Use of Learning Mechanisms to Improve Wind Farms Operation Conditions Monitoring

Ana Rita Sabudin Nunes

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisors: Prof. José Alberto Rodrigues Pereira Sardinha
Prof. Hugo Gabriel Valente Morais

Examination Committee

Chairperson: Prof. Daniel Jorge Viegas Gonçalves
Supervisor: Prof. José Alberto Rodrigues Pereira Sardinha
Member of the Committee: Dr. Vanessa Gonçalves Guedes

October 2021

Acknowledgments

I would like to thank all my family for the unconditional support. Specially to my mother for all the encouragement and support in pursuing my academic life and building my future and to my brother for being a role model.

My gratitude for all the guidance, insight and sharing of knowledge to my dissertation supervisors Prof. Hugo Morais and Prof. Alberto Sardinha.

Last but not least, to all my friends and to my partner that provided me an escape in the most stressful times during this last year.

Without all of you this project would not be possible.

To each and every one of you – Thank you.

Abstract

The contribution of this thesis is to take advantage of state-of-the-art machine learning techniques and apply them in wind energy problems, boosting wind turbines availability. To accomplish this, the work was focused on improving the existing techniques in predictive maintenance, that consist of acting before a major failure occurs using condition monitoring techniques. In particular, six regression models that learned from data belonging to normal periods of operation were developed to determine regular operating performance of wind turbines. One model for the normal behaviour of power generation, and five for specific turbine components, namely, the hydraulic group, generator, generator bearing, gearbox and power transformer. These models are used to detect deviations in predictions during the antecedent days of a failure, and consequently, predict its occurrence. The implementation of this thesis comprised the use of techniques to tackle the challenges of each of the stages of machine learning, such as data pre-processing, feature engineering and the selection of the best-suited model.

By the use of specific evaluation metrics, the results showed an improvement in the early prediction of a future fault with an average of seven days' notice and a minimum of one day. Additionally, a second improvement was on the accuracy of the normal behaviour models, substantiating their predictions. These improvements allow an alert to be given to the wind farm operator that through maintenance is able to prevent or decrease the fault occurrence repercussions. Consequently, having an increase in the availability of the turbine and therefore, in energy production.

Keywords

Condition Monitoring ; Fault Prediction ; Machine Learning ; Wind Turbine

Resumo

As metodologias propostas nesta tese tiram proveito do estado da arte de técnicas de aprendizagem automática e aplicam-nas a problemas de energia eólica, aumentando a disponibilidade dos aerogeradores. Assim sendo, a metodologia desenvolvida consiste numa ferramenta de suporte à manutenção preditiva, tendo como base a previsão das condições operacionais, ou seja, identificar alterações no modo de operação normal que sejam indicativas de uma futura falha. Mais concretamente, foram desenvolvidos seis modelos de regressão para determinar o funcionamento normal, treinados com dados pertencentes a períodos de operação sem falhas da turbina. Destes modelos, um permite determinar o comportamento normal de produção, e cinco determinam o funcionamento normal de componentes específicos da turbina, nomeadamente, grupo hidráulico, gerador, rolamento do gerador, caixa de velocidade e transformador de potência. Estes modelos são usados para detectar desvios nas previsões dos dias antecedentes à falha, e consequentemente, prevendo a sua ocorrência. A principal contribuição desta tese é o desenvolvimento de técnicas para superar os desafios de cada etapa da aprendizagem automática, i.e., pre-processamento de dados, seleção de variáveis e do modelo mais adequado.

Os resultados demonstraram uma melhoria na previsão de falhas com uma média de sete dias de antecedência e um mínimo de um dia. Uma segunda contribuição desta tese foi o aumento da precisão dos modelos de funcionamento normal, substanciando as suas previsões. Estas contribuições permitem a emissão de alarmes para o operador do parque eólico que através de manutenção poderá impedir ou diminuir as repercussões da ocorrência da falha. Tal terá como consequência o aumento da disponibilidade das turbinas eólicas e por sua vez da produção de energia.

Palavras Chave

Aprendizagem Automática ; Monitorização de Condições; Parque Eólico ; Previsão de Falhas

Contents

1 Introduction **1**

1.1 Motivation 3

1.2 Contributions 4

1.3 Organization of the Document 6

2 Background **7**

2.1 Wind Turbine 9

2.2 Condition Monitoring of Wind Turbines 10

2.3 State-Of-The-Art Machine Learning 12

3 Related Work **15**

3.1 Data Pre-Processing 17

3.2 Feature Selection and Extraction 18

3.3 Models 19

3.3.1 Turbine performance assessment 20

3.3.2 Power curve monitoring 20

3.3.3 Wind turbine conditions monitoring 22

3.3.4 Fault detection, diagnosis, and prediction of generator faults 24

3.3.5 Pitch System monitoring 25

3.3.6 Generator temperature monitoring 26

3.3.7 Generator bearing failure prediction 28

3.3.8 Gearbox monitoring 29

3.3.9 Transformer monitoring 30

4 Methodology **31**

4.1 Description of Data And Faults 33

4.1.1 Operational Data 34

4.1.2 Status Data 34

4.1.3 Faults 35

4.2 Data Acquisition and Pre-processing 35

4.2.1 Dealing with NaNs 36
4.2.2 Missing Data 36
4.2.3 Turbine Performance Assessment 37
4.2.4 Removal of Outliers 38
4.2.5 Data Labelling 39
4.3 Feature Selection 39
4.4 Model Selection 40
4.5 Validation 41
5 Results and Discussion **43**
5.1 Data Acquisition and Pre-processing 45
5.1.1 Missing Data 45
5.1.2 Removal of non-production periods 46
5.1.3 Turbine Performance Assessment 47
5.1.4 Removal of Outliers 49
5.1.5 Data Labelling 52
5.2 Feature Selection 52
5.2.1 Normal Behaviour of Production 52
5.2.2 Normal Behaviour of Turbine Components 53
5.3 Model Selection and Case Studies 56
5.3.1 Thresholds Assessment 56
5.3.2 Normal Behaviour of Production 58
5.3.3 Normal Behaviour of Hydraulic Group 60
5.3.4 Normal Behaviour of Generator Temperature 62
5.3.5 Normal Behaviour of Generator Bearing Temperature 64
5.3.6 Normal Behaviour of Gearbox 67
5.3.7 Normal Behaviour of Transformer 68
5.3.8 Concluding Remarks of Results 69
6 Conclusion **71**
6.1 Conclusions 73
6.2 Limitations and Future Work 74
Bibliography **75**
A Code of Project **81**

List of Figures

1.1 Summary of Thesis Implementation 5

2.1 Wind Turbine Components [1] 9

4.1 Machine Learning Phases 33
4.2 Dataset Partitions 36
4.3 Power Curve Turbine One (raw data) 38

5.1 All Turbines Raw Data 48
5.2 Outlier Removal - Normal behaviour of production model of the wind farm 51
5.3 Polynomial model - All turbines data 58
5.4 Turbine eleven pitch system faults 61
5.5 Turbine six - generator stator winding temperature phase two 63
5.6 Fault prediction of the second fault of turbine six 63
5.7 Turbine seven - generator bearing first and second fault 65
5.8 Turbine nine - generator bearing fourth fault 66
5.9 Transformer Faults - Turbine one and Turbine seven (second occurrence) 69

List of Tables

2.1 Most common types of failures with their respective problems and current condition monitoring techniques [2] 11

4.1 Signals file for the EDP dataset: five first entries 34
4.2 Failure logs file for the EDP dataset: five first entries 34
4.3 Number of failures by component and turbine 35

5.1 Methods for Missing data, estimated value, classification and model's errors 45
5.2 Results of removal of non-production periods by component and turbine 46
5.3 Number of rows: (1) initially, (2) with production under zero kW, (3) ten days before and after each failure, (4) after the removal of failures rows, and (5) previous column plus the removal of non-production periods 47
5.4 Mean and Standard Deviation of Production [kW] for the data before removals (1), the failure periods (3) and after removals (5) 48
5.5 Methods for Outlier Removal - Normal Behaviour of Production 50
5.6 Results of outlier removal by turbine component 52
5.7 Model's MSE [kW] for each Feature Selection Algorithm 53
5.8 Feature Selection - First Approach 54
5.9 Feature Selection by Turbine Component 55
5.10 Threshold Approach and Constant 57
5.11 Summary of results 70

Listings

A.1 Code for dealing with missing data 81

Acronyms

AI	Artificial Intelligence
ANFIS	Adaptive Neuro Fuzzy Inference System
CNN	Convolutional Neural Network
CM	Condition Monitoring
CCFL	Cluster Center Fuzzy Logic
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
EDP	Energias De Portugal
EWMA	Exponentially Weighted Moving Average
FiT	Feed-in Tariff
GBRT	Gradient-Boosting Regression Trees
GMM	Gaussian Mixture Model
GMR	Generalized Mapping Regressor
GPR	Gaussian Process Regression
GRNN	General Regression Neural Network
GW	GigaWatt
K-NN	K-Nearest Neighbor
kW	kiloWatt
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short-Term Memory

MD	Mahalanobis Distance
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MI	Mutual Information
MSE	Mean Square Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MW	MegaWatt
MWh	MegaWatt hour
NSET	Non-linear State Estimation Technique
NN	Neural Network
OM	Operation and Maintenance
PARAFAC	PARAllel FACtor
PCA	Principal Component Analysis
RF	Random Forest
RMSE	Root Mean Square Error
RMSPE	Root Mean Squared Percent Error
RNN	Recurrent Neural Network
RPM	Revolutions Per Minute
SCADA	Supervisory Control And Data Acquisition
SFS	Sequential Forward Selection
SOM	Self-Organizing Map
SVM	Support Vector Machine
SVR	Support Vector Regression
WT	Wind Turbine

1 Introduction

Contents

1.1 Motivation 3
1.2 Contributions 4
1.3 Organization of the Document 6

1.1 Motivation

As a consequence of an increasing climate change awareness, research on subjects such as renewable energy are of extreme importance. An event that played a big role in enhancing the need to rapidly find practical solutions to stop those changes was the Paris Agreement. The long term goal of that agreement was to limit temperature rises, therefore, countries needed to adapt their current practices to reduce carbon emissions [1].

Since the power sector is one of the main contributors to global greenhouse gas emissions [2], new technologies for cleaner energy production have been developed to replace conventional production (e.g., production based on fossil fuels). In order to create incentives for stakeholders to invest in cleaner energy technologies, countries (e.g., China, Russia, India, Japan and Brazil) have proposed national actions plans. By analysing those national plans, it can be observed that some of them highlighted that wind energy can have a major contribution, as a zero-emission energy source[3].

Following this, the Portuguese government came up with an initiative called "Roadmap for Carbon Neutrality 2050 (RNC2050)" [4]. The main objective of the RNC2050 is to identify and analyze energy source solutions and possible alternative trajectories to consumption. Two examples of those decarbonisation solutions are: eliminate coal-based power generation by 2030, incorporating endogenous renewable energy sources into final energy consumption, and promote electrification of end-uses. The elimination of coal-based power has already taken place by the shut down of Sines coal plant, leaving Portugal with only one remaining coal power station in operation, which is also already scheduled for closure. The new alternatives should be feasible and economically viable, allowing the Portuguese economy to reach carbon neutrality by 2050. To this end, it is essential to create an energy system that is safe, affordable and compatible with environmental protection. On that plan, the Portuguese government states that *"the path towards carbon neutrality will lead to a much wider use of endogenous renewable energy resources of which over two-thirds are sun and wind, accounting for over 80% of primary energy consumption by 2050"*. In addition, the production based on renewable sources will drastically decrease the dependency on the energy sources of foreign countries, creating new opportunities in the Portuguese economy.

As referred in the previous initiative, the future of energy will profoundly depend on renewable energies, such as wind and photovoltaic energy. This thesis focused on wind technology due to its increasing importance in the last few years. The cumulative installed capacity of both off and on-shore wind energy has been growing, with global yearly new installations rounding 50 GigaWatt (GW) from 2016 to 2019. In 2020, it even surpassed 90 GW, a 53% growth compared to 2019, bringing the total installed capacity to

[1] https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement
[2] https://www.c2es.org/content/international-emissions
[3] Only considering wind power plants lifetime emission
[4] https://unfccc.int/sites/default/files/resource/RNC2050_EN_PT%20Long%20Term %20Strategy.pdf

743 GW [5]. According to [3], the global installed capacity of on-shore wind power would increase (taking 2018 as reference) three-fold by 2030 (to 1 787 GW) and ten-fold by 2050 (to 5 044 GW). Concerning off-shore installations, the predictions are similar, with the global off-shore wind capacity rising 228 GW in 2030 and near 1 000 GW in 2050, comparing with 23 GW installed in 2018 [3].

Wind farms are usually connected in remote areas far from the big cities and large consumption centers. A key reason for energy producers to select remote locations, such as off-shore, is the better wind conditions. Near big cities, wind suffers a lot of interference, which is not ideal for the wind generators. In off-shore wind farms, higher and more consistent wind speeds lead to the potential to generate more electricity at a steadier rate. Adding to that, we avoid visual impact and land use issues. However, some disadvantages arise from this, such as the connection to the transmission network, transportation and installation of giant turbines and finally, Operation and Maintenance (OM). Nevertheless, knowing the previously mentioned benefits, it might reveal that it is worth investing in the transmission infrastructure to access them [4]. To reinforce the need for OM, Wind Turbines (WTs) are exposed to unpredictable and harsh weather conditions which result in highly variable operational conditions that lead to intense mechanical stress (the description of more specific problems of WTs are covered in Table 2.1).

Since it is complicated to reduce the initial costs of transportation and installation, most of the research has focused on maintenance, more specifically in the use of Condition Monitoring (CM). CM consists of monitoring the components of a WT to identify changes in operation that can be indicative of a developing fault and preventing it through maintenance. CM increases the availability of the wind farm and, as a result, the production of electricity, decreasing the global cost of the project.

1.2 Contributions

The main contribution of this work is a methodology for monitoring WTs efficiently through predictive (conditioned-based) maintenance. Predictive maintenance tries to identify changes in operation that can be indicative of a future fault before they occur [5]. Other types of maintenance are described in Section 2.2.

Predictive maintenance has two main advantages: (i) increasing the availability of wind generators and (ii) decreasing the costs of wind farms maintenance, reducing corrective maintenance actions. The first advantage plays quite an important role. For example, a WT of 2.0 MegaWatt (MW) can generate 48 MegaWatt hour (MWh) during a day (in maximum) and generate a revenue of 3 600€ when considering an average Feed-in Tariff (FiT) [6] of 75€/MWh.

Until recently, monitoring has only relied on manual and straightforward analysis of specific measure-

[5]https://gwec.net/wp-content/uploads/2021/03/GWEC-Global-Wind-Report-2021.pdf

[6]Policy mechanism designed to accelerate investment in renewable energy technologies by offering long-term contracts to renewable energy producers

ments and aspects of operation [6]. However, this type of analysis is inefficient when detecting electrical (including power electronics), mechanical, and hydraulic problems. At present, developments in sensors and signal processing systems have improved the quality and quantity of data obtained [7]. This data combined with big data management, Machine Learning (ML) and improvements in computational capabilities have opened up opportunities for informed, reliable, cost-effective and robust decision-making in CM.

The key contribution of this thesis is to take advantage of these previously mentioned advances, replacing manual analysis for ML when detecting faults. By removing human judgement from this task it is less prone to error and more efficient since it handles more data in real-time.

Using real data obtained from Supervisory Control And Data Acquisition (SCADA) systems, this contribution was translated into an ML regression problem able to identify patterns in data, in the form of deviations from normal behaviour, that can represent a future failure. The second contribution is to improve the existing work on the subject by combining the strengths of the different methods proposed in each article. For instance, the models that yield better results and the pre-processing techniques that improve the performance of the model. This way, achieving better results than the ones using those approaches isolated. An additional contribution was a published review regarding the existing literature on this thesis' topic [8].

The implementation of this thesis consisted of developing six regression models, one capable of defining the normal behaviour of production and five others for the normal behaviour of specific turbine components. As seen in Figure 1.1, primarily, the models are trained considering the data from normal operation periods. Afterwards, deviations from the predicted values were detected when testing with failure periods. Those deviations were analysed in order to determine if they were indicative of a failure and with how many days in advance. Reaching the objective of predictive maintenance, detecting and preventing the failure before it occurs.

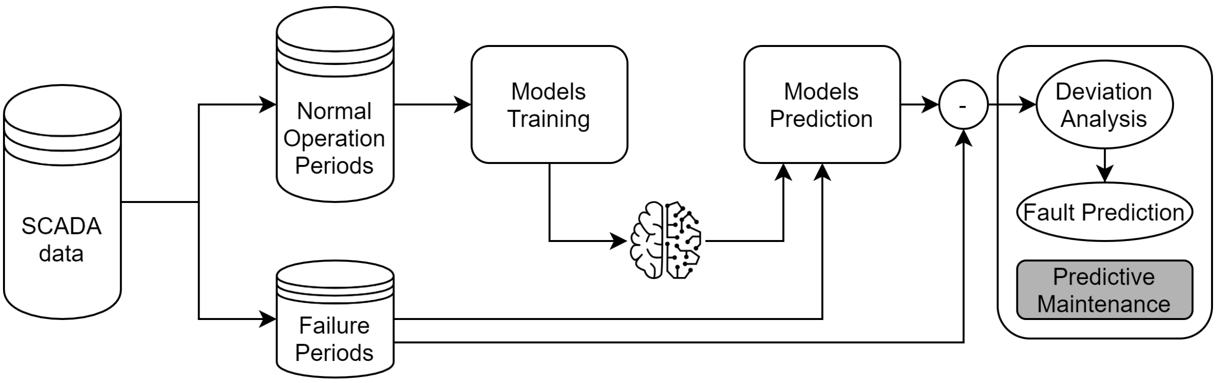


Figure 1.1: Summary of Thesis Implementation

1.3 Organization of the Document

This thesis is organized as follows: Chapter 2 comprises a brief background on the covered subjects to guarantee a better understanding of the following sections. Chapter 3 presents the relevant related work on each of the ML phases. Chapter 4 presents the methodology for the different phases that comprise the implementation of the solution, followed by Chapter 5, in which the results are described and discussed for each of the same phases. Chapter 6 summarizes the most important findings and results, ending with possible future work.

2

Background

Contents

2.1 Wind Turbine 9
2.2 Condition Monitoring of Wind Turbines 10
2.3 State-Of-The-Art Machine Learning 12

In order to better understand the following sections, Section 2.1 describes, in a superficial way, how a wind turbine works and its key components. Then, Section 2.2, presents an introduction regarding condition monitoring of wind turbines. Finally, in Section 2.3, a review of the state-of-the-art machine learning is provided in order to help to substantiate the chosen methodology.

2.1 Wind Turbine

Before describing how a wind turbine works, a short description of its key components is provided.

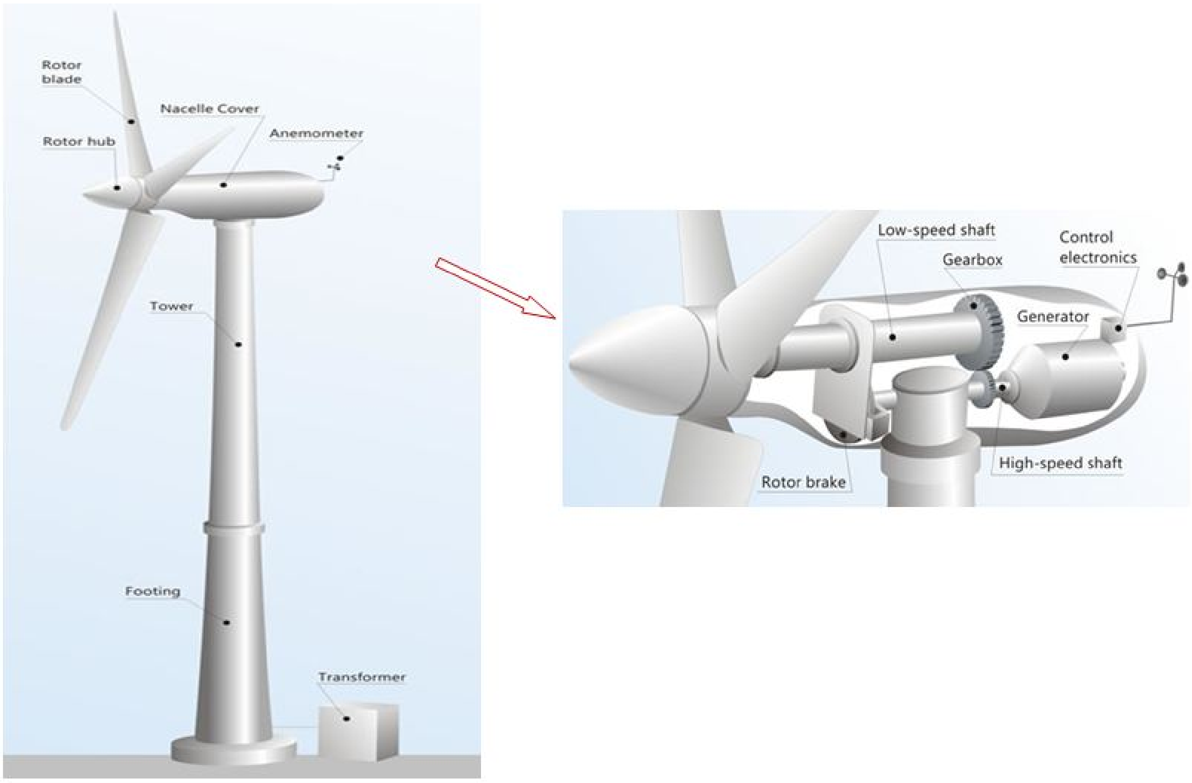


Figure 2.1: Wind Turbine Components [1]

As presented in Figure 2.1, the key components are:

- **Rotor blades** - The wind forces the blades to move, transferring some of its energy to the rotor;
- **Rotor** — Captures the power from the blades and converts it to kinetic mechanical power;
- **Transmission System**:
  - **Shaft** - When the rotor spins, the shaft spins as well. In this way, the rotor transfers its mechanical, rotational energy to the shaft, which enters an electrical generator at the other end;

  - **Mechanical brakes** - Used as a backup system for the aerodynamic braking system;
  - **Gearbox** - System of gears that converts the slow high torque rotation of the rotor into a faster rotation;

- **Generator** — Electromechanical component that converts the mechanical power into electrical power.

