MSc thesis in sustainable energy technology

The development of a two day ahead power forecasting model for an offshore windpark.

Jules Zweekhorst

2020

THE DEVELOPMENT OF A TWO DAY AHEAD POWER FORECASTING MODEL FOR AN OFFSHORE WINDPARK.

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Sustainable Energy Technology

This thesis is under embargo until: 22-09-2022

by

Jules Adriaan Zweekhorst

September 2020

ISBN 999–99–9999–999–9

The work in this thesis was made in the:

Intelligent Electrical Power Grids
Department of Electrical sustainable energy
Faculty of Electrical engineering, mathematics and computer science
Delft University of Technology

Supervisors:	Prof.dr. Peter Palensky
	Dr. Simon Tindemans
	ir. Wouter tromp
Co-reader:	Dr. Rudi Santbergen

ABSTRACT

The growing fleet of offshore wind farms in the North Sea imposes a lot of stress on the energy grid due to its volatile power production. Accurate forecasting models that can predict the power generation of these wind parks could help grid operators to keep the energy grid stable and reduce the re-dispatch of energy. Therefore the objective of this thesis was to find how to make an accurate power forecasting model for an offshore wind turbine park.

To find the most suitable model, five different models were made. The first model tested was a physical model which calculates the power output through a deterministic approach. Four other models were made through the use of machine learning algorithms. A Gradient Boosting model, a support vector regressor, a feed-forward neural network, and a recurrent neural network were built to find out which one was most suitable. The next aspect looked at was what input data could be beneficial to make the power predictions, the numerical weather prediction model ICON-EU was chosen as input data. The support vector regressor algorithm showed the best results with a score of relative mean absolute error 14.35%, which is the product of the mean absolute error divided by the difference between the minimum and maximum power production of the period in question. Since the performance of the model met the requirements it can be concluded that the forecasting model can be beneficial for the grid operator to reduce the amount of re-dispatched energy.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ANN	Artificial Neural Network
BRP	Balance Responsible Party
CV	Cross Validation
D-2	Two days ahead
DPE	Digital Process Excellence
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecasts
EMS	Energy Measurement System
FFNN	Feed Forward Neural Network
GB	Gradient Boosting
GFS	Global Forecasting System
ICON	Isocahedral Non Hydrostatic Model
KNMI	Koninklijk Nederlands Meteorologisch instituut
KPI	Key Performance Indicator
LSTM	Long Short Term Memory
NWP	Numerical Weather Prediction
MAE	Mean absolute error
MCP	Market Clearing Price
MI	Mutual Information
MSE	Mean Squared Error
PCC	Pearson Correlation Coefficient
R2	Coefficient of determination
RBF	Radial Basis Function
RNN	Recurrent Neural network
rMAE	Relative Mean Absolute Error
RMSE	Root Mean Squared Error
SCADA	Supervisory Control And Data Acquisition
SO	System Operations
SVM	Support Vector Machine
SVR	Support Vector Regression
TSO	Transmission system operator
XGB	Extreme Gradient Boost

1 INTRODUCTION

On 12 December 2015, all parties of the UNFCCC joined hands in the Paris agreement to tackle climate change. The contributing parties pledged themselves to stop global warming and limit the increase in average temperate to 1.5 C^o UNFCCC [2016]. To reach this goal, many sectors including the energy sector need to become more sustainable. The energy sector needs to make the so-called "energy transition" from fossil fuel-based energy sources to sustainable energy sources like wind and solar energy.

To meet the criteria set by the Paris agreement, the Netherlands needs to heavily invest in these renewable energy sources (RES). At the moment, the Netherlands is behind schedule to meet the required target and well below Europe's average as can be seen in figure 1.1. Due to its geographical location, off-shore wind energy will predominantly be used to meet the required quota and needs to generate at least 49 TWh of electricity in 2030 Wim van 't Hof [2016]. This large fleet of new offshore wind turbines imposes new challenges on the current electricity grid. Due to the volatile nature of wind energy, it is particularly challenging to accurately predict the quantity of power production, which may stress the electricity grid which constantly needs to be in balance. This is largely taken care of by the energy market.

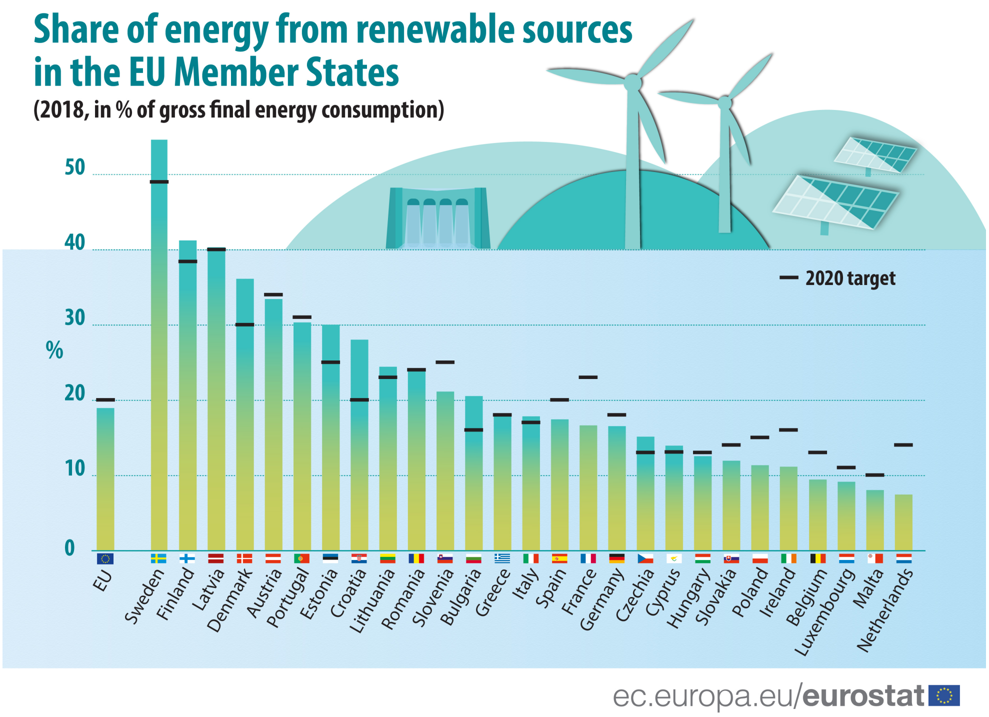


Figure 1.1: The share of energy from renewable sources from EU members Eurostat [2020].

1.1 THE ELECTRICITY MARKET

The short term electricity market works in the form of an auction. The main markets where most energy is traded are the day-ahead market and the intraday market. However when the energy flow is not in balance the balancing market operated by the TSO will correct the power flow.

On the day ahead market electricity is traded for each hour of the following day. Here the power producers offer the quantity electricity they are willing to generate at a certain price and the consumers bid on the amount of energy they want to buy for a certain price as depicted in figure 1.2. The market participants of the power exchange enter their orders before the market closes at 12:00. The energy price which is called the market clearing price (MCP) for each hour of the following day is then calculated and presented by the power exchange.

On the intraday market trades are made with delivery on the same date. The intraday market is more detailed where electricity can be traded up to five minutes before delivery. The energy is traded in hourly half-hourly, or quarter-hourly contracts Epexspot [2020]. This allows market participants to make last minute adjustments to their schedule allowing a high level of flexibility.

When the market participants do not fulfil their promise on either consuming or producing electricity, this will result in an imbalance in the power grid that needs to be fixed. The market that trades the power that will correct the power flow is called the balancing market. This market is operated by the TSO (TenneT) of the Netherlands. The market parties want to prevent having to deal with the balancing market as much as possible due to the additional costs it introduces.

The growing share of RES in the portfolio of electricity providers introduces new challenges in the energy markets due to their intermittent volatile nature. One of the challenges is when inaccurate RES power production forecasts are made and the market participant either sold too much or too little electricity they will be fined on the balancing market increasing the cost for the electricity provider. Another side effect is that the generation of RES can range between full capacity or nothing at all. If this is not managed properly it can result in congestions in the grid. To prevent grid congestions from happening the TSO can interfere with the orders of market participants to limit the electricity production at a critical point and produce it elsewhere, this process is called 'redispatch'.

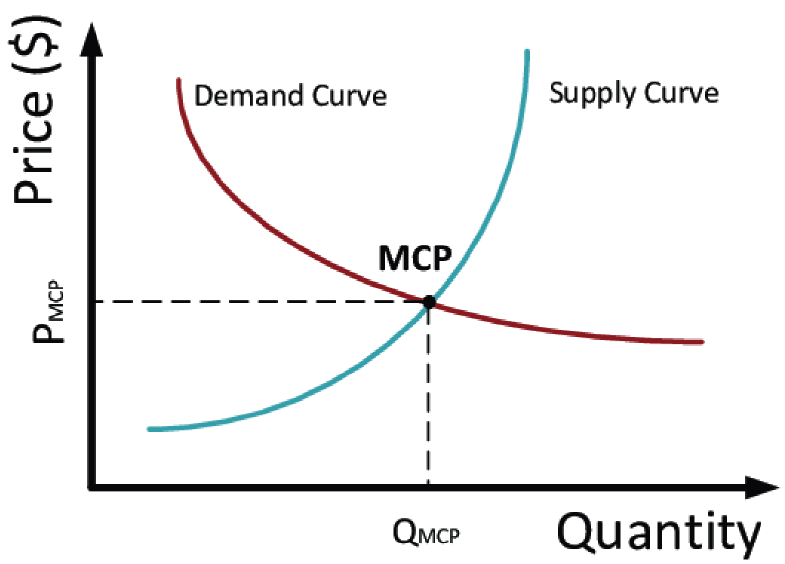


Figure 1.2: Schematic diagram of the market clearing Fouladfar et al. [2019]

1.2 DISPATCH AND REDISPATCH

Before the large implementation of RES in the electricity grid a constant balance was relatively easy to maintain. The daily energy consumption was relatively easy to predict and to meet this demand large centrally orientated power plants were scheduled to produce the required amount of energy to meet this demand. This process is called the 'dispatch' which tries to obtain the optimal economically-lucrative operation schedule through resource planning at the power plants obtained through the energy market. After selling the energy on the electricity market the power plant operators are required to send their operating schedule to the transmission system operator (TSO).

Now a lot of RES has been introduced to the electricity grid the situation changed. Due to the volatile power production from RES its production can range between maximal production or nothing at all. When the power plants have submitted their production schedules the TSO needs to do a grid safety calculation based on the next day's power flow. The safety calculation will determine whether the power flow will remain below the maximum allowed power flow in the transportation lines and is used to determine whether the energy grid will be secure in order to keep the number of grid stabilising interventions as small as possible. To prevent grid congestions on vulnerable grid points where the power flow may become too high, the TSO can ask power plant operators to change their scheduled power production based on the safety calculation. This change in electricity generation is called 'redispatch'.

E.g. this can occur when the wind is strong in the north of the country and wind turbines will generate a lot of power, while at the same time an operator of a power plant in the north of the country also sold a lot of energy. If most of the consumption would be in the south of the country, this would result in the power flow in the transmission lines from the north to the south becoming too high.

The grid operators would need to interfere and would have three possible actions in order to prevent the power flow from becoming too high. The first would be to ask the power plant operator up north to ramp down and produce less energy. The second action could be to temporally shut down the wind turbines, which is called curtailment. The third option would be to order the power plant operator in the south to generate more electricity.

When the power plant or wind turbines would be shut off, the operators would need to be compensated for their lost profit and the power plant operator that would need to produce additional power would need to be paid more than the market price. The costs that result from the redispach are paid for by the TSO. These costs will eventually reach the consumers since the price they pay in order to be connected to the grid is based on the expenses of the grid operators.

More accurate forecasts RES generation would reduce the improbability in the safety calculations of the TSO, resulting in less redispatch cases. Thus accurate power forecasting modes for renewable energy sources are of major importance in order to prevent the electricity price from going up. This is the reason the European Union imposed the regulation on TSO's to use meteorological forecast in order to make two days ahead predictions about the RES generation.

1.3 AIM OF THE PROJECT

This thesis aims to develop a model that can forecast the two days ahead (D-2) power generation for a wind park in the north sea of which the name will not be disclosed in this thesis. The System Operations (SO) department of TenneT TSO B.V. initiated the use case together with the Digital Process Excellence (DPE) department. This, to comply with the regulation imposed on TenneT by the European

Union intending to reduce the amount of re-dispatched energy and to get insights into RES forecasting .

1.4 RESEARCH QUESTIONS

To serve as a guideline during the thesis, research questions have been formulated. These will help to maintain the focus in the right direction. The main research question that was formulated states:

How should a D-2 power generation forecasting model of an offshore wind park be designed and built?

1.4.1 Sub-questions

The main research question is quite elaborate and difficult to answer on its own. To focus on specific relevant task, three sub-questions were derived in order to split the project up.

- **What requirements does a D-2 power generation forecast model need to meet, to conform to the use cases of TenneT TSO B.V.?**
- **What input data are beneficial to the machine learning forecasting model?**
- **What type of models and algorithms are effective at forecasting the D-2 power production of an offshore wind park?**

1.5 OUTLINE OF THE REPORT

The outline of this thesis is as follows. In the second chapter, a literature study is conducted to obtain background knowledge about forecasting of power production of wind turbines. In the third chapter, the requirements the model needs to meet are set up. The fourth chapter then elaborates upon the input data which are available and will be used. In the fifth chapter, the process of feature selection and feature engineering is worked out to find on what input data the model will be trained. The sixth chapter presents a way to validate the model predictions. Chapter seven, explains how the algorithms being used are optimised, so they can perform optimally. Once the models are trained and the data set is known, the results will be presented and discussed in the eight chapter. Finally once the results of the models are known the findings made in this thesis will be concluded in the ninth chapter. The last and tenth chapter presents recommendations for further research.

2 LITERATURE STUDY

In order to gain knowledge regarding power forecasting of offshore wind farms, literature research has been conducted. The necessity of the literature research is to lay a foundation of background knowledge in order to know the possible model choices. Topics of interests are model types and which algorithms to use.

2.1 MODEL TYPE

There exist various types of model methodologies which can be used to make predictions about the power generation of a wind farm. Lei et al. [2009] gives a survey of research and developments in the field of wind speed and wind power forecasting and states the different kind of models available. There are three main types of model structures that can be used to design a wind power forecasting model.

- Physical models, which calculate the parameter of interest in a deterministic way. These models are so called 'white-box' models since they calculate the parameter of interest in a transparent way that is easy to follow.
- Statistical models find patterns in data and use these to make predictions. These models are also known as 'black-box' models since the processes that calculate the parameter of interest are not known.
- Hybrid models are a combination of physical and statistical models. These these type of models then can be defined as 'grey-box' models, since some processes that calculate the parameter of interest are known and some are not.

2.1.1 Physical Models

The first model structure discussed is the physical model type, which makes a forecast of the power generation in a deterministic way. A physical model can be defined as the physical reproduction of the parameter one wants to model and is driven by a process. Thus the physical model tries to calculate the electricity production for the occurring meteorological conditions with a deterministic approach. This model is then fed with meteorological forecast data in order to make predictions about the future energy production.

Physical models are the most white-box models because its behaviour and are easily explainable due to the deterministic or causal approach Briggs [2016]. When the input data is known, the process of how the model converts it to an output is straightforward because of the formulas it uses. For calculating the power production of a wind turbine park this would be the formulas that convert the meteorological conditions to power production. However, the straightforwardness of the physical models can also lead to errors. Some physical phenomena are too complex (i.e. the influence of turbulence in the wind) or not physically analysable (i.e. the influence of the wake effect for every wind direction), thus assumptions need to be made. These assumptions introduce uncertainties into the model which leads to errors.

2.1.2 Statistical Models

The second model type discussed is the statistical model. Statistical data driven models work through the use of statistics on historical data which are used to make predictions about the future Barak Rishan [2018]. They derive relationships between the input and the output without looking at the underlying process. The model is fed the training data in order to learn the trends and correlations between the data points and then generalises these. In general, statistical models require a lot of historical input data to train the model on, because before the model can predict an event the model needs to have seen something similar before.

2.1.3 Hybrid Models

The last model type discussed is the hybrid model type. Both the physical and statistical model types have their advantages and disadvantages, thus these models can be combined to strengthen each other Foley A [2012]. Physical models are based on theory and therefore can only mimic the behaviour of the wind park in accordance with the known relationships that convert meteorological conditions to a power output. However, some processes are too complex or not (easily) physically analysable. These processes and relations can then be found by combining the physical model with a statistical model to enhance the model's accuracy Makoto. Nakaya, Xinchun. Li [2018].

