

**TRƯỜNG CAO ĐẲNG CÔNG NGHỆ THỦ ĐỨC**  
**KHOA CÔNG NGHỆ THÔNG TIN**



**LẬP TRÌNH PYTHON NÂNG CAO**  
**ĐỀ TÀI: DỰ ĐOÁN VÀ PHÂN TÍCH GIÁ BITCOIN**

**GVHD: PHAN THỊ THỂ**

**CÁC THÀNH VIÊN TRONG NHÓM 2:**

- 1. NGUYỄN ĐỨC SÁNG (NHÓM TRƯỞNG)**
- 2. HOÀNG VĂN DŨNG**



*Tp. Hồ Chí Minh, Tháng 11/2024*

## MỤC LỤC

<b>Chương 1: Dữ liệu .....</b>	<b>1</b>
<b>1.1. Nguồn gốc dữ liệu.....</b>	<b>1</b>
<b>1.2. Các biến số (Features).....</b>	<b>1</b>
<b>1.3. Xử lý dữ liệu.....</b>	<b>2</b>
<b>1.4 Kết quả sau khi xử lý dữ liệu .....</b>	<b>4</b>
<b>Chương 2: Trực quan hóa dữ liệu .....</b>	<b>5</b>
<b>2.1. Mục tiêu trực quan hóa dữ liệu .....</b>	<b>5</b>
<b>2.2. Các loại biểu đồ đã sử dụng .....</b>	<b>5</b>
<b>Chương 3: Mô hình hóa dữ liệu .....</b>	<b>10</b>
<b>3.1. Giới thiệu về mô hình hóa dữ liệu .....</b>	<b>10</b>
<b>3.2. Lựa chọn mô hình hóa dữ liệu .....</b>	<b>10</b>
<b>3.3. Chuẩn bị dữ liệu .....</b>	<b>10</b>
<b>3.4. Xây dựng và huấn luyện mô hình.....</b>	<b>10</b>
<b>3.5. Đánh giá mô hình .....</b>	<b>10</b>
<b>Chương 4: Thực nghiệm, kết quả và thảo luận .....</b>	<b>12</b>
<b>4.1. Các bước thực hiện thực nghiệm .....</b>	<b>12</b>
<b>4.2. Các chỉ số đánh giá.....</b>	<b>13</b>
<b>4.3. Kết quả .....</b>	<b>14</b>
<b>4.4. Kết Luận .....</b>	<b>15</b>
<b>4.5. Vẽ Biểu Đồ .....</b>	<b>15</b>
<b>4.5.1. Biểu đồ sánh giá thực tế và dự đoán (Random Forest).....</b>	<b>15</b>
<b>4.5.2. Biểu đồ hộp cho giá và theo năm .....</b>	<b>16</b>
<b>4.5.3. Biểu đồ hồi quy giữa giá và CPI.....</b>	<b>18</b>
<b>4.5.4. Biểu đồ giá và lãi xuất DFF theo thời gian .....</b>	<b>19</b>
<b>4.5.5. Biểu đồ nhiệt mối tương quan của giá, CPI và DFF .....</b>	<b>20</b>
<b>Chương 5: Kết luận .....</b>	<b>22</b>

## MỤC LỤC HÌNH

Hình 2: Loại bỏ dấu % ra khỏi dữ liệu của cột Changes %.....	2
Hình 3: Xử lý cột Vol. để chuyển đổi “K”, “M” và “B” thành số.....	2
Hình 4: Chuyển đổi các cột Price, High, Open và Low sang kiểu số.....	2
Hình 5: Chuyển đổi cột Date sang kiểu datetime và sắp xếp .....	3
Hình 6: Kết hợp dữ liệu .....	3
Hình 7: Các tính đặc trưng bổ sung.....	3
Hình 8: Tạo các cột mới dựa trên cột Date .....	3
Hình 9: Loại bỏ các hàng có giá trị NaN .....	4
Hình 10: Tính trung bình giá theo năm.....	4
Hình 11: Data trước khi chưa xử lý .....	4
Hình 12: Data sau khi xử lý .....	4
Hình 13: Code chia dữ liệu thành tập huấn và kiểm tra giá với CPI.....	12
Hình 14: Code chia dữ liệu thành tập huấn và kiểm tra giá với DFF .....	12
Hình 15: Code chia dữ liệu thành tập huấn và kiểm tra giá với CPI và DFF .....	12
Hình 16: Code huấn luyện và đánh giá mô hình .....	13
Hình 17: Kết Quả huấn luyện và kiểm tra giá với CPI.....	14
Hình 18: Kết quả huấn luyện và kiểm tra giá với lãi suất của FED .....	14
Hình 19: Kết quả huấn luyện và kiểm tra giá với cpi và lãi suất của Fed .....	14
Hình 20: Code vẽ biểu đồ so sánh giá thực tế với giá dự đoán (Random Forest) .....	15
Hình 21: Biểu đồ giá thực tế và dự đoán (Random Forest).....	16
Hình 22: Code vẽ biểu đồ hộp cho giá theo năm .....	16
Hình 23: Biểu đồ phân giá theo năm .....	17
Hình 24: Code vẽ biểu đồ hồi quy giữ giá và CPI.....	18
Hình 25: Biểu đồ hồi quy giữa giá và CPI.....	18
Hình 26: Code biểu đồ giá và lãi suất DFF theo thời gian .....	19
Hình 27: Biểu đồ Giá và lãi suất DFF theo thời gian .....	19
Hình 28: Code của biểu đồ nhiệt mối tương quan của giá, CPI và DFF .....	20
Hình 29: Biểu đồ nhiệt mối tương quan của giá, CPI và DFF .....	21

## ĐÓNG GÓP

Stt	Mã sinh viên	Họ tên	Công việc	Mức độ đóng góp
1	22211TT0690	Nguyễn Đức Sáng (Nhóm trưởng)	<ul style="list-style-type: none"><li>– Tìm Kiếm các Data khác nhau</li><li>– Làm sạch dữ liệu</li><li>– Thực nghiệm</li><li>– Làm Powerpoint</li><li>– Viết Báo Cáo</li></ul>	60%
2	21211TT4642	Hoàng Văn Dũng	<ul style="list-style-type: none"><li>– Làm sạch dữ liệu</li><li>– Trực quan hóa dữ liệu</li><li>– Mô Hình Hóa Dữ Liệu</li><li>– Thực nghiệm</li></ul>	40%

DANH MỤC TỪ VIẾT TẮT

Kí hiệu chữ viết tắt	Chữ viết đầy đủ
CPI	Consumer Price Index
FED	Federal Reserve
DFF	Effective Federal Funds Rate
NAN	Not a Number
MAE	Mean Absolute Error
MSE	Mean Squared Error
$R^2$	R-Squared
ATH	All-Time High

# DANH MỤC THUẬT NGỮ

Tiếng Anh	Thuật Ngữ Tiếng Việt
Consumer Price Index	Là một chỉ số đo lường sự thay đổi trung bình theo thời gian trong các giá cả mà người tiêu dùng phải trả cho một bộ sản phẩm và dịch vụ tiêu dùng
Federal Reserve	Ngân hàng Dự trữ Liên bang Hoa Kỳ
Effective Federal Funds Rate	Là tỷ lệ lãi suất mà các tổ chức tài chính (như ngân hàng) trao đổi tiền gửi lẫn nhau qua các ngân hàng trung ương Hoa Kỳ trong một đêm
All-Time High	Đây là thuật ngữ được sử dụng trong lĩnh vực tài chính và đầu tư để chỉ giá trị cao nhất mà một tài sản (như cổ phiếu, tiền điện tử, hàng hóa,...) đã đạt được kể từ khi nó bắt đầu giao dịch

# MỞ ĐẦU

Trong những năm gần đây, Bitcoin đã nổi lên như một loại tài sản kỹ thuật số quan trọng và đã thu hút sự chú ý của rất nhiều nhà đầu tư, nhà phân tích tài chính, và những người yêu công nghệ trên toàn thế giới. Sự biến động giá của Bitcoin mang đến cả cơ hội và thách thức lớn khi giá Bitcoin vừa phá ATH cũ tại giá 99500 đô vào ngày 23 tháng 11 năm 2024 lại khiến đồng tiền này hấp dẫn hơn, đặc biệt đối với những ai đang tìm kiếm cách dự đoán xu hướng giá để đưa ra các quyết định đầu tư thông minh.

