International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 13th Aug 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 08

DOI: 10.48047/IJIEMR/V11/ISSUE 08/11

Title **Power Consumption Estimation of Household Electric Appliances**

Volume 11, ISSUE 08, Pages: 85-93

Paper Authors
D. Naga Swetha

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Power Consumption Estimation of Household Electric Appliances

D. Naga Swetha
Assistant Professor, Department of CSE,
G. Narayanamma Institute of Technology & Science (for Women), Telangana, India.

Abstract – A non-intrusive monitoring system estimates the behaviour of individual electric appliances from the measurement of the total household load demand curve. The total load demand curve is measured at the entrance of the power line into the house. The power consumption of individual appliances can be estimated using several machine learning techniques by analyzing the characteristic frequency contents from the load curve of the household. Monitoring system of ON/OFF states have already been developed. This system could establish sufficient accuracy. In the next phase, the monitoring system should be able to estimate the power consumption for an air conditioner with an inverter circuit. The results of applying several regression methods such as multi-layered perceptrons (MLP), radial basis function networks (RBFN) and support vector regressors (SVR) to estimate the power consumption of an air conditioner. Experiments show that RBFN can achieve the best accuracy for the non-intrusive monitoring system.
Keywords – household appliances; energy efficiency; power consumption; electricity bill.

I. INTRODUCTION

The development of a monitoring system, which can monitor each household electric appliance, is very important to acquire basic information for building the policy of the energy conservation, the forecast of electric energy demand to plan building new power plants, and creating better new customer services. It is expected that the monitoring system for household electric appliances is inexpensive and non-intrusive, because the conventional monitoring method which is set up measuring equipment for each appliance is expensive and forces the customers inconvenience. The nonintrusive monitoring means the measurement of power load is done in the outside of a house. Non-intrusive monitoring system have already been developed for conventional appliances[1]. This system, increase for a higher energy efficiency of inverter circuits. We have developed a non-intrusive monitoring system, which can estimate the ON/OFF states of household electric appliances include inverter. For the system users, the system can achieve a sufficient accuracy using large margin classifiers. Hence, the system is able to estimate the power consumption of conventional electric appliances, which show a constant power consumption, if we know the power consumption of each household electric appliance[2]. However, an electric appliance with an inverter circuit (inverter type appliance) such as an air conditioner does not show a constant power consumption. Therefore, we have to develop the non-intrusive monitoring system, which can estimate the power consumption of inverter type appliances.

1.1 Existing System

Power consumption is predicted using the K-nearest neighbour model. The K nearest neighbour algorithm is a supervised classification algorithm. It takes a bunch of labelled points and uses them to learn how to label other points. To label a new point, it looks at the labelled point closest to the new point which are its nearest neighbours and has those neighbours vote. So which ever label, the most neighbours have is the label for our new point[1]. Here, "K" in K-NN is the number of neighbours it checks. Its supervised because you are trying to classify a point basing on the known classification of other points. The K-nearest neighbour clusters all the similar instances into a cluster. The K-NN forecasting method was approached using set of historical observation (daily load curve) and their successors. K-nearest model clusters only the similar instances. Any instance is not similar or far away doesn't belong to any cluster and is ignored[3]. K-NN forecasting method will not include all instances which may result in losing the information. K-NN doesn't work on large datasets.

1.2 Disadvantages

There are many disadvantages in existing system[4]:

1. K-NN is a lazy learner

Because it doesn't learn a discriminative function from the training data but "memorizes" the training dataset instead.

2. Doesn't work well with a large dataset:

Since KNN is a distance-based algorithm, the cost of calculating distance between a new point and each existing point is very high which in turn degrades the performance of the algorithm.

3. Doesn't work well with a high number of dimensions:

Again, the same reason as above. In higher dimensional space, the cost to calculate distance becomes expensive and hence impacts the performance.

4. Sensitive to outliers and missing values:

KNN is sensitive to outliers and missing values and hence we first need to impute the missing values and get rid of the outliers before applying the KNN algorithm.

