



Time Series Data Analysis and Forecasting

Muhammad Rasyid Ridha



9 November 2020

Agenda

- **Konsep time series dan forecasting**
 - Pengertian time series dan forecasting
 - Tujuan forecasting
 - Contoh data time series
 - Pola data time series
- **Metode forecasting**
 - Tahapan membuat model forecasting
 - Metode tradisional (statistika)
 - Machine learning
 - Kelemahan forecasting
- **Use case**
 - Menggunakan data dari Google Trend
- **Tanya jawab**

Konsep Time Series dan Forecasting



Pengertian Time Series dan Forecasting

- **Time series**
 - Data yang berurutan berdasarkan waktu yang biasanya dapat diperoleh dalam interval waktu yang sama (jam, harian, mingguan, bulanan dan sebagainya)
- **Forecasting**
 - Proses membuat model berdasarkan nilai di masa lampau untuk memprediksi nilai masa depan



Tujuan Forecasting

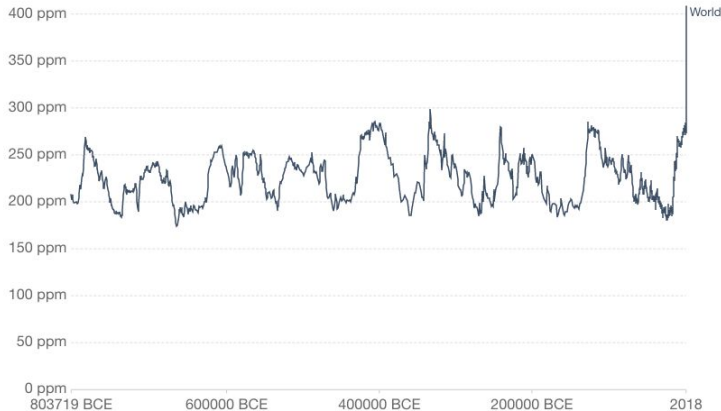
- Strategi dan Perencanaan Bisnis
- Deteksi dan Mitigasi Bencana
- Supply Chain
- Membuat data sintetis sebagai kontrol untuk kuasi-eksperimen

Contoh Data Time Series

- Data perekonomian
- Data permintaan (*demand*)
- Data penjualan (*supply*)
- Data terkait alam (curah hujan, suhu)

Atmospheric CO₂ concentration

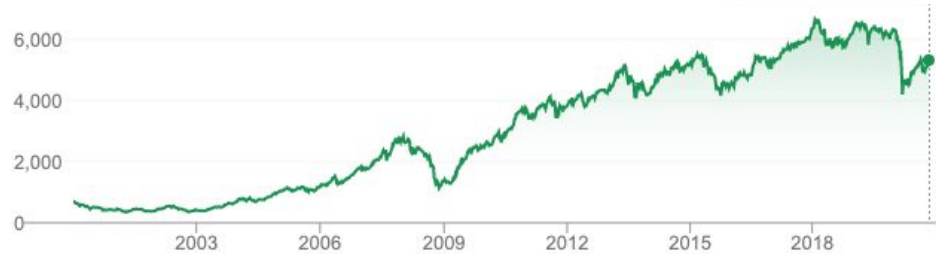
Global average long-term atmospheric concentration of carbon dioxide (CO₂), measured in parts per million (ppm). Long-term trends in CO₂ concentrations can be measured at high-resolution using preserved air samples from ice cores.



