Energy Management Research Journal
Vol. 2, No. 1; 2019
Published by CanSRG
http://cansrg.com/journals/emrj/

Energy Consumption and Saving Calculations Using Nearest Neighbor and Artificial Neural Network Models

Kevin Eaton[1,*], Nabil Nassif [2] Pyrian Rai[1] and Alexander Rodrigues[2]

[1]*Department of Computer Science, University of Cincinnati, Cincinnati, Ohio, United States 45221*

[2] *Department of Civil and Architectural Engineering and Construction Management, University of Cincinnati, 2850 Campus Way Drive, 790 Rhodes Hall, Cincinnati, Oh 45221*

Abstract

Demand for building units with advanced saving techniques has grown and continues to grow. Any company or organization has plenty of need for saving energy, whether the motive behind energy saving is saving money or aiding in saving the environment. Buildings now have the technology to track and produce highly accurate electoral outputs. These can be used for improving the functionality of energy models. This paper discusses typical data-based building energy models and proposes new improvements by utilizing an artificial neural network. Using sub-hourly and hourly electric energy consumptions, five different data-based models are utilized and compared to one another. The first two models are linear one regressor models. The first is a linear fit model and the second is a linear change point fit. The third model is a two regressor model using a linear fit. The fourth model is a proposed Classification Learning model using three regressors. The fifth model is a proposed Artificial Neural Network model using three regressors. The two types of data collected are simulation data and actual data. There are four buildings in total; two with simulation data and two with actual data. The results show that the proposed Artificial Neural Network model and K Nearest Neighbor model can provide accurate predictions for the data as compared to traditional linear modeling techniques. These models are then utilized to calculate saving percentage, which is then compared to the actual percentage.

Keywords: *energy saving; energy prediction; artificial neural network; classification learner; linear estimation.*

1. Introduction

With a need for energy savings rising, more and more look to create more efficient building systems has grown with it. This requires the use of modeling techniques to predict energy consumption. After creating an accurate model, the model can be then be used as a prediction. Most modern buildings include electric power meters that can read out the energy data at sub-hourly time intervals. This data is useful, and, in most buildings, it is not used to is full potential [7]. The energy consumption data, used with the temperature and what day of the week, is useful for creating a prediction model for how much power a building consumes. The single or multivariate regression model is widely employed as a means of identifying energy efficiency measures, monitoring energy consumptions, and measurement and verification projects [2, 8]. These models, however, are not as accurate as they could be. More advanced models are available that are much more accurate and useful for energy consumption modeling. The goal is to have intelligent applications in the building system to find efficiency, optimization, energy assessment, and fault detection [2, 7, 8, 10]. This paper investigates data-based building energy models and

* Corresponding author
Submitted: December 13, 2018 Accepted: March 23, 2019

proposes new improvements by utilizing an artificial neural network and a nearest neighbor model. Applying these advanced prediction techniques can help predict in many areas. The KNN model. stands for K Nearest Neighbor which is a classification learning technique [3]. The nearest neighbor determines several samples that can be grouped together in a certain classification [11]. Usually, this technique is used with binary data sets where there are few classifications it can be described to, such as a "+" or "-". For the energy consumption data, the KNN model sorts the data into small pockets of many possible outcomes. This is the reason a "Fine" method is used. The "Fine" version of the KNN only places data into a category if the data is very specific and fits well. Since there are many outcomes to be trained for in energy consumption, the algorithm must be certain that the predicted value is part of the classification [13]. KNN classifiers have been shown to achieve low false positive rates as well as more accurate than other methods on different types of data [4, 12]. Artificial Neural Network model is a model that uses advanced learning techniques to create an output based on inputs given along with validation data. Neural network applications are useful for pattern classification and prediction which are useful in energy model development [12]. The artificial neural network uses artificial neurons to take inputs and use synaptic connections between "cells" to interpret and learn from data [4]. It then creates a map between the inputs and the target values that should be reached [7]. The artificial neural network uses 25 neurons to learn from training and validation data. Previous papers have used an artificial neural network on individual aspects of a building's energy consumption using four neurons [9]. While this is helpful to learn trends of individual factors, it would not predict the total consumption by the building as well as using more neurons. The training data is run into the neurons and the machine learns through the training data. Then, using validation data, the network checks to make sure the validation data fits well [5]. Validation is an important step up that model 5 has over the other 4. Neural networks can check the model and make sure that it is learning in the correct direction instead of teaching itself incorrect patterns [6, 7]. This makes artificial neural networks excellent for training and testing on evolving data, such as building energy consumption [1]. The artificial neural network is a fitting tool. Instead of using regression learning, this artificial neural network attempts to fit the trend in a way that is designed to fit trends that can be represented in graphs [7, 8]. While regression models can be effective, a neural network focused on fitting to a graph is effective. These predictions will aid in being able to calculate savings. The saving can be calculated after simulating optimizations to the air conditioning systems. Saving can then be calculated and estimated to determine which model creates the best estimation for saving calculation.