Models can be combined in multiple ways. Two main methods discussed are the stacking and bagging of models.

The stacking method uses an ensemble of models which predict the parameter of interest in order to make a meta-model which is the eventual model that will be used, stacking can be used for both classification or regression problems Vadim Smolyakov [2017]; Khairalla et al. [2018]. The meta-model is trained on the output of the base level models, which learns the superior characteristics of each model and therefore can reach higher accuracies than individual models .

Another means of combining different kind of models is trough bagging. The bagging model that often consists out of homogeneous weak learners which are trained independently from each other Joseph Rocca [2019]. The results of these models are then combined and averaged. So the bagging method only averages the weak learners, this does not change the output of the input models however it reduces their variance. Due to the simple approach in aggregating the models Ting and Witten [1999] found that the implementation of stacked generalisation achieved better predictive accuracy than when combining models than when done with the bagging method.

2.2 ALGORITHMS AND METHODS

Now the model types are known in order to design a forecasting model for the power production of an offshore wind farm. The possible methods and algorithms that are going to be used need to be selected.

2.2.1 Power curves

The power curve is a graph that indicates how much power a wind turbine produces for certain wind speeds and is one of the most basic ways to convert meteorological data to a power output. Therefore power curves are an important tool in the wind energy industry to assess and monitor the power forecasting of a wind turbine Shokrzadeh et al. [2014]. Figure 2.1 depicts a typical wind turbine power curve where can be seen how certain wind speeds are converted to their respective outputs. Manufacturers of wind turbines provide these power curves with their wind turbines. However, power curves only hold in perfect conditions which is not the case in reality. When wind turbines are grouped in a wind park phenomena like wake effects occur, or the sudden deviation in wind speed will deviate the power production of the wind park from its theoretical power curve.

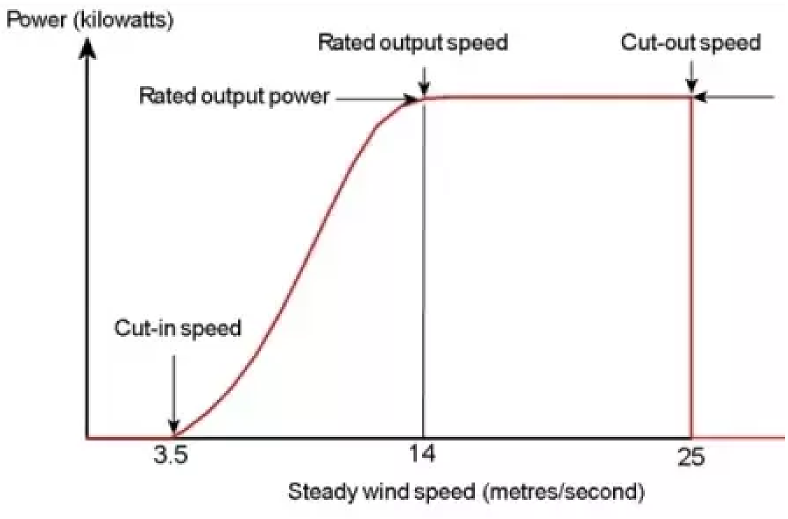


Figure 2.1: A typical wind turbine power curve Ron Wiseman [2017].

Therefore when enough data is available it is a better option to fit an empirical power curve to take into account external phenomena like wake effects which differ per wind park Shokrzadeh et al. [2014]. Niayifar and Porté-Agel [2015] studied different ways of modelling wind turbine power curves using parametric and non-parametric methods when fitting a power curve of a wind park empirically with enough data.

However, implementing a fitted power curve in the forecast model does not come without risk of reducing the forecast accuracy compared to a forecasting model based on a provided power curve Goretti et al. [2017]. Since an incorrect power curve can introduce a significant error in the wind park power forecasting if not properly modelled or when inaccurate data was used to derive it. When insufficient or inaccurate data is available its better to use the power curve provided by the manufacturer.

The Windpowerlib Python package

The windpowerlib is a python toolbox designed to calculate the power output of wind farms Haas et al. [2019]. It is derived from the feedinlib (wind power and pv) to build up a community concentrating on wind power models. This feedinlib was split up due to the increasing size of the windpowerlib. It uses many modes in order to mimic the behaviour of the wind farm are accurate as possible and calculates the power output in a detailed way. Inputs for the library would be the physical parameters from the wind park along with data regarding the current meteorological conditions.

The package parameters of interest in the are the capacity of the wind park, the hub height of the wind turbine and the power curve of the wind turbine. The power curve was determined empirically, the power output of the wind park was determined by the wind speed in steps of 0.2 m/s.

When the wind turbine is initialised the modelchain needs to be defined, here all the model features provided by windpowerlib can be selected. When no function is selected the default parameter is used. The functions selected are for the wind speed model, density model, temperature model, power output model, density correction and the hellman exponent.

2.2.2 Gradient boosting method

Another promising machine learning algorithm is Gradient boosting Chen and Guestrin [2016]. Competitions on forecasting renewable energy in the wind sector have been won using this algorithm Landry et al. [2016]. Gradient boosting is a technique that can be used for solving regression or classification problems and was developed by Friedman [2002].

It produces a model of an ensemble of weak prediction models which typically are decision trees as can be seen in figure 2.2. The algorithm tries to predict values (Y) of the model F in the form $Y = F(x)$. The way these models are trained through the minimisation of their prediction error.

The gradient boosting method used is XGBoost, this is an optimised distributed gradient boosting library developed to be efficient and flexible on implementation. Therefore gradient boosting is very suited for the prediction of the wind farm power output.

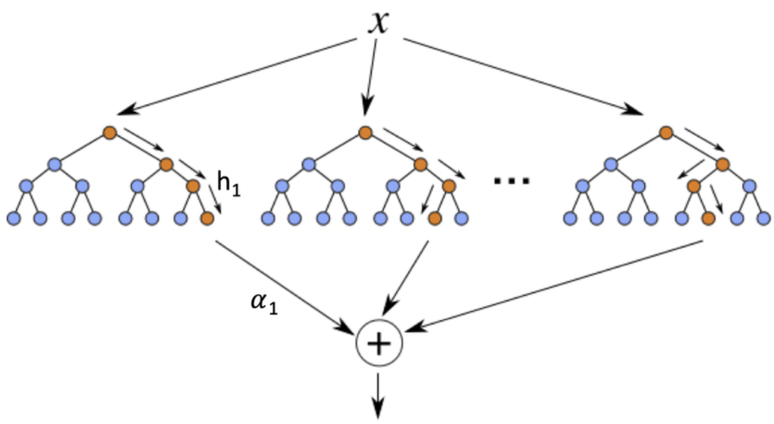


Figure 2.2: Visualization of the gradient boosting decision tree.

2.2.3 Support vector machine regression

Support vector machines (SVMs) are generally used for classification problems although they are very much suited to solve regression problems as well, when used for regression problems they are called support vector regressors (SVR). The objective of the SVM is to find a hyperplane in an N-dimensional space that fits the data points appropriately, with N being the number of features used in the regression problem.

Kong et al. [2015] researched the topic of forecasting wind speed with the use of support vector regression (SVR) in order to solve problems of security, stability and quality caused by wind energy production systems. They found that SVR can be a powerful machine learning model with a good generalization capability which prevents the model from overfitting, which is favourable in forecasting power from wind turbines. Furthermore, SVR ensures fast training of the model with a low computational burden.

The objective function of SVR is displayed in 2.1 where w is the l2-norm of the coefficient vector, ε is the allowed error, and ξ is the deviation from the allowed error.

With the constraints displayed in equation 2.2. Figure 2.3 depicts a visualization of SVR with the error margin and the slack variables.

$$MIN\frac{1}{2}\|w\|^2 + c\sum_{i=1}|\xi_i| \tag{2.1}$$

$$|y_i - w_i x_i| \leq \varepsilon + |\xi| \tag{2.2}$$

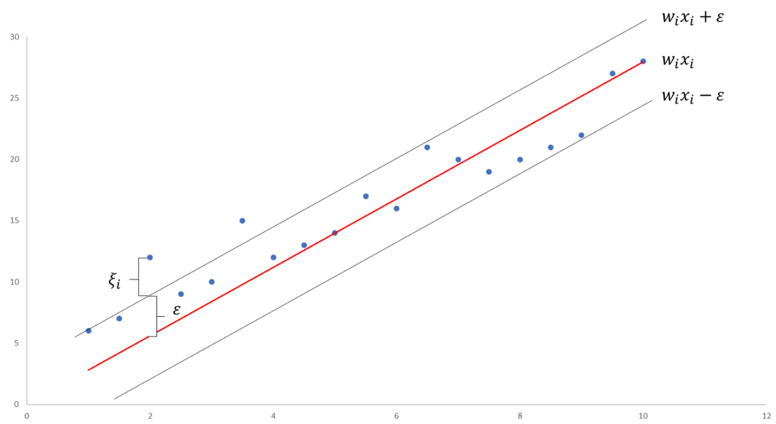


Figure 2.3: Visualisation of SVR with slack variables. Tom Sharp [2020]

2.2.4 Fuzzy logic

Another type of forecasting models is based on fuzzy logic Damousis et al. [2004]. This is a machine learning area which is based on approximate reasoning and computational intelligence. Fuzzy models can be used in cases where it is difficult to exactly model the situation or where uncertainty is encountered in the problem definition. The principle of fuzzy logic is depicted in figure 2.4. As can be seen with Boolean logic can only be either cold or not cold, whereas Boolean logic present a more gray scale where the definition is not set in stone and the value van be partially true and partially false.

The fuzzy logic could be interesting on for example an edge case with high wind speeds which are higher than the cutoff wind speed. In this case the wind turbine park is not always cut of as a whole, but sometimes only a part of the wind park is cut off.

However what has to be noted is that fuzzy models always approximate the solution and thus care needs to be taken when interpreting their results.

2.2.5 Artificial Neural network

Artificial neural networks (ANNs) are computing systems inspired on the brain, which consist of a network of nodes or so-called 'neurons' A. Foley [2010]. These neurons are all connected to each other by weights and transmit a signal to other neurons. This way the Neural network is able to convert the input data into the output data. This network is able to learn certain tasks through feeding it on input data without it having prior knowledge about a subject. ANNs are excellent at finding complex patterns in data sets and therefore used extensively throughout the machine learning field in the forecasting of energy.

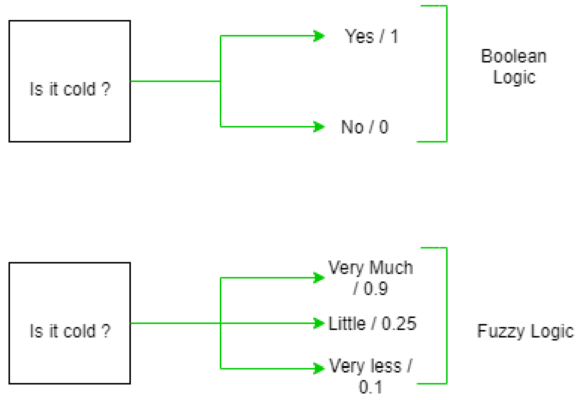


Figure 2.4: Visualisation of fuzzy knowlegde. Surya Priy, Abhishek Rajput [2019]

Input data is transformed through the ANN due to the way the nodes are connected to each other. Every node receives input data which then gets multiplied by a weight value in order to transform the data this will then possibly receives a bias before it reaches the output of the node as seen in figure 2.5. Whether the output of the node will be used and the neuron will activate (fired), is based on relevancy of the neuron's input data for the models predictions and is determined by the activation function. So the activation function determines the output of the neuron given a set of inputs. The relevancy of the activation function can be explained though how the human brain works. If the eyes see a flower the visual neurons which will process this data will be triggered in order to recognise the flower. However if the eyes perceive a tree the input data is different and the neurons which could recognise flowers would not need to be triggered but the ones which can recognise trees do. Thus the activation functions determines the relevancy of the input data for the neurons.

Neural networks consist out of three types of layers namely: the input layer, hidden layers and the output layer Géron [2019]. The number of neurons in the input and the output layer are fixed since the in and output of the model are defined. With the number of input neurons being equal to the amount of input features and the number of output neurons equal to the parameter(s) of interest. The amount of neurons per hidden layers can still be optimised, there can be as many hidden layers as desired. More hidden layers result in a more complex network that can recognise complex patterns. Deeper neural networks also have a higher parameter efficiency compared to shallow networks, resulting in a better performance of deep networks compared to shallow networks when the input data is limited. However, if the model has too little neurons the model will be underfit and cannot recognise all the patterns. When the model has too many neurons the model will be overfit and sees correlations which are not there or is not trained enough on the amount of training data provided.

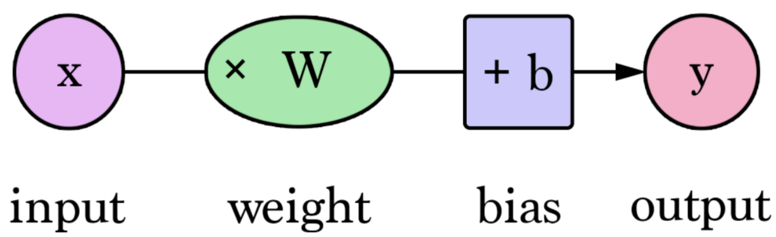


Figure 2.5: Visualisation of the node connection in an ANN Jay Alammar [2019].

ANNs need to be trained to find the patterns which are hidden in data sets in order to make predictions of the output from the input data Géron [2019]. This is done through splitting up the training data in training batches and sending these batches through the model. Each time the model makes a predicting the outcome is compared with the actual value the prediction should be. The difference between the prediction and the actual value is called the 'loss' which then is used by the loss function in order to update the weights in which the Neurons are connected through each other in a process that is called backpropagation.

The number of times the model is trained on the whole training set are called the Epochs. Once the error of the model is within an acceptable range or the final epoch is trained the model is ready to make the forecast.

Neural networks are widely used for the prediction of time-series data and thus suitable for the prediction of the power production of an offshore windpark. Bhaskar and Singh [2012] used a Feedforward neural network (FFNN) in order to map the non-linear correlation between windspeed and power output, due to the strong non-linear pattern recognition capabilities of neural networks.

The use of recurrent neural networks (RNN) in order to predict windspeed is discussed by Balluff et al. [2015] and concludes that RNN's can increase the forecast accuracy when predicting wind energy trough the storage of relevant information in the model.

Feed forward neural network

The first type of neural network chosen is a feed-forward neural network (FFNN) as used by A. Foley [2010]. It is one of the most basic types off ANN, also known as the "vanilla" neural network. The direction of the information flow is unidirectional so it passes from the input nodes through the hidden layers into the output layer. A visualisation of this type of network is shown in figure 2.6.

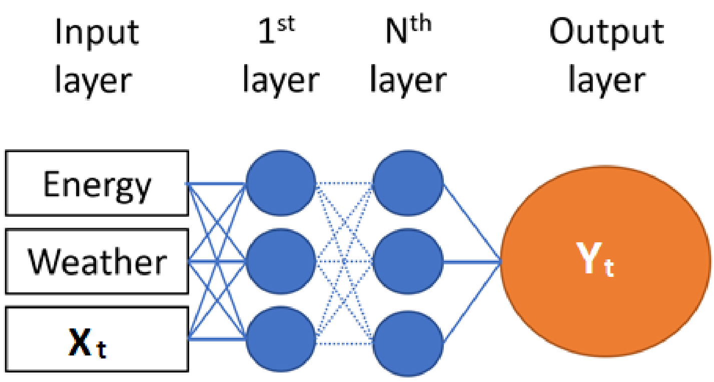


Figure 2.6: Visualisation of a FFNN.

Recurrent neural network

RNNs are proficient in predicting the next value in a sequence through learning arbitrary patterns of the input data. The output of values of the neurons are stored and fed back into the model for reiteration as seen in figure 2.7. Classic RNNs have the problem that due to the unknown amount of data points (gap length) between important events in the data the exploding or vanishing gradient problem can occur. Therefore the Long Short Term Memory (LSTM) network is used. These have the advantage that there is an insensitivity to gap length in the time series. Wu et al. [2016]; Hochreiter and Schmidhuber [1997] This aspect makes it a suitable choice for the forecasting of energy generation.

