

# LINEAR REGRESSION

# Anggota Kelompok

M. Nabil Maulana (2208107010011)

Irfan Rizadi (2208107010062)

Maulana Fikri (2208107010042)

Indriani Miza Alfiyanti (2208107010026)

Raihan Firyal (2208107010084)

# Latar Belakang

- Prediksi suhu penting untuk sektor pertanian, penerbangan, kesehatan, dll.
- Digunakan metode: Linear Regression & Polynomial Regression.
- Tujuan: prediksi suhu berdasarkan kelembaban, tekanan, dan kecepatan angin.

# Tujuan Penelitian

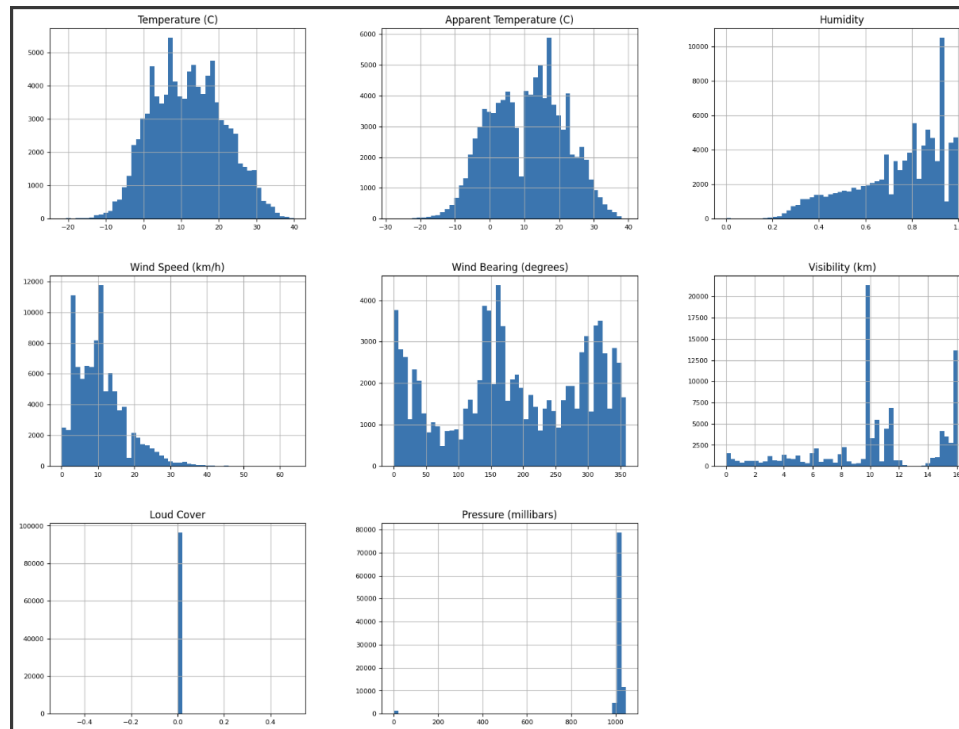
- Memahami pengaruh parameter cuaca terhadap suhu.
- Membangun dan membandingkan 2 model regresi.
- Evaluasi performa model (MSE, MAE,  $R^2$ ).
- Menilai efektivitas prediksi suhu.

# Dataset & Statistik Deskriptif

- Total data: 96.453 entri cuaca historis.
- Target: Suhu (Temperature °C).
- Variabel: Kelembaban, Kecepatan Angin, Tekanan, dll.
- Fitur Cloud Cover tidak bervariasi.

