Verena Niederwanger, BSc

A data-driven approach to determine product quality in manufacturing

Master's Thesis

to achieve the university degree of

Diplom-Ingenieur or international Master of Science

Master's degree programme: Software Engineering and Management

submitted to

Graz University of Technology

Supervisor

Dipl-Ing. Johannes Pan// Dipl-Ing. Clemens Gutschi// Dipl-Ing. Dr.techn. Nikolaus Furian

Institute of Engineering and Business Informatics
Head: Univ.-Prof. Dipl-Ing. Dr.techn. Siegfried Vössner

Graz, May 2019

This master thesis is written in cooperation with GKN Driveline Bruneck, an automotive supplier for drive systems located in South Tyrol, Italy. The company's homepage may be found here: https://www.gknbruneck.com.

This thesis is written in order to achieve the Austrian academic degree *'Diplom-Ingenieur'* (Dipl.Ing.), equivalent to the international *'Master of Science'* (MSc).

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date

Signature

Acknowledgement

For me it was always clear that I want to do an academic research within a field, which is not only interesting for me, but has also relevancy for companies. Therefore, I want to thank the management of GKN Driveline Bruneck for giving me the opportunity to write my master thesis in cooperation with the Institute of Engineering and Business Informatics about this very interesting project and for all the support I got during that time.

A special thank goes to my GKN-internal supervisor Thomas Seiwald for his ideas and the continuous support. I want to thank also Dietmar Tasser and Matthias Voppichler for their help towards machine connection and IT-related questions. A big thank goes also to the operators in gear manufacturing for supporting this project in the best way possible by giving me a helping hand during the experimental setup.

Thanks go also to Präwema Antriebstechnik GmbH, for the sharing their machine and process knowledge with me and the permission to use certain images of their machine.

I would sincerley like to thank also my mentors from the Institute, especially my main supervisor Johannes Pan and co-supervisor Clemens Gutschi, for their guiding and all the helpful suggestions.

Finally, my thanks go to my family and friends for their never-stopping support and patience throughout the years and especially during the last few months.

Abstract

Product quality is a very important evaluation criteria for companies all over the world. Customers expect and demand from their suppliers high quality products to satisfy their needs. Especially in the field of automotive, military and aerospace, a high level of quality is important to ensure the safe operation of products. If the required product quality is not fulfilled, this can cause component or system breakdowns and in the worst case cost people their lives. This will end up throwing a bad light on the company, destroying its reputation and finally can lead to the closure of the company. Therefore, it is crucial to ensure the required product quality. To do so, continuous measurements and tests are executed. In a production environment it is not always possible to perform a 100% testing of all components, due to performance and logistic reasons. Therefore, in this thesis a research was conducted, in order to determine the product quality in manufacturing based on machine data. This would then lead to a more efficient production process and detect scrap parts already when they are produced.

This thesis was conducted in cooperation with GKN Driveline Bruneck. The first step was to select a suitable monitoring method and parameters, which provide relevant information about the machines condition. A framework for data acquisition and ingestion was then developed in order to transfer the data continuously from the machine to the data storage. Experiments were conducted in order to gather relevant data, which could be analysed later. For the analysis different data-driven approaches and methods were used.

Contents

List of Figures

List of Tables

Abbreviations

API Application Programming Interface

CMS Condition Monitoring System

CPS Cyber-physical System

ELT Extract, Load, Transform

EOL End-of-Line

ETL Extract, Transform, Load

ETM Electronic Torque Management

EV Electric Vehicle

FFT Fast Fourier Transform

HDFS Hadoop Distributed File System

HEV Hybrid Electric Vehicle

Hz Hertz

IoT Internet of Things

I/O Input/Output

KDD Knowledge Discovery in Databases

MAD Magnetic, Agile, Deep

MEMS Microelectromechanical System

MPP Massive Parallel Processing

NoSQL Not Only Structured Query Language

NVH Noise, Vibration, Harshness

PB Petabyte

PLC Programmable Logic Controller

RFE Recursive Feature Elimination

RPM Rounds per Minute

RSS Rich Site Summary

TB Terabyte

VSD VarioSpeedDresser

XML eXtensible Markup Language

1 Introduction

At the moment, the term Industry 4.0 is on everyone's lips and all companies want to drive it, because they expect an increase in productivity, efficiency and cost savings. Important parts of Industry 4.0 are condition monitoring and process optimization. Both methods provide the possibility of saving effort and money. Condition monitoring becomes increasingly important due to the rising complexity and automation of machines and assembly lines, the overall shortage of skilled personnel and the high wage costs. Condition monitoring can help to prevent longer machine downtimes and make maintenance tasks more efficient. For condition monitoring it is necessary to equip the machines with sensors and to continuously retrieve the relevant parameters from the machine. The data is then collected, stored and can then be visualized and analysed accordingly.

Process optimization instead aims for example to increase machine and tool utilization and to decrease production time. This helps the companies to reduce costs for production, machines and toolings. To detect optimization possibilities is is necessary to retrieve over sensors or the PLC certain parameters from the machine. By analysing these attributes inconsistencies or deviations can be detected and later on corrected. Several researches already showed, that machine data can help to optimize the production processes and detect potential failures. Since the analysis of the data is at times quite complicated and takes therefore some time and effort it is crucial to select the right machine parameters and methods. This can then prevent not meaningful analysis results and additional work.

1.1 Objective of the thesis

The main objective of this thesis was to examine if it is possible to detect the product quality in manufacturing based on data originating from the machine. The research was conducted in the gear manufacturing of GKN Driveline Bruneck. Different experiments were done on a machine, where the influence of tool wear on the final product quality was examined. The aim was to optimize not only the product quality, but also to detect scrap parts in advance, before they are further machined. This can then increase the customer satisfaction, reduce production and scrapping costs.

1.2 Structure of the thesis

The thesis is structured into two parts: a theoretical and a practical part. The first part is divided into different chapters, which give a theoretical background over the different discussed topics. The topics cover a wide range from maintenance to condition monitoring, big data and different manufacturing methods for gears.
The second part instead describes the practical work which was conducted during the course of this thesis at GKN Driveline Bruneck. At the beginning a short introduction about the GKN Group and especially GKN Driveline Bruneck is given. Then the importance of the use case is described and why the gear manufacturing was chosen for this research. The influence of tool wear on the product quality and the conducted experiments to proof the influence and gather relevant data are described. The performed data analyses are also further explained. All results and outcomes are then collected and presented. At the end an outlook for further projects and research is given to conclude the research.

2 Maintenance

Maintenance is an important department in each company, since it does not only influence directly the machine downtimes and therefore costs, but affects also the quality of the products. In recent years more and more companies realised the impact of maintenance and try to improve their overall maintenance activities and processes. This chapter describes not only the main tasks and goals of maintenance, but also the main strategies are further explained. Another important discussed aspect is wear, especially tool wear, which relevancy regarding product quality is highlighted.

2.1 Maintenance tasks

To the tasks of maintenance belong not only the upkeep, inspection, repair and optimization of machines, but also the analysis of the breakdown behaviour, the detection of potential faults and the active prevention of malfunctions as described in the DIN 31051.

Upkeep

All the measures to maintain a machine in its original state or which delay the wear belong to the task of upkeeping a machine. (Schenk, 2010, pp.23) The main goals are the prolongation of machine or parts lifetime and the preservation of occupational safety. Some of the upkeeping actions are: cleaning, preserving, readjusting, greasing, replenishing and replacing. (Matyas, 2016, pp.38)

Inspection

Inspection comprises all actions to define and evaluate the current state of a machine, including the reasons for equipment wearing. (Schenk, 2010, pp.24) It is very important for those checks that the operational and environmental conditions do not change significantly. The process for an inspection should always consist of the following steps:

- Determine the condition
- Evaluate the condition
- Analyse the condition information
- Cause analysis
- Failure analysis
- Other required actions

Diagnostic systems can support inspections by providing offline or online monitoring services which replace some manual checks or simplify inspection tasks. Additionally, diagnostic systems are also able to detect failures when they occur and predict errors before they happen. (Matyas, 2016, pp.35-37)

Repair

All measures to get a machine back in its functioning state by either mending or replacing machine parts are labelled as repair. According to the time and planability of the service actions, this task can be divided into:

- **Planned repair** envisages that all maintenance measures are planned and executed according to time and extent to preserve or restore full operability. Maintenance intervals are scheduled mainly on experience or when the probability of a machine impairment is high.

- **Prepared repair** provides that all maintenance actions are preplanned, but the execution time is not known.
- **Unplanned repair** labels all maintenance measures where neither the execution time, the extend nor the kind of work are known. (Matyas, 2016, pp.39-40)

Improvement

This task comprises all combinations of technical, administrative and managerial actions to increase the functional reliability of a machine without changing its required function. (Schenk, 2010, pp.24) Starting with the failure analysis, an examination takes place, analysing if the object causing the error is improvable and the enhancement economically reasonable. (Matyas, 2016, pp.40-41)

2.2 Objectives of maintenance

General objectives of maintenance are targets centred around availability, cost reduction, product quality, safety, environment and asset value preservation. (DIN 13306, 2010)
Like any other business unit, maintenance has to make its contribution to reach the overall goal: the minimization of total operating costs. This means to find an optimum between the preventive maintenance costs and the costs of machine downtimes to guarantee a high machine reliability at the lowest costs possible. Since some of the cost drivers are not directly quantifiable, like the risk minimization, adherence to delivery dates and the preservation of product quality, this can be at times a difficult task. (Matyas, 2016, pp.32)

2.3 Maintenance strategies

Maintenance strategies are methods to reach the maintenance targets. The strategies define when and which maintenance actions take place how often and for which

machine. Before deciding upon a maintenance strategy, it is necessary to consider also legal, security relevant, technical and economical aspects. The choice of the right strategy influences not only heavily the reliability of technical plants, but also the costs for maintenance. Due to the fact that changes in the maintenance strategy have mid and long term consequences and are often overshadowed by other factors, the impact on cost reduction is not clearly visible. It is proven instead that there exist connections between the maintenance strategy and the downtime of machines and the utilization of technical equipment.
As seen in figure 2.1, according to the conduction time of the maintenance measures the DIN EN 13306:2010 specifies the following maintenance strategies.

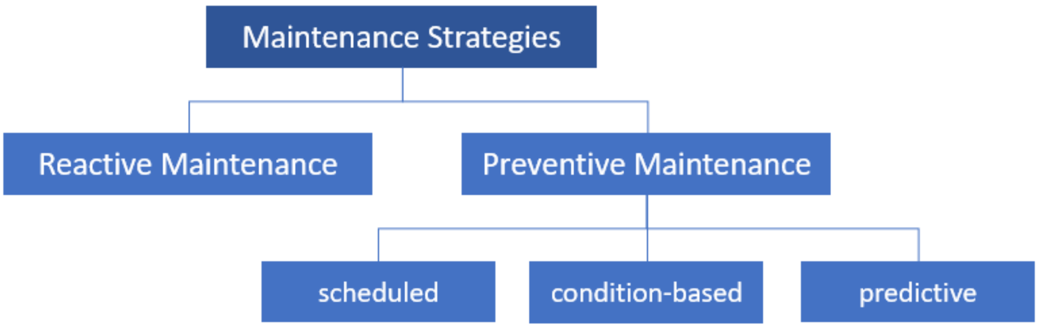

Figure 2.1: Maintenance strategies (Schenk, 2010, pp.27)

2.3.1 Reactive maintenance

This strategy is also known as corrective maintenance and takes place when a component or machine breakdown happened. No inspection or upkeep activities are performed. This approach is actually no strategy, since no planning is done before the breakdown. As soon as a failure is detected, the maintenance staff has to react quickly and spontaneously. Since often resources like spare parts, tooling and personnel are not immediately available, this approach can lead to long downtimes and is therefore the most cost-intensive approach. (Schenk, 2010, pp.23-28)

2.3.2 Preventive maintenance

Using this maintenance approach, constant inspections, replacements of components and overhauls of machines take place to prevent breakdowns and therefore unplanned downtimes. This strategy ensures a high machine reliability and availability. According to the time when maintenance activities are performed, the strategy can be divided into three sub-strategies:

Scheduled maintenance

All the maintenance measures are timed upon a fixed time schedule or a specified number of produced parts or used units. (Fredriksson et al., 2012, pp.29-30) The actual condition of the machine, component or tool is not taken into account. Maintenance actions can be planned beforehand and performed together with other maintenance activities when no production takes place. A drawback of this approach is that most of the time machine components are replaced too early and therefore, the expenditure of spare parts is relative high compared to other strategies. The maintenance interval has to be well chosen, to neither waste components nor risk unplanned downtimes. (Schenk, 2010, pp.28-29)

Condition-based maintenance

Compared to the scheduled maintenance strategy, the condition-based approach monitors either continuously, on request or at planned times machine performance and parameters. Based on this information, maintenance actions can take place when they are really necessary. (Fredriksson et al., 2012, pp.30) The monitoring can be performed by staff which inspects the machine manually or with the help of condition monitoring systems (CMS). The aim should be to monitor a maximum of components using as few sensors as possible to keep the investment costs for a CMS system low and to not introduce new potential points of failure. (Schenk, 2010, pp.30-31)

Predictive maintenance

This strategy is a development of the condition-based maintenance method. By analysing and evaluating the machine parameters derived through the monitoring process, a forecast is made when certain failures will happen. Based on this predictions, the maintenance measures are applied before the potential breakdown. (Fredriksson et al., 2012, pp.31)

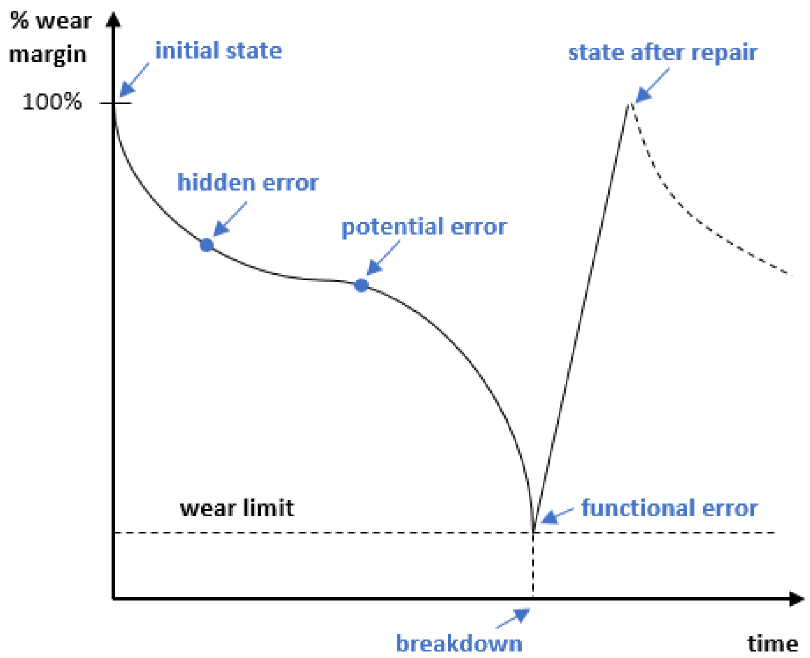

Figure 2.2: The wear curve (Matyas, 2016; Pawellek, 2013; Schenk, 2010, pp.34; pp.18; pp.146)

2.4 The wear curve

During its lifetime, a machine or system will undergo a certain wear. Initially, the new machine will have the full capacity, accuracy and safety, but due to wear which occurs over the service life, the unit will get less efficient, precise and safe. This wear follows a certain curve, which is depicted in figure 2.2. At a certain point in time an error occurs, but stays undiscovered since it has no influence on the machine's performance yet. After some time, the error starts to be observable on the machine. If the potential error is not fixed, it leads to functional failures and in the end to a machine breakdown. The goal is to detect errors as soon as possible by carrying out constant inspections. These checks have no influence on the service life, but if deviations in the machine's condition are observed, certain maintenance measurements can be executed. Constant upkeep actions will delay the machine's wear, but the initial state of the machine can be restored only with specific repair tasks. (Matyas, 2016; Pawellek, 2013; Schenk, 2010, pp.34-35; pp.18-19; pp.145-146)

2.5 Tool wear

The cutting, division and friction processes during the machine operation lead to high compressive stress, high cutting velocity and high temperatures. These mechanical and thermic loads act directly on the tooling and cause then the continuously progressing tool wear. Based on the different loads and tooling materials, the wear can have various forms and characteristics. According to (Zanger, 2013) there are four main mechanisms for tool wear:

- Surface breakdown is the damage accumulation of micro-contacts, which have to transmit periodical forces. Therefore, the root cause for surface breakdown are mechanical loads.
- Abrasion is also a mechanically dominated mechanism, in which softer particles are worn away by hard particles under relative movement of two contact objects.
- Adhesion instead is based on both mechanical and thermal processes. When friction partners touch each other directly, mechanical and thermal stresses occur on the surface roughness hills. There, high pressures and temperatures lead to so-called micro-welding, which is eliminated by relative movements, whereby near-surface volume particles are torn out or displaced from the adjacent bodies of the friction partners.
- Tribochemical reactions are caused by tribological stresses which lead to thermal activated chemical reactions. An example for this are diffusion processes.

The time which it takes from sharpening the tool until the attainment of the maximal wear criteria is called tool life. A quantitative, on the tooling measurable wear measurand can be such a criteria. If the tool wear is not directly, or only with a lot of effort measurable, then product features like the surface roughness are taken as criteria. (Hirsch, 2016, pp.16-17) For cutting processes can be generally said, that an increase in cutting speed causes rising temperatures at the touch points of the tool and the workpiece, which leads then to a drastic decrease of the tool's service life. (Haber Guerra et al., 2004)

2.5.1 Influence on product quality

Various researches show, that the machine tooling has a big influence on the product quality. Tool wear does therefore not only influence the surface, but also the geometrical and subsurface quality of a workpiece. These criteria then influence hugely the stress load. Most of the times the final surface roughness is taken as main quality criteria for workpieces resulting from cutting processes.
If worn out tools are used in production they cause thermal effects, which lead to an additional increase in shape deviations. Therefore it is necessary to react in time, change or reshape the tool and prevent so also damages on the machine. (Scheffer et al., 2003) Replacing worn out tools timely increases the product quality and reduces at the same time scrap rates. Since the expenses for the machine tools are a not to underestimate cost factor, it is crucial to find an optimal trade-off between tool wear and product quality, e.g. the surface quality. (Luan et al., 2018) The tooling costs do not only comprise the acquisition costs for the tooling, but also the replacement costs and the costs for the machine downtime. Maximizing the service life of a tool and prevent tool breakage is also directly related with the optimization of the entire machining process. (Haber Guerra et al., 2004)

2.5.2 Detection of tool wear

Unfortunately, there exists no single decision criterion which determines when a tool needs to be replaced or resharpened. Changes in the tooling geometry and catastrophic failures are the two most known and used criteria. There exist also other criteria like the sudden change of important process variables e.g. vibration and cutting force, the degradation of the tool surface, an increase in power consumption, scrap parts and when the machine starts to make abnormal noises. (Haber Guerra et al., 2004)
In order to detect tool wear based on changes in tooling geometry, important process variables, power consumption or the degradation of the tool surface it is crucial to monitor the relevant process continuously. Otherwise the, at times, sudden changes can not be detected. (Aliustaoglu et al., 2009)

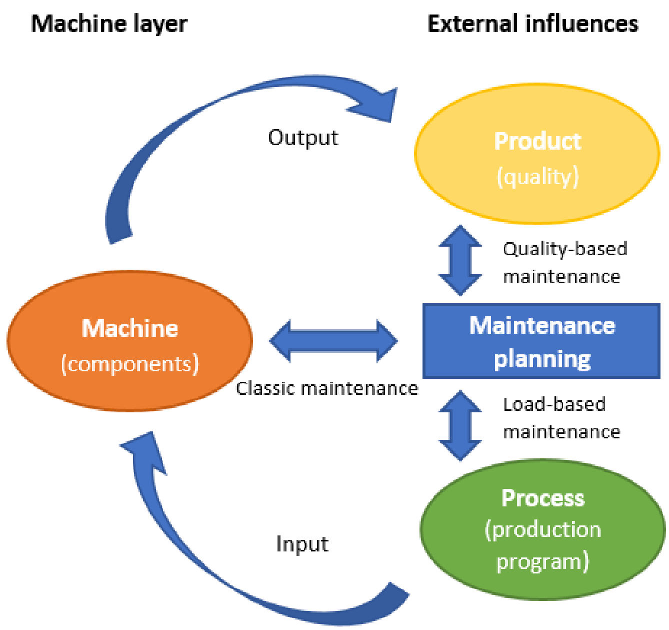

Figure 2.3: Holistic view for maintenance planning (Matyas, 2016, pp.140)

2.6 Smart maintenance and anticipative quality and maintenance planning

The creation of so-called cyber-physical systems (CPS), where the physical and the virtual world fuse together, is characteristic for the evolvement of Industry 4.0. This development makes it possible to connect all kinds of resources, information, humans and devices and creates a lot of new functionalities, services and capabilities. To fully exploit them it is necessary to change the way of thinking and operating also in maintenance.

Information about the current state of the machine are often late and incomplete and make it difficult to provide exact and timely optimal maintenance interventions which take also the actual production condition and quality into consideration. Requirement for an integrated, anticipative maintenance method is a holistic view over the three layers machine, product and process as can be seen in figure 2.3.

Correlations of machine condition data with quality data, monitored machine load and already known breakdown patterns can build a concrete decision basis for optimizing the timing of maintenance measures, product quality and energy consumption. Data mining is a quite useful tool to find correlations or behavioural patterns in the data and to predict possible, future machine breakdowns. The holistic, anticipative method makes predictions about the remaining service life based upon changes in the product quality. The challenge is to extract out of the measurements the relevant quality features. (Matyas, 2016, pp.138-141)

3 Condition monitoring

Condition monitoring is a technological evolvement which can simplify the maintenance task of inspection. Through the rise of Industry 4.0 it is now possible to monitor machines' remote, without human interference. Since it is difficult to find competent employees and personnel is always a huge cost factor, with condition monitoring it is possible to deploy the work force only where it is really necessary. In this chapter the general condition monitoring process and some of the most important diagnostic methods together with their applicability are described.

3.1 Condition monitoring process

Condition monitoring can be either done by humans and their natural senses or with the aid of sensors. Although the monitoring by humans is the simplest approach, it can work quite well when operators are able to hear changes in the machine's condition. Of course those observations are rather subjective and it is not possible to store the findings automatically. When the monitoring is done with the aid of sensors, the data acquisition can be continuous or discontinuous. The main difference between the offline and online approach is the data acquisition frequency. For the discontinuous method, also called offline system, a portable datalogger is used for a time period of four to six weeks. For the online, continuous system, sensors are firmly installed into the machine and therefore data about the unit's condition is permanently available. Using this approach it is also possible to detect possible failures very early and to create a trend analysis. (Matyas, 2016, pp.128-129)
Generally, the condition monitoring should be done during normal production

conditions and without stopping the machine. The schematic approach of a condition monitoring system is depicted in figure 3.1. (Schenk, 2010, pp.132-133)

The data transmission itself to the computer unit or the technician with a measurement device can be either tethered or wireless. Some possible wireless transmission systems for short distances are IrDA and Bluetooth, for a few hundred of meters WLAN is the preferred solution and for distances in kilometre range UMTS, GSM/GPRS and WiMAX should be used. (Pawellek, 2013, pp.157)

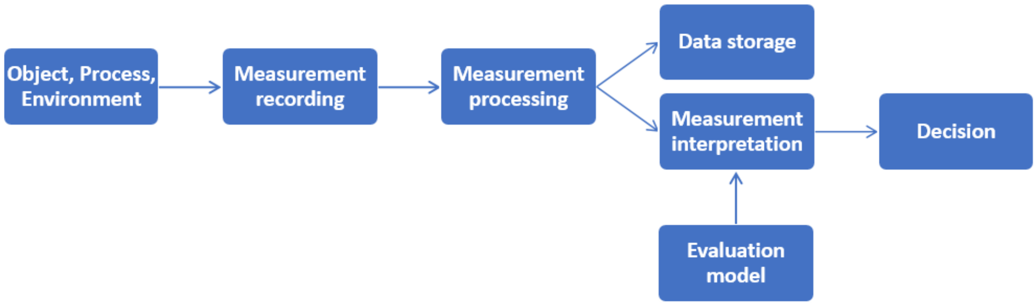

Figure 3.1: Schematic structure of a condition monitoring system (Schenk, 2010, pp.133)

3.2 Diagnostic methods

To determine a machine's condition, a broad variety of measurands can be taken into consideration. Therefore, for each machine the right parameters have to be chosen, respecting also technical and economical aspects. (Pawellek, 2013, pp.155) The goal should be to monitor with as few sensors as many components as possible. (Matyas, 2016, pp.131)

Figure 3.2 shows only a selection of the different measurands for detecting machine failures. In the following, some of the possible diagnostic methods are further explained.

