

Anomaly Detection System (For Unsupervised Machine Learning):

This document explains the method that I have used for the Anomaly Detection. The method used is primarily density-based method for anomaly detection but there one key change. For the purpose of threshold measuring, you should either have labeled data or some prior information about the expected Anomaly rate(Maybe you are selling some product and your customers complaint for the degraded quality of some of your products). So you know that on general, you expect X% of your products to be anomalous. You use this information and after finding densities of all the data points using pdf for normal distribution(using pdf for normal distribution because we normalize the data first for this density- based method), we use percentiles for determining the threshold.

For example: Let's say 5% anomalies are expected. So when we compute the densities, we use 5th percentile to find the density below which lies our 5% data and we set this density value as our threshold value.

This method comes as a rescue when we don't have labeled data for determining threshold (using true positives and false negatives).

Limitations: One limitation for this method is that the accuracy of the model cannot be specifically determined because we only apply this method where we don't have labeled data. This leaves us with nothing to compare with to determine accuracy. Yet, the model is a strong tool for Anomaly Detection and has been proved successful in the test runs.