2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019

Predicting Electric Energy Use of a Low Energy House: A Machine Learning Approach

Risul Islam Rasel
Faculty of Computer Science and Engineering, Chittagong Independent University, Bangladesh
risul@ciu.edu.bd

Nasrin Sultana
Faculty of Computer Sciece and Engineering, University of Chittagong, Bangladesh
nasrin1219@gmail.com

Shairin Akther
Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh
shairin_cse@yahoo.com

Amran Haroon
School of Science and Engineering, Chittagong Independent University, Bangladesh
amran@ciu.edu.bd

***Abstract*— Electricity is one of the greatest inventions of modern science. Now-a-days, without electricity, life is unthinkable. However, the production and distribution of electricity is costly. So, it is necessary to use this facility with great care. Moreover, people in many of the countries using pre-paid electric meters in their houses, offices, stores, even in factories, where they have to buy electricity in pre-paid basis. To do so, they have to estimates the future use of electricity in their vicinity. In this study, we have focused on predicting electric energy use of home appliances in a low energy consumption house. Two very popular and computationally effective machine learning algorithms, namely Support vector regression (SVR) and Artificial neural network with back-propagation (BP-ANN) are applied with cross validation approach. Principal component analysis (PCA) is used to solve data dimensionality problems to analyze and select input features. In addition, F-test and correlation analysis are also done to check the dependencies of dependent variables (target label) on independent variables (predictors). Finally, the proposed model able to predict electric energy uses with more that 98% accuracy.**

Keywords—electric energy prediction; support vector machine; neural network; back-propagation

I. INTRODUCTION

The use of electricity varies among domestic buildings. It depends on numerous factors, such as home architecture, number of occupants, number of electric appliances, indoor environment and outdoor environment of the vicinity, for example: temperature, light, noise etc [1]. All these factors are interrelated strongly. These relations can be analyzed using regression models to understand the relationship among the factors [2,3,4]. Moreover, electric energy demands are changed in weekdays and weekend days due to the staying time of home residents. So, the demand loads are fluctuated during regular days and weekdays. Regression analysis helps to reveal these patterns changing in load demands also [5-9]. Many researches have been done to identify these demand loads patterns. Candanedo et.al.[1], Wang et.al.[4], Arghira et.al.[5], Cetin et.al.[7], and Kavousian et.al.[12], they all worked for predicting appliances energy use into a low energy house. They have used wireless sensor data for environmental assessments inside the house and outside the house, and smart electric meter's data to analyze energy demand loads. All the works above tried to design a load demand model for predicting future electric use into a home using big data driven analysis. On the other hand, many studies have been undertaken to analyze occupant's behaviors into home or office in order to rank the appliance efficiency in household use [2,3,6] . These works tried to analyze occupant's behaviors in home appliance using during their staying at home or office. They used regression analysis and probabilistic analysis to identify the load patterns in different behaviors of occupant's. In addition, some works were done to predict accurate occupancy numbers by analyzing the appliance use behaviors [11]. However, all these works above proved one common thing that, appliance energy use in home or office depends on many factors, such as occupant's numbers, internal and external environment of home or office, building architecture, geo-location of the building etc. Without counting these factors it is not possible to estimate the appliance energy use in a home or office. In this study, we have used environmental wireless sensor data, such as humidity, temperature, wind speed, dew points and visibility in order to understand the internal and external environments of the building and energy use data from smart electric meter. The experimental dataset is a secondary dataset, which is collected from UCI machine learning repository for research purpose abiding by the copyright instructions. Author of the dataset is Luis M. Candanedo [1].

II. METHODOLOGY

A. Data description

The experiment dataset was prepared by combining two types of information. One was environmental data, like: temperature, humidity, wind speed, dew point, and visibility. Other was energy use information for lights and appliances.

978-1-5386-9111-3/19/$31.00 ©2019 IEEE

Temperature and humidity data were collected by using wireless sensor network (ZigBee), which contained 4.5 months data for each 10 minutes slot, which means hourly 6 data and 144 data daily. The energy use data was also collected in 10 minutes interval using m-bus meters. However, wind speed, dew point and visibility data were collected from nearest weather station. Then all data from different sources were merged together to prepare experiment data. Finally, We had separated the main dataset into five parts, like: mooring, afternoon, evening, night, and midnight. Table I shows the data set attributes with their measurement units information.

