



INSTITUT TEKNOLOGI SUMATERA

WORKSHOP DEEP LEARNING • SESI 2

RNN, LSTM & GRU

Model Bermemori untuk Data Deret Waktu: Studi Kasus Prediksi Trafik Seluler

Martin C.T. Manullang

Workshop Deep Learning untuk Mahasiswa Teknik Telekomunikasi — Institut Teknologi Sumatera
Semester Ganjil 2026/2027

mctm.web.id/dl-workshop

Capaian Pembelajaran Sesi Ini

Apa yang bisa kamu lakukan setelah 105 menit ini?

Setelah sesi ini, peserta mampu:

1. Menjelaskan mengapa **data berurutan** butuh model **bermemori**.
2. Memahami **intuisi** sel RNN serta **gerbang** pada LSTM dan GRU.
3. Menyiapkan data deret waktu dengan **windowing** dan **baseline naif**.
4. **Membandingkan** ANN, RNN, LSTM, dan GRU secara adil untuk prediksi trafik seluler.

Menit	Agenda
0–10	Motivasi: trafik seluler
10–30	RNN & masalah ingatan
30–45	LSTM & GRU
45–60	Data trafik Milan
60–75	Hasil perbandingan
75–85	Uji ingatan
85–105	Hands-on Notebook

Peta Materi Hari Ini

Dari ANN di Sesi 1 menuju model bermemori

- Motivasi
- Recurrent Neural Network
- LSTM dan GRU
- Studi Kasus: Trafik Milan
- Hasil Perbandingan
- Uji Ingatan
- Penutup

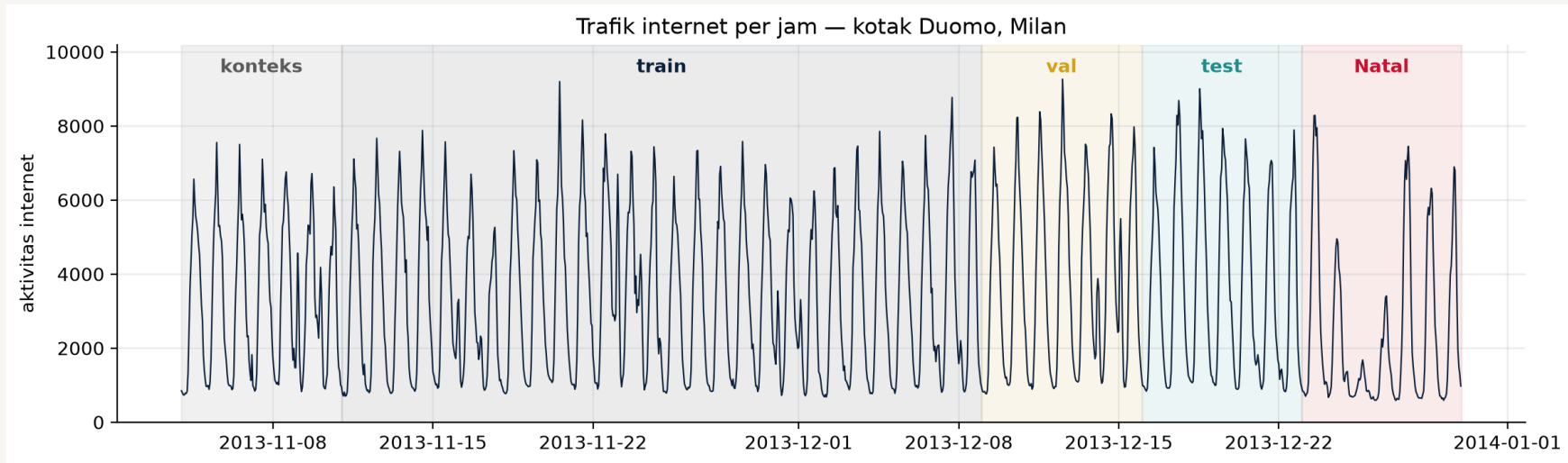


BAGIAN 1

Motivasi

Trafik Seluler Itu Berdenyut

Delapan minggu trafik internet di satu sel pusat kota Milan



1 Kapasitas

Menambah kanal/*carrier* sebelum jam sibuk.

2 Hemat energi

Mematikan sebagian sel (*sleep mode*) saat sepi.

3 Deteksi anomali

Trafik jauh dari ramalan = ada yang tidak beres.

Urutan menyimpan informasi

Sesi 1 (NSL-KDD): setiap koneksi berdiri sendiri.
Urutan baris tidak penting, boleh diacak.

Sesi 2 (trafik): nilai jam ini **bergantung** pada jam-jam sebelumnya.

- Pukul 07.00 trafik naik → pukul 08.00 hampir pasti lebih tinggi.
- Senin mirip Senin minggu lalu.
- Mengacak urutan = membuang informasi.

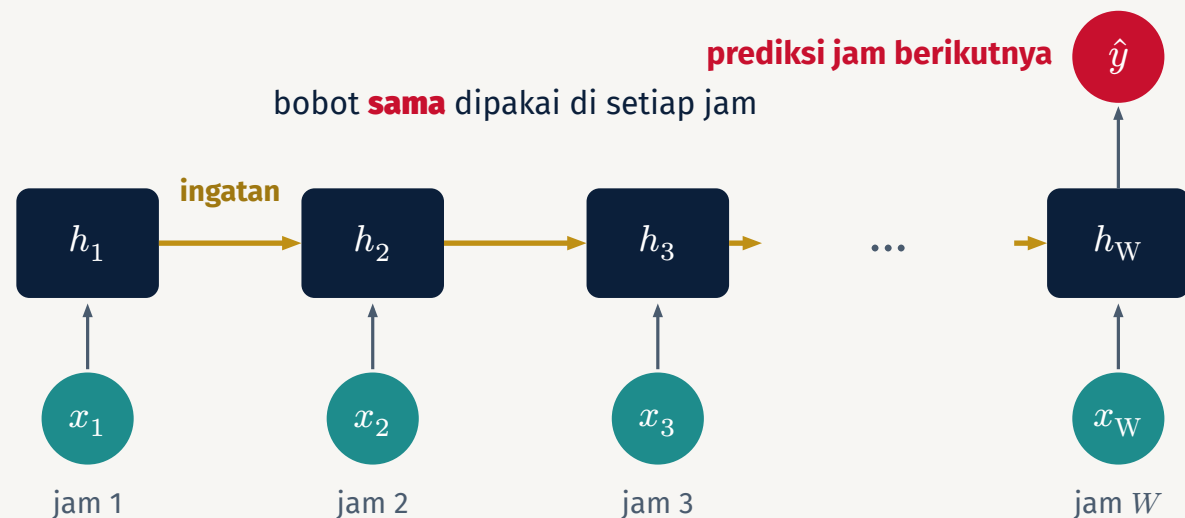
Kalau tetap memakai ANN?

