UNIVERSIDAD DE VALLADOLID

ESCUELA DE INGENIERIAS INDUSTRIALES

Grado en Ingeniería de Organización Industrial

Mantenimiento Predictivo: Historia, una guía de implementación y enfoques actuales

Autor:

Santalices Pérez, Saúl

Responsable de Intercambio en la Uva

Ángel Manuel Gento Municio

Universidad de destino

Hochschule Albstadt-Sigmaringen

Valladolid, julio de 2019.

TFG REALIZADO EN PROGRAMA DE INTERCAMBIO

TÍTULO: Predictive Maintenance: History, an implementation guideline and current approaches

ALUMNO: Saúl Santalices Pérez

FECHA: 24 de julio de 2019

CENTRO: Hochschule Albstadt-Sigmaringen

TUTOR: Bernd StauB

RESUMEN:

Debido al aumento del número de sensores utilizados en las plantas de producción, la posibilidad de obtener datos de estas ha incrementado considerablemente. Esto conlleva la posibilidad de detectar fallos antes de que estos ocurran y futuras paradas que afecten a las plantas de producción. Las tecnologías de mantenimiento predictivo permiten predecir eventos futuros, convirtiéndolas en herramientas para afrontar los retos que surjan en los mercados competitivos. Esta tesis está dividida en cinco partes.
La primera, describe el mantenimiento a lo largo de la historia, mientras que la segunda está enfocada en el mantenimiento predictivo. El tercer punto es una guía de implementación de un programa de mantenimiento predictivo para cualquier organización interesada en el tema. Finalmente, las dos últimas partes hacen referencia a los enfoques más comunes en inteligencia artificial donde se explican técnicas importantes como “Artificial Neural Networks” y “Machine Learning”, describiendo algunos ejemplos donde fueron usadas para realizar mantenimiento predictivo.

KEYWORDS:

Mantenimiento Predictivo, Machine Learning, Artificial Intelligence, Artificial Neural Networks, Industria 4.0.

Hochschule
Albstadt-Sigmaringen

Albstadt-Sigmaringen University

Faculty of Informatics

Course of Industrial Engineering

Topic

Predictive Maintenance: History, an implementation guideline and current approaches

Bachelor-Thesis

Submitted in fulfilment of the requirements for the academical degree
Bachelor of Science

submitted
in the spring semester 2019
on 14.07.2019

Inquiries about this publication should be addressed to

Saúl Santalices Pérez
Student number: 89558
PostraBe 22/1
72458 Albstadt
Course of Industrial Engineering
Semester 7
E-Mail: saulsantalicesperez@gmail.com

1. Supervisor/Examiner: Prof. Dr. Bernd Stauß, University of Albstadt-Sigmaringen
2. Examiner: Prof. Dr. German Nemirovski, University of Albstadt-Sigmaringen

Table of Contents

Table of Contents I
List of Figures V
List of Tables VII
List of Abbreviations VIII
Academic Honesty Declaration IX
Abstract X
1 Fundamental Maintenance Theory 1
1.1 Maintenance 1
1.2 Evolution of Maintenance 2
1.2.1 Origins 2
1.2.2 First Generation. 3
1.2.3 Second Generation. 4
1.2.4 Third Generation. 7
1.2.5 Comparison of the three generations. 9
1.3 Fourth Generation 10
1.3.1 From third to fourth generation- Managing the transition 10
1.3.2 Emerging trends and challenges- Shaping the future of Maintenance. 18
2 Predictive Maintenance 20
2.1 Introduction of Predictive Maintenance 20
2.1.1 The scope of Predictive Maintenance 20
2.2 Differences with other Maintenance types 21
2.2.1 Corrective Maintenance. 21
2.2.2 Preventive Maintenance. 21
2.2.3 Condition Based Monitoring. 21
2.3 Important factors in Predictive Maintenance 22

2.4 Predictive Maintenance and the Industry 4.0 22

2.4.1 Industry 4.0. 22

2.4.2 Important trends of the Industry 4.0. 23

2.4.3 Predictive Maintenance in the Industry 4.0. 25

2.5 How does Predictive Maintenance work? 27

2.5.1 Predictive Maintenance architecture. 27

2.5.2 Tools, technologies and techniques required. 28

2.6 Challenges when establishing Predictive Maintenance 33

3 Implementing a Predictive Maintenance Program 36

3.1 Business Analysis 36

3.1.1 Analysing the current situation. 36

3.1.2 Knowing the goals of the project. 42

3.1.3 Determine the goals of the data analysis. 42

3.2 Collecting and Analysing the data 42

3.2.1 Data Requirements. 42

3.2.2 Collecting the data. 46

3.3 Data Processing. 50

3.3.1 Signal Processing. 50

3.3.2 Data Cleaning. 57

3.3.3 Data Fusion. 58

3.4 Feature Engineering. 59

3.4.1 Feature Selection. 60

3.4.2 Feature Extraction. 61

3.4.3 Feature Construction. 63

3.5 Modelling in Predictive Maintenance 63

3.5.1 Binary Classification models. 64

3.5.2 Regression models. 64

3.5.3 Multiclass classification for Predictive Maintenance. 65
3.6 Learning, validation and testing the methods. 66
3.6.1 Cross-Validation. 67
3.6.2 Performance test of the model. 68
3.7 Model Implementation 68
3.8 Model Evaluation 69
3.8.1 Standard Metrics. 69
3.8.2 Prognostics Metrics. 70
4 Common AI approaches in Predictive Maintenance 73
4.1 Rule-Based systems 73
4.1.1 Expert Models. 73
4.1.2 Fuzzy-Logic systems. 74
4.2 Machine Learning Systems 74
4.2.1 Supervised Learning. 74
4.2.2 Unsupervised Learning. 75
4.2.3 Reinforcement Learning. 75
4.3 Machine Learning Approaches 75
4.3.1 Parametric Approaches. 76
4.3.2 Non-Parametric Approaches. 82
5 AI applications in Predictive Maintenance 97
5.1 Artificial Neural Network applications 97
5.1.1 Mechanical damage and crack detection. 97
5.1.2 Early detection of faulty electrical devices. 100
5.1.3 Fault detection on pneumatic systems. 101
5.1.4 Robotic manipulator monitoring. 103
5.2 Application´s conclusion 105
6 Conclusion 107

REFERENCES 108

List of Figures

Figure 1. Input- output model of enterprise with respect to Maintenance 1
Figure 2. Views on failure of equipment during First generation of maintenance 3
Figure 3. Typical bath-tub curve 5
Figure 4. Nowlan and Heap Failure patterns as explained in (Moubray,1997) 6
Figure 5. Expectations of maintenance during the three generations 9
Figure 6. View of equipment failure during the three generations 10
Figure 7. Changing maintenance techniques during the three generations 10
Figure 8. Some important reason for safety problems in maintenance. 12
Figure 9. Maintenance in product life cycle 13
Figure 10. Maintenance methods and techniques and their environmental impact 13
Figure 11. `Abandonment´ phase in a product´s life cycle 14
Figure 12. Maintenance strategy continuum related to OEE 15
Figure 13. Cyber-Physical systems 23
Figure 14. From interconnectivity to value creation: the four cornerstones of digitalization . 27
Figure 15. PdM Architecture 28
Figure 16. Block Diagram of a Sensor Node (Wired or Wireless) for Vibration Analysis 29
Figure 17. PdM model in Big Data framework 31
Figure 18. The three dimensions (Occurrence), (Severity) and (Detection) of System FMEA 41
Figure 19. Example of FMEA analysis general table 41
Figure 20. Using Simulink to model a transmission. This model can be used to synthesize fault data.. 44
Figure 21. A Conceptual Framework of Data Quality 45
Figure 22. Push- and pull-based data acquisition techniques 48
Figure 23. An example of a waveform (blue line) for a given random variable 51
Figure 24. An example of a signal in time domain (left) and frequency domain after a Fast Fourier Transformation 52
Figure 25. Example of Short Time Fourier Transformation 52
Figure 26. Shifting Process 54
Figure 27. Starting signal and the residual obtained after calculating the first empirical mode 55
Figure 28. Example of Modal Empirical Decomposition 55

Figure 29. Labelling for binary classification 64
Figure 30. Labelling for regression 65
Figure 31. Labelling of the multiclass classification to predict the time of error 65
Figure 32. Labelling of the multiclass classification to predict the cause of the error 66
Figure 33. Hierarchical design of the prognostic's metrics 71
Figure 34. Example of prognostic horizon of two different PdM systems (red and blue lines) 71
Figure 35. An illustration of a discriminant function in two dimensions (left) and a plot on the right which samples two classes (red and blue) by using the Fisher linear discriminant 78
Figure 36. An example of a Dynamic Bayesian Network unfolded over the time 81
Figure 37. Example of a semi-Markov model structure 82
Figure 38. An example of a Support Vector Machine approach for two features and two classes 83
Figure 39. Example of a Decision Tree 84
Figure 40. Comparison between training error rate and validation error rate depending on the k value 87
Figure 41. Structure of a NN with two hidden layers 88
Figure 42. Backpropagation example for a cost function 90
Figure 43. Probabilistic NN architecture 92
Figure 44. Adaptative Neuro-Fuzzy inference system for the given rules 94
Figure 45. Classification results 98
Figure 46. Multilayer Perceptron performance 99
Figure 47. Grey image of thermogram 100
Figure 48. Training and testing results with different fault conditions 103
Figure 49. Schematic representations of the SOMNN analyser and the RBNN analyser 104
Figure 50. Table with the structural and training parameters of the feedforward NNs with two learning algorithms 105

List of Tables

Table 1. Characteristics of the First Generation of Maintenance 3
Table 2. Characteristics of the Second Generation of Maintenance 5
Table 3. Characteristics of the Third Generation of Maintenance 7
Table 4. Applications of Machine Learning Approaches in PdM´s recent literature 96
Table 5. Overview of used ANN designs with achieved results...105

List of Abbreviations

PdM	Predictive *Maintenance*
RCM	*Reliability-Centered Maintenance*
CBM	Condition Based Monitoring
IoT	Internet of Things
ML	*Machine Learning*
ANN	*Artificial Neural Network*
AI	*Artificial Intelligence*
NN	*Neural Networks*
ERP	*Enterprise Resource Management*
MES	*Manufacturing Execution System*
KBS	*Knowledge Based System*
RUL	*Remaining Useful Life*
PCA	*Principal Component Analysis*
RBNN	*Radial Basis Neural Network*
SOMNN	*Self-Organizing Maps Neural Network*

Academic Honesty Declaration

I declare that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for an assignment at this or any other University.

Formulations and ideas taken from other sources are cited as such. This work has not been published.

Albstadt, July 11th, 2019

Saúl Santalices Pérez

....

Abstract

This report regards a Bachelor´s thesis conducted at the University of Albstadt-Sigmaringen in the spring semester of 2019. The purpose of this thesis was to investigate and to write a report about Predictive Maintenance related to the Industry 4.0, as well as, to carry out an implementation guide for any organisation interested in such program and to present the current methodologies used nowadays.

Because of the increasing number of embedded sensing computer systems used in production plants and machines, the possibility to monitor the data from such systems have increased considerably. Such developments have given risen to detect failures before they occur and predict future breakdowns which affect production plants. Predictive Maintenance technologies allow the organizations to predict the future, which become them in powerful competitive tools to face future problems and challenges that arise in such competitive markets. This thesis is divided in 5 chapters related to the main topic.

The first chapter describes the evolution of maintenance along the history, since its first appearances until nowadays, comparing the different generations from the same terms. Second chapter is focused on Predictive Maintenance, its architecture and its importance in the Industry 4.0, where it is described how are they related to each other. The third chapter, the longest of them, tries to be a guide about how to implement a Predictive Maintenance program in an organization, where all the steps, which must be carried out to achieve successful results, are described in detail. Fourth chapter describes the most common Artificial Intelligence approaches where, some Machine Learning techniques like Artificial Neural Networks acquire much importance, as they are usually used nowadays to solve Predictive Maintenance problems. Finally, in the last chapter, some examples about Artificial Neural Networks set up in Predictive Maintenance and failure type detection problems are described.

Key words: Predictive Maintenance, Machine Learning, Artificial Neural Networks, Artificial Intelligence, Industry 4.0.

1 Fundamental Maintenance Theory

This chapter presents the general theory behind the history of the maintenance, the different types of maintenance and the importance of this issue in the day by day in any company, industry and sector.

The scope of this project involves research of different maintenance approaches and concepts but otherwise, it is focused mainly in the Predictive Maintenance (PdM), on how to implement a PdM program as well as the different AI approaches to carry out it and the most modern techniques used at the present.

1.1 Maintenance

There are a lot of definitions of maintenance, all of them are true but ones are explained better than others, but if I had to choose one, without doubt I would take this one:

Maintenance is defined as ``The combination of all technical and associated administrative actions intended to retain an item in, or restore it to, a state in which it can perform its required function´´ British standard 3811 as cited in (Luxhøj, Riis, & Thorsteinsson, 1997)

Maintenance acts as a support for the production process, where the production input is converted into specified production output. Industrial maintenance comes as a secondary process, which must contribute for obtaining the objectives of production. Maintenance must be able to retain or restore the systems for carrying out a perfect production function (Gits, 2010, pp.5-17).

Maintenance is an important part of the entire organisation; thus, its goals must be related an in accordance with organisation objectives and purposes.

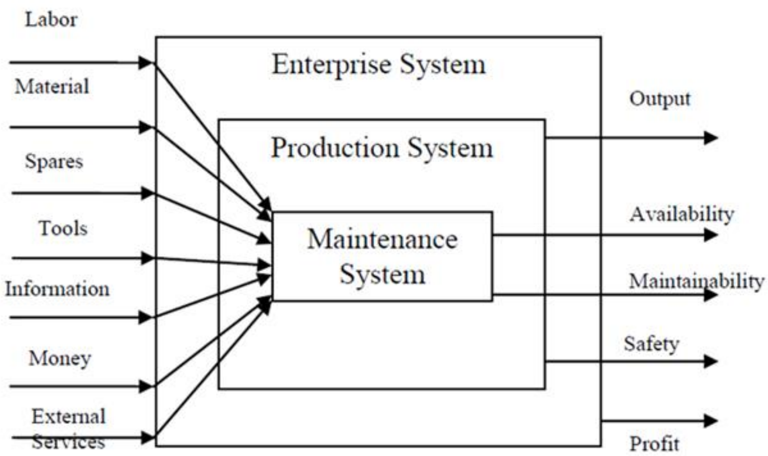


Figure 1. Input- output model of enterprise with respect to Maintenance (Al-Turki, 2011)

As described from the systematic view of maintenance, there are four strategic dimensions of maintenance (Tsang, 2002, pp. 7-39). The dimensions start with the called ``Service- delivery´´ options which are related with some inputs as labour, material, spares and external services. This explains the choice within the inside capability and outsourcing. The second and third dimensions are related to the design and selection of maintenance methodologies. The performance will play a major role on the output, some of them are productivity output, safety, maintainability and the profit of the whole enterprise. The fourth dimension is related to the support systems which is explained as the design that is supporting maintenance (Tsang, 2002, pp. 7-39).

In all cases, the performance of the maintenance function can be judged by the condition of machinery, which the following factors indicate:

- *Performance*, this is the ability of the machine to perform its functions.
- *Downtime* refers to a period that a system fails to provide or perform its primary function. Operation of the machine must be within an acceptable level of downtime.
- *Service life*, before replacement of the machine is necessary; it must provide a good return on investment.
- *Efficiency,* the level of efficiency of the machine must be acceptable.
- *Safety,* the machine must be safe to the personnel.
- *Environmental impact*, the operation of the machine must be friendly to the environment and other equipment, trying to reduce the number of emissions harmful to the atmosphere.
- *Cost*, it is expected to have a maintenance cost within an acceptable level.

The goal of maintenance is to ensure that machinery performance is satisfactory, considering the above factors. This chapter covers the brief history of traditional machine maintenance and maintenance strategies.

1.2 Evolution of Maintenance

Throughout the course of the last century, the focus on the maintenance of industrial production and process systems has seen major development. (Moubray, 1997, p. 445) suggested that there had been three distinct ``generations´´ of maintenance, but, as he wrote this more than twenty years ago, it is considered that there is a fourth generation.

Each generation can be dated in different periods of the last century;

- the period between (1930-1950) which is referred to as the *First Generation*,
- between 1950 to 1970 as the *Second Generation*,
- up to 1970 to last years of the century as the *Third Generation*,
- from 2000 to nowadays as the *Fourth Generation*.

The growth in maintenance efficiency has become more complex due to equipment automation, the development of the industry which nowadays is considered as the industry 4.0 where Big Data, Internet of Things, ML… have become important

1.2.1 Origins.

Maintenance has been related for many years along with other fields. Really, the concept of maintenance was heartily considered in a high degree after the industrial revolution. Through the literature research on maintenance not many papers were written on the evolution and developments of maintenance.

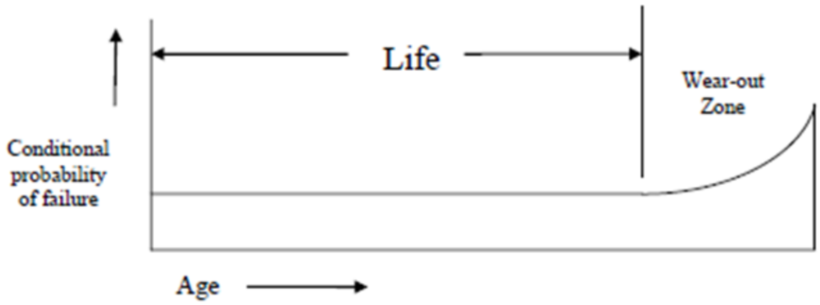
According to (Sherwin,2000, pp. 138-164), before the industrial revolution, maintenance generally is considered that begun around 1750 and firstly used by carpenters, smiths', coopers' wheelwrights, masons, etc. There was no control, and repairing, means the replacement of the spare part with other new parts. The crafts-man would replace the repaired part by a new strong part which would give a long life. As the methods for stress calculation were not present, the design and repairs were closely integrated. A strong part would be fit for the replaceable part, thus, changing the design for the next similar machine.

1.2.2 First Generation.

The first generation describes the earlier days of industrialization where mechanization was low. Most equipment in the factory was basic and the repairing and restoration process was done in a very short time. Thus, the term downtime did not matter much and there was no need for managers to put maintenance as a high priority issue.

(Moubray,1997) suggested that the first generation of maintenance, before and up to the Second World War, could be described in the following terms:

Table 1. Characteristics of the First Generation of Maintenance

FIRST GENERATION	
Expectations of Maintenance	• ***Fix equipment when it breaks*:** fixing the damage or broken parts after it was broken
Views on Equipment Failure	• ***Simple machines*:** slow and simple machines to work as well in its design. The mechanization of machines was simple in design and operations. • ***Necessary evil*:** Maintenance was one of the production tasks. The repairing or replacement of parts stopped the whole work so maintenance was related to cost consuming but necessary as the repair or replacement must be done. • ***Every machine wears out:*** the views about the failure of equipment was described as each machine or item wears out (Dunn, 2003):  *Figure 2. Views on failure of equipment during First generation of maintenance (Moubray,1997)*

Maintenance Techniques	• Fundamental Repair Skills: – The skills of maintenance were basically the repair or replacement of the part when it broke so before this happen nothing was done. – The cause of failure was normally not considered, and maintenance activities focused on restoring and repairing the asset to fully operational conditions. This type of reactive approach is commonly referred to as *"run-to-failure" or Corrective Maintenance.* – People could easily understand why equipment needed fixing when it was broken, and it was difficult for them to comprehend any other form of maintenance

After the end of the Second World War an increase in product demand and a booming market in the west pushed the development of the industrial production systems. The equipment and processes started to become bigger, faster and in many cases, better. Moreover, the industrial equipment started to focus on production costs and the role of maintenance became increasingly important. The "fix when it breaks" mentality, which characterize the corrective maintenance approach, largely shifted towards a preventive maintenance mindset with scheduled maintenance activities aiming to prevent equipment failure from occurring (Alsyouf, 2006, pp. 70-78).

The impact of this was immediately felt by the maintenance engineers and artisans, and they started introducing the concept of planned shutdown periods when these consumable parts were replaced. These planned shutdowns did not necessarily eliminate breakdowns, but engineers started to find that by replacing the key components the number of breakdowns was reduced. This was the introduction of the second generation of maintenance, and equipment started to be maintained on fixed frequencie.

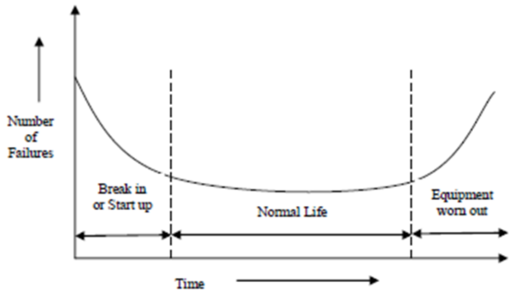
1.2.3 Second Generation.

The second generation emerged as the results of growing complexity in equipment and plant design. This led to an increase in mechanization, and industry began to depend on these complex machines. On these days, the reparation and restoration had become more difficult. Special skills emerged and the time to mend the machinery became larger.

Because of the increasing of this dependence, downtime became an important problem and then received more attention from management. Since then, it started to think that these failures could be avoided by the concept of preventive maintenance. As maintenance cost started to rise sharply relative to other operating costs, there was a rising interest in the field of maintenance planning and control systems.

(Moubray,1997) suggested that the second generation of maintenance, after the Second World War, and up to the 1980´s could be described in the following terms:

Table 2. Characteristics of the Second Generation of Maintenance

SECOND GENERATION	
Expectations of Maintenance	• ***Higher equipment availability*:** the availability of equipment was important and had high level of expectation for maintenance, because the downtime became one of the most important criteria for the production. • ***Longer equipment life:*** longer life span of equipment was mostly expected from maintenance (Sherwin, 2000, pp. 138–164) • **Lower maintenance Costs:** maintenance cost increased relative to other costs which made to be planned and controlled. These planned outages started becoming an unwanted focus of attention as their costs spiralled outside of the budgets.
Views on Equipment Failure	• All equipment complies with the ***``bath-tub´´*** curve:  *Figure 3. Typical bath-tub curve (Mobley 2002)* The three sections forming the bath-tube curve is explained by (Mobley, 2002) as the new machine has a high number of failures at the first section or starting up because of the problems in installations. After some operations are carried out, the probability of failure decreases, and the machine can work in a steady normal life. After a certain amount of time the machine can start to work worse and this is the final stage, when the probability of a failure occurs gets higher.
Maintenance Techniques	• ***Scheduled Overhauls:*** personal used typically scheduled overhauls to prevent the failure of the equipment. (Moubray, 1997) explains two terms related to this issue. – *Scheduled Restoration Tasks*: restore a component before a specific age limit, regardless of its condition currently. – *Scheduled Discard Tasks:* discarding a component at or before a specific age limit, regardless of its condition at the time. • Systems for planning and controlling work (PERT, GANTT) • ***Big, slow computers:*** The second generation in maintenance concur with the first second and third generations of computers. Although a lot of improvements were carried out through these generations, computers of this time were bigger and slower than nowadays and computers from third generation of maintenance.

1.2.3.1 Nowlan and Heap experiment, a change of mindset.

In the 1978, after years of research the Us department of defense published “Reliability Centered Maintenance” by Nowlan and Heap. In their seminar work, Nowlan and Heap defined failure and potential failure as…

“A failure is an unsatisfactory condition. In other words, a failure is an identifiable deviation from the original condition which is unsatisfactory to a particular user.”

“A potential failure is an identifiable physical condition which indicates a functional failure is imminent and is usually identified by a Maintenance Technician using predictive or quantitative preventive maintenance”

With this publication they explained the term Predictive or Condition Based Maintenance which is based on the concept that there is enough time between the potential failure is detected and the functional failure occurs for the organization to react and prevent the functional failure. This interval of time is known as the p-f interval. Thanks to this research organizations could decide what constitutes an unacceptable condition. The decision impacts whether an organization will be able to eliminate all functional failures except for those that have decided to accept by making a run-to-failure, or no scheduled maintenance decision.

Over a period of years of a lot of studies conducted on aircraft components the authors revealed the six-basic age-reliability relationships shown on next figure. The Y axis represents the conditional probability of failure and the X axis represents time in service after installation or overhaul.

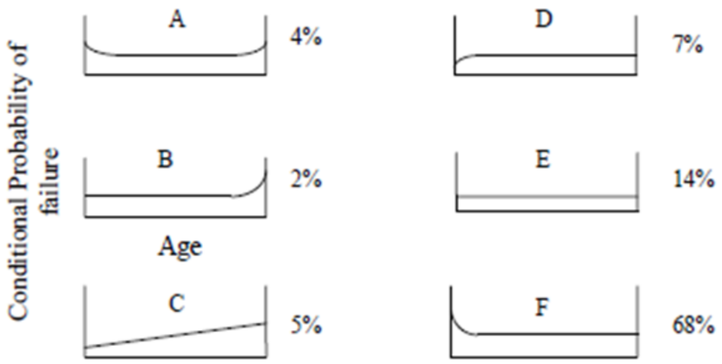


Figure 4. Nowlan and Heap Failure patterns as explained in (Moubray,1997)

Only patterns A and B which represent only a six percent of the items studied display the wear-out region by a rapidly increasing conditional probability of failure. On the study it cans observe that ninety five percent of the items presented a least some regions of random failures. This means that 95% of the equipment in the study may benefit from some form of Condition Monitoring (random failures) and that only 6% may benefit from time-based replacement or overhaul (Preventive Maintenance).

The main issue is related to the 68% percent obtained of the pattern F, or Infant Mortality. This high percentage means that a lot of items suffered a failure after installation or Preventive Maintenance. Most item failures were being induced by activities directly related to time-based replacements and overhauls (Preventive Maintenance).

These findings exposed a need for a new way to approach equipment failures. As the first industry to confront the problems of multiple failure patterns and increasing complexity in maintenance decision-making, the aviation industry in the United States started to develop a framework. It came to become a comprehensive decision-making process, later known as RCM (Reliability-Centered Maintenance).

According to (Hinchcliffe & Smith 2003), RCM consists of four main features; defining system function, identifying potential loss of function, prioritize functions and identifying candidate actions. The general view of RCM is that, if implemented correctly, several positive effects can be achieved. For example, safety and environmental hazards can be reduced, operating performance increased, and maintenance can become more cost effective.

1.2.4 Third Generation.

Reliability had become vital in the maintenance circle from the middle 70 ‘s: failure of machinery would be harmful to reach high levels of productivity and profitability. At this time a machine breakdown could have a fatal effect on a plant and its operations. The complexity of machinery and the automation system had been living an incredible increase during this period. Since the mid-seventies, because of the changes, the industry had reached a great momentum. These changes were related to new expectations, new research and new techniques.

Mindset started to change, and maintenance started to consider as a ``profit contributor´´. The better maintenance models helped to produce better quality, safety and maintainability and hence making it one of the main areas for increasing profit. In the manufacturing industry, the effects of downtime were aggravated by the just-in-time systems, where a small breakdown, likely stopped the whole plant. In this generation the aims of maintenance where focused on the following;

Table 3. Characteristics of the Third Generation of Maintenance

THIRD GENERATION	
Expectations of Maintenance	• ***Higher equipment availability:*** industries became much more specialized and they started to get closer to the nowadays industries and availability became more important. Little breakdowns could cause big loses on production and budgets. • ***Higher equipment reliability:*** Related to the availability of the components is the reliability of them. Both issues go together, and one implies the other. • ***Greater safety and no environmental damage:*** (Moubray,1997) emphasize that more and more failures had serious safety and environmental consequences, at a time when standards in these areas were rising rapidly. • ***Better Product Quality:*** organizations had to compete against others and began to difference from the others • Longer equipment Life • ***Greater Cost Effectiveness:*** cost of maintenance itself was still rising, in absolute terms and as a proportion of total expenditure
Views on Equipment Failure	• ***6 failures patterns, following the research of Nowlan and Heap:*** the Nowlan and Heap experiment set a before and after in the maintenance operations. The experiment showed that there is less and less connection between the operating age of most assets and how likely they are to fail.
	• ***Condition Monitoring:*** strategy that monitors the actual condition of an asset and, depending on the results, you will have to take the better decisions on terms of maintenance to allow the asset works as its normal life. Collecting and monitoring the condition of the

<table>
<tr>
<td>Maintenance Techniques</td>
<td>
equipment allowed industries to calculate whether when a fault or a breakdown might occur. Some important monitoring techniques are:
– Vibration Monitoring-monitoring and measuring the vibration signals to detect the malfunction of the equipment.
– Oil based Monitoring-Analysing a sample of an engine, transmission or hydraulic oil.
– Infrared thermography-detecting the invisible thermal energy and converting it to a visible image on a screen which can be analysed to identify any abnormality.
– The acoustic emission (AE): monitoring the condition of rotating machinery with the advantage of a significant improvement in signal to noise ratio.
• Design for Reliability: probability that a system will perform a specified function within prescribed limits, under given environmental conditions for a specific time (Stapelberg, 2009).
The method integrates functional failure as well as functional performance criteria so that a maximum safety margin is achieved with respect to acceptable limits of performance. The aim is to produce a design that has the highest possible safety margin with respect to all constraints.

• Design for maintainability: the method makes use of maintainability prediction techniques as well as specific quantitative maintainability analysis models relating to the operational requirements of the design.
The predictions can aid in design decisions where several design options need to be considered and the analysis considers the assessment and evaluation of maintainability from the point of view of maintenance and logistics support concepts.

• Hazard studies: involves the identification of hazards at a facility and evaluating possible scenarios leading to unwanted consequences. This stage is a very important part of the risk management process, as no action can be made to avoid, or reduce, the effects of unidentified hazards.
Many methods and techniques can be used to perform this task at various stages through the life cycle of the process. These techniques try to identify fault conditions that lead to hazards and reduce the chance of missing hazardous events. (Glossop, 2000)

The hazard identification techniques have been divided into four categories depending on the area in which they are predominantly applied:

– Process hazards identification
– Hardware hazards identification
– Control hazards identification
– Human hazards identification
The stage in the process life in which the safer process has been carried out is important as the cost of alterations in the plant vary.

• Failures Modes and Effect Analysis (FMEA, FMECA):
FMEA: is a step-by-step approach for identifying all possible failures in a design, a manufacturing or assembly process, or a product service.

– ``Failure modes´´ means the ways or modes in which something might fail. Failures are any errors or defects, especially ones that affect directly the customer, and can be potential or actual.
– ``Effect analysis´´ is the study of the consequences of those failures.
To evaluate the FMEA, there is a method called…

– Risk Priority Number (RPN): A method that provides an alternate evaluation approach to Criticality Analysis. This method provides a qualitative numerical estimate of design risk. RPN is composed by three independent factors: Severity (S), Occurrence (O) and Detection (D).
$$RPN= (S) * (O) *(D)$$
</td>
</tr>
</table>

	FMECA is the result of two steps: – Failure Mode and Effect Analysis (FMEA) – *Criticality Analysis (CA):* quantitative and qualitative. Calculate the criticality for *each potential failure* mode by obtaining the product of the three factors: *Mode Criticality = Item Unreliability x Mode Ratio of Unreliability x Probability of Loss* Calculate the criticality for *each item* by obtaining the sum of the criticalities for each failure mode that has been identified for the item. (Donald W. Benbow, et al, 2002) *Item Criticality = SUM of Mode Criticalities* To use the *qualitative* criticality analysis method to evaluate risk and prioritize corrective actions, the analysis team must: Rate the severity of the potential effects of failure; rate the likelihood of occurrence for each potential failure mode. Compare failure modes via a Criticality Matrix, which identifies occurrence on the horizontal axis and severity on the vertical axis (Hot Wire, 2004) • ***Small, fast computers***: During this generation, the fourth generation of computers was developed. On this computer generation the microprocessor appeared, and this led to the apparition of the first Personal Computers

1.2.5 Comparison of the three generations.

Here a quickly review about how the expectations of maintenance, views of equipment failures and Maintenance techniques is done. For representing this, some figures from (Moubray, 1997) have been used.

Moubray describes de development of the expectations of maintenance alongside the three generations as follows:

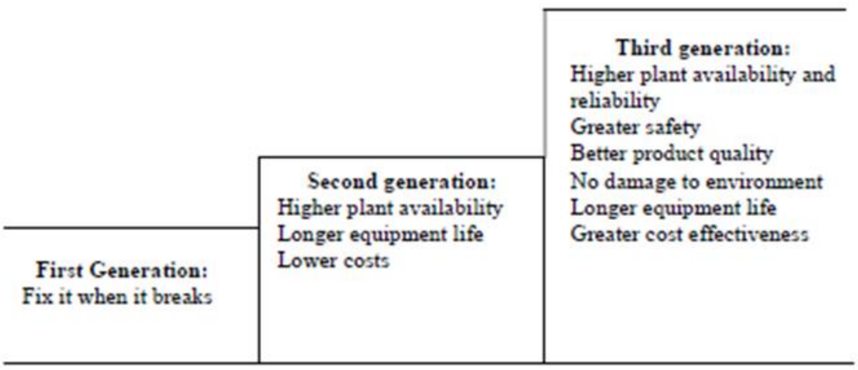


Figure 5. Expectations of maintenance during the three generations (Moubray, 1997)

Now how view of equipment failure changed during the three generation is showed in the figure:

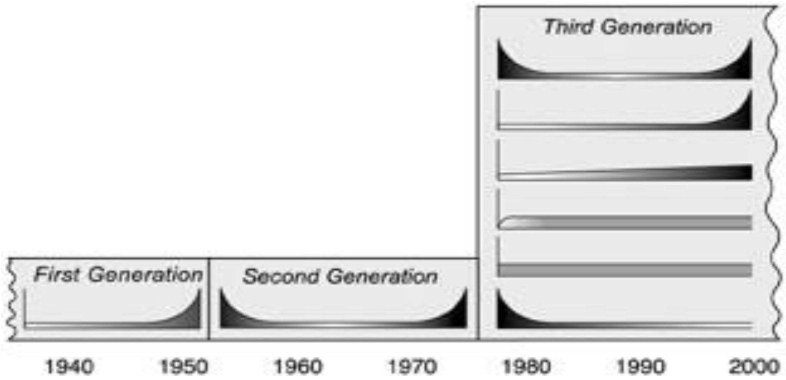


Figure 6. *View of equipment failure during the three generations (Moubray, 1997)*

Finally, in the next figure a comparison between the techniques developed and carried out in all the three generations is presented

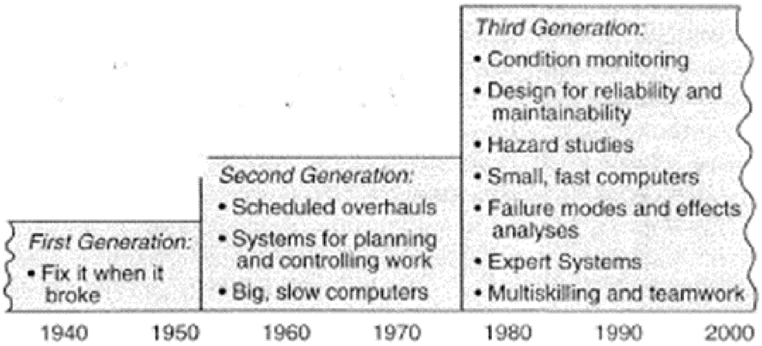


Figure 7. Changing maintenance techniques during the three generations (Moubray, 1997)

1.3 **Fourth Generation**

1.3.1 From third to fourth generation- Managing the transition

During the history, the humanity has lived an evolutionary development, new ideas, technics, methods, knowledges, skills… have appeared which have led to a continuous state of changing methods. Talking about maintenance. It can be observed something similar since maintenance and the development of maintenance practices has been an evolutionary process thus far, and if the trends still appearing and being developed, will continue to be so for the foreseeable future.

Moubray wrote his report between 1991 and 1997 during the third generation of maintenance, but since then the things have changed and new developments have appeared. A review of each of the influencing factors that may shape the fourth generation of maintenance are made. As always, this part is divided in different parts as expectation of maintenance, views on equipment failure and maintenance techniques. A discussion about how these issues have changed in the last years will be performed to try to justify the born of the fourth-gene (Dunn, 2003) list all the Moubray terms from the third generation of maintenance and try to show how the changes have occurred during the last years. He examines the current expectations in detail as showing now.

1.3.1.1 Expectations of Maintenance.

- **Equipment Availability.**

During the last twenty years the great importance that equipment availability achieved in the third-generation of maintenance has not decreased, in contrary, organization have still putting more and more effort to improve this issue and try to get levels of 100% in equipment availability.

Related to equipment availability, in the last years a new term has surged which called OEE (Overall Equipment Effectiveness). (Clarke, 2018) states that OEE is a method of analyzing the performance of a machine or piece of equipment compared to its theoretical maximum capacity. He says that OEE includes three main factors which are availability, performance and quality.

Availability takes an important role and impacts directly on OEE. Planned stops in production for setups and adjustments such as planned maintenance, cleaning, and quality inspections must be made to avoid unplanned failures. Unplanned stops in production, typically because of breakdowns, also negatively impacts OEE as well as short stops.

- **Equipment Reliability.**

Equipment availability and reliability are really related to maintenance. (Dunn, 2003) pointed that it has a great perception that there is now a far greater awareness amongst operations and maintenance personnel of the impact that low level of equipment reliability can have on continuous manufacturing processes, even when availability is relatively high.

(Mobley, 2019) explains that during the last years poor maintenance practices have been perceived as the dominant factor that limits production capacity, product quality and profitability. Many of the perceived maintenance problems are outside of the maintenance function. Improper operating procedures, poor design or improper scheduling of production is the real sources of many plant reliability problems.

Mobley points some plants functions that assume an active role in equipment reliability, so organizations should put more effort to performance very well on these aspects to achieve a good equipment reliability:

- Asset dependability begins with the specification and selection process
- Purchasing must also use good judgement when selecting replacement components for both maintenance and production.
- Production has the greatest role to play.
- Sales and marketing direct affect equipment reliability.
- Employee skills are also a critical part of equipment reliability.

- **Greater safety.**

(Dunn, 2003) suggests that typically the focused was on high frequency, low consequence events, but in the last decade, has been expanded to include low frequency high consequence events which resulting in industrial catastrophes and disasters. According to (Dhillon, 2002), in addition to the general safety considerations, other factor that influence the safety dimensions of maintenance tasks include:

- Numerous maintenance tasks or jobs are in direct response to the needs of working safely. Consequently, safety needs augment maintenance tasks or jobs.
- Numerous maintenance tasks or jobs are hazardous and lead to hazardous solutions. Thus, maintenance work is a cause of safety-related problems.