5. Limited in forecasting future value as it only identifies similar instances.

1.3 Proposed System

The power consumption of individual appliances can be estimated using several machine learning techniques. Machine learning (ML) is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks[5]. To get the most value from the ML you have to know how to pair the best algorithms with the right tools and processes. Supervised learning, unsupervised learning and reinforcement learning are the most widely adapted machine learning methods. The main difference with machine learning is that just like statistical models, the goal is to understand the structure of the data – fit theoretical distributions to the data that are well understood. So, with statistical models there is a theory behind the model that is mathematically proven, but this requires that data meets certain strong assumptions too. Machine learning has developed based on the ability to use computers to probe the data for structure, even if we do not have a theory of what that structure looks like. The test for a machine learning model is a validation error on new data, not a theoretical test that proves a null hypothesis. Because machine learning often uses an iterative approach to learn from data, the learning can be easily automated. Passes are run through the data until a robust pattern is found.

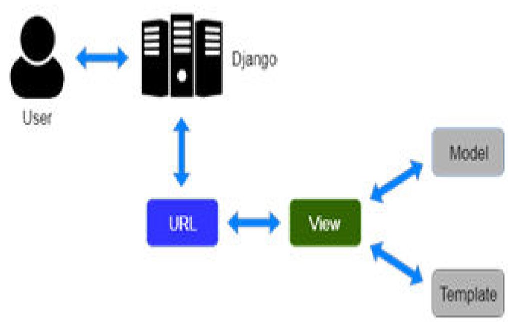


Fig.1: Django uses MVT

1.4 Advantages

1. Separate measuring equipment is not needed.
2. It's inexpensive
3. Its non-intrusive monitoring system.
4. Power consumption can be reduced by monitoring the usage of

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

individual appliances

II. METHODOLOGY

The overall system is divided into two sections. The first one is admin and the other is user. Admin can perform many functions like uploading the dataset, training the dataset with three algorithms and testing the dataset[6]. Admin will upload the dataset which consists of 1500 to 2000 attributes and watts, number of hours consumed are taken as parameters. Admin can train the dataset with different algorithms. The algorithms used are Naïve Bayes algorithm, Support Vector Machine, Decision Trees.

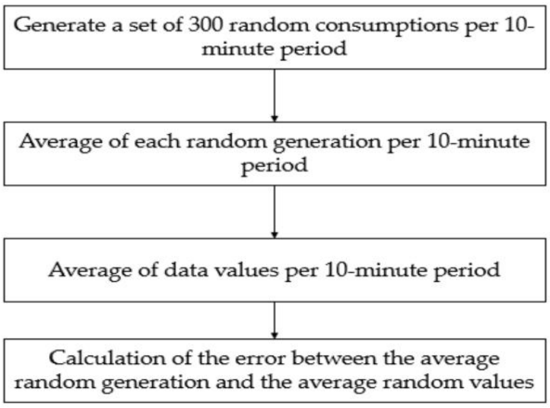


Fig. 2: Flowchart of the comparison between random and real consumptions.

2.1 Discontinuous Power Consumption Household Appliances

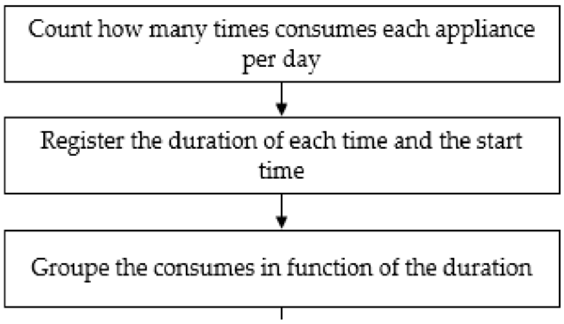
The number of times each household appliance is used per day is counted, as well as the duration of each time and in which periods of the day it happens. The washing machine, dishwasher and dryer only consume energy while they are doing their functions. The first step consists of counting how many times each appliance consumes per day. The duration of each time and its start hour are registered as well. The times with the same consumption duration are grouped together and evaluated in the same way as the continuous consumption appliances[7]. The essential difference is that, in the latter case, just the moments where consumption is taking place are considered, and consequently shorter time intervals, whereas in the previous process the whole day was evaluated.

