**Cài đặt thuật toán ARIMA và thuật toán Học máy trên bộ dữ liệu Gas & Electric**

## Thống kê bộ dữ liệu

### Thống kê mô tả

|  |  |  |
| --- | --- | --- |
|  | **Gas Use** | **Electric Use** |
| **Count** | 24.00 | 24.00 |
| **Mean** | 107.79 | 945.25 |
| **Std** | 86.54 | 255.46 |
| **Min** | 26.00 | 672.00 |
| **25%** | 32.25 | 800.25 |
| **50%** | 70.50 | 854.00 |
| **75%** | 190.75 | 1006.75 |
| **Max** | 247.00 | 1445.00 |

A picture containing text, screenshot, diagram, plot

Description automatically generated

A picture containing text, screenshot, diagram, line

Description automatically generated

### Vẽ biểu đồ chuỗi thời gian

* **Gas Use**

A picture containing diagram, plot, line, text

Description automatically generated

* **Electric Use**

A picture containing diagram, plot, line, text

Description automatically generated

### Kiểm tra tính dừng của chuỗi thời gian

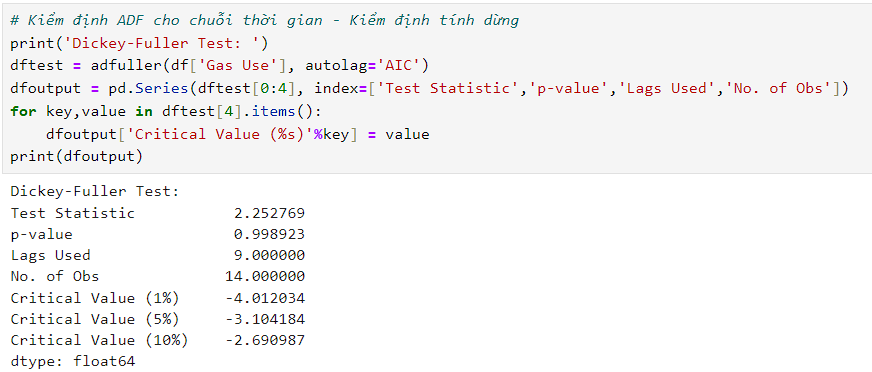
Để kiểm tra tính dừng của chuỗi thời gian, ta sử dụng phép kiểm định Augmented Dicky-Fuller

Phát biểu giả thuyết không và đối thuyết:

* : Chuỗi thời gian không có tính dừng
* : Chuỗi thời gian có tính dừng

#### Gas Use:

* Sử dụng hàm adfuller trong thư viện statmodels bằng ngôn ngữ Python ta thu được kết quả:



* Sử dụng hàm ndiffs trong thư viện pmdarima bằng ngôn ngữ Python ta thu được kết quả

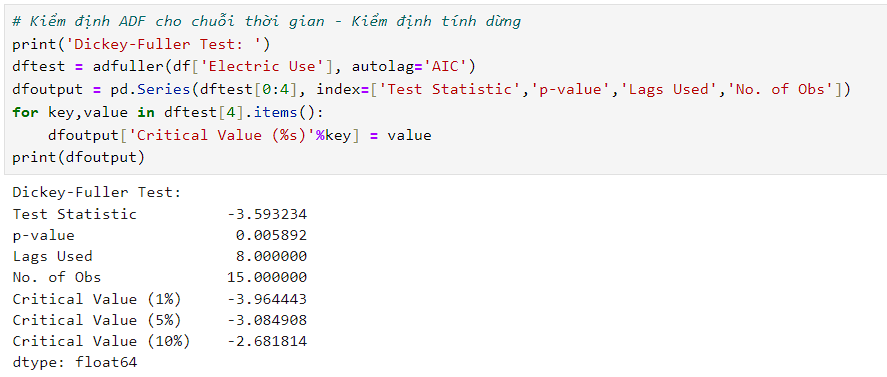
A close-up of a computer screen

Description automatically generated

* Với p-value = 0.9989 > 0.05, ta không thể bác bỏ
* **Kết luận:** Chuỗi thời gian Gas Use không có tính dừng ***(Non-stationary)*** và bậc sai phân phù hợp là 2 ***(d = 2)***

#### Electric Use:

* Sử dụng hàm adfuller trong thư viện statmodels bằng ngôn ngữ Python ta thu được kết quả:



* Sử dụng hàm ndiffs trong thư viện pmdarima bằng ngôn ngữ Python ta thu được kết quả

