Machine Learning with Applications 6 (2021) 100182

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Technical indicators for energy market trading☆

Catherine McHugh *, Sonya Coleman, Dermot Kerr

Intelligent Systems Research Centre (ISRC), Ulster University, Northern Ireland, UK

ARTICLE INFO

Keywords:
Energy market
Hourly price forecasting
Regression machine learning
Technical indicators

ABSTRACT

Technical indicators have been widely applied to the financial trading market, often combined with machine learning algorithms, to predict future stock market prices. The characteristics of energy market data are comparable to financial trading data; hence this research derives eight price prediction technical indicators for hourly electricity prices from the Irish Integrated Single Electricity Market. The proposed indicators consider the three key types of price indicators: trend, oscillator, and momentum. Building the technical indicators from raw electricity price data helps to capture market behaviours and find information to predict future profitable prices. The electricity price data for the proposed indicators were collected from February 2019 until March 2020. Three machine learning regression algorithms were trained with the technical indicators: Extreme Gradient Boosting, Gradient Boosting, and Random Forest. The results demonstrate that the price prediction models perform much better when trained using the proposed technical indicators when compared with baseline raw price data models.

1. Introduction

Time series prediction models are trained using historical data to analyse patterns and to help with forecasts. Electricity prices exhibit differing characteristics leading to price fluctuations in response to supply and demand, making energy prices quite challenging to predict (Mosbah & El-Hawary, 2016). Accurate prediction models would be beneficial for energy traders to observe electricity price trends and over time reduce trading costs. In particular, machine learning could be considered as these algorithms aim to create optimal models that try to reflect the market trend (Gao, Lo, & Fan, 2017). Short-term forecasting models are preferable in energy trading to manage the volatility which is present in the market (Amjady & Hemmati, 2006; Pandey & Upadhyay, 2016). Technical indicators have been widely used in the financial trading market to help investors decide whether to buy, sell or hold price units (Tanaka-Yamawaki & Tokuoka, 2007). Therefore, technical analysis would be an appropriate tool to consider for electricity price forecasting to aid energy traders in making a decision on when to buy or sell electricity price units within the energy market.

The Integrated Single Electricity Market (ISEM) is a recent development across Ireland which gives energy traders more control and greater flexibility. Given the popularity of technical indicators in the financial market, there is a need to design original technical indicators for the energy market which is the main contribution of this paper. These technical indicators observe market trends and support price forecasting decisions. Therefore, this paper examines day-ahead electricity data with the goal of building a novel system to reduce purchasing and trading costs within the ISEM. In Section 2 we present related work on technical indicators in both the financial and energy markets. The proposed eight electricity price technical indicators are defined in Section 3. Section 4 describes the three machine learning methodologies used with both the raw price data and the proposed technical indicators. Section 5 presents and discusses the results, highlighting the accuracy of each approach. Finally, Section 6 outlines the key findings and concludes with possible future work.

2. Related work

Fundamental (economic) and technical (derived from raw data) indicators influence price forecasting (Pandey & Upadhyay, 2016). Previous literature showed that technical indicators are desirable preferred tools for short-term prediction over fundamental approaches (Shynkevich, 2016). In Teixeira and De Oliveira (2010) financial technical indicators were used as inputs with statistical approaches and successfully forecast stock prices. Technical indicators are especially common with stock trading, building them from raw stock price and following the price movements over time to capture trends (Diego, Ignacio, Francisco, & Jose, 2009). The financial technical indicators use previous price information as inputs, often alongside machine learning algorithms to find relationships and thus forecast stock prices to achieve profitable returns (Gerlein, McGinnity, Belatreche, & Coleman, 2016).

☆ This work was supported by DfE, UK CAST scholarship in collaboration with Click Energy.

* Corresponding author.
E-mail addresses: mchugh-c24@ulster.ac.uk (C. McHugh), sa.coleman@ulster.ac.uk (S. Coleman), d.kerr@ulster.ac.uk (D. Kerr).

https://doi.org/10.1016/j.mlwa.2021.100182
Received 2 April 2021; Received in revised form 23 September 2021; Accepted 8 October 2021
Available online 16 October 2021

The three key types of price prediction indicators are (i) Trend, (ii) Oscillator, and (iii) Momentum (Tanaka-Yamawaki & Tokuoka, 2007). A similar approach of developing electricity price technical indicators and using these for day-ahead prediction is a natural progression in the advent of the energy trading market.

Fundamental indicators have been used as inputs into energy prediction models and robust correlation was noted between data from the same hour (Li, Arci, Reilly, Curran, & Belatreche, 2016). A study using separate hourly models for forecasting with the Spanish energy market demonstrated homogeneity was observed among hourly models compared with a single 24-hour model (García-Martos, Rodríguez, & Sánchez, 2007). Recently energy technical indicators have been created for day-ahead energy market prediction (Demir, Mincev, Kok, & Paterakis, 2020) using hourly models.

One of the difficulties with technical analysis is determining parameter optimisation. For example, a key feature to consider is the sliding window size as this corresponds to the amount of historical data used to calculate the technical indicators (Shynkevich, McGinnity, Coleman, Belatreche, & Li, 2017) and this can influence overall accuracy. If the prediction required is for day-ahead forecasting, generally the window size is set to 24 h (García-Martos et al., 2007; Li et al., 2016). Nonetheless, weekly or monthly forecasts could also be considered, and if so, the window size would become 168 h or 720 h respectively (Mei, He, Harley, Habetler, & Qu, 2014).

This paper presents eight novel electricity price technical indicators. Building on the concept in Demir et al. (2020) we generate optimal versions of our technical indicators for each hour. The newly developed indicators are included as inputs into three machine learning algorithms to predict short-term electricity prices. Initially we create one model that predicts for all 24 h and then we create 24 individual 1-hourly models. These are analysed to determine which approach is most accurate for modelling market trends and price predictions. A persistence model (baseline) forecasts prices on the assumption that the conditions stay unchanged from the current to the future time (Pedro & Coimbra, 2012). Hence, persistence machine learning models are trained using only raw price data. For each of the different models, performance and accuracy are evaluated and compared to determine if the use of technical indicators enhances forecasting accuracy when compared with raw price data.

3. Technical indicators

The proposed technical indicators are motivated by the standard financial trading indicators. This paper focusses on price-based technical indicators and as the ISEM requirement is day-ahead prediction, we are interested in technical indicators that improve short-term prediction accuracy. As discussed in Section 2, the main technical indicator classes are trend, oscillator, and momentum. The trend indicators focus on moving averages to indicate whether price movement is increasing or decreasing, the oscillator price indicators represent periodic patterns and the momentum indicator signifies market power and expectation level (Tanaka-Yamawaki & Tokuoka, 2007).