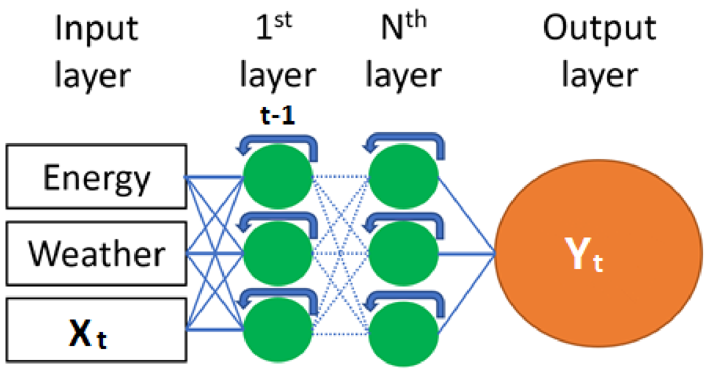


Figure 2.7: Visualisation of a RRN.

2.3 TEMPORAL MODEL CLASSIFICATION

Wind power forecasting can be classified in various ways. The temporal model category determines what kind of models are possible and what input data can be used. In terms of time scale, the models can be categorised in three different categories Wang et al. [2011].

- Immediate-short-term forecasting (until 8 hours ahead)
- Short-term forecasting (day-ahead)
- long-term forecasting (multiple days ahead)

The project aims at making a forecast with the time span of 48 hours, therefore the model would be considered a long-term forecasting model.

2.4 MODEL OUTPUT: POINT FORECAST VS PROBABILISTIC FORECAST

Other than the time horizon, models can also be classified based on the output data of the model models, an ensemble NWP model, or through data about wind volatility Juban et al. [2008], Liu et al. [2013]. A point forecast predicts only the most likely value for the variable of interest. Another method is ensemble forecasting, which can be used to make probabilistic forecasts, it runs different models parallel to each other to predict multiple outcomes. The results of these models are then evaluated and a confidence interval can be made accordingly to estimate the likelihood the predictions will become reality Foley A [2012]. Another method of making an ensemble forecast is to run one model with input data from multiple data sources. This can be from various NWP.

2.5 MODEL SELECTION

Now the possible model types and algorithms are known it is possible to choose which will be used for the forecasting model to predict the power production of the offshore windpark. Five promising models will be built in order to find which one is most suited to predict the power output.

- The first model that will be made is a physical model based on windpowerlib. Since this is a physical model it will prove as a good baseline.
- The second model that will be build is a gradient boosting model based on the XGBoost algorithm. This algorithm has won awards with energy forecasting competitions and thus is a very promising model Landry et al. [2016].
- The third model made is based on the support vector regression algorithm. Due to its great generalisation characteristics it is a suitable candidate.
- The fourth model set up is the FFNN since it has good pattern recognisable capabilities.
- The last and fifth model made is the RNN model based on a LSTM architecture. Just like the FFNN it has great pattern recognisable capabilities although it should be more suitable fore time series problems.

2.6 CONCLUSION

A thorough background research was conducted to find the possible model types. The relevant models which will be built are: A gradient boosting model, a SVM model, a FFNN, a RNN, and a physical model based on windpowerlib.

The next step is to determine what input data to use for the models. This is done in the following chapter where will be explained how feature selection is used to determine the importance of the input data.

3 REQUIREMENTS OF THE PROJECT

Before the model can be built its scope and all the requirements it needs to fulfil have to be set up in order to make the two day ahead power forecast of a wind park. The necessary quality indicators need to be known as guidance to design and build the model on. This chapter elaborates on the model requirements and the process of designing key performance indicators (KPI) for rating the model. The model requirements were set up in accordance with the case owners through the conduction of interviews. Along with the interviews, statistical error estimation tools which indicate the model performance have been elaborated. Both the requirements that came from the interviews along with the result from other studies formed the basis on the design of the KPI's.

First, the model output requirements are presented. Followed by the error measurement tools which can be used to evaluate the model. Lastly, the findings of this chapter are presented in the conclusions.

3.1 THE MODEL OUTPUT REQUIREMENTS

All the model requirements are defined in this section. First, the specifications of the output data are defined, then the possibility of introducing bias is discussed and finally the decision to choose between a point or probabilistic forecast is made.

3.1.1 Output data

The model aims to predict the D-2 power output of a wind park in the North Sea. This means the forecast needs to be at least 48 hours in the future. This prediction needs to contain the average power production at a temporal resolution of one hour. The objective function of the forecasting model could be described by equation 3.1. Where Y_{48} is the actual power generation of the wind farm 48 hours in the future. The function F then uses the input dataset X_0 which is available at T=0 to predict the power production 48 hours in the future which is denoted as $\hat{Y}_{48}$.

$$Y_{i+48} \approx F(X_i) = \hat{Y}_{i+48} \tag{3.1}$$

Due to the uncertainty in the meteorological predictions or a model error, the actual power production can deviate substantially from the predicted value. When this results in the initial predictions being much lower than the actual power production it might result in grid congestions. This problem could be prevented to a certain degree through changing the objective function of the model in such a way that the predicted power output is always slightly higher to prevent the actual power production from being greater. Although the introduction of a bias makes the overall accuracy of the model decrease resulting in shifting the problem elsewhere along the grid. Thus the bias is not introduced in this model. This model aims at predicting the power generation of the wind turbine park at the highest average accuracy.

There is also the possibility to create an ensemble model which is able to make a probabilistic forecast. With an ensemble model, multiple predictions are made in parallel, with these outputs a confidential interval can be made. This gives an estimation of the probability the forecast will happen and can give a lot of insights for the system operators to see where problematic situations might occur. Right now this option is not necessary. However, in future projects the possibility to implement this feature might arise.

For now, the requirements on the output data of the model are to make a point forecast as stated in equation 3.1. This point forecast should be as accurate as possible without the introduction of a bias.

3.2 STATISTICAL ERROR MEASUREMENT TOOLS

In order to determine the quality if the forecasts, error measurement tools need to be selected in order to evaluate the models' accuracy. Many statistical tools can be used to evaluate the forecast accuracy. This section elaborates relevant types of these statistical tools used in literature or throughout TenneT TSO B.V. Botchkarev [2018] Madsen et al. [2004] along with an explanation on how these can measurements tools be of use for model accuracy evaluation.

Mean absolute error (MAE)

The MAE is the absolute average value of the forecaster energy generation data point are deviating from the actual generated value. It gives a solid overview of how well the model performs by giving the average magnitude of the errors without considering their direction. Thus it does not provide information on whether the model has a certain bias but only gives the size of the error. Formula 3.2 gives the mathematical expression of how the MAE is calculated, where Y_i is the actual power generation and $_i$ the predicted power generation.

$$MAE = (\frac{1}{n}) \sum_{i=1}^{n} |Y_i - \hat{Y}_i| \tag{3.2}$$

Relative mean absolute error (rMAE)

The rMAE is used to put the MAE into perspective and displays the parameter in a percentage of the normalisation value. The scale the parameter needs to be normalised to needs to be chosen carefully though to give an accurate perspective. At TenneT the normalisation factor is the difference between the minimum and maximum generation in the respective month as shown in equation 3.3. Where Y_{max} is the maximum generation in the month and Y_{min} The minimum generation in the month.

$$rMAE = (\frac{1}{n}) \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_{max} - Y_{min}} \right| \tag{3.3}$$

Mean squared error (MSE)

The MSE measures the average of the squared forecast errors. It is a measurement tool to measure the quality of the forecast which emphasises the fact that large errors are punished harder than small errors as seen in equation 3.4 Meaning this tool can be used to analyse whether the model is more prone to make large mistakes.

$$MSE = \frac{1}{n} \Sigma_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{3.4}$$

Maximal error

The maximum error is also one of the indication tools to measure how well the model performs. It can be used to verify if the predictions of a model conform to the standards it needs to meet if a certain error magnitude is not allowed.

Coefficient of determination (R^2)

R^2 is a statistical measure that represents the proportion of the variance for the actual power output that is explained by the predicted output. Where the correlation explains the strength of the relationship between these variables. It is a tool predominantly used for testing forecasts making is a suitable candidate to rate the model on. Formula 3.5 depicts how the R^2 can be calculated where SS_{tot} is the total sum of squares which is the squared product of the energy generation Y_i minus the average energy generation $\bar{Y}_i$. SS_{res} is the residual sum of squares calculated by taking the square of the energy generation Y_i minus the predicted energy generation $\hat{Y}_i$.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \qquad SS_{tot} = \sum_i^n (Y_i - \bar{Y}_i)^2, \qquad SS_{res} = \sum_i^n (Y_i - \hat{Y}_i)^2 \tag{3.5}$$

Bias of the estimator

The bias of an estimator is the average error the model makes. It is the summation of the model errors taking into account whether the value is positive or negative as seen in equation 3.6. So it measures how much the predictions tend to go to either side of the mean.

$$bias = \sum_i^n (Y_i - \hat{Y}_i) \tag{3.6}$$

3.3 KEY PERFORMANCE INDICATOR

In order to assess whether the model performs according to the standard of TenneT a means to measure the quality needs to be derived. This is done through a KPI.

Through an interview with the case owners of SO, the requirements of the quality of the model have been discussed. The main error estimation tool of importance is the rMAE, therefore the goal of the project is to reduce the rMAE as much as possible. Foley A [2012] presented an in-depth review of the current methods and advances in wind power forecasting and prediction, and found that for day-ahead predicting of the power production of wind farms a rMAE of 10-15% of the maximum capacity has been obtained.

The final KPI that the model needs to meet is derived through the baseline which states that the maximum MAE of the day ahead model was 15% of the nameplate capacity. Thus it would be decided that this same measure would be taken for the two day ahead model. Thus a maximum rMAE allowed would be 15%.

3.4 CONCLUSION

The model is going to forecast the power output of a wind park in the north sea at a temporal resolution of 15 minutes, while trying to minimise the average output error. The statistical tool used to estimate these errors will be the rMAE. The rMAE can scale the MAE so it can be compared with different wind parks.

Now the scope of the project is elucidated, the requirements have been defined, the statistical tools necessary to evaluate the model are known. Thus the next step is that the possible model structures and algorithms need to be selected which will be elaborated in the next chapter.

4 DATA SOURCES

Accurate and relevant input data is arguably the most important facet in machine learning. This chapter delves deeper into the possible data sources that can be accessed to feed and train the model. First, the available internal, energy-related data sources from TenneT will be discussed. Secondly, a description of Numerical Weather Prediction (NWP) models which make forecasts about meteorological conditions is given. After the explanation of NWP models, meteorological realisation data is discussed, followed by other relevant data sources that will be discussed. The concluding chapter will state the chosen data sets and how these will be fed into the model.

4.1 RELEVANT DATA SOURCES

Due to the complexity of long-term forecasting models, the use of NWP models needs to be considered Wang et al. [2011]; Xu et al. [2015]. Models not using NWP might perform within an acceptable error margin for the first few hours however they become very inaccurate for longer predictions. Even though about 80% of the model error can be explained due to errors in the NWP model. Sideratos and Hatziargyriou [2007] researched advanced statistical methods and did not only use NWP models but also the historical power production of the wind park.

4.2 ENERGY RELATED DATA SOURCES

This section discusses energy-related data sources that can be used to feed into the model. The sources used are the measured energy generation of the wind park along with T-prognosis data. An overview of the energy-related data can be seen in table 4.1

4.2.1 Power output of the wind park

A reliable source for the actual output power of the wind turbine park needs to be used. The exact amount of power the wind park produces is measured at the point where the offshore wind farm is connected to the Dutch energy grid. The Energy Measurement System (EMS) system monitors power flow in the Dutch energy grid at all times. Thus the EMS is able to provide accurate data regarding the quantity of energy delivered on the energy grid.

The EMS system measures the power flow at each grid node every few second. This data is then aggregated and stored as the average energy transported in KWh per 15 minutes. These measured values are not directly used due to the possibility of mistakes in the measured values in the power flow. The measured data is thoroughly checked to verify its accuracy by the Supervisory Control And Data Acquisition (SCADA) system before being stored in the database, this verified data is named: the settlement data. The settlement data are available a few days after the actual measurements of the power flow, due to the verification process of the accuracy.

4.2.2 T-prognosis

Once a wind park becomes larger than one megawatt is connected to the energy grid the owner of the asset becomes a balance responsible party (BRP) and is obligated to report how much energy they consume or generate to the electricity grid to the TSO. These prognoses are so-called Transport prognosis (T-prognosis). This prognoses contains information about the production of the coming 24 hours and is delivered by the wind park to the TSO the day ahead before 15:30 hours. These files contain information about the average amount of energy delivered at an fifteen minute interval and whether the wind park is fully operational. Tennet TSO [2020] These T-prognosis do not yet have a guideline in the form of a KPI to which they have to conform to ans the quality of these prognoses are not determined via a monitoring process. Therefore this data has been qualified as not to be suitable input data for the model.

Table 4.1: Overview of the energy related data

Data	Interval	Source	Unit	Availability
Settlement data	15 minutes	SCADA	[J/15 min]	Post measurement
T-prognosis	15 minutes	BRP	[J/15 min]	Day ahead

4.3 NUMERICAL WEATHER PREDICTION MODELS

Numerical weather prediction (NWP) models solve a set of mathematical equations that calculate the changes in the atmosphere, in order to make predictions about the future meteorological conditions. There are many different NWP models that are used throughout the world, all with their own characteristics. Therefore a selection of these models has to be made in order to find the most useful NWP model for the forecasting of the D-2 energy generation of an offshore wind turbine park. This section delves into the different NWP models and compares them in order to find the most suitable for the use case. First KPIs are derived to rate the different NWP models on. Secondly a summary of different relevant NWP models are given and finally the NWP models are compared to each other.

4.3.1 The Key performance indicators for the selection of NWP models

To obtain the most useful forecasts as input data, the NWP models need to be rated according to their usefulness for the prediction of the energy production by wind farms. The NWP model results are very case sensitive and each model does not suit every application optimally Kalverla et al. [2019]. For these reason KPIs have been developed to rate the NWPs appropriately. This way the strengths and weaknesses of each model can be brought into clear perspective in order to select the most useful output data of each model. The following criteria have been selected to rate the models on and turn into KPIs.

- The time span of the model. This is one of the most dominant criteria to select the NWP model on due to the fact that the model needs to be able to forecast the energy production of a wind park 48 hours ahead. Thus in order to make use of the predicted meteorological data the NWP model it is desirable to have a time span that is longer than 60 hours. This extra time is necessary to acquire the most recent forecast and process it.

- The size of the time steps the NWP model uses for its output data is also of importance. The output of the energy generation forecast model will be in time steps of one hour. Thus a NWP model with an interval of one hour would be ideal so no data needs to interpolated and the closer all the data points will resemble reality.

- The spatial resolution the model uses is of importance as well. The closer the nodes of the grid are to each other the more the NWP model can resemble reality. This way the interpolation from a grid node to the actual location of the wind park has less influence on the outcome of the NPW model.
- The cost is also of importance when selecting the model. To obtain actual NWP forecasts from certain providers, a licence needs to be bought. The licence cost of using a specific weather forecasting model should be proportionate to the value it creates.

4.3.2 Description of the different NWP models

This section delves into the different models that make meteorological predictions of the area of interest: the North sea. A brief summarisation of the models is given along with their characteristics.

HARMONIE

The HARMONIE model is a limited-area NWP model created in 2012 by the Koninklijk Nederlands Meteorologisch Instituut (KNMI) KNMI [2020]. This model is a short term forecasting model that produces weather predictions of up to 48 hours on an hourly interval. Its focus is on predicting the meteorological changes on the north west region of Europe. Due to this limited area, the grid it uses is relatively fine with a spatial resolution of 2.5X2.5km. The KNMI runs the model four times a day to update the input parameters and give new predictions. The HARMONIE model is not a global model thus it needs to be fed on initial and boundary condition, these are provided by the global ECMWF model L. Bengtsson [2017].

ECMWf-HRES

The European Centre for Medium-Range Weather Forecasts (ECMWF) model. The ECMWf-HRES is a global NWP model. It is a medium-range model which can make predictions about the atmosphere for up to 15 days. The model is designed in such a way that the first 90 time steps have an hourly interval making it more precise for the short term forecast. The model is run four times a day, with the initial times 00, 06, 12, and 18 h UTC.The grid size the high resolution model uses has a resolution of $9X9km$ ECMWF [2019]. The basic model is an point forecast although there are also ensemble forecasts with an aggregation of 52 individual forecasts ECMWF [2020].

ICON-EU

The Isocahedral Non hydrostatic model (ICON) modelling system is developed by the Deutscher Wetterdienst (DWD) in collaboration with the Max-Planck Institute for meteorology in Hamburg D. Reinert and Zängl [2019]. It is a short and medium range weather forecasting system that can make predictions about the weather up to 78 hours on an hourly interval with its most detailed version. ICON operates on a global scale and therefore its native grid resolution is 13km. It does have a refined version over the European model with a grid length of 6.5km which is called the ICON-EU model. The model is run 4 times a day by the DWD. The Icon also has a ensemble forecast, the ICON ensemble suite generates 40 short to medium range forecasts.