On the one hand, maintenance techniques lead to increase safety in organization but, on the other hand, these maintenance techniques or methods can put staff at risk, decreasing the safety in the jobs.

The next figure shows the most important reason associated for safety problems in maintenance:

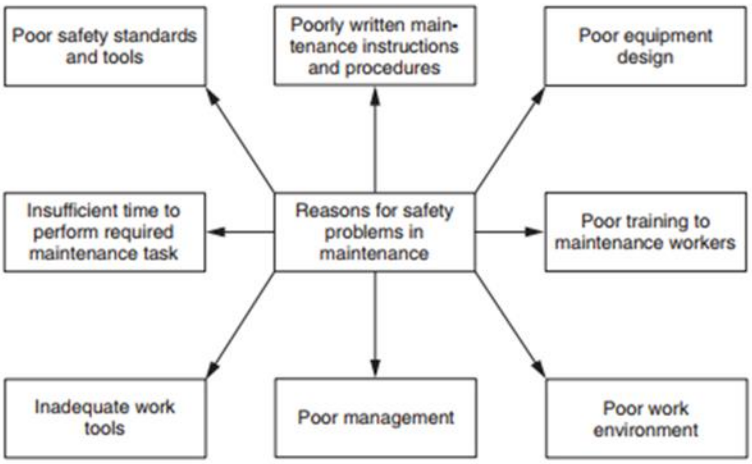


Figure 8. Some important reason for safety problems in maintenance. (Dhillon, 2002)

- **No environmental damage.**

During the last years there has been an increasing focus in the relevant industries, on minimising the environmental impact of operations. There is no doubt that equipment reliability has a large part to play in ensuring compliance with environmental standards and regulations, and there is an ever-increasing expectation that maintenance will ensure that equipment permits compliance with these standards and regulations (Dunn, 2003).

New concepts of business management (Lean Manufacturing, Green Manufacturing or Sustainable Manufacturing) resulted in modified perception of maintenance. The two areas, maintenance and environment, are inter-dependable, both in terms of results of actions and the effects. (Jasiulewicz Kaczmarek & Drożyner, 2013). Regular professional maintenance ensures the most eco-efficient use of equipment and the longest, cleanest life cycle, with the smallest environmental impact. Thus, inclusion of the maintenance function in the pro-environmental strategy is equally indispensable at the stage of investing in new machines and equipment and later during operation and abandonment phases (Napiórkowski & Szczyglak 2011, pp. 145–152; Napiórkowski et al. 2011).

With the new paradigm of sustainable development which has affected the product life cycle management, maintenance can be included in the chain of values of the entire organisation. To highlight and justify the new way of perceiving maintenance, Takata introduced the expression of \`maintenance values chain´ (Takata et al. 2004, pp. 643–656).

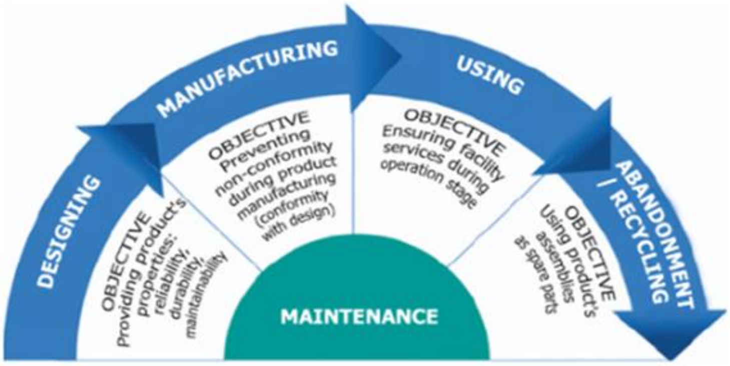


Figure 9. Maintenance in product life cycle (Jasiulewicz, Małgorzata, Drożyner & Przemysław., 2013)

Although maintenance services have not a direct impact neither on power consumption nor mount of generated waste, they may actively contribute to the reduction of environmental aspects identified in the organisation and to the improvement of its efficiency.

On the next figure it is shown that the maintenance management has a variety of tools that enable participation of technical services of an organisation in all phases of a machine life cycle, and thus, participation in the participation of pro-environmental strategy of the organisation.

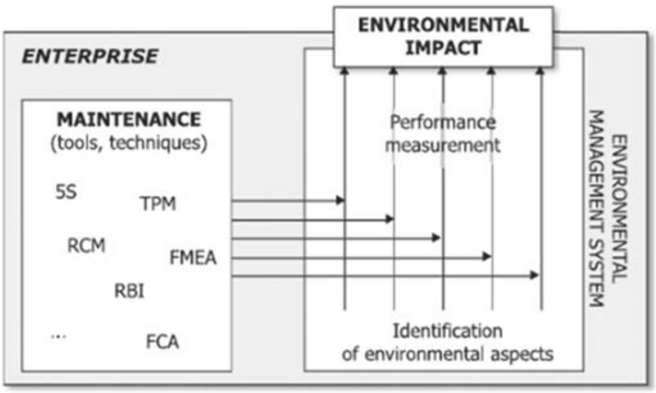


Figure 10. Maintenance methods and techniques and their environmental impact (Jasiulewicz, Małgorzata, Drożyner & Przemysław., 2013)

The first stage of a product's life cycle is its design. Basic pro-environmental prerequisites of the designing phase of a technical object include implementation of the '3R' principle (reduce, reuse, recycle), and in particular: selection of structural materials about environmental burden following their degradation (Cempel, Kasprzak & Kłos, 2006).

Total Productive Maintenance (TPM) is a maintenance strategy developed to meet the new maintenance needs (Shetty, Ali, Chapdelaine ,2009, pp. 117–136) (Ahuja & Khamba 2008, pp. 709–756). TPM is based on a ''Zero-loss'' concept with zero breakdown, accident and defects, to achieve high reliability, flexibility of equipment and reduce cost through minimizing wastage of manpower, raw material, energy, consumables, etc.

The next phase of a product's life cycle on which maintenance services have a great impact. From the ecological point of view, maintenance of infrastructure in the phase of equipment operation is focused on ensuring systems, procedures and trainings that build the operational knowledge and skills as well as functional possibilities of the systems to prevent, manage and eliminate losses and environmental incidents. A solution used in organisations to build the 'cleanness culture' is the Japanese 5S practice. A 5S cornerstone is ''the right thing in the right place at the right time''; anything else should be disposed of in a safe and environmentally correct manner.

Finally, abandonment is the last phase. It is when the machine reached its limit value of wear and further operation is impossible or uneconomical. The problem of managing of the worn parts and materials from the abandoned machine with the minimum burden to environment and certain economic result arises.

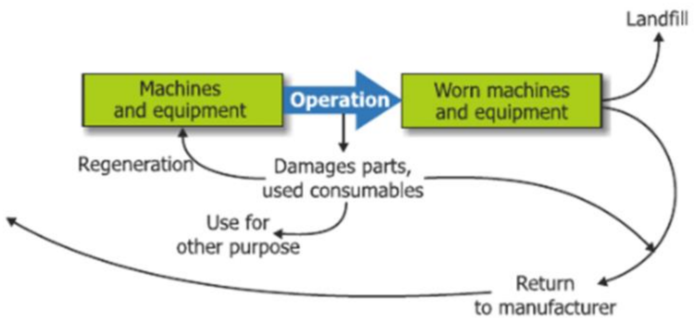


Figure 11. `Abandonment´ phase in a product´s life cycle (Jasiulewicz, Małgorzata, Drożyner & Przemysław., 2013)

Eco-design is an approach where complex technical objects are designed so that recycling of materials, from multiple uses and reuse of elements in several generations of machines for repair and modernisation is possible. The task of maintenance personnel is to appropriately assess the fitness of assemblies and parts for other machines and equipment owned by the organisation and repairing or regenerating thereof for further use.

- **Better product quality.**

Nowadays Global markets have become so competitive that to obtain a product, there are thousands of options. Markets have become literally in battlefields where competitors `fight´ to differentiate from each other's, and this difference can be reached through the quality of the product.

The anxiety of connection between maintenance and quality was discussed by (Daya & Duffuaa, 1995, pp. 20-26). They proposed a broad framework for modelling the maintenance and quality relationship. Then, the idea of incorporating maintenance in quality philosophies discussed in many literatures.

In the Total Quality Control approach (TQM), introduced by (Feigenbaum, 1991) maintenance is one of the characteristics of the total composite product, which will meet the expectations of the customer. In TQM, maintenance plays a role in quality (Lillrank, 1990, pp.277). Moreover, they pointed out that the best equipment will not work satisfactorily unless it is cared for. The ISO 9000 series states that process and equipment maintenance should be performed in a planned manner guaranteeing continuous performance. Moreover, the quality rate is becoming one of the three components affecting OEE.

Concluding, performance of equipment is substantial to preserve the performance of production process, as the increasing trend on mechanization and automation. Since the current market improved, customer demand high-quality product. The well-maintained equipment ensures the product quality will conform to requirements. It is worth to note, the importance of quality assurance through equipment maintenance becomes increasingly indispensable (Kurniati, Ruey-Huei, Yeh, Jong & Lin, 2015).

- **Longer equipment´s life.**

The fourth generation of maintenance is tightly related to the development of PdM techniques which are based mainly in continuously monitoring of the machines. With the help of the rise of digital systems, Big Data analysis, internet of Things technologies, PdM has become not only practical but critical to optimize operations and reduce costs. PdM means become the equipment more efficiently and to optimize the initial investment, by keeping those units running for longer than they otherwise would have.

- **Greater cost effectiveness.**

Fourth generation means PdM which is characterized with the rise of new connected technologies. This connection between all the equipment of a factory can enable machines to do tasks as the identification of failure modes or mitigation of downtime by themselves, which lead to maximizing the useful life of machine components as well avoiding machine failure. Today, poor maintenance strategies can reduce a plant’s overall productive capacity between 5 and 20 percent (Wollenhaupt, 2016). Recent studies also show that unplanned downtime costs industrial manufacturers an estimated $50 billion each year (Emerson, 2017).

OEE (Overall Equipment Effectiveness) can be related to cost effectiveness, thus, on the next figure it is shown that PdM is the best option to improve the OEE and cost effectiveness.

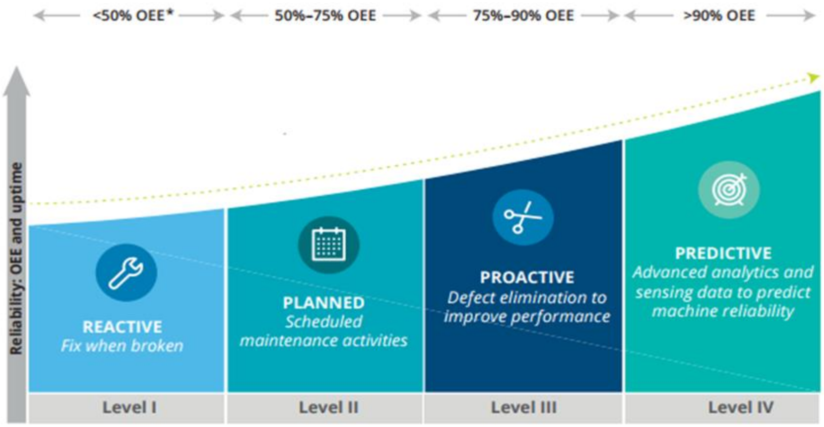


Figure 12. Maintenance strategy continuum related to OEE (Coleman, Damodaran, Chandramouli & Deuel 2017)

- **Risk Management.**

During the fourth generation on maintenance, organisations have started to focus on identifying and managing potential high consequence, low probability events, particularly in those organisations operating in hazardous industries. Maintenance is mainly related to this process and is being a key participant at the time to solve these problems.

On his report (Dunn, 2003) mentioned that in the past, these types of events were simply an extension of the routine Safety Management or Environmental Management systems in place at most organisations. However, in the last years managers have realised that these systems are not very good at effectively managing high consequence, low probability events. These types of systems are more suitable to manage high frequency low probability events.

These problems could be related to both quantitative risk assessment approaches and to those that do not formally rely on statistics, but instead rely on intuitive risk assessments approaches. Fundamentally, as human beings, we tend to underestimate the risks associated with high consequence, low probability events.

Defences against high consequence, low probability events are many layered and consist of various devices, systems and procedure. (Reason, 2001) and (Latino & Latino, 1999), have noted some of them that are intended to serve one or more of the following functions:

- To create understanding and awareness of the risks
- To give clear guidance and how to operate in such a manner as to avoid the risks
- To provide alarms and warnings when danger is imminent
- To restore the system to a safe state in an abnormal imminent
- To interpose safety barriers between the hazards and the potential losses

The defences are divided in hard and soft defences, some of them could be:

- Hard defences: Automated Safety Features, Physical Barriers, alarms…
- Soft defences: Legislation, Rules and Procedures, training, licensing supervision…

The defences are many-layered, and this generally means that no singles failure will result in a catastrophic failure, but any holes in these defensive layers must simultaneously line up at the same time as some initiating event, before a catastrophic occurs.

Successful defence requires the establishment and maintenance of a risk-aware, reliability- focused organisational culture, and this has more to do with the effective management of people, than it has to do with the analytical tool that one uses for assessing risks.

All the information provided about expectations of maintenance nowadays has helped to prove that the change from third to fourth generation of maintenance is real and organisations are trying to adapt to these new challenges. PdM is one of the most important characteristics of the fourth generation, most of the organisations recognise it and, therefore, implementing methods related to this is a priority for them.

1.3.1.2 Views on Equipment Failure.

Nowlan and Heap experiment was used to warn of the dangers of indiscriminately applying fixed interval change-out techniques for equipment and components. As (Dunn, 2003) mentioned, it is important to note that according to the results, significantly more than 50% of components experience early-life failures. This important factor means that whenever one replaces or repair a component, there is more than a 50% chance that it will fail early in its life. Moreover, on his report discuss the results of the study of Nuclear Power which also indicated that more than half of identified nuclear power plant performance problems were associated with maintenance, testing and calibration activities (quoted in Reason, p.92).

In the fourth generation of maintenance exist a necessity to reduce the proportion of maintenance repair tasks which, many times result in Failure Pattern F type probability distributions (failure pattern in the first stages of the machine running).

(Dunn, 2003) mentioned a second point regarding the knowledge of equipment failures. Since Nowlan and Heap presented their studies, most of the maintenance personnel have paid attention to the distribution patterns and they simply dedicated on predicting or preventing the failures. The author propose that a great advance would be also proactively seeking to eliminate those failures. As mentioned, in the fourth generation, maybe, the main goal of every maintenance manager should be to make the maintenance department (including him) redundant, and this only can be reached by eliminating totally failure modes. To eliminate proactively every failure, failure causes must be eliminated too, and this mainly implies a knowledge of those failure causes.

(Dunn, 2003) pointed that to apply PdM it will be required some tools and methodologies as:

- Organisations must ensure that equipment is designed in a way to ensure that it is *`fit for purpose´´*. This means that when designing or selecting the equipment, a consideration of the maintenance requirements, its maintainability and life cycle cost must to be more important that simply the initial capital cost.
- Equipment must be operating always within its design limitations. The goal of every equipment should be that every time the production process is optimised to give greater output, or better quality, or lower costs.
- Spare parts management processes are very important in maintenance thus, it is important that the processes are in place to ensure that the right spare parts are obtained and that they are adequately cared when it transited or stored.
- Equipment is in the right place is one of the key factors of every maintenance repair procedures, as well as that they are rigidly adhered to. Appropriate repair quality standards may need to be established.

Finally, it can be observed that the views on equipment failure have changed in the fourth generation. Low probability, high consequence failures will characterise the maintenance world and zero maintenance tasks will be one of the first goals.

1.3.1.3 Maintenance Techniques- Introduction to Predictive Maintenance

(Moubray,1997) listed many tools and techniques that were available to maintenance managers. These tools, methods and techniques were related to the third generation when PdM techniques yet had not lived their major development.

The origins of PdM begin with Condition Based Monitoring (CBM) which had its first appearance during the third generation of maintenance. The purpose of using preventive maintenance techniques was to prevent unplanned failures, but as the Nowlan and Heap report showed, at the end, more unplanned failures were created.

CBM technologies brought with it a huge level of specialization, and people needed to be trained in the use of the equipment and how to interpret the results. In the first years of the new century, information technology has developed in leaps and bounds. With the development of internet, big amount of information is almost freely available and can be in our hands only with searching quickly on this platform.

All these developments led the apparition of PdM techniques and methods. This new type of maintenance can be described as the next level of CBM, maintenance activities are determined by analysis of the historical process data and forecasts based on data trends (Swedish Standard Institute, 2001).

The difference between CBM techniques and PdM is that by analyzing historical data, greater knowledge and control of the equipment status can be obtained, instead of only looking at the current state, as CBM do (Mobley, 2002, pp. 99-113). PdM uses process data and advanced analytical methods to predict failures well before immediate action must be taken. Thus, this new technique is related to the development of Industry 4.0 or Smart Factory, which are composed by additional process that become data available.

1.3.2 Emerging trends and challenges- Shaping the future of Maintenance.

The trends will shape the future of production are described by (Finnin, Shipp, Gupta & Lal, 2012, p. 248) in their report "Emerging global trends in advanced manufacturing" for IDA explain the four major trends which are to be faced by production. The identified four trends are explained as follows;

1.3.2.1 Information Technology.

This takes an important paper in the maintenance tasks as every machine will have sensors which collect all the important data that describe the performance of them. The data collected will be analyzed and if any problem could appear, quickly a maintenance task would be done. As explained by (Sanders 2011), the evolution and usage of information technology improvises the production process which enables smart manufacturing.

1.3.2.2 Modelling and simulation.

Simulation and modelling enable a quick process from design stage to production (Finnin, Shipp, Gupta, & Lal, 2012). About maintenance, in the modelling phase the development and modelling of the product must consider the maintenance tasks. Thus, the modelling department and maintenance specialists should be in contact each a certain period to discuss what is the better for the interest of the organization.

1.3.2.3 Innovation in supply-chain management.

The supply chain management increased the usage of computer systems with innovative new technologies which include ERP software, RFID, Internet of Things and Predictive Analyses.

Maintenance management plays a major role in keeping the supply chain from crumbling down. In one way or another, maintenance activities can relate to every stop in the supply chain. The supply chain can be divided in some stages or phases which are transportation, production and storage. A failure in one of these stages could affect considerably the life cycle of the product and delays may appear. Thus, maintenance management plays an important role in the supply chain and as other areas of the organization, both must be connected from each other.

1.3.2.4 Rapid changeability in manufacturing.

The rapid changeability in manufacturing defines to meet customer needs and respond to external impediments as described by (Wiendhal et al. 2007). The changeability in manufacturing is divided from various evolutions as reconfigurability, flexibility, transformability and agility. Inevitable shift to leaner, smarter, more flexible production will have significant impacts on factories design, operation and control, supply chains and nature of work.

These trends will shape the world of production in the next years and maintenance managers must be prepared to face the new challenges that will appear. PdM will play an important role in the organizations thus, investment on this issue will help to improve the level, safety, environment, wealth of them.

2 Predictive Maintenance

2.1 Introduction of Predictive Maintenance

There are lots of definition about what PdM is, all of them are correct and express how the authors understand the role of PdM inside the organizations. (Mobley, 2002, pp.99-113), states that…

``*the common premise of PdM is that regular monitoring of the actual mechanical condition, operating efficiency, and other indicator of the operating condition of machine -trains and process systems will provide the data required to ensure the maximum interval between repairs and minimize the number and cost of unscheduled outages created by machine-train failures*´´

However, as he points PdM is much more…

``*It is the means of improving productivity, product quality, and overall effectiveness of manufacturing and production plants*´´

The analyses of data to predict what can happen in the future not only have to be transferred to the manufacturing of the product but also to other stages such as transportation of materials, storage of spare parts, final products…

As (Hashemian, 2011, pp. 226-236) mentioned, PdM works on the philosophy ``*Monitor, earn, predict and repair*´´, so, before attaining the eventual state of breakdown, equipment or machines show signs of imminent failures if pertinent actions are not performed timely.

PdM not only allows to achieve ``*Performance maintenance tasks, only when required*´´ but also it benefits by executing maintenance tasks without conflict with the production schedules which is very important for the organization as can they benefited from the perspective of cost savings, increase in the profit and higher machine availability for production (Jain, 2018).

2.1.1 The scope of Predictive Maintenance

Although PdM is the most efficiency program to implement in an organization it is important to consider that it has not to be a replacement of other techniques or methods. The necessity of preventive and corrective actions will not be eliminated upon the implementation of PdM; however, a PdM program can serve as a powerful complement to the existing maintenance strategy.

Basically, all maintenance activities initiated to prevent failures based on predictions can be labelled as PdM, but how the predictions are produced varies and ranges from very simple and analogous methods to highly advanced and technological ones. The basic premise is that this prediction should be based on regular monitoring of the condition of the machine (Mobley, 2002, pp.99-113).

Traditionally, PdM programs are focused on maintenance optimization efforts. However, the scope of predictive analytics can be extended to include more than maintenance management. It is a philosophy that can cover every part of the organizations and with the development of the Industry 4.0 and Internet of Things all these parts can be connected at real time which help to implement all maintenance strategies as well as of improving productivity, quality and overall effectiveness of the manufacturing and production plants (Smith, 2001).

In terms of plant optimization, PdM techniques can provide valuable information regarding the optimal production practices. Nowadays this is possible thanks to the implementation of modern techniques and the developing of microchips which are monitoring all machines in the plant and sending the results to either a web platform or a database. The data can be used to analyze the effect on the production system for different operating modes which allow to experiment different production modes in a safer way. This is a valuable capability since it provides a solid basis for decision-making regarding new investments. Further, it gives the opportunity to avoid costly corrective maintenance activities and reduction in useful life of critical equipment.

2.2 Differences with other Maintenance types

On the first episode of this thesis an historical analysis of the maintenance performance alongside the four generations of maintenance is made. After describing each one of the generations, it can be pointed that three main maintenance techniques or strategies have appeared over the time: corrective maintenance, preventive maintenance and PdM which is the aim of this study.

2.2.1 Corrective Maintenance.

Corrective Maintenance or ``run to failure´´ method appeared during the first generation of maintenance and can be defined as “Maintenance carried out after fault recognition and intended to put an item into a state in which it can perform a required function.” (prEN13306, 1998). It is also the most expensive one due to high machine downtime, low production availability, high overtime labor costs and high spare parts inventory cost (Mobley, 2004). The main difference between corrective and PdM is that the first one is a reactive tool while PdM is considered as a proactive method.

2.2.2 Preventive Maintenance.

The definition of Preventive Maintenance from the European standard (prEN 13306, 1998) is presented as: “Maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of an item.”. Preventive Maintenance seeks to decrease the likelihood of a machine´s failure through the performance of regular maintenance without considering any type of historical data about the performance of the machine, as said before it is based on prescribed criteria. It shares with PdM one aspect which is that both are proactive techniques, they are implemented before the failure took place.

2.2.3 Condition Based Monitoring.

Another important method that has been confused many times with PdM is CBM. According to the Professional’s Guide to Maintenance & Reliability Terminology, “CBM is the direction of maintenance actions based on indications of asset health as determined from non-invasive measurement of operation and condition indicators. CBM allows preventative and correction actions to be optimized by avoiding traditional calendar or run based maintenance.” A lot of disagreement appeared when differing CBM and PdM, but for me the main difference relies on that CBM is based on the actual condition of the equipment which means that it is close to analyzing the current state and depending on the results some measures will be taken. In contrary, PdM is based both on current and historical results and after collecting the data, using trends, algorithms and more tools, the failure mode can be predictive in time.

2.3 Important factors in Predictive Maintenance

Real time sensory data is important as by utilizing it, the intelligent maintenance prediction support system monitors the machine status. Advance maintenance in PdM policy can provide insights for maintenance scheduling in advance, to eliminate unanticipated machine breakdowns and minimize maintenance costs as well as downtime, before the occurrence of random machine failure (Garcia, Sanz-Bobi, & del Pico, 2006, pp. 552-568).

(Lee et al 2017), pointed some important factors in Maintenance Policy Management (MPM) as:

- Critically in Failures: In reliability studies, the critical assets of the organizations must have a higher rank in priority when formulating the PdM strategy to predict the most likely time for the next machine breakdown or random error, as this will have the greatest impact on the production operations.
- Reliability: Some targets of PdM policy are maintaining critical performance and leveraging the overall cost to sustain the production requirements. The system must provide the correct measures and reliable performance at the time of predicting feasible and foreseeable machine failure as well as building confidence in operation.
- Timeliness: Confidence is important at the time of predicting the maintenance modules before the undesired event occurs as well as data size and data transmission speed must be administrated in a timely manner. The time series of schedule and delivery of maintenance should be taken into consideration by the maintenance management to facilitate the production, with zero tolerance of equipment failure.
- Relevance: PdM systems needs to be developed based on the opinions of experts. To improve data quality, extraction of relevance data for maintenance decision making is crucial regarding engineering aspects. Thus, inappropriate integration of sensors and machines may cause poor estimation which led to an inaccuracy prediction of the current machine performance.
- Knowledge-Objective Oriented Strategy: The concept of the PdM strategy involves a belief that implicit knowledge from collaboration of sensory information contribute to the maintenance in advanced. The knowledge-transfer-system facilitates the disclosure of implicit information to maximize production efficiency and minimizes the adverse impact of idling time under maintenance and unawareness of potential failure. The decision of PdM policy could be assessed by the involvement of Big Data Mining Techniques to detect and defeat anomalies at an early stage.

2.4 Predictive Maintenance and the Industry 4.0

2.4.1 Industry 4.0.

During the last years the world of manufacturing has changed. There are a lot of terms to describe this new era as Industry 4.0, the 4th Industrial Revolution, the Internet of Things (IoT)… Anyway, all of them referred to the same attributes: doing things just in time, concurrently, more efficiently, with greater flexibility and in a safer and environmentally friendly manner.

The introduction of new technologies and services related to the Internet of Things is revolutionizing most of the industrial applications. Initiatives such as those in factories regarding automation and industrial PdM and initiatives to build smarter working environments are creating opportunities for new entrants and traditional players to offer innovative solutions that change business models (Tomarchio, Faulisi, Neumann & Janousek, 2018).

The fourth industrial revolution takes the automation of manufacturing processes to a new level by introducing customized and flexible mass production technologies. This means that machines will operate independently or cooperate with humans in creating a customer-oriented production field that constantly works on maintaining itself. The machine rather becomes an independent entity that can collect data, analyze it, and advise upon it. This becomes possible by introducing self-optimization, self-cognition, and self-customization into the industry. The manufacturers will be able to communicate with computers rather than operate them.

2.4.2 Important trends of the Industry 4.0.

There are a lot of terms related to the Industry 4.0, but here the most important or those which are commonly used and linked with PdM are the following:

2.4.2.1 Cyber Physical Systems.

Cyber-Physical systems are a basic pillar for the development of the Industry 4.0. These are defined as informatic and communication technologies incorporated in all types of devices, provided them with intelligence and a high efficiency.

Cyber-Physical systems combine communication functions, data and mechanical components using key technologies which include network sensors, communication infrastructures via Internet, event management smart in real time, providing with services for the "Big Data", functions of embedded software to solve logic and automatized and management operations of the activities from the system through the entire company.

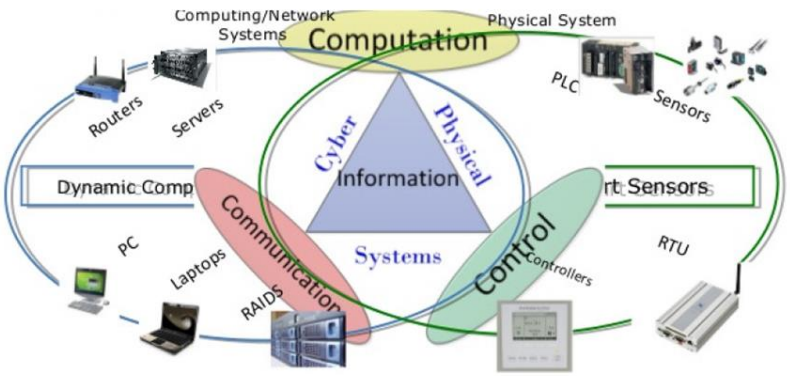


Figure 13. Cyber-Physical systems (Martinez García, 2015)

Some solutions implemented in real industries could be:

- *Control equipment* of a machine tool to optimize its Real-time performance.
- *Monitoring* of the state of the machine or, in general, of a system and the optimization of its operation and maintenance strategy.
- *Robots* that collaborate, consider context information and learn from each other.

– *Vehicles* that communicate with others and with the road infrastructure for determine the right speed or route.

2.4.2.2 Internet of Things.

There are hundreds of definitions about what is the IoT, but this term can be main described as a concept of connecting any device to the Internet and to other connected devices.

The IoT is a giant network of connected things and people, all of which collect and share data about the way they are used and about the environment around them. Nowadays machines, tools, objects… used in the industry are equipped with electric components, software, sensors and network connectivity which allow them to collect and change data through the Internet. This is mainly related with PdM as this technique basically works in this way.

The IoT is intrinsically related to Big Data, which allows to transform data into information.

2.4.2.3 Big Data.

It is the management and analysis of data whose quantities would be impossible to process with conventional tools and human resources. Big Data manages to provide valuable information about the behavior of different processes and services, which can be used to prevent problems, among other purposes.

Industrial automation systems integrate more and more sensors and communication capabilities, so factories must be able to gather enough data and interoperability between their processes. But to achieve real improvements in manufacturing efficiency and flexibility, manufacturers must be able to manage and analyze these large quantities of data.

The development of powerful solutions in the analysis of large quantities of data and the management of this knowledge is becoming a determinant factor in organizations. Companies must be implementing Big Data systems capable of processing, filtering and interpreting large quantities of data from the manufacturing environment.

2.4.2.4 Artificial Intelligence (AI).

It is the general term to define a set of computer systems that can feel, think, learn and take actions in response to what they are feeling and their objectives. This is achieved using tools such as bio-inspired algorithms, probabilistic reasoning, Machine Learning (ML) and Artificial Neural Networks (ANN).

AI elements need to be embedded and synchronized on a deterministic basis in the data across the digital core enabling the firm to derive new insights, such as predicting outcomes and automating actions to ensure optimize the outcomes.

2.4.2.5 Machine Learning.

Branch of AI that refers to the ability of a machine equipped with AI to manage and, above all, to learn automatically. Based on the identification and extraction of complex patterns among millions of data, an algorithm can extract information and obtain high value predictions about future behaviors for better decision making. This implies that, automatically, the systems are improved autonomously over time, without human intervention (Saez, 2018).

In the recent years, ML has become increasingly important in computer science because data could be collected and stored much more easily. The data collected is usually so extensive that it is not practical to analyse the data manually. In such a scenario, the ML technique plays a key role.

Another reason for the growing popularity of ML is the decreasing of computational costs. With the evolution of hardware in recent years, the usage of ML approaches has become efficient in terms of both time and money, especially for failure type detection and PdM.

2.4.2.6 Smart Factories.

Industry 4.0 combined with increasingly sophisticated analytics, is playing a huge role in driving the smart factory movement. Smart Factories are rapidly becoming the future of manufacturing, offering a new level of efficiency and productivity to those investing in them. Manufacturing executives and engineers no longer see a factory as a mass of machinery operating as part of one or more individual production lines. Instead, they see an interconnected network of moving parts, something related to a living and breathing organism, that can be fine-tuned to optimize performance.

With the development of new technologies such as Big Data prognostics, AI and the cloud can ensure that those who manage and maintain manufacturing environments are able to get on the front foot and proactively manage these environments, with relatively low levels of investment.

The concept of a Smart Factory is entirely dependent on the connectivity enabled by Industry 4.0. Machines that can sense and communicate provide a vast amount of valuable data. However, this data needs to be filtered and analyzed to be translated into actionable insight for manufacturers. Historically this analysis would have been a highly manual endeavor, involving teams of expensive data scientists.

More recently, however, organizations have developed intelligent software to automate this activity, creating a bespoke algorithm to identify problems and, crucially, spot the signs that indicate when a machine will fail in the future. This prognostic approach allows engineers to undertake precisely the right maintenance activities during periods of planned downtime and fix problems before they can affect production.

2.4.3 Predictive Maintenance in the Industry 4.0.

PdM plays an important role in the development of the Industry 4.0. Today´s industrial applications require a high level of automation and this only can be achieved by ensuring that the equipment is much more efficient, intelligent, aware of context and more connected. Moreover, humans still being equal or more important as equipment must be more robust and ensure greater safety for the humans interacting with them. Anyone running a factory wants to keep it running at optimal speed with minimal downtime. They are also aware that any machine with moving parts suffers wear and tear and inevitably requires servicing and repair. Thus, PdM is the solution as can predict maintenance requirements in advance avoiding equipment failure.

PdM combines condition monitoring with a dynamic predictive model for failure modes. This approach promises maximum protection of machinery and minimum productivity impact, without necessarily increasing overall system complexity. Until recently, fault indicators like vibration and sound emissions were almost exclusively monitored with discrete portable sensor probes based on piezoelectric or ultrasound sensing principles. Not only was diagnostic equipment based on these technologies relatively complex and expensive, there were also limitations related to repeatability, data management and analysis.

Nowadays, with the development of the Industry 4.0 and Smart Factories a new approach is possible to be carried out thanks to the state-of-the-art in electronics and advanced algorithms. Most modern industries can apply small, low-power sensing devices directly on machines and monitor multiple parameters, pre-process acquired signal data, and send data to the local, remote or cloud-based analysis and control facilities. These compact smart sensors nodes consist on sensors, a microcontroller, power management circuitry and wired or wireless connectivity, which offer many advantages over traditional condition monitoring equipment (Tomarchio et al. 2018):

- *Cost:* smart sensors with autonomous operation are cheaper than portable piezoelectric probes as well as do not have the necessity to employ skilled technicians required to operate them.
- *Ability to trigger immediate local action:* with the implementation of intelligent algorithms it is possible to analyse data locally on the node and trigger immediate actions to protect the equipment and ensure worker safety.
- *Repeatable, reliable and timely data:* one of the main advantages of these devices is that using continuous measurement and analysis during machine operation, it is no possible to miss early or critical failure signs, as with scheduled maintenance use to happen.
- This type of sensors can be used both in local database systems and in cloud networks to gather large amounts of data for deeper PdM analyses.

PdM has clearly marked a turning point in the world of industrial services. Unlike previous approaches, such as reactive service models, Preventive Maintenance, and Condition-Based Maintenance, PdM adds a critical edge to the use of sensors collecting and measuring datasets (Feldman, Buechele & Preveden, 2018).

(Sciban, 2017) presented two key reasons for rising PdM growth.

First reason is related to the modern machinery which nowadays, usually integrates embedded computer chips for reading and controlling the state of the machines, developing a potential system for data capture. Secondly, the author pointed that the cost of implementing and using embedded sensor and other new technologies for recording datasets is continuously being reduced. Moreover, (Feldman et al. 2018) wrote that the drivers for PdM are already well developed in the areas of sensor technology, data and signal processing, and condition monitoring and diagnosis.

However, the real challenge for this generation lies in predictive ability, process and decision support, and the resulting opportunities in service and business models. PdM today is highly related to Smart Factories which has carried out to that, this type of maintenance is built on the four cornerstones of digitalization: interconnectivity, digital data, automation and value creation.

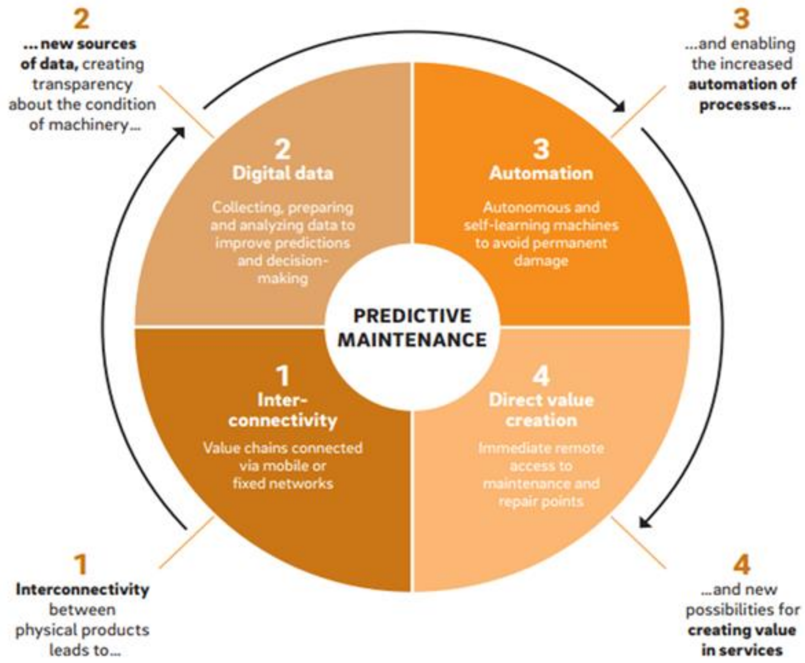


Figure 14. From interconnectivity to value creation: the four cornerstones of digitalization (Feldman, Buechele & Preveden, 2018)

2.5 How does Predictive Maintenance work?

Condition Monitoring and PdM systems include several smart sensor nodes connected in a network via a gateway to an edge server or to a cloud service.

Depending on the ERP architecture, there are different ways of. processing data. Many organizations use sensor nodes which have embedded analytics so that data can be processed immediately. On the other hand, nowadays have appeared modern methods which use on remote servers or cloud infrastructure which can process and correlate many data sets over time. In these cases, systems are usually connected to the Maintenance and Procurement component of ERP systems for timely ordering of spare parts.

Failure modes can be detected from some ways. If machine deterioration is detected at the node levels, immediate corrective actions can be set up in the machine which will allow to prevent further damage and failure. When failure modes or machine deterioration cannot be found at sensor levels, longer term analysis and actions are managed on the cloud, allowing more complex analytics on large amounts of pre-processed data, which can be useful to determine trends and optimize local analysis models.

2.5.1 Predictive Maintenance architecture.

PdM architecture is based on an availability of a production asset data which is streamed from the sensors to a central repository using industrial communication protocols and gateways. After that, bussies data from ERP and MES systems, together with manufacturing process flows, are integrated into the central data repository to provide context to the production asset data. Then, predictive analytics algorithms are applied to provide insights for reducing downtime, predict possible machine failures… which are investigating using different techniques as ML, NNs, Data Mining, root cause analysis…

Next figure shows the common PdM architecture that can be found in many organisations which belong to the industry 4.0.