Wind turbine operation can be summarized in three important steps:

1. Wind force against the blades causes them to rotate and propel the rotor. The rotor is connected to the main shaft and is responsible for moving the generator;
2. Inside the turbine, there is a speed multiplier (gearbox), allowing the generator to transform mechanical energy into electrical energy;
3. The electricity is conducted through the interior of the tower to the outside power lines.

A wind turbine will start working when the wind reaches the cut-in speed. There is no justified energy conversion below the cut-in speed. The turbine's power is also limited to its rated power. Whenever it reaches the cut-out wind speed, it stops working to prevent extra mechanical stress.

2.2 Condition Monitoring of Wind Turbines

This section shows how to apply CM on wind turbines and presents the different approaches.

The maintenance cost of components of a wind turbine strongly depends on how the problem is addressed. Reactive maintenance, consists on only replacing the component when it fails and does not use CM, resulting in a more expensive approach. Small wind farms typically rely on this approach because they do not have a permanent maintenance team. On the other hand, predictive maintenance, through CM, enables an operator to know when to replace a component before a fault occurs. This prevents major failures, decreasing the costs and saving up to 20-25% of maintenance costs of wind turbines [9]. Consequently leading us to another aspect: how do we choose the turbine's element(s) to be monitored? A good strategy would be to prioritize the components that are more likely to fail or lead to long down periods. Previous studies [10] concluded that components such as the rotor and transmission system tend to have a higher rate of failure. Carroll *et al.* [11] also concluded that generators tend to have a higher rate of failure in offshore wind turbines than on onshore ones, which may be due to higher average wind speeds, harder maintenance access and higher rated power.

There are also various ways of performing CM, ones more intrusive (wear out the component) than others, including acoustic emission measurement, power quality (harmonics measurements) and temperature monitoring, oil debris monitoring, and vibration analysis [12].

Finally, we can use CM for diagnosis, fault detection in real-time, or we can use it for prognosis, i.e., fault prediction. For instance, Table 2.1 presents three of the most common failures, together with some of the most common problems and respective current condition monitoring used.

Type of Failure	Common Problems	Condition Monitoring Techniques
Blade Failure	Deterioration, cracking, manufacturing flaws and adjustment error	Ultrasound, thermography, vibration, torque and visual inspection monitoring
Generator Failure	Wearing, electrical problems, rotor asymmetries, overheating and over speed	Temperature, vibration, torque, current, voltage, power signal analysis, process parameters, performance monitoring and thermography
Gearbox Failure	Wearing, fatigue, oil leakage and insufficient lubrication	Temperature, vibration, particles in oil, acoustic emission, power signal analysis, thermography and performance monitoring

Table 2.1: Most common types of failures with their respective problems and current condition monitoring techniques [2]

Tchakoua et al. [2] discuss some of the limitations and possible improvements on current CM techniques:

- Determine the most cost-effective measurement or monitoring strategy;
- Automate the “experts” in data interpretation to automate actionable recommendations;
- Develop reliable and accurate prognostic techniques;
- Improve the use of collected data (normally only stored at 10 or 15 minutes intervals) to provide a more reliable, flexible, accurate, and efficient tool for automatic monitoring;
- Develop smart, wireless, and energy-efficient sensors that will offer opportunities for placing sensors in difficult-to-reach locations (e.g. blades);
- Focus on providing the newest and industry-proven signal processing algorithms for extracting the key features of a signal to predict machine component health.

Although these future research areas may appear challenging to address, they also represent great opportunities for CM to boost the wind industry’s success by reducing the cost of energy and increasing its competitiveness.

2.3 State-Of-The-Art Machine Learning

To better understand how ML can be helpful, this section presents a brief introduction about the subject and the state-of-the-art regarding condition monitoring to address fault detection and diagnosis issues.

Recently, the field of condition monitoring has moved from the use of conventional techniques to Artificial Intelligence (AI) techniques [13]. The conventional methods consisted of sensing technologies or analysing physical quantities (last column of Table 2.1), having the major problem of needing an expert to diagnose. AI tries to automatize this diagnosis, removing human error, while handling more data in real-time. AI through ML techniques has been widely used to improve the accuracy and efficiency of fault detection and diagnosis [14].

ML models can follow two different approaches, supervised learning predicts an output variable using labelled input data or unsupervised learning that draws inferences from data without labelled inputs. In addition, for supervised learning, we distinguish between models that predict a numeric variable (regression) or a categorical variable (classifiers) [15]. The ML model selection step is particularly significant as it is the core functionality that learns from past data and generalizes into the future. For example, Neural Networks (NNs) and Support Vector Machines (SVMs) are two popular models that have been used in ML for diagnostics and prognostics.

A NN is a computational structure inspired by the data processing and learning ability of biological neurons in the brain. NNs are arranged in layers, and each layer is composed of a set of artificial neurons. Each neuron receives an input signal, manipulates it, and then the output is forwarded to the next layer of neurons [16]. NNs have been evolving rapidly over time. In the beginning, these models could only solve linear classification problems, which in the majority of the cases, cannot be applied in fault detection. Then, NNs evolved to multi-layered architecture that could solve non-linear problems such as the feed-forward multi-layered method [17] [18], in which no feedback from the previous signal is provided to the next. Another example is Recurrent Neural Networks (RNNs) [19], which have feedback connections, and past signals are used to identify new features.

In this thesis' research field, Self-Organizing Maps (SOMs), another type of NN, are also used [17]. SOMs are trained using unsupervised learning. As a result, they produce a low-dimensional (typically two-dimensional) and discretized representation of the input space of the training samples. For this reason, they can be called maps and are typically used for dimensionality reduction. Self-organizing maps differ from other neural networks as they apply competitive learning as opposed to error-correction learning.

The availability of larger data sets, better initialization algorithms, a variety of activation functions and stronger computational power made it possible to add hidden layers (layers that allow a NN to filter/transform the data). This approach is called deep learning and has started to be used in the wind energy field. NNs can be used for a variety of tasks, such as control (e.g. wind turbine power control), fault diagnosis and forecasting (e.g. wind speed forecasting) [20], [21].

As for SVMs, they are often used in fault detection [19], [22], [23], [24]. SVMs work by finding decision boundary hyperplanes that best separate classes of instances, i.e. by leaving the widest possible margin to the instances closest to the boundary. They evolved from performing only linear classification or regression to non-linear problems by adding polynomial features created from existing ones. This method makes the problem linearly separable in a higher-dimensional space. They have gained significant importance recently because of their superior ability to generate an accurate representation of the relationship between the input and the output from a small amount of training information.

3

Related Work

Contents

3.1 Data Pre-Processing 17
3.2 Feature Selection and Extraction 18
3.3 Models 19

At the end of the previous section, a summary of the evolution of the use of ML in CM was provided. In this section, it will be covered in detail the recent research on the subject of this thesis, including possible limitations and suggested improvements. Before using ML methods, we typically use pre-processing techniques on the data. Hence, it might be helpful to first look into work related to those initial tasks. After that, models for specific tasks will be covered.

3.1 Data Pre-Processing

The data from most existing CM models comes from SCADA systems. This is an advantage because using data from SCADA turns out to be a cheap alternative (e.g., does not require any extra hardware investment) [25]. This type of system has been integrated into wind farms and wind turbines by using sensors, controlling electricity generation, and providing time-series signals in regular intervals. Unfortunately, there is still a high non-conformity between sets of SCADA signals and taxonomies [26] used by different turbine manufacturers, which makes it challenging to compare existing research.

Another challenge to be faced is that typically a wind farm has hundreds of sensors in each turbine, all of them producing signals at a high rate; this results in a "big data" problem [27].

After dealing with the previously mentioned problems, we can start by looking into the raw SCADA data collected and perform pre-processing. SCADA data are not only influenced by the structural integrity of the turbine but also affected by many other factors. For example, temperature spikes could arise due to an increase in ambient temperature and not due to an internal malfunction of a component. This task is called outlier identification and removal.

At first, one could expect that a simple outlier removal technique might solve the problem, but Marti-Puig *et al.* [28] showed that this is not the case. Although these methods decrease the error on the training data set, they also increase the error of the test data set, meaning that most of the values considered outliers by the simpler methods are true failures. Consequently, Marti-Puig *et al.* suggest the aid of an expert on the subject to define absolute and relative ranges.

Lapira *et al.* [29] applied three changes to the SCADA data obtained to filter outlier samples:

- Removal of any instances in which the output power was below zero watts or when the measured wind speed was below the rated cut-in wind speed;
- Segmenting the data into week-long intervals so that the health value could be calculated every week;
- Normalizing the data by subtracting the mean value and dividing by the standard deviation.

Yang *et al.* [30] also developed a method to pre-process raw SCADA data based on expected value calculation. The advantage of that method is that the expected value reduces the statistic error caused by outliers in the SCADA data. Moreover, methods based on the average value, as previously mentioned, may fail to consider the probability distributions of outliers.

Segmenting and normalizing the data as Lapira *et al.* [29] suggested was implemented. For some models, the non-production periods were also removed, as will be seen in Section 5.1.2. Computing the expected value as Yang *et al.* [30] was tried out but was too heavy to run on the whole data set and therefore it was not pursued. The additional methods used for data pre-processig will be presented in Section 4.2.

3.2 Feature Selection and Extraction

There is no conventional method for feature selection when using ML on CM because it depends on the component being monitored. However, it can be as simple as asking an expert if it is more useful to focus on the acoustic sensor or the generator's vibration or going beyond that and using an automatic method. After feature selection, the next step is feature extraction, i.e., compress high-dimensional time-series, reducing the amount of data that must be processed while accurately and completely describing the original data set.

Auto-encoders or Principal Component Analysis (PCA) can reduce the extracted features or combine them. An autoencoder is a type of neural network used to learn efficient data codings in an unsupervised manner. They are useful for dimensionality reduction since they learn a representation (encoding) for a set of data by training the network to ignore signal "noise" [31]. PCA is the process of computing the principal components and projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible [32]. Auto-encoders can perform similarly when the activation functions are linear, and the cost function is the mean squared error. However, when comparing dimensionality-reducing techniques, non-linear techniques are often incapable of outperforming traditional linear techniques such as PCA.

Wang *et al.* [33] proposed a variable selection algorithm based on PCA with multiple selection criteria, selecting a set of features that better identify fault signals without altering the variety of data in the original data set. Moreover, it also has the advantage of reducing the number of sensors installed by identifying the least significant variables. More specifically, the selection method proposed in the paper is the T selection method, which targets a specific fault signal [34]. This algorithm maximizes variance and maintains the independence among the selected variables and seeks to preserve the underlying features regarding the fault signal/variable within the retained data set.

W. Zhang and X. Ma [35] proposed a model that uses PARAllel FACtor (PARAFAC) analysis for fault detection and sensor selection of wind turbines based on SCADA data. PARAFAC, in resemblance with other decomposition methods, belongs to the same family of bi-linear or multi-linear methods of decomposing multi-way data into a set of loading and score matrices [36]. The difference in PARAFAC is the use of few degrees of freedom than in the other mentioned methods. This fact presents an advantage since it leads to simpler models while excluding noise and insignificant or redundant information. PARAFAC has attracted increasing interest because it is a processing technique capable of simultaneously determining the pure contributions to the data set and optimising each factor at a time in trilinear systems. By using an appropriate clustering method, measurement samples can be classified and the sensor array can be optimised. This method has firstly been applied to condition monitoring of wind turbines by [35]. More recently, in [37], they proposed the use of PARAFAC and sequential probability ratio test for multi-source and multi-fault condition monitoring, nevertheless not specific to the wind farms domain.

Peng *et al.* [38] proposed a method called Mahalanobis Distance (MD) to reduce the input variable number of the prediction model. MD method is a minimum-redundancy, maximum-relevance feature approach. MD measures the distance between a point and a distribution considering the effect of different units used for the measurement. Thus, MD methods can detect correlations between variables in a process or a system. In addition, the MD method provides a value for the univariate distance containing the main features of multivariate data. This advantage of the MD method is ideal for reducing the number of input variables to the prediction model. Furthermore, most wind farms are in remote locations and the data collected is usually transmitted to an analysis center or cloud service by wireless or optical fibre networks. Therefore, fewer input variables decrease the communication load.

As previously mentioned, we do not have a conventional method for feature selection, which can be proved by the number of different approaches in the literature. With that in mind, the followed approach was, as a starting point, to test different algorithms, beginning with simpler methods such as the use of domain knowledge, followed by the methods specified in Section 4.3.

3.3 Models

Many research works have used different ML methods for various CM tasks, each with advantages and drawbacks. In this section, a couple of the most relevant papers will be discussed. Starting with more general issues, such as "Turbine performance assessment" and "Power curve monitoring", that are not specific to a turbine component. Moving to "Wind turbine conditions monitoring" that covers multiple turbine components. Then focusing on specific faults: "Fault detection, diagnosis, and prediction of generator faults", "Hydraulic group monitoring", "Generator temperature monitoring", "Generator bearing failure prediction", "Gearbox temperature monitoring" and "Transformer temperature monitoring".

3.3.1 Turbine performance assessment

Lapira *et al.* [29] used the SCADA data from a large-scale on-shore wind turbine to assess which of the three selected models better captures the turbine's performance and degradation. The methods used to pre-process and filter outlier samples were already mentioned in Section 3.1.

The important SCADA parameters were chosen to model the wind turbine's system performance (wind speed and the average active power), splitting them into two steps: multi-regime (dynamic wind turbine operating regimes) partitioning and baseline comparison. Finally, a confidence value was generated during the baseline comparison step, which describes the health state of the wind turbine. The multi-regime models being tested were SOM and Gaussian Mixture Model (GMM). GMM is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Finally, feed-forward NNs, which used an approach based on residuals being greater than a given threshold during a given time segment. A comparison between the first two, unsupervised models and the last one, a regression model, was a major conclusion of the paper.

They found that the GMM presents a more gradual health change, being more suitable in performance prediction. Nevertheless, the other two methods can be used for fault or anomaly detection. The suggested future work was to predict the progression of the degradation using predictive techniques, computing the remaining operational time before a future downtime.

The most interesting feature of this paper is the use of unsupervised methods since most data sets composed by SCADA signals are not labelled as fault or not. As the paper states, an interesting approach is to use SOM and NNs on fault detection to label the data. The paper's addition to the existing literature is to produce a standard for manufacturers to compare performance. In this thesis NNs were also tried out for fault detection, however, a direct comparison with this paper will not be possible since they used it for a different goal. As for turbine performance assessment, only an analysis of some statistics taken from raw data was carried out. More complex techniques such as using the GMM were not justified since it is not the main objective of this thesis.

3.3.2 Power curve monitoring

The predicted power usually does not meet the reality due to several reasons. For instance, the wind speed on-site is not uniform horizontally across the face of the turbine, the vertical wind profile and the air density are different than during the calibration, and the wind data available on-site are not always measured at the height of the turbine's hub [39]. This fact is true both for a single turbine or for a whole wind farm, making it hard to assess a prediction of the energy output of a wind farm. A reliable wind power forecasting model, capable of giving rapid answers, is important for energy management.

Wind power forecasting and prediction tools enable better dispatch, scheduling, and unit commitment of thermal generators, hydro and energy storage plants, reducing the financial and technical risk of uncertainty of wind power production for all electricity market participants. Even though this is not why this tool is helpful for CM, it was probably a good reason for investing in it in terms of the market.

Marvuglia *et al.* [40] presented a data-driven approach for building a steady-state model of a wind farm's power curve under normal operating conditions. This approach allows the creation of quality control charts that can be used as a reference profile for detecting anomalous functioning conditions of the wind farm and power forecasts.

The paper compares three different ML models to estimate the relationship between the wind speed and the generated power in a wind farm:

- A self-supervised neural network called Generalized Mapping Regressor (GMR), a novel incremental network that can approximate every multidimensional function or relation presenting any kind of discontinuity. The basic idea of GMR is to transform the function approximation problem into a pattern recognition problem under an unsupervised framework.
- A General Regression Neural Network (GRNN), which is a novel incremental self-organizing competitive neural network. GRNNs belong to the family of kernel neural networks. The ordinary GRNN training procedure is a Mean Square Error (MSE) minimization, accomplished using a cross-validation (leave one out) approach.
- A feed-forward Multi-Layer Perceptron (MLP). MLPs are the most widely used state-of-the-art neural network model. The approach to regularization applied in the training phase of the MLPs used in this paper is based on the weight decay technique.

This paper has the novelty of applying power curve models to an entire wind farm and is focused on GMR. When looking into the results, the first two non-parametric methods provided more accurate results when compared with the classical parametric MLP.

Regarding future work, the paper states that labelled data classified as normal or abnormal could lead to various improvements. One of those possible improvements is the utilization of this type of algorithm to perform the prediction and diagnosis of wind turbine faults. In this case, the ML approach should be used to build a steady-state model of the reference power curve of the wind farm under normal operating conditions and to predict the occurrence and category of a fault ahead of time.

The paper [40] also covers a problem already mentioned, the lack of labelled data, being the learning focused on determining what are normal and abnormal behaviours (fault detection) and not on fault prediction. Nevertheless, the approach of considering the wind farm as a whole could be extended to other tasks (e.g. obtaining more general statistics that could indicate a possible fault not detected by a single turbine). The fact that it focuses on the whole wind farm is one of the points that was added by this paper; the other point is that it uses GMR, a novel incremental self-organizing competitive neural network. In resemblance, this thesis will also model the normal behaviour of production of the wind farm.

When modelling power curves, wind speed may not be the only dependent variable used. For example, Schlechtingen *et al.* [41] compared two classes of models: one using only wind speed as the dependent variable and one also using wind direction and ambient temperature. After searching among the several comparative studies found in literature, they selected the best performing and recommended models for wind turbine power curve monitoring and applied them for each class. Those models were Cluster Center Fuzzy Logic (CCFL), K-Nearest Neighbor (K-NN) and Adaptive Neuro Fuzzy Inference System (ANFIS). The K-NN algorithm uses feature similarity to predict the values of new data points, which further means that a new data point will be assigned a value based on how closely it matches the points in the training set. ANFIS integrates both neural networks and fuzzy logic principles. Therefore, it captures the strengths of both in a single framework. Its inference system is based on a set of fuzzy IF-THEN rules, with the learning ability to approximate non-linear functions.

Schlechtingen *et al.* [41] have shown that by adding wind direction and ambient temperature, the models fit the data better, reducing the variance in the prediction errors. This finding made it possible for earlier detection of abnormal turbine performance. Specifically, for the available data set, the anomaly is detected up to five days earlier than with the models using only the wind speed. The ANFIS model showed the best performance in terms of this metric and in terms of abnormal power output detection, whereas the K-NN model performed worst. The paper's explanation for the poor performance of the K-NN model was that the number of considered neighbors decreased by increasing the dimension of the space by adding wind direction and ambient temperature. As a consequence, the predictions become more sensitive to outliers in the reference data set.

In contrast, with the first paper [40], the previous used the presence of labelled data to predict errors having best results using the ANFIS model, which allows the incorporation of *a priori* knowledge in the form of rules. In addition to the previously mentioned model, another novelty added to the literature was including wind direction and ambient temperature as input variables. This addition was also tried out to conclude if it also improves the detection of abnormal turbine performance. However, for the dataset used in this thesis, these variables did not yield better results. The goal of assessing the power curve's normal behaviour is to detect anomalies when the power deviates from the expected. As will be seen in Chapter 4, this approach was not only followed for the normal behaviour of production but also for other wind turbine variables.

3.3.3 Wind turbine conditions monitoring

In this section, will be covered existing articles in the literature that modelled the behaviour of various turbine components simultaneously.

