

Article

A Novel Feature Selection and Short-Term Price Forecasting Based on a Decision Tree (J48) Model

Ankit Kumar Srivastava ^{1,*}, Devender Singh ², Ajay Shekhar Pandey ³ and Tarun Maini ²

¹ Electrical Engineering Department, Institute of Engineering & Technology, Dr. Rammanohar Lohia Avadh University, Ayodhya 224001, India

² Electrical Engineering Department, Indian Institute of Technology (BHU), Varanasi 221005, India; dsingh.eee@itbhu.ac.in (D.S.); tmaini.rs.eee13@itbhu.ac.in (T.M.)

³ Electrical Engineering Department, Kamla Nehru Institute of Technology, Sultanpur 228118, India; ajayshekharpandey@gmail.com

* Correspondence: aankit.srivastava32@gmail.com; Tel.: +91-9451110115

Received: 26 July 2019; Accepted: 11 September 2019; Published: 25 September 2019



Abstract: A novel feature selection method based on a decision tree (J48) for price forecasting is proposed in this work. The method uses a genetic algorithm along with a decision tree classifier to obtain the minimum number of features giving an optimum forecast accuracy. The usefulness of the proposed approach is established through the performance test of the forecaster using the feature selected by this approach. It is found that the forecast with the selected feature consistently out-performed than that having larger feature set.

Keywords: price forecasting; J48 classifier; feature selection; elite genetic algorithm; confidence interval

1. Introduction

In the de-regulated scenario, generators, as well as consumers, are free to buy and sell electricity as per their choice. Every market participant needs to know the accurate electricity price for each load block to achieve maximum profitability. If the electricity market prices can be predicted accurately, generating companies and large-scale enterprises can reduce their risks and further maximize outcomes [1]. Price forecasting differs with load forecasting due to uncertainties involved in operation and bidding strategies [2]. It is more complex than load forecasting. In the current deregulated environment, the price forecasting has emerged as one of the major challenges for researchers and academics [3]. Researchers are continuously working to develop efficient tools and algorithms for electricity price forecasting.

Although load forecasting and price forecasting are mutually dependent, electricity price forecasting is a much more complex process due to its unique characteristics, such as non-constant mean and variance, high frequency, calendar effect, multiple seasonality, high level of volatility and high percentage of unusual price movements. These characteristics are due to various reasons such as, non-storable nature of electrical energy, inelastic nature of demand over short time period, balance between demand and supply, and oligopolistic generation side. Due to these unique aspects, electricity price forecasting methods need sophisticated and modern techniques and tools to cater to the demand of market players. Electricity price forecasting can be further divided on the basis of time span and tools used.

Alanis et al. [4] proposed, recurrent neural network which is based on the Kalman filter and which included stability proof using the Lyapunov methodology for cases of one-step ahead and n-step ahead electric energy price prediction. Rafiei et al. [5] used a probabilistic approach for the hourly electricity price forecasting and have used a bootstrapping technique to implement the uncertainty

in the forecasting model as an uncertainty factor. A generalized learning method is applied for fast and low computational cost for daily forecasting. Kim et al. [6] introduced a cuckoo search Levenberg–Marquardt (CSLM)-trained, CSLM feedforward neural network (CSLM-FFNN) for the electricity price forecasting process, which combines the improved Levenberg–Marquardt and cuckoo search algorithms using actual power generation and system load as input sets. Darudi et al. [7] used a hybrid electricity price forecasting methodology with a new data fusion algorithm combining artificial neural network (ANN), adaptive neuro-fuzzy inference system, and autoregressive moving average methods to extract the advantages of these forecasting engines. Sarikprueck et al. [8] presented a novel hybrid market price forecasting method with data clustering techniques to predict very short time price forecasting for non-spike and spike wholesale market prices. The authors used support vector classification for spike price occurrences and then support vector regression to forecast the value of both non-spike and spike prices. Furthermore, three clustering techniques including classification and regression trees, K-means, and stratification methods are used to minimize high error spike magnitude evaluation. Wu et al. [9] proposed a hybrid, two-stage integrated price and load forecasting model to predict the integrated day-ahead electricity price as well as load in a smart grid. Wan et al. [10] discussed, a comprehensive evaluation model for probabilistic price forecasting. In this a hybrid method is used to construct prediction intervals of MCPs with a two-stage formulation. Miranian et al. [11] presented a singular spectrum analysis (SSA) technique-based method for day-ahead electricity price forecasting. SSA is used to decompose the original electricity price series into trend, periodic and noisy components. Wu et al. [12] used functional principal component analysis (FPCA) which is a sophisticated tool of multivariate analysis to forecast the electricity prices. They further used a recursive dynamic factor analysis (RDFA) algorithm to reduce the arithmetic complexity. Mandal et al. [13] developed a hybrid intelligent algorithm using data filtering based on wavelet transform. In this an optimization technique based on the firefly (FF) algorithm along with a soft computing model based on fuzzy ARTMAP (FA) network has been effectively incorporated to forecast day-ahead electricity prices in the Ontario market. Chen et al. [14] introduced a method using extreme learning machine and bootstrapping for electricity price forecasting. They have used a fast method of single hidden layer feed-forward neural networks instead of the slow method of the back-propagation (BP) approach. The extreme learning machine (ELM) is used to overcome the drawbacks of ANNs. Catalao et al. [15] proposed a combined model of wavelet transform, particle swarm optimization, and an adaptive-network-based fuzzy inference system to predict the electricity price in case study of electricity market of mainland Spain. Saini et al. [16] emphasized parameter selection in electricity price forecasting. They used a support vector machine (SVM) tool for function approximation and genetic algorithm (GA) for optimization of electricity price forecasting model. Huang et al. [17] presented, a data mining approach for electricity price classification. The authors focus on price threshold prediction for operation decisions for demand side management. Areekul et al. [18] developed a hybrid methodology that combines both artificial neural network (ANN) and autoregressive integrated moving average (ARIMA) models for electricity price forecasting.

The process of selecting a subset of relevant features from a large dataset is termed feature selection [19]. Computational complexity of the system models can be effectively handled with the help of feature selection. Numbers of feature selection techniques are available for the reduction of dimensionality and optimization of the features [20]. Feature selection is problem-dependent. In [21,22], Feature ranking and subset selection are the two techniques used for feature selection. Amjady et al. [23] proposed an improved version of a relief algorithm for feature selection and a hybrid neural network for forecasting of electricity price. In another paper, Amjady et al. [24] developed a combination of a feature selection technique and cascaded neuro-evolutionary algorithm (CNEA). They used an improved version of the mutual information (MI) technique for feature selection, along with CNEA, which is composed of cascaded forecasters, a neural network (NN), and an evolutionary algorithm (EA). Abedinia et al. [25] proposed a feature selection method for load and price forecasting. Here, modeling of the interaction, in addition to relevancy and redundancy,

based on information-theoretic criteria along with a hybrid filter-wrapper approach for feature selection, has been used. Tahmasebifar et al. [26] employed a new hybrid method to estimate point forecasts. The hybrid employed combination of wavelet transformation (WT), extreme learning machine (ELM), feature selection based on mutual information (MI), and bootstrap approaches in an ensemble structure. Abedinia et al. [27] presented a novel forecast approach based on combination of a neural network with a meta-heuristic algorithm as the hybrid forecasting engine. Gonzalez et al. [28] discussed regression tree-based models, like bagging and random forests, to identify the inputs dominating the marginal price and highlighted the effectiveness of the proposed ensemble of tree-based models. Baek et al. [29] presented a next day forecast for total daily solar irradiation through an ensemble of multiple machine learning algorithms using forecasted weather scenarios from numerical weather prediction (NWP) models. Many data trimming techniques, such as outlier detection, input data clustering, and input data pre-processing, are developed and compared. Gao et al. [30] proposed a short-term electricity load forecasting model based on an empirical mode decomposition-gated recurrent unit (EMD-GRU) with feature selection (FS-EMD-GRU). First, the original load series is decomposed into several sub-series by EMD. Then, they analyzed the correlation between the sub-series and the original load series through the Pearson correlation coefficient method.

