Lappeenranta University of Technology
LUT School of Business and Management
Industrial Engineering and Management
Business Analytics

Miska Valkonen

PREDICTIVE MAINTENANCE FOR VALMET'S BREAST ROLL SHAKER

Author: Miska Valkonen
Examiners: Professor Pasi Luukka
Post Doctoral Researcher Jan Stoklasa

Supervisors: Global Product Manager Markku Savioja

TIIVISTELMÄ

Lappeenrannan teknillinen yliopisto
LUT School of Business and Management
Tuotantotalous
Business Analytics

Miska Valkonen

Ennakoiva kunnossapito Valmetin rintatelan ravistimelle

Diplomityö

2019

86 sivua, 36 kuvaa, 7 taulukkoa ja 7 liitettä

Tarkastajat: Professori Pasi Luukka
Tutkijatohtori Jan Stoklasa

Hakusanat: Business analytiikka, ennakoiva kunnossapito, koneoppiminen, ennakoiva mallintaminen, pilvilaskenta, teollinen internet, tekoäly

Digitalisaatio ja teollinen internet mahdollistavat suuren datamäärän keräyksen ja analysoinnin. Tätä suurta dataa ei yleensä hyödynnetä optimaalisesti, etenkään kunnonvalvonnan ja huoltotoiminnan näkökulmasta. Tarpeeton huoltotoiminta lisää kustannuksia, mutta toisaalta huoltamisen laiminlyönti voi aiheuttaa merkittävää vahinkoa laitteeseen tai koneeseen. Tämän diplomityön tavoitteena on luoda ennakoivaa huoltoa tukevia malleja hyödyntämällä laitteistosta kerättyä dataa. Mallit mahdollistavat tarpeettomien huoltojen minimoinnin sekä yllättävien vikatilanteiden ehkäisyn.

Ennakoivan huollon mallit luotiin kohdeyrityksen laitteistolle. Onnistuneen mallin luonti vaatii laitteiston syvää tuntemusta, prosessiosaamista sekä sopivien analyyttisten mallien sovittamista. Tässä diplomityössä tutkitaan laitteiston toimintaperiaatteita ja pyritään tunnistamaan ongelmakohdat sekä prosessissa että huoltotoiminnassa. Työ sisältää kattavan kirjallisuuskatsauksen ennakoivan kunnossapidon menetelmiin, algoritmeihin ja ennustusmenetelmiin teollisuudessa. Sopivimmat menetelmät valittiin kirjallisuuskatsaukseen perustuen ja ennakoivan kunnossapidon mahdollistavat mallit luodaan.

Malleja kehitetään sekä aikasarjadatan ennustamiseen että moniulotteisen luokitteluongelman ratkaisemiseen. Aikasarjadatan ennustamisessa useita menetelmiä testataan kuten lineaariregressio, ARMA ja VAR. Luokittelu-ongelmaan hyödynnetään koneoppimisen menetelmiä ja valittu menetelmä perustuu erialisiin variaatioihin päätöspuu algoritmeista. Ennakoivat mallit sijoitettiin serverless AWS-ympäristöön ja analyysien tulokset lasketaan reaaliajassa. Tulokset visualisoidaan Tableau-työkalulla. Erityyppisiä malleja käytetään löydettyihin ongelmakohtiin ja luotettavia ennustuksia luodaan laitteiston kunnon mallintamiseksi sekä yllättävien vikatilanteiden ehkäisemiseen.

ABSTRACT

Lappeenranta University of Technology
LUT School of Business and Management
Industrial Engineering and Management
Business Analytics

Miska Valkonen

Predictive maintenance for Valmet's Breast Roll Shaker

Master's Thesis

2019

86 pages, 36 figures, 7 tables, and 7 appendices

Examiners: Professor Pasi Luukka
Post Doctoral Researcher Jan Stoklasa

Keywords: Business Analytics, Predictive Maintenance, Machine Learning, Predictive Modeling, Cloud Computing, Industrial Internet, Artificial Intelligence

Digitalization and the Industrial Internet of Things (IIoT) enables the collection and analysis of vast amounts of data. Big data is often not utilized optimally, especially when regarding the condition and maintenance of different equipment. Unnecessary maintenance increases costs, and lack of maintenance may cause significant damage to the equipment and production loss. The goal of this thesis is to utilize the acquired data to create a predictive maintenance model to prevent unplanned shutdowns and decrease unnecessary maintenance costs.

A predictive maintenance model is created for the case company's equipment. A successful predictive maintenance model requires knowledge of the equipment, the underlying process, and appropriate analytical methods. This thesis researches the equipment in question thoroughly to identify key issues in the current maintenance plan and to understand how the equipment behaves. An exhaustive literature research is performed to review the current industrial applications, models, and algorithms of predictive maintenance and prediction of remaining useful life (RUL). The most suitable approaches are selected for the equipment in question and predictive maintenance models are created.

Models are developed to solve both time series regression and classification tasks. To predict the time series data efficiently, multiple models are compared such as linear regression, ARMA, and VAR. For the classification task, modern machine learning methods are applied, and the most accurate model is selected. In this case, different variations of a decision tree classifier are used. The predictive maintenance models are deployed to AWS Lambda to run serverless and real-time, and the results are visualized using the Business Intelligence (BI) tool Tableau. Different models are used for different key issues and confident predictions are made of the equipment's condition to prevent unplanned shutdowns.

ACKNOWLEDGEMENTS

I would first like to thank my supervisors Professor Pasi Luukka, Markku Savioja, and Post Doctoral Researcher Jan Stoklasa for their invaluable guidance through the course of this thesis. My sincere gratitude to everyone at Valmet who assisted and encouraged me to pursue and overcome the challenges I so often faced. Thanks to my fellow graduate students who made the time I spent at Lappeenranta unforgettable. And finally, I would like to thank my family and friends for all the love and support before, during and beyond this research.

Machine learning, artificial intelligence, predictive modeling, and the industrial internet appear intimidating when faced without a context or knowledge about the subject. This thesis taught me both that during my studies at Lappeenranta and career at Valmet, I have barely scratched the surface of the topics and there is much more to learn. However, they are not intimidating anymore, I am looking forward to learning more.

"All models are wrong, but some are useful"

- *George Box*

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

AI	Artificial Intelligence
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARMA	Autoregressive Moving Average
AWS	Amazon Web Services
BRS	Breast Roll Shaker
BI	Business Intelligence
CI/CD	Continuous Improvement / Continuous Development
CNN	Convolutional Neural Network
CSV	Comma Separated Values
DSC	Distributed Control System
DT	Decision Tree
EDW	Enterprise Data Warehouse
ETL	Extract, Transform, and Load
FaaS	Function as a service
FFN	Feedforward neural network
FM	Form Master
GNB	Gaussian Naïve Bayes
IaaS	Infrastructure as a service
IIoT	Industrial Internet of Things
IoT	Internet of Things
JDBC	Java Database Connectivity
JSON	Javascript Object Notation
KNN	K-Nearest Neighbors
KPI	Key Performance Indicator
MCS	Machine Control System
ML	Machine Learning
MLP	Multi-Layer Perceptron
NN	Neural Network

PaaS	Platform as a service
PE	Paris-Erdogan
ODBC	Open Database Connectivity
PDF	Probability Distribution Function
PdM	Predictive Maintenance
PH	Proportional Hazards
RBAC	Role Based Access Control
RNN	Recurrent Neural Network
RUL	Remaining Useful Life
R&D	Research and Development
SaaS	Software as a service
S3	Amazon Simple Storage Service
SF	Snowflake
SFTP	Secure File Transfer Protocol
SNS	Simple Notification System
SQL	Structured Query Language
SVM	Support Vector Machine
SVR	Support Vector Regression
VAR	Vector Autoregression
VII	Valmet Industrial Internet

1 INTRODUCTION

In the era of digitalization and digital transformation, many industrial companies are developing different industrial internet solutions to gain competitive advantages as suppliers or service providers. The industrial internet, cloud computing, data analytics, machine learning, performance optimization, and predictive maintenance are currently trending topics and consequently focus points for companies that can benefit widely from utilizing these tools and services. Increased connectivity of industrial equipment and components rigged with a large number of different sensors provide continuous data, that allows for a comprehensive and extensive understanding of the equipment's state and condition. This enables optimization of the equipment's performance by means of analyzing and visualizing the accumulated data. (Collin & Saarelainen 2016)

Predictive maintenance (PdM) is one crucial output that can be derived from the sensor data that is collected. Typical industrial and process plant maintenance techniques such as run-to-failure management and scheduled maintenance have major disadvantages; either high costs from production plant downtime or high costs from unnecessary maintenance. The goal of predictive maintenance is to improve the overall effectiveness of manufacturing and production plans - this is achieved by using the actual operating condition of different components to optimize total plant operation. Scheduling maintenance activities on an as-needed basis improves productivity and product quality by minimizing downtime and maintenance costs. There are numerous ways to develop a predictive maintenance model, but there are also requirements to get a successful predictive maintenance concept to work optimally. Connectivity in the sense of data variety, volume and velocity are essential. A sufficient set of tools and services are needed for data collection, analytics, and visualization. Predictive models can be based on first principles or data-driven. Physics based-models are based on deep knowledge of the process and known behavior of data when faults are about to occur. Data-driven models are taught to learn these behaviors and complex relationships using historical data and are then used to identify possible faults in the future. (Lei et al. 2018 & Mobley 2002)

1.1 Background

The subject for this master's Thesis was chosen by the strong incentives of benefits and profits that a functional predictive maintenance concept delivers. The Industrial Internet of Things (IIoT) is the enabler for data collection, storage, and utilization. Typically, data is collected and monitored on-site, or examined by process experts remotely to identify unhealthy behavior of the machine part or equipment. However, this can be time-consuming and when there is a large number of machines and a limited number of professionals for specific equipment, it is almost impossible to monitor all the machinery simultaneously.

Valmet is the leading global developer and supplier of automation, services, and technologies for the paper, pulp, and energy industries. Valmet's extensive technology portfolio consists of pulp-, board-, tissue-, and paper mills and bioenergy power plants. Valmet focuses on enhancing customers' productivity and value with new cost-efficient equipment and solutions for optimizing raw material and energy usage, automation solutions and plant upgrades and rebuilds. Must-Wins pursued by Valmet are 'customer excellence, 'leader in technology and innovation', 'excellence in processes', and 'winning team'. (Valmet 2018) Valmet has an industrial history of over 200-years, the company was reformed in the December of 2013, when the pulp-, paper- and power plant business lines were separated from Metso Oyj. Valmet's net sales in 2018 were around 3.3 billion euros and the company employed around 12 000 professionals all around the world. (Valmet 2018)

Valmet's breast roll shaker is a typical component in a paper or board machine line. In short, it is used to increase the quality of the end product by mixing the raw material fibers to a better formation. This is achieved by a "shaking" effect in the process, thus the name "breast roll shaker". Data from this component have been available and collected for a few years. Maintenance management for the breast roll shaker is scheduled in a preventive manner; at intervals spaced weekly, monthly, or annually there is maintenance performed either by the customer or by Valmet personnel. Corrective maintenance is also utilized - in the case of a fault, or if there is a problem with the component, the collected data is examined to see what possible causes for the problem or fault could be. Valmet utilizes cloud services for data collection and storage. The business intelligence tool Tableau is used for data visualization. The vast amount

of data that is currently available is not thoroughly utilized and there is an opportunity to create value for both Valmet and its customers. The reader is assumed to be familiar with the basics of databases and structured query language SQL, fundamentals of the papermaking process, elements of programming with python, and principles of statistical computing.

1.2 Objectives and limitations

This thesis explores existing predictive maintenance approaches in industrial environments and finds the most suitable methods to be used for the case company. The goal is to *develop and deploy a predictive maintenance model to prevent an unplanned shutdown of the equipment*. The goal is achieved by following four objectives of this thesis:

- *Research different predictive models and algorithms, their strengths and weaknesses, to find the most suitable methods to create a predictive maintenance model for the breast roll shaker.*
- *Identify current maintenance plan weaknesses and possible key degradation components to focus on.*
- *Study of the breast roll shaker mechanics and mode of operation.*
- *Research and utilization of the case company's industrial internet architecture for data extraction, analysis, and visualization.*

The limitations of this thesis come from the extensive but restrictive scope, data security for both the case company and its customers, and literature available for similar cases. While the goal is to create an end-to-end solution for the case company, every part of the systematic approach cannot be covered in the scope of this thesis and thus the research is focused mainly on predictive modeling. The solution is created for the case company, utilizing data from its customers – some limitations such as visualization of the data might be limited due to data security and privacy. As the equipment in question has been recently developed and the custom end-to-end solution created, a limited amount of literature is available.

1.3 Approach

Historical data that is available from the breast roll shaker makes this thesis possible and is at the core of the approach. Depending on the customer and the type, age and location of the component, approximately 1 - 2 years of historical data is accessible in the database. Additionally, the expertise of Valmet's personnel regarding the product and the process of this component is available and utilized. During this thesis, multiple experts have been interviewed and they have provided useful information and guidance in all aspects of this project. From the research and development (R&D) engineers who designed the breast roll shaker to the technology managers developing the latest industrial internet applications, all have been part of this project. The interviews are not recorded and presented directly, but have without a doubt affected the outcome of this thesis. Also, access to most of Valmet's internal databases such as product manuals, maintenance plans, and sales material is utilized.

The groundwork for this thesis starts by learning how the breast roll shaker operates, how maintenance is performed, and what is measured from the component and the process. The recognition of why different sensors are installed and what the measurements actually mean is crucial. Valmet's industrial internet architecture and way-of-work are introduced and followed. After a thorough understanding of the breast roll shaker and Valmet's industrial internet architecture, it is possible to start developing predictive maintenance models. Before theoretical frameworks are utilized, some exploratory analysis is performed to get an insight into the data in question such as resolution, size, type, and volume.

1.4 Structure of the main report

The thesis structure is divided into six chapters. The structure and contents are presented in figure 1. Each of the chapters is interconnected and the theoretical framework provides requirements for a successful approach in the empirical section.

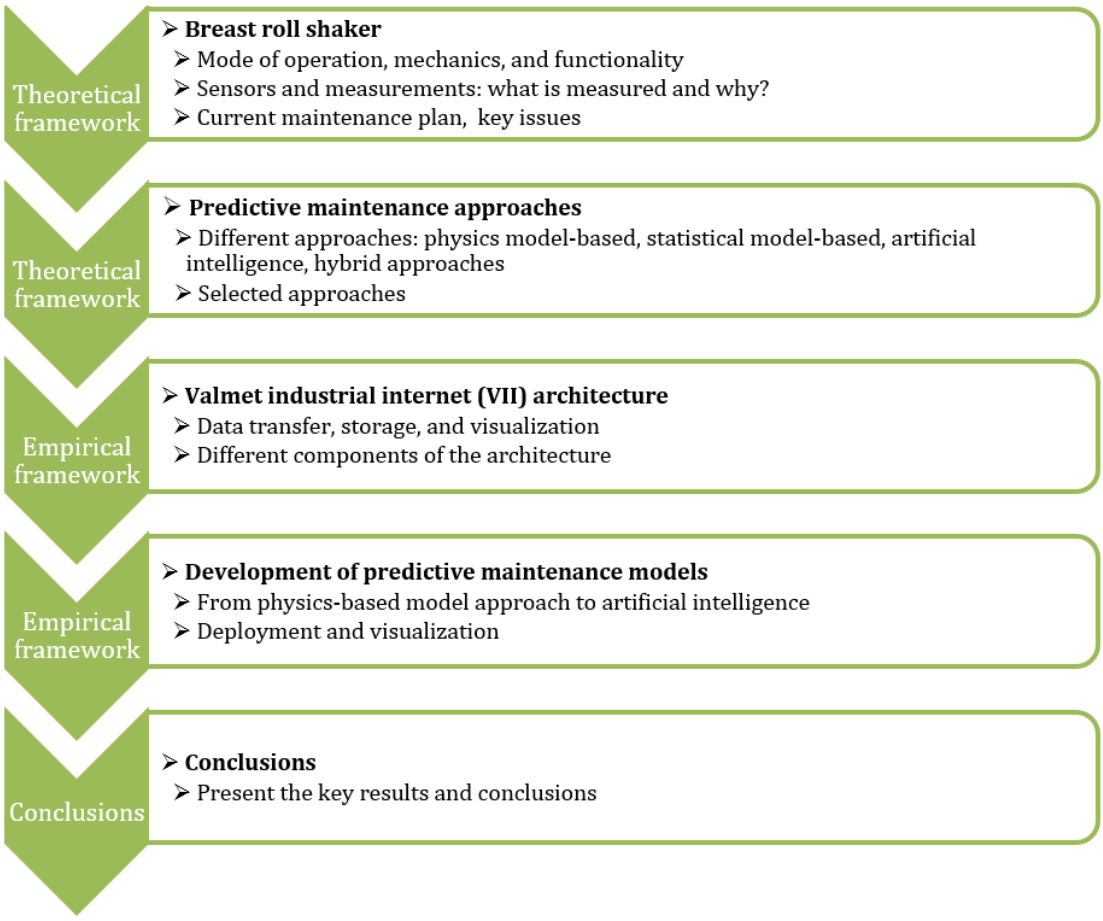


Figure 1. Structure of the report starting from the theoretical framework to conclusions.

As seen in figure 1, this study begins with the introduction of the equipment in question: the breast roll shaker. The first chapter provides knowledge of the mode of operation, different sensors and the current maintenance plan, all of which are important considerations when developing a predictive maintenance model. The literature review begins to explore different predictive maintenance approaches in the industry, introducing the methods found briefly, and focusing on the selected methods in this study. In the third chapter, Valmet's industrial internet architecture is presented, defining the different components that enable predictive maintenance. The fourth chapter is dedicated to the development and deployment of the selected methods found in the theoretical framework, utilizing the knowledge of the machine gained in chapter 2 and the architecture defined in chapter 3. The study ends with conclusions about the predictive maintenance models created, together with the key results from the empirical framework.

2 BREAST ROLL SHAKER

Predictive maintenance models require an understanding of the underlying process mechanics. This section introduces the breast roll shaker in such a level of detail, that the reader can comprehend what it is used for, the basics of the operating mode, what is measured and how maintenance is currently performed.

One of the most important structural properties of paper and paperboard is formation - strength and visual properties are strongly affected by it. To improve the sheet forming process the Breast Roll Shaker (BRS) is an effective solution. By creating shear forces on the web, it improves formation by breaking up large fiber flocs to give a better orientation. Breast roll is shaken cross-directionally by the breast roll shaker. Two pairs of rotating eccentric masses are used to generate the shaking forces, while no reaction forces are transferred to the foundation of the machine. By using a BRS, the same strength and quality can be achieved with less refining and preprocessing. Cost savings for the customer are achieved when less raw material and chemicals are needed to improve the strength properties of the paper or paperboard. (BRS Manual)

Valmet provides remote monitoring and support for the breast roll shaker. This service is provided by Valmet’s experts via a remote connection. Connectivity is established utilizing secure data transfer protocols, and the connectivity specifications depend on what kind of system the customer's machine has. Some of the previously connected breast roll shakers are now offline and some have issues somewhere in the data pipeline, and thus are not providing data even close to real time. In the development of predictive maintenance models, all the connected shakers that are still online are used.

2.1 Mode of operation

The operating principle is quite simple: with the help of rotating eccentric masses, the breast roll shaker shakes the breast roll. The breast roll is connected to the carriage where these

eccentric masses are rotating. Shaking frequency and stroke length can be adjusted. By altering the phase shift between two mutually adjustable mass pairs the shaking force, or stroke, is created. Zero-degree phase shift means that no shaking force is created, this is demonstrated on the left side of figure 2. The right side of figure 2 shows how the maximum stroke length is generated, with a phase shift of 90 degrees. A phase shift unit is used to mechanically adjust the phase shift to the desired degree. While the phase shift controls the stroke length, the frequency is controlled with the electric drive motor speeds' frequency converter. Frequency and phase shift are used to generate the optimal stroke length and frequency combination for the best formation of the fibers. Torque from the electric drive motor is transmitted via a Schmidt-coupling™. (BRS Manual)

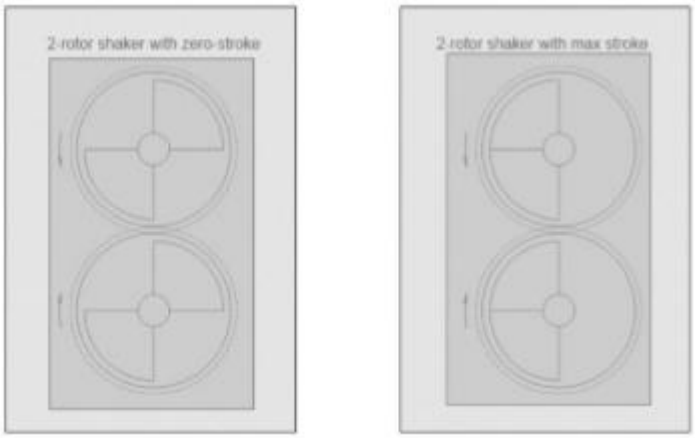


Figure 2. Phase angle operating principle of the breast roll shaker.

The carriage, that holds the eccentric mass pairs moves on a hydraulic oil film and therefore reaction forces are not conveyed to the units' foundation. The connecting rod is used to connect the internal carriage of the shaker to the breast roll shaft. Mounted on bearings, the rod is used to transfer the shaking force to the rotating breast roll shaft. A general overview of the breast roll shaker is shown in figure 3 below. For the breast roll shaker to work properly, everything must operate as expected. (BRS Manual)

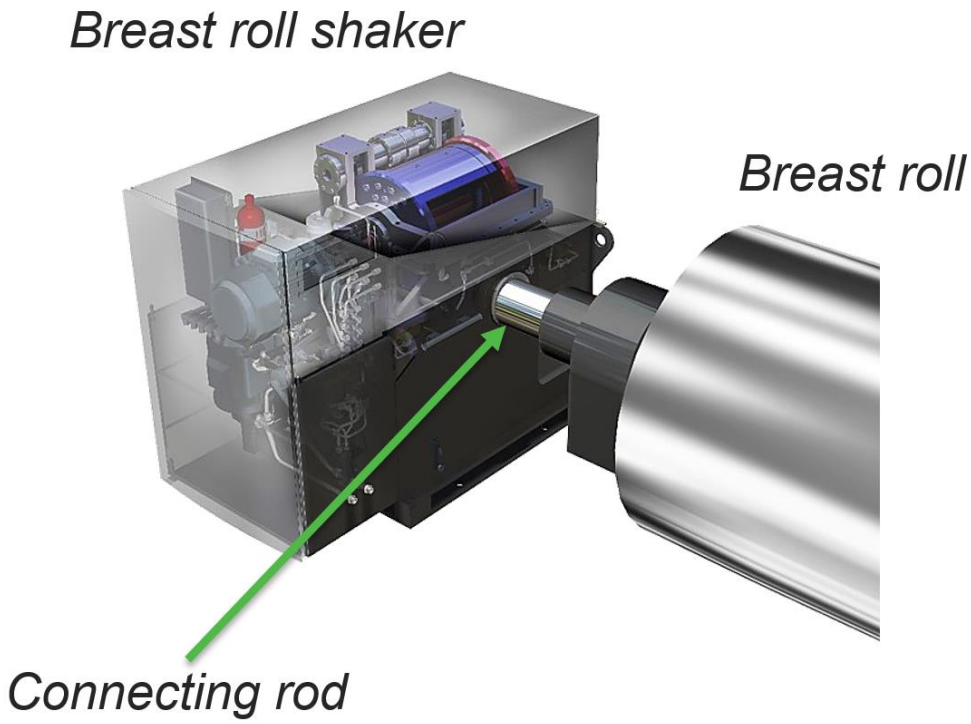


Figure 3. Breast roll shaker, breast roll, and the connecting rod.

Figure 3 presents the main operating components. The BRS includes the eccentric mass pairs, hydraulic system, lubrication system, motors, and pumps. The connecting rod is shown with the green arrow, which connects the breast roll to the BRS.

