Search

Write

Guide to Time Series Analysis with Python — 3: Autoregressive Process

Buse Köseoğlu · Follow

4 min read · Aug 8, 2023

117

In the previous article, we examined the moving average process. This time we will examine another statistical model, the autoregressive process.

You can find previous articles below.

1.Guide to Time Series Analysis with Python — 1: Analysis Techniques and Baseline Model

2.Guide to Time Series Analysis with Python — 2: Moving Average Process

If you're ready, let's start.

What is Autoregressive Process (AR)?

Autoregressive process is a statistical model used for time series to predict the future by looking at historical data. Just as the moving average process uses the effect of past errors, the autoregressive process says that the past values in the time series affect the present. We can see how this is used in the equation below. Here, the parameter we want to determine is the "p" parameter. This p-value decides how far back we go. It is important to find the optimum number.

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

Constant

The magnitude of of influence of past values on present ones
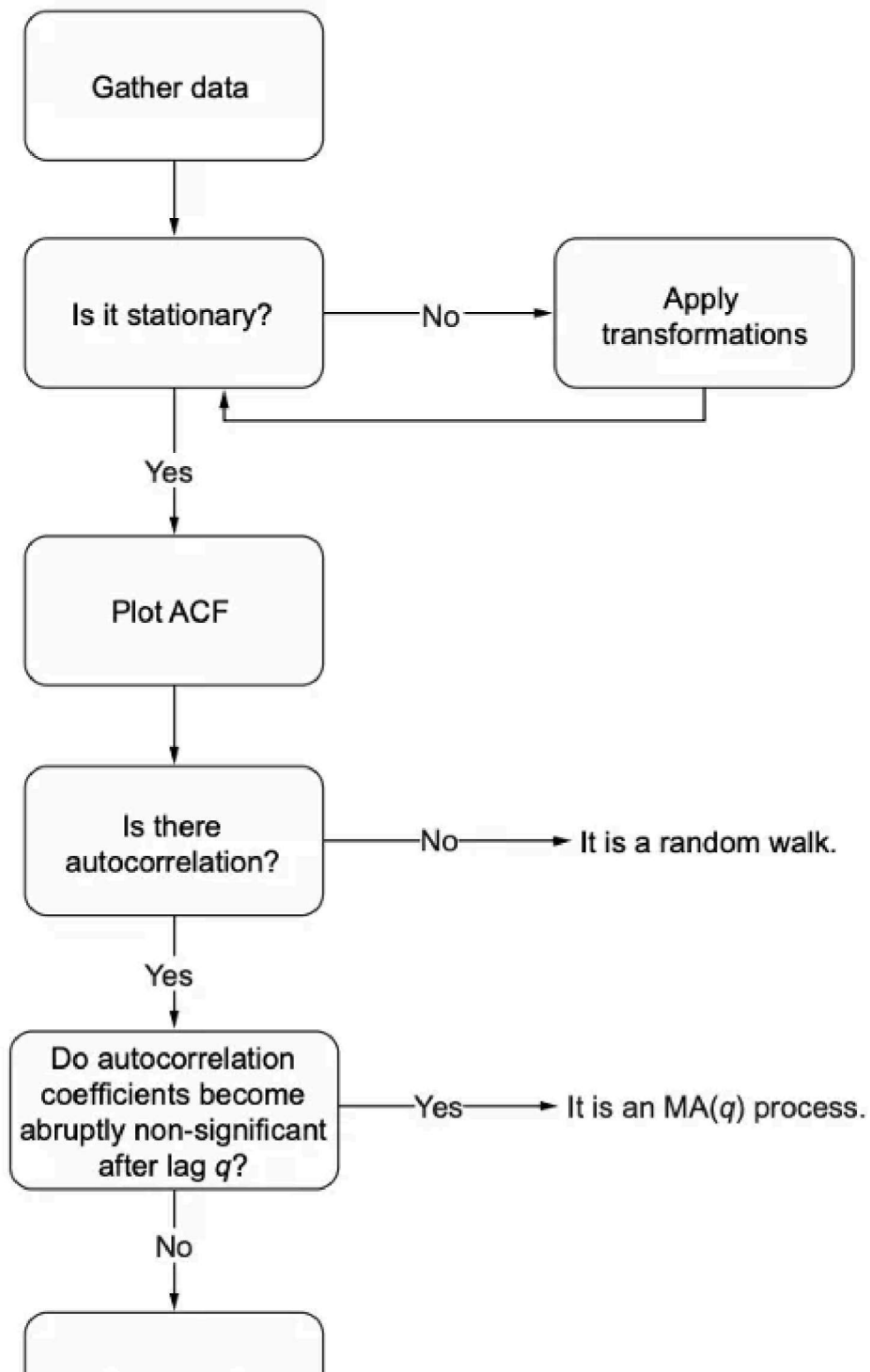
present error term

