

# DỰ BÁO PHÁT TRIỂN CỦA XE ĐIỆN TRÊN THẾ GIỚI BẰNG MÔ HÌNH ARIMA

Môn học: Lập trình phân tích dữ liệu 2

GVHD: TS. Lê Trọng Ngọc

Thành viên nhóm:

21028411 – Trương Dương Minh Nhật

21017351 - Trương Quốc Bảo

21024541 - Nguyễn Đức Tài

21080071 - Nguyễn Thanh Thuận

22643391 - Nguyễn Hải Tâm

GIỚI THIỆU


# Lý do chọn đề tài

- Phương tiện di chuyển sử dụng điện ngày càng phổ biến dần
- Từ sự phổ biến, có thể xét sự ảnh hưởng ở các thời điểm và dự báo sự phát triển ở tương lai
- Có thể dự báo doanh thu xe điện và phương tiện nào sẽ là trend trong tương lai sắp tới

# Bộ dữ liệu

## Tổng quan

- Bộ dữ liệu gồm 12654 dòng và 9 cột (bộ dữ liệu có sẵn), bộ dữ liệu gồm 3 biến định lượng và 6 biến định tính.

region	category	parameter	mode	powertrain	year	unit	value	percentage
Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing
Count	Count	Count	Count	Count	Count	Count	Count	Count
World	Historical	EV stock	Cars	EV	36%	Vehicles	54.1%	100,00%
Europe	Projection-APS	EV sales	Buses	BEV	25.3%	percent	25.1%	200,00%
Other	Projection-STEPS	Other	Other	Other	36%	Other	20.9%	97.2%
								
Austria	Historical	EV stock	Cars	BEV	2010	Vehicles	3.500000e+02	35000,00%
Austria	Historical	EV stock share	Cars	EV	2010	percent	7.900000e+11	78999996185300,00%
Belgium	Historical	EV stock	Buses	BEV	2010	Vehicles	3.000000e+00	300,00%
Belgium	Historical	EV sales	Vans	BEV	2010	Vehicles	7.000000e+00	700,00%
Belgium	Historical	EV stock	Vans	BEV	2010	Vehicles	6.200000e+01	6200,00%
Belgium	Historical	EV sales	Cars	BEV	2010	Vehicles	4.800000e+01	4800,00%
Belgium	Historical	EV stock	Cars	BEV	2010	Vehicles	6.100000e+01	6100,00%

# Bộ dữ liệu

Ý nghĩa từng cột

- Region (khu vực): là biến dung để mô tả khu vực kinh doanh xe điện
- Category (loại): phân loại danh mục dữ liệu
- Parameter (tham số): phân loại kiểu của dữ liệu
- Mode (chế độ): mô tả loại xe / phương tiện xe nào
- Powertrain (năng lượng chạy): mô tả loại năng lượng nào mà xe / phương tiện đó sử dụng
- Year (năm): năm kinh doanh
- Unit (đơn vị): đơn vị đo
- Value (giá trị): mang ý nghĩa thu thập
- Percentage (phần trăm): cũng giống như value nhưng nó mô tả dưới dạng phần trăm (recipe:  $\text{value} * 100 = \text{percentage}$ )

# Các mô hình sử dụng

- Từ yêu cầu và bộ dữ liệu, nhóm sử dụng các mô hình dưới để phân tích và dự báo:
  1. Hồi quy tuyến tính (Linear Regression)
  2. ARIMA (Time – Series)\*

Từ các mô hình đã áp dụng, nhóm em quyết định sử dụng mô hình ARIMA cho đề tài này

