

Muhammad Harits

1103223159

Bab 2: Proyek Machine Learning End-to-End

Ringkasan Bab

Bab ini merupakan panduan praktis yang menunjukkan seluruh proses penyusunan proyek Machine Learning secara end-to-end dengan studi kasus regression. Fokus proyek adalah memprediksi harga rumah di berbagai distrik di California menggunakan data sensus.

Langkah-langkah yang dibahas meliputi:

1. Mengunduh dan menyiapkan data
Termasuk pembuatan train/test split untuk memastikan evaluasi model yang objektif.
2. Eksplorasi dan visualisasi data
Bertujuan untuk memahami struktur data dan mendapatkan wawasan awal (insight) sebelum pemodelan.
3. Persiapan data (data preprocessing)
 - o Membersihkan data (data cleaning)
 - o Menangani atribut teks
 - o Membuat custom transformers
 - o Melakukan feature scaling
4. Membangun pipeline preprocessing
Pipeline memungkinkan proses transformasi data dilakukan secara konsisten dan otomatis pada tahap pelatihan maupun prediksi.
5. Pelatihan beberapa model ML
Menggunakan beberapa algoritma regresi, seperti:
 - o Linear Regression
 - o Decision Tree
 - o Random Forest
6. Evaluasi model
Menggunakan cross-validation untuk mengukur performa model secara lebih akurat dan mengurangi risiko hasil evaluasi yang bias.
7. Fine-tuning model
Mengoptimalkan performa model menggunakan GridSearchCV untuk menemukan kombinasi hyperparameter terbaik.

Dengan mengikuti seluruh langkah ini, pembaca memahami tidak hanya cara melatih model, tetapi juga bagaimana mengelola dataset dunia nyata dan membangun alur kerja ML yang reproduktif dan terstruktur dari awal sampai akhir.