Forecasting electricity price in different time horizons: an application to the Italian electricity market

(Article begins on next page)

Forecasting electricity price in different time horizons: an application to the Italian electricity market

Mahmood Hosseini Imani, Ettore Bompard, Pietro Colella, Tao Huang

***Abstract*— Electricity price is a crucial element for market players to maximize their profits. In this context, the forecast of the hour-ahead, day-ahead, and week-ahead electricity prices plays a crucial role. The more accurate the prediction is, the lower the market risk is. In this paper, several machine learning algorithms (Support Vector Machine, Gaussian Processes Regression, Regression Trees, and Multi-Layer Perceptron) with different structures have been adopted to forecast Italian wholesale electricity prices. Considering different time horizons (hourly, daily, and weekly), their performances have been compared through several performance metrics, including Mean Absolute Error (MAE), R-index, Mean Absolute Percentage Error (MAPE), and the number of anomalies in which the forecast error passes a threshold.**

The investigation reveals that, in general, SVM and Tree-based models outperform other models at different time horizons.

***Index Terms*— Electricity price prediction, Different forecasting horizons, Italian electricity market, Machine learning, Prediction error distribution, PUN**

NOMENCLATURE

ANN	Artificial Neural Network
GPR	Gaussian Process Regression
GPR M.5/2	GPR Matern 5/2
GPR R.Q	GPR Rational Quadratic
GPR S.E	GPR Squared Exponential
ML	Machine Learning
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MGP	Italian day-ahead market
MLP	Multi-Layer Perceptron
PUN	National Single Price
NG	Natural Gas
SVM	Support Vector Machines
SVM C.G	SVM Coarse Gaussian
SVM F.G	SVM Fine Gaussian
SVR	Support Vector Regression

Mahmood Hosseini Imani, Ettore Bompard, Pietro Colella, Tao Huang are with the Politecnico di Torino, Dipartimento Energia, Turin, 10129, Italy.(e-mail:mahmood.hosseiniimani@polito.it, ettore.bompard@polito.it, pietro.colella@polito.it, and tao.huang@polito.it).

I. INTRODUCTION

ELECTRICITY market price has huge volatility, and this increases the risk for market players. Electricity Price forecasts from one hour to few days ahead have become of great interest to power portfolio organizers, generation companies (GENCOs), and consumers [1].

Due to the importance of the electricity market price prediction, several approaches have been proposed so far. Artificial Intelligence (AI) has attracted considerable attention over the previous years in electricity price prediction. Many researchers focused on how a particular class of AI, Machine Learning methods (ML), can improve the time series[2]–[5]. In the field of prediction, the objective of ML methods is improving prediction accuracy by minimizing some loss function, and in most cases, the sum of squared errors [6], [7].

One of the most widely used methods is the Artificial Neural Network (ANN). The ANN computational procedures imitate the learning method of the human brain. The performance of an ANN is based on the structure of the network [8]. The most common artificial neural network utilized in the price forecasting process is the Multi-Layer Perceptron (MLP) [9]. Other algorithms adopted in this field are Support Vector Machines (SVM) and Gaussian Process Regression (GPR). They are powerful kernel-based ML methods for data analysis [10]. SVM and GPR are commonly supervised learning algorithms [11], [12]. Support vector regression (SVR) is a generalization of the SVM to the regression problem [13]. Moreover, some studies have utilized Decision Tree Algorithms, particularly Regression Trees, for price prediction problems [14]. To improve the performance of Tree-Based models, researchers have applied Boosting and Bagging techniques in the prediction process [14], [15].

In the last years, several multi-step frameworks to forecast electricity prices were developed. In particular, an attractive combined model that includes variational mode decomposition, mixed data modeling, feature selection, generalized regression neural network, and gravitational search algorithm is proposed in [16]. Another example is the hybrid deep-learning framework, which includes the feature preprocessing module, the deep learning-based point prediction step, the error compensation, and the probabilistic prediction modules [17].

Another method available in literature adopts the components estimation techniques. This approach requires filtering out the deterministic structural components from the original time series and modeling the residual component utilizing some stochastic process. The final forecast is obtained by combining the predictions of both these components [18].

Large-scale studies comparing the performance of machine learning models for regression problems have focused almost exclusively on the RESs generation prediction and load consumption prediction [19]–[21]. However, there are very few comprehensive comparison studies (if any) for electricity price prediction on different time horizons.

In the previous paper [22], we have compared the main machine learning algorithms described above. In the present paper, we have extended the comparison considering different time horizons, including hourly, daily, and weekly. We ranked the algorithms based on four performance metrics, MAE, R, MAPE, and the number of anomalies. Moreover, by analyzing the error tails using the probability density function, we determined the overall prediction performance [23], [24]. In addition, we evaluated the impact of cyclical input data on prediction performance.

This paper is organized as follows. Section II proposed an overview of conducted methods and strategies for price prediction. In Section III, we described the metrics adopted to compare the forecasting methods. Section IV provides information about the input data of the chosen case study. In Section V, the outcome of the comparisons is discussed.

II. Algorithms and Strategy For Price Prediction

In this section, we provided a brief description of the adopted methodologies.

A. *Support Vector Regression (SVR)*

SVR is categorized as a supervised learning method performed by a Support Vector Machine (SVM) to solve regression problems [10]. The principal object of SVR is determining hyperplanes that maximize the margin between classes [25].

SVR is categorized based on the type of kernel [26], [27] , which can be: Linear, Quadratic, Cubic, and Gaussian (Fine, and Coarse).

B. *Gaussian Process Regression (GPR)*

Gaussian Process is a collection of random variables; each subset of variables has a joint Gaussian distribution.

GP is characterized by the mean function $m(x)$ and the covariance function $k(x_1, x_2)$ [28]. Eq. (1)- (3) can explain a real process $f(x)$ as a GP.

$$f(x) \sim Gp(m(x), k(x_1, x_2)) \tag{1}$$

where:

$$m(x) = \mathbb{E}[f(x)] \tag{2}$$

$$k(x_1, x_2) = \mathbb{E}[(f(x_1) - m(x_1))(f(x_2) - m(x_2))] \tag{3}$$

In the regression model, considering a dataset D with N observations; $D = \{ (x_i, y_i) | \, i = 1, \dots, N\}$, with $x_i \in \mathbb{R}^D$ and $y_i \in \mathbb{R}$ the goal is to predict new y_* given x_* using $f(x)$ such that: $y_i = f(x_i) + \delta_i$ where δ_i is Gaussian noise with mean zero and variance δ^2.

In the GPR method, various types of kernel classes can be used. The most important types of kernel used in this article are described below [28], [29]:

- Exponential Covariance
- Squared Exponential Covariance
- Rational Quadratic Covariance
- Matern Class Covariance

C. *Tree-based Methods*

Tree-based methods or Decision tree algorithms have a flowchart-like structure. Regression trees are a type of decision tree where the target variable can take numerical values. Regression trees models can be applied to models having both a large number of observations and many variables.

Each decision tree consists of "root nodes", “decision nodes”, and Leaf or Terminal nodes [30]. The minimum number of leaf size is a key factor to tune a Tree-based model. If it is too large, the accuracy of the model will be reduced. Vice-versa, if it is too small, the risk of overfitting will increase.

Some techniques, which are often called “ensemble methods”, have been used to improve the performance of Tree-based models. Among them, Bagging and Boosting are very popular, and we investigated them in this paper [31].

Bootstrap aggregation or Bagging Trees is a learning method for improving the forecast by reducing the variance related to prediction; it averages the results to achieve an overall forecast.

Boosting Tree is another method for improving the result of the prediction. Different from bagging, boosting tree methods use weighted average outcomes to achieve the forecast. Moreover, in boosting methods, in each step, each tree is grown based on the information related to previously grown trees [31].

D. *Multi-Layer Perceptron*

MLP is a feed-forward neural network. The artificial neurons, which are also called nodes, are organized as the Input layer, hidden layer (one or more), and output layer. In the input layer, neurons receive the input data; in the intermediate hidden layers, they elaborate them, and in the output layer, they provide the output variables. Output nodes and hidden neurons use an activation function, such as Sigmoid tangent functions [32]. For training the model, we adopted the Levenberg-Marquardt BackproPagation (BP) algorithm to solve the non-linear least squares problem [10].

