Predictive methods of electricity price: An application to the Italian electricity market

Predictive methods of electricity price: an application to the Italian electricity market

Mahmood Hosseini Imani, Ettore Bompard, Pietro Colella, Tao Huang
Department of Energy
Politecnico di Torino
Torino, Italy
{mahmood.hosseiniimani, ettore.bompard, pietro.colella, tao huang }@polito.it

***Abstract*— Price forecasting is a crucial element for the members of the electricity markets and business decision-making to maximize their profits. The electricity prices have an impact on the behavior of market participants, and thus, predicting prices for generation companies, and consumers is essential for both the short-term profits in the Day-Ahead, Intra-Day and Ancillary markets, and the long-term benefits in the future planning, investment, and risk management. Therefore, participants in the electricity market need to accurately and effectively predict the price signal to manage market risk. In this paper, different forecasting models have been compared, and the most promising ones have been employed to forecast the short term Italian electricity market clearing price for achieving forecasting accuracy. In particular, simulations are performed for four principal regression methods, including Support Vector Machine, Gaussian Processes Regression, Regression Trees, and Multi-Layer Perceptron. The performance of predicted models is compared through several performance metrics, including MAE, RMSE, R, and the total number of percentage error anomalies. The results indicate the SVM is the best choice for forecasting the electricity market price on the Italian case study.**

Keywords*— *Electricity price prediction, Italian electricity market, PUN, Artificial neural network, Machine learning

I. INTRODUCTION

Following the restructuring of the electricity market in many countries, prices in a competitive market are influenced by market players. Price is determined by the intersection of the supply curve and demand curve.

Electricity market price has huge volatility, and this increases the risk for market players. Therefore, forecasting prices for generation companies (GENCOs) and consumers is essential.[1].

Due to the importance of the electricity market price prediction, several approaches have been proposed so far. The time series model can refer to "dynamic regression and transfer function models" [2], "autoregressive conditional heteroscedasticity" [3], and "conditional variance forecasts using a dynamic model" [4]. Although these methods are taken into consideration for the sake of simplicity in implementation and linearity, they are not efficient in the nonlinear system and dramatically increase the prediction error.

In the last years, several Machine Learning (ML) algorithms have been adopted to forecast the electricity price. Identifiyng patterns and trends easily, continuous improvement, and handling multi-variety data are the advantage of the ML algorithm.

One of the most widely used methods is the Artificial Neural Network (ANN). The most common types of artificial neural networks used in the price forecasting process are the Multi-Layer Perceptron (MLP) [5], Radial Basis Function (RBF) [6] and Self-Organized Maps (SOM) [7].

Another method that is commonly adopted in the field of electricity market prediction is the Support Vector Machines (SVM) [8]-[9]. When the algorithm has been adopted for regression, it is called Support Vector Regression (SVR) [9].

Moreover, in [10], tree-based techniques are applied for price prediction. Among the Tree-based methods, Random Forests (RF), Bagged Regression, and Boosted Trees are the most commonly used [10,11].

Gaussian Process Regression (GPR) is another supervised learning algorithm used for predicting market price and stock trends [12,13].

In this paper, twenty potential models based on the main machine learning algorithms described above (i.e., SVM including Linear, Quadratic, Coarse Gaussian and Cubic, GPR including Exponential, Squared Exponential, Rational Quadratic and Matern 52, MLP including four different structures, as well as Tree-based methods including bagged and boosted trees) are extensively investigated. Besides, three priority lists of diverse methods based on the performance metrics, MAE, R, total anomalies, are provided.

The rest of the paper is organized as follows: Section 2 proposed an overview of conducted methods for price prediction. An application of predictive methods of electricity price to the Italian electricity market is presented in Section 3. The paper is concluded in Section 4.

II. CONDUCTED METHODS FOR PRICE PREDICTION

In this section, the detailed information of used methodologies is described.

A. Support Vector Machine

SVM is categorized as a supervised learning method for the application of Regression and classification [14]. The principal object of SVM is determining hyperplanes that maximize the margin between classes.

