International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org
Vol. 10 Issue 3, March-2020,
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196 |

SUPPORT VECTOR REGRESSION FOR PREDICTING ENERGY USAGE IN MULTI-FAMILY DWELLINGS

Battula Sridhararao[1], Dr. Rajeev Yadav[2]
Department of Computer Science
[1,2]Sunrise University, Alwar, Rajasthan

Abstract

Energy consumption and air pollution are primarily the result of the construction industry in urban areas. If we want to reduce emissions through urban energy management and efficiency measures, need to be able to accurately predict and characterise a building's energy usage. As smart metering technology has advanced, researchers have developed "sensor based" strategies for estimating a building's future energy needs with significantly less information than traditional approaches. This study applies the well-known machine learning method of Support Vector Regression (SVR) to an empirical data-set gathered from a high-rise apartment complex in one of India's main cities. The study also looks at how using daily, hourly, and 10-minute intervals affects the single-step model's ability to predict the future. The results demonstrate that sensor-based predictionsimulations can be used in high-density residential buildings, and that floor-level and hourly monitoring granularity are optimal for identifying potential problems. Practical applications include facilitating the distribution and installation of state-of-the-art smart metres, while theoretical applications include aiding in the creation of home energy prediction models. If we want to create reliable and affordable energy projections on a global scale, it is imperative that next-generation energy efficiency programs account for not only the methodology;however, the ranges at which data can be transformed into intelligence data.

Keywords: Energy Conversion,Monitoring forecastingresidential areas SVR

1. Introduction

Energy usage in the constructedatmosphere is equivalent to 40% of total depletion in India and other countries. Greater than seventy-five percent of all GHG emissions come from buildings in high-density metropolitan areas like Delhi, Mumbai, and other metro cities **(Petersen, &**

Svendsen, 2010). That's why, if metropolitan areas want to cut back on their overall energy use, they need to characterise, model, and anticipate the energy consumption of their buildings. Energy management and efficiency applications benefit greatly from accurate modelling and forecasting of building energy demand, including informing initial design decisions, predicting advances in building energy consumption, optimising the ventilation system in a building, and planning the urban electricity grid**(Howard et al 2010)**

In the past, engineers have used software like Energy Plus to predict building-wide energy needs by compiling extensive data about the building's geometry, geometry, and materials. Widespread energy forecasting is hampered by the difficulties of acquiring and validating such data. Consequently, there has been a rise in the popularity of "sensor based" methods, which eschew the stringent conceptual interpretations that demand feedback of a more manageable collection of granular, historically-based measurements.

Smart metre, building management system (BMS), and weather station data is fed into a machine learning algorithm to help uncover hidden patterns in energy consumption as a function of factors like temperature, occupancy, and time of day. Researchers have discovered that sensor-based energy forecasting methodologies are as accurate as more conventional engineering-based methods, and even more so in some cases, and require considerably less user input. Sensor-based systems have grown in applicability and cost-effectiveness as low-cost, easily available solutions for energy metering have rapidly multiplied in recent years. This research expands on the benefits of sensor-based energy prediction over more conventional technicalapproaches**(Edwards, & Parker 2012).**

Sensor-based prediction has found broad usage in non-residential dwellings, its potential utility in the home construction sector has been the subject of less research. In the limited amount of writing that has been done about the residential sector, single-family residences have received the bulk of attention, whereas multi-family dwellings have been mostly ignored. To effectively execute citywide energy efficiency and conservation measures, it is essential to precisely anticipate the consumption of multi-family buildings, which make up the bulk of the housing stock in dense metropolitan locations.

2. Objective of the study

The objectiveof the studyis to balance the monetary and opportunity expenses of mounting and working monitoring equipment against the forecasting accuracy gained by finer-grained energy usage monitoring.

3. Methodology

One of our primary focuses is figuring out whether or not sensor-based, one-step energy consumption forecasting is applicable to the multi-family home market. It decided to construct a model using Support Vector Regression (SVR) **(Do, &Thi, 2021**) and validate it against the outcomes of the Great Energy Predictor Shootout with the help of the Scikit-learn module **(Rehman et al., 2020),** a Python interface to LIBSVM, a popular Support Vector Machine library.