TABLE I. DATA ATTRIBUTE LIST

Attribute Name	***Unit***
Date time year-month-day	hour:minute:second
Appliances, energy use	Wh
lights, energy use of lights	Wh
T1, Temperature in kitchen	ºC
T2, Temperature in living room	ºC
T3, Temperature in laundry room	ºC
T4, Temperature in office room	ºC
T5, Temperature in bathroom	ºC
T6, Temperature outside the building	ºC
T7, Temperature in ironing room	ºC
T8, Temperature in teenager room	ºC
T9, Temperature in parents room	ºC
T_out, Temperature outside (from weather station)	ºC
RH_1, Humidity in kitchen	%
RH_2, Humidity in living room	%
RH_3, Humidity in laundry room	%
RH_4, Humidity in office room	%
RH_5, Humidity in bathroom	%
RH_6, Humidity outside the building	%
RH_7, Humidity in ironing room	%
RH_8, Humidity in teenager room	%
RH_9, Humidity in parents room	%
RH_out, Humidity outside (from weather station)	%
Pressure (from weather station)	mm Hg
Wind speed (from weather station)	m/s

B. Data vidualization

Figure 1 and Figure 2 show the visual representation of appliances and lights energy use data respectively. From these two graph, there is a clear indication of uncertainty in energy use prediction into a low energy house. These data are only about a specific home. The spikes of the line graphs show that the demand of power fluctuates in time varying. It is because of many correlated things; like: presence of household people, room temperature, outside temperature, light condition into and outside the rooms, weather conditions, Humidity etc. Light energy use relatively lower than the appliances, because lights remains turned off most of time in day time. But, appliances need more energy use because some household appliances, like: Refrigerator, Air condition, Electric Fan, Electric cooker, Room heater, Other electronic stuffs, etc remain in use very frequently in whole day.

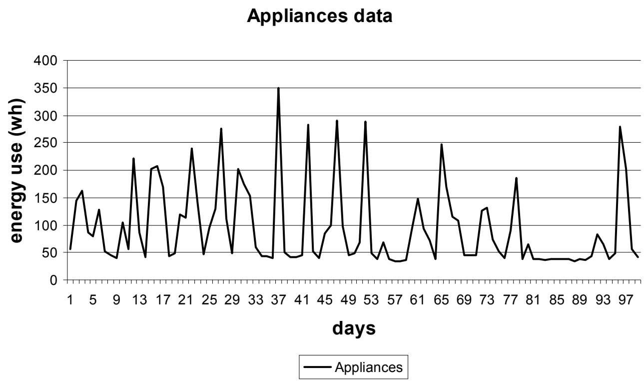

Fig. 1. Daily average energy use data by home appliances

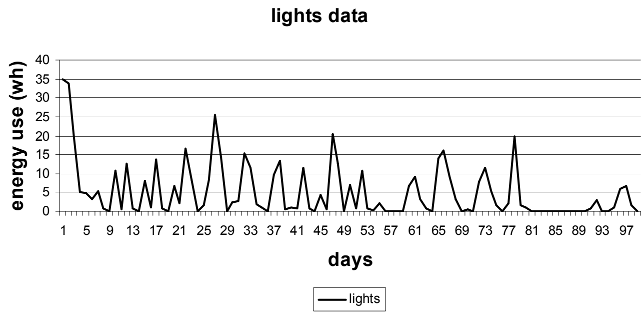

Fig. 2. Daily average energy use data by lights

C. Data preparation

Though the raw data set was collected from online repository [1], it needed to do some major data pre-process before preparing our experiment data. Such as, the main data set contained energy usage time series data for each 10 minutes interval. But we have calculated some derived attributes, like; daily average, daily minimum, daily maximum, weekday use, weekend use, morning time use, afternoon time use, evening time use, night and midnight use. All these attributes are prepared in order to identify the different patterns of electric energy use in different times in a day. These analyses help to identify correlations among the predictor attributes and the target attribute (label). Moreover, we have done 5 different types of analysis, such as morning time use, afternoon time use, evening time use, night and midnight use. So, dataset are prepared accordingly for these five types of experiments. Table II shows the PCA analysis results for choosing correlated feature attributes as predictors. Here Positive (+) means strong correlation and Negative (-) means relatively less correlated.