2. Data Collection

The data is collected from the entire energy consumption of the buildings. Building 1 and 2 are collected from a simulation data using the energy simulation program eQUEST [3]. Building 1 is a simulated building with specific dimensions in New York and building 2 is the same building in Greensboro, North Carolina. Buildings 3 and 4 are both real buildings using their actual data in North Carolina. Both buildings 3 and 4 have different dimensions and attributes. Figure 1 shows the whole building electric consumptions for the four buildings used as functions of dry bulb temperature.

Building 1 and 2, as seen in Figure 1, have similar looking data points. The data differs at the higher dry bulb temperatures due to location. Building 2 is in North Carolina which has higher temperatures. This affects the energy consumption and creates many higher energy output data points around higher temperatures when the air conditioning must be turned on. For the real data of building 3 and 4, They also have much different looking shapes. This is because they are two different buildings with different energy consumption. Building 4 consumes much less energy which is because of the way the building is laid out, the amount of people in the building, etc.

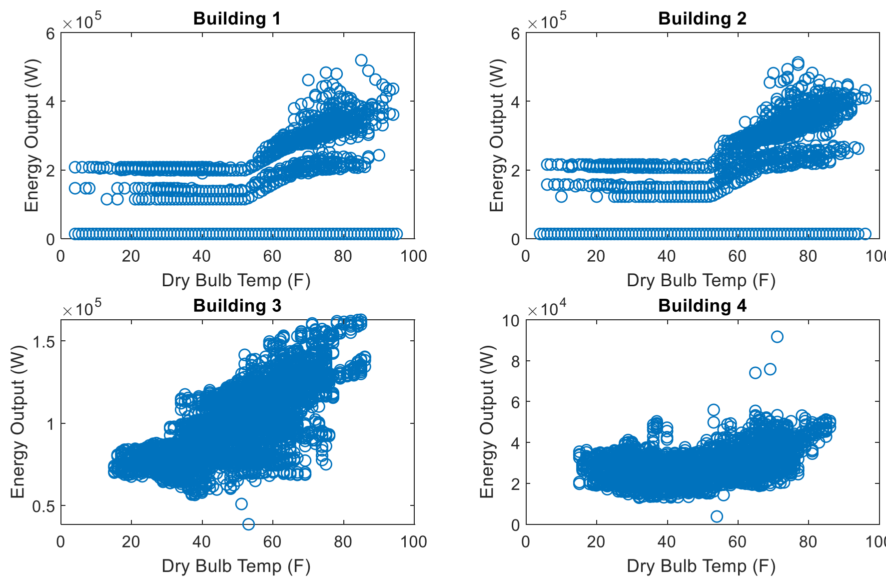

Figure 1. Whole building electric consumptions of buildings being investigated as functions of outside dry air temperature.

3. Building Estimation Models

There are two main ways for energy consumption to be approached in modeling, a forward method and an inverse method. The forward method is to use the detailed parts of the existing system to predict how the system will act in the building. The inverse method is using the actual data of an existing system and making models based on the data acquired by the system [7]. This paper will utilize the inverse method. The inverse method can use multiple regressors or just one. There is no single model that is appropriate for all buildings. One method may be best for a different type of building compared to another. This paper investigates new methods to estimate the energy consumption on a sub-hourly basis for each building as well as the type of day and dry bulb temperature. The five different models listed in Table 1 are discussed. The first two are linear models used by systems today. The third is a two regressor model using temperature and dew point as regressors. The fourth model is a Classification Learner model. The fifth is a proposed Artificial Neural Network model.

The single-variant model (Model 1) is a linear fit model. This model is based on only one regressor, the dry bulb temperature t_a. The a and b values are based on the best fit method within a curve fitting algorithm. The single-variant change-point model (Model 2) uses the dry bulb temperature t_a as the only regressor variable but at 55°F, the model will be more than just the "a" value for the plot. The "+" indicates that the interior of the parenthesis will become zero if outside temperature, t_a, is less than t_1. Model 3 uses two different regressors, dry bulb temperature, t_a, and dew point, t_{dp}. Model 3 then uses a simple one-degree poly fit to estimate what the "a" and "b" values are for the fit equation. These models (Model 1, 2, and 3) are commonly used across commercial and residential areas. These models are easily applied to assist with energy analysis and have data that is easily accessible. Data for these models can be obtained easily from electricity bills or simple output algorithms within the building. These models do not consider advanced factors, such as the day of the week or hours of each day when the air conditioning is on. These factors are considered in Model 4 and 5. Model 4 is a Fine KNN. KNN stands for K Nearest Neighbor which is a classification learning technique. Model 5 is an artificial neural network model. The Model 4 and 5 take multiple inputs, as shown by Figure 2.