GFS

The Global Forecasting system (GFS) is NWP created by the national center for environmental prediction. It is a medium range forecast model who predicts the weather out to 16 days in the future at a three hourly interval. The GFS is a global forecasting model with a spatial resolution of 28 km between grid points. The GFS is a coupled model which combines four different models to create an accurate picture of the weather conditions. National Center For Environmental Prediction [2020]

4.3.3 Comparing the NWP models

Now basics of the more relevant NWP models are known along with the KPIs on which the models are rated. The most suitable can be chosen to provide the meteorological forecasts. The KPIs have to be divided in two different groups. The first group being the group that can be answered using literature research. Containing the KPIs about: The horizon of the model, the time steps, the spatial resolution, and the frequency at which it is run. The second group will be answered in the feature selection of the model containing the accuracy of the NPW model and the relevant parameters. An overview of the characteristics of the different NWP models can be seen in table 4.2.

Table 4.2: NWP performance matrix

Parameter/ model	Horizon [hour]	Temporal resolution [hour]	Spatial resolution [km]	Frequency [hour]	Cost [€]
HARMONIE	48	1	2.5	6	Free
ECMWF-HRES	90	1	9	3	100,000
ICON-EU	78	1	6.5	3	Free
GFS	384	3	28	6	Free

Forecasting horizon

As mentioned in chapter one, the forecasting horizon of the model needs to be 48 hours. Meaning it would be desirable that the predictions of the NWP models have a forecast horizon that exceeds these 48 hours plus the update frequency of the NWP model and the processing time. To provide some slack for the model a horizon of at least 60 hours is desired. The only NWP model that does not meet this criteria is the Harmonie model with a forecast horizon of 48 hours.

Time steps

The required output of the model is the average energy production over and interval of 15 minutes. Thus the closer output of the NWP models is to these 15 minutes the more accurate the actual results will be. Due to the effect that the interpolation will have on the accuracy of the predictions.
The HARMONIE, ECMWF, and the ICON model have a hourly output frequency, except for the GFS model which has an output frequency of 3 hours.

Spatial Resolution

The smaller the spatial resolution of the NWP model is the more detailed the predictions of the model can be. Not only the more detailed forecast is beneficial for the model due to the fact that the wind park most probably will not lay exactly

on a node. Thus the finer the spatial resolution the less influence the interpolation between the forecast nodes and the wind park will be.
The Harmonie model has the finest spatial resolution of 2.5 km followed by the ICON model which has a resolution of 6.5 km. The ECMWF has a spatial resolution of 9 km and the GSF has the most course grid with a spatial resolution of 27km.

Cost of the model

The cost of the model should be proportionate to the value it creates. The HARMONIE, ICON, and GFS model can be obtained for at only the cost for the data transfer. The ECMWF-HRES model which cost up to 100.000 euros per year is too expensive for this use case and therefore is not one of the possibilities.

4.3.4 Model choice

The model chosen to provide the meteorological input data is the ICON model from the DWD. This model scores well on all and is available for free. It is also easy to obtain and therefore a suitable candidate to provide the NWP data. What has to be noted is that historical data set bought from the DWD does not have four model runs per day. Only the 00:00 and 12:00 runs are stored in the data set. This results in the model having to forecast 48 + 12 hours instead of 48 + 6 hours.

4.3.5 Meteorological parameters chosen

Now the source of the meteorological data are chosen it is important that the potential relevant parameters on which the model will be trained on are selected. Ideally the more parameters are collected in this part of the project the better since not all the data sets are public and need to be purchased. However not all the parameters might be beneficial for the forecasting model or possible to obtain. When the relation between the power output and the respective parameter is too low the model could see the data source as noise which might reduce the accuracy and increase the computational time. Therefore the feature selection will be further elaborated on in chapter 5. Right now the value and availability of certain parameters needs to be determined, these parameters are available in all the meteorological data sources.

The following parameters have been selected and can be seen in table 4.3 and 4.4. Table 4.3 depicts the single level parameters that will be collected, this means that the ICON-EU model only calculates these parameters at a single height.

The multi level parameters which the ICON-EU model calculates for different height levels are depicted in table 4.4. For the wind speed the ICON-EU model only provides the U and the V component. U is the zonal wind speed which goes from the east to the west and V is the meridonal wind speed which goes from the south to the north.

Table 4.3: Single level parameters

Parameter	Description	Unit
PS	Surface pressure	Pa
WSMAX	Wind gusts	m/s
SR	surface roughness	m
2r/RELHUM	Relative humidity	$\frac{kg}{kg}$

More height layers might have been beneficial to the model since the turbines are approximately 100m in the air. However the only multilevel parameters made available are stated in table 4.4.

Table 4.4: Multi level Parameters

Parameter	Height [m]	Description	Unit
U	10	U component of wind speed	m/s
V	10	V component of wind speed	m/s
T	2	Temperature	K

4.4 METEOROLOGICAL DATA FOR MODEL TRAINING

Now the available parameters from the NWP model are known the data on which the model will be trained is selected. This is meteorological realisation data, which contains information about how the meteorological conditions were at the wind farm on a certain time. This data then can explain the resulting power production of the wind farm so the model can learn what meteorological conditions result in a certain power output. In most cases the training data would be the same as what would be used for the predictions. However the ICON-EU model is a forecasting model which is not entirely accurate, resulting in the model being trained on data which already has an error introduced. For this reason two other data sets which could give a better representation of the actual meteorological conditions at a certain time. Thus three sources for training the data are discussed, namely data from the ICON-EU model, measured data, and data from the reanalysis model era5.

4.4.1 ICON-EU training set

In general machine learning models perform best when trained on the same data as which it will make predictions with. Thus the first dataset on which the model will be trained is the ICON-EU dataset which makes predictions from 48 hours in the future until the next run. SO X_{48+i} where X is the data from the NWP model and i is the hours from zero until the next forecast run, this dataset will be called ICON-EU-48.

When making the D-2 forecast the ICON-EU model has a lot of uncertainty because the predictions are still far in the future. This uncertainty decreases as the forecast approaches to zero hours in the future it becomes easier to predict how the weather will be. Therefore for training the model on the ICON-EU dataset an aditional set is created with not the prediction of 48 hours in the future, but the predictions from the first prediction until the next model run. So from X_{0+i} where X is again the data from the NWP model and i is the hours from zero until the next forecast run, this dataset will be called the ICON-EU-zero.

4.4.2 Measured training set

The KNMI which has weather stations throughout the North sea to measure the current meteorological conditions. The windpark has a weather station in its proximity which can be used as a source for the meteorological realisation data. A downside to this source is that not all windparks in the north sea have a weather station is in its proximity. Thus this data can only be used for the windpark which have a weather station in its proximity without having to interpolate the data. Another downside of the dataset is that the data is rounded and does not contain decimals, making the data relatively imprecise. Due to this aspect the data might not be suitable for training the model.

4.4.3 Meteorological reanalysis training set

The second type of meteorological realisation data discussed is meteorological reanalysis data. Meteorological reanalysis models try to describe the recent occurred conditions in the atmosphere, land surface and the oceans. These models are made through the use of forecast models and data assimilation's systems such as weather towers to make a 'reanalysis' of these occurred conditions.

The main advantage of using reanalysis data is that it its data covers the whole world without gaps, making it possible to obtain information about meteorological conditions where no weather stations are in its proximity. Making it possible to use this data source for all windparks in the North Sea. Thus reanalysis is used throughout the world for research, educational, and commercial applications Ramon et al. [2019].

There are many different meteorological reanalysis models which are designed by different institutes. Analysing these and selecting the most suitable for the use case therefore is quite the challenge. Auspiciously Ramon et al. [2019] made a comparison between the most prominent reanalysis models and found that ERA5 had the most accurate values for the winds speeds. Wind speed is by far the most important parameter in the development of a power forecast for a wind park.

The DWD also has a reanalyses model next to their ICON NWP model which is called the REA6 model. Since the NWP model chosen also was made by the DWD the models might be able to work well with each other. However this the coordinate system used for this model has a rotated grid. It proved too much work to convert the rotated coordinate system to the coordinate system of the ICON-EU model, thus was decided to use the ERA5 model as the source for providing the reanalysis data.

The ERA5 model is created by the ECMWF, it combines a vast amount of historical observations and uses these to make estimations on how the meteorological conditions at the site are. Strengths of the model are that it has many variables, it has global data sets, and the data sets are relatively straightforward to handle.

The model also has some limitations which are: that the observational constraints per site can considerably vary and hence the reliability is location dependent, and the changing mix of observations, biases in observations, and models can introduce spurious variability and trends into the output of the ERA5 data UCAR [2016]

4.5 OTHER INPUT DATA

The power output of the wind park is not only defined by the current meteorological conditions and other factors also need to be taken into account. Thus this section will discuss other input data sources that can be beneficial to the model. The first type of input data discussed is the availability of the wind park.

4.5.1 Availability of the wind park (REMIT)

An important feature that needs to be introduced into the model is the availability of the wind park. When the wind park is in maintenance or has a defect it is not able to produce energy at its full capacity. The time span of these outages can range between a few hours for minor issues up to a few days for large repairs. Therefore it is crucial to introduce information about these outages to the model.

REMIT is an European regulation designed to increase the transparency and stability of the European energy markets while combating insider trading and market manipulation. The data contains information about whether the windpark is operational. When there is an outage or schedules maintenance the windpark needs to publicly publish this information to prevent insider trading. Autoriteit Comsument Markt [2020]

Legislation forces the owners of the wind park to publicly post data regarding maintenance online. Information about maintenance of the wind park and the time the outage takes needs to be posted at least a day ahead. The data about the availability will be used to modify the possible amount of maximum possible output of the wind park.

When the windpark is not fully operational the meteorological conditions can not explain the decrease in power production. If these data points remain in the dataset they may 'spoil' the data and the models would be falsely punished because of something they could not predict. Thus when there is a case of an outage which is stored in REMIT the data point is removed from the dataset.

4.5.2 Market data

The electricity market has a huge role in the production of energy. Once the energy price becomes too negative to make it profitable to keep on running the wind parks they can be curtailed. Therefore the energy prices could contribute in making the model more accurate.
As market data available on a D-2 forecasting window is severely limited, the use of market data can only reasonably used in explaining the past power output of a windfarm. This applies to spot trading data (Day-Ahead and Intraday), but also imbalance prices. As such, it has been decided to not use any market data within the scope of this thesis.

4.5.3 Physical model

In order to make a hybrid model type a statistical an physical model need to be combined. This can be done by using the output of the physical model discussed in chapter 2 as the input for the statistical models. Thus the implementation then will be through a back to back model by feeding the output of the physical model into to the statistical model.

4.6 CONCLUSION

A brief summary with the data that will be used is presented to conclude this chapter. For the energy related data the settlement data is used. To make the predictions the NWP model that will be used is the ICON-EU model. Also the meteorological data on which the models are going to be trained is selected. The data sources for the meteorological data are the measurements by the KNMI weather stations, ERA5 reanalysis data and the ICON-EU model. Finally data about the availability of the wind park and the output physical model is used to feed into the models.

Now the input data to train the model on is known. The optimal dataset needs to be acquired through feature selection. This is done in the following chapter.

5 FEATURE SELECTION AND FEATURE ENGINEERING

This chapter delves deeper into feature engineering and feature selection methods in order to obtain the most optimal data subset for the model within the available data. Feature engineering is the process of building new features from available data sources for training purposes. Then feature selection is the process of choosing the input variables which will feed and train the machine learning model. It is important since not all data sources models are relevant or less important in order to calculate the power output of the wind park. When a feature is irrelevant or only partially relevant it can harm the model performance, therefore choosing the right data set to feed the model has a huge impact on the performance of the model.

5.1 FEATURE ENGINEERING

This section delves deeper into how features can be engineered and which features where derived from the available data sources. It can be grouped into two different categories: namely transformations and aggregations.

A transformation acts on an input variable and converts it into a new feature. For example, the date-time stamp could be converted into the new feature hour to give the daily variations as an input. Aggregations are performed on input variables and uses them to calculate a new feature, e.g. the U and V component of the wind speed are used to calculate the wind speed and wind direction.

Once a new set of features has been created they need to be verified whether they have a possible influence on the predictions of the model. The verification of the engineered features will be done on the feature selection section. Now the process of feature engineering is cleared up, the created features are elaborated.

5.1.1 Polynomial features

To calculate the power from wind equation 5.1 is used. Where P is the power of the wind, ρ is the air density, A is the surface of the wind turbine, and v is the wind speed.

$$P = \frac{1}{2}\rho A v^3 \tag{5.1}$$

Since the correlation between the power output and the wind speed is to the power three it could be beneficial to include the squared wind speed and the wind speed to the power three. Thus the features 'Ws_ power2' for the wind speed squared and 'ws_ power3' for the wind speed to the third power have been created.

5.1.2 Zero generation features

In some conditions, the wind park will not generate any power. Either from the wind speed being too low or when the conditions become too harsh for the wind park so that it will be turned off.
The wind turbines start producing power from when the wind speed is higher than

the cut-in wind speed, which is 3.6 [m/s], this becomes the feature 'below_cutin'.

The wind can also become too strong for the turbine to operate safely. Thus when the wind exceeds the cut-out wind speed the turbine blades are pitched out of the wind and the wind turbine is put to a full stop. The wind speed at which this happens is at 25 [m/s], the feature created from this limit is: 'high_windspeed'.

The temperatures are also of great importance for the wind turbines. When the ambient temperature falls out of the operating temperature range the turbines will be put to a stop to prevent mechanical failure. The temperatures in which the turbines operate is between 0 and 31 [C^o]. Due to this operational range the features 'belowzero' and 'above31' were created.

5.1.3 Temporal features

Meteorological conditions have temporal trends e.g. the wind speeds is generally higher at night than during the day. Therefore it is wise to implement a certain time feature into the model. Three different features have been created to cover these temporal trends. First the feature 'Hour' has been created which ranges from 0 until 23 to represent each hour of the day covering the daily trends. The second feature created is the 'Month' feature which ranges from 0 to 11 to represent the month in a year and it covers the yearly trends. The third feature created is the 'year_ in_ percent' feature. Which if the product of which day of the year it is divided by the total amount of days in a year. This feature also covers the yearly trend although more detailed than the month feature.

5.1.4 Wind buckets

The direction of wind also has a major influence on the power production of the wind farm. The dominant wind direction of the location needs to be taken into account. Wake effects are phenomena that in a wind park the first wind turbines make the wind after it passes trough turbulent resulting in lower production for the turbines behind these. The strength of this phenomena differs per wind direction due to the orientation of the wind park Neustadter and Spera [1985]. Thus when grouping wind directions in so-called buckets the model might find it easier to find these correlations. Thus eight wind buckets have been created which each have a range of 45 degrees. Which resulted in the feature 'wd'

5.1.5 Air density

As seen in equation 5.1, the air density (ρ) one of the dependencies for calculating the maximum power wind can produce. Because this value can be calculated from the relative humidity, air pressure, and the air temperature.

5.1.6 Features created

The available data has been manipulated, which resulted in the following features:

These engineered features will now be put to the test in order to verify whether they have an positive influence on the model performance.

ws_power2
ws_power3
below_cutin
above_cutin
belowzero
above31
Hour
month
wd
rho

5.2 FEATURE SELECTION METHODS

In this section first, the possible methods for the feature selection will be elaborated. Followed by the selection of the method is chosen and applied for the models. Finally, the features of interest will be chosen. An exhaustive search for the best feature subset of a given data set is practically impossible in a lot of situations. Although a good estimation can be obtained through a variety of feature selection techniques. There are two main kinds of feature selection methods which will be elaborated, namely wrapper methods, and filter methods.

Features can generally be divided into three kind of categories being:

- Relevant: these features are of importance in finding the right output.
- Irrelevant features: These features do no play any part in finding the right output and therefore should be discarded.
- Redundant features: These are features which could be explained through other features which are already fed into the model.

Thus features are selected based on the principle of Occam's razor, which is to make the model as simple as possible while still being able to make accurate predictions. Non-informative features introduce more noise decreasing the output quality and want to be removed from the dataset.

The model benefits from feature selection through the following aspects Xue et al. [2016].

- The chance to over-fit the data is reduced since less noise is introduced to make throw the model off course.
- The accuracy of the model can improve because when there is less misleading data.
- The curse of dimensionality occurs when the dimensionality of the features space increases. Since the number of possible configurations grow exponentially this phenomena obstructs the model from reaching a higher accuracy since not all subsets can be tested which might result in stagnation in local optima.
- The computational force required to train the model is reduced along with the training time.