Specifying the p Parameter

We use a method like the method in the moving average process to determine this parameter. But this time we will use PACF (Partial

Autocorrelation Function) plot instead of ACF plot. PACF gives the partial correlation of the time series with its lagged values.

If we examine the roadmap below, we can see how we can reach the result step by step. Here, first of all, we need to see if the data fits the moving average process. If it does not fit, then we need to check whether it is suitable for the autoregressive process with the PACF graph.
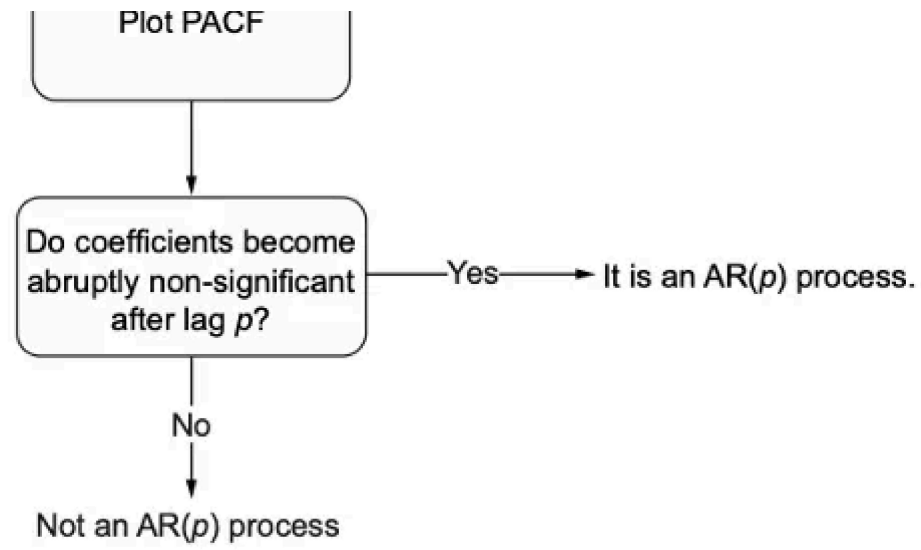
```
┌─────────────────┐
│   Gather data   │
└─────────────────┘
         │
         ▼
┌─────────────────┐      ┌──────────────────┐
│  Is it          │─No──▶│     Apply        │
│  stationary?    │      │  transformations │