4. Description of Model

Inputs M1 and M2 will be defined for the model in this research; M1 will be used to validate the model with figures from the GEPS, and M2 will be used to test the model with our own empirical data. The following conditions must hold true for M1 to be consistent with its inputs:

$$M1\colon \overrightarrow{x(t)} = [y(t-1), y(t-2), T(t), s, sh, ch] \ldots \ldots .. (1)$$

The sine of the current hour is denoted by sh, and the cosine by ch; the current temperature is denoted by $T(t)$and the solar flux at this time by $S(t)$the indicator variable s can be used to determine whether or not it is a weekend or holiday.

Based on previous work **(Li. 2011)** and the unavailability of solar flux data, we apply the following modifications to the inputs in order to test the model on the Watt Hall data set:

$$M2\colon \overrightarrow{x(t)} = [y(t-1), y(t-2), T(t), s, sh, ch] \ldots \ldots \ldots .. (2)$$

CV Metric Definition

In this equation, N represents the total number of observations, $\hat{y}_i$represents the predicted value, $y_i$represents the observed value, y represents the median outcome, and $\bar{y}$ is the median outcome.

$$\mathrm{CV} = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{\bar{y}} * 100 \ldots\ldots\ldots\ldots(3)$$

4.1SVR and model Assessment

4.1.1 SVRMethod

Support Vector Regression (SVR) stands out among methods for predicting continuous variables due to its remarkable generalisation when used with new data. With the help of the support vectors, SVR is able to achieve excellent consistency with only a subset of the training data. SVR can distinguish the support vectors from the other training data because it uses a discerningdamagetask that does not penalise residuals below a tolerance. Predictions for a specific hypothesis and are thus unaffected by data that are restricted within the "-tube" surrounding the hypothesis. An illustration of a so-called "e-tube" is shown in **Fig. 1**.

Minimize:

$$\frac{1}{2} \parallel \omega \parallel^2 + C\frac{1}{T}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \ldots\ldots\ldots(4)$$

to which the vector of feature weights ω refers. Reducing the meanω of is indicative of a simpler, "flatter," theory. Both ξ_i and ξ_i^*measure how much residuals exceed the tolerance set by the problem, and they ensure that the problem can be solved for all ε(illustrated in **Fig. 1**). The amount of linear penalty given to the residual excess $\xi_i^{(*)}$.is controlled by the regularisation term *C*. Most commonly, SVR uses the "kernel trick" to transfer the input space into a higher

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org
Vol. 10 Issue 3, March-2020,
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196 |

dimensional feature space using a kernel function ϕ, allowing for a computationally efficient non-linear scenario. Therefore, the following limitations apply to Eq. (4):

$$y_i - \omega \cdot \phi(x_i) - b \leqslant \varepsilon + \xi_i \dots \dots \dots (5)$$

$$\omega \cdot \phi(x_i) + b - y_i \leqslant \varepsilon + \xi_i^*, i = 1, \dots, l \dots \dots \dots \dots (6)$$

$$\xi_i^{(*)} \geqslant 0 \dots \dots \dots (7)$$

This study was conducted using an epsilon-SVR model, which is worth noting. The RBF kernel is described by the following equations **(Zhao &Magoulès, 2012)**:

$$\varphi(x, \chi') = \exp(-\gamma|x - x'|^2), \quad \gamma > 0$$

the kernel parameter γ being denoted by For each data point, γ defines its effective radius. Notably, in the RBF implementation, C, ε and γ are all user-defined variables that significantly affect the SVR result.

