NTNU
Norwegian University of
Science and Technology

Utilisation of Machine Learning in Power Transformer Asset Management

Martine Ukkelberg

Master of Energy and Environmental Engineering
Submission date: June 2018
Supervisor: Eivind Solvang, IEL
Co-supervisor: Maren Istad, SINTEF Energy Research
Espen Eberg, SINTEF Energy Research

Norwegian University of Science and Technology
Department of Electric Power Engineering

Preface

This Master's thesis is the conclusion of my Master of Science degree in Energy and Environmental Engineering with the Department of Electric Power Engineering at the Norwegian University of Science and Technology (NTNU). The work was carried out in the spring semester of 2018, and has been performed in collaboration with SINTEF Energy Research and Statnett. The thesis is a continuation of my specialisation project, and involves examining potential opportunities obtained by better management of big data and utilisation of machine learning for power transformer asset management purposes.

The assumed background and knowledge expected of the reader are that of a fellow 5th-year electric power engineering student.

Firstly, I would like to thank Arne Smisethjell at Statnett and the project "Smarter Asset Management using Big Data (SAMBA)" for giving me the opportunity to undertake this work. I would also like to thank Boye Annfelt Høverstad at Statnett for providing data and helping me in the approach to understand and use them.

I am grateful to Eivind Solvang, my supervisor at NTNU. He has been my main source of guidance and support during the development of this thesis. I would also like to thank Espen Eberg, Maren Istad and Gjert Hovland Rosenlund, Research Scientists at SINTEF Energy Research, for their helpful contributions and advice.

Finally, I would like to thank my family and friends for their love, support and motivation during my 5 years at NTNU.

Trondheim, June 2018

Martine Ukkelberg

Martine Ukkelberg

Abstract

Ageing assets in the power system increase the need for maintenance and reinvestments. There is currently a shortage of adequate data and analysis systems available for estimation of condition and remaining lifetime, to facilitate decision-making. It is a challenge to restructure asset management with regard to collecting relevant data and to introduce new systems for handling and analysing the data. There is a large potential for increased value creation through more online and automatic collection and analysis of condition values.

This thesis examines potential opportunities obtained by better management of big data and utilisation of machine learning for power transformer asset management purposes. Power transformers are expensive and important components of the power system, and an increasing volume of data concerning their condition is becoming available. Transformers also have the potential for longer lifetimes. The introduction of data analysis using machine learning and management of big data enables condition monitoring of components to a greater extent than was previously possible. As a result, power companies can potentially optimise maintenance, and postpone reinvestments by adding resources where needed.

After presenting the theoretical framework, two case studies are performed. A machine learning model was developed to predict transformer hot-spot temperature for a chosen transformer, using available data. Furthermore, predicted hot-spot temperatures were used to estimate winding insulation degradation for different scenarios, including increased load and increased and decreased ambient temperature. These estimates were further used to determine the associated remaining lifetimes. The emerged utility value is also presented.

The results show that machine learning models are able to predict transformer hot-spot temperature with satisfactory accuracy, compared to the hot-spot temperature measured by a fiber optic sensor. The models can also be used to determine the maximum acceptable loading. Change of hot-spot temperature has a large impact on the estimated remaining lifetime, with higher hot-spot temperature leading to accelerated ageing. The predicted hot-spot temperature was also proved to correspond better with the measured hot-spot temperature than the ones obtained by the commonly used loading guide.

Better management of big data and utilisation of machine learning creates many new opportunities. Machine learning can be implemented in already existing activities to increase efficiency and accuracy, and to reduce uncertainty. New applications can also arise as the power system becomes more digital, with emerging use of sensors. Examples are normal behaviour models, sensor verification, and loading determination. Good ICT-structure and data of quality are necessary for these purposes. It is important to emphasise that predictions of condition and lifetime obtained using machine learning are estimates. This needs to be considered when making decisions involving maintenance and reinvestments.

Sammendrag

Aldring av komponenter i kraftsystemet øker behovet for vedlikehold og reinvesteringer. I dag er det mangel på tilstrekkelige data og analysesystemer for estimering av tilstand og restlevetid for å legge til rette for beslutningstaking. Det er en utfordring å omstrukturere anleggsforvaltning med hensyn til innsamling av relevante data og å innføre nye systemer for å håndtere og analysere dataene. Det er et stort potensial for økt verdiskaping gjennom mer online og automatisk innsamling og analyse av tilstandsverdier.

Denne masteroppgaven undersøker potensielle muligheter ved bedre håndtering av store datamengder og utnyttelse av maskinlæring for forvaltning av krafttransformatorer. Krafttransformatorer er kostbare og viktige komponenter i kraftsystemet, og en økende mengde data om deres tilstand blir tilgjengelig. Transformatorer har også potensial for lengre levetid. Innføringen av dataanalyse ved bruk av maskinlæring gir mulighet for tilstandsovervåking av komponenter i større grad enn tidligere mulig. Som et resultat kan kraftselskaper potensielt optimalisere vedlikehold og utsette reinvesteringer ved å øke ressurser ved behov.

Etter å ha presentert det teoretiske rammeverket, er to case-studier utført. En maskinlæringsmodell er utviklet for å predikere transformator hot-spot temperatur for en valgt transformator, ved bruk av tilgjengelige data. Videre ble predikerte hot-spot temperaturer benyttet til å estimere viklingsisolasjonsdegradering for forskjellige scenarier, inkludert økt last og økt og redusert omgivelsestemperatur. Disse estimatene ble ytterlige brukt til å fastslå tilhørende restlevetider. Den fremkomne nytteverdien er også presentert.

Resultatene viser at maskinlæringsmodeller kan predikere transformator hot-spot temperatur med tilfredsstillende nøyaktighet, sammenlignet med den målte hot-spot temperaturen fra en fiberoptisk sensor. Modellene kan også brukes til å fastslå maksimal akseptabel last. Endring i hot-spot temperatur har stor innvirkning på estimert levetid, der høyere hot-spot temperatur fører til raskere aldring. Den predikerte hot-spot temperaturen viste seg også å korrespondere bedre med den målte hot-spot temperaturen enn de som ble oppnådd ved bruk av den ofte brukte modellen "loading guide".

Bedre håndtering av store datamengder og utnyttelse av maskinlæring åpner for mange muligheter. Maskinlæring kan implementeres i allerede eksisterende aktiviteter for å øke effektiviteten og nøyaktigheten, og redusere usikkerheten. Nye applikasjoner kan også oppstå ettersom kraftsystemet blir mer digitalt med større bruk av sensorer. Eksempler er normal-oppførselsmodeller, sensor verifisering og bestemmelse av lastgrense. God IKT-struktur og data kvalitet er nødvendig for disse formålene. Det er viktig å vektlegge at prediksjoner av tilstand og levetid oppnådd ved bruk av maskinlæring er estimater. Dette må tas med i betraktning når en tar beslutninger om vedlikehold og reinvesteringer.

Contents

List of Figures

List of Tables

Listings

Acronyms

AI	Artificial Intelligence
ANNs	Artificial Neural Networks
CRISP-DM	Cross-Industry Standard Process for Data Mining
DP	Degree of Polymerisation
FAT	Factory Acceptance Test
GLM	Generalised Linear Models
ICT	Information and Communication Technology
IED	Intelligent Electronic Device
MAE	Mean Absolute Error
NaN	Not a Number
ONAF	Oil Natural Air Forced
ONAN	Oil Natural Air Natural
PV	Present Value
RMSE	Root Mean Squared Error

Nomenclature

A	Environment factor [h^{-1}]
$\overline{A}$	Average environment factor [h^{-1}]
α	Complexion parameter
α_N	Discounting factor
DP_{new}	Degree of polymerisation when the transformer was new
DP_{today}	Degree of polymerisation today
E_a	Activation energy [J mol^{-1}]
ϵ	Penalty distance
ε_T	Annuity factor
g_r	Average-winding-to average-oil [K]
H	Hot-spot factor
K	Load factor (load current/rated current)
M	Number of hot-spot temperature measurements
N	Analysis period [years]
O	Multiplication factor for the presence of oxygen = 2.5
R	Loss ratio
R^2	Coefficient of determination
R_g	Gas constant = 8.31 [J mol^{-1} K^{-1}]
RC	Reinvestment cost
T	Economic lifetime [years]
t	Time [hours]
$w_{o,n}$	Concentration of water in oil [%]

$w_{p,n}$	Concentration of water in paper [%]
x	Oil exponent
y	Winding exponent
y_i	Measured value
$\hat{y}_i$	Predicted value
$\overline{y}$	Mean of the measured values
θ_a	Ambient temperature [°C]
θ_h	Hot-spot temperature [°C]
θ_o	Top-oil temperature [°C]
$\Delta\theta_{or}$	Rated top-oil temperature rise [K]
λ_N	Capitalisation factor
ρ	L1 ratio
τ	Age of transformer [hours]
ϕ	Incremental factor for the concentration of water in paper
ω	Coefficients
ω_o	Intercept

Chapter 1

Introduction

1.1 Background

Power companies are in a period where many ageing assets in the power system increase the need for maintenance and reinvestments. This further leads to increased requirements for utilisation of resources and economic savings. Today, there is a shortage of adequate data and analysis systems available for estimation of condition and remaining lifetime, to facilitate decision-making. There is a large potential for increased value creation through more online and automatic collection and analysis of condition values. These data will be essential for estimating asset's condition, the probability of failure and the remaining lifetime [1].

Improvements in asset management imply increased, systematic, predictive and risk-based maintenance as well as reinvestment decisions. This can lead to major economic gains [1].

It is a challenging task for power companies including Statnett to restructure asset management with regard to collecting relevant data and to introduce new systems for handling and analysing the data. Based on this, the three-year research project "Smarter Asset Management with Big Data (SAMBA)" was initiated by Statnett. The project has several partners and will create an arena for knowledge sharing between Statnett, SINTEF Energy Research and some of the largest providers of systems for asset management. In this way, the project will help to improve the decision-making processes within asset management in Statnett. This will be done by combining innovations in information and communication technology (ICT), and electric engineering knowledge [1].

Statnett is planning to build approximately 20 new power transformer stations by 2021, while several of the existing transformers require reinvestment as a result of reaching remaining lifetime expectancy [2]. Power transformers are expensive and important components of the power system, but they have the potential for long lifetimes, around 80 years, under ideal conditions with low load, proper cooling and low water penetration

[3]. Such long lifetimes can be safely utilised with improved asset management. This will potentially lead to an improved estimation of condition and remaining lifetime, as well as improved decision-making processes for maintenance and reinvestment. The correct measuring equipment and analysis systems must be selected to reach these objectives. Choosing unsuitable equipment and systems can later lead to the wrong decisions being made.

An increasing volume of data from power transformers concerning their condition is currently becoming available. The introduction of data analysis using machine learning and management of big data enables condition monitoring of components to a greater extent than was previously possible. Today, maintenance in Statnett is normally performed on a regular basis with given routines and time intervals. In addition, some components are replaced after a certain period of time, generally based on a predetermined technical age. By more efficient condition monitoring of its components, Statnett can optimise maintenance, and postpone reinvestments by adding resources where needed. The utilisation of machine learning and better management of big data on components can therefore potentially lead to considerable economic savings.

Machine learning is a part of data science which is defined by [4] as "the extraction of actionable knowledge directly from data through a process of discovery, or hypothesis formulation and hypothesis testing". This process is called data mining, while machine learning is a commonly used method. Figure 1.1 shows the skills needed in data science. As an electric power engineering student, the domain expertise will be the focus of this thesis work.

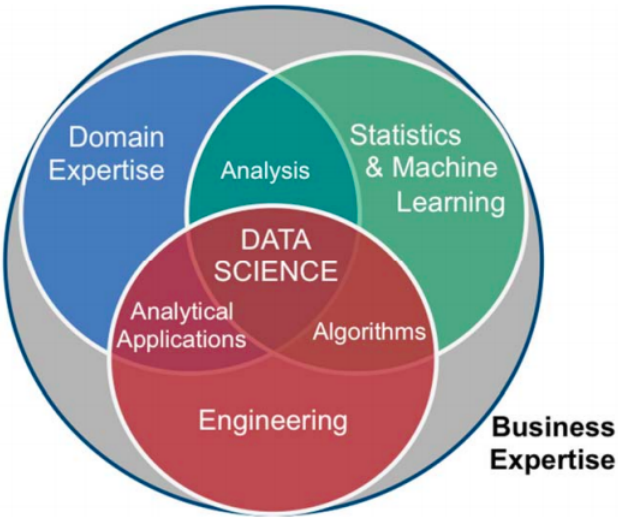

Figure 1.1: Skills needed in data science [4].

1.2 Problem Description

This thesis work is a part of the SAMBA research project and involves examining potential opportunities obtained by better management of big data and utilisation of machine learning for power transformer asset management purposes. This is done in order to determine if machine learning is an alternative method to already existing models. It can optionally be one of several tools used.

The motivation is to develop machine learning models to describe physical conditions based on big data. Using such models actively in decision-making will help developing safe and efficient operations of power transformers.

The aim is that such a model can be able to use real-time data from power transformers in order to predict the temperature development of the given transformer in a more efficient way than already existing methods. This can further be used to estimate the condition and remaining lifetime of the transformer for asset management purposes. For this purpose, there exists a useful data base in addition to the correlations being relatively clear, making it possible to have a physical understanding of the relationships and results.

The objectives of this thesis are:

- To describe the method of asset management involving condition monitoring, measurements, estimation, and instruments.
- To describe typical ageing mechanisms and faults of power transformers.
- To obtain information about measuring methods and models used for power transformers.
- To describe the concept and method of machine learning.
- To obtain information about data mining processes.
- The development of machine learning models to predict the transformer temperature response.
- The estimation of transformer condition and remaining lifetime for different scenarios.
- To evaluate machine learning as an alternative method to already existing models used for power transformers.
- To evaluate the utility value and benefits of using machine learning for power transformer asset management purposes.

1.3 Approach and Limitations

To approach and meet the stated objectives, the thesis work started with a literature review, and the study of relevant theory for asset management, power transformers, machine learning and a relevant data mining process. In the thesis work, it has been important to use both Norwegian and international sources and standards to get the best possible access to information, while at the same time focusing on Norwegian conditions. Books, articles and conversations with my supervisors, SINTEF Energy Research researchers and Statnett employees, have all been sources of information. For information concerning power transformers, IEEE *Xplore* Digital Library has been visited frequently. As machine learning is a field of study that is developing rapidly, online sources including tutorials have also been used as background material, as traditional textbooks were not sufficient. Furthermore, the literature and theory from various sources have been combined and processed.

Remark: The thesis work builds on the theory and literature review carried out as part of my specialisation project [5], and as such, there is an extensive usage of the content therefrom.

Based on the knowledge obtained from the literature and theory, it was decided to focus on the insulation condition of power transformers. Aged insulation is one of the main causes of failure. As the insulating paper is subject to faster ageing than the insulating oil, the strength of the paper is considered to be the final limiting factor for the remaining lifetime of the transformer. The strength of the paper is strongly influenced by hot-spot temperature. In this thesis work, there is therefore an emphasis on methods and models used for determining the ageing of the insulating paper, and the hot-spot temperature development.

Furthermore, two case studies have been performed - for developing machine learning models to predict transformer temperature development, and for the estimation of condition and remaining lifetime. Data provided by Statnett has been used, together with the processed literature and theory. Cross-Industry Standard Process for Data Mining (CRISP-DM) was chosen as the data mining method, as it is it a leading methodology within data mining. In addition, Python was chosen as the programming language for examining the data due to the format of the data, Statnetts experience with Python, and Python's ability to work with data science and machine learning.

Some limitations are included in the approach to the objectives. The input data for the case studies consists of measured values, which can be subject to measurement errors. In addition, as explained in Section 1.1, machine learning is a part of data science that requires several skills. As the domain expertise (electric power engineering) is the focus of the thesis work, insufficient machine learning skills can be a limitation for the accuracy of the results. Finally, a few simplifications and assumptions had to be made when using different methods and models on the data. These are presented during the case studies and in the discussion.

1.4 Outline

The outline of the thesis is shown in Figure 1.2. The thesis is built around the case studies in Chapter 3, where the goal is to give an answer to the problem description. Chapter 2 is used as a basis to better understand the content of these case studies. Chapter 4 discusses the utility value and benefits which emerged from the case studies. The results are further discussed in Chapter 5 before the final conclusion is given in Chapter 6.

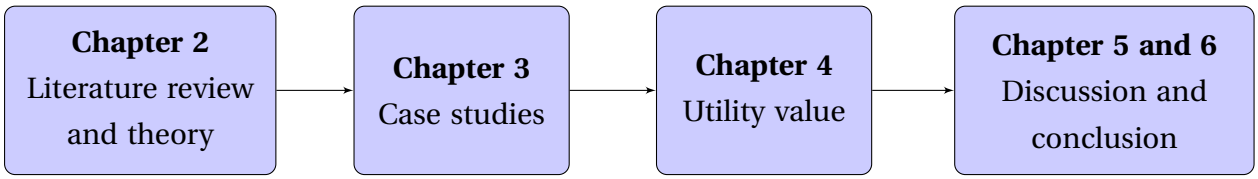

Figure 1.2: Thesis outline.

In **Chapter 2**, the method of asset management involving condition monitoring, measurements, estimation, and instruments has been described. This is done to create a fundamental understanding of the general procedures needed when estimating condition and remaining lifetime, and making decisions. Furthermore, typical ageing mechanisms and faults of power transformers are described in order to understand the underlying reasons and physics behind the relevant measuring methods and models used to detect the ageing and/or faults, which are also presented. Finally, the concept and method of machine learning are described in addition to a relevant data mining process.

Chapter 3 is based on the knowledge obtained from Chapter 2 and includes two case studies performed to demonstrate the applicability of machine learning in power transformer asset management. The aim is to determine if better management of big data and utilisation of machine learning is an alternative method to already existing models for power transformer asset management purposes. A machine learning model has been developed to predict transformer hot-spot temperature for a chosen transformer, using available data. The maximum acceptable loading was also determined using the model. Furthermore, predicted hot-spot temperatures were used to estimate winding insulation degradation for different scenarios, including increased load and increased and decreased ambient temperature. These estimations were further used to determine the associated remaining lifetimes. The predicted and measured hot-spot temperatures were also compared to calculated hot-spot temperatures obtained using the loading guide.

Chapter 4 presents the utility value and benefits which emerged from the case studies in Chapter 3. Three uses of machine learning are proposed: as a normal behaviour model, for loading determination and as an alternative to the loading guide. In addition, one small economic analysis has been performed.

The results obtained are discussed in **Chapter 5** before the final conclusion and recommendations for further work are given in **Chapter 6**.

Chapter 2

Literature Review and Theory

2.1 Asset Management

Asset management involves balancing costs, opportunities and risks against the desired performance of assets, to achieve the organisational objectives. It enables the application of analytical approaches towards managing an asset over the different stages of its life cycle. This includes the conception of the need for the asset, through to its disposal [6]. The general life cycle of an asset is shown in Figure 2.1.

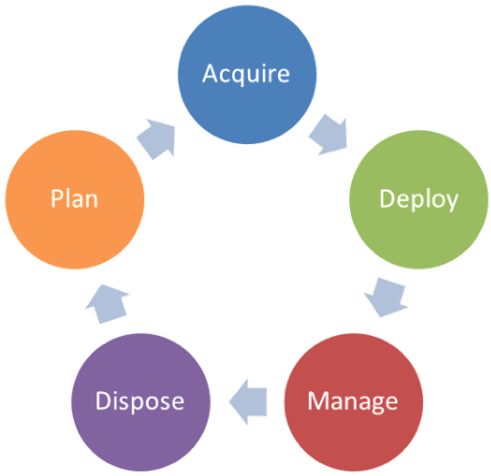

Figure 2.1: General life cycle of an asset [7].

Some of the benefits of asset management include improved financial performance, improved services, and outputs, managed risk and informed asset investment decisions. It gives an organisation the opportunity to improve its decision-making and effectively balance costs, risks, opportunities and performance. Asset management involves making the right decisions and optimising the delivery of value based on measurements and monitoring [6].

A more specific example of an asset life cycle is shown in Figure 2.2. The figure illustrates the technical condition of an asset over time, with maintenance and reinvestment performed to increase the performance and life expectancy of the asset.

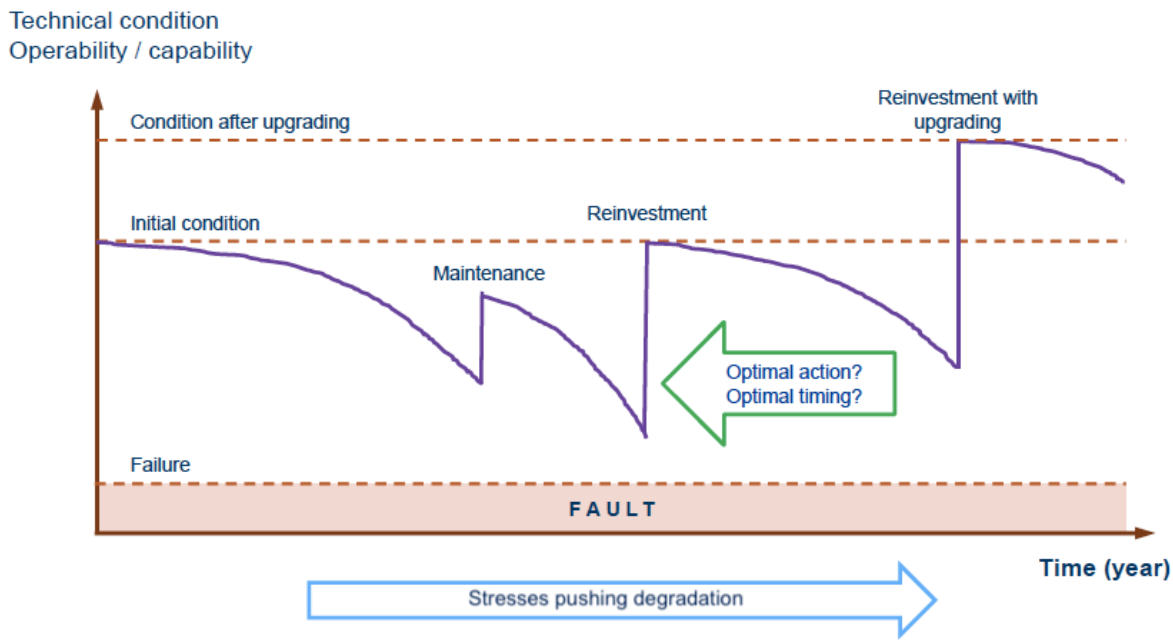

Figure 2.2: Specific example of an asset life cycle [8].

Maintenance is defined by [9] as "the combination of all technical, administrative and managerial actions during the life cycle of an asset intended to retain it in, or restore it to, a state in which it can perform the required function".

Asset management is an important task performed by power companies to ensure a reliable and secure operation of the power system. The goal is to repair and upgrade power system components in an efficient and timely manner such that component failures, and consequently power outages, are reduced. The topic is today more important than ever as the power system infrastructure, mainly built in the 1950s and 60s (both in Norway and internationally), is ageing and the consumers' expectations of a reliable and high-quality service are high [10].

The volume of measurements and other data available for asset management purposes is currently increasing and becoming very large. Big data are data sets so voluminous and complex that traditional data processing applications are inadequate. Big data is normally described by the following characteristics [11]:

- Volume - The quantity of generated and stored data.
- Velocity - The speed at which new data is generated and processed.
- Variety - The different types of data are available from different sources, in different

formats, and of different quality.

- Veracity - The data quality of captured data.
- Value - The ability to turn data into value.

Challenges include capturing data, data storage and data analysis. Big data will nevertheless also lead to great opportunities, such as the use of machine learning [12]. In contrast, when data are missing, asset management decisions are made difficult and more uncertain. Modern asset management based on comprehensive measurement and monitoring of technical condition (big data) depends on good and accurate measurements used for estimating the condition and remaining lifetime.

Transmission of accurate, reliable measurements is important for a safe, efficient and economical operation of the power system. The power system is undergoing extensive digitalisation, enabling digital substations. Digital substations are based on control and protection intelligent electronic devices (IED). By using digital substations, real-time management signals and status information can be transferred throughout the substation without complex wiring schemes. The increasing amount of data available in digital substations need better solutions to transform this data into actionable information, and to make sure the data is properly and securely managed [13], [14].

2.1.1 Condition Monitoring

Condition monitoring plays an important role in asset management and is defined by [15] as "continuous or periodic measurement and interpretation of data to indicate the degraded condition (potential failure) of an item and the need for maintenance".

Condition monitoring is normally carried out with the item in operation, in an operating state, or removed, but not subject to dismantling [15].

Typical methods for condition monitoring are measurements of material properties, and chemical, electrical or mechanical responses. In some cases, large amounts of data are collected. In order to use this information effectively, it is important that the analysis and visualisations are good and well-organised [16].

Some of the advantages of condition monitoring are, as stated in [16]:

- Increased reliability and availability.
- More efficient maintenance processes.
- Optimised maintenance planning.
- Improved resource planning.

The collected data is analysed to find patterns that may indicate that failure is incipient or that the asset is already in a faulty state. Knowing the condition of an asset makes it possible to try to predict failure. Thus, maintenance planning can be done in advance of failure. These factors lead to reduced downtime and greater opportunities for the organisation to plan resource utilisation. This increases the maintenance capacity and reliability [16]. There is, however, considerable uncertainty associated with condition monitoring due to inaccurate measurements and estimations.

2.1.2 Measurement and Estimation

Measurement is necessary for condition monitoring and is the process of collecting information from a physical state and comparing the information with standards. Measurements are performed by using instruments, which are designed and manufactured to accomplish specific tasks. Often, several instruments are used to collect information about the process under investigation. They are then connected to the station using wired, optical or wireless networks [17].

When measurements are made, different types of data analysis methods may be used to extract useful information about the condition of the asset. Examples of different types of data analysis include statistical methods, curve fitting and selecting or discarding subsets of data [17]. Machine learning is also being used more commonly, particularly when the measurements result in big data.

However, there may be many sources of inaccuracy in measurements. Incorrect measurements can lead to the wrong asset management decisions being made and result in reduced asset lifetime and interruptions. It is therefore important to identify these sources by considering different factors, as stated in [17]:

- Imperfections in electrical and mechanical components.
- Changes in component performances (for example changes in chemistry, and ageing).
- External and ambient influences (for example temperature, pressure, and humidity).
- Inherent physical fundamental laws (for example thermal and other electrical noises, Brownian motion in materials, and radiation).

Estimation is a process of finding an approximation. This value can be used for some purposes even if the data from measurements may be incomplete, uncertain or unstable [18]. An estimate that turns out to be incorrect will be an overestimate or an underestimate. Over- and underestimation of condition values and remaining lifetime can cause severe and costly consequences if the wrong decisions are being made. The technical condition of an asset can unexpectedly lead to failure, or the asset can be taken out of service unnecessarily early.

2.1.3 Instruments and Sensors

Instruments are devices used for determining the value of a quantity/variable when measuring. They are designed to maintain prescribed relationships between the measured parameters and the physical variables under investigation, by sensing and generating signals. The design of the instruments is based on existing knowledge, which is obtained either from peoples' experiences of the physical process, or from a structured understanding of the process. Ideas conceived about an instrument are transformed into hardware and software that can perform well within the expected standards and be easily accepted by the end users [17].

There are many different ways to design and construct instruments. They can be analog, digital or hybrid, and design usually requires many multidisciplinary activities. The ideal instrument would have a perfect sensitivity (detection of small changes), reliability (low failure rate) and repeatability (narrow variation of results), while being within the acceptable standards. However, in many cases, there will be imprecise and inaccurate results because of internal and external factors. The performance of an instrument needs to be determined by considering the factors mentioned in Subsection 2.1.2, and it is essential to appreciate how errors arise [17].

The general construction of an instrument is illustrated in Figure 2.3. An instrument will typically have a sensor or transducer stage, a signal-conditioning stage, and an output or termination stage. All instruments have some or all of these functional blocks [17].

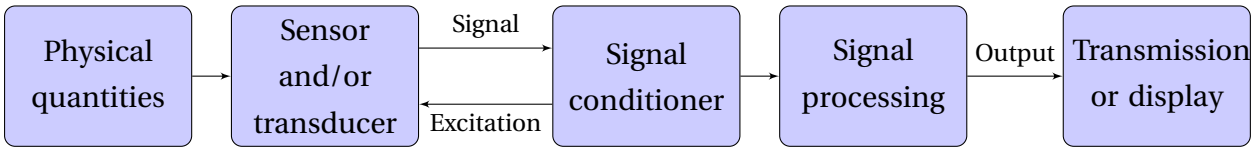

Figure 2.3: Structure of a typical instrument [17].

