Forecasting Energy Demands based on Ensemble of Classifiers

Soundaryadevi M, Jayashree L.S
RVS College of Engineering and Technology
Banglore
India
soundarya1705@gmail.com, jayashreecls@yahoo.co.in

ABSTRACT: *Analysis of time series data and accurate prediction of future values are among the most challenging tasks that the data analysts face in many fields. Forecasting of energy demands is very essential because both insufficient and excess energy production may lead to a significant reduction of benefits and high storage costs respectively. In order to discover the regularities in dynamic and non stationary data, improved time series forecasting requires a model that combines multiple prediction models. The Ensemble approach performs better than single learning model and discovers the dynamic patterns in Energy time series data. In this paper, we compare the performance of two different Ensemble learning techniques; Bagging (Bootstrap Aggregating) and stacking in forecasting energy time series data. Stacking technique used in this paper, combines different classifiers like Radial Basis Function (RBF), Multilayer perceptron (MLP) and Support Vector Machine (SVM).*

Keywords: Forecasting, Ensemble, Neural Networks

Received: 25 September 2015, Revised 29 October 2015, Accepted 2 November 2015

1. Introduction

Time series analysis and forecasting plays a significant role in various fields of science like electricity demand forecasting, financial forecasting, weather forecasting, etc., Forecasting is the method of predicting future events based on the historical values. Forecasting problems can be categorized in to short-term forecasting, medium-term forecasting, and long-term forecasting. Short-term forecasting is the process of predicting events with short time periods such as days, weeks, months. Medium-term forecasting involves prediction of future values which can extend up to two years whereas long-term forecasting problems can be extended for many years [3].

Energy demand planning is an important factor for any country's economic development. Forecasting allows decision makers to make right decisions and to get competitive advantage. Improving the accuracy of the forecasting technique is the major issue in time series analysis. Efficient Energy planning requires accurate prediction of future energy demands.

There are several statistical methods for time series forecasting such as ARX (Auto-Regressive Exogenous), ARIMA (Auto-Regressive Integrated Moving Average), ARMA (Auto-Regressive Moving Average), GARCH (Generalized Autoregressive Conditional Heteroskedasticity), Box and Jenkins and Smoothing techniques.

Machine Learning Techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees are also widely used for time series prediction. The forecasting accuracy of each model is different from each other. The selection of a particular model is based on the accuracy needed the computational overhead incurred.In [2] it is reported that ANN provides better forecasting accuracy than statistical techniques such as ARIMA and Multiple Linear Regression (MLR). But the computational complexity of ANN was higher over other two techniques.

Each forecasting method has its own advantage as well as disadvantage in time series prediction. Ensemble methods improve the forecasting accuracy by combining the hypothesis generated from individual base learners for the same training data set [7].

Ensemble learning is a supervised learning, the main objective of which is to improve the performance of classifier or predictor. Model Selection is an important task in machine learning. Accuracy depends on type of classifier used and which realization of the classifier to be chosen. For example, in ANN different initialization of parameters gives different outcomes.

An ensemble model consists of a group of learners which are called as the base learners. The generalization ability of the base learners is usually weaker than an ensemble. Ensemble learning is more advantageous because it can boost weak learners to strong learners which makes highly accurate forecast. Most ensemble methods such as Bagging and Boosting generate a homogeneous base learner by making use of a single learning algorithm. There are also some other methods such as Stacking and Voting which produces a heterogeneous learner using multiple learning algorithms [7].

An Ensemble can be constructed using two different steps. Generation of base learner can be in sequential or parallel manner, while the generation of the base learner has influence on the generation of subsequent learners. In order to get a more accurate ensemble, the base learners should be as more accurate as possible. There are many effective steps for evaluating the accuracy of learners, such as cross-validation, Random Sampling, hold-out test etc.

The advantages of various individual techniques are combined in Ensemble Learning, thereby improving the accuracy of predicted values. Accuracy is also based on the Meta learner to be chosen in the process of creating an Ensemble. This paper compares the performance of the ensemble with that of individual methods in prediction of energy demand.

2. Literature Survey

From literature, a number of pattern recognition techniques are used for time series forecasting. Those techniques include ARIMA and GARCH models, artificial neural networks, fuzzy logic, genetic algorithms and Support Vector Machines (SVM).

