Research Project - Draft 1:

"Predicting Electricity Prices with power plant outage and alternative energy supply data"

Abstract: The authors apply traditional and machine learning models to electricity price forecasting with power outages as one of the input variables. For the Irish electricity markets in the latter half of 2020, the paper finds a relation between power outages and price exists. However, only two out of three models improve significantly upon inclusion of power outages as an input feature. The authors also find that the random forest type model and multi-layer perceptron model perform similarly. Although these models both outperform a naive benchmark model, they do not outperform the ARX model.

Keywords: Electricity price, Forecasting, Power outages, Neural network, Random forest, ARX

Professor:
S. Borovkova

Group 4:
Daniel van Duin – (2578439)
Camilo Suarez – (2629514)
Maarten van Twisk – (2578039)

1. Introduction

1.1 Disclaimer

The topic at hand was brought to the attention of the authors by Nabla Analytics. Nabla Analytics is a data analytics company providing consulting services in the energy industry. Its customers are primarily traders in the electricity market. The goal set out in this paper is twofold: to assess the relationship between power outages and electricity price movements; and to forecast electricity prices using power outages and other data. The data used in this research has been supplied by Nabla Analytics, and is specific to the Irish electricity market.

1.2 Research question

Theoretically, we can surmise that an increase in power outages is correlated with an increase in price due to supply constraints. Summarizing the goals set out, the research question at hand is set up as follows: *"What is the effect of power plant outages on electricity prices and can future electricity prices be forecasted using future power plant outage data and alternative energy supply data?"*

1.3 Relevance

The practical relevance of this paper is twofold. First, the ability to anticipate price movements is useful to a trader in any type of market, including the electricity market. If a trader can accurately forecast future prices, this means they can take a position in order to profit from the spread.

Wholesale participants also make use of future prices (forecasts), mainly in decision making and risk management. (Weron, 2006). Electricity price forecasting is an important component of optimizing operational and planning activities of wholesale participants in the electricity market (Zareipour, 2012). Its use is key in determining the optimal bid.

1.4 Electricity market structure

Firstly, it is important to understand the electricity markets. Electricity markets are very complex markets. Since the end of the previous century restructuring, privatization, and deregulation in the power sector have changed the traditional setting of a governmental and monopolistic power sector (Bhattacharya, 2001). Because of this change in the power sector electricity prices in many countries (like Ireland) are now determined by the supply and demand of electricity prices. Moreover, these electricity prices are now traded under market rules using spot and derivatives contracts (Weron, 2014).

Lago et al (2018) illustrate several dynamics that make the market so complex: balance between production and consumption; dependence consumption on time; power generation influenced by weather conditions; and influence of neighbouring markets (through interconnectors). Weron (2014) also mentions extreme volatility in prices and non-stationarity (although we do not observe non-stationarity in graph 1 below). The increasing adoption of renewable energy sources is considered a positive development, but it has aggravated aforementioned characteristics and comes with higher volatility and sudden price jumps in the market (Lago et al, 2018). As we can see in our data, it even makes negative prices not uncommon in the case of excess generation caused by optimal weather conditions. Lago et al (2018) also claim that the increased volatility leads to more unpredictable behavior of market agents and an unstable electrical grid. Electricity price forecasting can mediate some of these problems, and has therefore become a center of attention in research on energy markets (Lago et al, 2018).

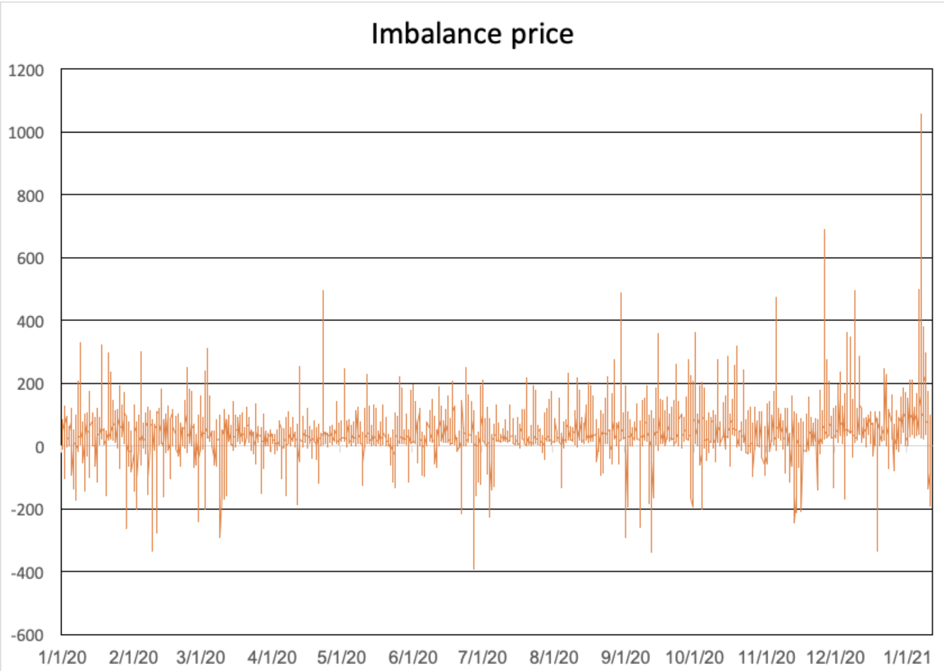


Figure 1. This graph plots the imbalance price of electricity in the Irish market over time from January 1st 2020 until January 1st 2021. Prices are expressed in euro. The interval is hourly and the data is obtained from Nabla Analytics.

In the Irish energy market (which includes both the Republic of Ireland and North Ireland), increased volatility is expected due to the recent Brexit and the neighbouring United Kingdom's decoupling from Euphemia (Nabla Analytics, 2020). Euphemia is a pan-European algorithm that calculates energy allocation and energy prices across Europe in order to maximize social welfare and increase price transparency (NEMO, 2019). This may present some issues for the external validity and extrapolating results of this research into the future. However, in electricity price forecasting it is common to continuously optimize models.

1.5 Electricity market trading

The far majority of trades in the electricity markets are made on the day ahead market, where electricity is sold at binding day-ahead prices for the following day (ISO New England, 2020). To clarify, the day ahead auction works as follows. Bids are allowed up until the moment the day ahead auction closes, which is at 12 PM on any given day. Market participants can place these price-quantity bids for all 24 delivery hours of the following calendar day. If a position is not closed it is settled for the imbalance price, which is what the trading customers of Nabla Analytics usually do. This means that the trading profit is effectively equal to the spread between the imbalance price and the day ahead price. The auction trading mechanism is illustrated in the graph below by Weron (2016):

1.6 Contribution

The focal points of the research are its usage of both modern techniques like neural networks and traditional models to forecast electricity prices. Moreover, besides traditional drivers of electricity prices the paper uses private data that tracks power plant outages in the Irish energy market. Since this data is not publicly available, this may be considered an under researched part of electricity price drivers.

The paper's empirical findings include the existence of a relation between power plant outages and electricity prices. It also finds that for two out of three forecast model types, the inclusion of power plant outage data significantly improves the forecasting accuracy. The paper cannot confirm, however, the forecasting outperformance of a neural network type model as compared to a random forest model or ARX type model. This contrasts with some findings in other literature. Since some of these findings are inconclusive, further research into the subject may be warranted.

2. Literature Review

2.1 Determinants of electricity prices

2.1.1 Feature selection

Through the selection of features we obtain a subset of the most informative features from all the features in the dataset that potentially have explanatory power (Zareipour, 2012). The selected features serve as input for the electricity pricing forecast model in this paper.

In existing literature many drivers of electricity prices have been discussed and used in forecasting of electricity prices. Many drivers will come from public databases. Below follows a list of several key drivers mentioned in literature: wind generation (Hong et al, 2020); demand (Zareipour, 2012); supply (Zareipour, 2012); weather conditions (Girish & Vijayalakshmi, 2013); time indices (Girish & Vijayalakshmi, 2013); and also the type of generation resources (Lora, Gonzales, Contreras et al.).

A more under discussed driver of electricity prices is power outages, although there is some literature in this area (see Rodriguez et al, 2004). While theoretically very sensible - it can be surmised that a (un)scheduled power outage leads to increased prices due to a lower supply (all else equal) - this data is not always public. For this research, we have access to a private database on (un)scheduled power outages in the Irish electricity market. This means we are able to look into power plant outages as a potential driver of electricity prices.

2.1.2 Selection method

This paper aims to predict one day-ahead electricity prices in the Irish electricity whole-sale market by using relevant features. For forecasting the electricity prices we first need to assess which futures are important for explaining the variation in electricity prices.

