Electric Price Forecast using Interbreed Approach of Linear Regression and SVM

Deepak Saini, Akash Saxena
Departement of Electrical Engineering
Swami Keshvanand Institute of Technology Management & Gramothan, Jaipur, India
*Corresponding author, e-mail: deepak.92@outlook.com, akash@skit.ac.in

Abstract

Electricity price forecasting is a hypercritical issue due to the involvement of consumers and producers in electricity markets. Price forecasting plays an important role in planning and managing economic operations related with the electrical power (bidding, trading) and other decisions related with load shedding and generation rescheduling. It is also useful for optimization in electrical energy trade. This paper explores an interbreed technique based on Support Vector Machine (SVM) and linear regression to predict the day ahead electricity price using historical data as a raw insert. Different 27 linear regression models are formed to create initial framework for forecasting engine. Comparison of the performance of different forecasting engines is carried out on the basis of error indices namely Mean Square Error (MSE), Sum Square Error (SSE) and other conventional error indices. A detailed explanation of linear regression system based model is presented and simulation results exhibit that the proposed learning method is able to forecast electricity price in an effective manner.

Keywords*: Historical price data, linear regression model, Price Forecasting, Support Vector Machine*

1. Introduction

In modern power market, price of electricity plays an important role. With the deregulation scenario, increasing population and addition of smart grid technology, the competitive business environment has emerged. Hence, the electricity price is an important denominator. Prediction of electricity price is quite troublesome due to its nonlinear, non-stationary and volatile nature. There is an acute need of supervised learning paradigm which not only can predict the price of the electricity accurately but also is easy to implement. Literature survey shows that many researchers have come forward with new supervised learning techniques, time series modeling techniques and regression methods in order to find the price of electricity with higher accuracy. In [1] authors employed similar days method to forecast day ahead electricity price for PJM electricity markets. In this method the correlation between price and load is formulated and further the Neural Network (NN) is employed. The statistical approach for interval forecasting is employed in [2]. Forecasting the prediction interval is essential for understanding the uncertainty involved in the price. Point and interval of prediction forecasting is performed with Functional Principal Component Analysis (FPCA) in approach [3]. Modified Relief Algorithm is employed with the hybrid neural network by N. Amjady et.al. [4]. Radial Basis Function Neural Network (BRFNN) based on fuzzy means and differential evolution is tested for electricity price forecasting [5]. After a careful investigation of the literature, it is concluded that the supervised learning models namely Feed Forward Neural Network (FFNN), RBFNN and SVM are employed as regression agents to forecast the electricity price in different regions. The ability of the neural networks to deal with the variable mapping problems is quite good and the results presented in previous approaches are promising. However, the computational time and determination of optimal set of parameters for deciding the macro and micro structure of the neural networks raise a question on the reliability of the approaches. To address this problem, this paper presents linear regression based approach for forecasting the electricity prices. 27 different linear regression models from the historical data of electricity price are carried out to predict the price of year 2010. The data for this regression modeling is taken from Spain Electricity markets (Iberian Energy Derivatives Exchange) [6]. Prediction by the regression models are put forward as an input features of

Received March 4, 2016; Revised May 9, 2016; Accepted May 20, 2016

Support Vector Machine (SVM). The comparison of the performance of various supervised learning models namely FFNN, Nonlinear Auto Regressive Exogenous (NARX) and Probabilistic Neural Network (PNN) on the basis of error indices is evaluated. The definition of the error indices are taken from [7]. The remaining part of this paper is organized as follows: in section 2 methodology for regression is explained, in section 3, a hybrid model for price forecast by SVM is explained, in section 4 simulated results are exhibited. Section 5 summarizes the results in a conclusive form.

2. Proposed Method

To develop different regression models from the past data is a tedious data mining process. In this section we explain the methodology adopted for the formation of the regression patterns. The data from Iberian stock markets are chosen for carrying out the price prediction of year 2010. The electricity price pattern of a year 2009, trend of price on New Year and a working day are exhibited in figure 1 to 3.

