A Novel Hybrid Feature Selection Method for Day-Ahead Electricity Price Forecasting

Ankit Kumar Srivastava [1,*]**, Ajay Shekhar Pandey** [2]**, Rajvikram Madurai Elavarasan** [3]**,**
Umashankar Subramaniam [4]**, Saad Mekhilef** [5] **and Lucian Mihet-Popa** [6,*]

1   Department of Electrical Engineering, Dr. Rammanohar Lohia Avdh University, Ayodhya 224001, India
2   Department of Electrical Engineering, Kamla Nehru Institute of Technology, Sultanpur 228118, India;
    ajayshekharpandey@gmail.com
3   Department of Electrical and Electronics Engineering, Thiagarajar College of Engineering,
    Madurai 625015, India; rajvikram787@gmail.com
4   Renewable Energy Laboratory, Department of Communications and Networks, College of Engineering,
    Prince Sultan University, Riyadh 11586, Saudi Arabia; usubramaniam@psu.edu.sa
5   School of Science, Computing and Engineering Technologies, Swinburne University of Technology,
    Hawthorn, VIC 3122, Australia; saad@um.edu.my
6   Faculty of Electrical Engineering, Ostfold University College, 1757 Halden, Norway
*   Correspondence: aankit.srivastava32@gmail.com (A.K.S.); lucian.mihet@hiof.no (L.M.-P.)

Abstract: The paper proposes a novel hybrid feature selection (FS) method for day-ahead electricity price forecasting. The work presents a novel hybrid FS algorithm for obtaining optimal feature set to gain optimal forecast accuracy. The performance of the proposed forecaster is compared with forecasters based on classification tree and regression tree. A hybrid FS method based on the elitist genetic algorithm (GA) and a tree-based method is applied for FS. Making use of selected features, aperformance test of the forecaster was carried out to establish the usefulness of the proposed approach. By way of analyzing and forecasts for day-ahead electricity prices in the Australian electricity markets, the proposed approach is evaluated and it has been established that, with the selected feature, the proposed forecaster consistently outperforms the forecaster with a larger feature set. The proposed method is simulated in MATLAB and WEKA software.

Keywords: price forecasting; feature selection; elitist genetic algorithm; SMO regression; confidence interval

1. Introduction

Efficient and consistent electricity price forecasting is vital for market participants in the preparation of appropriate risk management plans in an electricity market. The higher the complexity in the market, the greater the peril of producing an error in forecast. An appropriate forecasting method allows suppliers and buyers to detail the bidding strategies for moderate losses and to increase profit. The research for developing price forecasting tools in non-intervention markets is at an intermediary stage and a variety of forecasting models covering many free trade markets have emerged in recent years [1–6]. Since a price chain is extremely volatile with non-constant mean and variance, because of the mobile nature and inflexible condition of establishing real-time stability of demand and supply of electricity, short-term price prediction is a difficult task. Market clearing prices (MCP) are volatile in a deregulated power market because of being an auction market price forecasting is an important tool for such markets. The companies that do business in electricity markets make broad use of price forecasting methods either to propose or to evade against volatility while auctioning in a pool system. Market contenders are requested to communicate the bids in terms of quantities and prices. A corporation can regulate its own production/price schedule on its own cast and hourly pool prices. A good quality MCP forecast and its confidence interval evaluation can help utilities and autonomous

power producers to present efficient bids. Electricity price forecasting has been focussed of several researchers in the field of electricity and various authors and researchers have published on this topic.

Time series, artificial neural networks (ANNs), regression, semi-parametric and non-parametric methods [7–14] are well known approaches used for electricity price forecasting. W K. Hubicka et al. [15] suggested a novel idea in vitality forecasting and have shown that averaging next day electricity price forecasts of a prescient model across 28- to 728-day alignment windows yields preferred outcomes over choosing only one 'ideal' window length. A. Y. Alanis et al. [16] presented, a recurrent neural network for forecasting electricity price based on the extended Kalman filter with both cases one step ahead and n-step ahead and also include the stability proof using the Lyapunov methodology.

L. Wang et al. [17] represented a model for day ahead hourly price forecasting using stacked denoising autoencoders (SDA) and deep neural networks. The data used are from the market of Arkansas, Texas, Louisiana, Nebraska, and Indiana hub of US. Authors in [17] used forecasting of two types: day ahead hourly forecasting and the online forecasting. The results show that the SDA model is competent in forecasting the electricity prices accurately. H. Mosbah et al. [18] used a multilayer neural network in composite topologies for accurate forecasting using data of 2005 as a training set while predicting prices of January 2006 in the Australian market. MAPE and MSE are calculated to compare the best performance indices using diverse composite topologies.

P. Sarikprueck et al. [19] proposed a hybrid method for both spike and non-spike power market prices using clustering methods of three types including CART, stratification and K-means methods and validated the performance of the proposed model with the Electric Reliability Commission of Texas wholesale market price. The outcomes given in the suggested technique show a vital up-gradation over other usual methodologies. J. P. González et al. [20] introduced a recently developed forecasting technique which strives to generalize the ARMAX time series method to the L2Hilbert space. The proposed model consists of linear regression where useful parameters work on functional variables. The variables used are either historical values, or exogenous variables, or past observed innovations. German and Spanish electricity markets are considered for forecasting in the work.

Anamika et al. [21] presented, a combination of various methodology like feed forward neural networks (FFNN), wavelet transform (WT), fuzzy adaptive particle swarm optimization (FA-PSO) for price forecasting on Spanish electricity markets for the year 2002. K. Wang et al. [22] suggested a novel approach for electricity price forecasting which consists of three models, first they combined the relief-F algorithm and random forest (RF) and proposed a hybrid FS method based on GCA to remove the feature redundancy. Secondly, a combination of KPCA was utilized in feature extraction procedure to understand the dimensionality reduction. Finally, a support vector machine (SVM) classifier based on differential evolution (DE) was used to predict electricity price.

P. Li et al. [23] addressed the dynamic similar sub-series technique for the consideration of cumulative effect which is due to collection of effecting factors during several consecutive time intervals and which makes relatively large forecasting errors. R. Tahmasebifar et al. [24] considered the probabilistic interval variables which can affect future electricity price forecasting. The method consists of two steps, in the initial step, a new hybrid method is introduced to predict point forecasts which is a combination of ELM, WT, FS based on mutual information (MI), and bootstrap approaches and in second step, PSO algorithm is used for improving the forecast accuracy and evaluating the variance of the model.

O. Abedinia et al. [25] suggested an FS approach based on criteria of informationtheory with important contribution of modeling interaction in combination with redundancy and relevancy. The hybrid filter-wrapper approach is another fundamental outcome of this work. C. González et al. [26] proposed a method to efficiently find the important variables that dominate the forecasting of electricity price forecasting. CART, RF and bagging

approaches were used for the accurate forecasting of electricity price in Iberian market. Historical data of 22 dominating factor such as load, price, hydro wind energy production, thermal generation hours etc., are used to efficiently forecast electricity price.

