Analysis of Individual Household Electricity Consumption Forecasting Using ARIMA Model, CNN and LSTM Model

B. N. Shwetha(✉) , R. Sapna , and S. Pravinth Raja

Presidency University, Bengaluru, India
shwethanrupathunga@gmail.com

Abstract. The electricity has a domineering part in residential as well as in industrial area. The electricity consumption has been increasing gradually over decades. The electricity consumption is the actual electricity demand from the power supply. Electricity demand forecasting can be used to plan electricity output. To make efficient use of electricity consumption or to reduce electricity consumption, it is important to understand the future electricity usage. Predicting power demand is imperative in the electric division because it provides a foundation for making decisions about power system planning and operation. The statistical model and deep learning models can forecast the electricity consumption. This paper provides the analysis of electricity forecasting using ARIMA model, CNN model and LSTM model. The paper gives idea, to predict weekly electricity consumption of single house using above models, the comparisons of three models in terms of forecasting and RMSE scores is also presented.

Keywords: Electricity forecasting · CNN model · ARIMA model · LSTM model · Time series forecasting

1 Introduction

Electricity consumption plays a crucial role in both industrial and residential sector. The need of electricity power is mandatory for industry to carry out the necessary activities and also power is mandatory in a house to satisfy the needs of human beings. In residential sector, the electricity consumption refers to the power consumed by electrical appliances at home. The electricity consumption is the actual electricity demand from the power supply. Electricity consumption analysis and forecasting is one of the major challenge. The electric power supplied from the power supply should be efficiently used by all the electrical appliances at home. The efficient use of electricity consumption can improve the economic growth of the country. The demand for electricity has been growing at a compound annual rate of growth of nearly 8% in India from 1999 [1].

The concept of power forecasting has been in presence for quite a long time to predict the upcoming electricity demand. The power utility will be able to make better decisions on load switching and purchasing and generating electric power. This includes the precise estimation of the measure plus geological areas of electric load above the

various times across the planned limit. Electricity foretelling is acritical element in the smart grid. This partakes to draw high rate of scholarly interest. The electrical energy is important in all developing countries; different models were used to forecast turkey's electricity consumption in [2].

The efficient responses and informed decisions of electricity demand can be enabled using forecasting. The 113 different case studies of forecasting models are presented in [3]. The electricity forecasting is the basic step, which gives clarity towards future electricity consumption. The prediction of future electricity consumption can help to know the future electricity demand and it also helps to have proper plan before optimizing power [4]. The load forecasting period of long term could be month or year, medium-term forecasts and for the short-term forecasting period of day or hour [5]. To predict the electricity consumption, for solitary household and group of houses for interval of time. The small term, midterm and extended term prediction covers daily, trimester and 13 months [6]. Time series estimation can be used to predict the load forecasting of electricity usage.

Electricity is supplied to both industrial and residential sectors every day from the power utility. The Electrical Energy Management Side always executes the production and delivery of electric power. The electric power of high reliability and good standards is supplied by the power supply. The main functionalities of power utility are retailing, generation, distribution and transmission of electric power. Water can be reserved in the storage unit and used at a later time. Electric power is the non-storable product. The amount of electricity generated should not be greater than the required electricity demand. The demand of consumers in the areas residential, industrial, commercial keeps varying with respect to time. The demand for electricity is increasing all over the world because of which it is becoming a scarce resource. The preserved electricity should be distributed in an efficient manner to meet the demand for electricity. Electrical energy management (EEM) supports the growing demand. It is a good action and provides enough time for utility companies to invest in new generating power plants [7].

Forecasting includes taking models fit on recorded information and use them to foresee future information. Through the analysis of sequence of time the future events are predicted in Time series forecasting based on assumption that the upcoming events will hold based on past records. The time series forecasting technique is used to recognize models as of the previous data. Presuming that data will bear resemblance in future days, forecasting is made with prevailing data. With the availability of many existing prediction methods with different accuracy, we need to choose the best of them. As we know the accuracy of the methods is generalized by the error as minimum as possible. The applicable prediction approaches are deliberated considering quite a few dynamics like prediction recess, prediction duration, characteristic of time series, and time series scope. The foremostintention of time series model is to gather and judiciously examine the bygone observations of a time series to progress a suitable model which labels the natural assembly of the series.