A sensor is a component that detects and responds to events or changes. A transducer is a device that transfers energy from one form to another. There is a diverse range of sensors and transducers available to meet the measurement requirements of a physical system. They can be categorised depending on the energy input and output, input variables, sensing elements, and electric or physical principles [17].

A conventional sensor measures physical, biological or chemical parameters and converts these parameters into electric signals. These sensors require many external circuits and components for signal processing and display. Smart sensors are a newer and different class of sensors, that consist of IEDs. They can convert a raw sensor signal to a level that makes them more convenient to use. This increases the quality of information rather than just

passing the raw signal. They can also perform functions such as self-identification and self-testing. They can consult tables and calibration curves, and have the ability to communicate with other devices. These functions are made possible by the integration of sensors with micro-controllers, microprocessors or logic circuits in the same chip [17].

Fiber optic sensors are smart sensors that use optical fibers to detect quantities such as mechanical strain or temperature, pressure, vibrations or displacements. An optical fiber is a cylindrical waveguide, composed of a core where light propagates, a cladding with a lower refractive index, and a coating layer for mechanical resistance as shown in Figure 2.4 [19]. They work on the principle that a pulse of light energy from a laser or any super-luminescent source, is transmitted by the optical fiber. The light will experience changes in its parameters and reach a detector that measures these changes. Fiber optic sensors are resistant to electromagnetic interference, and they do not conduct electricity. They can, therefore, be used for applications involving high voltage electricity, and they are mainly used in remote sensing applications [20].

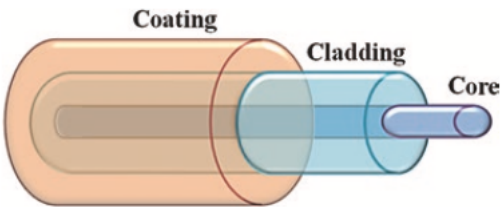

Figure 2.4: Basic structure of an optical fiber [19].

2.2 Power Transformers

Power transformer reliability is very important for the performance of the power system. They have an important function, they are costly, and if they fail, the outage time may be extensive. The main function of transformers is to transform energy from one voltage level to another. An important issue is the long unavailability and replacement time in case of severe failure. Against the background of the increasing age of the power system, failure probability and ageing of important components such as power transformers become important issues for asset management, including maintenance and reinvestment planning [3]. Studies show that power transformers have the potential for long lifetimes, around 80 years, under ideal conditions with low load, proper cooling and low water penetration [3]. Manufacturers however often define the expected lifetime of power transformers to be between 25 and 40 years [21]. Statnett is planning to build approximately 20 new power transformer stations by 2021, while several of the existing transformers require reinvestment as a result of reaching remaining lifetime expectancy [2]. According to the expert in the field of power transformers in Statnett, Statnett expects a minimum lifetime of 60 years during normal operation, and with maintenance performed on the transformer. Some parts of the transformer (such as the bushings) however normally need to be replaced after about 30 years. The time for reinvestment is determined by condition monitoring and assessment. Figure 2.5 shows an illustration of a power transformer and its main components.

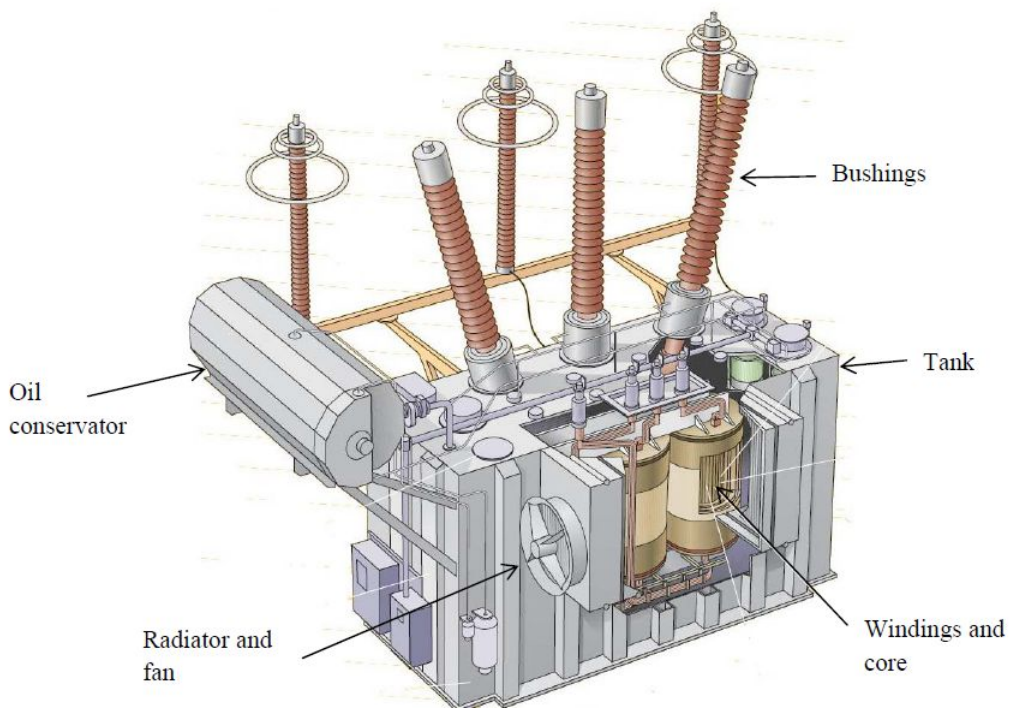

Figure 2.5: Illustration of a power transformer and its main components [22].

The power transformer normally consists of two windings (coils) per phase, wound on a common core of magnetically suitable material such as iron or steel. One winding receives the energy from an AC source through the transformer bushings, and is known as the primary winding. The other, which receives the energy by a mutual inductance from the primary, delivers that energy back through the transformer bushings, to the load, and is known as the secondary winding. The core provides a pathway for the easy flow of the magnetic lines of force or magnetic flux. The windings are isolated from each other and the core by an insulating paper that mainly consists of cellulose. The windings and core assembly is placed in an enclosure, often a steel tank, for protection. The tank is filled with oil, normally mineral oil, to provide insulation and cooling. A tap-changer is also connected to the transformer to regulate the output voltage to desired levels by varying the turns ratio [23].

During operation, power transformers are exposed to electrical, chemical and thermal stresses which often lead to higher temperatures. High temperatures result in decomposition and ageing of the insulating paper and oil in the transformer. Oil molecule bonds can break, and generate free particles. The interactions between these particles and external molecules form by-products. Common products are water, acids and dissolved gases formed due to contact with oxygen or copper and iron. This results in a reduction of the oil quality. The acids also trigger the decomposition of paper, which results in the production of water. This whole process is self-reinforcing as acids also cleave the oil molecules, and water in oil reduces the strength and further degenerates the oil. This leads to reduced service lifetime and reliability of the transformer. Actions including filtering or drying of the oil, or alternatively replacement, are done to improve the oil quality. On the other hand, the paper can not be repaired. A radical rehabilitation or replacement of the transformer is therefore necessary if the paper has poor quality. The paper is also subject to faster ageing than the oil. The strength of the paper is therefore considered to be the final limiting factor for the remaining lifetime of the transformer [24].

Figure 2.6 shows the distribution of failure causes of large transformers, according to an analysis performed by CIGRÉ [25]. As can be seen from the figure, aged insulation is one of the main causes of failure. Failures of for example tap-changers and bushing may also occur, but these failures can usually be repaired faster than failures originating from the windings. However, failures in the bushings can lead to considerable consequences such as fires and explosions [25].

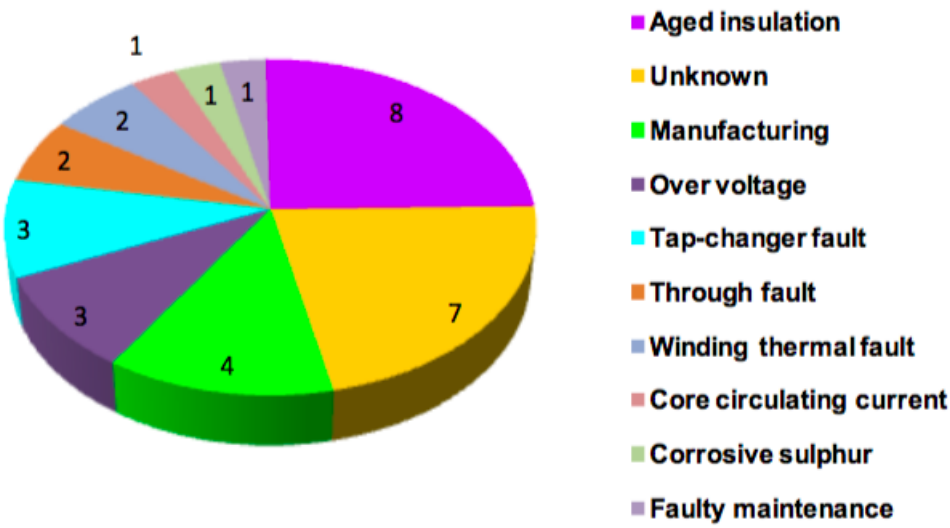

Figure 2.6: Failure causes of large transformers [25].

The lifetime of a transformer is therefore highly dependent on its insulation condition. The ageing rate for oil and paper is, as explained, strongly influenced by temperature. It is therefore important to keep the operating temperature as low as possible by ensuring that the cooling system is functioning properly. There are different cooling methods available for a transformer but the most common ones are the "oil natural air natural" (ONAN) and "oil natural air forced" (ONAF) cooling systems. For ONAN, the oil and air is circulated without any forced measure, but for ONAN, the air is circulated through fans. Some transformers can switch between these two methods [10], [26].

A critical temperature is the winding temperature, as the windings are a source of heat caused by the resistive losses, in addition to being surrounded by paper. The winding is the component which is subjected to the fastest temperature increase as the load increases. Another important temperature is the oil temperature. The oil temperature can be quite different in the top and bottom of the transformer. The top-oil has the highest temperature, as oil with high temperature will rise. This results in the winding temperature varying in the vertical direction with the upper winding temperatures being the most critical. These are commonly called hot-spots. The temperature of hot-spots is also governed by transformer loading and ambient temperature. These influence the transformer ageing and should be considered in transformer asset management. Large transformers lose valuable lifetime when the hot-spot temperature exceeds a critical maximum value, even for a few hours during short-term overloads. It is widely accepted that the remaining lifetime of the transformer is halved with a hot-spot temperature increase of 7°C [10], [27]. It is standard to have an upper limit of 105°C for the hot-spot temperature, as it is assumed a maximum average winding temperature rise limit of 65°C and a maximum ambient temperature of 40°C [28].

Disconnections and outages of power transformers in order to perform analysis are very expensive. Routine analysis of temperatures, in addition to dissolved gas and moisture content, are thus important to reveal developing faults as early as possible, so that disconnections for preventive action can be done at an operationally favourable time. For this thesis work, there is an emphasis on the ageing of the insulating paper and control of hot-spot temperature development, as it is considered to be the final limiting factor for the remaining lifetime of the transformer.

2.2.1 Maximum Temperature Estimation

Control of transformer temperatures is important, as discussed in Section 2.2. Transformer operating temperatures can be measured using installed temperature indicator thermometers. A commonly used thermometer is Qualitrol AKM 34/35 Gen 2 OTI/WTI, shown in Figure 2.7. It is a capillary-based, mechanical, remote indicating thermometer with configurations for oil temperature measurement and indirect winding temperature simulation. It features adjustable switches for alarm, trip, and cooling system functions [29].

Figure 2.7: Illustration of AKM 34/35 Gen 2 OTI/WTI [29].

Traditionally, it has been a challenge to directly measure the hot-spot temperature. Transformer manufacturers, utilities, and research organisations have for a long time been experimentally installing fiber optic sensors (explained in Subsection 2.1.3) in the windings of transformers, with the aim to directly measure the actual hot-spot temperature. The technology has today developed into equipment that is able to perform in the harsh environment inside the transformer. The equipment is reliable and able to precisely identify hot-spot locations in the windings [26].

The main principle behind detecting hot-spot temperature by fiber optic sensors is based on a pulse of light energy being sent over the optical fiber cable into the windings. A returned signal is then captured and converted to temperature in an optical electronic conversion. There are several technologies available [26]. Statnett has installed the fiber optic sensor Qualitrol/Neoptix T/Guard 408 in some of their transformers.

Fiber optic sensors placed into the windings of transformers are being accepted for research, verification of temperature rise tests and emergency conditions. They have become very important for transformer manufacturers and users. Manufacturers are now acknowledging fiber optic sensors as the proper equipment to gather data during factory testing, to provide feedback for design, information for maintenance, and for maximising loading [26].

The hot-spot temperature can also be estimated by calculations. IEC 60076-7 "Loading guide for oil-immersed power transformers" [30] describes how the temperature distribution can be illustrated as shown in Figure 2.8. Top- and bottom-oil temperature can be measured using temperature indicator thermometers, and the average winding temperature can be calculated from resistance measurements on the winding. The hot-spot factor, H, is a factor used to calculate the hot-spot temperature. This factor is specific for each design, winding and cooling mode, and should be requested from the manufacturer when buying the transformer. It normally varies in the range 1.0-2.6. If this factor is not known for the transformer, the guide suggests using a hot-spot factor of 1.3 for large power transformers [3], [30].

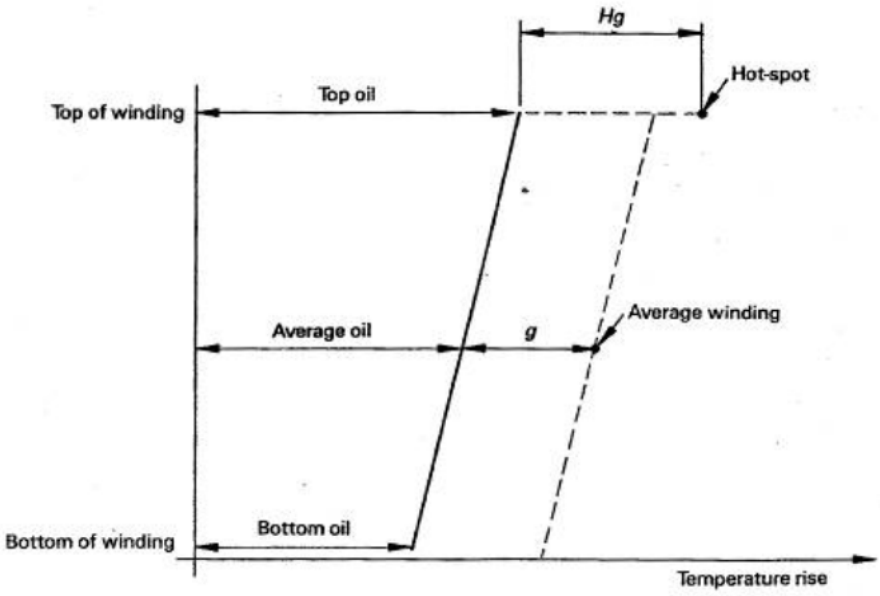

Figure 2.8: Thermal diagram for transformers [3].

Hot-spot temperature of a transformer at a given load can then be calculated from the data available from a standard heat run test. The simplified formula for the hot-spot temperature

for steady state, θ_h [°C], is described by Equation 2.1 [3]:

$$\theta_h = \theta_a + \Delta\theta_{or} \cdot [(1 + R \cdot K^2)/(1 + R)]^x + H \cdot g_r \cdot K^y \tag{2.1}$$

Equation 2.1 shows that θ_h is influenced by:

- θ_a, ambient temperature [°C]
- $\Delta\theta_{or}$, rated top-oil temperature rise [K]
- R, loss ratio
- K, load factor (load current/rated current)
- H, hot-spot factor
- g_r, average-winding-to average-oil [K]
- x, oil exponent
- y, winding exponent

These values are individual for every transformer and also for whichever cooling system is used (ONAN or ONAF). If the top-oil temperature of the transformer is known, Equation 2.1 can be further simplified to:

$$\theta_h = \theta_o + H \cdot g_r \cdot K^y \tag{2.2}$$

Equation 2.2 is assumed to give a more accurate estimate. This is because Equation 2.1 includes calculating the top-oil temperature, θ_o [°C].

2.2.2 Winding Insulation Degradation Model

As discussed in Section 2.2, the transformer liftime depends highly on its insulation condition, which again is influenced by hot-spot temperature. The insulation paper in the transformer windings degrades over time, as cellulose, which is the main part of the paper, is decomposed into shorter cellulose molecules. The two main ageing processes for cellulose are oxidation and hydrolysis. Paper with shorter cellulose molecules has reduced mechanical strength, and if the mechanical strength becomes too weak, the transformer may fail when exposed to high mechanical stresses, for example, if there are short circuits present. The length of the cellulose molecules is expressed as a degree of polymerisation (DP) - that is the average number of monosaccharide units in each cellulose molecule. New transformer paper has a DP-value of about 1200, but this decreases in the production process (drying of the materials) to around 1000. In Norway, it is common to define the end

of the lifetime of the paper, and thus the transformer, as the time when the DP-value reaches 200. A reduction in the DP is irreversible. The DP can be measured on samples of paper in laboratories, but this is difficult while it is in use [31].

The DP can also be estimated by calculations from information on the temperatures to which the windings have been exposed (hot-spot temperature). This requires access to detailed temperature and load history [31].

The change in the DP from any time "start" to any time "end" is described and estimated by Equation 2.3 [3]:

$$\frac{1}{DP_{end}} - \frac{1}{DP_{start}} = \int_{start}^{end} A(t) \cdot e^{\frac{-E_a}{R_g \cdot (\theta_h(t)+273)}} dt \tag{2.3}$$

Equation 2.3 shows that the DP is influenced by:

- t, time [hours]
- A, environment factor [h^{-1}]
- E_a, activation energy [J mol^{-1}]
- $R_g = 8.31$, gas constant [J mol^{-1} K^{-1}]
- θ_h, hot-spot temperature [°C]

A expresses the probability that a reaction will take place, while E_a determines the temperature dependence of the material ageing process. E_a and A have been determined by experiments and depend upon the type of paper used. It is suggested that E_a should be equal to 111 kJ/mol for normal Kraft-paper, and 86 kJ/mol for thermally upgraded paper of the Insuldur-type. A is dependent on the contamination content of the cellulose (water, acids and oxygen), which may vary with time [3].

A can also be calculated by using Equation 2.4 [3]:

$$A = \begin{cases} 4 \cdot 10^8 w_{p,n} \cdot O & \text{for Kraft-paper} \\ (1300 w_{p,n} + 14000) \cdot O & \text{for Insuldur-paper} \end{cases} \tag{2.4}$$

Here $w_{p,n}$ is the concentration of water in paper [%] and O is a multiplication factor for the presence of oxygen, which is assumed to be equal to 2.5 for all conservators [3].

$w_{p,n}$ can be calculated from the concentration of water in oil [%], $w_{o,n}$, which is typically measured during yearly oil analysis. The relationship between these quantities at equilibrium between paper and oil can be described by Equation 2.5 [31]:

$$w_{p,n} = (w_{o,n} \cdot 2.24 e^{-0.04\theta_h})^{0.63} \tag{2.5}$$

Here $w_{o,n}$ is the concentration of water in oil at hot-spot temperature θ_h when the oil analysis was taken [31].

Access to detailed temperature and load history may not be available for the entire transformer lifetime. If there exists detailed temperature and load history for one year, it may be assumed that this is similar to the history for the transformer for subsequent years [31]. Equation 2.3 can then be rewritten as:

$$DP_{today} = \frac{DP_{new}}{1 + DP_{new} \cdot \tau \cdot \overline{A} \cdot \frac{\sum_{m=1}^{M} e^{\frac{-E_a}{R_g \cdot (\theta_m(t)+273)}}}{M}} \tag{2.6}$$

Here DP_{new} = 1000 is the DP-value when the paper was new, τ is the transformers age today [hours] and $\theta_h(m)$ is the hot-spot temperature at the operating condition m. It is assumed that $\theta_h(m)$ has been measured a total number of M times at a fixed interval (every hour) over one year. By assuming that the one year is similar to the other years, one can calculate the average environment factor, $\overline{A}$, from when the transformer was new until today, assuming it increases linearly [31]. This can be done using Equation 2.7:

$$\overline{A} = \frac{A_{new} + A_{today}}{2} \tag{2.7}$$

When the transformer is new, the value of $w_{p,n}$ is assumed to be 0.5%, which can be used in the equation for A_{new} [31].

By using Equation 2.6, it is also possible to estimate the remaining lifetime of the transformer, by solving for τ. $\overline{A}$ will then also be dependent on τ, as $w_{p,n}$ will increase over time, shown in Equation 2.8:

$$\overline{A} = \frac{A_{new} + (13000 \cdot (0.5 + \phi \cdot \tau) + 14000) \cdot O}{2} \tag{2.8}$$

Here ϕ represents the incremental factor for the concentration of water in paper. It will be individual for each transformer but can be calculated by using the relationship in Equation 2.9 when τ and $w_{p,n}$ are known:

$$\phi = \frac{w_{p,n} - 0.5}{\tau} \tag{2.9}$$

2.3 Machine Learning

Machine learning was first defined by Arthur Samuel as early as 1959 as "the field of study that gives computers the ability to learn without being explicitly programmed" [32]. In recent years, several more definitions have emerged, but the common content of these is that machines have the ability to learn from historical data by finding a context and thus making a prediction. Machine learning is an approach, or subset, of artificial intelligence (AI), with an emphasis on learning rather than just computer programming. These terms are both a part of data science.

Machine learning follows the step by step approach shown in Figure 2.9. Machine learning is built on a data base, often big data, to learn from. In this context, learning means making the right decisions based on historical data and events. It is not a prerequisite to have big data to generate good results, but big data will support a finding to a greater extent than fewer data. Small data sets can reduce the relevance or correctness of the results obtained by analysis. The data needs to be cleaned, prepared and corrected for missing or faulty data. It is then used as input in learning algorithms to make models for the machine to be able to find relationships and make decisions. The machine learns to perform a task by studying a training set of examples and then performs the same task with the data it has not encountered before, so-called test data. Based on the result, the model can be improved.

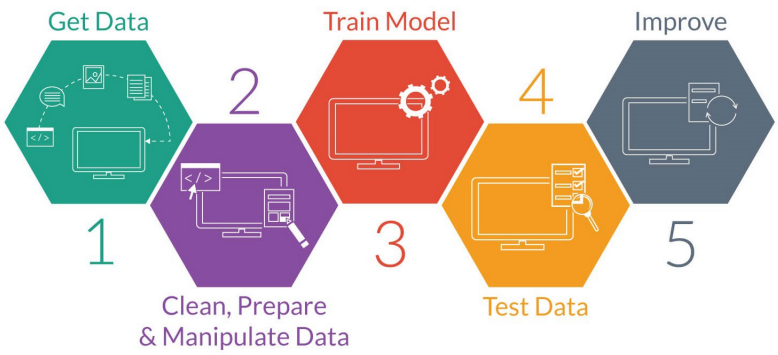

Figure 2.9: Machine learning step by step approach [33].

Machine learning is not a new concept. The first related algorithms appeared in the 1970s. However, the increase in computing power has allowed machine learning to handle more complex problems, while the increase of data available due to better sensor technology has allowed the application of machine learning to an expanding range of domains [34]. According to [35], less than 1% of global big data is analysed today. Machine learning can potentially be the solution for increasing this percentage for asset management purposes.

Today, machine learning is already well implemented in our everyday life. Examples include voice recognition systems such as Siri for iPhone/iPad and Cortana for Windows, filtering of spam for emails, and online recommendations [36], [37]. Power companies can also benefit from the use of machine learning technology, and some have already started implementing it. For example, the power company TrønderEnergi, with help from the information technology company Powel, has used measurement data to predict production of wind power to reduce the production uncertainty [38].

2.3.1 Data Processing

Pre-processing of data is an important task that normally needs to be done before the data set can be effectively used for machine learning. Raw data is often noisy and unreliable. In some cases, it also may be missing values. The use of such data for modelling can give misleading results [39].

Typical data quality issues are, as stated in [39]:

- **Incomplete**: Data lacks attributes or contains missing values.
- **Noisy**: Data contains erroneous records or outliers.
- **Inconsistent**: Data contains conflicting records or discrepancies.

Quality data is essential for quality predictive models. It is therefore important to perform a quality control to discover data issues early and further decide on the corresponding data processing and cleaning steps. If there appear issues with the data, processing steps are necessary. These often involve [39]:

- **Data cleaning**: Fill in missing values, detect and remove noisy data and outliers.
- **Data transformation**: Normalise data to reduce dimensions and noise.
- **Data reduction**: Sample data records or features for easier data handling.

2.3.2 Learning Strategies

The two most common learning strategies within machine learning are supervised and unsupervised, as shown in Figure 2.10. Both learning strategies include several learning algorithms. Which method that is the best to use depends on what type of data is available, how this data is structured and what is to be predicted.

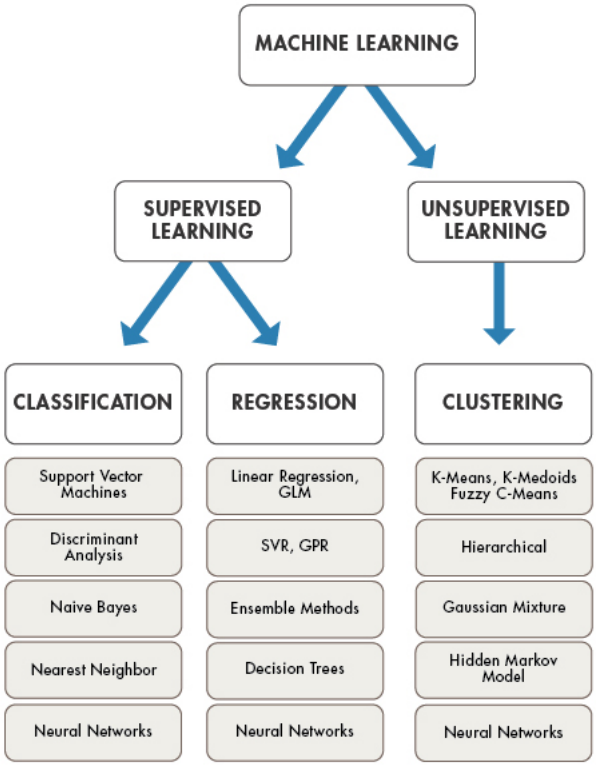

Figure 2.10: Machine learning strategies [40].

Choosing the right learning strategy and algorithm can be a challenge. There are many algorithms, and each one uses a different approach to learning. There is not a best overall method, nor one that fits all, and finding the right algorithm is partly based on trial and error. Choosing the right algorithm also normally requires trading off one benefit for another, for example, model speed, accuracy, and complexity [40].

Supervised Learning

Supervised learning trains a predictive model based on a set of input data and known responses to the (output) data. The goal is to learn a general rule that maps inputs to outputs, making it possible to predict the response to new data [40]. The input and output data are also called predictor and target data, respectively.

One way to think of supervised learning is by means of the simple Equation 2.10:

$$Y = f(x) \tag{2.10}$$

The machine will try to find a function f for the output Y by basing it on the input data x. The function is thus based on variables where the desired output is known. When using new input data with the same variables, where the output is unknown, the machine can use the

function to predict the output of the new data.

Supervised learning uses classification and regression techniques, as shown in Figure 2.10, to develop predictive models [40]:

- Classification: predicts categorical responses. It takes as input a data set and the class of each data such that the machine can learn how to classify new data into categories. An example is whether an email is genuine or spam. Typical applications include image and speech recognition.
- Regression: predicts continuous responses. It takes as input a data set and investigates the relationships between the input (predictor) and output (target) data. Based on these relationships, the model will be able to predict quantities of new data. Examples include changes in temperature or fluctuations in power demand. A typical application is electricity load forecasting.

Unsupervised Learning

For unsupervised learning, the data set consists of input data without known output data. The goal is to find the underlying structure or distribution of the data to learn more about it. There is often a need for large amounts of data in order to group data optimally [40].

The most common unsupervised learning technique is clustering, as shown in Figure 2.10. It takes as input a data set covering various dimensions, and partitions it into clusters satisfying certain criteria, as illustrated in Figure 2.11. It is used for exploratory data analysis to find hidden patterns or groupings in data [34], [40].

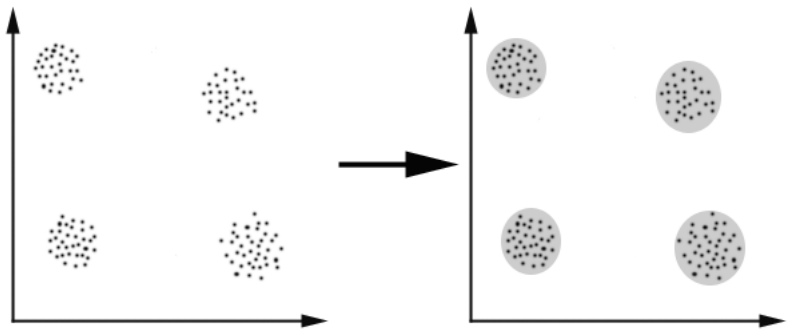

Figure 2.11: Illustration of a clustering example [41].