5.2.1 Wrapper method

Wrapper methods try to evaluate the different subset of features on the machine learning algorithms and only use the relevant features. It finds the optimal features trough training the model on a subset of the complete data set and selects the relevant parameters based on the model's score.

In order to find the optimal subset of features, the wrapper methods can use various methods. The two most used wrapper methods are the forward selection and the backward elimination.

The forward selection is the simplest method which starts with an empty set and greedily adds attributes one at a time. Thus the model is first trained on an empty data-set and then is trained for each possible combination of features. The main problem with forward selection is that it might not include interdependent features, since it ads features one at a time. On the bright side, it can locate small effective subsets early on, as the early iterations have few variables.

The backward elimination starts with all the features and greedily removes them one at a time Liu H [1998]. Thus the mode is first trained with all the features and then removes them in such a way all the combinations are tested and therefore is the complete opposite of forward selection. Therefore the pros and cons are turned around as well. The interdependencies are handled quite well, however early iterations are computationally expensive.

Feature wrappers often achieve better results compared to the filter methods due to the fact they are tuned to the specific interaction between a certain algorithm and the data it's trained on Karagiannopoulos et al. [2007]. Because different feature sets could be optimal for different algorithms. On the downside the require a lot of computational force since the model needs to be trained contiguously on the different subsets of data in order to find the optimal features.

Frénay et al. [2013] stated about feature selection: "It can be very computationally demanding since many prediction models with different feature subsets have to be built. Then, the results of the wrapper strategy lack generality as their use is limited to a specific regression model. To circumvent both problems, filter methods are often used in practice"

5.2.2 Filter methods

Filter feature selection methods use statistical techniques to evaluate the relationship between each input variable and the target variable, these scores are used as the basis to choose those input variables that will be used in the model. The features with the most relevance to the target variable will be kept and thus the features are which are irrelevant are filtered out before the models are trained.

Mutual information

Mutual information (MI) is a commonly used filter feature selection technique. MI is a measure between two random variables X and Y, which quantifies the amount of information that can be obtained from the target variable through the other random variable.

$$MI(X;Y) = \int_X \int_Y p(x,y) log \frac{p(x,y)}{p(x), p(y)} dxdy \tag{5.2}$$

The formula for how the MI is calculated can be seen in formula 5.2. Where p(x,y) is the joint probability density function of X and Y, p(x) and p(y) are the marginal density functions. Thus the MI determines how similar the joint distribution is to the products of the factored marginal distributions p(x,y). Beraha et al. [2019]

Pearson's coefficient for linear relation

The Pearson correlation coefficient (PCC) is a statistical tool that measures the linear relationship between two variables: X and Y. It's value ranges between minus one and one. A score of 1 means a total positive linear correlation when its zero it means no linear correlation, and when is minus one there is a fully negative linear correlation. The function for the Pearson's coefficient can be seen in equation 5.3.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{5.3}$$

Were $cov(X,Y)$ is the covariance between X and Y, σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

Spearman's rank coefficient

Spearman's rank coefficient (SRC) it measures the statistical dependence between the ranking of two variables and assesses the relationship between two variables using a monotonic function whether linear or not. When the SRC will be high when the rank between observations are similar and low then the observations are dissimilar. The function can be seen in equation 5.4

$$r_s = \rho_{rg_X, rg_y} = \frac{cov(rg_x, rg_y)}{\sigma_{rg_X} \sigma_{rg_Y}} \tag{5.4}$$

Where ρ is the PCC, $cov(rg_x rg_Y)$ is the covariance between the rank variables of X and Y, σ_{rg_X} is the standard deviation of the rank of X, and σ_{rg_Y} is the standard deviation of the rank of Y.

Filters are powerful yet simple methods to quickly select the most relevant features for the data set on which the model is trained Frénay et al. [2013]. The biggest advantages of filter methods are that they can be used with any machine learning algorithm. As well as that they are not computationally expensive, therefore a lot of features can be processed in a short period. This makes that filter methods are generally much faster than wrapper methods.

5.3 CHOOSING THE FEATURES

The relevance for the features and the parameters from the NWP model can now be verified. The process in obtaining the optimal subset of features uses the methods as described in the previous section.

The wrapper methods most likely will result in the best set of features for each algorithm at the expense of a high computational loss. On the other hand, the filter methods are able to quickly determine a suitable set of features which then is likely to be less optimal than the outcome of the wrapper method.

The wrapper method forward selection was initially chosen as the method that would determine the subset of features for each algorithm. This method would be able to find the best features whilst quickly giving an indication of which features performed well on each model.

However when the dimensionality of the data set increased and combinations of multiple features where tested the computational time the algorithm needed in order to obtain the increased so much that it would not be a practical solution. Since each feature added the number of simulations increased exponentially. Hence the decision was made to use a filter method which would create one data set which would be used for all algorithms.

The filter methods chosen is the mutual information method combined with SRC. Since MI is adequate choosing a subset of data which minimized the MAE in regression problems Frénay et al. [2013]. The downside with MI is that it does not look at the interdependencies between features. Thus when the most important features are selected with MI a correlation matrix with the SRC is made to verify which information is redundant. Table 5.1 show the results for MI feature selection were the cuttoff criterion was if a feature contributes less than five percent, the features that scored lower were dropped and are not depicted in the table.

Table 5.1: The results from the MI feature selection.

wind_speed_10	1.0244
ws_power2	1.0233
Calculated_power	1.0201
ws_power3	1.0158
instantaneous_wind_gust_10	0.9348
prec_in_year	0.5656
u_10	0.424610
v_10	0.267747
below_cutin	0.208835
month	0.145504
temperature_2	0.110900
mean_sea_level_pressure	0.090962
wind_direction_10	0.088155
rho	0.070096

A correlation matrix was made with the result from the MI feature selection. The now the redundant features which resembled each other too much were removed. The results from the feature selection method are seen below. These features are going to be used as input data for the model, all the other features which did not contribute positively have been removed.

ws_power2
wind_speed_10
Calculated_power
ws_power3
instantaneous_wind_gust_10
prec_in_year
u_10
v_10
below_cutin
month
temperature_2
mean_sea_level_pressure
wind_direction_10

5.4 CONCLUSION

From the data sources, various features have been engineered in order to create valuable input data for the machine learning model. Then the relevance of these features for the model was analysed through the filter feature selection methods which is called mutual information.

Since the filter method, MI does not check the interdependence between the features this might result in features containing similar information which can result in introducing noise in the models, distorting the learning process of the models. Thus to verify whether features are interdependent a correlation matrix is made in which the correlation between each feature is displayed. The method used to calculate the correlation matrix is the SRC due in order to find the non-linear relationships between the features.

The next step into designing the machine learning model would be to optimise the hyperparameters in the model. Which is done in the following chapter.

6 VALIDATION OF THE MODEL

The forecast-accuracy evaluation of the model is an important part to be able to verify its performance. The fact that the model makes a regression of a time series exacerbates the validation problem, due to the fact that time series do not hold on to the assumption of identically distributed independence of the datapoints. Identically distributed means that there are no overall trends and the data have the same probability distribution. In the time series the data points are also not independent from each other since they are connected to each other in such a way that past datapoints may contain information about future datapoints. For example when it rains the temperature decreases as well, so if it started raining at a certain time step at the next time step the temperature can be decreased.

An error estimation could be made for the model after training the model on the full data set. However this only gives an indication on how well the model performs on data it's trained on. The problem that arises here is that with this evaluation technique it is possible to over fit the data without knowing. A better way to validate the model is through cross validation, which is the process of dividing the dataset in an train and test set. The following techniques discussed state possible ways to validate the model in a better way, as well as a proposition is given on how to split the dataset into a train and test set.

6.0.1 Hold out method

The holdout method is one of the most basic cross validation (CV) techniques and splits the data up in a training set and a test set. The training set is then again divided into a training set and a validation set as depicted in figure 6.1. This first training set is used to train the model and the validation set is then used to tune the hyper parameters of the model. Once the hyper parameters are trained the model is trained on the full training set before it makes an accuracy evaluation of the model on the test set. The hold-out validation does not have a fixed persentage of data that needs to be held out for testing. However if the percentage is 10% or less then it might result in the model to overfit if the testing data differs greatly from the training data Yadav and Shukla [2016].

This method is already an improvement whilst comparing it with no separate training and testing set, albeit it still suffers from issues of high variance Prashant Gupta [2017]. Other arising problems are that once a part of the training set is used as a test set the reduction in training data poses a risk in under fitting the model and the important patterns which are stored in the training set are lost. To overcome these problems K-fold CV is introduced.

6.0.2 K-fold cross validation

K-fold CV divides the data in to K subsets and repeats the holdout method K times so that each of the K subsets are used as the test set while the remaining data is used for training the model as depicted in figure 6.2. Using K-fold CV the bias is significantly reduced since most of the data set is used for training the model. The variance of the model is also used since most of the data is also being used in the

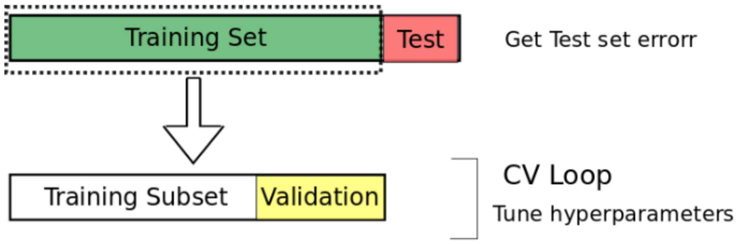


Figure 6.1: How to split the data set in the training set, validation set, and test set.

validation set Prashant Gupta [2017].

When repeating the process of splitting the data into a train and validation set multiple times, it allows multiple models to be trained the aggregated results of these models will provide a more robust estimate of the regression's accuracy estimation. The robustness of the forecast-accuracy evaluation increases although the computational time of training the models increases.

K-fold CV is already an improvement on the holdout method. However it still imposes problems because it only holds for independent identically distributed random variables which is not the case for the time series data used. Another reason K-fold CV is not an ideal way to validate a model that predicts time series data is the order in which the data is evaluated. When the forecast is based on a model trained by data from before the forecasted event, the forecast-accuracy evaluation is identical to the real-world-forecasting environment. However once data about events after the forecasted events are shown to the model is basically 'peeking at future' which pollutes the evaluation environment. Tashman [2000]

Altough, due to the complexity of the meteorological conditions the consequences of this matter are questionable. If the meteorological conditions of next month are known the weather of tomorrow can not be predicted using that data alone. Thus the assumption is made that due to the complexity of the data used k-fold cross validation does not pollute the test set.

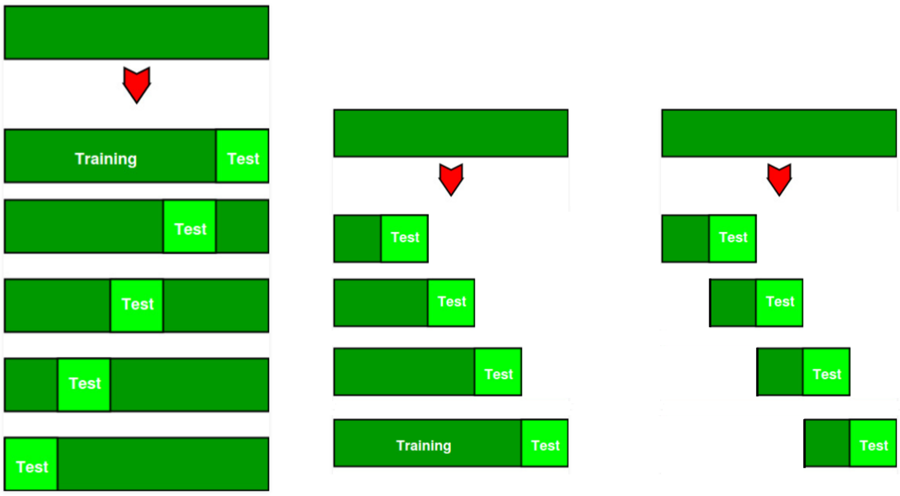
6.0.3 Time series cross validation

In order to overcome the problems in the previous subsection regarding the CV of the time series forecasting model nested CV also known as time series CV can be used. Just like with the K-fold CV the data set is split up in N parts. However these parts are used for the training and testing in chronological order. Figure 6.2 gives an overview of how the time series works. So for the first iteration the training set will be K1 and the test set will be K1 + 1. For the second iteration the training set is K1 + K2 and the test set K2 +1 and so on for the following iterations. The quality of the model is estimated by averaging the test scores of the subsets.

A nested CV procedure provides an almost unbiased estimate of the true error. Sudhir Varma [2006] Since the model is a rolling forecast more data is obtained over time. This increase in data can be used to train the model train and validate the model again improving the forecast evaluation accuracy. However the first batches might not be very representative since the amount of data the model has been trained on is still limited. Thus as the time series cross validation continues the estimation of the true error will become better as well as the model results.

6.0.4 Time series walk cross validation

The last CV method discussed is the time series walk CV. Which has lot of resemblance with the nested CV, although the full data set is not fully used when training the model. As seen in figure 6.2 time series walk discards previous training sets in further iterations. Because not the full training set is used the computational force is less that for the nested CV. Thus making it suitable for models which can be trained where large quantities of training data are available.



(a) K-fold cross validation (b) Time series cross validation (c) Time series walk cross validation

Figure 6.2: Visualisation of different cross validation methods.

6.1 SPLITTING OF THE DATASET

The total dataset as prescribed in chapter 5 contains data for the period of 01-01-2017 until 01-03-2020. In order to properly evaluate the performance of the model the dataset needs to be split up in a train and test set. Ideally the model would get as much training data as possible so it can learn all the existing patterns, however the test set needs to be large enough to make draw valid conclusions from the results.

To verify that the model is able to perform accurately during the seasonal changes a test set of a year needs to be held out in order to check the consistency of the model during the year. Therefore the period of 01-03-2019 until 01-03-2020 will be used to test the data on.

Now the training set and the test set are defined the straining set needs to be split up into the training and validation set in order to optimize the models. In order to make sure the models are not optimized on data it will be tested on the dataset the models will be trained on is from 01-01-2017 until 01-03-2019. To ensure the models are properly trained, K-fold cross validation with five folds will be used to ensure the models do not over fit the data.

For the cross validation the K-fold cross validation will be used. Due to the robustness it brings in the error estimation of the model. Time series cross validation could also have been a valid option, although its questionable on how to compare each batch with another since the first batch has less train data than the last batch trained.

6.2 CONCLUSION

The validation method that will be used to asses the model performance is the K-fold cross validation with a K of five. The train set ranges from 01-01-2017 until 01-03-2019 and the test set ranges from 01-03-2019 until 01-03-2020.

7 OPTIMIZATION OF THE MODELS

Now the input data for the models has been chosen the hyperparameters of the models can be tuned to optimize the models' performance and minimize the error criteria. This is one of the core components in machine learning and it's paramount that it is done correctly, thus this is done separately for each model. Hyperparameters are the parameters which can be arbitrarily changed by the user in order to enhance the models' performance, in contrast to the model parameters which are learned during the training of the model (e.g. decision trees in XGBoost). So hyperparameters define the structure of the model. The optimization of these hyperparameters is an optimization problem which aims to find the right combination to minimize the error of the model. This is done trough the Loss, which is the penalty for the model when it makes wrong predictions. The goal of training the model is to find the right model parameters which minimize the loss.

If the hyperparameters are not selected properly the model will not work properly and the data can be under- or over-fit as depicted in figure 7.1. The hyperparameters will be configured on the train and validation set, so the dataset that will be used is the ERA5 dataset which ranges from 01-01-2017 until 01-03-2019. In order to make the data unbiased a K-fold cross validation with five folds is used. The average score of the cross validated models will be used to decide which hyperparameters are optimal.

The models will not only be scored on the validation data but also on the training data. This is in order to check whether the models made a good generalisation of the underlying patterns in the data and were not overfit. Thus the models will be rated on their validation and train score. The final hyperparameter configuration of each model can be found in appendix A.