The calculations used to derive each technical indicator for both the 24-hour model and the hourly models are described here. For the 24-hour model, any calculation that involves a moving average is calculated using the previous 24 h (window size = 24) and does not include the current hour value.

1. Percentage Price Change Moving Average (PPCMA): A trend energy market indicator in which we calculate price change as the difference between the current price ($Price_i$) and the price from the same time period in the previous day ($Price_{Lag24}$) to capture daily trend, all divided by $Price_{Lag24}$. For the 24-hour model, the percentage price change moving average is calculated using a rolling 24-hour window ($s = 24$). For the hourly models, rolling windows from 1-hour to 150-hour ($s = 1, \ldots, 150$) are used. PPCMA is defined as follows:
$$PPCMA_s = \frac{1}{s}\sum_{i=1}^{s} PPC_i \tag{1}$$
where
$$PPC_i = \frac{Price_i - Price_{Lag24}}{Price_{Lag24}} * 100 \tag{2}$$
2. Moving Average Deviation (MAD): A trend energy market indicator that uses Price Change Moving Average (PCMA) to calculate the deviation rate of the current electricity price from PCMA. For the 24-hour model, the moving average deviation is calculated using a rolling 24-hour window ($s = 24$) and for the hourly models, $s = 1, \ldots, 150$. MAD is defined as:
$$MAD_s = \frac{Price_s - PCMA_s}{PCMA_s} \tag{3}$$
where
$$PCMA_s = \frac{1}{s}\sum_{i=1}^{s} \frac{Price_i - Price_{Lag24}}{Price_{Lag24}} \tag{4}$$
3. Percentage Range (PR): An oscillator energy market indicator that finds a relationship between current electricity price and the highest/lowest prices. This indicator oscillates between 0 and 100, with a value tending towards 100 signifying that the current electricity price is closer to the lowest price and a value towards 0 signifying that the current electricity price is tending towards the highest price. For the 24-hour model, the percentage range is calculated over a 24-hour window ($s = 24$). For the hourly models, $s = 1, \ldots, 150$ is considered to calculate the highest and lowest prices within a given window:
$$PR_s = \left[\frac{HighestPrice_s - Price_i}{HighestPrice_s - LowestPrice_s}\right] * 100 \tag{5}$$
where $Price_i$ is current price.
4. Average True Range (ATR): A trend energy market indicator measuring price volatility. Over a 24-hour window, there are three different values calculated: (a) highest price minus lowest price; (b) highest price minus a lagged (n) electricity price denoted as $Price_n$; and (c) lowest price minus a lagged (n) electricity price ($Price_n$). The maximum of these three values is selected for each trading hour and averaged over a rolling 24-hour window ($s = 24$ and $n = 24$). For the hourly models, a rolling window of length 1-hour to 150-hour is used and n is set to 24 to capture daily trend:
$$ATR_s = \frac{1}{s}\sum_{i=1}^{s} TR_i \tag{6}$$
where
$$TR_i = MAX\{A_s, B_s, C_s\} \tag{7}$$
and
$$A_s = HighestPrice_s - LowestPrice_s \tag{8}$$
$$B_s = |HighestPrice_s - Price_n| \tag{9}$$
$$C_s = |LowestPrice_s - Price_n| \tag{10}$$
5. Relative Strength Index (RSI): An oscillator energy market indicator that compares recent price gains to recent price losses. This indicator oscillates between 0 and 100, with a value close to 100 signifying that the majority of electricity price units within the period are *Price Up* and a value close to 0 signifying that the majority of electricity price units are *Price Down*. For the 24-hour model ($s = 24$), *Price Up* is the average of the previous 24 h when price difference increased, and *Price Down* is the average of the previous 24 h when price difference decreased. For the

hourly models, *Price Up* and *Price Down* were calculated from the average of the previous s hours with $s = 1, \ldots, 150$ and n was set to 24 to capture daily trend. *Price Up* and *Price Down* were determined by a piecewise function: if price difference was greater than 0 then *Price Up*, else *Price Down*:

$$RSI_s = 100 - \left[\frac{100}{D_s}\right] \tag{11}$$

where

$$D_s = \left(1 - \frac{\frac{1}{s}\sum_{i=1}^{s} PriceUp\left[Price_i - Price_n\right]}{\frac{1}{s}\sum_{i=1}^{s} PriceDown\left[Price_i - Price_n\right]}\right) \tag{12}$$

6. Average Directional Movement Index (ADX): A trend energy market indicator measuring the strength of the trend, grouping the two directional movement indexes depending on whether price change, calculated as current electricity price minus either previous 24-hour price (24-hour model) or previous s -hour price (hourly models), is grouped as a *Price Up* (positive) change or *Price Down* (negative) change. The two indexes are combined and smoothed with a moving average. For the 24-hour model, $s = 24$ and $n = 24$. For the hourly models, ($s = 1, \ldots, 150$) and n=24:

$$ADX_s = \frac{|DXUp_s - DXDown_s|}{|DXUp_s + DXDown_s|} \tag{13}$$

where

$$DXUp_s = \frac{\frac{1}{s}\sum_{i=1}^{s} PriceUp[Price_i - Price_n]}{ATR_s} \tag{14}$$

$$DXDown_s = \frac{\frac{1}{s}\sum_{i=1}^{s} PriceDown[Price_i - Price_n]}{ATR_s} \tag{15}$$

7. Moving Average Convergence/Divergence (MACD): An oscillator energy market indicator that takes into consideration the strength, direction, and duration of the trend as well as price momentum through moving averages of previous price values with rolling window sizes of $a = 12$ and $b = 24$ for the 24-hour model and rolling window sizes of $a = 168$ and $b = 336$ for the hourly models:

$$MACD = \frac{1}{a}\sum_{i=1}^{a} Price_a - \frac{1}{b}\sum_{i=1}^{b} Price_b \tag{16}$$

8. Price Momentum (PMOM): A momentum energy market indicator that evaluates the power of the market by examining the current electricity price with the previous trading value (1 h before) for the 24-hour model ($n = 1$). For the hourly models $s = 1, \ldots, 150$ and the current electricity price is considered with the previous trading value (n is equal to s minus the window size):

$$PMOM_s = Price_s - Price_n \tag{17}$$

4. Machine learning models

Machine learning models can be used to determine interesting patterns among data and make future predictions (Gupta, Mohanta, Chakraborty, & Ghosh, 2017). For machine learning prediction to be successful, a desirable model is one that avoids overfitting and provides transparency (Jansen, 2018). To predict the energy prices, the eight novel technical indicators (Section 3) were used as inputs to train three models using Gradient Boosting, Extreme Gradient Boosting (XGBoost), and Random Forest.

