Project Report

DLOps

Ayush Abrol
Neehal Prakash Bajaj
Aryan Tiwari
11 May 2023

Table of Content

Overview

Audio classification is a challenging task in the field of deep learning. It involves learning to classify sounds and predict the category of that sound. This type of problem can be applied to many practical applications such as noise monitoring, speech recognition, and music genre classification. In this project, we aim to classify urban sounds using the urbansound_8k dataset. The dataset contains 8732 labeled sound excerpts of urban sounds from 10 classes.

Dataset

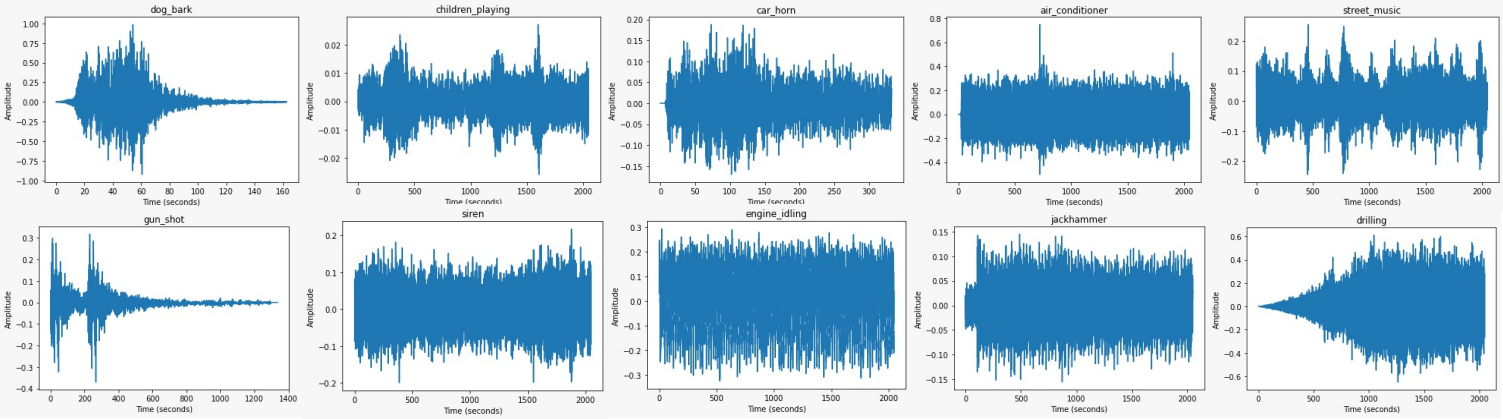
UrbanSound8K contains over 8000 sound files separated by categories of sounds typically found in an urban setting. Instead of exclusive categorical labels in its original state, the dataset has been recreated with the purpose of exploring how a successful architecture might perform on multi-label samples rather than simply uni-label, multi-class sounds. Thus, this notebook investigates how techniques in classifying an environmental noise data might generalize to a Multi-Label scenario with a new Audio data composed of overlaid sound pairs. The original dataset, UrbanSound8K, contains 8732 .wav sounds sourced from Freesound.org across 10 classes of urban sounds. These audio files were saved across 10 different folds, with each fold containing audio files from different classes, the distribution of the files in the each folder was done such a way that the CSV file contained the name of file, the fold in which it is located and its respective label. :

- Air Conditioner
- Car Horn
- Children Playing
- Dog Bark
- Drilling
- Engine Idling
- Gun Shot
- Jackhammer
- Siren
- Street Music

These audio files were saved across 10 different folds, with each fold containing audio files from different classes, the distribution of the files in the each folder was done such a way that the CSV file contained the name of file, the fold in which it is located and its respective label.