Traditionally the Ordinary Least Squares regression method is used for assessing the relationship between a dependent and an independent variable. The method computes the coefficients that best fit the data for the minimum residual sum of squares. Díaz(2019) defined an Ordinary Least Squares method for assessing the relationship between a dependent variable the 24 hour day ahead price and the number of relevant explanatory variables. However, they mentioned that publicly available data related to energy contains variables that are inherently highly correlated. This increases the risk of multicollinearity for the independent variables. Brooks(2019) states that one of the assumptions for the OLS to be unbiased is no perfect-multicollinearity in the independent variables. If there is no-perfect multicollinearity the expected estimated coefficient equals the true coefficient, but when this is not the case, the OLS estimator will be biased and we cannot trust the estimates anymore. Thus, to establish possible relationships through OLS we will have to keep the feature correlations in mind.

2.2 Forecasting models

In 2014, a comprehensive review of the, at that time, current methods for electricity price forecasting was published by Weron. The areas he identified for forecasting were (i) game theory models, (ii) fundamental methods, (iii) reduced-form models, (iv) statistical models, and (v) machine learning methods. It was concluded that the statistical models and machine learning models yielded the best results. Recently it was found that almost all machine learning methods perform better than statistical models. Especially deep neural networks (DNN) perform well in forecasting electricity prices (Lago et al., 2018). In a study where 15 statistical models and 11 machine learning methods were tested for their power to predict the day-ahead market in Belgium, the *European power exchange* (EPEX)-

Belgium, in a period ranging from 01/01/2010 to 31/11/2016. The top 8 performing methods of forecasting were machine learning methods. Among these machine learning methods, the multilayer perceptron (MLP), long-short term memory (LSTM), gated recurrent unit (GRU) and the DNN perform the best based on symmetric mean absolute percentage error (sMAPE). Note that three of these methods (LSTM, GRU and DNN) are part of the subclass deep learning.

The articles of Weron (2014) and Lago et al. (2018) give solid argumentation for using the four above mentioned forecasting methods for forecasting electricity prices. Since our time and our knowledge of deep learning is limited, and given the importance of explainability stressed in section 1, the MLP makes the best candidate to be applied in this article. Additionally, we will use a random forest type model since these models are often used in electricity price forecasting and less affected by the black-box challenge.

2.3 SHAP

However machine learning models tend to outperform traditional models in predictions, these models are more complex and harder to interpret. The problem with complex machine learning models is that they are treated as a black box (Ribiero 2016). For this research it is important that we understand what takes place under the hood of machine learning methods such as the Multi Layer Perceptron and the Random Forest that are used to predict the imbalance price. It is important that we can rely on the outcome of our model and that we better understand the reasons for the outcome and its predictions. Shapley Additive explanation (SHAP) and Lime(Local Interpretable Model Agnostic Explanations) explore the property of local explainability to build surrogate models to black-box machine learning models to provide them interpretability (Nayak, 2020). By deploying these models we gain greater transparency of what is going on inside the algorithm, why they make a prediction and which features are important (Gall,2019). In this research we perform the SHAP method, which is broadly more accepted than Lime because of its simplicity and theoretical guarantee (Nayak, 2020). SHAP is a game theoretic approach used for explaining the output of machine learning models.

2.4 Benchmark model

It is a common issue that energy forecasting that introduced models are not compared with classic models (Hong et al., 2020). A comparison with a benchmark model is necessary to be able to evaluate a forecasting model. A common approach to forecasting electricity prices is using auto regressive (AR) -type models (e.g. see Weron & Misiorek (2005)) and extensions of these models with regressors like load of the electricity grid, so called ARX models (Weron & Misiorek (2005). The common usage and the statistical inference framework of AR and ARX models make these types of models explainable. Since this paper aims to give traders and managers at trading firms and energy firms insights in the value of an improved forecasting model (either through a better method or through not before added data), an easy to explain benchmark model is key to assessing the model and results of this paper. This makes AR and ARX type models suited as benchmark models for this study.

Another model that is often explicitly used as a benchmark model for electricity price forecasting is the *similar day method* (Weron 2014). This model looks for days in the historical dataset with similar characteristics to the forecasted day and uses this to estimate the price of the forecasted day (Shahidehpour et al., 2002, Weron, 2006). A simple version of this which is frequently used is the *naive method* (Nogales et al., 2002, Weron, 2014). In this method only the most recent observation is taken into account and used as the forecast (Hyndman, 2005). This naive model is in a way incorporated into our forecast accuracy measure, which is explained in section 4.4.

2.5 Hypotheses

Given the information provided by the literature and the rationale provided in this section, a number of hypotheses are proposed.

H1: there is a relationship between power plant outages and electricity prices

H2: electricity prices can be forecasted more accurately by using power plant outage data.

H3: machine learning methods outperform statistical methods in forecasting electricity prices.

3. Data overview

3.1 General data overview

The availability of data differs for each variable collected. Since the power outage tracker is only fully backlogged past June 2020, however, we will not use data from before that period. Additionally, while historical data goes back further the day ahead forecasted values are only available for the past three months. However, since we will use forecasted data to test the model and not train it this should not prove a problem. To summarize, the data used in this research ranges from 01/06/2020 to 01/01/2021.

The data is restricted to the Irish electricity market, as this is the only market where the power outage tracker of Nabla Analytics is launched so far. In this paper we refer to the combined electricity market of the Republic of Ireland and North Ireland as the Irish market. Note that Brexit and the UK's uncoupling from Euphemia may affect applicability of the results outside of the academic framework if due to uncertainties around changes in the Irish electricity markets.

The frequency of available data ranges from 15 minute observations to hourly, but will be resampled to the lowest common denominator (60 minutes). The relevant timezone is UTC. Price variables are given in euro/MWh and the other variables in plain MW.

3.2 Sources

The data is retrieved from a mix of public and private sources. The power outage data is tracked by a product from Nabla Analytics, and is therefore private. Other data has been directly obtained from the publicly available database of market providers. The *historic wind*, *historic demand*, and *interconnector* variables are sourced from the EirGrid group. The *day ahead wind forecast* and *day ahead demand forecast* as well as the *imbalance price* are from SEMO. Lastly, the *day ahead prices* were gathered from SEMOpx.

3.3 Variables

Demand

The historical demand is the actual aggregate demand of Ireland and Northern Ireland resampled to the lowest common denominator (60 minutes) expressed in megawatt from the period June 1st 2020 until January 2021. The day ahead demand contains the aggregated demand of the Republic of Ireland and Northern Ireland for the period October 2020 until January 2021. The historical demand is used to fit the forecasting models and the day ahead demand forecast is used to test its forecasting performance.

Wind

The historical wind is the actual aggregated wind generation of Ireland and Northern Ireland resampled to the lowest common denominator (60 minutes) expressed in megawatt from the period June 1st 2020 until January 2021. The day ahead wind contains the aggregated wind generation of the Republic of Ireland and Northern Ireland for the period October 2020 until January 2021. This historical demand is used to fit the model and the day ahead demand is used to test its performance in the forecasting model in section 4.

Interconnector

The historical interconnector contains the actual net value of the interconnector flow of energy between electricity markets resampled to the lowest common denominator (60 minutes) from January 2020 until January 2021 expressed in megawatt. The relevant interconnectors for the Irish electricity market are the East-West interconnector, which connects the Irish and British electricity market, and the Moyle interconnector that connects the North Irish and Scottish electricity market.

Power outage forecast

We use the power outage of Nabla Analytics in megawatt from July 2020 until January first 2021. This variable is resampled to the lowest common denominator (60 minutes). A negative value of the *aggregated power outage* variable at a certain hour indicates an outage of a power plant with a corresponding decrease in power plant capacity in megawatt.

Prices

The day ahead price is a position for a quantity of energy that the bidder at the day ahead auction is willing to pay for a certain hour on the next calendar day. The day ahead auction takes place every day at 12:00 pm for the following calendar day. If the position a bidder places in the auction is not closed on the following calendar day, the bidder is settled for the imbalance price. The spread is computed by taking the difference between Day ahead price and the Imbalance price. The variables are denoted in euros and are taken over the period June 2020 until January 2021.

3.4 Descriptive Statistics

The summary statistics for the variables described above have been provided below in table 1. Note that these variables are accounted for hourly and daily seasonality, apart from the power outage and interconnector variable.

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
historic demand	5127	-3.004	1607.904	-12445.4	6710.667
historic wind	5127	1.159	3969.729	-6947.3	10265
interconnector	5127	-477.828	2044.97	-4535	3909
power outage	5127	-3531.387	926.693	-6110	-1840
spread	5127	84.282	49.736	-13.837	172.454
imbalance price	5127	0	46.209	-285.161	570.267
forecasted demand	1873	0	523.032	2733.909	1430.727
forecasted wind	1873	0	4945.846	-8858.637	10485.182

Table 1. This summary statistics table shows us the number of observations, mean, standard deviation, minimum and the maximum of the data used in this research. The variables described in the summary statistics table are the historical demand for electricity, the historical wind generation, the forecasted wind, the forecasted demand, the interconnector, the forecasted power plant outage expressed in megawatt. Note that this concerns the aggregated value for the Republic of Ireland and North Ireland. The imbalance price and the spread are expressed in euro/MWH corrected for hourly and weekly seasonality. Other variables are expressed in MW. All variables are hourly resampled to the lowest common denominator (60 minutes). The observed variables are taken from the period June 1st 2020 until January 1st 2021.