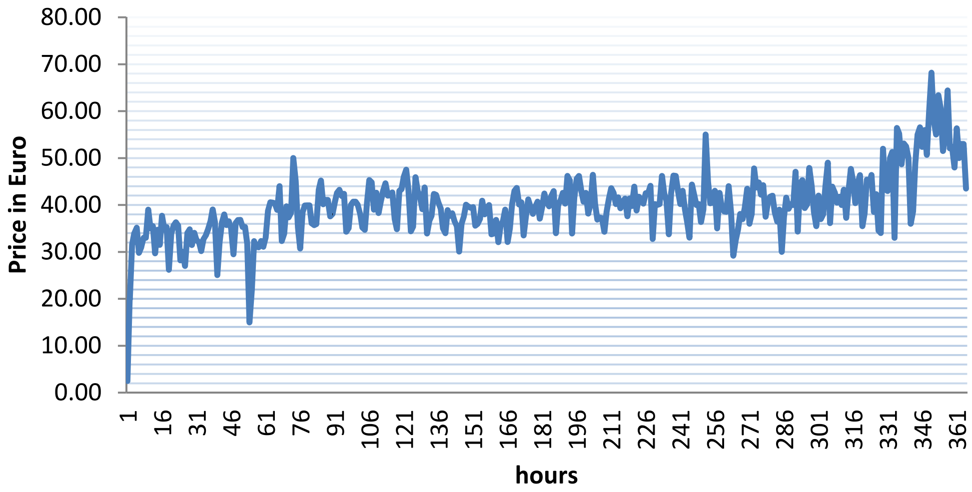

Figure 1. Electricity price pattern of year 2009

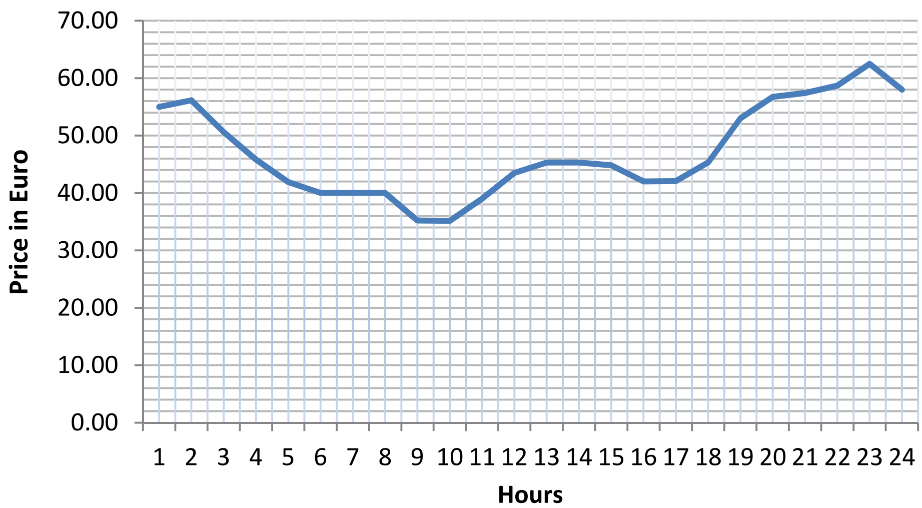

Figure 2. Electricity price pattern of 1-1-2009

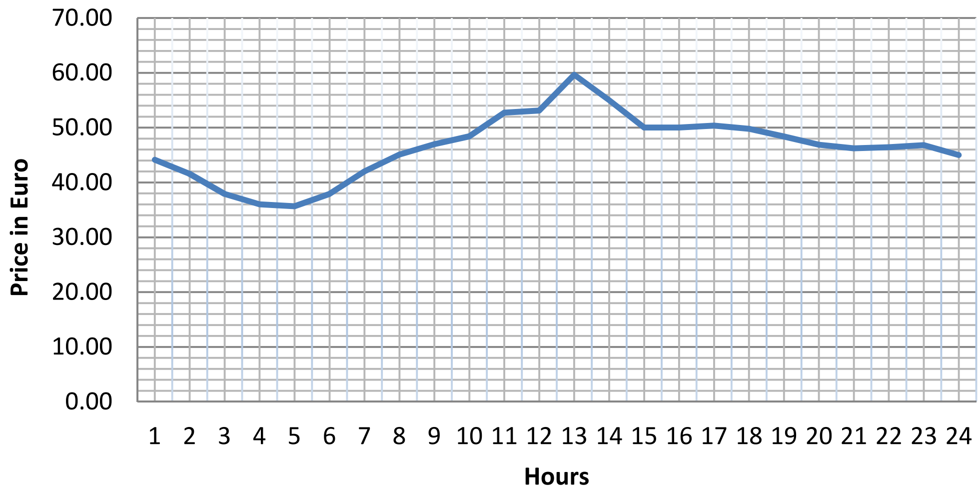

Figure 3. Electricity price pattern for a working day

To develop the regression model the current price is considered by the abbreviation (DMYH). The table 1 shows different variables and their combinations. For example, design 8 is employed to forecast the data of any day of 2010 of specific hour.

