A Research Study on Unsupervised Machine Learning Algorithms for Early Fault Detection in Predictive Maintenance

Nagdev Amruthnath
Department of IEE and EDMM
Western Michigan University
Kalamazoo, Michigan, USA
nagdev.amruthnath@wmich.edu

Tarun Gupta
Department of IEE and EDMM
Western Michigan University
Kalamazoo, Michigan, USA
tarun.gupta@wmich.edu

Abstract— **the area of predictive maintenance has taken a lot of prominence in the last couple of years due to various reasons. With new algorithms and methodologies growing across different learning methods, it has remained a challenge for industries to adopt which method is fit, robust and provide most accurate detection. Fault detection is one of the critical components of predictive maintenance; it is very much needed for industries to detect faults early and accurately. In a production environment, to minimize the cost of maintenance, sometimes it is required to build a model with minimal or no historical data. In such cases, unsupervised learning would be a better option model building. In this paper, we have chosen a simple vibration data collected from an exhaust fan, and have fit different unsupervised learning algorithms such as PCA T^2 statistic, Hierarchical clustering, K-Means, Fuzzy C-Means clustering and model-based clustering to test its accuracy, performance, and robustness. In the end, we have proposed a methodology to benchmark different algorithms and choosing the final model**

Keywords- Predictive maintenance, Fault detection, manufacturing, machine learning, Just in Time

I. Introduction

The concept of predictive maintenance (PdM) was proposed a few decades ago. PdM is also a subset of planned maintenance. PdM did not gain prominence until the recent decade. This rapid advance is mainly due to emerging internet technologies, connected sensors, systems capable of handling big data sets and realizing the need to use these techniques. The abrupt growth can also be theorized due to the demand for high-quality products, at the least cost and with shortest lead time. Every year, it is estimated that U.S. industry spends \$200 billion on maintenance of plant equipment and facilities and the result of ineffective maintenance leads to a loss of more than \$60 billion [1]. In food and beverage industry it was estimated that failures and downtime accounted for 18% of OEE [2]. Over the years, different architecture, algorithms, and methodologies have been proposed. One of the most prominent methods is watchdog agent, a design enclosed with various machine learning algorithms [3] [11]. Some of the other architectures are an OSA-CBM architecture [4], SIMAP Architecture [5], and predictive maintenance framework [6]. Emerging technologies such as the Internet of things (IoT) devices have formed a gateway to connect to machines and its subcomponents to not only collect the process data and its parameters but also to collect the physical health aspects of the machine such as vibration, pressure, temperature, acoustics, viscosity, flow rate and many as such. This information is widely used for early fault detection, fault identification, health assessment of the machine and predict the future state of the machine. Some of this is made possible due to machine learning algorithms available across different learning domains.

Machine learning is a subsection of Artificial Intelligence Figure 1. Machine learning can be defined a program or an algorithm that is capable of learning with minimum or no additional support. Machine learning helps in solving many problems such as big data, vision, speech recognition, and robotics [7]. Machine learning is classified into three types. In supervised learning, the predictors and response variables are known for building the model, in unsupervised learning, , only response variables are known, and in reinforced learning, the agent learns actions and consequences by interacting with the environment. In this research, the main focus will be on unsupervised learning methodology. One of the most commonly used approaches in unsupervised learning is clustering where, response variables are grouped into clusters either user-defined or model based on the distance, model, density, class, or characteristic of that variable. For this research, vibration data has been used. Data collection, feature selection, and extraction will be described in the later sections.

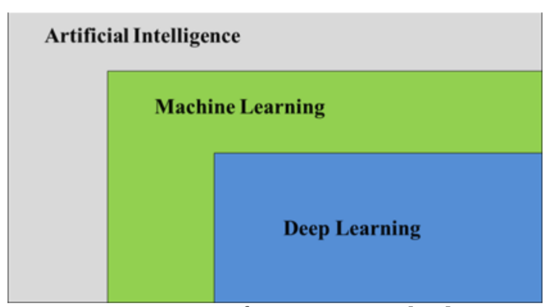

Figure 1: Structure of Learning Methods

All the programming in this research is performed in a statistical tool called as R- Programming. R- Program is open source software and was designed by Ross Ihaka and Robert Gentleman in August 1993. As of today, there are over 10,000 packages which include thousands of different algorithms contributed by various authors for different applications.