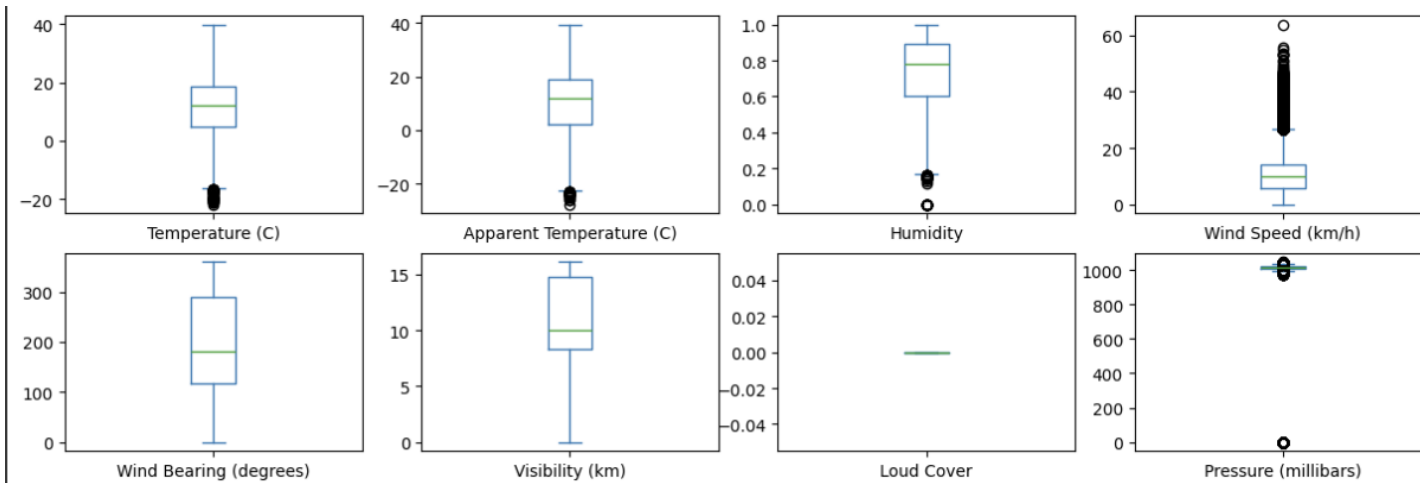
# Visualisasi Data

- Histogram menunjukkan distribusi normal suhu.



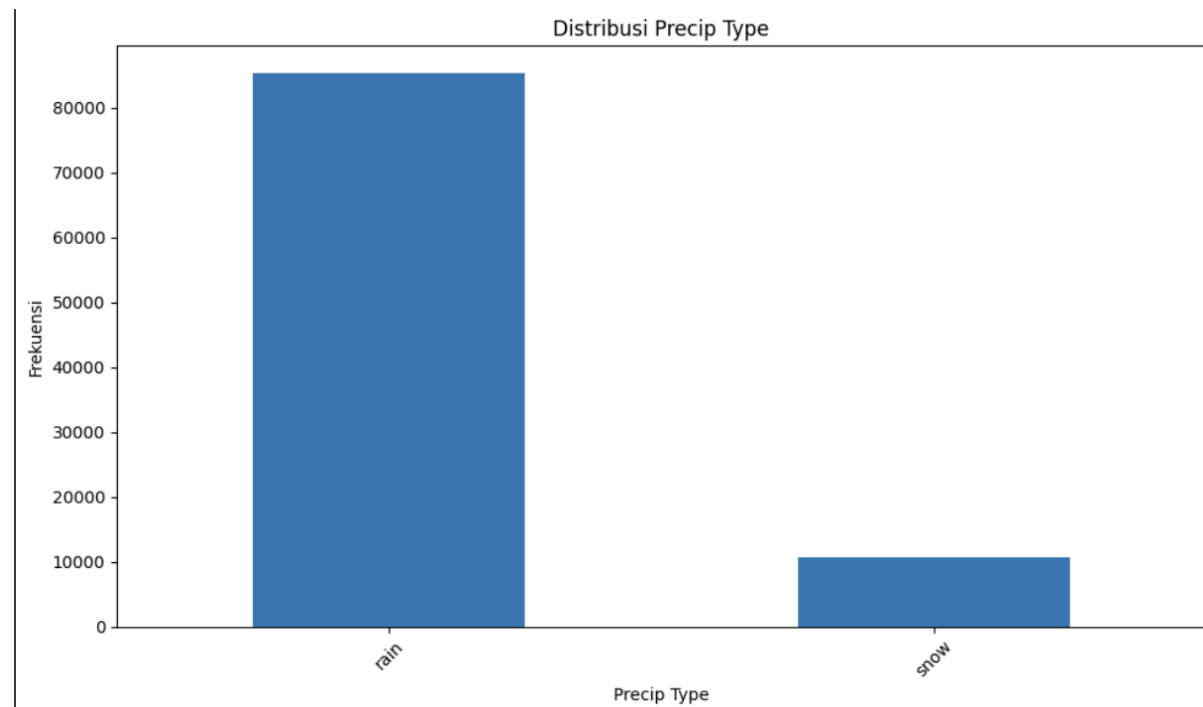
# Visualisasi Data

- Boxplot Kolom Numerik



# Visualisasi Data

- Distribusi Kolom Kategorikal (Bar Plot)