1. Jendela 24 jam $\times$ 7 fitur **diratakan** menjadi 168 angka: ANN tidak tahu mana yang “kemarin” dan mana yang “barusan”.
2. Jendela diperpanjang → **parameter membengkak** (24 jam: 6.5 ribu, 168 jam: 38.8 ribu).
3. Panjang input harus **tetap**.

BAGIAN 2

Recurrent Neural Network

Membaca jam demi jam sambil membawa “catatan”



- Setiap jam, sel menerima **input jam ini** x_t dan **catatan sebelumnya** h_{t-1} .
- Hasilnya **catatan baru** h_t (*hidden state*), diteruskan ke jam berikutnya.
- Catatan terakhir h_W merangkum seluruh jendela → dipakai untuk prediksi.
- **Bobotnya sama** di setiap jam → parameter **tidak** bertambah walau jendela diperpanjang.

Rumusnya mirip neuron di Sesi 1

Perbarui catatan

$$h_t = \tanh(W_x x_t + W_h h_{t-1} + b)$$

Prediksi di akhir jendela

$$\hat{y} = W_y h_W + c$$

- $W_x x_t$: informasi **baru** dari jam ini.
- $W_h h_{t-1}$: informasi **lama** dari catatan.
- $\tanh$: menjaga catatan tetap di rentang -1 s.d. 1 .

Analogi

Seperti membaca laporan trafik per jam sambil memperbarui **satu kertas ringkasan**. Kertasnya terbatas, jadi yang lama perlahan tertimpa.

Vanishing gradient sepanjang waktu

Saat *backpropagation*, gradien dari jam terakhir harus **mengalir mundur** melewati setiap jam. Di setiap langkah, gradien dikali faktor yang biasanya **< 1** .

Mundur sejauh	Sisa sinyal (0.9^k)
1 jam	0.9
24 jam	0.08
100 jam	0.00003
168 jam	0.00000002

Akibatnya

RNN biasa **sulit belajar** hubungan dengan kejadian yang **jauh di masa lalu**. Ia praktis hanya “ingat” beberapa puluh langkah terakhir.

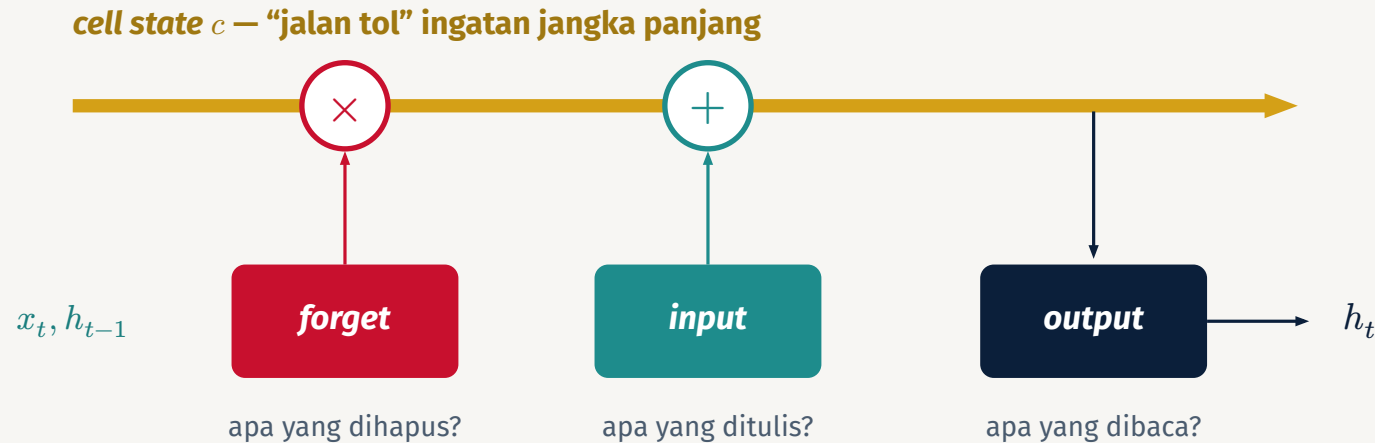
Analogi: pesan berantai

Bisikkan pesan melewati 100 orang. Di ujung, pesan aslinya hampir hilang.

BAGIAN 3

LSTM dan GRU

Long Short-Term Memory (Hochreiter & Schmidhuber, 1997)



Selain h_t , LSTM membawa **cell state** c_t . Informasi di jalur ini hanya **dikali gerbang** dan **ditambah**, bukan diremas berulang kali oleh $\tanh$. Karena itu gradien bisa mengalir jauh ke belakang tanpa memudar.

Setiap gerbang adalah “keran” bernilai 0–1

Gerbang	Tugas	Contoh trafik
Forget	Menghapus ingatan yang tidak relevan lagi	Hari berganti → lupakan pola kemarin sore
Input	Menulis informasi baru yang penting	Lonjakan pagi ini → simpan
Output	Memilih bagian ingatan yang dipakai sekarang	Ambil “pola jam sibuk” saat mendekati siang

Intuisi kunci

Gerbang bernilai **0** = tutup, **1** = buka penuh. Nilainya **dipelajari** dari data, sama seperti bobot di Sesi 1.

Bila *forget gate* ≈ 1 , ingatan dibawa **utuh** dari langkah ke langkah.

Gated Recurrent Unit (Cho dkk., 2014)

- Hanya **dua gerbang**:
 - ▶ **Update**: seberapa banyak catatan lama dipertahankan vs diganti.
 - ▶ **Reset**: seberapa banyak catatan lama dipakai saat membuat kandidat baru.
- Tidak ada *cell state* terpisah: h_t sekaligus menjadi ingatan.
- Lebih sedikit parameter, sering **sama baiknya** dengan LSTM.