Table 1. Definitions of different historical patterns for factorial design

1	(D-1,M,Y,H)A	(D,M-1,Y,H)B	(D,M,Y-1,H)C
2	(D-1,M,Y,H-1)D	(D,M-1,Y,H-1)E	(D,M,Y-1,H-1)F
3	(D-1,M,Y,H+1)G	(D,M-1,Y,H+1)H	(D,M,Y-1,H+1)I
4	(D-1,M,Y,H)A	(D-1,M,Y,H-1)D	(D,M-1,Y,H-1)E
5	(D,M,Y-1,H-1)F	(D,M-1,Y,H-1)E	(D-1,M,Y,H)A
6	(D,M-1,Y,H) B	(D-1,M,Y,H-1)D	(D,M,Y-1,H-1)F
7	(D-1,M,Y,H-1)D	(D,M-1,Y,H)B	(D,M-1,Y,H-1)E
8	(D,M,Y-1,H)C	(D,M-1,Y,H-1)E	(D-1,M,Y,H-1)D
9	(D,M,Y-1,H-1)F	(D,M,Y-1,H)C	(D,M-1,Y,H-1)E
10	(D-1,M,Y,H+1)G	(D-1,M,Y,H) A	(D,M,Y-1,H+1) I
11	(D,M-1,Y,H)B	(D,M-1,Y,H+1)H	(D-1,M,Y,H+1)G
12	(D,M,Y-1,H)C	(D-1,M,Y,H+1)G	(D,M-1,Y,H+1)H
13	(D-1,M,Y,H-1)D	(D,M-1,Y,H)B	(D,M,Y-1,H) C
14	(D-1,M,Y,H)A	(D,M-1,Y,H-1) E	(D,M-1,Y,H)B
15	(D,M,Y-1,H-1) F	(D,M-1,Y,H)B	(D,M,Y-1,H)C
16	(D,M-1,Y,H-1)E	(D-1,M,Y,H+1)G	(D,M,Y-1,H+1) I
17	(D,M,Y-1,H+1)I	(D,M,Y-1,H-1)F	(D,M-1,Y,H+1)H
18	(D-1,M,Y,H+1)G	(D-1,M,Y,H)A	(D,M-1,Y,H)B
19	(D,M,Y-1,H)C	(D-1,M,Y,H+1)G	(D-1,M,Y,H)A
20	(D-1,M,Y,H)A	(D,M-1,Y,H)B	(D-1,M,Y,H-1)D
21	(D,M-1,Y,H+1)H	(D,M,Y-1,H)C	(D-1,M,Y,H)A
22	(D-1,M,Y,H-1)D	(D,M-1,Y,H+1)H	(D,M,Y-1,H-1)F
23	(D,M-1,Y,H-1)E	(D,M,Y-1,H-1)F	(D,M-1,Y,H+1)H
24	(D,M,Y-1,H-1) F	(D,M,Y-1,H+1)I	(D,M,Y-1,H)C
25	(D,M,Y-1,H+1)I	(D-1,M,Y,H-1)D	(D,M-1,Y,H-1) E
26	(D-1,M,Y,H-1)D	(D,M-1,Y,H)C	(D,M,Y-1,H+1) I
27	(D,M,Y-1,H-1)F	(D,M-1,Y,H+1)H	(D,M,Y-1,H)C

This design regresses three variables at a time, the data of the same day, same month, previous year and same hour considered as variable (C), same day, previous month, same year and previous hour as variable (E) and previous day, same month, same year and previous hour as variable (D). Variable C, E and D are employed to predict the price of the same day, same month, same year and same hour. Similarly, in this fashion, 27 different linear regression models are prepared and comparative analysis of these patterns is carried out on the basis of Mean Square Error (MSE). The error in prediction is shown in figure 4.