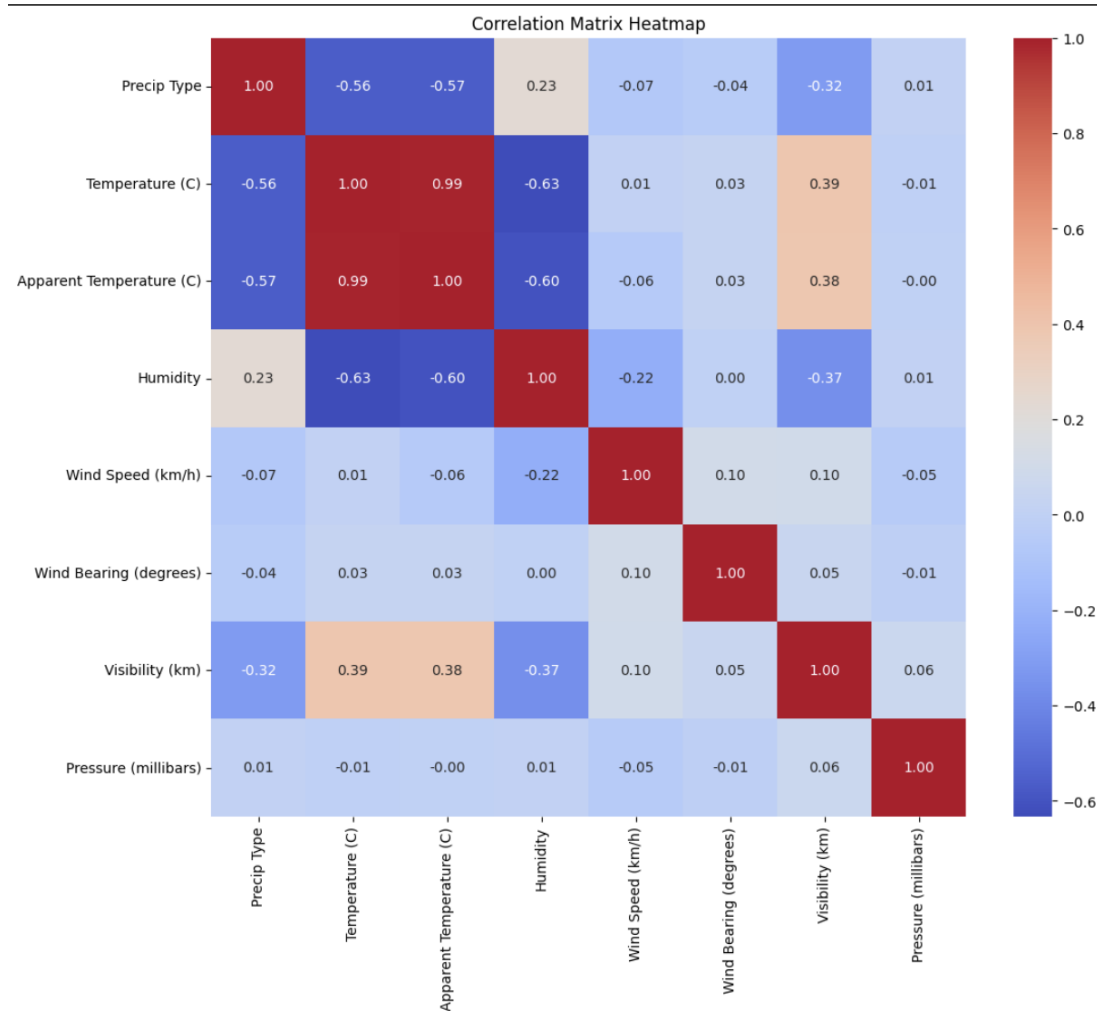




# Eksplorasi & Preprocessing

- Cek & hapus missing value dan duplikasi.
- Encoding Precip Type (rain=0, snow=1).
- Drop kolom dengan korelasi tinggi.
- Outlier handling pakai IQR.
- Scaling: RobustScaler.