Once the power consumption curves are associated with each distribution, the following process is carried out:

1. Simulation of the number of power consumptions of each element, where the probability of each integer value is based on the ratio of the previous count.
2. If the above-simulated value is greater than or equal to 1, the duration of each count and its start time are simulated, also based on the data previously collected.
3. Simulation of 300 sets of random consumptions according to the duration of the consumptions and their associated distribution, comparing their average value with the average of the real values.
4. Calculation of the percentage of error.

Figure below shows a flowchart of the entire process of the treatment of the punctual consumptions, from the different counts to the calculation of errors, including the evaluation of distributions.

The simulation was carried out with the consumption data of a house, with 199m2 and three occupants, located in Vancouver[8]. The extrapolation of results can be done in homes with an equivalent surface area, an equal number of inhabitants and a similar climatic zone. If these conditions change, the results would be different.



International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

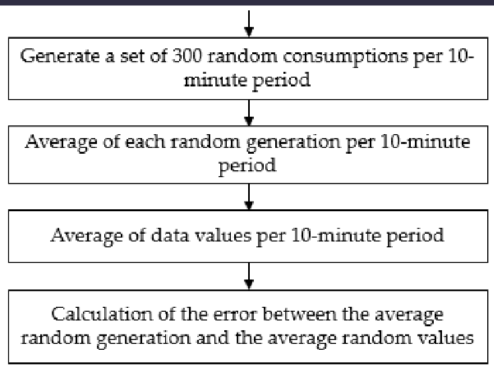


Fig.3: Flowchart of the treatment of the punctual consumptions.

Thus, if the surface area was larger, the consumption would increase. If it were a single dwelling, consumption would be lower, on the other hand. However, the methodology proposed in this paper can be applied if the number of consumption values available are significant and correspond to the same appliances that have been evaluated. Moreover, in some cases, such as the HVAC, the climate data influences the models obtained. This fact is not taken into account in this research work but can be of interest for future work.

The methodology explained in the previous section was applied to the instantaneous consumption data per second of different household appliances, extracted from [3], namely lighting, refrigerator, HVAC, dryer, washing machine and dishwasher. The input data are the consumption of different household appliances in a house located in Vancouver. The powers were measured for 63 days, from 6 March 2016 to 7 May 2016. The dwelling consists of two floors, with a total of 199 m2 and three people living in it. Before starting the statistical evaluation, the consumption was grouped according to a 10-min period of power and then separated into three groups, i.e., working days, Saturdays and Sundays. The last step beforehand was to distinguish between continuous and discontinuous consumption, given that the treatment of the latter is more complex. The reason for evaluating the 10-min periods is a matter of trade-off in accuracy and computational cost. On one hand, in a 10-min period, an appliance is not very likely to experience a large number of consumption changes. If it were 20 min, there is a risk of adding a larger error, as it could cover more than two phases of operation, e.g., a dishwasher. On the other hand, if the period is reduced to 5 min, the accuracy would increase, but the computational cost would be doubled, as twice as many periods would have to be modeled. Thus, a period of 10 min shows an adequate balance. The refrigerator, lighting and HVAC consume a considerable minimum at all times, while the dryer, dishwasher and washing machine only consume a considerable amount when they are running, thus they are considered point loads, or discontinuous consumptions.

III. SYSTEM DESIGN

3.1 Architecture

The overall architecture is shown in the figure. As shown in the figure admin can login into the system. Admin can train and test the dataset with different algorithms. Admin can also analyse the performance[9]. The performance is analysed based on the accuracy of each algorithm. All the four metrics precision, recall, fscore and accuracy are calculated which will be show in the database but only accuracy is presented in the form of a graph. Admin can know which algorithm is the best among the three given algorithms.