Price forecasting is a tricky problem as it depends not only on combination of several factors, such as time of day and type of weather (temperature, humidity etc.), but also on the bid pricing and market dynamics. The proposed method rests on the motivation that the electricity prices are a function of combinations of feature values rather than assuming that each of the features contribute a specific amount to the prices, i.e., regression with time series and other variables.

Tree-based classification J48 has been used in several problems. In this paper an attempt has been made to study its applicability and usefulness for the price forecasting problem. In earlier literature it was observed that various classifiers based on SVM, ANN, fuzzy, etc., had been used. However, the method based on a tree-based classifier based on J48 had not been used earlier. In this paper the following novel contribution have been made:

1. Using a combination of GA and J48 for feature selection in the price forecasting problem.
2. Using the J48 classifier for prediction of Australian price data.
3. Applying the confidence interval for fixing the error margins in the prices forecasted.
4. Season-wise feature selection is attempted to draw certain insight on the number and type of features affecting the price in different seasons.

This paper focuses on feature selection for short-term price forecasting. Therefore, only the forecasting of day-ahead prices is considered with and without selected features. Although the method proposed is general in nature, in this work, the forecast of electricity price on a half-hourly basis and for each day is considered. The method proposes the decision tree algorithm combined with genetic algorithms for feature selection for price forecasting. Later, these selected features are used with decision trees to forecast the prices. The usefulness of the proposed algorithm is established by comparing the forecasts obtained using the full feature set with that of the reduced feature set. The work proposed in the paper can lead to some new insights on the type and the seasonality of features and their effects on electricity prices.

Section 2 of the paper explains the working of the decision tree classifier. The application of the classifier for the task of price forecasting is explained with an example. Section 3 presents the method of feature selection using a GA and J-48 classifier. Section 4 present the results of feature selection and the performance of the proposed method in detail. The conclusion and findings are summarized in Section 5.

2. Decision Tree Classifier (J48)

The J48 classifier is a decision tree classifier. By applying J48 one can predict the class label of a new record in a dataset out of a list of dependent and independent variables. The attribute which is to

be predicted is known as the dependent variable and the other attributes which help in predicting it are known as independent variables. The decision tree models the classification process through symbols of nodes and branches. The nodes of the tree denote different attributes and branches represents the splitting of the attributes based on their values and leaves denotes the classes of the dependent variable. The node at a particular level is found on the basis of highest information gain ratio obtained on the set of available attributes, and the same attribute is selected for further branching. Splitting is done on the basis of the highest information gain for a selected node attribute. It creates a decision tree based on the attribute values of the available training set.

Here, a stratified 10-fold cross-validation (10-FCV) classification accuracy using J48 has been used as the fitness function in the genetic algorithm (GA) for feature selection. The standard way of forecasting the error rate of a learning technique given a single, fixed sample of data is to use stratified 10-FCV. Data is divided into ten parts randomly and the class is represented in approximately the same proportions as in the full dataset. The error rate is calculated on the holdout set after each part is held out in turn and the learning scheme trained on the remaining nine. The learning process is executed ten times on different training sets. The 10 error estimates are finally averaged to obtain an overall estimate of the classification error [31].

Dataset $S = (X_1, \dots, X_n, C_i)$, where n is the number of independent variables and C_i is the dependent variable, $i = 1, 2, \dots, K$, where K is the number of classes of dependent variable. A new node is added to the decision tree for every partition. In a partition S , a test attribute X is selected for further partitioning the set into S_1, S_2, \dots, S_L . New nodes for S_1, S_2, \dots, S_L are created and these are added to the decision tree as children of the node for S . The node for S is labelled with test X , and partitions S_1, S_2, \dots, S_L are then recursively partitioned. A partition in which all the records have identical class labels is not partitioned further and the leaf corresponding to it is labeled with the dependent variable.

Decision Tree Algorithm (J48): The algorithm to construct the decision tree using J48 takes following steps [32].

Step 1: Calculate $Entropy(S)$ of the training set S as follows:

$$Entropy(S) = - \sum_{i=1}^K \left\{ \left[\frac{freq(C_i, S)}{|S|} \right] \log_2 \left[\frac{freq(C_i, S)}{|S|} \right] \right\} \quad (1)$$

where $|S|$ is the number of sample in the training set. C_i is a dependent variable, $i = 1, 2, \dots, K$, K is the number of classes of the dependent variable, and $freq(C_i, S)$ is the number of samples included in class C_i .

Step 2: Calculate the $Information\ Gain_X(S)$ for test attribute X to partition:

$$Information\ Gain_X(S) = Entropy(S) - \sum_{i=1}^L \left[\left(\frac{|S_i|}{|S|} \right) Entropy(S_i) \right] \quad (2)$$

where L is the number of test outputs, X, S_i is a subset of S corresponding to i^{th} output, and $|S_i|$ is the number of dependent variables of subset S_i . For a particular attribute partition the subset which provides maximum information gain will be selected as the threshold. The two partitions S and $S-S_i$ will be the branch of the node. In case the instances belong to the same class, the tree represents a leaf; so the leaf is returned by labeling with the same dependent variables (class).

Step 3: Calculate the partition information value $Split\ Info(X)$ acquiring for S partitioned into L subsets:

$$Split\ Info(X) = - \sum_{i=1}^L \left[\left(\frac{|S_i|}{|S|} \right) \log_2 \left(\frac{|S_i|}{|S|} \right) + \left(1 - \left(\frac{|S_i|}{|S|} \right) \right) \log_2 \left(1 - \left(\frac{|S_i|}{|S|} \right) \right) \right] \quad (3)$$

Step 4: Calculate the *Gain Ratio*(X):

$$\text{Gain Ratio}(X) = \frac{\text{Information Gain}_X(S)}{\text{Split Info}(X)} \quad (4)$$

Step 5: The attribute having the highest gain ratio will be designated as the root node and the same computation from step 1 to step 4 is done for every intermediate node until all the instances are exhausted and it reaches the leaf node according to step 2.

Following are the characteristic features of the J48 algorithm:

- (1) It handles classification with the missing values in the data.
- (2) It can be applied to both discrete and continuous variables.
- (3) It also performs the pruning of the tree.
- (4) It can handle high dimensional data.
- (5) It replaces internal node with a leaf node and thus reduces the error rate.

3. Input Feature Selection Using a Genetic Algorithm

Genetic Algorithm (GA) is a heuristic search technique based on the Darwinian theory of natural evolution and genetics of survival of the fittest. To generate useful solutions to optimization problems, this heuristic is used. While searching the space for feature selection, no assumption of the relationships among features involved has been considered in this approach. GA can easily encode decisions as sequences of boolean values, allowing for exploration of the feature space by retaining the decisions that benefit the classification task. It also avoids local optimums due to their intrinsic randomness simultaneously [33,34]. Also, it generates the solutions to optimization problems using operators inspired by natural evolution, like selection, crossover, and mutation [35,36].