To construct forecasts, this model is then utilized to generate future values for the series. The concept of forecasting the forthcoming by studying the previous data is known as time series forecasting. Considering time series forecasting important in many practical sectors, such as business, banking, sciences and engineering, etc. Fitting an

appropriate model to the underlying time series should be done with caution. A proper model fitting is required for good time series forecasting [8]. Researchers have put in a lot of time and effort over many years to develop efficient models to increase predicting accuracy. Within the realm of decision-making, time series prediction plays a crucial role. Analysis are made from previous data to construct a mathematically immediate series, in order to try to predict future values with a reasonable margin of error by the decision maker. [9]. The time series data can use forecast algorithms to predict the future data. In this paper, the deep learning models namely, CNNas well as LSTM, statistical model ARIMA are taken to forecast forthcoming electricity usage using time series model.

George Box and Gwilym Jenkins' ARIMA model gives the best fit results. The ARIMA model can also be called as Box-Jenkins models, its main focus is on forecasting using time series analysis. The ARIMA model forecasts future values based on its past observations, which includes lags and lagged forecast errors. It verifies the finest model from a general class of ARIMA models using a 3 step iterative tactic of model identification, parameter approximation, and diagnostic inspection. These steps of the model is repeated n times until a good model is obtained [8, 9, 17].

Convolutional Neural Network is kind of feed-forward neural network. CNN entailslayers like input, hidden and output layers. The input to the layer of convolution are the yield of the preceding layer of convolution or pooling. The CNN contains rare features like pooling layers and fully connected layers. In traditional neural network hidden layers' count were less and the hidden layers' count is more. Its better to have more number of hidden layers to have improved feature extraction and it will be able to recognize or predict values very easily. Convolutional Neural Network can be taken for image classification, pattern recognition and forecast of continuous data. CNN model is used to predict electrical energy usage with the time series data. The week ahead consumption is predicted in [11]. The forecasting accuracy is measured in terms of RMSE score. The rmse score achieved using cnn model is 404.11 kilowatts.

LSTM is a recurrent neural network design that is more sophisticated. Because the activation of a recurrent hidden state in every time phase is reliant on the hidden state of a president time step, RNN can handle time series data, whereas a traditional neural network transfers information to the next layer without regard for the previous time step. The primary idea behind LSTM is to substitute a typical neuron in an RNN through a memory cell made up of three sigmoid layers called input, output, and forget gates. The forget gate determines whether the cell state's value should be retained or dismissed in the first step. The gate generates a ranges from 0 to 1 in this process by referring to the input and the hidden state at time t and t −1, which it dismisses and retains, accordingly. To complete this operation, the resulting value is multiplied by the hidden state at time t −1. The outputs from the next two procedures in the state of the cell are then multiplied to determine freshly created information to be stored. The input gate creates a value to represent the updated state of the cell at time t, and the tanh layer yields a different candidate vector. The vector is utilised to update cell's status in this case. The outputs of these operations are multiplied, and the outcome is then summed with the output of the forget gate. The output gate uses the updated hidden state at time t to translate the memory cell's output. The output gate determines which information to consider as the

output based on the cell's state during the output generation process. Eventually, the memory cell's output is formed by multiplying the vector from the output gate and the cell's value through the tanh layer element-by-element. [18].

The forecasting accuracy of each model can be measured in terms RMSE score, MAPE. Here, RMSE score is used to predict the better performance of the model. The RMSE is the square root of the residuals' variance. It shows how near the observed data points are to the predicted values of the model, indicating the model's absolute fit to the data. RMSE is an absolute amount of fit, however R-squared is a relative amount of fit. Lower the values of RMSE score, which indicates it is a better fit.

2   Related Work

Electricity load estimating is vital in forecasting future demands. Existing papers on load forecasting are mentioned below:

A research work with the application of ARIMA and ARMA model to forecast electricity consumption was proposed [10]. Here, to measure the forecasting method AIC (Akaike Information Criterion) and RMSE (Root Mean Square Error) were used and the smaller value of AIC and RMSE were considered as better forecasting models with good accuracy. The superlative model to forecast monthly and quarterly duration is ARIMA model and The superlative model for determining the best predicting duration in daily and weekly forecasts was that of ARMA.

To forecast the electrical load for all days in a week, the researchers suggested a Deep Convolution Neural Network (DCNN) model [11]. The electrical load readings used on that particular day for the preceding 90 days was evaluated to determine the power usage for one day of the week. The predicting models viz., recurrent neural network, extreme learning machine, CNN and auto regressive integrated moving average performance comparisons was mentioned to show the importance of DCNN model. The proposed DCNN has the lowest mean absolute percentage error, mean absolute error, and root mean square error of 2.1%, 138.771%, and 116.417%, respectively, according to the results.

