UKRAINIAN CATHOLIC UNIVERSITY

BACHELOR THESIS

Recommendation System for Inspection of a Global Wind Turbine Fleet

Author:
Kateryna MAKAROVA

Supervisor:
Bohdan KULYK

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2021

Declaration of Authorship

I, Kateryna MAKAROVA, declare that this thesis titled, "Recommendation System for Inspection of a Global Wind Turbine Fleet" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Bachelor of Science

Recommendation System for Inspection of a Global Wind Turbine Fleet

by Kateryna MAKAROVA

Abstract

Using wind turbines as a means to create electric energy and other sources of renewable energy has been gaining popularity over the last few years. To keep them working safely and stably, maintenance should be done regularly.

With that, the need for affordable good quality inspections is rising, now more than ever. The supply of such services is catching up with the demand. However, unfortunately, due to a lack of experience and data, it is hard to decide, so the inspections are delayed, leading to severe consequences and money losses.

In this work, I will analyze the data available to me on this topic to create a recommendation system that will advise owners of wind turbines on how to maintain them best.

Acknowledgements

I owe my deepest gratitude to Iryna Kostyshyn for continuing support and invaluable help throughout the entire project (and life).

And, of course, my parents who have always been by my side through the long journey of getting higher education.

I also wish to thank the Ukrainian Catholic University for everything I have learned.

Contents

List of Figures

List of Abbreviations

WT	**W**ind **T**urbine
LEP	**L**eading **E**dge **P**rotection
Biax	**B**i-axially oriented fiberglass layers
LE	**L**eading **E**dge
UD	**U**nidirectional Fiberglass Layers
LPS	**L**ightning **P**rotection **s**ystem
TE	**T**railing **E**dge
VG	**V**ortex **G**enerator
PS	**P**ressure **S**ide
SS	**S**uction **S**ide

Dedicated to my dear grandmother

Chapter 1

Introduction

1.1 Motivation

Wind energy is one of the most rapidly growing renewable sources of energy today. This is mainly since it has a minor negative impact on the environment. It is a unique source as it can be produced in any quantities, and it is not dependant on fossils. Wind turbines are now found in increasingly remote areas worldwide, face a wide range of environmental pressures, and will have to operate for more than 120,000 hours over their lifetime. By the end of 2017, their total power had surpassed 540 GW. According to the Global Wind Energy Council, total global wind power could provide 17–19 percent of the world's electricity needs by 2020 [12].

To meet the growing demand, wind turbines are being installed all over the globe. Offshore wind is seen as a vital component of the sustainable energy supply. However, maintaining turbines is complicated because they are always subject to hourly or seasonal fluctuations in wind speed and direction. Unlike most industrial machines that operate under more or less static conditions, Turbines are subjected to random charge [1]. The maintenance of offshore wind turbine systems (OWTs) equipment is complicated by limited accessibility and severe failure effects, influenced by weather conditions [7]. The sizes of the blades are getting bigger and bigger, the power ratings are growing as well. New advanced technologies require higher standards of maintenance that have to be performed regularly.

Developing cost-effective inspection and maintenance programs for wind energy farms is a difficult task filled with risk due to the diversity of equipment and their corresponding damage mechanisms and failure modes, weather-dependent transportation conditions, unpredictable spare parts demand, insufficient space or limited access for maintenance and repair, and limited availability of resources. Many researchers and practitioners from different sectors of the wind energy industry, including manufacturers, component suppliers, maintenance contractors, and others, have been interested in maintenance optimization in recent times [10].

There are different methods of performing the inspections currently available on the market, such as manual inspections when a trained professional has to climb the blade and inspections performed by the drones. The photos from such inspections are later analyzed by qualified personnel and are given recommendations on necessary repairments. Ideally, regular checks should be performed once a year to prevent severe defects from going undetected and causing severe damage to the turbine or even the whole farm or surrounding objects. In reality, such inspections can be costly, and turbine owners might neglect to perform them in time. They can be pursued to do so by giving them the optimal way to complete the inspection, which does not lead to unnecessary costs.

1.2 Goal

As the size of the blades increases, the structural loads of the turbine become more dominant, causing increased stress on the turbine components, which can lead to early failure. Because of this, an important area of focus in wind energy is lowering production costs to give it a competitive edge over other alternative power sources.

In this work, I analyze historical data from multiple suppliers of maintenance services for a wind turbine to determine the correlation between different types of blades, sites, quality of inspections, and the time it takes to complete them. Once this data is presented, it is easy to calculate the costs for each new inspection.

The cost of energy produced by offshore wind is heavily influenced by maintenance costs. The cost of maintenance is primarily determined by the strategy used to conduct it [9]. In this work, I will propose a formula for calculating these costs to make the process of planning inspections easier.

Chapter 2

Background information and theory

In this chapter, I will present the information needed to understand the topic of this work better. Mainly I will talk about basic knowledge of WT energy, such as the structure of turbines and how the inspections are analyzed. I will also go into detail about the methods used for analyzing data and building models.

2.1 Blade Finding Categorization

Firstly, to understand the importance of the specific type of maintenance inspections described and analyzed in this work, one has to learn about the structure of wind turbines and the possible defects that can happen.

If we classified the damages to wind turbines by the cause of failure, there would be two of them: mechanical damages and non-mechanical damages. Ultimate and cyclic loadings are the leading causes of mechanical injury. Ultimate strength loadings (which cause gradual deterioration or sudden failure) commonly cause overstress and buckling, while cyclic loadings cause fatigue cracks. Non-mechanical damages are described as material degradation caused by non-mechanical or indirect mechanical actions such as erosion, flaking, lightning [14]. Depending on the types of defects, they can lead to effects of various severity. That is important as, depending on the severity, different actions need to be taken. So to better classify the defects, a five-level damage categorization scheme is applied and is based on a similar scheme used by WT designers, operators, and owners.

By severity ranking, each finding is ranked in accordance with general principles. Severity ranking can often be combined with an action plan, which shall be determined by blade type, maintenance strategy. Explanation of ranking and results of Findings is shown in Figure 2.1

2.2 Data Analysis and Exploration

When the goal and importance of the task are understood, the central part of this work is interpreting historical data that is available to make recommendations for future inspection. Before working on creating machine learning models, data analysis should be done - initial and one of the most human-centric parts of the data science process. The critical steps in a data science process are shown in Figure 2.2.