2.2 Sensors and measurements

The breast roll shaker is equipped with multiple different sensors for remote monitoring and control of the machine. For example, the stroke length and frequency are important parameters considering runnability and can be monitored or adjusted remotely. There are a total of 50 signals measured or generated from the BRS, these signals and their descriptions are listed in Appendix 1, including the typical operating range. Some signals are important from the operating point of view, such as the signals for stroke length set point and measured value - when the difference between the set point and the measured value is too high, adjustments are needed. Crucial signals from the predictive maintenance point of view are the signals used to monitor that operating conditions are healthy. For example, it is important that the oil

temperature to varies within acceptable limits. Depending on the measurement, the data resolution varies from 1 second to 600 seconds.

Stroke length and stroke frequency are the most important parameters when considering the operation of the BRS. These measurements are used to calculate the shake number (SN), which indicates the effect of shaking on stock formation. The wire speed also affects the shake number. SN increases as stroke length and shaking frequency increase and decreases as wire section speed increases. The shake number is calculated using the equation:

$$SN = \frac{n^2 \cdot S}{v_{wire}} \tag{1}$$

Where SN is the shake number, n is the shake frequency $\left[\frac{1}{min}\right]$, S is the stroke length $[mm]$, and v_{wire} is the wire section speed $\left[\frac{m}{min}\right]$. The wire section speed is the peripheral speed of the breast roll. Oil pressures and temperatures are the main measurements used for remote monitoring of the breast roll shakers condition. There are a total of 5 different lines where the oil pressures (MPa) are measured: coupling -, bearings -, carriage -, mesh -, and phase angle oil pressures. Oil temperature (C°) is measured from the hydraulic oil tank. Coupling -, bearing -, and mesh pressure measurements are in lubrication lines. As mentioned previously, the carriage moves on a thin oil film to avoid conveying forces to the foundation, thus the carriage flotation pressure is also measured. To adjust the phase angle of the eccentric mass pair, a high oil pressure is maintained in the hydraulic system. Phase angle control pressure is measured from the line used to control the phase angle. There is a pressure differential measurement in the lubrication line over the oil filter immediately after a hydraulic oil pump unit. The operating states of the breast roll shaker and its hydraulic system are also measured, states ‘zero’ or ‘one’ corresponding to states ‘off’ and ‘on’, respectively. The control system facilitates troubleshooting and the number of different alarms is comprehensive. For example, high and low oil pressures or differences between the setpoint and measured values can cause an alarm or fault of the breast roll shaker. The reason for the alarm or interlocking can be directly seen from the operating display at the customer site, but also deciphered from the numerical decimal vectors created by the control system and saved in tags MsgW_BrS_HelpW_M1_1,

MsgW_BrS_HelpW_M1_2, and MsgW_BrS_HelpW_M2. Appendix 2 lists the possible faults, corresponding measurement value, and a description of the fault.

2.3 Current maintenance plan

To get a better understanding of the benefits of predictive maintenance it's important to recognize the differences between the different categories of maintenance. Different types of maintenance are defined in the maintenance terminology standards EN-13306 and figure 4 illustrates the most common maintenance types, and how they are connected and classified into maintenance categories.

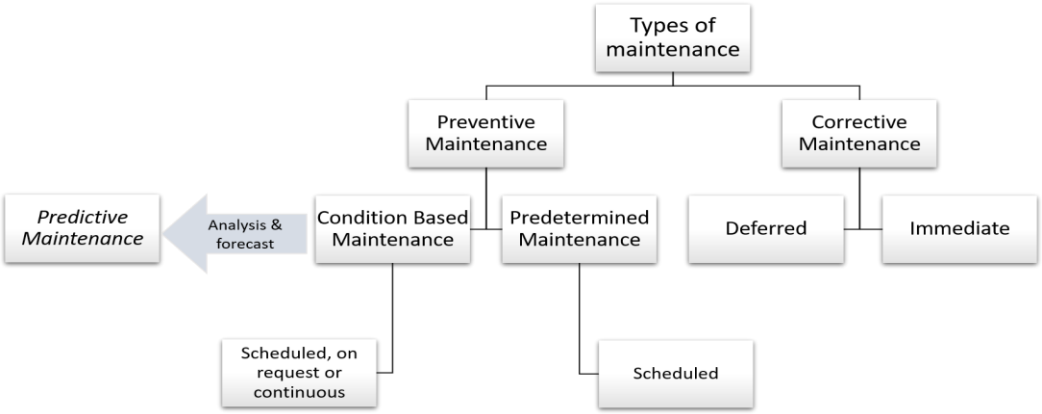


Figure 4. Different types of maintenance based on EN 13306:2010. (CEN 2010)

Different types of maintenance are divided into corrective and preventive maintenance. Corrective maintenance is usually referred to as run-to-failure (RTF), reactive maintenance, or breakdown maintenance. In other words, maintenance is carried out after the failure of equipment either immediately or after a certain amount of time. Preventive maintenance is usually time-driven and carried out based on elapsed time or hours of operation and is further categorized as condition-based maintenance (CBM) and predetermined maintenance. CBM is a type of maintenance where the condition is monitored, inspected, and tested to plan maintenance actions. For example, monitoring the pressure difference over an oil filter to ensure necessary filter replacement. Predetermined maintenance is a scheduled maintenance type,

where maintenance is carried out with established intervals of time - the interval length may be established from previous knowledge of failure mechanisms. Predictive maintenance is derived from condition-based maintenance by performing repeated analyses based on known characteristics of the significant parameters of the failure mechanisms. Forecasts are made based on these analyses to estimate the degradation of the equipment and plan maintenance accordingly. (CEN 2010, Mobley 2002)

Breast roll shakers' current maintenance plan is mostly composed of predetermined maintenance, but also includes condition-based maintenance and corrective maintenance. The predetermined maintenance plan defines the position name, maintenance specification, maintenance interval, job, and method with possible additional notes. This maintenance plan is shown in Appendix 3. As seen from the maintenance plan, the maintenance interval varies between 1 week to 104 weeks, from small visual checks to major overall maintenance. Some maintenance tasks can be performed when the machine is operational, such as checks for leaks and vibrations, while other maintenance operations require a shutdown, such as oil replacement. Condition-based maintenance is performed with the oil filter, where the replacement of the filter is performed according to the differential pressure transmitter alarm. When the oil pressure differential measurement is high, it corresponds to a clogged oil filter and thus is a clear signal for maintenance. Corrective maintenance is usually immediate if possible, but depending on available spare parts and available experts, maintenance cannot be always immediate. After an unsuspected fault occurs, the data of the breast roll shaker is analyzed to see potential reasons for the fault and corrective actions are performed accordingly.

The goal of the predictive maintenance model is to eliminate unnecessary maintenance and unplanned shutdowns by utilizing the available historical knowledge of mechanical faults and current sensor data. After interviewing Valmet service experts concerning the most common issues with the breast roll shaker, the hydraulic oil pump unit and the Schmidt coupling distinctively stand out. The axle transferring the torque in the coupling or the double hydraulic gear pump is prone to faults, and it is an issue in some breast roll shakers. Some experts believe that cavitation in the oil pump causes degradation and is the reason for the breakage. After review of the maintenance documents performed after a coupling failure, there are multiple different reasons deduced for the coupling failure such as faulty lubrication, increased strain by

running the BRS over its operational limits, or mechanical misalignment causing damage to the axle or bearings. However, even though the faults mentioned are the more frequent, they still occur very rarely. With the operation principles defined in section 2.1, sensors mentioned in section 2.2, and faults recognized previously, it is clear that at least the following models can be tested:

- *Oil leaks: Phase angle control pressure indicates the calibration of the phase angle in the case of a high difference between the set point and measured value of the stroke length. If the calibration frequency increases, without changing the setpoint value, it is a clear indication of leaks somewhere in the valves, cylinder or in the hydraulic system.*

- *Pump unit: Cavitation could be detected in the oil pressure measurements after the oil pump unit. If there is data from a known failure, a model can be developed to examine if cavitation can be detected as an increased variation in the oil pressure measurements.*

- *Alarms/Faults: Predictive models can be created for measurements that have alarm or fault (interlock) limits. By utilizing historical data, time series forecasts can be made to observe if the measurement is within acceptable limits.*

- *Classification of degradation stage. If faults are present in the historical data, different machine learning methods can be tested to classify the different stages of degradation.*

With the help of the predictive maintenance models, some of the predetermined maintenance tasks can be rescheduled. For example, if oil pump unit faults can be predicted from the pressure measurements, the monthly maintenance task (*pump unit: check sound, temperature, and vibration*) can be performed only when necessary. This methodology can also be applied to oil leak checks if the leaks can be detected from the different measurements. Predictive models can be created for the different important measurements to forecast possible alarms and faults. Modern artificial intelligence-based methods can be used, such as machine learning, to create accurate models with high interpretability. For example, the ‘alarm’ limit for bearing lubrication pressure is 3.5 MPa, and an alarm is triggered if higher values are measured. If the measurement surpasses the systems ‘fault’ limit, the value of 4.0 MPa, the breast roll shaker is shut down and

a fault is triggered: *Breast roll shaker fault: Bearing lubrication oil pressure*. A simple visualization of the measurement in question can be seen in figure 5. Some of these alarms and faults can be predicted with advanced data analytics. There are no direct measurements that indicate, for example, the failure of couplings or bearings, but it may be indicated through the complex relationships of different measurements. These relationships can be detected with modern classification methods and failures possibly prevented.

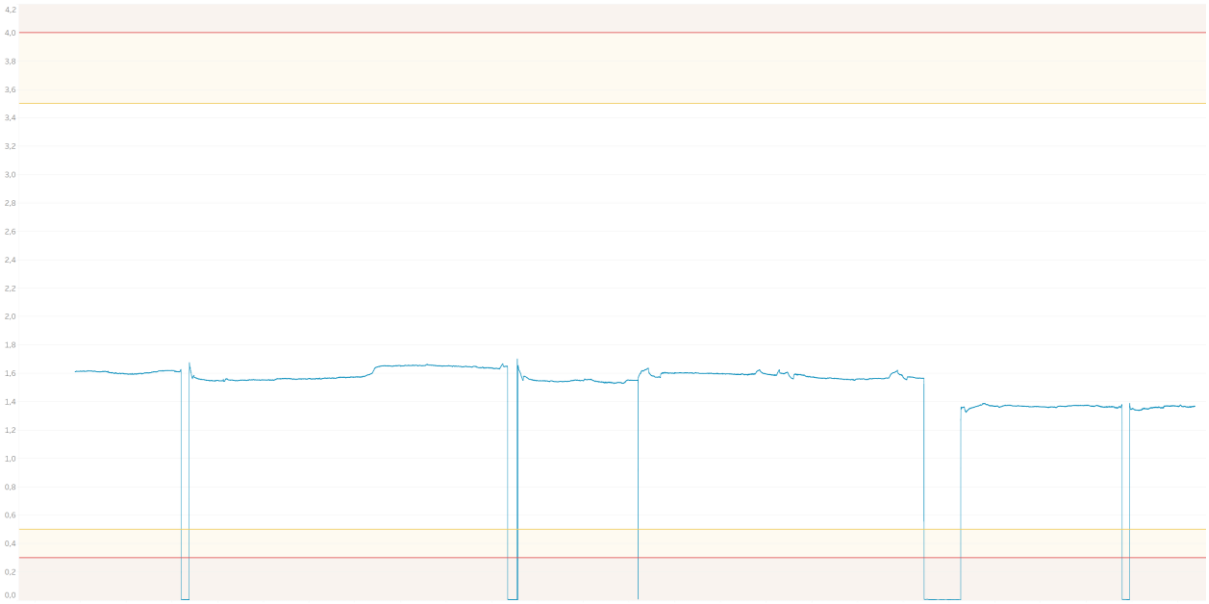


Figure 5. The trend of pressure measurement and acceptable limits. The y-axis is oil pressure (MPa) and the x-axis is time. The blue line is the measured pressure.

In figure 5 the blue line depicts the pressure measurement. Yellow and red horizontal lines are the alarm and fault limits where the pressure measurement can vary, respectively. When the measurement is zero, the machine is shut down and should not trigger an alarm. The trend was created with a business intelligence tool as a part of this thesis, as one goal is to predict the future value for the measurement to predict possible alarms and faults. For example, if the oil pressure is predicted to decrease below the lower limits, oil leak is the most probable cause.

3 PREDICTIVE MAINTENANCE APPROACHES

This chapter includes a comprehensive literature review of the methods previously used for predictive maintenance and the estimation of remaining useful life (RUL). Additionally, the different process sections of creating a machinery prognostic solution are presented. The goal isn't to fully define and formulate each method used as there are numerous different algorithms, functions, and approaches with some introducing exquisitely complex mathematical representations. This literature review gives insight into (i): what kind of industrial components are typically within the scope of predictive maintenance and the type of data extracted from them and (ii): what are the methods used, in the form of feature selection and models. Of the reviewed methods, the most suitable solutions are selected, explained and deployed in the empirical section of this thesis for Valmet's breast roll shaker. The suitability is measured as models' accuracy, performance, and transparency. The scope of this literature review and this thesis is to focus on developing a successful predictive maintenance model, thus this section of the literature review mainly targets the different RUL prediction approaches. Useful life of an equipment is the time when the equipment is operational. RUL is defined as "the length from the current time to the end of the useful life" (Si et al. 2011).

A predictive maintenance program can be divided into four technical processes: Data acquisition, health indicator construction, health stage division and finally prediction of remaining useful life. In the core of predictive maintenance is data acquisition. Information from the underlying process is captured with sensors such as accelerometers, thermometers, barometers, flowmeters, etc. The data is then transmitted into storage for further prognostic analysis. Signals of the degree of damage are not often measured directly, but the signals measured usually contain this information indirectly such as vibration or pressure signals, which are common measurements that are monitored. Historical maintenance data is crucial information, especially if there are expected lifetime estimates and maintenance intervals for the component. These signals can contain a lot of information about the health condition of the machine. Health indicators are constructed using different signal processing techniques such as, simple moving averages, moving standard deviations, or complex artificial intelligence techniques. After the health indicators are constructed, different health stages can be

recognized. These stages present the degree of degradation typically in two more different stages and can be as simple as: “healthy stage” and “unhealthy stage”. Example of a two-stage division is shown in figure 6 and a possible three-stage division in figure 7. Finally, after the health stages are defined and failure thresholds are specified, the remaining useful life can be predicted by analyzing the health indicators. (Lei et al. 2018)

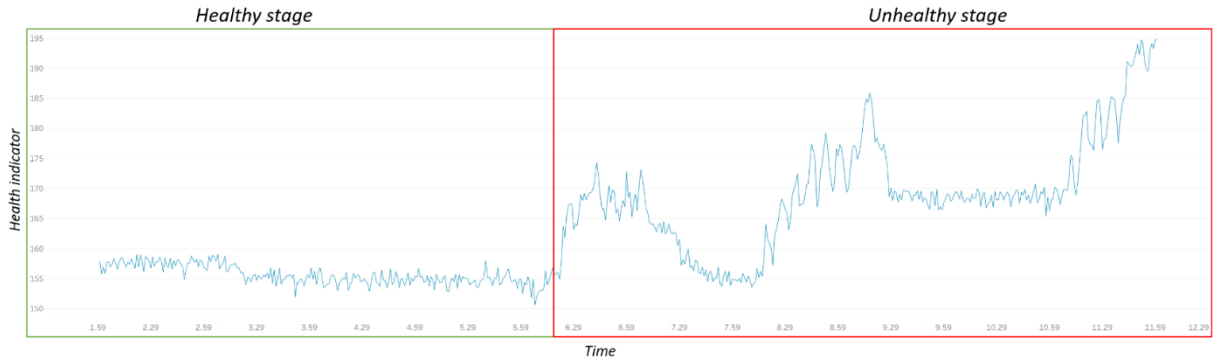


Figure 6. Two-stage health division. The y-axis presents the health indicator, the x-axis is time.

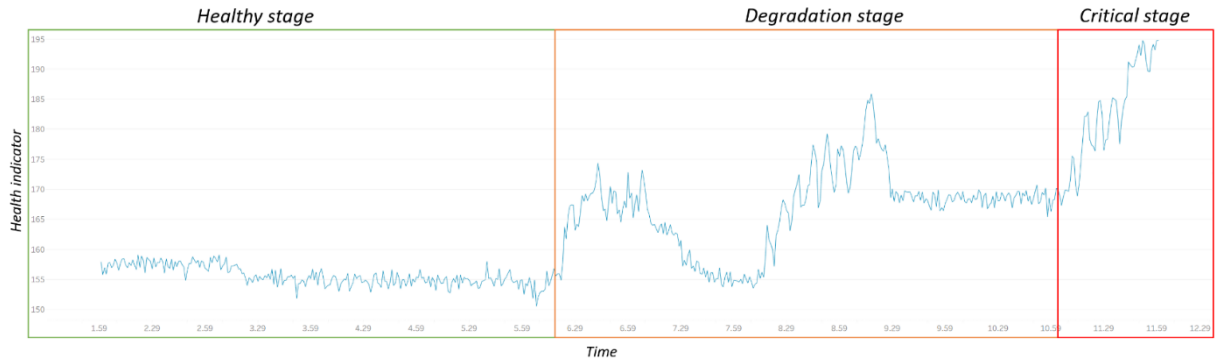


Figure 7. Three-stage health division. The y-axis presents the health indicator, the x-axis is time.

As seen from the example figures 6 and 7 the same health indicator can be divided into multiple different stages. The squared green area presents a normal healthy signal. An increase to the signal may indicate degradation and be classified as unhealthy. The division should be made based on the severity of the fault or degradation. At the end of the critical stage or the unhealthy stage a failure threshold can be specified; the value after which a failure occurs. Remaining useful life prediction can be made by analyzing the health indicator trend and predicting future values, as shown in figure 8.

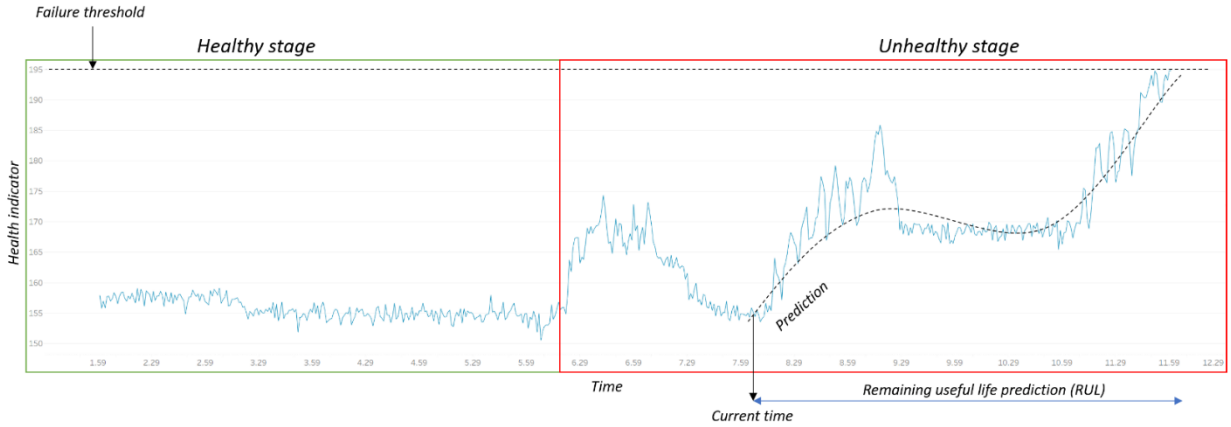


Figure 8. Example of a two-stage health division with failure threshold and remaining useful life prediction. The y-axis presents the health indicator, the x-axis is time.

Figure 8 shows the basic principle of remaining useful life prediction. The example shows the estimation from the current time to the end of useful life as a dotted line – the prediction. However, health indicators cannot be always constructed as clearly as in this example. The underlying degradation of the machine can be a combination of multiple signals with linear or non-linear interdependency.

The prediction of remaining useful life and predictive maintenance has received more attention over the past years. After reviewing numerous papers such as the works of Jardine et al. (2006), Heng et al. (2009), and Kan et al. (2015) RUL approaches are classified ambiguously into various categories, i.e., knowledge-based models, physical models, and heuristic models. Naming conventions are not clear, which may lead to confusion when classifying different approaches. For example, the knowledge-based models and physical models are both based on deep knowledge of the process and built based on a complete understanding of the failure mechanisms. After an exhaustive literature review of predictive maintenance approaches and prediction of remaining useful life approaches, these can be divided into four different categories. The recent work of Lei et al. (2018) covers more than 250 scientific publications related to RUL approaches and classifies the approaches to the following categories; physics model-based approaches, statistical model-based approaches, AI approaches, and hybrid approaches. The categories are shown in table 1, including the percentage and number of publications related to the various categories.

Table 1. Different categories of RUL approach publications. (Lei et al. 2018)

RUL prediction approach	***Percentage and number of publications related to the approach***
Physics model-based approaches	*10 % / 28*
Statistical model-based approaches	*56 % / 144*
AI approaches	*26 % / 81*
Hybrid approaches	*8 % / 21*

The next chapters answer what are the differences between the approaches described in table 1, and what kind of applications or use cases can be found for each approach. The goal is to have a good overview of the different methods of use cases, instead of a detailed and thorough explanation of how the different models work.

3.1 Physics model-based approaches

Ten percent of the publications related to RUL approaches are physics model-based approaches. These first principle approaches try to describe the degradation of machinery or processes based on physical and mathematical models of the failure mechanisms, the first principles of damage. For example, the parameters of physics models can be related to material properties, stress levels, or other laws of physics. The identification of stress levels and material degradation can be done by using finite element analysis, specific experiments or other suitable techniques. These models require a complete understanding of the machinery and processes and provide accurate estimates of the remaining useful life or condition of the equipment. However, it is often difficult to understand and model complex mechanical systems and processes to get a good estimation of model parameters. (Lei et al. 2018) One of the most widely used physics models is the Paris-Erdogan (PE) model that is used to describe the crack growth presented in the works of Paris & Erdogan (1963). For example, Li et al. (1999) used a variation of the PE model for adaptive prognostics for rolling element bearing condition.

A good example of a physics model-based approach is the prediction of cavitation made by Jacobs (1961), where an extensive model is created to detect cavitation of a centrifugal pump. Cavitation occurs when the local static pressure drops below the local vapor pressure and fluid vaporizes. These small vapor bubbles collapse causing noise and damage to the equipment. The model is based on fluid mechanics and examines the effect of cavitation on different performance characteristics of a pump such as the head-capacity and calibration curves. The work of Jacobs (1961) is not directly applicable to the case of this thesis, but clearly states that cavitation causes changes in the performance characteristics of a pump and thus may be detected in the pressure measurements used in this case. (Jacobs 1961) The work of Samanipour et al. (2017) examines the potential detection of cavitation in centrifugal pumps, achieving an 88 % accuracy for cavitation detection. The predictive model for cavitation is based on the pressure and torque measurements of the centrifugal pump by using a self-organizing map (SOM) approach to classify healthy and unhealthy states

3.2 Statistical model-based approaches

As shown in table 1. the most common approach to RUL prediction is a *statistical model-based approach* covering 56 % of the publications. These statistical models are based on empirical knowledge and typically calculate conditional probability distribution function (PDF) to present the RUL without relying on knowledge about the nature or physics of the machine or processes. The available observations are fitted into stochastic process models. Typical models in this category are Autoregressive (AR) models, random coefficient models, Wiener process models, Markov models, and Proportional Hazards (PH) models. These models are commonly utilized to predict the RUL of machinery such as bearings, conveyor belt system, or aluminum plates fatigue cracks (Tang & Su 2008, Gebraeel et al. 2009, Caesarendra et al. 2011).

Autoregressive models are built by assuming that the state in the future is a linear function of past observations and random errors (Sikorska et al. 2011). AR and linear models are often used because of their simplicity. Enhanced versions, in particular, such as Autoregressive Moving Average (ARMA) and Vector Autoregression (VAR) models are widely utilized in time series forecasting and therefore in RUL prediction of machinery such as bearing degradation

prognostics (Caesarendra et al. 2011). However, the dependency on the trend information of historical observations is a major disadvantage as it is not common for the historical observations to have trending components.

Random coefficient model assume the degradation models to be normally distributed and add a random coefficient into the models to describe the stochasticity. These coefficients are assumed to follow a Gaussian distribution, and this may restrict the applications in which these models are used. These models are suited to predict the remaining useful life probability distribution, such as done in the work of Jin et al. (2016) for prognostics of bearing failure.