3.5 Seasonality

In electricity price forecasting, it is important to appropriately account for seasonality. If the seasonality is not taken into account, it is very likely that a model will merely forecast seasonal trends as opposed to the unknown variations we are interested in. There are multiple types of seasonality, and Weron (2014) states that for short-term forecasting the annual seasonality may be ignored, but the daily and weekly seasonalities should be accounted for. Since our models are built for short-term forecasting, we will remove only the daily and weekly seasonalities present in the forecast target: the imbalance price.

Following Weron (2006), the data can be rearranged in a matrix with a row length of T. T is to be all the hours in a week: 168. The average of this data should be taken, resulting in a vector of means. This vector is considered the seasonal component, and is then subtracted from the original observations of each respective hour of the week. Aside from the necessary deseasonalizing the forecasting target (electricity price), in this paper we have also deseasonalized the demand and wind features since this enhanced the performance of our models.

3.6 Standardization

While standardization of features is not necessary for the ARX type model or the random forest type model, it is useful for the Multi-Layer Perceptron type model. Thus, we used standardized input features only for the latter model type.

4. Methodology

4.1 Feature selection

First, we aim to obtain a selection of informative features that will be input of the electricity price forecasting model. The features are selected by performing an Ordinary Least Squares Regression and the necessary measures to complete the assumptions of the model.

The OLS regression offers an effective way of assessing the relationship between the potential informative features and the dependent variable (Diaz 2019). The OLS model that assess the relationship between the imbalance price and the regressors denoted in equation:

$$Y_t = \alpha + \beta_1 Demand_t + \beta_2 Wind_t + \ \beta_3\ Outage_t + \beta_4\ Interconnector_t + \ \ \varepsilon_t$$

,where the subscript t indicates that we use time series data, parameter alpha denotes the intercept, 1 denotes the estimated coefficient of the demand parameter, 2 is the estimated coefficient of the wind generation parameter, 3 is the estimated coefficient of the power outage parameter, 4 is the estimated coefficient of the interconnector parameter and 1 is the error term in the model.

4.2 Correlation

To account for the OLS assumption of unbiased betas of the regressors we assess for multicollinearity in the independent variables through testing the correlation between the regressors in the OLS model. Cohen(1988) set a framework for assessing the correlation of variables in which variables that have a correlation higher than 0.5 are highly correlated. To overcome the multicollinearity problem in the regressors that cause biased betas, regressors with high correlation are excluded from the OLS model.

Variable	historic demand	historic wind	aggregated power outage	interconnector	Intercept
historic demand	1				
historic wind	-0.4239	1			
aggregated power outage	-0.0312	-0.0332	1		
interconnector	-0.4063	0.7026	0.0788	1	
Intercept	-0.0633	0.0267	0.965	0.1588	1

Table 2. This table shows the correlation between the variables in regression model 1. These variables are historic demand, historic wind generation, interconnector and powerplant outage. All values are expressed in megawatt over the period June 2020 until January 2021.

From the table we observe that the independent variable interconnector has a correlation of 0.70 which is rather high. The high correlation indicates multicollinearity in the model and could lead to biased betas that cannot be interpreted. Therefore the OLS model for assessing the relationship between the imbalance price and the regressors should be respecified. This is done by excluding the interconnector regressors. Hence, we define the model as follows:

$$Y_t = \alpha + \beta_1 Demand_t + \beta_2 Wind_t + \ \beta_3\ Outage_t + \ \varepsilon$$

4.3 Forecasting methodology

In this section, three methods for forecasting the one day ahead electricity price will be employed. Two methods are machine learning methods, namely forecasting using a Multilayer Perceptron (MLP) and Random Forest. The third method is a statistical method, as Hong et al. (2020) stress the importance of a traditional benchmark model. This statistical method is forecasting using an ARX model. Since the interest of this research is the predictive power of power plant outages, two predictions are done for each forecasting method: one prediction with the power plant outages in the dataset, and one without.

For each method, the model is trained using historical data, whereas the models are tested using 1 day ahead forecasts. This way a situation is simulated where traders try to find the price of tomorrow in order to make trading decisions. For the tests, the forecasted data is presented as a theoretical time series with observations placed at the day and hour for which they are forecasted (and not at the time that they are forecasted). E.g. a forecast of t+1 is in the dataset at t+t.

4.3.1 K-fold cross validation

In estimating models, an important thing to be aware of is overfitting. Training and testing a model on the same data for example leads to overfitting. The model will have a very low bias but if it is used to forecast out of sample it would likely not be very accurate. One method to prevent overfitting from interfering with accurate forecasts is K-fold cross validation (Scikit-Learn, 2020). K-fold cross validation strengthens the external validity of a fitted model.

The first step is to split the data in a training set and a test set. The test set data will be used only for the final forecast and evaluation, whereas the training set data is used to find the best hyperparameters and optimization of the model. The training data is split into a number of folds (*K*), and the model is estimated with *K-1* folds as training data and 1 fold as test set. This is done so that every fold is used as the test set once, and then the K-fold cross validation estimated parameter values are averaged and returned. As a last validity test, the model with these parameters is then used to forecast on the test set data that was originally left out. This ultimate forecasting estimation is used to determine the forecasting accuracy.

In this paper we will use K-fold cross validation for both the random forest and multi-layered perceptron type model. To maximize the data used, the training data set consists of the historical values for the features (demand, wind, power outage), whereas the test data consists of the day ahead forecasted features (demand, wind, power outage). Moreover, the in sample training set data time period ranges from 01/06/2020 to 01/12/2020 while the out of sample test set uses data ranging from 01/12/2020 to 01/01/2020. The amount of folds required can vary and is dependent on the dataset size. Since we have a relatively large amount of observations, 5 folds should be sufficient. The K-fold cross validation with 5 folds is illustrated for additional clarity in the figure below (Scikit-Learn, 2020).

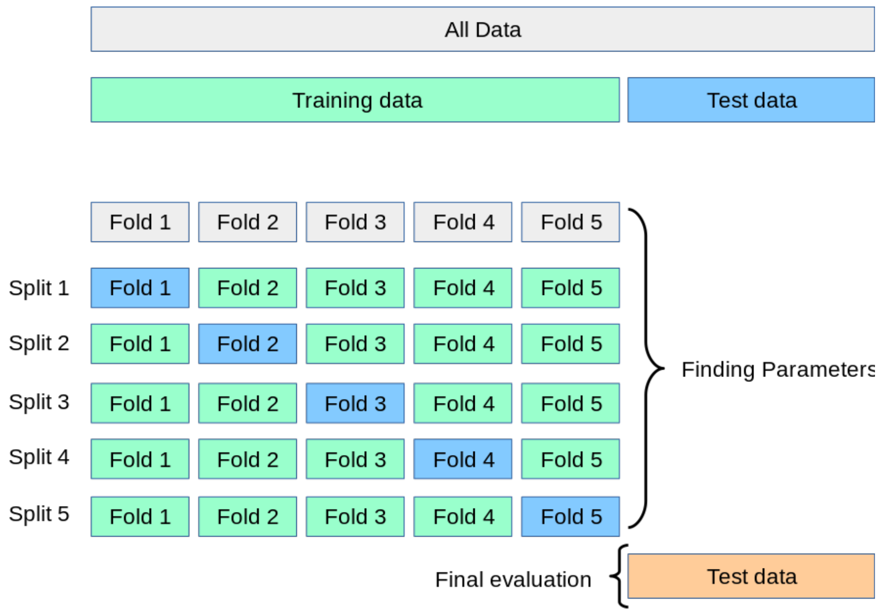


Figure 2. K-fold cross validation as illustrated in the scikit-learn.org manual.

4.3.2 Random Forest

Introduction

The random forest is a machine learning method that was introduced by Breiman in 2001 and is based on decision trees. A decision tree as a sole prediction model is known to be not optimally accurate. As quoted by Hastie, Tibishirani and Friedman in chapter 10 of The Elements of Statistical Learning (2001): “Trees have one aspect that prevents them from being the ideal tool for predictive learning, namely inaccuracy”. The random forest offers a lower variance than individual decision trees by combining low bias, high variance decision trees in an ensemble learning method (Hastie et al. 2001, chapter 10). The random forest employed in this research is based on a certain type of decision trees: regression trees. Regression trees are elaborated below, given their importance to understanding the random forest used in this research.

Regression tree

The regression tree is an algorithm that predicts the values of observations of a series based on the values of features that correspond to the observations of the values in that series. A decision tree is created where the sample is split into two smaller samples as a result from a decision. In turn, each sample is then again split into smaller samples. Each point where the sample is split is called a “node”. The node at the first decision is called the “root node”. If a node can no longer be split into two samples, this node is called a “leaf node”.