In this work electricity price forecasting is done using sequential minimization optimization (SMO) regression. The SMO regression algorithm followed by regression analysis is a powerful tool of decomposition for training SVM without the requirement of a quadratic programming (QP) solver. It is probable that SMO is the only SVM optimizer that exclusively exploits the quadratic form of the objective function and simultaneously uses the analytic solution of the size two cases. Since most of the objective functions have minimization criteria characteristics with set of constraints, for optimization purposes, such objective functions are represented in the form of Lagrange's multipliers. These Lagrange multipliers contain a dual form of the primal set of the predefined objective functions minus linear constraints. The only problem with SMO is its convergence accuracy for non-sparse data sets, but a few modifications have been suggested by the researchers to cope with such limitations. The novel contributions made in this work are:

1. A novel hybrid method based on elitist GA and tree based method for input FS in price prediction,
2. SMO regression base SVM is used for prediction of price and result obtained is compared with classification tree (J48) [27] and regression tree [28] (bagging and M5P) with and without feature selection,
3. Fixing the error margins during price prediction by applying the confidence interval, and
4. Season-wise optimize FS to have a better forecasting accuracy.

The paper is organized as follows. The proposed methodology for price forecasting is summarized in Section 2. Section 3 of the paper explains the SMO regression. The methodology adopted for input FS using a novel hybrid method based on elitist GA and tree-based method is explained briefly in Section 4. Section 5 shows the outcomes and performance of the proposed methodology. Findings and concluding remarks are provided in Section 6.

2. Proposed Methodology

Forecasting of day-ahead prices with FS and without feature selection (WoFS) has been considered in this work. The present work emphasizes on an elitist GA and tree-based FS method for price forecasting. Forecast of electricity price on half hourly basis and for each day and all season, week-wise is considered. A comparison of forecasting accuracy has been made by using full feature set and that of reduce feature set, which validates the usefulness of the forecasting with a reduced feature set. The proposed methodology for comparison of different methods is shown in Figure 1. Figure 1 shows that electricity price forecasting is performed using SMO regression employing a full feature set and in parallel the same SMO regression is used with selected feature set. These two forecasting modes were compared to ascertain the superiority of FS. The result obtained from SMO Regression, WoFS and with FS methods is also compared with classification tree and regression tree.
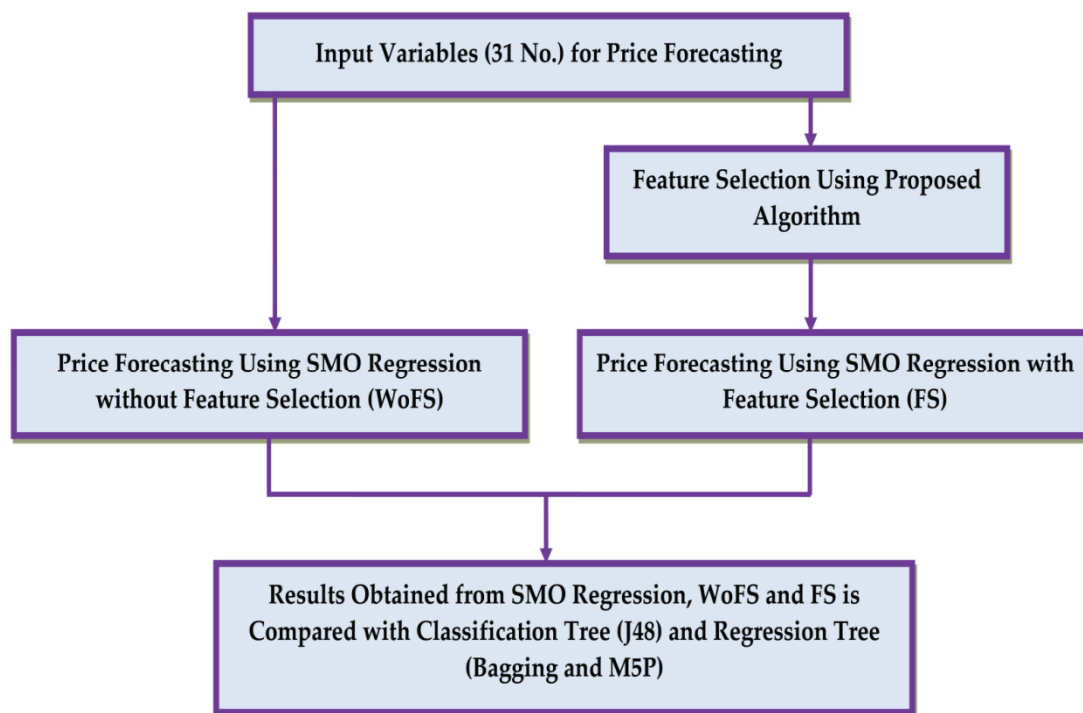
Figure 1. Proposed methodology.

3. Sequential Minimal Optimization (SMO) Regression Algorithm

SMO regression was proposed by J. Platt (1998) as the SVM classifier design for the training of SVM (Vladimir Vapnik, 1979) by using the LIBSVM tool. SMO is used to optimize solutions of large QP without using extra matrix storage and QP optimization steps. It is probable that SMO is the only SVM optimizer that exclusively exploits the quadratic form of the objective function and it breaks large QP into sets of the smallest possible QP by using Osuna's theorem. SMO consist two components one is analytic method and another is heuristic process. During the computation, at every step SMO uses only two Lagrange's multiplier and solves it by using analytic method. Then by heuristic process it selects the best result to update the system to get new optimal value. This selection method needs very short and simple C coding. Therefore, it fastens the speed and omits the requirement of numerical QP optimization which requires an entire QP library routine or complex matrix algorithms.

3.1. Methodology

The SMO algorithm basically works with two steps and until convergence, the algorithm keeps repeating these steps in each iteration:

Step 1: Breaking of large QP problems into series of smallest possible QP problem. Find the most promising pair (μ_1 and μ_2).

Step 2: It solves small QP problems in a very fast manner compared to the QP optimization process because it consumes more time due to inner loops. It is important point to consider that it requires memory in proportion to the smallest possible samples taken under step 1. This enables it to handle a lot of training sets i.e., very large QP problems. Optimize μ_1 and μ_2 keeping other μ's fixed.

A general quadratic problem has a quadratic objective function with linear constraints. A basic form is expressed by Equation (1).

$$\underset{A}{minimize}\,\frac{1}{2} * A_j^M * B * A_j + d^M * A_j; \ subject\ to\ 0 \leq A_j \leq C, j = 1, 2, \ldots, n. \quad (1)$$

where, $A = \begin{vmatrix} a_1 \\ a_2 \end{vmatrix}$ is a vector of decision variable, $C = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix}$ contain the constant coefficients that would multiply the squared term and the a1 times a2 terms. Therefore, C needs to be multiplied two times by the column vector A. the linear term of model is defined by $d = \begin{vmatrix} d_1 \\ d_2 \end{vmatrix}$ the coefficients are d_1 and d_2 in the vector. 0 defines the lower bound C is higher bound or upper bound in the constraint.

3.2. Calculation of SMO Regression

See Appendix A.