II. Literature Review

The primary goal of PdM is to reduce the cost of a product or service and to have a competitive advantage in the market to survive. Today business analytics are embedded across PdM to realize the need for it and to make appropriate decisions. Business analytics can be viewed in three different prospective (i) Descriptive analytics (ii) Predictive analytics and (iii) Prescriptive analytics [16]. Descriptive analytics is a process of answering questions like what happened in the past? This is done by analyzing historical data and summarizing them in charts. In maintenance, this step is performed using control charts. Predictive analytics is an extension to descriptive analytics where historical data is analyzed to predict the future outcomes. In maintenance, it is used predict type of failure and time to complete failure. Finally, prescriptive analytics is a process of optimization to identify the best alternatives to minimize or maximize the objective. This also answers the questions such as what can be done? In maintenance, this can be used to optimize the maintenance schedules to minimize the cost of maintenance. In this paper, our primary focus will be on descriptive and predictive analytics to detect the faults.

Predictive analytics has spread its applications into various applications such as railway track maintenance, vehicle monitoring [23], automotive subcomponents [8], utility systems [19], computer systems, electrical grids [13], aircraft maintenance [21], oil and gas industry, computational finance and many more.

Fault detection is one of the concepts in predictive maintenance which is well accepted in the industry. Early Failure detection could potentially eliminate catastrophic machine failures. In one of the recent research studies, this process is classified into different methods such as quantitative model-based methods, qualitative model-based methods, and process history based methods [25].

Principle component analysis (PCA) is one of the oldest and most prominent algorithms that are widely used today. It was first invented by Karl Pearson in 1901. Since then, they have been many hybrid approaches to PCA for fault detection such as using Kernel PCA [17], adaptive threshold using Exponential weight moving average for T^2 and Q statistic [9], multiscale neighborhood normalization-based multiple dynamic principal component analysis (MNN-MDPCA) method [27], Independent Component Analysis. Another common method used for fault detection is clustering method. Similar to PCA, there are various algorithms such as neural net clustering algorithm neural networks and subtractive clustering [28], K-means [10], Gaussian mixture model [15], C-Means, Hierarchical Clustering [22], and Modified Rank Order clustering (MROC) [33].

III. Fault Detection

Fault detection is one of the most critical components of predictive maintenance. Fault detection can be defined as a process of identifying the abnormal behavior of a subsystem. Any deviation from a standard behavior can be categorized as a failure. In this section, we will discuss different algorithms such as Principle Component Analysis (PCA) T^2 statistic, Hierarchical clustering, K- Means clustering, C-Means, and Model-based clustering for fault detection and benchmark its results for vibration monitoring data.

A. Data Collection

Vibration data is one of the most commonly used technique to detect any abnormalities in a submachine. In this research paper, a vibration monitor sensor was set up on an exhaust fan. The vibration was collected every 240 minutes for 12 days at a sampling frequency of 2048 Hz on both X and Y axis. From the following data, different features were extracted such as peak acceleration, peak velocity, turning speed, RMS Velocity, and Damage accumulation. Figure 2 is the time series plots of the data.

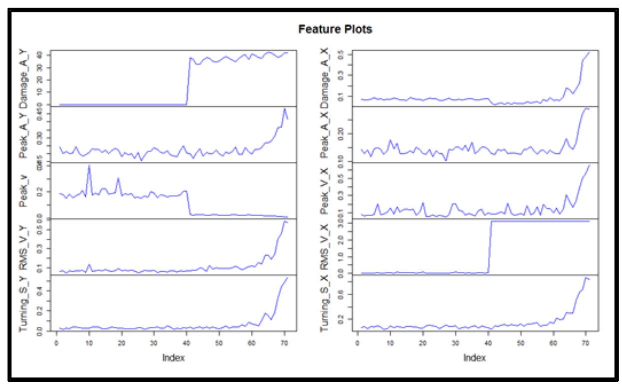

Figure 2: Feature data plot

In Figure 2, we can see a trend line generating closer to index 60[th] observation. In this paper, we will test to see how different algorithms help in detecting this fault earlier.

B. Feature Selection using PCA

Not all features extracted provide a true correlation. If right features are not selected, then a significant amount of noise would be added to the final model and hence, reduce the accuracy of the model. One of the most prominent algorithms for that is used for dimensionality reduction is Principle component analysis. Principal component analysis (PCA) is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation (information) in the data set [18]. In a simple context, it is an algorithm to identify patterns in data and expressing such a way to showcase those similarities and differences [29].

