International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

A Hybrid Arima and Discrete Wavelet Transform Model for Predicting the Electricity Consumption of Punjab

Harveen Kaur, Sachin Ahuja

Abstract: ***The significant increase in the world population increases the demand for energy which seems to be alarming for the electricity production boards in the existing time. In the last decade, there are various engineering, simulation tools, and artificial intelligence-based methods such as Support Vector Machine and Artificial Neural Network proposed in the literature to forecast the optimal electricity demand. But these models seldom to work with the linear data. In this paper, a reliable prediction model using the linear time series data of the previous years from January 2013 to December 2017 has been presented to forecast the electricity consumption in Punjab, India. Initially, Discrete Wavelet Transform (DWT) analysis presented to extract the upper and lower limit of the previous year dataand then Auto-Regressive Integrated Moving Average (ARIMA) model has been applied to extract the forecast values. The experimental results compared the original and predicted value using the proposed model to evaluate the effectiveness of the proposed approach. The results show that the difference between the original and proposed modelis only 9% while that of ARIMA only it is 11%. Thus, the proposed model using ARIMA and DWT provides effective results in predicting the forecast value.***

I.INTRODUCTION

Electricity is the main source which satisfies the consumer needs in terms of running various machines by supplying adequate energy to them (Reddy, 2018). The electricity demand is derived from goods and services which is directly linked with the human beings such as production and its usage. However, it seems to be difficult to determine the electricity usage by the consumers and its production. So, various prediction models were implemented in the literature to forecast the energy consumption, crop production, and weather information in various domains such as business processes, marketing, and variations in climatic conditions. The time-series data has been used to predict the output, which may be volatile or non-volatile.

For instance, the speed of the wind is volatile while data attained for global temperature is non-volatile. There are various time series prediction techniques developed in the past studies which entirely depends upon the data. The techniques having linear nature are ARIMA and non-linear models are SVM and ANN (Al-Musaylh et al., 2018; Raza and Khosravi, 2015). ARIMA model forecast based on the past observations and errors. This model has been widely used to forecast the data such as stock index, wheat yield, forecasting wind speed, and electricity prices of the coming days.

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
Harveen Kaur, Department of Computer Science and Engineering, Chitkara University, India
Sachin Ahuja, Department of Computer Science and Engineering, Chitkara University, India

In addition, this prediction model has been widely used for many applications which prove fruitful to predict the future crude oil prices, a load of electricity, electricity consumption and power of the wind. But, there must be some input for these models to predict the future. The time and frequency domain series such as Discrete Wavelet Transform (DWT) compute the output in the form of year, past data, and lag time. The calculated output finally considers as an input for these prediction models. However, analytical methods such as quadratic trend, double exponential techniques, and linear trend etcetera have been used to predict the electricity consumption, and its production in different regions such as India, Pakistan, and Philippines etcetera. However, adequate supply of the energy is a national priority in every state of the country. The issue becomes more severe when electricity energy has been considered rather than other energy sources. In the economic point of view, the demand of the users must be met in a given time. In such a scenario, the prediction of the accurate load seems to be an utmost important in the electrical sector. For instance, when the bearer consumes more than actually required energy then it not only wastes the useful resources but also burdens the expense of cost. The main focus of conducting this study is Punjab, India. This is due to the tremendous use of the fertile land across the border state of the nation. The agriculture growth has been excelled in this side which not only provides growth in the agriculture sector but also develops the energy sector. Thus, Punjab considers as an interesting place to forecast the energy consumption. In the last decade, there are various approaches proposed in the literature to estimate the energy consumption by predicting the electrical energy usage (Hong and Fan, 2016). These methods have been fragmented into two sections such as short terms and long term. However, first is related to the time frame from minutes to hours and the later related to the few weeks to years. But, it is a challenging task to forecast the electrical energy consumption due to the non-linear and multi-dimensional nature of the data. The rapid development in the energy sector has surged the significance of the studies in the power sector. These studies are aggregative due to geographical changes in nature. Currently, the average temperature in Punjab is 25 °C during the summers and 15 °C during the winters. It is estimated that the annual average temperature has been increased by 0.7 to 2.5 °C in 2050. Such a rise in temperature reduces the environmental stability for the energy production. Moreover, natural resources are scarce such as coal mines, wind power, natural gas, and oil are not available in Punjab. These resources transported from the far-away places. In addition, there is not any nuclear plan installed near the border area of the Punjab.

