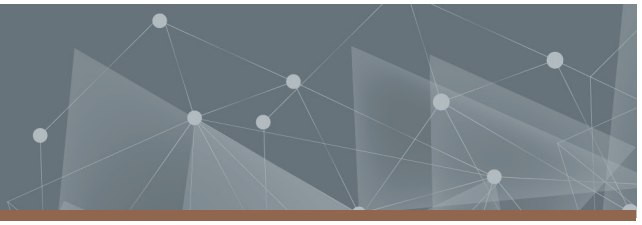
CHALMERS
UNIVERSITY OF TECHNOLOGY



# Predictive Maintenance in HVAC System utilizing Machine Learning

## A study to determine if Machine Learning is applicable

Master's thesis in Product Development & Production Engineering

IBRAHIM ABDULLE
RICHARD DANG

DEPARTMENT OF INDUSTRIAL AND MATERIAL SCIENCE
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

Master's thesis 2021

# Predictive Maintenance in HVAC System utilizing Machine Learning

A study to determine if Machine Learning is applicable

IBRAHIM ABDULLE
RICHARD DANG


CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Industrial and Materials Science
*Product Development & Production Engineering*
Chalmers University of Technology
Gothenburg, Sweden 2021

Predictive Maintenance in HVAC System utilizing machine learning
A study to determine if machine learning is applicable
IBRAHIM ABDULLE, RICHARD DANG



Supervisor: Jon Bokrantz, Chalmers University of Technology
John Wibrand, Swegon Group AB
Examiner: Anders Skoogh, Chalmers University of Technology

Master's Thesis 2021
Department of Industrial and Materials Science
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Predictive Maintenance in HVAC System utilizing machine learning
A Study to determine if machine learning is applicable
IBRAHIM ABDULLE
RICHARD DANG
Department of Industrial and Material Science
Chalmers University of Technology

# Abstract

The fourth industrial revolution is present in today's landscape of industrial engineering and digitalization has emerged to be a vital part of an organization's product portfolio. Industry 4,0 endorses companies an opportunity to make a superior informed fact-based decision. Digitalization and creating more data-driven decision making is considered to be lucrative and innovative enough to push organisations a step closer to Industry 4,0.

Swegon aims to investigate if it is feasible to implement a predictive type of maintenance to forecast when the wreckage is approaching in the HVAC systems. To guide Swegon AB closer to the ideal Industry 4,0, a current situation analysis was conducted to examine if the predictive type of maintenance is viable on Swegon current data by utilizing Machine Learning.

A collaboration of Cross-Industry Process for Data Mining (CRISP-DM) and Product Development methodologies have been utilized to prepare and create an understanding of the input data, to build the Machine Learning model based on input data from Swegon and also to measure the overall potential of the input data. An analysis of the Machine Learning model was conducted and this resulted in several recommendations for not only Swegon but every company trying to implement predictive maintenance using machine learning to continue pushing the organisation closer towards Industry 4.0 and to accomplish a predictive type of maintenance

Keywords: Predictive Maintenance, HVAC System, machine learning, wreckage

# Acknowledgements

First and foremost, we would like to express our gratitude towards Anders Skoogh who is a Professor of Production Maintenance and Director of Master's Programme in Production Engineering for his support and constructive feedback additionally the gratitude is extended to our academic supervisor Jon Bokrantz for his continuous support and guidance throughout this master thesis. His knowledge of project management within a machine learning thesis has helped us tremendously to assist to overcome several hurdles. We would also like to thank Martin Dahl who is a post-doctoral researcher at Chalmers University of Technology at the Department of Electrical Engineering for his input and valuable insight in data management in machine learning projects. We would also extend our gratitude towards Ebru Turanoglu Bekar who is a post-doctoral researcher at the Department of Industrial and Materials Science at Chalmers University of Technology for her technical guidance in machine learning.

Last but not least, we would like to express our gratitude towards our industrial supervisor John Wibrand who is a System Development Director, Digital Services at Swegon AB for his continuous support, professional guidance and for allowing us to collaborate with a global company as Swegon AB. Finally, we would also like to thank Tommy Stenkvist who is a Development Engineer at Swegon AB for his professional guidance and technical support.

Ibrahim Abdulle, Gothenburg, June 2021
Richard Dang, Gothenburg, June 2021

# Contents

# List of Figures

# List of Tables

# 1
# Introduction

This chapter will present the background, aim, limitations of this project.

## 1.1 Background

The realm of manufacturing is currently in an age of digitisation. The fourth industrial revolution, the Internet of Things and Big Data, is founded on the notion of cyber-physical systems and internet technologies employed in industrial production systems. These areas all endeavour to observe, collect and allocate data from different parts of an organization to facilitate an agile cognizant decision-making process throughout the enterprise (Drath and Horch, 2014). For the future industry to ensure strong interconnectivity between systems, the following aspects are essential: (1) The collection of data will be facilitated by the aid of smart sensors; (2) Structure and save data by utilizing IT system, and (3) Interpretation of a decision by utilizing analysis with machine learning algorithms. This collocation is described as the cyber-physical production system (Hashem et al., 2015).

Corresponding with the immense possibilities of Big data are the large significant deficits. The collection of big data can become overwhelming for the personnel responsible for the utilization of the data, hence why a good sample size of data often remains untouched. The consequences are investment costs of implementation, maintenance, and storage of data without gaining any business value (Gölzer and Fritzsche, 2017).

The solution to the predicament of Big Data along with benefiting the opportunities, machine learning can be applied. Machine learning is an exceedingly mature technology that has added value to many different fields for decades. However, within the manufacturing domain, machine learning still has a lot of unexploited potential due to the focus has been on the machine level rather than the system level. Manufacturing is a convoluted territory where the interdependence between different segments is exquisite and too complex for a human to analyse with many shifting pieces. Thus, this thesis aims at displaying an approach towards anomaly detection for one type of manufacturing system by applying machine learning to examine the manufacturing domain on a unique component level to determine if predictive maintenance is germane (Witten et al., 2016).

Swegon Group AB, owned by Investment AB Latour, listed on the Stockholm Stock

Exchange, is a market-leading supplier in the field of an indoor environment, offering solutions for ventilation, heating, cooling and climate optimisation, as well as connected services and expert technical support. Swegon has subsidiaries and distributors all over the world and 16 production plants in Europe, North America and India. The company employs more than 2 600 people and in 2019 had a turnover of 6 billion SEK.

Swegon is an indoor climate company. Swegon manufactures ventilation equipment such as an Air Handling Unit (AHU) which is the heart of the system. It is the AHU that delivers fresh and climatized air into the building. The air is transported by a centrifugal fan, which is assembled inside the AHU. Swegon has been accumulating data for a long period, especially their AHU, the PX-unit. The PX-unit is mainly used in an industrial setting especially areas where regulation and cleanliness of the airflow is required. Additionally, the PX-unit is able to remove strong fragrances from the air. Therefore, as one of their most popular product, Swegon is interested in utilizing the data available to investigate whenever a Predictive type of maintenance is viable with the current accumulation of data Swegon has obtained so far in the PX-unit.

## 1.2 Aim

This thesis aims to present a study to determine if a predictive type of maintenance is viable by using Machine Learning based on Swegon's current accumulation of data. The purpose of this aim is to familiarize Swegon with machine learning and provide valuable experience. Thus, Swegon can implement this study in other aspects within their systems to determine if predictive maintenance is applicable and any other company can use this study as a benchmark to determine if their system is ready for a predictive type of maintenance.

## 1.3 Limitations

Identifying anomalies of an industrial HVAC system will require extensive data analysis and several algorithms testing. Therefore this thesis will primarily handle the anomalies related to the dampers of the PX-unit. As the dampers are considered to be a very critical component of the PX-unit. Furthermore, regarding the root causes of the anomalies, this thesis will only handle anomalies related to product development and manufacturing process aspects. This thesis will be conducted remotely, this means no visits to Swegon's manufacturing plant. The time limit of the thesis is 16 February to 16 June.

## 1.4 Objectives

This project will investigate if optimal service intervals to deploy predictive maintenance, especially service intervals related to dampers. Achieving this objective will

result in increased utilization of the PX-unit and enable other components of the PX-unit for a predictive type of maintenance.

## 1.5 Research Questions

The following research questions have been established for this thesis.

**RQ1:** What type of machine learning model is more suitable for a predictive type of maintenance?
**RQ2:** What is required of a data-set to apply a predictive type of machine learning?

The following research question will hopefully report to Swegon if the final machine learning model reflects a predictive type of maintenance and explore general benchmark requirements of a data-set to implement this type of solution.

# 2
# Frame of reference

This chapter will present an introduction to a HVAC system and the different theories used to answer the research questions combined with different methodologies to implement machine learning. These can be split into three areas, AI & Machine Learning, Maintenances, Methods for product development & data analytics.

## 2.1 HVAC System

A heating, ventilation and air conditioning system is utilized to supply good indoor air quality and appropriate temperature. The device consists of an exhaust fan and an inlet fan which control the air circulation in the HVAC system. Dampers regulate the air velocity as well as opening and closing the dampers to control the airflow in the dampers. Filters that filter the air and cooling coils that adjust the air if a cool down or calefaction is necessary, are also included in the system.

An HVAC system usually pumps in the air from the outside into the HVAC system where the air passes by the filter detection to get filtered and to further mix with current air circulation inside the system. The coils in the system heat or cool the air when the air is passing by to achieve the requested temperature for the air channel. A fan pumps the air into the air channel where the air later travels onto the zones in the building that requests to be supplied by air. To ensure a comfortable thermal level, a reheating coil is located in each zone as well as an outlet fan pumps air out of the room to maintain the air circulating. Usually, the majority of air drained out of the room is mixed with incoming air from the outside but simultaneously the same amount of air that came into the air channels is extracted to the atmosphere (Ye and Yu. 2017). See figure 2.2 for a typical HVAC system.

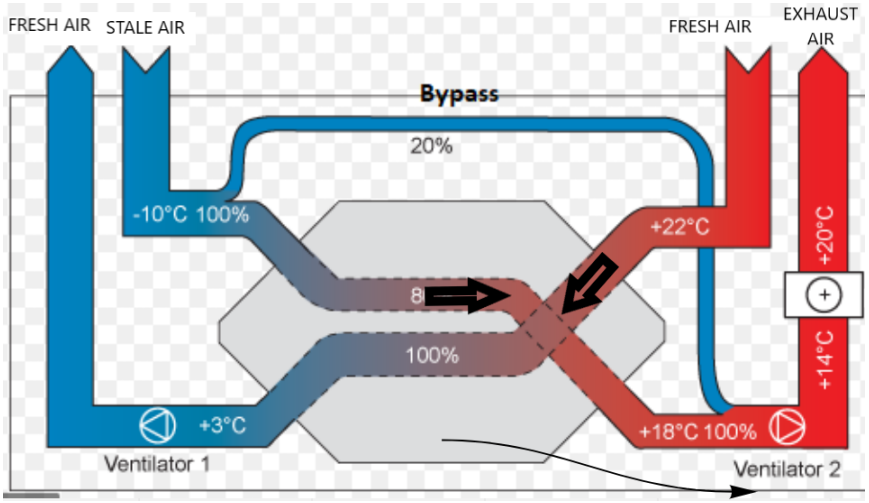



**Figure 2.1:** Figure illustrating a typical HVAC System

A crossflow heat exchanger consists of thin layers of aluminium that creates crossing air ducts. The heated exhaust air warms the lamellas and conveys heat to the colder supply air. In conclusion, a crossflow heat exchanger operates with the principle of the warm channel flows perpendicular against the colder supply air.

## 2.2 Machine Learning

Machine learning is a subsidiary of Artificial Intelligence. By computing, systems are designed to learn from data in an aspect of preparation. Through learning and improving with experience, the goal is to process a model that can be used to predict the consequences of questions based on patterns from previous learning (Bell 2014).

## 2.3 Artificial Intelligence and Machine Learning

Artificial Intelligence was first introduced to the public in 1950 when Alan Turing published the paper Computing Machinery and Intelligence, and later became notorious for participating in a competition where the goal was to create the most human-like chatbot, known as “the Turing test”. However, the duo of McCulloch and Pitts (ref) in 1944 created a computer model inspired by the function of brain neurons, aiming to mimic the intelligent model created by nature. This event is usually seen as the introduction of AI for mankind.

According to Russell and Norvig (2002, p.2), the definition of artificial intelligence can be summarized in four big categories:

1. Systems that act like humans - Premise is to create machines that perform functions that require intelligence when performed by people.
2. Systems that act rationally – Artificial intelligence’s only concern is the intelligence behaviour of the product. Hence why, computational intelligence is the framework of intelligent artificial agents.

3. Systems that think like humans – The prevailing endeavour to make computers think.
4. Systems that think rationally – By utilizing computational models, study of mental capacity is achieved.

### 2.3.1 Types of Machine Learning

There are different types of Machine Learning, but supervised learning and unsupervised learning are the 2 most common types of subareas of ML.

**Supervised Learning:** This type of ML refers to working with a set of identified training data. An input and output object is required in every training data to develop a predictive model. Furthermore, supervised learning can be categorized into *Classification* and *Regression*. Classification techniques calculate categorical responses whereas regression methods predict continuous responses (Jimenez, Juan José Montero, et al.. 2020)

The *bias-variance dilemma* is the most common issue that occurs with supervised learning. The dilemma is how the machine learning models cope with different training sets to subsequently perform meticulously. Complexity is the technique high variance models utilize to comprehend boisterous training data whilst high bias models accommodate restricted learning sets. A trade-off is presented between the two models. Additionally, it is essential to pinpoint where to settle with the trade-off and simultaneously know when to apply which of the two models.