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making [13]. And the initial step of data analysis is data exploration.

Severity	Finding	Type of Findings	Possible results
0	A Finding of a cosmetic nature e.g. surface contamination.	If blades are contaminated	This is only for info.
1	A Finding in a component/feature that is adhered to the outer surface of the blade, such as the aerodynamic add-ons or PowerEdge™. For the avoidance of doubt, this is a Finding type categorization only and does not reflect an assessment of blade condition.	• Add-ons with Findings, cracks and/or missing • Observation is an area of interest • PowerEdge torn, loss of adhesion, pitting/penetration	• Can cause noise • Loss of power
2	A Finding where the irregularity is in the top coat or/and filler with no impact on laminate.	• LE protection is missing/damaged • Paint is missing/ damaged	Further degradation of the surface layers.
3	A Finding where the irregularity is in the top coat and filler resulting in laminate exposure and includes damage to the 1^{st} or 2^{nd} Biax layers	• If filler is damaged (laminate exposed) • Damage through topcoat/ filler (1^{st} or 2^{nd} Biax layers damaged)	• Further degradation and laminate damages over time • Can be first indication of Findings in underlying layers
4	A Finding of minor laminate damage (Structural – in the UD) in the blade	If fiber/laminate is damaged	• Structural failure • Water in structure • Frost damages
5	A Finding of major laminate damage (Structural – in the UD) in the blade, impacting blade structural integrity.	Critical Findings shall be investigated as a priority	Blade failure

FIGURE 2.1: Explanation of ranking and results of Findings
Source: Siemens Gamesa. Blade Finding Categorization – External [2]

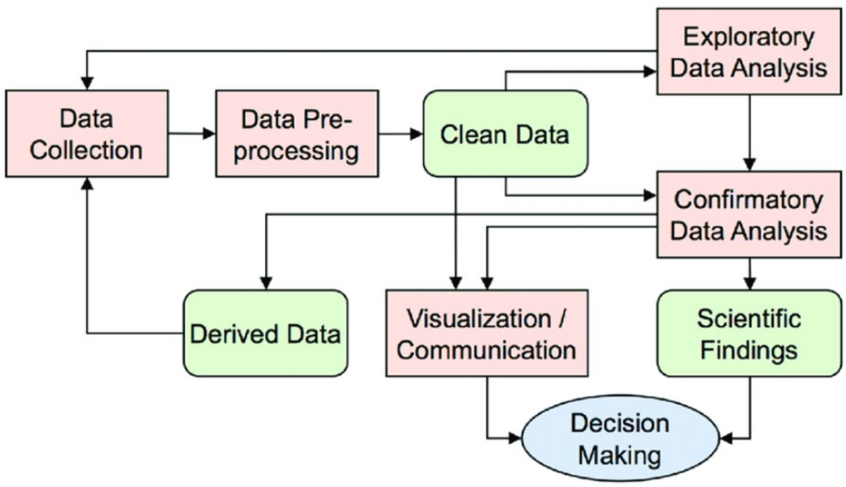

FIGURE 2.2: Key steps in a data science process
Source: https://www.researchgate.net

Data exploration is the first step of data analysis. Data analysts use data visualization and statistical techniques to depict dataset characterizations such as size, quantity, and accuracy to better understand the data's nature [4]. The purpose of data exploration is efficiently extracting knowledge from data even if we do not know exactly what we are looking for [6].

Data exploration may include manual and automated approaches, such as data visualizations, charts, and preliminary reports. This method facilitates more profound research by assisting in targeting prospective searches and initiating the process of eliminating unnecessary data points and search paths that might yield no results. More importantly, it aids in developing familiarity with existing data, which makes seeking better answers much more accessible. Visualization is often used in data discovery because it provides a more concise view of data sets than looking at thousands of unique numbers or names.

The manual and automated aspects of any data exploration look at opposite sides of the same coin. Manual research allows users to become more acquainted with knowledge and can reveal large patterns. These approaches are often unstructured by design, allowing users to analyze the entire collection without bias. On the other hand, automated tools excel at weeding out irrelevant data points, reorganizing data into easier-to-understand collections, and scrubbing data sets to make their results more critical.

You will start discovering similarities, trends and determining whether a particular directive is worth exploring or if the knowledge is less available by taking the time to do an honest exploration of the data using visualization software. Data exploration can also aid in the reduction of work time and the discovery of more valuable and actionable ideas from the outset, and the presentation of simple paths to better analysis.

2.2.1 Data exploration in Machine Learning

The following are data discovery steps to take before constructing a machine learning model:

- Identification of variables: describe each variable and its role in the dataset.
- Univariate analysis: create box plots or histograms for each variable separately for continuous variables; build bar charts to display the frequencies for categorical variables.
- Bi-variable analysis: Determine the relationship between variables using visualization methods
- Detect and treat missing values
- Detect and treat outliers

This allows making unexpected data findings. Also, promoting a comprehensive understanding of data as a foundation for effective and efficient data science projects.

2.3 Regression in Machine Learning

Regression analysis [8] is a set of machine learning methods for predicting a continuous outcome variable (y) based on the values of one or more predictor variables

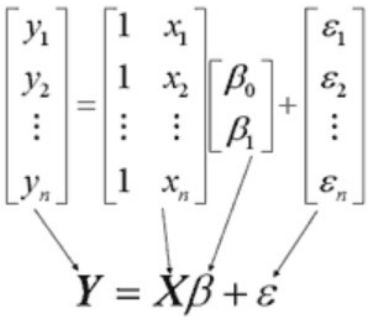

FIGURE 2.3: Linear regression function in matrix notation
Source: https://www.stat.psu.edu

(x). A regression model aims to create a mathematical equation that defines y as a function of x variables in a nutshell. Then, using new values for the predictor variables, this equation can be used to predict the outcome (y) (x). The most basic and widely used technique for predicting a continuous variable is linear regression. It implies that the outcome and predictor variables have a linear relationship. The linear regression equation can be written as

$$y = b_0 + b * x + e$$

where:

- **b0** is the intercept,
- **b** is the regression weight or coefficient associated with the predictor variable x,
- **e** is the residual error

Linear regression can also be presented in the matrix format. For more complex cases it makes it clearer. The formula is shown in Figure 2.3.