Wiener process models and Markov models are very commonly used stochastic process models. Both are based on the assumption of Markov property, meaning that the future state is independent of past behavior and only dependent on the current state of the process. RUL prediction applications for both Wiener and Markov process can be found for different kinds of machinery, such as the degradation of cylinder liners (Giorgio et al., 2011). However, the inconsistency of the Markov property leads to inconsistency in real life applications.

Proportional hazards models are a class of survival models in statistics, relating to the 'survival' time that passes before failure while associating parameters with that quantity of time. PH models are composed of two multiplicative factors: the baseline hazard function and covariate function. Simply stated, the hazard at any given time t is the baseline hazard rate and a multiplicative factor through the covariate function, integrating both the event data and related monitoring data. Vlok et al. (2002) created optimal component replacement decisions with the PH model and Banjevic et al. (2001) created a condition monitoring solution for machinery to predict the RUL.

3.3 Artificial intelligence approaches

Statistical models and physics models are often not capable of dealing with highly complex mechanical systems and processes. Complex nonlinear relationships between multiple different systems are usually not known and artificial intelligence approaches are suited to deal with

these kinds of problems and therefore are recently attracting more attention. There are numerous AI techniques used in the field of predictive maintenance such as Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Support Vector Regression (SVR), K-Nearest Neighbor (KNN), and different kinds of decision tree (DT) models. Some of the models lack transparency and are therefore called "black box" models, meaning that the reasons for achieving the results are not always identified and the root causes of faults are therefore not recognized. These models are typically called machine learning models and can be divided into supervised and unsupervised models. The focus in this thesis is on the supervised models, which can be further split into two rough purpose-based categories: classification and regression. The purpose of classification is to determine which particular discrete class a data point belongs to, while regression is used to predict continuous values. Typically, in a machine learning function estimation problem the system consists of response variables y, and a set of explanatory variables x. By using the training sample or "historical data" of known (y, x) values the goal is to estimate the function $\hat{y}(x)$, mapping x to y, while minimizing a specified loss function. In unsupervised learning, prior knowledge of the known y is not available and the goal is to infer possible structures present in the data. (Friedman 2001, Lei et al. 2018)

Artificial Neural Networks are the most commonly used in the AI field for machinery RUL prediction. There especially Artificial Neural Networks, such as feed-forward neural networks (FFNN), recurrent neural networks (RNN), and convolutional neural networks (CNN) are frequently used. These techniques are inspired by the biological neural networks where many connected nodes and edges are in a complex layered structure. These layers are used to map specific inputs to specific outputs. Hu et al. (2015) used a dataset with both known failures and suspension data to predict the RUL of both rolling-element bearings and an electric cooling fan with the help of both feedforward neural networks. Mahamad et al. (2010) used the same kind of approach - feedforward neural networks to predict the RUL of bearings and bearing failure. Modern work of Li et al. (2018) used deep convolutional neural networks to estimate the RUL of turbofan engines. Also, multiple papers (Liu et al. 2014, Peng et al. 2013, Hu et al. 2012) can be found utilizing recurrent neural networks for different RUL prediction purposes such as health indication, lifetime prediction, and machinery prognostics. ANNs can learn complex non-linear relationships in the data, but often have low transparency because of their complex layered structure. In addition to the low transparency, ANNs usually require a large volume of

good quality training data and may have problems with generalization in the case of unseen data.

Support Vector Machine (SVM), Support Vector Regression (SVR), or Support Vector Classification (SVC) can be used for classifying both regression or data, respectively. The principle is in finding a linear hyperplane that separates the data with minimal error. With non-linear high dimensional data, a kernel function can be used to reduce the dimensionality. As with the previous methods, multiple papers and research can be found using SVC or SVR in the field of machinery prognostics. Some good examples can be found from the works of Kimotho et al. (2013) where support vector machine is used for machinery prognostics or Sloukia et al. (2013) where SVM is used for bearing prognostics. Support vector models can be trained to good accuracy with relatively small sample sizes when compared to other artificial intelligence approaches. However, the limitations of these models are the lack of probabilistic predictions and performance is highly dependent on the selected kernel function.

Different decision tree and KNN models are best suited for classification purposes such as classifying different stages of degradation. Decision trees map the input data with labeled output data with a tree-like model where each branch from root to leaf represent the classification path. Non-linear problems can be solved with boosting and bagging the decision trees, meaning that the input data is upsampled and multiple different trees are created to solve the input-output task, respectively. This approach is transparent and thus called a “white box” model, where the root cause for failure and feature importance can be inspected by following the branches of the tree. Enhanced decision tree models are widely used as they allow a methodology similar in structure to deep learning, but with a white box approach for thorough post-analysis (Friedman 2001). Usually the best performing solution in data analytic competitions, such as the famous Kaggle competitions, is a combination of different decision tree models (Kaggle).

3.4 Hybrid approaches

The methods mentioned in the previous section have their own benefits and weaknesses, and by combining some of the approaches the advantages of each individual model can be

integrated. These methods are the least common with the fewest number of publications. In hybrid approaches usually “black box” methods are used to estimate parameters of another model such as probability distributions or parameters of autoregressive models to estimate the likelihood of failure. Also, a combination of different approaches can be used as a voting system to systematically combine the predictions of multiple methods and use the most effective combination of models. Interesting studies can be found such as Martinsson (2016), where RNNs are used to estimate the parameters of Weibulls time to event probability distribution with good accuracy in jet-engine degradation prognostics. Liao & Kottig (2014) write a thorough review of hybrid prognostics approaches and the paper proposes a model to predict battery degradation and failure.

3.5 Selected methods and model evaluation

In the case of developing a successful data analytics solution, multiple different models and approaches are tested to get the best performance. Therefore, after a review of different machinery prognostic methods, and taking into consideration the nature of the current issue, the following models are selected for further inspection: linear regression, autoregressive moving average (ARMA), vector autoregression (VAR), decision tree models, and simple physics-based models for cavitation. Each of these models has its own benefits and weaknesses, which are identified in this section. To stay within the scope of this thesis the basic functionality is clarified in a way that the reader understands how the models work and the models can be further utilized in the empirical section of this thesis.

Linear regression is chosen for its simplicity and an easy way to test a proof-of-concept with possibly a good fit. The model is computationally effective to create and the deployment to the architecture can be tested and finally, how the visualization of the results looks like to the end customer. Simple linear regression is used to model the relationship between a scalar response and one or more explanatory features. This is practically useful in this case for the purpose when a measurement is constantly moving closer to the limits of an alarm or fault, a simple linear prediction can detect this behavior. The linear regression model can be expressed as:

$$\hat{\gamma} = f(x_1, x_2 \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots, + \beta_k x_k \quad (2)$$

Where $\hat{\gamma}$ denotes the response variable, β_0, $\beta_1 \ldots$, β_k are parameters, and x_1, $x_2 \ldots$, x_k the explanatory features. With a single explanatory feature, the *k* equals 1. The goal is to find values for β, so the model fits the underlying data with minimal error. The error, or residual, can be calculated with the distance between the response variable $\hat{\gamma}$ and actual measured values. (George et al. 2003) Thus, linear least squares find the parameters that minimizes the sum function:

$$S = \sum_{i=1}^{n} r_i^2 \quad (3)$$

Where S is the sum of residuals, r_i are the residuals, and n is the number of observations. The residuals are calculated as difference between the actual value and the value predicted by the model. (George et al. 2003)

Autoregressive moving average and vector autoregression models are both widely used in time series forecasting. These models are selected for the same purpose as the linear regression - to make predictions of the possible alarms or faults. Difference to the linear regression is that these models use the lagged (or past) values of the explanatory features, in the case of univariate ARMA model - their own lagged values. The ARMA model consists of two polynomial terms, the autoregression term and moving average term. (Hamilton 1994) The autoregressive model is expressed as:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \quad (4)$$

Where c is a constant term, p denotes the model order, $\varphi_i \ldots$, φ_p are parameters and ε_t is white noise. Autoregressive models requires the mean and variance of the process to be stationary (Hamilton 1994). The moving average model can be expressed as:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \quad (5)$$

Where μ is the expected value of X_t, q denotes the model order, θ_i ..., θ_p are the parameters of the model and ε_{t-i} values are white noise, generally assumed to be independent identically distributed random variables sampled from Gaussian distribution $\varepsilon \sim N(0, \sigma^2)$ (Hamilton 1994). Thus the ARMA model with the order of q and p can be expressed as the combination of the equations (4) and (5) above:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \quad (6)$$

Finding the optimal model order (p, q) is facilitated by plotting the autocorrelation or partial autocorrelation functions for different estimates of p and q (Hamilton 1994). However, modern programming tools allow for parallel automatic testing of multiple different models and thus it is not necessary to go into detail on the specifics how to find the optimal model. The quality of different models is tested relative to each other and the best model is selected based on the goodness of fit and the models' simplicity using the Akaike Information Criterion (AIC). Additionally, stationarity tests are automatically performed when building different models. If the observed values are dependent on their own lagged values, the ARMA model can make potentially accurate predictions for future values.

Vector autoregression is like the autoregression model defined above, but extended to capture the linear dependencies among multiple variables and their own lagged values. However, without the moving average term. This model is tested for the possible interdependency for example, with oil pressure and temperature. Say, for example, lagged values of oil temperature measured in the system may explain the current oil pressure value. General matrix notation for vector autoregression with *k*-number of variables:

$$y_t = c + A_1 y_{t-1} + \cdots, + A_p y_{t-p} + \varepsilon_t \quad (7)$$

where y_t is the k-length vector response value for each variable, c is a vector of constants, $A_1, A_2 \ldots, A_p$ are a size $k * k$ matrix, and ε_t is a k-vector error term. The order of integration (stationarity) for the variables should be equal or else the variables are cointegrated and the error term is included, or the variables are differenced (Hamilton 1994). As with the ARMA model, modern programming tools can be used to automatically find optimal parameters and stationarity tests for the VAR model.

The decision tree model is selected as an artificial intelligence approach. This approach can be easily automated without the need for thorough feature selection and feature transformation processes. Additionally, decision trees allow post-analysis of the important features by estimating the number of times a feature is used to make key decisions with decision trees. The decision tree variant selected is the implementation of gradient boosted trees originally introduced by Friedman (2001). Going through the highly detailed mathematical representation of gradient boosted trees could be the subject of a single thesis, but again to stay within the scope of this thesis, the structure is explained more ambiguously. To understand the principle of gradient boosting the reader is assumed to understand one of the simplest and widely used numerical minimization methods "steepest-descent" also known as line search.

The general principle of decision trees is to numerically optimize an objective function, instead of parameters. By following the work of Friedman (2001) the general form can be expressed as

$$F^*(x) = \sum_{m=0}^{M} f_m(x) \tag{8}$$

Where M denotes the number of trees and f is a function in a functional space, $f_0(x)$ is an initial guess. f_m is defined by the optimization method of steepest descent:

$$f_m(x) = -\rho_m g_m(x) \tag{9}$$

Where

$$g_m(x) = \left[\frac{\partial \emptyset\big(F(x)\big)}{\partial F(X)}\right]_{F(X)=F_{m-1}(x)} = \left[\frac{\partial E_y[(L(y,F(x) \parallel x]}{\partial F(X)}\right]_{F(X)=F_{m-1}(x)} \quad (10)$$

and

$$F_{m-1}(x) = \sum_{i=0}^{m-1} f_i(x) \quad (11)$$

With the assumption of interchange of differentiation and integration, this becomes

$$g_m(x) = E_y\left[\frac{\partial L\big(y,F(X)\big)}{\partial F(X)} \parallel x\right]_{F(X)=F_{m-1}(x)} \quad (12)$$

and the line search gives the multiplier to equation (9)

$$\rho_m = \arg\min_{\rho} E_{y,x}\, L\big(y, F_{m-1}(x) - \rho g_m(x)\big) \quad (13)$$

However, with finite data, the approach breaks down, and one cannot approximate values at x outside the training data. To overcome this, a parametrized form needs to be obtained and optimize the parametrized equation such as

$$\{\beta_m, a_m\}_1^M = \arg\min_{\{\beta'_m, a'_m\}_1^M} \sum_{i=1}^{N} L\left(y_i, \sum_{m=1}^{M} \beta'_m h(x_i; a'_m)\right) \quad (14)$$

The loss function L includes the generic function $(x_i; a_m)$, a simple parametrized function of input variables x characterized by parameters $a = \{a_1, a_2 \ldots\}$ and β_m is a finite set of parameters whose joint values identify individual class members. Of interest here is that each of the functions $(x_i; a_m)$ is a small regression tree. A "greedy-stagewise" approach can be used here for $m = 1,2 \ldots, M$

$$\{\beta_m, a_m\} = arg\min_{\beta,a} \sum_{i=1}^{N} L\big(y_i, F_{m-1}(x_i) + \beta h(x_i; a)\big) \tag{15}$$

and

$$F_m(x) = F_{m-1}(x) + \beta_m h(x, a_m) \tag{16}$$

where the $L(y_i, F(x))$ is the squared-error loss and $h(x_i, a)$ are the "basis functions" also known as "weak learners" or "base learners" - a classification tree. Now an unconstrained negative gradient gives the best steepest-descent step direction

$$-g_m(x_i) = -\left[\frac{\partial L\big(y_i, F(x_i)\big)}{\partial F(x_i)}\right]_{F(X)=F_{m-1}(x)} \tag{17}$$

To allow generalization to other x-values the most highly correlated $h(x; a)$ with $-g_m(x)$ over the data distribution can be obtained from the solution

$$a_m = arg\min_{a,\beta} \sum_{i=1}^{N} [-g_m(x_i) - \beta h(x_i; a)]^2 \tag{18}$$

Now the unconstrained negative gradient is constrained, and the line search (13) is performed

$$\rho_m = arg\min_{\rho} \sum_{i=1}^{N} L\big(y_i, F_{m-1}(x_i) + \rho h(x_i; a_m)\big) \tag{19}$$

And the approximation is updated

$$F_m(x) = F_{m-1}(x) + \rho_m h(x, a_m) \tag{20}$$

Now the generic gradient boosting algorithm using steepest-descent can be expressed as the following pseudo-code

Algorithm: Gradient boost

1. $F_0(x) = arg\min_{\rho} \sum_{i=1}^{N} L(y_i, \rho)$
2. $For\ m = 1\ to\ M\ do$:
3. $\hat{\gamma}_i = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(X)=F_{m-1}(x)}, i = 1, N$
4. $a_m = arg\min_{a,\beta} \sum_{i=1}^{N} [\hat{\gamma}_i - \beta h(x_i; a)]^2$
5. $\rho_m = arg\min_{\rho} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \rho h(x_i; a_m))$
6. $F_m(x) = F_{m-1}(x) + \rho_m h(x, a_m)$
7. $end\ For$
8. $end\ Algorithm$

Figure 9. Pseudo-code for gradient boosting algorithm using steepest-descent. (Friedman 2001, 5)

The gradient boost algorithm (figure 9) structure changes depending on the type of loss criteria, which in turn are based on the type of problem, regression or classification. Possible loss criteria are, for example, least squares, least-absolute-deviation, Huber, or logistic binomial log-likelihood. (Friedman 2001) In machine learning models fitting the data perfectly is not the optimal solution. A model with an almost perfect fit in the training set is counterproductive since when going beyond the training data the expected loss increases, the model is overfitted. This can be prevented to some degree by using an independent test set or cross-validation when selecting the best model. Cross-validation partitions the data into training and testing sets and model validation is performed on multiple different subsets to reduce variability. Additionally, regularization can be performed by introducing shrinkage to the model. For example, to the generic gradient boosting algorithm introduced previously (equation 21), the shrinkage can be expressed as simply as

$$F_m(x) = F_{m-1}(x) + v \cdot \rho_m h(x, a_m), \qquad 0 < v \leq 1 \tag{21}$$

Now the updates are now scaled by the "learning rate" parameter v. The regularization can be adjusted using two parameters, the learning rate v and the number of trees M. The best combination of parameters can be found using grid-search or random-search methods. In the grid-search method, a parameter grid of suitable range is defined, and models are trained exhaustively with every parameter combination with cross-validation – the best model is selected out of all possible combinations. Random-search is a similar approach, but parameters are tested randomly instead of using brute force and going through all possible combinations to reduce computational time. (Barga et al. 2015)

This gradient boosting algorithm is further improved by Friedman by using a "bagging" procedure introduced by Breiman (1996). There is a lot more to the model than meets the eye here. The programming community and data analysts have refined the model and multiple different versions are available for use. A novel algorithm created by Chen & Guestrin (2016), an Extreme Gradient Boosting (XGBoost) is used in the empirical section, this scalable tree boosting system originates from the gradient boosting concept. As defined in the beginning of section 3, the classifier can be used, for example, to classify different stages of degradation of the breast roll shaker.

Multiple metrics are available for models' performance evaluation. Common metrics for the time series prediction models of continuous future values are a root-mean-squared error (RMSE) and R-squared (R^2). For classification tasks possible evaluation metrics are, for example, Area Under Curve (AUC), where the curve is the Receiver Operating Curve (ROC) and different metrics calculated from a Confusion Matrix. (Barga et al. 2015)

RMSE is simply the square root of the average of squared errors. For example, the linear regression equation (2) introduced previously, the model's root mean squared error can be calculated with the equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{22}$$

Where $\hat{y}$ is the predicted response variable, y denotes the actual response value, and n is the number of observations. (Devore 2011)

R-squared is used to explain the proportion of the variance – how well the actual response variables are replicated with the model. The R^2 can be calculated with the equation:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{23}$$

Where SS_{res} is the sum of squares of residuals and SS_{tot} is the total sum of squares, when the actual observed values are subtracted from the mean of the observed data. (Devore 2011)

For models that are used to predict or estimate discrete classes instead of continuous values, confusion matrix and AUC can be used to evaluate the models' performance. The confusion matrix is a representation of the results in terms of true and false values of both the predicted class and the actual class. Table 2 represents a confusion matrix, where possible classes are faulty or healthy, as an example.

Table 2. Confusion matrix, columns are actual classes and rows show predicted classes.

		Actual class	
		Healthy	*Unhealthy*
Predicted class	*Healthy*	True positive (tp)	False positive (fp)
	Unhealthy	False negative (fn)	True negative (tn)

Different performance metrics from the confusion matrix (table 2) can be calculated with equations 24-27:

$$Precision = \frac{tp}{tp + fp} \tag{24}$$

$$Recall = \frac{tp}{tp + fn} \tag{25}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{26}$$

$$False\ positive\ rate = \frac{fp}{tn + fp} \tag{27}$$

Accuracy measures the ability of the classifier to correctly label each data point in the separate testing set. Precision measures the ability of the classifier to not label a negative sample as positive. Therefore, this aims to avoid false positives. Recall measures the ability of the classifier to find all the positive samples. In this case, precision and recall are more important than accuracy, as it is more important to reduce the false positives in a system than to accurately predict all samples. Precision is also more important than recall, as although it is important to correctly identify the true positives and true negatives, it is better to have false negatives than false positives. The Receiver Operating Curve (ROC) can be created by plotting the true positive rate against false positive rate. True positive rate equals to precision and is calculated with equation 24 and false positive rate is defined as in equation 27. The area under the ROC curve (AUC) of 1 means that the model classifies all classes correctly – higher the AUC, better the model. (Barga et al. 2015)

4 VALMET INDUSTRIAL INTERNET (VII) ARCHITECTURE

This chapter introduces Valmet's industrial internet architecture. Valmet offers industrial internet solutions to its various customer segments such as energy, pulp, paper, tissue, and board. The foundation of Valmet's industrial internet is know-how in automation, process technology, information technology platform, applications, and services. Valmet operates globally and thus there are multiple different components such as data sources, data transfer methods and databases in the VII-architecture. The focus is specifically on the services and applications that are utilized in this thesis to understand how the data is transferred, processed, analyzed and finally presented visually to the end customer.

It is very common that data analytics, data warehousing and business intelligence projects take a long time to deliver tangible results. Valmet industrial internet (VII) tools, services, and applications allow developers to provide quick results and proofs-of-concept. Amazon Web Services (AWS) are widely utilized and its different service models such as *function-*, *platform-*, *software-*, and *infrastructure as a service solution* (*FaaS, PaaS, SaaS, and IaaS*) are presented.

4.1 AWS S3, DynamoDB, Lambda, EC2, CloudWatch, and SNS

AWS S3 and DynamoDB are databases, that are used to store all incoming data to the VII-platform. In S3 the data is stored in buckets as objects with the specified naming convention. S3 is good for storing large amounts of data cost-efficiently (Amazon Web Services a). DynamoDB is a NoSQL database containing configuration information for further processing such as, what files are read from S3 and where the files are saved. AWS Lambda is a serverless computing service, which can be used, for example, for data transfer or transforming the data into a different format. The user is billed by the computational time and Lambdas support common programming languages such as Python and Java. Additionally, when a Lambda using python runtime is deployed, the size is limited to 250 MB. This should include all necessary

packages the python code requires to run, for example, pandas, numpy and scikit-learn (Amazon Web Services b). Amazon Elastic Compute Cloud (EC2) provides scalable computing capacity as virtual computing environments, known as instances. All instances can have specific configurations of CPU, storage, memory, and networking capacity. AWS CloudWatch is used to monitor the services run on AWS and provides important metrics and logs such as CPU utilization and memory (Amazon Web Services c). AWS SNS (Simple Notification Service) is used to send notifications from the cloud to users, devices, and applications. SNS can be attached to various AWS services such as S3 and furthermore Lambdas can be subscribed to the SNS to receive information when new data arrives at S3 (Amazon Web Services d).

4.2 Data transfer, storage and visualization

Depending on the control system configuration, breast roll shaker data can be extracted either from the mill's machine control system (MCS) or through a separate logic device called eWON. The data source defines the pipeline architecture via which the data is transferred into the VII-platform. Most of the breast roll shakers that are connected to the Valmet cloud use the eWON gateway. Both gateways utilize different AWS services extensively. The data pipelines are shown in figures 10 and 11.

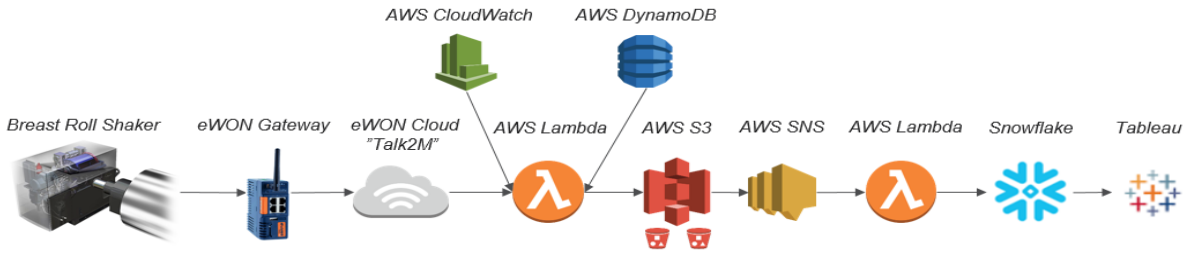


Figure 10. A simplified version of the "eWON" data pipeline solution for the breast roll shaker

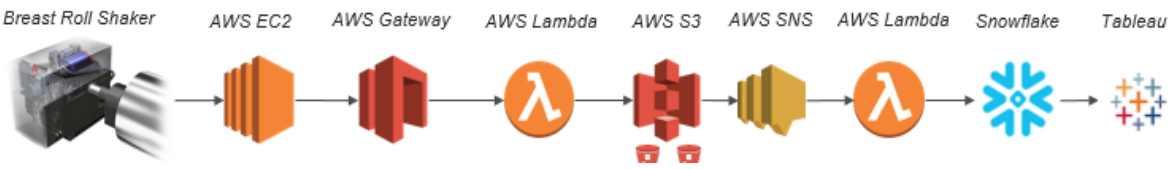


Figure 11. A simplified version of the MCS data pipeline integrated into the existing system.