**Unsupervised Learning:** This type of machine learning allows an algorithm to find a hidden pattern in a data set and is applied when it is not clear which type of information is going to be discovered. The principle of unsupervised learning is to run the machine learning algorithm and discover patterns and outcomes. Hence why the goal is not to find a right or wrong answer, unlike supervised learning. Clustering is a subcategory within unsupervised learning where it is applied for exploratory data analysis to find groupings in data or hidden patterns (Jimenez, Juan José Montero, et al.. 2020).

## 2.4 Application of Machine Learning

Witten et al. (2016) stated that where there is data there is the possibility to implement machine learning. Toto implements ML into business projects, an investigation is needed to determine if the data mining will be fruitful for the business. ML has received more attention in recent years due to its profitability (Witten et al. 2016). It is crucial for corporations that seek good business solutions to understand when applying Machine learning that:

1. The solution requires that the problem is based on prediction rather than usual conclusions. By using statistical averages as a bedrock, the key is to comprehend how particular characteristics of the data relate to one another.

2. The problem is efficiently isolated and not influenced from the outside. The premise is to feed the learning algorithm most of the required data from the problem. Hence, if the data is not fed adequate information when trying to predict changes by matching prior patterns in the data, the algorithm will not know how to operate further.

A good practical example of when machine learning is a good application is a Urkund database. The database employs machine learning to detect plagiarism by comparing new essays to already published essays. A bad pragmatic example is to hypothesize next year's sales in a company by having last year's sales as a reference. This is doomed to fail from the start due to the dynamic nature of business markets and new competitors always arising into space.

## 2.5 Choosing Machine Learning

Before determining if the utilization of ML as a tool to solve the problems, it is vital to consider the importance of data, the financial aspects and which problems are suited for ML. This is further discussed below.

### 2.5.1 Significance of Data

When choosing Machine Learning as a methodology, it is by default a demand that the ML models require a large amount of data to be able to make predictions and give perceptions (Hashem et al., 2015; Al-Jarrah et al., 2015). According to Witten et al. (2016), the growth in data and the requirement of big data hints that in the future, ML models will likely be applicable in more business areas, given that the knowledge of how to implement it is present. It is crucial to understand the different ML models to apply the right model with the right data. Thus, the variety of the variables and the correlations can lead to discovering patterns that can have huge business values for companies (Grover et al., 2018). Additionally, due to the nature of imperfect data, it is important to construct the algorithm in a manner to deal with the imperfections of the data, to find symmetries that are still valuable but could be vague (Witten et al., 2016).

### 2.5.2 Requirements for implementation of Machine Learning

According to Ng,(2019), a business understanding investigation is required for corporations to be involved in Machine learning projects. It is usually decided by examining if the data mining will be profitable. Lucrative ML business solutions are often found on the premise of these two following problems types:

- Solutions demand prediction instead of casual interference.
- Sufficient isolation from outside interference.

The first problem type is centred on predicting new data based on previous data rather than seeking answers in the correlation between the data. Hence why it is recommended to have a big data sample.

The second one implies that the isolation of the data is essential due to the data fed to the learning algorithm, consisting of almost everything there is to the problem. Therefore, if the factors that are sought after change, the algorithm can no longer match the new information with prior patterns due to the sudden change and the algorithm won't know what to do with the information.

Examples of business applications of ML are content recommendations on Youtube, autonomous drive, marketing, machine automation(Kourou et al., 2015; Libbrecht and Noble, 2015; Witten et al., 2016).

### 2.5.3 Financial Investment and Business Case

According to (Kashyap, 2017), to start an ML project, creating a data infrastructure with investments in the data purchase, along with pairing the right competence with the right resources to create a data culture within the company. However, data science-related projects usually come with a high risk of failure due to the nature of the projects. The demands from the projects such as multiple simulations often require a lot of time and money and the trade-off is not always positive (Larson, 2018). The majority of the big tech companies in the world devote a lot of research into AI and investments in new applications. The reason behind this is the value of big data and the financial benefits if applied correctly. A good example is Google's approach by providing open-source Machine learning tools to other organisations without compensation (AI, 2018).

### 2.5.4 Team competence for in-house development

Machine Learning projects are of the nature of technology hence why it is necessary that the members of the project possess mathematical, computational, statistical, data management and programming skills. Additionally, it is essential to have a business understanding in order to answer the reason behind a problem. A small, dedicated team is required to work in close collaboration with the product in an isolated and protected environment, in the spirit of nullifying the potential of having outside interference from organizational bureaucracy and hierarchy (Lenfle 2014), (Kashyap 2017). However, constant support from the senior management is crucial for the team's success. (Larson, 2018)

Investing in Data Scientist is considered to be a relatively new occupation due to Data Scientist being inadequate. According to Davenport and Patil, (2012), the best Data Scientists embrace big data challenges, to explore their curiosity, expand their field of expertise by deep diving into complex problems in order to test their theories. Other valuable skills such as data cleaning, organization of large data sets are precious due to data often coming in unstructured formats. According to

McAfee and Brynjolfsson (2012), the best Data Scientists are comfortable speaking in business terminology in order to support their superordinate effort to "redefine their business tasks in a language that big data can interpret".

Kashyap (2017) suggests searching internally in the company for inquisitive data engineers or developers and teach them the fundamentals of Machine Learning. Outsourcing competence does not always lead to success due to not meeting business objectives because state-of-art-techniques usually are not applicable (Mathur, 2018). Crosby (1992) implies the role of an external analyst can play a valuable role in decision making, even in bias situations such as project evaluation.

### 2.5.5 Acquisition of external expertise and vendors

Davenport and Patil, (2012) The advantages of utilizing third-party service providers are the access to Machine Learning experts and data scientists. Especially, during times when skilled personnel is limited along with the expense of hiring and the time consumption. Though, business domain expertise is required to build profitable ML solutions. Additionally, if the external competence does not acquire knowledge in the same field of expertise that is required, expected consequences are such as exchanges with business units along with time consumption to compromise for the lack of specific competence in the requested area (Ahrens, 2014).

A general perspective to consider according to Nasir and Ivanouskaya (2018), is the risk of losing control of the organization when outsourcing parts of the operation. Organizations should evade outsourcing parts of the core business. Hence why, determining a long term strategic role of AI and ML in the organisation is recommended, to recognize if outsourcing to a third-party service provider is sufficient (Burgess, 2018).

## 2.6 Machine learning requirements

With the nature of how ML models are constructed, sometimes an ML model can recognize patterns that are tedious but simultaneously come across patterns coincidentally in the particular data set of users. This type of issue can be linked to project failures, hence why the project manager/team must establish a procedure to cope with this potential problem to have continuous faith from stakeholders.

Machine learning requires a high-quality data set for the algorithm to perform at a high satisfactory level. Thus, it is a disadvantage to have insufficient amounts of data. In most cases, insufficient data is a result of data not being stored primarily, unauthorized data or data that is too expensive to obtain. Hence why a mediocre ML algorithm with the backing of big data usually outperforms a better algorithm but with an inferior data set. The conclusion to be drawn is that the biggest requirement for a machine learning model is having a big sample size of data (Witten et al., 2016).

### 2.6.1 Machine Learning solution acquistition

According to Krensky and Linden. (2016), if a company makes the choice to apply ML, it faces an acquisition decision with the following three alternatives.

- Internal development of a solution with the company's own Machine Learning team.
- Outsource developing a solution to an analytic service provider.
- The following chapters present the development of customized solutions, both in-house and outsourced.

## 2.7 Anomalies

To understand how to detect anomalies, it is important to present what an anomaly is and the characteristics of an anomaly. Anomalies are patterns in a data set that diverges from the expected behaviour in the data set. The source of an anomaly mostly depends on the data and operation (Chandola, Banerjee, and Kumar. 2009).

### 2.7.1 Anomaly Detection Methods

In this chapter, different types of detection techniques regarding anomalies will be introduced along with a discussion of their respective assets and liabilities. The denouement will be the foundation of the selections made in this thesis. Based on different machine learning algorithms, anomaly detection methods are largely categorized under the following two headings.

### 2.7.2 Supervised Learning Methods

This supervised anomaly detection method demands the existence of a characterised data set that includes both anomalous and normal data points. Under their ability to encrypt any reliance between variables and involving previous data in some forecasting model (Witten et al., 2016). The most common supervised method that will be further described below is Classification Based along with the advantages and disadvantages. See table 2.1 for advantages and disadvantages of supervised learning methods

**Table 2.1:** Table describing the advantages and disadvantages of a supervised method.

| **Advantages** | **Disadvantages** |
|---|---|
| Numerous effective algorithms | Dependent on labelled data |
| Agile in the testing phase | Requires abundant computation training time |
| Applicable in real-time | Anomaly detection is not first priority |
| Easy visualization | Prone to over-fitting |

**Classification Based:** The premise with classification-based methods is to directly label a log as either normal or anomalous (Witten et al., 2016). Usually, classifiers trained on a data set that exemplifies all possible typical behaviour without including any anomalies. If new data cannot be categorized into any one of the classified normal behaviours, inevitably it is classified as an anomaly (Jimenez, Juan José Montero, et al.. 2020). The classifier can also be trained by utilizing a data set containing both anomalous and normal data. To create a classifier able to accurately classify future logs as either anomalous or normal instantly. The downside with the majority of classification-based methods (except human decided rules-based) is that they are dependent on a large set of data, also these labelled data sets often require to be generated due to being endured in the first place. Examples of classifiers are Decision Tree, Naive Bayes Classifier, Artificial Neural Networks, and rule-based classifiers where the idea is to set rules by utilizing human expertise or machine learning.

**Regression Based:** A regression-based method is commonly used for predicting time series and model the relationship between one or several inputs, usually time as a function of a variable. The result of a regression model is the expression of future values based on input values (Witten et al., 2016). This type of method is usually used in forecasting or more generally predicting future trends of the data. Examples of regression-based algorithms are Linear Regression, Lasso Regression, Logistic Regression, etc.

### 2.7.3 Unsupervised Learning Methods

Unsupervised anomaly detection does not require any training data with manual classification (Jimenez, Juan, et al.. 2020). The majority of data are normal and only a minor percentage is abnormal data. Hence why unsupervised learning methods are based on statistical assumptions, see table 2.2 for advantages and disadvantages.

**Table 2.2:** Table describing the advantages and disadvantages of a unsupervised method.

| **Advantages** | **Disadvantages** |
|---|---|
| Unsupervised | Defining distance between points |
| Agile in the testing phase | Requires abundant computation training time |
| Applicable in real-time | Anomaly detection is not first priority |
| Easy visualization | Prone to over-fitting |

Additionally, estimation of any anomalous data would be flagged as a potential threat statistically when compared to normal data with these methods. *Clustering methods* and *Nearest Neighbour Based* are two of the most popular unsupervised methods that will be further described below, along with the advantages and disadvantages as well.

**Clustering Methods:** This type of anomaly detection method compares new logs with defined normal data instances labelled as clusters. It is identified that there are

two ways to approach anomalies with cluster analysis. The first one is to cluster the data set as well as analysing the frequency of each cluster. Consequently, normal logs are considered to be a part of the frequent large data set whilst abnormal logs are located in inadequate clusters. The second approach is to assume the distinction that normal logs are located more closely to the centre than anomalous logs that are assumed to be located further from the centre (Liu and Deng 2021). Concerning testing, clustering analysis is considered the fastest one in the testing phase compared to other unsupervised learning methods.

**K-Means Clustering:** The K-Means anomaly detection method attempts to split the data sets into a selected number of k clusters. The k is called centroids, the main gist of the centroids is to attempt to centre itself in the middle of the k clusters selected. As soon as the centroids are stabilized, the K-Means clustering method is finished. K-means clustering can be evaluated through the inertia value and using the elbow technique. The inertia value can be described as the variance based on the number of clusters, and the elbow technique is used to select the recommended amount of clusters for a specific data set (Liu and Deng. 2021).

The elbow technique method is based to find the specific point where the variance drastically levels out and remains largely unchanged based on the number of clusters. This specific point corresponds to the recommended amount of clusters, see figure 2.1 for an example.

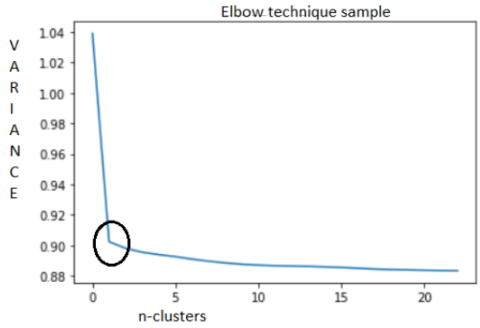



**Figure 2.2:** Figure illustrating how to select the amount of clusters based on the inertia value

## 2.8 Maintenance Strategies

Machine failures are a common occurrence and at the same time very costly in every production system. Every failure requires maintenance to fix the machine failures and to subsequently increase the utilization of the machines and whilst decreasing the number of downtimes, but every production system resorts to many different types of a strategic approach to elevate the problem. The most common strategic approaches can be categorized into four different strategies (Gackowiec. 2019).

- Preventive Maintenance
- Corrective Maintenance

- Condition Based Maintenance
- Predictive Maintenance

**Preventative Maintenance:** This strategic approach is to have planned repairs regularly over a specific timeframe, usually required to collect statistical data of failures to provide an estimation of the breakdown (Gackowiec, Paulina. 2019). Compared to Reactive Maintenance, this approach does not take into consideration that the component might break in between the specific timeframes or is still working as intended and does not require any maintenance. This results in increased maintenance costs compared to predictive maintenance but at a lower investment cost.

