By: Marwan Harajli
Last Edit:10/15/2018
Subject: Capstone Project 1 Proposal.

Urban Sound Classification

Problem Statement:
Classifying a given .wav sound file as one of the following noises: Air conditioner, car horn, children playing, dog barking, drilling, engine idling, gun shot, jackhammer, siren, and street music.

Project Importance:
This project presents itself as an educational opportunity in tackling unstructured data. Since the data is unstructured, useful features need to be found and extracted. Moreover, due to the nature of the unstructured data (being audio data), the project also presents an opportunity to delve into audio processing. The latter has a wide array of applications that ranges from speech processing to recommending music to radio stations. With that, the project yields the potential to be a good data science exercise that also opens the door to the exciting field of audio processing.

Potential Client:
A possible client can be the local government who plans to use microphones located in congested areas to identify an emergency situation (such an gunshot, an explosion, ambulance siren, etc.) and accordingly send first responders or other relevant parties.

The Dataset:
A dataset of 8732 .wav files (each representing one of the 10 noises mentioned above) can be downloaded from the link below:

https://datahack.analyticsvidhya.com/contest/practice-problem-urban-sound-classification/

Each .wav files is less than 4s (length is not fixed over all sound files), and around 5400 of the files have been labeled (i.e. they can be used for a training a classification model). The label of the remaining wav files is not explicitly given, but a solution checker is given in the link above (so we can check accuracy of classification algorithm).

Methodology:
Every .wav file is to be converted into a numpy array that represents the sound as a time series (noise amplitude vs. time). This can be done using either the librosa or scipy libraries. Librosa, a python package for sound analysis, will then be used to extract features from these numpy arrays. These features will include: mel-scaled spectrogram, RMS energy for each time frame,

roll-off frequency, and many more. A full list of features that can be extracted using librosa can be found in the link below:

https://librosa.github.io/librosa/feature.html

These features will then be used to train a classification model. Several models will be tested (SVM, ANN, GMM, and others).

Deliverables:

A report is to be written that summarizes:

1. The process of feature extraction.
2. Visualisations of these features and an assessment of their relevance.
3. A description of each classification model used.
4. A comparison of the performance of each classification model used.

The report will also have an appendix that includes and discusses the code used to classify the sounds.

The files for this project can be found in the following GitHub repository:

https://github.com/harajlim/Urban_Noises