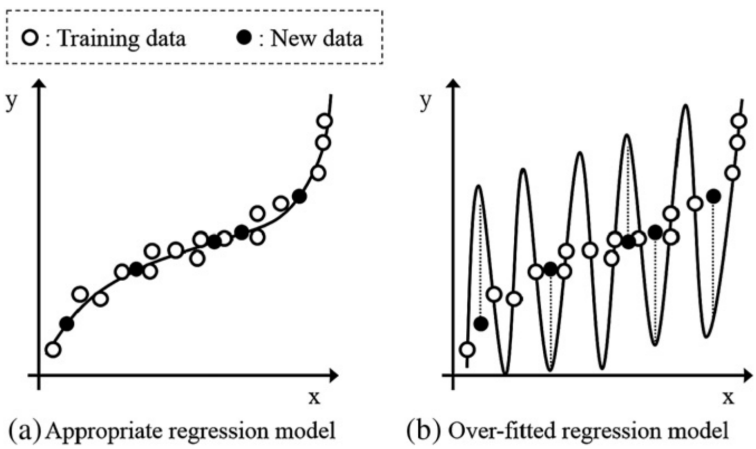


Figure 7.1: results of under or over-fitting of a machine learning model. Kaneko and Funatsu [2015]

7.1 OPTIMIZATION METHODS

There are various ways to find the right hyperparameters for each model. To find them in the right way a structured method needs to be followed.

7.1.1 Manual search

Manual search can be used when the data scientist already has experience with the algorithm. The hyperparameters are chosen based on an estimation of the right values. When the hyperparameters are initialised the model is trained and the results are evaluated, this is repeated until the desired result is obtained.

When properly used it can quickly give an estimation on what the right hyperparameters of the model are. However, this method will not likely result in an optimal configuration and can backfire when the algorithm is not understood well enough, resulting in a model that is not working properly.

7.1.2 Random search

With random search, a grid of hyperparameters is made. The model is then trained on a random combination of these hyperparameters. The hyperparameter combination which performs the best is then is set as its optimal configuration. Sometimes random search is the first step before grid search which is explained in the next paragraph.

7.1.3 Grid search

With grid search, a grid of hyperparameters is set up just like with random search. However in this case it is way more refined zooming in on the values that worked with random search. This way it is easy to verify whether a better combination of hyperparameters is possible. However, the disadvantage of this method is that the amount of combinations exponentially grows with each added parameter. Thus if six parameters need to be tuned and ten different combinations per parameter this would result in a 1,000,000 different combinations on which the model needs to be trained on.

7.2 OPTIMIZATION OF THE GRADIENT BOOSTING ALGORITHM

The gradient boosting algorithm has a wide range of hyperparameters that can be optimized. This makes the algorithm very flexible however, due to this flexibility the complexity of obtaining the right hyperparameters also increases. Finding the optimal hyperparameters can be a crux due to the interdependence of the hyperparameters if one is tuned the other also need to be tuned accordingly.

The following hyperparameters are of interest when optimizing the XGBoost model xgboost developers [2020].

1. Trees: The number of decision trees the model will create.
2. learning_rate: The rate at which the mode will learn characteristics from the data.
3. max_depth: The maximum depth of a tree, thus how much decisions steps can be in one tree. When the max depth is increased the model becomes more complex and prone to overfitting.

4. Gamma: Specifies the minimum loss reduction required for the tree to make a split. The higher the value the mode conservative the model will be.

5. min_ child_ weight: The minimum sum of instance weight needed in before a child is created. Meaning when the weight of the portion of a tree step results in a node the weight needs to be higher than min child weight in order for the building process to continue.

6. Sub-sample: Is the ratio of the training instances, when it is set to 0.5 means this that the algorithm would sample half of the data before the trees are grown.

7. co-sample_bytree: is the sub-sample ratio of columns when constructing a tree.

Aarshay Jain [2016] provides a detailed guide on how the hyperparameters can be optimized in a manner of grid search. The following steps have to be taken in order to obtain the optimal hyperparameter configuration.

1. According to the guide first a high learning rate needs to be selected, in this case 0.1 was used. Then using this learning rate the optimal number of trees needs to be selected.

2. The second procedure is to tune the tree specific parameters (max depth, min child weight, gamma, sub-sample, and co-sample by tree).

3. Once the optimal values for the tree specific parameters were selected the third step could be initiated to reduce the learning rate to its optimum and find the optimal amount of trees using grid search.

The model was optimized resulting in the hyperparamter configuration which is presented in apendix A. The model scores on the validation and the train set are shown in table 7.1. The train and the validation score are not identical, which might suggest that the model is slightly over fit. However this configuration resulted in the best test score and will thus be used.

Table 7.1: Optimization results of the XGB model

	Validation	Train
rMAE [%]	9.29	6.41

7.3 OPTIMIZATION OF THE SUPPORT VECTOR REGRESSION ALGORITHM

Compared to the XGboost model the SVR algorithm is a lot easier to optimize. Since there are fewer hyperparameters which can be tuned. The following hyperparameters need to be found in order to optimize the model.

- The kernel function: This can be seen in a SVR as a similarity function. Given an input the kernel gives an output of a similarity score which can be used to make a prediction. This is often referred to the kernel trick. Which is to solve a nonlinear problem with a linear solution through introducing higher dimensions to solve the problem. The kernel used for the SVR model was the radial basis function (rbf) kernel.

- The C value: This is the regularization parameter and determines the strength of the regularization of the training data.

- The gamma value: Is the kernel coefficient and determines how much influence a single training example has. The larger gamma is, the closer other examples must be to have any effect.

Kaneko and Funatsu [2015] states that a grid solution will give the best results although when choosing too many steps this will be computationally expensive. A grid search has been conducted with values for C ranging between $2^{-}3$ and 2^8 in 12 steps. The values for gamma ranged between 2^{-5} and 2^4 in 10 steps.

The resulting scores of the SVR model on the cross-validated dataset can be seen in table 7.2. Since the train and the validation set are close to one and other it can be concluded that the hyperparameter optimization was successful.

Table 7.2: Optimization results of the SVR model

	Validation	Train
rMAE [%]	8.68	7.34

7.4 OPTIMISATION OF THE NEURAL NETWORKS

Neural networks are highly complex and therefore very hard to optimise properly. The many hyperparameters need to be tuned making this a very time consuming, complex process. The following steps need to be taken in order to optimise the neural network Matthew Steward [2019].

1. The first step is to define a basic model.
2. When a basic neural network is initialised a grid search needs to be performed the learning rate to find the optimal value
3. The third step is to choose what optimiser and loss function to use.
4. The fourth step is to find the right batch size, number of epochs, the number of neurons per hidden layer, and the number of hidden layers which are ideal.

Once these steps are taken, the neural network should be optimised. What does need to be noted is that when a neural network is trained the randomisation factor ensures no model will be the same. This makes choosing the right hyperparameters extra difficult because during the cross-validation will perform better on a certain validation set.

7.4.1 Feedforward neural network

The first neural network that was optimised is the FFNN. Following the procedure to optimise the neural networks, it was optimised. the result from optimising the model are presented in table 7.3. As seen the validation and train score are nearly identical meaning the hyper parameters are chosen correctly and the model does not over fit the data.

Table 7.3: Optimization results of the FFNN model

	Validation	Train
rMAE [%]	9.12	9.09

7.4.2 Recurrent neural network

The recurrent neural network is optimised almost the same was as the feed-forward neural network. However, the main difference between the two is that there is time embedded in the model. This is done through time steps, just like the other hyperparameters these time steps will also be optimised trough grid search. The validation and train score for the RNN model can be seen in table 7.4. Like most cases the train score is a bit lower than the validation score, however since the difference is not substantial it is assumed that the hyper parameters are correctly chosen.

Table 7.4: Optimisation results of the RNN model

	Validation	Train
rMAE [%]	8.68	7.34

7.5 CONCLUSION

This chapter presented various methods in which machine learning models can be optimised. Using these methods the machine learning models that were optimised and its results validated through K-fold cross-validation.

All the machine learning models have been optimised through the selection of the right hyperparameters. Therefore its time to put the models to the test and use them to predict the power production on the test set as done in chapter 8.

8 RESULTS AND DISCUSSION

Now the models are optimised and trained they can be used to make predictions about the power generation in order to assess their performance. This chapter presents the results that were found with the different models and discusses their relevance.

The models will be tested on the test set which contains data from 01-03-2019 until 01-03-2020, while being trained on the full train set which ranges from 01-01-2017 until 01-03-2019. The test set is data the model has never seen before thus can be used in order to asses the quality of the models.

First, the general performance of the models trained on different data sets is covered where the average scores are presented. When the general performance of the models are known the edge cases are discussed in where the wind gust speed is higher than the cutoff speed or lower than the cut-in speed.

8.1 GENERAL MODEL PERFORMANCE

The first results of the model that are discussed are the general performance of the models. The total test data set of one year is used which ranges from 2019-03-01 until 2020-03-01. On this data set, the rMAE is calculated with the predictions from the model and the actual power output from the wind park.

The models trained on the ICON-EU data will first be evaluated. Secondly, the models trained on measurement data are discussed. The remaining models which are trained on the era-5 data set are discussed thirdly.

8.1.1 Model performance when trained on icon data

This section presents the results from the models trained on the ICON-EU data. First the findings of the models trained on the data set ICON-EU-48hour will be presented and secondly the findings when the models are trained on the ICON-EU-zero data set.

ICON-EU-48hour

Table 8.1 presents the results from the models when trained on the ICON-EU-48hours data set. Most models do not perform well enough with only the SVR being able to meet the set KPI.

Table 8.1: The models' scores when trained on ICON-EU-48hour data

	physical	XGB	SVR	FFNN	RNN
rMAE [%]	15.47	15.64	14.42	17.36	17.82

When looking at the box plot in figure 8.1 it becomes apparent that the SVR performs better than the other modes. The whiskers have a smaller spread meaning that it has fewer outliers and is the most steady model. The median of the SVR is also the lowest indicating that the it has the most predictions have the best score of all the models.

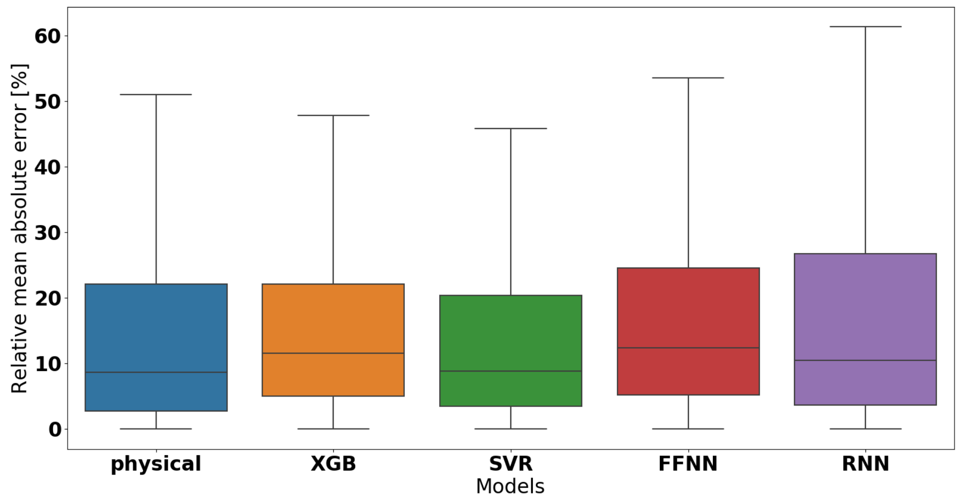


Figure 8.1: Box plot of the models' results trained on ICON-EU-48hour data.

The SO department calculates each month whether the forecast has met the required standard. Thus the monthly rMAE is calculated for the whole training set to see whether the models meet the set standard each month of the year. The results of the models are portrayed in figure 8.2. The SVR outperforms the other models each month of the year with the SVR and the Physical model having a similar score . The neural networks are not able to properly forecast the wind power of the model, with the RNN having an outlier with an rMAE of more than 22 percent in November.

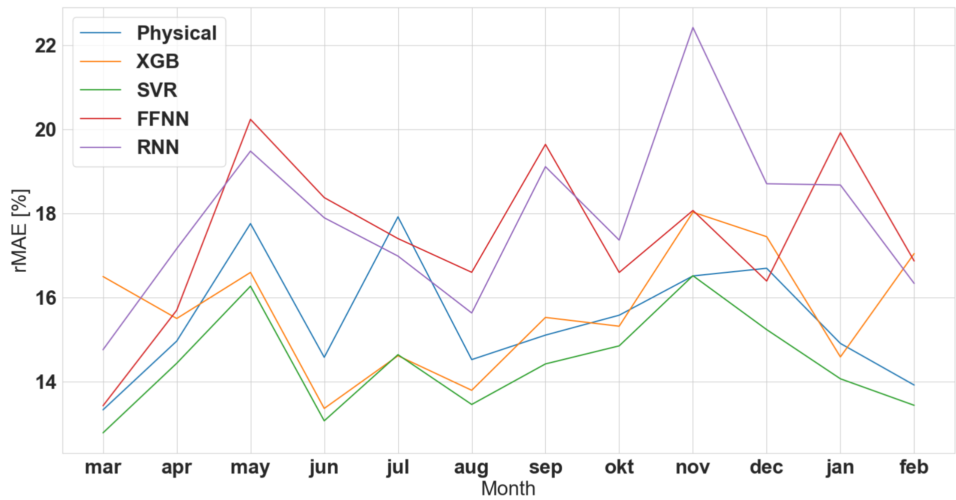


Figure 8.2: The rMAE score of the models trained on ICON-EU-48hour data for each month of the test set.

ICON-EU-zero

The model performance when trained on the ICON-EU-zero data can be seen in table 8.2. Three models conform to the standard set for the use case and have a rMAE lower than 15 percent. The models which conform to the standards are the XGB, SVR, and FFNN, these have a rMAE of 14.7, 14.35, and 14.99 respectively.

In figure 8.3 the box plot of the model results can be seen. As seen the boxes and whiskers of the different models do not differ too much from each other. Although can be seen that the box of the SVR is lower than the other and its median is the lowers of all the models.

Table 8.2: The models' scores when trained on ICON-EU-zero data

	physical	XGB	SVR	FFNN	RNN
rMAE [%]	15.47	14.7	14.35	14.99	15.42

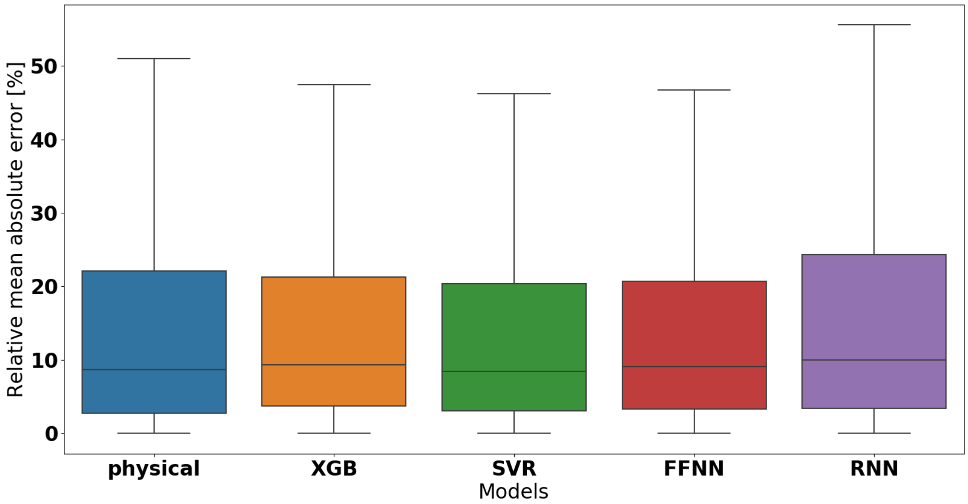


Figure 8.3: Box plot of the models' results trained on ICON-EU data.

The monthly scores of the model are depicted in figure 8.4. Here can be seen that none of the models are are meeting the KPIs that have been defined by SO, even though the average score for the XGB, SVR, and FFNN are good enough.

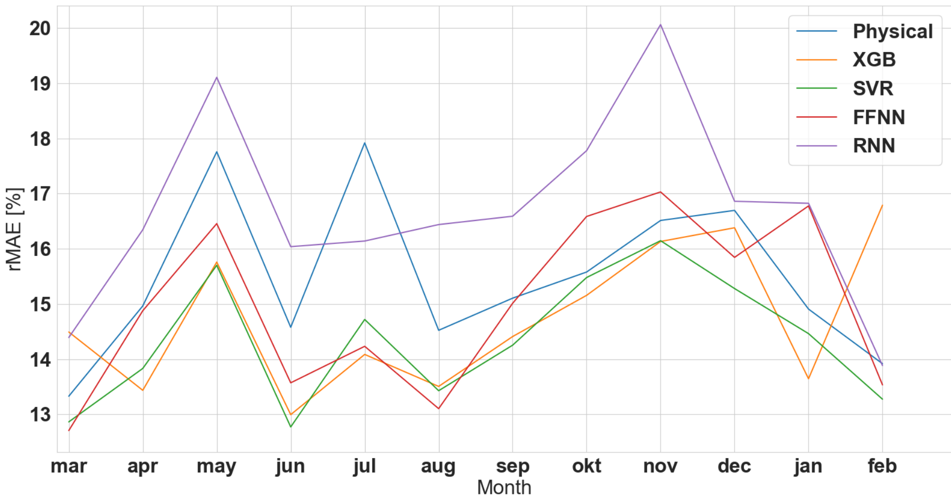


Figure 8.4: The rMAE score of the models trained on ICON-EU data for each month of the test set.