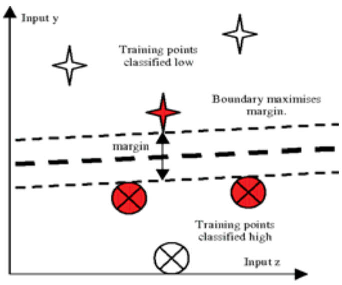

FIGURE 1: MAXIMUM MARGIN OF SUPPORT VECTOR MACHINE [9]

SVM is categorized based on the type of kernel [15,16].

- Linear
- Quadratic
- Cubic
- Gaussian (Fine, Medium and Coarse)

B. Gaussian process regression

GPR is a commonly used method in machine learning (ML) and is a joint Gaussian distribution over time. Gaussian Process is specialized with a mean function $m(x)$ and covariance function $k(x_1, x_2)$ [12]. To explain a real process $f(x)$ as a GPR, we have:

$$f(x) \sim Gp(m(x), (x_1, x_2)) \quad (1)$$

where,

$$m(x) = E[f(x)] \quad (2)$$

$$k(x_1, x_2) = E\left[\left(f(x_1) - m(x_1)\right)\left(f(x_2 - m(x_2))\right)\right] \quad (3)$$

In the regression model, considering a dataset D with N observations;

$D = \{(x_i, y_i | \; i = 1, \ldots, N\}$, with $x_i \in R^D$ and $y_i \in R$ 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.

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:

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

C. Tree-based Methods

Tree-based regression is nonparametric methods that can be applied to models having both a large number of observations and a large number of variables. There are three types of Tree-based methods. In this article, two of the most popular models, including Bagging, and Boosting, are investigated [17].

- Bagging Tree

Bagging or Bootstrap aggregation is a learning method for improving prediction by reducing the associated with forecasting. Bagging tree uses a simple averaging of results to achieve an overall forecast.

- Boosting Tree

Another method for improving the result of the prediction is the Boosting tree. Like bagging, boosting tree uses a committee-based approach and weighted average results to obtain the forecast. Moreover, in each step, each tree is grown based on the information related to previously grown trees [17].
In both bagging and boosting methods, the lowermost node on the tree is called Leaf or terminal node.

D. Multilayer perceptron

MLP is a feed-forward neural network consisting of the nodes can be organized as follows:

- Input layer
- one or more layers as a hidden layer
- Output layer

Output nodes and hidden nodes are neurons that use an activation function, for instance, Sigmoid, hyperbolic tangent, and so on [18]. The MLP model uses the Levenberg-Marquardt backpropagation (BP) algorithm for model training.

$$\Delta x = [J^T(x)J(x) + \mu I]^{-1} J^T(x) e(x) \quad (4)$$

where x is the parameter vector, $J(x)$ is the Jacobian matrix, and μ is the parameter [19].

The structure of the MLP is represented in Figure 2.

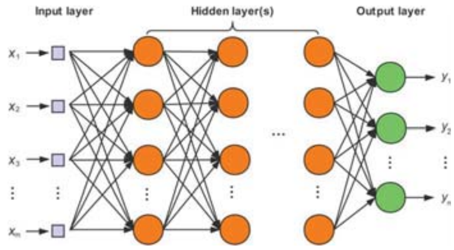

FIGURE 2: MLP STRUCTURE [17]

III. ITALIAN ELECTRICITY MARKET CASE STUDY

In this section, an introduction about the Italian power market is presented (subsection A). Detail information of the adopted data is provided in subsection B. A descriptive statistics on the Italian electricity market is provided in subsection C. The method for determining the train and test set of the data, brief information of adopted performance metrics, and a comprehensive discussion about results are provided in subsections D,E, and F, respectively.

A. Introducing the Italian Power Market

In the Italian Power Market, the geographical market includes seven foreign virtual zones, six geographical zones, fand five poles of limited production (national virtual zones).

In the Italian day-ahead market[1] (MGP), the submission from market participants takes place between the ninth day before the day of physical delivery (opens at 8 a m) and the day before the day of delivery(closes at 12 p.m) [20]. In this study, to predict the National Single Price [2] (PUN), four principal prediction methods, including SVM, GPR, Tree-based method, and MLP, are applied to the MGP.

