International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-7, May 2021

Energy Audit System for Households using Machine Learning

A.Nagesh

Abstract: ***the growth in population and economics the global demand for energy is increased considerably. The large amount of energy demand comes from houses. Because of this the energy efficiency in houses in considered most important aspect towards the global sustainability. The machine learning algorithms contributed heavily in predicting the amount of energy consumed in household level. In this paper, a energy audit system using machine learning are developed to estimate the amount of energy consumed at household level in order to identify probable areas to plug wastage of energy in household. Each energy audit system is trained using one machine leaning algorithm with previous power consumption history of training data. By converting this data into knowledge, gratification of analysis of energy consumption is attained. The performance of energy audit Linear Regression system is 82%, Decision Tree system is 86% and Random Forest 91% are predicted energy consumption and the performance of learning methods were evaluated based on the heir predictive accuracy, ease of learning and user friendly characteristics. The Random Forest energy audit system is superior when compare to other energy audit system.***

Keywords: Energy Prediction, Linear Regression, Machine Learning, Decision Tree, Random Forest

I. INTRODUCTION

With startling rise of India's growing population and increased access to electricity, it has been projected that energy consumption trends in India reflects increasing surges in demand for energy. It has also been noted that nation's electricity demand will almost triple between 2018 and 2040. With identified energy needs nation has been taking several policy measures to promote energy efficiency and conservation. As a result energy auditing is considered as the first step in the process of energy conservation. Industries and big buildings are usually energy efficient as energy audits are regularly conducted and measures are taken to reduce the energy wastage. However it is not true at household levels. Most households may not go for energy audit various reasons to assess the energy efficiency in their homes. Hence it is important to identify areas requiring major investment by incorporating modern energy efficient equipment and by upgrading the existing equipment in homes. For this purpose several methods are developed and special attention is paid towards machine learning algorithms such as linear regression, decision tree, multivariate regression and random forest.

Manuscript received on May 03, 2021.
Revised Manuscript received on May 08, 2021.
Manuscript published on May 30, 2021.
* Correspondence Author
Dr. A.Nagesh*, Professor, Computer Science & Engineering, Mahatma Gandhi Institute of Technology Hyderabad (Telangana), India. Email: akknagesh@rediffmail.com

The machine learning algorithms are best suited to predict the energy consumption at household level to suggest the energy efficiency and conservation. The machine learning algorithm based energy audit systems outperformed all the existing energy audit systems based on other approaches. Thus machine learning models are used to predict the energy consumed at household level.

In the energy systems, one of the most essential topics is the energy consumption. First time the energy consumption has come under preview after the energy crisis in 1970[1]. It is also observed that the consumption of the energy is also increasing rapidly throughout the world[2]. So, we try to use as less energy as possible at each level such as household, farms, vehicles and industrial needs. The energy consumption of these levels are influenced by different factors. Using multiple factors to predict the energy consumption is a complex task [3]. The machine learning [ML] algorithms predicts the energy consumption easily and efficiently [4]. Therefore, using ML models such as linear regression, decision tree and random forest predicted the amount of energy is consumed in houses accurately and efficiently [5][6].

The paper is organized as follows, in section ii the data used to train the models are discussed. In the next section iii the various machine learning algorithms such as linear regression, decision tree and random forest are discussed in detail. Followed by in section iv energy audit model creation using linear regression, decision tree and random forest are explained. Finally the prediction of the energy audit system and performance analysis done in section v.

II. HOUSEHOLD DATA SET

For any machine learning task consists of three stages. They are data pre-processing, training and testing [7]. In the preprocessing stage, the data is checked for weather it may contain noise, missed values to correct it. In the training phase, we train the system using train data to create the models. In the testing phase, we predict the accuracy of the system. To develop the household audit system, data of different houses energy consumed is collected. The data used to train the system is taken from www.data.gov web site, which is collected during residential energy consumption survey. This is a national sample survey that collects energy related data for household. The data consists of 12083 households. The data consists of 15 attributes.