# Lý do chọn mô hình

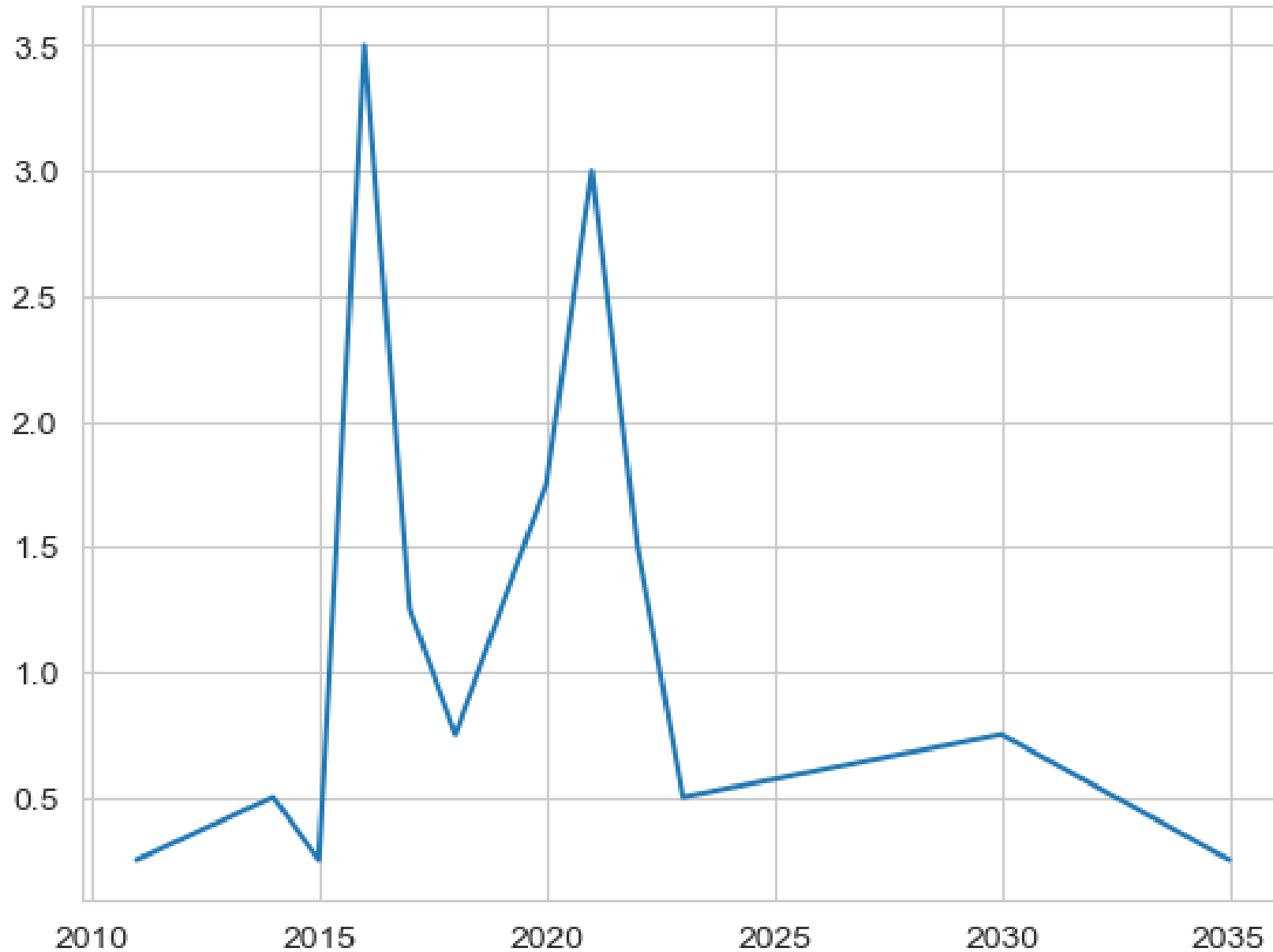
- Bộ dữ liệu này là dữ liệu doanh thu bán xe điện qua các năm là một chuỗi thời gian, thông thường mô hình ARIMA rất phù hợp cho việc xử lý các bộ dữ liệu theo thời gian (Time – Series).
- Mô hình ARIMA sử dụng dữ liệu thông tin trong quá khứ (past value) để có thể dự báo giá trị trong tương lai.
- ARIMA Model rất phù hợp cho xử lý, phân tích và dự báo cho các bộ dữ liệu liên quan đến thương mại, điều này giúp cho việc xác định chiến lược kinh doanh trong tương lai.

