

BÁO CÁO ĐỒ ÁN CUỐI KHÓA KHOA HỌC DỮ LIỆU – DSP305

DỰ BÁO GIÁ CỦA NĂM ĐỒNG TIỀN MÃ HÓA TRONG NGẮN HẠN SỬ DỤNG MÔ HÌNH MACHINE LEARNING VÀ DEEP LEARNING

Mentor: Lê Văn Công | Học viên: Nguyễn Thùy Linh – FX18208

Mục lục

01 Phân Tích Dữ Liệu Khám Phá

1. Hiểu Biết Về Nghiệp Vụ
2. Hiểu Biết Về Dữ Liệu
3. Phân Tích Dữ Liệu
4. Phương Pháp Giải Quyết Vấn Đề

02 Lập Mô Hình Và Đánh Giá

1. Mô Hình ARIMA
2. Mô Hình ARIMA – GARCH
3. Mô Hình LSTM

03 Cải Thiện Mô Hình

1. Chuẩn Bị Dữ Liệu
2. Lập Mô Hình
3. Đánh Giá Mô Hình

04 Kết Luận Và Dự Báo

1. Kết Luận
2. Dự báo giá trong thời gian 30 ngày
3. Đánh Giá Mô Hình

I. Phân Tích Dữ Liệu Khám Phá



Hiểu Biết Về Nghiệp Vụ



Hiểu Biết Về Dữ Liệu



Phân Tích Dữ Liệu



Phương Pháp Giải Quyết Vấn Đề

1. Hiểu Biết Về Nghiệp Vụ

Chủ đề

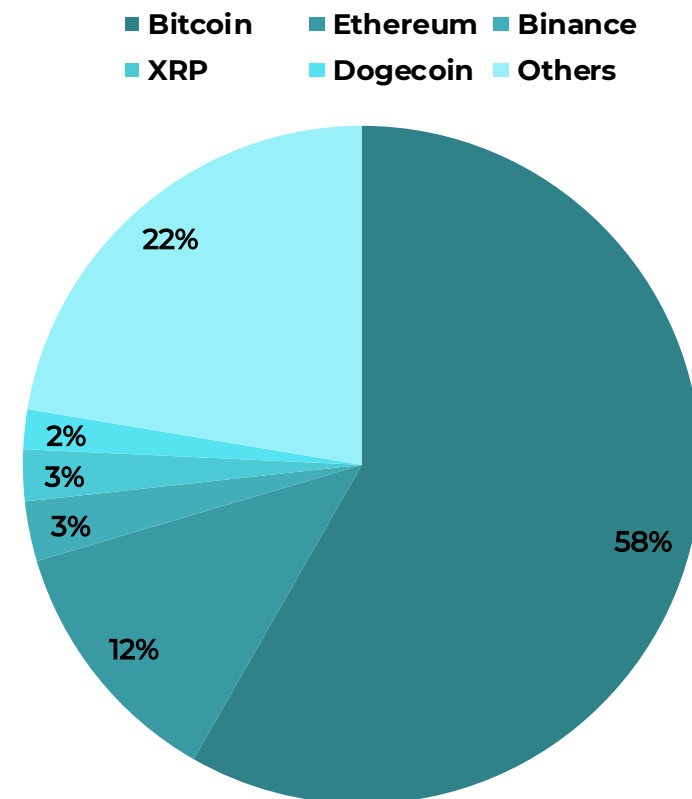
Thị trường tiền mã hóa bùng nổ trong những năm gần đây đã thu hút sự chú ý của nhiều nhà kinh tế, nhà đầu tư và nhà hoạch định chính sách.

Việc dự báo giá cả của tiền mã hóa trở thành đề tài hấp dẫn. Dự báo chính xác không chỉ giúp những nhà đầu tư sở hữu chúng đưa ra được quyết định đầu tư đúng đắn mà còn giúp những nhà kinh tế hiểu được biến động trong ngắn hạn của thị trường tiền mã hóa, từ đó đưa ra những chính sách kinh tế phù hợp.

Lĩnh vực nghiên cứu của đồ án là giá cả và những biến động về giá cả của năm đồng tiền mã hóa lớn: Bitcoin, Ethereum, Binance Coin, XRP và Dogecoin.

Giá trị vốn hóa của chúng chiếm khoảng 78% tổng giá trị vốn hóa của thị trường (tính tại thời điểm tháng 11.2024), quyết định sức khỏe của toàn bộ thị trường tiền mã hóa.

Giá Trị Vốn Hóa



1. Hiểu Biết Về Nghiệp Vụ

Mục tiêu

Sử dụng mô hình **ARIMA**, **ARIMA-GARCH** và **LSTM** để dự báo về giá cả của năm đồng tiền mã hóa lớn, bao gồm: Bitcoin, Ethereum, Binance Coin, XRP và DogeCoin, trong 30 ngày, từ ngày **30 – 11 – 2024** đến **29 – 12 – 2024**.

Vấn đề cần giải quyết

Xây dựng thước đo để đánh giá chất lượng một mô hình dự báo hiệu quả

Phát triển mô hình học máy và học sâu dự báo giá của năm đồng tiền mã hóa dựa trên dữ liệu lịch sử.

Đánh giá hiệu quả và tính chính xác của mô hình

Cải tiến mô hình

1. Hiểu Biết Về Nghiệp Vụ

Tiêu Chí Đánh Giá

Kết quả của mô hình được đánh giá bằng chỉ số: **Mean Absolute Percentage Error – MAPE**

| Tác giả | Bộ dữ liệu | MAPE |
|---|------------|---|
| Leonardo Felizardo và cộng sự (2019) | Bitcoin | MAPE của ARIMA = 17.586% MAPE của LSTM = 14.121% |
| Hashem Fekry Nematallah và cộng sự (2022) | Bitcoin | MAPE của LSTM = 2.34% |
| Karakoyun và Çibikdiken (2018) | Bitcoin | MAPE của ARIMA = 11.86% MAPE của LSTM = 1.40% |
| Guzalxon Belalova và cộng sự (2023) | Bitcoin | MAPE của ARIMA = 9.26% MAPE của LSTM = 1.48% |

So sánh kết quả các mô hình đã được thực hiện



Đồ án được đánh giá là đạt yêu cầu khi **MAPE** của mô hình tối ưu nhỏ hơn hoặc bằng **2.34%**

2. Hiểu Biết Về Dữ Liệu

| ID | Tên trường | Giải thích |
|----|------------|--|
| 1 | Date | Ngày quan sát |
| 2 | Open | Mức giá mở cửa ban đầu trước khi giao dịch |
| 3 | High | Mức giá cao nhất trong ngày |
| 4 | Low | Mức giá thấp nhất trong ngày |
| 5 | Close | Mức giá đóng cửa cuối cùng sau giao dịch |
| 6 | Volume | Số lượng giao dịch trong ngày |

1 Nguồn dữ liệu

Bộ dữ liệu Crypto-Currency trên Kaggle
Bao gồm giá và số lượng giao dịch **theo ngày** của 5 đồng tiền mã hóa lớn: Bitcoin, Ethereum, Binance Coin, XRP và Dogecoin

2 Thời gian nghiên cứu

Từ ngày 09 – 11 – 2017 đến ngày 29 – 11 – 2024

3 Số quan sát

Mỗi chuỗi thời gian có 2578 quan sát

4 Mô tả biến

Bộ dữ liệu gồm 6 chỉ số quan trọng của thị trường.
Biến phụ thuộc được sử dụng là biến **Close**

3. Phân Tích Dữ Liệu

Mô tả dữ liệu

Dữ liệu của 5 chuỗi thời gian đều toàn vẹn, không có giá trị khuyết.

Mô tả biến phụ thuộc **Close (giá đóng cửa)** trong bảng:

| | Bitcoin | Ethereum | Binance Coin | XRP | Dogecoin |
|-------|-----------|----------|--------------|-------|----------|
| count | 2,578 | 2,578 | 2,578 | 2,578 | 2,578 |
| mean | 26,890.75 | 1,464.75 | 213.20 | 0.53 | 0.073 |
| std | 20,714.00 | 1,210.00 | 206.18 | 0.32 | 0.092 |
| min | 3,236.76 | 84.21 | 1.51 | 0.14 | 0.001 |
| max | 98,977.66 | 4,812.09 | 710.46 | 3.38 | 0.685 |

Nhận xét

Mức giá trung bình của 5 đồng tiền mã hóa không đồng nhất

Độ lệch chuẩn của mỗi chuỗi lớn, mức biến động về giá cao, khoảng giá rộng.

3. Phân Tích Dữ Liệu

Khám phá dữ liệu

► Xu hướng chung

Giá tăng mạnh trong các giai đoạn **2017-2018**, **2020 – 2021**, **2023-2024**. Biên độ tăng lớn.

Xu hướng tăng giá qua các năm của Bitcoin, Ethereum và Binance Coin rõ ràng hơn so với các đồng còn lại.