The algorithm can be summarized in a number of steps:

1. Cut-off values are calculated between the values of the features of all observations. E.g. if a feature has a vector of values [1, 2, 3, 5], then the cut-off values would be [1.5, 2.5, 4].
2. All combinations of features and cut-off values are considered a base to split the sample on. From each hypothetical split, the MSE of all observations in the two resulting nodes is calculated as $MSE = \frac{1}{n} \Sigma \left(y - \widehat{y} \right)^2$,
   where y-hat is the predicted value (from OLS) of y for the observations in the sample in the sample before the split and y is a vector of the observed value of y for each observation in the sample at each two resulting nodes. The MSE is calculated for all combinations of features and cut-off values. The combination of feature and cut of values that leads to the lowest MSE is chosen as the root node and the sample is split based on this cut-off and sent to two new nodes.
3. This process is repeated at each node, so that the sample at each node is again split based on the combination of a feature and cut-off value that yields the lowest MSE.

This process is repeats itself until one of two situations is present:

a. The nodes cannot be split further since the observations in each last node have the same y as the other observations in that node

b. The nodes can no longer be split since a minimum number of observations is set in each split, and this number is reached in each split.

In both cases, the last nodes cannot be split further and the process is terminated. Since no new nodes will be created, the last nodes will become leaf nodes. In case b), the output of a leaf node is the average y of the observations in that node. For this research, the y_t is the electricity price at t. An example of a hypothetical regression tree in the random forest employed in this research can be found in figure 3.

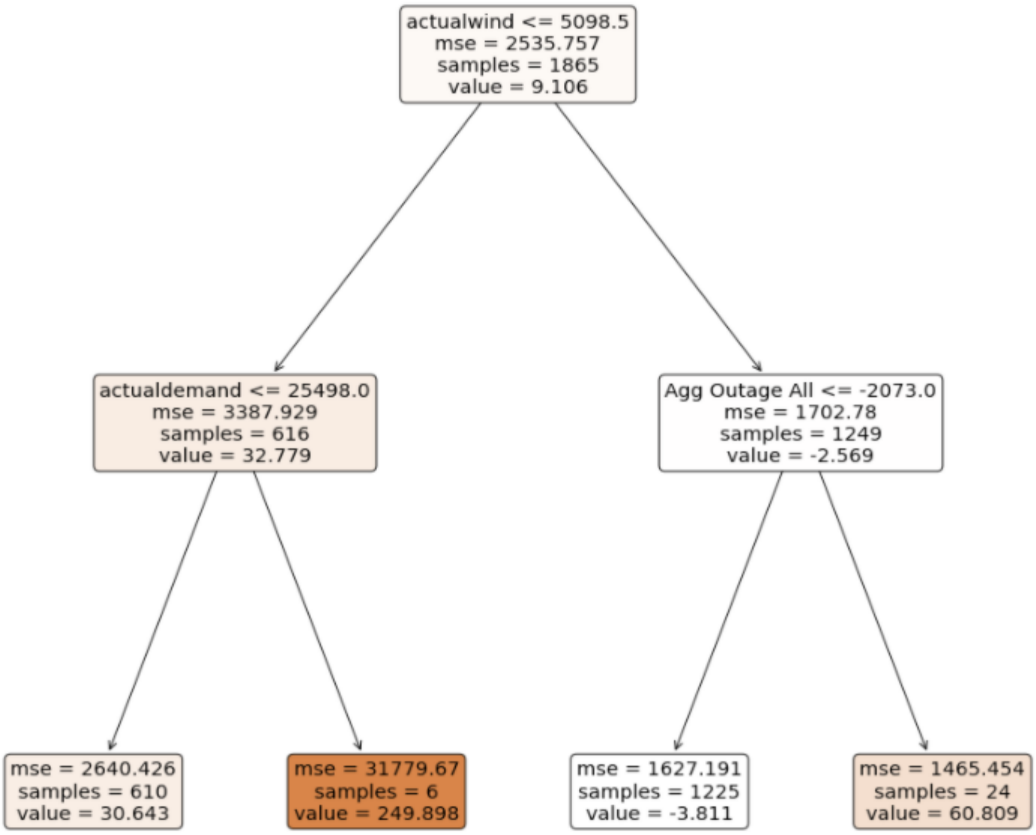


Figure 3. This figure shows a regression tree based on the data used in this research. The features “actual demand”, “actual wind” and “aggregated power outage” are used to predict the electricity price. In this figure “samples” is the number of observations left in that node. “value” is the value of the predicted value of the electricity price in that node. The maximum depth of the tree is set at 2 and the minimum number in each node resulting from a split is set to the default of 1.

Random Forest

A random forest regression is an algorithm that computes many regression trees according to the following equation:

$$\hat{f}_{\mathrm{rf}}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

It has the possibility to do so because it creates a “bootstrapped dataset” out of the original dataset. The bootstrapped dataset is created by randomly selecting rows from the original dataset and adding them to the bootstrapped dataset. The random forest then uses a random subset of features and uses this to create the root node. The subsequent nodes are made using the features that are not yet used to make a decision, up until no features are longer available. This process is repeated many times with many different bootstrapped samples so that many different trees are created.

Then the average of the prediction for each data point is taken from all trees. This will be the predicted y for each observation, in this research the electricity price. The power of estimating many trees and taking the average prediction of these trees lies in the fact that the over estimations cancel out the underestimations, leading to a more accurate prediction (Hastie et al 2001).

Hyperparameters

The random forest regression offers a number of hyperparameters that affect the quality of its forecasts. First of all, a maximum depth of the trees inside the random forest can be set. A higher depth of trees results in less bias, however it can lead to larger variance and overfitting (Hastie et al. 2001). The model is trained and tested for different tree depths through K-fold cross validation. K-fold cross validation is further discussed in section 4.3.1. In graph **4.** Below we can see the overfitting in the training data (orange lines) but for the test data (blue line) the optimal maximum tree depth is only 3. This is according to our forecast accuracy measure MASE, which is further explained in section 4.4.

Another hyperparameter that can be set to prevent overfitting is the minimum number of observations at a node required to split. It is set to the default number of observations of 2. Testing for different values no improved MASE is found when setting a higher number of minimum observations. Lastly, the max number of features is that is considered to create the optimal split at the trees inside the random forest. We have only three features available. The best in sample MASE is found at 2 for the model with power plant outages and at 1 for the model without power plant outages.

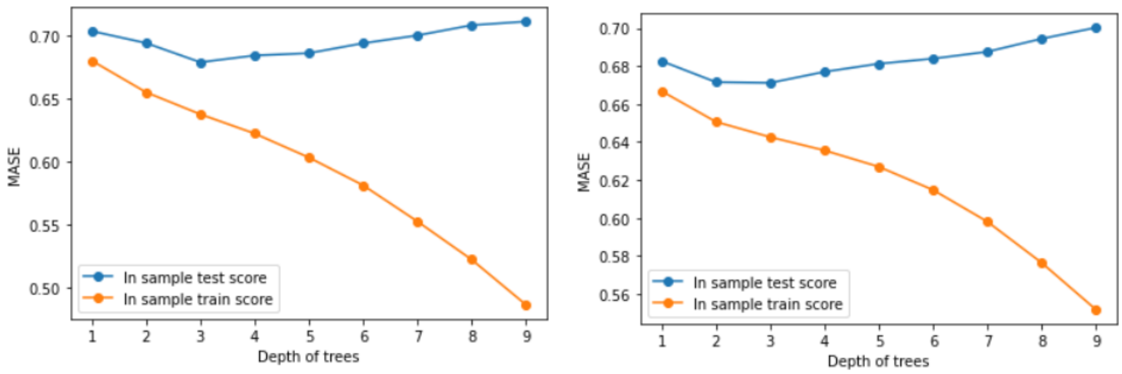


Figure 4. These graph plots the K-fold cross validated in-sample mean test and mean train score for different maximum tree depths, for the random forest type model with power outages (left panel) and the random forest type model without power outages (right panel). The score is the MASE, for which a lower value is considered better as it reflects the MAE of the model against the MAE of a naive model. For both models the optimal tree depth is considered 3. This accuracy measure is further explained in section 4.4.

4.3.3 Multi-Layer Perceptron

The multi-layer perceptron is a neural network type machine learning model. Neural networks are often considered to be somewhat of a black box, meaning people find it hard to understand the inner workings of a neural network based model. This section is aimed to relieve some of that black box feeling.

In the figure below a visualization of a neural network is illustrated. In this case the neural network consists of an *X*□ amount of input nodes (also features), *a*□ number of nodes in the hidden layer, and the output node *f(X)* (also the target/forecast). Information passes from the input nodes to the hidden layer to form the output. The neural network illustrated has only a single hidden layer but some neural networks will use more hidden layers. This is called deep learning. While there are some rules of thumb in regards to the number of nodes and hidden layers, this is generally based on trial-and-error (StatQuest, 2020).