2.4 Linear ML algorithms

2.4.1 Linear Regression

It is a standard algorithm that can be found in the Linear Regression class. One or more output variables are predicted using a single input variable (the significant one), assuming that the input variables are uncorrelated. There can be a loss in output unless there is an exact line connecting the dependent and independent variables, generally calculated as the square of the difference between the expected and actual output, i.e., the loss function.

Multiple linear regression is when you use more than one independent variable to get the results. The drawback of this type of model is that it assumes a linear relationship between the given function and the output. Linear Regression model is shown in Figure 2.4.

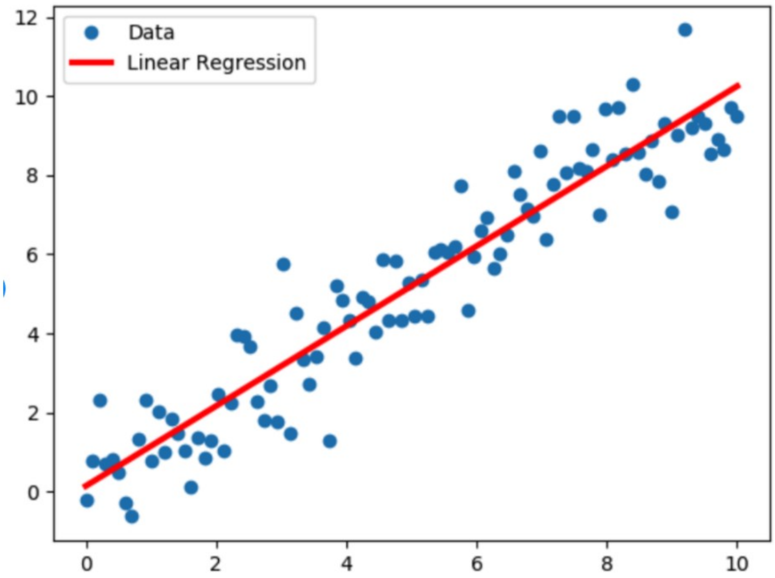

FIGURE 2.4: Linear Regression model illustration
Source: https://www.researchgate.net

2.5 Non-Linear ML algorithms

2.5.1 Decision Tree Regression

By dividing a data set into smaller and smaller subsets, it creates a tree with decision nodes and leaf nodes. The goal here is to plot a value for each new data point that connects the two sides of the problem. The parameters and algorithm decide how the split is carried out, and the split is completed when the minimum amount of data to be added has been reached. Decision trees often produce good results, but any slight change in data causes the entire structure to change, making the models unstable. Basic structure of a decision tree is shown in Figure 2.5.

A decision tree's basic structure. Recursion is used to build all decision trees. Different assumptions, such as normal distribution, are not used in decision trees, and collinearity or association between explanatory variables can also be ignored.

2.5.2 Random Forest

Random forests [3] are a collection of tree predictors in which the values of a random vector sampled independently and with the same distribution for all trees in the forest are used to predict the behavior of each tree. If the number of trees in a forest grows larger, the generalization error converges a.s. to a limit. The intensity of individual trees in the forest and their association determine the generalization error of a forest of tree classifiers.When a random set of features is used to separate each node, the error rates are comparable to Adaboost [5], but they are more resilient to noise. Internal calculations are used to display the answer to increasing the number of features used in the splitting by monitoring error, strength, and correlation. Internal calculations are also used to determine the value of variables. These concepts can also be applied to regression. Schematic of random forest algorithm is shown in Figure 2.6.

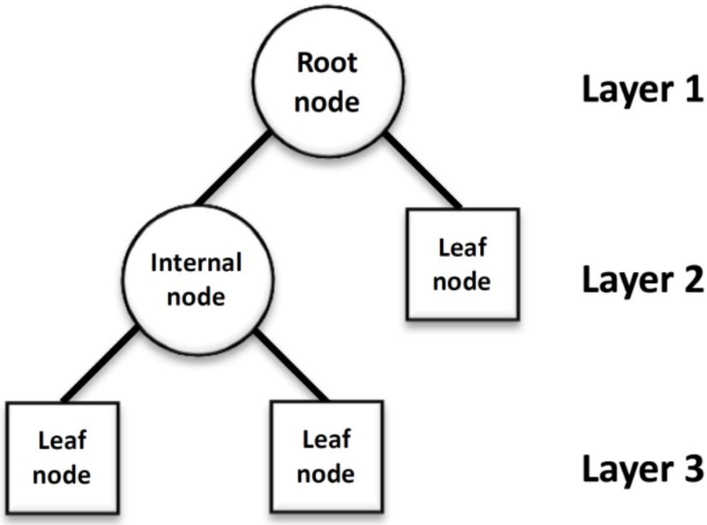

FIGURE 2.5: Basic structure of a decision tree
Source: https://www.researchgate.net

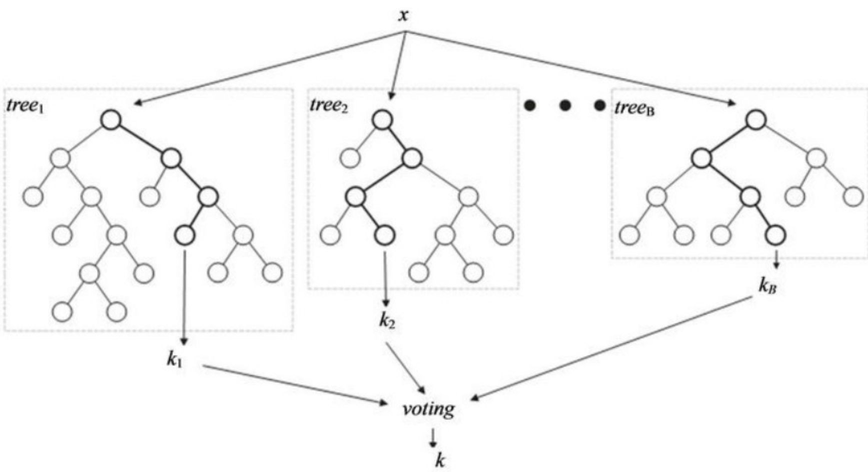

FIGURE 2.6: Schematic of random forest algorithm
Source: https://www.researchgate.net