Table 1. Models 1 - 4 and detailed description of each model.

Note	Models	Regressor (t_a, t_{dp})	Parameter (a, b, c)	Equations
Model 1	Single Variant	1	2	$Y = a + b.t_a$
Model 2	Single Variant Change-Point t_1	1	2	$Y = a + b.(t_a - t_1)^+$
Model 3	Double-Variant	2	3	$Y = a + bt_a + ct_{dp}$
Model 4	Classification Learner	3	-	-
Model 5	Artificial Neural Network	3	-	-

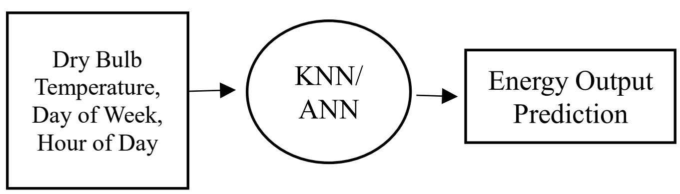

Figure 2. Inputs and outputs that define the Classification Learner and the Artificial Neural Network.

4. Results

The one-year data collected are divided between training data that is for the first nine months of the data and testing data covering three months. All models are developed by taking the training data and placing the model against the testing data. In Model 5, part of the data is also taken as validation data in order to properly train the artificial neural network. An example of the training and testing is shown below in Figure 3.

To discuss these results, the performance of each model will be analyzed for all buildings. As shown in figure 9, the linear models (Models 1, Model 2, and Model 3) have a lower training and testing R-Squared value as compared to the Classification Learner and Artificial Neural Network. In figure 9, model 1 shows that the test values fall below 0.2 for all buildings. This demonstrates how the single linear model does not fit well onto the testing data. This can also be seen in the figure 4 with the poor fit onto both buildings. The data is spread widely for all the dry bulb points and having a single linear line to fit to the data does not predict the data well, especially in the simulation data. The two times where the R-Squared was high for model 1, buildings 3 and 4, the data collected had much noise on it due to the acquired data being from a real building with noisy data points. This accounts for a raise in r-squared value in figure 9.

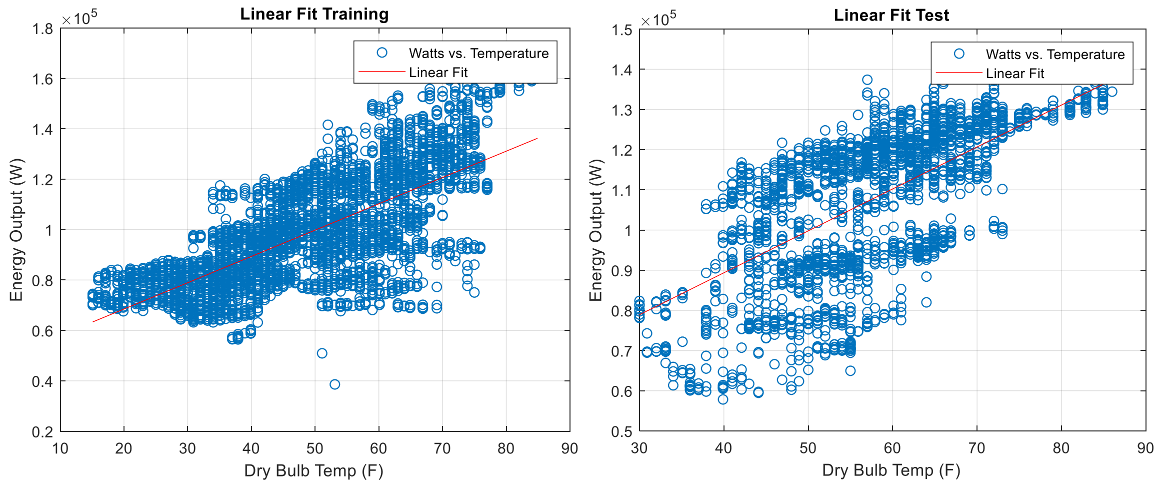

Figure 3. Training and Testing data with a linear fit line for Building 3.