4. Input Feature Selection Using Proposed Algorithm

GA is based on the Darwinian theory of natural evolution. It is basically a heuristic search technique. In this approach no assumptions are made for relationship among features, while searching the space for feature selection. The genetic algorithm, when selecting any feature as per their importance to define the given data, decides by giving a sequence of Boolean values, allowing exploration of the feature space. It retains the features that benefit the classification task. By doing this it simultaneously avoids any local optima due to their intrinsic randomness. By using operators inspired by natural evolution such as selection, mutation and crossover, GA finds the solutions to optimize the problems. For any given dataset, FS is used to select dominating features among the pool of various feature so that handled data would be less and efficient. It is performed using an elitist technique of GA and tree based method.

In this method we select the dominant features of the input data set, which affects the process of forecasting on a priority basis. The 20% elite data from the large dataset to be presented in next generation has been selected, so that the next population is with features having classification accuracy not less than the previous generation. For the purpose of FS string of '0' and '1' are used in the manner of chromosome segments. In this chromosome '0' shows that the specific feature corresponding to that index is not chosen and '1' shows that the specific features is chosen. Here, the length of the chromosome is same as the number of features in given dataset. Computation is undertaken using data mining workbench WEKA. The fitness function is derived through tree and will be the stratified with 10-fold cross validation (10-FCV) classification accuracy. Consequently, we apply genetic operators, iteration by iteration until we met the stopping criteria. The set of selected features is the obtained chromosomes with optimal fitness function.

The mathematical definition of the fitness function for the proposed approach for feature selection is given below:

Fitness function = Classification accuracy

$$\text{Classification accuracy} = \left(\frac{\text{No.} of \text{ instancescorrectlyclassified}}{\text{Totalno. ofinstanceinthedataset}} \right) * 100 \tag{2}$$

In this work, single site crossover with 0:5 probabilities is performed in every step through roulette wheel selection. The mutation is performed with probability of 0:005. Further, we include the elite chromosomes by keeping 20% elite chromosomes for the next generation. In this way it can be seen that the resulting population has the best chromosomes which have optimal or, best classification accuracy of the given data set. The flowchart of elitist GA and the tree-based method for FS is shown in Figure 2.
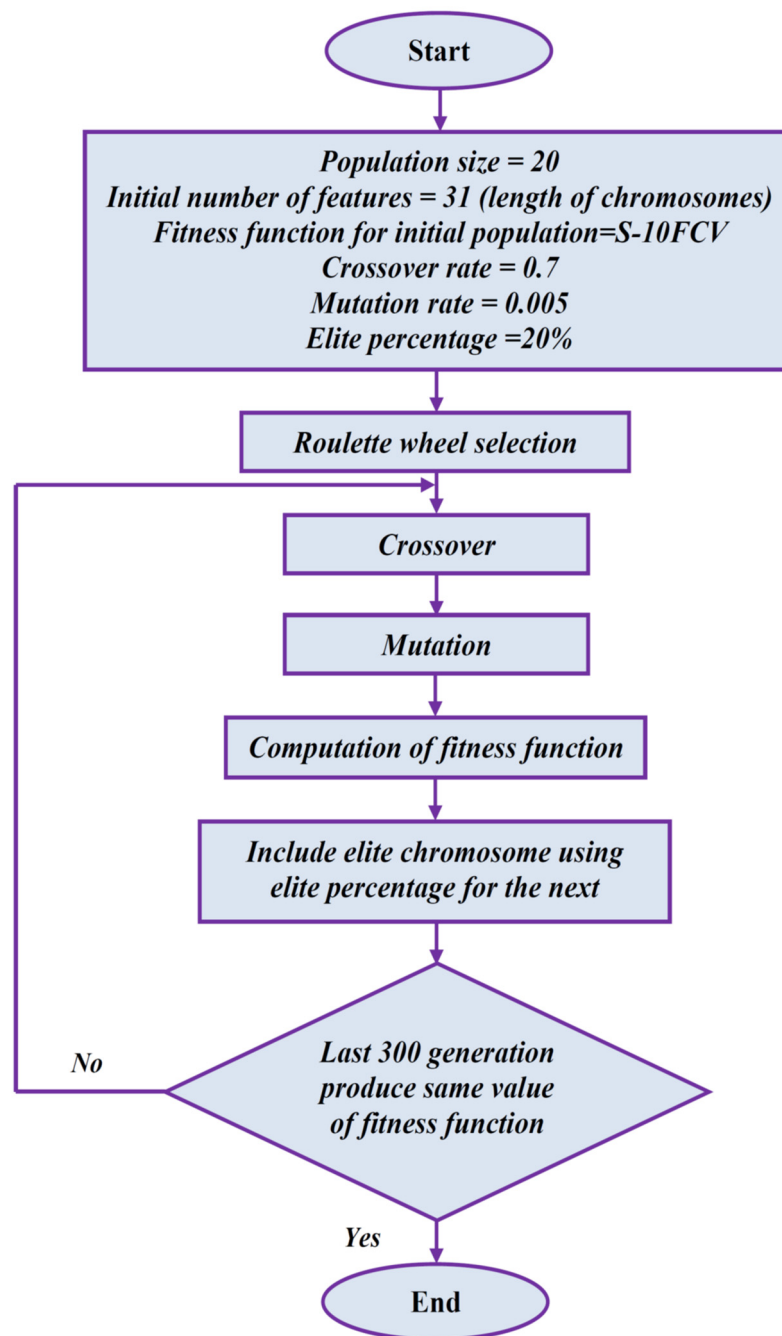
Figure 2. Flow chart of feature selection using proposed algorithm.

5. Result and Discussion

The half hourly historical load and price data of New South Wales, Australia (taken from Australian Energy Market Operator (AEMO)), and weather data of Sydney City (www.weatherzone.com/au, accessed on 18 August 2019) has been taken from January 2014 to June 2016 for the forecasting. The elitist GA algorithm is performed in MATLAB software, whereas the formation of optimal regression tree was performed in WEKA software. The WEKA software was interfaced with MATLAB for performing all regression tree calculations. The final forecast was made using MATLAB Software. Electricity price, load, wind speed, temperature, and humidity are considered as input variable in the present study. Table 1 shows the list of input variables that may affect day ahead price forecasting. Table 1 shows the input feature set and the time delay relative to forecast hour. The input feature set is taken on the basis of various literatures.

Table 1. List of input variables which affect electricity price forecasting.