EDA DỮ LIỆU

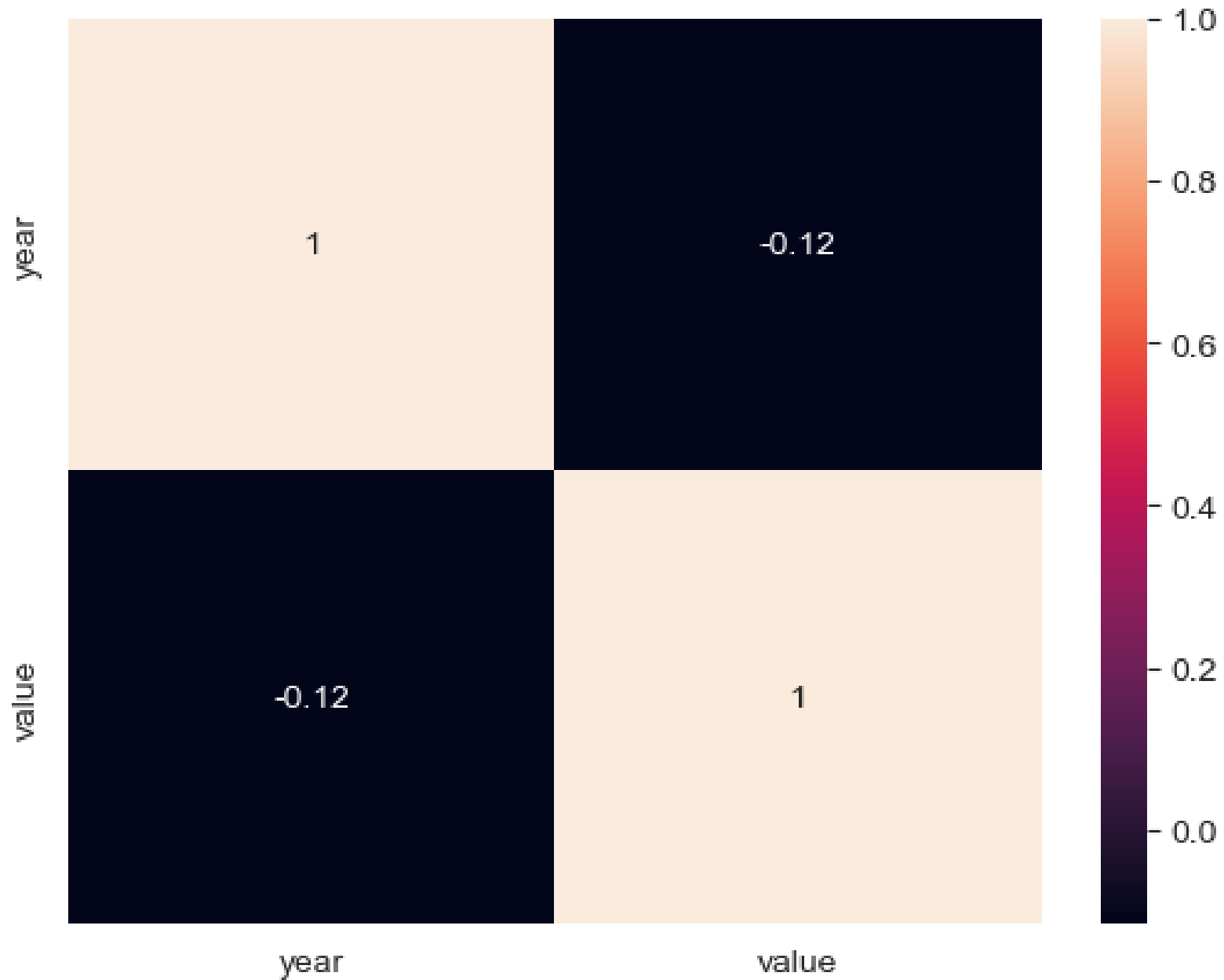


# 1. Trực quan doanh thu

Visualize EV Sales follow each years



## 2. Biểu đồ tương quan



ỨNG DỤNG MÔ HÌNH

# 1. Chia tập train/tập test để áp dụng mô hình

```
train = data_group[data_group["year"] <= 2023]
test = data_group[data_group["year"] >= 2023]
train, test
✓ [10] 14ms
```

2	2015	0.25
3	2016	3.50
4	2017	1.25
5	2018	0.75
6	2019	1.25
7	2020	1.75
8	2021	3.00
9	2022	1.50
10	2023	0.50,
	year	value
10	2023	0.50
11	2030	0.75
12	2035	0.25)

## 2. Kiểm tra tính dừng lần 1.

```
adfuller(train["value"])      trai
```

```
✓ [11] 282ms
```

```
(-2.781185328699488,  
 0.061011975232483,  
 1,  
 9,  
 {'1%': -4.473135048010974,  
  '5%': -3.28988060356653,  
  '10%': -2.7723823456790124},  
 19.272979508238105)
```

