**Dokumentasi Proses**

Proyek ini bertujuan mengklasifikasikan jenis suara urban (seperti suara anjing menggonggong, klakson mobil, dsb.) menggunakan **Convolutional Neural Network (CNN)** dengan fitur **MFCC (Mel-Frequency Cepstral Coefficients)** sebagai representasi fitur audio. Dataset yang digunakan adalah **UrbanSound8K**, berisi lebih dari 8000 potongan audio berdurasi maksimal 4 detik.

Langkah-langkah utama:

1. **Import Library**: Digunakan librosa untuk ekstraksi audio, TensorFlow/Keras untuk deep learning, serta matplotlib dan seaborn untuk visualisasi.
2. **Load Metadata**: Dibaca dari file .csv, digunakan hanya data dari fold1–fold9 (80% data).
3. **Ekstraksi Fitur**: Audio dijadikan array MFCC berdimensi tetap (173, 40) dengan padding/trimming, kemudian diubah menjadi tensor 4D untuk CNN.
4. **Label Encoding**: Label diubah menjadi format one-hot vector menggunakan LabelEncoder dan to\_categorical.
5. **Split Dataset**: Data dibagi 80% untuk pelatihan dan 20% untuk pengujian dengan stratifikasi.
6. **Bangun CNN**: CNN terdiri dari 2 blok Conv2D + MaxPooling + Dropout, dilanjutkan Flatten dan Dense layer untuk klasifikasi.
7. **Training Model**: Model dilatih selama 30 epoch, diverifikasi dengan data validasi, dan disimpan sebagai file .h5.
8. **Evaluasi**: Akurasi model dan confusion matrix dievaluasi, serta dilakukan prediksi untuk 1 audio secara individu.

**Implementasi Operasi**

Operasi-operasi yang dilakukan:

* Ekstraksi sinyal audio dan padding jika durasi kurang.
* Ekstraksi fitur MFCC dengan n\_mfcc=40, padding matriks hasil agar seragam.
* Encode label menggunakan LabelEncoder.
* Pembagian dataset menggunakan train\_test\_split.
* Pembuatan model CNN dengan 2 layer konvolusional, dropout, flatten, dan softmax.
* Pelatihan menggunakan model.fit dengan validation\_data.
* Evaluasi akurasi dan kerugian (loss), prediksi model, dan visualisasi hasil.
* Menyimpan model .h5 dan melakukan prediksi file individual.

Tiap tahap diterapkan secara sistematis, menunjukkan penguasaan proses end-to-end pada audio classification.

**Visualisasi Hasil**

Model menyertakan berbagai visualisasi:

1. **Loss Curve** – menunjukkan penurunan loss training dan validasi yang stabil selama 30 epoch.
2. **Accuracy Curve** – memperlihatkan peningkatan akurasi secara bertahap hingga ~93% training dan ~82% validasi.
3. **Confusion Matrix** – visual interaktif yang menunjukkan performa per kelas (misal: dog\_bark vs car\_horn).
4. **Waveform** – visualisasi bentuk gelombang dari salah satu file audio contoh.
5. **Prediksi File Tunggal** – ditampilkan label hasil prediksi dan file sumbernya.

**Interpretasi & Analisis**

* Model CNN mampu mencapai **akurasi validasi 81.95%**, menunjukkan kemampuan cukup tinggi dalam membedakan suara urban.
* Dari confusion matrix, beberapa kelas sering tertukar, terutama antara suara-suara dengan karakteristik serupa (misal: air\_conditioner vs engine\_idling).
* Loss validation cukup stabil setelah epoch ke-10, walau sedikit fluktuatif, menunjukkan kemungkinan **early stopping** bisa dimanfaatkan untuk efisiensi.
* Hasil prediksi untuk file 102105-3-0-0.wav berhasil diklasifikasikan sebagai **dog\_bark**, menunjukkan performa model yang tepat pada kasus individual.

**Refleksi Pribadi**

Dalam proyek ini, saya memahami bahwa:

* Ekstraksi fitur audio memerlukan **konsistensi dimensi** agar dapat diproses oleh CNN.
* **Padding dan trimming** sinyal sangat krusial untuk menyamakan input time-step.
* **Visualisasi** memberikan wawasan penting untuk debugging model dan memahami performa.
* Dibandingkan MLP, CNN jauh lebih efektif dalam menangkap pola spasial dari fitur MFCC.
* Tantangan terbesar adalah menyeimbangkan jumlah data antar kelas dan menghindari overfitting.

Pengalaman ini memperkuat pemahaman saya tentang **deep learning untuk data audio**, serta pentingnya preprocessing yang cermat untuk keberhasilan model.