Schlechtingen *et al.* [42] proposed using multiple ANFIS normal behaviour models in order to detect faults in the main components of a turbine. By the use of the prediction error the models are able to detect components' abnormal behaviours and inform the engineers of a possible future fault. For each of the developed models, they selected the input features that better captured the behaviour of the output feature. They started by using a genetic algorithm combined with a partial least squares regression to do a primary selection of relevant features. Then, they trimmed the most important features using expert knowledge. Another interesting approach was using two different types of ANFIS models for the same features. More specifically, one model using signals from another turbine subsystem, and the other using, for instance, the temperature from the other phases of that component. This will allow a more informed diagnosis. The models' prediction errors are presented in two different ways. One in terms of standard deviation for SCADA data entry and another using that deviation on the daily average of the error. This averaging decreases the fault detection sensitivity and therefore the false alarms. In addition, it enhances the fault patterns leading to earlier and reliant detection. To strongly point out the advantages of ANFIS, the training times and errors were compared with a NN, showing a superior efficiency on both.

Sun *et al.* [43] developed prediction models for different condition parameters of a WT in order to detect anomalies. More specifically, parameters that are dependent on environmental conditions, which is the case of rotor speed, output power and temperatures of several components. They used a fuzzy theory method to integrate the prediction result of multiple models. For each model, the input parameters are selected based on expert knowledge. Secondly, NNs are used as the prediction model. A study carried out in this paper was to see the influence of using different types of training data. Using current data instead of historical data yielded better performance since the latter included more components' degradation. For this purpose, the Mean Absolute Error (MAE) index, was used to evaluate the different models' performance for each of the training samples. As for the quantification of the prediction error and therefore fault detection, another novelty was using a new metric called abnormal level index. Compared with threshold methods, this new metric is able to more accurately identify anomalies.

An inconvenient of having a model for each component is that each of the models needs to be updated and maintained. Meyer [44] suggested multi-target regression models in order to deal with this problem. A multi-target regression model receives as input a set of features and outputs multiple target values simultaneously. Meaning that, for example, instead of having two separate models for predicting the power and the generator temperature, we could have only one model. This technique decreases the time and work of having to do the pre-processing tasks, train and select the thresholds for multiple models. They developed six multi-target regression models, some using deep neural networks and others classical ML algorithms. Secondly, they compared the model's prediction error with the single target models. They also investigate if using models that take into consideration past observations, as

Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), present better results than the ones considering only present observations (K-NN and MLP). CNN is a NN, however, it can filter and pool the input data to create a feature map that summarizes the important features in the input. LSTM is a type of RNN, but instead of taking as input a single data point, it can process entire sequences of data. The results showed that the multi-target models achieved similar, and in some cases, even smaller, predictive errors, than single-target models. Another interesting conclusion was that taking into consideration past observations as input, did not improve the performance of the model when the target variables were strongly correlated. This is a novel promising approach since the authors were able to reach the same performance as using multiple models. The model with the best performance, LSTM, achieved an Root Mean Square Error (RMSE) of 0.51. Therefore, it will be analyzed in this thesis if the developed individual models can achieve a lower error.

Liu *et al.* [45] developed a novel approach for monitoring the overall health status of a WT taking into consideration all temperature variables in the SCADA data. First, the pre-processing of the data deals with the non-stationarity using a non-parametric regression technique. Then, a statistical process control method is applied to detect faults. When an alarm is raised, the fault is isolated using a variable selection method to determine which variables contributed to its occurrence. The SCADA data used in this study was the same as the one used in this thesis, therefore, in Section 5.3, the results obtained in both are going to be compared.

In resemblance with the previous papers, this thesis also focused on developing models for the normal behaviour of various turbine' components. As Schlechtingen *et al.* [42], a combination of algorithms for feature selection and the use of domain knowledge was tried out, yielding better results. Another similarity is in the use of daily averages, with the addition of also considering hourly averages. Sun *et al.* [43] also tried NNs as the model and Meyer [44] proposed a model that unifies all target variables, a technique that will be compared and discussed in Chapter 6.

3.3.4 Fault detection, diagnosis, and prediction of generator faults

Looking into literature that covers conditions monitoring and fault prediction, the prediction of more than a half-hour in advance is currently very poor for minor faults. Even though they are minor faults, they occur quite often, contributing to failures related to the power system. In a study carried out by the EU FP7 ReliaWind project[1], the study found that just under 40% of overall turbine downtime can be attributed to power system failures [46].

Leahy *et al.* [23] focused on fault detection, fault diagnosis, and fault prediction of minor faults. The first classification level, fault detection, is distinguishing between two classes: "fault" and "no-fault". Fault diagnosis represents a more advanced level of classification than simply fault detection. Fault diagnosis

[1] https://cordis.europa.eu/project/id/212966/reporting

aims to identify specific faults from the rest of the data. Faults were labelled in five classes, including generator heating, power feeder cable, generator excitation, air cooling malfunction faults, and others. The last level was fault prediction/prognosis that has the objective of predicting the fault before it occurs. The predictions focused only on generator heating and excitation faults, as these showed the most promising results for early detection. The data used came from a SCADA system, and 29 features were selected to be used in classification using SVM as the ML classification model. Several scoring metrics were used to evaluate final performance. The precision score being one of them, as high number of false positives can lead to unnecessary checks or corrections carried out on the turbine or wind farm. A high number of false negatives, on the other hand, can lead to failure of the component with no detection has taken place, and the recall score captures this.

For fault detection, the recall was high (78% to 95%), but precision was low (2%-4%), suggesting a high number of false positives. High recall and low precision were also found for the diagnostic and prognostic cases. As for fault diagnosis, generator heating faults showed a low proportion of false positives and correctly predicted 89% of faults. In fault prediction, the best performance was achieved with SVM trained with the addition of class weight, using a linear kernel. In general, for fault detection and diagnosis, the recall scores were above 80% and prediction up to 24 hours in advance of specific faults, representing a significant improvement over previous techniques.

Possible improvements, excluding adding more data, are using feature selection methods to find only the relevant features and speed up training time. In addition, a possible avenue for future research is determining whether trained models will still be accurate after a significant change in the turbine, e.g. after replacing a major component.

The most interesting feature in this article was how they used operational and status data to label the data, i.e. fault detection. For example, they considered an operational data point as faulty if it occurred in a time frame of 10 minutes before or after a fault present in the status data. This approach was followed in this thesis. Conversely, as the authors stated, a technique that could be improved is the feature selection, as it was based on a personal judgment that is always prone to error. In general, the paper presents simple yet efficient solutions for the three different levels of fault monitoring.

3.3.5 Pitch System monitoring

In this section I will start by revising the performances achieved for the previous works that developed multiple models. Schlechtingen *et al.* [42] modelled the normal behaviour of the pitch angle. Reaching to a standard deviation error of 0.44° and as for the daily average error, the value was 0.08°.

The pitch system presents a high failure rate leading to substantial downtime [46]. Yang *et al.* [47] proposed a method for modelling the normal behaviour of the pitch system. This paper does not require fault data, using only historical data from healthy periods. They tried three different feature selection methods to select the most important features to predict the pith motor temperature: (i) Wrapper method: Sequential Forward Selection (SFS); (ii) Embedded method: Gradient-Boosting Regression Trees (GBRT); and (iii) Filter method: Mutual Information (MI). In SFS, the model's MSE is computed and we sequentially add features to the set that contribute to the minimum of that error until it does not decrease any more. GBRT is used as automatic feature selection to compute the importance of the features. MI selected the features with a high correlation with the monitored variable. Secondly, they tried six regression models: ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO), K-NN, Random Forest (RF), NN, and Support Vector Regression (SVR). Ridge estimates the model's coefficients while performing L2 regularization. LASSO shrinks the data to a central data point, for example, the mean. RF is an ensemble of unpruned classification or regression trees, trained from bootstrap samples of the training data. SVR is a version of SVM for predicting discrete values. Contrary to SVMs, the best fit hyperplane for SVRs is the one with the maximum points. The pitch motor temperature is estimated using the trained model and if a deviation occurs, there is a possible future fault. The SVR model led to the lowest MSE. Control charts using Exponentially Weighted Moving Average (EWMA) were constructed to perceive changes in the normal behaviour of the pitch. By using an average of past and current observations they were able to smooth the effect of noise. The outliers of the control chart were indicative of a future fault. Several case studies indicated that the model can predict faults with six to seven days notice for faults as the limit switch and angle-encoder. As for the slip-ring failures, the model can detect the fault 37.3 hours earlier than the time it was indicated in the historical fault data.

It should be noted that this paper presents a solution for an electric-pitch system and an hydraulic system was considered in this thesis. However, this thesis' model intends to surpass the the disadvantage of the hydraulic pitch system's oil leakage and maintenance issues.

3.3.6 Generator temperature monitoring

Most of the generator over-temperature events and failures occur in spring and autumn, especially spring. This fact is due to the increases in the ambient temperatures in springtime and the high wind speed. If this causes a fault on the generator that leads to a shut down in the wind turbine, significant energy generation will be lost due to the time required to change/repair the generator.

Schlechtingen *et al.* [42] modelled the normal behaviour of the temperatures of the three phases of the generator. Reaching a standard deviation error of 4.98°C for the first phase and 4.92°C for the remaining two. As for the daily average standard deviation error, the values were 1.90°C, 1.87°C and 1.36°C. Sun *et al.* [43], reached an MAE and Mean Absolute Percentage Error (MAPE) of 0.72°C and 1.23°C for the first phase, of 0.71°C and 1.22°C for the second and of 0.72°C and 1.25°C for the third.

Guo *et al.* [48] proposed a new condition monitoring method, consisting of a temperature trend analysis method based on the Non-linear State Estimation Technique (NSET). NSET is used to construct the normal operating model for each wind turbine generator temperature, and then, at each time step, the model is used to predict the generator temperature. In addition, a new and improved memory matrix construction method is adopted to achieve better coverage of the generator's normal operational space.

The time series of residuals between the real measured temperature and the estimate is smoothed using a moving average window to reduce the method's sensitivity to isolated model errors, thereby improving its robustness. The average and standard deviation computed by that moving window are used to detect potential faults early, when significant changes occur, exceeding predefined thresholds, an incipient failure is flagged.

The model uses SCADA data from a wind farm that records all wind turbine parameters every 10s. In total, 47 parameters are recorded for each turbine. At the same time, the SCADA system keeps a record of wind turbine operation and fault information, such as start-up, shutdown, generator over-temperature, pitch system fault, etc. Nevertheless, only five variables were considered relevant (stored in an observation vector): power, ambient temperature, nacelle temperature, and the generator cooling air temperature.

The results showed that the new approach to the memory matrix increased the model's accuracy. The model can identify dangerous generator over-temperature before damage has occurred, resulting in a complete shutdown of the turbine. In order to compare with the NSET method, a NN was developed and then used to model the normal behaviour of the same wind turbine. Results showed that NSET achieves considerably higher accuracy modelling the normal behaviour of the wind turbine generator temperature. Moreover, NSET has another benefit compared with the neural network; it can more easily adapt to a new normal working condition.

In similarity with this paper, this thesis implementation also uses a model to define the normal operating behaviour of the generator temperature. Nevertheless, as will be seen in Section 5.2, different features were used.

Tautz-Weinert *et al.* [19] compared different approaches to normal behaviour modelling of bearing and generator temperature based on 6 months of 10 minutes SCADA data from 100 turbines. The different approaches were: linear regression, SVMs, a MLP with one hidden layer of six neurons, and an RNN with two recurrence steps, ANFIS and Gaussian Process Regression (GPR). GPR is a non-parametric Bayesian approach to regression. The input variables are found by analysing cross-correlations between SCADA variables and the target variables.

The authors use only two input variables in their baseline configuration and add further ones for a sensitivity analysis. They find the performance of RNN to be quite similar to the MLP, with both NN types usually outperforming other approaches. GPR and SVM techniques, however, do not perform as accurately as the other models. SVM and ANFIS tend to have larger errors with more inputs. GPR modelling works well for the generator temperature prediction but less well for the bearing temperature prediction. The authors noted that adding interactions to linear models is beneficial, whereas introducing recurrence in the NN model seems to be helpful for some turbines but leads to inferior performance for others.

A combination of the two suggested feature selection methods was tried out in this thesis, domain knowledge to select some initial features and analysing correlations between variables at a later stage.

3.3.7 Generator bearing failure prediction

Starting with the papers covering multiple components previously mentioned, Schlechtingen *et al.* [42] modelled the normal behaviour of the generator bearing temperatures. Reaching a standard deviation error of 1.44°C for the non-drive end and 1.86°C for the drive end. As for the daily average, the values were 0.73° and 0.75°C. Sun *et al.* [43] reached an MAE of 0.17°C and MAPE of 0.43°C for the non-drive end and 0.21° and 0.45° for the drive end.

Schlechtingen *et al.* [49] compared the performance of two artificial intelligence approaches (autoregressive NNs and full signal reconstruction NNs (non-linear NNs)) to a regression-based approach when learning to approximate the normal bearing temperature. The regression models used as SCADA input signals, the power output, nacelle temperature, generator speed, and generator stator temperature. This task also used data smoothing techniques in combination with the learning techniques. By using a smoothing filter, the higher-order variations can be filtered and the prediction error reduced.

Although NNs can handle fuzzy or incomplete data, they are sensitive to invalid data. Therefore, one must typically use a pre-processing technique, which is particularly important when training a network. The network might not give an optimal generalization otherwise. The principal pre-processes applied were: Validity check - data is checked for their ranges and their consistency by filtering extreme outliers and data with unexpected high gradients; Data scaling; Missing data processing; Lag removal - Wind turbine signals usually do not respond immediately to changes of operational conditions. Many wind turbine signals can be closely correlated to other simultaneously measured signals, and only some are related to the output signal (bearing temperature). The related signals and their lag to the desired signal can be found by using linear cross-correlation.

In [49], the authors found that the non-linear NN approaches outperform the regression model. However, they are more challenging to interpret. In comparison to the regression model, the full signal reconstruction NN had an averaged error with reduced amplitude and was more accurate, leading to reduced alarm limits. An alarm is triggered 30 days before the bearing breaks. The autoregressive model has a very high accuracy, thus detecting very small changes in the autoregression of the temperature signal (50 days in advance).

Kusiak and Verma [50] estimated an expected behaviour model of a generator bearing by training an MLP to predict generator bearing temperature. The model is trained on high-frequency (10 s) SCADA data from 24 wind turbines of the same type and location. Two turbines that showed over-temperature faults were used for testing and model validation. Some of the input variables were selected by domain knowledge (selecting 50 out of 100) and subsequently by applying three different data-mining algorithms: wrapper with genetic search, wrapper with best-first search, and boosting tree algorithm. The residuals were smoothed with a moving average filter (window size 1 hour). An alarm is triggered if the smoothed residuals exceed two standard deviations on the training set. The authors find that their method can predict over-temperature fault on average 1.5 h ahead of the fault occurrence.

Both papers, [49] and [50], used NNs to detect faults on the generator bearing. However, the first paper [49] used more complex approaches resulting in an earlier prediction of the fault when compared with the second paper. The first paper [49] presented interesting ways of pre-processing the data, and the second paper [50] strength is the three different feature selection algorithms. In resemblance, this thesis also tried NNs, and combines the previously mentioned strengths of both articles, pre-processing and feature selection.

3.3.8 Gearbox monitoring

Starting with the multiple models' papers previously mentioned, Schlechtingen *et al.* [42] modelled the normal behaviour of the gearbox bearing temperature. Reaching to a standard deviation error of 0.79°C and for the daily average, the value was 0.35°C. Sun *et al.* [43] reached to a MAE of 0.11°C and a MAPE of 0.29°C.

Zhang *et al.* [51] proposed using a NN to define the normal behaviour of the gearbox bearing temperature. When the deviation from the actual values exceeds a certain threshold, the WT operator was flagged. By using domain knowledge, the variables used as input were active power output, nacelle temperature and turbine speed. The average error found was 0.026°C and the RMSE was 0.2. Since these values are quite low, a direct comparison with the actual temperature to detect faults is pursued. From a case study, they were able to give a warning to the WT operator with three months notice and an alarm with 10 days notice.

Orozco *et al.* [52] used an unsupervised regression approach to diagnose WT gearbox failures. They adjusted the gearbox temperature signal to neglect rises in temperature caused by ambient temperature or increases in power production. A novel addition was only taking as input features the ambient temperature and power output to model the gearbox temperature. The paper suggested four regression models to define the normal behaviour of the gearbox: linear regression, multivariate polynomial regression, RF and NN. Multivariate polynomial regression is characterized for having as linear variables, the coefficients of the polynomial expression. The models that leaded to the lowest value of RMSE were linear regression and a multivariate polynomial regression. The tried models were able to detect turbine failures caused by high gearbox temperature that leaded to a downtime. Additionally, by using unsupervised learning, they do not require failure logs.

Fu *et al.* [53] proposed a model for the normal behaviour of the gearbox bearing temperature. This paper used as feature selection a method called adaptive elastic network. After selecting the relevant variables, they use a CNN for feature extraction and dimension reduction. At last, LSTM uses this reduced data to predict the gearbox bearing temperature. They use two metrics, RMSE and the MAPE to evaluate prediction results. By computing these metrics for different lengths of input data, they concluded that for the biggest length tried out, they achieved an RMSE of 2.023°C and a MAPE of 4.623%.

3.3.9 Transformer monitoring

Schlechtingen *et al.* [42] modeled the normal behaviour of the temperatures of the three phases of the transformer. Reaching to a standard deviation error of 3.59°C for the first phase, 2.88°C for the second and 3.25°C for the third. As for the daily average, the values were 1.64°C, 1.32°C and 1.47°C.

In the best of my knowledge, there is no recent research on a regression model for the normal behaviour of the transformer. The majority of the papers used a classification model, which would be difficult to directly compare to this thesis' model.

4

Methodology

Contents

4.1 Description of Data And Faults 33
4.2 Data Acquisition and Pre-processing 35
4.3 Feature Selection 39
4.4 Model Selection 40
4.5 Validation 41

The contribution of this thesis is to take advantage of state-of-the-art machine learning and apply it in the field of wind energy, boosting wind turbines availability. To accomplish this, the focus was on improving the existent techniques in predictive maintenance, which act before a major failure occur using CM. Since ML was used, that was translated into regression models capable of establishing the expected normal behaviour of the turbine. Therefore the logical extension was, to detect deviations from that expected normal behaviour that are indicative of a future fault. As it can be seen in Figure 4.1, ML comprises four stages and all of them have challenges to be faced. Therefore, the course of the thesis followed the same stages, using a combination of the strengths found in each article in the existent literature. That being said, this section will present this thesis' methodology divided in the following subsections: "Description of Data And Faults", "Data Acquisition and Pre-processing", "Feature Selection", "Model Selection" and "Validation".

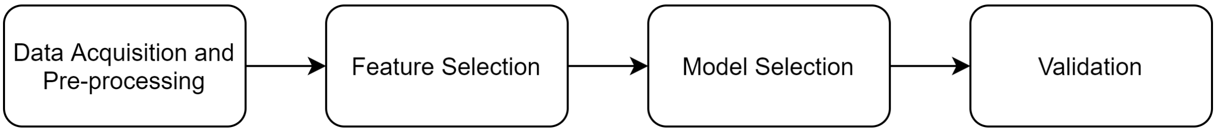


Figure 4.1: Machine Learning Phases

Using real data obtained from SCADA systems, described in Section 4.1, the first task, Data Acquisition and Pre-processing will deal with the variety and quantity of collected data. In other words, this task does the necessary pre-processing of the data to remove outliers and treat missing values, as delineated in Section 4.2. Feature Selection, present in Section 4.3, selects the features that better represent the patterns in data, removing unnecessary noise. According to the obtained data, the third stage, Model Selection, aims to select the best ML model to detect and predict faults, as described in Section 4.4. Finally, in the Validation stage, Section 4.5, it is assessed the accuracy of the model on identifying new data as representative of a failure in the turbine or not.

4.1 Description of Data And Faults

As already mentioned in Section 3.1, the data used on most of the existent CM models are obtained through SCADA systems. Among the available open-source datasets, the one that proved to be more complete was the one from Energias De Portugal (EDP) [1]. Therefore, the implementation of this thesis used the EDP dataset and insights on the data will be given in this section. The data refers to two years, from 2016-01-01 to 2017-08-31, providing SCADA records from five wind turbines.

[1] https://opendata.edp.com/explore/?sort=modified&refine.theme=Wind

4.1.1 Operational Data

The operational data is composed by the signals file that includes SCADA signals for each wind turbine's most important components and production values. An example of some entries for the signals file is given below in Table 4.1. As it can be seen, for the same time entrance there is different information for each one of the five turbines. The SCADA data are sampled at the frequency of ten minutes and 81 parameters are recorded for each turbine.

Timestamp	Turbine_ID	Gen_RPM_Max	...	Nac_Direction_Avg
2016-01-01 00:00:00+00:00	1	1277.4	...	218.5
	6	1270.0	...	204.6
	7	1317.5	...	197.3
	9	1376.7	...	214.0
	11	1339.4	...	206.9
2016-01-01 00:10:00+00:00	...	...	...	...

Table 4.1: Signals file for the EDP dataset: five first entries

4.1.2 Status Data

The status data consists of a failure logs file, which is the historical failure logbook for the five turbines, with logs such as replacement and repair processes, errors, high signal values and component failures. Once more, an example of some entries is provided in Table 4.2.