B. Data Classification

The variables adopted to forecast PUN are listed in Table 1. Data are gathered on an hourly basis, except for Natural Gas (NG) price, for 2015, 2016, 2017, and 2018. All the data, including load consumption, electricity price, and NG price, are taken from Gestore dei Mercati Energetici (GME) [19]. NG price data are daily data where the data are copied to the corresponding hour of the day. Each year consists of 8,760

[1] Mercato del Giorno Prima

[2] Prezzo Unico Nazionale

hours (except 2016, which was a leap year, which consisted of one additional day making 8,784 hours), resulting in 35064 observations for the given period. These data are divided into a training set, which consists of 26,304 observations from Jan 1, 2015–Dec 31, 2017, and a test data set which consists of 8760 observations representing Jan 1, 2018–Dec 31, 2018.

In the current research, two dummy variables are considered for the type of hour. The hours from 08:00 to 20:00 are considered as peak hours, and the hours from 00:00 to 08:00 and from 20:00 to 24:00 are considered as off-peak hours [20]. Two dummy variables are also considered for day type, 1 for weekday, and 2 for the weekends and holidays.

TABLE 1: ALL THE VARIABLE FOR PUN PREDICTION

Variable	Units
Hour	1-24
Hour type	1(Peak), 2(Off-Peak)
Week Day	1-7
Day type	1(weekday) 2(weekend)
Season type	1-4
Current 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
Previous week same hour price	€/MWh
Previous 24 hours average price	€/MWh
Previous day NG price	€/MWh
Previous week average NG price	€/MWh

Due to the different behaviour of the load and price profile in each hour of a day, 24 dummy variables are considered for the hour.

C. *Descriptive Statistics on the Italian electricity market*

The raw data of hourly PUN from 2015 to 2018 are depicted in Figure 3.

Based on the box plots in Figure 4 and Table 2, after 2015, it can be clearly seen that there is an increasing trend of load, which leads to an increase in the PUN trend. The positive trend of the mean value of load and PUN confirms the previous claims.

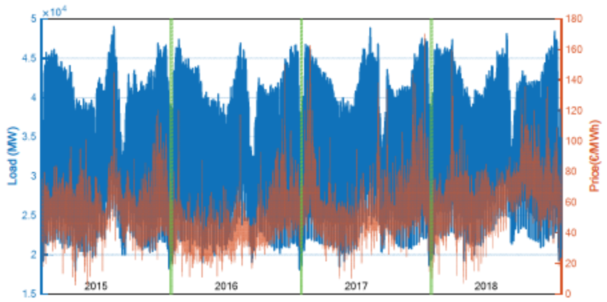

FIGURE 3: ITALIAN NATIONAL SINGLE PRICE AND LOAD, 2015 – 2018

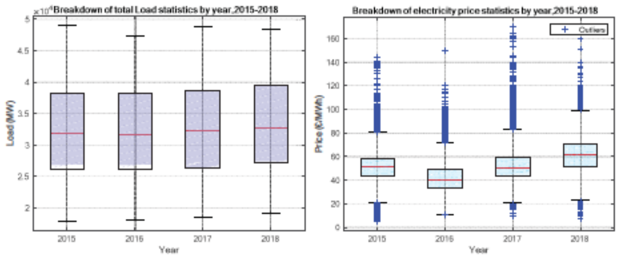

FIGURE 4: BREAKDOWN OF ANNUAL TOTAL LOAD AND PUN STATISTICS (2015 – 2018)

TABLE 2: DESCRIPTIVE STATISTICS OF ANNUAL LOAD AND PRICES (2015-2018)

Year		2015	2016	2017	2018
Load	Max	55157	49441	50333	52412
	Min	17268	16716	17212	17610
	Mean	32935.6	32247.51	33271.01	33934.27
	Std	7648.84	7287.92	7567.55	7554.62
Electricity price (PUN)	Max	144.57	150	170	159.4
	Min	5.62	10.95	10	6.97
	Mean	52.31	42.78	53.95	61.31
	Std	13.32	13.1	16.46	14.84

D. *Determining training and testing sets*

A training set is used to train the model, while a validation set is used to evaluate how that model performs on unseen data. The hold-out method dataset is divided into train and validation sets. In this method, X% of the train set is considered an actual train set, and the remaining (100-X) %, to be considered as the validation set. In this article, the dataset in the period of 2015 to 2017 is chosen as train and validation. In this regard, 25% of the data for the given period are considered unseen data.