A white rectangular object with text

Description automatically generated with medium confidence

* Với p-value = 0.0058 < 0.05, bác bỏ , chấp nhập
* **Kết luận:** Chuỗi thời gian Electric Use có tính dừng ***(Stationary)*** và bậc sai phân phù hợp là 0 ***(d = 0)***

## Cài đặt thuật toán ARIMA

### MS Excel

#### Gas Use

* **Bước 1:** Sử dụng Real Statstics để tính ACF và PACF
* Tính ACF:

A screenshot of a graph

Description automatically generated

* Tính PACF:

A screenshot of a graph

Description automatically generated

🡪 Ta thấy có ACF(1) và PACF(1) vượt ra khỏi dải tin cậy nên p và q = 1🡪 Mô hình ARIMA sử dụng sẽ là ARIMA(1, 1, 1)

* **Bước 2:** Sử dụng Real Statistics để tính **ARIMA(1, 1, 1)**

A screenshot of a computer

Description automatically generated

* Kết quả sau sử khi tính:

A white sheet of paper with black text

Description automatically generated

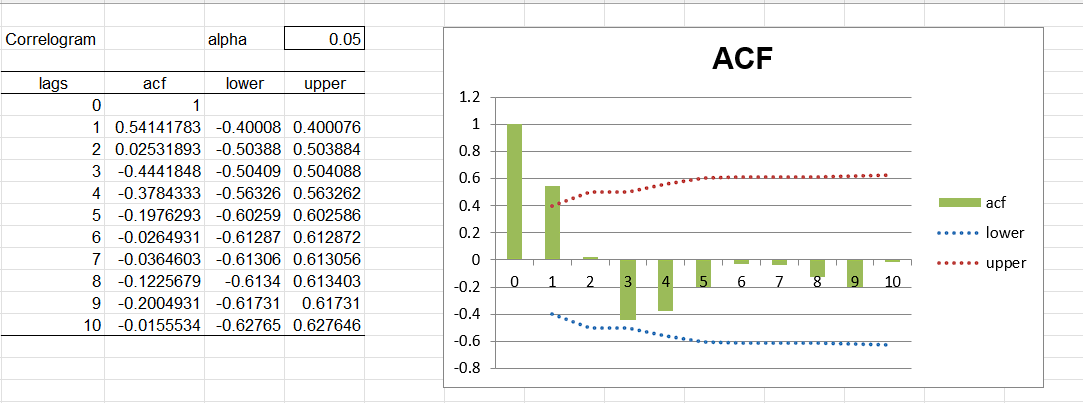
* 🡪 Vậy phương trình có dạng:
* Kết quả dự đoán 10 tháng tiếp theo:

