MASTER

Road to predictive maintenance for capital goods
Condition-based maintenance on machine priority components

Thelosen, F.J.M.M.

Award date:
2021

Department of Industrial Engineering & Innovation Sciences
Mastertrack Manufacturing Systems Engineering

Road to predictive maintenance for capital goods

Condition-based maintenance on machine priority components

F.J.M.M. (Frank) Thelosen

Supervisors:
Rob Basten (TU/e)
Nima Manafzadeh Dizbin (TU/e)

Third Assessor:
Alp Akçay (TU/e)

Eindhoven, July 2021

Eindhoven University of Technology, School of Industrial Engineering
Series Master Theses Manufacturing Systems Engineering

Keywords: Condition based maintenance, Machine learning, Failure mode and effect analysis, Anomaly detection, Anomaly diagnosis, CRISP-DM

It should be noted that all names (i.e. *Company X, Machine Y, Customer Z, etc.*) mentioned in this thesis are fictitious due to confidentiality reasons

Preface

Voila, this thesis is the result of my graduation project that concludes my Masters in Operations Management and Logistics with a specialization in Manufacturing Systems Engineering at the Eindhoven University of Technology. Already during my Bachelors, maintenance optimization in manufacturing industries fascinated me. Moreover, during my Masters, I noticed data science is such a crucial skill for an industrial engineer. Therefore, I was glad to find a project where these two subjects were combined. I am proud of the lessons learned, the progress in research independence, and applying knowledge gained over the years in a business perspective. Conducting research for almost a year is not easy, especially since COVID-19 threw a spanner in the works regarding social interactions with colleagues. Therefore, I would like to thank everyone who encouraged me along the way in finalizing my 'student' chapter of life. Now I am fully prepared for making the first step of my professional career.

First, I would like to thank Rob Basten, my mentor during my Master studies, and supervisor of both this and my Bachelors project. He inspires me with his commitment, positive and motivational attitude. Our conversations always helped me to stay focused on the overall goal of the project and the feasibility of the research. Furthermore, I want to thank Nima Manafzadeh Dizbin for being my second supervisor and helping me in improving my machine learning skills. The data mining discussions we had truly improved the quality of this research. I would like to thank Alp Akçay for being the third assessor of this thesis.

This master thesis finalizes the student chapter of my life. Looking back, this amazing time made me see I developed myself in numerous personal and professional skills. I had to overcome quite some challenges but always won. Perseverance and critically-thinking are life lessons that cannot be taken away from me anymore. Last but certainly not least, successfully finalizing this chapter of life would not be possible without the amazing support of my parents, grandparents, girlfriend, brothers, and friends. I am looking forward to what the future brings!

Frank Thelosen, Eindhoven, July 2021

Abstract

Predictive maintenance in the form of condition-based maintenance has attracted a lot of attention in the industry due to the development of advanced sensor technology and measurement equipment (Peng, 2016). Condition-based maintenance is a preventive maintenance policy that makes use of the information about the actual health status or degradation of a component, rather than relying only on the usage or age of a component. A maintenance decision support system that considers the actual condition of an asset has proven to be transparent and accurate in planning maintenance actions (Jardine et al., 2006). In this research, we develop condition-based maintenance methods for anomaly detection and anomaly diagnosis purposes. In anomaly detection, we use models (i.e. isolation forest, local outlier factor, and random forest regression) to find potential anomalies that are validated by domain knowledge. Conversely, in anomaly diagnosis, we use multiclass classification algorithms (i.e. k-nearest neighbor, random forests, and artificial neural networks) to predict root causes for specific sensor behaviors. We show the anomaly diagnosis model can reach an F_{β}-score of 0.888.

Executive summary

Business problem

Unexpected failures of physical assets are one of the key operational hazards in the capital goods industry because they can disrupt supply chains and impose substantial costs due to lost productivity (LaRiviere et al., 2016). It is therefore important to plan maintenance actions before a failure occurs, but also not too early from an efficiency perspective (Tiddens, 2018). Condition-based maintenance (CBM) is a preventive maintenance policy that makes use of the information about the actual health status or degradation of a component, rather than relying only on the usage or age of a component. This maintenance decision support system has proven to be transparent and accurate in planning maintenance actions (Jardine et al., 2006).

Company X mainly uses corrective and time-based maintenance where the replacement periods are based on the knowledge of field service engineers (FSEs) which prevents unscheduled downtime to a large extent. However, Company X sees opportunities in CBM to further improve on maintenance optimization. At this time, Company X has a limited picture of a (sub)system degradation path. It might be the case that some service parts are replaced too early or too late since it is unclear if the threshold for time-based maintenance has been set accurately or not. Furthermore, if a failure occurs at a client and the FSE wants to know what might have happened with Machine Y, a data specialist is necessary to detect and solve issues. This process takes time and could be automated by building condition-based anomaly detection and diagnostic models. The main question that is answered by this research is formulated as follows:

How can CBM be applied on critical components of Machine Y and what is the added value compared to the current situation?

Analysis

The analysis started off by acquiring a reference point that can be compared with the expected costs and downtime when CBM is used. Components included in the preventive maintenance schedule (PMS) are truly replaced on a fixed time interval, which is not always the case for inspection-based maintenance and can therefore be used as a reference point to compare with the CBM predicted replacement period. However, since the maintenance information sources are not complete in archiving all Machine Y failures, the data lacks quality on the *frequency of component replacement*, the consequential *amount of downtime*, and the *yield effects* of sub-optimal processing machines. This denotes we need to utilize domain knowledge to find the most suitable components for CBM.

As we cannot investigate CBM for each Machine Y component, we found suitable candidates (i.e. main air distribution fan, conveyor belt, and conveyor belt drives) for CBM by applying the three-stage funnel developed by Tiddens et al. (2018). However, two major risks that were identified appeared to have consequences for the remaining of this research. The fan has scarce maintenance data available and seems to be a robust component. Furthermore, the conveyor belt has no direct sensors monitoring the behavior. Conversely, the conveyor belt drives had sufficient failure frequencies and related sensor information to perform anomaly detection and diagnosis.

We built anomaly detection models to automatically detect a failed or under-performing priority component. According to Hawkins (1980), "An anomaly is an observation that deviates so much from the other observations as to arouse suspicions that it was generated by a non-optimal operating or failed priority component of Machine Y." (p.1). We assumed anomaly detection models take advantage of two quantitative properties of anomalies: (i) they are the minority consisting of few instances, and (ii) they have attribute-values that are very different from those of normal instances (Liu et al., 2012). As the anomaly detection models (i.e. isolation forest (IF), local outlier factor (LOF), and random forest (RF) regression) have been formulated to find potential anomalies, validating them is an important next step. Frequently, anomaly detection algorithms not only find anomalies but also outliers that are not interesting to the application-specific situation. By utilizing expert knowledge with the help of visualization tools to show anomaly behavior, we validated the potential anomalies.

Only two anomaly types have been distinguished related to the fan which is not sufficient for further evaluation. Conversely, the belt drive motors were further evaluated as they show fourteen types of interesting anomalies all connected to anomalous power consumption behavior. Interpreting the anomaly detection scores made us conclude the models can distinguish fluctuating power behavior from normal data points. However, the anomaly detection model cannot distinguish the difference between interesting and uninteresting anomalies since they both contain fluctuating power behavior. Therefore, anomaly diagnosis is important to learn the model how to distinguish interesting from uninteresting machine behavior. Furthermore, evaluating the models resulted in the finding that IF and LOF can better distinguish fluctuating power behavior from normal data points than the RF. This is because the IF and LOF can include temporal information (i.e. rolling window calculations) and no specific favor of one modeling technique over the other was found. However, we do have to take into account this score is based on an evaluation of a debatable test set. This is because the distinguished anomaly types have been found and evaluated by the same modeling technique, no cross-validation has been used because of a small test set, and the recall score that cannot be reliably interpreted as the test set does not contain all different types of anomalies.

We built anomaly diagnosis models to automatically find the root cause of a failed or under-performing conveyor belt. There are multiple root causes, so we have to deal with a multiclass classification problem. Multiple off-the-shelve models are available that can deal with multiclass classification. Three of these models (i.e. k-Nearest Neighbors (KNN), Random forest (RF), and artificial neural networks(ANN)) are selected that are evaluated and compared to each other. The RF scored higher (F_β-score = 0.888) than the ANN and KNN. Therefore, the RF model is used to investigate how well the distinguished root causes can be identified. Three root causes were able to be distinguished with great certainty. Four root causes scored well for precision and recall. Only two root causes had an F_β-score around 0.5. Hence, we can conclude the old situation where data specialists at Company X had to specifically look at parameter behavior can be transformed into a situation where the anomaly diagnosis model can reliably classify some root causes. This saves time of the Company X data specialists.

Recommendation

The analysis results in the following distinct recommendations when willing to take more steps in PdM and to reproduce this research on other machines. The first and second recommendation is company-specific whereas the third and fourth are general recommendations. (i) Improve the data acquisition process in four ways. These ways are creating awareness of documentation importance among FSEs and customers, inspecting replaced components of the PMS, and add more information sources to CXMDS. (ii) Implement two sensor types to increase the predictive value of a component's remaining useful lifetime. These sensor types are sensors that monitor the length and tension of the conveyor belt and sensors that monitor the vibration of motors. (iii) Perform time-series analysis in anomaly detection to find gradual wear and tear. Finally, (iv) focus on more frequently failed components to reach a richer ML environment.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
CBM	Condition-based maintenance
FMEA	Failure Mode and Effect Analysis
CXGMC	Company X Global Maintenance Concept
FSE	Field Service Engineer
PdM	Predictive Maintenance
SME	Subject Matter Expert
PMS	Preventive Maintenance Schedule
RPN	Risk Priority Number
AHP	Analytical Hierarchy Process
PS	Potential Showstopper
ERP	Enterprise Resource Planning
RPN	Risk Priority Number
LCC	Life Cycle Costing
MCDM	Multi-Criteria Decision Making
IF	Isolation Forest
LOF	Local Outlier Factor
RF	Random Forest
ML	Machine Learning
CRISP-DM	Cross-Industry Standard Process of Data Mining
CMD	Customer Maintenance Documentation
CXMDS	Company X Maintenance Documentation System
BT	Base Tower
ST	Secondary Tower
ANN	Artificial Neural Network
KNN	k-Nearest Neighbors

Chapter 1

Introduction

Unexpected failures of physical assets are one of the key operational hazards in the capital goods industry because they can disrupt supply chains and impose substantial costs due to lost productivity (LaRiviere et al., 2016). It is therefore important to plan maintenance actions before a failure occurs, but also not too early from an efficiency perspective (Tiddens, 2018). However, most current maintenance decisions still rely on previous experiences and expert knowledge which are often not transparent and accurate (Tinga, 2010). On the other hand, predictive maintenance (PdM) in the form of condition-based maintenance has attracted a lot of attention in the industry due to the development of advanced sensor technology and measurement equipment (Peng, 2016). The term predictive maintenance refers to a maintenance policy that triggers maintenance activities by predictions of failures (Tiddens, 2018). On the other hand, CBM is a preventive maintenance policy that makes use of the information about the actual health status or degradation of a component, rather than relying only on the usage or age of a component. A maintenance decision support system that considers the actual condition of an asset has proven to be transparent and accurate in planning maintenance actions (Jardine et al., 2006). Hence, more successful prevention of unexpected failures and reducing total costs of maintenance can be achieved. This thesis is the result of research regarding anomaly detection and diagnostics for enabling CBM.

This section introduces the problem by describing the research environment in Section 1.1 which helps in better understanding the problem. This research environment contains a company and Machine Y description together with the companies vision regarding maintenance. Subsequently, Section 1.2 discusses the research design by elaborating on the multiple existing maintenance strategies with a more detailed explanation of condition-based maintenance. Furthermore, the research design elaborates how the road to predictive maintenance can be defined and presenting the problem statement, research goal, research scope, and research questions. Lastly, the outline of this thesis is formulated in Section 1.3.

1.1 Research environment

1.1.1 Company X description

This research is performed at Company X which is a leading global provider of advanced food processing equipment, systems, software, and services. By continuously transforming food processing, Company X enables its customers to increase yield and throughput, ensure food safety, and improve sustainability in food production. Company X has three main industries: Industry X, Industry Y, and Industry Z. The business unit in which this research is performed assists these three main industries.

1.1.2 Machine Y

Company X offers multiple types of machines suitable for different capacities and applications. One of these machines, which is developed by the business unit in which this research is performed, is the Machine Y where this research is focused on. Machine Y can process a large variety of products at high capacity and is one of Company X's best-known products. Machine Y consists of two isolated towers with controls for air temperature, humidity, and velocity. This gives the necessary adjustable functionality to process in two zones. A conveyor belt transports the products through the whole machine, runs spiral-shaped through the two climate zones, and returns underneath the machine. The two main purposes of Machine Y are creating an attractive surface appearance on the product and increasing food safety. The products have to be heated up until the temperature at the core of the product will exceed a certain limit for a certain time. This reduces the number of micro-organisms that can cause foodborne diseases.

1.1.3 Company X Global Maintenance Concept

According to Park (2002), the 'Operating & Maintenance' phase of a machine's life cycle is responsible for the highest annual cost in the stages of life cycle costing (LCC). Improvements in this phase can lead to significant advantages for companies and customers. LCC is an important concept for Company X and its customers since it is the sum of all funds expended in support of the item from its conception and fabrication through its operation to the end of its useful life (White and Ostwald, 1976). With the installed base continuing to grow worldwide, a significant proportion of Company X's revenues (around 40%) results from recurring service and spare parts revenues. Company X is transforming from a reactive to a proactive maintenance strategy. More and more customers of Company X demand a higher standard in terms of machine availability because they see the benefits of plannable maintenance that facilitate scheduled downtime.

Company X's vision to invest in proactive maintenance should lead to lower downtime and repair costs and fewer performance issues decreasing the LCC. Long-term trends in the business environment show that it is getting more complicated for customers to perform maintenance on their systems. Additionally, users require higher availabilities and lower costs. Therefore, these customers mainly outsource the maintenance to third parties or Original Equipment Manufacturers. Company X's goal is to be the organization to whom this maintenance is outsourced. Therefore, Company X needs to show its added value in supporting the whole lifecycle in a uniform, efficient and standardized way in the form of the Company X Global Maintenance Concept (CXGMC). The CXGMC is a gathering of maintenance activities for a particular asset that provides added value for the customer by (i) maximizing uptime, (ii) improving the performance and the yield level, (iii) minimizing reoccurring failures in the form of continuous improvement, (iv) implementing plannable maintenance, (v) coming up with more accurate and transparent pre-scheduled maintenance budgets.

1.2 Research design

1.2.1 Introduction to maintenance policies

Figure 1.1 shows that maintenance policies fall into one of three categories: reactive maintenance, aggressive maintenance, and proactive maintenance (Tiddens et al., 2018). Each with its advantages, disadvantages, and challenges. Reactive maintenance allows parts to run to failure. In this policy, the maximum utilization of the components is reached since they are used until the very limits. On the other hand, reactive maintenance can lead to undesired downtime costs. Aggressive maintenance improves the design of a component to increase its performance. This maintenance action is typically project-based and thus non-recurring. Proactive maintenance tries to prevent problems before they occur and can be subdivided into preventive and opportunistic maintenance. Opportunistic maintenance clusters maintenance activities based on production planning. Preventive maintenance can be sub-divided into the traditional way of prescribing maintenance actions based on time or usage. Contrarily, maintenance can be preventively performed based on the actual condition of a component (striped rectangle in Figure 1.1). CBM policies can be subdivided into policies that use the measured condition of the asset and policies that use the calculated condition of an asset. Hence, the measure phase monitors the condition of particular components of a machine with sensors and uses thresholds in maintenance-decision making. Contrarily, the calculation-based condition of an asset uses models with data analytics to detect, diagnose, and prognose failures and come up with decision support for maintenance actions which is further elaborated in Section 1.2.2.

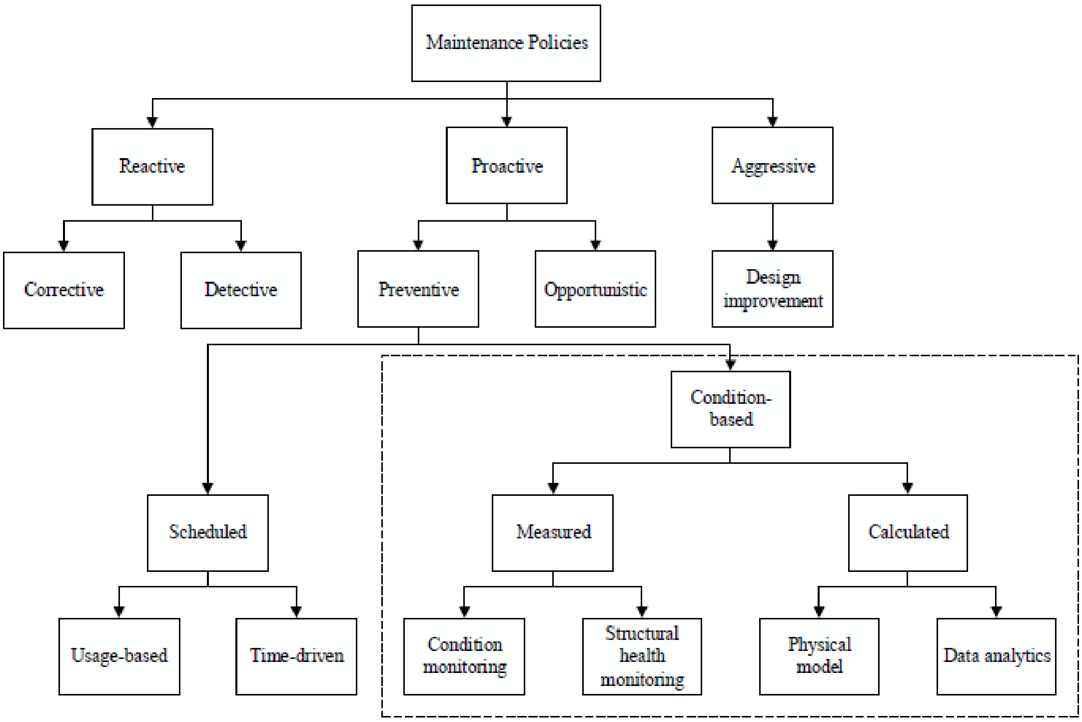


Figure 1.1: Maintenance policies (Tiddens, 2018).

1.2.2 Condition-based maintenance

As CBM is the main subject of this thesis, further elaboration is provided in this section. According to Jardine et al. (2006), there are two categories of maintenance decision support: diagnostics and prognostics. Diagnostics deals with fault detection, isolation, and identification when it occurs. Fault detection is a task to indicate whether something is going wrong in the monitored system (i.e. anomaly detection), fault isolation is a task to locate the faulty component, and fault identification is a task to determine the nature of the fault when it is detected. Prognostics deals with fault prediction before it occurs. Fault prediction is a task to determine whether a fault is impending and estimate how soon and how likely a fault will occur.

According to Jardine et al. (2006), prognostics is much more efficient than diagnostics to achieve zero-downtime performance. However, diagnostics is necessary for building a knowledge base to know how components degrade and to distinguish interesting anomalous patterns. Furthermore, the power of ML algorithms is to recognize patterns and make predictions. Therefore, huge amounts of data are desired as more data means better learning opportunities. Diagnostics can help in preparing more accurate event data that can be used for improving the prediction algorithm.

This research focuses on anomaly detection to find data points representing a failed or non-optimal performing machine. When we have sufficient confidence in the anomaly detection models, the output of these models can be labeled by experts to build an anomaly diagnosis model. Prognostics is also important for CBM but is not part of the scope (Section 1.2.6). Recommendations will be given regarding prognostic purposes in Section 8.2.

1.2.3 Road to predictive maintenance

The primary objective of this research is to bring Company X a step further in predictive maintenance. According to Haarman et al. (2017), there are four levels of maturity in PdM. Level 1 (visual inspections) consists of periodic physical inspections where maintenance activities are based solely on the inspector's knowledge, experience, and intuition. Level 2 (instrument inspections) includes periodic inspections where conclusions are based on a combination of the inspector's expertise and instrument read-outs. Level 2 contains more specific and objective information about the condition of the asset in question than Level 1. Level 3 (real-time condition monitoring) monitors in real-time the assets condition, where alerts are given based on pre-established rules or critical levels. Level 4 (predictive maintenance 4.0) is about predicting failures in assets and ultimately prescribing the most effective preventive measure by applying advanced analytical techniques on big data about anything that may correlate with the performance of an asset (technical condition, usage, environment, and the maintenance history).

Nowadays, Company X is situated in the first level of predictive maintenance maturity as the preventive maintenance schedule (Section 2.2) is created by periodic physical inspections with checklists where conclusions are solely based on the expertise of the field service engineer (FSE). This research aims to show CBM applies to Company X and provides added value to Company X and its customers which aims to bring Company X from level one to level three of the PdM Maturity Matrix (Haarman et al., 2017). By automatically detecting and diagnosing interesting anomalies for CBM purposes we aim to improve the maintenance decision making process.

1.2.4 Problem statement

Company X mainly uses corrective and time-based maintenance where the replacement periods are based on the knowledge of FSEs which prevents unscheduled downtime to a large extent. However, Company X sees opportunities in CBM to further improve on maintenance optimization. Figure 1.2 shows a degradation path where the blue dots represent good states and orange dots represent degraded states. At this time, Company X has a limited picture of a (sub)system degradation path. It might be the case that some service parts are replaced too early or too late since it is unclear if the threshold for time-based maintenance has been set accurately or not. Replacing parts too early can lead to higher maintenance costs since components do not use their maximum lifetime. Moreover, replacing too early is not sustainable because of over-consuming spare parts. On the other hand, replacing parts too late can lead to downtime.

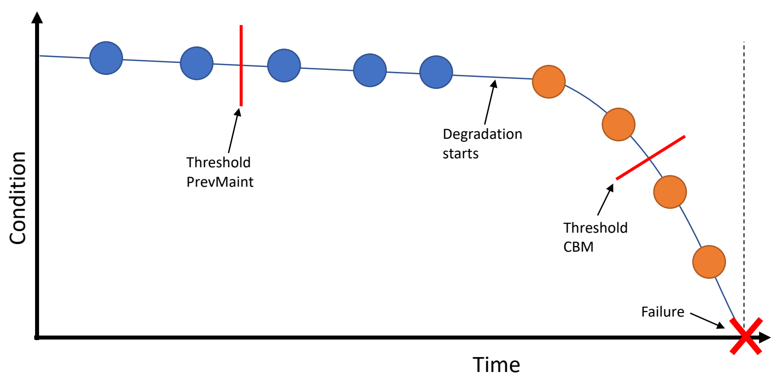


Figure 1.2: Degradation path example.

By interviewing multiple subject matter experts (SME) at Company X, we found that servitization and introducing full-service contracts are transformations that have to be taken into account to stay competitive in the food industry. Servitization is known as the process through which manufacturers can base their competitive strategies on services since product innovation alone does not produce sufficient competitive advantage and growth (Baines and Lightfoot, 2014). However, Company X does not sell full-service contracts yet, but sells its systems and provides service to the customer charging this separately.

If a failure occurs at a client and the FSE wants to know what might have happened with Machine Y, a data specialist is necessary to detect and solve issues. This process takes time and could be automated by building anomaly detection and diagnostic models. The model should support the service employees in finding issues and root causes on their own and save time from the data specialists.

CBM has the potential to maximize the useful lifetime of parts and find solutions to unscheduled downtime at the same time. CBM intends to create an accurate degradation path that can be used for selecting an optimal CBM threshold taking costs and availability into account. The main advantage is that, since failures can be better predicted, maintenance actions can be scheduled during hours where the machine is in its idle state resulting in less downtime. Additionally, CBM provides more transparent maintenance support since decisions are based on data. This property of CBM helps to make up full-service contracts and move up in the transformation to servitization. Additionally, automatic detection and diagnostics can save time for data specialists because they do not have to dive deep into particular failure cases in the field. However, Company X does not use CBM nowadays, leading to the following problem statement:

It is unclear which Machine Y components have the highest potential for conducting CBM, and how to automatically detect and diagnose failures to enable CBM.

1.2.5 Research goal

The research goal is to build models, based on available data, that can detect and diagnose failures or under-performing circumstances of distinguished Machine Y priority components to enable CBM. Therefore, we aim to prescribe more transparent and accurate maintenance decision support to agree on the ambitions of CXGMC (Section 1.1.3). Furthermore, this research aims to bring Company X a step further in PdM by transforming from maintenance based on visual inspections to maintenance based on real-time condition data for the priority components (Section 1.2.3). Moreover, this project helps Company X moving up in the transformation to servitization in the form of full-service contracts by providing transparent data-based maintenance decision support. Lastly, achieving the defined goal should result in automatic fault diagnostics preventing Company X data-scientists to look at each case individually which saves time.

1.2.6 Research scope

The scope of this research includes condition-based maintenance on three Machine Y priority components. Primarily, this research will be based on sensor data that is already available (i.e. no focus on finding new interesting sensors). According to Jardine et al. (2006), techniques for maintenance decision support in a CBM program can be divided into three main categories: detection, diagnostics, and prognostics. This research focuses on automatic anomaly detection and diagnostics. Recommendations are provided for anomaly prognostic purposes (Section 8.2).

1.2.7 Research questions

The overall model that will be followed to structure the data analysis project is the Cross-Industry Standard Process of Data Mining (CRISP-DM) (Chapman et al., 2000). The CRISP-DM process model, consisting out of six stages shown in Figure 1.3, aims to make large data mining projects, less costly, more reliable, more repeatable, more manageable, and faster (Wirth and Hipp, 2000).

The following main question is formulated to solve the problems mentioned in the problem statement and reach the research goals:

How can CBM be applied on critical components of Machine Y and what is the added value compared to the current situation?

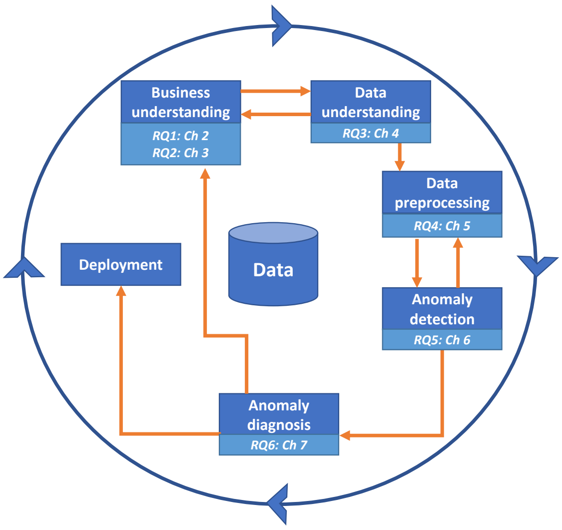


Figure 1.3: Phases of the CRISP-DM process model.

This main question can be divided into research questions that fit into the phases of the CRISP-DM model (Figure 1.3). By interviewing service managers and coordinators and investigating how maintenance is performed on Machine Y in the field (qualitative research) and providing clarity among the costs and downtime by analysing maintenance data (quantitative research), we acquire a reference point that can be compared with the situation when CBM is used. Hence, by applying qualitative and quantitative research we aim to find an answer to the following research question:

RQ1. How is maintenance performed on Machine Y nowadays and what are the corresponding costs and downtime?

When having answered RQ1, identification of suitable Machine Y candidates for CBM is necessary as we cannot investigate CBM for each Machine Y component. By performing a failure modes and effects analysis (FMEA) with SMEs this research aims to systematically identify possible root causes, failure modes, and their relative risks to answer the following research question:

RQ2. What are suitable Machine Y candidate components for CBM?

As we elaborated what the research problem is in the introduction, explained how the research environment looks like, investigated how maintenance is performed on Machine Y nowadays and what suitable candidate components are for CBM, we can accomplish the business understanding phase of the CRISP-DM model. The second step of the CRISP-DM model aims to understand if the available data can help in reaching our research goal (defined in Section 1.2.5) leading to the following research question:

RQ3. Can the available data enable CBM for the priority components of Machine Y?

When we have sufficient knowledge to conclude the available data can help us in reaching our research goal, the raw data can be prepared for modeling. Tasks include attribute selection, data cleaning, construction of new features, and transformation of data for modeling tools. Data understanding in combination with data preparation is a time-consuming task, for which it is important to understand Machine Y process well. Therefore, close collaboration with functional experts is necessary to answer the following research question:

RQ4. How to preprocess the available raw data so it can be used for anomaly detection and diagnostics?

Sufficient and efficient decision support would be crucial to maintenance personnel's decisions on taking maintenance actions. When the data preprocessing steps have been accomplished, the models can be built for enabling automatic anomaly detection and diagnostics for the selected candidates for CBM. Different modeling techniques apply to anomaly detection and diagnostics and are therefore split up in RQ5 and RQ6, respectively. Research question 5 aims to build anomaly detection models for the distinguished priority components. Anomaly detection is binary by nature, it indicates whether a system is healthy or faulty (Tiddens et al., 2020). By selecting the right dataset, the right modeling techniques, and utilizing expert knowledge to validate the potential anomalies, we aim to find an answer to the following research question:

RQ5. How to detect a failed or under-performing priority component?

After investigating how interesting anomalies look like and evaluating the anomaly detection modeling techniques, we aim to diagnose faults in priority components. The diagnostic phase focuses to understand why something happened. According to Karim et al. (2016), diagnostics can be divided into fault isolation (i.e. determining the cause and the related component) and fault identification (i.e. designating the type and the nature of the fault.). The final determination of why a type of failure has occurred is made by recognizing patterns between the feature extraction results and the knowledge base, which may be derived from expert knowledge, physical models, and historical data (Jardine et al., 2006). Therefore, we aim to label the distinguished anomaly types in RQ5 and build a diagnostic model to automatically provide the Company X FSE with a probability of specific fault isolation and identification in the following research question:

RQ6. How to find the root cause of a failed or under-performing priority component?

In the end, the models should be evaluated to show their performance on failure detection and diagnostics (i.e. fifth step in the CRISP-DM model). Hereafter, a discussion is written where we conclude what the added value of the new CBM situation is compared to the current situation.

1.3 Project outline

Chapter 2 elaborates the current Machine Y state. Chapter 3 explains how to select the suitable CBM candidates for Machine Y. Chapter 4 aims to understand the data and answer the question if the available data can help us in reaching our research goals. Chapter 5 discusses how the data should be prepared and reduced so it can be used for modeling. Chapter 6 discusses how to detect interesting anomalous behavior of Machine Y. Chapter 7 elaborates how to find the root cause of failures or less-performing situations of Machine Y. This thesis finalizes with a discussion in Chapter 8 where we conclude, provide recommendations, elaborate the limitations, and give directions for future research.

Chapter 2

Current Machine Y state

By investigating how maintenance is performed on Machine Y in the field (qualitative research) and providing clarity on the corresponding costs and downtime (quantitative research), we aim to acquire a reference point that can be compared with the expected costs and downtime when CBM is used. Therefore, this chapter answers Research Question 1:

How is maintenance performed on Machine Y nowadays and what are the corresponding costs and downtime?

Section 2.1 discusses the customer-specific environment in which Machine Y has to operate. Section 2.2 explains how Company X performs maintenance on Machine Y nowadays. Section 2.3 discusses the maintenance data that is available. At the end of this chapter, a conclusion is drawn in Section 2.4.

2.1 Customer-specific environment

Company X offers its processing lines in two different ways (i.e. traditional and innovative). The traditional processing line contains the following sequential steps: forming, coating, frying, and cooking. This implies the products enter Machine Y already fried resulting in oily products that automatically lubricate the conveyor belt. On the other hand, the products enter Machine Y without being fried in the innovative way. Thus, a lubrication system is needed for the conveyor belt of Machine Y. These two types of processing lines probably have their influence on Machine Y degradation behavior (Table 2.1).

Approximately fifty machines are installed in the field. However, Company X is real-time connected to thirteen machines by a VPN connection. Three different machine types are included in this research: (i) Machine Type A (ii) Machine Type B (iii) Machine Type C (Table 2.1). Machine Type A is the oldest version with a belt width of x centimeters. Machine Type B is the newer version of Machine Type A with the same belt width. Machine Type C is the newer version of Machine Type A with a belt width of y centimeters. The machine types with wider belt widths need more powerful motors to operate well. A significant difference between Machine Types A,B on one side and Machine Type C on the other side can be distinguished resulting in separately modeling them in the remaining of this research.

Lastly, every user treats its system differently. Some customers use their machines more often than others as can be seen in Figure B.1. Moreover, some customers use their fan in a stable process (i.e. only one or two different airspeeds) while others use the fan in an unstable process (i.e. many different airspeeds). As can be derived from the fluctuations in Figure B.2, some customers show constant power consumption while others show a high diversity in power consumption that probably means the machine has to process multiple types of products influencing the degradation behavior of Machine Y. Furthermore, Figure B.3 shows that some customers value the amount of cleaning the system higher than others.

ID	Machine type	Processing line
1	A	Traditional
2	C	Traditional
3	A	Traditional
4	C	Traditional
5	A	Innovative
6	C	Innovative
7	A	Traditional
8	A	Traditional
9	C	Innovative
10	C	Innovative
11	A	Traditional
12	B	Innovative
13	B	Innovative

Table 2.1: Customers included in this research.

2.2 Preventive maintenance schedules and inspections

Company X maintains its Machine Y in two different ways. The first way is providing a customized preventive maintenance schedule (PMS) to the customer and the second way is agreeing with the customer to provide visual inspections and replace components when necessary. A PMS is a predefined schedule of overhauls needed to secure the technical functioning of the equipment consisting of wear and tear parts. An overhaul is a planned repair and maintenance of Company X equipment and fits in the *Scheduled time-driven* box in Figure 1.1. Field Service Engineers (FSE) determine the replacement period of wear and tear parts by experience taking the criticality and failure behavior into account. This PMS can be updated when an FSE gets new insights regarding the lifetime of a component. The advantage of Machine Y users when choosing a PMS is the fixed maintenance costs which can be budgeted already in advance. The disadvantage of a PMS is some components are replaced even though the FSE can hardly distinguish any wear and tear.

Inspection schedules are similar to a PMS. However, the FSE is free to keep the component in the system when having confidence in another faultless inspection interval. FSEs prefer this way of maintenance over PMS since replacing components where no degradation can be visually noticed is hard to explain to the customer.

2.3 Maintenance data

To find out the current maintenance costs and downtime per Machine Y, we need the (i) frequency of component replacement, (ii) prices of components, (iii) amount of downtime per Machine Y, (iv) downtime costs (i.e. lost production, reputation loss, employee salaries, travel expenses, building expenses during non-working hours), and (v) yield effects. This section aims to find this specific data in the Company X maintenance software system environment (i.e. Company X maintenance documentation system, ERP system, and customer maintenance documentation). As elaborated in Section 2.2, customers who have a PMS replace their components on a fixed replacement interval. However, since FSEs made clear components are replaced even though hardly any wear and tear can be distinguished, the PMS replacements cannot be used as information to know when a specific failure happened to a customer. Therefore, the PMS replacement intervals are not included in the maintenance data that we use to validate our distinguished anomalies and diagnostics in Chapter 6 and 7, respectively.

2.3.1 Customer maintenance documentation

The customers themselves are the ones who have the most information on their maintenance actions and corresponding downtimes. Therefore, every customer sharing real-time sensor data with Company X has been asked to also share Machine Y failure data. However, not every customer has a high-quality customer maintenance documentation (CMD). About half of the customers shared their maintenance data for this research (Table 2.2). Reactions of customers that did not share their maintenance data are related to confidentiality issues or that they do not document their replacements or maintenance actions. As can be seen in Table 2.2, the customer data lists lack quality as not every customer shared their replacement data and, if they do, they are not always specifically related to machine failures.

			Available data records			**Available years of data**		
Customer	Quality customer failure data	Actions to structure customer failure data	CMD	CXMDS	ERP	CMD	CXMDS	ERP
1	Not willing to share	-	-	4	-	-	2018-2020	-
2	Not willing to share	-	-	-	-	-	-	-
3	Not willing to share	-	-	6	2	-	2017-2020	2015-2020
4	Just started collecting maintenance data (no fan or belt related issues)	-	0	-	1	Nov 2020 - Dec 2020	-	2015
5	Good	-	6	-	1	2019- 2020	-	2011
6	They do not collect maintenance data	-	-	-	-	-	-	-
7	They do not collect maintenance data	-	-	-	-	-	-	-
8	Not willing to share	-	-	1	1	-	2018	2013
9	Good	Translate German to English	21	1	-	2014 - 2020	2020	-
10	Good	Translate German to English	5	1	-	2014 - 2020	2020	-
11	Good	Translate Swedish to English	10	1	-	2017 - 2020	2020	-
12	Mainly performance issues (no failure issues)	Translate German to English	71	1	1	2018 - 2020	2018	2018
13	Mainly performance issues (no failure issues)	Translate German to English	55	-	-	2018 - 2020	-	-

Table 2.2: Maintenance data available.

2.3.2 Company X maintenance documentation system

Company X makes use of a maintenance documentation system (CXMDS) in which the FSEs can document the maintenance actions performed. This source of information can be used in addition to the maintenance data the customers shared. It should be taken into account that these documentations only exist out of replacements performed by Company X. It could also be possible that some major replacements/refurbishments have been outsourced by the customer to a third party or have been replaced by the customer themselves. Furthermore, by interviewing SMEs (i.e. service managers, coordinators, and FSEs), we found that maintenance actions are documented occasionally. As it takes more effort for the FSEs to report their actions and they do not realize the importance of documenting, it frequently occurs the FSE does not archive the replacements in CXMDS resulting in an information loss. Moreover, CXMDS archives the starting date when the customer calls to mention an issue and the end date when the specific issue is solved. This implies, we only know the time span in which the maintenance action is performed. But we do not know the specific moment in time when the replacement took place.

2.3.3 Enterprise resource planning

The Enterprise Resource Planning (ERP) system of Company X has been consulted to look for component replacements of Machine Y. This system provides information when specific components have been sold and can give an indication when a component has been replaced. However, uncertainties arise as we do not know if this data source provides reliable failure data. It can be the case customers buy parts for inventory purposes and lay the components on stock. Moreover, the ERP order date does not mean the failure of that order took place on the same day. Therefore, uncertainty arises on what specific moment the archived component failed.

2.3.4 Conclusion

After analysing the available maintenance data, we can conclude there are seven known issues regarding the fan of Machine Y. Since not every customer shared their maintenance lists and the quality of documentation in CXMDS is rather low, we can assume more failures occurred. However, this denotes we can validate our anomaly detection and diagnostics model with only a few known issues. On the other hand, failures linked to the conveyor belt are more frequent (i.e. 181 known issues in total) that makes validating the models more reliable. Furthermore, hardly any customer shared the corresponding downtime to a specific failure which makes it difficult to determine the criticality of failures. Moreover, customers do not know if their machine operates sub-optimal resulting in a lower yield and can therefore not be numerically expressed for comparing with the CBM solution.

2.4 Conclusion

In this section, we answer research question 1: *How is maintenance performed on Machine Y nowadays and what are the corresponding costs and downtime?* As we deal with a specific Machine Y environment at each customer (i.e. different processing lines, Machine Y types, utilization rates, cleaning times), also specific maintenance decisions are advised. Currently, Company X offers a PMS and inspection-based maintenance which are customer-specific fixed time replacement maintenance advises. However, the advises are based on FSE experiences and opinions rather than data. Nowadays, FSEs sometimes have to replace components that do hardly show any wear and tear due to the fixed replacement periods in a PMS. This makes it hard for an FSE to explain the component replacement to the client. CBM enables showing the real-time degradation level of a component which makes it more transparent for an FSE to explain the component replacement to a customer. Therefore, a maintenance policy based on the real-time condition of a component suits better than a fixed time replacement period.

Components included in the PMS are truly replaced on a fixed time interval, which is not always the case for inspection-based maintenance, and can therefore be used as a reference point to compare with the CBM predicted replacement period. This fulfills the aim of this research question which was to acquire a reference point that can be compared with the CBM scenario.

Historical maintenance and downtime costs are necessary for founding a business case where the added value of CBM can be expressed in monetary numbers. After analysing the available maintenance data, we found seven known issues regarding the fan and 181 known issues of the conveyor belt that can be used for validating our models. However, since the maintenance information sources are not complete in archiving all Machine Y failures, the data lacks quality on the *frequency of component replacement*, the consequential *amount of downtime*, and the *yield effects* of sub-optimal processing machines. This lack in data quality directly results in the current maintenance strategy of Company X that cannot be quantified in costs, component criticality that cannot be numerically expressed based on downtime information, and the output of the anomaly detector cannot be fully validated with data. This denotes we need to utilize domain knowledge to find the most suitable components for CBM and for validating purposes of the built anomaly detection and diagnosis models. In Section 8.2, recommendations are given to improve the data acquisition process.

