Tampere University

Anton Miettinen

DATA-DRIVEN PROGNOSTICS IN INDUSTRIAL SERVICE BUSINESS

Faculty of Engineering and Natural Sciences
Master of Science Thesis
May 2019

ABSTRACT

Anton Miettinen: Data-Driven Prognostics in Industrial Service Business
Master of Science Thesis, 74 pages
Tampere University
Master's Degree Programme in Mechanical Engineering
Examiners: Professor Miia Martinsuo, Professor Kari Koskinen
May 2019

There is a shift in the manufacturing industries in which original equipment manufacturers (OEM) are gaining increasingly large portion of their revenue from services rather than the manufacturing of goods. This change is called servitisation. Additionally, the advancements in information technology are opening new possibilities and opportunities, such as in how data can be processed, analysed and used to create data-driven applications to support the business functions. The possibilities are, however, still largely unexploited especially in the field of maintenance services. The data-driven prognostics could not only enhance the existing maintenance activities, but also create new ways of partnership and service development between the OEMs and their clients. This could induce further growth and increase in the servitisation level. However, there is lack of insight of how the methods could be applied to practice; especially case studies are few in quantity. Hence, this study aims to increase understanding of the practical application of the data to support maintenance service business.

This study examines the application of data-driven methods, mainly machine learning, to aid valve maintenance business of a service providing OEM. The aim is to create a data-driven system to forecast failures in devices and generate automated service recommendations. The forecasting was based on idea that the failures would induce a detectable pattern in the measured data prior a failure. The chosen machine learning method, the neural networks, excel in this kind of task and hence can predict failures. The study is conducted in practical setting as a case study with real data.

Various systems and processes were examined, and data was extracted for analysis. With this data several models for prediction were built. However, the accuracy of these was ultimately deemed insufficient for generation of service recommendations and hence all the set goals were not fully reached. As the greatest contributing factors for the poor performance of the forecasts, the data itself and the operations related to it were identified. The data was hard to access and lacking both in quality and quantity as it is recorded, stored and managed with day-to-day operations in mind. As result, we found that significant portions of data were deleted or were recorded with accuracy insufficient for this research. However, through the analysis of these factors several concrete points of development emerged.

The outcome of this study also confirms the inherent challenges regarding service partnering and intercompany data-transfer presented in literature. A need for standardised and light-weight legal frameworks and methods of data sharing was identified. Without these, the potential may not be fully realisable in practice and hence more case practically oriented studies on the subject are required.

To conclude, the OEM had too optimistic view of the availability, quality and quantity of data, which resulted in an attempt, which did not reach all the set goals. On the other hand, the academic literature shows that there is great potential in these methods. Data refined into wisdom which may support decisions and actions can facilitate value generation in services. The findings encourage OEM to improve the collection, storage and management of data and other organisations to carefully evaluate whether their capabilities are sufficient.

Keywords: Machine learning, Maintenance, Reliability, Servitisation, Value

TIIVISTELMÄ

Anton Miettinen: Datapohjaisilla menetelmillä ennustaminen teollisessa palveluliiketoiminnassa
Diplomityö, 74 sivua
Tampereen yliopisto
Konetekniikan diplomi-insinöörin tutkinto-ohjelma
Tarkastajat: Professori Miia Martinsuo, Professori Kari Koskinen
Toukokuu 2019

Yhä kasvava osuus perinteisen valmistavan teollisuuden laitevalmistajien liikevaihdosta syntyy teollisista palveluista varsinaisen tuotevalmistuksen sijaan. Lisäksi tietotekniikan ja datan käsittelyn kehittyminen ovat mahdollistaneet uusia tapoja toimia ja kehittää liiketoimintaa. Tiedon käsittelyn sekä analyysin ja dataan pohjautuvien sovellusten tuomat mahdollisuudet ovat kuitenkin suurelta osin hyödyntämättä erityisesti huoltoliiketoiminnassa. Dataan pohjautuva vikaantumisten ennustaminen voi parantaa jo olemassa olevia prosesseja tai jopa mahdollistaa uusia tapoja huollon organisoinnissa ja huoltopalveluliiketoiminnassa sekä lisätä laitevalmistajien ja loppukäyttäjien välistä yhteistyötä. Käytännön sovelluksia kuitenkin on vielä vähän. Tämä työ käsitteleekin data-pohjaisten menetelmien, pääasiassa koneoppimisen, soveltamista käytäntöön venttiilien huoltoliiketoiminnan hyödyttämiseksi ja tiedon lisäämiseksi käytännön soveltamisen haasteista.

Työn tarkoituksena oli luoda data-pohjainen järjestelmä, joka pystyisi ennustamaan laitteiden tulevia vikaantumisia ja luomaan tähän ennusteeseen pohjautuvia huoltosuosituksia automaattisesti. Tulevat vikaantumiset näkyvät mittausdatassa jo ennen varsinaista vikaantumista johtuen kunnon heikkenemisestä syntyvistä kaavamaisista poikkeamista laitteen toiminnassa. Tunnistamalla nämä poikkeamat pystytään ennustamaan näitä seuraava vikaantuminen. Koneoppimismenetelmät, erityisesti neuroverkot, suoriutuvat tällaisista tehtävistä mainiosti ja siten niitä pystytään hyödyntämään myös vikaantumisen ennustuksessa. Tutkimus tehtiin laitevalmistajan toimeksiannosta oikealla mitatulla datalla.

Tutkimuksessa data kerättiin sekä analysoitiin ja sen pohjalta luotiin 30 neuroverkkoihin pohjautuvaa vikaantumisia ennustavaa mallia. Työssä kuitenkin havaittiin ennusteiden olevan liian epätarkkoja niiden hyödyntämiseksi liiketoiminnan tukena. Datan ja vikaantumisten välinen yhteys jäi siis lopulta osoittamatta ja asetettu tavoite tältä osin saavuttamatta. Kuitenkin, ennusteiden epätarkkuuden syyt pystyttiin analysoimaan ja analyysin pohjalta esitettiin kehitysehdotuksia. Suurimmaksi syyksi epäonnistumisellemme määriteltiin data itsesänsä. Vaikakin data oli tallennettu vuosikymmenen verran, olisi se vaikeasti saatavilla sekä laadullisesti ja määrällisesti huonoa. Data on tuotettu, tallennettu ja hallittu päivittäisiä tarpeita varten mistä johtuen suuria osia siitä oli ajan saatossa tuhottu tai kirjattu ylös tätä tutkimusta ajatellen riittämättömällä tarkkuudella.

Tutkimuksen tulokset tukevat kirjallisuudessa esitettyjä palveluliiketoimintaan ja yritysten väliseen yhteistyöhön liittyvien haasteiden suhteen esitettyjä väittämiä. On olemassa selkeä tarve standardoituille ja helppokäyttöisille yritysten väliseen datan jakoon soveltuville käytännöille, prosesseille ja järjestelmille. Lisäksi datapohjaisten menetelmien ja datan käytön käytännön sovelluksia tulee tutkia lisää, jotta näiden tarjoamat mahdollisuudet tulevat hyödynnetyiksi.

Yhteenvetona laitevalmistajalla oli liian optimistinen näkemys datan saatavuudesta, sen määrästä ja laadusta. Tämä johti siihen, että kaikki tutkimukselle asetetut tavoitteet eivät täysin täyttyneet. Kuitenkin, kirjallisuuden perusteella voidaan väittää, että menetelmät voivat olla toimivia. Lisäksi datan jalostus toimintoja tukevaksi tiedoksi voi toimia perustana asiakasarvoa luoville toimille. Täten, tulokset kertovat, että laitevalmistajan tulisi parantaa tiedon keruuta, tallennusta sekä hallintaa. Yritysten olisi suotavaa myöskin realistisesti arvioida omia kyvykkyyksiään sekä aktiivisesti etsiä kehityskohteita nykytilan parantamiseksi yhteistyössä sekä asiakkaiden että laitetoimittajien kanssa.

Avainsanat: Koneoppiminen, Huolto, Käyttövarmuus, Palvelullistuminen, Asiakasarvo

CONTENTS

1. INTRODUCTION 1
   1.1 Background 1
   1.2 Scope and objectives of the work 3
   1.3 Structure of the thesis 4
2. LITERATURE REVIEW 6
   2.1 Industrial services and data 6
   2.2 Service value 9
   2.3 Condition-based and predictive maintenance 15
   2.4 Prognostics and fault detection with machine learning 20
   2.5 Synthesis 25
3. RESEARCH METHODS 29
   3.1 Methodology 29
   3.2 Case description 30
   3.3 Collection of data 32
   3.4 Machine learning methods 34
   3.5 Data pre-processing 37
4. RESULTS 42
   4.1 Evaluation of the built predictive models 42
   4.2 Analysis of the inaccuracy of the predictive models 48
   4.3 The application of the model to service 51
   4.4 Development suggestions 53
      4.4.1 Increasing the availability of data 53
      4.4.2 Other possibilities of forecasting 56
      4.4.3 Services and data management 57
      4.4.4 Summary and timeline 58
5. DISCUSSION 61
   5.1 How can the OEM use available data to forecast failures of field devices and to generate service recommendations? 61
   5.2 In what ways can the OEM use the forecast to develop MRO services offered to customers? 64
6. CONCLUSIONS 67
REFERENCES 70

LIST OF FIGURES

Figure 1: Illustration of different levels of servitisation. As organization adds more use or result oriented services to their portfolio, the level of servitisation increases. 7
Figure 2: Value exchange model adapted from Khalifa (2004). In the figure the relation of value to various costs or sacrifices is illustrated. 10
Figure 3: Pattern of deterioration with the detection and functional limits. Adapted from (Takata et al. 2004) 19
Figure 4: Connection of raw data and the value and the constituent steps between them in business context 28
Figure 5: Examples of tested networks. Shown is the layers and their arrangement. Networks #2 and #20 are convolutional networks, while #12 and #29 are LSTM. Networks #2 and #29 have classification output and the two others regression. 35
Figure 6: The typical measurements in the vicinity of a device (1 and 2) and by the device itself (3). These include pressure (P), temperature (T), and flow (V) before (1) and after the device (2). In addition, positioner of the assembly collects self-diagnostics data (3). 38
Figure 7: Illustration of how the data is trimmed into packets for use as inputs. Firstly, the point of failure t_f is identified from the event data and the data after it is cut off. Secondly, the RUL is given a value and the time window is defined with a constant t_l. Lastly, all data outside time window is discarded and the data packet consists of the data within time window and the value of RUL. 40
Figure 8: A selection of histograms of the prediction errors of the contrived predictive neural networks. The y-axes represent the number of predictions which fell into each of the 20 bins. The x-axis represents the prediction errors. 47
Figure 9: The predictions of RUL by selected example networks plotted against the actual values. The actual value – prediction -pairs were sorted based on the actual value in order from smallest to biggest. 48
Figure 10: Bar graph showing RMSEs of the tested networks. Network #24 has been excluded from this figure as the bar extends way beyond the upper limit of the graph. 49

Table 1: A selection of published research on data-driven prognostics. 22
Table 2: Comparison of results obtained in various studies with various methods of failure prediction on C-MAPPS data. Adapted from Li et al. (2018) 23
Table 3: Types of data collected for the purpose of this research, as well as their sources and description of quantity. 33
Table 4: The recorded variables from the ND9-series positioners that were used in this study as well as their descriptions. The descriptions are adapted from the Metso ND9 Intelligent Valve Controller User's Guide (Metso, 2019). Avg refers to averaged value, Std to the standard deviation and Cum to the cumulative value. 42
Table 5: Tabulated results of the first testing. In the columns the number of the network, the type of the network, and the calculated values of RMSE, median, mean and standard deviation.............................. 45
Table 6: Tabulated results of improvement and optimisation effort. Presented are the number of the network, its type, the best calculated RMSE, the vector of parameters of the network, and increase in per cents compared to results presented in the table 5. 46
Table 7: Summary of development suggestions... 59

LIST OF SYMBOLS AND ABBREVIATIONS

CBM	Condition Based Maintenance
CCEB	Current Condition Evaluation-Based
C-MAPSS	Modular Aero-Propulsion System Simulations
CNN	Convolution Neural Network
DCOM	Distributed Component Object Model
DIKW	Data, Information, Knowledge, Wisdom
FCPB	Future Condition Prediction-Based
G-D	Goods-Dominant
HART	Highway Addressable Remote Transducer
IoT	Internet of Things
LSTM	Long Short Term Memory
MRO	Maintenance and repair operations
NN	Neural Network
OEM	Original Equipment Manufacturer
OPC	Open Platform Communications
OPC HDA	OPC Historic Data Access
OSP	Original Service Provider
PSS	Product Service System
RMSE	Root Mean Square Error
RUL	Remaining Useful Life
S-D	Service-Dominant
SVM	Support Vector Machine
SQL	Structured Query Language
TBM	Time Based Maintenance

1. INTRODUCTION

1.1 Background

Maintenance and repair operations (MRO) incur costs from both the labour and material as each of the devices targeted by the operation need to be checked and serviced, worn parts need to be changed and faulty devices need to be replaced. In the past, maintenance and related activities have been viewed as an inevitable costs (Fraser et al. 2015; Ali-Marttila et al. 2017). It is estimated that the cost of maintenance can amount from 15% to 40% or even 70% (Lofsten 1999; Muthu et al. 2000; Efthymiou et al. 2012) of the total production costs during the life time of the plant. As the technological complexity increases, so does the associated maintenance costs (Lofsten 1999). In addition to direct costs, inadequate maintenance results also to low performance, low productivity and high downtime (Fraser et al. 2015). As efficiently planned maintenance can reduce these expenses, instead of seeing MRO as a cost, modern scholars view it as a value generating function (Ali-Marttila et al. 2017).

It is increasingly common for plant operators to outsource the maintenance of the facilities. Even though it is a vital function for the success of the operator, it is not the core business function for the operator and therefore can be outsourced to a specialist maintenance companies and service provider (Toossi et al. 2013). This provides great cost savings, as the operator does not need to keep an own extensively trained maintenance workforce on the payroll. Moreover, Original Equipment Manufacturers (OEM) are expressing great interest to partake in the service business as the profit margins of the goods diminish (Kowalkowski 2006; Kindström et al. 2013).

The high availability of devices is ensured with proper maintenance and care and there are several ways in which the maintenance can be organised. The most rudimentary method is to maintain the devices when they fail. This practice is common in e.g. paper industry, where the normal operation requires breaks in the process during which the maintenance can be conducted. In other industries, such as in oil and gas industry, preventive maintenance, which is also known as scheduled maintenance, is the prevailing maintenance practise. This industry is characterised by high volumes of production and continuous operation of the facilities and the main competitive factor of the plants is the low cost of production per unit of products. This is achieved with high total volume and efficient use of capacity. Moreover, any downtime in process carries risk of damaging the process equipment. Due to this, it is essential that the plant does not suffer from unintended downtime, meaning that the availability of each individual device or subsystem must be high.

Preventive maintenance strategy aims to minimise unplanned downtime by servicing the equipment periodically on a predefined schedule at regular intervals (Ahmad & Kamaruddin 2012). For example, an oil refinery may perform total plant shutdowns annually or biannually. During the shutdowns the process is halted, and all daily activities cease for a plant wide maintenance. This practise significantly reduces unwanted and costly failures, but on the other hand, the amount of unnecessary checks and maintenance increases as perfectly functioning devices get called to maintenance (Susto et al. 2012).

Before a plant enters shutdown, a maintenance plan and schedule are made by an expert engineer or a technician with the help of recommendations of the original equipment manufacturer (Ahmad & Kamaruddin 2012). For this, information on the condition and type of the devices are needed (Efthymiou et al. 2012). In all cases the information is not easily available or is missing, thus the knowledge is gathered from physical device audit. However, during the audit it is not possible to inspect the interior parts of the device as the equipment is still in use and it is common for the devices to be covered in insulation or be in hard to reach places. If the information, or a forecast, on condition of the devices were available in the planning phase, the planning process could be made more efficient and plans more accurate. A prospective measure for improving this process is data-driven prognostics and condition analysis.

Forecasting the condition of the devices and predicting failures is called prognostics. This is a well-researched field that has traditionally relied on analytic examination of physical phenomena, statistics and other more conventional methods (Costello et al. 2017). Data-driven prognostics are a recent development in the field of condition management. Data-driven methods are methods of analysing and processing data. They can be used to model systems by essentially mapping the input data to the results. Data-driven approaches focus on establishing the connection directly from the available data.

The development is made possible by the rapid development of data-driven methods, especially machine learning, in the last decade. Several researchers have applied the data-driven methods successfully in for this purpose, but also in application in other fields, such as detection of diabetes (Lekha & Suchetha 2018), image recognition and modelling of waste water hydraulics (Granata & De Marinis 2017).

The advantage of this approach is that the engineer does not necessarily need to know the exact cause and effect relations to form the relation. This has both advantages and disadvantages. On one hand, the relation remains opaque and thus it is not entirely obvious as to how a certain set of inputs affects the outcome. This is disadvantageous, if there is any need to explain how the model functions. On the other hand, forming explicit relations on a complex multivariable system is time consuming at best and impossible at worst. Considering this, data-driven methods have potential to save time in building an accurate model and hence serve as a valuable aid in maintenance planning.

However, the possibilities of value generation by processing the data are not fully exploited (Kunttu et al. 2017) and especially data-driven methods are largely underutilised in the industry due to their novelty. On the other hand, the findings suggest that there is great potential if these methods were applied to help the management and planning of activities. This research aims to provide insight on practical applications of data-driven prognostics and how it can be used to support planning and business decisions regarding MRO activities. The research is commissioned by Metso Oy, who manufactures field devices and provides field services to the customers.

1.2 Scope and objectives of the work

The aim of the research and development efforts of this work is to improve the maintenance planning by developing a data-driven system which can forecast failures. The objectives are twofold. First objective is to contrive a system which automatically generates forecasts on the future condition of devices and generates device specific service recommendations. Second objective is to assess the feasibility of the system and conduct an analysis on the possibilities of service development using the system.

Essentially, the research questions lie within the objectives:

1. How can the OEM use available data to forecast failures of field devices and to generate service recommendations?
2. In what ways can the OEM use the forecast to develop MRO services offered to customers?

The focus of this study is in the practical evaluation to discover the extent of possibilities and what is attainable in an actual business setting. However, fit the study within the allocated timeframe the scope must be limited. There is a myriad of different ways in which a prediction of failures can be made and hence, only those which can be expected to perform the best in this application are concentrate on. This shall be established as a part of the literature review. Once the methods are selected, the predictive model is built.

Furthermore, in depth research on data-driven models and maintenance management falls out of scope. Instead the focus is on the practical application of the methods that are proven to work using readily available tools. The system is to be made as a minimum viable product. The main criterion is the sufficient accuracy of predictions and the usability of service recommendations. The goal is to is to investigate the ways in which such a module can be made. At later point, should the effort be considered financially worthwhile and in line with strategic focus and resource usage, the optimisation and improvement can be done.

Data is inevitably needed to build and test the predictive system. Real data, which is recorded in an actual production facility operated by a customer of the OEM is used. With the use of real data, an accurate view on what data is available is obtained and it is ensured that the practical focus is not lost. This will allow the better application of the findings in practice and suggest concrete development actions.

The model needs to support business goals. Therefore, the business goals have to be defined. Moreover, it is vital that the results are something which the customers need and value. To assess the quantity of the generated value, the places where the potential of value generation exist must be found. To support this, a literature review on industrial services and value generation is done.

Special attention is directed to predictive maintenance especially as a service. Ways to build a service offering including predictive maintenance are investigated and discussion on how to integrate new models to existing are examined. The end result should be mutually beneficial to the customer and to the OEM.

1.3 Structure of the thesis

The thesis starts with a literature review to establish an understanding of previous relevant academic research on relevant subjects. These subjects are the industrial services, value in industrial services, predictive maintenance and data-driven methods employed in prognostics. The key findings from the literature are compiled into a synthesis. Some of the most noteworthy points discovered included the Data, Information, Knowledge, Wisdom (DIKW) -hierarchy, which was supplemented by the concept of subjectively experienced value that stems from the well-planned actions supported by the wisdom. The link between raw data and value which this chain of data refinement provides is examined through a case. The current understanding of the service and the customer value in service business to understand the subjective qualities of value is reviewed. Moreover, by reviewing the literature on prognostics and machine learning, it was possible to establish that the failures can be identified by patterns that are induced in measurements and that neural networks are the most capable machine learning method in detecting these.

The case focuses on refining data obtained from a client that operates a large production facility to develop value generating services through refining this data into knowledge and wisdom. The data is condition monitoring data and the desired knowledge reliable forecasts on the future condition of the devices. As the literature indicates that the neural networks perform the best in prognostics tasks, they are employed in contriving a system capable of predicting the upcoming failures. It is assumed that with the help of this knowledge service recommendations could be generated for each of the individual devices automatically. The focus is heavily on the practical work and thus realistic evaluation is selected as the methodology. These are all discussed in detail in the third chapter.

The results of the work as well as analysis on the results and suggestions on the development are presented in the fourth chapter. The preciseness of the built predictive systems is evaluated to assess the usability of the forecasts to determine how well they can be used to advance development of data-driven services and especially the service recommendations. Unfortunately, the built predictive systems do not reach suitable level of accuracy. Nevertheless, there is insight to be gained through analysis of the results and causes for the inadequate performance of the predictive systems. Based on the analysis and observations made during the study concrete suggestions for development are presented. These improvements mainly focus on increasing the availability of data for future research similar to this one so that greater success may be found later on.

In the fifth chapter, the results are discussed in the light of previous research. Here the results are compared to the findings of other researchers. The fact that most of the research on data-driven prognostics are done on simulated rather than real data is recognised, and in the light of this study whether the findings of these studies represents the reality well enough are questioned. Moreover, the challenges of sharing the data between various organisations are discussed. Additionally, concise answers to the research questions are provided. In the sixth chapter the conclusions are presented.

2. LITERATURE REVIEW

2.1 Industrial services and data

One contemporary definition of service is the application of operant resources, that is, knowledge and skill, in benefit for another party (Lusch & Vargo 2014). The definition is very broad, but it reflects the broad nature of services. Services can be offered to both consumers as well as to companies in business to business (B2B) market. Of the latter, some services are industrial services. Kowalkowski (2006) defines industrial services as processes, which support the customers industrial production process in a value generating way. Some typical examples of industrial services are maintenance, life cycle services and facility modernisation services (Ali-Marttila et al. 2017). The industrial services have gained attention in manufacturing industry as the manufacturing companies are seeing the margins of produced good diminish, new contenders rising and technology getting commoditised, and the old ways cannot provide suitable answers in the changing world (Kowalkowski 2006; Kindström et al. 2013).

