

# Kinerja Model Regresi Linear dan Polinomial dalam Memprediksi Konsumsi Daya Rumah Tangga

Disusun oleh

- Willy Jonathan Arsyad
- Agil Mughni
- Alfi Zamriza
- T.M Fadlul Ihsan
- M. Arkan Haris



# Sumber dan Deskripsi Dataset

## Sumber Data

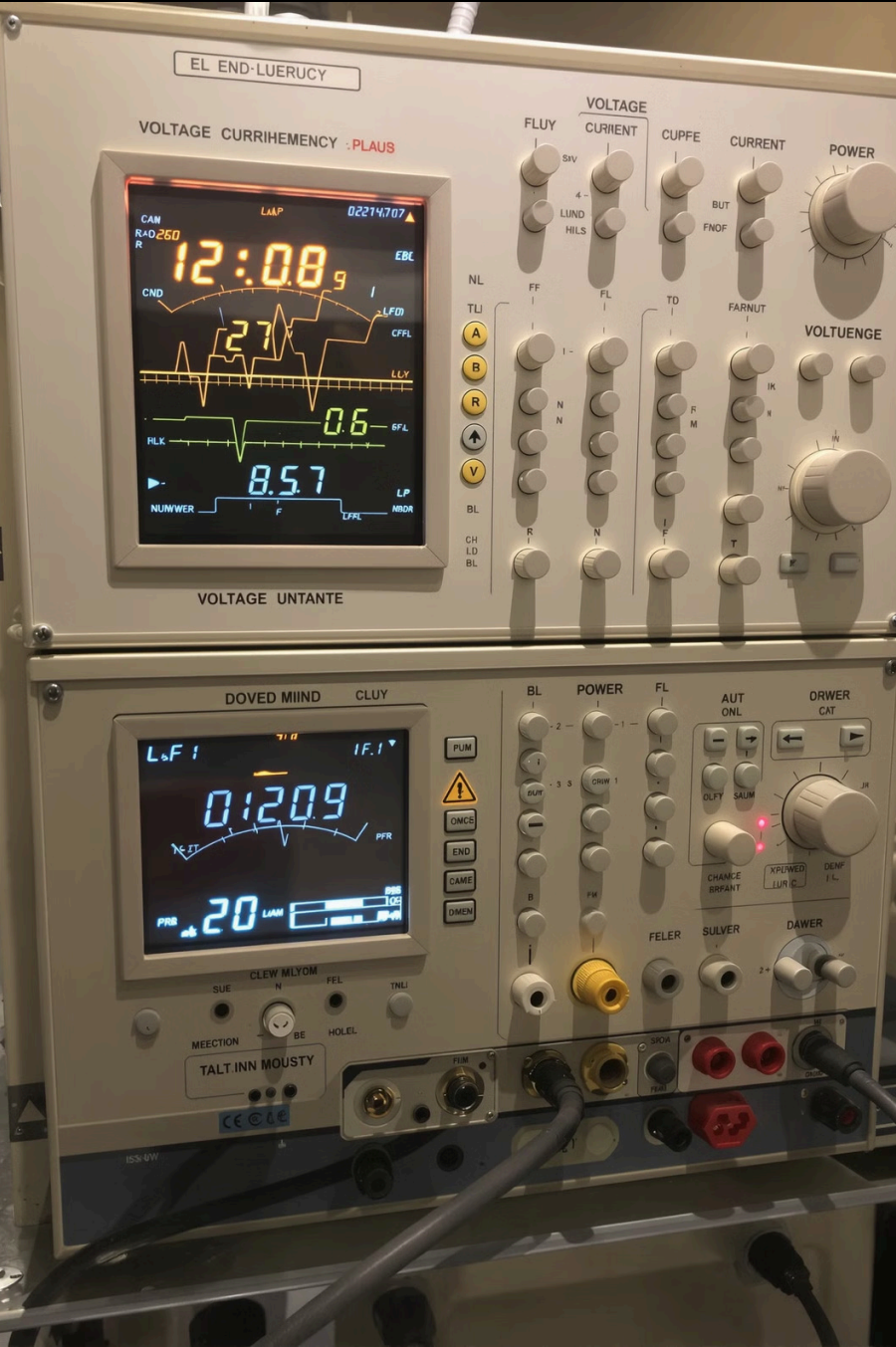
Dataset "Individual household electric power consumption" dari UCI Machine Learning Repository.