Đề án này nhằm mục đích xây dựng một mô hình dự đoán giá Bitcoin bằng cách sử dụng các mô hình dự đoán khác nhau. Mục tiêu cuối cùng là phát triển một công cụ dự đoán hiệu quả, giúp các nhà đầu tư có thêm dữ liệu và góc nhìn để đưa ra các quyết định chính xác hơn.

Chúng em sẽ bắt đầu bằng việc thu thập và xử lý dữ liệu lịch sử giá Bitcoin, CPI và lãi suất của FED, sau đó xây dựng mô hình khác nhau. Cuối cùng, chúng em sẽ đánh giá hiệu suất của mô hình dựa trên các chỉ số thống kê và trực quan hóa các kết quả dự đoán so với giá trị thực tế.

Đề án này không chỉ giúp chúng em hiểu rõ hơn về cách áp dụng các kỹ thuật học máy đơn giản vào bài toán thực tế mà còn mở ra cánh cửa để khám phá các phương pháp tiên tiến hơn trong việc dự đoán giá trị tài sản kỹ thuật số trong tương lai.

# TÓM TẮT

## **Xây dựng các mô hình thuật toán:**

- Sử dụng các thuật toán khác nhau để thiết lập mối quan hệ giữa các biến độc lập (thời gian, khối lượng giao dịch, CPI, lãi suất Fed) và biến phụ thuộc (giá Bitcoin).
- Đánh giá hiệu suất của mô hình hồi quy tuyến tính thông qua các chỉ số thống kê như độ chính xác, sai số trung bình tuyệt đối (MAE).

## **Thử nghiệm các thuật toán học máy khác:**

- Hồi quy Lasso và Ridge.
- Rừng Ngẫu nhiên (Random Forest).
- Hồi quy tuyến tính (Linear Regression)

## **Đánh giá và so sánh các mô hình:**

- So sánh hiệu suất của các mô hình khác nhau dựa trên các chỉ số đánh giá để tìm ra mô hình phù hợp nhất cho dự đoán giá Bitcoin.
- Trực quan hóa kết quả dự đoán của các mô hình và so sánh với giá trị thực tế để hiểu rõ hơn về ưu điểm và hạn chế của từng mô hình.

## **Phân tích và kết luận:**

- Phân tích kết quả từ các mô hình khác nhau để rút ra kết luận về hiệu quả của từng mô hình trong việc dự đoán giá Bitcoin.
- Đề xuất các phương pháp cải tiến và hướng phát triển tiếp theo cho việc dự đoán giá tài sản kỹ thuật số



## Chương 1: Dữ liệu

### 1.1. Nguồn gốc dữ liệu

- Dữ liệu lịch sử giá của Bitcoin được lấy từ [investing](https://investing.com) từ 18/07/2021 đến 21/11/2024. Investing là một trang web tài chính uy tín và được nhiều người sử dụng. Trang web này cung cấp thông tin thời gian thực về các chỉ số tài chính, lịch sử giao dịch, và các báo cáo kinh tế từ các nguồn đáng tin cậy như Thomson Reuters, FXStreet, và OANDA.
- Dữ liệu lịch sử lãi suất của Fed được lấy từ [fred.stlouisfed.org](https://fred.stlouisfed.org) từ 01/07/1954 đến 21/11/2024.
- Dữ liệu lịch sử chỉ CPI của Hoa Kỳ được lấy từ [fred.stlouisfed.org](https://fred.stlouisfed.org) từ 01/01/1974 đến 01/10/2024
- FRED là một trang web uy tín do Ngân hàng Liên bang St. Louis quản lý. Trang web này cung cấp một bộ dữ liệu lớn về các dữ liệu kinh tế từ nhiều nguồn khác nhau, bao gồm các cơ quan chính phủ như Bộ Ngân hàng Hoa Kỳ, Bộ Censo, và Bộ Thống kê Lao động. FRED đã tồn tại từ năm 1991 và được nhiều nhà nghiên cứu, nhà đầu tư, và các chuyên gia tài chính tin dùng.

### 1.2. Các biến số (Features)

- Giá mở cửa (Open)
- Giá đóng cửa (Price)
- Giá cao nhất (High)
- Giá thấp nhất (Low)
- Giá đóng cửa (Close)
- Ngày dữ liệu được ghi nhận (Date)
- Khối lượng giao dịch (Vol)
- Tỷ lệ CPI
- Lãi suất Fed
- Tỷ lệ thay đổi giá so với ngày hôm trước (Change %)

### 1.3. Xử lý dữ liệu

```
# Xử lý các cột dữ liệu
pd_btc['Change %'] = pd_btc['Change %'].str.replace('%', '').astype(float)
```

Hình 1: Loại bỏ dấu % ra khỏi dữ liệu của cột Changes %

```
# Xử lý cột 'Vol.' để chuyển đổi từ 'K', 'M', và 'B' thành số
def convert_volume(val):
    if isinstance(val, str): # Kiểm tra nếu giá trị là chuỗi
        if 'K' in val:
            return float(val.replace('K', '')) * 1000
        elif 'M' in val:
            return float(val.replace('M', '')) * 1000000
        elif 'B' in val:
            return float(val.replace('B', '')) * 1000000000
    return float(val)
pd_btc['Vol.'] = pd_btc['Vol.'].apply(convert_volume)
```

Hình 2: Xử lý cột Vol. để chuyển đổi “K”, “M” và “B” thành số

```
# Chuyển đổi các cột Price, High, Low và Open sang số
pd_btc['Price'] = pd_btc['Price'].str.replace(',', '').astype(float)
pd_btc['Open'] = pd_btc['Open'].str.replace(',', '').astype(float)
pd_btc['High'] = pd_btc['High'].str.replace(',', '').astype(float)
pd_btc['Low'] = pd_btc['Low'].str.replace(',', '').astype(float)
```

Hình 3: Chuyển đổi các cột Price, High, Open và Low sang kiểu số

```
# Chuyển đổi cột Date sang kiểu datetime
pd_btc['Date'] = pd.to_datetime(pd_btc['Date'])
cpi_data['Date'] = pd.to_datetime(cpi_data['Date'])
fed_rate_data['Date'] = pd.to_datetime(fed_rate_data['Date'])
df = pd_btc.sort_values('Date') # Sắp xếp theo thời gian
```

*Hình 4: Chuyển đổi cột Date sang kiểu datetime và sắp xếp*

```
# Kết hợp dữ liệu
combined_data = pd.merge(pd_btc, cpi_data, on='Date', how='inner')
combined_data = pd.merge(combined_data, fed_rate_data, on='Date', how='inner')
```

*Hình 5: Kết hợp dữ liệu*

```
# Tính các đặc trưng bổ sung
combined_data['Prev_Price'] = combined_data['Price'].shift(1)
combined_data['Price_Change'] = combined_data['Price'] - combined_data['Prev_Price']
combined_data['Year'] = combined_data['Date'].dt.year
combined_data['Month'] = combined_data['Date'].dt.month
combined_data['Day'] = combined_data['Date'].dt.day
combined_data['DayOfWeek'] = combined_data['Date'].dt.dayofweek
```

*Hình 6: Các tính đặc trưng bổ sung*

```
# Tạo các cột mới dựa trên cột Date
df['Year'] = df['Date'].dt.year
df['Month'] = df['Date'].dt.month
df['Day'] = df['Date'].dt.day
df['DayOfWeek'] = df['Date'].dt.dayofweek
```