Timestamp	Turbine_ID	Component	Remarks
2016-03-03 19:00:00+00:00	11	GENERATOR	Electric circuit error in generator
2016-01-01 00:10:00+00:00	6	HYDRAULIC_GROUP	Error in pitch regulation
2016-04-30 12:40:00+00:00	7	GENERATOR_BEARING	High temperature in generator bearing
2016-06-07 16:59:00+00:00	9	GENERATOR_BEARING	High temperature generator bearing
2016-07-10 03:46:00+00:00	7	TRANSFORMER	High temperature transformer

Table 4.2: Failure logs file for the EDP dataset: five first entries

Table 4.3, summarizes the number of times each type of failure occurred in each of the turbines. This table presents the case studies that are going to be analysed for each of the models developed in Section 5.3. It can already be concluded that the models for the normal behaviour of gearbox and transformer temperatures are the ones with fewer cases.

Component	Turbine	Nr of failures
Hydraulic Group	T6	2
	T7	1
	T11	2
Generator	T6	5
	T7	1
	T11	1
Generator Bearing	T7	2
	T9	4
Gearbox	T1	1
	T9	1
Transformer	T1	1
	T7	2
	Total	23

Table 4.3: Number of failures by component and turbine

4.1.3 Faults

As a starting point, the focus was on fault detection, meaning that it was a simple regression problem turned into a classification problem with only two possible outputs: Fault or Non-Fault. For this case only looking into the turbine's production behaviour, determining if, for a certain wind condition, the turbine is producing as expected. At a later stage, the focus was on detecting the different types of faults, moving from fault detection to fault diagnosis. To achieve this goal, it was supervised the behaviour of specific components, which are represented by variables that can be obtained in the signals file. When looking into the status data, it is possible to identify five different types of faults associated with specific components: generator, hydraulic group, generator bearing, transformer and gearbox. Therefore, having five different models, one for each component.

4.2 Data Acquisition and Pre-processing

Primarily, handling the SCADA data consisted of analysing how the information is being displayed and the specificity of the data. This step let us perceive the extent of pre-processing needed and determine how detailed the prediction can be. The dataset is initially stored in a *CSV* file, therefore, to import and analyse the data in a programming environment, in this case, Python, there is a need to save the data into an appropriate format. For this end, pandas library [2] was used since it is already prepared to load such files into an efficient data frame. Since the process of preparing the data is yet quite extensive, Pickle library [3] was used avoiding the need to run it every time the pre-processed dataset is needed, saving it for later use.

[2] https://pandas.pydata.org/pandas-docs/stable/index.html
[3] https://docs.python.org/3/library/pickle.html

To help differentiate which part of the data is affected by the model's training, testing and each of the pre-processing phases, we can look at Figure 4.2. The first task is to determine the turbine's normal behaviour, therefore, the training set is obtained by removing the ten days before and after each failure. The choice of the ten days was made taking into consideration a balance between providing enough time for the turbine's operator to schedule maintenance and not removing too much data from the training set. The test set is only the ten days before the failure, and the ten days after are either not part of the training set. The reason for the latter is that after the failure, the turbine might take some time to reestablish the normal operation conditions, not being able to include it in the training set, and for prediction purposes, it is also not relevant, therefore, it is also not part of the test set. Dealing with NaNs, missing data and data labelling are done using the whole dataset. Turbine performance assessment and removal of outliers are only applied on the training set for reasons given in Section 4.2.3 and Section 4.2.4, respectively.

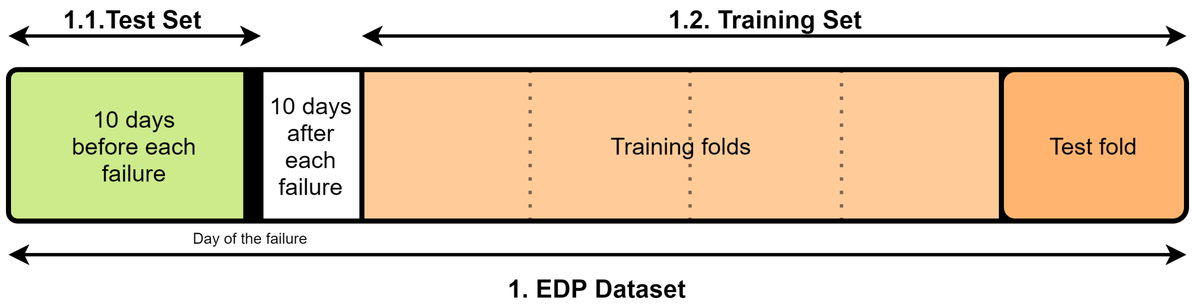


Figure 4.2: Dataset Partitions

4.2.1 Dealing with NaNs

Although this is a common problem in most of the datasets, when looking into the EDP's dataset, it does not occur. Nevertheless, if that was not the case, the approach that was going to be followed is the same as the one that was used for missing data (Section 4.2.2), in the following subsection. If the aim was to increase the performance even further, it would be used a different approach taking into consideration the variable in which the NaN occurred. For instance, developing a regression model that takes as inputs the variables with a direct relation to the NaN variable, helping to predict it.

4.2.2 Missing Data

Gaps in SCADA data exist due to occasions when the turbine is inactive during periods of low and high wind speeds. Additional gaps occur due to the occurrence of scheduled maintenance and faults [54].

The first task, dealing with missing data, was tackled following multiple approaches and choosing the one that yielded better results. Nevertheless, beforehand, it is important to determine the acceptable number of sequential missing values that justify the use of such approaches. For instance, if the following ten entries for that variable are missing as well, it may indicate an error in the sensor. As will be seen in Section 5.1.1., this case did not occur but if it did, instead of filling it with a suitable value for that variable, it would be used a value distinct from all other values, representing an error in the sensor. Subsequently, as a first approach, the missing values were filled with a random value from the ones present in the data for that variable. Since this is the simplest method, it was considered as a baseline. The following methods tested were: (1) Use the last value obtained for that field; (2) Mean for that column and (3) Value estimated by another predictive model. The first two methods have been taken from Pandas documentation[4], the library used to load the dataset. The last method was an improvement to the interpolation method provided by the same library. Instead of using a linear function to predict a missing value, was used a regression model. The code for the implementation of these methods can be found in Appendix A, Listing A.1.

The previous techniques refer to a missing variable in a row. When looking into the cases where entire rows are missing the same approach was not used since the noise that would be introduced by inserting an estimation for a whole row would affect the model's accuracy.

4.2.3 Turbine Performance Assessment

When developing the model, it was taken into consideration the normal behaviour of production to label data points close to a failure as normal or also a failure. Therefore, as a starting point, by looking into the raw data, some conclusions on the normal behaviour of the wind farm can be taken. For instance, if the individual turbines perform differently from each other, tracing a profile for each one. Knowing in advance if some turbines yield better performance can be helpful for the subsequent steps. To better visualize the differences in performance, the followed approach was to plot the output active power for different wind speeds. Primarily for the whole wind farm and secondly, for each turbine individually. Since only normal periods of operation are relevant for this step, for both approaches only power outputs greater than zero kiloWatt (kW) were taken into consideration and the periods of ten days before and after a fault were removed. An example of the mentioned plots is presented below in Figure 4.3, for turbine one. More examples will be found in Section 5.1.3.

Analysing how the turbine is working near a failure was also helpful to establish if the normal behaviour of production can be used to detect the failure. If the behaviour appears to be normal, there is a need to look at the specific component related to the fault, as will be seen in the example of Section 5.3.3. It was also important to see if the developed method of detecting outliers was able to remove points that deviate from the power curve, as seen in Section 5.1.4.

[4] https://pandas.pydata.org/pandas-docs/stable/user_guide/missing_data.html

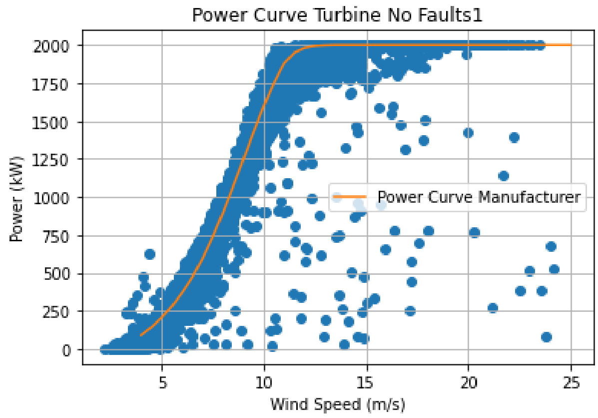


Figure 4.3: Power Curve Turbine One (raw data)

4.2.4 Removal of Outliers

Removal of outliers is probably the most challenging task since it is quite difficult to identify an outlier in such volatile data, having the risk of removing points considered as outliers that in fact are indicative of a future failure. For that reason, a data point considered an outlier was not removed if it occurred ten days before a fault. That being said, the followed method was to test several approaches, identifying the removals that yield better performance. Starting with simple methods using statistic measures, Z-score and IQR-score were tried [55]. The Z-score is the signed number of standard deviations by which the value of an observation or data point is above the mean value of what is being observed or measured. The number of standard deviations used was also defined by finding the one that yielded better results. IQR-score is the first quartile subtracted from the third quartile.

Afterwards, more complex methods were tried out, for instance, using clusters to detect outliers in each turbine running the algorithm for each day. The first approach tried out was to separate the data into two clusters using agglomerative clustering[5], one to represent normal data and another to represent the outliers. The entries comprised in the clusters considered as outliers were removed. The reason for running the algorithm for each day is that the behaviour of wind turbines is extremely volatile. Leading to a second experience, instead of separating the data by day, it was separated in intervals of 1m/s of wind average speed from 0 to 25 m/s. Since was already taken into consideration the volatility of the wind by dividing the data into wind bins, it was not divided into days. Lastly, a different clustering algorithm was tried out, Density-Based Spatial Clustering of Applications with Noise (DBSCAN)[6] . This algorithm allows data points to not have a class, meaning that they were the ones considered as outliers. The approach of using bins of 1m/s of wind speed was also tried for this algorithm.

For reasons that will be seen in Section 5.1.4, the simpler methods, Z-score and IQR-score required

[5] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html
[6] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

an additional modification. The data were differentiated into three zones, each one representing a different turbine's operating state. The first one is from 0 m/s till the cut-in wind speed, the second one from the cut-in to the rated wind speed and the last one from the rated to the cut-out wind speed. Only nine rows in the entire dataset exceed the cut-out wind speed so that zone was neglected.

The evaluation and choice between the different methods of outlier removal were done taking into consideration the balance between not removing too many data points, decreasing the model's error and providing good results graphically as mentioned in the last part of the previous section.

4.2.5 Data Labelling

Another challenge related to data is due to the fact that, normally, each signal is not labelled as normal or abnormal. For this particular dataset, there is a file that contains the existent faults, as described in Section 4.1.2, nevertheless, there is still a need to cross that information with the signals file to label the data. Consequently, starting by crossing the timestamp of each failure and labelling entries in the signals file that occurred in that time frame as abnormal, similarly to the method followed by Leahy *et al.* [23]. More specifically, labelling a data point as faulty if it occurred ten minutes previous or after each fault. Nevertheless, this technique is not accurate in labelling the signals prior to the fault that are indicative of possible failures. For this problem, the models described in Section 4.4 were developed.

4.3 Feature Selection

Subsequently, the second Machine Learning phase and quite an important one, feature selection. All the datasets comprising SCADA data have dozens of features resulting from having multiple sensors in each turbine. In this thesis specific case, the EDP dataset includes 81 features per turbine, the normal deduction is that only some of them will be relevant for predicting a failure. The remaining should not be taken into consideration since they will only be working as noise, leading to a decrease in the accuracy of the model. As mentioned in Section 3.2, a conventional method for this problem does not exist. Therefore the followed approach was to test several methods, in order to find the one that yielded better performance. K-best and wrapper methods, such as forward and backward selection [56] with different numbers of variables were implemented and analysed. K-best selects features according to the K highest scores. In this case the scoring function chosen was "f_regression" from the scikit-learn implementation[7], which uses F-statistic. The F-statistic is the ratio between the variation in sample means and the variation within samples. Additionally, the correlation between features, visual analysis and the use of domain knowledge from existent literature [30] were also tried out. The visual analysis

[7] https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html

consists of looking at the value of each variable at the time of the fault, for faults that occurred more than one time, to see if there is some strange behaviour in common.

As it will be seen in the next section, multiple models have been developed and consequently, different features for each of them need to be selected. The model considered to be the most important was the regression model of the normal behaviour of production, since different faults can be detected by a decrease in production. For that reason, all the feature selection methods were tried out. For the remaining models, only filter methods, methods that do not require choosing a model, were used. This is justified by the best results being found for a combination of variables selected by multiple methods. Filter methods, the correlation between features, visual analysis and the use of domain knowledge ends up by providing a bigger set of features to try out than the ones that would be obtained by using wrapper methods. Additionally, they do not change according to the model or model parameters, which is time-consuming to optimize.

4.4 Model Selection

The third stage of ML and therefore the third phase of this thesis was to choose the model that better captures the normal behaviour of WTs. The optimal conditions to develop a supervised classification model are not met, since there is only a small amount of data points labelled as fault, as it will be seen in Section 5.1.5, therefore, a regression model was implemented. Nevertheless, this regression model was able to classify the data by looking into the deviation between the predicted and the true value of the dependent variable. High deviations were classified as faulty, and negligible deviations were classified as normal behaviour. Six models were developed, a regression model to establish the normal behaviour of production, and five regression models for specific turbine's components behaviour. These five models are for the hydraulic group, gearbox, transformer, generator temperature and generator bearing temperature. The regression model to determine the normal behaviour of production was tried out for the entire dataset and for each turbine isolated. Conversely, the models for a specific component were trained for certain turbines only, since not all faults occur on multiple turbines, and none of the faults occurs in all turbines. Primarily, to complement the work done in Section 4.2.3, "Turbine Performance Assessment", monitoring of the power curve was implemented. As suggested in the literature [57], a polynomial model[8] was tried out, establishing the degree that yielded better results. Secondly, the most popular model in the literature was experimented with, Neural Networks (NNs), most specifically, the MLP[9]. For this model, hyperparameters optimization consisted of choosing the number of neurons and layers that yielded better results. At last K-NN[10] was also implemented, choosing the number of

[8] https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html
[9] https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html
[10] https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html

neighbours that resulted in the smallest model's error. Therefore, for the normal behaviour of production, three different models were tried out. Conversely, for each turbine's specific component only the last two models were tested.

Firstly, each model was trained under normal operation conditions (part 1.2. of Figure 4.2), meaning that it was excluded the days before and after the failure. The objective of all the models was to predict the expected value under the normal operation of that component and compare it with its true value. This prediction is done for the day of the failure and the antecedent days (part 1.1. of Figure 4.2). If the error of the prediction exceeded a certain threshold, a fault is pointed out. Therefore, this phase also includes the determination of that threshold. This method is applied for each data point with the goal of detecting faults. In terms of fault prediction, a more coarse approach was followed to be more robust to isolated points that exceeded the threshold. Instead of analysing the error for each data point, the daily average error was taken into account, i.e., if the daily average error on the eighth day before the failure exceeded the threshold, it can be concluded that the fault is predicted with eight days notice. If that prediction occurs only on the day of the failure, the hourly error was used in order to establish with how many hours in advance the fault is detected.

As will be seen in Section 5.3, some of the results for the daily error presented isolated days above the threshold a few days before the failure, returning to normal after that. Since that might be due to the turbine's generator being under a different mechanical stress state, it was opted for dividing the data into three different groups, each one representing a different level of turbine's stress. One for the days where the mean daily production was under 50% of the rated capacity of the turbine (level one), the second one for the days between 50% and 80% (level two) and the last one for above 80% (level three). This meant that three different models and three different thresholds were obtained, being each of the days previous to the failure, classified taking into consideration the level in which they are inserted. Finally, since the impact of production was studied, the impact of the ambient temperature was also considered, analysing if an isolated daily average error above the threshold was caused by an increase in ambient temperature.

4.5 Validation

Lastly, there is a need to substantiate the decisions made with concrete proof that they improve the model's performance. Since it is a regression problem, there is a need to look into the model's error when predicting the desired values. The selected implementation of the MLP regressor uses the squared loss to optimize the model, therefore MSE is going to be the main metric used to compare models and parameters. The remaining metrics are computed in order to compare with the ones used in existent work on the subject, them being the MAE, MAPE and Root Mean Squared Percent Error (RMSPE). When comparing MSE and MAE, the latter is not concerned if we are predicting a lower or higher value than supposed since it is absolute. On the other hand, MSE assigns more weight to the bigger errors, which in our case is important since faults are going to deviate more from the expected value than normal data, being easier to detect using MSE than MAE. The same logic can be applied for MAPE and RMSPE since they are only the percentage of the previous metrics, with the difference that the root of the MSE is used for RMSPE. RMSPE or RMSE are more widely used than MSE in literature to evaluate the performance of the regression model compared with other models, as it has the same units as the dependent variable. The formulas for these metrics can be found below. Being $\hat{Y}(w)$ the predicted output and $Y(w)$ the actual output, as a function of some variable, for example, wind speed *w*. All the metrics are the mean of a total of *N* samples.

$$MSE = \frac{1}{N}\sum_{N}^{w=1}(\hat{Y}(w) - Y(w))^2$$

$$MAE = \frac{1}{N}\sum_{N}^{w=1}|\hat{Y}(w) - Y(w)|$$

$$MAPE = \frac{1}{N}\sum_{N}^{w=1}\frac{|\hat{Y}(w) - Y(w)|}{\hat{Y}(w)} * 100$$

$$RMPSE = \sqrt{\frac{1}{N}\sum_{N}^{w=1}(\frac{\hat{Y}(w) - Y(w)}{\hat{Y}(w)})^2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{N}^{w=1}(\hat{Y}(w) - Y(w))^2}$$

These metrics are going to be computed during cross-validation. Cross-validation consists of dividing the data into folds, as can be seen in Figure 4.2. The training set was divided into five folds. Four folds are for training the model and one fold is used as validation, the test fold of Figure 4.2. This procedure is repeated five times, choosing a different fold as the validation set. Since the model was not trained using that set, if it is able to predict the correct values, it can be concluded that the model is capable to generalize to new data. The mean of the computed metrics for each of the five folds was used. The hyperparameter optimization of all the models was done using cross-validation to ensure generalization to different parts of the dataset and prevent the overfitting of the training set.

In addition to using the previous metrics to prove the efficiency of the model, it was also used graphics and statistics to validate the choices made as will be seen in the next section.

5

Results and Discussion

Contents

5.1 Data Acquisition and Pre-processing 45
5.2 Feature Selection 52
5.3 Model Selection and Case Studies 56

In this chapter the results of the implementation of the methods mentioned in the previous section will be discussed. The chapter will have a similar structure to Chapter 4, concluding, for each step of the proposed methodology, which approach yielded the best results.

5.1 Data Acquisition and Pre-processing

5.1.1 Missing Data

This task consists of filling the gaps in EDP's dataset so that it can be processed by the machine learning model. By analysing the missing values, it can be concluded that there are seven rows, each one with two variables missing, i.e., a total of fourteen empty entries. The two variables were the same for all the rows, more specifically, 'average temperature in generator bearing' and 'average actual phase displacement'. For this particular dataset, since there are isolated missing values, the rule mentioned in Section 4.2.2, for checking if a sensor error is causing sequential missing values, was not applied.

By looking into the file containing the failures for each turbine, it can be concluded that some of the missing values will not have an impact on the models. For instance, the first row with two missing values is at turbine one, which happens to do not present any problem related to those variables. The same can be said for the rows that happened in turbines six and eleven, one in each. Leaving four relevant rows to analyse. The analysis was done for the first row and applied for the remaining. The row that was analysed is from turbine seven, occurring at timestamp '2017-08-17 11:30:00', three days before a generator bearing failure that occurred on day 2017-08-20, meaning that it will affect the period before the failure. In Table 5.1, can be seen the values estimated for each of the methods mentioned in Section 4.2.2 for the variable 'average temperature in generator bearing' and if using that value, the point was considered as a fault or not.

Method	Value [°C]	Fault	Model's MAE [°C]	Model's MSE [°C]
Random	36	Yes	3.7282	32.2760
Last Value	48	No	3.7255	32.2439
Mean	44.48	No	**3.7252**	**32.2429**
Regression Model	47.99	No	3.7255	32.2439

Table 5.1: Methods for Missing data, estimated value, classification and model's errors

As we can observe in Table 5.1, the random method estimated a value way different from the other methods, causing the point to be classified as faulty, meaning that it can be excluded as the approach to be followed. To choose between the remaining three methods, it was considered the MAE and the MSE of the first attempt of the model developed in Section 5.3.5, for the previous ten days of that failure, without removing outliers. By a small difference, using the mean of the column leads to a smaller error. Adding to that, it is the most robust method since the last value might be missing or wrong and the regression model is not optimized for each variable. Therefore, the column's mean was used to fill all the missing values.