E. *Price Prediction Strategy*

There are different methods to forecast multiple horizons [33]. In this study, we used the Iterative Forecasting method: the first subsequent horizon is predicted using the full set of input variables; then, the predicted value is used as an input to forecast the next horizon. The process is carried on until the end of the forecast horizon. Due to using the previous forecasted value, for a longer horizon, the results may deteriorate.

III. Performance Evaluation Metrics

The common evaluation metrics for a regression problem are Mean Absolute Error (MAE), Pearson correlation coefficient (R), and Mean Absolute Percentage Error (MAPE). Eq. (4), eq (5), and eq. (6) defines MAE, R, and MAPE respectively [20],

[34], where:

- x_j corresponds to the actual value;
- y_j is the forecasted value;
- $\bar{x}$ is the average of the actual output;
- $\bar{y}$ is the average of the predicted output;
- n is the number of observations.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|x_j - y_j\right| \tag{4}$$

$$R = \frac{\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2}\sqrt{\sum_{j=1}^{n}(y_j - \bar{y})^2}} \tag{5}$$

$$MAPE = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{x_j - y_j}{x_j}\right| \tag{6}$$

For the prediction j, the error is defined by eq. (7).

$$Error_j = \left|x_j - y_j\right| \tag{7}$$

In this paper, for analyzing and comparing the performance of the forecasting methods, we defined an index named "anomaly index", which represents the number of forecasts characterized by an error greater than a specific threshold, which is defined in two steps: first, a line (the red one in Fig. 1) is fitted on the ascending percentage error plot; second, the threshold is set considered the maximum value of the red line. The percentage errors higher than the threshold line (yellow) are considered as anomalies. The rationale is to provide a different kind of signal to market participants, for which the forecast reliability is crucial. A model with a higher number of anomalies causes a serious risk to decision-makers, who are loss-averse. This kind of participant can adjust its bidding strategy using a model with a lower total number of anomalies in price prediction.

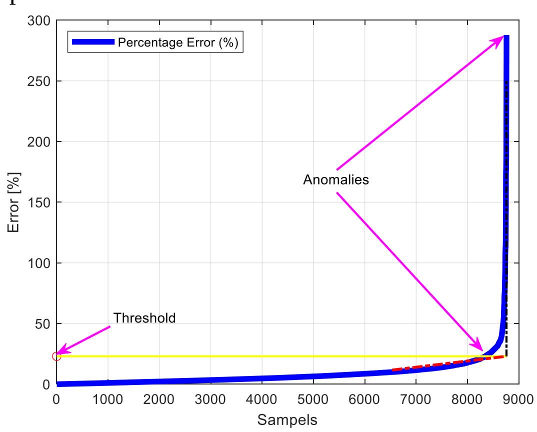

Fig. 1. Finding anomalies

IV. Italian Electricity Market Case Study: Input Data and Settings

In this section, an introduction to the Italian power market is presented (subsection A). Detailed information on the processed input data is provided in subsection B. A descriptive statistic on the Italian electricity market is provided in subsection C. Preprocessing of the data is presented in subsection D. The method for determining the train and test set of the data is provided in subsection E.

A. Introducing the Italian Power Market

In the Italian Power Market, the geographical market includes 13 foreign virtual zones, 6 geographical zones, and 1 pole of limited production (national virtual zone) [35]. In the Italian day-ahead market (MGP), the submissions of market participants take place between the ninth day before the day of physical delivery (opens at 8 AM) and the day before the day of delivery (closes at 12 p.m) [35].

The MGP National Single Price (PUN) formation follows the following process:

• in the first step, the supply and demand curves are constructed based on the merit order criterion by collecting all the supply and demand bids from all the zones;

• in the second step, the interzonal transmission line limits are checked. If the interzonal flows are within the line limits, the market-clearing price is identified by the intersection of the demand and supply curves [36].

Currently, considering taxes and dispatching costs, the MGP National Single Price (PUN) is around 35% of the Italian final electricity bills.

In this study, several algorithms are adopted to predict the MGP National Single Price [1](PUN), which is defined as the average of zonal prices in the MGP, weighted for total purchases and net of purchases for pumped-storage units and of purchases by neighboring countries' zones [35].

B. Input Data

The electricity wholesale price depends on several factors, such as load forecasts, fuel costs, weather data, scheduled maintenance. In our study, we selected only relevant and public data based on an extensive literature review [3], [9]. The variables adopted to forecast PUN are listed in Table 1.

All the data, including load consumption, electricity price, and Natural Gas (NG) price, are taken from Gestore dei Mercati Energetici (GME) for 2017, 2018, 2019, and 2020 [35]. Data are gathered on an hourly basis, except for NG price, which is gathered on a daily basis. In our dataset, all variables were complete, and there was not any gap in the data.

As mentioned before, in the iterative forecasting method to predict the next step in the longer horizon, the model uses the previous forecasted value instead of the actual value. In the daily forecast, the data related to the previous 24 hours average price will be updated in each step. On a weekly horizon, the previous 24 hours average price and the previous day same hour price will be updated in the next step. Due to inaccessibility to

[1] Prezzo Unico Nazionale

data related to the previous day NG price in the weekly forecast, this input variable is removed in the week-ahead prediction.

Each year consists of 8,760 hours (except 2020, which was a leap year, which consisted of one additional day making 8,784 hours), resulting in 35064 observations for the given period.

The hours from 08:00 to 20:00 on working days are considered as peak hours, and all the hours of non-working days and the hours from 00:00 to 08:00 and from 20:00 to 24:00 on working days are considered as off-peak hours [35].

In the hourly forecast (one-step-ahead), all the predictors are considered in the training process. In the daily forecast, the predictor related to the previous 24 hours average price is updated in each iteration, replaced by a new forecasted price. In the weekly forecast, the previous day's natural gas price is unavailable. Previous 24 hours average electricity price and previous day same hour electricity price are updated by new predicted prices.

TABLE 1

ALL VARIABLES FOR PUN PREDICTION (✓: AVAILABLE, ✗: NOT AVAILABLE, AND **U**: UPDATED IN EACH ITERATION)

Input Variables	Units	Hourly	Daily	Weekly
Hour	1-24	✓	✓	✓
Hour type	0(Peak), 1(Off-Peak)	✓	✓	✓
Week Day	1-7	✓	✓	✓
Day type	0(weekday) 1(weekend)	✓	✓	✓
Season type	1-4	✓	✓	✓
Forecasted load	MWh	✓	✓	✓
Previous day same hour load	MWh	✓	✓	✓
Previous week same hour load	MWh	✓	✓	✓
Previous 24 hours average load	MWh	✓	✓	✓
Previous day same hour price	€/MWh	✓	✓	**U**
Previous week same hour price	€/MWh	✓	✓	✓
Previous 24 hours average price	€/MWh	✓	**U**	**U**
Previous day NG price	€/MWh	✓	✓	✗
Previous week average NG price	€/MWh	✓	✓	✓

C. *Descriptive Statistics on the Italian electricity market*

The raw data of hourly PUN and load consumption in each year from 2017 to 2020 are depicted in Fig. 2. The correlation between PUN and load is also presented in this figure.

The boxplot of annual PUN statistics and annual load statistics from 2017 to 2020 are plotted in Fig. 3. TABLE 2 presents a descriptive statistic of annual load and price for four years.

Based on Fig. 3 and Table 2, after 2018, it can be seen that there is a decreasing trend of load, which leads to a decrease in the PUN trend.

D. *Data Preprocessing*

To maximize the performance of the models, we preprocess the data as described below.

1) *Data normalization*

We normalized the data so that each value is bounded between 0 and 1 based on eq. (8) [37]:

$$q^* = \frac{q - Q_{min}}{Q_{max} - Q_{min}} \tag{8}$$

where q^* is the scaled value of q, and Q_{max} and Q_{min} are the maximum and minimum value of dataset Q.