Ensemble of multiple classifiers can improve the prediction power rather than using a single classifier. Combining the predicted results of a number of classifiers will significantly improve the accuracy of the prediction algorithm.

Mart´inez-´ Alvarez, proposed the Pattern Sequence-based Forecasting (PSF) algorithm [3], which predicts the future values of a time series based on pattern sequence similarity. PSF produced a better prediction of energy time series compared to other well known techniques.

Karin Kandananond, [6] compared the performance of ANN approach with that of ARIMA and Multiple Linear Regression (MLR) models. ANN model outperforms MLR and ARMIA models in terms of accuracy. They reported that even though accuracy of ANN is higher the computational overhead of ANN is also higher compared to other models.

Wen Shen, Vahan Babushkin,[1] proposed an algorithm Pattern Forecasting Ensemble Model (PFEM) for day ahead Energy Demand Forecasting. They combined various clustering techniques KMeans, SOM, Hierarchical Clustering, Fuzzy C Means, KMediods and applied forecasting models for the clustered output prediction of the day ahead energy demand. The performance of PFEM was proved to better than other models.

Pasapitch Chujai, Nittaya Kerdprasop, [4] compared the performance of ARIMA and ARMA models in forecasting house hold power consumption patterns. ARIMA model is proved to be better for monthly and quarterly analysis and ARMA model performs better for daily and weekly analysis.

Reinaldo C. Garcia, Javier Contreras,[8] forecasted day ahead electricity prices of Spain and California Electrcity markets with GARCH model.

Chitra [5] used an Ensemble of multiple classifiers such as Self Organizing Map, K-Nearest Neighbours and Radial Basis Function for time series prediction. The performance of individual learners with that of Ensemble model is compared using three different data sets.

In this paper, the performances of two different Ensemble learning techniques were compared by combining various classifiers such as RBF, SVM and MLP.

3. Existing Methods

3.1 Radial Basis Function (RBF) Network

A radial basis function network is an ANN in which radial basis functions are used as activation functions. The output from RBF network is a linear combination of radial basis functions of the inputs and parameters of Neurons [5].

Radial basis function networks have a variety of applications, including time series prediction and classification. The network training in RBF can be divided into two different stages: first the weights from the input to hidden layer are determined, and then the weights from the hidden to output layer is found.

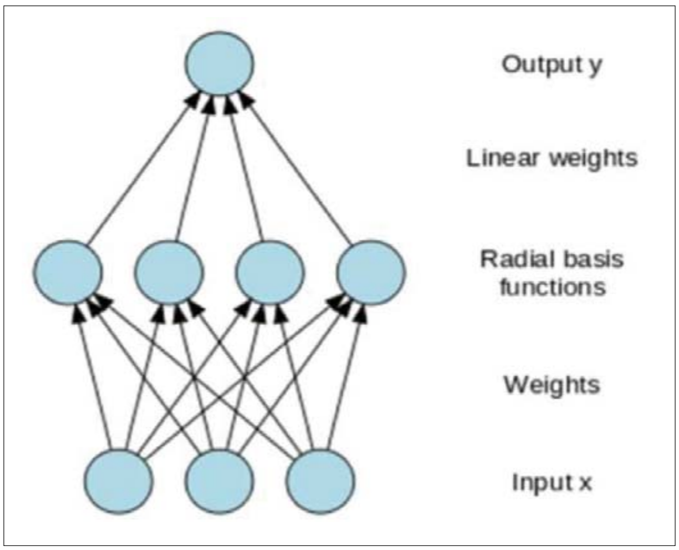

Figure 1. Radial Basis Function Network

Advantage of RBF networks is that it can train extremely faster with a few training samples. RBF also has drawbacks such as difficulty in using Gaussian functions to approximate constant values if a function has nearly constant values in some intervals.

The Prediction error strongly depends on the selection of the number, centres and widths of the radial basis functions which are used as the activation function [16].

3.2 Multilayer Perceptron (MLP)

A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps input data into a set of outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a processing element with a nonlinear activation function [16].

MLP uses back propagation technique for training the network. Back propagation is a supervised learning technique. MLP can distinguish data that is not linearly separable since it is a modification of the standard linear perceptron .

As shown in the Fig 2 multilayer perceptron (MLP) consists of three or more layers, an input and an output layer with one or more hidden layers of nonlinearly-activating nodes.