2.3.3 Model Theory

Training and Testing

The two phases training and testing, as illustrated in Figure 2.12, are essential parts of machine learning. The training phase uses one or more machine learning algorithms to make models from the training data set. Later, the testing phase uses new data from the test data set to evaluate the accuracy of these models. Test data will thus be perceived as real data by the model and provide an indicator of how the model treats data sets differently from the original training data set [42].

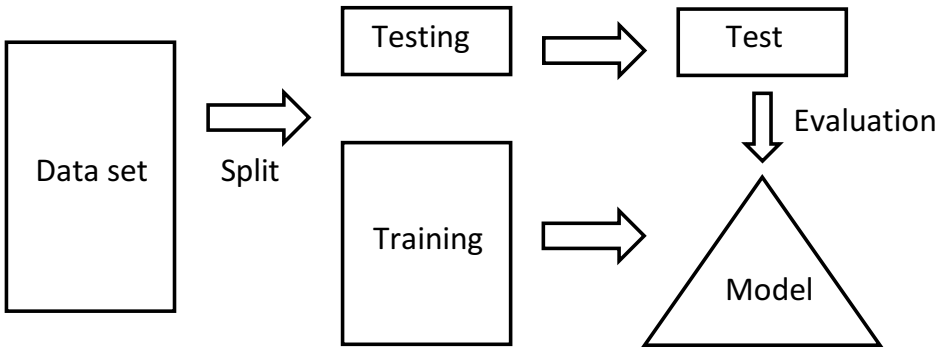

Figure 2.12: Illustration of how a data set is split to evaluate the model [42].

If only one data set is available, it is necessary to split the data set into two parts that constitute the training and testing data set. Ideally, more data sets are used, but this depends on the type of data available [42]. There is no general rule on how the data set should be split, but a common starting point is an 80% train and 20% test split [43].

Overfitting and Underfitting

When the machine has made a model based on the training data set, it is important to be critical of the results obtained. An important problem that should be considered is the bias-variance dilemma.

In data sets, not all variables (or features), will be equally important in order to predict the desired output. If the machine learning model is not configured correctly, the model can cater for more information than needed, for example noise and inaccuracies. This will result in a very complex model that fits the training data very well, but not the testing data. This results in high variance and constitutes overfitting, which is one of the major challenges in machine learning. A possible solution to this problem is to decrease the number of features in the training set [44], [45]. Simpler models do not usually overfit.

However, when the model is trained on data with too few features, it results in a model that performs poorly on both the training and testing data. This results in high bias and constitutes underfitting. In this case, it may be necessary to check if the data set is of good enough quality or if the wrong machine learning algorithm is used. Possible solutions to this problem are to train longer or to obtain more features. Underfitting is not as big a problem as overfitting as these errors are easier to detect [44], [45]. Figure 2.13 illustrates how the same data set can be fitted in three different ways, including overfitting and underfitting.

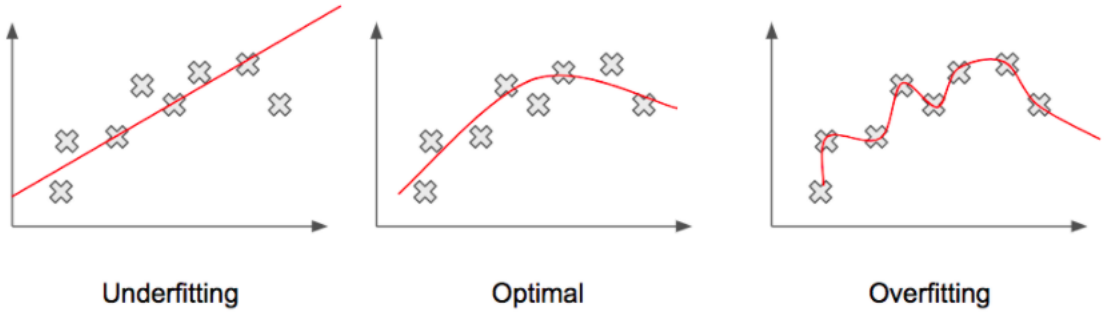

Figure 2.13: Illustration of how the same data set can be fitted three different ways [46].

Feature Selection

Not all features are equally important in order to predict the desired output. It would therefore be smart to omit less relevant features from the training process. This will reduce the processing power needed to train the model, in addition to reducing the chance of overfitting. It can be analysed which features have the highest variance, i.e. carry the most information. The lowest ranking features can then be removed without losing much information. This will reduce the complexity of the model, which again makes the model easier to understand and explain [47].

Several different feature selection algorithms exist. Which one to use depends on the data set and model. The algorithms can be divided into three categories: filter methods, wrapper methods and embedded methods. Filter methods select the features that best perform in different statistical tests for their correlation with the output. Wrapper methods test different sets of features and select the model that scores the best. This has quite a high computational cost for the machine. Embedded methods are a combination of these two, and there are several popular feature selection algorithms that use this. Embedded methods learn which features best contribute to the accuracy of the model while the model is being created [47].

The most common type of embedded feature selection methods is regularisation methods. Regularisation methods add a penalty as model complexity increases for predictive algorithms (regression). It constrains/regularises or shrinks the coefficient estimates of the

different features towards zero. In this way, the technique avoids the risk of overfitting by biasing the model toward lower complexity and flexibility [48].

Model Evaluation

When a model has been made, it is important to evaluate its performance. This determines if it can be used further, if the configurations should be changed, or if the entire model must be rejected. There are a number of metrics used to evaluate predictions, depending on the model used.

For classification models, a desirable evaluation metric is the classification accuracy, which is the number of correct predictions made as a ratio of all predictions made [49].

For regression models, there are several statistical metrics, where three of the most important ones are summarised in Table 2.1 [49].

Table 2.1: Model evaluation for regression.

Name	Definition	Explanation
Mean absolute error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Evaluation with even weight on all types of errors.
Root mean squared error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	Evaluation with high weight on major errors. Should be used if large errors are particularly undesirable.
Coefficient of determination (R^2)	$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$	A general evaluation of accuracy. Can be interpreted as how much variance the model explains.

Here y_i is the measured value, $\hat{y}_i$ is the predicted value, and $\overline{y}$ is the mean of the measured values.

For MAE and $RMSE$, a value of 0 indicates no error or perfect prediction. R^2 compares the fit of the model with a horizontal straight line. It returns a value between 0 and 1 for no fit and perfect fit respectively [49]. This means, when R^2 equals 0, a horizontal straight line explains the data equally as well as the model. If the model fits the data worse than a horizontal straight line, the value of R^2 can also be negative. A negative value should however generally be interpreted as 0.

2.3.4 Essential Tools

As machine learning has grown in popularity, a variety of tools have become available. Table 2.2 compares some popular machine learning tools. R is generally more popular among people with a strong statistical background. It has a large volume of machine learning and statistical inference libraries. Python is more popular among people with a computer science background. Although Python is not made specifically for machine learning or statistics, it has extensive libraries for numerical computing (NumPy), scientific computing (SciPy), statistics (StatsModels) and machine learning (scikit-learn). MATLAB has also added machine learning algorithms and implementations, and might be a natural choice for people familiar with MATLAB. Recently, Google also released as open source its TensorFlow library for working with Artificial Neural Networks (ANNs), which can be accessed through Python [34].

Table 2.2: Overview of popular machine learning tools [34].

	Python	R	MATLAB	TensorFlow
License	Open source	Open source	Proprietary	Open source
Distributed	No	No	No	No
Visualisation	Yes	Yes	Yes	No
Supported languages	Python	R	MATLAB	Python and C++
Variety of machine learning models	High	High	High	Low
Suitability as a general-purpose tool	High	Medium	High	Low
Maturity	High	Very high	Very high	Low

In addition, there are several other tools which can be helpful when using machine learning. Jupyter Notebook is an open-source web application that allows the creation and sharing of documents that contain live code, equations and visualisations. The notebook has support for many programming languages, including Python and R [50].

2.4 Cross-Industry Standard Process for Data Mining

Cross-Industry Standard Process for Data Mining (CRISP-DM) is a data mining process model which provides a structured approach to obtain better and faster results from data mining projects [51]. Data mining is the process of discovering interesting and useful patterns and relationships in large volumes of data, for example by the use of machine learning [52].

CRISP-DM was conceived in 1996 as a result of the need for a data mining process model that would standardise the industry. At that time, there was a big explosion in interest for data mining, but no widely accepted approach existed. The development of a non-proprietary, documented and freely available model would enable better results from data mining and encourage best practices in the industry [51]. CRISP-DM is now the most popular methodology for analytics, data mining and data science project according to KDnuggets, a leading site on data mining [53].

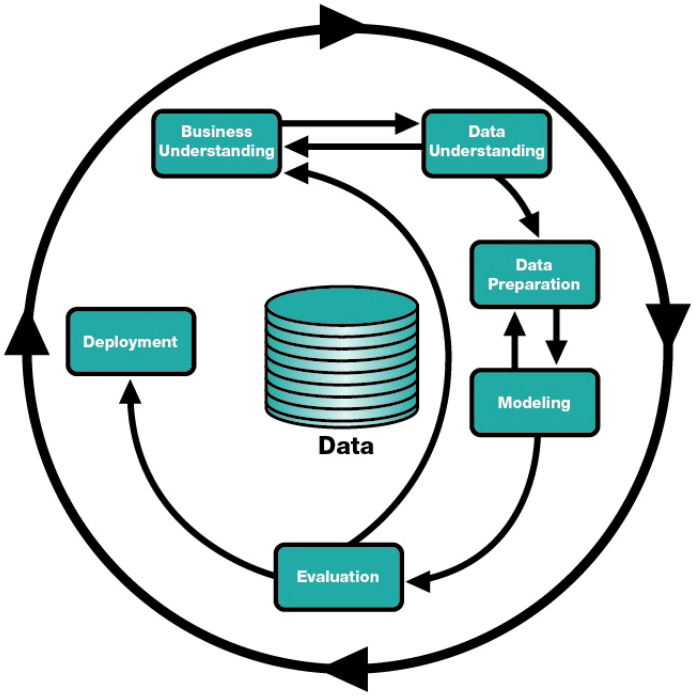

Figure 2.14: Process diagram showing the relationship between the different phases of CRISP-DM [54].

As seen in Figure 2.14, CRISP-DM organises the data mining process into six phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. These phases will provide an understanding of the data mining process in addition to providing a procedure for planning and carrying out the data mining project. It is normal to go back and forth between the different phases. The tasks involved with each phase are described below [51].

Business Understanding

The first phase involves understanding the project objectives from a business perspective, with the aim to transform this knowledge into a data mining problem definition. A plan should then be developed to achieve the objectives.

Data Understanding

The next phase starts with an initial data collection and continues with activities performed to increase familiarity with the data. This includes describing and exploring the data in addition to verifying data quality.

Data Preparation

The data preparation phase contains activities to construct the final dataset from the initial raw data. Steps include selecting, cleansing and formatting data.

Modelling

In the modelling phase, different modelling strategies are normally selected and applied. There are typically several techniques applicable to the same data mining problem. Some techniques have special requirements for the form of data. This often requires a return to the data preparation phase.

Evaluation

When the model has been built, it is important to thoroughly evaluate its performance. The construction of the model should also be reviewed to make sure it achieves the objectives. A decision has to be made as to whether the model needs altering or can continue to be used unaltered.

Deployment

The knowledge obtained must be organised and presented in such a way that it is usable. This includes planning deployment, monitoring and maintenance and reviewing the project.

Chapter 3

Case Studies

In total 203 data sets containing time series for 154 power transformers from Statnett were made available for case studies in this thesis work. CRISP-DM was chosen as the method to approach the data, as it from theory (Section 2.4) is a leading methodology within data mining. The motivation was to predict temperature development for a given transformer in order to estimate the condition and remaining lifetime for asset management purposes, using machine learning.

Two case studies were performed to demonstrate the applicability of machine learning in power transformer asset management:

1. **Hot-spot Temperature Prediction**:
   The aim of this case study was to predict transformer hot-spot temperature using the available data and machine learning. Measured hot-spot temperature by a fiber optic sensor was used as known output (target) data in supervised learning. The purpose of this case study was to predict the maximum temperature in the transformer in a relatively short-term period (months). This will enable Statnett to determine how much they can load a transformer short-term, for example in cold periods, or when there is a need for emergency loading due to a reconstruction of the power system or a redirection of the load. This case study has an operational focus.

2. **Winding Insulation Degradation Estimation**:
   The aim of this case study was to use predicted hot-spot temperatures to estimate winding insulation degradation for different scenarios. The scenarios include increased load over time compared with the actual load and both increased and decreased ambient temperatures. The predicted and measured hot-spot temperatures were also compared to calculated hot-spot temperatures obtained using the loading guide. The purpose of this case study was to estimate how much different scenarios could affect the estimated remaining lifetime of the transformer. In this way, Statnett can evaluate the economics of whether to increase the load of the transformer

as an alternative, or to evaluate other alternatives, for example, the purchase of a new transformer. This is more long-term (several years) prediction and the case study has an asset management focus.

The structure of this chapter is shown in Figure 3.1. The CRISP-DM phase *business understanding* is already conducted in Chapter 1. This chapter therefore starts with the phase *data understanding* and also includes the phases *data preparation* and *modelling*. *Evaluation* is performed during the *modelling* phase, and is also discussed in more detail in Chapter 5. The last phase *deployment* is not included in this thesis as it will be the responsibility of Statnett to implement. As the case studies are based on the same data sets, the phases *data understanding* and *data preparation* are equal for both cases. During the thesis work, it was necessary to alternate back and forth between the different phases.

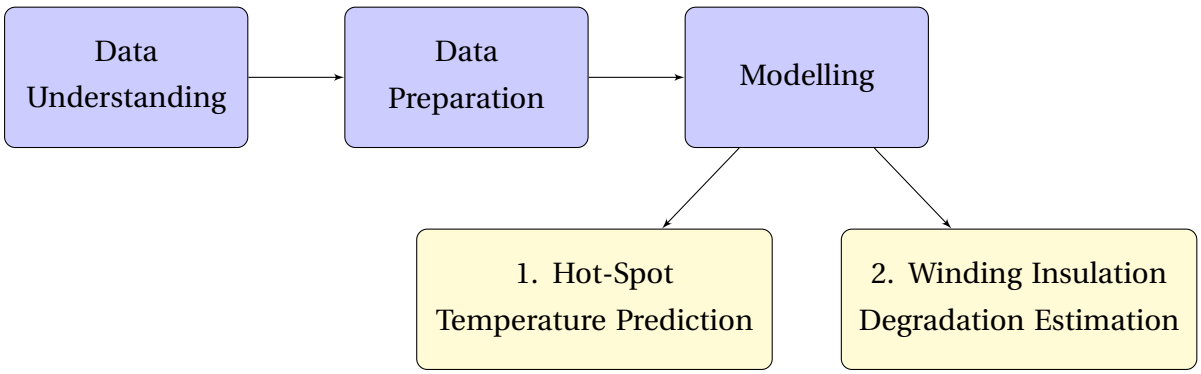

Figure 3.1: Structure of Chapter 3.

3.1 Data Understanding

To build a model using machine learning, it is important to have sufficient data. Statnett has collected and stored operational sensor measurements from its stations for many years, which provides a good starting point.

Initially, 50 data sets for 50 power transformers from Statnett were made available for this thesis work, described in Table 3.1.

Table 3.1: Description of the data sets initially provided by Statnett.

Category	Description	
Data	Obtained from	*Innsikt*
	Number of data sets	50
	Format	HDF5
Transformers	Number of power transformers	50
	Voltage level	420 kV or 300 kV
Time	Time period	01.01.2012 - 01.01.2017
	Sampling rate	One-minute
Measurements	Criteria	Gas concentration
		Winding temperature
		Oil temperature
		Line current (load)

The data sets were obtained from Statnetts data base *Innsikt*. The data was stored as time series in HDF5-format, where each file corresponded to one transformer. Each sensor had its own column in the file. The data was collected from the time period 01.01.2012 to 01.01.2017, at a one-minute sampling rate. The transformers were chosen from the criteria of being 420 kV or 300 kV, and have measurements for gas concentration, winding temperature, oil temperature and line current (load). In addition to these measurements, all other available sensor measurements were included:

- Moisture concentration
- Three phase power
- Switch position
- Three phase active power
- Tap position
- Line to line voltage
- Three phase reactive power
- Phase voltage

To examine the data, it was necessary to visualise the data. It was advised from Statnett not to use the default HDFView because of its complexity. Statnett has good knowledge with Python, and already uses it for other machine learning applications. As HDF5-files can be read directly from Python, it was decided to use Jupyter Notebook (described in Subsection 2.3.4) with Python as the programming language for examining the data. The main code used for this thesis work is provided in Appendix A.

Table 3.2 lists an excerpt of rows and columns from one example data set, obtained using Jupyter Notebook. As can be seen from the table, each sensor has an individual name. The name does not contain any information about what the sensor is measuring. To obtain this information, a separate CSV-file was provided by Statnett. Consequently, each sensor name had to be manually looked up and changed in Jupyter Notebook in order to understand the data set. Each data set contains on average 15 sensors. Not all data sets contain measurements for all the sensors in the CSV-file, as empty time series were dropped when the data sets were created (the availability of a sensor does not necessarily mean that there is data available for the selected time period and time resolution). It can also be seen from the table that the last column consists of NaNs (not a number). This is a result of missing data due to the sensor not measuring, which is undesirable when performing data science. This means that the sensors for each data set may not be measuring during the whole time period.

Table 3.2: Excerpt of rows and columns from one example data set.

Timestamp	e98fd2df-7d0c-f924-e040-1e828c94d40c	e9dddc4a-6ffc-b133-e040-1e828c948d72	e9dddc4a-700a-b133-e040-1e828c948d72	e9dddc4a-7026-b133-e040-1e828c948d72
2015-01-06 09:07:00	12.000	19.297	-0.275	NaN
2015-01-06 09:08:00	12.000	19.297	-0.275	NaN
2015-01-06 09:09:00	12.000	19.297	-0.175	NaN
2015-01-06 09:10:00	12.000	19.297	-0.125	NaN
2015-01-06 09:11:00	12.000	19.297	-0.325	NaN

Figure 3.2 shows further examination of the same example data set as above, with the sensor names changed. There are missing values, values being equal to zero, constant values and outliers. There is also a big change in the development of the values, at the time indicated by the arrow. The data in the time period before the arrow has a recognisable development, with the temperatures having daily and seasonal variations (due to a variable load and ambient temperature). However, it is physically unlikely for the oil temperatures to reach negative values, and the winding temperatures are also unlikely to be so low. During the time period after the arrow, the temperature values have constant values for longer periods of time, and also suddenly fall or rise. This deviates from what is physically reasonable. It is also physically unlikely that the moisture concentration decreases. This indicates that the sensors are not measuring correctly. The data set, therefore, does not seem to be of good enough quality to perform data science. In addition, it does not include measurements for line current, which was one of the measurements criteria in Table 3.1. Neither is hot-spot temperature measured by a fiber optic sensor present, which is necessary as output (target) values when predicting hot-spot temperature.

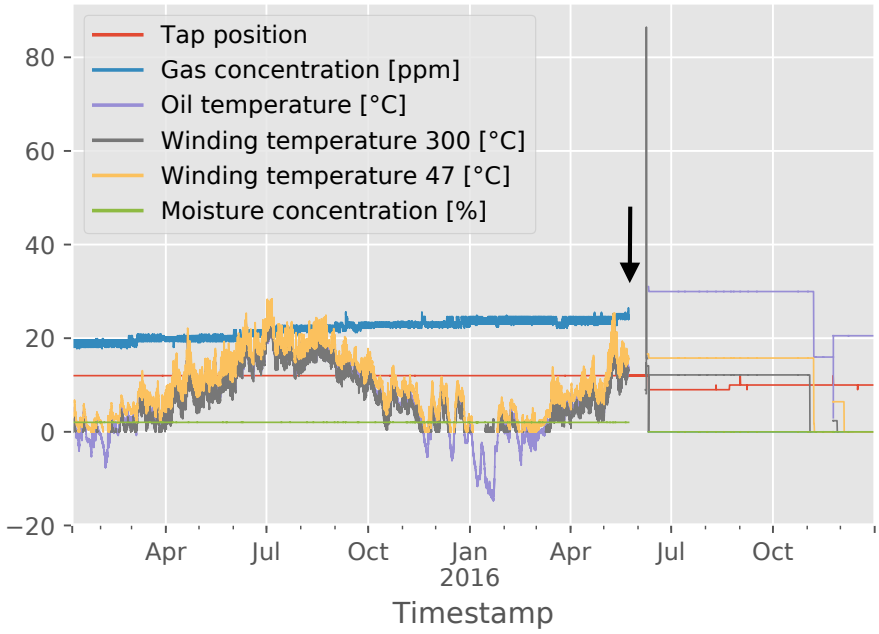

Figure 3.2: Plot of one example data set, with the sensor names changed.

Based on these examinations, some requirements were made to select data sets to be used further in this thesis work. These requirements are listed in Table 3.3. The requirements were made based on the knowledge obtained from theory (Subsection 2.3.1) about how important data quality is for data science. In addition, from theory (Section 2.2), the hot-spot temperature is mainly influenced by oil temperature and line current (load). As the goal is to predict hot-spot temperature, these measurements are therefore essential.

Table 3.3: Requirements for the data sets to be used further in the thesis work.

Description	Requirement
Data quality	Data which has been measured continuously for as long as possible (preferably several years)
	As few NaNs as possible
Measurements	Hot-spot temperature measured by a fiber optic sensor
	Line current (load)
	Oil temperature

3.2 Data Preparation

To be able to select which data sets to use further in this thesis work, all the data sets were visualised as plots. In addition, the separate CSV-file containing sensor information was consulted to find which data sets that were containing hot-spot temperatures measured by fiber optic sensors (this is one of the requirements in Table 3.3). Out of the 50 data sets, only 11 transformers had a sensor installed supposedly measuring hot-spot temperature. However, two of these sensors were not actually measuring. By studying the remaining plots, three data sets seemed of good enough quality for the purpose of this thesis work and met all the requirements made in Table 3.3. There was one data set where the data had been measured continuously for a longer time period than the two others. As it is important to have sufficient data when building a model by machine learning, it was decided that this data set was to be used further. The selected data set is called Data Set X for this thesis work, described in Table 3.4. Data Set X contains data from a 300 MVA 420/132/22 kV power transformer, called Transformer X for this thesis work. The data set contains measurements from several different sensors for the time period 06.01.2015 to 31.12.2016. The sensor measurements will be called features for this thesis work.

Table 3.4: Description of Data Set X.

Description	
Transformer X	300 MVA 420/132/22 kV
Time period	06.01.2015 - 31.12.2016
Features	*Oil temperature hot-spot* *Line current L1 420* *Line current L2 420* *Tap position* *Oil temperature* *Gas concentration* *Moisture concentration* *Winding temperature 420* *Reactive power 420* *Winding temperature 132* *Line current L3 420* *Winding temperature hot-spot 132* *Winding temperature hot-spot 420*

Figure 3.3 shows the plot of the initial raw Data Set X, with the sensor names changed to their feature names obtained by the provided CSV-file. Not all of these features are necessary when predicting hot-spot temperature, as they do not influence the hot-spot temperature of much importance. Therefore, only the features *Line current L2 420* and *Oil temperature* and *Winding temperature hot-spot 420*, requirements from Table 3.3, were chosen. This was done to reduce the complexity of the machine learning model. Only one of the line currents was chosen, as the currents of the three phases are assumed to be well balanced. This was also confirmed by plotting and comparing.

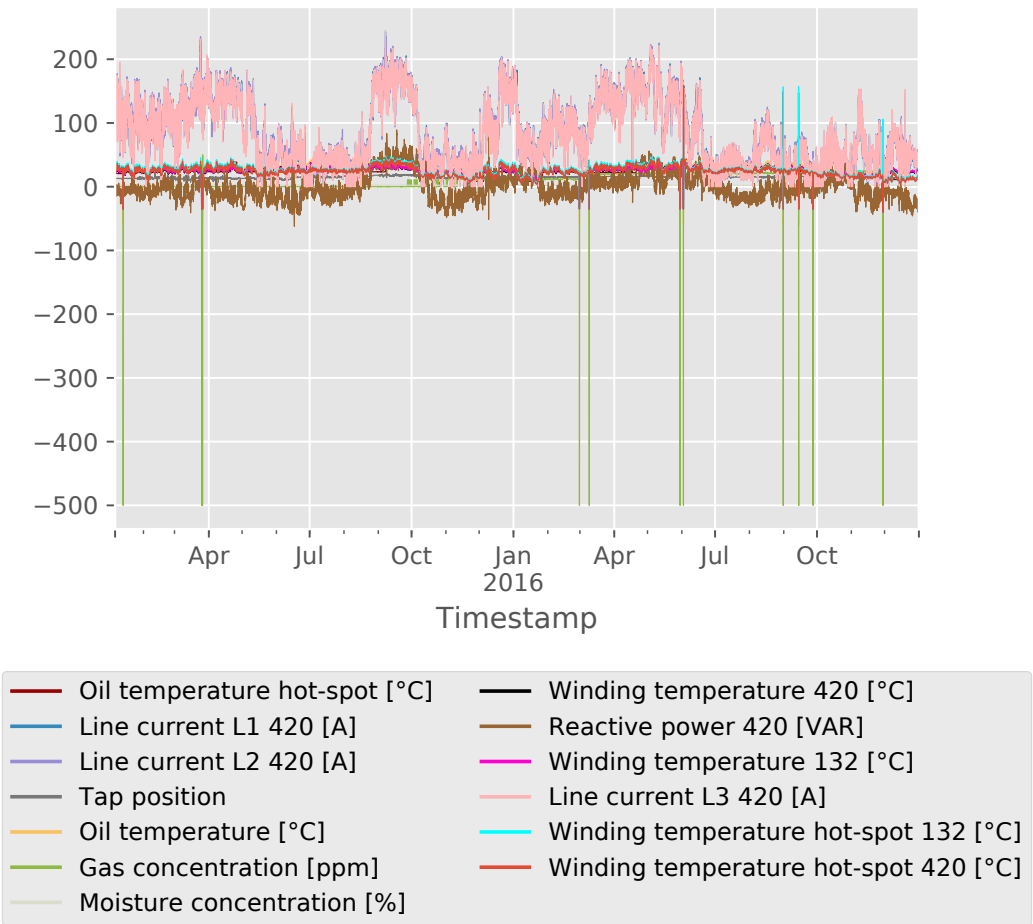

Figure 3.3: Raw plot of Data Set X, with the sensor names changed.

Figure 3.4 shows the plot of Data Set X, with the selected features. As can be seen, the oil temperature and hot-spot temperature contain both high and low outliers. These outliers are probably measurement errors as it is physically unlikely for the temperatures to reach those values. For the accuracy of the machine learning model, these values were cleaned by replacing them with the mean of the previous and proceeding values. As the data set is a time series containing continuous values, this method should be appropriate because the values are dependent on previous and proceeding values. This principle was also used to fill NaNs.

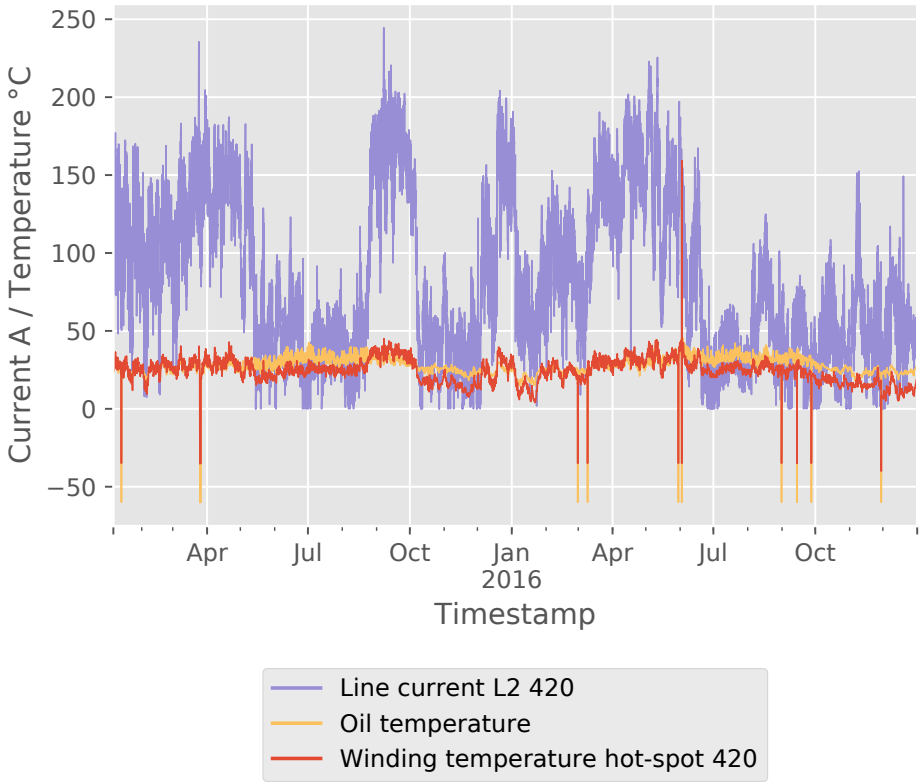

Figure 3.4: Plot of Data Set X with selected features.

