

# DATA SCIENCE and BUSINESS INTELLIGENT

---

Author: Egi Safitri

Meeting 11



# Kenali Analisis Time Series

- Metode statistik menjadi salah satu bagian penting dari Data Science. Berbagai metode pengolahan data yang digunakan untuk mengolah big data dan machine learning, ternyata menggunakan dasar ilmu statistik. Metode statistik yang beragam akan cocok dengan berbagai jenis data yang ada dari big data. Dalam metode statistik ini, ada tiga jenis data yang bisa digunakan dalam analisis, yaitu data cross section, data time series, dan data panel (gabungan cross section dan time series).
- Data time series merupakan data yang biasanya hanya memiliki satu variabel dengan rentang waktu tertentu, misalnya data saham. Untuk mengolah data time series, tentu data analyst membutuhkan metode analisis yang tepat. Kita bisa menggunakan metode Analisis Time Series yang terbagi menjadi beberapa metode seperti MA, ARMA, dll.

# Pengertian Analisis Time Series

Analisis Time Series atau Analisis Runtun Waktu merupakan analisis yang digunakan untuk mengolah data time series. Dimana analisis ini akan melibatkan penggunaan data tersebut untuk membuat model yang akan digunakan sebagai dasar peramalan. Dengan analisis ini, kita dapat mengetahui apa yang menyebabkan terjadinya suatu trend, apakah karena sentimen tertentu atau karena hal lain. Hal-hal inilah yang bisa kita jadikan pertimbangan dalam pembuatan keputusan

## Uji Stationer Data Time Series

Uji stasioner termasuk bagian penting dalam time series. Untuk mengetahui apakah data yang dimiliki stasioner atau tidak, kita bisa melihatnya dengan tiga cara, yaitu dengan mengamati grafik, melihat correlogram, serta uji unit root yang bisa menggunakan ADF-Test dan Phillips-Perron Test. Data time series akan stasioner jika nilai mean dan variansinya konstan atau tidak mengalami perubahan sepanjang waktu secara sistematis.

## Apa yang Harus Dilakukan Jika Tidak Stasioner?

Data time series yang tidak stasioner merupakan salah satu masalah besar. Jika terjadi ketidak stasioneran data, kemungkinan disebabkan karena adanya Spurious regression, dimana regresi memiliki nilai r-squared yang tinggi namun tidak memiliki hubungan yang cukup berarti. Untuk mengatasi data yang tidak stasioner di dalam model, tentu kita harus membuatnya stasioner terlebih dahulu dengan cara melakukan diferensiasi pada derajat  $d$ .

## Model yang Time Series yang Bisa Digunakan

Sama halnya dengan regresi, analisis time series juga memiliki berbagai jenis metode yang bisa digunakan. Pemilihan metode tentu harus disesuaikan dengan data yang dimiliki serta tujuan analisis, karena meskipun sama-sama merupakan data time series, ada keadaan tertentu yang membuat kita bisa menggunakan metode yang berbeda-beda. Model time series yang bisa dijadikan pilihan adalah Autoregressive (AR), Moving Average (MA), ARIMA, dan SARIMA.

# Komponen Data Time Series

- Teknik analisis data menjadi bagian terpenting untuk mengubah data mentah menjadi sumber informasi. Hingga saat ini, ada banyak sekali jenis teknik analisis data yang bisa digunakan. Tentunya setiap teknik analisis data hanya bisa digunakan untuk data dengan kriteria tertentu saja.
- Misalkan untuk melihat pengaruh antar variabel dan untuk meramalkan keadaan selanjutnya, sementara data yang digunakan hanya ada satu waktu pengambilan namun memiliki banyak variabel, maka bisa menggunakan analisis regresi.