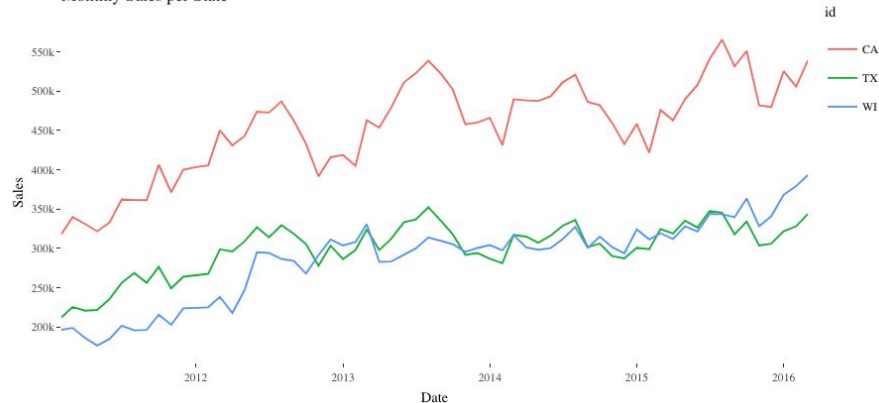
Source: EPICA Dome C CO₂ record (2015) & NOAA (2018)

OurWorldinData.org/co2-and-other-greenhouse-gas-emissions • CC BY

Indeks harga saham gabungan (IHSG)



Monthly Sales per State

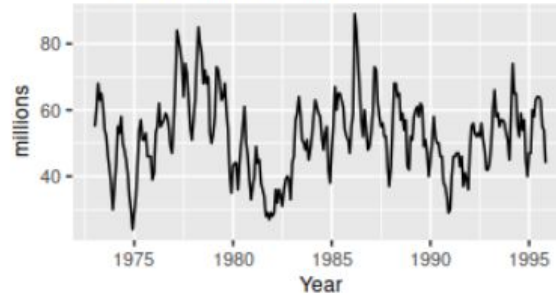


Pola Data Time Series

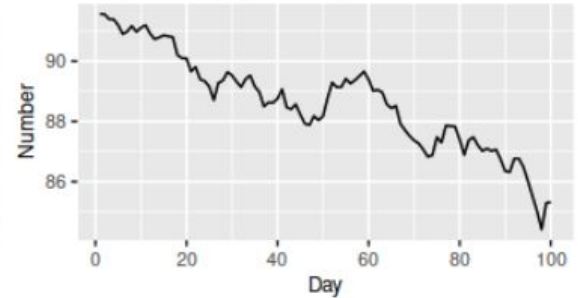
- Tren
- Musiman
- Siklik
- *Point change*