TABLE II. ATTRIBUTE RELATIONSHIP SUMMARY AFTER PCA

Attributes	*Appliance*	*Lights*
T1	Positive (+)	Negative (-)
T2	Positive (+)	Negative (-)
T3	Positive (+)	Negative (-)
T4	Positive (+)	Negative (-)
T5	Positive (+)	Negative (-)
T6	Positive (+)	Negative (-)
T7	Positive (+)	Negative (-)
T8	Positive (+)	Negative (-)
T9	Positive (+)	Negative (-)
T_out	Positive (+)	Negative (-)
RH_1	Positive (+)	Positive
RH_2	Negative (-)	Positive
RH_3	Positive (+)	Positive
RH_4	Positive (+)	Positive
RH_5	Positive(+)	Positive
RH_6	negative(-)	Positive
RH_7	negative(-)	Positive
RH_8	negative(-)	Positive
RH_9	negative(-)	Negative(-)
RH_out	negative(-)	Positive
Pressure	negative(-)	Negative(-)
Wind speed	positive(+)	Positive

D. *Model Flow chart*

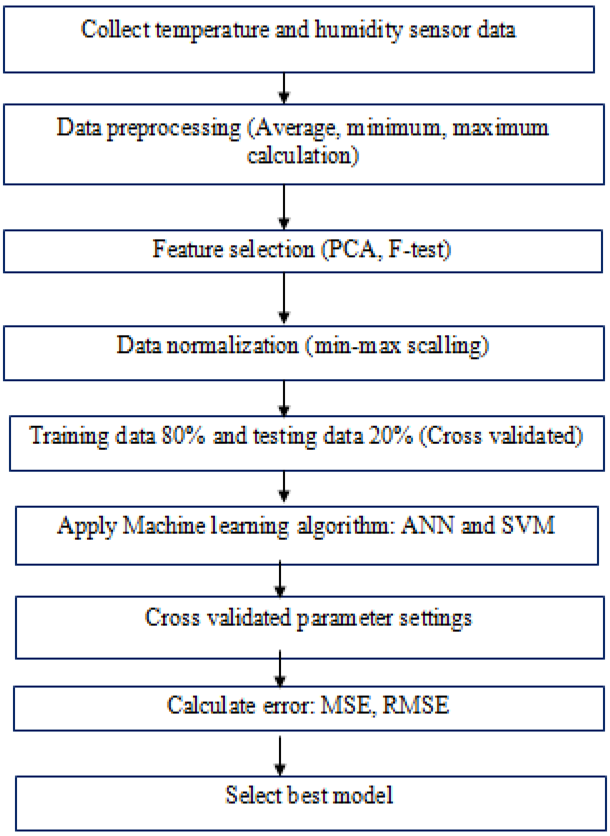

Fig. 3. Model flow chart

Figure 3 shows the flow chart of our experiment process. For feature selection we applied PCA and F-test analysis in order to identify necessary input attributes from the dataset. Min-max normalization techniques are applied such that the ranges of values of different attributes remain into a certain range, like: 0 to 1. It helps to create equal opportunities of dominance to all predictor variables into predictive model. Otherwise, attributes with larger values will always dominate in weight gaining during model training, which creates hurdle to ensure significance of lower valued attributes into trained model. As a result, model faces over-fitting and under-fitting problems in prediction. We have separated our dataset into two parts, 80% of total data into training data and rest of the 20% into test data. SVR and BP-ANN were applied as machine learning algorithms. Cross validations were applied to identify fine tuned parameters combinations of those algorithms. Finally, best model was selected based on their predictive performance.

III. EXPERIMENT RESULTS

A. *Support vector machine model parameters for appliances*

Support Vector Regression (SVR) works very good to identify patterns of dataset with numeric feature values. It has the ability to perform computationally effective analysis to reveal knowledge from input dataset. As our experiments are related to time series data analysis for pattern recognition and future value prediction, we choose SVR to apply with different types of kernel tricks, like; radial, polynomial, and sigmoid. Different kernel tricks help to identify and learn those hidden patterns and train the model with these knowledge such that model can produce good predictive results with minimum error rate. Table III show the cross validated information of SVR analysis with three different kernel techniques for appliance energy use prediction. All the parameters are cross validated and only the best combinations are tabulated with their error performances. From the Table III we see radial bias function performs better than the others, though the differences are very small.