In figures 10 and 11, the arrows show the flow of data. In both figures, the measurement data is transferred to the same databases in both AWS S3 and Snowflake. AWS Lambda serverless computing services are used to transfer and transform the data between the different components. Lambdas are subscribed to the AWS SNS to initialize data transfer each time notification of new data arrives. Data from the site (or the Breast Roll Shaker) is first transferred using SFTP (Secure File Transfer Protocol). Figure 10 is the eWON pipeline, where data is transferred to eWONs own cloud database "Talk2M". The data is crawled from the eWON cloud with AWS Lambda to the AWS S3 data storage. The devices are listed in the DynamoDB database to upkeep what devices are crawled from the cloud. In figure 11 the data is transferred via Amazon EC2 and Elastic File System (EFS) to the same AWS S3 data storage as in the figure 10 pipeline. Measurement data from both pipelines can be visualized using a business intelligence tool Tableau, which is connected to the Snowflake cloud database.

4.2.1 Snowflake (SF)

Storage and transfer of data is a crucial part of a functional predictive maintenance system. To get an understanding of the procedure, Snowflake and its components are introduced briefly. Snowflake is a high-performance cloud-based compatible and complete SQL data warehouse, that provides flexibility and scalability. SF uses clusters of computer resources i.e. CPU, memory, and temporary storage, to perform different database-oriented tasks in the warehouse and separates storage from the computations. This cloud-based enterprise data warehouse (EDW) delivers an easy to use web interface that can be used to perform the SQL task such as the creation of tables and views, select statements, updating tables, loading data into tables, unloading data from tables, and monitoring of queries. The data can be semi-structured, such as JSON and XML, and loaded directly into a traditional relational table and queried with SQL. Performing operations consumes Snowflake credits, which are billed on a per-second basis based on the number of virtual warehouses in use, how long they run, and their size. Data in Snowflake is stored in tables, they consist of *facts* and *dimensions*, where the facts contain information of particular business process (measurements) and dimension tables contain descriptive attributes, such as details about each instance of a measurement in a fact table, dimensions of a fact (Snowflake Manual).

Valmet uses a database structure as visualized in figure 12. The three-layered structure consists of a staging layer, enterprise data warehouse or dimensional layer, and an application specific layer.

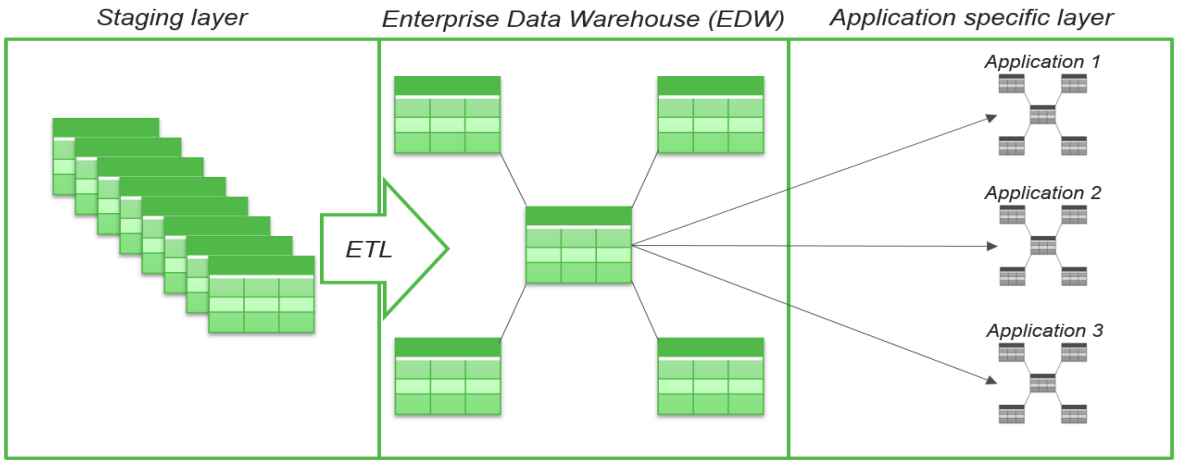


Figure 12. Three-layered structure for Valmet's data processing and storage.

The staging layer is common for all applications and the data is loaded forward from there (figure 12). Some of the incoming data to Snowflake require the Extract, Transform, and Load (ETL) process: data is extracted, transformed, and loaded to the appropriate tables in the dimensional layer. *Matillion ETL for Snowflake* is used for this purpose. The EDW dimensional layer has common dimensions for the data. Facts may vary between tables for example, by granularity and type, but the main principle is that similar data is in the same fact table. EDW is a centralized and restricted layer that follows common Valmet rules and guidelines for secure data storage. The application specific layer consists of fact tables or views where only the relevant data is transferred for specific applications. These applications can be anywhere from complex analytics to just simply visualizing the measurement data with a business intelligence (BI) tool. Snowflake achieves the benefits of the data vault model, such as historical tracking, by using persistent staging and slowly changing dimensions in the dimensional model. Persistent staging means that the staging area is not cleaned after loads to EDW layer, thus the historical data is persistent (Snowflake).

The structure of Snowflake allows the combination of measurement data from a Breast Roll Shaker to be connected to data from other sources such as customer relationship management (CRM) software. The combined information can include, for example, the customer's site, company and mill location. Additionally, Snowflake provides an interface for developing Python applications by using JDBC or ODBC drivers. This enables all standard operations to be performed via Python and loading data from the cloud to a local hard drive or memory.

4.2.2 Tableau

Tableau is a business intelligence tool for easy data visualization. Tableau is designed for fast analytics, big data and ease of use. The Tableau workbook consists of worksheets, dashboards and stories. Users can create visualizations such as charts, heat maps, scatter plots and other analyses on worksheets and combine multiple sheets into a single dashboard to coordinate the visualization purposes. Multiple dashboards with navigation can be combined with stories. Each worksheet can connect to a different data source such as Snowflake or S3. Data sources can be extracted, or a live connection can be established to visualize data in real time, too.

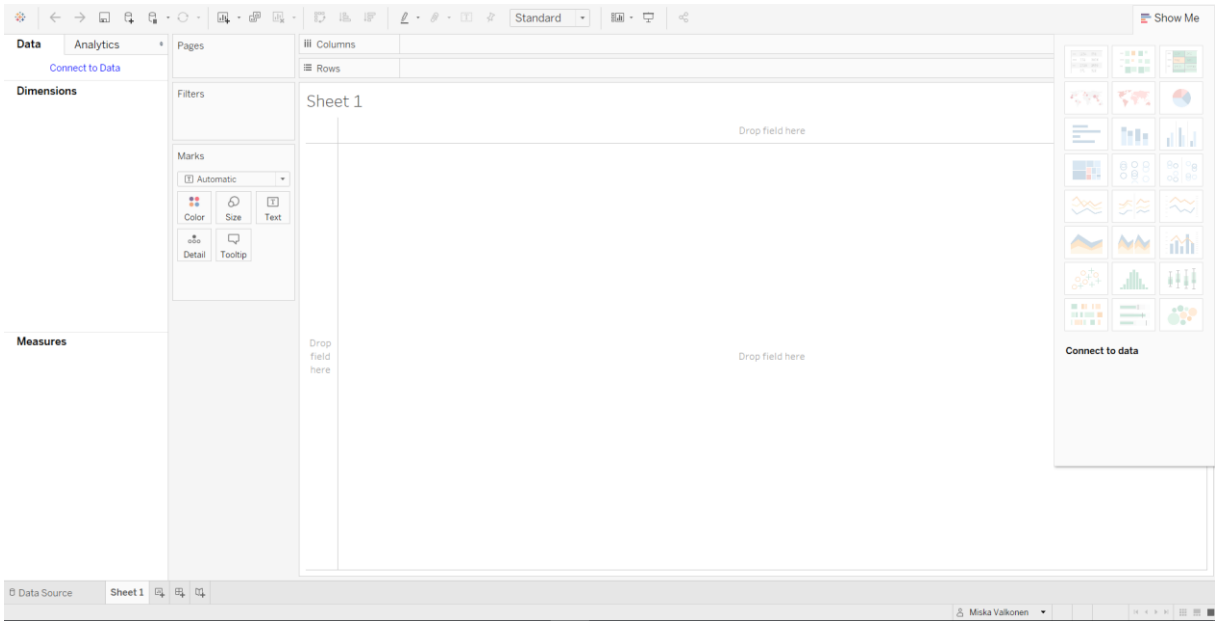


Figure 13. Tableau worksheet.

Figure 13 shows the worksheet where the visualizations are created. Tableau automatically classifies the data sources' data types to numerical measures and categorical dimensions. Based on the visualization purposes the dimensions and measures are placed in the columns or rows in the worksheet. Filters can be also created in the worksheet and the marks can be used to modify the visualization type. The filters can be set, for example, to specific dates, mill, or in this thesis, a specific breast roll shaker. The marks card can convert numeric values to simple visual symbols such as converting zeroes and ones to green and red circles to represent the status of the data, respectively.

4.2.3 Python, Jenkins, and Valmet Bitbucket

Python is used to load data from Snowflake, data processing, analytics, and saving analytical results to Snowflake. Python codes are compiled to AWS Lambda functions for serverless and automatic data analytics. The source code is kept in Valmet Bitbucket (Stash). Jenkins is the CI/CD tool used to build and deploy the source code from stash to AWS automatically after modifications to the source code.

5 DEVELOPMENT OF PREDICTIVE MAINTENANCE MODELS

The development phase includes several stages: data transfer, data preprocessing, model development, and visualization. First, the measurement data is loaded from Snowflake. The required data preprocessing and filtering is performed before model development. Models described in section 3.5 are developed and tested. Successful models are selected and the results uploaded to Snowflake. Finally, the results are visualized in Tableau.

5.1 Data transfer

As defined in section 4.2, all measurement data from the breast roll shakers is stored in Snowflake. Snowflake connector for Python is utilized to load the measurement data to the local device. A Python function is created to establish an ODBC connection to Snowflake. After the connection is established the data can be queried using SQL clauses. Most of the signals from the breast roll shakers are measured with a 1-second resolution and depending on when the data pipeline was established, the available historical data varies up to a maximum of 2 years. Another Python function is created to create the necessary SQL clauses that create the query by specifying the breast roll shaker in question, queried tags, and time period (timestamps). Finally, a data loading function is created which utilizes the previously created functions to load the data from Snowflake for further analytics – upsampling the lower 600-second resolution to a maximum of 1 second by filling the missing measurement times with empty values to get the multiple time series to be in comparable units. The observations are saved in rows and different columns represent different measurements; tags. Additionally, data can be loaded to Snowflake using the same ODBC connection and specifying the SQL clause for data uploading.

All tags are queried from all available breast roll shakers and the data is saved as Comma Separated Value (CSV) text files for further processing. For example, a single breast roll shaker

dataset may consist of 45 million rows (observations) and 44 columns (measurements, tags) with a file size of 1 GB. A snapshot of the data in the text file can be seen in Appendix 4.

5.2 Model development

Models defined in section 3.5 are developed and tested using appropriate evaluation metrics. Python modules are used for data preprocessing and data analytics such as pandas, numpy, and scikit-learn. The development starts with exploratory data analytics to get an overview of the situation. First, physics-based-models are developed to predict pump failures due to cavitation. Additionally, the phase angle calibration frequency is observed to see if it is possible to detect oil leaks in the hydraulic system with increased calibration frequency. The development continues with statistics-based models and time series prediction. Important measurements, that have limits in the system, are analyzed and predictive models are developed to forecast possible alarms and faults. Machine learning models are trained with known historical faults to predict and prevent similar faults in the future.

5.2.1 Physics based-models

As described in section 2.3, known issues with the breast roll shakers are the double gear hydraulic oil pump and couplings. It is suspected that the axle failure is caused by cavitation in the pump unit and faulty lubrication causing damage to the coupling. In the available data, there are no coupling failures present, thus this section focuses on the pump unit. Cavitation might be detected as an increased variation in the pressure measurements after the oil pump unit, from tags PT1-235_me, PT2-235_me, PT3-235_me, PT4-235_me, and PT5-235_me. The hydraulic system layout shows that measurement PT1-235_me is behind an additional valve and measurement PT5-235_me is in the high-pressure hydraulic line, therefore the analysis focuses on measurements PT2-235_me, PT3-235_me, and PT4-235_me (BRS Manual). There are two similar pump failures in the dataset constructed from the breast roll shakers shown in figure x with a dashed red line. A moving standard deviation is calculated for the pressure measurements

and the pump failures are visualized in the same trend. An example of the moving standard deviation calculation results of PT4-235_me measurement is also shown in figure 14.

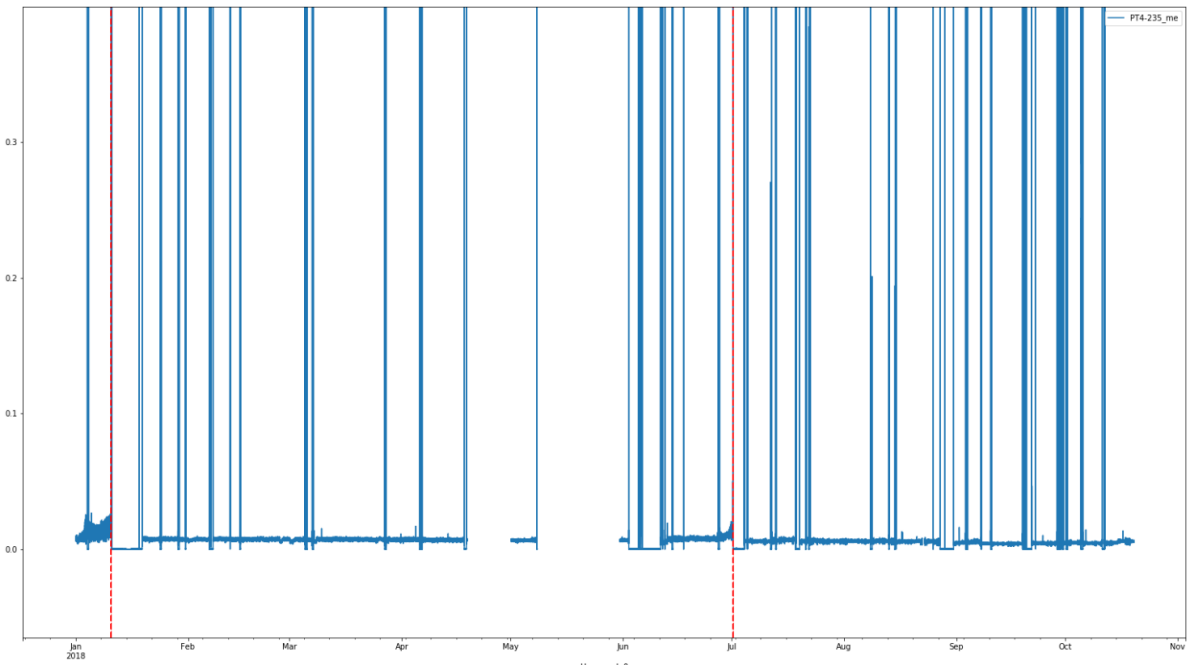


Figure 14. The blue line shows the moving standard deviation of PT4-235_me measurement on the y-axis. The x-axis is time and the vertical dashed red lines show the pump failures.

As seen in figure 14, an increased variation in the pressure measurement can be detected both times before the pump failure. Vertical blue lines are planned shutdowns and startups of the breast roll shaker hydraulic system, this is indicated in measurement M2-232_r, where values one and zero corresponds to “on” and “off” states of the hydraulics respectively. Measurements PT2-235_me and PT3-235_me indicate similar behavior. As expected, the variation is not increased in measurements PT1-235_me or PT5-235_me. Occasional smaller spikes of increased variation can also be seen in figure 14. The spikes are caused by calibration or adjustment of the stroke length, readjusting the phase angle. This calibration is indicated as an increased pressure in measurement PT5-235_me, which is the phase angle control pressure measurement.

To create an indicator that detects increased variation before pump failure, the data needs filtering to prevent false alarms. Variation caused by normal operation needs to be filtered. If the hydraulic system is not on (M2-232_r = 0) or the PT5-235_me measurement indicates an

adjustment in the phase angle, the pressure data is discarded during these timestamps. Additionally, it seems that the pressure variation increases for a while after each startup, phase angle adjustment and sometimes before planned shutdowns. Therefore, the previously mentioned discarding bases are extended. Approximately 5 minutes of each "off" state of the M2-232_r measurement is bilaterally discarded. The same procedure is repeated for the phase angle adjustment filter.

After filtering the data, a trailing moving standard deviation with a 10-minute window is calculated from the remaining data. A mean of the moving standard deviation is calculated to detect when the variation is abnormal. The moving standard deviation is then compared by ratio to the mean of the moving standard deviation. Results of this calculation for the PT4-235_me measurement can be seen in figure 15.

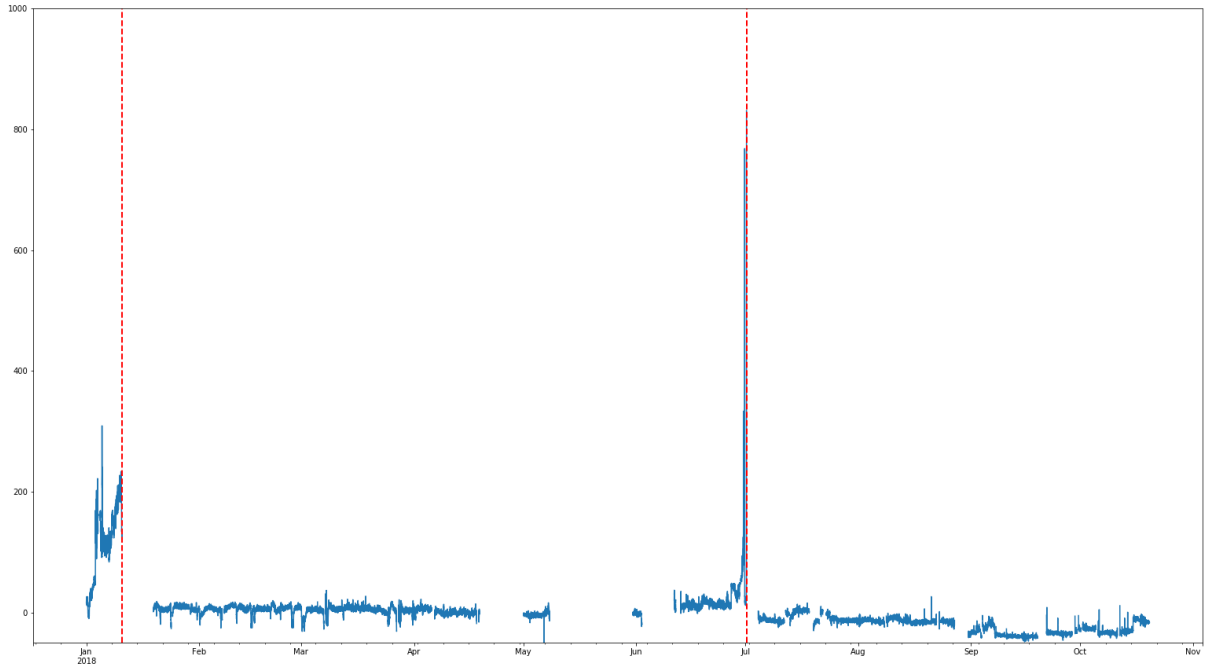


Figure 15. The ratio of moving standard deviation and the mean of the moving standard deviation. Blue is the calculation result and vertical red dashed lines are pump failures.

From figure 15, it can be seen that pump failures could have been detected from the data. To further increase the models' accuracy and reduce false alarms, a threshold is selected. The pressure deviation increased by more than 50 % before the failures when compared to the mean value. Negative values indicate lower pressure deviation when compared to the mean value of

the deviation. To indicate possible cavitation, the calculation is presented after the pressure deviation has been over 50 % for 30 minutes. This model is tested on every breast roll shaker and does not cause false alarms, but it will detect an abnormal pressure variation and possible pump failure in the future.

As described in section 2.3, the phase angle pressure measurement (PT5-235_me) can be analyzed to detect oil leaks. A visual inspection of the measurement is shown in figure 16.

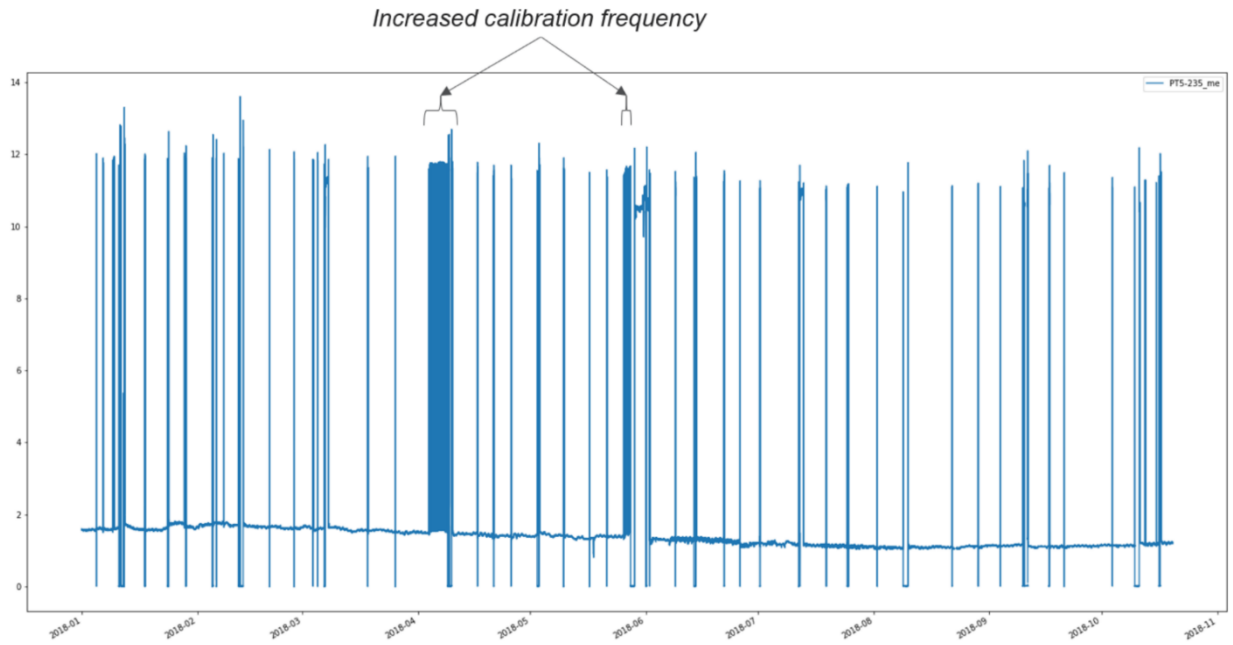


Figure 16. Phase angle control pressure (PT5-235_me) measurement visual inspection.

Figure 16 implies timeframes when the calibration frequency has clearly increased due to oil leaks in the hydraulic system. The control pressure should only increase when the equipment is shutting down, starting up, or the stroke length is adjusted by the operator. The stroke length is automatically adjusted (calibrated) when the difference between the measured stroke length value and the setpoint value is too high. The stroke length setpoint value measurement is collected with tag GY-235_me. All automatically started calibrations can be extracted by inspecting the measurement PT5-235_me at times when the equipment is running and there are no adjustments to the stroke length setpoint value. These are the calibrations due to oil leaks in the hydraulic system. The calibrations are part of normal operation, but when they occur with increased frequency it may indicate a severe oil leakage in the hydraulic system.

To create an indicator that detects increased calibration frequency, the PT5-235_me measurement is filtered based on the condition of the equipment state (M2-232_r) and the setpoint measurement (GY-235_me). The calibrations only account for situations when the equipment is running (M2-232_r = 1), the setpoint value (GY-235_me) is not adjusted and the control pressure has increased momentarily. The calibrations are visualized as the daily running sum to indicate an abnormal amount of calibrations. This visualization can be seen in figure 17.