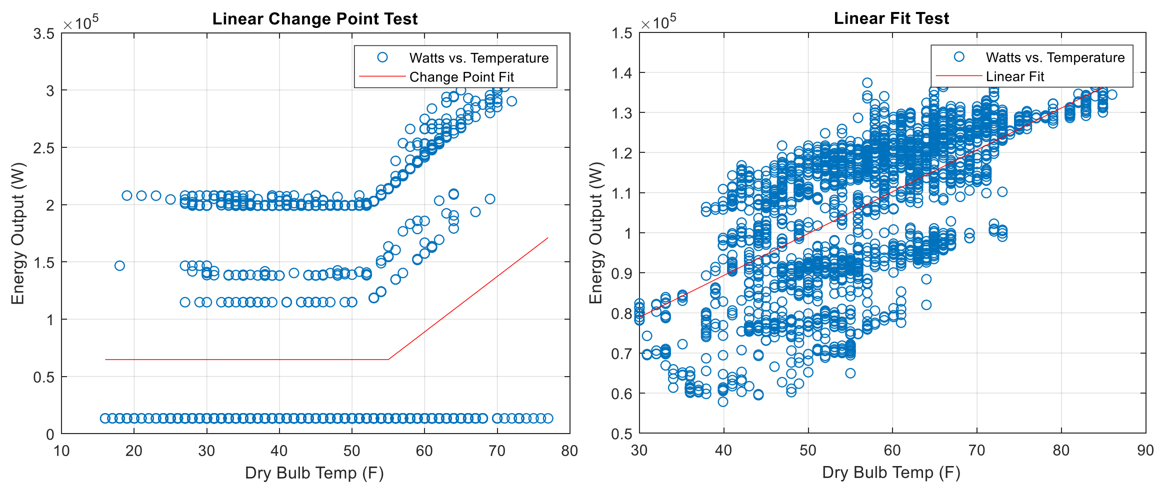

Figure 4. Results of Model 1 on Building 1 and Building 3.

Model 2 creates an overall better fit than Model 1. In every building, the testing results of model 2 increase from model 1's r-squared result. This improvement comes from the change point considering how the air conditioning system works. The air conditioning systems come on at around 55 degrees Fahrenheit which increases energy consumption. Figure 5 shows how the testing fit on the real and simulated buildings appears accurate than in figure 9 but still not as accurate as possible. In the simulated data, the change point fit only is accurate for a small fraction of the data.

Model 3 creates a better fit than Model 1 and 2 in all buildings except Building 3. This may be due to how R-Squared is calculated. Since it falls within the middle of the data and R-Squared calculations take the difference from each data point, the R-Squared value may be giving a misleading number when calculated. When looking at figure 6, it appears that the second regressor adds in important data gaps that models 1 and 2 leave out. The second graph in figure 6 appears linear slightly. It is not linear though because it does slightly spread in the middle. The data does not spread as much in the real data because the data points are closer in relation to each other without much jumping as in the simulation data.

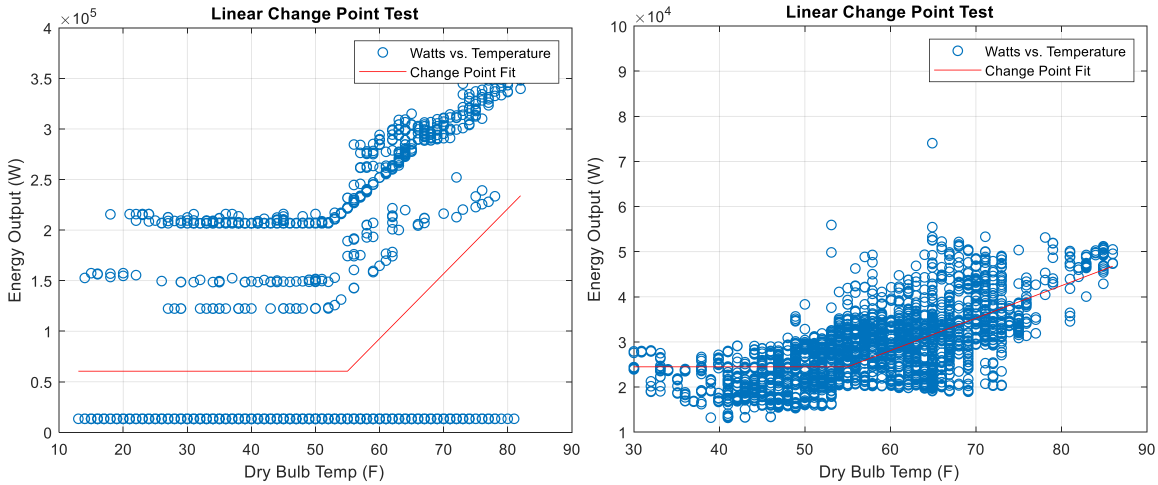

Figure 5. Results of Model 2 on Building 2 and Building 4.