TABLE III. CROSS VALIDATED RESULT SUMMARY OF SVM MODEL WITH DIFFERENT KERNEL FOR APPLIANCES

kernel	*Kernel parameters*					*MSE*	*RMSE*
	gamma	*cost*	*Epsilon*	*degree*	*Coef0*		
radial	0.01	2	0.0001	----	----	1.32%	11.49%
polynomial	0.01	2	0.0001	2	0	1.56%	12.51%
sigmoid	0.01	2	0.0001	----	0	1.54%	12.42%

Table IV shows the fine tuned SVR parameters with three different kernels for lights energy use prediction. Here polynomial and sigmoid kernels very slightly outperformed radial bias function.

TABLE IV. CROSS VALIDATED RESULT SUMMARY OF SVM MODEL WITH DIFFERENT KERNEL FOR LIGHTS

kernel	*Kernel parameters*					*MSE*	*RMSE*
	gamma	*cost*	*Epsilon*	*degree*	*Coef0*		
radial	0.01	2	0.0001	----	----	0.39%	6.25%
polynomial	0.01	2	0.0001	2	0	0.37%	6.07%
sigmoid	0.01	2	0.0001	----	0	0.37%	6.11%

TABLE V. ACTUAL VS PREDICTED VALUES OF APPLIANCES USING SVR

Attributes	*Date*	*Actual Value*	*Predicted value*	*Difference*
Appliances, T1,RH_1, T2,T3, RH_3,T4, RH_4,T5, RH_5,T6, T7,T8,T9, T_out, Windspeed	4/20/2016 morning	0.168	0.108	0.060
	4/20/2016 afternoon	0.284	0.069	0.215
	4/20/2016 evening	0.034	0.051	-0.017
	4/20/2016 night	0.038	0.006	0.033
	4/21/2016 midnight	0.052	0.163	-0.111
	4/21/2016 morning	0.135	0.225	-0.090
	4/21/2016 afternoon	0.262	0.240	0.022
	4/21/2016 evening	0.884	0.152	0.733
	4/21/2016, night	0.078	0.112	-0.034

Table V and VI show the actual vs Predicted values of energy usage of home appliances and lights respectively using support vector regression model. The predictors were chosen by using PCA analysis to feed input data into SVR model. There are five types of predicted values, like; morning, afternoon, evening, night and midnight time prediction of energy usage by home appliances and lights. In some cases of light energy prediction, some predicted values are produced in negative amount. This is because the previous inputs pattern contained many zero values, which means no lights were used. So the model predicts some negative energy values which could be treat as "0", as the energy use amounts can not be negative ever; either zero or some positive values.

TABLE VI. ACTUAL VS PREDICTED VALUES OF LIGHTS USING SVR

Attributes	*Date*	*Actual Value*	*Predicted value*	*Difference*
Lights, RH_1, RH_2, RH_3, RH_4, RH_5, RH_6, RH_7, RH_8, RH_out, Windspeed	4/20/2016 morning	0.063	0.007	0.057
	4/20/2016 afternoon	0.333	0.001	0.333
	4/20/2016 evening	0.000	-0.038	0.038
	4/20/2016 night	0.010	-0.079	0.089
	4/21/2016 midnight	0.041	0.009	0.032
	4/21/2016 morning	0.008	-0.019	0.027
	4/21/2016 afternoon	0.036	-0.002	0.038
	4/21/2016 evening	0.000	-0.106	0.106
	4/21/2016 night	0.000	-0.140	0.140

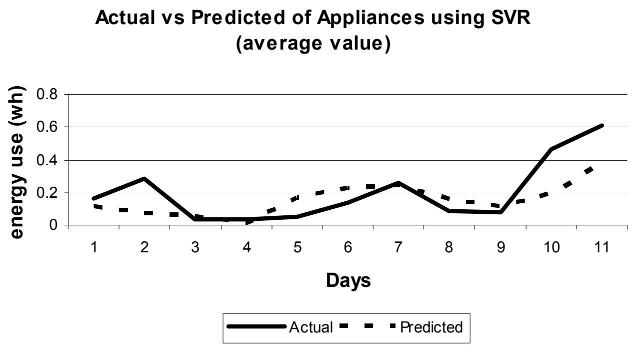

Fig. 4. Actual vs Predicted electric energy use values for home appliances using SVR

Fig. 4 and Fig. 5, show the graphical illustrations of actual and predicted energy use values for home appliances and lights respectively using support vector regression model. From Fig. 4, it is clear that energy usage in home appliances always has some values rather than "0", and there is no gradual increase or decrease happens all in a sudden. So, the proposed model can identify and predict the pattern of appliances power use almost correctly. But, in Fig. 5, there is a clear indication of irregular and inconsistent fluctuations of energy use by lights into the house. So, the pattern identification is little bit critical than the appliances use. But, after the rigorous and careful training, our proposed model able to solve the problem with margin of error between actual and predicted values.