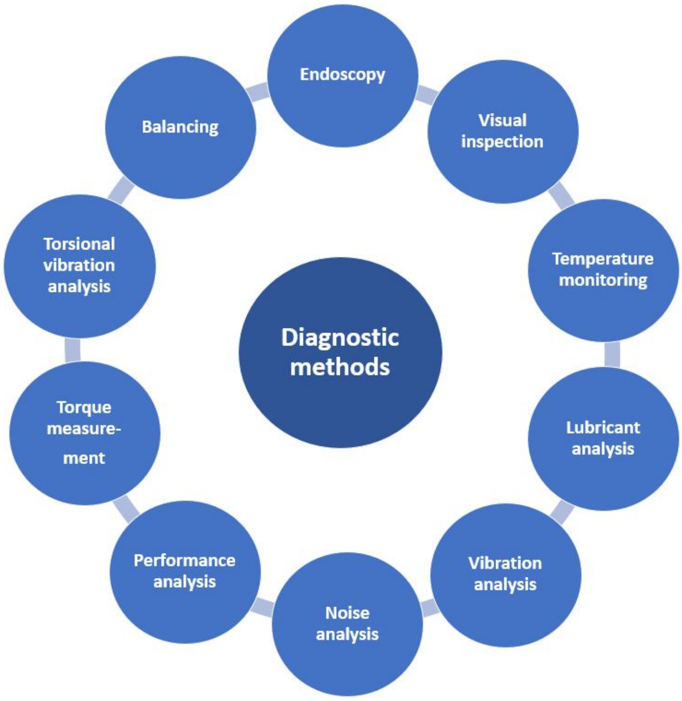

Figure 3.2: Possible diagnostic methods (Pawellek, 2013, pp.155)

3.2.1 Vibration

The basic concept for the vibration analysis is the fact that all mechanical operations in machines provoke power transmission processes which are then forwarded and eventually visible at the surface of the housing. These vibrations have a periodic character and therefore occur repeatedly at fixed intervals. Vibration analysis is mainly used to monitor the structure-borne sound, to detect mechanical loads in the form of unbalance, misalignment, striking or loose parts, fit issues, shaft damage, electrical effects, local and revolving gearing defects and rolling bearing malfunctions. Therefore, the vibration signals produced by the machine are recorded and examined for their composition from individual signals and their measured variables using modern analysis techniques. An example is the Fast Fourier transform (FFT) analysis for the conversion of a time signal into a frequency signal or the representation of spectra. It is then possible to detect characteristic frequencies, which do not occur during normal, undisturbed production. The effective values of the signals can serve as reference points for the degree of degradation. If additional, kinematic data like rotational speed, amount of rolling bearings or the teeth number of the gear stage is available, an allocation which component is damaged can be made. (Schenk, 2010, pp.133-134) Figure 3.3 depicts how a time waveform signal originating from a vibration sensor is transformed with the aid of FFT into spectral components. It is also shown how the resulting characteristic frequencies can then be related to the different machine components. (Coronando et al., 2015)

Despite modern analysis technologies and due to the many influencing factors that affect a technical system, the description of the damage processes still depends highly on the knowledge and experience of the operator or diagnostician. (Schenk, 2010, pp.133-134)

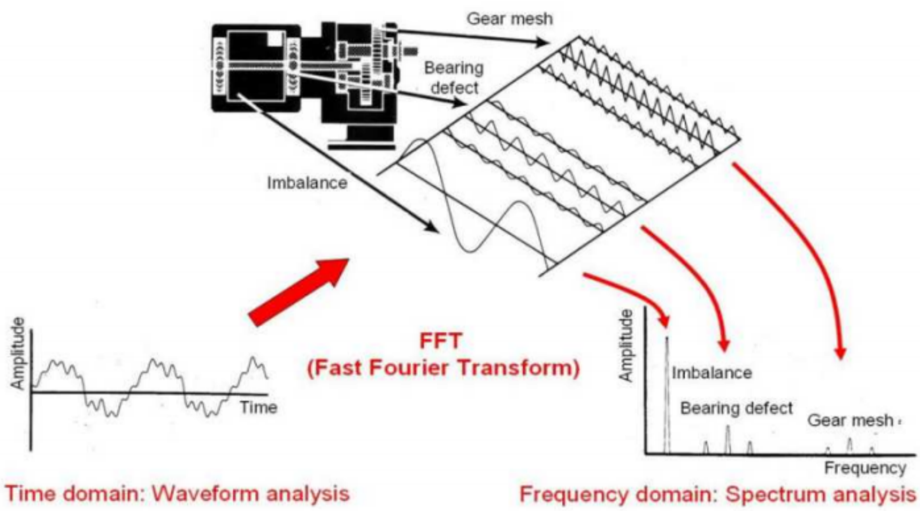

Figure 3.3: Vibration analysis method (Coronando et al., 2015)

3.2.2 Temperature

Not only vibrations and pressures lead to changes in a machine's condition, but also thermal load. In the classical approach, thermal sensors are used, which check and analyse the temperature constantly. Based on the height of temperatures and the duration of exposure, it is possible to derive statements about the condition changes. The determination of temperatures by using thermal sensors has the advantage that due to the small size of the sensors measurement is possible at almost any location and that specific points can be monitored effectively. However, for the temperature measurement a direct contact to the relevant object is necessary. Another drawback is that for monitoring large areas a high number of sensors is required and the evaluation of temperature propagation is difficult with this measuring principle.

Another approach is the method of thermography, which is a contactless measuring principle. It makes thermal radiation visible by using a thermal or infrared camera and is therefore the preferred solution when large areas have to be monitored. It is in interest for industrial applications because by using this principle temperatures can be represented surface-wise and also moving objects can be measured without contact. With the help of modern acquisition systems, the images from these cameras can be automatically evaluated, analysed and the results stored for other uses. A drawback of this method are the relatively high acquisition and operation costs and the limitations of use in places where many heat sources influence each other. The professional operation of the devices and the interpretation of the results require a great deal of experience of the users and are therefore limited to a few technical experts. (Schenk, 2010, pp.134-135)

3.2.3 Power & current

Power & current monitoring is preferably done for detecting flaws regarding the motor of a machine. With this method it is possible to detect problems like broken rotor bars, broken/cracked end rings, flow or machine output restrictions and machine misalignment. In a power lead at the motor starter or control centre have to be firmly installed sensors, which measure the current flow. The variations in the current flow signalise changes in the machine's condition. Through recording the sensor data and performing a trend analysis maintenance measurements can be scheduled accordingly. Motor defects can also discovered by comparing lines with motor frequencies. (American Bureau of Shipping, 2016, pp.76-77) An advantage of this approach is that in most cases it is possible to access the current data via the digital controller. Only when this is not possible, additional sensors have to be installed. Also for this method expert knowledge is required to properly interpret the data. (Pawellek, 2013, pp.159)

3.2.4 Noise

This is one of the most often used monitoring techniques in industry since a broad selection of potential errors can be revealed. Only by listening to a machinery worn bearings, steam or coupling leaks, pressure reliefs, excessive loads, misalignments of equipment etc. can be detected. Changes in a machine's condition can be identified in this method by practised and experienced employees, since for the human hearing it is really easy to detect new or changed noises. As support system, also small microphones can be installed. In high noise areas this monitoring method is nearly impossible, additionally, the reaction time for maintenance activities is short, since errors are detected rather late. (American Bureau of Shipping, 2016, pp.79-80)

3.2.5 Oil

Different types of oil, like lubrication, hydraulic and electrical insulation oils, can be examined with an oil analysis. Through oil monitoring statements about the machine degradation, the oil contamination, consistency and deterioration can be made. There exists no general guideline regarding the sampling frequency, only that examinations should be done regularly to provide valuable results. For determining an analysis frequency a good basis is to start with the machine manufacturer's recommendations, the criticality, risk factors and historical machine and equipment data.
Oil has three aspects: the lubricant condition, the contaminants and machine wear. By checking the lubricant condition maintenance can decide if the oil has to be replaced, filtered, dewatered or can still be used. Fluids and particles which come from the surrounding environment and enter into the oil are considered as contaminants. A high oil pollution can lead to massive machine wear, therefore, monitoring the contamination and taking maintenance measurements if necessary is an important aspect. Particles originating from the machine components themselves are referred as machine wear. By analysing the particles, their provenance can be determined and

decisions be made if the relevant components should be either replaced or serviced. The oil analysis can be done either inhouse with portable equipment or sensors, or by sending the samples to external laboratories. When sourcing the competence out to a contractor, it is really important to keep the time aspect in mind, so that in case maintenance activities can be scheduled in time. (American Bureau of Shipping, 2016, pp.44-45)
With regular oil monitoring possible failures can be detected in early stages as well as it simplifies to find complex failure correlations. Drawbacks are that performing oil analysis are rather expensive and that expert knowledge is required to interpret the results. (Pawellek, 2013, pp.159)

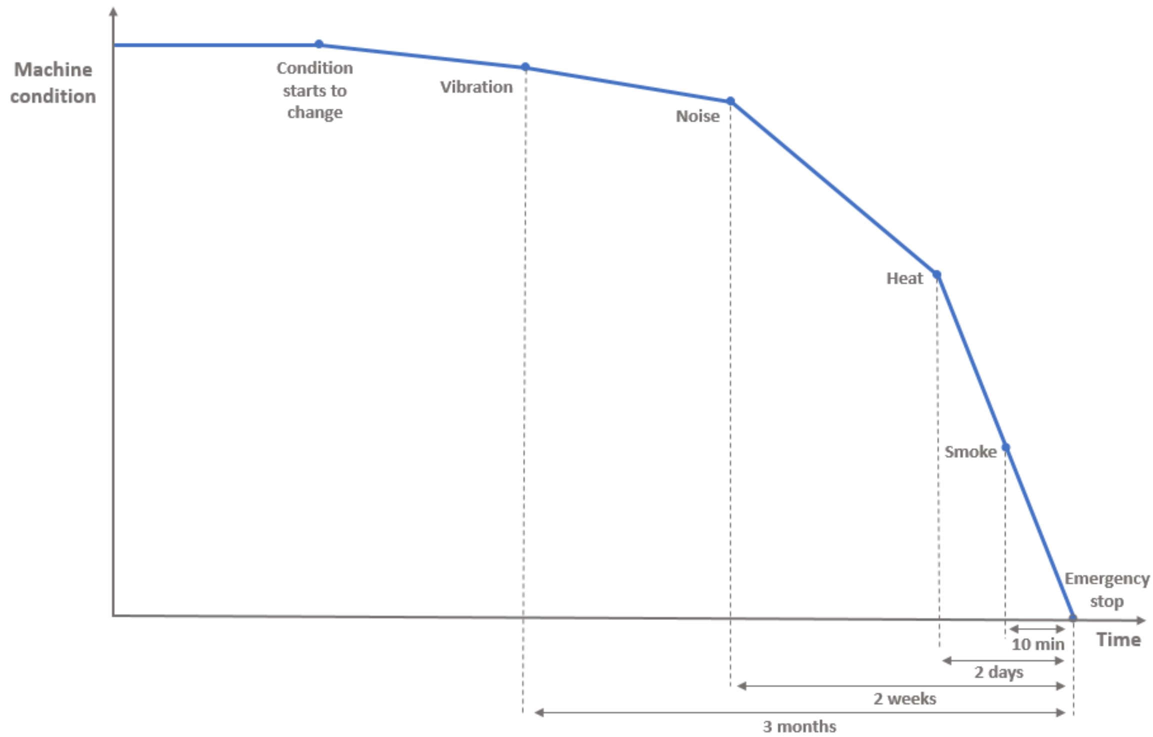

Figure 3.4: The warning signs of a machine failure (National Instruments, 2019a)

3.3 Applicability

Figure 3.4 shows the general development of a machine's condition over time and when and which warning signs occur until the machine breakdown. Vibrations are observable shortly after a condition change and months before the actual breakdown happens. The next warning signals which arise about two weeks before the malfunction are noises, followed by temperature and smoke. The latter two are detectable only days or hours before the breakdown and are therefore not well suited for a preventive strategy, since the time horizon for maintenance activities is too small. (National Instruments, 2019a)

3.4 Goal

The main goal of condition monitoring, and technical diagnosis in general, is to detect failures as soon as possible before malfunctions or damages happen. Otherwise, this can lead to high subsequent costs. In addition, the service life and the availability of the machine can be increased and condition monitoring is one of the basic building blocks for the condition-based maintenance. (Pawellek, 2013, pp.154)

4 Big data

Nowadays it is possible, through the developments in storage capacity and data collection, to generate, collect and store huge amounts of data. Because of the increase in connected devices and sensors more and more data is created. The data has to be efficiently handled, stored and furthermore also analysed, to extract useful knowledge. To provide this, new methods of data handling, storage and analysis had to be created and are explained in the following chapter. (Elgendy et al., 2014)

4.1 Characteristics

Data sets which are generated quickly and grow very fast are generally referred to as big data. Big data is characterised by the 4 V's: volume, variety, velocity and veracity. With volume the pure size of the data is described, normally a big data set size is specified in Terabytes (TB) or Petabytes (PB). Big data does not originate from only one data source, but from quite a number of different resources like clickstreams, sensors, logs and social media. Therefore, to the normally structured data, semi-structured data, like Rich Site Summary (RSS) feeds and eXtensible Markup Language (XML), and unstructured data, originating from audio and video, is added. The sheer number of different data sources and included data types and formats is referred as variety. Velocity describes the speed or frequency of data generation or change. The fourth V, veracity, discusses the quality of the data. It can either be good, bad or, because of incompleteness, deception, inconsistencies, approximations, latency and ambiguity, undefined. (Elgendy et al., 2014)

4.2 Data acquisition process

The process of capturing electrical or physical phenomena like temperature, pressure, vibration, is called data acquisition. Nowadays there exist two main approaches for data acquisition: a PC-based and PLC-based method. A typical PC-based data acquisition system has three main components: sensors, a data capturing device and a computer with the necessary software. Figure 4.1 shows this PC-based data acquisition process. The PC-based concept is not only cheap, but provides also a high performance. Drawbacks are that this method requires an external signal conditioning, is poorly expandable and the connection to sensors is sometimes difficult.
In the Programmable-Logic-Controller(PLC)-based approach no extra PC is needed, because the controller and its I/O are able to do the sensing, signal conditioning, measuring and analysing. Since the PLC is also used to control also the machine and the process, this method is the simplest and also most cost effective. Some basic data acquisition is already done because of the PLCs control tasks, if there is the need for specific data it can be easily gathered by adding extra I/O devices e.g. sensors. The PLC is also able to log or store the data locally, but also transferring it via Ethernet to other systems is no problem. (National Instruments, 2019b; Payne, 2013)

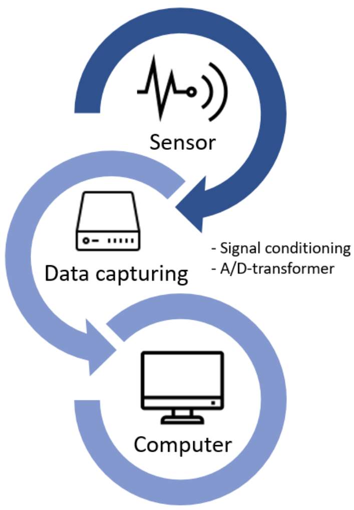

Figure 4.1: The PC-based data acquisition process for a PC-based (National Instruments, 2019b)

4.2.1 Sensors

The measurement of a physical phenomenon, e.g. temperature in a room, intensity of a light source, or the applied force on an object, starts always with a sensor. It is possible to transform the physical happenings into measurable electrical signals, therefore, sensors are also called transducers. The type of the sensor defines if the outcome is a tension, current, resistance or electrical value, which changes over time. Some sensors require additional components and circuits to correctly create a signal, which can then be precisely read from a data capturing device.

4.2.2 Data capturing device

The hardware interface between a computer and the signals of the environment is called data capturing device. The main function of such a device is to convert the ingoing analogue signals into digital ones so that a computer is able to process and analyse the data. The three main functions of a data capturing device are the circuits for the signal conditioning, the analogue-digital converter and the computerbus.

Signal conditioning

Some signals from sensors or the environment are sometimes too loud or too dangerous for a direct measurement. The circuits for the signal conditioning transform the signals so that they are suitable as input for the analog-digital converter. These circuits can comprise reinforcement, absorption, filtration and isolation.

Analogue digital converter

Analogue signals from sensors have to be converted into digital signals, otherwise, other digital devices like a computer are not able to process the data. An analogue-digital converter is a chip, which provides a digital representation of an analogue signal at a specific point in time. Normally, an analogue signal is changing constantly over time and an A/D-converter captures the samples with a specified frequency.

These samples are then transferred over a bus system to a computer, which reconstructs then the original signal with a software.

Computerbus

Data capturing devices use a computerbus system to communicate with the computer and to transfer relevant measurement data and instructions. Supported are common bus systems like USB, PCI, PCI Express and Ethernet, but also wireless solutions can be used. The market offers a lot of different bus systems, which all have different advantages for varying applications.

4.2.3 Computer

The computer is managing the operation of the data capturing device and is used for the processing, visualisation, analysis and storage of the measurement data. Driver software ensures that the application software can communicate with the data capturing device, mostly using an application programming interface (API) and certain data transfer protocols. The application software instead facilitates the interaction between the user and the computer for data processing, visualization and analysis. This software can either be a prefabricated application with a fixed functionality or a programming environment where customized functions can be created by the user. (National Instruments, 2019b)

4.3 Data ingestion

All activities regarding the process of loading, processing, transferring and transforming data originating from various data sources with different sampling rates is referred as data ingestion. During the process, data is imported and loaded from various sources, and then transformed, converted or formatted to fulfil specific needs for analytics or storage reasons. Normally, this is done by using multiple data transport protocols to support a broad range of data sources. Data can be ingested in real-time or in batches, the latter means to import data in chunks at certain intervals. Real-time ingestion instead loads data as soon as it is produced by the source. In the end, the data is stored in a data storage system or directly ingested in data analytics process. (PAT Research, 2018)

4.4 Storage and management

Due to the different characteristics of big data, new ways of data storage and management had to be created. Classical methods of data storage comprise data warehouses, relational databases and data marts. Before uploading the data into such a traditional storage, the data has to be processed according to the Extract, Transform, Load (ETL) or Extract, Load, Transform (ELT) principles. Therefore, the data needs to be extracted from the different sources, transformed according to the needs and then loaded into the storage system. Following these principles guarantees that the data is transformed, cleaned and catalogued before operations are executed upon it.
Big data instead has other requirements for storage systems and follows the Magnetic, Agile, Deep (MAD) principle. The magnetic aspect indicates that not only already cleaned data with good quality is integrated into the storage. Furthermore, big data storages should allow to easily produce and adapt data, as well as to provide the possibility to be used as algorithmic runtime engine. Developed storage solutions for big data comprise methods from in-memory or non-relational databases to distributed systems and Massive Parallel Processing (MPP) databases. (Elgendy et al., 2014)

4.4.1 Non-relational databases

To handle unstructured or non-relational data, specific database systems were developed. An example for that kind of databases is Not Only SQL (NoSQL), which provides flexible data models, high scaling functionalities and a simple development and deployment of applications. NoSQL databases divide data storage and management and the focus lies on high performance and scalability. (Elgendy et al., 2014)

4.4.2 In-memory database

In in-memory databases the data is handled in the server memory, therefore, disk input/output (I/O) is not needed anymore and real-time responses originating directly from the database are possible. The database is stored in the main memory and not on a disk drive, has therefore a higher performance and enables the development of totally new applications. (Elgendy et al., 2014)

4.4.3 Hadoop

Another popular, alternative storage system for big data is the Hadoop framework. It combines the big data storage and analytics parts and is therefore a reliable, scalable and easily manageable solution. For the storage functionality, the Hadoop distributed file system (HDFS) is used, which is optimized for large files, and provides redundancy as well as reliability. The basic architecture of a HDFS framework is shown in figure 4.2. HDFS works by splitting single files into blocks and distributing them then across cluster nodes. To provide also a high reliability and availability, replication is used. The nodes used in a HDFS can be divided into Data Nodes and Name Nodes. Data Nodes contain the file blocks with the data, while the Name Nodes direct the client to the specific Data Nodes, where the requested data is stored.

