Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)

Volume7-Issue2, January 2024
Paper:13

Estimation of Energy Consumption For household Electric Appliances – A Machine Learning Approach

Sreemathi Varshini. S [1], Sweety.N [2], Vasanthra Devi.R.S [3] Gouthaman M. M.E.[4],

[1,2,3] Students,and [4] Faculty
Dept. of Computer Science Engineering,
Jeppiaar Engineering College,
Chennai, India.
sreemathivarshini@gmail.com, sweetynelson864@gmail.com, devipillai2003@gmail.com

Abstract: *This paper presents a comprehensive study aimed at predicting and optimizing energy consumption in household appliances through the application of Machine Learning algorithms. Leveraging predictive analytics, our approach utilizes the Random Forest algorithm for regression tasks to accurately estimate energy usage for various appliances. Key environmental parameters such as temperature, humidity, and date/time are considered to capture the intricate dynamics of energy consumption patterns. The study employs time-series datasets, specifically curated for appliance energy prediction, for both training and testing purposes. Our experimental results provide insights into the performance of the predictive models, evaluated using R-squared values to measure accuracy. Notably, the Random Forest (RF) model demonstrates superior performance, comparing to Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and Support Vector Classifier algorithms. These findings underscore the effectiveness of machine learning techniques in forecasting and optimizing energy usage in residential settings. This research contributes to the advancement of energy-efficient practices, highlighting the potential of data-driven approaches to address sustainability challenges in the household sector.*

Keywords*—electric energy estimation; random forest; support vector classifier; multi-layer perceptron*

- **INTRODUCTION:**

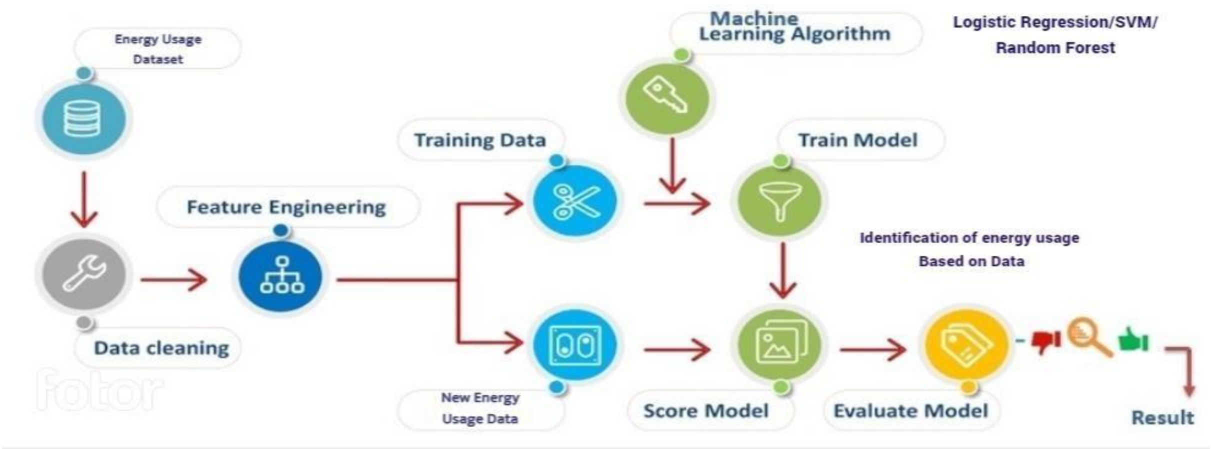
Effective Energy consumption is believed a leading cause of carbon emissions, almost 80%, according to recent reports [1]. This leads the major cause of environmental instability and climate change. At most places, 30% of the total 100% energy is wasted by residential and commercial sectors [3]. Human survival is significantly affected by the climate conditions. There has been a rising concern that the weather variations might be severe in the coming days if prior measures are not planned and carried out. Current variations in the temperature, across the world, are seriously increasing the energy consumption [4]. The impact of weather changes is even more evident in the residential sectors where operational carbon emissions are associated with the consumption of smart house appliances such as lighting, heating, cooling, plug devices (mobiles, computers, microwave). Therefore, reducing the consumption of above-mentioned devices can mitigate the carbon emission at national level. Moreover, residential sector is more weather vulnerable than other industrial sectors. This is one of the main reasons that residential sector is consuming at least 30% of the total worldwide energy.