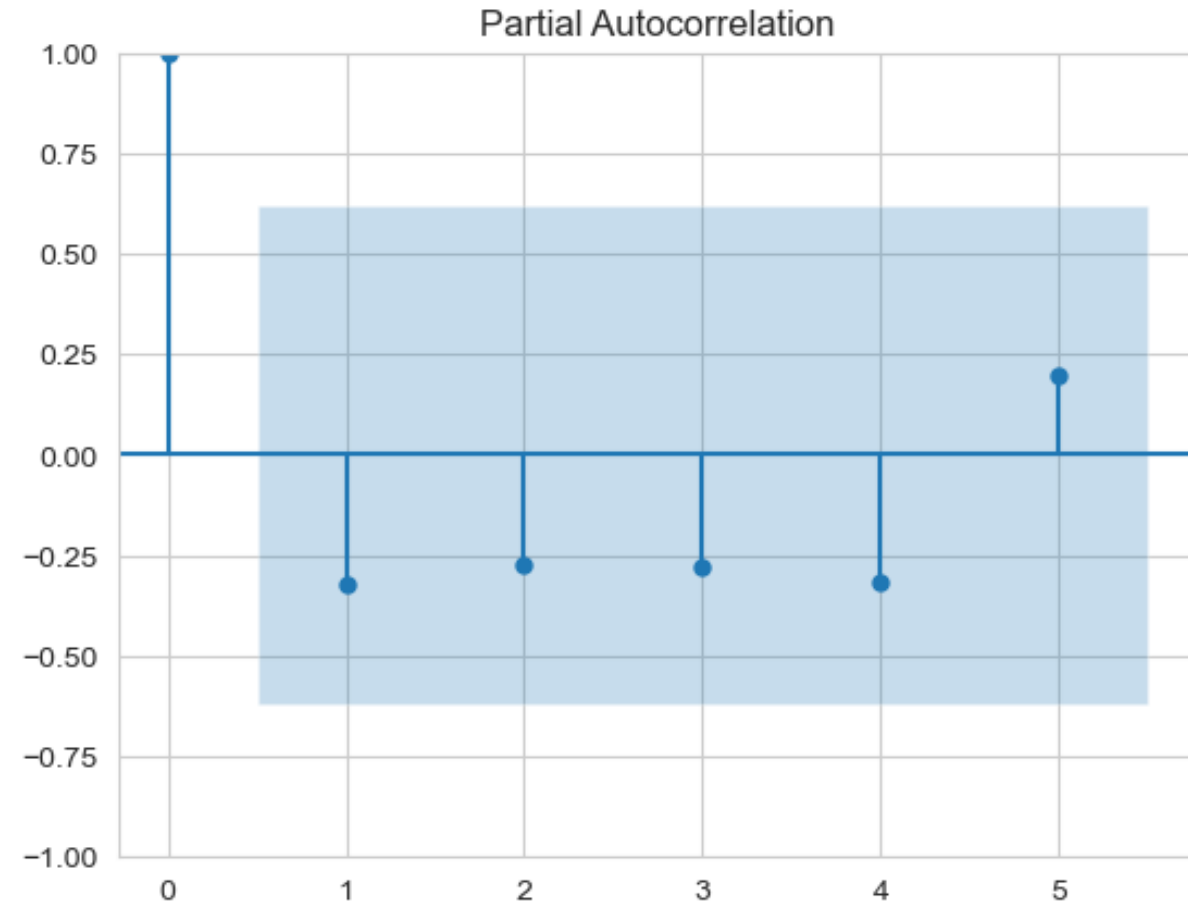
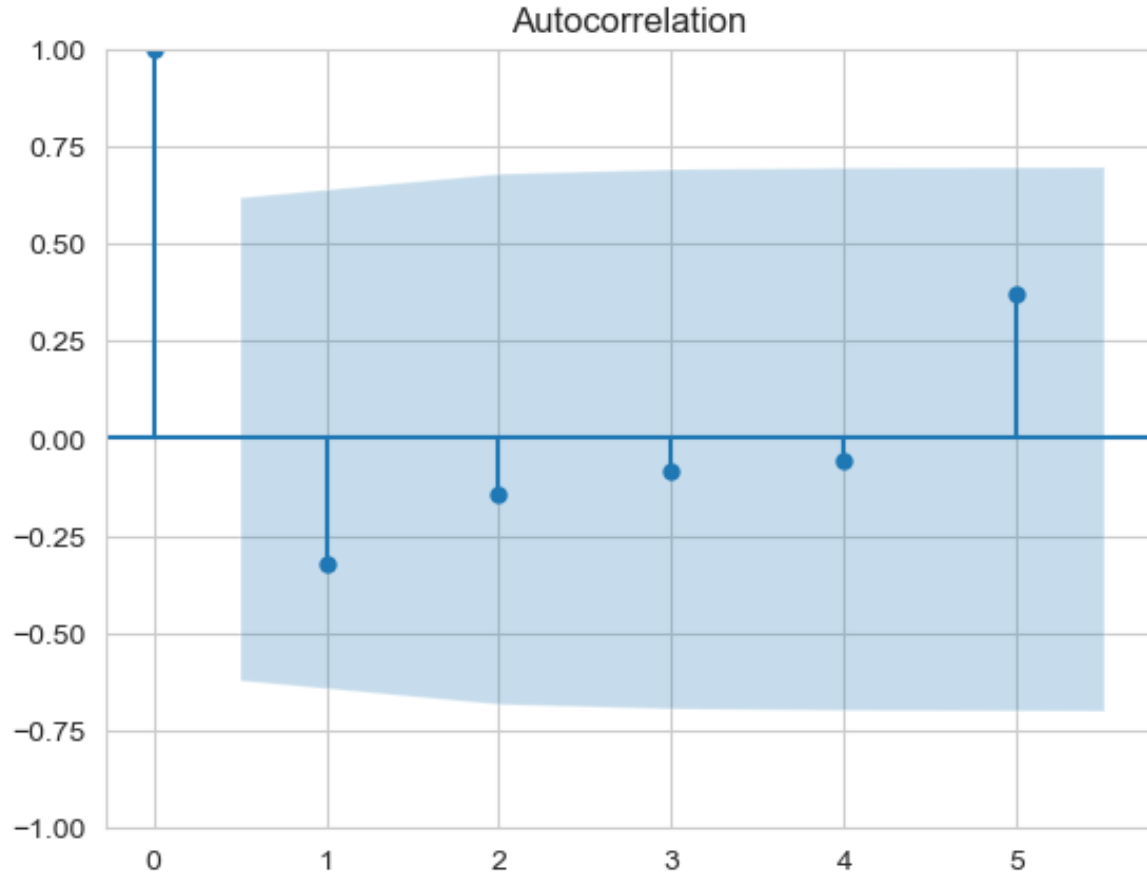
### 3. Lấy sai phân lần 1 và kiểm tra tính dừng lần 2

```
train_diff_1 = train["value"].diff().dropna()  
adfuller(train_diff_1)
```