Algorithm:

Step 1: Consider a data matrix X

$$[X]_{mxn} \tag{1}$$

Where, X is the matrix, m is a row, and n is a column

Step 2: Subtract the mean from each dimension

$$[\bar{X}]_n - [X]_n \tag{2}$$

Step 3: Calculate the covariance matrix

$$[c]_{mxn} \tag{3}$$

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix

$$([c]_{nxn} - I_{nxn}\lambda)\{X\}_{nx1} = \{0\} \tag{4}$$

Step 5: Store the eigenvector in a matrix

$$[P]_{nxn} = [\{X_1\}\{X_2\}\{X_3\} \dots . \{X_n\}] \tag{5}$$

Step 6: Store eigenvalues in a diagonal matrix

$$[Eigen]_{nxn} \tag{6}$$

Where [Eigen] is the eigenvalues corresponding to the principal components, and P contains the loading vectors

Step 7: Rank eigenvalues in decreasing order and choose top "r" vectors to retain

$$[Eigen]_{rxr} \tag{7}$$

Step 8: Retain "r" eigenvectors

$$[P]_{nxr} = [\{X_1\}\{X_2\}\{X_3\} \dots . \{X_r\}] \tag{8}$$

Step 9: Calculate the principal components [U] which is projected in data matrix

$$[X]_{mxn}[P]_{mxr} = [U]_{mxr} \tag{9}$$

Summary of the PCA indicates that the first two principal components show 95.65% of variance compared to rest of the components.

A scree plot can be plotted for Eigenvalues versus principle components as shown in Figure 4. This plot can be used to define the components that show significant variance in the data.

From summary data and scree plot, we can conclude that the first two principal components present maximum variation compared to the rest of the principal components.

C. T^2 Statistic

T^2 Statistic is a multivariate statistical analysis. The T^2 statistic for the data observation xcan be calculated by [12]

$$T^2 = \sum_{i=l}^{l} \frac{t_i^2}{\lambda_i} \tag{10}$$

The upper confidence limit for T^2 is obtained using the F-distribution:

$$T^2_{l,n,\propto} = \frac{l(n-1)}{n-l} F_{l,n-l,\propto} \tag{11}$$

```
Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4
Standard deviation     2.7179536 1.4760982 0.36372749 0.33442159
Proportion of Variance 0.7387699 0.2178992 0.01323053 0.01118443
Cumulative Proportion  0.7387699 0.9566690 0.96989956 0.98108399
                           Comp.5      Comp.6      Comp.7       Comp.8
Standard deviation     0.28769049 0.250105624 0.165748777 0.0878586294
Proportion of Variance 0.00827706 0.006255644 0.002747424 0.0007719585
Cumulative Proportion  0.98936105 0.995616694 0.998364118 0.9991360765
                             Comp.9      Comp.10
Standard deviation     0.0772115724 0.0517407897
Proportion of Variance 0.0005961971 0.0002677264
Cumulative Proportion  0.9997322736 1.0000000000
```

Figure 3: Summary of PCA

Where n is the number of samples in the data, a is the number of principal components, and α is the level of significance [24]. This statistic can be used to measure the values against the threshold and any values above the threshold; can be concluded as out of control data. In this case, it is going to be faulty data. The results for the vibration data are shown the Figure 5.

Based on the results from T^2 statistic in Figure 5, we can observe that the faults can be detected as early as 41 observations. Hence, this early detection would help the maintenance teams to monitor these process changes and take corrective actions accordingly.

D. Cluster Analysis

Clustering analysis is one of the unsupervised learning methods. In cluster analysis, similar data are grouped into different clusters. Some of the most prominent cluster analyses are K-Means clustering, C-Means clustering, and hierarchical clustering. There are various merging principles in hierarchical clustering. They are iterative, hierarchical, density based, Metasearch controlled and stochastic. In this paper, we will be discussing one of the commonly used hierarchical clusterings.

E. Optimal number of Clusters

In cluster analysis, we need to know the optimal number of clusters that can be formed. Although we know that, we have healthy data and faulty data, identifying the number of optimal cluster formations in our data would help in understanding different states in the data and representing the data more accurately. To identify the number of clusters, there are many procedures available such as elbow method, Bayesian Inference Criterion method and nbClust package in R. The results for elbow method is shown in Figure 6 and using nbClust [30] is shown in Figure 7.