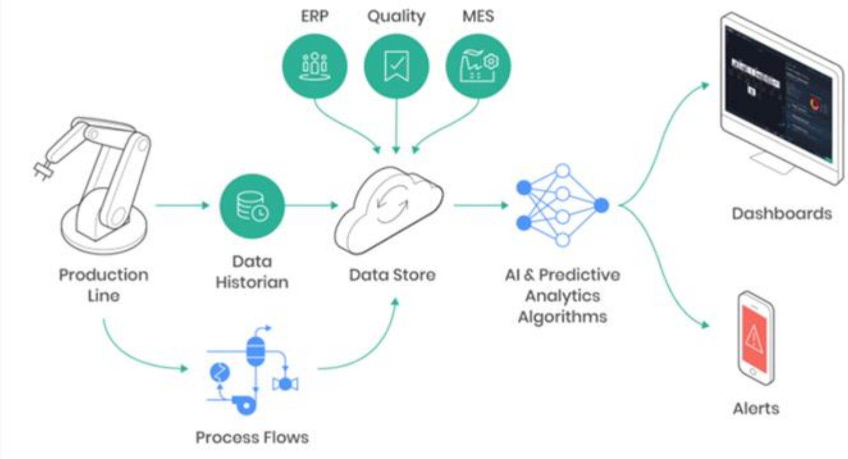


Figure 15. PdM Architecture (Seebo, 2019)

This connected system is based on the manufacturing assets and sensors, the business systems, communication protocols, gateways, cloud, predictive analytics and visualization.

Engineering can graphically capture the production process using a visual IoT modeler with rules that monitor and alert to maintenance issues. This modeler will generate a system blueprint critical at the time of getting accurate predictive analytics. Then, predictive analytics will be applied to the machine data, and the system blueprint data will try to predict conditions of upcoming failures. A dashboard for predictive analytics will synthesize operational data, allowing process and maintenance engineering to address actionable insights in the form of corrective actions.

2.5.2 Tools, technologies and techniques required.

At the time of implementing a PdM program it will be important that, to be carried out on an industrial asset, several components will be required:

2.5.2.1 Importance of sensors.

PdM techniques are based mainly on collecting large amount of data from the machines, for after that, analyzing them in a timely manner to try to find possible failure modes and carried out the necessary actions. Thus, sensors play an important role in the stage of collecting data which will be crucial to the next steps.

Sensors of various types can be used, including temperature, pressure, humidity sensors for environmental data, accelerometers for vibration measurement, current sending devices, microphones for ultrasound… among others. Moreover, it is essential to periodically verify proper sensor operation, and the self- test feature allows verification of the functionality of any sensor, regardless of where it is positioned, without moving it.

Next figure shows the numerous components in a smart sensor node which is described through a block diagram.

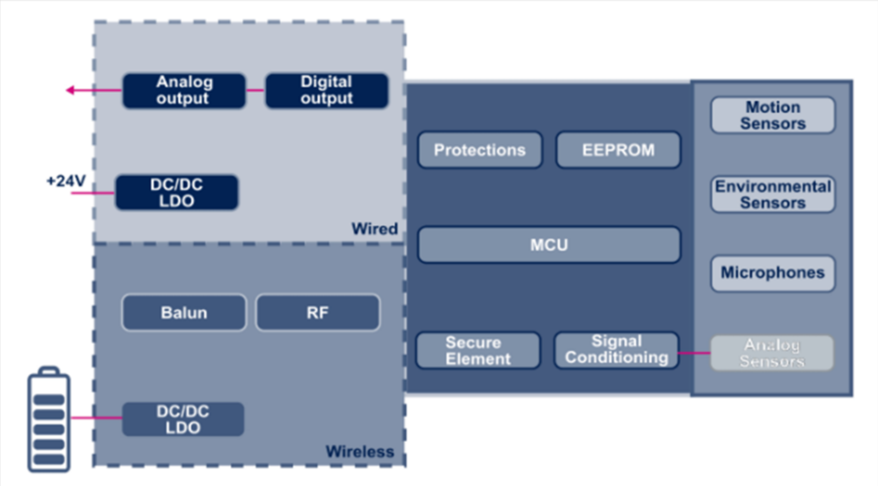


Figure 16. Block Diagram of a Sensor Node (Wired or Wireless) for Vibration Analysis (Tomarchio et al. 2018)

- **Common Sensors.**

- Accelerometers: common type of sensor used particularly for failure detection in rotating machines, where acceleration, velocity and amplitude of a vibration signal are very useful as they can provide warning signs for imminent failures.
- Microelectromechanical systems (MEMS): used in smartphones and other consumer services are gaining popularity in industrial applications. What makes so special is that these devices cost less, offer more flexibility and they are competing with piezoelectric sensors in terms of accuracy, stability, bandwidth, temperature and dynamic range.
- Sound Emission Analyzers: using to detect several faults as gas leaks and fan and motor imbalance, analyzing emission performed in both the acoustic and ultrasonic spectre.
- Microcontrollers: provide local data processing capability with data capture, processing and communication handling. There are lots of different types of them which offer different features in terms of processing power, memory and interfaces. According to application requirements, one can select the type of Microcontroller which adjust better to them. Perform Embedded Analytics is one of the main processing tasks of the microcontrollers, commonly domain and frequency spectrum analysis (FTT) require a high-quality processing, as well as other embedded analytics.

2.5.2.2 Data Communication.

PdM requires an Internet of Things infrastructure, the backbone that wirelessly connects your assets to your maintenance data center and enables the collection and distribution of sensor data. Setting up the proper IoT infrastructure for a company involves choosing the right protocols for wireless connectivity, data encryption and security. Connectivity is managed through standard wired and industrial protocols (such as IO-Link, Industrial Ethernet, and Modbus) as well as wireless (Bluetooth Low Energy, Wi-Fi, and cellular). There is no unique solution, and the factory infrastructure and other circumstances will determine the choice.

The communication system that allows data to securely flow between the monitored asset and the central data store will be needed. Fast development of ICT (Informatic and Communication

Technologies), including industrial Big Data analysis, AI, Cloud Computing, Edge Computing, 5G Communication… are powerful enablers to PdM.

2.5.2.3 Central Data Store.

The central data hub in which asset data (from OT systems), and business data (from IT systems) are stored, processed and analyzed; either on-premise or on-cloud. Data are stored and labelled in different folders so, here is important to try to have a high level of organization so that data can be available in the shortest possible time.

(Lavi, 2018) on his article wrote that a robust ERP system that can handle unstructured data is the key to unlocking the value of collected data. There is a need to integrate and aggregate data collected from separate solutions from different product manufacturers to create meaningful data to create insights. After aggregating the necessary data, a system that can store the data, process it and analyse it, will be needed. Integration platforms can be a good alternative to hand-coded integration, because they provide a more flexible environment that can handle multiple integrations with different systems that frequently require updating.

An integrated ERP-MES platform empowers Industry 4.0 by connecting core systems to enable manufacturers to have all the necessary data combined to gain the needed insights to achieve higher levels of quality and productivity. Management ERP manages the business of manufacturing products while MES controls the production process itself. Integrating ERP and MES systems can enable alerts for immediate actions to prevent downtime. Maintenance management is much more effective if brought to the operations level, where it can be integrated with production processes closer to where things happen.

2.5.2.4 Big Data Analytics.

(Lee, Cao & Kam, 2017) wrote that Big Data Analytics techniques identify the common characteristics of data with the purpose of finding patterns and relationships existed in the data. The descriptive functions of big data mining include classification analysis, clustering analysis, association analysis and logistic regression between others.

- Classification Analysis: this type of analysis is a typical learning model which aims to build a model for making prediction on data features from predefined set of classes according to certain criteria. In some cases, a rule-based classification is used to extract IF-THEN rules to classify as different categories. The examples include NNs (which will be explained later in detail), Decision Trees and Support Vector Machine.
- Clustering Analysis: this type of analysis can be defined as the process of grouping data into separate cluster of similar objects, which helps to segment and find the data features. Data is divided into different subgroups according to their characteristics. Moreover, appropriate strategies for forming different clusters might be formulated. The common example of clustering technique are K-means Algorithm, Self-organizing Map, Hill Climbing Algorithm and Density-based Spatial Clustering.
- Association Analysis: Association models are very useful to recognize groups of items that occur synchronously. Association algorithms are developed for searching frequent sets of items with a minimum of specified confidence level. The criteria support and confidence levels help to identify the most important relationships among the related items.
- Regression Analysis: This technique is used to represent the logical relationship of the historical data. The focus is to measure and to identify the dependent variable, given one or several

independent variables, which are related to the dependent variable and support the conditional estimation of expected outcome using regression functions. There are a wide variety of statistical regression functions, but the most known and used are; Linear Regression, Non-linear Regression and Exponential Regression.

- **Predictive Maintenance in Big Data framework.**

Next figure shows the framework of a big data platform in PdM for closer of data acquisition and the Maintenance Decision Support System (MDSS), which highlights the dataflow process in diagnostics and prognostics modelling for PdM.

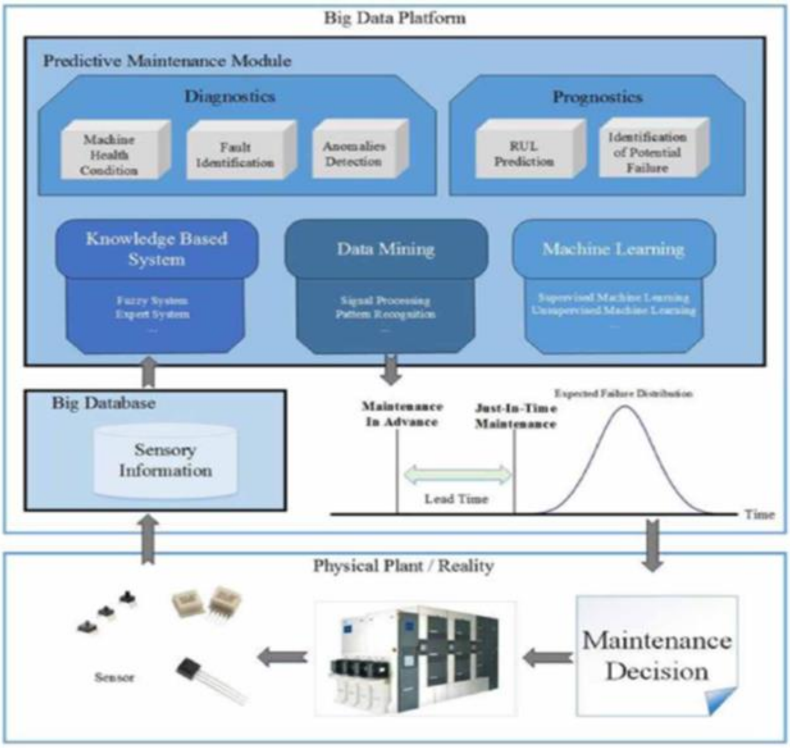


Figure 17. PdM model in Big Data framework (Lee, Cao & Kam, 2017)

The advantage of Big Data Architecture is the capacity to manage huge units of data and perform ETL (Extract, Transform and Load) in a timely manner by using appropriate data processing algorithms such as Map- Reduces techniques, ML algorithms…

The PdM module supervises the machine condition and frequently aids the analytics process of diagnosis and prognosis of failures. At the time of performing predictive analytics, AI is a modern tool which will give some results that will be processed and evaluated by the Maintenance Decision Support System (MDSS). Maintenance Decision Making must perform certain operational guidance and estimation of failure events, such as foreseeable situations, time to breakdown and estimated downtime.

It is practical for a decision maker to select appropriate analytic processes and recognize the functionalities of algorithms for maintenance planning. The common practices of AI techniques and Predictive Analytics are Knowledge Based System (KBS), Data Mining (DM) and ML (ML*)* (Faiz & Edirisinghe, 2009, pp. 23-36).

2.5.2.5 Predictive Analytics.

Predictive Analytics can be described as the branch of the advanced analytics which are used to make predictions about unknown future events (Imanuel, 2018).

Basically, these analytics try to find patterns in historical and transactional data and through them identifying risks and opportunities for the future. Predictive analytics models capture relationships among many factors to assess risk with a set of conditions to assign them a score or weightage. Techniques as Data Mining and Text Analytics along with Statistics, allow to create predict intelligence by uncovering patterns and relationships in both the structured and unstructured data.

Structured Data is referred to data which can be used readily for analysis. Unstructured Data is data which need to be extracted from the text, social media content, images…. And then used them in the model building process (Imanuel, 2018). Predictive Analytics allows organizations to become proactive, forward looking, anticipating outcomes, behaviors and failure modes, based upon the data and not on a hunch or assumptions. Between all the analytical techniques here are some of the most important:

- **Knowledge-Based Systems.**

These types of analytics process require logical deduction and cognitive reasoning to resolve complex problems and support decision making. The main characteristic of KBS is its attempt to extract rules for algorithm contexts by human intelligence and expert opinion, which are of practical significance.

KBS is becoming more and more necessary and its popularity is increasing due to the advancement of sensor based PdM. This technique set a more flexible way to increase quality in problem solving and in extracting relevant data into knowledge for decision making, being commonly used in machines to failure identification and in classification of maintenance policies. A variation of sensory information causes data-booming in the analytics process. The rule-based and inference engine expert system can simulate a human expert in reducing the complexity in Manufacturing Process Management and in discovering hidden machine failures.

- **Data Mining.**

(Imanuel, 2018) pointed that Data Mining can be described as the computational process of discovering patterns, trends and behaviours in large data sets using some of the most modern techniques like AI, ML, Statistics…The main goal of Data Mining is to extract information from a data set and transform it into an understandable structure for further use.

Data Mining is an essential step in the process of predictive analytics as allow mining and extraction of useful information from the existing data which is an interdisciplinary work since involves mathematicians, statisticians, data scientists, computer programmes…

This Data Mining plan has 5 steps which are as follows:

1. Sample: referred to extract a portion of a large data collected which is big enough to contain the significant information and small enough to manipulate quickly.
2. Exploration: searching for unanticipated trends and anomalies for gaining understanding and ideas.
3. Modification: not all data collected is necessary and thus, creating, selecting and transforming the variables is a key of success to focus the model construction process.

4. Modelling: Searching automatically for a variable combination that reliably predicts a desired outcome. Later different modelling techniques for PdM will be explained.
5. Assessment: Key step for evaluating the usefulness and reliability of findings from the data mining process.

- **Machine Learning.**

ML is considered as another dimension of the analytics process. The two techniques mentioned before (KBS and Data Mining) are more focused on discovering knowledge and insight beforehand for the working process of the algorithm, which is concerned information and knowledge extraction from massive data.

ML, on the other hand, deals with automatic reasoning and artificial cognitive resolution by an intelligence agent (Imanuel, 2018). This tool works as an online measurement of a health detection system to reveal machine degradation and anomalies from the models. Self-Learning and reinforcement in ML together, with Normal Degradation, allow the forecasting of random machine failure effectively and efficiently to plan for the best before failure occurs.

As ML gets important relevance in PdM, a large chapter in this thesis will be dedicated to it.

- **Artificial Neural Networks.**

Machine failures can be caused by normal degradation and random failure. (Lee, Cao & Kam, 2017) mentioned, basic KBS and Data Mining are incapable of distinguishing between random failure and normal degradation, since machine degradation follows time progression instead of rules. Therefore, an adoptive approach of ML to prognosticate failure is required.

The ANN model is a supervised learning model which can estimate the Remaining Useful Lifetime (RUL) of a machine from degraded failure, as well as making a classification of failure types, or to differentiate anomalies from normal machine behavior. ANN can be used as a time series or forecasting of unanticipated machine failure with online sensory data, which is able to learn patterns from the training data set to distinguish the normal machine behavior and any anomalies.

The prediction in machine failure involves Big Data management with the purpose of strengthening the data quality. Continuous data collection from different sensors installed in the machine support the quality of prediction.

Because of the importance of this issue in PdM, ANN are explained more in detail in chapter 4.

2.6 Challenges when establishing Predictive Maintenance

(Lavi, 2018) wrote that there are several pieces of the puzzle that need to be put in place before starting to implement a PdM program. Machines, devices, sensors and people must be connected and communicated all together in a perfect way. As there is a necessity to have a virtual copy of the physical world to make sense of all the data to conceptualize the information, the most sophisticated technologies need to be used. These technologies will help to support decision making and problem solving, making cyber systems as autonomous as possible.

(Cousineau, 2018) wrote some specific requirements or challenges to consider when implementing the PdM program:

– Bringing the organization into the Industry 4.0

To implement the PdM techniques, organizations need to develop some important tools, systems, technologies of the Industry 4.0. Starting with state-of-the-art sensors whose technologies can be linked to large amounts of data in almost real time is a crucial step. These sensors need to be able to monitor conditions with a high level of reliability in real time to provide meaningful data. Many sensors which probably the organisation already have, are limited by antiquated technology and they will need to be updated.

– Data monitoring and evaluating

One crucial step to achieve a successful PdM strategy is to have the ability of collecting and analysing data to accurately predicting failure patterns. PdM strategies will not last very long if predictive technologies cannot be used on critical assets or if no one can monitor, evaluate and act on the data.

Evaluating equipment and ensuring that PdM software technology can capture data used to identify failures is one of the first steps. Moreover, algorithms must be created to predict failures and outcomes which sometimes can become a difficult issue.

Finally, data cleaning and mapping must be done, and a plan must be put in place to constantly collect, monitor and analyse data.

– Cost and availability of predictive technology

PdM can be very expensive as this program requires a wide variety of different technologies to run efficiently (Cousineau, 2018). These programmes sometimes may require high amounts of investment to upgrade decades old equipment with smart assets or to integrate predictive technology into these old machines. So, it is crucial to know if the technology even exists or is available for the type of measurements and assets that you are trying to analyse.

However, sometimes it may be the case that even although the cost and installation of technology can be supported, the software, hardware and algorithms used are still in the early steps compared with other maintenance solutions, which can lead to an increase in the cost.

– Expertise, training and workplace culture

The entire organisation must be concerned about creating new skill sets to build, monitor, analyse and maintain the system. PdM not only concerns one part of a company, the different areas must work together at the time of developing the solutions. When creating algorithms, data scientists, reliability engineers and other members from different teams will be required to work together.

The skills to carry out the PdM solutions are often hard to find and require from a wide variety of experts in a lot of different techniques which can require partnering with multiple outside providers.

Finally, there is a necessity of willingness from all areas of the organization to embrace PdM strategies and methods, from operators and technicians to the C-suite.

– Security

PdM normally use together different assets and digital systems using Cloud Technology and IoT techniques, thus it can require a great investment in protecting assets. Between all the measurements that can be taken some of the most important are safeguarding access and the adoption proactive stance toward cybersecurity.

Time, skill, effort and financial resources must all always be available to ensure that the PdM technologies are safe and secure.

3 Implementing a Predictive Maintenance Program

Increasing equipment reliability and reduce unscheduled downtime are some of the main goals of companies nowadays. To try to achieve that, many organisations have taken the proactive step of implementing a PdM Program. Unfortunately, it has been discovered that only some of these initiatives actual achieve the anticipated results due to common mistakes that companies used to do. Thus, to achieve good results at the time of implementing PdM techniques some issues must be considered.

There are two stages of implementation. The first phase is analysis and design, which consists in points such as: data collection, assessment of condition´s equipment, the view of possible repair actions, renewal of machinery and design of inspection routines. The second stage deals with the allocation of resources by the company to the maintenance unit and the execution of predictive techniques.

Implementing a PdM Program is not an easy task and there are some steps that must be done to achieve accurate results and improve the effectiveness of the company. In this chapter, the basics steps to follow when implementing such program and get a successful result are explained.

3.1 Business Analysis

First, at the time of starting to implement a PdM Program, as in every project, the first thing is the understanding from the business area to know their needs, their current situation and their expectations.

It is important that during the analysis the following issues are considered:

3.1.1 Analysing the current situation.

When starting to implement a PdM Program to avoid any types of upcoming failures and to detect them, a key factor is to evaluate what has led to this situation, which facts gave rise to start such a complex project. Doing this will help to understand which the current situation of the company is, and which problems have happened related to the machines or the production line.

Moreover, another important task in this step is to know which is the significant equipment and indicators that are related to the maintenance problems, from which parts or machines data will be needed, as well as whether the data is available or not. Some features about this data must be known as how often the data is updated, the people who manages them, if there have been any failure in the components life and if it was measured… must be known.

3.1.1.1 Identification of significant equipment and indicators.

Focusing on the most critical equipment and enabling the maintenance department to quantify the cost benefits of the system is one of the mandatory steps to be carried out.

Defining what machinery and equipment are going to be included when setting up the solution will be the scope of this step. This decision should be based on the results of a critically analysis as (Groba, Cech, Rosenthal & Gössling, 2007) wrote.

- **Criticality Analysis (Analyzing the equipment)**

Criticality analysis can assess the importance of the concerned equipment to the purpose for which it is being utilized. Criticality must define the priority of equipment that is used for allocating the maintenance budget. This step is important as allow to know which components a severe impact on the production output in case of failure will have, which will be related to a high level of criticality. However, assessing the impact of failure to production is not an easy task since it involves being able to know several key factors such as cost of equipment, failure frequency, cost of replacement, maintainability of equipment and safety issues related to equipment failure (Gomez de Leon Hijes & Cartagena. 2006, pp. 444-451).

(Duraccio et al. 2014) proposed a methodology base on the research in the main areas of maintenance. The method that is going to be suggested aims to quantify units' criticalities, identifying the most relevant to ensure the production. Such criticalities are described by several factors with their indices to quantify the importance of all of them.

Factors related to critically of Machines

For the application of the proposed methodology, factors considered critical to the maintenance of the machines need to be analysed, which are numerical described by a structure of indices.

1. Level of use.

As not all the units have the same relevance in a plant and ones are more involved than others in the production process, the criticality will depend on such factor. The unavailability of a highly used machine put on risk the whole process. Such level depends on two factors that can be numerical described by its own indices:

– Working time of machine compared to running time of the plant.

This technical indicator measures the importance that the unit or the machine covers in the production process. Such factor is described by the *Flow Index* (FL_i) for the machine i-th which is equal to the ratio between the flow processed by the unit (F_i) and the total materials flow of the plant (F_{tot}). It always takes positive values between 0 and 1 being those close to 1 who represent indispensable units of production.

$$FL_i = \frac{Fi}{Ftot}$$

In calculating the index, it is necessary to measure the flow of materials in a way suited to the specific production process (number of pieces, weight, capacity). (Duraccio et al. 2014) explain more in detail the method for some cases.

– Flow of material crossing the unit, compared to the overall flow of the plant.

This is measured by the Time Index (T_i) for the unit or machine i-th which is the ratio between the working time of the machine for a unit of product and the takt time of the product. In case of multi-product company, where the same machine is used to produce different products, the index formula is different. $P_1...P_n$ are the products made by the company**.** $TC_1...TC_n$ are the takt times for each product and T_{ij} is the time in which the i-th machine works the product j-th; the machine time index is between 0 and 1 and is:

$$\mathrm{Ti} = \frac{\sum_{j=1}^{n} \mathrm{Tij}}{\sum_{j=1}^{n} \mathrm{TCj}}$$

2. Maintenance time.

To measure how important maintenance is in each machine the *Maintenance Index (M_i)* is given which is the ratio between the average number of hours spent in maintenance in the i-th station (TM_i) and the average number of hours of maintenance time, calculated for the unit with the highest maintenance time (TM_{max}).

$$\mathrm{Mi} = \frac{TMi}{TMmax}$$

3. Maintenance cost.

Maintenance is an investment which depends on production volumes and the importance of the machine in the production cycle. There are many types of costs related to main types of activities like the cost of spare plants and equipment, internal and external labour cost, cost of image… thus, maintenance´s investment must be weighted and focus on the departments whose unavailability can compromise the production process. To measure it, the *Cost Index (C_i)* is given, which is the ratio between the annual cost of maintenance for the i-th unit *(CM_i)* and the total cost of maintenance of the system (CM_{tot}) in a year, such index only considers the costs of industrial maintenance.

$$Ci = \frac{CMi}{CMtot}$$

4. Variety of Faults.

Variety of faults affects the maintenance operations thus; different types of faults and the structural complexity of unit generally is a synonym of high complexity maintenance operations. To measure it, the *Failure Index (Fa_i)* is introduced which is the ratio between the failure modes occurring in the i-th machine (Fm_i) and all the different failure modes (Fm_{tot}) of the plant (in a year). Such term refers to the diversity of faults and not their number thus, if the same fault occurs more than once, the index it should be counted at as one.

$$Fai = \frac{Fmi}{Fmtot}$$

- **Maintenance Critical Analysis steps.**

The introduced methodology, called Maintenance Critical Analysis, aims to identify units or machines that have the most critical maintenance and require major investments. The MCA method can be summarized in some steps (Duraccio et al. 2014):

1. Plant breakdown.

The first step is defining the units of machines that break down the system to be analysed. Here, some basics rules must be observed:

- The number of components must be the adequate to be able to describe the system, but it must not create problems of data management due to too detailed breakdown.
- For such components, the parameters used for the analysis must be uniquely and easily identified.

2. Data collection.

For such analysis the data which describes the selected components need to be collected to calculate the indices mentioned before. This step use to require an extended period.

3. Indices Calculation.

Once the necessary data are been collected, the indices for the analysed components are calculated.

4. Maintenance Priority Index (M.P.I) calculation.

The calculated indices are multiplied and result in a single index, called the Maintenance Priority index (MPI) which formula is described as:

$$\mathrm{MPIi} = \mathrm{FLi} \times \mathrm{Ti} \times \mathrm{Mi} \times \mathrm{Fai} \times \mathrm{Ci}$$

Such index can take values between 0 and 1, directly proportional to the criticality related to the unit or the machine.

5. Analysis of the results.

Finally, by ordering the MPI values, a list with the priority maintenance components can be made. According to these priorities, maintenance must be designed, and budgets must be allocated. Units with higher values of MPI are the ones in which more resources should be used in terms of maintenance activities. Their unavailability may in fact seriously undermine the success of the production process (Duraccio et al. 2014).

After making a description about the different steps on the mentioned methods, it is important to highlight some important features (Duraccio et al. 2014).:

First, it is important to consider that such method is a tool for performing preliminary analysis, which can be described by for its simplicity, both in mathematical treatment and in the application. This allows a profitable use, without excessive use of resources.

As well, since the used indicators are characterised by being very general and simples to understand, this method can be easy to apply in any industrial plant, regardless its peculiarities. By repeating the maintenance analysis on selected units after the first application, the method can be also used as comparison and to check the results.

Moreover, it must be considered that there are other types of method which can be used to analyse the equipment of the plant. FMEA is another technique which has a very similar logic to the Criticality Analysis, although using different indicators. FMEA is focused on the analysing of failure modes while Criticality Analysis considers many factors and its main aim is related to the analysis of units and machine in general.

- **Other techniques.**

– Multi-criterion Classification of Critical Equipment (MCCE):

is used to calculate a critically index which is based on several factors related to equipment failure, each of them unique to the concerned process and company. Each factor is linked with a weight, indicating its relative importance to the other factors. (Gomez de Leon Hijes & Cartagena 2006, pp. 444-451) in their book explain more in detail the structure of such technique and how it works.

– ABC Critically Rating Method:

On the other hand, the ABC critically rating method relies on an Activity-Based Costing (ABC) analysis, which consists on a method for determining the actual cost of necessary activities to produce a certain product. Each potential failure mode for all plant equipment is analyzed and the calculated associated cost determines the criticality of the equipment (Sondalini 2004).

– FMEA:

Finally, as mentioned before FMEA is a very useful method to describe the failure modes of the equipment in the industrial plants, as there is a lot of literature written about this methodology and can be used in different ways, here there is not a detailed explanation about such technique although some main points are going to be described.

A typical FMEA analysis incorporates some methods to evaluate the risk associated with the potential problems identified through the analysis. The most common method of evaluation of risk is Risk Priority Number which can be performed in the components of the industrial plant.

(RPN) Risk Priority Number method to assess risk is based on the analysis done by the team who must rate the Severity of each effect of the failure, rate the likelihood of Occurrence for each cause of failure, rate the likelihood of prior Detection for each cause of failure (likelihood of detecting the Problem before it reaches the end user or customer), and then, calculating the RPN by obtaining the product of the three ratings. Depending on the RPN result of the equipment, this can be classified in a level of risk so that the importance of such tool is assessed.

$$RPN = Severity \times Occurrence \times Detection.$$

The three indicated parameters use to be classified in tables where each of the parameters are evaluated given them a certain value depending on the rate of each of them. Some examples about the tables are showed in the next figure:

Rating	Severity of Effect	Likelihood of Occurence	Ability to Detect
10	Hazardous without warning	Very high: Failure is almost inevitable	Can not detect
9	Hazardous with warning		Very remote chance of detection
8	Loss of Primary function	High: Repeated failures	Remote chance of detection
7	Reduced primary function performance		Very low chance of detection
6	Loss of secondary function	Moderate: Occasional failures	Low chance of detection
5	Reduced secondary function performance		Moderate chance of detection
4	Minor defect noticed by most customers		Moderately high chance of detection
3	Minor defect noticed by some customers	Low: Relatively few failures	High chance of detection
2	Minor defect noticed by discriminating customers		Very high chance of detection
1	No effect	Remote: Failure is unlikely	Almost certain detection

Figure 18. The three dimensions (Occurrence), (Severity) and (Detection) of System FMEA (FMEA Ratings) (SPCCONSULTING, 2012)

Then, usually organisations have developed their own tables where they describe the *FMECA analysis*, incorporating there the part, component, or machine assessed, the failure mode, the cause of the failure and the effects, as well as the parameters indicated and in some cases any recommendations. Next figure shows an example of it:

Process Function	Potential Failure Mode	Potential Effect(s) of Failure	Sev	Class	Potential Cause(s)/ Mechanism(s) of Failure	Occur	Current Process Controls	Detec	RPN	Recommended Action(s)	Responsibility and Target Completion Date	Action Results				
												Actions Taken	Sev	Occ	Det	RPN
Drill Blind Hole	Hole to deep	Break through bottom of plate	7		Improper machine set up	3	Operator training and instructions	3	63							0
	Hole not deep enough	Incomplete thread form	5		Improper machine set up	3	Operator training and instructions	3	45							0
			5		Broken Drill	5	None	9	225	Install Tool Detectors	J. Doe	3/1/2008	5	5	1	25
									0							0
									0							0
									0							0
									0							0
									0							0
									0							0
									0							0
									0							0
									0							0
									0							0

Figure 19. Example of FMEA analysis general table (SPCCONSULTING, 2012, "Six Sigma Tools: FMEA types")

The Failure Modes, Effects and Analysis (FMEA) procedure is a tool that has been adapted in many ways for various purposes. The tool can be used to establish and optimize maintenance plans for repairable systems and/or contribute to control plans and other quality assurance procedures. It provides a knowledge base of failure mode and corrective action information that can be used as a resource in future trouble shooting efforts. Indeed, this technique provides a thorough knowledge of the functioning and interactions of a system, by the systematic analysis of cause-effect relationships. The information obtained is used as part of risk management, with primary concern obtaining a good level of dependability of operational system (Sahoo, Sarkar & Sarkar, 2014).

It allows to:

- Know the most important elements (functions and components)
- Find, evaluate and rank the weaknesses, faults and malfunctions of the system; -
- Manage the critical points and specify the corrective action

3.1.2 Knowing the goals of the project.

To know which the main objectives of the PdM program are, a key factor is evaluating what has led to this problem, that is, which facts gave rise to start to implement such program. Examples of it could be, the high percentage of failures in a machine, the high maintenance cost of the year, the high rate of fail parts that a component produces… Once the maintenance team have discovered what were the causes which produce the start of implementing a PdM program, a key factor is to stablish the main goals of such program.

This must be made in accordance with the interests of the entire organisation and managers of the different areas can be take part of it. When deciding the goals of the maintenance program, the first step which was analysing the critical equipment as well as the knowledge of previous events which led to it, are very useful tools which can help to perform it.

After knowing the goals of the project, it is needed something to measure it, or an estimation of a number which will be related to the success of the project.

An example may be to know which components are going to fail in the next N days. The success of the project could be to reduce part failures by X% (since knowing this information in advance we can change the part before it fails).

3.1.3 Determine the goals of the data analysis.

Here, the task is to translate business objectives into data analysis goals.

For example, in the case of wanting to know which components are going to fail in the next N days, a first goal could be to determine which characteristics of the sensors cause the component to fail, with the subsequent goal of determining the time it takes to fail the component once these conditions occur. Going a step further could establish the goal of knowing what actions and at what time they apply to reduce the times when the machine is stopped.

3.2 Collecting and Analysing the data

One of the most important steps when implementing the PdM Program is the acquisition of data. Depending on which PdM approach is going to be used a different data acquisition technique will be needed.

But before starting to explain the different techniques of collecting and analysing data it is crucial to point that this data need to comply some specific requirements.

3.2.1 Data Requirements.

Predictive models learn patterns from historical data and predict future results with some probability based on these observed patterns. The accuracy of the prediction of a model depends on the relevance, adequacy and quality of the learning and test data. The new data that is "punctuated" by this model must have the same characteristics and the same schema as the learning and test data. The characteristics of the function (type, density, distribution, etc.) of the new data must match those of the learning and test datasets. The objective of this section is in those data requirements.

3.2.1.1 Relevant data.

After being able to recognize which machines or components are the most critical in the production system and knowing which failure types or modes are the most common in them, it is the time of collect the data. But before starting to collect any type of data, ensuring which data is related to the failure mode of the machine it is very important.

The data must be relevant to the problem. For example, considering the case of use of machine breakdown, the learning data must contain characteristics of it. Relevant data about some parameters like the vibration, speed rotation, state of the oil… the learning data should describe all the different components of this system.

As mentioned before, FMEA methodology can be very helpful as for each component that have failed, such method will allow to know the failure mode and the causes and consequences of the fails. Knowing which the causes of the different fails are, led to know what relevant data needs to be collected and where the sensors must be put. In more advanced versions, the diagnostic and prognostic algorithms that could identify the failure modes would also be listed, some of them are explained in the chapter of ML approaches.

3.2.1.2 Enough data.

Examining the accessibility and quality of the data is an important part, i.e. determining whether the necessary data is available and in a proper condition to build a reliable model.

At the time of ensuring that there is enough data to implement the PdM program it is important to ensure that the data which must be collected is accessible, readable and the organization is legally allowed to collect it. Some data from components might not be available to the end-user and other times the user is not legally allowed to use the data, so obtaining access to the data in these cases is an important step.

After, obtaining the access to the data, two common issues appear related to the amount of data, the number of necessary events to train the model and the enough amount of records to collect, but there is no definitive answer, only general rules. However, it is true that the more error events collected, the better the model will be, but the exact number of error events varies depending on the data and the problem faced. On the other hand, if a machine presents errors too often, the company will replace it, which will reduce the instances of error.

As PdM programs relies on ML algorithms, enough data must be collected to create an accurate model. This data usually comes from machine sensors, so engineers should avoid conditions where their systems operate in modes where little data is collected until a failure occurs. To prevent this, organizations can change the data logging option to record more data.

Another option to increase the amount of data available is related to generate test data using simulation tools by creating models, that cover the mechanical, electrical or other physical systems to be monitored and then validating against measured data. To do this, there are some tools or programs as Simulink which creates models to generate fault and healthy data to develop condition monitoring algorithms. As this is a really extended field, this thesis not shows the steps to create such models, but the team of

MathWorks (MATLAB) in their article “Using Simulink to Generate Fault Data” show the steps to do it in detail. Next figure shows an example of a model created in Simulink:

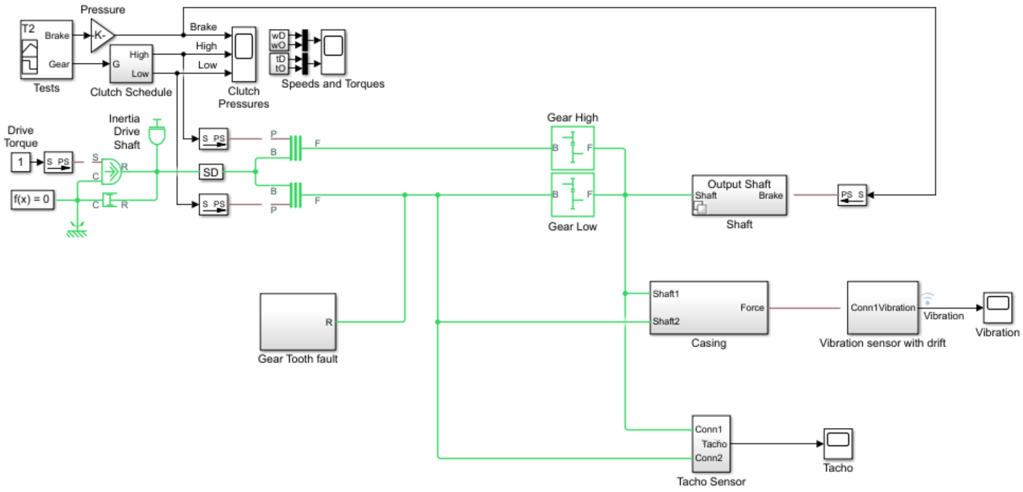


Figure 20. Using Simulink to model a transmission. This model can be used to synthesize fault data. © 1984–2018 The MathWorks, Inc.

However, even without the availability of such failure data, using Unsupervised Learning techniques normal and faulty behaviour can be identified. Techniques as *Principal Component Analysis* can be used to reduce the amount of data into a low-dimensional representation for visualization and analysis, where healthy equipment data may be centred around a normal operating point and unhealthy equipment may be moving away from normal conditions (Deland, 2018).

3.2.1.3 Data quality.

The quality of the data is fundamental, so that each prediction attribute value must be precise along with the value of the target variable. The quality of the data is surprisingly bad when they take advantage of systems that have not been designed with such intention. In the industrial sector this happens because organisations usually think that quality is implicit with the precision and calibration of the sensors which collect the data (Hannah & Starr, 2001, pp. 275–282). However, it is common to find inconsistencies in exploratory analysis like: negative times, erroneous numbers by several orders of magnitude, variables that should be related are not, the essential variables are not found among the dozens that have been stored, variables that do not appear at the frequency necessary, impossibility of relating variables between different databases, etc.

Some authors have defined a series of contextual and intrinsic dimensions which help to understand when it is considered that the data have quality (Hannah & Starr, 2001, pp. 275–282).