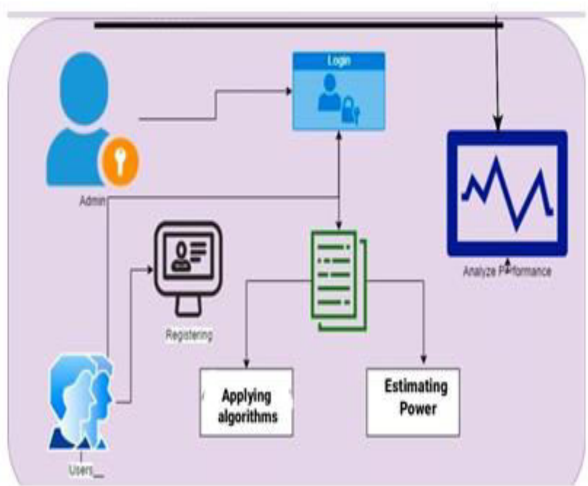


Fig.4: System Architecture

A new user can register with the system with his or her credentials. Already registered user can login into

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

the system with valid credentials and estimate the power consumption by giving the valid inputs like watts, no. of hours used. Invalid credentials are not accepted. The user can know the number of units an each appliance has consumed and can monitor the usage.

3.2 Module Description

The 3 modules we are using are data pre-processing, training and testing, and validation.

Data pre-process: This step includes the collection of proper data set and reducing the noise in the dataset. We collected our data from dataworld. The dataset contains around 1500 -2000 records or instances with 3 different attributes. The noise in the dataset should be reduced. Noisy data includes missing values, null values and faulty data. These values should be replaced by new values which is the mean of all the values or we may replace it with a common null value.

Training or testing: This step contains training our dataset with different algorithms. We are using Naive bayes, Support vector machine and Decision trees. Whenever the dataset is trained with different algorithms, an object for that algorithm is created. This creation of object helps in avoiding repeated training of the dataset with the same algorithm until the dataset is updated.

Validation: This step includes finding the output of the problem and finding the accuracy of the algorithms we are using[3]. Accuracy is found by comparing the predicted and actual values. Other factors like recall, precision, f-score are also being calculated. As accuracy is our main concern we are representing only accuracy in our graph.

The inputs are to be given and we get the output based on the model we have built. In our case decision tree algorithm gives the most accuracy. However this accuracy may differ slightly by increasing or decreasing the number of instances in the dataset. It may also increase or decrease if the dataset with same number of instances is updated.

Table 1. Summary of the evaluated appliances with their type of consumption, type of days and
amount of data.

Appliance	Type of Consumption	Type of Days	Amount of Data
Lightning	Continuous	Working days	45
		Saturdays	9
		Sundays	9
Refrigerator	Continuous	Working days	45
		Saturdays	9
		Sundays	9
HVAC	Continuous	Working days	45
		Saturdays	9
		Sundays	9
Dryer	Occasional	Working days	45
Washing machine	Occasional	Working days	45
Dishwasher	Occasional	Working days	45

IV. IMPLEMENTATION

4.1 Machine Learning

Machine learning (ML) is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

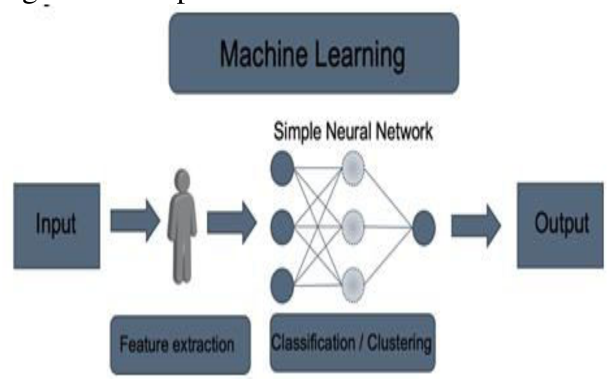


Fig.5: Machine Learning working

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Machine learning approaches are traditionally divided into three broad categories, depending on the nature of the "signal" or "feedback" available to the learning system:

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.