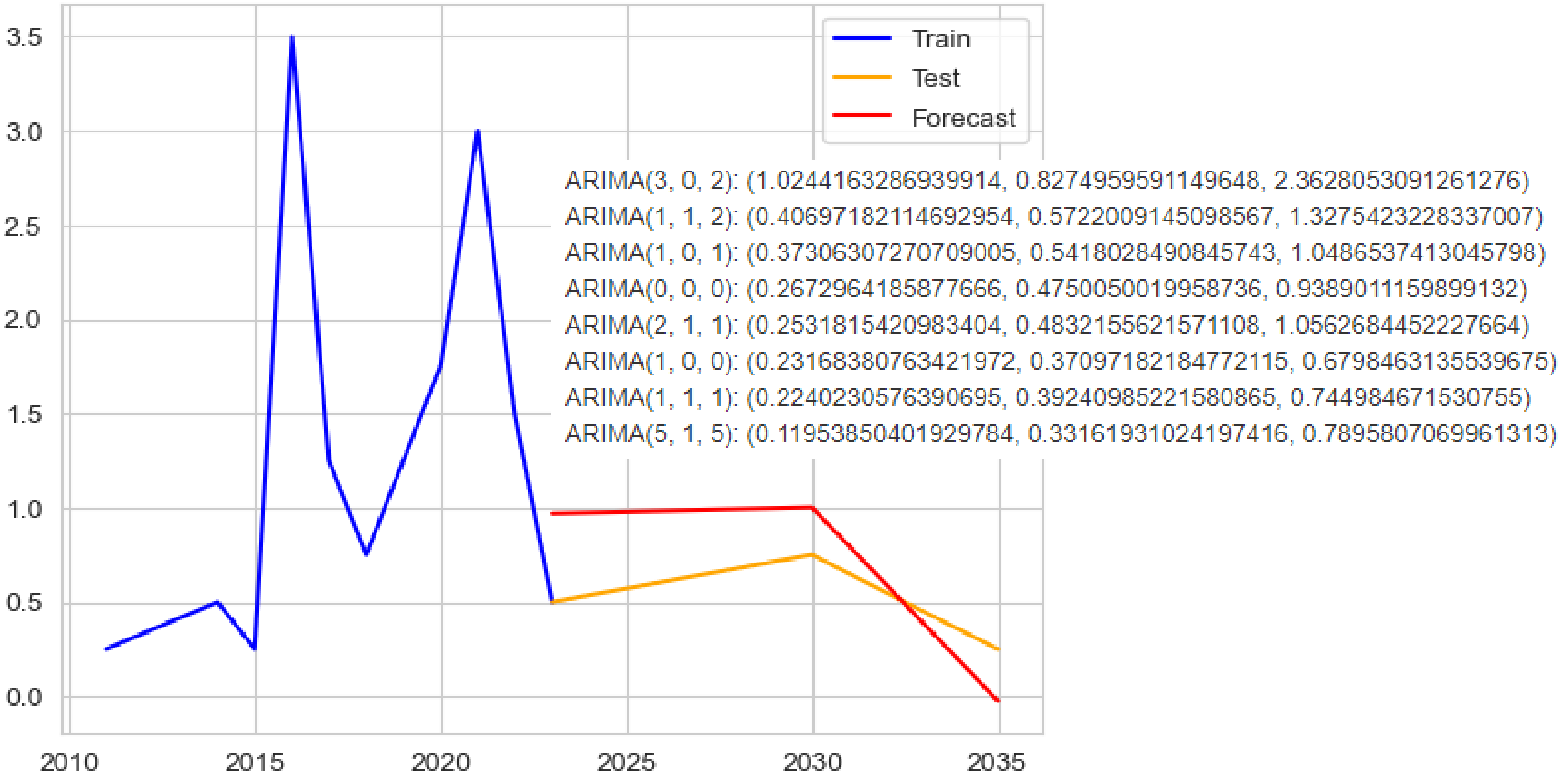
```
✓ [12] 19ms
```

```
(-3.322376869678028,  
 0.013907381219126266,  
 3,  
 6,  
 {'1%': -5.354256481481482,  
  '5%': -3.6462381481481483,  
  '10%': -2.9011977777777778},  
 13.386246362519806)
```