Chapter 3

Suitable CBM candidates selection

As we cannot investigate CBM for each Machine Y component, we aim to find suitable candidates for performing CBM. Research Question 2 helps in assessing where CBM would provide the greatest benefit in guaranteeing availability and reducing downtime costs:

What are suitable Machine Y candidate components for CBM?

This chapter begins with a literature review regarding suitable candidate selection methods for CBM in Section 3.1. The advantages and disadvantages of multiple selection methods are compared to find the most appropriate selection methodology in Section 3.2. Section 3.3 elaborates the selection procedure of suitable candidates for CBM at Company X after which a conclusion is drawn in Section 3.4.

3.1 Literature review

3.1.1 Failure mode and effect analysis

According to Gouriveau et al. (2016), an often applied approach is to define critical components as a component whose failure leads to unavailability of the whole system, and/or a component that has a high failure rate. The Failure Mode and Effect Analysis (FMEA) methodology has been used as a powerful tool for risk assessment and reliability analysis in a wide range of industries (Liu et al., 2019). By performing an FMEA with a group of experts this method aims to systematically identify possible root causes, failure modes, and their relative risks to distinguish the most critical components (Arabian-Hoseynabadi et al., 2010). The group of experts must be cross-functional and multi-disciplined and the team members must be willing to contribute to having an effective brainstorm. An FMEA prioritizes the identified failure modes according to the risk priority number (RPN). The RPN is the product of the ordinal values of the frequency of occurrence, severity, downtime, and detection (Table 3.1). Occurrence is the frequency of the failure. Severity is the seriousness of the failure for the customer (i.e. food safety, yield/ performance issues). Downtime is the amount of time the machine cannot operate due to a failure. Detection is the ability to detect the failure. The failure modes with higher RPN values are viewed as more important and thus should be given greater attention for risk mitigation. A quantitative approach that follows this logic is to select the top **X** cost drivers or availability killers for a PdM policy, as done in for example the degrade analysis of Banks et al. (2008). This approach has similarities with an often applied Pareto analysis based on maintenance costs, failures, downtime, or safety.

3.1.2 Analytical hierarchical process

Gupta and Mishra (2018) came with a solution to prioritize the critical components for reliability-centered maintenance. In this paper, an attempt is made to find out the key factors associated

Rating		Severity	Occurence	Downtime	Detection
Very high	10	No guarantee for food safety	Failure is almost inevitable. One occurrence per week	Repairment takes at least 1 day	Control can detect the effect
High	7	Failure renders the unit inoperable or unfit for use. (downtime)	Repeated failures. One occurrence per month	Failure can be repaired in 4 - 8 hours	Control can detect the failure mode
Moderate	3	Yield/performance issues	Infrequent failures. One occurrence per year	Failure can be repaired in 1 - 4 hours	Control can detect the cause
Very low	1	Minor defect	No failure within three years	Failure can be repaired in a matter of minutes	Control can prevent cause from happening

Table 3.1: FMEA rating definitions.

with the criticality of components. The distinguished five major criteria affecting the criticality of components are cost, functional dependency, complexity, maintainability, and safety impact. To identify the critical components, which is a multi-criteria decision-making (MCDM) problem, an analytic hierarchy process (AHP) is used (Figure 3.1). MCDM does not only focus on decision-making but also aims to provide insight into the decision process leading to a better understanding (Goodwin and Wright, 2014). AHP is an MCMD method in which the five major criteria are arranged in a hierarchical structure that is known as a well-established MCDM method in both academia and industry. Additionally, AHP is designed to integrate objective, subjective, qualitative, and quantitative information and produces plausible and defensible results (Goossens and Basten, 2015). Saaty (2008) developed the AHP methodology which works as follows: at the top of the hierarchy in an AHP is the decision-making goal. The criteria are on the next level, which can be decomposed to the sub-criteria (and further decomposed to the lower levels). On the last level are the alternatives. Weights are calculated by using pairwise comparisons and local priorities of alternatives. Then, global priorities can be calculated of alternatives and decisions can be made regarding priority components for CBM.

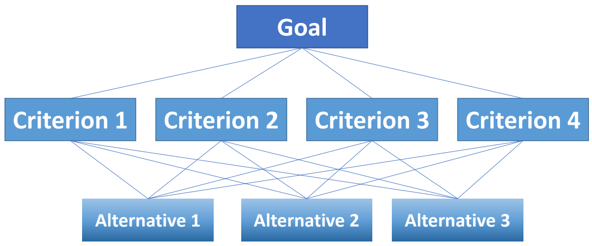


Figure 3.1: Analytic hierarchy process.

3.1.3 Three-stage funnel-based selection method

Another methodology in selecting suitable CBM candidates is the method of Tiddens et al. (2018) that consists of three stages: the criticality classification, the identification of showstoppers, and a focused feasibility study (Figure 3.2). The initial criticality classification acts as a filter to significantly reduce the number of possible candidates. The four-quadrant chart based on the work of Lee et al. (2009) helps to bring focus to only the most promising candidates, namely those with a low frequency of failure and a high associated failure consequence (e.g. failure, costs, or downtime). When the most critical components have been distinguished, a showstopper identification is necessary to check if the concepts of CBM can be beneficial for these components. Showstoppers are factors that can make the CBM approach infeasible or providing no added value. The third step contains a focused feasibility study where the showstopper factors will be studied in more detail. An economic feasibility study is necessary which focuses on whether developing the prognostic model is beneficial to the firm, from a strategic point of view. The technical feasibility study focuses on whether the firm can develop and implement the desired prognostic approach.

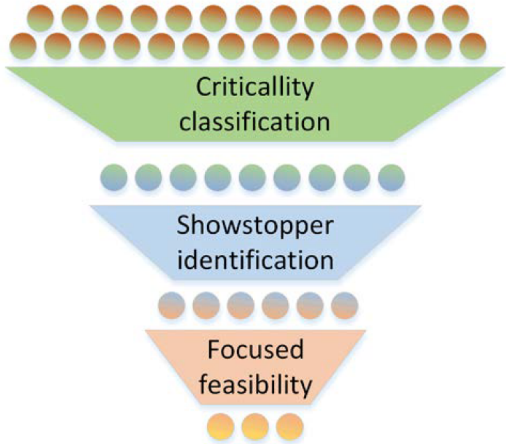


Figure 3.2: Funnel approach to identify suitable candidates for PdM (Tiddens et al., 2018).

3.2 Methodology selection

The method of Gupta and Mishra (2018) and an FMEA do not consider technical and economical feasibility in their methods. Although an imminent failure might be beneficial to monitor, predicting or detecting the anomaly before a failure occurs is not always possible. The methodology of Tiddens et al. (2018) does take this feasibility check into account. Additionally, the business case for predictive maintenance could be negative because: (i) introducing PdM is more costly than preventive or corrective maintenance or (ii) the probability of correct prediction of time to failure is not important as part will be replaced anyhow before each inspection. The methodology of Tiddens et al. (2018) does take this business case into account while that is not the case in the methodology of Gupta and Mishra (2018) and the FMEA. Furthermore, the method of FMEA and Gupta and Mishra (2018) do not take into account the consideration of organizational feasibility, whereas the method of Tiddens et al. (2018) does. Even though PdM might be economically and technically feasible, it should also fit the organization. Issues such as a lack of system and domain knowledge, a lack of trust in monitoring systems, and the organization not being ready to implement PdM can hinder the adoption of PdM. All in all, the listed advantages of the funnel approach over the other elaborated methods are the reasons why the methodology of Tiddens et al. (2018) is used to identify the most suitable candidates for CBM at Company X in Section 3.3.

3.3 Select suitable candidates for CBM at Company X

Together with the service and innovation department of Company X, we concluded the scope of this research includes three priority components (i.e. main air distribution fan, conveyor belt, and conveyor belt drive motors) that will be further researched for applicability in CBM. This denotes we do not have to check the criticality of all components of Machine Y but we only focus on the distinguished priority components.

3.3.1 Stage 1 - criticality classification

An FMEA (Section 3.1.1) is conducted to find the frequency of Machine Y component failures and consequential downtime. These results (Appendix A) can be plotted into the four-quadrant chart (Figure 3.3) to initialize the first filter to find suitable candidates for CBM (Lee et al., 2009). The four-quadrant matrix is a powerful method for identifying critical components that displays the frequency of failure versus the average downtime associated with failure (Tiddens et al., 2018). The quadrant with a high frequency of failure and high downtime corresponds to a maintenance strategy where redesigning the component would be recommended. The maintenance strategy recommended in the top-left quadrant is to have more spare parts on hand. The quadrant with a low frequency of failure and low downtime does not need any changes regarding its maintenance strategy since the current maintenance practices are working for these components. Lastly, the failures in the bottom-right quadrant should be maintained based on its real-time condition. Hence, in this approach, a condition-based concept is only applied to those components that have a low frequency of failure but a high associated downtime. The results of the FMEA prove this is the case for our priority components.

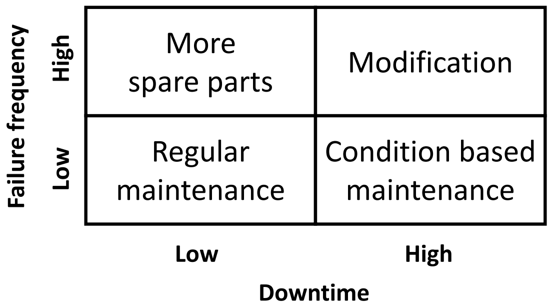


Figure 3.3: Four-quadrant chart.

3.3.2 Stage 2 - showstopper identification

A showstopper identification is necessary to check if the concepts of CBM can be beneficial for the priority components. First of all, three maintenance activities are differentiated to explore the possibilities and impossibilities by recognizing the potential showstoppers: detection, diagnostics, and prognostics. Our research is focused mainly on detection and diagnostics. However, prognostics is discussed in the recommendations (Section 8) and is therefore also taken into account in the showstopper identification analysis for the priority components. Analysing data sets without knowing the underlying failure mechanisms can lead to incorrect results. The main technical precondition for applying CBM is that it must be feasible to detect a gradual loss of function (or degradation). It is this gradual loss of function that leads to a predictable situation. Company X uses mostly mechanical components in its equipment since electrical components are more sensitive to the harsh environment in a processing plant. Many mechanical components have a clear degradation pattern, making them suitable for CBM. Table 3.2 can be used for identifying potential showstoppers and check if the concepts of CBM can be beneficial for the distinguished critical components. *Det* stands for detection, *Diag* stands for diagnostics, and *Prog* stands for prognostics. The grey regions in Table 3.2 mean these potential showstoppers are not relevant for the specific maintenance activities (i.e. Det, Diag, or Prog). The possible outcomes of the showstoppers can be *Yes* (it is a showstopper), *No*, or *Maybe*. In the final step of the funnel approach, the economic and technical feasibility are further examined for those components where a *Maybe* has been selected for one or more showstoppers.
Per priority component and ambition level of prognosis (i.e. detection, diagnosis, prognosis) it is determined if the potential showstopper (PS) is present. First, it is checked whether the priority components have a match with the production or inspection planning (c1 in Table 3.2). This implies the maintenance activity can be planned with a minimum of the duration of one operational period in advance. Applying this to Machine Y situation, we should be able to make a prognostic model with a predictive value for at least one day since some urgent failures can be repaired in the evening or night when the machine does not operate. This is a question we aim to answer in this research and therefore is labeled as a *maybe* showstopper. Secondly, the showstopper analysis looks if the priority components have a match with technical clustering.

					Main Air Distribution Fan			Conveyor Belt Drive			Conveyor Belt Itself		
Label	Potential Showstoppers (PS)	Det	Diag	Prog	Det	Diag	Prog	Det	Diag	Prog	Det	Diag	Prog
Clustering													
c1	No match with production or inspection planning		PS	PS		M	M		M	M		M	M
c2	No match with technical clustering		PS	PS		N	N		N	N		N	N
Technical Feasibility													
t1a	Failure cannot be detected with existing technology	PS	PS		M	M		M	M		M	M	
t1b	Failure cannot be predicted with existing technology		PS	PS		M	M		M	M		M	M
t2a	Failure cannot be detected with additional research	PS	PS		M	M		M	M		M	M	
t2b	Failure cannot be predicted with additional research		PS	PS		M	M		M	M		M	M
Economic Feasibility													
e1	Insufficient financial resources	PS	PS	PS	N	N	N	N	N	N	N	N	N
e2	Not enough failures (during life time) for positive business case	PS	PS	PS	N	N	N	N	N	N	N	N	N
Organizational Feasibility													
o1	No trust in monitoring system	PS	PS	PS	N	N	N	N	N	N	N	N	N
o2	No fit to personnel	PS	PS	PS	N	N	N	N	N	N	N	N	N
o3	No fit to operational task		PS	PS		N	N		N	N		N	N
o4	No fit to relations and policies		PS	PS		N	N		N	N		N	N
o5	No fit to the spare parts		PS	PS		N	N		N	N		N	N

Table 3.2: Identification of potential showstoppers.

This is not the case since the priority components do not have to be replaced if another priority component needs a replacement. Therefore showstopper *c2* is not a thread for the diagnostics and prognostics of the priority components.

Concerning the technical feasibility of our priority components, we can conclude detection, diagnostics, and prognostics might be possible by researching the existing sensors (*t1a, t1b*). It is a project goal to detect, diagnose, and prognose failures based on the current set of sensors. Therefore, this identified showstopper does not make the CBM approach infeasible for our priority components. However, we do have to take into account that no direct sensors are measuring the length or the tension of the conveyor belt. This denotes we are dependent on the model to link specific parameter behavior related to the belt drive motors to a failure caused by the conveyor belt. By adding new sensors we enhance the predictive value that might result in a more reliable detection and prognostics model (*t2a, t2b*). Also, this showstopper does not make us think CBM might not be the right maintenance policy.

Next, the economic feasibility is important to take into account. First of all, the sensor information used in this study is already implemented on the machine which results in showstopper *e1* not being a thread. Additionally, enough failures are necessary for a positive business case. By analysing the available maintenance data, the fan seems to be failing sporadically. However, if the fan fails during a production session this could result in huge downtime costs that are hard to quantify. Therefore, the main air distribution fan is assumed to have sufficient failures for a positive business case. On the contrary, more replacements are known that can be linked to the failure or malfunction of the conveyor belt. Also, their downtime costs are high resulting in showstopper *e2* not being a thread.

Finally, the organizational feasibility is tested for our priority components. Company X has an innovative digitization program, which brings together experts at Company X from a variety of disciplines tasked with delivering proactive services built on the Internet of Things and data science. This shows Company X is proactive in data-based solutions for maintenance. Therefore, CBM fits the personnel (*o2*), the operational task (*o3*), and the relations and policies (*o4*). The 'Data Understanding' phase of the CRISP-DM (Chapter 4) is consulted to assess the quality of the monitoring system. In general, the parameters show reliable values resulting in showstopper *o1* being no threat. Lastly, this project assumes sufficient spare parts are available when components should be replaced based on maintenance decision-making by consulting CBM (*o5*).

3.3.3 Stage 3 - focused feasibility study

The final stage of the method for selecting suitable candidates for CBM studies some components where the showstoppers cannot be clearly identified (i.e. *Maybe* should be transformed in either a *Yes* or a *No*). For these components, an economic and technical feasibility study is performed. The economic feasibility study focuses on whether developing the CBM model is beneficial to the firm from a strategic point of view. Discussing the possible gains in comparison to the possible investment costs is important (Bengtsson, 2008). On the other hand, the technical feasibility study focuses on whether the firm can develop and implement the desired approach. The following stages should be accomplished to assess the technical feasibility: data acquisition, data processing, detection, diagnostics, prognostics, and decision analysis (Tiddens et al., 2018).

The economic feasibility study is accomplished with the FMEA results and the statement given in the scope of this research. This statement is to focus on candidates that can already be monitored with existing sensors in the Machine Y leading to minimal investment costs. Sequentially, as already mentioned in Section 3.1.1, the higher the RPN, the more critical the failure is. Therefore, these failures should be given greater attention for risk mitigation. This project mainly focuses on improving detection by implementing automatic detection and diagnostics models. The FMEA detection scores attached to specific failures are given in Figure 3.4. If the cause can be prevented from happening the detection score is one. A detection score of three is established when the cause can be detected. A detection score of seven is given when the failure mode can be detected. Lastly, when only the effect can be detected a score of ten is given. This denotes a failure can be detected very late in the process. As can be seen in Figure 3.5, the RPN reduces when automatic detection and diagnostic models can be implemented. This insight gives confidence in economically feasible priority components regarding the implementation of CBM. The technical feasibility is discussed in the 'Data Understanding' phase of the CRISP-DM model (Chapter 4).

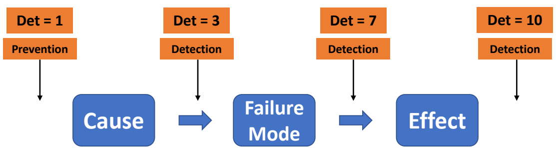


Figure 3.4: Detection scores.

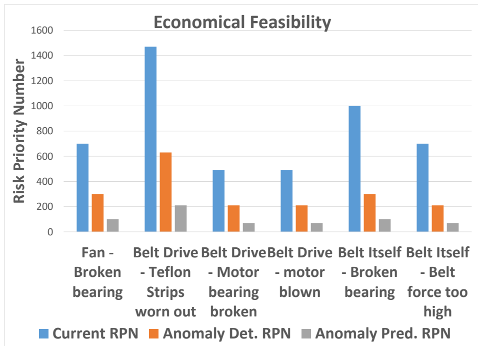


Figure 3.5: Economical feasibility.

3.4 Conclusion

In this section, we answer research question 2: *What are suitable Machine Y candidate components for CBM?* Together with the service and innovation department of Company X, we distinguished three priority components (i.e. main air distribution fan, conveyor belt, and conveyor belt drive motors) that were assessed on their suitability for CBM purposes. After analysing various methodologies to find suitable candidates for CBM, the three-stage funnel developed by Tiddens et al. (2018) has been argued to best fit our research.

The basic idea behind Figure 3.2 is to filter out components that do not seem to be beneficial for CBM in each stage of the funnel. The first stage is successfully accomplished by all priority components as they do not fail frequently and have a corresponding high downtime (cost). However, the main air distribution fan has extremely scarce maintenance data available and seems to be a robust component. As we know that not all customers shared their maintenance data and the FSEs do not always archived their maintenance actions, we assume the frequency of failure of this priority component is sufficient for a positive business case. The second stage of the funnel identified potential showstoppers mainly related to technical feasibility. However, since the project goal is to research if enabling CBM is possible, we do not label this potential showstopper as a showstopper but as a reason why we conduct this research. The third stage of the funnel shows a *numerical* advantage (i.e. risk mitigation of priority component failure) of automatic detection and diagnostics over the current situation.

The FMEA was conducted to support the statements made in the three stages of the funnel of Figure 3.2. However, the FMEA appeared to have additional advantages that can be used in the remaining of this research. These benefits are providing structure in finding sensors that have the highest potential in increasing the predictive value that is further discussed in the recommendations (Section 8.2). Another benefit is building a knowledge base for anomaly diagnostics as the failure modes can be the labels for anomaly diagnosis modeling (Section 7). The FMEA also identified a risk for the conveyor belt as no direct sensors were available for measuring its tension and length. This implies we are dependent on the diagnostic model that should link specific belt drive motor behavior to the root cause of the conveyor belt failure.

The risks of enabling CBM on the distinguished priority components have been refuted and each priority component has successfully accomplished the process displayed in Figure 3.2. Therefore, it can be concluded the priority components *main air distribution fan*, *conveyor belt*, and the *conveyor belt drives* are suitable for CBM purposes. However, we do have to take into account two out of three distinguished priority components have associated risks that can complicate enabling CBM for these components.

Chapter 4

Data understanding

Understanding the strengths and limitations of the data is key since there is rarely an exact match with the problem. Historical data often are collected for purposes unrelated to the current business problem, or for no explicit purpose at all (Provost and Fawcett, 2013). Company X's goal to monitor Machine Y behavior was to better understand how the parameters behave when a failure happened (i.e. not for CBM purposes). When customers have Machine Y problems, data experts first look at the parameter behavior to diagnose the fault and (if possible) remotely help the customer to solve the issue before they fly over. Company X has software systems available, containing data of the condition and component replacements of Machine Y. In this research, we utilize available data (Section 1.2.6) and do not focus on collecting new sensor data. Therefore, this chapter aims to find an answer to Research Question 3:

Can the available data enable CBM for the priority components of Machine Y?

The data understanding phase in the CRISP-DM model starts by collecting initial data and get familiar with it. Section 4.1 describes the available data generated by Machine Y. Section 4.2 applies descriptive analytics to explore the data concerning the priority components and list the key findings of exploring the data. Lastly, Section 4.3 concludes and provides a risk assessment of the data understanding stage.

4.1 Condition and event data

According to Jardine et al. (2006), data collected for CBM purposes can be categorized into two main types: event data and condition monitoring data. Condition monitoring data are the measurements related to the health condition/state of the physical asset. Event data include the information on what happened (e.g., installation, breakdown, overhaul, warnings, error codes, etc.) and what was done (e.g., minor repair, preventive maintenance, oil change, etc.) to the targeted physical asset. Company X possesses an IT infrastructure that allows collecting condition and event data from a processing plant and sending it to Company X control tower for analysis. Data has been collected from thirteen Machine Y systems over the years 2017 through 2020 (Table 4.1).

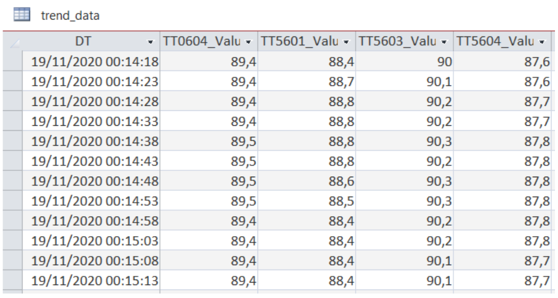
Every five seconds each day of the week about 200 condition-related Machine Y parameters are logged. In total 104,369,270 condition-related data records are available for this data project. These condition monitoring data consist of measurements in Machine Y regarding temperatures, dew point, pressures, motor power consumption, motor current consumption, motor speed, target belt speeds, target dwell times, etc. An illustration of how the condition dataframe looks like is given in Figure 4.1. This Figure shows that every five seconds a record is logged consisting out of parameter names with their corresponding monitored value. The most important parameter names for this research are related to the state of the machine and the priority components monitoring sensors. Machine Y can be in ten different machine states that can be found in Table 5.4. Furthermore, the conveyor belt can be in multiple belt states as well. Moreover, the selected

Customer	Years of data available	Number of daily files available	Data records available
1	Feb 2020 - Nov 2020	240	3,968,392
2	Feb 2020 - Nov 2020	270	4,653,796
3	Jan 2019 - Aug 2019	224	2,322,553
4	Jan 2019 - Nov 2020	603	10,383,103
5	May 2019- Jan 2020	227	3,915,413
6	Jan 2019 - Nov 2020	673	11,526,508
7	Nov 2019 - Nov 2020	384	6,612,924
8	Jun 2018 - Nov 2020	308	5,262,755
9	Sep 2019 - Nov 2020	383	6,684,561
10	Sep 2019 - Nov 2020	357	6,130,494
11	Jan 2017 - Dec 2020	995	16,931,907
12	Apr 2018 - Sep 2020	893	15,374,303
13	Apr 2018 - Jan 2020	617	10,602,561

Table 4.1: Available condition data.

parameters related to the fan and conveyor belt can be found in Table 5.2 and 5.3, respectively, and are extensively elaborated in Section 5.2.1.

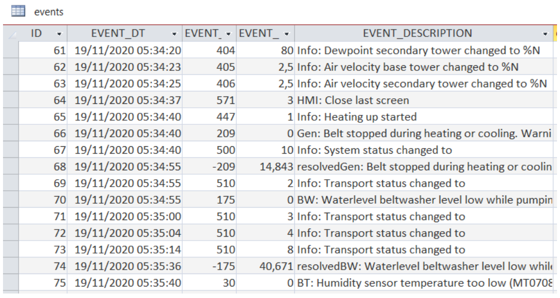
Moreover, Machine Y generates event data that includes operator machine program setting changes and errors. Examples of machine program setting changes can be modifications in target belt speeds, temperatures, dewpoints, or dwell times because of different products being processed by the machine. Event data also includes moments when operators press a button to stop the belt or stop the machine completely. Furthermore, some rules are implemented in the software to automatically provide alerts or perform actions if a specific threshold has been reached. An example of such an event is a belt detection sensor that stops the belt when the sensor is triggered. An illustration of how Machine Y event dataframe looks like is given in Figure 4.2. This Figure shows that the specific moment in time is logged of a specific event. Besides, the event data includes the maintenance data from the CMD, ERP system, and CXMDS (Section 2.3).



trend_data

DT	TT0604_Valu	TT5601_Valu	TT5603_Valu	TT5604_Valu
19/11/2020 00:14:18	89,4	88,4	90	87,6
19/11/2020 00:14:23	89,4	88,7	90,1	87,6
19/11/2020 00:14:28	89,4	88,8	90,2	87,7
19/11/2020 00:14:33	89,4	88,8	90,2	87,7
19/11/2020 00:14:38	89,5	88,8	90,3	87,8
19/11/2020 00:14:43	89,5	88,8	90,2	87,8
19/11/2020 00:14:48	89,5	88,6	90,3	87,8
19/11/2020 00:14:53	89,5	88,5	90,3	87,8
19/11/2020 00:14:58	89,4	88,4	90,2	87,8
19/11/2020 00:15:03	89,4	88,4	90,2	87,8
19/11/2020 00:15:08	89,4	88,4	90,1	87,7
19/11/2020 00:15:13	89,4	88,4	90,1	87,7

Figure 4.1: Condition data example.



events

ID	EVENT_DT	EVENT_	EVENT_	EVENT_DESCRIPTION
61	19/11/2020 05:34:20	404	80	Info: Dewpoint secondary tower changed to %N
62	19/11/2020 05:34:23	405	2,5	Info: Air velocity base tower changed to %N
63	19/11/2020 05:34:25	406	2,5	Info: Air velocity secondary tower changed to %N
64	19/11/2020 05:34:37	571	3	HMI: Close last screen
65	19/11/2020 05:34:40	447	1	Info: Heating up started
66	19/11/2020 05:34:40	209	0	Gen: Belt stopped during heating or cooling. Warni
67	19/11/2020 05:34:40	500	10	Info: System status changed to
68	19/11/2020 05:34:55	-209	14,843	resolvedGen: Belt stopped during heating or coolin
69	19/11/2020 05:34:55	510	2	Info: Transport status changed to
70	19/11/2020 05:34:55	175	0	BW: Waterlevel beltwasher level low while pumpin
71	19/11/2020 05:35:00	510	3	Info: Transport status changed to
72	19/11/2020 05:35:04	510	4	Info: Transport status changed to
73	19/11/2020 05:35:14	510	8	Info: Transport status changed to
74	19/11/2020 05:35:36	-175	40,671	resolvedBW: Waterlevel beltwasher level low whil
75	19/11/2020 05:35:40	30	0	BT: Humidity sensor temperature too low (MT070

Figure 4.2: Event data example.

4.2 Data exploration

An important question that should be asked at the beginning of every data project is if the past represents the future. Since there are no major modifications planned and the analysed machines do not reach their end-of-life in the near future, we can assume the historical data represent the future behavior of Machine Y. The main goal of exploring the data is to get a better representation of the available data. Descriptive analytics are useful in exploring the available data (Olson and Lauhoff, 2019).

Different tools have been used to get an overview of the available data. Line charts, boxplots, scatterplots, and correlation matrices have been consulted to find if we have enough data available to for enabling CBM for the priority components, if the sensors record the correct values, if there are outliers in the data, if the timestamps are correctly logged in Machine Y data files, etc. Outliers can be detected but being cautious in making conclusions is key. Sections 5.1.2 and 5.1.3 dive deeper into how to specifically deal with potential outliers and procedural errors in our research environment. However, we can conclude the dataframes are well-structured in general since the data has a clear homogeneous structure where the elements are of the same data type. The condition data are saved at each customer daily, no redundant parameters are distinguished and only a small fraction of the data consists out of outliers due to procedural error.

4.2.1 Priority components

The first step taken in understanding the data of the priority components is finding the sensors corresponding to the specific priority component. Utilizing expert knowledge and a correlation analysis resulted in the selection of parameters for the distinguished priority components. This selection procedure is extensively elaborated in Section 5.2.1 and can be used to make the observations explained in the next three paragraphs. These paragraphs are important for assessing the data quality. However, Chapter 5 further elaborates the specific actions performed related to the observations made in the next three paragraphs.

4.2.1.1 Main air distribution fan

In the data exploration phase, we found that the parameters of the current and power of Customer 4's base tower fan recorded wrong values (Appendix B). Figures B.4, B.5 show reliable values until July 2019 after which the current and power behavior show stable values that cannot represent reality according to normal fan behavior. Apart from this observation, scatterplots of the fan motor gave us the insight data points exist that have a power higher than zero while receiving no motor current which is physically not possible according to Ohm's law (Figure B.6, B.7).

4.2.1.2 Conveyor belt

When a failure is distinguished that is connected to the conveyor belt it can either be caused by broken belt drives or by a broken conveyor belt. Unfortunately, there is no sensor information about the length or the tension of the conveyor belt. This denotes that we are dependent on the models to find a relation between the sensor data of the conveyor belt drives and the maintenance data of a conveyor belt-related replacement or failure.

4.2.1.3 Conveyor belt drive

After analysing the line charts of the motor speed, two customers have been found from which the base and secondary drum tower speed sensors are broken (Figure B.8, B.9). Besides, one customer has a period of time where it structurally monitored too high values for the motor speed of the base drum tower (Figure B.10). The encoder of that tower could be broken according to an expert. Analysing the line plots in Figure B.8 and B.9 made us aware another type of procedural error arises. Some customers have data available from the year 1998 which is not possible thus means a wrong timestamp is linked to the logging. Moreover, not all customers monitor the current of the drum motors equally. Two clients have a scaled measure that is calculated by dividing the power by the speed. The remaining clients have the correct way of logging the current of the drum towers. Utilizing domain knowledge made clear no translation is possible for these scaled parameters to return to their natural form which means some data points cannot be used for this research.

4.2.2 Key findings

4.2.2.1 Multicollinearity in data

The priority components have some highly correlated parameters as can be seen in the heatmaps shown in Appendix C.1 and C.2. Data, where the independent variables are highly correlated, is said to have multicollinearity (Piramuthu, 2008). Section 6.2.5 makes use of a supervised machine learning technique in the form of a regression analysis. If the goal is to understand how the single independent (X) variables impact the dependent variable (Y), then multicollinearity is a problem. One problem is that the individual p-values can be misleading (a p-value can be high, even though the variable is important). The second problem is that the confidence intervals on the regression coefficients will be very wide. The confidence intervals may even include zero, which means one can't even be confident whether an increase in the X value is associated with an increase, or a decrease, in Y. Because the confidence intervals are so wide, excluding a subject (or adding a new one) can change the coefficients dramatically and may even change their signs. However, if the goal is simply to predict Y from a set of X variables, then multicollinearity is not a problem. The predictions will still be accurate, and the overall R^2 (or adjusted R^2) quantifies how well the model predicts the Y values (Paul, 2006).

The selection of relevant predictors could be an alternative for eliminating multicollinearity in the data (Section 5.2.1). The presence of multicollinearity is not of concern for decision trees (Piramuthu, 2008) and for logistic regression (Ray, 2019). The attempt to remove multicollinearity resulted in poor model performance for these models. On the other hand, Artificial Neural Networks (ANN) cannot automate the selection of relevant predictors and hence is not appropriate to be applied on multicollinearity in the data (Garg and Tai, 2012). The key finding of multicollinearity in parameters should be taken into account in the data and model selection in Section 6.2.1 and 6.2.2, respectively.

4.2.2.2 Measurement time frequency interval

After visually inspecting the time-series data of sensor behavior, we found that some peaks can be distinguished. Figure 4.3 shows an example of such a peak which gave us the impression even more peaks could be distinguished if we shorten the time interval of measurement frequency. Since Company X started monitoring the Machine Y condition, it automatically logged the parameter behavior every five seconds. To assess if all interesting behavior is included in this sample frequency of five seconds, a test has been set up where the frequency of measuring has been increased to every 100 milliseconds. Two customers (i.e. Customer 11 and Customer 5) have increased their monitoring frequency for two weeks (i.e. 16-3-21 - 30-3-21). A test period of two weeks among two customers is assumed to be sufficient to find significant differences between 0.2 and 10 Hertz measurements. Furthermore, 10 Hertz has been chosen as shorter time interval of measurement frequency as the VPN infrastructure at Company X cannot transfer data at higher Hertz nowadays. These 10 Hertz data contain more detailed information and transitions in parameter behavior are smoother than the original measuring frequency (i.e. 0.2 Hertz). However, comparing the time-series data of five-second to the 100-millisecond frequency of measurement did not lead to significant changes in interesting peaks but the higher Hertz the more information can be captured. As the 100-millisecond option is more memory intensive than the five-second option and we have not distinguished significant advantages in the more frequent measuring interval, we can state the sampling frequency of five seconds provide us the right information for finding anomalous behavior of priority components.

However, the test period of a different frequency of measurement interval is rather short which could mean no interesting anomalous behavior took place during this two-week test period which resulted in seeing no difference between the two time frequency measurement intervals.

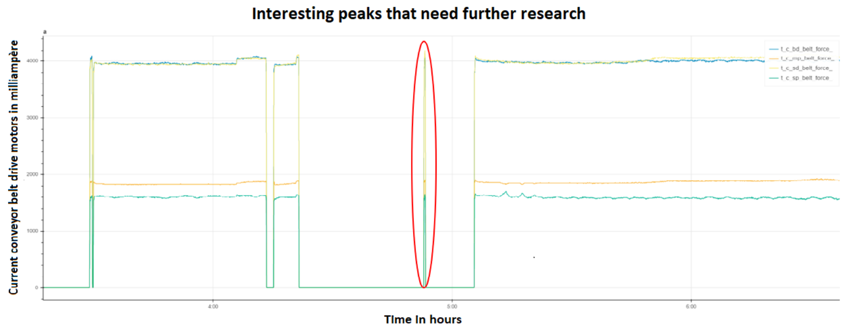


Figure 4.3: Short peaks in parameter behavior that need further research.

4.3 Conclusion

In this section, we answer research question 3: *Can the available data enable CBM for the priority components of Machine Y?* by listing the strengths and limitations of the dataset. We can conclude the condition-monitoring data is of high quality and quantity and sensors are available that monitor the behavior of the priority components (i.e. main air distribution fan and the belt drive motors). Furthermore, we assess the historical data represents the future behavior of Machine Y and the available data consisting of five-second frequency of measurements contains sufficient information for finding anomalous behavior.

Extra attention should be paid to the data limitations as they can harm the results of this research. A key data limitation is the lack of sensors to monitor the conveyor belt behavior (Section 3.4). Furthermore, the risk of a sporadically malfunctioning main air disturbing fan can result in too little learning instances for anomaly detection and diagnosing (Section 3.4). The distinguished priority components have multicollinearity in their parameter values which can still result in accurate predictions but feature importances cannot be reliably interpreted. Moreover, the statement made in Section 2.1 that multiple machine types exist is not specifically a limitation but a situation to take into account in the modeling stage.

Succeeding this risk assessment in the data understanding phase and refuting the limitations gives us confidence in reliable data that can be used to solve the business problems which enables the answer: *Yes, the available data can enable CBM for the priority components of Machine Y.*

Chapter 5

Data preprocessing

The analytic technologies we use in Chapter 6 and 7 are powerful but they need data that fulfills certain requirements. They often require data to be converted in a form different from how the data are provided naturally. Therefore, this chapter aims to answer Research Question 4:

How to preprocess the available raw data so it can be used for anomaly detection and diagnostics?

According to García et al. (2015), data preprocessing consists of data preparation and data reduction. An overview of the data preprocessing stage is given in Figure 5.1 where the blue blocks represent data preparation steps and the orange blocks represent data reduction steps. The data preprocessing phase in the CRISP-DM model starts with Section 5.1 where all the sets of techniques are explained and applied to prepare the dataset for modeling. Hereafter, Section 5.2 explains and applies the set of techniques to obtain a reduced representation of the original data. Lastly, Section 5.3 concludes the data preprocessing stage.

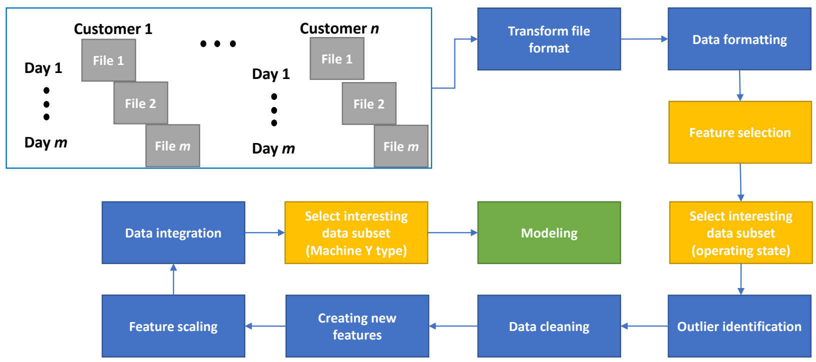


Figure 5.1: Data preprocessing steps.

5.1 Data preparation

According to García et al. (2015), data preparation comprises all the sets of techniques that initialize the data properly to serve as input for a certain data mining algorithm. The set of techniques are data cleaning, data transformation, data integration, data normalization, missing data imputation, and outlier identification. Hence, data preparation converts raw data into new data that fits into a data mining process. The following data preparation steps have been studied in our case. First, the daily generated files by every customer are transformed to a format that is faster to read-in and has a higher data capacity. Hereafter, the data quality should be raised to the level required by the models explained in Chapter 6 and 7. Therefore, the following sections describe the data preparation steps necessary to create an integrated dataframe that is suitable for modeling purposes.

5.1.1 Data formatting

Every daily file containing Machine Y data that is generated by a customer gets a file name where the specific day is mentioned. This process went well for all the customers except for Customer 3. This customer's specific timestamps of a file were not coherent with the file name. Therefore, the names of the files are formatted to the most frequently occurring day in a specific file. It seemed that many records contain the same timestamp due to documentation error. This reduced the total of 224 files to 134 relevant files linked to Customer 3. Apart from that, every timestamp has been labeled to the specific customer for transparency purposes.

5.1.2 Outlier identification

Investigating the influence of errors, outliers, and noise is key when aiming to build good-performing models. In this section, we define the multiple classifications of outliers and design a way how to deal with these types of outliers. If the data contains errors, outliers, and noise (e.g., due to poor quality measurements), difficulty arises for the system to detect the underlying patterns, so the system is less likely to perform well. These errors and noise can have harmful effects on data mining analyses (Osborne and Overbay, 2004). According to Hair (2009), outliers can be classified into one of three classes based on the source of their uniqueness (Table 5.1).

Outlier types	**Procedural error**	**Extraordinary event**	**Unique in their combination**
More detailed explanation of outlier types	e.g. a data entry error or a mistake in coding.	This type of outlier is the observation that occurs as the result of an exceptional event, which accounts for the uniqueness of the observation.	This type of outlier contains observations that fall within the ordinary range of values on each of the variables. These observations are not particularly high or low on the variables, but are unique in their combination of values across the variables.
Performed action	This type of outlier should be fixed or removed in the data cleaning stage (Osborne and Overbay, 2004)	Research is necessary to distinguish interesting extraordinary events (i.e. machine malfunction) from uninteresting extraordinary events (i.e. operator or climate changes)	Research is necessary to distinguish interesting events (i.e. machine malfunction) from uninteresting events (i.e. operator or climate changes)
Section where action is performed	Data cleaning (Section 5.1.3)	Anomaly detection (Chapter 6)	Anomaly detection (Chapter 6)

Table 5.1: Outlier types with their application-specific action.

