

Zaman Serisi Nedir?

Zamana göre sıralanmış gözlem değerlerinden oluşan veridir. Zaman serisi verileri, belirli bir değişkenin zaman içinde nasıl değiştiğini incelemek ve gelecekteki modelleri tahmin etmek için kullanılabilir.

Zaman Serisi Hangi Problemlerde Kullanılmaktadır?

- **Tahmin:** Gelecek veri noktalarını tahmin etmek. Örneğin, gelecek yılın satış tahminleri, hava durumu tahminleri.
- **Anomali Tespiti:** Anormal veya beklenmedik olayları belirlemek. Örneğin, finansal dolandırıcılık tespiti, makina arızalarının öngörülmesi.
- **Desen ve Trend Analizi:** Verinin zaman içindeki desenlerini ve trendlerini incelemek. Örneğin, mevsimsel satış dalgalanmaları, uzun dönemli ekonomik trendler.
- **Kontrol ve Optimizasyon:** Sistemlerin kontrolü ve optimizasyonu. Örneğin, enerji tüketimi yönetimi, stok yönetimi.

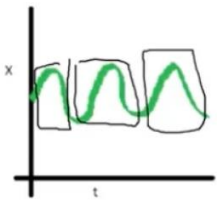
Sıcaklık, Hisse senedi fiyatları, Hisse senedi hacmi, Elektrik tüketimi, Hastalık vaka sayıları veya diğer çeşitli ölçümleri içeren problemlerde kullanılabilir. Her bir gözlem zaman sıralıdır ve genellikle eşit zaman aralıklarına sahiptir (gün, hafta, ay, yıl vb.).

Zaman serilerini tanımlamak için kullanılan bazı kavramlar vardır. Bu kavramlar; **Stationary, Trend, Seasonality, Cycle**.

1. Durağanlık (Stationary)

Bir serinin istatistiksel özelliklerinin zaman içerisinde değişmemesidir. Bir zaman serisinin ortalaması, varyansı ve kovaryansı zaman boyunca sabit kalıyorsa, serinin durağan olduğu ifade edilir.

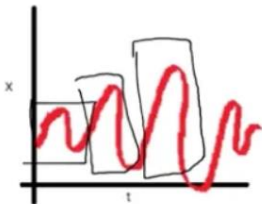
Stationary (durağan) zaman serisi verileri genellikle daha kolay modellemek ve tahmin yapmak için kullanılır çünkü istatistiksel özellikleri sabittir. Birçok zaman serisi tahmin modeli stationary varsayımını kabul eder.



Stationary series

Durağan

Her bir kutucuk bir zaman periodu. Herbir kutu için ortalama alındığı düşünüldüğünde veya standart sapması alındığında her bir kutucuk değeri birbirlerine yakın çıkacaktır. Durağanlık belirli bir pattern, belirli bir model, belirli bir tahmin edilebilirlikte olursa, daha tahmin edilebilir olur.



Non-Stationary series

Ortalama varyans ya da kovaryansın biri zamana göre değişiyorsa veri kümesi durağan olmayan olarak adlandırılır. Durağan olmayan zaman serisi verileri, istatistiksel özellikleri sabit olmadığı için modellemek ve tahmin yapmak daha zor olabilir. Durağan olmayan verileri, durağan hale getirmek için dönüştürme veya fark alma gibi işlemler uygulanabilir. Durağan olmayan bir yapıya baktığımızda, açıkça görülmektedir ki farklılıklar vardır. Kutucukların ortalaması alındığında fark fazla olacaktır.

Durağanlığı Nasıl Test Edebiliriz?

Bir zaman serisinin durağanlığını iki şekilde test edebilirsiniz:

Sezgisel yaklaşım: Görsel değerlendirme

İstatistiksel yaklaşım: Unit Root testi

Durağanlığı kontrol etmek için kullanabileceğiniz çeşitli birim kök testleri vardır. En popüler olanlar:

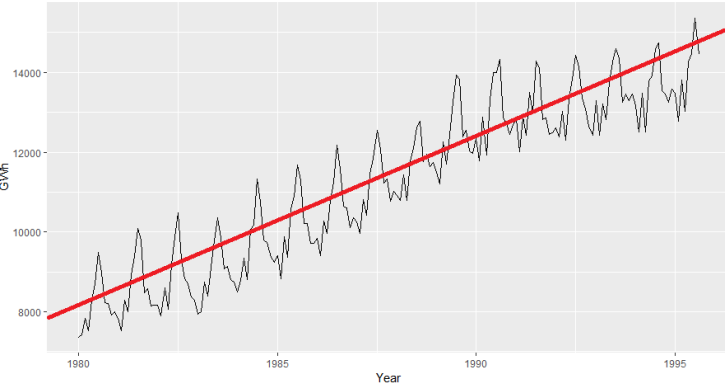
Augmented Dickey-Fuller test (ADF) & Kwiatkowski-Phillips-Schmidt-Shin test(KPSS)

2.Trend

Bir zaman serisinin uzun vadedeki artış ya da azalışının gösterdiği yapıya trend denir. Trend, zaman içinde ortalamanın yükseliş veya düşüş yönünü gösterir. Trend, zaman serileri analizinde önemlidir çünkü gelecekteki değerlerin tahmin edilmesinde veya verilerin anlaşılmasında etkilidir.

