

Prepared By: Marwan Harajli

Last Edit: 2018-09-26

Subject: Springboard Bootcamp Capstone 1 Potential projects

## **Potential Capstone Projects:**

Note: These projects are listed in order of preference. They are also ordered in decreasing difficulty (where the level of difficulty is a purely subjective thing at the moment).

### **1- Identifying Urban Sounds:**

A dataset of 8732 sound clips each 4s long is given in the link below:

[https://datahack.analyticsvidhya.com/contest/practice-problem-urban-sound-classification/?utm\\_source=auto-email](https://datahack.analyticsvidhya.com/contest/practice-problem-urban-sound-classification/?utm_source=auto-email)

Each of these clips represents one of the following urban sounds: Air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren, and street music.

Features (discussed below) are to be extracted from each sound clip and used to train a classification model. Different classification models are to be compared for accuracy.

I have started tackling this problem and you can monitor my progress on:

[https://github.com/harajlim/Urban\\_Noises](https://github.com/harajlim/Urban_Noises)

### **2- Women's E-commerce:**

23500 reviews for various items are given in the dataset below:

<https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>

The dataset contains the text for the review, a rating out of 5, the ID of the item being reviewed, the department that sells the reviewed items, reviewer age, and other descriptors.

A fun exercise would be to try and guess the rating based on the text (using something like a Naive Bayes classifier).

### **3-MNIST Number Recognition:**

MNIST provides a database of 50,000+ images. Each image shows a single handwritten digit (0,1,2...9). The database does provide what the number is in each image. A classification algorithm can thus be trained and tested.

I have tackled this problem using Neural networks following the book in the link below and gotten good results. It would be interesting to test other classification algorithms and compare efficiency.

<http://neuralnetworksanddeeplearning.com/chap1.html>