Several approaches have been studied in the literature to deal with noisy data and to obtain higher classification accuracies on test data (García et al., 2015). Among them, the most important are: (i) Robust learners: These are techniques characterized by being less influenced by noisy data explained in Chapter 6. (ii) Noise filters: identify noisy instances which can be eliminated from the data explained in Section 5.1.3. Filtering the data has also one major drawback: some instances will be dropped from the data sets, even if they are valuable. According to Hair (2009), outliers should be retained unless demonstrable proof indicates that they are truly aberrant and not representative of any observations in the population. If they do portray a representative element or segment of the population, they should be retained to ensure generalizability to the entire population.

5.1.3 Data cleaning

Data cleaning is the process of fixing or removing procedural errors from the dataset (Table 5.1). Therefore, domain knowledge is utilized for discussing the findings in the data exploration phase (Section 4.2) to clean the data. The fan and belt drives have different actions related to cleaning the data as can be seen in Figure 5.3 and Figure 5.4, respectively. We aim to transform the specific procedural error (i.e. corresponding to one or a few values out of the whole data record) to an empty value instead of deleting the whole records to keep maximal relevant data in the dataframe (Figure 5.2).

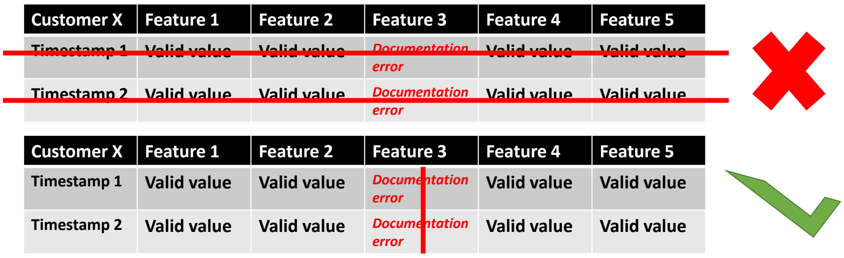


Figure 5.2: Data cleaning process.

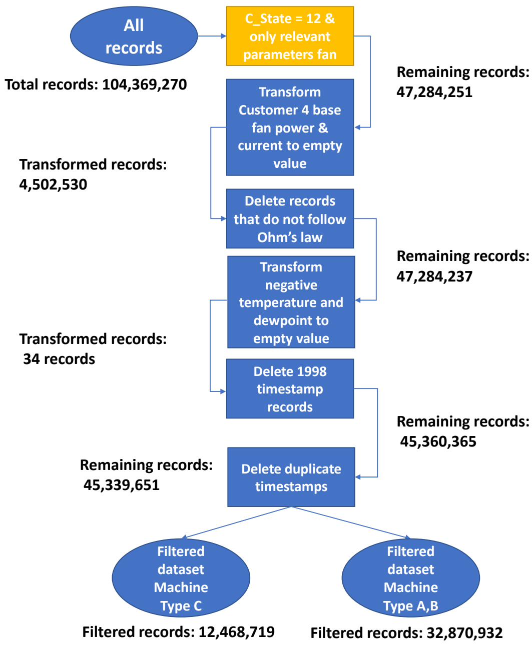


Figure 5.3: Data filtering process fan.

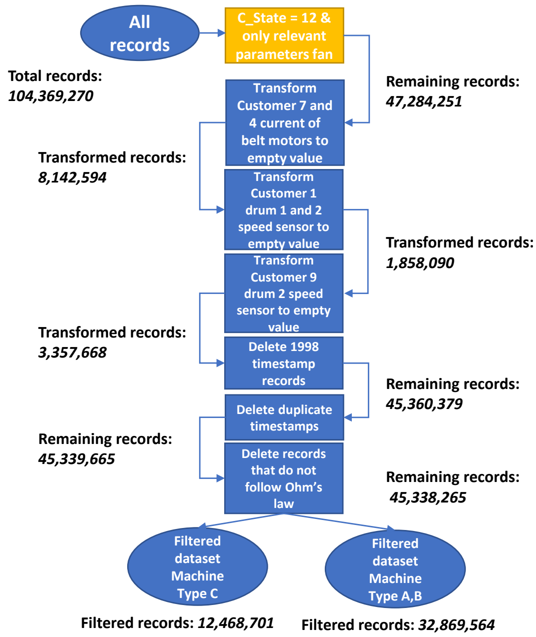


Figure 5.4: Data filtering process belt.

We transformed Customer 4's data regarding the consumed base fan motor power and current to empty values as the original values do not represent the true consumed power and current. Regarding the conveyor belt drive motors, Customer 7 and Customer 4 calculated the current consumption in different ways which was impossible to recover according to experts. This implies the current values of two customers are transformed to empty values. In addition, Customer 9 and Customer 1 had a broken sensor of the speed measurement of the base and secondary drum. This is no problem for the customer since the drum engines can operate in an 'open loop' which means they cannot be controlled but still function well. However, for the data analysis these values for speed monitoring of Customer 9 and 1 are transformed to empty values.

Furthermore, some customers recorded their parameters with the wrong timestamp. Some customers have records of data in 1998 while that is not possible since Machine Y did not exist back then. Since the timestamps are wrongly monitored they cannot be compared to the maintenance data and therefore deleted from the dataframe. Another procedural error identified was that some customers contain duplicate timestamps. Rolling window calculations (Section 5.1.4.1) cannot deal with duplicate values so only the first occurrence of duplicate timestamps is used while others are deleted from the dataframe. According to Ohm's law, the power cannot be positive if the current and voltage are zero. Therefore, these datapoints are deleted from the dataframe since they can be categorized as documentation errors due to the physical property of power. Additionally, the temperature and dew point sensors sometimes record a negative value which is not possible in Machine Y. This denotes the sensor is broken and will therefore be transformed to empty values.

5.1.4 Creating new features

Feature engineering is the act of extracting features from raw data and transforming them into formats that are suitable for the machine learning model (Zheng and Casari, 2018). This is a crucial step in machine learning because the right features can ease modeling, and therefore enable to output results of higher quality. The right features can only be defined in the context of both the model and the data. A feature is a numeric representation of raw data.

5.1.4.1 Rolling window univariate statistics

A conclusion drawn from the FMEA (Section 3.1.1) was that a fluctuating power could be an indicator of a failed priority component. Therefore, for the two fan motors and four belt drive motors, a rolling window standard deviation and min-max range of the power is calculated and added to the dataframe that takes the fluctuation of powers into account (Zivot and Wang, 2003). When the rolling window univariate statistics are implemented, the machine learning model can make use of a temporal feature. The standard deviation is a measure of the amount of variation or dispersion of a set of values (Bland and Altman, 1996), while the min-max range calculates the difference between the highest and lowest value of a set of values. It is assumed these two motions contain sufficient information to describe the fluctuating parameter behavior. The hyper-parameter *set of values* is chosen with the intention to find both small and larger peaks and gabs. We can take the parameter behavior into account of the last *5* minutes to find small peaks and/or falls and the rolling window of the last *30* minutes to find larger peaks and/or falls. Section 6.4 discusses which motion performs better in detecting anomalies. The created rolling window features are listed for the fan and belt drive motors in Table C.2 and C.4, respectively.

The rolling window calculations can best be explained according to the following formal definition. Consider the analysis of a univariate time series parameter y_a^x over a sample from $a \in \{1, ..., T\}$ where y is the power parameter and x is the specific motor of the priority component. To assess the fluctuation, let n denote the width of a window and define the rolling sample mean $\hat{\mu}_t^x(n)$, rolling sample standard deviations $\hat{\sigma}_t^x(n)$, and rolling sample min-max range $\hat{\omega}_t^x(n)$ for windows $t = n, ..., T$. The rolling standard deviation and min-max range estimates at time t with window width n are the usual sample estimates using the most recent n observations. Two window widths are defined $n = 60$ and $n = 360$ corresponding to the amount of data records of a 5 minute-window and 30 minute-window. Provided the windows are rolled through the sample one observation at a time, there will be $T - n + 1$ rolling estimates of each parameter.

$$\hat{\mu}_t^x(n) = \frac{1}{n} \sum_{i=0}^{n-1} y_{t-i}^x \tag{5.1}$$

$$\hat{\sigma}_t^{x^2}(n) = \frac{1}{n-1} \sum_{i=0}^{n-1} (y_{t-i}^x - \hat{\mu}_t^x(n))^2 \tag{5.2}$$

$$\hat{\sigma}_t^x(n) = \sqrt{\hat{\sigma}_t^{x^2}(n)} \tag{5.3}$$

$$\hat{\omega}_t^x(n) = \max_{t \in n} y_t^x - \min_{t \in n} y_t^x \tag{5.4}$$

5.1.4.2 Operating time after being in standby mode

Analysing the time-series data when Machine Y, following a traditional method of processing, starts after being in a standby state provides the insight that belt motors need more power at the starting phase (Figure 5.5). As the belt moves for a longer time in an operating state, the belt gets automatically lubricated by the products processed resulting in a smoother movement through the machine. Therefore, an additional feature has been created that includes the operating time of the machine when it enters an operating state after being in a standby state (Table C.4 record one). The machine learning model can use this information for prediction purposes of the power of the belt motors.

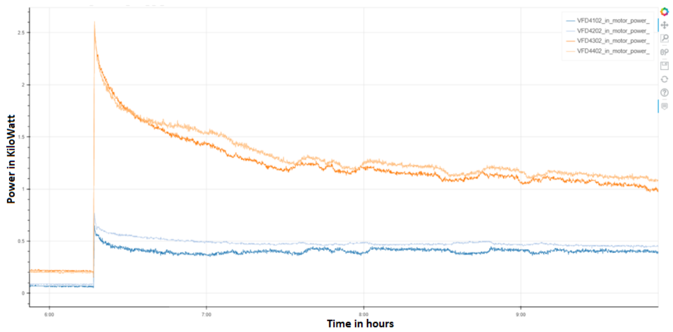


Figure 5.5: Power decrease after lubrication process.

5.1.4.3 Relative power belt drive motors

An extra feature that applies to the belt drive motors but not to the main air distribution fan motors is the dependency of the motors. The fans are implemented in two towers of Machine Y that can reach different climates and conditions. Therefore, the fan motors cannot be compared to each other as there are different conditions applicable. However, the drawing in Figure 5.6 shows the belt drive motors are all connected because of the conveyor belt. Therefore, the relative power of the motors should be the same in a normal situation. The slave pickup motor (orange encircled) is located at the infeed side of Machine Y. Thereafter, the belt is sequentially moved by the drum tower 1 motor (yellow encircled), drum tower 2 motor (blue encircled), and the master pickup motor (green encircled). Hence, four new features are added that correspond to the relative power of belt drive motors to each other (Table C.4 number 18 until 21).

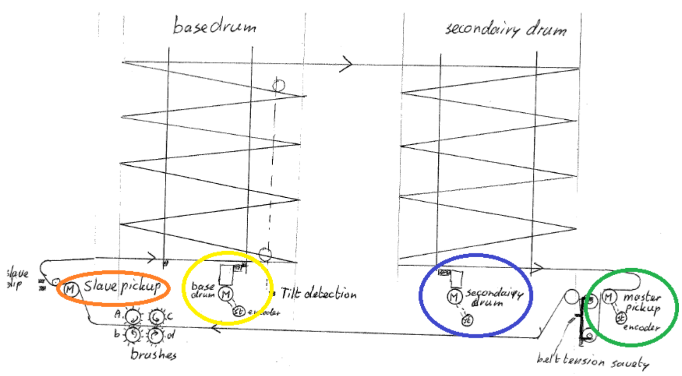


Figure 5.6: Conveyor belt movement.

5.1.5 Feature scaling

Some ML algorithms need their features to be scaled in order to work properly. Especially, ML techniques that calculate distances between data take advantage of this important data transformation. Several modeling techniques are explained in Chapter 6 and 7. All these models include a discussion in their description whether scaling is necessary or not. According to Géron (2019), there are two common ways to scale the data: (i) min-max scaling and (ii) standardization. Min-max scaling (also named normalization) shifts values and rescale them so they end up ranging from 0 to 1 which is done by subtracting the min value and dividing by the max minus the min. Min-max scaling is very sensitive to outliers. Conversely, standardization subtracts the mean value and then divides by the standard deviation so that the resulting distribution has unit variance. Standardization is less affected by outliers. As standardization is less affected by outliers than min-max scaling, the standardization methodology is used for scaling.

5.1.6 Data integration

After the previously elaborated data preparing steps have been succeeded, the data of all customers having the same type of Machine Y should be integrated into one clear dataframe that can be used for modeling. So the integrated dataframe contains a dataframe that has been formatted and cleaned. Furthermore, the new features found in Section 5.1.4 are also integrated into the final dataframe. Now, we have reached a richer dataset that can be used for modeling.

5.2 Data reduction

Data reduction comprises the set of techniques that obtain a reduced representation of the original data. The set of techniques connected to data reduction are feature selection and instance selection that are elaborated in Section 5.2.1 and 5.2.2, respectively.

5.2.1 Feature selection

The goal of feature selection is to identify the features in the data set which are important. It facilitates the understanding of the pattern extracted and increases the speed of the learning stage. Features can be selected by utilizing expert knowledge (Section 5.2.1.1) or automatically selected by data analysis (Section 5.2.1.2). A conclusion on what method fits best to our data is given in Section 5.2.1.3.

5.2.1.1 Utilizing expert knowledge

While the purpose of machine learning in many cases is to avoid having to create a set of expert-designed rules, that does not mean that prior knowledge of the application or domain should be discarded. Domain experts can help in identifying useful features that are much more informative than the initial representation of the data (Müller and Guido, 2016). Interviewing subject matter experts resulted in interesting attributes related to the main air distribution fan and the belt drive motors that are shown in Table 5.2 and 5.3, respectively.

5.2.1.2 Automatic feature selection

Since there may be complex and unknown relationships between the variables in the dataset that the experts are not familiar with, automatic feature selection methods are consulted. Two automatic feature selection methods (i.e. univariate statistics, model-based selection) are discussed in this section.

In univariate statistics, we compute whether there is a statistically significant relationship between each feature and the target (Müller and Guido, 2016). We chose an important feature as the target that came out of the FMEA analysis (Section 3.1.1): the power of the fan motor and

Attribute	Description	Type
VFD0502_in_Motor_current	Actual current main fan zone 1	Ampere
VFD5502_in_Motor_current	Actual current main fan zone 2	Ampere
VFD0502_in_motor_power	Actual power main fan zone 1	KiloWatt
VFD5502_in_motor_power	Actual power main fan zone 2	KiloWatt
VFD0502_in_motor_speed_hz	Actual frequency main fan zone 1	Hertz
VFD5502_in_motor_speed_hz	Actual frequency main fan zone 2	Hertz
R_active_recipe_airspeed_BT_value	Target airspeed zone 1	MetersPerSecond
R_active_recipe_airspeed_ST_value	Target airspeed zone 2	MetersPerSecond
B_T_T_air	Actual temperature zone 1	Celsius
S_T_T_air	Actual temperature zone 2	Celsius
mt0708_dew_value	Actual dewpoint zone 1	Celsius
mt5708_dew_value	Actual dewpoint zone 2	Celsius

Table 5.2: Fan important parameters utilizing expert knowledge.

the power of the belt drive motors. Automatic feature selection consulting univariate statistics looks for highly positively or negatively correlated features that have a predictive value towards the target variable (Provost and Fawcett, 2013). Hence, we took a random sample ($n = 100,000$) of states where Machine Y is operating of all customers and performed a correlation analysis of all parameters related to the power consumption of the fan motors and belt drive motors. The five parameters with the highest positive or negative correlation are selected resulting in Table C.1 for the fan and Table C.3 for the belt drives. Both tables show the experts distinguished the correct parameters linked to the behavior of the priority components as the parameters that score high in the correlation analysis are also mentioned by the experts. So, performing univariate statistical analysis did not result in new features compared to interviews with experts.

On the other hand, model-based feature selection is also frequently used in literature. This feature selection method uses a supervised machine learning model to judge the importance of each feature and keeps only the most important ones (Müller and Guido, 2016). Decision trees and decision tree–based models provide a feature importance attribute, which directly encodes the importance of each feature. In contrast to the univariate selection, model-based selection considers all features at once, and so can capture interactions. Firstly, the important features for the fans and the belt drives are tested for multicollinearity as Section 4.2.2.1 elaborates datasets with multicollinearity cannot provide reliable feature importances. To assess multicollinearity among parameters, literature prescribes to create a correlation matrix (Figure C.1 for the fan and Figure C.2 for the belt drives) and look for highly correlated parameters. As the power, current, and speed of the fan and belt drive motors are highly correlated, we deal with multicollinearity resulting in the model-based feature selection method that cannot be used to find important parameters for our priority components.

5.2.1.3 Conclusion

It is important to not only look at the correlation-based feature selection as features with low correlation to the target can still have a predictive value in some circumstances. Moreover, the correlation-based feature selection method only takes into account linear relationships while there might be non-linear relationships that can explain the power behavior of our priority component. Therefore, the main emphasis for selecting the important features for the priority components is laid on utilizing expert knowledge while correlation-based feature selection is used for validation purposes.

Attribute	Description	Type
T_sd_dev_block_rpm_pv	Process value rpm drum motor zone 2	RevolutionsPerMinute
T_bd_dev_block_rpm_pv	Process value rpm drum motor zone 1	RevolutionsPerMinute
T_mp_dev_block_rpm_pv	Process value rpm master pick-up motor	RevolutionsPerMinute
T_SP_H_out_hz	Target rpm slave pick-up motor	RevolutionsPerMinute
R_config_Belt_Overdrive	Belt overdrive	Percentage
t_c_bd_belt_force	Actual current drum drive zone 1	MilliAmpere
t_c_mp_belt_force	Actual current master pickup drive	MilliAmpere
t_c_sd_belt_force	Actual current drum drive zone 2	MilliAmpere
t_c_sp_belt_force	Actual current slave pickup drive	MilliAmpere
T_B_Beltspeed_m_s	Target belt speed	MetersPerSecond
C_State	Status of the machine	Integer
T_c_status	Status of transport	Integer
T_A_Dwelltime	Actual dwell time	Seconds
R_active_recipe_dwelltime_Value	Target dwell time	Seconds
VFD4102_in_motor_power	Master pickup motor power	KiloWatt
VFD4202_in_motor_power	Slave pickup motor power	KiloWatt
VFD4302_in_motor_power	Drum motor zone 1 power	KiloWatt
VFD4402_in_motor_power	Drum motor zone 2 power	KiloWatt
FV0712_value	Control valve steam zone 1	Percentage
FV5712_value	Control valve steam zone 2	Percentage

Table 5.3: Belt drive motors important parameters utilizing expert knowledge.

5.2.2 Instance selection

This phase of the data reduction stage consists of choosing a subset of the total available data to achieve the original purpose of the data mining application as if the whole data has been used (García et al., 2015). According to Liu and Motoda (2002), instance selection has the following functions: (i) Focusing on the relevant part of the whole amount of data which is explained in Section 5.2.2.1. (ii) Section 5.2.2.2 elaborates how to deal with the phenomenon when the data set is too huge and it may not be possible to run a data mining algorithm, or the data mining task might not be able to be effectively performed.

5.2.2.1 Interesting data subset

The severity of a failure of Machine Y got the highest value when the core temperature of a product cannot be reached during an operating cycle (Table 3.1). Therefore, the data analysis primarily searches for anomalies and diagnosing them for timestamps where the machine is in the operating state. This implies the dataset containing information about all machine states (Table 5.4) can be filtered on $C\_State = 12$.

Furthermore, we made a distinction in modeling between Machine Y types because of different motor types and the conveyor belt width which is different for both Machine Y types probably resulting in other degradation paths for our priority components.

5.2.2.2 Sampling

A sample has been taken since the computer cannot analyse the full dataset. By sampling, this research aims to select some part of the full dataset so that one may estimate something about the full dataset. There are some important aspects to consider in collecting a data sample: (i) sample goal, (ii) population, (iii) selection criteria, (iv) sample size.

Chapter 6 elaborates we first aim to find interesting anomalies out of the whole dataset. Thereafter, an evaluation based on a ground-truth dataset is given in Section 6.4. Chapter 7 aims to build a reliable automatic diagnostic model for identifying failure causes. All these modeling

C_State	Description	Detailed Description
0	Off	Machine is turned off
2	Emergency stop	Operator pressed emergency button to stop Machine Y
4	Powering up	Machine state just after standby where Machine Y is powering up
6	Restore	State that tests if all electrical devices of the machine work properly
8	Standby	Machine is in standby state
10	Transition	The machine state between standby and the production state where the machine prepares for processing products
12	Production	Machine state where Machine Y can process products
14	Forced Cooling	Operator pressed forced cooling button where Machine Y actively cools down by maximizing the airspeed
16	Passive Cooling	Cooling state that takes more time to cool down Machine Y, however takes less energy
18	Emergency Cooling	Operator pressed emergency cooling button where Machine Y actively cools down by maximizing the airspeed
20	Cleaning	Machine state where Machine Y runs its cleaning program

Table 5.4: Status of Machine Y.

objectives need a specific sample each with its own sample goal, population, selection criteria, and sample size as input. Therefore, each modeling purpose explained in Chapter 6 and 7 includes a paragraph where the sample is extensively elaborated.

5.3 Conclusion

In this section, we answer research question 4: *How to preprocess the available raw data so it can be used for anomaly detection and diagnostics?* by preparing and reducing the dataset. Data preparation is necessary to raise the data quality to a level that suits modeling in Chapter 6 and 7. Therefore, we formatted the data, cleaned the data, and created new features. We conclude that investigating the influence of noise and outliers is key to reach for good-performing models. Hence, the data has been cleaned and filtered for both the fan and conveyor belt components. In the end, we found 12,468,719 clean records for Machine Type C fan; 32,870,932 records for the Machine Types A,B fan; 12,468,701 clean records for the Machine Type C belt drive motors and 32,869,564 clean records for the Machine Types A,B belt drive motors. These clean records include newly created features (i.e. rolling window calculations, operating times after being in standby mode, and relative powers of belt drive motors) to facilitate higher quality output results.

Data reduction comprises feature and instance selection. By utilizing expert knowledge wemanaged to select the most important parameters linked to our priority components. We applied an automatic feature selection technique (i.e. correlation analysis) and validated these distinguished parameters were correct. Moreover, we prioritized data records representing an operating machine state as the severity of a failure is the highest in this state. Additionally, different models are made for different Machine Y types as the motors are different. Hence, by preparing and reducing the dataset, we preprocessed the available raw data that can be directly used for anomaly detection and diagnostics.

Chapter 6

Anomaly detection

Anomaly detection is a set of techniques and systems to find unusual behaviors and/or states in systems and their observable signals (Jinka and Schwartz, 2016). An anomaly should first be detected before it can be isolated, labeled, and used for building a failure diagnostic model for Machine Y. As anomaly detection is used for learning to distinguish multiple types of failures (i.e. anomaly diagnosis), it is not necessary to create an *online* anomaly detection model that can update automatically with new data and perform novelty detection. This research aims to build anomaly detection and diagnostic models based on *offline* settings. In *offline* settings, the entire history of the data is available for analysis. Therefore, this chapter aims to answer Research Question 5:

How to detect a failed or under-performing priority component?

This chapter discusses the steps performed in anomaly detection and starts with Section 6.1 where we aim to elaborate on what anomalous behavior is. Section 6.2 extensively elaborates the models consulted for finding anomalies. Here we explain the type of data that is used for finding anomalies, the methodology of the selected models, and the visualization techniques used for utilizing expert knowledge to validate anomalies. Hereafter, Section 6.3 explains the distinguished interesting anomaly types of the priority components. Section 6.4 evaluates the performance of the anomaly detection models and this chapter is concluded in Section 6.5.

6.1 Understanding anomalies

An anomaly refers to a special kind of outlier that has application-specific importance. This application-specific importance is a state of a Machine Y priority component that is not performing optimal or in a failed state. Therefore, we come up with an application-specific definition of an anomaly based on Hawkins (1980) (p.1) definition of an anomaly:

"An anomaly is an observation that deviates so much from the other observations as to arouse suspicions that it was generated by a non-optimal operating or failed priority component of Machine Y."

According to Chandola et al. (2009), an important aspect of an anomaly detection technique is the nature of the desired anomaly. Anomalies can be classified into three categories (i.e. point anomalies, contextual anomalies, collective anomalies). The majority of anomaly detection research is focused on point anomalies (= global anomaly). This type of anomaly can be defined as an individual data point that behaves anomalously compared to the rest of the data. For example, in Figure 6.1, the data points P_3 and P_4 are considered a global outlier.

Contextual anomalies (= local anomaly) is a type of anomaly that is only anomalous in a specific context, but not otherwise (Chandola et al., 2009). Local outliers (points P_1 and P_2 in Figure 6.1) can be objects that are outlying relative to their local neighborhoods, particularly with respect to the densities of the neighborhoods.

Collective anomalies (Figure 6.2) is a type of anomaly where the collection of related data instances is anomalous with respect to the entire data set. The individual data instances in a collective anomaly may not be anomalies by themselves, but their occurrence together as a collection is anomalous. It should be noted that collective anomalies can occur only in data sets in which data instances are time-related.

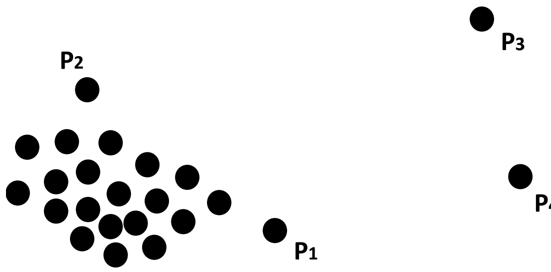


Figure 6.1: Illustration global and local outliers.

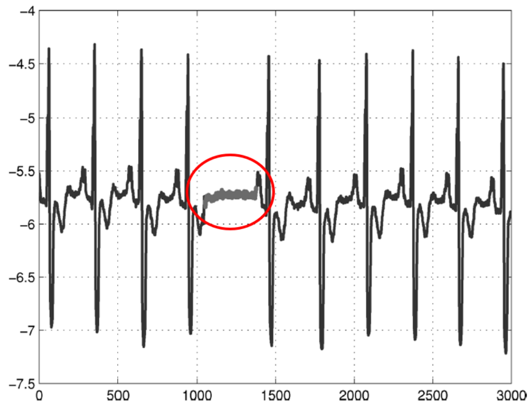


Figure 6.2: Illustration collective anomaly.

Inspecting previous examples of interesting anomalies are the supervision facet to enable building a robust anomaly detection model that can detect events of interest. Therefore, time-series data are inspected near documented failures for both priority components leading to the following statements: (i) Anomalous situations arise when the fan is turned off while the machine is in an operating state. (ii) As the maintenance data for the belt is richer than the fan data, we can see anomalous behavior near a documented failure. The first type of anomaly was found by being in an operating state while turning off the belt drive motors. Secondly, for some documented failures we saw fluctuating power behavior of all the motors linked to the belt drive.

As the maintenance data lacks quality and quantity we are confident in finding more types of anomalies by building unsupervised models to detect anomalies and utilize expert knowledge to validate if the potential anomalies out of the model are true anomalies (Aggarwal, 2015). Hence, this research uses a supervision process utilizing expert knowledge to define interesting application-specific anomalies. Section 6.3 elaborates on unsupervised methods that are used in an exploratory setting. We do have to take into account that many outliers found correspond to noise or other uninteresting phenomena. It has been observed that interesting anomalies are often highly specific to particular types of abnormal activity in the underlying application (Aggarwal, 2015).

6.2 How to find anomalies?

Since there are different types of anomalies, the core principle of discovering them is based on assumptions about the structure of the normal patterns in a given data set. The *given dataset* is explained in Section 6.2.1 while the *discovering anomalies* models are explained in Section 6.2.2. An overview of the process in finding interesting anomalies is given in Figure 6.3. This Figure shows that utilizing domain knowledge is important to distinguish anomalous situations from normal situation.

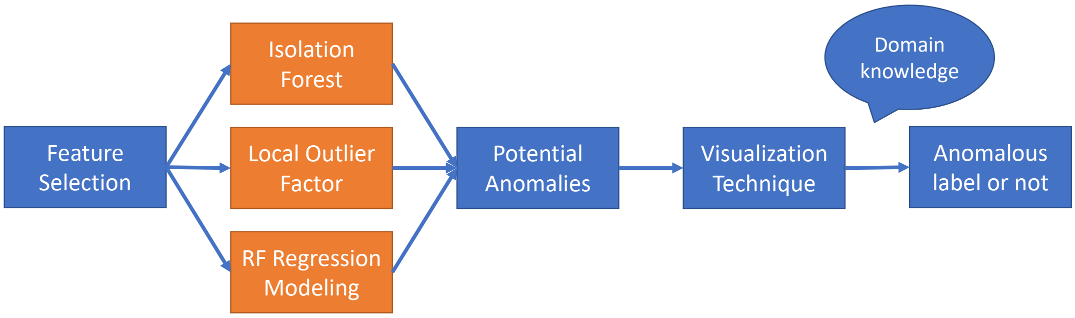


Figure 6.3: Validate potential anomalies process.

6.2.1 Data selection

The input data types used for modeling purposes influence the multiple types of anomalies (Section 6.1) that can be found. Two types of data dominate in anomaly detection models: multidimensional data where data records are time-independent of one another and time-series data. As we assume anomaly detection models take advantage of two quantitative properties of anomalies: (i) they are the minority consisting of few instances, and (ii) they have attribute-values that are very different from those of normal instances (Liu et al., 2012), anomalies can be defined as global or local outliers. Global and local outliers can be found with multidimensional data where the data records are time-independent. Feature engineering (Section 5.1.4), especially rolling window calculations, helps in finding more interesting anomalies as the model can take advantage of the temporal patterns. Furthermore, analysis based on multidimensional time-independent data is less complicated and more accessible than time-series analysis. As the interesting anomaly types can be found with multidimensional data where records are time-independent and it is more accessible than time-series analysis, we primarily focus on anomaly detection with time-independent data records as input.

6.2.2 Model selection

6.2.2.1 What is a normal machine state?

The choice of the 'normal' model depends highly on the understanding of the natural data patterns of Machine Y. According to Aggarwal (2015): "The more complex the data is, the more the analyst has to make prior inferences of what is considered normal for modeling purposes." (p.275). As concluded in Section 2.1, Machine Y has a high diversity in usage among multiple customers. It could be that a customer produces a product for one or two days with specific climate parameters and fan/ belt speeds that are rarely occurring compared to the majority of the available data. This situation agrees upon the two assumptions of being anomalous (Section 6.2.1). However, detecting such situations are not interesting to our customer since it does not represent a failed or underperforming priority component. Furthermore, parameters of priority components can sometimes show weird peaks and gaps just after an operator action which is not interesting to detect. In short, the two quantitative properties of anomalies are debatable and should be carefully dealt with when drawing conclusions about data points being interesting anomalies or not. Because of the wide variations and multiple facets that have to be taken into considerations to make the distinction between anomaly and normal, we cannot use off-the-shelf models and have to adapt the models so they can be applied to the problem-specific situation.

6.2.2.2 Bias-variance tradeoff

Whenever we discuss model prediction, it is important to understand prediction errors (i.e. bias and variance). Thoroughly understanding this phenomenon would help us not only to build accurate models but also to avoid the mistake of overfitting and underfitting.

The bias-variance tradeoff can be viewed as a sliding scale that controls how closely a learning algorithm adheres to its training data (Briscoe and Feldman, 2011). Models at one extreme of the scale (high variance) can entertain complicated hypotheses, allowing them to fit a wide variety of data (also outliers) very closely. However, they might generalize poorly as a result, a phenomenon known as overfitting. Contrarily, at the other end of the scale (high bias), models make relatively simple and inflexible assumptions, and as a result may fit the data poorly, called underfitting. Therefore, a simple model that is constructed with a good intuitive understanding of the data is likely to lead to much better results. Hence, the initial stage of selecting the data model is crucial in anomaly analysis.

6.2.2.3 Predictive learning

In common usage, prediction means to forecast a future event. In data science, prediction means to estimate an unknown value (Provost and Fawcett, 2013). This value could be something in the future, but it could also be something in the present or the past. This research defines *unknown value* as an anomaly score of a datapoint. Many forms of predictive learning use one of the following methods: instance-based learning methods, isolation-based learning methods, or explicit generalization methods.

In instance-based methods, a training model is not constructed upfront. Rather, for a given test instance, one computes the most relevant instances of the training data and makes predictions on the test instance using these related instances. The Local Outlier Factor (Section 6.2.4) is a popular and successful instance-based method for outlier detection. Furthermore, isolation-based learning methods isolate anomalies instead of profiles normal points and are further elaborated in Section 6.2.3. On the other hand, explicit generalization methods need a summarized model to be created upfront. In general, explicit generalization methods use a two-step process on the dataset: (i) Create a model of the normal data using the original data set. (ii) Score each point in the dataset based on its deviation from this model of normal data. A famous explicit generalization method is regression-based modeling which is explained in Section 6.2.5.

The well-known unsupervised outlier detection algorithms, Isolation Forest (Chapter 6.2.3) and Local Outlier Factor (Chapter 6.2.4) have been widely used in literature. However, Isolation Forest (IF) is only sensitive to global outliers and is weak in dealing with local outliers. Contrarily, the Local Outlier Factor (LOF) performs well in local outlier detection but has a high time complexity (Cheng et al., 2019). Therefore, both methods are researched and evaluated to see which method best fits our data in Section 6.4. As explicit generalization models are also frequently mentioned in anomaly detection literature, we aim to build anomaly detection models based on regression analysis (Section 6.2.5). By applying these different types of models that follow different methodologies we aim to find the anomaly situations in Machine Y.

6.2.3 Isolation forest

In recent years, the IF proposed by Liu et al. (2008) has attracted attention from the industry and academia due to its low time complexity and high accuracy. This methodology proposes to explicitly isolate anomalies instead of profiling normal points. IF take advantage of two anomalies' quantitative properties: (i) they are the minority consisting of fewer instances and (ii) they have attribute-values that are very different from those of normal instances. In other words, anomalies are 'few and different', which makes them more susceptible to isolation than normal points. The anomaly score of this method is based on this phenomenon. IF does not need to know the underlying distribution of the data and works well with multidimensional data where the records are time-independent of one another and consist out of continuous-valued attributes. The idea of identifying normal versus abnormal observations can be observed in Figure 6.4. A normal point (on the left) requires more partitions to be identified than an abnormal point (right). A formal definition of the IF can be found in Appendix D.1.

6.2.4 Local outlier factor

Breunig et al. (2000) came up with an unsupervised approach named 'Local Outlier Factor' in the instance-based proximity methods for searching the anomaly based on an anomaly score. The LOF evaluates data points according to a degree of measurement and does not require any explicit notion of clusters. The outlier factor is local in the sense that only a restricted neighborhood of each object is taken into account. The idea in proximity-based methods is to model anomalies as points that are isolated from the remaining data based on similarity or distance functions (Figure 6.5). A density-based analysis is one of the most common ways of defining proximity for anomaly detection and will therefore be further researched. In density-based anomaly detection, the number of other points within a specified local region (grid region or distance-based region) of a data point, is used to define local density. These local density values may be converted into anomaly scores. The LOF agrees upon the assumptions made at the end of Section 6.1 and can therefore be used in this research. This method does not need to know the underlying data distribution before it can function well and assumes that normal points occur in dense regions, while anomalies occur in sparse regions (Chandola et al., 2009). Attribute scaling (Section 5.1.5) is necessary since the LOF is based on the nearest neighbor approach. Normalization is recommended for most cases where similarity measures (like nearest neighbor) are being used. A formal explanation of LOF can be found in Appendix D.2.

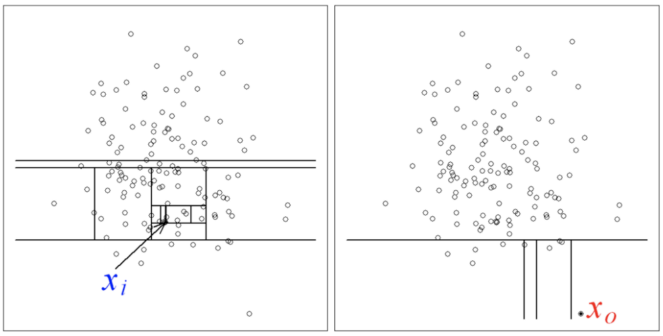


Figure 6.4: Isolation forest.

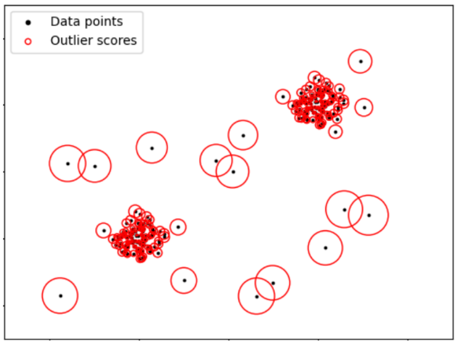


Figure 6.5: Local outlier factor.

6.2.5 Regression analysis

A regression procedure is a statistical method to model the relationship between a dependent (i.e. target) and one or more independent (i.e. predictor) variables (Provost and Fawcett, 2013). The notion of regression-based prediction and anomaly detection are intimately related. Anomalies are values that deviate from expected values based on a particular model (Paulheim and Meusel, 2015). Supervised methods can also be used for unsupervised anomaly detection by decomposing the anomaly detection problem into a set of regression modeling problems. The basic idea is to use regression modeling to predict an important attribute of the priority component from the remaining attributes that came out of the feature selection procedure (Section 5.2.1) and then compare the predicted attribute value to the actual attribute value to create a potential anomaly definition (Aggarwal, 2015). By the assumption that normal examples far outnumber anomalous examples, we can 'pretend' that the entire data set contains the normal class and create a model of the normal data. Based on the FMEA analysis (Section 3.1.1), we can assume anomalous points arise as the power of a priority component motor shows fluctuating behavior. We aim to predict the power based on independent variables related to a specific Machine Y motor.

Multiple *off-the-shelf* regression models can be used for effective supervised fan motor power predictions (i.e. nearest neighbor, linear regression, random forest, and neural networks). After the regression model predicted the values of the priority component's power, a visualization technique is necessary for distinguishing potential anomalies. Therefore, a scatterplot is made showing the actual power consumed on the x-axis and the predicted power consumed on the y-axis (Figure E). Higher deviations from the diagonal of the scatterplot that exceed some predefined threshold are distinguished as potential anomalous data points (red encircled in Figure 6.6). Regression models can deal with multidimensional data where the records can be independent of each other. Furthermore, we do not assume any particular distribution of the data (Paulheim and Meusel, 2015). However, the regression-based anomaly detector does not know how to deal with rare situations not seen before. Hence, the model needs to be trained with many different production states of the machine in a balanced way.

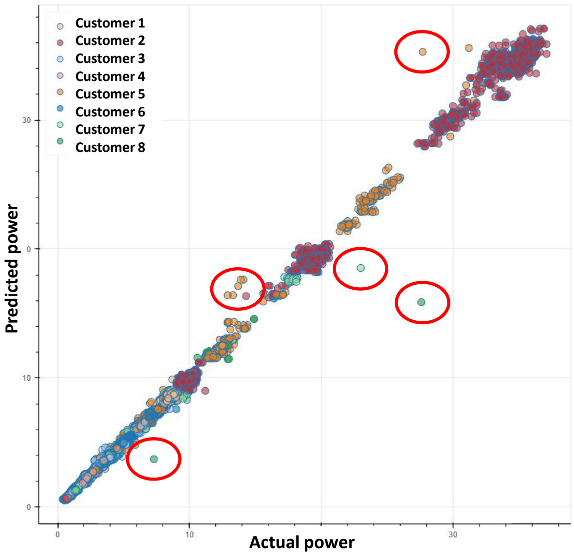


Figure 6.6: Regression-based anomaly detection.

6.2.5.1 Random forest

This research uses random forest (RF) regression models to predict the priority component motor power since these models are currently among the most widely used machine learning methods (Müller and Guido, 2016). They are robust, powerful, often work well without heavy tuning of the parameters, and do not require scaling of the data (Fernández-Delgado et al., 2014). Random forests are robust because of their implicit use of multiple locally relevant subspaces (Aggarwal, 2013). A RF is essentially a collection of decision trees, where each tree is slightly different from the others. The idea behind RF is that each tree might do a relatively good job of predicting, but will likely overfit on part of the data. If we build many trees, all of which work well and overfit in different ways, we can reduce the amount of overfitting by averaging their results.