	RNN	LSTM	GRU
Gerbang	0	3	2
Blok bobot	1×	4×	3×
Ingatan	h	$h + c$	h
Hidden kita	68	33	38
Parameter	5.3 rb	5.6 rb	5.4 rb

Hidden dipilih agar parameter ketiganya **setara**.

nn.RNN, nn.LSTM, dan nn.GRU punya antarmuka yang sama

```
class Rekuren(nn.Module):
    def __init__(self, jenis, n_fitur, hidden):
        super().__init__()
        sel = {"RNN": nn.RNN, "LSTM": nn.LSTM,
              "GRU": nn.GRU}[jenis]
        self.rnn = sel(n_fitur, hidden,
                      batch_first=True)
        self.head = nn.Linear(hidden, 1)

    def forward(self, x):          # [batch, W, 7]
        out, _ = self.rnn(x)      # [batch, W, hidden]
        return self.head(out[:, -1]) # jam terakhir
```

- Input berbentuk **[batch, waktu, fitur]**.
- `out[:, -1]`: *hidden state* pada jam terakhir.
- head: lapisan linear biasa (seperti Sesi 1) → 1 angka.

Loss & training

Regresi → loss **MSE**. Optimizer Adam, *mini-batch*, *early stopping*: sama seperti Sesi 1.

BAGIAN 4

Studi Kasus: Trafik Milan

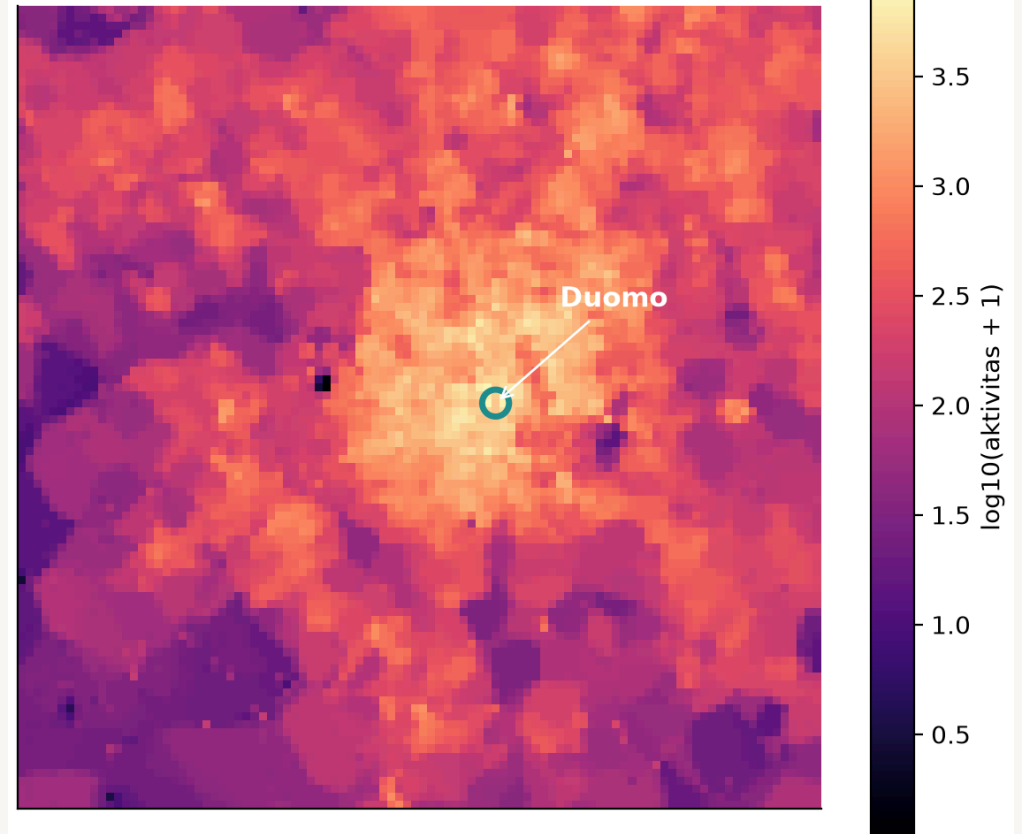
Big Data Challenge 2014 (Barlacchi dkk., 2015)

- Kota Milan dibagi **grid 100 × 100** kotak, masing-masing $\pm$ **235 m × 235 m**.
- Aktivitas **internet**, **telepon**, dan **SMS** per kotak, 1 Nov – 31 Des 2013.
- Versi per jam (308 MB) dari Zhang dkk. (2021): **8 minggu penuh**, 4 Nov – 29 Des.
- Kita pakai **satu kotak**: sekitar **Katedral Duomo** (pusat kota, kotak ke-95 tersibuk).

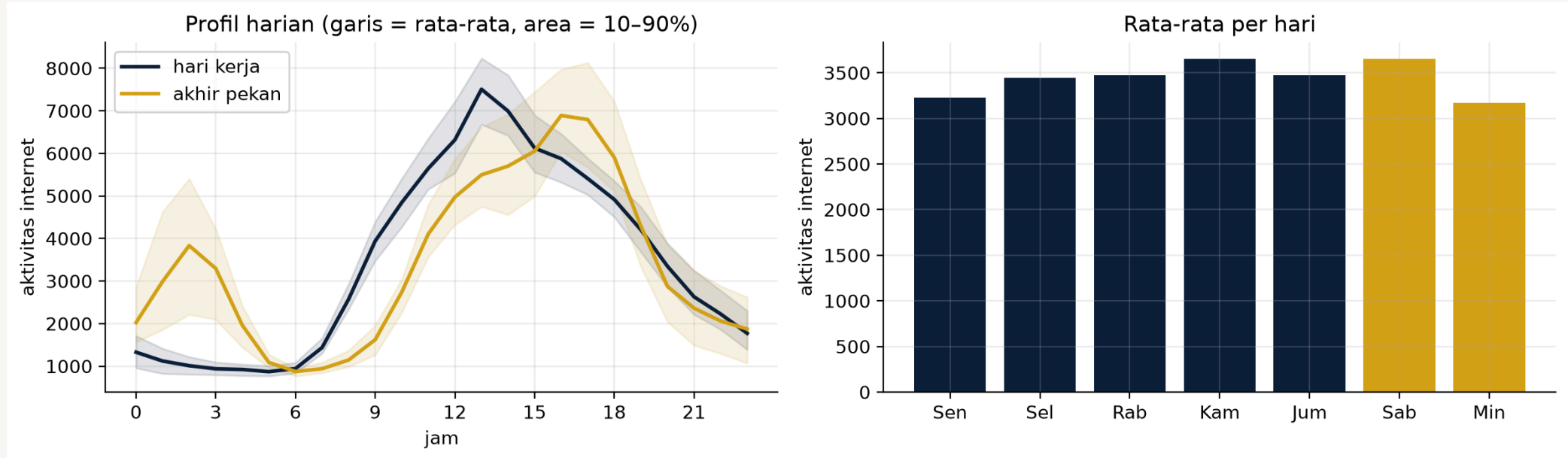
Catatan satuan

Nilai berupa **“aktivitas”** yang dinormalisasi operator, **bukan** Mbps. Cocok untuk pola, bukan untuk dibaca sebagai kecepatan.