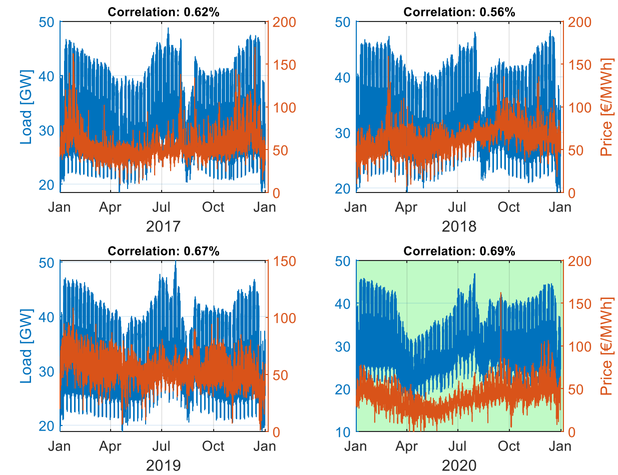

Fig. 2. Electricity load consumption and PUN in each year

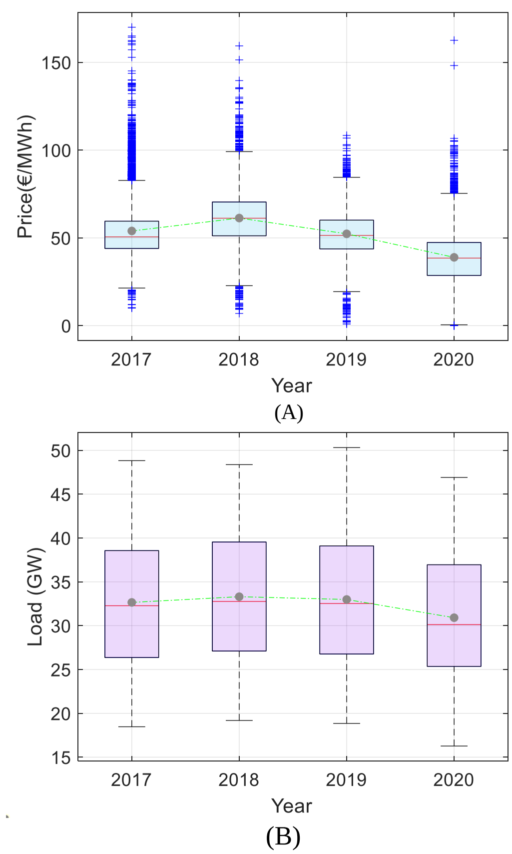

Fig. 3. Boxplot of annual PUN statistics (A) and annual load statistics (B). On each box central red line indicates the median; the grey circle indicates the mean, and the upper and lower edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to 1.5 times the interquartile range, and outlier appears as a blue + sign, is the value away from whiskers range, green dashed line indicated the yearly trend

TABLE 2
DESCRIPTIVE STATISTICS OF ANNUAL LOAD AND PRICES (2017-2020)[38]

Year		2017	2018	2019	2020
Load (MW)	Max	50333	52412	53824	49964
	Min	17212	17610	17014	15334
	Mean	33271	33934	33579	31102
	Std	7568	7555	7731	7449
PUN (€/MWh)	Max	170	159.4	108.4	162.6
	Min	10	6.97	1	0
	Mean	53.95	61.31	52.32	38.91
	Std	16.46	14.84	12.68	14.65

2) Handling cyclical input data

Hours of the day, days of the week, seasons in a year are all examples of input data that are cyclical. There is one major issue with these cyclical input data. In the prediction methods, the cyclical feature such as the hour of a day does not hold any numerical ordering ('1' < '2' < '3'), and some machine learning methods will not be able to understand that naturally. To overcome this issue, there are some different methods to handle such input data. They involve creating dummy variables and representing the cyclical input data as (x, y) coordinates on a circle: the first approach, dummy encoding, a dummy variable takes values of 0 and 1, where the values indicate the presence or absence of a group membership. in the second method, trigonometric encoding, each cyclical variable be mapped onto a circle which the lowest value for that variable appears right next to the largest value. The transferred values of that point can be calculated using cos and sin trigonometric functions [39].

$$x_1' = \cos(\frac{2\pi x}{C_n}) \tag{9}$$

$$x_2' = \sin(\frac{2\pi x}{C_n}) \tag{10}$$

where x is the sample before conversion, and C_n is the period of the cycle.

In this study, we have evaluated the impact of these methods to increase forecast accuracy.

E. Determining training and testing sets

Training and validation sets are used to fit the model and tune the model hyperparameters, respectively. A test set is utilized to provide an unbiased estimation of the final model [40].

In our study, by using the holdout method, we split up the dataset into train, validation, and test (prediction) sets. 55% of the dataset is considered a train, 20% validation, and 25% a test set. In other words, the period Jan 1, 2017, to Dec 31, 2019, is considered a train and validation, and the entire period of 2020 is considered the test set.

V. DISCUSSION ON THE RESULTS

The objective of this study is to compare the performance of the methods described in section II for predicting electrical prices at different time horizons. To carry out a more extensive analysis, we compared different versions/structures of the same algorithm. For the algorithms SVM and GPR, the different kernels are described in subsection II.A and II.B were implemented, respectively. For MLP, we built neural networks characterized by two hidden layers, and we changed the number of neurons per layer: the notation MLP {x,y} means that the MLP model has x neurons in the first hidden layer and y neurons in the second one. Based on the literature, there is currently no theoretical reason to use neural networks with more than two hidden layers [41]. Carrying out some preliminary tests, we found that two hidden layers can generally minimize errors and improve prediction accuracy. There are many methods for defining the correct number of neurons in the hidden layers [42]–[45]. Based on the criteria in the mentioned literature and the trial-and-error approach, we selected the number of neurons in each hidden layer between 5 to 10 in the first hidden layer and 5 to 15 in the second hidden layer. For the sake of completeness and to show the impact of the number of neurons, we analyzed six different structures in the paper.

Finally, Tree-based methods described in subsection II.C were implemented, considering the different number of leaf sizes; the notation adopted to describe the properties of tree-based methods is composed by the name of the algorithm (Bagged Trees or Boosted Trees) followed by the minimum number of leaf size. The leaf size is a crucial factor in tuning a Tree-based model. If it is too large, the accuracy of the model will be reduced. Vice-versa, if it is too small, the risk of overfitting will increase.

Table 3 ranks the compared forecasting models on different performance metrics in the hour-ahead horizon. The models are implemented in MATLAB software on an Intel(R) Core(TM) i7-8700 @ 3.20GHz, RAM 32GB system.

Each numerical value is the mean of the metrics computed for each test data sample (8784 hours – year 2020). The results in Table 3 show that the SVM methods generally outperform the remaining ML methods in all performance metrics, except the total number of anomalies.

As we explain in section II.E, we used the iterative forecasting method to obtain forecasted values for daily and weekly horizons. All performance comparisons reported in Table 4 are based on the average performance of 366 days forecast of test data. As expected, by increasing the forecasting horizon, from hour-ahead to day-ahead, the accuracy of the new forecasts depends on the accuracy of the previous ones, which results in deteriorated predictions on a longer horizon. Based on Table 4, considering MAE, the group of GPR method was more sensitive to increasing the prediction horizon, an average increase of 3.28 €/MWh on MAE. Vice-versa, the SVM methods were less sensitive with an average increment of 1.5 €/MWh on MAE. TABLE 5 allows us to investigate the performance obtained by each method across the weekly horizon. In the transition from hourly horizon to weekly horizon, on average, the MAE of the SVM group increased by 1.7 €/MWh. The deterioration in the MAE for MLP, Tree-based, and GPR is different, with an increment of 2.3, 1.51, and 4.41 €/MWh, respectively.

It is also worth mentioning that the results are the same as the previous one in the transition from day-ahead to a week ahead horizon. In addition, to seek more insight into the results listed in the previous tables, we have compared the total anomalies in percentage error for each method in different horizons.