6.2.5.2 Cross-validation

To further reduce the probability of overfitting, we train the individual predictive models in different folds, using cross-validation. Cross-validation is a statistical method of evaluating the generalization performance where the data is split repeatedly and multiple models are trained. According to Müller and Guido (2016), the most commonly used version of cross-validation is k-fold cross-validation, where the data is first partitioned into k parts of equal size. Hereafter, a sequence of models is trained visualized in Figure 6.7. As a simple random sample is used for regression modeling based on multidimensional data where the records are time-independent, k-fold cross-validation can be reliably used. However, for classification purposes (Chapter 7) stratified k-fold cross-validation is used which is extensively elaborated in Section 7.3.1.2.

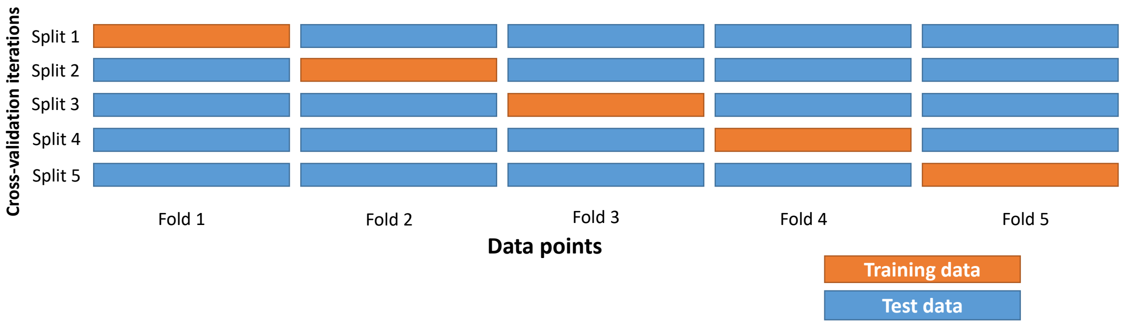


Figure 6.7: Data splitting in k-fold cross-validation.

6.2.6 Utilizing expert knowledge to validate anomalies

As the anomaly detection models have been formulated to find potential anomalies validating them is an important next step. Frequently, anomaly detection algorithms not only find anomalies but also outliers that are not interesting to the application-specific situation. An ideal situation to effectively and efficiently validate potential anomalies would be analysing high-quality maintenance data. However, as Section 6.1 elaborates that inspecting data points near a documented failure is not sufficient for completely understanding what kind of anomalies exist for the priority components, we had to find an additional way of reliably validating the potential anomalies. By utilizing expert knowledge with the help of visualization tools to show anomaly behavior we aim to validate the potential anomalies. Three visualization tools have been consulted to provide the expert with knowledge about the specific situation which has been detected by the unsupervised anomaly detection algorithms (Sections 6.2.3, 6.2.4, and 6.2.5). The expert has to formulate a classification if a specific datapoint or a sequence of datapoints are anomalous or not based on the visualization tools that are explained in Sections 6.2.6.1, 6.2.6.2, and 6.2.6.3.

6.2.6.1 Behavior parameters relative to each other

The multidimensional data record that is given an anomaly score by the anomaly detection model can be visualized where all the attributes are shown relative to each other. This visualization techniques only takes into account a specific moment in time at a customer that the anomaly detection models give as an output. To provide the expert with an extra information source the mean behavior of the parameter in an operating state of the machine among all customers has been added.

6.2.6.2 Time-series plot

Time series visualizations that show the sensor behavior in time provide the expert with knowledge about the machine behavior before and after a potential anomalous data point. Furthermore, the visualization technique shows the relation of the interesting parameters to each other of a specific timestamp at a customer.

6.2.6.3 Density plot

The visualization techniques in Section 6.2.6.1 and 6.2.6.2 provide the expert with knowledge about a specific situation at a customer where we can take the mutual behavior of parameters into account over time. Conversely, the visualization technique explained in this section provides the expert with knowledge about parameters behavior among multiple customers. Utilizing domain knowledge made clear a fan should absorb a specific motor power at a specific setting (i.e. situation with a fixed temperature, airspeed, and dew point) as the fan power behavior is independent of the belt load. Creating a density plot enables us to compare power absorption at multiple customers. If a customer has a significantly higher or lower power consumption than others in the same setting, the experts have an extra source of information to base their anomaly label on.

6.3 Finding Machine Y priority component anomalies

We aim to use anomaly detection models (Section 6.2.2) to explore potential anomalies. The potential anomalies that the model provides as output are provided to Machine Y expert for further examination of their application-specific importance (Aggarwal, 2015). As it is expensive and time-consuming to manually label potential anomaly data points, the expert is asked to assess only pre-filtered candidate points for labeling. Therefore, it is important to present carefully chosen examples to the expert so that the decision boundary between the rare and normal classes is learned with as few examples as possible. Examples that are predicted to be clearly positive or negative are not the only interesting set of points that should be assessed by experts. Also, points with a greater uncertainty or ambiguity in anomaly probabilities should be presented to the expert for labeling. The models that have been built to find potential anomalies and the results after utilizing domain knowledge are discussed in Section 6.3.1 and 6.3.2. A conclusion of finding interesting anomalies is provided in Section 6.3.3.

6.3.1 Main air distribution fans

6.3.1.1 Input data

Chapter 5 describes how to get a clean dataframe related to the fan. Table 5.2 shows the parameters describing the fan behavior that has been found by utilizing expert knowledge and a correlation analysis (Section 5.2.1). As the two fans in Machine Y operate independently we have to model them separately for both Machine type C and Machine type A,B. Following Figure 5.3, we have 12,468,719 Machine type C fan base and secondary tower data records. Furthermore, we have 32,870,932 Machine type A,B fan base and secondary tower data records. However, some of these records contain empty values because of procedural error elaborated in Section 5.1.3. We are not imputing these missing values since we have sufficient condition data available consisting of good measurements. We keep the missing values in the overall dataframe as missing values. However, ML models cannot deal with missing values. Hence, when creating a reduced sample of the total dataframe we do not take into account these empty values. The sample goal in finding interesting anomalies out of the whole dataframe is to reduce the dataset to a manageable size for the computer for making calculations that would be as good as making calculations for the whole dataset.

The population (i.e. whole dataset) for finding interesting anomalies for the fan starts with the filtered dataframe of Figure 5.3 as can be seen in Figure 6.8. However, since the elaborated anomaly detection models mainly found uninteresting outliers (i.e. parameter value transitions after arriving in a new machine state or transitions after an operator change in setpoint), we further filtered the dataframe to only the most interesting part. This denotes the population is further reduced to only the part that does not include transitions because of machine state changes or operator changes. Specifically, data records within a time span of one minute after a machine state transition or an operator transition are not included in the population. Furthermore, some anomaly detection models make use of the rolling window feature (Table C.2) which means these are added to the input data. Now, we have to take a sample from the remaining interesting population to facilitate modeling purposes as the whole population is too huge for analysis (i.e. it may not be possible to run a data mining algorithm, or the data mining task might not be able to be effectively performed). Additionally, including the whole population for analysis has no added value as data records are very similar to each other as they are monitored every five second. Furthermore, the rolling window feature includes information of previous time measurements which also favor analysing a sample over the whole population.

The three elaborated anomaly detection models all have their characteristics. For the regression-based analysis, we chose a sample size of 30,000 (10,000 for training purposes and 20,000 for test purposes). In literature, a rule-of-thumb is to use 75% of data for training and 25% for test purposes (Müller and Guido, 2016). However, we assume 10,000 data points should be sufficient for the RF to learn what a normal machine represents. The larger test set is used to find interesting anomalies as only a very small percentage of the total dataset is assumed to be anomalous. For the IF and LOF, we created a sample size of 20,000. We assumed these sample sizes are large enough for drawing reliable conclusions for the whole population. All these samples are randomly selected where all examples have an equal chance to be included. Concluding, we have four data samples related to the fan that are used as input for the modeling techniques elaborated in Section 6.3.1.2.

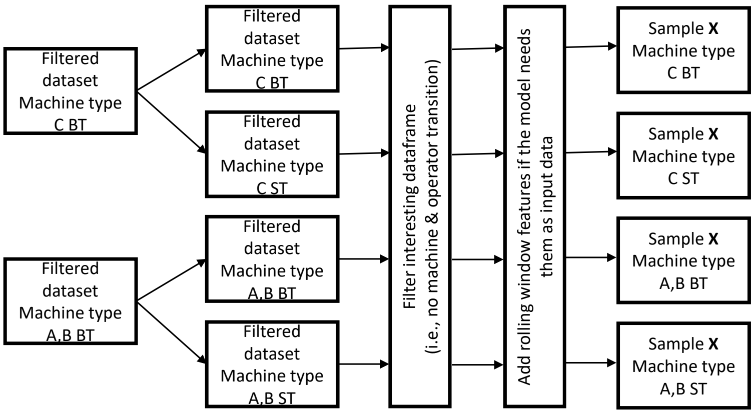


Figure 6.8: Input data fan.

6.3.1.2 Anomaly detection modeling techniques

Table 6.1 summarizes all anomaly detection models that have been built to find interesting anomalies related to the fan. The RF models 1, 2, and 3 are supervised as they aim to predict the power consumption of the motor. However, the general anomaly detection model, that uses the results of RF regression, is unsupervised as there are no labels available (i.e. normal or anomalous). In total, each model shown in Table 6.1 has been built for the distinguished four samples of Figure 6.8. The *Model name* labels each model with its specific input parameters. *Input parameters* show which parameters have been used from the input data sample for finding potential anomaly data records. *Input sample* elaborates which sample size has been used. Column name *Output values* show high accuracies for the regression-based models after performing cross-validation (Section 6.2.5.2). This denotes the power can be predicted accurately based on the provided input parameters. The anomaly score is based on the difference between the predicted and actual power consumption. As can be seen in Appendix E, RF1 hardly contains any data point from the diagonal. RF2 and RF3 have fewer input parameters to predict the power consumption resulting in more data points located from the diagonal. Especially, the RF regression model for Machine type C secondary tower contains significantly more data points from the diagonal. These points from the diagonal have been visualized in time-series plots but the experts were not able to distinguish anomalous behavior. This implies, the data points further away from the diagonal can be a consequence of insufficient training data to learn Customer 4's varying machine behavior correctly. The other anomaly detection models (i.e. LOF and IF) provide anomaly scores that can be ranked and the most interesting anomalies can be shown to experts to validate if the data record can be categorized as anomalous. Some remarks have been written down with the main findings of that particular model in column *Remarks*. Lastly, *Results* refer to the Appendix where the regression-based visualizations can be found of potential anomalies.

Model name	Input parameters	Input sample	Output values	Remarks	Results
RF 1	current, motor speed, airspeed, temperature, dew value	sample Machine type C & A,B (Figure 6.8) train: 10,000 records test: 20,000 records	* Accuracy scores: Machine type . C BT: 0.9993 A,B BT: 0.9989 C ST: 0.9995 A,B ST: 0.9997	* Very high accuracy resulting in hardly any distinguished potential anomalies. * Mainly detects fluctuations in climate parameters	Appendix E
RF 2	motor speed, airspeed, temperature, dew value	sample Machine type C & A,B (Figure 6.8) train: 10,000 records test: 20,000 records	* Accuracy scores: Machine type . C BT: 0.9969 A,B BT: 0.9978 C ST: 0.9904 A,B ST: 0.9987	* Very high accuracy resulting in hardly any distinguished potential anomalies. * Mainly detects fluctuations in climate parameters	Appendix E
RF 3	airspeed, temperature, dew value	sample Machine type C & A,B (Figure 6.8) train: 10,000 records test: 20,000 records	* Accuracy scores: Machine type . C BT: 0.9979 A,B BT: 0.9961 C ST: 0.9877 A,B ST: 0.9987	* Very high accuracy resulting in hardly any distinguished potential anomalies. * Mainly detects fluctuations in climate parameters	Appendix E
IF + LOF 1	power, current, motor speed, airspeed, temperature, dew value	sample Machine type C & A,B (Figure 6.8) total dataset: 20,000 records	Anomaly scores	* Mainly finds uninteresting climate fluctuations, therefore we build models only taking into account motor-related parameters from this moment	
IF + LOF 2	power, current, motor speed	sample Machine type C & A,B (Figure 6.8) total dataset: 20,000 records	Anomaly score	* As we look at single time stamps it is difficult to take into account temporal trends. Therefore, the next elaborated models use window calculations	
IF + LOF 3	power, current, motor speed, rolling window std 5	sample Machine type C & A,B (Figure 6.8) total dataset: 20,000 records	Anomaly score	* Detect small peaks (no big peaks or gaps) * Therefore, rolling window of 30 minutes has been added the next model	
IF + LOF 4	power, current, motor speed, rolling window std 30	sample Machine type C & A,B (Figure 6.8) total dataset: 20,000 records	Anomaly score	* No big peaks or gaps can be distinguished	
IF + LOF 5	power, current, motor speed, rolling window min-max 5	sample Machine type C & A,B (Figure 6.8) total dataset: 20,000 records	Anomaly score	* Detect small peaks (no big peaks or gaps) * Therefore, rolling window of 30 minutes has been added the next model	
IF + LOF 6	power, current, motor speed, rolling window min-max 30	sample Machine type C & A,B (Figure 6.8) total dataset: 20,000 records	Anomaly score	* No big peaks or gaps can be distinguished	

Table 6.1: Unsupervised modeling main air distribution fan.

6.3.1.3 Anomaly types distinguished

The results of the anomaly detection modeling techniques (Section 6.3.1.2) in combination with experts' opinions if the potential anomalies are true anomalies instead of uninteresting outliers are discussed in this section. Two types of anomalies can be distinguished for the main air distribution fan. The first type is a situation where the machine is in an operating state and the fan stops blowing air through the machine. This directly results in fluctuating climate parameters that are not conducive for the products that are being baked or cooked. Moreover, a motor power fluctuation analysis has been performed to look for anomalous behavior in the power consumption.

After finding weird power consumption behavior, we concluded only one type was not caused by normal behavior of the fan and is therefore an interesting anomaly detection (Table 6.2). The parameter behaviors listed in Table 6.3 are not interesting since they represent fluctuating fan motor powers that can be explained according to setpoint changes of one of the climate parameters and do not represent a non-optimal performing or failed fan motor.

	Parameter behavior	Diagnosis expert	Visualization
1	Fan turned off while being in an operating state	Fan motor frequency converter failure	Figure E.13
2	Fluctuations in motor power and current consumption without a setpoint change in either airspeed, dewpoint, or temperature	Undefined	Figure E.14

Table 6.2: Interesting anomalies fan.

	Parameter behavior	Diagnosis expert	Visualization
1	Fluctuations in motor power and current consumption	low temperature or dewpoint and high airspeed result in fluctuating power fan	Figure E.15
2	Fluctuations in motor power and current consumption	Changes in temperature, dew value and airspeed cause fluctuating power	Figure E.16
3	Fluctuations in motor power and current consumption	Broken dewpoint sensor causing fluctuations power	Figure E.17

Table 6.3: Uninteresting anomalies fan.

6.3.2 Conveyor belt drives

6.3.2.1 Input data

Chapter 5 describes how to get a clean dataframe related to the belt drive motors (i.e. master, slave, base drum, secondary drum). Table 5.3 shows the parameters describing the belt drive motors that have been found by utilizing expert knowledge and a correlation analysis (Section 5.2.1). As the four belt drive motors in Machine Y operate interdependently (Section 5.1.4.3), we can both separately and collectively model them for Machine type C and Machine type A,B (*separately* and *collectively* in Figure 6.9). Separately modeling means to only model based on input parameters directly connected to one specific motor. Conversely, collectively modeling means building the model based on input parameters connected to all belt drive motors. Following Figure 5.4, we have 12,468,701 Machine type C belt drive motors data records. Furthermore, we have 32,869,564 Machine type A,B belt drive motors data records. However, some of these records contain empty values because of procedural error elaborated in Section 5.1.3. The same reasoning as for the fan (Section 6.3.1.1) can be applied to the belt drives why we are not imputing the missing values, create a sample without these empty values, and what the sample goal is to find interesting anomalies.

The population for finding interesting anomalies for the belt drive motors starts with the filtered dataframe of Figure 5.4 as can be seen in Figure 6.9. This population has to be further reduced to get only the interesting part of the whole dataset. We found out the belt can be turned off in the operating state of the machine. Hence, we further reduced the population to only situations where the machine is in an operating state and the belt is running. Hereafter, we deleted the first ten minutes after a machine has been started up because of uninteresting power fluctuations due to initial lubrication processes that do not represent a failed belt drive motor. Now, we ended with a remaining population that should contain stable parameter behavior in a normal Machine Y situation and can therefore be used for anomaly detection. Furthermore, some anomaly detection models make use of the rolling window feature (Section 5.1.4.1), operating time after being in standby mode (Section 5.1.4.2), and the relative belt drive motor powers (Section 5.1.4.3) which mean these are added to the input data. Now, we have to take a sample from the remaining interesting population to facilitate modeling purposes. For the three elaborated anomaly detection models we took the same sample sizes as for the fan (Section 6.3.1.1). All these samples are randomly selected where all examples have an equal chance to be included. Concluding, we have ten samples related to the belt drive motors that are used as input for the modeling techniques elaborated in Section 6.3.2.2.

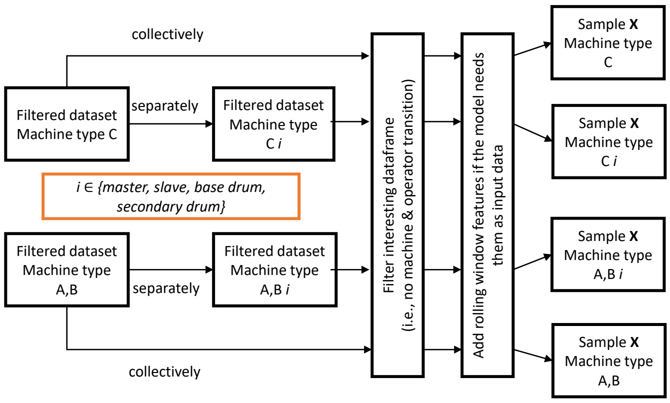


Figure 6.9: Input data belt drive motors.

6.3.2.2 Anomaly detection modeling techniques

Table 6.4 summarizes all anomaly detection models that have been built to find interesting anomalies related to the belt conveyor drive motors. Column *separate/ collective?* shows if each belt drive motor has been modeled separately or collectively. Moreover, the number of built models is mentioned between brackets. When separately modeled we took the sample paths in Figure 6.9 labeled with *seperately*. Conversely, when collectively modeled we took the sample paths labeled with *collectively*. As can be seen in Table 6.4, the last two model names were only built two times since all data of all belt motor drives were necessary as input values and only a distinction can be made between Machine type C and Machine type A,B.

The accuracy scores of *RF 1* and *RF 3* in Table 6.4 are high. This denotes most data points have been accurately predicted compared to the actual value. *RF 3* shows more data points located from the diagonal than in *RF 1* and are inspected by experts if these points are anomalous or not. Conversely, *RF 2* shows lower accuracy scores as it does not use specific motor input parameters (i.e. current and motor speed) but only general input information (i.e. overall beltspeed and overdrive). This denotes the input parameters connected to *RF 2* are insufficient for reliably predicting the power consumption of the belt drive motors. Nevertheless, the data points that are located from the diagonal in Appendix E were inspected and assessed by experts resulting in no interesting anomalous situations.

The remark *Deal with procedural errors* is mentioned frequently in Table 6.4. This remark is a result of not all customers logging all their parameters in the right way as explained in Section 4. ML models cannot deal with missing values so when for example the current is necessary as input value for modeling, only the customer data are included from which the current is monitored in the right way.

The advantage of the LOF and the IF over the regression-based anomaly detection algorithms is that they can take into account temporal data. When the IF and LOF include the rolling window calculations (Section 5.1.4.1), better potential anomalies can be found. Rolling window calculations of 5 minutes resulted in small peaks that can be distinguished in power consumption while calculations of 30 minutes resulted in wider gaps and peaks that can be distinguished.

6.3.2.3 Anomaly types distinguished

The results of the anomaly detection modeling techniques in combination with experts' opinions if the potential anomalies are true anomalies instead of uninteresting outliers are discussed in this section for the belt drive motors. Two main types of anomalies can be distinguished for the belt drive motors. The first type is a situation where the machine is in an operating state and the belt is not running. The second main type of anomalies can be distinguished by looking at the interesting unstable behavior of the power consumption of the belt drive motors.

The first type of anomaly needs further research to distinguish interesting from uninteresting anomalies. An event analysis has been performed to find out why the belt drive motors stop running while being in an operating machine state. Hence, first all data points have been filtered on the rule that Machine Y is in an operating state and the belt is not running. Hereafter, we have been looking at what events were connected to the time range in which the rule holds. Table F.1 contains interesting events that can be used to identify if a datapoint is truly anomalous or if a datapoint is an uninteresting outlier. Event one until eighteen in Table F.1 are interesting anomalous events as they represent a failed or malfunctioning Machine Y while events nineteen and twenty are uninteresting events since they represent an operator pressing the button to stop the belt from running. After conducting the event analysis, we concluded the type of anomaly where the machine is in an operating state and the belt does not run is not interesting for evaluating purposes as they follow a simple rule that is already known by Company X.

The second main type of anomalies can be distinguished by looking at the interesting unstable behavior of the power consumption of the belt drive motors. As can be seen in Table 6.5, we found fourteen interesting belt drive anomaly types. Most of the results shown in Table 6.5 are not detected by present-day machine rules. On the other hand, we also found fluctuating power behavior that can be explained by healthy machine behavior which is shown in Table 6.6.

6.3.3 Conclusion

Finding interesting anomalies for the priority components is an iterative and incremental process. The right visualization techniques are crucial to provide experts information so they can classify anomalous from normal situations. The visualization technique of parameter behavior relative to each other (Section 6.2.6) appeared to contain not enough information for categorizing data points as anomalous or not. Conversely, time-series visualization techniques were better information sources for the expert. The main advantage of IF and LOF over the regression-based anomaly detection model is that these techniques can take into account temporal data information in the form of rolling window calculations (Section 5.1.4.1).

In the end, we were able to distinguish two types of interesting anomalies for the fan (Table 6.2). The first type of anomaly is a fan turned off while being in a machine operating state. This anomaly can be found with a rule $(fan(power\ OR\ current\ OR\ speed) == 0 \mid C\_State = 12)$. The second type of anomaly shows weird behavior of the power and current consumption. However, utilizing expert knowledge resulted in an undefined diagnostic as it seems hard to label a specific cause to this fan behavior. Experts are not sure if fluctuations in climate parameters (i.e. temperature, dewpoint, airspeed) cause fluctuations in power and current consumption of the fan or if it is the other way around. Anomaly type 2 in Table 6.2 would be an interesting anomaly if and only if the fluctuations in the fan motor parameters cause the fluctuations in climate parameters. All in all, as the first type of anomaly can be found with a simple rule and the second type of anomaly does not have a diagnosis, it is not interesting to further evaluate how well the anomaly detection models (Table 6.1) score related to the main air distribution fan.

Conversely, the belt drive motors show fourteen types of interesting anomalies all connected to anomalous power consumption behavior (Table 6.5). This denotes the belt drive motors can be further evaluated how the anomaly detection models (Table 6.4) perform in finding these distinguished anomaly types based on a ground truth dataset (Section 6.4.1).

Model name	Separate/ Collective?	Input parameters	Input sample	Output values	Remarks	Results
RF 1	Separately (8 models)	current, motor speed, beltspeed, overdrive, time in operating state	sample Machine type C & A,B (Figure 6.9) train: 10,000 records test: 20,000 records	* Accuracy scores: Machine type ...... C..........A,B... 1: 0.9999 1: 0.9999 2: 0.9997 2: 0.9989 3: 1.0000 3: 0.9999 4: 1.0000 4: 0.9999	* Very high accuracy resulting in hardly any distinguished potential anomalies. * Deal with procedural errors	Appendix E
RF 2	Separately (8 models)	beltspeed, overdrive, time in operating state	sample Machine type C & A,B (Figure 6.9) train: 10,000 records test: 20,000 records	* Accuracy scores: Machine type ....C..........A,B... 1: 0.9801 1: 0.9263 2: 0.9771 2: 0.9197 3: 0.9736 3: 0.9601 4: 0.9975 4: 0.9573	* Machine type A,B has a lower accuracy than Machine type C.	Appendix E
RF 3	Collectively (8 models)	power other belt drive motors, beltspeed, overdrive, time in operating state	sample Machine type C & A,B (Figure 6.9) train: 10,000 records test: 20,000 records	* Accuracy scores: Machine type ....C..........A,B... 1: 0.9998 1: 0.9936 2: 0.9994 2: 0.9940 3: 0.9996 3: 0.9970 4: 0.9997 4: 0.9988	* Very high accuracy resulting in hardly any distinguished potential anomalies.	Appendix E
IF + LOF 1	Separately (8 models)	power, current, motor speed, beltspeed, overdrive, relative power among connected belt drive motors	sample Machine type C & A,B (Figure 6.9) total dataset: 20,000 records	Anomaly scores	* Mainly finds data points where Machine Y is in operating state, but belt does not run. * Deal with procedural error	
IF + LOF 2	Separately (8 models)	power, current, motor speed, power rol. win. 5 min (stdev, min-max), relative power among connected belt drive motors	sample Machine type C & A,B (Figure 6.9) total dataset: 20,000 records	Anomaly score	* Detects small peaks in power because window is 5 min * Deal with procedural error	
IF + LOF 3	Separately (8 models)	power, current, motor speed, power rol. win. 30 min (stdev, min-max), relative power among connected belt drive motors	sample Machine type C & A,B (Figure 6.9) total dataset: 20,000 records	Anomaly score	* Mainly detects wider peaks and gaps in power because window is 30 minutes * Deal with procedural error	
IF + LOF 4	Collectively (2 models)	all motors: beltspeed, overdrive, power, power rol. win. 5 min. (stdev, min-max), relative power among connected belt drive motors	sample Machine type C & A,B (Figure 6.9) total dataset: 20,000 records	Anomaly score	* Detects small peaks in power because window is 5 minutes * As some customers had missing values in their current and speed loggings, we only looked at the power here of the motor parameters	
IF + LOF 5	Collectively (2 models)	all motors: beltspeed, overdrive, power, power rol. win. 30 min. (stdev, min-max), relative power among connected belt drive motors	sample Machine type C & A,B (Figure 6.9) total dataset: 20,000 records	Anomaly score	* Detects wider peaks in power because window is 30 min * As some customers had missing values in their current and speed loggings, we only looked at the power here of the motor parameter	

Table 6.4: Unsupervised modeling belt drive motors

	Parameter behavior	Diagnosis expert	Visualization
1	Sudden power increase master motor, a few moments later the secondary drum motor power also increased	Conveyor belt got stuck	Figure E.42
2	Sudden decrease in slave pickup power consumption, while simultaneously the base drum motor increased its power consumption	Belt tension mechanism got stuck	Figure E.43
3	Recurring base drum peaks and gaps after a specific time horizon	Base drum got stuck	Figure E.46
4	Fluctuating behavior base drum motor and slave motor while there is not transition in setpoints	Belt tension mechanism got stuck	Figure E.45
5	Fluctuating behavior of all belt drive motor power consumptions where the period of one sine is equal to the dwell time	Conveyor belt got stuck	Figure E.48
6	As the base drum power consumption gets higher, the secondary drum power consumption gets lower	Drum pulls other drum which mean the overdrive has been set wrongly	Figure E.49
7	Fluctuating behavior of all belt drive motor power consumptions starting with a base and secondary drum power consumption close to each other and as the time proceeds the difference of power gets wider	Drum pulls other drum which mean the overdrive has been set wrongly	Figure E.50
8	Sudden recurring decreases in power consumption of all belt drive motors except the slave motor	Undefined	Figure E.51
9	Spikes in power consumption base drum motor and slave motor	Belt tension mechanism got stuck	Figure E.53
10	Sudden change in sine period power consumption drum motors	Undefined	Figure E.54
11	During a day structurally recurring gaps in power consumption of all belt drive motors except the slave motor	Undefined	Figure E.56
12	One decreasing spike of all belt drive motor powers except the slave motor power	Belt tension mechanism got stuck	Figure E.57
13	Highly fluctuating slave power consumption	Belt in return section got stuck	Figure E.59
14	Weird peaks in master power consumption where the other belt drive motors do not show weird behavior	Undefined	Figure E.60

Table 6.5: Interesting anomalies conveyor belt drive motors.

	Parameter behavior	Diagnosis expert	Visualization
1	Sudden increase in master pickup power consumption, while simultaneously the secondary drum motor increased its power consumption	Product weight transition due to control valve steam zone increase and no transition in setpoints	Figure E.44
2	Sudden increase of all belt drive motors	Product weight transition due to control valve steam zone increase and no transition in setpoints	Figure E.47
3	Sudden decrease and increase in power consumption all belt drive motors	Product weight transition due to control valve steam zone increase and no transition in setpoints	Figure E.55
4	Sudden decrease of all belt drive motors powers except the slave motor	Conveyor belt got lubricated	Figure E.58

Table 6.6: Uninteresting anomalies conveyor belt drive motors.

6.4 Evaluation and model improvement

This section aims to show the effectiveness of the belt drive anomaly detection algorithms listed in Table 6.3.2.2 based on scoring metrics suitable to our available data. To compare the formulated anomaly detection algorithms to each other, we score the models based on a uniform ground-truth dataset (Section 6.4.1). A grid search is performed to gain maximum performance and a risk assessment of our evaluation method is given in Section 6.4.4.

6.4.1 Ground-truth dataset

In the unsupervised scenario where we did not know if data records are anomalous or not, difficulty arises in judging the effectiveness of the underlying algorithms transparently. Therefore, we use the validated distinguished anomaly types in and bring them together in a ground-truth dataset. A ground-truth dataset is a reference set where we know which records are anomalous and which records are normal that can be used for evaluating purposes. Two different ground-truth datasets have been built for assessing the validity of scoring. One ground-truth dataset is built with interesting anomalies (Tabel 6.5) and normal data points (i.e. low-scored IF scores). Another ground-truth dataset is built with interesting anomalies, uninteresting anomalies (Table 6.6), and low-scored IF scores (i.e. normal Machine Y situation). The reason why we built this second ground-truth dataset is to validate how well the anomaly detection model can categorize the uninteresting anomalies as a normal situation. Both ground-truth datasets contain 540 anomaly records and 4,860 normal records (i.e. 10% anomalies, 90 % normal data records). Referring to Section 5.2.2.2, our sampling goal for evaluating is to create a supervised anomaly detection situation. This can be reached by creating a population of records that have been assessed by experts as either anomalous or normal. In the end, we got 5,400 records in our ground-truth dataset which is assumed to be sufficient for making valid evaluation statements. Hence, we transformed a situation with no labeled data to an expert validated ground truth labeled dataset that can be used for scoring the defined models in Table 6.4.

Since anomalies are far less common than normal points, we arrive in a situation where we have to deal with an imbalanced dataset. An imbalanced dataset is a dataset in which one class is much more frequently occurring than the other in a binary classification setting. A characteristic of imbalanced data is that accuracy is not a reliable scoring method since machine learning models can predict all test points as normal points and still achieve excellent accuracy (Müller and Guido, 2016). Therefore, the classifier is tuned, so that errors in classification of the anomalous class are penalized more heavily than the errors in classification of the majority class (Aggarwal, 2015).

6.4.2 Scoring models

6.4.2.1 Error types

If the model does not detect an anomaly, Machine Y is assumed to be healthy, while if the model detects an anomaly, the FSE gets an alarm to solve the issue. We cannot guarantee an anomaly detection model that always works perfectly. Therefore, selecting the right measure of predictive performance starts by assessing the consequences of the model output. One possible mistake is that a healthy priority component of Machine Y will be classified as anomalous (i.e. false positive), leading to the FSE getting a notification. This results in inconvenience of the FSE as there seems to be no issue and ultimately leads to the anomaly detector that will not be used anymore because of a confidentiality issue. Especially, in the beginning stage of implementing an anomaly detector, it is important to gain the confidence of the FSEs by providing reliable maintenance decision support. Furthermore, a possible mistake can be a failed or non-optimally performing Machine Y priority component to be predicted as not anomalous (i.e. false negative). In this case, it could be that food safety cannot be guaranteed anymore, or much yield is lost, or company's revenue can decrease due to unscheduled downtime. Hence, false negatives lead to customers having less confidence in the anomaly detector. Now the consequences of the model output have been assessed, we face the dilemma between which error type is more harmful to Company X.

Together with our stakeholders, we determined reducing the false positives to a minimum and gaining confidence of the operators and customers is key at the beginning stage of an anomaly detection implementation. Not prioritizing this error type would result in customers not using the tool anymore leading to data loss for further improving the detection and diagnosis models. Furthermore, preventing false negatives from happening is also key as food safety cannot be guaranteed anymore. Conversely, preventing false positives from happening is key to keep the confidence in the anomaly detection model high from the FSEs. However, if this error type occurs we can provide the FSE with information that the ML model has to learn from this mistake probably resulting in better predictions in the future. This implies the detection model improves as time (and amount of machine failures) proceeds. This insight is taken into consideration when selecting the scoring methods in Section 6.4.2.2.

6.4.2.2 Confusion matrices

One of the most comprehensive ways to represent the result of evaluating binary classification is using confusion matrices (Müller and Guido, 2016). The output of a confusion matrix (Figure 6.10) is a two-by-two array, where the rows correspond to the true classes and the columns correspond to the predicted classes. Entries on the main diagonal of the confusion matrix correspond to correct classifications, while other entries tell us how many samples of one class got mistakenly classified as another class.

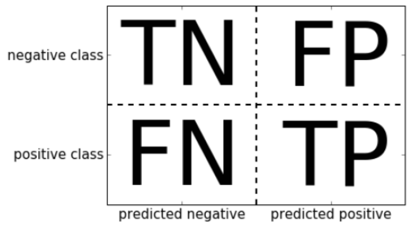


Figure 6.10: Confusion matrix example.

In the machine learning community, precision and recall are arguably the most commonly used measures for binary classification and summarizing the confusion matrix (Müller and Guido, 2016). Precision (Equation: 6.1) measures how many of the samples predicted as positive are positive. Recall (Equation 6.2) measures how many of the positive samples are captured by the positive predictions. However, while precision and recall are very important measures, looking at only one of them will not provide the full picture. While performing grid search (Section 6.4.3), the most suitable hyper-parameter combination can be found by ranking a combination score of the precision and recall. According to Sasaki et al. (2007), one way to summarize the precision and recall is the F_β*-score* (Equation: 6.3). β is chosen such that recall is considered β times as important as precision. The F_β*-score* has been widely used in machine learning where more emphasis is paid on either the precision or recall (Li et al., 2008).

Now the scoring methods are defined and the dilemma between false positives and negatives is discussed in Section 6.4.2.1, we decided together with our company stakeholders to reach a precision of 80% and reach 95 % recall before continuing to the diagnostic stage (Chapter 7). The recall score is higher than the precision score as we prioritize the score that measures how many of the positive samples are captured by the positive predictions. This denotes we can formulate β to be $\frac{95}{80} = 1,1875$.

$$Precision = \frac{TP}{TP + FP} \tag{6.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{6.2}$$

$$F_\beta\text{-}score = (1+\beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall} \tag{6.3}$$

6.4.3 Grid search

Now we defined how to score our models in Table 6.3.2.2, we aim to improve the model's performance by tuning its parameters using grid search. Grid search means trying multiple combinations of the parameters of interest (i.e. threshold what scores are anomalous and model hyper-parameters) to reach the highest model performance. The amount of contamination (i.e. the proportion of anomalies in the data set) of the data set has a direct influence on the scoring metrics. On the other hand, much of the research literature related to the elaborated models of Section 6.2.2 use case studies to provide intuitive and qualitative advice for setting the hyper-parameters.

6.4.3.1 Hyper-parameter selection

The RF regressor class that is used to perform regression-based anomaly detection (Section 6.2.5) has multiple hyper-parameters to adapt to specify the model to our application. According to Müller and Guido (2016), the important parameters to adjust are *n_estimators*, *max_features*, and *max_depth*. The *n_estimators* represent how many trees to build (i.e. larger is always better, but also more time complex). The *max_features* parameter represents the number of features that are selected in each node of the decision tree and determines how random each tree is. Including all features of the training sample is recommended to use for the parameter *max_features* in regression analysis. The *max_depth* parameter represents the maximum number of levels in each decision tree to reduce the complexity of each tree.

The IF has the following important parameters to adjust *n_estimators* and the *max_samples*. Compared to the random forest *max_depth* is not an important hyper-parameter since the limit is automatically set by the average tree height of the sub-sampling size. The *n_estimators* parameter represents how many trees to build. Liu et al. (2008) described *n_estimators* equal to 100 is sufficient usually. The *max_samples* parameter represents the number of samples to draw from the training data to train each decision tree. Empirically, Liu et al. (2008) found that setting *max_samples* to 2^8 generally provides enough details to perform anomaly detection across a wide range of data.

One of the advantages of LOF is that there is only one important hyper-parameter *n_neighbors* to adapt to achieve an increase in model performance. The *n_neighbors* parameter represents which is the number of nearest neighbors used in defining the local neighborhood of the object. According to Breunig et al. (2000), this hyper-parameter value could be application-dependent, but picking 10 to 20 appears to work well in general.

Hyper-parameters	**Grid RF**
n_estimators	[50, 100, 500]
max_features	[None, 25, 100]
max_depth	['auto', 'log2']
threshold setting	[0.09, 0.1, 0.11, 0.15]

Table 6.7: Grid RF.

Hyper-parameters	**Grid IF**
n_estimators	[50, 100, 500]
max_samples	[100, 256, 3000]
threshold setting	[0.09, 0.1, 0.11,0.15]

Table 6.8: Grid IF.

Hyper-parameters	**Grid LOF**
n_neighbors	[10, 20, 50, 500, 1000]
threshold setting	[0.09, 0.1, 0.11, 0.15]

Table 6.9: Grid LOF.

6.4.3.2 Anomaly detection model grid search

Reliable comparisons of the built models in Table 6.3.2.2 can only be made when the models are trained with the same training data and tested with the same test data. The ground-truth dataset contains data of both Machine Y types (i.e. C and A,B). Therefore, we created a training dataframe consisting of data of both Machine type C and Machine Type A,B. A simple random sample of 10,000 filtered records of Figure 6.9 is used as training data. It is assumed this sample contains sufficient data for building the anomaly detector. The RF and IF models in Table 6.3.2.2

are trained based on the same dataset. The LOF algorithm does not need any training and can be used just for the test set. We used RF 2, IF/LOF 2, and IF/LOF3 in Table 6.3.2.2 for finding potential anomalies since looking at individual priority component motors can be interesting for finding interesting anomaly types. However, for scoring we have data records available containing data of all priority component motors and therefore priority is given to models in Table 6.3.2.2 that have been collectively analysed. Hence, all models where *collectively* is written in Table 6.3.2.2 have been scored based on the grid parameters provided in Section 6.4.3. Additionally, IF1 has been scored since we also want to know how the IF and LOF score when no temporal data is included (i.e. no rolling window feature). Lastly, RF1 has also been scored because we wanted to compare the scores of a RF regressor where the power prediction has been based on the parameters of the individual motor or where the power prediction has been based on the parameters of all other motors. Appendix G contains all precision, recall, and f-score scores regarding the specific input hyper-parameters. The best scores per model have been summarized in Tables 6.10 and 6.11 based on the two different ground-truth datasets elaborated in Section 6.4.1.

Model	**precision**	**recall**	**F_β-score**
RF1	0.69	0.62	0.65
RF3	0.67	0.75	0.72
IF1	0.88	0.98	0.94
LOF1	0.9	1.0	0.95
IF4	0.99	1.0	1.0
LOF4	0.99	0.99	0.99
IF5	1.0	1.0	1.0
LOF5	0.96	0.96	0.96

Table 6.10: Scoring anomaly detection models belt drives without uninteresting anomalies.

Model	**precision**	**recall**	**F_β-score**
RF1	0.67	0.75	0.72
RF3	0.65	0.65	0.65
IF1	0.66	0.90	0.78
LOF1	0.62	0.85	0.74
IF4	0.67	1.0	0.83
LOF4	0.67	1.0	0.83
IF5	0.67	1.0	0.83
LOF5	0.67	1.0	0.83

Table 6.11: Scoring anomaly detection models belt drives including uninteresting anomalies.