The main power source in this state is thermal and hydro-electricity. The prediction of these sources is vital. Thus, forecasting is a robust approach which helps the decision-makers in making decisions with respect to such temperature variations considering the economic and non-economic perspective with high accuracy. In addition, the determination of the forecast value from the past behavior provides a way to check the effectiveness of the time series model.

A. Discrete Wavelet Transform (DWT) Analysis

DWT is a popular tool which helps the predictors to obtain the meaningful information about the signal in-depth in both time and frequency domain series (Nury et al., 2017). This property of the Wavelet avoids the localization problem of Fourier transforms in the frequency domain. It is a mathematical tool which transfers the input signal into the different domain to process and analyze the signal. This time and frequency domain series is suitable to predict the non-stationary processes in which autocorrelation function and mean are not stable with time. The climatic and financial data both are non-stationary which continuously changes over time. DWT has been used to predict such type of data. Wavelet function has been constructed from a single entity as defined in Eqn. 1.

$$ℶ_{c,d}(t) = \frac{1}{\sqrt{|c|}} ℶ\left(\frac{t-d}{c}\right); c, d \in J; c \neq 0 \qquad (1)$$

In the given equation 4, c is the scaling factor which generally computes the compression value. The parameter d is called the translation parameter, which calculates the wavelet time location. Here condition executes, such as when $|c| < 1$, then wavelet is a compressed version which is related to the higher frequencies (a large number of time-domain cycles). On the other side, if $|c| > 1$, then time width of this function $ℶ_{c,d}(t)$ is larger than the $ℶ(t)$ which directly linked to lower frequencies. Thus, DWT is an important tool used to analyze the time series using wavelets and analogous to discrete. It is based on the coding of a sub-band and fast computation process of the wavelet transform. The implementation of the DWT is simple and easy with reduced resources and computation time. The original dataset (Ap) has been considered where scaling parameter such as 'c' has been signified in the form of 2-p and other parameters' represented as L2-p, where p, L€I. The discrete wavelet function represented as:

$$ℶ_{p,L}(t) = \frac{1}{2^{-p}} ℶ\left(\frac{1 - L2^{-p}}{2^{-p}}\right) \qquad (2)$$

B. ARIMA Model

ARIMA model was implemented by Jenkins by integrating the autoregressive and moving average model, an acronym for ARIMA (Box and Jenkins, 1976). In mathematical form, it is represented as:-

$$z_t = K + \vartheta_1 z_{t-1} + \vartheta_2 z_{t-2} + \cdots + \vartheta_m z_{t-m} + n_t \qquad (3)$$

Where, ϑ_1 represents the intercept term at 1st place, ϑ_2 depicts the intercept term at 2nd place and so on. K is constant and n_t is the white Gaussian noise. Equation 3 in the form of regression in which values are lagged. These values are often used as predictors (Mudelsee, 2014). However, Moving Average (MA) model mathematically represented as given below:-

$$z_t = K + n_t + \varphi_1 n_{t-1} + \varphi_2 n_{t-2} + \cdots + \varphi_r n_{t-r} \qquad (4)$$

Moving average is also a regression in which errors are lagged. The composition of the MA filter help to decompose the time series data (Aradhya et al., (2018)). The MA filter of order 'f' given as:-

$$M_1 = \frac{1}{f} \sum_{u=-i}^{i} s_t + u \qquad (5)$$

Where, f=2i+1. Consequently, the time series has been averaged with i periods of t to acquire the trend cycle. The observational values obtained seems to be near the value which expels the randomness. However, M2 is the detrend section which is acquired given as $M - M_2$. In this equation, M2 represents the detrend part and M1 represents the trend section and M signifies the original series. The output from the MA filter further combined with the autoregression output to constitute the ARIMA model. Mathematically, it is represented as follows:-

$$z_t = K + \vartheta_1 z_{t-1} + \vartheta_2 z_{t-2} + \cdots + \vartheta_m z_{t-m} + \varphi_1 n_{t-1} + \varphi_2 n_{t-2} + \cdots + \varphi_r n_{t-r} + n_t \qquad (6)$$

The parameters of the ARIMA model represented as given below:-

$$\text{ARIMA}(m, d, r) \qquad (7)$$

Where m represents the AR model, r represents the moving average model and d represents the lagging value from the previous years. ARIMA model predicts the value in a time series which is a linear combination of past values. Typically, there are three types of processes such as (a) autoregression process (b) differencing in value (c) moving average part (Taieb et al., 2016). The given ARIMA model has been further used to predict the energy consumption in the coming years in Punjab. However, the developed model is applicable for various applications to forecast the weather information, climatic conditions, and etcetera.