(Wang & Strong, 1996, pp. 5-33) in their research developed a framework that captured the aspects of data quality that were important to data consumers. Specifically, 118 data quality attributes collected from data consumers are consolidated into twenty dimensions, which in turn are grouped into four categories. They point that using this framework, information systems professionals will be able to better understand and meet their data consumers' data quality needs.

The resulting framework had four data quality categories which are; intrinsic, contextual, representational and accessibility, which can be viewed in next figure:

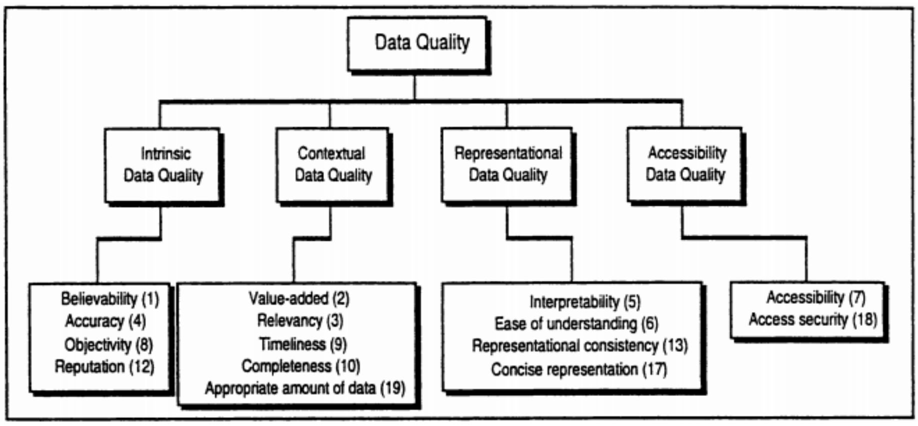


Figure 21. A Conceptual Framework of Data Quality (Wang & Strong, 1996)

- **Intrinsic Data Quality.**

Intrinsic category includes not only accuracy and objectivity, which were evident to IS professionals, but also believability and reputation. This suggests that, contrary to the traditional development view, data consumers also view believability and reputation as an integral part of the category; accuracy and objectivity alone are not enough for data to be considered of high quality (Wang & Strong, 1996, pp. 5-33).

- **Contextual Data Quality.**

Contextual category consists of value-added, relevancy, timeliness, completeness, and appropriate amount of data. This category refers to the information that is provided regarding subjective attributes of the context in which the data will be used.

- **Representational Data Quality.**

Representational DQ includes aspects related to the format of the data {concise and consistent representation) and meaning of data (interpretability and ease of understanding). These two aspects suggest that for data consumers to conclude that data are well represented, they must not only be concise and consistently represented, but also interpretable and easy to understand (Wang & Strong, 1996, pp. 5-33) One focus of current research in that area is context interchange among heterogeneous database system (Sciore, Siegel & Rosenthal, 1994, pp. 254-290).

- **Accessibility Data Quality.**

Accessibility category consists of accessibility and access security which emphasize the importance of the role of systems.

Using this framework, information systems professionals will be able to better understand and meet their data consumers' data quality needs. Moreover, in this framework the two requirements mentioned before (relevant and enough data) appeared but in the other part they are explained more in detail.

As the authors wrote in the research based on this framework, several research directions can be pursued.

First, a questionnaire could be developed by the organizations to measure perceived data quality and the data quality categories and their underlying dimensions in this framework would provide the constructs to be measured. Second, methods for improving the quality of data could be developed in concordance with this framework, which as well, could be used as a checklist during the analysis of data requirements.

If the existing historical data were not enough, or its quality was not that necessary to apply the chosen techniques, it would be necessary to capture new fault data. In this context, it would be necessary to design a test plan, execute it, and assess the statistical significance of the data collected. Approaches such as the design of experiments are very useful in this acquisition process.

3.2.2 Collecting the data.

3.2.2.1 Data Sources.

Once it is known which components are more likely to fail and which data need to be collected, maintenance team can start to collect the data if it is available. Relevant data sources for PdM include, but are not limited to:

- **Failure History.**

As mentioned before, failure events are rare in PdM applications as usually such events are avoided by preventive techniques. However, when compiling prediction models using ML techniques, the algorithm must know the normal operating pattern of a component, as well as its error patterns. Therefore, the learning data should include enough examples from both categories. Maintenance records and parts replacement history are good sources to look for error events. In case of not being able to collect data of from both categories, techniques such as simulation can be used as described before.

- **Maintenance history of repairs.**

The maintenance history of a resource includes details of the replaced components, repair activities, intervals of failures, repayment time etc. These events record the degradation patterns. If this fundamental information is not available in the learning data, deceptive results can be obtained for the model.

- **Operating conditions of the machine.**

A key assumption of PdM is that the maintenance status of a machine degrades over time during its routine operation. It is expected that the data contain characteristics that vary with time and that capture this expiration pattern, in addition to all the anomalies that lead to degradation. Thus, it is very important to have enough data from all the different features which describe the components in each one of the different operation conditions in which it has work. For example, the failure mode of a machine can be detected through the vibration and the temperature of it, then having data of such features in both health and failure states of the machine is key to the success of the program. Based on these data points, for example, the ML algorithm learns to predict how many more units of time a machine can continue to run before an error occurs which is described as the RUL of the component.

- **Static Characteristics data.**

Static characteristics are metadata about the equipment. Some examples are the brand of the equipment, the model, the date of manufacture, the date of start of the service, the location of the system and other technical specifications.

3.2.2.2 Types of data.

Data acquisition systems, as well as the signal processing needs to know the type of the data they are dealing with. In general, all collected data can be subdivided into two groups (Davies & Greenough, 2000):

- Events: The events include information about what has happened (failures, repairs, causes, actions of previous maintenance, etc.), what caused the event and what was done. Such group is more economical and accurate when using filtered or processed data. In return, there is a risk that the information that occurred between two successive inspections will be lost. In addition, the interval to which it applies must be chosen. There are multiple publications that study how to optimize the interval by minimizing the cost (for example, (Christer &. Wang 1995, pp. 258-269).
- Condition Monitoring: Measurements related to the health state of the machine, i.e. vibration data, temperature, pressure, oil debris analysis data, etc. Monitoring refers to physical measurements taken on the health status of a resource. Its drawbacks are the high cost and noise that sometimes occurs in the signals.

However, when it is talked about the main groups of data, data also can be classified by the type of component used to collect them. The two most important types of data for failure type detection and PdM are data gathered by sensors and data gathered by logfiles:

- **Sensor Data.**

Sensors have become smarter, smaller, easier to implement in existing systems, as well as cheaper and more reliable (Van Hoof, Baert & Witvrouw 2004, pp. 986-987). Sensors convert physical values into electrical values (voltage, current or resistance). Usually, one sensor measures one mechanical value (acceleration, pressure, flow, torque). With this mechanical value, one can interpret the vibration data, acoustic data, temperature, humidity, weather, altitude, etc. (Andrew, Daming & Banjevic, 2006, pp.1483–1510) point that this data falls into three categories:

- Value type data: Data collected at a specific time epoch for a condition, whereby monitoring variables are of a single value (i.e. temperature, pressure, oil, debris analysis data, etc)
- Waveform type data: Data in the form of time series, such as vibrations or sounds.
- Multidimension type: Multidimensional data collected at a specific time (i.e. different images like X-ray, thermographs, etc)

- **Logfiles.**

On the other hand, maintenance management systems (CMMS, ERP, etc.) or application log files can provide information about events that affect the system (Davies & Greenough, 2000). Events of a system can be recorded to a logfile. In this manner, the declaration of an event is very broad. A changing value can be such an event, as it executes a maintenance action. Logfiles can only contain fused sensor values. Descriptive log files can be written by humans describing maintenance actions, or any failures or errors detected (Jahnke, 2015).

3.2.2.3 Data Acquisition Techniques.

When implementing the PdM program, two main techniques exist for data acquisition, either push or pull based data techniques. A prominent research area heavily interested in push and pull is content distribution. (Deolasee, Katkar, Panchbudhe, Ramamritham & Shenoy, 2001) provide a good analysis of the roles of push and pull.

- **Pull- Based protocols.**

In pull protocols, the user connects to the server, makes a requirement of data or defines a query for the system, which will then return the matched value and finally, closes the connection and disconnects from the server. In this mode, the frequency the query must be executed can be determined in the query itself, or by submitting the query in the desired frequency (Deolasee et al. 2001). An example of it related to PdM, would be when the ML algorithm needs the values of a temperature of a machine of the last two weeks, here this would be the query and the system would give such amount of data.

Pull protocols are great when the device that one is trying to collect data from has a lot of different types of data, and one may be interested in a small subset of it.

- **Push- Based protocols.**

In push protocols, a former described state or a behavior will be communicated by system autonomously whenever values recently changed. The server will send (push) all new events or changes of the value to the user using that single always-on connection. An example of this related to PdM would be when the Predictive team must know when a temperature reaches a threshold, here the temperature would be the behavior and reaching or overtaking the threshold would be the change of value and the system would send an alert to the team.

Push is great when the device wants to report a small subset of data at high rates and in real time.

Next figure retrieved from (Jahnke, 2015) shows how these protocols work, here, the user of the system could be identified as a Data expert or even a ML platform which collects data for implement for developing the algorithm for PdM. The system could be viewed as the data collected by the sensors set in the machines. For pull protocol a query about any data is sent and the data system answer with such values. For push protocol, the user wants to know when a change occurs for a precise value and when it occurs the system send an alarm:

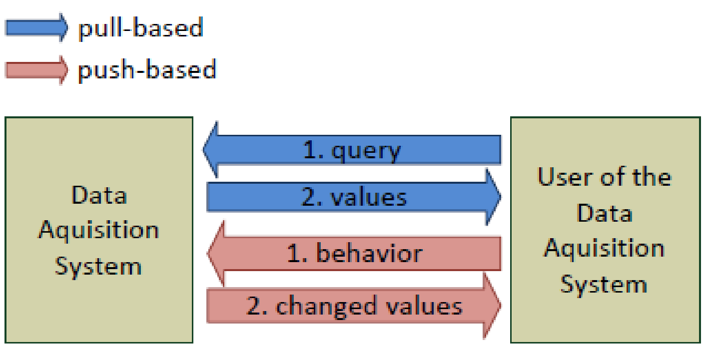


Figure 22. Push- and pull-based data acquisition techniques (Patrick Jahnke, 2015)

The most common pull and push methods for data acquisition of PdM which are related to the types of data (events and condition monitoring) discussed above, are the following (Jahnke P. 2015):

– Real time:

The real-time data acquisition technique is pull-based. Although, there is no system that can process data without any delay, in terms of failure type detection and PdM, real-time means that there is no perceivable delay. Such technique can be considered as the simplest and most generic data acquisition system. All available data are transferred in real time, where the subsequent processes selects the relevant data. Such a data acquisition technique increases the communication and computational costs of the failure type detection and PdM technique. The costs of energy efficiency and communication are high in real-time systems (Jahnke, 2015). Moreover, a real-time system may transfer a considerable amount of useless information to the system.

– Interval:

Sometimes either it is not possible or not necessary to use a real-time technique. For these cases, interval data acquisition systems can be used which sends the acquired data in time intervals. This system represents also a pull-based system. The time interval for transferring the data to the user can be of different lengths, especially when the transfer is limited by thresholds of values. Such method is used when communication costs are high.

This type of system can be simply designed with a fix interval, where the computational and the communication costs depend on the length of the interval. A long interval has the drawback that critical issues may be detected too late, whereas a short interval increases the computational and communication costs. A dynamic interval could solve the described problem as such, but will then increase the implementation costs, as well as requiring domain knowledge (Jahnke, 2015).

– Event driven:

The type of data acquisition technique is push-based. The data acquisition system publishes the value changes detected (events) and sends them to all users registered to this value or the properties of the value change. In terms of PdM, an event in this type can have many interpretations, as it can represent a value change in a critical way, or it can be a combination of value changes with or without any special order. An example here, could be the event of a change in temperature value, reaching a limit given. When it occurred, the system would send an alert informing of that. Usually, an event does not represent a state change of the real-world system, although it can lead to a state change. This is a specific method for a given PdM task.

A well-designed event-based system can reduce the volume of data, and thus the computational costs. But the effort for designing and implementing such a system is high. In addition, expert knowledge is required to determine when and what kind of event will occur. Moreover, sometimes, such method can increase the memory consumption and processing time of the system, because a complex event depending on several parameters may need to store values for a long time and may have many rules to check (Buchmann & Koldehofe, 2009, pp. 241–242).

3.3 **Data Processing.**

For most PdM applications, the values obtained from the data acquisition system must typically be pre-processed before transforming them into a new space of variables for better performance of the ML algorithm (Zhang, Xiong, Liu, Zou, & Guo. 2010, pp. 6077–6085). These pre-processed values from the data acquisition system are an example of a feature vector.

Data must be cleaned, eliminating errors that have occurred in the shot. Errors in the data may be due to human failures (especially in the introduction of events), or to sensor failures (Xu & Kwan. 2003, pp. 12-23). Cleaning is a complicated task, which often requires manual procedures to examine the data.

Some of the methods used for data processing include (Bengtsson, Olsson, Funk & Jackson, 2004)

- Representative sampling of a large data population.
- Transformation to manipulate raw data and produce a single entry (minable table).
- Eliminate noisy data.
- Normalize, organizing the data for more efficient access.
- Extraction of characteristics to identify the most significant data.

3.3.1 Signal Processing.

Signal processing uses mathematical, statistical, computational, heuristic and/or linguistic representations, formalisms, modelling techniques and algorithms for generating, transforming, transmitting, and learning from analog or digital signals. It consists of the theory, algorithms, architecture, implementation, and applications related to processing information, contained in a variety of different formats, broadly designated as signals. Signal processing may involve filtering, recovery, enhancement, translation, detection, and decomposition (Vaseghi, 2008). This thesis will focus on waveform signal processing because many ML approaches discussed in the last chapter are based upon waveform data.

In PdM, sensor signals which have time and frequency components (i.e. vibration signals, acoustic signals), are the most common types of signals for detecting the state of the system. Thus, signal processing techniques can be described into three types: time domain, frequency domain and time-frequency domain.

Although time-frequency analysis is the most common technique, for a better understanding, the time-domain and frequency-domain analysis will be introduced.

3.3.1.1 Time domain analysis.

This type of analysis consisted on the analysis of a waveform, which is a chronological sequence of the value of a random variable. This random variable has expected value, variance skewness and kurtosis whose values are the results of such analysis.

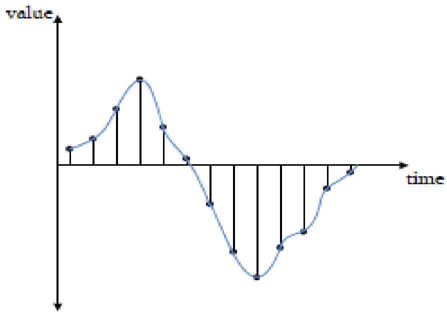


Figure 23. An example of a waveform (blue line) for a given random variable (Patrick Jahnke, 2015)

The main idea consist on fitting the waveform data to a parametric time series model thus, the (ARMA: autoregressive moving average) model is a common time series model, which specifies that the output variable depends linearly on its previous values (Said & David, 1984). On the other hand, the fractal time series model, detects the dependency of two waveforms, for long-range dependence or short-range dependence, and global or local self-similarity (Li, 2009).

Vibration signals are very common among signals in PdM approaches. As with all sensor signals, noise is a component of vibration signals which can distort the result. To denoise a given signal, it need to be filtered. The (TSA: time synchronous average) technique can be used, as it tries to denoise a vibration signal by analysing examples from the real-world system. However, this technique leads to a large computational cost when filtering the signal in the time domain. Such filtering operations are simple multiplications in the frequency domain which resulted in a lower computational effort.

3.3.1.2 Frequency domain analysis.

A signal in the time-domain can be transformed into the frequency domain and back again. Such a transformation has two reasons. It is important to know the distribution of the frequency share, and the signal needs to be filtered due to noise components in sensor signals (Jahnke, 2015). Filtering is usually performed in the frequency domain, because filtering causes convolution in the time domain and multiplication in the frequency domain. Therefore, it is considerably more efficient to transfer a signal from the time domain to the frequency domain, perform the filtering (multiplication), and transform the filtered signal back to the time domain (Jahnke, 2015). However, signals are most available in the time domain and thus, there are some techniques which perform this transformation. Between the most important are Fourier Transformation, Laplace Transformation and Z Transformation (Föllinger & Kluwe, 1977).

- **From time to frequency- Fast Fourier Transformation**

Fast Fourier Transformation (FFT) is an algorithm to efficiently compute a discrete Fourier Transformation (Duhamel & Vetterli, 1990, pp. 259-299). There are some different algorithms for the Fast Fourier Transformation, and all of them work with interim results and reduce the number of arithmetic operations. The Fast Fourier Transformation of a discrete, finite signal is explained in detail in (Welch, 1967, pp. 70-73). Such technique is applied to all values of the signal. Therefore, the frequencies with a very small percentage of the overall signal can be excluded. In PdM approaches, an upcoming failure is detected with a changing frequency, but if the percentage of the new frequency does not significantly increase in relation to the overall signal, this could be also interpreted as noise. This is main drawback of this method in PdM (Jahnke, 2015).

Nowadays both time and frequency are relent in PdM approaches. Accordingly, the frequency domain analysis is almost irrelevant. Next figure shows a transformation of this type.

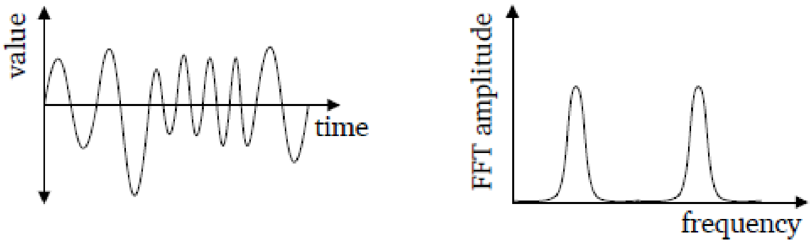


Figure 24. An example of a signal in time domain (left) and frequency domain after a Fast Fourier Transformation (right) (Jahnke, 2015)

3.3.1.3 Time-frequency domain.

PdM approach is based on analyzing a given signal and trying to determine the state of the system analyzed. The detection of significant signal changes is a time-based problem. However, in many cases, the relevant information is hidden in the frequency content of signals, so that processing techniques in both time and frequency domain is widely used. (Jahnke, 2015) described the most common time-frequency domain techniques used in PdM, who also displayed a table with the most used signal-processing techniques in recent literature, which resulted on being the Hilbert Huang and Wavelet transformation. But as the Wavelet transformation has the drawback that leakage effects can occur at the beginning or end of the wavelet (Kiencke, Schwarz & Weickert, 2008), in this thesis it is considered more relevant the Hilbert Huang transformation, which is the most recent technique as well as, have the ability of analysing non-linear and non-stationary signals which are very common in this type of problems, therefore, is explained more in detail than the others.

- **Short-time Fourier transformation.**

This technique, in contrast with the Fast Fourier Transformation analyses a signal for a defined time window only. The transformation evaluates the sinusoidal frequency and phase content of local sections of a signal as it changes over time. The resolution of this method depends on the width of the window between the time and frequency domain. Having a wide window implies a bad resolution, and a narrow window a good. In the next figure, the left figure has a good resolution in time domain (narrow) and bas in the frequency domain (wide).

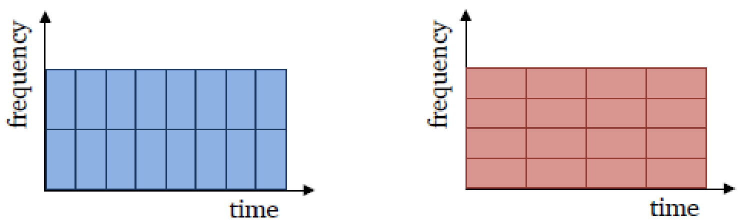


Figure 25. Example of Short Time Fourier Transformation (Jahnke, 2015)

Usually, the complexity of the such technique is *Θ (n log n)*, but there exists some modification of the algorithm that can achieve the complexity of *Θ (n)* (Covell & Richardson, 1991, pp. 2041–2044).

- **Wavelet transformation.**

As explained, transforming signals into the time-frequency domain has its weakness in the resolution, which is always a trade-off between good time or good frequency resolution. With this method, it is possible to get a poor frequency and good time resolution at high frequencies and good frequency and poor time resolution at low frequencies since it has a shifted and scaled version of the original signal. Such temporal resolution is key using this transformation, which can be divided into two categories, continuous or discrete wavelet transformation (Jahnke, 2015).

The continuous wavelet transformation can change the scale continuously in the window at every iteration, which means that both frequency and time resolution can be changed at every iteration to fit the signal as well as possible. However, at the time of computing it is more efficient to discretize such transformation, which is called discrete wavelet transformation.

Discrete wavelet transformation usually uses a dynamic method to do the transformation. (Jahnke, 2015) explains more in detail such transformation technique and points that, in PdM, a signal mostly needs a high time domain resolution and low frequency resolution for high frequencies, and a low time domain resolution and high frequency resolution for low frequencies of the given signals, because failures are often represented as a combination of short recurring frequency components in a signal. To ascertain such properties, it is useful to create a wavelet packet tree. Some combinations of wavelets with different resolutions can be significant for a failure type detection and PdM.

- **Hilbert Huang transformation.**

This transformation is a recent approach in signal processing and was proposed by (Huang et al 1998). The approach analyses non-linear and non-stationary signals which are very common in PdM problems. Therefore, an implementation into such approach is very useful. It has been validated empirically and has proven to be very effective for a wide variety of different disciplines as (Huang & Attoh-Okine, 2005) (Huang & Shen, 2005) (Huang N.E. et al. 2003, pp. 245-268). The method to obtain such transforming is composed in two parts: the decomposition in empirical modes of the signal (Empirical Mode Decomposition EMD) and the analysis of each one of the modes through the Hilbert transformation (Hilbert Spectral Analysis HAS)

1. Empirical Mode Decomposition (EMD).

Empirical Modal Decomposition assumes that any data set can be divided into a set of intrinsic oscillation modes. Each mode of vibration, linear or not, represents a simple oscillation that has the same number of relative ends and cut points with the OX axis. In addition, the oscillation will be symmetric with respect to its local average. That is, in a moment of time the data has many oscillation modes, whose sum is the state of total vibration. Each of the vibration modes is called Intrinsic Modal Function (IMF) and can be defined as follows (Huang, 2005, pp. 1-26):

- The number of extreme (minima and maxima) and the number of zero-crossing must be equal to or differ by one in the whole dataset.
- At any point, the mean value between the local minima and the local maxima is zero.

To obtain the modes, firstly the local maxima must be obtained and using the cubic spline interpolation, its envelope is got. With the minimums, the procedure is the same and after obtaining both envelopes, the average between them is obtained. Thus, starting from a certain signal, its envelopes and the average of them are obtained as shown in next figure:

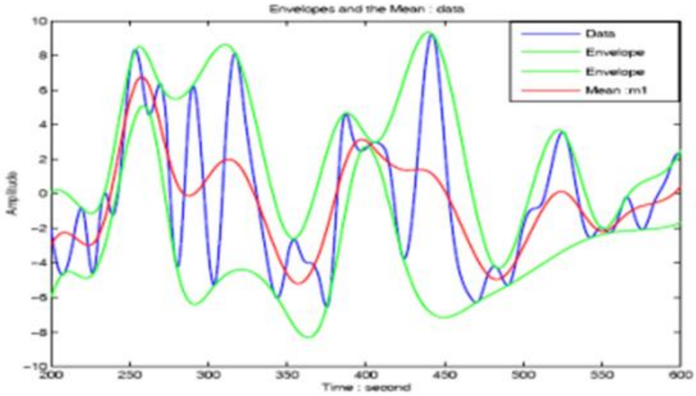


Figure 26. Shifting Process (Somonte, 2010)

Subtract that mean from the total data set by obtaining h_1 (t), which could be the first signal mode:

$$h_1\ (t) = x\ (t)- m_1\ (t)$$

To know if h_1 (t) is the first empirical mode of the signal x (t) it is necessary that it fulfils the properties previously stated. If it does not comply, it is necessary to continue carrying out the process with h_1 (t) of obtaining envelopes and averages, which is called sifting process, until obtaining the first desired empirical mode. So, the next step would be to consider that h_1 (t) is the starting signal and repeat the sifting process:

$$h_{11}\ (t) = h_1\ (t)- m_{11}\ (t)$$

Being m_{11} the average of the envelopes of h_1 (t) and being h_{11} (t) the possible new vibration mode. After repeating this process, the stop will be when the number of ceros and extremes differ in no more than consecutively during several fixed times. If were necessary k times to achieve the first mode, the first mode h_{1k} (t) is reached as:

$$h_{1k}\ (t)= h1_{(k-1)}\ (t)- m_{1k}(t)$$

And it is denoted as:

$$c_1\ (t)= h_{1k}\ (t)$$

As the signal yet has information, it is necessary to continue decomposing it in its empiricism modes, thus, the named residual r_1 (t) is obtained as:

$$r_1\ (t)= x\ (t)- c_1(t)$$

Such residual is submitted to the sifting process mentioned before. Next figure shows the starting signal and the residual after calculating the first empirical mode for the example:

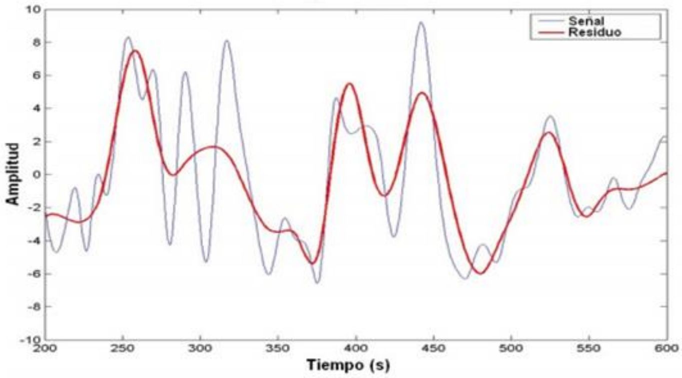


Figure 27. Starting signal and the residual obtained after calculating the first empirical mode (Montalvo, 2010)

The decomposition will continue until one of the following two criteria is obtained, either the value of the residual is very small or that it is a function without relative extremes from which no more empirical mode can be extracted.

Obviously, the initial signal can be recomposed by simply adding up the obtained empirical mode and the final residues:

$$x\ (t)= \sum_{j=1}^{n} cj\ (t) + r_n(t)$$

Where n is the total number of obtained modes. Next figure shows an example of decomposition modal empirical signal and the residual:

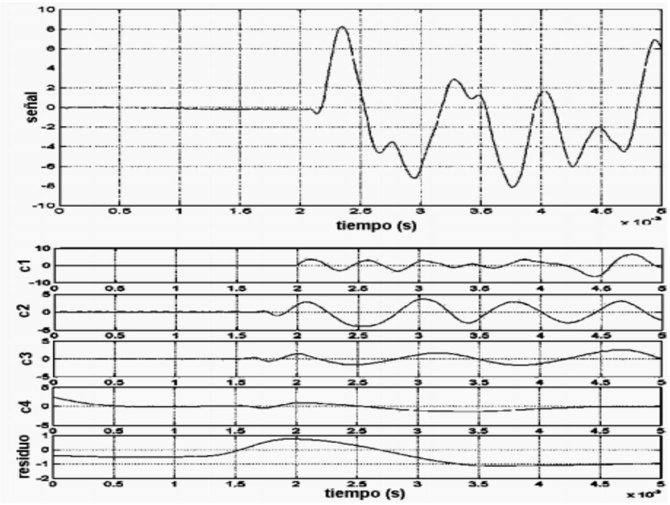


Figure 28. Example of Modal Empirical Decomposition (Montalvo, 2010)

Once the Empirical Modal Decomposition is done, the Hilbert Transform can be applied to each one of the obtained (IMF)

2. Hilbert Spectral Analysis.

The reason for doing the decomposition is to be able to find the instantaneous frequencies of the signal, before Huang introduced DME, the use of the Hilbert Transform was limited to signals with the same number of extremes and zeros (Huang, 2005, pp. 1-26).

The Hilbert transform is defined as:

$$y(t) = H[x(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x(u)}{t-u} du$$

The analytic signal is defined as:

$$z(t) = x(t) + jy(t) = a(t)e^{j\Theta(t)}$$

Where:

$$a(t) = \sqrt{x^2(t) + y^2(t)}$$

$$\Theta(t) = arctg\frac{\mathrm{y}(t)}{\mathrm{x}(t)}$$

Finally, the instant frequency can be found using the next equation:

$$\omega(t) = \frac{d\Theta}{dt}$$

With these data a large amount of analysis can be done, all the tests indicate that the Hilbert-Huang Transform is a very powerful tool for the time-frequency analysis of non-linear and non-stationary signals. It is based on an adaptive base and the frequency is defined through the Hilbert Transform (Somonte, 2010).

As discussed, the Hilbert Huang transform distributes the original signal in components called intrinsic mode functions, which are orthogonal. This implies that an intrinsic mode function represents a physical meaning, and thus practical for PdM (Wu, Chen, Xiaoxue & Wang 2012, pp. 103–122), but one of its drawbacks is that it is also expensive, and it is very special for a real-world system to define an HHT which interprets the physical meaning.

- **Wigner Ville distribution.**

This is one of the most studied methods in this type of analysis. In contrast to other methods, this technique has no function window which means that it has no leak and the best spectral resolution of all-time frequency methods (Jahnke, 2015). The Wigner Ville distribution is a quadratic integral transformation, meaning that it is a two-dimensional Fourier transformation of the ambiguity function in relation to time and frequency.

Two signals that are compared by the distribution give rise to cross terms. These cross terms are generally, not preferable, but they could be interesting in terms of failure type detection and PdM. The cross terms can also be helpful because they have significant information about the two signals and thus, it could also be helpful for PdM. For example, if the vibration signals are different, the cross-term interferences are also different, which will also be helpful to distinguish both signals (Wang, Zhang, & Zhong 2008, pp. 1981–1993).

Signal processing is an important step in PdM applications for real-world systems monitored by sensor data. For sensor data it is important to know which information is relevant. If only change of the sensor value over time is involved, the time domain analysis can be used to extract more information as a new value but, if only the change in the frequency of sensor value is relevant, the frequency domain analysis must be used. As mentioned before, the most discussed signal processing technique is the time frequency analysis. Here, the value change over time and the frequency of value changes are relevant. For more information about these techniques, in his thesis (Jahnke P. 2015), explained them more in detail. Moreover, this author, as mentioned before, shows in a table which are the most common used method nowadays in the recent literature for PdM, being the Wavelet transformation and Hilbert-Huang transformation the most preferred by the authors.

3.3.2 Data Cleaning.

The data should have quality, being sufficiently representative and significant so that later, useful analyses can be carried out (Koksal, Batmaz & Testik. 2011, pp. 13448-13467). If this were not the case, it would be necessary to complete the existing data with new captures using a statistical experiment design. Data from real word systems used to present some problems that can be solved by cleaning techniques.

3.3.2.1 Incomplete data and missed values.

Many times, sensors do not work correctly, and a large amount of data is not collected, so the models learned lose accuracy. Although some techniques can treat this problem, it is advisable to offer solutions before beginning the training of the models (Khoshgoftaar & Hulse, 2009, pp.1513–1542). Incomplete data leads to less robust and less accurate models. In these cases, their values are usually interpolated, or the samples eliminated as described in (Khoshgoftaar & Hulse, 2008, pp.563–600).

3.3.2.2 Noisy data.

Noisy data is known to affect the way any data mining system behaves (Brodley & Friedl 1999, pp.131–167). Focusing on the scenario of imbalanced data, the presence of noise has a greater impact on the minority classes than on usual cases, since the positive class has fewer examples to begin with, it will take fewer "noisy" examples to impact the learned sub concept. Irrelevant, outliers or meaningless data hinder training or even cause overtraining. These samples could be detected by clustering or measures of similarity (Khoshgoftaar & Van Hulse, 2011, pp. 552–68).

There are a lot of techniques to solve such problem, for example, (Batuwita & Palade 2010, pp.558-571) developed the FSVM-CIL algorithm, a synergy between Support Vector Machines and Fuzzy Logic aimed to reflect the within-class importance of different training examples to suppress the effect of outliers and noise. (Khoshgoftaar & Hulse, 2011, pp. 552–68), the authors presented a similar study on the significance of noise and imbalance data using bagging and boosting techniques. Their results show the goodness of the bagging approach without replacement, and they recommend the use of noise reduction techniques prior to the application of boosting procedures. For more information about these techniques (López, Fernandez, Garcia, Palade & Herrera. 2013, pp. 113-141) explain these more in detail.

3.3.2.3 Unbalanced data.

In classification problems, it occurs when one of the classes has many more samples than another. This can be alleviated by applying sampling techniques which are very common in this type of problems. When applying PdM solutions this problem often occurs but can be solved with the following techniques.

- **Sampling methods.**

Imbalanced learning involves the use of sampling methods to modify the training data set to a balanced data set. Sampling methods are not applied to the test set. Although there are several sampling techniques, most used ones are Random Oversampling and Under sampling.

- Random Oversampling: Involves selecting a random sample from minority class, replicating these examples, and adding them to training data set. Consequently, the number of examples in minority class is increased, and eventually balance the number of examples of different classes. A drawback of oversampling is that multiple instances of certain examples can cause the classifier to become too specific, leading to over-fitting. The model may show high training accuracy, but its performance on unseen test data may be suboptimal. To deal with the mentioned problems, more sophisticated methods have been proposed (MarkTab, 2018). The "Synthetic Minority Oversampling Technique" (SMOTE) (Chawla, Bowyer, Hall & Kegelmeyer 2002, pp.321–357) has become one of the most renowned approaches in this area. In brief, its main idea is to create new minority class examples by interpolating several minority class instances that lie together for oversampling the training set.

- Random Under Sampling: Involves selecting a random sample from a majority class and removing those examples from training data set. However, removing examples from majority class may cause the classifier to miss important concepts pertaining to the majority class. Other examples here are, the Wilson's Edited Nearest Neighbour (ENN) (Wilson, 1972, pp. 408–421) rule, which removes examples that differ from two of its three nearest neighbours or the One-Sided Selection (OSS) (Kubat & Matwin, 1997, pp. 179–186.), an integration method between the Condensed Nearest Neighbour Rule (Hart, 1968, pp. 515–516) (Tomek, 1976, pp. 769–772). and the Neighbourhood Cleaning Rule (Laurikkala, 2001, pp. 63–66.), which is based on the ENN technique.

Hybrid sampling where minority class is over-sampled and majority class is under-sampled at the same time is another viable approach. There are many sophisticated sampling techniques. The technique chosen depends on the data properties and results of iterative experiments by the data scientist.

- **Cost sensitive learning.**

In PdM, failures that constitute the minority class are of more interest than normal examples. So, the focus is mainly on the algorithm's performance on failures. Incorrectly predicting a positive class as a negative class can cost more than vice-versa. This situation is commonly referred as unequal loss or asymmetric cost of mis-classifying elements to different classes. The ideal classifier should deliver high prediction accuracy over the minority class, without compromising the accuracy for the majority class (MarkTab, 2018).

There are multiple ways to achieve this balance. To mitigate the problem of unequal loss, assigning a high cost to mis-classification of the minority class, and trying to minimize the overall cost can help. Algorithms like SVMs (Support Vector Machines) adopt this method inherently, by allowing cost of positive and negative examples to be specified during training. Similarly, boosting methods such as Boosted Decision Trees usually show good performance with imbalanced data.

3.3.3 Data Fusion.

In a complex system, a single sensor is unable to collect enough information to accurately apply diagnostic or prognostic techniques. When multiple sensors are installed, the measurements collected of each sensor contain partial information that must be combined. Three approximations are usually followed to group the information (data fusion): at the level of data, of characteristics or of decisions (Hall & Llinas,2001) (Hall & McMullen, 2004). Several articles study them in diagnostic or prognostic applications (Hannah, Starr, Bryanston, 2001, pp. 275–282) (Liu & Wang, 2001, pp. 203-210).

When merging data from several sources, it is common that in each of the sources data had been sampled at a different frequency, thus a phase of temporal alignment must be carried out so that, some sources can be related with others (Winters & Silipo, 2015).

Another related problem is to select only the sources of data relevant to the diagnosis or forecast, known as Blind Separation of Sources (BSS) (Haykin, 2000). The objective is to find the separation function that can be applied to the observed signals to obtain an estimate of the original signals and the generated noise. For this, it is possible to apply techniques like Principal Component Analysis (PCA) or second-order statistical techniques.

3.4 Feature Engineering.

After the data acquisition and signal processing, all features of the system are available, but, that feature set can be enormous. Sometimes to increase the accuracy and to reduce the computational effort of the ML approach in PdM, it is necessary a reduction in the number of features and an increase of the information content in these features (avoiding the so-called curse of dimensionality) and reduce the computational cost of later modelling (having to handle a smaller amount of data) (Koksal, Batmaz & Testik. 2011 pp. 13448- 13467).

Data sets from real applications often contain lots of features, many of which are redundant or irrelevant to certain tasks (Fayyad, Shapiro & Smyth 1996, pp. 27-34).

This happens especially when it is unknown what features are relevant to an application or when knowledge of the domain is incomplete. The presence of irrelevant or redundant features may mask the distribution of the relevant ones, degrading the quality of the classification models. In addition, as the dimensionality increases, it will increase the noise and complexity of the interactions between the features, hindering the effective and efficient application of the subsequent analysis (Piramuthu, Ragavann & Shaw 1998, pp. 416-430). On the other hand, there is usually interaction between the features, since they are not isolated, so they are combined in some way with others (Markovitch and Rosenstein, 2002, pp. 59-98). To solve all these problems, various techniques have been proposed for the selection, extraction and construction of characteristics (Liu and Motoda, 1998) (Zhao, Atish, Sinha & Wei Ge, 2009, pp. 2633-2644)

In general, the techniques to reduce the number of dimensions are classified into two groups: selection of the most important variables ignoring the rest; or projection of the original set of variables in a space of smaller dimensions (also known as extraction of characteristics).

The selection of variables chooses a subset of the most representative features, while the extraction transforms the original space generating new information when combining features. Even if a significant reduction of the features space is achieved, the extraction can be a non-linear process, so the results are not clearly interpretable.

For its part, the construction of characteristics deals with the problem of interaction by proposing advantageous combinations of the original characteristics.

Often, it is opted for an automation of these processes to look for the best representation of the input characteristics (Murthy, 1998, pp. 345-389).