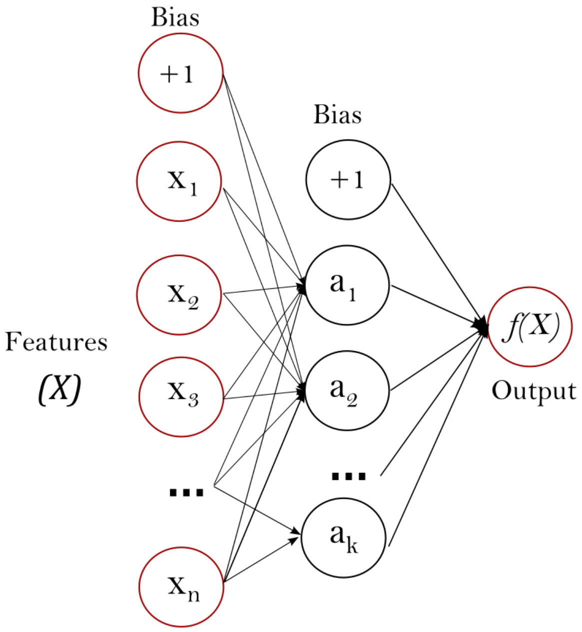


Figure 5. A neural network as illustrated in the scikit-learn.org manual.

A neural network generally works as follows (StatQuest, 2020). The value of the input node is taken and transformed by multiplying it with the weight parameter and also adding a bias value. Both these parameters are estimated through backpropagation. Backpropagation conceptually means it estimates the last parameters first (the weights and biases closer to the output). Each input-node relation has its own estimated parameter values. This transformation gives a value that is considered the x-coordinate and it is plugged into the node of the hidden layer. The x-coordinate is used to look up the y-coordinate of the activation function in the hidden layer. There are several types of activation functions and the corresponding y-value is of course dependent on this choice. An example of an activation function is the softplus function: $f(x) = ln(1+exp\ x)$. Ultimately each node recreates a portion of its activation function with the y-values it receives from the input features. Moving from the node to the output, this line is then rescaled and transformed using (again) estimated weight and bias parameters. The resulting functions of all the nodes are combined to form the fitted line. The fitted line is the output *f(X)* in the figure above, and in terms of this paper would be the electricity price forecast.

The primary upside of neural networks is the ability to fit non-linear models (Du & Swamy, 2014). However, there are also some downsides to be considered. Neural networks can become very computationally expensive and requires tuning of many types of hyperparameters, which we will discuss further below.

Hyperparameters that we tuned/optimized

The activation function that yields the optimal K-fold cross-validated in sample MASE value in this research is the logistic sigmoid function defined as $f(x) = 1 / (1 + exp(-x))$. The best solver for the weight optimization is the "adam" solver which is based on a stochastic gradient-based optimizer proposed by Kingma et al. (2014). The optimal L2 regularization term is the default for both the models with and without power plant outage data. For the size of the hidden layers, Jeff Heaton (2008) suggests several rules of thumb. Heaton (2005) suggests a maximum of two hidden layers, and to use ⅔ the size of the input layer plus the size of the output layer. There are 3 and 2 features in the input layer of our models, respectively. Thus, two hidden layers with 3 nodes each are chosen. More and larger layers can lead to more precision, however they are also more computationally expensive or overfitting. The accuracy gain from more and larger layers (e.g. a layer structure of (10,10,10,10) is also very marginal.

4.3.4 ARX Model

An ARX model is fitted following the methodology of Weron and Misiorek (2005). An ARX model is an extended AR type model that includes exogenous input variables X (Weron, 2014). Time series models have been extensively applied to short term electricity price forecasting according to Weron (2014). The AutoRegressive part of the specification basically means that the current value of price reflects the past one or more price values. Usually this is indicated as AR(p), with p the number of lags that is considered. For the purpose of this paper, we will follow Weron and Misiorek (2005) who identify that the best forecasting results are found when lag 24, lag 48, and lag 168 are in the model. However, since lag 24 is not always known to traders at the time of forecasting, it is not incorporated into our model. Together with forecasts of wind and demand and the power plant outage data this concludes to the following model:

$$p\_t = c + p\_t\text{-}48 + p\_t\text{-}48 + WINDFC\_t + DEMANDFC\_t + (POWEROUTAGE\_t) + u\_t$$

Where the brackets next to the power outage variable signal its inclusion in only the model that includes power plant outage data and not the other.

4.4 Measure to evaluate forecasts: MASE

A measure to evaluate the forecast has to be carefully selected as selecting the wrong measure may invalidate the results and conclusion. An often advocated measure is the MAPE (Mean Absolute

Percentage Error). However, the MAPE has a few disadvantages: it is infinite if the target value is 0; has a very skewed distribution for target values close to 0; it assumes the 0 value is meaningful; and lastly penalizes negative errors more than positive errors (Hyndman, 2005). Our electricity price data does indeed have negative values and values that are close to 0, so this measure would pose a problem. To address some of the issues with the MAPE, Makridakis (1993) introduced the symmetric MAPE. Hyndman (2005), however, notes that the sMAPE mostly lessens these issues but does not remove them completely.

Ultimately, in his paper Hyndman (2005) proposes a new forecasting accuracy measure that does not suffer these biases: the MASE (Mean Absolute Scaled Error). Effectively, the MASE compares the MAE of the estimated forecasting model to the MAE of the in-sample naive model. The naive model is a model that takes the most recent observation and uses that as a forecast. In our paper, the forecasting models are estimated using data available from before the closing of the day ahead auction at 12PM the calendar day before. To give our forecasting models a fair chance, for the naive model we will use the observation of 24 hours before. This is imperfect, but the hourly forecasts for each calendar day range from 12 hours to 36 hours away from closing of the auction - so on average it works out. Note that often the forecast input data is from earlier in the morning before the auction closes, giving the naive model a slight advantage.

Hyndman (2005) defines the MASE as follows below:

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|} \qquad \text{MASE} = \text{mean}(|q_t|). \tag{x}$$

Fortunately, the MASE is also very easy to interpret. A MASE of 1 means the estimated model is no better or worse than a naive model. If the MASE > 1, this means that the estimated model has worse forecasting accuracy than the naive model. If the MASE < 1, this means the estimated model is better than the naive model. This would be the optimal outcome in terms of forecasting accuracy.

Lastly, use of the MASE should also make comparison of the forecasting method to other papers using the MASE (even with other datasets), possible (Hyndman, 2005). This strengthens the academic contribution of this paper, even if some of the data used itself is private.

4.5 Establishing possible statistically significant superiority between models

After assessment of the individual accuracy of all three model types, in the final step of this paper the predictions of the forecasting methods are compared in order to find whether the addition of power plant outage data can lead to a better prediction of electricity prices. The Diebold-Mariano is used to compare the accuracy scores of the predictions (Diebold & Mariano, 1995).

The test statistic DM is calculated as DM = $\frac{\bar{d}_{12}}{\hat{\sigma} i_{12}}$ where $\bar{d}_{12} = \frac{1}{T} \sum_1^T (l_{1,t} - l_{2,t})$ with l being the loss function $\mathrm{MASE} = \mathrm{mean}(|q_t|)$. with $q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^{n} |Y_i - Y_{i-1}|}$ and $\hat{\sigma}_{\bar{d}_{12}}$ as a consistent estimate of the standard deviation of $\bar{d}_{12}$. For the Diebold-Mariano test statistic to be valid, the value of the test statistic DM should hold *DM ~ N(0, 1)*. Franses (2015) suggests that this does not necessarily hold for measures such as the MAPE or the sMAPE, but does hold for the MASE which is used in this research.

We test the difference in performance for model 1 (with power plant outage data) and model 2 (without power plant outage data) for the random forest and the MLP with

H0: $\bar{d}_{12}$ *= 0 and model 1 and 2 have equal forecast accuracy.*

and with

H1: $\bar{d}_{12}$ *≠ 0 and model 1 has better forecast accuracy than model 2.*

For this test, the models with power plant outage data are labeled as "model 1", the models without power plant outage data are labeled "model 2". The results of this test can be found in section x.

Furthermore, the methods are compared by performing a Diebold-Mariano test on the models of different methods with the same variables, e.g. the MASE of model 1 predicted with a random forest is compared with the MASE of model 1 predicted with a MLP. The corresponding hypotheses are:

H0: the random forest and the MLP have equal forecasting accuracy in predicting one day ahead electricity prices.

with *H1: either the random forest or the MLP has better forecasting accuracy than the other in predicting one day ahead electricity prices.*

The results of this can also be found in section x.

4.6 Interpretability of machine learning models

Lundberg and Lee (2016) proposed the SHAP value as a unified framework to explain the output of any machine learning model, interpreting predictions. The SHAP provides global interpretability. It shows us how much each feature contributes positively or negatively to the target variable. Where traditional variable importance algorithms such as the OLS regression show a global result, the SHAP method has local interpretability, each observation has its own set of shap values. This increases the transparency of the method and reduces the black box problem of complex machine learning methods.

Shapley(1953) designed a model in which that computes the contribution of each player in a coalition game. This model can be applied to machine learning models for creating an unified framework that calculates the marginal contribution of each feature.

