

Analisis *notebook*

Nama : Muhammad Abyan Ridhan Siregar

NIM : 1103210053

Kelas : TK-45-01

Notebook 1: Worksheet_of_7_TugasXGBoost_1.ipynb

Dataset yang digunakan dalam notebook ini adalah tugasxgboost.csv.

Analisis Output

1. Data Preprocessing:
 - a. Jika dataset berisi fitur-fitur seperti X1, X2, ... X11 dan Class sebagai target, pastikan tidak ada nilai yang hilang (missing values) dan semua kolom memiliki tipe data yang sesuai (numerik).
 - b. Pembagian data ke dalam training dan testing set akan memberikan dasar untuk mengevaluasi performa model.
2. Training dan Evaluasi Model:
 - a. Akurasi: Output akurasi model akan menunjukkan seberapa baik model dapat memprediksi label Class. Nilai akurasi yang tinggi (mendekati 1 atau 100%) menunjukkan performa yang baik, namun tetap harus ditinjau apakah dataset memiliki bias.
 - b. Confusion Matrix: Ini memberikan rincian jumlah prediksi benar dan salah untuk setiap kelas. Perhatikan apakah ada ketidakseimbangan dalam memprediksi kelas tertentu.
 - c. Classification Report: Laporan ini mencakup metrik seperti precision, recall, dan F1-score untuk setiap kelas dalam Class. Lihat apakah metrik ini konsisten atau ada kelas tertentu yang sulit diprediksi dengan baik.
3. Feature Importance:
 - a. Grafik feature importance akan menunjukkan fitur mana yang paling berpengaruh dalam menentukan kelas. Misalnya, jika X1 dan X5 memiliki nilai penting yang tinggi, ini menunjukkan bahwa fitur tersebut sangat berkaitan dengan prediksi kelas Class.

Kesimpulan Notebook 1

Jika model memberikan hasil akurasi yang tinggi dan grafik feature importance menunjukkan fitur-fitur tertentu yang dominan, Anda bisa mempertimbangkan untuk fokus pada fitur-fitur tersebut dalam analisis lebih lanjut atau eksperimen lanjutan.

Notebook 2: Worksheet_of_7_TugasXGBoost_2.ipynb

Dataset yang digunakan adalah tugasxgboost2.csv.

Analisis Output

1. Data Preprocessing:
 - a. Jika dataset mirip dengan tugasxgboost.csv, pastikan bahwa semua kolom terpisah dengan baik dan formatnya sudah sesuai. Periksa juga apakah ada kolom dengan nilai kosong atau yang perlu dikonversi.
2. Model Training dan Evaluasi:
 - a. Akurasi Model: Nilai akurasi akan menunjukkan performa model dalam memprediksi target. Jika akurasi rendah, ini bisa disebabkan oleh fitur yang kurang informatif atau perlu penyesuaian parameter.
 - b. Confusion Matrix dan Classification Report: Lihat apakah model kesulitan memprediksi kelas tertentu. Jika recall atau precision untuk kelas tertentu sangat rendah, mungkin perlu dilakukan balancing atau penyesuaian model.
 - c. Feature Importance: Perhatikan fitur mana yang paling berpengaruh terhadap prediksi. Jika ada fitur yang memiliki nilai penting jauh lebih tinggi daripada yang lain, ini menunjukkan bahwa fitur tersebut sangat menentukan hasil prediksi model.

Kesimpulan Notebook 2

Jika ada beberapa fitur dominan dalam feature importance dan model memberikan hasil yang stabil, Anda bisa mempertimbangkan untuk mengurangi fitur yang kurang penting guna menyederhanakan model.

Notebook 3: Worksheet_of_7_XGBoost_data_melb.ipynb

Dataset yang digunakan adalah melb_data.csv, kemungkinan berisi data properti di Melbourne.

Analisis Output

1. Data Preprocessing:
 - a. Dataset ini mungkin mengandung berbagai jenis kolom seperti harga, lokasi, luas tanah, dan lain-lain. Pastikan bahwa semua kolom yang diperlukan telah diubah ke tipe data yang sesuai (misalnya, harga sebagai float, kategori sebagai numerik jika perlu).
 - b. Pisahkan dataset ke dalam fitur dan target (misalnya, harga properti sebagai target).
2. Training dan Evaluasi Model:
 - a. Akurasi Model: Dalam prediksi harga atau regresi, nilai error (misalnya MSE, RMSE) adalah metrik yang penting. Semakin rendah nilai error, semakin baik performa model dalam memprediksi harga properti.
 - b. Confusion Matrix dan Classification Report: Jika masalahnya adalah klasifikasi harga dalam kategori, lihat distribusi prediksi. Jika ini masalah regresi, confusion matrix tidak akan tersedia.
3. Feature Importance:
 - a. Analisis feature importance untuk melihat faktor mana yang paling mempengaruhi harga properti. Fitur seperti lokasi, luas tanah, dan jumlah kamar mungkin memiliki nilai penting yang lebih tinggi, menunjukkan bahwa fitur tersebut adalah penentu utama dalam penentuan harga.

Kesimpulan Notebook 3

Fokus pada fitur-fitur yang memiliki pengaruh besar terhadap harga properti. Jika model memiliki error yang cukup rendah, model ini dapat membantu memperkirakan harga properti di Melbourne dengan cukup akurat.

Rekomendasi Umum untuk Semua Notebook :

1. Hyperparameter Tuning: Jika akurasi model tidak memuaskan, pertimbangkan untuk melakukan tuning hyperparameter pada model XGBoost.
2. Balancing Data: Jika dataset memiliki distribusi kelas yang tidak seimbang, pertimbangkan untuk menggunakan teknik oversampling atau undersampling.
3. Feature Engineering: Cobalah membuat fitur baru atau menormalkan fitur-fitur yang ada jika relevan.