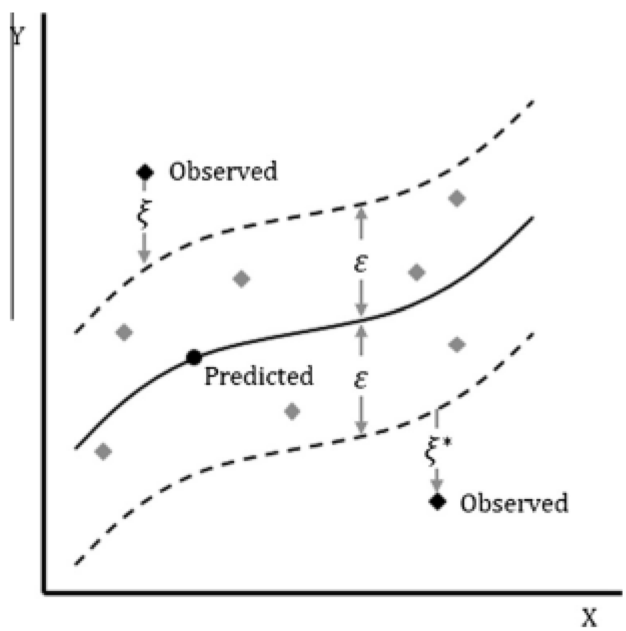

Fig. 1. SVR Variables**(Li et al 2009).**

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org
Vol. 10 Issue 3, March-2020,
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196 |

4.1.2 Model optimization

To be more precise, the training tolerance and the regularisation term *C* are two parameters that may be adjusted by the user to affect how well SVR achieves its aim function. For the Gaussian RBF to work, we also need a value for, which establishes how much weight to give to each support vector. Our model was found to be rather unaffected in the range $(10^{-5} \leqslant \varepsilon \leqslant 10^{-1})$, in both the GEPS and theexperimental datasets, while both, showed affective and required free fitting. While there are heuristic methods for selecting parameters in tiny dimensions, most practitioners instead resort to brute-force procedures, which involve the exhaustive exploration of a grid of parameter combinations and the subsequent grading of each occurrence based on the model's performance. Using k-fold cross validation on all possible combinations, we were able to avoid overfitting. k-fold cross-validation alternates between using the validation set and the training set by subdividing the training set into *k-1*progressively smaller subsets of almost equivalent size. Deterministic k-folds methodology, in contrast to randomised cross validation processes, enables robust comparison of results from diverse permutations of *C* and. Cumulative variance (CV) will be used across all folds (with=5) to determine relative performance. To do this, a recursive algorithm was created that iteratively moves through grids with progressively lower cell sizes if the difference in $|\Delta \mathrm{CV}_{\min}|$between rounds is more than a specific threshold (i.e., $|\Delta \mathrm{CV}_{\min}| > \varphi$ where$\varphi = 0.05$).

4.1.3 Algorithm overview

Figure 2 shown the Algorithm overview

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org
Vol. 10 Issue 3, March-2020,
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196 |