E. *Performance Evaluation Metrics*

The common evaluation metrics for a regression problem are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and correlation coefficient (R). The metrics define as follow [21, 22]:

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|x_j - y_j| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - y_j)^2} \tag{6}$$

$$R = \frac{\sum_{j=1}^{n}(x_j-\bar{x})(y_j-\bar{y})}{\sqrt{\sum_{j=1}^{n}(x_j-\bar{x})^2}\sqrt{\sum_{j=1}^{n}(y_j-\bar{y})^2}} \tag{7}$$

Where $x_j$corresponds to the actual value, $y_j$is the forecasted value, $\bar{x}$, $\bar{y}$ are the average of the actual and predicted output, and n is the number of observations. The value of $|x_j - y_j|$ is considered as an error.

In this paper, for analyzing and comparing the performance of the forecasting methods, we defined an index named "anomaly index". In this regard, the recorded errors which upper than a specific value are considered as anomalies.The detail information about anomalies is depicted in Figure 5.

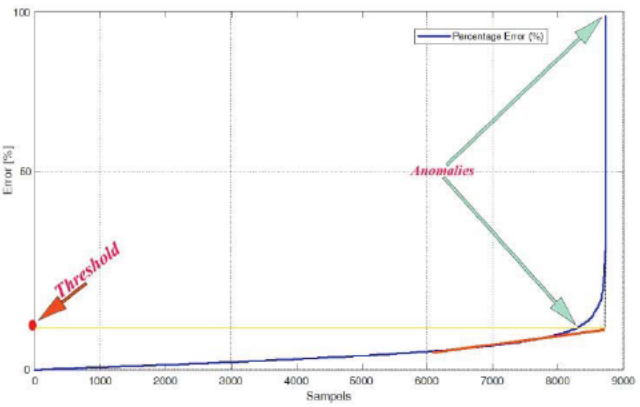

FIGURE 5: FINDING ANOMALIES

In this figure, the percentage error of each forecast is depicted in blue in ascending order. To identify anomalies, a threshold has been defined: first, a line (the red one in Figure 5) is fitted on the 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.

F. Discussion on the Results

The objective of this study is to compare the performance of SVM, GPR, MLP, and Tree-based methods for predicting short term electricity prices. In Figure 6 to Figure 9, the models are ordered from the best to the worst.

The strength of the mentioned methods is assessed using MAE, RMSE, R, and the total number of anomalies. The achieved statistical analysis for the PUN prediction using SVM methods is depicted in Figure 6. Based on this figure Linear, Quadratic ,and Coarse Gaussian SVM has the best accuracy, with respect to MAE, RMSE, and R statistics, while the Fine Gaussian SVM provides the best results for prediction price with fewer anomalies in percentage error.

