

**NAMA : RHEZA ILHAM**  
**NIM : 1103204033**

## **Laporan Teknis: Pembuatan Dataset Kustom di PyTorch**

### **Konsep Dataset di PyTorch**

PyTorch menyediakan dua kelas utama untuk penanganan dataset: `Dataset` dan `DataLoader`. Kelas `Dataset` adalah kelas abstrak untuk mewakili sebuah dataset. Subkelas dari `Dataset` harus mengimplementasikan metode `__len__` dan `__getitem__`. Sementara `DataLoader` digunakan untuk membungkus iterable di sekitar `Dataset`.

### **Membuat Dataset Kustom**

Untuk membuat dataset kustom, pertama-tama harus mewarisi kelas `Dataset` dari PyTorch. Setelah itu, mendefinisikan metode `__init__`, `__len__`, dan `__getitem__`. Metode `__init__` digunakan untuk inisialisasi dataset, `__len__` untuk memberikan jumlah total data dalam kumpulan, dan `__getitem__` untuk mengambil sampel data spesifik.

### **Struktur Data**

Data dalam dataset kustom dapat berasal dari berbagai sumber seperti file gambar, teks, atau file audio. Penting untuk mengatur struktur data dengan cara yang memudahkan akses dan manipulasi data. Biasanya, data disimpan dalam struktur folder atau dalam format file seperti CSV atau JSON.

### **Pengolahan Data Awal**

Sebelum data dimasukkan ke dalam model, seringkali diperlukan langkah pra-pengolahan atau transformasi data. PyTorch menyediakan modul `transforms` yang memungkinkan operasi seperti normalisasi, rotasi, dan skala pada data.

### **Manajemen Memori**

Salah satu tantangan dalam pembuatan dataset kustom adalah manajemen memori, terutama saat bekerja dengan dataset yang sangat besar. Penting untuk memastikan bahwa data yang tidak diperlukan segera dibuang dari memori untuk mencegah pemborosan sumber daya.

### Contoh Dataset Kustom

Sebagai contoh, bisa membuat dataset gambar kustom. Dalam `__init__`, akan mengumpulkan path dari semua gambar. Kemudian, dalam `__getitem__`, gambar di-load dan ditransformasi menjadi tensor, bersama dengan labelnya.

### Integrasi dengan `DataLoader`

Setelah dataset kustom dibuat, ia dapat diintegrasikan dengan `DataLoader` untuk efisiensi pelatihan. `DataLoader` memungkinkan pembuatan batch, pengacakan data, dan paralelisasi dalam pemuatan data.

### Kesimpulan

Pembuatan dataset kustom di PyTorch adalah proses yang penting dan serbaguna untuk aplikasi pembelajaran mesin. Melalui dataset kustom, pengembang dapat memanfaatkan jenis data yang unik dan spesifik untuk kebutuhan penelitian atau pengembangan mereka. Proses ini membutuhkan pemahaman yang baik tentang struktur data dan cara PyTorch mengelola data dan memori.