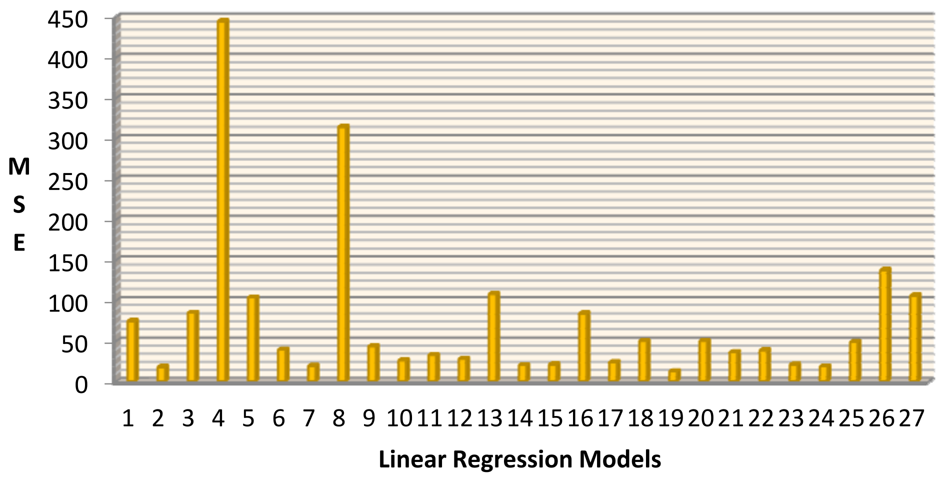

Figure 4. Linear regression model for historical electric price

It is observed that pattern 10, 11 and 19 give lowest MSE. For prediction of the price, these forecast inputs are utilized as input features of SVM.

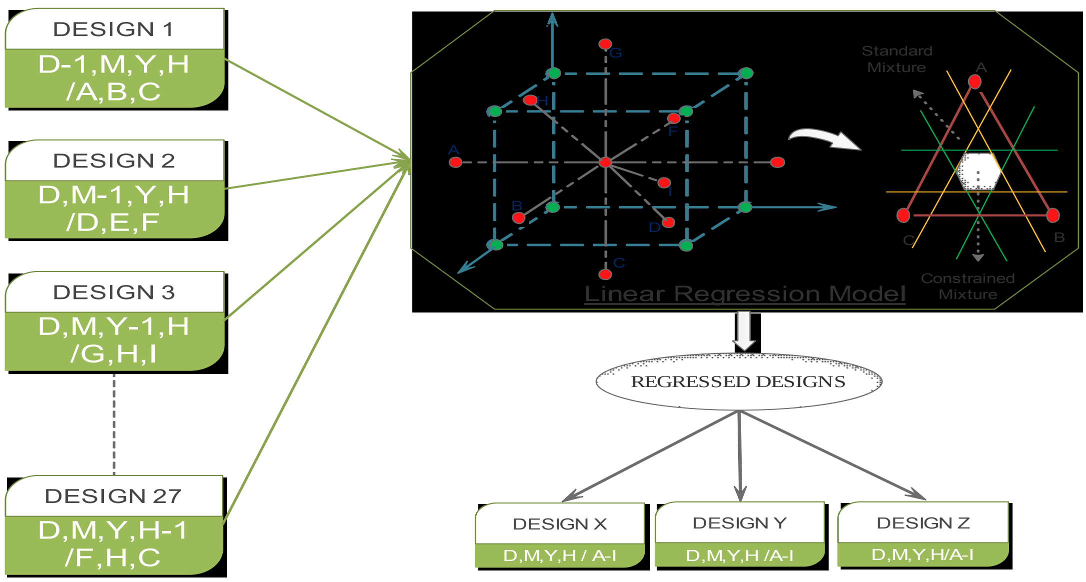

Figure 5. Linear Regression model for Price Forecasting

Figure 5 shows the linear regression modeling between historical data sets and formation of different data patterns (A to I). By the combination of these models with their specific attributes in terms of day, month, year and hour, forecasts are predicted.The mathematical relationship between these variables is shown in appendix.

3. Support Vector Machine (SVM)

In recent years the application of SVMs has increased in nonlinear mapping and classification problems due to its extra ordinary capacity of data matching and regression. Recently a comparison of the performance of the SVM with other relevant neural network topologies has been carried out for contingency ranking and classification in [8]. SVMs are applied for classification of power quality events [9], multi-dimensional data classification [10], classification of microarrays [11], wind speed prediction [12], voltage stability monitoring [13] and many more [14]. The main reason behind this popularity of the SVM as a classifier is that SVM can handle large feature space. Structure of Least Square Support Vector Machine is shown in the figure 6. In this paper Radial Basis Function kernel is used. The choice of RBF is obvious due to its higher accuracy. To design the SVM data of electricity price is sub divided into five subsets. Cross validation technique is employed to train, test and validate the model. To train the supervised learning model electricity price data has been taken for 2006-2009 from the Iberian Market (Spain Electricity Markets) and prediction of the hourly load of year 2010 is carried out. Supervised learning models use 70 % data for training and remaining 30% for testing and validation purpose.