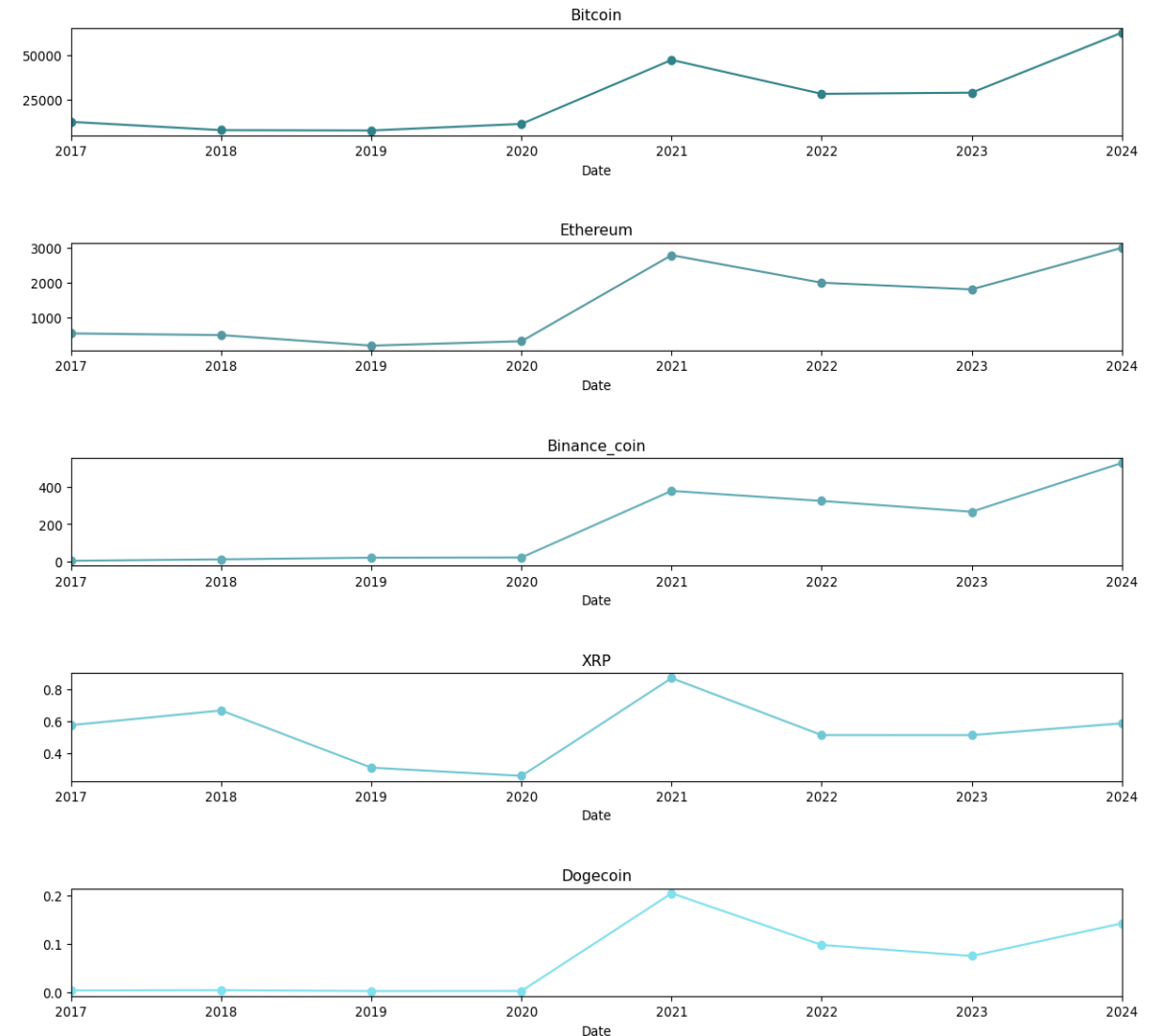
► Nguyên nhân biến động

Quy định pháp lý

Sự chấp nhận của thị trường và các tổ chức lớn

Tâm lý thị trường và đầu cơ

Tình hình kinh tế toàn cầu



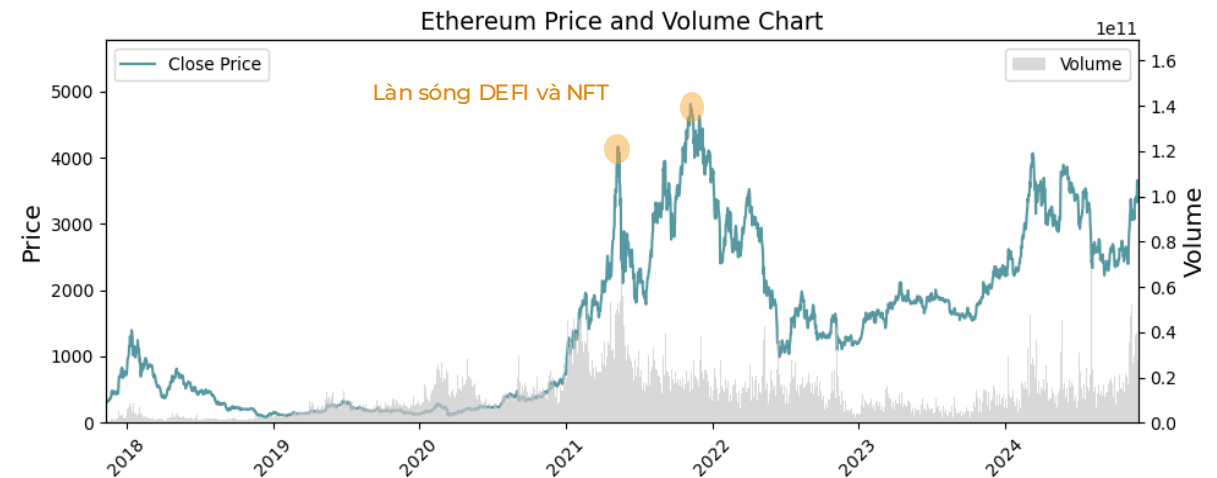
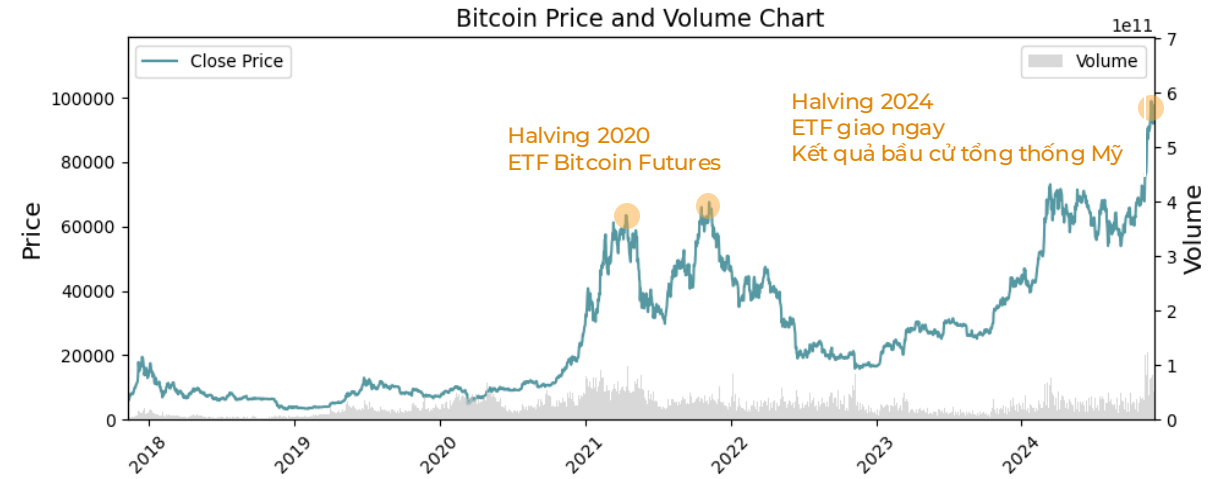
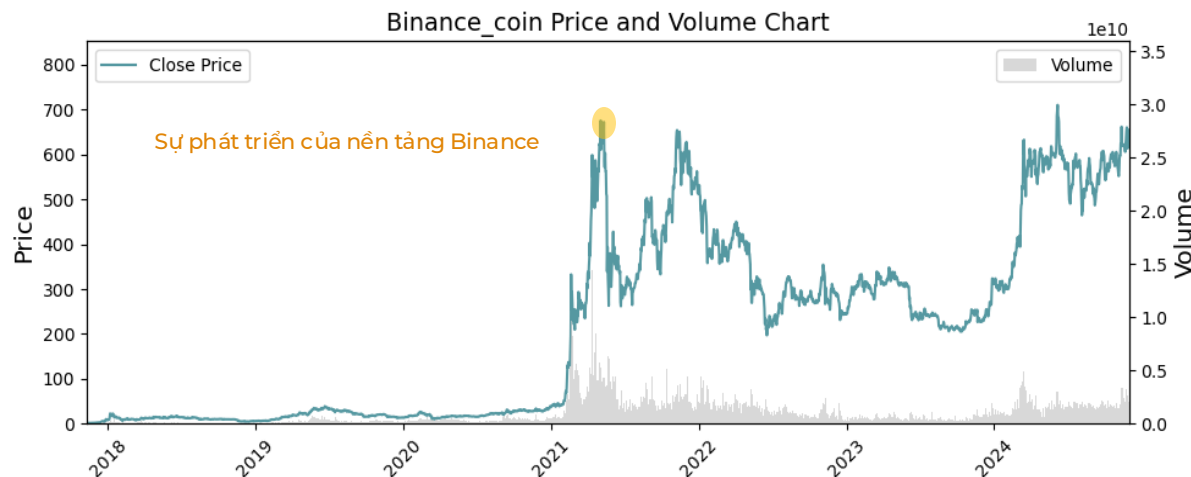
3. Phân Tích Dữ Liệu

Bitcoin, Ethereum, Binance Coin

Ethereum, Binance Coin có biến động giá tương tự Bitcoin, phản ứng đồng pha trước các sự kiện lớn.

Thường di chuyển sau Bitcoin trong nhiều chu kỳ thị trường (có độ trễ so với Bitcoin).

Số lượng giao dịch duy trì cao ổn định.

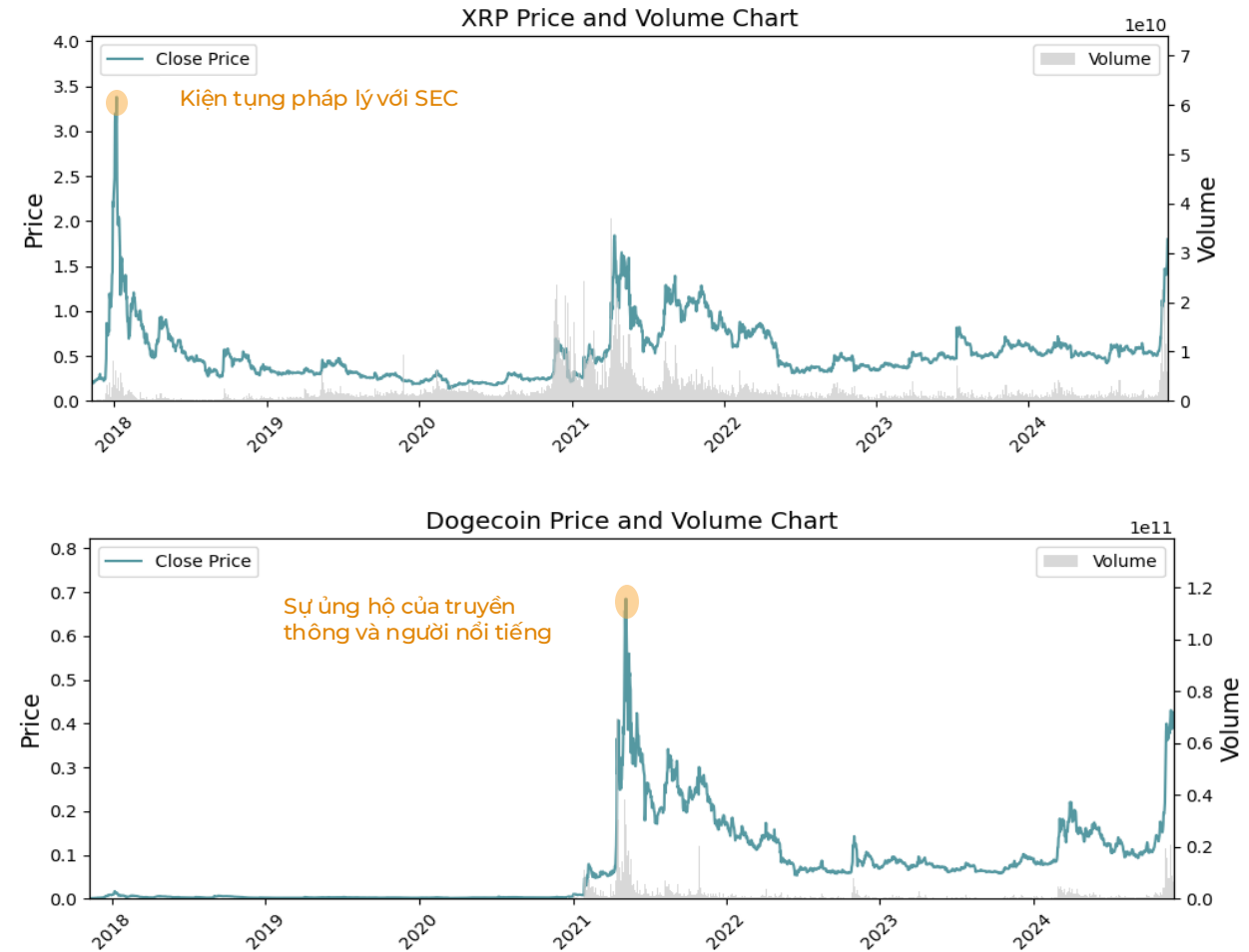


3. Phân Tích Dữ Liệu

XRP, Dogecoin

XRP có sự tăng trưởng rất mạnh trong giai đoạn 2017 – 2018. Tuy nhiên vụ kiện pháp lý với SEC đã làm mức giá XRP giảm mạnh và chưa có dấu hiệu phục hồi thật sự rõ ràng.

Dogecoin có mức giá ổn định trong giai đoạn đầu ra mắt, từ 2017 đến 2020 và bùng nổ mạnh mẽ vào năm 2021.



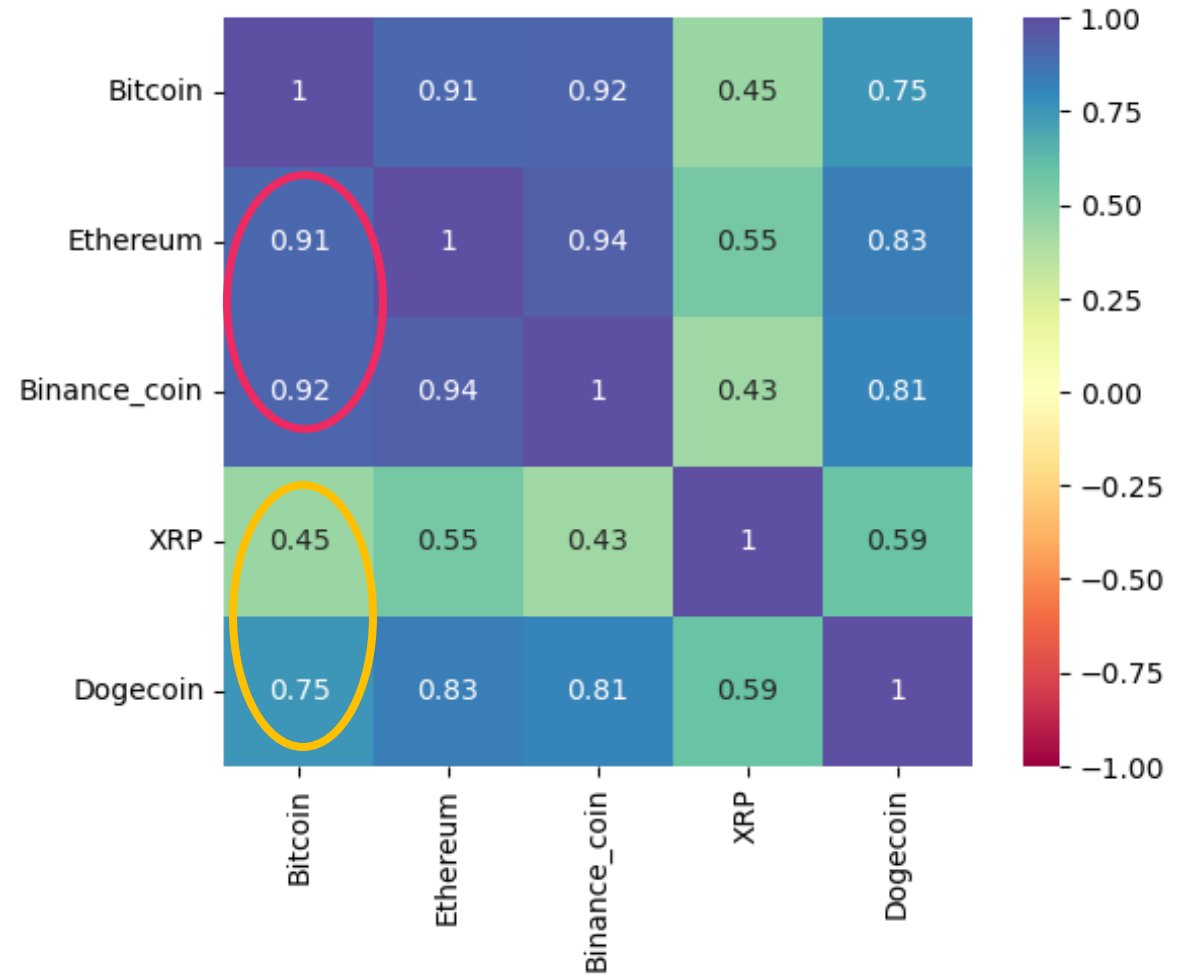
3. Phân Tích Dữ Liệu

Ma trận hệ số tương quan

Hệ số tương quan dương, thể hiện mối quan hệ cùng chiều giữa các đồng tiền số.

ETH và **BNB** có mức tương quan cao nhất với **BTC**, thể hiện chúng chịu ảnh hưởng mạnh từ hành vi giá của **BTC**.

Tương quan của **DOGE** và **XRP** với Bitcoin ở mức thấp hơn, thể hiện ngoài tác động về giá của Bitcoin, chúng còn chịu tác động của các yếu tố phi thị trường khác.