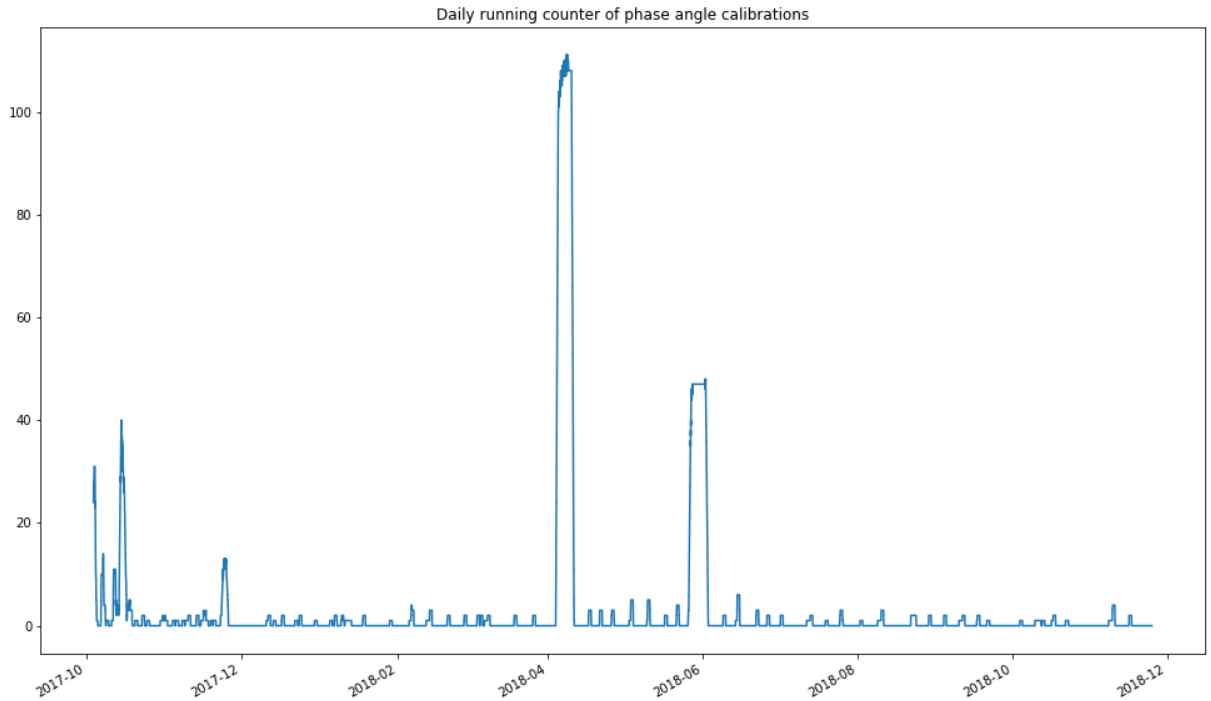


Figure 17. The daily running counter of calibrations. Calibrations that occurred due to equipment shutdown, startup, or operator adjustment are excluded.

The calibration frequency increases momentarily between April and June 2018 (figure 17). This is a good indication to check the hydraulic system for oil leaks. A threshold is calculated by taking the average of the daily running counter of all available data. The current running counter is then compared to the threshold to indicate the condition.

5.2.2 Statistics-based models

The time series nature of the data gives the fundamental idea for statistics-based models. Future predictions are made from the available data based on historical behavior. The main goal is to

predict the future measurement values and, based on the prediction, whether the measurement will cause an alarm or failure of the equipment in the future. The models to be tested are introduced in section 3.5: linear regression, ARMA, and VAR. This section focuses on measurements that have alarm or fault limits in the system, which are TT1-235_me, PT1-235_me, PT2-235_me, PT3-235_me, PT4-235_me, PT5-235_me and PDT1-235_me. The alarm and fault limits for the measurements are shown in table 3.

Table 3. Measurements, units, and limits.

Measurement	***Unit***	***Alarm limit (low)***	***Fault limit (low)***	***Alarm limit (high)***	***Fault limit (high)***
TT1-235_me	C°	31	29	49	55
PT1-235_me	MPa	0.2	0.1	1.4	1.5
PT2-235_me	MPa	0.5	0.3	3.5	4
PT3-235_me	MPa	2.2	2	7	8
PT4-235_me	MPa	0.5	0.3	5.4	5.8
PT5-235_me	MPa	3.5	2.5	14	15
PDT1-235_me	kPa	-	-	450	480

There are a total of 7 different measurements, which are both the response variable and explanatory feature for the predictions. A model is trained for each measurement in table 3, and predictions are calculated based on the historical measurement data. Measurements except PDT1-235_me have both low and high limits, which indicate when an alarm or fault is triggered (table 3). As an example, visual examination is presented for the PT2-235_me measurement in figure 18.

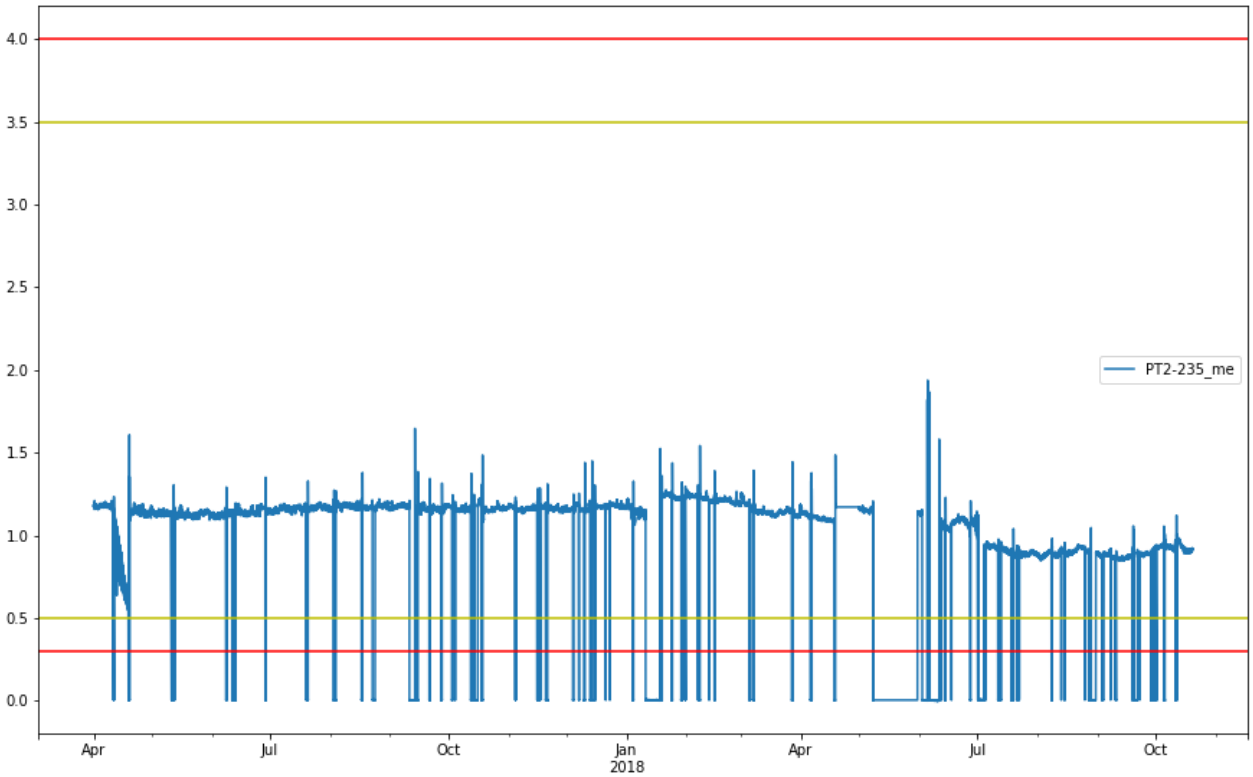


Figure 18. PT2-235_me measurement value with limits. The blue line is the measurement value, the yellow and red lines are the alarm and fault limits, respectively.

Of interest are the situations, when the measurement is trending near the limits, possibly causing an alarm or fault. Unfortunately, these kinds of situation are rare and occurred only once during the historical data: April 2017 (figure 18). Before training the models to predict the series' future values, the data needs to be preprocessed. The shutdowns, seen as spikes in the data, are discarded. The goal is not to fit a model utilizing the whole dataset, but to create a model that would generate real-time predictions over a specified time window. Specifically detecting the event in April 2017, while not generating false alarms during other times. False alarms are counted as a prediction of an alarm or fault outside the event during April 2017 (11.4.2017 – 19.4.2017). The performance, in this case, is assessed by trying to predict as early as possible the event while avoiding false alarms. Therefore, different time windows and prediction horizons are tested, starting with the least squares linear regression. The delay in the data pipeline is approximately two hours, thus the prediction horizon should be at least twice as long as the delay. A parameter grid for the prediction horizon and the historical data used by the linear regression is set up: starting from one hour of historical data and eight-hour prediction horizon all the way up to 6 hours of historical data and 11 days of the prediction horizon. For

example, in the beginning, the model is trained with one hour of data and a prediction of future values is made for the next eight hours. To suppress false alarms, constraints are set in place: to create an alarm or fault there must be more than one consecutive predicted alarm or fault. Additionally, both consecutive predictions must pass either the high or low limit twice before triggering an alarm. These filters suppress most false alarms. Finally, it is examined to see whether the predicted value is over the specified alarm or fault limits. The results of all combinations are presented in Appendix 5, where the prediction horizon, historical data usage, RMSE, fault detection rate and false alarms are presented. As seen in the Appendix 5, using less historical data and a longer prediction period increases the false alarm rate. This is caused by a poor fit of the data and high variation of the slope of the linear fit. The best subset of the results is selected based on the false alarm rate and shown in table 4.

Table 4. A subset of the results for the linear regression model, where false alarms are not generated and thus suitable for the final model.

Prediction horizon (minutes)	***Historical data used (minutes)***	***RMSE***	***Detection rate (11.04.2017 - 19.4.2017)***	***False alarms***
480	60	163,82	0	0
480	180	25,15	0	0
480	300	10,97	0	0
480	420	6,77	0	0
480	540	5,27	0	0
480	660	4,36	0	0
2400	180	140,35	0	0
2400	300	63,97	0	0
2400	420	37,94	0	0
2400	540	25,57	0	0
2400	660	17,40	0	0
4320	300	118,49	1	0
4320	420	71,07	4	0
4320	540	48,91	0	0
4320	660	33,96	0	0
6240	540	72,17	4	0
6240	660	50,45	3	0
8160	540	95,46	6	0
8160	660	67,20	8	0
10080	660	83,95	11	0

The best results, when related to the detection of the fault and false alarms, are achieved when the model is trained with 6 hours of available data and prediction is made up to 10 days ahead (table 4). With these parameters, the model detects the decreasing trend (April 2017) 7 days before the alarm occurs and creates the fewest false alarms. A visual presentation of the models' predictions is shown in figure 19.

Figure 19. PT2-235_me measurement, limits, and the running predictions of the least squares regression. The green lines are predictions without alarms, the yellow lines are alarms and red lines are faults.

The least squares linear regression can be used as a computationally efficient and confident predictor. The linear regression predicts the degrading condition and does not create false alarms when appropriately filtered. In figure 19 the linear regression is trained on 6 hours of historical data and a prediction is made 10 days into the future. If the predicted value is below the alarm or fault limit, the line is colored green. The next prediction is made with the next 6 hours of data, and another 10-day prediction is created. If the alarm or fault limits are passed for two consecutive predictions, the line is colored yellow or red, respectively. This allows for computationally efficient simple real-time predictions. Additionally, the date when the measurement will cause an alarm or fault if no preventive actions are taken can be estimated.

A similar approach is taken with the ARIMA and VAR models. A feature set is selected for both models including all measurements presented in table 3. The Akaike Information Criterion

is used in both cases to select the optimal model to make predictions. For the ARIMA model, exhaustive training with different parameters for (p, q) in the equation (6) are tested. The order of integration is automatically tested to make the series stationary. To reduce noise in the data and decrease the computational time, a daily mean is calculated for each measurement, shown in figure 20.

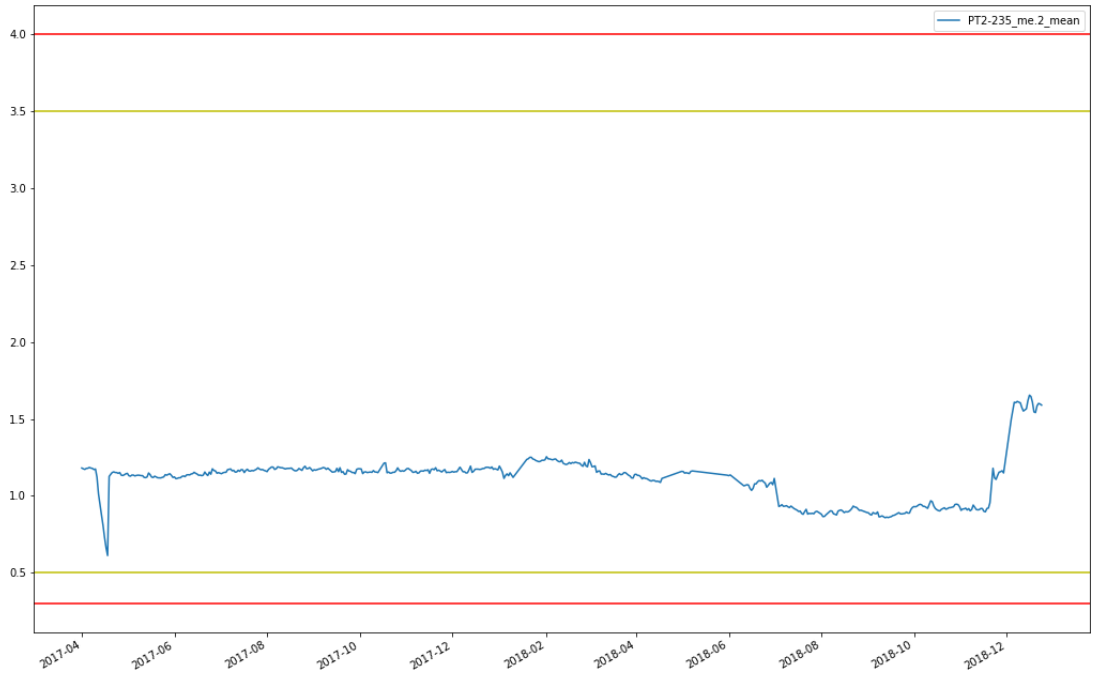


Figure 20. Daily mean calculated for PT2-235_me measurement.

Additional data acquired during this thesis is also included in the model. Figure 20 contains more recent data, November and December 2018, when compared to figures 18 & 19. Each measurement is tested separately for the ARIMA model. The goal with the ARIMA model is, like the linear regression, the prediction of future values using historical measurement data. Data introduced to the model is changed from hourly to daily averages and thus fewer combinations can be tested. Forecasts for multiple steps ahead are made iteratively, applying recursively the forecast results into the model, until the desired prediction horizon is reached. The measurements rarely include trending components (figure 20), thus the performance of the ARIMA models are poor and additionally, the iterative forecasting is computationally ineffective. The daily averages are used to reduce the computational time.

Table 5. A subset of the results for the ARIMA model, where false alarms are not generated and thus suitable for the final model.

Prediction horizon (Days)	*Historical data used (Days)*	*RMSE*	*Detection rate (11.04.2017 - 19.4.2017)*	*False alarms*
1	4	87,81	0	0
1	5	12,61	0	0
1	6	6,43	0	0
1	7	3,40	0	0
1	8	3,13	0	0
3	4	2,21	0	0
3	5	76,75	0	0
3	6	34,18	0	0
3	7	21,87	0	0
3	8	13,61	0	0
5	4	8,96	0	0
5	5	67,50	0	0
5	6	41,58	0	0
5	7	24,79	0	0
5	8	19,12	0	0
7	4	40,59	2	0
7	5	29,97	2	0
7	6	52,93	1	0
7	7	36,85	1	0
7	8	46,53	1	0

The ARIMA model achieves a better RMSE score when related to the linear regression (table 4 & table 5), but performs poorly with the detection of the fault during April 2017. However, this is expected as the model uses daily averages and thus fewer predictions are made. The VAR model is tested to search for relationships between the measurements to predict the future values using multiple measurements as inputs. The hypothesis is that oil pressure and temperature, and their own lagged values, could be used to make accurate predictions of future values. Another exhaustive training and testing is performed with the VAR model: the lag order p in equation (7), k-length vector of input variables, available training data, and prediction horizons are brute-force grid-tested. VAR model performance is on the same level as the ARIMA model, achieving better RMSE scores than linear regression, but performing poorly with fault detection. Unfortunately, the time series data's lack of trending components caused poor performance with the autoregressive models. Additionally, ARIMA and VAR require code from the python package *statsmodels* and this can cause problems when deploying the python codes to AWS Lambda. Better performance could have been achieved utilizing the hourly data, but due to computationally inefficient iterative forecasting, the approach was discarded. However, this leaves room for future research and improvements.

5.2.3 Machine learning models

This section focuses on classifying the degradation states described in section 3. Modern machine learning techniques are used to classify different running conditions into healthy or unhealthy stages, using a two-stage division. Historical data from different faults, shown in Appendix 2, are utilized. First, exploratory visual analysis of the tags MsgW_BrS_HelpW_M1_1, MsgW_BrS_HelpW_M1_2, and MsgW_BrS_HelpW_M2, which display the occurrences of faults is performed, shown in figure 21. The goal for the models in this section is different from the linear regression, ARIMA, and VAR models. Instead of predicting future measurement values to predict a fault, historical fault data are utilized, and the machine learning models are trained to recognize similar behavior in the future. Therefore, the problem to be solved in this section is of classification instead of regression.

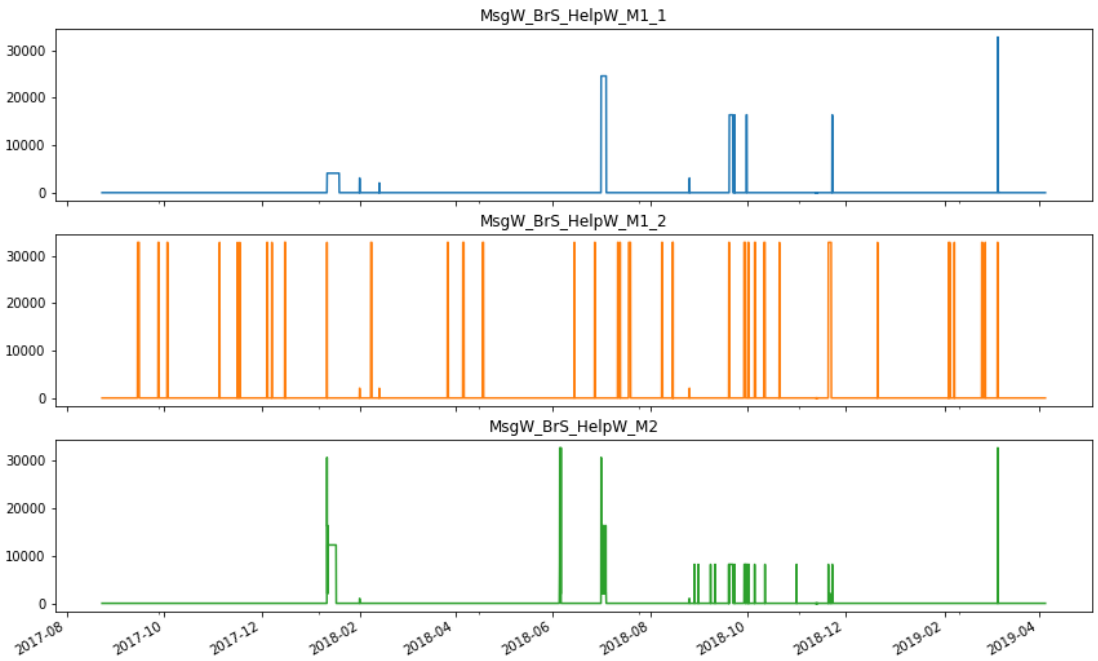


Figure 21. Tags MsgW_BrS_HelpW_M1_1, MsgW_BrS_HelpW_M1_2, and MsgW_BrS_HelpW_M2. Values on the y-axis represent different faults, shown in Appendix 2.

Tag MsgW_BrS_HelpW_M1_2 contains the most often occurring fault with a corresponding value of 32768 (Figure 21). The description for this alarm is *Fault: Wire not run* (Appendix 2), which indicates the breast roll shaker interlocking due to the low wire speed. This indicates

inadequate operator actions, as the breast roll shaker should be shut down before shutting down the wire, but this is not an actual fault that can be predicted (Valmet internal source). Due to the massive amount of data, around 50 million observations of each measurement, the data is filtered when extracted from the database. An hourly mean, standard deviation, minimum, and maximum values are calculated when queried from the database. The hourly mean is calculated by truncating the timestamps hourly and calculating the mean value of all the measurements for the truncated period. Snowflake scalable warehouses can be utilized in the query, which reduces the volume of the data significantly. This removes unimportant noise from the signal while preserving the features within the signal over time that may be informational. Additionally, filtering for the data is done according to the M1-232_r signal to discard measurement values when the breast roll shaker is not operational. Faults related to the wire speed are discarded. According to Valmet internal sources, the most important features in the data consist of 10 different measurements, shown in table 6.

Table 6. Most important features, units, and descriptions.

Measurement	***Unit***	***Description***
TT1-235_me	C°	Hydraulic tank oil temperature
PT1-235_me	MPa	Coupling lubrication pressure
PT2-235_me	MPa	Bearing lubrication pressure
PT3-235_me	MPa	Carriage flotation pressure
PT4-235_me	MPa	Mesh lubrication pressure
PT5-235_me	MPa	Phase angle control pressure
PDT1-235_me	kPa	Lubrication line filter pressure differential
GI-233.1_me	mm	Carriage center point
GIY-235_me	mm	Stroke length measurement
SIY-235_me	Hz	Stroke frequency measurement

The measurements in table 6 are used as features. For each measurement, the hourly average, standard deviation, minimum, and maximum values are calculated, thus totaling 40 different features. Most of the important features are different pressure measurements (Table 4). As an example, a single fault is extracted to examine possible models for the fault. Starting with the most frequent fault from measurement MsgW_BrS_HelpW_M1_1, the value of 16384 (figure 21), which corresponds to *Fault: Carriage Flotation Oil Pressure*. The measurement value is

changed to one or zero depending if the value of the measurement is 16384 or not. Furthermore, only the timestamps when the fault first occurred each time are extracted. The extracted carriage flotation oil pressure faults are shown in figure 22.

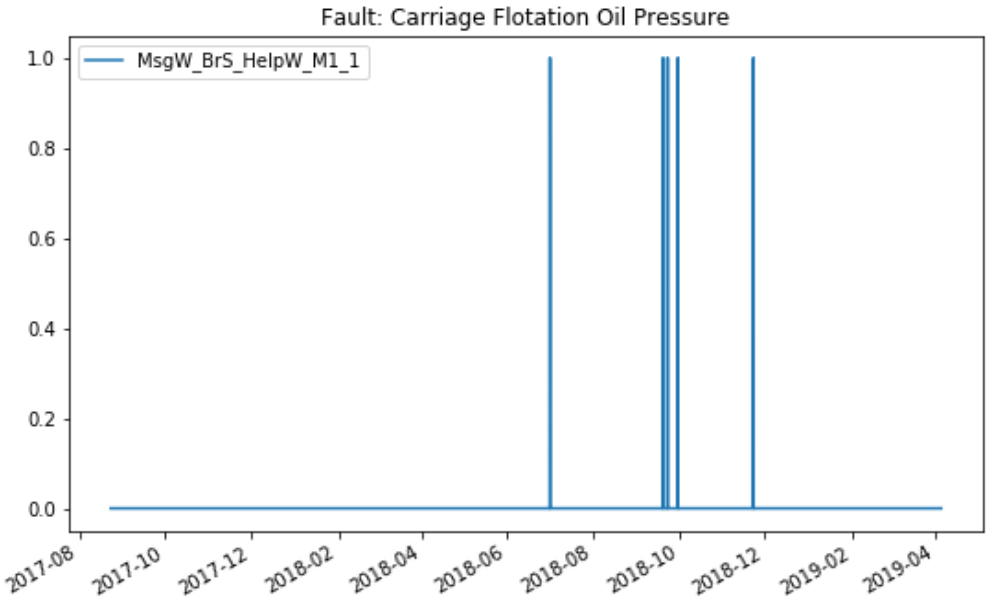


Figure 22. Carriage flotation oil pressure faults that occurred during 2018. The y-axis value of 1 represents the fault.

The carriage flotation oil pressure fault occurred a total of five times during the observation period (figure 22). The exact timestamps are presented in table 7.

Table 7. Carriage flotation oil pressure fault timestamps numbered in chronological order.