6.4.4 Conclusion

The ideal situation would be a dataset where each condition-data record has been labeled by either *anomalous* or *normal*. Then, we know with certainty the recall and precision score for the built models. However, more than 100,000,000 data records cannot be all labeled by utilizing expert knowledge and therefore we created two ground-truth datasets that should represent the population (as far as possible). Nevertheless, we know for sure this ground-truth dataset does not contain all different types of anomalies. This implies the recall score cannot be reliably interpreted as this scoring method looks at how well the anomaly detection model can detect all different types of anomalies. As we do not know all different types of anomalies, we can only base the recall score on the distinguished anomaly types. Conversely, the precision can be reliably interpreted as this scoring method defines how well the model can detect the known anomalies. Hence, we have to take these general remarks into account when discussing the results of Table 6.10 and 6.11.

The first important distinction that can be made based on Tables 6.10 and 6.11 is that the anomaly detection models score higher when uninteresting anomalies are not included in the ground-truth dataset. This implies the anomaly detection models can distinguish fluctuating power behavior from normal data points of Machine Y. However, the anomaly detection model cannot distinguish the difference between interesting and uninteresting anomalies since they both contain fluctuating power behavior. Therefore, anomaly diagnosis is important to learn the model how to distinguish interesting from uninteresting machine behavior.

Furthermore, Table 6.10 shows the IF and LOF score high in detecting anomalies compared to the RF regressor. This denotes the IF and LOF can better distinguish fluctuating power behavior from normal data points than the RF. Moreover, IF/LOF 1 (i.e. no rolling window feature) score lower on the F_β-score than IF/LOF 4 and IF/LOF 5 (i.e. rolling window feature included) which means the rolling window feature adds key information as input to the anomaly detector to make better predictions. Additionally, the difference between IF/LOF 4 and IF/LOF 5 is the time horizon of the rolling window. As these models show hardly any difference in performance we can state the rolling window of five minutes does not make a difference compared to a rolling window of thirty minutes. Moreover, we used the standard deviation and the min-max range of a predefined time period as a rolling window calculation. As the scores of these motions do not differ significantly, we can state that distinguishing between min-max and the standard deviation does not make a difference and are both performing well in detecting anomalies.

We do have to take some debatable statements into account when interpreting the results of Table 6.10 and 6.11. (i) The distinguished anomaly types have been found and evaluated by consulting the same unsupervised anomaly techniques (Table 6.4). This implies we are in a circle that starts with distinguishing interesting points by unsupervised modeling, validate these points with experts, and then evaluating these points with the same unsupervised models resulting in high scores (Table 6.10). (ii) We did not use cross-validation as the ground-truth dataset is too small. This denotes we cannot assess the generalization performance of the anomaly detector. (iii) As the ground truth dataset does not contain all different types of anomalies we cannot interpret the recall score reliably as elaborated at the beginning of this section.

6.5 Conclusion

In this section, we answer research question 5: *How to detect a failed or under-performing priority component?* by going through the stages of Figure 6.3. We stated application-specific interesting anomalies (i.e. global and local outliers) should be able to find with multidimensional data where the data records are time-independent as we take advantage of the assumption that interesting anomalies are 'few and different' compared to the majority of the data. This assumption does not always hold as operator changes can also lead to 'few and different' situations which do not represent a failed or under-performing priority component. Therefore, we reduce the input data (Figure 6.8 and 6.9) to only data that contain no uninteresting anomalous behavior situations (i.e. operator changes, machine state changes). This reduced input dataframe is used for building and evaluating multiple unsupervised anomaly detection models.

Two interesting anomaly types were distinguished related to the fan (Figure 6.2). As the first type of anomaly can be found with a simple rule and the second type of anomaly does not have a diagnosis, it is not interesting to further evaluate how well the anomaly detection models (Table 6.3.1.2) score related to the main air distribution fan. On the other hand, the belt motors failed more frequently and different types of anomalies were distinguished in Table 6.5. This argues the belt drive motors anomaly detection methods (Table 6.4) are interesting for further evaluation.

Interpreting the anomaly detection scores made us conclude the models can distinguish fluctuating power behavior from normal data points. However, the anomaly detection model cannot distinguish the difference between interesting and uninteresting anomalies since they both contain fluctuating power behavior. Therefore, anomaly diagnosis is important to learn the model how to distinguish interesting from uninteresting machine behavior. Furthermore, evaluating the models resulted in the finding that IF and LOF can better distinguish fluctuating power behavior from normal data points than the RF. This is because the IF and LOF can include temporal information (i.e. rolling window calculations) and no specific favor of one modeling technique over the other was found. Furthermore, it appeared the hyper-parameters of five and thirty minutes rolling window calculations and the two different motions standard deviation and min-max all scored high in detecting anomalies and did not lead to favors to a particular combination.

Hence, we found high-scoring models that can detect a failed or under-performing priority component. The models score higher than the predefined precision and recall scores which are 80% and 95%, respectively. However, we do have to take into account this score is based on an evaluation of a debatable ground-truth dataset. This is because the distinguished anomaly types have been found and evaluated by the same modeling technique, no cross-validation has been used because of a small ground-truth dataset, and the recall score that cannot be reliably interpreted as the ground-truth dataset does not contain all different types of anomalies.

Chapter 7

Anomaly diagnosis

Anomaly diagnosis modeling is a set of techniques and systems to automatically find the root cause of a failed or under-performing priority component. Diagnosis deals with fault detection, isolation, and identification. Fault detection is a task to indicate whether something is going wrong in the monitored system, fault isolation is a task to locate the cause and component that is faulty, and fault identification is a task to determine the nature of the fault when it is detected (Jardine et al., 2006). In practice, the identification phase appears rarely and is sometimes connected with fault isolation. Thus, the diagnostic process includes usually only two phases: fault detection and isolation (Isermann and Ballé, 1996). Fault detection has already been extensively elaborated and discussed in Chapter 6. Automatic fault isolation is explained in this chapter by answering research question 6:

How to find the root cause of a failed or under-performing priority component?

This chapter starts with elaborating on the labeled conveyor belt drive input dataset in Section 7.1. Hereafter, Section 7.2 explains the models used for diagnosing purposes. The distinguished anomaly diagnosis models are evaluated and improved in Section 7.3. Finally, this chapter is concluded in Section 7.4.

7.1 Data selection

Chapter 6 concluded by stating high scoring models have been found that can detect failed or under-performing conveyor belt drives. The ground-truth dataset (Section 6.4.1) consists out of only the labels *anomaly* and *normal* and contains insufficient instances for learning purposes in finding root causes automatically. The more labeled data, the better for learning purposes of the ML model. Therefore, the highest-scoring models (i.e. IF 4, LOF4, IF5, LOF5) have been selected to find potential anomalies that can be labeled with a root cause by utilizing domain knowledge to increase the size of the labeled dataset. Hence, during the interviewing sessions with experts, we did not only wanted to know the *anomalous* or *normal* label, but also the specific diagnosis label. Similar to the ground-truth dataset of anomaly detection (Section 6.4.1), we aim to build a dataframe consisting out of 90 % normal data records and 10% anomalous data records since the machine is working properly most of the time. The interesting and uninteresting (i.e. lubrication and belt load changes) anomalies have been found by consulting the anomaly detector and analysing consequential time-series plots. The remaining normal datapoints have been found by picking the lowest IF scored data records (similar way as in Section 6.4.1).

Some anomalous parameter behavior appeared over a longer time range which means these anomalies also contain more sequential datapoints than anomaly types that appeared over a short time range. To minimize the imbalance of the dataset as much as possible, we took a simple random sample (n=100) from each distinguished anomaly case. In the end, we build a labeled dataframe consisting out of the following root causes with the consequential amounts of cases distinguished (Table 7.1).

Anomaly/ Normal	**Diagnosis**	**Amount of cases distinguished**	**Data records**
Anomaly	Base drum got stuck	9	900
Anomaly	Belt got stuck	5	500
Anomaly	Belt in return section got stuck	10	1.000
Anomaly	Belt tension mechanism got stuck	12	1.200
Normal	Conveyor belt got lubricated	9	900
Anomaly	Drum pulls other drum	17	1.700
Normal	Product weight transition	20	2.000
Anomaly	Undefined	7	700
Normal	Normal data records	-	51.100

Table 7.1: Amount of distinguished diagnosis.

This dataframe (Table 7.1) has the same input features as for anomaly detection. These features are (i) important parameters related to the conveyor belt (Table 5.3), (ii) rolling window calculations, (iii) relative power behavior of the conveyor belt, (iv) and the time in the operating state which can all be found in Table C.4. Two dataframes have been built out of Table 7.1 as the table contains procedural error, which influences the number of empty values in an input feature (Section 5.1.3). Dataframe 1 and 2 consist out of all the input features explained at the beginning of this paragraph, except the current and speed of the conveyor belt drive motors have not been taken into account in dataframe 1. These parameters are not included since they contain procedural errors for some customers. The data records of dataframe 2 containing empty values (i.e. procedural error) are deleted as the ML models cannot deal with empty values.

7.2 Model selection

The upper part of the process in finding root causes automatically (Figure 7.1) is elaborated in Chapter 6. Conversely, building the classifier to enable automatic anomaly diagnosis is elaborated in this chapter. There are multiple root causes, so we have to deal with a multiclass classification problem. By recognizing patterns between the input features and the root cause labels, we aim to build a model that can predict a root cause based on an input vector (Jardine et al., 2006).

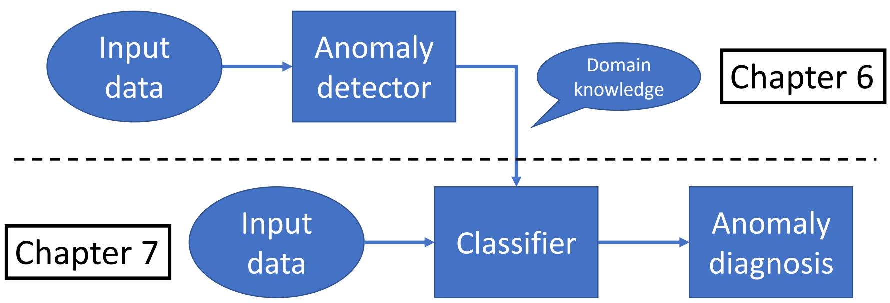


Figure 7.1: Finding root causes process.

Analytical, mathematical, or ML models are most frequently mentioned as techniques to build classifiers in literature (Jardine et al., 2006). The application of analytical or mathematical models is usually limited to linear systems that cannot fully describe the complexity of the Machine Y parameter behavior over time which can lead to dissatisfying results. In these cases, ML techniques based on a knowledge base become an attractive tool for constructing desired models that show improved performance over conventional approaches (Korbicz et al., 2012; Jardine et al., 2006; Siddique et al., 2003). Therefore, this thesis consults ML techniques to build the classifier for automatically finding root causes.

Multiple off-the-shelve models are available that can deal with multiclass classification. Three of these models (i.e. k-Nearest Neighbors, Random forest, and neural networks) are selected that can be evaluated and compared to each other. Sections 7.2.1, 6.2.5.1, and 7.2.2 discuss why these three models have been selected and how they work.

7.2.1 k-Nearest neighbors

k-Nearest Neighbors (KNN) is arguably the simplest ML model (Müller and Guido, 2016). Building the model consists only of storing the training dataset. To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset (i.e. nearest neighbors). The parameter k considers the number of neighbors that are used to make a majority class prediction. One of the strengths of KNN is that the model is very easy to understand, and often gives reasonable performance without a lot of adjustments. Using this algorithm is a good baseline method to try before considering more advanced techniques.

7.2.2 Artificial neural networks

An Artificial Neural Network (ANN) is a computational model of the brain and can capture domain knowledge from examples (Pham and Pham, 1999). ANN can be viewed as generalizations of linear models that perform multiple stages of processing to come to a decision (Müller and Guido, 2016). ANNs assume that computation is distributed over several simple units called neurons, which are interconnected and operate in parallel (Siddique et al., 2003). Implicit knowledge is built into a neural network by training it. The ANNs can be trained by typical input patterns and the corresponding expected output patterns. The error between the actual and expected outputs is used to update the weights and biases to increase the model's accuracy. ANNs have reemerged as state-of-the-art models in many applications of machine learning. They can handle both continuous and discrete data and have good generalization capability. A major limitation of ANN is the difficulty to have physical explanations of the trained model.

7.3 Evaluation and model improvement

As 90% of the diagnosis dataframe (Section 7.1) represents normal data points and only 10 % represents anomalous behavior, we deal with an imbalanced dataset. Therefore, the scoring methods (i.e. precision, recall, and F_{beta}-score) explained in Section 6.4.2 also apply to the labeled diagnosis dataframe. A grid search is performed to find the hyper-parameters resulting in the highest model score.

7.3.1 Grid search

A grid search is performed by trying multiple combinations of hyper-parameters (Section 7.3.1.1) to reach for the highest model performance. Cross-validation is performed to evaluate the generalization performance of the models (Section 7.3.1.2).

7.3.1.1 Hyper-parameter selection

According to Müller and Guido (2016), the following parameter is important to adjust *n_neighbors* in KNN (Table 7.2). This parameter represents the number of neighbors that are used for making the majority class prediction. Conversely, the explanation of the hyper-parameters connected to a RF (Table 7.3) can be found in Section 6.4.3.1. Finally, according to Müller and Guido (2016), the most important parameters to adjust for ANN are: *hidden_layer_sizes, solver, alpha*, and max_iter (Table 7.4). The *hidden_layer_sizes* parameter consists out of vectors where the ith element represents the number of neurons in the ith hidden layer. Moreover, the *solver* parameter represents the solver for weight optimization. The parameter value *adam* refers to a stochastic gradient-based optimizer proposed by Kingma and Ba (2014). The parameter value *lbfgs* refers to an optimizer in the family of quasi-Newton methods. Additionally, the hyper-parameter *alpha* is the learning rate. Finally, the *max_iter* hyper-parameter corresponds to the maximum number of solver iterations if no convergence has already been reached. In principle, more iterations result in better behaviour, however a trade-off should be made with computation time.

Hyper-parameters	Grid KNN
n_neighbors	[(10,),(10,10), (100,100)]

Table 7.2: Grid KNN.

Hyper-parameters	Grid RF
n_estimators	[50, 100, 500]
max_depth	[None, 25, 100]
max_features	['auto', 'log2']

Table 7.3: Grid RF.

Hyper-parameters	Grid ANN
hidden_layer_sizes	[10,100,300]
solver	['adam','lbfgs']
alpha	[0.0001, 0.01]
max_iter	[200, 400]

Table 7.4: Grid ANN.

7.3.1.2 Grid search with cross-validation

The data is split into training and test data, to test the accuracy of the model on previously unseen data. Cross-validating is a statistical method of evaluating the generalization performance. When performing a simple grid search there is a danger of overfitting the parameters and validation set as for each combination of hyper-parameters a classifier is trained and evaluated only once. In this situation, no independent dataset is used for evaluating purposes which results in a classifier that can overfit the training data resulting in a biased test score. Therefore, grid search with stratified 5-fold cross-validation (Figure 7.2) is utilized to measure how well the model generalizes to new data. The stratified cross-validation takes into account the proportions between classes are the same in each fold as they are in the whole dataset (Müller and Guido, 2016). Hence, this type of cross-validation is slightly different as the normal k-fold cross-validation (Section 6.2.5.2). For each parameter setting, five F_β-scores are calculated, one for each split in stratified cross-validation. Then the mean F_β-score is computed for each parameter setting. The parameters with the highest mean validation F_β-score are chosen. The final evaluation is performed by another stratified 5-fold cross-validation where the retrained model based on the best parameters is scored based on the unseen test data resulting in Table 7.5 and 7.6.

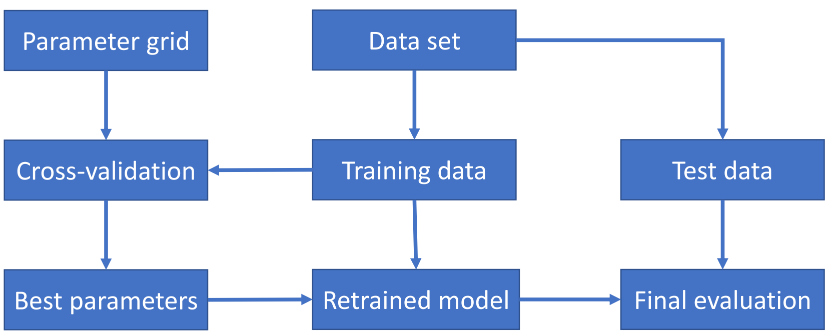


Figure 7.2: Overview of model evaluation with grid search and cross-validation.

As can be seen in Table 7.5 and 7.6, the anomaly diagnosis models score slightly better, based on dataframe 1. RF models perform the best compared to the KNN and ANN. A stratified 5-fold cross-validated F_β-score of 0.888 is high, but further research is necessary to state which distinguished diagnoses (Table 7.1) can be classified with great certainty and which cannot. Therefore, we took the best anomaly diagnosis model (i.e. RF based on dataframe 1) and investigated the stratified 5-fold cross-validation scores for all diagnosis types resulting in Table 7.7.

When looking at the diagnoses in particular (Table 7.7), we can see the following labels can be classified with great certainty: *belt got stuck*, *belt in return section got stuck*, and *normal data records*. Possible reasons for these certain classifications are (i) sufficient amount of training data or (ii) easily distinguishable patterns connected to a root cause. The other diagnoses perform slightly worse compared to previously mentioned diagnoses. The *product weight transition* and *drum pulls other drum* are diagnoses that are harder to classify for the diagnosis model. Possible reasons why these diagnoses score lower are (i) they need more training data to recognize the specific pattern, (ii) the input feature is not sufficient in providing information of the specific diagnosis, or (ii) the expert labeled these diagnoses not consistently.

Model	**precision**	**recall**	**F_β-score**
KNN	0.886	0.837	0.857
RF	0.920	0.867	0.888
ANN	0.887	0.832	0.854

Table 7.5: Anomaly diagnosis scores based on dataframe 1.

Model	**precision**	**recall**	**F_β-score**
KNN	0.876	0.826	0.846
RF	0.903	0.841	0.866
ANN	0.828	0.837	0.833

Table 7.6: Anomaly diagnosis scores based on dataframe 2.

Diagnosis	**precision**	**recall**	**F_β-score**
Base drum got stuck	0.772	0.834	0.807
Belt got stuck	0.952	0.991	0.974
Belt in return section got stuck	0.994	0.998	0.996
Belt tension mechanism got stuck	0.657	0.742	0.704
Conveyor belt got lubricated	0.689	0.789	0.744
Drum pulls other drum	0.431	0.623	0.526
Product weight transition	0.402	0.563	0.483
Undefined	0.675	0.752	0.718
Normal data records	0.999	0.999	0.999

Table 7.7: Scores anomaly diagnosis based on RF dataframe 1.

7.4 Conclusion

In this section, we answer research question 6: *How to find the root cause of a failed or under-performing priority component?* ML techniques (i.e. KNN, RF, and ANN) have been used to recognize conveyor belt drive patterns between the input vector (i.e. sensor data) and the label (i.e. root cause). The RF scored higher (F_β-score = 0.888) than the ANN and KNN. Therefore, the RF model is used to investigate how well the distinguished root causes can be identified. Three root causes were able to be distinguished with great certainty. Four root causes scored well for precision and recall. Only two root causes had a F_β-score around 0.5. Hence, we can conclude the old situation where data specialists at Company X had to specifically look at parameter behavior can be transformed to a situation where the anomaly diagnosis model can reliably classify some root causes. This saves time of the Company X data specialists.

Chapter 8

Discussion

This chapter draws overall conclusions in Section 8.1. Recommendations that contain specific solutions and directions based on the research findings are provided in Section 8.2. Finally, limitations and future research directions are explained in Section 8.3. This section addresses the influences that cannot be controlled because of time or available data restrictions. Future research directions are given to counterattack these shortcomings.

8.1 Conclusions

In this section, we answer the main research question in the first paragraph: *How can CBM be applied on critical components of Machine Y and what is the added value compared to the current situation?* Furthermore, multiple additional conclusions can be drawn which are elaborated in the sequential paragraphs.

CBM is applied on Machine Y priority components and adds value compared to the current situation

Firstly, priority components are distinguished for Machine Y that are suitable for CBM. Secondly, an anomaly detection model is built and evaluated based on the present-day sensor data. This anomaly detector facilitates building and evaluating the anomaly diagnosis model which succeeds an important step of enabling CBM (Jardine et al., 2006). The added value compared to the current Machine Y situation is that knowledge has increased about how a failure or malfunction behaves of a priority component. Where initially the knowledge of failures was only in expert's minds, the knowledge has been made tangible and transparent now. Furthermore, the added value of this study is an anomaly detector and anomaly diagnosis model. These models can assist the Company X data specialist in automatically classifying root causes based on sensor data behavior. However, the added value of this research is not explainable in monetary terms as we concluded in Chapter 2 no sufficient data was available to make a reliable prediction about the current average maintenance costs. Additionally, we did not manage to build a fully working CBM decision support model that can be compared to the current situation.

Priority components have been distinguished that are suitable for CBM

As we cannot investigate CBM for each Machine Y component, we found suitable candidates (i.e. main air distribution fan, conveyor belt, and conveyor belt drives) for CBM by applying the three-stage funnel developed by Tiddens et al. (2018) shown in Figure 3.2. However, two major risks that were identified appeared to have consequences for the remaining of this research. The main air distribution fan has scarce maintenance data available and seems to be a robust component. Performing anomaly detection resulted in only two types of interesting anomalies that were not sufficient for further evaluation. Furthermore, the conveyor belt has no direct sensors monitoring the behavior. In anomaly diagnosing we concluded too little maintenance data was available to classify the right conveyor belt related failures. Conversely, the conveyor belt drives had sufficient failure frequencies and related sensor information to perform anomaly detection and diagnosis.

Knowledge has been increased about a failed or malfunctioning Machine Y component

We used the advise of Aggarwal (2015) to run an unsupervised anomaly detector, select a pre-filtered potential anomaly dataframe, and validate these points by utilizing expert knowledge. This validation step results in the added value that experts knowledge is not only in their minds, but can also be made tangible and transparent by showing time-series parameter behavior with the expert's label (Table 6.5 and 6.2). This transparent knowledge results in a better understanding of how Machine Y can fail or lose performance that is important for making prognostics. Furthermore, these anomalous parameter behaviors that come as output from the unsupervised anomaly detectors are not detected by the present-day rules that are implemented in Machine Y. Therefore, we reached an improvement in making the event data more accurate that can help in improving the prediction algorithm as well.

LOF and IF are higher scoring anomaly detectors than RF regression

The previous conclusion elaborated a situation is created with a ground-truth dataset that is used for scoring the anomaly detection models. This paragraph interprets the data-driven anomaly detection algorithms results that were enabled by understanding and preprocessing the data firstly. The main difference between LOF and IF on one side and the RF regressor on the other side is the inclusion of rolling window features (i.e. temporal data) in the LOF and IF models. We found the LOF and IF scored higher in precision, recall, and the f-score than the RF regressor. However, we do have to take into account this score is based on an evaluation of a debatable ground-truth dataset. This is because the distinguished anomaly types have been found and evaluated by the same modeling technique, no cross-validation has been used because of a small ground-truth dataset, and the recall score that cannot be reliably interpreted as the ground-truth dataset does not contain all different types of anomalies.

High-quality data is a must when effectively and efficiently applying ML to enable CBM

ML maximizes its relevance when high-quality data is provided. The ideal situation would be to exactly know when a failure or malfunction took place at a customer. Many years of high-quality data management, arises a situation where the model can learn based on the huge amounts of labeled data what specific patterns arise at a specific failure. This is where CBM can achieve its highest performance. Unfortunately, this high-quality dataset was not available in this research meaning we did not have many instances that the ML algorithm can use to recognize patterns. Furthermore, labeling a specific root cause of a failure is difficult for experts even though the parameter behavior has been shown in time-series plots. This way of labeling can increase the bias in the input data as categorizing a root cause to a specific parameter behavior is subjective. Therefore, this research concludes the statement of Jardine et al. (2006) about equal importance of quality in event and condition monitoring data is correct. Having a high-quality condition monitoring system would not automatically result in a scenario where CBM can be effectively and efficiently enabled, accurate event data is also key.

Evaluated an anomaly detection method that has not been extensively elaborated in literature

Regression-based anomaly detection based on FMEA findings has not been extensively elaborated in literature as far as the author is aware. Aggarwal (2015) wrote a more or less similar approach. However, that approach repeatedly applies a regression model by selecting an attribute as the dependent variable and the remaining attributes as the independent variables. Hence, the method found by Aggarwal (2015) builds as many regression models as there are input values and aggregates the scores at the end. Section 6.6 contains a detailed explanation of how the regression-based anomaly detection works for our case. An important parameter distinguished via an FMEA is used as label for the RF regressor. Other relevant input parameters are used to predict the important parameter. This methodology has been evaluated and compared to the other anomaly detection algorithms. However, Figure 6.10 shows the new anomaly detection model scores lower than the LOF and IF.

8.2 Recommendations

The conclusions result in the following distinct recommendations when willing to take more steps in PdM and to reproduce this research on other machines. The first and second recommendation is company-specific whereas the third and fourth recommendation is more general.

Improve the data acquisition process in four ways

In our research we consulted domain knowledge to validate if potential anomalies are truly interesting anomalies or not. This process is time-consuming and we noticed the experts sometimes doubted to classify a specific root cause to specific parameter behavior. Therefore, the quality of both condition data and event data should be high to use ML in maintenance decision making. Furthermore, the business case, which is not numerically assessed in this research because of a lack of data (Section 2), should be a key incentive for enabling CBM. Hence, improving the data acquisition process is a must when willing to mature in PdM and can be reached in the following ways:

1. *Create awareness of documentation importance among FSEs*
   Currently, FSEs do not see the importance of high-detail documentation and therefore they mostly do not archive well, but it is one of the key factors in ML.

2. *Inspect replaced components of the PMS*
   In Chapter 2, we stated since FSEs made clear components are replaced even though hardly any wear and tear can be distinguished, the PMS replacements cannot be used as information to know when a specific failure happened at a customer. Consequently, the PMS replacement intervals are not included in the maintenance data that we use to validate our distinguished anomalies and diagnostics in Chapter 6 and 7, respectively. Therefore, we recommend inspecting the replaced components of the PMS on the remaining quality after it has been replaced. Categorizing the replaced component on a scale that makes clear what degradation level can be distinguished. This information is useful for determining what the replacement period of a component should be and we can use the data for ML purposes. Some components are truly almost at the end of their lives and can be included in the maintenance data of (almost) failed components, while components that hardly show any wear and tear should not be included in the maintenance data as an (almost) failed component.

3. *Add more information sources to CXMDS*
   Documenting downtime in the asset management system is important to assess the criticality of components. There should be added a column in the CXMDS tool so FSE can document this information as well. Another feature that should be added to CXMDS is the date of replacement. Nowadays, only the starting and ending date of a specific case is archived. However, this time range can be very long and therefore an extra column in CXMDS containing information about the specific date of replacement is desired. Lastly, an extra column in CXMDS where the FSE can document if corrective or preventive maintenance is performed improves the maintenance data quality.

4. *Create awareness of documentation importance among customers*
   Sometimes maintenance is performed by the customer themselves or a third party. The recommendation that can be given in this case is to make the customer aware of a highly detailed maintenance list that agrees upon CXMDS standards (i.e. with additional recommendations) with the message this information can lead to better performance of Machine Y in the end. Company X can also be proactive in this case by calling the customer if a potential anomaly has been distinguished and ask them what the reason is.

Implement two sensor types to increase the predictive value of a component's remaining useful lifetime

Two specific sensor types are distinguished that can be implemented on Machine Y for prognosis purposes.

1. *Sensors that monitor the length and tension of the conveyor belt*
   There is no sensor information about the length or the tension of the conveyor belt. This denotes that we are dependent on the ML models to find a relation between the sensor data of the conveyor belt drives and the maintenance data of a conveyor belt-related failures or malfunctions. As the maintenance data is rather scarce we did not succeed in building an anomaly detection model for the conveyor belt (i.e. only for the conveyor belt drive motors). Furthermore, even if there would be high quantity maintenance data regarding conveyor belt failures, more interesting condition data should be measured as components can show degradation before failing. Investigating the run-up to a failure is key and can be transparently discussed with experts by analysing condition data. Hence, we recommend adding the following sensors to Machine Y: conveyor belt load sensor and continuously measuring the position of the belt tension mechanism. Instead of continuously measuring the position of the belt tension mechanism, a sensor that measures the belt length would also satisfy. This sensor measures the time between some fixed distance points of the conveyor belt.

2. *Sensors that monitor the vibration of motors*
   Vibration sensors are mentioned frequently in literature to increase the predictive value of a remaining useful lifetime of a component (Carden and Fanning, 2004). Vibration analysis works as follows, a machine in standard condition has a certain vibration signature. Fault development changes that signature in a way that can be related to the fault (Randall, 2021). According to Farrar and Doebling (1999), the most mature and successful application of vibration-based damage detection technology has been in the monitoring of rotating machinery (which is the case in Machine Y). Furthermore, vibration analysis reacts immediately to change. Moreover, many powerful signal processing techniques can be applied to vibration signals to extract even very weak fault indications from noise and other masking signals (Randall, 2021).

Perform time-series analysis in anomaly detection to find gradual wear and tear

Anomalies could occur in one of two possible ways (Aggarwal, 2015): (i) The values and trends in the data stream change slowly over time. In such cases, the anomalies can only be detected by careful analysis over a longer period of time, and is not immediately obvious in many circumstances. (ii) The values and trends in the data stream change abruptly. By utilizing multi-dimensional data where records are time-independent and assuming anomalies were 'few and different' we mainly found abrupt anomalies and fluctuating behaviors of power. However, we did not find anomalies of gradual wear and tear as they do not follow the assumption of 'few and different'. Nevertheless, these gradual wear and tear circumstances are interesting anomalies to find since (i) experts expect the power of priority component motors and the length of the conveyor belt to gradually increase as the machine gets older. (ii) It is this gradual loss of function that leads to a predictable situation for failure prognostics. (iii) The temporal aspects of the time-series data are important to the extent that the anomalies are defined with respect to the history rather than the entire data set (i.e. collective anomalies in Section 6.1). These gradual anomalies cannot be distinguished based on data where records are time-independent and time-series analysis is key to find this type of interesting anomalies. However, research time restrictions prevent us from starting this type of data analysis.

Our research focused on operating machine states (Figure 5.4). However, since the data linked to these operating machine states are exposed to many different circumstances (i.e. changes in climate setpoints, product weights, and belt speeds) that complicate the time-series analysis, we recommend starting to perform time-series analysis on a more stable machine state (i.e. cleaning state) that should behave more similarly among customers.

Focus on more frequently failed components to reach a richer ML environment

We did not find sufficient interesting anomaly types for the main air distribution fan as this priority component is robust and hardly fails. This implies the impact of failure has been taken more into account than the frequency of failure when distinguishing this component as a high priority component for CBM. We learned that since this is a pilot study towards making the first steps in PdM, we better could have focused on components that fail more frequently. Then we were better able to show the added value of CBM since we had more examples to learn from in detection, diagnosing, and prognosing. Hence, this research recommends investigating priority components that fail more frequently instead of highly prioritizing the impact of a failure in the beginning stages of CBM. After the added value of CBM has been shown we can further implement the maintenance policy to components that fail less frequently but have a higher impact. Tiddens et al. (2018) did mention a lower limit on the failure frequency to help selecting only those candidates that fail often enough for a positive business case in his funnel approach (Figure 3.2). However, we interpreted huge downtime costs in combination with lower failure frequencies as a positive business case at that point during this research.

8.3 Limitations and future research

During this research, multiple limitations have been distinguished that should be taken into account when using the results of this thesis. Future research directions are provided to solve these issues.

Not fully satisfying the transparency principle of CBM

One of the main advantages of CBM is that decisions can be based on transparent data rather than expert opinions. As we had insufficient numerical data available regarding components' frequency of failure and corresponding downtimes (Chapter 2), we had to perform an FMEA to assess which components were suitable for CBM. Moreover, since the priority components did not have sufficient maintenance data available to reliably find interesting anomaly types, we validated potential anomalies with domain knowledge. This implies we needed to utilize domain knowledge instead of data-based decision making to proceed in CBM. Consequentially, we cannot completely satisfy one of the advantages of CBM (i.e. transparency). Future research that prevents this limitation from happening can be to perform multiple actions to improve the data acquisition process (Section 8.2).

Debatable ground truth dataset

More than 100,000,000 data records cannot be all labeled by utilizing expert knowledge and therefore we created a ground-truth dataset that should represent the population and enables a test set for scoring purposes. However, two limitations can be distinguished inherent to the ground truth dataset. The first limitation is that the distinguished anomaly types have been found and evaluated by consulting the same unsupervised anomaly techniques (Table 6.4). This implies we arrived in a repetitive circle that starts with distinguishing interesting points by unsupervised modeling, validate these points with experts, and then evaluating these points with the same unsupervised models resulting in (biased) high scores (Table 6.10). The second debatable point is the incompleteness of the ground truth dataset as we did not analyze other machine states than the operating state and the machine can also break while being in another state of Table 5.4. Due to research time restrictions, we can only provide future research directions to analyze the data related to other machine states than the operating state. Furthermore, we did not manage to find gradual wear and tear as elaborated in the recommendations (Section 8.2). This incomplete ground-truth dataset has the following consequences: (i) we stated the precision is still reliable, but the recall is not since this scoring method takes into account *false negatives*. Therefore, it is necessary to know all different types of anomalies since that is the only way to know how accurate the anomaly detection model is. (ii) We did not use cross-validation as the ground-truth dataset is too small. This denotes we cannot assess the generalization performance of the anomaly detector. Future research in satisfying all recommendations of Section 8.2 is necessary to enrich the ground truth dataset.

FMEA only built for distinguished priority components

As performing an FMEA on whole Machine Y would be too time-consuming, we focused ourselves on priority components distinguished together with our stakeholders. However, performing an FMEA for the whole machine could have led to other components being distinguished as important as they fail more frequently or have a higher consequential downtime than the fan or belt drive motors. Future research is necessary to perform the FMEA on all interesting components of Machine Y to distinguish a higher amount of suitable components for CBM.

Bibliography

Aggarwal, C. C. (2013). Outlier ensembles: position paper. *ACM SIGKDD Explorations Newsletter*, 14(2):49–58. 38

Aggarwal, C. C. (2015). Outlier analysis. In *Data mining*, pages 237–263. Springer. 34, 35, 37, 40, 48, 59, 60, 61

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In *International conference on database theory*, pages 420–434. Springer. 77

Arabian-Hoseynabadi, H., Oraee, H., and Tavner, P. (2010). Failure modes and effects analysis (fmea) for wind turbines. *International Journal of Electrical Power & Energy Systems*, 32(7):817–824. 12

Baines, T. and Lightfoot, H. W. (2014). Servitization of the manufacturing firm. *International Journal of Operations & Production Management*. 5

Banks, J. C., Reichard, K. M., Hines, J. A., and Brought, M. S. (2008). Platform degrader analysis for the design and development of vehicle health management systems. In *2008 International Conference on Prognostics and Health Management*, pages 1–12. IEEE. 12

Bengtsson, M. (2008). A method for implementing condition based maintenance in industrial settings. In *18th international conference on flexible automation and intelligent manufacturing*, pages 348–356. University of Skövde. 17

Bland, J. M. and Altman, D. G. (1996). Statistics notes: measurement error proportional to the mean. *Bmj*, 313(7049):106. 27

Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104. 37, 50

Briscoe, E. and Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, 118(1):2–16. 36

Carden, E. P. and Fanning, P. (2004). Vibration based condition monitoring: a review. *Structural health monitoring*, 3(4):355–377. 61

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58. 33, 34, 37

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., et al. (2000). Crisp-dm 1.0: Step-by-step data mining guide. *SPSS inc*, 9:13. 5

Cheng, Z., Zou, C., and Dong, J. (2019). Outlier detection using isolation forest and local outlier factor. In *Proceedings of the conference on research in adaptive and convergent systems*, pages 161–168. 36

Farrar, C. R. and Doebling, S. W. (1999). Damage detection and evaluation ii. In *Modal analysis and testing*, pages 345–378. Springer. 61

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, 15(1):3133–3181. 38

García, S., Luengo, J., and Herrera, F. (2015). *Data preprocessing in data mining*, volume 72. Springer. 24, 25, 31

Garg, A. and Tai, K. (2012). Comparison of regression analysis, artificial neural network and genetic programming in handling the multicollinearity problem. In *2012 Proceedings of International Conference on Modelling, Identification and Control*, pages 353–358. IEEE. 22

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media. 29

Goodwin, P. and Wright, G. (2014). *Decision analysis for management judgment.* John Wiley & Sons. 13

Goossens, A. J. and Basten, R. J. (2015). Exploring maintenance policy selection using the analytic hierarchy process; an application for naval ships. *Reliability Engineering & System Safety*, 142:31–41. 13

Gouriveau, R., Medjaher, K., and Zerhouni, N. (2016). *From prognostics and health systems management to predictive maintenance 1: Monitoring and prognostics.* John Wiley & Sons. 12

Gupta, G. and Mishra, R. (2018). Identification of critical components using anp for implementation of reliability centered maintenance. *Procedia CIRP*, 69:905–909. 12, 14

Haarman, M., Mulders, M., and Vassiliadis, C. (2017). Predictive maintenance 4.0-predict the unpredictable. *PwC and Mainnovation.* 4

Hair, J. F. (2009). Multivariate data analysis. 25

Hawkins, D. M. (1980). *Identification of outliers*, volume 11. Springer. vii, 33

Isermann, R. and Ballé, P. (1996). Trends in the application of model based fault detection and diagnosis of technical processes. *IFAC Proceedings Volumes*, 29(1):6325–6336. 53

Jardine, A. K., Lin, D., and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510. v, vi, 1, 3, 5, 7, 19, 53, 54, 55, 58, 59

Jinka, P. and Schwartz, B. (2016). *Anomaly Detection for Monitoring.* O'Reilly Media, Incorporated. 33

Karim, R., Westerberg, J., Galar, D., and Kumar, U. (2016). Maintenance analytics–the new know in maintenance. *IFAC-PapersOnLine*, 49(28):214–219. 7

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.* 56

Korbicz, J., Koscielny, J. M., Kowalczuk, Z., and Cholewa, W. (2012). *Fault diagnosis: models, artificial intelligence, applications.* Springer Science & Business Media. 55

LaRiviere, J., McAfee, P., Rao, J., Narayanan, V. K., and Sun, W. (2016). Where predictive analytics is having the biggest impact. *Harvard business review*, 25. vi, 1

Lee, J., Liao, L., Lapira, E., Ni, J., and Li, L. (2009). Informatics platform for designing and deploying e-manufacturing systems. In *Collaborative Design and Planning for Digital Manufacturing*, pages 1–35. Springer. 13, 14

Li, X., Wang, Y.-Y., and Acero, A. (2008). Learning query intent from regularized click graphs. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 339–346. 49

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE. 36, 50

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1):1–39. vii, 35

Liu, H. and Motoda, H. (2002). On issues of instance selection. *Data Mining and Knowledge Discovery*, 6(2):115. 31

Liu, H.-C., Chen, X.-Q., Duan, C.-Y., and Wang, Y.-M. (2019). Failure mode and effect analysis using multi-criteria decision making methods: A systematic literature review. *Computers & Industrial Engineering*, 135:881–897. 12

Müller, A. C. and Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists.* " O'Reilly Media, Inc.". 29, 30, 38, 41, 48, 49, 50, 55, 56

Olson, D. L. and Lauhoff, G. (2019). Descriptive data mining. In *Descriptive Data Mining*, pages 129–130. Springer. 20

Osborne, J. W. and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(1):6. 25

Park, C. S. (2002). *Contemporary engineering economics*, volume 4. Prentice Hall Upper Saddle River, NJ. 2

Paul, R. K. (2006). Multicollinearity: Causes, effects and remedies. *IASRI, New Delhi*, 1(1):58–65. 22

Paulheim, H. and Meusel, R. (2015). A decomposition of the outlier detection problem into a set of supervised learning problems. *Machine Learning*, 100(2):509–531. 37, 38

Peng, H. (2016). Condition-based maintenance: Capital goods industry. In *Analyzing Risk through Probabilistic Modeling in Operations Research*, pages 292–320. IGI Global. v, 1

Pham, D. and Pham, P. (1999). Artificial intelligence in engineering. *International Journal of Machine Tools and Manufacture*, 39(6):937–949. 55

