

Efraim Herstic

Fall 2020

MAIS 202 - PROJECT DELIVERABLE 1

Project Idea

The idea behind this project is to create a web-app that, given a ten-second audio clip input, can predict the classification of the inputted sound. I plan on using the google research AudioSet dataset (<https://research.google.com/audioset/>) to train, validate, and test my model to classify different ten-second sound clips. I am choosing to use this dataset because of its scale and variety of annotated sound clips, classified according to over 527 labels. I do not plan to use all of those classes, for multiple reasons. Firstly, some of the labels have low sound quality estimates, and so I will pre-process my data to only include categories of sounds with 100% quality estimates. This includes almost 80 different categories of sounds, with around a couple thousand sound clips using each of those labels. Furthermore, during pre-processing, I plan to limit the number of labels from the Music and Speech labels, to only a few thousand, instead of the around 1 million clips that each of these categories contains, in order to more equally balance the number of clips with different labels in my dataset. Since this proposed project can be classified as a classification problem with many different labels, I plan to use a multi-nomial logistic regression model. I plan to utilize a Confusion matrix and accuracy metrics on each classifier as my evaluation metrics for this model. Finally, after training, validating and testing my model, I plan on integrating my model into a simple landing-page webapp, where users are able to upload a ten-second audio clip, and my model will predict the classification of that sound clip according to some of the around 80 labels I plan on training it on.