Each node in one layer connects with a certain weight "*w*" to every node in the following layer.

Multilayer perceptron using a back propagation algorithm is a standard algorithm for any supervised learning pattern recognition process and the subject of ongoing research in computational neuroscience and parallel distributed processing. They are useful in research in terms of their ability to solve problems stochastically, which often allows one to get approximate solutions for extremely complex problems like fitness approximation [13].

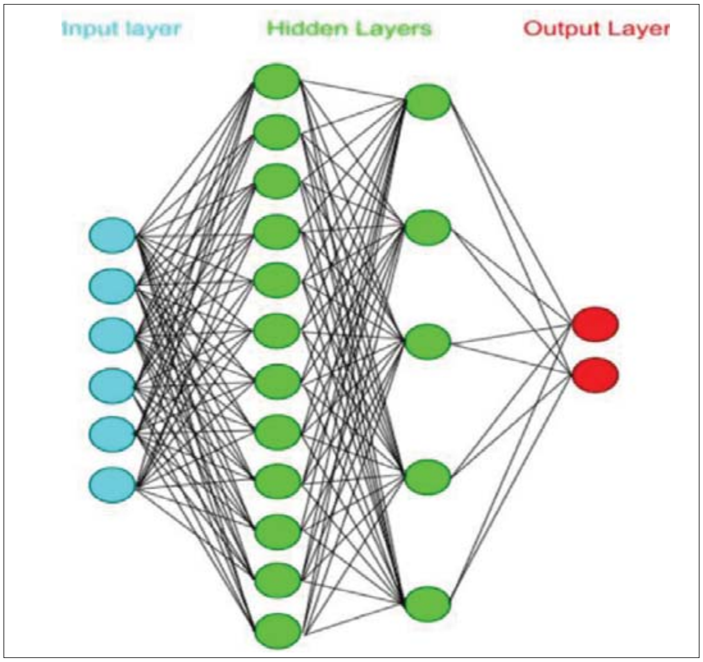

Figure 2. Multilayer Perceptron

3.3 Support Vector Machine (SVM)

Support Vector Machines are most promising methods for both linear and nonlinear data. SVM makes use of a nonlinear mapping to transform original data into a higher dimension. Within the higher dimension, this technique searches for optimal separating hyper plane. A hyper plane is a decision boundary separating tuples of a class from another [15].

As shown in the Figure 3 any training tuple which fall on the hyper plane is called as a Support Vectors.

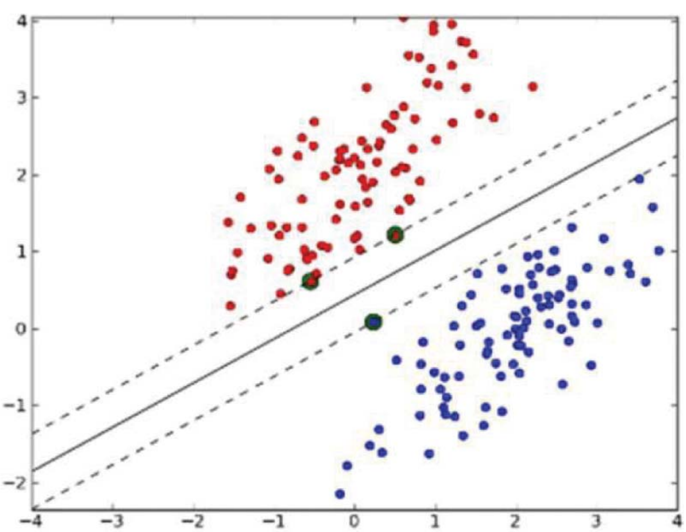

Figure 3. Support Vectors and Hyper plane

Support Vector Machine (SVM) is primarily a classifier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables.

SVM can model complex, real-world problems such as text and image classification, hand-writing recognition, and bioinformatics and bio sequence analysis.

SVM performs well on data sets that have many attributes, even if there are very few cases on which to train the model. There is no upper limit on the number of attributes; the only constraints are those imposed by hardware. Traditional neural nets do not perform well under these circumstances [16].

4. Methodolgy

Several types of classifiers are available for pattern prediction. RBF, MLP, SVM, are considered as the best methods for non-linear time series analysis and prediction.