**Corrective Maintenance:** Corrective maintenance can be defined as performing maintenance mainly when a breakdown occurs (Gackowiec. 2019). Corrective maintenance can be considered the most simple strategy, however quite expensive in the long term due to being mainly used as a reaction towards breakdown, and all consequences as a result.

**Condition Based Maintenance:** Similar to Preventative, this strategy emphasizes more on keeping a real-time status of the system to decide when maintenance is required (Gackowiec. 2019). This is usually done when key parameters of the system reach an unacceptable level or when performance starts to degrade as a result of faulty tools.

**Predictive Maintenance:** The main purpose of this strategic approach is to monitor the key parameters affecting the performance of the components in real-time, identifying key areas where parameters are affected negatively and areas where the parameters are working as intended (Okoh, et al.. 2016). Combining this with sensors that measure the key parameters, a pattern will be discovered when the failures usually occur. Compared to Condition Based Maintenance, this strategy is able to predict future failures in advance. Finally, this strategy aims to implement maintenance just before this failure will occur based on the pattern discovered. This type of strategy has the potential to decrease downtimes, cut maintenance costs and increase the lifespan of the component. However, this type of strategy is considered to be very complex and will require high investment and extensive data volume.

## 2.9 CRISP-DM

To tackle the anomalies, Swegon is having a predicament, the chosen methodology about its relevance to the problem is CRISP-DM. The method is a common working approach within machine learning.

The Cross-Industry Standard Process for Data Mining (CRISP-DM) is a process model utilized to characterize data mining (Piatetsky, 2014). The process model was established in the 1900s and consists of six different phases with the purpose to facilitate and deliver business value (Wirth and Hipp, 2000). The six high-

level phases are the following: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment, see figure 2.3. To guarantee the success of the project, it is vital to shift between the different phases in the process model.

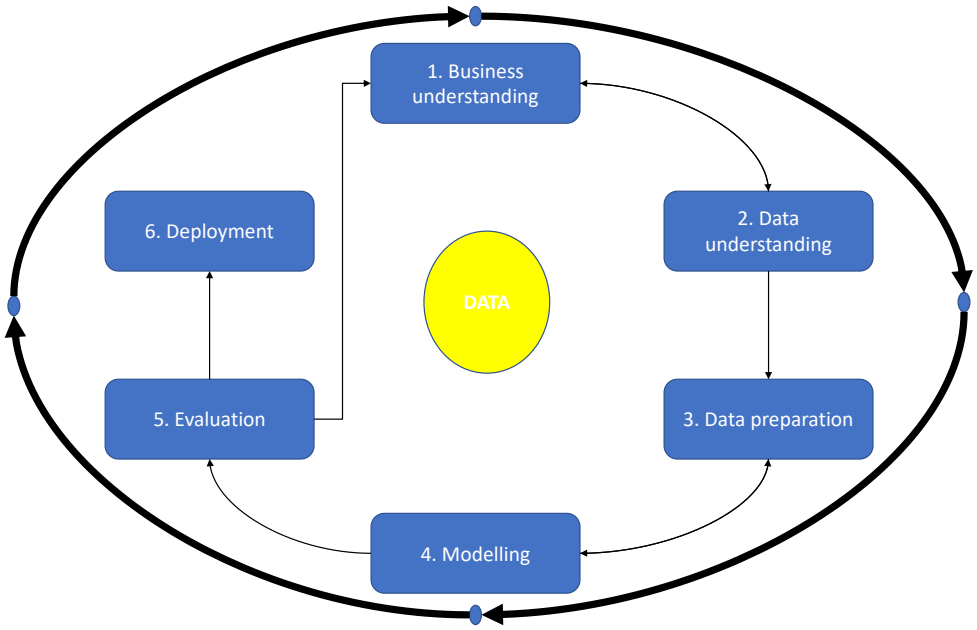



**Figure 2.3:** An illustration similar to the process of CRISP-DM (Wirth and Hipp, 2000.)

1. Business understanding: This first step is about defining the requirements and project goals from a broad business perspective. Usually, a project plan is established, followed by the derivation of specific data mining goals, evaluating the situation, and deciding the required tasks to reach the business objectives.
2. Data understanding: This phase aims to gather the initial data and identify the quality of the data by analysing, to understand the underlying structure to accomplish the project goals.
3. Data preparation: This phase is the most crucial part of this process model as it covers approximately 80% of the entire project. This part builds the final data set by following these steps, data selection, reduction, reprocessing and data cleaning.
4. Modelling: This part consists of the selection and application of different models or different model techniques on the data. Besides, calibration of improving parameters is essential in this phase, which is the shortest phase of the process.
5. Evaluation: Concerning the initial business and data mining objectives, evaluation is essential whether a model can be utilized or requires to be polished. The consequence of this phase will culminate in deciding which model aligns best with the objectives.
6. Deployment: The goal of this phase is to develop a deployment plan to implement all the required steps of the proposed model in production. In addition,

this step also requires a maintenance strategy as well as a monitoring process plan and review of the process model.

## 2.10 Product development methods

This chapter explains the theory behind product development methods when selecting concepts without including the traditional process steps for product development such as product planning, market research and concept generation. Due to the already established concept in the form of a machine learning algorithm model, the application from the traditional process step for product development is not necessary but the selection of the right method will be vital. Hence why the product development methodology for concept selection is the most objective choice.

### 2.10.1 Requirement specification

A requirements specification is conducted after the customer needs have been identified so that the concept generation has a goal frame to relate to. Additionally, own requirements and wishes are presented as well as the customers and other stakeholders. There are five different categories where all requirements and wishes are divided, the categories are; Function, Performance, Strength, Economy and Maintenance and Environment. These categories also constitute a foundation for the selection matrices created later in the project. Some of the requirements and wishes are measurable and have target values to be achieved.
The wishes that are asked should be weighted to examine which ones require the most focus and which ones are the most favourable for the product. The weighting is done using a weight determination matrix and earlier experiences. In the matrix, all wishes are compared with each other and the wish that is considered to be of greater weight gets one point while what is judged to be less advantageous gets zero points. A half points are awarded for both requests if they are considered equivalent. The points that each wish receives is summed in rows and divided by the maximum number of points in the matrix. (Johannesson 2013).

After the criteria have been weighted, they should be written on a scale of 1-3, where 3 is the one that weighs the most and 1 is considered to be the least important. This is done with the equation:

$$w_i = (\sigma_i / \sigma i_{max}) w_{imax} \tag{2.1}$$

where $\sigma_i$ is the number that obtained by dividing the sum by the total and $\sigma_{imax}$ is the largest of these figures. $w_i$ is the weighting and $w_{imax}$ is the maximum value on the scale.

After the requirement specification list is conducted, in order to determine which model has fulfilled the requirements, the model faces an elimination and evaluation phase. The purpose is to objectively decide which model has fulfilled the requirements which concepts has not by applying evaluation method within product devel-

opment.

The evaluation consists of three phases in form of an elimination matrix, Pugh matrix and Kesselring matrix. An elimination matrix is a method to make a first draft screening of models created in the model generation phase(Johannesson, 2011). A Pugh Matrix aims to screen out the number of concepts and further improve the models if required. (Ulrich and Eppinger, 2008). The Kesselring matrix operates in the same way as the Pugh matrix although the difference is that it is more time consuming than the Pugh matrix. The benefit of the Kesselring matrix in contrast to the Pugh matrix is that the results from the Kesselring matrix reflect an impartial comparison as well as eliminate unachievable concepts that the Pugh matrix is unable to do. (Ulrich and Eppinger, 2008)

# 3
# Methodology

The following chapter presents a detailed overview of the methodology applied to the thesis. The thesis has focused on utilizing CRISP-DM and product development methods to describe how the problem will be managed. The flowchart in figure 3.1 presents an overview of the approach to the problem.

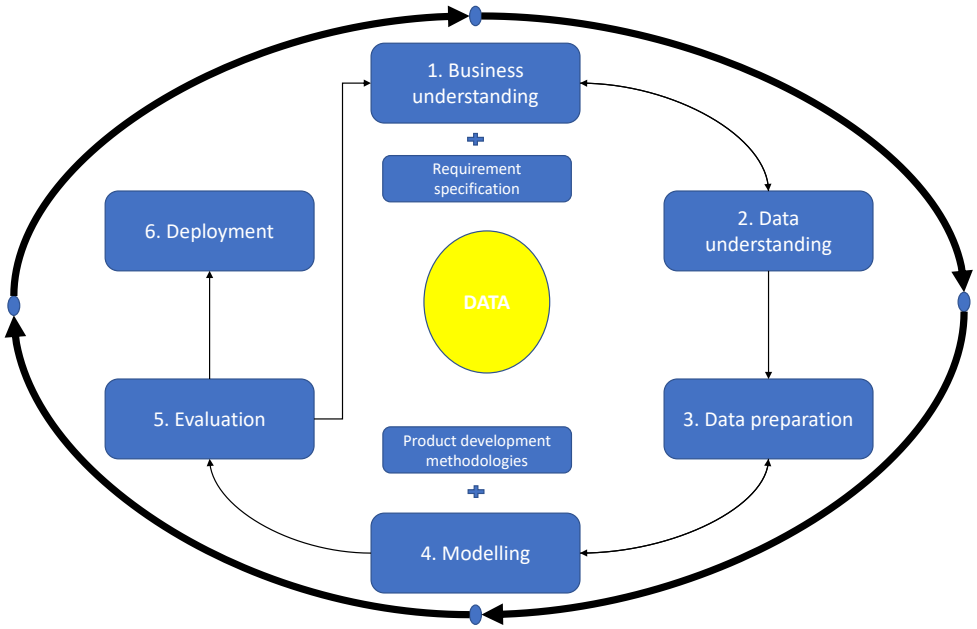



**Figure 3.1:** A diagram depicting the flow of the thesis methodology, combining CRISP-DM (Wirth and Hipp, 2000) with product development methodology.

## 3.1 Business Understanding

Business understanding is about creating a broader understanding of the thesis's requirements and objectives. Usually, it includes establishing a project plan, determining data mining goals, assessing the situation, and determining business objectives for the thesis.

¨ In this instance, the priority is to define the problem definition together with Swegon to understand the business requirements and conclude with assessing the situation and which methodology is most suitable and conduct workshops and semi-structured interviews to gather opinions, perceptions, background information and expectations of machine learning. The semi-structured interview was done by following the guidelines set by (Harrell, Margaret and Bradley. 2009), Some of the questions were: what are your expectations of machine learning?, how often do you collect data?, what is predictive maintenance for Swegon?

To accomplish this, a specific data mining goal was established in combination with workshops and meetings including Swegon. This resulted in required data mining tasks to reach the established business objectives.

To get a broader view regarding Swegon's requirements with Machine Learning, a mature problem definition is created. Good mature problem definition often has a time perspective and an ML perspective to find the respective tools to accomplish this. Bearing in mind that the company desires to have the framework predictive machine learning solution.

## 3.2 Requirement Specification

The requirements specification for the development of the models is established from the requirements of the business understanding along with the demands from the CRISP-DM model. The requirements are divided into the categories data, performance, and model.

The requirement specification has been developed throughout the process. A preliminary version was generated early in the process and with time it was updated due to new knowledge as well as new demands from Swegon. The requirement specification was applied as guidance during the whole development process to evaluate models and the data. This approach was to ensure the requirements were abided by.

This was done by creating a requirements specification on three key areas: Data, Model, Performance. The requirements were created based on available theory and the set business objective. This will lead to a comparison between the model and the requirement specification to perform product development methods for evaluation

to decide the best algorithm models.

## 3.3 Data Understanding

Data understanding consists of obtaining access to data, explorations of the data by applying visualization methods to better understand the data. This will determine the overall ability of the data to implement Machine Learning.

### 3.3.1 Collection of initial data

This step was done to trace the method Swegon uses to extract raw data from the PX-unit into a common computer. This was done by conducting workshops with a development engineer representing Swegon. In hindsight, this method aims to replicate Swegons method to extract data. This leads to a collection of the data in this thesis to be similar to Swegon method to collect data.

### 3.3.2 Data Description

Data description was conducted to get a wider understanding of the data, more specifically more insight towards the chosen parameters. Swegon has an extensive library of available data within the bin files. The data format and the data background was found based on Swegon's library of data information and further details regarding the parameters were easily accessed through the development engineer within Swegon.

### 3.3.3 Data visualization

Visualization of the specific parameters was performed to obtain better clarity of the data and highlight specific intervals where the data is valuable. In this case, an alarm was visualized because of the business understanding to find the interval from start to alarm. Additionally, the relevant data was visualized to find possible correlations between the parameters.

### 3.3.4 Data quality report

A Data quality report was done to determine the overall quality of the data based on the set business objectives and to filter the unwanted data sets.

A data quality report can usually contain several attributes of the data-sets depending on the business objective. Based on a study of a Data quality Model in machine learning with several attributes (Rudraraju. and Boyanapally. 2019), attributes were selected which matches the business objective and requirement specification. This data quality report was based on grading the following aspects in conjunction with the requirement specification:

- Accuracy

- Completeness
- Relevancy
- Size
- Traceability

## 3.4 Data preparation

Several parameters required preprocessing to increase the quality. The Alarm parameters' original state is stored in a 16-bit binary value with several different types of alarm. Based on the business understanding, this model will only handle Damper specific alarms. Therefore, all alarms unrelated to dampers were removed from the data sets.