Mencatat konsumsi listrik rumah tangga di Prancis selama Desember 2006 hingga November 2010.

## Interval Pengumpulan

Data dikumpulkan setiap satu menit selama periode empat tahun.

Total data mencapai lebih dari 2 juta baris observasi.



# Variabel Penting dalam Dataset

Nama Variabel	Deskripsi
Date	Tanggal pencatatan (format: dd/mm/yyyy)
Time	Waktu pencatatan (format: hh:mm:ss)
Global_active_power	Konsumsi daya aktif global (kilowatt) - target regresi
Global_reactive_power	Konsumsi daya reaktif global (kilowatt)
Voltage	Tegangan (volt)
Global_intensity	Intensitas arus global (ampere)



# Statistika Deskriptif Dataset

## Konsumsi Daya Aktif

Rata-rata: 1.090 kW

Minimum: 0.076 kW

Maksimum: 11.122 kW

## Tegangan Listrik

Rata-rata: 237.837 V

Minimum: 1.000 V

Maksimum: 254.150 V

## Intensitas Arus

Rata-rata: 4.582 A

Minimum: 0.200 A

Maksimum: 48.400 A



# Pra-pemrosesan Data

## Penggabungan Kolom Date dan Time

Menggabungkan kolom Date dan Time menjadi satu kolom DateTime.

Memudahkan analisis berbasis waktu (harian, mingguan, bulanan).

## Konversi Tipe Data

Mengubah tipe data kolom menjadi numerik dengan errors='coerce'.

Nilai tidak valid diubah menjadi NaN untuk analisis yang akurat.

## Penanganan Missing Values

Terdapat 25.979 nilai NaN pada setiap kolom numerik.

Nilai ini perlu ditangani sebelum pemodelan dilakukan.



# Perubahan Nama Kolom dan Ukuran Dataset



## Perubahan Nama Kolom

Sub\_metering\_1 menjadi Sub\_metering\_kitchen untuk area dapur.

Sub\_metering\_2 menjadi Sub\_metering\_laundry untuk ruang cuci.