Trend iki şekilde görülebilir: Lineer Trend ve Non-Lineer Trend. Lineer Trend, verilerin doğrusal bir eğilim gösterdiği durumlarda görülür ve veriler arasındaki ilişkiyi doğrusal bir denklemle ifade eder. Non-Lineer Trend ise, verilerin doğrusal olmayan bir eğilim gösterdiği durumlarda görülür.

- *Bir zaman serisinde trend görülürse durağanlık çok düşük bir ihtimaldir.*



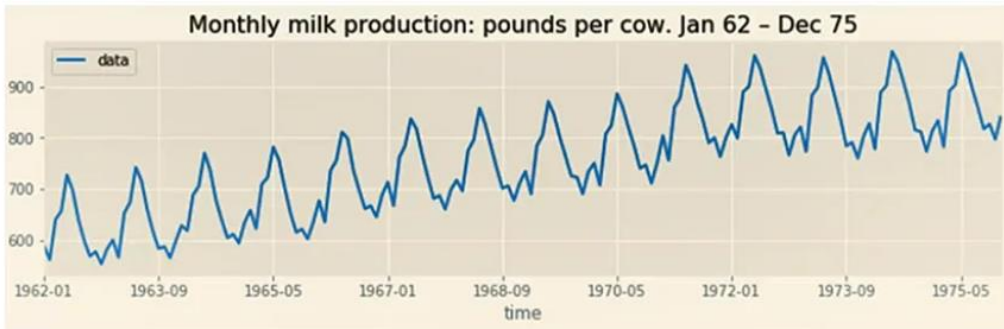
ARTAN TREND

AZALAN TREND



3.Mevsimsellik (Seasonality)

Zaman serisinin belirli bir davranışı belirli periyotlarla tekrar etmesi durumuna mevsimsellik denir. Örnek olarak Yaz, kurban bayramı gibi zamanlarda yazlık alanlarda ev kiralının uçuşması fakat Kış Sonbahar gibi zamanlarda ev fiyatlarında düşme durumu verilebilir.



Bu grafiđi yorumlayalım;

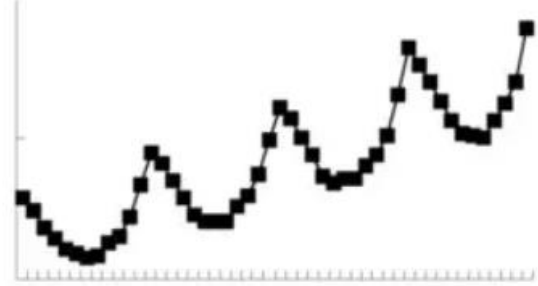
- İlk bakışta artan bir trende sahip olduğundan söz edebiliriz.
- Trend artmaya devam ederken iniş ve çıkış durumunun **kendi içinde düzenli bir şekilde tekrarlanması** bize mevsimsellik olduğuna dair ipucu veriyor.
- Bu serinin istatistiksel özellikleri zaman içinde değışkenlik gösterdiğinden durağanlık yoktur diyebiliriz.

Döngü (Cycle)

- Az önce de değindiğimiz gibi mevsimsellik tekrar eden düzen barındırır. Daha belirgin kısa vadeli ve düzenli aralıklarla gün, hafta, mevsim, yıl gibi zamanlarla örtüşecek şekilde ilerler.
- Döngüsellik daha uzun vadeli bir davranış sergiler. Tamamen mevsimsel olmamak ile birlikte kısmi mevsimsellik barındırır. Bir siyasetçinin konuşması, teknoloji devlerinin yeni teknolojiler duyurması, okulların açılması gibi olaylar döngüsellığı çok etkiler.



Cyclical Variation



Seasonal Variation

Zaman Serisi Analizinde Kullanılan Algoritmalar

Zaman serisi verilerini tahmin ederken amaç, gözlem dizisinin gelecekte nasıl devam edeceğini tahmin etmektir. Zaman serilerinde aşağıdaki belli başlı yöntemleri kullanıp, tahmin başarısı en yüksek olan yöntemle modellemeye gidilmektedir.

Zaman serisi analizinde smoothing (düzeltme) yöntemleri

Verilerdeki gürültüyü azaltmak, trendleri veya desenleri daha iyi ortaya çıkarmak için kullanılan istatistiksel tekniklerdir, Holt-Winters de denilir.

