International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072

Energy Consumption Model for Educational Institutes in Kerala

Abhijith K N[1], Ashwin Ramesh[2], Sidharth Vasumithran[3], Sooryanunny A[4]

[1-4]*Students, Dept. of Electrical Engineering, Mar Athanasius College of Engineering, Kerala, India*

Abstract -*Due to the rapid economic growth around the world, energy consumption has risen to levels never seen before. Energy producers are struggling to keep up with the increasing demand. Therefore, there is an emphasis to improve the efficiency of energy usage. Improving efficiency helps in complying with regulations on energy use as well. Also, another added benefit of improving the energy efficiency of buildings is that Energy costs incurred by the user can be reduced as well.Energy modeling can help us in achieving that goal of increased energy efficiency since the models help to predict the energy consumption of a building to be constructed. However, the models or simulations currently available has few limitations such as accuracy due to the parameters they have considered for the formulation. In our project to overcome these limitations, this study has taken into account almost all the factors that influence the energy consumption of an institutional building. These factors are identified as independent variables in the model which is generated by machine learning tools. The result is that the model thus created will be quicker and also be highly accurate thereby satisfying the needs of the user.*

***Key Words*: Energy modelling, Machine Learning, Ridge linear Regression, Correlation Analysis, Matlab, Root Mean Square Error(RMSE), R-Squared**

1.INTRODUCTION

Around the world, energy consumption is very high. Most of it is due to the inefficient use of energy. In institutional buildings, this leads to an increase in the carbon footprint as well as operational costs. Studies show that energy cost is the second most significant expense of educational institutes, the first being the salary paid to staff. School buildings represent a major portion of government buildings that are under the domain of governmental finance. So engineers and designers are focusing on making the buildings as energy-efficient as possible. There is currently a lack of research in energy consumption modelling based in Kerala. Therefore the aim of this study is to correct this by developing a tool that is well suited to the conditions of Kerala and perform the following objectives.

- Offering a simple, accurate, and better tool for improving the future prediction of energy consumption, which contributes to more precise budget allocations.
- Improving energy control by offering a more accurate prediction model of future energy consumption.
- Preparing the ground for further studies in the area of the energy consumption of school buildings for both the public and private sectors.
- Provide Architects and Engineers the means to reduce energy consumption and improve energy efficiency by running multiple simulations and selecting that uses the least energy.
- Clean the collected energy data which is often noisy Hopefully, this effort can be carried over to other sectors as well and in the end, improve the energy sector of the country as a whole.

2. LITERATURE REVIEW

School buildings are vital assets that play an important role in the educational process. These buildings host students and stay throughout the day in different climate conditions. In light of that, the owners of such buildings spend a considerable portion of the assigned operation and maintenance budget keeping these buildings running in a healthy environment. Energy consumption represents a significant total running cost of this class of facilities.

2.1PREVIOUS STUDIES ON ELECTRICAL ENERGY CONSUMPTION

Most of the studies conducted earlier were focused on minimal number of factors that will affect the energy consumption of an educational institution. Also the number of sample institutions taken was not sufficient for an accurate model. The previous works mainly focused on the relationship of energy consumption with the gross floor area of the building[1]. But there are many other important factors which will have a significant impact in the energy consumption of an educational institution.

2.2 APPLICATION OF ARTIFICIAL INTELLIGENCE FOR PREDICTING ENERGY CONSUMPTION

Due to the advancement of machine learning and information technologies, novel computer application tools have recently advanced in construction and building management. In different fields the machine learning tools have received attention as one of the more attractive utilized tools for developing prediction models. The accuracy of the anticipated energy consumption is an important consideration with any method of estimation. A literature review suggested that machine learning tools

have been intensively adopted in the development of forecasting energy consumption models[1].

3. DATA DESCRIPTION

This Phase is concerned with the process of identifying the required factors for the study and the process of obtaining the relevant data. Also, the various characteristics and patterns of the collected data like the proportion of schools visited among others are observed and finally, the process of analyzing the data using various techniques like correlation analysis is looked at.

3.1 DATA IDENTIFICATION

Before moving onto Data Collection there was a need to identify what type of data is to be collected. That was done with the help of an extensive literature review and consultation with professionals working in this field. In the end, a list of factors including Occupancy, Floor Area, Ventilation Area, Connected Load, Temperature, Rainfall among others was drawn up. The next process was to chart a course on how the data was to be collected. Initially, Online Sources including Government web portals was accessed but it was not fruitful as these websites neither had all the data required nor was there any mechanism to ensure that this data was accurate and regularly updated. Therefore the only option was to physically go to the institutions and with their permission obtain the required Data.