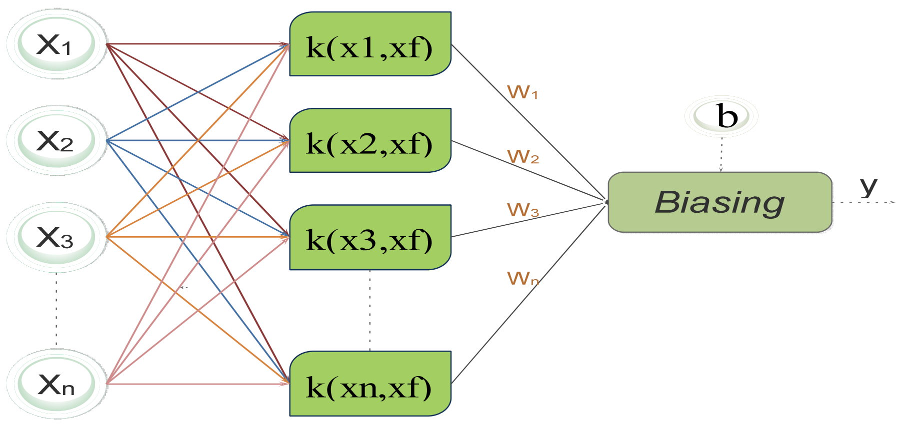

Figure 6. Support Vector Machine

4. Results and Analysis

This section presents the analysis of electricity price forecast derived by the supervised learning method using linear regression model. To draw a fair comparison of the prediction performance by SVM, three other networks are chosen which are FFNN, NARX and Probabilistic Neural Network (PNN). The linear regression model consists of 27 different design sets of merge historical data of Spain electricity price market. This model provides three resultant optimal design sets which are less erroneous and give an enormous relation between raw historical inserts to predicted data set.

Table 2. Comparison of the Network Topologies

Error Indices	FFNN	NARX	Probabilistic	SVM
Mean Square Error	0.0062	0.0059	0.4049	0.0009
Root Mean Square Error	0.0878	0.0768	0.6300	0.030
Mean Absolute Error	0.0605	0.0585	0.6233	0.0155
Sum of Squared Error	4.5396	4.3743	301.24	0.6765
Sum of Absolute Error	55.027	43.49	463.70	11.5680

Following points are emerged from table 2:

a. After careful observations it is concluded that SVM has low values of errors that advocate the efficacy of the proposed approach in the prediction of the electricity price. These errors are plotted in figure 7 and 8. The obvious choice for solving this forecasting problem is SVM due to lower values of errors.