4. Phương Pháp Giải Quyết Vấn Đề

Thuật toán

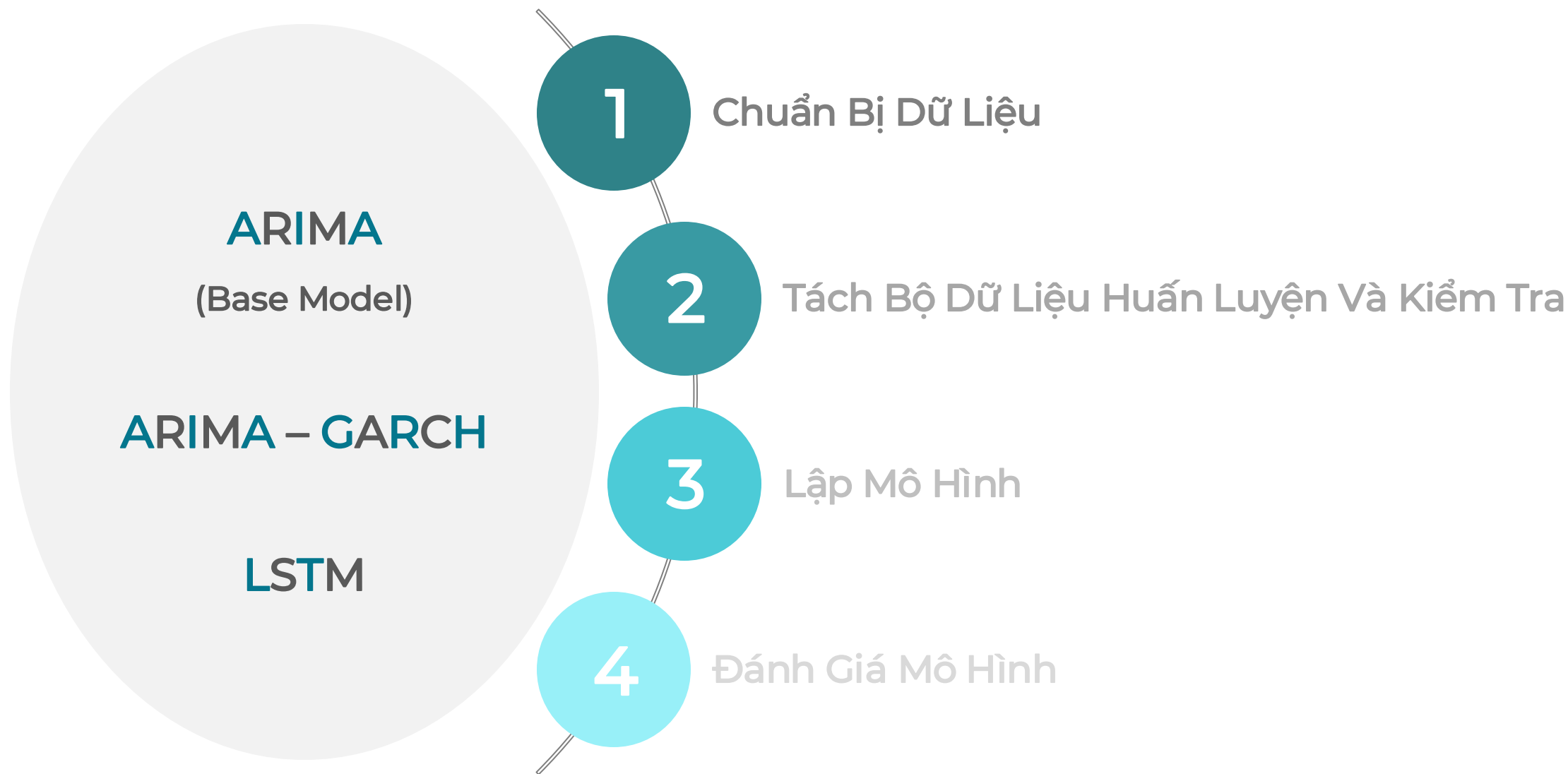


4. Phương Pháp Giải Quyết Vấn Đề

Môi Trường Làm Việc

| | Packages | Version |
|---|--------------|---------|
| 1 | python | 3.12.2 |
| 2 | pandas | 2.2.2 |
| 3 | numpy | 1.26.4 |
| 4 | matplotlib | 3.9.2 |
| 5 | scikit-learn | 1.5.1 |
| 6 | scipy | 1.13.1 |
| 7 | tensorflow | 2.17.0 |
| 8 | pmdarima | 2.0.4 |
| 9 | arch | 7.2.0 |

II. Lập Mô Hình Và Đánh Giá



1. Mô Hình ARIMA

Chuẩn Bị Dữ Liệu



- ☐ Phân Rã Chuỗi Thời Gian Theo Mùa
- ☐ Kiểm Định Tính Dừng
- ☐ ACF – PACF
- ☐ Xử Lý Dữ Liệu

Tách Bộ Dữ Liệu



- ☐ Dữ Liệu Huấn Luyện
- ☐ Dữ Liệu Kiểm Tra

Lập Mô Hình



- ☐ Lựa Chọn Tham Số Cho Mô Hình

Đánh Giá Mô Hình



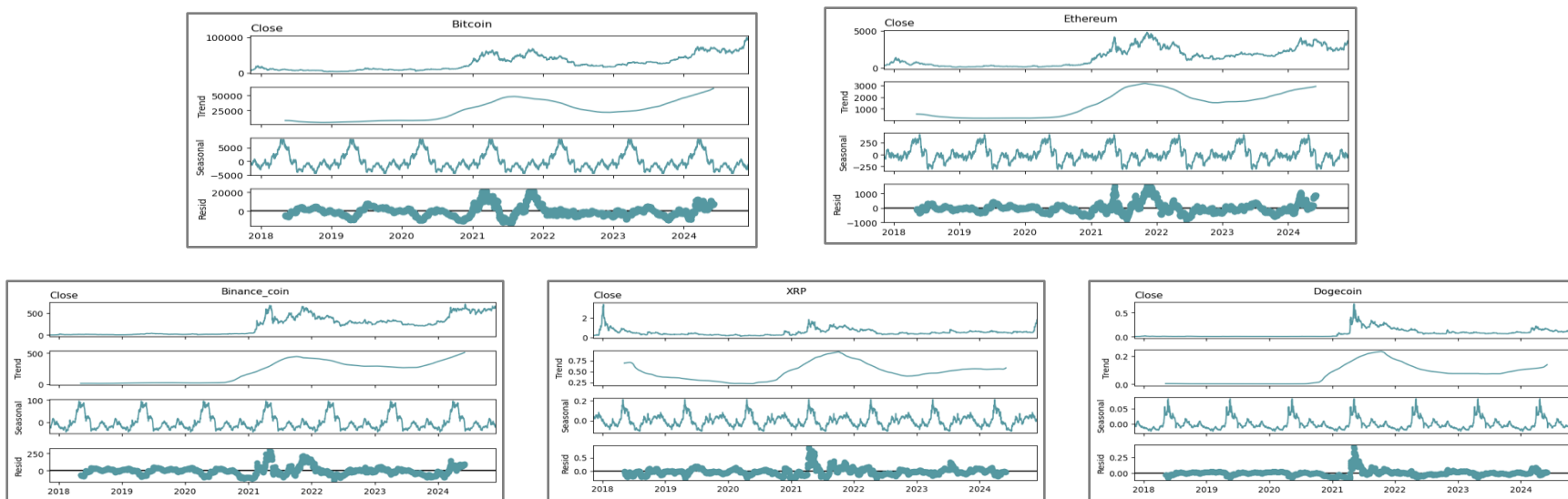
- ☐ Step = 15
- ☐ Step = 30
- ☐ Step = 120
- ☐ Test Data

Chuẩn Bị Dữ Liệu

Phân rã chuỗi thời gian theo mùa – Seasonal Decomposition

Phạm vi của seasonal – effect là rất nhỏ so với khoảng giá của 5 đồng tiền mã hóa.

Tác động của seasonal effect tới giá không đủ lớn.



Nhận xét

Tính thời vụ không xuất hiện trong cả 5 chuỗi thời gian.

Chuẩn Bị Dữ Liệu

Kiểm định tính dừng của dữ liệu

Thực hiện kiểm định **Augmented Dickey – Fuller test** (ADF test) để kiểm tra tính dừng của các chuỗi thời gian

| | P-value tại dữ liệu gốc | P-value tại sai phân bậc 1 |
|--------------|----------------------------|-------------------------------|
| Bitcoin | 0.976188 | 4.442514e-11 |
| Ethereum | 0.676049 | 1.431639e-23 |
| Binance Coin | 0.793634 | 5.955849e-17 |
| XRP | 0.000975 | 7.363113e-18 |
| Dogecoin | 0.216541 | 8.091548e-11 |

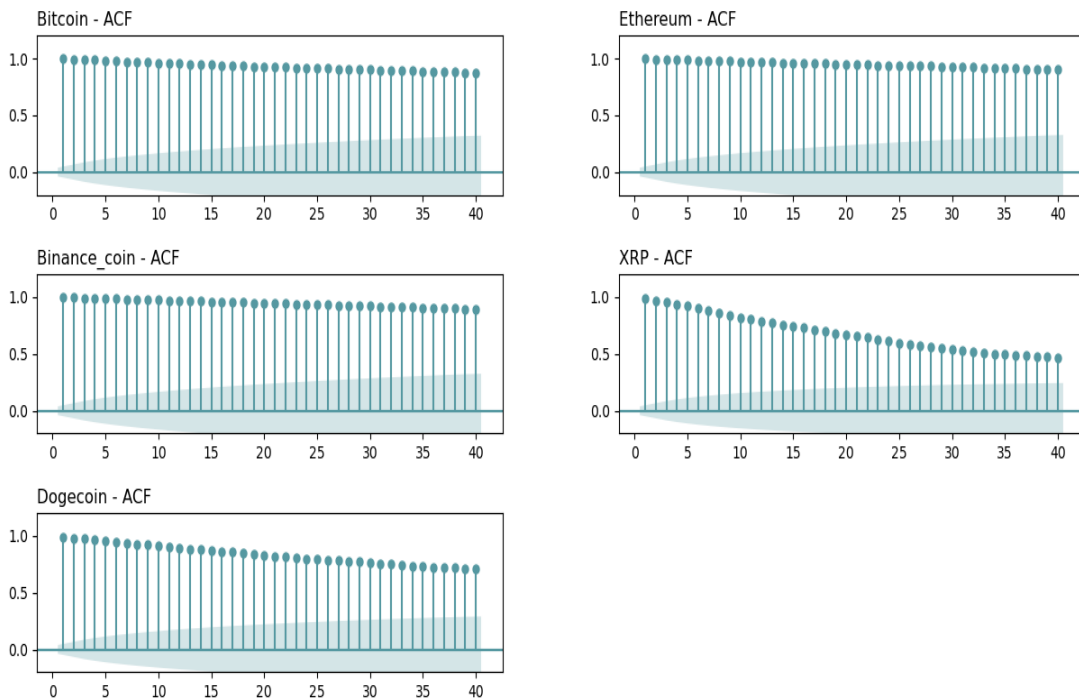
Kết quả:

Chuỗi thời gian **XRP** có tính dừng tại dữ liệu gốc.

Chuỗi thời gian **Bitcoin, Ethereum, Binance Coin và Dogecoin** không có tính dừng tại dữ liệu gốc, có tính dừng tại sai phân bậc 1.

Chuẩn Bị Dữ Liệu

Auto Correlation Function (ACF)



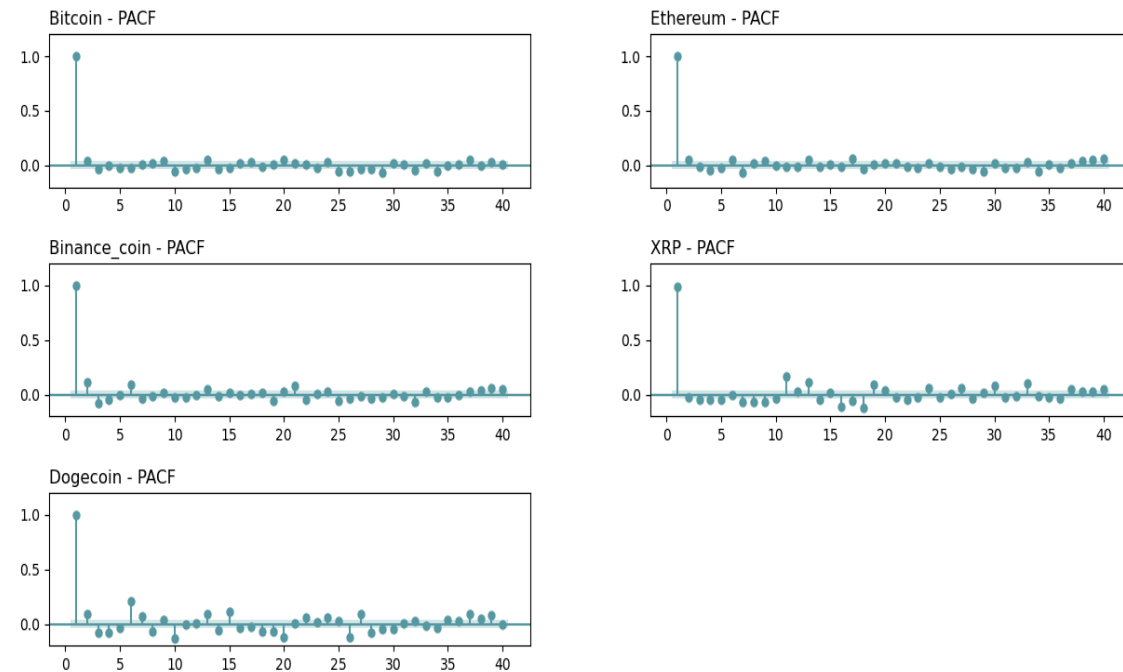
Hệ số tự tương quan cao và có ý nghĩa tại tất cả 40 lags.

⇒ **Chuỗi có yếu tố tự tương quan dài hạn**

Xu hướng giảm chậm và liên tục khi độ trễ tăng diễn ra tại BTC, ETH, BNB và DOGE

⇒ **Đặc điểm của chuỗi không dừng**

Partial Autocorrelation function (PACF)



Giá trị cao (xấp xỉ 1) tại lag=1 và cắt giảm nhanh sau đó.

Kể từ lag=2 trở đi, giá trị PACF không khác biệt đáng kể với 0.

⇒ **Đặc điểm của mô hình AR(1) hoặc có yếu tố phi tuyến trong chuỗi**

Chuẩn Bị Dữ Liệu

Xử lý dữ liệu

Gộp mức giá đóng cửa của 5 đồng tiền mã hóa vào một Data Frame

Index là giá trị 'Date'

column là giá trị 'Close' (giá đóng cửa)



Thực hiện biến đổi logarit trên giá đóng cửa

Giúp ổn định phương sai, giảm tác động của các giá trị ngoại lai và những thay đổi bất thường về giá.



Thực hiện sai phân bậc 1 trên dữ liệu logarit để kiểm tra tính dừng.