Boosting algorithms use sequential learning to create one strong model with a small error rate by combining many weak models (Gandhi, 2018a). A Gradient Boosting algorithm builds a strong prediction model by optimising a loss function with weak learner models (Dey, Kumar, Saha, & Basak, 2016) to minimise the error residuals (Qin, Wang, Li, & Ge, 2013). In this approach for predicting energy prices, a Gradient Boosting algorithm is implemented with the parameters determined to be 1000 trees, the minimum sample leaf set to 1, the minimum sample split set to 2, and the learning rate set to 0.1. An XGBoost regression algorithm is an advanced technique with extra features due to its ability to train with large datasets and its speed (Pathak, 2019). It functions through ensemble learning with weighted predictors (Dey et al., 2016). XGBoost is similar to Gradient Boosting, however it applies Newton boosting for approximation which improves model performance by applying an extra parameter for randomising to reduce correlation (Gandhi, 2018b). This research implements an XGBoost algorithm with 1000 trees, the fraction of column is set to 0.6, the fraction of observations is set to 0.8, the maximum depth of a tree is set to 4, and the learning rate is set to 0.05.

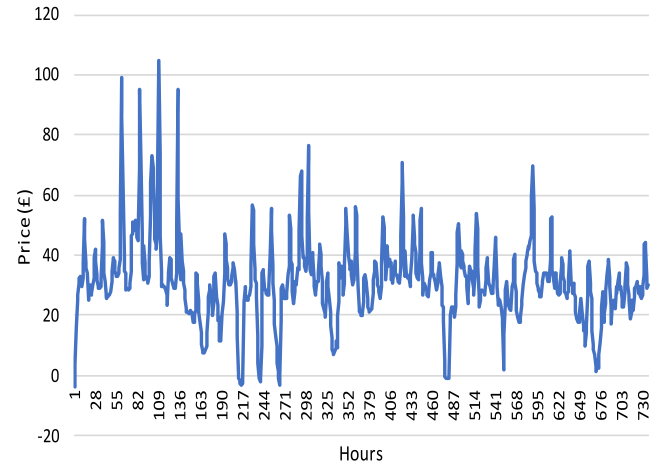

Fig. 1. Testing period actual electricity price data.

A Random Forest algorithm is another efficient ensemble regression technique which is robust to outliers, is expandable thus avoiding overfitting, adaptable, and is simple to tune (Mei et al., 2014). The input data are split by a tuning parameter which decides recursively when to generate new classifiers (Mulrennan, Donovan, Tormey, & Macpherson, 2018). During the training stage, no single tree sees the full training data as the training consists of bagging with multiple decision trees each divided at the nodes (Khaidem, Saha, & Dey, 2016). Feature importance is ranked from the inclusion of multiple trees and the accuracy improves as more trees are added (Jansen, 2018). The final prediction value is the average of the predictions from each of the individual trees (Pórtoles, González, & Moguerza, 2018). Here the parameters for the Random Forest algorithm are determined to be 1000 trees, the minimum sample leaf is set to 1, the minimum sample split is set to 2, and no pruning is included.

When the models are split by hour and trained using the technical indicators, each hourly model (0–23) has different parameters. The method of selecting an optimal lag factor (n) and the span (s) is based on the approach in Demir et al. (2020). In our approach, n is used for five of the proposed technical indicators in Section 3 (ADX, ATR, PMOM, PR, and RSI) and s is used for two of the proposed technical indicators (MAD and PPCMA). To find the optimal hourly technical indicators, a grid-search is applied with all the possible combinations of hyperparameters n and s ranging from 1 to 150 for the respective technical indicators. The Random Forest algorithm was used to determine the optimal n and s values for each individual hour. All possible combinations of n and s were generated and ranked from lowest to highest with respect to Root Mean Square Error (RMSE) which is calculated as the difference between the actual and the predicted electricity price. The n and s values that result in the smallest RMSE during testing were chosen as the optimal values.

Table 1
Summary results for 24-hour training models.

Model	Algorithm	EVS	RMSE	RMSLE	MedAE
Baseline	Gradient boosting	0.66	11.28	0.32	5.43
	XGBoost	0.59	12.44	0.34	5.94
	Random forest	0.83	8.07	0.27	3.45
Technical indicators	Gradient boosting	0.99	1.76	0.091	1.12
	XGBoost	0.99	2.21	0.093	1.30
	Random forest	0.99	2.18	0.088	0.77

EVS = Explained Variance Score, RMSE = Root Mean Square Error, RMSLE = Root Mean Square Log Error, MedAE = Median Absolute Error.

5. Results and discussion

The ISEM electricity price data used in the experiments were hourly records ranging from 1st February 2019 to 31st March 2020 retrieved from the Single Electricity Market Operator website SEMOpx. The proposed technical indicators, outlined in Section 3, are utilised as model inputs for the regression algorithms to predict electricity prices. First, the technical indicators are calculated using data from every hour and the window size is 24 h for each of the three machine learning algorithms. The 24-hour models were trained using the eight technical indicators with 85% of the data (9th September 2019 Hour 0 to 1st March 2020 Hour 5) and tested using the technical indicators with the remaining 15% of the data (1st March 2020 Hour 6 to 31st March 2020 Hour 23). The input variables are the technical indicators at time T and the target variable is the actual electricity price at time T+24. For comparison, and to determine if technical indicators improve prediction models, a persistence 24-hour model is trained as a baseline with the same train/test split. The input variable in this case is actual electricity price at time T and the output variable is actual electricity price at time T+24.