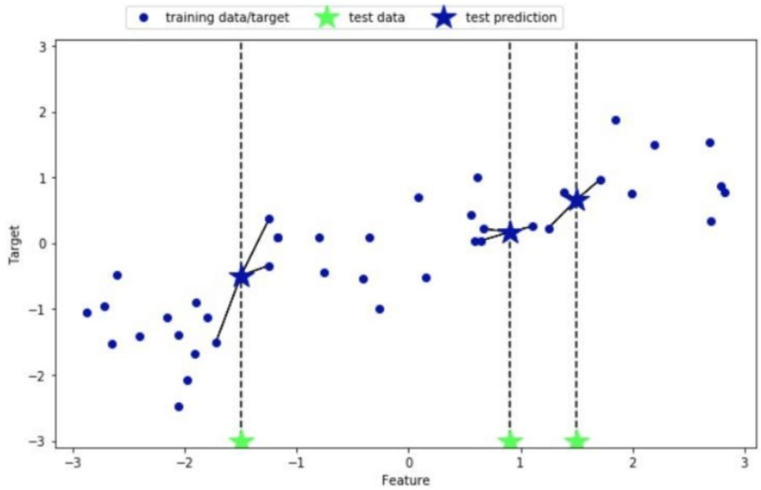

FIGURE 2.7: Example of KNN for Regression
Source: https://www.researchgate.net

2.5.3 K Nearest Neighbors(KNN model)

KNN regression is a non-parametric method that approximates the relationship between independent variables and continuous outcomes by averaging observations in the same neighborhood intuitively.

It can be imported from the KNearestNeighbors class. These are straightforward and quick to implement. The K Nearest Neighbors help find the k most similar instances in training set for an input inserted in the data set. The value for that input is either the average or the median of the neighbors. Example of KNN for Regression is shown in Figure 2.7.

Chapter 3

Related Works

In this chapter, I talk about some approaches on this topic that set requirements for the project and some that propose different ways of solving the task of performing inspections for the wind turbines.

3.1 Wind Turbine Reliability: Understanding and Minimizing Wind Turbine Operation and Maintenance Costs

When faced with the task of optimization the costs of the wind turbine farm maintenance, it is essential how you calculate them and which actions seem more important for achieving this goal [11]. Different farm owners may choose different strategies, but there is some research available on this topic.

The wind energy research community is tightly intertwined with the commercial side of the energy business. So to make their research more applicable, the cost of energy (COE) is a crucial project evaluation metric. This metric accounts for both predictable and unpredictable events. In the first category, we can put initial capital investment and scheduled maintenance and operating expenses—the second, costs associated with component failures. Unanticipated failures (mainly serial failures) can have a significant impact on the economics of a project. To get a clear picture, here is the calculation method for a wind turbine system that has been adopted by the Department of Energy in the Low-Speed Wind Turbine (LWST) program.

The formula for calculating COE is shown in Figure 3.1.

.

$$COE = \frac{ICC * FCR + LRC}{AEP_{NET}} + O\&M$$

$$AEP_{NET} = AEP_{GROSS} * Availability * (1 - Loss)$$

COE	Cost of Energy ($/kWh)
ICC	Initial Capital Cost ($)
FCR	Fixed Charge Rate (%/year)
LRC	Levelized Replacement Cost ($/year)
O&M	Operations and Maintenance Costs ($/kWh)
AEP	Annual Energy Production (kWh/year)

FIGURE 3.1: COE Formula Explained

Here are the ways how COE is affected by the unscheduled maintenance:

- AEP is affected by equipment reliability through turbine downtime associated with both scheduled and unscheduled maintenance.
- O+M consists of both scheduled (preventive) and unscheduled (repair) maintenance costs, including expenditures for replacement parts, consumables, manpower and equipment. OM costs can account for 10 – 20 percent of the total COE for a wind project.
- LRC costs are associated with major overhauls and component replacements over the life of a wind turbine.

And here are presented the recommendations for cost reduction:
1. Improving System Reliability

- Identify Critical Components
- Characterize Failure Modes
- Determine the Root Cause

2. Reducing Maintenance Costs

- Develop Logistics Plan
- Identify Opportunities for Redundancy
- Improve Training
- Improve Maintainability
- Implement Condition Monitoring

3.2 An Approach for Condition-Based Maintenance Optimization Applied to Wind Turbine Blades

Unexpected failure incidents and low availability are significant problems for wind-power operators to overcome. Uncertainties about the economic returns on wind projects may limit the rate of growth required to meet goals set by the European Wind Energy Association (180 GW in 2020) and the U.S. Department of Energy (more than 300 GW by 2030). Although optimizing maintenance techniques and making maintenance decisions has the potential to cut operational costs drastically, this topic has received little attention.

This study aims to offer a method for optimizing the maintenance of components whose deterioration can be categorized based on the severity of the damage. Different condition-based maintenance procedures, such as visual inspection, inspection with condition monitoring, or an online condition-monitoring system that can continually monitor the status of the component, can be used to maintain these components. The approach for estimating maintenance costs suggested in this paper is based on Monte Carlo simulation. The method is being used to optimize the maintenance of wind turbine blades.

.
.
.

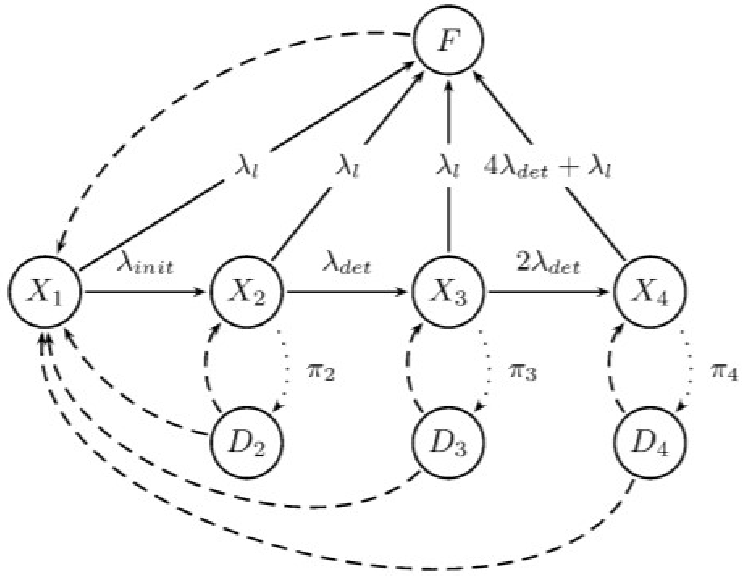

FIGURE 3.2: Proposed deterioration and maintenance model for the example. X_i: deterioration states; F: failure state; D_i: decision state. *Fullarrow*: deterioration transition; *dottedarrow*: inspection transition; *dashedarrows*: maintenance transition. Maintenance decisions occur at inspection or at deterioration state transition for online condition-monitoring system if the deterioration state is observed.