Since the Damper parameters are dependent to trigger the alarm, a new parameter was created to combine the two parameters into one. This new parameter "Delta1" and "Delta2" improves the accessibility and creates an easier understanding of the Dampers function based on the temperature and humidity.

Finally, the data sets were split up into several different data sets to enable clustering analysis on every available data sets.

### 3.4.1 Feature extraction

The intended purpose of feature extraction is to explore options to improve the performance by extracting features from the existing data sets. In this type of project, to reduce the dimension of the data and at the same time keep the majority of the distribution. The MinMaxScaler in Python is a powerful tool used to normalize the different parameters into one identical range of values but still keeping the distribution of the specific parameter.

The feature extraction will greatly improve the quality of the data by transforming the data to have better distribution, removing linear dependencies, etc and is deemed crucial to increase the data quality (Singh and Singh. 2019).

## 3.5 Model

This model is built and generated in Python, utilizing libraries such as Pandas, Numpy, scikit, etc. This model was first built by performing an unsupervised learning, k-means clustering to generate clusters to label the data-sets into three labels 0-2, 0 indicates low risk towards failure and 2 indicates high risk towards failure.

Afterwards, the results of the clusters and data sets are combined into several data sets to test the available algorithms. One regression algorithm and three classification algorithms were tested.

- Logistic Regression
- Linear Discrimination Analysis
- K Neighbors Classifier
- Decision Tree Classifier

## 3.6 Evaluation Methods

This project will emphasize using product development methods to evaluate the different algorithm models based on accuracy and deviation to select the most optimal algorithm model and the clusters were evaluated by a built-in tool to plot the inertia value and then using the elbow technique to approximately find the optimal amount of clusters.

## 3.7 Deployment

Usually, deployment consists of monitoring the model in a real-life environment, but this thesis will exclude the deployment step as the model is not implemented and will not be reviewed based on real-life performance.

# 4
# Results

In this chapter, the results from the model are presented and the evaluation process to decide if the machine learning model is applicable.

## 4.1 Business Understanding

Based on the information gathered by meetings and workshops the data mining goals and mature problem definition was defined. This case considers the prediction of alarm failures and the root causes of the dampers jamming based on the PX-unit.

The following sub-goal was created in conjunction with Swegon based on the business understanding: “Analysis of the correlation between the behaviour of Temperature combined with Humidity of the Exhaust Air with the Damper using data from the sensors.

To get a clear goal of the following mature problem definition was established:
“I want to predict the remaining useful life of the PX-unit’s Damper in regards to the exhaust air moisture regulator. I need a classification model that can predict the life span of the PX-unit’s Damper by displaying the probability of failure of the unit in its current state.

### 4.1.1 Requirements established from business understanding

The requirements and wishes were established primarily from the business understanding chapter in the CRISP-DM along with the needs from Swegon AB, see figure 4.1. It was evident that a big data sample is necessary to have a foundation to build the model. Hence, the data collection needs to be relevant and suitable to the mature problem definition in the business understanding. Additionally, the data set must consist of the occurrence of alarms otherwise a vital element is missing. The desires regarding the data are based on the data being more plausible.

| **Swegon** | | **Document type** | **Requirement specification** | | | | |
|---|---|---|---|---|---|---|---|
| | | ***Project*** | Master thesis | | | | |
| ***Issuer: Richard & Ibrahim*** | | | ***Created: 2021-04-21*** | | | | |
| | | | ***Modified*** | | | | |
| | Criteria | | ***Aim value*** | ***R/W*** | ***Weight*** | ***Verification Method*** | ***Reference*** |
| **1.** | **Data** | | | | | | |
| | 1.1 | Large Sample Size | > 365 bin files | R | | Data understanding | Theory |
| | | *Large Sample Size* | > 750 bin files | W | 1 | | |
| | 1.2 | Data collected for extended period | > 365 days in a row | R | | Data understanding | Group + Theory |
| | 1.3 | *Data relevant with objective* | | R | | Data vizualization | Theory |
| | 1.4 | *Past maintenance and service history* | | W | 2 | Swegon | Group |
| | | *Sufficient isolation from outside interference* | | R | | Data Understanding | Theory |
| | 1.5 | *Data set must include relevant alarms* | | R | | Data understanding | Group |
| **2.** | **Model** | | | | | | |
| | 2.1 | Clustering | | R | | Modelling | Group |
| | | Clusters | 3 or 4 | R | | Modelling | Group |
| | 2.2 | Clustering accuracy | > recommend 3 or 4 clusters | R | | Modelling | Group |
| | 2.3 | Classification model | | R | | Modelling | Group |
| | 2.4 | *Classification Accuracy* | >85% | R | | Modelling | Group |
| **3.** | **Performance** | | | | | | |
| | 3.1 | Script runtime | < 1 minute | W | 3 | Model Evaluation | Swegon |
| | 3.2 | User friendly format | | W | 1 | Model Evaluation | Swegon |

**Figure 4.1:** The established requirement specification. Weight is within a scale of 1-3.

The performance of the requested classification model is set to at least predict 85% of the remaining useful lifespan of the PX-damper by displaying how many hours

until debris is present, given the current production load. The model should also predict failure given a time window along with the root cause for failure. If a specification of the range of time to failure is developed, it will bring value to the model.

The correlation between alarms and relevant parameters is indispensable since if there is no correlation between them, the model can not operate efficiently. The prediction rate of the model will determine how applicable the model can be to the requirements from Swegon. If the model could be applicable in different fields in the company, the solution could be more lucrative for the company than anticipated.

## 4.2 Data Understanding

This chapter will present the results from data understanding following the subchapters from the methodology.

### 4.2.1 Collection of initial data

The collection of the data sets was initiated by extracting a bin file obtained by Swegon. The bin files contain several data and the data is structured by three categories, Header, Metadata, Raw-Data.

**Table 4.1:** Table describing the different categories within a bin file.

| | |
|---|---|
| Header | File-version: MAC-address, SW version. |
| Metadata | Information about parameters: Name, unit, scale. |
| Raw-Data | Parameter-ID, Value of Parameter-ID |

The bin file was extracted by the software "Sympathy for data". The software allows the user to decide what type of data to extract. Applying Sympathy for data, the bin files from Swegon were extracted and structured accordingly to the chosen options in a .csv format. One bin file contains one day of data accumulation. The user can decide to extract every second/minute/hour of the file.
In total, data were collected from three PX-units and two different periods from each PX-unit were extracted, which resulted in a total of six CSV files with $\tilde{8}0$ 000 rows each.

### 4.2.2 Data description

The six data-sets collected can be described as following:

**Table 4.2:** This table describes the different data sets.

| CSV File | Row Counts | Size |
|---|---|---|
| VBG99a | 82 081 | 5 419 KB |
| VBG99b | 82 171 | 4 565 KB |
| VBG00a | 174 241 | 10 211 KB |
| VBG00b | 168 731 | 9 391 KB |
| VBG01a | 74 881 | 4 107 KB |
| VBG01b | 73 441 | 4 401 KB |

Every data set contains approximately 2-3 months of collected data and a unique number represents a PX-unit but at a different geographic position.
And every data-set contains the following parameters:

Dampers [%]
- IO3_ID3_AI1_PX_Damper1FB
- IO3_ID3_AO1_PX_Damper1
- IO3_ID3_AI2_PX_Damper2FB
- IO3_ID3_AO2_PX_Damper2

The Dampers are being measured by a sensor detecting the opening percentage, 100% indicates the Dampers are fully open whereas 0% indicates Dampers are fully closed. One PX-Unit consists of 2 Dampers. Parameters with “FB” correlates to the real-time value of the Damper, whereas parameters without “FB” correlates to the expected value of the Damper.

Moisture status in PX-Unit
- HM_HX_TEMP [C]
- HM_HX_HUMIDITY [%]

HM_HX is a moisture sensor located at the exhaust-air. This sensor measures the temperature and the relative humidity within the exhaust-air area of the PX-unit

Occurrence of alarm
- AL_Triggered14

This output includes all of the different types of alarm of the PX-unit and this output is stored as a 16-bit binary value. Where the 10th and 9th position counted from the right corresponds to a Damper failure. For example, 0000 0011 0000 0000 indicates an alarm has occurred on both dampers.

The condition for a damper failure is based on the values of the parameters IO3_ID3_AI1_PX_Damper1FB and IO3_ID3_AO1_PX_Damper1. Thus the

alarm is triggered if the values correspond to this conditional statement during 10 minutes.

$$AL\_Triggered14 = 1 \quad if \begin{cases} IO3\_ID3\_AI1\_PX\_Damper1FB < 85\% \ or \\ IO3\_ID3\_AO1\_PX\_Damper1 < 3\% \ and \\ IO3\_ID3\_AI1\_PX\_Damper1FB > 15\%. \\ Same\ handling\ for\ Damper2. \end{cases} \tag{4.1}$$

Simplified, this statement checks on the value of the feedback value signal and the expected value signal. If the difference between these two parameters exceeds certain values during a period of 10 minutes, the alarm will trigger.

### 4.2.3 Data visualization

If an alarm occurred around 40 000 minutes, this indicates that the interval between 0 til 40 000 minutes is important, see figure 4.2. Since after the alarm the machine will break and not continue running until fixed. Hence, this example indicates that every data set should only contain values from start to alarm. See appendix I to VI for more alarm plots.

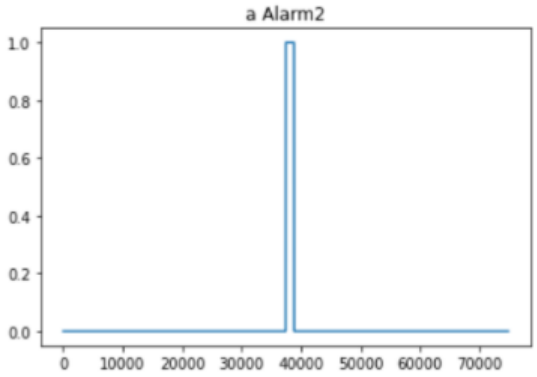



**Figure 4.2:** This example shows the moment the alarm occurs within the data set. Y-axis = Alarm, X-axis row count (minutes).

Furthermore, the different parameters were plotted towards each other to find any relevance. This resulted in temperature and humidity not always having a relevance towards the dampers failures see appendix VII - XII. See appendix VII & VIII for plots where the parameters are relevant.

Finally, every data set except VBG00a and VBG01b shows faulty behaviour from the start of the data collection till the alarm, see appendix XIII. For the opposite behaviour see appendix XIIII.

### 4.2.4 Data quality report

The data quality is analysed based on the Accuracy, Completeness, Relevancy, Size and Traceability in the data set, see figure 4.3. This data quality report considers the business objective and requirement specification.

| Analysis dimension | Description | Checklist | | | | | |
|---|---|---|---|---|---|---|---|
| | | VBG99a | VBG99b | VBG00a | VBG00b | VBG01a | VBG01b |
| Accuracy | No missing values | x | x | x | x | x | x |
| Completeness | Not missing the critical values | | | x | | x | |
| Relevancy | The data is applicable based on the objective | | | x | | x | |
| Size | Amount of data is sufficient | | | | | | |
| Traceability | The raw data extraction is free of error | | | | | | |

**Figure 4.3:** Evaluation of the data quality in several analysis dimension

Every data set except VBG00a and VBG01a fails on relevancy since no correlation between temperature combined with humidity with Delta 1 & 2 could be determined. None of the data sets passed the size criteria due to not being sufficient enough to meet the requirement specification.

Another point to regard is Swegon's approach to the extraction of the data invites the risk of collecting insignificant data due to the lack of a filter to remove unwarranted data and the given parameters did not measure the time since the previous breakdown. Swegon's approach to gathering information inside an SD-card by applying a sensor to collect information every second, minute and hour, without filtering out excessive information is one of the reasons the data quality is considered poor and therefore requires extensive preprocessing and is the reason every data set fails on traceability.

Every data set except VBG00a and VBG01a failed on completeness since the alarm is dependent on the damper output, the data sets must include several working dampers before the alarm occurs. However, some data sets did not include any working dampers before the alarm occurred. Hence, lowering the overall quality of

completeness.

Additionally, every data set failed to pass the requirement specification due to data size and data collection over an extended period being insufficient. Thus, only VBG00a and VBG01a will be investigated further to find the potential of the data set and no further product development methodologies will be implemented.

## 4.3 Data preparation

A total of 22 unrelated damper alarms were found in the chosen data sets and were subsequently removed.

Thus, the current format of the data set includes Delta1 and HM_HX_TEMP into one data-set and respectively Delta1 and HM_HX_HUMIDITY, same format handling for Delta2 as well. See table 4.3 for a sample data set.

This leads to a new alarm statement based on:

$$12\% < Delta < 100\% \; During\ a\ period\ of\ 10\ minutes. \tag{4.2}$$

**Table 4.3:** VBG00a sample structure

| Delta1 | Delta2 | HM_HX_HUMIDITY | HM_HX_TEMP |
|---|---|---|---|
| 2 | 2 | 52.04 | 22.79 |
| 2 | 2 | 52.02 | 22.80 |

To prepare for the clustering, the data sets were split up into four different data sets and each containing:

1. Delta1 & Temperature
2. Delta1 & Humidity
3. Delta2 & Temperature
4. Delta2 & Humidity

This resulted in two data sets VBG00a and VBG01a being transformed into 8 total data sets each containing Delta1 & Temperature, Delta1 & Humidity, etc.

Finally, the MinMaxScaler resulted in a scale of 0-1 in the x-axis, y-axis to prepare for clustering.