## 4. Kiểm tra đồ thị tương quan và tương quan 1 phần

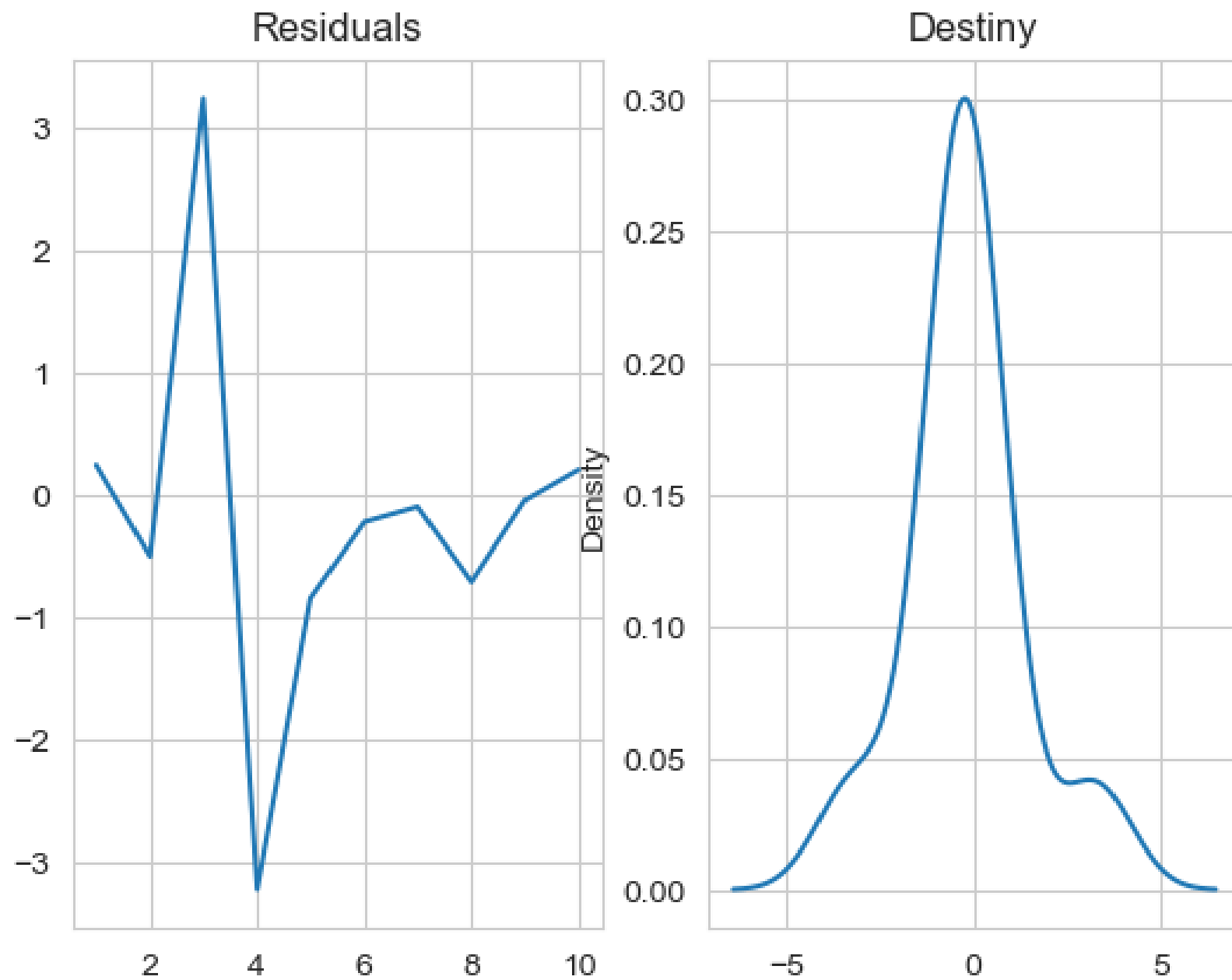


5. Tiến hành train và đánh giá chất lượng mô hình