Name of Variable	Name of Input	Timing of Variable
Load (Lo)	Lo_6	$Lo_{(D-23:00)}$
	Lo_5	$Lo_{(D-23:30)}$
	Lo_4	$Lo_{(D-24:00)}$
	Lo_3	$Lo_{(D-01:30)}$
	Lo_2	$Lo_{(D-01:00)}$
	Lo_1	$Lo_{(D-00.30)}$
Price (Pr)	Pr_6	$Pr_{(D-23:00)}$
	Pr_5	$Pr_{(D-23:30)}$
	Pr_4	$Pr_{(D-24:00)}$
	Pr_3	$Pr_{(D-01:30)}$
	Pr_2	$Pr_{(D-01:00)}$
	Pr_1	$Pr_{(D-00.30)}$
Wind Speed (Wi)	Wi_6	$Wi_{(D-23:00)}$
	Wi_5	$Wi_{(D-23:30)}$
	Wi_4	$Wi_{(D-24:00)}$
	Wi_3	$Wi_{(D-01:30)}$
	Wi_2	$Wi_{(D-01:00)}$
	Wi_1	$Wi_{(D-00.30)}$
Temperature (Te)	Te_6	$Te_{(D-23:00)}$
	Te_5	$Te_{(D-23:30)}$
	Te_4	$Te_{(D-24:00)}$
	Te_3	$Te_{(D-01:30)}$
	Te_2	$Te_{(D-01:00)}$
	Te_1	$Te_{(D-00.30)}$
Humidity (Hu)	Hu_6	$Hu_{(D-23:00)}$
	Hu_5	$Hu_{(D-23:30)}$
	Hu_4	$Hu_{(D-24:00)}$
	Hu_3	$Hu_{(D-01:30)}$
	Hu_2	$Hu_{(D-01:00)}$
	Hu_1	$Hu_{(D-00.30)}$
Day Timing (Hto)	Ht_o	$Ht_{o(D-00.00)}$

Each data set consists of 31 input features. 2016 data sets are used in the one training set to predict the electricity price. The results obtained from the proposed work are explained in two sections. The importance of FS is discussed in first section and forecast accuracy in second section. The input variables for FS of electricity price forecasting are taken from Table 1 as input feature set. FS is performed in a weekly manner. 2016 data sets are used in the one training set to forecast the electricity price and was considered on the concept of similar weeks. Every data set consists of five similar weeks of the months of the previous year and the preceding week of the same year, e.g., if the input FS for price forecasting is to be done for the week of 22–28 August 2015, the training set would consist of the data corresponding to 15–21 August 2015, 22–28 August 2014, 15–21 August 2014, 8–14 August 2014, 29 August–04 September 2014, 05–11 September 2014. For obtaining the FS the classification accuracy of data for the classifier is calculated using 10-FCV. It means at least once complete data is tested by this method. In this way FS is the only method that can be used for conducting feature analysis. By the present studies a detailed feature analysis can be undertaken.

From Table 2 it is clear that the input variable price of present-day (Pr_1) was selected in all the runs i.e., 36 times. At the same time, the price of the previous day (Pr_4) was selected 22 times and the hour type (Ht_o) 24 times, which shows their relative importance accordingly. The load of immediate hour (Lo_1) and previous day load (Lo_4) was an important feature and selected 20 and 23 times respectively. The temperatures of the present day (Te_1) selected more times than on the previous day. Wind speed of the previous day (Wi_6) was selected more number of times than the same day. The humidity of the previous day (Hu_6)

was selected more times than on the same day. Effects of features can also be analyzed according to seasons. Table 2 also indicates the top 10 features selected most times.

Table 2. Number of feature selected year wise highlighted on the top 10 features selected most times.

Feature Name	Number of Time Selected	Feature Name	Number of Time Selected
Lo_6	16	Wi_2	20
Lo_5	10	Wi_1	17
Lo_4	23	Te_6	14
Lo_3	16	Te_5	19
Lo_2	18	Te_4	19
Lo_1	20	Te_3	11
Pr_6	15	Te_2	16
Pr_5	18	Te_1	21
Pr_4	22	Hu_6	23
Pr_3	17	Hu_5	20
Pr_2	18	Hu_4	15
Pr_1	36	Hu_3	10
Wi_6	21	Hu_2	18
Wi_5	13	Hu_1	18
Wi_4	18	Ht_o	24
Wi_3	15		

Table 3 indicates the season-wise importance of different features. From the Table 3 it is observed that Lo_6 and Lo_1 etc. are features which are assumed to have more importance during the winter season. Hu_6 and Pr_5 etc. are features which are assumed to have more importance during the spring season. Lo_4 and Te_5 etc. are features which are assumed to have more importance during the summer season. Pr_1 is seem to be feature regardless of the season. These analysis indicated the relative importance of feature in terms of seasonal variations.

In this work, numerous error measures like root-mean square error (*RMSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*) and error variance (EV) are used for numerical accuracy assessment of the price forecasting as follows:

$$MAE = \frac{1}{K}\sum_{p=1}^{K}\left|(A_P - F_p)\right| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{K}\sum_{p=1}^{K}(A_p - F_p)^2} \tag{4}$$

$$MAPE = \frac{100}{K}\sum_{p=1}^{K}\frac{\left|A_p - F_p\right|}{A_p} \tag{5}$$

$$\sigma_{e,\,week}^2 = \frac{1}{K}\sum_{p=1}^{K}\left(\frac{\left|A_p - F_p\right|}{A_{week}} - (e_{week})\right)^2 \tag{6}$$

$$e_{week} = \frac{1}{K}\sum_{p=1}^{K}\frac{\left|A_p - F_p\right|}{A_{week}} \tag{7}$$

where A_p and F_p are actual and forecasted value of electricity price at time p and K is the length of forecast horizon and $A_{week} = \frac{1}{K}\sum_{p=1}^{K}A_p$.

Table 3. Number of times a feature selected season wise.

Feature Name	Winter	Spring	Summer
Lo_6	8	4	4
Lo_5	3	4	3
Lo_4	5	8	10
Lo_3	4	6	6
Lo_2	3	7	8
Lo_1	9	5	6
Pr_6	3	6	6
Pr_5	4	8	6
Pr_4	9	7	6
Pr_3	7	5	5
Pr_2	5	7	6
Pr_1	12	12	12
Wi_6	7	7	7
Wi_5	4	5	4
Wi_4	7	6	5
Wi_3	6	4	5
Wi_2	5	7	8
Wi_1	5	7	5
Te_6	3	7	4
Te_5	6	5	8
Te_4	6	6	7
Te_3	4	2	5
Te_2	6	7	3
Te_1	6	7	8
Hu_6	6	10	7
Hu_5	6	7	6
Hu_4	6	3	6
Hu_3	4	3	3
Hu_2	6	5	7
Hu_1	7	6	5
Ht_0	7	9	8

The electricity price forecasting is undertaken using SMO regression with the FS method for the New South Wales (NSW) electricity market. The average errors of each method for all seasons are calculated week-wise. Table 4 shows comparison between the SMO regression approach and seven other approaches (SMO regression, M5P, M5P +FS, Bagging, Bagging +FS, J48, J48+FS), in terms of different error measures in terms of MAPE, RMSE, MAE and EV. It also summarizes the overall mean performance for each method, in the last column. Results show that the SMO regression + FS methods perform well over all other methods used for the comparison. The error of four prior weeks are evaluated and arranged at the regular interval of half an hour from 00:00, 00:30, 01:00,...., up to 23:30 to evaluate the confidence interval for a day. Then, hourly standard deviation (δ) and 2δ are calculated for 95% confidence interval. The upper limit and lower limit calculation is as given below:

$$Upper\ Limit = Forecasted\ value + 2\delta \tag{8}$$

$$Lower\ Limit = Forecasted\ value - 2\delta \tag{9}$$

Table 4. Comparison of forecast accuracy using different error measures.

