

## THỰC HÀNH BUỔI 5

(Data Collection and Preprocessing)

**Bài 1:** Bộ dữ liệu [Urban Sound 8K](#) trong bài này. Nó bao gồm 10 classes là các loại âm thanh khác nhau như tiếng chó sủa, tiếng còi báo động, tiếng máy khoan etc.

Sau khi tải về, chúng ta thấy bộ dữ liệu này gồm 2 phần:

- Các Audio files trong 10 sub-folders có tên từ *fold1* đến *fold10*. Trong mỗi sub-folders đó đều chứa các Audio files. Ví dụ: *fold1/103074-7-1-0.wav*. Độ dài của mỗi file Audio khoảng 4s.
- File *UrbanSound8K.csv* chứa thông tin về mỗi Audio files trong bộ dữ liệu: Tên file, nhãn, ... Nhãn của mỗi file Audio được quy định là theo ID từ 0 đến 9.

**Yêu cầu:** Sinh viên thực hiện các bước tiền xử lý dữ liệu chuyển sang dạng Spectrogram (hoặc *Mel Spectrogram*, hoặc *MFCC*, hoặc *STFT*) (có thể tham khảo trên website: [https://tiensu.github.io/blog/70\\_audio\\_deep\\_learning\\_part\\_4/](https://tiensu.github.io/blog/70_audio_deep_learning_part_4/) )

**Bài 2:** Từ tập dữ liệu Video, tiến hành chuyển đổi dữ liệu này thành dạng văn bản để có thể thực hiện các phân tích và xử lý dữ liệu.

- Sử dụng thư viện *pydub* hỗ trợ mạnh mẽ để xử lý âm thanh. Sử dụng lớp *AudioSegment* từ thư viện *pydub* để chuyển đổi video thành định dạng âm thanh.
- Sau khi đã có âm thanh từ video, sử dụng thư viện *speech\_recognition* để thực hiện nhận dạng giọng nói và chuyển đổi nó thành dạng văn bản (Để tối ưu hóa quá trình chuyển đổi và tiết kiệm thời gian, sử dụng lớp *ThreadPoolExecutor* từ thư viện *concurrent.futures*)