Ensemble learning is one of the machine learning techniques in which multiple learners are trained to solve a single problem. There are various ensemble methods: Bagging, Boosting, Stacking and voting.

Model Selection is an important factor in Supervised learning. Which model is best suited for the given problem? There are two different ways for interpretation of this question i) what type of model is to be chosen among many competing models, such as multilayer perceptron (MLP), support vector machines (SVM),etc; ii) given a particular prediction algorithm, which realization of this algorithm is to be chosen. ie., different initializations of SVMs gives rise to different decision boundaries, even if all other parameters are kept constant.

The most commonly used technique is to choose the algorithm which gives the smallest error and high accuracy. But it is a flawed one because most of the techniques evaluating the accuracy such as cross validation may be misleading because it provides less accuracy for previously unseen data.

So Ensemble learning is the best alternative for this model selection problem. The following section describes two different Ensemble learning techniques Stacking and Bagging respectively.

4.1 Stacking

The block diagram of Stacking is shown in the following Figure 5. The ensemble process is done with the following four steps: Data Processing, Applying Individual Learner, and Classification of the new data set with the Meta Learner and Evaluation of predicted output. It consists of three different classifiers Radial Basis Function Neural networks, Multilayer perceptron and Support Vector Machines.

In the Stacking algorithm, a number of first-level individual learners are generated from the training data set D by employing different first level learning algorithms $L1$, I2,...L_T.

Those individual learners are then combined by a second-level learner '*L*' which is called as meta-learner.

Ensemble Learners are superior to single learners because of various reasons. The training data might not contain sufficient inputs for choosing a single learner. The search process of theEnsemble learners is better than individual learners. The hypothesis space might not have actual target function. Ensemble learners are used where ever machine learning techniques can be applied.

Algorithm:

Input: Data Set *D*
First Level Learning algorithms RBF, MLP, SVM
Second Level Learning algorithm *L*

Process

1. Train first-level individual learners h_1, h_2, h_3 by applying the first-level learning algorithms RBF, MLP and SVM to the original data set *D*.
2. Generate a new data set D'
3. Apply Individual learners h_1, h_2, h_3 to train the data set D
4. Assign the result of first level learning to the new data set D'.
5. Train the second level learner h' by applying second level learning algorithm to the new data set D'.

Output

Output of Stacking is the final predicted output of two level Ensemble.

The above algorithm gives steps of the two level ensemble approaches with the set of base learners and the Meta learner. As a first step the input data is preprocessed and given to the set of base learners. Finally the resultant data set is trained with the Meta learner. Hence it uses the two level heterogeneous ensemble approach, the generalization ability of the ensemble is much higher than that of individual learners.

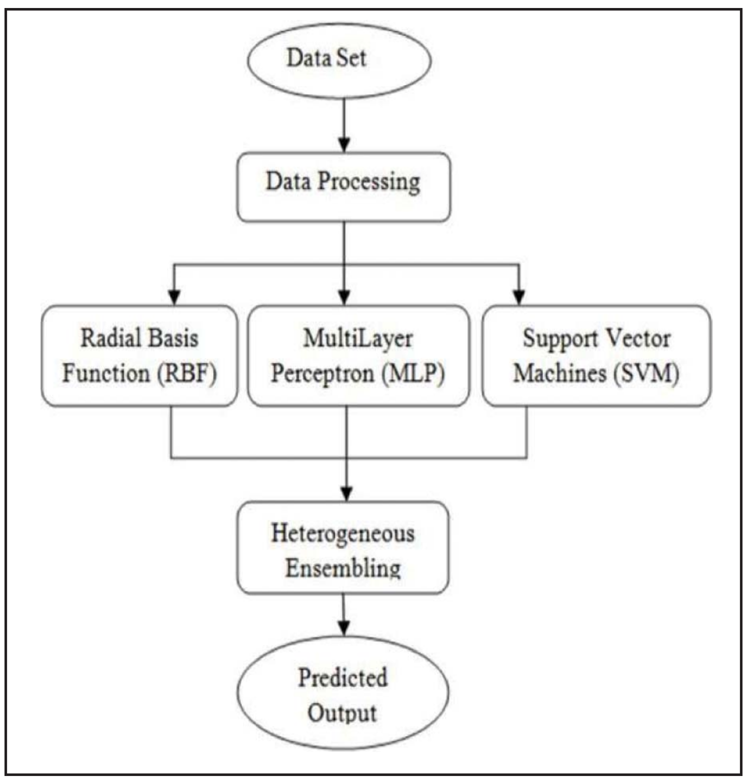

Figure 4. Stacking of Classifiers

Bagging

Bagging trains various base learners with the help of bootstrap samples obtained from the training data set. Bootstrap sample is obtained from creating subsamples from the original data set with replacement. The size of the bootstrap sample is same as that of the training set.