4 chuỗi thời gian: Bitcoin, Ethereum, Binance Coin, Dogecoin có tính dừng tại sai phân bậc 1

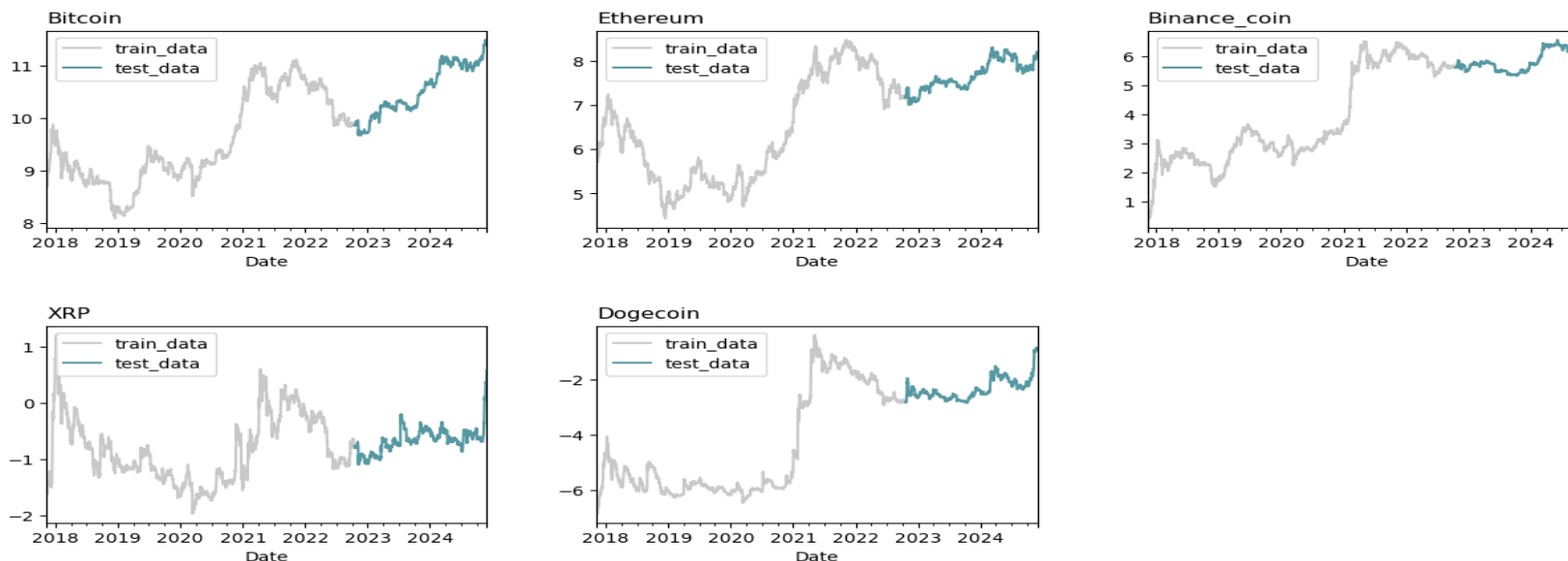
XRP có tính dừng tại dữ liệu gốc

Tách Bộ Dữ Liệu Huấn Luyện – Kiểm Tra

Dữ liệu huấn luyện (**train_data**) và dữ liệu kiểm tra (**test_data**) được tách từ bộ dữ liệu ban đầu với tỷ lệ 0.7/ 0.3. Trong đó:

Dữ liệu huấn luyện: là logarit mức giá đóng cửa của năm đồng tiền mã hóa trong khoảng thời gian từ ngày 09 – 11 – 2017 đến ngày 17 – 10 – 2022 (1804 quan sát)

Dữ liệu kiểm tra: là logarit mức giá đóng cửa của năm đồng tiền mã hóa trong khoảng thời gian từ ngày 18 – 10 – 2022 đến ngày 29 – 11 – 2024 (774 quan sát)



Lập Mô Hình

Lựa chọn tham số cho mô hình

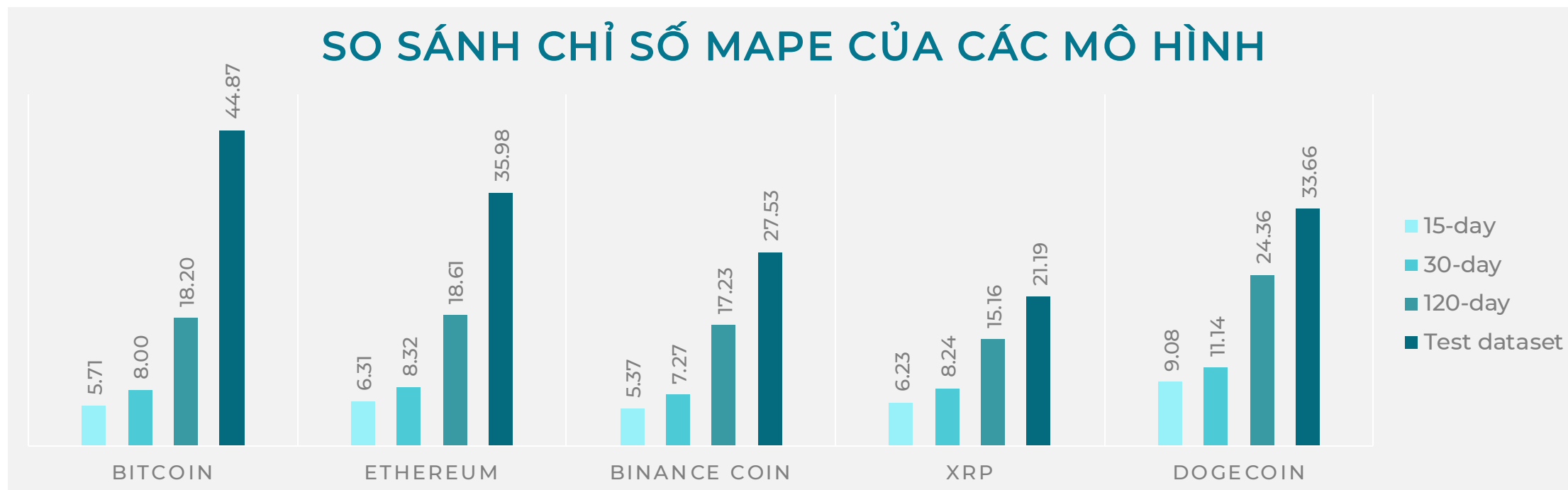
d được lựa chọn thông qua ADF test

(p, q) được lựa chọn bằng cách sử dụng `auto_arima()`, so sánh chỉ số **Akaike Information Criterion** (AIC) giữa các mô hình với nhau. Dựa trên kết quả đồ thị ACF và PACF, đồ án sẽ chọn p, q thuộc vùng giá trị **[0, 5]** cho cả 5 chuỗi thời gian.

Kết quả:

| | Mô hình ARIMA tối ưu | AIC |
|--------------|----------------------|--------------|
| Bitcoin | (2, 1, 0) | -6427.757589 |
| Ethereum | (2, 1, 0) | -5574.334059 |
| Binance Coin | (2, 1, 2) | -5073.777402 |
| XRP | (3, 0, 2) | -4778.437829 |
| Dogecoin | (3, 1, 2) | -4075.557053 |

Đánh Giá Mô Hình



Nhận xét

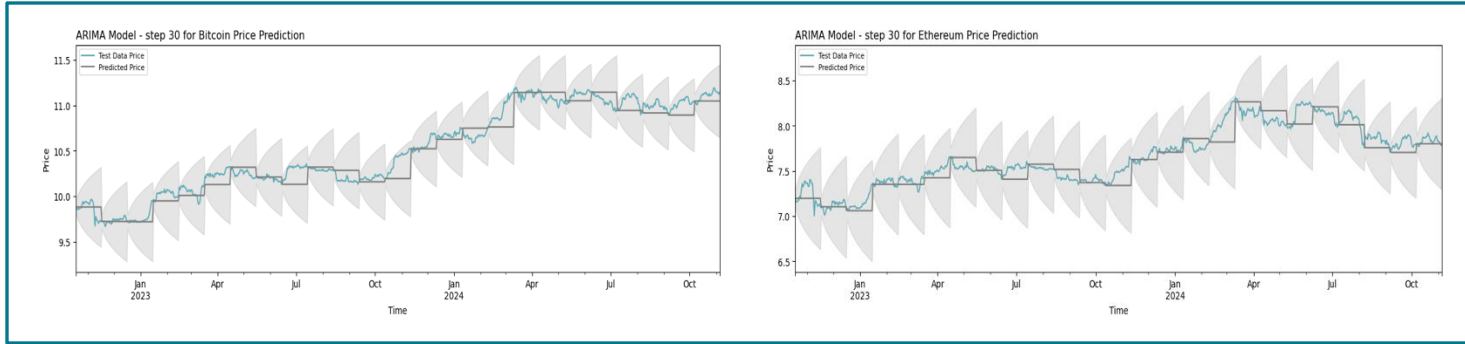
Độ chính xác của mô hình giảm dần khi khoảng thời gian dự báo tăng lên.

Mô hình ARIMA dự báo tốt trong những khoảng thời gian ngắn ngày nhưng không khái quát tốt cho khoảng thời gian dài hơn (đặc biệt là trên toàn bộ Test set).

Mô hình dự báo giá Bitcoin và Ethereum có độ chính xác thấp nhất trong dài hạn.

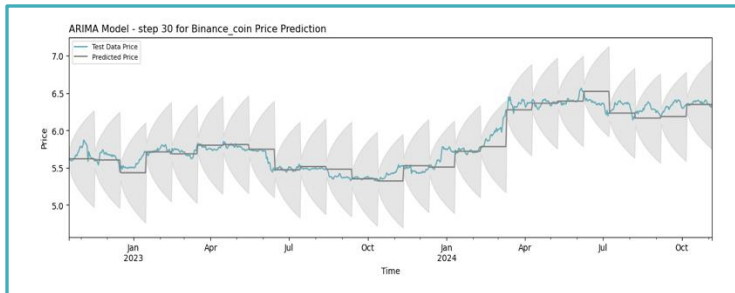
Đánh Giá Mô Hình

Step 30 Giá dự báo khá sát với giá thực tế, mô hình dự báo tốt trên tập kiểm tra.

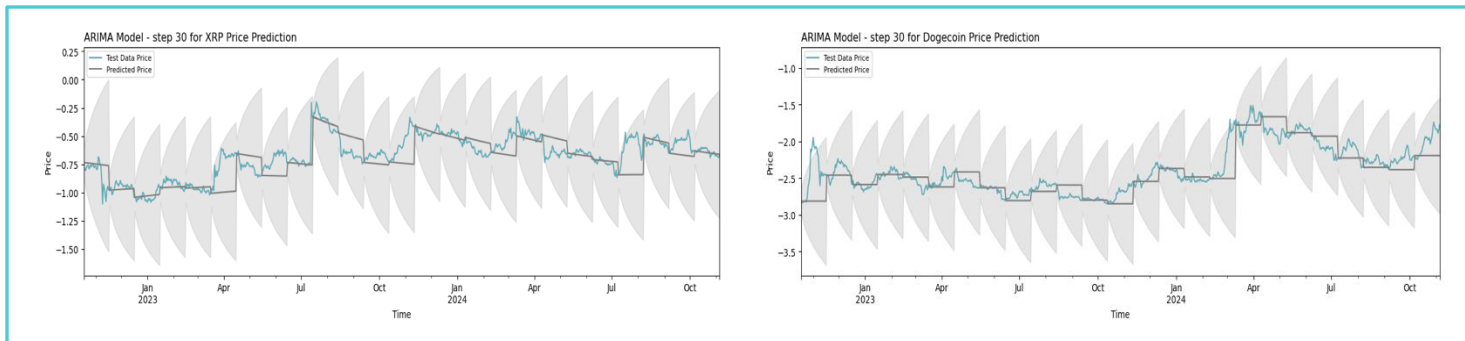


Giá dự báo bám khá sát nhưng có độ trễ so với giá thực tế.

Khoảng tin cậy tương đối ổn định.



Giá dự báo có độ trễ lớn hơn, đặc biệt tại một số điểm giá dao động mạnh.

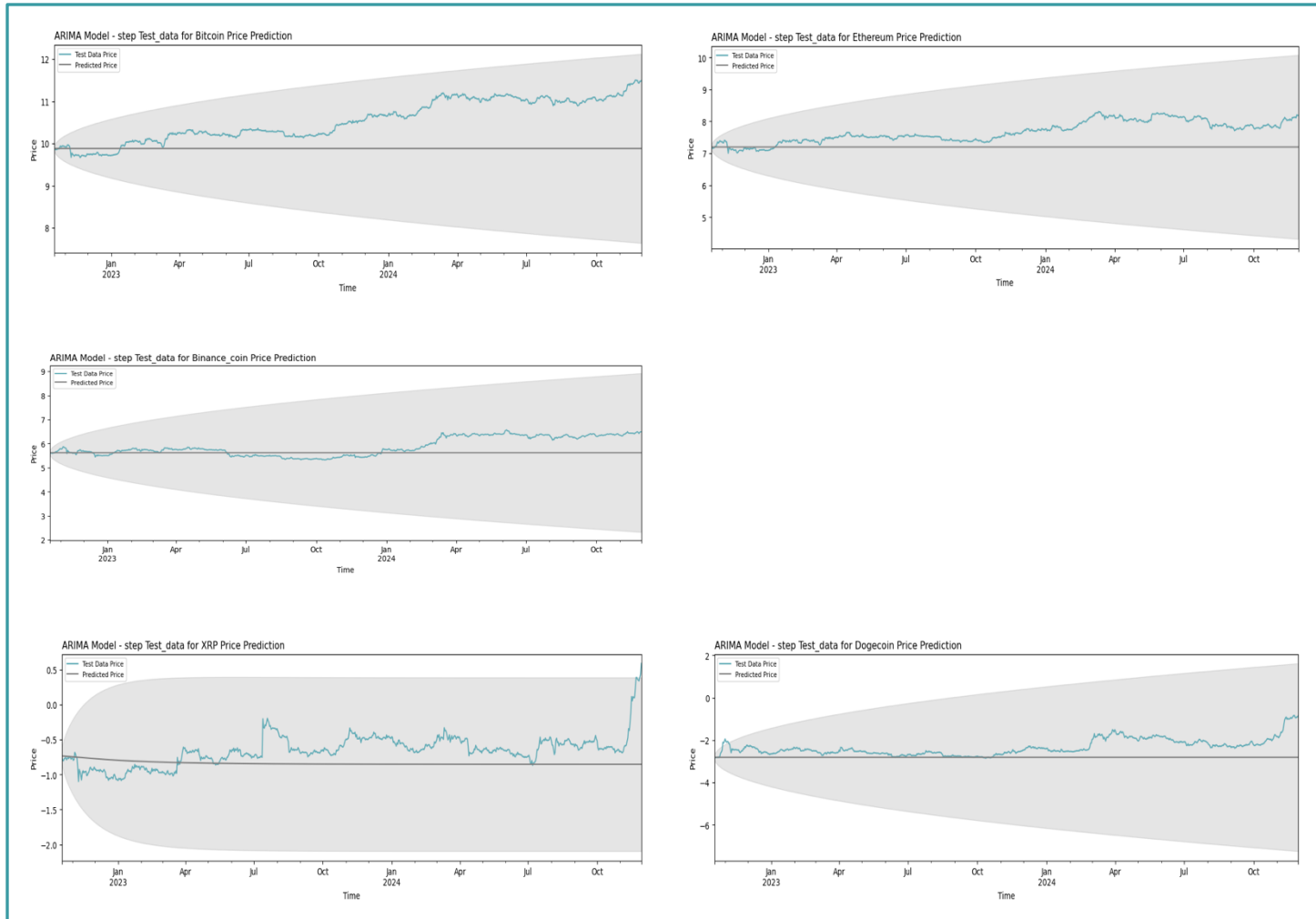


Giá dự báo không phản ánh được chính xác biến động tại một số điểm giá dao động mạnh hoặc đảo chiều.

Khoảng tin cậy rộng.

Đánh Giá Mô Hình

Test Data Giá dự báo không phản ánh được xu hướng thực tế, khả năng dự báo kém.



Giá dự báo đi ngang

Không phản ánh được dao động và xu hướng giá trong thực tế

Khoảng tin cậy tăng nhanh theo thời gian.