Parameters in the Example:
λ_s Failure rate for sudden failures.
λ_{init} Crack initiation rate.
T_{crack} Average time from crack initiation to failure.
λ_{det} Deterioration rates in the case study
.

Deterioration Model

It is expected that the condition of the blade is classified after each inspection based on the findings of the condition-monitoring analysis. The proposed classification is Good, Minor degradation, Advanced degradation, Major degradation, or Failure.

A continuous time Markov chain is used to represent the deteriorating process. The proposed deterioration classification for blades has five states, with X_1 indicating a Good state, X_2 - minor degradation state, and so on, with F and representing the failure state.

$$X = \{X_1, X_2, X_3, X_n - 1, F\}$$

denotes the deterioration state space and is used to index the states. To model possible unexpected failures, a direct transition to the failure state is used. The model is shown in Figure 3.2.

Maintenance Model

The horizon of the maintenance model assumed in this paper is finite. The model is enhanced to allow for maintenance evaluations based on online condition monitoring. If the online condition-monitoring system detects deterioration, an inspection is performed, according to this maintenance method. The inspection is required to identify the deterioration state and to determine whether or not maintenance is required.

Inspection maintenance strategies involve making maintenance decisions at predetermined intervals based on the degree of deterioration at the time of inspection. Between two inspections or until failure, the deterioration follows the Markov chain. Furthermore, the state and the next inspection date are updated based on the current state's maintenance decision. If maintenance is performed, the system is in an "as good as new" condition after maintenance. In the event of a failure, the component will be replaced.

Evaluation Method

For a fixed set of decision variables, the Monte Carlo simulation is used to generate scenarios and estimate maintenance costs. Maintenance decisions based on deterioration state for each maintenance approach and inspection interval for inspection-based maintenance are among the decision variables.

Implementation and results

For inspection-based maintenance strategies and maintenance based on online condition monitoring, a simulation method was proposed to evaluate expected life cycle maintenance costs.

Matlab was used to implement the model. For each set of input parameters for one maintenance approach, 100 000 simulations were run to evaluate the expected life cycle maintenance cost. The number of simulations was determined experimentally for the coefficient of variation of the results to be lower than 1 percent. The method has the disadvantage of being time-consuming because it requires a large number of simulations to produce reliable findings.

The results suggest that the best inspection interval for a condition-monitoring methodology was 6 months and 3 months for visual examination in the most basic scenario. Compared to a yearly visual examination, examinations with a condition-monitoring system or an online condition-monitoring system would be beneficial.

Chapter 4

Data

To perform the analysis in this work, I have been given access to data of one of the companies performing the reviews of inspections of WT. The data consist of several parts, which were analyzed separately:

- Pictures of the blades taken during the inspections (with metadata)
- Validation data for the pictures (validation scores that represent the quality of the photos)
- Logs of server time needed to process the photos
- Logs of annotator time (manual work of detecting Findings)

For the most part, I have worked with the images of blades. To perform fundamental analysis, I had to create a module to extract EXIF data which contained crucial data, such as timestamps of the pictures, which allowed me to calculate the time needed to perform the inspection by certain suppliers.

For the actual analysis, I have transformed the data into a data frame that still had to be processed for the analysis to be possible. The initial data had 1190335 unique pictures.

Chapter 5

Experiments

5.1 Data Exploration

To better understand the data that I had, the first step was to present it in a form that could be analyzed.

I have tried several libraries to extract EXIF data, such as Pillow, PyExifTool, and ExifRead. After some comparison, Pillow proved to work best for the task at hand.

I have created a python module that extracted EXIF data and, using the unique ID, combined it with other information available for each image, such as validation scores. The scores are from 1 to 5 and are given based on the quality of the picture. The data frame that I have created using all of the data is 1190335 rows x 24 columns. Some of the columns are used for internal identification, but some are useful data for later. The data is explained in Figure 5.1.

As a part of the data exploration step, I have created multiple visualizations (Figures 5.2, 5.3, 5.4, 5.5, 5.6 and 5.7), which I will present here. These can be used with different parameters to filter by site location, inspection type, site type, supplier, or time of the inspection. The visualizations that use duration can be chosen to use sequence, turbine, or inspection duration.

After closely inspecting the data, I have made some assumptions. It seems that while some suppliers have higher scores and/or shorter inspection duration, both of which facts lead to lower costs, there is not the correlation between scores and duration by themselves.

5.2 Formula for price calculation

To make predictions about optimal ways to perform an inspection, a formula for price calculation needs to be proposed. This formula is based on my knowledge of the whole cycle that goes into performing them.

.

Price =
Price per turbine * number of turbines +
Turbine duration * Idle time cost +
Time between turbines * Idle time cost +
Distance to site * Mobility costs +
Vesselprice

Column name	Description
inspection_type	How the inspection was performed (ex: from ground or using different types of drones)
site_type	Location of WT: onshore or offshore
blade_length	How long is the blade of WT (meters)
supplier	Name of the supplier that performed the inspection (the names in this work has been changed)
validation_score	Score of the picture quality (1 - 5)
image_timestamp	Image timestamp
focal_length	Focal length of each pic. This comes from EXIF data as well.
distance_to_blade	How far away each picture is from the beginning of the blade.
start_timestamp	Timestamp of the first picture in sequence.
sequence_duration	Duration of sequence (start_timestamp to last picture) in seconds
turbine_duration	Duration of inspecting the whole turbine (in seconds)
inspection_duration	Total inspection duration (in seconds)