# 1. Trend Data

- Trend data menunjukkan pergerakan data time series dalam jangka waktu yang lama. Nantinya kita bisa melihat apakah terjadi kenaikan atau penurunan pada trend nya. Dari trend data ini, kita akan bisa melihat pola, biasanya pola ini akan terus berulang.
- Sayangnya, kita memerlukan waktu yang cukup lama agar pola tersebut bisa terlihat, bisa sampai 15 atau 20 tahun. Pada dasarnya, trend data ini tidak selalu linier atau lurus, sehingga kita bisa menggunakan perhitungan matematika untuk melihat trend datanya.



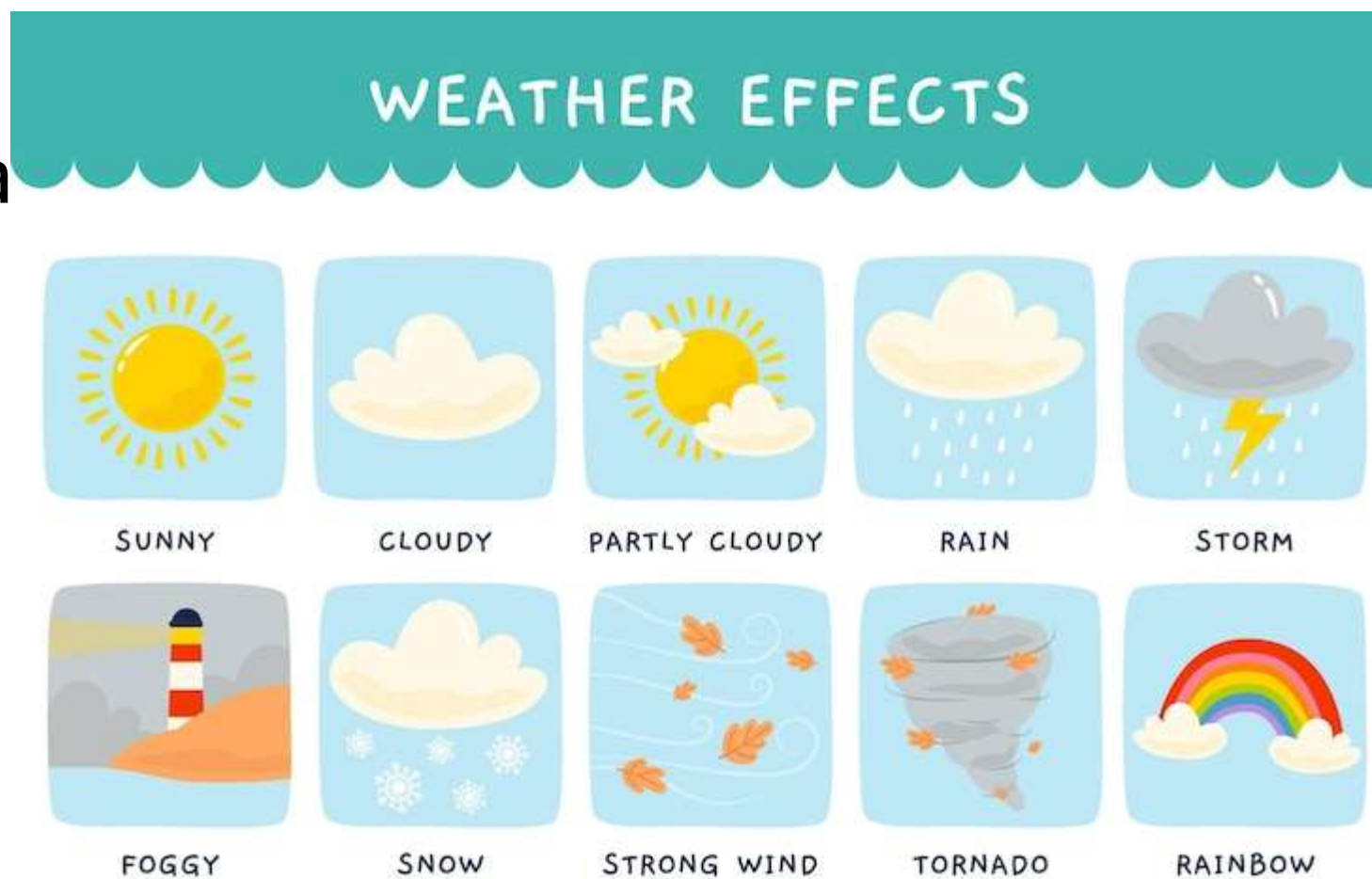
## 2. Siklus

- Jika dibandingkan dengan trend, siklus ini memiliki jangka waktu yang lebih pendek, sekitar 2 sampai dengan 10 tahun. Pada dasarnya, siklus ini merupakan fluktuasi bisnis. Lama dan besarnya fluktuasi tentu akan berbeda pada masing-masing perusahaan dan juga industri.
- Untuk melakukan analisis siklus, kita bisa melihat dengan menghitung persentase trend:  $\% \text{ Trend} = Y/Y_t \times 100$
- Dimana Y merupakan data aktual tahunan (data sesungguhnya), sementara  $Y_t$  merupakan data trend yang dihitung berdasarkan persamaan trend.



## 3. Musiman

- Musiman merupakan fluktuasi data yang terjadi namun dalam rentang waktu yang lebih singkat dibandingkan trend dan juga siklus, dimana hanya akan membutuhkan waktu dalam lingkup satu tahun. Ada beberapa faktor yang membuat fluktuasi musiman ini terjadi, misalnya karena peristiwa tertentu (lebaran, natal, tahun baru, dll) atau karena cuaca (musim hujan dan musim kemarau).
- Analisis musiman dilakukan dapat dilakukan untuk beberapa hal, seperti mengatur alokasi anggaran penjualan dalam kuartalan yang mungkin akan dipengaruhi oleh musim.