*Hình 7: Tạo các cột mới dựa trên cột Date*

```
# Loại bỏ các hàng NaN
combined_data.dropna(inplace=True)
```

Hình 8: Loại bỏ các hàng có giá trị NaN

```
# Tính trung bình giá theo năm
df['Year'] = df['Date'].dt.year
yearly_data = df.groupby('Year').agg({'Price': 'mean'}).reset_index()
yearly_data.rename(columns={'Price': 'Average_Price'}, inplace=True)
```

Hình 9: Tính trung bình giá theo năm

## 1.4 Kết quả sau khi xử lý dữ liệu

	Date	Price	Open	High	Low	Vol.	Change %
0	3/24/2024	67,211.90	64,036.50	67,587.80	63,812.90	65.59K	4.96%
1	3/23/2024	64,037.80	63,785.60	65,972.40	63,074.90	35.11K	0.40%
2	3/22/2024	63,785.50	65,501.50	66,633.30	62,328.30	72.43K	-2.62%
3	3/21/2024	65,503.80	67,860.00	68,161.70	64,616.10	75.26K	-3.46%
4	3/20/2024	67,854.00	62,046.80	68,029.50	60,850.90	133.53K	9.35%

Hình 10: Data trước khi chưa xử lý

	Date	Price	Open	High	Low	Vol.	Change %
0	2024-03-24	67211.9	64036.5	67587.8	63812.9	65590.0	4.96
1	2024-03-23	64037.8	63785.6	65972.4	63074.9	35110.0	0.40
2	2024-03-22	63785.5	65501.5	66633.3	62328.3	72430.0	-2.62
3	2024-03-21	65503.8	67860.0	68161.7	64616.1	75260.0	-3.46
4	2024-03-20	67854.0	62046.8	68029.5	60850.9	133530.0	9.35

Hình 11: Data sau khi xử lý

## Chương 2: Trực quan hóa dữ liệu

### 2.1. Mục tiêu trực quan hóa dữ liệu

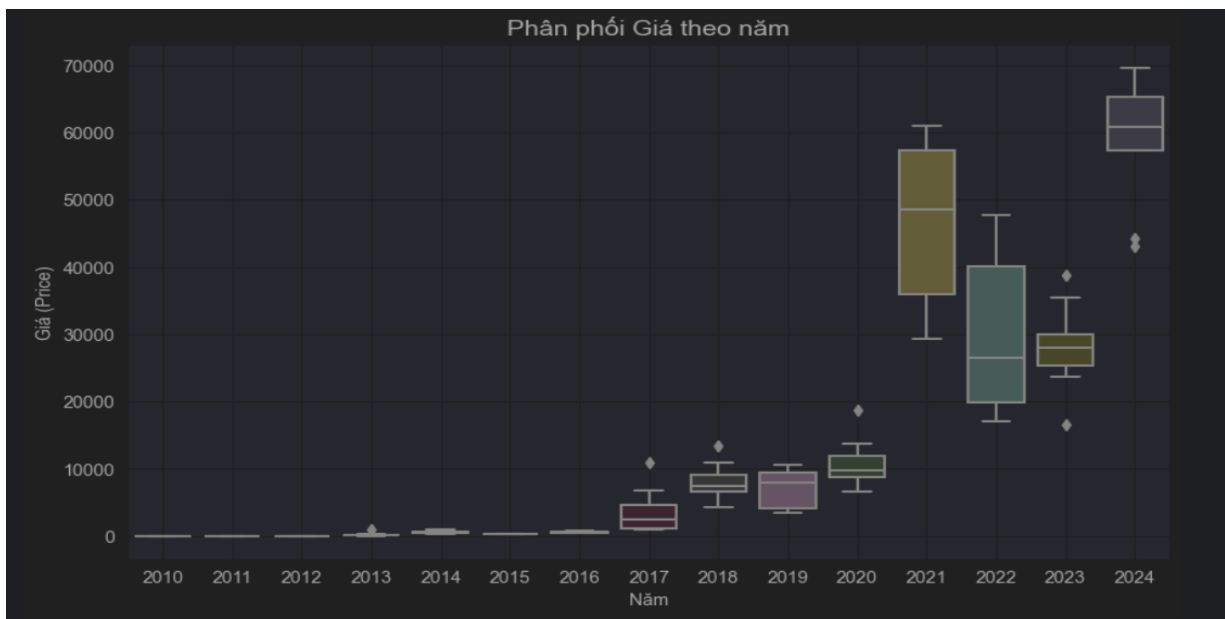
- Mục tiêu của chúng em là hiểu rõ hơn về sự biến động giá Bitcoin theo thời gian và các yếu tố ảnh hưởng đến giá như là CPI và lãi suất của FED.

### 2.2. Các loại biểu đồ đã sử dụng

- Biểu đồ đường (Line Chart): So sánh giá thực tế và giá dự đoán (Random Forest).
- Biểu đồ hộp (Box Plot): Phân phối giá theo năm.
- Biểu đồ đường kép (Dual-axis Line Plot): Giá và lãi suất DFF theo thời gian
- Biểu đồ phân tán (Scatter Plot): Mối quan hệ hồi quy giữa giá và CPI.
- Biểu đồ nhiệt (Heatmap): Mối tương quan của giá, CPI và DFF.

### 2.3. Phân tích sự biến động của giá Bitcoin qua thời gian

- Mục Đích: Sử dụng biểu đồ hộp (Box Plot) cho phép nhận biết sự biến động và xu hướng giá trong từng năm, giúp phân tích và đưa ra các quyết định dựa trên dữ liệu lịch sử.



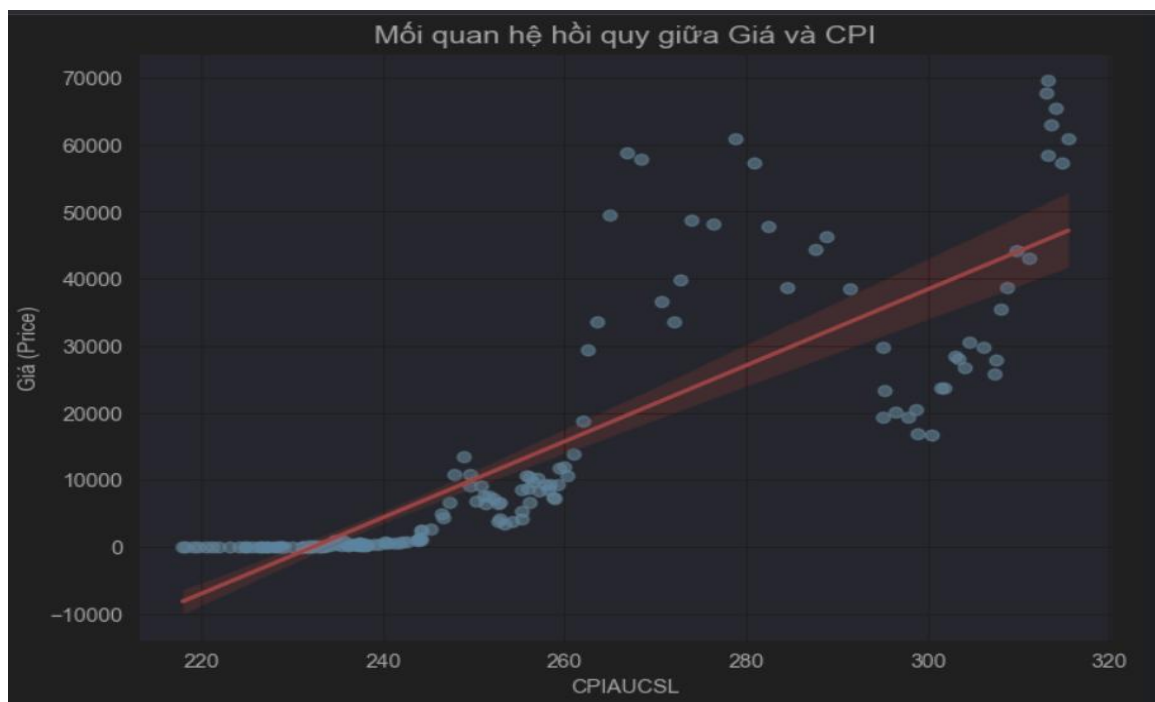
Hình 12: Biểu đồ hộp (Box Plot) để hiển thị giá Bitcoin theo thời gian

- Nhận xét: Xu hướng tăng giá mạnh mẽ từ năm 2021 có thể là do Fed hạ lãi suất cho vay dẫn đến giá cả trong những năm này tăng mạnh.