The services are the basis for the *service-dominant logic* introduced by Vargo & Lusch in their article from 2004. The service dominant (S-D) logic has been presented as a replacement for the older product centric, or Good-Dominant (G-D) logic. In G-D logic the main factor in competitiveness and the source of the customer value was the product itself (Smith et al. 2014). Hence, the aspects and features such as the technological superiority over competition were the key-points of the marketing and development. Essentially, it was thought that a good product alone will sell itself. In comparison to the G-D logic, the roles of goods and services are reversed in S-D logic. Whereas in G-D logic the services are supplements to the physical products, in the S-D logic the goods are subordinate to the services. Indeed, the service dominant logic places the service as the fundamental basis for all exchange and the goods as a distribution mechanism with which the service is delivered (Lusch & Vargo 2014).

To fit services in G-D logic, they are thought of as immaterial goods. Essentially this separates the tangible goods and intangible services to different categories. The commonly mentioned key characteristics that distinguish services from goods are intangibility, heterogeneity, inseparability and perishability; together these are known as IHIP characteristics. In S-D logic the division is not necessary (Lusch & Vargo 2014) and it is also reported that many of the services exhibit opposite characteristics what IHIP suggests (Lovelock & Gummerson 2004).

Servitisation, or the Product-Service transition, is the transition from pure manufacture to pure services offering (Pawar et al. 2009). The transition phase itself between the two states contain the Product Service System (PSS), which is a variable mixture of product centric thinking and services (Pawar et al. 2009). The process is continuous and currently in effect in many companies. Assuming, that operating under the S-D logic provides superior results as opposed to G-D logic, as Lusch & Vargo (2014) claim, it seems natural, that it is in the best interest of companies to proceed rapidly with the transition.

According to Tukker (2004) and Smith et al. (2014) the PSS can be broadly divided into three categories, which are the product-oriented services, the use-oriented services and the result-oriented services. The categories, in this order, represent the spectrum of possible combination between full product centricity and models consisting of only service. When moving further towards the service end of the spectrum, the level of abstraction in the objective or the defined needs increases, but so does the freedom of the provider to choose the way of execution or the means with which to accomplish the desired result (Tukker 2004). The greater freedom of the provider allows them to organise the activities as they see fit and produce solutions in a greater scale. As the provider are specialists in the field they operate in, they should have greater capabilities and insight on as to what is the best way to achieve the desired results.

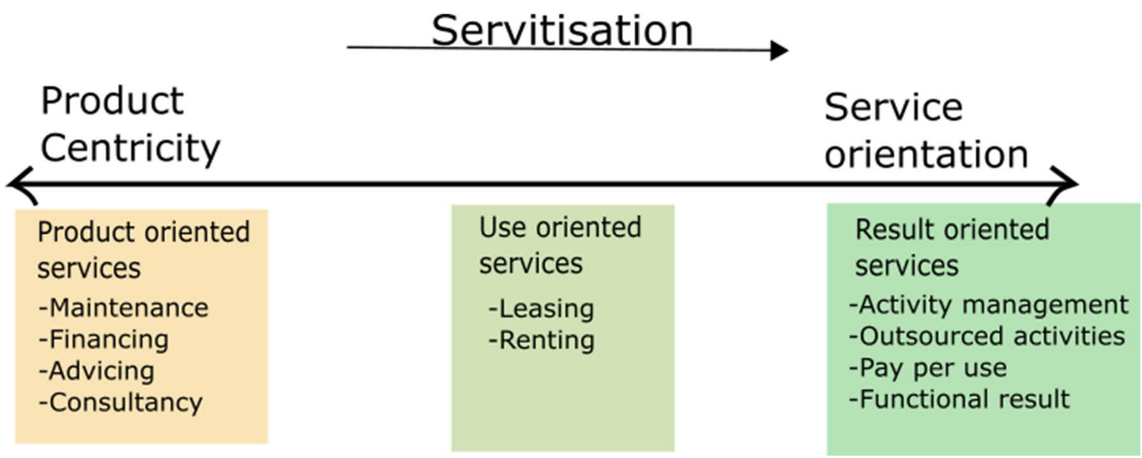

Figure 1: Illustration of different levels of servitisation. As organization adds more use or result oriented services to their portfolio, the level of servitisation increases.

In product-oriented services, the physical product is owned by the customer. The service agreement attached ensures the utility of the item (Smith et al. 2014). This category can be further divided into product-related services, such as maintenance or financing options, and into advice and consultancy (Tukker 2004). This is the most common of the PSS's in the area of industrial flow control equipment.

In contrast, in use-oriented services, the supplier or OEM retains the ownership of the devices and sells the function. Leasing, renting and similar agreements are common contracts of this type (Tukker 2004; Smith et al. 2014). Leasing or renting contracts of field devices are almost non-existent. Unlike in the case of for example cars or property, which are typical examples of leasing and renting respectively, with field devices the manufacturer can not know for certain how the device will be used or under what sort of stress it will experience under its life cycle, therefore the level of risk is uncertain. In addition, the devices are integral parts of the facility, thus not only the operators have interests to keep them in their own possession, but also, they are hard to repossess in the event the of non-payment. Moreover, a leasing agreement warrants a financial institution as middleman as otherwise the OEM would have to bear increased financial risk for no benefit. This in turn only adds cost. Nevertheless, with careful planning, a leasing or renting agreement of field devices could be a lucrative venture.

In result-oriented services the provider sells results of some type and typically no pre-determined product involved (Smith et al. 2014). Result-oriented services include models of activity management or outsourcing, pay per service unit, or pay per use, and functional result (Tukker 2004). In activity management a part of an activity of a company is outsourced; typically catering, cleaning or some other non-core, but still essential activity. Pay per service unit refers to a practice where the user buys the output of a product according to the level of use, such as in the pay-per-print schemes, where the provider of the printing machine facilitates the activity and the user pays for each print (Tukker 2004). Functional result is a hands-free approach. The provider is in this model free to deliver the result in any way it deems appropriate, thus the wished result is often specified in abstract terms (Tukker 2004). This extreme freedom also implies the need for provider to have excellent operant resources at their disposal. Both the skill to formulate the service that fulfils the abstract need and the capability to provide it are needed. For example, if the client requests a safe and green solution, the provider has to not only determine what safe and green are, but to also be able to deliver a result that is both.

In the ever-changing, dynamic field, the companies constantly need to develop new resources, roles and processes to be able to identify the opportunities of service provision (Kindström et al. 2013). Both the technologies available and the needs of the customer can change, and the company must innovate to match their service offering to suit the situation (Kindström et al. 2013). This implies that the company and its organisation must transform with the business environment.

In the perspective of a manufacturing company that engages in service provision, the servitisation is a process that transforms the company from OEM to OSP; an Original Solution Provider (Schnürmacher et al. 2015). Successful transition has several requirements. As creating reciprocal value is the basis of the business (Grönroos & Ravald 2011), it is not surprising that several of the requirements are about the relationship between the participants and organisational aspects. These are e.g. service-oriented culture, relationship marketing and trust between the provider and supplier (Schnürmacher et al. 2015).

A second category that emerges from the study of Schnürmacher et al. (2015) is data and how it is collected, processed and analysed. The challenge is not with the recording the data, but in how it can be made available to the provider legally for analysis

(Schnürmacher et al. 2015) as the unavailability of data hinders the decision making and development of services (Kunttu et al. 2017). Schnürmacher et al. (2015) highlight the legal part of the data sharing by calling obtaining the allowance of data recording and usage essential as the data legally belongs to the customer. To gain the allowance, a good, trustful relationship is required and to maintain the relationship, the supplier must be able to demonstrate, that the data is not used to harm the customer by passing it to competition or by using it for internal purposes that are not explicitly agreed for.

In addition to gathering the data, it has to be analysed and processed for it to provide any benefit to either party (Schnürmacher et al. 2015). Preferably, this process should be automated. Analysis refines the raw data first into information and then further to knowledge and wisdom. This data, information, knowledge, wisdom (DIKW)-hierarchy was originally presented by Ackoff in 1989 (Kunttu et al. 2017).

Information is data refined to format, which is understandable to humans, such as characteristic values or graphs, knowledge is the capabilities to interpret the information and recognise the need for actions and wisdom is the skills to combine information and knowledge from different sources to support decisions and to compare alternative actions (Kunttu et al. 2017). Schnürmacher et al. (2015), however, presents the last two steps of the hierarchy as the competence and safe actions, but as they do not present substantial description on what these categories contain or represent, it is reasonable to assume that they are identical to the categories used by Kunttu et al. (2017). The wisdom gained from data-analysis is the precursor for creating functional service provision. All steps of the DIKW-hierarchy may also be offered as stand-alone services (Kunttu et al. 2017) or they can be used to design new PPS offerings.

2.2 Service value

Value, in the context of services, can be defined as achieving customers outcome, purpose or objective with service (Macdonald et al. 2011). Consideration of the value of the service provision is of utmost importance when designing the service offering as it provides the basis for the activities. Therefore, it is vital to determine how the value is created and how the creation can be supported.

The paradigm shift to service dominant thinking is reflected also in the discussion on value. The research that is published before the release of the original article on service-dominant logic by Vargo & Lusch (2004), or that is heavily influenced by the older paradigm, draw their ideas and premise from a foundation, which several scholars do no longer consider completely valid. However, this does not render the older material invalid as the key findings and concepts can still be applicable if the differences in the underlaying narrative are recognised.

The old, goods dominant, way is that the value is created in the manufacture of a product and destroyed at the use. This view, that stems from the economics, implies that the items would have an intrinsic value that can be measured in currencies. This can be seen from e.g. Khalifa (2004)s value exchange model that is also illustrated in the figure 2.

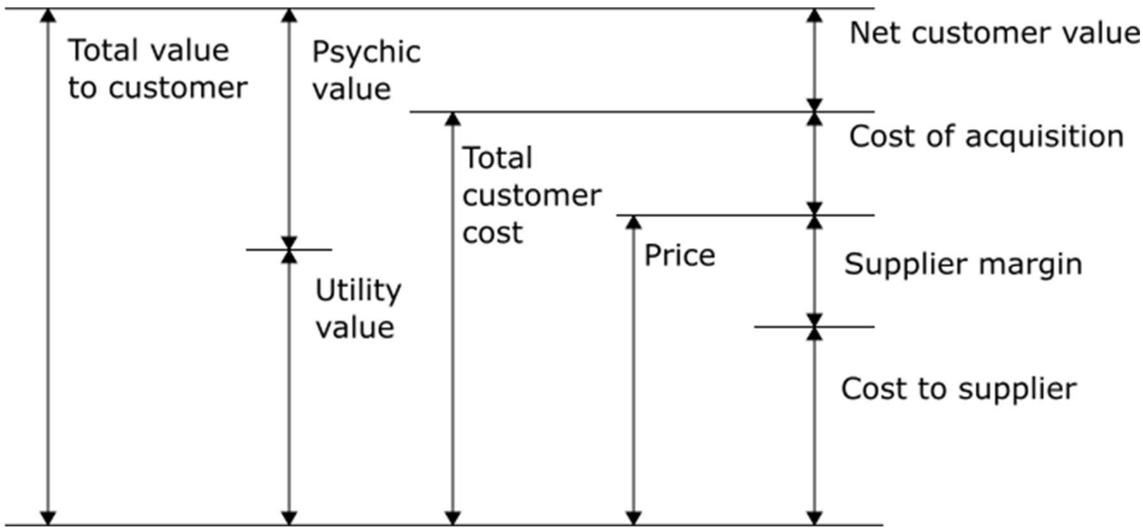

Figure 2: Value exchange model adapted from Khalifa (2004). In the figure the relation of value to various costs or sacrifices is illustrated

As can be seen from the figure 2, the total value to customer is the sum of various costs that the author also calls sacrifices and a net customer value, but it is also the sum of utility value and a rather vague physic value. The value-exchange model is accompanied with a notion, that the customer arrives to a purchasing decision only if the net customer value is greater than zero. This implies, that the customer is rational and infallible in the determination of the value and consistently is measuring the value of various value offerings and propositions. Yet, at the same time this must be done in a way that the cost of search does not expand and thus consume the net customer value. Moreover, to do so requires the physic value be assigned a monetary value, which can be rather difficult.

The value was equalled to revenue in the old paradigm, and thus added value was the same as the increase in revenue. Interestingly this led to notions that adding services to accompany the product would incentive the customer to pay more (Smith et al. 2014). While it is undeniable that service provision can increase revenue, it is not necessary for the compensation to come directly from the customer as a transaction or payment. Even a service that is free to the user can be profitable for the provider as Google and Facebook have proved with their array of digital services.

Modern scholars e.g. Pawar et al. (2009), Lusch & Vargo (2014) and Smith et al. (2014) view the subject differently compared to the old notions of G-D logic. Pawar et al. (2009) state the following: "Customers value of a product could lie in the benefits they attain from the product instead of product ownership". This is in line with the ideas of value embedded in S-D logic (Lusch & Vargo 2014). Pawar et al. (2009) continue with the implications of the view: “the provider could shift focus from the means of achieving such benefits to the benefits themselves”. This is essentially inverted paradigm of what was before, as prior value creation was focused on satisfying the needs of the customers predominantly with the manufactured products (Smith et al. 2014).

Instead of being intrinsic property of a physical item, the value is seen in S-D logic as something, which is co-created by multiple actors, including the customer, in the use of the service and the accompanying physical products. Moreover, the supplier can not deliver value, only take part in the creation and offer value propositions, or in other words facilitate the generation of value. The facilitation of value, which is the prerequisite for the value creation, encompasses the production, delivery and the back and front office activities (Grönroos & Ravald 2011). This implies that the industrial equipment is valueless until they are given function, that is, e.g. put to use or allocated to the stock of readily available spare devices. If this was not the case, the customer should be able to maximise the value they receive by simply purchasing a vast quantity of unnecessary devices.

The tenth founding premise of the S-D logic states that “value is uniquely and phenomenologically determined by the beneficiary” (Lusch & Vargo 2014). This implies, that the same service, and by extension product, is of different value to different users. Therefore, in combination with the dynamic nature of service provision, this means that ideally the value must be gauged constantly for each of the customers individually. In addition, even for the same customer, the use cases and requirements can vary therefore making each case of service provision unique event that requires a unique solution.

According to Pawar et al. (2009), three steps for the creation of value can be identified in the context of product-service systems. Firstly, the value is defined. This includes the identification of what is valuable to customer as well as what the customer needs and the cost of providing satisfactory solution to meet the needs. Secondly, the value offering is designed and the needed capabilities for providing the service which are required to be sourced or available within the organisation are identified. Lastly, the value is delivered through a network of partners, which are coordinated and controlled to ensure the performance of the delivery process.

However, as Smith et al. (2014) note, this model does not fully capture the S-D logic as it implies the value is determined by the producer. In addition, Pawar et al. (2009) discuss of delivering the value through a network of partners whereas Lusch & Vargo (2014) claim that “Actors cannot deliver value but can participate in the creation and offering of value propositions”. Moreover, in the model, the customer is a passive receiver of value, which contrasts with the idea of co-creation of value in S-D logic. According to Smith et al. (2014) customer needs to be treated as an active and accountable participant for them to be a co-creator of value.

Toossi et al. (2013) describe three different types of information, which must be obtained to satisfy the customer requirements. First of these is what customers are trying to get done when using the service and secondly, what customers are trying to achieve with the service. Thirdly the constraints or roadblocks that stand in the way must be considered. The two first points describe the function or utility from where the value emerges. Wisely the authors separated these points, as what the customer is trying to get done and what the customer tries to achieve are not always the same. This highlights the requirement of activity, collaboration and co-operation of the customer in the value co-generation.

As the customer should be an active participant in value co-creation, the resources of the customer are central to achieving the benefits and the end goals (Smith et al. 2014). Moreover, Pawar et al. (2009) state that the needs of the customer should determine how the service offering is built instead of the internal resources and capabilities, therefore the operant resources not present must be developed self or obtained elsewhere. This is in line with what Kindström et al. (2013) voice about the dynamic nature of service innovation where capabilities and resources need to be constantly improved. Considering the ideally active and accountable nature of the customer, this implies, that the customer also must improve their internal resources and capabilities to get best possible benefit. Moreover, as the value is uniquely and phenomenologically determined by the beneficiary (Lusch & Vargo 2014), the beneficiary, or the customer, needs to also evaluate their own performance (Smith et al. 2014). For this, knowledge on the activities is needed from all involved parties.

The knowledge has implications for both the customer and the supplier. On one hand, knowledge and skills to evaluate the performance of the services also helps the customer to better understand the value of the service, and therefore make better decisions regarding service purchases (Toossi et al. 2013). On the other hand, even though supplier is fundamentally the facilitator of the value, it may through interaction become co-creator of value (Grönroos & Ravald 2011). This is essential, as like Grönroos & Ravald (2011) report: "Value for supplier cannot be expected to be created from business engagement unless customers value generation is supported". Therefore, the customers should be contacted early in the service design process of innovative concepts so that the new ideas and expectations can be jointly refined (Kindström et al. 2013).

The focus on identifying and fulfilling the needs and requirements with right services is important concern, but it is only one side of the challenge; namely the market pull. The technology push, finding applications for new technologies, is equally lucrative. As the customer might not know how to achieve a certain function, or fully recognise their needs or possibilities offered by new technologies and methods, it is vital for the supplier to be able to communicate the value to the customer in a value offer. This value offering can serve as the basis for further innovation and latent needs can be discovered by examining the possibilities that can be realised with new technologies. Interestingly, study of Ali-Marttila et al. (2017) on what managers of manufacturing companies value in the maintenance services found out that communication and relations is not a dimension that the managers in general value. They do note that the well-functioning relationships are valued, however the elements that result in positive synergy are not recognised.

According to Ali-Marttila et al. (2017), the customers and providers of maintenance service fall into 3 distinct categories in terms of relationship and contract related attributes they value in a service. These are the collaboration-oriented partners, the basic partners and the quality-oriented partners. Collaboration-oriented partners are interested in forming deep relations with the service providers and recognise the value of doing things together, synergy and co-development of service. Basic partners, on the other hand, see the service as transaction without any special value. They simply order maintenance and wish to see it delivered as ordered. The third group, quality-oriented partners, are interested in the outcome of the service solution, but not so much in the co-development of services and relationships behind them.

It is important to note the composition of the categories, or what kind of companies fall into each of them. Collaboration-oriented partners were mostly medium sized service provider, while the customers fell into the latter two categories (Ali-Marttila et al. 2017). This suggests, that the service companies have embraced the new paradigm of service dominant logic, while it has not yet completely been adopted by the customers, or that the service providers are more eager to offer and develop services and relations than the potential customers to receive and use them. This is logical, as development of services is the core business for the maintenance companies, thus the companies have greater incentive to investigate new ways of working than the customers, who focus on their own core activities. Also the companies, which were doing financially well, were more interested in developing ideas and common projects, suggesting that if business is running smoothly, more time can be spent on such ventures (Ali-Marttila et al. 2017). As there could be a sort of asymmetry in the spread of the S-D logic, perhaps the suppliers could try to hasten the diffusion of the paradigm among the clients to enhance their amenability for engagement in mutually beneficial service agreements. The latter statement implies that financially affluent companies could be among the most receptive, thus they provide a suitable starting point or focus.

Toossi et al. (2013) studied how important certain value dimensions are for customers of maintenance service providers. Specialist knowledge was found out to be the most important aspect. Suppliers understanding of customer business and the customers desire to be in control of the activities were also considered important. This is also something that Vaittinen et al. (2017) found in their study. Of financial imperatives, cost savings have only low to medium importance, however the price is considered important. This suggest that the customers are focused more on the price rather than the potential savings. Keeping the customer informed with constant feedback and reporting can help them recognise the value. The quality of the maintenance was either considered indifferent or extremely important and the authors reason that if all of the providers in field can provide quality maintenance, the importance of the aspect diminished and vice versa. However, the service orientation, relational dynamic, interpersonal relations, and consistency in service were found out to be relevant to the customers. The wide range of products and service offerings was one of the key points along with accessibility, that is ease of access to services, and delivery. Locality, such as having an on-site representative was highlighted. (Toossi et al. 2013)

Laurila (2017) discusses the sources of customer value in services based on industrial internet as part of her thesis. As the industrial internet is the product and thus the vehicle of the service, the points should be applicable to any other service that provides value in same areas. A desire for value should be universal regardless of the vehicle of delivery. Indeed, as the author notes, customers are not willing to pay for the industrial internet itself, but for the benefit and concrete solutions that can be provided with it. Through interviews in a case study, some important aspects were identified. The streamlining, performance increases and the enhancing of the activities as well as the quality and risk management were considered important. The subfactors in these include energy and resource savings, remote use, the forecasting of the maintenance needs, faster reaction times in problem situations and the optimal, consistent quality.

It is important to note that there were differences in how the different companies and people answered in the studies of Ali-Marttila et al. (2017), Laurila (2017) and Toossi et al. (2013). In research of Ali-Marttila et al. (2017) this manifested in the discovery of the three categories of companies, and in study of Toossi et al. (2013) in the different magnitudes of importance the companies reported for each of the dimensions; such as one customer giving high value for quality, while the other gave it very low value. Laurilas (2017) study, on the other hand, identified concrete sources of value in a specific, yet insightful, case. This confirms the unique nature of value, which the beneficiary defines in use and suggests that even-though broad categorisation and ranking of value dimensions can be done to support general planning, what the customer actually values or needs must be considered individually for each case.

The concept of value in S-D logic resembles the postmodernist school of thought in philosophy and arts. Just as the truth in postmodernism, the value S-D logic is dynamic, relativistic and subjective (Weiss 2000; Iannone 2017). No value dimension is more valuable than other as no medium is more artistic than others and nothing is inherently more valuable than others. In contrast the value in G-D logic is more modernistic. It is something that can be objectively defined and quantified and it is inherently present in products.