3.4.1 Feature Selection.

The objective is to keep a subset of the initial variables, eliminating those that do not provide information or that even degrade the models, maximizing their accuracy and simplifying their calculation (Gang. 2017). Feature Selection methods in ML are classified into three main types.

3.4.1.1 Filter methods.

Filter methods are used to select a subset of relevant features independent of any model or ML algorithm. Many of the filter methods are univariate and provide statistical test scores for each feature-outcome combination. Examples in this category include:

- Pearson's Correlation: It is used as a measure for quantifying linear dependence between two continuous variables X and Y. Its value varies from -1 to +1. Pearson's correlation is given as:

$$\rho_{XY} = \frac{cov(X,Y}{\sigma_X \sigma_Y}$$

- Linear discriminant analysis: It is used to find linear combination of features that characterizes or separates two or more classes (or levels) of a categorical variable.
- ANOVA: Analysis of variance. It is like Linear Discriminant Analysis except for the fact that it is operated using one or more categorical independent features and one continuous dependent feature. It provides a statistical test of whether the means of several groups are equal or not.
- Chi-Square: It is a is a statistical test applied to the groups of categorical features to evaluate the likelihood of correlation or association between them using their frequency distribution.
- mRMR: Maximal-relevance and minimal-redundancy identifies features which are most relevant to the outcome but are not highly correlated among themselves (Ding & Peng 2005, pp. 185–205). The goal of this algorithm is to find a subset of features which better describe the target class; in other terms, the subset of features the target class is most statistically dependent on. The complexity of this task increases with the dimensionality, particularly if the number of samples available is insufficient. As such, the maximum dependency calculation can be enhanced with two different criteria: maximum relevance and minimum redundancy.

3.4.1.2 Wrapper methods.

Wrapper methods try to search for the best feature combination by training a predictive model repeatedly for various feature subsets and keep aside the best or worst performing subsets. Therefore, wrapper methods provide the best performing feature combination on that predictive model. Some common examples of wrapper methods are:

- Forward Feature Selection: In the first phase of the step forward feature selection, the performance of the classifier is evaluated with respect to each feature. The feature that performs the best is selected out of all the features. In the second step, the first feature is tried in combination with all the other features (Saurav Kaushik, 2016). The combination of two features that yield the best algorithm performance is selected. In each iteration, it keeps adding the feature which best improves our model till an addition of a new variable does not improve the performance of the model.

– Backward Elimination: it starts with all the features and removes the least significant feature at each iteration which improves the performance of the model. The process is repeated until no improvement is observed on removal of features (Kaushik, 2016).
– Recursive Feature elimination: It is a greedy optimization algorithm which aims to find the best performing feature subset. It repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It constructs the next model with the left features until all the features are exhausted. It then ranks the features based on the order of their elimination (Kaushik, 2016).

The drawback is that Wrapper methods can be computationally expensive on a large dataset.

3.4.1.3 Embedded methods.

Embedded methods are in between filter and wrapper methods in terms of computational complexity. Embedded methods combine the qualities of filter and wrapper methods. These are the algorithms with built-in feature selection methods, i.e., they perform feature selection as a step toward predictive model building. Some of the most popular examples of these methods are LASSO and RIDGE regression which have inbuilt penalization functions to reduce overfitting.

– LASSO: "Least absolute shrinkage and selection operator" is a popular embedded Feature Selection method due to its simplicity. It is essentially a linear regression method with an L1-penalty (regularization) which shrinks many of the coefficients to zero. The features with non-zero coefficients in LASSO are considered relevant variables and it forces weak features to have zero as coefficients (Dorugade & Kashid., 2010, pp. 447 – 456). However, when the features are correlated, LASSO tends to randomly pick only one feature.
– Ridge Regression: LASSO does not differentiate "important" from "less-important" predictors in a model, so it includes almost all of them. This leads to overfitting a model and failure to find unique solutions. Ridge regression avoids all these problems. It works in part because it does not require unbiased estimators. Ridge regression adds just enough bias to make the estimates reasonably reliable approximations to true population values. Moreover, it uses a type of shrinkage estimator called a ridge estimator which produces new estimators that are shrunk closer to the "true" population parameters. The ridge estimator is especially good at improving the least-squares estimate when multicollinearity is present. Such method uses L2 regularization which avoid the problem of the L1 regularization, which limits the size of the coefficients by adding an L1 penalty equal to the absolute value of the magnitude of coefficients that can result in the elimination of some coefficients altogether, which can yield sparse models. L2 regularization adds a L2 penalty, which equals the square of the magnitude of coefficients. All coefficients are shrunk by the same factor (so none are eliminated). Unlike L1 regularization, L2 will not result in sparse models (Dorugade & Kashid., 2010, pp. 447 – 456).

3.4.2 Feature Extraction.

Feature extraction creates new variables as combinations of others to reduce the dimensionality of the selected features. Feature extraction methods can be divided into linear or non-linear, depending on the nature of the mapping function (Gang. 2017). Another option is to divide them into supervised or unsupervised, depending on whether they take classification information.

3.4.2.1 Linear methods.

Linear feature extraction assumes that the data lies on a lower-dimensional linear subspace. Some of the most known linear methods are:

– Principal Component Analyses: PCA is the most well-known dimensionality reduction algorithm. Using the covariance matrix and its eigenvalues and eigenvectors, PCA finds the "principal components" in the data which are uncorrelated eigenvectors each representing some proportion of variance in the data. It has been argued (Bair, Hastie, Paul & Tibshirani, 2006, pp. 119–137) (Bair & Tibshirani, 2004, pp. 511–522) that when computing the PCs of a dataset there is no guarantee that the principal components will be related to the class variable. Therefore, Supervised Principal Component Analysis (SPCA) was proposed, which selects the PCs based on the class variables. Even though the supervised version of PCA performs better than the unsupervised, PCA has an important limitation: it cannot capture nonlinear relationships that often exist in data, especially in complex systems.

3.4.2.2 Non-linear methods.

Nonlinear dimensionality reduction works in different ways. For example, a low-dimensional surface can be mapped on a high-dimensional space so that a nonlinear relationship among the features can be found. In theory, a lifting function can be used to map the features onto a higher-dimensional space. On a higher space the relationship among the features can be viewed as linear and therefore is easily detected. This is then mapped back on the lower-dimensional space and the relationship can be viewed as nonlinear (Zena & Duncan, 2015)

– Using Manifolds: It assumes that the data (genes of interest) lie on an embedded nonlinear manifold which has lower dimension than the raw data space and lies within it. Several algorithms exist working in the manifold space and applied to microarrays. A commonly used method of finding an appropriate manifold, Isomap (Tenenbaum, Silva & Langford, 2000 pp. 2319–2323), constructs the manifold by joining each point only to its nearest neighbours. Distances between points are then taken as geodesic distances on the resulting graph. Many variants of Isomap have been used; for example, (Balasubramanian & Schwartz, 2002, p. 7). proposed a tree connected version which differs in the way the neighbourhood graph is constructed. The k-nearest points are found by constructing a minimum spanning tree using an -radius hypersphere. This method aims to overcome the drawbacks expressed by (Orsenigo & Vercellis, 2012, pp. 9-16) regarding the robustness of the Isomap algorithm when it comes to noise and outliers. These could cause potential problems with the neighbouring graph, especially when the graph is not fully connected. Isomap has been applied on microarray data with some very good results (Orsenigo & Vercellis, 2012, pp. 9-16). Compared to PCA, Isomap was able to extract more structural information about the data.
– Neural methods: Thy also be used for dimensionality reduction like Self Organizing Maps (Kohonen, 1988, pp. 509–521) (SOMs) or Kohonen maps that create a lower-dimensional mapping of an input by preserving its topological characteristics. They are composed of nodes or neurons and each node is associated with its own weight vector. Autoencoders are feed-forward NNs which are trained to approximate a function by which data can be classified. For every training input the difference between the input and the output is measured (using square

error) and it is back-propagated through the NN to perform the weight updates to the different layers (Zena & Duncan, 2015).

- Kernel PCA: It has been widely used since dimensionality reduction helps with the interpretability of the results. It does have an important limitation in terms of space complexity since it stores all the dot products of the training set and therefore the size of the matrix increases quadratically with the number of data points (Liu & Yang, 2009).
- Independent component analysis: ICA finds the correlation among the data and decorrelates the data by maximizing or minimizing the contrast information. This is called "whitening." The whitened matrix is then rotated to minimise the Gaussianity of the projection and in effect retrieve statistically independent data. It can be applied in combination with PCA. It is said that ICA works better if the data has been pre-processed with PCA (Cao, Chua, Chong, Lee & Gu. 2003, pp. 321–336). This could merely be due to the decrease in computational load caused by the high dimension.

3.4.3 Feature Construction.

The construction consists of inventing new characteristics that express better the behaviour of the system than the original characteristics. In general, they try to take advantage of the interaction between them.

The problem is that the number of possible interactions between all the characteristics increases exponentially with respect to the original number. In addition, the potential search space is almost infinite, since new features can be built in multiple ways. Some authors opt for voracious heuristics to construct and traverse characteristic spaces (for example, with evolutionary strategies using Genetic Algorithms (Krawiec, 2002, pp. 329-343)). In general, it is observed that non-exhaustive heuristics work better and guarantee optimal solutions (Piramuthu, Ragavan & Shaw 1998, pp. 416-430). Another important factor that can facilitate construction is the use of domain knowledge and application (Feelders, Daniels & Holsheimer, 2000, pp. 271-281).

After feature engineering, the learning process starts. The next section discusses opportunities of labelling historic data. This process is dependent on the modelling technique. In turn, the modelling technique depends on the business problem and nature of the available data. Data labelling, like in feature engineering, also have influence on the accuracy and computational costs.

3.5 **Modelling in Predictive Maintenance**

Predictive analytics is defined as the application of mathematical models to estimate business data that are unknown or uncertain. A mathematical model in the field of predictive analytics is a set of transformations or functions applied to input data and transformed them into a value of the response variable.

Depending on the type of business objective, one type of technique or another will be applied. (Mark Tab, 2018) on his article describes the main modelling techniques for PdM problems, along with their methods of constructing specific labels.

3.5.1 Binary Classification models.

These techniques are used to predict the probability that a component presents an error in a future period, which is known as the future horizon period X. The X depends on the business problem and the data in question, according with the domain expert.

These models look for patterns that have given rise to failures in the past (historical data measured in the sensors along with a fault history) to be able to predict future errors in the machines. In this technique, two types of learning examples are identified. A positive example, indicating an error, with the label = 1 and a negative, indicating normal operation, label = 0. The destination variable, and therefore the label values, are categories.

3.5.1.1 Data labelling for binary classification.

The aim here is to find the probability that a resource can present a failure in the next X units of time. The label X records before the error of a resource as "near the error" (label = 1) and label all other records as "normal" (label = 0).

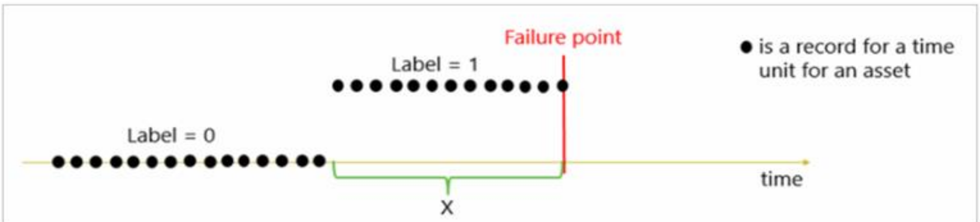


Figure 29. Labelling for binary classification (Mark Tab, 2018)

3.5.2 Regression models.

The Residual Useful Lifetime (RUL) prediction is a regression problem only relevant in PdM scenarios as it returns one continuous value, which represents the rest of useful lifetime and not the type of failure. The remaining lifetime is defined as the amount of time a resource is operational before the next error occurs. Each learning example is a record that belongs to a unit of time “nY” of a resource, where n is the multiple. The model must calculate the RUL of each new example as a continuous number. This number indicates the period that remains before the error (Mark Tab, 2018).

3.5.2.1 Data labelling for regression models.

The aim here is to find the remaining time until a component can present a failure. For each record before the error, calculate the label to be the number of time units remaining before the next error. In this method, labels are continuous variables. Labelling is done by reference to an error point. It cannot be calculated without knowing how long the resource has survived before an error. Therefore, unlike the binary classification, resources without errors in the data cannot be used for modelling (Mark Tab, 2018).

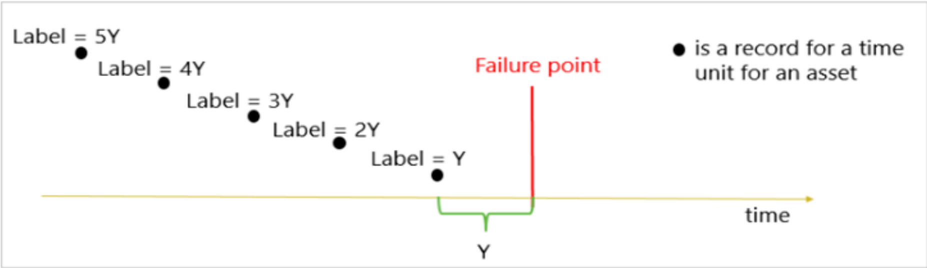


Figure 30. Labelling for regression (Mark Tab, 2018)

3.5.3 Multiclass classification for Predictive Maintenance.

Multiclass classification techniques can be used in PdM solutions in two scenarios:

- *Predict two future results:* the first result indicates the time interval of error of a resource. The resource is assigned to one of several possible time periods. The second result is the probability of error in a future period due to one of the different main causes.
- *Predict the most likely root cause of an error*: This result is used to recommend the correct set of maintenance actions to correct an error. A classified list of the main causes and recommended repairs can help technicians to prioritize their repair actions after an error.

3.5.3.1 Construction of labels for Multiclass classification.

- *Predict two future results:* The aim here is to find the probability that a resource presents an error in the next “nZ” units of time, being “n”, the number of periods selected. Label “nZ” records before the error of a resource by time deposits (3Z, 2Z, Z). Label all other records as "normal" (label = 0). In this method, the destination variable contains the category values.

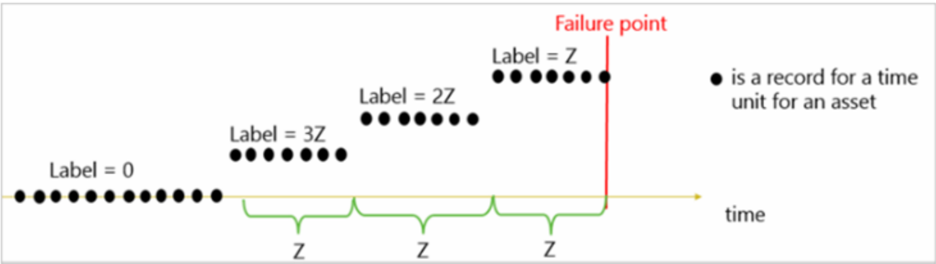


Figure 31. Labelling of the multiclass classification to predict the time of error (Mark Tab, 2018)

- *Predict the most likely root cause of an error:* The aim here is to find the probability that a resource present a problem in the next X time units due to a cause or problem Pi, where i is the number of possible main causes. To answer this question, label X records before the error of a resource such as "near the error due to the main cause “Pi" (label = Pi). Label all other records as "normal" (label = 0). In this method, the labels are also category.

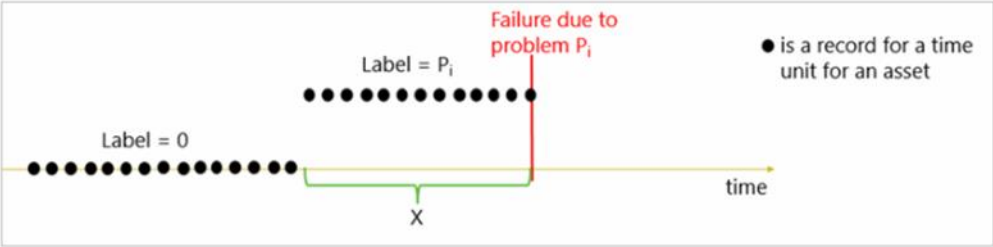


Figure 32. Labelling of the multiclass classification to predict the cause of the error (Mark Tab, 2018)

In the chapter four of this thesis some common AI approaches used for PdM problems will be described. Some of them use classification model techniques and other regression model techniques, or even both. Some Supervised diagnosis methods use classification models like the k- Nearest Neighbour technique. Other classic supervised learning techniques can be proposed, such as linear, quadratic, Bayesian classification, etc. Also, more advanced techniques such as Self-Organizing Maps, Radial NNs, Vectoral Support Maps, Decision Trees or Random Forest are commonly used in PdM problems.

Support Vector Machines (SVM) has also proposed as supervised applications in multivariate processes for detection and identification of faults. Some of the advantages of SVM are that they can get similar results to NNs, with fewer parameters to adjust, with Big Data potential and with categorical and continuous data simultaneously.

Moreover, for Unsupervised Learning other techniques like Clustering methods can be proposed for PdM problems. In summary, Clustering analysis is a multivariate statistical analysis method that groups the signs in different categories of situations according to the similarity of their characteristics.

On the other hand, the prediction of the remaining lifetime is a regression problem that calculates a continuous value. A Logistic Regression to calculate the probability of failure given some monitoring variables is a common method to face PdM problems. Different types of NNs have been applied to estimate the evolution of sensors (Wang, Golnaraghi & Ismail, 2004, pp. 813-831) in the process of propagation of faults and the RUL (Zhang & Ganesan,1997, pp. 378-384). Temporary patterns of vibration characteristics or particles in oil are often used. Other stochastic alternatives are autoregressive, bilinear models, adaptive multivariate regression splines, etc. (Gang 2017).

All these different approaches can be found in the chapter four where the most related to PdM problems are described.

3.6 Learning, validation and testing the methods.

Once the type of modelling to be used has been determined, it will be necessary to choose the specific model that will be applied and how the model will be parameterized. Each identified group will have a series of models that apply different techniques to try to answer the problem. Often, when historical fault and fault data are available in terms of temporal graphs of various signals, it is difficult to determine a predictive physical model. In these cases, it is preferable to apply non-linear approaches such as NNs.

To choose the model and the parameterization with which the forecast will be made it, several models are tested on the data with several parameterizations and the one with the highest precision will be chosen. If there is none that reaches a reasonable precision, it is considered whether the necessity to make any further transformation in the data or include any other data with which it has not counted.

In Predictive models, if all the data is used and the model is adjusted, it could be the case that a model is found that predicted the historical data very well, but that when applied to future data its accuracy is bad. For this reason, these models are adjusted with a subset of the data (usually 80%) called training data and their accuracy is measured using the part of the data with which they have not been trained, called test data. Sometimes a third subset is used to improve parameterization of the model, called validation data.

3.6.1 Cross-Validation.

Cross-validation is a popular technique used to evaluate and validate the model. The principle of using separate datasets for testing and training applies here: The training data is used to build the model; the model is run against the testing set to predict data which has not seen before, which is one way to evaluate its accuracy.

The historical data is divided into X numbers of subsets. Each time a subset is chosen to be used as test data, the rest of the subsets are used as training data. Then, on the next run, the former test set becomes one of the training sets and one of the former training sets becomes the test set. This technique allows to use every data point for both training and testing which is more effective than just dividing the historical data into two sets, the one with the most data for training and the other for testing.

Many ML algorithms depend on different parameters that can significantly change the performance of the model. The optimal values of these parameters are not automatically calculated when training the model. They must be specified by data scientists. There are several ways of finding good values of parameters (Mark Tab, 2018).

One of the most known methods is the *k-fold cross-validation* which divides the examples randomly into k folds. For each set of parameters values, the learning algorithm is run k times. At each interaction, the examples in the current fold are used as a validation set, and the rest of them as a training set. Then the algorithm is trained over training examples and computed performance metrics over validation examples. At the end of these loop, the average of k performance metrics is computed and then, for each set of parameter values, the ones that have the best average performance are chosen (Mark Tab, 2018).

However, in PdM problems, data is collected as a time series of events that come from several data sources (different components inside the organisation). These records use to be ordered according to the time of labelling but, if the dataset is split randomly into training and validation set, some of the training samples may be later in time than some of the validation examples. This led to the future performance values will be estimated based on some data that was collected before model was trained. Such estimations of parameters might be overly optimistic, especially if the time-series is not stationary and evolves over time. Such problem will cause that the chosen parameter values might be suboptimal (Mark Tab, 2018).

Such author recommend splitting the examples into training and validation set in a time-dependent manner, where all validation examples are later in time than all training examples. He also points that, parameter values chosen by train/validation split result in better future model performance than with the values chosen randomly by cross-validation.

3.6.2 Performance test of the model.

Once a model is generated, an estimate of its future performance in new data is required. A suitable estimate is the performance metric of the parameter values calculated with the validation set or an average performance metric calculated by cross-validation as mentioned before. However, these estimations are often overly optimistic, thus, other guidelines can be used to test the model.

3.6.2.1 Time-dependent split.

The recommended method for PdM is to divide the examples into learning, validation and test datasets in a time-dependent manner. All test examples must be later in time to all learning and validation examples. After the division, generate the model and measure its performance, as described above.

When time-series are stationary and easy to predict, both random (Cross validation) and time-dependent methods generate similar estimations of future performance but, when time-series are non-stationary and/or hard to predict as in many cases, time-dependent approaches generate more realistic estimates of future performance.

At the time of implementing time-dependent split is important that training and test data have separate labelling time frames to prevent label information leakage.

Regression models used for predicting RUL are more severely affected by the leakage problem. Using the random split method leads to extreme over-fitting. For regression problems, the split should be such that the records belonging to assets with failures before a chosen time (cut off which divides training data from testing data), go into the training set. Records of assets that have failures after the cut off go into the test set.

Another best practice for splitting data for training and testing is to use a split by asset ID. The split should be such that none of the assets used in the training set are used in testing the model performance. Using this approach, a model has a better chance of providing more realistic results with new assets (Mark Tab, 2018).

3.7 **Model Implementation**

Then, the model must be implemented in the company's systems to create predictions based on new data, not previously seen. The new data must comply exactly with the signature of the trained model in two ways:

- All the characteristics must be present in each logical instance (such as a row of a table) of the new data.
- New data must be pre-processed, and each of the features designed, in the same way as the learning data.

To comply with the signature of the model, the characteristics of the new data must be designed in the same way as the learning data. In the case of typical large datasets for new data, characteristics are aggregated for periods of time and assigned a score.

3.8 **Model Evaluation**

Misclassification is a major problem in PdM scenarios where the cost of false alarms is high for the business. Therefore, the evaluation of models with the correct performance metrics with respect to the new test data is fundamental in PdM approaches.

3.8.1 Standard Metrics.

There are many common ways to check the quality of Classification and Remaining Useful Life methods, (Dhingra, 2018) in his article explain some standard metrics as follow:

– True positive rate*:* refers to percentage of correctly classified failure instances.

$$TP_{rate} = \frac{tp}{(tp + fn)}$$

– True negative rate*:* refers to percentage of correctly classified normal instances.

$$TN_{rate} = \frac{tn}{(tn + fp)}$$

– False positive rate: refers to percentage of incorrectly classified failure instances.

$$FP_{rate} = \frac{fp}{(fp + tn)}$$

– False negative rate*:* refers to percentage of incorrectly classified normal instances.

$$FN_{rate} = \frac{fn}{(fn + tp)}$$

– Accuracy*:* refers to the degree of correctness for all trials. It is the ratio of correct predictions to (over) total examples. Accuracy is a good measure if the incidence of the multiple classes is balanced although not a good metric if the total number of failures is very small.

$$Accuracy = \frac{tn + tp}{(tp + tn + fp + fn)}$$

– Precision: is related to the rate of false alarms, the value give the fraction of correct predictions.

$$Precision = \frac{tp}{(tp + fp)}$$

– Recall: is also called sensitivity and measures the percentage of trials that a positive prediction holds true.

$$recall = \frac{tp}{(tp + fn)}$$

Where…

– tp = true positive= correctly identified
– tn = true negative= correctly rejected
– fp = false positive= incorrectly identified
– fn = false negative= incorrectly rejected

- F1 score*:* is the harmonic average of precision and recall, with its value ranging between 0 (worst) to 1 (best).

$$F1 = \frac{2 \times (precison \times recall)}{(precison + recall)}$$

- Mean absolute percentage error*:* measures the difference between the predicted value and the actual value and gets average difference.
- Root mean square error*:* is like MAE, except that it penalizes large errors more. RMSE squares the differences between predicted and actual value, calculates the average difference, and then take the square root of that average difference.

In recent literature, the most commonly used metrics in PdM are accuracy, precision, mean square error and mean absolute percentage error (Saxena, Celaya, Saha, Saha, & Goebel, 2009, pp.1-13). However, in PdM scenarios, the requirements for an evaluation metric are different from those of a common classification problem. The main reasons are (Jahnke, 2015):

- Causality*:* Some failures in a real-world system are difficult to predict, because the knowledge about the future operation modes of environmental conditions is not controllable.
- Failure data from real applications: Failure data in a real-world system is very rare, because faults are generally expensive and a safety risk.
- Off-line performance evaluation: For obtaining missing failure data, experiments are sometimes conducted in a controlled run-to-failure scenario. However, these experiments may not have the same conditions as the real execution, and therefore, the data may not be precise.

Another problem for all ML algorithms is the uncertainty in prognostics. A good ML technique, and especially in a PdM system, should provide accurate and precise estimations. However, it is also important to obtain information from such a system about the reliability of the prediction.

Approaches which try to predict the state of the real-world system n-states ahead (fixed horizon) as classification methods, have better performance of the metrics mentioned above than approaches with a Residual Useful Life estimation (moving horizon) that say nothing about the accuracy close to the end of its life, which is most important (Jahnke, 2015).

Such standard metrics for ML systems, are not efficient enough to assess performance of them for PdM (Jahnke, 2015) but, unfortunately, PdM systems use to be evaluated with them which gives rise to difficulties comparing those systems to each other. Thus, next section describes four metrics that can be used in a hierarchical manner.

3.8.2 Prognostics Metrics.

(Saxena, Celaya, Saha, Saha, & Goebel, 2010, pp.1-20) introduces a new metrics of prognostic performance for estimating residual useful lifetime. These metrics are more compliant to the requirements of a PdM System. The metrics has a hierarchical design and will be discussed below. Such metrics help to detect weak spots in the models as transient events which can change the behaviour of the model, changes in the operational conditions and conditions which have not being detected in the model. The Hierarchical design of the prognosis metrics can be viewed in next figure:

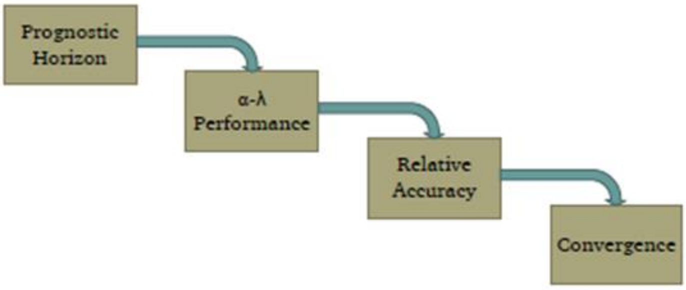


Figure 33. *Hierarchical design of the prognostic's metrics (Saxena, Celaya, Saha, Saha, & Goebel, 2010)*

3.8.2.1 Prognostic Horizon.

The prediction horizon is the difference between the time $i_{\alpha\lambda}$ when the predictions first meet a specified performance criterion, and the time index for the *End of Life, EoL.* The prediction horizon (PH) can expressed as:

$$PH = t_{EoL} - t_{i\alpha\lambda}$$

The prediction horizon is the time between the *End of Life (EoL)* and the time when the prediction of the PdM system for the first time fulfils the performance criteria. This allows to identify when the system first enters the specified error margin around the actual end-of-life and gives information about how much time there is for any corrective action to be taken.

Next figure shows an example of two different prognostics and the prognostic horizon (Jahnke, 2015). The black line shows the residual useful lifetime, degrading and linear with increasing time. The grey lines represent the border of the performance criteria. The red and the blue lines represent the results of the predictions of two different PdM systems. The red and blue points represent new predictions on grounds of new data available from the real system. The prediction horizon is the time between the End of Life (EoL) and the time when the prediction of the PdM system for the first time fulfills the performance criteria (prediction is between the black and the grey line). That implies, that PH^2 is the prediction horizon of the blue PdM system and PH^1 is the prediction horizon of the red PdM system.

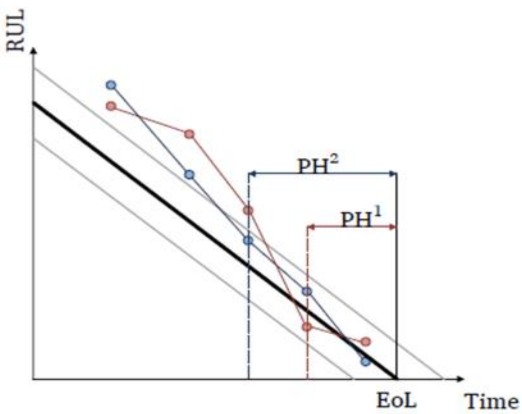


Figure 34. Example of prognostic horizon of two different PdM systems (red and blue lines) (Saxena, Celaya, Saha, Saha, & Goebel, 2010)

3.8.2.2 α-λ Performance.

The α-λ accuracy is defined as a binary metric, which evaluates the accuracy and confidence at a time t_λ:

$$\alpha - \lambda\ accuracy = \begin{cases} 1\ if\ \pi[r(i_\lambda)]_{-\alpha}^{+\alpha} \geq \beta \\ 0\ otherwise \end{cases}$$

Where β is a threshold and the formula compared is the probability mass of the prediction probability density function within the bounds +α to -α. Here, the probability mass of the prediction probability density function represents the confidence of the prediction and the α bounds the error margins of the prediction. So, the smaller the α bounds the higher accuracy is achieved. Such prediction needs to have greater or equal than β of probability mass within the α bounds. For a given time instant, this prediction identifies whether the algorithm evaluated perform within the desired error margins of the actual residual useful time.

3.8.2.3 Relative Accuracy.

Such method gives information about accuracy at a given time, without the accuracy level (as the method before). (Jahnke, 2015) points that is needed because in PdM systems, the accuracy of the prediction must increase with time to the End of Life´s component. So, it will help to distinguish between estimations at the beginning or the ending lifetime´s component. Such method can be expressed as follows:

$$relative\ accuracy = 1 - \frac{|r_*(i_\lambda) - (r(i_\lambda))|}{r_*(i_\lambda)}$$

Where i_λ is the time index, $r_*(i_\lambda)$ is the ground truth residual useful lifetime at i_λ and $r(i_\lambda)$ is the predicted residual useful time at such time index.

3.8.2.4 Convergence.

The convergence define de distance between the origin and the prediction result and can be expressed as:

$$convergence = \sqrt{(x_p - x_0)^2 + (y_p - y_0)^2}$$

Where (x_p, y_p) are the coordinates of the prediction and (x_0, y_0) are the origin coordinates, being the x- axis the lifetime and the y- axis the residual useful lifetime. So, the End of Life correspond with a value of 0 in the y-axis, y_0 =0. Such metric quantifies how fast the system converges when it satisfied all previous metrics.

With these four metrics, an assumption of the quality of a PdM system is much more descriptive than using the standard metrics pointed before and comparisons between systems are more sensible.

4 Common AI approaches in Predictive Maintenance

In recent literature, a variety of authors have focused on PdM techniques. Some authors like (Sheng Si, Wang, Hu & Zhou, 2011, pp. 1-14) who provide a review of statistical driven approaches, writing about parametric ML approaches. Others like (Peng, Dong & Zuo, 2010, pp. 297-313) describe the process of PdM approach, focusing on data-driven methodologies, where knowledge-based techniques are introduced.

However, in this chapter the main approaches carried out nowadays related to PdM will be introduced, as well as, putting special attention and focusing on ML algorithms as it is the NN methodology.

There are several approaches to implement AI which can be divided in Rule-Based techniques and ML techniques (Jahnke, 2015).

4.1 Rule-Based systems

Rule-based systems mainly use rules as the knowledge representation. These rules are coded into the system in the form of *if-then-else* statements. The main idea of a rule-based system is to capture the knowledge of a human expert in a specialized domain and embody it within a computer system. Focusing on PdM, these systems rely on sensors to continuously collect data about the assets, for after that, sending alerts according to predefined rules. These rules provide a level of automated PdM, but they must depend on a team of experts who understand which parts or environment elements require measuring.

Ruled-based systems also called Knowledge-Based Models are created by domain experts, although mathematical models are not used to cover the physical behaviour. The quality of these systems depend on the knowledge of such experts which are totally responsible of the success of them. The most common Rule-Based system technologies in the field of PdM are Expert Models and Fuzzy Logic (Jahnke, 2015).

4.1.1 Expert Models.

These techniques solve problems making use of specialists who define rules and describe the state of the real system. Such rules are written in the form: IF condition, THEN consequence, ELSE consequence. Conditions describe the real system with the consequence of a result or another rule.

The main problem of these types of systems relies on the limit in the ability to simulate intelligence as it is said to have a rigid intelligence. Expert Models cannot handle situations that are not covered in the rule-based system which means that when the system encounters a problem for which no rules have been designed, it will not be able to solve the problem. Moreover, if the number of rules increase dramatically, the model can run into a combinational explosion. These issues, the high cost and time consuming of maintaining these systems added to the difficulty of solving complex systems, make them not very used in PdM (Zhou, Hu, Zhang, Xu & Chen, 2013, pp. 402-413). However, sometimes Expert Models can be combined with ML techniques which achieves better results than using the techniques alone.

This thesis will not focused on this type of models, but other authors have written about them. (Butler, 1996, pp. 321–326) proposed a framework, based on the Expert Model for PdM. (Biagetti & Sciubba, 2004, pp. 2553–2572) created a prognostic and intelligent monitoring expert system, which detects faults in real time and provides forecasts for detected and likely faults.

4.1.2 Fuzzy-Logic systems.

Fuzzy logic (Lotfi, 1965, pp. 338–353) methodology can be simply described as a situation where it is not possible to say "Yes" or "No" because of the necessity of more information. This technique tries to describe the real system based on imprecise or noisy inputs. Therefore, the Fuzzy-logic system makes a description in a continuous manner, which is more intuitive and less specific than a mathematical description.

Every input value belongs to a fuzzy rule in a certain degree of membership covered from 0 (not being a member) to 1 (a member). This means that all degrees of membership from the input value can take part in the rule and the result, which describes the state of the system. For a given element, the fuzzy logic returns the degree of membership to a set. With a fuzzy logic model, states can be described in a continuous and overlapping manner, like the state transitions in a real system. Therefore, a fuzzy logic model can create simpler, more intuitive and better-behaved models (Peng, Dong, & Zuo, 2010, pp. 297–313).

In PdM, these systems are used to develop Fuzzy Risk Assessment in components of the organisations. The inputs variables describe characteristics of the system like temperature, vibration… in values between 0 and 1. Then by using If-then fuzzy rules, the system provides output variables which referred to some parameters like the life expectancy in some cases. As is not the main topic in this thesis, more emphasis on this type of system will not be made but, for more information (Sudarsanam, 2014) in his article describes more in detail this topic using a clear example.

4.2 Machine Learning Systems

These types of systems rely on statistical and learning techniques. In contrast to Rule-based systems, these techniques have a very ambitious goal instead of a concrete vision, which is implementing AI through the learning capability of these systems. Whereas Rule-based systems are based on explicit stated static models of a domain, ML systems create and adapt their own systems during the time. Moreover, instead of relying in human experts or explicit programming as the previous systems, these techniques just automate the process of learning and improve their own performance by collecting data based upon experience for achieving the desired result.

Nowadays, with the development of AI and the growing number of smart sensors in real systems, most approaches in recent literature of PdM account for ML models. On this thesis, a description of the common ML techniques is made as well as focusing on ANNs.

In general, ML techniques are often divided according to the degree of data monitoring.

4.2.1 Supervised Learning.

This type of systems rely on data collected from the past which is referred to as historical data and then it is filtered so that it can be used as learning examples. Then the data is labelled according to its expected result which is called training data. After that, ML techniques process the data, examining the relationship between the data recorded and the labelled output which leads to the creation of the ML model. When the models are created, for any new data, it will give the best result based upon the data learned (Jahnke, 2015).

Supervised learning problems are generally categorized into "regression" and "classification" problems which were described before. In PdM this type of systems are the most common as the real system is monitored and data is available. Common algorithms in supervised learning include Logistic Regression, Naive Bayes, Support Vector Machines, ANNs, and Random Forests.

4.2.2 Unsupervised Learning.

Sometimes, the historic data has not corresponding target values, in these cases the aim of the system is to find or discover groups of similar examples in the historic data, which is known as Clustering. Other times, the aim is to find rules that describe large portions of data also called Association Analysis. Other common tasks within these techniques are representation learning, and density estimation, which tries to determine the distribution of data within an input space. In all these cases, the main goal is to learn the inherent structure of our data without using explicitly-provided labels. Some common algorithms include k-means Clustering, Principal Component Analysis, and Autoencoders.

In PdM problems, the Unsupervised Learning technique is an unusual learning technique, because the clustering and density estimation of the historic data is not efficient for precise failure type detection and PdM (Jahnke, 2015).

4.2.3 Reinforcement Learning.

This technique unlike the other learning methods has not learning phase for the historic data collected. The ML algorithm relies on a trial and error scheme. The essence of Reinforcement Learning is learning through interaction, which is implemented by putting the machine in certain state, then take an action to after, bringing it to another state. The goal of this technique is to define the final desired state. When the action taken by the systems is near the final state, it is stored as a positive action but, when is further, it is stored as a negative action (Arulkumaran, Deisenroth, Brundage & Bharath, 2017, pp. 26-38.). An example of Reinforcement Learning is the Markov Decision Process.

In recent literature, there are not cases of Reinforcement Learning techniques used in PdM because the learning process happen under runtime condition, which implies that a wrong estimate of the state of the real system could have disastrous consequences (Jahnke, 2015).

4.3 Machine Learning Approaches

ML techniques are mainly based on different programming techniques to assess the data collected and the experience of domain techniques who evaluate such techniques. The principal task of such approaches is to identify the relation between the data collected in form of vector and a result. Such relation is expressed in a model and then, the ML algorithm must optimize the parameters of the model created.

ML has become important in these years as data can be collected and stored much more easily. Such techniques play an important role since the data collected is so extensive that is not possible to analyse manually and here is just where ML takes part. Moreover, these methods have become very popular thanks to the evolution of hardware in recent years, using these techniques is much more efficient in terms of both time and money especially when applied in PdM tasks.