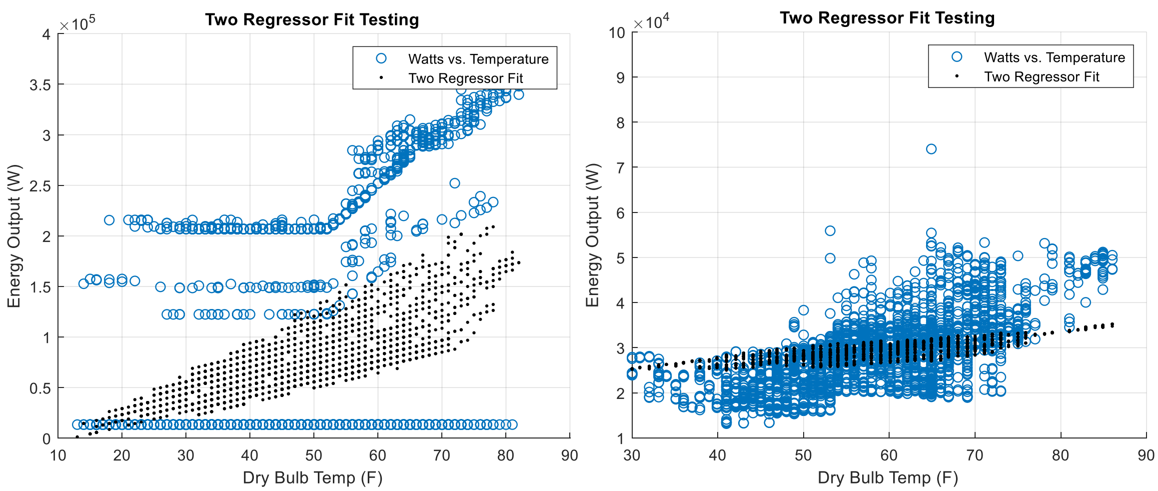

Figure 6. Results of Model 3 on Building 2 and 4.

Model 4 does much better when compared to the other three models. By training the Classification Learner with more than one regressor, the R-Squared values were greater than the linear model. Figure 6 shows how the testing fit onto the data matches the fit and more accurately represents the data. The data points predicted to fit onto the data set are visibly closer than a single line of fit. The data also backs this up by showing that all R-Squared values are greater in the Classification Model as compared to the single and double linear fit models, as seen in figure 9.

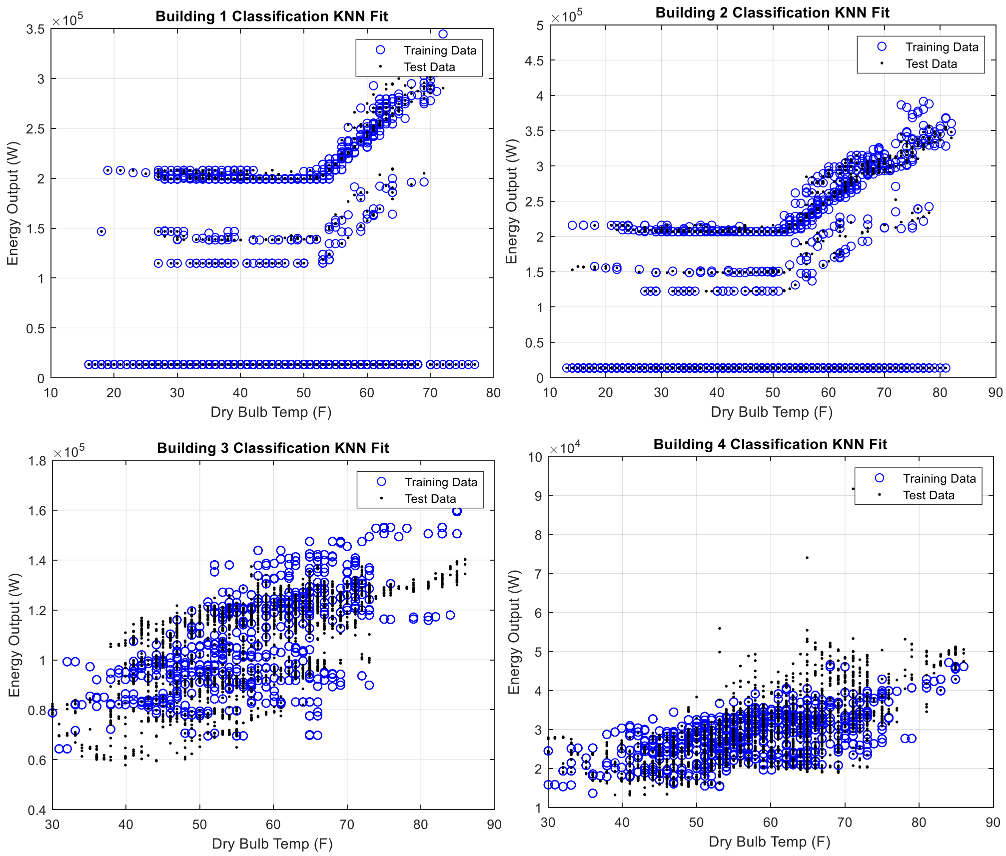

Figure 7. All proposed model results on Buildings 1-4.