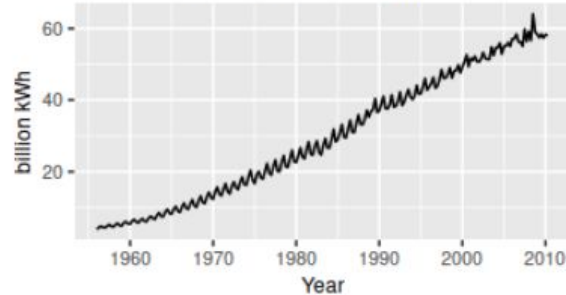
Sales of new one-family houses, USA



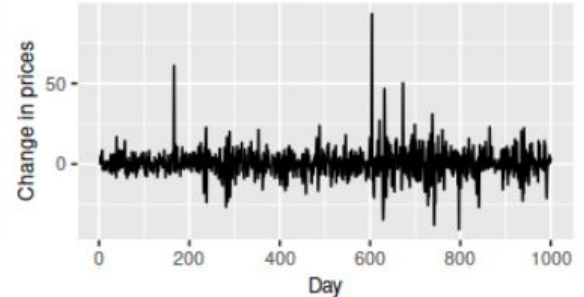
US treasury bill contracts



Australian quarterly electricity production

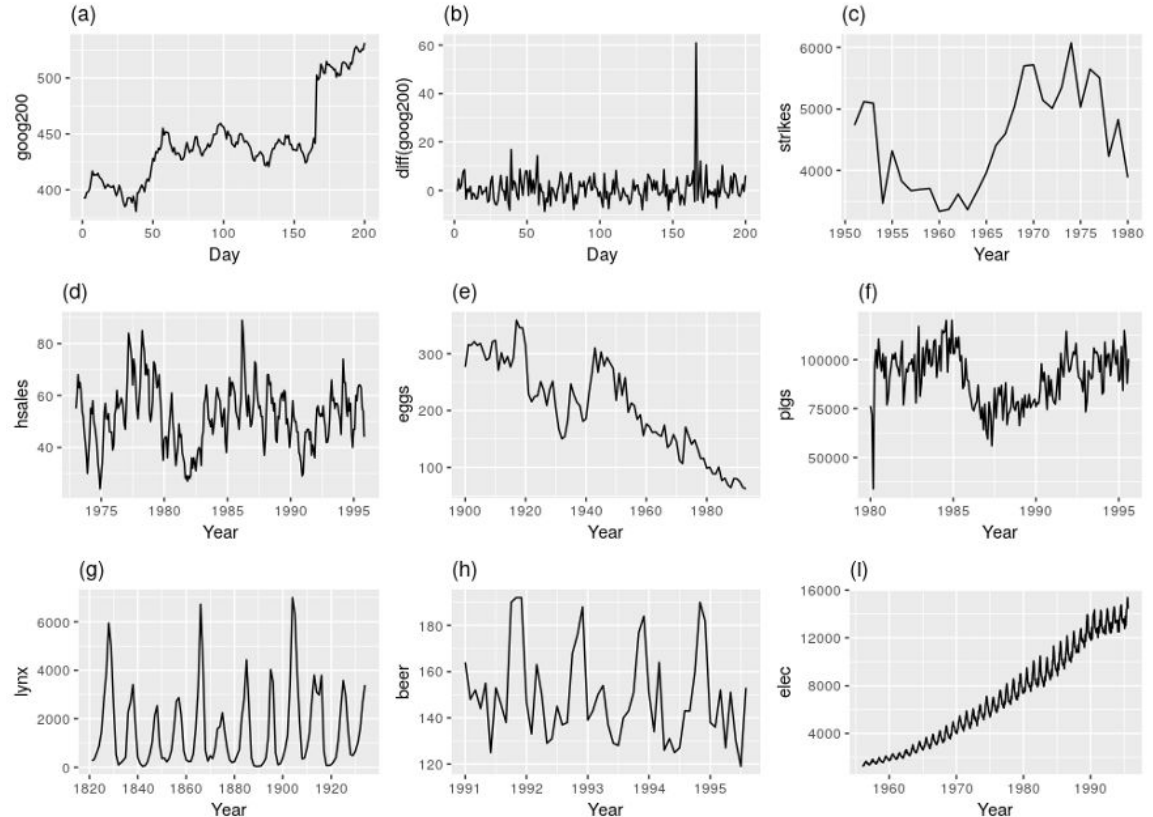


Google daily changes in closing stock price



Pola Data Time Series (Konsep Stasioneritas)

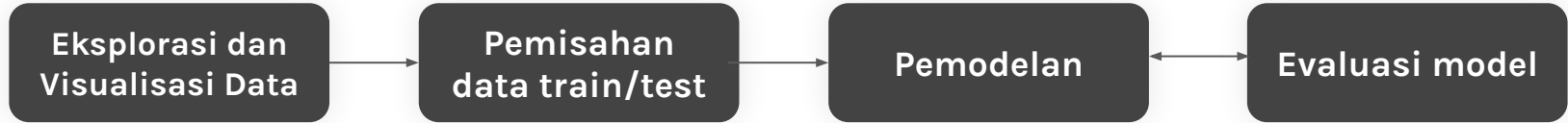
- Stasioner → rata-rata dan varians data tidak tergantung waktu
- Tren dan musiman → tidak stasioner
- Siklik → bisa stasioner, bisa tidak stasioner
- (d),(h),(i) musiman
- (a),(c),(e),(f),(i) tren
- (b) dan (g) stasioner