2. Mô Hình ARIMA – GARCH

Lập Mô Hình



☐ Lựa chọn tham số

Đánh Giá Mô Hình



☐ Step 15

☐ Step 30

☐ Step 120

☐ Test data

Lập Mô Hình

Lựa chọn tham số cho mô hình

(**p, d, q**) được lựa chọn bằng cách sử dụng hàm `forecast.auto_arima()` trên bộ dữ liệu `train_data`.

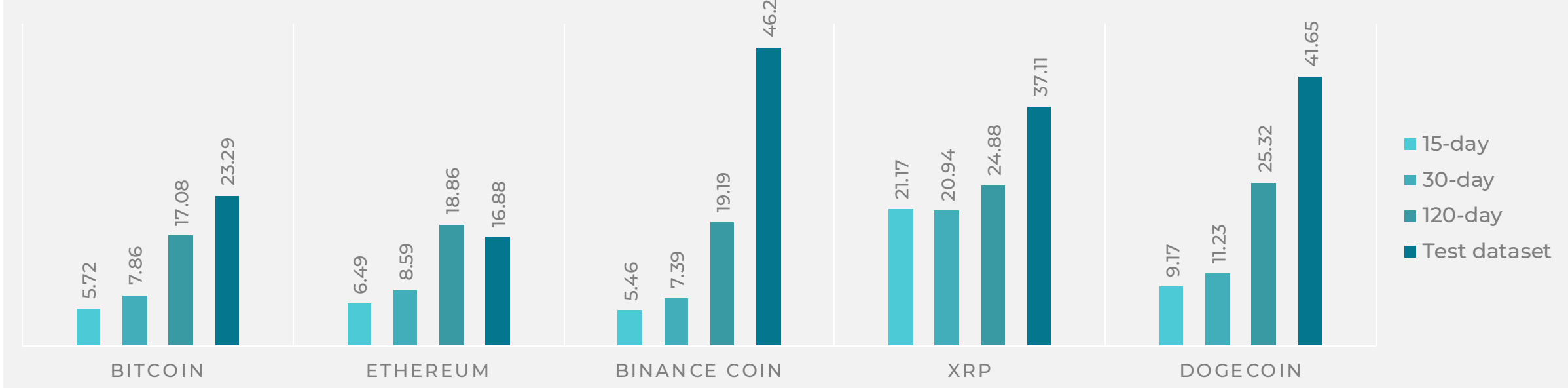
Mô hình GARCH được sử dụng là **GARCH (1, 1)**.

Kết quả:

| | Mô hình ARIMA-GARCH tối ưu | AIC |
|--------------|--------------------------------|--------------|
| Bitcoin | ARIMA (4, 1, 1) – GARCH (1, 1) | -6425.749322 |
| Ethereum | ARIMA (2, 1, 1) – GARCH (1, 1) | -5576.360408 |
| Binance Coin | ARIMA (3, 1, 1) – GARCH (1, 1) | -5070.881562 |
| XRP | ARIMA (0, 0, 0) – GARCH (1, 1) | -4779.778612 |
| Dogecoin | ARIMA (1, 1, 4) – GARCH (1, 1) | -4069.10033 |

Đánh Giá Mô Hình

SO SÁNH CHỈ SỐ MAPE CỦA CÁC MÔ HÌNH



Nhận xét

Đối với Bitcoin, Ethereum, Binance Coin và Dogecoin:

Mô hình dự báo tốt trong ngắn hạn, độ chính xác giảm dần khi thời gian dự báo tăng lên.

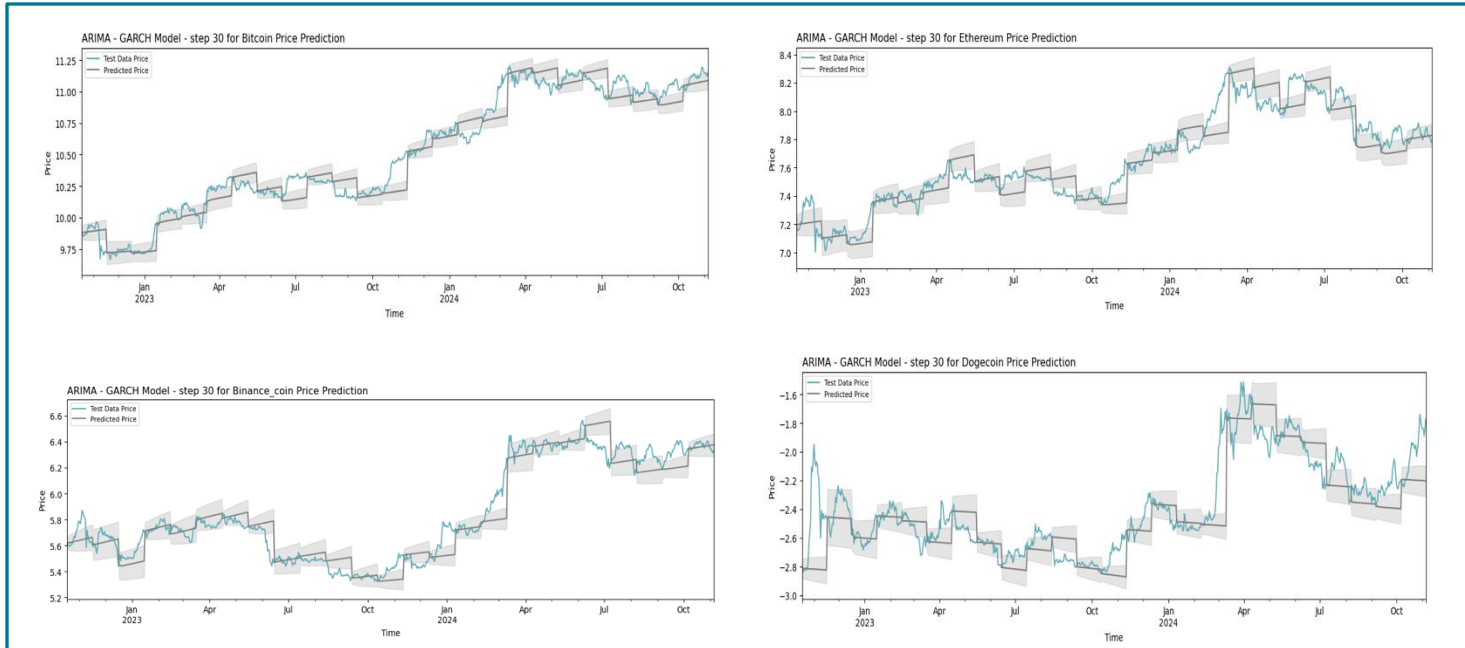
Mô hình dự báo được xu hướng tổng thể trong dài hạn của Bitcoin, Ethereum, những đồng có xu hướng tăng giá rõ ràng và ít dao động đột ngột. Nhưng không hiệu quả trong việc nắm bắt các xu hướng dài hạn của Binance Coin và Dogecoin, những đồng vốn có tính chất biến động và chịu nhiều ảnh hưởng của các yếu tố phi thị trường.

Đối với XRP:

Mô hình dự báo không nắm bắt được biến động cả trong ngắn hạn và dài hạn.

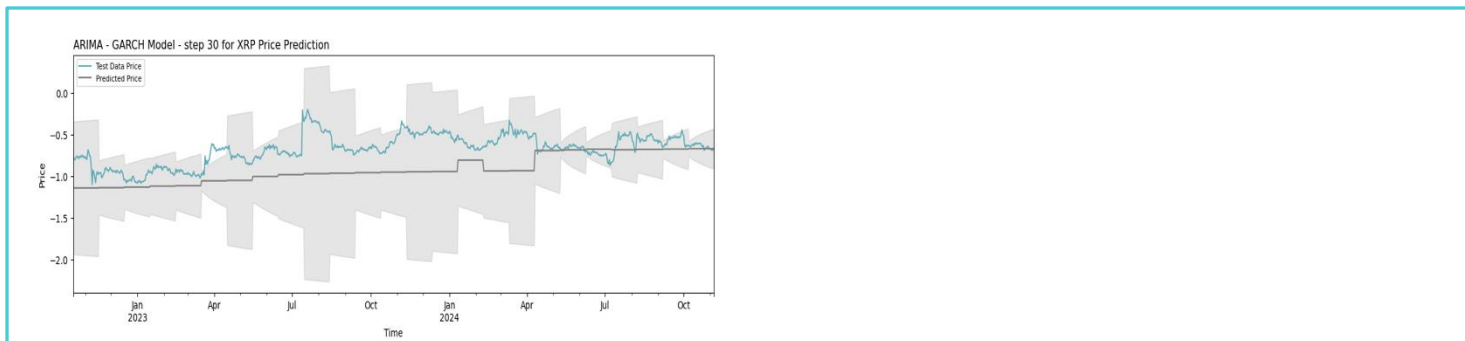
Đánh Giá Mô Hình

Step 30 Mô hình dự báo kém đối với XRP. Với các đồng còn lại, giá dự báo sát với giá thực tế, mô hình dự báo tốt trên tập kiểm tra.



Giá dự báo bám tương đối sát biến động thực tế.

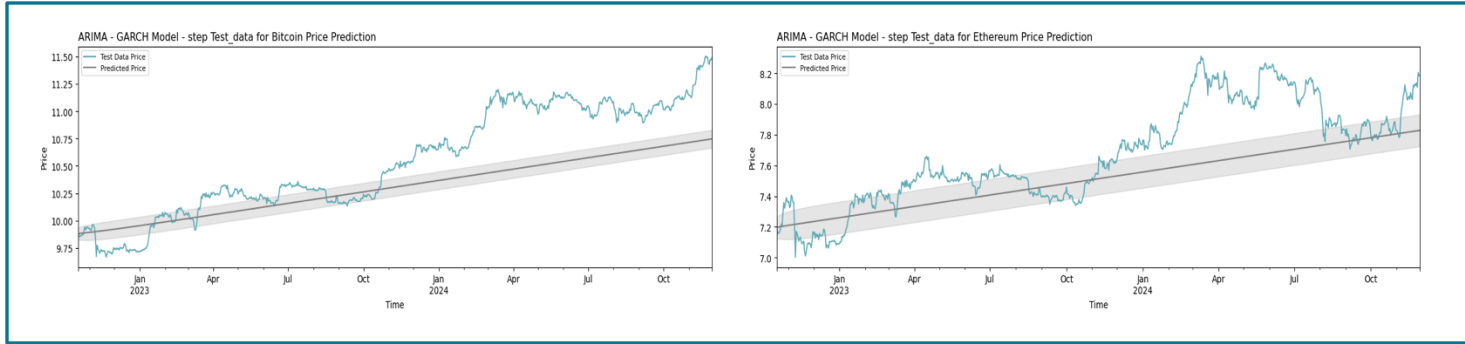
Riêng với Dogecoin, có độ trễ khi có biến động giá lớn



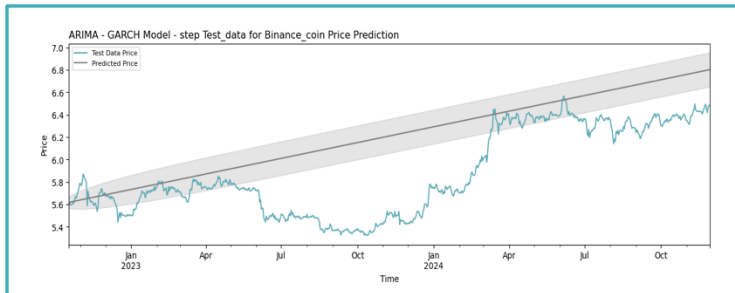
Mô hình dự báo kém
Không phản ánh được biến động thực tế của giá
Khoảng tin cậy lớn

Đánh Giá Mô Hình

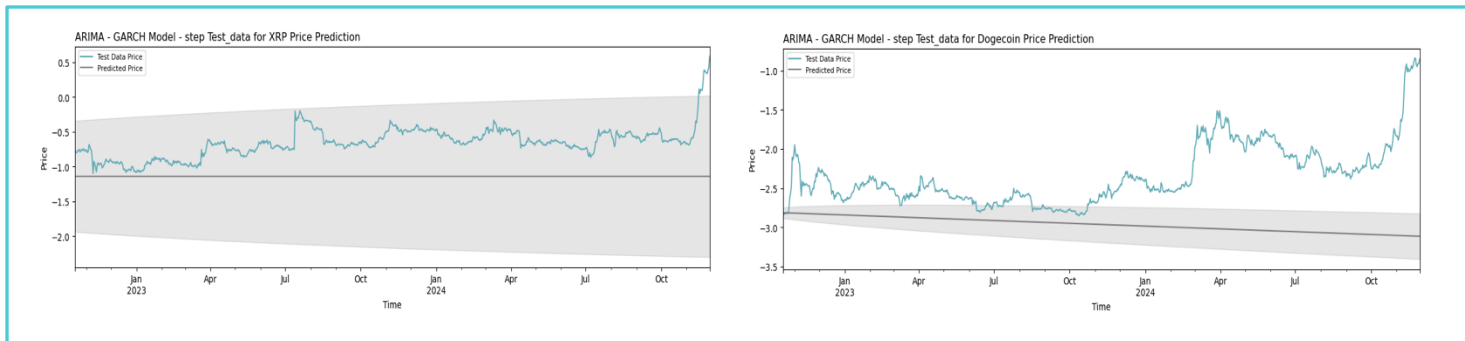
Test Data Mô hình hoạt động không ổn định, không nhất quán trên các bộ dữ liệu khác nhau



→ Dự báo được xu hướng tổng thể dài hạn
Không nắm bắt được biến động chi tiết



→ Không dự báo tốt các biến động giảm giá



→ Dự báo xu hướng ngược chiều/ đi ngang so với thực tế
Khoảng tin cậy rộng, sai số lớn

3. Mô Hình LSTM

Chuẩn Bị Dữ Liệu



☐ Xử Lý Dữ Liệu

Đánh Giá Mô Hình



☐ Step 15

☐ Step 30

☐ Step 120

☐ Test data

Chuẩn Bị Dữ Liệu

Xử lý dữ liệu

Chuyển đổi chuỗi thời gian thành dạng cửa sổ trượt (sliding window)

Target là giá trị Close (giá đóng cửa)

Feature là giá trị Close của 15 ngày trước đó

Thực hiện chuẩn hoá MinMaxScaler()

Chuyển giá trị dữ liệu về khoảng (0, 1)