└─────────────────┘      └──────────────────┘
         │     ▲                   │
        Yes    └───────────────────┘
         │
         ▼
┌─────────────────┐
│    Plot ACF     │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│  Is there       │─No──▶ It is a random walk.
│  autocorrelation?│
└─────────────────┘
         │
        Yes
         │
         ▼
┌──────────────────────┐
│ Do autocorrelation   │
│ coefficients become  │─Yes──▶ It is an MA(q) process.
│ abruptly non-significant│
│ after lag q?         │
└──────────────────────┘
         │
         No
         │
         ▼
```
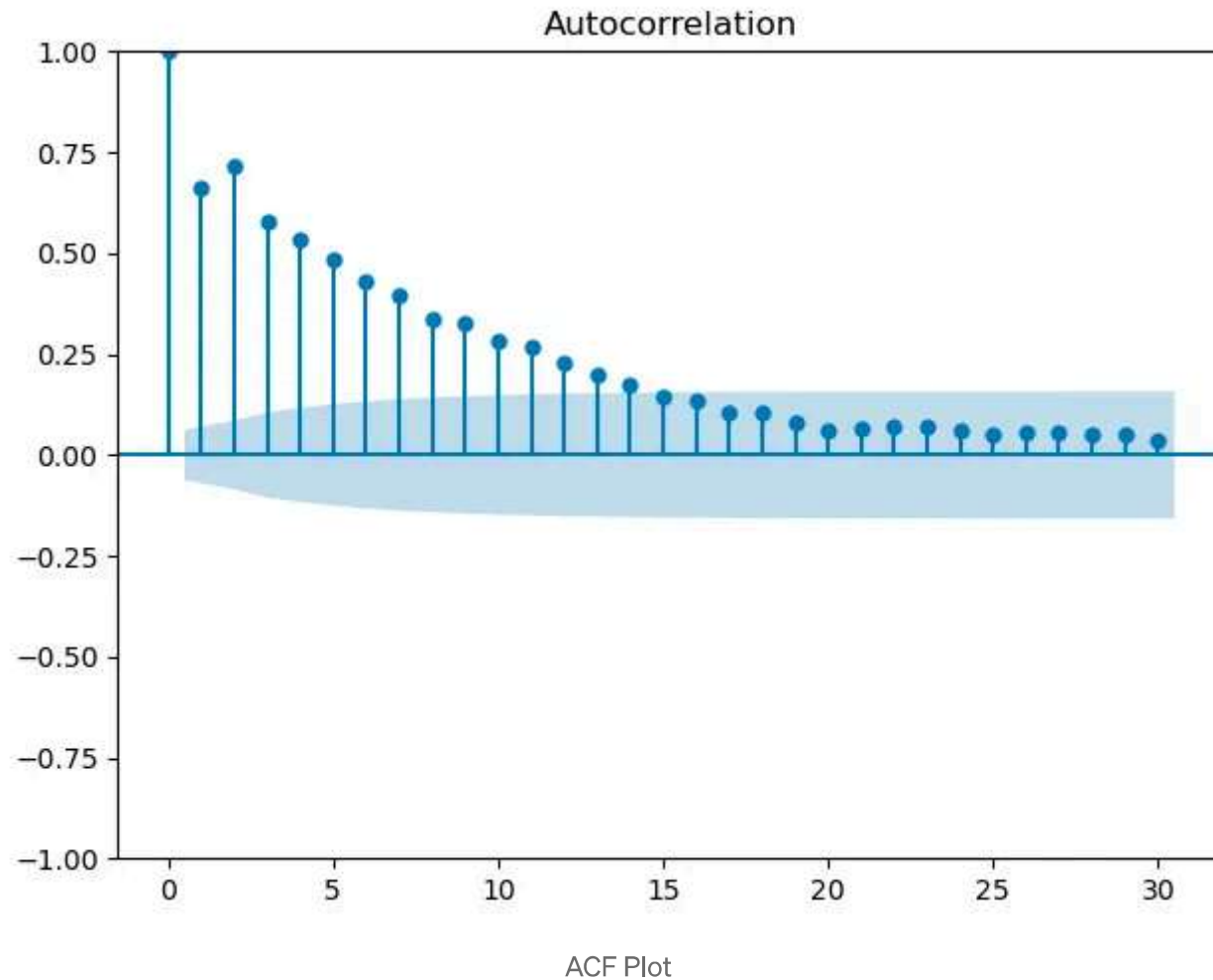
Figure from Time series forecasting in python — Marco Peixerio

Let's say we have stationary data. Then, let's draw an ACF plot with this data, and as a result, we get the following graph.

```
plot_acf(ar2_process)

plt.tight_layout()
```