These two schools of thought are distinct, and the difference is apparent and when the value is communicated. Following the idea of subjectively and phenomenologically defined value, especially important is how it is communicated. Should the customer think like a modernist, they would see the value objective and determinable and thus it should be communicated with charts, digits and calculations; evidence that supports the value. For a more postmodernist customer, the focus should be more on the experience and subjective side of things. However, these are extremes and most customers fall into a spectrum between them. While choices can be rationally justified through e.g. utility value or financial considerations, often the decision is made based on real, perceived or predicted subjective experience (Kaasinen & Liinasuo 2017). Indeed, service experience can be sustainable source of competitive advantage (Kaasinen & Liinasuo 2017).

2.3 Condition-based and predictive maintenance

Maintenance is a typical example of industrial services, as the operations are often at least partially outsourced (Ali-Marttila et al. 2017). In addition, it is one of the major services in context of life cycle management (Takata et al. 2004).

Lenahan (2011) defines maintenance as "the sum of activities performed to protect the reliability of the plant". For an individual item or device, the definition, of activities, which are required to keep the device in proper condition can be used (Kowalkowski 2006). The definition of Lenahan (2011) and Kowalkowski (2006), however, interestingly exclude the traditional reactive maintenance from the definition. In reactive maintenance, the device is serviced only after the failure has occurred, thus the actions to take are to bring the device back to the proper condition and restore the functionality of the plant and not necessarily to keep it in condition and protect reliability. Nevertheless, the definitions of Lenahan (2011) and Kowalkowski (2006) reflect well the modern view of maintenance as an important value creating function. Maintenance is usually grouped with related, but distinct activities of repair and operation or overhaul to form the concept of MRO. Interestingly, the "O" can be either operations or overhaul depending on context with the former used in managerial context and latter in technical context. For this work, the former context seems more appropriate.

If the maintenance is defined through the activities, it is necessary to specify what exactly these actions are or could be. Takata et al. (2004) list the following activities as constituents of maintenance: maintainability design, maintenance strategy planning, maintenance task control, evaluation of maintenance results, improvement of maintenance and products and dismantling planning and execution. Maintainability design refers to improving the design in product development phase and providing design data for maintenance strategy planning and task control. In maintenance strategy planning a strategy for maintenance is selected and in task control the actions and capacities are planned, scheduled and executed based on strategy. The results are evaluated to determine if the taken actions are appropriate and the results of evaluation are used to improve the strategy, maintenance process and importantly even the product. Lastly, the planning and execution of dismantling the product at the end of the life cycle as well as replacing it with new solution are also parts of maintenance (Takata et al. 2004). When defined in this manner, the maintenance not only covers the whole life cycle, but also overlaps with life cycle management.

The effectiveness of maintenance is highly dependent on the maintenance strategy planning (Takata et al. 2004). This is a paramount concern as efficient planning leads to lower cost of ownership (Banks et al. 2009). The function of the strategy planning is to select the best combination of maintenance strategies. In addition to the reactive, or run-to-failure, maintenance, the options for maintenance strategy include the scheduled, or Time-Based Maintenance (TBM), and the Condition-Based Maintenance (CBM) (Takata et al. 2004).

In condition-based maintenance (CBM) the devices are monitored, and maintenance actions are scheduled once the condition deteriorates below a set threshold (Efthymiou et al. 2012a). Predictive maintenance is similar to the condition-based maintenance, only the time-horizon is different and according to Susto et al. (2012), several authors combine the categories and indeed e.g. Mrad et al. (2013) treat the concepts as synonyms. However, while what applies to CBM also applies to predictive maintenance, the predictive maintenance has enough unique characteristics to be constituted as a separate practice. On the other hand, due to the similarities, it is reasonable to also treat predictive maintenance as a refined form of CBM.

The goal of CBM is organise and plan the maintenance with real-time assessment of the condition of the devices (Ahmad & Kamaruddin 2012). In addition, in predictive maintenance a prognosis, prediction of damage that is yet to occur, is made (Ahmad & Kamaruddin 2012; Efthymiou et al. 2012). Prognostics consists of prediction of remaining useful life (RUL) and the estimation of the confidence interval of the prediction to determine the accuracy of the prediction (Efthymiou et al. 2012). Of the methods, the predictive maintenance is more effective as it allows prevention of unexpected failures (Ahmad & Kamaruddin 2012).

Yet, while CBM is highly regarded, it is not the best method of maintenance; not even in the cost effectiveness according to Takata et al. (2004). When the failures of the devices are not critical, the breakdown maintenance can be allowed (Takata et al. 2004). This requires, however, the criticality to be first determined and then assessed. Nevertheless, looking at the extremely non-critical devices one might find in a working environment, such as easily replaceable hand tools or the coffee maker in break room, it is clear, that highly advanced techniques are not warranted in all cases. In addition, when the lives of the devices can be precisely estimated, TBM is the most effective of the options (Takata et al. 2004), yet the precise estimation can be challenging. Very few systems degrade linearly with time, or in respect to any other variable in fact. On the other hand, the condition-based predictions that are linked to CBM do provide estimates that can be precise and thus might be usable with the scheduled maintenance as well. Nevertheless, authors such as Hakanen et al. (2017) claim that CBM creates concrete and easily verifiable savings for customers in spare part and maintenance expenses.

The CBM encompasses three constituent steps; namely, the data acquisition, the data processing and the maintenance decision-making (Efthymiou et al. 2012). The first two of these can be described as monitoring, which is the observation of the actual state of the devices (Ahmad & Kamaruddin 2012). Data acquisition is the collection of the information and its purpose is to obtain relevant data about the health of the system. Two main categories of data can be identified. First of these is the event data, which consists of observations on events, such as failures, and the reasons for them as well as the actions done in response to the event. Second category is the data which indicates the condition or health of the devices and this can be measured parameters such as temperatures or pressures (Efthymiou et al. 2012). The data collection can be performed either on-line, that is during the operation of the equipment, or off-line, when the equipment is not in use (Ahmad & Kamaruddin 2012). Monitoring and observations can be done with fixed measurement equipment, hand-held meters, or even with human senses (Ahmad & Kamaruddin 2012).

The monitored properties can include quantitative measurements such as vibration, acoustic, electric and temperature measurements and oil analyses, and qualitative measurements, such as dirtiness, leaks, or colour aberrations detected with human senses (Ahmad & Kamaruddin 2012). Essentially anything can be measured, but due to data storage and transfer considerations, the focus should be on measurements that can be reasonably assumed to be linked to the device condition. On the other hand, a large amount of seemingly unrelated or useless data may prove to be useful in the future in the same or different application. Efthymiou et al. (2012) report that the data is usually not systematically stored, or it cannot be easily retrieved. The automated data collecting and storage capabilities of internet of things (IoT) or industrial internet could potentially rectify the problems.

According to Efthymiou et al. (2012), the data acquisition is hindered by the simplicity of the sensory systems. In addition, the measuring devices are expensive due to specialised system required and the large data-flow is susceptible to noise (Ahmad & Kamaruddin 2012). However, these two statements seem contradictory as the advantage of simple measuring devices in general is their low relative cost.

Data processing is the handling and analysis of the collected data (Efthymiou et al. 2012). The key function is to improve the understanding and interpretability of the data. Data is first cleaned to errors and noise after which it is analysed with selected methods, such as principal component analysis, expert systems or AI methods. The AI methods have been found out to outperform the conventional ones (Efthymiou et al. 2012). Data processing, along with acquisition, is important consideration not only in the CBM, but also in the service planning. Thus, if condition management was provided as a service, these steps would serve both the service planning and reporting as well as the maintenance management. Therefore, to have capabilities to reliably and consistently collect and analyse data should be of great interest to both the client and the provider of services.

The maintenance decision-making is about the selection of the appropriate actions. This can be split to diagnostics and prognostics (Ahmad & Kamaruddin 2012; Efthymiou et al. 2012). If considered as separate practices, this is where CBM and predictive maintenance would differ. Diagnostics is the identification of patterns that are related to the immediate failure or abnormal operation of the equipment. This includes three steps, which are the fault detection, the fault isolation to determine which component is failing and fault identification to determine the nature and magnitude of the fault (Efthymiou et al. 2012). However, for devices or assemblies with few components that are not easily repairable without complete disassembly, the isolation of the failing component is rather unnecessary. When the device is maintained, also the other components are replaced or repaired as they also have been subject to wear and degradation. Once the abnormal operation condition of the equipment has been detected, the equipment can still be run for a certain amount of time before the device fails, that is fails to perform the intended function in adequate manner (Ahmad & Kamaruddin 2012).

The decision making can be carried out with two distinct methods, the current condition evaluation-based (CCEB) and the future condition prediction-based (FCPB) (Ahmad & Kamaruddin 2012). In CCEB, the current condition of the device is evaluated and compared to a predefined failure limit. If the condition exceeds or meets the limit, the maintenance is performed. The difference to diagnostics is that the failure has not yet occurred. In FCPB the future trend of the deterioration of the equipment is formed and if the trend crosses a set limit, the maintenance is scheduled.

Both the CCEB and FCPB have limitations. In CCEB, the timeframe between the deterioration surpassing the limit and the failure can be too short for any proper planning of maintenance if the interval of updating the data and calculation is too long. FCPB on the other hand is useful only for short-term predictions if the prediction is unreliable (Ahmad & Kamaruddin 2012). Therefore, it is essential that the prediction is accurate for a long timeframe so that proper decisions can be made, and maintenance actions be planned well ahead. According to Ahmad & Kamaruddin (2012), the CCEB and FCPB both rely on set limits for the decision. How the limits are set is however left ambiguous. However, it can be assumed that the limits have to link to the degradation of the device. Takata et al. (2004), in their article, present a figure that is also presented here as figure 3. In the figure there are two limits, the detection and the functional limit. The detection limit refers to the first moment when the abnormal condition of the device can be detected and the functional limit to the point where the device can be considered as failed. For FCPB, the functional limit, with added margin to cover for the errors in prediction, can be considered as the set limit. The limit for CCEB, however, is not so obvious. Setting it at T_d, the first sign of abnormal deterioration, can be too early if the deterioration pattern can be detected at a very early stage. Hence, the limit needs to be defined with the help of the functional limit and a desired margin. The margin in this case should be long enough so that it covers the desired maintenance interval or be long enough to for the management to be able to schedule appropriate actions to avoid the aforementioned limitation of too short timeframe. How long the timeframe should be at minimum depends on the case and sets an important boundary condition for designing the predictive model.

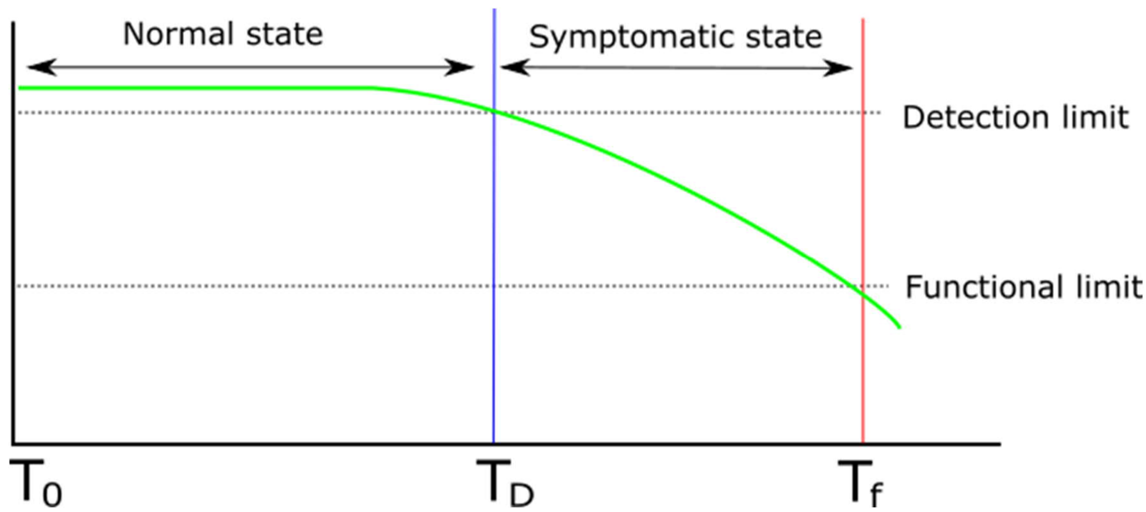

Figure 3: Pattern of deterioration with the detection and functional limits. Adapted from Takata et al. (2004)

The prediction can be done in multiple ways. According to (Ahmad & Kamaruddin 2012) various methods e.g. vibration signal analysis, neural networks, voltage mismatch techniques and state-space models with filtering have been successfully applied to various use cases for prognostics ranging from aircraft to electric motors. However, any general model, that would work for all different cases is not present in literature, suggesting that the selection of the prediction methods is highly case specific and therefore, must be adapted to the unique characteristics.

The methods can be classified into qualitative, quantitative, history-based and hybrid methods (Ribeiro & Barata 2011). Qualitative methods are based on examining deviations or inconsistencies between the output of the actual control system and expected output produced by a model describing said system. Downside of this approach is the futility of building a model that replicates the actual system exactly due to unexpected reaction and unperceived interactions. The difficulty of accurate modelling increases rapidly as the system gets physically complicated or the quantity of variables increase (Ribeiro & Barata 2011). The natural conclusion is to keep the number of variables low. This could be done, for example, by splitting the system to smaller constituent parts for separate evaluations and forming the total assessment of the system as a sum of the smaller parts.

Qualitative methods produce models, which are not mathematically describable. Instead, the information is represented in logic or symbolic way (Ribeiro & Barata 2011).

Qualitative models can be formed with abduction, which is the generation of explanations from observations, and induction, that is learning with historic data. The model can be then constructed to represent the information is structured or functional way. Structured models represent the connections between the system and subsystems, while functional representation displays means-ends relationships between the subsystems. Structured models are used in situations where relationships are easy to map out, while functional ones perform better if some part is too complex to represent as a coherent structure. Notable methods of qualitative modelling are the expert systems and fault tree analysis. (Ribeiro & Barata 2011)

History-based methods form the model from existing data. Ribeiro & Barata (2011) report the availability of historic data of systems as one of the advantages. The other advantage is the speed of modelling, which is faster than building the models manually based on the experiences of field expert or domain knowledge. History-based methods have some overlap with quantitative methods as they include learning expert systems. Some notable history-based methods are machine-learning algorithms such as neural networks and hidden Markov models.

2.4 Prognostics and fault detection with machine learning

Prognostics is widely researched field and several studies have been done to search for methods, which can be used to predict future failures. The approaches employed by the researchers are various, including for example physical modelling, signal analysis and data-driven methods. It is widely reported, by e.g. Zhang et al. (2017), that while approaches based on phenomenological modelling can be more accurate when well designed, they do not provide cost-efficient results due to the inherent complexity of the physical world. Hence, in practise, such models are generally unavailable. However, at the same time, data-driven methods have been found out to be both cost effective and sufficiently accurate.

Data-driven methods for fault prediction and prognostics are a relatively new field as advanced machine learning capabilities have become readily available only recently. Several different approaches have been attempted. Data-driven model can be constructed by traditional statistical methods or by utilising machine learning. Statistical methods include various methods from basic linear regression to more complex filtering and prediction. Some examples of application of these methods are regression analysis with particle filters (Susto et al. 2012), on-demand regression (Yang et al. 2017), support vector regression (Gokulachandran & Mohandas 2015) and logistic regression (Costello et al. 2017). Machine learning methods range from basic support vector machines to complex neural networks. The difference between statistical methods and machine learning is that the statistical methods require the user to impose a specific structure for the model and estimate suitable parameters that could fit the data, while machine learning methods do not (Raza et al. 2016; Marinelli et al. 2014).

Machine learning based classifiers are rudimentary machine learning methods. They split the data into distinct groups based on the rules that are learned directly from data. Some examples of machine learning classifiers employed in prognostics and fault detection are support vector machines (SVM) (Costello et al. 2017) and k-Nearest Neighbour (k-NN) (Susto et al. 2015). Several classifiers can be used together to detect different faults simultaneously like Susto et al. (2015) do in their study. In addition, some classifiers, such as decision trees, can be used as an ensemble, which is a single classifier that is built out of several separate classifiers of same type.

Artificial neural networks (ANN), or just neural networks (NN) for short, are more complex machine learning methods than simple classifiers. Neural networks vaguely replicate the functionality of a brain to an extent. A NN consists of nodes arranged in layers and connections between the nodes. Nodes and connections are analogous to synapses and their connections in a brain. The nodes receive their inputs through these connections, perform a function or an operation on the inputs and then output the results through the connections to following or in some cases preceding or the same layer. Simple feed forward artificial neural networks, such as single or multilayer perceptrons in prognostics have been researched by e.g. Raza et al. (2016) and Costello et al. (2017). Neural networks can also be used in conjunction with fuzzy logic such as in the studies of Jayaswal et al. (2016) and Gokulachandran & Mohandas (2015).

Advanced neural networks that employ convolution and memory, such as LSTM (Long Short-term Memory), have been discussed by e.g. Li et al. (2018) and Wu et al. (2018) respectively. Convolutional Neural Networks (CNN) use layers that apply a convolutional operation to the inputs, while the LSTM has capability to store and forget values. Convolution is widely used in image detection applications as a way to extract features automatically, while the LSTM has been used in time variant scenarios, where the desired output is dependent on the current input as well as historic events.

In table 1 a selection of studies that resemble the case of this work in one or more ways is presented. The main criterion was to pick studies that employed advanced methods on cases where several measurements were taken on multiple devices or machines. In addition, the studies collectively encompass a wide range of machine learning methods from simple ones to more advanced. It is important to note that, like e.g. Costello et al. (2017) report of their findings, the models built with data-driven methods are highly case specific, even to the extent that the model is valid only for a single unit. However, general observations on the methods and their relative strengths, performance and advantages can still be made.

Table 1: A selection of published research on data-driven prognostics.

Source	Methods and Application	Findings	Gaps
Raza et al. (2016)	Comparison study of PCA, MLP, Feed-forward NN, SVM and LR on strainers of oil pumps with 6 measured variables.	NN outperforms the alternatives both in accuracy and versatility. The approach can potentially be useful for planning maintenance tasks.	The insight on using the results to aid management and planning is vague as it is not the main topic of the research.
Yang et al. (2017)	On demand regression on starter motors of an aircraft auxiliary power units with 6 measurements and 12 other attributes.	The authors report good prognostics performance and significant improvement on basic regression analysis.	The method requires data pre-processing and feature selection. The applicability of the findings and method to practice is questionable and performance is not compared to other advanced methods.
Chen et al. (2017)	Application of fault trees and fuzzy NN in an IoT application in aquaculture with a total of 22 direct measurements and derived attributes.	The applied model can diagnose most kinds of faults and provide users information and suggestions for maintenance.	The model cannot provide predictions of future faults. The authors report the data gathering and model refinement as potential areas of future development.
Li et al. (2018)	Comparison study of convolutional NN to other methods using C-MAPSS data. Comparison was done both with results form literature as well as calculated ones.	The presented convolutional NN outperformed other alternatives, including other state of the art methods. The accuracy of predictions was excellent.	Data was normalised, and constant sensor data was removed. The study was done on simulated data. No insight on practical implementation of presented method.
Wu et al. (2018)	LSTM NN with C-MAPPS data.	LSTM NN provides high accuracy and versatility in prognostics.	Data was normalised and pre-processed. The research process generalised as much as possible and is therefore not easily transferable to other sets of data or applied in practise.

In the table 1 three of the entries are aircraft engine related. The aircraft engines are a significant area of research in prognostics, as the aviation industry is very capital intensive and technologically advanced. In addition, Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset, is a popular and open dataset that contains simulated data on aircraft engine failures provided by NASA (Saxena & Goebel 2008).

The set consists of four sets of simulated turbofan condition monitoring data. In the simulated data, the defined failing point is preceded by a steadily growing indicator, more specifically an increase in the amplitude of the monitored signals, such as ever-increasing temperature swings. This reflects the actual deterioration patters with reasonable accuracy; thus it is usable for testing and comparing different methods.

Li et al. (2018) compiled from literature a comparison of methods. The studies of several authors all used the same data set, namely the C-MAPPS, provided by NASA. The findings are presented in the table 2. In the columns of the table, the method employed, the original source, the root means square errors (RMSE) obtained for four different datasets (FD001 to FD004) and the R early are listed. The RMSE indicates the error in the prediction, which negatively correlates with the accuracy of the prediction. The R early describes the earliest timestep that the researchers specified as the borderline between normal operation and abnormal operation that indicates upcoming failure.

Table 2: Comparison of results obtained in various studies with various methods of failure prediction on C-MAPPS data. Adapted from Li et al. (2018)

Method		**RMSE**				
	Authors	**FD001**	**FD002**	**FD003**	**FD004**	**R early**
Echo state NN with Kalman filter	Peng et al. (2012)	63.46	N/A	N/A	N/A	N/A
Support vector machine classifier	Louen et al. (2013)	29,82	N/A	N/A	N/A	N/A
RULCLIPPER	Ramasso (2014)	13.27	22.89	16.00	24.33	135
LSTM	Malhotra et al. (2016)	12.81	N/A	N/A	N/A	125
Deep CNN	Badu et al. (2016)	18.45	30.29	19.82	29.16	N/A
Time window based NN	Lim et al. (2016)	15.16	N/A	N/A	N/A	N/A
Multi-objective deep belief networks ensemble	Zhang et al. (2017)	15.04	25.05	12.51	28.66	N/A
Random forest	Zhang et al. (2017)	17.91	29.59	20.27	31.12	N/A
Gradient Boosting	Zhang et al. (2017)	15.67	29.09	16.84	29.01	N/A
Convolution with rectified labels	Li et al. (2018)	12.61	22.36	12.64	23.31	125
Convolution without rectified labels	Li et al. (2018)	13.32	24.86	14.02	29.44	N/A

From the table 2 several observations can be made. Firstly, neural networks and other more advanced methods, apart from echo state NN, outperformed the basic SVM classifier. The findings of the comparative study of Raza et al. (2016) also support the notion of high performance of neural networks over simpler methods. However, in their study, the NN employed was a rather simple feed forward network, which required manual feature extraction to function. With feature extraction, the amount of data needed to train the network can be lowered (Wu et al. 2018), and it is a requirement for some methods to function. However, the selection of features to be extracted needs to be carefully planned and executed, which takes extra effort. More advanced convolutional networks, such as the ones used by Li et al. (2018), learn the patterns of feature extraction directly from data.

Secondly, ensemble methods, the random forest and the gradient boosting, are very close to the performance of the neural networks. However, these methods require handcrafted features as input, which limits their versatility and adaptability to new scenarios unless the features are redefined.