Tăng tốc độ hội tụ và tăng khả năng dự báo của mô hình

Riêng với Dogecoin, thực hiện thêm biến đổi logarit trước khi chuẩn hóa

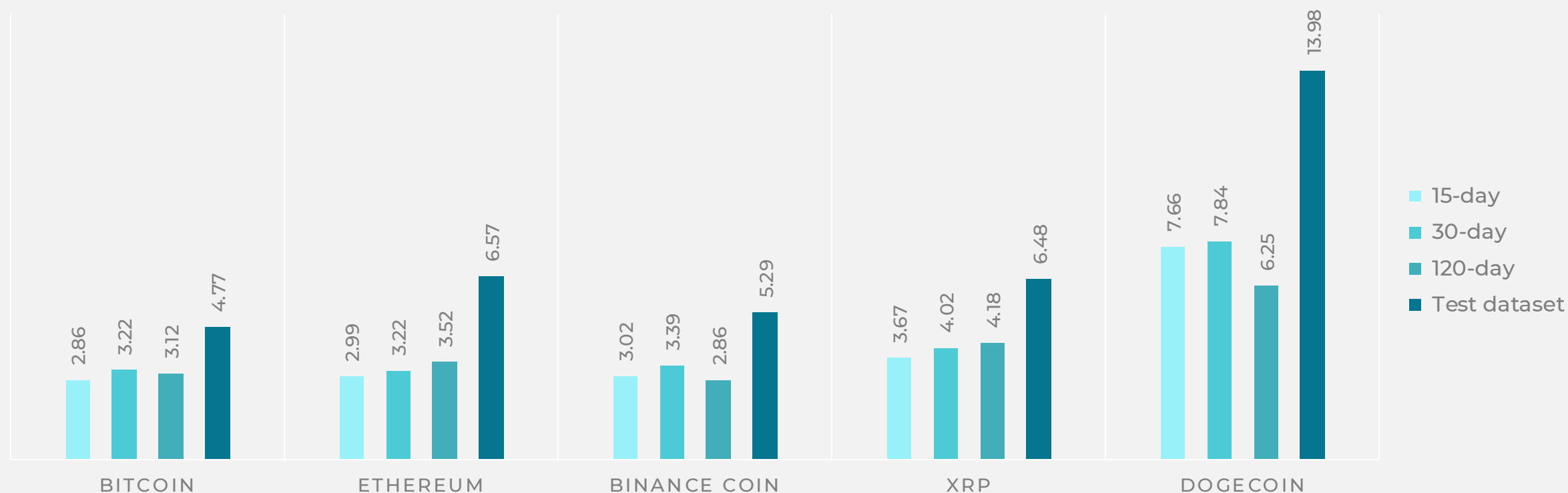
Tách bộ dữ liệu huấn luyện/ kiểm tra

Dữ liệu training: 70% bộ dữ liệu

Dữ liệu testing: 30% bộ dữ liệu

Đánh Giá Mô Hình

SO SÁNH CHỈ SỐ MAPE CỦA CÁC MÔ HÌNH

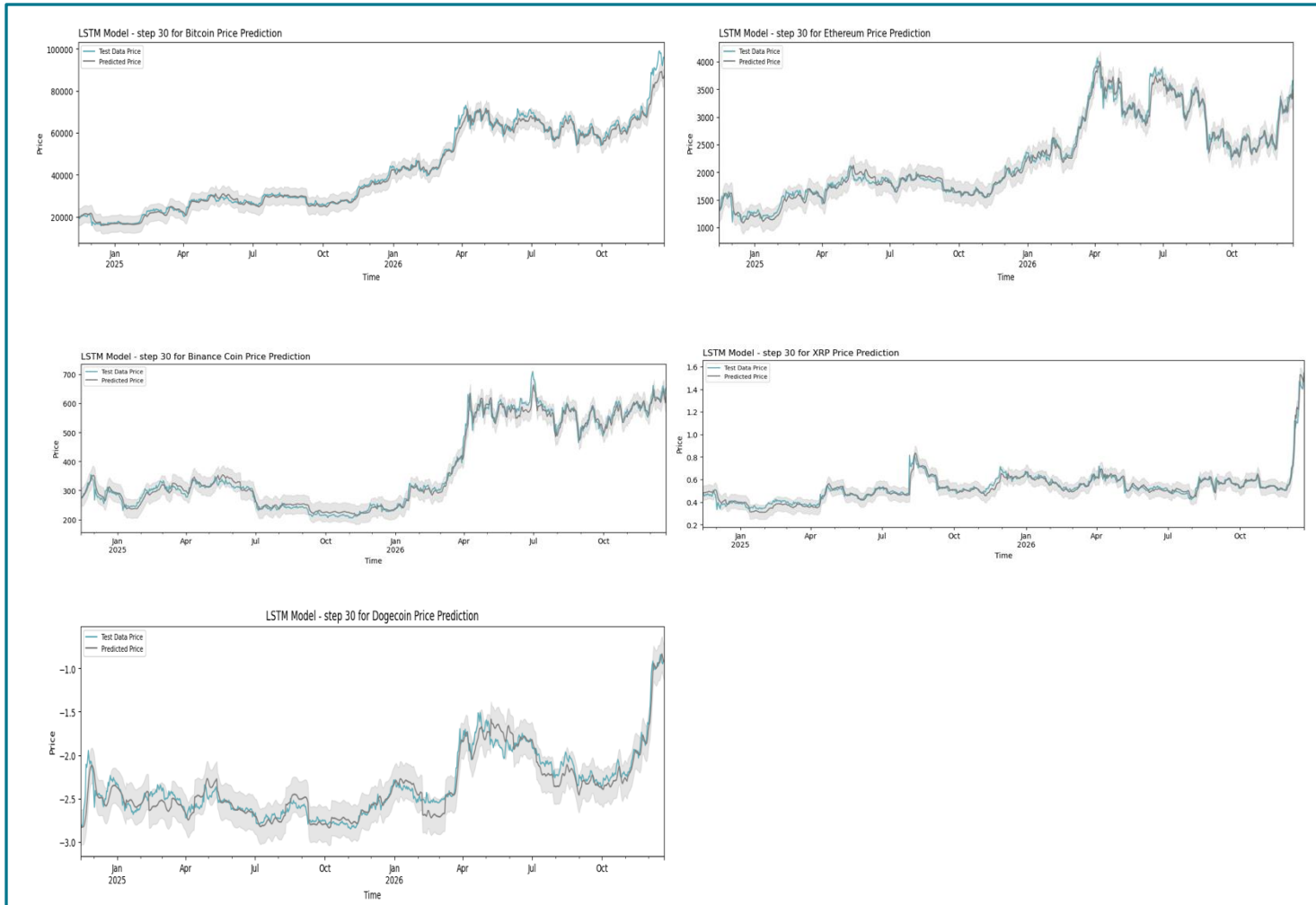


Nhận xét

Độ chính xác của mô hình dự báo cao. Mô hình dự báo tốt trong cả ngắn hạn và dài hạn.
Độ chính xác giảm dần khi thời gian dự báo tăng lên.

Đánh Giá Mô Hình

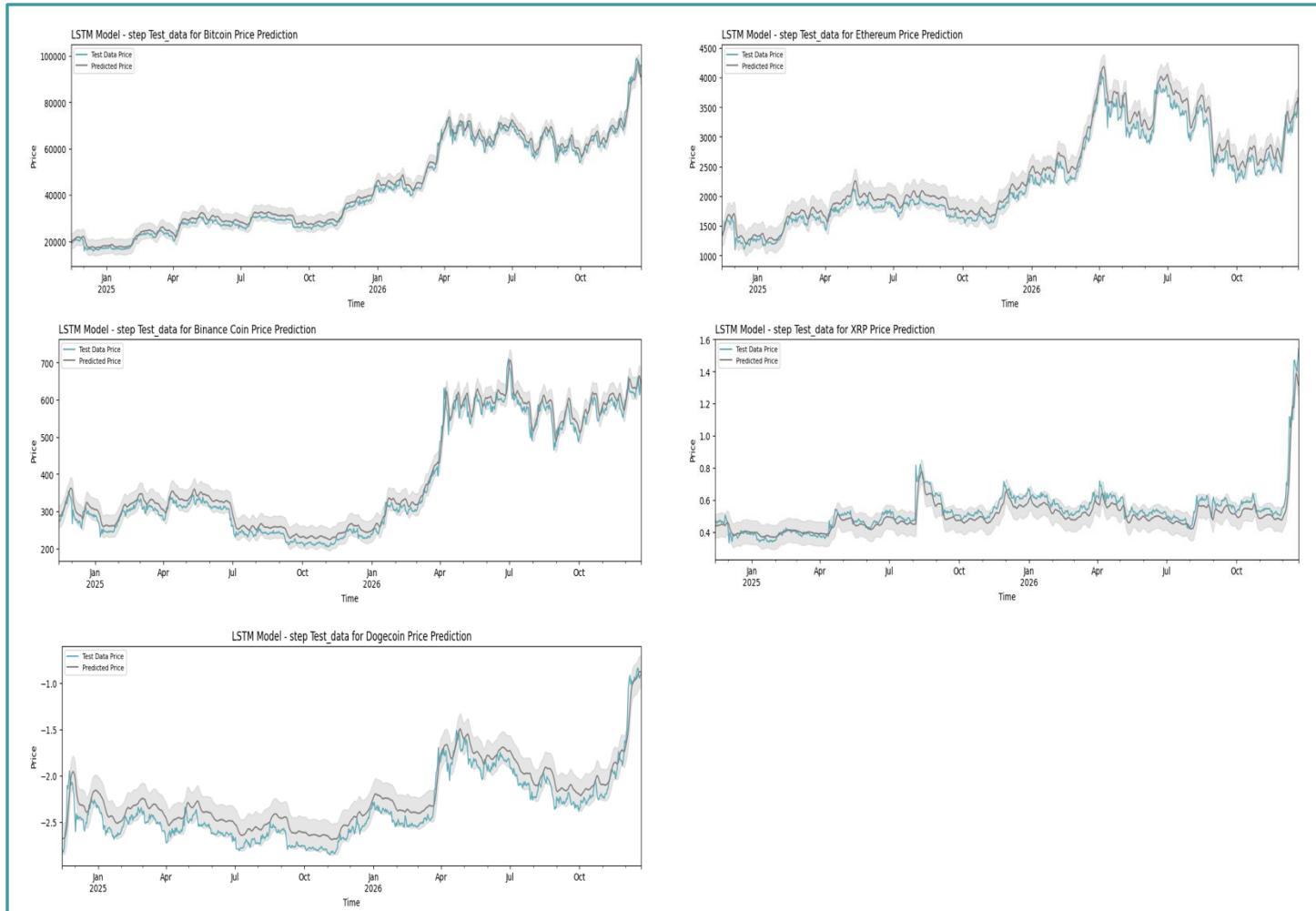
Step 30 Độ chính xác của mô hình dự báo cao, bám sát giá thực tế kể cả với những biến động lớn.



→ Khoảng tin cậy nhỏ và ổn định

Đánh Giá Mô Hình

Test Data Dự báo tốt xu hướng chung và diễn biến chi tiết của giá.

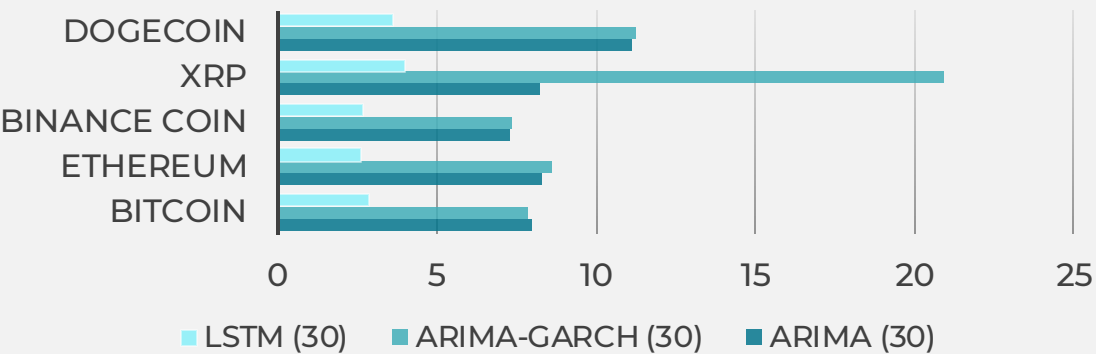


Giá dự báo phản ánh đúng dao động và xu hướng giá thực tế.

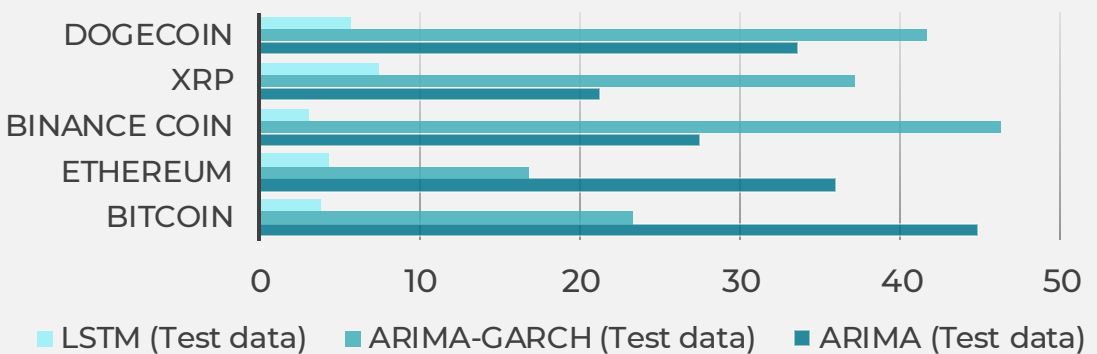
→ Có độ trễ trong dự báo, đặc biệt đối với những giai đoạn biến động lớn (điển hình là Dogecoin)

KẾT LUẬN

Giá trị MAPE với Step=30



Giá trị MAPE trên Test Data



| | ARIMA | ARIMA-GARCH | LSTM |
|-----------------|---|--|---|
| Dự báo ngắn hạn | Dự báo tốt. | Dự báo tốt (ngoại trừ XRP) | Dự báo tốt |
| Dự báo dài hạn | Không nắm bắt được xu hướng chính. Phù hợp khi mức giá có nhiều dao động đột ngột nhưng biên độ dao động nhỏ (XRP, Dogecoin). | Nắm bắt được xu hướng chính Phù hợp khi mức giá tương đối ổn định và có xu hướng tăng giảm rõ ràng (Bitcoin, Ethereum). | Nắm bắt được xu hướng chính và dao động giá chi tiết. |
| Kết luận | Mô hình LSTM là mô hình có performance tốt nhất và được lựa chọn làm mô hình để cải tiến. Mô hình LSTM có thể dự đoán xu hướng chung và các giá trị đỉnh/ đáy của thị trường với độ chính xác cao. Mô hình có thể học những pattern phức tạp và phi tuyến của thị trường tiền mã hoá và đưa ra dự đoán tốt trong ngắn hạn và dài hạn MAPE tại step 30: 2.86% > 2.34% => Mô hình chưa đáp ứng được độ chính xác yêu cầu. Cần cải tiến thêm | | |