ACF Plot

When the above ACF graph is examined, we can see that there are coefficients that are significant after lag 0. When examining this graph in the moving average process, we expected only 1 significant coefficient and the other coefficients to remain within the confidence interval (blue area). Since there is no such situation here, we can say that this data is not suitable for the moving average process and draw the PACF graph.

```
plot_pacf(ar2_process, lags=20)

plt.tight_layout()
```


PACF Plot
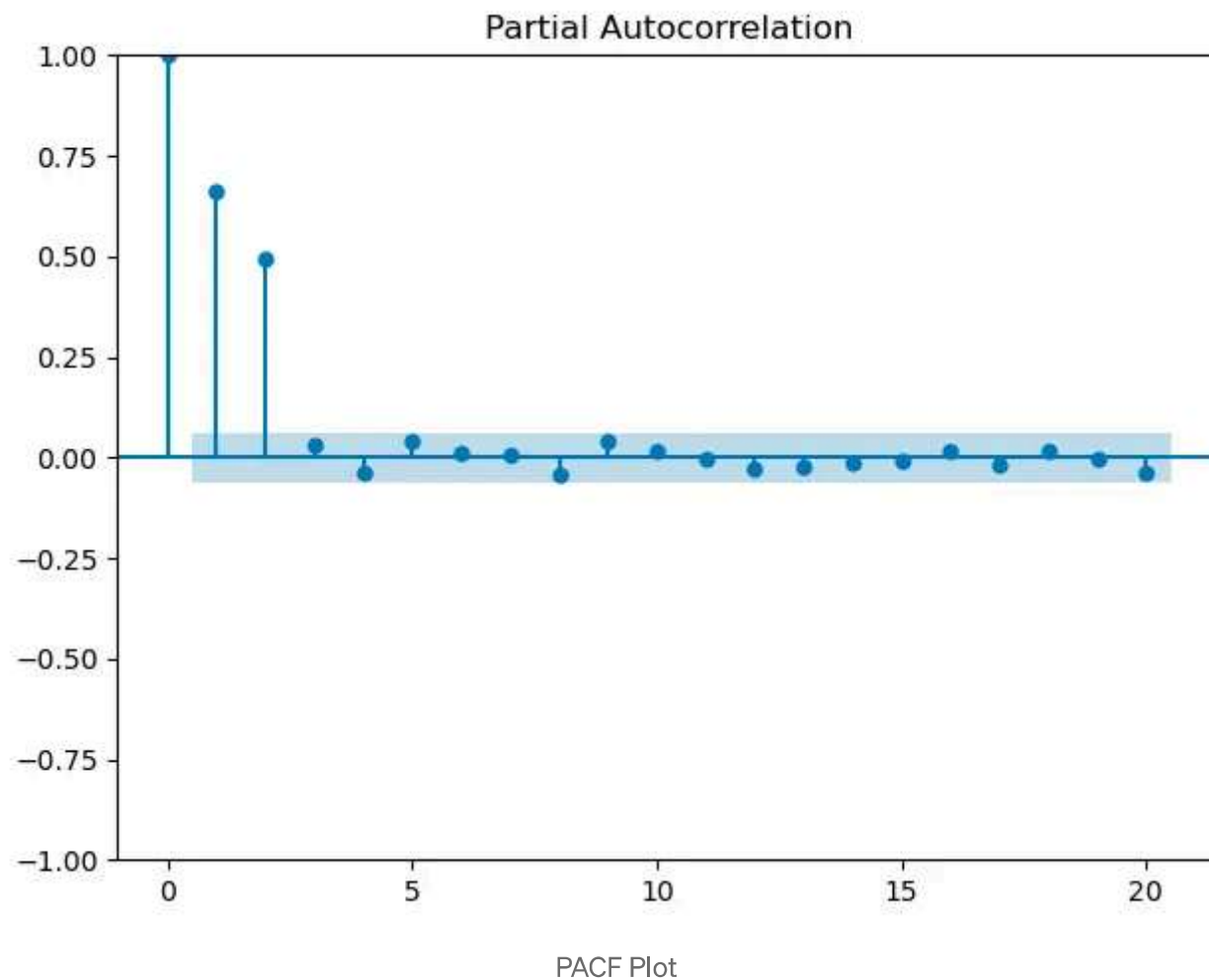
If we examine the PACF graph above, we can see that the coefficients after lag 2 are non-significant. At the same time, with this graph, we determine that the "p" parameter is also 2, and we have the AR(2) process.