4.2 Machine Learning Models

Performing machine learning involves creating a model, which is trained on some training data and then can process additional data to make predictions. Various types of models have been used and researched for machine learning systems. Artificial neural networks (ANNs), or connectionist systems, are computing systems vaguely inspired by the biological neural networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task- specific rules. An ANN is a model based on a collection of connected units or nodes called "artificial neurons", which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit information, a "signal", from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. In common ANN implementations, the signal at a connection between artificial neurons is a real number, and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs. The connections between artificial neurons are called "edges". Artificial neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Artificial neurons may have a threshold such that the signal is only sent if the aggregate signal crosses that threshold. Typically, artificial neurons are aggregated into layers. Different layers may perform different kinds of transformations on their inputs. Signals travel from the first layer (the input layer) to the last layer (the output layer), possibly after traversing the layers multiple times.

The original goal of the ANN approach was to solve problems in the same way that a human brain would. However, over time, attention moved to performing specific tasks, leading to deviations from biology. Artificial neural networks have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis.

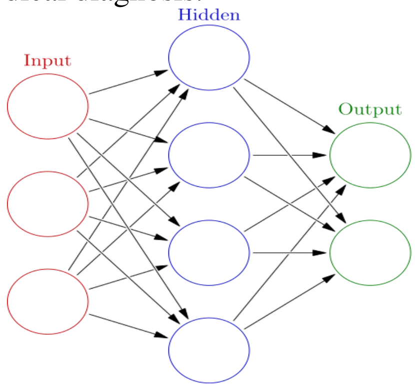


Fig. 6: Artificial Neural Network

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Regression analysis: Regression analysis encompasses a large variety of statistical methods to estimate the relationship between input variables and their associated features. Its most common form is linear regression, where a single line is drawn to best fit the given data according to a mathematical criterion such as ordinary least squares. The latter is often extended by regularization (mathematics) methods to mitigate overfitting and bias, as in ridge regression. When dealing with non-linear problems, go-to models include polynomial regression (for example, used for trendline fitting in Microsoft Excel), logistic regression (often used in statistical classification) or even kernel regression, which introduces non-linearity by taking advantage of the kernel trick to implicitly map input variables to higher- dimensional space.

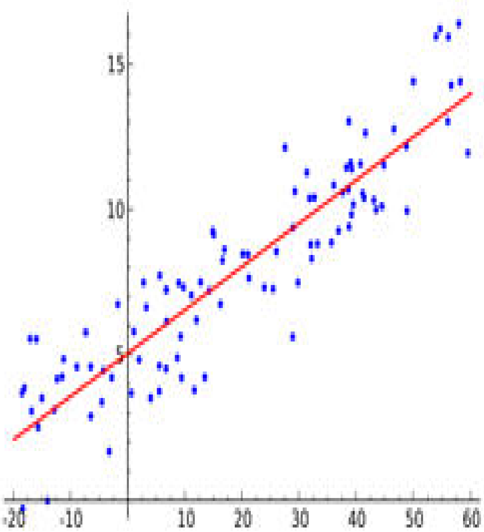


Fig. 7: Illustration of Linear Regression on dataset

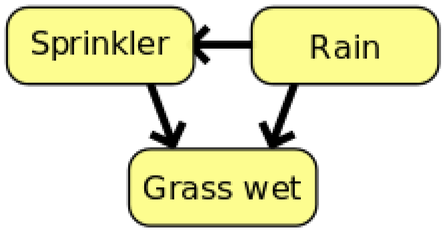


Fig. 8: Simple Bayesian Network

Bayesian networks:

A Bayesian network, belief network, or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independence with a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. Efficient algorithms exist that perform inference and learning. Bayesian networks that model sequences of variables, like speech signals or protein sequences, are called dynamic Bayesian networks. Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams.