The paper is organized as section 1 illustrates the Introduction with a basic explanation of ARIMA and DWT. Section 2 elaborates the literature review which is followed by the section 3 with the proposed methodology. Experimental Results explained in section 4 and finally concluded in section 5.

II. RELATED WORK

Temperature Variations in Punjab ranges from 16°C to 31°C reduces the stability of the environment which results that electricity production affected. Therefore, it is vital to predicting the electricity consumption to preserve the electricity in the future. These models have been developed by analyzing the previous data which is acquired from Punjab state Electricity board. Hence, various prediction models developed in the literature to predict the electricity usage as per country wise. For instance, scholars attempted to predict the future electricity emission and its generation in Pakistan (Mengal et al., 2019). However, energy production in India is quite different as the growth of the energy sector and the installation of hydro and electricity plants reduces the load.

Retrieval Number: K21240981119/19©BEIESP
DOI: 10.35940/ijitee.K2124.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

But, it also poses a threat to the wastage of such useful resource (Tiwari and Menegaki, 2019). Therefore, researchers use the prediction models to reduce the wastage such as ARIMA model in conjunction with a neural network help to predict the temperature variations. The training and testing dataset of the previous year has been considered to build the model and estimating the performance. In addition, the proposed model has been evaluated to determine the predictive ability of the proposed model. The wavelet-ARIMA and wavelet-ANN approach further compared with each other to check the effectiveness. The desired results obtained but with higher complexity due to variations in data (Nury et al., 2015). Alternatively, scholars employed the different prediction models such as random walk, moving average, ARIMA, deep learning framework, Gaussian processes etcetera to forecast the electricity consumption (Meer et al., 2018; Bedi and Toshniwal, 2019; Ding et al., 2019). The best model has been chosen based on the performance of the prediction model. For instance, Gaussian processes to forecast the electricity consumption in households helps to understand the net demand of the electricity. The covariance functions need to be analyzed to enhance the accuracy. But, the sharper Gaussian processes provide an accurate value with low error. Thus, the best model selected based on the best-fitted value. Specifically, the time series data beneficial to predict the consumption of power in the future. The various time series models such as Linear, Quadratic, S-curve, Exponential, MA, ARIMA and smoothing estimate the power production. The presented research considers two main objectives such as considering the linear, non-linear and time series models and secondly forecast the data using the best-fitted method. The statistical analysis and predicted data using the ARIMA evaluated considering the mean square error, and root means approximation error. The actual and forecast value has been compared considering the lower and upper limits acquired through the wavelet analysis. The results show that ARIMA outperforms the other model in predicting the electricity consumption.

More specifically, scholars propose the hybrid time series model to forecast the energy (Dong et al., 2016). In addition, researchers combine the linear such as ARIMA and non-linear model such as GARCH model to forecast the solar irradiance(David et al., 2016). The performance of the proposed model evaluated considering the MSE, and mean absolute error in which conventional models have been considered to determine the effectiveness of the trend pattern. In addition, the performance has been improved by hybrid the linear and non-linear models, which not only enhances the accuracy but also reduces the overall error (Aradhya et al., 2018). In this paper, an attempt has been made to combine the wavelet transform analysis an ARIMA model to determine the forecasted value.

III.PROPOSED METHODOLOGY

The proposed methodology has been constructed into two phases in which the first phase use the Wavelet model to analyze the time-series data. The output of wavelet analysis further fed as an input to the ARIMA model which is linear time series data of higher and lower range. The dataset has been considered from Punjab State Power Corporation Limited (PSPCL). The flowchart of the proposed methodology shown in Fig. 1.