Thirdly, RULCLIPPER, which is not a machine learning method but rather a case-specific purpose-built prognostics model based on computational geometry, matches the performance of the neural networks (Ramasso 2014). However, modifying the purpose-built program to suit other sets of data or application can be difficult. Therefore, while RULCLIPPER functions well in this case, computational geometry as a method cannot be concluded to be particularly lucrative approach. However, its success could warrant additional investigation in a separate research.

Fourthly, the NNs with either LSTM or convolutional layers provide the best performance. This is not surprising, as both are deemed to be among the state of the art in machine learning currently. In addition, LSTM and convolution are among the most versatile methods as they do not require manual feature engineering. However, while the convolution NNs seems to only have a small advantage over LSTM, recent study of Bai et al. (2018) indicates that in general time series modelling tasks networks employing highly advanced convolution perform significantly better. The researchers do note, however, that until very recently convolutional networks were weaker, which might explain the small difference in the presented comparison.

The novel methods, e.g. the on-demand regression presented by Yang et al. (2017) and combination of fault trees, NNs and fuzzy logic by Chen et al. (2017), are not present in any performance comparison that would provide results that are comparable to the more usual methods. In case of study of Chen et al. (2017), this is due to the highly case specific nature of the implementation as the system was developed with the single application in mind and the details of implementing a similar system to other cases would result in a vastly different performance. In case of Yang et al. (2017), a reliable comparison to other advanced methods is simply absent. The findings are compared to simple global regression, which other research already deems to be a wholly insufficient method for accurate prognostics. Despite the claimed success of these methods, there does not seem to be any incentive or substantial reason to attempt to adapt these to practise over other proven methods.

Data-driven methods require data and computational power to train the model. The more complex the model is, the higher the requirements are Wu et al. (2018). While the complexity of the model is not directly proportional or related to the accuracy, the more complex models tend to achieve higher precision. Some authors e.g. Ramasso (2014) list the speed of training as one of the potential development area, which is an interesting concern as that can be rectified with increase in computational resources, although at an expense of increased cost. In addition, training of the model is only done once, after which the model can be used for predictions indefinitely. The more crucial matter is the data, as it can neither be conjured from thin air nor sourced as easily as calculating power.

The application of the methods to practise outside the respective cases is not a point which gathers great attention in the select research. This is understandable, as the foci of the studies is either the application in the select system or general evaluation of the capabilities of different methods. In addition, the insight on financial aspects and business goals is wholly absent. This is unfortunate, as according to Mottaghinejad (2017), the choice of a model is affected greatly by the business goal as the outcome has to answer the problem at hand. On the other hand, the model built, and methods employed greatly affect the output and therefore limit the possible business objectives; insufficient model may not be able to provide the knowledge that is needed and conversely a great model may be able to provide insight more than was originally anticipated.

Moreover, the selection process of the methods employed and the criteria on which the authors chose the models to use is not presented in detail in most the studies. Costello et al. (2017) selected the methods in their study based on what had already been investigated and found out to work in similar cases. However, this approach is problematic as the field of machine learning is developing fast. Any method that is presented in literature, even in the recent publications, can already be subpar when compared to the current state of the art. Indeed, in their research tried and functional, but underperforming, methods were used.

2.5 Synthesis

The literature review consisted of four distinct, but interlinked subjects. This work handles the topic of maintenance as a service, on which the concept of industrial services provides the context, the value of the service the purpose, the maintenance practises the framework of actions, and the data-driven prognostics the means. Interestingly the past research on the subjects is focused on only one field at a time with other getting very little or no mention. Thus, while there is extensive research on each of the subjects, the insight of how these all link to each other in a practical situation is absent.

Industrial services, defined as application of operant resources to the benefit of the customer (Lusch & Vargo 2014), are a strong area of development and growth for many manufacturing companies that embrace the prospects of servitisation. Incorporating physical products, data analytics and activities, such as maintenance, into a coherent service offering as a PSS can change the way the business is conducted and thought about while increasing not only customer satisfaction but also revenue and green values.

Provision of a maintenance service based on data requires data from the customer. This is an obvious statement, yet it adequately reflects a major hurdle. The challenge is twofold; on one hand, the data must exist and on the other hand the customer must be willing to share it. The former requires adequate data gathering and storage capabilities and the latter forming an agreement of intercompany data transfer and a trustful reciprocal relationship. While the former is often in sufficient state, the latter requires more attention (Kunttu et al. 2017). Deep and functional relations to customers are therefore necessary to gain access to data.

The data is then processed or analysed to produce information, which in this case is information on degradation and remaining life of field devices. For data-driven prognostics, the analytic focus of this work, the literature suggests that the most efficient ways are machine learning methods, particularly neural networks. Especially the more advanced ones, CNNs and LSTM show great promise. Hence, these should be most capable methods for searching for and subsequently constructing a prognostics model.

The information can then be further refined to knowledge and be used to aid the design of services, aid decision making and support planning. It is indicated, that the management of maintenance is one possible area that can benefit from this kind of information. The systems and practices vary between the clients, thus the information that could be beneficial varies also. Therefore, the services provided should be either tailored on customer basis and the knowledge to support the service design should either be case specific or be general enough to provide usable information in wide array of cases.

Value is something that is uniquely and phenomenologically defined by the customer. This further reinforces the need to tailor the details of services on customer basis. Certain characteristics, such as a desire to manage the risks and to stay in control along with the great efficiency and optimal quality of the operation can be identified as something that can be deemed almost universally beneficial, desirable and thus valuable. On the other side, the loss of control and uncertainty were found as some of the major adverse, value reducing, dimensions. In addition, there is a so-called utility value, that can be easily calculated as increase or decrease in costs and which is a useful tool in communicating the value proposition by supporting the rational side of the decision making. Nevertheless, the often the subjective experience of service is what influences the decision the most.

These are also some advantages condition-based maintenance offers over the scheduled and reactive maintenance practises, yet, due to the established nature of value, a logical conclusion that the CBM is the desirable solution in a general case is unwarranted; this must be examined independently for each of the cases.

From what is stated previously can be inferred that data has connection with value in a way that is illustrated in the figure 4. Even though the connection leads from data to value, the two are far from each other. Therefore, it can be hard to see the potential value that can be gained from refining the data and hence the obvious incentive to improve data collection and subsequent steps can be low. Nevertheless, as stated, these actions are required for the service provision and thus the co-creation of the value. This implies, that the potential value must be communicated well to establish the proper incentive to develop the required capabilities of data collection, analysis and decision making. How the value is communicated, presented or appraised depends on the needs of the clients. It is noteworthy, that the value rises from the function, result and solution and not from the technical details. That is, what is done matters and not how it is done.

Existing capabilities, or the lack of them, should not restrict the search of optimal service solutions; what the provider can currently provide, is a secondary consideration. The capabilities and offerings that are found lacking should be developed. On the other hand, the existing capabilities can provide a sufficient starting point and they can be used to demonstrate the know-how of the supplier and potentially easily gained value and to form a service provision, which can easily be implemented in multiple cases.

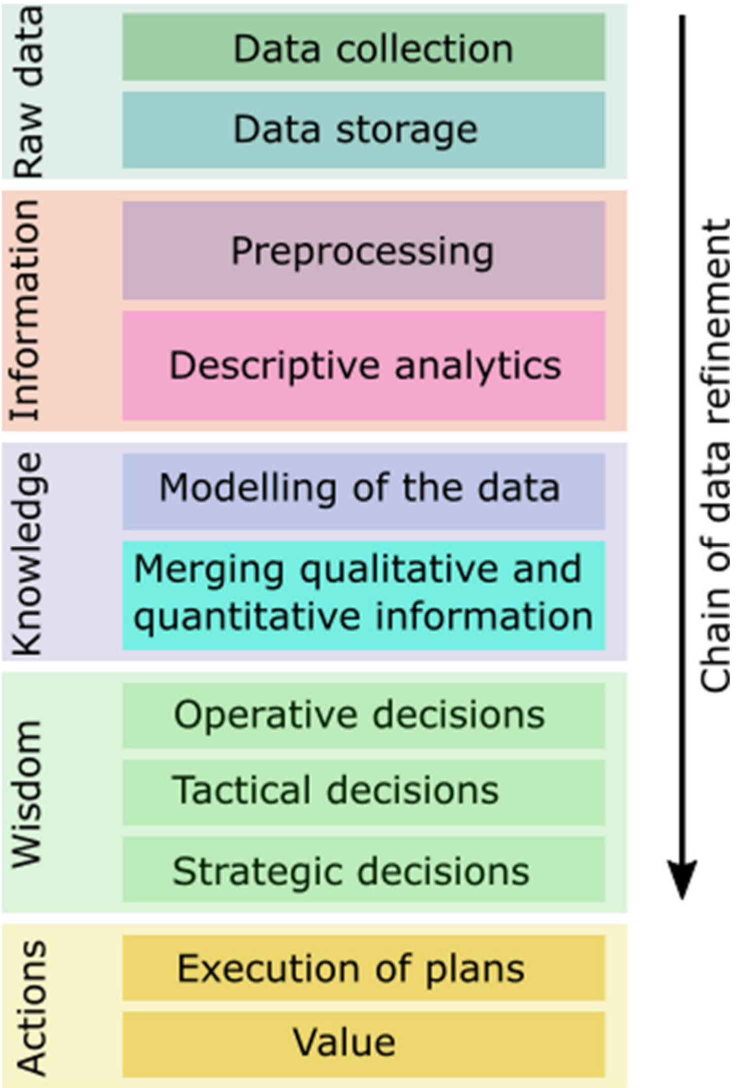

Figure 4: Connection of raw data and the value and the constituent steps between them in business context

Another point that can be made, is that the chain of figure 4 must be robust. Any missing or subpar performing link will jeopardise the whole process and therefore also the value generation. An extensive analytics capability is of no value, if the data acquisition is crippled or if it does not produce information that can be used in the latter phases of the chain. It is therefore necessary to have all the constituent steps in a balance and focus the development on the weakest of links. Moreover, to be able to communicate the value, the supplier must be able to demonstrate that all of the steps are actually doable in a scope that provides sufficient value creation that outweighs the effort that is put to develop the capabilities. Simply, the system has to be proven to be feasible and worthwhile.

3. RESEARCH METHODS

3.1 Methodology

It is undeniable, that the adoption of machine learning has resulted in considerable advances and significant innovations in several fields. The data and advanced analytics have potential not only enhance and support the business but also explore new ways of working and new business opportunities. Moreover, a growing interest in digitalisation is detectable. However, the rate of adoption has not been as rapid in all fields and industries as it has been in e.g. IT. Additionally, a large portion of the research is either theoretical or has been done using simulated data sets with practical case studies being in minority. This study examines how applicable the machine learning techniques are in the field of valve prognostics and how they can be used to support service business through a practical case study. The details of the case are further discussed in 3.2.

The study follows the chain of data refinement presented in chapter 2.5. By following it in practise, the robustness of the chain is examined in this applied case and thus the possibilities in which the data can support value creation for customers. The focus of the work is on the generation of service recommendations and objective of is to develop a rudimentary working predictive model, minimum viable product of sorts, to examine the ways through which one may be contrived. This includes a very high level of practical focus; thus, the realistic evaluation is chosen as the methodology for this work.

Realistic evaluation is a methodology, which mixes theoretical and practical thinking and where the conclusions are made based on the observed facts and set evaluation criteria. The realistic evaluation relies on pragmatism and realism. Together these define the truth as something which works, while rejecting the impractical and fictional. Anything that may be considered true, must be proved to be so through empirical practical demonstration; this is by examining the different choices, executing the actions while discussing the made decisions and the process which lead to them (Anttila 2007).

The realistic evaluation accepts context sensitivity, meaning, that that which might work in one setting, might not function in different time, place or situation. Hence the effort can be directed away from finding a general solution that works in all situations, to finding solutions that work e.g. in a defined setting for a defined group of customers (Anttila 2007). This, along with the focus on practical and proving the solutions through empirical demonstration make the realistic evaluation suitable methodology for commercially oriented development within the industries; a category under which this work also falls. Diverting the focus away from the general or universal solution allows the research of the practical process of building data-driven applications. This is important, as the challenges are not fully known. Moreover, by establishing the process on one specific case allows the subsequent examination how the methods can potentially be scaled up to incorporate several cases at once and what areas of interest should be developed for that.

The process presented as chain of data refinement is reflected in the structure of this research. Firstly, the data is collected, then processed and lastly analysed and modelled. The specific details on how these steps were done are presented in subchapters 3.3, 3.4 and 3.5 respectively.

3.2 Case description

The research questions and the problems are approach as a case study consisting of a single case. This study is an intrinsic case study as the case is selected due to its interest to the researcher and organisations involved (Crowe et al. 2011). The case represents a situation in a typical company in the field of manufacture rather than a special or critical case as described by e.g. Siggelkow (2007). This allows the study of practical capabilities, challenges and opportunities in a typical situation in the industry as opposed to the niche, cutting edge or atypical scenarios. Moreover, this allows the focus on the practise rather than purely theoretical, as the needs and desires of the customers, the data which is available and the situation itself is very real. This allow the investigation of how the data-driven prognostics and condition analysis can be offered as a service in an actual business environment. It is this latter part that has been absent from large number of the previous research.

The case organisation is the Valve Services department of Metso. Metso is a world-leading manufacturing and services company, which offers equipment and services in the mining, aggregates, recycling and process industries. Globally Metso employs over 12,000 people in more than 50 countries. Within the company, Metso Valve Services provides support on field services and services sales globally through the sales and services channels, but also directly engages with the clients. Valve Services is also responsible of developing new service processes and concepts as well as tasks related spare parts of field devices.

Field devices referred to in this work are industrial valves and related equipment, such as actuator, positioners and instrumentation equipment. Other field equipment, such as pumps, are not considered in the scope of this research. The function of a valve is to regulate and control the direction and volume of the flow. Actuator, which may work with pneumatic, hydraulic or electric power, is responsible of opening and closing the valve. Positioner communicates with the process control systems and drives the actuator if a change in valve position is required. Instrumentation equipment, which includes for example limit switches and pneumatic booster packs, enhance the capabilities of other equipment or provide additional signal inputs for positioner and process control systems.

The Metso Valve Services have developed a state-of-the-art installed base data storage and management systems and thus are among the industry leaders in installed base data management. The data is already used to automate certain parts of service processes, such as to automate the selection spare parts for devices customers have installed.

The case organisation is seeking ways to improve the maintenance services that are offered to the customers of the equipment with data analytics. The area of interest in which this research focuses is the provision of better service recommendations to support maintenance planning. The assumption is, that the data from various sources could be used in conjunction with machine learning to create device specific recommendations automatically if the condition of the device can be deduced from the data and maintenance need can be forecasted well in advance. This, in turn, should not only reduce manual labour in planning phase, but also increase the quality of service to customers.

From the onset of the research, it was clear, that data, that is related to the condition of the devices was needed. This is data that is produced by various sensors and devices within plants in productional use and stored in the data systems of said facilities. The data is intellectual property of the owners of the plants, thus explicit permission to use it was needed.

It was not deemed necessary for all the data to come from the same customer. As the data is intellectual property of the plant owners, a specific permission and assistance to access the system was needed to collect the data. This also limited the possibility of combining similar data from different plants. This proved to be a slight limitation to the extent of data that was available for use in this research. It was nevertheless deemed to be worthwhile to carry out the research with the data from a single plant. Should a promising result that warrants further development and research be found, the results can be used as leverage to secure more permissions to use data.

The main requirement for selecting the customers that were contacted was the suspected availability of collected historical data for long enough period for substantial number of devices as data-driven methods rely on large amounts of data to successfully train a working predictive model. Moreover, the customers to which the case organisation had the closest relations and that were suspected to be willing to participate in this endeavour were prioritised. Thus, only select few operators of large productional facilities, 8 in total, were considered and contacted. Selected customers were contacted through the sales channels and sales representatives. The customers that were contacted were selected based on few important key characteristics, namely they had to operate large facilities, operate Metso FieldCare and preferably Expertune PlantTriage, and have in the past expressed interest in sharing data, using analytics or have been collaborative in other ways.

The Metso FieldCare is a solution comprising of software and necessary communication interfaces that can read data from the valve positioners through the automation system using standard communication protocols, such as the Highway Addressable Remote Transducer (HART) protocol. It is used for monitoring anomalies and error messages during normal operation and it alerts the users should one occur. The FieldCare is capable of reading data from several different models of positioners, each of which records and sends different measurements. Some examples are the supply line pressure or the counter of valve travel.

The FieldCare is used in conjunction with APM-ticketing system, where the technicians can record the events of anomalies and the actions taken to resolve them. The PlantTriage is a system which collects measured data from the plant instrumentation, such as signals from temperature and pressure instruments. It is used for fine tuning the process to optimise it for better efficiency. The requirement to have both FieldCare and PlantTriage in use were set so that the data that could be collected would be as extensive as possible in both quantity and quality. The APM data is further referred to as event data while the FieldCare and PlantTriage data is referred together as measured data.

Ultimately, only a single customer agreed to grant the access to the condition monitoring dataset. This dataset consisted of actual recorded data collected during the operation of the plant. This customer operates in an industry in which a significant quantity of Metso products is used; and in this sense the customers facility represents a typical operating environment for the devices. The customer is known for their efforts in research and innovation as well as the quality of the end products. Moreover, the practices and systems they employ are among the most advanced globally. Therefore, in the context of this study, the customer represents an optimal case of a typical customer within this industry segment and hence the data and observations should be well generalisable to represent the industry segment as whole. It was agreed to destroy the data after the study was finished.

3.3 Collection of data

The research material consists of several different sets of data, each of which are collected from different sources with different methods. The first of these sets is the condition monitoring data, which is collected from an actual operating production plant. The first set is supplemented with records on the plant available internally through known dimensional data as well as installed base.

The collection of data was carried out by a service engineer of the case organisation, who was familiar with the customer’s site and the systems. The engineer created backups of the data to an external hard drive and had the drive delivered for analysis. The alternative would have been to use a remote connection, however transferring the backups with this method would have taken over 10 days; significantly longer than what it took with an external drive. Moreover, the stability of the remote connection could not be guaranteed, thus should the connection fail even momentarily, the transfer would have needed to be started again from the start. Accessing data directly from the live production servers was out of question due to various firewalls, data security and the risk of disrupting the production.

The types of data that was collected from the systems of customer are represented in table 3. This dataset consists of various measurements from field devices as well as from automation circuits and maintenance observation and action reports. The plant is a modern and well performing production facility, which uses typical processes and raw materials for its industry segment. In addition, the automation network and the collection tools used by the plant rely on industry wide standards. Therefore, the plant and by extension also the dataset can be considered representatives of a typical, generalisable, plant.

Table 3: Types of data collected for the purpose of this research, as well as their sources and description of quantity.

Source & database type	Type of data	Quantity
APM MS SQL Server	Event data consisting of reports of observations of anomalies and completed corrective actions. This includes maintenance operations.	A total of over 4000 event reports linked to 1200 tickets from a timespan of over 10 years.
FieldCare MS SQL Server	Measurement data from devices, such as Metso ND9-series positioners collected by the Metso FieldCare software suite.	Over 4000 unique process nodes with a total of 55,5 million rows of collected data spread across 7 different database backups.
PlantTriage CanaryLabs historian	Recorded values collected directly from automation circuits. The set consists of raw measurements of e.g. temperature and pressure from instruments.	2500 registered circuits, from each of which a measurement is stored every second. The database spans from May to September 2018

This set of data was supplemented with the records of Metso Installed Base -database (IB). The IB is a database that contains the most accurate available data on which devices are installed on the device positions of all customer plants globally and the technical specifications of the positions. The data is collected from e.g. sales records, service projects and from customer records directly. From this set of data, the type codes of devices and some of the dimensional data were used.

3.4 Machine learning methods

In this evaluative study, machine learning methods are used as the method of descriptive analytics, but also as the way to model the data. The purpose is to construct a model, which can detect the patterns of upcoming failures within the data and give estimates of RUL as output, which is also known as response. The process relies on automatic detection of the symptomatic state within the data. Subsequently the future condition of the device is estimated by the model as well as the point in future where the condition will below the functional limit. Essentially the model functions as FCEB-model. It is important to note that there are several different mechanisms of failure, each of which have induce different patterns to the monitored data and each of which occur at different rates. However, given enough data, the model is able to detect patterns of several different types of failures and give an according estimate.

The preliminary step of the practical examination of the topic was the selection of the tools to be used. The topic was discussed internally with experts in the field of machine learning. In this discussion three possible tools, or environments suitable for the task, were identified; namely Azure by Microsoft, Matlab by MathWorks and open source environments using Python -programming language. Azure is a cloud solution, Matlab a mathematics-oriented environment and Python a very common programming language used in machine learning coding and data handling. The Azure had been used before within the company in a different business area to develop data-driven application and it offered advantage of storing the data for future use. However, since the permission was obtained to only use the data within this research, this was not relevant. It was deemed that none of the platforms offered significant advantages over the others and Matlab was chosen as the researcher had the most experience in it.

Data-driven models are built by first selecting the architecture of the model and then training it. In the training a special training algorithm is used to tune the model so that the input correlates with the output. In this case, the pre-processed monitoring data is used as the input and the known corresponding RUL as the output. The correlation essentially is the detection of symptomatic limit and the calculation of RUL. Once the training is complete, a system or a model capable of prediction is ready. When new data, that is data that was not present in the training phase, is given to the trained system as an input, it should output an accurate estimate of RUL. In this sub-chapter, how the predictive models were built are discussed.

The first step is the selection of architecture. As was found out in the review of the literature, the neural networks are currently the best performing and the most versatile machine learning methods for this application. Especially networks, which include LSTM and convolution layers, are found to perform the best in e.g. studies of Wu et al. (2018) and Li et al. (2018). On these grounds, neural networks were chosen as the primary method of machine learning to be employed in this study.

A neural network is built of layers, each of which apply a mathematical operation to the input to produce an output. As the number, type, and order of these layers can vary, and the only way to find sufficiently well performing network architecture is through iteration and educated guesses. Additionally, it is not entirely obvious whether LSTM or convolutional network would perform better in this exact application. Hence, several networks of varying configurations were built to test differing architectures. In total 30 networks were built. Some of these are presented in figure 5. It is to be noted that this study does not attempt to compare or rank the different network architectures, but rather to thoroughly examine different variations in the construction while striving for the best result for the evaluation of the whole process.