FIGURE 5.1: Data frame columns explained

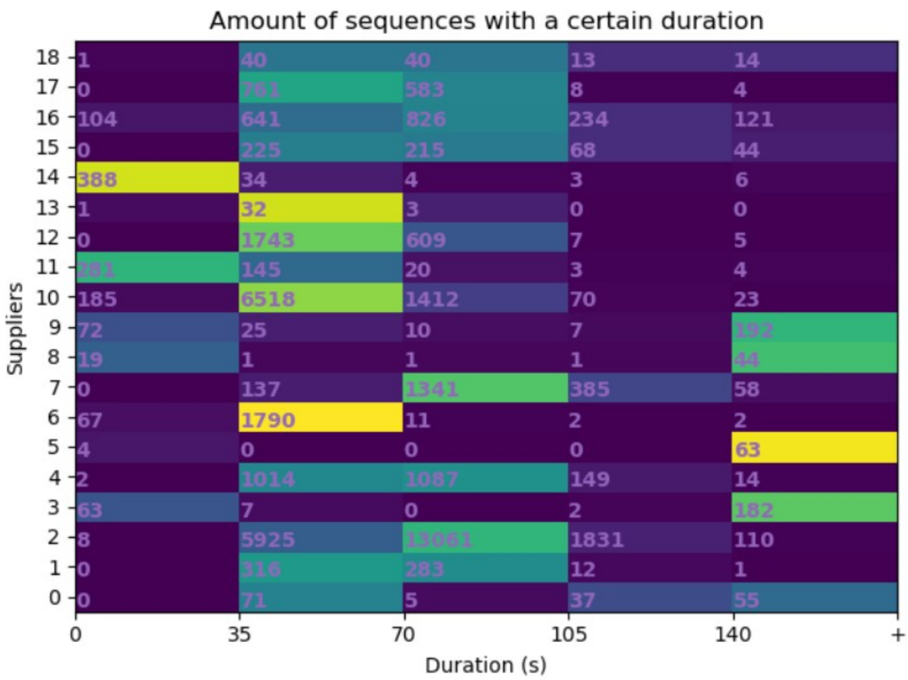

FIGURE 5.2: Heatmap showing the amount of sequences with certain duration by supplier

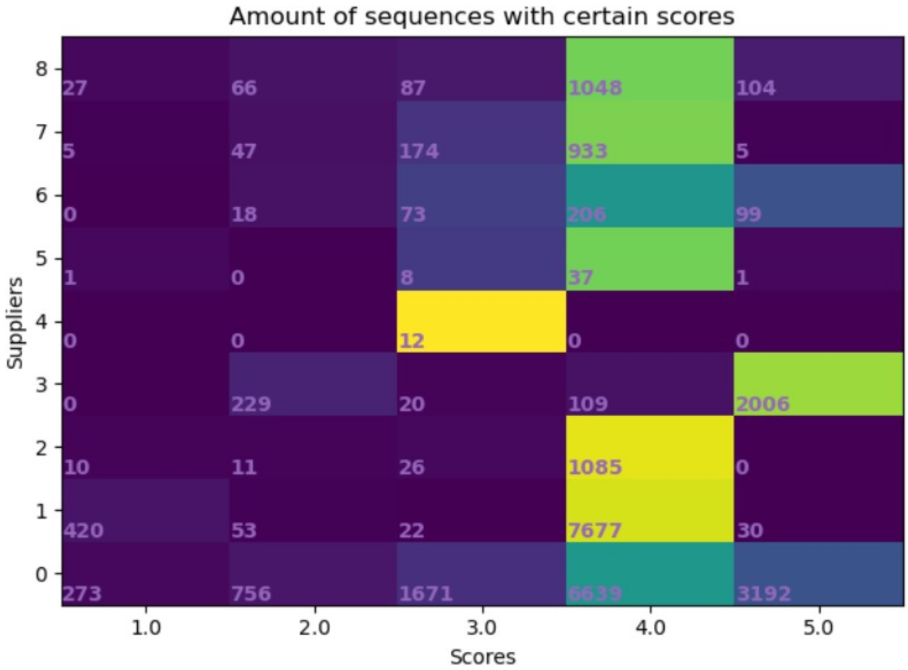

FIGURE 5.3: Heatmap showing the amount of sequences with certain scores by supplier

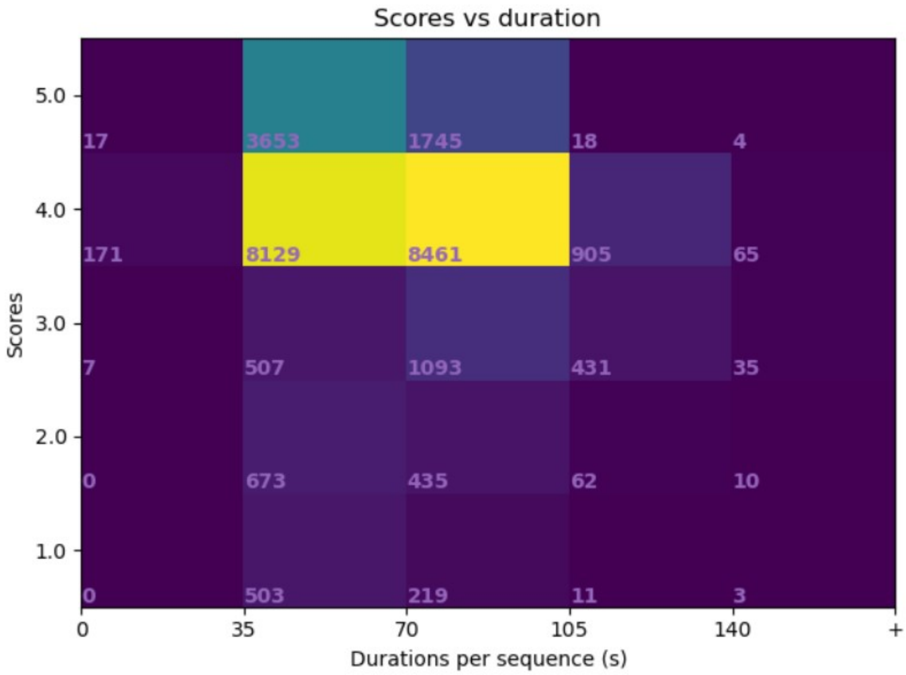

FIGURE 5.4: Heatmap showing the amount of sequences with certain scores and duration