Modeling AR(2)

The next processes are very similar to the moving average process. We can implement the AR(2) process with only minor changes. After separating the data as train and test, we can use the following function for modeling.

```python
def rolling_forecast(df: pd.DataFrame, train_len: int, horizon: int, window: int

    total_len = train_len + horizon
    end_idx = train_len

    if method == 'mean':
        pred_mean = []

        for i in range(train_len, total_len, window):
            mean = np.mean(df[:i].values)
            pred_mean.extend(mean for _ in range(window))

        return pred_mean

    elif method == 'last':
        pred_last_value = []

        for i in range(train_len, total_len, window):
            last_value = df[:i].iloc[-1].values[0]
            pred_last_value.extend(last_value for _ in range(window))
```

```python
        return pred_last_value

    elif method == 'AR':
        pred_AR = []

        for i in range(train_len, total_len, window):
            model = SARIMAX(df[:i], order=(2,0,0))
            res = model.fit(disp=False)
            predictions = res.get_prediction(0, i + window - 1)
            oos_pred = predictions.predicted_mean.iloc[-window:]
            pred_AR.extend(oos_pred)

        return pred_AR
```

The important part here is the SARIMAX(df[:i], order=(2,0,0)) part in the "AR" condition. In the MA(2) process, the order was written as (0,0,2). In AR(2), it is written as (2,0,0).

Actually, the process is that simple. Both MA and AR contain very similar implementations. You can find the codes for both in the github repo. If we examine the AR model and base model results, we can see the graph below. Here we can see that the autoregressive process captures highs and lows better than the base models and works better than the base model as we expected.
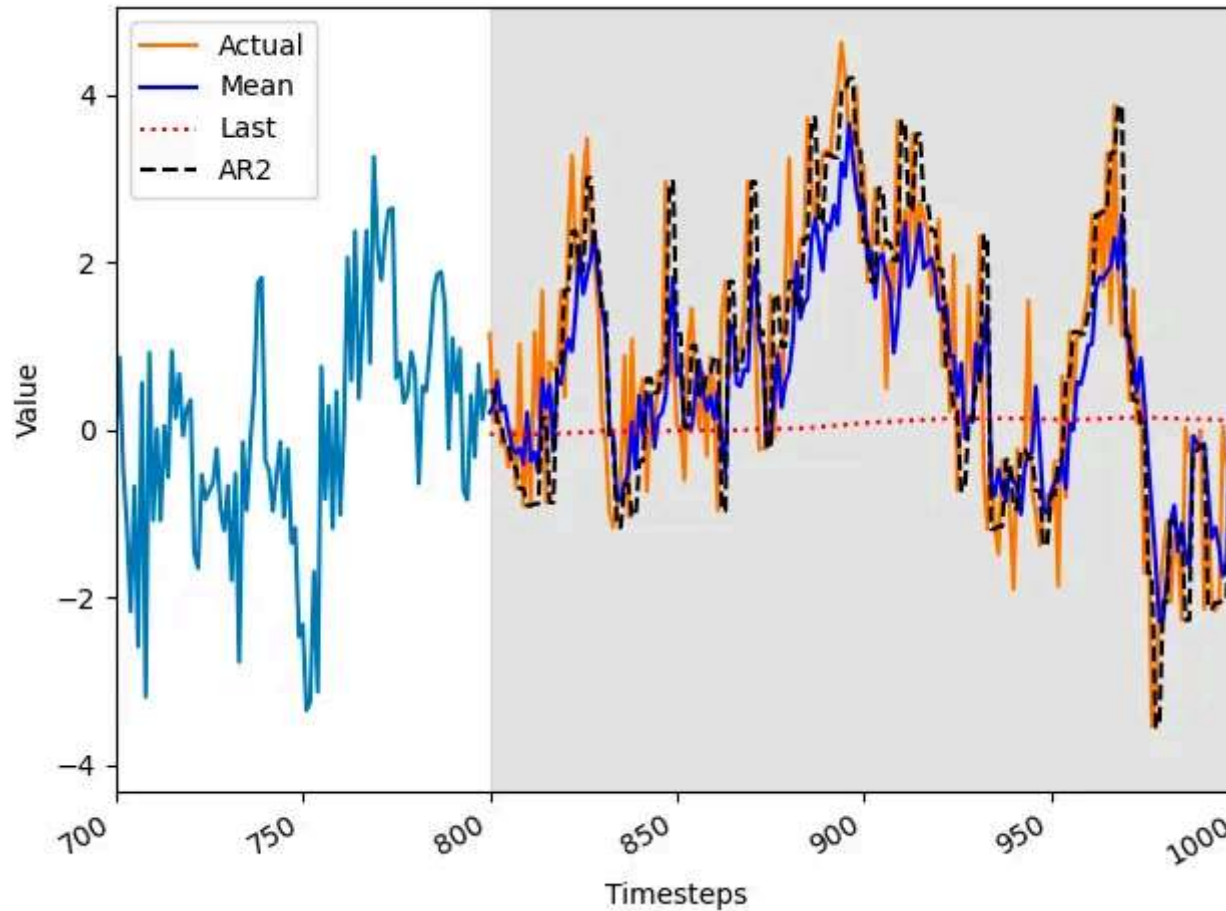
Thank you for reading. See you in the next article.