The present method of decision trees is used for two purposes viz. (i) feature selection and (ii) to classify the target class, i.e., the price for the given feature set. The regression trees have normally been used for predicting the data given the feature values. Unlike regression trees, which are normally used to predict a value for given set of features, the present method of using a decision tree requires that the target classes be fixed beforehand for building the decision tree. Thus, we need to have discretized data for all features, and the data are discretized in a much wider range as there may be cases which may be beyond any values. If the data could go beyond any value, then it is fixed to the maximum value of the minimum value taken for that particular variable. In the present work the variables were discretized at 2.5% of the range. The value was arrived through experimenting with values in the range of 1–5%.

In the present work, feature selection is performed using an evolutionary elitist Genetic algorithm [34]. The significant features of input dataset, which affects the forecasting process meritoriously, are selected simultaneously.

In the elitist GA the 20% elite population is passed to the next generation, by which the next generation has the population of feature sets whose classification accuracy is not less than the previous generation. Parameters for this method are same as in [37], and the fitness function used is the classification accuracy (stratified 10-fold cross-validation classification accuracy) of the given dataset. Strings of 0s and 1s are taken as chromosome segments in the present problem: in these chromosomes 1 shows that the particular feature is selected and 0 shows that the feature corresponding to that index is not selected. The length of the string is the same as the number of features in the dataset. All the computation of the fitness function is the stratified 10 FCV classification accuracy, which is being computed using the WEKA data mining workbench [31], through a decision tree-based J48 classifier. Classification accuracy shows an estimate of the number of instances correctly classified. Resulting

chromosomes with an optimal fitness value is the set of selected features. The mathematical definition of the fitness function for the GA for feature selection is given below:

$$\text{Fitness function} = \text{Classification accuracy} = \left(\frac{\text{No. of instances correctly classified}}{\text{Total No. of instance in the data set}} \right) * 100 \quad (5)$$

Roulette wheel selection has been used in this work and then single-site crossover with 0.7 probability is performed in every step. The mutation operator is also performed with a mutation probability 0.005. Furthermore, 20% of the elite population was passed on to the next generation as it is. The final combination or encoding of chromosomes provide the best set of selected features for the forecasting of the electricity price.

The flowchart of the GA for feature selection is shown in Figure 1.

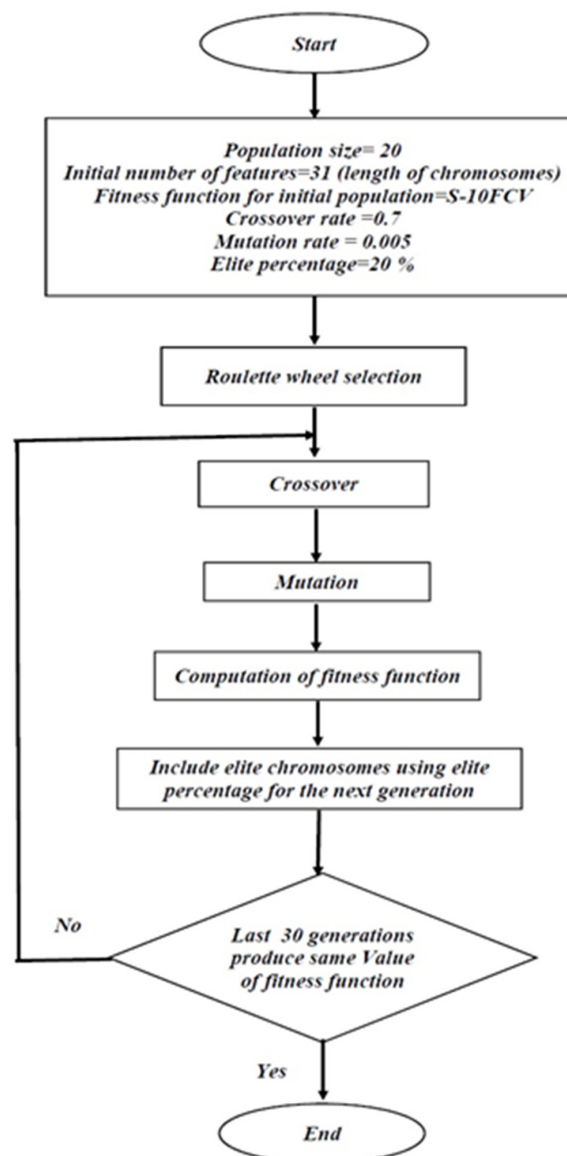


Figure 1. Flowchart of feature selection using a GA.

4. Results

For short-term price forecasting data is collected from the Australian Energy Market Operator for New South Wales, Australia. The data consists of half-hourly load and prices of all seasons from

January 2014 to June 2016. The weather data of Sydney City is taken from www.weatherzone.com/au. Weather data considered in the present studies are half-hourly wind speed, temperature, and humidity. All the data were quantized at 40 levels, each level consisting of 2.5% of the range. Thus, for a particular week, all the data have been classified to have only 40 discrete values.

Table 1 shows the list of features which are assumed to affect half-hourly electricity prices. It has been observed that there is similarity in energy prices for the same hours of the day, although, there may be small shift in the price of similar hours. It is quite common to assume the phenomena of “similar hours” of the day as input in load and price forecasting literature. The set of input features thus consisted of 31 features. The training set for the classifier consisted of 2016 datasets. The results of the present work are discussed in two parts. The first part discusses the results of feature selection and its significance. The second part discusses the accuracy of the forecasted results.

Table 1. List of features which are assumed affect half-hourly electricity.

Variable	Variable Timing	Feature Name
Load (L)	L(t–23:00)	L6
	L(t–23:30)	L5
	L(t–24:00)	L4
	L(t–01:30)	L3
	L(t–01:00)	L2
	L(t–00:30)	L1
Price (P)	P(t–23:00)	P6
	P(t–23:30)	P5
	P(t–24:00)	P4
	P(t–01:30)	P3
	P(t–01:00)	P2
	P(t–00:30)	P1
Wind Speed (W)	W(t–23:00)	W6
	W(t–23:30)	W5
	W(t–24:00)	W4
	W(t–01:30)	W3
	W(t–01:00)	W2
	W(t–00:30)	W1
Temperature (T)	T(t–23:00)	T6
	T(t–23:30)	T5
	T(t–24:00)	T4
	T(t–01:30)	T3
	T(t–01:00)	T2
	T(t–00:30)	T1
Humidity (H)	H(t–23:00)	H6
	H(t–23:30)	H5
	H(t–24:00)	H4
	H(t–01:30)	H3
	H(t–01:00)	H2
	H(t–00:30)	H1
Day Timing (Ho)	H(t–00:00)	Ho

Feature selection: The method of simultaneous feature selection using an elitist genetic algorithm is employed in this work. The features listed in Table 1 are taken as the input feature set from which the feature selection is done. The process of feature selection was performed weekly. The training set for the classifier consists of 2016 datasets. The graph showing the evolution of the fitness function with generations of the first week of February 2016 is shown in Figure 2.

The training set was taken on the concept of similar weeks. The dataset corresponds to the five similar weeks of the months of the last year and the past week the same year. For example, if the feature

selection is to be performed for the week of 1–7 July 2015, the training set would include the data corresponding to 24–30 June 2015, 1–7 July 2014, 24–30 June 2014, 17–23 June 2014, 8–14 July 2014, and 15–21 July 2014. The training set for classifier consists of 2016 datasets. The classifier uses stratified 10-fold cross-validation classification accuracy methodology for finalizing the feature selection. Thus, the whole of the data is tested in this method at least once. The major advantage of feature selection is that the method can be employed for conducting feature analysis, as to which feature or the component is affecting the forecast or consumption pattern more significantly. A detailed feature analysis is performed on the basis of the present studies.