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(N - |S| - 1)!}{N!}(v(S \cup \{i\}) - v(S))$$

Assume there are N features and S is a subset of the N features. Let $v(S)$ be the total value of the S features. When feature i joins the S features, feature i's marginal contribution is $v(S \cup \{i\}) - v(S)$. The contribution of each feature i is obtained by taking the average of the contribution over the possible different permutations in which the coalition can be formed. The model works for tree models because the variables enter the machine learning model sequentially or repeatedly in the trees of the model. In every step of tree growing, the algorithms will evaluate each of all the variables equally and will settle with the variable that contributes the most to the target. In the random forest multiple trees are constructed and because of this there will be various permutations of the variables available for the features. In this way we obtain the marginal contribution of each variable that can be calculated for machine learning models such as the Random Forest Regression.

5. Results

5.1 OLS regression

For the selection of the relevant features in the forecasting model this paper assesses the relationship between the imbalance price and the variables historic demand, historic wind, interconnector and powerplant outage. Two regressions models are estimated in table 3. For feature selection we focus on model 2 that is corrected for multicollinearity by excluding the feature interconnector.

	(1)	(2)
VARIABLES	imbalance price	imbalance price
historic demand	0.012***	0.009***
	(0.000)	(0.000)
historic wind	-0.006***	-0.004***

	(0.000)	(0.000)
power outage	-0.003***	-0.002***
	(0.001)	(0.001)
Interconnector	-0.005***	
	(0.000)	
Constant	-13.586***	-8.710***
	(2.295)	(2.305)
Observations	5,127	5,127
R-squared	0.218	0.191

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3. This table contains the results of the estimation of OLS model 1 an OLS model 2: Both models are estimated using OLS. Where demand is the historic demand for electricity, historic wind is the wind, historic power plant outage is the power plant outage, and interconnector is the aggregated interconnector. The observed variables concern Ireland. The variables are expressed in megawatt and Euro per megawatt taken from the period June 2020 until January 2021.

Historic demand

From table 3 we find that the coefficient of the feature *historic demand* in model 2 has a positive significant value of 0.009 for a level of significance of $p < 0.01$. This indicates that there is a positive relationship between the independent variable *historic* demand and the dependent variable imbalance price. When the historic demand for electricity increases with 1 megawatt the imbalance price will increase with EUR per megawatt, ceteris paribus. Note that the average imbalance price is equal to 0 EUR per megawatt and that the

Historic wind

From table 3 we find that the coefficient of the feature *historic wind generation* in model 2 has a negative significant value of -0.004 for a level of significance of $p < 0.01$. This indicates that there is a negative relationship between the independent variable *historic* wind generation and the dependent variable imbalance price. When there is an increase in wind generation of 1 megawatt the imbalance price will decrease with EUR 0.004 per megawatt, ceteris paribus.

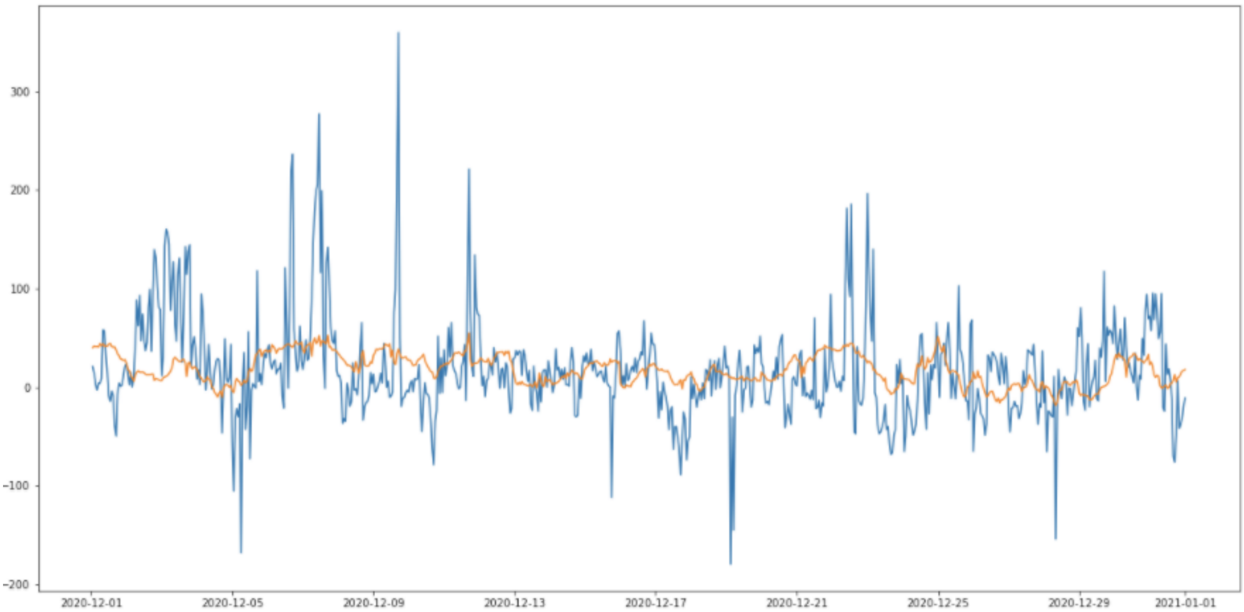
Powerplant outage

From table 3 we find that the coefficient of the feature *powerplant outage* in model 2 has a neg significant value of -0.002 for a level of significance of $p < 0.01$. This indicates that there is a negative relationship between the independent variable *power plant outage* and the dependent variable imbalance price. An increase of the value of the feature *power plant outage* is a decrease in *power plant outage*. This indicates that when the independent variable power plant outage increases with 1 megawatt the imbalance price decreases with EUR 0.002 per megawatt, ceteris paribus.

From the output of the OLS regression model we find that the features *historic demand*, *historic wind* and *power plant outage* have significant explanatory power in explaining the dependent variable *imbalance price.* Therefore we chose these features in the forecasting methods ARX, the Multi-Layer Perceptron, and the Random Forest Regression.

5.2 ARX forecasting results

The ARX models deliver a more accurate forecast than the MLP and the random forest, with and without power plant outage data. Furthermore, adding the power plant outage data does not improve the forecast accuracy of the ARX models.



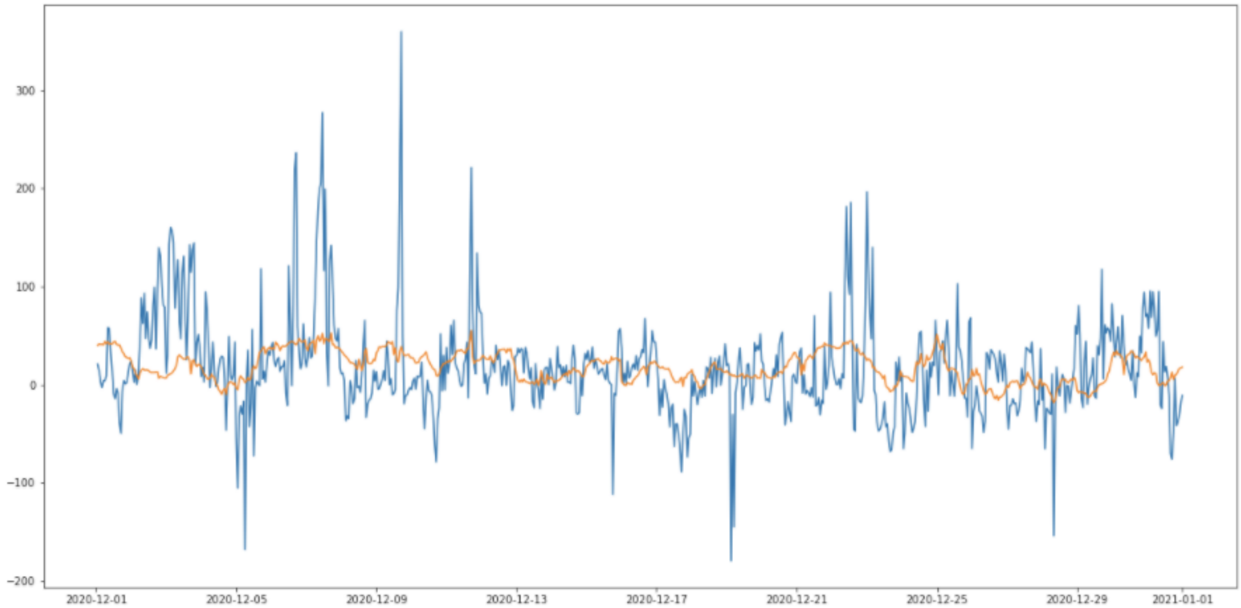


Figure 7. Upper panel presents the out-of-sample forecast with power outage data, the lower panel presents the out-of-sample forecast without power outage data. Both models are the ARX type model. Time period ranges from 01/12/2020 to 01/01/2021, and observations are hourly.

5.3 Random forest forecasting results

Following the methodology as laid out in section 4, the electricity price forecasts of the random forest type model are as presented in figure 7 below. The MASE compares the forecasting error of these models to that of the naive model. For the out of sample random forest type forecast with power outage data, the MASE is 0.885. For the random forest type forecast without power outage input, the MASE comes to only 0.984. This indicates that the random forest type forecast without power outage is essentially equal to the naive model. Furthermore, the MASE of the model with power outage data included indicates better forecast accuracy than both the naive model and the model without power outage data. This will be formally tested in section 5.6 with the Diebold-Mariano test.