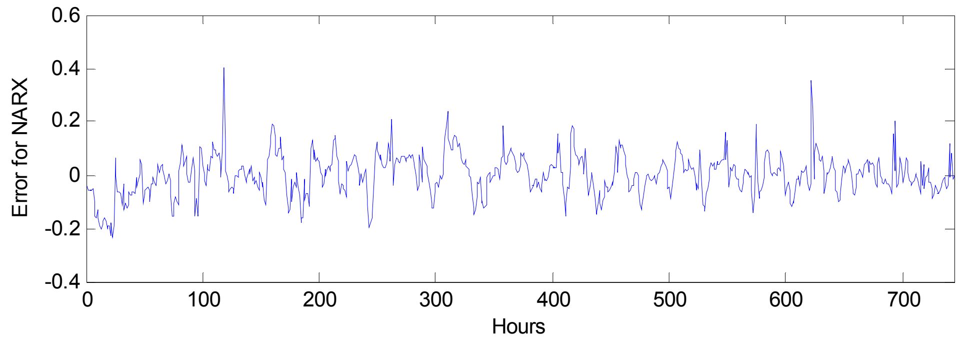

Figure 7. Nonlinear Auto Regressive Exogenous (NARX) Error

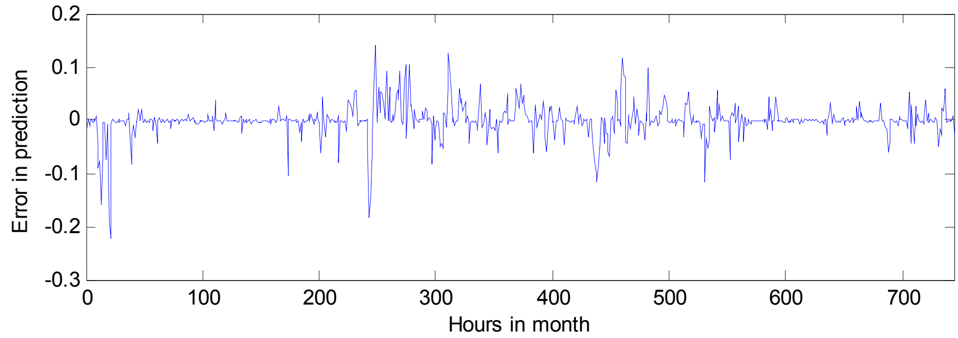

Figure 8. Support Vector Machine Error

b. The error indices are shown in figure 9 and 10. It is also concluded that PNN is not a good choice for prediction of the electricity price. In this particular problem, this topology not only gives a weak performance of the prediction task but also requires high computational time. The prediction of electricity price is a critical task hence, any approach which requires high computational time is not suitable for prediction.

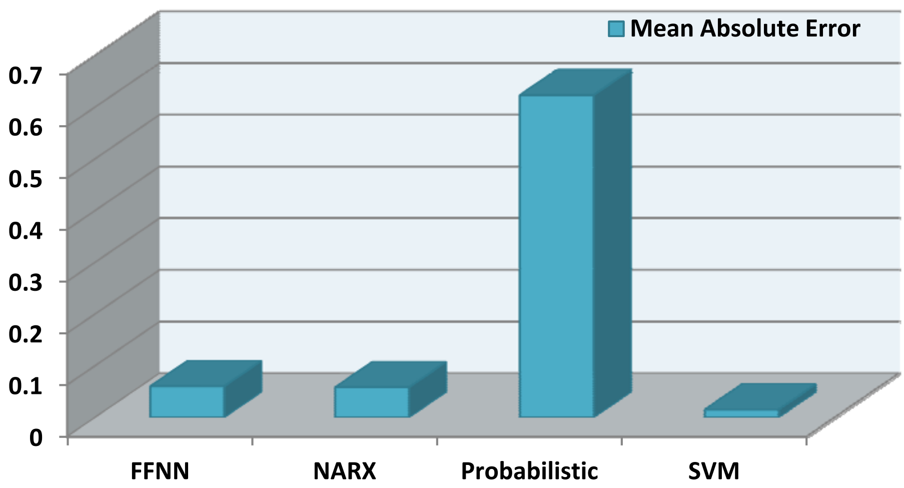

Figure 9. Mean absolute Error for prediction models

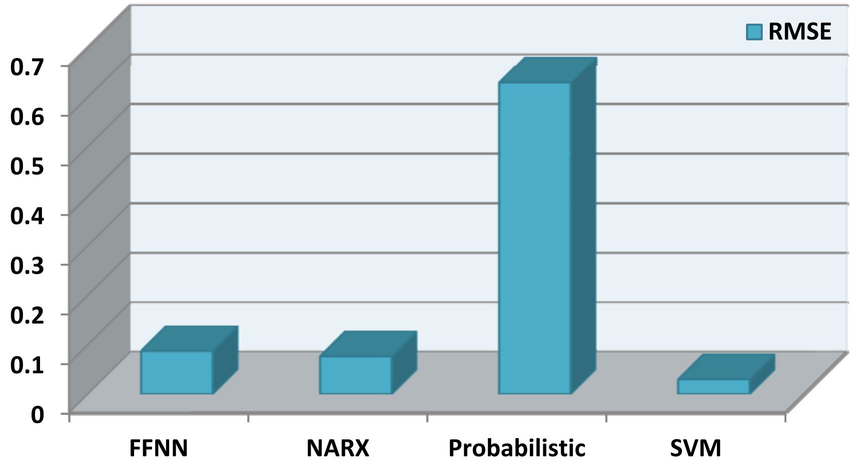

Figure 10. Root Mean Squared Error for prediction models

c. These simulation results show a futuristic solution for different sections of a complex power system network. The error plots of January month of 2010 prediction is shown in figure 10, it is observed that FFNN and NARX performs marginally well and show almost same response in terms of prediction errors. The lower values are observed in the prediction by SVM. This shows the efficacy of this supervised learning model over rest of the conventional topologies.