## **2.4. Mối quan hệ giữa giá Bitcoin và các yếu tố kinh tế vĩ mô**

### **2.4.1. Mối quan hệ giữa giá Bitcoin và CPIAUCSL**

- Mục đích: Sử dụng biểu đồ phân tán (Scatter Plot) để minh họa mối quan hệ giữa CPI và giá Bitcoin. Phân tích xem mối quan hệ có tuyến tính không, có sự tương quan mạnh hay yếu.

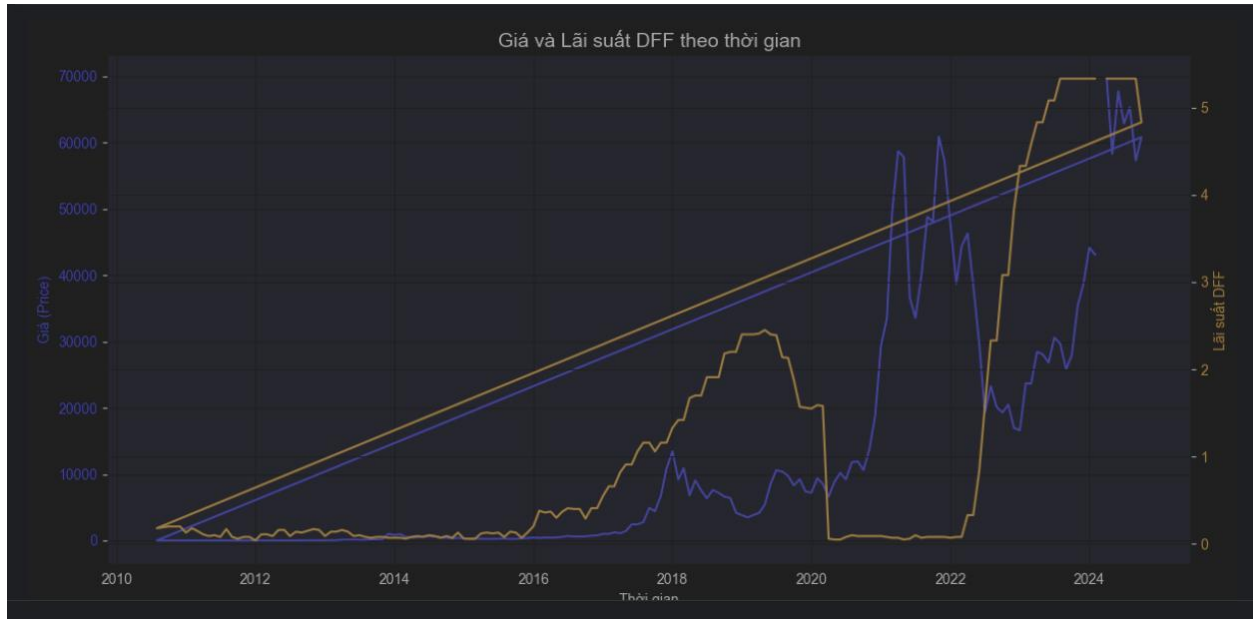


*Hình 13: Biểu đồ mối quan hệ hồi quy giữa giá Bitcoin và CPI*

- Nhận xét: Biểu đồ cho thấy mối quan hệ giữa Giá và CPIAUCSL là tuyến tính, được minh họa rõ ràng bởi đường hồi quy và đang trong xu hướng tăng.

### 2.4.2. Mối quan hệ giữa giá Bitcoin và DFF

- Mục đích: Sử dụng biểu đồ đường kép (Dual-axis Line Plot) để phân tích mối quan hệ giữa giá Bitcoin và DFF và xem thử mối quan hệ này có ngược chiều nhau không.

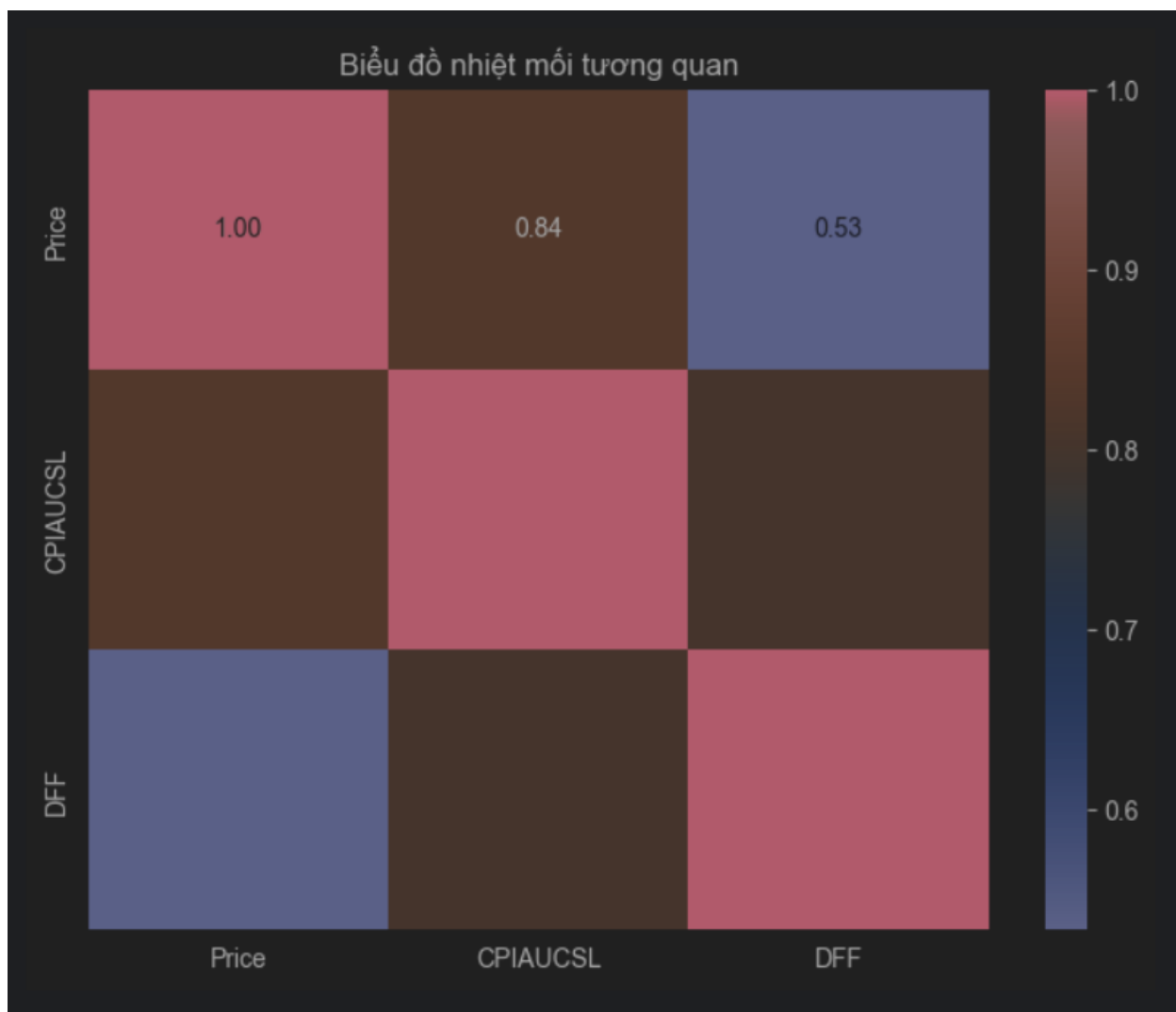


Hình 14: Biểu đồ giá và lãi suất DFF theo thời gian

- Nhận xét: Biểu đồ cho thấy mối quan hệ giữa DFF là ngược chiều nhau, khi DFF tăng thì giá giảm và ngược lại