8.2 MODEL PERFORMANCE WHEN TRAINED ON MEASUREMENT DATA

Secondly, the performance of the models are analysed when trained on measured data. The model scores can be seen in table 8.3. When the model is trained on the measurements data none of the models are able to meet the set KPI. The errors for the XGB and SVR are substantially higher than for the FFNN and RNN. As seen the Physical model is the best performing model in this case. This indicates that the models are not able to make proper generalisations from the measurement data in order to predict the power production.

Table 8.3: The models scores when trained on measurement data

	physical	XGB	SVR	FFNN	RNN
rMAE [%]	15.47	18.64	19.30	16.29	16.54

Figure 8.5 displays the box plot of the models' errors when trained on the measurements data. As expected the boxes and whiskers for the XGB and SVR are the largest since their average error was the highest.

Figure 8.6 shows the rMAE score for the five models trained on the measurements data set for each month of the test set. As seen almost none of the models are able to meet the required standard and the physical model which is not trained outperforms the other models.

8.2.1 Models trained on the ERA5 data set

Lastly, the performance of the five different models trained on the ERA5 data set will be evaluated. The performance of each model on the data set can be seen in table 8.4 in the chosen error measurement tools. As seen in table 8.4, the SVR performs best with an rMAE of 14.89 and therefore is the only model which satisfies the KPI of an rMAE of 15 percent.

Table 8.4: The models performance when trained on ERA5 data

	physical	XGB	SVR	FFNN	RNN
rMAE [%]	15.47	15.35	14.89	15.26	16.16

Figure 8.7 depicts box plots of the different models trained on the ERA5 data source. Here can be seen that the FFNN makes the most steady predictions because the box is slightly lower than for the other models and its top whisker is lower as well.

When looking at figure 8.8, where the monthly rMAE scores for the models trained on the era5 data set are shown. It can be seen that the scores are heavily fluctuating along with the quality of the forecast. Which can be seen through the score of the physical model. When the physical model performs well the other models performs relatively well and vice-versa.

8.2.2 Comparing the models

When it comes to choosing the best data set on which the models are trained it can easily be seen that the measured data set can be discarded. The models trained on this data set have a significantly higher rMAE than when trained on the other data sets.

The scores of the models when trained on the era5 data set or the ICON-EU data set are relatively close to each other. With the best score for the models trained on the era5 data set being a rMAE of 14.89 and when trained on the ICON-EU data set 14.35. However when looking at graphs: 8.8, and 8.4 it can be seen that the monthly scores for the models are better when trained on ICON-EU data. What can be seen is that when the model is trained on ERA5 data set the model accuracy is coupled stronger to the accuracy of the models than when trained on the ICON data. This is seen through the physical model which is a direct representative of the NWP accuracy since it is not trained. When trained on era5 data the model scores are relatively worse when the NWP is less accurate than compared to when the models

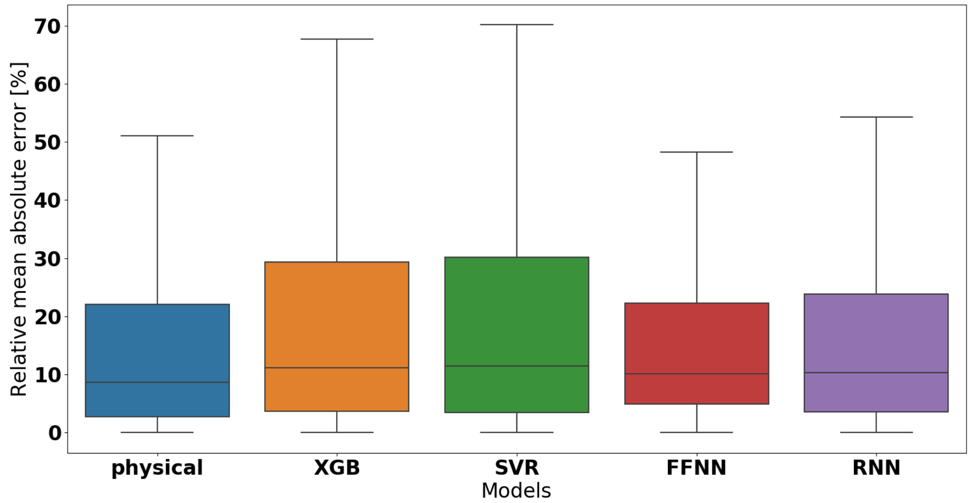


Figure 8.5: Box plot of the model results trained on the measurement data.

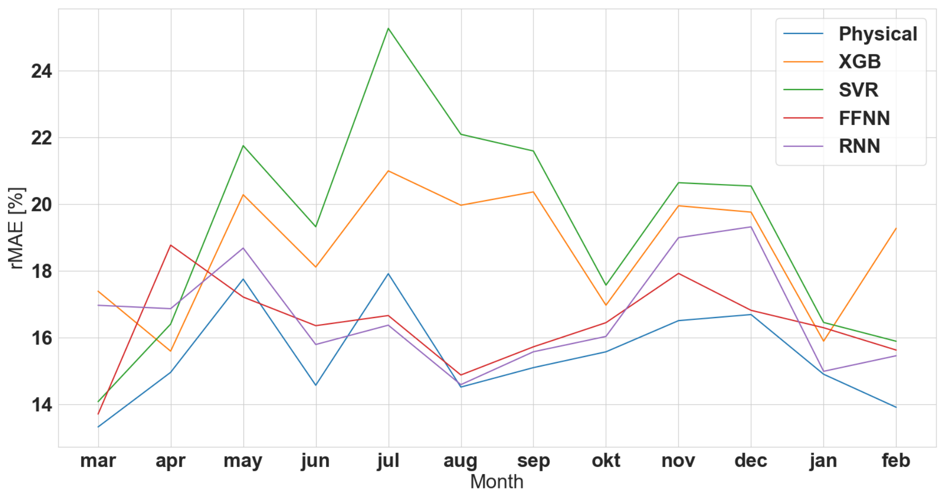


Figure 8.6: The rMAE score of the models trained on ICON-EU data for each month of the test set.

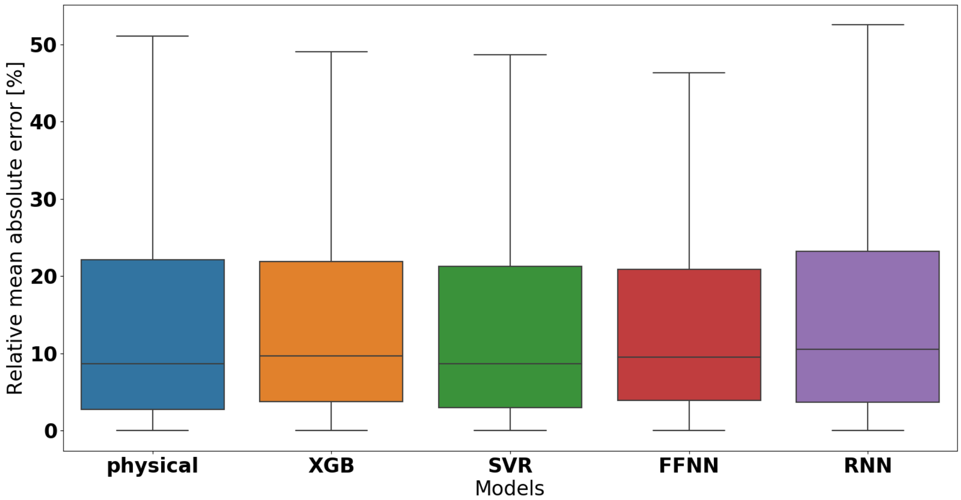


Figure 8.7: Box plot of the model results trained on era5 data.

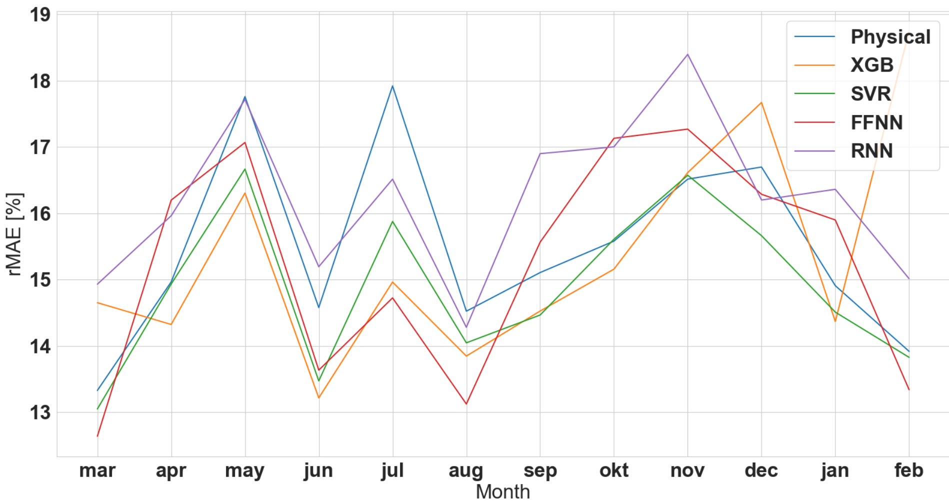


Figure 8.8: The rMAE score of the models trained on era5 data for each month of the test set.

are trained on the ICON data set. This behaviour can be explained through the fact that the models are able to recognise patterns in the ICON data set which cannot be found in the era5 data set.

Thus can be concluded that the training data set which is most beneficial to the model is the ICON-EU data set.

Now the most suitable training set is found, the best algorithm can be selected. When looking at the model scores in table 8.2 the XGB, SVR and the FFNN meet the required KPI. However, the project aims to design a model that can make the best forecast on average. Thus the most suitable model for this application would be the SVR because its mean rMAE is the lowest with a value of 14.35%.

8.3 EDGE CASES

When the general performance of the model is discussed, it is interesting to look at special cases which might be troublesome for the model. In these so-called edge cases, the models trained on the ICON-EU data set will be discussed since it showed the best results.

The first edge case looked at is the case for high wind speeds which are above the cut off wind speed, followed by cases where the wind speeds are below cut-in speed.

8.3.1 High wind speeds

The first edge case looked at is for cases where the predicted wind gusts are higher than the cut off speed where the wind gusts are higher than 25 $\frac{m}{s}$. These cases might be troublesome for the model to predict since there is a high uncertainty on whether the turbines are going to be shut off. Thus since the wind turbines will either produce at full capacity or none at all give these cases an almost 100% uncertainty.

In figure 8.9 the predictions of the models are depicted for the cases where the wind gusts are higher than the cutoff speed. It shows that the wind turbines are only shut off completely once around data point 120. The only model that was able to predict this outage was the physical model. For these points, almost all the other models knew a reduction in power production was coming although they were not able to respond as drastically.

Because these high wind speeds do not occur very often it might be possible that the machine learning models were not adequately trained for these cases and therefore not able to recognise when the wind park would be turned off.

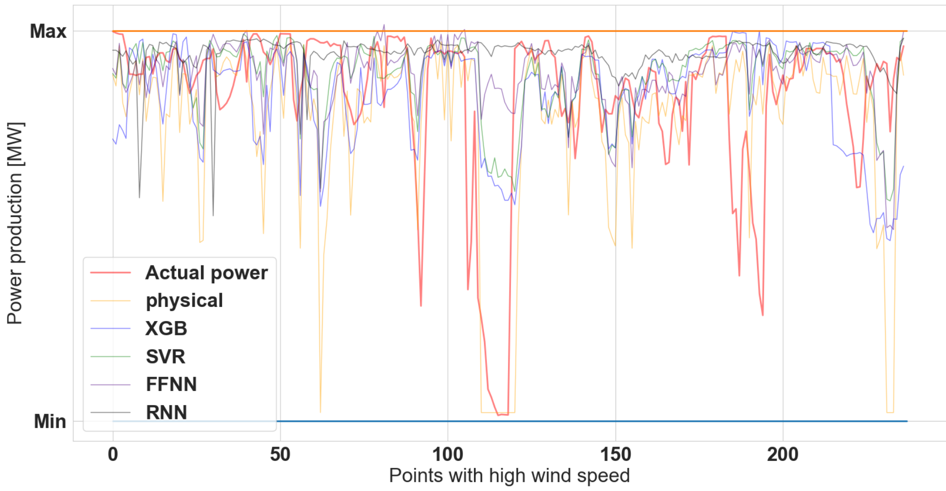


Figure 8.9: Model predictions for when the wind gusts are higher than the cutoff speed.

Table 8.5 shows the model scores for the edge cases with high wind speeds. Here can be seen that the SVR is yet again the best model. Another remarkable fact is that even though the high uncertainty all the MAE values for this edge case were lower than the general MAE.

What can be seen as well is that even though the physical model was able to know when the wind farm would complete shut down its score whilst predicting the power generation at high wind speeds was one of the lower ones. The main reason for this low score is due to the forecast errors since the model predicted no generation for some cases while actually the power generation was at its maximum.

Therefore the ability of the model to completely turn off the power generation acts like a double-edged sword.

Table 8.5: Models scores for high wind speed

	physical	XGB	SVR	FFNN	RNN
rMAE	14.08	14.45	12.40	13.24	13.27

8.3.2 Low wind speeds

Now the behaviour of the models are known at high wind speeds it is interesting to have a look at the other side of the spectrum and look at their behaviour for low wind speeds. Thus the behaviour of the models are monitored for when the predicted wind speed is lower than the cut-in wind speed.

Figure 8.10 depicts the models' predictions at low wind speeds. What can be seen is that the predominant mistake made is through the error in the NWP model. Since there are various large spikes in the actual power production where the predicted power production of the physical model stays around zero.
Another interesting observation is when looking at figure **??**. It can be seen that the RNN was able to partially predict the peak in power production a few times despite the error in the input data.

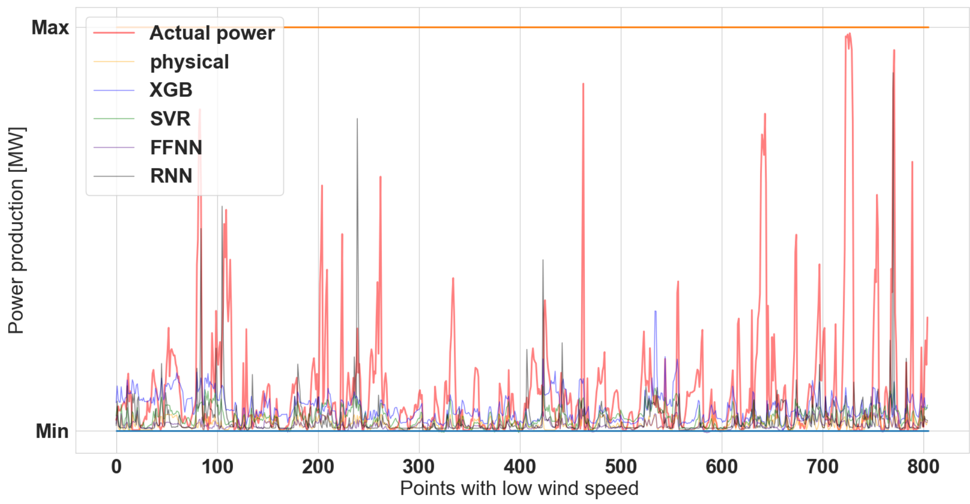


Figure 8.10: The error of the model when prediction for low wind speeds.

The model scores for low wind speeds can be seen in table 8.6. Now the XGB model is able to make the most accurate predictions for these kind of conditions. Even though the RNN is able to predict high power consumption without an accurate forecast it still does not rank high. This is because sometimes it also predicts these spikes at wrong times lowering its accuracy.

Table 8.6: Model scores for low wind speeds

	physical	XGB	SVR	FFNN	RNN
rMAE	53.16	47.95	49.13	54.34	52.17

8.4 CONCLUSION

Now the results of the models are analysed a conclusion can be drawn about the most beneficial data set and what algorithm works best for forecasting the power generation of the wind park.

The best model results were obtained through the use of training the models on the ICON-EU data set. Therefore the conclusion can be made that the most suitable training data set was the same as the testing set.

