020 – 6/6/2022.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 4. Làm sạch dữ liệu.

* Làm sạch dữ liệu, loại bỏ cột Volume, Currency.
* Đồng thời kiểm tra từng phần tử nếu có dòng rỗng sẽ trả về None.
* Kiểm tra từng dòng của df (file dataset) cột Open (giá cổ phiếu phiên mở) nếu có giá trị None sẽ loại bỏ khỏi df.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 5. Kiểm tra dữ liệu.

* Kiểm tra dữ liệu trước lúc làm việc với bảng df. Nếu dữ liệu Index có tên biến cũng như kiểu dữ liệu khác với bảng sau khi dự đoán thì sẽ có lỗi phát sinh.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 6. Thêm thư viện ARIMA.

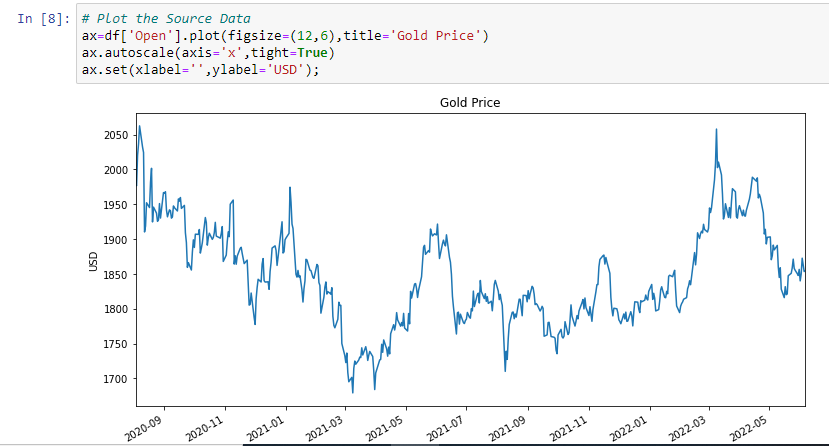
* Thêm thư viện để xây dựng mô hình ARIMA.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 7. Thêm thư viện ADF và kiểm tra hợp lệ.

* Thêm thư viện Dickey-Fuller Test cùng với thông báo để kiểm tra tính dừng của mô hình ARIMA. Mô hình ARIMA phải có tính dừng.



Hình 8. Biểu đồ giá vàng.

* Thống kê dữ liệu giá vàng từ ngày 01/08/2020 – 06/06/2022.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 9. Kiểm tra mô hình thông qua 4 biểu đồ.

* *Mục đích*: kiểm tra tính mùa vụ.
* *Biểu đồ 1:* Observations: biểu đồ giá Open.
* *Biểu đồ 2:* Trend: xu hướng biểu đồ giá Open.
* *Biểu đồ 3:* Sesonal: mùa vụ. Có thể thấy không có yếu tố mùa vụ ở đây có thể phân biệt như hình dưới.

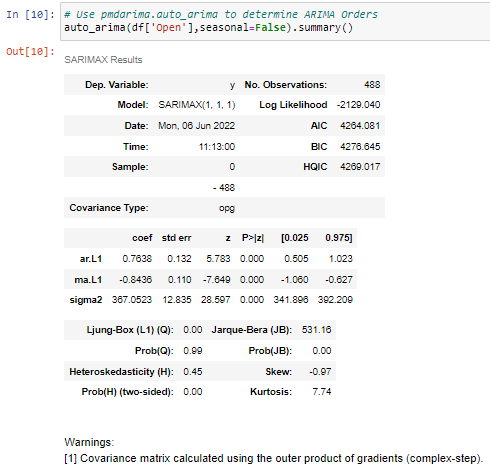
Ảnh có chứa văn bản, thước đo

Mô tả được tạo tự động

Hình 10. Biểu đồ tính mùa vụ.

[[1]](#footnote-1)

* *Biểu đồ 4:* Resid: phần dư của biểu đồ giá Open.