Fig. 1 illustrates the proposed methodology of the presented research. The time-series data from the electricity board of the previous year considered as an input to the wavelet transform. The basic model of the DWT is given in Fig. 3

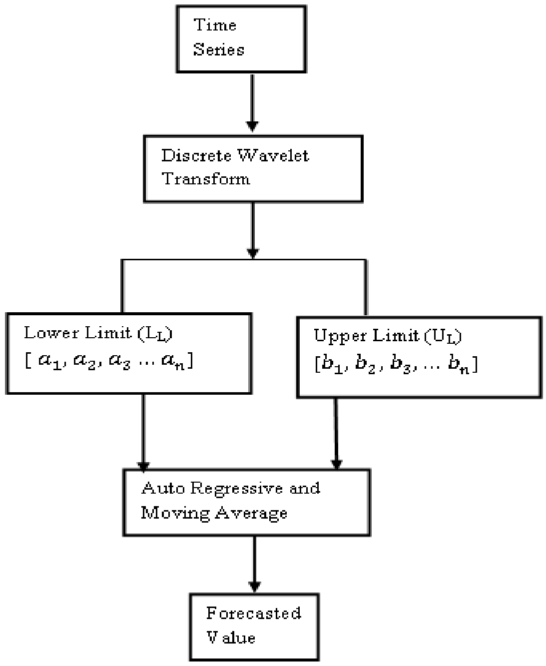

Fig. 1 Proposed Methodology

Fig. 2 depicts the Wavelet analysis of the time series data. The previous year data from the electricity board of Punjab has been considered as an input (Si) signal. The low pass filter passes the lower values of the input and high pass the higher range of the input data. The output from the filters further fed to down sampling. The upper block converts the low values into coefficients such as $a_1, a_2, a_3, \dots \dots a_n$. Similarly, the lower block transforms the higher values into coefficients such as $b_1, b_2, b_3, \dots \dots \dots \dots b_n$. Thus, the lower and upper range of the linear time series data has been arranged in a concise manner. The linear time series data obtained at the output of DWT further considered as an input to the ARIMA model.

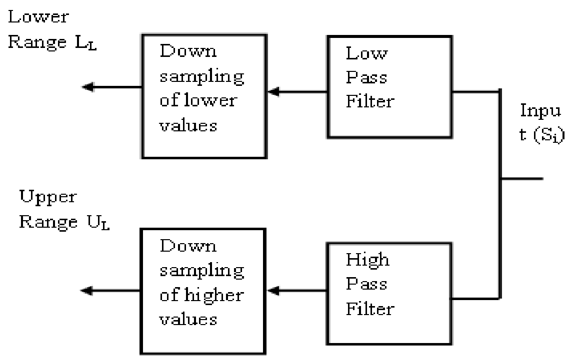

Fig. 2 Discrete Wavelet Transform Output

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The ARIMA model has been applied in the following steps given as:-

Step 1. Time series data consider as an input to the DWT. DWT transforms the input into the upper and lower limit to determine the range of the forecasted value.

Step 2. The upper limit represented by the coefficients $b_1, b_2, b_3, \dots \dots \dots \dots b_n$ and lower limit represented by the coefficients such as $a_1, a_2, a_3, \dots \dots a_n$.

Step 3. The obtained DWT function such as $\beth_{p,L}(t)$ directly linked with the ARIMA model.

Step 4. The regression value has been computed from Eqn. 3 and Moving Average has been computed from Eqn. 4. These equations compute the AR(m) and MA(r). In addition, the Moving Average filter converts the trend part and detrend part from the original series. Eqn. 5 computed the filtered values.

Step 5. Finally, Eqn. 6 combines the MA and regression value to constitute the ARIMA model.

Step 6. The forecasted value has been finally computed.

Step 7. The performance criteria have been defined to determine the applicability of the designed model.

IV.RESULTS

The results have been computed based on the proposed methodology. The electricity consumption original consumed by the consumers and expected has been expressed in the given below graph.