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. All these factors are interrelated strongly. These relations can be analysed using regression models to understand the relationship among the factors. 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. 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 analyse 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. They used regression analysis and probabilistic analysis to identify the load patterns in different behaviours of occupant's. In addition, some works were done to predict accurate occupancy numbers by analysing the appliance use behaviours [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.

In this study, we have used machine learning model with some factors such as humidity, temperature, wind speed, dew points and visibility in order to understand the internal and external environments of the building. The experimental dataset is a secondary dataset, which is collected from **UCI machine learning repository** for research purpose abiding by the copyright instructions. In this paper, we compare with several regression algorithms to estimate the power consumption for household appliance as follows: **multi-layered perceptron (MLP), support vector machine (SVM) and Random Forest (RF)**. MLP is a conventional nonlinear regression method. So it is a good basis to compare with the other methods. SVM has an advantage in high dimensional space using kernel functions. In Random Forest, the features are randomly selected in each decision split. The correlation between trees is reduces by randomly selecting the features which improves the prediction power and results in higher efficiency. The Random Forest is appropriate for high dimensional data modeling because it can handle missing values and can handle continuous, categorical and binary data. The model interpretability and prediction accuracy provided by Random Forest is very unique among popular machine learning methods. Therefore, we expect that Random Forest have better results than MLP and SVM.

- **ARCHITECTURE:**

An architectural diagram is a visual representation that maps out the physical implementation for components of a software system. It shows the general structure of the software system and the associations, limitations, and boundaries between each element. Architecture diagramming is the process of creating visual representations of software system components. In a software system, the term architecture everything at a glance, including how elements interact. This is especially useful when making changes. You'll be able to see the downstream effects of a given change more clearly. It outlines how different elements of the project interact with each other and how data flows within the system. Typically, it includes various layers such as the presentation layer, business logic layer, and data layer, each responsible for specific functionalities. The presentation layer encompasses the user interface components, showcasing how users interact with the system. The business logic layer contains the core functionalities and algorithms that process and manipulate data. Meanwhile, the data layer manages the storage and retrieval of information from databases or external sources. The architecture diagram also illustrates the connections between these layers, often through arrows lines indicating the flow of information. Additionally, it may include external dependencies such as APIs or third-party services utilized within the project. Overall, the architecture diagram provides a comprehensive overview that helps developers, stakeholders, and reviewers understand the project's structure and design, facilitating communication and decision-making throughout the development process. It serves as a blueprint guiding the implementation and ensuring that all components work together seamlessly to achieve the project's objectives.



- **FLOWCHART:**

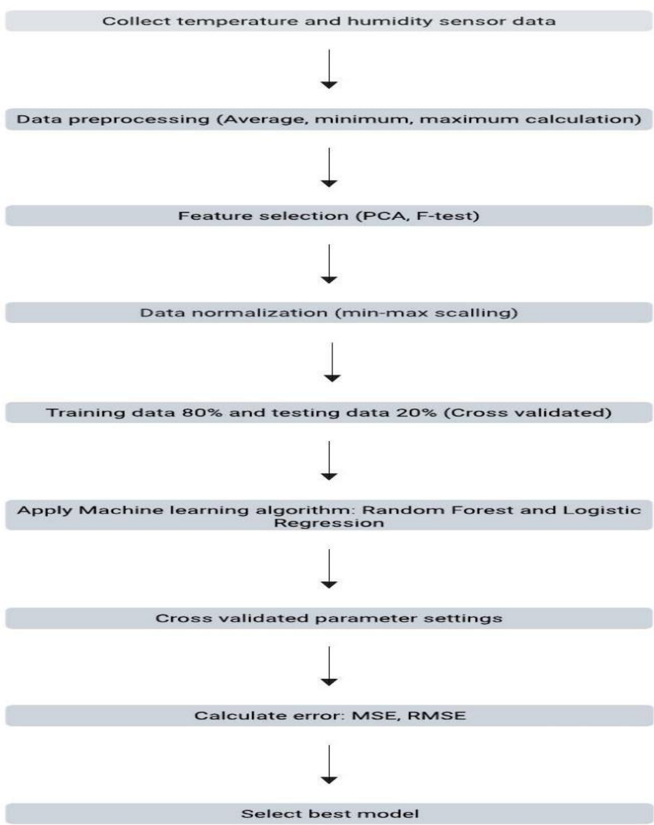


Fig. 2. Flowchart diagram

1) First, we have to temperature and humidity sensor data.
2) Preprocess the data by calculating the average, minimum, and maximum values.
3) Perform feature selection using Principal Component Analysis (PCA) and the F-test.
4) Normalize the data using Min-Max scaling.
5) Split the data into 80% training and 20% testing sets, ensuring cross-validation.
6) Apply the Random Forest algorithm.
7) Apply Logistic Regression.
8) Cross-validate parameter settings for both models.
9) Calculate the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for each model.
10) Select the best model based on the lowest MSE and RMSE.