3.2 DATA COLLECTION

Initially the target was to collect data from about 250 schools. But due to the current pandemic situation, it was only possible to visit 65 educational institutes, which include lower primary, high school, and higher secondary school. The percentage of different schools that covered is represented by a pie chart shown below

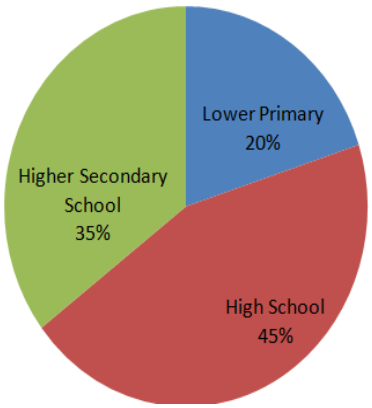


Fig 1:Type of Schools

The study mainly focused on different Schools in the districts of Kollam and Ernakulam. The data collected include the gross floor area, the height of rooms, and area of ventilation, number of occupants, temperature, connected load, and working hours. The consumer number was also collected from respective institutions and with that we collected the energy consumption of institutions in 2018 from the Kerala State Electricity Board (KSEB) office. For an easy application of the model, the data collection was organized in Worksheets in Microsoft Excel 2013.

The data were randomly divided, with 60 percent used as training data and 40 percent used for testing the model, while new data was used for model validation.

A sample of data collected from different institutions is shown below

SAMPLE OF DATA COLLECTED						
Name	Rooms	Area(m^2)	Occupancy	Vent. area(m^2)	Total load(kW)	Energy(kWh)
Veliyam West LP	8	130	120	21	3	85
Kottara LPS	6	219.8	70	27.5	2.769	40.83
Ayilara HS	35	1346	575	177.1	25.483	597.45
Perumkulam HS	36	1377	691	203.35	28.466	667.02
Azheekkal HS	37	1426.54	553	203.35	25.09	600.18
Chirakkara HS	30	1084	892	175.5	25.491	533.16
Vakkanad HSS	39	1463.78	532	160.5	17.035	318.17
Pallickal HSS	52	1966.435	996	252.75	28.593	593.38
Pattanakkad HSS	39	1442.6	1132	243	38.091	564.66

3.3 DATA ANALYSIS

Data Analysis/Correlation Analysis was performed and the purpose of the target(energy consumption) vs. input variable correlations analysis was to investigate the correlation between the model output, which was the energy consumption, and the model input variables. The analysis computed the relationship or "correlation coefficient value" between all the model inputs and the output (target) as can be seen in the below table.

	ENERGY CONSUMPTION
ROOMS	0.888500449
GROSS FLOOR AREA	0.8868947
OCCUPANCY	0.887171958
VENTILATION AREA	0.886791
CONNECTED LOAD	0.797293923
WORKING HOURS	0.43365
HEIGHT OF ROOM	0.24322
TEMPERATURE	0.224566

Table 3: Correlation Analysis

The weakest correlation was 0.2 for "Temperature". The maximum correlation was 0.89 for "number of classrooms". It should be noted that a positive value indicates a positive relationship, which means an increase in one variable leads to an increase in the other. On the other hand, a negative value indicates a negative relationship, which means an increase in one variable will lead to a decrease in the other. Also the reason "temperature" and "working hours" have weak correlation is because they have similar values for all institutions and this is because they fall under the same climatic zone and have fixed working hours.

4. MODEL FORMULATION AND VALIDATION

Due to the advancement of machine learning and information technologies, novel computer application tools have recently advanced in construction and building management. In various fields, machine learning has received attention as one of the more attractive and utilized tools for developing prediction models. Machine learning is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data to make predictions or decisions without being explicitly programmed to do so[2].

The application that was used for modeling was MATLAB. MATLAB is popular and widely adopted software that offers a whole lot of functionalities. Some toolboxes offer professionally developed, rigorously tested, field-hardened, and fully documented functionality for a wide range of scientific and engineering applications, apps that are interactive applications that combine direct access to large collections of algorithms with immediate visual feedback, and most importantly it is user friendly[7].

Before moving onto the modeling part, the collected data was divided into two sets the "test" dataset and the "train" dataset. The "train" dataset was used to train the model as the name suggests and the "test" dataset was used to gauge its accuracy. To find the right algorithm to be used for modeling using machine learning, a trial and error process was performed. Bayesian, Support Vector Machines, Decision Trees and many others were tried out but the best one was linear regression.