Figure 7. Upper panel presents the out-of-sample forecast with power outage data, the lower panel presents the out-of-sample forecast without power outage data. Both models are the random forest type model. Time period ranges from 01/12/2020 to 01/01/2021, and observations are hourly.

5.4 SHAP

From the graph of the unified framework (SHAP) we observe that in the Random Forest Regression the feature demand contributes the most to the target value followed by wind and powerplant outage. We observe that a high level of wind (red) has a high and negative contribution to the imbalance price. Also we observe that a high level of the feature demand has a high positive impact on the imbalance price. Furthermore we observe that a high level of the feature power outages(negative value) has a positive contribution to the imbalance price

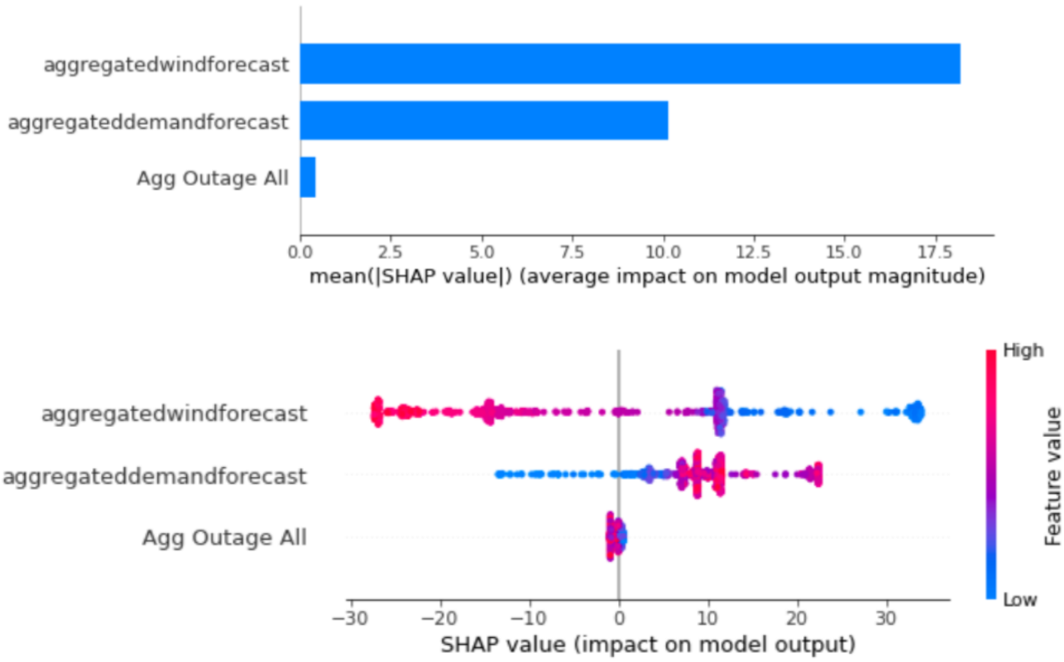


Figure 8: The upper graph shows us the contribution of the features to the target value(descending) by taking the mean absolute SHAP value per feature. The lower graph sorts the features by the sum of SHAP value magnitudes over all samples, and uses SHAP values to show the distribution of the impacts of each future on the model output. The location on the x-axis shows the effect of the value on the prediction, the color red and blue tell us whether the feature is high or low for that observation. Both graphs concern the random forest model including the power plant outage feature. Time period ranges from 01/12/2020 to 01/01/2021, and observations are hourly.

5.5 Multi-Layer Perceptron forecasting results

The electricity price forecasts of the Multi-Layer Perceptron type model are as presented in figure 9 below. The MASE compares the forecasting error of these models to that of the naive model. For the out of sample random forest type forecast with power outage data, the MASE is 0.903. For the Multi-Layer Perceptron forecast without power outage input, the MASE comes to only 1.022. This indicates that the random forest type forecast without power outage is essentially equal to the naive model (~1). Furthermore, the MASE of the model with power outage data included indicates better forecast accuracy than both the naive model and the model without power outage data. This will be formally tested with Diebold-Mariano in the next section.

Figure 9. Upper panel presents the out-of-sample forecast with power outage data, the lower panel presents the out-of-sample forecast without power outage data. Both models are the Multi-Layer Perceptron type model. Time period ranges from 01/12/2020 to 01/01/2021, and observations are hourly.

5.5.1 SHAP

The SHAP unified framework in appendix 1 shows us how the features contribute to the level of the target value. We find that the importance of the features are in line with the outcome of the SHAP summary of the Random forest: the demand contributes the most followed by the wind generation and the level of power outage.

5.6 Diebold-Mariano testing

5.6.1 Evaluation power outage inclusion

The Diebold-Mariano tests for the random forest forecasts and the MLP forecasts with

H0: "model 1 and 2 have equal forecast accuracy"

and

H1: "model 1 has better forecast accuracy than model 2".

Where the models with power plant outage data are considered "model 1" and models without power plant outage data are considered "model 2", the following results are found:

a. The test on the random forest models yields a test statistic DM = |-4.07| > DM_crit, which means that the *H0* is rejected.

b. The test on the MLP models yields a test statistic DM = |-8.87| > DM_crit, thus is this case *H0* is also rejected.

For both the random forest and the MLP the H1 is adopted and therefore the forecast accuracy increases when power plant outage data is added. For the exact MASE values please consult table 4. However, this does not seem to hold for the ARX type model, which has similar MASE values for both the model with power plant outage and without power plant outage.

Target	Power outage feature	Naive model	ARX model	Random Forest model	Multi-Layer Perceptron model
price	no	1	0.819	0.984	1.026
	yes	1	0.819	0.885	0.912

Table 4. This table presents the MASE forecast accuracy measure for the out-of-sample forecasted electricity prices for various forecasting models.

5.6.2 Evaluation and comparison of forecasting methods

For the Diebold-Mariano test with

H0: the random forest and the MLP have equal forecasting accuracy in predicting one day ahead electricity prices

and

H1: either the random forest or the MLP has better forecasting accuracy than the other in predicting one day ahead electricity prices.

It is found that for type model 1 *H0* is rejected, but there is no statistical difference in forecast accuracy when testing this for type model 2. Associated test statisitc for type model 1 is -2.31 with p-value of 0.021. For type model 2 this is a test stat of 0.59 and p-value of 0.554.

This indicates that the random forest type model and the MLP type model do not perform significantly differently in their best performing model specification.

A Diebold-Mariano test is done for similar hypotheses:

H0: the random forest and the ARX have equal forecasting accuracy in predicting one day ahead electricity prices

and

H1: either the random forest or the ARX has better forecasting accuracy than the other in predicting one day ahead electricity prices.

For this test it is found that for model 1, the ARX is statistically more accurate than the random forest with a test statistic of -4.05 with p-value 0.000. However, for model 2 the difference in accuracy is not statistically significant. The corresponding test statistic is -1.12 with a p-value of 0.262

6. Conclusions

Hypothesis 1: there is a relationship between power plant outages and electricity prices in the Irish electricity market

From the first part of this research we can conclude that demand for electricity, wind generation and power plant outage have significant power in explaining the electricity prices in the Irish electricity market. The results from the traditional OLS regression tells us that demand for electricity has a significant positive effect on the imbalance price per megawatt, where wind generation has a significant negative effect on the imbalance price per megawatt. These findings are in line with (Zareipour, 2012). We find that power outage has a negative significant effect on the imbalance price per megawatt. The private data on power plant outage tells us that an unscheduled power plant outage that leads to lower electricity supply, results in a rise of electricity prices. Because of this we can state that it is likely that power plant outages are a drive for electricity prices in the Irish electricity market. This finding confirms hypothesis 1 of this research.

Hypothesis 2: electricity prices can be forecasted more accurately by using power plant outage data.

The Diebold-Mariano test formally confirmed that inclusion of power plant outage data improves the forecasting accuracy of both the random forest type model and the multi-layered perceptron type model, but we do not see the same for the ARX type model. Thus, it is tough to either confirm or deny our second hypothesis outright. If we go by majority rule, it is in line with our expectations and the theoretical relation between supply constraints and price. It can be taken from this that the power outage tracker of Nabla Analytics should give its customers some advantageous knowledge in predicting electricity prices, although using it to forecast accurately remains imprecise.

Hypothesis 3: machine learning methods outperform statistical methods in forecasting electricity prices.

Since the hypothesis of equal forecasting accuracy for the random forest and multi-layer perceptron type model could not be rejected with the Diebold-Mariano test, it can not convincingly be said that the MLP performs better in forecasting electricity prices than the random forest type model. Thus, we reject our third hypothesis that machine learning methods outperform statistical methods in forecasting electricity prices. This means our findings are not necessarily in line with the findings of Lago et al (2018) and perhaps further research and finetuning is warranted.