Bootstrap samples are generated from the given training set as random samples. So there is a possibility of choosing the same tuples again.

For instance, the given data contains *d* tuples, the data set is sampled *d* times there by generating *d* bootstrap samples. Therefore, each tuple has a probability of 1/*d* to be chosen and has the probability of 1-1/d not to be chosen [15].

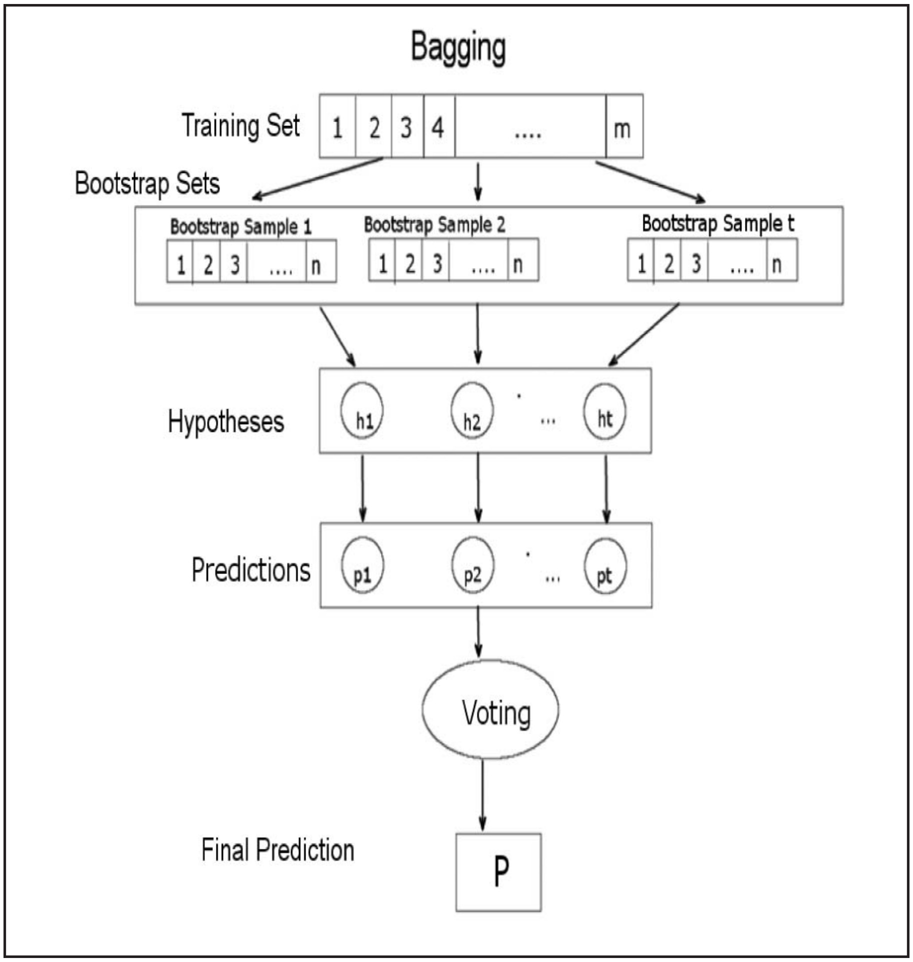

Figure 5. Bagging of Classifiers

Algorithm:

Input: Data Set *D*
Base Learning Algorithm L
Number of rounds I

Process

For *i* = 1, 2, … I
1. Create Bootstrap sample D_i from the data set *D*.
2. Train the base learner h_i with the bootstrap sample D_i

End

Output
Output of bagging is the average of predictions from the base learner.

Like in Voting Ensemble Learning, for numeric prediction, predictions of base learners are averaged. For classification problems, predictions of base learners are taken majority voting.

5. Results and Discussions

The data set was obtained from US department of Energy. The data set contains the monthly energy consumption data from 1973 January to 2013 October.

