This repo is about a concept study on the MIMII dataset to detect anomaly of machines or machine parts like fans, slider, pump and valves by means of machine learning and deep learning methods.

In condition monitoring of machinery, it is common to use structure-borne sound and order tracking (rpm, etc.) to detect malfunctions, for various reasons like ease of retrofitting or need for a mobile solution and size of the machine part or operational needs like zero downtime, airborne sound anomaly detection could be preferable. The proof of concept study conducted here, shows that by applying machine learning anomaly detection on acoustic sensing, an AI sensor can be constructed the has similar performance as trained technician when detecting anomalous by listen to the machine, with the benefit of staying 24/7 at the machine part of interest. The development can be seen as ground work for an embedding solution of smart sensor into an IoT plant supervision system.

Anomaly detection with machine learning means mostly unsupervised learning as the base assumption is that none normal operation is unknown, is just different from normal operations and will sound different. This means a smart senor microphone system needs to be trained by being placed for a reasonable training time in front of a healthy machine part to then be able to become an armed detector.

would be placers the following techniques are explored:

* Stochastic model (a multi-dimensional normal distribution is found and outlier defined by significant)
* Random Isolation trees (a decision tree depth is taken assuming outlier need only view decisions to be found as spatial separated in one or more feature)
* Auto encoder (an underrepresented auto encoder reconstruction error is taken)
* Pseudo supervision (where normal observation is augmented/ distorted to train a binary classifier)

In order to make any machine learning algorithm able to work with audio it is also necessary to use various signal processing steps that may be of classic nature or also take use of machine learning methods like clustering.