

Nama : Izzan Muhammad Fa'iz (1103210126)

Kelas : TK4505

UTS I Regresi (RegresiUTSTelkom)

1. Jika model seperti Linear Regression atau Decision Tree Regressor mengalami underfitting, artinya model terlalu sederhana untuk menangkap pola dalam data, maka terdapat dua pendekatan berbeda yang dapat dilakukan untuk meningkatkan performanya.
 - Pendekatan pertama adalah melakukan transformasi fitur, seperti menambahkan fitur interaksi atau menerapkan Polynomial Features. Hal ini memungkinkan model yang awalnya linier menangkap hubungan non-linier. Transformasi ini secara teknikal menurunkan bias karena model menjadi lebih fleksibel, tetapi juga meningkatkan varians, sehingga perlu pengendalian melalui regularisasi atau validasi silang.
 - Pendekatan kedua adalah mengganti model ke algoritma yang lebih kompleks seperti Gradient Boosting Regressor atau Support Vector Regressor (SVR). Model-model ini memiliki kapasitas yang lebih tinggi untuk menangkap pola data yang kompleks dan non-linier. Namun, model ini juga lebih berisiko mengalami overfitting jika tidak dikontrol dengan baik.

Dalam konteks bias-variance tradeoff, pendekatan pertama menambah fleksibilitas model secara eksplisit, sedangkan pendekatan kedua mengandalkan kompleksitas algoritma untuk mengurangi bias.

2. Selain menggunakan Mean Squared Error (MSE) sebagai loss function, terdapat dua alternatif penting yaitu Mean Absolute Error (MAE) dan Huber Loss.
 - MAE mengukur rata-rata selisih absolut antara nilai prediksi dan aktual, sehingga lebih tahan terhadap outlier. Namun, MAE kurang sensitif terhadap variasi kecil, dan tidak dapat diturunkan secara analitik dengan mudah seperti MSE.
 - Sebaliknya, Huber Loss menggabungkan sifat MSE dan MAE, di mana ia bertindak seperti MSE untuk error kecil, dan seperti MAE untuk error besar. Ini

menjadikannya ideal untuk situasi di mana terdapat outlier dalam jumlah terbatas, atau ketika target memiliki distribusi non-Gaussian.

Pada skenario loss function, MAE lebih cocok digunakan jika data banyak mengandung nilai ekstrem, sedangkan Huber Loss lebih stabil dan fleksibel saat distribusi error bervariasi.

3. Jika nama fitur tidak diketahui tetapi kita tetap ingin mengetahui pentingnya suatu fitur dalam model, maka terdapat dua metode teknikal yang umum digunakan. Pertama-tama, pada model linier kita dapat melihat nilai koefisien regresi sebagai indikator pengaruh fitur. Semakin besar nilai absolut koefisien, maka semakin besar kontribusi fitur terhadap target. Namun, metode ini sangat bergantung pada skala fitur dan mengasumsikan hubungan linier. Kedua, pada model pohon keputusan atau ensemble seperti Random Forest, kita dapat menggunakan metrik feature importance yang dihitung dari penurunan impurity (misalnya Gini atau Variance). Semakin besar pengurangan impurity yang disumbangkan oleh fitur, semakin penting fitur tersebut. Metode ini tidak bergantung pada skala, namun bisa bias jika menghadapi fitur dengan banyak kategori atau nilai unik.
4. Untuk memilih hyperparameter optimal, desain eksperimen dapat dilakukan melalui Grid Search atau Randomized Search dengan validasi silang (cross-validation). Cara yang dilakukan yaitu menentukan rentang nilai untuk setiap parameter, lalu lakukan eksplorasi dengan k-fold cross-validation untuk menghindari overfitting dan memastikan model stabil di berbagai subset data. Semakin banyak kombinasi yang diuji, maka semakin tinggi biaya komputasi yang dibutuhkan. Namun, eksplorasi yang luas dapat meningkatkan peluang untuk menemukan konfigurasi optimal yang dapat menyeimbangkan akurasi dan generalisasi.
5. Jika residual plot dari model Linear Regression menunjukkan pola non-linear atau heteroskedastisitas (varian residual tidak konstan), maka beberapa langkah dapat dilakukan. Caranya yaitu kita dapat melakukan transformasi pada variabel target menggunakan logaritma, akar kuadrat, atau transformasi Box-Cox. Setelah itu, kita bisa mentransformasi fitur atau mengganti model ke algoritma non-linear seperti Decision Tree, Random Forest, atau Support Vector Regressor (SVR).

Model-model ini lebih fleksibel dan mampu menangkap pola non-linier yang tidak bisa ditangkap oleh regresi linier. Dalam kasus tertentu, pendekatan regularisasi

seperti Ridge atau Lasso Regression juga bisa digunakan untuk mengurangi varian berlebih dan memperbaiki kestabilan model.