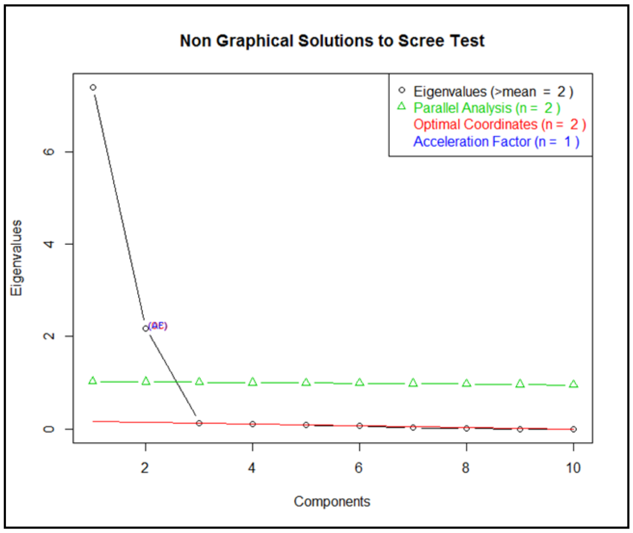

Figure 4: Scree plot to determine the variation between principal components

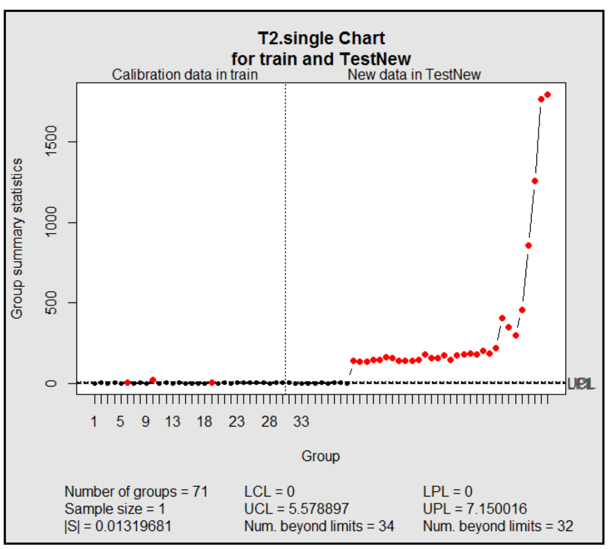

Figure 5: T^2 statistic results for training dataset and testing dataset

From both the procedures shown in Figure 6 and Figure 7, we can identify that 3 clusters are the optimal number of clusters. For fault detection, we can use three clusters and theorize each cluster represents a normal condition, warning condition, and faulty condition. In the next section of cluster analysis, we can observe how each of the clustering algorithms provides the results.

From both the procedures shown in Figure 6 and Figure 7, we can identify that 3 clusters are the optimal number of clusters. For fault detection, we can use three clusters and theorize each cluster represents a normal condition, warning condition, and faulty condition. In the next section of cluster analysis, we can observe how each of the clustering algorithms provides the results.

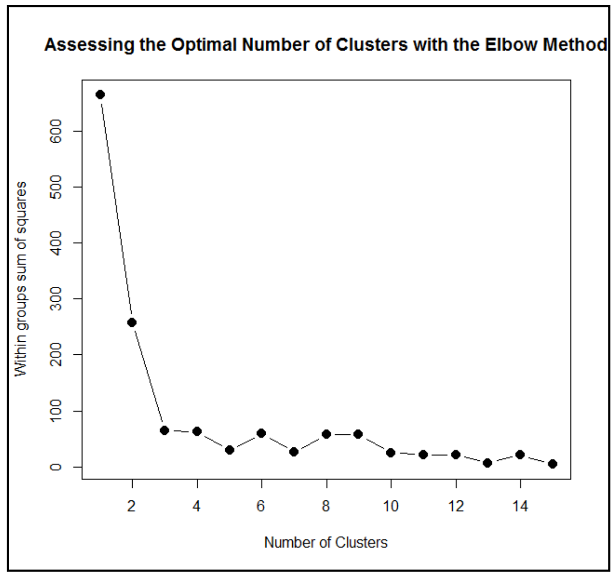

Figure 6: Determining the optimal number of clusters based on Elbow Method

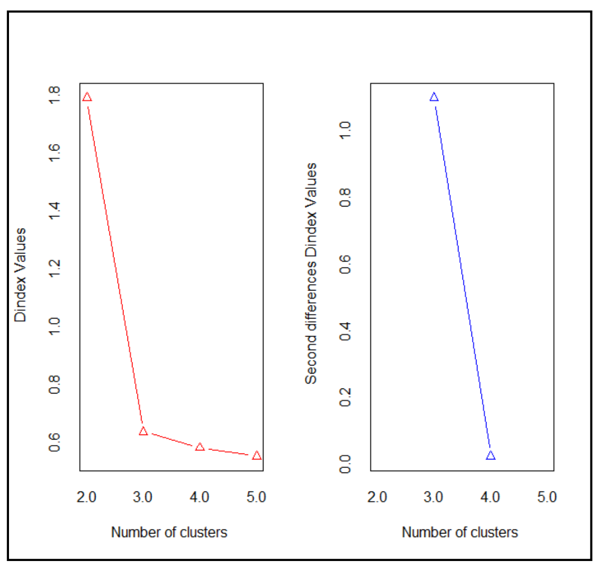

Figure 7: Determining the number of Clusters using nbClust Package

F. Heirarchical Clustering

Start by assigning each item to its own cluster, so that if you have N items, you now have N clusters, each containing just one item. Let the distances (similarities) between the clusters equal the distances (similarities) between the items they contain [24].

Algorithm:

Step 1: Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one less cluster.