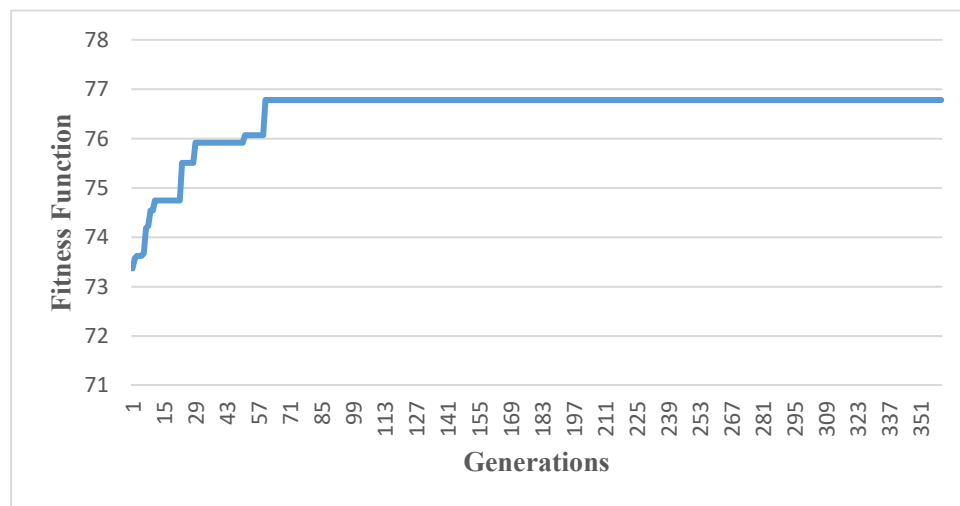


Figure 2. Fitness function with generations.

Table 2 shows that, out of a total of 36 times for which the feature selections were made, the number of times a particular feature was selected. It is observed from Table 2 that the feature price of the present day (P1) is selected in all the runs, the hour type (Ho) was the second most important feature and was selected 25 times. It is notable that the humidity of previous day (H5 and H6) was selected a greater number of times than the same day (H1 and H2). The wind speed and temperatures were selected fewer times than humidity. The load of the immediate hour (L1) was selected only 12 times compared to load of the 24-hours before (L3), which was selected 16 times. The effects of the features can also be analyzed according to the season.

Table 2. Number of times a feature is selected year-wise.

Feature Name	Number of Time Selected	Feature Name	Number of Time Selected
L6	8	W2	11
L5	13	W1	14
L4	16	T6	13
L3	10	T5	9
L2	11	T4	7
L1	12	T3	8
P6	10	T2	8
P5	11	T1	10
P4	16	H6	15
P3	17	H5	16
P2	17	H4	11
P1	36	H3	10
W6	8	H2	11
W5	8	H1	12
W4	5	Ho	25
W3	8		

Table 3 shows the number of times a feature was selected in a particular season. From the table it can be observed that the load previous hour load value is invariably selected in winters whereas, in spring, the previous day load values are not selected. Similarly, the weather variables, such as wind speed, temperature, and humidity, are selected quite regularly in winter as compared to that of summer and spring seasons.

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

Feature Name	Winter	Spring	Summer
L6	4	0	4
L5	4	2	7
L4	4	7	5
L3	5	4	1
L2	6	3	2
L1	4	4	4
P6	4	4	2
P5	5	4	2
P4	6	4	6
P3	7	5	4
P2	6	5	6
P1	12	12	12
W6	3	2	3
W5	5	1	2
W4	3	2	0
W3	5	0	3
W2	2	5	4
W1	3	6	5
T6	5	3	5
T5	2	4	3
T4	3	3	1
T3	4	1	3
T2	2	4	2
T1	5	1	4
H6	6	4	5
H5	7	5	4
H4	2	5	4
H3	3	2	5
H2	3	5	3
H1	4	5	3
Ho	9	8	8

The maximum and minimum number of features selected for each of the seasons are depicted in Table 4.

Table 4. Maximum and minimum number of features selected in all season.

Season	Duration	Maximum Features Selected	Duration	Minimum Features Selected
Winter	Aug. 08–14, 2015	14	July 08–14, 2015	5
Spring	Sept. 1–7, 2015	15	Sept. 08–14, 2015	2
Summer	Feb. 08–14, 2016	19	Dec. 08–14, 2015	3

Table 4 also shows the week for which the minimum and maximum number of features were selected. It is observed that the minimum number of features selected were five in the winter season during July 8–14, 2015. The forecasts obtained for same period are depicted in Figures 3 and 4. The mean absolute percentage error (MAPE) for this week with feature selection (FS) is 9.38, whereas without feature selection (WoFS) it is 11.72. It is observed that, despite the very small number of

features in the case of FS, the results are better, compared to that of WoFS, which has all 31 features. However, the maximum number of features selected were 14 in the winter season during Aug 8–14, 2015. The forecasts obtained for Aug 8–14, 2015 are depicted in Figures 5 and 6. The MAPE for this week with FS is 10.22, whereas the same for WoFS is 10.47. It is observed that the results obtained from FS are better, compared to that of WoFS, which has all the 31 features.

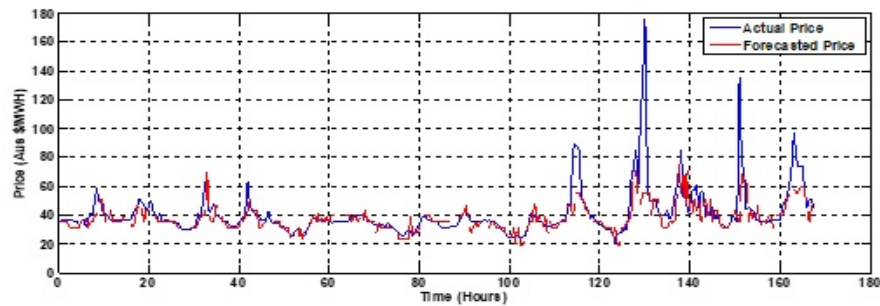


Figure 3. July 8–14, 2015 forecasting without feature selection (MAPE = 11.72).

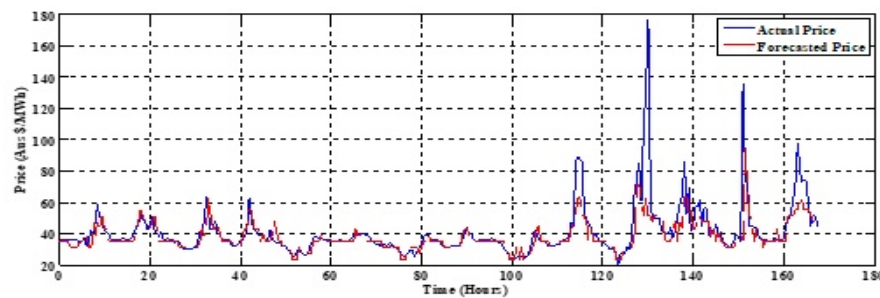


Figure 4. July 8–14, 2015 forecasting while, minimum number of features selected in the winter season (MAPE = 9.38).