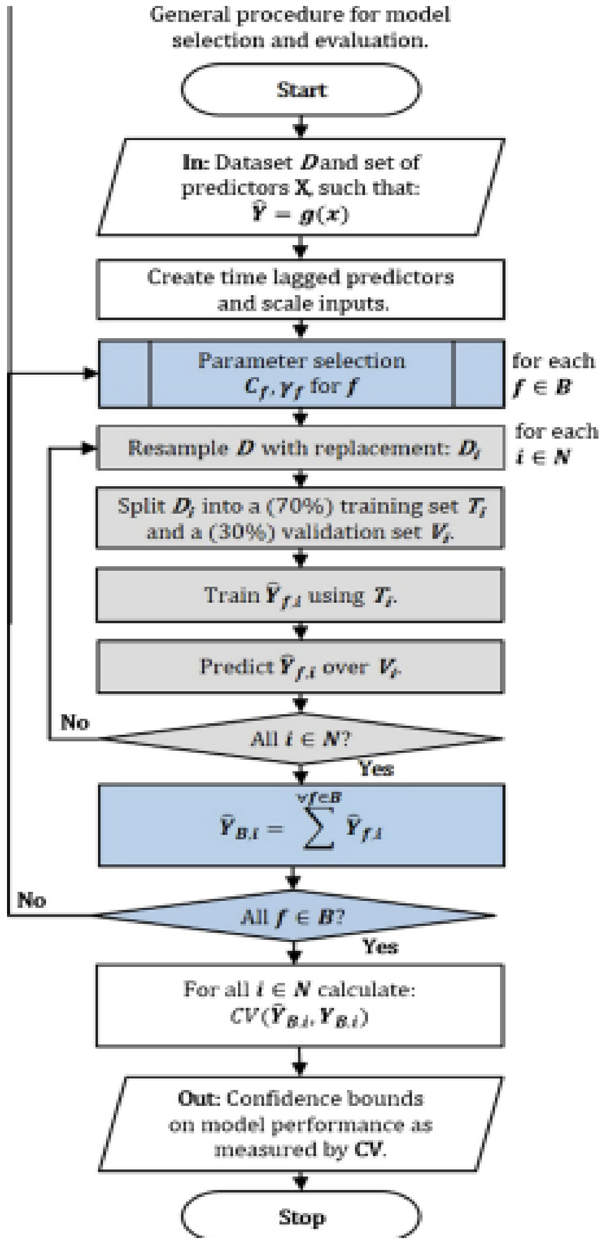

Fig. 2. Algorithm for selecting and validation models.

5. Results & discussion

Data from the GEPS was included in the programme, and its results were checked against those produced by Edwards to make sure they were accurate Starting in the middle of September, the GEPS dataset contains hourly data on WBE, temperature, sunlight, humidity, and wind speed. Both the authors' technique and the inputs of humidity and wind speed were overlooked (given in Eq. 1)

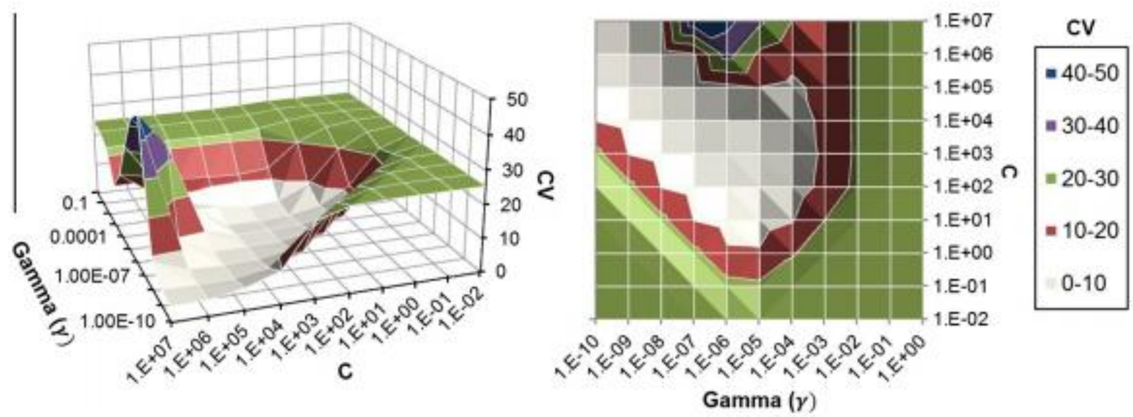

Fig. 4. Amplification of GEPS data using a grid search

In line with previous studies, we trained the model using 75% of the GEPS dataset and tested it using 25%. Successfully achieved results of “10,000 and $1\text{E}-5$ for C andγ", respectively, using grid-search optimization techniques outlined earlier. Outcomes are shown in Figure Figures 5 and 6 show the expected values (red bars) compared to the actual data (black bars). (break in the line) Results from the model showed a CV of 3.30 percent, which is less than what was found using the GEPS data set, demonstrating the superiority of our approach. For all save Case 3C, the results from the one-step predictive model fell within this reasonable margin of error. In light of these findings, it is clear that sensor-based forecasting algorithms, building on previous research, can be successfully applied to multi-family residential complexes.