The following section sets a revision of the different methods of analysis using ML algorithms. Such techniques can be divided into two categories: Parametric and non-parametric approaches.

4.3.1 Parametric Approaches.

ML algorithms are based on the information obtained from samples of the data collected. Parametric Approaches assume that such samples are distributed in known models which can be described by some parameters of fixed size, which means that they are independent of the number of training examples. The number of parameters will not change although more data will be sent. The advantages of such approach rely on that the model is based in few parameters, as can be the mean or the variance, as well as, its simplicity and speed. ML plays the role of estimating the parameters in respect of the data collected, then, the assumption that all data complies with this distribution is made it.

Although parametric approaches are faster to compute, they need exact assumptions for the distribution to achieve a successful result. If the assumption about the parameters is not the same, the result will seem to be interpreted arbitrarily (Jahnke, 2015).

Next, the most used Parametric ML or data mining techniques in literature are reviewed, highlighting those that are usually apply to the maintenance field.

4.3.1.1 Regression Models.

Regression Models types are mainly based on modelling the relations between the input variables or features and the output variables, iteratively refining the measure of the error in the predictions made by the model (Plasmatic, 2017).

- Ordinary Least Squares Regression (OLSR)
- Linear Regression
- Stepwise Regression
- Multivariate Adaptive Regression Splines (MARS)
- Locally Estimated Scatterplot Smoothing (LOESS)
- Logistic Regression

As this type of methods are not very common in PdM problems, in this thesis they are not explained in detail.

4.3.1.2 Dimension Reduction models.

Such approaches are like Clustering in the sense that they seek to exploit the structure of the data. They try to summarize the data using minimal information which can be useful to visualize multiple dimensions or to simplify data. Many of these methods can be adapted for classification or regression. The most common techniques used in ML for PdM are:

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Principal Component Regression (PCR)
- Partial Least Squares Regression (PLSR)
- Sammon Mapping
- Multidimensional Scaling (MDS)

– Projection Pursuit
– Mixture Discriminant Analysis (MDA)
– Quadratic Discriminant Analysis (QDA)
– Flexible Discriminant Analysis (FDA)

As Linear Discriminant Analysis is one of the most popular methods from all of them, it is explained more in detail in this thesis.

- **Linear Discriminant Analysis (LDA).**

This type of method is commonly used for dimensional reduction and classification, although it can appear at the time of feature extraction. Logistic Regression is commonly limited to only two-class classification problems, so that, when there are more than two classes Linear Discriminant Analysis is preferred.

Linear Discriminant Analysis (LDA) is used to find a combination linear analysis of the variables that reduces the total dimension of the problem and allows to describe the differences between groups to subsequently provide a tool to classify new observations. This supervised learning technique aims to project the data in a smaller dimension obtaining more separation between the data of the different classes.

The simplest representation of a linear discriminant analysis is the linear function:

$$y(x) = w^t x + w_0$$

Where x is the input vector (features), w^t is weighted vector and w_0 is a bias. For a problem of two classes, C_1 and C_2, a classification of the input vector x will be C_1 if $y(x) \geq 0$, otherwise C_2. The projection from x to w^t is a dimensional reduction from d to 1. Vector w is orthogonal to every vector lying within the decision surface. Similarly, if y(x)=0, this implies that x is a point on a decision surface, the normal distance from the origin to the decision surface is:

$$\frac{w^t x}{||w||} = -\frac{w_0}{||w||}$$

Next figure shows the geometry for a two-dimensional function in two dimensions. The decision surface (red), is orthogonal to w, which origin´s displacement is given by w_0 . Orthogonal distance of a point x from the decision surface is $\frac{y(x)}{||w||}$.

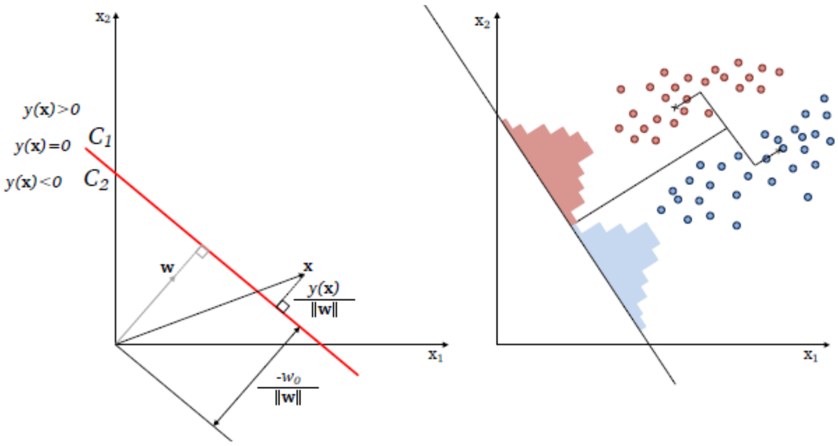


Figure 35. An illustration of a discriminant function in two dimensions (left) and a plot on the right which samples two classes (red and blue) by using the Fisher linear discriminant (Bishop, 2006)

If there were more than two classes, defining k linear functions for k classes it would be key

$$y_k(x) = w_k^t x + w_{k0}$$

A vector x is assigned to a class C_k if

$$y_k(x) > y_j(x)$$

The Fisher´s linear discriminant can be used to define a weighted vector w^t that maximizes the class separation. For a C_1 class with N_1 points and C_2 class with N_2 points, the mean vector of such classes is:

$$m_1 = \frac{1}{N_1} \sum_{n \in C_1} x_n$$

$$m_2 = \frac{1}{N_2} \sum_{n \in C_2} x_n$$

And choosing w to maximize:

$$m_1 - m_2 = w^t(m_1 - m_2)$$

Where:

$$m_k = w^t m_k$$

Then the variance of the transformed data from class C_k can be expressed as:

$$s_k^2 = \sum_{n \in C_k} (y_n - m_k)^2$$

Where $y_n = w^t x_n$, being the distance between classes as large as possible to have a small variance. Being the desirable case having a high value for $|m_1 - m_2|$ and low for $s_1^2 + s_2^2$ for two classes. With Fisher´s criterion the aim is maximizing w to fulfil:

$$J(w) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

Such criterion with the other equations can be rewritten as:

$$J(w) = \frac{w^t S_B w}{w^t S_w w}$$

Where S_B is the between class covariance matrix defined as:

$$S_B = (m_1 - m_2)\,(m_1 - m_2)^T$$

and S_w is the within-class covariance matrix, expressed as:

$$S_w = \sum_{n \in C_1} (x_n - m_1)\,(x_n - m_1)^T + \sum_{n \in C_2} (x_n - m_2)\,(x_n - m_2)^T$$

Then differentiating J(w) equation with respect to w, it is maximized when

$$(w^t S_B w) S_w w = (w^t S_w w) S_B w$$

Where $(w^t S_B w)$ and $(w^t S_w w)$ are scalar factors which do not provide information

about the direction of **w** and can be dropped. Last equation can multiply both sides with S_w^{-1}, so the result of the Fisher linear discriminant can be expressed as:

$$w = S_w^{-1}(m_1 - m_2)$$

The right-hand side of Equation shows the variance of the mean value and the Fisher linear discriminant of an example separation problem. The linear discriminant analysis is an easy and useful classification technique, when the classes can be linearly separated from each other (Bishop, 2006). One of its drawbacks is the creation of a weighted vector, which is difficult, because they need a linear dependency on each other, and that is not automatically given in real-world systems. But still it is possible to have such linear dependencies in the data from a data acquisition system (Jahnke, 2015).

4.3.1.3 Bayesian Models.

These techniques apply the Bayesian theorem to solve problems of both classification and regression. Some techniques or models used in PdM are the following:

- Naïve Bayes
- Gaussian Naïve Bayes
- Multinomial Naïve Bayes
- Averaged One-Dependence Estimators (AODE)
- Bayesian Belief Network (BBN)
- Bayesian Network (BN)

The most used techniques used in PdM of these models are de Naïve Bayes and the Bayesian Network (Wu, Chen, & Wang, 2012, pp. 103-122) which are explained in the following section, showing the main characteristics of such models and how they work.

- **Naïve Bayes.**

In ML, Naive Bayes classifiers are a family of probabilistic classifiers based on Bayes theorem. It assumes that a value of a feature is independent of a value of any other feature, given the class variable

(Russell, S. et al 2003). This assumption is often violated in practice but even though Naive Bayes classifier is still powerful classification techniques.

These models are based on the calculation of probabilities from the training data set. The probability to be estimated is a conditional probability $P\ (c_j|x_1, ..., x_d)$ where c_j refers to each class and $X= (x_1, \ldots, x_d)$ are the different features.

After using the Bayes rule the posterior probability can be expressed by…

$$P\ (c_j|x_1,\ldots,x_d)=\frac{P\ (c_j)P(x_1,\ldots,x_d|c_j)}{P(x_1,\ldots,x_d)}$$

Where here…

- *P(c|X)* is the posterior probability of class (target) given predictor (feature).
- *P(c)* is the prior probability of class.
- *P(x|c)* is the likelihood which is the probability of feature given class.
- *P(x)* is the prior probability of feature.

The resulting model is represented by prior probabilities of each class and likelihood probabilities for each combination of class and feature. Finally, assuming the independence between al features the likelihood of each class can be calculated as follows:

$$P\left(x_1,\ldots,x_d\middle|c_j\right)=\prod_{i=1}^{n}P(x_i\ |c_j)$$

The class with the highest posterior probability is the outcome of prediction.

- **Bayesian Networks.**

Bayesian networks are a widely-used class of probabilistic graphical models. (Cózar, Puerta & Gámez, 2016, pp. 176 – 195). They consist of two parts: a structure and parameters.

The structure is a Directed Acyclic Graph (DAG) that expresses conditional independencies and dependencies among random features associated with nodes. The parameters consist of conditional probability distributions associated with each node. A Bayesian network is a compact, flexible and interpretable representation of a joint probability distribution. It is also a useful tool in knowledge discovery as DAG allows representing causal relations between features. Typically, a Bayesian network is learned from data.

Next figure shows an example graph representing a Bayesian network with six random features.

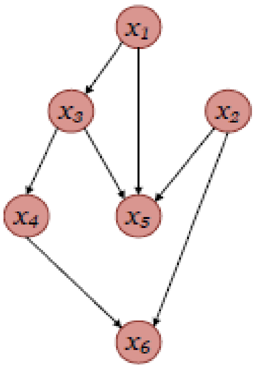


Figure 35. Example of a Bayesian network with six variables or nodes (Jahnke, 2015)

In this case, the probability distribution of the combination of these random variables is called joint distribution which here is given by:

$$p(x_1)\, p(x_2) p(x_3|x_1) p(x_4|x_3) p(x_5|x_1, x_2, x_3)\, p(x_6|x_2, x_4)$$

Generally, for a graph with *k* nodes or features, the joint distribution can generally be expressed as:

$$P(x) = \prod_{k=1}^{n} P(x_k\, |pa_k)$$

where pa_k are the parent nodes of x_k. The probabilities $p(x_k|pa_k)$ of the possible combinations can be listed in a conditional probability table.

An important and attractive issue of Bayesian Networks is their ability to incorporate the temporal dimension, allowing in this way reasoning over time. *Dynamic Bayesian Networks* allow to represent different instances of the same variables over time, as well as temporal relations between them. Next figure shows the most used way of them, which consists of a basic structure, which represents a static network, together with a set of temporal relations representing the dependences from time $t-1$ to time t. This structure is unfolded as many times as needed to forecast the values of variables at time $t+k$ (Cózar, Puerta & Gámez, 2016, pp. 176 – 195).

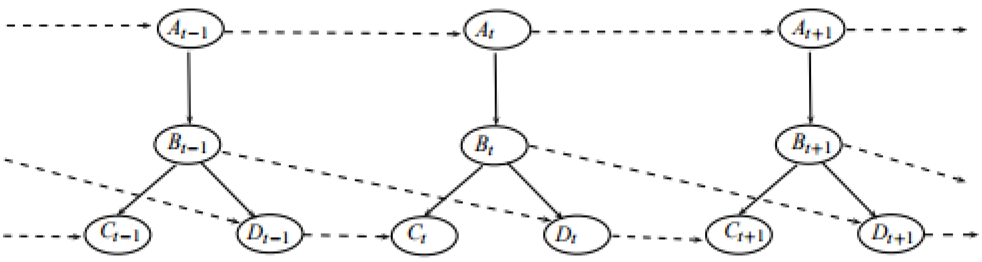


Figure 36. An example of a Dynamic Bayesian Network unfolded over the time (Cózar, Puerta & Gámez, 2016)

Bayesian Networks represent an important tool for PdM problems as can build a model from a set of examples to predict the health status of the real system. There are two ways to learn about a Bayesian network. First, it exists as the structure of a network and so it is not necessary to mention the conditional independencies. It is only necessary to calculate the conditional probability on every node. The second way is to create a network structure with conditional independencies and a conditional probability on every node (Jahnke, 2015). Although, such technique use to face problems where data is labelled, and the class is known (normal behaviour or failing), sometimes this information is not available and other approaches have to be used. However, (Cózar, Puerta & Gámez, 2016, pp. 176 – 195) in their article show how to apply this technique to propose a failure detection tool whose goal is to detect generic failures from the information collected in a sensored system and can be included in classification-based anomaly detection techniques.

4.3.1.4 Hidden Markov Models.

Hidden Markov models assume that the system can be modelled by a Markov chain with hidden states (not observed) and transitions between them. The Semi-Markov models also add a duration to each state, being also able to be estimated with non-parametric models with lower computational cost, which make them be preferred (Jardine, Lin, & Banjevic, 2006, pp. 1483–1510).

In terms of PdM, a state represents a health state of the real system with its own duration in time units, which can be represented as time usage, working cycles, maintenance intervals… Each state is based on single states which are observations of the real system. Moreover, these models can be used to make prediction of the Residual Useful lifetime as showed in (Dong & He, 2007, pp. 858–878).

Next figure shows a sample of a semi-Markov model structure, where the blue cycles are states of the model. The black arrows are the transitions between the states. Every state has its own duration (d_1, ..., d_n) of time units (1, ..., q_n). These time units are observations of the real system (o_1, ..., o_{qN}) and every observation represent a single state (s_1, ..., s_{qN}).

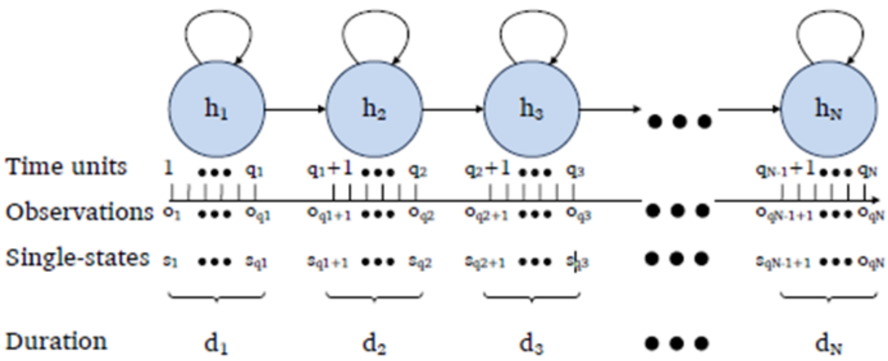


Figure 37. Example of a semi-Markov model structure (Jahnke, 2015)

4.3.2 Non-Parametric Approaches.

These approaches rely on the assumption that are close to one another must have the same result, where that “close to” makes the difference between the other techniques. Such methods are very useful when there are high amounts of data and no prior knowledge about the system studied, here the experts do not need to worry about choosing just the right features or parameters (Jahnke, 2015). Therefore, the main difference between this approach and the parametric one is that there is not a finite number of parameters and the complexity of the model grows with the number of training data.

Some investigations in recent literature carried out by (Ackermann, Bartlett, Kaesbauer, Sienel & Steinhauser, 1993) (Isermann & Münchhof, 2010) and (Reinelt & Ljung, 2003, pp. 373–380) indicate that such approaches have more stable result than Parametric ML techniques. This can happen because of the statistics satisfying a PdM process are too difficult to be estimates exactly, as most of real systems are non-linear and use to have uncertainties due to noises, incomplete knowledge…

The main disadvantages of such approaches are the necessity of more data to estimate the function, the slowness due to the increase in the number of parameters and the difficulty of interpreting them.

4.3.2.1 Support Vector Machines (SVM).

This technique is a nonparametric ML technique that is useful when the underlying process of the real system is not known, or the mathematical relation is too expensive to be obtained due to the increased influence by several interdependent factors (Jahnke, 2015). This type of technique are probably the most popular approach in classification, thanks to their high classification accuracy, even for non-linear problems, and to the availability of optimized algorithms for their computation (Chang & Lin, 2011, pp. *27*).

In this algorithm, each data item is plotted as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a coordinate. Then some lines that splits the data between the differently classified groups of data are found. This is the line such that the distances from the closest point in each of the groups is the farthest away. Next figure shows an example for two class and two features problem.

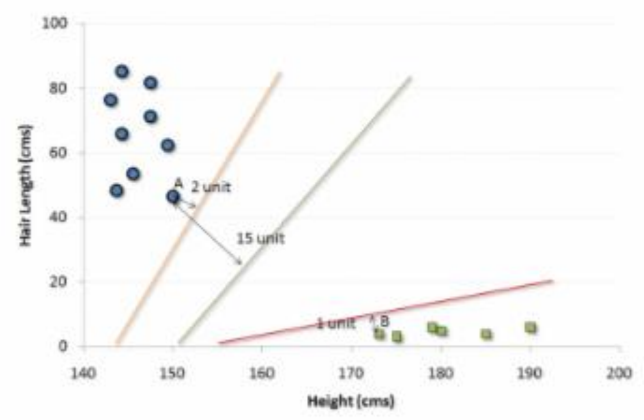


Figure 38. An example of a Support Vector Machine approach for two features and two classes (Ray, 2017)

The line which splits the data into two differently classified groups is the black line, since the two closest points are the farthest apart from the line. This line is the classifier. Then, depending on where the testing data lands on either side of the line, that is what class one can classify the new data as.

SVM are usually employed in combination with Kernel Methods to further enhance the classification performance by allowing non-linear solutions. In general, the boundary between two classes is non-linear in the feature space, so it must also be the multidimensional hyperplane. To reduce the complexity of your estimate, a Kernel function is used, that maps the input vector to another with more dimensions, where it is easier to calculate the boundary between classes based on historical data, or support vectors (Plasmatic, 2017). The technique with the maximal margin hyperplane focuses on a trade-off between model accuracy and model ability in predicting future values (Moura, Zio, Lins & Droguett, 2011, pp. 1527–1534). This characteristic tends to improve the tool effectiveness to forecast upcoming outputs, which is a main requirement for a ML approach in PdM. For more detailed presentations of these approaches (Scholkopf & Smola 2002) and (Suykens & Vandewalle, 1999, pp. 293–300) focus on them.

4.3.2.2 Decision Tree models.

Decision trees can be viewed as a set of non-overlapping rules, which can incorporate expert knowledge and historical data. These techniques are based on the construction of a model in the form of a tree based on the real values of the data. They are used in problems of classification and regression, being fast and precise (Jahnke, 2015). Decision Tree methods empower predictive models with high accuracy, stability and ease of interpretation. Next figure shows an example of a Decision Tree with three variables or features, rules in the ellipses which results in the arrows and the result of the tree in the leaves:

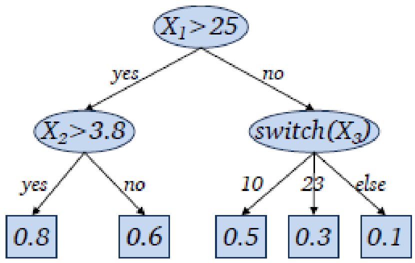


Figure 39. Example of a Decision Tree (Jahnke, 2015)

In general, after its construction, the Decision tree model is usually pruned. The most used techniques for their learning are C4.5 and C5.0, which evolves the first one achieving great purity, speed of computation and efficiency in the use of memory, besides including boosting and costs of poor classifications (Tomasz, Riaz, & Pedersen, 2012, pp. 237–241). (Plasmatic, 2017) marks other learning techniques which are the following:

- Classification and Regression Tree (CART)
- Iterative Dichotomiser 3 (ID3)
- Chi-squared Automatic Interaction Detection (CHAID)
- Decision Stump
- M5
- Conditional Decision Trees

No approach combines PdM in one decision tree in current research. (He, He, & Wang, 2013, pp. 25–34) implemented two decision trees: a decision tree identifying the state of the system as well as any failure in the near future, and a tree identifying the type of upcoming failures. Moreover, a decision tree can also be used for feature selection, as (He et al. 2013, pp. 25–34) uses the resulting rules of such algorithm to identify the most used features, by which a NN is learned.

One of the most important steps when the Decision Tree is being created is related to know whether the tree is going to split or not. The decision of making strategic splits will affect heavily the tree´s accuracy, such decision criteria is different in classification and regression trees. To face this difficult step these models, use multiple algorithms to decide to split a node in two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. Thus, such methods splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes.

In their article, (Analytics Vidhya Content Team, 2016) explain how work the most popular algorithms in Decision Trees, which are the following:

- Gini algorithm
- Chi- Square algorithm
- Information gain
- Reduction in Variance

However, when building the Decision Trees overfitting is one of the key challenges to face. Preventing overfitting is essential while modelling the tree and it can be done in two ways. First way is by setting constraints on tree size which can be done by using various parameters when defining the tree and applying them using a greedy-approach. The other way is the so-called *Tree Pruning* which is different from the other because is based on looking at a few steps ahead when modelling and making the best choice.

Between the main advantages of this method are the ease to understand, the usefulness in data exploration and the small amount of data cleaning required compared to other techniques.

In maintenance, these techniques are often used for the classification of the state of a system, but not for the regression of the residual useful life, since the tree can only have one finite number of sheets, representing the possible results. For example, (He et al. 2013, pp. 25–34) create trees to identify the most useful features, know the state of the system, predict future failures, and establish the type of failure.

4.3.2.3 Random Forest.

Random Forest is a versatile ML method capable of performing both regression and classification tasks. It also undertakes dimensional reduction methods, treats missing values, outlier values and other essential steps of data exploration, and does a fairly good job. It is a type of ensemble learning method, where a group of weak models combine to form a powerful model (Analytics Vidhya Content Team, 2016).

Random forest trains and combines different models, building a set of decision trees, staying with the most chosen class (classification) or with the average of the results (regression). In general, each tree is trained only with a subset of the features, which reduces the learning time. The combination also prevents overtraining. Despite Decision Trees are easily interpretable, this is not the case when they are combined in random forests or other ensembles, although there are methods to extract the importance of variables, dependencies and proximities (Aldrich & Auret, 2013).

Some of the most used combination strategies are based on votes, minimum, maximum or average. More elaborately, it considers information from the training phase using Bayesian methods, behavioural knowledge spaces, Dempster-Shafer theory, or correlations between the results of each model (Gang, 2017).

One of the most important advantages of Random Forests relies on their ability to determine feature´s importance by randomly permuting (shuffling) a given feature. In this way, the feature should have no relationship with the response. A statistic measuring the difference in the random forest prediction accuracies using the original data and that of random forest predictions using the shuffled feature is then calculated. A single feature importance measure is computed as the average of these differences across every tree in the forest. The process is repeated for each feature. The features may be ranked according to this difference measure, the largest difference indicating a feature furthest from a random shuffling and thus most important (Breiman, 2001, pp. 5-32).

Other advantages of this method are the power of handle large data set with higher dimensionality as can handle thousands of input variables, become it in one of the dimensionality reduction methods. Moreover, it has methods for balancing errors in data sets where classes are imbalanced, and it can be extended to unlabelled data, leading to unsupervised clustering, data views and outlier detection.

Although Decision Trees, as NNs, are unstable statistical models, in the sense that a small change in the training set or initialization can dramatically affect the final model (Breiman, 1996, pp. 123–140) Precisely, this instability can be exploited using the theory of ensembles.

According to recent literature on PdM, the Random Forest ML approach is used due to its low computational cost with large data and stable results (Jahnke, 2015).

As pointed, Random Forest method belongs to a series of techniques called *Ensembles* which independently train multiple simple models, whose predictions combine to arrive at a global prediction, other techniques in the field of ensembles are:

- Bootstrapped Aggregation (Bagging)
- AdaBoost
- Stacked Generalization (blending)
- Gradient Boosting Machines (GBM)
- Gradient Boosted Regression Trees (GBRT)

For more specific information about how these techniques works and how can be implemented using different programming tools, (Analytics Vidhya Content Team, 2016) and (Srivastava, 2014) have several articles describing all these process as well as (Rey, 2018) is his bachelor´s thesis does.

4.3.2.4 K- Nearest Neighbour.

It can be used for both classification and regression problems. However, it is more widely used in classification problems in the industry. K-nearest neighbours is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbours. The case being assigned to the class is most common amongst its K nearest neighbours measured by a distance function.

The search for nearest neighbours estimates the probability of a density function (Bishop, 2006). A distance measure such as Euclidean, Hamming or Manhattan is used to identify the neighbourhood of a feature vector, that is, the closest vectors seen previously.

The most important step of the technique is at the time of choosing the k value, which determines the number of neighbours the algorithm look at when assigning the value to any observation. Here there are two important concepts which are the training error rate and the validation error rate for different values of k. Next figure shows an example of both values from a problem where the technique is implemented.

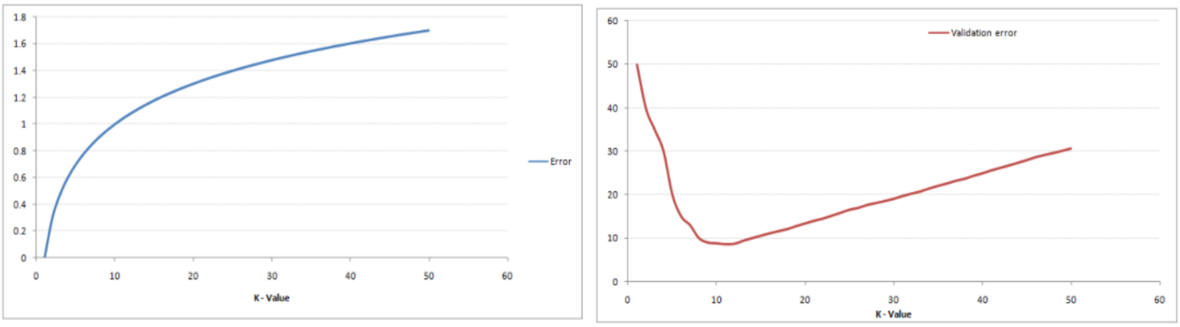


Figure 40. Comparison between training error rate and validation error rate depending on the k value (Srivastava, 2018)

After studying both rates, the problem becomes clearer. In this case, the validation error rate reaches the minima point and after, starting to increase with k. Such minima will be the value chosen of k for all prediction in the sample.

This algorithm is one of the simplest classification algorithms. Even with such simplicity, it can give highly competitive results. It can also be used for regression problems. The only difference from the discussed methodology will be using averages of nearest neighbours rather than voting from nearest neighbours (Srivastava, 2018).

This technique has some drawbacks that need to be mentioned with its use for failure type detection and PdM (Jahnke, 2015):

- Training sets are sometimes not equally distributed, in this case correct classification cannot be achieved or can result in misclassification.
- The type of distance metric has a significant influence on the result. To achieve a good result, the non-linear behaviour of real system needs different distance metrics for the lifetime.

Despite the drawbacks, the nearest neighbour technique is used PdM, due to the lazy learning technique and simple usage (Jahnke, 2015).

4.3.2.5 Clustering methods.

Clustering methods are other important techniques of unsupervised learning, trying to find a structure in a collection of unsupervised data (Gang, 2017).

Intuitively, such techniques seek to organize the objects into groups whose members are similar according to some criteria (often, depending on the user's requirements). Clustering techniques can belong to both parametric and non-parametric approaches.

The main problems of clustering are (Gang 2017):

- The desired requirements are not met.
- Complexity depending on the dimensions of the data.
- Effectiveness depending on the definition of distance (in complex cases, it must be settled).
- The result is not always interpretable

The data are organized based on inherent structures deduced from them. They can follow a hierarchical or centroid approach. The most common types are the following:

- k-Means
- k-Medians

– Expectation Maximization (EM)
– Hierarchical Clustering

In this thesis, such methods are not explained in detail as, it focused more on other type of techniques.

4.3.2.6 Artificial Neural Networks (ANN).

An ANN consists of interconnection of neurons. The neurons are usually assembled in layers (Barad et al., 2012, pp. 729-742). Each layer has several simple, neuron processing elements called nodes or neurons that interact with each other by using numerically weighted connections (Peng, Dong & Zuo, 2010, pp. 297-313).

Generally, NNs consist of n layers of neurons of which two are input and output layers, respectively. The former is the first and the only layer which receives and transmits external signals (values of the features) while the latter is the last and the one that sends out the results of the computations. The n-2 inner ones are called hidden layers which extract, in relays, relevant features or patterns from received signals. Next figure show the structure of a NN with two hidden layers.

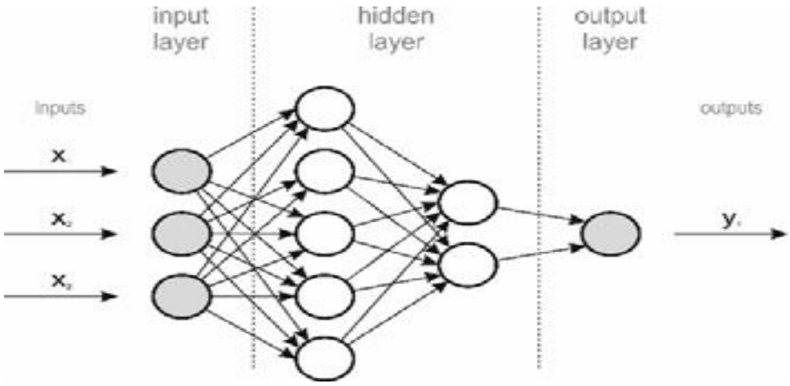


Figure 41. Structure of a NN with two hidden layers (Křenek et al. 2016)

Every neuron has an activation function. The most common activation function is the sigmoid and radial base function (Hrushikesh, Mhaskar & Micchelli, 1992, pp. 350–373). The usual way of designing an ANN is as a feed-forward NN, which can be represented as a directed acyclic graph from the input to the output nodes. Theoretically, an ANN can learn by the following operations:

– Creation of new neurons with new connections, removal of existing neurons and their connections
– Creation of new connections between existing neurons, removing existing connections between neurons
– Changing of the weights between neurons
– Adjustment of the neuron's thresholds
– Changing the properties of the activation function

The basic principle of learning is gradient descent, which compares the output of the network with the labelled output. This comparison is measured typically by the Mean Square Error and, by readjusting the weights of the neurons, this error is minimized.

- **Main Concepts.**

In their article, (MissingLink Team, 2018) pointed the main concepts which need to be understood to how these types of techniques work are the following:

– Inputs*:* Source data fed into the NN with the goal of deciding or prediction about the data. In PdM problems these inputs are the different features represented by the data collected. Each value is fed into one of the neurons in the input layer.
– Outputs*:* Prediction generated by the NN. Each output is generated by one of the neurons in the input layer. This outcome can be a value or several values. ANN can work in both classification and regression problems. For classification the value or values are between 0 and 1 while, for regression it can be any number.
– Training Set: consider in a set of input for which the correct outputs are known, used to train the NN.
– Activation Function*:* this is one of the most important concepts in NNs which can be described as a mathematical equation that determines the output of each neuron. They determine the output of the model, its accuracy and the computational efficiency of training the model. These functions determine whether the neuron should be activated or not, based on the relevance of the input as well as also hep to normalize the output of each neuron to a determined range. Most of actual NNs use non-linear activation functions as can help to learn complex data, compute and learn almost any type of function. The most common activation function used are…
    - Sigmoid function: converts the output of the neuron (z) to a number between 0 and 1.
    $$Sigmoid = \frac{1}{1 - e^{-z}}$$
    - Rectified Linear Unit (ReLU): which outputs the highest of either zero or z.
    $$ReLU = \max(0, w * x + b) = \max(0, z)$$
– Weight Space*:* Each neuron is given a numeric weight which together with the activation function define each neuron´s output. These weights express the importance or relevance of a change in a given feature. NNs are trained to find the optimal value of such weights which minimizes the error between the output value and the real value, generating the most accurate prediction. Next equation represents the output of a neuron which *input x*, *weight w* and *bias b*.
$$predicted = w * x + b$$
– Bias*:* or also called *bias neuron* is part of the NN. In each layer a bias neuron is added with value of 1. Such neuron is multiplied by a value b which is added to the output prediction as described in the later equation. This value make it possible to move the activation function left, right, up or down on the number graph. When setting up a NN, all weights and biases are given an initial value which will be discussed in advance later.

– Forward pass*:* is the action of taking the set of inputs, passing them through the network and allowing each neuron to react to a fraction of the input data set. Neuron generates their outputs, apply the activation functions and pass the outcomes to the next layer.
– Error function: represents the error obtained by comparing outputs with a desired result to modify weights with a specific training algorithm. The training performance is evaluated using the following performance measure, Mean Square Error or average sum of squared error represented in the next equation. Also called Cost or Loss function when training the model, the aim is to minimize it changing the values of w and b:

$$J(w, b) = C(w, b) = \frac{1}{2m} \sum_{m} (y - predicted)^2$$

- Backpropagation: also called *Gradient Descent* is an algorithm commonly used to train the NN which helps to adjust the weights of the neurons so that the error functions will be minimized. Next figure shoes a convex cost function, where there is a global minimum, but there are other scenarios which are non-convex or local minimums. To find the values of w and b some equations as the partial derivates can be used to find the minimized error for those values.

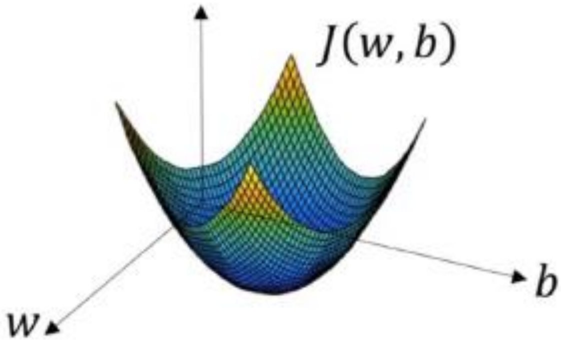


Figure 42. Backpropagation example for a cost function (Mourri, 2017)

- Bias and Variance: the meaning of bias here is different that the mentioned before. Here bias is a statistical concept which reflects how well the model fits the training set. A high bias means that the NN is not able to generate correct predictions even for the examples trained.
  Variance is related to how well the model fits the examples used in the validation set. A high value of variance means that the NN is not able to correctly predict outcomes for new features have not seen before.
- Overfitting and Underfitting: Related to the concepts mentioned just before, these problems appeared in NNs.
  Overfitting is based on a good accuracy when performing the training set, but it is not able to generalize its predictions to additional new samples, which is featured by a low bias and high variance of the data set.
  Underfitting is related to a low accuracy of the prediction in both the training set and validation set. It is featured by high bias and high variance.
- Hyperparameters: External parameters set by the operator of the NN. Hyperparameters have a huge impact on the accuracy of the predictions of the networks and there are different methods to discover the which fit better in the models. The common hyperparameters in NNs are the following:
  - Number of hidden layers*:* generally, more hidden layers improve accuracy until a certain limit which differs depending on the model.
  - Activation function: which impacts the network´s ability to converge and learn for different ranges of input data as well as its training speed.
  - Weights and bias initialization: The initialization of values can greatly affect the process of gradient decent and thus the training process of the NN. To initialize W with zeros will cause a problem called symmetry breaking problem, resulting in the network being equivalent to a linear model. A common procedure is to initialize W with random

numbers and then multiplying it with a low number such as 0.01. *b* is usually initialized with zeros, as they do no cause the symmetry breaking problem. The values of initialization are dependent on the system and activation function (Mourri, 2017). An example is the typical sigmoid function where higher values of Z, in other words, higher values of W and b, will result in a slow gradient descent and thus slow learning.

- Learning rate: depending on its value the speed of how the backpropagation algorithm performs gradient descent will be affected.
- Epoch, iterations and batch size: such parameters are related to the rate at which samples are fed to the model for training. AN is a group of samples which passed forward the networks and then run through backpropagation to determine the weights. Sometimes the epoch cannot be run all together as the high number of samples and it is split into batches, so the epoch is run in two or more interactions. The number of epochs and batches per epoch affect the accuracy of the models.

4.3.2.7 Artificial Neural Networks types.

There are many different varieties of ANN and many different techniques that can be used. Each version boasts certain benefits and/or drawbacks. There are multiple variants, but amongst all, the most known are:

- Perceptron
- Back- Propagation
- Hopfield Network
- Radial Basis Function Network (RBFN)

In the field of maintenance, multiple variants of NNs have been applied, such as back-propagation, probabilistic, generalized regression, or adaptive neuro-fuzzy inference system.

- **Probabilistic Neural Networks.**

Probabilistic NNs is a feedforward NN form of Radial Basis Function, widely used in classification and pattern recognition problems. These networks use a Kernel estimation and a non-parametric function to approximate the probability distribution function of each class and then the class of new data inputs is estimated using Baye´s rule. These types of neurons only differs in the replacement of the sigmoid activation function with an index function (Wang, Zhang & Zhong. 2008, pp. 1981–1993). Results of Probabilistic NNs can approximate the nonlinear decision boundary of the Bayes optimal decision surface. These networks do not need the iterative training procedure. They converge fast, even with a low number of training samples (Wang et al. 2008, pp. 1981–1993). Therefore, the training time and classification time are also acceptable, even for high-dimension input vectors.

The architecture of a Probabilistic NN is shown in the next figure:

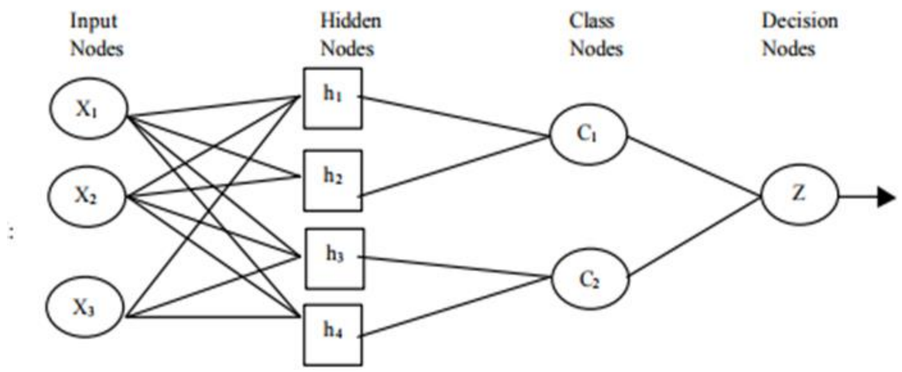


Figure 43. Probabilistic NN architecture (El Emary & Ramakrishnan ,2008)

(El Emary & Ramakrishnan ,2008, pp. 772-780), carry out a deep description of how these networks work, explaining in detail its structure and pointing the most common applications where to use Probabilistic NNs.