## 4. Ketidakteraturan

- Komponen data time series yang selanjutnya adalah ketidakteraturan. Fluktuasi jenis biasanya terjadi dalam waktu jangka pendek, dan tidak begitu jelas polanya. Faktor-faktor yang menyebabkan fluktuasi jenis ini juga munculnya tidak teratur, misalkan tiba-tiba ada masalah dalam perusahaan yang tidak bisa diprediksi seperti kebakaran atau ada masalah pada pimpinan yang diketahui publik. Tentu hal ini akan sangat mempengaruhi keuntungan dari perusahaan.



# Jenis data menurut waktu

1. Cross-section Data : Terdiri dari beberapa objek data pada satu waktu tertentu. Misalnya data penduduk dan pendapatan perkapita tingkat kabupaten pada tahun 2021.
2. Time-series Data : Merupakan data yang terdiri atas satu objek tetapi meliputi beberapa periode waktu yaitu harian, bulanan, mingguan, tahunan, dll. Misalnya data jumlah penduduk kabupaten A pada tiga tahun terakhir.
3. Panel Data : Merupakan data yang menggabungkan antara data time-series dan data cross-section. Sehingga data panel akan memiliki beberapa objek dan beberapa periode waktu.

# Jenis Data

Jenis Data	Objek	Waktu
Cross-section Data	Banyak	Satu
Time-series Data	Satu	Banyak
Panel Data	Banyak	Banyak

# Konsep Penting

1. **Stasioneritas:** Analisis deret waktu biasanya dimulai dengan memeriksa stasioneritas data. Stasioneritas berarti bahwa statistik dasar seperti rata-rata, varians, dan kovarians tidak berubah seiring waktu. Jika data tidak stasioner, langkah-langkah seperti diferensiasi dapat digunakan untuk membuatnya stasioner.
2. **Pola dan Tren:** Analisis deret waktu mencari pola dan tren dalam data. Pola adalah fluktuasi teratur yang dapat berulang dalam data, sedangkan tren adalah perubahan panjang waktu yang menunjukkan peningkatan atau penurunan secara keseluruhan.