Figure 8 shows the plots for Model 5. For the simulation data, the neural net does not fit the data as well as the KNN does. The data also backs this up in figure 9. This is due to how the classification learner operates. Since it is much better at classifying the data it can tell which level of energy consumption the data will be at and create a model to fit it. In the actual data, the artificial neural network demonstrates its advantage over the KNN model. The artificial neural network is a much better predictor of the trends of the actual data. This is due to its advanced modeling techniques and precision due to its ability to learn with the data. This can also be supported by the data in figure 9.

Figure 10 shows how the models fit as compared to the actual data for one week of a simulation building. The proposed model, 5, fits very closely to the actual data unlike model 1, 2, and 3. Model 4 does seem to do a slightly better job than model 5 in the simulation data for the week. This is due to the classification learner working better in simulation data due to its classification techniques, as mentioned earlier. This figure demonstrates most how much better model 4 and 5 do at identifying the correct data points as compared to models 1, 2, and 3.

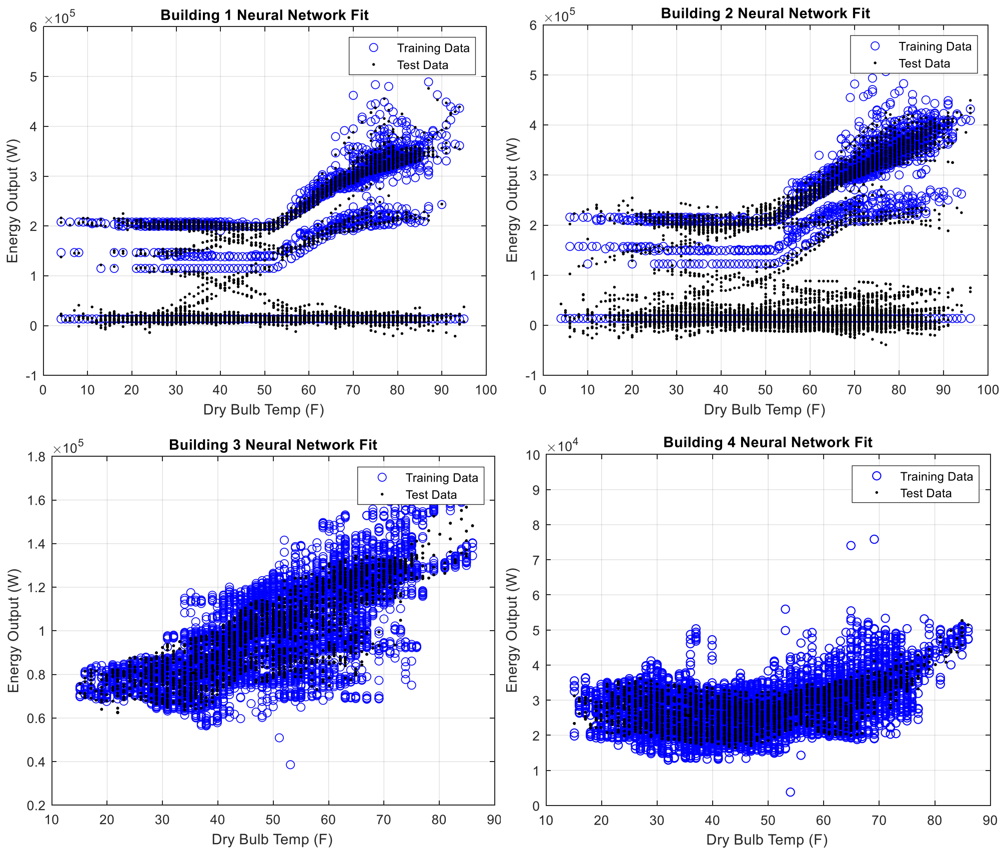

Figure 8. Artificial Neural Network Fit on Buildings 1-4.