Piramuthu, S. (2008). Input data for decision trees. *Expert Systems with applications*, 34(2):1220–1226. 22

Preiss, B. R. (1999). *Data structures and algorithms.* John Wiley & Sons, Inc. 76

Provost, F. and Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking.* " O'Reilly Media, Inc.". 19, 30, 36, 37

Randall, R. B. (2021). *Vibration-based condition monitoring: industrial, automotive and aerospace applications.* John Wiley & Sons. 61

Ray, S. (2019). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 35–39. IEEE. 22

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1):83–98. 13

Sasaki, Y. et al. (2007). The truth of the f-measure. 2007. *URL: https://www. cs. odu. edu/˜ mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07. pdf [accessed 2021-05-26].* 49

Siddique, A., Yadava, G., and Singh, B. (2003). Applications of artificial intelligence techniques for induction machine stator fault diagnostics. In *4th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, 2003. SDEMPED 2003.*, pages 29–34. IEEE. 55

Tiddens, W., Braaksma, J., and Tinga, T. (2020). Exploring predictive maintenance applications in industry. *Journal of Quality in Maintenance Engineering.* 7

Tiddens, W. W. (2018). Setting sail towards predictive maintenance: Developing tools to conquer difficulties in the implementation of maintenance analytics. vi, xi, 1, 3

Tiddens, W. W., Braaksma, A. J. J., and Tinga, T. (2018). Selecting suitable candidates for predictive maintenance. *Int. J. Prognostics Health Manag*, 9(1):020. vi, xi, 2, 13, 14, 17, 58, 62

Tinga, T. (2010). Application of physical failure models to enable usage and load based maintenance. *Reliability Engineering & System Safety*, 95(10):1061–1075. 1

White, G. and Ostwald, P. (1976). Life cycle costing. *Management Accounting (US)*, pages 39–42. 2

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pages 29–39. Springer-Verlag London, UK. 5

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2005). Practical machine learning tools and techniques. *Morgan Kaufmann*, page 578. 77

Zheng, A. and Casari, A. (2018). *Feature engineering for machine learning: principles and techniques for data scientists.* " O'Reilly Media, Inc.". 27

Zivot, E. and Wang, J. (2003). Rolling analysis of time series. In *Modeling Financial Time Series with S-Plus®*, pages 299–346. Springer. 27

Appendix A

Failure mode and effect analysis

Due to confidentiality this content is removed

Appendix B

Data Exploration

Customer-specific environment

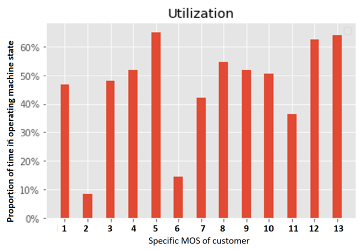


Figure B.1: Machine Y utilization

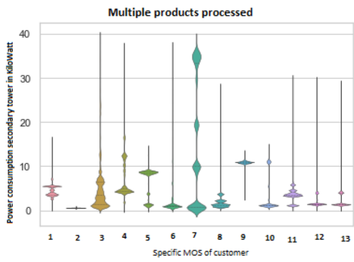


Figure B.2: Product diversity

Figure B.3: Amount of cleaning

Main air distribution fan

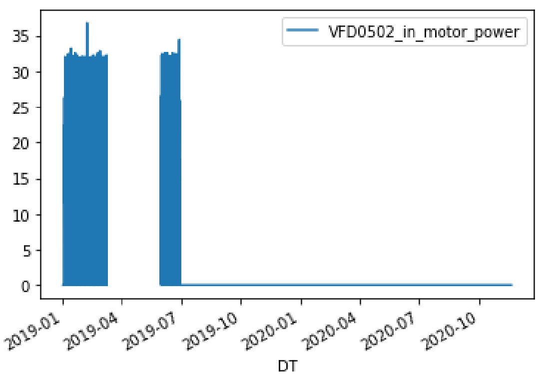


Figure B.4: Power fan Customer 4

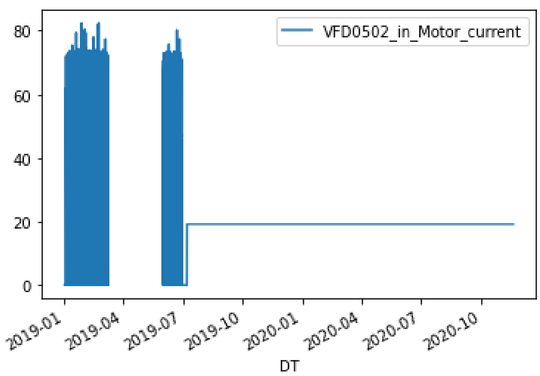


Figure B.5: Current fan Customer 4

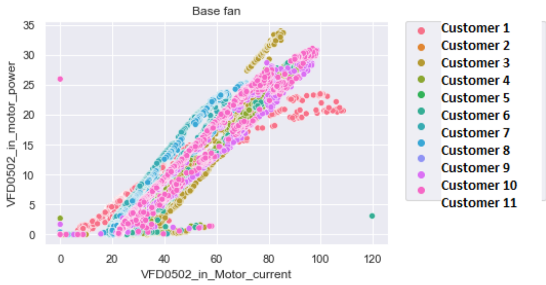


Figure B.6: scatter base fan

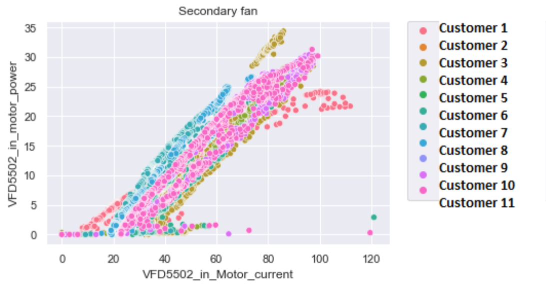


Figure B.7: scatter secondary fan

Conveyor belt drives

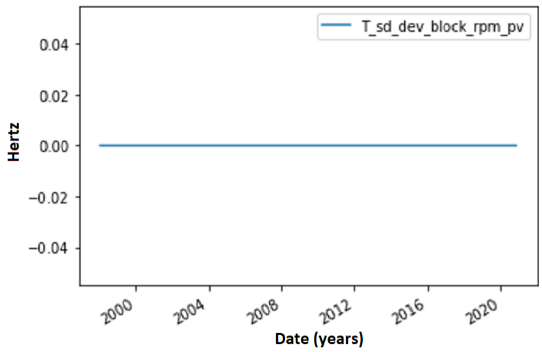


Figure B.8: Wrongly monitored speed in hertz Customer 1

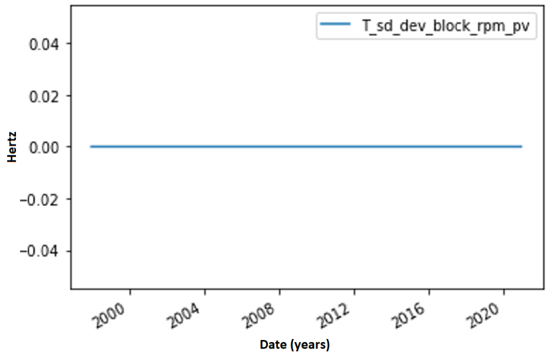


Figure B.9: Wrongly monitored speed in hertz Customer 9

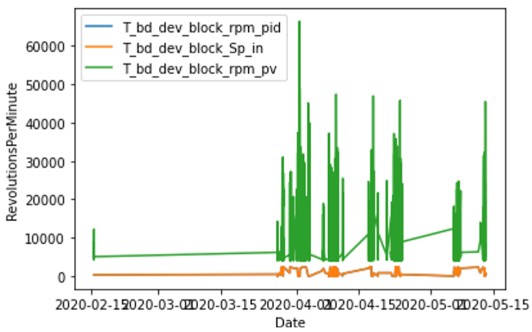


Figure B.10: Structurally too high speed Customer 7

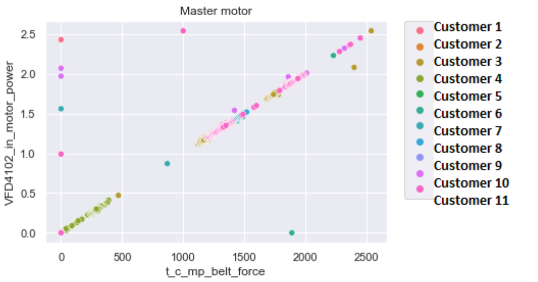


Figure B.11: scatter master motor

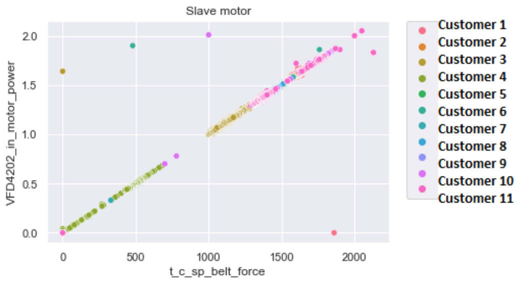


Figure B.12: scatter slave motor

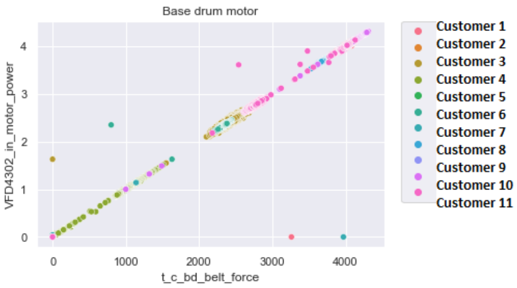


Figure B.13: scatter base drum

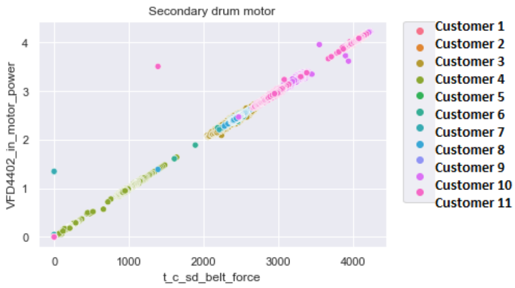


Figure B.14: scatter secondary drum

Appendix C

Priority components important parameters analysis

Fan: important parameters utilizing correlation analysis

Parameter:	R_active_recipe _airspeed_BT_value	VFD0502_in_ motor_speed_hz	VFD0502_in_ Motor_current	mt0708_ dew_value	R_active_recipe_ airspeed_ST_value
VFD0502_in_ motor_power	0.98	0.98	0.95	-0.91	0.75
Parameter:	VFD0502_in_ Motor_current	R_active_recipe _airspeed_BT_value	VFD0502_in_ motor_speed_hz	VFD0502_in_ motor_speed_hz	VFD0502_in_ Motor_current
VFD5502_in_ motor_power	0.98	0.97	0.97	0.65	0.65

Table C.1: Select 5 highest correlated features

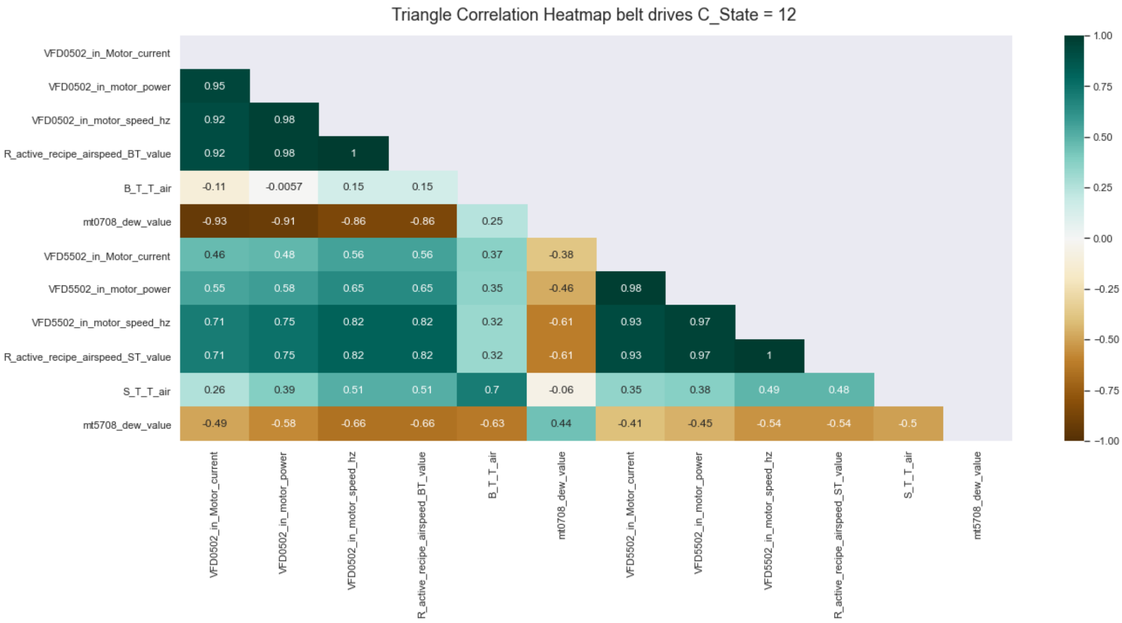


Figure C.1: Correlation heatmap fan

Fan: Extra features created

Attribute	Description	Type
VFD0502_in_motor_power_rol_win_std_5_min	Rolling window calculations for base drum motor where the standard deviation of the power is calculated based on the instance's 5-minute historical data	Float
VFD5502_in_motor_power_rol_win_std_5_min	Rolling window calculations for secondary drum motor where the standard deviation of the power is calculated based on the instance's 5-minute historical data	Float
VFD0502_in_motor_power_rol_win_std_30_min	Rolling window calculations for base drum motor where the standard deviation of the power is calculated based on the instance's 30-minute historical data	Float
VFD5502_in_motor_power_rol_win_std_30_min	Rolling window calculations for secondary drum motor where the standard deviation of the power is calculated based on the instance's 30-minute historical data	Float
VFD0502_in_motor_power_rol_win_min-max_5_min	Rolling window calculations for base drum motor where the minimal power is subtracted from the maximal power based on the instance's 5-minute historical data	Float
VFD5502_in_motor_power_rol_win_min-max_5_min	Rolling window calculations for secondary drum motor where the minimal power is subtracted from the maximal power based on the instance's 5-minute historical data	Float
VFD0502_in_motor_power_rol_win_min-max_30_min	Rolling window calculations for base drum motor where the minimal power is subtracted from the maximal power based on the instance's 30-minute historical data	Float
VFD5502_in_motor_power_rol_win_min-max_30_min	Rolling window calculations for secondary drum motor where the minimal power is subtracted from the maximal power based on the instance's 30-minute historical data	Float

Table C.2: Created features fan

Belt drive motors: important parameters utilizing correlation analysis

parameter:	t_c_mp_belt_force	t_c_sp_belt_force	VFD4202_in_motor_power	VFD4302_in_motor_power	t_c_bd_belt_force
VFD4102_in_motor_power	1	0.95	0.95	0.93	0.93
parameter:	t_c_sp_belt_force	VFD4102_in_motor_power	VFD4202_in_motor_power	t_c_bd_belt_force	VFD4302_in_motor_power
VFD4202_in_motor_power	1	0.95	0.95	0.93	0.93
parameter:	t_c_bd_belt_force	t_c_sd_belt_force	VFD4402_in_motor_power	t_c_sp_belt_force	t_c_mp_belt_force
VFD4302_in_motor_power	1	0.99	0.99	0.93	0.93
parameter:	t_c_sd_belt_force	t_c_bd_belt_force	VFD4302_in_motor_power	t_c_sp_belt_force	t_c_mp_belt_force
VFD4402_in_motor_power	1	0.99	0.99	0.91	0.91

Table C.3: Select 5 highest correlated features

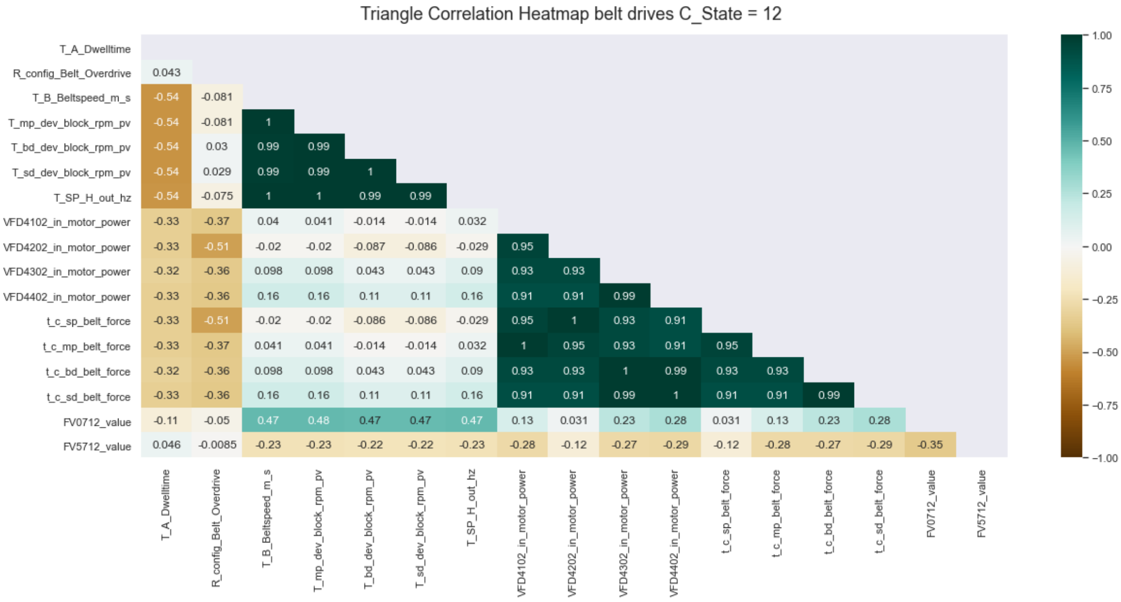


Figure C.2: Correlation heatmap belt drives

Belt drive motors: Extra features created

	Attribute	**Description**	**Type**
1	C_12_time	Feature that counts what time Machine Y is in an operating state	Seconds
2	C_12_5_VFD4102_power_rol_win_std	Rolling window calculations for master motor where the standard deviation of the power is calculated based on the instance's 5-minute historical data	Float
3	C_12_5_VFD4202_power_rol_win_std	Rolling window calculations for slave motor where the standard deviation of the power is calculated based on the instance's 5-minute historical data	Float
4	C_12_5_VFD4302_power_rol_win_std	Rolling window calculations for base drum motor where the standard deviation of the power is calculated based on the instance's 5-minute historical data	Float
5	C_12_5_VFD4402_power_rol_win_std	Rolling window calculations for secondary drum motor where the standard deviation of the power is calculated based on the instance's 5-minute historical data	Float
6	C_12_5_VFD4102_power_rol_win_std_30_min	Rolling window calculations for master motor where the standard deviation of the power is calculated based on the instance's 30-minute historical data	Float
7	C_12_5_VFD4202_power_rol_win_std_30_min	Rolling window calculations for slave motor where the standard deviation of the power is calculated based on the instance's 30-minute historical data	Float
8	C_12_5_VFD4302_power_rol_win_std_30_min	Rolling window calculations for base drum motor where the standard deviation of the power is calculated based on the instance's 30-minute historical data	Float
9	C_12_5_VFD4402_power_rol_win_std_30_min	Rolling window calculations for secondary drum motor where the standard deviation of the power is calculated based on the instance's 30-minute historical data	Float
10	C_12_5_VFD4102_power_rol_win_min-max	Rolling window calculations for master motor where the minimal power is subtracted from the maximal power based on the instance's 5-minute historical data	Float
11	C_12_5_VFD4202_power_rol_win_min-max	Rolling window calculations for slave motor where the minimal power is subtracted from the maximal power based on the instance's 5-minute historical data	Float
12	C_12_5_VFD4302_power_rol_win_min-max	Rolling window calculations for base drum motor where the minimal power is subtracted from the maximal power based on the instance's 5-minute historical data	Float
13	C_12_5_VFD4402_power_rol_win_min-max	Rolling window calculations for secondary drum motor where the minimal power is subtracted from the maximal power based on the instance's 5-minute historical data	Float
14	C_12_5_VFD4102_power_rol_win_min-max_30_min	Rolling window calculations for master motor where the minimal power is subtracted from the maximal power based on the instance's 30-minute historical data	Float
15	C_12_5_VFD4202_power_rol_win_min-max_30_min	Rolling window calculations for slave motor where the minimal power is subtracted from the maximal power based on the instance's 30-minute historical data	Float
16	C_12_5_VFD4302_power_rol_win_min-max_30_min	Rolling window calculations for base drum motor where the minimal power is subtracted from the maximal power based on the instance's 30-minute historical data	Float
17	C_12_5_VFD4402_power_rol_win_min-max_30_min	Rolling window calculations for secondary drum motor where the minimal power is subtracted from the maximal power based on the instance's 30-minute historical data	Float
18	bd&sd_in_motor_power	(power base drum motor)/(power secondary drum motor) As the conveyor belt of the base drum is directly connected to the secondary drum, the relative power consumption should be relatively similar in normal situations	Float
19	mp&sd_in_motor_power	(power master motor)/(power secondary drum motor) As the conveyor belt of the master motor is directly connected to the secondary drum, the relative power consumption should be relatively similar in normal situations	Float
20	mp&sp_in_motor_power	(power master motor)/(power slave motor) As the conveyor belt of the master motor is directly connected to the slave motor in the return section, the relative power consumption should be relatively similar in normal situations	Float
21	sp&bd_in_motor_power	(power slave motor)/(power base drum motor) As the conveyor belt of the slave motor is directly connected to the base drum motor in the return section, the relative power consumption should be relatively similar in normal situations	Float

Table C.4: Created features belt drive motors

Appendix D

Anomaly detection models

D.1 Isolation forest

An IF is an ensemble combination of a set of isolation trees. In an isolation tree, the data is recursively partitioned with axis-parallel cuts at randomly chosen partition points in randomly selected attributes, so as to isolate the instances into nodes with fewer and fewer instances until the points are isolated into singleton nodes containing one instance. In such cases, the tree branches containing anomalies are noticeably less deep, because these data points are located in sparse regions. Therefore, the distance of the leaf to the root is used as the anomaly score. The final combination step is performed by averaging the path lengths of the data points in the different trees of the IF.

We define path length and anomaly score as follows. Path length $h(x)$ of a point x is measured by the number of edges x traverses an isolation tree from the root node until the traversal is terminated at an external node. On the other hand, the anomaly score can be defined based on (i) the average path length of all instances, and (ii) the expected path length of point x. So, given a data set of n instances, Section 10.3.3 of Preiss (1999) gives the average path length as Equation D.1 where $H(n)$ is the harmonic number and it can be estimated by $ln(n) + 0.5772156649$ (Euler's constant).

$$c(n) = 2H(n-1) - (2(n-1)/n) \tag{D.1}$$

As $c(n)$ is the average of $h(x)$ given n, we use it to normalise $h(x)$. The anomaly score s of an instance x is defined as Equation D.2 where $E(h(x))$ is the average of $h(x)$ from a collection of isolation trees.

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{D.2}$$

Using the anomaly score we can make the following assessment: (i) if instances return s very close to 1, then they are definitely anomalies. (ii) If instances have s much smaller than 0.5, then they can safely be regarded as normal instances. (iii) If all the instances return s = 0.5, then the entire sample does not have any distinct anomaly.

D.2 Local outlier factor

Throughout the formal explanation of LOF, we use o, p, q as objects in a dataset D. An object is equal to a multidimensional sensor data record with a particular timestamp and customer. We use the notation $d(p, q)$ to denote the distance between objects p and q. For a set of objects, we use C. To simplify our notation, we use $d(p, C)$ to denote the minimum distance between p and object q in C, i.e. $d(p, C) = min\{d(p, q) \mid q \in C\}$. The LOF can be calculated according to the following five steps:

- **Step 1**: (calculate k-distance of a data point p)
  According to Aggarwal et al. (2001) the choice of the distance metric is not obvious in most high dimensional applications and the notion for the calculation of similarity is very heuristical. There is very little literature on providing guidance for choosing the correct distance measure which results in the most meaningful notion of proximity between two records. Many high dimensional indexing structures and algorithms, especially in instance-based learners (Witten et al., 2005), use the euclidean distance metric and will be applied in this project as well. So, for given two data objects (p, o), the distance between them is measured with the Euclidean distance in n-dimensional space. So, p_i and o_i correspond to a continuous-valued attribute (sensor data point i) of a particular time at a customer.

$$d(p, o) = \sqrt{\sum_{i=1}^{n} (p_i - o_i)^2} \tag{D.3}$$

  For any positive integer k, the k-distance of object p, denoted as $k - distance(p)$, is defined as the distance $d(p, o)$ between p and an object $o \in D$ such that:

  - for at least k objects $o' \in D \backslash \{p\}$ it holds that $d(p, o') \leq d(p, o)$, and
  - for at most k-1 objects $o' \in D \backslash \{p\}$ it holds that $d(p, o') < d(p, o)$.

- **Step 2**: (Calculate k-distance neighborhood of an object p)
  Given the k-distance of p, the *k-distance neighborhood of p* contains every object whose distance from p is not greater than the k-distance (Equation D.4). These objects q are called the k-nearest neighbors of p.

$$N_{k-distance(p)}(p) = \{q \in D \backslash \{p\} \mid d(p, q) \leq k - distance(p)\} \tag{D.4}$$

- **Step 3**: (Calculate reachability distance of object p with respect to object o)
  Let k be a natural number. The *reachability distance* of object p with respect to object o is defined as:

$$reach - dist_k(p, o) = max\{k - distance(o), d(p, o)\} \tag{D.5}$$

- **Step 4**: (Calculate local reachability density of object p)
  To detect density-based outliers, it is necessary to compare the densities of different sets of objects, which means that we have to determine the density of sets of objects dynamically. Therefore, we use $MinPts$, which is the number of nearest neighbors used in defining the local neighborhood of the object, as the only parameter. As a measure of the volume to determine the density in the neighborhood of an object p, we use the values $reach - dist_{MinPts}(p, o)$, for $o \in N_{MinPts}(p)$. The *local reachability density* of p is defined as:

$$lrd_{MinPts}(p) = 1 / \left[\frac{\sum_{o \in N_{MinPts}(p)} reach - dist_{MinPts}(p, o)}{\mid N_{MinPts}(p) \mid} \right] \tag{D.6}$$

- **Step 5**: (The local outlier factor (LOF) of p) The *(local) outlier factor* of p is defined as:

$$LOF_{MinPts}(p) = \frac{\sum_{o \in N_{MinPts}(p)} \frac{lrd_{MinPts}(o)}{lrd_{MinPts}(p)}}{\mid N_{MinPts}(p) \mid} \tag{D.7}$$

For objects deep inside a cluster, the LOF value is approximately 1. For other objects, we give tight lower and upper bounds on the LOF value.

Appendix E

Exploring anomalies

Fan: regression modeling

Regression modeling 1 fan

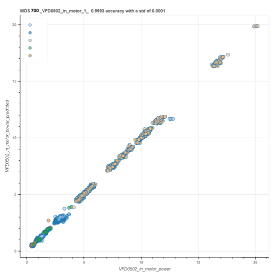


Figure E.1: BT Machine type C

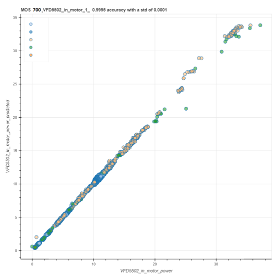


Figure E.2: ST Machine type C

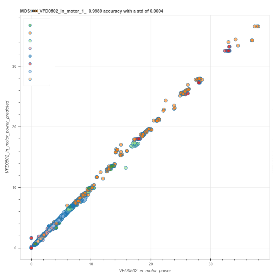


Figure E.3: BT Machine type A,B

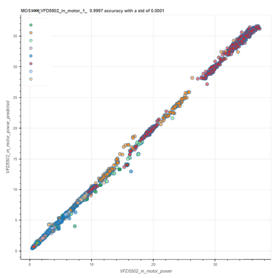


Figure E.4: ST Machine type A,B

Regression modeling 2 fan

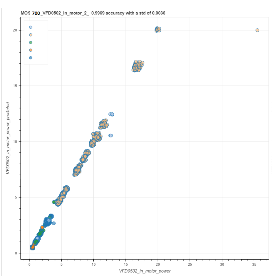


Figure E.5: BT Machine type C

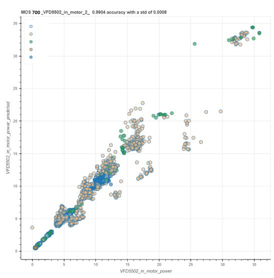


Figure E.6: ST Machine type C

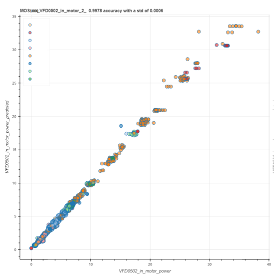


Figure E.7: BT Machine type A,B

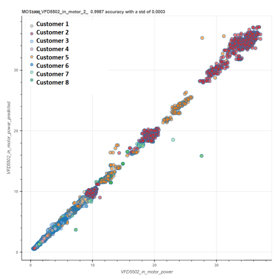


Figure E.8: ST Machine type A,B

Regression modeling 3 fan

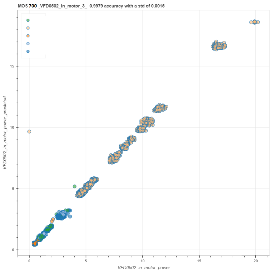


Figure E.9: BT Machine type C

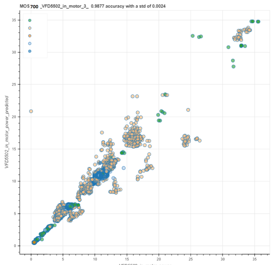


Figure E.10: ST Machine type C

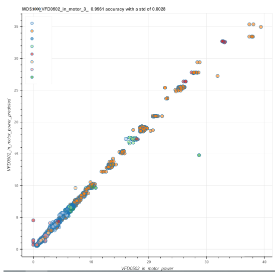


Figure E.11: BT Machine type A,B

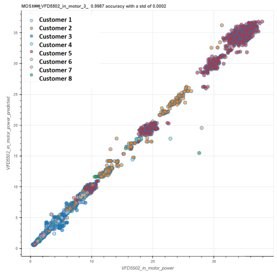


Figure E.12: ST Machine type A,B

Fan: distinguished anomaly types

Fan off while being in operating state

Diagnosis: Fan motor frequency converter failure

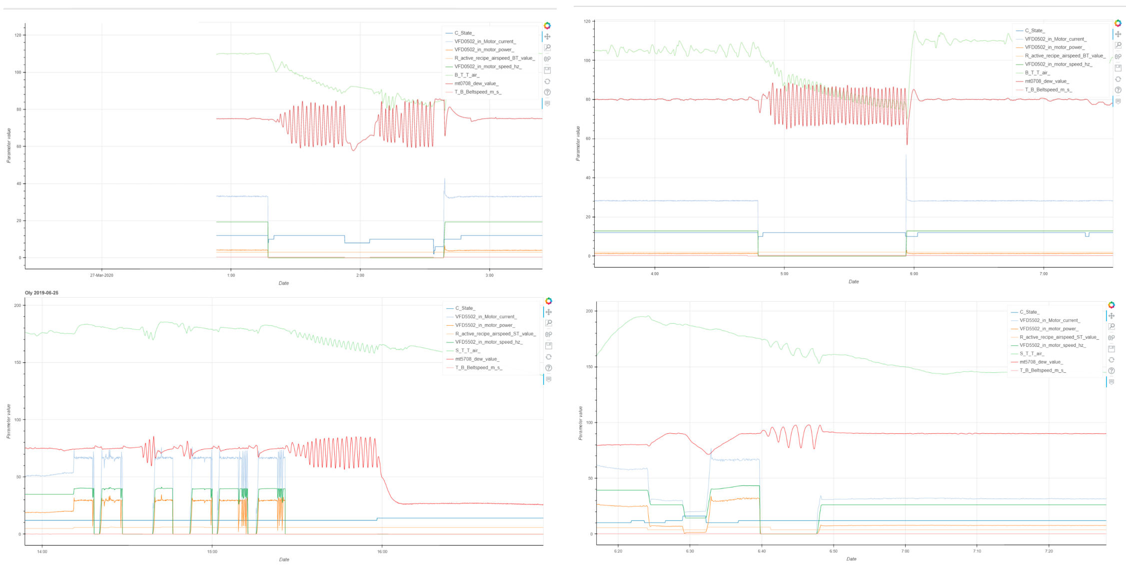


Figure E.13: Fan off while being in operating state

Fluctuating behavior of the power

Diagnosis: Undefined

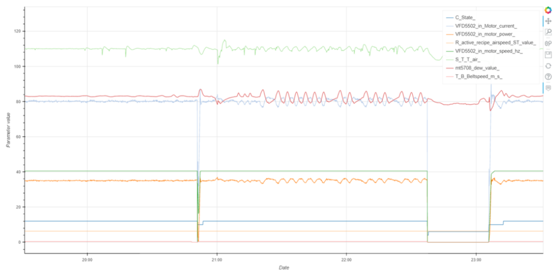


Figure E.14: Inexplicable fluctuating behavior of the power

Diagnosis: low temperature or dewpoint and high airspeed result in fluctuating power fan

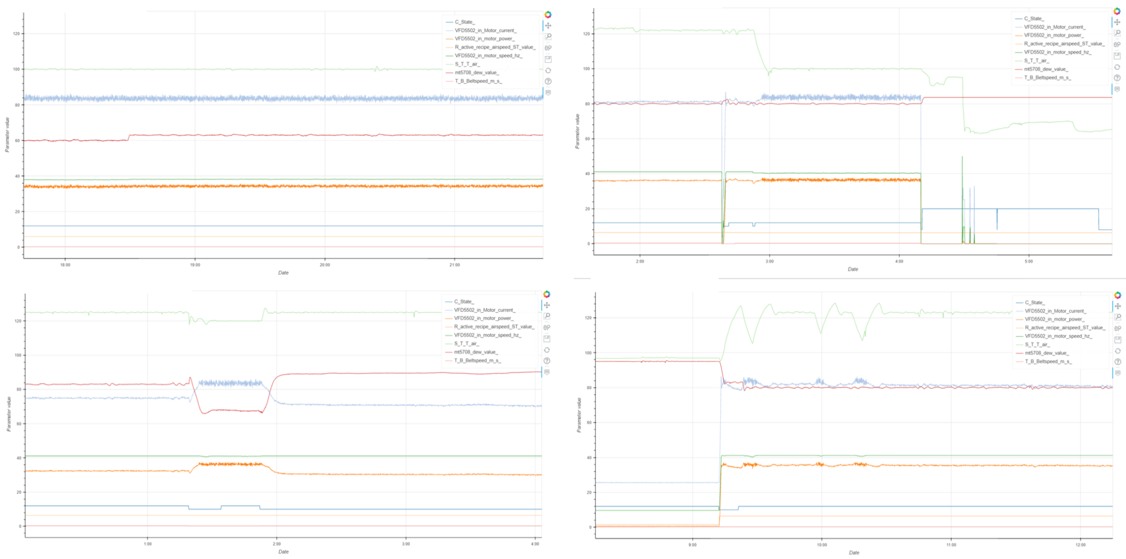


Figure E.15: Low temperature or dewpoint and high airspeed result in fluctuating power fan

Diagnosis: Changes in temperature, dew value and airspeed cause fluctuating power

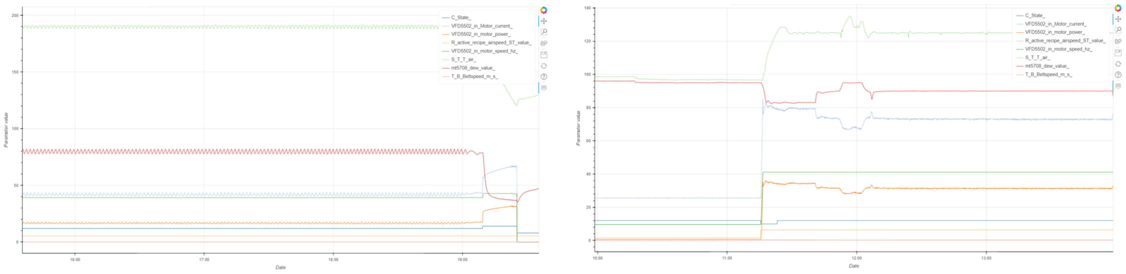


Figure E.16: Changes in temperature, dew value and airspeed cause fluctuating power