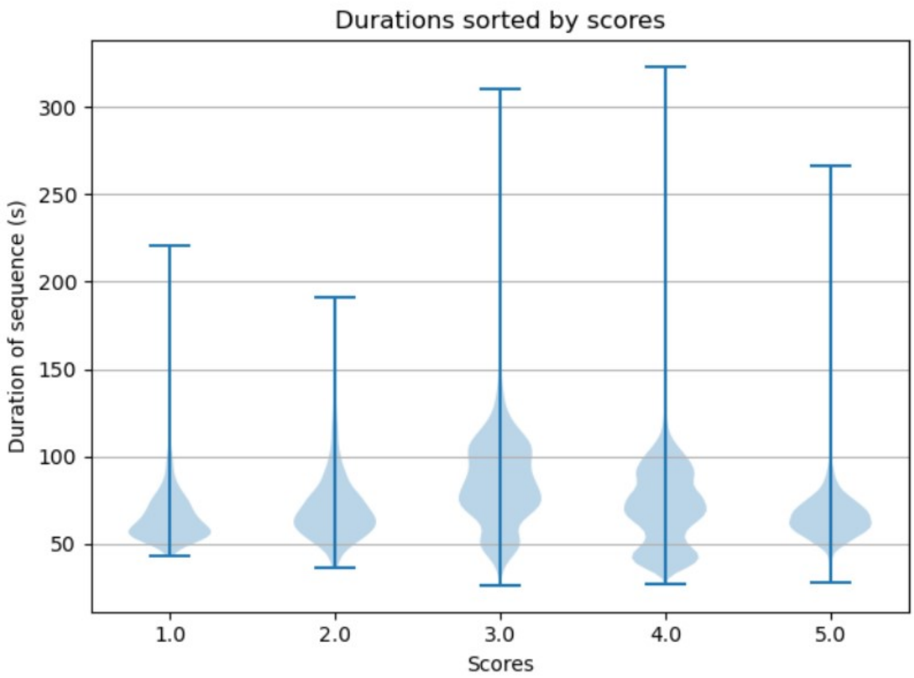

FIGURE 5.5: Violin plot showing the amount of sequences for certain scores

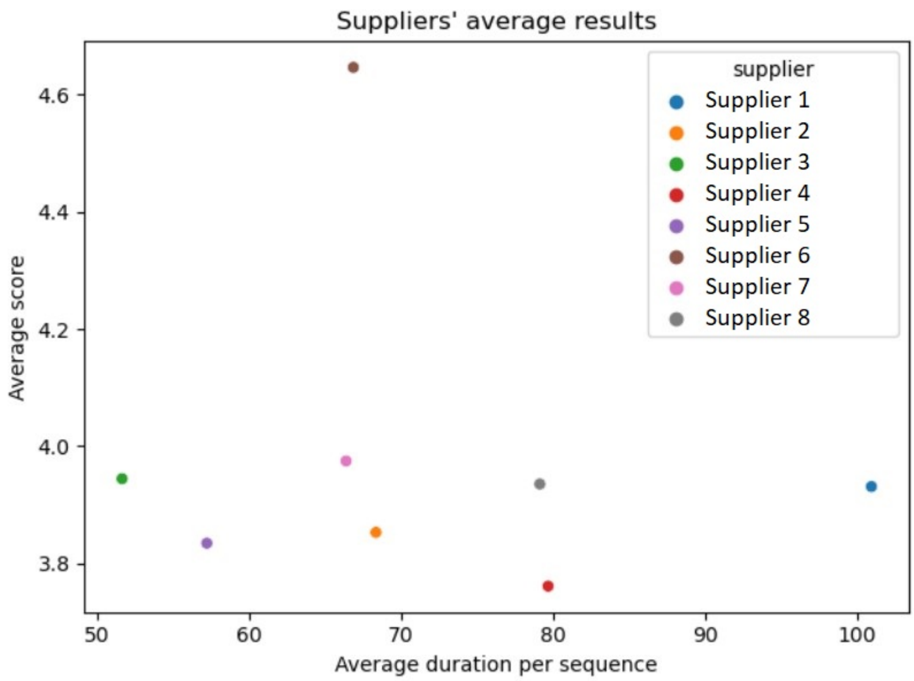

FIGURE 5.6: Scatter plot showing the average results for each supplier

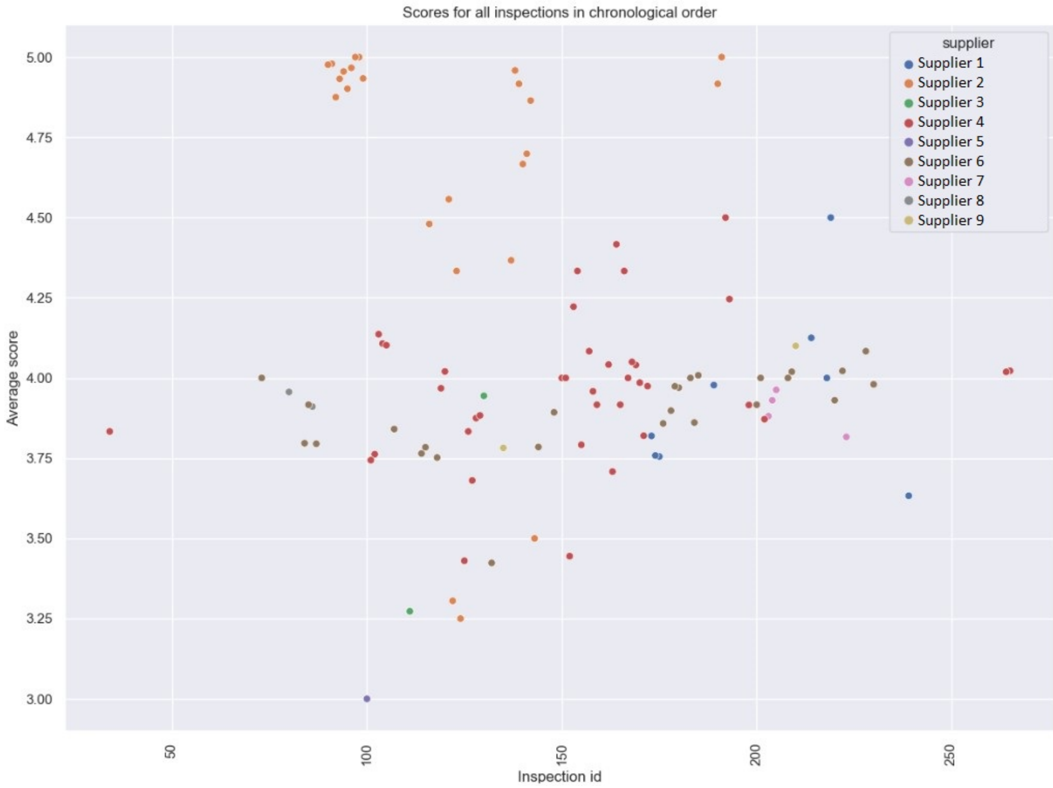

FIGURE 5.7: Scatter plot showing all results for each supplier sorted in chronological order