TABLE 3

RANKING OF DIFFERENT PREDICTION METHODS IN HOUR-AHEAD FORECAST CONSIDERING SEVERAL INDEXES: MAE, R, MAPE [20], [34], AND ANOMALIES [22]

Rank	Methods order by MAE	MAE	Methods order by R	R	Methods order by MAPE	MAPE	Methods order by Anomalies	Anomalies	Methods	References
1	SVM Quadratic	3.703	SVM Quadratic	0.945	SVMLinear	15.833	Bagged Trees 30	82	Bagged Trees 10	[14], [31]
2	SVMLinear	3.769	SVM C.G	0.944	GPR M.5/2	15.979	Bagged Trees 10	84	Bagged Trees 20	[14], [31]
3	MLP {5 5}	3.873	SVMLinear	0.941	SVM Quadratic	16.162	Bagged Trees 20	95	Bagged Trees 30	[14], [31]
4	SVM C.G	3.885	SVM Cubic	0.936	MLP {5 5}	16.308	Bagged Trees 40	96	Bagged Trees 40	[14], [31]
5	MLP {10 5}	3.901	GPR Exponential	0.936	SVM Cubic	16.421	SVM F.G	108	Boosted Tress 10	[14], [31]
6	MLP {10 15}	3.952	MLP {10 5}	0.935	MLP {10 15}	16.66	GPR Exponential	118	GPR Exponential	[28]
7	SVM Cubic	4.093	GPR M.5/2	0.934	MLP {10 5}	16.797	GPR R.Q	121	GPR Matern 52	[28]
8	MLP {5 15}	4.183	MLP {5 5}	0.931	MLP {10 10}	16.887	Boosted Tress 10	126	GPR R.Q	[28]
9	GPR M.5/2	4.305	MLP {5 15}	0.928	MLP {5 15}	16.91	MLP {5 5}	132	GPR S.E	[28]
10	MLP {10 10}	4.325	GPR S.E	0.925	GPR S.E	17.037	SVM C.G	144	MLP {5 5}	[32]
11	MLP {5 10}	4.349	MLP {5 10}	0.924	MLP {5 10}	17.097	SVM Quadratic	145	MLP {5 10}	[32]
12	GPR Exponential	4.537	MLP {10 10}	0.921	SVM C.G	19.656	SVMLinear	148	MLP {5 15}	[32]
13	GPR S.E	4.729	GPR R.Q	0.919	GPR Exponential	21.549	MLP {5 15}	149	MLP {10 5}	[32]
14	Boosted Tress 10	5.063	MLP {10 15}	0.919	Bagged Trees 30	24.674	SVM Cubic	179	MLP {10 10}	[32]
15	GPR R.Q	5.229	Boosted Tress 10	0.915	Bagged Trees 40	24.947	MLP {10 10}	179	MLP {10 15}	[32]
16	Bagged Trees 40	5.255	Bagged Trees 40	0.906	GPR R.Q	27.139	MLP {10 5}	186	SVM C.G	[26], [27]
17	Bagged Trees 30	5.5	Bagged Trees 30	0.905	Bagged Trees 20	27.369	GPR S.E	200	SVM Cubic	[26], [27]
18	Bagged Trees 10	5.765	Bagged Trees 20	0.894	Boosted Tress 10	27.696	GPR M.5/2	201	SVM F.G	[26], [27]
19	Bagged Trees 20	5.774	Bagged Trees 10	0.887	Bagged Trees 10	30.106	MLP {10 15}	225	SVM Quadratic	[26], [27]
20	SVM F.G	19.23	SVM F.G	0.549	SVM F.G	103.61	MLP {5 10}	254	SVMLinear	[26], [27]

TABLE 4

RANKING OF DIFFERENT PREDICTION METHODS IN DAY-AHEAD FORECAST CONSIDERING SEVERAL INDEXES: MAE, R, MAPE [20], [34], AND ANOMALIES [22]

Rank	Methods order by MAE	MAE	Methods order by R	R	Methods order by MAPE	MAPE	Methods order by Anomalies	Anomalies	Methods	References
1	SVMLinear	4.359	SVMLinear	0.921	MLP {5 15}	17.251	Bagged Trees 20	98	Bagged Trees 10	[14], [31]
2	SVM Quadratic	4.403	SVM C.G	0.921	MLP {5 10}	17.312	Bagged Trees 10	101	Bagged Trees 20	[14], [31]
3	MLP {5 5}	4.419	SVM Quadratic	0.92	MLP {5 5}	17.476	Bagged Trees 30	103	Bagged Trees 30	[14], [31]
4	MLP {5 15}	4.515	MLP {5 5}	0.914	MLP {10 15}	17.773	SVM F.G	114	Bagged Trees 40	[14], [31]
5	MLP {10 15}	4.637	MLP {5 10}	0.913	MLP {10 10}	18.019	Bagged Trees 40	119	Boosted Tress 10	[14], [31]
6	MLP {5 10}	4.702	MLP {5 15}	0.91	MLP {10 5}	18.592	Boosted Tress 10	127	GPR Exponential	[28]
7	MLP {10 5}	4.714	MLP {10 15}	0.91	SVM Quadratic	18.938	GPR R.Q	142	GPR Matern 52	[28]
8	SVM C.G	4.725	MLP {10 5}	0.906	SVMLinear	19.181	GPR Exponential	142	GPR R.Q	[28]
9	MLP {10 10}	4.859	MLP {10 10}	0.904	SVM Cubic	22.917	GPR Matern 52	144	GPR S.E	[28]
10	SVM Cubic	6.009	Boosted Tress 10	0.887	GPR S.E	23.888	SVM C.G	151	MLP {5 5}	[32]
11	Boosted Tress 10	6.116	SVM Cubic	0.867	SVM C.G	24.37	SVMLinear	156	MLP {5 10}	[32]
12	Bagged Trees 40	6.663	Bagged Trees 40	0.867	Bagged Trees 40	28.652	SVM Quadratic	158	MLP {5 15}	[32]
13	Bagged Trees 10	6.725	Bagged Trees 30	0.858	GPR R.Q	28.966	MLP {10 5}	181	MLP {10 5}	[32]
14	Bagged Trees 20	7.001	Bagged Trees 20	0.853	Bagged Trees 30	30.082	MLP {5 5}	182	MLP {10 10}	[32]
15	GPR S.E	7.032	Bagged Trees 10	0.845	Bagged Trees 20	31.02	MLP {5 10}	183	MLP {10 15}	[32]
16	Bagged Trees 30	7.355	GPR S.E	0.835	Boosted Tress 10	31.693	MLP {10 15}	189	SVM C.G	[26], [27]
17	GPR R.Q	7.395	GPR R.Q	0.813	Bagged Trees 10	32.826	SVM Cubic	197	SVM Cubic	[26], [27]
18	GPR Matern 52	7.985	GPR Exponential	0.79	GPR Matern 52	32.954	MLP {10 10}	202	SVM F.G	[26], [27]
19	GPR Exponential	9.371	GPR Matern 52	0.789	GPR Exponential	39.325	GPR S.E	205	SVM Quadratic	[26], [27]
20	SVM F.G	20.3	SVM F.G	0.502	SVM F.G	106.64	MLP {5 15}	223	SVMLinear	[26], [27]

TABLE 5

RANKING OF DIFFERENT PREDICTION METHODS IN WEEK-AHEAD FORECAST CONSIDERING SEVERAL INDEXES: MAE, R, MAPE [20], [34],AND ANOMALIES [22]