5.1.2 Removal of non-production periods

When considering the regression model for the normal behaviour of production (Section 5.3.2), data points where the power production was lower than zero kW were removed since they do not represent periods when the turbine is producing energy. For the regression models that trace the normal behaviour of a specific component of the turbine, that same removal was tried out.

Effects of only considering production greater than zero kW for a specific turbine's component:

Component	Turbine	MSE before removal	MSE after removal
Hydraulic Group	T6	7.335	0.902
	T7	7.010	1.004
	T11	8.785	0.934
Generator	T6	19.356	23.883
Generator Bearing	T7	10.020	13.445
	T9	72.143	95.929
Gearbox	T1	1.711	1.081
	T9	2.813	3.013
Transformer	T1	32.629	33.264
	T7	49.293	46.000

Table 5.2: Results of removal of non-production periods by component and turbine

After the removal of production under zero kW, the choice of the model and hyperparameter optimization was repeated. Therefore, the MSE values presented in table 5.2 are already optimal ones, moreover, they are presented without the normalization since that would lead to smaller values, and, consequently, would be less interesting to analyse. Except for the hydraulic group all the other models representing specific components showed worse results by removing non-production periods. It is believed that it might be due to the hydraulic group having a different behaviour when the turbine is under the necessary conditions to produce and when it is not. In other words, the hydraulic group is responsible for adapting the nacelle position to the wind direction and the pitch angle, according to the wind speed, to the optimal angle in order to achieve the desired production. This behaviour occurs when the turbine is under the necessary operational conditions to produce. When it is at rest, the angle is set to 90°, preventing the blades to rotate, avoiding unnecessary mechanical stress. For the cases where temperatures are being monitored, such as the generator phases and generator bearing temperatures, the periods when the turbine is not producing probably do not affect the error of the model since the temperatures are just lower at those periods. Therefore, that periods only contribute with more data points that are easy to predict and decrease the model's error. A more complicated case was the gearbox and transformer component that showed better results for one turbine and worse for the other. Therefore, there is a need for another criterion, different from the model's error to decide if the non-production periods should be removed or not. For the transformer, the fault prediction (days notice) showed worse results, therefore the non-production periods were kept. The detection of outliers showed that the non-production periods should be removed for the gearbox component. Additionally, the prediction of failures also presented better results when removing those periods. As a matter of fact, even for the generator cases, the outlier removal phase will remove periods with production equal to zero kW and in that stage, it will increase the model's accuracy. Therefore, it can be concluded that only some data points of that periods should be removed. For instance, low production causes lower WT components temperature and that behaviour should be considered as normal and may be important for the model's capability of detecting true failures.

5.1.3 Turbine Performance Assessment

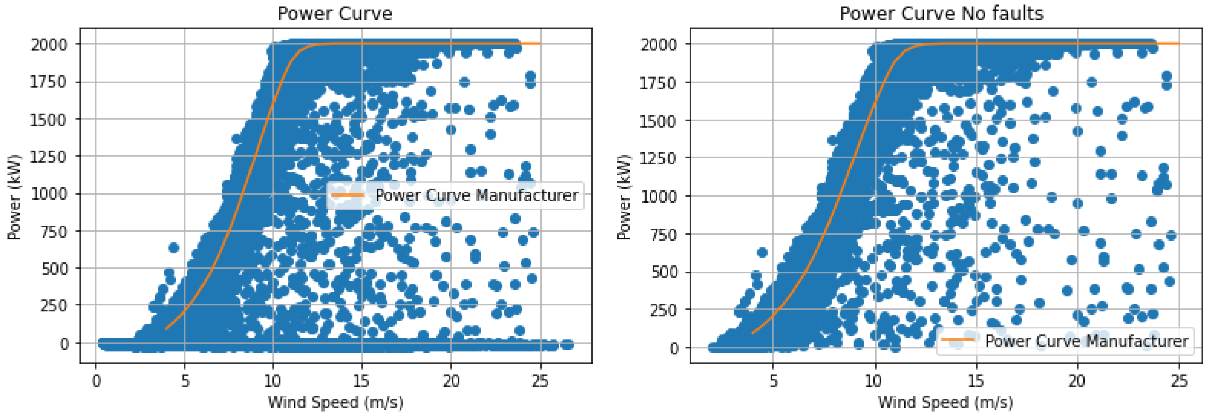
Since it is being dealt with a dataset with information from five different turbines, before developing a regression model that learns the normal behaviour of production of a turbine, it might be interesting to see if it should be expected different results from each turbine.

To compare if there are differences between treating the data as equal or separating it into turbines, primarily, the data was considered as a whole, i.e., as a wind farm. To ensure that it is only being considered data representative of normal behaviour, the ten days after and before a failure and the periods with no production were removed. Afterwards, the same approach was followed for each turbine. In Table 5.3, it is provided some statistics regarding these removals.

	Total (1)	Non Production Periods (2)	Failures (3)	After failures removal (4)	After non production removal (5)
Wind farm	434 115	131 740	57 934	376 181	266014
T1	87 140	27 737	5 740	81 400	55 763
T6	85 382	26 856	17 039	68 343	47 593
T7	87 201	27 160	14 385	72 816	51 121
T9	87 150	28 027	12 146	75 004	52 703
T11	87 242	21 960	8 624	78 618	58 834

Table 5.3: Number of rows: (1) initially, (2) with production under zero kW, (3) ten days before and after each failure, (4) after the removal of failures rows, and (5) previous column plus the removal of non-production periods

As it is possible to see in the first column (1), there are approximately the same number of rows per turbine. Turbine T6 has less rows due to the replacement of the generator that caused 1 372 missing rows. Turbine T11 is the one with fewer rows with production under zero kW, as shown in column (2). A possible explanation for this decrease might be due to different elevations between turbines, i.e., higher locations lead to higher wind speeds; unfortunately, that information is not available. The data after the two removals, column (5), is the one that was used for the regression model of the normal behaviour of production. As can be seen, turbine T1 and turbine T11 are the ones that present more data. The reason for that is the reduced number of failures and therefore the reduced number of removed rows in column (3). Turbine T1 presents only two failures and turbine T11 presents three. As for turbines T6, T7 and T9, present seven, six and five failures, respectively.



(a) Before removal

(b) After removal

Figure 5.1: All Turbines Raw Data

As we can see in Figure 5.1, by removing production under zero kW and periods close to failures, are removed a lot of data points far away from the manufacturer power curve. The same happens for each of the turbines. As mentioned in Section 4.2.3, similar graphs were used to show the efficiency of outliers removal (ex.: Figure 5.2). Through deviations from the power curve, the graphs were also used to determine if the fault should be detected using the model of the normal behaviour of production or of a specific component (ex.: Figure 5.4(b)).

	Mean Production			Standard Deviation of Production		
	(1)	(3)	(5)	(1)	(3)	(5)
Wind farm	2673.390	755.897	2975.433	3605.741	3978.642	3952.221
T1	524.719	646.997	756.269	236.192	549.583	81.138
T6	546.605	494.559	806.274	203.381	351.334	90.908
T7	550.081	625.508	765.341	282.517	519.451	88.821
T9	512.532	372.007	765.328	193.421	243.113	91.635
T11	555.563	535.612	747.372	89.728	73.418	77.751

Table 5.4: Mean and Standard Deviation of Production [kW] for the data before removals (1), the failure periods (3) and after removals (5)

To compare how much, on average, the wind farm and each one of the turbines are producing we can look at Table 5.4. The active power ranges from 0 to 2 000 kW, leading to high standard deviations when considering the mean. Therefore, the standard deviation is also presented in Table 5.4, however, not considering the deviation to the mean, but to the expected theoretical value given by the manufacturer power curve. Before the removals (column (1)), the mean production of each one of the turbines is approximately the same, and the wind farm is five times that production, as expected. Even so, all the turbines presented a low mean production when compared to the rated power of 2 000 kW. That difference, in percentage, is called capacity factor, rounding 27% for the wind farm, representing its poor use of the installed power. As for the failure periods (column (2)), turbines T6 and T9 have lower mean production, which is due to the majority of the failures on that turbines directly affecting the generator, causing the turbines to produce less in average. Turbine T1 and turbine T7 have fewer failures related to the generator, therefore, present high mean production on the failure periods. After removals (column (5)), the mean production of the wind farm ceases to be approximately five times bigger than the individual production of the turbines. The reason behind it is that the periods where each turbine is not producing or fails, are different, affecting multiple timestamps. Leading to an overall decrease in production when considering the bigger picture, i.e, the wind farm. Which, combined with the low capacity factor, also explains the high standard deviation for the wind farm for column (5). The low capacity factor has an higher impact on all wind farm columns since the sum of the production for the five turbines in each timestamp is being compared with five times the expected theoretical value for the power curve. It can be seen that the failure periods (column (2)) deviate the most from the power curve, yielding their removal a decrease in standard deviation for all turbines (column (5)). Turbine T11 has a more stable behaviour regarding the power curve in all the three periods, which can be due to multiple reasons (e.g. steadier wind speeds resulting from different elevations, geographical positions, etc [58]).

5.1.4 Removal of Outliers

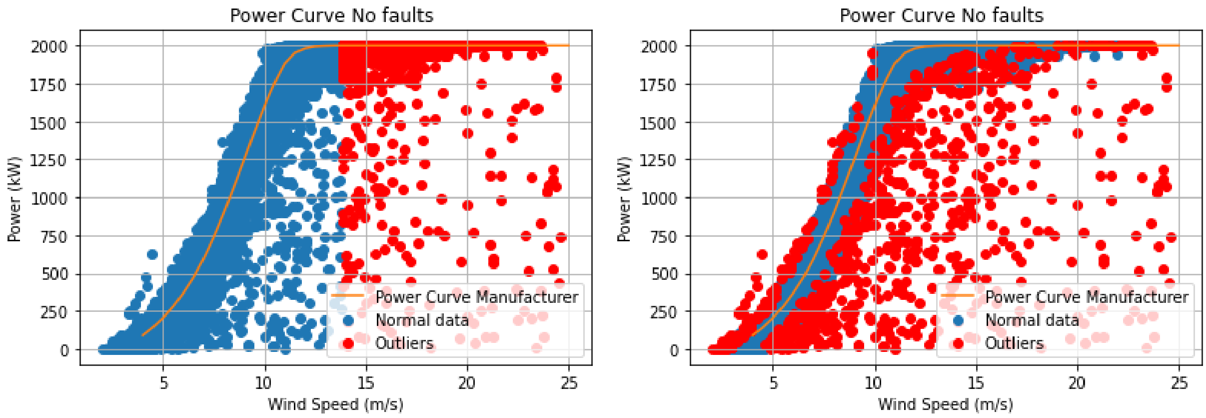
This ML stage was done in an unusual way when compared with the literature. The normal sequence of events is to first remove outliers and then do feature selection; however, since EDP's dataset have 81 features and multiple models were developed, this approach was not leading to great results. Too many entries were being considered as outliers since an abnormal variable among the 81 variables could cause the point to be considered as an outlier, and the variable responsible for the removal could not even be important for the model. Therefore, as the first approach to feature selection, domain knowledge was used to select only the most important features. For the model of the normal behaviour of production, only two variables were considered, the output power and the wind speed, since they are the ones used in power curve monitoring [40] [41]. Recalling that, as mentioned in Section 3.3.2, following the approach of [41], adding wind direction and ambient temperature, did not yield improvements. For the normal behaviour of a specific component, domain knowledge and correlation were used to select some features, more details will be presented in Section 5.2. The hyperparameter optimization and choice of the model were done before and after outlier removal.

Starting with the normal behaviour of production, Table 5.5, summarizes the model's error for each of the methods and the respective number of outliers removed. As mentioned before, the model's error is the MSE without normalization. In Table 5.5, the number within parenthesis after the Z-score is the number of standard deviations used by the algorithm from which a point was considered an outlier.

Method	Nr. Outliers	Model's MSE [kW]
No removal	-	5470.533
Clustering	18 899	4020.438
Clustering divided by wind speed	284	4066.757
DBSCAN	6 265	3852.541
DBSCAN divided by wind speed	**1 004**	**3519.129**
Z-score(2)	11 964	4761.931
Z-score(3)	2 681	5066.571
IQR score	4 603	4977.846
Z-score(3), 3 sections	951	4230.078
IQR score, 3 sections	3 900	4174.504

Table 5.5: Methods for Outlier Removal - Normal Behaviour of Production

The results using all the variables were omitted since, as mentioned in the beginning of this subsection, they yielded too many removals. Z-score and IQR-Score were not working properly as we can see in Figure 5.2(a). The algorithm considered points above certain wind speeds as outliers, due to the fact that high wind speeds are rare and therefore deviate from the mean. To solve this problem, the implemented solution was dividing the data into three different sections, each one representing low, medium and high wind speeds, as described in Section 4.2.4. Since the clusters already had an implementation that considered wind speed bins, this approach was only followed for Z-Score and IQR-Score. Moreover, for these two methods, this approach presented better results than separating by wind bins. As we can see, by only removing 951 points, it decreases, even more, the model's error than by removing 11 964 data points without considering those sections. The two clustering solutions presented quite different results from one another. The first technique that only divided the data into two clusters, did not remove all the points dispersed from the power curve. The reason behind it is that the two clusters were not enough to cover all the different behaviours of production, even when dividing into bins of wind speed. The solution for that could be computing the silhouette to find the optimal number of clusters but that was not viable for such a big dataset. The value of silhouette represents simultaneously how well separated are the clusters from different classes and how cohesive are the clusters from the same class. Conversely, DBSCAN, showed great results when divided into wind speed bins. By only removing 1 004 data points, it achieved the smallest model's error from all the techniques. Figure 5.2(b), also confirms that this method presents better results graphically, removing data points far away from the power curve.



(a) Wrong outlier removal Z-score (2)

(b) DBSCAN divided by wind speed

Figure 5.2: Outlier Removal - Normal behaviour of production model of the wind farm

When running the algorithms for the normal behaviour of production of each turbine, in addition to trying each outlier removal method on the turbines isolated, it was also tried removing the ones already found. More specifically, the ones found for each of the turbines using DBSCAN on the whole wind farm. All the turbines showed the best results by isolated removing outliers using the DBSCAN algorithm divided by wind speed. Removing the data points considered as outliers by the same algorithm but for the whole wind farm also presented good results. Nevertheless, some outliers specific to each turbine were not identified and removed, yielding worse results.

Concerning the outlier removal regarding the models for a specific component, as seen in Table 5.6, all the turbines components presented better results using Z-Score divided by the three wind sections. The reason why hydraulic group and gearbox presented fewer outliers was that the non-production periods were already removed. For the remaining components, the algorithm removed some of the data points where we had no production, causing the number of outliers to be bigger. Moreover, reducing the number of considered variables also affects the number of outliers. The hydraulic group and gearbox components were the only ones with less than ten selected variables (Table 5.8). The transformer and generator had to take into consideration the temperature of the three phases as features. As for the generator bearing, the temperature of the non-drive end and the drive end was taken into consideration. Therefore, these three components have more features and consequently more outliers. By looking at the model's MSE, all the components presented better results by removing outliers. Nevertheless, when choosing the method, a balance between decreasing the model's error and not removing too many data points was always taken into consideration. In addition, methods that graphically did not present the best results were not chosen, in resemblance to Figure 5.2.

Component	Turbine	Chosen Method	Nr. Outliers	MSE before removal	MSE after removal
Hydraulic Group	T6	Z-score(4), 3 sections	552	0.902	0.178
	T7	Z-score(4), 3 sections	836	1.004	0.177
	T11	Z-score(4), 3 sections	744	0.934	0.156
Generator	T6	Z-score(3), 3 sections	2906	19.356	18.319
Generator Bearing	T7	Z-score(3), 3 sections	2615	10.019	9.936
	T9	Z-score(3), 3 sections	2978	72.144	65.217
Gearbox	T1	Z-score(3), 3 sections	558	1.081	1.000
	T9	Z-score(3), 3 sections	567	3.013	2.872
Transformer	T1	Z-score(3), 3 sections	2855	32.629	31.664
	T7	Z-score(3), 3 sections	2701	49.293	46.647

Table 5.6: Results of outlier removal by turbine component

5.1.5 Data Labelling

As described in Section 4.2.5, the timestamp of the failures file was crossed with the signals file to label the ten minutes before and after a failure. The 23 entries of the failure file resulted in 44 entries labelled as faults in the signals file. At this stage, the type of failure was not specified since the developed models should be able to determine that. The expected number of labelled entries would be 46 since the signals file have entries with ten minutes interval. However, two of the faults occurred at existent signals file timestamps, leading to three labelled entries. In one case the turbine shut down before the failure, having zero labelled entries. And in two other occasions, the turbine shut down after the failure, leading to only one labelled entry, the one before the failure.

5.2 Feature Selection

In this section, the choice of features for each model, choosing the ones that lead to the best results will be deepened. The methods used for the model of the normal behaviour of production and for the normal behaviour of a specific component were different, as mentioned in Section 4.3. Therefore they were separated into two subsections, Section 5.2.1, for the production model and Section 5.2.2 for the turbine components.

5.2.1 Normal Behaviour of Production

As described in Section 4.3, for the monitoring of the normal behaviour of production were tried some filter and wrapper algorithms. When using the polynomial model, there is no need to select features, since the behaviour of production is going to be fitted only according to the wind speed. Therefore there are only two relevant variables, output power and wind speed. Until the moment, those two variables were the ones being taken into consideration for the remaining tried models. However, in this section, the features were optimized in order to achieve the best possible model for the normal behaviour of production. First trying the features with a correlation above 0.8 with the dependent variable. Secondly, the K most important features are selected by the K-best algorithm, with values of K ranging from one to ten. The features selected by these two methods will be equally tried on the suggested models. The wrapper algorithms, forward and backward selection, being dependent on the model were experimented for each of the tried models isolated. The backward selection algorithm was too computationally heavy to run on the entire dataset, therefore it was only used a sample of 20%. Finally, a combination of the features selected by each of the algorithms isolated was tried.

Model	Correlation	K-best	Forward Selection	Backward Selection	Combination
MLP	1582.921	589.435 (K = 1)	392.253	556.679	346.935
K-NN	1491.197	415.050 (K = 6)	139.899	172.847	**132.983**

Table 5.7: Model's MSE [kW] for each Feature Selection Algorithm

The hyperparameter optimization was done for each of the methods, being the model's MSE the optimal one. As seen in Table 5.7, the method that yielded the best results was using a combination of features. For MLP the selected features were wind speed, rotor speed and blades pitch angle. As for K-NN, the selected features were **wind speed, generator speed and blades pitch angle**.

For the models of the normal behaviour of production for each turbine, the same features were used.

Since it was noticed that a combination of features selected using multiple algorithms yielded better results, this technique was applied in the next subsection, for the models of a specific turbine component.

5.2.2 Normal Behaviour of Turbine Components

The first attempt at feature selection started at Section 5.1.1, Section 5.1.2 and Section 5.1.4, where there was a need to present which approaches led to the best results. Therefore, having to reduce the number of features used in each model. The followed approach was to select features based on domain knowledge present in existent literature.

Yang *et al.* [30] suggested some features and gave some insight into each of the turbine's components. In the case of the hydraulic group, there are two applications for this component. It is used for controlling the position of the nacelle, being relevant to look at miscorrelations between the wind direction and the direction of the nacelle. The second application is to control the angle of the pitch according to the wind speed. For EDP's dataset, the faults were only related to the pith angle. Therefore, the considered features included wind speed and blade pitch angle. Additionally, as suggested by the paper [30], generator speed, rotor speed and the output power were added. As for the generator component, since the existent faults are related to its temperature, the temperature of the three phases of the stator windings were considered. It was also added generator speed and power, as suggested in [30]. For the generator bearings, they suggested as features: 1) Temperature of main bearing; 2) Grease level for lubrication of main bearing; 3) Rotor speed and 4) Generator power. EDP's dataset does not contain feature 2), so the remaining three were considered. In terms of domain knowledge, the gearbox component is responsible for adapting the turbine speed to the generator speed. Consequently, the rotor and generator Revolutions Per Minute (RPM) were taken into consideration. The authors [30] suggested gearbox vibrations, oil temperature, pressure and level as features, however, there is only oil temperature and gearbox bearing temperature in the signals file. As for the transformer, Yang *et al.* [30] suggested the use of transformer temperature, current and voltage, both at grid and turbine side. Therefore, they were added as features to take into consideration for that component.

Secondly, the features that had a correlation above 0.9 with the dependent value of that model were added. The selection of that dependent variable will be further discussed in Section 5.3. In Table 5.8, are summarized the chosen features by this approach, enhancing in bold the dependent variable. For the outlier removal and all the previous results, these were the variables considered.

