

Nama : Mauricio Bethoven Tigauw
NIM : 1103204099
Kelas : TK-44-04

PyTorch Custom Datasets

PyTorch custom datasets merujuk pada kemampuan untuk membuat dataset khusus atau disesuaikan sesuai dengan kebutuhan pengguna dalam pengembangan model menggunakan framework PyTorch. Dalam konteks machine learning, dataset adalah kumpulan data yang digunakan untuk melatih atau menguji model. PyTorch menyediakan fasilitas untuk membuat dataset khusus agar dapat diintegrasikan secara efisien dengan alur kerja pelatihan model.

Untuk membuat dataset khusus di PyTorch, pengguna perlu membuat kelas yang mengimplementasikan antarmuka PyTorch **`torch.utils.data.Dataset`**. Kelas ini harus menyediakan metode khusus, yaitu **`__len__`** untuk mengembalikan panjang dataset dan **`__getitem__`** untuk mengakses item individu dari dataset berdasarkan indeks.

Dalam metode **`__getitem__`**, pengguna dapat mengakses data mentah dari sumber daya eksternal seperti file gambar, file teks, atau sumber daya lainnya. Selanjutnya, data tersebut dapat diproses, diubah, atau disesuaikan sesuai dengan kebutuhan spesifik proyek.

Contoh penggunaan PyTorch custom datasets dapat mencakup situasi di mana data tidak tersedia dalam format standar atau memerlukan pra-pemrosesan yang kompleks sebelum digunakan dalam pelatihan model. Dengan menggunakan custom datasets, pengguna dapat dengan mudah mengintegrasikan data yang unik atau kompleks ke dalam alur kerja PyTorch tanpa kehilangan fleksibilitas.

Secara umum, custom datasets memungkinkan pengguna untuk lebih bebas dalam menyesuaikan proses pengolahan data sesuai dengan kebutuhan proyek tertentu, memungkinkan fleksibilitas yang lebih besar dalam membangun model machine learning dengan PyTorch.