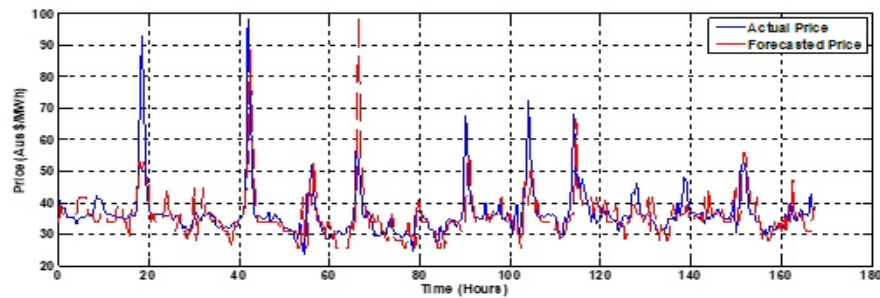


Figure 5. August 8–14, 2015 forecasting without feature selection (MAPE = 10.47).

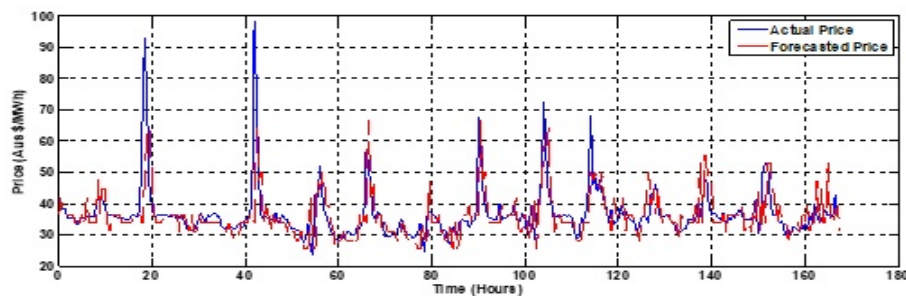


Figure 6. August 8–14, 2015 forecasting, while maximum number of features selected in the winter season (MAPE = 10.22).

It is observed that the minimum number of feature selected were two in the spring season during September 8–14, 2015. The forecasts obtained for September 8–14, 2015 is depicted in Figures 7 and 8. The MAPE for this week with FS is 8.73, whereas for WoFS it is 10.05. It is also observed that, despite very small number of features in the case of FS, the results are better, compared to that of WoFS, which has all the 31 features. However, the maximum number of feature selected were 15 in the winter season during Sept. 1–7, 2015. The forecasts obtained for Sept. 1–7, 2015 is depicted in Figures 9 and 10. The MAPE for this week with FS is 11.23, whereas for WoFS it is 11.70. It is observed that, with the number of features in the case of FS, the results are better, compared to that of WoFS, which has all the 31 features.

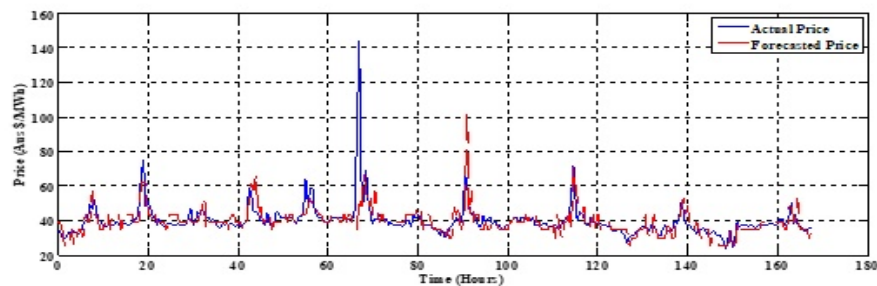


Figure 7. September 8–14, 2015 forecasting without feature selection in the spring season (MAPE = 10.05).

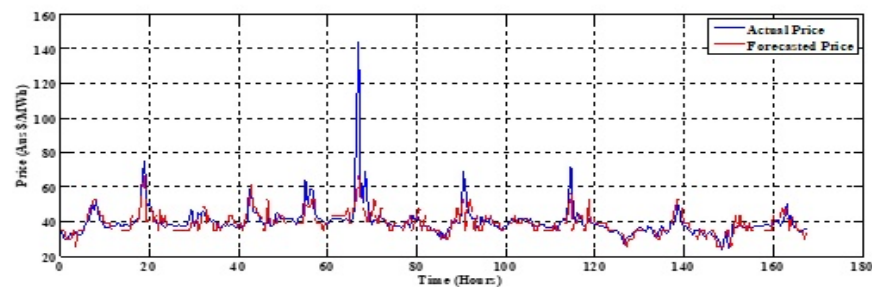


Figure 8. September 8–14, 2015 forecasting while minimum number of features selected in the spring season (MAPE = 8.73).

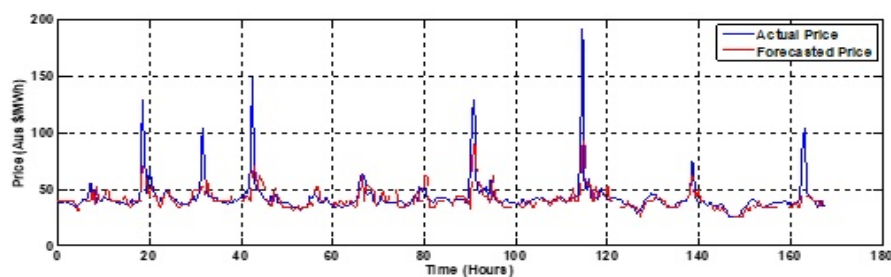


Figure 9. September 1–7, 2015 forecasting without feature selection in the spring season (MAPE = 11.70).

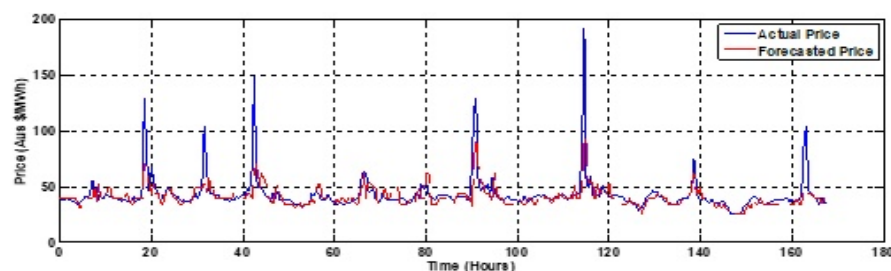


Figure 10. September 1–7, 2015 forecasting with the maximum number of features selected in the spring season (MAPE = 11.23).

It is observed that the minimum number of features selected was three in the summer season during December 8–14, 2015. The forecasts obtained for December 8–14, 2015 is depicted in Figures 11 and 12. The MAPE for this week with FS is 10.69, whereas for WoFS it is 13.08. It is observed that despite the very small number of features in the case of FS, the results are better, compared to that of WoFS, which has all 31 features. However, the maximum number of features selected were 19 in the summer season during Feb 8–14, 2016. The forecasts obtained for Feb 8–14, 2016 is depicted in Figures 13 and 14. The MAPE for this week with FS is 13.70, whereas for WoFS it is 12.81.

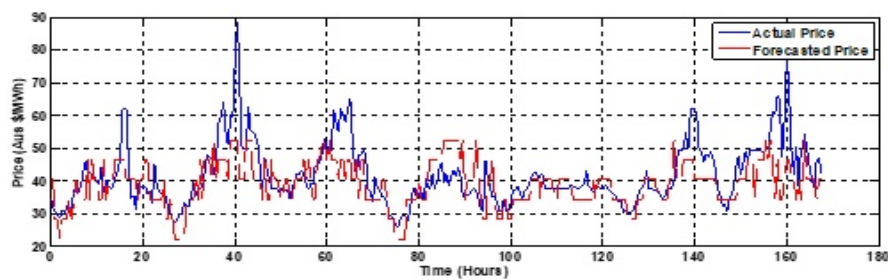


Figure 11. December 8–14, 2015 forecasting without feature selection in the summer season (MAPE = 13.08).