## 4.4 Model

Firstly, k-means clustering was performed on 8 data sets to group every data set into three clusters, 0-2, see figure 4.4, see appendix VII - XII for other clusters.

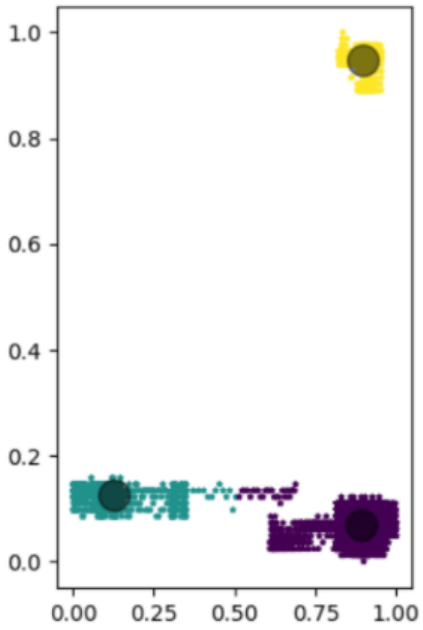



**Figure 4.4:** An example of clusters of VBG00a. Green = 0, purple = 1, Yellow = 2.

Afterwards, the different data sets with the same parameters were combined into a total of four data sets including the clusters to increase the overall size of the PX-unit data set, the results of the combined data sets will be presented in table 4.4.

**Table 4.4:** Describing the amount of clusters in the combined data sets.

| Data set | Cluster | Instances |
|---|---|---|
| Delta1 & Temperature | 0 | 146911 |
| | 1 | 82868 |
| | 2 | 25221 |
| Delta1 & Humidity | 0 | 96322 |
| | 1 | 101870 |
| | 2 | 1808 |
| Delta2 & Temperature | 0 | 75621 |
| | 1 | 104282 |
| | 2 | 20097 |
| Delta2 & Humidity | 0 | 140337 |
| | 1 | 55694 |
| | 2 | 3969 |

The model was built to classify a grade from 0-2, where 0 equals low risk and 2

equals high risk. Based on the input values of temperature or humidity, the model will give an output value of 0-2. See table 4.5 for a clearer depiction of every model input values and output.

**Table 4.5:** Describing the models input and output.

| Model | Input | Output |
|---|---|---|
| 1 | Delta1 & Temperature | Risk factor 0-2 |
| 2 | Delta1 & Humidity | Risk factor 0-2 |
| 3 | Delta2 & Temperature | Risk factor 0-2 |
| 4 | Delta2 & Humidity | Risk factor 0-2 |

## 4.5 Evaluation of Model

The clusters were evaluated by plotting the inertia value, see appendix X for a plot of the inertia values on every model.

Evaluating the difference between the inertia value and the number of clusters combined with the elbow technique results in a recommended amount of clusters. The following table represents the recommended amount of clusters in every model:

**Table 4.6:** Result of the elbow technique.

| Model | Recommended amount of clusters |
|---|---|
| 1 | 3 |
| 2 | 2 |
| 3 | 1 |
| 4 | 1 |

This cluster evaluation indirectly indicates, model 3 and model 4 clusters were inaccurate since three clusters were wanted. Therefore the classification model of model 3 and 4 should be considered as unreliable.

The model was evaluated by testing several different algorithms, the following tables 4.7-4.10 will be presented and then evaluated using the requirement specification to see if any of the algorithms of every model will pass the requirement specification in terms of data, model, performance aspects.

**Table 4.7:** Model 1 Algorithms.

| Algorithm | Accuracy [%] | Deviation |
|---|---|---|
| Linear Regression | 73.5 | 0.0026 |
| Linear Discrimination Analysis | 62.6 | 0.0032 |
| K Neighbors Classifier | 89.7 | 0.0021 |
| Decision Tree Classifier | 89.3 | 0.0021 |

**Table 4.8:** Model 2 Algorithms.

| Algorithm | Accuracy [%] | Deviation |
|---|---|---|
| Linear Regression | 91.7 | 0.0014 |
| Linear Discrimination Analysis | 91.4 | 0.0015 |
| K Neighbors Classifier | 99.2 | 0.0061 |
| Decision Tree Classifier | 99.3 | 0.0053 |

**Table 4.9:** Model 3 Algorithms.

| Algorithm | Accuracy [%] | Deviation |
|---|---|---|
| Linear Regression | 58.0 | 0.00041 |
| Linear Discrimination Analysis | 57.8 | 0.00048 |
| K Neighbors Classifier | 80.5 | 0.0022 |
| Decision Tree Classifier | 78.1 | 0.0020 |

**Table 4.10:** Model 4 Algorithms.

| Algorithm | Accuracy [%] | Deviation |
|---|---|---|
| Linear Regression | 54.2 | 0.0018 |
| Linear Discrimination Analysis | 54.2 | 0.0020 |
| K Neighbors Classifier | 61.2 | 0.0032 |
| Decision Tree Classifier | 67.6 | 0.0023 |

The Linear Regression was tested to see any feasibility of a regression type model but the accuracy was low compared to the other algorithms. The other three algorithms were of a classification type and based on model 1 & model 2, the algorithm K Neighbors Classifier had the best Accuracy. As previously mentioned, the results from model 3 & model 4 does not represent a valid model as the results of a classification model is solely based on the previously performed clustering.

# 5
# Discussion

This chapter will discuss CRISP-DM methodology, data aspects, model results.

## 5.1 Analysis of results

This chapter presents an analysis of the research conducted by comparing empirical data with theory, requirement specification and business objectives.

### 5.1.1 Relationship between business understanding and data understanding

The business understanding was built upon establishing a script to predict PX-damper problems based on gathered data from old damper problems. The positive outcome would be that Swegon will have a reliable basis on how to handle future PX-damper problems. To achieve the goal a requirement list was conducted whereas the most vital requirement was a large relevant sample size of data (Witten et al.. 2016). The correlation between business understanding and data understanding is important due to being dependent on each other.

After the requirements list was conducted and the first requirement was not fulfilled due to the lack of relevant data to accomplish the objectives of the business understanding, the solution was to seek if it was viable to either compensate the business understanding of the data understanding with questions like.

- How can the objectives with the business understanding be changed if it does not fulfil the requirements of the data understanding?
- How can the data be modified in order to fulfil the objectives of the business understanding?
- Are there other business problems the available gathered data can be applicable with?

The answers to these questions were unfortunately answered as soon as the data collected from Swegon did not match the requirements of their business objectives. The requirement list was based on Swegon's business objectives and a high percentage classification model is not possible with the current data. The question that arose was.

- What can be achieved with the available data?

A classification model was conducted with the available data and with the results of accuracy of 50%<, the conclusion was simply that the model is not trustworthy to determine and predict the potential failure rate of a PX-damper in the future.

### 5.1.2 Relationship between requirements and data understanding

The requirement list foundation was based on the determined business objectives. Thus, this entailed that the collected data had requisites in order to achieve the business objectives. The demand for the gathered data is also based on the foundations of machine learning which is the requirement of accumulating big relevant data. As stated by Ng,(2019), profitable machine learning business solutions are often found on obtaining a big data sample as well as having sufficient isolations from outside interference.

The obtained data was big but not relevant in some regard due to not being sufficiently isolated. The isolation is essential because the data fed into the learning algorithm should consist of nearly everything there is to the problem. Otherwise, if factors that are being pursued change, the accuracy of the model decreases, Ng, (2019). The collected data suffered from not being sufficiently isolated because of the data not only presenting alarms from the PX-damper. Occasionally, combinations of alarms with no associations with the PX-damper were discovered. The immediate consequences were that the data quality, as well as the potential accuracy of the machine learning model, sank. Additionally, this led to automatically disqualifying any potential machine learning models due to the foundation of a good machine learning model requiring a relevant sample size of data along with being efficiently isolated.

Bearing in mind that the requirement of relevant clean data is essential, further consequences arose such as not implementing product development methods to examine which machine learning model is the most suitable for the business objectives. The reason is because of the data failing in the previous phase before the implementation of product development methods.

The following questions surfaced to find a middle ground between the requirements and the data understanding:

- Can the requirement specification be modified in order to achieve the business objectives?

The answer is no because of the conflict of interest between the business objectives and the fundamentals of a profitable machine learning model. The demands on the data can not be compromised, otherwise, the business objectives would be compromised as well as the efficiency of the machine learning model.

## 5.2 Data discussion

As portrayed in a previous study by Popenici & Kerr (2017), it was presented that established technology companies are investing in machine learning research and contributing their findings for free to support the advancement of machine learning. This implies that data is the main concern for machine learning projects, not the machine learning models. The business objectives were consistently in the view of understanding that available data is crucially important as well as the importance of possessing large relevant data samples. As stated in earlier studies by Witten et al. (2016) and Clemmedsson (2018), machine learning models require a large amount of data to operate but unfortunately, the data collection is inadequate in the real world. Evident consequences are the necessity for special treatment in the design of machine learning models, data creation, and data management. Hence why, a strong guideline from the industry is that Big Data is defined by its volume, variety, and velocity. This should be considered to be the most important obligation before launching an machine learning project, Clemmedsson (2018).

It was discovered that a big issue with the data quality is that it was not considered to be an asset, leading to not storing data efficiently and sufficiently when the problem is resolved. In a broader perspective, not many people in organizations understand the potential value of their data which leads to not storing data proficiently with filters and indulgence.

Even though considering data as a valuable asset is important, it cannot be determined beforehand what is valuable data and what is not before. As stated by Witten et al. (2016) if there is accessible data, there is a possibility of applying machine learning. Hence why the quality of the data is the centre of attention.

A discovered mistake which is a common mistake when being inexperienced with machine learning is starting to work with algorithms and models without ensuring the quality & quantity of the data is adequate. The mistake can both come from the management above in an organisation that can demand a model without understanding the fundamentals of machine learning or inexperience in the personnel handling the data. Usually, when people learn to use machine learning, they often handle complete and perfect data sets, trusting that machine learning is mostly about creating models and algorithms. Instead, 90% of the effort is about working with the data by labelling, creating, and maintaining the data. The data requires evaluation and depending on the evaluation, the data can demand to be labelled, new data is required, or the algorithm necessitates to be adjusted to be able to strongly manage the defects in the data. Though, it should be considered to always test the algorithm in the early stages to achieve a current status rapport.

As presented in the results, no produced model matched the requirements established from the business objectives sufficiently. It is believed that the fundamental cause is not authorizing a goal with the data collection. As stated by Passi and Barocas, (2019) machine learning requires clear and specific definitions that refer to

something measurable to be able to predict specific outcomes or quality of interest. Hence why it is crucial to establish a goal by having a solid business understanding before setting the goal. By collecting data without a goal, how is the machine learning model supposed to have a goal since the data reflect the model?

The thought process behind the development of the models VBG99B and VBG01B was to visualize for the company the potential with machine learning when data is handled correctly. The reason VBG99B and VBG01B were successful in comparison to the rest of the models is because of possessing accuracy, completeness, and relevancy. But failed due to lack of size in the data as well as the raw data was not free of error. To increase the sufficiency of the data size, a combination of model VBG00a and VBG01b was established. Why the model was combined was because of the models possessing the same prerequisites and the graphs showed the same attributes, ee appendix x. Further, the other models could not be combined with VGB00a and VBG01a with regard to plots, see appendix x that shows the conditions were not the same because of reasons such as geographical differences that could be a source of error, as well as no valuable information, could have been gathered because of the time span of the collection of data is faulty due to PX-unit being broken from the start. Thus, it does not matter if the alarm is alerted.

Another discovery and reason why the data was considered to be insufficient were that the alarm data is not trustworthy due to a delay between when the PX-unit is malfunctioning and when the alarm alerts. The potential consequences are that it could affect the accuracy of the clustering model due to these conditions.

## 5.3 Does the model reflect predictive maintenance?

The model was built as a classification model based on the unsupervised clustering method and this was performed to measure the potential of the current data sets. To incorporate the product development methodology in the data quality report and the model evaluation the requirement specification was created based on theory and the business objectives. In product development, the requirement specification is seen as the foundation. If no model is able to meet the demands of the requirement specification no further evaluation can be conducted since the other methods such as elimination matrix and Kesselring matrix required at least one model to pass the requirement specification. Therefore, these presented models in the results should only reflect the possible potential of the data sets.

Another aspect to consider is this model is a classification model based on unsupervised clustering, this type of model does not directly represent a predictive type of maintenance. The output of this model only classifies a risk factor of 0-2, whereas a predictive type of maintenance is intended to find the specific time frame to perform maintenance or a real-time update on the machine status. This leads to uncertainties if the current model truly reflects a predictive type of maintenance.

## 5.4 Reflection on the chosen methodology

The CRISP-DM methodology was used throughout the course of this thesis. The main reason CRISP-DM was a good option for this thesis was due to the fact that CRISP-DM was seen as a method easier to follow as other methods such as Knowledge Discovery Databases (KDD) are widely used by data mining experts (Shafique, Umair, and Qaiser. 2014). Additionally, since this thesis was done in the Department of Industrial and Material Science further decreased the option to use KDD. Furthermore, CRISP-DM also has elements of product development in the process, therefore CRISP-DM was used throughout the course of this thesis.