Step 2: Compute distances (similarities) between the new cluster and each of the old clusters.

Step 3: Repeat steps 2 and 3 until all items are clustered into a single cluster of size N.

In Figure 8, the cluster is formed based on the feature data using Ward's method. Irrespective of feature data and Principle components, the results were identical. Three clusters were formed, where the first cluster includes observations from 1 to 40, the second cluster includes observations 41 to 67 and finally, the third cluster includes observations from 68 to 71. Based on the domain knowledge, we can represent cluster 1 as healthy dataset, cluster 2 as warning dataset and finally cluster 3 as faulty data set.

G. K-Means and Fuzzy C-means Clustering

K-means is one of the most common unsupervised learning clustering algorithms. This most straightforward algorithm's goal is to divide the dataset into pre-determined clusters based on distance. Here, we have used Euclidian distance. The graphical results as shown in Figure 9

C-means is a data clustering technique where each data point belongs to every cluster at some degree. Fuzzy C means was first introduced by Bezdek [14]. Fuzzy C-Means has been applied in various applications such as agricultural, engineering, astronomy, chemistry, geology, image analysis [14], medical diagnosis, and shape analysis and target recognition [26]. The graphical results for C-Means is as shown in Figure 9

<u>*Summary of K-Means and C-Means Clustering*</u>

Table 1: Cluster Means of K-Means Algorithm

	1	**2**
1	-9.665	-1.609
2	-0.497	1.856
3	1.301	-1.092

Within cluster sum of squares by cluster:
[1] 16.758705 39.575966 8.823486
(between_SS / total_SS = 90.2 %)

Table 2: Fuzzy C-Means Cluster Centers with 3 clusters

	1	**2**
1	1.275	-1.071
2	-0.289	1.920
3	-9.935	-1.723

From Table 3 summary of K-means and C-means clustering, we can observe that clusters of sizes 4, 27 and 40 are formed. Observation 1 to 40 formed one cluster, 41 to 67 formed second cluster and the third cluster with 68 to 71 observations. These results are same as hierarchical clustering.

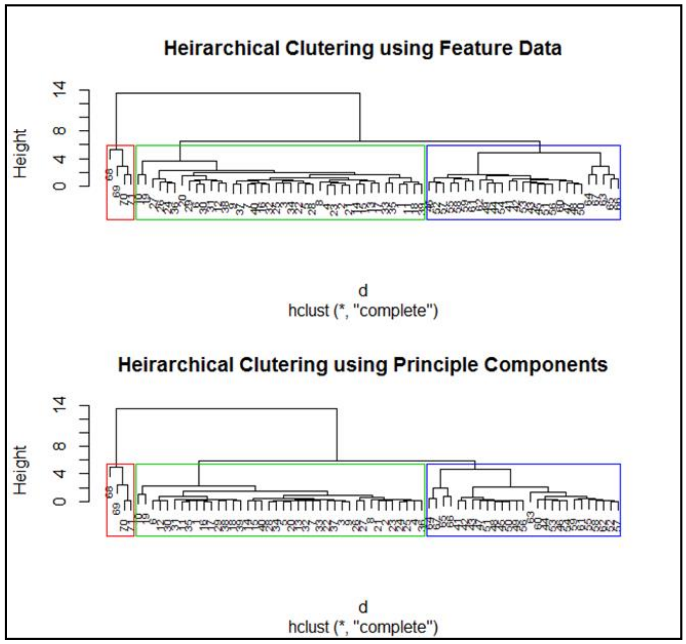

Figure 8: Hierarchical Clustering Solution for Fault Identification

H. Model-Based Clustering

A Gaussian mixture model (GMM) is used for modeling data that comes from one of the several groups: the groups might be different from each other, but data points within the same group can be well-modeled by a Gaussian distribution [20]. Gaussian finite mixture model fitted by EM algorithm is an iterative algorithm where some initial random estimate starts and updates every iterate until convergence is detected [31] [32]. Initialization can be started based on a set of initial parameters and start E-step or set of initial weights and proceed to M-step. This step can be either set randomly or could be chosen based on some method.