The researchers also suggested a CNN-LSTM neural network that can obtain spatial and temporal characteristics to foresee dwelling energy use well [12]. Combining a convolutional neural network (CNN) and long short-term memory can obtain the intricate aspects of energy use (LSTM). The CNN layer can obtain features amongst numerous variables that influence energy intake, whereas the LSTM layer is good for modelling temporal data of uneven drifts in time series modules. The CNN-LSTM model can achieve very good prediction compared to previous models, where as it was very difficult to predict consumption. The forecasting accuracy of CNN-LSTM model is measured with root mean square error and it has very small value of root mean square error when in comparison with other predicting model like linear regression and LSTM model.

Another proposal of a hybridized model combining DL with a Convolutional Neural Network (CNN) and an AI-Tuned Support Vector Machine (SVM) fusion was made. This addresses the short-term load predicting test. Deep Learning (DL) is a top methodology for extracting features, learning from filters, and classifying the output [13]. The suggested DL CNN and AI-finely tuned SVM combination can efficiently handle the

electricity consumption time series data's nonlinear intricacies and short-term dependences. The major problem of deep learning is over fitting, which can be relaxed using machine learning. The results show that prediction errors is minimal with 0. 514% of MAE and 0. 688% of RMSE.

A study on ARIMA model for time series prediction to calculate forecast accuracy by means of Mean absolute percentage error (MAPE). The seasonal ARIMA (0, 1, 0) _(2, 0, 1, 12) model turned as the best model for forecasting electricity usage in IIT(ISM) for the years 2004–2008, and thus, anticipate usage during 2008–2009 with a MAPE of 6. 63 percent [14].

The investigation with two models ARIMA model and NAR model was made to forecast electricity consumption for a given period of time [15]. The performance of both the models is compared, both are capable of producing reliable estimates and can be used to forecast energy effectively. The ARIMA model is best compared to NAR model in terms of predictive errors. The forecasting performance of ARIMA RMSE score is 0. 0084 and NAR RMSE score is 0. 0104.

The researchers in their work have proposed Electric Energy Consumption Prediction model coalescing Convolutional Neural Network (CNN) and Bi-directional Long Short-Term Memory (Bi-LSTM) to foresee the usage of electricity consumption. In the first module, two CNNs are utilized to obtain the key values from many variables for the individual home electric power consumption (IHEPC) dataset [16]. The foregoing information, and the trends of time series in forward and backward states, is then used by two Bi-LSTM layers to create predictions. The values collected in the Bi-LSTM module will be transmitted to the final unit, which is made up of two fully connected layers and will estimate future electric energy usage. The experiments were carried out to evaluate the suggested model's prediction performance to that of contemporary models for the IHEPC dataset with numerous modifications. The above model outperforms the consumption prediction on different variants of data set in all terms of real-time, short-range, medium-range and long-range time taken. Even with lot of work in this field, we find there is still much required explorations to be done for betterment of our life with ease [17–20]. There are much more applications of the CNN, ARIMA and LSTM model like in medicine field, educational sector and many more [21–24]. We can integrate the semantic web technologies and its quality aspects to have a better yield [25, 26].

3   System Architecture

3.1   Data Set Description

The electricity consumption of single house. The dataset contains multivariate time series data the power consumption of single house for duration of four years. The dataset contains 2075259 records collected from solitary house situated at Sceaux which is 7km of Paris in France. The records were collected from December 2006 to November 2010 which accounted to 47 months duration. The multivariate data contains seven variables, which is listed below Table 1. Data set was collected from UCL machine learning repository [27] (Fig. 1; Table 1–3).

The data set contains minutely power consumption records of single day, so the total records will be 1440 slots per day [27]. The 47 months (2075259 records) power

Table 1. Data set.

Variable	Description	Units
Global_active_power	The total active power utilized at home	kilowatts
Global_reactive_power	The total reactive power utilized at home	kilowatts
Voltage	Average voltage	Volts
Global intensity	Average current intensity	Amps
Sub_metering_1	Active energy for pantry	watt-hours of active
Sub_metering_2	Active energy for washing clothes	watt-hours of active
Sub_metering_3	Active energy for climate control	watt-hours of active

Table 2. RMSE score of three predictive models

Variable	RMSE score
ARIMA model	381.636 kilowatts
CNN model	404.11 kilowatts
LSTM model n input $= 7$	399.456 kilowatts