### 2.4.3. Ma trận tương quan giữa các biến

- Mục đích: Sử dụng heatmap để trực quan hóa tương quan giữa các biến quan trọng: Price, CPIAUCSL, DFF, xem thử biến nào có tương quan mạnh với giá, biến nào ít bị ảnh hưởng



Hình 15: Biểu đồ nhiệt mối tương quan giữa các giá trị Price, DFF, CPI

- Nhận xét: Giá và CPIAUCSL có mối tương quan dương rất mạnh: Hệ số tương quan gần bằng 0.84 cho thấy khi chỉ số giá tiêu dùng (CPIAUCSL) tăng thì giá cả cũng có xu hướng tăng theo. CPIAUCSL và DFF có mối tương quan dương: Hệ số tương quan bằng 0.84 cho thấy mối quan hệ rất mạnh giữa chỉ số giá tiêu dùng và lãi suất quỹ liên bang

## 2.5. Kết Luận

- Giá Bitcoin có biến động lớn theo thời gian, với giai đoạn tăng mạnh từ 2017 đến 2021.



- CPIAUCSL và DFF đều có mối quan hệ rõ ràng với giá Bitcoin, lần lượt là mối quan hệ thuận và nghịch.

## Chương 3: Mô hình hóa dữ liệu

### 3.1. Giới thiệu về mô hình hóa dữ liệu

- Mô hình hóa dữ liệu là bước quan trọng trong việc dự đoán giá Bitcoin, cho phép chúng em nắm bắt các mối quan hệ phức tạp giữa các biến số và đưa ra các dự đoán chính xác về giá Bitcoin dựa trên các yếu tố kinh tế vĩ mô.

### 3.2. Lựa chọn mô hình hóa dữ liệu

- Hồi quy tuyến tính: Đơn giản, phù hợp cho mối quan hệ tuyến tính.
- Lasso và Ridge: Kiểm soát overfitting, lựa chọn đặc trưng quan trọng.
- Random Forest: Mô hình mạnh mẽ cho dữ liệu phi tuyến

### 3.3. Chuẩn bị dữ liệu

- Dữ liệu được chia thành tập huấn luyện (80%) và tập kiểm tra (20%) để đảm bảo tính nhất quán và hiệu quả của các thuật toán
- Chia dữ liệu tập huấn và kiểm tra giá với CPI
- Chia dữ liệu tập huấn và kiểm tra giá với DFF
- Chia dữ liệu tập huấn và kiểm tra giá, CPI và DFF

### 3.4. Xây dựng và huấn luyện mô hình

- Hồi quy tuyến tính (Linear Regression): Chúng em sử dụng thư viện scikit-learn để xây dựng và huấn luyện mô hình hồi quy tuyến tính.
- Hồi quy Lasso và Ridge: Các mô hình hồi quy tuyến tính có điều chỉnh để giảm thiểu overfitting.
- Rừng ngẫu nhiên (Random Forest): Mô hình học máy mạnh mẽ dựa trên nhiều cây quyết định

### 3.5. Đánh giá mô hình

- Chúng em đánh giá mô hình bằng các sai số trung bình tuyệt đối (MAE), sai số trung bình bình phương (MSE), và hệ số xác định ( $R^2$ ).



## Chương 4: Thực nghiệm, kết quả và thảo luận

### 4.1. Các bước thực hiện thực nghiệm

- Dữ liệu được chia thành tập huấn luyện (80%) và tập kiểm tra (20%) và sử dụng các mô hình sau đây để dự đoán giá Bitcoin: Hồi quy tuyến tính, Hồi quy Lasso và Ridge, Rừng ngẫu nhiên.

```
# Chia dữ liệu thành tập huấn luyện và kiểm tra giá với cpi
X_train, X_test, y_train, y_test = train_test_split(X_cpi, y, test_size=0.2, random_state=42)
```

Hình 16: Code chia dữ liệu thành tập huấn và kiểm tra giá với CPI

```
# Chia dữ liệu thành tập huấn luyện và kiểm tra giá với DFF
X_train, X_test, y_train, y_test = train_test_split(X_dff, y, test_size=0.2, random_state=42)
```

Hình 17: Code chia dữ liệu thành tập huấn và kiểm tra giá với DFF

```
# Chia dữ liệu thành tập huấn luyện và kiểm tra giá với cpi và lãi suất của Fed
X_train, X_test, y_train, y_test = train_test_split(X_cpi_dff, y, test_size=0.2, random_state=42)
```

Hình 18: Code chia dữ liệu thành tập huấn và kiểm tra giá với CPI và DFF

```
models = {
    "Linear Regression": LinearRegression(),
    "Lasso Regression": Lasso(alpha=0.1),
    "Ridge Regression": Ridge(alpha=1.0),
    "Random Forest": RandomForestRegressor(n_estimators=100, random_state=42)
}

# Huấn luyện và đánh giá các mô hình
results = []
for name, model in models.items():
    # Huấn luyện mô hình
    model.fit(X_train, y_train)
    # Dự đoán trên tập kiểm tra
    y_pred = model.predict(X_test)
    # Đánh giá mô hình
    mse = mean_squared_error(y_test, y_pred)
    rmse = np.sqrt(mse)
    mae = mean_absolute_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)
    # Lưu kết quả
    results.append({
        "Model": name,
        "R2": r2,
        "MAE": mae,
        "RMSE": rmse
    })
```

Hình 19: Code huấn luyện và đánh giá mô hình

## 4.2. Các chỉ số đánh giá

- Sai số trung bình tuyệt đối (MAE)
- Sai số trung bình bình phương (MSE)
- Hệ số xác định ( $R^2$ )

### 4.3. Kết quả

Giá với CPI				
	Model	$R^2$	MAE	RMSE
0	Linear Regression	0.681500	7785.770529	10036.808803
1	Lasso Regression	0.681500	7785.770220	10036.808182
2	Ridge Regression	0.681501	7785.757384	10036.782386
3	Random Forest	0.962318	1967.028324	3452.301139

Hình 20: Kết Quả huấn luyện và kiểm tra giá với CPI

Giá với lãi suất của FED				
	Model	$R^2$	MAE	RMSE
0	Linear Regression	0.292387	10929.904170	14960.244668
1	Lasso Regression	0.292386	10929.904244	14960.247001
2	Ridge Regression	0.292291	10929.935233	14961.252689
3	Random Forest	0.345892	10006.167619	14383.526114

Hình 21: Kết quả huấn luyện và kiểm tra giá với lãi suất của FED

Giá với lãi suất của FED và CPI				
	Model	$R^2$	MAE	RMSE
0	Linear Regression	0.672692	7519.634518	10174.637287
1	Lasso Regression	0.672694	7519.618489	10174.605058
2	Ridge Regression	0.673373	7514.339807	10164.051010
3	Random Forest	0.962398	2038.932882	3448.617138

Hình 22: Kết quả huấn luyện và kiểm tra giá với cpi và lãi suất của Fed

## 4.4. Kết Luận

- Mô hình Random Forest tỏ ra vượt trội: Trong cả 3 trường hợp (dự đoán giá dựa trên CPI, DFF và cả CPI lẫn DFF), mô hình Random Forest đều đạt được độ chính xác cao nhất (đánh giá qua R-squared) và các chỉ số sai số (MAE, RMSE) thấp nhất. Điều này cho thấy Random Forest có khả năng bắt giữ các mối quan hệ phức tạp giữa các biến tốt hơn so với các mô hình hồi quy tuyến tính (Linear Regression, Lasso Regression, Ridge Regression).
- CPI có mối liên hệ chặt chẽ hơn với giá so với DFF: Khi chỉ sử dụng CPI để dự đoán giá, mô hình đạt được độ chính xác cao hơn so với khi chỉ sử dụng DFF. Điều này cho thấy CPI có thể là một yếu tố quan trọng ảnh hưởng đến giá.
- Kết hợp cả CPI và DFF cải thiện độ chính xác: Khi sử dụng cả CPI và DFF để dự đoán giá, mô hình Random Forest đạt được độ chính xác cao nhất, cho thấy việc kết hợp cả hai yếu tố này có thể giúp cải thiện đáng kể khả năng dự đoán của mô hình.