<u>*Summary of Classification*</u>

Mclust EVV (ellipsoidal, equal volume) model with five components:

log.likelihood	n	df	BIC	ICL
-57.23501	71	25	-221.037	-222.0734

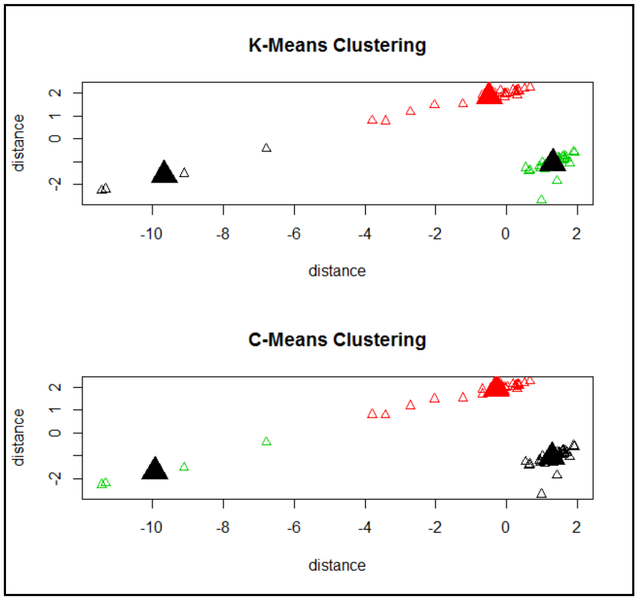

Figure 9: K-Means and C-Means clustering for fault identification

The results are summarized in Table 3. The results from Gaussian finite mixture model fitted by EM algorithm Classification, there was a total of 5 groups of components are formed. Component 1 and two are assigned to observation 1 to 40, component group 3 consists of observation 41 to 63, component group 4 consist of observations 64 to 67 and finally component 5 consists of observations 68 to 71. It is interesting to note that, the critical fault detection which is accurately predicted similarly to other clustering algorithms as well.

IV. RESULTS

In this research, initially, we were hypothesized that two states in data. One is healthy data set, and the other is unhealthy data set. Using PCA and T^2 statistic, we were able to fit our hypothesis states and able to detect the faults 31 observations ahead. Whereas, without a tool and just based on data plots we could observe the trends only 11 observations ahead. As we moved on to fitting different unsupervised clustering algorithms, we found most of the clustering algorithms provided much more than the T^2 statistic.

Using elbow method and nbClust package, we were able to identify that the most optimal number of clusters that could be formed was three. Based on these results, when data was fitted in hierarchical clustering, K-means, and C-means, the results were nearly identical. Based on the previous knowledge of the data, we were able to identify each of three states. The first state was identified as healthy state (since it was calibrated for healthy data), second state was identified as the warning state and finally the third state was identified as faulty state. It would not be surprising to obtain the following results as all these algorithms were based on a distance measure.

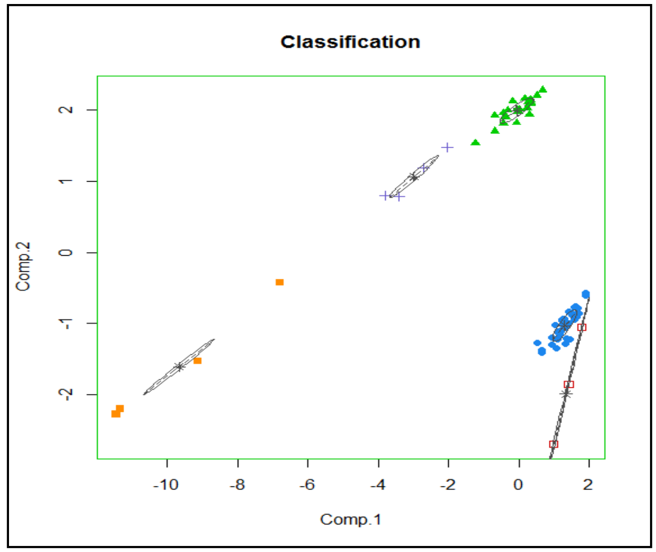

Figure 10: Gaussian finite mixture model fitted by EM algorithm Classification

For our final model, Gaussian finite mixture model fitted by EM algorithm was used. Unlike providing the number of clusters, this model identifies optimal clusters and accordingly classifies the observations into groups. Here, the model recognized a total of 5 components. Although with five components, upon closer investigation, we could observe that, there is an overlap of component 1 and 2 and component 3 and 4. When these components are reorganized we can observe much similar pattern to the previous cluster analysis.

V. CONCLUSION

This research started out as a test bed to benchmark different machine learning algorithms for early fault detection using unsupervised learning. In our results, T^2 statistic provided more accurate results compared to GMM method, and no hypothesis was required to identify the relationship between cluster and state. One of the main benefits of this method is that, even when this is deployed to the manufacturing environment, with minimum or no domain knowledge, one can identify fault or critical condition when compared to clustering analysis. On the other hand in clustering, some information about the data is needed to name the clusters as healthy, warning or critical. Clustering methodology is undoubtedly a better tool in detecting different levels of faults where T^2 statistic would be challenging after certain levels. To emphasize this, when the cost machine maintenance is expensive, clustering would be a flexible option where machine health can be monitored continuously until a critical level is reached.