The chosen methodology for model selection was to implement product development methodologies in order to objectively choose the right model (Johannesson, 2011). Unfortunately, this step was never viable as soon as it was discovered that the gathered data and the business understanding had a mismatch. Due to the nature of these methods, the vulnerability of being dependent on the requirement specification list became evident here. The data did not pass the most vital requirements which means the procedure can not be progressed to the next stage, which is implementing an elimination matrix, Pugh matrix and finally a Kesselring matrix to present the final concept. This was considered to be a limitation with the chosen methodology due to being dependent on the previous stage of the process, but the foundation of the methods is built on the requirement specification list.

What could have been done differently is a question that is difficult to answer because of the developing situation of not being able to implement the methods? Thus, CRISP-DM in combination with product development methods was the right approach to this thesis even though it is an unusual combination, but the results could have been fruitful based on the knowledge of these concept selection methods has been successful methods within product development (Johannesson, 2011).

## 5.5 Sustainability

Sustainability is the ability to endure in a relatively ongoing way across various domains of life (Paul. 2010). Sustainability development consists of three dimensions: social sustainability, economic sustainability, and ecological sustainability. The dimension of ecological sustainability is based on knowledge of what strain humans and ecosystems can withstand and how we can avoid damage. This chapter presents how this thesis approaches sustainability.

When putting this thesis into the context above, implementing a predictive type of maintenance by applying machine learning is sustainable. The economical benefits are evident, due to predicting wreckage which is expensive for companies. Thus, it will reduce maintenance cost. The environmental benefits will be less travel to service the AHU which will lead to not wasting energy. The social aspect is that machine learning can benefit the day-to-day operation by planning in relation to

the prediction of alarm. This will lead to structuring the workday efficiently and increasing the satisfaction level of the job.

Any company intending to pursue a predictive type of maintenance using machine learning can expect several sustainable benefits to continue striving towards Industry 4.0. These are following:

- Lower maintenance costs
- Less unexpected failures
- Less reserve parts inventory
- Increased utilization
- Increased life cycle

## 5.6 Research Questions

The following research questions have been established for this thesis to answer in this chapter.

**RQ1:** What type of machine learning model is more suitable for a predictive type of maintenance?

A regression type model is more suitable for a predictive type of maintenance due to the main point being to monitor key parameters affecting the system. Thus, implementing maintenance based on the output given is the most suitable type of machine learning for a predictive type of maintenance based on the research in this thesis.

**RQ2:** What is required of a data-set to apply a predictive type of machine learning?

The results presented that what is required of the dataset in order to apply a predictive type of machine learning is to first and foremost match the business objective to the raw data collection. This means establishing a goal with what type of data is being gathered. Introducing filters to the raw data collection is an approach to further cement a firm foundation for the application of predictive machine learning. The last requirement discovered from the results is collecting data within the critical time frame is the approach to increase the quality of completeness and the size of data.

# 6
# Recommendation

This chapter presents the necessary steps Swegon is recommended to apply to meet their business objectives in machine learning projects, this can also be used for other companies trying to begin exploring predictive maintenance using machine learning.

## 6.1 Data management

To answer RQ2 more in detail, the following statements regarding data should be considered.

**Match the business objective to the raw data collection:** The established requirement specification list that was based on the business objectives had a mismatch with the data Swegon collected. Hence why the data segment in the requirement specification list did not match the goals of the business objective and resulted in no realistic model that could be generated except the model that was created to envision the small potential in the data when applying machine learning. The purpose of the project was to determine if it was viable to implement predictive type machine learning and the results presented how crucial it is to invest in establishing data culture with strategic data collection before attempting to apply machine learning in order to automate processes. A feasible explanation could be the current AI hype that inflates the expectations on machine learning and a seeming pressure to rapidly react to evolving technologies.

**Introduce filters to the raw data collection:** Investments in systematic data collection based on clear business objectives are essential to building a firm foundation in the application of predictive machine learning. The next step is to collect the data based on the business objectives to determine what type of data to gather and filter unwanted data that could potentially contaminate the quality of the data. For example, if a temperature failure is detected in an HVAC system at the exhaust area, the data of interest is solely in the exhaust area, not the entire HVAC system. The reason why this is important is to maintain isolation from outside interference such as data from the rest of the HVAC system that could contaminate the quality of data. Thus, introducing filters into the data collection will increase the data quality to ensure the business objectives will be achieved.

**Collect data within critical time frames:** To increase the quality of the completeness and size of the data, a complete time frame is required. The data needs

to be collected during a specific time frame, this time frame should be between the moment a damper related breakdown is fixed to the next moment a damper related alarm occurs.

**Outsource machine learning projects:** If a company has a strong desire to apply machine learning, the idea of having an in-house machine learning team applied machine learning projects is suggested. To enhance the journey, it has been efficiently proven to outsource machine learning, in the beginning, to build momentum and captivate machine learning talent to develop an in-house team (Davenport and Patil, 2012). If outsourcing is not a possibility, Lenfle (2014) recommends creating a small, dedicated machine learning team for in-house development in an isolated environment with the support from the senior management which is important for the results the team can produce (Larson, 2018). It is also essential to understand that implementation of machine learning is time-consuming and should be not expected to give quick and effective results.

## 6.2 Model

To answer RQ1 more in detail, the following paragraph will expand and give more information regarding the answer to RQ1.

**The future model should consider Delta 1 & Delta 2 based on time instead of temperature and humidity.** This recommendation is heavily dependent on the quality of the data in terms of completeness and size.

To properly become a predictive type of maintenance a regression model is recommended. **A regression model** based on Delta and time will present the Delta's performance based on time and provide an output of the remaining time until an alarm occurs based on the values of Delta, additionally, the future model can even include temperature over time and humidity over time to increase the models capacity. The main purpose to switch to a time-related parameter instead of temperature or humidity is sole because the time-related parameter will take several other aspects into consideration and is less complex than taking several other parameters into consideration. This model can also be considered as a more holistic approach to Damper related issues.

Thus answering RQ1, a regression type model is recommended for a predictive type of maintenance. As a predictive maintenance main point is to monitor key parameters affecting the system and implement maintenance based on the output given.

# 7
# Conclusion

The purpose of the thesis was to determine if a Predictive Type of maintenance is viable by using machine learning based on Swegon's current data. The purpose of this aim is to familiarize Swegon with machine learning and provide valuable experience.

One of the discoveries from this study implied that digital transformations have strong relevancy for machine learning projects. The majority of machine learning projects are currently executed as the drivers of the digital transformation which could be seen as perilous. Establishing a data culture with strategic data collection before applying machine learning is a lucrative option for the business objectives and the right approach when applying machine learning.

Implementing machine learning is considered to be a crucial stepping point for companies striving towards Industry 4.0. This type of solution goes well into the digitalization aspect of Industry 4.0 considering incorporating Data-Driven Decision Making as a basis of machine learning. This goes without saying implementing machine learning to predictive maintenance is very complex and difficult and requires several iterations of CRISP-DM to fully accomplish predictive maintenance. This thesis comprises of one iteration to evaluate the current data-set to find its potential and give feedback in terms of recommendation on how to improve the current data-set into a fully-fledged data-set applicable towards predictive maintenance. Combining the frame of reference, data quality report and the model's results, the potential of the data sets does not reflect the set business objective. Several steps are required to increase the quality of the data set to create a model to accomplish the business objectives. These steps are described as recommendations which Swegon or any company should take into consideration before implementing a predictive type of maintenance using machine learning. These recommendations are split into two categories, data and model aspects.

Model
- Regression model to predict future breakdowns
- Time-based model results in a more holistic approach

Data
- Match business objective with raw data collection
- Implement several filters into the raw data collection
- Collect data within the critical time frames
- Outsource machine learning projects to gain experience

# References

AI, Google (2018). Publication database. url: https://ai.google/research/pubs/ (visited on 08/05/2021).

Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." ACM computing surveys (CSUR) 41.3 (2009): 1-58.

Clemmedsson, Elin. "Identifying Pitfalls in Machine Learning Implementation Projects: A Case Study of Four Technology-Intensive Organizations." (2018).

Davenport, Thomas H. and D. J. Patil. "Data Scientist: The Sexiest Job Of the 21st Century." (2012) In: Harvard Business Review 90.10, pp. 70–76.

Drath, Rainer, and Alexander Horch. "Industrie 4.0: Hit or hype?[industry forum]." IEEE industrial electronics magazine 8.2 (2014): 56-58.

Gackowiec, Paulina. "General overview of maintenance strategies–concepts and approaches." Multidisciplinary Aspects of Production Engineering 2.1 (2019): 126-139.

Grover, Varun, et al. "Creating strategic business value from big data analytics: A research framework." Journal of Management Information Systems 35.2 (2018): 388-423.

Gölzer, Philipp, and Albrecht Fritzsche. "Data-driven operations management: organisational implications of the digital transformation in industrial practice." Production Planning & Control 28.16 (2017): 1332-1343.

Harrell, Margaret C., and Melissa A. Bradley. Data collection methods. Semi-structured interviews and focus groups. Rand National Defense Research Inst santa monica ca, (2009).

Hashem, Ibrahim Abaker Targio, et al. "The rise of "big data" on cloud computing: Review and open research issues." Information systems 47 (2015): 98-115.

James, Paul. Urban sustainability in theory and practice: circles of sustainability. Routledge, 2014.

Jimenez, Juan José Montero, et al. "Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics." Journal of Manufacturing Systems 56 (2020): 539-557.

Johannesson, Hans, Jan-Gunnar Persson, and Dennis Pettersson Produktutveckling. "Effektiva metoder för konstruktion och design." Liber, andra upplagan (2013).

Krensky, Peter and Linden, Alexander. (2016). Machine-Learning and Data Science Solutions: Build, Buy or Outsource? url: https://www.gartner.com/en/documents/3531217 (visited on 05/07/2021).

Kasey, Panetta. (2018). 5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018. url: https://www.gartner.com/smarterwithgartner/5–trendsemerge-in-gartner-hype-cycle-for-emerging-technologies-2018/ (visited on 04/05/2021).

Kashyap, Patanjali. "Do not forget me: the human side of machine learning." Machine learning for decision makers. Apress, Berkeley, CA, (2017). 281-314.

Kourou, Konstantina, et al. "Machine learning applications in cancer prognosis and prediction." Computational and structural biotechnology journal 13 (2015): 8-17..

Larson, Deanne. "Exploring communication success factors in data science and analytics projects." The ISM Journal of International Business 2.2 (2018): 30-38.

Lenfle, Sylvain. "Toward a genealogy of project management: Sidewinder and the management of exploratory projects." International Journal of Project Management 32.6 (2014): 921-931.

Lenfle, Sylvain. "Floating in space? On the strangeness of exploratory projects." Project Management Journal 47.2 (2016): 47-61.

Liu, Fan, and Yong Deng. "Determine the number of unknown targets in Open World based on Elbow method." IEEE Transactions on Fuzzy Systems (2020).

Lwakatare, Lucy Ellen, et al. "A taxonomy of software engineering challenges for machine learning systems: An empirical investigation." International Conference on Agile Software Development. Springer, Cham, (2019).

McAfee, Andrew, et al. "Big data: the management revolution." Harvard business review 90.10 (2012): 60-68.

Nasir, Muhammad Omer, and I. V. Ivanouskaya. The role of outsourcing in modern business practices. Phd thesis. (2018).

Nitesh Varma Rudraraju, Nitesh, and Varun Varun Boyanapally. "Data Quality Model for Machine learning." (2019).

Okoh, Caxton, Rajkumar Roy, and Jorn Mehnen. "Predictive maintenance modelling for through-life engineering services." Procedia CIRP 59 (2017): 196-201.

Russell, Stuart, and Peter Norvig. "Artificial intelligence: a modern approach." (2002).

Shafique, Umair, and Haseeb Qaiser. "A comparative study of data mining process models (KDD, CRISP-DM and SEMMA)." International Journal of Innovation and Scientific Research 12.1 (2014): 217-222.

Singh, Dalwinder, and Birmohan Singh. "Investigating the impact of data normalization on classification performance." Applied Soft Computing 97 (2020): 105524.

Ulrich, K. T., and S. D. Eppinger. "Product Design and Development–McGraw-Hill Irwin." New York, NY (2011)..

Wirth, Rüdiger, and Jochen Hipp. "CRISP-DM: Towards a standard process model for data mining." Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. Vol. 1. London, UK: Springer-Verlag, (2000).

Witten, Ian H., and Eibe Frank. "Data mining: practical machine learning tools and techniques with Java implementations." Acm Sigmod Record 31.1 (2002): 76-77.

Yao, Ye, and Yuebin Yu. Modeling and control in air-conditioning systems. Springer Berlin Heidelberg, (2017).

