Linear and Classification Learner Models for Building Energy Predictions and Predicting Saving Estimations

Kevin Eaton, Nabil Nassif, Pyrian Rai and Alexander Rodrigues
University of Cincinnati
Cincinnati, United States
{eatonko, nassifnl, raipy, rodrigax}@mail.uc.edu

ABSTRACT
The need for creating building systems with smart systems is growing. Saving energy in buildings is both important in aiding the environment and saving money for the companies and organizations who run those buildings. Most buildings are now equipped with technology to produce accurate electrical outputs that can be used for improving the accuracy of energy models. This paper discusses typical data-based building energy models and proposes new improvements by utilizing a classification learner. Estimating sub-hourly and hourly electric energy consumptions are discussed using four different data-based models. The first model is a linear fit model using one regressor, the second is a linear change point fit using one regressor, and the third model is a two regressor model using a linear fit. The fourth model is a proposed Classification Learning model using three regressors. Two different types of data were collected: simulation and actual data. There are four buildings total: two with simulation and two with actual data. The results show that the proposed Classification Fine KNN 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.

1 INTRODUCTION
As energy concerns continue to grow, the need for creating more efficient building systems has increased. This requires us to use modeling techniques to save energy more accurately and efficiently. Most modern buildings include electric power meters that can read out the energy data at specific time intervals. This data is extremely useful, and, in most buildings, it is underutilized. This data paired with the temperature and what day of the week is useful for predicting how much energy a building will consume. There is a need to investigate how modern computing techniques can benefit these buildings. 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 [1][3]. These models, however, may not be as accurate as using advanced techniques. The hope is to have intelligent applications in the building system to find efficiency, optimization, energy assessment, and fault detection [1][3][4][5]. Appling these advanced prediction techniques can aid in many areas. The model being proposed in this paper is the Fine KNN model. A KNN model is an algorithm that interprets the data based on the inputs given and what the output should be and creates a fit based on the data. A Nearest Neighbor model was chosen to be a proposed model because it evolves based on the data and surrounding points. This makes it excellent for training and testing on evolving data, such as building energy consumption. 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 [2]. 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 have different dimensions and attributes. Figure 1 shows the whole building electric consumptions for the four buildings used as functions of dry bulb temperature.

SimAUD 2019 April 07-09 Atlanta, Georgia

Building 1 and 2, as seen in Figure 1, have similar looking data. The data differs at the higher dry bulb temperatures due to location. Building 2 is located in North Carolina which has a higher temperature more often. This affects the energy consumption and creates many higher energy output data points around higher temperatures when the air conditioning has to 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 needs. Building 4 consumes much less energy which may be because of the way the building is laid out or the amount of people in the building, etc.

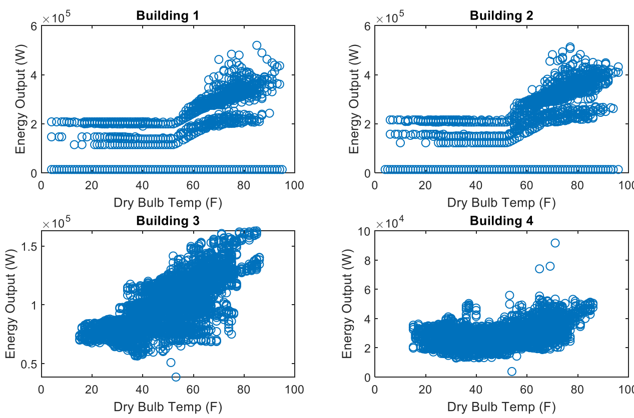

Figure 1. Whole building electric consumptions of buildings being investigated as functions of outside dry air temperature.

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 [3]. The inverse method is using the actual data of an existing system and making models based on the data acquired by the system. This paper will investigate the inverse method. The inverse method can use multiple regressors or just one. There is no single model that is appropriate for all buildings and system. This paper investigates new techniques to estimate the energy consumption on a sub-hourly or hourly basis for each building as well as the type of day and dry bulb temperature. The four 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. The fourth model is a proposed Classification Learner 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	$Y = a + bt_a + ct_d$
Model 4	Classification Learner	3	-	-

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

The Single-Variant Model (Model 1) is a simple linear 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. The single-variant change-point model (Model 2) uses the dry bulb temperature t_a as the only regressor variable but only at 55°F will the model 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 energy consumption would be. These models (Model 1, 2, and 3) are used widely across commercial and residential areas. These models are easily applied and have data that is easily accessible. Data for these models can be obtained easily from electricity bills and simple output algorithms within the building. These models do not consider the day of the week or hours of each day when the air conditioning is on or not. These factors are considered in Model 4. Model 4 is a Fine KNN. KNN stands for K Nearest Neighbor which is a classification learning technique. The nearest neighbor determines several samples that can be classified as a certain group (Yong, 2013). 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 predicts 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 close and fits well. Since there are many outcomes to be trained for, the algorithm must be certain that the predicted value is correct. "Fine" describes how far away the neighbors are that the data is trained off. The "Fine" version of the KNN uses close data around it to train and trusts them to be accurate. The fine setting was chosen because the data is reliable enough that using nearest neighbors with a fine detail and create an accurate model. The learner takes multiple inputs, as shown by Figure 2.