In conclusion of this study, although most algorithms provided nearly similar results, each algorithm provided deeper insight into the data. Hence, if the application is just to detect the faults, T^2 statistic would be an excellent tool. But if fault detection needs to be performed under different levels then, clustering algorithms would be a better choice.

Table 3: Summary results of all Models

Obs	Actual	T2	Heirarchical	K-Means	C-Means	Model-Based	Obs	Actual	T2	Heirarchical	K-Means	C-Means	Model-Based
1	H	0	1	1	1	1	37	H	0	1	1	1	1
2	H	0	1	1	1	1	38	H	0	1	1	1	1
3	H	0	1	1	1	1	39	H	0	1	1	1	1
4	H	0	1	1	1	1	40	H	0	1	1	1	1
5	H	0	1	1	1	1	41	F	1	2	2	2	3
6	H	1	1	1	1	1	42	F	1	2	2	2	3
7	H	0	1	1	1	1	43	F	1	2	2	2	3
8	H	0	1	1	1	1	44	F	1	2	2	2	3
9	H	0	1	1	1	1	45	F	1	2	2	2	3
10	H	1	1	1	1	1	46	F	1	2	2	2	3
11	H	0	1	1	1	1	47	F	1	2	2	2	3
12	H	0	1	1	1	1	48	F	1	2	2	2	3
13	H	0	1	1	1	1	49	F	1	2	2	2	3
14	H	0	1	1	1	1	50	F	1	2	2	2	3
15	H	0	1	1	1	1	51	F	1	2	2	2	3
16	H	0	1	1	1	1	52	F	1	2	2	2	3
17	H	0	1	1	1	1	53	F	1	2	2	2	3
18	H	0	1	1	1	1	54	F	1	2	2	2	3
19	H	1	1	1	1	1	55	F	1	2	2	2	3
20	H	0	1	1	1	1	56	F	1	2	2	2	3
21	H	0	1	1	1	1	57	F	1	2	2	2	3
22	H	0	1	1	1	1	58	F	1	2	2	2	3
23	H	0	1	1	1	1	59	F	1	2	2	2	3
24	H	0	1	1	1	1	60	F	1	2	2	2	3
25	H	0	1	1	1	1	61	F	1	2	2	2	3
26	H	0	1	1	1	1	62	F	1	2	2	2	3
27	H	0	1	1	1	1	63	F	1	2	2	2	3
28	H	0	1	1	1	1	64	F	1	2	2	2	3
29	H	0	1	1	1	1	65	F	1	2	2	2	4
30	H	0	1	1	1	1	66	F	1	2	2	2	4
31	H	0	1	1	1	1	67	F	1	2	2	2	4
32	H	0	1	1	1	1	68	F	1	2	2	2	4
33	H	0	1	1	1	1	69	F	1	3	3	3	5
34	H	0	1	1	1	1	70	F	1	3	3	3	5
35	H	0	1	1	1	1	71	F	1	3	3	3	5
36	H	0	1	1	1	1	72	F	1	3	3	3	5

VI. Future Scope of Work

Fault detection is one of the preliminary analytics for predictive maintenance. Hence, detecting the fault accurately is regarded important. This work is currently performed for vibration data. The scope of this research can be extended out to other physics-based parameters and combination of these parameters. It would also be interesting to observe the detection accuracy for bigger sample size and multiple fault states.

References

[1] Mobley, R Keith, "An Introduction to predictive maintenance", 2002, 2nd ed, ISBN 0-7506-7531-4

[2] Battini, D., Calzavara, M., Persona, A., and Sgarbossa, F. (2016) "Sustainable Packaging Development for Fresh Food Supply Chains. Package." Technol. Sci., 29: 25–43. doi: 10.1002/pts.2185.

[3] Jay Lee, Hung-An Kao, Shanhu Yang, (2014) "Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment", Procedia CIRP Volume 16, 2014, Pages 3-8

[4] Lebold M, Thurston M. "Open standards for condition-based maintenance and Prognostic systems". In: Proceedings of MARCON 2001—fifth annual maintenance and reliability conference, Gatlinburg, USA, 2001.

[5] Garcia E, Guyennet H, Lapayre J-C, Zerhouni N. "A new industrial cooperative tele-maintenance platform". Comput Ind Eng 2004;46(4): 851–64.

[6] Groba. C, Cech. S, Rosenthal. F., Gossling. A, "Architecture of the predictive maintenance framework", 6th International Conference on Computer Information Systems and Industrial Management Applications, 2007, IEEE

[7] Ethem Alpaydin, "Introduction," in Introduction to Machine Learning,3rd ed. Cambridge

[8] Ahmed, M., Baqqar, M., Gu, F., Ball, A.D., 2012. "Fault detection and diagnosis using principal component analysis of vibration data from a reciprocating compressor", in: Proceedings of the UKACC International Conference on Control, 3-5 September 2012, IEEE Press.