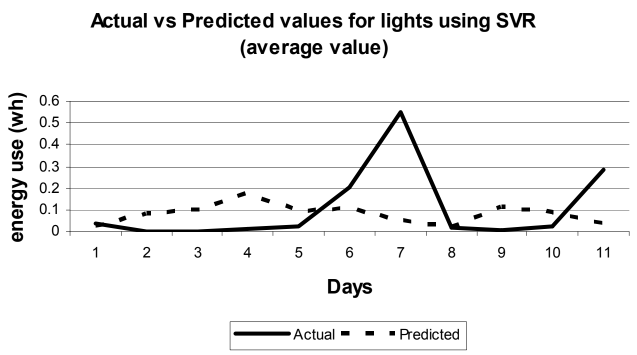

Fig. 5. Actual vs Predicted electric energy use values for home lights using SVR

C. *Artificial Neural Network with Back Propagation*

Artificial neural network is one the best machine learning approaches for identifying patterns and predictions. It has many variations such as feed forward, back-propagation, recurrent etc which make it pioneer in computational intelligence era. Moreover, the flexibility in ANN architectural design, such as choosing hidden layer number, processing unit number, activation function variations, threshold values setup etc make it fantastic in computational analysis for identifying very critical patterns which is very tough task for other AI techniques.

In this study, we have applied feed forward and back-propagation artificial neural networks with different network architectures using logistic (sigmoid) and tanH activation functions for activating processing units (neurons) for triggering information in forward directions and sum squared error (sse) functions as convergence of error to minimize error generated in each iteration. Table VII and VIII show the experiments summaries using ANN with back propagation technique. The tabulated values of parameters were cross validated and only the fine tuned combination of parameters are shown. From the Table VII and VIII, it is very clear to see that sigmoid or logistic activation function with threshold value 0.05 and learning rate 0.05 can produce the best results among the all. TanH function also able to predict energy use for appliances and lights with marginal error rates. But logistic function outperforms all. And the network architecture is 15 inputs, one hidden layer with 2 neurons and 1 output (15-2-1).

TABLE VII. BP-ANN MODEL PARAMETERS WITH ERROR RATE FOR APPLIANCES (CROSS VALIDATED)

Hidden Layer	*Activation function*	*Error Function*	*Threshold value*	*Learning rate*	*MAE*	*RMSE*
2	logistic	sse	0.01	0.01	0.29	0.36
			0.02	0.02	0.24	0.30
			0.02	0.03	0.25	0.30
			0.03	0.04	0.18	0.20
			0.04	0.05	0.16	0.19
			0.05	0.05	0.14	0.18
3	tanh	sse	0.01	0.01	0.48	0.50
			0.02	0.02	0.48	0.50
			0.02	0.03	0.48	0.50
			0.03	0.04	0.48	0.50
			0.04	0.05	0.48	0.50
			0.05	0.05	0.48	0.50

TABLE VIII. BP-ANN MODEL PARAMETERS WITH ERROR RATE FOR LIGHTS (CROSS VALIDATED)

Hidden Layer	*Error Function*	*Activation function*	*Threshold value*	*Learning rate*	*MAE*	*RMSE*
2	logistic	sse	0.01	0.01	0.08	0.13
			0.02	0.02	0.07	0.12
			0.02	0.03	0.07	0.12
			0.03	0.04	0.06	0.11
			0.04	0.05	0.06	0.11
			0.05	0.05	0.06	0.11
3	tanh	sse	0.01	0.01	0.07	0.13
			0.02	0.02	0.07	0.13
			0.02	0.03	0.07	0.13
			0.03	0.04	0.07	0.12
			0.04	0.05	0.07	0.12
			0.05	0.05	0.08	0.12

Table IX and X show the actual vs Predicted values using neural network model with back propagation approach. The predictors were chosen by using PCA analysis to feed input data into BP-ANN model. There are five types of predicted values, like; morning, afternoon, evening, night and midnight time prediction of energy usage by home appliances and lights. Difference between actual vs predicted values is very marginal in each time slot, which is the indication of model performance stability in different situation or time.