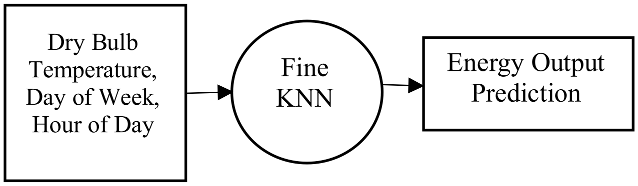

Figure 2. Inputs and outputs that define the Classification Learner

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. An example of the training and testing is shown below in Figure 3.

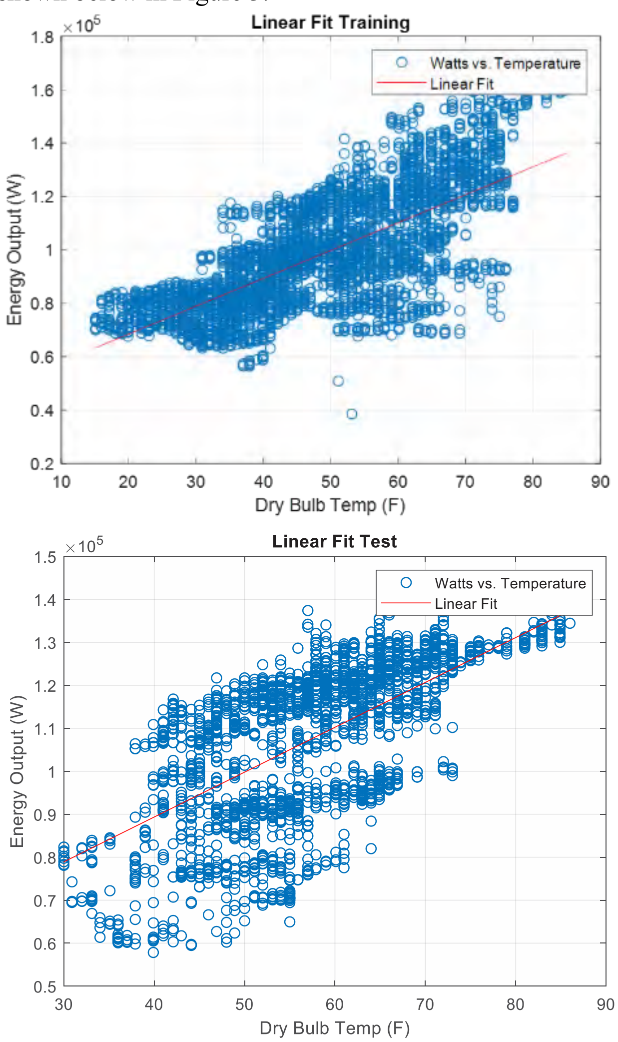

Figure 3. Training and Testing data with a linear fit line for Building 3.

To discuss these results, the performance of each model will be analyzed for all buildings. In a general sense, 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.

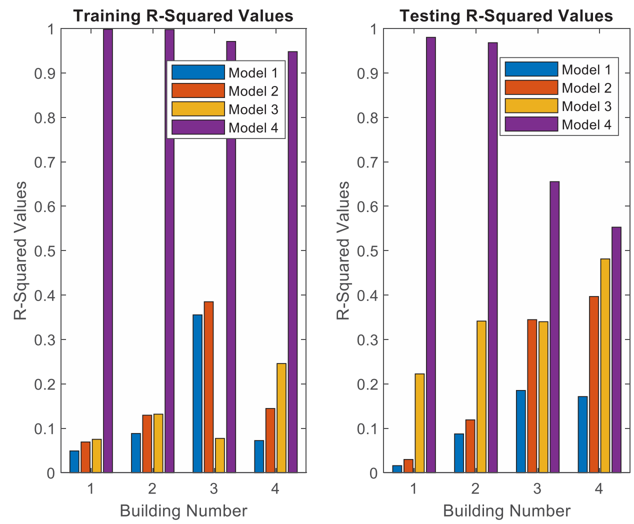

Figure 4. R-Squared Values for each building and model.

Looking at Model 1, we see that the test values fall below 0.2 for all buildings. This demonstrates how the single linear model does not fit the data very well. This can also be seen in the Figure 5 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. 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 fluctuations in the amount of total energy consumed.