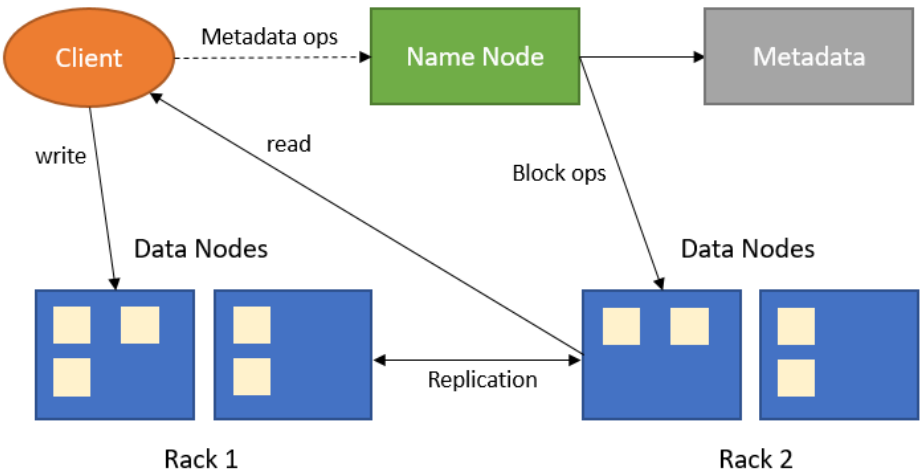

Figure 4.2: HDFS Architecture (EDUCBA, 2019)

Instead for the big data processing and analytics is the MapReduce paradigm used, which is depicted in figure 4.3. Instead of increasing the storage capacity or power of a single unit, MapReduce focuses on working in parallel with multiple computers. The idea is to divide a specific task into multiple smaller tasks and to execute them then in parallel. This approach speeds up the entire computation process. The first step is the "Map" function, which maps input values to keys. These key/value pairs are then shuffled and sorted in an intermediate step, which groups all keys and creates a list of the according values. The "Reduce" function takes the map with the key and the list of values and aggregates then the different value lists and performs computations on it. (Elgendy et al., 2014)

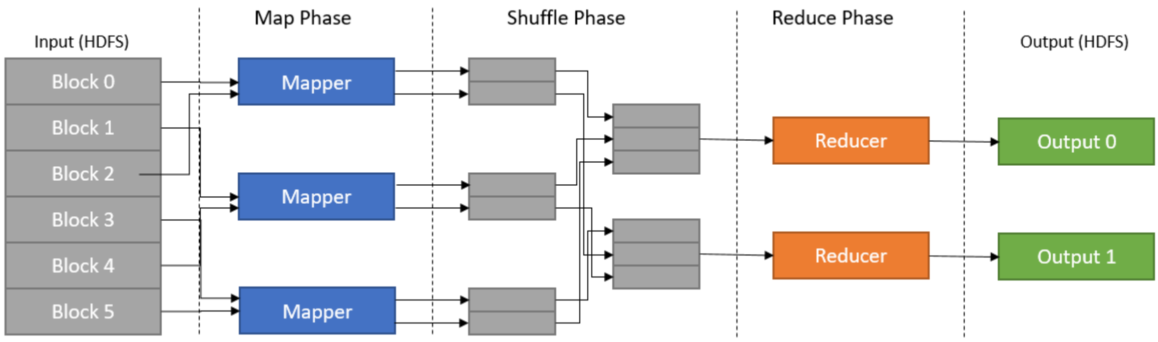

Figure 4.3: The MapReduce Framework (EDUCBA, 2019)

4.5 Knowledge discovery in databases

All activities of retrieving valuable information out of big data are known as knowledge discovery in databases (KDD). Due to the nowadays huge amounts of data, humans are not anymore able to process the data and extract knowledge from it. With the help of KDD potentially useful information and patterns can be detected. Therefore, KDD is an indispensable process when working with big data. Figure 4.4 shows the different steps of a knowledge discovery process, which are also further explained in the following subsections. (Kayaalp et al., 2018)

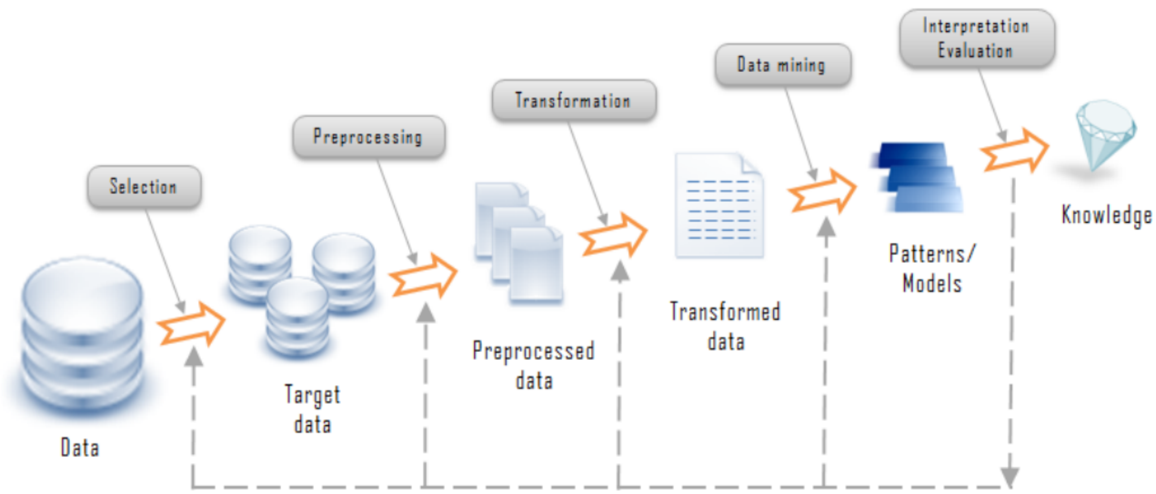

Figure 4.4: The knowledge discovery process (Kayaalp et al., 2018)

4.5.1 Selection

Before data can be processed, transformed and analysed the relevant data has to be determined, selected and extracted. What data is chosen depends always upon the specific use case and analysis task. The selection can be done by a domain expert, who has knowledge about the specific use case. This task can take a lot of time and effort when the specific knowledge about the data is not in place. The selection is considered as crucial, because when keeping irrelevant or redundant data or leaving out relevant data, the obtained patterns can be distorted. Another factor which has to be considered is the fact that too many data attributes slow down the analysis process. Therefore, the overall goal is to select the optimal number of data attributes, which provide a valuable but also easily understandable pattern. To determine and select this optimum number of data attributes, specific heuristic methods e.g. forward selection, backward elimination, decision tree induction, are often used. (Han et al., 2012, pp.8, 103-104)

4.5.2 Preprocessing

Due to the fact that big data comes from multiple different heterogeneous data sources, the data can be inconsistent, incomplete and noisy. These factors can influence the efficiency and accuracy of the data mining process and therefore, the data quality has to be improved by performing certain activities to remove and clean the inconsistencies. Data preprocessing can be divided into multiple subtasks:

- **Data cleaning** tries to smooth noises in data, clear inconsistencies, detect and remove the outliers and fill in missing values.
- **Data integration** instead focuses on including data originating from multiple, different sources. Redundancies have to be removed and inconsistencies e.g. in the naming scheme resolved.
- **Data reduction** is used when a data set is rather huge and therefore the mining process slow. This task tries to reduces the size of the data set while obtaining the same analytical results. Popular strategies for data reduction include numerosity reduction and dimensionality reduction.

4.5.3 Transformation

Transforming the data into a certain format should make the mining process more efficient and the found patterns easier to understand. There exist various methods for data transformation:

- **Smoothing** is used to remove noise from data with strategies like clustering, binning, regression.
- **Attribute construction**, also called feature construction, tries to create new data attributes out of given data sets.
- **Aggregation**, which applies summarization and aggregation operations on the data.
- **Normalization** scales the attribute data on a smaller range e.g. 0.0 to 1.0.
- **Discretization** replaces the values of numeric attributes like the age with interval labels e.g. 30-40 or conceptual labels e.g. adult.

- **Concept hierarchy generation for nominal data** generalizes specific attributes to concepts on a higher level. (Han et al., 2012, pp.82-87, 111-112)

4.5.4 Data mining

After the data is prepared and preprocessed, the essential step for a KDD process can take place. Through data mining patterns in the data can be detected with the aid of smart methods. Data mining functionalities describe the types of patterns which can be identified with data mining. In general, such functionalities can be either descriptive or predictive. Descriptive functionalities characterize specific data, while predictive methods search for general statements in data and make predictions based on patterns. The different data mining functionalities are further described in the following paragraphs.

Detection of frequent patterns, correlations and associations

This method searches in the data for frequently occurring patterns, associations and correlations. Frequent item sets, subsequences and substructures are some types of the frequent patterns which can be found in data sets. By identifying such patterns, interesting correlations and associations can be discovered too. At times it can also lead to interesting results when correlations between associated data are examined. Figure 4.5 shows an example for a) a positive correlation and b) a negative correlation in the data sets, while in figure 4.6 no correlations between the features are observable.

Characterization and discrimination

These functionalities are descriptive methods since they provide class or concept descriptions. Generally, there exist associations between data entries and concepts or classes. At times it can be useful to have a summarized, term-wise but concise description of classes or concepts. Characterization is one of the possible methods to derive such a description by summarizing in general terms the features and

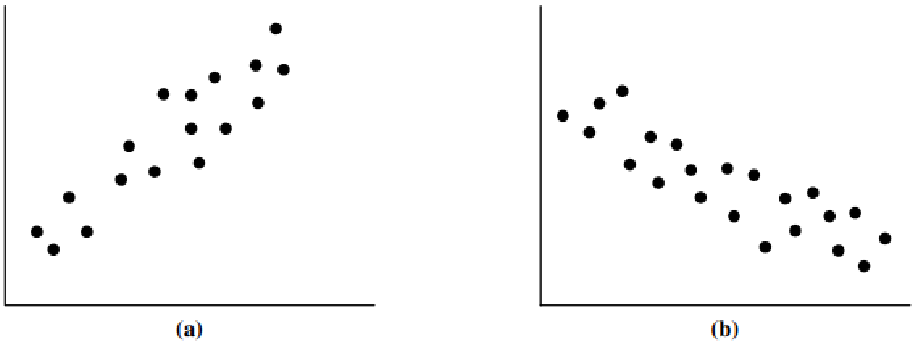

Figure 4.5: An example for a) positive and b) negative correlation between features (Han et al., 2012)

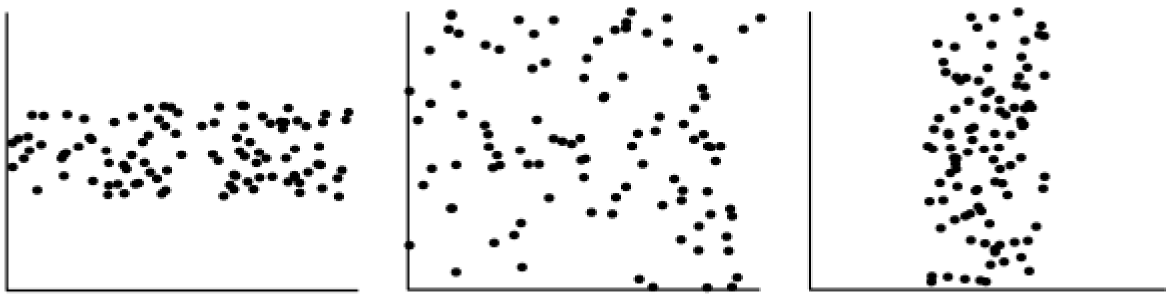

Figure 4.6: An example for three data sets where no correlations can be observed (Han et al., 2012)

characteristics of the class of data.
Another method is data discrimination, where the main features of contrasting classes were compared. Results of a characterization and discrimination can be visualized as pie charts, curves and multidimensional tables.

Classification and regression

The process of finding a function or model which best describes data concepts or classes and is able to distinguish between them is called classification. In order to make classifications a so-called training data set is needed, where the class labels of the data entries are already known. The classification model is created by analysing the training data set. Based on this function or model, class labels can be predicted for data objects with unknown class labels. A classification model can be represented as classification rules, mathematical formulae, decision trees or neural networks.

Regression instead is able to model continuous-valued functions, which means to make predictions about missing numerical values. This method can also be used to identify distribution trends on the basis of available data.
For both methods, classification and regression, it may be necessary to perform beforehand a relevance analysis. By doing so the relevant attributes for the classification and regression process are determined and irrelevant features are neglected.

Clustering analysis

In contrast to classification and regression, the clustering method does not try to predict class labels but can be used to create class labels. Clustering always follows the principle of minimizing the interclass similarity and maximizing the intraclass similarity. Therefore, objects within a cluster are highly similar, but the similarity between objects of different clusters is low. Clusters generated according to this principle are also called classes.

Outlier detection and analysis

Due to the veracity characteristic of big data it is normal that a data set contains also values which are not compliant with the general tendency of the data. These non-compliant data objects are referred to as outliers. Most of the data mining algorithms neglect these outliers and see them as exceptions or noise, because they can falsify the outcomes. Nevertheless, in some cases a further look into such outliers can be more valuable than analysing the normal data behaviour. Such examinations are called outlier analysis or anomaly mining. (Han et al., 2012, pp. 8, 15-21)

4.5.5 Interpretation, evaluation & presentation

With data mining potentially can be found thousands of different patterns in the data. Therefore, the last step of the knowledge discovery process is to identify the really relevant and valuable patterns and to represent then the mined knowledge accordingly with certain techniques.

Generally, only a small fraction of the detected patterns is really interesting and provides value. There exist several different measures which indicate if a pattern is valuable or not. The pattern has to be:

- easily understandable for humans,
- potentially useful,
- novel,
- and valid on test or new data.

Patterns can also be valuable if they validate a hypothesis which a user wanted to confirm. Such interesting patterns always contain knowledge.
With the aid of data visualization it is possible to represent data in a way that relationships are easily detectable. There exist certain techniques like pixel-oriented, geometric-based, hierarchical or icon-based. (Han et al., 2012, pp.8, 21, 55-56)

5 Gear manufacturing

The requirements for a gear vary depending on the application. Gears can be manufactured in various sizes and quantities, which is why different manufacturing processes are used, which differ with regard to the achievable manufacturing quality or productivity. In automotive industry, gears have to be produced in large series with high quality. In the past decades, these requirements led to further development of the technology. Special machines and processes evolved to meet the needs of the automotive industry, offering high productivity and good manufacturing quality at the same time. The component costs vary from a few cents up to a unit price of several thousand euros.

5.1 Production methods

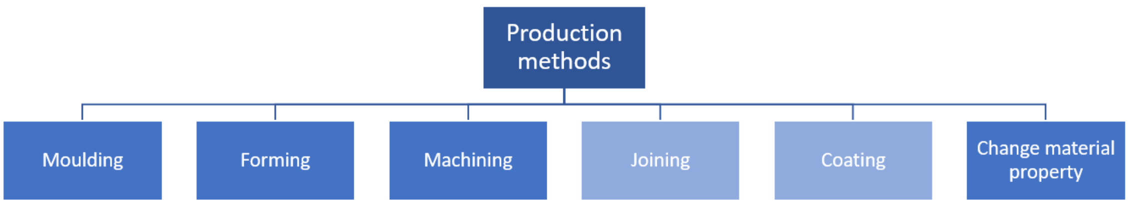

Figure 5.1: Production methods (Klocke et al., 2016, pp.160)

All production methods depicted in figure 5.1, except joining and coating processes, can be used for the gear manufacturing process.
In moulding and forming processes, the entire workpiece is projected in one form. In each work cycle, the mould is filled with an amorphous raw material (e.g. melt, powder) which solidifies into a gear contour in the die or is formed into a gear

contour. Casting, forging and sintering count to the moulding or forming processes. The machining or separating processes play a dominant role in the manufacturing of high performance gears. Hobbing is one of the most productive soft machining processes. Other processes include shape milling, gear shaping or shaving, and broaching. Hard finishing methods with undefined cutting edges, such as profile or gear grinding, are also among the machining processes. Depending on the process, there are advantages and disadvantages in the applicability, productivity and achievable component quality of the various processes.
Gears which have to endure high stress are heat-treated. The process of heat-treatment is used to change the material properties of the component and belong therefore to the last production method visible in figure 5.1. Heat treatment by case hardening, nitriding or quenching and tempering, for example, changes not only the component strength but also the geometry of the workpiece. In many cases, case hardening reduces the component quality due to hardening distortions. For this reason, gears with the highest quality requirements are hard-finished after heat treatment. (Klocke et al., 2016, pp.159-160)

5.1.1 Honing

According to (Klocke et al., 2016, pp.159-162) the hard finishing process determines mostly the final surface quality of the product. As (Klink, 2015, pp.207) mentioned that the hard finishing manufacturing process which provides the best surface quality is the honing process, in the following, this specific production step is further explained.

Honing methods

Honing is a cutting production process and belongs to the superfinishing methods, with which it is possible to generate a high surface quality. This leads then to an improvement in the dimensional and shape accuracy. As can be seen in figure 5.2, the honing process can be divided according to the kinematics of the movement into long stroke honing, formerly draw grinding, and short stroke honing, often also called superfinishing. A special honing method is the gear honing for the surface finishing of gears. (Klink, 2015, pp.1-6) Long-stroke honing machines are used for the production of cylinder blocks and liners, connecting rods, hydraulic valves, compressors and tax valves, while short-stroke honing machines come into operation when other production machines are not able to provide the necessary surface quality and shape accuracy. The main application areas are the production of rolling bears, motors and gearboxes. (Perovic, 2009, pp. 249, 256-257)

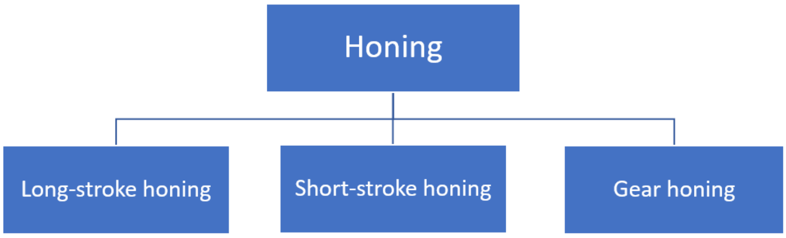

Figure 5.2: Overview of honing methods (Klink, 2015, pp.6)

Honing process

Honing is a machining process with geometrically indeterminate cutting edges in which one component of the cutting speed performs an oscillating motion. The cutting speed consists of two separate speeds: the peripheral speed of the workpiece and the oscillation movement with the speed of the axial stroke. By superimposing the two cutting speed components, a high quality surface with intersecting machining tracks is achieved. The oscillating movement with the speed of the axial stroke is carried out either by the honing stone or by the workpiece. Since honing is carried out with small oscillation amplitudes, an axial feed movement is initiated at the feed speed so that the workpiece can be machined over its entire length. During machining, the required pressure is achieved through the contact pressure of the honing stone. (Perovic, 2009, pp. 249, 256-257)

5.2 Quality measurement

The design requirements for gear units with regard to the power density to be transmitted are increasing. In smaller installation spaces, always higher power ratings are required, with a longer service life of the gear unit and low noise emission. Due to that, a smooth running, accurate angular transmission of the rotary motion and the required load capacity of gears must be ensured. For this purpose, the deviations of all the parameters of the gears must be maintained within certain tolerances. In order to meet these requirements, it is essential to test and analyse the quality-determining, production-related component properties before the final assembly of the gearbox. The operating behaviour of a gear is not only dependant from its macro-geometry and the surface roughness, also called micro-geometry, but also from the physical and chemical properties of the material. All three aspects have to be considered for evaluating the quality of a gear. Meeting the requirements is necessary in order to ensure the load-bearing capacity and to detect and prevent a potential, cost-intensive failure of a gearbox due to damage of the gear during the manufacturing process.

5.2.1 Geometric measurement

During the geometric inspection of gearings, in addition to compliance with the known tolerances, such as form, metrological measurands specially defined for gearings are also checked. In addition, the shape deviations of the surface are recorded, especially irregularities, roughness and waviness. Geometric measurement can be divided into:

- Macro-geometric measurements are used to measure the dimensional and form deviations. They can be further divided into individual and cumulative defects. Important characteristics are the tooth thickness, profile and flank line deviation as well as the pitch and concentricity deviation.
- Micro-geometric measurements, whose task it is to describe the three-dimensional geometry topographically. To limit the effort, often only characteristic values for the surface roughness are used.

5.2.2 Metallographic analysis

The physical and chemical properties of the material in the core and near-surface edge area also influence the quality of gearings. In industrial practice, a distinction is made between non-destructive and destructive processes. Methods for the metallographic analysis are crack examinations, visual and ultrasonic inspections as well as tensile and hardness testing. (Klocke et al., 2016, pp.283-284,300,307-309)

6 Use Case at GKN Driveline Bruneck

This chapter provides a short overview of the GKN Group as well as GKN Driveline Bruneck, were the following use case was executed. The described production shift and the new requirements which have to be challenged, led to the following use case, where the influence of the honing super-finishing process on the End-of-Line (EOL) results was further investigated. In the process, a study was made to check if it is possible to make a statement about the product quality based on the vibration data during the honing process. To continuously monitor and analyse the occurring vibrations and other machine data, a connection between the machine and a HDFS was built. This data ingestion process is also described in the following chapter, as well as the gear manufacturing process and the honing machine itself. All experiments, the executed analyses as well as the obtained results and findings, are collected in this chapter.

6.1 The GKN Group

The GKN plc group is a world leading manufacturing company listed on the London stock exchange. Founded in 1759 as iron and steel work in Wales, it developed to a global player for aerospace, automotive, agricultural and military vehicle components. In 2017, the company employed 58,000 people overall in 30 countries and had a revenue of £9,671 billion.

The GKN group is divided into 5 main divisions:

- Aerospace: is the biggest division and one of the world's best known aerospace tier 1 suppliers, serving over 90% of all aircraft and engine manufacturers, with over 50 plants in 14 countries.
- Automotive: which is the market leader in the field of contemporary and electrified driveline systems. It consists of two subdivisions:

  Driveline: is a supplier for automotive driveline components and products.

  ePowertrain: provides all-wheel and electrified driveline systems.
- Powder Metallurgy: offers high precision metal products, which can be integrated into automotive and industrial systems.
- Off-Highway Powertrain: is a global tier 1 supplier for agricultural and off-highway systems and components.
- Wheels & Structures: supplies off-highway wheels and structural assemblies. (Melrose Industries PLC, 2019)

6.1.1 GKN Driveline Bruneck

Initially, GKN Driveline Bruneck was founded 1963 as Birfield Trasmissioni S.p.A. and produced drive and double-joint shafts. After GKN took over the company in 1970, they started to produce also cardan shafts, universal joints and visco couplings. Shortly after the millennial change followed the first movements towards more complex technologic drive solutions like the electronic torque management (ETM). 2011 was the year in which the first electrified driveline systems were developed and serially produced at GKN Driveline Bruneck. In 2013 the production of the first gearbox for the electric engine for the well known hybrid sports car BMW i8 started. Since 2018, GKN Driveline Bruneck is part of the GKN ePowertrain division, which has sealed the decision to be a leader for electrified driveline systems in the future.

GKN Driveline Bruneck currently employs 663 people and had a revenue of 167.7 million € in the year 2017. (GKN Driveline Bruneck, 2019)

6.2 Noise, vibration and harshness analysis

Due to the product shift in GKN Driveline Bruneck from simple driveshaft components to complex, electrified powertrain systems, the company has to face new challenges in the production process. One of those challenges are the increased noise, vibration and harshness (NVH) drivetrain requirements for full electric (EV) and hybrid vehicles (HEV). (Kotthoff, 2018) Internal combustion engines provide a higher noise masking, in electrified vehicles this masking is reduced. Therefore gear whine can be heard when the gear quality is not optimal. (Wilson, 2015)

Since the main root cause for NVH conspicuous gearboxes are gears with a high surface roughness, it is crucial to take a further look into the gear manufacturing process. Gears for gearboxes require a target-oriented, i.e. economically and technically coordinated choice of gear geometry. The demand for an optimum application behaviour of the gear geometry is countered by the necessity of cost-effective production. (Klocke et al., 2016, pp.11)

6.3 Gear manufacturing

Due to the fact that NVH is mostly caused by higher roughnesses or irregularities on the gear surface, in the following chapter the gear manufacturing process at GKN Driveline Bruneck is further described. The single production steps are mentioned and the honing production process, which influences the surface quality the most, is explained in more detail. The different quality measurements which are executed for the gear production are aggregated at the end of the chapter.

6.3.1 Manufacturing process at GKN Driveline Bruneck

Different process chains are used for gear production depending on the required profile. While choosing the optimal production process, it must be taken into account that the individual process steps along the chain influence each other. The achievable gear quality is the result of the entire process chain. Figure 6.1 shows a selection of common process chains in gear manufacturing and, marked in blue, the process chain which is applied at GKN Driveline Bruneck. In the following, these relevant process steps are further explained.