From theory (Section 2.2), the ambient temperature also affects the hot-spot temperature. It was therefore decided to add ambient temperature to the data set by downloading weather data from met.com, by using a Python code provided by SINTEF Energy Research. In addition, the sampling rate of the data set was changed from one-minute intervals to one-hour intervals. This was done in order to get a continuous data set which describes the condition of the transformer better than the individual measurements do due to thermal time delays. Figure 3.5 shows the plot of the final Data Set X, with cleaned values, the feature *Ambient temperature* added and a sampling rate of one hour.

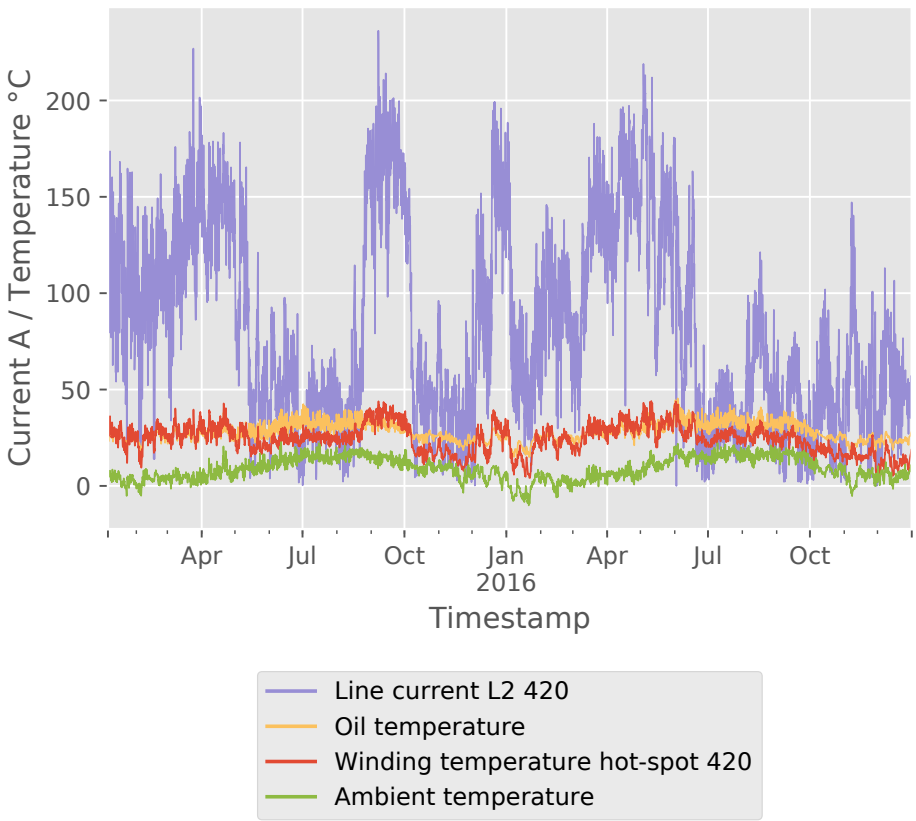

Figure 3.5: Final plot of Data Set X with cleaned values, the feature *Ambient temperature* added and a sampling rate of one hour.

Figure 3.6 plots the different features separately to better illustrate the seasonal variations of the values. It can be observed that there are similarities between the value developments of the features *Winding temperature hot-spot 420* and *Line current L2 420*, and to an even greater degree, between *Oil temperature* and *Ambient temperature*.

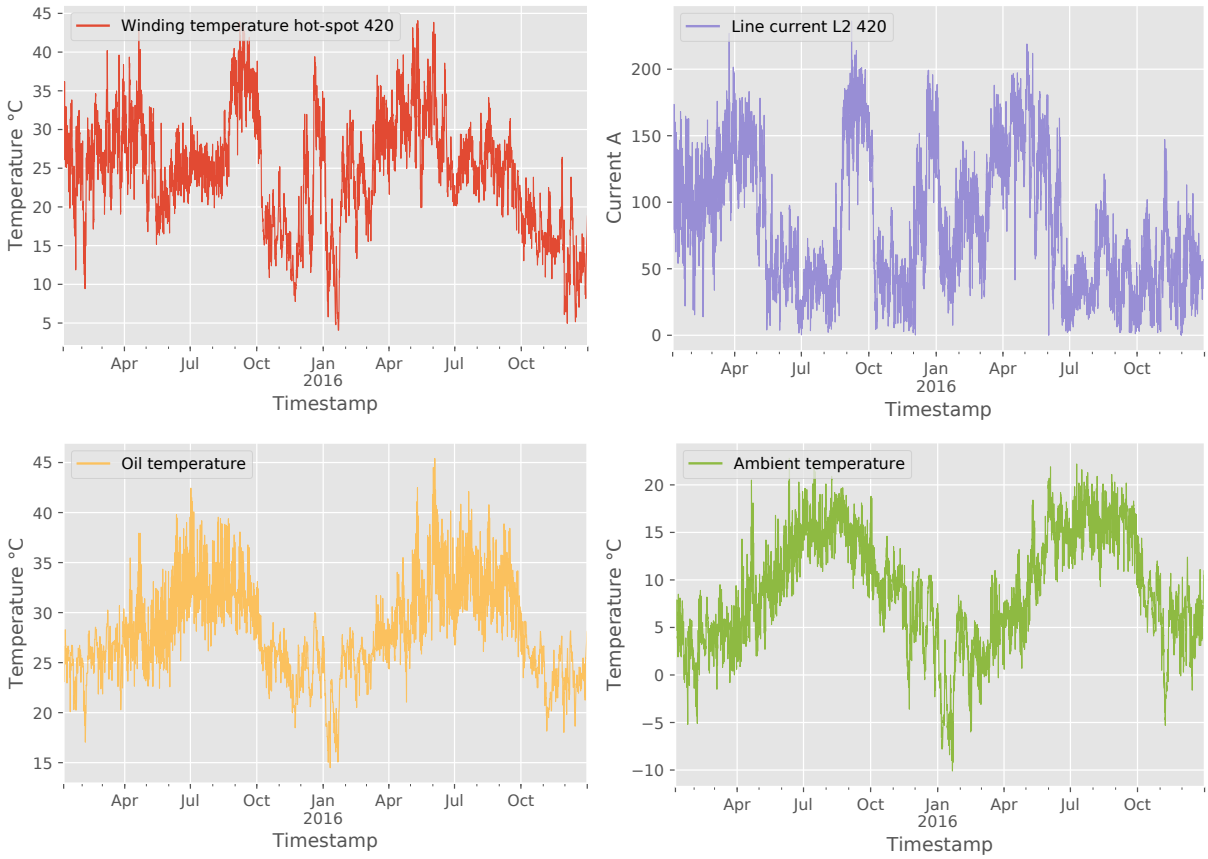

Figure 3.6: Separate plots of the different features of Data Set X.

A description of the different features of Data Set X from Jupyter Notebook is shown in Table 3.5. These values are within the normal range of transformer values.

Table 3.5: Description of the features of Data Set X.

	Line current L2 420	Oil temperature	Winding temperature hot-spot 420	Ambient temperature
Count	17415	17415	17415	17415
Mean	82.393	28.016	24.386	8.943
Standard deviation	50.826	4.439	7.332	5.899
Minimum	0.000	14.490	4.064	-10.100
Maximum	236.172	45.410	44.069	22.800

Figure 3.7 shows the correlation matrix of Data Set X, where a value of 1 indicates perfect correlation. The features *Line current L2 420*, *Oil temperature* and *Ambient temperature* correlate well with *Winding temperature hot-spot 420*. As they increase, the hot-spot temperature also increases. This corresponds to the physical principles of the transformer. The features also correlate to each other to some degree. As the ambient temperature decreases, the line current increases. This is explained by the need for a higher load when it gets colder outside. The oil temperature should also correlate to the line current, as from theory (Section 2.2) the windings will create more heat and thus warm up the oil as the load increases. However, this is not the case for this data set. This might be due to a time delay, but it also depends on how the oil temperature measurements are taken. If they are taken close to the wall of the transformer tank and further away from the windings, the oil temperature will be more influenced by the ambient temperature. As can be seen in the figure, there is a high correlation between the oil temperature and the ambient temperature. This is also supported by their value developments in Figure 3.6.

Figure 3.7: Correlation matrix of Data Set X.

3.3 Modelling

As mentioned in Subsection 2.3.4, Python has an extensive library for machine learning called scikit-learn. This library contains simple and efficient tools for data mining and data analysis. It provides a range of supervised and unsupervised learning algorithms including classification, regression, and clustering.

Time series prediction can be framed as a regression supervised learning problem, as it is desired to predict a quantity. The goal is to fit a curve/line to the data points, in such a manner that the differences between the distances of data points from the curve or line are minimised. As the data set contains several features, the process is called multiple regression. There are various kinds of regression techniques available for making predictions. As written in Subsection 2.3.2, choosing the right algorithm can be challenging. Scikit-learn provides a rough guide on how to approach problems with regard to which algorithm to try on the data. Figure 3.8 shows the guide for regression supervised learning [55].

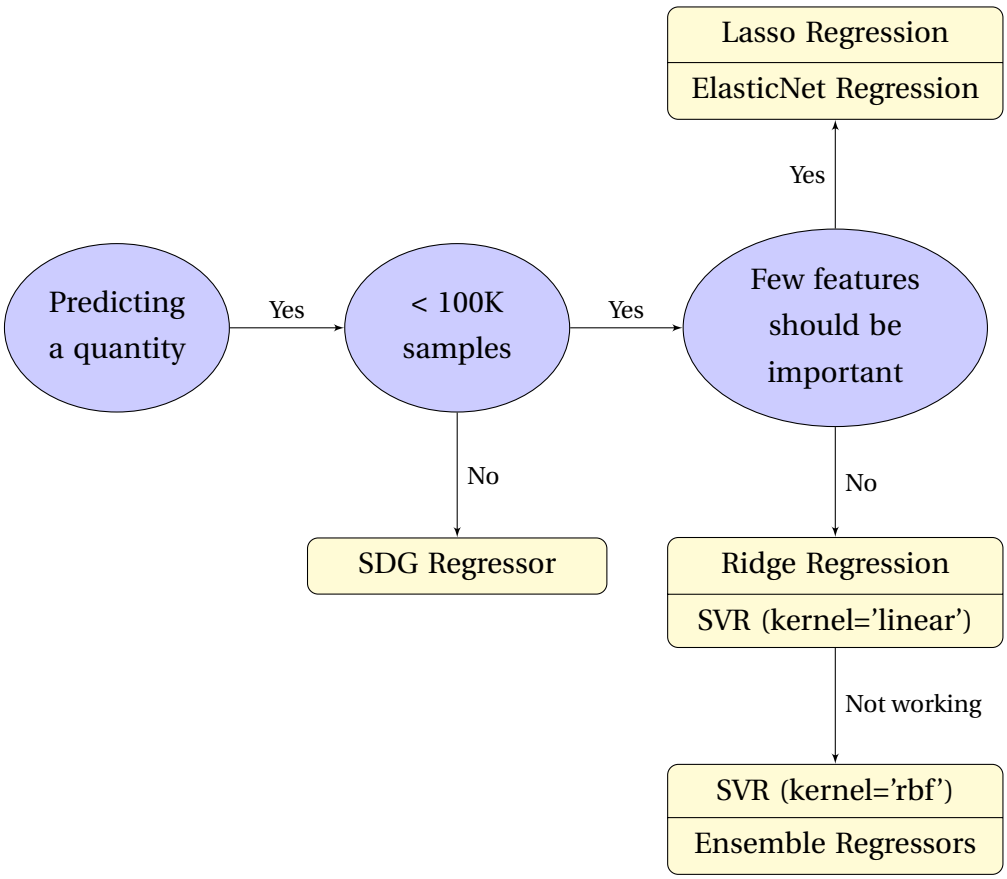

Figure 3.8: Guide for choosing the right regression algorithm [55].

3.3.1 Hot-spot Temperature Prediction

For this case study, the goal was to predict transformer hot-spot temperature by developing a machine learning model. The prepared data set Data Set X from Section 3.2 was chosen, and the feature *Winding temperature hot-spot 420* was used as known output (target) data, while the features *Line current L2 420*, *Oil temperature* and *Ambient temperature* were initially used as input (predictor) data.

Before machine learning algorithms could be applied to the data set, the data set had to be split into two parts that constitute the training and testing data. From theory (Subsection 2.3.3), there is no general rule on how the data set should be split, but a common starting point is an 80% train and 20% test split. However, before the data set could be split into training and testing data, it had to be split into predictor and target features. Listing 3.1 shows how this was done for Data Set X.

Listing 3.1: Split of Data Set X into predictor and target features.

```
x = data.drop(['Winding temperature hot-spot 420'], axis=1) #Predictors
y = data['Winding temperature hot-spot 420'] #Target
```

Data is here a data frame (tabular data structure with labelled axes) containing the prepared Data Set X, with cleaned values, selected features and a sampling rate of one hour. The function *drop* is applied to *data* to remove the target feature, *Winding temp hot-spot 420*, and stores the predictor features in a new data frame called *x*. The target feature is stored in a series (one-dimensional array with labelled axes) called *y*.

Data Set X was then split into training and testing data sets with 80% train and 20% test split, as shown in Listing 3.2.

Listing 3.2: Split of Data Set X into training and testing subsets.

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,
↪ shuffle=False)
```

The function *train_test_split* splits *x* and *y* into training and testing subsets. The parameter *test_size* represents the portion of the data set to include in the test split, which in this case was set to 0.2, being equal to 20%. The parameter *shuffle* determines whether or not to shuffle the data before splitting. As Data Set X is a time series containing continuous values, *shuffle* was set equal to *False* as there are dependencies between the previous and proceeding values. The training subset then contains the first 80% of Data Set X (06.01.2015 to 08.08.2016), while the testing subset contains the last 20% (09.08.2016 to 31.12.2016).

The next step was to apply machine learning algorithms to Data Set X. Figure 2.10, in addition to Figure 3.8, shows various kinds of regression algorithms. Figure 2.10 includes among others Linear Regression. Linear Regression is one of the most widely known modelling techniques, and is usually among the first to be applied when performing predictive modelling [56]. It was therefore decided to initially apply Linear Regression to Data Set X. Afterwards, several other regression algorithms were also applied according to Figure 2.10 and Figure 3.8, with the aim to find the algorithms that perform best on Data Set X. The approach and results of Linear Regression are first described in detail, while the results of the other regression algorithms are compared at the end of this subsection. The plots of the results from the other regression algorithms are provided in Appendix B.

Linear Regression

Linear Regression is a part of generalised linear models (GLM). These are methods intended for regression where the target value is expected to be a linear combination of the predictor variables. They are represented by Equation 3.1 [57]:

$$\hat{y}(\omega, x) = \omega_0 + \omega_1 \cdot x_1 + ... + \omega_p \cdot x_p \tag{3.1}$$

Here $\hat{y}$ represents the predicted value, $\omega = (\omega_1, ..., \omega_p)$ are coefficients, ω_0 is the intercept and $X = (x_1, ..., x_p)$ represent the predictor variables [57].

Linear Regression uses ordinary least squares and fits a linear model with coefficients $\omega = (\omega_1, ..., \omega_p)$ to minimise the residual sum of squares between the observed data points in the data set, and the data points predicted by the linear approximation. Mathematically it solves a problem of the form of Equation 3.2, also called the loss function [57]:

$$\min_{\omega} \| X \cdot \omega - y \|_2^{\,2} \tag{3.2}$$

However, coefficient estimates for Linear Regression rely on the independence of the predictor variables. When the predictor variables are correlated, the least-squares estimate becomes highly sensitive to inaccuracies, producing a large variance. This situation is called multicollinearity [57]. Having more features may be a good way to improve the accuracy of the trained model as it will be more flexible and take into account more parameters. However, it is then important to be careful not to overfit the model. Inaccuracies in the data may result in the model memorising the noise instead of learning the trend of the data.

Listing 3.3 shows how Linear Regression was applied to Data Set X. There are different setting parameters that could have been inputs for the algorithm, but the default settings were used in this case. The method *fit* fits the linear model based on the training subset, and the method *predict* predicts the target values using the linear model based on the predictor

variables in the testing subset. Model evaluation was also performed based on the statistical metrics in Table 2.1. The method *score* determines the coefficient of determination (R^2), interpreted as the accuracy of the model. In addition, functions for calculating the root mean squared error ($RMSE$) and mean absolute error (MAE) were used.

Listing 3.3: Application of Linear Regression to Data Set X.