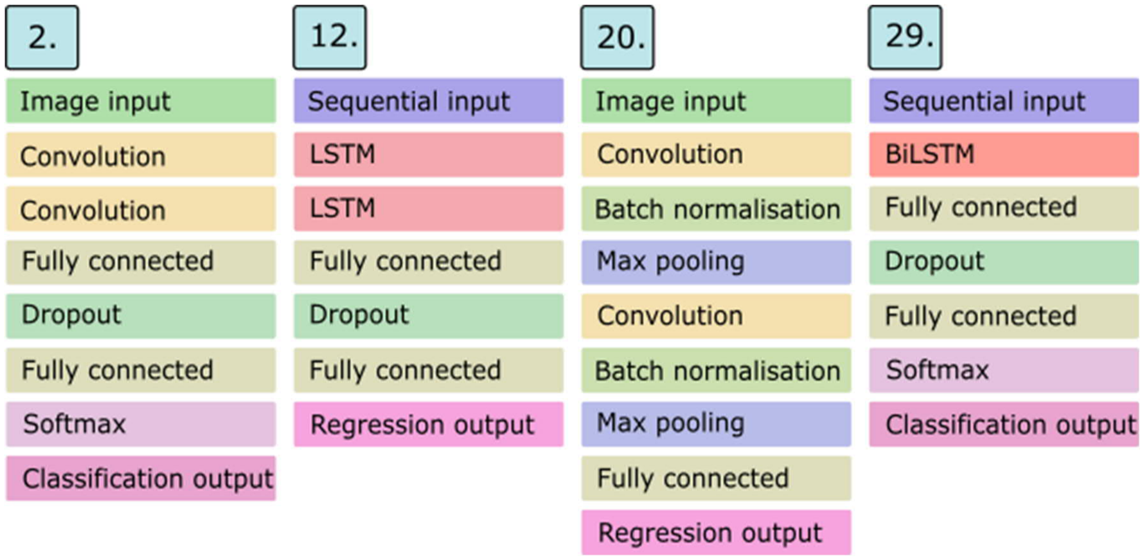

Figure 5: Examples of tested networks. Shown is the layers and their arrangement. Networks #2 and #20 are convolutional networks, while #12 and #29 are LSTM. Networks #2 and #29 have classification output and the two others regression.

The start and the end of each of the networks of the same type are the same as they are imposed by the design and technical considerations. Networks with convolution begin with image input layer, while LSTM networks start with sequential input layers. Current version of Matlab, R2018b, does not allow any other configurations using standard layers of the neural network toolbox. Investigation of custom-built layers and esoteric mathematical operations is, on the other hand, outside the scope of this research. The end of the networks is dictated by the output layer, which can either be classification output or regression output layer.

On the other hand, the middle section, or the section between the first and the last layer varies. There are myriad ways to arrange different types of layers. In this study the goal was to evaluate several different types of networks in the search ranging from very simple to moderately complex. This search was not to compare the different arrangements of layers, but simply to find something which works in this case. Highly complex networks of more than 3 convolution or LSTM layers were excluded from the search due to their extreme demand of computational resources and quantity of data. These were to be researched should longer networks produce significantly greater results than shorter ones.

Each of the layers of a neural network have also internal parameters. These include variables such as the number of nodes the layer has or the size of the convolutional filter of convolutional layer. The second step is the adjustment of these parameters through iterative optimisation. Only the architectures of the predictive models, which perform the best, are selected for this step as the iterative process is very time-consuming. The goal is to find the optimal parameters, that is, with which the precision is at the highest. To do this a function, which forms a network defined by the input parameters and subsequently trains and tests it, is written as a Matlab script. The script takes the parameters in as a vector of 4 parametric values.

To reduce the calculation time, a quasi gradient descent algorithm is used. As the accuracy negatively correlates with the error, a minimum for the error, determined by the RMSE, is searched. Gradient descent is defined for function F(x) as:

$$\boldsymbol{a}_{n+1} = \boldsymbol{a}_n - \gamma \nabla F(\boldsymbol{a}_n) \quad [1]$$

where a_n is the starting point, γ the step size, and a_{n+1} the end point. The start and the end points are defined in n-dimensional space, where n corresponds to the number of parameters in function F(x). The negative gradient defines the highest descent and for a small enough step size, the value of the function at the end point is therefore smaller than at the start point.

In this case, each of the parametric values correspond to a value in one of the distinct dimensions in a 4-dimensional space. The value, which the algorithm attempts to minimise, adds a fifth dimension. As a computer script, the function is neither differentiable nor explicitly definable in mathematical form and hence the gradient cannot be explicitly defined. Therefore, the calculation of gradient is replaced with an operation that probes the neighbourhood of the starting point at a distance of γ. By calculating the difference between the probed point and starting point, the descent or ascent in that direction can be estimated. The steepest descent, and therefore the approximation of the negative gradient, is in the direction of the minimum value at distance of γ.

Once the direction that can be considered approximation of the negative gradient is known, the starting point is moved in that direction by distance γ in the 5-dimensional space. This process is repeated until the minimum is found, that is, a point where there is only ascending gradient in all directions. At this point is also the minimum for the error of prediction and thus the highest accuracy for that type of neural network.

3.5 Data pre-processing

To access the data on the database back-ups, servers had to be set up with Microsoft SQL Server (MS SQL Server) and Canary Labs -database services and the backups restored to these. The MS SQL server database was restored on an existing server with the help of an IT admin. For the proprietary Canary Labs -database, the server software was set up on a virtual server in a remote location with the help of performance solutions manager who had experience working with the database.

The data on MS SQL server databases was accessed in Matlab using standard Java Database Connectivity (JDBC) drivers and Structured Query Language (SQL)-queries. The data on the Canary Labs database was accessed by installing a second copy of Matlab on the server and connecting it locally to the database through Open Platform Communications Historic Data Access -standard (OPC HDA). This instance of Matlab was used to collect the data, which was subsequently sent over the remote connection to the instance of Matlab used for analysis. The possibility of connecting Matlab directly over intranet to the database was also investigated, however, this was deemed to be unnecessarily complicated due to the Microsoft Distributed Component Object Model (DCOM), that the OPC HDA connections use to authenticate users and to establish the link.

The FieldCare databases, restored on the MS SQL Server, were obtained as 7 separate backups from different time periods. These databases contained the historic information gathered from devices by the Metso FieldCare software. As different types of devices produce different types of data, which are incomparable from one device type to another, ND9-series was chosen as that series was the most numerous among all devices. In total, there were 669 unique ND9-devices recorded in the database.

The machine learning methods require uniform input; thus the data was filtered to include only 40 common measurements produced by ND9-series of positioners. Interestingly, not all ND9s had all these measurements recorded. Removing the devices without all 40 measurements left 598 unique devices.

The PlantTriage dataset consists of measurements recorded directly from the control loops. This includes measurements of temperature, pressure and flow. In addition, PlantTriage system performs analytics, results of which are stored in a separate database. Unfortunately, during the work it was found out that the PlantTriage data was not usable for this project. The obtained data was of too short time span. Only a select few failure events fell within this time span and hence the amount of data would have had to be reduced to a minimal number if this set of data would have been included. It was clear, that this would not have been enough to develop a predictive model with machine learning.

The event data, within which failures, malfunctions and replacements were recorded, was fetched from the SQL server and then processed by hand to identify which of the events were related to failures. The processing by hand was warranted as the events were filed in with a description in a free text field. This sort of data is not processable directly by computer. In the database the events were stored with greatly variable level of depth in the content of the description. Some events were accompanied with detailed and accurate descriptions, others not; few even contained short internal discussions. For some events, the descriptions were wholly absent. All events, which could be broadly identified as failure or replacement related were interpreted as failures. As there was only a limited number of records, the individual records were not further categorised. This decision limits the prediction only to the RUL while forgoing the possibility of predicting the individual types of failure, however, it allowed the supplementation of the records with maintenance data from internal data systems of OEM.

The data from the different sources was pooled and then split and arranged into sets so that a single set consists of measurements, parameters and event data that can be linked to a single device assembly. A set of this kind is further referred to as a packet. It is possible omit the split and use all data at once, however, there are several reasons why this is not done. Firstly, this would result in a model of the whole plant, which, while it would take into account complex relations of the multitude of devices working in conjunction, it would need complete data of several plants to train the model. Secondly, the time taken to train the system, as well as the complexity of the model, would increase dramatically. Thirdly, a far-removed measurement point, such as one at the very opposite end of the process, is unlikely to provide any indication on the condition of a particular device. The linking should be done by identifying the measurements that are taken in the near vicinity, such as right before or after the device or by the positioner of the device itself. This is illustrated in the figure 6.

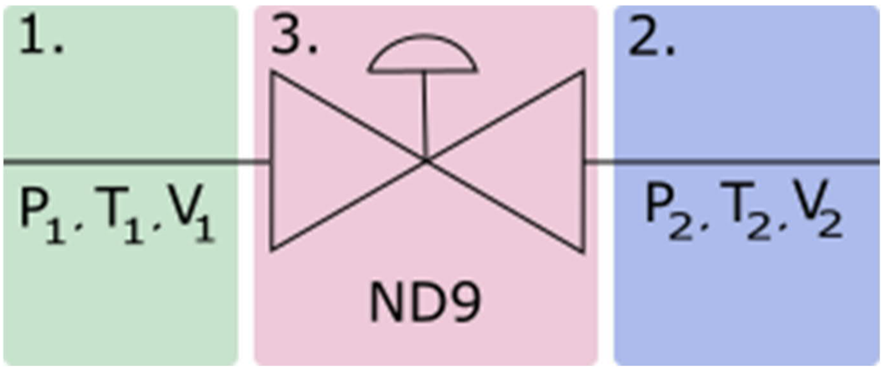

Figure 6: The typical measurements in the vicinity of a device (1 and 2) and by the device itself (3). These include pressure (P), temperature (T), and flow (V) before (1) and after the device (2). In addition, positioner of the assembly collects self-diagnostics data (3).

In literature, e.g. in the studies of Li et al. (2018) and Wu et al. (2018), the data is usually normalised. The normalisation is done due to technical considerations by making the signals and measurements of different sizes comparable with one another. There is a possibility that normalisation will induce an error, however, considering the results of the aforementioned studies, the error that would stem from this assumption seems to be rather miniscule, therefore normalisation is done also in this study. Like in the studies, also in this work standard score is used as the method or normalisation.

The normalisation scales the measurements to specified range and distribution. Using standard score, the result of scaling is the same regardless of the physical quantity and its magnitude. As the phenomenon of degradation and subsequent failure of field devices are inherently related to the fluid and solid mechanics, this introduces an interesting question of similitude and proper scaling of variables. However, this would mainly be concern for measurements of temperature, pressure etc., which were excluded from this research as a part of rest of the PlantTriage data set.

The networks that were described in the section 3.4 require the same data to be fed in slightly differing formats. These formats are illustrated in the figure 7. Most of the networks apart from LSTM regression requires the inputs to be in a form of packets which are of uniform and fixed size of N by M, where N is the number of different measurements and M the number of observations in the time series. Here N is 40 and M is 641. As the time-step in the time series 12 hours was chosen, and hence 641 time-steps corresponds with 320 days. 320 days was selected as the length of the time window as long time increases the possibility of a failure detection limit falling within it. The LSTM networks, on the other hand, do not impose any limit on the length of the time series, thus, in the form of N by M, the M can be of arbitrary length. Hence, for the LSTM networks, the whole time series was given as input.

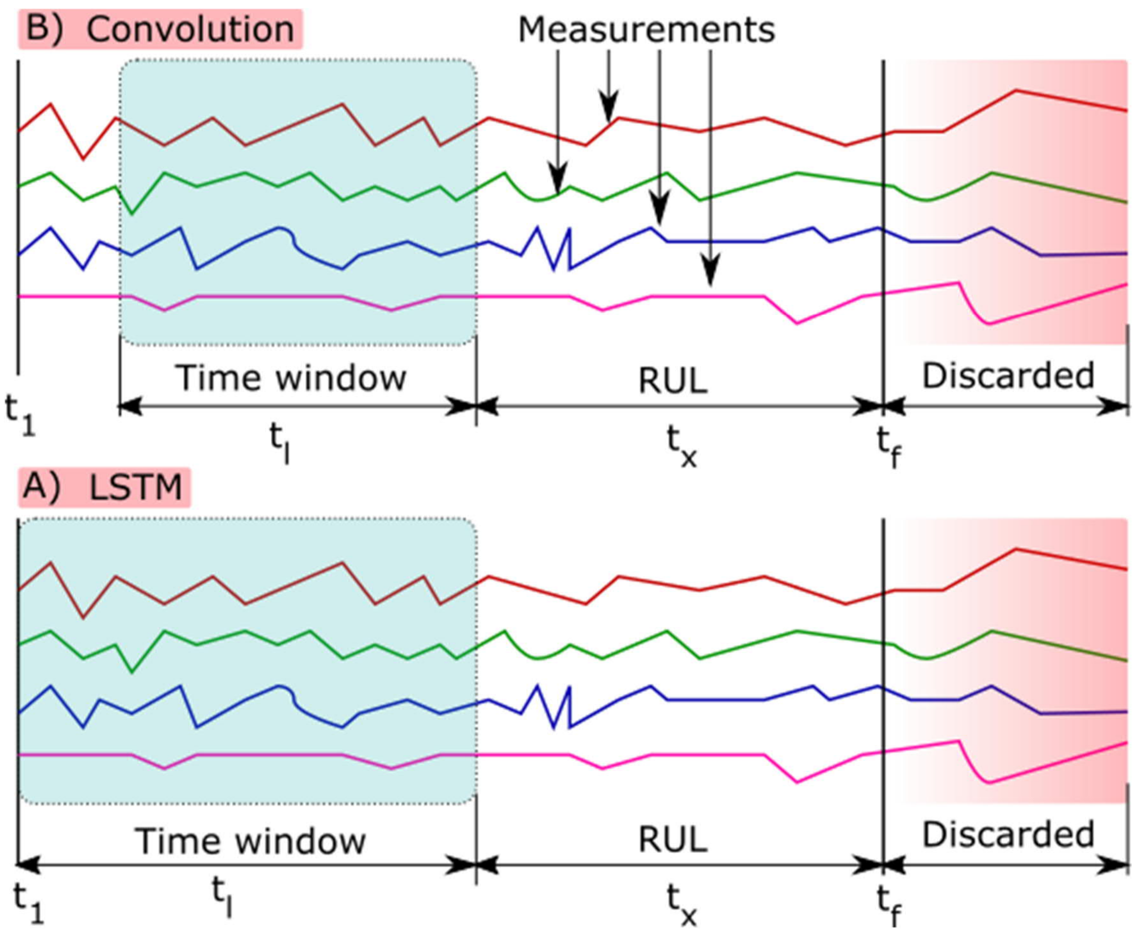

Figure 7: Illustration of how the data is trimmed into packets for use as inputs. Firstly, the point of failure t_f is identified from the event data and the data after it is cut off. Secondly, the RUL is given a value and the time window is defined with a constant t_l. Lastly, all data outside time window is discarded and the data packet consists of the data within time window and the value of RUL.

Time step of 12 hours is a compromise between the size of the packet for a given length of time and the level of detail within it. A shorter time step would have allowed the inclusion of more detail, but at the same time, it would have increased the size of the packets, which is not desirable if it can be avoided as it increases the training time of the networks. Moreover, the loss of detail was considered negligible as only ND9-diagnostics data was used. Had the PlantTriage data, which is polled at much higher frequency than the ND9-diagnostics data, been used, the time step would have had to be significantly shorter.

The formatting of data was done by first linking the time stamp of the failure event to the data and discarding any data after the failure; only the data that is produced prior to a failure is of interest. This point was considered the functional limit. The data recording does not stop at the functional limit, thus the data after the limit might be data from failed device or from a new replacement device. Hence, discarding the extra is important.

Next, the values for t_x are set and the length of data starting from the functional limit backwards is cut out. To augment data, 20 different values for t_x were set. This allowed the splitting of each of the data packets in 20 different ways, which increased the number of packets that could be used as input by a factor of 20. The t_x also corresponds to the RUL as it is the time between the last point of data that the predictive model will see and the failure. The value for t_x was included in the packets; thus, they now consist of monitoring data and a value for RUL. This RUL value is used in the training as well as in the subsequent testing of the models.

The training and testing of the neural networks both require similar, but distinct data sets. Hence, the full set of data was split with ratio 7:3 into training and test sets. The split was done with a script and was randomised. These actions should reduce the risk of overfitting as well as any bias that might occur while at the same time providing enough data for both training and testing. As described prior, both the training and testing datasets contain the monitoring data and RUL values.

To evaluate the performance of the built models and their capability to produce useful information, key indicators of accuracy are calculated. Accuracy is sum of two factors: precision and trueness. The precision relates to how close the predicted values are to the mean while trueness is the difference of the mean of grouping to the desired values. The closer to zero both factors are, the more accuracy there is. The standard deviation relates to precision, and the mean to the trueness. The median, on the other hand, gives indication whether the quantity of predictions is focused on either side of the median.

The RMSE is a performance indicator consisting of a single value, with lower value being better. It is scale dependant and very simple indicator; however, it is suitable for roughly the networks in this case. To better evaluate the accuracy, the analysis needs to be based also on the other indicators.

4. RESULTS

4.1 Evaluation of the built predictive models

In total 30 different predictive models were built in a way described in chapter 3. These are all based on different kinds of neural networks and all were trained using 70% of the data set that was available. The remaining 30% is used to test and evaluate the models and to validate the results. The details of the content of the diagnostic part of the used dataset are presented here in the table 4. Additionally, the set of data contained the temporal points of failure which were used as reference points in the calculation of know RUL values as described in chapter 3.

Table 1: The recorded variables from the ND9-series positioners that were used in this study as well as their descriptions. The descriptions are adapted from the Intelligent Valve Controller ND9000F Device Revision 6 User's Guide (2019). Std refers to standard deviation and Avg and Cum to average and cumulative values.

Recorded variable	Description
Dynamic Deviation, Avg Dynamic Deviation, Std Dynamic Deviation, Cum	Dynamic state deviation is used to estimate valve dynamics such as response times. It is updated whenever the setpoint changes and the valve is expected to move accordingly. Updating continues until steady state has been reached.
Travel Histogram, values recorded in 11 bins	Travel Histograms of the valve operation areas. Total operation range of 100% is split into 10 bins of 10% range each. 11th bin is the bin for the closed position.
Spool Valve Position, Avg Spool Valve Position, Std	Position of the spool valve measured as percentage of the movement range.
Steady State Deviation, Avg Steady State Deviation, Cum Steady State Deviation, Std	Steady state deviation is used to determine basic control accuracy of the valve. It is updated when the setpoint has reached the desired position as precisely as possible.
Stiction, in total 6 recorded values	Stiction is a pneumatic load measurement. It can be used to estimate internal frictions of the control valve package.
Supply Pressure, Avg Supply Pressure, Std Supply Pressure, Cum	The air pressure in the pneumatic air intake of the positioner.
Temperature, Avg Temperature, Std Temperature, Cum	Device temperature measured within the positioner
Total Operation Time	Device total operation time in hours
Actuator, Setpoint, Spool valve and Valve reversals Actuator, Setpoint, Spool valve and Valve Travel	The amount of changes in the direction and the total distance travelled for Actuator, Setpoint, Spool valve and the valve.

All of data that was collected was not ultimately used. Most notably the PlantTriage data that would have contained the measurements of e.g. pressure and temperature in the vicinity of the devices are absent. The decision to omit this part of the dataset was made because of the meagre temporal range the records covered.

The training of a neural network sets the vast number internal hidden parameters so, that a connection between the predictor and response is formed. In this case, the predictor is measured data and the response is the RUL. Output of the training processes are the trained predictive models. Essentially, in the training process, the training program seeks the patterns that precede the failures as well as the rate at which these develop. For example, degradation in the closing element or seals of a valve would cause leakage, which would result in a discrepancy between the pressure difference over the valve and the opening angle. Likewise, valve getting stuck due to residue deposits from the flow medium would be seen in the increase the moment needed to turn the valve which would be seen in e.g. increase in stiction. Both would result in the device being unable to perform its assigned function, that is, in failure. Aside from these examples, there are several other methods in which a valve assembly may fail. Each of the methods of failure affect the surrounding process and operation of the device in a distinct way and thus have their own patterns of failure. Given rich enough data the neural networks should be able to learn to detect all of these and give an accurate estimate of how long it will take for the condition to worsen to the point of failure, that is the RUL.

In the DIKW-ladder the RUL would represent information. This combined with e.g. the information of customers planned maintenance shutdowns could help in selecting the most appropriate times at which the each of the devices should be serviced, and therefore help the OEM to better offer the maintenance services. Additionally, the information can be combined in several different ways, such as to help select the proper replacement devices or spare parts to stock in preparation to maintenance. The full extent of the technical possibilities depends highly on how accurate and reliable the predictions are. However, other factors, such as the needs of the customer and the method of service delivery need to be considered. The assessment of the accuracy is done by testing their accuracy by comparing the responses of the networks to the actual known values. This testing will give indication on how well the contriving of the data-driven models succeeded and to what extent they may be used in service business development and in generation of service recommendations.

The criteria for the evaluation are based on the desired outcome, that is to what extent the predictions are usable as an aid for decision making and hence value creation. Practically, this means that the DIKW-ladder is followed. Therefore, first the networks are evaluated and improved iteratively until the best outcome obtainable in this setting has been reached. This is the step from data to information or knowledge. Subsequently, the ways in which the information or knowledge that was gained can be used to improve MRO services and how it can help to generate service recommendations shall be assessed.

The testing was done with the test data by using the built networks to predict the RUL and comparing the values from prediction to the actual values that were known. When data, which is similar to what was used to train the neural network, is fed to the predictive model as input, it outputs a response, which is of similar nature to the responses that the network was trained with. In this case, the networks were trained with data pairs consisting of monitoring data and a known RUL value, hence a trained model will take monitoring data as input and give an estimate of RUL as the response, that is, as the output. As described in 3.5, the test data consisted of 30% of the total dataset.

In the ideal case the response should match the known value. The ideal case is never reached, as due to the inherent nature of predictions, there will always be some difference, or error, between the actual known value and the prediction. The error is directly related to the accuracy of the prediction and, therefore, also to the usability of the prediction. Hence, it is vital to ascertain the magnitude of the error. This is done by first calculating the prediction error as the difference between the output value and the anticipated real value for a large number of individual inputs of monitoring data, after which the Root Mean Square Error (RMSE), mean, median and standard deviation are calculated from the batch of prediction errors. These are presented in the table 5. The unit for all the values is days.