Linear regression is simply a linear approach to modeling the relationship between a scalar response and one or more explanatory variables (also known as dependent and

independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regressions. It is an attractive model because the representation is so simple

The representation is a linear equation that combines a specific set of input values or dependent variables (X) the solution to which is the predicted output or independent variables for that set of input values (Y). As such, both the input values and the output value are numeric. The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta (B). One additional coefficient is also added, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient. The equation takes the form of

$Y_P=B_0+X_1* B_1+X_2*B_2\ldots\ldots\ldots\ldots+X_n*B_n$

The terms of this equation are selected in such a way so as to reduce the loss function which is shown below

Loss function=$\sum(Y_T-Y_P)^2$

Where,
Y_T=True Value
Y_P=Predicted Value

However, all Machine learning algorithms tend to overfit the data. That is, they perform well on certain datasets but do poorly in others. This is because they closely mirror the patterns in test data sets, therefore losing their ability to generalize[2]. To avoid this issue, a variant of linear regression called ridge linear regression. It imposes a penalty term on the model to prevent it from overfitting the data[4]. The mathematical representation of it is

Loss function= $\Sigma(Y_T-B_0-B_i)^2$ – lambda $*$ ΣB_i^2
Where lambda can take any value from zero to positive infinity. This concept was implemented in Matlab using code shown in the following section.

5. RESULTS AND DISCUSSION

The model which is developed in Matlab using Ridge linear regression is
Energy Consumed= -28.2176 + (25.5573*Rooms) - (0.5124*Area) + (0.141*Occupancy) + (0.1094*Ventilation Area) + (12.1546*Connected Load)

5.1 EVALUATION OF MODEL USING STATISTICS

To evaluate the effectiveness of our model certain statistical parameters are used .Which are ,

R-squared is also known as the coefficient of regression is a statistical measure that represents the goodness of fit of a regression model[5]. The ideal value for r-square is 1. The closer the value of r-square to 1, the better is the model fitted. It can be computed using the following formula. Before computing R-squared we first need to calculate the sum of squared deviations from the mean of predicted value X(SSX), The sum of the squared deviations of Y from the mean of real value Y(SSY) and measures the correlation between y and x in terms of the corrected sum of products(SSXY). The equations for these terms and R-squared are given below

$SSX=\Sigma(X-Mx)^2$
Where X=predicted value and Mx =mean of X

$SSY=\Sigma(Y-My)^2$
Where Y=Actual value and My=mean of Y
$SSXY=\Sigma((X-Mx)(Y-My))$
R squared = SSXY/sqrt(SSX*SSY)

Rsquared value obtained for the model is 0.9758 which means the model was able to fit the data very well. **Adjusted Rsquared** is a refined version of Rsquared [5] and was calculated using the equation
$\mathbf{R_{adjusted}}$=1-((13*(sqrt(SSX*SSY)SSXY))/(8*sqrt(SSX*SSY)))

The value obtained was **0.961**

Residuals in a statistical or machine learning model are the differences between observed and predicted values of data. They are a diagnostic measure used when assessing the quality of a model. They are also known as errors. Residuals are important when determining the quality of a model. Ideally, residuals should be zero and not should form any patterns. Although our residual values were nonzero, there were no patterns in the residual.

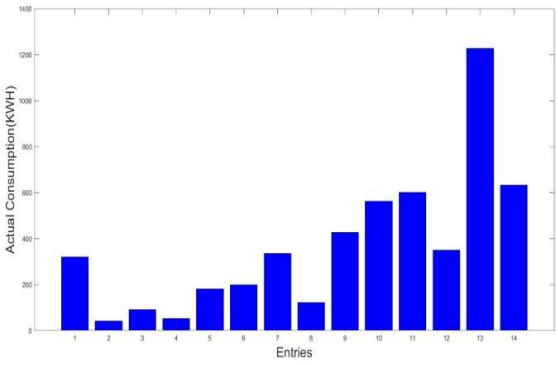
Root mean square error or root mean square deviation(RMSE) is one of the most commonly used measures for evaluating the quality of predictions. It shows how far predictions fall from measured true values using Euclidean distance[6]. In machine learning, it is extremely helpful to have a single number to judge a model's performance, whether it be during training, cross-validation, or monitoring after deployment. Root mean square error is one of the most widely used measures for this. It is a proper scoring rule that is intuitive to understand and compatible with some of the most common statistical assumptions. It can be calculated using the following equation