This type of NN belongs to the group of classifying ML techniques and they are not suitable for regression problems as Residual Useful Lifetime estimations. Improvements of the probabilistic NN, over a standard back propagation NN, have made the probabilistic NN suitable for a failure type detection and PdM system, where a ML labelling technique is used.

These techniques have been very successful in maintenance applications, since they have interesting properties for both diagnosis and prognosis (Pooria & Jazayeri-Rad, 2014, pp.21-32).

– They can handle non-linear and indeterminate processes for which the model is not known, learning the diagnosis from the information of the training data.
– They are tolerant to noise.

Their main drawbacks are (Jian-Da & Liu, 2009, pp. 4278–4286):

– Convergence to local minimums, without achieving the global minimum.
– Need for multiple iterations for good learning.
– Decision of the size of the recognition window

- **Generalized Regression Neural Networks.**

Generalized Regression Neural Network (GRNN) proposed by (Specht, 1991, pp.568–576) does not require an iterative training procedure as in back propagation method. This method is based on the approximation of any arbitrary function between input and output vectors, drawing the function estimate directly from the training data. Moreover, such technique is consistent, as the training set size becomes large, the estimation error approaches zero, with only mild restrictions on the function.

GRNN is used for estimating continuous variables, as in standard regression techniques. The regression of a dependent variable y on an independent x estimates the most probable value for y, given x and a training set. The regression method will produce the estimated value of y which minimizes the mean-squared error (Jahnke, 2015).

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (\overline{yi} - y_i)^2$$

Where n is the number of examples, $\overline{y\iota}$ is the i^th^ estimated value and y_i is the i^th^ real value.

The difference between this NN technique and standard regression ML techniques relies on the characteristics of the data used. If the data is complex sometimes, standard models cannot model it accurately and NNs can create a more complex mathematical structure to represent the data.

Such NN type, is a method for estimating the joint probability density function of x and y, given only a training set. Because the such function is derived from the data with no preconceptions about its form, the system is perfectly general.

Being f(x,y) the known joint continuous probability density function of random vector x, and y is a scalar random variable, the regression of y on X is given by:

$$E[y|X] = \frac{\int_{-\infty}^{+\infty} yf(X,y)dy}{\int_{-\infty}^{+\infty} f(X,y)dy}$$

When the density $f(X,y)$) is not known, it must usually be estimated from a sample of observations of x and y which is described in (Kerem Cigizoglu &Alp, 2006, pp. 63-68).

The structure of a GRNN is the same as the Probabilistic Network with the input, pattern, summation and output layers. The only difference between such methodologies relies on the type of kernel functions used.

- **Adaptative Neuro-Fuzzy inference system.**

Adaptive neuro-fuzzy inference system is one type of artificial NN, which is based on the Takagi and Sugeno Fuzzi Inference System (Petković, & Ćojbašič, 2013, pp. 191-195). This method comprises NNs and Fuzzy Inference Systems, which means that it has the benefits of both methods in one framework. The system has the ability of learning by which nonlinear functions can be approximated (Petković, Pavlović & Ž. Ćojbašić, 2016, pp. 215-221) (Zhou, Wu & Chan 2011, pp. 2066-2073).

An adaptive Neuro-Fuzzy Inference System optimizes a final fuzzy inference system, which uses fuzzy set theory (Zadeh. 1965. pp. 338-353) to map an input vector to an output, with the ANN training. It maps the input vector by inputting membership functions and associated parameters, prior to mapping the normalized results to outputs by using output membership functions. An adaptive neuro-fuzzy inference system can be used for classification and regression problems.

Such technique refines the fuzzy IF-THEN-ELSE rules and membership functions, according to labelled historical data. (Jang. 1993. pp. 66-685) showed that even if domain expert knowledge is not available, it is possible to set up membership functions, and then employ the ANN training process on labelled historical data, to generate a set of fuzzy IF-THEN ELSE rules. Two examples of fuzzy IF-THEN-ELSE rules, based upon a first-order Sugeno's model, show the adaptive neuro-fuzzy inference system architecture:

Rule 1: IF (x_1 is A_1) and (x_2 is B_1) THEN ($y_1=p_1x_1+q_1x_2+r_1$)

Rule 2: IF (x_1 is A_2) and (x_2 is B_2) THEN ($y_2=p_2x_1+q_2x_2+r_2$)

Here, x_1 and x_2 are the inputs feature vector and A_i and B_i are the fuzzy sets, y_1 and y_2 are outputs, while p_i, q_i and r_i are the parameters defined by the training process. Next figure shows the neuron network structure for the given rules.

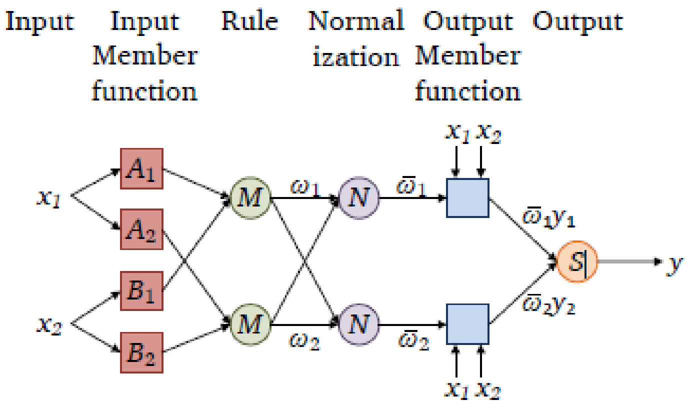


Figure 44. Adaptative Neuro-Fuzzy inference system for the given rules (Lei, He, and Zi, 2008)

(Jang, 1993, pp. 665-685) explain the layers as following:

1. Input member function layer.

The node in this layer is adaptive, and the layer performs fuzzification. In this layer, x is the input to the node i and A_i is the linguistic label. This implies that the outputs of this layer represent the membership grade of the inputs. There are various forms of membership functions to be used and parameters in this layer are referred to as premise parameters.

Next function is often chosen to be bell-shaped with maximum equal to 1 and minimum equal to 0, as below;

$$u_{A_i}(x) = \frac{1}{1 + |\frac{x - c_i}{a_i}|2b_i}$$

Here, (a_i, b_i, c_i) is the set of parameters. If these parameters change, the bell-shaped functions change too. Parameters in this layer are referred to as *premise parameters*.

2. Rule layer.

These nodes are simple multipliers. The inputs are the outputs of the layer explained before, the outputs of this layer represent the fuzzy strengths or the firing strength of a rule.

$$w_i = u_{A_i}(x) \times u_{B_i}(y); i = 1,2$$

3. Normalization layer.

Every node in this layer is a fixed node "Norm". The i_{th} node calculates the ratio of the i_{th} rule's firing strength to the sum of all rules' firing strength. Outputs are called normalized firing strengths.

$$\overline{w_i} = \frac{w_i}{w_1 + w_2}; i = 1,2$$

4. Output member function layer.

Every node "i" in this layer is a square node with a node function. The output of this layer is given by:

$$\overline{w_i} y_i$$

Where

$$y_i = p_i x_1 + q_i x_2 + r_i$$

and (p_i, q_i, r_i) is the parameter set of this node. These are referred to as *consequent parameters*.

5. Output layer.

This node performs the summation of all incoming final signals.

The premise parameters and the consequent parameters are calculated during the training process. The adaptive neuro-fuzzy inference system can be useful for failure type detection and PdM, especially in combination with the knowledge of a domain expert. A good definition of the IF-THEN ELSE rules can significantly reduce the labelled historical data needed. The simple structure also allows the fast computation of results (Jahnke,2015).

4.3.2.8 Deep Learning.

They are a subclass of NNs, although they are usually categorized separated by the recent explosion of this field. They build larger and more complex NNs.

- Deep Boltzmann Machine (DBM)
- Deep Belief Networks (DBN)
- Convolutional NN (CNN)
- Recurrent NN (RNN)
- Stacked Auto-Encoders

Thus, "Deep learning" is an evolution of NNs with multiple hidden layers (Schmidhuber, 2015, pp. 85-117). Some of its main implementations are:

- Fully connected NNs (FCNs), whose history goes back to the 80s, when the backpropagation algorithm was proposed (Hecht-Nielsen, 1989, pp. 593–605).
- Convolutional NNs (CNNs), which reduce the number of total parameters of each layer using kernels, demonstrating their effectiveness in image recognition.
- Recurrent NNs (RNNs), which allow cyclical connections, so they keep information passed in memory. In addition, to facilitate their training, the Long-short term memory (LSTM) (Hochreiter & Schmidhuber, 1997) was proposed, which appropriately discards the information. These networks have proven effective in speech recognition and natural language processing.

Another of the variants of NNs is the "autoencoder" (Aldrich & Auret, 2013), which is characterized by having the same number of neurons in the output layer as in the input layer, while the hidden layers are composed of a smaller number. In the training phase, it is usually fed with each sample both at the input and at the output. In this way, the network is forced to decrease the dimensions in the hidden layers, compressing the data to decompress them again and reconstruct the sample in the output layer. Autoencoders have multiple applications, such as reduction of dimensions, learning of generative models or detection of anomalies.

4.3.2.9 Machine Learning in Predictive Maintenance.

All the techniques described before in ML are classified into Parametric or Non- Parametric approaches. Parametric algorithms make assumptions about the distribution of data which feature them to be of low computational effort and very fast when being applied. However, the accuracy of such techniques defines the accuracy of the result, which become them not being so used for failure type detection and PdM scenarios where data are usually non-linear in the lifetime of real systems and therefore, it is hard to make assumptions about the distributions.

On the other hand, Non-Parametric approaches assumes that inputs close to one another must have the same result and, this “close to another” is that makes the difference between all the Non-Parametric techniques.

In his thesis, (Jahnke, 2015) provides an overview of the ML approaches used in recent literature of failure type detection and PdM. Such approaches were introduced before in this thesis, discussing their advantages and drawbacks. For calculating the usage of ML approaches in recent literature the author choose 32 sources among which are articles, books, scientific articles… The result of it is showed in the next table in terms of percentage, being the most common approaches those that have a high value:

Table 4. Applications of Machine Learning Approaches in PdM´s recent literature

ML Approach	Application Percentage
NNs	56%
Support Vector Machines	34%
Decision Trees and Random Forest	19%
K Nearest Neighbour	9%
Markov Models	9%
Bayesian Networks	3%
Linear Discriminant Analysis	3%

Scenarios differ considerably from each other in terms of computational resources, memory resources, availability of data and quality of data, thus, it is no possible, from an analytic point of view, to decide on what ML approach is suitable for a given failure type detection and PdM problem. The reason is that the accuracy in these areas is too important and needs the correct analytic choice of a ML technique (Jahnke, 2015).

5 AI applications in Predictive Maintenance

5.1 Artificial Neural Network applications

As mentioned before ANN currently are the most common techniques in the field of failure type detection and PdM. These techniques can cover noisy or incomplete data or are even suitable to predict unknown data based on defined parameters (Cheng & Sutariya, 2012).

ANN show promising results as a robust tool for evaluation collected data to support PdM activities. There exist a lot of papers focused on application of ANN in maintenance. Mainly multi-layer Perceptron (MLP) are used for fault diagnosis of bearings, induction motors, non-destructive evaluation of check valve performance and degradation and in robotic systems (Meireles, Almeida & Simoes, 2003, pp. 585-601)

In their book (Křenek. et al. 2016, pp.75-86) point some of the ANN application in the field of PdM which are the following.

5.1.1 Mechanical damage and crack detection.

Recent literature has shown that ANN are very common in the field of mechanical detection.

5.1.1.1 Cracks identification in large crankshafts.

Stiffness and damping data based on engine load of large diesel engine crankshafts were used for early detection of cracks by (Villanueva, Espadafor, Cruz-Peragon & García, 2011, pp. 3168-3185) whose method was capable to evaluate risk of forming cracks using Radial Basis Function NN as well as suggesting an implementation possibility of on-line monitoring. Radial Basis Function was used as a classification method which was able to identify cracks and their depth in crankshaft which led to an improvement in the PdM strategy of diesel power plants.

The method used was based on the development of several system models of the engine linked to a NN classifier. To make the classification they used a net architecture composed of a hidden layer with Gaussian activation functions. The inputs of the PdM model were the percentage of current load torque and the angular speed. Such inputs were scaled in the range (0,1) so that the overall variance in the data set was maximised. The reason of this is that NNs have better results if the values for associated vectors are in the same domain (Gill, Murray & Wright, 1997).

The output layer of the network consisted of a competitive layer that classified the crack damage derived from growth area percentage in relation to the journal section, which was divided in three area percentage intervals. Such consideration allowed to classify failure damage into the following scale: 1 (minimum risk), 2 (medium risk) and 3(crack risk), which were the only possible output values of the network. The network structure were established with 288 modelled conditions (samples) and from them 9 conditions were used to train the net (training set), the rest of operation conditions were used to validate the net. The trained net assured a high accuracy of the risk level between model results of the training patterns (assuming the role of real values) and the network output. Depending on the risk level provided by the network, data was marked with ‘x’ (level1), ‘o’ (level2) or ‘square’ (level3) as the next figure shows.

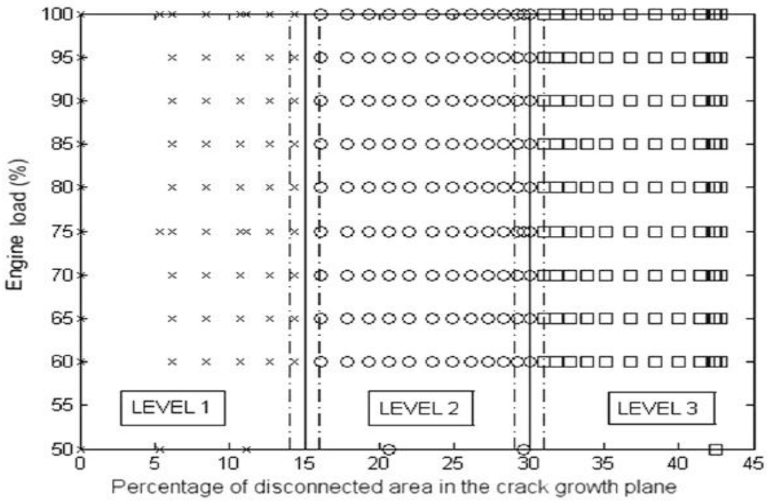


Figure 45. Classification results (Villanueva, Espadafor, Cruz-Peragon & García, 2011)

Finally, the result demonstrated that more than 96% of the validation patterns belonged to its corresponding risk level. The incorrect prediction were 8 situations where percentage growth area took a value of 30,1%, classifying them into the second risk level, while the real situations corresponded with the maximum level (Villanueva, Espadafor, Cruz-Peragon & García, 2011, pp. 3168-3185).

5.1.1.2 Incipient bearing fault diagnosis using DWT for feature extraction.

(Castejón, Lara & García-Prada. 2009, pp. 289-299) applied ANN beside the multiresolution analysis to diagnose damage of rolling bearings which are a frequent cause of machinery breakdowns and added radial load data to their model. The data that featured the crankshaft rotation was collected by sensors as accelerometers and photo tachometers. The simulations were done with the new bearing and with bearings combinating artificial damage of inner race, outer race and a ball. The extracted features from the Daubechies Wavelet Transform (DWT*)* were used as inputs in a NN for classification purposes.

Two important phases were carried out to implement the fault diagnosis process: the first was the signal processing for feature extraction and noise diminishing, and the second phase consisted of the signal classification based on the characteristics obtained in the previous step. For fault classification, Multi-layer Perceptron NN was successfully used. Such network was trained with the extracted features from the (DWT), training data was set of 75% including the validation set, while the test set represented the 25% of all data.

Network´s architecture was composed by an input layer of 18 neurons, a hidden layer with 10, 20 and 30 neurons with the aim of choosing the best option, and the output layer had 4 neurons. The transfer function used was the hyperbolic tangent sigmoid. Such structure was the most popular one and proved to be able to learn complicated functions (Chow, 1997).

Four sets of data were obtained from the experimental system: under normal conditions; inner race faults; outer race faults and ball fault. A total of 196 bearings were tested, 49 per each condition. To train the model, back-propagation momentum method was used, and the topping criteria was based in the minimum error reached, in that case the 5%. The training was limited to 25000 epochs and the initial weights of the network were random values between [-0.1, 0.1] (Castejón, Lara & García-Prada. 2009, pp. 289-299).

Finally, the results showed that normal condition and ball fault condition were easily diagnosed, and they did not show any dependence with the number of hidden neurons. However, outer race fault was diagnosed with less success rate, and worked better with higher number of hidden neurons Moreover, inner race fault could not be classified at good success rate, but also improved its behaviour at higher complexity (Castejón, Lara & García-Prada. 2009, pp. 289-299). For fault classification, multi-layer perceptron with the conformance of 80% was successfully used. Next figure shows the performance percentage with different number of hidden layers.

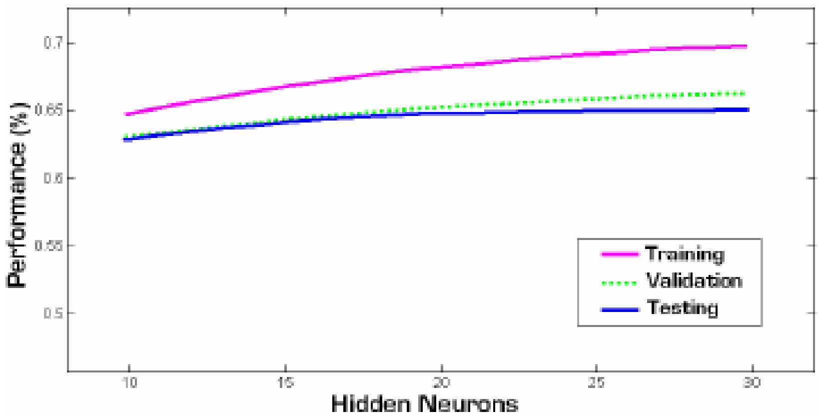


Figure 46. Multilayer Perceptron performance (Castejón, Lara & García-Prada, 2009)

5.1.1.3 Other examples.

(Saxena & Saad 2015, pp. 441-454) used combination of accelerometers and an acoustic emission sensor for the diagnosis of mechanical bearings. In the investigation, they used Genetic Algorithms (GA) for identifying near optimal design parameters of diagnostic systems as could be described as a non-conventional non-linear search algorithm able to obtain fast results (Saxena & Saad 2015, pp. 441-454) and then an ANN was used for fault classification. Three accelerometers and one acoustic emission sensor were employed to get the signal from all eight faulty bearings and one bearing without any defect. Features such as mean, variance, standard deviation and kurtosis (normalized fourth central moment) were the most common features employed for rotating mechanical components

For the Network´s structure, the best results were obtained with 7 hidden nodes and with 10 input nodes (the most important features). The network was trained for 50-100 epochs before using the test data to measure its performance and a resilient back-propagation training algorithm showed to work faster on larger neurons, which was used (Saxena & Saad 2015, pp. 441-454). A batch training strategy was employed with tan-sigmoid activation functions for the input and hidden layers and a pure-linear activation function for the output layer. Finally, the conformity rate of the measurement was higher than 95%.

(Pacheco & Pinto, 2013) used for bearing fault detection a method of acoustic measurement with conformity rate 95% and higher. (Tian, 2012, pp.227-237) suggests that the late in the bearing life is the prediction accuracy more important as the aging could affect decision making about whether the bearing should be replaced or not and use for computations Feed Forward ANNs.

5.1.2 Early detection of faulty electrical devices.

Infrared thermography in PdM is used for detection of heat distribution in the machine parts. This method profits from the phenomenon that every object emits infrared radiation. In industrial maintenance, the application of infrared thermography, is suitable for example for detection of heat losses in detection of heat of the electrical equipment, transformer load, etc (Al-Kassir, Fernandez, Tinaut & Castro 2005, pp. 183-190) (Meola & Carlomagno, 2004).

Thermal snapshot of machine electrical control cabinet can quickly and easily display invisible temperature irregularities of electrical devices such as circuit breakers, transformers, fuses, relays etc. Abnormal heat concentration on these devices is caused by faulty, oxidised or loosen contacts, damaged or cracked insulations or broken surge arrestors (Lizák, & Kolcun, 2008, pp. 60-63).

5.1.2.1 Infrared thermography for Predictive Maintenance in electrical equipment.

(Huda & Taib, 2013, pp. 220-227) used the ANN Technique Multi-Layer Perceptron with one hidden layer and 5-fold cross validation method for the evaluation of temperature differences from ambient temperature of a component with the similarly expected temperature value. A total of 368 thermal images of circuit breakers operating at the same loading condition having 500 hotspots were captured with the infrared camera. Such temperatures were classified in three classes: no defect with temperature difference less than 5 °C; probable deficiency with difference above 5 °C and a defect with major discrepancy with temperature difference above 15 °C.

Firstly, the original thermal image was converted into greyscale image and the regions of interest for both the references and hotpots were selected manually in a rectangular shape. Next figure shows such image where white or brighter region indicates a defect.

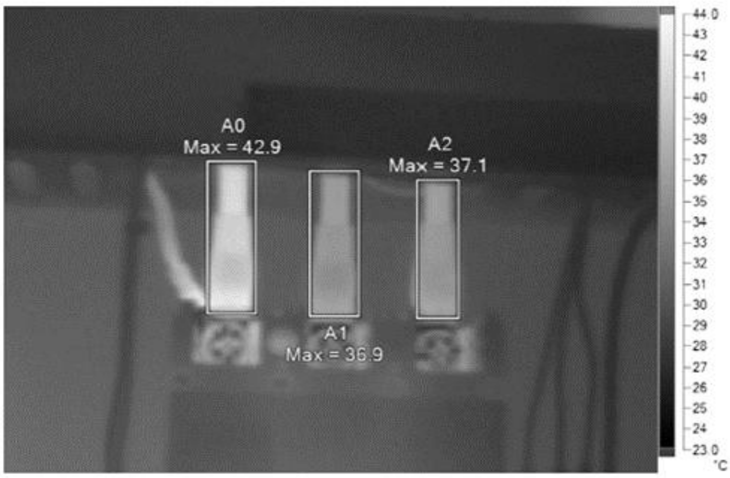


Figure 47. Grey image of thermogram (Huda & Taib, 2013)

Fifteen features were evaluated. The two first order features were kurtosis and standard deviation. The second order features were entropy, contrast, homogeneity and energy. On the other hand, a total of nine mutual features were extracted. In the case of first order, were average, grey intensity entropy, highest grey value, kurtosis, skewness, standard deviation, and variance. From second order, were entropy and energy.

Discriminant analysis was used to select optimum features, determine the performance of each class and to calculate the overall performance. The univariate and multivariate tests were applied to analyse the performance of features. After performing such tests, the optimum number of features was 10 which were the most suitable features to use during the formation of discriminant function. Apart from determining the suitable features, the discriminant analysis was also able to predict the linear regression model of each class taking the account of classification function coefficients and to classify the thermal class (i.e. determine the number of correctly classified data, the number of misclassified data and the accuracy for each class) (Huda & Taib, 2013, pp. 220-227). The overall data were 500; each data had 10 optimum features. The classification accuracy was 81.1% for 'defect' class (258 thermal conditions correctly classified out of 318) and 84.6% for 'no defect' (154 out of 182). The overall accuracy was 82,4 %.

Then, Multi-layered perceptron network was used. Ten input nodes and one output node were used. The target outputs were classified into two categories, 'defect' or 'no defect'. The sigmoid activation function was used, and the weights and threshold were unidentified and chosen for reducing the prediction error. The 500 samples were divided into training and testing dataset

This procedure advances from the fact, that it is robust against the differences of subject emissivity on measuring its temperature. Normalisation of the images was done by conversion to greyscale bit-maps where focused regions were selected for detailed analysis. The 5-fold cross validation analysis method was used (Schaffer, 1993, pp.135-143), where entire data was partitioned into 5 segments or folds. Five iterations of training and testing were performed. In each iteration, a single fold data was selected for testing and the remaining four folds were used for training the network. Finally, the average value of each fold performance was assumed as the network performance.

The network was trained varying hidden nodes from 1 to 20, with 300 epochs and a learning rate of 0,005. For analysing the network, they used the accuracy, specificity and sensitivity indicators, obtaining values in the training phase of 88.20%, 89.789% and 85.418% for each of them. Similarly, in the testing phase, the same network provided accuracy 80.40%, sensitivity 83.98% and specificity 75.29%. Therefore, it was stated that the MLP network trained using Levenberg Marquart algorithm and the proposed 10 features had high capacity to detect the defect in equipment (Huda & Taib, 2013, pp. 220-227).

5.1.3 Fault detection on pneumatic systems.

Pneumatic systems consisting from various pneumatic cylinders which are used as a cheap and reliable instrument for machine driven parts where there is no requirement for precise positioning (Křenek. Et al. 2016, pp.75-86).

(Demetgu, Tansel & Taskin 2009, pp. 10512-10519) focused on the system of the pneumatic cylinder and air valves block. In the study, the operation of a pneumatic system was monitored by using eight sensors. Two ANNs (Adaptive Resonance Theory 2 (ART2) (Carpenter & Grossberg, 1987), and Back propagation (Rumelhart, Hilton, & Williams, 1986) which were used to interpret the encoded data of the sensors.

To represent the characteristics of the system, the sensory signals were encoded by selecting their most descriptive features. The 25 most descriptive features of the sensory signals were calculated and used for each case. The pneumatic system was operated at the normal conditions and at 11 different faulty conditions. The normal and faulty modes were detected by using the signals of eight transducers including three pressure sensors, a linear potentiometer and four P/E (pneumatic to electric) switches. Most of the sensors had unique patterns for the problems except the pressure sensor of the main system and the optic sensor. Almost the signal of each one of these sensors was informative enough to detect the defects (Demetgu et al. 2009, pp. 10512-10519). Since the sensory data was almost identical, the training and testing cases were almost the same too.

When the data of the sensors or encoded parameters were identical or extremely close to each other, as here, using data from multiple tests did not improve the quality of the training and might even be considered as presenting duplicate cases. Therefore, in this study, to fill the input variable space better, synthetic cases were generated by increasing and decreasing the encoded parameters by 2%, 4%, 6%, and 8%. This process enabled to work with nine times more data than the original cases and to train the networks within a larger band which covers the ±8% of the input variable space (Demetgu et al. 2009, pp. 10512-10519). Performance of the ART2 was studied first without considering the generated synthetic cases. Later, synthetic cases were also considered to evaluate the robustness of the ANN when the encoded parameters deviates ±8% from the experimental ones.

Although the fault modes used in this study were set to 11, authors recommend reducing the number of fault modes below 5 to improve NN´s classification reliability due to experience of misidentification of some faults (Křenek. et al. 2016, pp.75-86).

The generated synthetic data was essential for the training and testing of the Back-Propagation type ANNs with the extremely repetitive data. The network had 12 inputs and one output. The value of the output was determined to be 1 for perfect case. For the faults the output was selected as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.1.

15 nodes were used at the hidden layer after several tests to obtain a compact network with acceptable accuracy which was tested by using 60 cases, after it was trained with 48 cases. For training the network, the encoded values of the original data and the artificial cases were generated with +4%, -4% and +8% deviations. Then, +2%, +6%, 8%, -2% and -6% deviations were used for testing the network (Demetgu, Tansel & Taskin 2009, pp. 10512-10519).

The estimation accuracy of the Back Propagations network was excellent for the training cases. The perfect operating condition and every fault were accurately identified without any uncertainty. The performance was also excellent when the network was tested on the synthetic data with +2% deviations from the encoded experimental data. For 8% and -6% deviations, the perfect and faulty operation modes were always accurately identified. However, at these extreme values the Bp started to confuse some of the faults with others (Demetgu, Tansel & Taskin 2009, pp. 10512-10519).

Next figure shows the results at the training and testing phase with different fault conditions.

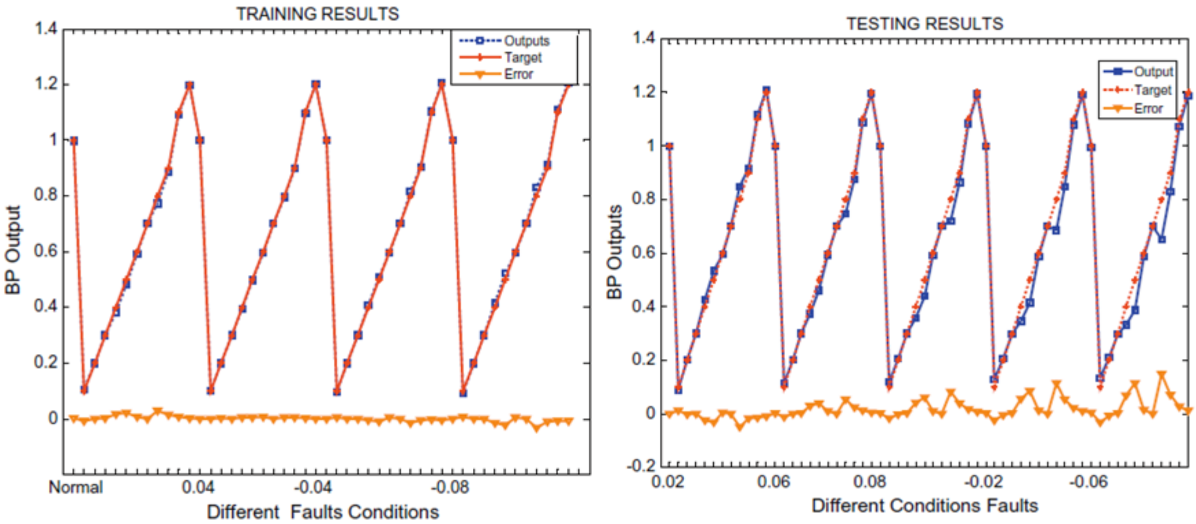


Figure 48. Training and testing results with different fault conditions (Demetgu, Tansel & Taskin, 2009)

The performance of the ANN was excellent with less than 1% average estimation error (respect to the range) for the training cases. For the test data, the average estimation error increased to 2.1%. The perfect and faulty modes were distinguished without the slightest uncertainty for the training and test sets. At the test cases, some of the 11 possible faults were misidentified in less than 10 cases which is an excellent level of performance for a Back-Propagation network (Demetgu, Tansel & Taskin 2009, pp. 10512-10519).

5.1.4 Robotic manipulator monitoring.

Robotic manipulators are commonly used nowadays for precise positioning of manufactured parts during transfer to the following process step. These robots have numerous joints with rotational axis for wide movement and positioning flexibility. Due to complicated design and implementation in fully automated process, these must be well maintained and monitored mainly for early risk of fault analysis (Křenek. Et al. 2016, pp.75-86).

Recently, different fault detection, fault diagnoses, fault detection and isolation, fault identification, fault-tolerance... techniques for robot manipulators have been proposed in the literature (Eski, Erkaya, Setrac & Yildirim, 2011, pp. 115-123). Some of the important survey papers have been presented using NN techniques for robot manipulators. For example, to stabilize the reinforcement-learning environment, (Yasuda, Ohkura & Ueda, 2006, pp. 301-311) used a NN for predicting the average of the other robots' postures at the next time step. In another investigation, Notash and Huang presented methodologies for the design of fault tolerant parallel manipulators based on the failure analysis of manipulators and optimum fault tolerant configuration of each class identified (Notash & Huang 2003, pp. 85-101).

(Eski, Erkaya, Setrac & Yildirim, 2011, pp. 115-123) presented an experimental investigation on a robot manipulator, using NN for analysing the vibration condition on joints. Firstly, robot manipulator's joints were tested with prescribed of trajectory end-effectors for the different joint's speeds. Furthermore, noise and vibration of each joint were measured. And then, the related parameters were tested with a NN predictor to predict servicing period. To find robust and adaptive network´s structure, two types of neural predictors were employed, in the investigation they deployed Radial Basis NN (RBNN) and a Kohonen Self-Organising Maps NN (SOMNN) for analysing of joint accelerometer data and compared performance of the two methods with the approach of vibration prediction on industrial robots. Next figure shows the structure of both networks.

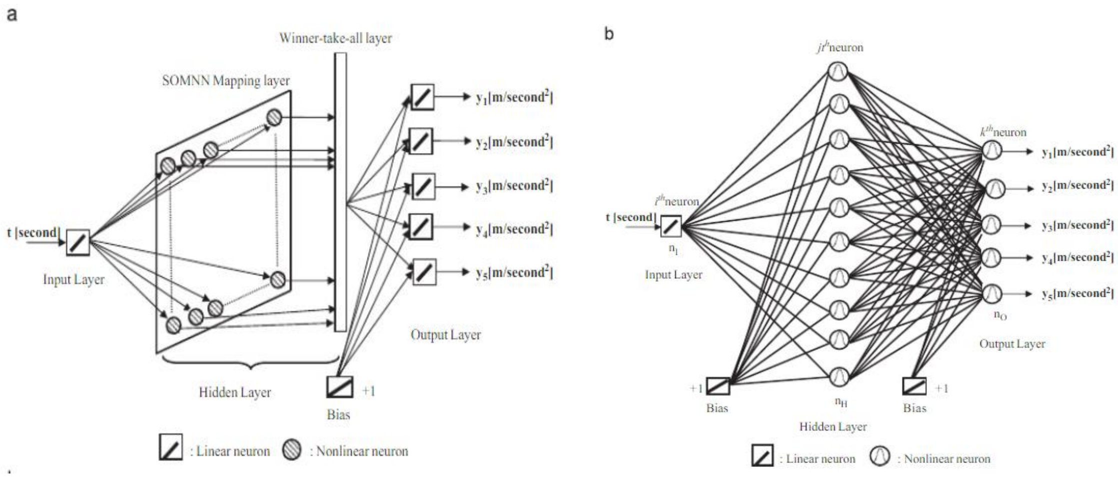


Figure 49. Schematic representations of the SOMNN analyser and the RBNN analyser (Eski, Erkaya, Setrac, & Yildirim, 2011)

Experimental and simulation investigation on robot manipulator's fault detection was carried out in two stages. First stage was experimental measurements on robot manipulator's joints. Such process consists of an intelligent data acquisition, 5 accelerometers, a microphone and a PC. Then, accelerations of welding robot, which had six degrees of freedom, were collected and analysed.

On the second step, the measured experimental accelerations values were used as desired signals of NNs, which were tested to predict exact signals of the two types of learning algorithms. In this step the two algorithms were used and compared. Experimental and simulation results were obtained, by considering the 1 rpm running speed or the desired trajectory. With the SOMM algorithm, there were differences between the experiment and the algorithm result. On the other hand, the RBNN approach was robust stable to analyse the vibration parameters of such welding robot manipulator joints for different prescribed and joint speeds. Noise variations for 1 and 2 rpm were evaluated together and the main factor was not the robot.

The results of SOMNN approach give poor performance. Again, the RBNN structure was used to predict acceleration of the welding robot manipulator joints with 2 rpm running speed. The proposed RBNN approach had superior performance for predicting accelerations variation of the welding robot manipulator joints for different running speeds. Next figure shows a table with the structural, training parameters used in both networks as well as the (RMSEs) from both. The structure and explanation of the network is not explained in this thesis, but for more information about it (Eski et al. 2011, pp. 115-123) developed in detail.

Running speed of end-effectors	NN type	η	γ	N	n_I	n_H	n_O	RMSEs
1 rpm	SOMNN	0.4	0.5	400,000,000	1	10	5	0.0600
1 rpm	RBNN	0.4	0.5	400,000,000	1	10	5	**0.0004**
2 rpm	SOMNN	0.4	0.5	400,000,000	1	10	5	0.0439
2 rpm	RBNN	0.4	0.5	400,000,000	1	10	5	**0.0004**

Figure 50. Table with the structural and training parameters of the feedforward NNs with two learning algorithms (Eski, et al. 2011)

In Kohonen Self-Organising Maps the Root Mean Square Error from 0,04 till 0,06. On the other hand, Radial Basis function demonstrated superior performance on accelerations variation descriptions for different running speeds with a Root Mean Square Error below 4×10^{-4}.

The major advantage of an RBNN which uses Gaussian function is that gives a response that drops off rapidly as the distance between the hidden unit and the input vector increases and is symmetrical about the radial axis. It has been scientifically proved that the RBNN structure gives the best approach rather than other structure (Eski, et al. 2011, pp. 115-123).

5.2 Application´s conclusion

From the applications discovered, Multilayer Perceptron Artificial Neuron Networks method is the most used as amongst other advantages, to its simplicity. Such technique is suitable for classification of faults on the machines condition and in the case of being used for fault detection caused by mechanical damage, it would be convenient to use boundary sample in the training phase.

Next table retrieved from (Křenek. et al. 2016, pp.75-86) shows the results of the ANN methods used to solve the problems related above.

Table 5. Overview of used ANN designs with achieved results

PdM Task	ANN design	Result
Crankshaft crack	Radial Basis Function NN	96%
Bearing condition	Multilayer Perceptron Neuron Network	80%
Bearing condition	Genetic Algorithm	93%
Bearing condition	Multilayer Perceptron Neuron Network	95%
Electrical fault	Multilayer Perceptron Neuron Network	83%
Pneumatic system	Adaptive Resonance Theory	93%
Robotic manipulator	Kohonen Self-Organising Maps	RMSE 0,0004
Robotic manipulator	Radial Basis Function NN	RMSE 0,04

In Artificial Neuron Networks, sometimes the usage of artificially produced faults which might lead to phenomenon, where an undiscovered fault in an early stage could cause irreversible damage on other parts. This could lead to an accelerated wear of the machine if the first damage is not recognised correctly and at the right time. This approach is useful in cases where there is expected as a fault a destroyed part or machine (Křenek. et al. 2016, pp.75-86).

Moreover, such techniques are very useful in PdM to use continuous data as an output of the Artificial NN, that would for example give an estimation of Remaining Useful Life of the parts as suggested by study. If the focus of PdM is on the extensively worn machine parts, it is more helpful to use continuous data as an output of such ANNs, that would for example give an estimation of remaining useful life of the parts as suggested by study (Tian, 2012, pp.227-237).