# Eksplorasi & Preprocessing

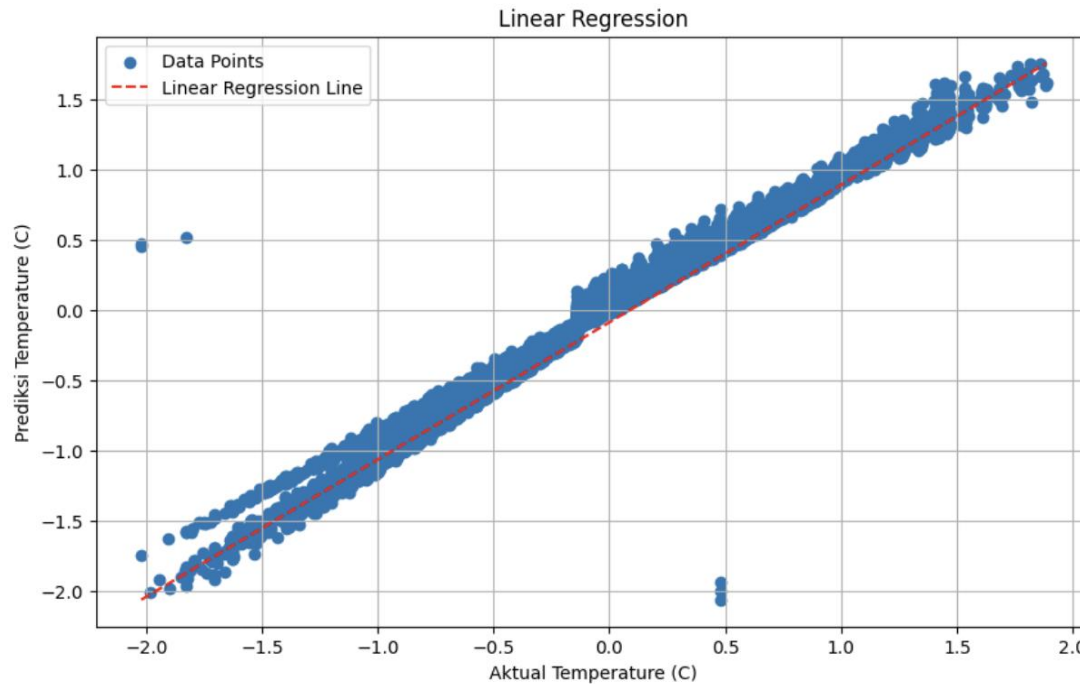


# Evaluasi Model

- Linear Regression:  $MSE=0.0062$ ,  $R^2=0.9862$ ,  $MAE=0.0521$
- Polynomial Regression:  $MSE=0.0015$ ,  $R^2=0.9966$ ,  $MAE=0.0184$