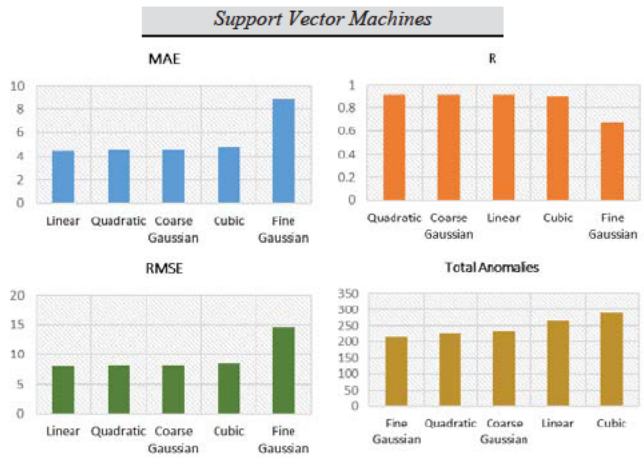

FIGURE 6: PERFORMANCE INDEXES FOR THE SVM METHOD

Figure 7 presents the performance indexes in PUN prediction by Bagged Tree Methods with considering different leaf sizes. According to MAE, R and RMSE, it is found that increasing the leaf sizes does not enhance the prediction performance but decreases it. Figure 8 displays the performance of GPR with different kernel classes. Based on all performance metrics, the results show that the Rational Quadratic kernel and Squared Exponential generally perform better than the other kernels.

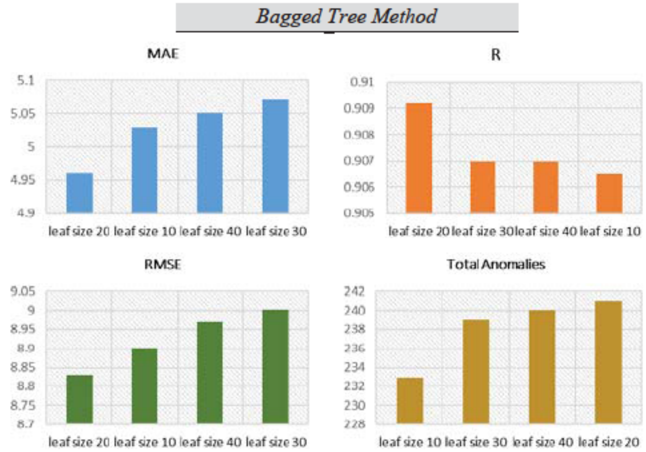

FIGURE 7: PERFORMANCE INDEXES FOR BAGGED TREES METHOD

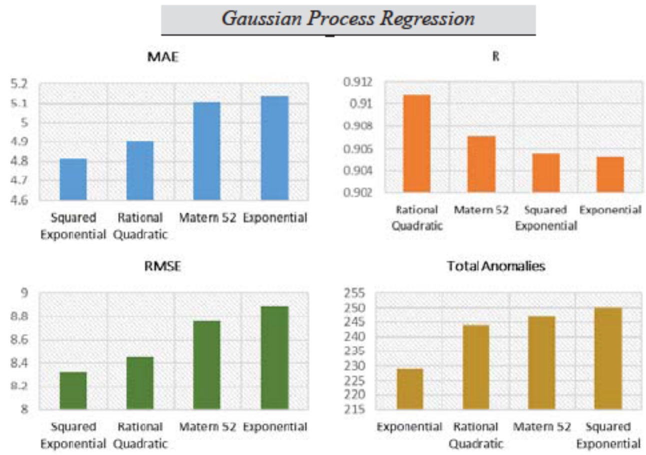

FIGURE 8: PERFORMANCE INDEXES FOR GPR METHOD

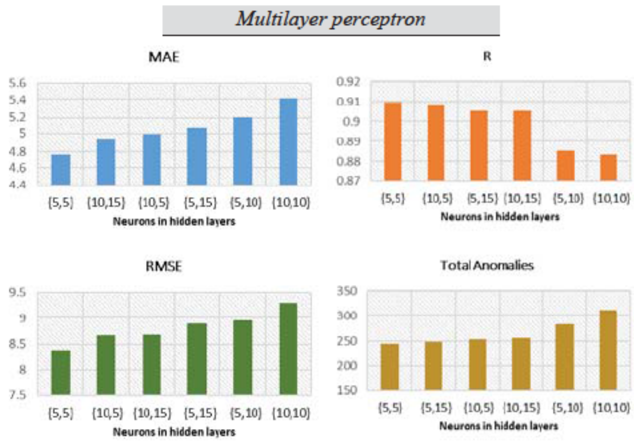

FIGURE 9: PERFORMANCE INDEXES OF MLP METHOD

In Figure 9, the different hidden node numbers versus performance metrics for the MLP model are depicted. The results illustrate that the lower the layer, the higher the efficiency. Results show that an unreasonable increase in the complexity of the model and the number of perceptron in the hidden layer do not increase the accuracy of the model.