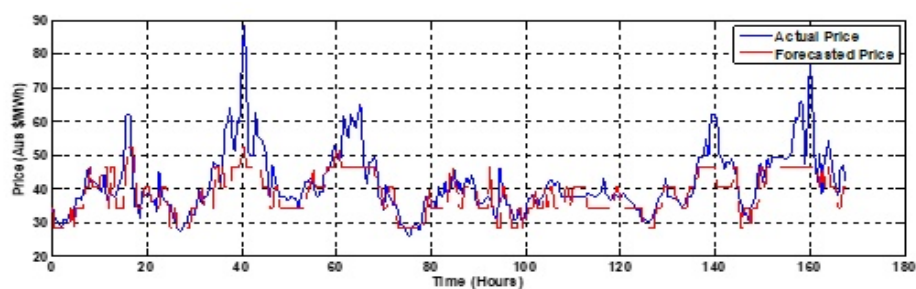


Figure 12. December 8–14, 2015 forecasting with the minimum number of features selected in summer season (MAPE = 11.69).

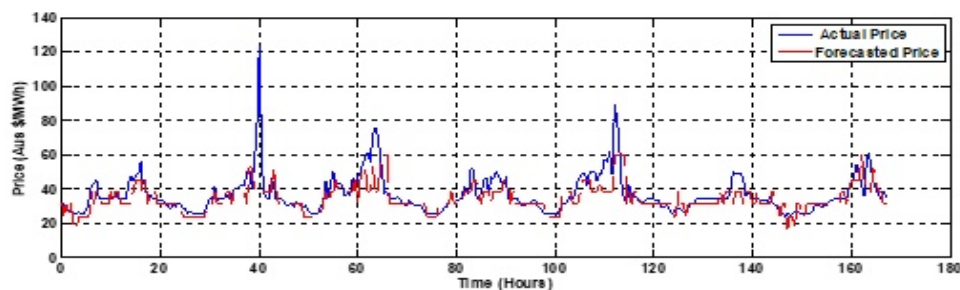


Figure 13. February 08–14, 2016 forecasting without feature selection in the summer season (MAPE = 12.81).

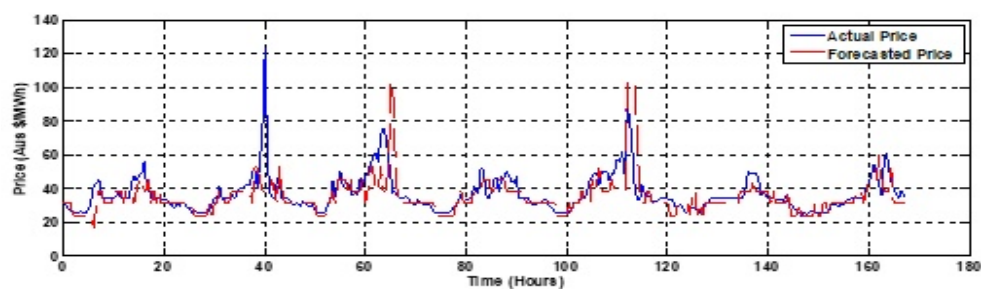


Figure 14. February 8–14, 2016 forecasting, with the maximum number of features selected in summer (MAPE = 13.70).

The feature selection was made season-wise. The MAPE recorded for whole of season of summer (i.e., from Dec. to Feb.) with feature selection was 12.48% whereas, the accuracy without feature selection was found to be 14.06%. The cases depicted in Figure 13 (without feature selection) and Figure 14 (with feature selection) had MAPEs of 12.81 and 13.70, respectively. Thus, for a particular week the error was higher in the case of feature selection to that without feature selection. This specific case was considered as a worst case and, therefore, depicted in Figures 13 and 14.

Forecast accuracy after forming the training set and testing set data, as well as the daily forecast using J48 classifier, are computed. The MAPE is calculated for the entire week. The MAPEs calculated without feature selection (WoFS) and with feature selection (FS) for three seasons are shown in Table 5. The monthly data is categorized in the three seasons viz. summer, winter, and spring. The data from June 2015 to August 2015 falls under summer, the data from September 2015 to November 2015 under spring, and the data from December 15 to February 16 under winter. This result is further obtained weekly without feature selection (WoFS) and with feature selection (FS). Additionally, month-wise, season-wise, and yearly averages are determined for the WoFS and FS.

Table 5. Mean absolute percentage error (MAPE) with FS and WoFS.

Sr. No.	Season	Month	MAPE									
			First (1–7)		Second (8–14)		Third (15–21)		Fourth (22–28)		Mean	
			WoFS	FS	WoFS	FS	WoFS	FS	WoFS	FS	WoFS	FS
1	Winter	June	15.69	11.81	10.15	10.27	8.62	8.26	11.73	12.6	11.54	10.73
		July	9.06	8.61	11.72	9.38	15.64	14.52	12.27	10.95	12.17	10.86
		Aug	12.31	13.25	10.47	10.22	9.39	11.64	12.45	11.4	11.15	11.62
	Winter Mean		12.35	11.22	10.78	9.95	11.21	14.14	12.15	11.65	11.62	11.07
2	Spring	Sep	11.7	11.23	10.05	8.73	10.25	10.92	17.93	14.21	12.48	11.27
		Oct	11.64	10.49	8.91	9.15	9.76	10.3	9.3	8.41	9.9	9.58
		Nov	11.96	9.37	10.99	10.45	25.3	25.81	12.34	12.24	15.14	14.46
	Spring Mean		11.77	10.36	9.98	9.44	15.1	15.68	13.19	11.62	12.51	11.77
3	Summer	Dec	13.27	11.67	13.08	10.69	25.48	17.61	16.23	13.26	17.01	13.3
		Jan	7.9	7.44	17.49	14.62	13.95	13.04	10.42	8.86	12.44	10.99
		Feb	8.93	8.14	12.81	13.7	12.31	11.92	16.89	18.81	12.73	13.14
	Summer Mean		10.03	9.08	14.46	13	17.24	14.19	14.51	13.64	14.06	12.48
Over all Mean			11.38	10.22	11.74	10.8	14.52	13.78	13.28	12.3	12.73	11.77

When the MAPE is compared season-wise, it is observed that the FS based method provides a lower MAPE (11.07) than that WoFS (11.62) for winter. This is found for all the three seasons. However, the effect of feature selection is more pronounced in case of summer season. In summer, FS methods yields a MAPE of 12.48 as compared to 14.06 to that of WoFS. It is observed that the overall mean MAPE is lower with FS as compared to that of WoFS in three weeks, i.e., first, second, and fourth. However, for the third week the MAPEs are comparable for FS and WoFS.

The MAPE obtained for weekdays and weekend days are depicted in Table 6. It is observed that, in general, prices of weekdays can be forecasted less accurately as compared to that of weekend days. The Maximum MAPE is obtained for Wednesdays. Mondays in the summer season yielded the maximum MAPE. It is observed that when the FS is employed the forecasts improves quite significantly for weekend days.

Table 6. Mean absolute percentage error (MAPE) with FS and WoFS, day-wise.