Fig. 1 displays the actual electricity price we are aiming to predict for the complete testing stage (738 h). The summary results are collected for both the training and testing stages. Well-known summary metrics are used to evaluate model performance. These are Explained Variance Score (EVS) where a score of 1 indicates the best possible outcome (Pedregosa et al., 2011), Root Mean Squared Error (RMSE) which is ideally close to zero, Root Mean Squared Log Error (RMSLE) a value closer to zero indicates improved performance, and Median Absolute Error (MedAE) which removes extreme values and reduces bias and should tend towards zero. The RMSLE metric was selected as one of the performance metrics as it applies relative error, includes a large penalty for underestimating, and is robust to outliers (Saxena, 2019). The EVS metric (Pedregosa et al., 2011) is calculated as follows:

$$EVS = 1 - \left[\frac{Var(Actual - Predicted)}{Var(Actual)}\right] \tag{18}$$

with values close to 1 indicating excellent model fit and lower values, in particular negative values, indicating that the model is overfitting. The RMSE metric (Saxena, 2019) is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Actual_i - Predicted_i\right)^2} \tag{19}$$

with a small RMSE value signifying that the predicted values on average match the actual values. The RMSLE metric (Saxena, 2019) is calculated as follows:

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(Actual_i + 1\right) - \log\left(Predicted_i + 1\right)\right)^2} \tag{20}$$

with a RMSLE value close to zero indicating robustness and an excellent model fit. The MedAE metric (Pedregosa et al., 2011) is calculated as follows:

$$MedAE = median\left(\left|A_1 - P_1\right|, \ldots, \left|A_n - P_n\right|\right) \tag{21}$$

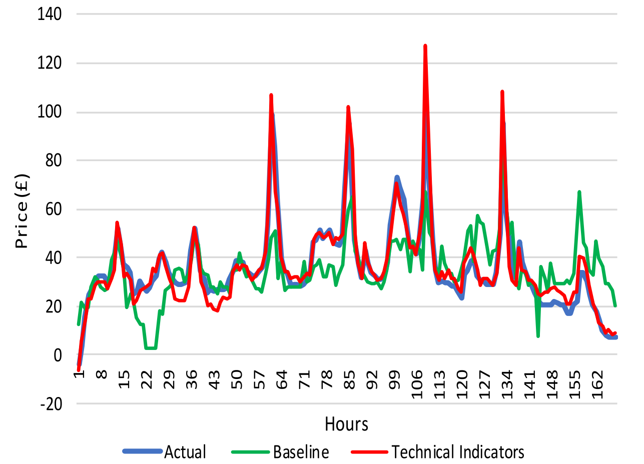

Fig. 2. 24-hour testing model gradient boosting.

Table 2
Summary results for 24-hour testing models.

Model	Algorithm	EVS	RMSE	RMSLE	MedAE
Baseline	Gradient boosting	0.03	13.39	0.39	7.90
	XGBoost	0.10	12.89	0.37	7.53
	Random forest	−0.29	15.43	0.41	9.85
Technical indicators	Gradient boosting	0.87	5.02	0.17	2.97
	XGBoost	0.84	5.59	0.19	3.24
	Random forest	0.82	6.17	0.20	3.30

EVS = Explained Variance Score, RMSE = Root Mean Square Error, RMSLE = Root Mean Square Log Error, MedAE = Median Absolute Error.

where A=Actual Price and P=Predicted Price. The lower the MedAE value, the less bias among the actual and predicted values, and the better the model fit.

Table 1 presents the results for both the baseline and technical indicator 24-hour models during the training stage for Gradient Boosting, XGBoost, and Random Forest. The EVS values for training ranged from 0.59 to 0.83 for the baseline models and was 0.99 for each of the technical indicator models. The baseline Random Forest model had the lowest RMSE value of 8.07 and the technical indicators Gradient Boosting model had the lowest RMSE value of 1.76. The Random Forest algorithm outputted the lowest RMSLE and MedAE values for both models (baseline and technical indicators). Observing both the baseline and technical indicator models, Random Forest performed the best overall.

Table 2 presents the results for both the baseline and technical indicator 24-hour models during testing for Gradient Boosting, XGBoost, and Random Forest. The testing EVS ranged from −0.29 to 0.10 for the baseline models and from 0.82 to 0.87 for the technical indicator models. Observing the baseline models, XGBoost performed the best with a RMSE value of 12.89, a RMSLE of 0.37, and a MedAE value of 7.53. When comparing the technical indicator models, Gradient Boosting performed the best with a RMSE value of 5.02, a RMSLE of 0.17, and a MedAE value of 2.97. Both the training and testing results indicate that including technical indicators as model inputs significantly improves forecasting performance.

Figs. 2, 3, and 4 illustrate the actual price values plotted against the baseline predicted price and technical indicators predicted price values in the first week of the testing period (168 h) for Gradient Boosting, XGBoost, and Random Forest respectively. Each of the figures shows how the technical indicators perform better than the baseline models with their values generally following the same trend as the actual price values and displaying a very close fit overall.

The next set of experiments involves the data being split by hour to enable individual hourly models to be trained using the three regression algorithms. To select optimal technical indicators for each hour,

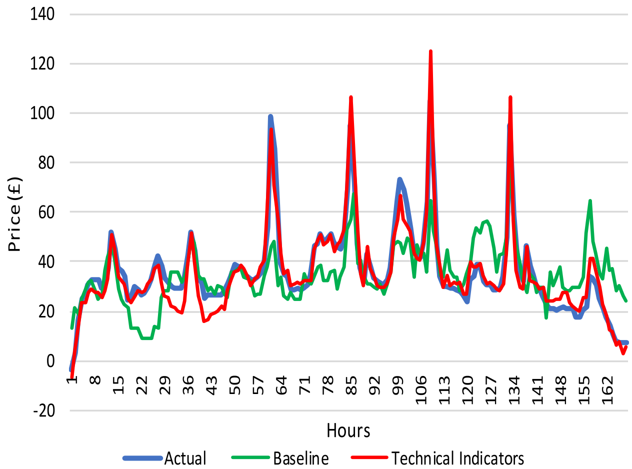

Fig. 3. 24-hour testing model XGBoost.

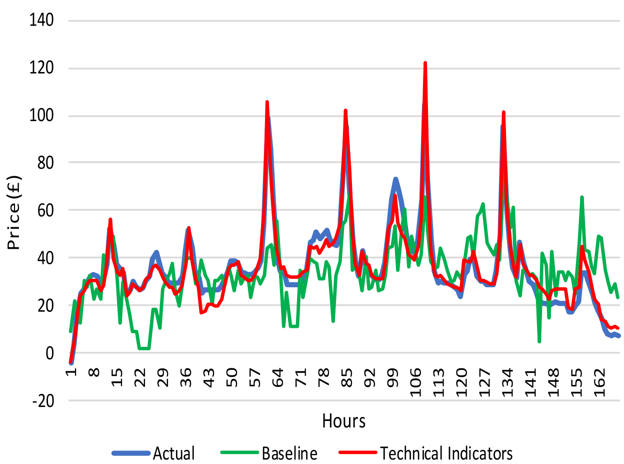

Fig. 4. 24-hour testing model random forest.

multiple versions of each indicator with varying n and s values are considered using a grid-search and calculated using data from February 2019 to March 2020. The optimal n and s, which are chosen using the Random Forest algorithm and by selecting the values that provided the lowest RMSE for each of the hourly technical indicator models, are displayed in Table 3.