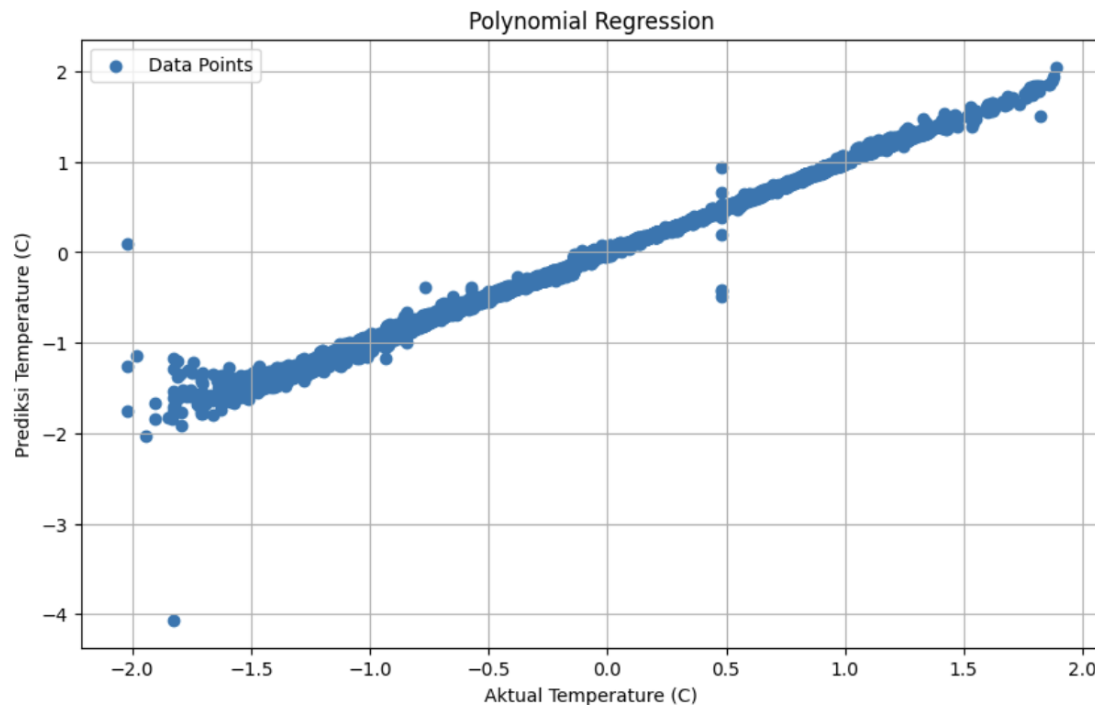
# Linear Regression

- Data split: 80% train – 20% test.
- Variabel target: Temperature (C).
- Model Linear Regression dilatih menggunakan  $X_{\text{train}}$ ,  $y_{\text{train}}$ .



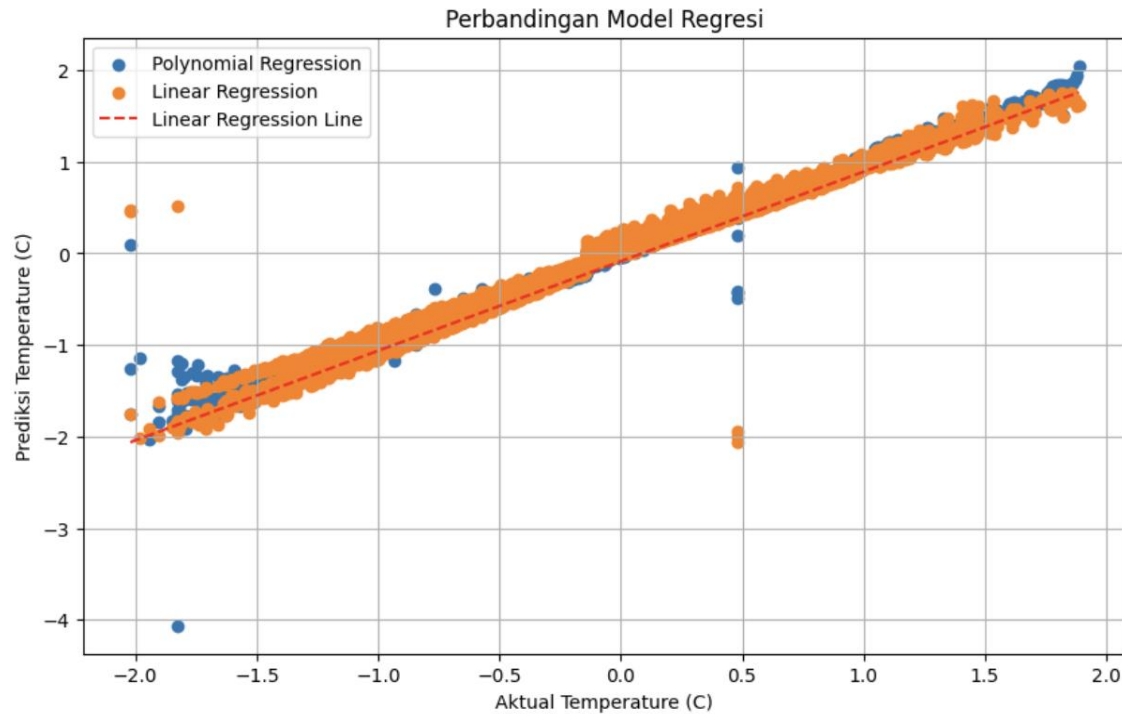
# Polynomial Regression

- Model Ridge Regression dalam pipeline.
- Optimasi hyperparameter: Degree = 2, 3, 4 | Alpha = 0.1, 1, 10
- Terbaik: Degree 4, Alpha 0.1



# Model Regresi

- variabel “Apparent Temperature (C)”
- koefisien = 1.0041
- Intercept = 0.0545



# Perbandingan Model

- Linear Regression: lebih stabil, generalisasi baik.
- Polynomial Regression: lebih akurat, tapi sensitif.
- Variabel paling berpengaruh: Apparent Temperature (C)

# Kesimpulan

- Kedua model efektif, namun Polynomial Regression lebih akurat.
- Linear Regression lebih stabil dan cocok untuk generalisasi.
- Pentingnya preprocessing dan pemilihan fitur yang tepat.