References

- Time series forecasting in python — Marco Peixerio

- Tan,Y.-F.;Ong,L.-Y.;Leow, M.-C.; Goh, Y.-X. Exploring Time-Series
  Forecasting Models for Dynamic Pricing in Digital Signage Advertising.

Future Internet 2021, 13, 241. https://doi.org/10.3390/ fi13100241

- Machine learning with time series — Markus Löning

- https://www.influxdata.com/time-series-forecasting-methods/

- https://builtin.com/data-science/time-series-forecasting-python

- https://www.machinelearningplus.com/time-series/time-series-analysis-python/

- https://towardsdatascience.com/interpreting-acf-and-pacf-plots-for-time-series-forecasting-af0d6db4061c

- https://en.wikipedia.org/wiki/Partial_autocorrelation_function

Time Series Forecasting   Autoregressive   Autoregressive Model   Time Series Analysis

Correlation

Written by Buse Köseoğlu

448 Followers

BAU Software Engineering | Data Science | https://www.linkedin.com/in/busekoseoglu/

More from Buse Köseoğlu

- The formula for R^2 is

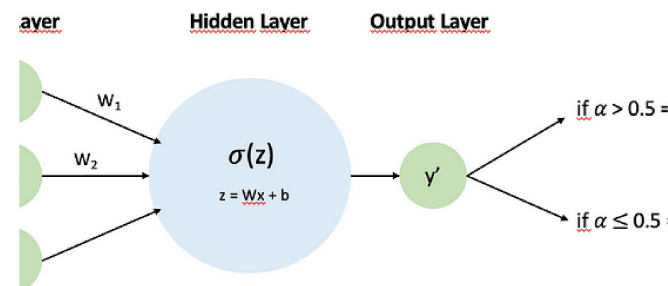$$R^2 = 1 - \frac{SS_{Residual}}{SS_{Total}}$$

while the formula for adjusted R^2 is

$$R^2_{adj} = 1 - \frac{MS_{Residual}}{MS_{Total}}$$

Buse Köseoğlu in BAU Yazılım ve Bilişim Kulübü

Model Performansını Değerlendirmek: Regresyon

Regresyonda bağımlı X değişkenlerini alarak bağımsız Y değişkenini tahmin etmeye…

Buse Köseoğlu

One Layer Neural Network From Scratch — Classification

Classification is a supervised learning method. In supervised learning, we have…

3 min read · Feb 11, 2021   7 min read · Jul 28, 2023

👤 Buse Köseoğlu   👤 Buse Köseoğlu in BAU Yazılım ve Bilişim Kulübü

TEMEL SWITCH KONFİGÜRASYONU

Bu yazımda iki switch ve iki bilgisayar ile kurulmuş bir topolojide çalışacağız....