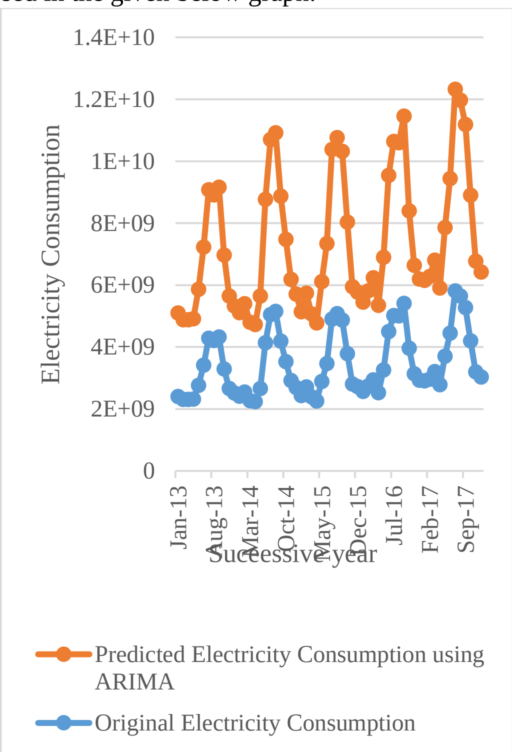

Fig. 3 Electricity Consumption predicted from previous year data

Fig. 3 represents the original electricity consumption from 2013 to 2017which is less than the electricity predicted by the ARIMA model for the same consecutive years. It is clearly seen that the electricity consumed in January 2013 is almost two thousand 40 crore Watts and predicted electricity consumption using the ARIMA model is two thousand sixty-nine crore Watts. Consequently, electricity consumption for the following years from 2013 to 2017 follows the zig-zag path. However, people consume electricity from January 2013 to 2017 ranges from 2402929222 to 3030487191. The predicted electricity consumption using ARIMA model falls in the range of 2691280729 to 3394145654. The actual difference between the original and predicted electricity consumption is around about 288351506.6.

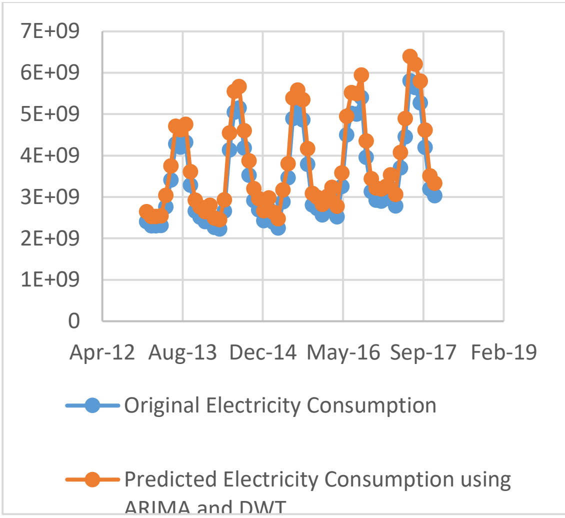

Fig. 4 Electricity Consumption predicted using prediction technique

Fig. 4 clearly shows that there is not much difference between the original electricity consumed by the consumers and predicted electricity consumption. However, the average original electricity consumed by the consumers was 3464988567 approximately. But, predicted electricity consumed by the users using the ARIMA and DWT model will be 3811487423 approx. The original and forecasted value falls in the same range with little difference.

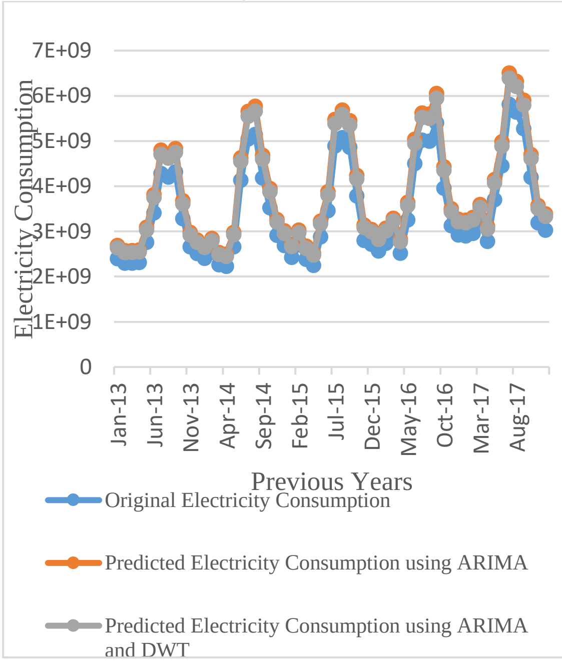

Fig. 5 Comparison of power consumed using prediction technique and original

Retrieval Number: K21240981119/19©BEIESP
DOI: 10.35940/ijitee.K2124.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Fig. 5 clearly depicts the predicted electricity consumption and the original, which follow the same trend. There is not much difference between the original and predicted power consumption by the users. The graph clearly shows that both graphs follow the same trend. The zig-zag path followed in the given graph. However, prediction using the ARIMA and DWT provides better results than the ARIMA only. The average computed value of the original, ARIMA, and ARIMA and DWT is 3464988567, 3880787195, and 3811487423. Thus, the average difference between the original and ARIMA model is 11% and that of the proposed model is 9%. Thus, our proposed model using ARIMA and DWT provides better results than the ARIMA only.