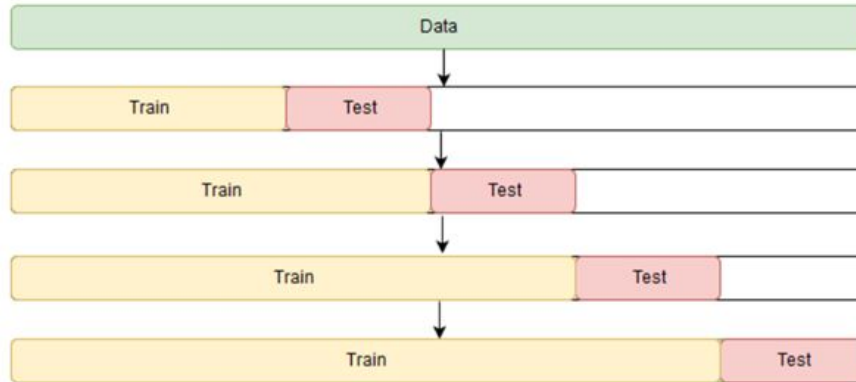
Metode Forecasting



Tahapan Pengembangan Model Forecasting



Time series cross validation



Time series error metrics

- RMSE
- MAE
- MAPE
- sMAPE
- MASE
- RMSSE

Metode Tradisional (Statistika)

- Model forecasting sederhana
 - Metode rata-rata
 - Metode Naive
 - Metode Naive musiman
- Exponential Smoothing (ETS)
- ARIMA

Model forecasting sederhana

- Metode rata-rata

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T)/T$$

- Metode Naive

$$\hat{y}_{T+h|T} = y_T$$

- Metode Naive musiman

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}$$

- Model paling sederhana seperti ini bisa dicoba pertama kali sebagai *benchmark* awal
- Untuk kasus tertentu, model yang lebih kompleks bisa lebih buruk dibandingkan model di atas