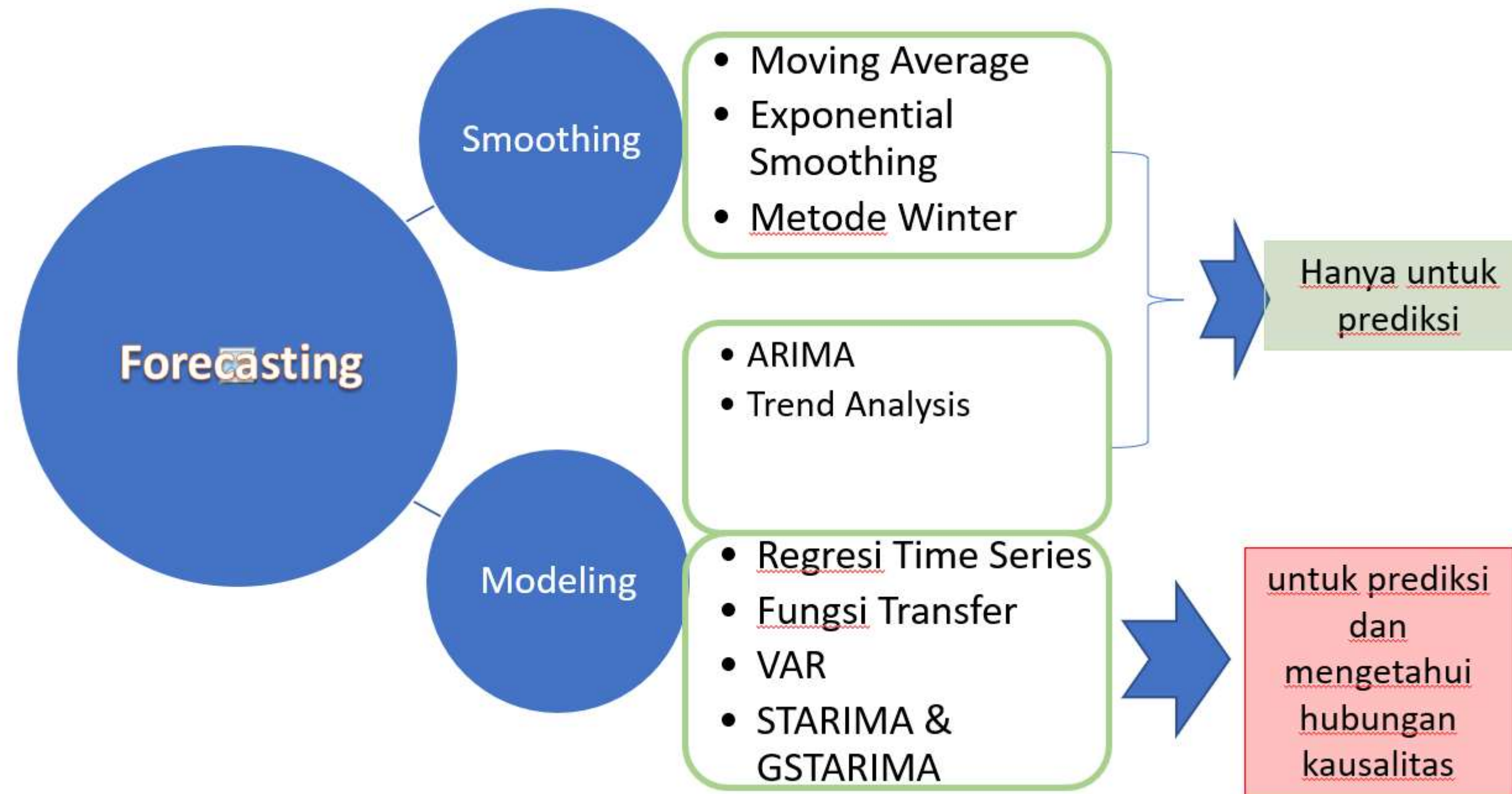
# Konsep Penting

3. **Komponen Deret Waktu:** Deret waktu dapat dibagi menjadi tiga komponen utama: komponen musiman (seasonal), komponen tren (trend), dan komponen residual (error). Komponen musiman adalah fluktuasi berulang dalam data yang terkait dengan faktor-faktor musiman seperti musim panas dan musim dingin. Komponen tren adalah perubahan panjang waktu dalam data. Komponen residual adalah bagian dari data yang tidak dapat dijelaskan oleh musim atau tren.
4. **Metode Peramalan (*Forecasting*):** Salah satu tujuan utama analisis deret waktu adalah melakukan peramalan, yaitu memprediksi nilai-nilai di masa depan berdasarkan pola dan tren dalam data historis. Metode peramalan melibatkan penggunaan model statistik atau teknik matematika seperti Moving Average, Exponential Smoothing, atau model ARIMA (AutoRegressive Integrated Moving Average).

## Konsep Penting

5. **Evaluasi Model:** Setelah membangun model peramalan, perlu dievaluasi untuk memastikan kinerjanya yang baik. Metrik evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE) digunakan untuk mengukur sejauh mana model cocok dengan data sebenarnya.
6. **Analisis Spektral:** Analisis deret waktu juga dapat mencakup analisis spektral, yang digunakan untuk mengidentifikasi frekuensi dominan atau siklus dalam data deret waktu. Ini sering digunakan dalam aplikasi seperti analisis gelombang dan pengolahan sinyal.
7. **Kompleksitas Model:** Dalam analisis deret waktu, perlu mempertimbangkan kompleksitas model. Model yang terlalu sederhana mungkin tidak mampu menangkap pola yang kompleks, sedangkan model yang terlalu rumit dapat mengalami overfitting.