| Sr. No. | Error Measures | Methods | Season | | | Average |
			Winter (22–28 August 2015)	Spring (22–28 October 2015)	Summer (22–28 January 2016)	
1	Mean Absolute Percentage Error (MAPE)	J48	12.45	9.30	10.42	10.72
		J48+ FS	11.40	8.41	8.86	9.56
		Bagging	7.86	7.67	8.61	8.05
		Bagging+ FS	7.98	7.31	8.43	7.91
		M5P	6.40	6.12	7.68	6.73
		M5P+ FS	6.05	5.77	6.47	6.10
		SMO regression	5.25	5.73	5.35	5.44
		SMO regression+FS	4.85	5.26	5.28	5.13
2	Mean Absolute Error (MAE)	J48	4.09	3.34	3.88	3.77
		J48+ FS	3.74	3.07	3.29	3.37
		Bagging	2.62	2.76	3.24	2.87
		Bagging+ FS	2.68	2.57	3.19	2.81
		M5P	2.14	2.19	3.65	2.66
		M5P+ FS	2.03	2.07	2.74	2.28
		SMO regression	1.75	1.99	2.11	1.95
		SMO regression+ FS	1.60	1.82	2.07	1.83
3	Root Mean Square Error (RMSE)	J48	5.37	4.35	6.91	5.55
		J48+FS	4.96	4.05	5.20	4.74
		Bagging	3.47	3.67	5.10	4.08
		Bagging+FS	3.55	3.54	5.12	4.07
		M5P	3.56	2.96	13.06	6.53
		M5P+FS	3.04	2.86	6.83	4.24
		SMO regression	2.62	2.75	4.07	3.15
		SMO regression+FS	2.39	2.49	4.00	2.96
4	Error Variance (EV)	J48	0.0102	0.0062	0.0271	0.0145
		J48+FS	0.009	0.0055	0.0134	0.0093
		Bagging	0.0044	0.0046	0.0129	0.0073
		Bagging+FS	0.0046	0.0047	0.0133	0.0075
		M5P	0.0068	0.0032	0.1307	0.0469
		M5P+FS	0.0043	0.0031	0.0325	0.0133
		SMO regression	0.0032	0.0029	0.01	0.0054
		SMO regression+FS	0.0027	0.0023	0.0097	0.0049

The results of proposed model for winter, spring summer of NSW electricity market is depicted in Figures 3–5 respectively. From the Table 4 it is clear that the proposed method (SMO regression+FS) outperforms other methods in all seasons and for all error measures. Table 5 shows percentage improvement achieved by SMO+FS over the considered approaches. It is observed that the SMO Reg + FS gives 52.16% improvement over the J48 method. It can be seen that the SMO Reg + FS has improved forecasting accuracy over all the methods considered for comparison.
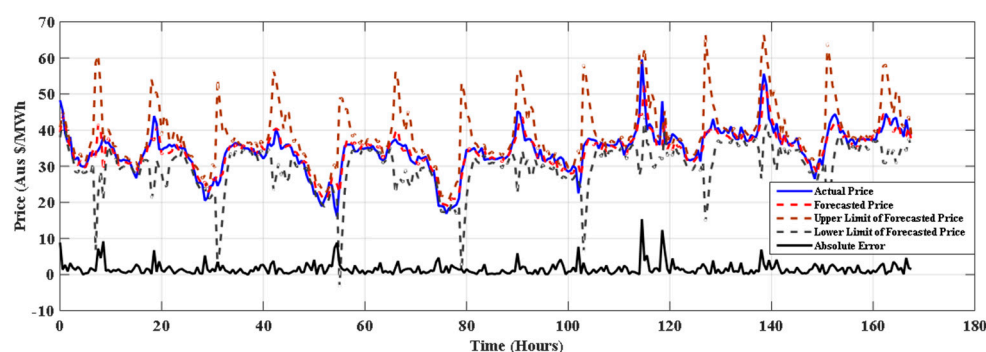
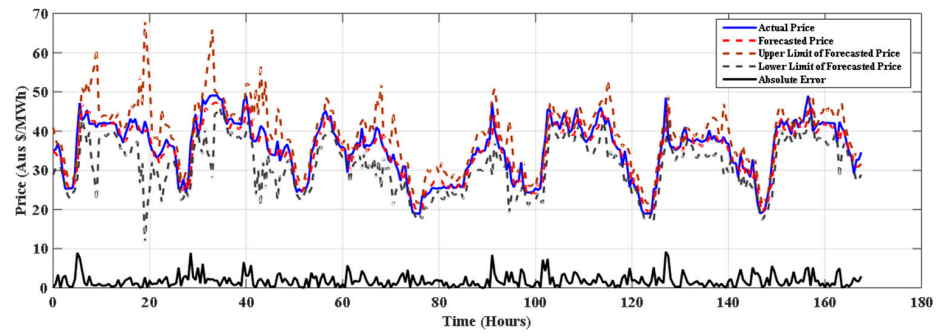**Figure 3.** Forecasting for 22–28 August 2015 in winter with SMO regression+FS.

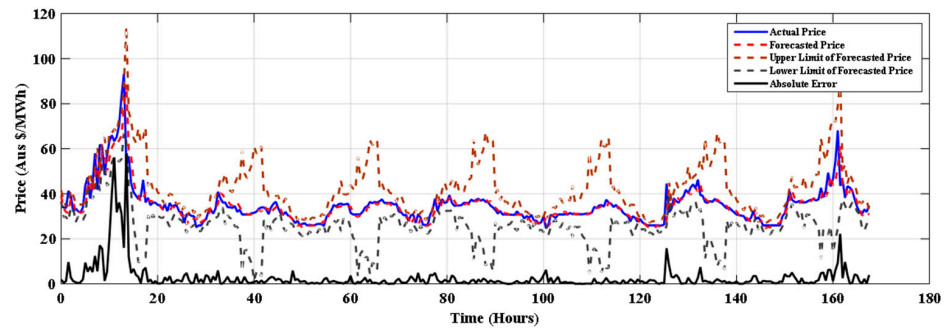**Figure 4.** Forecasting for 22–28 October 2015 in spring with SMO regression+ FS.

Figure 5. Forecasting for 22–28 January 2016 in summer with SMO regression+ FS.
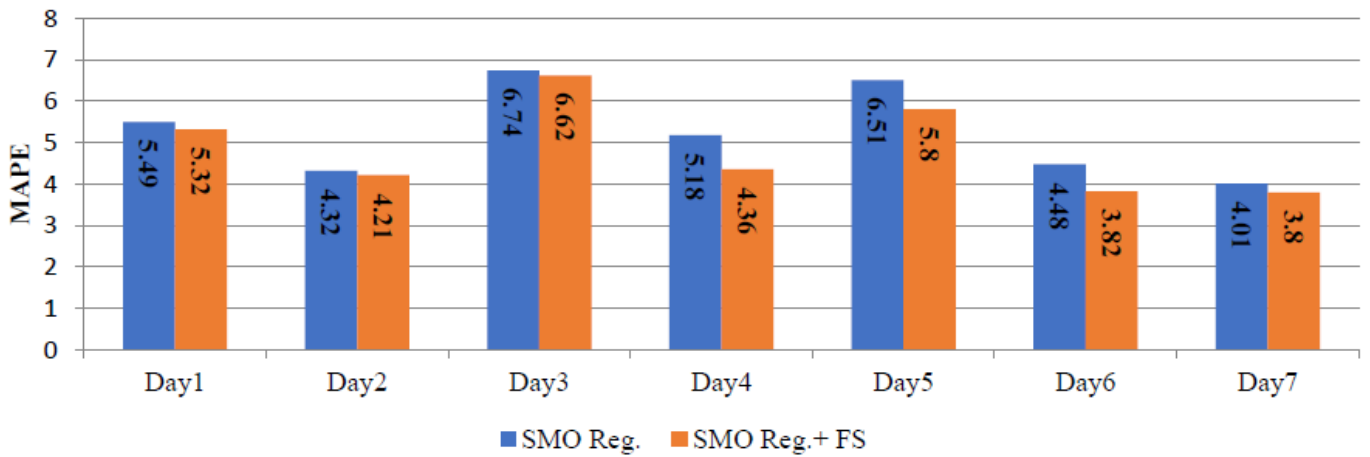

Figure 6. Daily MAPE corresponding to SMO reg. and SMO reg.+ FS for winter season.