Time	***Fault number***
2018-07-01 07:00:00	1
2018-09-19 12:00:00	2
2018-09-22 19:00:00	3
2018-09-30 03:00:00	4
2018-11-12 19:00:00	5

To visually represent the faults, measurement data, and the degradation stages, the measurement data is scaled between zero and one, where zero represents the measurement's minimum value of the observation period and one represents the maximum value. The visualization of all

carriage flotation oil pressure faults and the measurement data is shown in figure 23. Figure 24 shows fault number 1 with arbitrarily chosen degradation stages.

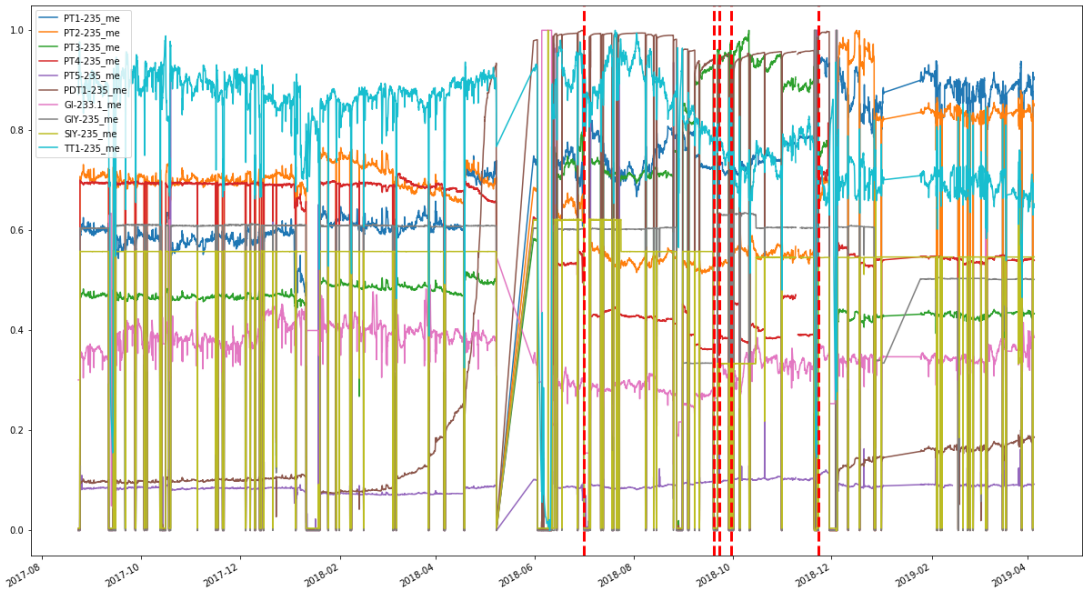


Figure 23. The most important feature data and carriage flotation oil pressure faults. The dashed red vertical lines represent the faults. The different colored lines represent different measurement values scaled between 0-1.

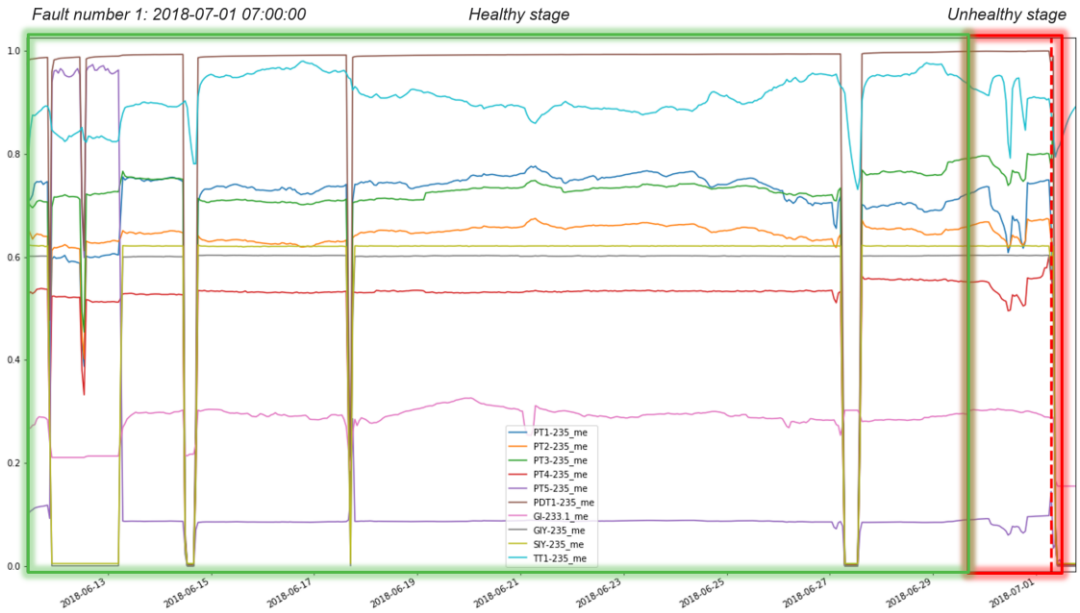


Figure 24. Two-stage degradation stage division. The green box represents a healthy stage. The red unhealthy stage is approximately 48 hours before the fault.

Figure 23 shows all five faults caused by carriage flotation oil pressure. Figure 24 represents, how possible degradation stages for the first fault could be set for the predictive model. The measurement data shown is the hourly average queried from the Snowflake database. The hourly standard deviation, minimum and maximum values are not visualized, but are used in training the models. According to Valmet internal sources, a realistic prediction horizon (unhealthy stage) is around 24 – 48 hours before the fault. Faults shown in figure 22 are extended 24 hours before the faults occur and this set is used as the response variable. Additional sets are created for 36 hours and 48 hours before the failure. Explanatory features are the features specified in table 4, including the hourly averages, minimums, maximums and standard deviations for each measurement. Therefore, the data introduced for the models consists of 40 explanatory features and 1 response variable, in this case, the carriage flotation oil pressure fault data shown in figure 22. However, the data in figure 22 is expanded as shown in figure 24 to use the model as a binary classifier for healthy and unhealthy stages.

Initial model validation is done using the holdout test set of the data. Multiple different models are trained, specifying the parameter grid for each model's parameters. K-Nearest Neighbors, Support Vector Classification, Multi-Layer Perceptron (MLP), Gaussian Naïve Bayes (GNB), and Gradient Boosting models are trained and tested. A holdout test set of 20 % is extracted from the data. During the training, multiple different parameters are tested for each model and for each parameter combination, the training set is 3-fold cross-validated. The best parameter combination is selected from the cross-validation based on precision, and the 20 % holdout test set is used to evaluate the performance of each model. Figure 25 presents this model comparison workflow.

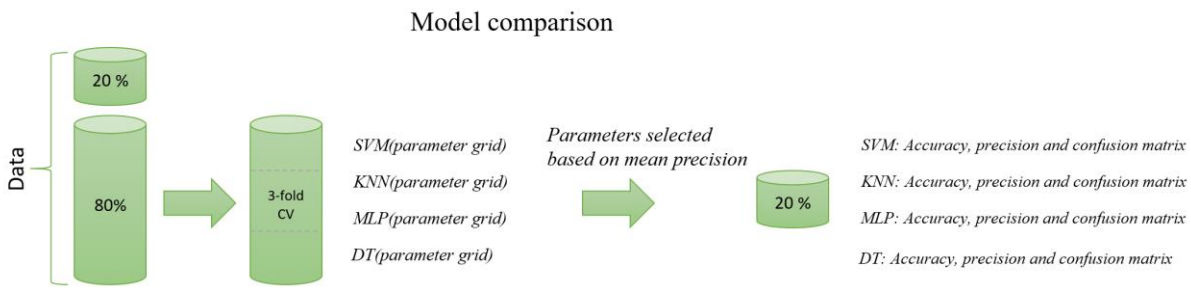
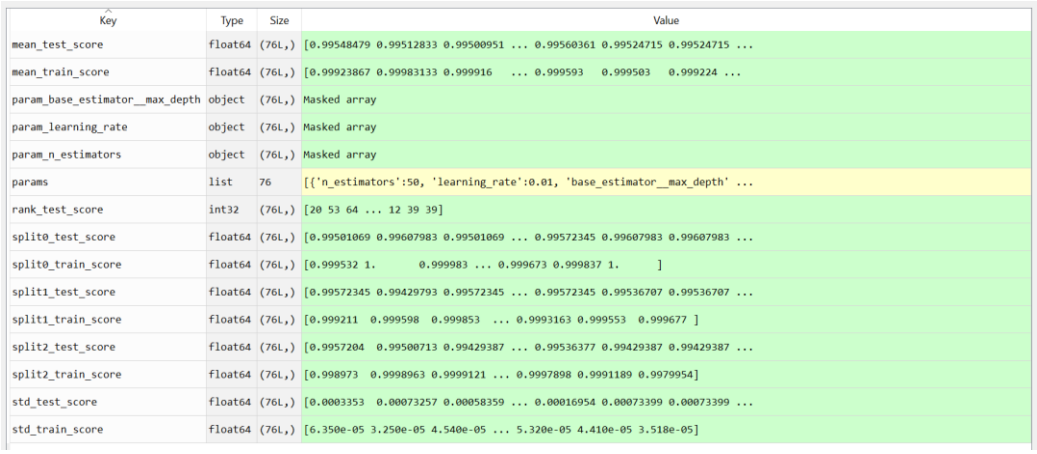


Figure 25. Model comparison workflow.

In figure 25, the parameter grid includes a wide range of possible hyperparameters for each model. For each parameter combination, the mean precision of the folds is calculated and the parameter combination with the highest precision is selected. For example, different results of the decision tree classifier between the folds in the cross-validation phase are shown in figure 26.



Key	Type	Size	Value
mean_test_score	float64	(76L,)	[0.99548479 0.99512833 0.99500951 ... 0.99560361 0.99524715 0.99524715 ...
mean_train_score	float64	(76L,)	[0.99923867 0.99983133 0.999916 ... 0.999593 0.999503 0.999224 ...
param_base_estimator__max_depth	object	(76L,)	Masked array
param_learning_rate	object	(76L,)	Masked array
param_n_estimators	object	(76L,)	Masked array
params	list	76	[{'n_estimators':50, 'learning_rate':0.01, 'base_estimator__max_depth' ...
rank_test_score	int32	(76L,)	[20 53 64 ... 12 39 39]
split0_test_score	float64	(76L,)	[0.99501069 0.99607983 0.99501069 ... 0.99572345 0.99607983 0.99607983 ...
split0_train_score	float64	(76L,)	[0.999532 1. 0.999983 ... 0.999673 0.999837 1.]
split1_test_score	float64	(76L,)	[0.99572345 0.99429793 0.99572345 ... 0.99572345 0.99536707 0.99536707 ...
split1_train_score	float64	(76L,)	[0.999211 0.999598 0.999853 ... 0.9993163 0.999553 0.999677]
split2_test_score	float64	(76L,)	[0.9957204 0.99500713 0.99429387 ... 0.99536377 0.99429387 0.99429387 ...
split2_train_score	float64	(76L,)	[0.998973 0.9998963 0.9999121 ... 0.9997898 0.9991189 0.9979954]
std_test_score	float64	(76L,)	[0.0003353 0.00073257 0.00058359 ... 0.00016954 0.00073399 0.00073399 ...
std_train_score	float64	(76L,)	[6.350e-05 3.250e-05 4.540e-05 ... 5.320e-05 4.410e-05 3.518e-05]

Figure 26. Results from grid-search cross-validation to select the optimal parameters for the decision tree classifier in the model comparison phase.

A total of 76 different parameter combinations were tested in the cross-validation for the gradient boosted decision trees (figure 26). Of interest here is the deviation of the results between the folds in both training and testing, and the mean score of each set: the standard deviation of the cross-validation test set scores is negligibly small.

A similar approach is repeated with KNN, SVM & MLP models to select the optimal parameters for each model. Finally, the 20 % holdout set is used, and classification is performed with each model. Performance metrics calculated from the confusion matrix are accuracy and precision. The confusion matrix and metrics of each optimal model after cross-validation for the holdout test set are shown in figure 27.

```
Training DecisionTreeClassifier GradientBoost with parameter grid search and cross-validation...
Holdout test set metrics
('Accuracy: ', 0.9973691902952353)
('Precision: ', 0.9973337072700814)
Confusion Matrix:
[[3344    3]
 [   6   68]]

Training k Nearest Neighbours model with parameter grid search and cross-validation...
Holdout test set metrics
('Accuracy: ', 0.9947383805904706)
('Precision: ', 0.9945899039879509)
Confusion Matrix:
[[3341    6]
 [  12   62]]

Training Support Vector Classification model with parameter grid search and cross-validation...
Holdout test set metrics
('Accuracy: ', 0.9909383221280328)
('Precision: ', 0.9905515871641838)
Confusion Matrix:
[[3344    3]
 [  28   46]]

Training Gaussian Naive Bayes with holdout test set...
Holdout test set metrics
('Accuracy: ', 0.7997661502484653)
('Precision: ', 0.971606511267115)
Confusion Matrix:
[[2685  662]
 [  23   51]]

Training Multi-Layer Perceptron model with parameter grid search and cross-validation...
Holdout test set metrics
('Accuracy: ', 0.9783688979830459)
('Precision: ', 0.9572057005405596)
Confusion Matrix:
[[3347    0]
 [  74    0]]
```

Figure 27. Performance metrics precision and accuracy of multiple different machine learning models on test set data.

The advantage of the decision tree classifier over the other machine learning classifiers is the robustness to noise and unimportant features. The implementation of scalable tree boosting of gradient boosting discards features with zero scores of importance automatically. (Chen & Guestrin 2016) As seen from figure 27 the decision tree classifier outperforms other common machine learning models when related to model precision. KNN and SVM algorithms perform adequately, the Gaussian Naïve Bayes is overconfident with a large amount of false positive classifications. The dataset is highly imbalanced and thus the MLP neural network classifier achieved the best performance by simply assigning all classes to the majority class. The imbalance issue could be solved by utilizing different data re-sampling techniques, but to stay within the scope of this thesis, the available dataset is used as is.

The implementation of a gradient boosting algorithm for decision trees, introduced in section 3.5 figure 9, is selected as the predictive model. Different parameters of the model are the learning rate v, decision tree depth M and the number of decision trees used. As a decision tree is chosen as the final model, the parameters are optimized even further. The decision tree depth and the learning rate is chosen by running an exhaustive grid-search with 5-fold cross-validation over the five 24-, 36-, and 48-hour periods using precision (equation 24) and accuracy (equation 26) to select the best parameters. To select the optimal number of trees, the cross-validation test error for each estimator is calculated, and the best score is chosen as the optimal value. Visualization of this assessment is presented in figure 28.

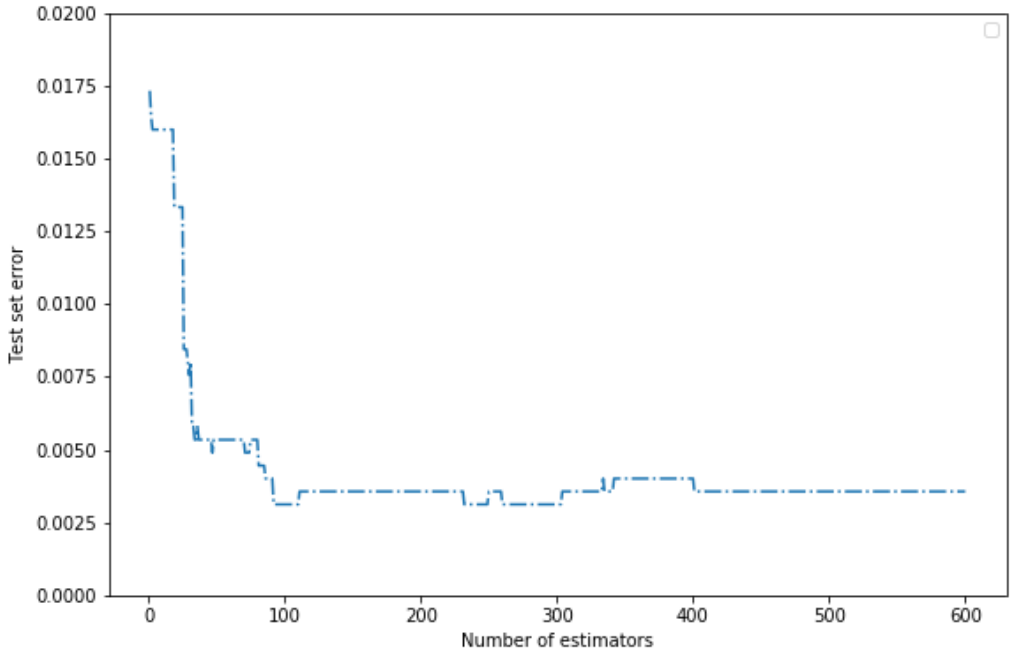


Figure 28. Cross-validation test set error for each number of estimators. The y-axis represents the error on the test set and the x-axis is the number of trees.

After around 100 estimators, there is not much improvement in the test set error. An optimal number of trees for the model predicting carriage flotation oil pressure faults is 98 (Figure 28). A snapshot of the python console during the exhaustive grid search 5-fold cross-validation training for a total of 405 parameter combinations is shown in figure 29.

```
Fitting 5 folds for each of 405 candidates, totalling 2025 fits
[Parallel(n_jobs=4)]: Using backend LokyBackend with 4 concurrent workers.
[Parallel(n_jobs=4)]: Done  24 tasks      | elapsed:   57.5s
[Parallel(n_jobs=4)]: Done 120 tasks      | elapsed:  4.6min
[Parallel(n_jobs=4)]: Done 280 tasks      | elapsed: 10.6min
[Parallel(n_jobs=4)]: Done 504 tasks      | elapsed: 19.0min
[Parallel(n_jobs=4)]: Done 792 tasks      | elapsed: 30.7min
[Parallel(n_jobs=4)]: Done 1144 tasks      | elapsed: 46.2min
[Parallel(n_jobs=4)]: Done 1560 tasks      | elapsed: 65.9min
[Parallel(n_jobs=4)]: Done 2025 out of 2025 | elapsed: 89.4min finished

 Time taken: 1 hours 29 minutes and 39.39 seconds.
```

Figure 29. Verbose during the training of the model for carriage flotation oil pressure faults.

The total time of the training procedure for the 24-hour prediction model was 89.4 minutes (Figure 29). Iterative training of the model is unnecessary; thus, the long training time of the model is irrelevant. Predictions on the test set return probabilities of the data belonging to class 0 or 1, meaning if a fault is predicted in the coming 24-hour window. Again, a confusion matrix is created from the holdout test set to visually observe the model's capability to predict true positives and true negatives on previously unseen data, shown in figure 30.

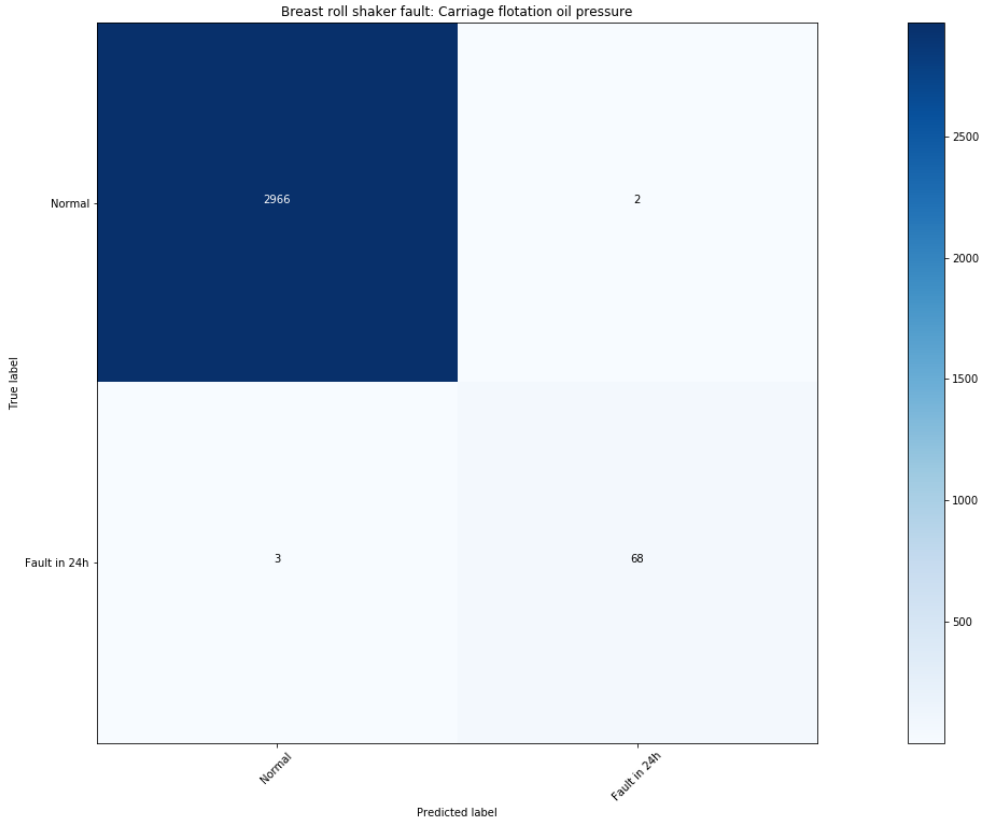


Figure 30. Confusion matrix of the model's performance on unseen test data to predict carriage flotation oil pressure fault.

As seen in figure 30, the model's performance increased slightly when compared to the performance seen in figure 27. The model is surprisingly accurate with only two false positives and three false negatives. The classes are highly imbalanced because the faults occur rarely. However, it does not seem to affect the model's performance negatively as 68 out of 71 actual unhealthy operation stages are classified correctly. The importance of the selected features for the model can also be examined. The 30 most important features are shown in figure 31.

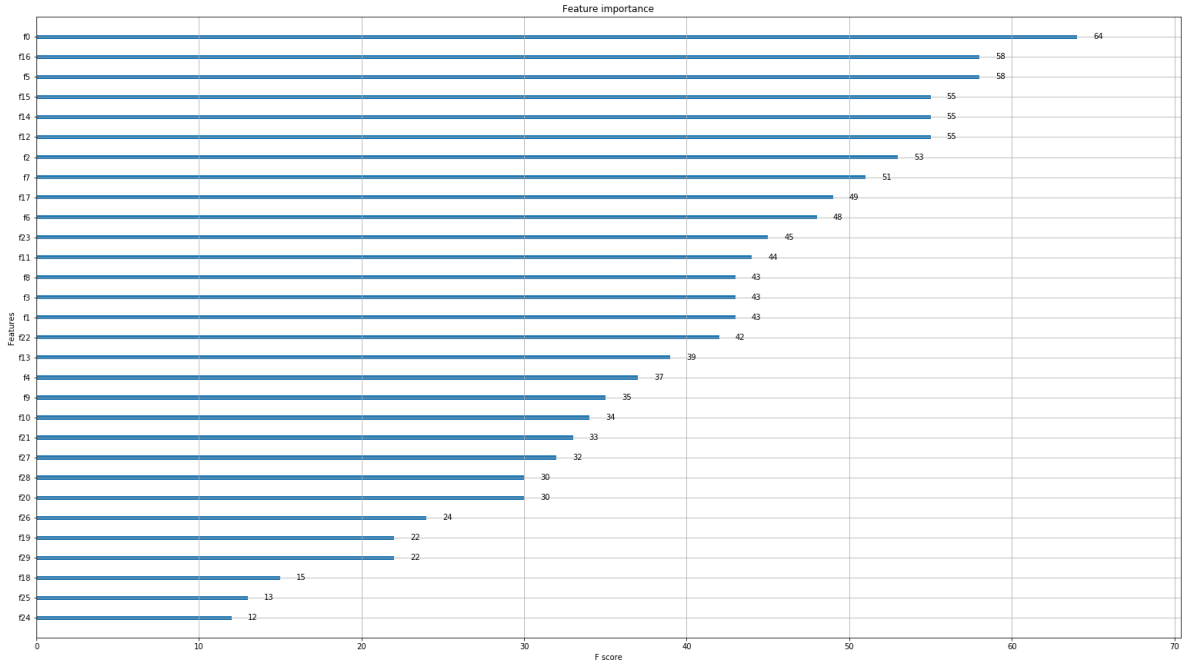


Figure 31. Feature importance score of the selected measurements. Features f0-f30 are on the y-axis and the feature importance is shown on the x-axis.