Pre-Processing & Visualisation

Visualisation



Pre-Processing

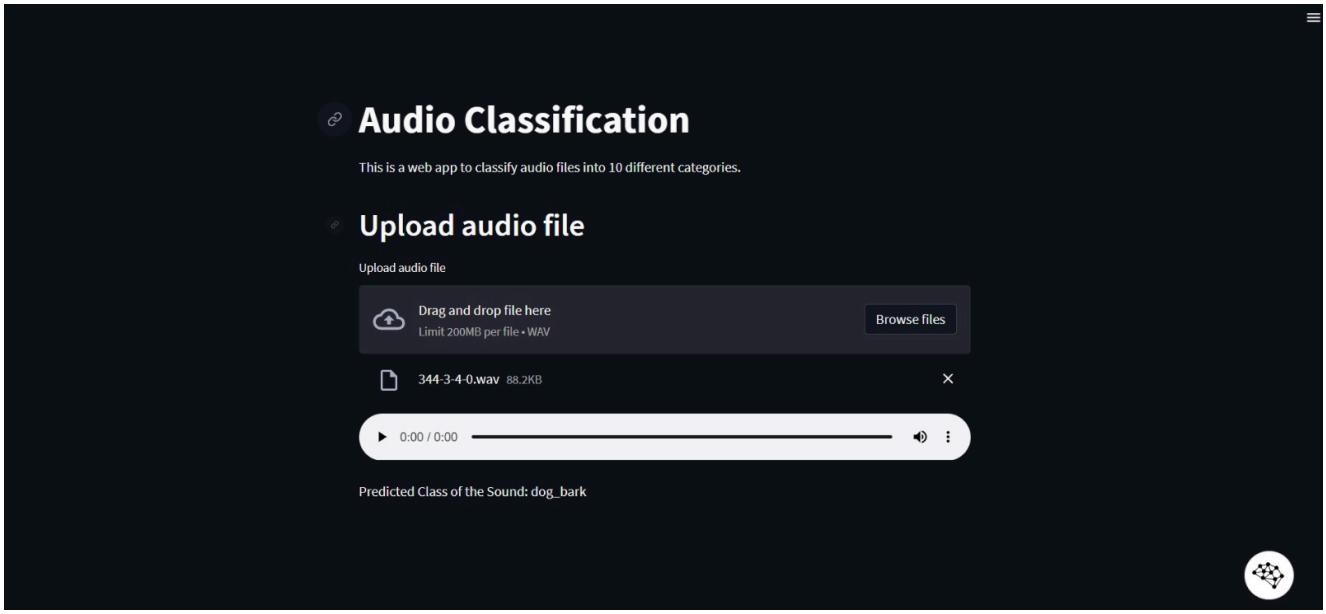
- For audio samples, one of the main features involved is the power spectrum. One type of power specturm is called the Mel Frequency Cepstrum.We calculate the coefficients for Mel Frequency Cepstrum (MFCC) using librosa.
- Take the mean of the mfccs for each of the frame and thereby audio samples into data
- Converting the features and labels to a tensor dataset
- Finally spitting into train and test sets

MFCC stands for Mel Frequency Cepstral Coefficients. It is a commonly used feature extraction technique in speech and audio processing.The MFCC algorithm involves several steps:
a)Pre-emphasis: A high-pass filter is applied to the signal to amplify the high-frequency components.
b)Framing: The signal is divided into small frames of equal duration.
c)Windowing: A window function is applied to each frame to reduce spectral leakage.
d)Fast Fourier Transform (FFT): The power spectrum of each frame is calculated using the FFT.
e)Mel Frequency Wrapping: The power spectrum is converted into the Mel frequency scale, which is a non-linear scale that is more closely related to human perception of sound.
f)Cepstral Analysis: The logarithm of the Mel power spectrum is calculated and then transformed using a Discrete Cosine Transform (DCT) to obtain a set of coefficients called Mel Frequency Cepstral Coefficients (MFCCs).

Model and Training

- **Model A:-** We have implemented a MultiLayerPerceptron model for audio classification purposes here which uses 3 fully connected layers
- **Model B:-** We have implemented a MultiLayerPerceptron model for audio classification purposes here which uses 5 fully connected layers

Depolyment and Prediction



Predicting Classes for different user intended audio files

References:-

- **Dataset:-** https://urbansounddataset.weebly.com/download-urbansound8k.html
- **Reference:**- https://towardsdatascience.com/urban-sound-classification-using-neural-networks-9b6fcd8a9150