Hình 11. Làm cho mô hình ARIMA có tính dừng.

* Dùng thuật toán để làm cho mô hình dừng.

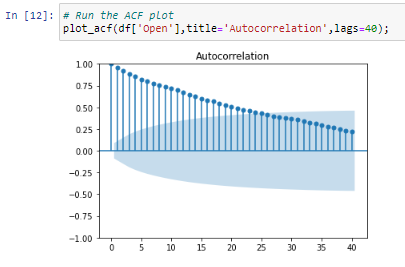
Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 12. Kiểm tra tính dừng của mô hình.

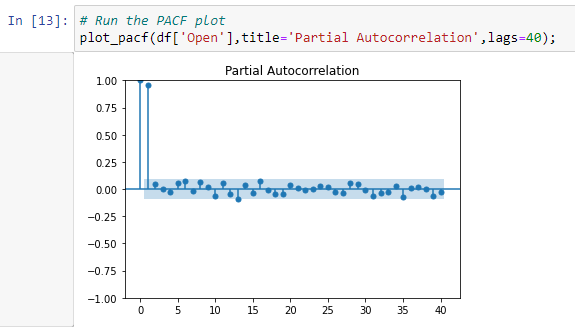
* Kiểm tra tính dừng của mô hình. P-value <= 0.05 mới tính là dừng.

### ***7.3. Nhận dạng p, d, q trong mô hình ARIMA:***



Hình 13. Mô hình ACF.

* Mô hình ACF: nhằm xác định vị trí độ trễ có giá trị nằm ngoài khoảng tin cậy để tìm ra q trong mô hình ARIMA (p, d, q).



Hình 14. Mô hình PACF.

* Mô hình PACF: nhằm xác định vị trí độ trễ có giá trị nằm ngoài khoảng tin cậy để tìm ra p trong mô hình ARIMA (p, d, q).

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 15.Mô hình ARIMA tốt nhất.

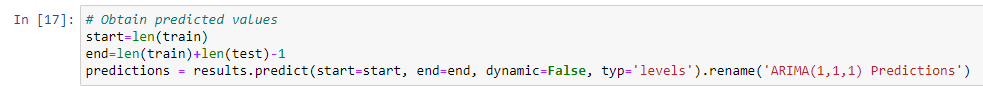
* Dùng thuật hàm auto\_arima để lựa tìm kiếm giá trị ARIMA (p, d, q) tốt nhất.
* Giá trị p, q thường nằm khoảng 0 – 5 nhưng trên mô hình PACF có giá trị là 0 và 1 nên max\_p, q lấy là 2.
* Giá trị m = 7 là lấy theo chu kì ngày. Nếu lấy 12 là tháng và 54 là tuần.
* Chọn được mô hình tốt nhất là ARIMA (1, 1, 1).

### ***7.4. Dự đoán tại thời điểm dữ liệu test:***



Hình 16. Chia dữ liệu và huấn luyện.

* Chia df ra làm dữ liệu train (< 30/5/2022) và dữ liệu test (>= 30/5/2022).
* Huấn luyện cho dữ liệu train theo mô hình ARIMA theo mô hình tốt nhất ARIMA (1, 1, 1).



Hình 17. Dự đoán.

* Tiến hành dự đoán giá trị trong khoảng thời gian của test 30/5/2022 – 6/6/2022.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 18. Chỉnh lại cột index (datetime).