Diagnosis: Broken dewpoint sensor causing fluctuations power

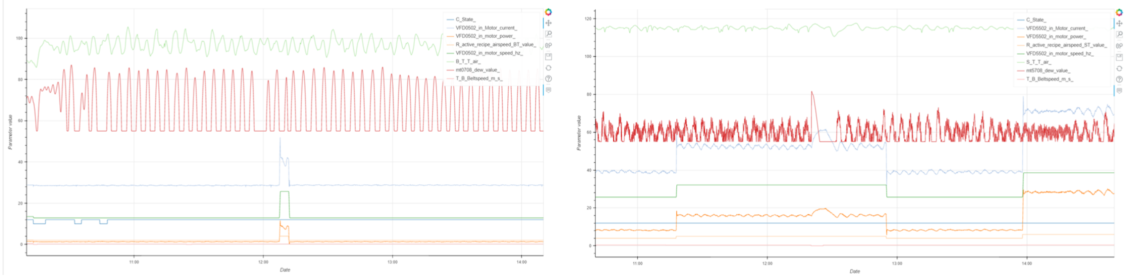


Figure E.17: Broken dewpoint sensor causing fluctuations power

Belt drives: regression modeling

Regression modeling 1 belt drives

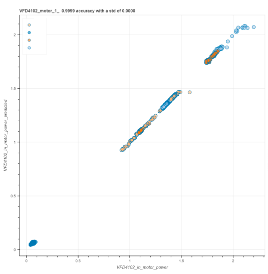


Figure E.18: Machine type C master

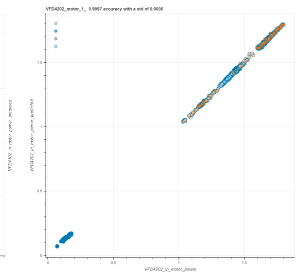


Figure E.19: Machine type C slave

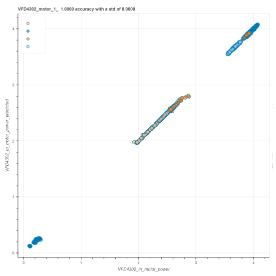


Figure E.20: Machine type C drum 1

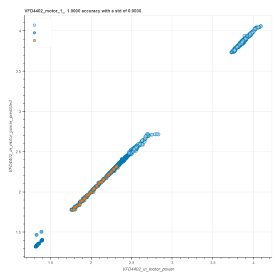


Figure E.21: Machine type C drum 2

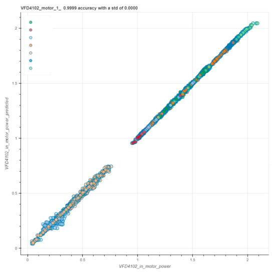


Figure E.22: Machine type A,B master

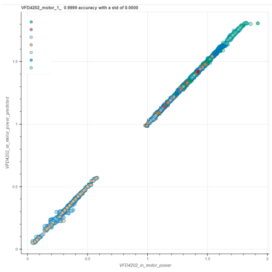


Figure E.23: Machine type A,B slave

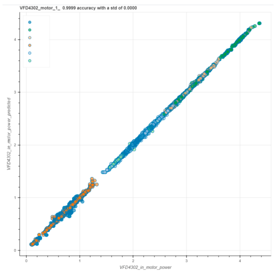


Figure E.24: Machine type A,B drum 1

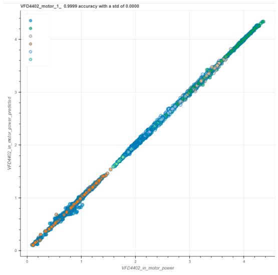


Figure E.25: Machine type A,B drum 2

Regression modeling 2 belt drives

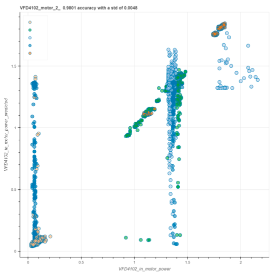


Figure E.26: Machine type C master

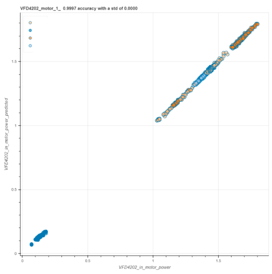


Figure E.27: Machine type C slave

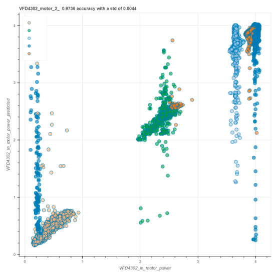


Figure E.28: Machine type C drum 1

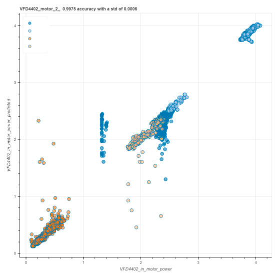


Figure E.29: Machine type C drum 2

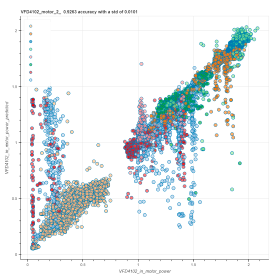


Figure E.30: Machine type A,B master

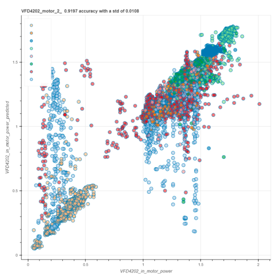


Figure E.31: Machine type A,B slave

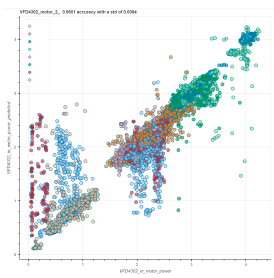


Figure E.32: Machine type A,B drum 1

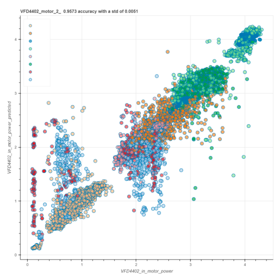


Figure E.33: Machine type A,B drum 2

Regression modeling 3 belt drives

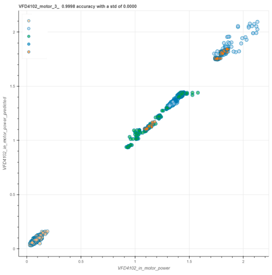


Figure E.34: Machine type C master

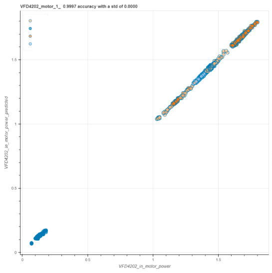


Figure E.35: Machine type C slave

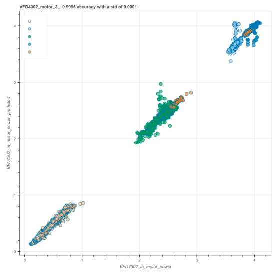


Figure E.36: Machine type C drum 1

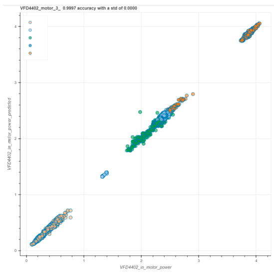


Figure E.37: Machine type C drum 2

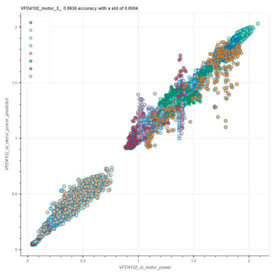


Figure E.38: Machine type A,B master

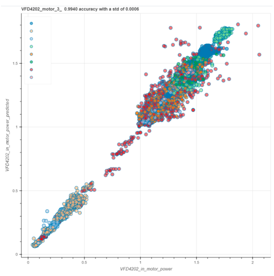


Figure E.39: Machine type A,B slave

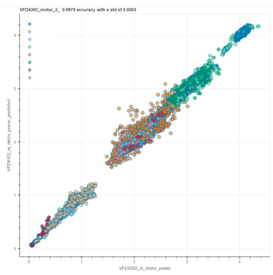


Figure E.40: Machine type A,B drum 1

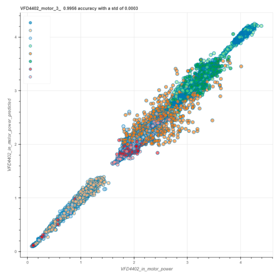


Figure E.41: Machine type A,B drum 2

Belt drives: distinguished anomaly types

Power increase master pickup

Diagnosis: Belt got stuck

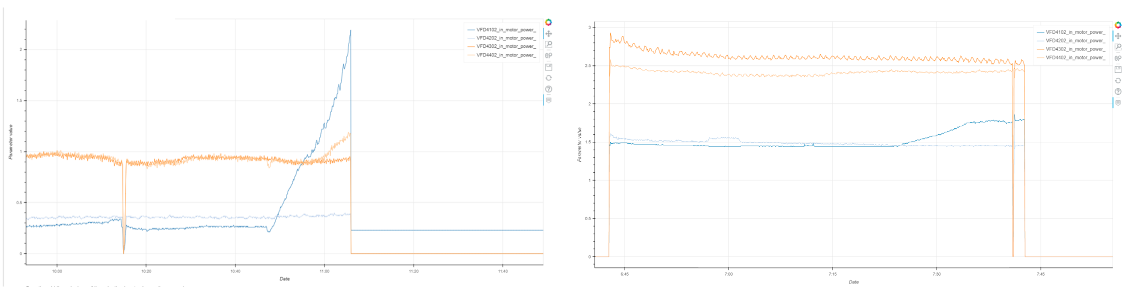


Figure E.42: Power increase master pickup

Slave pickup less power, base drum more power

Diagnosis: Belt tension mechanism stuck

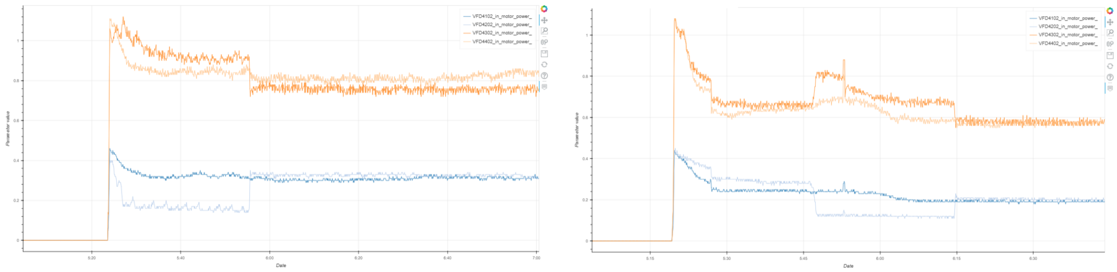


Figure E.43: Slave pickup less power, base drum more power

Master more power, secondary drum more power

Diagnosis: Product weight transition

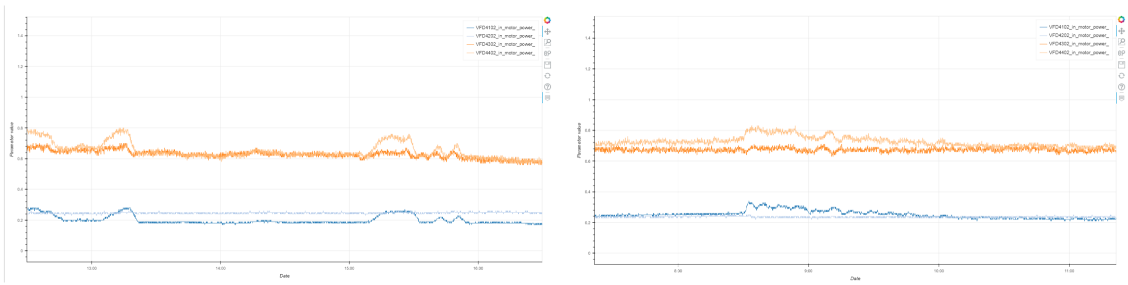


Figure E.44: Master more power, secondary drum more power

Base drum peaks

Diagnosis: 2 types of diagnosis

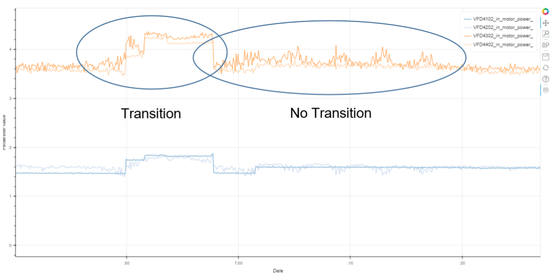


Figure E.45: Belt tension mechanism stuck

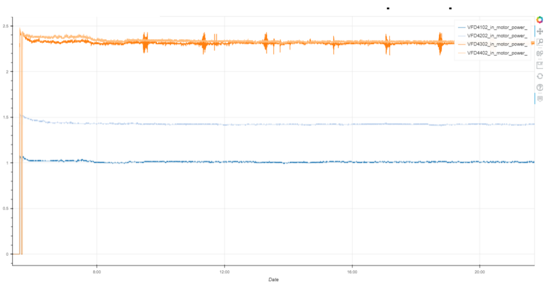


Figure E.46: Base drum stuck

Fluctuating powers

Diagnosis: 10 types of diagnosis

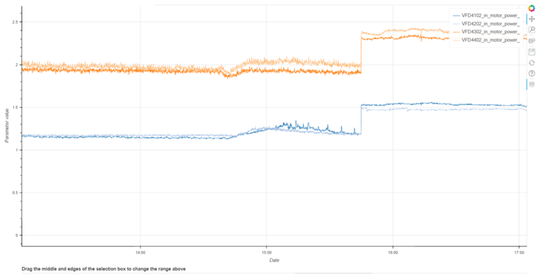


Figure E.47: Product weight transition (Not interesting)

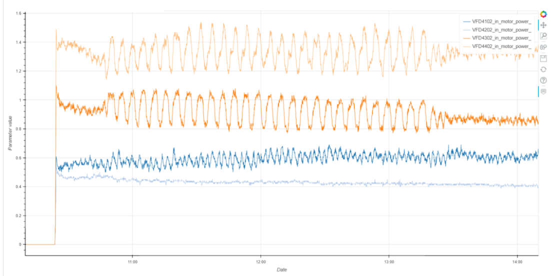


Figure E.48: Belt got stuck

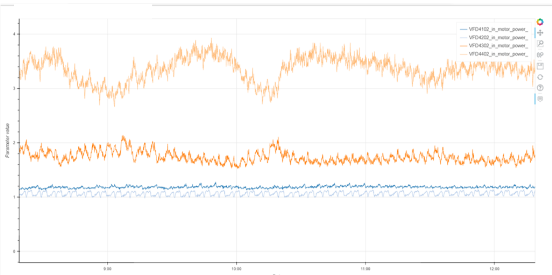


Figure E.49: drum pulls the other drum

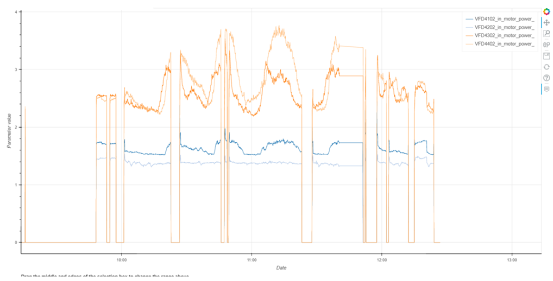


Figure E.50: 1 drum pulls the other drum

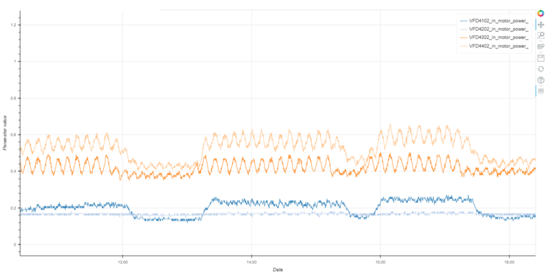


Figure E.51: Diagnosis: Undefined

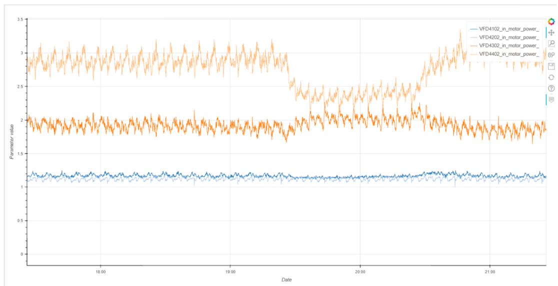


Figure E.52: 1 drum pulls the other drum

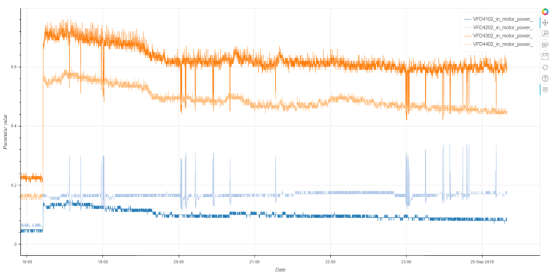


Figure E.53: Belt tension mechanism stuck

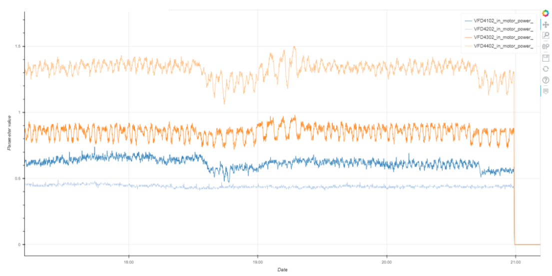


Figure E.54: Diagnosis: Undefined

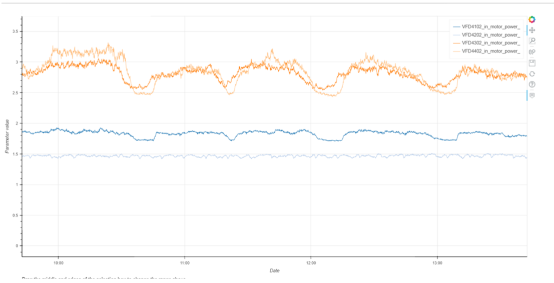


Figure E.55: Diagnosis: Product weight transition (Not interesting)

Figure E.56: Diagnosis: undefined

Sudden decrease power all motors

Diagnosis: 2 types of diagnosis

Figure E.57: Diagnosis: Belt tension mechanism stuck

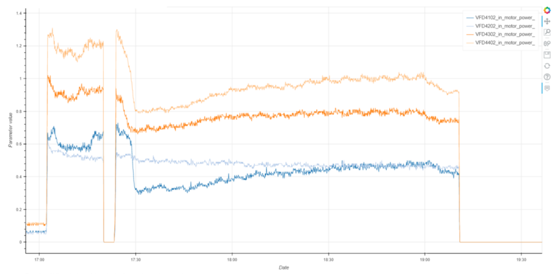


Figure E.58: Diagnosis: Belt lubricated

Fluctuating behavior slave pickup

Diagnosis: Belt in return section gets stuck

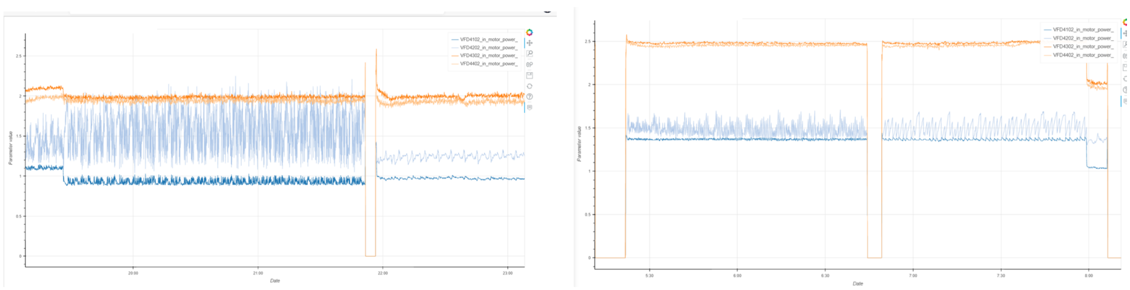


Figure E.59: Fluctuating behavior slave pickup

Fluctuating behavior master power consumption

Diagnosis: Undefined

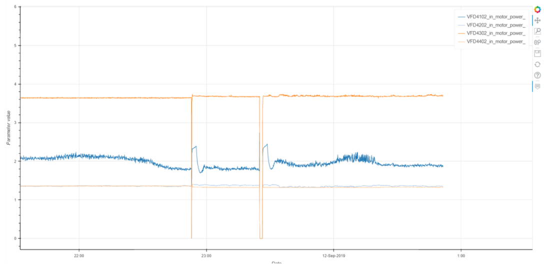


Figure E.60: Weird peak power consumption master motor

Appendix F

Event analysis

	Description
1	event;10;BT: Slave pick-up motor frequency converter failure (VFD4202);;
2	event;14;BT: Drum motor frequency converter failure (VFD4302);;
3	event;19;BT: Slave pick-up motor power overload (M4201);;
4	event;20;BT: OLD ALARM: Base: ZS4002: Transport belt tilted;;
5	event;21;BT: Drum motor power overload (M4301);;
6	event;110;ST: Master pick-up motor frequency converter failure (VFD4102);;
7	event;111;ST: Lost communication with Master pick-up frequency controler (VFD4102);;
8	event;112;ST: Master pick-up motor circuit breaker tripped (CB4103);;
9	event;113;ST: Master pick-up motor overheated (TS4105);;
10	event;114;ST: Drum motor frequency converter failure (VFD4402);;
11	event;117;ST: Drum motor overheated (TS4405);;
12	event;119;ST: Master pick-up motor overpower (M4101);;
13	event;120;ST: Belt tension detection activated (LS4001);;
14	event;121;ST: Transport Secondary drum overpower (M4401);;
15	event;205;Gen: One of the belt motors did not run;;
16	event;206;Gen: One of belt motors out of synchronization;;
17	event;208;Gen: Master pick-up motor encoder motor failure, Belt runs in open loop (ST4106);;
18	event;209;Gen: Belt stopped during heating or cooling. Warning, running belt is required!;;
19	event;211;Gen: Belt stop button at outfeed pressed;;
20	event;227;Belt stop button at infeed pressed;;

Table F.1: Interesting events conveyor belt drive motors

Appendix G

Grid search

G.1 IF grid search

Ground-truth dataset without uninteresting anomalies

	name	motion	n_estimators	max_samples	contamination	precision	recall	fscore
IF	1	#N/A	50	100	0.09	0.86	0.96	0.91
IF	1	#N/A	50	100	0.1	0.79	0.97	0.89
IF	1	#N/A	50	100	0.11	0.72	0.97	0.85
IF	1	#N/A	50	100	0.15	0.51	0.98	0.71
IF	1	#N/A	50	256	0.09	0.67	0.75	0.72
IF	1	#N/A	50	256	0.1	0.61	0.76	0.69
IF	1	#N/A	50	256	0.11	0.58	0.79	0.69
IF	1	#N/A	50	256	0.15	0.47	0.87	0.64
IF	1	#N/A	50	3000	0.09	0.89	0.99	0.94
IF	1	#N/A	50	3000	0.1	0.81	1.0	0.91
IF	1	#N/A	50	3000	0.11	0.73	1.0	0.87
IF	1	#N/A	50	3000	0.15	0.54	1.0	0.74
IF	1	#N/A	100	100	0.09	0.75	0.84	0.8
IF	1	#N/A	100	100	0.1	0.7	0.86	0.79
IF	1	#N/A	100	100	0.11	0.65	0.88	0.76
IF	1	#N/A	100	100	0.15	0.52	0.97	0.71
IF	1	#N/A	100	256	0.09	0.68	0.76	0.72
IF	1	#N/A	100	256	0.1	0.63	0.78	0.71
IF	1	#N/A	100	256	0.11	0.6	0.82	0.71
IF	1	#N/A	100	256	0.15	0.51	0.95	0.7
IF	1	#N/A	100	3000	0.09	0.88	0.98	0.94
IF	1	#N/A	100	3000	0.1	0.81	0.99	0.91
IF	1	#N/A	100	3000	0.11	0.73	1.0	0.87
IF	1	#N/A	100	3000	0.15	0.54	1.0	0.74
IF	1	#N/A	500	100	0.09	0.67	0.74	0.71
IF	1	#N/A	500	100	0.1	0.6	0.75	0.68
IF	1	#N/A	500	100	0.11	0.55	0.75	0.65
IF	1	#N/A	500	100	0.15	0.41	0.76	0.56
IF	1	#N/A	500	256	0.09	0.7	0.78	0.75
IF	1	#N/A	500	256	0.1	0.66	0.82	0.74
IF	1	#N/A	500	256	0.11	0.64	0.86	0.75
IF	1	#N/A	500	256	0.15	0.54	0.99	0.73
IF	1	#N/A	500	3000	0.09	0.84	0.94	0.89
IF	1	#N/A	500	3000	0.1	0.78	0.96	0.87
IF	1	#N/A	500	3000	0.11	0.72	0.98	0.86

IF	1	#N/A	500	3000	0.15	0.54	1.0	0.74
IF	4	std	50	100	0.09	1.0	0.91	0.94
IF	4	std	50	100	0.1	0.99	0.99	0.99
IF	4	std	50	100	0.11	0.91	1.0	0.96
IF	4	std	50	100	0.15	0.67	1.0	0.83
IF	4	std	50	256	0.09	1.0	0.9	0.94
IF	4	std	50	256	0.1	0.99	0.99	0.99
IF	4	std	50	256	0.11	0.9	0.99	0.95
IF	4	std	50	256	0.15	0.66	0.99	0.82
IF	4	std	50	3000	0.09	1.0	0.9	0.94
IF	4	std	50	3000	0.1	0.97	0.98	0.98
IF	4	std	50	3000	0.11	0.89	0.98	0.94
IF	4	std	50	3000	0.15	0.65	0.98	0.81
IF	4	std	100	100	0.09	1.0	0.9	0.94
IF	4	std	100	100	0.1	0.99	0.99	0.99
IF	4	std	100	100	0.11	0.9	0.99	0.95
IF	4	std	100	100	0.15	0.67	1.0	0.83
IF	4	std	100	256	0.09	1.0	0.9	0.94
IF	4	std	100	256	0.1	0.99	1.0	1.0
IF	4	std	100	256	0.11	0.9	1.0	0.96
IF	4	std	100	256	0.15	0.67	1.0	0.83
IF	4	std	100	3000	0.09	1.0	0.9	0.94
IF	4	std	100	3000	0.1	0.98	0.98	0.98
IF	4	std	100	3000	0.11	0.9	0.99	0.95
IF	4	std	100	3000	0.15	0.66	0.99	0.82
IF	4	std	500	100	0.09	1.0	0.9	0.94
IF	4	std	500	100	0.1	0.99	0.99	0.99
IF	4	std	500	100	0.11	0.91	1.0	0.96
IF	4	std	500	100	0.15	0.67	1.0	0.83
IF	4	std	500	256	0.09	1.0	0.9	0.94
IF	4	std	500	256	0.1	0.99	1.0	1.0
IF	4	std	500	256	0.11	0.91	1.0	0.96
IF	4	std	500	256	0.15	0.67	1.0	0.83
IF	4	std	500	3000	0.09	1.0	0.9	0.94
IF	4	std	500	3000	0.1	0.99	0.99	0.99
IF	4	std	500	3000	0.11	0.91	1.0	0.96
IF	4	std	500	3000	0.15	0.67	1.0	0.83
IF	4	min-max	50	100	0.09	1.0	0.9	0.94
IF	4	min-max	50	100	0.1	0.98	0.98	0.98
IF	4	min-max	50	100	0.11	0.89	0.98	0.94
IF	4	min-max	50	100	0.15	0.66	0.99	0.82
IF	4	min-max	50	256	0.09	1.0	0.9	0.94
IF	4	min-max	50	256	0.1	0.99	0.99	0.99
IF	4	min-max	50	256	0.11	0.9	1.0	0.96
IF	4	min-max	50	256	0.15	0.66	1.0	0.82
IF	4	min-max	50	3000	0.09	1.0	0.9	0.94
IF	4	min-max	50	3000	0.1	0.98	0.99	0.98
IF	4	min-max	50	3000	0.11	0.9	0.99	0.95
IF	4	min-max	50	3000	0.15	0.66	1.0	0.82
IF	4	min-max	100	100	0.09	1.0	0.9	0.94
IF	4	min-max	100	100	0.1	0.98	0.98	0.98
IF	4	min-max	100	100	0.11	0.9	0.99	0.95
IF	4	min-max	100	100	0.15	0.66	0.99	0.82
IF	4	min-max	100	256	0.09	1.0	0.9	0.94
IF	4	min-max	100	256	0.1	0.99	0.99	0.99
IF	4	min-max	100	256	0.11	0.9	1.0	0.96
IF	4	min-max	100	256	0.15	0.67	1.0	0.83

IF	4	min-max	100	3000	0.09	1.0	0.9	0.94
IF	4	min-max	100	3000	0.1	0.98	0.99	0.98
IF	4	min-max	100	3000	0.11	0.9	0.99	0.95
IF	4	min-max	100	3000	0.15	0.66	0.99	0.82
IF	4	min-max	500	100	0.09	1.0	0.9	0.94
IF	4	min-max	500	100	0.1	0.99	0.99	0.99
IF	4	min-max	500	100	0.11	0.9	0.99	0.95
IF	4	min-max	500	100	0.15	0.66	1.0	0.82
IF	4	min-max	500	256	0.09	1.0	0.9	0.94
IF	4	min-max	500	256	0.1	0.99	0.99	0.99
IF	4	min-max	500	256	0.11	0.9	1.0	0.96
IF	4	min-max	500	256	0.15	0.66	1.0	0.83
IF	4	min-max	500	3000	0.09	1.0	0.9	0.94
IF	4	min-max	500	3000	0.1	0.99	0.99	0.99
IF	4	min-max	500	3000	0.11	0.91	1.0	0.96
IF	4	min-max	500	3000	0.15	0.67	1.0	0.83
IF	5	std	50	100	0.09	0.96	0.87	0.91
IF	5	std	50	100	0.1	0.94	0.94	0.94
IF	5	std	50	100	0.11	0.87	0.96	0.92
IF	5	std	50	100	0.15	0.66	0.99	0.82
IF	5	std	50	256	0.09	1.0	0.9	0.94
IF	5	std	50	256	0.1	1.0	1.0	1.0
IF	5	std	50	256	0.11	0.91	1.0	0.96
IF	5	std	50	256	0.15	0.67	1.0	0.83
IF	5	std	50	3000	0.09	1.0	0.9	0.94
IF	5	std	50	3000	0.1	0.99	0.99	0.99
IF	5	std	50	3000	0.11	0.9	0.99	0.95
IF	5	std	50	3000	0.15	0.66	0.99	0.82
IF	5	std	100	100	0.09	1.0	0.9	0.94
IF	5	std	100	100	0.1	0.95	0.96	0.95
IF	5	std	100	100	0.11	0.89	0.98	0.94
IF	5	std	100	100	0.15	0.66	0.99	0.82
IF	5	std	100	256	0.09	1.0	0.9	0.94
IF	5	std	100	256	0.1	0.99	0.99	0.99
IF	5	std	100	256	0.11	0.9	0.99	0.95
IF	5	std	100	256	0.15	0.67	1.0	0.83
IF	5	std	100	3000	0.09	1.0	0.9	0.94
IF	5	std	100	3000	0.1	0.99	0.99	0.99
IF	5	std	100	3000	0.11	0.9	0.99	0.95
IF	5	std	100	3000	0.15	0.67	1.0	0.83
IF	5	std	500	100	0.09	1.0	0.9	0.94
IF	5	std	500	100	0.1	0.99	0.99	0.99
IF	5	std	500	100	0.11	0.9	0.99	0.95
IF	5	std	500	100	0.15	0.66	1.0	0.83
IF	5	std	500	256	0.09	1.0	0.9	0.94
IF	5	std	500	256	0.1	0.99	0.99	0.99
IF	5	std	500	256	0.11	0.91	1.0	0.96
IF	5	std	500	256	0.15	0.67	1.0	0.83
IF	5	std	500	3000	0.09	1.0	0.9	0.94
IF	5	std	500	3000	0.1	1.0	1.0	1.0
IF	5	std	500	3000	0.11	0.91	1.0	0.96
IF	5	std	500	3000	0.15	0.67	1.0	0.83
IF	5	min-max	50	100	0.09	1.0	0.9	0.94
IF	5	min-max	50	100	0.1	0.98	0.98	0.98
IF	5	min-max	50	100	0.11	0.9	0.99	0.95
IF	5	min-max	50	100	0.15	0.66	0.99	0.82
IF	5	min-max	50	256	0.09	1.0	0.91	0.94

IF	5	min-max	50	256	0.1	0.99	0.99	0.99
IF	5	min-max	50	256	0.11	0.9	0.99	0.95
IF	5	min-max	50	256	0.15	0.66	0.99	0.82
IF	5	min-max	50	3000	0.09	1.0	0.9	0.94
IF	5	min-max	50	3000	0.1	0.95	0.96	0.95
IF	5	min-max	50	3000	0.11	0.87	0.96	0.92
IF	5	min-max	50	3000	0.15	0.65	0.97	0.8
IF	5	min-max	100	100	0.09	1.0	0.9	0.94
IF	5	min-max	100	100	0.1	0.99	0.99	0.99
IF	5	min-max	100	100	0.11	0.9	0.99	0.95
IF	5	min-max	100	100	0.15	0.66	0.99	0.82
IF	5	min-max	100	256	0.09	1.0	0.91	0.95
IF	5	min-max	100	256	0.1	0.99	0.99	0.99
IF	5	min-max	100	256	0.11	0.9	0.99	0.95
IF	5	min-max	100	256	0.15	0.66	1.0	0.82
IF	5	min-max	100	3000	0.09	1.0	0.9	0.94
IF	5	min-max	100	3000	0.1	0.97	0.97	0.97
IF	5	min-max	100	3000	0.11	0.89	0.98	0.94
IF	5	min-max	100	3000	0.15	0.66	0.99	0.82
IF	5	min-max	500	100	0.09	1.0	0.9	0.94
IF	5	min-max	500	100	0.1	0.99	0.99	0.99
IF	5	min-max	500	100	0.11	0.9	1.0	0.96
IF	5	min-max	500	100	0.15	0.66	1.0	0.82
IF	5	min-max	500	256	0.09	1.0	0.9	0.94
IF	5	min-max	500	256	0.1	0.99	0.99	0.99
IF	5	min-max	500	256	0.11	0.9	1.0	0.96
IF	5	min-max	500	256	0.15	0.66	1.0	0.82
IF	5	min-max	500	3000	0.09	1.0	0.9	0.94
IF	5	min-max	500	3000	0.1	0.99	0.99	0.99
IF	5	min-max	500	3000	0.11	0.9	0.99	0.95
IF	5	min-max	500	3000	0.15	0.66	1.0	0.82

Table G.1: IF grid search based on ground-truth dataset without uninteresting anomalies.

Ground-truth dataset including uninteresting anomalies

	name	motion	n_estimators	max_samples	contamination	precision	recall	fscore
IF	1	#N/A	50	100	0.09	0.67	0.76	0.72
IF	1	#N/A	50	100	0.1	0.68	0.84	0.77
IF	1	#N/A	50	100	0.11	0.66	0.91	0.79
IF	1	#N/A	50	100	0.15	0.52	0.97	0.72
IF	1	#N/A	50	256	0.09	0.65	0.73	0.69
IF	1	#N/A	50	256	0.1	0.6	0.75	0.68
IF	1	#N/A	50	256	0.11	0.54	0.75	0.65
IF	1	#N/A	50	256	0.15	0.44	0.82	0.6
IF	1	#N/A	50	3000	0.09	0.61	0.68	0.65
IF	1	#N/A	50	3000	0.1	0.64	0.8	0.72
IF	1	#N/A	50	3000	0.11	0.66	0.91	0.79
IF	1	#N/A	50	3000	0.15	0.54	1.0	0.73
IF	1	#N/A	100	100	0.09	0.65	0.73	0.7
IF	1	#N/A	100	100	0.1	0.6	0.75	0.68
IF	1	#N/A	100	100	0.11	0.59	0.81	0.7
IF	1	#N/A	100	100	0.15	0.48	0.89	0.66
IF	1	#N/A	100	256	0.09	0.67	0.74	0.71
IF	1	#N/A	100	256	0.1	0.6	0.75	0.68
IF	1	#N/A	100	256	0.11	0.55	0.75	0.65

IF	1	#N/A	100	256	0.15	0.47	0.87	0.64
IF	1	#N/A	100	3000	0.09	0.6	0.67	0.64
IF	1	#N/A	100	3000	0.1	0.64	0.79	0.72
IF	1	#N/A	100	3000	0.11	0.66	0.9	0.78
IF	1	#N/A	100	3000	0.15	0.54	1.0	0.73
IF	1	#N/A	500	100	0.09	0.66	0.73	0.7
IF	1	#N/A	500	100	0.1	0.6	0.74	0.67
IF	1	#N/A	500	100	0.11	0.54	0.74	0.65
IF	1	#N/A	500	100	0.15	0.4	0.75	0.55
IF	1	#N/A	500	256	0.09	0.66	0.73	0.7
IF	1	#N/A	500	256	0.1	0.61	0.75	0.69
IF	1	#N/A	500	256	0.11	0.55	0.76	0.66
IF	1	#N/A	500	256	0.15	0.48	0.89	0.66
IF	1	#N/A	500	3000	0.09	0.6	0.67	0.64
IF	1	#N/A	500	3000	0.1	0.62	0.77	0.7
IF	1	#N/A	500	3000	0.11	0.62	0.84	0.73
IF	1	#N/A	500	3000	0.15	0.53	0.98	0.72
IF	4	std	50	100	0.09	0.76	0.69	0.72
IF	4	std	50	100	0.1	0.74	0.74	0.74
IF	4	std	50	100	0.11	0.71	0.79	0.76
IF	4	std	50	100	0.15	0.66	1.0	0.83
IF	4	std	50	256	0.09	0.76	0.69	0.71
IF	4	std	50	256	0.1	0.71	0.71	0.71
IF	4	std	50	256	0.11	0.7	0.77	0.74
IF	4	std	50	256	0.15	0.66	0.99	0.82
IF	4	std	50	3000	0.09	0.76	0.68	0.71
IF	4	std	50	3000	0.1	0.73	0.73	0.73
IF	4	std	50	3000	0.11	0.71	0.79	0.76
IF	4	std	50	3000	0.15	0.65	0.98	0.81
IF	4	std	100	100	0.09	0.76	0.68	0.71
IF	4	std	100	100	0.1	0.74	0.74	0.74
IF	4	std	100	100	0.11	0.71	0.78	0.75
IF	4	std	100	100	0.15	0.66	0.99	0.82
IF	4	std	100	256	0.09	0.77	0.69	0.72
IF	4	std	100	256	0.1	0.72	0.73	0.73
IF	4	std	100	256	0.11	0.7	0.77	0.74
IF	4	std	100	256	0.15	0.66	1.0	0.82
IF	4	std	100	3000	0.09	0.78	0.7	0.73
IF	4	std	100	3000	0.1	0.72	0.73	0.73
IF	4	std	100	3000	0.11	0.71	0.79	0.76
IF	4	std	100	3000	0.15	0.66	0.99	0.82
IF	4	std	500	100	0.09	0.75	0.68	0.71
IF	4	std	500	100	0.1	0.71	0.71	0.71
IF	4	std	500	100	0.11	0.69	0.76	0.73
IF	4	std	500	100	0.15	0.67	1.0	0.83
IF	4	std	500	256	0.09	0.77	0.69	0.72
IF	4	std	500	256	0.1	0.74	0.74	0.74
IF	4	std	500	256	0.11	0.71	0.78	0.75
IF	4	std	500	256	0.15	0.67	1.0	0.83
IF	4	std	500	3000	0.09	0.78	0.7	0.73
IF	4	std	500	3000	0.1	0.74	0.74	0.74
IF	4	std	500	3000	0.11	0.73	0.8	0.77
IF	4	std	500	3000	0.15	0.66	1.0	0.83
IF	4	min-max	50	100	0.09	0.8	0.73	0.76
IF	4	min-max	50	100	0.1	0.76	0.76	0.76
IF	4	min-max	50	100	0.11	0.7	0.77	0.74
IF	4	min-max	50	100	0.15	0.65	0.98	0.81

IF	4	min-max	50	256	0.09	0.77	0.7	0.73
IF	4	min-max	50	256	0.1	0.72	0.72	0.72
IF	4	min-max	50	256	0.11	0.71	0.79	0.75
IF	4	min-max	50	256	0.15	0.66	1.0	0.82
IF	4	min-max	50	3000	0.09	0.82	0.74	0.77
IF	4	min-max	50	3000	0.1	0.8	0.8	0.8
IF	4	min-max	50	3000	0.11	0.76	0.84	0.8
IF	4	min-max	50	3000	0.15	0.66	0.99	0.82
IF	4	min-max	100	100	0.09	0.77	0.69	0.72
IF	4	min-max	100	100	0.1	0.74	0.75	0.75
IF	4	min-max	100	100	0.11	0.7	0.77	0.74
IF	4	min-max	100	100	0.15	0.66	0.99	0.82
IF	4	min-max	100	256	0.09	0.79	0.71	0.74
IF	4	min-max	100	256	0.1	0.75	0.75	0.75
IF	4	min-max	100	256	0.11	0.72	0.8	0.77
IF	4	min-max	100	256	0.15	0.66	1.0	0.82
IF	4	min-max	100	3000	0.09	0.78	0.71	0.74
IF	4	min-max	100	3000	0.1	0.73	0.73	0.73
IF	4	min-max	100	3000	0.11	0.71	0.78	0.75
IF	4	min-max	100	3000	0.15	0.66	0.99	0.82
IF	4	min-max	500	100	0.09	0.75	0.68	0.71
IF	4	min-max	500	100	0.1	0.72	0.73	0.73
IF	4	min-max	500	100	0.11	0.7	0.77	0.74
IF	4	min-max	500	100	0.15	0.66	0.99	0.82
IF	4	min-max	500	256	0.09	0.78	0.7	0.73
IF	4	min-max	500	256	0.1	0.74	0.75	0.75
IF	4	min-max	500	256	0.11	0.71	0.79	0.75
IF	4	min-max	500	256	0.15	0.66	1.0	0.82
IF	4	min-max	500	3000	0.09	0.78	0.7	0.73
IF	4	min-max	500	3000	0.1	0.74	0.74	0.74
IF	4	min-max	500	3000	0.11	0.72	0.79	0.76
IF	4	min-max	500	3000	0.15	0.67	1.0	0.83
IF	5	std	50	100	0.09	0.7	0.64	0.66
IF	5	std	50	100	0.1	0.67	0.68	0.68
IF	5	std	50	100	0.11	0.68	0.75	0.72
IF	5	std	50	100	0.15	0.64	0.96	0.8
IF	5	std	50	256	0.09	0.7	0.63	0.66
IF	5	std	50	256	0.1	0.7	0.7	0.7
IF	5	std	50	256	0.11	0.67	0.74	0.71
IF	5	std	50	256	0.15	0.67	1.0	0.83
IF	5	std	50	3000	0.09	0.75	0.68	0.71
IF	5	std	50	3000	0.1	0.71	0.71	0.71
IF	5	std	50	3000	0.11	0.69	0.76	0.73
IF	5	std	50	3000	0.15	0.66	0.99	0.82
IF	5	std	100	100	0.09	0.7	0.63	0.65
IF	5	std	100	100	0.1	0.68	0.68	0.68
IF	5	std	100	100	0.11	0.67	0.74	0.71
IF	5	std	100	100	0.15	0.65	0.98	0.81
IF	5	std	100	256	0.09	0.7	0.63	0.66
IF	5	std	100	256	0.1	0.71	0.71	0.71
IF	5	std	100	256	0.11	0.67	0.73	0.7
IF	5	std	100	256	0.15	0.66	0.99	0.82
IF	5	std	100	3000	0.09	0.76	0.68	0.71
IF	5	std	100	3000	0.1	0.72	0.72	0.72
IF	5	std	100	3000	0.11	0.7	0.77	0.74
IF	5	std	100	3000	0.15	0.66	0.99	0.82
IF	5	std	500	100	0.09	0.68	0.61	0.64

IF	5	std	500	100	0.1	0.68	0.68	0.68
IF	5	std	500	100	0.11	0.68	0.75	0.72
IF	5	std	500	100	0.15	0.66	0.99	0.82
IF	5	std	500	256	0.09	0.71	0.64	0.67
IF	5	std	500	256	0.1	0.71	0.71	0.71
IF	5	std	500	256	0.11	0.68	0.75	0.72
IF	5	std	500	256	0.15	0.67	1.0	0.83
IF	5	std	500	3000	0.09	0.74	0.66	0.69
IF	5	std	500	3000	0.1	0.71	0.71	0.71
IF	5	std	500	3000	0.11	0.69	0.76	0.73
IF	5	std	500	3000	0.15	0.67	1.0	0.83
IF	5	min-max	50	100	0.09	0.72	0.65	0.68
IF	5	min-max	50	100	0.1	0.68	0.68	0.68
IF	5	min-max	50	100	0.11	0.69	0.76	0.72
IF	5	min-max	50	100	0.15	0.66	0.99	0.82
IF	5	min-max	50	256	0.09	0.76	0.69	0.72
IF	5	min-max	50	256	0.1	0.71	0.71	0.71
IF	5	min-max	50	256	0.11	0.68	0.75	0.72
IF	5	min-max	50	256	0.15	0.66	0.99	0.82
IF	5	min-max	50	3000	0.09	0.76	0.69	0.72
IF	5	min-max	50	3000	0.1	0.7	0.7	0.7
IF	5	min-max	50	3000	0.11	0.68	0.75	0.72
IF	5	min-max	50	3000	0.15	0.64	0.96	0.8
IF	5	min-max	100	100	0.09	0.72	0.65	0.68
IF	5	min-max	100	100	0.1	0.69	0.69	0.69
IF	5	min-max	100	100	0.11	0.68	0.75	0.72
IF	5	min-max	100	100	0.15	0.66	0.99	0.82
IF	5	min-max	100	256	0.09	0.77	0.69	0.72
IF	5	min-max	100	256	0.1	0.71	0.71	0.71
IF	5	min-max	100	256	0.11	0.68	0.75	0.72
IF	5	min-max	100	256	0.15	0.66	0.99	0.82
IF	5	min-max	100	3000	0.09	0.77	0.69	0.72
IF	5	min-max	100	3000	0.1	0.73	0.73	0.73
IF	5	min-max	100	3000	0.11	0.69	0.76	0.73
IF	5	min-max	100	3000	0.15	0.65	0.98	0.81
IF	5	min-max	500	100	0.09	0.75	0.67	0.7
IF	5	min-max	500	100	0.1	0.7	0.7	0.7
IF	5	min-max	500	100	0.11	0.69	0.76	0.72
IF	5	min-max	500	100	0.15	0.66	1.0	0.82
IF	5	min-max	500	256	0.09	0.74	0.67	0.7
IF	5	min-max	500	256	0.1	0.7	0.7	0.7
IF	5	min-max	500	256	0.11	0.68	0.75	0.72
IF	5	min-max	500	256	0.15	0.66	1.0	0.82
IF	5	min-max	500	3000	0.09	0.78	0.7	0.73
IF	5	min-max	500	3000	0.1	0.73	0.73	0.73
IF	5	min-max	500	3000	0.11	0.71	0.78	0.75
IF	5	min-max	500	3000	0.15	0.66	0.99	0.82

Table G.2: IF grid search based on ground-truth dataset including uninteresting anomalies.