Rank	Methods order by MAE	MAE	Methods order by R	R	Methods order by MAPE	MAPE	Methods order by Anomalies	Anomalies	Methods	References
1	SVM Quadratic	5.057	SVM Quadratic	0.899	SVMLinear	22.153	SVM F.G	113	Bagged Trees 10	[14], [31]
2	SVMLinear	5.112	SVMLinear	0.897	SVM Quadratic	22.585	Bagged Trees 30	115	Bagged Trees 20	[14], [31]
3	MLP {5 10}	5.407	SVM C.G	0.894	MLP {5 10}	24.219	MLP {10 5}	118	Bagged Trees 30	[14], [31]
4	SVM C.G	5.56	MLP {5 15}	0.885	MLP {10 15}	24.346	Bagged Trees 20	125	Bagged Trees 40	[14], [31]
5	MLP {5 15}	5.676	MLP {5 10}	0.884	MLP {5 5}	24.939	Bagged Trees 40	125	Boosted Tress 10	[14], [31]
6	MLP {10 15}	6.319	Bagged Trees 10	0.867	SVM Cubic	25.292	Boosted Tress 10	130	GPR Exponential	[28]
7	Bagged Trees 10	6.513	MLP {10 5}	0.866	MLP {5 15}	25.466	GPR R.Q	132	GPR Matern 52	[28]
8	MLP {10 10}	6.537	MLP {5 5}	0.864	GPR Matern 52	28.181	GPR Exponential	132	GPR R.Q	[28]
9	MLP {5 5}	6.541	Boosted Tress 10	0.859	SVM C.G	28.795	Bagged Trees 10	137	GPR S.E	[28]
10	SVM Cubic	6.713	MLP {10 15}	0.856	MLP {10 10}	29.448	GPR Matern 52	138	MLP {5 5}	[32]
11	Boosted Tress 10	6.88	MLP {10 10}	0.854	MLP {10 5}	32.223	SVM C.G	152	MLP {5 10}	[32]
12	Bagged Trees 40	7.196	Bagged Trees 40	0.849	GPR R.Q	34.072	SVMLinear	153	MLP {5 15}	[32]
13	Bagged Trees 30	7.309	SVM Cubic	0.848	Boosted Tress 10	34.564	MLP {5 10}	158	MLP {10 5}	[32]
14	Bagged Trees 20	7.311	Bagged Trees 20	0.845	Bagged Trees 10	34.867	MLP {10 10}	158	MLP {10 10}	[32]
15	GPR Matern 52	7.354	Bagged Trees 30	0.838	GPR S.E	36.416	SVM Quadratic	164	MLP {10 15}	[32]
16	MLP {10 5}	8.156	GPR Matern 52	0.819	Bagged Trees 20	42.39	MLP {10 15}	181	SVM C.G	[26], [27]
17	GPR R.Q	8.244	GPR S.E	0.817	Bagged Trees 30	43.811	GPR S.E	188	SVM Cubic	[26], [27]
18	GPR S.E	9.685	GPR R.Q	0.796	Bagged Trees 40	43.855	SVM Cubic	192	SVM F.G	[26], [27]
19	GPR Exponential	11.196	GPR Exponential	0.758	GPR Exponential	45.88	MLP {5 15}	195	SVM Quadratic	[26], [27]
20	SVM F.G	20.716	SVM F.G	0.504	SVM F.G	107.9	MLP {5 5}	255	SVMLinear	[26], [27]

The results show that having better performance in MAE and R index does not necessarily mean having fewer outliers in the percentage error. As seen in the results, methods with low MAE might result in higher total anomalies.

To make easier the visualization of the performance, the radar graph for four considered indicators are depicted in Fig. 4 - Fig. 7. It is necessary to mention that the indicators are normalized between 0 and 1. Also, to have a better overview, we inverted the MAE, the Total number of anomalies, and MAPE, i.e., we apply the 1−normalizes metric. In this regard, the models with the best performance should present a performance closer to 1 for all indicators. The best model is the one with the largest sides.

Due to the worse performance of the SVM fine Gaussian method compared to other methods in the R index, the values related to this method are eliminated in the radar graph illustration. Looking at the results, we can see that the SVM model kept good performances as the prediction horizon grew considering MAE, R index, and MAPE. Moreover, on longer horizon, after SVM, Tree-based models have an impressive improvement. Fig. 4 also affirms, in most cases in the MLP model, an inverse relationship between the number of neurons and performance improvement. The higher the number of neurons, the lower the accuracy. GPR models have worse performance in MAE, R, and MAPE metrics, particularly in the longer horizon. Bagged Trees have better performance in considering the number of anomalies.

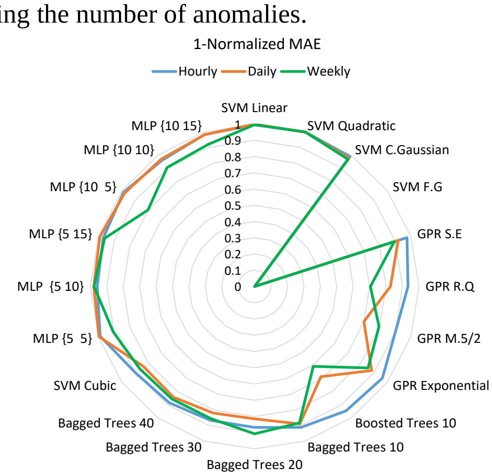

Fig. 4. 1-normalized MAE; best performance = 1, worst performance= 0.

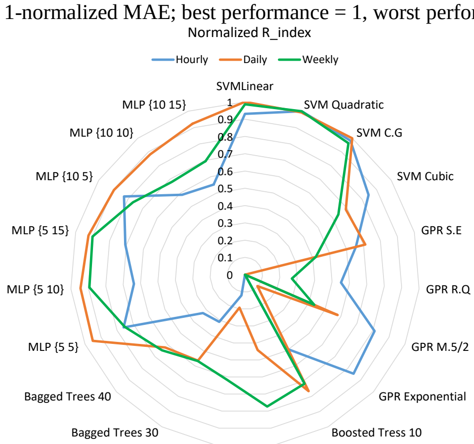

Fig. 5. R index; best performance = 1, worst performance= 0.

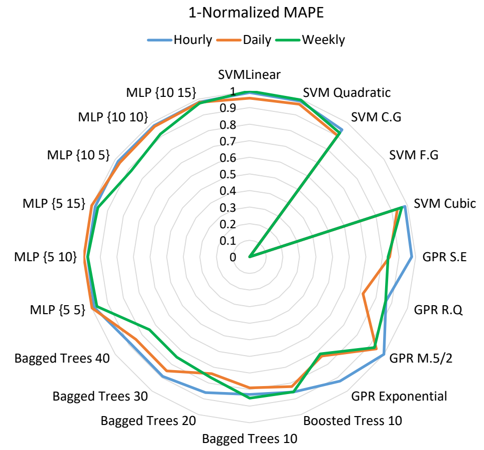

Fig. 6. 1-normalized MAPE; best performance = 1, worst performance= 0..

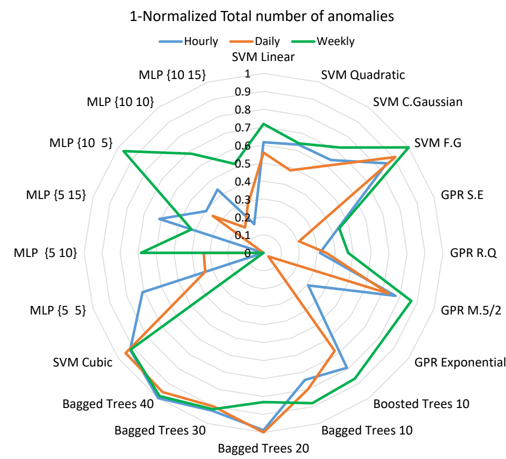

Fig. 7. 1-normalized total number of anomalies; best performance = 1, worst performance= 0.

o present the performance of the models in the day-ahead forecast, Fig. 8 shows the prediction results of four methods as a representative of each group, that is, SVM Linear, Bagged Regression Trees with 40 leaves, MLP with 5 neurons in the first hidden layer and 5 neurons in the second hidden layer, and GPR Squared Exponential. The results are produced on test data for an example day,11-Nov-2020. Based on this figure, in the peak price, the Bagged tree always follows the same price of the previous day, but other methods adjusted the predicted price based on the average of the previous week and the previous day. It is necessary to mention that the behavior of prediction methods is different in other seasons.

Fig. 9 draws the results of the week-ahead prediction on test data for an example week, 46th week of 2020.

As discussed in TABLE 5, bagged SVM and MLP models outperform GPR and bagged trees in the week-ahead horizon. In comparison to the day-ahead forecast, the performance of GPR is more sensitive to increasing the prediction horizon and much weaker than other algorithms. Although the prediction accuracy of GPR decreased by increasing the horizon, it follows the trend of the actual price.

Distributions of model errors for hourly, daily, and weekly horizon forecast are visualized in Fig. 10 to Fig. 12. In these figures, prediction error distributions of the four selected

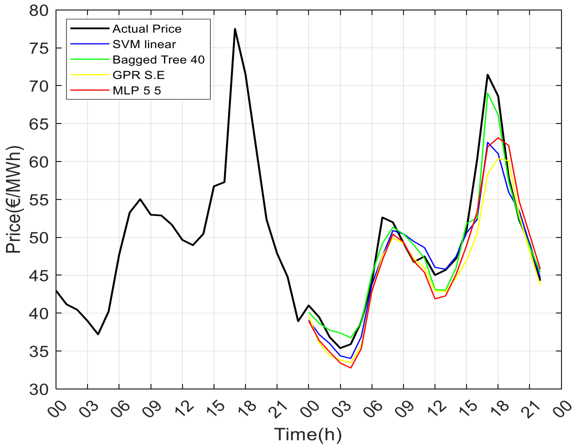

Fig. 8. The forecasts of four methods (SVM Linear, GPR S.E, Bagged Regression tree40, and MLP {5 5}) as a representative of each group for 316th day of 2020 (11-Nov-2020).