To summarize, we conclude in answer to our research question that there is a relation between power plant outages and electricity prices and that inclusion of this information improves the forecast accuracy of our estimated forecasts depending on model type. This divided outcome perhaps indicates further research is warranted, which will be discussed in the following section.

7. Discussion

No research is perfect, and it is important to discuss some of the limitations and angles for future research that flow from this paper. There were two main limitations with regards to the research. First is the optimization of hyperparameters which is expensive both computationally and in time. There is no guarantee that the parameters are fully optimized, although they are to the best under the circumstances. The research was limited in the data provided, but could possibly benefit from the availability of data, both in terms of an increased number of features and in terms of a longer time period.

Following from these limitations is also future research that could follow from this paper. As mentioned, more control variables including perhaps a revolving price lags, and a longer dated timeset could further the forecasting accuracy and external validity of the model. Moreover, future research could also look into multi-target classifier predicting. This could still be useful for traders, and this method may alleviate some of the problems with forecasting the extreme peak electricity price values present in the dataset.

8. References

Bhattacharyya, S. (2018). Ridge and Lasso Regression: L1 and L2 Regularization. Retrieved from: *https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b*

Blake S., (2020) Regularization with Ridge, Lasso, and Elastic Net Regression. Retrieved from: https://towardsdatascience.com/what-is-regularization-and-how-do-i-use-it-f7008b5a68c6

Bhattacharya K, Bollen MH, Dallder JE (2001) Operation of restructured power systems. Kluwer, Boston *https://books.google.nl/books?hl=en&lr=&id=PtLgBwAAQBAJ&oi=fnd&pg=PP11&ots=PiWGg2jvr1&sig=CSRuCL72aK4h0DAA82iQAfiAZW4&redir_esc=y#v=onepage&q&f=false*

Brooks, C(2019): Introductory Econometric for finance. Cambridge University Press, 4rd ed.

Cohen, J. (1988). Set correlation and contingency tables. Applied psychological measurement, 12(4), 425-434.

Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. IEEE transactions on power systems, 18(3), 1014-1020.

Díaz, G., Coto, J., & Gómez-Aleixandre, J. (2019). Prediction and explanation of the formation of the Spanish day-ahead electricity price through machine learning regression. Applied Energy, 239, 610-625. Prediction and explanation of the formation of the Spanish day-ahead electricity price through machine learning regression - ScienceDirect (oclc.org)

Diebold F.X. (2012). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. NBER Working Paper No. 18391 Retrieved from http://www.nber.org/papers/w18391.

Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. J Bus Econ Stat 1995;13(3):253–63. http://dx.doi.org/10.1080/07350015.1995.10524599.

Du, K-L., & Swamy, M.N.S. (2014). Neural Networks and Statistical Learning. London: Springer.

Franses, P. (2015). A note on the Mean Absolute Scaled Error. International Journal of Forecasting, 32, p. 20- 22. doi:10.1016/j.ijforecast.2015.03.008

Gall, R. (2019). Machine learning explainability vs interpretability: Two concepts that could help restore trust in AI. KDnuggets News, 19(1).

González, A. M., Roque, A. S., & García-González, J. (2005). Modeling and forecasting electricity prices with input/output hidden Markov models. IEEE Transactions on power systems, 20(1), 13-24.

Heaton, J. (2008). Introduction to Neural Networks with Java. St. Louis: Heaton Research, 2nd ed.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy Forecasting: A Review and Outlook. IEEE Open Access Journal of Power and Energy: 7, pp. 376-388.

Hyndman, R.J. (2005). Another look at measures of forecast accuracy. Retrieved from: https://robjhyndman.com/papers/mase.pdf.

ISO New England (2020). FAQ: Day-Ahead Energy Market—Commitment, Scheduling, and Dispatch. Retrieved from: https://www.iso-ne.com/participate/support/faq/da-market-commitment#:~:text=The%20Day%2DAhead%20Energy%20Market%20(day%2Dahead%20market),prices%20for%20the%20following%20day

Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied Energy: 221, pp. 386-405.

Lora, A. T., Santos, J. M. R., Expósito, A. G., Ramos, J. L. M., & Santos, J. C. R. (2007). Electricity market price forecasting based on weighted nearest neighbors techniques. IEEE Transactions on Power Systems, 22(3), 1294-1301.

Lundberg and Lee: Lundberg, S., & Lee, S. I. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. Int J Forecast 1993;9(4):527–9. http://dx.doi.org/10.1016/0169-2070(93)90079-3.

Nayak, A. (2019, 22 december). Idea Behind LIME and SHAP - Towards Data Science. Medium. https://towardsdatascience.com/idea-behind-lime-and-shap-b603d35d34eb

NEMO (2019). EUPHEMIA Public Description: Single Price Coupling Algorithm. Retrieved from: http://www.nemo-committee.eu/assets/files/190410_Euphemia%20Public%20Description%20version%20NEMO%20Committee.pdf

Nogales FJ, Contreras J, Conejo AJ, Espínola R. Forecasting next-day electricity prices by time series models. IEEE Trans Power Syst 2002;17(2):342–8. [http://dx. doi.org/10.1109/MPER.2002.4312063.](http://dx.doi.org/10.1109/MPER.2002.4312063)

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

Rodriguez, C. P., & Anders, G. J. (2004). Energy price forecasting in the Ontario competitive power system market. IEEE transactions on power systems, 19(1), 366-374.

Shahidehpour, M., Yamin, H., & Li, Z. (2002). Market operations in electric power systems: forecasting, scheduling, and risk management. Wiley.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288. *Regression Shrinkage and Selection Via the Lasso - Tibshirani - 1996 - Journal of the Royal Statistical Society: Series B (Methodological) - Wiley Online Library*

Weron, R. (2014). "Electricity price forecasting: A review of the state-of-the-art with a look into the future." International journal of forecasting 30.4 (2014): 1030-1081.

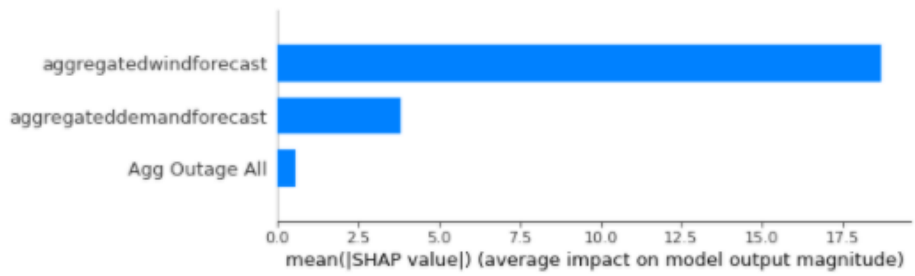
Weron, R, & Misiorek, A. (2005). Forecasting spot electricity prices with time series models. In: Proceedings of the European electricity market EEM-05 conference; 2005. p. 133–41.

Weron, R., & Misiorek, A. (2006). Short-term electricity price forecasting with time series models: A review and evaluation. In W. Mielczarski (Ed.), Complex electricity markets (pp. 231–254). Łódź: IEPŁ& SEP. *https://www-nber-org.vu-nl.idm.oclc.org/system/files/working_papers/w8442/w8442.pdf*

Zareipour, H. (2012). Short-term electricity market prices: A review of characteristics and forecasting methods. In Handbook of networks in power systems I (pp. 89-121). Springer, Berlin, Heidelberg.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), pp. 301-320.

APPENDIX

Appendix 1. SHAP Multilayer Perceptron



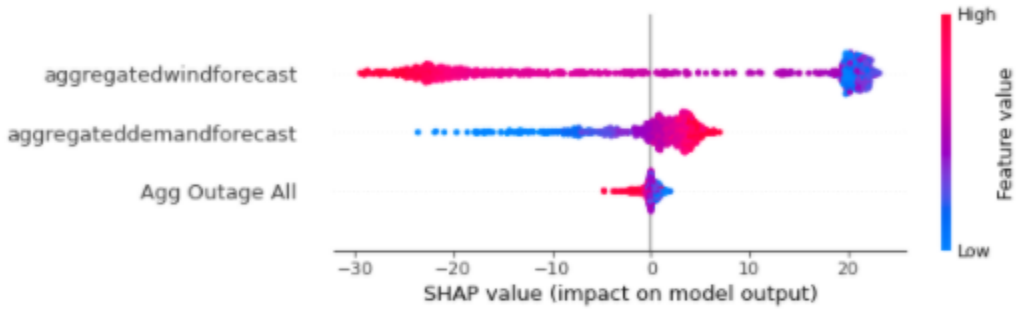


Figure 8: The upper graph shows us the contribution of the features to the target value(descending) by taking the mean absolute SHAP value per feature. The lower graph sorts the features by the sum of SHAP value magnitudes over all samples, and uses SHAP values to show the distribution of the impacts of each future on the model output. The location on the x-axis shows the effect of the value on the prediction, the color red and blue tell us whether the feature is high or low for that observation. Both graphs concern the multilayer perceptron model including the power plant outage feature. Time period ranges from 01/12/2020 to 01/01/2021, and observations are hourly.