For the SVM, GPR, MLP, and Bagged trees models, the MAE values range 4.487–8.801, 4.816–5.137, 4.764-5.41, 4.961-5.071 €/MWh, respectively. The total recorded

TABLE 3: RANKING OF DIFFERENT PREDICTION METHODS

	Model	MAE	Normalized MAE	Model	R	Model	Total Anomalies	Normalized Total Anomalies
1	SVM Linear	4.4873	1	SVM Quadratic	0.918	SVM Fine Gaussian	216	1
2	SVM Quadratic	4.5829	1.021	SVM Corse Gaussian	0.9179	SVM Quadratic	228	1.056
3	SVM Corse Gaussian	4.5875	1.022	SVM Linear	0.9173	GPR Exponential	229	1.060
4	MLP {5 5}	4.7647	1.062	GPR Rational Quadratic	0.9108	Bagged Regression Trees 10	233	1.079
5	SVM Cubic	4.8078	1.071	MLP {5 5}	0.9098	SVM Corse Gaussian	234	1.083
6	GPR Squared Exponential	4.8162	1.073	SVM Cubic	0.9095	Bagged Regression Trees 30	239	1.106
7	GPR Rational Quadratic	4.9065	1.093	Bagged Regression Trees 20	0.9092	Bagged Regression Trees 40	240	1.111
8	MLP {10 15}	4.9449	1.102	MLP {10 5}	0.9085	Bagged Regression Trees 20	241	1.116
9	Bagged Regression Trees 20	4.9611	1.106	GPR Matern 52	0.9071	GPR Rational Quadratic	244	1.130
10	MLP {10 5}	4.9891	1.112	Bagged Regression Trees 30	0.907	MLP {5 5}	246	1.139
11	Bagged Regression Trees 10	5.0299	1.121	Bagged Regression Trees 40	0.907	GPR Matern 52	247	1.144
12	Bagged Regression Trees 40	5.0508	1.126	Bagged Regression Trees 10	0.9065	MLP {5 15}	249	1.153
13	MLP {5 15}	5.0648	1.129	GPR Squared Exponential	0.9056	Boosted Tress 8	250	1.157
14	Bagged Regression Trees 30	5.0714	1.130	MLP {5 15}	0.9053	GPR Squared Exponential	250	1.157
15	GPR Matern 52	5.1068	1.138	GPR Exponential	0.9052	MLP {10 5}	254	1.176
16	GPR Exponential	5.1372	1.145	MLP {10 15}	0.9051	MLP {10 15}	256	1.185
17	MLP {5 10}	5.2025	1.159	Boosted Tress 8	0.9051	SVM Linear	265	1.227
18	Boosted Tress 8	5.3955	1.202	MLP {5 10}	0.8854	MLP {5 10}	284	1.315
19	MLP {10 10}	5.4107	1.206	MLP { 10 10}	0.8837	SVM Cubic	290	1.343
20	SVM Fine Gaussian	8.8012	1.961	SVM Fine Gaussian	0.6766	MLP { 10 10}	311	1.440

anomalies value is given in Table 3 range 216-290, 229-250, 246-311, 233-241 for the SVM, GPR, MLP, and Bagged trees models, respectively. SVM model produces the lowest anomaly error in two kernels (Fine Gaussian and Quadratic) and highest accuracy in three kernels (Linear, Quadratic ,and Corse Gaussian).

Comparisons of four principal prediction methods with regard to their different structures based on the various performance metrics are made in Table 3.

Therefore, Table 3 clearly indicates that ,in general, the SVM model performs superior to the MLP and Tree-based models in predicting electricity prices in the Italian market.

However, the difference between the accuracy of the prediction methods is not so huge.