Dataset Description: The dataset is split into two parts- training and testing. 70% is used for training and 30% for testing. Dataset is trained with 3 algorithms Naïve bayes, SVM and decision tree. When the dataset is trained an object is formed for that particular algorithm. So the repeated training of the dataset is not necessary. Dataset is taken from data world repository. Database consists of 1500 to 2000 instances or records. Each record consists of 3 attributes. The 3 attributes are no. of hours, watts and the result. Here, we considered 10 different appliances. Each appliance having about 150 records.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

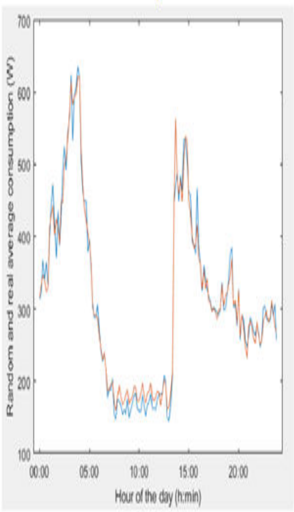
	HomeAppl	Watt	Hours	Units
1	HomeAppl	Watt	Hours	Units
2	Television	17	19	3.23
3	Television	17	16	2.72
4	Television	40	21	8.4
5	Television	30	18	5.4
6	Television	25	1	0.25
7	Television	17	6	1.02
8	Television	17	11	1.87
9	Television	30	14	4.2
10	Television	17	11	1.87
11	Television	30	10	3
12	Television	25	13	3.25
13	Television	30	9	2.7
14	Television	30	9	2.7
15	Television	40	20	8
16	Television	17	9	1.53
17	Television	25	11	2.75
18	Television	25	1	0.25
19	Television	25	11	2.75
20	Television	25	5	1.25
21	Television	25	2	0.5
22	Television	40	16	6.4
23	Television	40	16	6.4
24	Television	25	2	0.5
25	Television	25	4	1
26	Television	30	16	4.8
27	Television	25	6	1.5
28	Television	30	4	1.2
29	Television	17	10	1.7

Sheet1 | Sheet2 | Sheet3

Fig. 9: Dataset

V. EXPERIMENTAL RESULTS

In general, the trends in both graphs are very similar, since the off-peak hours, between 08:00 and 13:00, coincide, and the peak hours, around 07:00, are the same.



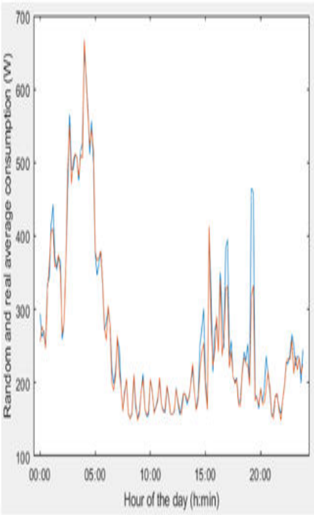
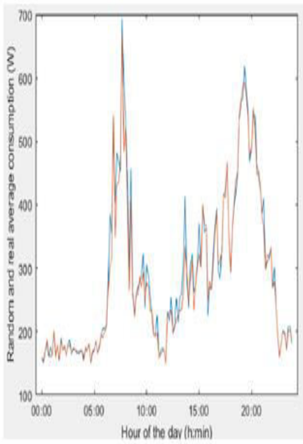


Fig.10: average consumption of the lighting during working days. Fig.11: average consumption of the lighting during Saturdays

However, there are some significant differences between the simulation and the measured situation, especially at 15:00, just at the end of the hours of minimum consumption. The error of the simulation with respect to the measured values is 3.81%, an acceptable value, which makes the simulation valid and perfectly close to the real situation. Figure 5 shows the difference in lighting on Saturdays. Just as there are several more coinciding values in the previous graph, with the same trend, there are also two or three consumption peaks with a somewhat more significant difference than in the previous case, which means that the error increases to 4.26%, but it is still a valid simulation that is close to reality. In the case of Sundays, there are also slightly unequal trends between the two curves (Figure 6), but also some significant differences in some specific points. The error produced in this simulation is 4.27%, very close to that of Saturdays thus all the simulations for the three lighting cases are acceptable.