Table 3. RMSE score table

Predictive models	MON	TUE	WED	THUR	FRI	SAT
ARIMA model	398.9	357.0	377.2	393.9	306.1	432.2
CNN model	400.6	346.2	388.2	405.5	326.0	502.9
LSTM model	422.1	384.5	395.1	403.9	317.7	441.5

consumption raw data is loaded in CSV format. Data preprocessing converts raw data into understandable format, which is useful for further processing. In Data preprocessing, the data will be passed through processes for cleaning it, in terms of filling in missing values or resolving the inconsistencies in the data. The missing values in the data set is replaced with the power consumption score exactly 24 h before and the date, time columns is combined into single column date_time as part of data cleaning [20]. Data reduction is the process of decreasing the volume or capacity of data. The minutely power consumption can be resampled into hourly, weekly, monthly and yearly power consumption. But here, the power consumption is resampled into total power consumption of each day.
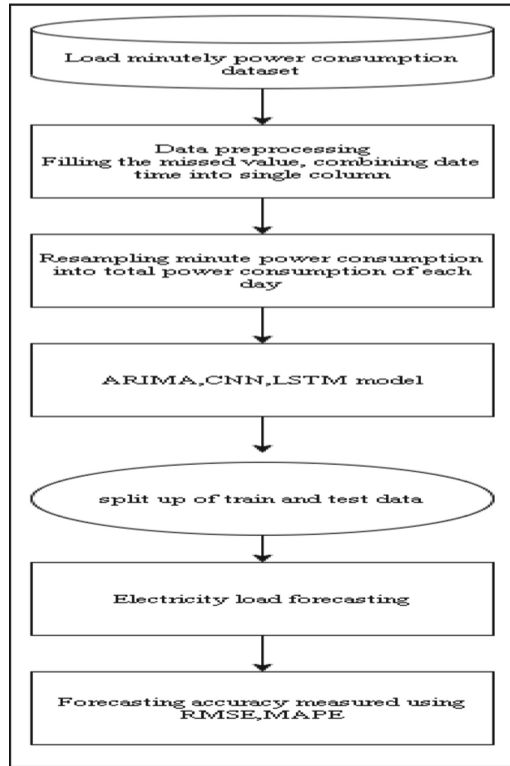
Fig. 1. Steps involved in forecasting schemes

To predict week ahead electricity consumption ARIMA, CNN and LSTM model is used in this paper. The train and test data is obtained from the data set by splitting up information into précised weeks. Considering the data in terms of weeks, full 159 weeks is given for train of prediction model. Finally, 159 weeks is considered for training and 46 weeks is considered for testing. The total active power consumption of each day is considered in terms of daily basis, the electricity consumption of few weeks (159 weeks) is given as training for the model based on the analysis of past observations the expected electric power usage, the week ahead usage should be predicted. The performance of three algorithms are correlated with reference to forecasting accuracy metric Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The error metrics are measured in terms of kilowatts as the total power are also measured in kilowatts. An array of rmse score can be obtained based on actual and predicted electricity consumption per week. The performance of a model is measured with a solitary score, RMSE score for all forecasting for all the three models is enclosed in the Table 1. The forecast consists of seven values, each value represents the days of week ahead. The rmse score for all days in a week is enclosed in Table 2. The actual active power and predicted active power consumption is represented in the form of graph in figure given below.

4   Data Understanding and Visualization

The 4 years electricity consumption is visualized in terms of different graphs. The graph gives the visualization of data over time.