## Pengecekan Duplikat

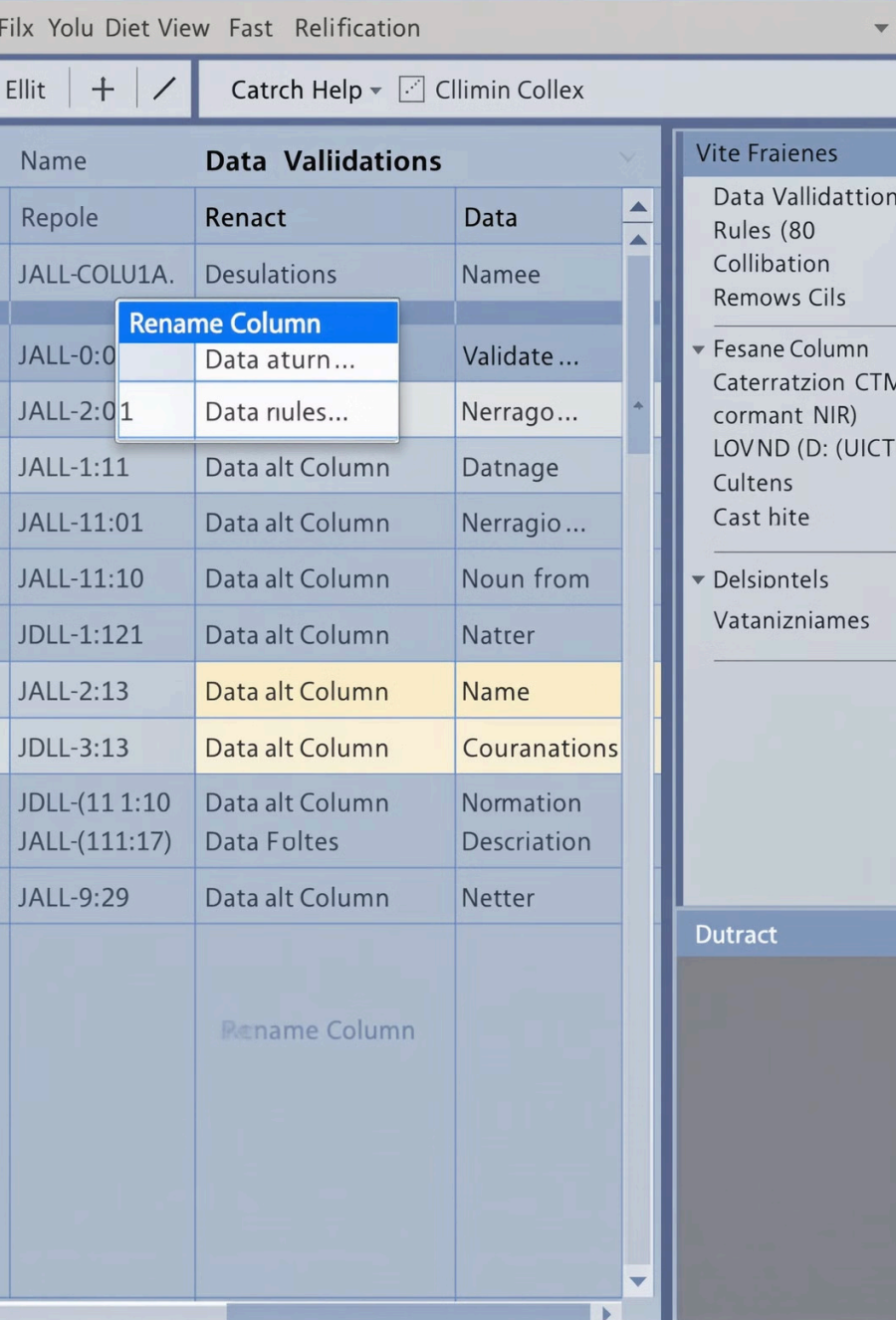
Memeriksa dan menangani data duplikat untuk menjaga kualitas data.  
Memastikan setiap observasi unik dan valid untuk analisis.



## Ukuran Dataset

Total 2.075.259 baris dan 8 kolom setelah pra-pemrosesan.

Dataset besar memungkinkan analisis pola konsumsi yang detail.



# Electricity Consumption Patterns



## Analisis Time Series Konsumsi Daya

1

### Pola Jangka Panjang

Menganalisis konsumsi daya rumah tangga seiring waktu dalam bentuk time series.

2

### Identifikasi Tren

Memeriksa pola atau tren konsumsi daya yang terjadi sepanjang periode empat tahun.

3

### Deteksi Anomali

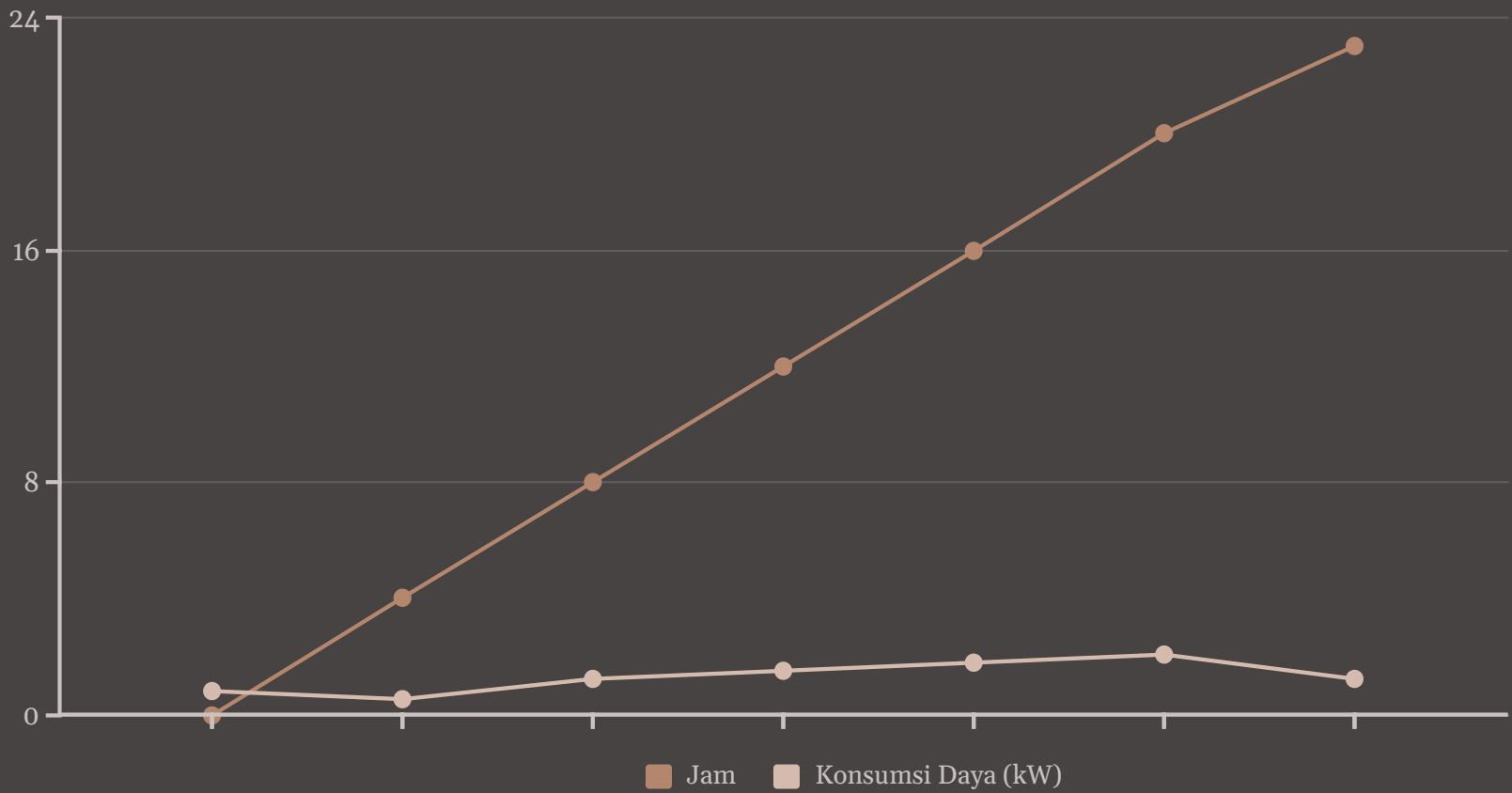
Menemukan penyimpangan atau lonjakan konsumsi daya yang tidak biasa.

4

### Analisis Musiman

Mengidentifikasi pola musiman dalam konsumsi daya (bulanan, tahunan).

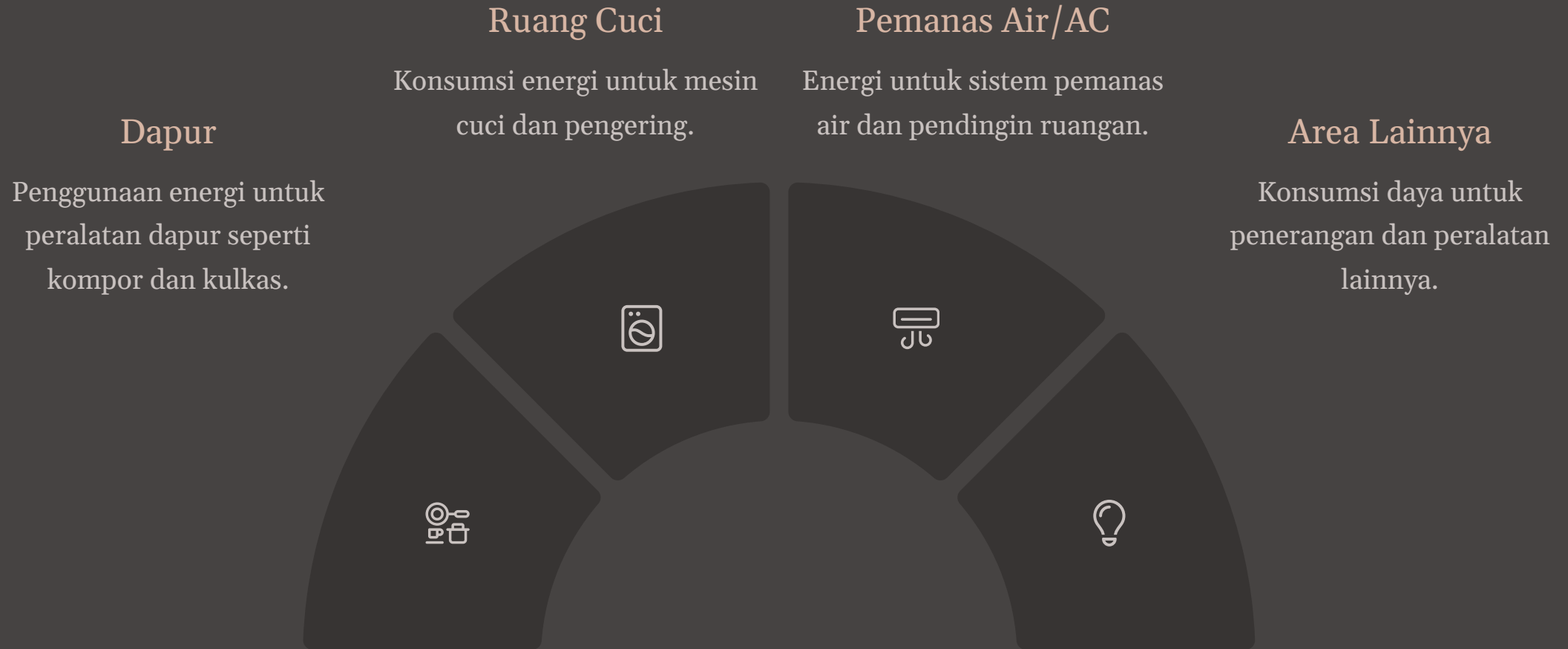
# Pola Harian Konsumsi Daya



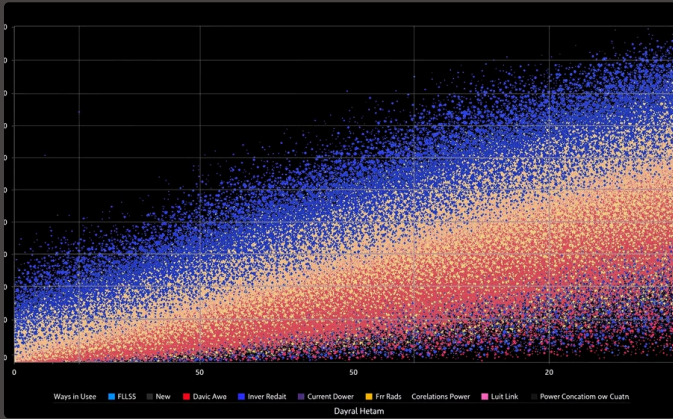
Analisis pola konsumsi daya berdasarkan waktu dalam sehari menunjukkan peningkatan pada pagi hari, puncak di sore/malam hari, dan penurunan di tengah malam.



# Perbandingan Sub-metering Area Rumah



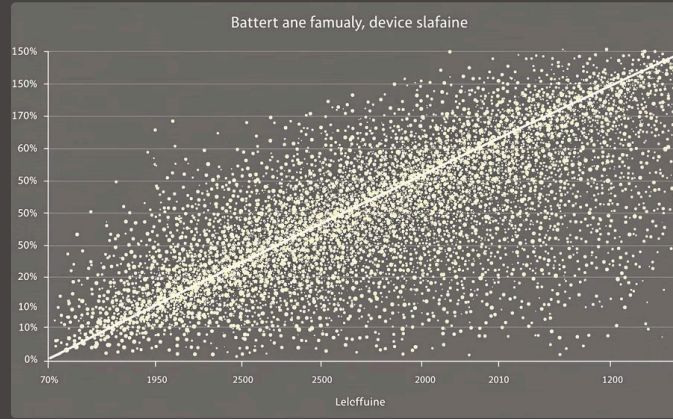
# Korelasi Antar Variabel Konsumsi Daya



## Korelasi Kuat

Global\_active\_power dan Global\_intensity memiliki korelasi sangat kuat (1.00).

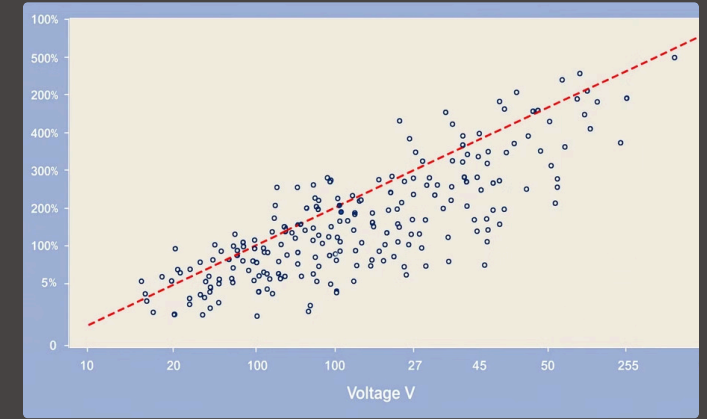
Menunjukkan hubungan linear yang hampir sempurna antara kedua variabel.



## Korelasi Sedang

Sub\_metering\_3 berkorelasi cukup kuat dengan Global\_active\_power (0.64).

Pemanas air/AC berkontribusi signifikan terhadap konsumsi daya total.



## Korelasi Negatif

Voltage dan Global\_reactive\_power memiliki korelasi negatif (-0.66).

Peningkatan tegangan cenderung menurunkan daya reaktif.

# Perbandingan Hari Kerja vs Akhir Pekan

## Pola Hari Kerja

Konsumsi daya pada hari kerja (Senin-Jumat) menunjukkan pola yang lebih teratur.

Terdapat puncak di pagi hari saat persiapan berangkat kerja.

Puncak tertinggi terjadi di malam hari saat semua anggota keluarga di rumah.

## Pola Akhir Pekan

Konsumsi daya pada akhir pekan (Sabtu-Minggu) lebih merata sepanjang hari.

Tidak ada puncak pagi yang signifikan seperti pada hari kerja.

Konsumsi total cenderung lebih tinggi karena penghuni lebih banyak di rumah.

# Linear Regression Model implementation workflow



1. data preparation nasei  
resrecton insectching data



2. Model training.

2. model evaluation is check ship  
model deploysion deploymen.

4. model devuiatio as  
a rund infilidute  
or gagamment



# Implementasi Model Regresi Linear



## Persiapan Data

Membuat fitur waktu tambahan: jam, hari dalam minggu, dan bulan.

Memisahkan fitur (X) dan target (Global\_active\_power).



## Pembagian Dataset

80% data untuk training dan 20% untuk testing.

Memastikan model diuji pada data yang belum pernah dilihat.



## Normalisasi Fitur

Menggunakan StandardScaler untuk standarisasi data fitur.

Mengubah data agar memiliki rata-rata 0 dan standar deviasi 1.



## Pelatihan Model

Membangun model LinearRegression dari scikit-learn.

Melatih model pada data training yang telah distandarisasi.

# Implementasi Model Regresi Polinomial



## Transformasi Fitur Polinomial

Menggunakan `PolynomialFeatures(degree=2)` untuk menghasilkan fitur polinomial.



## Perluasan Fitur

Menghasilkan kuadrat dan interaksi antar fitur ( $X_1^2$ ,  $X_2^2$ ,  $X_1 \cdot X_2$ ).



## Pelatihan Model Polinomial

Melatih model regresi linear pada fitur polinomial yang telah dibuat.

# Evaluasi dan Perbandingan Model

0.9980

$R^2$  Linear

Koefisien determinasi model regresi linear.

0.9988

$R^2$  Polinomial

Koefisien determinasi model regresi polinomial.

0.0022

MSE Linear

Mean Squared Error model linear.

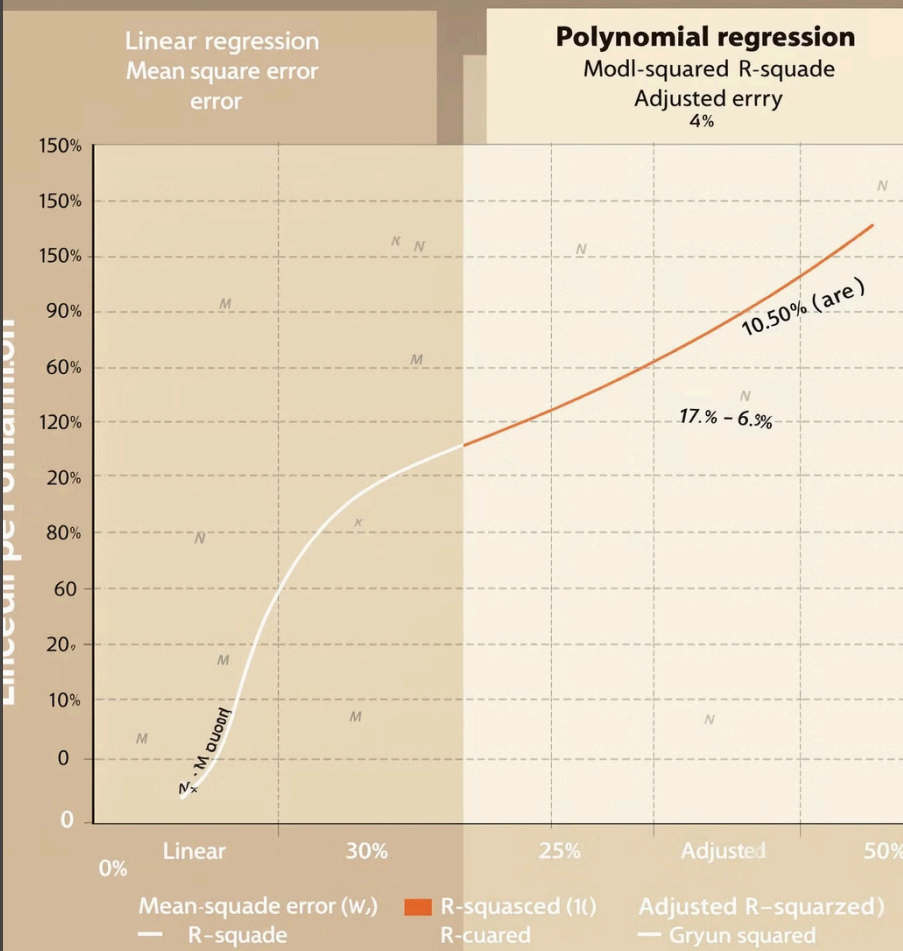
0.0013

MSE Polinomial

Mean Squared Error model polinomial.

## Linear vs. polynomial regression

### Linear vs Fine Regression



Meimome for Exgresrections



# Kesimpulan dan Fitur Penting