```
#Linear Regression
lin = LinearRegression()
lin.fit(x_train, y_train)
lin_pred = lin.predict(x_test)
#Evaluation
lin.score(x_train, y_train) #Train
lin.score(x_test, y_test) #Test
rmse = np.sqrt(mean_squared_error(y_test, lin_pred))
mae = mean_absolute_error(y_test, lin_pred)
```

Figure 3.9 shows plots of the predicted values and the measured values of hot-spot temperature from the testing subset. The values do partly correspond, but the accuracy could advantageously be improved as there is a time period with large deviations, as illustrated by the arrow.

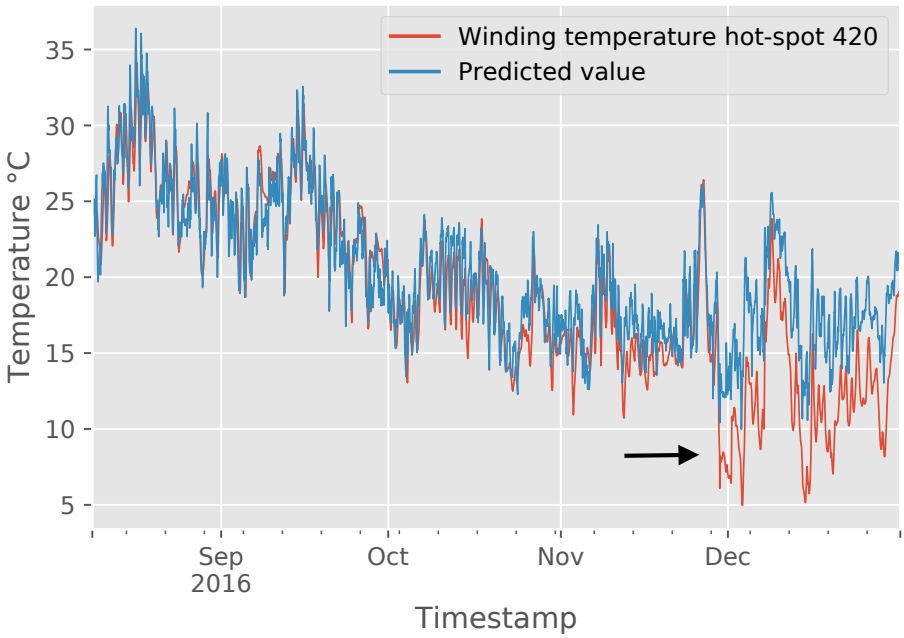

Figure 3.9: Plots of predicted values and measured values of hot-spot temperature for Data Set X using Linear Regression with 20% test size.

Table 3.6 lists the results of the model evaluation. From theory (Subsection 2.3.3), it is desired that R^2 is close to 1, while $RMSE$ and MAE, on the other hand, are desired to be close to 0. The model scores (R^2) well on the training subset (close to 1), but not as well on the testing subset. $RMSE$ and MAE are also higher than desired, as from theory (Section 2.2), the remaining lifetime of the transformer is halved with a hot-spot temperature increase of 7°C. It is therefore desirable for these metrics to be as low as possible to avoid making decisions based on a faulty basis.

Table 3.6: Model evaluation for Linear Regression applied to Data Set X, with a test size of 20%.

Name	Value
R^2 for training	0.917
R^2 for testing	0.747
$RMSE$	3.00°C
MAE	2.25°C

It was decided to try to improve the model. As can be observed from Figure 3.7, there are high correlations between the predictor features, resulting in the Linear Regression model being sensitive to inaccuracies. However, as the data set was cleaned by removing outliers prior to the machine learning algorithm being applied, there is most likely a different reason for the inaccuracy.

Different splits of Data Set X into training and testing subsets were therefore performed to see if the accuracy improved. Figure 3.10 shows plots of the scores for R^2 for the training and testing subsets with different values for the parameter $test\_size$. Looking at the figure, the scores on both the training and testing subsets are better when $test\_size$ equals 0.5. The parameter $test\_size$ in Listing 3.2 was therefore changed to 0.5, resulting in the training subset containing the first 50% of Data Set X (06.01.2015 to 03.01.2016), and the testing subset containing the last 50% (04.01.2016 to 31.12.2016). The measurements of Data Set X were taken over a time period of approximately two years, and contain seasonal variations. As seen from the figure, the test sizes 0.4-0.6 gives approximately the same scores. An explanation could be that the model performs best when trained for the seasonal variations of one year, and then tested for the one remaining year.

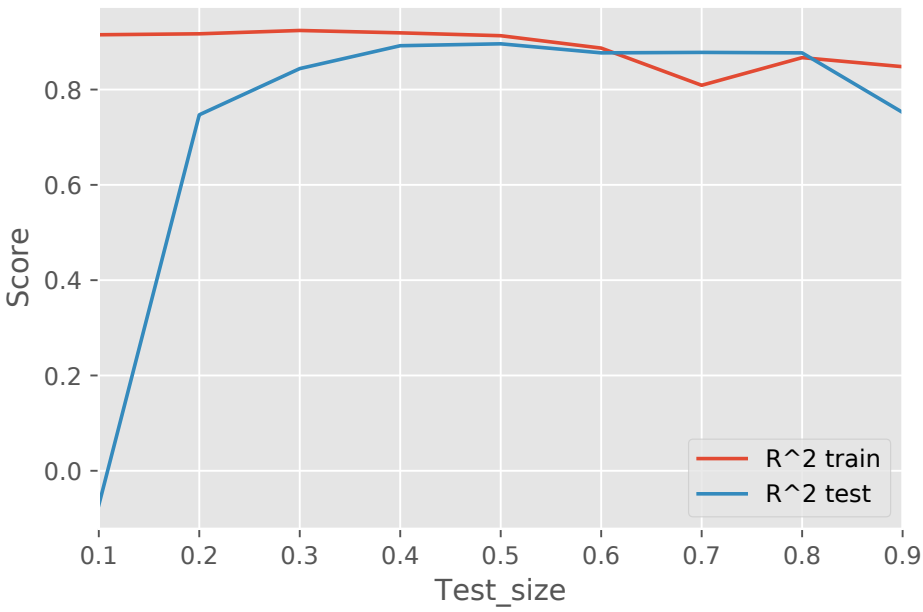

Figure 3.10: Plots of scores for the training and testing subsets with different values for the parameter *test_size* for Linear Regression applied to Data Set X.

Figure 3.11 shows plots of the predicted values and the measured values of hot-spot temperature from the testing subset, with the parameter *test_size* changed to 0.5. The values are now corresponding better and the accuracy of the model seems to have improved.

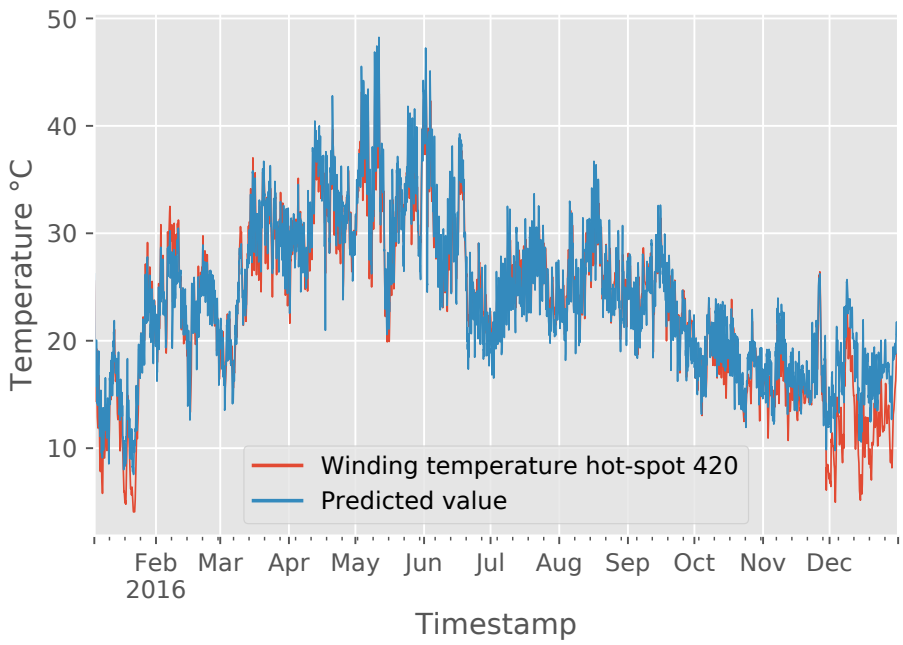

Figure 3.11: Plots of predicted values and measured values of hot-spot temperature for Data Set X using Linear Regression with 50% test size.

Figure 3.12 shows an excerpt of the same plots, of approximately one week, to better illustrate the details.

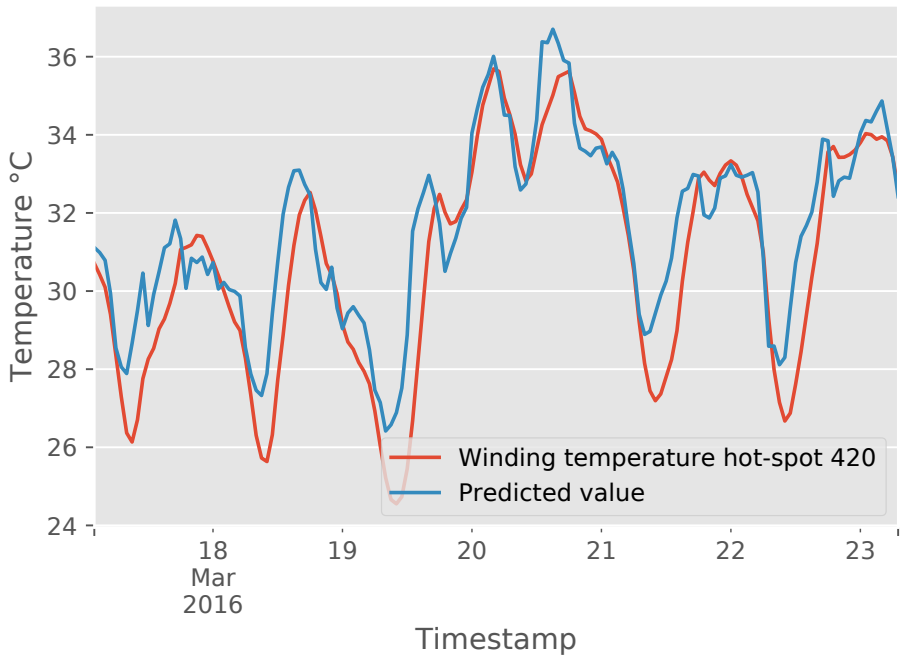

Figure 3.12: Excerpt of the plots of predicted values and measured values of hot-spot temperature for Data Set X using Linear Regression with 50% test size.

Table 3.7 lists the results of the model evaluation. By comparing these results with Table 3.6, it can be seen that the model now performs better. R^2 for the testing subset is now higher, and $RMSE$ and MAE are lower. R^2 for the training subset has however decreased somewhat. It is although more important that the R^2s are approximately equal for consistency reasons.

Table 3.7: Model evaluation for Linear Regression applied to Data Set X, with a test size of 50%.

Name	Value
R^2 for training	0.913
R^2 for testing	0.896
$RMSE$	2.48°C
MAE	1.86°C

It is not desirable to predict a lower hot-spot temperature than the actual value. This is because the purpose of this case study was to predict the maximum temperature in the transformer, making it possible for Statnett to determine how much they can load a transformer short-term. As from theory (Section 2.2), the hot-spot temperature increases as the load increases, and the transformer lifetime is reduced. Therefore, predicting a hot-spot temperature lower than the actual may result in false security.

In this case, the mean difference between the predicted values and the measured values of hot-spot temperature was 0.80°C, with the model generally predicting a higher value than the measured value, which is more desirable than predicting a lower value.

The parameter *test_size* in Listing 3.2 was changed and set equal to 0.5 before further algorithms were applied.

Ridge Regression

Ridge Regression is an extension of Linear Regression, and addresses some of the problems with multicollinearity. It is a regularised regression model and imposes a penalty size on the coefficients. The ridge coefficients minimise a penalised residual sum of squares, called L2 regularisation, as shown in Equation 3.3 [57]:

$$\min_{\omega} \left\| X \cdot \omega - y \right\|_2^{\,2} + \alpha \cdot \|\omega\|_2^{\,2} \tag{3.3}$$

Here, α is a complexity parameter that controls the amount of shrinkage - the larger the value of α, the greater the amount of shrinkage, thus making the coefficients more robust to multicollinearity, and reducing overfitting [57]. However, if α becomes too large, it will lead to underfitting. α forces the coefficient to be lower, but it does not force them to zero. Therefore, it will not remove irrelevant features, but rather minimise their impact on the trained model.

As can be seen from Figure 3.8, Ridge Regression is recommended when the data set contains less than 100K measurements, and when not only a few features are important. Data Set X contains 17415 measurements, as can be seen in Table 3.5 (count).

Ridge Regression was applied to Data Set X, using the same method as in Listing 3.3. Normalisation is important when using algorithms with regularisation. This is because the scale of the predictor variables affects how much regularisation will be applied to the specific predictor variable. The Ridge Regression algorithm has an input setting parameter called *normalize* which was set equal to *True*. In addition, α had to be specified by the parameter *alpha*. Different values of α were applied, and plots of the scores for R^2 for the training and testing subsets, with different values of α, is shown in Figure 3.13.

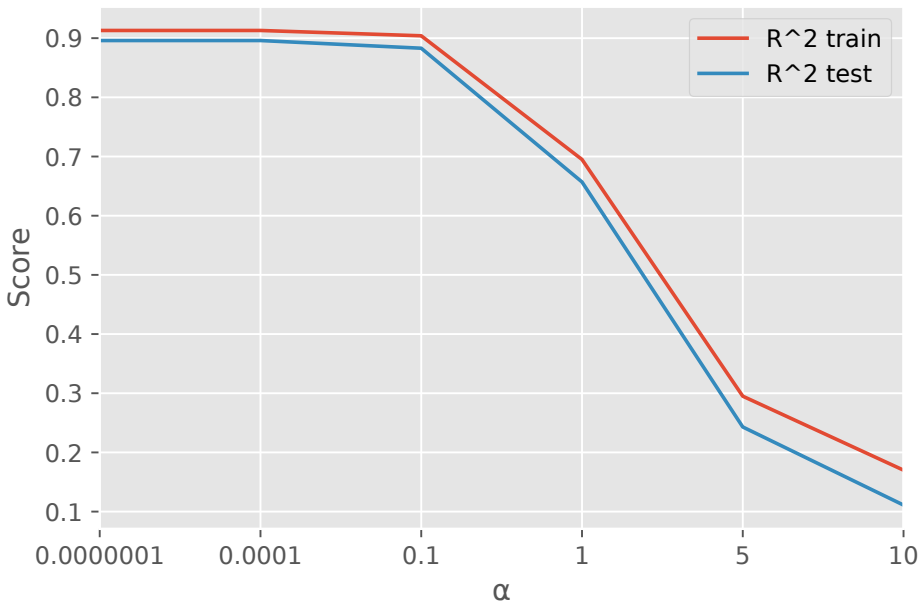

Figure 3.13: Plots of scores for the training and testing subsets with different values for α for Ridge Regression applied to Data Set X, with a test size of 50%.

The figure shows that the performance of the model is reduced as α increases. The model performs best when α is as small as possible, which results in the model being equal to the one obtained from Linear Regression. This indicates that the model was not overfitted.

Lasso Regression

Lasso Regression is another extension of Linear Regression. Similar to Ridge Regression, Lasso Regression is a regularised regression model that imposes a penalty size on the coefficients. Lasso Regression uses absolute values in the penalty function, called L1 regularisation, instead of squares, as shown in Equation 3.4 [57]:

$$\min_{\omega} \frac{1}{2n_{samples}} \cdot \left\| X \cdot \omega - y \right\|_2^2 + \alpha \cdot \|\omega\|_1 \tag{3.4}$$

As absolute values are used, the coefficients can become exactly zero. As α increases, the coefficients regress towards absolute zero. This results in feature selection out of n given features. Figure 3.8 recommends Lasso Regression when few features should be important.

Lasso Regression was applied to Data Set X. As with Ridge Regression, *normalize* was set equal to *True*. Figure 3.14 shows plots of the coefficients for different values of α. It can be seen that all the coefficients become zero when α is equal to 0.1, giving a model which is unable to predict. When α equals 0.0000001 to 0.001, it results in the coefficients being approximately of the same value as for Linear Regression, and consequently having the same model performance. However, when α equals 0.01, some penalisation has occurred.

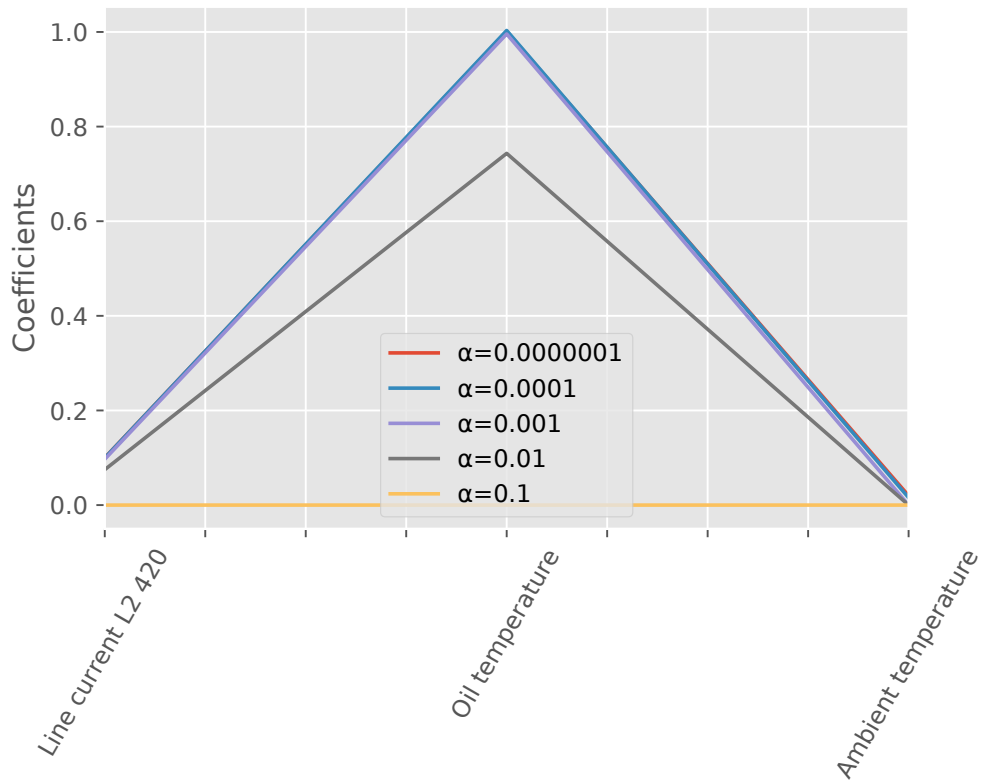

Figure 3.14: Plots of coefficients with different values for α for Lasso Regression applied to Data Set X, with a test size of 50%.

ElasticNet Regression

ElasticNet is a combination of Ridge and Lasso Regression. The combination allows for creating a model where few of the coefficients can be zero as in Lasso Regression, while still maintaining the regularisation properties of Ridge Regression. Equation 3.5 shows the function to minimise in this case [57]:

$$\min_{\omega} \frac{1}{2n_{samples}} \cdot \left\| X \cdot \omega - y \right\|_2^{\,2} + \alpha \cdot \rho \cdot \|\omega\|_1 + \frac{\alpha \cdot (1-\rho)}{2} \cdot \|\omega\|_2^{\,2} \tag{3.5}$$

The combination between L1 and L2 regularisation is controlled by using a parameter called $l1\_ratio$, which is symbolised as ρ in the equation.

ElasticNet Regression was applied to Data Set X. α was set equal to 0.0000001 as that gave the best results for both Ridge and Lasso Regression. The parameter *normalize* was also set equal to *True*. The default value for $l1\_ratio$ is equal to 0.5, which also was used for this work.

SVR

Support Vector Regression (SVR) is a type of learning algorithm that instead of minimising the observed training error, attempts to minimise the generalised error, to achieve better general performance. SVR is based on the computation of a linear regression function in a high dimensional feature space where the input data are mapped by a nonlinear function [58]. SVR penalises values predicted outside a certain distance, ϵ, while no penalty is given to the predicted values within the distance ϵ from the actual values.

As can be seen from Figure 3.8, SVR is recommended when not only a few features should be important. It also recommends the parameter *kernel* to be set to *linear* or *rbf* (radial basis function) if *linear* is not working. The kernel functions transform the data into the higher dimensional feature space to make it possible to perform the linear separation.

SVR was applied to Data Set X. The parameter *kernel* was initially set equal to *linear* before *rbf* also was applied.

Ensemble Regressors

Ensemble Regressors combine the predictions of several base estimators built with a given learning algorithm in order to improve generalisability/robustness of a single estimator [59].

There are two methods used for ensemble regressors:

- **Averaging methods**: the principle is to build several estimators independently and then to average their predictions. On average, the combined estimator is usually better than any of the single base estimators as its variance is reduced.
- **Boosting methods**: by contrast, base estimators are built sequentially and one tries to reduce the bias of the combined estimator. The aim is to combine several weak models to produce a powerful ensemble.

Random Forest Regressor is an ensemble regressor using the averaging method. It is based on decision trees, where the goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. It fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and to control overfitting [59].

AdaBoost Regressor is an ensemble regressor using the boosting method. The principle is to fit a sequence of weak learners on repeatedly modified versions of the data. The predictions are then combined through a weighted majority sum to produce the final prediction [59].

Both Random Forest Regressor and AdaBoost Regressor were applied to Data Set X. All the input setting parameters were set equal to the default except the parameter *random_state*, which was set equal to 0, for reasons of reproducibility.

Comparison of Results

Table 3.8 lists the results of the model evaluation for the different machine learning regression algorithms being applied to Data Set X. Almost all of the algorithms perform well with approximately the same results. An exception is SVR (kernel='rbf'), which performs worse than the others. Ridge Regression, Lasso Regression, and ElasticNet Regression are extensions of Linear Regression, and they all perform similarly. As feature selection and cleaning of data were performed prior to applying the algorithms, the extension algorithms did not improve the performance on Data Set X. All of these algorithms are recommended by scikit-learn when using regression supervised learning, as shown in Figure 3.8. It is therefore not surprising that most of the algorithms perform well. Other algorithms could also have been applied, but as the ones applied are performing well, it was decided to stay with these. As explained in Section 1.1 and shown in Figure 1.1, there are several skills needed in data science. The domain expertise is the focus of this thesis work, and the opportunities gained are therefore of more importance than applying advanced algorithms to improve the accuracy of the models.

Table 3.8: Results of the model evaluation for different regression algorithms being applied to Data Set X, with a test size of 50%.

Regression algorithm	R^2 for training	R^2 for testing	$RMSE$	MAE
Linear Regression	0.913	0.896	2.48°C	1.86°C
Ridge Regression	0.913	0.896	2.48°C	1.86°C
Lasso Regression	0.913	0.896	2.48°C	1.86°C
ElasticNet Regression	0.913	0.896	2.48°C	1.86°C
SVR (kernel='linear')	0.912	0.896	2.47°C	1.88°C
SVR (kernel = 'rbf')	0.772	0.422	5.82°C	3.98°C
Random Forest Regressor	0.989	0.893	2.51°C	1.91°C
AdaBoost Regressor	0.910	0.870	2.76°C	2.07°C

This case study shows that it is possible to predict transformer hot-spot temperature when using the line current (load), oil temperature and ambient temperature, which are the main influencers of the hot-spot temperature. The performance of several learning algorithms shows good accuracy for this case. However, to be able to use these results further in the case study, some adjustments had to be done. Firstly, in every transformer, there is a different dependency between the oil temperature and the hot-spot temperature. This depends on the design of the transformer, and also how the oil flows. In addition, there is a dependency between the oil temperature and the load. To be able to determine how much a transformer can be loaded short-term, the hot-spot temperature should be predicted without considering the oil temperature. It will then be possible to increase the line current (load), while not taking into account how that influences the oil temperature. Therefore, it was decided to try to predict the hot-spot temperature without using the oil temperature.

Exclusion of Oil Temperature

The feature *Oil Temperature* was excluded from Data Set X. The different algorithms were applied to the data set, using the same method as in Listing 3.3. Table 3.9 lists the results of the model evaluation. By comparing the results to Table 3.8, it can be seen that the algorithms now perform less well than when the oil temperature was included. There is approximately a 10% difference in the scores for R^2. The $RMSE$ and MAE have also increased. However, all the algorithms have a score of around 0.8, which is much closer to 1 than 0, and satisfactory for the purpose of this thesis as the domain expertise is the focus.

Table 3.9: Results of the model evaluation for different regression algorithms being applied to Data Set X, with a test size of 50% and the feature *Oil Temperature* excluded.

Regression algorithm	R^2 for training	R^2 for testing	$RMSE$	MAE
Linear Regression	0.813	0.813	3.32°C	2.58°C
Ridge Regression	0.813	0.813	3.32°C	2.58°C
Lasso Regression	0.813	0.813	3.32°C	2.58°C
ElasticNet Regression	0.813	0.813	3.32°C	2.58°C
SVR(kernel='linear')	0.813	0.810	3.34°C	2.59°C
SVR(kernel = 'rbf')	0.767	0.465	5.61°C	3.94°C
Random Forest Regressor	0.970	0.787	3.54°C	2.74°C
AdaBoost Regressor	0.832	0.780	3.59°C	2.76°C

This shows that it is possible to predict hot-spot temperature using only the line current and the ambient temperature. The accuracy could however be improved. This can possibly be done by optimising the algorithms or applying other algorithms, or by implementing the other skills needed in data science, as shown in Figure 1.1. As mentioned, Linear Regression is one of the most commonly used modelling techniques. As it is also performing well on Data Set X, it was decided to continue using this algorithm. Figure 3.15 shows plots of the predicted values and the measured values of hot-spot temperature from the testing subset for Data Set X using Linear Regression with 50% test size, with the feature *Oil temperature*

removed. By comparing this figure with Figure 3.11, it can be seen that there now are larger deviations between the predicted values and measured values of hot-spot temperature. However, there is a satisfactory correspondence. The model is generally predicting a higher value than the measured value - the mean difference being 0.18°C. As previously discussed, this is more desirable than predicting a lower value.

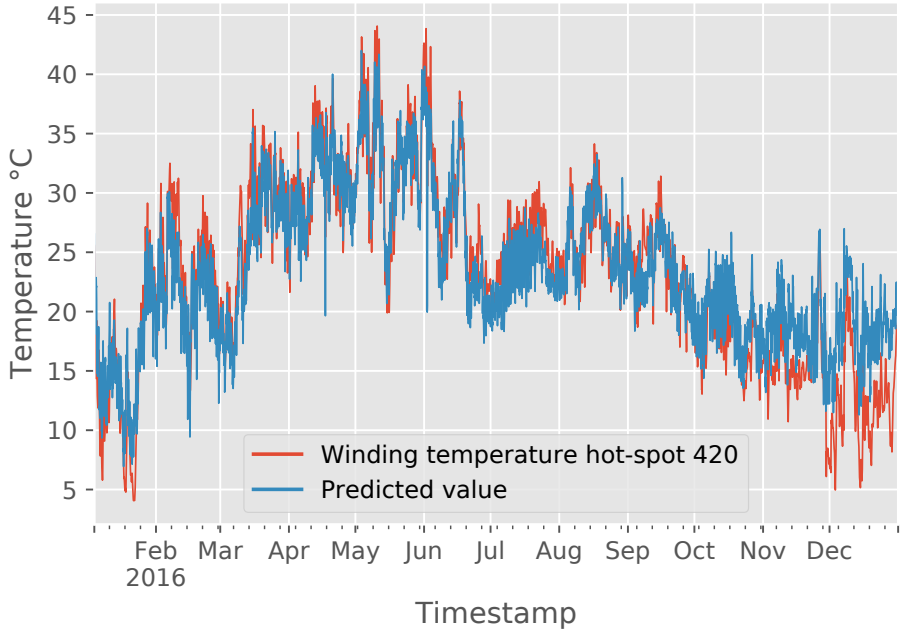

Figure 3.15: Plots of predicted values and measured values of hot-spot temperature for Data Set X using Linear Regression with 50% test size, with the feature *Oil temperature* removed.

Figure 3.16 shows an excerpt of the same plots, of approximately one week, to better illustrate the details. Comparing this figure with Figure 3.12, it is easy to see that the correspondence is now reduced.

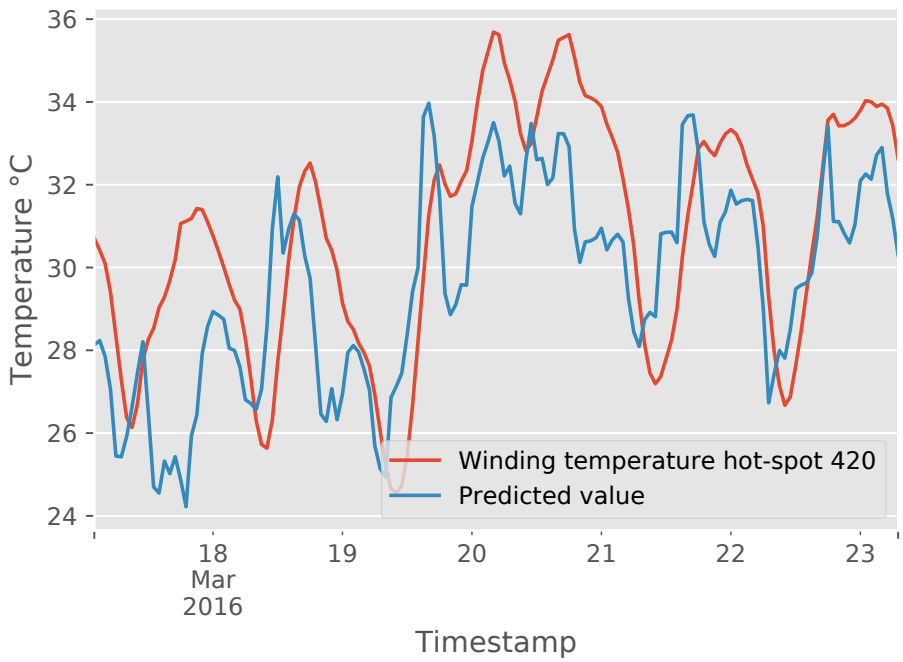

Figure 3.16: Excerpt of the plots of predicted values and measured values of hot-spot temperature for Data Set X using Linear Regression with 50% test size, with the feature *Oil temperature* removed.

Table 3.10 summarises the characteristics of the model, called Model X, which was made and chosen to be utilised further in this thesis work.

Table 3.10: Characteristics of Model X.

Description	Characteristic
Transformer	Transformer X
Data set	Data Set X
Learning algorithm	Linear Regression
$test\_size$	0.5
Features	*Winding temperature hot-spot 420* (target)
	Line current L2 420 (predictor)
	Ambient temperature (predictor)

Maximum Acceptable Loading

Being able to predict the hot-spot temperature in a transformer creates many opportunities. According to theory (Section 2.2), the hot-spot temperature is a critical temperature in the transformer and is given a standard upper limit of 105°C. Based on this, it is possible to determine how much the transformer can be acceptably loaded, for example in cold periods, or when there is a need for emergency loading.

The feature *Line Current L2 420* in Data Set X was therefore increased, before Model X described in Table 3.10 was applied. The load was increased until the highest predicted hot-spot temperature reached the upper limit of 105°C in the testing subset. The feature *Line Current L2 420* was multiplied by 3.5 to reach a maximum predicted hot-spot temperature of 105°C. This indicates that Statnett could have increased the load of Transformer X by 250% before the hot-spot temperature reached 105°C in the testing subset. Figure 3.17 shows plots of the predicted values and the measured values of hot-spot temperature from the testing subset, when *Line Current L2 420* was increased by 250%. The arrow indicates the hot-spot temperature of 105°C.

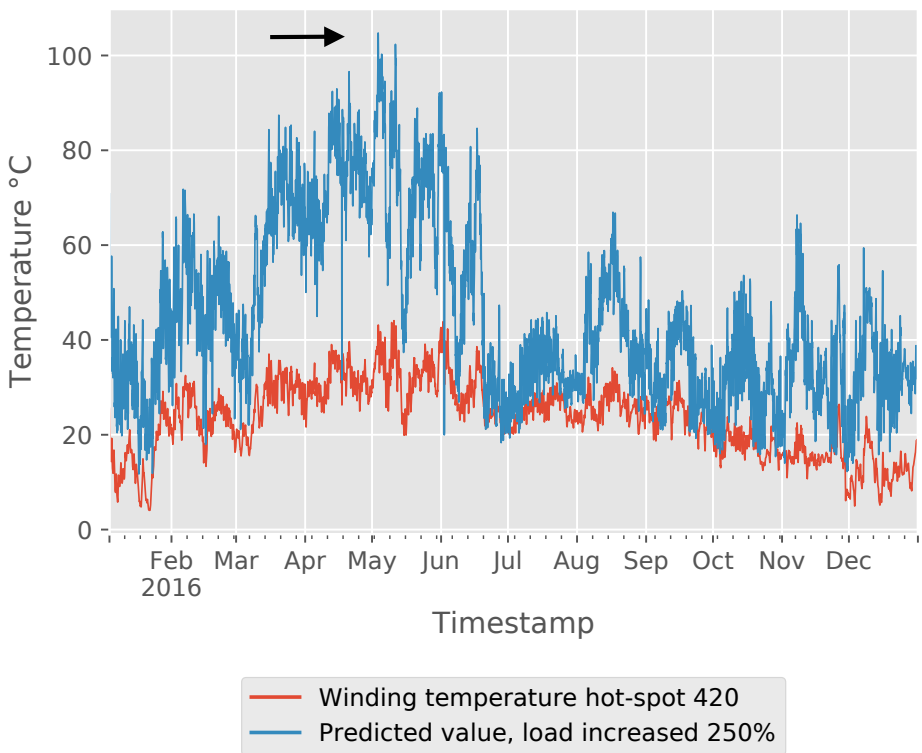

Figure 3.17: Plots of predicted values and measured values of hot-spot temperature for Data Set X, when *Line Current L2 420* is increased by 250%.

However, 250% is quite a large increase of the load. It is possible that the model is not valid for such big changes, as it has not been trained properly on data for those changes.

Application on other Data Sets

During the thesis work, more data sets containing hot-spot temperature by fiber optic sensors were requested from Statnett, with the aim of determining how successful machine learning would be when applied to other data sets. 153 new data sets containing time series for 153 power transformers were then provided, described in Table 3.11.

Table 3.11: Description of data sets provided by Statnett by request.

Category	Description	
Data	Obtained from	*Innsikt*
	Number of data sets	153
	Format	HDF5
Transformers	Number of power transformers	153
	Voltage level	420 kV or 300 kV
Time	Time period	01.01.2005 - 01.01.2018
	Sampling rate	One-hour
Measurements	Criteria	Winding temperature
		Oil temperature
		Line current (load)

These data sets were obtained in the same way, and were similar to the previously provided data sets described in Table 3.1, but were collected during the time period 01.01.2005 to 01.01.2018 at a one-hour sampling rate. Additionally, there were no criteria for gas concentration measurements, resulting in the inclusion of more data sets having measurements for hot-spot temperature. A new CSV-file was also provided from Statnett, describing what the sensors were measuring. 49 of these data sets were for transformers which were also included in the previously provided data sets, but now collected for a different time period and sampling rate.

Out of the 153 new data sets, only 29 transformers had a sensor installed supposedly measuring hot-spot temperature. However, 12 of these sensors were not actually measuring. Figure 3.18 shows the measured hot-spot temperature of four of the remaining data sets. These hot-spot temperatures are not of good enough quality for data science because they show values of hot-spot temperature having instantaneous changes of considerable magnitude or long periods of constant value. This is either physically impossible, or exceedingly unlikely.

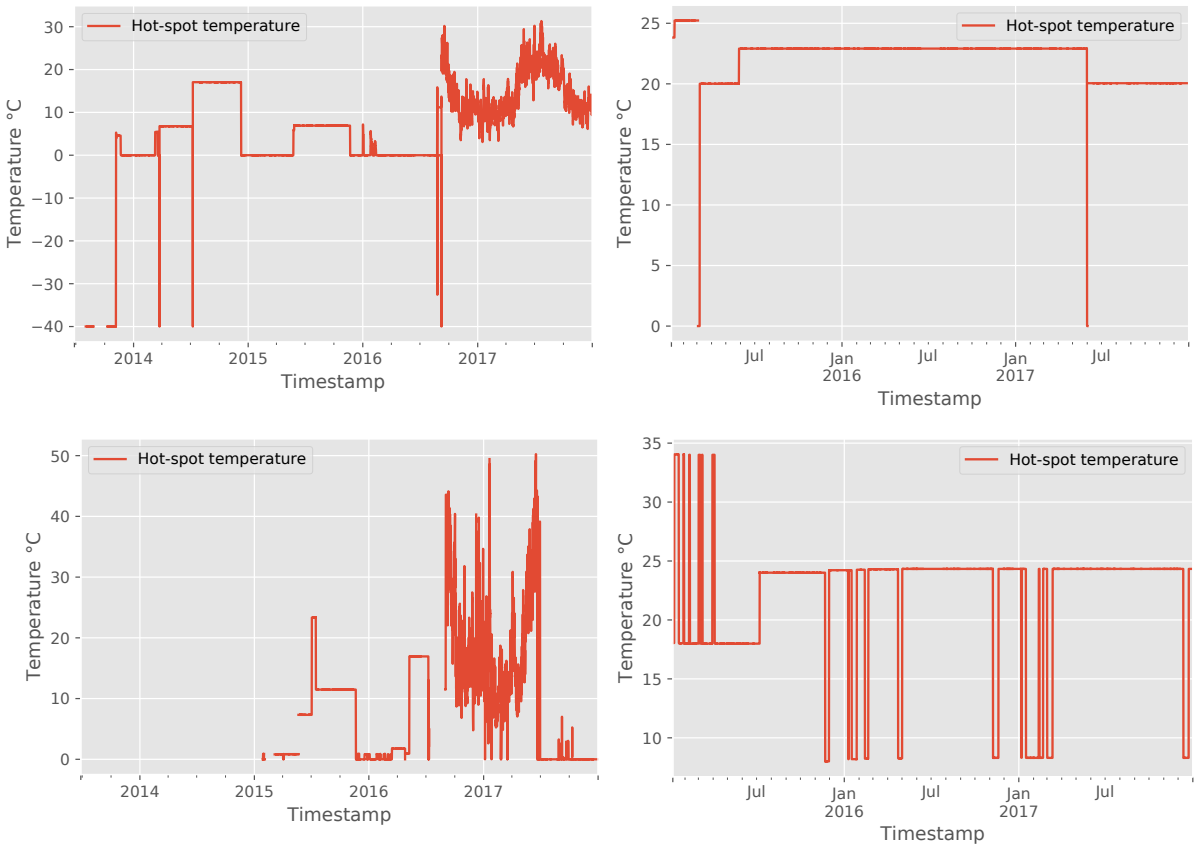

Figure 3.18: Examples of unreliable hot-spot temperature measurements.

Only two of the remaining data sets seemed of good enough quality for applying machine learning to make models, and met all of the requirements in Table 3.3. These sets are called Data Set Y and Data Set Z for this thesis work. The other remaining data sets had very short measuring periods, making them less suitable for applying machine learning.

It was therefore decided to make new machine learning models on Data Set Y and Data Set Z. In addition, it was decided to try use Model X described in Table 3.10 to attempt to predict hot-spot temperatures from the data sets.

Data Set Y contains data from a 300 MVA 420/132/22 kV power transformer, called Transformer Y for this thesis work. The measurements of quality are taken in the time period 19.04.2015 to 31.12.2017. The data set was prepared as in Section 3.2, with the features *Line current L2 420*, and *Winding temperature hot-spot 420* chosen, with the feature *Ambient temperature* downloaded from met.com added. The data set was also cleaned for outliers and NaNs. Linear Regression was applied to Data Set Y. Different test sizes were applied to determine which gave the best result. Table 3.12 lists the best obtained results, which occurred when the parameter $test\_size$ was set equal to 0.2.

Table 3.12: Model evaluation for Linear Regression applied to Data Set Y, with a test size of 20%.

Name	Value
R^2 for training	0.726
R^2 for testing	0.741
$RMSE$	2.66°C
MAE	2.17°C

As it can be observed from the table, Linear Regression performs well on Data Set Y. However, comparing these results with the results in Table 3.9, the scores of R^2 of this model are poorer than the results of Model X made on Data Set X described in Table 3.10. The $RMSE$ and MAE are however lower, which is desirable. The best results were obtained when there was an 80% train and 20% test split. This is from theory (Subsection 2.3.3) the most common starting point. The measurements of Data Set Y were taken over a longer time period (approximately 9 months) than Data Set X, which may be the reason behind the need for the different split. Figure 3.19 shows plots of the predicted values and the measured values of hot-spot temperature.

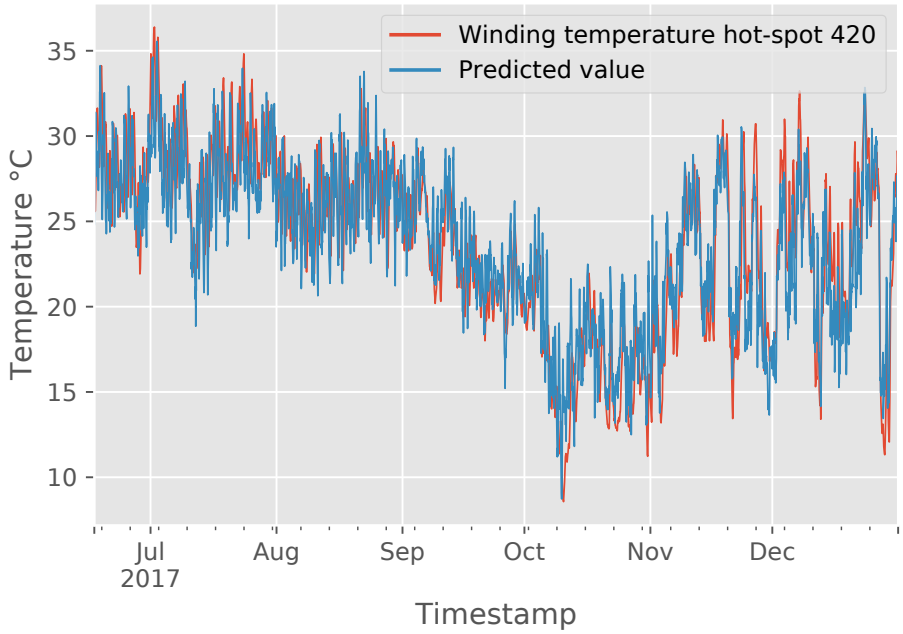

Figure 3.19: Plots of predicted values and measured values of hot-spot temperature for Data Set Y using Linear Regression with 20% test size.

Data Set Z contains data from a 200 MVA 300/47/11 kV power transformer, called Transformer Z for this thesis work. The measurements of quality are taken during the time period 01.01.2016 to 31.12.2017. This data set was also prepared as in Section 3.2, with the features *Line current L2 300* and *Winding temperature hot-spot 300* chosen, and with the added feature *Ambient temperature* downloaded from met.com. The data set was also cleaned for outliers and NaNs. Linear Regression was applied to Data Set Z. Different test sizes were applied to obtain the best result. Table 3.13 lists the best obtained results, which occurred when the parameter $test\_size$ was set equal to 0.3.

Table 3.13: Model evaluation for Linear Regression applied to Data Set Z, with a test size of 30%.

Name	Value
R^2 for training	0.707
R^2 for testing	0.812
$RMSE$	0.92°C
MAE	0.73°C

It can be observed that Linear Regression also performs well on Data Set Z. Figure 3.20 shows

plots of the predicted values and the measured values of hot-spot temperature.

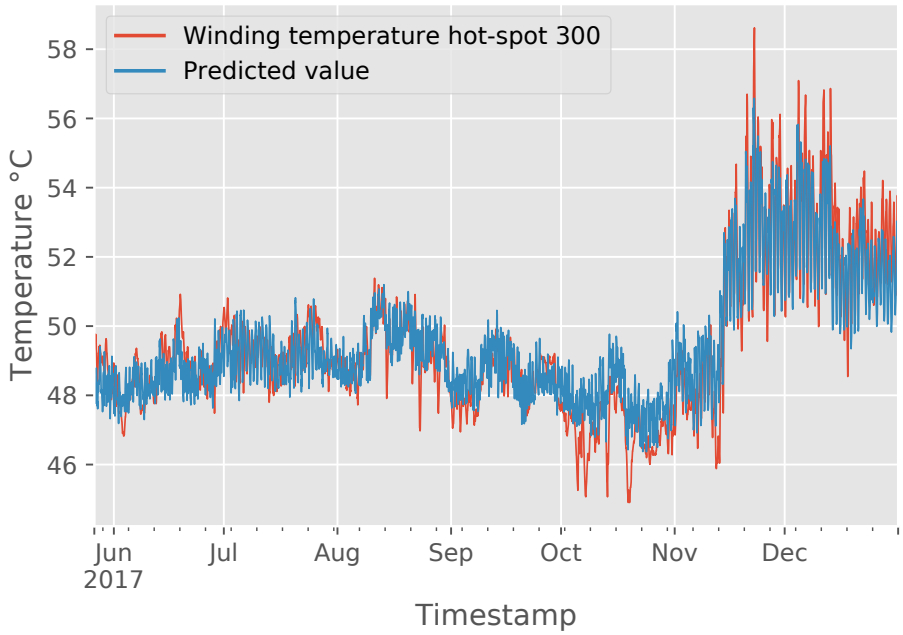

Figure 3.20: Plots of predicted values and measured values of hot-spot temperature for Data Set Z using Linear Regression with 30% test size.

This shows that it is possible to apply machine learning to several data sets to make models for predicting hot-spot temperature with satisfactory accuracy. It was also decided to apply Model X described in Table 3.10 to Data set Y and Data Set Z, to see how the model would perform on other transformer data sets. Table 3.14 lists the results of the model evaluation.

Table 3.14: Model evaluation for Model X applied to other data sets.

Data Set	R^2 for testing	$RMSE$	MAE
Data Set Y	0.614	3.33°C	2.69°C
Data Set Z	-216.05	26.96°C	26.81°C

Comparing the results for Data Set Y in Table 3.14 and Table 3.12, it can be seen that the results are worse (lower R^2 and higher $RMSE$ and MAE) when Model X is applied to the data set, rather than using the model specifically made for Transformer Y. For Data Set Z, the results are even worse, with a negative R^2. This indicates from theory (Subsection 2.3.3) that a horizontal straight line explains the data equally as well as the model. This illustrates that a model made specifically for one transformer should not be applied to another transformer. One individual model must be made for each transformer. This is due to the dissimilar relationships between different parameters in a transformer.

3.3.2 Winding Insulation Degradation Estimation

For this case study, the goal was to use predicted hot-spot temperatures for Transformer X obtained using Model X, described in Table 3.10, to estimate winding insulation degradation for different scenarios. This was done to estimate how much the different scenarios could affect the estimated remaining lifetime of the transformer. The predicted and measured hot-spot temperatures were also compared to calculated hot-spot temperatures using the loading guide.

Degree of Polymerisation

Using the predicted hot-spot temperature for Transformer X obtained using Model X, the degree of polymerisation, DP, can be calculated using the winding insulation degradation model explained in Subsection 2.2.2. Detailed temperature and load history for Transformer X is as seen in Table 3.4 given for approximately two years, 06.01.2015 to 31.12.2016. As the parameter *test_size* for Model X was chosen to be 0.5, the temperatures were predicted for approximately one year, 04.01.2016 to 31.12.2016. Therefore, Equation 2.6 has to be used, assuming the temperature and load history for one year is similar for subsequent years. In order to be able to use this equation, the average environment factor, $\overline{A}$, and the concentration of water in paper, $w_{p,n}$, has to be calculated. This case study aims to calculate the DP-value for today (11.06.2018), this means that the value of $w_{p,n}$ for today has to be calculated.

The value of $w_{p,n}$ is assumed from theory (Subsection 2.2.2) to have been 0.5% when Transformer X was new on 01.10.2014. The measurements from Data Set X start on 06.01.2015, and as can be seen from Table 3.4, include the feature *Moisture concentration*, which measures the concentration of water in oil, $w_{o,n}$. On 06.01.2015, $w_{o,n}$ was measured to be 2.275%. This can be used to calculate the value of $w_{p,n}$, using Equation 2.5. The hot-spot temperature, θ_h, at the time of the measurement is also needed for the equation, and is included in the feature *Winding temperature hot-spot 420*. On 06.01.2015, the hot-spot temperature was measured as 24.91°C.

The concentration of water in paper for Transformer X on 06.01.2015, $w_{p,n,2015}$, can then be calculated using Equation 2.5:

$$w_{p,n,2015} = (2.275 \cdot 2.24e^{-0.04 \cdot 24.91})^{0.63} = 1.489 \text{ [\%]} \tag{3.6}$$

There are 97 days between when the transformer was new on 01.10.2014, and 06.01.2015, which means that the age of the transformer, $\tau_{2014,2015}$, equals 2328 hours on 06.01.2015. Using Equation 2.9, the incremental factor for the concentration of water in paper, ϕ, can be found:

$$\phi = \frac{1.489 - 0.5}{2328} = 4.249 \cdot 10^{-4} \tag{3.7}$$

The number of days between 01.10.2014 and today, 11.06.2018, are 1349, which means that the age of the transformer, $\tau_{2014,2018}$, equals 32376 hours on 11.06.2018. The concentration of water in paper, $w_{p,n,2018}$, for Transformer X on 11.06.2018 can then be calculated by using Equation 2.9 and solving for $w_{p,n,2018}$ instead of $\tau_{2014,2018}$:

$$w_{p,n,2018} = 0.5 + 4.249 \cdot 10^{-4} \cdot 32376 = 14.256 \text{ [\%]} \tag{3.8}$$

Having calculated the value for $w_{p,n,2018}$, the average environment factor, $\overline{A}$, from 01.10.2014 to 11.06.2018 can be calculated using Equation 2.8, with the multiplication factor for oxygen, O, being equal to 2.5 from theory (Subsection 2.2.2):

$$\overline{A} = \frac{(13000 \cdot 0.5 + 14000) \cdot 2.5 + (13000 \cdot 14.256 + 14000) \cdot 2.5}{2} = 274786.8 \tag{3.9}$$

The DP-value for Transformer X on 11.06.2018 can now be calculated using Equation 2.6. Transformer X is constructed with thermally upgraded paper of the Insuldur type, whose activation energy, E_a, is 86 kJ/mol from theory (Subsection 2.2.2). The DP-value from when the transformer was new in 2014, DP_{new}, is assumed to be 1000 from theory (Subsection 2.2.2). The hot-spot temperature $\theta_h(m)$ at the operating condition m has been predicted a total number of M = 8708 times at one-hour sampling rate during approximately one year. The values needed for use in Equation 2.6 are summarised in Table 3.15:

Table 3.15: Values for Transformer X used in Equation 2.6.

Description	Symbol	Value
Degree of polymerisation when the transformer was new	DP_{new}	1000
Average environment factor	$\overline{A}$	274786.8 [h^{-1}]
Age of transformer	$\tau_{2014,2018}$	32376 [hours]
Activation energy	E_a	86000 [J mol^{-1}]
Gas constant	R_g	8.31 [J mol^{-1} K^{-1}]
Number of hot-spot temperature measurements	M	8708

Equation 2.6 can then be written as:

$$DP_{today} = \frac{1000}{1 + 1000 \cdot 32376 \cdot 274786.8 \cdot \frac{\sum_{m=1}^{8708} e^{\frac{-86000}{8.31 \cdot (\theta_h(m)+273)}}}{8708}} \tag{3.10}$$

The result for DP_{today} based on the predicted hot-spot temperature is listed in Table 3.16. DP_{today} based on the measured values of hot-spot temperature is also shown for comparison. In addition, DP_{today} is shown for hot-spot temperatures predicted for when the load is increased and the ambient temperature both increased and decreased during the whole lifetime of the transformer.

Table 3.16: DP_{today} values of Transformer X based on different hot-spot temperatures.

DP based on different hot-spot temperatures:	Symbol	Value
Measured	$DP_{today,\ measured}$	991.4
Predicted	$DP_{today,\ predicted}$	992.1
Predicted, load increased 20%	$DP_{today,\ predicted,\ load+20\%}$	989.6
Predicted, load increased 80%	$DP_{today,\ predicted,\ load+80\%}$	975.2
Predicted, load increased 250%	$DP_{today,\ predicted,\ load+250\%}$	743.3
Predicted, ambient temperature increased 2°C	$DP_{today,\ predicted,\ ambient+2°C}$	991.0
Predicted, ambient temperature increased 5°C	$DP_{today,\ predicted,\ ambient+5°C}$	988.9
Predicted, ambient temperature decreased 2°C	$DP_{today,\ predicted,\ ambient-2°C}$	993.1
Predicted, ambient temperature decreased 5°C	$DP_{today,\ predicted,\ ambient-5°C}$	994.5

As can be seen from Table 3.16, the estimated DP-values vary according to which hot-spot temperature they are based on. The DP-values of the measured and predicted hot-spot temperatures are close, which is satisfactory. They are also close to 1000, as this is a new transformer. As the load and ambient temperature are increased, the predicted hot-spot temperatures also increase. Higher hot-spot temperature leads to accelerated ageing of the transformer and a lower DP-value. This is confirmed by the results in the table. As the ambient temperature is decreased, the hot-spot temperature also decreases, leading to slower ageing and consequently higher DP-value.

Remaining Lifetime

The remaining lifetime of Transformer X can also be estimated using Equation 2.6. From theory (Subsection 2.2.2), it is common in Norway to define the end of the lifetime of the paper as the time when the DP-value reaches 200. Having already calculated the values that are included in Equation 2.6, the fact that the value of DP_{today} will be 200 at the end of the lifetime of the paper can be used.

Using the values in Table 3.15 and DP_{today}=200, the age of the transformer, τ, can be calculated. The concentration of water in paper, $w_{p,n}$, will also be a function of τ, used in the equation for the average environment factor, $\overline{A}$. Equation 2.6 will then be as follows, where 51250 is the calculated value of A_{new}:

$$200 = \frac{1000}{1 + 1000 \cdot \tau \cdot \frac{51250 + (13000 \cdot (0.5 + 4.249 \cdot 10^{-4} \cdot \tau) + 14000) \cdot 2.5}{2} \cdot \frac{\sum_{m=1}^{8708} e^{\frac{-86000}{8.31 \cdot (\theta_h(m) + 273)}}}{8708}} \tag{3.11}$$

MATLAB was used to solve for τ. The estimated ages of Transformer X at which the value of DP_{today} equals 200, based on the diversely obtained hot-spot temperatures, are listed in Table 3.17 (ages converted to years). The associated remaining lifetimes of Transformer X can also be estimated based on these estimated ages. From when the transformer was new on 01.10.2014 until today, 11.06.2018, there are 3.7 years. Subtracting 3.7 from the estimated ages consequently gives the remaining lifetimes of Transformer X as of today. The associated remaining lifetimes as of 11.06.2018 are also listed in Table 3.17.

Table 3.17: Age of Transformer X when DP_{today}=200, and remaining lifetime as of 11.06.2018 [years].

Based on different hot-spot temperatures:	Age [years]	Remaining lifetime [years]
Measured	87.9	84.2
Predicted	91.6	87.9
Predicted, load increased 20%	79.6	75.9
Predicted, load increased 80%	51.0	47.3
Predicted, load increased 250%	13.7	10.0
Predicted, ambient temperature increased 2°C	85.5	81.8
Predicted, ambient temperature increased 5°C	77.0	73.3
Predicted, ambient temperature decreased 2°C	98.3	94.6
Predicted, ambient temperature decreased 5°C	109.4	105.7

As can be seen, there are considerable variations in the estimated remaining lifetime for Transformer X depending on which hot-spot temperature is used to determine it. The results correlate with the DP-values in Table 3.16. They show that it is the hot-spot temperature with the highest DP-value (lowest hot-spot temperature) that gives the longest remaining lifetime.

The predicted hot-spot temperature results in a longer estimated remaining lifetime than the measured hot-spot temperature - an increase of 4.4%. Looking at Figure 3.15, it can be observed that the predicted hot-spot temperature is lower than the measured hot-spot temperature when the hot-spot temperatures are at their highest. Although the difference is quite modest, it has an effect on the estimated remaining lifetime. Figure 3.21 shows a plot of the remaining lifetime of Transformer X with increasing loads.

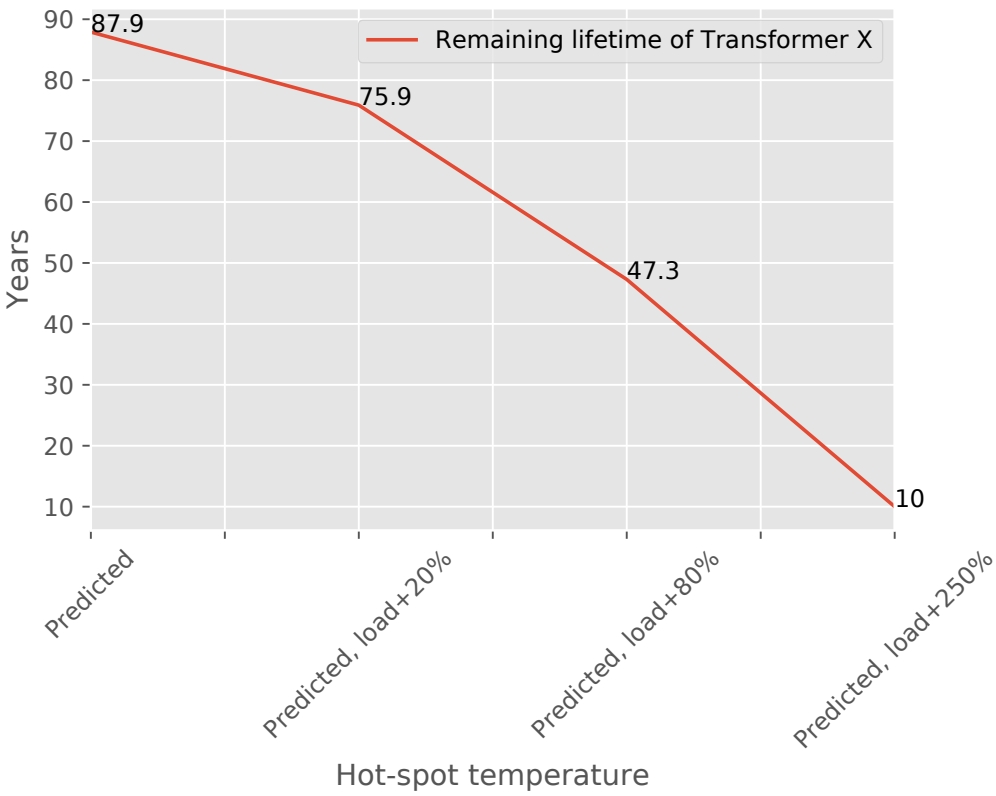

Figure 3.21: Plot of remaining lifetime of Transformer X with increasing loads.

The shortest estimated remaining lifetime is obtained from the predicted hot-spot temperature where the load is increased by 250%. This was the maximum acceptable loading before the hot-spot temperature reached the upper limit of 105°C. This results in a decrease in DP-value of 25.1% which further results in a decrease in remaining lifetime of 88.6%, compared to the results from the original predicted hot-spot temperature (without increased load). However, as already mentioned in Subsection 3.3.1, 250% is quite a large increase of the load. It is far outside the range of values where the machine learning model is trained. It is therefore likely that the results are not as accurate for such big increases of the load.

Comparison with Loading Guide

It is also possible to calculate the hot-spot temperature using the loading guide explained in Subsection 2.2.1. By calculating the hot-spot temperature using the loading guide based on the ambient temperature and the top-oil temperature, the results can be compared to the ones predicted using machine learning. Equations 2.1 and 2.2 were used for this purpose.

The specific values needed in Equations 2.1 and 2.2 were taken from the temperature rise test in a factory acceptance test (FAT) for the transformer type of Transformer X, and are listed in Table 3.18. To reduce the complexity, it was assumed only ONAN cooling. For medium and large power transformers with ONAN cooling, the IEC 60076-2 "Temperature rise for liquid-immersed transformers" [28] recommends the oil exponent and winding exponent to be 0.9 and 1.6, respectively.

Table 3.18: Specific values for Transformer X from temperature rise test.

Description	Symbol	Value
Rated top-oil temperature rise	$\Delta\theta_{or}$	46.8 [K]
Loss ratio	R	2.55
Hot-spot factor	H	1.32
Average-winding-to-average-oil	g_r	3.9 [K]
Oil exponent	x	0.9
Winding exponent	y	1.6

In addition, the ambient temperatures, θ_a, top-oil temperatures, θ_o, and the load factor, K, were obtained from Data Set X.

Equation 2.1 for calculating hot-spot temperature from ambient temperature then becomes as follows:

$$\theta_h = \theta_a + 46.8 \cdot [(1 + 2.55 \cdot K^2)/(1 + 2.55)]^{0.9} + 1.32 \cdot 3.9 \cdot K^{1.6} \tag{3.12}$$

Further, Equation 2.2 for calculation hot-spot temperature from top-oil temperature becomes as follows:

$$\theta_h = \theta_o + 1.32 \cdot 3.9 \cdot K^{1.6} \tag{3.13}$$

Figure 3.22 shows a comparison of these results and the measured hot-spot temperature by a fiber optic sensor, for the whole time period of Data Set X (06.01.2015 to 31.12.2016). It can be seen that the hot-spot temperatures calculated from ambient temperature and top-oil temperature, using the loading guide, correspond well with each other. However, they do not correspond as well with the measured hot-spot temperature obtained by a fiber optic sensor. Ideally, the three methods should give exactly the same results, but considerable differences can be seen. Small differences should however be expected when using different techniques. In addition, assumptions (such as only ONAN cooling), affects accuracy.

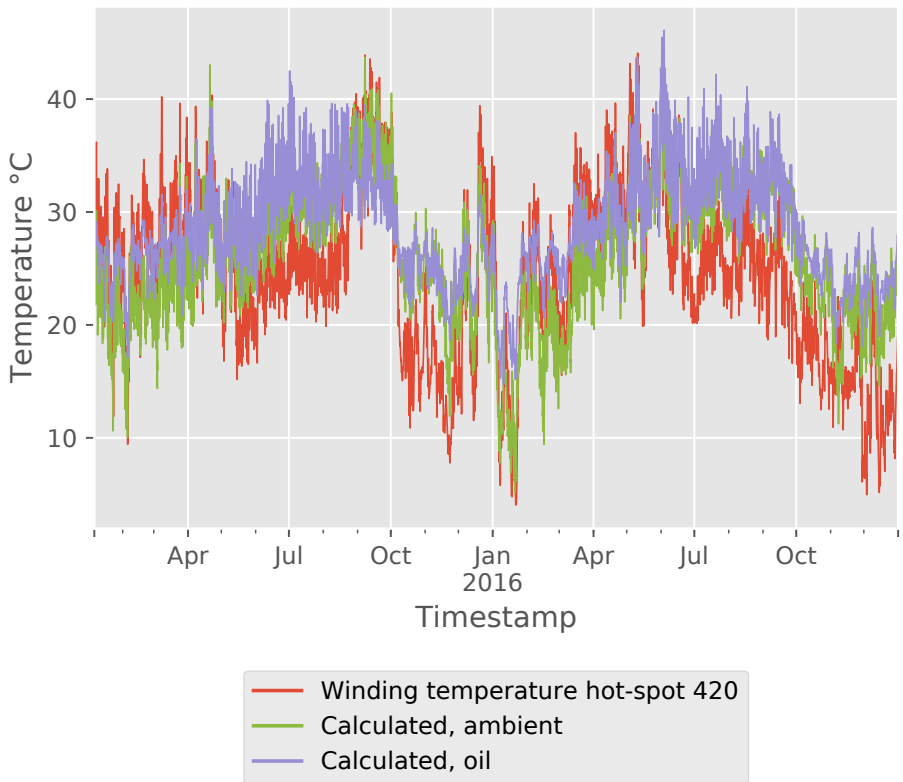

Figure 3.22: Plots of calculated and measured values of hot-spot temperature for Transformer X.

Figure 3.23 in addition shows the predicted hot-spot temperature for the time of the predictions (04.01.2016 to 31.12.2016). As can be seen, the predicted hot-spot temperature corresponds better with the measured hot-spot temperature than with the calculated.

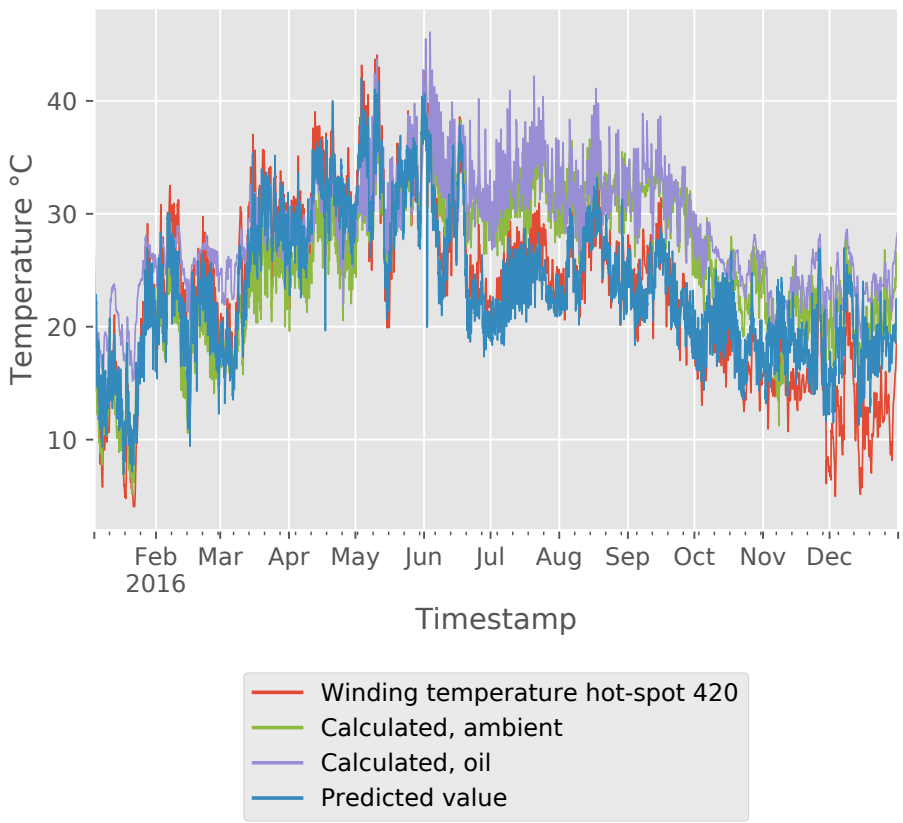

Figure 3.23: Plots of calculated, predicted and measured values of hot-spot temperature for Transformer X.

Table 3.19 lists the results of the evaluation of different methods of obtaining hot-spot temperature for Transformer X, compared to the measured temperature by a fiber optic sensor. In addition, the remaining lifetimes based on the diverse hot-spot temperatures are shown. The results confirm that the predicted hot-spot temperature corresponds better with the measured temperature than the calculated ones. The difference between the estimated remaining lifetime based on the predicted and the measured hot-spot temperature is as previously written 4.4%. The difference is 11.2% and 15.8% between the estimated remaining lifetime based on hot-spot temperatures calculated from ambient temperature and oil temperature, and measured hot-spot temperature, respectively.

Table 3.19: Evaluation of different methods of obtaining hot-spot temperature for Transformer X.

Hot-spot temperature	R^2	$RMSE$	MAE	Remaining lifetime [years]
Measured	-	-	-	84.2
Predicted	0.813	3.32°C	2.58°C	87.9
Calculated, ambient	0.358	6.14°C	5.35°C	74.8
Calculated, oil	0.154	7.05°C	6.06°C	70.9

Chapter 4

Utility Value

Based on the case studies in Chapter 3, several benefits have emerged from utilising machine learning on the power transformer component. The major benefit of machine learning is its capability of utilising the information contained in historical data to investigate properties of future data. Three uses of machine learning in power transformer asset management are proposed in this chapter.

4.1 Normal Behaviour Model

During the development of the machine learning models, it was assumed that the transformers were behaving normally. It was also assumed that the fiber optic sensors were measuring correctly. This means, it was assumed that the data used as input for the machine learning algorithms represent the normal condition of healthy transformers without anomalies and faults. As a result, the machine learning models can be used as normal behaviour models for Statnett. For this purpose, it is necessary with fiber optic sensors installed in the transformers measuring hot-spot temperature. If there are deviations between the hot-spot temperatures obtained from machine learning models and the hot-spot temperatures measured by fiber optic sensors, it can indicate that:

- The machine learning model is incorrect. For example: the wrong machine learning algorithm is being used or the model is not trained well enough. The model should then be altered or a new model could be made.
- The fiber optic sensor is malfunctioning. Sensors may fail, or be incorrectly installed. Calibrations can thus be performed if errors are discovered.
- There are anomalies in the transformer. Something is not as expected compared to the condition during normal behaviour of the transformer. For example, the cooling system is not functioning correctly.

- There is an emerging fault in the transformer. For example: the transformer has been loaded heavily, resulting in accelerated ageing of the transformer insulation.

An online overview of the machine learning models can be continuously monitored. Increased information and knowledge of the condition of the transformer will reduce the risk, and may result in Statnett gaining the confidence to operate a transformer for a longer time period (even after reaching remaining lifetime expectancy). Statnett then has the opportunity to increase its expected minimum lifetime of 60 years for power transformers closer to the potential of 80 years. A reinvestment can thus be postponed. The alternative would be to replace the transformer earlier due to insufficient assurance of its condition. On the other hand, Statnett can also discover that the transformer is in a poorer condition than initially assumed. A replacement program can then be initiated. The benefits of using machine learning as normal behaviour models are:

- Optimised maintenance can be performed if needed (if deviations in hot-spot temperature appear).
- A possible reinvestment can be postponed (if no significant deviations).
- Unnecessary resources spent on healthy transformers can be reduced.
- Verification that fiber optic sensors are measuring correctly.

The normal behaviour models can also be used as a protection support system for Statnett. The system could give a warning signal when the deviations in hot-spot temperatures reach a predetermined value or percentage. In this way, Statnett can be notified early when there is a probability of an emerging fault in the transformer and further add resources for inspection of the transformer. Compared to other protection devices for transformers, for example, a Buchholz relay[1] or Hydran[2], a machine learning support system has the potential to give an earlier warning, and may also be more precise. Other protection devices normally give a warning when a fault has already developed.

The power system is undergoing extensive digitalisation. Digital substations are based on IEDs and more use of sensors. Machine learning can be used to reveal insufficient routines on sensors. With more advanced use of machine learning, it can also be used to recognise and distinguish different emerging faults of transformers.

To illustrate the potential economic savings by using machine learning to extend the liftime of a power transformer and further postpone reinvestment, a simple economic analysis has been performed. The analysis is based on a transformer of the type 300 MVA 420/132 kV. This is a common power transformer, and the case studies have also been performed on two transformers of this type. The reinvestment cost, RC, of such a transformer is assumed to be approximately 75 MNOK, including engineering costs, purchase, assembly

[1] mechanical fault detector for electrical faults in oil-immersed transformers.
[2] online dissolved gas analysis monitoring device.

and transportation. The economic lifetime of the transformer, T, is set to 35 years, as used in Norges vassdrags- og energidirektorat (NVE) [60]. The analysis period, N, is also set to 35 years. Two alternatives (A2 and A3) for using machine learning to extend the lifetime of the transformer and further postpone reinvestment have been used: by 5 years and 10 years, compared to the alternative (A1) of performing a reinvestment in year 0 of the analysis period. This is illustrated in Figure 4.1.

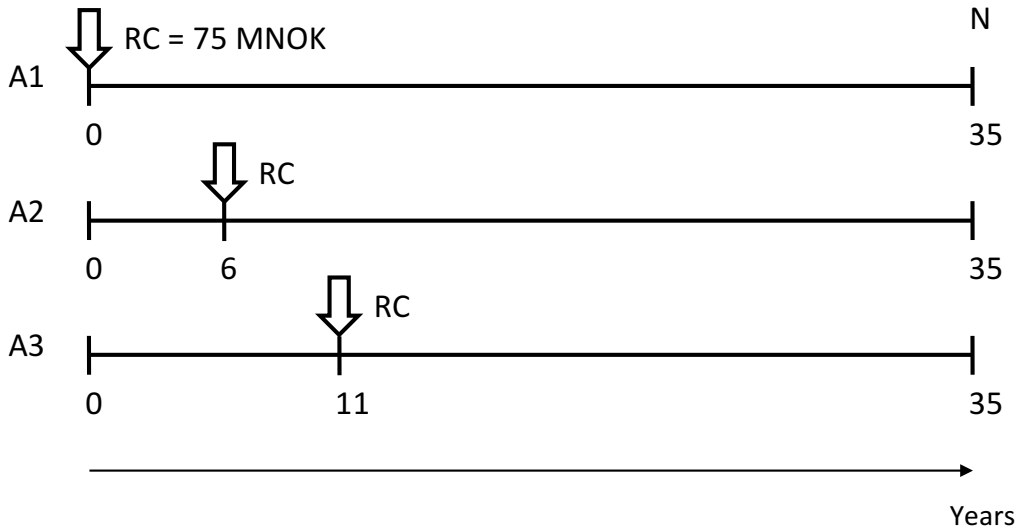

Figure 4.1: Illustration of alternatives for using machine learning to extend the lifetime of a transformer and further postpone reinvestment.

Furthermore, the economic savings are estimated at present value (PV), using Equation 4.1:

$$PV = RC \cdot \varepsilon_T \cdot \lambda_N \cdot \alpha_N \tag{4.1}$$

Here, ε_T is the annuity factor, λ_N is the capitalisation factor and α_N is the discounting factor.

The discount rate is set to 4%, also as used in Norges vassdrags- og energidirektorat (NVE) [60]. The different factors for each alternative are calculated as in [61]. The results from the analysis of economic savings from extending the lifetime and further postponing reinvestment of a power transformer using machine learning are shown in Table 4.1.

Table 4.1: Economic savings from extending the lifetime and further postponing reinvestment of a power transformer using machine learning.

	Without machine learning		With machine learning (5 years postponed)		With machine learning (10 years postponed)	
Costs	kr	75 000 000	kr	57 526 848	kr	42 764 436
Savings	kr	-	kr	17 473 152	kr	32 235 564

The results of the analysis show the potential for considerable savings. When extending the lifetime and further postponing reinvestment of the transformer 10 years, the savings are approximately 32 MNOK. If machine learning is used on several transformers, the savings will of course increase. However, the analysis is superficial and does not take into account many variables that may affect the final result. These include for example maintenance costs and that the probability of a fault is kept constant. The analysis still indicates a considerable potential for savings, since Statnett currently operates 135 power transformers of the type 400-453 kV, amongst many other types. The value of with certainty being able to postpone reinvestments based on machine learning is therefore big.

4.2 Loading Determination

Machine learning can also be used to determine how much a change of load will affect the hot-spot temperature in the transformer. The predicted hot-spot temperature can further be used as input in the winding insulation degradation model to determine the remaining lifetime of the transformer, based on the load change. Statnett can in this way evaluate the economy of whether to increase the load of the transformer as an alternative, or to evaluate other alternatives, for example, the purchase of a new transformer. This can be useful for both short-term (months) and long-term (several years) purposes. Statnett can determine how much they can load a transformer short-term in different operational situations, based on determining how much the load can be increased until the hot-spot temperature reaches the standard upper limit of 105°C. Examples of different operational situations are:

- For seasonal variations, especially cold periods.
- When there is a need for emergency loading. For example: due to a reconstruction of the power system or a redirection of the load.
- When there is a poor condition in the transformer.

For long-term evaluation, the winding insulation degradation model can be used to determine how much the remaining lifetime is affected by the increased load during its

lifetime. It can for example be more cost effective to reduce the remaining lifetime of one transformer than purchasing a new transformer.

4.3 Alternative to Loading Guide

The predicted hot-spot temperature obtained from the machine learning model made in this thesis corresponds well with the measured hot-spot temperature by a fiber optic sensor. Furthermore, the results show that the predicted hot-spot temperature actually corresponds better with the measured temperature than the temperatures calculated using the loading guide. The loading guide is a common model for determining the hot-spot temperature in a transformer. However, the loading guide model has significant uncertainty and constraints due to assumptions. The specific values needed for the calculations, usually obtained from a temperature rise test, are difficult to obtain and are uncertain, or in the worst case, do not exist at all. This thesis indicates that machine learning models can be an alternative to the loading guide, as they give better accuracy and will thus reduce uncertainty.

Chapter 5

Discussion

Power transformers have the potential for longer lifetimes. Long lifetimes can be safely utilised with improved asset management. The introduction of data analysis using machine learning and management of big data enables condition monitoring of power transformers to a greater extent than was previously possible. This will potentially lead to an improved estimation of condition and remaining lifetime, as well as improved decision-making processes for maintenance and reinvestment needs.

Major power transformer failures generally originate from the windings. Therefore, condition monitoring of the insulation system is important to secure the reliability of transformers. Monitoring of hot-spot temperature is a method used for this purpose.

5.1 Hot-spot Temperature Prediction

In the first case study of this thesis, machine learning models have been made to predict transformer hot-spot temperature, using the available data from Statnett. Several learning algorithms were applied to the chosen and prepared Data Set X, containing measurements from Transformer X, in order to find the algorithms that performed best on the data set. The made machine learning model, Model X, was further used to determine the maximum acceptable loading of the transformer. Finally, the efficiency of machine learning models was tested when they were applied to other data sets.

The results show good accuracy for predicted hot-spot temperature for several different learning algorithms being applied to Data Set X. The best results are obtained when the predictions are based on the line current (load), oil temperature and ambient temperature, with the hot-spot temperature measured by a fiber optic sensor as output, using the learning algorithm Linear Regression. In this case, Linear Regression gives a score for the coefficient of determination, R^2, of 0.896 for the testing subset. The hot-spot temperature was also predicted without the oil temperature, in order to determine the maximum acceptable

loading. Linear Regression then gives a score for R^2 of 0.813 for the testing subset. As the oil temperature correlates well with the hot-spot temperature, the accuracy naturally has decreased somewhat - approximately 10%. However, the accuracy is still satisfactory for the purpose of this thesis as the domain expertise was the focus. The made Model X based on Data Set X, using the learning algorithm Linear Regression and the line current (load) and ambient temperature as predictors, was further used to determine the maximum acceptable loading of Transformer X. The results indicate that the load of Transformer X could have been increased 250% before the hot-spot temperature reached the upper limit of 105°C in the testing subset. This makes it possible to determine how much the transformer can be loaded, for example in cold periods, or when there is a need for emergency loading due to a reconstruction of the power system or a redirection of the load. However, 250% is quite a large increase of load. It is possible that the model is not valid for such big changes, as it has not been trained properly on data for those changes. Other factors can begin to have an effect, for example more or different oil convection. New machine learning models were also made based on two other data sets, with satisfactory accuracy. Linear Regression gives scores for R^2 of 0.741 and 0.812 for the testing subsets of Data Set Y and Data Set Z, respectively. Model X was also applied to Data Set Y and Data Set X to see how the model would perform on other transformer data sets. The results show poor accuracy and indicate that a model made specifically for one transformer should not be applied to another transformer. An individual model should be made for each transformer.

Data of quality is a necessity while performing data science and applying machine learning. According to Statnett, the process for extracting the data sets from their data base was not straightforward. The data sets are also cumbersome to work with due to their format. Based on examinations during the case study, it seems that the data is not used to a large extent, resulting in potential errors or missing data not being discovered. Several data sets contain missing values, values being equal to zero, constant values, and outliers. Many measurements are physically unlikely for transformers, and indicate that the sensors are not measuring correctly. This may be due to the sensors failing or not being installed correctly. The data sets therefore had to be studied in detail and well prepared before machine learning could be applied. Most of the time spent while performing the case study was used on handling the data.

The results from the case study show that machine learning can be used to predict transformer hot-spot temperature development with good accuracy. However, power transformers are complex, and they also possess inertia. Many of their parameters are dependent on each other and correlate, and there are also time delays. This is a limitation for the accuracy of the machine learning models. Transformers are also constructed in different ways with dissimilar relationships between different parameters (such as temperature and load correlations). This makes the transformers individual. As a result, a machine learning model based on measurements from one transformer can not be applied to another

transformer. Individual models must be made for each transformer. This requires sufficient data from the transformer, in addition to data processing and preparation. However, for distribution transformers, standardisation can possibly be done and a general model made, as these transformers are much smaller and less complex than power transformers.

The case study has been performed on data sets containing measurements from relatively new transformers. This is because it is generally new transformers which have fiber optic sensors installed, to measure hot-spot temperature, which was necessary as output in the machine learning algorithms. This results in the data sets being relatively small. It would have been more desirable to use bigger data sets containing measurements for several years, to apply machine learning and to make models. This is because machine learning models which are trained longer on bigger data sets will support a finding to a greater extent than those using smaller data sets. The accuracy of the case study would have been improved, as the model would have learned several more output responses to the input during training. Testing over longer periods of time would also help validate the accuracy of the machine learning models. Finally, machine learning will give more utility value when transformers are old or in poor condition, as they then have a higher need for condition monitoring due to a higher probability of failure and thus the need for maintenance and reinvestments.

5.2 Winding Insulation Degradation Estimation

In the second case study of this thesis, the transformer winding insulation degradation has been estimated based on diversely obtained hot-spot temperatures. The machine learning model Model X, made in the first case study, was used to predict the hot-spot temperature for different scenarios, including increased load, and both increased and decreased ambient temperature. These hot-spot temperatures were further used to calculate the degree of polymerisation, DP, which again has been used to estimate the associated remaining lifetime. The results were compared with results obtained from measured hot-spot temperature by a fiber optic sensor. The predicted and measured hot-spot temperatures were also compared to calculated hot-spot temperatures obtained using the loading guide.

The results show that a higher hot-spot temperature leads to accelerated ageing, which again leads to a lower DP-value and shorter estimated remaining lifetime of the transformer. This corresponds with theory (Section 2.2). The difference in the hot-spot temperature does not have to be very large to have a serious impact on the estimated remaining lifetime. The DP-values of the measured and predicted hot-spot temperatures are almost equal, 991.4 and 992.1, respectively, which is desirable. This further leads to estimated remaining lifetimes of 84.2 years and 87.9 years, respectively. This is a difference of 4.4%. The predicted hot-spot temperature is lower than the measured hot-spot temperature when the hot-spot temperatures are at their highest, which results in the remaining lifetime based on predicted

hot-spot temperature being longer. This is not desirable as it can lead to false security. As the load and ambient temperature are increased, the predicted hot-spot temperatures increase as expected, and this results in lower DP-values and shorter remaining lifetimes. As the ambient temperature is decreased, the hot-spot temperatures decrease, resulting in higher DP-values and longer remaining lifetimes. The change of the load has however a bigger effect on the hot-spot temperature than a change of the ambient temperature. This corresponds with the correlation matrix shown in Figure 3.7, which shows that the load correlates better with the hot-spot temperature than the ambient temperature does. The shortest estimated remaining lifetime is obtained from the predicted hot-spot temperature when the load is increased by 250%. This was the maximum loading before the predicted hot-spot temperature reached the upper limit of 105°C. This results in a decrease in the DP-value of 25.1% which further leads to a decrease in the estimated remaining lifetime of 88.6%, compared to the results from the original predicted hot-spot temperature (without increased load). It is however likely that the results are not as accurate for such big increases of load as it is far outside the range of values where the machine learning model has been trained. The results also show that the predicted hot-spot temperature corresponds better with the measured temperature than the calculated ones based on ambient temperature and oil temperature, using the loading guide. The difference is 11.2% and 15.2% between the estimated remaining lifetime based on the hot-spot temperatures calculated from ambient temperature and oil temperature, and measured hot-spot temperature, respectively. This illustrates how much effect the method of obtaining the temperature can have on the estimate of the ageing of transformers. It is therefore very important that the temperatures are correct when estimating the remaining lifetime. Inaccurate temperature measurements, predictions, and calculations will lead to an incorrect basis for decisions concerning maintenance and reinvestment.

The purpose of this case study was to estimate how much different scenarios would affect the estimated remaining lifetime of the transformer. Predicted hot-spot temperatures obtained from a machine learning model were used. The results show that there is a small difference between the estimated remaining lifetime based on the predicted hot-spot temperature as against the remaining lifetime based on the measured temperature. It is therefore important to emphasise that these are estimations, and this should be taken into account when making decisions. More advanced learning algorithms could be used to improve the accuracy of the machine learning model, to make the predicted hot-spot temperature more equal the measured hot-spot temperature. However, it is unlikely that hot-spot temperature can be predicted with 100% accuracy. The winding degradation model also has uncertainties. It is used to give an estimate. Assumptions, including that the temperature and load history for one year are similar to the history of the transformer for subsequent years, and that $\overline{A}$ increases linearly, are limitations. These assumptions will however not change the fact that higher hot-spot temperature leads to shorter remaining lifetime, they will only affect the accuracy of the results. This is also important to consider

when increasing the load. The increased load may affect other parameters in the winding insulation degradation model than just the hot-spot temperature, for example $\overline{A}$, the average environment factor. It is therefore important for Statnett to consider these limitations and to emphasise that the results are estimations when making maintenance and reinvestment decisions. However, when it is possible to predict hot-spot temperature with good accuracy, the winding insulation degradation model should be improved, such that the estimations represent reality as close as possible. More research is necessary for this purpose.

The assumption of having only ONAN cooling is also a limitation, and may be a reason for the difference between the predicted hot-spot temperature and the ones calculated ones using the loading guide. In reality, the cooling switches between ONAN and ONAF cooling. However, the loading guide has significant uncertainty, and the specific values needed are also uncertain and difficult to obtain. The difference is therefore more likely to be due to the uncertainty of the loading guide than to the assumption of having only ONAN cooling.

This case study has modelled the ageing of the transformer component which has the highest failure rate - the insulating paper of the windings. However, the estimated remaining lifetimes will not completely represent reality as there are also other factors influencing it. Failures from for example tap-changers and bushings may also occur, but these failures can usually be repaired faster than failures originating from the windings. However, the estimated remaining lifetimes calculated in this thesis are the upper limits for the maximum potential. Ideally, it is achievable if satisfactory maintenance is performed on all the other components of the transformer. It is most likely that other components will need to be replaced during the transformer lifetime.

5.3 Utility Value

The utility value and benefits which emerged from the case studies were presented. Three uses of machine learning were proposed: as a normal behaviour model, for loading determination and as an alternative to the loading guide. In addition, one small economic analysis was performed.

The results show that there are considerable opportunities to be obtained by better management of big data and utilisation of machine learning for power transformer asset management. The results from the economic analysis also show the potential for big savings, approximately 32 MNOK by extending the lifetime and further postponing reinvestment of a transformer 10 years, using machine learning. This shows that the value of with certainty being able to postpone reinvestments based on machine learning is big. A good starting point is to incorporate machine learning into already existing activities and to learn from improvements. Existing activities can thus be performed more efficiently, with better accuracy, and less uncertainty. For example, using machine learning to predict

hot-spot temperature instead of calculating it using the loading guide. It is also interesting to consider how machine learning can be used for new applications - for example as a normal behaviour model or for loading determination.

However, it is important to take into account the limitations that are included. Some prerequisites must be fulfilled for machine learning to be implemented and to provide adequate and reliable results. Good ICT-structure and data of quality are necessary. Good quality data implies both high enough frequency of measurements, the desired parameters being measured and variations in data. There must exist data on the parameters which are to be predicted.

Chapter 6

Conclusion and Further Work

6.1 Conclusion

This thesis started by describing the method of asset management involving condition monitoring, measurements, estimation and instruments, used for decision making. Typical ageing mechanisms and faults of power transformers were described, in addition to relevant measuring methods and models used to detect ageing and/or faults. The concept and method of machine learning were also described, in addition to a relevant data mining process. Finally, case studies were performed and presented to demonstrate the applicability of machine learning in power transformer asset management. The utility value and benefits which emerging from the case studies have also been presented.

The purpose of this thesis was to examine potential opportunities obtained by better management of big data and utilisation of machine learning for power transformer asset management purposes. A machine learning model was developed to predict transformer hot-spot temperature with satisfactory accuracy, compared to the hot-spot temperature as measured by a fiber optic sensor. The maximum acceptable loading has also been determined using the machine learning model. The machine learning model has in addition been used to predict hot-spot temperatures for different scenarios, including increased load and increased and decreased ambient temperature. These hot-spot temperatures have further been used to estimate the condition and remaining lifetime of the transformer for the different scenarios, using the winding degradation model. Comparing the predicted hot-spot temperature with hot-spot temperatures obtained by the commonly used loading guide, indicates that the predicted hot-spot temperature corresponds better with the measured hot-spot temperature than the hot-spot temperatures obtained from the loading guide do.

In this thesis, it can be concluded that better management of big data and utilisation of machine learning for power transformer asset management creates many new

opportunities. It is possible to construct machine learning models to apply to data from transformer measurements to predict diverse parameters which are of great relevance to asset management. The major benefit of machine learning is its capability of utilising the information contained in historical data to investigate properties of future data. Firstly, it can be implemented in already existing activities to increase the efficiency and accuracy, and to reduce the uncertainty. Machine learning results, in this thesis, in predictions of greater accuracy than those in current use, giving better asset management possibilities, and potential for considerable economic gains. For example, hot-spot temperatures can be predicted more accurately by machine learning than by the current practice of using the loading guide. In addition, it can be used for new applications. For example, as a normal behaviour model or for loading determination. Optimised maintenance can thus be performed as needed, and a possible replacement can be postponed. Unnecessary resources spent on healthy transformers can also be reduced. This will lead to considerable economic savings. Creativity and curiosity can further lead to many other applications. As the power system is becoming more digital, machine learning has good opportunities to be implemented in several activities. With emerging use of sensors, machine learning can play an important role for sensor verification.

Power transformers are complex components in the power system. They are constructed in different ways with dissimilar relationships between different parameters. They are therefore individual, requiring an individual machine learning model to be made for every power transformer. In addition, decisions involving maintenance and reinvestments should not be based solely on results obtained from machine learning. As with many other methods and models, some degree of uncertainty is inherent in machine learning and has to be accepted. For example, unsuitable machine learning algorithms can be used, or the models can be insufficiently trained, or trained on poor data. When using predictions obtained from machine learning to estimate condition and remaining lifetime, it is important to emphasise that these are estimates. This needs to be considered when making decisions involving maintenance and reinvestments.

Good ICT-structure and data of quality are necessary for handling big data and applying machine learning. Machine learning models need a considerable amount of data as input, and the data used must be of quality. Based on the examinations in this thesis, it is recommended that Statnett monitors their data more frequently, and introduce a quality control regime. It would be a great advantage for Statnett to have a framework that could deliver the desired data sets more easily, in addition to having a built-in tool for data preparation and processing, where the generated data sets can be downloaded directly with the desired format. This would make it easier to quality check the data, and to utilise them. It is recommended that machine learning is initially used by Statnett as a support system, and as one of several tools used for decision making. As the machine learning models are used over longer periods of time, Statnett will gain more experience and the models will be

improved. Statnett will then be able to rely to a greater extent on the results obtained by the models. Experience-sharing with other power companies, machine learning experts, and research organisations is also recommended.

6.2 Recommendations for Further Work

Based on the discussion and conclusion, some considerations for further work and evaluation are:

- Application of machine learning on other components in the power system. Data from power transformers were used in this thesis because an increasing volume of data concerning their condition is becoming available. However, application of machine learning on data from other components will further illustrate potential opportunities to be obtained from machine learning.
- Creation of machine learning models to distinguish different emerging faults of power transformers. This thesis illustrates how machine learning models can be used for normal behaviour models. With more advanced use of machine learning, it is also possible to train the model to recognice and distinguish different faults of the transformer. Such models can possibly replace currently used protection devices.
- Investigation of which criteria Statnett uses for power transformer reinvestment. Reinvestments are economically challenging and should be based on several significant criteria. It can be a challenge to determine and specify good criteria.
- Evaluation if physical models, for example the loading guide and the winding insulation degradation model, are good enough for intended use. Being able to predict hot-spot temperature with good accuracy requires that the models used further are of good quality.
- A comprehensive analysis of potential savings obtainable by the use of machine learning. However, precise figures may be difficult to specify.
- Establishing a framework for collecting, processing and organising measurements and data. This can be helpful for Statnett, as the infrastructure around the data will be improved, making them more readily available for further use.
- The creation of a research project in Statnett that explores which divisions in Statnett can benefit from machine learning, and what needs to be done to implement different initiatives.

Bibliography

[1] A. Smisethjell, "Short description of the SAMBA project," 2016.

[2] *Statnett Nettutviklingsplan 2017*, Statnett, 2017. [Online]. Available: http://www.statnett.no/Global/Dokumenter/NUP%202017-endelig/Nettutviklingsplan%202017.pdf

[3] L. E. Lundgaard, K. B. Liland, D. Linhjell, D. Susa, M.-H. G. Ese, C. M. Selsbak, M. Kes, and C. M. Lesaint, "Transformer Windings. Ageing, diagnosis and asset management," SINTEF Energy, Tech. Rep. TR 17099, September 2015.

[4] "NIST Big Data Interoperability Framework: Volume 1, Definitions," *NIST Special Publication 1500-1*, September 2015. [Online]. Available: https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf

[5] M. Ukkelberg, "Inaccuracy of Technical Condition Values from Online Measurements of Power Transformers," Specialisation project, NTNU, December 2017.

[6] *Forvaltning av anlegg og verdier - Oversikt, prinsipper og terminologi*, NS-ISO 55000:2014 Std., 2014.

[7] J. Voskuil, "Did you mean ALM or ALM or ALM?" July 2012. [Online]. Available: https://virtualdutchman.com/2012/07/26/did-you-mean-alm-or-alm-or-alm/

[8] E. Solvang, "Risk based maintenance and reinvestment," *Lecture slide in the course ELK30 at NTNU*, 2016.

[9] *Vedlikehold - Vedlikeholdsterminologi*, NS-EN 13306:2017 Std., 2018.

[10] A. Majzoobi, M. Mahoor, and A. Khodaei, "Machine Learning Applications in Estimating Transformer Loss of Life," in *2017 IEEE Power Energy Society General Meeting*, July 2017, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8274564

[11] X. L. Dong and D. Srivastava, "Big Data Integration," in *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, April 2013, pp. 1245–1248. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6544914

[12] M. Hilbert, "Big Data for Development: A Review of Promises and Challenges," *Development Policy Review*, vol. 34, no. 1, pp. 135–174, 2016. [Online]. Available: http://dx.doi.org/10.1111/dpr.12142

[13] *Informasjonsteknologi og elektroteknikk - Det digitale energiskiftet.* NEF Teknisk Møte, SINTEF Energi AS, April 2017.

[14] *ABB Digital Substations*, ABB, July 2016.

[15] *Risk based maintenance and consequence classification*, NORSOK Z-008:2017 Std., 2017.

[16] T. K. S. Karlsen, "Trender og konsepter innen vedlikehold," Master's thesis, NTNU, March 2017.

[17] J. G. Webster and H. Eren, Eds., *Measurement, Instrumentation and Sensors Handbook*, 2nd ed. CRC Press, 2014.

[18] C. L. Enloe, E. Garnett, J. Miles, and S. Swanson, *Physical Science: What the Technology Proffesional Needs to Know.* John Wiley & Sons, Inc., 2001.

[19] M. F. F. Domingues and A. Radwan, *Optical Fiber Sensors for IoT and Smart Devices*, 1st ed. Springer, 2017.

[20] *What are Fiber Optic Sensors?*, AZoSensors, September 2014. [Online]. Available: https://www.azosensors.com/article.aspx?ArticleID=333

[21] W. D. A. G. Hillary, K. L. I. M. P. B. Jayarathna, L. I. Ranasinghe, S. M. B. P. Samarakoon, N. M. T. N. Rathnayake, J. R. Lucas, and R. Samarasinghe, "A tool for estimating remaining life time of a power transformer," in *2017 Moratuwa Engineering Research Conference (MERCon)*, May 2017, pp. 373–378.

[22] P. Dvorak, "Report update: Large Power Transformers and the U.S. Electrical Grid," *Windpower Engineering & Development*, June 2014. [Online]. Available: http://www.windpowerengineering.com/electrical/report-update-large-power-transformers-u-s-electric-grid/

[23] A. J. Pansini, *Electrical transformers and power equipment*, 3rd ed. The Fairmont Press, Inc., 1999.

[24] B. Nyberg, "Tank og olje," *Håndbok i Vedlikehold av Krafttransformator*, April 2006.

[25] "Transformer Reliability Survey," CIGRÉ, Tech. Rep. WG A2.37, 2015.

[26] T. Lundquist and W. Knuth, "Finally Fiber Optic Sensing of Power Transformer "Hotspots" Is Paying Off," in *2008 International Conference on Condition Monitoring and Diagnosis*, Beijing, April 2008, pp. 819–824. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4580410&isnumber=4580203

[27] R. F. Tillman, *The Electric Power Engineering Handbook*. CRC Press, 2001, ch. Loading Power Transformers.

[28] *Power transformers - Part 2: Temperature rise for liquid-immersed transformers*, IEC 60076-2:2011 Std., 2011.

[29] *AKM 345 Gen2 Instruction Manual*, Qualitrol.

[30] *Power transformers - Part 7: Loading guide for mineral-oil-immersed power transformers*, IEC 60076-7:2018 Std., 2018.

[31] J. Foros, "Trafotiltak - Analyse av behov for diagnose, vedlikehold eller utskifting av krafttransformatorer," SINTEF Energy, Tech. Rep. 502000704, March 2017.

[32] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal*, vol. 3, no. 3, July 1959. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.2254&rep=rep1&type=pdf

[33] *Machine Learning*, ITERON, 2013. [Online]. Available: http://www.iteron.ch/machinelearning/

[34] P. Louridas and C. Ebert, "Machine Learning," *IEEE Software*, vol. 33, no. 5, pp. 110–115, September 2016. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7548905

[35] J. Gantz and D. Reinsel, "THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," *IDC Digital Universe*, December 2012. [Online]. Available: https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

[36] C. de Looper, "What is machine learning and how is it used today?" *PMECH*, July 2017. [Online]. Available: https://www.pcmech.com/article/machine-learning-used-today/

[37] S. H. Somanchi, "The mail you want, not the spam you don't," *Official Gmail Blog*, July 2015. [Online]. Available: https://gmail.googleblog.com/2015/07/the-mail-you-want-not-spam-you-dont.html

[38] M. Sprenger, "Vindkraft-produsenten sparer millioner på å forutsi været mer nøyaktig enn metrologene," *Teknisk Ukeblad*, April 2017. [Online]. Available: https://www.tu.no/artikler/vindkraft-%20produsenten-sparer-millioner-pa-a-forutsi-vaeret-mer-noyaktig-enn-meteorologene/379833

[39] D. GahuThakurta, R. B. Severtson, and G. Ericson, "Tasks to prepare data for enhanced machine learning," *Microsoft Azure*, September 2017.

[Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/prepare-data#references

[40] *Machine Learning in MATLAB*, MathWorks, 2018. [Online]. Available: https://se.mathworks.com/help/stats/machine-learning-in-matlab.html

[41] S. Mishra, "Unsupervised Learning and Data Clustering," *Towards Data Science*, May 2017. [Online]. Available: https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a

[42] F. S. Dietrichson, H. L. Steen, J. W. Friis, J. Hole, K. K. Krossholm, and S. Måseidvåg, "Maskinlæring," *Kube2017, Statnett*, august 2017.

[43] A. Bronshtein, "Train/Test Split and Cross Validation in Python," *Towards Data Science*, May 2017. [Online]. Available: https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6

[44] D. Geng and S. Shih, "Machine Learning Crash Course: Part 4: The Bias-Variance Dilemma," *Machine Learning @ Berkeley*, July 2017. [Online]. Available: https://ml.berkeley.edu/blog/2017/07/13/tutorial-4/

[45] J. Brownlee, "Overfitting and Underfitting With Machine Learning Algorithm," *Machine Learning Algorithms*, March 2016. [Online]. Available: https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/

[46] M. Deshpande, "A Guide to Improving Deep Learning's Performance," *Zenva*, October 2017. [Online]. Available: https://pythonmachinelearning.pro/a-guide-to-improving-deep-learnings-performance/

[47] J. Brownlee, "An introduction to Feature Selection," *Machine Learning Process*, October 2014. [Online]. Available: https://machinelearningmastery.com/an-introduction-to-feature-selection/

[48] P. Gupta, "Regularization in Machine Learning," *Towards Data Science*, November 2017. [Online]. Available: https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a

[49] J. Brownlee, "Metrics to Evaluate Machine Learning Algorithms in Python," *Python Machine Learning*, May 2016. [Online]. Available: https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/

[50] *Jupyter*, Project Jupyter, February 2018. [Online]. Available: http://jupyter.org/index.html

[51] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing*, vol. 5, no. 4, 2000. [Online]. Available: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf

[52] C. Clifton, "Data mining," *Britannia Academic*, November 2009. [Online]. Available: http://academic.eb.com/levels/collegiate/article/data-mining/437561

[53] G. Piatetsky, "CRISP-DM, still the top methodology for analytics, data mining, or data science projects," *KDnuggets*, October 2014. [Online]. Available: https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html

[54] *Data Mining Phases*, Smart Vision Europe, 2015. [Online]. Available: http://crisp-dm.eu/reference-model/

[55] *Choosing the right estimator*, scikit-learn, 2017. [Online]. Available: http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

[56] S. Ray, "7 Types of Regression Techniques you should know!" *Business Analytics*, August 2015. [Online]. Available: https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/

[57] *1.1 Generalized Linear Models*, scikit-learn, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/linear_model.html#ridge-regression

[58] D. Basak, S. Pal, and D. C. Patranabis, "Support Vector Regression," *Neural Information Processing - Letters nd Reviews*, vol. 11, no. 10, October 2017. [Online]. Available: https://pdfs.semanticscholar.org/c5a9/67eaded74a9fc414de4ad5120b0b66acd2c3.pdf

[59] *1.11. Ensemble methods*, scikit-learn, 2017. [Online]. Available: http://scikit-learn.org/stable/modules/ensemble.html#

[60] O.-P. Kordahl, H. M. Kvile, R. Amundsveen, and T. Langset, "Forslag til nytt vektsystem i modellen for å fastsette konstnadsnormer i regionalnettene," *NVE*, no. 5, September 2014. [Online]. Available: http://publikasjoner.nve.no/hoeringsdokument/2014/hoeringsdokument2014_05.pdf

[61] E. Solvang, "Nåverdier og kostnadselementer (investeringer, tap og avbrudd)," *Lecture slide in the course ELK11 at NTNU*, 2017.

Appendix A

Python Code

Listing A.1 shows the main Python code used for this thesis. The listing shows the code specifically used on Data Set X, but the code was also used on Data Set Y and Data Set Z, with small alterations:

- Line 29: change of filename
- Line 42, 50, 62-65 and 107-108: change of sensor names
- Line 111: change of *test_size*

In addition, the SINTEF Energy Research code used to download ambient temperature from met.com was altered to get the ambient temperature from the area of the transformers.

Listing A.1: Python code.