The data set is available at the following site http://www.eia.gov/totalenergy/data/monthly/index.cfm#consumption. The performance of the above algorithms is evaluated with the following metrics.

Performance Metrics

The following are various metrics that are used for evaluating the performance of the classifiers with that of Ensemble Model.

Mean Absolute Error (MAE)

It measures how the forecasted value is closer to the eventual outcome [5]..

$$\mathrm{MAE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{p_i} - \mathrm{a_i}| \tag{1}$$

Where p_i and a_i is the predicted value and is the actual value observed during the time *i*.

Root Mean Squared Error (RMSE)

RMSE also called as Root Mean Squared Deviation(RMSD)

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{p_i} - \mathrm{a_i}|^2} \tag{2}$$

Mean Squared Error (MSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{p_i} - \mathrm{a_i}|^2} \tag{3}$$

Mean Absolute Percentage Error (MAPE)

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (|\mathrm{p_i} - \mathrm{a_i}|/\mathrm{a_i}) \tag{4}$$

A comparison of Prediction Accuracy, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of the algorithms RBF, MLP, SVM and Stacking and bagging is given below.

The data set is evaluated with hold out evaluation technique, where the given data is randomly partitioned in *training* and testing sets. Therefore one third of the data is used as the test set. The data set consists of 490 records in which 327 records are chosen as training set and remaining 163 records are used as test set.

	MAE	RMSE	MSE	MAPE
RBF	0.2854	0.3783	0.1431	3.3788
MLP	0.395	0.499	0.249	4.8151
SVM	0.2153	0.2792	0.0779	2.6585
Stacking	0.2861	0.3477	0.1209	3.4222

Table 1. performace comparision of classifiers with ensemble

RBF Ensemble is the Ensemble obtained as a result of applying bagging to RBF neural network. Similarly SVM Ensemble and MLP Ensemble are the results of bagging of SVM and MLP respectively.

	MAE	RMSE	MSE	MAPE
RBF Ensemble	0.2989	0.3956	0.1565	3.5394
MLP Ensemble	0.2643	0.3563	0.1269	3.1497
SVM Ensemble	0.1864	0.2482	0.0616	2.2767

Table 2. Performace Comparision of Bagging of Different Classifiers

The following chart shows the accuracy of the different algorithms in forecasting of energy time series data.

As the shown in the above figure the prediction accuracy of the Ensemble model is higher than other models. The above Figure 7 shows the comparison of accuracy of classifiers after applying bagging. SVM Ensemble outperforms other techniques.