Rata-rata aktivitas internet per kotak (skala log)



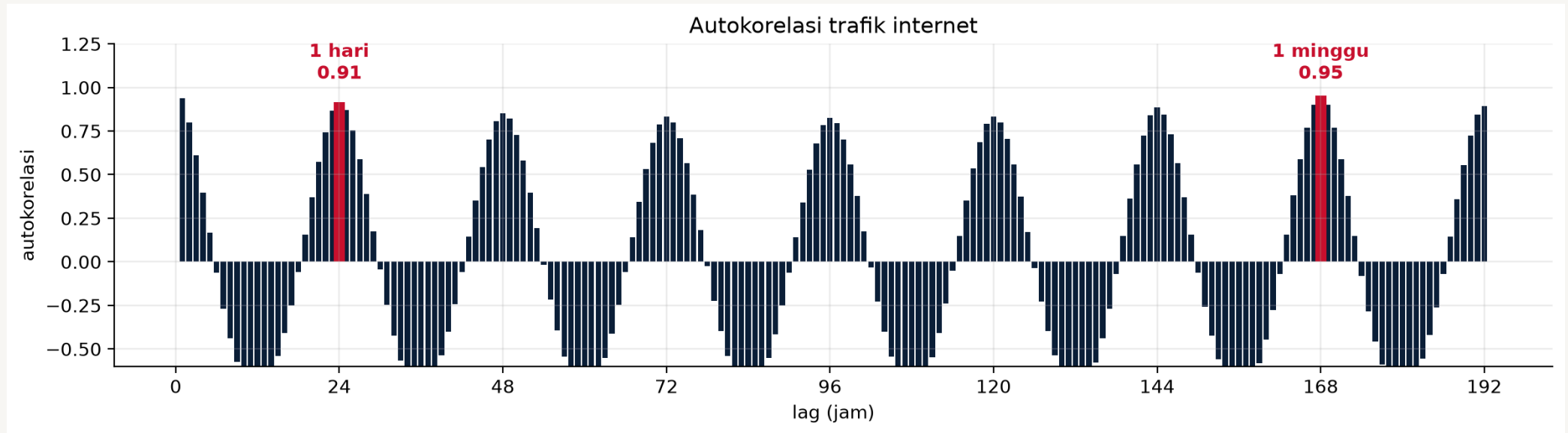
Sepi menjelang subuh, puncak siang hingga sore



Hari kerja dan akhir pekan berbeda sedikit. Area 10–90% menunjukkan variasi antar-hari: polanya **kuat tetapi tidak identik**.

Berapa Jauh Perlu Mengingat?

Autokorelasi: kemiripan dengan k jam sebelumnya



Puncak di **24 jam** (0.91) dan **168 jam** (0.95): trafik sekarang sangat mirip trafik **kemarin** dan **minggu lalu** pada jam yang sama. Maka kita coba dua jendela: **$W = 24$** dan **$W = 168$** jam.

Jendela geser sepanjang deret waktu

```
jam:  0 1 2 ... 23 | 24  ← target
      └─ X (W=24) ─┘   y
jam:   1 2 ... 24 | 25
      └─ X ─┘         y   ... dst.
```

7 fitur per jam

net, call, sms, jam (sin, cos), hari (sin, cos)

Bentuk tensor

X: **[672, W, 7]** y: **[672, 1]**

Kenapa sin & cos?

Jam 23 dan jam 0 bertetangga, tetapi angka 23 dan 0 berjauhan. Sebagai titik pada lingkaran, keduanya berdekatan.

$$\text{jam}_{\sin} = \sin\left(2\pi \cdot \frac{\text{jam}}{24}\right)$$

Latih dengan masa lalu, uji pada masa depan — jangan diacak

Bagian	Tanggal (2013)	Peran
Konteks	4 – 10 Nov	Hanya sebagai input jendela 168 jam
Train	11 Nov – 8 Des	Melatih bobot (4 minggu, 672 sampel)
Validasi	9 – 15 Des	<i>Early stopping</i>
Test	16 – 22 Des	Minggu normal — angka utama
Natal	23 – 29 Des	Uji tambahan: pola berubah drastis

Perbandingan yang adil

Semua model memakai **data, fitur, split, dan normalisasi yang sama**. Model rekuren diberi **parameter setara** (± 5.5 ribu). Setiap konfigurasi dilatih dengan **3 seed** lalu dirata-rata.

Apakah tebakan sederhana sudah cukup?

Tebakan	MAE	MAPE
Nilai jam sebelumnya	601	18.4%
Jam yang sama kemarin	552	23.8%
Jam yang sama minggu lalu	468	19.9%

Minggu test, prediksi 1 jam ke depan.

■ **MAE**: rata-rata |error| (satuan aktivitas).

■ **MAPE**: rata-rata error dalam **persen**.