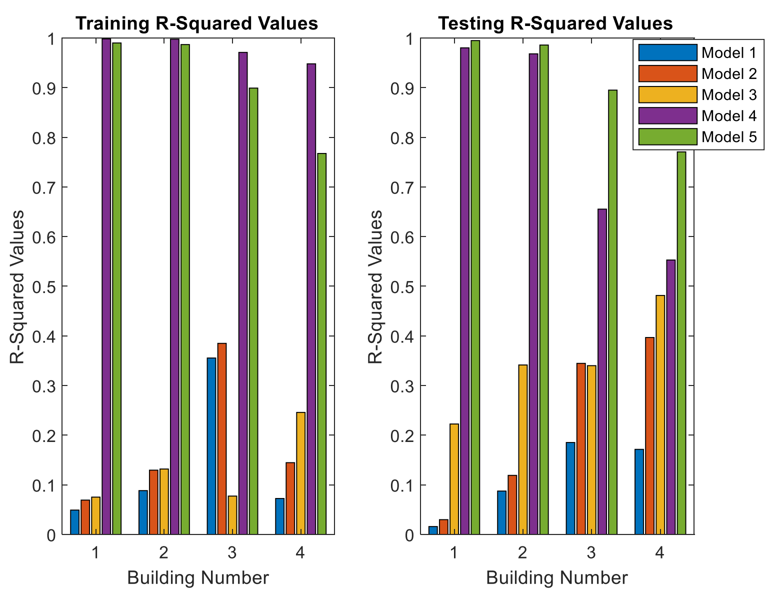

Figure 9. R-Squared Values for each building and model.

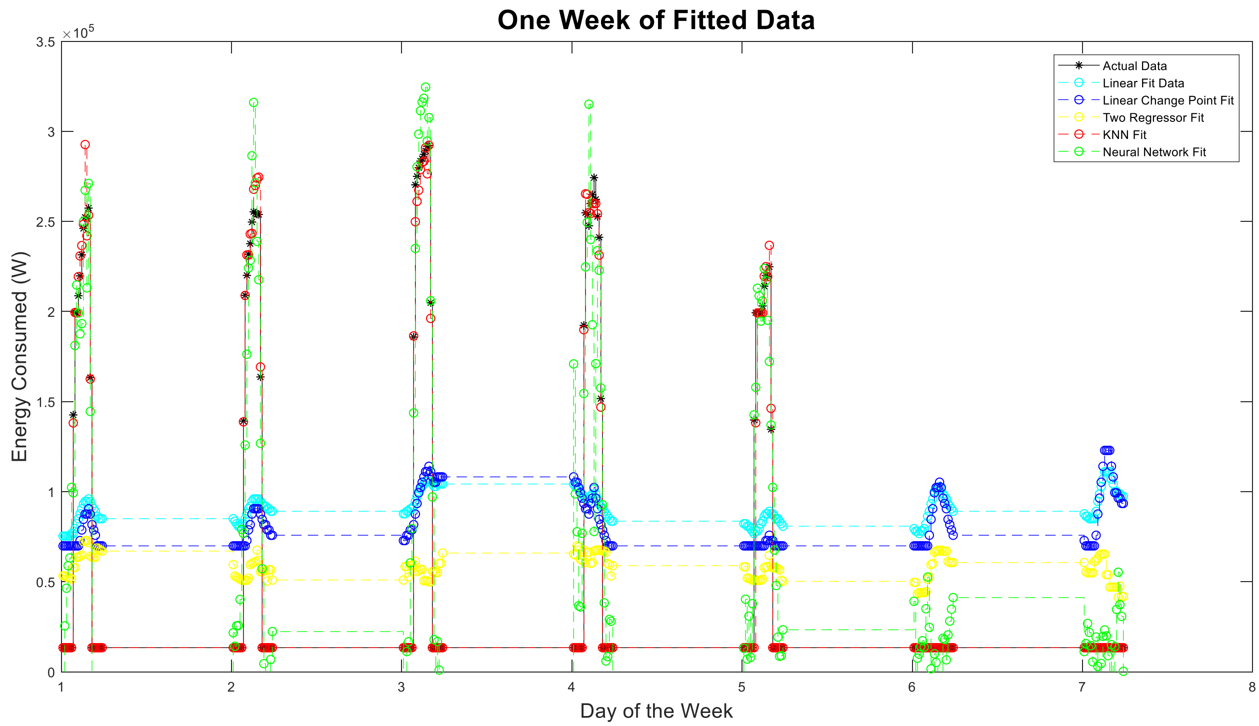

Figure 10. One week of testing data for Building 1 (not negative).

5. Saving Calculation

Annual saving calculations were done on the simulation buildings using simulated optimization data. The eQuest simulation program was used to simulate optimized energy consumption. The optimized factors were a light change from 1.2 to 0.8 watts per square foot, a Fan Premium, and Optimal Supply Air Temperature. The estimated data was created by training the all the models from Table 1 with the non-optimized data using nine months of training and three months of testing. As seen in Table 2, the saving percentage is all positive. That is because the actual data used more energy than the optimized data, so the saving percent is positive. If the estimated data in Table 2 is within two percent of the actual data, showing that using the classification learner and the artificial neural network are good fits of estimating actual savings for the simulation data. Model 1 outperforms Model 4 in Building 2, but this may come because of circumstance of the data due to Model 1 being much less accurate for Building 1. Model 5 calculates the saving data the most accurately. It falls less than one percent from both buildings, better than every other model. For saving calculation of simulation data, model 5 preforms the best.