A table with numbers on it

Description automatically generated

#### Electric Use

* **Bước 1:** Sử dụng Real Statistics để tính ACF và PACF
* Sơ đồ ACF

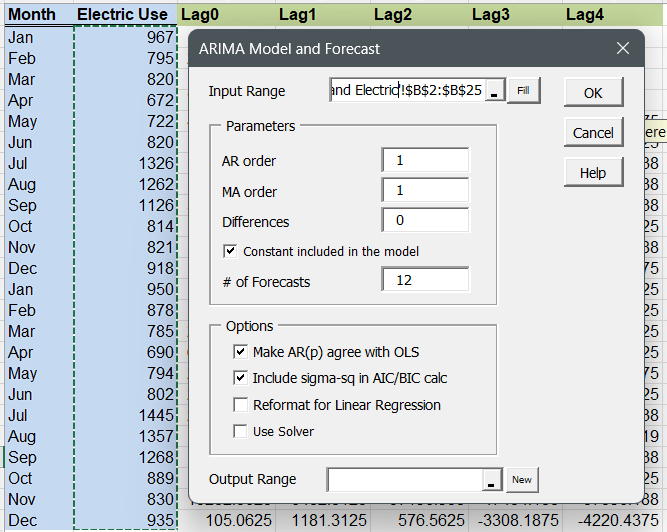


* Sơ đồ PACF

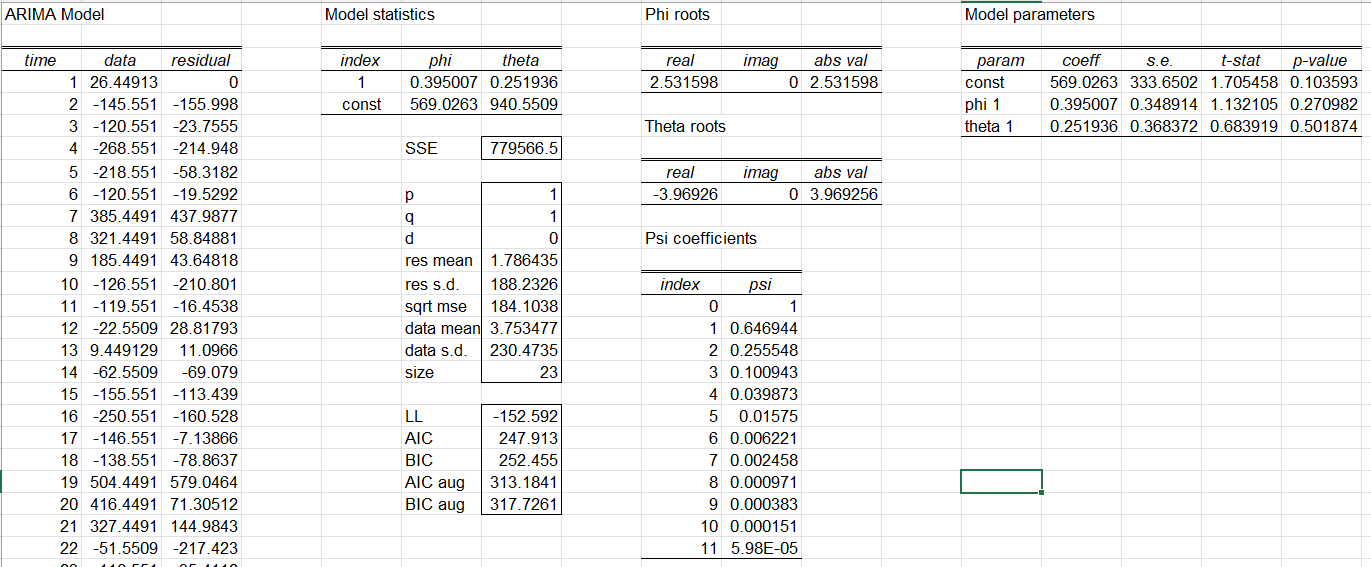
A graph on a white sheet

Description automatically generated

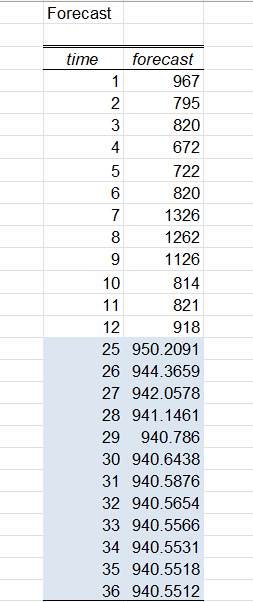
* Vậy ta thấy ACF(1) và PACF(1) vượt ra khỏi dải tin cậy nên ta có mô hình ARIMA(1,0,1)
* **Bước 2:** Sử dụng Real Statistics để dự đoán giá trị (***ARIMA(1, 0, 1)***)



* Đọc kết quả từ Real Statistics



* Vậy phương trình có dạng:
* Kết quả dự đoán cho 12 tháng tiếp theo



### Python

#### Gas Use

* **Bước 1:** Import các thư viện và đọc file

A screenshot of a computer code

Description automatically generated

* **Bước 2:** Chia tập dữ liệu thành train - test theo tỉ lệ 7:3 và sử dụng hàm auto\_arima để tìm bộ số tốt nhất

A screenshot of a computer

Description automatically generated

* Mô hình ARIMA tốt nhất

A screenshot of a computer code

Description automatically generated

* **Bước 3**: Tiến hành testing dữ liệu và dự đoán trong 12 tháng tiếp theo

A screenshot of a computer

Description automatically generated

**Bước 4:** Vẽ sơ đồ

A graph with lines and text

Description automatically generated with medium confidence

#### Electric Use

* **Bước 1:** Import các thư viện và đọc file

A screenshot of a computer program

Description automatically generated

* **Bước 2:** Chia tập dữ liệu thành train - test theo tỉ lệ 7:3 và sử dụng hàm auto\_arima để tìm bộ số tốt nhất

