

Nama : Fadel Muhammad

NIM : 1103213062

1. Jika model linear regression atau decision tree mengalami underfitting pada dataset ini, strategi apa yang akan digunakan untuk meningkatkan performanya? Bandingkan setidaknya dua pendekatan berbeda (misal: transformasi fitur, penambahan features, atau perubahan model ke algoritma yang lebih kompleks), dan jelaskan bagaimana setiap solusi memengaruhi bias-variance tradeoff!

#### Strategi Mengatasi Underfitting

- Transformasi fitur: Tambahkan fitur polinomial atau interaksi antar fitur. Ini membantu menangkap pola non-linear → *bias turun, varian naik*.
  - Ganti model lebih kompleks: Linear → Polynomial Regression, Decision Tree → Random Forest. Ini membuat model lebih fleksibel → *bias turun, varian bisa naik*.
2. Selain MSE, jelaskan dua alternatif loss function untuk masalah regresi (misal: MAE, Huber loss) dan bandingkan keunggulan serta kelemahannya. Dalam skenario apa setiap loss function lebih cocok digunakan? (Contoh: data dengan outlier, distribusi target non-Gaussian, atau kebutuhan interpretasi model).

Loss	Kelebihan	Kekurangan	Cocok untuk
MAE	Tahan terhadap outlier	Sulit dioptimasi (tidak halus)	Data banyak outlier
Huber	Kombinasi MAE & MSE, stabil	Butuh pilih parameter delta	Data outlier sedang, stabilitas

3. Tanpa mengetahui nama fitur, metode apa yang dapat digunakan untuk mengukur pentingnya setiap fitur dalam model? Jelaskan prinsip teknikal di balik metode tersebut (misal: koefisien regresi, feature importance berdasarkan impurity reduction) serta keterbatasannya!

#### Cara Mengetahui Pentingnya Fitur Tanpa Tahu Namanya

- Regresi: Lihat besar koefisien ( $|\beta|$ ) Semakin besar, semakin penting.
- Decision Tree/Random Forest: Lihat *feature importance* dari pengurangan impurity (Gini/entropy).

Kelemahan: Bisa bias jika fitur punya skala berbeda atau korelasi tinggi antar fitur.

4. Bagaimana mendesain eksperimen untuk memilih hyperparameter optimal (misal: learning rate untuk SGDRegressor, max\_depth untuk Decision Tree) pada dataset ini? Sertakan analisis tradeoff antara komputasi, stabilitas pelatihan, dan generalisasi model!

#### Cara Pilih Hyperparameter (misalnya max\_depth, learning\_rate)

- Gunakan Grid Search atau Random Search dengan Cross Validation.
  - Pertimbangkan:
    - Komputasi: Random Search lebih cepat.
    - Stabilitas: Cross-validation bantu deteksi overfitting.
    - Generalisasi: Pilih parameter dengan skor validasi terbaik, bukan hanya train.
5. Jika menggunakan model linear regression dan residual plot menunjukkan pola non-linear serta heteroskedastisitas, langkah-langkah apa yang akan diambil? (contohnya: Transformasi data/ubah model yang akan dipakai/etc)

#### Residual Linear Regression Tunjukkan Pola Non-Linear

Langkah-langkah:

- Lakukan transformasi fitur (log, sqrt, polinomial).
- Ubah ke model non-linear: Decision Tree, Random Forest, Gradient Boosting.

- Jika heteroskedastisitas: gunakan Weighted Least Squares atau transformasi target.