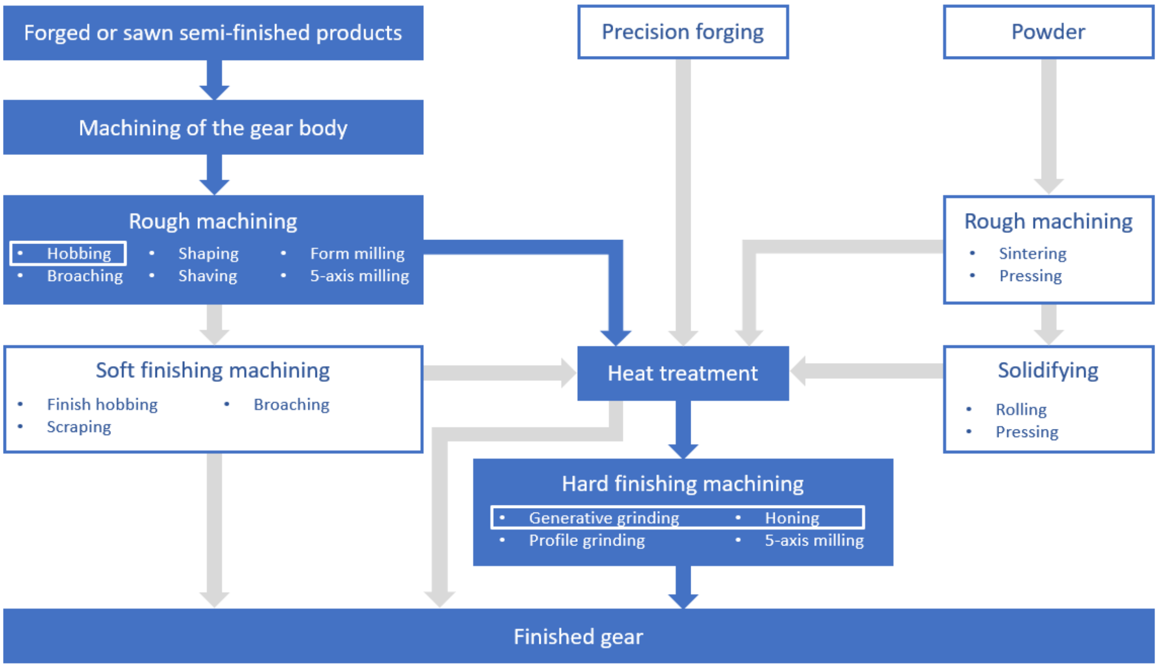

Figure 6.1: Gear manufacturing process chain (Klocke et al., 2016, pp.162)

To start with the gear manufacturing process, already forged and turned semi-finished products are bought from suppliers. The production flow follows the conventional process chain, since the gear produced is an intermediate shaft and has to endure high loads. Therefore, the next step is to tooth the turned blank. Afterwards, the workpiece is case hardened to increase the strength to ensure sufficient load-bearing capacity and wear resistance. Due to the distorting resulting from the hardening process, in the next step hard finishing is carried out to achieve the required geometric accuracy. According to the requirements, the hard finishing process is either generative grinding or honing. This manufacturing process causes rather high logistic effort and costs, but ensures a good final surface quality. (Klocke et al., 2016, pp.159-162)

6.3.2 Gear honing

Since one of the core competences of GKN Driveline Bruneck is the production of gears for gearboxes, all the honing machines there follow the gear honing method.

Gear honing is also known as rolling honing and used to improve the quality of premachined, hardened and toothed shafts. The quality improvements essentially result in noise reduction and increased service life with improved efficiency. In the case of gear honing, not all process characteristics are comparable with conventional honing. The main distinguishing feature is the tool geometry. A basic distinction is made between spiral- and gear-shaped tools as can be seen in figure 6.2. In large-scale production, however, the variant with spiral tools was not able to assert itself, while the gear-shaped tools are nowadays used in various modifications for hard fine machining of tooth flanks. Depending on machining task and manufacturing concept, either external or internal toothed tools are used. During the process, the tool and the workpiece comb together with a small axis angle offset.

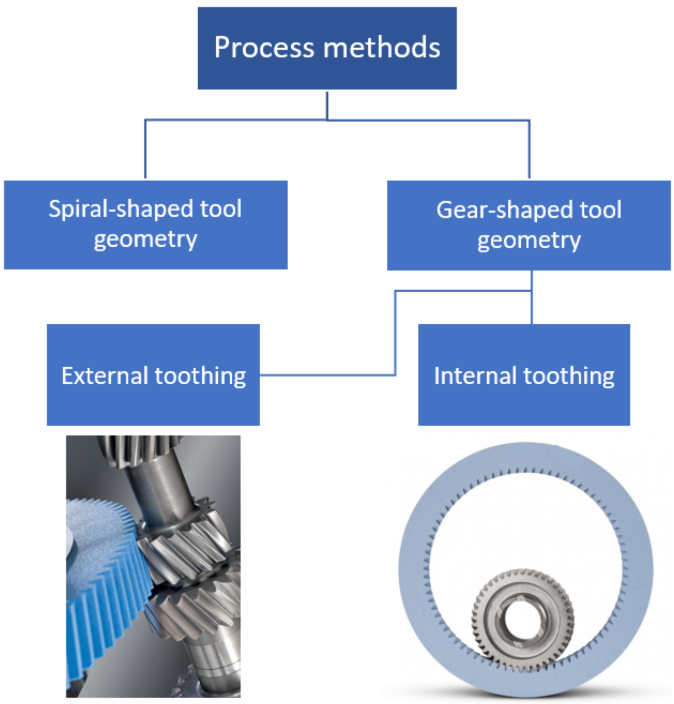

Figure 6.2: Honing process methods (Klink, 2015; Gleason, 2019; Konradin Mediengruppe, 2012)

Honing tools

Internal or external toothed wheels are used as toolings, which are equipped with bonded abrasive deposits. These coatings are comparable to the classic cutting bodies used for grinding and short-stroke honing. A distinction between dressable and non-dressable machining tools can be made.

Process kinematics

During these process-specific kinematics, the honing tool rolls off as an internally toothed honing ring under an inclined axis to the workpiece in the toothing. Under this kind of hobbing process, material is removed over the entire tooth flank, which causes tooth correction and creates the gearing specific surface. Gear honing produces a fish burr pattern on the individual teeth.

Dressing

If the honing process is carried out with conventional cutting tools, dressing must be carried out at certain intervals before use and in series operation. The tool wear resulting from material removal must be readjusted to the nominal dimension and shape within narrow tolerances by dressing. In mass production the required quality consistency can be ensured by dressing with diamond tools. Dressing takes place in two process steps: head and flank dressing. In head dressing with a diamond dressing roller, the tooth height is set and in flank dressing with a diamond dressing gear the flank profile is set. (Klink, 2015, pp.207-211)

6.3.3 The honing machine

Due to the fact that gear honing machines influence the gear surface the most, for this scientific work only honing machines are taken into consideration to find correlations between the gear quality and the condition of the honing machine. Since at GKN Driveline Bruneck only honing machines of the type Synchrofine 205 HS (W) from the German manufacturer Präwema Antriebstechnik GmbH are used, one of these machines was selected for the relevant use case. The concerned machine type is illustrated in figure 6.3.

Figure 6.3: Präwema Synchrofine 205 HS (W) (IndustryArena, 2019)

6.3.4 Tooling

As toolings for the honing process, dressable, diamond-coated, internal toothed wheels are used. Due to the honing ring wear resulting from the continuous removal of material, the tooling has to be dressed at certain intervals. For the dressing process a specific tool, called VarioSpeedDresser (VSD) from Präwema Antriebstechnik is used. This VSD uses only the front, completely defined cutting edge to resharpen the honing ring. Therefore, during the dressing, the high surface quality of the VSD is perfectly transferred to the honing ring and later on to the workpiece. Figure 6.4 depicts the dressing of a honing ring with a VSD. (Präwema Antriebstechnik, 2016) The dressing intervals are determined by internal experts according to the required product quality and may vary from product to product due to the different specifications. The dressing intervals are assessed manually and visually by experts and stay then stable for months until the next inspections are done. The goal is to optimize the dressing interval and therefore, to find an optimum between product quality and tooling costs. Another important point is that the process has to be standardized, which means that employees can follow a given work cycle.

Figure 6.4: The dressing of the honing ring with a VSD (DVS Tooling, 2019)

6.3.5 Quality measurement

To assess the quality of the gearings at GKN Driveline Bruneck all three aspects: the macro-geometry, the micro-geometry and the metallurgical composition, are constantly inspected throughout the manufacturing process.

Geometric measurement

Both, macro-geometric and micro-geometric measurements, are done during the various gear production steps. For the measurement, a device from the company Klingelnberg is used, which is able to calculate not only the profile and flank line deviation, but also pitch and concentricity failure, surface waviness and convexity. The machine uses a tactile probe to measure lines along the gear surface and compares them with the target values specified by the internal experts. Figure 6.5 depicts a typical Klingelnberg device, as it is used at GKN Driveline Bruneck. Such gearing measurement centres provide a high accuracy, low operator influence and they are rather economical compared to other measuring methods. A drawback of the Klingelnberg device is the long measurement time, which makes a 100% testing of all gears impossible. Furthermore, the interpretation of the measurements requires a lot of experience and knowledge from the operators. The Klingelnberg device is also very sensitive to environmental influences such as temperature, vibrations, etc. and therefore requires a constant working environment. (Klocke et al., 2016, pp.284-286)

Figure 6.5: A Klingelnberg gearing measurement centre (Klocke et al., 2016, pp.285)

Metallographic analysis

Also metallographic analyses are performed, in order to check the physical and chemical material properties. Especially the core and near-surface edge area are examined. The specific analyses are executed after the case-hardening of the gears in order to check for the required metallurgic composition. For this type of inspection, destructive as well as non-destructive methods are used. A hardness test takes place, where the hardness of both, the surface and core, are examined. Other important measurands are the hardness penetration depth, which is inspected through specimens, and the metallographic structural analysis.

6.4 Tool wear condition monitoring

There exist various different condition monitoring possibilities for the honing process. Since the tooling has a huge influence on the surface quality of the workpiece, tool wear monitoring is one of the most relevant monitoring methods. The goal should be to achieve an optimum dressing interval through constantly monitoring the tool wear. In the end, this should lead to a stable product quality and lower tooling costs. Furthermore, studies showed that the downtime for machines with a tool monitoring and detection system decreased by 75%, the productivity increased by 10%-60% and the machine utilization enhanced by more than 50%. The reason behind that is the fact that the main factor for machine breakdowns and downtimes are tool failures. (Chen, 2011)

6.4.1 Monitoring methods

Generally, the tool wear monitoring methods can be divided into direct and indirect measurements. To the direct methods belong the discharge current measurement, micro-structure of coating, optical fibre measurement, ray measurement, computer image processing and the resistance measurement. The indirect methods instead make use of physical quantities which occur during the cutting process, e.g. vibration or noise intensity, work piece geometry, torque, cutting force and chip shape. Figure 6.6 shows some of the indirect tool monitoring methods. In comparison to the indirect methods, the direct measurements have the disadvantage that the machine has to be stopped to investigate the tool state. Furthermore, sudden changes or damages which occur during the process can't be detected. Due to these reasons, the indirect methods are nowadays the preferred solution for tool wear monitoring. (Chen, 2011)

Vibration monitoring

Cyclic variances in the movement of the cutting force components provoke vibrations. These vibrations can be visible as small irregularities on the workpiece surface. Vibration signals caused by metal cutting processes can comprise free, periodic, forced

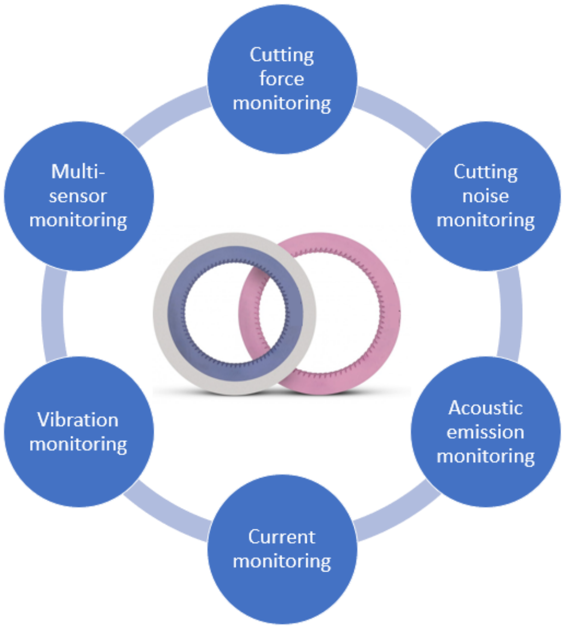

Figure 6.6: Indirect tool monitoring methods (Chen, 2011; Gleason, 2019; Chelladurai et al., 2008)

and random vibration types. It is difficult to measure vibration directly because the vibration mode is strongly dependent of the frequency. Therefore similar parameters, like the acceleration are measured. The vibration characteristics are then extracted out of the acquired data. Studies showed that with vibration monitoring it is possible to detect both, tool wear and breakage. (Dimla, 2000)

Cutting force monitoring

Studies showed that the main cutting force and the feed force stand in a strong relationship with the tool wear. The force itself can be assessed through piezoelectric sensors. Due to the fact that cutting is a complex process, a broad variety of factors influence the cutting force. Therefore, it is difficult to create an accurate and precise cutting force model. Furthermore, due to the complexity, it is difficult to understand if a change was caused by modifications of the cutting parameters, a tool break or other reasons. Also the re-equipment and maintenance of the sensors is rather inconvenient. (Chen, 2011)

Cutting noise monitoring

Sound signals are useful for tool monitoring since they reveal a lot of information about the cutting state and therefore also the tool state. When the wear starts, the noise level increases rapidly but becomes stable later on. With rising cutting velocity, the noise level decreases. Generally, the tool wear has a very good correlation with the sound pressure level. This method is rarely used, since normal production environments are rather loud and thus complicate the implementation. (Chen, 2011)

Current monitoring

A rather popular method for tool monitoring is the supervision of motor current data. If the tool wears out or breaks, the cutting force also changes instantly. The advantages of this method are the easy installation and the low influence of the processing environment. (Chen, 2011)

Acoustic emission monitoring

The contact during the cutting process causes a plastic deformation of the workpiece. Due to the deformation, energy is released, which is also known as acoustic emission. Other causes for the energy release can be friction mechanisms, phase transformations, extension fractures or crack formations. Several researches showed that this approach works well for determining tool breakage or fractures, but is not that well suited for monitoring tool wear. Another drawback is the fact that the interpretation of acoustic emission data is rather complex. (Dimla, 2000)

Multi-sensor monitoring

One of the best methods is to combine multiple sensors, since a single sensor can only provide limited and partial information. Using this approach, higher reliability and accuracy can be reached. A drawback is the higher maintenance effort due to multiple different sensors. (Chen, 2011)

6.4.2 Condition monitoring process

The structure for indirect condition monitoring processes may change from application to application slightly, but consists generally of four main steps:

1. Acquire, collect and process the sensor data like vibration, cutting force, motor current, temperature.
2. Extract the relevant features out of the signals. For feature extraction exist various algorithms which facilitate the selection.
3. Classify or estimate tool wear by using frequent pattern recognition, neural networks, fuzzy logic or regression analysis.
4. Adapt the machining process according to the knowledge gained from the 3^{rd} step.

Through such a monitoring process, the current state of a tool can be assessed and necessary maintenance measurements can take place in time. The chosen sensor and its placement have a big influence on the outcome of such a tool monitoring system. Therefore it is necessary to handle the sensor selection and placement carefully. Generally, the best position to place the sensors is the nearest possible point to the tool, which should be monitored. (Chelladurai et al., 2008)

6.4.3 Acceleration sensors

The special characteristics of the honing cutting process require a monitoring method which provides a high accuracy, simple installation and an easy interpretation of the outcomes. Since similar analyses were already conducted for other cutting processes like grinding and milling, the vibration monitoring approach was used to detect tool wear.
In the following, the sensor type and the placement are further described, since these properties have a huge influence on the outcome of the monitoring system. (Schmitt et al., 2015)

Sensor equipment

For the vibration measurement, acceleration sensors from the company ifm were used. For each sensor an own processing unit, also from ifm, had to be installed, which provides an on-the-fly calculation of the Fast Fourier transform (FFT).
These devices were chosen due to the high reliability and experience ifm has already in the field of condition monitoring. Another important point was that similar devices of ifm are already in use in different machines, therefore, a certain knowledge about the installation, maintenance and handling is already in place. Table 6.1 gives further details about the used technologies.

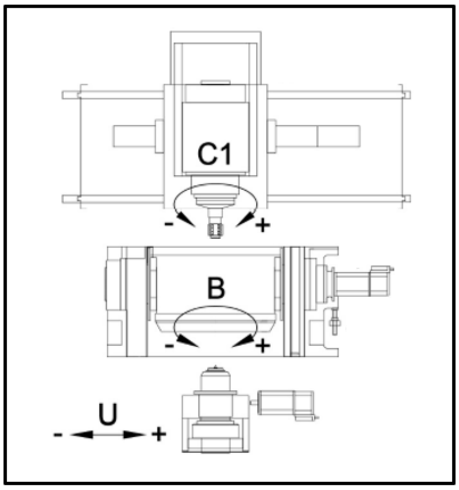

Figure 6.7: The three monitored axes of the honing machine (Präwema Antriebstechnik, 2018)

Image	General Information	Further Details
Vibration sensor: ifm Accelerometer VSA001 (ifm electronic, 2019a)		
	• Type: microelectromechanical system (MEMS) • Weight: 50g • Housing material: stainless steel • Operating voltage: 7.2...10.8 DC V	• Measurement range: ±25g • Frequency range: 0...6000 HZ • Temperature range: -30...125 °C
Processing unit: ifm VSE001 (ifm electronic, 2019b)		
	• Dimensions: 100 x 25.4 x 103.4 mm • Weight: 230g • Housing material: plastic • Operating voltage: 20 DC V	• Sampling rate: 100 kSamples • Frequency range: 0...12000 HZ • Temperature range: 0...70 °C • Communication base: Ethernet • Protocol: TCP/IP

Table 6.1: Sensor equipment

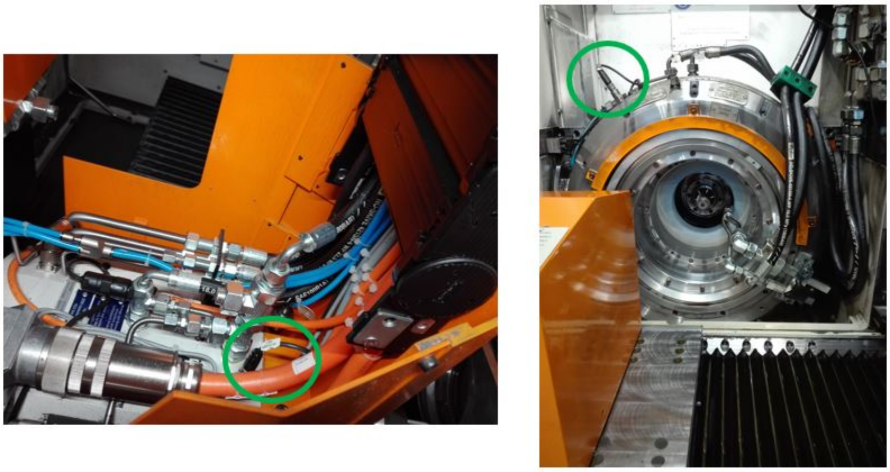

Figure 6.8: The installed sensors, marked in green, on the left on the workpiece and right on the honing ring axis

Placement

A study of (Schmitt et al., 2015) showed that it is necessary to install the sensors as close as possible to the honing process. In total, three sensors were installed at three different axes to cover all potentially important positions. As suggested in the research, one sensor was installed on the top of the workpiece fixture, one on the honing spindle and one on the counter part of the workpiece. Figure 6.7 depicts the machine's axes, where C1 is the workpiece spindle, B the honing spindle and U is the axis with the counterpart of the workpiece. The sensor installation for both, the workpiece axis and the tooling axis is shown in figure 6.8. Due to the fact that the sensor for the workpiece counterpart is mounted in a box, it is not directly visible. (Schmitt et al., 2015)

FFT calculation

Since the raw vibration data is not that descriptive and provides only a small amount of information, it is necessary to do a signal analysis. Therefore, a FFT calculation has to be done, which works best if it is done directly on the processing unit. In order to cope with the vast amount of data, it was necessary to install three processing units, one for each sensor. The processing units continuously calculate the FFT for the incoming vibration data and provide them as spectra with 850 frequencies, which can then be used for the signal analysis.

6.5 Data acquisition

Data analysis and monitoring is only possible if the relevant data is selected, collected from the various sources and then processed to a central storage point. In order to process the data from the machine and the sensors to a database system, a certain ingestion process has to be followed. Before the data is processed, the relevant parameters had to be selected. In the following, the selection of the necessary parameters and the pursued data ingestion process are presented.

6.5.1 Parameter selection

In order to determine the product quality, it is not enough to collect only the data coming from the vibration sensors, but also process parameters of the machine have to be taken into consideration. This is because according to the process parameters different vibrations occur. Therefore, it was necessary to identify potentially relevant machine data for the honing process. Table 6.2 gives a short overview of which parameters were considered relevant and are therefore collected from the machine. A distinction is made between data originating from the machine's PLC and the data coming from the vibration sensors, because for both sources a slightly different acquisition process is persecuted.

Due to the fact that the gear quality is determined after the production on the Klingelnberg measurement device, the quality data had to be collected manually and was added in a later moment to the other data.

Parameter	Type	Description
Machine data		
Product	Integer	The product code of the currently produced part
Mode	Integer	Current machine mode e.g. if the machine is honing or dressing
Process step	Integer	Single machining steps which occur during the process
Dressing interval	Integer	Shows the dressing interval for the current product
Gear counter	Integer	Number of products until the next dressing will occur (starts always at dressing interval)
Date and Time	Timestamp	Production time
Vibration data		
Vibration at tool axis	Float	Spectra with 850 frequencies resulting out of the FFT calculation
Vibration at workpiece axis	Float	Spectra with 850 frequencies resulting out of the FFT calculation
Vibration at counterpart axis	Float	Spectra with 850 frequencies resulting out of the FFT calculation

Table 6.2: Collected parameters

6.5.2 Data ingestion process

To transfer the data from the sensor or machine to a storage and in the end to a visualization or analysis tool, a certain routing has to be followed. This data ingestion process is depicted in figure 6.9. Starting from the machine on the left, the selected parameters are transferred over the machine's PLC to the connectivity platform KEPServerEX, which collects all the data provided by the machines in the entire shopfloor. The main advantage of such a connectivity platform is that it offers a single collection point for data originating from different sources. Therefore, such systems are able to handle various protocols. In this case, the machine parameters are sent from the machine's PLC via the protocol OPC UA to the KEPServerEX. The vibration data, with the already performed FFT analysis on the processing unit, is instead sent directly to the connectivity platform via the publish/subscribe protocol MQTT. Therefore, on the machine's computer a MQTT broker is publishing the vibration data, while on the KEPServerEX a client is subscribed and consumes the published data. The routing for the vibration data is different because the machine's PLC is not able to cope with the vast amount of sensor data. MQTT instead is the perfect protocol for data generated from sensors, since it is open, light weight and simple. It is the preferred solution for machine to machine communication and Internet of Things (IoT) applications where different devices have to be connected. (OASIS Open, 2014)

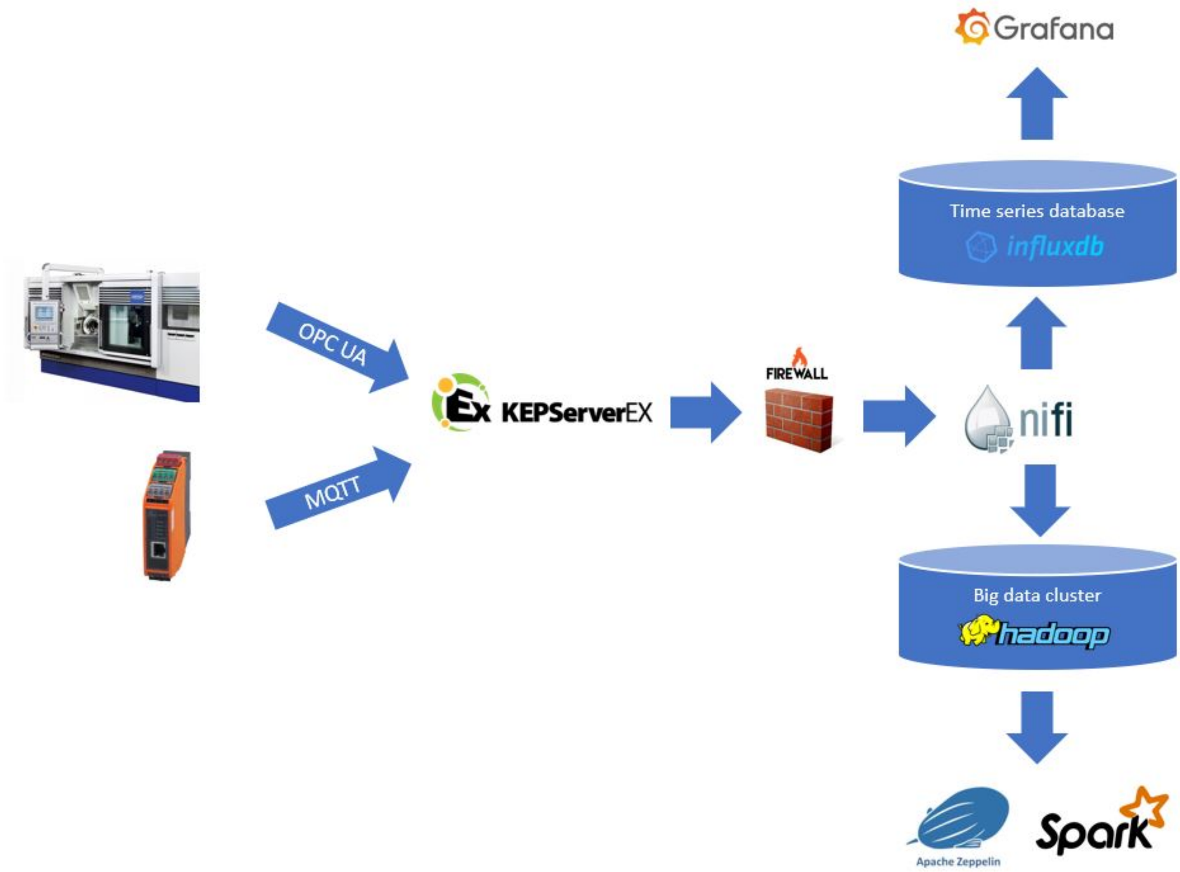

Figure 6.9: The data ingestion process

The collected data is then transferred from the KEPServerEX over an IoT-gateway via MQTT to a server were Apache NiFi is running. In this case, the gateway acts as publisher and Apache NiFi is subscribed and consumes the data. During this step,

the data passes also the GKN Bruneck Firewall in order to access the server which is used by the whole GKN Automotive Group.
Apache NiFi is a open source data distribution and processing system from the well known Apache Software Foundation, which provides not only high performance, but is also reliable. The advantages of Apache NiFi are clearly the good scalability and configurability as well as the web-based user interface and the high security standards. (The Apache Software Foundation, 2018a)
With the aid of Apache NiFi, the the data coming from the IoT-gateway is consumed and transformed in order to extract the machine parameters out of the JSON-formatted data. Afterwards, the data is routed forward into two different storage systems: the Hadoop HDFS and an InfluxDB. These two routings and storage systems are needed because the applications which later use the data have different requirements. The first routing stores the data in the big data cluster, namely the Hadoop HDFS. Before storing the data into the HDFS, it is buffered in a queue and as soon as the queue reaches a certain size, the data is written as one file into the HDFS. If the data wouldn't be buffered, NiFi would write each parameter as single file into the HDFS. This would then cause a lot of single files in the HDFS and lead after some time to a reduced performance during the analysis. Due to the buffering, the data reaches the HDFS with a certain latency, which is not wanted for condition monitoring systems, which require a live streaming of the data. Therefore, the second routing stores the data in a time series database, called InfluxDB, which is optimised for data originating from sensors and therefore suitable for monitoring applications. This NoSQL database is open source, provides a high availability and a fast retrieval. (InfluxData, 2019)
The data stored in the InfluxDB is then used for the live monitoring and visualisation of the machine parameters. As monitoring and visualization tool Grafana is used, which offers not only the possibility to visualize data originating from different sources, but also to monitor parameters and alert the responsible persons if necessary. (Grafana Labs, 2019) Grafana is not able to load huge amount of data, therefore, data which is older than two days is deleted from the InfluxDB. This ensures that there are no two databases which contain the same data.
The data source for the analysis part is the HDFS, since it contains all historical data. In order to provide fast and flexible analyses Apache Spark is used, which is

a good scaling, open source analytics engine for data processing. Spark offers the possibility to access structured data with SQL queries, specific dataframes and the machine learning library MLib. (The Apache Software Foundation, 2018b) Web-based notebooks are provided by Apache Zeppelin, another open source tool from the Apache Software Foundation. In the notebooks, users are able to program specific analytics and visualisations according to their needs. These analytic researches are always based on Apache Spark. (The Apache Software Foundation, 2019)

Process design

For the overall ingestion process it was important to rely mostly on open source software, and therefore, prevent a lock-in. If a company is locked-in into a specific software, this can cause huge costs. Not only the yearly licensing is expensive, but also the switching costs rise. Often it is also not possible to switch from one system to another without a lot of effort and losing data or knowledge. Therefore, all of the used software systems are open source platforms, except the KEPServerEX. The reason behind the usage of this specific platform is the fact that it is the leading connectivity system and is already in use at the shopfloor for several other machines. Therefore, the existing knowledge could be leveraged and it was already known that KEPServerEX is easy to use and offers a good performance and reliability.
Due to the specific design of the ingestion process and the fact, that most of the platforms are open source, single systems can be exchanged without problems. Furthermore, if in the future other machine parameters are required, they can be added easily without causing a lot of changes. The entire process is also flexible enough that it can be used without major changes also for other machines or applications. If the requirements change, the process can also be easily adapted according to the new needs, since all modules work independently from each other. For example, if in the future the visualisation part is not needed anymore, it can be removed easily without interfering the analysis part.

6.6 Experiments

In order to determine the product quality based on machine data not only the relevant machine parameters have to be known, but also the quality of the produced workpieces. Since the quality of the products is measured on the Klingelnberg device after the honing process, an experimental setup had to be executed to gather the relevant information. In the following, the experimental setup, as well as the experiments and the limitations are further described.

6.6.1 Experimental setup

To acquire representative data, four experiments were conducted, where always a pinion gearing for one specific intermediate shaft was honed. The respective intermediate shaft is depicted in figure 6.10, where A) marks the pinion gearing and B) the intermediate gearing. The limitation to one product was necessary because some process parameters change significantly across the different honed gearings.
Normally, the tooling is dressed for this gearing after 70 pieces. This dressing interval was determined from internal experts based on manual and optical investigations of the tooling and the workpieces. Since the process is influenced by a lot of factors, the experts decided upon a rather small dressing interval to ensure that even if factors change the product quality will still be good.
The goal of the conducted experiments for this research was to provide representative data, which then could be analysed in order to proof whether it is possible to determine the product quality based on machine parameters. Therefore, the experimental approach was to produce gearings without dressing the honing ring after 70 pieces until scrap pieces were detected on the Klingelnberg device. To do so, it was necessary to execute measurements continuously over the production process. In order to link the measurement results with the machine data, all workpieces had to be marked manually. The marking was also necessary because the gears were forwarded after the honing process to the assembly line, where they were installed into gearboxes. Only with the marking it was possible to trace the gears and to connect the data of the honing process with the gears' performance on the End-of-Line (EOL) test rig.

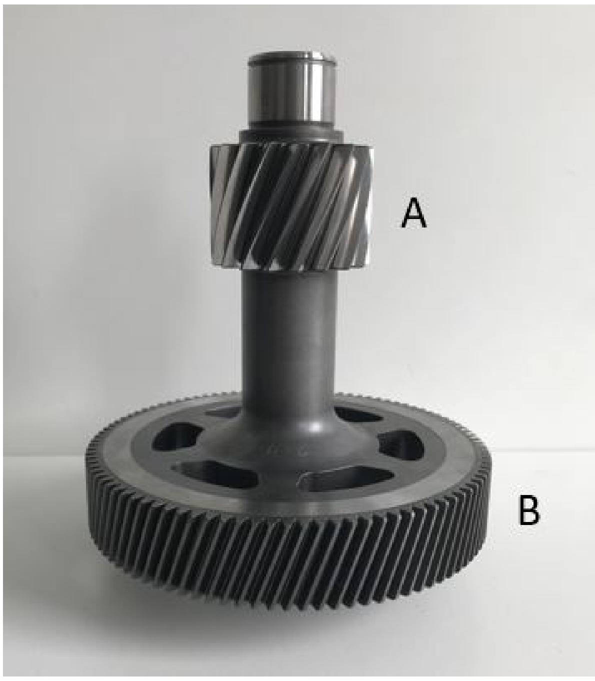

Figure 6.10: The relevant intermediate shaft with A) the pinion gearing and B) the intermediate gearing

6.6.2 Honing experiment

Table 6.3 summarises the conducted four experiments in more detail regarding the honing process. During all trials, the gearings were constantly measured on the Klingelnberg device. A 100% testing was not possible, since the measurement of the flank and profile line and their deviations, the surface waviness and convexity as well as the pitch and concentricity failure takes about 4 minutes per part. Therefore, the gearings were measured as often as possible, mostly every 5 to 10 pieces. The classification of good and scrap gearings out of the Klingelnberg measurements is a complex task, since it is not enough to consider only the calculated measurands, but also the graphical representation of the flank and profile lines. Therefore, the classification into good and scrap parts had to be done by experienced operators and experts and due to that, the valuations can be very subjective and not always clear. In order to avoid any falsifications on the basis of those facts, the classifications for this experiments were done by three different domain experts independently. Nevertheless, it occurred that some measurements lay in grey zones, where no clear statements could be made.

Total honed gearings	Good parts	Scrap parts	Percentage of good parts
Experiment 1			
140	140	0	100%
Experiment 2			
160	160	0	100%
Experiment 3			
128	120	8	93,75%
Experiment 4			
200	94	106	47%
Total			
628	514	114	81,85%

Table 6.3: Experiment details about the honing process

All four experiments were conducted one after the other. At the beginning of the trial, a totally new honing ring was used, which was dressed after each experiment. The dressing interval has to be entered into the machine before the manufacturing process starts and is then not changeable during production. If necessary, the machine provides the possibility to dress the tooling before the selected interval is reached. For the first two experiments the chosen dressing interval, once 140 and once 160, was too small, therefore, the tooling had to be dressed even if the process was still perfectly running. During both experiments, the process operated stable and the surface quality was superior. Of course there were some decreases in product quality over time, but even the last parts were far away of the specified quality limits.
As learning from the first two experiments the dressing interval for experiments 3 and 4 was 900 pieces. It was already known that it is impossible to produce that many gearings without dressing the tooling, but due to the selection of such a high interval, there was no possibility to miss the point when the process starts to get worse and the product quality declines.
From the very beginning, the workpieces from experiment 3 showed a worse surface quality compared to trials 1 and 2. The reason for this can be manifold, for example that the initial configurations done by the machine operator were not that precise

and optimal like for the previous tests. Starting from the 100^{th} piece it became visible that the product quality started to get worse and worse and therefore, the production had to be stopped at part 128. The last eight parts were on the limit of the acceptable surface roughness. Even if they were not really scrap parts, due to the fact that the process did not operate accurately anymore and the quality was at the limit, they are considered as scrap.
The fourth and final experiment started with some machine configuration problems. Therefore, the first six parts had some conspicuous measurands on the Klingelnberg device and were classified as scrap parts. Also for this experiment it became soon visible that the initial configurations were not optimal and starting from piece 80, the quality decreased constantly. Beginning with the 100^{th} part, all following gearings were considered as scrap parts because the process was not stable anymore and the product quality was worse than the internal limits.
Overall, the difficulties of the conducted experiments lay in the classification of the product quality and that the overall process, from marking to honing, to finally measuring the gears, requires a lot of time, effort, experience and coordination. Therefore, and due to the fact that in a running production environment time and employees are always short, during the period of this research not more experiments could be conducted. Another point which became evident was that the production of clearly scrap parts is not easy and depends also strongly on the initial machine configuration which the operators make.

6.6.3 EOL performance experiment

Since the relevant intermediate shaft has two different gearings, after the honing of the pinion gearing, the second gearing, called intermediate gearing, had to be grinded. The hardfinishing process for both gearings is different because they have diverse requirements regarding the surface quality, construction and load. After the grinding process, the gears are then washed before they reach the assembly line of the gearbox.
The internal quality standards for the honing process are much higher than the customer specified standards. Therefore, none of the parts considered as scrap pieces after the honing process had a quality inferior to the required customer quality. Due

to that fact, all honed gearings, except one, from the four experiments were assembled into gearboxes. Only one intermediate shaft, the sixth part of the fourth experiment, has been retained for further investigations and tests.
On the assembly line, the relevant intermediate shaft is assembled together with all the other components like the e-motor, the input shaft, the output shaft and bearings into the gearbox. During the entire assembly process, components are checked regarding the correct positioning. After all parts are installed, the gearbox is tested for leakages and goes then into the EOL test bench. The EOL is the last and final decision point, afterwards, the gearboxes considered as good, go directly to the customer.
At the EOL test rig, the gearboxes are examined towards the fulfilment of the requirements given by the customer. Therefore, various tests are executed to evaluate the engagement and disengagement time of the dog clutch, with which the electric motor can be connected and disconnected from the gearbox, the gear whine noise, the differential drag torque, the damaging of the gear flanks etc.
The most relevant tests regarding NVH behaviour are the Gear Whine Noise test and the Flank Damaging Test. For both tests, the speed is ramped up until a certain number of rounds per minute (RPM) in forward (coast) and reverse (drive) direction. Any occurring vibrations and noises can then be observed with the vibration sensors installed on the test bench. Through these tests, any gear whining or clattering can be discovered. Furthermore, it is also possible to detect broken or damaged bearings or other defects.
For the evaluation, a FFT computation is performed, which outputs an order analysis. Over the years, a certain knowledge could be built up and nowadays it is possible to link peaks at certain orders to specific components. Therefore, abnormal behaviour can easily be detected and leads to a negative valuation of the gearbox. If the information is available which component caused the failure, it is often enough to exchange only the defect part. This leads to an increase in productivity and also lower scrapping costs.

Table 6.4 summarizes the overall performance on the EOL test rig of the gearboxes. As can be seen, the percentage of scrap gearboxes was really small. None of the gearboxes containing intermediate shafts originating from the first three honing experiments was considered conspicuous on the test bench. Both gearboxes which

Total assembled gears	Good gearboxes	Scrap gearboxes	Percentage of good gearboxes
Experiment 1			
140	140	0	100%
Experiment 2			
160	160	0	100%
Experiment 3			
128	128	0	100%
Experiment 4			
199	197	2	98,99%
Total			
627	625	2	99,68%

Table 6.4: Experiment details about the EOL performance

showed abnormal behaviour on the EOL test originated from the fourth experiment. Further investigations showed that both abnormalities were in connection with the corresponding intermediate shafts. At one of the two gears the intermediate gearing was not grinded. The surface quality is without hardfinishing not good enough and therefore, the existing roughness causes vibrations, which can easily be observed with the EOL tests. Since the reason for the failure was caused by humans and is in no relation with the honing process, this scrap part was not considered as relevant for the research.

Also the second gearbox which delivered bad results on the EOL test, originated from the fourth honing experiment. Further research showed that at the 52^{nd} order of the FFT analysis an abnormal peak appeared during the tests. For this specific gearbox, the 52^{nd} order refers to the connection between the intermediate shaft and the differential. Therefore, either the pinion gear on the intermediate shaft or the ring gear on the differential was defected. Unfortunately, it was not possible to detect after the EOL test which gearing caused the defect. In order to determine which gearing triggered the fault, during the data analysis an attempt has been made to find differences in the machine parameters during the honing process between the potentially faulty gearing and the other gearings.

6.6.4 Limitations

Of course experimental setups always have certain limitations and drawbacks. In this research it has been tried to carry out the tests as close to the real production process and environment as possible. Nevertheless, some restrictions had been made to ensure that the production was not stopped, the retrieved data was representable and that the study would not exceed its scope.
The first deviation from normal production was that each intermediate shaft was marked manually to provide a linkage between the honing, measurement and assembly process. Through this additional effort, the production was not as fast as usual. Nevertheless, it does not influence the general outcome of the research negatively. At the moment, no single piece traceability system is used to mark gears, but the first steps towards such a system are already made.
In order to retrieve representative data, it was decided to draw the focus on one specific, high-running honed gearing. Since the honing machine is currently used for multiple different gearings, a suitable time slot was searched to make sure that at a certain time all involved parties and machines were available. Normally, one operator has to work on multiple machines. For this research, during the entire experiment, one operator was responsible only for the honing machine and the Klingelnberg device. This ensured that all arising problems could be fixed immediately and the gearings were continuously measured.
A 100% testing is not possible during normal production, since the measurement time on the Klingelnberg device for the relevant gearing takes around 4 minutes. Due to that, normally, only the first workpiece after dressing and the last piece before dressing are measured. This works quite well, since the process works quite stable and the chosen dressing intervals are rather small. During the experiments, the gearings were measured as often as possible, especially the gearings which were produced above the normal dressing interval of 70 parts.
It became also evident during the experiments, that the classification into good and scrap parts is difficult, since it is not possible to draw a clear line between the two. Due to the fact that it is necessary that the classification task is done by human experts, the decisions are sometimes a bit subjective. To restrict this falsification, three different experts did the classification independently. Nevertheless, a certain kind of

subjectivity remains and influences also the research result.
The honing process itself is really sensitive for external influences. The environment can influence the process e.g. if there occur certain vibrations or if temperature changes happen outside the machine. Also these factors could have affected the result of the experiments, even if they were not noticeable for human senses.
Of course, also the pre-machining and the material have certain influences on the honing process, on the quality and in the end on the NVH behaviour of the gears. If, for example, failures or deviations occur at the hobbing process or the case-hardening, then this will affect the honing process. Therefore, the arising vibrations can be different. These changes affected of course also the results of these experiments.
Another influence factor are the initial configurations, which are done by the operators. During the conduction of the experiment it became evident that the honing process was not operating stable anymore the more parts were produced if the configurations were not optimal.
Also, the selection of the tooling influenced the experiments. Different types of honing rings from various suppliers affect the honing process differently. Based on the honing ring, the vibrations arising during the process can differ significantly. For the conducted experiments always the same honing ring from one supplier was used. It became visible, that also the condition of the tooling affected the vibrations. The experiment was started with a new honing ring, which was then dressed after each experiment. Of course, the outcome of the research would be different with a more worn tooling.
The results of the EOL tests are influenced by a lot of components and factors. The test rig is for example highly sensitive for environmental changes. Due to the fact that multiple components and gearings are assembled into a single gearbox, there are a lot of parts which could have caused a failure. For this research it became visible that the gearboxes considered as defect on the EOL had either non-conforming pinion gearings on the intermediate shaft or ring gearings on the differential. Therefore, also other components which could have been faulty have affected the conducted research.

6.7 Data analysis

During the experiments, the data generated from the machine and different sensors was collected and processed. The knowledge discovery process, which was followed, is depicted in figure 4.4 and further described in section 4.5.
For the analyse, an Apache Zeppelin notebook with Apache Spark was used. The data was retrieved out of the HDFS, preprocessed, transformed and then mined in order to search for correlations between the product quality and vibrations. In the end it was tried to interpret the found patterns and to find a suitable presentation. The approach is further described in the following.

Selection

The real data selection was already done in the parameter selection process described in subsection 6.5.1. During that process it was decided in cooperation with the machine supplier and different domain experts, which machine data could be relevant. The main reason behind this approach was to only retrieve the data that could be relevant and not all the generated parameters. Due to that it was avoided to store a lot of unnecessary data. Furthermore, due to the flexible nature of the data ingestion process, it is still easily possible to add parameters, which could be considered relevant in the future.
By using this approach it was not anymore necessary to do much data selection, since all available attributes were already considered as relevant. The data was retrieved out of the HDFS as dataframe and was then ready to be preprocessed. The only small selection process which took place was to consider only the data resulting from the three main process steps as relevant. This was done because in the other four production process steps the machine is not in direct contact with the workpiece and therefore, the obtained vibrations are not regarded as significant for the product quality. Therefore, the data for the process steps considered as not meaningful was excluded for this research.

Preprocessing

The three main tasks for data preprocessing are data cleaning, integration and reduction. For this research, two of the three tasks were executed.
At first, the data was casted into the corresponding data-types, since they were stored as JSON strings. Then, the data was ready to be cleaned. During this data cleaning, incomplete data points were removed or missing values filled in and inconsistencies resolved.
Due to the structure of the data ingestion process, all relevant data is stored in the HDFS, therefore, no integration of different data sources had to be done. The only pieces of information, which had to be added manually to the parameters were the quality labels for each produced part.
Data reduction was not needed for this research because the data set was not that huge.

Transformation

The next step was then the transformation of the cleaned and preprocessed data. It was decided that the data had only to be normalized, no other additional transformation task was required. The reason for that was the fact, that the amplitudes of the vibration spectra vary in a wide numerical range.
Three different algorithms were used for the task of normalization: the MinMaxScaler, the Normalizer and the RobustScaler. All algorithms are included in the scikit-learn library. The MinMaxScaler scales each feature individually to a given range e.g. [0,1]. This works quite well, but is very sensitive to outliers because all features are compressed into a narrow range. The Normalizer instead rescales the data vector in samples, e.g. in rows into unit norm. The RobustScaler centres and scales the data based on percentiles and is therefore not that sensitive for outliers. (scikit-learn developers, 2018a) These three different algorithms were selected and tested because the normalization has a huge influence on the final result and it was therefore crucial to determine the optimal algorithm. The Normalizer function proved to be the best algorithm for this application, because the RobustScaler and the MinMaxScaler tended to lose the relevant information. MinMaxScaler scales the data down so that all points

lie between 0 and 1. RobustScaler instead doesn't count much on possible outliers, but for this application they appear to contain the most information. Therefore, after the normalization with both algorithms, the datapoints appeared to lie mostly on the same spots, therefore, no interpretation of the data was possible. Figure 8.1, which can be found in the appendix, shows the similarity of the results for the SelectKBest algorithm with the f_regression scoring function when normalizing the data before with the A) MinMaxScaler and B) RobustScaler. The diagrams present on the x- and y-axes the features, the frequencies, which were selected as the most important features by the used feature extraction algorithm. In the plots the calculated correlations between the single features are depicted. The orange data points are data samples of good parts, while the blue data points mark scrap parts. In the main diagonal the statistical distributions of the calculated correlations are shown. It appears that there is nearly no difference when applying the mentioned algorithms. The result for the Normalizer function is instead depicted in figure 6.12 later in the chapter. It is clearly different from the other two functions and provides more information about the data. Therefore the Normalizer function is used throughout the research.

Data mining

After that the right data was selected, preprocessed and transformed, the essential step of data mining could take place. There exist multiple data mining possibilities, which are depicted in figure 6.11 and are described in section 4.5.4. Out of all the possible methods, the detection of frequent patterns, correlations and associations was chosen in order to find dependencies of the different features based on the product quality.

Since the selected data consisted of vibration spectra with 850 frequencies, it was necessary to detect the most relevant frequencies. Otherwise, the irrelevant features can falsify the analysis results. Therefore, a feature extraction took place to determine the features which contain the most information. By selecting the most important features, overfitting is reduced, while the accuracy is improved. Additionally, for smaller datasets an eventual training of the data performs faster. For feature importance calculation, a wide range of different algorithms are available, each providing

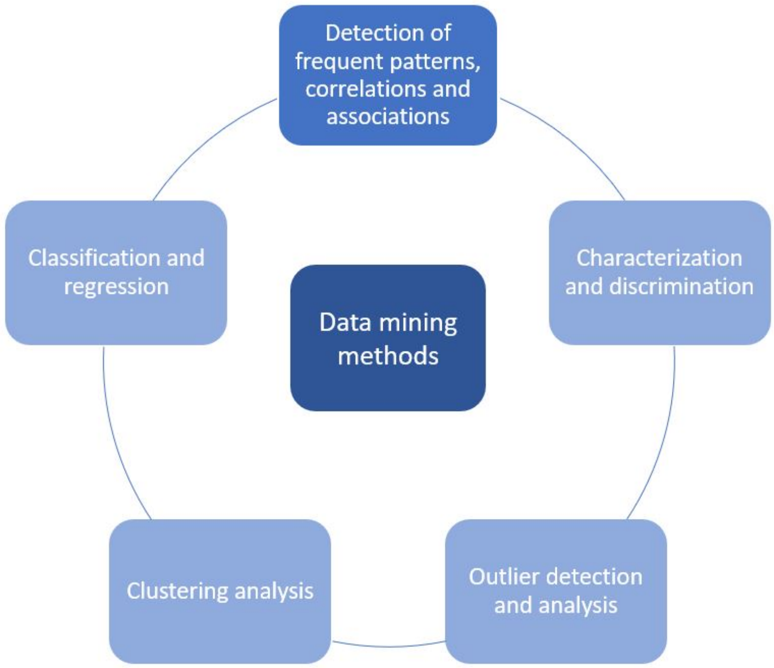

Figure 6.11: The data mining possibilities and the selected method.

varying advantages and drawbacks. During this research, a small selection of these algorithms was tried out and evaluated. In the following, the different approaches are presented and further explained. (Brownlee, 2018)

Univariate feature selection

The first tested algorithm was the univariate feature selection provided by the scikit-learn library. When using this algorithm, the features which have the strongest relation with a certain output variable are selected based on univariate statistical tests. There exist different implementations which base on the univariate feature selection, the most relevant are: the SelectKBest, the SelectPercentile, the GenericUnivariateSelect. SelectKBest selects the k features with the highest score and removes all other features. For SelectPercentile the user instead can enter a certain percentage of relevant features. SelectPercentile calculates the scores for all features and removes all but the percentile of highest-scoring features. GenericUnivariateSelect allows to configure the strategy of the univariate feature selection with a hyper-parameter estimator. All

of these objects have an input parameter, which is a scoring function. The task of the scoring function is to calculate and return the univariate scores and the p-values. Based on the task the user wants to perform, the functions can be divided into scoring methods for regression and classification. For regression, the scoring functions are f_regression and mutual_info_regression, while for classification tasks chi2, f_classif and mutual_info_classif are used. (scikit-learn developers, 2018d)
For this research, the SelectKBest and the SelectPercentile were tried out with the scoring functions f_regression and mutual_info_regression. The scoring function for regression tasks were used, since the task was to examine the relationship and dependence from the features. Linear dependencies are visible with f_regression, while f_mutual_info captures all dependencies between the features. If the result is 1 or close to 1, then the features are highly dependent from each other. On the other side, if the values are close to 0, then the values are independent from each other. (scikit-learn developers, 2018b; scikit-learn developers, 2018e)
The top 5 features were selected at first with the SelectKBest algorithm, using the f_regression scoring function. The input parameter for the corresponding output variable is in this case the quality of the corresponding piece. For normalization,the Normalizer function was used before. Then, the correlation was calculated between the features, which resulted as k best features. The same was then done a second time, but as scoring function the mutual_info_regression function was used. Figure 6.12 shows exemplary the difference in the obtained results between the two scoring functions. In order to give more detail the same plot is depicted in a higher resolution in the appendix as figure 8.2. The diagrams present on the x- and y-axes the features, the frequencies, which were selected as the most important features by the used feature extraction algorithm. In the plots the calculated correlations between the single features are depicted. The orange data points are data samples of good parts, while the blue data points mark scrap parts. In the main diagonal the statistical distributions of the calculated correlations are shown. It is clearly visible that both variants mark different features as relevant. Part A shows the results from the f_regression function, it is visible that some correlations exist and the selected features are not that close to each other. The results obtained from the f_mutual_info function show that the algorithm selects subsequent frequencies, which is not that useful because frequencies which are close to each other are similar and have therefore also strong

relationships. For both applied algorithms is to say that the found correlations are too weak or not that useful to interpret. Furthermore, a distinction between the good and scrap parts can't be done easily.

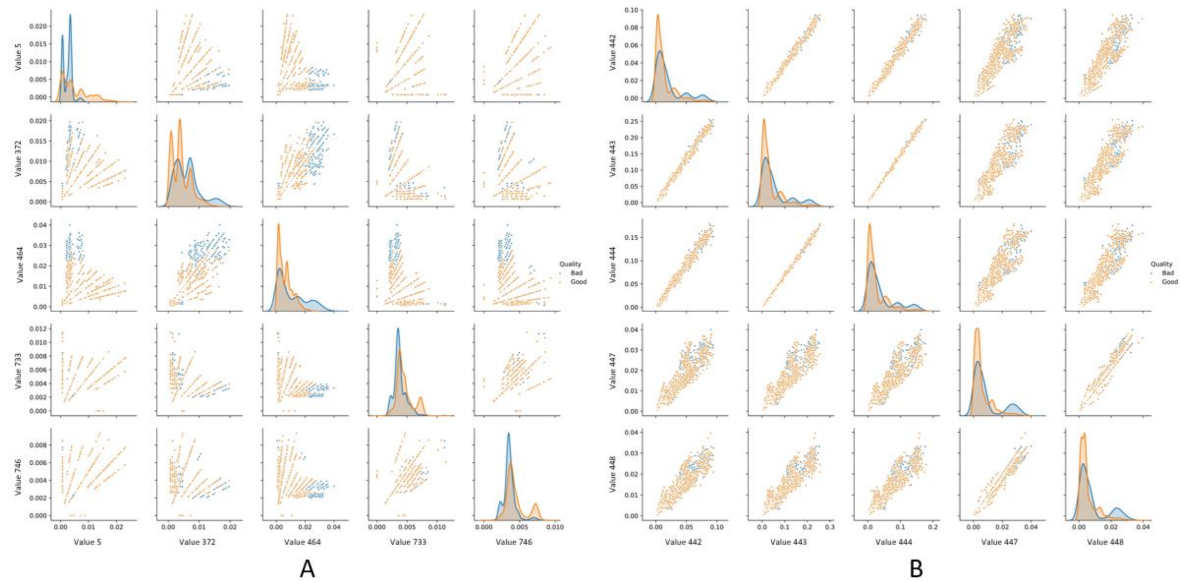

Figure 6.12: Feature selection using the SelectKBest function with A) f_regression and B) f_mutual_info.

The SelectPercentile function works similar to the SelectKBest, the only difference is that not a specific number k but a certain percentage of features are selected. Since the first highest-scoring percent returned already 9 features, for convenience reasons only SelectKBest was used.

Variance threshold

The next performed feature selection algorithm was the variance threshold. The basic principle of this selection method is to remove all features with a low-variance. The function takes a certain threshold as input and all features with a lower variance are removed. The variance threshold function works without inserting a desired output and is therefore often used for unsupervised learning applications. For this research, a threshold of 0.0291 was chosen, in order to retrieve at least 5 relevant features. This was done in order to provide a certain comparability to the SelectKBest function.

(scikit-learn developers, 2018f)
Figure 8.3 in the appendix shows the correlation result obtained for the features determined with the variance threshold selector. The diagram presents on the x- and y-axes the features, the frequencies, which were selected as the most important features by the used feature extraction algorithm. In the plot the calculated correlations between the single features are depicted. The orange data points are data samples of good parts, while the blue data points mark scrap parts. In the main diagonal the statistical distributions of the calculated correlations are shown. Compared to the results of the SelectKBest function, the resulting features are totally different. For most of the features no correlation is visible, furthermore, also a distinction between good and scrap parts is difficult to make.

Extra-trees classifier

Another algorithm for determining feature importance is the extra-trees classifier. This algorithm is based on decision trees and calculates the importance of each feature. A certain number of decision trees is fitted on different sub-samples originating from the main dataset. For the fitting, a special estimator is used. The higher the estimated score for the single features is, the more important they are. Through averaging, a certain accuracy can be ensured and overfitting can be prevented. (Brownlee, 2018; scikit-learn developers, 2018c)

In an input parameter of the extra-trees classifier function, the number of trees in the forest can be specified. A crucial part when using the classifier is to find an optimal number of trees in order to get the best possible results in a reasonable time. Figure 6.13 shows the obtained results for two extra-trees classifiers, with A) 100 and B) 200 used trees. The plot can be found in a higher resolution also in the appendix as figure 8.4. The diagrams present on the x- and y-axes the features, the frequencies, which were selected as the most important features by the used feature extraction algorithm. In the plots the calculated correlations between the single features are depicted. The orange data points are data samples of good parts, while the blue data points mark scrap parts. In the main diagonal the statistical distributions of the calculated correlations are shown. The graph A) looks really similar to the obtained

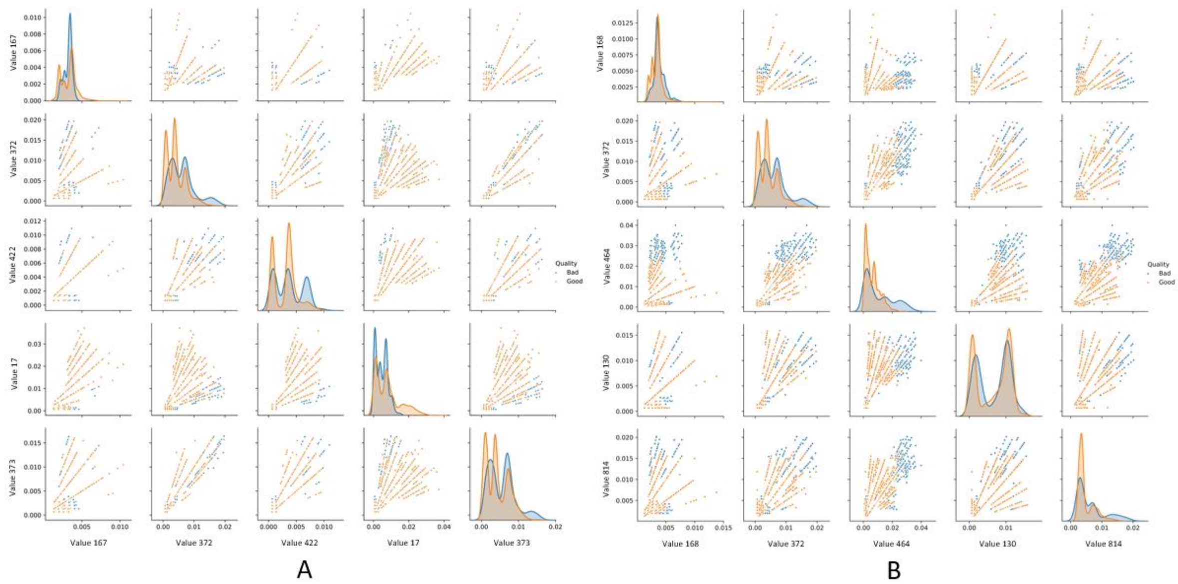

Figure 6.13: Feature importance calculation with the extra-trees classifier with A) 100 and B) 200 estimators.

results for the SelectKBest with the f_regression scoring function. Graph B) instead uses more trees for the importance computation and shows better results because the good and scrap parts appear to sometimes build cluster-similar concepts. For the classifier with 100 trees instead, the graph is more sparse. It was also tried to further increase the number of trees to 1000, but the results do not get significantly better.

Recursive feature elimination

The last algorithm, which was tried for feature extraction, was the recursive feature elimination (RFE). The basic principle behind this method is to remove attributes recursively and to build then a model based on the remaining features. The algorithm takes as input a target output and with the aid of the model accuracy, the features which make the highest contribute to predict the target remain. The recursive feature elimination is often used in combination with the algorithm for logistic regression in order to select the most important features. (Brownlee, 2018)
Also for this research, the RFE method is used with the logistic regression algorithm. The quality labels are the target attributes for the algorithm. The correlation results for the five selected features are depicted in figure 8.5, which can be found in the

appendix. The diagram present on the x- and y-axes the features, the frequencies, which were selected as the most important features by the used feature extraction algorithm. In the plot the calculated correlations between the single features are depicted. The orange data points are data samples of good parts, while the blue data points mark scrap parts. In the main diagonal the statistical distributions of the calculated correlations are shown. It is visible that the distinction between good and scrap parts can't be done easily, because the datapoints are spread evenly. A drawback of this method is the long computing time, which is caused by the recursive strategy.

Interpretation, evaluation & presentation

As presentation of the retrieved results, a correlation matrix based on the pairplot function from seaborn was used. With this function it is easy to graphically represent correlations of different features. The main diagonal of the plot shows the statistical distribution for each feature. (Michael Waskom, 2018)
Using the graphical representation of the data, correlations and clusterings between the different features were searched. Unfortunately, none of the used algorithms and methods returned an explicit correlation or distinct clusters of good and scrap parts. It could be observed, that there exit some transition areas were good and scrap parts are mixed. These transition areas can be seen especially good in the plot resulting out of the extra-tree classifier with 200 estimators. Also when using the SelectKBest algorithm with the f_regression scoring function and the extra-trees classifier with 100 trees the transition areas are visible. Rather difficult to interpret are the variance threshold method, the SelectKBest with the scoring function f_mutual_info and the recursive feature elimination.
A heatmap plot for the correlations between the features was considered as not enough for this research. The reason behind this decision was the fact, that some structures of the datapoints can falsify the calculations and clusters of samples can't be seen.

6.8 Outcomes

Different investigations of the retrieved data were made. Not only the correlation of the vibration data with the product quality was examined with big data algorithms, but also some researches were performed on the EOL test data. In the following section, the results from the different investigations are presented and explained in more detail. The data analysis is not the only outcome of this masters thesis. In the course of the research, a condition monitoring system was developed for the relevant honing machine. The corresponding dashboard is also presented in this chapter, as well as potential saving, which can be accomplished by optimizing the tool change process.

6.8.1 Correlation of honing vibrations and product quality

In order to fulfil the main goal of this research to determine product quality in manufacturing based on machine data, it was tried to find relationships and correlations in the data. In the previous section, the different data-driven approaches and used methods are described, which were used to analyse the vibrations of the honing machine and the quality of the workpieces.
The analysis showed, that there exists no clear clustering or delimitation between good and scrap parts. Nonetheless, some accumulations of only good and only scrap parts could be observed. There exist also some transition areas, where good and scrap parts are mixed. These observations give the opportunity to specify certain limitations at very low limits in order to retrieve only good parts. The drawback of such an approach is that the limits are for sure chosen to low, which causes the tooling to be dressed and in further course exchanged more often. The resulting low utilization rate of the tooling would then increase the cost, which is not economic at all.

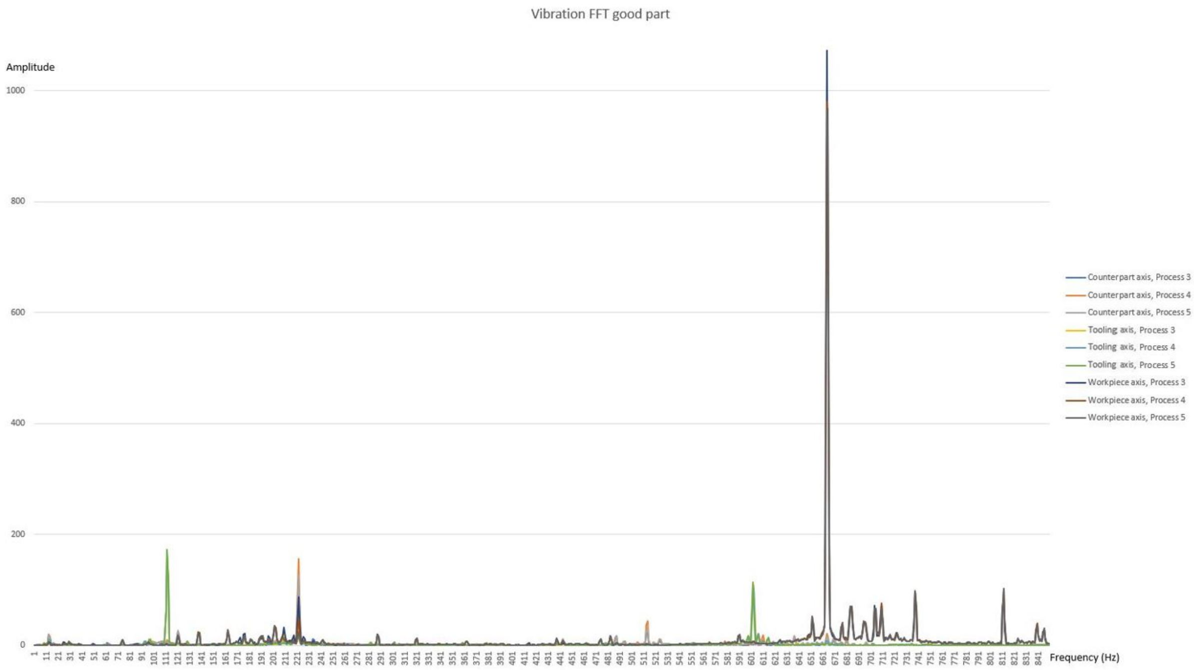

Figure 6.14: Comparison of the FFT spectra obtained from the different axes and processes for a good part.

6.8.2 FFT comparison of gearings

To determine if it is possible to detect if the produced scrap parts have significantly other vibrations during the honing process, the detected vibrations are compared with comparable vibrations when good parts were produced. A good starting point for such a comparison is to calculate a FFT analysis out of the occurring vibrations and to compare then the resulting spectra.

The observation is done for the three different axes of the tooling, the workpiece and the counterpart. The selected production processes during which the vibrations are recorded were chosen, because during this main three processes the workpiece and the tooling are in direct contact. During the remaining process steps, which are considered as irrelevant, the tooling and the workpiece do not interact with each other. Therefore, the relevant processes are process 3, process 4 and process 5. Process 3 is the initial way of the workpiece spindle towards the honing ring with a very high speed. Process 4 is the main production step, where the workpiece moves inside the honing ring but with lower speed. Process 5 is the last step before finishing, where

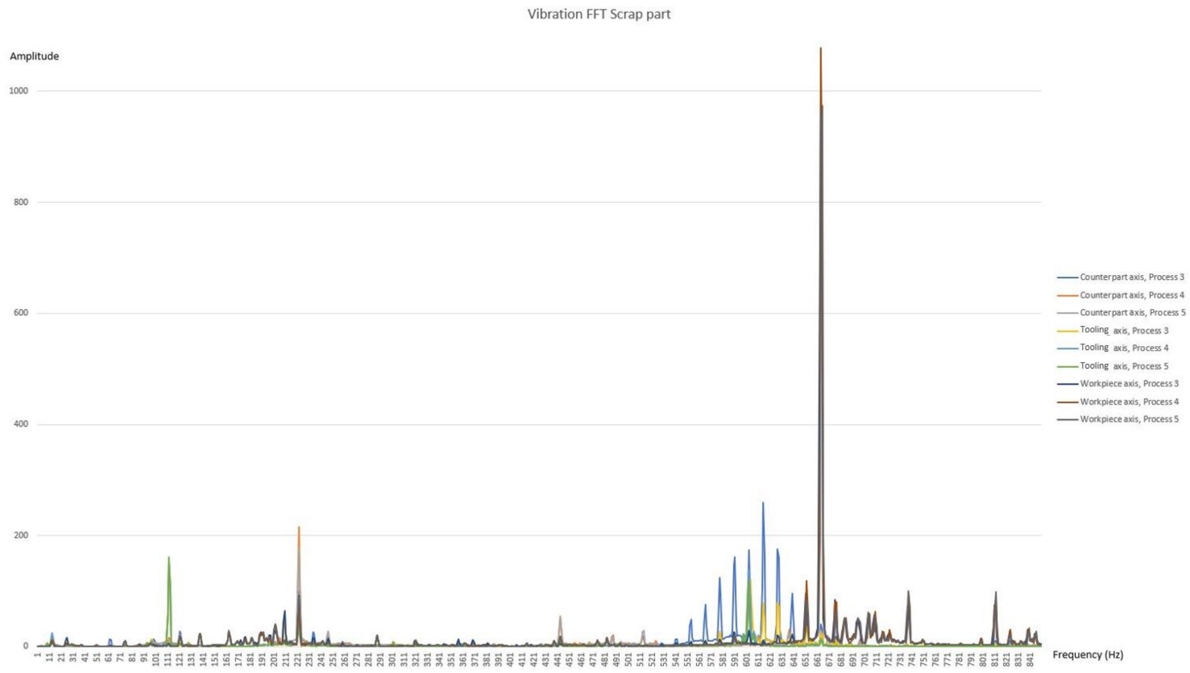

Figure 6.15: Comparison of the FFT spectra obtained from the different axes and processes for a scrap part.

the speed is more and more decreasing until the workpiece is detached from the tooling.

The different vibration spectra obtained for a good part can be seen in figure 6.14, while figure 6.15 shows the spectra for a scrap part. In order to provide more detail both figures are depicted in a higher resolution in the appendix as figure 8.6 and figure 8.7. The two plots show the vibration spectra obtained with the FFT calculation. The y-axes depict the amplitude, while the x-axes show the frequency (Hz) of the samples. The differently colored spectras show the various spectras obtained from the three different sensors during the three different processes. It is clearly visible that all vibrations, except the ones recorded from the workpiece axis, lie below amplitude 200. The vibration of the workpiece axis instead has a very high peak at the frequencies 661 and 671. This is equal for both, the good and the scrap part. The difference in the vibration of the two different parts is visible when observing the counterpart axis when process 3 (spectra in blue) is executed. Increased vibrations are also visible for the tooling axis also during process 3 (spectra in yellow) and for the counterpart axis during process 4 (spectra in orange). Process 5 and the workpiece axis seem to be rather irrelevant and can be neglected for the detection of scrap parts. The vibrations which seem to give the best hints about the product quality seem to originate from the counterpart axis and when process 3 is executed.

6.8.3 Detection of tool wear influence on EOL results

In the previous sections the influence of honing tool wear on the surface quality of the gearings was investigated. The mentioned surface roughness of the gears is known to have a huge impact on the NVH behaviour of the gearbox. Therefore, also the influence of the tool wear on the final performance of the entire gearbox on the End-of-Line test rig was examined.

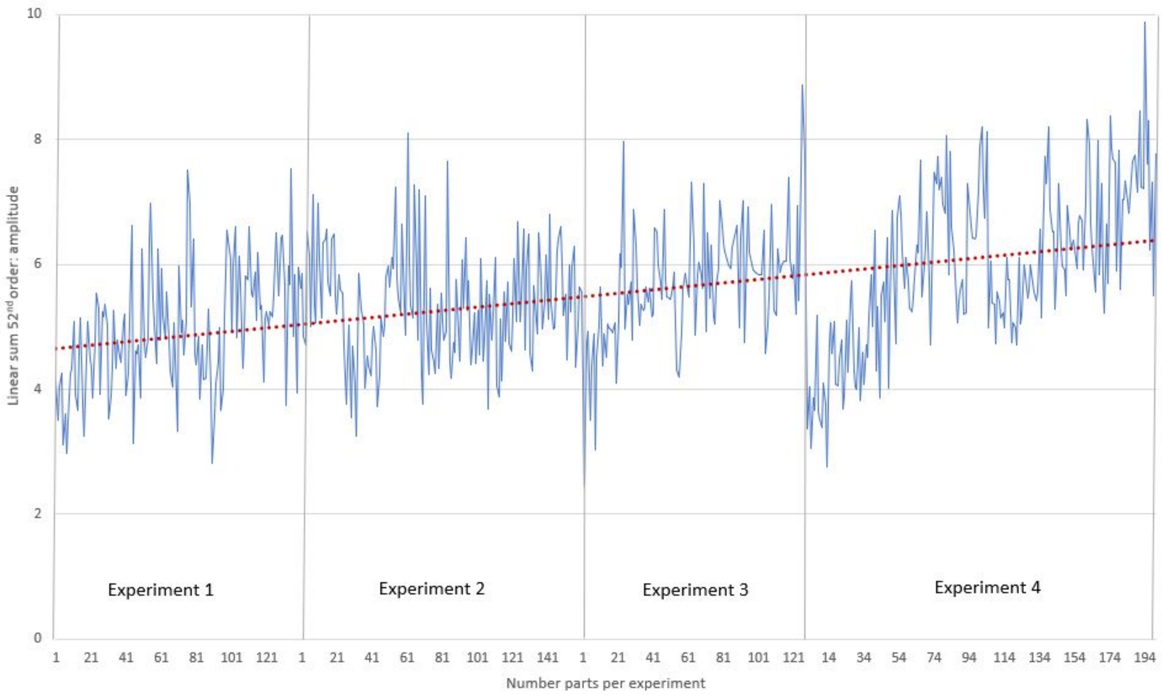

Figure 6.16: Tool wear influence on the EOL results for the 52^{nd} order.

The gears which were produced for the honing experiment were marked manually and could then be assembled into the gearboxes in sequential order. All the data generated by the EOL test rig is stored in a special database and can then be further

investigated. Since it was already known that the 52^{nd} order of the FFT analysis performed on the test rig corresponds to the relevant intermediate shaft, the linear sum of the 52^{nd} order was used for the analysis. The created diagram is visible in figure 6.16. On the y-axis the linear sum of the amplitude of the 52^{nd} order is depicted, while the x-axis shows the number of parts produced over the different experiments. The blue line shows the behaviour of the linear sum over the produced parts of the experiments. The dotted line in red is the linear trend line calculated out of the blue linear sum line. The plot depicts not only that the linear sum is constantly increasing between the experiments, but also that after dressing the tool, the linear sum starts again at a lower point. This trend is especially visible between the experiments 2 and 3 and the experiments 3 and 4. Furthermore, also an overall trend is visible, which shows that the linear sum is constantly increasing the more the tool is worn out. It is here necessary to take into consideration that the 52^{nd} order is not only associated with the intermediate shaft, but also with the differential. Therefore the results can also be falsified by the performance of the differential.
This result confirms that the honing process and the tool wear have a huge influence on the final NVH behaviour of a gearbox. The more the tooling is worn out, the more vibrations are caused on the EOL test and the worse the NVH behaviour is.

6.8.4 Potential savings

Optimizing the tool life helps a company to save money and make the production more efficient. During the four conducted experiments on the honing machine it became evident, that the chosen dressing interval was too small. For each experiment the 70 pieces were exceeded by far, while the product quality was mainly good. This shows, that extensive saving are still possible in this area and in the following a short summary for potential cost savings is given.

Table 6.5 gives an overview over the main cost drivers for the tooling of the honing process. It is evident, that the number of honed gearings will increase in the next year drastically. The tool life of a honing ring depends on the product, the material of the honing ring and the configurations on the machine. During production the tool wears out and has to be resharpen. These so called tool dressings resharpen the tool

Description	Number
Number of honed gears in the year 2019	241,350 #
Number of honed gears in the year 2020	980,000 #
Tool life of honing rings	800 - 2,700 #
Acquisition costs for honing rings	450 - 500 €

Table 6.5: Tooling cost details for the honing process

by cutting away material. This approach causes, that the tooling forms get worse after each dressing, so they also increase the tool wear. Therefore, after a certain number of tool dressings the tool has to be exchanged, otherwise the required product quality can't be reached. The tool life of the honing ring is therefore between 800 and 2,700 produced parts, depending on the different influence factors. Also the costs for a honing ring depend on the the product and on the material of the tooling, therefore they lie between 450 - 500 €.

This calculation wants to give only a rough and simple overview about the saving potential for tooling acquisition costs. Therefore, machine downtimes, current and maintenance costs are neglected. Furthermore, the exact life time of a honing ring is not yet exactly known. Therefore, for this calculation some assumptions were made in order to give an exemplary overview.

During all four experiments the minimum number of produced parts which were still good where 100 pieces. Therefore, it was assumed for the calculation, that the dressing interval was increased from 70 to 100 pieces. That would correspond to 42.86% of additionally produced parts for each dressing interval. For the tooling acquisition price the average cost of 475 € and for the tool life of a honing ring the average life time of 1.750 produced parts are assumed.

For the year 2019 the tooling costs for a process without optimization are

$$\begin{gathered}\frac{241,350}{1,750} = 137.91 = \lceil 140 \rceil \\ 475\,€ * 140 = 66{,}500\,€\end{gathered} \tag{6.1}$$

In comparison to that, the tooling costs with optimized dressing intervals are

$$\begin{aligned} 1,750 * 42.86\% &= \lfloor 750 \rfloor \\ \frac{241,350}{2,500} = 96,54 &= \lceil 100 \rceil \\ 475\,€ * 100 &= 47{,}500\,€ \end{aligned} \tag{6.2}$$

That means, that the potential savings for the year 2019 are

$$66{,}500\,€ - 47{,}500\,€ = 19{,}000\,€ \tag{6.3}$$

The savings for the year 2019 are rather small, but for the year 2020 the number of honed gears is nearly tripled. Therefore, the savings for the year 2020 are also calculated in the following.
The tooling costs for a non optimized honing process for the year 2020 are

$$\begin{aligned} \frac{980,000}{1,750} &= 560 \\ 475\,€ * 560 &= 266{,}000\,€ \end{aligned} \tag{6.4}$$

While the tooling costs for the optimized dressing intervals are

$$\begin{aligned} \frac{980,000}{2,500} &= 392 \\ 475\,€ * 392 &= 186{,}200\,€ \end{aligned} \tag{6.5}$$

This leads to potential cost savings for the year 2020 of

$$266{,}000\,€ - 186{,}200\,€ = 79{,}800\,€ \tag{6.6}$$

Therefore, it can be seen that for the short term the optimization results in not high savings. Due to the fact, that the number of honed gears increases rapidly and the more gears are produced the more savings can be made.

6.8.5 Condition monitoring dashboard

Condition monitoring systems can help to detect potential machine failures, which will arise in the future. With the aid of a condition monitoring system maintenance employees don't have to do all inspections continuously on the machine itself, but can rely on the information provided by the system. If necessary the responsive maintenance people will be notified about the detected discrepancies. In the course of the research such a monitoring system was developed.

Figure 6.17: Condition monitoring system for the honing machine.

Figure 6.17 shows the developed condition monitoring system for the honing machine used in this research. The dashboard was created with Grafana and the visualized data is retrieved out from the InfluxDB. In order to make a user-friendly interface, it was very important to keep the visualization simple and to get constant feedback from the users. The visualized parameters were selected in cooperation with the users, especially maintenance employees and machine operators. These experts know the best, which parameters are relevant to determine the condition of a machine.
To describe the machine condition a wide range of parameters is retrieved and visualized. The current product which is produced, the program, in which the machine is operating as well as the process step are displayed. The number of parts till dressing,

the dressing interval, the maximum tool life and the remaining tool life are also visualized. Additionally, different thermal information like the temperature of the motor winding or the cooling of the spindle for the various axes are shown. On the bottom of the dashboard the vibration data of the tooling, workpiece and counterpart axes is visualized as time domain signals.
In cooperation with the maintenance employees certain limits for the machine parameters were determined. When the attributes exceed the defined borders, the responsible people are directly notified via mail.

7 Conclusion & further work

The several conducted experiments and the different data analyses showed, that the determination of product quality based on machine data is generally possible. The selection of the right machine parameters is crucial in order to get the relevant data. Due to the fact that different approaches exist to determine the tool wear, the most important methods were described and in the end vibration monitoring was selected for this research.
In order to get the data from the machine to the data storage point and to different application tools, a framework for data acquisition and ingestion was developed. This framework was held open and flexible, so that additional parameters can be added easily. Additionally, the framework can also be used for other, similar machines or applications. Parts or single systems within the framework can also be exchanged, without influencing the other components.
When the data was stored, various analyses were performed upon the different parameters. A big-data approach was utilized to find correlations and associations within the data. Therefore, data selection, preprocessing, transformation and mining were executed in order to extract relevant knowledge. Due to the fact that many features were available, it was necessary to find the most relevant features during the data mining step. To do so, the different feature extraction and selection methods were tried. The results showed that no really strong and interpretable correlations exist within the data. Also no clustering between good and scrap parts was visible. This can have various reasons e.g. not enough data points in general, no really scrap parts were produced, subjectivity of quality determination, to narrow selection of machine parameters. For further work it is considered, to execute more experiments and add additional machine parameters like temperatures, force and energy consumption.
The examination of the FFT spectra for a good and scrap part showed, that differences

in the vibration exist. The conducted analyses showed, that for the detection the counterpart and the tooling axes gave the most hints, while the workpiece axis showed no difference in behaviour. Also the three main processes of the machine, where the workpiece is in direct contact with the tooling, were examined. It became evident, that especially the starting process, were the workpiece and the tooling interact with a very high speed contains the most information. For further work in this area, these investigations can give a good starting point. A very interesting approach would also be, to try with a big-data approach to extract the relevant features out of exactly these axes and process steps.

Since the gear surface quality is the determining factor for the NVH behaviour of gearboxes, also an investigation of the honed gear performance on the EOL test rig was conducted. There it became evident that the tool wear is visible in the performance of the gearbox of the EOL. Even if the relevant frequency order is also associated with the differential, the found trend is strong enough to be caused by the tool wear of the honing ring.

A calculation was performed to assess the cost saving potential for the tooling acquisition costs for an optimized dressing interval. The results showed that for the year 2019 the potential cost savings for the tooling are 19,000 €. Due to the fact, that the number of honed gears is 2020 nearly tripled, there the cost savings total add up to 79,800 €. This makes evident, that in this area a lot of money can be saved. Therefore, a recommendation is to make further experiments, find a suitable parametric limitation and optimize then the dressing interval.

A condition monitoring system was developed in order to provide the employees with information about the machine and to optimize the maintenance work. Certain inspections can now be done with the aid of the monitoring system. It is possible to define also limits for certain machine parameters and when these limitations are exceeded the system will notify the responsible people.

Overall can be said that monitoring and optimizing the machine and the tooling can save a lot of money, effort and time. Nevertheless, the usage of the developed systems and knowledge is necessary to add value. Therefore, a companies strategy has to be adapted and the employees sensitized for the aid which the systems can deliver.

8 Appendix

In this chapter the diagrams, which resulted out of the feature selection and extraction process described in section 6.7 and the comparison of the vibration spectra explained in section 6.8.2, are depicted in a higher resolution in order to give more detail.

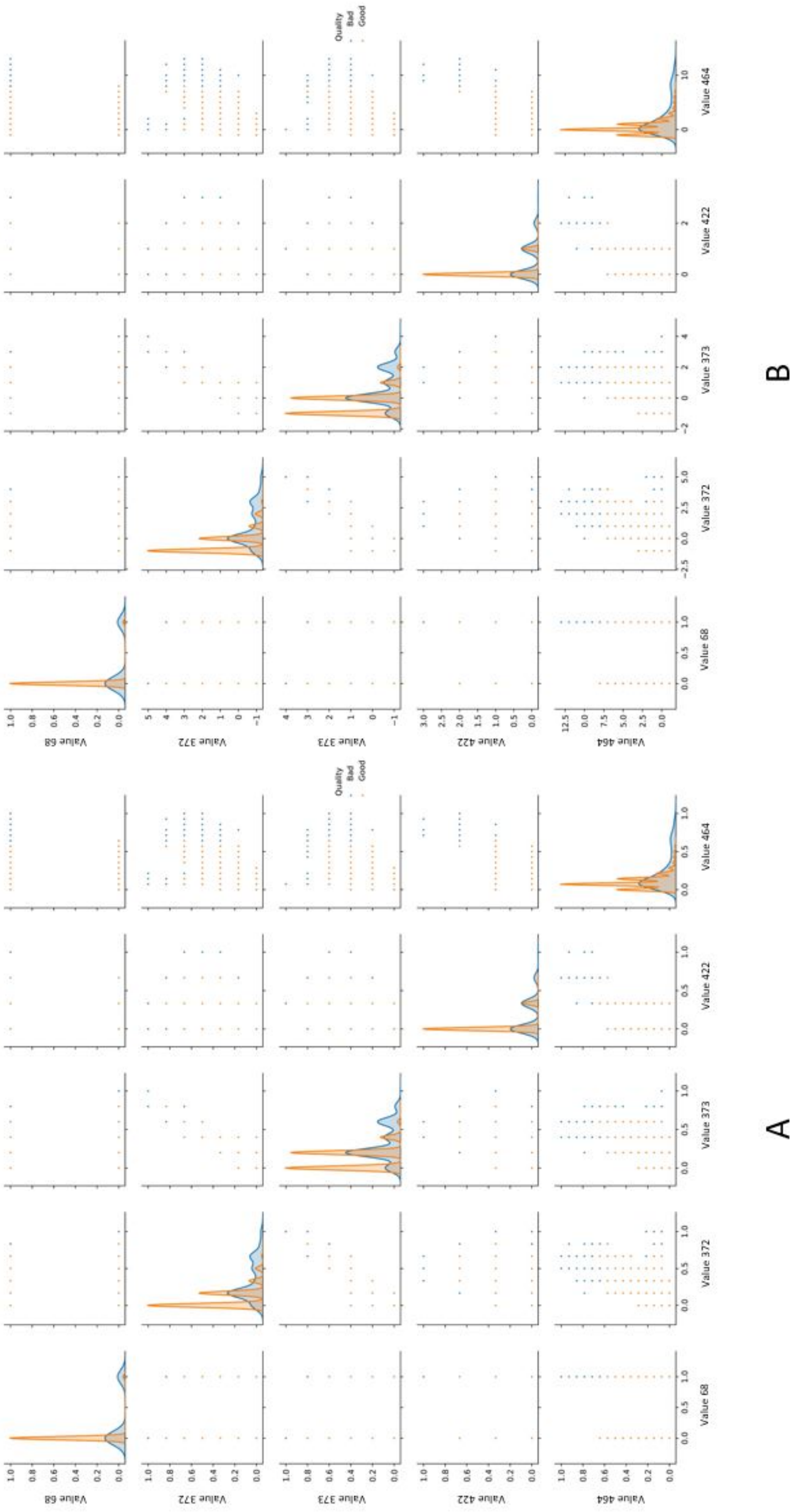

Figure 8.1: SelectKBest with f_regression for A) MinMaxScaler and B) RobustScaler normalization.

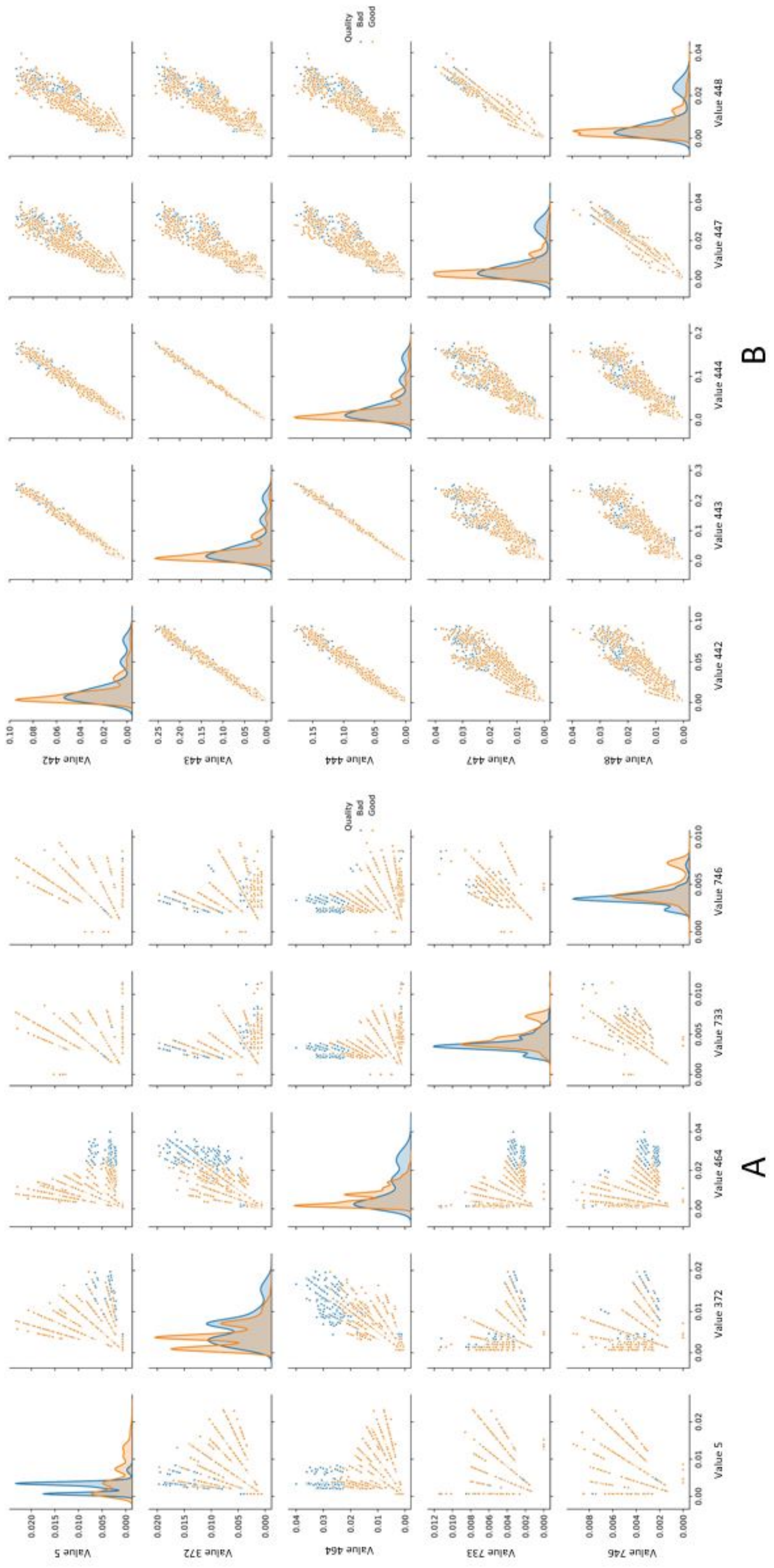

Figure 8.2: Feature selection using the SelectKBest function with A) f_regression and B) f_mutual_info.

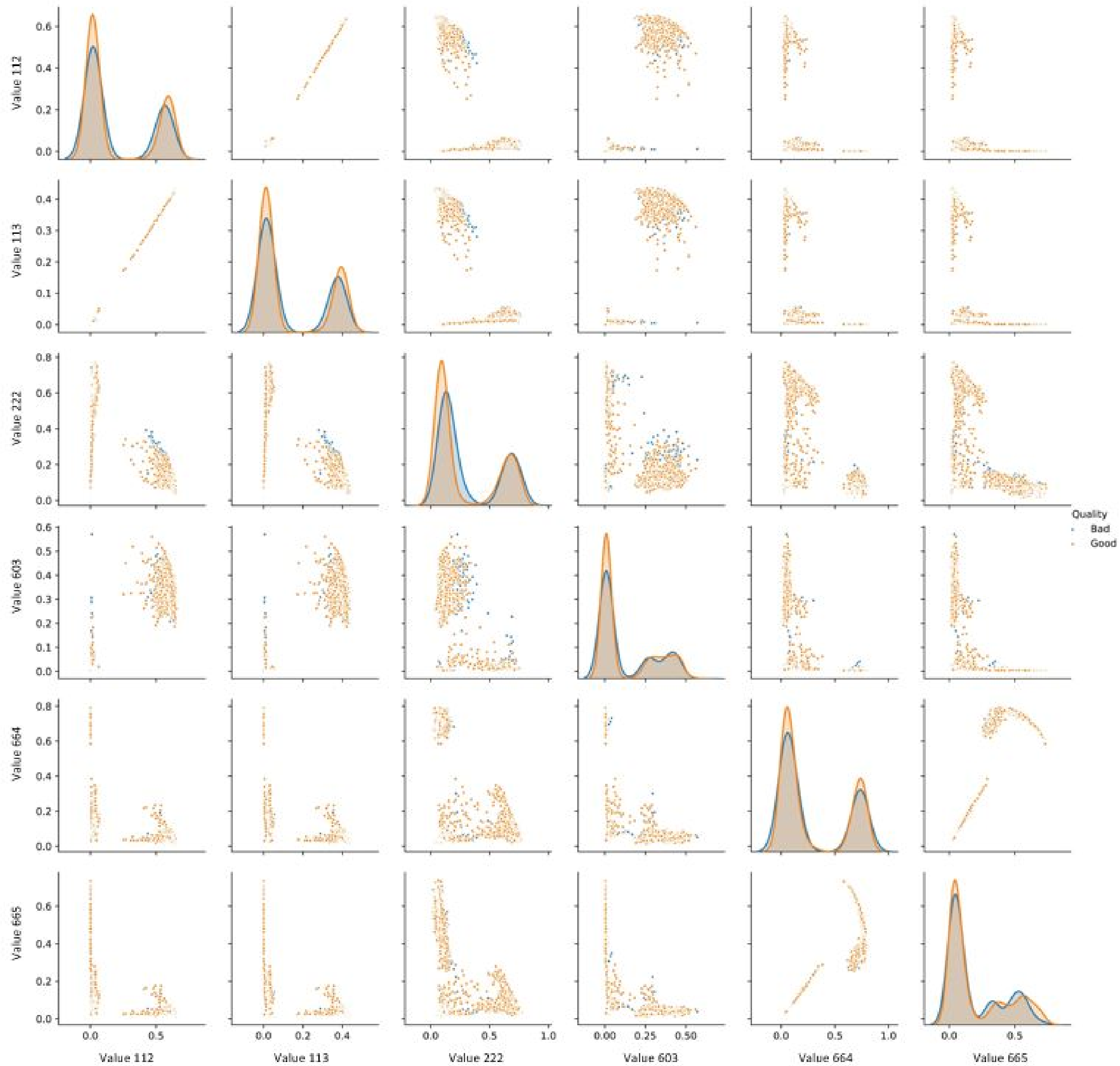

Figure 8.3: Feature selection using the variance threshold function with a threshold of 0.0291.

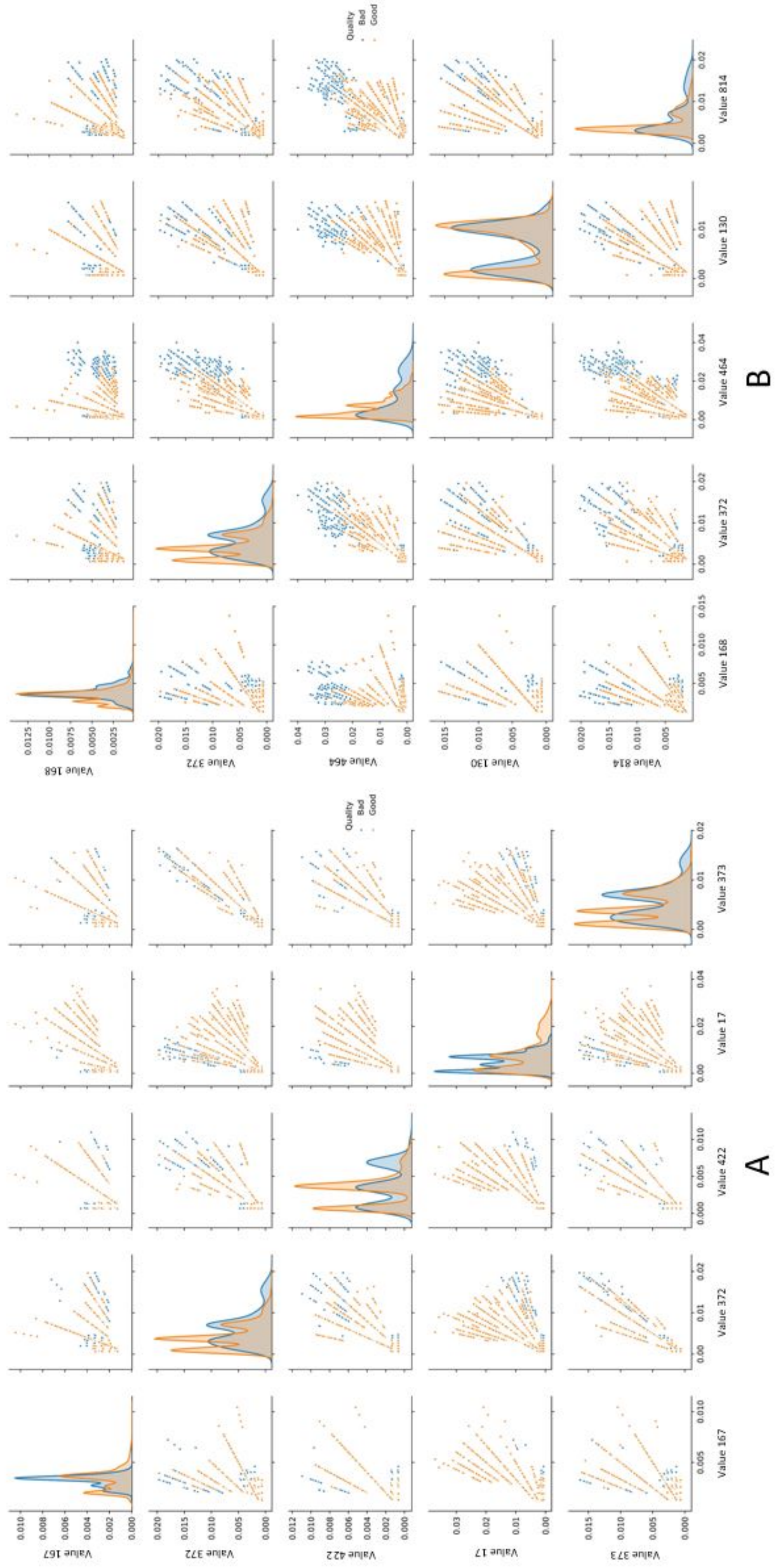

Figure 8.4: Feature importance calculation with the extra-trees classifier with A) 100 and B) 200 estimators.