Aturan emas

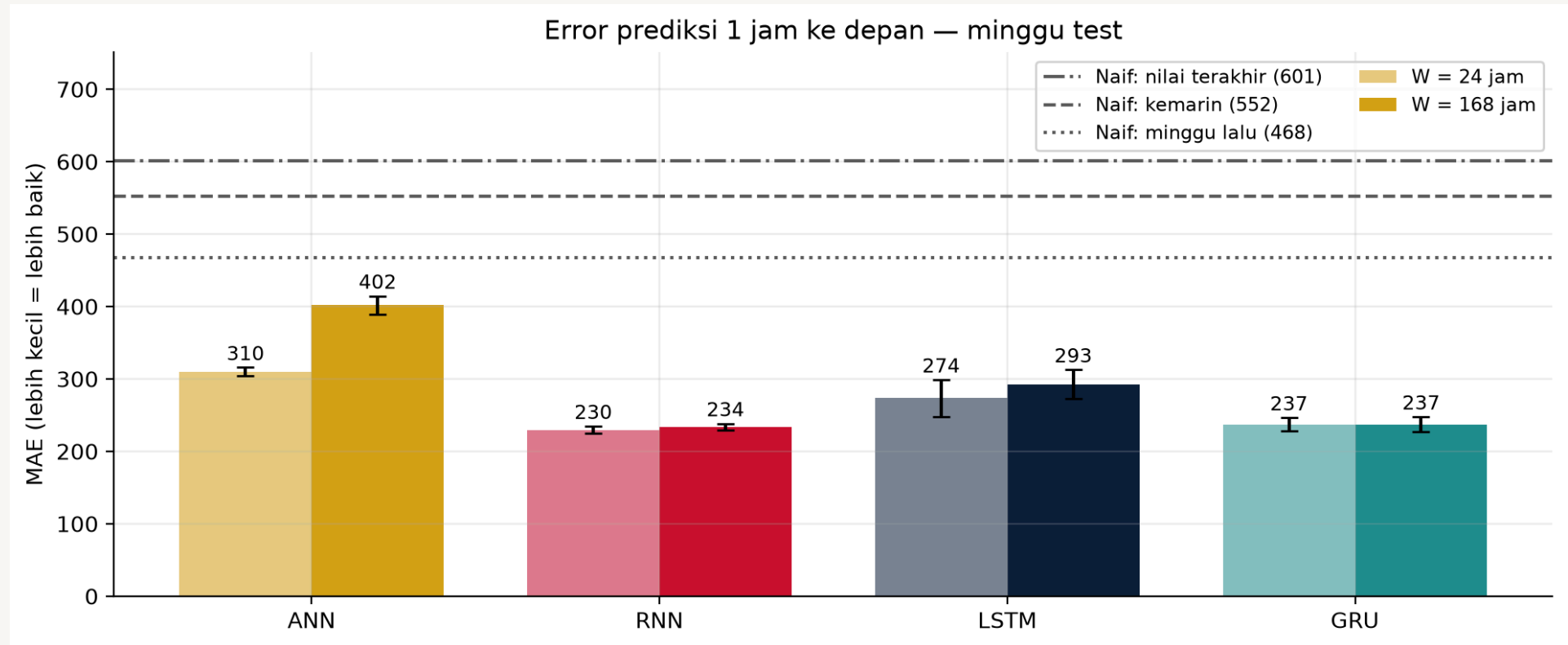
Model *deep learning* yang **tidak** mengalahkan baseline naif **tidak berguna**, sebagus apa pun arsitekturnya.

BAGIAN 5

Hasil Perbandingan

Prediksi 1 Jam ke Depan

MAE minggu test, rata-rata 3 seed (GPU RTX 5070)



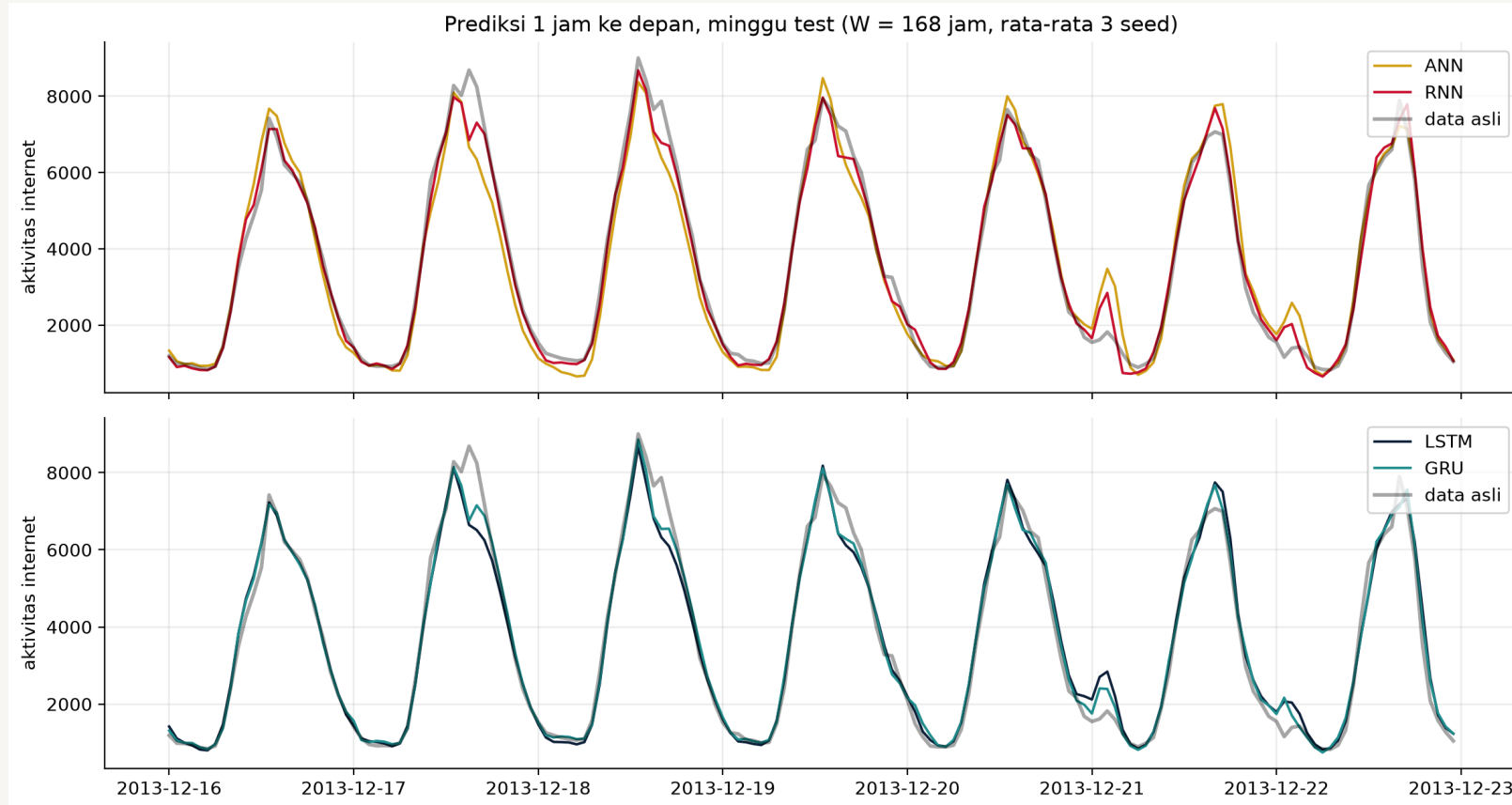
Semua model jauh mengalahkan baseline terbaik (468). MAPE turun dari $\pm 20\%$ menjadi **8–11%**.