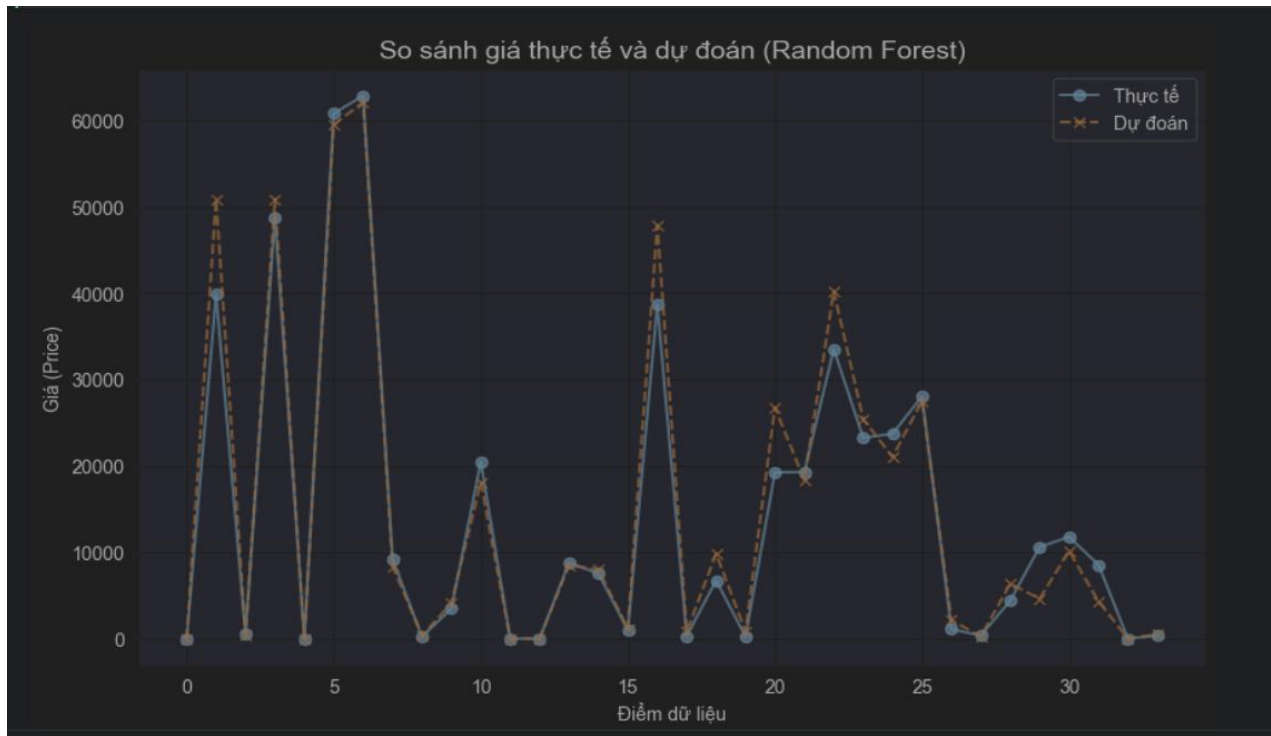
## 4.5. Vẽ Biểu Đồ

### 4.5.1. Biểu đồ sánh giá thực tế và dự đoán (Random Forest)

```
# Dự đoán giá với mô hình Random Forest
rf_model = models["Random Forest"]
y_pred_rf = rf_model.predict(X_test)

# 1. Biểu đồ so sánh giá thực tế và giá dự đoán
plt.figure(figsize=(10, 6))
plt.plot(y_test.values, label="Thực tế", linestyle='-', marker='o', alpha=0.7)
plt.plot(y_pred_rf, label="Dự đoán", linestyle='--', marker='x', alpha=0.7)
plt.title("So sánh giá thực tế và dự đoán (Random Forest)", fontsize=14)
plt.xlabel("Điểm dữ liệu")
plt.ylabel("Giá (Price)")
plt.legend()
plt.grid(True)
plt.show()
```

Hình 23: Code vẽ biểu đồ so sánh giá thực tế với giá dự đoán (Random Forest)



Hình 24: Biểu đồ giá thực tế và dự đoán (Random Forest)

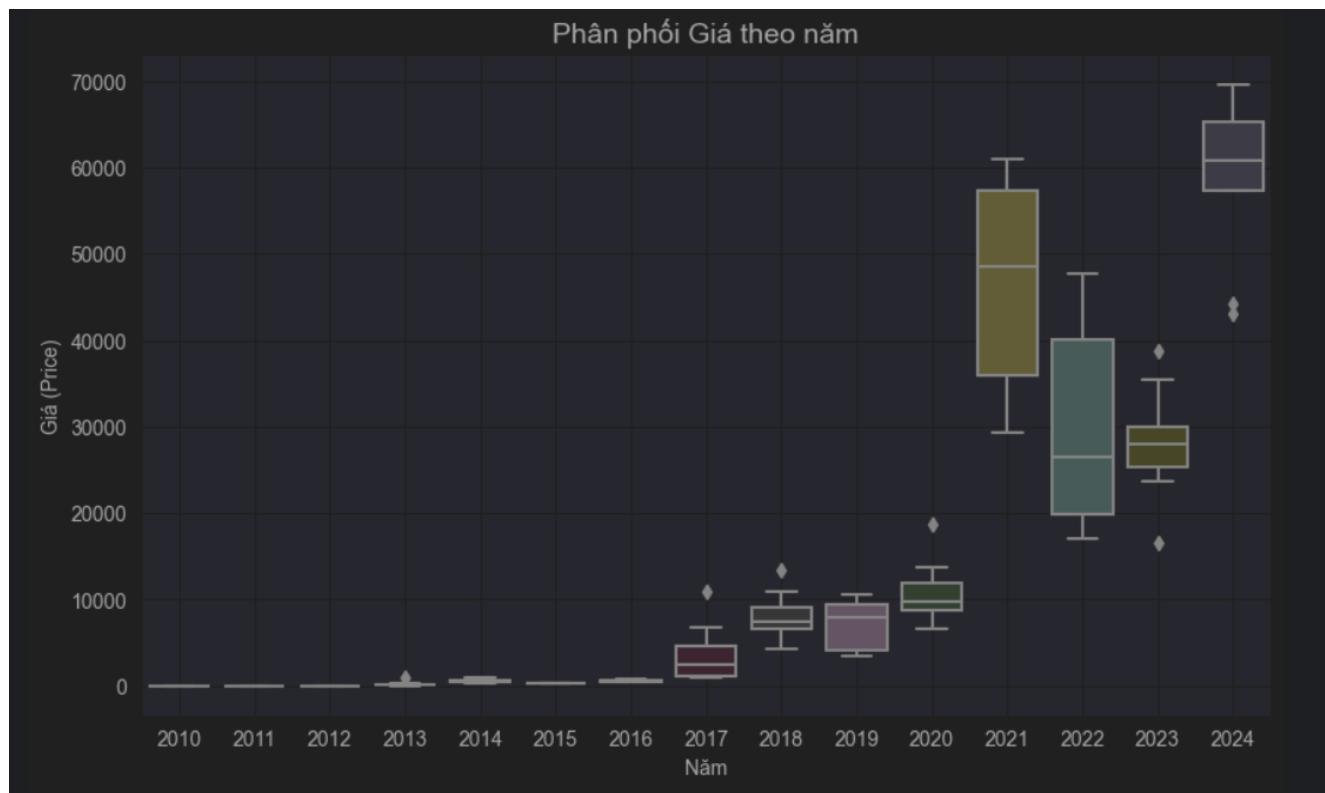
**Nhận xét biểu đồ:** Nhìn chung, đường biểu diễn giá dự đoán có xu hướng đi theo đường biểu diễn giá thực tế, cho thấy mô hình đã bắt được khá tốt xu hướng chung của dữ liệu. Tuy nhiên, vẫn có một số điểm dữ liệu mà giá dự đoán lệch khá ít so với giá thực tế.