For example, the optimal n is 24 and s is 91 for Hour 0 which means the derived technical indicators outlined in Section 3 ($PPCMA_s$, MAD_s, PR_n, ATR_n, RSI_n, ADX_n, MACD, and $PMOM_n$) are now $PPCMA_{91}$, MAD_{91}, PR_{24}, ATR_{24}, RSI_{24}, ADX_{24}, MACD, and $PMOM_{24}$ for Hour 0. As seen in the 24-hour models, no matter which regression algorithm is used, the technical indicator models still outperform the baseline models, therefore the optimal n and s chosen from the Random Forest algorithm are used for all hourly technical indicator experiments.

Similar to the 24-hour models, the hourly technical indicator models are trained with the eight technical indicators calculated using 85% of the data (9th September 2019 Hour 0 to 1st March 2020 Hour 5) and tested with the technical indicators using the remaining 15% of the data (1st March 2020 Hour 6 to 31st March 2020 Hour 23). The input variables are the technical indicators at time T and the target variable was the electricity price at time T+i. For comparison, hourly baseline models are trained with the same train/test split. This time for each hour, the actual electricity price at time T is the input and the actual electricity price at time T+24 is the output.

Tables 4, 5, and 6 present the testing results for Gradient Boosting, XGBoost, and Random Forest respectively for both the hourly baseline

Table 3
Optimal N and S values.

Hour	Optimal N	Optimal S
0	24	91
1	12	71
2	14	35
3	55	1
4	47	32
5	19	78
6	47	97
7	15	15
8	41	63
9	12	77
10	73	69
11	81	64
12	69	63
13	80	14
14	110	3
15	96	67
16	99	92
17	89	95
18	91	62
19	84	37
20	37	82
21	9	5
22	62	138
23	12	42

and hourly technical indicator models. In Table 4, the testing EVS ranges from −4.20 to −0.30 for Gradient Boosting persistence model and ranged from 0.49 to 0.99 for the technical indicator models. The RMSLE metric is robust to outliers and examines relative error introducing large penalties if the model underestimates, therefore it is also chosen to determine hour model performance. Hour 10 provides the best model performance out of the baseline models with a RMSLE value of 0.042. Hour 19 provides the best performance for the technical indicator models with a RMSLE value of 0.013. In Table 5, the testing EVS for XGBoost ranges from −2.61 to −0.30 in the baseline models and ranged from 0.64 to 0.96 in the technical indicator models. Hour 21 provides the best model performance out of the baseline models with a RMSLE value of 0.095. Hour 19 provides the best performance for the technical indicator models with a RMSLE value of 0.023. In Table 6 the testing EVS for Random Forest ranges from −2.16 to −0.22 for the baseline models and ranges from 0.83 to 0.99 for the technical indicator models. Hour 21 provides the best model performance out of the baseline models with a RMSLE value of 0.090. Hour 20 provides the best out of the technical indicator models with a RMSLE value of 0.010.

Reviewing Tables 4–6, the use of technical indicators significantly improves the price prediction performance compared with using the raw price data. Random Forest is the top performing machine learning algorithm, but all three regression algorithms perform well with the inclusion of technical indicators as inputs. This is noted with EVS changing from negative values for each of the best baseline models to values close to 1 for the best technical indicator models. Each of the four summary metrics (EVS, RMSE, RMSLE, and MedAE) also improves with the inclusion of technical indicators. Each of the tables demonstrates promising results and highlights that hourly models improve performance during electricity price forecasting.

Figs. 5, 6, and 7 illustrate the actual prices plotted against the predicted prices for each of the top performing hourly technical indicator models using the RMSLE during the testing period. Fig. 5 displays Hour 19 using Gradient Boosting, Fig. 6 displays Hour 19 using XGBoost, and Fig. 7 displays Hour 20 using Random Forest. Each plot displays the same hour of the day through the testing period. The figures show excellent similarity between the actual and predicted values. In particular, the Random Forest (Fig. 7) illustrates how well electricity prices can be predicted when using individual hourly models.

Table 4
Gradient boosting optimal testing models summary results.

	Baseline				Technical indicators			
Hour	EVS	RMSE	RMSLE	MedAE	EVS	RMSE	RMSLE	MedAE
0	−1.37	17.35	0.50	9.71	0.98	2.70	0.049	1.68
1	−1.20	17.20	0.49	12.46	0.98	2.74	0.083	2.67
2	−1.03	17.02	0.71	8.12	0.99	1.75	0.08	1.30
3	−0.78	16.26	0.81	7.69	0.99	1.49	0.24	0.96
4	−0.89	17.31	0.71	9.79	0.99	1.40	0.12	0.75
5	−0.49	15.75	0.64	7.78	0.97	2.24	0.27	1.25
6	−1.38	20.46	1.02	12.90	0.97	2.43	0.60	0.99
7	−1.12	20.65	0.64	15.83	0.87	5.38	0.44	4.31
8	−1.77	22.60	0.44	15.67	0.96	2.72	0.067	1.80
9	−0.30	14.43	0.22	9.60	0.89	4.38	0.052	2.30
10	−0.68	13.54	0.042	8.90	0.95	3.50	0.042	2.22
11	−1.08	13.91	0.17	8.78	0.88	3.41	0.055	2.89
12	−1.71	14.69	0.18	10.97	0.91	2.72	0.037	1.12
13	−1.81	13.86	0.17	7.14	0.89	2.86	0.041	2.01
14	−3.49	17.04	0.29	10.55	0.93	2.01	0.076	0.70
15	−4.20	18.79	0.21	8.88	0.95	1.63	0.027	1.01
16	−1.22	12.42	0.16	7.28	0.94	3.15	0.047	2.70
17	−1.49	20.63	0.20	16.14	0.94	3.40	0.041	2.78
18	−0.88	29.02	0.19	13.61	0.99	2.06	0.019	1.44
19	−1.81	22.36	0.18	11.03	0.98	2.04	0.013	0.76
20	−1.52	13.39	0.14	8.31	0.98	1.30	0.014	1.16
21	−3.38	10.53	0.11	4.71	0.49	3.26	0.042	1.70
22	−1.65	8.45	0.15	5.48	0.97	1.04	0.017	0.72
23	−1.55	14.19	0.25	12.13	0.89	3.12	0.048	1.71

EVS = Explained Variance Score, RMSE = Root Mean Square Error, RMSLE = Root Mean Square Log Error, MedAE = Median Absolute Error.