5. Conclusion

In the competitive business scenario, an accurate prediction of the electricity price is inevitable. Electricity as a commodity can't be stored and stockpiled. Trading of electricity has come in practice in the recent years.This paper presents a straight forward approach for establishing different regression models between historical data sets to the predicted data sets. These models are tested and the finest models are taken for the prediction. Predicted outputs are employed as an input features to SVM and the forecasting of hourly electricity price is carried out. It is observed that the proposed method is efficient and can be a beneficial tool at energy management center. To establish the efficacy of the proposed approach a fair comparison between the performances of networks are carried out. Long term forecasting of electricity price by this method lies in the scope of the future.

References

[1] P Mandal, T Senjyu, N Urasaki, T Funabashi and AK Srivastava. "A Novel Approach to Forecast Electricity Price for PJM Using Neural Network and Similar Days Method". in *IEEE Transactions on Power Systems*. 2007; 22(4): 2058-2065.
[2] JH Zhao, ZY Dong, Z Xu and KP Wong. "A Statistical Approach for Interval Forecasting of the Electricity Price". in *IEEE Transactions on Power Systems*. 2008; 23(2): 267-276.
[3] HC Wu, SC Chan, KM Tsui and Y Hou. "A New Recursive Dynamic Factor Analysis for Point and Interval Forecast of Electricity Price". in *IEEE Transactions on Power Systems*. 2013; 28(3): 2352-2365.
[4] N Amjady, A Daraeepour and F Keynia. "Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network". in *IET Generation, Transmission & Distribution*. 2010; 4(3): 432-444.
[5] K Meng, ZY Dong and KP Wong. "Self-adaptive radial basis function neural network for short-term electricity price forecasting". in *IET Generation, Transmission & Distribution*. 2009 3(4): 325-335.
[6] OMIP The Iberian Energy Derivatives Exchange, http://www.omip.pt/
[7] Nikita Mittal and AkashSaxena. "Layer Recurrent Neural Network based Power System Load Forecasting". *Telekomonika Indonesian Journal of Electrical Engineering. TELKOMNIKA*. 2015: 16(3): 423 – 430.

[8] Bhanu Pratap Soni, Akash Saxena and Vikas Gupta. "A least square support vector machine approach for contingency classification and ranking in a large power system". *Cogent Engineering*. 2016; 3: 1137201.
[9] D De Yong, S Bhowmik, F Magnago. An effective Power Quality classifier using Wavelet Transform and Support Vector Machines. *Expert Systems with Applications*. 2015; 42(15–16): 6075-6081.
[10] Bogusław Cyganek, Bartosz Krawczyk, Michał Woźniak. Multidimensional data classification with chordal distance based kernel and Support Vector Machines. *Engineering Applications of Artificial Intelligence*. 2015; 46(A): 10-22.
[11] M Kumar, S Kumar Rath. *Classification of microarray using Map Reduce based proximal support vector machine classifier*. Knowl. Based Syst. 2015.
[12] Xiaobing Kong, Xiangjie Liu, Ruifeng Shi, Kwang Y Lee, Wind speed prediction using reduced support vector machines with feature selection. *Neuro computing*. 2015; 169: 449-456.
[13] KS Sajan, Vishal Kumar, Barjeev Tyagi. "Genetic algorithm based support vector machine for on-line voltage stability monitoring". *International Journal of Electrical Power & Energy Systems*. 2015; 73: 200-208.
[14] Feng Lv, Fengning Kang, Hao Sun. The Predictive Method of Power Load Based on SVM. *Telekomonika Indonesian Journal of Electrical Engineering*. 2014; 12(4): 3068-3077.