#### 4.5.2. Biểu đồ hộp cho giá và theo năm

```
# Biểu đồ hộp cho Giá theo năm
plt.figure(figsize=(10, 6))
sns.boxplot(x='Year', y='Price', data=combined_data, palette="Set3")
plt.title("Phân phối Giá theo năm", fontsize=14)
plt.xlabel("Năm")
plt.ylabel("Giá (Price)")
plt.grid(True)
plt.show()
Executed at 2024.11.27 01:50:52 in 306ms
```

Hình 25: Code vẽ biểu đồ hộp cho giá theo năm





Hình 26: Biểu đồ phân giá theo năm

### Nhận xét biểu đồ:

- Xu hướng tăng giá: Nếu trung vị của hầu hết các năm có xu hướng tăng dần, điều đó cho thấy giá cả chung có xu hướng tăng qua các năm.
- Sự biến động: Nếu khoảng biến thiên của các năm khác nhau đáng kể, điều đó cho thấy mức độ biến động của giá cả có thể thay đổi qua các năm.
- Các năm có giá cả đặc biệt: Các năm có hộp cao hoặc thấp hơn so với các năm khác có thể có những đặc điểm riêng biệt về giá cả.
- Sự xuất hiện của giá trị ngoại lệ: Các giá trị ngoại lệ có thể cho thấy những sự kiện đặc biệt ảnh hưởng đến giá cả trong năm đó, chẳng hạn như khủng hoảng kinh tế, biến động thị trường.

### 4.5.3. Biểu đồ hồi quy giữa giá và CPI

```
# Biểu đồ hồi quy giữa Giá và CPI
plt.figure(figsize=(8, 6))
sns.regplot(x='CPIAUCSL', y='Price', data=combined_data, scatter_kws={'alpha':0.6}, line_kws={'color':'red'})
plt.title("Mối quan hệ hồi quy giữa Giá và CPI", fontsize=14)
plt.xlabel("CPIAUCSL")
plt.ylabel("Giá (Price)")
plt.grid(True)
plt.show()
Executed at 2024.11.27 01:16:00 in 303ms
```

Hình 27: Code vẽ biểu đồ hồi quy giữa giá và CPI



Hình 28: Biểu đồ hồi quy giữa giá và CPI

#### Nhận xét biểu đồ:

- Xu hướng chung: Đường hồi quy có độ dốc dương, cho thấy khi chỉ số CPI tăng thì giá cả cũng có xu hướng tăng. Điều này phù hợp với lý thuyết kinh tế, khi lạm phát (được đo bằng CPI) tăng thì giá cả hàng hóa và dịch vụ thường đi lên.
- Độ phân tán của dữ liệu: Các điểm dữ liệu phân tán xung quanh đường hồi quy cho thấy mối quan hệ giữa giá và CPI không hoàn toàn tuyến tính và có thể còn chịu ảnh hưởng của các yếu tố khác.
- Các điểm ngoại lệ: Một số điểm dữ liệu nằm khá xa đường hồi quy, có thể là các điểm ngoại lệ. Những điểm này cần được xem xét kỹ hơn để hiểu rõ nguyên nhân.

#### 4.5.4. Biểu đồ giá và lãi xuất DFF theo thời gian

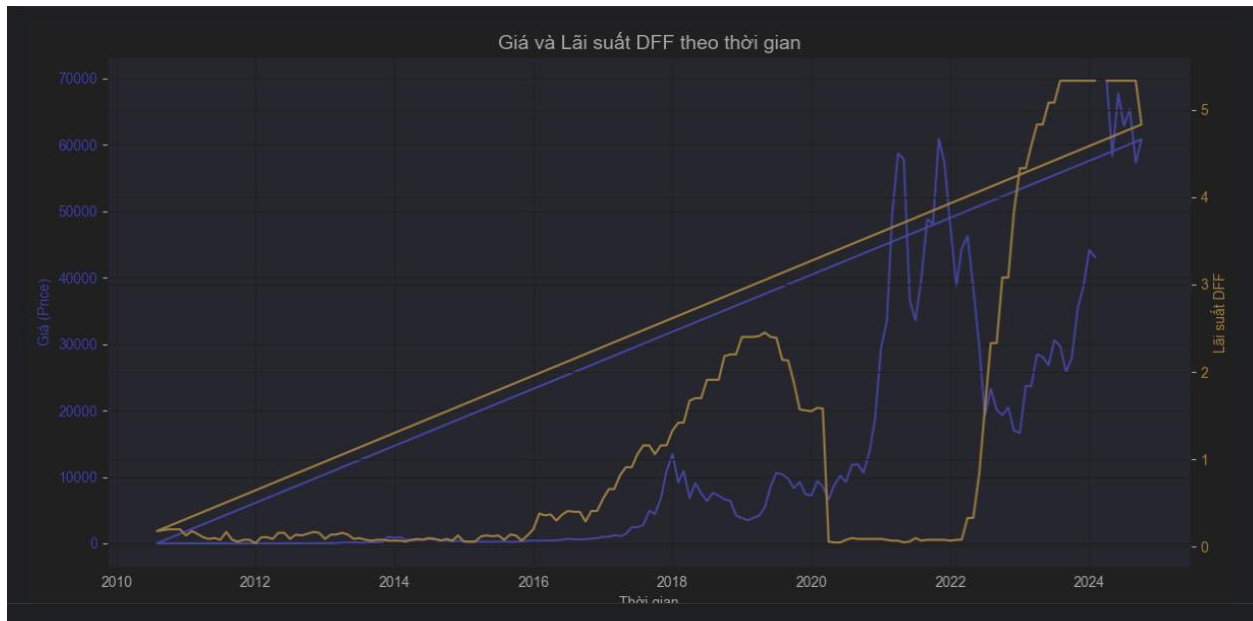
```
fig, ax1 = plt.subplots(figsize=(12, 6))

# Trục y đầu tiên cho Price
ax1.plot(combined_data['Date'], combined_data['Price'], color='blue', label="Price")
ax1.set_xlabel("Thời gian")
ax1.set_ylabel("Giá (Price)", color='blue')
ax1.tick_params(axis='y', labelcolor='blue')

# Trục y thứ hai cho DFF
ax2 = ax1.twinx()
ax2.plot(combined_data['Date'], combined_data['DFF'], color='orange', label="DFF")
ax2.set_ylabel("Lãi suất DFF", color='orange')
ax2.tick_params(axis='y', labelcolor='orange')

plt.title("Giá và Lãi suất DFF theo thời gian", fontsize=14)
fig.tight_layout()
plt.show()
```