Table 5. Improvement of error measures using SMO regression+ FS over other approaches.

Sr. No.	Method	Average MAPE	Improvement (%)
1.	SMO regression+ FS	5.13	-
2.	J48	10.72	52.16
3.	J48+ FS	9.55	46.32
4.	Bagging	8.086	36.27
5.	Bagging+ FS	7.86	35.15
6.	M5P	6.73	23.77
7.	M5P+ FS	6.09	15.90
8.	SMO regression	5.44	5.75
Sr. No.	**Method**	**Average MAE**	**Improvement (%)**
1.	SMO regression+ FS	1.83	-
2.	J48	3.76	51.33
3.	J48+FS	3.36	45.54
4.	Bagging	2.87	36.24
5.	Bagging+FS	2.81	34.89
6.	M5P	2.66	31.20
7.	M5P+FS	2.28	19.74
8.	SMO regression	1.95	6.15
Sr. No.	**Method**	**Average RMSE**	**Improvement (%)**
1.	SMO regression+ FS	2.96	-
2.	M5P	6.53	54.68
3.	J48	5.55	46.64
4.	J48+ FS	4.74	37.51
5.	M5P+ FS	4.24	30.23
6.	Bagging	4.08	27.44
7.	Bagging+ FS	4.07	27.29
8.	SMO regression	3.15	5.97

Table 6 summarizes the daily MAPE corresponding to SMO regression and SMO regression +FS. The daily errors for the winter, spring and summer seasons, using the SMO regression and SMO regression + FS are depicted in Figures 6–8. These results indicate that the performance of the SMO regression + FS is generally better than the performance of the SMO regression.

Table 6. Comparative daily MAPE corresponding to SMO regression and SMO regression +FS.

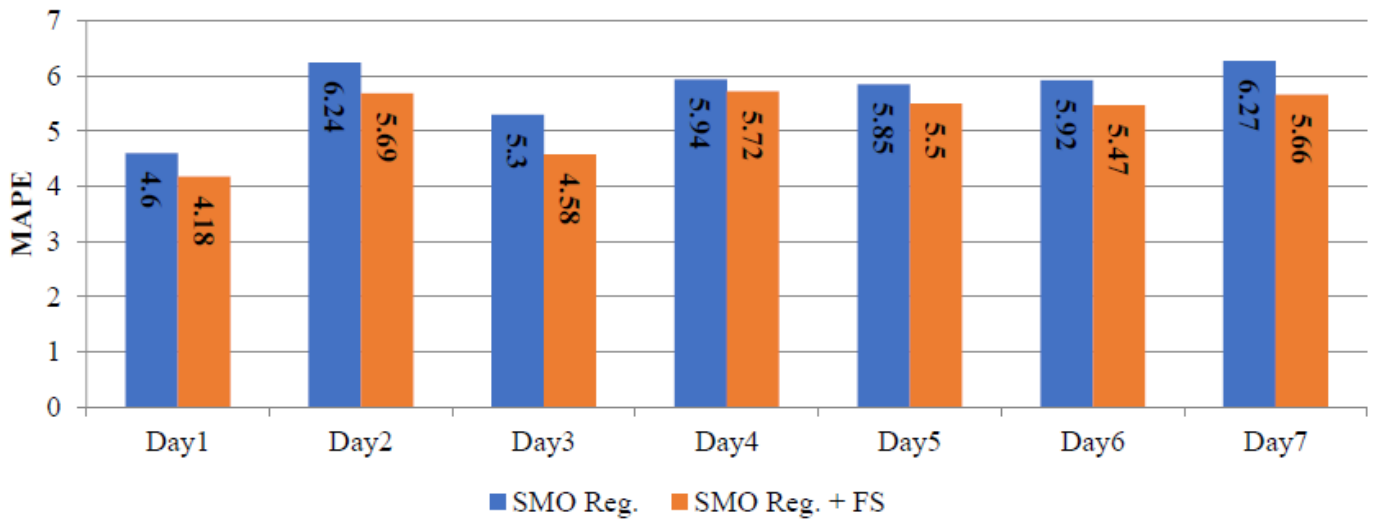| Sr. No.	Days	Winter (22–28 August 2015)		Spring (22–28 October 2015)		Summer (22–28 January 2016)	
		SMO Reg	SMO Reg+FS	SMO Reg	SMO Reg+FS	SMO Reg	SMO Reg+FS
1	Day1	5.49	5.32	4.60	4.18	10.96	10.72
2	Day2	4.32	4.21	6.24	5.69	4.86	4.85
3	Day3	6.74	6.62	5.30	4.58	3.1	3.01
4	Day4	5.18	4.36	5.94	5.72	4.37	4.20
5	Day5	6.51	5.80	5.85	5.50	3.3	3.29
6	Day6	4.48	3.82	5.92	5.47	4.64	4.62
7	Day7	4.01	3.80	6.27	5.66	6.27	6.24
Average		5.25	4.85	5.73	5.26	5.35	5.28

Figure 7. Daily MAPE corresponding to SMO reg. and SMO reg.+ FS for spring season.
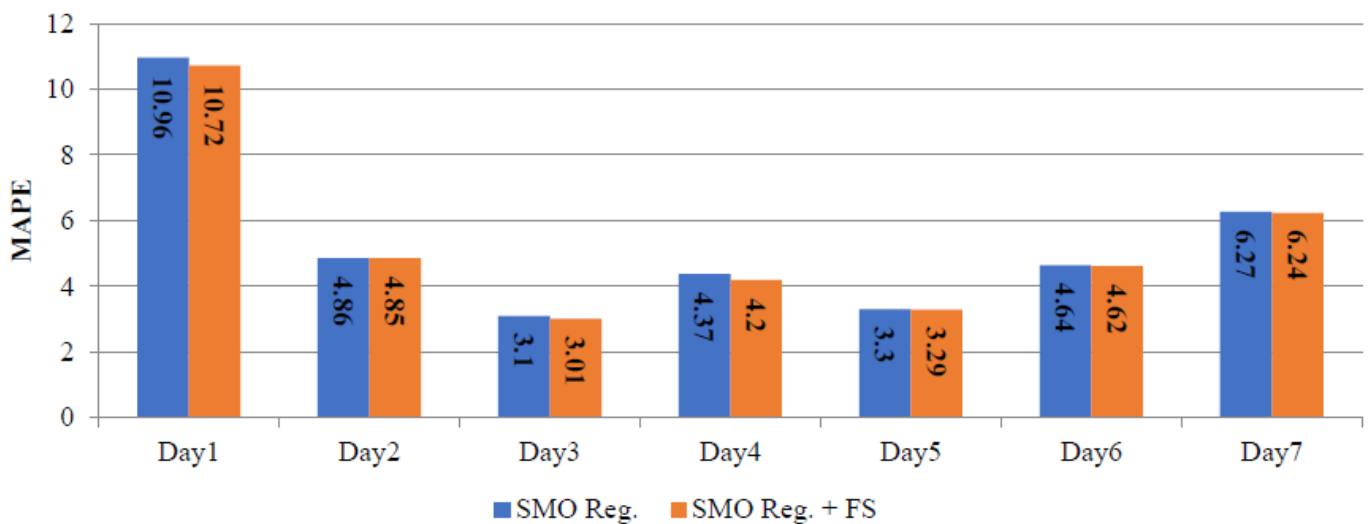