V.CONCLUSION

This paper presents a prediction model using the ARIMA and DWT techniques to forecast the electricity consumed by the consumers. There is a great significance to predict the reliable and error-free

prediction of the power. There are different prediction models proposed in the literature to forecast the efficient and accurate power consumption values. The linear and non-linear models provide an added advantage to predict the temperature, climate, electricity, etc. using the linear and non-linear data. The linear time series data of the previous year from 2013 to 2017 has been considered in this research to predict the value. The original and forecasted value has been compared to evaluate the results. The prediction helps to identify the demand for electricity in the future so that power shortage may be avoided. The experimental set up considering the January 2013 to December 2017 electricity consumption values. The outcomes conclude that there is not much difference between the original and predicted value. Thus, a reliable and efficient prediction model presented in this research to estimate the power consumed by the consumers. The developed model is efficient in predicting the future forecast value of the electricity consumption.

REFERENCES

1. Al-Musaylh, M.S., Deo, R.C., Adamowski, J.F. and Li, Y., 2018. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. Advanced Engineering Informatics, 35, pp.1-16.
2. Nury, A. H., Hasan, K., &Alam, M. J. B. (2017). Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in north eastern Bangladesh, Journal of King Saud University-Science, 29(1), 47–61.
3. Meer, V.D., D.W., Shepero, M., Svensson, A., Widén, J. and Munkhammar, J., 2018. Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes. Applied energy, 213, pp.195-207.
4. Bedi, J. and Toshniwal, D., 2019. Deep learning framework to forecast electricity demand. Applied energy, 238, pp.1312-1326.
5. Ding, S., Hipel, K.W. and Dang, Y.G., 2018. Forecasting China's electricity consumption using a new grey prediction model. Energy, 149, pp.314-328.
6. David M, Ramahatana F, Trombe PJ, Lauret P. 2016. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. Solar Energy;133:55–72.
7. Mengal, A., Mirjat, N.H., Das Walasai, G., Khatri, S.A., Harijan, K. and Uqaili, M.A., 2019. Modeling of Future Electricity Generation and Emissions Assessment for Pakistan. Processes, 7(4), p.212.
8. Aradhye, G., Rao, A. C. S., &Mastan Mohammed, M. D. (2018). A Novel Hybrid Approach for Time Series Data Forecasting Using Moving Average Filter and ARIMA-SVM. Emerging Technologies in Data Mining and Information Security, 369–381.
9. Tiwari, A.K. and Menegaki, A.N., 2019. A time-varying approach on the price elasticity of electricity in India during 1975–2013. Energy.
10. Reddy, B.S., 2018. Economic dynamics and technology diffusion in the Indian power sector. Energy policy, 120, pp.425-435.
11. Hong T, and Fan S. 2016, Probabilistic electric load forecasting: a tutorial review. International Journal of Forecast;32 (3):914–38.
12. RazaMQ, and Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew Sustain Energy Rev2015;50:1352–72.
13. Dong B, Li Z, Rahman SM, Vega R. A hybrid model approach for forecasting future residential electricity consumption. Energy Build 2016;117:341–51.
14. Taieb SB, Huser R, Hyndman RJ, Genton MG. Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. IEEE Trans Smart Grid 2016;7(5):2448–55.

AUTHORS PROFILE

Harveen Kaur is currently working as an assistant professor in Chitkara University. She is pursuing Ph.D in Computer Science and Engineering. Her research interest area is in Data mining, machine learning and Time series analysis.

Dr. Sachin Ahuja is currently working as a Director (Research) in Chitkara University. He has over 10 years of experience in teaching and research at UG and PG levels of Computer Science & Engineering. His research interests include: Data Mining, Education Data Mining, Prediction using Fuzzy Logic and Fuzzy Time series.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)