

Tugas Kelompok – Kelas B

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. Air Pollution Prediction

- Dataset: Tingkat polusi udara berdasarkan faktor lingkungan.
- Target: Indeks Kualitas Udara (*Air Quality Index* - AQI).
- Features: Kadar CO, NO₂, SO₂, suhu, kelembaban, kecepatan angin.

2. Traffic Volume Prediction

- Dataset: Volume lalu lintas berdasarkan faktor lingkungan dan waktu.
- Target: Jumlah kendaraan per jam.
- Features: Waktu dalam sehari, cuaca, hari kerja/libur, kecepatan rata-rata kendaraan.

3. Crop Yield Prediction

- Dataset: Produksi hasil pertanian berdasarkan faktor cuaca dan tanah.
- Target: Hasil panen per hektar.
- Features: Curah hujan, suhu, jenis tanah, pemakaian pupuk.

4. Electricity Consumption Forecasting

- Dataset: Konsumsi listrik rumah tangga atau industri.
- Target: Konsumsi listrik (kWh).
- Features: Waktu dalam sehari, suhu, jumlah penghuni, penggunaan perangkat elektronik.

5. Medical Cost Prediction

- Dataset: Biaya medis berdasarkan data pasien.
- Target: Biaya pengobatan.
- Features: Usia, indeks massa tubuh (BMI), jumlah anak, riwayat merokok, jenis kelamin.

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)*, *R² 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.
Laporan & Slide Presentasi	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