Audio Classification Project Report

1. Dataset Organization

- Audio files grouped into folders by class (class_1, class_2, ...).
- Each folder represents a unique sound category.

2. Preprocessing

- Audio loaded using librosa.
- Resampled to 22,050 Hz.
- Padded/truncated to 5 seconds.
- Converted into mel spectrograms.
- Normalization applied.
- SpecAugment used for data augmentation.

3. Model Architecture

- CNN with 3 convolutional blocks.
- MaxPooling for downsampling.
- AdaptiveAvgPool + Linear layer for classification.

4. Training

- Optimizer: Adam
- Loss: CrossEntropy
- Epochs: 12
- Batch size: 16
- Model saved as cnn_audio_model.pth

5. Evaluation

- Accuracy calculated from predictions.
- Classification report generated.
- Confusion matrix plotted.

6. Project Structure

- utils.py: audio loading + spectrogram + augmentation

- dataset.py: PyTorch dataset loader

- model.py: CNN model

- train.py: training loop

- evaluate.py: evaluation logic

- main.py: full pipeline runner

7. Outcome

- Generates a trained model for audio classification.

- Ready for deployment or inference.