Data was collected from 12,083 households' selected at random using multistage, area probability sample design. Dataset has 935 factors that affect energy consumption of a housing unit out of which 16 attributes are considered. All the independent and dependent attributes are numeric.

The target variables kwhcol, kwhwth, kwhrfg, kwhoth are continuous i.e, energy consumed for cooling, water heating, refrigerator and other purposes.

The next step was to explore and clean the data. In order to clean the data, we must first find out whether there are any missing values in the data considered. This can be done by calculating the mean of missing values with respect to each attribute in the dataset. This function yielded the value zero for all the attributes which means that there aren't any missing values.

The next task was to perform some analysis on data such as checking for the dimensions of data and getting various statistics. Finally, a heatmap was created to show whether there is strong correlation between target variables and any independent variables. This heatmap is a graphical representation which not only gives the correlation but helps us visualise in a better way through the use of different colors. If the value between any two variables is greater than or equal to 0.5, then we can say that those two variables are correlated. From this inference below Fig.1 can be carefully analysed to identify strongly correlated variables.

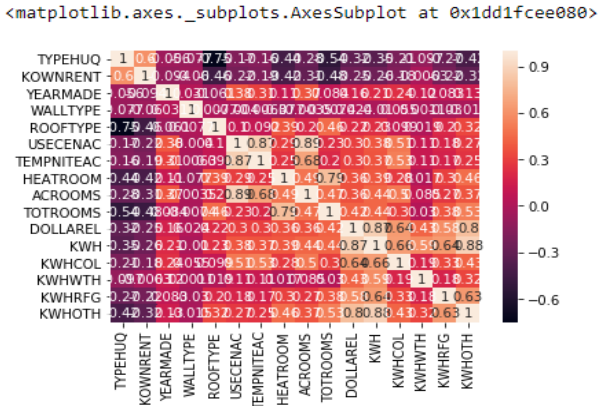

Fig.1. Heatmap of rankings for each variable

From Fig.1 it can be observed that the independent variables considered are (tempniteac, heatrooms, acrooms, totrooms, dollarel, kwh) i.e, temperature, number of rooms with heater, number of rooms with ac, total number of rooms, total cost for energy consumed and total energy consumed respectively since they are strongly correlated with target variables

The next step is to train the energy audit systems using machine learning algorithms. The machine learning models should be fit on training data to learn parameters of the model and then they can be run on test set to get the prediction [8] [9].

III. MACHINE LEARNING ALGORITHM FOR RIDICTING ENERGY CONSUMPTION

In this section the brief discussion of the machine learning algorithms such as linear regression, decision tree and random forest for energy consumption is discussed [10]. The different algorithms described below are different ways of deciding how to predict the energy consumed for various appliances [11].

Linear Regression*:* Linear Regression is based on supervised learning which performs a regression task. This method is fastest to predict the energy consumption in houses [12]. Regression models a target prediction variable based on independent variable. Linear Regression performs the task to find dependent variable value(Y) variable based on independent variable value(X) and finds out a linear relationship between x (input) and y (output). Mathematically the linear relationship can be described using

$$Y= \beta_0 + \beta_1 X+ \varepsilon$$

Where X and Y are input and output data respectively, β_0 is intercept and β_1 is coefficient of X. Once we find the best β_0 and $\beta_1$values using training data, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of Y for the input value X. In the considered dataset the values of X will be inputs (independent variables); i.e., tempniteac (temperature at night (summer)), heatroom (number of rooms heated), acrooms (number of rooms cooled), totrooms (total number of rooms), dollarel (electricity amount for entire energy consumption), kwh(total site electricity usage, in kilowatt-hours) and the corresponding Y values(dependent variables) will be kwhcol (electricity usage for air-conditioning, central and window/wall (room) in kilowatt-hours), kwhwth (electricity usage for water heating), kwthrfg (electricity usage for refrigerator), kwhoth (electricity usage for others(except col,wth,rfg). These four variables are dependent variables . For each dependent variable a linear model is created and thus (2) is used to compute first output (kwhcol) becomes

$$Y_1= \beta_0 + \beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\beta_4 X_4+\beta_5 X_5+\beta_6 X_6 \qquad (2)$$