# A

# Appendix I

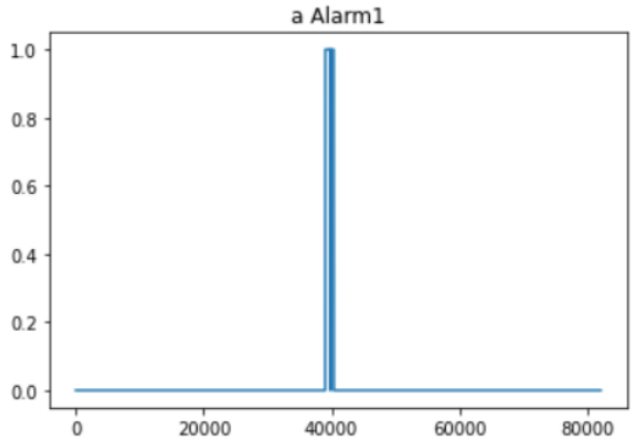



**Figure A.1:** VBG99a Alarms, X-axis = row counts(minutes) Y-axis = alarm

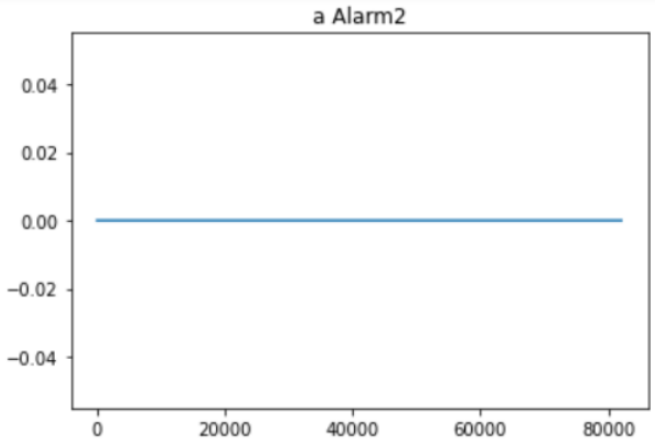



**Figure A.2:** VBG99a Alarms, X-axis = row counts(minutes) Y-axis = alarm

# B

# Appendix II

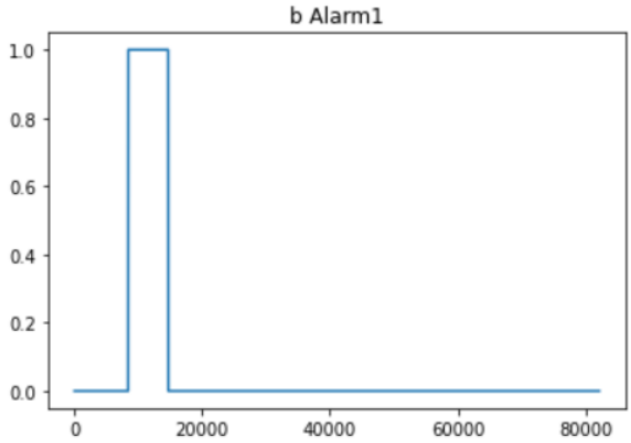



**Figure B.1:** VBG99b Alarms, X-axis = row counts(minutes) Y-axis = alarm

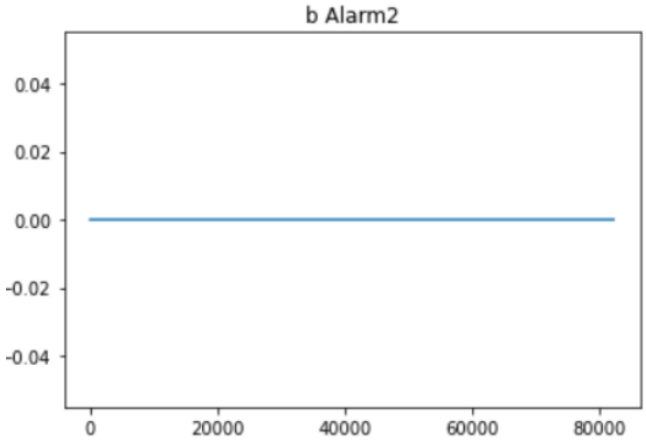



**Figure B.2:** VBG99b Alarms, X-axis = row counts(minutes) Y-axis = alarm

# C

# Appendix III

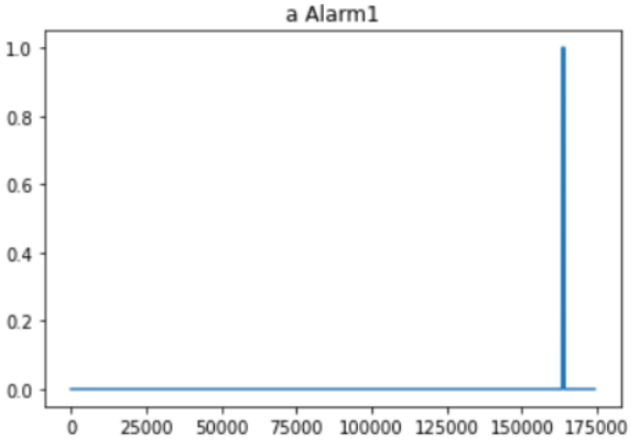



**Figure C.1:** VBG00a Alarms, X-axis = row counts(minutes) Y-axis = alarm

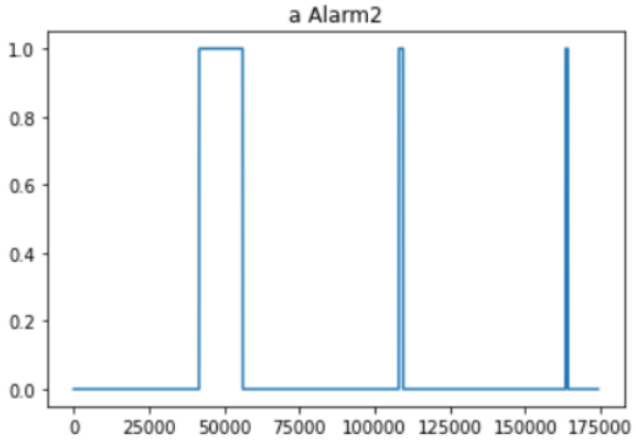



**Figure C.2:** VBG00a Alarms, X-axis = row counts(minutes) Y-axis = alarm

# D

# Appendix IIII

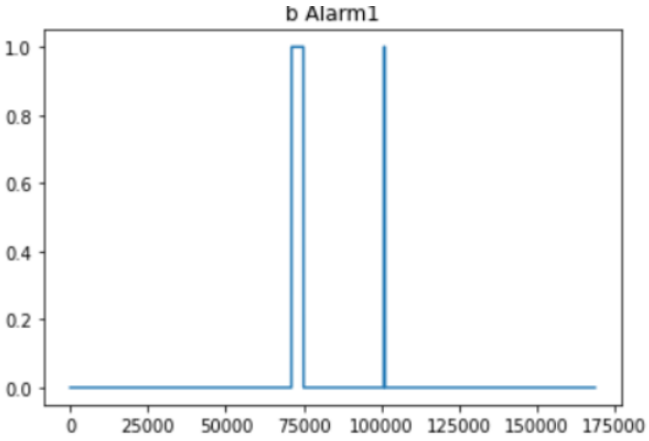



**Figure D.1:** VBG00b Alarms, X-axis = row counts(minutes) Y-axis = alarm

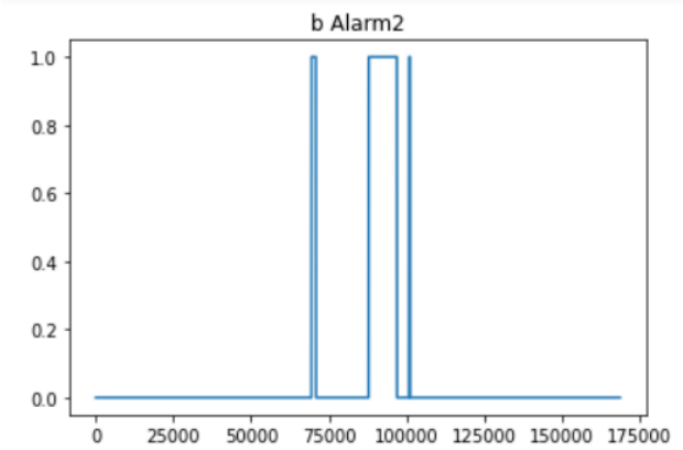



**Figure D.2:** VBG00b Alarms, X-axis = row counts(minutes) Y-axis = alarm

# E

# Appendix V

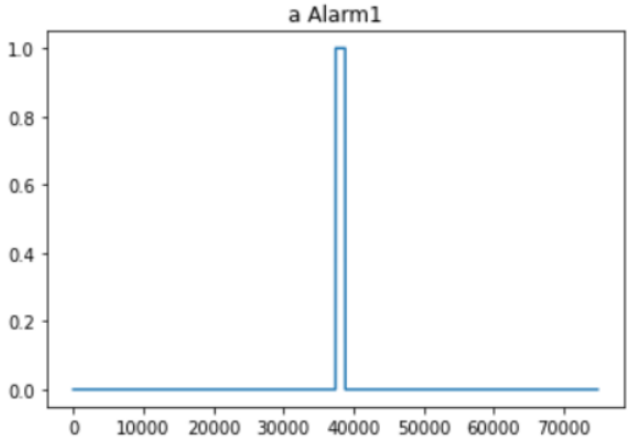



**Figure E.1:** VBG01a Alarms, X-axis = row counts(minutes) Y-axis = alarm

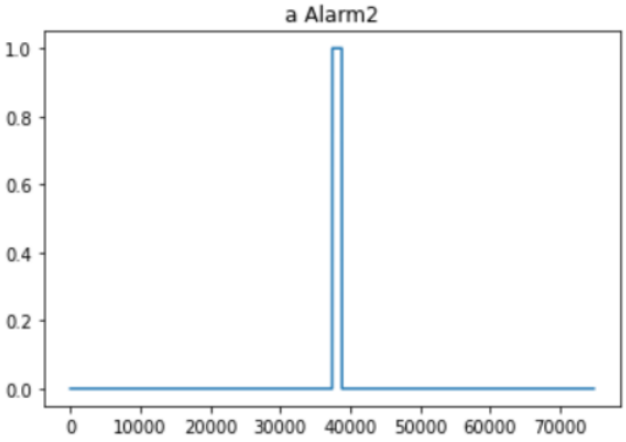



**Figure E.2:** VBG00a Alarms, X-axis = row counts(minutes) Y-axis = alarm

# F

# Appendix VI

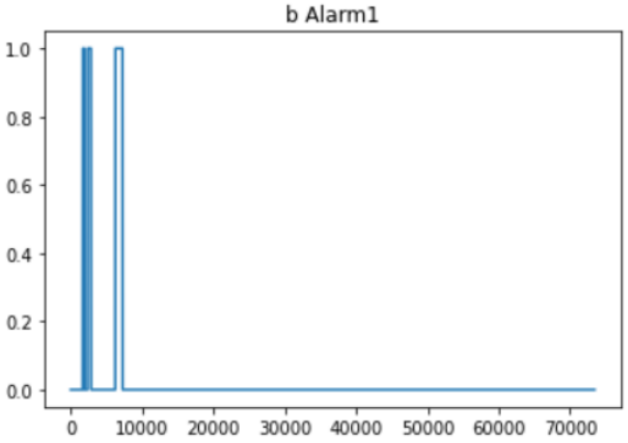



**Figure F.1:** VBG01b Alarms, X-axis = row counts(minutes) Y-axis = alarm

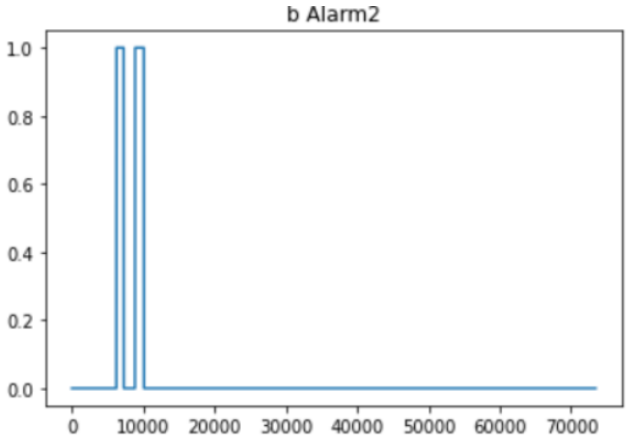



**Figure F.2:** VBG01b Alarms, X-axis = row counts(minutes) Y-axis = alarm

# G

# Appendix VII

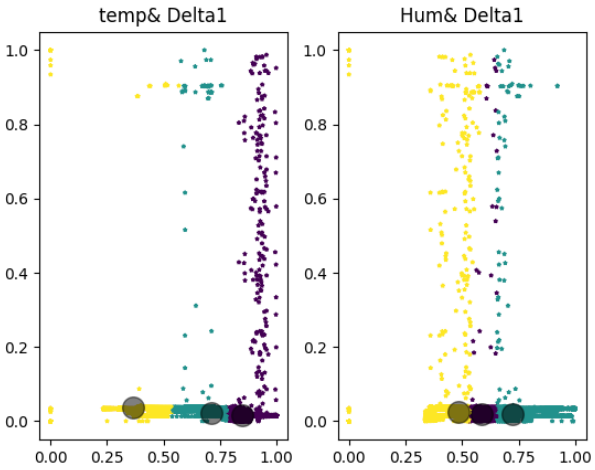