TABLE IX. ACTUAL VS PREDICTED VALUES OF APPLIANCES USING BP-ANN MODEL WITH LOGISTIC FUNCTION

Attributes	*Date*	*Actual value*	*Predicted value*	*Difference*
Appliances, T1,RH_1, T2,RH_2,T3,RH_3, T4,RH_4,T5,RH_5,T6, RH_6,T7,RH_7, T8,RH_8,T9,RH_9, T_out, Windspeed	4/20/2016 (afternoon)	0.1677	0.1385	0.0292
	4/20/2016 (evening)	0.2842	0.2128	0.0714
	4/20/2016 (night)	0.0337	0.0905	-0.0569
	4/20/2016 (midnight)	0.0383	0.0407	-0.0024
	4/21/2016 (morning)	0.0520	0.1224	-0.0704
	4/21/2016 (afternoon)	0.1347	0.1541	-0.0194
	4/21/2016 (evening)	0.2621	0.1876	0.0745
	4/21/2016 (night)	0.0884	0.0948	0.0121
	4/21/2016 (midnight)	0.0779	0.0658	0.0121

TABLE X. ACTUAL VS PREDICTED VALUES OF LIGHT USING BP-ANN MODEL WITH LOGSTIC FUNCTION

Attributes	*Date*	*Actual value*	*Predicted value*	*Difference*
Lights, T1,RH_1, T2,RH_2,T3,RH_3, T4,RH_4,T5,RH_5,T6, RH_6,T7,RH_7, T8,RH_8,T9,RH_9, RH_out, Windspeed ,	4/20/2016 (afternoon)	0.0635	0.0201	0.0434
	4/20/2016 (evening)	0.3333	0.0550	0.2783
	4/20/2016 (night)	0.0000	0.0566	-0.0566
	4/20/2016 (midnight)	0.0095	0.0554	-0.4586
	4/21/2016 (morning)	0.0408	0.0051	0.0357
	4/21/2016 (afternoon)	0.0079	0.0484	-0.0405
	4/21/2016 (evening)	0.0357	0.0544	-0.0187
	4/21/2016 (night)	0.0000	0.0560	-0.0560
	4/21/2016 (midnight)	0.0000	0.0557	0.0557

Fig. 6. and Fig. 7. show the graphical representations of actual and predicted energy use values for home appliances and lights respectively using BP-ANN model. From Fig. 6. it is

clear that our proposed BP-ANN model able to predict appliance energy use very effectively. The error between actual and predicted values is so marginal, almost negligible. But from Fig. 7. it is very easy to understand that model has limitation to predict actual pattern of light energy usage. It predicts almost a linear pattern of energy use for lights. This is because input data contained too many "0" values, as a result model could not reveal the actual pattern of energy usage of lights. Which means, the experiment dataset needs to be purified by applying some rectifications techniques to strengthen the input signals rather than "0" values to overcome the problem stated above. And, the predictive model still needs some fine tuning in parameter selections also. However, the error rate between actual and predicted values is still in tolerable range.

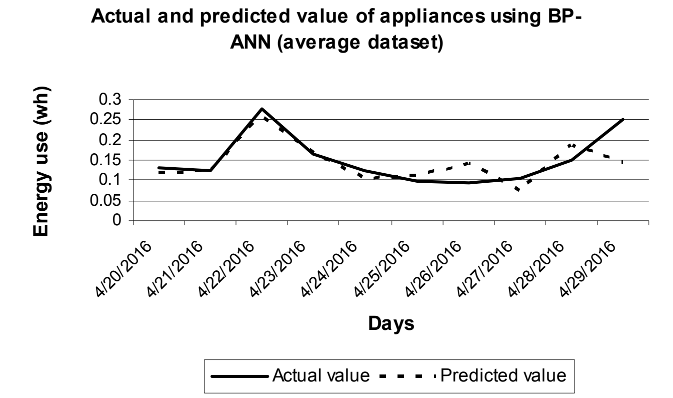

Fig. 6. Actual vs Predicted electric energy use values for home appliances using BP-ANN