Appendix

DESIGNS	REGRESSION EQUATIONS
DESGIN 1.	R = -4783 + 77.8 A + 81.0 B + 176.1 C - 1.296 A*B - 2.82 A*C - 2.96 B*C+ 0.0472 A*B*C
DESGIN 2.	R = -1109 + 17.8 D + 17.8 E + 31.2 F - 0.269 D*E - 0.471 D*F - 0.478 E*F+ 0.0072 D*E*F
DESGIN 3.	R = -1891 + 34.5 G + 34.1 H + 74.2 I - 0.599 G*H - 1.31 G*I - 1.32 H*I+ 0.0231 G*H*I
DESGIN 4.	R = 5565 - 89.1 A - 92.7 D - 91.8 E + 1.50 A*D + 1.47 A*E + 1.56 D*E- 0.0249 A*D*E
DESGIN 5.	R = -3145 + 102 F + 51.7 E + 51.6 A - 1.66 F*E - 1.63 F*A - 0.829 E*A + 0.0264 F*E*A
DESGIN 6.	R = -1326 + 23.5 B + 23.8 D + 37.9 F - 0.401 B*D - 0.65 B*F - 0.658 D*F + 0.0112 B*D*F
DESGIN 7.	R = -2201 + 32.6 D + 32.3 B + 36.6 E - 0.458 D*B - 0.529 D*E - 0.522 B*E + 0.00745 D*B*E
DESGIN 8.	R = -2794 + 100.4 C + 44.1 E + 50.3 D - 1.57 C*E - 1.77 C*D - 0.776 E*D + 0.0275 C*E*D
DESGIN 9.	R = -1704 + 60.2 F + 58.1 C + 28.8 E - 1.95 F*C - 0.971 F*E - 0.95 C*E + 0.0315 F*C*E
DESGIN 10.	R = 4008 - 67.0 G - 61.8 A - 146.6 I + 1.049 G*A + 2.51 G*I + 2.26 A*I - 0.0387 G*A*I
DESGIN 11.	R = -3564 + 74.7 B + 45.3 H + 60.7 G - 1.008 B*H - 1.249 B*G - 0.759 H*G + 0.0169 B*H*G
DESGIN 12.	R = -1740 + 61.3 C + 30.6 G + 30.7 H - 1.040 C*G - 1.06 C*H - 0.520 G*H + 0.0180 C*G*H
DESGIN 13.	R = -4595 + 81.0 D + 78.4 B + 166.7 C - 1.362 D*B - 2.89 D*C - 2.82 B*C + 0.0489 D*B*C
DESGIN 14.	R = -4755 + 77.5 A + 83.6 E + 71.5 B - 1.344 A*E - 1.153 A*B - 1.253 E*B + 0.0202 A*E*B
DESGIN 15.	R = 248 + 0.8 F - 2.9 B - 4.6 C - 0.013 F*B - 0.08 F*C+ 0.069 B*C + 0.0014 F*B*C
DESGIN 16.	R = -1380 + 21.9 E + 24.8 H + 51 I - 0.374 E*H - 0.79 E*I - 0.87 H*I + 0.0135 E*H*I
DESGIN 17.	R = 3779 - 132.8 I - 112.1 F - 65.4 H + 3.98 I*F + 2.334 I*H + 1.983 F*H- 0.0702 I*F*H
DESGIN 18.	R = 1324 - 6.8 G - 35.8 A - 24.5 B + 0.374 G*A + 0.18 G*B + 0.643 A*B- 0.0071 G*A*B
DESGIN 19.	R = 2658 - 86.7 C - 46.2 G - 40.4 A + 1.563 C*G + 1.320 C*A + 0.719 G*A- 0.0238 C*G*A
DESGIN 20.	R = 1324 - 35.8 A - 24.5 B - 6.8 D + 0.64 A*B + 0.374 A*D + 0.18 B*D- 0.0071 A*B*D
DESGIN 21.	R = -2240 + 39.7 H + 82.9 C + 38.9 A - 1.44 H*C - 0.668 H*A - 1.395 C*A+ 0.0242 H*C*A
DESGIN 22.	R = 725 - 8.8 D - 12.9 H - 26.8 F + 0.174 D*H + 0.368 D*F + 0.505 H*F- 0.0070 D*H*F
DESGIN 23.	R = 907 - 11.2 E - 28.8 F - 15.8 H + 0.40 E*F + 0.213 E*H + 0.53 F*H- 0.0074 E*F*H
DESGIN 24.	R = 643 - 8.8 F - 20.8 I - 31.0 C + 0.355 F*I + 0.626 F*C + 1.092 I*C- 0.0230 F*I*C
DESGIN 25.	R = -1575 + 60 I + 24.9 D + 24.6 E - 0.91 I*D - 0.91 I*E - 0.372 D*E + 0.0137 I*D*E
DESGIN 26.	R = 991 - 11.4 D - 16.3 C - 47 I + 0.205 D*C + 0.61 D*I + 0.81 C*I - 0.0106 D*C*I
DESGIN 27.	R = 2012 - 57.4 F - 34.1 H - 57.9 C + 1.007 F*H + 1.68 F*C + 1.013 H*C - 0.0296 F*H*C