# Metode Peramalan



# Metode Smoothing

**Metode smoothing (pemulusan)** dalam analisis deret waktu adalah teknik yang digunakan untuk menghaluskan fluktuasi acak atau variabilitas dalam data deret waktu, sehingga pola dan tren yang mendasarinya menjadi lebih jelas. Tujuannya adalah membuat data yang lebih mudah diinterpretasikan dan digunakan untuk peramalan atau analisis lainnya. Berikut adalah beberapa metode smoothing umum dalam analisis deret waktu:

# Metode Smoothing

1. **Simple Moving Average (SMA):** SMA adalah metode pemulusan yang paling dasar. Ini melibatkan perhitungan rata-rata dari nilai-nilai dalam jangka waktu tertentu. Contoh termudah adalah rata-rata bulanan penjualan selama 12 bulan untuk menghaluskan fluktuasi musiman. SMA memberikan bobot yang sama pada semua data dalam jangka waktu yang dipilih.
2. **Exponential Smoothing (Pemulusan Eksponensial):** Metode ini memberikan bobot yang berbeda pada nilai dalam deret waktu, dengan bobot yang semakin berkurang seiring berjalannya waktu. Pemulusan eksponensial berguna ketika nilai yang lebih baru dianggap lebih penting daripada yang lebih lama. Ada beberapa variasi dari metode ini, termasuk pemulusan eksponensial sederhana (SES), Holt-Winters (yang mempertimbangkan komponen musiman dan tren), dan metode eksponensial beradaptasi.

# Metode Smoothing

- 3. Double Exponential Smoothing (Holt's Linear Exponential Smoothing):**  
Metode ini adalah ekstensi dari pemulusan eksponensial sederhana yang juga mempertimbangkan tren dalam deret waktu. Ini berguna ketika data memiliki tren yang signifikan yang harus diperhitungkan selain fluktuasi musiman.
- 4. Triple Exponential Smoothing (Holt-Winters Exponential Smoothing):**  
Metode ini adalah ekstensi dari double exponential smoothing yang juga memasukkan komponen musiman. Ini cocok untuk data deret waktu yang memiliki pola musiman yang jelas selain tren.