Although has to be noted that the differences between the results of training the model on the ICON-EU or on the ERA5 data set were not large. Therefore it could be possible to substitute the ICON-EU train set for the ERA5 train set once an ICON-EU data set can not be acquired.

When it comes down to the algorithm that makes the best predictions, it can be clearly stated that the SVR is able to make a prediction with the lowest average error. The rMAE of the SVR was best when the models where trained on the ICON-EU data and the ERA5 data.

9 CONCLUSION

Now all the results of the models are analysed the research questions can be answered and all the conclusions can be drawn. The research questions that were derived to come up with the best possible forecasting model are revised so they can be answered. These answers combined then answer elucidate the main research question which is: How should a D-2 energy generation forecasting model of an offshore wind park be designed and built?

What requirements does a D-2 electricity generation forecast model need to meet, to conform to the use cases of TenneT TSO B.V.?

The model needed to predict the average power production of the wind park at an hourly interval. The rMAE error measurement tool was used to asses the model performance. To make it possible to compare wind parks of different sizes. Here the MAE was divided by the difference between the minimum and maximum power production. With the error measurements, tools selected the KPI on which the forecasting model needed to conform to was also derived, the maximum rMAE was set to 15 %.

What input data are beneficial to the machine learning forecasting model?

The data source used that contained information about the power production of the wind park was the settlement data. Which is measured by the EMS system and then checked by the SCADA system. The ICON-EU model proved to meet the desired requirements and was chosen as the NWP model for using its data to make the power predictions. The other used data sources to train the models on were from: the reanalysis model ERA5, measured data from a weather station near the wind park, and forecast data from the ICON-EU model. When put to the test it became evident that the most suitable data source to train the model on was from the ICON-EU model, since the highest model accuracy was obtained using this data set. The measurements data set was not beneficial since none of the models were able to conform to the set KPI. When the models were trained on the era5 data set, the SVR had a rMAE that did not score far below the SVR trained on ICON-EU data. Thus the when a training data set of the ICON-EU model would not be available, an era5 data set could be a suitable replacement.

What type of models and algorithms are effective at forecasting the D-2 power production of an offshore wind park?

In order to answer this question, five different models were made. The first being a physical model using windpowerlib, which was also functioned as a baseline model due to its direct correlation between the meteorological conditions and its simplicity. The other four models where machine learning algorithms being: A XGBoost model, a Support Vector Regression algorithm, a Feed-Forward Neural network, and Recurrent neural network.

The models that met the set KPI were the XGBoost model, the SVR, and the FFNN. Their rMAE scores are: 14.70%, 14.35%, and 14.99% respectively. Thus can be concluded that the SVR algorithm is most suited to make predictions about the power production of an offshore wind farm. However has to be noted that despite the average rMAE was lower than the set KPI, the rMAE calculated per month was not able to meet the KPI in some instances. This was largely because when the NWP model's predictions became less accurate predictions of the SVR also became less accurate.

10 RECOMMENDATIONS

Due to the limited time in which the thesis needed to be concluded various topics were not explored as extensively as desired. During the research it became evident that this topic can be explored from different angles this is why the following recommendations are made.

10.1 ICON-EU DATA SET

The ICON-EU purchased data set only contained two runs per day, instead of the 4 runs which can be downloaded from the internet. This resulted in the forecast being 48 + 12 hours instead of 48 + 6 hours. It would be wise to keep on storing this data and reevaluate the model once a test set with significant length has been obtained.

Even though for the month where the data was available the scores have been calculated. Table 10.1 display's the rMAE of the models when the ICON-EU data with 4 runs per day was used for testing. As seen the results improved significantly when compared with the scores from table 8.2. However since this is only a month of training data it could be possible that the NWP model was really accurate this particular month, thus no valid conclusions can be drawn from this result and it can only give an indication about what is to come.

Table 10.1: The models performance when tested om ICON-EU data with four runs.

	physical	XGB	SVR	FFNN	RNN
rMAE [%]	12.10	12.58	11.98	12.53	12.85

10.2 MODEL OPTIMISATION

The assumption was made that because of the resemblance between the data sets the hyperparameters of the models did not need to be optimised for each data set. Even though the data looks similar it might be wise to recalculate the hyperparameters when using a different data set.

10.3 MAKE FORECAST FOR ADDITIONAL PARKS

Sometimes the wind park showed behaviour which could not be explained by the meteorological conditions. When the forecast of a whole fleet of wind park would be analysed the forecasts can be compared, in order to check what the cause of this behaviour could be.

BIBLIOGRAPHY

A. Foley, P. Leahy, E. M. (2010). Wind power forecasting prediction methods. *Conference on Environment and Electrical Engineering*, 9:61–64.

Aarshay Jain (2016). Complete guide to parameter tuning in XGBoost with codes in python. https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/. Online; accessed 2020-06-25.

Autoriteit Comsument Markt (2020). Verplichtingen onder REMIT. https://www.acm.nl/nl/onderwerpen/energie/de-energiemarkt/remit/verplichtingen-onder-remit. Online; accessed 2020-07-19.

Balluff, S., Bendfeld, J., and Krauter, S. (2015). Short term wind and energy prediction for offshore wind farms using neural networks. In *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, pages 379–382.

Barak Rishan (2018). Data driven statistical models vs process droven physical models. https://medium.com/@b.bhaskaran/data-driven-statistical-models-vs-process-driven-physical-models-340f4dd4eea8. Online; accessed 2020-05-04.

Beraha, M., Metelli, A. M., Papini, M., Tirinzoni, A., and Restelli, M. (2019). Feature selection via mutual information: New theoretical insights. *CoRR*, abs/1907.07384.

Bhaskar, K. and Singh, S. (2012). Awnn-assisted wind power forecasting using feed-forward neural network. *IEEE transactions on sustainable energy*, 3(2):306–315.

Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint arXiv:1809.03006*.

Briggs, W. (2016). *Statistical and Physical Models. In: Uncertainty*. Springer, Cham.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA. ACM.

D. Reinert, F. Prill, H. F. M. D. and Zängl, G. (2019). *Database Reference Manual for ICON and ICON-EPS*. Deutscher Wetterdienst, Offenbach am Main.

Damousis, I. G., Alexiadis, M. C., Theocharis, J. B., and Dokopoulos, P. S. (2004). A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion*, 19(2):352–361.

ECMWF (2019). *IFS Documentation -cy46r1, Part V: Ensemble Prediction System*. ECMWF, Shinfield Park Reading RG2 9AX United Kingdom.

ECMWF (2020). Medium-range forecasts. https://www.ecmwf.int/en/forecasts/documentation-and-support/medium-range-forecasts. Online; accessed 2020-03-09.

Epexspot (2020). Basics of the power market. https://www.epexspot.com/en/basicspowermarket#day-ahead-and-intraday-the-backbone-of-the-european-spot-market. Online; accessed 2020-08-20.

Eurostat (2020). Renewable energy statistics. https://ec.europa.eu/eurostat/statistics-explained/index.php/Renewable_energy_statistics#:~:text=The%20EU%20seeks%20to%20have,each%20of%20the%20Member%20States. Online; accessed 2020-08-20.

Foley A, Leahy P, M. A. M. E. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37:1–8.

Fouladfar, M., Loni, A., Tookanlou, M., Marzband, M., Godina, R., Al-Sumaiti, A., and Pouresmaeil, E. (2019). The impact of demand response programs on reducing the emissions and cost of a neighborhood home microgrid. *Applied Sciences*, 9:2097.

Friedman, J. H. (2002). Greedy function approximation: A gradient boosting machine. *Ann. Statist*, 29(5):1189–1232.

Frénay, B., Doquire, G., and Verleysen, M. (2013). Is mutual information adequate for feature selection in regression? *Neural Networks*, 48:1 – 7.

Goretti, G., Duffy, A., and Lie, T. T. (2017). The impact of power curve estimation on commercial wind power forecasts — an empirical analysis. In *2017 14th International Conference on the European Energy Market (EEM)*, pages 1–4.

Géron, A. (2019). *Hands on Machine Learning with Scikit Learn Keras and TensorFlow*. O'Reilly Media Inc, Sebastopol.

Haas, S., Schachler, B., and Krien, U. (2019). windpowerlib - a python library to model wind power plants.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Jay Alammar (2019). A visual and interactive guide to the basics of Neural Networks. http://jalammar.github.io/visual-interactive-guide-basics-neural-networks/. Online; accessed 2020-09-03.

Joseph Rocca (2019). Ensemble methods: bagging, boosting, and stacking. https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205. Online; accessed 2020-09-03.

Juban, J., Fugon, L., and Kariniotakis, G. (2008). Uncertainty estimation of wind power forecasts. In *Proceedings of the European Wind Energy Conference EWEC08, Brussels, Belgium*.

Kalverla, P., Steeneveld, G., Ronda, R., and Holtslag, A. (2019). Evaluation of three mainstream numerical weather prediction models with observations from meteorological mast ijmuiden at the north sea. *Wind Energy*, 22:34–48.

Kaneko, H. and Funatsu, K. (2015). Fast optimization of hyperparameters for support vector regression models with highly predictive ability. *Chemometrics and Intelligent Laboratory Systems*, 142:64–69.

Karagiannopoulos, M., Anyfantis, D., Kotsiantis, S., and Pintelas, P. (2007). Feature selection for regression problems. *Proceedings of the 8th Hellenic European Research on Computer Mathematics & its Applications, Athens, Greece*, 2022.

Khairalla, M. A., Ning, X., Al-Jallad, N. T., and El-Faroug, M. O. (2018). Short-term forecasting for energy consumption through stacking heterogeneous ensemble learning model. *Energies*, 11(6):1605.

KNMI (2020). MS Windows NT Kernel Description. https://www.knmi.nl/kennis-en-datacentrum/uitleg/weermodellen. Online; accessed 2020-03-09.

Kong, X., Liu, X., Shi, R., and Lee, K. Y. (2015). Wind speed prediction using reduced support vector machines with feature selection. *Neurocomputing*, 169:449–456.

L. Bengtsson, U. Andrae, T. A. (2017). The harmonie-arome model configuration in the aladin-hirlam nwp system. *Monthly Weather Review*, 145:1919–1935.

Landry, M., Erlinger, T. P., Patschke, D., and Varrichio, C. (2016). Probabilistic gradient boosting machines for gefcom2014 wind forecasting. *International Journal of Forecasting*, 32(3):1061 – 1066.

Lei, M., Shiyan, L., Chuanwen, J., Hongling, L., and Yan, Z. (2009). A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915 – 920.

Liu, H., Shi, J., and Qu, X. (2013). Empirical investigation on using wind speed volatility to estimate the operation probability and power output of wind turbines. *Energy Conversion and Management*, 67:8–17.

Liu H, M. H. (1998). *Feature Selection for Knowledge Discovery and Data Mining*. Springer US, Boston, MA.

Madsen, H., Kariniotakis, G., Nielsen, H., Nielsen, T., and Pinson, P. (2004). A protocol for standardizing the performance evaluation of short-term wind power prediction models. In *Proceedings of the 2004 Global Windpower Conference and Exhibition*. 2004 Global Windpower Conference and Exhibition ; Conference date: 28-03-2004 Through 31-03-2004.

Makoto. Nakaya, Xinchun. Li (2018). Physical-Statistical hybrid model for MIRROR PLANT. https://web-material3.yokogawa.com/rd-te-r05601-005.pdf?_ga=2.36086051.2084764254.1588597845-719819792.1588597845. Online; accessed 2020-05-04.

Matthew Steward (2019). Simple guide to hyperparameter tunding in neural networks. https://towardsdatascience.com/simple-guide-to-hyperparameter-tuning-in-neural-networks-3fe03dad8594. Online; accessed 2020-07-19.

National Center For Environmental Prediction (2020). Global Forecasting System. https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs. Online; accessed 2020-03-17.

Neustadter, H. and Spera, D. (1985). Method for evaluating wind turbine wake effects on wind farm performance.

Niayifar, A. and Porté-Agel, F. (2015). A new analytical model for wind farm power prediction. In *Journal of physics: conference series*, volume 625, page 012039. IOP Publishing.

Prashant Gupta (2017). Cross-Validation in Machine Learning. https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f. Online; accessed 2020-04-16.

Ramon, J., Lledó, L., Torralba, V., Soret, A., and Doblas-Reyes, F. J. (2019). What global reanalysis best represents near-surface winds? *Quarterly Journal of the Royal Meteorological Society*, 145(724):3236–3251.

Ron Wiseman (2017). How do I draw wind turbine power curve? https://www.quora.com/How-do-I-draw-wind-turbine-power-curve. Online; accessed 2020-04-02.

Shokrzadeh, S., Jafari Jozani, M., and Bibeau, E. (2014). Wind turbine power curve modeling using advanced parametric and nonparametric methods. *IEEE Transactions on Sustainable Energy*, 5(4):1262–1269.

Sideratos, G. and Hatziargyriou, N. D. (2007). An advanced statistical method for wind power forecasting. *IEEE Transactions on Power Systems*, 22(1):258–265.

Sudhir Varma, R. S. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(91):1471–2105.

Surya Priy, Abhishek Rajput (2019). Fuzzy logic, introduction. https://www.geeksforgeeks.org/fuzzy-logic-introduction/. Online; accessed 2020-09-04.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450.

Tennet TSO (2020). T prognoses: nut en noodzaak. https://www.tennet.eu/fileadmin/user_upload/Company/Publications/Technical_Publications/Dutch/TenneT_Brochure_T-prognoses_AUG2016_NL.pdf. Online; accessed 2020-03-23.

Ting, K. M. and Witten, I. H. (1999). Issues in stacked generalization. *Journal of artificial intelligence research*, 10:271–289.

Tom Sharp (2020). An introduction to support vector regression (SVR). https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2. Online; accessed 2020-05-07.

UCAR (2016). Atmospheric Reanalysis: Overview Comparison Tables. https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables. Online; accessed 2020-03-23.

UNFCCC (2016). The Paris Agreement. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement. Online; accessed 2020-07-18.

Vadim Smolyakov (2017). Ensemble learning to improve machine learning results. https://blog.statsbot.co/ensemble-learning-d1dcd548e936. Online; accessed 2020-09-03.

Wang, X., Guo, P., and Huang, X. (2011). A review of wind power forecasting models. *Energy procedia*, 12:770–778.

Wim van 't Hof (2016). Energy transition in the Netherlands - phasing out of gas. https://ec.europa.eu/energy/sites/ener/files/documents/01.b.02_mf31_presentation_nl-fuel_switch-vanthof.pdf. Online; accessed 2020-07-09.

Wu, W., Chen, K., Qiao, Y., and Lu, Z. (2016). Probabilistic short-term wind power forecasting based on deep neural networks. pages 1–8.

xgboost developers (2020). python API reference. https://xgboost.readthedocs.io/en/latest/python/python_api.html. Online; accessed 2020-07-19.

Xu, Q., He, D., Zhang, N., Kang, C., Xia, Q., Bai, J., and Huang, J. (2015). A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. *IEEE Transactions on sustainable energy*, 6(4):1283–1291.

Xue, B., Zhang, M., Browne, W. N., and Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4):606–626.

Yadav, S. and Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In *2016 IEEE 6th International conference on advanced computing (IACC)*, pages 78–83. IEEE.

A MODEL OPTIMIZATIONS

This chapter discusses what the hyperparameter configuration of the models are along with their corresponding test score.

A.1 HYPER PARAMETERS OF THE XGBOOST

Following this guide the xgboost algorithm was optimized. The following hyperparameters have been found.

- Trees: 130
- Learning rate = 0.05
- Max depth = 5
- Min child weight = 5
- Gamma = 0
- Subsample = 0.7
- Co-sample by tree = 0.7

A.2 HYPER PARAMETERS OF THE SVR

The optimal values that have been found for the SVR algorithm with the rbf kernel function are:

- C: 64
- Gamma: 0.0625

A.3 HYPER PARAMETERS OF THE FFNN

The optimal values for the hyperparameters that were found are:

- Epochs: 50
- Batch size = 20
- Hidden layers = 4
- Neurons = 70
- Learning rate = 0.001
- Activation function = relu
- Loss function = Mean absolute error
- Optimizer = rmsprop

A.4 HYPER PARAMETERS OF THE RNN

Following the procedure to find the optimal hyperparameters for the neural networks the following values were found.

- Epochs: 20
- Batch size = 20
- Hidden layers = 3
- Neurons = 100
- Learning rate = 0.005
- Activation function = relu
- Loss function = Mean absolute error
- Optimizer = rmsprop
- Time steps = 12

COLOPHON

This document was typeset using $\LaTeX$. The document layout was generated using the arsclassica package by Lorenzo Pantieri, which is an adaption of the original classicthesis package from André Miede.