**Anomaly Detection Approach:**

**Approach:**

I have used a moving average window method in building the anomaly detection algorithm. I calculated the Moving Average Using Discrete Linear Convolution. The purpose of choosing moving window is that it smooths the short term fluctuations and highlights the long term ones.

The idea is to keep a window of the last few data points and calculate the average and the standard deviation.

If abs(data-avg) > std\*threshold

Anomalous

Else

Pass

After calculating the anomalous points, I removed the noise from these calculated data points. I considered any contiguous data points less than five as noise. After removing the noise, I grouped the anomalous points per file and printed the result on the console.

After trying the various size of moving windows and threshold, I received the best results when I used a window of length 60 and 4 as the threshold.

**Limitations:**

This method will not perform well when we have data with noise which might be similar to abnormal behaviour, because the boundary between normal and abnormal behaviour is often not precise.

**Next step [Prediction]:**

After reading many articles about time series data prediction, I came to know that Classification and regression trees (CART) are most effective machine learning techniques for anomaly detection. This will require to have labeled anomaly data points. Other way would be to try unsupervised learning and teach CART to predict next data point and have some confidence interval and prediction error.

Additionally, Recurrent Neural Networks may work well for given dataset. RNN with LSTM cells can keep track of previous values and it can make predictions based on it. We can feed last fifteen minutes of data to model so that it can understand the current pattern and make predictions accordingly.

-------------------------------------------------------------------------------------------------------------------------------