# Metode Smoothing

5. **Metode Keseluruhan (Holistic Methods):** Pendekatan ini mencoba untuk mengkombinasikan beberapa metode smoothing, seperti SMA dan metode eksponensial, untuk memanfaatkan kekuatan masing-masing metode dalam mengatasi berbagai jenis fluktuasi dalam data.
6. **Metode Jendela (Window Methods):** Metode ini melibatkan pemilihan jangka waktu tertentu (jendela) dan mengambil nilai rata-rata atau bobot lainnya di dalam jendela tersebut. Metode ini digunakan untuk menghaluskan fluktuasi harian atau mingguan dalam data.

## Metode Smoothing

7. **Lowess (Locally Weighted Scatterplot Smoothing):** Metode ini merupakan pendekatan nonparametrik yang menghaluskan data dengan cara memberikan bobot berdasarkan jarak relatif antara titik data dalam deret waktu. Ini cocok untuk data yang memiliki fluktuasi yang tidak teratur.

# Moving Average

Metode Moving Average (Rata-rata Bergerak) adalah salah satu metode smoothing yang umum digunakan dalam analisis deret waktu. Metode ini digunakan untuk menghaluskan fluktuasi acak dalam data deret waktu dan membantu dalam mengidentifikasi tren atau pola yang mendasarinya.

# Simple Moving Average

- SMA adalah metode pemulusan yang paling sederhana. Ini melibatkan penghitungan rata-rata dari sejumlah nilai dalam deret waktu yang sama.
- Contoh sederhana dari SMA adalah jika kita ingin menghaluskan data penjualan bulanan selama setahun, kita dapat menggunakan SMA 12-bulan, di mana kita menjumlahkan penjualan selama 12 bulan terakhir dan kemudian membaginya dengan 12.
- Cocok untuk pola data konstan/stasioner.

# Simple Moving Average

- Prinsip dasar:
  - Data smoothing pada periode ke- $t$  merupakan rata-rata dari  $m$  buah data dari data periode ke- $t$  hingga ke- $(t-m+1)$

$$S_t = \frac{1}{m} \sum_{i=t-m+1}^t X_i$$

- Data smoothing pada periode ke- $t$  berperan sebagai nilai forecasting pada periode ke- $t+1$ .

$$F_t = S_{t-1}$$

$$F_{n,h} = S_n$$

- MA dengan  $m$  yang lebih besar akan menghasilkan pola data yang lebih halus.

# Double Moving Average

Double Moving Average (DMA), juga dikenal sebagai Moving Average Crossover Method, adalah metode peramalan yang melibatkan penggunaan dua moving average dengan periode yang berbeda untuk menghasilkan sinyal peramalan. Metode ini umumnya digunakan dalam analisis deret waktu untuk mengidentifikasi perubahan tren dalam data.

- Proses penghalusan dengan rata-rata dilakukan dua kali.

- Tahap I:

$$S_{1,t} = \frac{1}{m} \sum_{i=t-m+1}^t X_i$$

- Tahap II:

$$S_{2,t} = \frac{1}{m} \sum_{i=t-m+1}^t S_{1,i}$$

# Double Moving Average

Double Moving Average (DMA), juga dikenal sebagai Moving Average Crossover Method, adalah metode peramalan yang melibatkan penggunaan dua moving average dengan periode yang berbeda untuk menghasilkan sinyal peramalan. Metode ini umumnya digunakan dalam analisis deret waktu untuk mengidentifikasi perubahan tren dalam data.

- Proses penghalusan dengan rata-rata dilakukan dua kali.