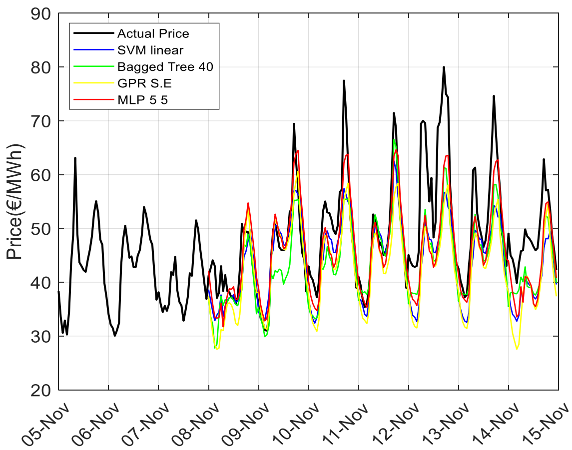

Fig. 9. The forecasts of four methods (SVM Linear, GPR S.E, Bagged Regression tree40, and MLP {5 5}) as a representative of each group for 46th week of 2020 (8-Nov-2020 to 15-Nov-2020).

methods are compared. Higher central error peak and thinner tails are the key features of a successful model in the forecast process [24].

In the electricity market, participants often prefer to use a model with lower and thinner tail errors in forecasting electricity prices. In fact, a model with fat error tails increases the probability of inaccurate predictions and causes a serious risk to market participants. It is also worth mentioning that a model with lower and thinner tail risk is more interesting to investigators in the electricity market; even the model has a higher general performance considering general performance indexes, e.g., MAE [24].

In the hourly-ahead forecast, Fig. 10, the SVM and MLP models have a higher central error peak. On the other hand, GPR has a lower central error peak and fat tail error. In general, the figure reflects TABLE 3.

In the daily horizon, as shown in Fig. 11, we can see the error distributions model got worse regarding the hourly-ahead forecast. Increasing the forecast horizon from one-hour-step to 24-hour-step yields fatter tails error, with decreasing central error peak. Based on this figure, a significantly lower central error peak and fatter tail error in the GPR model are apparent. The obtained performance of the GPR model indicates a certain inconsistency in the predictability of electricity price in the transition from hour-ahead to day-ahead forecast.

According to Fig. 12, the transition from day-ahead to week-ahead prediction makes a slightly rightward shift of the error distribution and increases error tail mass in all methods. The transition results in more deterioration of the GPR and MLP prediction performance. Unlike with GPR, increasing the time forecasting horizon causes less impact on the performance of the SVM model. The error distribution in Fig. 12 also reflects TABLE 5.

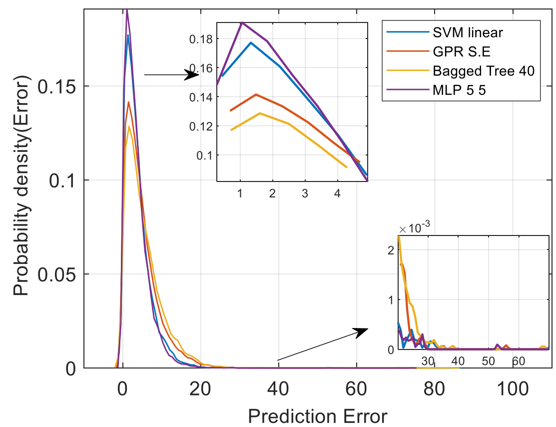

Fig. 10. Kernel density estimates of forecast Error for Hourly-Horizon

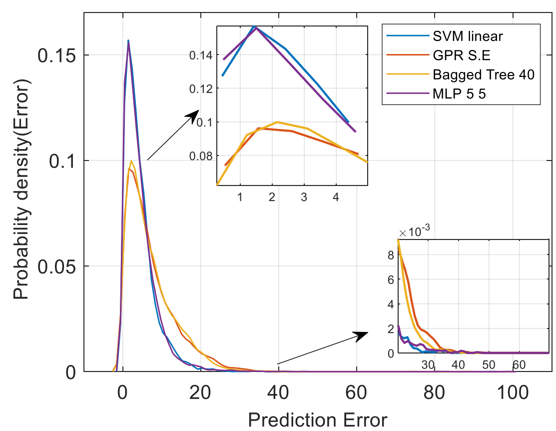

Fig. 11. Kernel density estimates of forecast Error for Daily-Horizon

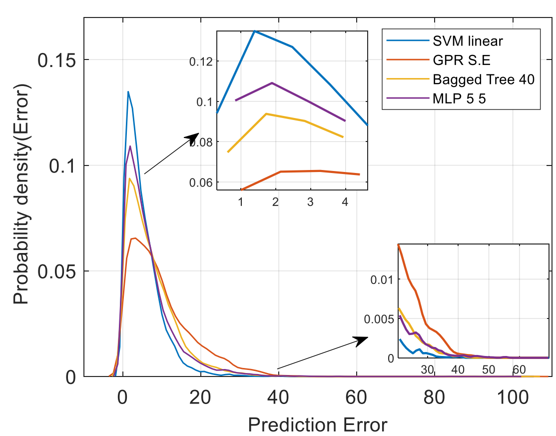

Fig. 12. Kernel density estimates of forecast Error for Weekly-Horizon

Computational complexity is a crucial aspect of comparing different machine learning algorithms. The Big O notation is used to rank the algorithms by how they operate depending on the dataset's dimension they are working on [46].

Based on the [47], for a training dataset of size N, SVM has training time complexity $O(N^2)$ or $O(N^3)$ depending on the kernel. On the other hand, GPR models suffer from $O(N^3)$ training computational complexity, making it difficult in the massive dataset [48].

In Decision tree algorithms, consider that the dataset has N observation, and d numerical features. When the numerical feature needs to be sorted, sorting one feature takes $O(N * log(N))$. Then sorting all the features take $O(N * log\, N * d)$.To calculate Information gain at each threshold, for one feature it takes $O(N)$, and for d features, it takes $O(N * d)$. Then computational complexity is $O(N * log(N) * d)$+ $O(N * d)$. Finally, in the Big O notation, $O(N * log(N) * d)$+ $O(N * d)$ will be almost equal to $O(N * log(N) * d)$ [49].

The computational complexity of MLP is related to the architecture of the model. The computational complexity for training MLP with two hidden layers is $O(N * w)$ where $w = i * j + j * k + k * l$ and i denotes the number of nodes of the input layer, j the number of nodes in the first hidden layer, k the number of nodes in the second hidden layer and l the number of nodes in the output layer [50]. The ranking of different machine learning algorithms is listed in Table 6.

Table 6
RANKING OF DIFFERENT PREDICTION METHODS BASED ON THE TIME COMPLEXITY (BIG O NOTATION)

Rank	Methods order by Time complexity	Big O notation	References
1	MLP	$O(N * w)$	[50]
2	Decision Tree	$O(N * \log(N) * d)$	[49]
3	SVM	$O(N^2)$ or $O(N^3)$	[47]
4	GPR	$O(N^3)$	[48]

Cyclical input data in the time series can have a different impact on the subsequent forecasting performance. Therefore, it is important to estimate it. As discussed in paragraph IV.D.2), two methods are popular to manage cyclical data: the dummy encoding and the trigonometric encoding. We applied both. For the first one, 24 dummies are created representing hours of a day, 7 dummies for days of a week, and 4 dummy variables for seasons of a year. For the second, a single variable is transformed into two variables by taking *cos* and *sin* of data.

We considered the day-ahead time horizon and four methods as a representative of each ML group.