Bandingkan dengan jujur, termasuk variasi antar-seed

W = 24	seed 0	seed 1	seed 2	MAPE
ANN	311	315	304	11.3%
RNN	225	235	229	8.5%
LSTM	250	300	271	9.1%
GRU	243	226	242	8.0%

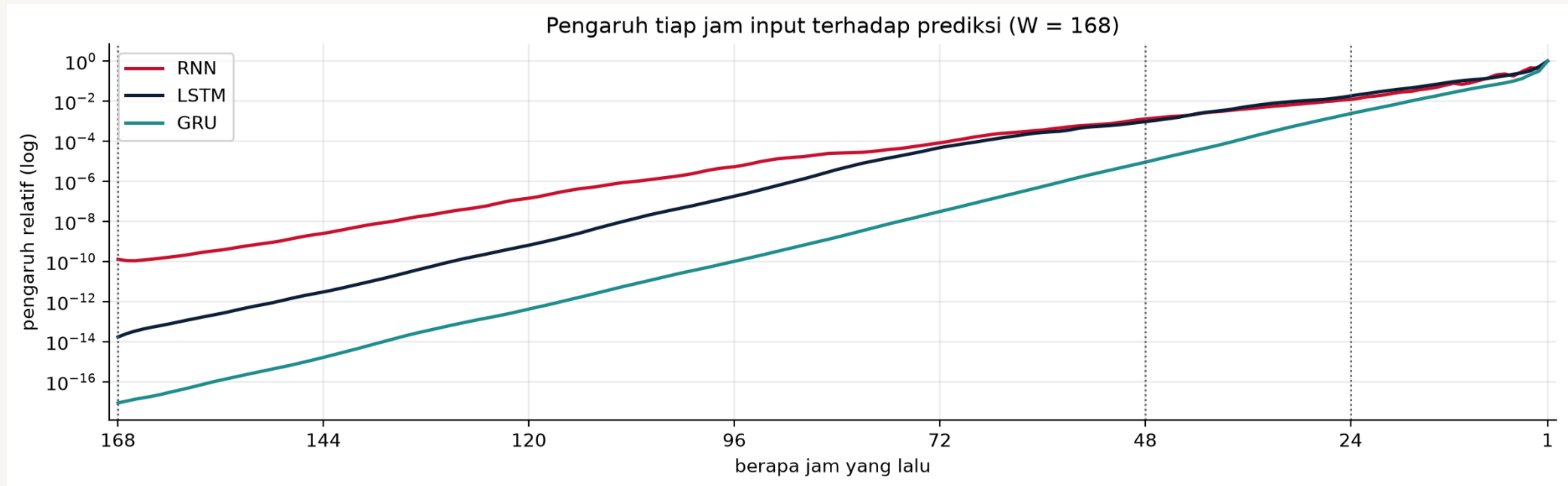
1. **Model rekuren > ANN**: setiap seed RNN/GRU lebih baik dari setiap seed ANN ($\pm 25\%$).
2. **RNN $\approx$ GRU**: selisihnya lebih kecil dari variasi antar-seed; menurut MAPE urutannya malah terbalik.
3. **LSTM** sedikit di bawah, dengan variasi besar.
4. **W = 168 tidak membantu**. ANN malah memburuk (38.8 ribu parameter, hanya 672 sampel $\rightarrow$ *overfitting*).

Minggu test 16–22 Des, $W = 168$, rata-rata 3 seed



Seberapa Jauh Model Melihat?

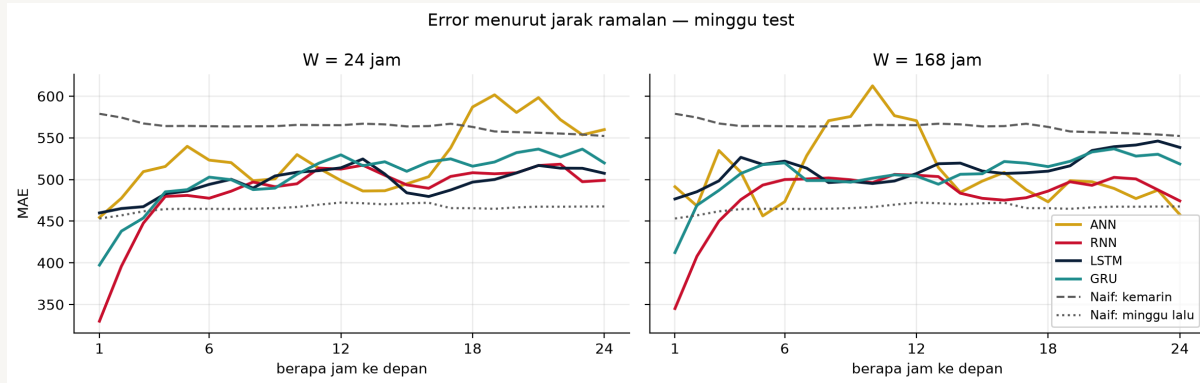
Pengaruh tiap jam input terhadap prediksi: $|\partial \hat{y} / \partial x_t|$



Pengaruh 24 jam lalu tinggal $\pm 1\%$ dan 168 jam lalu ≈ 0 . **Ketiga model hanya memakai ± 1 hari terakhir.** Untuk 1 jam ke depan, beberapa jam terakhir + fitur jam & hari sudah cukup, jadi gerbang LSTM/GRU tidak memberi keuntungan.

Ramalan 24 Jam ke Depan

Seluruh hari esok sekaligus (24 output)



W = 168	MAE	MAPE
Naif: minggu lalu	466	19.1%
RNN	481	16.8%
GRU	506	18.0%
LSTM	514	17.7%
ANN	510	20.3%

Model hanya unggul untuk **± 3 jam pertama**. Setelah itu, “jam yang sama minggu lalu” **lebih baik** menurut MAE.

