# **Synthetic data generation:**

As it is known collection of anomaly data is one of the challenging phase when carrying out predictive maintenance of the motors. Lifetime of the motors is so long and only 1 or 2 times problem can happen between 2-3 years. That`s why in this project I did my own anomaly, namely synthetic data generation. While doing researches I figured out that if we assume normal distribution in this case, the data that varies equal or more than 2 sigma from mean should be considered anomaly, as it is only 5%. In this part I did the same, but trick I did is to use our normal data to generate anomaly data. Despite, theory is known for generating synthetic data, reference to create is also difficult. In my case, I am getting one of the files and generating anomaly out of it.

# **Feature extraction:**

Once we have both normal and abnormal data, next step needs to be done is feature extraction. Signal processing is applied here and different type of variables is derived from the main data. The information varies from mean of data to waveform factor. Each file contains a lot of data. Nevertheless, 2048 data represents one time reading in case of adxl362, because this sensor reads 1024 value in 3 dimensions. That`s why after reading data I am separating file by 1024, and extracted features out of them. Example, one file contains 113664 data which is 111 separate reading. Feature extraction will be carried out by using 111 separate reading. 13 feature is derived per separate reading. Which makes up, 111\*13 variables.

Once previous step is done, I did test all the variables separately and combination of all of them to find which, combination gives best result. There are multiple findings I would like to not:

* Mean value itself gives reasonable result, as it contains reliable information about the motor.
* Mean value and skewness give anomaly after 4 std difference which is a little bit late. That`s why it needs to be improved
* Mean value, skewness and kurtosis works less powerful than previous one. It means that kurtosis is not powerful indicator to test the anomaly. But still useful.
* Best operation so far conducted by mean value, skewness and rms value. It gives anomaly exactly proposed time. As soon as vibration data hits 2 std value it means that anomaly is going to happen.
* Pulse index mixed with previous one gives bad result as skewness.

* Waveform factor is also one of the powerful indicator. While mixed with skewness mean and rms it even improves the signal results and removes most of the false alarms.

So as a result Waveform, mean, skewness, rms and margin factor combination gives perfect result as it reduces false alarm namely, improves efficiency of the algorithm and indicates all the anomaly cases which will result in motor failure.

All in all, I decided to use mentioned combination to achieve the best. Furthermore, what I found is, instead of merging 3 dimensional data to 1 dimension and doing feature extraction is not good way of handling it. Getting features for 3 dimension separately and merging them gives perfect result. That`s why the DataFrame variable I am using contains features for x,y,z axis.

# **ML implementation:**

For ML implementation I utilized two different algorithms: Isolated Forest and One class Support Vector Machine. Both are unsupervised learning so there is no need to label the data. After careful analysis, I came to conclusion that Isolated Forest algorithm gives best result.

# **Conclusion**

All in all, data is collected, processed, feature extracted. Research was done on finding best combination of the indicators, and algorithm. The best performing combination is: Waveform, mean, skewness, rms and margin factor. The best performing algorithm is Isolated Forest.