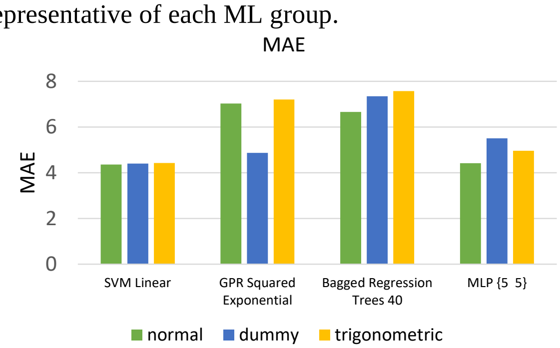

Fig. 13. Impact of encoding cyclical input data considering MAE index

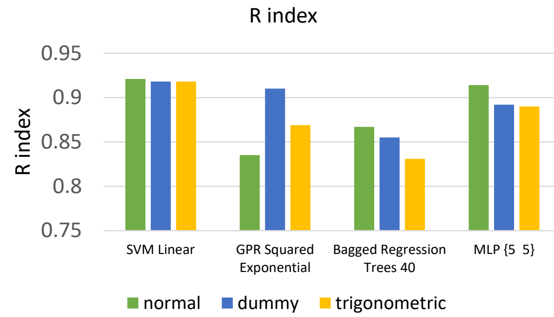

Fig. 14. Impact of encoding cyclical input data considering R index

Bar charts in Fig.14-Fig.17 depict the MAE, R, MAPE, and the number of anomalies, respectively. In the figures, three cases are represented: in the first, no algorithm is adopted to manage cyclical input data (green bars); in the second, the dummy encoding method is used (blue bars); in the third, the trigonometric encode method is adopted (yellow bars).

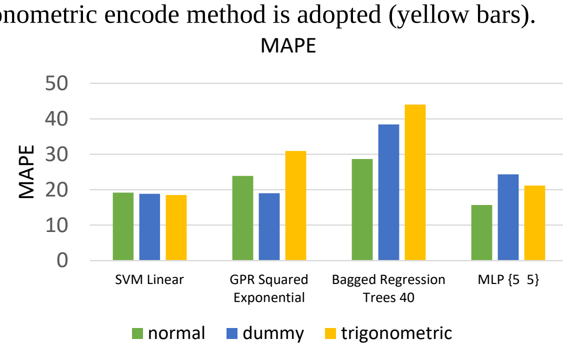

Fig. 15. Impact of encoding cyclical input data considering MAPE index

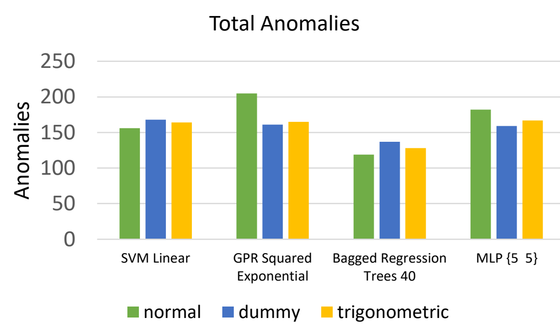

Fig. 16. Impact of encoding cyclical input data considering the total number of anomalies

Performance indicators recorded a statistically slight increment in forecast error. The exception is for GPR Squared Exponential Method, whose performance after transferring data using dummy variables is improved.

VI. CONCLUSION

In this study, the applicability and accuracy of four relevant predictive models (SVM, GPR, MLP, and Tree-Based methods) in hour-ahead, day-ahead, week-ahead electricity price prediction of the Italian electricity market were investigated. To obtain a multiple-horizon forecast, the Iterative forecasting method was applied. Data obtained from Gestore dei Mercati Energetici (GME), Italian energy markets manager, were used in the applications. Based on the dataset's availability, several inputs from the electricity market were considered, such as the electricity prices and the demand forecast. Besides, we considered the natural gas price, which specifically impacts the Italian electricity market. The outcomes were compared through several performance metrics, including MAE, R, MAPE, and the total number of anomalies. Also, the algorithms are ranked based on the computational time complexity (big O). Along with using the total number of anomalies as a performance indicator, the kernel smoothing function was used to analyze the error tails in each prediction method.

The investigation on the results at above twenty models has identified SVM models as suited to forecast hourly wholesale electricity price. Furthermore, it was found that an irrational increase in the complexity of the MLP model does not lead to increased accuracy.

Investigating the results in day-ahead prediction reveals that SVM and MLP model had a certain consistency in the prediction, transition from hour-ahead to day-ahead forecast.

Across the weekly horizon, comparison studies can show that SVM and MLP models are not only the appropriate models in performance accuracy but also well suited to reduce the error tails risk. In this regard, these two methods are the appropriate models for decision-makers in the Italian electricity market.

The impact of cyclical input data on subsequent forecasting performance was evaluated in the last part of this study. Considering different performance metrics, we have concluded that encoding the cyclical input data using dummies and trigonometric function leads to worse results than without encoding. The results of this study are advantageous for generator companies, utility companies, retailers, and large industrial consumers who can predict the volatile wholesale prices with a rational level of accuracy to adjust its bidding strategy, reduce the risk or maximize the profits in the day-ahead electricity market.

There exist several avenues for future studies. Since the comparison of the prediction methods of this work is based on the Italian day-ahead market (MGP), investigating the performance of the models on other electricity markets, e.g., Intra-Day Market (MI) and Ancillary Services Market (MSD), is more advantageous for market participants. Moreover, developing and comparing price prediction models would be interesting for the longer horizon, e.g., month-ahead horizon.

References

[1] X. R. Li, C. W. Yu, S. Y. Ren, C. H. Chiu, and K. Meng, “Day-ahead electricity price forecasting based on panel cointegration and particle filter,” *Electr. Power Syst. Res.*, vol. 95, pp. 66–76, 2013.

[2] A. Bello, D. W. Bunn, J. Reneses, and A. Muñoz, “Medium-term probabilistic forecasting of electricity prices: A hybrid approach,” *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 334–343, 2016.

[3] D. Keles, J. Scelle, F. Paraschiv, and W. Fichtner, “Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks,” *Appl. Energy*, vol. 162, pp. 218–230, 2016.

[4] G. Klæboe, A. L. Eriksrud, and S.-E. Fleten, “Benchmarking time series based forecasting models for electricity balancing market prices,” *Energy Syst.*, vol. 6, no. 1, pp. 43–61, 2015.

[5] M. Halužan, M. Verbič, and J. Zorić, “Performance of alternative electricity price forecasting methods: Findings from the Greek and Hungarian power exchanges,” *Appl. Energy*, vol. 277, p. 115599, 2020.

[6] E. Alpaydin, “Introduction to Machine Learning, Massachusetts.” Cambridge, MIT Press, 2004.

[7] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[8] M. M. Lotfinejad, R. Hafezi, M. Khanali, S. S. Hosseini, M. Mehrpooya, and S. Shamshirband, “A comparative assessment of predicting daily solar radiation using bat neural network (BNN), generalized regression neural network (GRNN), and neuro-fuzzy (NF) system: A case study,” *Energies*, vol. 11, no. 5, p. 1188, 2018.

[9] F. Keynia, “A new feature selection algorithm and composite neural network for electricity price forecasting,” *Eng. Appl. Artif. Intell.*, vol. 25, no. 8, pp. 1687–1697, 2012.

[10] J. Fan *et al.*, “Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China,” *Renew. Sustain. Energy Rev.*, vol. 100, pp. 186–212, 2019.

[11] D. C. Sansom, T. Downs, and T. K. Saha, “Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants,” *J. Electr. Electron. Eng. Aust.*, vol. 22, no. 3, p. 227, 2003.

[12] M. Mojaddady, M. Nabi, and S. Khadivi, “Stock market prediction using twin Gaussian process regression,” *Int. J. Adv. Comput. Res. Prepr.*, 2011.

[13] J. M. Moguerza and A. Muñoz, “Support vector machines with applications,” *Stat. Sci.*, vol. 21, no. 3, pp. 322–336, 2006.

[14] C. González, J. Mira-McWilliams, and I. Juárez, “Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests,” *IET Gener. Transm. Distrib.*, vol. 9, no. 11, pp. 1120–1128, 2015.

[15] L. Breiman, “Arcing classifier (with discussion and a rejoinder by the author),” *Ann. Stat.*, vol. 26, no. 3, pp. 801–849, 1998.