III. Cải Tiến Mô Hình

Chuẩn Bị Dữ Liệu

- Trích lọc đặc trưng
- Ma trận tương quan
- Xử lý dữ liệu

Lập Mô Hình

Cấu trúc mô hình

Đánh Giá Mô Hình

- Step = 15
- Step = 30
- Step = 120
- Test Data

1. Chuẩn Bị Dữ Liệu

Trích lọc đặc trưng

Đặc trưng bổ sung là các **chỉ báo kỹ thuật**, được lựa chọn dựa trên ba tiêu chí:

Được tính toán từ dữ liệu giá trong quá khứ

Đã được sử dụng trong các mô hình nghiên cứu trước đây

Có ý nghĩa giải thích cho xu hướng thị trường và những dịch chuyển giá trong ngắn và dài hạn

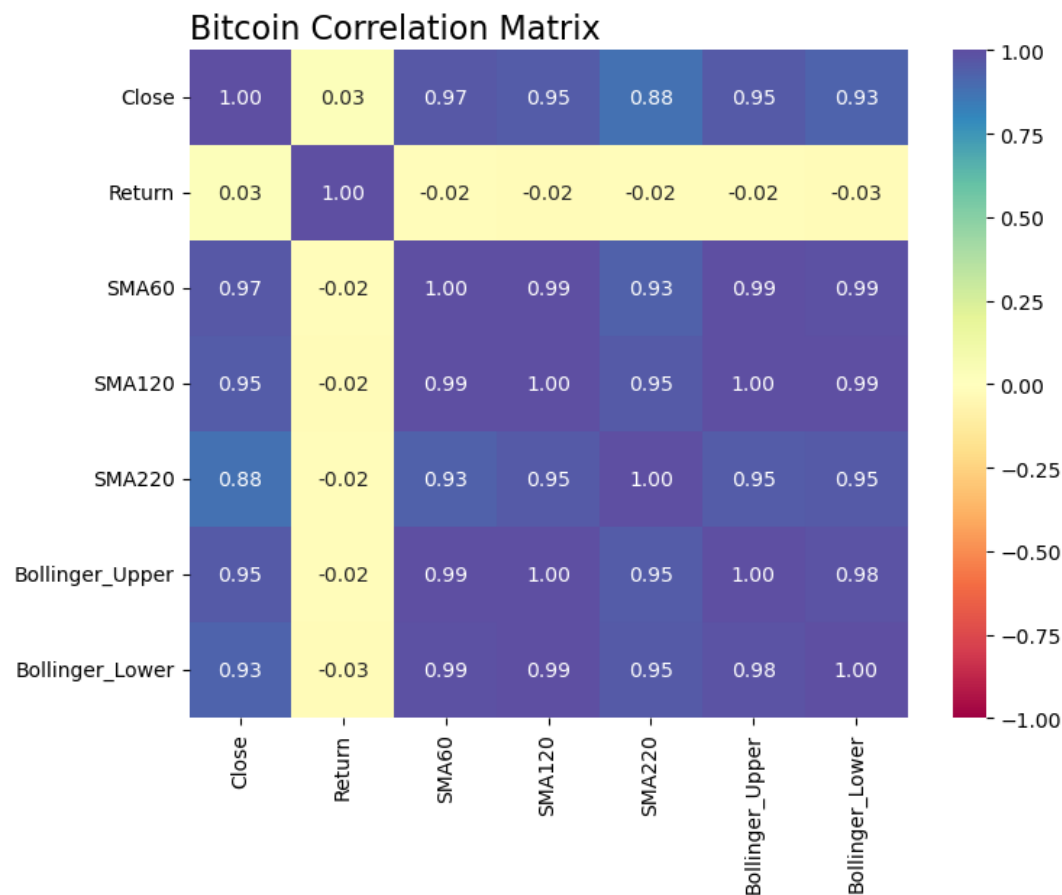
| Đặc trưng | Mô tả |
|-----------------|--|
| Close | Mức giá đóng cửa |
| Return | Phần trăm giá thay đổi so với ngày trước đó |
| SMA60 | Trung bình giá trong 60 ngày |
| SMA120 | Trung bình giá trong 120 ngày |
| SMA220 | Trung bình giá trong 220 ngày |
| Bollinger_Upper | Tính toán dựa trên độ lệch chuẩn của giá. Đại diện cho mức giá cao nhất/thấp nhất mà thị trường có thể đạt đến trong điều kiện biến động bình thường |
| Bollinger_Lower | |

1. Chuẩn Bị Dữ Liệu

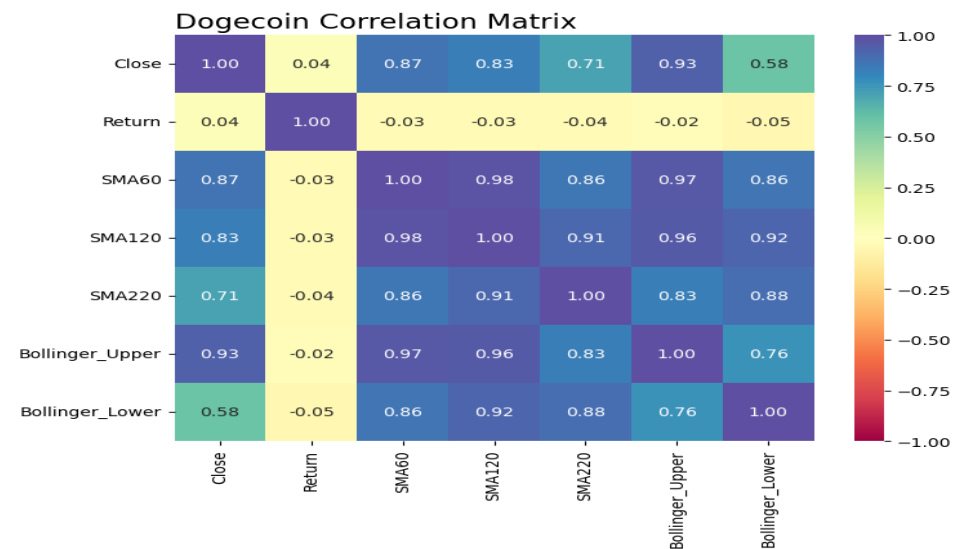
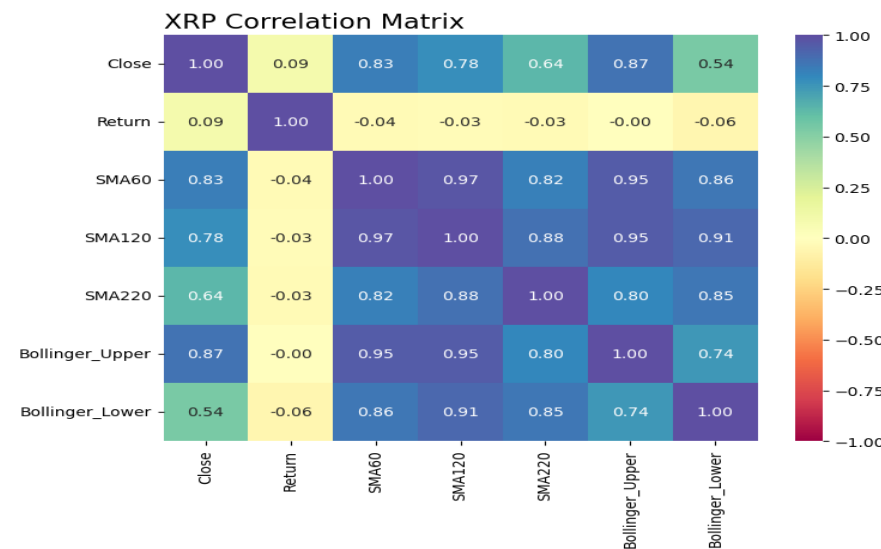
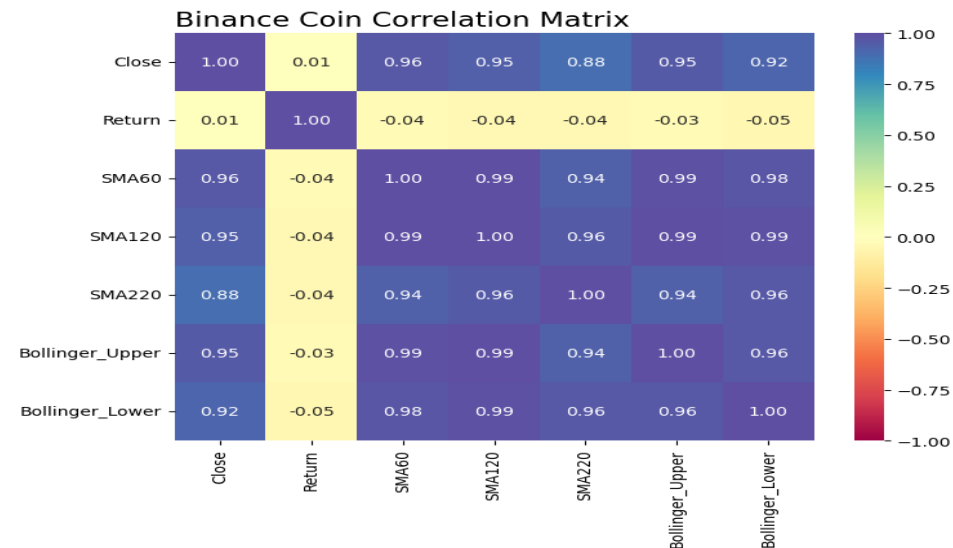
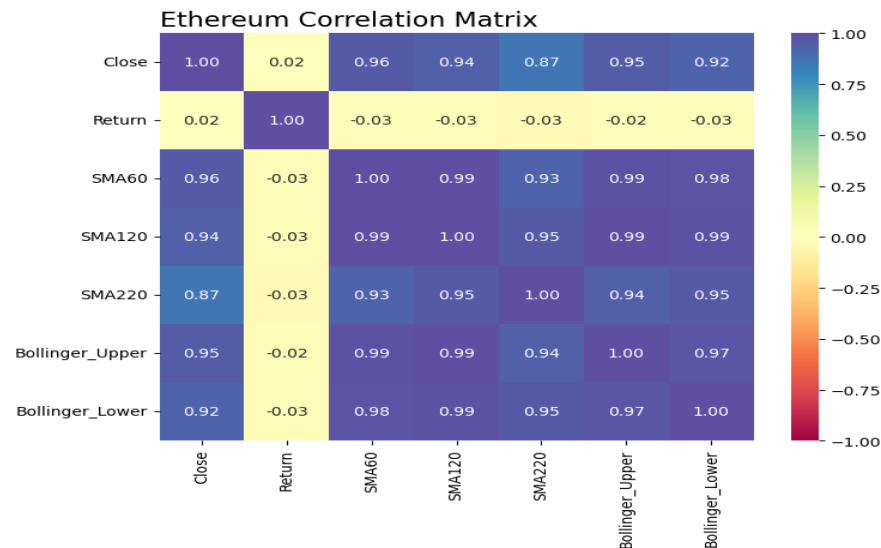
Ma trận tương quan giữa các đặc trưng với giá

SMA60, SMA120, SMA220, Bollinger_Upper, Bollinger_Lower có tương quan cao với nhau và với biến mục tiêu Close

Return có hệ số tương quan thấp, có thể đại diện cho một mối quan hệ phi tuyến tính với biến mục tiêu



1. Chuẩn Bị Dữ Liệu



1. Chuẩn Bị Dữ Liệu

Xử lý dữ liệu

Chuyển đổi chuỗi thời gian thành dạng cửa sổ trượt

Target là giá trị Close (giá đóng cửa)

Feature là giá trị Close của 15 ngày trước đó và **6 features được bổ sung**



Thực hiện chuẩn hoá MinMaxScaler()

Chuyển giá trị dữ liệu về khoảng (0, 1)

Tăng tốc độ hội tụ và tăng khả năng dự báo của mô hình

Riêng với Dogecoin, thực hiện thêm biến đổi logarit trước khi chuẩn hóa



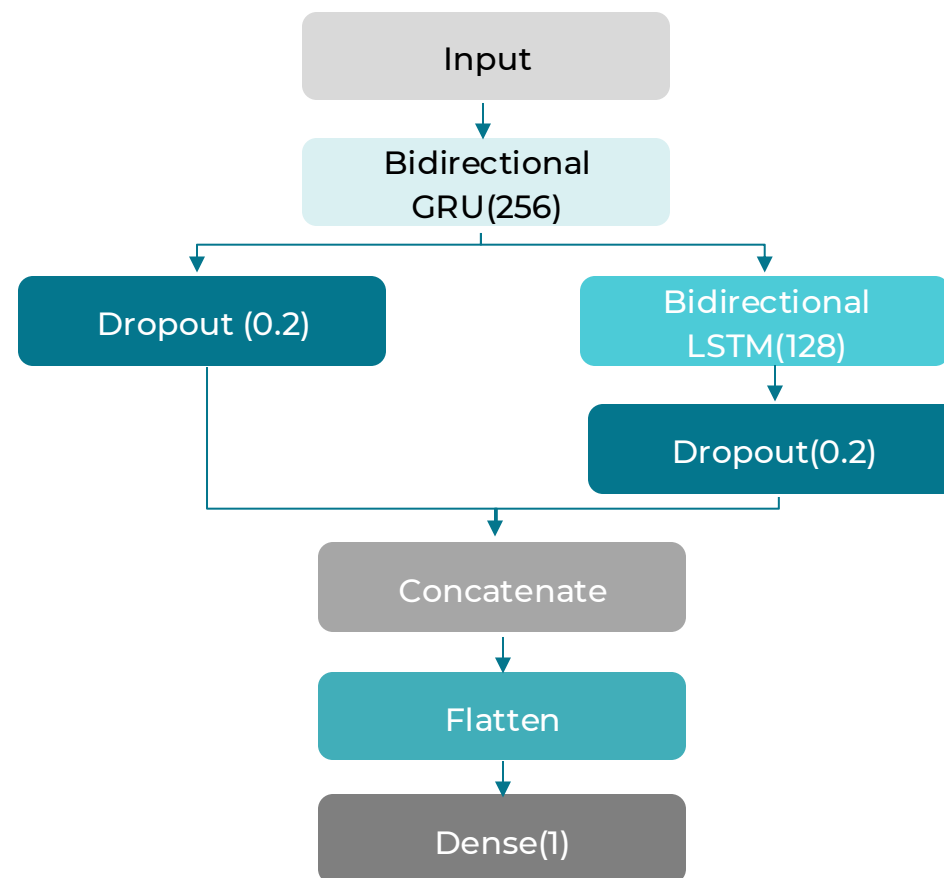
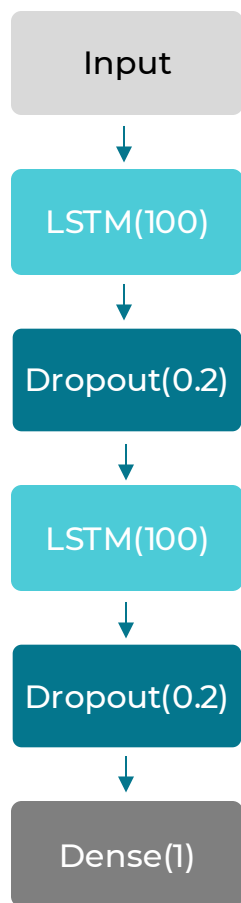
Tách bộ dữ liệu huấn luyện/ kiểm tra