From the table 5, it can first be seen that the standard deviation for most networks that were built it is in the range from 170 to 200 days, the RMSE in a range of 160 to 180, the median between -60 and +60, and the mean from -30 to +30. Secondly, a clear outlier is identifiable; namely the network #24, whose all calculated values are approximately hundredfold larger than for other networks. As in e.g. RMSE, smaller is better, it can be safely concluded that this particular network exhibits extremely imprecise prediction behaviour. It can be assumed, that something has not fully succeeded in training, and hence the network is excluded from further analysis.

The RMSE is not only scale-dependent but is also affected more greatly by larger values of error, and thus the scale needs to be considered in analysis. As described prior, the time-step is set as 12 hours and the length of each input to correspond to 320 days, hence the RMSEs are to be understood in this scale. The threshold for what is suitable would be set by the actual value-facilitating use case, but for the sake of assessing the scale, it shall be assumed that a constant mean error of 10 days would be acceptable. The 10-day prediction error would produce RMSE of 10, which is considerably smaller than what was obtained in this study. This indicates either a systematic inaccuracy in the built predictive systems or that there are few very large errors within more accurate predictions.

Table 2: Tabulated results of the first testing. In the columns the number of the network, the type of the network, and the calculated values of RMSE, median, mean and standard deviation.

#	Network type	RMSE [D]	Median [D]	Mean [D]	Standard deviation
1	Convolution Classifier	236,34	-60	-31,655	235,14
2	Convolution Classifier	198,23	30	24,805	197,45
3	Convolution Classifier	236,07	-30	-58,11	229,71
4	Convolution Classifier	210,44	-60	-80,55	195,18
5	Convolution Classifier	204,39	-60	-57,875	196,8
6	Convolution Classifier	187,52	-30	-61,89	177,715
7	Convolution Classifier	173,89	-30	-6,85	174,445
8	Convolution Classifier	205,255	-90	-90,945	184,735
9	Convolution Classifier	233,915	-120	-133,465	192,86
10	Convolution Classifier	174,645	30	33,07	172,165
11	LSTM Regression	154,45	12,215	22,845	153,325
12	LSTM Regression	153,775	28,56	21,465	152,84
13	LSTM Regression	145,41	4,385	10,385	145,585
14	LSTM Regression	161,415	6,37	20,68	160,685
15	LSTM Regression	186,185	-100,21	-87,21	165,115
16	LSTM Regression	278,13	-229,975	-227,415	160,72
17	Convolution Regression	172,765	20,9	-11,945	173,035
18	Convolution Regression	176,15	46,52	21,175	175,565
19	Convolution Regression	241,335	-100,255	-119,095	210,73
20	Convolution Regression	172,285	13,465	3,935	172,92
21	Convolution Regression	192,935	-5,51	-9,015	193,49
22	Convolution Regression	176,35	43,23	28,165	174,775
23	Convolution Regression	171,43	1,64	-10,05	171,815
24	Convolution Regression	14881,75	8707,875	10541,09	10546,46
25	LSTM Classifier	181,28	30	-9,35	181,7
26	LSTM Classifier	141,13	0	-23,045	139,74
27	LSTM Classifier	206,4	-30	-28,26	205,2
28	LSTM Classifier	181,785	-30	-27,175	180,395
29	LSTM Classifier	178,07	-30	-43,915	173,2
30	LSTM Classifier	151,58	-30	-1,955	152,12

For the output of a working predictive model, the mean and median are both expected to be fairly close to zero as this would indicate error which is evenly spread on the positive and negative, that is, good trueness. This is not the case with the built models as there are notable, and in some cases very large, deviations from zero. In addition, the standard deviations indicate a very poor precision for all the networks. The deviation is roughly tenfold too large. Therefore, it can be said that the predictions do not match the correct values at all at this stage and that there is room for improvement.

The networks, which provided the best results overall were selected for improvement and further optimisation to see if the performance can be improved by adjusting the variables of network layers. The optimisation was done using the quasi gradient method presented in chapter 3. Calculating this took a week on an above the average desktop PC. The results of this effort are presented in the table 6. Here, the emphasis is on the increase of RMSE to a level which can be accurate.

Table 3: Tabulated results of improvement and optimisation effort. Presented are the number of the network, its type, the best calculated RMSE, the vector of parameters of the network, and increase in per cents compared to results presented in the table 5.

#	Network type	RMSE	Parameter vector	Increase (%)
6	Convolution Classifier	176,963	[15,15,45,195]	5,629799
7	Convolution Classifier	175,454	[15,15,45,195]	-0,89942
10	Convolution Classifier	215,0053	[25,25,45,195]	-23,1099
11	LSTM Regression	148,55	[15,15,65,215]	3,820006
12	LSTM Regression	141,161	[5,15,55,215]	8,202894
13	LSTM Regression	141,083	[5,5,65,205]	2,975724
17	Convolution Regression	179,4778	[5,5,55,215]	-3,88551
20	Convolution Regression	180,5298	[25,5,65,195]	-4,78556
23	Convolution Regression	177,9603	[15,15,55,195]	-3,80931
26	LSTM Classifier	193,469	[15,5,55,195]	-37,0857
29	LSTM Classifier	163,923	[20,20,50,210]	7,944629

The table 6 lists the best result that was obtained for each of the network architectures which were selected for the optimisation. Here, the numbering of the networks corresponds to the numbering presented in table 5. The parameter vector describes the variables of the network layers, the specific details of which are unimportant. What is important is, that a great number of different combinations were tested, and only the single combination that is related to the best result is displayed.

Despite the effort, the increases are insubstantial, the result even decreased for some. Even, while the percentage increase of e.g. network 12 is at 8% noteworthy, the increase in absolute accuracy is far too low for all networks. The RMSEs are still relatively large, which indicates that the predictions are unusable. The iterative optimisation could have continued further, however, it was decided to stop here. It is not reasonable to expect any further improvement that would increase the accuracy substantially to a range where the predictions would be usable as the required leap in performance is simply too large.

The histograms of prediction errors of working predictive models should roughly follow the bell curve shape of normal distribution with small standard deviation. In this ideal case, the majority of prediction errors are small with the larger errors being in minority. As can be seen on the histograms in figure 8, this is not the case for the built networks.

On the contrary, in fact, as the prediction errors seem to be distributed rather evenly. This indicates that there are equal amounts of predictions which are correct and those which are extremely wrong. Considering also the RMSEs, it can be safely concluded that the built networks do not perform at all in this prediction task.

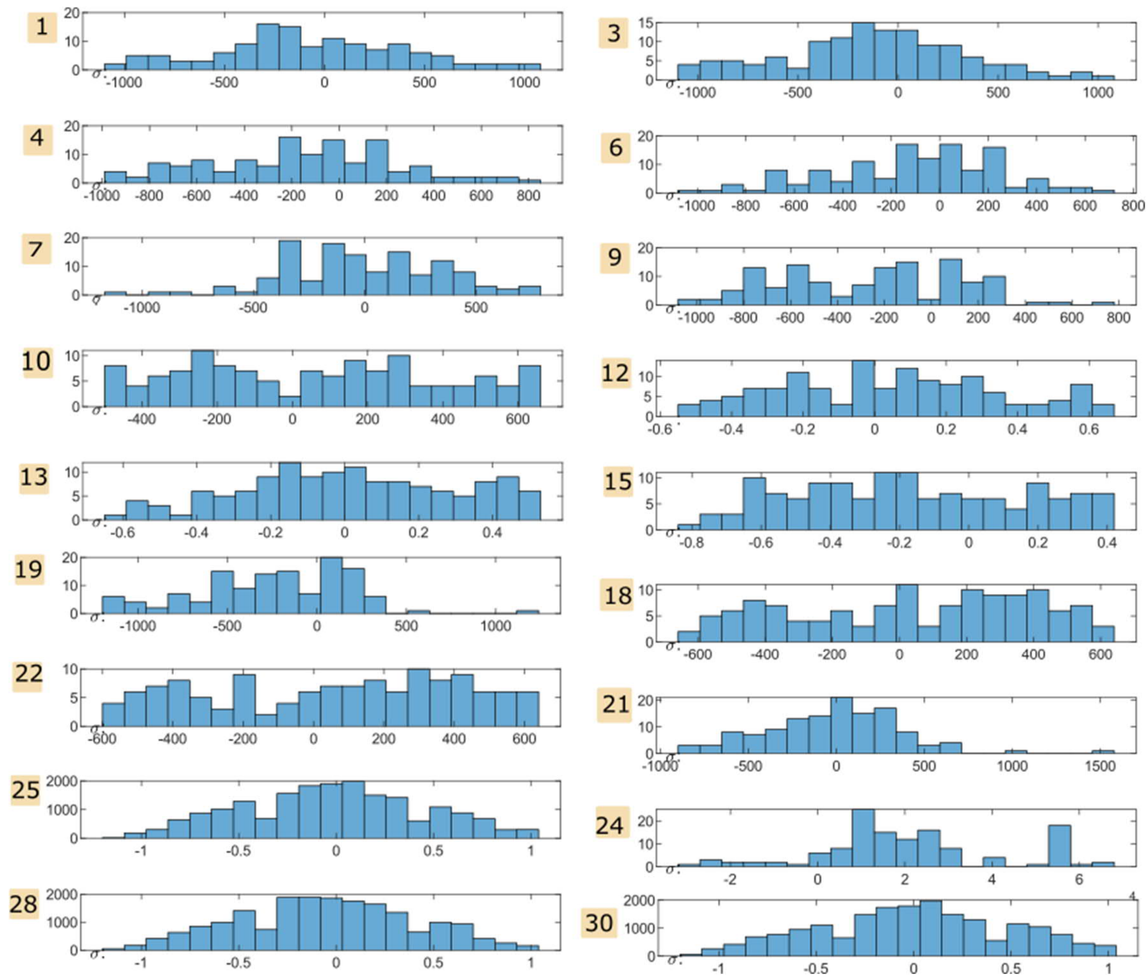

Figure 8: A selection of histograms of the prediction errors of the contrived predictive neural networks. The y-axes represent the number of predictions which fell into each of the 20 bins. The x-axis represents the prediction errors.

This can be further confirmed by looking at the figure 9, where predictions and actual values are plotted for each of the predictions. Here, the predictions are arranged in order based on the actual value so that the actual values form a diagonal line. Ideally, the predictions would fall very close to their associated actual values and hence form a similar line in the same place. As can be seen, this is not the case and hence there does not seem to be even a remote correlation between the prediction and actual value. It can therefore safely be concluded, that the refinement of data into information with the selected method from the available data did not provide useful result. The evaluation was stopped here, as no improvement could be expected, and the time constraints of this study do not allow additional experimentation with other methods. Had there been additional time allocated, alternative approaches and methods could have been tried.

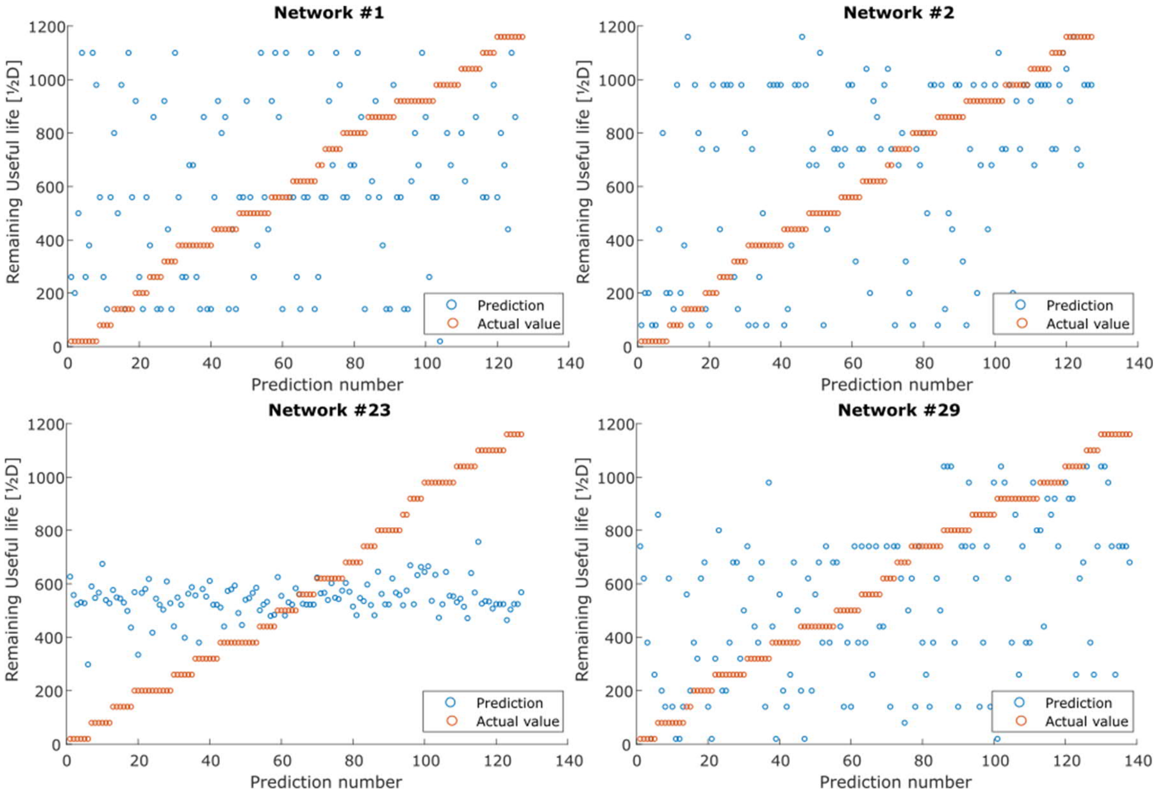

Figure 9: The predictions of RUL by selected example networks plotted against the actual values. The actual value – prediction -pairs were sorted based on the actual value in order from smallest to biggest.

From the high level of inaccuracy directly results that the predictions are not reliable and therefore unusable. It would be unwise and unreasonable to use these as a source of information or to base any sort of decisions on these.

4.2 Analysis of the inaccuracy of the predictive models

It was found out that the accuracy of the contrived predictive models is poor. This section is dedicated on analysing why the results ended up as they did. The thorough understanding of the results and the factors which affected them is essential in forming the suggestions for further development and possible success in the succeeding studies. There are several possible reasons for the dismal predictions, some of which are the methods, the data and the execution of the research itself.

The selected methods for the prediction were based on neural networks as these are according to the literature reviewed the best performing methods. Both the convolutional and LSTM NNs were tested, both of which are highly advanced networks in total of 30 different configurations. The configurations ranged from very simple to more complex and hence the tested networks should be representative enough of the method as whole.

The RMSEs of predictions from different networks are rather uniform as can be seen from the figure 10. The low variation between the performance of different networks indicates, that all the networks, and by extension the different architectures performed equally badly. If there had been a noticeable performance difference between some of the networks, it could have been concluded that certain architectures perform better than others. Subsequently, deeper research could have been done into those and perhaps through exhaustive iteration contrived an adequately performing predictive model. However, as this was not the case, as well as considering that the networks have performed well in studies by other authors, it can be concluded that the employed methods themselves were not the main reason for the bad forecasting accuracy.

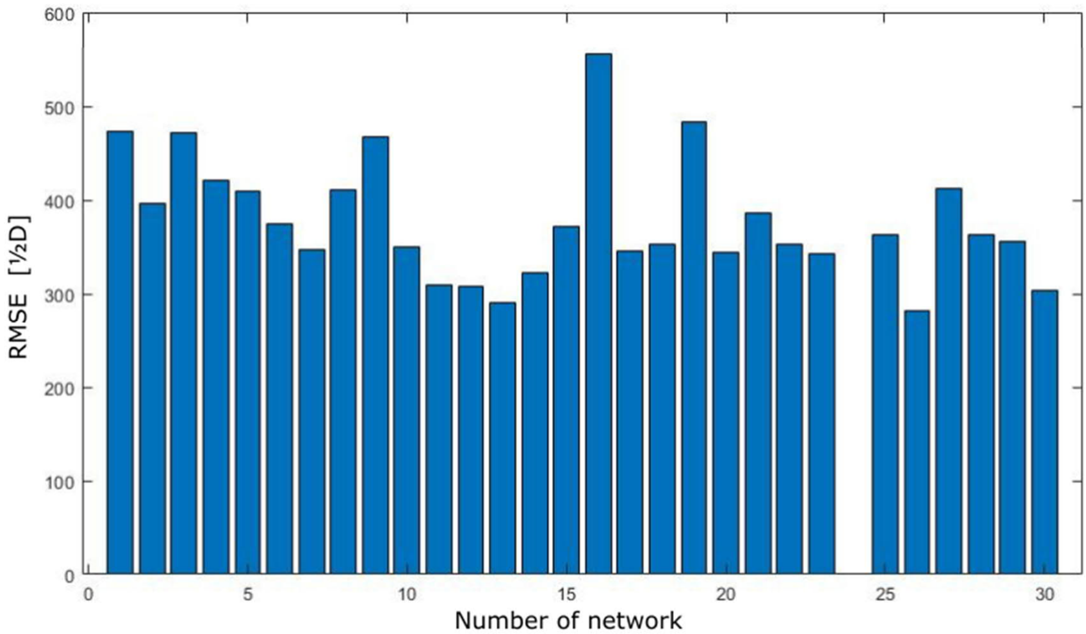

Figure 10: Bar graph showing RMSEs of the tested networks. Network #24 has been excluded from this figure as the bar extends way beyond the upper limit of the graph.

As the tool to contrive the predictive model, Matlab with Neural Network -toolbox was used. While the toolbox as a packaged commercial product does not represent the pinnacle of machine learning, it stands to reason that the toolbox is not wholly without merits. Hence, the selection of Matlab should have induced a greater error to the results that what is obtainable with model built specifically for this task. However, not in a scale that obtained in this study. Hence, the tools used can be excluded from the main reasons for inaccuracy.

The data itself is one of the most likely sources of the inadequate results. The event data is recorded with preciseness adequate for its use in its main function, that is day-to-day maintenance, it was found out that this was usable in this study. The free text fields in which the observations of the anomalies were recorded was filled with varying levels of elaboration. Original aspiration was to treat different failures of each of the components of the assembly as separate events and hence have the predictive model be able to distinguish the different failures. However, as the event data did not in all cases contain information in sufficient quality, all failures had to be treated simply as general failures.

The several different types of failures can be assumed to each induce different patterns in the monitoring data. This does not necessarily pose any challenge directly to the employed methods themselves even if they are all simply listed in the general category of failures, as a functioning neural network should be able to identify several different patterns at once. Nevertheless, if multiple different patterns are present, substantial amount of data is needed to train the network to detect each of the patterns.

However, data was not available in sufficient quantity. Dictated by the approach that was used, a portion of data from a pool that was not particularly large in the beginning had to be discarded. Firstly, data from only one plant was available for use as obtaining the data and the explicit permission to use it proved to be non-trivial task. Secondly, the unavailability of historic PlantTriage data forced the omittance of this set of data in its entirety. Thirdly, while there were over thousand tickets in the event data, only 52 of these were failures that could be linked to specific devices. Data to which no failure event could be linked had to be discarded as unusable. Each of these mandatory steps reduced the dataset significantly. This reduction of data was more than what was anticipated.

If a neural network, or any other type of machine learning, is trained with an insufficient amount of data, the training algorithm cannot find the patterns and the result underfitted model. The quantity of data that can be considered insufficient depends on the type and especially the depth and complexity of the network. In this study, networks of varying depth were used, with focus on the shallower ones. However, they all the networks can nevertheless be considered to be at the very least moderately complex.

Networks suffering from underfitting exhibit high amounts of inaccuracy and their responses bear little relation to correct values. This is the case also in the results that were obtained in this study and is observable in the figure 9. Here the RUL predictions from the built predictive models are plotted against the known correct values. Ideally, that is, when the model works as intended, the predictions match the correct values and thus the pairs of predicted and correct values should form a clean line starting from the origin and running in 45-degree angle respective to the axis. As can be seen, this is not the case here. Instead, the pairs are distributed either erratically or in a vertical line. The erratic predictions clearly indicate the lack of connection between the input and the response as the predictive model seems to assign a random value as response regardless of what the data contains. The vertical line, on the other hand, indicates that the network outputs an approximately constant value regardless of the data input. In both cases, the relation of data to response is absent.

Considering the results and the quantity of available data, it can be concluded, that the inaccuracies are caused by underfitting which resulted from the low amount of data. Moreover, it is possible that the employed methods increased exposure to greater risk of experiencing underfitting. As result, the accuracies of the predictive models are so bad, that they cannot be used for any application. Therefore, it is best to search for other possibilities or to develop the prerequisite stages of data collection before another attempt.

4.3 The application of the model to service

As the forecasts made from the available date are not reliable, the investigation of the potential improvements in service offering has to be conducted purely on theoretical level without any reasonable mean to test or evaluate the ideas in practice. As such, the ideas that are present here cannot be currently realised but may serve as inspiration or incentive for future research. Had the contriving of the forecasting system been more successful, predictions of the upcoming failures could be made with certain level of accuracy purely based on data. Let the accuracy of the predictions be considered sufficient and predictions reliable for all intents and purposes.

Ideally, the customers should be contacted early in the design process and the service offering designed to suit needs of the customers (Pawar et al. 2019; Smith et al. 2014). Unfortunately, the client, who generously provided the data for this study, declined to participate in interviews and mutual design process for service improvements. Other customers respectfully declined the invitations too. Additionally, there is no time within the limits of this study to seek out more participants. Hence, there is limited possibilities to reliably discover the needs of the customers or how the details of the service should be in objective way within this research. On the other hand, the needs of customers are not static and as such, they might be totally different in future when reliable data-driven predictions of device failures actually are doable. Therefore, details are omitted in this discussion; these are left to be determined in future research.

In the DIKW-ladder the predictions of RUL is information, that is, data which is refined into format that is understandable to humans (Kunttu et al. 2017). The step from the data to information is but the first step in the chain. The next two steps would be the capabilities to interpret the information and recognise the need for actions as well as skills to combine information and knowledge from different sources to support decisions; that is, the knowledge and wisdom respectively (Kunttu et al. 2017). The decisions, which lead to actions are what creates value.

This implies, that to reach the value additional sources of information and knowledge as well as human expertise and insight are needed. From this can be inferred, that what additional information is required depends on to whom they are presented to and conversely, different decision makers require different types of information and knowledge to support their activities.