Hareketli ortalama (moving average):

Belirli bir zaman aralığındaki verilerin aritmetik ortalamasını alarak verilerdeki dalgalanmaları düzleştiren bir yöntemdir. Trendi gözlemlemek için ya da feature türetiminde kullanılabilir. Zaman serisi teorisine göre, zaman serisi **en çok** kendisinden bir önceki değerden etkilenir. Hareketli ortalama formülünden anlaşılacağı üzere gelecek değer, kendisinin k adet önceki değerinin ortalamasıdır.

$$\hat{y}_t = \frac{1}{k} \sum_{n=1}^k y_{t-n}$$

Ağırlıklı ortalama (weighted average):

Hareketli ortalamaya benzer olmakla beraber son verilere daha fazla ağırlık veren bir yöntemdir. Gerçek değerlere hareketli ortalamaya göre daha başarılı yaklaşır. Zaman serisinin trendi ve mevsimselliğinin olmaması durumunda tahmin için kullanılacak en iyi yöntem ağırlıklı ortalamadır.

$$\hat{y}_t = \sum_{n=1}^k \omega_n y_{t+1-n}$$

Single exponential smoothing (SES/Single HWES):

Basit Üstel Düzeltme Yöntemi olan SES, sadece durağan serilerde başarılıdır, trend ve mevsimselliği modelleyemez. Geçmiş değerlere daha fazla ağırlık vererek üstel düzeltme yapar. *Gelecek, yakın geçmişten daha çok etkilenir* varsayımıyla geçmişin etkileri ağırlıklandırılır.

$$\hat{y}_t = \alpha \cdot y_{t-1} + (1 - \alpha) \cdot \hat{y}_{t-1}$$

α ; 0 ile 1 arasında bir değer alan smoothing faktördür ve değere ne kadar ağırlık verileceğini belirler. Geçmiş gerçek değer (learning) ile geçmiş tahmin edilen değerlerin (remembering) üstel olarak ağırlıklandırılmasıyla tahmin yapılır.

Doble exponential smoothing (DES/Double HWES):

Çift Üstel Düzeltme yöntemi DES; SES yönteminin yaptığına ek olarak trend etkisini de göz önünde bulundurur ve üstel düzeltme yapar. Ancak mevsimselliği modelleyemez.

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$$

$$\hat{y}_{t+1} = \ell_t + b_t$$

İlk formül level'i, ikinci formül ise trendi ifade eder. α parametresi level'a ilişkin geçmiş değer ve tahmin edilen değeri ne kadar ağırlıklandıracağımızla ilgiliydi. DES, α parametresine ek olarak trend bileşenine ilişkin optimize edilmesi gereken parametre olan β 'yi da kullanır.

Triple exponential smoothing (TES/Triple HWES):

Winters Üstel Düzeltme Yöntemi, DES yönteminin yaptığına ek olarak mevsimselliği de modeller. En gelişmiş smoothing yöntemidir.

$$\ell_t = \alpha(y_t - s_{t-p}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_t) + (1 - \gamma)s_{t-p}$$

$$\hat{y}_{t+m} = \ell_t + mb_t + s_{t-p+1+(m-1)modp}$$

Level ve trend'e ek olarak 3.formül mevsimselliği ifade eder. Burada α , mevsimsellik bileşenine ilişkin optimize edilmesi gereken parametredir.

Bu yöntem ile dinamik olarak level, trend ve mevsimsellik etkilerini değerlendirerek tahmin yapılmasına olanak sağlar.

Zaman Serilerinde Tahmin Modelleri

Zaman serisi tahmininde kullanılan çeşitli istatistiksel, matematiksel, makine öğrenmesi ve derin öğrenme tabanlı yöntemler vardır. Bu yöntemler arasında **AR, MA, ARIMA, SARIMA, SARIMAX, Cointegration, XGBoost, LSTM** gibi birçok farklı yaklaşım bulunmaktadır.

Zaman Serileri Analizinde ACF ve PACF Grafikleri Nelerdir?

Zaman serisi analizinde, ACF ve PACF grafikleri, bir zaman serisinin temel yapısını tanımlamak için en önemli araçlardan ikisidir.

- ACF grafiği, bir zaman serisinin farklı gecikmelerde kendisiyle olan korelasyonunu gösterirken, PACF grafiği, önceki gecikmelerin etkilerini çıkardıktan sonra, bir zaman serisinin farklı gecikmelerde kendisiyle olan korelasyonunu gösterir. ACF ve PACF grafikleri, bir zaman serisinin temel yapısını tanımlamak için kullanılabilir. Aşağıdakiler, ACF ve PACF grafiklerini yorumlamak için bazı genel yönergelerdir:
- ACF grafiği ilk birkaç gecikmede ani artışlar gösteriyorsa, bir AR modeli uygun olabilir.
- PACF grafiği ilk birkaç gecikmede ani artışlar gösteriyorsa, bir MA modeli uygun olabilir.
- ACF ve PACF grafiklerinin her ikisi de ilk birkaç gecikmede ani artışlar gösteriyorsa, bir ARMA modeli uygun olabilir.