Component	Theoretical Features	Added By Correlation
Hydraulic Group	**Blades Pitch Angle**, Wind Speed, Generator Speed, Rotor Speed, Power	-
Generator	**Stator Windings Temperatures (3 phases)**, Generator Speed, Power	Generator Bearing Temperature, Wind Speed, Transformer Temperatures (3 phases), Current (3 phases), Temperature in the Busbar Section
Generator Bearing	**Generator Bearing Temperature**, Rotor Speed, Power	Temperature in the Split Ring Chamber, Temperature in the Busbar Section Stator Windings Temperatures (3 phases)
Gearbox	**Gearbox Bearing Temperature**, Rotor Speed, Generator Speed, Gearbox Oil Temperature	-
Transformer	**Transformer Temperatures (3 phases)**, Voltage (3 phases), Current (3 phases)	Power, Stator Windings Temperatures (3 phases)

Table 5.8: Feature Selection - First Approach

Afterwards a more fine feature selection was carried out, collecting a set of relevant features and see which ones yield a decrease in the model's error. Starting by doing a visual analysis as described in Section 4.3. For each type of failure that occurred more than once, each variable was analysed to see if, in both occurrences, it deviates from the mean of that column (more than around one standard deviation). Since it was a manual analysis, the failures were narrowed to only the ones that had similar descriptions in addition to occurring on the same component of the turbine.

The hydraulic group component presented some failures with descriptions related to the brake circuit and oil leakage. For these faults, it was noted some deviations in the active and reactive power. Excluding the features already considered by the previous methods. The generator bearing presented high-temperature faults and even faults where the bearing was replaced. For those faults the features that presented abnormal behaviour were: generator speed, active power, current (3 phases), blade pitch angle and capacitive and inductive reactive power. The visual analysis for the transformer did not add new features. As for the gearbox and generator component, they did not present faults with similar descriptions in order to do a visual analysis. Although it is a primary and simple approach, the majority of the features that presented abnormal values were related to the component where the fault occurred.

Component	Turbine	Selected Features	Model's MSE	Model's MSE (Refined)
Hydraulic Group	T6	Rotor speed, Wind Speed, Active Power, Generator Speed, Power	0.178	0.116
	T7	-	0.177	0.105
	T11	-	0.156	0.111
Generator	T6	Generator Bearing Temperature, Transformer Temperatures (3 phases), Current (3 phases), Temperature in the Busbar Section, Gearbox Oil Temperature, Gearbox Bearing Temperature, Temperature measured by the IGBT-driver on the rotor side inverter (3 phases), Temperature in the top nacelle controller, Temperature in choke coils	18.319	11.479
Generator Bearing	T7	Power, Temperature in the Busbar Section, Temperature in the Split Ring Chamber, Stator Windings Temperatures (3 phases), Generator Speed, Wind Speed, Temperature in the top nacelle controller, Temperature in the choke coils	9.936	9.840
	T9	-	65.217	55.661
Gearbox	T1	Stator Windings Temperatures (3 phases), Ambient Temperature, Wind Speed, Gearbox Oil Temperature, Rotor Speed, Generator Speed	1.000	0.417
	T9	-	2.872	0.727
Transformer	T1	Power, Stator Windings Temperatures (3 phases), Generator Bearing Temperature, Temperature oil in hydraulic group, Temperature in the Busbar Section	31.664	26.522
	T7	Power, Stator Windings Temperatures (3 phases), Temperature in the Busbar Section, Active Power	46.647	44.579

Table 5.9: Feature Selection by Turbine Component

Afterwards, the correlations between the previously selected variables for each component and the rest of the available variables were considered, to see if more relevant features were found. This time correlations above 0.8 were considered. The last tried procedure was the K-best algorithm, where the ten most important features were selected for each of the components and added if they were not already taken into consideration.

After collecting all relevant features for each of the components, the result was sets of features with considerable sizes. The followed procedure was to test the eliminations or additions of features that resulted in a decrease of the model's error, leading to the smallest possible error. Reaching to the features present in Table 5.9. These tests were done for one of the turbines with failures on that component and generalized for the remaining. Therefore, only specifying the features for the first turbine. Nevertheless, it was one exception, the transformer model, more specifically, turbine T7, that had worse results using the features selected for turbine T1. Consequently, the feature selection was repeated for this case. That might have been due to the fault on turbine T1 being on the transformer fan and the faults from turbine T7 being related to the temperature of the transformer.

Comparing the model's error before the refined feature selection (fourth column of Table 5.9), with the present error (last column of Table 5.9), it can be seen a decrease in all model's error. The ones that presented a considerable change in the features compared to the ones previously selected, were the ones with a visible decrease in error. For instance, in the generator component was added and removed many features causing the model's error to decrease from 18.319 to 11.479.

5.3 Model Selection and Case Studies

This section presents the tried models for defining the normal behaviour of production and for each of the turbine components. Primarily, is specified the threshold used for each of the models to define from which value a data point should be considered as a failure. Afterwards, in each of the subsections, are presented the results of using the selected model for detecting and predicting faults. At last, conclusions of the global results are taken.

It should be noted that the comparisons made in this section, between the results from this thesis' models and the ones from existent literature, are limited. The considered datasets were different, with exception of [45], moreover, the data pre-processing approaches and models also differ. This limitation will be addressed in Section 6.2.

5.3.1 Thresholds Assessment

As mentioned in Section 4.4, an important part of the modelling of the normal behaviour of any component is to define from which values to consider a point as abnormal. There is a need to take into consideration that a lower threshold would cause more frequent alarms and therefore, more frequent maintenance actions. On the other hand, high thresholds may fail to identify small deviations and cause later predictions of the fault. This choice depends on the size of the maintenance team operating the wind farm and their priorities in terms of early detection or avoiding false alarms. Therefore, this thesis considered thresholds that better suited the trade-off between those two criteria for the available faults in EDP's dataset. More specifically, the used thresholds were the model's error plus the standard deviation of that error multiplied by a constant. The squared error was considered (expression (1) below) since, as mentioned in Section 4.5, MLP uses the squared loss for parameter optimization.

$$(1)\ SE = (\hat{Y}(w) - Y(w))^2$$

Squared Error (SE) of the model between the predicted output, Ŷ(w) and the actual output, Y(w)

$$(2)\ Threshold_{normal} = \text{MSE}_{train} + \sigma_{SE_{train}} * constant$$

$$(3)\ Threshold_{fault} = \text{MSE}_{fault} + \sigma_{SE_{fault}} * constant$$

As seen above, two approaches were considered, the model's error being the one corresponding to the error from the training using the periods of normal operation (expression (2)) or the error from the prediction of an existent fault using the ten days prior to its occurrence (expression (3)). The former is more scalable to changes in data and was used for the daily and hourly average error methods (prediction methods). However, the latter presented better results on detecting faulty data points for some models. Another parameter that needed to be tuned was the constant that was used to multiply the standard deviation, reflecting the mentioned trade-off. Therefore, this constant was obtained by manually trying different values, choosing the one that did not considered too many data points as failures, but capturing the ones that presented an higher deviation. These thresholds should be updated following changes in data, for instance, if the model's error is improved, the constant might need to be adjusted.

Component	Turbine	Chosen Approach	Detection Constant	Chosen Approach	Prediction Constant	Prediction w/ Levels Constant
Production	-	Normal	1	Normal	1	1
Hydraulic Group	T6	Fault	1	Normal	10	10
	T7	Fault	1	Normal	10	10
	T11	Fault	1	Normal	10	10
Generator	T6	Normal	10	Normal	10	10
Generator Bearing	T7	Fault	1	Normal	1	2
	T9	Fault	1	Normal	1	1
Gearbox	T1	Normal	3	Normal	2	3
	T9	Normal	3	Normal	2	3
Transformer	T1	Normal	4	Normal	4	2
	T7	Normal	6	Normal	10	6

Table 5.10: Threshold Approach and Constant

For the cases where the failure error was used as a threshold, one of the faults were selected for each turbine and applied for the remaining faults. As seen in Table 5.10, for those cases there was no need to multiply the detection standard deviation by any constant. For the hydraulic group, all turbines presented low values for the model's error during training, therefore, if that error was used, the constant would have to be extremely high to avoid false alarms. The values for the daily and hourly methods were not presented separately since the same constant was used.

5.3.2 Normal Behaviour of Production

The model for the normal behaviour of production can be used to detect any type of failure that directly or indirectly causes a decrease in production. For modelling the normal behaviour, three models were tried: (1) polynomial model for power curve monitoring; (2) MLP and (3) K-NN.

Starting with the polynomial model for power curve monitoring, the expected value of production was defined given the wind speed. Beginning by defining the polynomial degree that better captures the data. For the model considering the data from all the turbines, the best results were obtained using a polynomial degree of eleven. For the individual turbines, a similar degree was obtained, twelve for turbine one, thirteen for turbine six, eleven for turbine seven, seventeen for turbine nine, and turbine eleven achieved better results for a higher polynomial degree of nineteen. The results for the last two turbines could be an indicator of more volatile data for that turbines, needing a more complex model, and therefore a higher degree. Another possible reason might be more data points concentrated between the cut-in wind speed and the rated wind speed, the harder region to model. As we can see in Figure 5.3, this model, although simple, is able to capture the behaviour of the data, achieving a similar curve as the one provided by the manufactures.

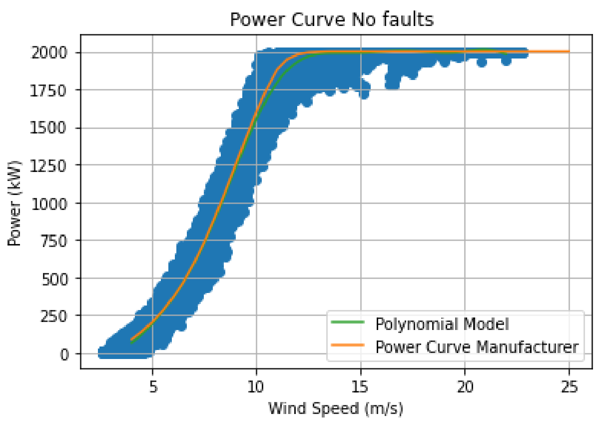


Figure 5.3: Polynomial model - All turbines data

Secondly, it was analysed if splitting the data into wind direction bins, achieves a better model performance. For each one of the wind direction 30° degree bins, the optimal polynomial degree was computed, obtaining lower degrees for bins with fewer data points. By doing the mean of the MSE of each bin, a value of 3725.042 kW was reached, as the previous model had an error of 3550.328 kW. Therefore, this division does not yield improvements.

By looking at Table 5.7, it can already be concluded that K-NN yields better results than the polynomial model and MLP. The polynomial model is a simple model that only considered one variable to define the normal behaviour of production, therefore, yielded worse results. As for the MLP, it might be due to its incapability to efficiently model the relations between the volatile input variables and the output variable. The K-NN model, even by only taking into consideration the five closest neighbours to generate a prediction, achieved good results. Moreover, K-NN provides a good trade-off between fast training and the model's performance. In addition, it does not vary over multiple runs, while for MLP there was a need to do a mean of ten runs, which is time-consuming.

For K-NN, considering five neighbors, was obtained an MAE of **7.65 kW** and a MAPE of **10.53 %**. As mentioned in Section 4.5, these results were obtained by doing cross-validation, therefore they are the mean of the results of the five validation folds. Comparing with the existent literature, [40] achieved the best results using GRNN model. More specifically, an MAE of 4.63 kW and a MAPE of 76.74%. Even though these results are scaled to a maximum of 100 kW, the obtained results in this thesis, unscaled, present better metric' values. The second paper, [41], achieved better performance for ANFIS, with an MAE of 32.01 kW, also unscaled, a worse result than the obtained. The obtained results by Sun *et al.* [43] were 39.12 kW and 11.52%. Therefore, this thesis' model achieved the best results.

The model developed in this section was used for the cases where the fault had a direct repercussion on the production. These were the cases of a replacement or repair of the generator and an electric circuit error on the generator. It was also used when the fault data points obtained in Section 4.2.5 deviated from the power curve, however, this will be covered in the next subsections. Even for the cases where the points are in concordance with the power curve, this model can be used to confirm the results obtained using the model for the specific faulty component.

Beginning with the generator replacement problems, there are three case studies, two in turbine six and one in turbine seven. Since the results for the model including all turbines and the ones from each turbine were similar, the results for the wind farm model will be presented. Using the method for fault detection mentioned in Section 4.4, **98.92%** of the data from the normal periods of operation were considered in fact normal. This is a good indicator that the model will only consider accurate deviations from the normal behaviour in the faulty period in order to detect faulty points. The method used for fault prediction considered zero days as faulty in a total of 609 days of normal data. For the same reasons, this also proves the accuracy of the method. For the first fault of turbine six, around 3% of the data points on the failure period were considered faulty, with incidence on the second day and the fault prediction method showed also a prediction with two days notice. As for the second fault, disperse data points were considered faulty with a total of 2% and it was not possible to predict the fault using this model. In Section 5.3.4, the generator temperature model was used to analyse if the generator replacement had to do with an over-temperature problem. As for turbine seven, 3.41% of faulty data points were found on the days before the fault, with incidence on one and eight days before the fault. This was confirmed by the prediction model, where an MSE above the threshold was found on days one, eight and ten.

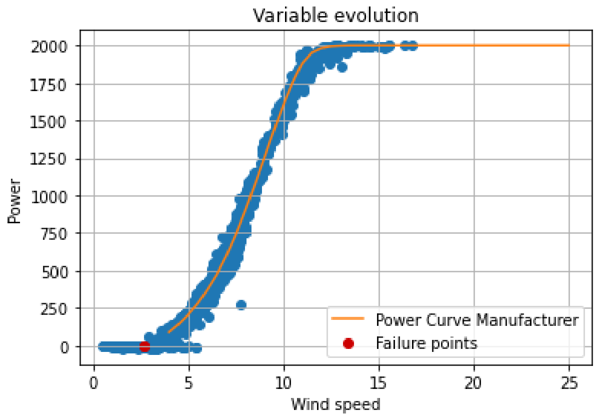
Secondly, the electric circuit error in the generator of turbine eleven. First, a model for the normal behaviour of the reactive power was tried out, as suggested in [30]. However, that did not lead to great results, therefore the model developed in this section was used instead. The detection method found around 2% of faulty data points around the sixth day before the failure. This was also observed using the prediction method where the daily average error on the sixth day was above the threshold.

5.3.3 Normal Behaviour of Hydraulic Group

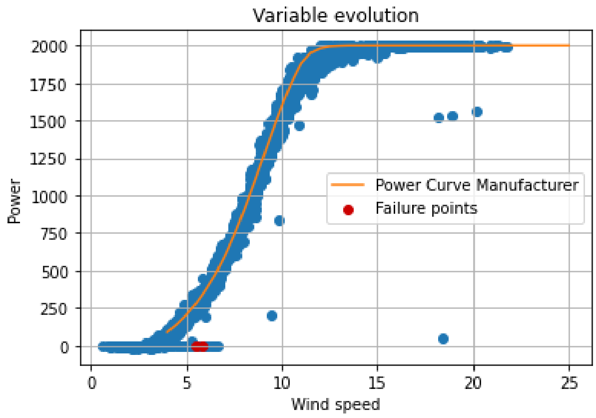
For the modelling of the normal behaviour of the hydraulic group, by observing the values of the variables on the time of the fault, it can be concluded that the blades pitch angle presented anomalies in the majority of the cases. Therefore, this variable was chosen as the one to monitor. There are five case studies for this component, as it was seen in Table 4.3. The values for the model's error on Table 5.9, are already the optimal ones and were found using K-NN for all the three turbines, with K equal to four for turbines six and seven, and equal to five for turbine eleven.

In Section 3.3.5, the results achieved in the literature were discussed. Comparing with Schlechtingen *et al.* [42], this thesis' results achieved a standard deviation error for the blade pitch angle of 0.34°, 0.32° and 0.33° for turbine six, seven and eleven, respectively, which surpassed the obtained 0.44° by [42]. As for the daily average standard deviation, all the turbines achieved a value around 0.004°, half of the value obtained in [42]. Yang *et al.* [47] were able to predict a fault with six to seven days for one specific type of pitch fault and with 37.3 hours for another type. Although having different types of specific pitch faults, the prediction was done with at least eight days notice. Two of the three cases were oil leakage faults, by being able to predict this type of fault, shows that the model was capable of surpassing the disadvantage of the hydraulic pitch system when compared with the electric pith used in [47].

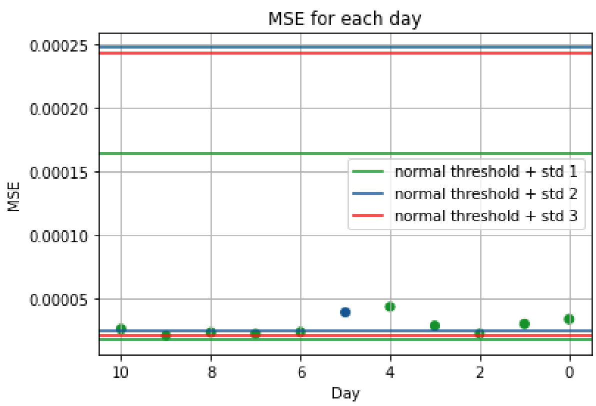
For all the turbines, **100%** of the data from the normal data period was considered normal and zero days were considered faulty, which are great indicators of accurate modelling of the normal behaviour. In addition, the models' error present in Table 5.9, is quite low for every turbine. Beginning with the first fault of turbine six, only 1.19% of the data from the ten days prior to the fault was considered faulty, with an incidence on the day of the fault. In terms of fault prediction, the MSE of the day of the fault and two, five and eight days before the fault was above the threshold. Meaning that it was possible to predict the fault with eight days notice. The second fault of turbine six presented 3.48% of faulty data points, also concentrated on the day of the failure. The prediction was also possible with eight days notice. For the only case of turbine seven, 7.84% of the data points were considered faulty, all of them on the day or the day before the fault. This result was confirmed by the prediction, with an abnormal daily average error on the day, the day before and ten days before the fault. Meaning that an alarm of a possible future fault can be given to the turbine operator with ten days notice. At last, turbine eleven, for the first fault 4.97% of data points distributed over the day, three and four days before the fault presented abnormal behaviour. The prediction showed an MSE above the threshold for all the days until day eight before the failure with exception of day five. Liu *et al.* [45] considered this fault as one of their study cases. They were able to detect the fault at 9:20h, giving only some hours notice to the operators, since the fault occurred at 17:44 on the same day. The last fault presented 4.03% faulty data points on the day of the fault and the prediction method was also able to detect the fault on that day. For this type of case, there is a need to resort to the model that takes into consideration the hourly error. Which allowed detecting abnormal behaviour on hours belonging to the tenth day before the failure. This model yielded only ten misclassified hours in a total of 14 616 hours of normal data, proving its ability to model the normal behaviour of the blades pith angle.



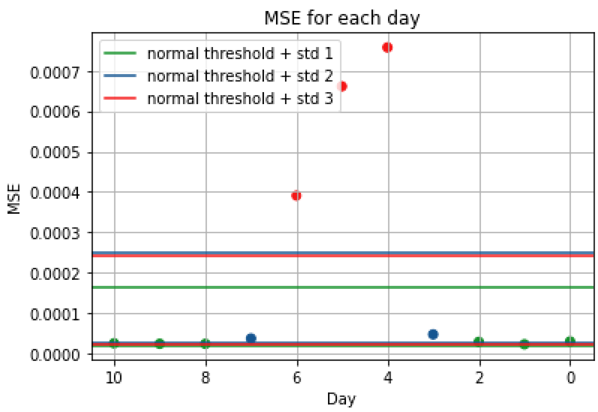
(a) First Fault Power Curve



(b) Second Fault Power Curve



(c) First Fault Prediction Method



(d) Second Fault Prediction Method

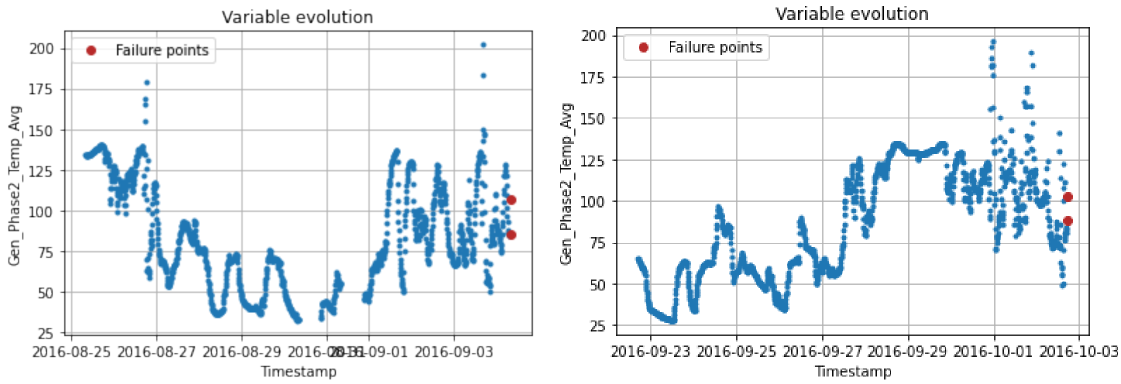
Figure 5.4: Turbine eleven pitch system faults

As mentioned in the previous section, to establish if the model for the normal behaviour of production can also be used, the procedure is to analyze the deviation of the points labelled in Section 4.2.5 to the power curve. These points are marked in red in Figure 5.4(a) and Figure 5.4(b), where the cases of the two faults of turbine eleven can be compared. The first one, Figure 5.4(a), presents no deviation and was not predicted by the model, as seen in Figure 5.4(c). The second one, Figure 5.4(b), presenting a deviation, shows abnormal behaviour on the fifth day before the failure using the model without levels. Using the model with the three different levels of mechanical stress, Figure 5.4(d), the results show that on days four, five and six before the failure, the turbine was working on level three and allowed a prediction of the fault. The first fault of turbine six and the fault on turbine seven presented deviations and were detected on the day of the failure. Therefore, using the hourly average error, leaded to a prediction of almost 24 hours and ten days, respectively. The second fault of turbine six was in concordance with the power curve and was not detected by the model of the previous section.