RMSE=sqrt(Σ(residual^2/n)
Where n is the number of observations

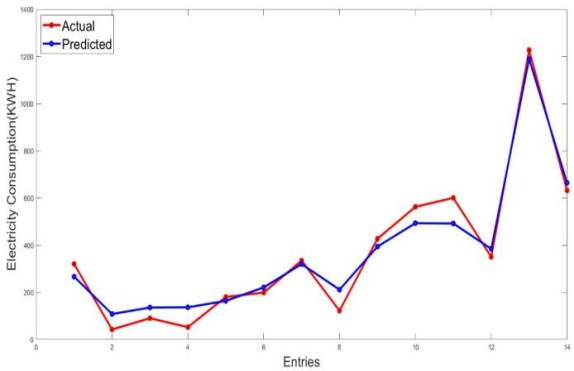
The RMSE value obtained was 57.5 which is a good indicator because if it was too low then the model would have been suspect of overfitting and if it was too high then the model would have been inaccurate.

5.2 EVALUATION OF MODEL USING GRAPHS

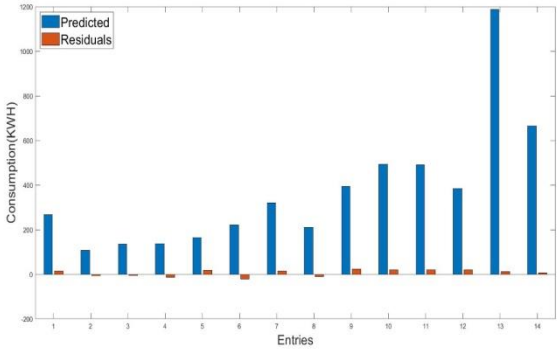
Besides Statistical analysis, we used various graphs and charts to visualize the results and better understand the effectiveness of our model. The first graph is a bar chart that shows the actual energy consumption of institutions in the test dataset. It can be deduced that there are all kinds of institutions in terms of electricity consumption with some schools having very low usage while others have large usages of energy. This shows that our dataset is evenly distributed and therefore is a good indicator of real-world scenarios.



The next graph plots the actual and predicted energy consumption of the test dataset against entries. Executed using the "plot" function, this graph perfectly sums up how accurate the model is. The predicted and actual curves almost overlap and therefore the values predicted by the model were fairly accurate.



Here we have the predicted value and residuals plotted side by side. Residuals are the errors in the predictions that is difference between predicted and actual values. This bar chart was executed using the "bar" function and it shows how insignificant the residuals are thereby reinforcing what the previous plot proved and shows the effectiveness of the model.



6. CONCLUSIONS

Due to the ever-increasing energy usage over the last decade, there has been a concerted effort by all parties involved to improve energy efficiency, decrease usage and accurately forecast future consumption to better prepare the power sector. Therefore there is a need for a simple yet accurate model that can predict future consumption and although there are a lot of models available, very few of them are focused on educational institutes and fewer still are focused on Kerala. This project aims to bridge that gap by creating a tool that is tailor made to the local conditions and can accurately forecast energy consumption among other uses. Data was collected from various types of schools and this data was dividing into two datasets with one of them being used to develop the model. The required parameters were shortlisted using correlation analysis. The parameters with weak correlation analysis (temperature, height of room and working hours) were discarded. The rest of the parameters with greater correlation were taken into account to develop the model. The forecasting model was developed in MATLAB using Ridge Regression. By testing the model, it was observed that the model forecasted energy consumption with good accuracy. Since the data that has been collected to develop the model was based on a wide variety of institutes, it can be used to predict the energy consumption of any educational institution that is going to be constructed in future with reasonable accuracy if the parameters in the model are available. Hopefully, this study has laid the groundwork for further research in this field.

REFERENCES

[1] Moon, J.; Park, J.; Hwang, E.; Jun, S. Forecasting power consumption for higher educational institutions based on machine learning. J. Supercomput. 2018, 74, 3778–3800. [CrossRef]

[2] Yves Kodratoff ; Ryszard Michalski. Machine Learning: An Artificial Intelligence Approach, Volume III,August 1990 [CrossRef]

[3] Helmuth Sp¨ath and Werner Rheinboldt; Mathematical Algorithms for Linear Regression December 1991.[CrossRef]

[4] Nurul Sima Mohammed Shariff; An Application of Proposed Ridge Regression Methods to Real Data Problem, November 2018

[5] https://doi.org/10.1080/02664760802553000

[6] DOI:10.5194/gmdd-7-1525-2014

[7]http://dx.doi.org/10.4028/www.scientific.net/AMM.328.239