SQL Nedir? Temel SQL Komutları

SQL (Structured Query Language — Yapılandırılmış Sorgu Dili) veritabanı yöneti...

6 min read · Aug 28, 2020   3 min read · Dec 15, 2020

See all from Buse Köseoğlu

Recommended from Medium
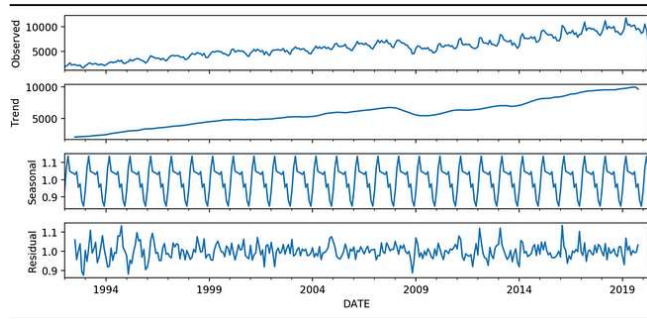
Roshmita Dey

Time Series Decomposition

Time series decomposition is a vital technique that helps in understanding the...

5 min read · Jan 13, 2024

 118   💬 1

Kaan Aslan

Time Series Forecasting with a Basic Transformer Model in...

Time series forecasting is an essential topic that's both challenging and rewarding, with ...

10 min read · Jan 12, 2024

 130   💬 2

Lists

Practical Guides to Machine Learning

10 stories · 1528 saves

Predictive Modeling w/ Python

20 stories · 1271 saves

ChatGPT prompts

48 stories · 1654 saves

Zaki Nurkholis

MindsDB team in MindsDB

A Practical Guide on Scikit-learn for Time Series Forecasting

While most machine learning algorithms available in scikit-learn (and various other...
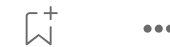
14 min read · Dec 12, 2023

How to pick the best-performing time-series AI model for your...
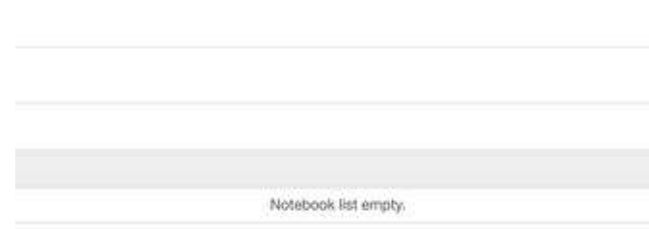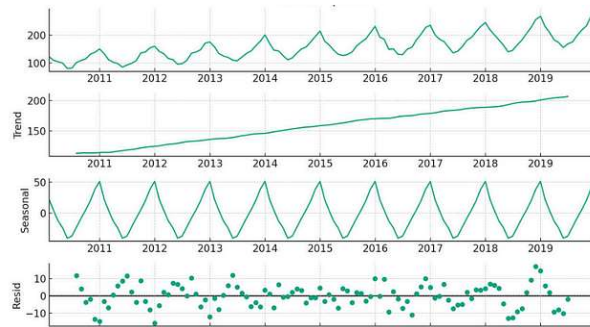
This practical guide will help you easily set up an automated framework to evaluate time-...

13 min read · Apr 27, 2024

96

28

Vansh Jatana in The Deep Hub

Max Zhang

Mastering Time Series Forecasting

Time series forecasting is an important aspect of Data science and is used in differe…

8 min read · Feb 18, 2024

A guide to time series forecasting with arima in python 3

Introduction

✦ · 14 min read · Jan 21, 2024

87   2

135

See more recommendations

Help   Status   About   Careers   Press   Blog   Privacy   Terms   Text to speech   Teams