Figure 8. Daily MAPE corresponding to SMO reg. and SMO reg.+ FS for summer season.

6. Conclusions

In this paper, a novel hybrid method for day ahead electricity price forecasting is presented. The elitist GA and tree-based method is used for input features selection. The forecasting is done for whole year with FS and WoFS. MAPE, RMES, MAE and EV have been calculated day-wise, week-wise and season-wise. The result obtained from SMO regression is compared with the results of classification tree (J48) and regression tree (bagging and M5P). It can be observed from the experimental results that SMO regression with FS method provides better forecast results of electricity prices than WoFS and SMO regression outperforms classification tree (J48) and regression tree (bagging and M5P)-based forecasters. It was observed that SMO regression + FS could give improved accuracy (MAPE) in the range of 23.77 and 36.27 (for M5P and bagging) to 52.16 (for J48).

Author Contributions: Conceptualization, methodology and software, A.K.S. and A.S.P.; writing—original draft and formal analysis, A.K.S., A.S.P., R.M.E.; supervision and visualization, A.S.P. and R.M.E.; data curation, validation and investigation, A.K.S. and R.M.E.; writing—review and editing: U.S., S.M. and L.M.-P.; All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The half hourly historical load and price data of New South Wales, Australia (taken from Australian Energy Market Operator (AEMO)) and weather data of Sydney City (www.weatherzone.com/au, accessed on 18 August 2019) has been taken for the forecasting.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

FS	Feature Selection
GA	Genetic Algorithm
MCP	Market Clearing Prices
ANN	Artificial Neural Network
SDA	Stacked Denoising Autoencoders
FFNN	Feed Forward Neural Networks
WT	Wavelet Transform:
FA-PSO	Fuzzy Adaptive Particle Swarm Optimization:
RF	Random Forest
SVM	Support Vector Machine
DE	Differential Evolution
MI	Mutual Information
SMO	Sequential Minimization Optimization
QP	Quadratic Programming
WoFS	Without Feature Selection
KKT	Karush–Kuhan–Tucker
10-FCV	10-Fold Cross Validation
AEMO	Australian Energy Market Operator
RMSE	Root-Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
EV	Error Variance

Appendix A

Calculation of SMO regression:

For a typical SMO algorithm, the dual form of the SVM minimized function and related calculations are given below from Equations (A1) to (A6).

$$
\begin{aligned}
W(\alpha^+, \alpha^-) = \varepsilon.\sum_{i=1}^{l}(\alpha_i{}^+ + \alpha_i{}^-) - \sum_{i=1}^{l}y_i(\alpha_i{}^+ + \alpha_i{}^-) + \\
\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_i(\alpha_i{}^+ + \alpha_i{}^-)(\alpha_j{}^+ + \alpha_j{}^-).k(x_i, x_j)
\end{aligned} \tag{A1}
$$

where, α^+, $\alpha^- = Positive and negative Lagrange's Multiplier$; K = kernel function subject to constraints:

$$
\sum_{i=1}^{l}(\alpha_i^+ + \alpha_i^-) = 0; \alpha_i^+ \geq 0; \alpha_i^- \leq -C \; for \; \forall_i \tag{A2}
$$

Transforming the above sets of equations by,

$$
\lambda_i = \alpha_i^+ - \alpha_i^- \; ; \; |\lambda_i| = \alpha_i^+ + \alpha_i^- \; ; \; k_{ij} = k(x_i, x_j)
$$

Then the model objective function can be represented as,

$$
W(\lambda) = \varepsilon.\sum_{i=1}^{l}|\lambda_i| - \sum_{i=1}^{l}\lambda_i y_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{i=1}^{l}\lambda_i\lambda_j k_{ij} \; such \; that \; \sum_{i=1}^{l}\lambda_i = 0; \tag{A3}
$$

To express analytically, the minimum of model objective function as a function of two parameters i.e., u and v, we have:

$$W(\lambda_u, \lambda_v) = \varepsilon.\left(\left|\lambda_u\right| + \left|\lambda_v\right|\right) - \lambda_u y_u - \lambda_v y_v + \tfrac{1}{2}\lambda_u^2 k_{uu} + \tfrac{1}{2}\lambda_v^2 k_{vv} \\ + \lambda_u \lambda_v k_{uv} + \lambda_u z_u^* + \lambda_v z_v^* + W_c \tag{A4}$$

where λ_u, λ_v are two constants, and

$$\lambda_i^* = \sum_{j \neq u,v}^{l} \lambda_j^*.k_{ij} = f_i^* - \lambda_u^*.k_{ui} - \lambda_v^*.k_{vi} - b* \, ; \, f_i^* = f(x_i, \lambda*, b)$$

Representing the model objective function in terms of a single Lagrange's multiplier by substituting $s* = \lambda_u + \lambda_v = \lambda_u^* + \lambda_v^*$ for the constraint to be true after a step in parametric space keeping summative values of 'λ' be fixed:

$$W(\lambda_v) = \varepsilon.\left|s* - \lambda_v\right| + \varepsilon\left|\lambda_v\right| - (s* - \lambda_v)y_u - \lambda_v y_v \tfrac{1}{2}(s* - \lambda_v)2k_{uu} + \\ \tfrac{1}{2}\lambda_v^2 k_{vv} + (s* - \lambda_v)\lambda_v k_{uv} + (s* - \lambda_v)z_u^* + \lambda_v z_v^* + W \tag{A5}$$

Taking partial derivative w.r.t. λ_v and equating to zero, $\frac{dW(\lambda_v)}{dx} = 0$, we obtain:

$$\lambda_v = \lambda_v^* + \frac{1}{\eta}\left(y_v - y_u + f_u^* - f_v^* + \varepsilon\left(sgn(\lambda_u) - sgn(\lambda_v)\right)\right) \tag{A6}$$

where $\eta = k_{vv} - k_{uu} - 2k_{uv}$; $sgn(\lambda_u) = \frac{d|\lambda_u|}{dx}$; $sgn(\lambda_v) = \frac{d|\lambda_v|}{dx}$

The steps involved will minimize the global objective function, if any of the parameter violates the KKT (Karush–Kuhan–Tucker) conditions of regression i.e.,

$$\lambda_i = 0 \Leftrightarrow |y_i - f_i| \leq \varepsilon - C < \lambda_i \neq 0 < C \Leftrightarrow |y_i - f_i| = \varepsilon \, and \, |\lambda_i| = C \Leftrightarrow |y_i - f_i| \leq \varepsilon$$

If none of the parameters violates the KKT condition, the global minima have been reached.

