CLAIMS

1. A computer-implemented method comprising:

inputting a new data sample to a failure prediction model, wherein the failure prediction model is trained using a labeled historical dataset, wherein respective data points are associated with a look-back window and a prediction horizon to create respective training samples, wherein the respective training samples are clustered in a plurality of clusters, and wherein the plurality of clusters are each associated with a normalcy score and an anomaly score; and

outputting a classification associated with the new data sample based on comparing a first anomaly score of a first cluster of the plurality of clusters that includes the new data sample to an average anomaly score of clusters of the plurality of clusters having the normalcy score greater than the anomaly score, wherein the classification is indicative of a likelihood of future failure or performance degradation of an asset associated with the new data sample.

2. The method of claim 1, wherein the classification is indicative of a likelihood of wear-related performance degradation of the asset associated with the new data sample.

3. The method of claim 1, wherein the look-back window defines a quantity of sequentially previous data points to include in each respective training sample.

4. The method of claim 1, wherein the prediction horizon defines a predefined amount of time in the future, and wherein respective labels of respective data points of the predefined amount of time in the future are associated with the respective training samples.

5. The method of claim 1, wherein the respective training samples are clustered using K-Means clustering.

6. The method of claim 1, wherein the method is performed by a failure prediction system according to software that is downloaded to the failure prediction system from a remote data processing system.

7. The method of claim 6, wherein the method further comprises:

metering a usage of the software; and

generating an invoice based on metering the usage.

8. A system comprising:

one or more processors; and

one or more computer-readable storage media storing program instructions which, when executed by the one or more processors, are configured to cause the one or more processors to perform a method comprising:

inputting a new data sample to a failure prediction model, wherein the failure prediction model is trained using a labeled historical dataset, wherein respective data points are associated with a look-back window and a prediction horizon to create respective training samples, wherein the respective training samples are clustered in a plurality of clusters, and wherein the plurality of clusters are each associated with a normalcy score and an anomaly score; and

outputting a classification associated with the new data sample based on comparing a first anomaly score of a first cluster of the plurality of clusters that includes the new data sample to an average anomaly score of clusters in the plurality of clusters having the normalcy score greater than the anomaly score, wherein the classification is indicative of a likelihood of future failure or performance degradation of an asset associated with the new data sample.

9. The system of claim 8, wherein the classification is indicative of a likelihood of wear-related performance degradation of the asset associated with the new data sample.

10. The system of claim 8, wherein the look-back window defines a quantity of sequentially previous data points to include in each respective training sample.

11. The system of claim 8, wherein the prediction horizon defines a predefined amount of time in the future, and wherein respective labels of respective data points of the predefined amount of time in the future are associated with the respective training samples.

12. The system of claim 8, wherein the respective training samples are clustered using K-Means clustering.

13. A computer program product comprising one or more computer readable storage media, and program instructions collectively stored on the one or more computer readable storage media, the program instructions comprising instructions configured to cause one or more processors to perform a method comprising:

inputting a new data sample to a failure prediction model, wherein the failure prediction model is trained using a labeled historical dataset, wherein respective data points are associated with a look-back window and a prediction horizon to create respective training samples, wherein the respective training samples are clustered in a plurality of clusters, and wherein the plurality of clusters are each associated with a normalcy score and an anomaly score; and

outputting a classification associated with the new data sample based on comparing a first anomaly score of a first cluster of the plurality of clusters that includes the new data sample to an average anomaly score of clusters in the plurality of clusters having the normalcy score greater than the anomaly score, wherein the classification is indicative of a likelihood of future failure or performance degradation of an asset associated with the new data sample.

14. The computer program product of claim 13, wherein the classification is indicative of a likelihood of wear-related performance degradation of the asset associated with the new data sample.

15. The computer program product of claim 13, wherein the look-back window defines a quantity of sequentially previous data points to include in each respective training sample.

16. The computer program product of claim 13, wherein the prediction horizon defines a predefined amount of time in the future, and wherein respective labels of respective data points of the predefined amount of time in the future are associated with the respective training samples.

17. The computer program product of claim 13, wherein the respective training samples are clustered using K-Means clustering.

18. A computer-implemented method for predicting wear-related deterioration of progressing cavity pumps (PCPs), the method comprising:

inputting a new data sample of a PCP to a model configured to predict wear-related deterioration of the PCP, wherein the model is trained using a labeled historical PCP dataset, wherein respective data points are associated with a look-back window and a prediction horizon to create respective training samples, wherein the respective training samples are clustered in a plurality of clusters, and wherein the plurality of clusters are each associated with a normalcy score and an anomaly score; and

outputting a classification associated with the new data sample based on comparing a first anomaly score of a first cluster of the plurality of clusters that includes the new data sample to an average anomaly score of clusters in the plurality of clusters having the normalcy score greater than the anomaly score, wherein the classification is indicative of the wear-related deterioration of the PCP.

19. A computer-implemented method for predicting wear-related deterioration of progressing cavity pumps (PCPs), the method comprising:

generating labeled historical data by performing binary labeling of historical data associated with one or more PCPs;

generating a plurality of training data samples by applying a look-back window and a prediction horizon to respective data points of the labeled historical data;

clustering the plurality of training data samples into a plurality of clusters;

calculating cluster scores for respective clusters of the plurality of clusters;

assigning a new data sample of a PCP to a first cluster of the plurality of clusters; and

assigning a classification to the new data sample based on cluster scores associated with the first cluster, wherein the classification is indicative of a likelihood of future wear-related deterioration of the PCP.

20. The method of claim 19, wherein the labeled historical data is labeled as faulty for a predetermined period of time prior to a known pump replacement date.

21. The method of claim 19, wherein the labeled historical data comprises pump speed data, pump torque data, casing pressure data, production rate data, and maintenance records.

22. The method of claim 19, wherein calculating the cluster scores for the respective clusters further comprises:

calculating a normalcy score for the first cluster, wherein the normalcy score is a first proportion of training data samples associated with a normal state in the first cluster divided by a second proportion of training data samples associated with the normal state in the plurality of training data samples; and

calculating an anomaly score for the first cluster, wherein the anomaly score is a third proportion of training data samples associated with a deteriorated state in the first cluster divided by a fourth proportion of training data samples associated with the deteriorated state in the plurality of training data samples.

23. The method of claim 22, wherein the classification is based on a larger value of the normalcy score or the anomaly score for the first cluster.

24. The method of claim 19, the method further comprising:

generating a failure signal for the new data sample, wherein the failure signal comprises an average anomaly score for the new data sample over a predetermined number of prior data points.

25. The method of claim 24, wherein generating the failure signal further comprises:

calculating a mean anomaly score for clusters of the plurality of clusters having a normalcy score greater than an anomaly score;

for each of the predetermined number of prior data points, associating a one value to data points having an anomaly score of the first cluster greater than the mean anomaly score, and associating a zero value to data points having an anomaly score of the first cluster less than the mean anomaly score; and

calculating the failure signal as an average of the one values and zero values associated with each of the predetermined number of prior data points.