Predictions were envisioned to be used in generating device specific service recommendations. It is necessary to specify what should be shown in such a recommendation and to whom they are presented and for what use. The prediction could serve as a basis for a general solution; a solution that can cover basic needs of several of customers. The need to keep the plant up and functioning is one example of such basic need as that is one of the foundations of the production and business. This requires maintenance, which is in several plants done periodically in shutdowns or during day-to-day operation should the need arise.

Before a plant enters shutdown, a maintenance plan and schedule are made by an expert engineer or a technician with the help of recommendations of the original equipment manufacturer (Ahmad & Kamaruddin 2012). For this, information on the condition and type of the devices are needed (Efthymiou et al. 2012). This information would be the RUL estimate that was intended to be automatically generable. Here, the prediction could aid expert engineers and technicians of the OEM to recommend the correct devices for maintenance and provide sound reasoning for the selection. Moreover, by observing the changes in the RUL estimates during the operation, the technicians might be able to spot devices which are failing sooner than anticipated due to e.g. changes in the process.

Alternatively, the estimate could be used a part of a tailored solution for a single plant. Here, the expert engineers of OEM could conduct various studies of the plant as well as the devices that are in use, such as device audits and plant criticality analyses to identify the most critical devices regarding the process, environment or safety. This would allow the OEM to assist the owners of the plant in long term maintenance planning and offer life cycle services.

Additionally, the service recommendation could provide better knowledge of the maintenance need to the customer. This would hence enhance their ability to evaluate the recommendations given by the engineers as these recommendations are essentially service propositions and value offerings. Moreover, properly presented knowledge could improve the service experience. Decisions are made based subjective experience (Kaasinen & Liinasuo 2017), and thus improving the service experience could improve the service sales.

All these suggested possible use cases would increase the level of servitisation and deepen the relations of the OEM with the customers. This would help with the stated strategic goal of the case organisation to be reliability partner. While the envisioned use case of using the prediction to help generate service recommendations and to improve service offering cannot be realised in any way, it can be concluded that there is incentive to continue development.

4.4 Development suggestions

The device forecasts of device condition were aspired to be reliable enough to automatically generate service recommendations based on them. However, as the accuracy of the models developed in this study were deemed too inaccurate, this prospect remains unrealised. From this directly results, that currently the forecast or the service recommendations cannot be used as a basis for development of MRO service offering in any way. Essentially, what was found out, is that the data-driven CBM methods cannot be implemented as the link between the available data and the condition or failures was not discovered.

Nevertheless, if the reasons which contributed to the egregious performance of the forecasts can be identified and rectified, better results could be obtainable in future research and development. This is especially important as the industry is transforming toward service orientated business models and strategies in which data-driven solutions will inevitably be valuable and as the organisation has deep desire to advance from being an OEM to a reliability partner. During the research several key factors that could serve as potential areas of future development were discovered. In this section the areas of development are discussed and in addition suggestions on how the development could be conducted are given. The order is discussion starts with data, then prognostics and lastly industrial services and business.

4.4.1 Increasing the availability of data

As was found out, the availability of data related to the condition of devices was significantly lower than what was expected. Firstly, there was trouble in gaining the access to data and the permission to use it for research. Secondly, it was found out that the historic PlantTriage data was missing, which considerably limited the quantity and quality of data that was ultimately usable. Thirdly, the event data was not in an easily machine-readable format and thus had to be manually interpreted. These three points highlight three different sectors which all need to be addressed; namely the permissions to gather and use data, management of data and its storage, as well as record keeping. Improvements in these areas should considerably increase the availability of data for future use, which is essential as CBM and related services cannot be implemented in any extent if reliable and exhaustive data on the condition of the devices are not readily available for use. Hence, development of an all-encompassing solution for gathering of condition data is suggested.

Lack in quantity of the data greatly hampered this study. Hence, the development should aim for a system which would be capable of combining raw data or information from several plants for mutual benefit; including plants of directly competing operators. The possibility of using data from several locations as one data set should tremendously increase the likelihood of developing data-driven applications. This is essential as it was found out that data from only one plant is insufficient. However, plans of pooling data from various customers will inevitably result in resistance from said customers are wont to keep their data private due to concerns of e.g. keeping the details of process secure.

The approach needs to be therefore subtle and greatly focused on the potential value as well as the integrity and reliability of the case organisation. Any possible risks of data breach need to be considered, addressed and demonstrated to be non-existent. Consideration needs to be placed also on the legal frameworks and contracts of data usage, so that the data is also legally available also for purposes that might not be yet envisioned.

In this research the transfer of data from the plant was done by manual copying of backups to an external drive and sending it with courier service. Every part of this process is inefficient. The manual copying required the service engineer to enter the facility of the customer, to attach the drive to the systems, and to locate the backups to copy. Subsequently, the drive had to be sent and once it has arrived the backups had to be restored on servers. This effort is acceptable for a single case, but it scales very poorly due to the vast number of drives required, the effort it takes to manage these and the risk of data loss in the transit. Moreover, the reliance on backups restricts the available data only to that which can be backed up and restored.

In addition, currently the main sources of monitoring data used in this study, FieldCare and PlantTriage, are separate systems. Some data is also stored in the factory automation systems. However, these are closed systems and thus not easily accessible. FieldCare and PlantTriage are designed ground up to collect data and perform different types of analytics. The problem with these systems is that neither of these have any easy way to extract the raw data or analytics results out of the systems and the plant. The new solution should be technologically advanced and be capable of collecting, transmitting and analysing vast quantities of data.

Hence, combining the monitoring data streams into a single and manageable stream is recommended. A single standardised stream or pipeline of data is way more manageable than the current systems and would provide a unified framework through which the data could be accessed thus truly making the data available. This could be either done by merging the FieldCare and PlantTriage into a single product or by adding a third standalone system which would only listen and compile data from the various other systems. The former option would allow the designing and implementation of exactly what is needed and potentially reduce the internal complexity of the factory data streams. However, development of such a system as well as implementing it in the productional facilities is very costly and time consuming. The latter option would be easier to implement as no significant adjustments to existing systems are needed, however, this would slightly increase the total complexity of the factory data network. The latter is still recommended at this stage as it could potentially be implemented even in facilities which currently neither employ FieldCare nor PlantTriage.

While in this study data which is already gathered for various other purposes was relied upon, this does not necessarily need to be so in the future. It is within realm of possibilities to add new measuring instruments and monitoring practises that are specifically designed for monitoring condition, especially if a new system of data gathering is designed and implemented. This would require the identification of which factors could and should be monitor. The identification could be done e.g. with the help of physical modelling. The additional monitored variables would increase the chances of success in future, however, does not come without a significant cost in design, manufacture and installation of the new instruments.

Several of the conducted studies on prognostics, such as the ones by Samanta et al. (2003), Costello et al. (2017) and Yunusa-kaltungo & Sinha (2017) rely on the measurements of vibration. Vibration is not something which is currently measured, and hence, considering the prevalence of the vibration measurements, adding such should be considered. Nevertheless, the cost and benefit still need to be assessed

Any measurements added will undoubtedly add congestion to the existing data transfer pathways. As currently the existing automation network is relied on to gather data, the congestion could negatively affect the primary function of the network, which is to control the devices. While the employment of the automation network can continue, it will limit the possibilities. The alternative is to build another network with technologies that are better suited to transfer of data. There are several well proven methods of data transfer, both wired and wireless, that can also reliably transmit vast quantities of data from multiple sources simultaneously and therefore can be applied in building a network of data transfer within a facility. The great challenge here is the cost, which makes building a data network unprofitable unless other applications wholly unrelated to device monitoring and analysis can use the same network.

The envisioned system would still be within the productional facility and behind their firewalls. In addition, as was found out in this study, the connection speeds to and from the facility can be low even in a developed country with good internet infrastructure. This presents problems, as the plant operators might be reluctant to open routes through the firewall as doing such might compromise the security of the plant and even if the route would be opened, the data infrastructure outside the plant could become a bottleneck. Thus, continuous transmission of all raw data from the plant could be infeasible. However, if some pre-processing or analytics was done already at the plant, a smaller amount of data could be sent outside for further analysis. Moreover, pre-processing can be used to anonymise data and thus obscure critical process information without compromising its usability in device level analysis.

Hence, the solution should include enough computing power to handle analytical tasks and data processing in addition to merely listening other systems and storing data. However, it is likely that the need of knowledge varies over time and thus varies the need for what the analytics produce. OEM thus needs to be able to implement new algorithms to the system. Essentially what is needed is a server, which can listen to the factory network to collect data and which can be ordered remotely to perform calculations and to send the results over internet to OEM. The data must also be received at the OEMs end. For this, cloud computing services, such as Microsoft Azure, could be used.

In addition to the technical capabilities, the processes and practices need to be addressed. It was found out, that the PlantTriage dataset that was available consisted only of around 6 months of information. The loss of this data was not caused by faults in gathering of data, but rather mistakes in how it was handled afterwards. Simply, sufficient backups were not made and achieved as there was seen no reason to store the old data and subsequently there was no systematic process to store the data in place. To avoid unfortunate loss of potentially valuable data in future, necessary process to achieve data needs to be formulated and the users, engineers and technicians trained to follow it.

4.4.2 Other possibilities of forecasting

In this research only one possible way to forecast faults was examined. There are multitude of other options which could be studied as well. While machine learning could be the way in future, these other methods could be used as intermediate methods while the data gathering and the machine learning methods are developed. The development of intermediate methods would allow the progression in the field and gain more insight while at the same time have working solutions albeit ones which might not be the most effective.

These methods could include for example physical simulation and basic statistic-based assessment methods. Physical simulation and analysis of the failure mechanism is challenging and developing a comprehensive model that addresses all the different mechanism might not be feasible. However, the scope could be limited to only certain mechanisms of failure, such as the ones that are easiest to model or the ones that are most common. If such a model is developed, the properties on which it depends still need to be measured warranting investments in data collection systems described in the previous subchapter. However, existence of a working model could act as a clear incentive to commit in resources and finances in development of data collection and hence later allow advances in other areas dependant of data.

The statistical analysis, on the other hand, can be based on several varieties of different data. Firstly, the approach could be similar as with this study, except with replacement of machine learning with more conventional statistical methods. Here, the dataset is the condition monitoring data and thus, also this approach would benefit from development of collection of data. However, unlike with the physics-based modelling, the model cannot be developed prior to the start of data collection and hence it cannot act as the primus motor of infrastructure investments.

CBM was focused on in this study. CBM inherently requires condition of individual devices to be monitored, and as currently there does seem to be deficiencies in the monitoring, it could be more lucrative to also examine the possibilities of TBM. TBM is inferior to CBM in many ways if the condition of the devices can be reliably assessed, but when it cannot, according to e.g. Takata et al. (2004) the TBM performs better. Time is, after all, much more trivial parameter to measure than condition. As we, as an OEM and service company can only incentivise investments in monitoring infrastructure rather than mandate them, the development of a TBM-based service offering could be lucrative due to comparative ease of implementation. Moreover, the OEM would not be dependent on data owned by customers.

While CBM links maintenance decisions to monitored condition, the TBM links the same decisions to time. Like CBM, the TBM requires a model to function as the link. A TBM-model to link the time to maintenance decisions could be developed using the same methods as CBM-based modelling. A physical model could be used in conjugation of e.g. Monte Carlo simulation to produce a statistic that could be used to then assess how various types of devices degrade in different conditions. The same can be done with empirical data e.g. through statistical analysis maintenance records of the OEMs own service centres. Here, the downsides are similar as with CBM except for the challenge of collecting the modelling data. If this approach is used, the physical model that could be used in simulation still needs to be developed and the statistic made. However, when using data from the OEMs own service centres, the company would have greater control over what data is collected and the data is inherently more readily available for use.

4.4.3 Services and data management

The research was set forth with an assumption that data-driven prognostics could be used as basis to improve services. This fundamentally made the effort a technology push approach and locked the investigation in only how the selected technology could answer to questions rather than finding the best alternative among multitude of technologies. Certainly, should there have been more success in creating a functional predictive model, some use for it could have been found and the value assessed. Nevertheless, it is reasonable to question the approach.

It was never certain if the wanted phenomenon, failures in devices, was even indicated in the data which was available. Moreover, the availability and quantity of data was thought to be higher than it was. The establishment of the former was desired in this study, while the latter highlights the lack of knowledge on the subjects in the case organisations. The knowledge on the possibilities and phenomena is vital for an organisation that thrives on expertise. Without such, the attempts to facilitate value are flawed. Thus, the case organisation should seek to acquire additional knowledge and skills on the subjects through training of employees and research projects.

Perhaps a strategy more akin to market pull would have been more fruitful. Instead of the used approach, perhaps a better starting point would have been to first find out what sort of information is most valuable, what data is available and then develop the analytics to support the end goals. This would allow the linking of the data directly to a business objective and select the best methods as well as the sources of data to support the necessary analytics. It is, after all, not necessary to develop CBM if CBM provides miniscule value to the OEM and the customers.

The steps, which lead to the creation of the value for customer, are a process. This could be a sales process, a maintenance process or some other; and they are known to result in value. If the focus would be on these processes, and how they can be supported with new knowledge, certainly something valuable would be done. Moreover, in all of them data is produced. Hence, by studying the processes in conjugation with one another, first the value facilitating processes, that could benefit from additional knowledge, could be identified and then identify possibilities of improved data gathering in the other processes.

Additionally, perhaps the chain of data to value should be seen more akin to a lattice, which interlinks different processes. This could potentially allow more holistic and all-encompassing approach to data management and improve the identification of possibilities in combining knowledge from different sources.

4.4.4 Summary and timeline

Several possibilities on how the development could go further were identified. These are summarised in table 7. However, it is not feasible to attempt all of them at once due to limited resources and the capital investment required. Moreover, some of the suggestions, such as improvement of connectivity within plants, will require deep co-operation between several participants and will incur considerable costs, and hence decisions to commit to such undertakings would be strategic. Here, a possible order of priority and a timeline is presented.

It is reasonable to first concentrate efforts on the prospects that are the most easily realisable and whose potential is the greatest. This leans the attention towards approaches which use data from the OEMs own sources, such as TBM based on statistical analysis on the service records. As the availability of the required data has already been verified, it can be expected that development of a model could take six to nine months.

Studies on the degradation and deterioration phenomena can be conducted concurrently as this and analysis of historic are not interdependent. It is unclear how long such an effort would take, but any possible findings can be in future used as a stand-alone model or to enhance other methods. It is reasonable to first investigate only certain parts of devices or certain mechanisms of failure to reduce complexity and later advance to study more complex mechanisms.

Table 7: Summary of development suggestions.

Identified challenges and good practices	Presented development suggestion
Accessing data is technically difficult	Development of an efficient, secure and easy to use solution to share large quantities of data to selected partners. Unification of the data sources to a single solution for better manageability. Review of what needs to be transferred and where the calculations should occur. Investigation of possibilities in cloud computing and data storage.
Firewalls block access	
Data is in multitude of different systems	
Data transfer speed is low from plants to outside	Improvements in the data transfer capabilities outside of the facilities.
Data transfer speed from devices is low	Implementation of industrial internet to increase transfer speeds and to alleviate congestion in automation system.
Data is missing or poorly recorded	Review of data management and recording processes. Better training of the data management skills of personnel.
Data is hard to process automatically	Review of what formats the data is stored in. Development of more appropriate machine-readable data recording methods.
Permissions and legal process is cumbersome	Development of legal frameworks and contracts of data usage, so that the data is also legally available also for purposes not yet envisioned.
Measured variables may not be fully linked to degradation phenomena	Increasing the quantity of measured variables could result in the data being better related to condition of devices. Vibration measurements are well represented in literature.
Other forecasting possibilities	If CBM cannot be realised now, development could go further using TBM methods and forecasting until better systems to track the condition of devices are implemented. Use of alternate sources of data and e.g. statistical analysis. Study of physical phenomena and development of a physical or hybrid degradation models.
Other ways to reach value	Identification of what is valuable and selection of the best methods to achieve the goals

The presented improvements for the gathering of monitoring data are second in priority and can be done step by step. First step should be to improve the current data handling process to ensure that all data which currently is produced is properly stored and managed as well as is readily available for analysis. This would include improvements to existing data gathering solutions to allow easier extraction of data as well as review of the current procedures related to data.

Second step is to add more monitoring instruments to the devices. This will require the identification of which factors could be monitored and which factors are worth monitoring; some indication for this could be gained from the studies on the physical phenomena and degradation. This would allow the cost of adding instruments to be focused on those which provide best benefit. Hence, while arbitrary number of different factors could be measured, it is better to first establish what should be measured. This would allow better monitoring of the condition of individual devices.

The next step would be to extend the monitoring to all devices within a plant and for this, the collection and storage of data needs to be improved. This requires the establishment of how much data needs to be transferred and then to select an appropriate technologies and systems to facilitate that. Once the technologies and methods are selected, a solution needs to be developed and implemented. Challenge here comes from the fact that the devices operate within customer plants, and hence the OEM does not have full control over the implementation. Hence, co-operation and co-development with customers is mandatory. Considering the prerequisite work that needs to be done, this step could be possible in next two to four years.

Once it is possible to efficiently collect data from a whole plant, the capabilities should be further extended to include collection, combining and analysis data from several plants, including ones which compete against one and other, simultaneously. Here, a big data capable cloud computing and storage solution is all but mandatory. Moreover, legal side, such as the permissions to use the data will need to be considered. While this step is akin to an end goal and is realisable at earliest in five to ten years, it needs to be considered in all previous steps to ensure that it will be possible to realise this in future. Practically this means that the legal side needs to be well defined from very early onwards and that the data systems are built to allow mass transfer of data.

5. DISCUSSION

The research was set forth to investigate the use of data-driven prognostics in the industrial service business. The objective was to see how machine learning could be applied to aid provision and development of industrial services. Research on the prognostics had been already published by several authors with claims of success in various degrees in various fields. However, insight on the application of the methods in practical business setting was scarce. More investigation was thus deemed necessary.

The empirical part of the research focused on a single intrinsic case. The case organisation represents a typical service organisation within an industrial OEM, which is seeking growth through digitalisation and related concepts. Examining the application of machine learning in this environment allowed the researcher not only to contribute to the operation of the organisation but also to contribute scientifically by studying how the methods presented in literature work in business setting.

The research was done with evaluative methodology with focus greatly in practical demonstration within the context of selected case and the case organisation. The aim was to develop a working predictive model, which could forecast failures in customer-owned valve assemblies using machine learning. Furthermore, the possibilities of using the model in development of maintenance services were to be examined.

5.1 How can the OEM use available data to forecast failures of field devices and to generate service recommendations?

The previous documented research does suggest that there are several methods of forecasting. The methods can be classified into qualitative, quantitative, history-based and hybrid methods (Ribeiro & Barata 2011). The method that was employed in this study, machine learning, is classified as history-based.

It is well accepted that machine learning based methods can, at least in some cases, be used to accurately develop models capable of forecasting faults. Out of myriad of different machine learning methods, the neural networks are considered the most sophisticated and the most accurate. For this reason, various forms of neural networks were chosen as the basis upon which the predictive models were built.

The machine learning based methods rely on detection of patterns from data. With pattern detection the symptomatic state can be identified in the data as it manifests as a pattern in data. Symptomatic state is the earliest point at which the fault can be detected. When the symptomatic state is detected, it is certain that the device is on course to a failure. Using this information, the remaining useful life can be estimated. The RUL estimate can then be used to estimate the maintenance need of a device and recommend as well as schedule service. Further, if the type of upcoming failure can also be detected, the type of service can also be recommended.

In this case, however, it was not possible to build a forecasting model with sufficient accuracy using the available data, and as a direct result the literary findings can not be affirmed in practice. The explaining factor here is the words “available data”. As it was found out, the availability of data was poor and thus the applicability of the methods could neither be proven nor disproven conclusively. Therefore, the availability of data should be drastically increased, and the study repeated if conclusive answer is desired.

Furthermore, due to the low accuracy, the built models are unsuitable to be used as a base for service recommendations. In DIKW-ladder service recommendations are on the level of knowledge or wisdom and hence depend on the data or information. This also means that the type and nature of any possible recommendation is defined by the underlaying data and how well it can be refined. Moreover, what sort of recommendation is useful or desirable, is also related to the business goals and value. These goals and value can, and will, change over time and hence, to discuss what the recommendation could potentially be is currently moot from the practical point.

It is to be noted, though, that the most promising research is focused on refining the machine learning methods themselves with tests being made on simulated dataset called C-MAPPS. In this dataset, the failures are specifically inserted as a continuously increasing disruption within the monitored values, thus it is known that a point where they can be detected exists. However, the simulated data, can be reasonably be likened to what data produced by actual measuring instruments would be for at least aviation engines. While it would be hasty to dismiss the value of the C-MAPPS studies and their like, it seems that there are practical challenges and complexities that are not fully represented in the simulated data.

On the other hand, several of the case studies, such as the ones by Samanta et al. (2003), Costello et al. (2017) and Yunusa-kaltungo & Sinha (2017) rely on the measurements of vibration. Vibration, however, is not something the devices which were examined are capable in measuring. Adding measurements of vibration is, however, a solid development suggestion to increase the variety and quantity of measurements.

To summarise, the OEM, in this case, can use the available data to neither forecast failures in field devices nor generate service recommendations as the availability of data is insufficient. On the other hand, as per Ribeiro & Barata (2011), high availability of data should be one of the primary advantages of history based methods such as this, in addition to the speed of building such a model. However, there was challenges in obtaining the monitoring data, which hindered the study. While data is indeed produced in vast quantities by various instruments and devices, this does not equate to high availability. Nevertheless, it should be recognised, that the OEM inherently has less access to data and thus lower availability of data of a single plant but might have higher possibilities in combining data from several customers should they agree to this. Considering the fault records and the closed nature of various systems, the plant operators would also have less than ideal access to various data. Hence, the high availability of data should not be taken as granted. Instead, the extent of availability should be investigated on case-by-case basis and measures taken to improve it to all parties.