Where X_1, X_2, X_3, X_4, X_5, X_6 are the values of inputs and Y_1 is the value of first output. Similarly values of Y_2, Y_3, and Y_4 are calculated. So for each output a linear model is built to compute the value.

Decision Tree Regression: Decision tree approach is most commonly used data mining method and consists of two steps, namely regression and classification [13]. During regression decision tree having a flow chart like structure, in which each internal (non leaf) node represents a test on an attribute, each branch shows the outcome of a test and each leaf (terminal) node holds a class label[14]. The topmost node in the decision tree represents the root node. In the decision tree process, it takes generation algorithm ID3 takes the training data as input and produces decision tree model as output. The decision tree regression process observes feature of training data and trains a model to predict data in the feature to produce meaningful continuous output. Decision tree regression can be done by importing DecisionTreeRegressor from sklearn and creating an instance of this regressor to fit the training data.

Random Forest Regression*:* The random forest approach is based on a decision tree [15]. Random forest use feature bagging techniques to tree learners, which decreases the variance of the model at the same time keeping the bias low. The random forest approach grows many decision trees in parallel and at the same time select random subset of features for split procedure. A decision tree predicts very fast but it suffers to over fitting [16][17].

Python provides the random forest regression package is sklearn. ensemble. Random Forest Regressor. Random forest repressors' are one of the ensemble learning methods for regression. It constructs multiple decision trees (a forest) at training time and to give an accurate result. The basic idea is to combine multiple decision trees in determining the output rather than relying on individual decision trees.

IV. ENERGY AUDIT SYSTEM MODELS CREATION

In any machine learning model that we build, data is divided into two parts; 70% training data and 30% testing data. Using training data the models are created i.e., a machine learns using the data given from training data. Then testing data is used to determine the model's performance and to draw conclusions about predictive capability. This can be done with a sklearn.model_selection by importing train_test_split call by specifying split ratio. The training set was then fit into various models. To predict the energy consumed, the machine learning system must use some criteria to make its decision.

Using the training data, linear regression, decision tree and random forest models are created. After creating the model, the each model is evaluated. Using test data the accuracy of the each model is calculated. This shows how well a models are trained to predict the values accurately.

Table1. The accuracy in p of the energy audit system for linear regression, Decision tree and Random forest models.

s.no	Algorithm used	Accuracy
1	Linear Regression	82.0%
2	Decision Tree	86.0%
3	Random Forests	91.0%

V. RESULTS AND PERFORMANCE ANALYSIS

After training the machine learning models, the testing is done against each model. First the linear regression energy audit model is predicts 82% energy consumption. Same procedure is repeated for decision tree model, and random forest models are predicted the energy consumption 86%, and 92% respectively.

The performance analysis is done on all the energy audit systems. From the analysis it is found that random forest model outperformed among the model. This is because any Random forest algorithm borrows the bagging concept of random feature selection to construct decision tree models with controlled variance. This method uses training datasets to create multiple decision trees on which results are evaluated and thus producing accurate results.

VI. CONCLUSION

In this paper, a energy audit system using linear regression, decision tree and random forest are proposed. The machine algorithms such as linear regression, decision tree, multivariate regression, random forest are used to create the energy audit models. For each system one model is created. The linear regression energy audit model is predicted 81% energy consumption. Same procedure is repeated for decision tree model, multivariate regression model, and random forest models are predicted the energy consumption 78%, 79.2% and 90% respectively.

The model's accuracy is subject to change with change in the infrastructure. Thus, the model needs to be updated with new data at finite intervals. It is observed that Random Forest gave good accuracy among all those models. Hence that model is used to predict the energy consumed.