The dataset contains minute electricity consumption records for all the seven variables. The minute records is resampled into total power consumption of each day, which is used for the below two graphs. The daily consumption of active power, global reactive power, global intensity, voltage, sub metering 1,2,3,4 of full data set will be available after resampling of consumption on daily basis. The daily power consumption of sub meter 1,2,3,4 over for 47 months is represented in the Figs. 2–5. The global active power available for 47 months is represented in the Fig. 6.
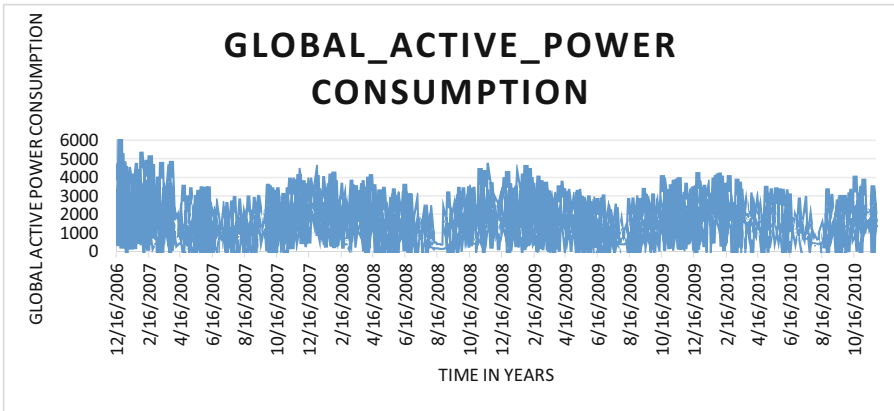
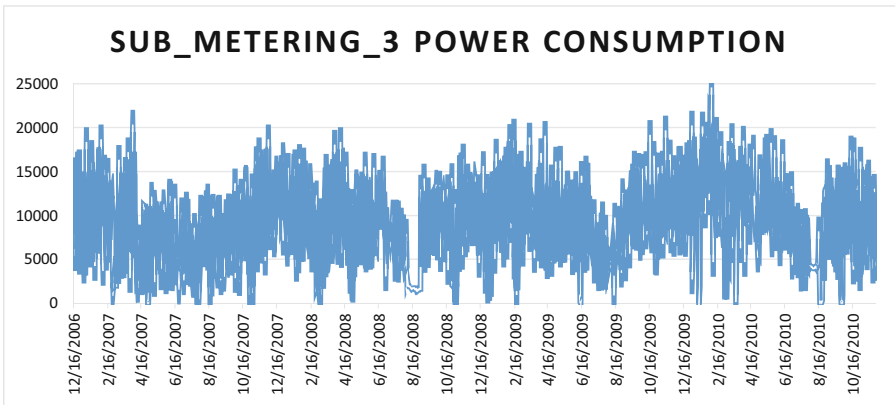**Fig. 2.** Line plot of global active power
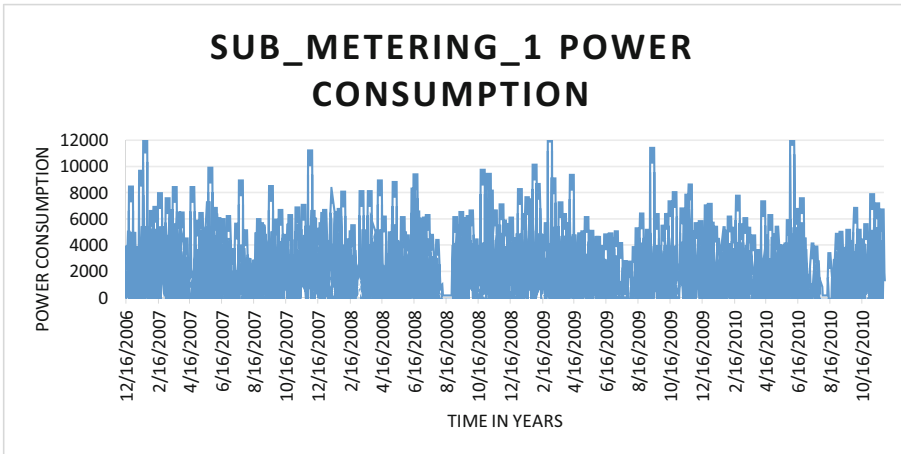
Fig. 3. Line plot of sub meter 1 power consumption on daily basis
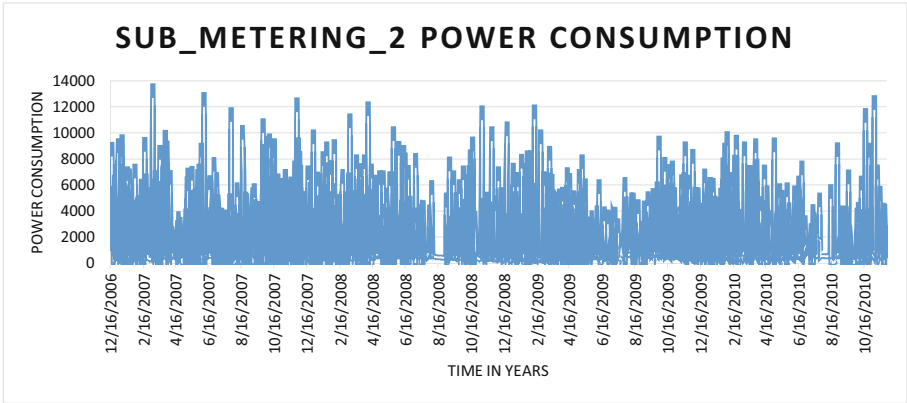
Fig. 4. Line plot of sub meter 2 power consumption on daily basis