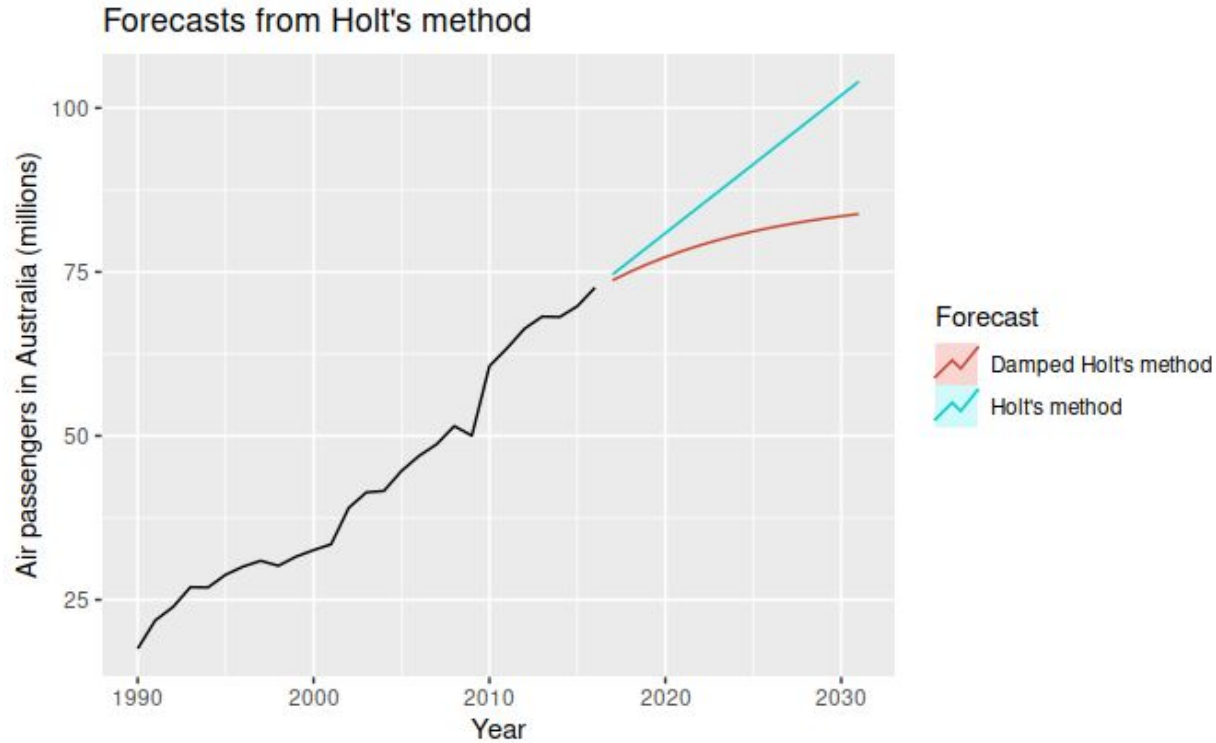
Exponential Smoothing

Komponen Tren	Komponen Musiman		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
Ad (Additive damped)	(Ad,N)	(Ad,A)	(Ad,M)

Parameter smoothing: *alpha* (level), *beta* (tren), *gamma* (musiman), *phi* (damped)

Notasi	Metode	Parameter				
		Alpha	Beta	Gamma	Phi	Seasonality (m)
(N,N)	Simple exponential smoothing	✓				
(A,N)	Holt's linear	✓	✓			
(Ad,N)	Additive damped trend	✓	✓		✓	
(A,A)	Additive Holt-Winters'	✓	✓	✓		✓
(A,M)	Multiplicative Holt-Winters'	✓	✓	✓		✓
(Ad,M)	Holt-Winters' damped	✓	✓	✓	✓	✓

Damped trend vs trend



ARIMA

- **AR(p) autoregressive model**
 - $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$
 - Regresi menggunakan data masa lalunya sendiri sebagai input
 - Ordo p ditentukan berdasarkan plot PACF (*partial autocorrelation function*)
- **I(d) integrated**
 - $y_t = y_{t-1} + \varepsilon_t$
 - Apabila data deret waktu tidak stasioner (ada tren dan musiman) perlu dilakukan *differencing* (*detrending*) hingga data deret waktu menjadi stasioner
 - Ordo d ditentukan berdasarkan berapa kali *differencing* dilakukan
- **MA(q) moving-average model**
 - $y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$
 - Regresi menggunakan eror di masa lampau sebagai input
 - Ordo q ditentukan berdasarkan plot ACF (*autocorrelation function*)
- Model musiman memiliki ordo tambahan
 - ARIMA(p, d, q)(P, D, Q)[m]

Machine Learning

- Model pohon
 - XGBoost
 - LightGBM
- *Deep learning*
 - RNN
 - LSTM
 - GRU

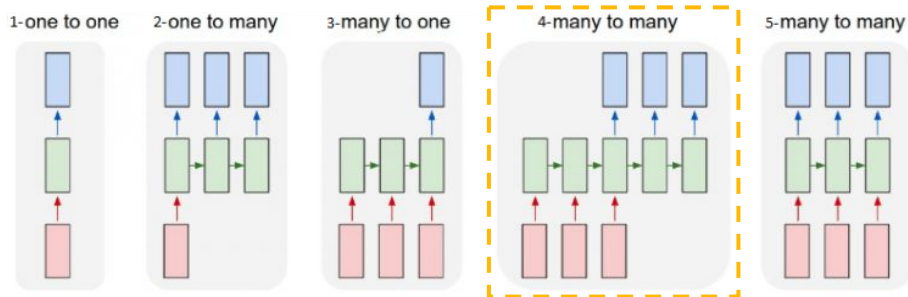
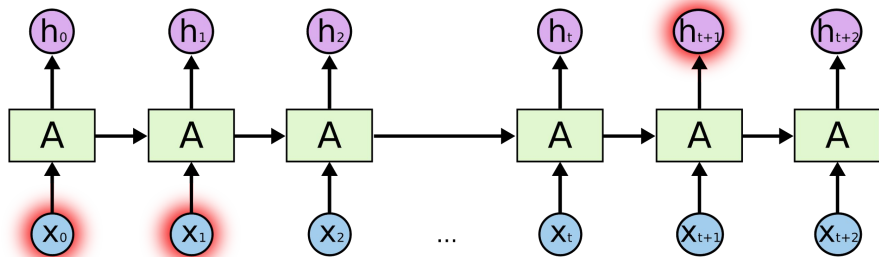
Model Pohon (XGBoost dan LightGBM)