- Tahap I:

$$S_{1,t} = \frac{1}{m} \sum_{i=t-m+1}^t X_i$$

- Tahap II:

$$S_{2,t} = \frac{1}{m} \sum_{i=t-m+1}^t S_{1,i}$$

- Forecasting dilakukan dengan formula:

$$F_{2,t,t+h} = A_t + B_t(h)$$

dengan

$$A_t = 2S_{1,t} - S_{2,t}$$

$$B_t = \frac{2}{m-1}(S_{1,t} - S_{2,t})$$

# Kelebihan Moving Average

Metode Moving Average memiliki beberapa keuntungan, seperti:

- **Penghalusan Data:** Metode ini membantu mengurangi fluktuasi acak dalam data, sehingga memudahkan pengamatan pola yang mendasari seperti tren atau musiman.
- **Ketahanan terhadap Outlier:** lebih tahan terhadap outlier daripada beberapa metode lainnya karena mereka mengambil rata-rata sejumlah nilai dalam jangka waktu tertentu.
- **Sederhana dan Mudah Dimengerti:** Metode ini relatif mudah dimengerti dan diimplementasikan.

# Kekurangan Moving Average

Namun, metode Moving Average juga memiliki beberapa kelemahan, termasuk:

- **Lag:** Salah satu kelemahan utama dari SMA adalah bahwa ia menghasilkan lag dalam perkiraan. Ini berarti bahwa prediksi SMA selalu tertinggal dibandingkan dengan tren aktual dalam data.
- **Kurang Responsif terhadap Perubahan:** SMA tidak responsif terhadap perubahan yang cepat dalam data karena mendasarkan prediksi pada nilai rata-rata dalam jangka waktu tertentu.
- Salah satu kelemahan utama DMA adalah bahwa ia dapat menghasilkan sinyal palsu ketika tren datar atau dalam kondisi pasar yang bergejolak. Selain itu, DMA cenderung memberikan sinyal dengan lag, yang berarti sinyal perubahan tren dapat terlambat.

# Exponential Smoothing

- Exponential Smoothing (Pemulusan Eksponensial) adalah metode peramalan yang digunakan dalam analisis deret waktu untuk menghaluskan fluktuasi data dan meramalkan nilai masa depan berdasarkan pola-pola dalam data historis. Metode ini memberikan bobot yang berkurang eksponensial seiring waktu kepada nilai dalam deret waktu, dengan memberikan lebih banyak bobot pada data yang lebih baru. Exponential Smoothing berguna untuk memperkirakan nilai masa depan berdasarkan perkiraan sebelumnya dan memberikan penekanan yang lebih besar pada data yang paling baru.

# Exponential Smoothing

Berikut adalah komponen utama dalam metode Exponential Smoothing:

## 1. Level (Tingkat):

- Tingkat adalah perkiraan rata-rata atau nilai tengah dari data deret waktu. Ini adalah representasi dari "tingkat" yang mendasari data. Dalam metode Exponential Smoothing, tingkat diperbarui pada setiap periode waktu berdasarkan peramalan sebelumnya.

## 2. Faktor Penghalusan (Smoothing Factor atau Alpha):

- Faktor penghalusan ( $\alpha$ , alpha) adalah angka antara 0 dan 1 yang menentukan sejauh mana bobot diberikan kepada data yang baru dibandingkan dengan data sebelumnya. Semakin besar nilai alpha, semakin besar bobotnya diberikan kepada data yang lebih baru.

## 3. Periode (Time Period atau P):

- Periode adalah jangka waktu antara dua pengamatan atau periode waktu di dalam data deret waktu. Periode ini digunakan untuk menghitung faktor penghalusan (alpha) yang optimal.

# Accuracy Measures

Beberapa ukuran yang dapat digunakan untuk mengevaluasi seberapa baik metode forecasting data:

**Mean Absolute Deviation (MAD)**

$$MAD = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{X}_t|$$

**Mean Squared Deviation (MSD)**

$$MSD = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2$$

**Mean Absolute Percentage Error (MAPE)**

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100$$