Table 2. Estimated Saving Percent for each model.

	Actual	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Building #1	27.82%	23.54%	33.43%	39.98%	26.93%	28.06%
Building #2	35.51%	36.81%	44.21%	52.36%	33.30%	35.22%

6. Conclusions

Five different data-based models of estimating energy consumptions were tested on four buildings. The models are two single regressor models, a two regressor model, a classification learning model, and an artificial neural network. The data was collected from simulation data and from real building data. The models were evaluated using this data. The testing results indicated that the regression model with the proposed classification learning could improve the models' accuracy. In Building 1, Model 1 had an R-Squared value of 0.06, Model 2 had a value of 0.09, Model 3 had a value of 0.33. These values are low and do not fit the data well, unlike the classification learner and artificial neural network which had an R-Squared testing value of 0.97 and 0.99. In Building 4, Model 1,2, and 3 had R-Squared values of 0.17, 0.35, and 0.46. Model 4 was only slightly greater at 0.55. Model 5 was much better of a fit at 0.77. This result comes from issues with models 1-4 with how noisy the data is in the actual buildings and affects the r-squared value. Overall, the proposed models performed better than any of the other three models. Looking at figure 9, model 4 does a better job predicting the trends of simulation and clean data while model 5 excels with real data. Due to its classification techniques, model 4's algorithm understands how the simulation data works better than the artificial neural network. This can be seen in figure 9 as well as figure 7 and figure 8. Model 5 demonstrates its benefit and usefulness in the real data. Model 5 is more consistent with its results because it is always above 0.75 r-squared value, which is substantially better than any of the other models tested. The proposed models are also effective because of the short training period they take, even though they are more complex than models 1 – 3. If used for saving calculation, both can be effective for all businesses to implement this prediction model and predict and optimize their building for calculations. Overall, using the Artificial Neural Network and K Nearest Neighbor compared to the linear models are more accurate and reliable at predicting the behavior of the data in real and simulated data. This can be applied to real buildings in order to optimize and save energy within buildings. Table 2 demonstrates that predicting the saving with the artificial neural network is the most valuable, within 1%, of the actual predicted savings when used with optimization data within Buildings 1 and 2. Model 4 is with in 2% of the actual predicted amount which is substantial, but not as effective as the artificial neural network.

References

[1] A. H. Neto and F. A. S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy and Buildings, vol 40(12), pp. 2169-2176, 2008.

[2] ASHRAE Handbook-Applications, ASHRAE. Chapter 41. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.

[3] eQuest. QUick Energy Simulation Tool, eQUEST Version 3.65.

[4] K. Fukushima, Neocognitron: A Self-Organizing Artificial Neural Network Model for a Mechanism of Visual Pattern Recognition, Biological Cybernetics, vol 45(4), pp 193-202, 1980.

[5] L. K. Hansen and P. Salamon, Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12(10), 1990.

[6] MATLAB and Artificial Neural Network Toolbox Release 2018a, The MathWorks, Inc., Natick, Massachusetts, United States.

[7] N. Nassif, Regression and Artificial Neural Network Models with Data Classifications for Building Energy Predictions, ASHRAE Transactions, vol 124(2), pp. 993-1001, 2018.

[8] J. E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energy and Buildings, vol 39(1), pp 52-58, 2007.

[9] S. L. Wong, K. W. Wan and N. T. Lam, Artificial neural networks for energy analysis of office buildings with daylighting, Applied Energy, vol 87(2), pp. 551-557, 2010.

[10] S. W. Wang, Q. Zhou and F. Xiao, A System-level Fault Detection and Diagnosis Strategy for HVAC Systems Involving Sensor Faults, Energy and Buildings, vol 42(4), pp. 477-490, 2010.

[11] W. Cedeño and D. K. Agrafiotis, Using particle swarms for the development of QSAR models based on K-nearest neighbor and kernel regression, Journal of Computer-Aided Molecular Design, vol 17(2-4), pp. 255-263, 2003.

[12] Y. Liao and V. R. Vemuri, Use of K-Nearest Neighbor classifier for intrusion detection, Computers & Security, vol 21(5), pp. 439-448, 2002.

[13] X. Yong, Q. Zhu, Z. Fan, M. Qiu, Y. Chen, and H. Liu, Coarse to fine K nearest neighbor classifier. Pattern Recognition Letters, vol 34(9), pp. 980-986, 2013.