Dữ liệu training: 70% bộ dữ liệu

Dữ liệu testing: 30% bộ dữ liệu

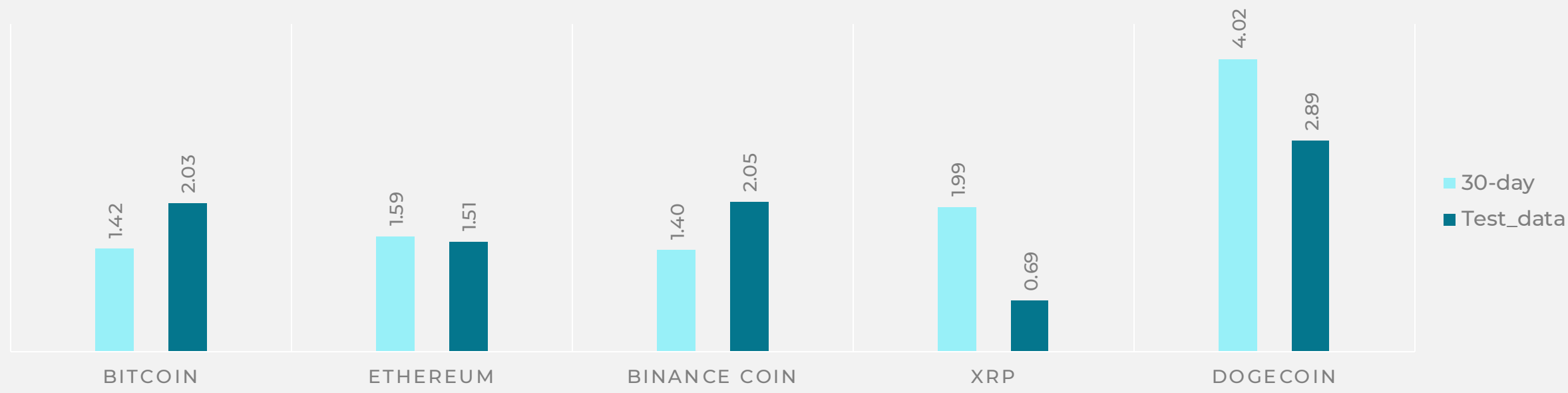
Lập Mô Hình

Cấu trúc mô hình



Đánh Giá Mô Hình

SO SÁNH CHỈ SỐ MAPE CỦA CÁC MÔ HÌNH



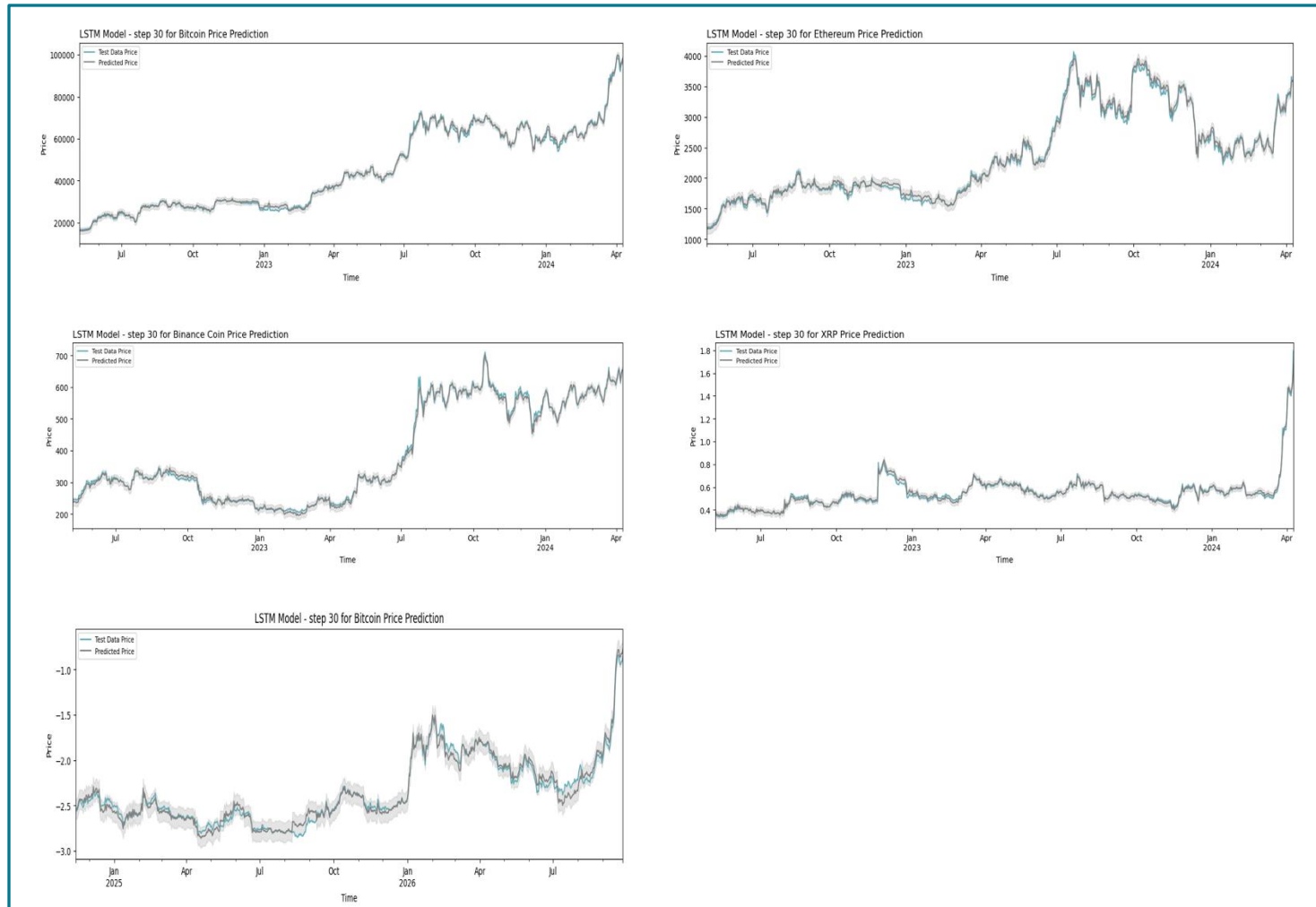
Nhận xét

Mô hình GRU – LSTM đạt được độ chính xác cao khi dự báo giá trong ngắn hạn và dài hạn.

MAPE tại step 30 của Bitcoin là $1.42\% < 2.34\%$ => Mô hình đã đáp ứng được tiêu chí về chất lượng.

Đánh Giá Mô Hình

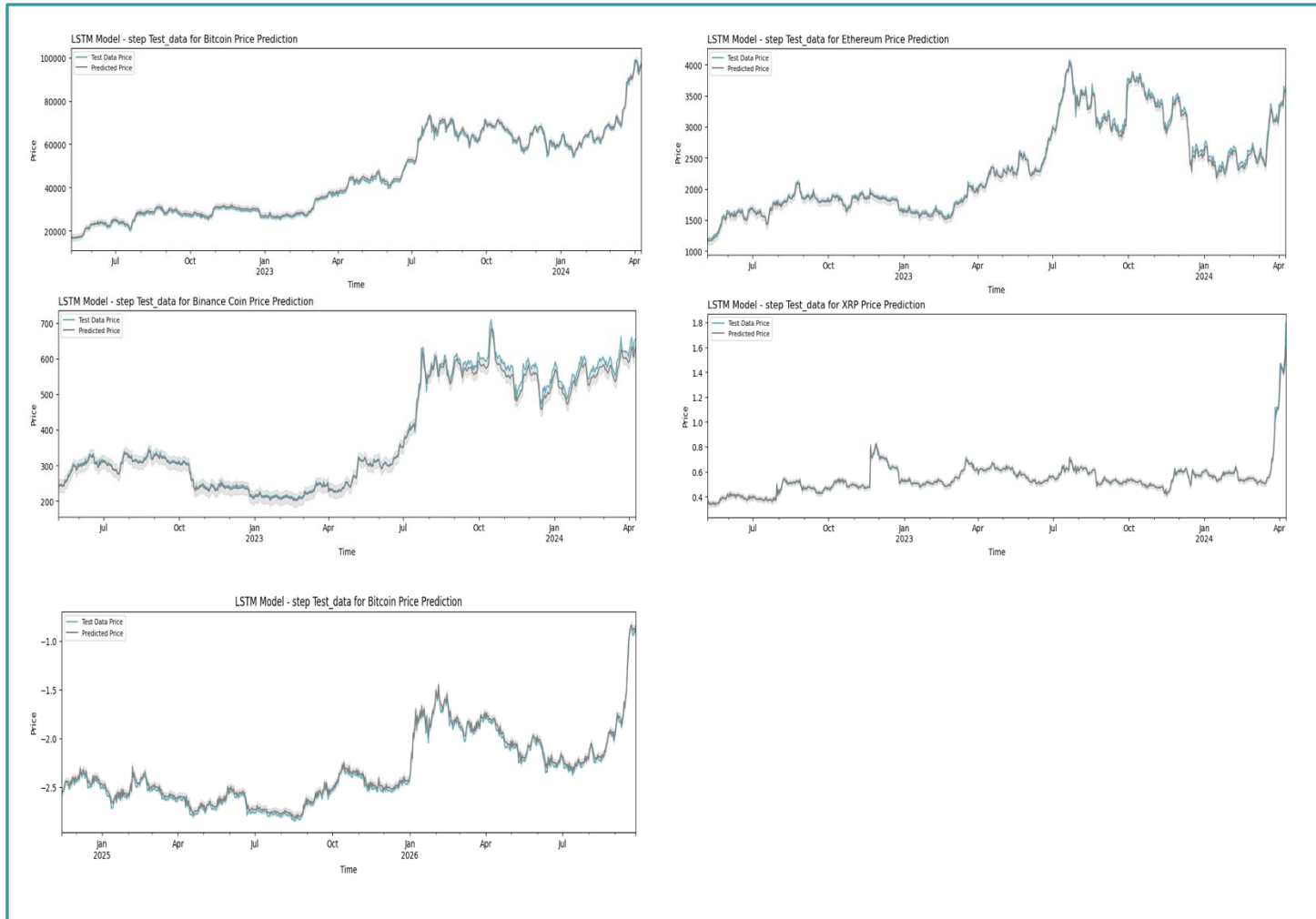
Step 30 Độ chính xác của mô hình dự báo cao, bám sát giá thực tế kể cả với những biến động lớn.



Khoảng tin cậy ổn định và thu hẹp đáng kể so với mô hình LSTM trước đó.

Đánh Giá Mô Hình

Test Data Dự báo tốt xu hướng chung và diễn biến chi tiết của giá.



Giá dự báo phản ánh đúng dao động và xu hướng giá thực tế.

Độ trễ trong dự báo giảm ở cả 5 đồng tiền mã hoá.

Kết Luận

Đánh giá

Mô hình được cải tiến đã đáp ứng được tiêu chí đánh giá đề ra ($MAPE = 1.42\% < 2.34\%$)

Các đặc trưng được chọn

21 đặc trưng bao gồm:

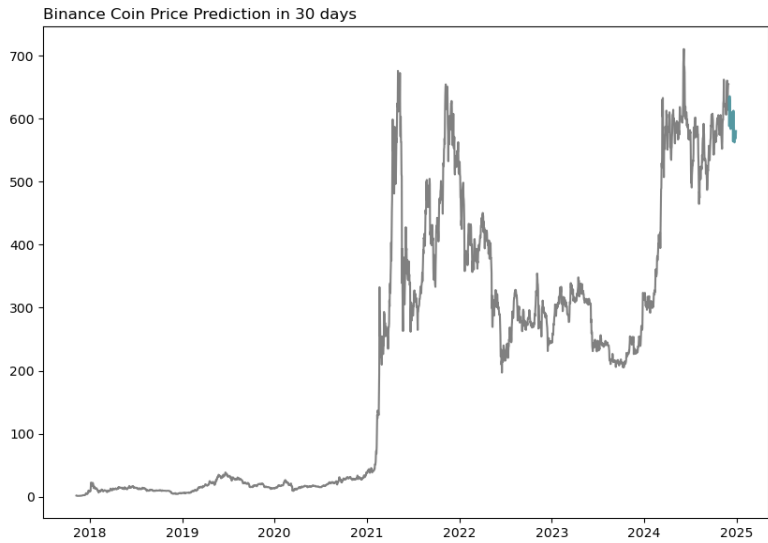
- Mức giá đóng cửa của 15 ngày trước

- 6 features được bổ sung: SMA60, SMA120, SMA220, Bollinger_Upper, Bollinger_Lower và Return

Mô hình được chọn

Mô hình GRU kết hợp với LSTM

Dự Báo Giá Trong Thời Gian 30 Ngày



Mở Rộng Và Định Hướng Trong Tương Lai

Dữ Liệu

Thêm dữ liệu về các yếu tố bên ngoài: Chỉ số kinh tế vĩ mô, chính sách tiền tệ

Thêm dữ liệu về phân tích cảm xúc (Sentiment Analysis)

Mô Hình

Kết hợp mô hình ARIMA – LSTM – GRU

Áp dụng mô hình khác (PELT, Transformer, ...)

Xây dựng mô hình riêng cho từng đồng tiền số

Phạm vi dự báo

Chia nhỏ dữ liệu thành nhiều thời kỳ khác nhau để dự báo

Dự báo cho quãng thời gian dài hơn

Tài Liệu Tham Khảo

1. Comparative study of Bitcoin price prediction using WaveNets, Recurrent Neural Networks and other Machine Learning Methods - Leonardo Felizardo; Roberth Oliveira; Emilio Del-Moral-Hernandez; Fabio Cozman - 2019
2. Comparison of ARIMA Time Series Model and LSTM Deep Learning Algorithm for Bitcoin Price Forecasting - Karakoyun, E. Ş. & Çıbıkdiken, A. O. - 2018
3. Bitcoin Price Trend Prediction Using Deep Neural Network - Hashem Fekry Nematallah1 , Ahmed Ahmed Hesham Sedky2 , Khaled Mohamed Mahar - 2022
4. The Future of Bitcoin Price Predictions Integrating Deep Learning and the Hybrid Model Method - Belalova Guzalxon*, Mannanova Shakhida and Karimov Botir. - 2023



XIN CẢM ƠN!