The difficulty of OEMs on getting the data is also something which is mentioned by Kunttu et al. (2017) in their article. Kunttu et al. (2017) mention lack of trust as the main reason in addition to technical reasons such as firewalls and frameworks, while according to Efthymiou et al. (2012), the data acquisition is hindered by the simplicity of the sensory systems. On the other hand Schnürmacher et al. (2015) claims that the challenge is not with the recording the data, but in how it can be made available to the provider legally for analysis. The lack of trust certainly can be true in some cases, but in light of this case, the technical and legal considerations are the primary challenge in obtaining the data. The case company as well as the clients have conducted business for several decades, which is not untypical for companies in more traditional industries. This is long enough time to establish both inter-company as well as inter-personal relations between the buyers, salespersons and managers. Nevertheless, Kunttu et al. (2017) seems to be right in that the unavailability of data hinders development of services.

The sensory systems can nevertheless be identified as a partial hinderance. The cause was, however, more due to the system being designed as closed systems and for a different purpose than due to their simplicity. The sensory systems in a plant are there for the day-to-day operation and process control and hence lack simple means to access the whole data outside of the systems. In fact, any systematic or standardised way to transfer large quantities of data seems to be absent.

As a solution to the data transfer problem Kunttu et al. (2017) present cloud services, in which a third party would handle the collection and storage of data and then distribute it to various industrial OEMs. The data management would be provided as a service. This, however, would only add an unnecessary middleman and thus increase costs for both the OEM and the client. Moreover, the framework over which the various data is transmitted still needs to be developed. There are benefits which such a centralised solution could bring, namely the ease of use and management. It should be noted, though, that currently no such service exists. Nevertheless, a cloud service without a middle-man could be a solution for the storage of data where the data would be readily available. However, this does not solve the problem of transmitting the data out of the plants.

As per Ali-Marttila et al. (2017)s classification, the clients fall mostly into the classes of basic and quality-oriented partners. The former group sees the services as mere transactions while the latter is interested in the outcome of a service solution, but not so much in the co-development of them. On the other hand, scholars e.g. Kindström et al. (2013) assert that customers should be contacted early in the service design process of innovative concepts so that the new ideas and expectations can be jointly refined. Should Ali-Marttila et al. (2017)s notions be accurate, there seems to be dilemma in here, at least considering the field of maintenance services. On one hand the services should be jointly refined starting from early, but on the other hand the very same clients with whom the refinement should be conducted do not exhibit great interest in such ventures. Tepid responses were also encountered in this study when customers were to participate in this study, however, this could be result of the unsolicited approach and that this study was a time constrained master's thesis. Moreover, the subject was approached with methods first as the ways in which the particular set of data could be refined and turned to value of unknown quality were sought. This, perhaps, did not give adequate understanding or reasoning of the prospects and possibilities of this study and hence did not entice participation.

It is also relevant to question whether the quantities that were attempted to be predicted are important. There is a great academic focus on the RUL prediction, and in some fields, RUL might indeed be an important measure, but it seems to be taken granted and any possible alternatives are left with lesser attention. Additionally, the focus has been greatly on the more expensive and complex pieces of machinery, such as aircraft engines, which due to these qualities are generally well monitored and maintained. Most of the devices are not aircraft engines, and some are in hard to reach places with very little monitoring and maintenance. Therefore, it is reasonable to question to what extent the research on the methods of failure detection and prognostics conducted in the context of engines are applicable to similar tasks in other types of devices. The devices that were dealt with certainly fall into the latter category, yet directed by the other studies, it was too decided that RUL is an important quantity to predict without really considering what other types of data driven information could have resulted in value.

5.2 In what ways can the OEM use the forecast to develop MRO services offered to customers?

Using the available data, it was ultimately not possible to develop reliable forecasts and hence, the ways in which it could have been used can not be assessed in practice. Additionally, it is not clear if such a forecast can even be made even if the quantity of data is increased without adding additional measurements or sets of data. Hence, there are currently no ways in which the OEM could use the forecast to develop MRO service offering without first placing effort in the development of required system of data collection and management.

According to the literature reviewed, refining data into information and to knowledge has great potential in service development for example through various reports and views that can be used to support decisions. These decisions in turn create value. A forecast of failures is but one possible way to refine data, the condition monitoring data but one source of data and the development of MRO services but one possible use case. Data is produced in various processes and if processes that can serve as sources of data were identified and the different the ways to analyse them were examined, more success might be found in the improvement of processes, including the MRO service offering. Nevertheless, how exactly the produced knowledge can aid the development is entirely dependent on what sort of knowledge is produced.

Considering the services offered to customers in general, according to e.g. Grönroos & Ravald (2011) report better understanding of the service and reasoning behind the proposed actions would help the customer to better evaluate the value of the service. Better capability to evaluate the value would, in turn, aid in deciding whether to commit to an unfamiliar agreement and hence enhance the odds that a given customer would accept a service offering. Additionally, understanding is a requirement for development, and thus, if one does not understand the value, they cannot participate in the co-creation of it.

Our valued customer, who kindly provided the data to conduct this research, did it so with the condition that the study would not constrain their resources. This imposed a limitation, as any interviews and workshops that were envisioned at the start could not be conducted. Hence, it was not possible to study the needs of the customer who provided the data. However, it is still reasonable to discuss the possibilities on more general level. Several of the customers of the OEM, like the one who provided the data, has a large installed base of valves, and hence they have thousands of components of varying degrees of condition installed and limited knowledge on which require maintenance. Therefore, a service recommendation and a prediction of RUL could, potentially, help them to be better prepared to evaluate the need of maintenance and hence also the value proposition of MRO services offered by the case organisation. This could reduce the effort of manual planning and increase the overall efficiency of maintenance and the reliability of the plant.

Considering, that several operators do not have a complete knowledge on even what is installed and where, one way to develop the service offering would be to develop and implement considerably more low-tech systems to gain additional information. In addition, fundamentals, such as reliable monitoring of condition, might tremendously increase the efficiency when allocating which devices need maintenance and when.

The increase in the reliable sources of information would also increase the ways in which it is possible to combine them. This is important, as wisdom, which is the skills to combine information and knowledge from different sources, is what results in decisions and hence to actions and ultimately value. Moreover, there are many different decision makers at various levels ranging from the technicians responsible of daily maintenance to the top management. All of these have not only their own functions but also their own ways and capabilities to understand and interpret the information and knowledge presented to them.

It is therefore that the prediction could have been employed as a part of several reports, recommendations and services, each of which view the same concepts of device condition, reliability and maintenance from different perspectives. However, it is important to remember that the activities cannot be hoped to create value unless the customers value generation is supported.

6. CONCLUSIONS

In this research the data-driven prognostics were investigated in the context of the industrial service business. The goal was to discover how the faults in field devices could be predicted with machine learning methods and how the results could be applied to improve service business. The subject was approached by practical examination within a single intrinsic case. Here, the summary, conclusions and key findings are presented.

It was hypothesised, that, should the upcoming failures be predictable with sufficient accuracy, the prediction could be used to help planning of services offered to clients. It was clear that the goal would be to improve the offered services in value generating way. Hence, the literature was pored over rigorously on the subjects of industrial services and value. It was found out, that the value stems from thoughtful decisions and actions, both of which could be supported by appropriate knowledge. This knowledge, in turn, could be refined from data. One of the key concepts here is the DIKW-ladder, which describes how the data can be turned to wisdom through steps of refinement. The DIKW-ladder was supplemented with the notion of value, which stems from actions supported by the wisdom and thus concluded that data can be an asset in facilitating customer value if the DIKW-hierarchy can be rigorously followed.

Failures can be detected from patterns the deterioration induces to the measured signals. The temporal point, in which the patterns appear, is known as the functional limit. The detection of this point is key in prognostics as it is the earliest sign of an upcoming failure and from this point onwards the Remaining Useful Life (RUL) can be estimated. As the pattern evolves towards the looming failure, the estimate can be progressively updated. Additionally, according to consensus established through the review literature, a forms of machine learning known as neural networks are the most capable data-driven method in pattern detection and in prognostics. Therefore, the approach was based on the detection of the functional limit from measurement data with neural networks.

Data that could be linked to failures was needed to build and test a system capable of predicting the failures. The appropriate available data set was condition monitoring data as well as records of failures. The focus was kept on practical evaluation, and hence the data was collected from an operational production facility of a valued client. An explicit permission to conduct investigative research was granted by the customer and as agreed, the data was destroyed after the analysis was complete. As the condition monitoring data recorded measurements from ND9-series of valve positioners were used. The records of failure were obtained from maintenance ticketing system.

Using Matlab 30 different neural networks were built. These networks were trained and subsequently tested with collected data to examine their capabilities in prediction. The best performing networks were further optimised to search for even better performance. The results were unfortunately vague and shallow; inconclusive. The precision of even the best of the models was deemed insufficient. The variance of the predictions was of such a scale, that none of the models could not be thought to be accurate. It shall lamented that for the quest for knowledge did not provide results that were hoped. Nevertheless, there was wisdom to be found within the valiant effort.

Several key factors which contributed to the failure to create a prognostic model were pinpointed, and measures with which to fix these were presented. The availability of the data was assumed to be greater than it was, its accessibility to be easier, and the data to be more numerous in quantity. Had the study not been attempted, the issues would not have been identified. Hence, while the evaluation of the phenomena in practice ended in a figurative dead-end, the efforts were not in vain.

Direction of development is clear if effective condition-based maintenance, automated service recommendations or data-driven planning aids are desired. The case organisation needs to develop the fundamental capabilities, systems and processes to be better prepared to evaluate the condition of devices as the current practises were found to be lacking. The OEM was too confident that the data was adequately available and a realistic view on the current situation and challenges was absent.

Our experiences reflect what is written in the academic literature. Firstly, there is no trusted standard for intercompany data transfer that could be easily leveraged to securely transit vast amounts of data from the devices operated in a production facility to an OEM. Secondly, in addition to technical barriers, there are legal and managemental barriers obstructing the flow of data from the devices to OEM. Thirdly, the existing capabilities and infrastructure are developed with day-to-day operation of the facilities in mind and hence are not fully adequate for intercompany data transfer. This hinders any meaningful research and analysis of the data that the devices already produce in large scale. The lack of data in turn negatively affects the innovation and service development.

Additionally, there is a lack of practically oriented research in the field of data-driven prognostics. Significant amount of the conducted studies are performed on the C-MAPPS dataset, which is a popular simulated set of data. There is a general assumption that this set of data represents the progression of degradation and subsequent failures adequately, but the validity of this assumption is rarely examined. Moreover, there is little insight in how the predictions could be incorporated as a part of condition management and maintenance practices, yet alone how such could be offered as a service. Considering the current paradigm of servitisation, this is slightly alarming.

As the OEMs in the more traditional mechanical fields are gaining an increasingly large portion of revenue from services rather than the products and installations, there is a need for closer partnerships between the organisations. Moreover, as digitalisation and utilization of data can be seen as one of the key drivers with great possibilities, whether the foundation is solid and what needs to be developed is something that organisations should consider as a first step. Processes should be examined to discover where data-driven knowledge could support value facilitating decisions and where the data could be collected. A holistic approach to data management and analysis can potentially change the way maintenance services are offered and deepen the relations between the OEMs and their customers. Additional research is warranted.

REFERENCES

Ahmad, R. & Kamaruddin, S. (2012). An overview of time-based and condition-based maintenance in industrial application, Computers & Industrial Engineering, Vol. 63(1), pp. 135-149.

Ali-Marttila, M., Marttonen-Arola, S., Kärri, T., Pekkarinen, O. & Saunila, M. (2017). Understand what your maintenance service partners value, Journal of Quality in Maintenance Engineering, Vol. 23(2), pp. 144-164.

Anttila, P. (2007). Realistinen evaluaation ja tuloksellinen kehittämistyö
Akatiimi Oy, Hamina.

Babu, G.S., Zhao, P. & Li, X. (2016). Deep convolutional neural network based regression approach for estimation of remaining useful life. In collection: Navathe, S., Weili W., Shekhar S., Du X., Wang, S., Xiong, H., Database systems for advanced applications: 21st international conference on Database Systems for Advances Applications Proceedings, Part 1, pp. 214-228.

Bai, S., Kolter, J.Z. & Koltun, V. (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, Available at: http://arxiv.org/abs/1803.01271.

Banks, J., Reichard, K., Crow, E. & Nickell, K. (2009). How engineers can conduct cost-benefit analysis for PHM systems, IEEE Aerospace and Electronic Systems Magazine, Vol. 24(3), pp. 22-30.

Chen, Y., Zhen, Z., Yu, H. & Xu, J. (2017). Application of Fault Tree Analysis and Fuzzy Neural Networks to Fault Diagnosis in the Internet of Things (IoT) for Aquaculture, Sensors, Vol. 17(1), pp. 1-15.

Costello, J.J.A., West, G.M. & McArthur, S.D.J. (2017). Machine Learning Model for Event-Based Prognostics in Gas Circulator Condition Monitoring, IEEE Transactions on Reliability, Vol. 66(4), pp. 1048-1057.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A.J. & Sheikh, A. (2011). The case study approach, BMC Medical Research Methodology, Vol. 11(1), pp. 100.

Efthymiou, K., Papakostas, N., Mourtzis, D. & Chryssolouris, G. (2012). On a Predictive Maintenance Platform for Production Systems, Procedia CIRP, Vol. 3 pp. 221-226.

Fraser, K., Hvolby, H. & Tseng, T. (2015). Maintenance management models: a study of the published literature to identify empirical evidence, International Journal of Quality & Reliability Management, Vol. 32(6), pp. 635-664.

Gokulachandran, J. & Mohandas, K. (2015). Prediction of cutting tool life based on Taguchi approach with fuzzy logic and support vector regression techniques, International Journal of Quality & Reliability Management, Vol. 32(3), pp. 270-290.

Granata, F. & De Marinis, G. (2017). Machine learning methods for wastewater hydraulics, Flow Measurement and Instrumentation, Vol. 57 pp.1-9.

Grönroos, C. & Ravald, A. (2011). Service as business logic: implications for value creation and marketing, Journal of Service Management, Vol. 22(1), pp. 5-22.

Hakanen, T., Mikkola, M. & Jähi, M. (2017). Palvelunäkökulma teollisen internetin liiketoimintamallien rakentamiseen. In collection: M. Martinsuo & T. Kärri: Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa, pp. 28-39. Kerava, Kunnossapitoyhdistys ProMaint.

Iannone, C. (2017). Postmodern Truth? Academic Questions, Vol. 30(2), pp. 129-133.

Intelligent Valve Controller ND9000F Device Revision 6 User's Guide. Available at: https://valveproducts.metso.com/documents/neles/IMOs/en/ND9000F_Users_guide.pdf Accessed 19.05.2019.

Jayaswal, Pratesh, Verma, S.N. & Wadhwani, A.K. (2016). Application of ANN, Fuzzy Logic and Wavel. Engineering Economics, Vol. 27(2). pp. 190-213.

Kaasinen, E. & Liinasuo, M. 2017. Ihmiseltä ihmiselle – kilpailuetua palvelukokemuksella. In collection: M. Martinsuo & T. Kärri: Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa, pp. 41-51. Kerava, Kunnossapitoyhdistys Promaint ry.

Khalifa, A.S. (2004). Customer value: a review of recent literature and an integrative configuration, Management Decision, Vol. 42(5), pp. 645-666.

Kindström, D., Kowalkowski, C. & Sandberg, E. (2013). Enabling service innovation: A dynamic capabilities approach, Journal of Business Research, Vol. 66(8), pp. 1063-1073.

Kowalkowski, C. (2006). Enhancing the industrial service offering, LiU-Tryck, Linköping, pp. 13-41.

Kunttu, S., Ahonen, T. & Kortelainen, H. (2017). Tiedon jalostusastetta nostaen parempia palveluita ja viisaampia päätöksiä. In collection: M. Martinsuo & T. Kärri: Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa, pp. 16-26. Kerava, Kunnossapitoyhdistys Promaint ry.

Laurila, F. (2017). Asiakasarvon ja ansaintalogiikan yhteensovittaminen teollisen internetin palveluliiketoiminnassa. Available at: http://URN.fi/URN:NBN:fi:tty-201709221945 .

Lekha, S. & Suchetha, M. (2018). A Novel 1-D Convolution Neural Network With SVM Architecture for Real-Time Detection Applications, IEEE Sensors Journal, Vol. 18(2), pp. 724-731.

Lenahan, T. (2011). Turnaround, Shutdown and Outage Management: Effective Planning and Step-by-Step Execution of Planned Maintenance Operations, Elsevier Science & Technology, Oxford.

Li, X., Ding, Q. & Sun, J. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks, Reliability Engineering and System Safety, Vol. 172 pp. 1-11.

Lim, P., Goh, C.K. & Tan, K.C. (2016). A time window neural network based framework for remaining useful life estimation, Proceedings of International Joint Conference on Neural Networks, pp. 1746-1753.

Louen, C., Ding, S.X. & Kandler, C. (2013). A new Framework for Remaining Useful Life Estimation Using Support Vector Machine Classification, Conference on Control and Fault-Tolerant Systems (SysTol), 2013, Nice, France, pp. 228-233.

Lusch & Vargo (2014). The Service Dominant Logic of Marketing: Dialog Debate and Directions. Routledge, New York.

Macdonald, E., Wilson, H., Martinez, V. & Toossi, A. (2011). Assessing value-in-use: A conceptual framework and exploratory study, Industrial Marketing Management, (40), pp. 671-682.

Malhotra, P., TV, V., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P. & Shroff, G. (2016). Multi-Sensor Prognostics using an Unsupervised Health Index based on LSTM Encoder-Decoder, available at: https://arxiv.org/abs/1608.06154. Accessed: 22.5.2019.

Marinelli, M., Lambropoulos, S. & Petroutsatou, K. (2014). Earthmoving trucks condition level prediction using neural networks, Journal of Quality in Maintenance Engineering, Vol. 20(2), pp. 182-192.

Mrad, N., Foote, P., Giurgiutiu, V. & Pinsonnault, J. (2013). Condition-Based Maintenance, International Journal of Aerospace Engineering, Vol. 2013, pp. 1-2.

Pawar, K.S., Beltagui, A. & Riedel, J.C.K.H. (2009). The PSO triangle: designing product, service and organisation to create value, International Journal of Operations & Production Management, Vol. 29(5), pp. 468-493.

Peng, Yu, Wang, Hong, Wang, Jianmin, Liu, Datong & Peng, Xiyuan (Jun 2012). A modified echo state network based remaining useful life estimation approach, 2012 IEEE Conference on Prognostics and Health Management, IEEE, pp. 1-7.

Ramasso, E. (2014). Investigating computational geometry for failure prognostics, International Journal of Prognostics and Health Management, Vol. 5(1), pp. 1-18.

Raza, Jawad, Liyanage, Jayantha, Al Atat, Hassan & Lay, Lee (2016). A comparative study of maintenance data classification based on neural networks, logistic regression and support vector machines, Engineering Economics, Vol. 27(2), pp. 303-318.

Ribeiro, L. & Barata, J. (2011). Re-thinking diagnosis for future automation systems: An analysis of current diagnostic practices and their applicability in emerging IT based production paradigms, Computers in Industry, Vol. 62(7), pp. 639-659.

Samanta, B., Al-Balushi, K.R. & Al-Araimi, S.A. (2003). Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection, Engineering Applications of Artificial Intelligence, Vol. 16(7), pp. 657-665.

Schnürmacher, C., Hayka, H. & Stark, R. (2015). Providing Product-Service-Systems - The Long Way from a Product OEM towards an Original Solution Provider (OSP), Procedia CIRP, Vol. 30 pp. 233-238.

Siggelkow, N. (2007). Persuasion with case studies, Revista Eletrônica de Estratégia & Negócios, The Academy of Management Journal Vol. 50(1), pp. 20-24.

Smith, L., Maull, R. & Ng, I.C.L. (2014). Servitization and operations management, International journal of operations & production management, Vol. 34(2), pp. 242-269.

Susto, G.A., Beghi, A. & De Luca, C. (2012). A Predictive Maintenance System for Epitaxy Processes Based on Filtering and Prediction Techniques, IEEE Transactions on Semiconductor Manufacturing, Vol. 25(4), pp. 638-649.

Susto, G.A., Schirru, A., Pampuri, S., McLoone, S. & Beghi, A. (2015). Machine Learning for Predictive Maintenance: A Multiple Classifier Approach, IEEE Transactions on Industrial Informatics, Vol. 11(3), pp. 812-820.

Takata, S., Kimura, F., Houten, F J A M van, Westkamper, E., Shpitalni, M., Ceglarek, D. & Lee, J. (2004). Maintenance: Changing Role in Life Cycle Management, CIRP Annals - Manufacturing Technology. Vol. 53. pp. 643-655.

Toossi, A., Louise Lockett, H., Z. Raja, J. & Martinez, V. (2013). Assessing the value dimensions of outsourced maintenance services, Journal of Quality in Maintenance Engineering, Vol. 19(4), pp. 348-363.

Tukker, A. (2004). Eight types of product–service system: eight ways to sustainability? Experiences from SusProNet, Business Strategy and the Environment, Vol. 13(4), pp. 246-260.

Saxena, A. & Goebel, K. (2008) Turbofan Engine Degradation Simulation Data Set. NASA Ames Research Center, Moffett Field, CA, USA. Available at: https://data.nasa.gov/widgets/vrks-gjie Accessed. 22.5.2019.

Vaittinen, E., Martinsuo, M. & Nenonen, S. (2017). Ratkaisua täydentävien palvelujen omaksuminen ja käytön edistäminen. In collection: M. Martinsuo & T. Kärri: Teollinen internet uudistaa palveluliiketoimintaa ja kunnossapitoa, pp. 52-62. Kerava, Kunnossapitoyhdistys Promaint ry.

Weiss, R. (2000). Taking Science out of Organization Science: How Would Postmodernism Reconstruct the Analysis of Organizations?, Organization Science, Vol. 11(6), pp. 709-731.

Wu, Y., Yuan, M., Dong, S., Lin, L. & Liu, Y. (2018). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks, Neurocomputing, Vol. 275 pp. 167-179.

Yang, C., Letourneau, S., Liu, J., Cheng, Q. & Yang, Y. (2017). Machine learning-based methods for TTF estimation with application to APU prognostics, Applied Intelligence, Vol. 46(1), pp. 227-239.

Yunusa-kaltungo, A. & Sinha, J. (2017). Effective vibration-based condition monitoring (eCVM) of rotating machines, Journal of Quality in Maintenance Engineering, Vol. 23(3), pp. 279-296.

Zhang, C., Lim, P., Qin, A.K. & Tan, K.C. (2017). Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics, IEEE Transactions on Neural Networks and Learning Systems, Vol. 28(10), pp. 2306-2318.