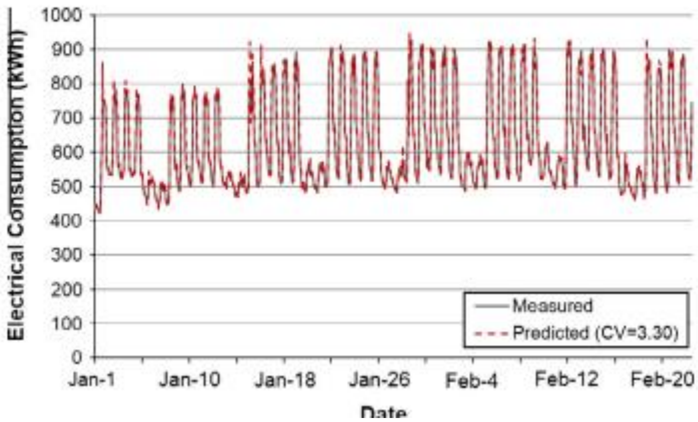

Fig. 5. The GEPS data collection, as predicted by the model.

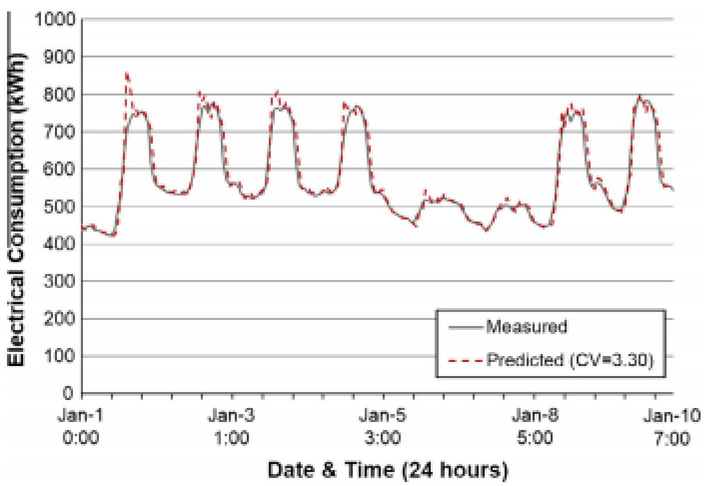

Fig. 6. Model's specific predictions for a limited time frame in the GEPS dataset.$(C = 10{,}000; y = 1\mathrm{e}-5)$

6. Conclusions

These results suggest that sensor-based energy forecasting models can be used to multi-family residential buildings, increasing the scope of earlier research into domestic energy forecasting beyond single-family homes. We went into the data to see how well sensor-based forecasting systems perform with more precise time and position information. Conclusions Hourly usage on a per-floor basis appears to be the most important factor in the accuracy of the models. Consequences for the widespread implementation of state-of-the-art smart metering technologies are also discussed, as these are essential for gathering the granular, high-resolution information on energy consumption that is required by the most precise prediction models. This study is an important first step toward more accurate energy forecasting in multi-family buildings using data collected from sensors.

Reference

Petersen, S., & Svendsen, S. (2010). Method and simulation program informed decisions in the early stages of building design. *Energy and buildings*, *42*(7), 1113-1119.

Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., & Modi, V. (2012). Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, *45*, 141-151.

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.

Rivero, C. R., Pucheta, J., Otaño, P., Juárez, G., Franco, L., Patiño, D., &Velazco, R. (2018, June). Bayesian enhanced ensemble approach (BEEA) for time series forecasting. In *2018 IEEE Biennial Congress of Argentina (ARGENCON)* (pp. 1-7). IEEE.

Rehman, A. U., Lie, T. T., Vallès, B., & Tito, S. R. (2020). Non-intrusive load monitoring of residential water-heating circuit using ensemble machine learning techniques. *Inventions*, *5*(4), 57.

Do, T. N., & Le Thi, H. A. (2021, December). Training Support Vector Machines for Dealing with the ImageNet Challenging Problem. In *International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences* (pp. 235-246). Springer, Cham.

Li K, Su H, Chu J. Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study. Energy Build 2011;43:2893–9.

Zhao H-x, Magoulès F. Feature selection for predicting building energy consumption based on statistical learning method. J Algor Comput Technol 2012;6:59–78

Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Applying support vector machine to predict hourly cooling load in the building. Appl Energy 2009;86:2249–56.