Table 5
XG boost optimal testing models summary results.

	Baseline				Technical indicators			
Hour	EVS	RMSE	RMSLE	MedAE	EVS	RMSE	RMSLE	MedAE
0	−0.82	15.14	0.35	8.53	0.89	4.01	0.23	2.23
1	−0.95	16.09	0.46	9.30	0.94	2.98	0.055	1.42
2	−0.40	14.13	0.54	5.88	0.93	3.12	0.11	1.86
3	−0.58	15.28	0.70	8.17	0.94	3.21	0.35	1.19
4	−0.33	14.54	0.60	9.81	0.94	3.01	0.23	1.01
5	−0.30	14.70	0.81	5.83	0.96	2.59	0.046	1.21
6	−0.86	18.16	0.93	9.83	0.95	3.91	0.57	3.49
7	−0.66	18.23	0.52	15.42	0.79	6.53	0.31	4.25
8	−1.28	20.55	0.29	11.26	0.91	4.23	0.051	2.51
9	−0.35	14.73	0.18	9.23	0.85	5.17	0.070	2.73
10	−0.30	11.96	0.14	6.63	0.93	3.79	0.040	2.36
11	−0.42	12.03	0.15	8.96	0.88	3.28	0.048	1.55
12	−1.01	13.07	0.16	7.54	0.93	2.19	0.029	1.15
13	−0.94	12.12	0.15	7.69	0.91	2.31	0.037	1.30
14	−2.18	14.97	0.21	9.80	0.83	3.35	0.037	1.46
15	−2.61	16.44	0.19	9.78	0.89	2.41	0.037	0.88
16	−0.59	10.99	0.14	6.51	0.83	3.89	0.054	3.05
17	−0.84	18.98	0.19	17.03	0.91	4.75	0.057	3.51
18	−0.56	26.70	0.18	12.24	0.96	4.63	0.030	2.74
19	−1.22	19.95	0.16	9.04	0.94	3.26	0.023	1.55
20	−0.72	11.15	0.11	6.99	0.95	2.15	0.024	0.95
21	−1.88	8.67	0.095	5.50	0.64	2.71	0.036	1.86
22	−0.88	7.08	0.12	4.45	0.94	2.28	0.041	1.78
23	−0.82	12.03	0.19	9.05	0.84	3.52	0.062	2.84

EVS = Explained Variance Score, RMSE = Root Mean Square Error, RMSLE = Root Mean Square Log Error, MedAE = Median Absolute Error.

Finally, sensitivity analysis was performed on the Hour 20 Random Forest model to explore the significance of each of the technical indicators. Sensitivity analysis examines robustness and model performance to determine which input parameters have the most influence (Kim, Kim, & Srebric, 2020). This analysis involved removing one technical indicator at a time from the model inputs and evaluating the model performance using the four summary metrics. This technique is known as parametric bootstrap as the factors are removed and the model re-evaluated after each replacement (Saltelli, 2002). The summary results for each technical indicator removed are displayed in Table 7.

Observing these results, PR was the most significant technical indicator as once removed the model accuracy significantly decreased (EVS=0.72, RMSE=4.40, RMSLE=0.047, MedAE=2.76). RSI was the least significant parameter as once removed the model accuracy remained similar to the original (EVS=0.99, RMSE=0.99, RMSLE=0.010, MedAE=0.61). PMOM also has low significance, although the RMSE has improved, the overall profile of the results is still not quite as good as using all the technical indicators. These findings suggest that Percentage Range is a strong technical indicator for electricity price forecasting.

6. Conclusion

This research presented eight novel energy technical indicators (PPCMA, MAD, PR, ATR, RSI, ADX, MACD, and PMOM) calculated from

Table 6
Random forest optimal testing models summary results.

Hour	Baseline				Technical indicators			
	EVS	RMSE	RMSLE	MedAE	EVS	RMSE	RMSLE	MedAE
0	−0.72	14.68	0.38	7.33	0.96	2.30	0.053	1.88
1	−0.92	15.97	0.46	9.44	0.97	1.91	0.14	1.38
2	−0.32	13.76	0.47	5.51	0.98	1.72	0.049	0.91
3	−0.30	13.84	0.71	7.79	0.99	1.45	0.21	0.99
4	−0.37	14.85	0.64	9.55	0.91	1.38	0.17	0.91
5	−0.23	14.32	0.69	7.51	0.99	1.43	0.19	0.84
6	−0.66	17.27	0.91	9.73	0.98	2.47	0.53	1.28
7	−0.61	17.97	0.49	13.99	0.90	4.60	0.24	3.02
8	−0.93	18.79	0.26	10.42	0.95	3.32	0.052	1.90
9	−0.32	14.57	0.17	10.22	0.92	3.62	0.050	1.95
10	−0.32	12.14	0.15	8.32	0.92	4.47	0.051	3.28
11	−0.39	11.81	0.15	8.60	0.84	3.79	0.050	2.72
12	−0.93	12.76	0.16	9.73	0.93	2.68	0.040	1.66
13	−0.86	12.22	0.15	6.29	0.92	2.07	0.033	1.39
14	−2.16	14.91	0.21	8.42	0.97	1.85	0.032	1.15
15	−1.76	14.4	0.18	9.60	0.96	1.44	0.023	0.84
16	−0.70	11.18	0.15	6.31	0.91	3.15	0.049	2.67
17	−0.60	19.16	0.19	16.04	0.92	3.80	0.045	3.42
18	−0.22	23.86	0.17	14.26	0.99	2.32	0.023	1.24
19	−0.81	18.04	0.15	10.19	0.99	1.71	0.014	1.17
20	−0.77	11.40	0.12	7.15	0.99	0.98	0.010	0.56
21	−1.18	7.73	0.090	5.99	0.83	2.26	0.029	1.46
22	−0.78	6.90	0.11	4.04	0.97	0.90	0.017	0.57
23	−0.81	11.98	0.18	9.11	0.89	2.96	0.044	2.03

EVS = Explained Variance Score, RMSE = Root Mean Square Error, RMSLE = Root Mean Square Log Error, MedAE = Median Absolute Error.