REFERENCES

1. Gonzalez-Briones, A., Hernandez, G., Corchado, J. M., Omatu, S., & Mohamad, M. S. "Machine Learning Models for Electricity Consumption Forecasting:" A Review. 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS).
2. Zhao, H.-x. and F. Magoulès, A review on the prediction of building energy 137 consumption. Renewable and Sustainable Energy Reviews, 2012. **16**(6): p. 3586-138 3592.
3. E. Barbour and M. González, "Enhancing household level load forecasts using daily load profile clustering," in Proceedings of the 5th Conference on Systems for Built Environments, 2018, pp. 107-115: ACM.
4. K. Gajowniczek and T. Zbkowski, "Short term electricity forecasting using individual smart meter data," Procedia Computer Science, vol. 35, pp. 589 – 597, 2014.
5. I.I. Attia and H. Ashour,Energy Saving Through Smart Home," The Online Journal on Power and Energy Engineering (OJPEE), vol. II, no. 3, July 2011.
6. Lombard, L., J. Ortiz, and C. Pout, A review on buildings energy consumption 129 information. Energy and buildings, 2008. 40(3): p. 394-398.
7. Jihoon Moon, Jinwoong Park, Eenjun Hwang, Forecasting power consumption for higher educational institutions based on machine learning. The Journal of Supercomputing August 2018; 74(8): 3778–3800
8. Chang, H.-C.; Kuo, C.-C.; Chen, Y.-T.;Wu,W.-B.; Piedad, E.J. Energy Consumption Level Prediction Based on Classification Approach with Machine Learning Technique. In Proceedings of the 4th World Congress on New Technologies (NewTech'18), Madrid, Spain, 19–21 August 2018; pp. 1–8.
9. K. Aurangzeb, "Short Term Power Load Forecasting using Machine Learning Models for energy management in a smart community," 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019,
10. H. Zhao and F. Magoul`es. "Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method". In: Journal of Algorithms & Computational Technology 6.1 (2012), pp. 59–77.
11. Edwards, R.E., New, J., Parker, L.E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, *49*, 591–603
12. LAM, J. C., CHAN, R. Y. C. & LI, D. H. W. 2002. Simple Regression Models for Fully Air-conditioned Public Sector Office Buildings in Subtropical Climates. Architectural Science Review, 45, 361-369.
13. Yu, Z., F. Haghighat, B. C. Fung, and H. Yoshino (2010). A decision tree method for building energy demand modeling. Energy and Buildings 42 (10), 1637-1646.
14. Tso, G. K. and K. K. Yau (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks.Energy 32 (9), 1761-1768.
15. Ahmad, T. and H. Chen, Nonlinear autoregressive and random forest approaches to 198 forecasting electricity load for utility energy management systems. Sustainable Cities 199 and Society, 2019. 45: p. 460-473
16. Ahmad, M.W., J. Reynolds, and Y. Rezgui, Predictive modeling for solar thermal 192 energy systems: A comparison of support vector regression, random forest, extra 193 trees and regression trees. Journal of Cleaner Production, 2018. 203: p.810-821.
17. G. Dudek. "Short-Term Load Forecasting Using Random Forests".In Intelligent Systems'2014: Proceedings of the 7th IEEE International Conference Intelligent Systems 2014, Warsaw, Poland 2 (2015), pp. 821–828

AUTHOR PROFILE

Dr. A. Nagesh, is currently working as a professor in CSE at MGIT, Hyderabad. He completed B.E and M.Tech from Osmania University, Hyderabad in 1996 and 2002 respectively. He did Ph.D in CSE from JNTUH, Hyderabad in the year 2012. He is having total 24 years of teaching experience. At present he his supervising five Ph.D students. Total he is having 60 publications in national & international journals. His research areas includes pattern recognition, speech processing and machine learning.