Baseline sederhana bisa sangat sulit dikalahkan

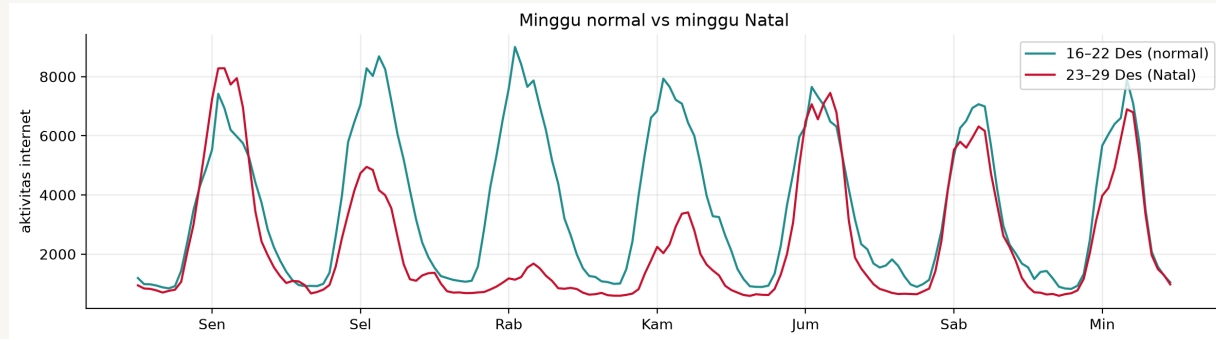
- Pola mingguan Duomo **sangat kuat** (autokorelasi 0.95).
- Data latih hanya **4 minggu**: model belum cukup contoh untuk mengalahkan “salin minggu lalu”.
- Menurut MAPE model unggul (lebih tepat saat trafik rendah di malam hari), tetapi **meleset di puncak siang**.

Untuk praktisi

- Selalu laporkan baseline naif.
- Pilih metrik sesuai kebutuhan: MAE untuk kapasitas puncak, MAPE untuk kesalahan relatif.
- Ide perbaikan: latih model memprediksi **selisih** terhadap minggu lalu (lihat Latihan).

Minggu Natal: Pola Berubah

Model dilatih dengan minggu normal, lalu diuji saat libur



1 jam ke depan	MAE
Naif: minggu lalu	1426
Naif: jam sebelumnya	431
RNN (W = 24)	421
GRU (W = 24)	468
ANN (W = 24)	680

“Salin minggu lalu” **runtuh**. Model yang membaca jam-jam terakhir masih bisa **menyesuaikan diri**.

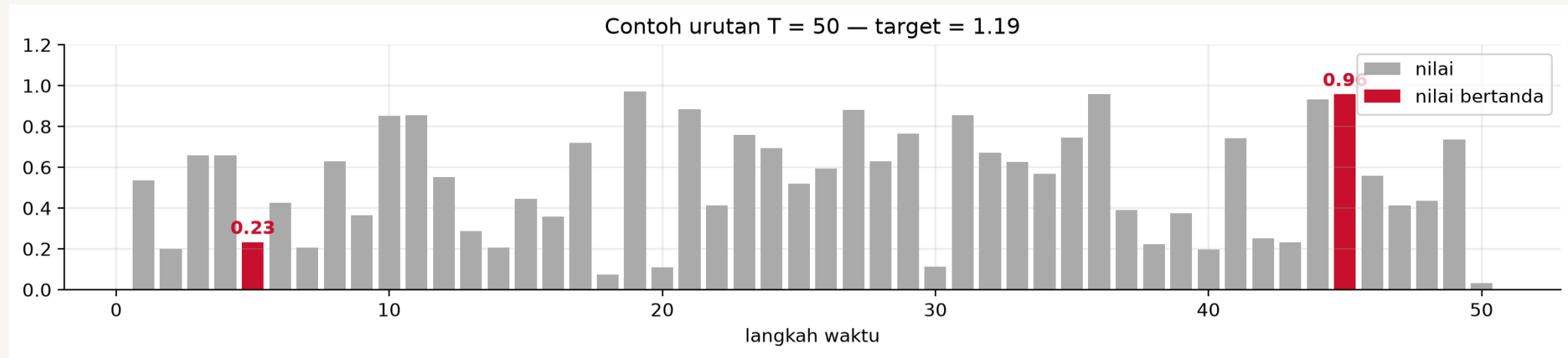
BAGIAN 6

Uji Ingatan

Di data trafik, RNN biasa sama baiknya dengan GRU.

Lalu kapan gerbang LSTM & GRU benar-benar penting?

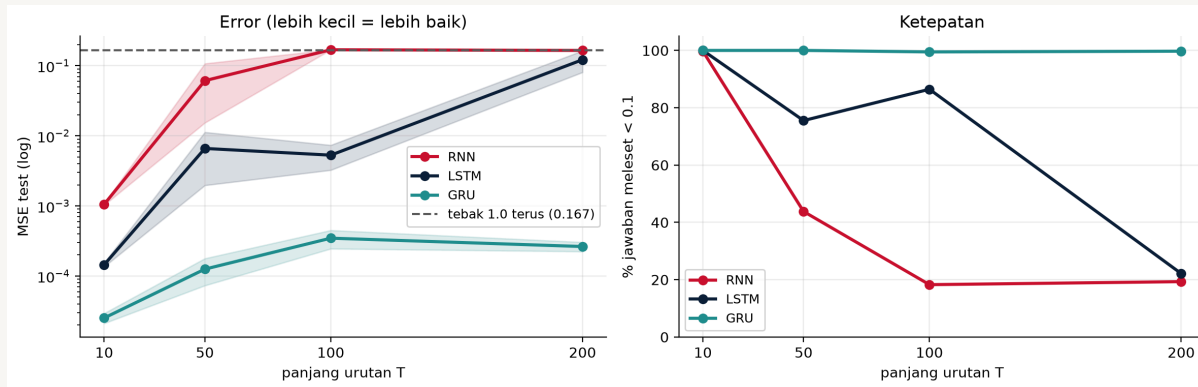
Tugas yang memaksa model mengingat jauh ke belakang



- Setiap langkah: **nilai** acak 0–1 dan **penanda** 0/1.
- Penanda pertama di **10% awal**, kedua di **paruh akhir**.
- Target: **jumlah dua nilai bertanda**.