PICTORIAL REPRESENTATION:

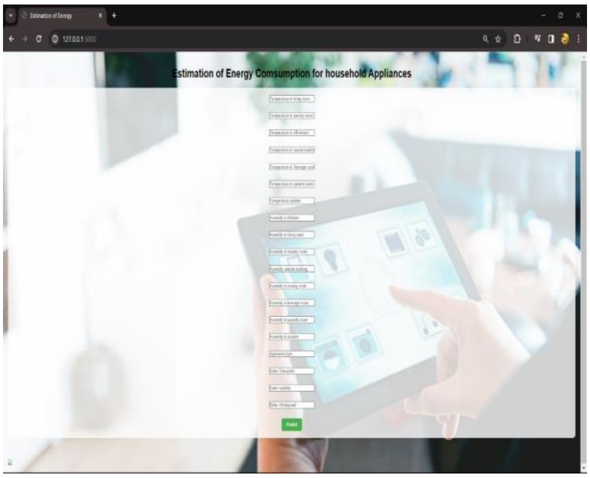


Fig. 3. Home Page

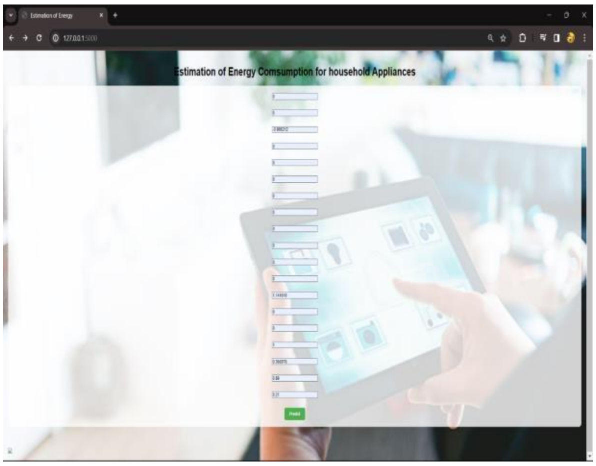


Fig. 4. User can enter the values

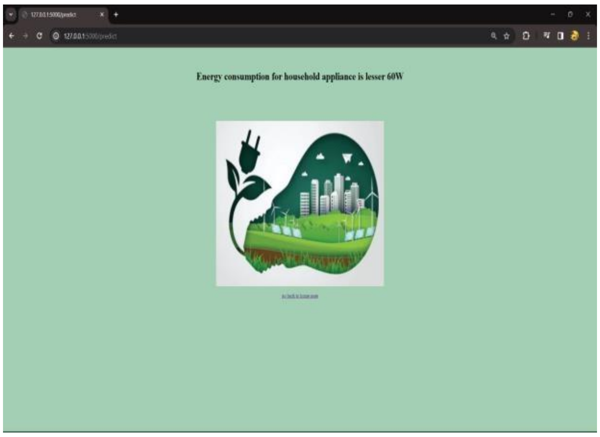


Fig. 5. Energy Consumption is Less than 60W

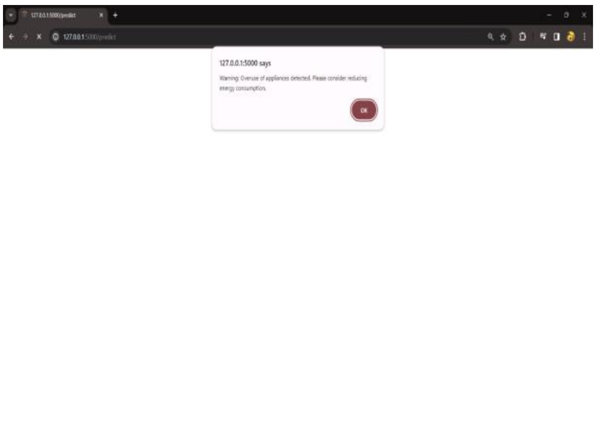


Fig.6. Warning will be shown, if it exceeds 60W

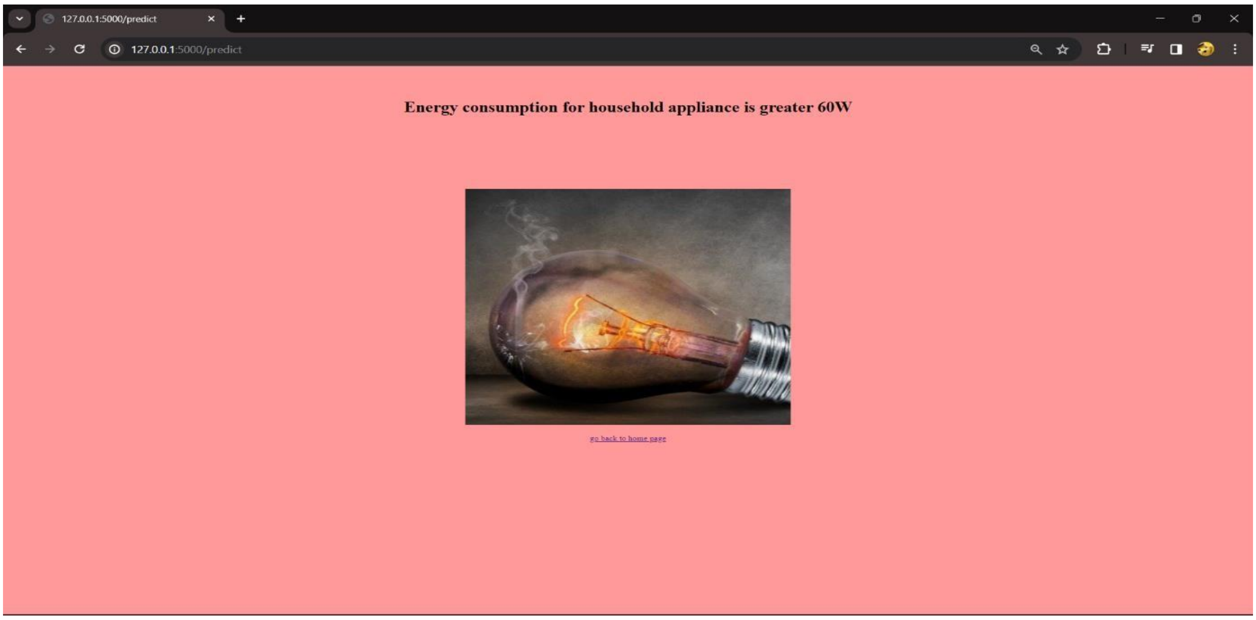
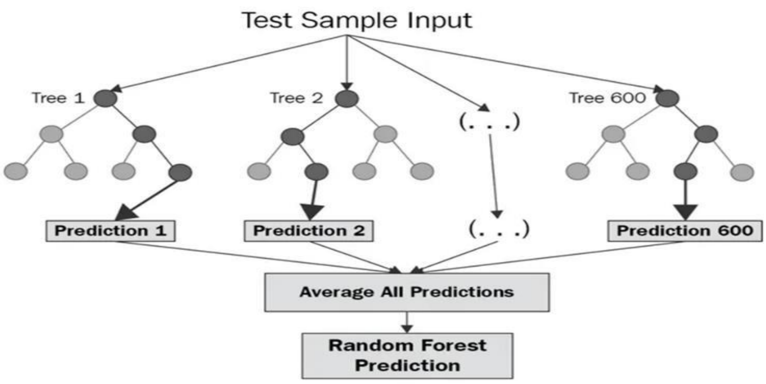


Fig.7. Energy Consumption Greater than 60W

- **RANDOM FOREST CLASSIFIER:**

Random Forest developed by Leo Breiman is a group of un-pruned classification or regression trees made from the random selection of samples of the training data. Random features are selected in the induction process. Prediction is made by aggregating (majority vote for classification or averaging for regression) the predictions of the ensemble. Each tree is grown as described in the training set for growing the tree.