Fig. 5. Line plot of sub meter 3 power consumption on daily basis

The dataset contains per minute consumption record for all the seven variables. The minute records is resampled into yearly data, which is used for the below two graphs. The yearly data contains the sum of voltage, global reactive power, active power, global intensity, sub metering 1, 2, 3, 4 of the full year. The yearly power consumption of sub meter 1, 2, 3, 4 over 2006 to 2010 years is represented in the Fig. 2. The global active power available for the years 2006 to 2010 is represented in the Fig. 3 (Figs. 7 and 8).
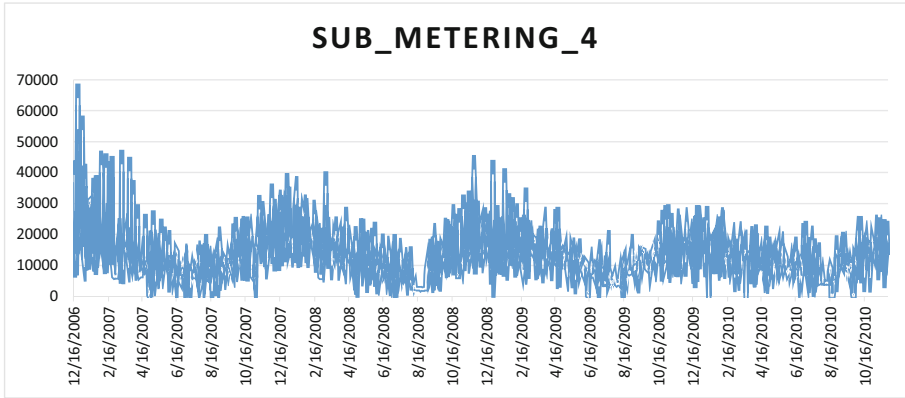
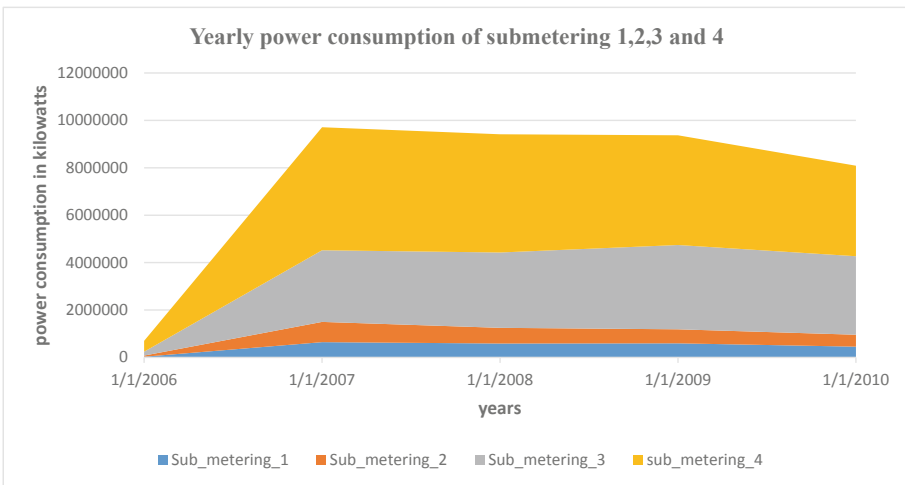**Fig. 6.** Line plot of sub meter 4 power consumption on daily basis


