Appendix for manuscript titled-"Papillon: a real-world prognostic toolkit for improving operational efficiency in Industry 4.0"

March 2, 2021

This appendix contains detailed description for the data-sets and the experiments that were carried out. Two data-sets are used for analyzing the prognostic capabilities of Papillon. Following is a detailed description of the data-set used.

1 Pump Data-set for an oil and gas upstream operation[1]

This data set comprises of the sensor data associate with 2 pumps that are installed in an upstream operation in oil upstream rig. The ids of these pumps are Pump-6 and Pump-7. Pump-6 data consist of 34 sensors data separated by a 3 min interval. Pump-7 data on the other hand consists of 35 sensor's data separated by a time interval, 3-minutes apart. For further analysis the data has been aggregated at an hourly level.these sensors measure various physical parameters of the machine like bearing temperature, Vibration etc. There is a column by the name *key-phasor*, associated with the data of both the pumps. If this parameter takes a value equal to zero then this represents that the machine is shut-down.

This shutdown can be a case of planned down-time or it can be because of normal machine failure. The instantaneous values of the sensor data 7 days prior to the instance when the key-phasor is down, are given the label 1, when the key-phasor remains down for a prolonged period, then in such cases the data is given the label-2 whereas, if the key-phasor is up, the data points are given the label-0 . Label-0 represents the normal functioning machinery, label-1 represents that the machine is in failing condition (or about to fail in 7 days) lastly, Label-2 represents that machine is not operational. For Prognostic applications, data corresponding to label-1 is of prime importance since, one would want the indication of failure few(7) days prior. After removing the data corresponding to label-2, the number of positive (or failing) class remaining in the data-set for Pump-6 are 657 out of total 6,672 points and for Pump-7 it is 331 out of 9072

[1]This data-set is private

total data points. As evident from the data the class corresponding to label-1 is very rare.The aim here is to detect the failing points in the time-series one week prior.

2 Pump sensor data for a smart city water distribution network

This is the data corresponding to water pump of a small area there are 7 system failures in the entire data-set. The data is a multivariate time-series containing data from 52 sensors.This is a public data-set available on Kaggle (Please visit `https://www.kaggle.com/nphantawee/pump-sensor-data` for more details). The labels in this data are given in the form of a string namely, *NORMAL*, *BROKEN*, *RECOVERING*. Values of the data-set 3 days prior to the state *BROKEN* are assigned the label-1, Values of the data-set Corresponding to the state *RECOVERING* are assigned the label-2 and rest of the point are assigned the label-0. In this case after removing the data points corresponding to label-2, the number of total data corresponding to label-1 is 7.67% in the entire data-set.The aim here is to detect the class-1, 7 days prior.

3 Algorithms Used

The occurrence of a failing class of data with label-1 is a rare phenomenon so it is important for an AI-based system to learn the pattern hidden in the data during normal mode of machine functioning so that when the model. Keeping this in mind both supervised and Unsupervised algorithms are used in the current study. Instead of using raw time-series data as the input for the model the quantiles of the *Haar* transform are used as an input features. The details of both types of algorithms are given below.

3.1 Supervised Learning

Four different binary classifiers are tested for their performances namely, Logistic regression, Random forest classifier, Support vector machines and XG-Boost Classifier. Finally XG-boost classifier with 200 gradient-boosted-decision trees gave the best performance for the 3 different data-sets. These classifiers were trained on the initial 70% of the time-series consisting of label-1 and label-0. The performance of the Algorithms are then tested using the entire data-set the detailed results are given in Table-1.

3.2 Unsupervised Learning

In this approach the data corresponding to the label-0 is used for training the models and the performance of these trained models are tested using the entire data-set. Isolation Forest, One-class SVM and Generative adversarial net-

Type of Algorithm	Data-set	Model-name	Precision	Recall	AUC-score
Supervised	ONG-PUMP-6	XG-BOOST	76.12	94.06	0.95
	ONG-PUMP-7	XG-BOOST	8.76	52.55	0.53
	Kaggle-PUMP-data-set	XG-BOOST	0	0	0.5
Unsupervised	ONG-PUMP-6	Isolaation-Forest	13.5	48.47	0.48
	ONG-PUMP-7	Isolaation-Forest	0	0	0.5
	Kaggle-PUMP-data-set	Isolaation-Forest	99.8	1	0.994
	ONG-PUMP-6	One-Class-SVM	8.76	54.09	0.532
	ONG-PUMP-7	One-Class-SVM	3.5	52.6	0.513
	Kaggle-data-set	One-Class-SVM	10.6	77.6	0.638
	ONG-PUMP-6	MLP-MLP GAN	**15.4**	**56.7**	**0.634**
	ONG-PUMP-7	MLP-MLP GAN	**8.8**	**100**	**0.803**
	Kaggle-PUMP-data-set	MLP-MLP GAN	**32.5**	**35.1**	**0.63**

Table 1: Results

works(GANs) are used for for testing in this approach. Results for all the 3 data-sets with Isolation Forest having 300 decision trees are shown in the table-1. Results for all the 3 data-sets with One-class SVM having linear kernel are shown in the table-1. Different Architecture of GANs are also used in this study. The GAN architecture with MLP as discriminator and generator network gave the best results compared to the other architecture where the discriminator and generator network were CNN or Bi-LSTM. the results for this GAN architecture are shown in the table-1.

4 Conclusion

The GAN model with MLP as generator and discriminator have a decent metrics for all the 3 data-sets considered in the study. Other model wile giving good results for on kind of a machine data can not be generalized for all three machines