- Nilai pertama harus **diingat** sampai akhir urutan.
- Tebak 1.0 terus → MSE 0.167 (**tidak mengingat apa pun**).
- Parameter ketiga model setara (± 17 ribu).

Persentase jawaban yang meleset < 0.1

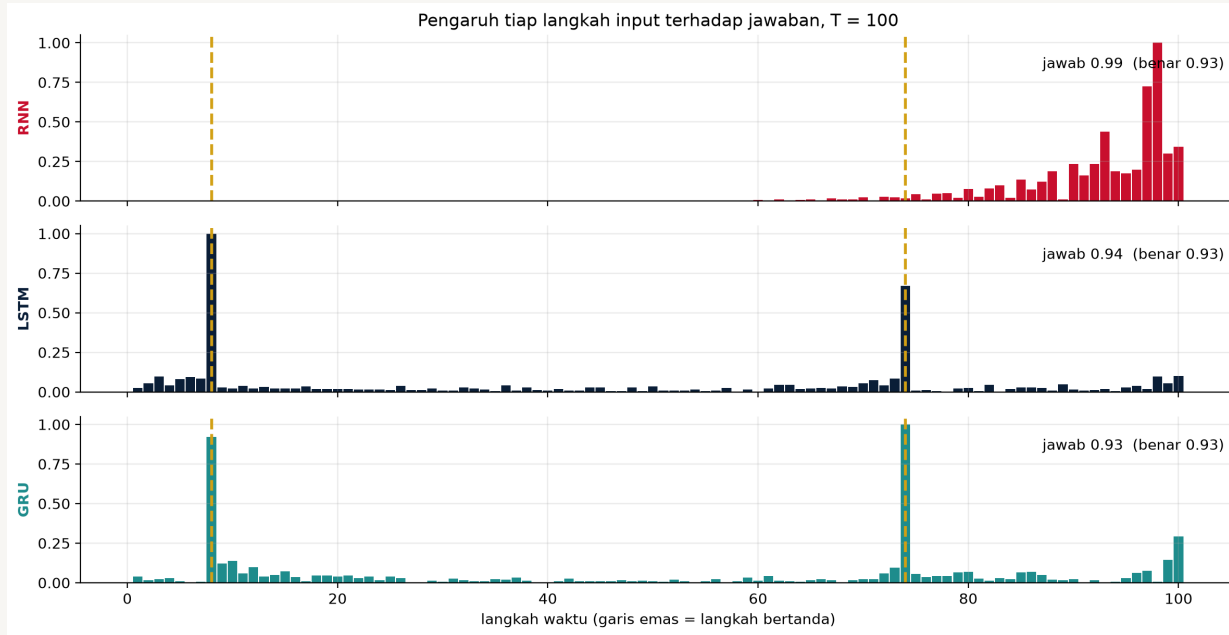


T	RNN	LSTM	GRU
10	100%	100%	100%
50	44%	76%	100%
100	18%	86%	99.5%
200	19%	22%	99.7%

RNN **gagal total** mulai $T = 100$. GRU tetap tepat hingga $T = 200$. LSTM belajar lebih lambat: dengan 4000 langkah ia belum berhasil di $T = 200$.

Ke Mana Model Melihat?

Pengaruh tiap langkah input terhadap jawaban, $T = 100$



- **RNN** hanya melihat ± 30 langkah terakhir. Nilai bertanda pertama (langkah 8) **tidak terlihat sama sekali**.
- **LSTM & GRU** menunjukkan lonjakan tajam **tepat di kedua langkah bertanda**.

Inilah fungsi gerbang

Menyimpan informasi penting, mengabaikan sisanya, lalu membawanya jauh tanpa memudar.

BAGIAN 7

Penutup

Empat pelajaran dari eksperimen hari ini

1 Urutan itu penting

Untuk 1 jam ke depan, RNN/LSTM/GRU mengalahkan ANN $\pm 25\%$ dengan parameter lebih sedikit.

2 Rumit $\neq$ lebih baik

Bila tugas hanya butuh ingatan pendek, RNN biasa sama baiknya dengan GRU.

3 Gerbang untuk ingatan panjang

Saat informasi penting ada jauh di belakang, GRU/LSTM berhasil dan RNN gagal.

4 Selalu pakai baseline

Untuk 24 jam ke depan, “jam yang sama minggu lalu” masih menang menurut MAE.

Repo yang sama dengan Sesi 1

```
cd dl-workshop
git pull          # ambil materi Sesi 2
uv sync
uv run jupyter lab
```

Tanpa GPU?

Set CEPAT = True di sel pengaturan: 1 *seed* dan epoch lebih sedikit, selesai dalam beberapa menit di laptop.

Notebook	Isi
02	Eksplorasi data trafik Milan
03	Baseline, ANN, RNN, LSTM, GRU — 1 jam ke depan
04	Perbandingan yang sama — 24 jam ke depan
05	Uji ingatan (<i>adding problem</i>)

01 hanya untuk pengajar (membuat CSV dari `milano.h5`).

Ubah satu hal, jalankan ulang, amati

1 Jendela

WINDOWS = [6, 48] di notebook 03. Model mana yang paling terpengaruh?

2 Fitur

Hapus call & sms dari FITUR. Apakah error naik?

3 Ukuran

Gandakan HIDDEN. Lebih akurat, atau *overfitting*?

4 Dua lapis

Tambahkan num_layers=2 pada sel rekuren.

5 Selisih

Notebook 04: prediksi selisih terhadap “minggu lalu”.

6 Lebih panjang

Notebook 05: PANJANG hingga 400. Masih mampu?

- Barlacchi, G., De Nadai, M., Larcher, R., dkk. (2015). A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Scientific Data*, 2, 150055.
- Cho, K., van Merriënboer, B., Gulcehre, C., dkk. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. *EMNLP*.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. otexts.com/fpp3
- Olah, C. (2015). Understanding LSTM networks. colah.github.io
- Zhang, C., Dang, S., Shihada, B., & Alouini, M.-S. (2021). Dual attention-based federated learning for wireless traffic prediction. *IEEE INFOCOM*.
- PyTorch Documentation: [Recurrent Layers](#)

Terima Kasih

Saatnya *hands-on* — buka notebook Sesi 2!

mctm.web.id/dl-workshop