The feature importance score implies how many times a feature is used to split the data across all decision trees (Chen & Guestrin 2016). In this case, the feature names are obscure as they are presented as f0, f1…f30. The number indicates the index at which they are presented to the model, for example, the three most important features for this model are the average of GI-233.1_me (f0), the average of PT3-235_me (f16), and standard deviation of the PT5-235_me (f5) measurement. The least important features, with feature score below 10, are not shown in figure 31. These consist mostly of the minimum and maximum values of the GIY-235_me, GI-233.1_me, and SIY-235_me measurements. These features are not of importance and thus are

not used to make decisions within the classifier. Feature selection similar to this procedure should be performed before training classifiers such as SVM, KNN, and MLP.

A similar training procedure is repeated for the 36- and 48-hour window degradation stages and the prediction window with the highest precision is selected. Furthermore, a similar training procedure is repeated for all faults and for all breast roll shakers. The best performing models for each fault are saved to be used in future real-time predictions.

5.3 Model deployment and visualization

The python codes and models created in section 5.2 are deployed to AWS Lambda to run continuously in a serverless environment. The results of each model and prediction are saved into Snowflake in the appropriate format, this is also done with python, which is deployed to Lambda. Tableau is connected to the Snowflake table where the results are saved to allow for visualization of the results and predictions. The values, limits and analysis results, are also saved in the Snowflake database application layer table, which allows for near real-time monitoring of the analyses, condition, and actual measurement values. Snapshots of the tables are shown in appendices 6 & 7. The different visualizations are created as Tableau worksheets, which are later combined to a single dashboard.

The first model, pump cavitation, is presented in Tableau as pressure variation where the analysis value indicates the ratio of the current moving pressure deviation when related to the average of the moving pressure deviation. The results of the phase angle calibrations are presented in a similar way. Both visualizations, created with Tableau, are shown in figure 32.

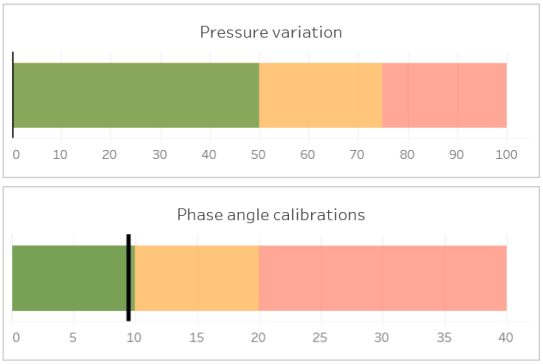


Figure 32. Calculation results visualized in Tableau. The green area indicates normal operating condition, the yellow area indicates above average conditions, and the red area is significantly above average conditions.

The simple visualization (figure 32) presents a quick overview of the condition. The values where the visualization changes from green to yellow and yellow to red are calculated as the average values and their quantiles. The black vertical line indicates the present analysis value of each model. In Appendix 6, specific columns can be found, which indicate the analysis type, limits and analysis value.

The time series prediction is visualized similarly to figure 18. The current measurement value is visualized along with alarm and fault limits. Additionally, the possible predicted date of an alarm or fault is visualized as a vertical line. This visualization is presented in figure 33.

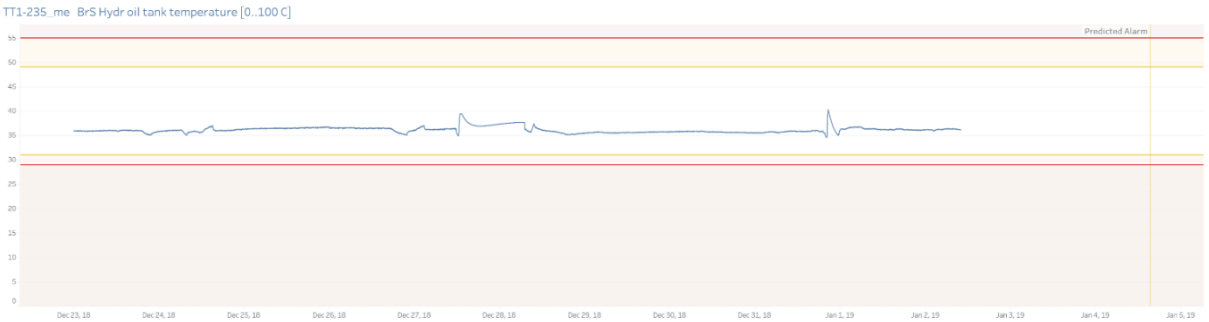


Figure 33. Tableau visualization of the time series prediction. The blue line is the measurement value. The horizontal yellow and red lines indicate the alarm and fault limits, respectively. The vertical yellow line is the predicted date for an alarm.

Figure 33 is an example of one of the measurements that are analyzed, and a prediction that is calculated. Similar visualizations are created for each measurement in table 3. The yellow vertical line is artificially created to test how the visualization works when an alarm or fault is predicted. The similar vertical line is colored red when a fault is predicted.

The results of the gradient boosting algorithm for decision trees are saved in a separate snowflake table. The probability of each timestamp belonging to the unhealthy stage is visualized, along with the important features for each model. In this case, the probability indicates the health of the equipment and visualization like figure 34 can be created.

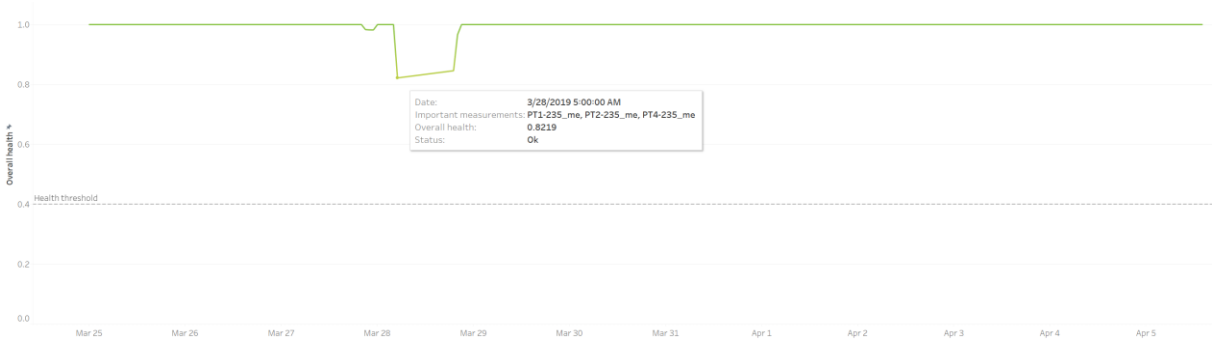


Figure 34. Visualization of the machine learning model results. The y-axis shows the complement probability of a fault, the x-axis is time and the green line indicates the current probability. The dashed vertical grey line indicates the threshold when the state is classified as unhealthy.

Appendix 7 presents the snowflake table format, which allows the visualization shown in figure 34. The model's complement probability of a fault is visualized; when the probability decreases below 0.5 the model classifies the observation as unhealthy and a fault is predicted. The tooltip on the trend shows the important measurements which contribute most to the prediction output.

As seen in Appendix 6 & 7, the status of the different analyses are saved in separate columns as binary information. For example, if the machine learning model predicts a fault, the "PREDICTED_FAULT" column is one, otherwise, the column has a value of zero. Similar columns are included for the pump cavitation, phase angle calibration, and linear regression models; STATUS_OK, STATUS_ALARM, and STATUS_FAULT. This enables simple

visualization of each analysis condition with simple color-coded visualizations, shown in figure 35.

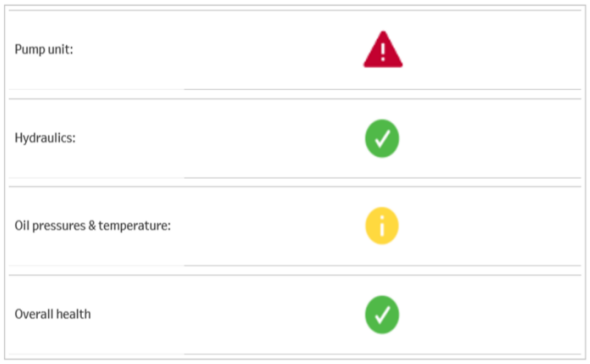


Figure 35. Simple color coding for the most recent analysis condition.

The indicator in figure 35 depends on the value of the different status columns in the snowflake database. Green indicates that the status is normal, yellow indicates an alarm, and red indicates a fault. The pump unit is related to the results of the pump cavitation calculation. Hydraulics indicate the condition of the phase angle calibrations – indicating possible oil leaks in the hydraulics. Oil pressures & temperature is the result of the linear regression for predicting alarms and faults for the different oil pressure & temperature measurements. Overall health is the status of the machine learning model results. Finally, the worksheet visualizations shown in figures 32-35 are combined into a single dashboard. The dashboard is shown in figure 36.

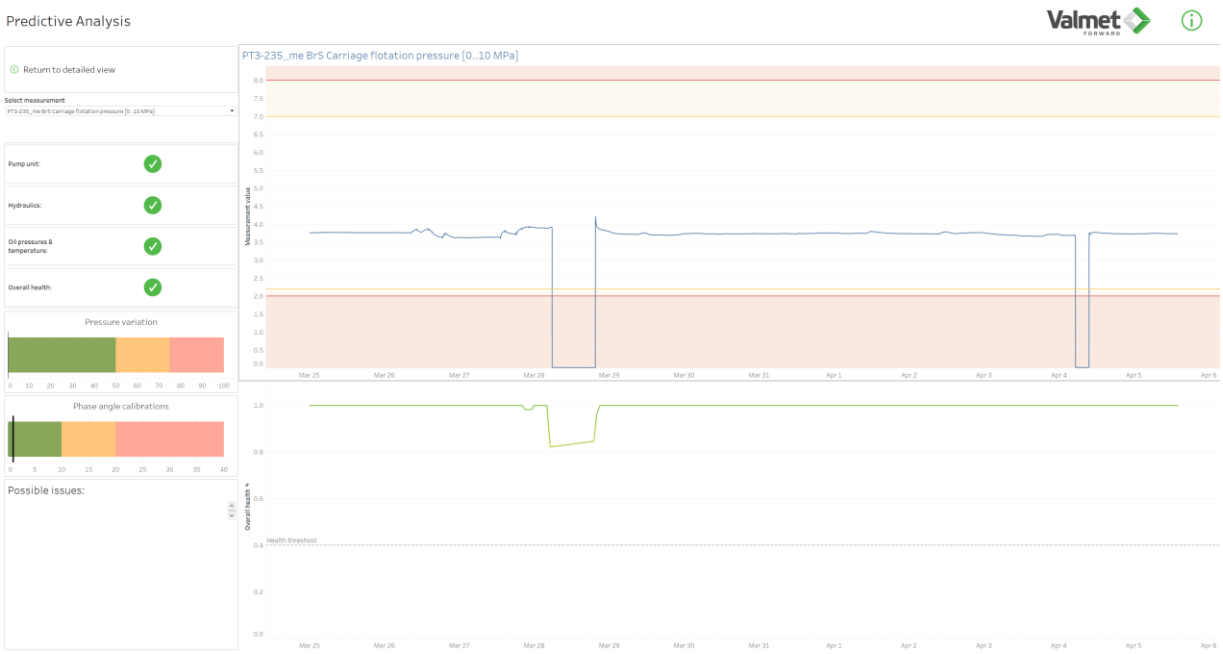


Figure 36. Combined worksheets to a single Tableau dashboard view.

The visualization shown in figure 36 includes the results of the different analyzes. The above trend visualization shows the current measurement value of the selected measurement, both limits and the possibly predicted alarm or fault date. Below the measurement trend is the visualization of the machine learning model results. On the left of these trend visualizations, there is a possibility to select a different measurement to explore, a quick overview of the most recent status of the different analyzes, and the current condition of the pump unit and phase angle calibration analyses. Possible issues list the tag if the linear regression predicts an alarm or fault, or the fault type of the machine learning model, for example, *Possible fault: Carriage flotation oil pressure*. The information saved in the FORMMASTER_DV_ID column (Appendix 6 & Appendix 7) allows a dimensional table to be joined to the fact table to get detailed information about the breast roll shaker in question. The visualization of this classified information is not shown in this work.

6 CONCLUSIONS AND FUTURE WORK

The goal of this thesis was to deploy a predictive maintenance model to prevent an unplanned shutdown of the equipment in question. The goal was achieved by following the four main objectives; identifying current maintenance plan weaknesses and possible key degradation components (chapter 2), studying the equipment's mechanics and operation (chapter 2), literature search of predictive models and algorithms (chapter 3), and utilizing the case company's industrial internet architecture (chapters 4 & 5).

The main findings in chapter 2 concerning the breast roll shaker were oil leaks, failure of the pump unit, acceptable limits of different measurements and historical data of faults. In chapters 4 & 5 each finding was explored with the suitable analytical method found in the theoretical framework in chapter 3.

The pump failures and oil leaks are detected with a model based on an understanding of the breast roll shaker and the underlying physics. By analyzing different important measurements from the process, physics-based models were developed to predict a possible pump failure and show the condition of the hydraulic system in terms of oil leakage.

To predict the future value of a measurement, multiple different models were comprehensively tested: least squares regression, ARIMA and VAR. The predictions are constructed to identify whether the measurement will remain within acceptable limits. Least squares regression implementation performed optimally with regards to prediction accuracy and computational efficiency.

The historical data of faults were utilized to train a machine learning model to predict a similar fault in the future. Multiple different models were tested, such as KNN, SVM, NN, and decision trees. The models are trained to predict a fault in both 24-hour and 48-hour windows before the occurrence of the fault. The decision tree model performed the most accurately and additionally, being a white box model where feature importance can be estimated, this model was selected as the optimal model. All predictive models developed are deployed to run in a serverless in

AWS Lambda environment. The results of each model are saved in a Snowflake database and visualized with Tableau. This allows for real-time monitoring of each connected breast roll shaker's condition and predicts unplanned shutdowns.

The models developed in this thesis performed suitably. However, spending more time on exhaustive research and testing could increase the accuracy of the models. A grid-search was performed, but increasing the number of parameters tested increases the computational time significantly. Bayesian optimization or another parameter search method could be tested to find the optimal parameters for the model.

Another interesting field for future research is hybrid models; combining multiple different models to compensate for each model's weaknesses and strengths. For example, by conflating the probabilities of overconfident and simple assumptions made from Gaussian distribution, combined with a highly accurate machine learning model, could give the most important features when health degrades. The accurate machine learning model cancels the false positives and false negatives, while the feature scores and importance are drawn from the Gaussian distribution for true positives.

Feature selection was not introduced in this thesis. The selected decision tree model automatically discards features with zero scores of importance during the training phase. However, using feature selection and possibly calculating new features could significantly increase the performance of the model. Currently, there is a lot of literature available for feature selection and that could be the sole subject of a single Thesis.

Finally, all models were tested separately on every breast roll shaker. In the future, it should be tested whether the data from the breast roll shaker pool can be utilized to train a generalized model. These generalized models can be used with new breast roll shaker deliveries to predict unnecessary shutdowns without the need for training new models from historical data.

REFERENCES

Amazon Web Services a, “Amazon S3” [online] https://aws.amazon.com/s3/ [Accessed: 1.1.2019]

Amazon Web Services b, “What is AWS Lambda? – AWS Lambda.” [online] https://docs.aws.amazon.com/lambda/latest/dg/welcome.html [Accessed: 1.1.2019]

Amazon Web Services c, “What is Amazon EC2?” [online] https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html [Accessed: 1.1.2019]

Amazon Web Services d, “Amazon Simple Notification Service Documentation” [online] https://docs.aws.amazon.com/sns/index.html#lang/en_us [Accessed: 1.1.2019]

Banjevic, D., Jardine, A., Makis, V. & Ennis, M. 2001. A control-limit policy and software for condition-based maintenance optimization. *Inform. Syst. Oper. Res*, vol 39, pp. 32-50.

Barga, R., Fontama, V., Tok, W. H. & Cabrera-Cordon, L. 2015. Predictive analytics with Microsoft Azure machine learning. Apress.

Breiman, L. 1996. Bagging predictors. *Machine Learning*, vol 26, pp. 123-140.

BRS Manual. Valmet’s internal source.

Caesarendra, W., Widodo, A., Thom, P.H., Yang, B.S. & Setiawan, J.D. 2011. Combined probability approach and indirect data-driven method for bearing degradation prognostics. *IEEE Trans. Reliab,* vol 60, pp. 14-20.

CEN 2010. Maintenance – Maintenance terminology. BSI Standards Publication [online]. Available: http://irma-award.ir/wp-content/uploads/2017/08/BS-EN-13306-2010.pdf [Accessed: 12.12.2018]

Chen, T. & Guestrin, C. 2016. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794.

Collin, J. & Saarelainen, A. 2016. Teollinen Internet, Helsinki: Talentum.

Devore, J. L. 2011. Probability and Statistics for Engineering and the Sciences. Cengage learning.

Friedman, J.H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of statistics,* pp. 1189-1232.

Gebraeel, N., Elwany, A. & Pan, J. 2009. Residual life predictions in the absence of prior degradation knowledge, *IEEE Trans. Reliab,* vol 58, pp. 106-117.

George, A., Seber, F. & Lee, A.J. 2003. Linear Regression Analysis. 2 ed. New Jersey: John Wiley & Sons.

Giorgio, M., Guida, M. & Pulcini, G. 2011. An age- and state-dependent Markov model for degradation process. *IIE Trans,* vol 42, pp. 621-632.

Hamilton, J.D. 1994. Time Series Analysis. 2 ed. New Jersey: Princeton university press.

Heng, A., Zhang, S., Tan, A.C.C. & Mathew, J. 2009. Rotating machinery prognostics: state of the art, challenges and opportunities. *Mechanical Systems and Signal Processing,* vol 23, pp. 724-734.

Hu, C., Youn, B.D., Kim, T. & Wang, P. 2015. A co-training-based approach for prediction of remaining useful life utilizing both failure and suspension data. *Mechanical Systems and Signal Processing,* vol 62-63, pp. 75-90.

Hu, C., Youn, B.D., Wang, P. & Yoon, J.T. 2012. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering and System Safety*, vol 13, pp. 120-135.

Jardine, A.K.S., Lin, D.M. & Banjevic, A. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, vol 20, pp. 1483-1510.

Jacobs, R.B. 1961. Prediction of Symptoms of Cavitation. *Engineering and Instrumentation*, vol 65C.

Jin, X., Sun, Y., Que, Z., Wang, Y. & Chow, T.W.S. 2016. Anomaly detection and fault prognosis for bearings, *IEEE Trans. Instrum. Meas.*, vol 65, pp. 2046-2054.

Kaggle, Winning solutions of kaggle competitions [online]. Available: https://www.kaggle.com/sudalairajkumar/winning-solutions-of-kaggle-competitions [Accessed: 1.1.2019]

Kan, M.S., Tan, A.C.C. & Mathew, J. 2015. A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, vol 63, pp. 1-20.

Kimotho, J.K., Sondermann-Woelke, C., Meyer, T. & Sextro, W. 2013. Machinery prognostic method based on multi-class support vector machines and hybrid differential evolution-particle swarm optimization. *Chern. Eng. Trans.*, vol 33, pp. 619-624.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T. & Lin, J. 2018. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, vol 104, pp. 799-834.

Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S. & Liang, S. 1999. Adaptive prognostics for rolling element bearing condition. *Mechanical Systems and Signal Processing,* vol 13, pp. 103-113.

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

Liao, L.X. & Kottig, F. 2014. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Trans. Reliab,* vol 63, pp. 191-207.

Liu, D., Xie, W., Liao, H. & Peng, Y. 2014. An integrated probabilistic approach to lithium-ion battery remaining useful life estimation, *IEEE Trans. Instrum. Meas.,* vol 64, pp. 660-670.

Mahamad, A.K., Saon, S. & Hivama, T. 2010. Predicting remaining useful life of rotating machinery based artificial neural network. *Computers and Mathematics with Applications,* vol 60, pp. 1078-1087.

Martinsson, E. 2016. WTTE-RNN: Weibull Time To Event Recurrent Neural Network. Master's Thesis, Chalmers University of Technology, Gothenburg, Sweden.

Mobley, R.K. 2002. An introduction to predictive maintenance. Elsevier.

Paris, P.C. & Erdogan, F. 1963. A critical analysis of crack propagation laws. *J. Fluids Eng,* vol 85, pp. 528-533.

Peng, Y., Wang, H., Wang, J., Liu, D. & Peng, X. 2013. A modified echo state network based remaining useful life estimation approach. *IEEE International Conference on Prognostics and Health management,* pp. 1-7.

Samanipour, P., Poshtan, J. & Sadeghi, H. 2017. Cavitation detection in centrifugal pumps using pressure time-domain features. *Turk. J. Elec. Eng. & Comp. Sci,* vol 25, pp. 4287-4298.

Si, X.S., Wang, W., Hu, C.H. & Zhou, D.H. 2011. Remaining useful life estimation – a review on the statistical data driven approach. *Eur. J. Oper. Res,* vol 213, pp. 1-14.

Sikorska, J.Z., Hodkiewicz M. & L. Ma. 2011. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing,* vol 25, pp. 1803-1836.

Sloukia, P., Aroussi, M.E., Medromi, H. & Wahbi, M. 2013. Bearings prognostic using mixture of Gaussian hidden markov model and support vector machine. *IEEE International Conference on Computer Systems and Applications,* pp. 1-4.

Snowflake, "Why Snowflake" [online]. Available: https://www.snowflake.com/product/why-snowflake/. [Accessed: 29.11.2018]

Snowflake Manual. Welcome to the Snowflake Documentation. [online] Available: https://docs.snowflake.net/manuals/index.html [Accessed: 29.11.2018]

Tang, J. & Su, T.S. 2008. Estimating failure time distribution and its parameters based on intermediate data from a Wiener degradation model. *Naval Res. Logist.,* vol 55, pp. 265-276.

Valmet 2018. Valmet's Annual Review, 2018. Espoo, Valmet Oyj.

Valmet internal source. Interviews, e-mail correspondence, and expert reviews.

Vlok, P., Coetzee, J., Banjevic, D., Jardine, A. & Makis, V. 2002. Optimal component replacement decision using vibration monitoring and the proportional hazards model. *J. Oper. Res. Soc,* vol 53, pp. 193-202.

APPENDICES

Appendix 1. Measured signals from the breast roll shaker. Tags, descriptions, typical range, and unit.

