**Assignment**

**Intro :**

This dataset contains 8732 labeled sound excerpts (<=4s) of urban sounds from 10 classes: air\_conditioner, car\_horn, children\_playing, dog\_bark, drilling, enginge\_idling, gun\_shot, jackhammer, siren, and street\_music. The classes are drawn from the urban sound taxonomy.

The files are pre-sorted into ten folds (folders named fold1-fold10) to help in the reproduction of and comparison with the automatic classification results. In Addition, to the sound excerpts, a CSV file containing the metadata about each excerpt is also provided.

### **Methodology :**

1. There are 3 basic methods to extract features from audio file :

a) Using the mffcs data of the audio files .

b) Using a spectogram image of the audio and then converting the same to datapoints.

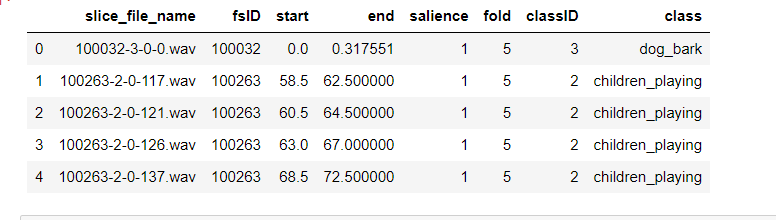
c) Combining both features to build a better model.

1. I have chosen to use the second method.
2. The labels have been converted to categorical data for classification.
3. CNN has been used as the primary layer to classify data.

**Importing the libraries :**

**Analysing Data Type And Format :**

**Analysing the csv data :**

****

##### ***Column Names***

* slice\_file\_name: The name of the audio file. The name takes the following format: [fsID]-[classID]-[occurrenceID]-[sliceID].wav, where: [fsID] = the Freesound ID of the recording from which this excerpt (slice) is taken [classID] = a numeric identifier of the sound class (see description of classID below for further details) [occurrenceID] = a numeric identifier to distinguish different occurrences of the sound within the original recording [sliceID] = a numeric identifier to distinguish different slices taken from the same occurrence
* fsID: The Freesound ID of the recording from which this excerpt (slice) is taken
* start The start time of the slice in the original Freesound recording
* end: The end time of slice in the original Freesound recording
* salience: A (subjective) salience rating of the sound. 1 = foreground, 2 = background.
* fold: The fold number (1-10) to which this file has been allocated.
* classID: A numeric identifier of the sound class: 0 = air\_conditioner 1 = car\_horn 2 = children\_playing 3 = dog\_bark 4 = drilling 5 = engine\_idling 6 = gun\_shot 7 = jackhammer 8 = siren 9 = street\_music
* class: The class name: air\_conditioner, car\_horn, children\_playing, dog\_bark, drilling, engine\_idling, gun\_shot, jackhammer, siren, street\_music.

**Using Librosa to analyse Random Sound Sample - Spectogram :**

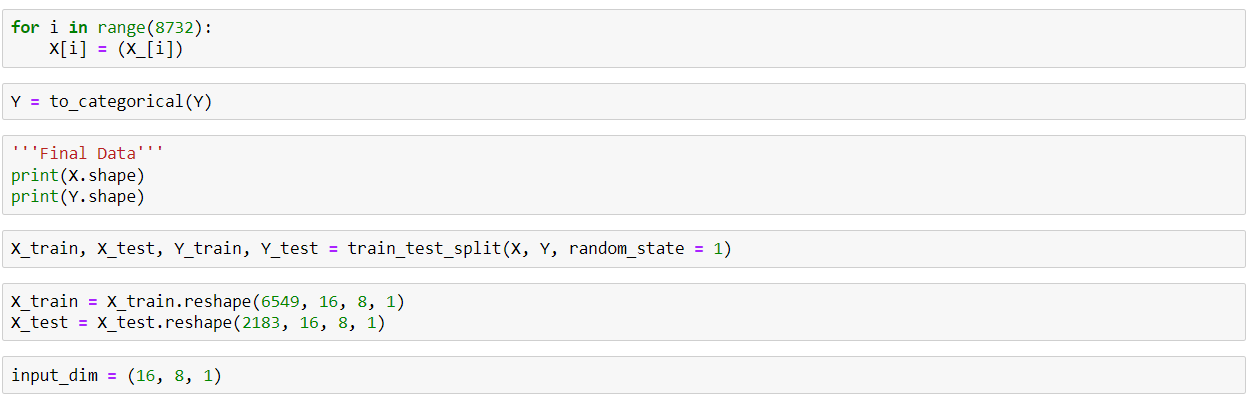


**Feature Extraction And Database Building :**

**Method**

1. I have used Librosa to extract features.
2. To do so, I will go through each fold and extract the data for each file. Then I have used the mel\_spectogram function of librosa to extract the spectogram data as a numpy array.
3. After reshaping and cleaning the data, 75-25 split has been performed.
4. Classes (Y) have been converted to Categorically Encoded Data usng Keras.utils.

****

****

**Creating Keras Model And Testing:**

**Model 1 :**

1. CNN 2D with 64 units and tanh activation.
2. MaxPool2D with 2\*2 window.
3. CNN 2D with 128 units and tanh activation.
4. MaxPool2D with 2\*2 window.
5. Dropout Layer with 0.2 drop probability.
6. DL with 1024 units and tanh activation.
7. DL 10 units with softmax activation.
8. Adam optimizer with categorical\_crossentropy loss function.

90 epochs have been used.

****