References

1. Tso, G.K.; Yau, K.K. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* **2007**, *32*, 1761–1768. [CrossRef]
2. Alghandoor, A.; Phelan, P.; Villalobos, R.; Phelan, B. US manufacturing aggregate energy intensity decomposition: The application of multivariate regression analysis. *Int. J. Energy Res.* **2008**, *32*, 91–106. [CrossRef]
3. Zaza, F.; Paoletti, C.; LoPresti, R.; Simonetti, E.; Pasquali, M. Multiple regression analysis of hydrogen sulphide poisoning in molten carbonate fuel cells used for waste-to-energy conversions. *Int. J. Hydrogen Energy* **2011**, *36*, 8119–8125. [CrossRef]
4. Kumar, U.; Jain, V. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* **2010**, *35*, 1709–1716. [CrossRef]
5. Rumbayan, M.; Abudureyimu, A.; Nagasaka, K. Mapping of solar energy potential in Indonesia using artificial neural network and geographical information system. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1437–1449. [CrossRef]
6. Matallanas, E.; Castillo-Cagigal, M.; Gutiérrez, A.; Monasterio-Huelin, F.; Caamano-Martin, E.; Bote, D.M.; Jimenez-Leube, J. Neural network controller for active demand-side management with PV energy in the residential sector. *Appl. Energy* **2012**, *91*, 90–97. [CrossRef]
7. Cheong, C.W. Parametric and non-parametric approaches in evaluating martingale hypothesis of energy spot markets. *Math. Comput. Model.* **2011**, *54*, 1499–1509. [CrossRef]
8. Wesseh, P.K., Jr.; Zoumara, B. Causal independence between energy consumption and economic growth in Liberia: Evidence from a non-parametric bootstrapped causality test. *Energy Policy* **2012**, *50*, 518–527. [CrossRef]
9. Hubicka, K.; Marcjasz, G.; Weron, R. A Note on Averaging Day-Ahead Electricity Price Forecasts across Calibration Windows. *IEEE Trans. Sustain. Energy* **2019**, *10*, 321–323. [CrossRef]
10. Alanis, A.Y. Electricity prices forecasting using artificial neural networks. *IEEE Lat. Am. Trans.* **2018**, *16*, 105–111. [CrossRef]
11. Wang, L.; Zhang, Z.; Chen, J. Short-term electricity price forecasting with stacked denoising autoencoders. *IEEE Trans. Power Syst.* **2017**, *32*, 2673–2681. [CrossRef]
12. Mosbah, H.; El-Hawary, M. Hourly electricity price forecasting for the next month using multilayer neural network. *Can. J. Electr. Comput. Eng.* **2016**, *39*, 283–291. [CrossRef]

13. Sarikprueck, P.; Lee, W.; Kulvanitchaiyanunt, A.; Chen, V.C.P.; Rosenberger, J. Novel Hybrid Market Price Forecasting Method With Data Clustering Techniques for EV Charging Station Application. *IEEE Trans. Ind. Appl.* **2015**, *51*, 1987–1996. [CrossRef]

14. González, J.P.; San Roque, A.M.; Perez, E.A. Forecasting functional time series with a new Hilbertian ARMAX model: Applica-tion to electricity price forecasting. *IEEE Trans. Power Syst.* **2017**, *33*, 545–556. [CrossRef]

15. Anamika; Peesapati, R.; Kumar, N. Electricity Price Forecasting and Classification Through Wavelet–Dynamic Weighted PSO–FFNN Approach. *IEEE Syst. J.* **2018**, *12*, 3075–3084. [CrossRef]

16. Wang, K.; Xu, C.; Zhang, Y.; Guo, S.; Zomaya, A.Y. Robust Big Data Analytics for Electricity Price Forecasting in the Smart Grid. *IEEE Trans. Big Data* **2019**, *5*, 34–45. [CrossRef]

17. Li, P.; Zhang, J.; Li, C.; Zhou, B.; Zhang, Y.; Zhu, M.; Ning, A. Dynamic Similar Sub-Series Selection Method for Time Series Forecasting. *IEEE Access* **2018**, *6*, 32532–32542. [CrossRef]

18. Tahmasebifar, R.; Sheikh-El-Eslami, M.K.; Kheirollahi, R. Point and interval forecasting of real-time and day-ahead electricity prices by a novel hybrid approach. *IET Gener. Transm. Distrib.* **2017**, *11*, 2173–2183. [CrossRef]

19. Abedinia, O.; Amjady, N.; Zareipour, H. A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems. *IEEE Trans. Power Syst.* **2017**, *32*, 62–74. [CrossRef]

20. González, C.; Mira-McWilliams, J.; Juárez, I. Important variable assessment and electricity price forecasting based on regression tree models: Classification and regression trees, Bagging and Random Forests. *IET Gener. Transm. Distrib.* **2015**, *9*, 1120–1128. [CrossRef]

21. Elattar, E.E.; Elsayed, S.K.; Farrag, T.A. Hybrid Local General Regression Neural Network and Harmony Search Algorithm for Electricity Price Forecasting. *IEEE Access* **2020**, *9*, 2044–2054. [CrossRef]

22. Zhang, C.; Li, R.; Shi, H.; Li, F. Deep learning for day-ahead electricity price forecasting. *IET Smart Grid* **2020**, *3*, 462–469. [CrossRef]

23. Zheng, K.; Wen, B.; Wang, Y.; Chen, Q. Impact of electricity price forecasting errors on bidding: A price-taker's perspective. *IET Gener. Transm. Distrib.* **2020**, *14*, 6259–6266. [CrossRef]

24. Zhang, D.; Li, Q.; Mugera, A.W.; Ling, L. A hybrid model considering cointegration for interval-valued pork price forecasting in China. *J. Forecast.* **2020**, *39*, 1324–1341. [CrossRef]

25. Furqan, A.M.; Younis, S. Multi-Horizon Electricity Load and Price Forecasting using an Interpretable Multi-Head Self-Attention and EEMD-Based Framework. *IEEE Access* **2021**, *9*, 85918–85932.

26. Huang, C.-J.; Shen, Y.; Chen, Y.-H.; Chen, H.-C. A novel hybrid deep neural network model for short-term electricity price forecasting. *Int. J. Energy Res.* **2021**, *45*, 2511–2532. [CrossRef]

27. Srivastava, A.K.; Singh, D.; Pandey, A.S.; Maini, T. A Novel Feature Selection and Short-Term Price Forecasting Based on a Decision Tree (J48) Model. *Energies* **2019**, *12*, 3665. [CrossRef]

28. Srivastava, A.K.; Singh, D.; Pandey, A.S. Short Term Load Forecasting Using Regression Trees: Ramdom Forest, Bagging & M5P. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 1898.