**Fig. 7.** Line plot of sub metering 1,2,3,4 yearly consumption

The below Fig. 9 contains the power consumption of sub metering 1,2,3 and 4 for all months of the year 2008. The resampled data from minute basis to sum of daily consumption is considered hereto analyze the months of year 2008 only 2008 year records are taken from resampled data set. The graph clearly shows that the consumption sub metering 4 is more. The global active power for all months of the year 2008 is represented in terms of graph in Fig. 10. The graph gives idea of the active power when it is high and low for the months of 2008 year.
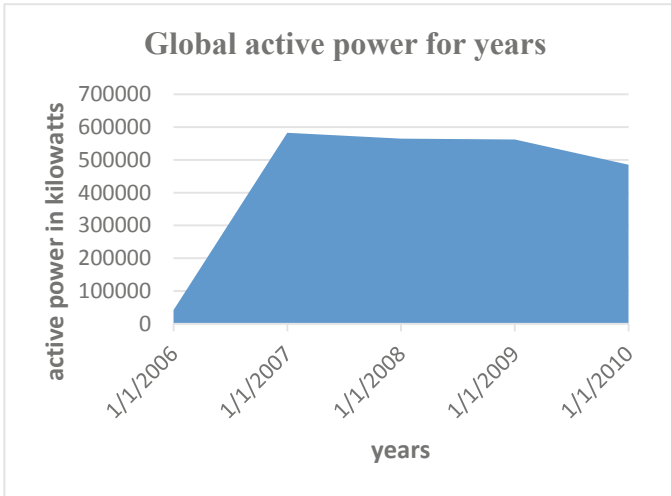
Fig. 8. Line plot of active power yearly consumption
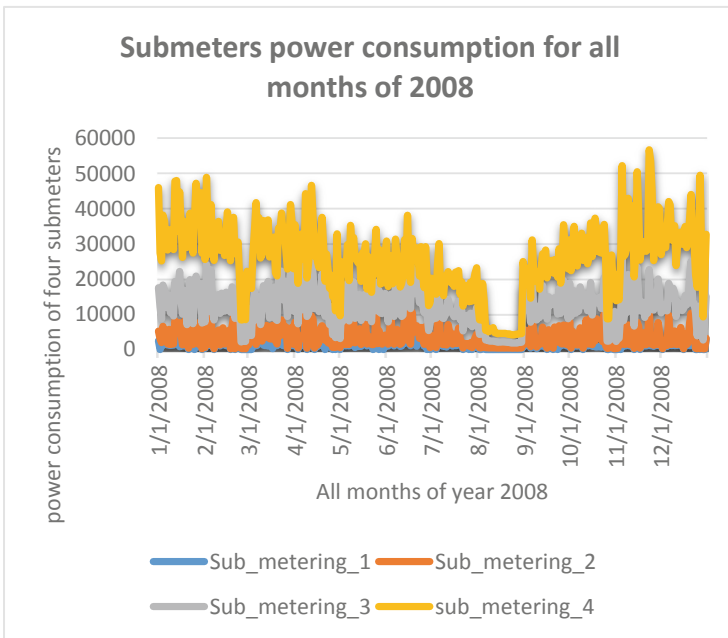
Fig. 9. Line plots for active power for all months in one year.
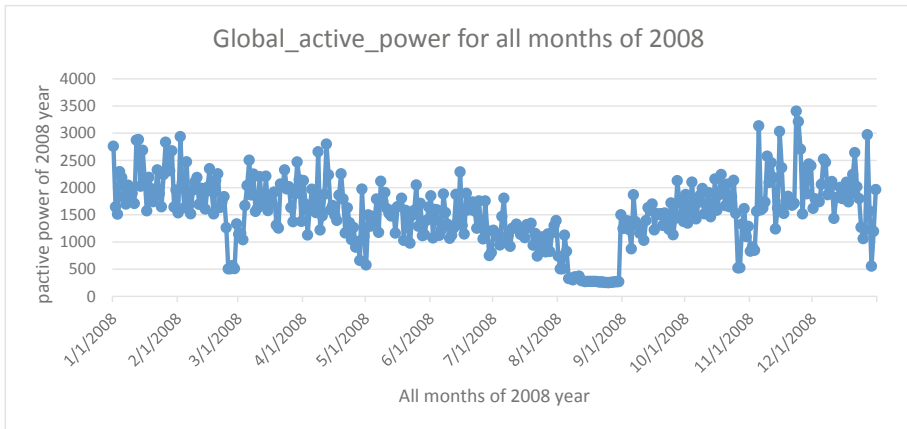
Fig. 10. Line plot of active power for all months of 2008 year.

5   Conclusions

The electricity has major influence in residential and industrial area. The power suppliers and others involved in electrical energy generation, transmission and distribution are dependent on load forecasting. The process and preparation of a power utility can be maintained well by precise models for electric power load forecasting. Electrical energy prediction plays a main role in estimating electricity production and gives clear idea of power system planning and operation. The deep learning models can be exploited to predict the electrical energy usage. The paper gives idea of electricity load forecasting with the help of LSTM, ARIMA and CNN model. All these three models can be used to forecast future load and the forecasting accuracy is measured in terms of RMSE score and MAPE. The ARIMA model has low RMSE score compared to other two models so, it can be considered as best predictive model compared to CNN and LSTM model as per the environment set up and data set used. The results need not be same for real time prediction, different type of data sets and variation in training and testing split up. The future work can be carried out by hyper tuning the parameters of models and varying train and test range.

References

1. Ghosh, S.: Electricity consumption and economic growth in India. Elsev. Energy Policy, 2002, vol. 30, pp. 125–129, 2002
2. Oğcu, G., Demirel, F., Zaim, S.: Forecasting electricity consumption with neural networks and support vector regression. Proc. Social Behav. Sci. **58**, 1576–1585 (2012)
3. Kuster, C., Rezgui, Y., Mourshed, M.: Electrical load forecasting models: A critical systematic review. Elsev. Sust. Cities Soc. **35**, 257–270 (2017)
4. Singh, A., Ibraheem, K., Muazzam, S.: An overview of electricity demand forecasting techniques. Proc. Nat. Conf. Emerg. Trends Electr. Instrum. Commun. Eng. **3**, 38–48 (2013)
5. Park, J.H., Park, Y.M., Lee, K.Y.: Composite modeling for adaptive short-term load forecasting. IEEE Trans. Power Syst. **6**(2), 450–457 (1991). https://doi.org/10.1109/59.76686