Finally, ANN show strong potential in industrial applications, especially in PdM tasks. Due to need of well-trained person for such computations, necessity of special software and availability of sensors able to capture and systematically store the collected data it must be properly evaluated whether the use in real production will be meaningful and profitable (Křenek. et al. 2016, pp.75-86). The advantage of these methods is the potential of effective equipment failure prevention and the increase of the Overall Effectiveness Equipment performance results.

6 Conclusion

This thesis is a report of Predictive Maintenance and all-important issues related with it. First two chapters aim to be a theorical summary about such topic. First, the concept of maintenance is introduced and an overview about its development along the history from different perspectives is carried out. Then the concept of Predictive Maintenance is introduced, where its structure, tools and technologies related with it are described, later, its relation with the Industry 4.0 is introduced. The third part aims to be an implementation guideline consisted on all steps to carry out for any organisation who wants to implement a Predictive Maintenance program. Here, some important methodologies and technologies are presented, and some advises are exposed to achieve successful results. Fourth part is related with the main Artificial Intelligence approaches implemented in Predictive Maintenance, where Artificial Neural Network become important to achieve good results. An overview of the different techniques is made it, as well as, focusing on which are commonly used. Finally, some successful cases of implementing different Artificial Neural Network techniques are presented, where it is shown its highly accuracy at the time of getting the results.

Although, there is a lot of work to do, Organisations have started to implement Predictive Maintenance programmes which have showed incredible results. Thus, such Predictive Maintenance has become in a key issue to the development of the industry of the future, where predict any type of problem before it happens, will made the difference between some organisations and others. In current markets which each day are becoming more and more competitive, implementing technologies like this, will affect highly the organisation´s performance.

REFERENCES

Ackermann, J., Bartlett, A., Kaesbauer, D., Sienel, W. & Steinhauser, R. (1993). *Robust control: the parameter space approach*. Springer Science & Business Media

Ahuja, I.P.S. & Khamba, J.S. (2008) Total productive maintenance: literature review and directions. IntJ Qual Reliab Manage 25(7):709–756

Aldrich, C. & Auret, L. (2013). *Unsupervised process monitoring and fault diagnosis with machine learning methods*. London: Springer,

Aldrich, C. & Auret, L. (2013) *Unsupervised process monitoring and fault diagnosis with machine learning methods.* London: Springer

Al-Kassir, A. R., Fernandez, J., Tinaut, F. V. & Castro, F. (2005). *Thermographic study of energetic installations. Applied Thermal Engineering*, 25(2), 183-190

Alsyouf I. (2006). The role of maintenance in improving companies: productivity and profitability ', International Journal of Production Economics, pp. 70-78.

Analytics Vidhya Content Team. (2016). *A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python).* Retrieved from https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/

Arulkumaran, K., Deisenroth, M. P., Brundage, M. & Bharath, A. A. (2017). *Deep reinforcementlearning: a brief survey.* IEEE Signal Processing Magazine, 26-38.

Arulkumaran, K., Deisenroth, M.C., Brundage, M. & Bharath, A.A. (2017). *Deep Reinforcement Learning: A Brief Survey.* IEEE Signal Processing Magazine. 34(6): p. 26-38.

Bair, E. & Tibshirani, R. (2004). *Semi-supervised methods to predict patient survival from gene expression data*, PLoS Biology, vol. 2, pp. 511–522, 2004.

Bair, E., Hastie, T., Paul, D. & Tibshirani, R. (2006*). Prediction by supervised principal components*, Journal of the American Statistical Association, vol. 101, no. 473, pp. 119–137

Balasubramanian, M. & Schwartz, E. L. (2002). *The isomap algorithm and topological stability,* Science, vol. 295, no. 5552, p. 7, 2002.

Barad, Sanjay & P.V., Ramaiah & Giridhar, R & G, Krishnaiah. (2012). *Neural network approach for a combined performance and mechanical health monitoring of a gas turbine engine*. Mechanical Systems and Signal Processing. 27. 729–742. 10.1016/j.ymssp.2011.09.011.

Batuwita, R. & Palade, V. (2010). *FSVM-CIL: fuzzy support vector machines for class imbalance learning*, IEEE Transactions on Fuzzy Systems 18 (3) 558–571. doi: 10.1109/TFUZZ.2010.2042721

Bengtsson, M., Olsson, E., Funk, P. & Jackson, M. (2004). *Technical design of condition-based maintenance system—a case study using sound analysis and case-based reasoning*. In Maintenance and reliability conference

Biagetti, T. & Sciubba, E. (2004). *Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems.* 29(12):2553–2572

Bishop, C. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer Science & Business Media

Breiman, L. (1996). *Bagging predictors. Machine Learning*. 24(2), 123–140.

Breiman, L. (2001). *Random Forest. Machine Learning* 45, 5-32.

Brodley, C.E. & Friedl, M.A. (1999). *Identifying mislabeled training data*, Journal of Artificial Intelligence Research 11, 131–167.

Buchmann, A. & Koldehofe, B. (2009). *Complex event processing. it- Information Technology Methoden und innovative Anwendungen der Informatik und Informationstechnik*, 51(5):241–242

Butler, K.L. (1996). *An expert system-based framework for an incipient failure detection and predictive maintenance system.* In International Conference on Intelligent Systems Applications to Power Systems (ISAP), pages 321–326

C. Davies, R.M. Greenough, The use of information systems in fault diagnosis, in: Proceedings of the 16th

Cao, L. J., Chua, K. S., Chong, W. K., Lee, H. P. & Gu, Q. M. (2003). *A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine*, Neurocomputing, vol. 55, no. 1-2, pp. 321–336

Carpenter, G. A., & Grossberg, S. (1987). ART 2: *Self organization of stable category recognition codes for analog input patterns*. Applied Optics, 26(23), 4919–4930.

Castejón, C., Lara, O. & García-Prada, J. C. (2009). *Automated diagnosis of rolling bearings using MRA and neural networks.* Mechanical Systems and Signal Processing, (24), 289-299

Cempel, C.Z., Kasprzak, J. & Kłos, Z. (2006). Ekoin_zynieria: ku holistycznemu projektowaniu izarza˛dzaniu s´rodowiskiem, Ogólnopolska konferencja naukowa Zrównowa_zony rozwój wteorii ekonomii i praktyce, Wrocław

Chang, C.C. & Lin, C.J. (2011). *Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3):27*

Chawla, N.V., Bowyer K.W., Hall, L.O. & Kegelmeyer, W.P. (2002). *SMOTE: synthetic minority over-sampling technique*, Journal of Artificial Intelligent Research 16, 321–357. doi: 10.1613/jair.953

Cheng, F. & Sutariya, V. (2012). *Applications of artificial neural network modeling in drug discovery.* Clin Exp Pharmacol, 2(3), 1-2

Chow, M. (1997). *Methodologies of Using Neural Network and Fuzzy-Logic Technologies for Motor Incipient Fault Detection*. World Scientific. ISBN 981-02-32659.

Christer, A.H. & Wang, W. (1995). *A simple condition monitoring model for a direct monitoring process*. European Journal of Operational Research, 82, pp. 258-269

Christopher, M.B. (2006). *Pattern recognition and machine learning*, volume 4. Springer Science & Business Media

Cigizoglu, H. K. & Alp, M. (2006). *Advances in Engineering Software*, ISSN: 0965-9978, Vol: 37, Issue: 2, Page: 63-68

Clarke, P. (2018). OEE And Why It's Important. Retrieved from https://slcontrols.com/oee-and-why-its-important/

Coleman,C., Damodaran, S., Chandramouli,M. & Deuel, E. (2017) ``*Making maintenance smarter: Predictive maintenance and the digital supply network*´´ Retrieved from https://www2.deloitte.com/insights/us/en/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html

Cousineau, M. 2018. *Implementing predictive maintenance with the help of preventive maintenance.* Retrieved from https://www.fiixsoftware.com/blog/implementing-predictive-maintenance-and-preventive-maintenance/#challenges

Covell, M.M. & Richardson, J. M. (1991). *A new, efficient structure for the short-time fourier transform, with an application in code-division sonar imaging.* In International Conference on Acoustics, Speech, and Signal Processing (ICASSP), volume 3, pp. 2041–2044

Cózar, J., Puerta, J. M. &. Gámez, J. A. (2016), *An Application of Dynamic Bayesian Networks to Condition Monitoring and Fault Prediction in a Sensored System*: *A Case Study*, Volume 10, Issue 1, pp. 176 – 195

Das Chagas Moura, M., Zio, E., Lins, I. D. & Droguett, E. (2011). *Failure and reliability prediction by support vector machines regression of time series data*. Reliability Engineering & System Safety, 96(11):1527–1534

Davies, C. & Greenough, R.M. (2000). *The use of information systems in fault diagnosis*, in: Proceedings of the 16th National Conference on Manufacturing Research, University of East London, UK

Daya, M.B. & Duffuaa, S.O. (1995) Maintenance and Quality: The Missing Link, *Journal of Quality in Maintenance Engineering* 1(1), pp. 20-26.

Deland, S. (2018). *Top Obstacles to Overcome when Implementing Predictive Maintenance.* Retrieved from https://www.kdnuggets.com/2018/10/mathworks-top-obstacles-predictive-maintenance.html

Demetgul, M., Tansel, I.N. & Taskin, S. (2009). *Fault diagnosis of pneumatic systems with artificial neural network algorithms.* Experts Systems with Applications, (36), 10512-10519

Deolasee, P., Katkar, A., Panchbudhe, A., Ramamritham, K. & Shenoy, P. (2001). *Adaptive Push-Pull: Disseminating Dynamic Web Data.* In Proc. of the 2001 Intl. WWW Conference, Hong Kong, China

Dhillon, B. (2002). *Engineering Maintenance: A Modern Approach.* 10.1201/9781420031843.

Dhingra, P. (2018). *A process for implementing industrial predictive maintenance: Part II*, Retrieved from https://cloud.google.com/blog/products/data-analytics/a-process-for-implementing-industrial-predictive-maintenance-part-ii

Ding, C. & Peng, H. (2005). *Minimum redundancy feature selection from microarray gene expression data*. J. Bioinform.Comput. Biol. 2005, 3, 185–205

Donald W. Benbow, Roger W. Berger, Ahmad K. Elshennawy & H. Fred Walker, 2002, the Certified Quality Engineer Handbook by ASQ.

Dong, M. & He, D. (2007). *Hidden semi-markov model-based methodology for multi-sensor equipment health diagnosis and prognosis*. European Journal of Operational Research, 178(3):858–878

Dorugade & Kashid, D. N. (2010). *Alternative Method for Choosing Ridge Parameter for Regression*. Applied Mathematical Sciences, Vol. 4, no. 9, 447 – 456. Retrieved from http://www.m-hikari.com/ams/ams-2010/ams-9-12-2010/dorugadeAMS9-12-2010.pdf.

Dunn, S. (2003). *The fourth generation of maintenance*. Retrieved from http://www.plant-maintenance.com/articles/4th_Generation_Maintenance.pdf

Duraccio, Vincenzo & Silvestri, Alessandro & Falcone, David & Forcina, Antonio & Cerbaso, Cristina. (2014). *Maintenance critical analysis and priority index: A new model for maintenance policy*. 26th European Modeling and Simulation Symposium, EMSS 2014.

El Emary, I. M. M. & Ramakrishnan, S. (2008). *On the Application of Various Probabilistic Neural Networks in Solving Different Pattern Classification Problems*. World Applied Sciences Journal 4 (6): 772-780, 2008ISSN 1818-4952

Emerson, "How manufacturers achieve top quartile performance," WSJ Custom Studios, Retrieved from http://partners.wsj.com/emerson/unlocking-performance/how-manufacturers-can-achieve-top-quartile-performance/

Eski, I., Erkaya, S., Setrac, S. & Yildirim, S. (2011). *Fault detection on robot manipulators using artificial neural networks*. Robotics and Computer-Integrated Manufacturing, (27), 115-123

Faiz, R., & Edirisinghe, E. A. (2009). *Decision making for predictive maintenance in asset information management. Interdisciplinary Journal of Information, Knowledge, and Management*, 4, 23–36.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). *The KDD process for extracting useful knowledge from volumes of data*. Communications of ACM, 39 (11), pp. 27-34

Feelders, A., Daniels, H. & Holsheimer, M. (2000). *Methodological and practical aspects of data mining*. *Information and Management*, 37, pp. 271-281

Feigenbaum, A.V. (1991) *Total Quality Control*. 3rd Edition, McGraw-Hill, NY, USA,

Feldman S., Buechele R. & Preveden V. (2018). *Predictive maintenance –From data collection to value creation*. Retrieved from https://www.rolandberger.com/publications/publication_pdf/roland_berger_predictive_maintenance_2018.pdf

Finnin, M., Shipp, S., Gupta, N., & Lal, B. (2012). *Emerging Global Trends in Advanced Manufacturing* (p. 248). Retrieved from http://www.stormingmedia.us/61/6168/A616855.html

FMEA Ratings. http://www.suggestkeywords.com/ Zm1lYSByYXRpbmdz/. (Accessed on 29.06.2019).

Föllinger, O. & Kluwe, M. (1997). *Laplace-und Fourier-Transformation*. Elitera Berlin

Garcia, M. C., Sanz-Bobi, M. A., & del Pico, J. (2006). *SIMAP: Intelligent Sys-tem for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox. Computers in Industry*, 57(6), 552–568.

Gill, P.E., Murray, W. & Wright, M.H. (1997). *Practical Optimization*, 11th Impression, Academic Press

Gits, C. W. (2010). Structuring Maintenance control systems. International Journal of Operations & Production Management, 14(1994), 5–17.

Glossop, M. (2000). Review of hazard identification techniques.

Gomez de Leon Hijes, F.C. & Cartagena, J.J.R. (2006). *Maintenance strategy based on a multicriterion classification of equipments', Reliability Engineering and System Safety*, pp. 444-451.

Groba, C., Cech, S., Rosenthal, F. & Gössling, A. (2007) *Architecture of a Predictive*

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006) *Multivariate data analysis*, volume 6. Pearson Prentice Hall Upper Saddle River

Hall D.L. & Llinas, J. (2001). *Handbook of Multisensor Data Fusion*. CRC Press, Boca Raton, FL

Hall D.L. & McMullen, S.A.H. (2004). Mathematical Techniques in Multi-Sensor Data Fusion. Artech House, Boston

Hannah, P., Starr, A. & Bryanston-Cross, P. (2001). *Condition monitoring and diagnostic engineering—A data fusion approach, in: Condition Monitoring and Diagnostic Engineering Management*, Amsterdam, pp. 275–282.

Hannah, P., Starr, A. & Bryanston-Cross, P. (2001). Condition monitoring and diagnostic engineering—A data fusion approach, in: Condition Monitoring and Diagnostic Engineering Management, Amsterdam, pp. 275–282.

Hart, P.E. (1968). *The condensed nearest neighbor rule*, IEEE Transactions on Information Theory 14, 515–516. doi: 10.1109/TIT.1968.1054155

Hashemian, H. M. (2011). State-of-the-Art Predictive Maintenance Techniques, in IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 1, pp. 226-236, Jan. 2011.doi: 10.1109/TIM.2010.2047662

Haykin, S.S. (2000). *Unsupervised Adaptive Filtering*. Wiley, New York

He, S. G., He, Z. & Wang, G. A. (2013). *Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques. Journal of Intelligent Manufacturing*, 24(1):25–34

Hecht-Nielsen, R. (1989). *Theory of the backpropagation neural network in Neural Networks.* IJCNN., International Joint Conference on. IEEE, 1989, pp. 593–605

Hinchcliffe, G. & Smith, A. (2003). RCM - Gateway to World Class Maintenance, Butterworth-Heinemann.

Hochreiter, S. & Schmidhuber, J. (1997). *Long short-term memory Neural computation*, vol. 9, no. 8, pp. 5–1780.

Hrushikesh N., Mhaskar and Charles A. Micchelli. (1992*). Approximation by superposition of sigmoidal and radial basis functions.* Advances in Applied mathematics, 13(3):350–373

Huang, N. E & Shen, S.S. (2005). *The Hibert-Huang transform and its applications*. World Scientific Pub Co Inc

Huang, N. E. &. Attoh-Okine, N.O. (2005). *The Hibert-Huang transform in engineering*. CRC Press

Huang, N. E. (2005). *Hilbert-Huang Transform and Its Applications*, pp. 1-26

Huang, N. E., Wu, N. L., Qu, W., Long, S.R., Shen, S.S.P. & Zhang, J.E. (2003) Aplications of Hilbert-Huang transform to non-stationary financial time series analysis. Applied Stochastic Models in Business and Industry, 19(3): 245-268

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. & Liu, H.H. (1998). *The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995

Huda, A.S.N. & Taib, S. (2013). *Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment*. Applied Thermal Engineering, (61), 220-227

Imanuel, (2018). *What is Data Mining?* Retrieved from https://www.predictiveanalyticstoday.com/what-is-data-mining/

Imanuel, (2018). *What is Predictive Analytics?* Retrieved from https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/

Jahnke, P. (2015). *Machine Learning Approaches for Failure Type Detection and Predictive Maintenance* (Master Thesis) University of Darmstadt

Jang, J. S. R. (1993). *Anfis: adaptive-network-based fuzzy inference system. Systems, Man and Cybernetics*, IEEE Transactions on, 23(3):665–685

Jang, J.S. (1993). *ANFIS: adaptive-network-based fuzzy inference system IEEE. Trans. Syst. Man. Cybern. Syst*., 23, pp. 665-685

Jardine, A.K.S., Lin, D. & Banjevic, D. (2006.) *A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical systems and signal processing,* 20(7):1483–1510

Jasiulewicz-kaczmarek, Małgorzata, Drożyner & Przemysław. (2013). The Role of Maintenance in Reducing the Negative Impact of a Business on the Environment. 10.1007/978-3-642-32081-1_7.

Jian-Da, W. & Liu, C.H. (2009). *An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network.* Expert systems with applications, 36(3):4278–4286

Kaushik, S. (2016). *Introduction to Feature Selection methods with an example (or how to select the right variables?)*. Retieved from https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/

Khoshgoftaar, T. M. & Hulse, J. V. (2008). Imputation techniques for multivariate missingness in software measurement data. Software Quality J. 16(4):563–600; [Online]. http://dx.doi.org/10.1007/s11219-008-9054-7.

Khoshgoftaar, T. M., Hulse, J. V. & Napolitano, A. (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. Syst Man Cybern Part a Syst Hum IEEE Trans.;41(3):552–68

Kiencke, U., Schwarz, M. & Weickert, T. (2008). *Signalverarbeitung: Zeitfrequenz-Analyse und Schätzverfahren*. Oldenbourg Verlag

Koksal G., Batmaz, I. & Testik, M. C. (2011). *A review of data mining applications for quality improvement in manufacturing industry*. Expert Syst. Appl., vol. 38, no. 10, pp. 13448-13467

Koksal, G., Batmaz, I. & Testik, M. C. (2011). *A review of data mining applications for quality improvement in manufacturing industry*. Expert Syst. Appl., vol. 38, no. 10, pp. 13448-13467

Krawiec, K. (2002). *Genetic programming-based construction of features for machine learning and knowledge discovery tasks*. Genetic Programming and Evolvable Machines 3.4, 329-343.

Křenek, Jiří & Kuca, Kamil & Blazek, Pavel & Krejcar, Ondrej & Jun, Daniel. (2016). *Application of Artificial Neural Networks in Condition Based Predictive Maintenance.* Retrieved from https://www.researchgate.net/publication/299533279_Application_of_Artificial_Neural_Networks_in_Condition_Based_Predictive_Maintenance

Kubat, M., Matwin, S. (1997). *Addressing the curse of imbalanced training sets: onesided selection*, in: Proceedings of the 14th International Conference on Machine Learning (ICML'97), pp. 179–186

Kurniati, Ruey-Huei, Y. & Lin, J.J. (2015) Quality Inspection and Maintenance: The Framework of Interaction. Retrieved from https://www.sciencedirect.com/science/article/pii/S2351978915011543

Laurikkala, J. (2001). *Improving identification of difficult small classes by balancing class distribution*, in: Proceedings of the 8th Conference on AI in Medicine in Europe: Artificial Intelligence Medicine (AIME'01), pp. 63–66.

Lavi, Y. (2018). *The Rewards and Challenges of Predictive Maintenance.* Retrieved fromhttps://www.infoq.com/articles/predictive-maintenance-industrial-iot/

Lee Y.W., Pipino, L., Strong, D.M. & Wang, R.Y. (2004). *Process-embedded data integrity. J. Database Manage,* 15 (1), p. 17 (87)

Lee, C., Cao, Yi. & Kam K.H. (2017). *Big Data Analytics for Predictive Maintenance Strategies.* 10.4018/978-1-5225-0956-1.ch004.

Lei, Y., He, Z. & Zi, Y. (2008). *A new approach to intelligent fault diagnosis of rotating machinery*. Expert Systems with Applications, 35(4):1593–1600

Li, M. (2010*) Fractal time series-a tutorial review. Mathematical Problems in Engineering*

Lillrank, P.M. (1990) *Laatumaa - Johdatus Japanin talouselämään laatujohtamisen näkökulmasta*, Gummerus. Kirjapaino Oy, Jyväskylä, Finland, pp. 277.

Liu Q. & Wang, H.P. (2001). *A case study on multisensor data fusion for imbalance diagnosis of rotating machinery*. (AI EDAM) Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 15, pp. 203-210

Liu, H. & Motoda, H. (1998). *Feature extraction, construction, and selection: A data ming perspective.* Kluwer

Liu, X. &. Yang, C. (2009). *Greedy kernel PCA for training data reduction and nonlinear feature extraction in classification*, in MIPPR: Automatic Target Recognition and Image Analysis, vol. 7495 of Proceedings of SPIE, Yichang, China

Lizák, F. & Kolcun, M. (2008). *Improving reliability and decreasing losses of electrical system with infrared thermography*. Acta Electrotechnica et Informatica, 8(1), 60-63

López, V., Fernandez, A., Garcia, S., Palade, V. & Herrera, F. (2013). *An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics*. Information Sciences 250, 113-141 doi: 10.1016/j.ins.2013.07.007

Luxhøj, J. T., Riis, J. O., & Thorsteinsson, U. (1997). Trends and perspectives in industrial maintenance management. Journal of Manufacturing Systems, 16(6), 437–453.doi:10.1016/S0278-6125(97)81701-3

Maintenance Framework, International Conference on Computer Information Systems and Industrial Management Applications, IEEE, Dresden.

Mark Tab, (2018). *Guía de Azure AI para soluciones de mantenimiento predictivo.* Retieved from https://docs.microsoft.com/es-es/azure/machine-learning/team-data-science-process/cortana-analytics-playbook-predictive-maintenance#feature-engineering

Markovitch, S. & Rosenstein, D. (2002). Feature generation using general constructor functions. Marching Learning, 49 (1), pp. 59-98

Mathworks Team. (2017). *Using Simulink to create fault data*, Retrieved from: https://de.mathworks.com/help/predmaint/examples/Use-Simulink-to-Generate-Fault-Data.html?s_eid=PEP_19715

Meireles, M.R.G., Almeida, P.E.M. & Simoes, M.G. (2003). *A comprehensive review for industrial applicability of artificial neural networks*. IEEE Transactions on industrial electronics, 50 (3), 585-601

Meola, C. & Carlomagno, G. M. (2004). *Recent advances in the use of infrared thermography*. Measurement science and technology, 15(9), R27

MissingLink Team, (2018). *The Complete Guide to Artificial Neural Networks: Concepts and Models*. Retrieved from https://missinglink.ai/guides/neural-network-concepts/complete-guide-artificial-neural-networks/

MissingLink Team. (2018). *Backpropagation in Neural Networks: Process, Example & Code*. Retrieved from https://missinglink.ai/guides/neural-network-concepts/backpropagation-neural-networks-process-examples-code-minus-math/

MissingLink Team. (2018). *Hyperparameters: Optimization Methods and Real-World Model Management*. Retrieved from https://missinglink.ai/guides/neural-network-concepts/hyperparameters-optimization-methods-and-real-world-model-management/

Mobley, K. (2002). *An Introduction to Predictive Maintenance*

Mobley, K. (2019) Improving availability is much more than maintenance. Retrieved from https://www.reliableplant.com/Read/19702/improving-availability-is-much-more-than-maintenance

Mobley, K.R. (2002), An Introduction to Predictive Maintenance, Butterworth-Heinemann, London.

Mobley, R. K. (2002). Predictive Maintenance Techniques. An Introduction to Predictive Maintenance, 99-113. doi:10.1016/b978-075067531-4/50006-3

Montalvo Martín, C. (2010). *Analysis of signals of mechanical measurements for advanced predictive maintenance*. Thesis (Doctoral), E.T.S.I. Mines (UPM)

Moubray, J. (1997). *Reliability-centered maintenance*. *ReliabilityCentered Maintenance* (p. 445). Industrial Press Inc. Retrieved from http://books.google.com.hk/books?id=bNCVF0B7vpIC

Mourri, B. (2017) Neural Networks and Deep Learning. Coursera:www.coursera.org.

Murthy, S.K. (1998). *Automatic construction of decision trees from data: A multi-disciplinary survey.* Data Mining and Knowledge Discovery, 2, pp. 345-389

Napiórkowski J. & Szczyglak, P. (2011) Analysis of factors in the decision process of purchasingfarm tractors. In_zynieria Rolnicza 9(134) 145–152

National Conference on Manufacturing Research, University of East London, UK, 2000.

Niu, G. (2017). *Data-Driven Technology for Engineering Systems Health Management*. Springer.

Notash, L. & Huang, L. (2003). *On the design of fault tolerant workspace and fault tolerant parallel manipulators* Mechanism and Machine Theory: 38:85-101

Orsenigo, C. & Vercellis, C. (2012). *An effective double-bounded tree-connected Isomap algorithm for microarray data classification*, Pattern Recognition Letters, vol. 33, no. 1, pp. 9–16

Pacheco, W.S. & Castro Pinto, F. A. N. (2013). *Classifier based on artificial neural networks and beamforming technique for bearing fault detection*. ABCM, 22nd International congress of mechanical engineering COBEM 2013, Brazil

Peng, Y., Dong, M. & Zuo, M. J. (2010). *Current Status of Machine Prognostics in Condition-Based Maintenance: A Review.* The International Journal of Advanced Manufacturing Technology, 50, 297-313.

Peng, Y., Dong, M. & Zuo, M.J. (2010). *Current status of machine prognostics in condition-based maintenance: a review.* The International Journal of Advanced Manufacturing Technology, 50(1-4):297–313

Petković, D. Ž. & Ćojbašič, V. N. (2013). Adaptive neuro-fuzzy approach for wind turbine power coefficient estimation Renew. Sustain. Energy Rev., 28, pp. 191-195

Petković, D., Pavlović, N.T. & Ćojbašić, Ž. (2016). *Wind farm efficiency by adaptive neuro-fuzzy strategy Int. J. Elec. Power*, 81, pp. 215-221

Piramuthu, S., Ragavan, H. & Shaw, M.J. (1998). *Using feature construction to improve performance of neural networks.* Management Science, 44 (3), pp. 416-430

Piramuthu, S., Ragavan, H. and Shaw, M.J. (1998). *Using feature construction to improve performance of neural networks.* Management Science, 44 (3), pp. 416-430

Plasmatic *(2017). Unsupervised Machine Learning Big Data technologies and techniques. Project carried out by PLASMATIC*

Pooria, K. & Jazayeri-Rad,H. (2014). *Comparing the fault diagnosis performances of single neural networks and two ensemble neural networks based on the boosting methods.* Journal of Automation and Control, 2(1):21–32

prEN 13306 (1998). Draft European Standard – Maintenance Terminology. Brussels: European Committee for standardization.

Ray, S. (2017) *Commonly used Machine Learning Algorithms (with Python and R Codes)* Retrieved from https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

Reason, J., (2001) ``Managing the Risks of Organisational Accidents´´, Ashgate, ISBN 184014015

Reinelt, W. & Ljung, L. (2003). *Robust control of identified models with mixed parametric and non-parametric uncertainties*. European journal of control, 9(4):373–380

Rey, C.C. (2018). *Classification and its implementation in* R. Bachelor tesis at the University of Sevilla

Robert J. Latino & Kenneth, C. Latino, (1999) ``Root Cuase Analysis- Improving Performance for Bottom- Line Results´´, CRC Press, ISBN 0849307732

Rolf Isermann and Marco Münchhof. Identification of dynamic systems: an introduction with applications. Springer Science & Business Media, 2010.

Rumelhart, D. E., Hilton, G., & Williams, R. J. (1986). *Learning internal representations by error propagation*. In E. Rumelhart & J. L. McClelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 1). MIT Press.

Russell, S. & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach.* Prentice Hall, second edition ISBN 978-0137903955.

Saeed, V. V. (2008). *Advanced digital signal processing and noise reduction.* John Wiley & Sons

Saez P. (2018). *Industrial Predictive Maintenance*. Retrieved from https://www.predictive-sigma.com/que-es-el-mantenimiento-predictivo/

Sahoo, T., Sarkar, P.K. & Sarkar, A.K. (2014). *Maintenance Optimization for Critical Equipments in process industries based on FMECA Method*, International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 10,

Said, E. & David, A. D. (1984) *Testing for unit roots in autoregressive-moving average models of unknown order*. Biometrika, 71(3):599–607

Sanders, Al. (2011). *Discussion with study team on general advanced manufacturing*. Washington, DC, July 15.

Saxena, A. & Saad, A. (2007). *Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems.* Applied Soft Computing, 7(1), 441-454

Saxena, A., Celaya, J., Saha, B., Saha, S. & Goebel, G. (2010). *Metrics for Offline Evaluation of Prognostic Performance*. International Journal of Prognostics and Health Management (IJPHM), 1:1–20, 2010.

Schaffer, C. (1993). *Selecting a classification method by cross validation*, Machine Learning 13, 135e143.

Schmidhuber, J. (2015). *Deep Learning in Neural Networks: An Overview*. Neural Networks. 61: 85–117.

Scholkopf, B. & Smola, A.J. (2002). *Learning with Kernels. MIT Press*

Sciban, R. (2017). *6 Tools for a Successful Predictive Maintenance Program*. Retrieved from:https://us.hitachi-solutions.com/blog/6-tools-for-a-successful-predictivemaintenance-program/.

Sciore, E., Siegel, M. & Rosenthal, A. (1994). *Using semantic values to facilitate interoperability among heterogeneous information systems.* ACM Transactions on Database Systems, 19,2, 254-290.

Sherwin, D. (2000). A review of overall models for maintenance management. *Journal of Quality in Maintenance Engineering*, *6*(3), 138–164. doi:10.1108/13552510010341171

Shetty, D., Ali A. & Chapdelaine, J.J. (2009) A model for the total manufacturing assessment and implementation. J Adv Manufact Syst 8(2):117–136

Shu-Guang, H., He, Z. & Wang, G.A. (2013). *Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques*. Journal of Intelligent Manufacturing, 24(1):25–34

Si, X.S., Wang, W., Hu, C.H. & Zhou, D.H. (2011). *Remaining useful life estimation–a review on the statistical data driven approaches.* European Journal of Operational Research, 213(1):1–14

Smith, R. (2001) Effective Maintenance Management in Plant Engineer's Handbook, Butterworth-Heinemann, Boston.

Somonte Martín, R. (2010). *Analysis of filters for the detection of faults in industrial bearings its your sign of vibration*, Bachelor Final Thesis, University of Madrid

Sondalini, M. (2004). *Win Production Over to Doing Better Maintenance with ABC Equipment Criticality Rating that Uses the Business-Wide Costs of Production Loss.*

SPCCONSULTING. (2012). *Six Sigma Tools: FMEA* types. Retrieved from https://www.spcconsulting.org/2012/08/6-sigma-tools-fmea-types/

Specht, D. F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6):568–576

Srivastava,T.(2014) *Introduction to Random forest – Simplified,* Retrieved fromhttps://www.analyticsvidhya.com/blog/2014/06/introduction-random-forest-simplified/

Srivastava, T. (2018). *Introduction to k-Nearest Neighbours: A powerful Machine Learning Algorithm (with implementation in Python*. Retrieved from https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

Stamatis, D.H. (2003). *Failure Mode and Effect Analysis: FMEA from Theory to Execution.*

Stanley Nowlan, F. and Howard Heap F. (1978). Reliability-centered maintenance, DOD Report Number A066-579, December 29, 1978.

Stapelberg, R.F. (2009). Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design.

Sudarsanam, S.K. (2014*). Fuzzy logic based predictive maintenance of complex equipments.* Retrieved from https://www.researchgate.net/publication/281277973_Fuzzy_logic_based_predictive_maintenance_of_complex_equipments

Suykens, J. & Vandewalle, J. (1999) *Least squares support vector machine classifiers. Neural Processing Letters, 9(3):293–300*

Swedish Standard Institute (2001), 'SS-EN 13306 - Maintenance terminology', Swedish Standard Institute, Stockholm.

Takata S., Kimura F., van Houten FJAM., Westkämper E., Shpitalni M., Ceglarek D. & Jay Lee J. (2004). *Maintenance: changing role in life cycle management*. Ann CIRP 53(2):643–656

Tenenbaum, J. B., de Silva, V. & Langford, J.C. (2000). *A global geometric framework for nonlinear dimensionality reduction*, Science, vol. 290, no. 5500, pp. 2319–2323, 2000.

Tomarchio, G., Faulisi, A., Neumann, W. & Janousek, V. (2018). Jumping into Industry 4.0 with Predictive Maintenance Solutions. Retrieved from https://www.eetimes.com/author.asp?section_id=36&doc_id=1333963

Tomasz, B., RiazT., & Pedersen, J.M. (2012). *A method for classification of network traffic based on c5.0 machine learning algorithm. In Computing, Networking and Communications (ICNC)*, pages 237–241

Tomek, I. (1976). *Two modifications of CNN*, IEEE Transactions on Systems Man and Communications 6, 769–772. doi: 10.1109/TSMC.1976.4309452

Tsang, A. H. C. (2002). Strategic dimensions of maintenance management. *Journal of Quality in Maintenance Engineering*, *8*(1), 7–39. doi:10.1108/13552510210420577

Van Hoof, C., Baert, K. & Witvrouw, A. (2004). *The best materials for tiny, clever sensors. Science*, 306(5698):986–987

Van Hulse J. & Khoshgoftaar, T. (2009). *Knowledge discovery from imbalanced and noisy data.* Data Knowl Eng;68(12):1513–42.

Villanueva, J. B., Espadafor, F. J., Cruz-Peragon, F. & García, M. T. (2011). *A methodology for cracks identification in large crankshafts.* Mechanical Systems and Signal Processing, 25(8), 3168-3185

Wang R.Y. & Strong, D.M. (1996). *Beyond accuracy: what data quality means to data consumers*. J. Manage. Inf. Syst., 12 (4), pp. 5-33

Wang, C., Zhang, Y. & Zhong, Z. (2008). *Fault diagnosis for diesel valve trains based on time–frequency images*. Mechanical Systems and Signal Processing, 22(8):1981–1993

Wang, W.Q., Golnaraghi, M.F & Ismail, F. (2004). *Prognosis of machine health condition using neuro-fuzzy systems. Mechanical Systems and Signal Processing*, 18, pp. 813-831

Wang, C., Zhang, Y. & Zhong, Z. (2008). *Fault diagnosis for diesel valve trains based on time–frequency images.* Mechanical Systems and Signal Processing, 22(8):1981–1993

Welch, P. (1967). *The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms*. IEEE Transactions on audio and electroacoustics, pages 70–73

Wessel N. van Wieringen. Lecture notes on RR. Retrieved from: https://arxiv.org/pdf/1509.09169.pdf

Wilson, D.L. (1972). *Asymptotic properties of nearest neighbour rules using edited data*, IEEE Transactions on Systems, Man and Cybernetics 2 (3) 408–421. doi: 10.1109/TSMC.1972.4309137

Winters, P. & Silipo, R. (2015). *Anomaly detection in predictive maintenance. Time alignment and visualization*. Knime white paper

Wire H., (December 2004). The eMagazine for the Reliability Professional by ReliaSoft Corporation available at www.weibull.com/hotwire/issue46/relbasics46.htm

Wollenhaupt, G.(2016) "IoT slashed downtime with predictive maintenance," PTC, Retrieved from http://www.ptc.com/productlifecycle-report/iot-slashes-downtime-with-predictive-maintenance

Wu, T. Y., Chen, J. & Wang, C. X. (2012). *Characterization of gear faults in variable rotating speed using hilbert-huang transform and instantaneous dimensionless frequency normalization*. Mechanical Systems and Signal Processing, 30:103–122

Xu R. & Kwan, C. (2003). Robust isolation of sensor failures. Asian Journal of Control, 5, pp. 12-23

Yasuda, T., Ohkura, K. & Ueda, K. (2006). Ahomogeneous mobile robot team that is fault tolerant. Advanced Engineering Informatics; 20:301-311 Retrieved from https://www.researchgate.net/publication/220371846_A_homogeneous_mobile_robot_team_that_is_fault-tolerant

Zadeh, L. A. (1965). *Fuzzy sets. Information and control,* 8(3):338–353

Zena M. H. & Gillies, D. f: (2015). *A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data*, Advances in Bioinformatics, vol. 2015, Article ID 198363, 13 pages, https://doi.org/10.1155/2015/198363.

Zhang, L., Xiong, G., Liu, H., Zou, H. & Guo, W. (2010). *Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference.* Expert Systems with Applications, 37(8):6077–6085

Zhang, S. & Ganesan, R. (1997). *Multivariable trend analysis using neural networks for intelligent diagnostics of rotating machinery*. Transactions of the ASME. Journal of Engineering for Gas Turbines and Power, 119, pp. 378-384

Zhao, H., Atish, P., Sinha & Wei Ge. (2009). Effects of feature construction on classification performance: An empirical study in bank failure prediction, In Expert Systems with Applications, Volume 36, Issue 2, Part 2, pp. 2633-2644, ISSN 0957-4174

Zhou, Q., Wu, Y., Chan, C.W. & Tontiwachwuthikul, P. (2011). *From neural network to neuro-fuzzy modeling: applications to the carbon dioxide capture process Energy*. Procedia., 4, pp. 2066-2073

Zhou, Z.J., Hu, C.H., Zhang, B. C., Xu, D. L. & Chen, Y. W. (2013). *Hidden behavior prediction of complex systems based on hybrid information.* IEEE Transactions on Cybernetics, 43(2):402–411