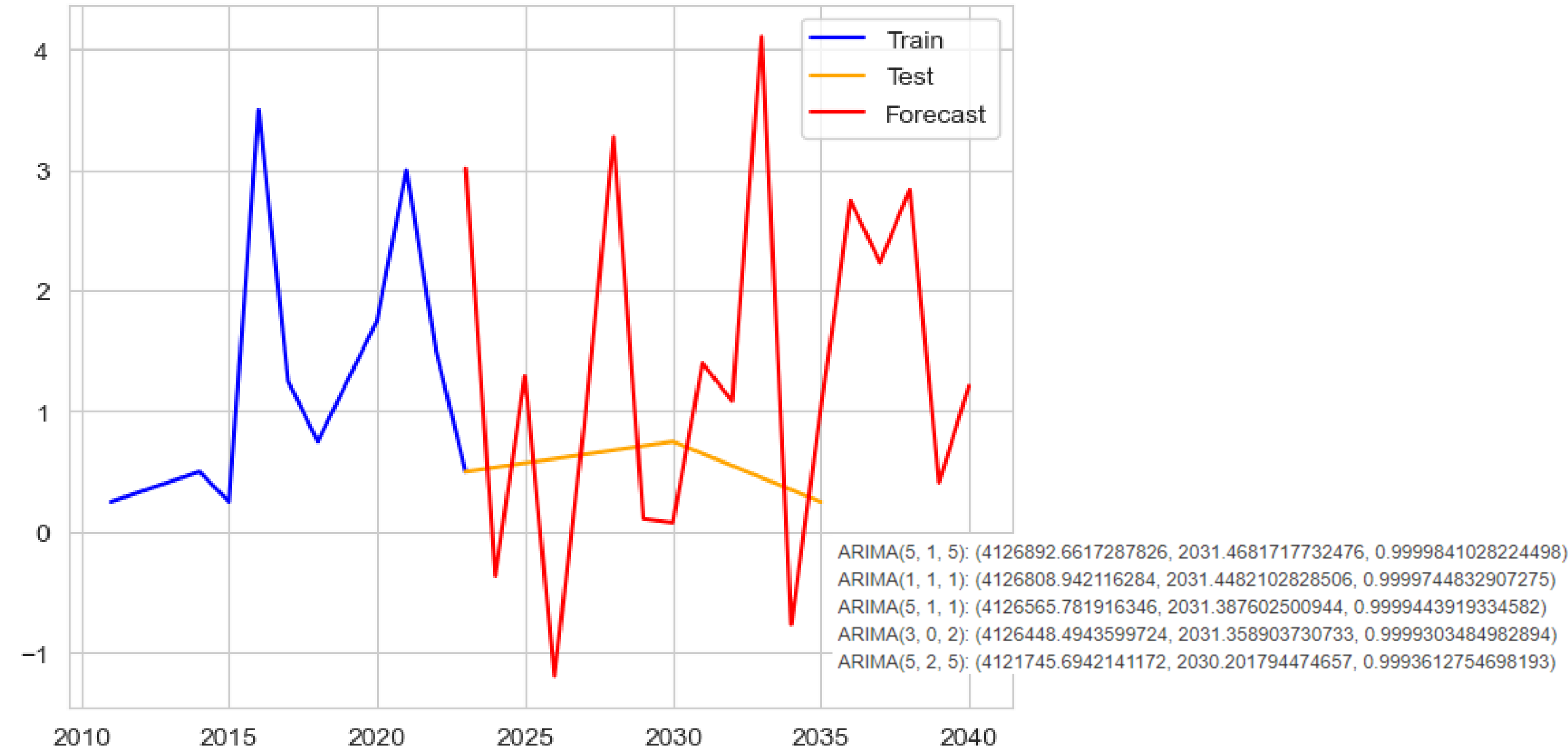


## 6. Kiểm tra phần dư sau khi train mô hình

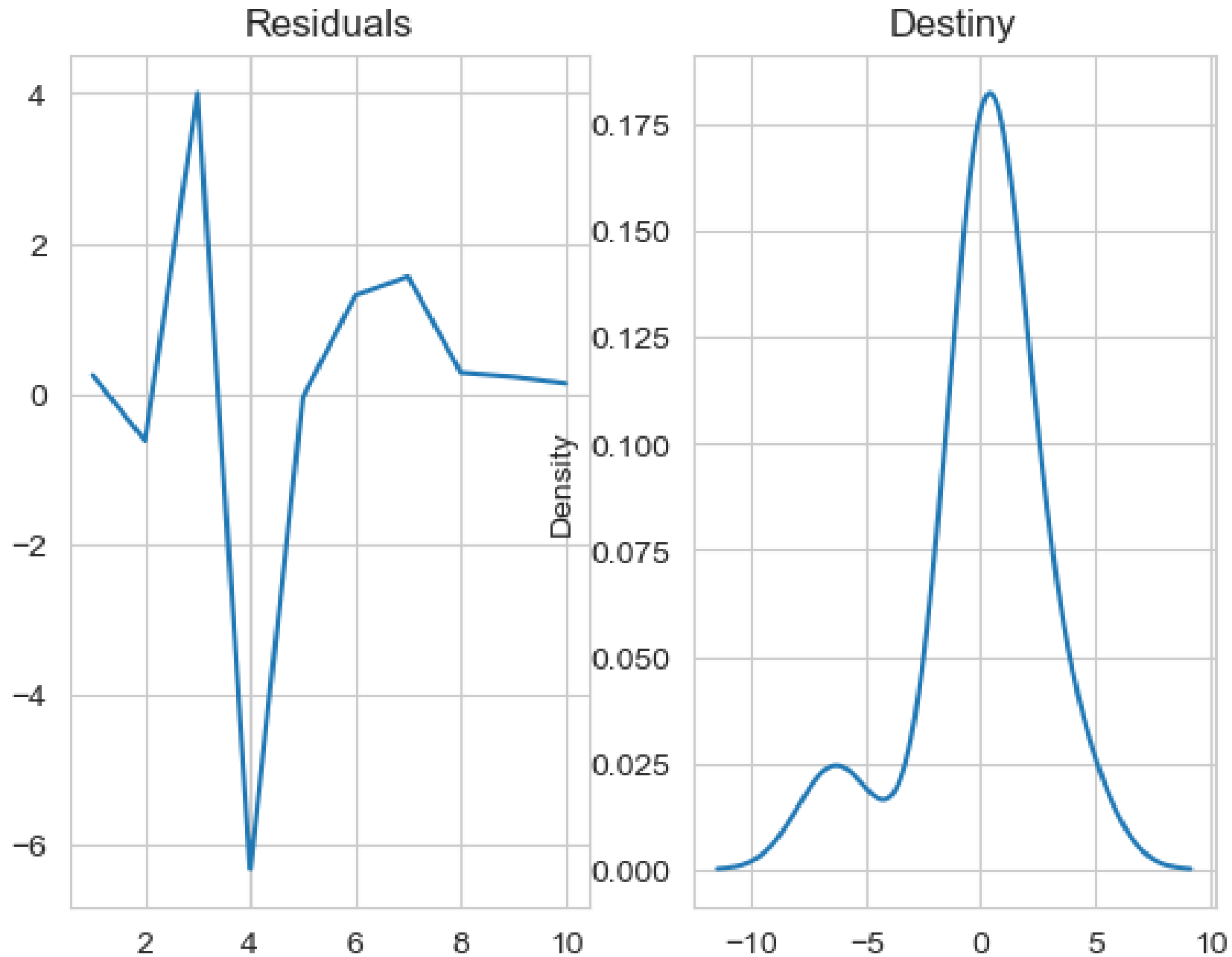


## 7. Dự báo tương lai

Forecasting EV Sales by ARIMA Modeling



## 8. Kiểm tra phần dư sau khi dự báo tương lai



THANK YOU FOR WATCHING!!!