Tag	Description, typical range, and unit
PT1-235_me	*BrS Coupling lubrication pressure [0..1.6 MPa]*
PT2-235_me	*BrS Bearing lubrication pressure [0..6 MPa]*
PT3-235_me	*BrS Carriage flotation pressure [0..10 MPa]*
PT4-235_me	*BrS Mesh lubrication pressure [0..6 MPa]*
PT5-235_me	*BrS Phase angle control pressure [0..16 MPa]*
PDT1-235_me	*BrS Lubrication line filter [0..500 kPa]*
GT1-233_me	*BrS Carriage position [0..xx mm]*
GT1-235_me	*BrS Phase angle position [0..104 mm]*
TT1-235_me	*BrS Hydr oil tank temperature [0..100 C]*
PCV1B-235_me	*BrS Phaseangle valve spool pos [-100..100 %]*
XZ1-237_me	*BrS Wire speed [0..1000 m/min]*
M1-232-SCF_me	*BrS Shaker motor speed feedback [0..600 rpm]*
WD1-237_me	*BrS Link watchdog waiting time [0..9999 s]*
H1-232_r	*BrS Hydraulic oil Heater running*
M1-232_r	*BrS Shaker motor running*
M2-232_r	*BrS Hydraulic oil pump Motor running*
GIC-235_me	*BrS Phase angle Control measurement [0..90 deg]*
GIC-235_out	*BrS Phase angle Control output [-100..100 %]*
GIC-235_spa	*BrS Phase angle Control active setpoint [0..90 deg]*
GI-233.1_me	*BrS Carriage center point [0..100 mm]*
GIY-235_me	*BrS Stroke length measurement [0..35 mm]*
SIY-235_me	*BrS Stroke frequency measurement [0..10 Hz]*
WY-235_me	*BrS Breast roll weight [0..4000 kg]*
XIY-235_me	*BrS Shake number [0..10000]*
GIY-235.Dev_me	*BrS Stroke length deviation [mm]*
SIY-235.3_me	*BrS Converter stroke freq feedback [0..10 Hz]*
SIY-235.Dev_me	*BrS Stroke frequency deviation [Hz]*
GIY-235.CENTER_set	*BrS Carriage center point [mm]*
GIC-235.4_set	*BrS Phase angle zero limit [deg]*
GIY-235.4_set	*BrS Stroke length deviation high limit [mm]*
GT1-235.min_Deg_set	*BrS Phase angle Min [deg]*
GY-235_set	*BrS Stroke length setpoint [0..35 mm]*
SIY-235.4_set	*BrS Stroke frequency deviation high limit [Hz]*
SY-235_set	*BrS Stroke frequency setpoint [2.5..10 Hz]*
TI-235_set	*BrS Oil temperature setpoint [32..49 C]*
GIY-235.CTOL_set	*BrS Carriage center point tolerance [mm]*
M2-232.STPD_set	*BrS Hydr pump stopping delay [s]*
GT1-235.gap_mm_set	*BrS Phase angle Gap [mm]*
FT1-238_me	*BrL Lubrication oil flow DS [0..2 l/min]*
FT2-238_me	*BrL Lubrication oil flow TS [0..2 l/min]*
TT1-236_me	*BrL Oil tank temperature [0..100 C]*
H1-236_r	*BrL Lubrication oil Heater*
M1-236_r	*BrL Main pump 1*
M2-236_r	*BrL Main pump 2*
TI-236_set	*BrL Oil temperature setpoint [48..52 C]*
BrS_Stop_by_user	*BrS Stopped by user*
BrS_ConditionLights	*BrS Condition traffic light; green = 1, yellow = 2, red =3*
MsgW_BrS_HelpW_M1_1	*BrS Help messages M1 1/2*
MsgW_BrS_HelpW_M1_2	*BrS Help messages M1 2/2*
MsgW_BrS_HelpW_M2	*BrS Help messages M2*

Appendix 2. Measurement tags BrS_HelpW_M1_1, BrS_HelpW_M1_2, and BrS_HelpW_M2: possible faults, their corresponding values, and descriptions.

Measurement Tag	Value	Description
MsgW_BrS_HelpW_M1_1	*1*	*Fault: Phase angle control oil pressure*
MsgW_BrS_HelpW_M1_1	*2*	*Fault: Lubrication line filter*
MsgW_BrS_HelpW_M1_1	*4*	*Fault: Control line filter*
MsgW_BrS_HelpW_M1_1	*8*	*Fault: Stroke freq setp-meas difference*
MsgW_BrS_HelpW_M1_1	*16*	*Fault: Stroke length setp-meas difference*
MsgW_BrS_HelpW_M1_1	*32*	*Fault: Phase angle measurement*
MsgW_BrS_HelpW_M1_1	*64*	*Fault: Phase angle control*
MsgW_BrS_HelpW_M1_1	*128*	*Fault: Carriage position measurement*
MsgW_BrS_HelpW_M1_1	*256*	*Fault: Motor center fault*
MsgW_BrS_HelpW_M1_1	*512*	*Fault: Safety switch fault*
MsgW_BrS_HelpW_M1_1	*1024*	*Fault: Emergency or quick stop*
MsgW_BrS_HelpW_M1_1	*2048*	*Fault: Hydraulic pump interlocked*
MsgW_BrS_HelpW_M1_1	*4096*	*Fault: Coupling lubrication oil pressure*
MsgW_BrS_HelpW_M1_1	*8192*	*Fault: Bearing lubrication oil pressure*
MsgW_BrS_HelpW_M1_1	*16384*	*Fault: Carriage flotation oil pressure*
MsgW_BrS_HelpW_M1_1	*32768*	*Fault: Mesh lubrication oil pressure*
MsgW_BrS_HelpW_M1_2	*1*	*Fault: Wire speed low*
MsgW_BrS_HelpW_M1_2	*2*	*Fault: Connecting rod state acknowledged*
MsgW_BrS_HelpW_M1_2	*256*	*Fault: carriage position extreme limit*
MsgW_BrS_HelpW_M1_2	*512*	*Fault: Oil temperature too high*
MsgW_BrS_HelpW_M1_2	*1024*	*Fault: Shaker motor stopped unwanted*
MsgW_BrS_HelpW_M1_2	*2048*	*Fault: Hydraulic pump not run with auto mode*
MsgW_BrS_HelpW_M1_2	*4096*	*Fault: (Start interlock, phase angle zero)*
MsgW_BrS_HelpW_M1_2	*8192*	*Fault: (Start interlock, Carriage at center)*
MsgW_BrS_HelpW_M1_2	*16384*	*Fault: (Start interlock, Oil too cold)*
MsgW_BrS_HelpW_M1_2	*32768*	*Fault: Wire not run*
MsgW_BrS_HelpW_M2	*1*	*Fault: Oil temperature too high*
MsgW_BrS_HelpW_M2	*256*	*Fault: Motor center fault*
MsgW_BrS_HelpW_M2	*512*	*Fault: Safety switch fault*
MsgW_BrS_HelpW_M2	*1024*	*Fault: Emergency or quick stop*
MsgW_BrS_HelpW_M2	*2048*	*Fault: Coupling lubrication oil pressure*
MsgW_BrS_HelpW_M2	*4096*	*Fault: Bearing lubrication oil pressure*
MsgW_BrS_HelpW_M2	*8192*	*Fault: Carriage flotation oil pressure*
MsgW_BrS_HelpW_M2	*16384*	*Fault: Mesh lubrication oil pressure*
MsgW_BrS_HelpW_M2	*32768*	*Fault: Phase angle control oil pressure*

Appendix 3. Preventive maintenance plan for the breast roll shaker

Position	Position name	Maintenance Item / Specification	Job Specification	No. of Maint. Items	Maint. Interval	Unit (h/wk)	Lube Method	Qty / Item	Unit	Lubricant	Operation / Shutdown
	FormMaster	Hydraulic and lubrication oil	Replace	1	104	wk	Pressure lubrication	160	l	MO20MI-150	Shutdown
	FormMaster	Hydraulic and lubrication oil	Sampling	1	52	wk					Shutdown
	FormMaster	Schmidt coupling	Clean and check, outside	2	13	wk	Pressure lubrication				Shutdown
	FormMaster	Schmidt coupling	Check operation	2	13	wk	Pressure lubrication				Shutdown
	FormMaster, hydraulic pump	AC motor ABB 7.5 kW	Clean and check, outside	1	52	wk					Operation
	FormMaster, hydraulic pump	AC motor ABB 7.5 kW	Check sound, temperature and vibration	1	26	wk					Operation
	FormMaster, drive motor	AC motor ABB 7.5 kW	Clean and check, outside	1	52	wk					Operation
	FormMaster, drive motor	AC motor ABB 7.5 kW	Check sound, temperature and vibration	1	26	wk					Operation
	FormMaster, carriage	Compression spring	Check operation	4	26	wk					Shutdown
	FormMaster, carriage	Compression spring	Replace	4	104	wk					Shutdown
	Limit switch	Protective bellows	Check wear	1	26	wk					Shutdown
	Limit switch	Protective bellows	Replace	1	104	wk					Shutdown
	Connecting rod	Spherical radial-thrust bearing	Check sound, temperature and vibration	1	4	wk					Operation
	Connecting rod	Spherical radial-thrust bearing	Lubricate	1	5	wk	Nipple	180	g	MG15MI-2	Operation/ Shutdown

Position	Position name	Maintenance Item / Specification	Job Specification	No. of Maint. Items	Maint. Interval	Unit (h/wk)	Lube Method	Qty / Item	Unit	Lubricant
	Connecting rod	Spherical radial-thrust bearing	Lubricate	1	1	wk	Nipple	60	g	MG15MI-2
	Connecting rod	Bellows	Check wear	1	26	wk				
	Connecting rod	Bellows	Replace	1	104	wk				
	Hydraulic system	Oil filter, lubrication filter	Replace	1	according to the differential pressure transmitter alarm.					
	Hydraulic system	Oil filter, loading filter	Replace	1	according to the differential pressure transmitter alarm.					
	Hydraulic system	Pump unit	Check sound, temperature and vibration	1	4	wk				
	Hydraulic system	Flow divider motor	Check sound, temperature and vibration	1	8	wk				
	Hydraulic system	Flow divider motor	Check leaks	1	8	wk				
	Hydraulic system	Air breather	Replace	1	52	wk				
	Hydraulic system	Pressure accumulator	Check leaks	1	26	wk				
	Hydraulic system	Pressure accumulator	Calibration	1	52	wk				
	Hydraulic system	Oil cooler	Check leaks	1	12	wk				

Position	Position name	Maintenance Item / Specification	Job Specification	No. of Maint. Items	Maint. Interval	Unit (h/wk)	Lube Method	Qty / Item	Unit	Lubricant	Operation / Shutdown
	Hydraulic system	Pressure relief valve	Calibration	2	52	wk					Shutdown

Appendix 4. Snapshot of the data queried from Snowflake and saved in text file.


```
GIC-235.4_set,GIC-235.4_set,GIC-235.4_set,PT1-235_me,PT1-235_me,PT1-235_me,PT2-235_me,PT2-235_me,PT2-235_me,PT3-235_me,PT3-235_me,PT3-235_me,PT4-235_me,PT4-235_me,PT4-235_me,PT5-235_me,PT5-
_2,MsgW_BrS_HelpW_M1_2,MsgW_BrS_HelpW_M2,MsgW_BrS_HelpW_M2,MsgW_BrS_HelpW_M2,PCV1B-235_me,PCV1B-235_me,PCV1B-235_me,PDT1-235_me,PDT1-235_me,PDT1-235_me,SIY-235.3_me,SIY-235.3_me,SIY-235.3_m
,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,max,mean,min,ma
2017-10-25 20:07:00,0.8000000011920929,0.8000000011920929,0.8000000011920929,0.0026041665114462376,0.0029513889458030462,0.0027900195698468 73,0.0028211805038154125,0.0034722222480922937,0.0031
8013775165,0.01844617910683155,0.02350983582437038 4,0.02108929779093999,4.976490020751953,4.981554031372 07,4.978910629565899,0.01844617910683155,0.02350983582437038 4,0.02108929779093999,5.0
2017-10-25 20:08:00,,,,0.0026041665114462376,0.0029513889458030462,0.0027669883664634266,0.0028211805038154125,0.0039062497671693563,0.0032294609394492 74,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
65728759766,43.51490020751953,43.499637603759766,0.0,0.0,0.0,0.0,900.0,900.0,900.0
2017-10-25 20:09:00,,,,0.0026041665114462376,0.0029513889458030462,0.0027388649117522 71,0.0023871527519077063,0.0039062497671693563,0.0029371706969586425,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
821041501802,0.0,0.0,0.0,0.0,900.0,900.0,900.0
2017-10-25 20:10:00,,,,0.002662037033587694,0.0029513889458030462,0.0027199073860390208,0.0023871527519077063,0.0039062497671693563,0.0031890006207119107,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
497020657,0.0,0.0,0.0,0.0,900.0,900.0,900.0
2017-10-25 20:11:00,,,,0.002662037033587694,0.0029513889458030462,0.0027434478663855184,0.0019531248835846782,0.0034722222480922937,0.0025894537556253 5,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
465858265505,0.0,0.0,0.0,0.0,900.0,900.0,900.0
2017-10-25 20:12:00,,,,0.0026041665114462376,0.0030671297572553158,0.0027218690907614 21,0.0019531248835846782,0.0036892360076308 25,0.0027586510977155323,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,
453803628,0.0,0.0,0.0,0.0,900.0,900.0,900.0
2017-10-25 20:13:00,,,,0.0026041665114462376,0.0029513889458030462,0.0027365818998571 17,0.0028211805038154125,0.0034722222480922937,0.0030198034522581404,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
43.45421076629121,0.0,0.0,0.0,0.0,900.0,900.0,900.0
2017-10-25 20:14:00,,,,0.0026041665114462376,0.0029513889458030462,0.0027199073702538924,0.0026041665114462376,0.0034722222480922937,0.0028726753455516053,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
,43.43784248222739,0.0,0.0,0.0,0.0,900.0,900.0,900.0
2017-10-25 20:15:00,,,,0.0026041665114462376,0.0029513889458030462,0.0027326584667346234,0.0026041665114462376,0.0039062497671693563,0.0029204919523859 43,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
340655,0.0,0.0,0.0,0.0,900.0,1000.0,969.4915254237288
2017-10-25 20:16:00,,,,0.0027199073228985 07,0.0029513889458030462,0.0028996100739102 09,0.0026041665114462376,0.0039062497671693563,0.0029963145334796423,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,
156,43.41089951364618,0.0,0.0,0.0,0.0,1000.0,1000.0,1000.0
2017-10-25 20:17:00,0.8000000011920929,0.8000000011920929,0.8000000011920929,0.0028356481343507767,0.0028935184236615896,0.0028601317182899667,0.0026041665114462376,0.0039062497671693563,0.003
0.020915408905308 98,4.976490020751953,4.981554031372 07,4.979084546749409,0.01844617910683155,0.02350983582437038 4,0.020908453358480565,5.0,5.0,5.0,43.388309478759766,43.410011291503906,43.3
2017-10-25 20:18:00,,,,0.0028356481343507767,0.0028935184236615896,0.0028523226244911806,0.0023871527519077063,0.0036892360076308 25,0.0028873881427729027,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
921183958,0.0,0.0,0.0,0.0,1000.0,1000.0,1000.0
2017-10-25 20:19:00,0.8000000011920929,0.8000000011920929,0.8000000011920929,0.0028356481343507767,0.0028935184236615896,0.0028683574283090625,0.0026041665114462376,0.0036892360076308 25,0.0029
020443337362097656,4.977213382720947,4.981554031372 07,4.979556643444559,0.01844617910683155,0.02278645895421505,0.020364709279459457,5.0,5.0,5.0,43.31958770751953,43.359375,43.3395591404127
2017-10-25 20:20:00,,,,0.0024884259328246117,0.0029513889458030462,0.0027699308761916 42,0.0026041665114462376,0.0039062497671693563,0.0031595750972312 89,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,
40259206,0.0,0.0,0.0,0.0,1000.0,1000.0,1000.0
2017-10-25 20:21:00,,,,0.0026041665114462376,0.0029513889458030462,0.0027159839647553736,0.0028211805038154125,0.0039062497671693563,0.0032846339410013063,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,
7900540626655,0.0,0.0,0.0,0.0,1000.0,1000.0,1000.0
2017-10-25 20:22:00,,,,0.002662037033587694,0.3237268626689911,0.0102314817253500 22,0.0026041665114462376,0.7584635019302368,0.0193446173286065 46,0.0,4.582248210906982,0.27196903228759767,0
7902374268,0.01844617910683155,0.0242332182824611 66,0.020732060074806213,,,,43.218315124511 72,43.268951416015625,43.24262077331543,0.0,0.0,0.0,0.0,1000.0,1000.0,1000.0
2017-10-25 20:23:00,,,,0.2482638806104 66,0.31747683882713 32,0.2663645645824 5295,0.71484375,0.7532552480697632,0.7252199478068594,4.2603440284729,4.407913684844971,4.299124774286303,1.801649
74769253,4.976128578186035,4.981554031372 07,4.978457919621872,0.01844617910683155,0.0238715279847383 5,0.021578781717 1696,,,,42.465999603271484,43.21831512451172,43.04341636269 78,0.0,0.0,0.0,
2017-10-25 20:24:00,,,,0.2407986223697662 4,0.24809026718139648,0.2440785859095848,0.7141926884651184,0.7202690839767456,0.7179628560098551,4.2324938774 10889,4.280237197875977,4.258633880291
84,0.021394871648843006,4.976490020751953,4.980830669403076,4.978605052172127,0.019169561564922333,0.0235098358243703 84,0.021425523305848494,,,,41.2109375,42.43706512451172,41.6645819453869
2017-10-25 20:25:00,,,,0.2369791567325592,0.2407986223697662 4,0.23853568141830378,0.7063802480697632,0.7141926884651184,0.7101554932265446,4.2053670883178 71,4.248408317565918,4.222921807190
.020572666849555564,4.976490020751953,4.982277393341064 5,4.979427312982493 5,0.017722802236676216,0.0235098358243703 84,0.020572666849555564,,,,41.142215728759766,41.38454818725586,41.209876948
2017-10-25 20:26:00,,,,0.2350115776062011 7,0.2369212955236435,0.2359669251462160 6,0.7009548544883728,0.7078992724418 64,0.7038459141375655,4.1778793333496094,4.2122392654418945,4.190973378844
,0.020990269542750666,4.976490020751953,4.981554031372 07,4.979009692951784,0.01844617910683155,0.0235098358243703 84,0.020978008917832778,,,,41.3953971862793,42.10430908203125,41.78571707515
```

Appendix 5. Results of linear least squares regression. Cells are colored based on false alarms column. Green color indicates the absence of false alarms.

Prediction horizon (Minutes)	Historical data used (Minutes)	RMSE	Detection rate (11.04.2017 - 19.4.2017)	False alarms
480	60	163,82	0	0
480	180	25,15	0	0
480	300	10,97	0	0
480	420	6,77	0	0
480	540	5,27	0	0
480	660	4,36	0	0
2400	60	833,70	20	5
2400	180	140,35	0	0
2400	300	63,97	0	0
2400	420	37,94	0	0
2400	540	25,57	0	0
2400	660	17,40	0	0
4320	60	1498,45	36	49
4320	180	255,51	4	1
4320	300	118,49	1	0
4320	420	71,07	4	0
4320	540	48,91	0	0
4320	660	33,96	0	0
6240	60	2159,05	42	114
6240	180	369,08	8	7
6240	300	172,39	7	1
6240	420	103,92	7	1
6240	540	72,17	4	0
6240	660	50,45	3	0
8160	60	2817,75	49	229
8160	180	482,15	19	15
8160	300	226,53	18	2
8160	420	136,76	10	1
8160	540	95,46	6	0
8160	660	67,20	8	0
10080	60	3474,21	49	362
10080	180	594,17	20	21
10080	300	280,29	18	3
10080	420	169,66	16	1
10080	540	118,72	13	1
10080	660	83,95	11	0
12000	60	4109,20	53	496
12000	180	702,36	22	40
12000	300	332,77	22	6
12000	420	201,09	17	1
12000	540	142,07	14	1
12000	660	100,40	13	1
13920	60	4757,36	54	665
13920	180	812,98	23	55
13920	300	386,47	24	9
13920	420	233,51	18	2
13920	540	163,90	15	1
13920	660	116,70	15	1
15840	60	5399,34	54	839
15840	180	923,68	23	72
15840	300	439,09	25	12
15840	420	265,32	18	2
15840	540	186,48	17	1
15840	660	132,90	15	1

Appendix 6. Snapshot of a snowflake fact table, where analysis results are saved.

Row	ANALYSIS_ID	FORMMASTER_DV_ID	ANALYSIS_DATE	PREDICTED_ALARM	PREDICTED_...	STATUS_OK	STATUS_ALA...	STATUS_FAULT	TAG	ALARMS_TOT...	TAG_H	TAG_HH	TAG_L	TAG_LL	ANALYSIS_VA...
6,166	linreg	01ccc6cf5ea9d42be8693015de312fb0	2019-03-12 12:13:00.000	2019-03-14 05:42:00.000	NULL	0	1	0	TT1-235_me	3	49.00000000	55.00000000	31.00000000	29.00000000	NULL
6,167	linreg	01ccc6cf5ea9d42be8693015de312fb0	2019-03-12 12:13:00.000	NULL	NULL	0	0	0	PT5-235_me	0	14.00000000	15.00000000	3.50000000	2.50000000	NULL
6,168	linreg	01ccc6cf5ea9d42be8693015de312fb0	2019-03-12 12:13:00.000	NULL	NULL	1	0	0	PT4-235_me	0	5.40000000	5.80000000	0.50000000	0.30000000	NULL
6,169	linreg	01ccc6cf5ea9d42be8693015de312fb0	2019-03-12 12:13:00.000	NULL	NULL	1	0	0	PT1-235_me	1	1.40000000	1.50000000	0.20000000	0.10000000	NULL
6,170	linreg	01ccc6cf5ea9d42be8693015de312fb0	2019-03-12 12:13:00.000	NULL	NULL	1	0	0	PT2-235_me	1	3.50000000	4.00000000	0.50000000	0.30000000	NULL
6,171	phaseangle	01ccc6cf5ea9d42be8693015de312fb0	2019-03-12 12:13:00.000	NULL	NULL	1	0	0	PHASEANGLE	NULL	10.00000000	20.00000000	NULL	NULL	5.00000000
6,172	linreg	01ccc6cf5ea9d42be8693015de312fb0	2019-03-12 12:13:00.000	NULL	NULL	1	0	0	PT3-235_me	1	7.00000000	8.00000000	2.20000000	2.00000000	NULL
6,173	cavitation	7e6ef0ef51c6e3e79be13a18d65a40...	2019-03-12 12:00:21.000	NULL	NULL	1	0	0	CAV	NULL	50.00000000	75.00000000	NULL	NULL	0.00000000
6,174	cavitation	30ac1717cb502e3302f25d0b8e561...	2019-03-12 12:00:21.000	NULL	NULL	1	0	0	CAV	NULL	50.00000000	75.00000000	NULL	NULL	0.00000000
6,175	cavitation	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	2019-03-12 12:00:12.000	NULL	NULL	1	0	0	CAV	NULL	50.00000000	75.00000000	NULL	NULL	0.00000000
6,176	cavitation	a9214b8f2f926ce0ac8b035e59b9f854	2019-03-12 12:00:12.000	NULL	NULL	1	0	0	CAV	NULL	50.00000000	75.00000000	NULL	NULL	1.26793317
6,177	linreg	30ac1717cb502e3302f25d0b8e561...	2019-03-12 12:00:00.000	NULL	NULL	1	0	0	PT3-235_me	0	7.00000000	8.00000000	2.20000000	2.00000000	NULL
6,178	phaseangle	30ac1717cb502e3302f25d0b8e561...	2019-03-12 12:00:00.000	NULL	NULL	1	0	0	PHASEANGLE	NULL	10.00000000	20.00000000	NULL	NULL	2.00000000
6,179	linreg	30ac1717cb502e3302f25d0b8e561...	2019-03-12 12:00:00.000	NULL	NULL	1	0	0	TT1-235_me	0	55.00000000	61.00000000	29.00000000	27.00000000	NULL
6,180	linreg	30ac1717cb502e3302f25d0b8e561...	2019-03-12 12:00:00.000	NULL	NULL	1	0	0	PT2-235_me	0	3.50000000	4.00000000	0.50000000	0.30000000	NULL

Appendix 7. Snapshot of a snowflake fact table, where the results of the machine learning model are saved.

Row	ANALYSIS_DATE	PREDICTED_FAULT	HEALTH	IMPORTANT_MEASUREMENTS	FORMMASTER_DV_ID	ANALYSIS_ID
31	2019-03-08 01:00:00.000	0	0.99999770	PT1-235_me, PT2-235_me, PT4-235_me	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	Possible fault Carriage flotation oil pressure
32	2019-03-08 02:00:00.000	0	0.99999770	PT1-235_me, PT2-235_me, PT4-235_me	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	Possible fault Carriage flotation oil pressure
33	2019-03-08 03:00:00.000	0	0.99999775	PT1-235_me, PT2-235_me, PT4-235_me	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	Possible fault Carriage flotation oil pressure
34	2019-03-08 04:00:00.000	0	0.99999775	PT1-235_me, PT2-235_me, PT4-235_me	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	Possible fault Carriage flotation oil pressure
35	2019-03-08 05:00:00.000	0	0.99999775	PT1-235_me, PT2-235_me, PT4-235_me	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	Possible fault Carriage flotation oil pressure
36	2019-03-08 06:00:00.000	0	0.99999775	PT1-235_me, PT2-235_me, PT4-235_me	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	Possible fault Carriage flotation oil pressure
37	2019-03-08 07:00:00.000	0	0.99999798	PT1-235_me, PT2-235_me, PT4-235_me	c74efbfbb5eb5dfc6dbf2aa7fd4c18cd	Possible fault Carriage flotation oil pressure