5.3.4 Normal Behaviour of Generator Temperature

Starting by comparing the obtained results in this thesis with four existent works in literature that modelled the normal behaviour of the three stator windings' temperature. Schlechtingen *et al.* [42] reached to a standard deviation error of 4.98°C for the first phase and 4.92°C for the remaining two phases. The obtained results in this thesis using K-NN were 0.68°C, 0.83°C and 0.72°C for each one of the phases, quite lower than the ones found in [42]. As for the daily average, the values were 1.90°C, 1.87°C and 1.36°C for [42] and 0.035°C, 0.036°C and 0.035°C for the developed model, showing great improvements. However, not being able to surpass the values obtained by Sun *et al.* [43], their values rounded 0.72°C for MAE and 1.23°C for MAPE, as the ones obtained in this thesis rounded 2.32°C and 3.94°C. By looking at the box plots from [43], as compared to the ones obtained using EDP's dataset, it is believed that the differences in errors might be due to a bigger range of stator windings' temperatures in EDP's dataset. Guo *et al.* [48] achieved a normalized RSME of 0.02 using a NN and 0.002 using NSET. Even though having achieved a normalized RMSE of 0.03 using K-NN, it surpassed the paper's [48] hourly range notice prediction, being able to predict the fault with days notice. Tautz-Weinert *et al.* [19] achieved an MAE of 2.61°C and an RSME of 4.00°C, values above the 2.33°C and 3.32°C, obtained in this thesis.

For this model, there are three faults, all of them in turbine six. By applying the model on the periods of normal data, **99.89%** of the data was considered in fact normal and zero days in a total 609 of normal days were identified as faulty. The first fault showed 12.03% of failure data points, concentrated on the day of the fault, three and four days before the fault. From day ten till day five before the fault, no detection was possible due to a turbine downtime. Therefore, the prediction confirmed the previous results, with an MSE above the threshold on days three and four, with no information from the previous days. For the second and third fault, only phase two of stator windings was monitored, since it was the one that presented abnormal behaviour, as seen in Figure 5.5. The red points represent the failure points on the day of the fault, already labelled in Section 5.1.5. It can be observed that there are over-temperature points on the day of the fault and on the antecedent days, reaching 200°C.



(a) Second Fault **(b)** Third Fault

Figure 5.5: Turbine six - generator stator winding temperature phase two

In the second fault, 9.61% detected faulty data points were focused on the day of the fault and eight days before the fault, which was confirmed by the prediction method. On day eight the production was on level two (mean daily production between 50% and 80% of the rated power), as seen in Figure 5.6, which might have caused mechanical stress and an aggravation of the future fault. On the other hand, the ambient temperature was uniform along the ten days, having no impact on causing this isolated day. Liu *et al.* [45], predicted this fault almost two months before it occurred, a prediction with such notice was not considered in this thesis.

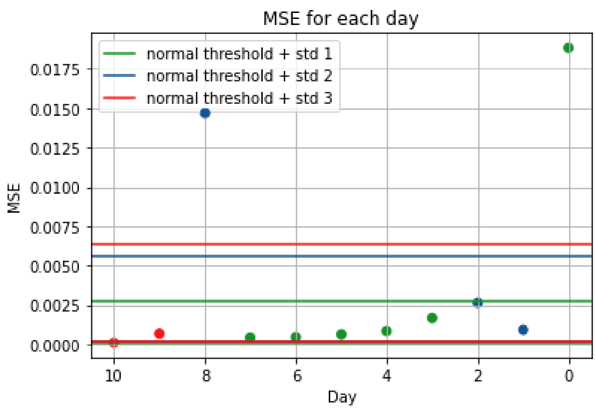


Figure 5.6: Fault prediction of the second fault of turbine six

Finally, on the third fault, 33.29% of the data from the ten days before the fault was considered faulty, more specifically, on the day of the fault and the day before. Once more, presenting the same results in the daily error method. Hence, it was possible to predict the three faults with four, eight and one day notice, respectively, giving the turbine' operators time to schedule maintenance before the fault. Using the model for the normal behaviour of production, the predictions were of eight and five days notice for the second and the third fault, respectively, as for the first one, no prediction was possible.

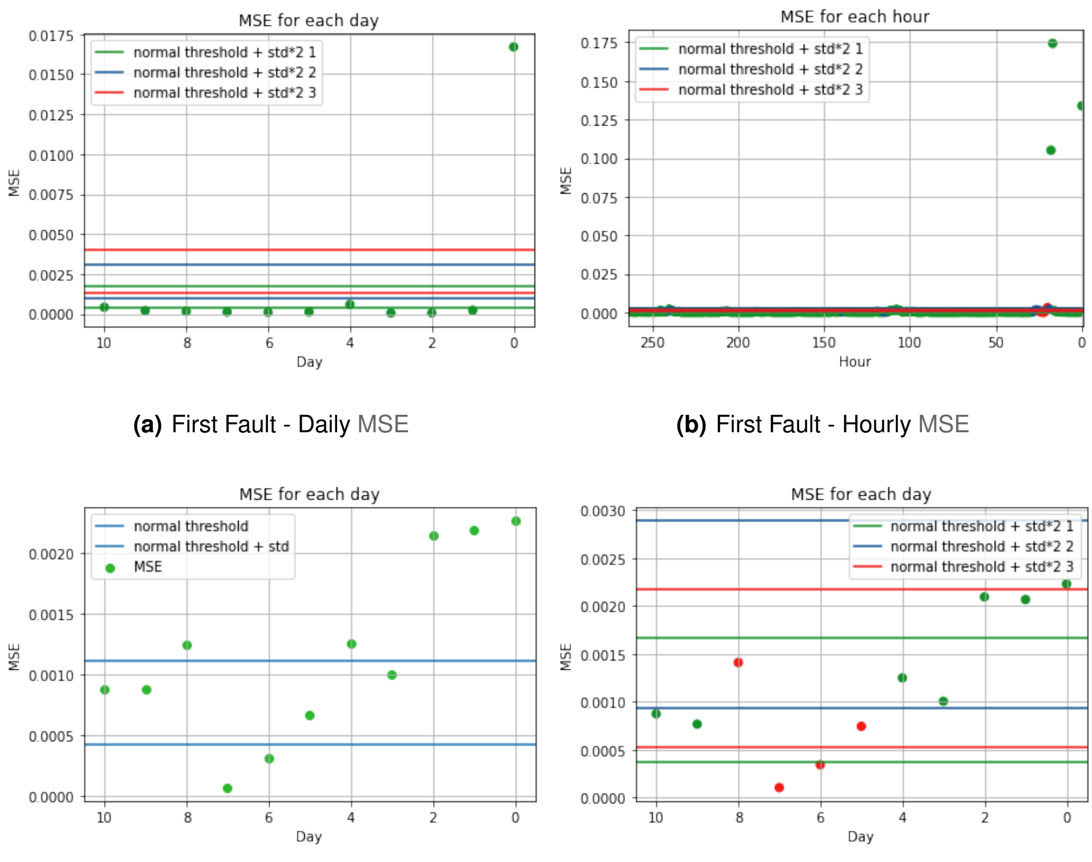
At last, this model was used to try to detect the fault caused by a replacement in the generator of turbine six, mentioned in the previous section. The detection method found 2.92% faulty data points concentrated on the day of the fault and four and eight days before its occurrence. An improvement from the model for the normal behaviour of production that had only disperse isolated points. However, this model was also not able to predict that fault using the daily average method.

5.3.5 Normal Behaviour of Generator Bearing Temperature

In agreement with the previous sections, the first step is going to be to compare the obtained results with the ones from existent literature. Schlechtingen *et al.* [42] obtained a standard deviation error of 1.44°C for the non-drive end and 1.86°C for the drive end. As for the results obtained using K-NN model in this thesis, the model was tested for two turbines that presented failures in this component, turbine seven and nine. For turbine seven the values were 0.28°C and 0.78°C for the non-drive end and drive end, respectively, as for turbine nine the values were 0.80°C and 0.64°C. In terms of the daily average, the values of 0.02°C for both ends of turbine seven and nine were lower than the 0.73°C and 0.75°C found by [42]. Tautz-Weinert *et al.* [19] reached an MAE of 1.99°C and an RMSE of 2.88°C, in the best of my understating, for the non-drive end. For these metrics the results obtained were quite different for the two turbines, 0.97°C and 1.84°C for the MAE and RMSE, respectively, of turbine seven and 6.63°C and 82.27°C for turbine nine. It is believed that the differences might be due to the feature selection being optimized for turbine seven. Therefore, only achieving better results than [19] for turbine seven. Compared with Sun *et al.* [43], both turbines failed to improve, believed to be due to the same reasons mentioned in Section 5.3.4. In terms of days notice, Schlechtingen *et al.* [49] was able to give an alarm to the turbine's operator with 30 days in advance using full signal reconstruction NN and with 50 days using an autoregressive NN. Kusiak and Verma [50] found that their method can predict over-temperature fault on average 1.5 h ahead of the fault occurrence. This thesis results are in between, exceeding the 1.5 hours from [50] but not reaching the 30 or 50 days from [49]. However, the prediction was done taking into consideration only ten days before the fault, being this thesis maximum notice; as future work, in Section 6.2, this subject will be discussed.

Beginning with turbine seven, the K-NN model considered **98.53%** data points as normal and zero days as faulty from the data corresponding to the normal operation periods. As for turbine nine, the values were **96.90%** and also zero days. The percentages were slightly lower than the ones from the previous models, indicating that the predictions made by this model can be less accurate. Therefore, at the end of this section, the results using this model will be compared with the ones obtained using the model for the normal behaviour of production.

Subsequently the analysis of the study cases was carried out, starting with the two cases from turbine seven. The first one only presented 0.97% of faulty data points, on the day of the fault, being the daily prediction also only possible on the day, as seen in Figure 5.7(a). Therefore, the hourly error was used to determine the time of the first error above the threshold. As seen in Figure 5.7(b), the most pronounced errors were on the day and the day before the fault. More specifically, the first point is at 16:40h from April 29th, having the fault occurred at 12:40h, April 30th. It is also possible to detect an isolated hourly average error above the threshold on days four and ten before the fault, however, if the goal was to avoid unnecessary checks, only a warning would be emitted to the turbine's operator that a small deviation has been detected. Moreover, the model considered 73 hours as faulty in a total of 14 616 hours of normal data, being a good indicator of its accuracy.



(a) First Fault - Daily MSE

(b) First Fault - Hourly MSE

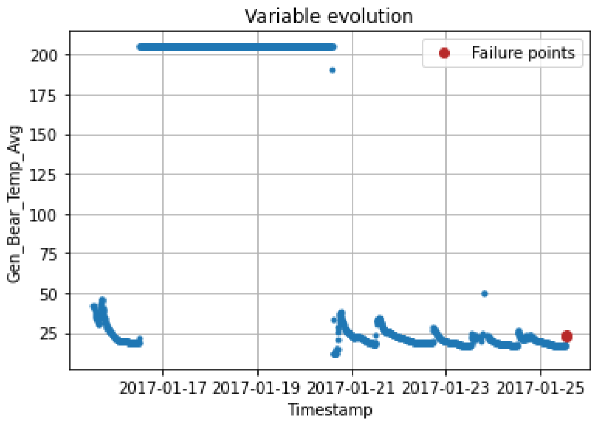
(c) Second Fault - Daily MSE

(d) Second Fault - Daily MSE with Levels

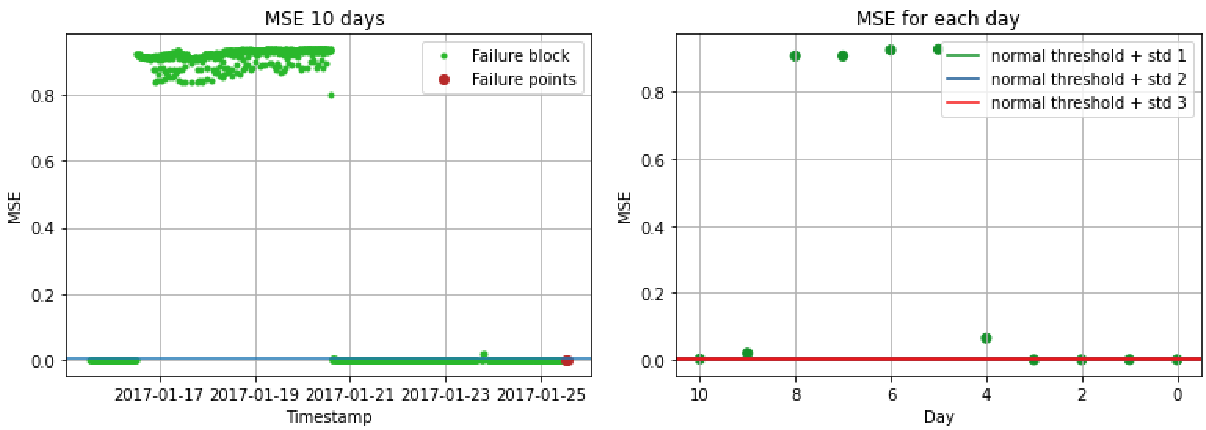
Figure 5.7: Turbine seven - generator bearing first and second fault

As for the second fault, the model that did not take into consideration the level of production, considered days eight, four, two and one before the fault as above the threshold, as seen in Figure 5.7(c). However, as observed in Figure 5.7(d), when using the model with levels, only from day two is possible to alarm the operator of the fault. In this case, the isolated point on day eight might have been caused by an increase in mechanical stress, causing the generator bearing temperature to increase. The ambient temperature was uniform during the ten days prior to the fault, therefore, had no impact on that isolated point.

Secondly, turbine nine presented four study cases with a generator bearing temperature fault. The first fault presented a daily average error from day six to four above the threshold. As for the second, only day eight before the fault was above the threshold, being on level three of mechanical stress (mean daily production above 80% of the rated power), most likely aggravating the temperature of the generator bearings. The third fault presented high daily MSE on days three to five before the fault.



(a) Monitored variable



(b) Detection method **(c)** Prediction method

Figure 5.8: Turbine nine - generator bearing fourth fault

Finally, the last case study reflected the behaviour of the monitored variable on the model's error. More specifically, the temperature of the non-drive end of the generator bearing presented abnormal behaviour from day nine to four before the fault, as see in Figure 5.8(a). In Figure 5.8(b), can be seen that the model detected faulty data points on the same days, and in Figure 5.8(c) the reflection on the daily average errors.

Finally, confirming the results using the model for the normal behaviour of production. For turbine seven, the first fault was also only detected on the day, as for the second fault, ten days notice were found, two more days than when using this model without levels. For turbine nine, the detection was only possible for the second fault, with ten days notice, once more, an improvement of two days.

5.3.6 Normal Behaviour of Gearbox

Five relevant existent works were found in the literature that focused on modelling the normal behaviour of the gearbox temperature. More specifically, the temperature of the gearbox bearing. Schlechtingen *et al.* [42] reached to a standard deviation error of 0.79°C and for the daily average error a value of 0.35°C. The values obtained in this thesis using K-NN model were 0.63°C and 0.008°C for turbine one, lower than the former. As for turbine nine the values were 0.84°C and 0.01°C, showing improvements only for the daily average error. As for Sun *et al.* [43], they reached an MAE of 0.11°C and a MAPE of 0.29°C. For both turbines the obtained results failed to show improvements, once again, believed to be due to the reasons mentioned in the previous sections. Orozco *et al.* [52] found an RMSE of 7.66°C, a value above the 0.64°C and 0.85°C obtained for turbine one and nine, respectively. Zhang *et al.* [51] were able to give a warning to the turbine operator with three months notice and an alarm with 10 days notice, with an RMSE of 0.2°C. Fu *et al.* [53] achieved an RMSE of 2.023°C. Therefore, the obtained results were able to improve in comparison to [53], yet failing to surpass [51].

For this type of fault, there are only two case studies, one for turbine one and another for turbine nine. The modelling of the normal behaviour of gearbox temperature correctly labelled **98.55%** of the normal period data as such for the first turbine and **98.33%** for the second. For both, only one day was considered faulty in a total of 609 days from normal data. For the fault of turbine one, a prediction with seven days notice was possible using the model without levels and only two days using the one with levels. As for turbine nine, for both cases, the daily average error was above the threshold as early as day nine before the fault.

Finally, comparing the predictions with the ones from the model of the normal behaviour of production. For turbine one, the data points close to the time of the failure presented a deviation from the power curve being suitable to be detected by the model. The model without levels presented an elevated MSE starting from day two before the failure, as for the one with levels, the prediction was extended till day three. A more reliable alarm could be given to the turbine operator merging the information provided by the two models. In that case only emitting an alarm three days before the fault, since both models agreed until that day. For turbine nine, no deviation from the power curve was found and no prediction was possible using the production model.

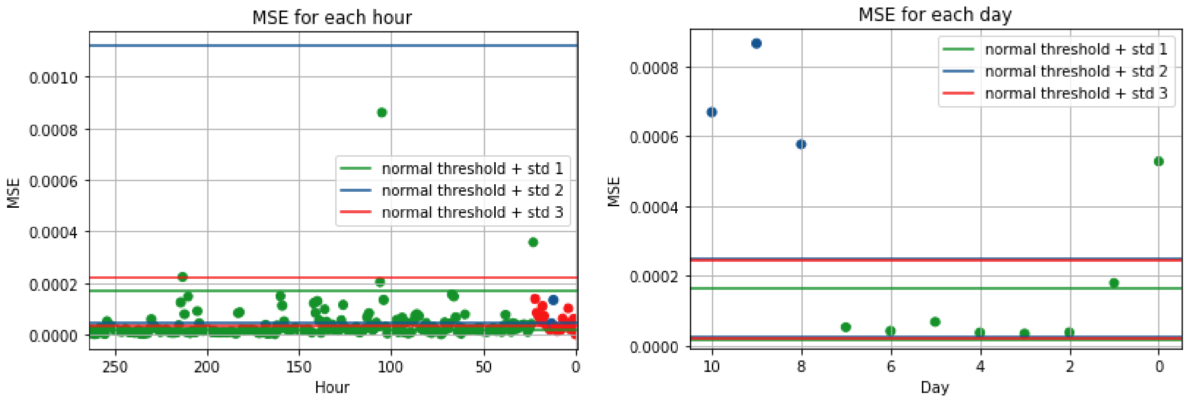
5.3.7 Normal Behaviour of Transformer

For the modelling of the normal behaviour of the transformer temperature, only one article that could be directly compared to this thesis' model was found. Schlechtingen *et al.* [42] obtained a standard deviation error of 3.59°C, 2.88°C and 3.25°C for each one of the three phases of the transformer. In comparison, this thesis results were 1.30°C, 0.98°C and 1.28°C for turbine one and 1.59°C, 1.22°C and 1.18°C for turbine seven. As for the daily average error, this thesis values were also inferiors to the ones obtained by [42].

In terms of accuracy of the developed model, K-NN correctly classified **98.70%** of the data points as normal for turbine one and **99.52%** for turbine seven. As for the days wrongly classified as faulty, turbine one found three days using the model without levels and nine with levels. As for turbine seven, the values were zero and two days, both considering a total of 609 days. Therefore, the predictions for turbine one may fail to be certain and should be compared to the ones from the production model.

There are three study cases related to transformer faults. Only the temperature of phase three was monitored for the two first cases, since was the one presenting abnormal behaviour. Starting with turbine one, the detection resulted in around 8% of failure data points on the day of the fault and a daily average error above the threshold also on the day of the fault. However, it was not possible to use the hourly error model since 323 hours were misclassified as faulty, compromising the accuracy of the model. Using the model for the normal behaviour of production for the whole wind farm, no prediction was possible. Therefore, there is a need to resort to the individual model specific for the normal behaviour of production of each turbine. This model identified an MSE above the threshold on the day of the fault, on level three of production, confirming the former result. As for the hourly model, 105 faulty hours were wrongly classified in a total of 14 616 hours corresponding to the normal periods of operation. Therefore, it can be assumed that the model is more trustworthy than the one developed in this section. As seen in Figure 5.9(a), only a few points were above the threshold, one on the day of the fault, two on the fourth day before the fault and one on the eighth day.

Finally, the last two study cases occurred in turbine seven, the first one being detected ten days before the failure. Liu *et al.* [45] also considered this case study and found that the most frequent isolated variable was the temperature of the third phase, sustaining the choice of using this variable as the target for the monitoring. Their prediction was with a month notice, exceeding the obtained in this thesis. The second case of this turbine was not detected using the developed model. However, the data points close to the failure presented deviations from the power curve, leading to experiment the model for the normal behaviour of production. The latter presented a value of MSE above the threshold from day ten to eight before the failure. As it is possible to see in Figure 5.9(b), all three days were inserted on level two of production, causing elevated mechanical stress that could have accentuated the presence of the fault.



