

UNIVERSITAS BENGKULU – FAKULTAS TEKNIK
JURUSAN TEKNIK ELEKTRO
PROBLEM SET 10: ANALISIS DERET WAKTU DAN PERAMALAN

Topik: Rekayasa Fitur Temporal, Model XGBoost, dan Validasi Silang Bebas Leakage

Informasi Tugas & Identitas Mahasiswa

Nama	:		NPM	:	
Prodi/Kelas	:	S1 Teknik Elektro	Hari/Tanggal	:	
Dosen	:	Ir. Novalio Daratha, Ph.D.	Batas Waktu	:	1 Minggu

Soal 1. [Level C1 – Mengingat / *Remembering*] (Bobot: 10%)

- a. Jelaskan apa yang dimaksud dengan teknik *Sliding Window* dalam mengubah data deret waktu sekuensial menjadi format matriks pembelajaran terawasi (*Supervised Learning Matrix*)!
- b. Mengapa fitur jam $H \in \{0, 1, \dots, 23\}$ perlu di-encode menggunakan fungsi trigonometri (sin dan cos)?

Soal 2. [Level C2 – Memahami / *Understanding*] (Bobot: 15%)

- a. Apa yang dimaksud dengan **Lookahead Data Leakage** dalam pemodelan deret waktu dan bagaimana kesalahan ini bisa terjadi jika kita menggunakan fungsi `StandardScaler.fit_transform()` sebelum membagi dataset train/test?
- b. Jelaskan keunggulan algoritma berbasis pohon (*Gradient Boosted Decision Trees* seperti XGBoost/LightGBM) dibandingkan regresi linier klasik dalam menangani relasi non-linier antara temperatur cuaca ekstrem dan konsumsi daya listrik pendingin gedung!

Soal 3. [Level C3 – Menerapkan / *Applying*] (Bobot: 20%)

Diberikan cuplikan data beban listrik harian (MW) selama 6 jam berturut-turut:

Jam (t)	08.00	09.00	10.00	11.00	12.00	13.00
Beban y_t (MW)	140	155	160	165	150	170

- a. Hitunglah nilai fitur *Lag-1* (y_{t-1}) dan *Lag-2* (y_{t-2}) untuk $t = 10.00, 11.00, 12.00, 13.00$.
- b. Hitunglah fitur statistik jendela bergerak (*Rolling Mean* 3 jam: $\bar{y}_{\text{roll}3,t} = \frac{1}{3}(y_{t-1} + y_{t-2} + y_{t-3})$) untuk target peramalan $y_{13.00}$. (Catatan: Pastikan tidak menyertakan $y_{13.00}$ dalam kalkulasi!).
- c. Hitung nilai fitur siklis $\sin\left(\frac{2\pi \times 13}{24}\right)$ dan $\cos\left(\frac{2\pi \times 13}{24}\right)$.

Soal 4. [Level C4 – Menganalisis / *Analyzing*]

(Bobot: 20%)

Sebuah model XGBoost dilatih untuk peramalan beban jangka pendek. Analisis nilai **Feature Importance** (Gain) menghasilkan peringkat lima fitur teratas:

- (a) `lag_24` (Gain: 58.2%)
- (b) `temperature` (Gain: 18.5%)
- (c) `lag_168` (Gain: 12.1%)
- (d) `hour_sin` (Gain: 6.4%)
- (e) `rolling_std_24` (Gain: 2.8%)

Tugas Analisis:

- a. Berikan interpretasi fisis kelistrikan mengapa `lag_24` dan `lag_168` memberikan kontribusi dominan (total > 70%) terhadap ketepatan prediksi model!
- b. Mengapa fitur temperatur tetap penting meskipun kontribusinya berada di bawah `lag_24`? Apa dampaknya pada peramalan saat terjadi gelombang panas mendadak di akhir pekan?

Soal 5. [Level C5 – Mengevaluasi / *Evaluating*]

(Bobot: 15%)

Dua insinyur mengimplementasikan skema peramalan multi-langkah (*Multi-Step Ahead Forecasting*) untuk 24 jam ke depan menggunakan XGBoost:

- **Insinyur A:** Menggunakan strategi *Recursive Multi-Step* (Satu model memprediksi $t + 1$, hasilnya diumpan balik sebagai fitur lag untuk memprediksi $t + 2$, dst).
- **Insinyur B:** Menggunakan strategi *Direct Multi-Step* (Melatih 24 model independen yang terpisah: Model 1 untuk $t + 1$, Model 2 untuk $t + 2$, ..., Model 24 untuk $t + 24$).

Evaluasilah komparasi kedua strategi tersebut ditinjau dari fenomena akumulasi kesalahan (*error accumulation*), kebutuhan waktu pelatihan komputasi, dan kestabilan variansi peramalan!

Soal 6. [Level C6 – Merancang / *Creating*]

(Bobot: 20%)

Rancanglah sebuah **pipeline Python lengkap** menggunakan `scikit-learn` dan `xgboost` yang mengimplementasikan:

- a. Pembuatan fitur lag otomatis ($t - 1$ hingga $t - 48$), fitur kalender siklis (`sin / cos`), dan fitur cuaca.
- b. Skema validasi `TimeSeriesSplit(n_splits=5)`.
- c. Penyetelan hiperparameter (*Hyperparameter Tuning* untuk `max_depth`, `learning_rate`, `n_estimators`) menggunakan metode `RandomizedSearchCV`.
- d. Menghitung dan menampilkan metrik evaluasi rata-rata (RMSE dan MAPE) dari seluruh lipatan (*folds*) validasi.