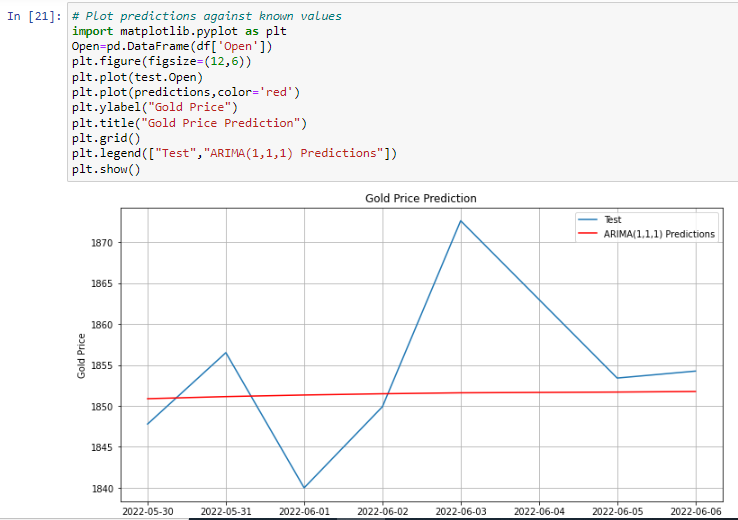
* Chỉnh lại cột Index của prediction về cột thời gian. Nhằm vẽ mô hình không bị lỗi.
* Sau đó tiến hành in giá trị dự đoán.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 19. So sánh dự đoán và thực tế.

* Tạo dataframe mới để so sánh giá trị dự đoán với giá trị thực của dữ liệu test.
* Có thể thấy giá trị dự đoán thứ 1 – 5 có khoảng cách khá nhỏ.



Hình 20. Dự báo giấ vàng trong khoảng thời gian của Test.

* Biểu đồ biểu thị giá trị vàng của df.Open (blue) và dự đoán giá trị dự đoán (red).

### ***7.5. Sai số:***

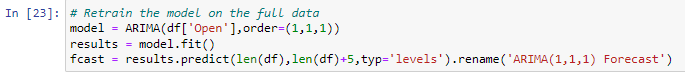
Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 21. Sai số MSE, RMSE.

* Đánh giá mô hình theo MSE, RMSE. Càng nhỏ càng tốt.

### ***7.6. Dự đoán 5 ngày tiếp theo:***



Hình 22. Tái huấn luyện và dự đoán 5 ngày kế tiếp.

* Tái huấn luyện lại dữ liệu (data huấn luyện df) và dự đoán 5 ngày tiếp theo. Lưu trữ tại fcast.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 23. Tạo file datetime.

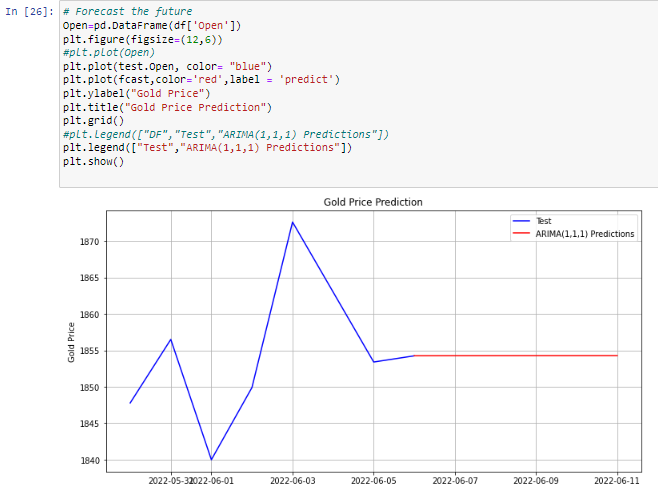
* Tạo file lưu trữ datetime cho 5 ngày tiếp theo.

Ảnh có chứa văn bản

Mô tả được tạo tự động

Hình 24. Chỉnh lại index của file fcast.

* Lưu trữ datatime cho fcast.index. Mục đích để vẽ hình dự đoán 5 ngày tiếp theo.



Hình 25. Biểu đồ dự đoán 5 ngày tiếp theo.

* Vẽ hình dữ liệu Test và dự đoán (do nếu là df thì sẽ rất lớn).