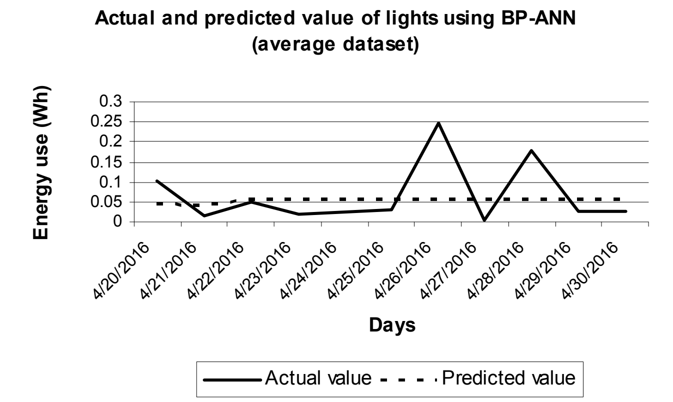

Fig. 7. Actual vs Predicted electric energy use values for home lights using BP-ANN

IV. CONCLUSION

The aim of this study is to propose a computational intelligence model to predict electric energy usage in a low energy house or apartment. To design the model two very computationally effective machine learning approaches are applied, named support vector machine (SVM) and Artificial neural network (ANN). Both models are trained and tested with a very clean and properly pre-processed dataset, which is collected from UCI, a very popular data repository for data mining and machine learning enthusiasts. The experiments are design to predict electric energy use in five dimensions, such as morning time use, afternoon time use, evening time use, night time use, and midnight time use. All are done in order to provide flexibility in electric energy use prediction in different times in a day. Moreover, use of electric energy by appliances and lights are experimented separately, by which our proposed model will help the users to understand the separate pattern of energy use by home appliances and lights. After a meticulous analysis, we find our proposed BP-ANN model is very efficient in predicting electric energy use in a low energy house or apartment. Though the BP-ANN model outperforms SVR model very slightly, but the SVR model also can produce good results in electric energy predicting.

References

[1] Luis M. Candanedo*, VéroniqueFeldheim, Dominique Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," Energy and Buildings 140 (2017) 81–97.

[2] T. Hong, S.C. Taylor-Lange, S. D'Oca, D. Yan, S.P. Corgnati, Advances in researchand applications of energy-related occupant behavior in buildings, EnergyBuild. 116 (2016) 694–702.

[3] A. Kavousian, R. Rajagopal, M. Fischer, Ranking appliance energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliance energy efficiency in residential buildings, Energy Build. 99 (2015) 220–230.

[4] Z. Guo, Z.J. Wang, A. Kashani, Home appliance load modeling from aggregatedsmart meter data, IEEE Trans. Power Syst. 30 (1) (2015) 254–262.

[5] N. Arghira, L. Hawarah, S. Ploix, M. Jacomino, Prediction of appliances energy use in smart homes, Energy 48 (1) (2012) 128–134.

[6] K.S. Cetin, Characterizing large residential appliance peak load reduction potential utilizing a probabilistic approach, Sci. Technol. Built Environ. 22 (6) (2016) 720–732.

[7] K.S. Cetin, P.C. Tabares-Velasco, A. Novoselac, Appliance daily energy use in new residential buildings: Use profiles and variation in time-of-use, Energy Build. 84 (2014) 716–726.

[8] R.V. Jones, A. Fuertes, K.J. Lomas, The socio-economic, dwelling and appliancerelated factors affecting electricity consumption in domestic buildings,Renew. Sustain. Energy Rev. 43 (2015) 901–917.

[9] H.-x. Zhao, F. Magoulès, A review on the prediction of building energyconsumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592.

[10] M.G. Fikru, L. Gautier, The impact of weather variation on energyconsumption in residential houses, Appl. Energy 144 (2015) 19–30.

[11] L.M. Candanedo, V. Feldheim, Accurate occupancy detection of an office roomfrom light, temperature, humidity and CO2measurements using statisticallearning models, Energy Build. 112 (2016) 28–39.

[12] A. Kavousian, R. Rajagopal, M. Fischer, Determinants of residential electricityconsumption: using smart meter data to examine the effect of climate,building characteristics, appliance stock, and occupants' behavior, Energy 55(2013) 184–194.