```
1  #Load libraries
2  %matplotlib inline
3  import os
4  import py.path
5  import glob
6  import numpy as np
7  import pandas as pd
8  import seaborn as sns
9  import json
10 from datetime import datetime
11 from urllib.parse import urlencode
12 from http.client import HTTPSConnection
13 from base64 import b64encode
14 import matplotlib.pyplot as plt
15 plt.style.use('ggplot')
16 from pandas.plotting import scatter_matrix
```

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso
from sklearn.linear_model import ElasticNet
from sklearn.svm import SVR
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import AdaBoostRegressor

#Load HDF using pandas
datapath = os.path.join(os.path.expanduser('~'), 'Documents', 'datasetx.h5')
data = pd.read_hdf(datapath)

#Plot data
data.plot(legend=False)

#Shape
print(data.shape)

#Rows and columns
data.head()

#Changing the sensor names
names = ['Oil temperature hot-spot', 'Line current L1 420', 'Line current L2
  ↪ 420', 'Tap position', 'Oil temperature', 'Hydran gas', 'Hydran moisture',
  ↪ 'Winding temperature 420', 'Reactive power 420', 'Winding temperature 132'
  ↪ , 'Line current L3 420', 'Winding temperature hot-spot 132', 'Winding
  ↪ temperature hot-spot 420']
data.columns = names

#Describe all sensors
data.describe()

#Choosing relevant sensors
#Oil temperature both included and excluded
data= data[['Line current L2 420', 'Oil temperature', 'Winding temperature hot
  ↪ -spot 420']]
data.head()

#Checking if any NaN
data1 = data[data.isnull().any(axis=1)]
```

```
print (data1)

#Fill NaN with mean of nearest neighbours
data=((data.fillna(method='ffill') + data.fillna(method='bfill'))/2)

#Data cleaning: fill outliers with mean of nearest neighbours
#Oil temperature both included and excluded
data[(data['Oil temperature'] < 0)] = np.nan
data[(data['Oil temperature'] > 70)] = np.nan
data[(data['Winding temperature hot-spot 420'] < 0)] = np.nan
data[(data['Winding temperature hot-spot 420'] > 100)] = np.nan
data = ((data.fillna(method='ffill') + data.fillna(method='bfill'))/2)

#Changing sampling rate from minutes to hours
data = data.resample('60T').mean()

#Plot cleaned data
ax=data.plot()
ax.set_ylabel('Current A / Temperature °C')

#Check when sampling starts/ends
data.head()
data.tail()

#New shape
print(data.shape)

#Using SINTEF Energy Research code to download ambient temperature from met.
  ↪ com

#Change format of the index and name of ambient temperature
ambienttemp.index.tz = None
names = ['Ambient temperature']
ambienttemp.columns = names
ambienttemp.head()

#Adding ambient temperature
data = pd.concat([data, ambienttemp], axis=1, join_axes=[data.index])
data.head()

#Describe all sensors
data.describe()