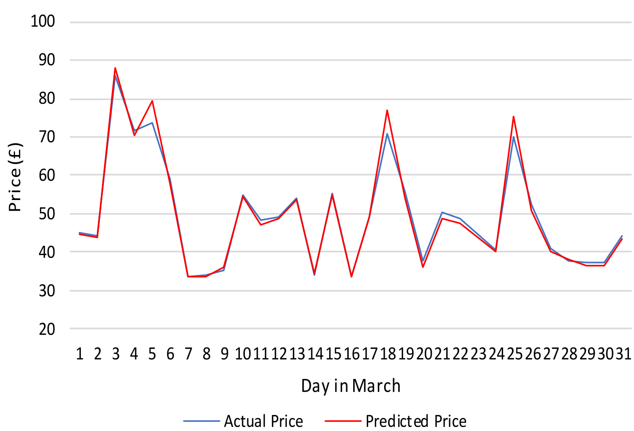

Fig. 5. Hour 19 testing model gradient boosting.

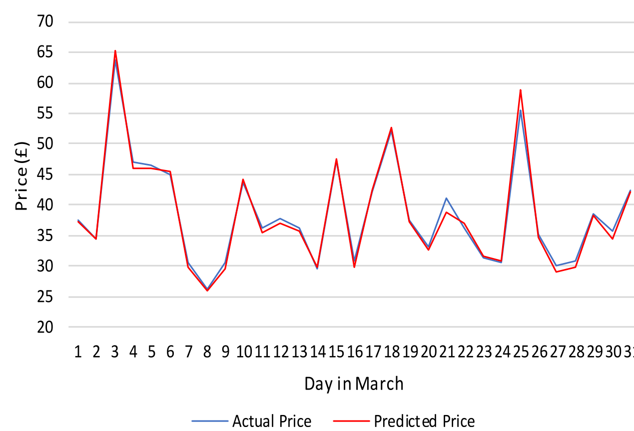

Fig. 7. Hour 20 testing model random forest.

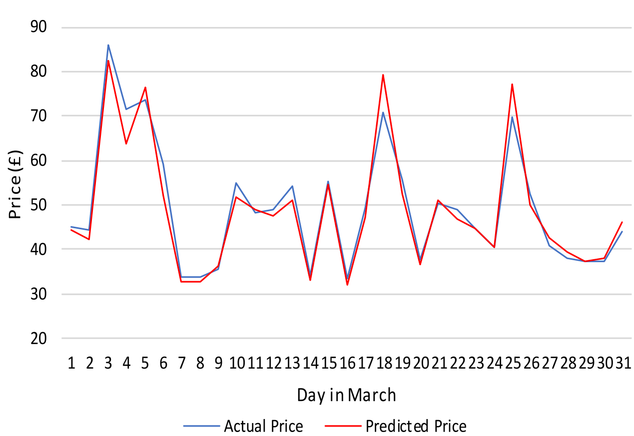

Fig. 6. Hour 19 testing model XGBoost.

Table 7
Random forest hour 20 sensitivity analysis.

Parameter removed	EVS	RMSE	RMSLE	MedAE
None	0.99	0.98	0.010	0.56
PPCMA	0.98	1.23	0.012	0.85
MAD	0.97	1.80	0.016	0.77
PR	0.72	4.40	0.047	2.76
ATR	0.98	1.02	0.010	0.69
RSI	0.99	0.99	0.010	0.61
ADX	0.99	1.03	0.010	0.68
MACD	0.99	1.03	0.010	0.72
PMOM	0.99	0.81	0.012	0.60

EVS = Explained Variance Score, RMSE = Root Mean Square Error, RMSLE = Root Mean Square Log Error, MedAE = Median Absolute Error.

raw electricity price data and were used as inputs into three regression algorithms (Gradient Boosting, XGBoost, and Random Forest) to predict electricity prices. The first set of experiments considered a 24-hour model approach and the second set of experiments focussed on hourly models to develop an optimal price prediction model. Both approaches were compared with baseline models which included only raw price data as input.

In the 24-hour approach, the three regression models EVS ranged between 0.59 to 0.83 for the baseline models and was 0.99 for each of the technical indicator models. The Random Forest training algorithm for both baseline and technical indicator models outputted the lowest RMSLE and MedAE values. The algorithms testing model EVS ranged between −0.29 to 0.10 for the baseline models and ranged between 0.82 to 0.87 for the technical indicator models. The XG Boost baseline testing model provided the lowest RMSE value of 12.89, the lowest RMSLE of 0.37, and the lowest MedAE value of 7.53. The Gradient Boosting technical indicators testing model provided the lowest RMSE value of 5.02, the lowest RMSLE of 0.17, and the lowest MedAE value of 2.97. The results concluded that including technical indicators as inputs significantly improves model performance. This was also confirmed in Figs. 2, 3, and 4 in which the predicted prices generally followed the same trend as the actual electricity price.

In the second approach, where each hour was modelled, the optimal n and s were selected for each hour during the testing period. The testing model EVS for each of the three regression algorithms ranged between −4.20 to −0.22 for the baseline hourly models and ranged between 0.49 to 0.99 for the technical indicator hourly models. Figs. 5, 6, and 7 illustrated excellent fits between actual and predicted electricity price values. These promising results highlight that hourly models do improve electricity prediction performance and therefore that they would be helpful in energy market trading. In particular, Percentage Range is an important technical indicator to include in an electricity price prediction model.

To conclude, energy technical indicators should be considered by energy traders to aid in capturing market trends and reducing costs. This will also benefit consumers in terms of savings because if energy traders can accurately predict electricity prices this can keep costs low in the retail market. A key foundation of the ISEM is the Day-Ahead market, thus both energy traders and consumers would profit from accurate price predictions (Mirakyan, Meyer-Renschhausen, & Koch, 2017). Possible future work will explore including other energy related factors such as wind generation or demand and developing associated technical indicators to determine if model performance can be improved further.

CRediT authorship contribution statement

Catherine McHugh: Conceptualisation, Methodology, Validation, Writing – original draft. **Sonya Coleman:** Conceptualisation, Supervision, Writing – review & editing. **Dermot Kerr:** Conceptualisation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was funded by DfE CAST scholarship in collaboration with Click Energy.

References

Amjady, N., & Hemmati, M. (2006). Energy price forecasting: Problems and proposals for such predictions. *IEEE Power and Energy Magazine*, *4*, 20–29. http://dx.doi.org/10.1109/MPAE.2006.1597990.

Demir, S., Mincev, K., Kok, K., & Paterakis, N. G. (2020). Introducing technical indicators to electricity price forecasting: A feature engineering study for linear, ensemble, and deep machine learning models. *Applied Sciences*, *10*(1), 255. http://dx.doi.org/10.3390/app10010255.

Dey, S., Kumar, Y., Saha, S., & Basak, S. (2016). Forecasting to classification: Predicting the direction of stock market price using xtreme gradient boosting. (pp. 1–10). http://dx.doi.org/10.13140/RG.2.2.15294.48968.

Diego, J., Ignacio, J., Francisco, J., & Jose, L. (2009). Multiobjective optimization of technical market indicators. In *Proceedings of the 11th annual conference companion on genetic and evolutionary computation conference* (pp. 1999–2004). http://dx.doi.org/10.1145/1570256.1570266.

Gandhi, R. (2018a). Boosting algorithms: AdaBoost, gradient boosting and XGBoost. https://hackernoon.com/boosting-algorithms-adaboost-gradient-boosting-and-xgboost-f74991cad38c. (Retrieved 6 January 2020).

Gandhi, R. (2018b). Gradient boosting and XGBoost. https://medium.com/hackernoon/gradient-boosting-and-xgboost-90862daa6c77. (Retrieved 6 January 2020).

Gao, G., Lo, K., & Fan, F. (2017). Comparison of ARIMA and ANN models used in electricity price forecasting for power market. *Energy and Power Engineering*, *09*(04), 120–126. http://dx.doi.org/10.4236/epe.2017.94B015.

García-Martos, C., Rodríguez, J., & Sánchez, M. J. (2007). Mixed models for short-run forecasting of electricity prices: Application for the Spanish market. *IEEE Transactions on Power Systems*, *22*(2), 544–552. http://dx.doi.org/10.1109/TPWRS.2007.894857.

Gerlein, E. A., McGinnity, M., Belatreche, A., & Coleman, S. (2016). Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications*, *54*, 193–207. http://dx.doi.org/10.1016/j.eswa.2016.01.018.

Gupta, S., Mohanta, S., Chakraborty, M., & Ghosh, S. (2017). Quantum machine learning-using quantum computation in artificial intelligence and deep neural networks: Quantum computation and machine learning in artificial intelligence. In *8th annual automation and electromechanical engineering conference* (pp. 268–274). http://dx.doi.org/10.1109/IEMECON.2017.8079602.

Jansen, S. (2018). *Hands-on machine learning for algorithmic trading* (pp. 309–358). Packt Publishing Ltd.

Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. (pp. 1–20). Retrieved from http://arxiv.org/abs/1605.00003.

Kim, M. K., Kim, Y. S., & Srebric, J. (2020). Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. *Sustainable Cities and Society*, *62*(June), Article 102385. http://dx.doi.org/10.1016/j.scs.2020.102385.

Li, P., Arci, F., Reilly, J., Curran, K., & Belatreche, A. (2016). Using artificial neural networks to predict short-term wholesale prices on the irish single electricity market. In *2016 27th Irish signals and systems conference* (pp. 1–10). http://dx.doi.org/10.1109/ISSC.2017.7983623.

Mei, J., He, D., Harley, R., Habetler, T., & Qu, G. (2014). A random forest method for real-time price forecasting in New York electricity market. In *IEEE power and energy society general meeting*, (pp. 1–5). http://dx.doi.org/10.1109/PESGM.2014.6939932.

Mirakyan, A., Meyer-Renschhausen, M., & Koch, A. (2017). Composite forecasting approach, application for next-day electricity price forecasting. *Energy Economics*, *66*, 228–237. http://dx.doi.org/10.1016/j.eneco.2017.06.020.

Mosbah, H., & El-Hawary, M. (2016). Hourly electricity price forecasting for the next month using multilayer neural network. *Canadian Journal of Electrical and Computer Engineering*, *39*, 283–291. http://dx.doi.org/10.1109/CJECE.2016.2586939.

Mulrennan, K., Donovan, J., Tormey, D., & Macpherson, R. (2018). A data science approach to modelling a manufacturing facility's electrical energy profile from plant production data. In *Proceedings - 2018 IEEE 5th international conference on data science and advanced analytics* (pp. 387–391). http://dx.doi.org/10.1109/DSAA.2018.00050.

Pandey, N., & Upadhyay, K. G. (2016). Different price forecasting techniques and their application in deregulated electricity market: A comprehensive study. In *International conference on emerging trends in electrical, electronics and sustainable energy systems* (pp. 1–4). http://dx.doi.org/10.1109/ICETEESES.2016.7581342.

Pathak, M. (2019). Using XGBoost in Python. https://www.datacamp.com/community/tutorials/xgboost-in-python. (Retrieved 6 January 2020).

Pedregosa, et al. (2011). Metrics and scoring: quantifying the quality of predictions. https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics. (Retrieved 5 November 2020).

Pedro, H. T. C., & Coimbra, C. F. M. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, *86*(7), 2017–2028. http://dx.doi.org/10.1016/j.solener.2012.04.004.

Pórtoles, J., González, C., & Moguerza, J. M. (2018). Electricity price forecasting with dynamic trees: A benchmark against the random forest approach. *Energies*, *11*(6), 1588. http://dx.doi.org/10.3390/en11061588.

Qin, Q., Wang, Q.-G., Li, J., & Ge, S. S. (2013). Linear and nonlinear trading models with gradient boosted random forests and application to Singapore stock market. *Journal of Intelligent Learning Systems and Applications*, *05*(01), 1–10. http://dx.doi.org/10.4236/jilsa.2013.51001.

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis*, *22*(3), 579–590. http://dx.doi.org/10.1111/0272-4332.00040.

Saxena, S. (2019). What's the difference between RMSE and RMSLE? https://medium.com/analytics-vidhya/root-mean-square-log-error-rmse-vs-rmlse-935c6cc1802a. (Retrieved 8 July 2020).

SEMOpx (2020). Day-ahead electricity price. https://www.semopx.com/market-data/market-results/. (Retrieved 2 November 2020).

Shynkevich, Y. (2016). Computational intelligence techniques for forecasting stock price movements from news articles and technical indicators (Thesis), Ulster University.

Shynkevich, Y., McGinnity, T. M., Coleman, S. A., Belatreche, A., & Li, Y. (2017). Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing*, *264*, 71–88. http://dx.doi.org/10.1016/j.neucom.2016.11.095.

Tanaka-Yamawaki, M., & Tokuoka, S. (2007). Adaptive use of technical indicators for the prediction of intra-day stock prices. *Statistical Mechanics and Its Applications*, *383*, 125–133. http://dx.doi.org/10.1016/j.physa.2007.04.126.

Teixeira, L. A., & De Oliveira, A. L. I. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Systems with Applications*, *37*(10), 6885–6890. http://dx.doi.org/10.1016/j.eswa.2010.03.033.