Hình 29: Code biểu đồ giá và lãi xuất DFF theo thời gian



Hình 30: Biểu đồ Giá và lãi suất DFF theo thời gian

### **Nhận xét biểu đồ:**

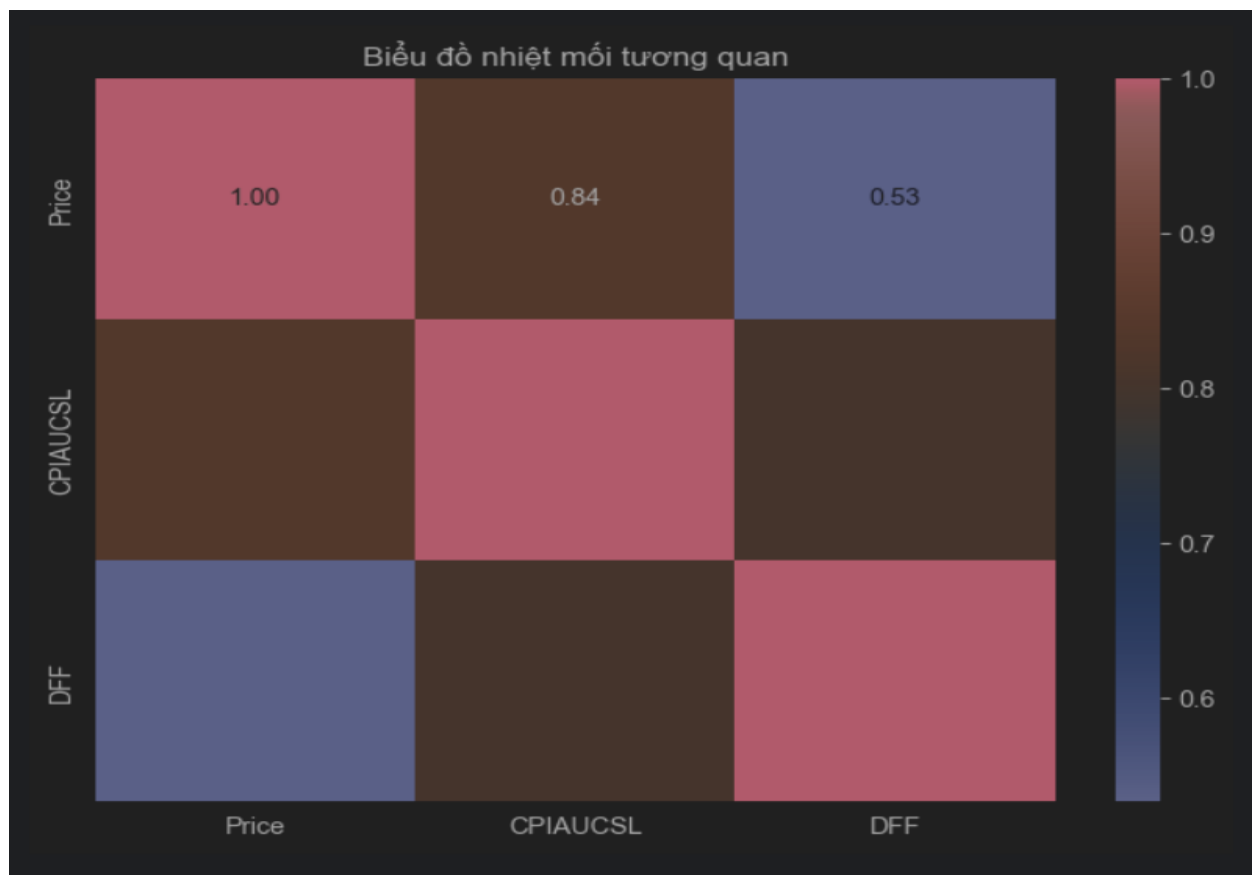
- Trục Y bên trái (Giá - Price): Giá có xu hướng tăng trưởng đáng kể trong giai đoạn từ năm 2010 đến 2021, đạt đỉnh trong khoảng 2021–2022. Sau đỉnh giá, có dấu hiệu giảm vào khoảng giữa năm 2022, sau đó biến động nhẹ. Xu hướng này phản ánh biến động lớn trên thị trường (có thể liên quan đến Bitcoin hoặc các tài sản khác được đo lường trong dữ liệu).
- Trục Y bên phải (Lãi suất DFF): Lãi suất DFF duy trì ở mức thấp từ 2010 đến khoảng 2021, với giá trị gần 0. Sau 2021, lãi suất bắt đầu tăng mạnh, đạt mức cao nhất vào khoảng 2023–2024.
- So sánh xu hướng giữa Giá và Lãi suất DFF: Trong giai đoạn lãi suất thấp (2010–2021), giá tăng mạnh. Điều này có thể gợi ý rằng môi trường lãi suất thấp thúc đẩy sự tăng trưởng giá (điển hình trong các thị trường tài sản rủi ro). Khi lãi suất tăng nhanh sau 2021, giá giảm đáng kể, cho thấy có thể có mối liên hệ nghịch chiều giữa lãi suất và giá Bitcoin trong giai đoạn này.

#### **4.5.5. Biểu đồ nhiệt mối tương quan của giá, CPI và DFF**

```
# Tính ma trận tương quan
correlation_matrix = combined_data[['Price', 'CPIAUCSL', 'DFF']].corr()

# Vẽ heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Biểu đồ nhiệt mối tương quan")
plt.show()
```

*Hình 31: Code của biểu đồ nhiệt mối tương quan của giá, CPI và DFF*



Hình 32: Biểu đồ nhiệt mối tương quan của giá, CPI và DFF

**Nhận xét biểu đồ:**

- Giá và CPIAUCSL có mối tương quan dương rất mạnh: Hệ số tương quan gần bằng 0.84 cho thấy khi chỉ số giá tiêu dùng (CPIAUCSL) tăng thì giá cả cũng có xu hướng tăng theo.
- CPIAUCSL và DFF có mối tương quan dương: Hệ số tương quan bằng 0.84 cho thấy mối quan hệ rất mạnh giữa chỉ số giá tiêu dùng và lãi suất quỹ liên bang. Điều này có thể do chính sách tiền tệ của ngân hàng trung ương: khi lạm phát (CPIAUCSL) tăng, ngân hàng trung ương thường tăng lãi suất (DFF) để kiềm chế lạm phát.

## **Chương 5: Kết luận**

- Mô hình Random Forest tỏ ra vượt trội hơn các mô hình còn lại trong việc dự đoán giá. Qua những thử nghiệm với các biểu đồ trên, ta thấy giá BitCoin phù hợp với nhiều yếu tố như là CPI, và lãi suất của Fed.
- Khi chỉ số CPI tăng thì giá Bitcoin tăng và ngược lại.
- Khi lạm phát (đo bằng CPI) tăng, giá trị đồng tiền mất đi, nhiều nhà đầu tư có xu hướng chuyển sang các tài sản có tính phòng ngừa lạm phát như Bitcoin. Ngoài ra, khi lạm phát cao, các ngân hàng trung ương thường tăng lãi suất, làm giảm sức hấp dẫn của các tài sản truyền thống và đẩy nhà đầu tư tìm đến các kênh đầu tư mới như tiền điện tử nên khiến giá Bitcoin tăng.
- Khi chỉ số lãi suất của Fed tăng thì giá Bitcoin giảm và ngược lại.
- Khi lãi suất DFF tăng, chi phí vay vốn của các nhà đầu tư tăng lên. Điều này có thể khiến một số nhà đầu tư rút vốn khỏi thị trường tiền điện tử để tìm kiếm kênh đầu tư có lợi nhuận ổn định hơn, gây áp lực giảm lên giá Bitcoin
- Giá Bitcoin còn phụ thuộc vào các yếu tố khác như là sự tham lam và sợ hãi của thị trường, tin tức, và nhiều yếu tố khác

Tài liệu tham khảo:

- <https://vietnambiz.vn/phan-tram-thay-doi-percentage-change-la-gi-cach-tinh-20200428234539481.htm>