## **8. Kết luận:**

* Sau khi nghiên cứu mô hình ARIMA, có thể những dữ liệu về mùa vụ mô hình sẽ dự đoán tốt hơn.
* Những dự đoán bên trong mô hình làm việc ở trên sẽ có dao động trong khoảng thời gian ngắn 1 – 5 ngày. Nếu dài hơn thì khoảng chênh lệch vô cùng nhỏ.
* Với dữ liệu giá vàng, xu hướng biến động khá cao. Nếu dự đoán dài hạn rất khó. Vì mô hình sẽ hồi quy lại những dữ liệu trước đó từ đó sẽ tạo nên một đường thẳng với một góc Ampla nhất định.

## **9. Kiến thức nhận được:**

* Kỹ năng về Python.
* Kiến thức về mô hình ARIMA.
* Mô hình dự đoán và thực tế sẽ rất khó đối với các trường dữ liệu đầy biến động thường xuyên như vàng ở bài làm nhóm. Nếu là giá cổ phiếu hay coin, thì áp dụng mô hình ARIMA sẽ khó có thể dự đoán hơn.
* Nếu kết hợp ARIMA với các mô hình khác thì hiệu quả dự đoán được sẽ chắc chắn cao hơn. Ví dụ: SVM, ANN.
* Kỹ năng làm việc nhóm, đúng người đúng việc.

# **BẢNG PHÂN CÔNG CÔNG VIỆC**

|  |  |  |
| --- | --- | --- |
| **Công việc** | **Quang** | **Nghĩa** |
| ***Code chính*** | **X** |  |
| ***Chỉnh sửa code*** |  | **X** |
| ***Word*** |  | **X** |
| ***Slide*** | **X** |  |
| ***Thuyết trình*** | **X** |  |
| ***Tìm tài liệu*** |  | **X** |

# **THAM KHẢO**

|  |  |
| --- | --- |
| [1] | F. University, "KINH TẾ LƯỢNG VỀ CHUỖI THỜI GIAN II:DỰ BÁO VỚI MÔ HÌNH ARIMA VÀ VAR". |
| [2] | H. C. Hoài, "Luận văn thạc sĩ; Ngành Khoa học máy tính; Mã số: 60.48.01.01; Kết hợp mô hình ARIMA và support vector machine (SVM) để dự báo tại công ty dịch vụ trực tuyến cộng đồng Việt". |
| [3] | T. G. Smith, "pmdarima.arima.auto\_arima," [Online]. Available: https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto\_arima.html. [Accessed 20 5 2022]. |
| [4] | J. Brownlee, "How to Make Out-of-Sample Forecasts with ARIMA in Python," 24 3 2017. [Online]. Available: https://machinelearningmastery.com/make-sample-forecasts-arima-python/. [Accessed 20 5 2022]. |
| [5] | geeksforgeeks, "geeksforgeeks," [Online]. Available: https://www.geeksforgeeks.org/. [Accessed 15 5 2022]. |
| [6] | P. R. J. Hyndman, "Seasonal-arima," [Online]. Available: https://otexts.com/fpp2/seasonal-arima.html. [Accessed 15 5 2022]. |
| [7] | F. Media, "Investing," [Online]. Available: https://www.investing.com/commodities/gold. [Accessed 13 5 2022]. |
| [8] | S. Pulagam, "Time Series forecasting using Auto ARIMA in python," 27 6 2020. [Online]. Available: https://towardsdatascience.com/time-series-forecasting-using-auto-arima-in-python-bb83e49210cd. [Accessed 14 5 2022]. |
| [9] | phamdinhkhanh, "Bài 19 - Mô hình ARIMA trong time series," 12 9 2019. [Online]. Available: https://phamdinhkhanh.github.io/2019/12/12/ARIMAmodel.html. [Accessed 14 5 2022]. |

1. Biểu đồ tính mùa vụ, địa chỉ: <https://phamdinhkhanh.github.io/2019/12/12/ARIMAmodel.html>. Ngày truy cập 27/5/2022 [↑](#footnote-ref-1)