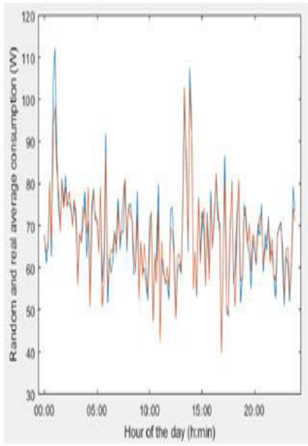


Fig.12: average consumption of the lighting during Sundays. Fig.13: average consumption of the refrigerator during working days

In the case of the refrigerator during the days of the week (Figure 7), the same trend and generally insignificant differences can be seen, except for the first peak. In many cases, the values are practically the same or very close. The error is 3.91%, thus the simulation is considered successful. On Saturdays (Figure 8), the simulation of refrigerator consumption is even tighter than on weekdays. Differences are generally minimal, leading to an error of 3.33%. For Sundays (Figure 9), the same happens as with the simulations for Saturdays, with minimal deviations and a high coincidence. Nevertheless, the error amounts to 3.61%, which is not very significant. The three simulation graphs for the refrigerator consumption fit perfectly with a real situation.

VI. CONCLUSION AND SCOPE FOR FUTURE ENHANCEMENTS

The research and development of scientific models of energy use and the precise prediction of the future energy supply and demand gap have important practical significance for the sustainable economic and social development of our country, the development of the energy industry, the rational use of energy resources, the creation of a conservation oriented society and the information of a national energy strategy. In this we compare with several regression algorithms to estimate the power consumption of an different electric appliances such as Naive base, Decision Tree and Support Vector Machine (SVM) is a conventional nonlinear regression method. We find the sufficient accuracy for monitoring system which can estimate the power consumption of electric appliances to actual fields. The system can be further developed by adding inverter feature. So, now the monitoring system shoulder be able to estimate the power consumption of any electric appliance with an inverter circuit. Power consumption can also be reduced by adding the feature- average usage of electric appliance per month. So, through this feature the person can understand the power consumed by each electric appliance and can reduce the usage of it.

REFERENCES

1. C. M. Bishop.: Newral Networks for Pattern Recognition. Oxford University Press, 1995.
2. W. Hart.: Non-intrusive appliance load monitoring. Proceedings of the IEEE, vol. 80, no.12, 1992.
3. J.Moody et al.: Fast learning in networks of locallytuned processing units. Neural Computation 1 (2), , 692498, 2001. 281-294, 1989.H.Murata et al.: Applying Kernel Based Subspace classification to a Non-Intrusive Monitoring System for Household Electric Appliances ICANN2001.
4. T.Onoda et al.: Applying Support Vector Machines and Boosting to a Non-Intrusive Monitoring System for Household Electric Appliances with Inverters. NC'2000, 2000.
5. D. E. Rumelhart et al.: Learning Internal Rep resent ations by Error Propagation. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. l: Foundations, 318-362, Bradford Books/MIT Press, Cambridge, MA., 1986.
6. V. N. Vapnik.: The Nature of Statistical Learning TheonJ. Springer, 1995.S. Zhai, Y. Cheng,W. Lu, and Z. Zhang, ‘‘Deep structured energy basedmodels for anomaly detection,’’ in Proc. Int. Conf. Mach. Learn., 2016,pp. 1–10.
7. C. Fan, F. Xiao, Y. Zhao, and J. Wang, ‘‘Analytical investigation ofautoencoder-based methods for unsupervised anomaly detection in build-ing energy data,’’ Appl. Energy, vol. 211, pp. 1123–1135, Feb. 2018.
8. E. Mocanu, P. H. Nguyen, M. Gibescu, and L. Wil Kling, ‘‘Deep learn-ing for estimating building energy consumption,’’ Sustain. Energy, GridsNetw., vol. 6, pp. 91–99, Jun. 2016.
9. R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, ‘‘Deep learning and its applications to machine health monitoring,’’ Mech. Syst. Signal Process., vol. 115, pp. 213– 237, Jan. 2019.