[16] A. Heydari, M. M. Nezhad, E. Pirshayan, D. A. Garcia, F. Keynia, and L. De Santoli, “Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm,” *Appl. Energy*, vol. 277, p. 115503, 2020.

[17] C. Zhang, R. Li, H. Shi, and F. Li, “Deep learning for day-ahead electricity price forecasting,” *IET Smart Grid*, vol. 3, no. 4, pp. 462–469, 2020.

[18] I. Shah, H. Bibi, S. Ali, L. Wang, and Z. Yue, “Forecasting one-day-ahead electricity prices for italian electricity market using parametric and nonparametric approaches,” *IEEE Access*, vol. 8, pp. 123104–123113, 2020.

[19] M. A. Mohandes, “Modeling global solar radiation using Particle Swarm Optimization (PSO),” *Sol. Energy*, vol. 86, no. 11, pp. 3137–3145, 2012.

[20] M. Guermoui, K. Gairaa, A. Rabehi, D. Djafer, and S. Benkaciali, “Estimation of the daily global solar radiation based on the Gaussian process regression methodology in the Saharan climate,” *Eur. Phys. J. Plus*, vol. 133, no. 6, pp. 1–17, 2018.

[21] J.-L. Chen, G.-S. Li, and S.-J. Wu, “Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration,” *Energy Convers. Manag.*, vol. 75, pp. 311–318, 2013.

[22] M. H. Imani, E. Bompard, P. Colella, and T. Huang, “Predictive methods of electricity price: an application to the Italian electricity market,” in *2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, 2020, pp. 1–6.

[23] N. S. Altman, “Kernel smoothing of data with correlated errors,” *J. Am. Stat. Assoc.*, vol. 85, no. 411, pp. 749–759, 1990.

[24] S. Demir, K. Mincev, K. Kok, and N. G. Paterakis, “Introducing technical indicators to electricity price forecasting: A feature engineering study for linear, ensemble, and deep machine learning models,” *Appl. Sci.*, vol. 10, no. 1, p. 255, 2020.

[25] A. Smola, “Support vector machines, regularization, optimization, and beyond,” *Learn. with Kernels*, 2002.

[26] W. Astuti, “Support vector machine and principal component analysis for microarray data classification,” in *Journal of Physics: Conference Series*, 2018, vol. 971, no. 1, p. 12003.

[27] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[28] C. E. Rasmussen and H. Nickisch, “Gaussian processes for machine learning (GPML) toolbox,” *J. Mach. Learn. Res.*, vol. 11, pp. 3011–3015, 2010.

[29] M. T. Farrell and A. Correa, “Gaussian process regression models for

predicting stock trends," *Relation*, vol. 10, p. 3414, 2007.

[30] B. Kamiński, M. Jakubczyk, and P. Szufel, "A framework for sensitivity analysis of decision trees," *Cent. Eur. J. Oper. Res.*, vol. 26, no. 1, pp. 135–159, 2018.

[31] C. D. Sutton, "Classification and regression trees, bagging, and boosting," *Handb. Stat.*, vol. 24, pp. 303–329, 2005.

[32] J. Lago, F. De Ridder, P. Vrancx, and B. De Schutter, "Forecasting day-ahead electricity prices in Europe: The importance of considering market integration," *Appl. Energy*, vol. 211, pp. 890–903, 2018.

[33] C. Hamzaçebi, D. Akay, and F. Kutay, "Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3839–3844, 2009.

[34] A. Rohani, M. Taki, and M. Abdollahpour, "A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)," *Renew. Energy*, vol. 115, pp. 411–422, 2018.

[35] "GME - Gestore dei Mercati Energetici SpA." https://www.mercatoelettrico.org.

[36] F. Ardian, "Empirical analysis of Italian electricity market," 2016.

[37] M. V. Selvi and S. Mishra, "Investigation of Performance of Electric Load Power Forecasting in Multiple Time Horizons With New Architecture Realized in Multivariate Linear Regression and Feed-Forward Neural Network Techniques," *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 5603–5612, 2020.

[38] "ENTSO-E." https://www.entsoe.eu/.

[39] A. A. P. Vamplew, "Encoding and decoding cyclic data," *S. Pac. J. Nat, Sci*, vol. 16, pp. 54–58, 1998.

[40] J. Stuart and P. Norvig, "Artificial Intelligence: a Modern Approach Prentice-Hall," 2010.

[41] S. Karsoliya, "Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture," *Int. J. Eng. Trends Technol.*, vol. 3, no. 6, pp. 714–717, 2012.

[42] J.-Y. Li, T. W. S. Chow, and Y.-L. Yu, "The estimation theory and optimization algorithm for the number of hidden units in the higher-order feedforward neural network," in *Proceedings of ICNN'95-International Conference on Neural Networks*, 1995, vol. 3, pp. 1229–1233.

[43] K. G. Sheela and S. N. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Math. Probl. Eng.*, vol. 2013, 2013.

[44] J. Ke and X. Liu, "Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction," in *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, 2008, vol. 2, pp. 828–832.

[45] B. Curry and P. H. Morgan, "Model selection in neural networks: some difficulties," *Eur. J. Oper. Res.*, vol. 170, no. 2, pp. 567–577, 2006.

[46] I. Chivers and J. Sleightholme, "An introduction to Algorithms and the Big O Notation," in *Introduction to programming with Fortran*, Springer, 2015, pp. 359–364.

[47] S. Wang, Z. Li, C. Liu, X. Zhang, and H. Zhang, "Training data reduction to speed up SVM training," *Appl. Intell.*, vol. 41, no. 2, pp. 405–420, 2014.

[48] H.-M. Lu, J.-S. Chen, and W.-C. Liao, "Nonparametric regression via variance-adjusted gradient boosting Gaussian process regression," *IEEE Trans. Knowl. Data Eng.*, 2019.

[49] H. M. Sani, C. Lei, and D. Neagu, "Computational complexity analysis of decision tree algorithms," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 2018, pp. 191–197.

[50] B.-M. Hsu, "Comparison of supervised classification models on textual data," *Mathematics*, vol. 8, no. 5, p. 851, 2020.

Mahmood Hosseini Imani (S'20) received the B.Sc. in electrical engineering from Babol Noshirvani University of Technology, formerly Institute of Mazandaran university, Iran, in 2011, the M.S. degree in electrical engineering from University of Guilan, Rasht, Iran, in 2013. He is currently pursuing the Ph.D. degree with the Department of Energy, Politecnico di Torino, Italy. His research interests include Data science, Machine Learning, Power system operation, Power Systems Economics, and Renewable Energy Integration.

Ettore Bompard (Member, IEEE) received the Ph.D. degree in electrical engineering from the Politecnico di Torino, Turin, Italy.

He has been a Visiting Assistant Professor with the University of Illinois, Urbana-Champaign (US), Champaign, IL, USA, from April to July 1999, January and October 2000) and the Senior Scientist with the Energy Security Unit of the JRC for Smart Electricity Systems and Interoperability, Institute for Energy and Transport, Power System and Critical Infrastructure in Petten, the Netherlands, from October 2012 to September 2014. He is a Professor of Power Systems with the Department of Energy of Politecnico di Torino and Scientific Director of the Energy Security Transition Lab @energy center, Torino, Italy. His research interests include electricity markets analysis and simulation, smart grids design and modelling, power system vulnerability assessment and security management, energy security, "science-based" support to policy decision-making, and data analytics applications to power systems.

Pietro Colella (S'15-M'16) received the M.Sc. and the Ph.D. degree in Electrical Engineering from Politecnico di Torino, Italy, in 2012 and 2016, respectively. He is currently an Assistant Professor at the same institute. His research interest includes power systems, electrical market, traction electrification systems and electrical safety. He is a Member of IEEE and AEIT - Associazione Italiana di Elettrotecnica, Elettronica, Automazione, Informatica e Telecomunicazioni. He is a Registered Professional Engineer in Italy.

Tao Huang (M'18) received his Ph.D. degree from Politecnico di Torino, Turin, Italy. He is currently a researcher with the Department of Energy, Politecnico di Torino, Italy. His research interests include vulnerability assessment, electricity markets, smart grids, and artificial intelligence in power systems.