(a) Hourly average error turbine one

(b) Detection method turbine seven

Figure 5.9: Transformer Faults - Turbine one and Turbine seven (second occurrence)

5.3.8 Concluding Remarks of Results

To summarize the results obtained for each of the models, Table 5.11, presents the 23 case studies. For each model, the third column reflects the percentage of data points correctly classified as normal among the data from normal periods of operation. Additionally, in the same column, for each fault, is also presented the percentage of data points considered abnormal regarding the data from the ten days before the fault. Excluding the first five rows, concerning the model for the normal behaviour of production, the remaining model's predictions (column four) were compared with the prediction of the former (column five). Finally, the last column presents the number of misclassified days and hours among the normal periods of operation, which contain a total of 609 days and 14 616 hours. For the faults where the hourly average was not necessary, the misclassified hours are not presented and are marked with a '-'. As for the cases where the values are equal to the ones from the row above, quotation marks (" ") were used.

All the models achieved a good percentage of correctly classified normal data points and almost no days misclassified as faults. In terms of fault detection, i.e., the percentage of failure data points, all faults were detected. Using only the model for the normal behaviour of production, it is possible to predict 15 of the 23 faults. Proving its generalization capability to detect and predict all types of faults that cause a decrease in production. Considering all the models, only one fault was not predicted. As for the ten cases where both the model of normal behaviour of production and of the model for the specific component were able to predict the fault, different conclusions can be taken. In three cases the models were in agreement, in four cases the normal behaviour of production decreased the prediction and in the other three, it increased it. If a more accurate alarm would to be given to the turbine operator, the prediction with less days notice would be chosen since both models agreed from that day on.

Component	Turbine	Normal / Failure Points [%]	Prediction Notice [day(s)]	Prediction Notice Production Model [day(s)]	Misclassified Days / Hours
Production	-	99 / -	-	-	0 / 99
	T6	- / 3	-	2	-
		- / 2	-	-	-
	T7	- / 3	-	1	-
	T11	- / 2	-	6	-
Hydraulic Group	T6	100 / 1	8	1	0 / -
		" " / 4	8	-	" "
	T7	100 / 8	10	10	0 / -
	T11	100 / 5	8	-	0 / 10
		" " / 4	10	6	" "
Generator	T6	100 / 12	4	-	0 / -
		" " / 10	8	8	" "
		" " / 33	1	5	" "
Generator Bearing	T7	99 / 1	1	1	0 / 73
		" " / 15	2	10	" "
	T9	97 / 10	6	-	0 / -
		" " / 4	8	10	" "
		" " / 9	5	-	" "
		" " / 42	9	-	" "
Gearbox	T1	99 / 19	7	3	1 / -
	T9	98 / 34	9	-	1 / -
Transformer	T1	99 / 8	-	8	9 / 323
	T7	100 / 35	10	3	2 / -
		" " / 4	-	10	" "

Table 5.11: Summary of results

6

Conclusion

Contents

6.1 Conclusions 73
6.2 Limitations and Future Work 74

6.1 Conclusions

In this thesis, the use of machine learning techniques to improve wind farms operation conditions monitoring was analysed and implemented. With this insight, in Chapter 3, some of the strengths presented in the existing literature were pointed out. More specifically, techniques regarding the pre-processing of the data (dealing with missing data, outliers, and data labelling), feature engineering, and finally, knowledge about the models that proved to be more accurate in predicting failures, helping the increase in the availability of the turbine generators.

The implemented methodology consisted of developing multiple accurate regression models capable of detecting significant deviations from the estimated values and indicative of a future fault. This process included the necessary pre-processing of the data from EDP's dataset. Multiple methods were tried out and in Section 5.1, the ones that led to the best results were selected. More specifically, the following procedures: (1) Missing Data: filling missing entries with the mean of the column for that variable; (2) Removal of Outliers: DBSCAN divided by wind speed bins of 1m/s and Z-score with three different sections of wind speed; and (3) Data Labelling: cross information of operational and status data to label as fault the points immediately close to the time of the fault. Additionally, in Section 5.2, feature selection was carried out, for each and every one of the developed regression models. By the use of cross-validation, multiple metrics were computed to evaluate the performance of the implemented methods and models enhancing the improvements when compared to the existent literature.

In Section 5.3, the obtained results from 23 case studies (faults) were presented followed by the formerly mentioned comparisons. In order to better substantiate the results while presenting them in a more visual and appealing way to the reader, multiple graphics were displayed. By looking at the results from the detection method, the majority of the faults presented concentrated faulty data points on the day of its occurrence or when the monitored variable started to present abnormal behaviour. These results supplemented with the high percentages of correctly classified data points among the data from normal periods of operation, prove the potential of the developed method. As for the method designed to predict faults, only one case study failed to be predicted by any of the developed models. The remaining were predicted on an average of seven days before the day of the fault and with a minimum of one day. This prediction allows an alarm to be given to the turbine operators to schedule maintenance in order to prevent or decrease the fault repercussions.

In comparison with existent literature, some of the models presented greater improvements than others. The models for the normal behaviour of production and for the hydraulic group were able to surpass all the considered existing works. As for the models for the normal behaviour of the generator temperature, generator bearing temperature and gearbox temperature failed to surpass some of the existing works. However, still presenting great potential taking into consideration the simplicity of the K-NN model. The last developed model, the modelling of the normal behaviour of the transformer, was only compared with one paper, which is not strong proof of its improvement but is still a promising result. At last, compared with A. Meyer [44] that obtained an RMSE of 0.51 for the multi-target normal behaviour model, all of our models obtained a lower RMSE. In fact, quite lower, more specifically, 0.006 for the normal behaviour of production, and ranging from 0.01 to 0.09 for the remaining models.

6.2 Limitations and Future Work

The first limitation found was the number of tried models and their complexity. Due to time constraints, only three different types of models were tried, being the initial idea to scale to more complex ones. By using more complex models, it is believed that the days notice could have been even extended to months. Therefore, as future work more models should be tried out and the test set broaden to include more than ten days before the fault. In agreement with the first limitation, all the models tried in existent literature should be included in order to provide a more accurate comparison between works.

Another point that could be improved is the smaller deviations in predictions being ignored in order to avoid false alarms, instead, they could be transformed into warnings. Therefore, in addition, a system that includes warnings prior to the alarms could also bring new possibilities for the developed WT condition monitoring.

At last, some of the papers that used supervised learning had problems with imbalanced data [24], having few classified faults and therefore, having issues with detecting fault patterns. Therefore, a possible future work would be to gather the best solutions in existent literature to deal with this problem, for example, data sampling. And then use them to implement a classification model to detect and predict WT faults. Additionally, those results could be compared with the ones obtained using unsupervised learning, i.e., regression models.

Bibliography

[1] W. Cao, Y. Xie, and Z. Tan, "Wind turbine generator technologies," *Advances in Wind Power*, vol. 1, no. 1, pp. 177–204, 2012.

[2] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe, and G. Ekemb, "Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges," *Energies*, vol. 7, no. 4, pp. 2595–2630, 2014.

[3] I. IRENA, "Future of wind: Deployment, investment, technology, grid integration and socio-economic aspects," 2019.

[4] M. Nick, G. Riahy, S. Hosseinian, and F. Fallahi, "Wind power optimal capacity allocation to remote areas taking into account transmission connection requirements," *IET renewable power generation*, vol. 5, no. 5, pp. 347–355, 2011.

[5] K. Wang, "Intelligent predictive maintenance (ipdm) system–industry 4.0 scenario," *WIT Transactions on Engineering Sciences*, vol. 113, pp. 259–268, 2016.

[6] K. Leahy, C. Gallagher, P. O'Donovan, and D. T. O'Sullivan, "Issues with data quality for wind turbine condition monitoring and reliability analyses," *Energies*, vol. 12, no. 2, p. 201, 2019.

[7] T. Ranchin, B. Furevik, M. Stette, H. Wensink, F. Van Hulle, C. Hasager, H. Johnsen, N. Fichaux, L. C. Christensen, P. B. Soerensen *et al.*, "Obtaining data for wind farm development and management: the eo-windfarm project," *Seatech week on" Technology for coastal environment monitoring", Brest*, 2004.

[8] A. R. Nunes, H. Morais, and A. Sardinha, "Use of learning mechanisms to improve the condition monitoring of wind turbine generators: A review," *Energies*, vol. 14, no. 21, p. 7129, 2021.

[9] J. Godwin and P. Matthews, "Classification and detection of wind turbine pitch faults through scada data analysis." *International journal of prognostics and health management.*, vol. 4, p. 016, July 2013. [Online]. Available: http://dro.dur.ac.uk/16546/

[10] J. M. P. Pérez, F. P. G. Márquez, A. Tobias, and M. Papaelias, “Wind turbine reliability analysis,” *Renewable and Sustainable Energy Reviews*, vol. 23, pp. 463–472, 2013.

[11] J. Carroll, A. McDonald, and D. McMillan, “Failure rate, repair time and unscheduled o&m cost analysis of offshore wind turbines,” *Wind Energy*, vol. 19, no. 6, pp. 1107–1119, 2016.

[12] F. Trojan and R. Marçal, “Proposal of maintenance-types classification to clarify maintenance concepts in production and operations management,” *Journal of Business Economics*, vol. 8, no. 7, pp. 560–572, 2017.

[13] Y. H. Ali, “Artificial intelligence application in machine condition monitoring and fault diagnosis,” *Artificial Intelligence: Emerging Trends and Applications*, p. 275, 2018.

[14] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic, “Machine learning methods for wind turbine condition monitoring: A review,” *Renewable Energy*, vol. 133, pp. 620–635, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S096014811831231X

[15] M. W. Berry, A. Mohamed, and B. W. Yap, *Supervised and unsupervised learning for data science*. Springer, 2019.

[16] E. Guresen and G. Kayakutlu, “Definition of artificial neural networks with comparison to other networks,” *Procedia Computer Science*, vol. 3, pp. 426–433, 2011.

[17] E. Lapira, D. Brisset, H. Davari, D. Siegel, and J. Lee, “Wind turbine performance assessment using multi-regime modeling approach,” *Renewable Energy*, vol. 45, p. 86–95, 09 2012.

[18] A. Marvuglia and A. Messineo, “Monitoring of wind farms’ power curves using machine learning techniques,” *Applied Energy*, vol. 98, p. 574–583, 10 2012.

[19] J. Tautz-Weinert and S. J. Watson, “Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection,” *Journal of Physics: Conference Series*, vol. 753, p. 072014, sep 2016. [Online]. Available: https://doi.org/10.1088/1742-6596/753/7/072014

[20] A. Kusiak and W. Li, “The prediction and diagnosis of wind turbine faults,” *Renewable energy*, vol. 36, no. 1, pp. 16–23, 2011.

[21] R. Ibrahim, J. Weinert, and S. Watson, “Neural networks for wind turbine fault detection via current signature analysis,” 2016.

[22] B. Tang, T. Song, F. Li, and L. Deng, “Fault diagnosis for a wind turbine transmission system based on manifold learning and shannon wavelet support vector machine,” *Renewable Energy*, vol. 62, pp. 1–9, 02 2014.

[23] K. Leahy, R. L. Hu, I. C. Konstantakopoulos, C. J. Spanos, A. M. Agogino, and D. T. O’Sullivan, “Diagnosing and predicting wind turbine faults from scada data using support vector machines,” *International Journal of Prognostics and Health Management*, vol. 9, no. 1, pp. 1–11, 2018.

[24] A. Verma and A. Kusiak, “Fault monitoring of wind turbine generator brushes: a data-mining approach,” *Journal of Solar Energy Engineering*, vol. 134, no. 2, 2012.

[25] M. Wilkinson, B. Darnell, T. Van Delft, and K. Harman, “Comparison of methods for wind turbine condition monitoring with scada data,” *IET Renewable Power Generation*, vol. 8, no. 4, pp. 390–397, 2014.

[26] M. S. A. Shafae, “Advancing the utility of manufacturing data for modeling, monitoring, and securing machining processes,” Ph.D. dissertation, Virginia Tech, 2018.

[27] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic, “Machine learning methods for wind turbine condition monitoring: A review,” *Renewable Energy*, vol. 133, pp. 620–635, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S096014811831231X

[28] P. Marti-Puig, A. Blanco-M, J. J. Cárdenas, J. Cusidó, and J. Solé-Casals, “Effects of the pre-processing algorithms in fault diagnosis of wind turbines,” *Environmental modelling & software*, vol. 110, pp. 119–128, 2018.

[29] E. Lapira, D. Brisset, H. Davari, D. Siegel, and J. Lee, “Wind turbine performance assessment using multi-regime modeling approach,” *Renewable Energy*, vol. 45, p. 86–95, 09 2012.

[30] W. Yang, R. Court, and J. Jiang, “Wind turbine condition monitoring by the approach of scada data analysis,” *Renewable Energy*, vol. 53, pp. 365–376, 2013.

[31] G. Jiang, H. He, P. Xie, and Y. Tang, “Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, 2017.

[32] M. Ringnér, “What is principal component analysis?” *Nature biotechnology*, vol. 26, no. 3, pp. 303–304, 2008.

[33] Y. Wang, X. Ma, and P. Qian, “Wind turbine fault detection and identification through pca-based optimal variable selection,” *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1627–1635, 2018.

[34] N. Zhou and L. Wang, “A modified t-test feature selection method and its application on the hapmap genotype data,” *Genomics, proteomics & bioinformatics*, vol. 5, no. 3-4, pp. 242–249, 2007.

[35] W. Zhang and X. Ma, “Simultaneous fault detection and sensor selection for condition monitoring of wind turbines,” *Energies*, vol. 9, no. 4, p. 280, 2016.

[36] M. Rajih, P. Comon, and R. A. Harshman, “Enhanced line search: A novel method to accelerate parafac,” *SIAM journal on matrix analysis and applications*, vol. 30, no. 3, pp. 1128–1147, 2008.

[37] L. Yang, H. Chen, Y. Ke, M. Li, L. Huang, and Y. Miao, “Multi-source and multi-fault condition monitoring based on parallel factor analysis and sequential probability ratio test,” 2021.

[38] P. Qian, X. Tian, J. Kanfoud, J. L. Y. Lee, and T.-H. Gan, “A novel condition monitoring method of wind turbines based on long short-term memory neural network,” *Energies*, vol. 12, no. 18, p. 3411, 2019.

[39] E. W. E. Association *et al.*, *Wind energy-the facts: a guide to the technology, economics and future of wind power*. Routledge, 2012.

[40] A. Marvuglia and A. Messineo, “Monitoring of wind farms’ power curves using machine learning techniques,” *Applied Energy*, vol. 98, p. 574–583, 10 2012.

[41] M. Schlechtingen, I. F. Santos, and S. Achiche, “Using data-mining approaches for wind turbine power curve monitoring: a comparative study,” *IEEE Transactions on Sustainable Energy*, vol. 4, no. 3, pp. 671–679, 2013.

[42] ——, “Wind turbine condition monitoring based on scada data using normal behavior models. part 1: System description,” *Applied Soft Computing*, vol. 13, no. 1, pp. 259–270, 2013.

[43] P. Sun, J. Li, C. Wang, and X. Lei, “A generalized model for wind turbine anomaly identification based on scada data,” *Applied Energy*, vol. 168, pp. 550–567, 2016.

[44] A. Meyer, “Multi-target normal behaviour models for wind farm condition monitoring,” *Applied Energy*, vol. 300, p. 117342, 2021.

[45] X. Liu, J. Du, and Z.-S. Ye, “A condition monitoring and fault isolation system for wind turbine based on scada data,” *IEEE Transactions on Industrial Informatics*, 2021.

[46] J. Gayo, “Reliawind project final report,” *Tech. Rep.*, 2011.

[47] C. Yang, Z. Qian, Y. Pei, and L. Wei, “A data-driven approach for condition monitoring of wind turbine pitch systems,” *Energies*, vol. 11, no. 8, p. 2142, 2018.

[48] P. Guo, D. Infield, and X. Yang, "Wind turbine generator condition-monitoring using temperature trend analysis," *IEEE Transactions on sustainable energy*, vol. 3, no. 1, pp. 124–133, 2011.

[49] M. Schlechtingen and I. F. Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mechanical systems and signal processing*, vol. 25, no. 5, pp. 1849–1875, 2011.

[50] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: A data-mining approach," *Renewable Energy*, vol. 48, pp. 110–116, 2012.

[51] Z.-Y. Zhang and K.-S. Wang, "Wind turbine fault detection based on scada data analysis using ann," *Advances in Manufacturing*, vol. 2, no. 1, pp. 70–78, 2014.

[52] R. Orozco, S. Sheng, and C. Phillips, "Diagnostic models for wind turbine gearbox components using scada time series data," in *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, 2018, pp. 1–9.

[53] J. Fu, J. Chu, P. Guo, and Z. Chen, "Condition monitoring of wind turbine gearbox bearing based on deep learning model," *Ieee Access*, vol. 7, pp. 57 078–57 087, 2019.

[54] P. Cross and X. Ma, "Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines," *International Journal of Automation and Computing*, vol. 12, no. 1, pp. 25–34, 2015.

[55] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, vol. 1, no. 1, pp. 73–79, 2011.

[56] V. Kumar and S. Minz, "Feature selection: a literature review," *SmartCR*, vol. 4, no. 3, pp. 211–229, 2014.

[57] A. C. de Barros Neiva, R. Linden, V. G. Guedes, C. L. S. Massa, and A. H. Macedo, "Development of an advanced power curve for performance monitoring of wind turbines," *Revista de Sistemas de Informação da FSMA*, no. 26, pp. 2–9, 2020.

[58] Y. Ding, N. Kumar, A. Prakash, A. E. Kio, X. Liu, L. Liu, and Q. Li, "A case study of space-time performance comparison of wind turbines on a wind farm," *Renewable Energy*, vol. 171, pp. 735–746, 2021.

A

Code of Project

In this section some snippets of code will be provided in order to clarify the implementation of the thesis.

Listing A.1: Code for dealing with missing data

```
""" Returns the N following entries for that column

Attributes:
    n (int): number of entries to be returned
    dfT (dataframe): dataframe containing only entries for a certain turbine
    idx (Timestamp): timestamp of the row that contains the missing value
    t (int): turbine's identifier
    col (string): name of the column that contains the missing value

Output:
        values (list): list of following N values for that column
"""
```

```

def getFollowingN(n,dfT,idx,t,col):
  count=0
  i=1
  values = []
  while count!=n:
    if (idx[0] + timedelta(minutes=10*i),t) in dfT.index: % check if a row with the following timestamp, for that turbine, exists
      value = dfT.loc[(idx[0] + timedelta(minutes=10*i),t)][col][0]
      values.append(value)
      count+=1
    i += 1

  return values

""" Replaces the missing values in the dataframe using the method received as argument

Attributes:
    dataset_t (dataframe): dataframe representing the signals file
    method (string): method to be used when selecting the new value for that entry
"""

def missing_data(dataset_t, method):
  turbines = dataset_t.index.unique(level='Turbine_ID')
  for t in turbines:
    dfT  = dataset_t.copy(deep=False).loc[(slice(None) , t),:] % selects only the entries that belong to the turbine t from the entire dataframe
    for idx,row in dfT.iterrows():
      if any(row.isnull()):
        nan_collumns = dfT.columns[np.where(row.isnull()==True)]
        for col in nan_collumns:
          values = getFollowingN(9,dfT,idx,t,col)
          if all(math.isnan(v) for v in values): % replace with a value that represents an error in the sensor
            dfT.loc[idx,col] = 123456789
            dataset_t.loc[idx,col] = 123456789
```

```
else: % not an error in sensor, find the estimated value
  new_value = None
  it = 1
  while new_value == None:
    if method == "random":  % replace with random value - Baseline
      rand = random.choice(dfT[col])
      if math.isnan(rand) == False:
          new_value = rand
    elif method == "last_value": % replace with last value - (1)
      if (idx[0] - timedelta(minutes=10*it),t) in dfT.index:
          if math.isnan(dfT.loc[(idx[0] - timedelta(minutes=10*it),t)][col][0]) == False:
            new_value = dfT.loc[(idx[0] - timedelta(minutes=10*it),t)][col][0]
      it += 1
    elif method == "mean":  % replace with column's mean - (2)
      new_value = np.mean(dfT[col])

    elif method == "new_model": % use another model to estimate the value - (3)
      dfT2 = dfT.copy()
      dfT2 =dfT2.dropna()
      X = dfT2.drop([col], axis = 1)
      scaler = MinMaxScaler().fit(X)
      X = scaler.transform(X)
      Y = dfT2[col]
      model = MLPRegressor()
      model.fit(X, Y)
      new_y = np.nan_to_num(dfT.loc[idx].drop([col], axis = 1))
      new_value = model.predict(scaler.transform(new_y))

  print("New value ",new_value)
  dfT.loc[idx,col] = new_value
  dataset_t.loc[idx,col] = new_value
```