```

```
#Plotdata
ax=data.plot()
ax.set_ylabel('Ampere / Temperature °C')

#Correlation matrix
f, ax = plt.subplots(figsize=(10, 8))
corr = data.corr()
sns.heatmap(corr, annot=True, mask=np.zeros_like(corr, dtype=np.bool), cmap=
  ↪ sns.diverging_palette(220, 10, as_cmap=True), square=True, ax=ax)

#Split data set into predictor and target values
x = data.drop(['Winding temperature hot-spot 420'], axis=1) #Predictors
y = data['Winding temperature hot-spot 420'] #Target

#Split data set into training and test: 50% test
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.5,
  ↪ shuffle=False)

#Linear Regression
lin = LinearRegression()
lin.fit(x_train, y_train)
lin_pred = lin.predict(x_test)
print("R^2 train (score): {}".format(lin.score(x_train, y_train)))
print("R^2 test (score): {}".format(lin.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, lin_pred))
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, lin_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
lin_pred=pd.Series(lin_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, lin_pred], axis=1)
ax=result.plot()
ax.set_ylabel('Temperature °C')
differance = [lin_pred-y_test]
np.mean(differance)

#Ridge Regression
ridge = Ridge(alpha=0.0000001, normalize=True)
ridge.fit(x_train, y_train)
ridge_pred = ridge.predict(x_test)
print("R^2 train (score): {}".format(ridge.score(x_train, y_train)))
print("R^2 test (score): {}".format(ridge.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, ridge_pred))
```

```
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, ridge_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
ridge_pred=pd.Series(ridge_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, ridge_pred], axis=1)
ax=result.plot()
ax.set_ylabel('Temperature °C')
differance = [ridge_pred-y_test]
np.mean(differance)

#Lasso Regression
lasso = Lasso(alpha=0.0000001, normalize = True)
lasso.fit(x_train, y_train)
lasso_pred = lasso.predict(x_test)
print("R^2 train (score): {}".format(lasso.score(x_train, y_train)))
print("R^2 test (score): {}".format(lasso.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, lasso_pred))
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, lasso_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
lasso_pred=pd.Series(lasso_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, lasso_pred], axis=1)
ax=result.plot()
ax.set_ylabel('Temperature °C')
differance = [lasso_pred-y_test]
np.mean(differance)

#ElasticNet Regression
en = ElasticNet(alpha=0.0000001, l1_ratio=0.5, normalize=True)
en.fit(x_train, y_train)
en_pred = en.predict(x_test)
print("R^2 train (score): {}".format(en.score(x_train, y_train)))
print("R^2 test (score): {}".format(en.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, en_pred))
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, en_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
en_pred=pd.Series(en_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, en_pred], axis=1)
ax=result.plot()
```

```
ax.set_ylabel('Temperature °C')
differance = [en_pred-y_test]
np.mean(differance)

#SVR linear
svr = SVR(kernel='linear')
svr.fit(x_train, y_train)
svr_pred = svr.predict(x_test)
print("R^2 train (score): {}".format(svr.score(x_train, y_train)))
print("R^2 test (score): {}".format(svr.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, svr_pred))
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, svr_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
svr_pred=pd.Series(svr_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, svr_pred], axis=1)
ax=result.plot()
ax.set_ylabel('Temperature °C')
differance = [svr_pred-y_test]
np.mean(differance)

#SVR rbf
svr = SVR(kernel='rbf')
svr.fit(x_train, y_train)
svr_pred = svr.predict(x_test)
print("R^2 train (score): {}".format(svr.score(x_train, y_train)))
print("R^2 test (score): {}".format(svr.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, svr_pred))
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, svr_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
svr_pred=pd.Series(svr_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, svr_pred], axis=1)
ax=result.plot()
ax.set_ylabel('Temperature °C')
differance = [svr_pred-y_test]
np.mean(differance)

#Random Forest Regressor
rf = RandomForestRegressor(random_state=0)
rf.fit(x_train, y_train)
```

```
rf_pred = rf.predict(x_test)
print("R^2 train (score): {}".format(rf.score(x_train, y_train)))
print("R^2 test (score): {}".format(rf.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, rf_pred))
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, rf_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
rf_pred=pd.Series(rf_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, rf_pred], axis=1)
ax=result.plot()
ax.set_ylabel('Temperature °C')
differance = [rf_pred-y_test]
np.mean(differance)

#AdaBoost Regressor
ada = AdaBoostRegressor(random_state=0)
ada.fit(x_train, y_train)
ada_pred = ada.predict(x_test)
print("R^2 train (score): {}".format(ada.score(x_train, y_train)))
print("R^2 test (score): {}".format(ada.score(x_test, y_test)))
rmse = np.sqrt(mean_squared_error(y_test, ada_pred))
print("Root Mean Squared Error (RMSE): {}".format(rmse))
mae = mean_absolute_error(y_test, ada_pred)
print("Mean Absolute Error (MAE): {}".format(mae))
#Plot measured and predicted
ada_pred=pd.Series(ada_pred, name='Predicted value', index= y_test.index)
result = pd.concat([y_test, ada_pred], axis=1)
ax=result.plot()
ax.set_ylabel('Temperature °C')
differance = [ada_pred-y_test]
np.mean(differance)
```