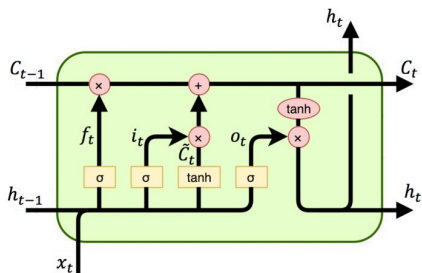
- Model nonlinear
- Gabungan dari banyak model **decision tree** (*weak learner*)
- Menggunakan algorithm **boosting** (model dilatih secara iteratif)
- Cepat dan lebih akurat dibandingkan satu model **decision tree**
- Flexibel (bisa menggunakan variabel kategorik sebagai input)
- Model pohon lainnya (Random Forest, Extra Trees, Catboost)

Deep Learning

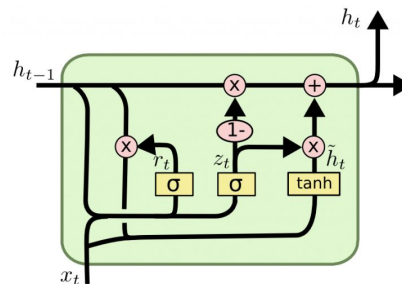
Recurrent Neural Network (RNN)



Long Short Term Memory (LSTM)



Gated Recurrent Unit (GRU)



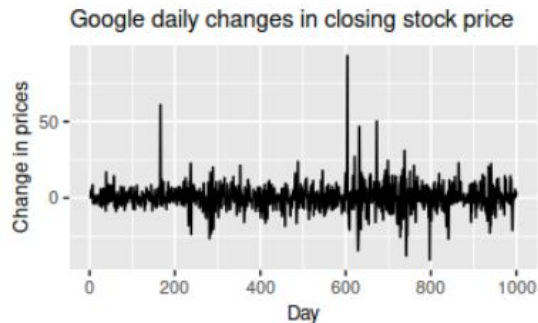
Referensi

<http://dprogrammer.org/rnn-lstm-gru>

<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Kelemahan Forecasting

- **Hasil prediksi tidak akurat** karena tidak ada pola sama sekali dalam deret waktu seperti musiman, tren maupun siklik
- **Hasil prediksi tidak akurat** karena hanya tergantung dari data masa lampau maupun data tertentu saja yang dimasukkan ke dalam model
- **Akurasi prediksi yang didapatkan tidak stabil dari waktu ke waktu** karena adanya faktor eksternal yang tidak ditangkap di dalam model (faktor musiman yang sifatnya tidak tentu)



Use Case



Use Case

- Mengambil data dari Google Trend <https://trends.google.co.id/trends/?geo=ID>
- Menggunakan R untuk membuat model
 - Memisahkan data menjadi dua bagian yaitu *train* dan *test*
 - Membuat model Naive menggunakan fungsi *naive*
 - Membuat model ARIMA menggunakan fungsi *auto.arima*
 - Membuat model ETS menggunakan fungsi *ets*

Terima Kasih



Muhammad Rasyid Ridha

rasyidstat@gmail.com

rasyidridha.com

[linkedin.com/in/rasyidridha](https://www.linkedin.com/in/rasyidridha)