1. AR (AutoRegressive) Model

AR modeli, bir zaman serisinin **geçmiş gözlem değerlerine dayanarak** gelecekteki değerleri tahmin etmeye çalışan temel bir istatistiksel modeldir. AR modeli, serinin geçmiş değerlerine dayanarak bir trendi veya deseni modelleyebilir.

2. MA (Moving Average) Model

MA (Moving Average) modeli, geçmiş veri noktalarının **belirli bir zaman dilimindeki ortalamalarını** ve **hata terimlerinin ağırlıklı ortalamasını** kullanarak gelecekteki değerleri tahmin etmeye çalışan temel bir istatistiksel modeldir.

3. ARIMA (AutoRegressive Integrated Moving Average) Model

ARMA (AutoRegressive Moving Average) modelleri, AR ve MA modeli ile birlikte kullanılarak oluşturulan bir zaman serisi modelidir. Ancak ARMA modeli günümüzde **güncelliğini**

ytirmiştir. Onun yerine aynı işlemleri yanında getirdiği ek faydalarla birlikte daha güçlü bir şekilde uygulayabilen ARIMA (AutoRegressive Integrated Moving Average) modeli kullanılır.

Buradaki Integrated (I) bileşeni, zaman serilerinde görülen trend ve mevsimsellik gibi **durağan olmayan durumları ortadan kaldırmak** için kullanılan fark alma işlemini temsil eder.

4. SARIMA (Seasonal Autoregressive Integrated Moving Average) Modelleri

SARIMA (Seasonal Autoregressive Integrated Moving Average) modeli, zaman serileri analizi ve tahmini için kullanılan bir istatistiksel modeldir. SARIMA, hem AR hem de MA bileşenlerini içermenin yanında ayrıca **mevsimsel bir bileşen de** içerir.

Buradaki Seasonal (S) bileşeni, modelde kullanılan **mevsimsellik periyodunu** temsil eder.

5. SARIMAX (Seasonal Autoregressive Integrated Moving Average Exogenous) Modelleri

SARIMAX (Seasonal Autoregressive Integrated Moving Average Exogenous) modeli, hem otoresif (AR) hem hareketli ortalama (MA) hem de mevsimsel bileşenleri içermenin yanında ayrıca **dışsal bir faktörü temsil eden eksojen (exogenous) bileşeni de** içerir. Başka bir deyişle, oluşturduğumuz zaman serisi modelimize eklenen **dışsal faktör de tahmin etmeye çalıştığımız değer ile bir regresyon ilişkisi** kurar.

6. Cointegration (Eş-Bütünleşme) Modelleri

Cointegration (Eş-Bütünleşme) modeli, iki veya daha fazla zaman serisi arasındaki uzun vadeli ilişkiyi analiz eden ve bu serilerin **birlikte hareket etme eğilimini** inceleyen istatistiksel modellerdir. Cointegration, seriler arasında bir uzun vadeli denge durumu olduğunu öne sürer.

7. Makine Öğrenmesi Tabanlı Modeller (XGBoost)

Makine Öğrenmesi modelleri özellikle zaman serisi verileri için üretilmemiştir. Çünkü, zaman serisi verilerinde genellikle bağımsız değişkenler yoktur. Ancak **zaman serinize ilişkin özellikleri (ay, yıl, gün, mevsimsellik gibi) bağımsız değişkenlere dönüştürerek** bunları zaman serilerine uyarlamak oldukça kolaydır.

Zaman Serisi Analizleri İçin Kullanılan Python Kütüphaneleri

from statsmodels.tsa.stattools import **adfuller** : *Augmented Dickey-Fuller testi ile zaman serisinin durağan olup olmadığını test eder.*

from statsmodels.tsa.arima_model import **ARIMA** : *ARIMA modelini oluşturur ve fit eder.*

from statsmodels.tsa.arima_model import **ARIMAResults** : *ARIMA modelinin sonuçlarını saklar ve analiz eder.*

from statsmodels.tsa.seasonal import **seasonal_decompose** : *Zaman serisini trend, mevsimsel ve rastgele bileşenlerine ayırır.*

from statsmodels.graphics.tsaplots import **plot_acf** : *Otokorelasyon fonksiyon grafiği çizer.*

from statsmodels.graphics.tsaplots import **plot_pacf** : *Kısmi otokorelasyon fonksiyon grafiği çizer.*