Sr. No.	Season		Week Days						
			Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	Winter	WoFS	12.86	13.33	11.73	10.41	10.76	9.74	10.02
		FS	12.3	12.67	11.86	9.77	9.95	8.8	10.02
2	Spring	WoFS	10.83	11.2	13.49	15.9	14.55	11.44	10.18
		FS	10.64	10.71	12.44	14.55	14.39	10.45	9.29
3	Summer	WoFS	14.23	15.59	16.27	14.34	13.91	11.76	12.37
		FS	13.15	13.41	14.66	14.16	12.88	9.11	10.06
4	Overall	WoFS	12.64	13.37	13.83	13.55	13.07	10.98	10.85
		FS	12.03	12.26	12.98	12.82	12.4	9.45	9.79

The comparison with other forecasting methods (ANN, SVM, and regression model) for the same data is computed for summer season and summarized in Table 7. The comparison shows that the proposed method J48 with FS performs competitively.

Table 7. Comparison of results of the proposed method with other forecasting methods.

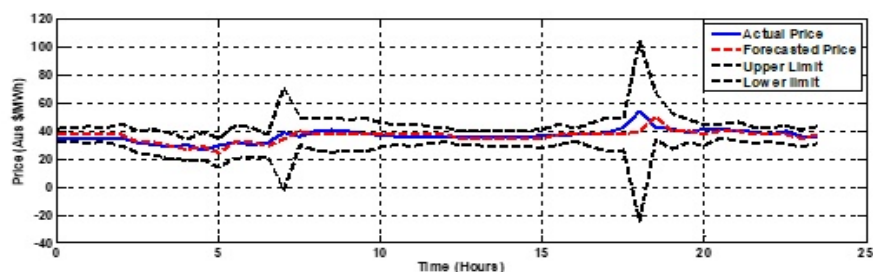
Season	MAPE				
	ANN Model	SVM Model	Regression Model	J48 Model WoFS	Proposed Model J48 with FS
Summer	13.87	10.80	9.98	10.42	8.86

To calculate the confidence interval for a particular day the error of four previous weeks are calculated and arranged at the regular interval of half an hour from 00:00, 00:30, 01:00, ..., up to 23:30. Then, the hourly standard deviation and 2δ are calculated for the 95% confidence interval. The calculation of upper limit and lower limit of the day is given below:

$$\text{Upper limit} = \text{forecasted value} + 2\delta$$

$$\text{Lower limit} = \text{forecasted value} - 2\delta$$

A typical forecast for weekdays of winter season with confidence interval is shown in Figures 15 and 16 for WoFS and FS. It is observed that the FS (6.03) method provides a lower MAPE than that of WoFS (6.88). A forecast for weekend days during the winter season with the confidence interval is shown in Figures 17 and 18 for WoFS and FS. It is observed that the FS (4.98) method provides a lower MAPE than that of WoFS (8.60).

**Figure 15.** Winter weekdays plot of Tuesday (7 July 2015) without selected features with a MAPE of 6.88.

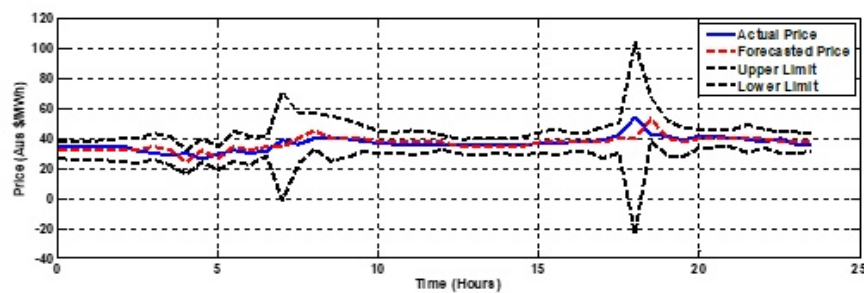


Figure 16. Winter weekdays plot of Tuesday (7 July 2015) with selected features with a MAPE of 6.03.

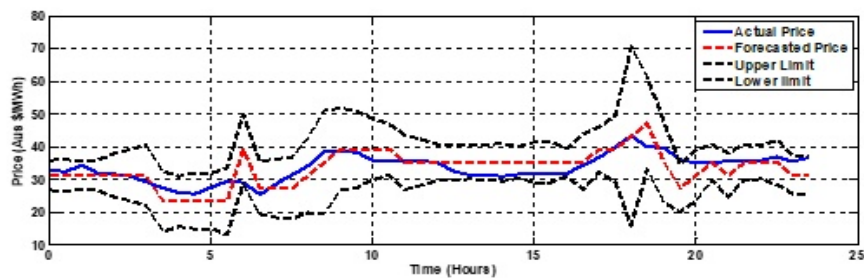


Figure 17. Winter weekend plot of Saturday (11 July 2015) without selected features with a MAPE of 8.60.

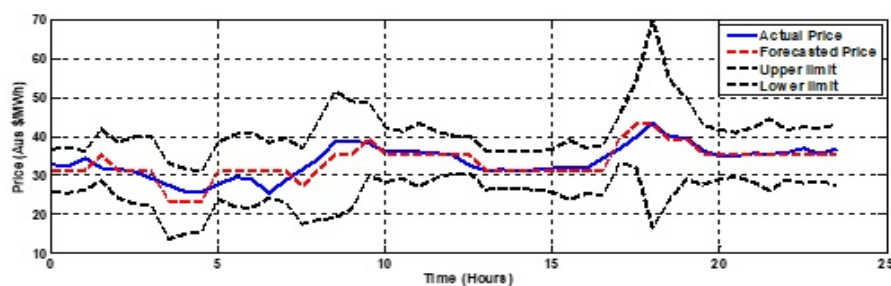


Figure 18. Winter weekends plot of Saturday (11 July 2015) with selected features with a MAPE of 4.98.

5. Conclusion

In this paper, a decision tree method with a novel feature selection is presented for predicting electricity prices. The method uses an elite genetic algorithm and decision tree classifier for feature selection and the process of feature selection has been performed weekly. This paper explains the effect of feature selection year-wise and season-wise and also analyzes the effect of feature selection when the maximum and minimum number of features are selected. It is observed that certain features are selected a greater number of times than the others depending on the season, and also that for the same system in a year the number of required features can be as low as 2 to as large as 19. The mean absolute percentage error (MAPE) has been calculated day-wise, week-wise, and season-wise without feature selection (WoFS) and with feature selection (FS), using predicted data for the whole year. It is established from the result that the proposed feature selection (FS) method provides better forecast accuracy of electricity prices in comparison to that using the full set.

Author Contributions: Conceptualization: A.K.S. and D.S.; methodology: A.K.S., D.S., and A.S.P.; software: A.K.S. and T.M.; validation: A.K.S., D.S., and A.S.P.

Funding: This research was funded by Technical Education Quality Improvement Program (TEQIP-III), IET, Dr. Rammanohar Lohia Avadh University, Ayodhya.