G.2 LOF grid search

Ground-truth dataset without uninteresting anomalies

	name	motion	n_neighbors	contamination	precision	recall	fscore
LOF	1	#N/A	10	0.09	0.06	0.07	0.06
LOF	1	#N/A	10	0.1	0.06	0.07	0.07
LOF	1	#N/A	10	0.11	0.06	0.08	0.07
LOF	1	#N/A	10	0.15	0.05	0.09	0.07
LOF	1	#N/A	20	0.09	0.06	0.07	0.07
LOF	1	#N/A	20	0.1	0.07	0.08	0.07
LOF	1	#N/A	20	0.11	0.06	0.08	0.07
LOF	1	#N/A	20	0.15	0.05	0.1	0.07
LOF	1	#N/A	50	0.09	0.26	0.29	0.28
LOF	1	#N/A	50	0.1	0.24	0.3	0.27
LOF	1	#N/A	50	0.11	0.23	0.31	0.27
LOF	1	#N/A	50	0.15	0.18	0.34	0.25
LOF	1	#N/A	500	0.09	0.86	0.96	0.92
LOF	1	#N/A	500	0.1	0.8	0.99	0.9
LOF	1	#N/A	500	0.11	0.73	0.99	0.87
LOF	1	#N/A	500	0.15	0.54	1.0	0.74
LOF	1	#N/A	1000	0.09	0.9	1.0	0.95
LOF	1	#N/A	1000	0.1	0.81	1.0	0.91
LOF	1	#N/A	1000	0.11	0.73	1.0	0.87
LOF	1	#N/A	1000	0.15	0.54	1.0	0.74
LOF	4	std	10	0.09	0.17	0.15	0.16
LOF	4	std	10	0.1	0.16	0.16	0.16
LOF	4	std	10	0.11	0.15	0.17	0.16
LOF	4	std	10	0.15	0.15	0.23	0.19
LOF	4	std	20	0.09	0.17	0.16	0.16
LOF	4	std	20	0.1	0.17	0.17	0.17
LOF	4	std	20	0.11	0.16	0.18	0.17
LOF	4	std	20	0.15	0.15	0.23	0.19
LOF	4	std	50	0.09	0.35	0.31	0.33
LOF	4	std	50	0.1	0.32	0.32	0.32
LOF	4	std	50	0.11	0.29	0.32	0.31
LOF	4	std	50	0.15	0.22	0.33	0.27
LOF	4	std	500	0.09	0.96	0.86	0.9
LOF	4	std	500	0.1	0.93	0.93	0.93
LOF	4	std	500	0.11	0.88	0.97	0.93
LOF	4	std	500	0.15	0.67	1.0	0.83
LOF	4	std	1000	0.09	1.0	0.9	0.94
LOF	4	std	1000	0.1	0.99	0.99	0.99
LOF	4	std	1000	0.11	0.91	1.0	0.96
LOF	4	std	1000	0.15	0.67	1.0	0.83
LOF	4	min-max	10	0.09	0.12	0.11	0.12
LOF	4	min-max	10	0.1	0.11	0.11	0.11
LOF	4	min-max	10	0.11	0.11	0.12	0.11
LOF	4	min-max	10	0.15	0.09	0.13	0.11
LOF	4	min-max	20	0.09	0.09	0.09	0.09
LOF	4	min-max	20	0.1	0.09	0.09	0.09
LOF	4	min-max	20	0.11	0.08	0.09	0.09
LOF	4	min-max	20	0.15	0.06	0.09	0.08
LOF	4	min-max	50	0.09	0.32	0.29	0.3
LOF	4	min-max	50	0.1	0.3	0.3	0.3
LOF	4	min-max	50	0.11	0.28	0.31	0.3

LOF	4	min-max	50	0.15	0.26	0.39	0.32
LOF	4	min-max	500	0.09	0.96	0.86	0.9
LOF	4	min-max	500	0.1	0.94	0.94	0.94
LOF	4	min-max	500	0.11	0.9	0.99	0.95
LOF	4	min-max	500	0.15	0.67	1.0	0.83
LOF	4	min-max	1000	0.09	1.0	0.9	0.94
LOF	4	min-max	1000	0.1	0.98	0.98	0.98
LOF	4	min-max	1000	0.11	0.91	1.0	0.96
LOF	4	min-max	1000	0.15	0.67	1.0	0.83
LOF	5	std	10	0.09	0.12	0.11	0.11
LOF	5	std	10	0.1	0.12	0.12	0.12
LOF	5	std	10	0.11	0.12	0.13	0.12
LOF	5	std	10	0.15	0.11	0.16	0.13
LOF	5	std	20	0.09	0.18	0.16	0.17
LOF	5	std	20	0.1	0.18	0.18	0.18
LOF	5	std	20	0.11	0.18	0.2	0.19
LOF	5	std	20	0.15	0.14	0.21	0.17
LOF	5	std	50	0.09	0.37	0.34	0.35
LOF	5	std	50	0.1	0.36	0.36	0.36
LOF	5	std	50	0.11	0.37	0.4	0.39
LOF	5	std	50	0.15	0.32	0.48	0.4
LOF	5	std	500	0.09	0.88	0.79	0.83
LOF	5	std	500	0.1	0.83	0.83	0.83
LOF	5	std	500	0.11	0.76	0.83	0.8
LOF	5	std	500	0.15	0.59	0.88	0.73
LOF	5	std	1000	0.09	0.97	0.87	0.91
LOF	5	std	1000	0.1	0.93	0.93	0.93
LOF	5	std	1000	0.11	0.91	1.0	0.96
LOF	5	std	1000	0.15	0.67	1.0	0.83
LOF	5	min-max	10	0.09	0.16	0.15	0.15
LOF	5	min-max	10	0.1	0.16	0.16	0.16
LOF	5	min-max	10	0.11	0.16	0.18	0.17
LOF	5	min-max	10	0.15	0.13	0.2	0.16
LOF	5	min-max	20	0.09	0.09	0.08	0.09
LOF	5	min-max	20	0.1	0.09	0.09	0.09
LOF	5	min-max	20	0.11	0.08	0.09	0.08
LOF	5	min-max	20	0.15	0.07	0.1	0.08
LOF	5	min-max	50	0.09	0.44	0.4	0.41
LOF	5	min-max	50	0.1	0.42	0.42	0.42
LOF	5	min-max	50	0.11	0.38	0.42	0.4
LOF	5	min-max	50	0.15	0.3	0.44	0.37
LOF	5	min-max	500	0.09	0.93	0.83	0.87
LOF	5	min-max	500	0.1	0.93	0.93	0.93
LOF	5	min-max	500	0.11	0.87	0.95	0.91
LOF	5	min-max	500	0.15	0.66	0.99	0.82
LOF	5	min-max	1000	0.09	1.0	0.9	0.94
LOF	5	min-max	1000	0.1	0.96	0.96	0.96
LOF	5	min-max	1000	0.11	0.91	1.0	0.96
LOF	5	min-max	1000	0.15	0.67	1.0	0.83

Table G.3: LOF grid search based on ground-truth dataset without uninteresting anomalies.

Ground-truth dataset including uninteresting anomalies.

	name	motion	n_neighbors	contamination	precision	recall	fscore
LOF	1	#N/A	10	0.09	0.07	0.08	0.08
LOF	1	#N/A	10	0.1	0.07	0.08	0.08
LOF	1	#N/A	10	0.11	0.06	0.09	0.08
LOF	1	#N/A	10	0.15	0.07	0.12	0.09
LOF	1	#N/A	20	0.09	0.09	0.1	0.1
LOF	1	#N/A	20	0.1	0.09	0.11	0.1
LOF	1	#N/A	20	0.11	0.08	0.11	0.09
LOF	1	#N/A	20	0.15	0.07	0.13	0.1
LOF	1	#N/A	50	0.09	0.31	0.35	0.33
LOF	1	#N/A	50	0.1	0.28	0.35	0.32
LOF	1	#N/A	50	0.11	0.26	0.35	0.31
LOF	1	#N/A	50	0.15	0.2	0.37	0.27
LOF	1	#N/A	500	0.09	0.6	0.67	0.64
LOF	1	#N/A	500	0.1	0.64	0.79	0.72
LOF	1	#N/A	500	0.11	0.61	0.83	0.72
LOF	1	#N/A	500	0.15	0.53	0.99	0.73
LOF	1	#N/A	1000	0.09	0.64	0.72	0.68
LOF	1	#N/A	1000	0.1	0.6	0.75	0.68
LOF	1	#N/A	1000	0.11	0.62	0.85	0.74
LOF	1	#N/A	1000	0.15	0.54	1.0	0.73
LOF	4	std	10	0.09	0.17	0.15	0.16
LOF	4	std	10	0.1	0.16	0.16	0.16
LOF	4	std	10	0.11	0.16	0.18	0.17
LOF	4	std	10	0.15	0.15	0.23	0.19
LOF	4	std	20	0.09	0.15	0.14	0.14
LOF	4	std	20	0.1	0.15	0.15	0.15
LOF	4	std	20	0.11	0.15	0.16	0.15
LOF	4	std	20	0.15	0.13	0.19	0.16
LOF	4	std	50	0.09	0.37	0.34	0.35
LOF	4	std	50	0.1	0.35	0.35	0.35
LOF	4	std	50	0.11	0.33	0.36	0.34
LOF	4	std	50	0.15	0.26	0.39	0.33
LOF	4	std	500	0.09	0.67	0.6	0.63
LOF	4	std	500	0.1	0.65	0.65	0.65
LOF	4	std	500	0.11	0.63	0.69	0.67
LOF	4	std	500	0.15	0.61	0.91	0.75
LOF	4	std	1000	0.09	0.74	0.67	0.7
LOF	4	std	1000	0.1	0.76	0.76	0.76
LOF	4	std	1000	0.11	0.76	0.83	0.8
LOF	4	std	1000	0.15	0.67	1.0	0.83
LOF	4	min-max	10	0.09	0.11	0.1	0.1
LOF	4	min-max	10	0.1	0.1	0.1	0.1
LOF	4	min-max	10	0.11	0.1	0.11	0.1
LOF	4	min-max	10	0.15	0.08	0.13	0.1
LOF	4	min-max	20	0.09	0.08	0.08	0.08
LOF	4	min-max	20	0.1	0.08	0.08	0.08
LOF	4	min-max	20	0.11	0.07	0.08	0.07
LOF	4	min-max	20	0.15	0.06	0.1	0.08
LOF	4	min-max	50	0.09	0.29	0.26	0.27

LOF	4	min-max	50	0.1	0.27	0.27	0.27
LOF	4	min-max	50	0.11	0.26	0.29	0.28
LOF	4	min-max	50	0.15	0.26	0.39	0.33
LOF	4	min-max	500	0.09	0.67	0.61	0.63
LOF	4	min-max	500	0.1	0.62	0.62	0.62
LOF	4	min-max	500	0.11	0.62	0.68	0.66
LOF	4	min-max	500	0.15	0.63	0.95	0.79
LOF	4	min-max	1000	0.09	0.75	0.67	0.7
LOF	4	min-max	1000	0.1	0.76	0.76	0.76
LOF	4	min-max	1000	0.11	0.75	0.82	0.79
LOF	4	min-max	1000	0.15	0.67	1.0	0.83
LOF	5	std	10	0.09	0.12	0.11	0.11
LOF	5	std	10	0.1	0.11	0.11	0.11
LOF	5	std	10	0.11	0.11	0.12	0.12
LOF	5	std	10	0.15	0.1	0.15	0.12
LOF	5	std	20	0.09	0.22	0.2	0.21
LOF	5	std	20	0.1	0.21	0.21	0.21
LOF	5	std	20	0.11	0.19	0.21	0.2
LOF	5	std	20	0.15	0.15	0.23	0.19
LOF	5	std	50	0.09	0.33	0.29	0.31
LOF	5	std	50	0.1	0.34	0.34	0.34
LOF	5	std	50	0.11	0.34	0.38	0.36
LOF	5	std	50	0.15	0.28	0.43	0.35
LOF	5	std	500	0.09	0.48	0.43	0.45
LOF	5	std	500	0.1	0.46	0.46	0.46
LOF	5	std	500	0.11	0.48	0.53	0.51
LOF	5	std	500	0.15	0.52	0.79	0.65
LOF	5	std	1000	0.09	0.7	0.63	0.66
LOF	5	std	1000	0.1	0.7	0.7	0.7
LOF	5	std	1000	0.11	0.71	0.78	0.75
LOF	5	std	1000	0.15	0.67	1.0	0.83
LOF	5	min-max	10	0.09	0.16	0.14	0.15
LOF	5	min-max	10	0.1	0.16	0.16	0.16
LOF	5	min-max	10	0.11	0.16	0.17	0.17
LOF	5	min-max	10	0.15	0.13	0.2	0.16
LOF	5	min-max	20	0.09	0.13	0.12	0.12
LOF	5	min-max	20	0.1	0.12	0.12	0.12
LOF	5	min-max	20	0.11	0.12	0.13	0.13
LOF	5	min-max	20	0.15	0.11	0.16	0.13
LOF	5	min-max	50	0.09	0.44	0.39	0.41
LOF	5	min-max	50	0.1	0.41	0.41	0.41
LOF	5	min-max	50	0.11	0.38	0.41	0.4
LOF	5	min-max	50	0.15	0.3	0.45	0.37
LOF	5	min-max	500	0.09	0.51	0.46	0.48
LOF	5	min-max	500	0.1	0.55	0.55	0.55
LOF	5	min-max	500	0.11	0.54	0.6	0.57
LOF	5	min-max	500	0.15	0.62	0.93	0.77
LOF	5	min-max	1000	0.09	0.71	0.64	0.66
LOF	5	min-max	1000	0.1	0.7	0.7	0.7
LOF	5	min-max	1000	0.11	0.73	0.8	0.77
LOF	5	min-max	1000	0.15	0.67	1.0	0.83

Table G.4: LOF grid search based on ground-truth dataset including uninteresting anomalies.

G.3 RF regressor grid search

Ground-truth dataset without uninteresting anomalies

	name	motion	n_estimators	max_depth	max_features	contamination	precision	recall	fscore
RF	1	#N/A	50	None	auto	0.09	0.57	0.63	0.6
RF	1	#N/A	50	None	auto	0.1	0.53	0.65	0.6
RF	1	#N/A	50	None	auto	0.11	0.5	0.67	0.59
RF	1	#N/A	50	None	auto	0.15	0.38	0.71	0.53
RF	1	#N/A	50	None	log2	0.09	0.64	0.71	0.68
RF	1	#N/A	50	None	log2	0.1	0.59	0.73	0.67
RF	1	#N/A	50	None	log2	0.11	0.54	0.73	0.64
RF	1	#N/A	50	None	log2	0.15	0.44	0.81	0.6
RF	1	#N/A	50	25	auto	0.09	0.57	0.63	0.6
RF	1	#N/A	50	25	auto	0.1	0.53	0.65	0.6
RF	1	#N/A	50	25	auto	0.11	0.5	0.67	0.59
RF	1	#N/A	50	25	auto	0.15	0.38	0.71	0.53
RF	1	#N/A	50	25	log2	0.09	0.64	0.71	0.68
RF	1	#N/A	50	25	log2	0.1	0.59	0.73	0.66
RF	1	#N/A	50	25	log2	0.11	0.54	0.73	0.64
RF	1	#N/A	50	25	log2	0.15	0.43	0.81	0.59
RF	1	#N/A	50	100	auto	0.09	0.57	0.63	0.6
RF	1	#N/A	50	100	auto	0.1	0.53	0.65	0.6
RF	1	#N/A	50	100	auto	0.11	0.5	0.67	0.59
RF	1	#N/A	50	100	auto	0.15	0.38	0.71	0.53
RF	1	#N/A	50	100	log2	0.09	0.64	0.71	0.68
RF	1	#N/A	50	100	log2	0.1	0.59	0.73	0.67
RF	1	#N/A	50	100	log2	0.11	0.54	0.73	0.64
RF	1	#N/A	50	100	log2	0.15	0.44	0.81	0.6
RF	1	#N/A	100	None	auto	0.09	0.57	0.64	0.61
RF	1	#N/A	100	None	auto	0.1	0.54	0.66	0.6
RF	1	#N/A	100	None	auto	0.11	0.5	0.67	0.59
RF	1	#N/A	100	None	auto	0.15	0.39	0.72	0.53
RF	1	#N/A	100	None	log2	0.09	0.65	0.72	0.69
RF	1	#N/A	100	None	log2	0.1	0.6	0.75	0.68
RF	1	#N/A	100	None	log2	0.11	0.57	0.77	0.67
RF	1	#N/A	100	None	log2	0.15	0.43	0.81	0.59
RF	1	#N/A	100	25	auto	0.09	0.57	0.64	0.61
RF	1	#N/A	100	25	auto	0.1	0.54	0.66	0.6
RF	1	#N/A	100	25	auto	0.11	0.5	0.67	0.59
RF	1	#N/A	100	25	auto	0.15	0.39	0.72	0.53
RF	1	#N/A	100	25	log2	0.09	0.64	0.71	0.68
RF	1	#N/A	100	25	log2	0.1	0.6	0.74	0.67
RF	1	#N/A	100	25	log2	0.11	0.55	0.75	0.65
RF	1	#N/A	100	25	log2	0.15	0.44	0.81	0.6
RF	1	#N/A	100	100	auto	0.09	0.57	0.64	0.61
RF	1	#N/A	100	100	auto	0.1	0.54	0.66	0.6
RF	1	#N/A	100	100	auto	0.11	0.5	0.67	0.59
RF	1	#N/A	100	100	auto	0.15	0.39	0.72	0.53
RF	1	#N/A	100	100	log2	0.09	0.65	0.72	0.69
RF	1	#N/A	100	100	log2	0.1	0.6	0.75	0.68
RF	1	#N/A	100	100	log2	0.11	0.57	0.77	0.67
RF	1	#N/A	100	100	log2	0.15	0.43	0.81	0.59
RF	1	#N/A	500	None	auto	0.09	0.55	0.62	0.59
RF	1	#N/A	500	None	auto	0.1	0.52	0.64	0.58
RF	1	#N/A	500	None	auto	0.11	0.48	0.65	0.57
RF	1	#N/A	500	None	auto	0.15	0.39	0.72	0.53
RF	1	#N/A	500	None	log2	0.09	0.67	0.75	0.71
RF	1	#N/A	500	None	log2	0.1	0.63	0.77	0.7
RF	1	#N/A	500	None	log2	0.11	0.58	0.79	0.68
RF	1	#N/A	500	None	log2	0.15	0.45	0.83	0.61
RF	1	#N/A	500	25	auto	0.09	0.55	0.62	0.59
RF	1	#N/A	500	25	auto	0.1	0.52	0.64	0.58
RF	1	#N/A	500	25	auto	0.11	0.48	0.65	0.57
RF	1	#N/A	500	25	auto	0.15	0.39	0.72	0.53

RF	1	#N/A	500	25	log2	0.09	0.67	0.75	0.72
RF	1	#N/A	500	25	log2	0.1	0.62	0.77	0.7
RF	1	#N/A	500	25	log2	0.11	0.58	0.78	0.68
RF	1	#N/A	500	25	log2	0.15	0.45	0.84	0.62
RF	1	#N/A	500	100	auto	0.09	0.55	0.62	0.59
RF	1	#N/A	500	100	auto	0.1	0.52	0.64	0.58
RF	1	#N/A	500	100	auto	0.11	0.48	0.65	0.57
RF	1	#N/A	500	100	auto	0.15	0.39	0.72	0.53
RF	1	#N/A	500	100	log2	0.09	0.67	0.75	0.71
RF	1	#N/A	500	100	log2	0.1	0.63	0.77	0.7
RF	1	#N/A	500	100	log2	0.11	0.58	0.79	0.68
RF	1	#N/A	500	100	log2	0.15	0.45	0.83	0.61
RF	3	#N/A	50	None	auto	0.09	0.69	0.62	0.65
RF	3	#N/A	50	None	auto	0.1	0.64	0.64	0.64
RF	3	#N/A	50	None	auto	0.11	0.6	0.66	0.63
RF	3	#N/A	50	None	auto	0.15	0.5	0.74	0.62
RF	3	#N/A	50	None	log2	0.09	0.64	0.58	0.6
RF	3	#N/A	50	None	log2	0.1	0.61	0.61	0.61
RF	3	#N/A	50	None	log2	0.11	0.58	0.64	0.61
RF	3	#N/A	50	None	log2	0.15	0.47	0.71	0.59
RF	3	#N/A	50	25	auto	0.09	0.68	0.62	0.64
RF	3	#N/A	50	25	auto	0.1	0.64	0.64	0.64
RF	3	#N/A	50	25	auto	0.11	0.6	0.66	0.63
RF	3	#N/A	50	25	auto	0.15	0.49	0.73	0.6
RF	3	#N/A	50	25	log2	0.09	0.65	0.59	0.61
RF	3	#N/A	50	25	log2	0.1	0.62	0.62	0.62
RF	3	#N/A	50	25	log2	0.11	0.59	0.65	0.62
RF	3	#N/A	50	25	log2	0.15	0.47	0.71	0.58
RF	3	#N/A	50	100	auto	0.09	0.69	0.62	0.65
RF	3	#N/A	50	100	auto	0.1	0.64	0.64	0.64
RF	3	#N/A	50	100	auto	0.11	0.6	0.66	0.63
RF	3	#N/A	50	100	auto	0.15	0.5	0.74	0.62
RF	3	#N/A	50	100	log2	0.09	0.64	0.58	0.6
RF	3	#N/A	50	100	log2	0.1	0.61	0.61	0.61
RF	3	#N/A	50	100	log2	0.11	0.58	0.64	0.61
RF	3	#N/A	50	100	log2	0.15	0.47	0.71	0.59
RF	3	#N/A	100	None	auto	0.09	0.69	0.63	0.65
RF	3	#N/A	100	None	auto	0.1	0.65	0.65	0.65
RF	3	#N/A	100	None	auto	0.11	0.61	0.67	0.64
RF	3	#N/A	100	None	auto	0.15	0.5	0.75	0.62
RF	3	#N/A	100	None	log2	0.09	0.66	0.6	0.62
RF	3	#N/A	100	None	log2	0.1	0.62	0.62	0.62
RF	3	#N/A	100	None	log2	0.11	0.58	0.64	0.62
RF	3	#N/A	100	None	log2	0.15	0.48	0.72	0.6
RF	3	#N/A	100	25	auto	0.09	0.69	0.62	0.65
RF	3	#N/A	100	25	auto	0.1	0.64	0.64	0.64
RF	3	#N/A	100	25	auto	0.11	0.61	0.67	0.64
RF	3	#N/A	100	25	auto	0.15	0.5	0.75	0.62
RF	3	#N/A	100	25	log2	0.09	0.66	0.59	0.62
RF	3	#N/A	100	25	log2	0.1	0.62	0.62	0.62
RF	3	#N/A	100	25	log2	0.11	0.58	0.64	0.61
RF	3	#N/A	100	25	log2	0.15	0.48	0.72	0.6
RF	3	#N/A	100	100	auto	0.09	0.69	0.63	0.65
RF	3	#N/A	100	100	auto	0.1	0.65	0.65	0.65
RF	3	#N/A	100	100	auto	0.11	0.61	0.67	0.64
RF	3	#N/A	100	100	auto	0.15	0.5	0.75	0.62
RF	3	#N/A	100	100	log2	0.09	0.66	0.6	0.62
RF	3	#N/A	100	100	log2	0.1	0.62	0.62	0.62
RF	3	#N/A	100	100	log2	0.11	0.58	0.64	0.62
RF	3	#N/A	100	100	log2	0.15	0.48	0.72	0.6
RF	3	#N/A	500	None	auto	0.09	0.68	0.62	0.64
RF	3	#N/A	500	None	auto	0.1	0.65	0.65	0.65
RF	3	#N/A	500	None	auto	0.11	0.61	0.67	0.64
RF	3	#N/A	500	None	auto	0.15	0.49	0.74	0.61
RF	3	#N/A	500	None	log2	0.09	0.67	0.6	0.63
RF	3	#N/A	500	None	log2	0.1	0.63	0.63	0.63

RF	3	#N/A	500	None	log2	0.11	0.59	0.65	0.63
RF	3	#N/A	500	None	log2	0.15	0.48	0.73	0.6
RF	3	#N/A	500	25	auto	0.09	0.69	0.62	0.64
RF	3	#N/A	500	25	auto	0.1	0.64	0.64	0.64
RF	3	#N/A	500	25	auto	0.11	0.61	0.67	0.64
RF	3	#N/A	500	25	auto	0.15	0.49	0.74	0.61
RF	3	#N/A	500	25	log2	0.09	0.67	0.6	0.63
RF	3	#N/A	500	25	log2	0.1	0.62	0.62	0.62
RF	3	#N/A	500	25	log2	0.11	0.59	0.65	0.62
RF	3	#N/A	500	25	log2	0.15	0.48	0.73	0.6
RF	3	#N/A	500	100	auto	0.09	0.68	0.62	0.64
RF	3	#N/A	500	100	auto	0.1	0.65	0.65	0.65
RF	3	#N/A	500	100	auto	0.11	0.61	0.67	0.64
RF	3	#N/A	500	100	auto	0.15	0.49	0.74	0.61
RF	3	#N/A	500	100	log2	0.09	0.67	0.6	0.63
RF	3	#N/A	500	100	log2	0.1	0.63	0.63	0.63
RF	3	#N/A	500	100	log2	0.11	0.59	0.65	0.63
RF	3	#N/A	500	100	log2	0.15	0.48	0.73	0.6

Table G.5: RF regressor grid search based on ground-truth dataset without uninteresting anomalies.

Ground-truth dataset including uninteresting anomalies

	name	motion	n_estimators	max_depth	max_features	contamination	precision	recall	fscore
RF	1	#N/A	50	None	auto	0.09	0.5	0.56	0.53
RF	1	#N/A	50	None	auto	0.1	0.47	0.59	0.53
RF	1	#N/A	50	None	auto	0.11	0.44	0.6	0.52
RF	1	#N/A	50	None	auto	0.15	0.37	0.68	0.5
RF	1	#N/A	50	None	log2	0.09	0.56	0.63	0.6
RF	1	#N/A	50	None	log2	0.1	0.53	0.65	0.59
RF	1	#N/A	50	None	log2	0.11	0.49	0.68	0.59
RF	1	#N/A	50	None	log2	0.15	0.4	0.74	0.54
RF	1	#N/A	50	25	auto	0.09	0.5	0.56	0.53
RF	1	#N/A	50	25	auto	0.1	0.47	0.59	0.53
RF	1	#N/A	50	25	auto	0.11	0.44	0.6	0.52
RF	1	#N/A	50	25	auto	0.15	0.37	0.68	0.5
RF	1	#N/A	50	25	log2	0.09	0.56	0.63	0.6
RF	1	#N/A	50	25	log2	0.1	0.52	0.65	0.59
RF	1	#N/A	50	25	log2	0.11	0.5	0.68	0.59
RF	1	#N/A	50	25	log2	0.15	0.4	0.74	0.54
RF	1	#N/A	50	100	auto	0.09	0.5	0.56	0.53
RF	1	#N/A	50	100	auto	0.1	0.47	0.59	0.53
RF	1	#N/A	50	100	auto	0.11	0.44	0.6	0.52
RF	1	#N/A	50	100	auto	0.15	0.37	0.68	0.5
RF	1	#N/A	50	100	log2	0.09	0.56	0.63	0.6
RF	1	#N/A	50	100	log2	0.1	0.53	0.65	0.59
RF	1	#N/A	50	100	log2	0.11	0.49	0.68	0.59
RF	1	#N/A	50	100	log2	0.15	0.4	0.74	0.54
RF	1	#N/A	100	None	auto	0.09	0.5	0.56	0.54
RF	1	#N/A	100	None	auto	0.1	0.47	0.59	0.53
RF	1	#N/A	100	None	auto	0.11	0.44	0.6	0.52
RF	1	#N/A	100	None	auto	0.15	0.37	0.69	0.5
RF	1	#N/A	100	None	log2	0.09	0.55	0.62	0.59
RF	1	#N/A	100	None	log2	0.1	0.52	0.65	0.59
RF	1	#N/A	100	None	log2	0.11	0.5	0.69	0.6
RF	1	#N/A	100	None	log2	0.15	0.42	0.78	0.57
RF	1	#N/A	100	25	auto	0.09	0.5	0.56	0.54
RF	1	#N/A	100	25	auto	0.1	0.47	0.59	0.53
RF	1	#N/A	100	25	auto	0.11	0.44	0.6	0.52
RF	1	#N/A	100	25	auto	0.15	0.37	0.69	0.5
RF	1	#N/A	100	25	log2	0.09	0.55	0.62	0.59
RF	1	#N/A	100	25	log2	0.1	0.52	0.64	0.58

RF	1	#N/A	100	25	log2	0.11	0.5	0.68	0.59
RF	1	#N/A	100	25	log2	0.15	0.41	0.76	0.56
RF	1	#N/A	100	100	auto	0.09	0.5	0.56	0.54
RF	1	#N/A	100	100	auto	0.1	0.47	0.59	0.53
RF	1	#N/A	100	100	auto	0.11	0.44	0.6	0.52
RF	1	#N/A	100	100	auto	0.15	0.37	0.69	0.5
RF	1	#N/A	100	100	log2	0.09	0.55	0.62	0.59
RF	1	#N/A	100	100	log2	0.1	0.52	0.65	0.59
RF	1	#N/A	100	100	log2	0.11	0.5	0.69	0.6
RF	1	#N/A	100	100	log2	0.15	0.42	0.78	0.57
RF	1	#N/A	500	None	auto	0.09	0.5	0.55	0.53
RF	1	#N/A	500	None	auto	0.1	0.45	0.56	0.51
RF	1	#N/A	500	None	auto	0.11	0.42	0.58	0.5
RF	1	#N/A	500	None	auto	0.15	0.36	0.67	0.49
RF	1	#N/A	500	None	log2	0.09	0.57	0.64	0.61
RF	1	#N/A	500	None	log2	0.1	0.55	0.68	0.62
RF	1	#N/A	500	None	log2	0.11	0.52	0.71	0.62
RF	1	#N/A	500	None	log2	0.15	0.42	0.79	0.58
RF	1	#N/A	500	25	auto	0.09	0.5	0.55	0.53
RF	1	#N/A	500	25	auto	0.1	0.45	0.56	0.51
RF	1	#N/A	500	25	auto	0.11	0.42	0.58	0.5
RF	1	#N/A	500	25	auto	0.15	0.36	0.67	0.49
RF	1	#N/A	500	25	log2	0.09	0.58	0.65	0.62
RF	1	#N/A	500	25	log2	0.1	0.55	0.68	0.62
RF	1	#N/A	500	25	log2	0.11	0.52	0.71	0.61
RF	1	#N/A	500	25	log2	0.15	0.42	0.79	0.58
RF	1	#N/A	500	100	auto	0.09	0.5	0.55	0.53
RF	1	#N/A	500	100	auto	0.1	0.45	0.56	0.51
RF	1	#N/A	500	100	auto	0.11	0.42	0.58	0.5
RF	1	#N/A	500	100	auto	0.15	0.36	0.67	0.49
RF	1	#N/A	500	100	log2	0.09	0.57	0.64	0.61
RF	1	#N/A	500	100	log2	0.1	0.55	0.68	0.62
RF	1	#N/A	500	100	log2	0.11	0.52	0.71	0.62
RF	1	#N/A	500	100	log2	0.15	0.42	0.79	0.58
RF	3	#N/A	50	None	auto	0.09	0.6	0.54	0.57
RF	3	#N/A	50	None	auto	0.1	0.59	0.59	0.59
RF	3	#N/A	50	None	auto	0.11	0.57	0.62	0.6
RF	3	#N/A	50	None	auto	0.15	0.47	0.71	0.58
RF	3	#N/A	50	None	log2	0.09	0.6	0.54	0.57
RF	3	#N/A	50	None	log2	0.1	0.56	0.56	0.56
RF	3	#N/A	50	None	log2	0.11	0.53	0.58	0.56
RF	3	#N/A	50	None	log2	0.15	0.45	0.67	0.56
RF	3	#N/A	50	25	auto	0.09	0.6	0.54	0.57
RF	3	#N/A	50	25	auto	0.1	0.59	0.59	0.59
RF	3	#N/A	50	25	auto	0.11	0.56	0.62	0.59
RF	3	#N/A	50	25	auto	0.15	0.46	0.69	0.57
RF	3	#N/A	50	25	log2	0.09	0.61	0.55	0.57
RF	3	#N/A	50	25	log2	0.1	0.57	0.57	0.57
RF	3	#N/A	50	25	log2	0.11	0.54	0.59	0.57
RF	3	#N/A	50	25	log2	0.15	0.45	0.68	0.56
RF	3	#N/A	50	100	auto	0.09	0.6	0.54	0.57
RF	3	#N/A	50	100	auto	0.1	0.59	0.59	0.59
RF	3	#N/A	50	100	auto	0.11	0.57	0.62	0.6
RF	3	#N/A	50	100	auto	0.15	0.47	0.71	0.58
RF	3	#N/A	50	100	log2	0.09	0.6	0.54	0.57
RF	3	#N/A	50	100	log2	0.1	0.56	0.56	0.56
RF	3	#N/A	50	100	log2	0.11	0.53	0.58	0.56
RF	3	#N/A	50	100	log2	0.15	0.45	0.67	0.56
RF	3	#N/A	100	None	auto	0.09	0.63	0.56	0.59
RF	3	#N/A	100	None	auto	0.1	0.6	0.6	0.6
RF	3	#N/A	100	None	auto	0.11	0.57	0.63	0.6
RF	3	#N/A	100	None	auto	0.15	0.47	0.71	0.59
RF	3	#N/A	100	None	log2	0.09	0.61	0.55	0.58
RF	3	#N/A	100	None	log2	0.1	0.57	0.58	0.58
RF	3	#N/A	100	None	log2	0.11	0.54	0.6	0.58
RF	3	#N/A	100	None	log2	0.15	0.45	0.68	0.56

RF	3	#N/A	100	25	auto	0.09	0.62	0.56	0.58
RF	3	#N/A	100	25	auto	0.1	0.6	0.6	0.6
RF	3	#N/A	100	25	auto	0.11	0.56	0.62	0.6
RF	3	#N/A	100	25	auto	0.15	0.47	0.71	0.59
RF	3	#N/A	100	25	log2	0.09	0.61	0.55	0.57
RF	3	#N/A	100	25	log2	0.1	0.57	0.57	0.57
RF	3	#N/A	100	25	log2	0.11	0.54	0.6	0.58
RF	3	#N/A	100	25	log2	0.15	0.45	0.68	0.56
RF	3	#N/A	100	100	auto	0.09	0.63	0.56	0.59
RF	3	#N/A	100	100	auto	0.1	0.6	0.6	0.6
RF	3	#N/A	100	100	auto	0.11	0.57	0.63	0.6
RF	3	#N/A	100	100	auto	0.15	0.47	0.71	0.59
RF	3	#N/A	100	100	log2	0.09	0.61	0.55	0.58
RF	3	#N/A	100	100	log2	0.1	0.57	0.58	0.58
RF	3	#N/A	100	100	log2	0.11	0.54	0.6	0.58
RF	3	#N/A	100	100	log2	0.15	0.45	0.68	0.56
RF	3	#N/A	500	None	auto	0.09	0.61	0.55	0.58
RF	3	#N/A	500	None	auto	0.1	0.59	0.59	0.59
RF	3	#N/A	500	None	auto	0.11	0.56	0.62	0.59
RF	3	#N/A	500	None	auto	0.15	0.47	0.71	0.59
RF	3	#N/A	500	None	log2	0.09	0.6	0.54	0.57
RF	3	#N/A	500	None	log2	0.1	0.57	0.58	0.58
RF	3	#N/A	500	None	log2	0.11	0.55	0.61	0.58
RF	3	#N/A	500	None	log2	0.15	0.46	0.7	0.58
RF	3	#N/A	500	25	auto	0.09	0.61	0.55	0.58
RF	3	#N/A	500	25	auto	0.1	0.59	0.59	0.59
RF	3	#N/A	500	25	auto	0.11	0.56	0.62	0.6
RF	3	#N/A	500	25	auto	0.15	0.47	0.71	0.59
RF	3	#N/A	500	25	log2	0.09	0.6	0.54	0.57
RF	3	#N/A	500	25	log2	0.1	0.57	0.57	0.57
RF	3	#N/A	500	25	log2	0.11	0.55	0.61	0.58
RF	3	#N/A	500	25	log2	0.15	0.46	0.69	0.57
RF	3	#N/A	500	100	auto	0.09	0.61	0.55	0.58
RF	3	#N/A	500	100	auto	0.1	0.59	0.59	0.59
RF	3	#N/A	500	100	auto	0.11	0.56	0.62	0.59
RF	3	#N/A	500	100	auto	0.15	0.47	0.71	0.59
RF	3	#N/A	500	100	log2	0.09	0.6	0.54	0.57
RF	3	#N/A	500	100	log2	0.1	0.57	0.58	0.58
RF	3	#N/A	500	100	log2	0.11	0.55	0.61	0.58
RF	3	#N/A	500	100	log2	0.15	0.46	0.7	0.58

Table G.6: RF regressor grid search based on ground-truth dataset including uninteresting anomalies.