A screenshot of a computer program

Description automatically generated

* Mô hình ARIMA tốt nhất

A blue and black text

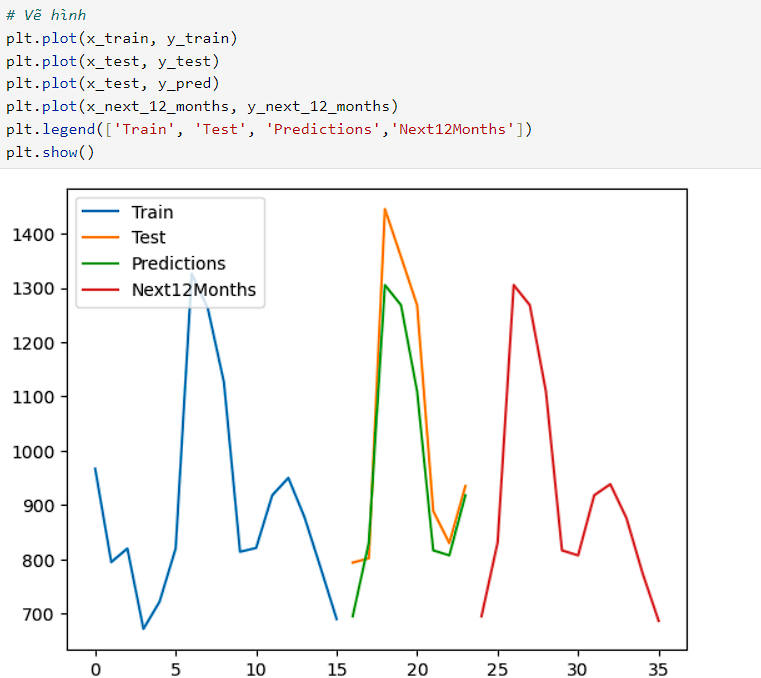
Description automatically generated

* **Bước 3**: Tiến hành testing dữ liệu và dự đoán trong 12 tháng tiếp theo

A screenshot of a computer code

Description automatically generated

* **Bước 4:** Vẽ sơ đồ



## Cài đặt thuật toán Machine Learning (SVR)

### Gas Use

* **Bước 1:** Import các thư viện và đọc file

A screenshot of a computer code

Description automatically generated

* **Bước 2:** Chia tập dữ liệu thành train - test theo tỉ lệ 8:2 và chuẩn hóa dữ liệu

A screenshot of a computer program

Description automatically generated

* **Bước 3:** Sử dụng hàm SVR để chọn mô hình dự đoán

A screenshot of a computer

Description automatically generated

* **Bước 4:** Tiến hành testing dữ liệu và dự đoán trong 12 tháng tiếp theo

A screenshot of a computer code

Description automatically generated

* **Bước 5:** Vẽ sơ đồ

A graph with numbers and lines

Description automatically generated with medium confidence

### Electric Use

* **Bước 1:** Import các thư viện và đọc file

A screenshot of a computer program

Description automatically generated

* **Bước 2:** Chia tập dữ liệu thành train - test theo tỉ lệ 8:2 và chuẩn hóa dữ liệu

A screenshot of a computer program

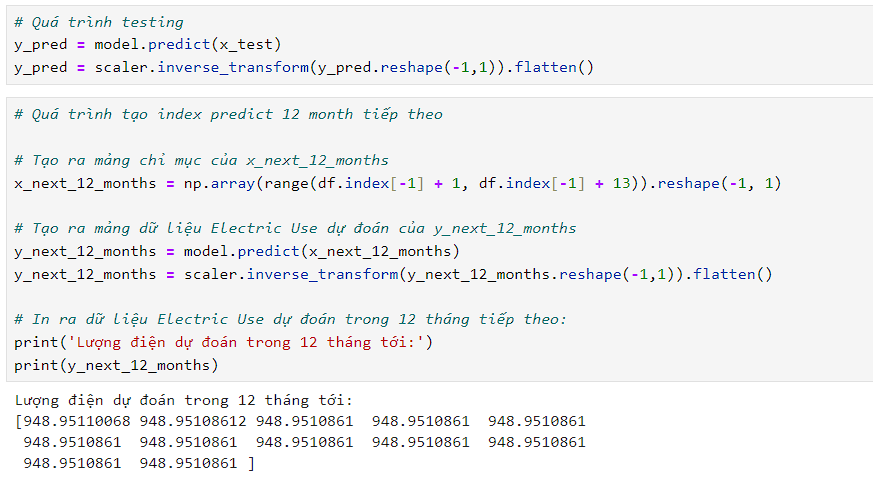
Description automatically generated

* **Bước 3:** Sử dụng hàm SVR để chọn mô hình dự đoán

A screenshot of a computer

Description automatically generated

* **Bước 4:** Tiến hành testing dữ liệu và dự đoán trong 12 tháng tiếp theo



* **Bước 5:** Vẽ sơ đồ

A graph with lines and numbers

Description automatically generated with medium confidence