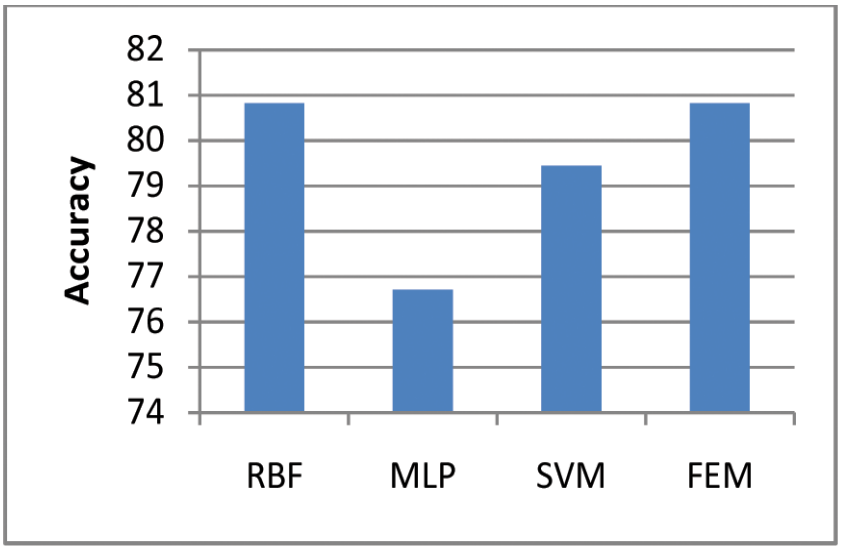

Figure 6. Prediction Accuracy of Classifiers in Time Series Forecasting

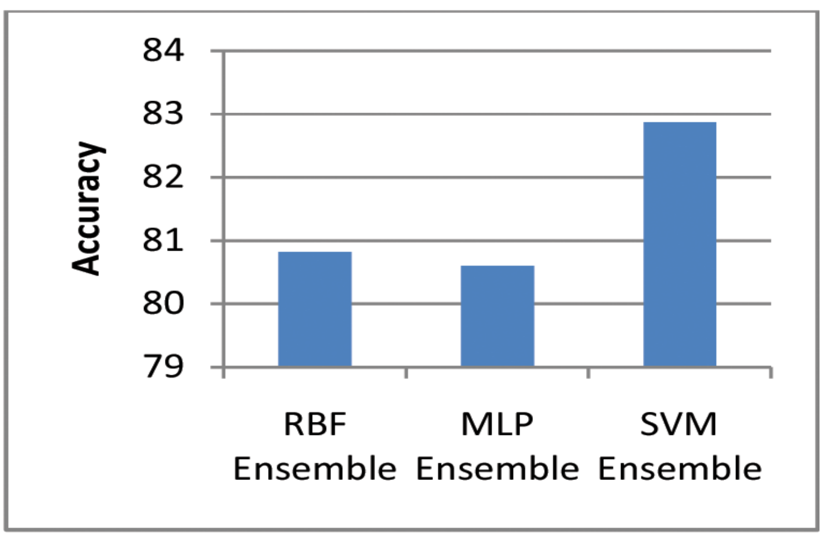

Figure 7. Accuracy of Classifiers after Bagging

6. Conclusion

The Energy demand forecasting is very essential for optimization of Energy resources and application of green trends. The proposed Ensemble model can be applied for prediction of both linear and non linear time series data and be used in wide range of applications in time series forecasting. As the experimental results show the Ensemble model outperforms all other individual learners.

References

[1] Wen, S., Vahan, B . (2013). An Ensemble Model for Day-ahead Electricity Demand Time Series Forecasting, e-Energy'13.

[2] Felix, L., Wolfgang, K. (2012). Classification and Clustering of Electricity Demand Patterns in Industrial Parks, Energies 5, 5215-5228.

[3] Francisco, A., Alicia, T. (2011). Energy Time Series Forecasting Based on Pattern Sequence Similarity. *IEEE Transactions on*, (23), 8, 1230-1243.

[4] Pasapitch, C., Nittaya, K. (2013). Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models. *In*: Proceedings of the International Multi Conference of Engineers and Computer Scientists, 295-300

[5] Chitra, A., Uma, S. (2010). An Ensemble Model of Multiple Classifiers for Time Series Prediction. *International Journal of Computer Theory and Engineering*, 2 (3).

[6] Karin. K. (2011). Forecasting Electricity Demand in Thailand with an Artificial Neural Network Approach. *Energies*, 4, 1246-125.

[7] Zhi-Hua, Z. Ensemble Learning

[8] Reinaldo, C, G., Javier, C. (2005). A GARCH Forecasting Model to Predict Day-AheadElectricity Prices. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 20(2), 867-874.

[9] Time Series forecasting with weka, http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis +and+Forecasting +with+Weka

[10] Weka, http://www.cs.waikato.ac.nz/ml/weka/

[11] Australian Energy Market Operator, http://www.aemo.com.au/Electricity

[12] Han, J., Kamber, M. (2012). Data mining concepts and techniques.

[13] Guoqiang, Z., Eddy, P. (1998). Forecasting with artificial neural networks:The state of the art. *International Journal of Forecasting Elsevier*, 35-62.

[14] Sandy, B., Keith, J. (2000). Automatic neural network modeling for univariate time series. *International Journal of Forecasting Elsevier*, 509-515

[15] Zhuo, Z. (2006). Boosting and Bagging of Neural Networks with Applications to Financial Time Series.

[16] Nesreen, A., Amir. A. An Empirical Comparison of Machine Learning Models for Time Series Forecasting.

[17] Belgin, T., Dilara, D. Electricity Load Forecasting Using Support Vector Machines. 49-53.

[18] Mark, H., Yoan, M. (2009). Adaptive Ensemble Models of Extreme Learning Machines for Time Series Prediction, Springer, 305-314.

[19] Yongning, M. (2012). A Survey of Time Series Prediction Using SVM.

[20] Gianluca, B. (2013). Machine Learning Strategies for Time Series Prediction.