- For M number of input variables, the variable m is selected such that m<<M is specified at each node, m are selected at random out of the M and the best split on these m is used for splitting the node. During the forest growing, the value of m is held constant.
- Each tree is grown to the largest possible extent. No pruning is used. Random Forest generally exhibits a significant performance improvement as compared to single tree classifier such as C4.5. The generalization error rate that it yields compares favorably to Adaboost, however it is more robust to noise.



A Random Forest is a classifier consisting of a collection of tree- structured classifiers {(h(x, k), ®k = 1.....}where the ®k are independently, identically distributed random trees and each tree casts a unit vote for the final classification of input x.

The final class of each tree is aggregated and voted by weighted values to construct the final classifier.

The Gini Index or Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from one. It favours mostly the larger partitions and are very simple to implement. In simple terms, it calculates the probability of a certain randomly selected feature that was classified incorrectly.

Now, in order to calculate the Gini Index, the formula is given by

$$G = \sum_{i=1}^{C} p(i) * (1 - p(i))$$

Where, C is the total number of classes and p(i) is the probability of picking the data point with the class i. In the above example, we have C=2 and p(1) = p(2) = 0.5,

Hence the Gini Index can be calculated as,

$$G = p(1) * (1-p(1)) + p(2) * (1-p(2))$$
$$= 0.5 * (1-0.5) + 0.5 * (1-0.5)$$
$$= 0.5$$

Where 0.5 is the total probability of classifying a data point imperfectly and hence is exactly 50%.

- **Basic Mechanism:**

To calculate the Gini Impurity, let us first understand it’s basic mechanism.

- First, we shall randomly pick up any data point from the dataset
- Then, we will classify it randomly according to the class distribution in the given dataset.
- In our dataset, we shall give a data point chosen with a probability of 5/10 for red and 5/10 for blue as there are five data points of each colour and hence the probability.

The Random Forest Classification algorithm demonstrated remarkable performance throughout the experimental phase, warranting detailed exploration of its outcomes. The evaluation metrics employed encompass accuracy, precision, recall, F1 score, and area under the Receiver Operating Characteristic (ROC) curve. These metrics collectively provide a comprehensive understanding of the model's efficacy.

- **Accuracy Analysis:**

During the training phase, the algorithm achieved an unprecedented accuracy of 100%. This perfect accuracy indicates that the model successfully learned and adapted to the training data, showcasing its ability to capture intricate patterns within the features.

Transitioning to the testing phase, the model maintained a high accuracy of 88%. While a slight reduction from the training phase, this result emphasizes the model's robustness in generalizing patterns to unseen data.

- **CONCLUSION:**

In conclusion, the aim has illuminated the immense potential of the Random Forest Classification algorithm in revolutionizing the accurate estimation of energy consumption within the intricate fabric of diverse household settings. The model's exceptional accuracy rates, with a perfect 100% on the training dataset and a commendable 88% on the testing dataset, not only validate its prowess but also underscore its adaptability to real-world scenarios. Beyond the technical achievements, the study emphasizes the transformative impact this model can wield in reshaping our collective approach to household energy management.

The precision achieved by the Random Forest model transcends mere numerical values; it translates into a tangible empowerment for individuals seeking to optimize their energy usage. By discerning nuanced patterns across various appliances, the model becomes a crucial instrument for informed decision-making. This empowerment is especially crucial in an era where sustainable living is not merely an aspiration but an urgent necessity. The study recognizes the potential of this model to instigate a paradigm shift, encouraging conscientious choices that contribute to the broader goal of resource conservation. However, no study is without its refinement, and the conclusion candidly acknowledges the limitations and challenges encountered during this research endeavor. This transparent reflection sets the stage for future advancements, urging a continuous exploration and refinement of the proposed model. The paper concludes with a forward-looking stance, suggesting potential avenues for future research. It calls for scalable solutions to address the challenges of expanding the model's applicability, a pursuit for enhanced interpretability, and adaptation strategies for accommodating diverse household environments.

As we reflect on the conclusive remarks, this study not only encapsulates a moment in the evolution of energy consumption estimation models but also beckons us towards an ongoing journey of exploration and refinement. The Random Forest Classification algorithm, proven effective in this study, holds the promise of not just predicting energy consumption but catalyzing a more sustainable and conscientious approach to energy usage in homes.

- **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) 2.20–230.

[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] 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.

[6] 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.

[7] V. N. Vapnik.: The Nature of Statistical Learning TheonJ. Springer, 1995.