6. Nugaliyadde, A., Somaratne, U., Wong, K.: Predicting Electricity Consumption using Deep Recurrent Neural Networks (2019)
7. Mohamed, A., Khan, M.: A review of electrical energy management techniques: Supply and consumer side (industries). J. Energy South. Afr. **20**, 14–21 (2009). https://doi.org/10.17159/2413-3051/2009/v20i3a3304
8. Farhath, Z.A., Arputhamary, B., Arockiam, L.: A survey on arima forecasting using time series model (2016)
9. Goswami, S.: Study of effectiveness of time series modeling (arima) in forecasting stock prices. Int. J. Comput. Sci. Eng. Appl. **4**, 13–29 (2014). https://doi.org/10.5121/ijcsea.2014.4202
10. Chujai, P., Nittaya, K., Kerdprasop, K.: Time series analysis of household electric consumption with ARIMA and ARMA Models. Lect. Notes Eng. Comput. Sci. **2203**, 295–300 (2013)
11. Khan, S., Javaid, N., Chand, A., Khan, A.B.M., Rashid, F., Afridi, I.U.: Electricity Load Forecasting for Each Day of Week Using Deep CNN. In: Barolli, L., Takizawa, M., Xhafa, F., Enokido, T. (eds.) WAINA 2019. AISC, vol. 927, pp. 1107–1119. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15035-8_107
12. Kim, T., Cho, S.: Predicting residential energy consumption using CNN-LSTM neural networks. Energy **182**, 72–81 (2019). https://doi.org/10.1016/j.energy.2019.05.230
13. Chan, S., Oktavianti, I., Puspita, V.: A deep learning CNN and AI-tuned SVM for electricity consumption forecasting: multivariate time series data. In,: IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Vancouver, BC, Canada, pp. 0488–0494 (2019). https://doi.org/10.1109/IEMCON.2019.8936260
14. Jain, P., Quamer, W., Pamula, R.: Electricity consumption forecasting using time series analysis second international conference. In: ICACDS 2018, Dehradun, India, pp. 327–335 (2018)
15. Nichiforov, C., Stamatescu, I., Făgărăşan, I., Stamatescu, G.: Energy consumption forecasting using ARIMA and neural network models. In: 5th Intenational Symposium on Electrical and Electronics Engineering (ISEEE). Galati, Romania, pp. 1–4 (2017). https://doi.org/10.1109/ISEEE.2017.8170657
16. Le, T., Vo, M.T. ; Vo, B., Hwang, E., Rho, B., Sung, W.: Improving electric energy consumption prediction using CNN and Bi-LSTM. Appl. Sci. **9**, 4237 (2019)
17. Mahia, F., Dey, A.R., Masud, M.A., Mahmud, M.S.: Forecasting electricity consumption using ARIMA model. In: 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, pp. 1–6 (2019)
18. Son, H., Kim, C.: A deep learning approach to forecasting monthly demand for residential–sector electricity. Sustainability **12**(8), 1–16 (2020)
19. Ali, U., Buccella, C., Cecati, C.: Households electricity consumption analysis with data mining techniques. In: IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, pp. 3966–3971 (2016). https://doi.org/10.1109/IECON.2016.7793118
20. Parate, A., Bhoite, S.: Individual household electric power consumption forecasting using machine learning algorithms. Int. J. Comput. Appl. Technol. Res. **8** (2018)
21. Balakrishna, K., Sandesh, N.G.: Design of Dynamic Induction Charging Vehicle for Glimpse of Future: Cutting Down the Need for High-Capacity Batteries and Charging Stations. In: Kalya, S., Kulkarni, M., Shivaprakasha, K.S. (eds.) Advances in VLSI, Signal Processing, Power Electronics, IoT, Communication and Embedded Systems. LNEE, vol. 752, pp. 197–204. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0443-0_16
22. Alzubaidi, L., et al.: Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. Big Data **8**(1), 1–74 (2021)

23. Sapna, R., Sheshappa, S.N.: An extensive study on machine learning paradigms towards medicinal plant classification on potential of medicinal properties. In International Conference on Image Processing and Capsule Networks, pp. 541–555. Springer, Cham (2022)
24. Sheela Sobana Rani, K., Pravinth Raja, S., Sinthuja, M., Vidhya Banu, B., Sapna, R., Dekeba, K.: Classification of EEG signals using neural network for predicting consumer choices. In: Computational Intelligence and Neuroscience (2022)
25. Sapna, R., Monikarani, H. G., Mishra, S.F.: Linked data through the lens of machine learning: an enterprise view. In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp. 1–6. IEEE (2019)
26. Mishra, S.: An investigative study on the quality aspects of linked open data. In: Proceedings of the 2018 International Conference on Cloud Computing and Internet of Things, pp. 33–39 (2018)
27. Individual household electric power consumption Data Set: https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption