

# Tugas Kelompok

## Linear dan Polynomial Regression

### 23 Maret 2025

#### Tujuan:

Mahasiswa memahami konsep dan implementasi regresi linear serta regresi polinomial melalui eksperimen dengan dataset nyata.

#### 1. Pembagian Dataset untuk 10 Kelompok

Setiap kelompok akan menggunakan dataset yang berbeda untuk menganalisis hubungan antara variabel dependen (*target*) dan variabel independen (*features*). Berikut adalah dataset yang dapat digunakan oleh setiap kelompok:

1. **House Prices**
  - Dataset: *Boston Housing* (atau dataset harga rumah lain)
  - Target: Harga rumah
  - Features: Luas rumah, jumlah kamar, usia bangunan, dll.
2. **Car Price Prediction**
  - Dataset: Harga mobil berdasarkan spesifikasi
  - Target: Harga mobil
  - Features: Tahun pembuatan, merek, tipe mesin, konsumsi BBM, dll.
3. **Salary Prediction**
  - Dataset: *Salary vs Experience*
  - Target: Gaji
  - Features: Pengalaman kerja, tingkat pendidikan, posisi pekerjaan.
4. **Startup Profit Analysis**
  - Dataset: *50 Startups*
  - Target: Profit perusahaan startup
  - Features: Modal awal, pengeluaran pemasaran, lokasi, dll.
5. **Student Performance**
  - Dataset: Nilai ujian siswa
  - Target: Skor akhir
  - Features: Jam belajar, kehadiran, keterlibatan di kelas, dll.
6. **Energy Efficiency**
  - Dataset: Konsumsi energi bangunan
  - Target: Efisiensi energi
  - Features: Luas jendela, bahan bangunan, ventilasi, dll.
7. **Health and BMI Prediction**
  - Dataset: Berat badan vs tinggi badan
  - Target: BMI
  - Features: Umur, tinggi badan, berat badan, pola makan.
8. **Weather and Temperature Forecasting**
  - Dataset: Suhu berdasarkan parameter cuaca
  - Target: Suhu
  - Features: Kelembaban, tekanan udara, kecepatan angin, dll.

### 9. E-commerce Sales Prediction

- Dataset: Penjualan produk online
- Target: Pendapatan
- Features: Harga, diskon, rating produk, jumlah ulasan.

### 10. CO2 Emission Prediction

- Dataset: Emisi karbon dari kendaraan
- Target: Emisi CO2
- Features: Jenis bahan bakar, konsumsi bahan bakar, kapasitas mesin.

## 2. Detail Project yang Harus Dikerjakan

Setiap kelompok harus melakukan tahapan berikut:

### 1. Pemahaman Dataset

- Menjelaskan sumber data dan variabel yang digunakan.
- Menampilkan statistik deskriptif dan visualisasi awal data.

### 2. Eksplorasi Data dan Pra-pemrosesan

- Menangani missing values (jika ada).
- Normalisasi atau standarisasi data jika diperlukan.
- Analisis korelasi antara variabel independen dan dependen.

### 3. Implementasi Model

- Membangun model *Linear Regression* dan *Polynomial Regression*.
- Memilih derajat polinomial yang sesuai dengan dataset.
- Menggunakan teknik validasi seperti *train-test split* atau *cross-validation*.

### 4. Evaluasi Model

- Menggunakan metrik evaluasi seperti *Mean Squared Error (MSE)*, *R<sup>2</sup> Score*, dan *Mean Absolute Error (MAE)*.
- Membandingkan kinerja regresi linear dan polinomial.

### 5. Analisis Hasil

- Menginterpretasikan koefisien regresi.
- Menyajikan grafik *regression line* dan *polynomial fit*.
- Menyimpulkan apakah model yang dibuat cukup baik dalam memprediksi target.

### 6. Laporan dan Presentasi

- Laporan dalam format PDF berisi seluruh hasil analisis dan implementasi.
- Slide presentasi per kelompok untuk menjelaskan hasil analisis.

## 3. Rubrik Penilaian

| Aspek                            | Bobot | Kriteria Penilaian  |
|----------------------------------|-------|---|
| Pemahaman Dataset                | 10%   | Memahami variabel, sumber data, dan konteks penggunaan dataset.                                   |
| Eksplorasi & Pra-pemrosesan Data | 15%   | Menganalisis data dengan visualisasi dan teknik pra-pemrosesan yang tepat.                        |
| Implementasi Model               | 25%   | Mengimplementasikan regresi linear dan polinomial dengan baik serta memilih parameter yang tepat. |
| Evaluasi Model                   | 20%   | Menggunakan metrik evaluasi dengan benar dan membandingkan performa model.                        |
| Analisis Hasil                   | 15%   | Menyajikan interpretasi yang jelas dari hasil model dengan visualisasi yang informatif.           |

|                                       |     |   |
|---------------------------------------|-----|---|
| <b>Laporan &amp; Slide Presentasi</b> | 15% | Laporan tersusun dengan baik, presentasi jelas dan komunikatif. |
|---------------------------------------|-----|---|

Total: **100%**

**Ketentuan Tambahan:**

- Kode program diimplementasikan menggunakan Jupyter Notebook (.ipynb) dan jelaskan setiap proses yang dilakukan menggunakan Markdown.
- Tugas dikumpulkan melalui Github repository dengan nama repo
  - **Kelompok\_<Nomor Kelompok>\_Tugas02\_Linear\_Regression**
  - Tugas dikumpulkan oleh repository ketua kelompok saja

**Batas Waktu Pengumpulan: 7 April 2025, Pukul 23:00 WIB**