5.3 Data Preparation

Before building the models, the data first needs to be cleaned. Although the size of the dataset was initially pretty impressive, after a close look, it is easy to see that there are a lot of outliers and other problems.

I had to drop all of the inspections for suppliers that only had performed one or two. Also, if suppliers typically perform only onshore inspections, the data for the offshore ones could not be used. I have also applied some of the assumptions I had about the importance of the available features and dropped unnecessary columns.

After performing all of these, the final size of the data frame was 40568 rows x 7 columns, which is significantly less than at the beginning. The amount of data affected the models' precision, which will be evident in the next part.

5.4 Predicting Inspection Duration

I have used several methods to make the predictions needed for the calculations of the price using the formula that was proposed earlier. The following have produced the best results:

- Linear Regression
- Random Forest Regressor
- Decision Tree Regressor
- KNeighbors Regressor

Model	Score
Linear Regression	0.35
Random Forest Regressor	**0.45**
Decision Tree Regressor	0.44
KNeighbors Regressor	0.34

FIGURE 5.8: Scores of all models

Initially, before cleaning the data sufficiently, I have gotten scores as low as 0.04. This led me to think that data quality was not good enough, or maybe there was no correlation for the prediction to be possible. But after additional cleaning and then tuning the models, I have gotten the final scores that were way higher. The results are shown in Figure 5.8.

As you can see, Random Forest Regressor works best here, with Decision Tree Regressor following it closely.

Unfortunately, these results are still not good enough for this work.

Chapter 6

Conclusions

The field of alternative energy is growing at a rapid pace. There are numerous works on ways to optimize the production of energy. In this work, I have analyzed the historical data on WT inspections to find if such inspections' duration could be predicted. Having that information would allow using the formula for price calculation so that the owners of WT parks can make more informed decisions about upcoming inspections, which would hopefully motivate them more and lower the risks of running the sites unchecked.

While working on this project, I have come to a conclusion that, at this time, such predictions cannot be made. Hopefully, it is an issue that can be fixed with more data which will be available at some point.

The exploration of the data is useful, but it can show the differences between suppliers available on the market. It also shows a correlation between different features or lack thereof. The data can be analyzed further if needed for other purposes.

Bibliography

[1] Tahar Benmessaoud, Kamal Mohammedi, and Y Smaili. "Influence of maintenance on the performance of a wind farm". In: *Przeglad Elektrotechniczny* 89 (Jan. 2013).

[2] *Blade Finding Categorization – External*.

[3] Adele Cutler, David Cutler, and John Stevens. "Random Forests". In: vol. 45. Jan. 2011, pp. 157–176. ISBN: 978-1-4419-9325-0. DOI: 10.1007/978-1-4419-9326-7_5.

[4] *Data Exploration - A Complete Introduction*. URL: https://www.omnisci.com/learn/data-exploration.

[5] Y. Freund and R. E. Schapire. "Experiments with a new boosting algorithm, Machine Learning: Proceedings of the Thirteenth International Conference". In: (1996), 148–156.

[6] Stratos Idreos, Olga Papaemmanouil, and Surajit Chaudhuri. "Overview of Data Exploration Techniques". In: *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. SIGMOD '15. Melbourne, Victoria, Australia: Association for Computing Machinery, 2015, 277–281. ISBN: 9781450327589. DOI: 10.1145/2723372.2731084. URL: https://doi.org/10.1145/2723372.2731084.

[7] Yang Lu, Liping Sun, and Yanzhuo Xue. "Research on a Comprehensive Maintenance Optimization Strategy for an Offshore Wind Farm". In: *Energies* 14.4 (2021). ISSN: 1996-1073. DOI: 10.3390/en14040965. URL: https://www.mdpi.com/1996-1073/14/4/965.

[8] *Regression Analysis Essentials For Machine Learning*. URL: http://www.sthda.com/english/wiki/regression-analysis-essentials-for-machine-learning.

[9] Bhaba R. Sarker and Tasnim Ibn Faiz. "Minimizing maintenance cost for offshore wind turbines following multi-level opportunistic preventive strategy". In: *Renewable Energy* 85 (2016), pp. 104–113. ISSN: 0960-1481. DOI: https://doi.org/10.1016/j.renene.2015.06.030. URL: https://www.sciencedirect.com/science/article/pii/S0960148115300549.

[10] Mahmood Shafiee and John Dalsgaard Sørensen. "Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies". In: *Reliability Engineering System Safety* 192 (2019). Complex Systems RAMS Optimization: Methods and Applications, p. 105993. ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2017.10.025. URL: https://www.sciencedirect.com/science/article/pii/S095183201630789X.

[11] Christopher A Walford. "Wind turbine reliability :understanding and minimizing wind turbine operation and maintenance costs." In: (Mar. 2006). DOI: 10.2172/882048. URL: https://www.osti.gov/biblio/882048.

[12] *Wind Turbine Service*. URL: https://www.siemensgamesa.com/en-int/products-and-services/service-wind.

[13] Belle Xia and Peng Gong. "Review of business intelligence through data analysis". In: *Benchmarking: An International Journal* 21 (Apr. 2014), pp. 300–311. DOI: 10.1108/BIJ-08-2012-0050.

[14] Yang Yi and John Dalsgaard Sørensen. *Reduction of Operation and Maintenance Cost for Wind Turbine Blades – Reliability Model*. English. DCE Technical Reports 260. Project No. 871062. Department of Civil Engineering, Aalborg University, 2019.