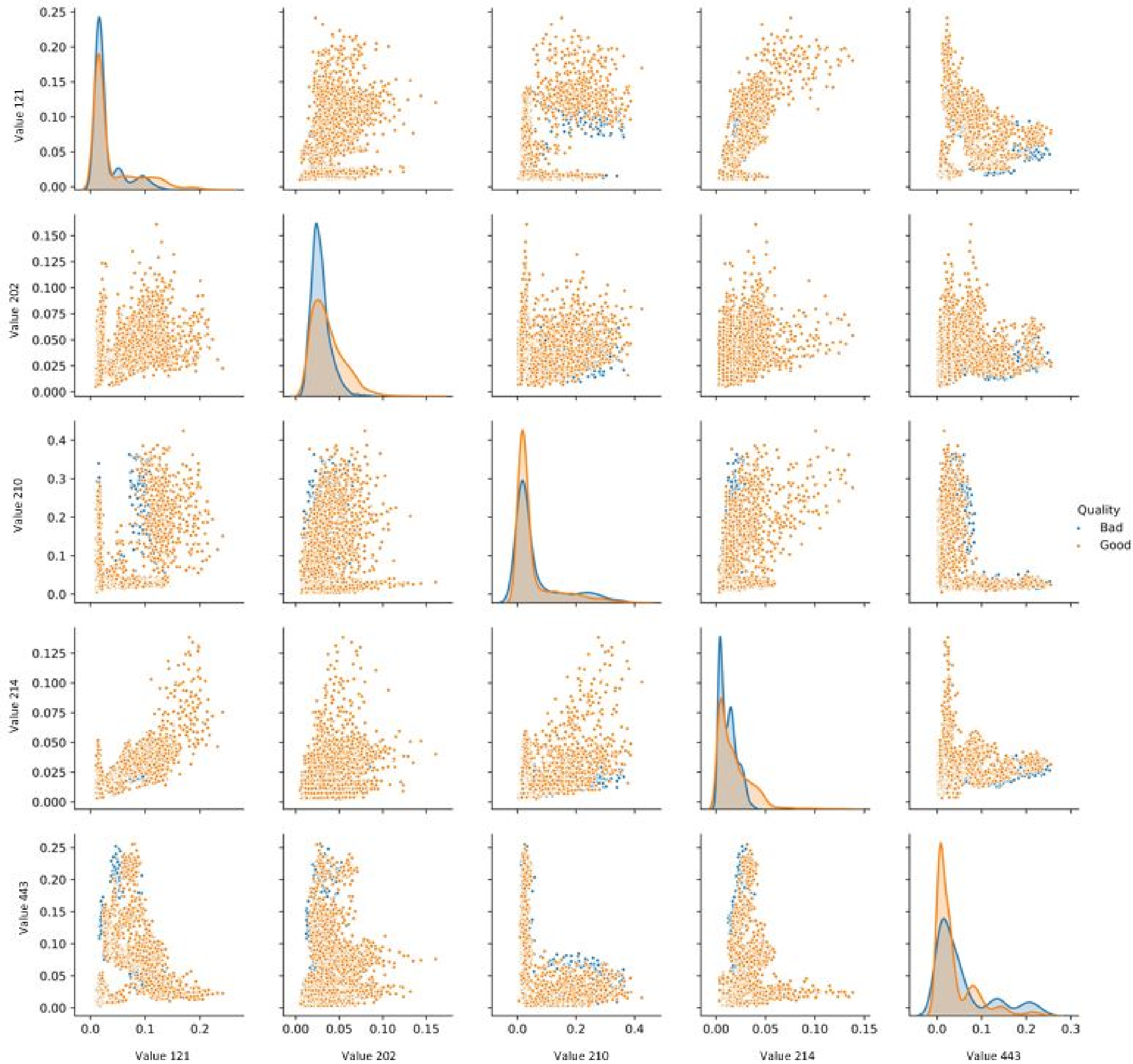

Figure 8.5: Feature extraction with the recursive feature elimination in combination with logistic regression.

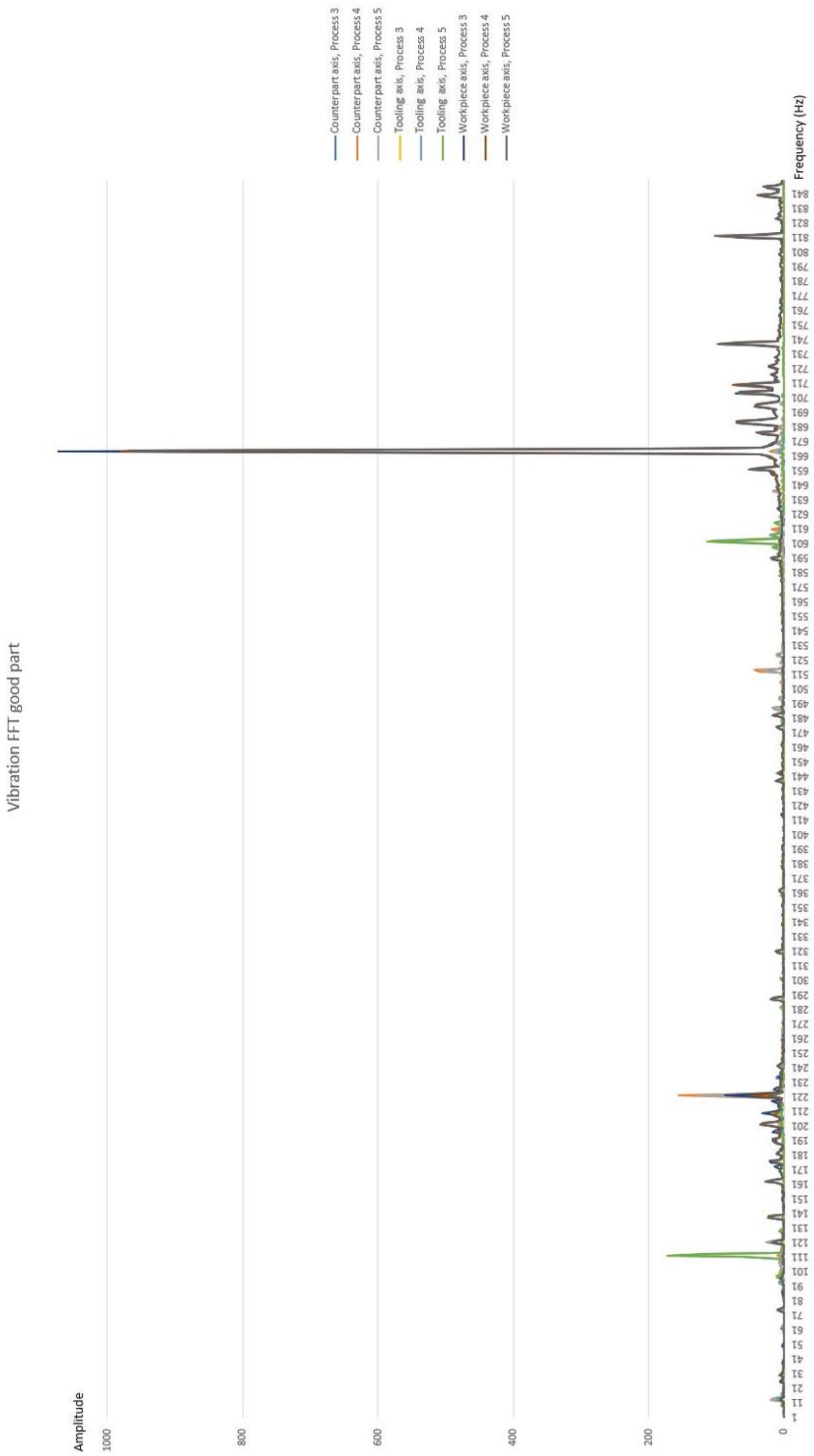

Figure 8.6: Comparison of the FFT spectra obtained from the different axes and processes for a good part.

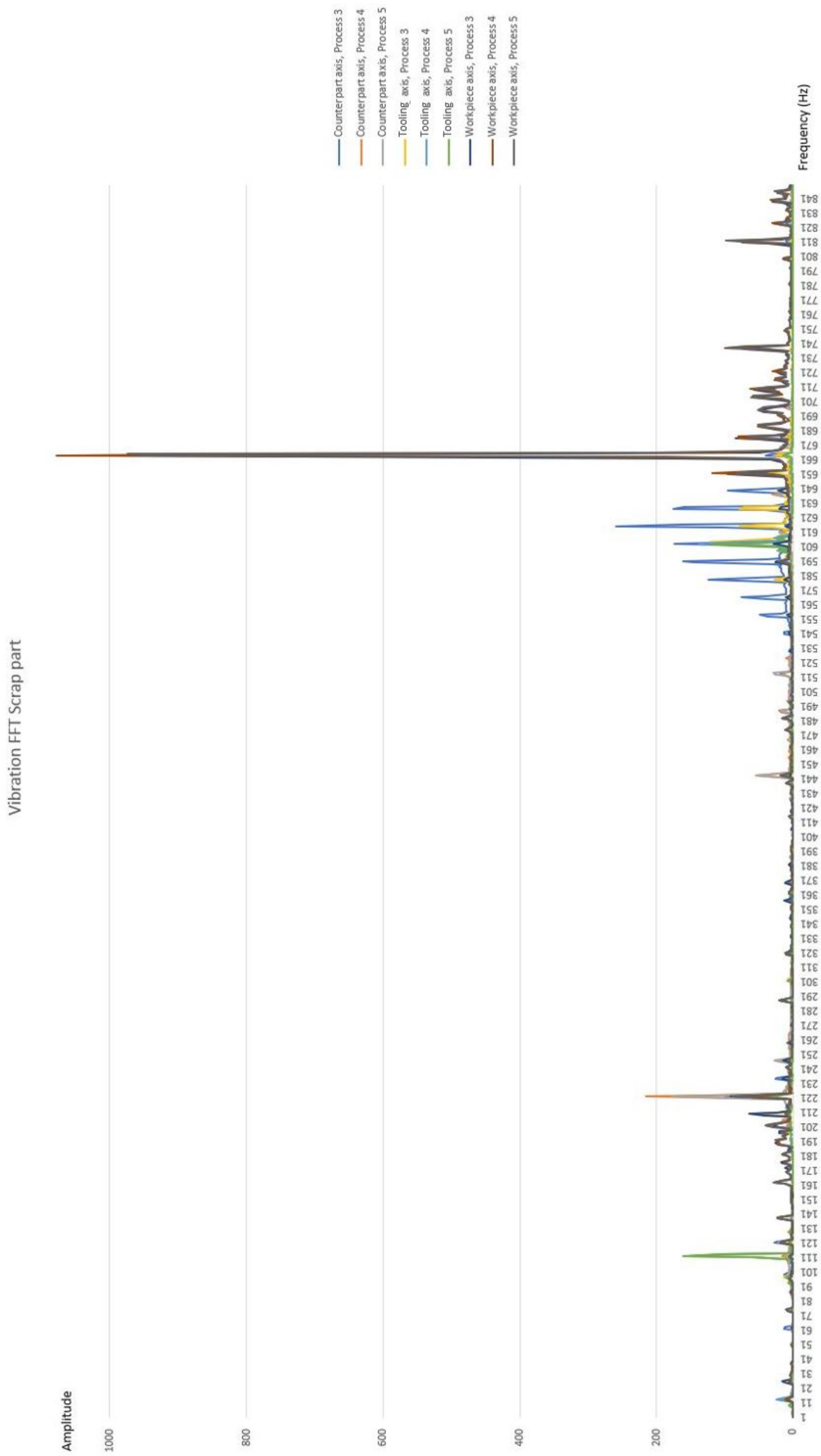

Figure 8.7: Comparison of the FFT spectra obtained from the different axes and processes for a scrap part.

Bibliography

Aliustaoglu, Cuneyt, H. Metin Ertunc, and Hasan Ocak (2009). "Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system." In: *Mechanical Systems and Signal Processing* 23.

American Bureau of Shipping (2016). *Equipment condition monitoring techniques*.

Brownlee, Jason (2018). *Feature Selection For Machine Learning in Python*. Also subsites are used. URL: `https://machinelearningmastery.com/feature-selection-machine-learning-python/` (visited on 04/08/2019).

Chelladurai H. and Jain, V. and N. Vyas (2008). "Development of a cutting tool condition monitoring system for high speed turning operation by vibration and strain analysis." In: *The International Journal of Advanced Manufacturing Technology* 37.5, pp. 471–485. URL: `https://doi.org/10.1007/s00170-007-0986-z`.

Chen, Hongjiang (2011). "Investigation of the Methods for Tool Wear On-Line Monitoring during the Cutting Process." In: *Computer and Computing Technologies in Agriculture IV*. Ed. by Daoliang Li, Yande Liu, and Yingyi Chen. Springer Berlin Heidelberg, pp. 215–220. ISBN: 978-3-642-18369-0.

Coronando, Diego and Katharina Fischer (2015). *Condition Monitoring of Wind Turbines: State of the Art, User Experience and Recommendations*.

Dimla, Eric (2000). "Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods." In: *International Journal of Machine Tools and Manufacture* 40.8, pp. 1073–1098. URL: `http://www.sciencedirect.com/science/article/pii/S0890695599001224`.

DIN 13306 (2010). *Maintenance - Maintenance terminology*. British Standard Institution.

DVS Tooling (2019). *Vario Speed Dresser. Finished with DVS laser cut. - Precise dressing with laser-finished VSD-dressing tools*. Also subsites are used. URL: `http://dvs-tooling.dvs-gruppe.com/index.php?id=1034&L=1` (visited on 03/28/2019).

EDUCBA (2019). Also subsites are used. URL: https://www.educba.com/hadoop-vs-mapreduce/ (visited on 03/11/2019).

Elgendy, Nada and Ahmed Elragal (2014). "Big Data Analytics: A Literature Review Paper." In: *Advances in Data Mining - Applications and Theoretical Aspects*. Ed. by Perner P. Springer International Publishing Switzerland. ISBN: 978-3-319-08975-1.

Fredriksson, Gustav and Hanna Larsson (2012). "An analysis of maintenance strategies and development of a model for strategy formulation - A case study." MA thesis. Chalmers University of Technology.

GKN Driveline Bruneck (2019). Also subsites are used. URL: https://www.gknbruneck.com/ (visited on 02/11/2019).

Gleason (2019). *Honing Tools*. Also subsites are used. URL: https://www.gleason.com/en/products/tools/cylindrical/honing/honing-tools (visited on 03/27/2019).

Grafana Labs (2019). *Grafana*. Also subsites are used. URL: https://grafana.com/ (visited on 04/03/2019).

Haber Guerra, Rodolfo Elias, Jose E. Jiménez, C.Ronei Peres, and Jose Alique (2004). "An investigation of tool-wear monitoring in a high-speed machining process." In: *Sensors and Actuators A: Physical* 116.

Han, Jiawei, Michelin Kamber, and Jian Pei (2012). *Data Mining. Concepts and Techniques*. 3rd ed. Elsevier Inc. ISBN: 978-0-12-381479-1.

Hirsch, Andreas (2016). *Werkzeugmaschinen. Anforderungen, Auslegung, Ausführungsbeispiele*. 3rd ed. Springer Fachmedien Wiesbaden. ISBN: 978-3-658-14248-3.

ifm electronic (2019a). *VSA001*. Also subsites are used. URL: https://www.ifm.com/de/en/product/VSA001 (visited on 04/01/2019).

— (2019b). *VSE002*. Also subsites are used. URL: https://www.ifm.com/de/en/product/VSE002 (visited on 04/01/2019).

IndustryArena (2019). *Präwema Synchrofine 205 HS (W)*. Also subsites are used. URL: https://de.industryarena.com/dvstechnology/produkte/honmaschinen--268/praewema-synchrofine-205-hs-w--13266 (visited on 03/27/2019).

InfluxData (2019). *InfluxData Product Overview*. Also subsites are used. URL: https://www.influxdata.com/products/ (visited on 04/03/2019).

Kayaalp, Fatih and Muhammet Başarslan (Apr. 2018). "Open Source Data Mining Programs: A Case Study on R." In: *Duzce University Journal of Science and Technology* 6, pp. 455–468.

Klink, Ulrich (2015). *Honen. Umweltbewusst und kostengünstig Fertigen*. Carl Hanser Verlag. ISBN: 978-3-446-44192-7.

Klocke, Fritz and Christian Brecher (2016). *Zahnrad- und Getriebetechnik. Auslegung - Herstellung - Untersuchung - Simulation*. Carl Hanser Verlag. ISBN: 978-3-446-43068-6.

Konradin Mediengruppe (2012). *Honmaschine beschleunigt die Produktion von Zahnrädern*. Also subsites are used. URL: https://mav.industrie.de/fertigung/honmaschine-beschleunigt-die-produktion-von-zahnraedern-2/ (visited on 03/27/2019).

Kotthoff, Gerd (2018). *NVH Potential of PM Gears for Electrified Drivetrains*. URL: https://www.geartechnology.com/issues/0918x/nvh-pm-gears.pdf.

Luan, Xiaona, Song Zhang, Jianfeng Li, Gamini Mendis, Fu Zhao, and John W. Sutherland (2018). "Trade-off analysis of tool wear, machining quality and energy efficiency of alloy cast iron milling process." In: *Procedia Manufacturing* 26. 46th SME North American Manufacturing Research Conference, NAMRC 46, Texas, USA, pp. 383–393.

Matyas, Kurt (2016). *Instandhaltungslogistik - Qualität und Produktivität steigern*. Carl Hanser Verlag München. ISBN: 978-3-446-44616-8.

Melrose Industries PLC (2019). Also subsites are used. URL: https://www.gkn.com/ (visited on 02/11/2019).

Michael Waskom (2018). *Seaborn Pairplot*. Also subsites are used. URL: https://seaborn.pydata.org/generated/seaborn.pairplot.html (visited on 04/29/2019).

National Instruments (2019a). Also subsites are used. URL: http://www.ni.com/example/6511/en/ (visited on 03/06/2019).

— (2019b). Also subsites are used. URL: https://www.ni.com/data-acquisition/what-is/d/ (visited on 03/11/2019).

OASIS Open (2014). *MQTT Version 3.1.1*. Also subsites are used. URL: http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html (visited on 04/03/2019).

PAT Research (2018). Also subsites are used. URL: https://www.predictiveanalyticstoday.com/data-ingestion-tools/ (visited on 03/11/2019).

Pawellek, Günther (2013). *Integrierte Instandhaltung und Ersatzteillogistik - Vorgehensweisen, Methoden, Tools*. Springer-Verlag Berlin Heidelberg. ISBN: 978-3-662-48666-5.

Payne, Jeff (2013). *PLCs and PACs Simplify Data Acquisition: Powerful PLCs and PACs have built-in data acquisition, data logging and data analysis functionality*. URL: https://cdn.automationdirect.com/static/press/plcs-and-pacs-white-paper.pdf (visited on 03/11/2019).

Perovic, Bozina (2009). *Spanende Werkzeugmaschinen*. Springer Berlin Heidelberg. ISBN: 978-3-540-89951-8.

Präwema Antriebstechnik (2016). *Präwema Gear Honing SynchroFine*. Also subsites are used. URL: http://praewema.dvs-gruppe.com/uploads/tx_xpctypedownloadssimple/PRAEWEMA-Antriebstechnik-Verzahnungshonen-SynchroFine-EN.pdf (visited on 03/28/2019).

— (2018). *Orginalbetriebsanleitung*. Mit freundlicher Genehmigung von Präwema Antriebstechnik GmbH.

Scheffer, C., H. Kratz, P.S. Heyns, and F. Klocke (2003). “Development of a tool wear-monitoring system for hard turning.” In: *International Journal of Machine Tools and Manufacture* 43.10.

Schenk, Michael (2010). *Instandhaltung technischer Systeme*. Springer-Verlag Berlin Heidelberg. ISBN: 978-3-642-03948-5.

Schmitt, Christina, Sven Klein, and Dirk Bähre (2015). “An Introduction to the Vibration Analysis for the Precision Honing of Bores.” In: *Procedia Manufacturing* 1. 43rd North American Manufacturing Research Conference, pp. 637–643. URL: http://www.sciencedirect.com/science/article/pii/S2351978915010562.

scikit-learn developers (2018a). *Compare the effect of different scalers on data with outliers*. Also subsites are used. URL: https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html#sphx-glr-auto-examples-preprocessing-plot-all-scaling-py (visited on 04/08/2019).

— (2018b). *Comparision of F-test and mutual information*. Also subsites are used. URL: https://scikit-learn.org/stable/auto_examples/feature_selection/plot_f_test_vs_mi.html#sphx-glr-auto-examples-feature-selection-plot-f-test-vs-mi-py (visited on 04/08/2019).

scikit-learn developers (2018c). *Extra-trees Classifier*. Also subsites are used. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html (visited on 04/08/2019).

— (2018d). *Feature selection*. Also subsites are used. URL: https://scikit-learn.org/stable/modules/feature_selection.html (visited on 04/08/2019).

— (2018e). $Mutual_info_regression$. Also subsites are used. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_regression.html#sklearn.feature_selection.mutual_info_regression (visited on 04/08/2019).

— (2018f). *Variance Threshold*. Also subsites are used. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.VarianceThreshold.html (visited on 04/08/2019).

The Apache Software Foundation (2018a). *Apache NiFi*. Also subsites are used. URL: https://nifi.apache.org/ (visited on 04/03/2019).

— (2018b). *Apache Spark*. Also subsites are used. URL: https://spark.apache.org/ (visited on 04/03/2019).

— (2019). *Apache Zeppelin*. Also subsites are used. URL: https://zeppelin.apache.org/ (visited on 04/03/2019).

Wilson, Brian (2015). *NVH Analysis Within the Design Process*. Also subsites are used. URL: https://www.geartechnology.com/articles/0515/NVH_Analysis_Within_the_Design_Process/ (visited on 03/25/2019).

Zanger, Frederik (2013). *Segemtspanbildung, Werkzeugverschleiß, Randschichtzustand und Bauteileigenschaften: Numerische Analysen zur Optimierung des Zerspanungsprozesses am Beispiel von Ti-6AI-4V*.