**Figure G.1:** VBG00a Plots, X-axis = Temp Y-axis = Delta

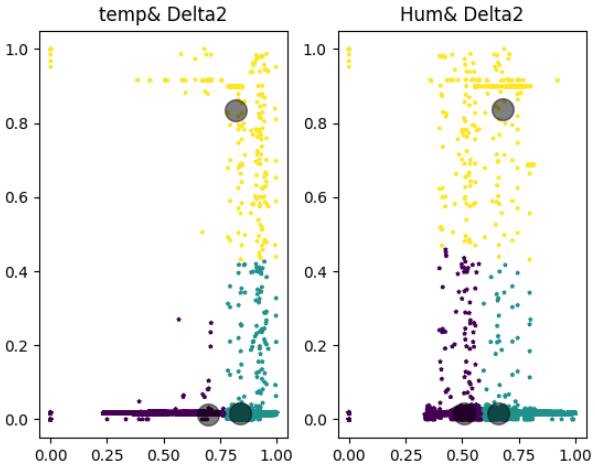



**Figure G.2:** VBG00a Plots, X-axis = Temp Y-axis = Delta

# H

# Appendix VIII

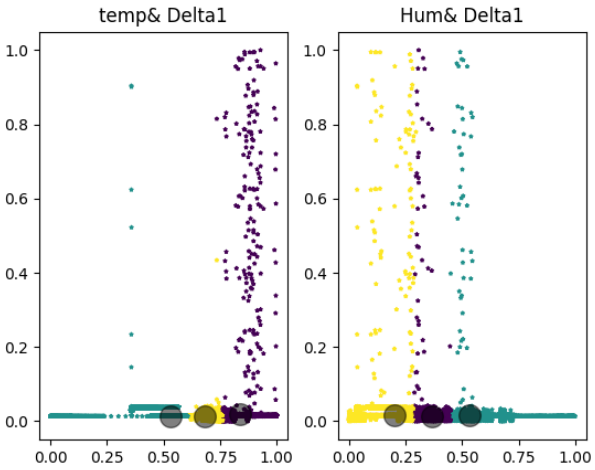



**Figure H.1:** VBG00b Plots, X-axis = Temp Y-axis = Delta

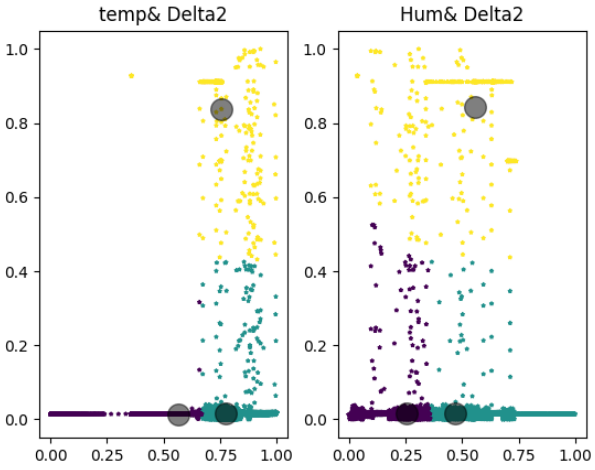



**Figure H.2:** VBG00b Plots, X-axis = Temp Y-axis = Delta

# I

# Appendix VIIII

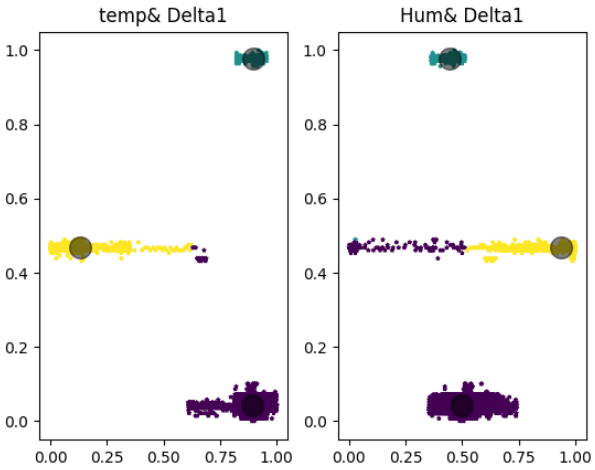



**Figure I.1:** VBG01a Plots, X-axis = Temp Y-axis = Delta

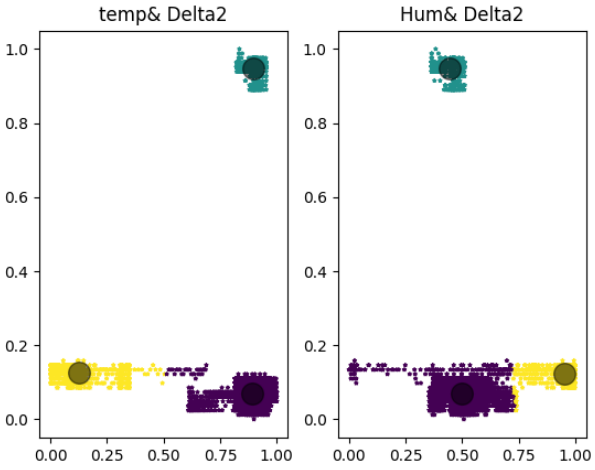



**Figure I.2:** VBG01a Plots, X-axis = Temp Y-axis = Delta

# J

# Appendix X

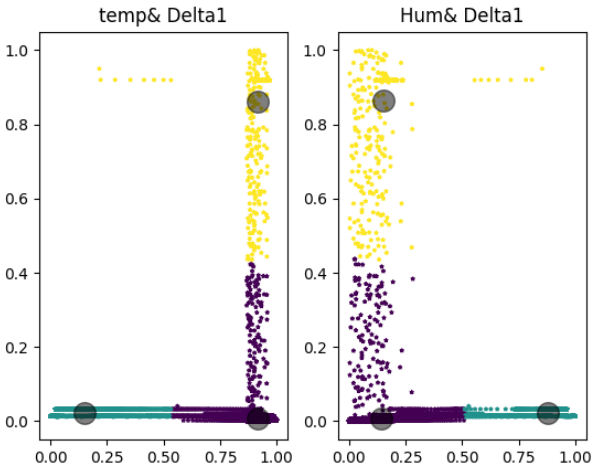



**Figure J.1:** VBG01b Plots, X-axis = Temp Y-axis = Delta

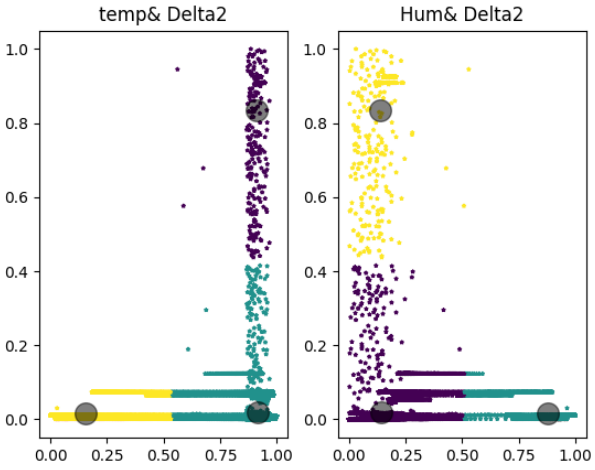



**Figure J.2:** VBG01b Plots, X-axis = Temp Y-axis = Delta

# K

# Appendix XI

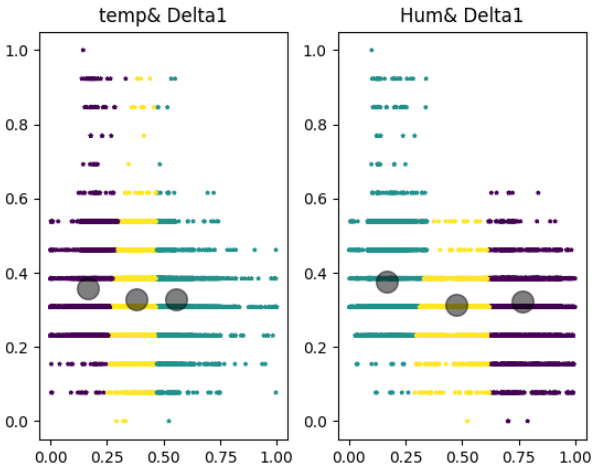



**Figure K.1:** VBG99a Plots, X-axis = Temp Y-axis = Delta

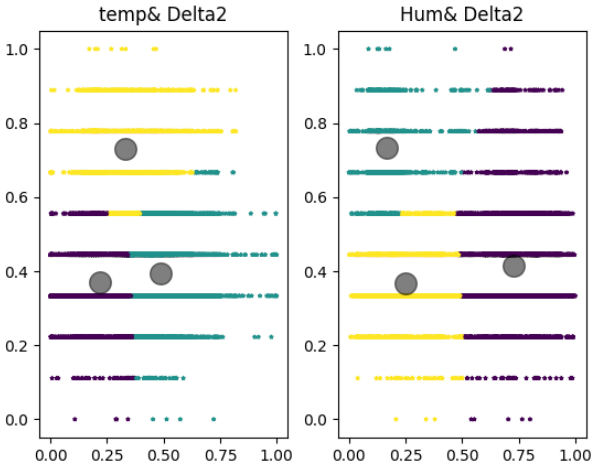



**Figure K.2:** VBG99a Plots, X-axis = Temp Y-axis = Delta

# L

# Appendix XII

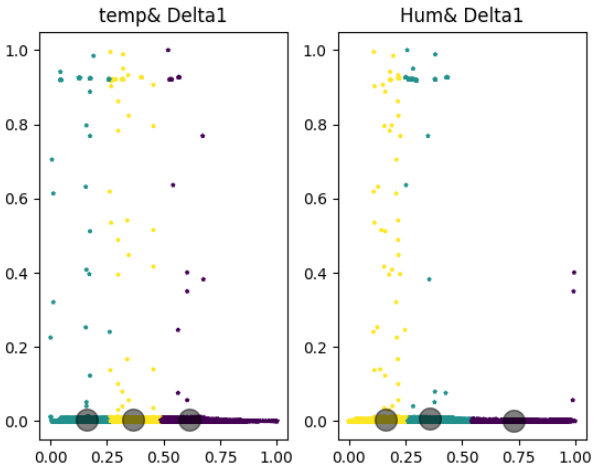



**Figure L.1:** VBG99b Plots, X-axis = Temp Y-axis = Delta

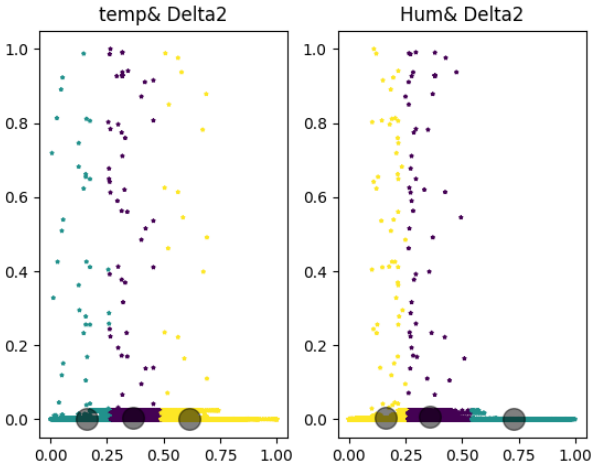



**Figure L.2:** VBG99b Plots, X-axis = Temp Y-axis = Delta

# M

# Appendix XIII

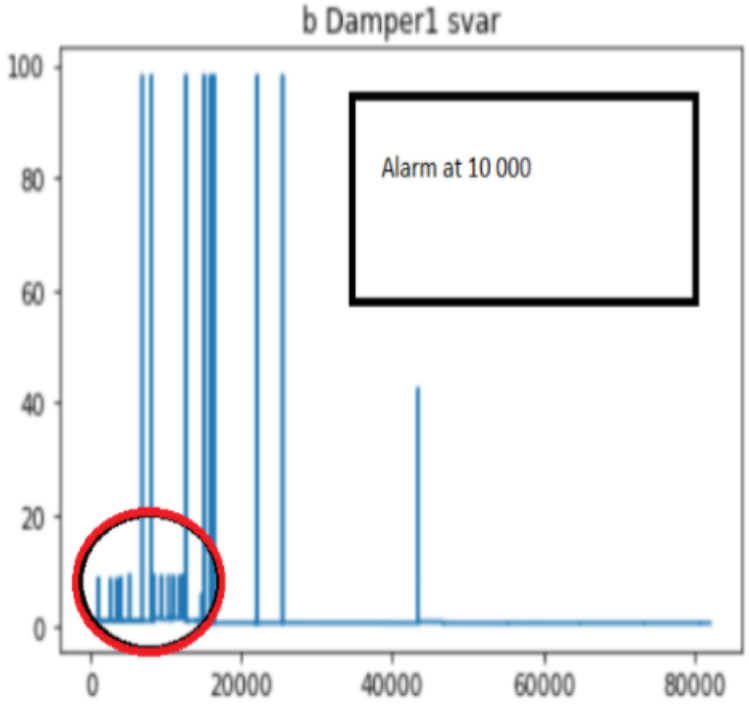



**Figure M.1:** Broken damper from start to failure

# N

# Appendix XIII

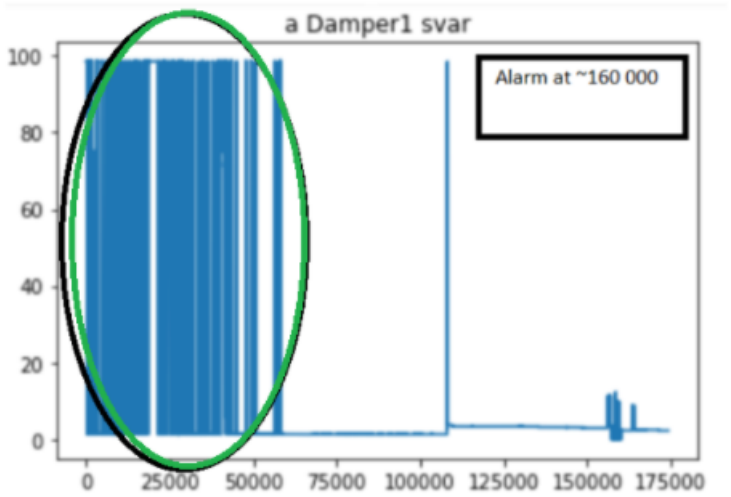



**Figure N.1:** Working Damper til alarm

**DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY**
**CHALMERS UNIVERSITY OF TECHNOLOGY**
Gothenburg, Sweden
www.chalmers.se


CHALMERS
UNIVERSITY OF TECHNOLOGY