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

Abbreviations

CSLM	Cuckoo search Levenberg–Marquardt
CSLM-FFNN	Cuckoo search Levenberg–Marquardt feed-forward neural network
SSA	Singular spectrum analysis
FPCA	Functional principal component analysis
RDFA	Recursive dynamic factor analysis
WT	Wavelet transform
BP	Back-propagation
ELM	Extreme learning machine
SVM	Support vector machine
GA	Genetic algorithm
CNEA	Cascaded neuro-evolutionary algorithm
MI	Mutual information
EA	Evolutionary algorithm
FCV	Fold cross-validation
MAPE	Mean absolute percentage error
FS	Feature selection
WoFS	Without feature selection

References

- Vahidinasab, V.; Jadid, S.; Kazemi, A. Day-ahead price forecasting in restructured power systems using artificial neural networks. *Electr. Power Syst. Res.* **2008**, *78*, 1332–1342. [[CrossRef](#)]
- Voronin, S.; Partanen, J. A Hybrid electricity price forecasting model for the Finnish electricity spot market. In Proceedings of the 32st Annual International Symposium on Forecasting, Boston, MA, USA, 24–27 June 2012.
- Bunn, D.W. Forecasting loads and prices in competitive power markets. *Proc. IEEE* **2000**, *88*, 163–169. [[CrossRef](#)]
- Alanis, A.Y. Electricity Prices Forecasting using Artificial Neural Networks. *IEEE Lat. Am. Trans.* **2018**, *16*, 105–111. [[CrossRef](#)]
- Rafiei, M.; Niknam, T.; Khooban, M.H. Probabilistic forecasting of hourly electricity price by generalization of elm for usage in improved wavelet neural network. *IEEE Trans. Ind. Inf.* **2017**, *13*, 71–79. [[CrossRef](#)]
- Kim, M.K. Short-term price forecasting of Nordic power market by combination Levenberg-Marquardt and Cuckoo search algorithms. *IET Gener. Transm. Distrib.* **2015**, *9*, 1553–1563. [[CrossRef](#)]
- Darudi, A.; Bashari, M.; Javidi, M.H. Electricity price forecasting using a new data fusion algorithm. *IET Gener. Transm. Distrib.* **2015**, *9*, 1382–1390. [[CrossRef](#)]
- Sarikprueck, P.; Lee, W.J.; Kulvanitchaiyanunt, A.; Chen, V.C.; Rosenberger, J. Novel hybrid market price forecasting method with data clustering techniques for EV charging station application. In Proceedings of the IEEE Industry Applications Society Annual Meeting, Vancouver, BC, Canada, 5–9 October 2014; pp. 1–9.
- Wu, L.; Shahidehpour, M. A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid. *IET Gener. Transm. Distrib.* **2014**, *8*, 1937–1950. [[CrossRef](#)]
- Wan, C.; Xu, Z.; Wang, Y.; Dong, Z.Y.; Wong, K.P. A hybrid approach for probabilistic forecasting of electricity price. *IEEE Trans. Smart Grid* **2014**, *5*, 463–470. [[CrossRef](#)]
- Miranian, A.; Abdollahzade, M.; Hassani, H. Day-ahead electricity price analysis and forecasting by singular spectrum analysis. *IET Gener. Transm. Distrib.* **2013**, *7*, 337–346. [[CrossRef](#)]
- Wu, H.; Chan, S.; Tsui, K.; Hou, Y. A new recursive dynamic factor analysis for point and interval forecast of electricity price. *IEEE Trans. Power Syst.* **2013**, *28*, 2352–2365. [[CrossRef](#)]
- Mandal, P.; Haque, A.U.; Meng, J.; Srivastava, A.K.; Martinez, R. A novel hybrid approach using wavelet firefly algorithm and fuzzy art map for day ahead electricity price forecasting. *IEEE Trans. Power Syst.* **2013**, *28*, 2352–2365. [[CrossRef](#)]
- Chen, X.; Dong, Z.Y.; Meng, K.; Xu, Y.; Wong, K.P.; Ngan, H. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Trans. Power Syst.* **2012**, *27*, 2055–2062. [[CrossRef](#)]
- Catalao, J.P.D.S.; Pousinho, H.M.I. Hybrid wavelet PSO-ANFIS approach for short-term electricity prices forecasting. *IEEE Trans. Power Syst.* **2011**, *26*, 137–144. [[CrossRef](#)]

16. Saini, L.M.; Aggarwal, S.K.; Kumar, A. Parameter optimization using genetic algorithm for support vector machine-based price-forecasting model in national electricity market. *IET Gener. Transm. Distrib.* **2010**, *4*, 36–49. [\[CrossRef\]](#)
17. Huang, D.; Zareipour, H.; Rosehart, W.D.; Amjady, N. Data mining for electricity price classification and the application to demand-side management. *IEEE Trans. Smart Grid* **2012**, *3*, 808–817. [\[CrossRef\]](#)
18. Areekul, P.; Senjyu, T.; Toyama, H.; Yona, A. A Hybrid ARIMA and Neural Network Model for Short-Term Price Forecasting in Deregulated Market. *IEEE Trans. Power Syst.* **2010**, *25*, 524–530. [\[CrossRef\]](#)
19. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
20. Verma, N.K.; Maini, T.; Salour, A. Acoustic signature based intelligent health monitoring of air compressors with selected features. In Proceedings of the 2012 Ninth International Conference on Information Technology–New Generations (ITNG), IEEE, Las Vegas, NV, USA, 16–18 April 2012; pp. 839–845.
21. Ren, D.; Ma, A.Y. Research on feature extraction from remote sensing image. In Proceedings of the 2010 International Conference in Computer Application and System Modeling (ICCSM), Taiyuan, China, 22–24 October 2010; pp. 144–148.
22. Saeys, Y.; Inza, I.; Larranaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [\[CrossRef\]](#)
23. Amjady, N.; Daraeepour, A.; Keynia, F. Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. *IET Gener. Transm. Distrib.* **2012**, *3*, 808–817. [\[CrossRef\]](#)
24. Amjady, N.; Keynia, F. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *IEEE Trans. Power Syst.* **2009**, *24*, 306–318. [\[CrossRef\]](#)
25. 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. [\[CrossRef\]](#)
26. 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. [\[CrossRef\]](#)
27. Abedinia, O.; Amjady, N.; Ghadimi, N. Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm. *Comput. Intell.* **2018**, *34*, 241–260. [\[CrossRef\]](#)
28. Gonzalez, C.; Mira-McWilliams, J.; Juarez, 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. [\[CrossRef\]](#)
29. Baek, M.K.; Lee, D. Spatial and temporal day-ahead total daily solar irradiation forecasting: Ensemble forecasting based on the empirical biasing. *Energies* **2017**, *11*, 70. [\[CrossRef\]](#)
30. Gao, X.; Li, X.; Zhao, B.; Ji, W.; Jing, X.; He, Y. Short-Term Electricity Load Forecasting Model Based on EMD-GRU with Feature Selection. *Energies* **2019**, *12*, 1140. [\[CrossRef\]](#)
31. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2016.
32. Sun, W.; Chen, J.; Li, J. Decision tree and PCA-based fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2007**, *21*, 1300–1317. [\[CrossRef\]](#)
33. Raymer, M.L.; Punch, W.F.; Goodman, E.D.; Kuhn, L.A.; Jain, A.K. Dimensionality reduction using genetic algorithms. *IEEE Trans. Evol. Comput.* **2000**, *4*, 164–171. [\[CrossRef\]](#)
34. Ghamisi, P.; Benediktsson, J.A. Feature selection based on hybridization of genetic algorithm and particle swarm optimization. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 309–313. [\[CrossRef\]](#)
35. Beasley, D.; Bull, D.R. An overview of genetic algorithms: Part 2, Research Topics. *Univ. Comput.* **1993**, *15*, 170–181.
36. Deb, K. *Optimization for Engineering Design: Algorithms and Examples*; PHI Learning Pvt. Ltd.: New Delhi, India, 2012.
37. Maini, T.; Mishra, R.K.; Singh, D. Optimal feature selection using elitist genetic algorithm. In Proceedings of the 2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI), Kanpur, India, 14–17 December 2015; pp. 1–5.