IV. CONCLUSION

The applicability and accuracy of four relevant predictive models (SVM, GPR, MLP, and Tree-Based methods) in short-time electricity price prediction of the Italian electricity market were investigated. Data obtained from Gestore dei Mercati Energetici (GME), Italian Energy Markets Manager, were used in the applications. Various useful variables were tried as inputs to build the models, and the outcomes were compared through several performance metrics, including MAE, RMSE, R, and the total number of anomalies. The model results at above twenty models indicate that the SVM and GPR models generally provide better accuracies than MLP and Tree-Based models in forecasting PUN, even if the differences are not huge. Furthermore, it was found that an irrational increase in the complexity of the MLP model does not lead to increased accuracy. It is observed that the SVM Linear model produces the best accuracy concerning MAE values. Also, SVM Quadratic results best performance with respect to the R index.

REFERENCES

[1] Li, X. R., et al. "Day-ahead electricity price forecasting based on panel cointegration and particle filter." Electric Power Systems Research 95 (2013): 66-76.

[2] Nogales, Francisco Javier, et al. "Forecasting next-day electricity prices by time series models." IEEE Transactions on power systems 17.2 (2002): 342-348.

[3] Bowden, Nicholas, and James E. Payne. "Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models." Energy Economics 30.6 (2008): 3186-3197.

[4] Diongue, Abdou Kâ, Dominique Guegan, and Bertrand Vignal. "Forecasting electricity spot market prices with a k-factor GIGARCH process." Applied energy 86.4 (2009): 505-510.

[5] Keynia, Farshid. "A new feature selection algorithm and composite neural network for electricity price forecasting." Engineering Applications of Artificial Intelligence 25.8 (2012): 1687-1697.

[6] Akbilgic, Oguz, Hamparsum Bozdogan, and M. Erdal Balaban. "A novel Hybrid RBF Neural Networks model as a forecaster." Statistics and Computing 24.3 (2014): 365-375.

[7] Anbazhagan, S., and Narayanan Kumarappan. "Day-ahead deregulated electricity market price forecasting using recurrent neural network." IEEE Systems Journal 7.4 (2012): 866-872.

[8] Moguerza, Javier M., and Alberto Muñoz. "Support vector machines with applications." Statistical Science 21.3 (2006): 322-336.

[9] Sansom, Damien C., Tom Downs, and Tapan K. Saha. "Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants." Journal of Electrical & Electronics Engineering, Australia 22.3 (2003): 227.

[10] González, Camino, José Mira-McWilliams, and Isabel Juárez. "Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests." IET Generation, Transmission & Distribution 9.11 (2015): 1120-1128.

[11] Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

[12] Mojaddady, Mohammad, Moin Nabi, and Shahram Khadivi. "Stock market prediction using twin Gaussian process regression." International Journal for Advances in Computer Research (JACR) preprint (2011).

[13] Farrell, M. Todd, and Andrew Correa. "Gaussian process regression models for predicting stock trends." Relation 10 (2007): 3414.

[14] Gao, Ciwei, et al. "Bidding strategy with forecast technology based on support vector machine in the electricity market." Physica A: Statistical Mechanics and its Applications 387.15 (2008): 3874-3881.

[15] Astuti, Widi. "Support vector machine and principal component analysis for microarray data classification." Journal of Physics: Conference Series. Vol. 971. No. 1. IOP Publishing, 2018.

[16] Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

[17] Sutton, Clifton D. "Classification and regression trees, bagging, and boosting." Handbook of statistics 24 (2005): 303-329.

[18] Lago, Jesus, et al. "Forecasting day-ahead electricity prices in Europe: the importance of considering market integration." Applied energy 211 (2018): 890-903.

[19] Fan, Junliang, et al. "Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China." Renewable and Sustainable Energy Reviews 100 (2019): 186-212.

[20] https://www.mercatoelettrico.org

[21] Rohani, Abbas, Morteza Taki, and Masoumeh Abdollahpour. "A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)." Renewable Energy 115 (2018): 411-422.

[22] Guermoui, Mawloud, et al. "Estimation of the daily global solar radiation based on the Gaussian process regression methodology in the Saharan climate." The European Physical Journal Plus 133.6 (2018): 211.