Appendix B

Algorithm Results

The following figures shows the plots of predicted values and measured values of hot-spot temperature for Data Set X using different regression algorithms, with a test size of 50%.

Ridge Regression

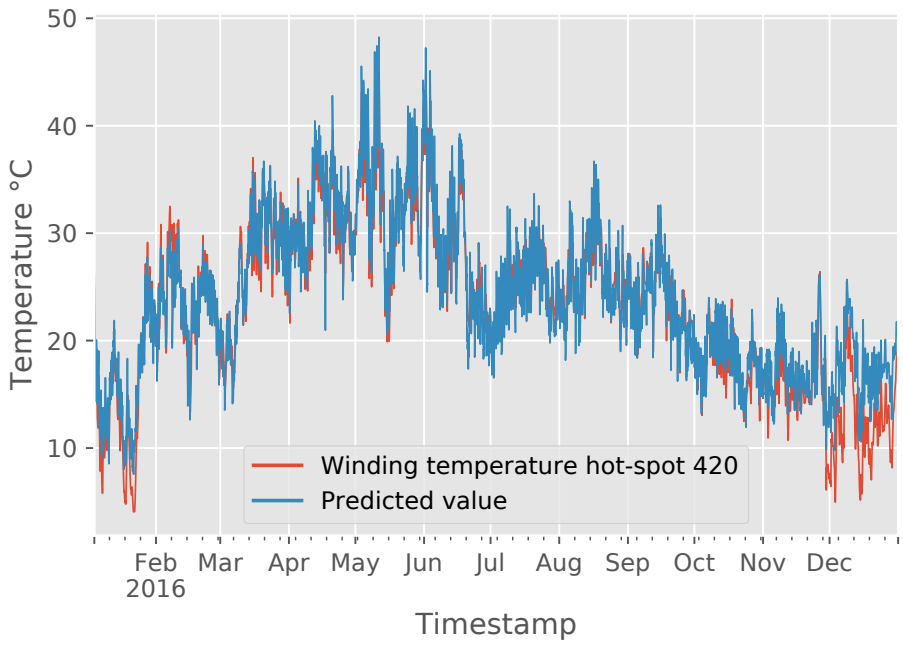

Figure B.1: Plots of predicted values and measured values of hot-spot temperature for Data Set X using Ridge Regression with 50% test size.

Lasso Regression

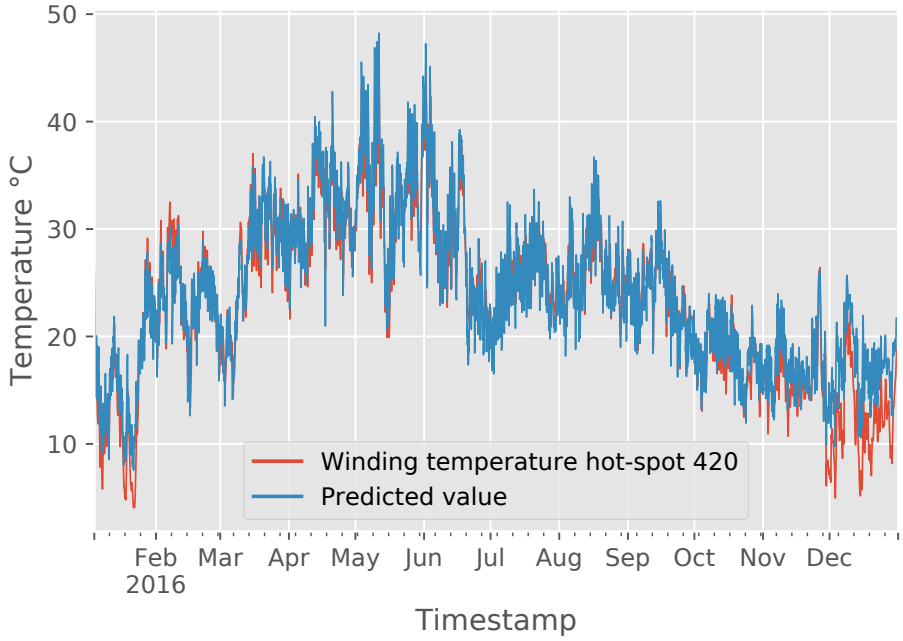

Figure B.2: Plots of predicted values and measured values of hot-spot temperature for Data Set X using Lasso Regression with 50% test size.

ElasticNet Regression

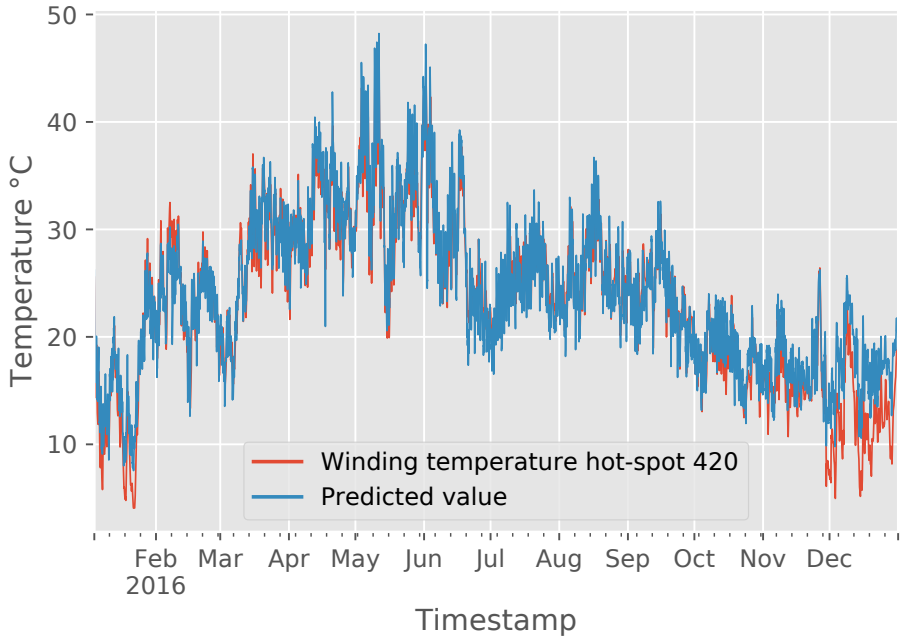

Figure B.3: Plots of predicted values and measured values of hot-spot temperature for Data Set X using ElasticNet Regression with 50% test size.

SVR (kernel='linear')

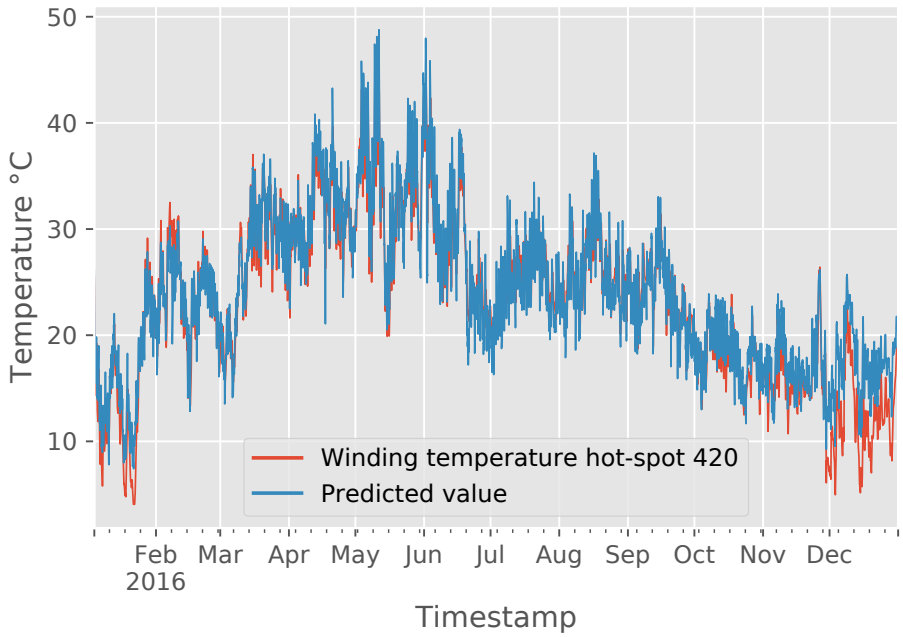

Figure B.4: Plots of predicted values and measured values of hot-spot temperature for Data Set X using SVR (kernel='linear') with 50% test size.

SVR (kernel='rbf')

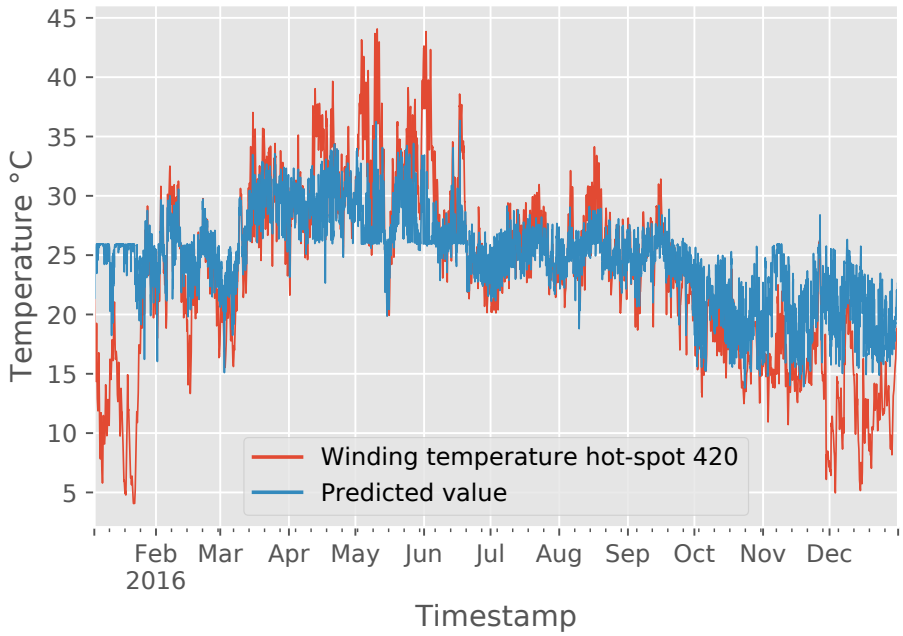

Figure B.5: Plots of predicted values and measured values of hot-spot temperature for Data Set X using SVR (kernel='rbf') with 50% test size.

Random Forest Regressor

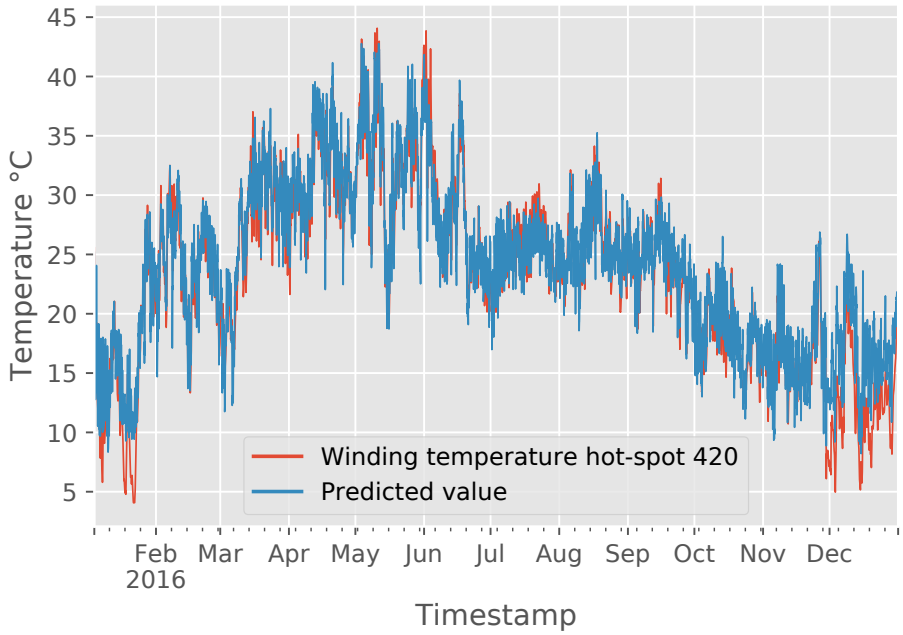

Figure B.6: Plots of predicted values and measured values of hot-spot temperature for Data Set X using Random Forest Regressor with 50% test size.

AdaBoost Regressor

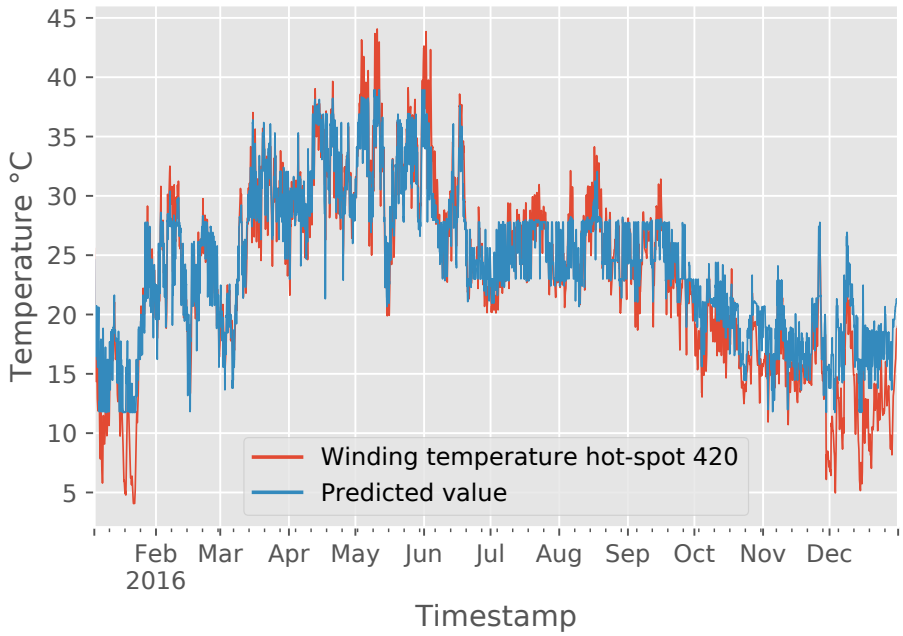

Figure B.7: Plots of predicted values and measured values of hot-spot temperature for Data Set X using AdaBoost Regressor with 50% test size.