[9] Azzeddine Bakdi, Abdelmalek Kouadri, Abderazak Bensmail, "Fault detection and diagnosis in a cement rotary kiln using PCA with EWMA-based adaptive threshold monitoring scheme", Control Engineering Practice, Volume 66, September 2017, Pages 64-75

[10] C. T. Yiakopoulos, K. C. Gryllias. I. A. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a K-means clustering approach", Expert Systems with Applications Volume 38, Issue 3, March 2011, Pages 2888-2911

[11] Djurdjanovic D, Lee J, Ni J. "Watchdog agent, an infotronics-based prognostics approach for product performance degradation assessment and prediction". Advance Engineering Informatics 2003; 17(3–4): 109–25.

[12] Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components." Journal of Educational Psychology, 24, 417–441.

[13] IBM Israel, "Israel Electric corporation moves towards smarter maintenance", 2013, retrieved from www.IBM.com

[14] J. C. sridek, "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum Press, 1981.

[15] Jacob Goldberger, Sam Roweis, "Hierarchical Clustering of a Mixture Model", Neural Information Processing Systems Conference

[16] James R. Evans, Carl H. Lindner (2012), "Business Analytics: The Next Frontier for Decision Sciences," College of Business, University of Cincinnati, Decision Science Institute

[17] Jingli Yang, Yinsheng Chen, Zhen Sun, "A real-time fault detection and isolation strategy for gas sensor arrays", Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International, 22-25 May 2017, 10.1109/I2MTC.2017.7969906

[18] Jolliffe, "I.T. Principal Component Analysis", Springer, New York, 2002.

[19] P. Liggan and D. Lyons, "Applying Predictive maintenance techniques to Utility Systems, Retrieved from Pharmaceutical Engineering", Official Magazine of ISPE, Nov/Dec 2011, Vol 31 No.6

[20] Ramesh Sridharan, "Gaussian mixture models and the EM algorithm", retrieved from https://people.csail.mit.edu/rameshvs/content/gmm-em.pdf
[21] Samaranayake, P. & Kiridena, S. (2012). "Aircraft maintenance planning and scheduling: An integrated framework". Journal of Quality in Maintenance Engineering, 18 (4), 432-453.
[22] Stephen P. Borgatti, "How to Explain Hierarchical Clustering", 1994, Connections 17(2):78-80 Copyright 1994 INSNA
[23] T. R¨ognvaldsson, S. Byttner, R. Prytz, S. Nowaczyk and M. Svensson, "Wisdom of Crowds for Self-organized Intelligent Monitoring of Vehicle Fleets" 2014, IEEE
[24] Thamara Villegas, María Jesús Fuente, Miguel Rodríguez,"Principal Component Analysis for Fault Detection and Diagnosis. Experience with a pilot plant", Advances in Computational Intelligence, Man-Machine Systems and Cybernetics, ISBN: 978-960-474-257-8
[25] Venkat Venkatasubramanian, Raghunathan Rengaswamy b, KewenYinc, .Surya N. Kavurid, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods", Computers & Chemical Engineering Volume 27, Issue 3, 15 March 2003, Pages 293-311
[26] Y. Yong, Z. Chongxun and L. Pan, "A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding", Measurement Science Review, vol. 4, no.1, 2004.
[27] Yajun Wang, Fuming Sun, Bo Li, "Multiscale Neighborhood Normalization-Based Multiple Dynamic PCA Monitoring Method for Batch Processes With Frequent Operations", IEEE Transactions on Automation Science and Engineering (Volume: PP, Issue: 99)
[28] Zhimin Du, .Bo Fan..Xinqiao Jin, Jinlei Chi, "Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis ", Building and Environment Volume 73, March 2014, Pages 1-11
[29] Lindsay I Smith, "A tutorial on Principal Components Analysis", February 26, 2002, page 2-8.
[30] Malika Charrad, Nadia Ghazzali, Veronique Boiteau, Azam Niknafs (2014). "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software", 61(6), 1-36. URL http://www.jstatsoft.org/v61/i06/
[31] Chris Fraley, Adrian E. Raftery, T. Brendan Murphy, and Luca Scrucca (2012) "mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation Technical Report No. 597", Department of Statistics, University of Washington
[32] Chris Fraley and Adrian E. Raftery (2002) "Model-based Clustering, Discriminant Analysis and Density Estimation" Journal of the American Statistical Association 97:611-631
[33] Nagdev Amruthnath, Tarun Gupta (2016), "Modified Rank Order Clustering Algorithm Approach by Including Manufacturing Data", 4th IFAC International Conference on Intelligent Control and Automation Sciences, Reims, France, June 1-3, 2016