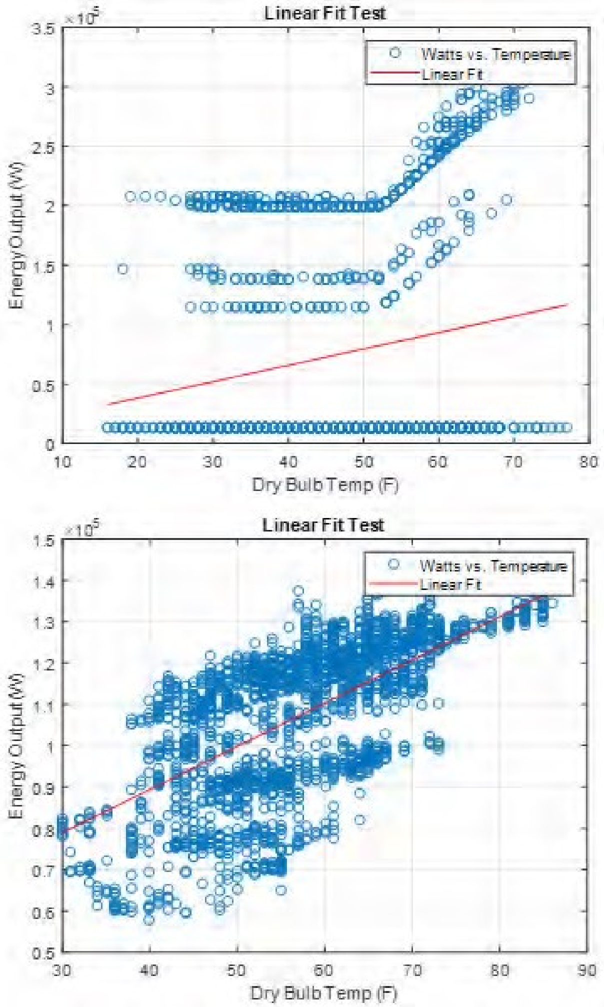

Figure 5. Results of Model 1 on Building 1 and Building 3

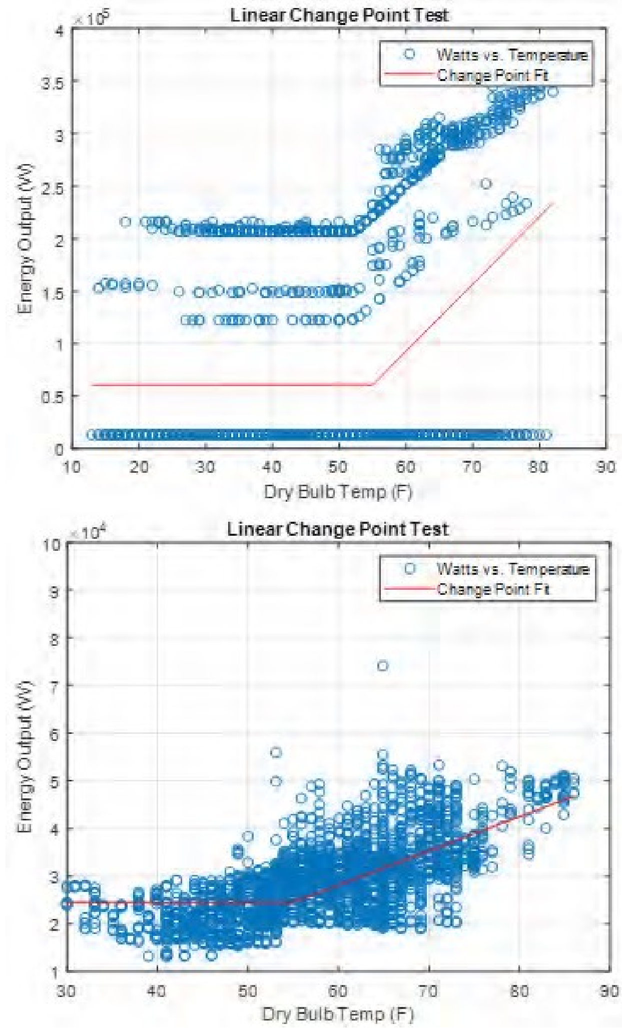

Figure 6. Results of Model 2 on Building 2 and Building 4

Model 2 creates a better fit than Model 1. In every building the testing results of Model 2 is an increase on Model 1 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 6 shows how the testing fit on the real and simulated buildings is more accurate than in Figure 5 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. Figure 4 illustrates this point as well because the R-Squared for the testing data for Model 1 is less than Model 2 for both Building 2 and 4.

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 7, it appears that the second graph is only linear. 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.

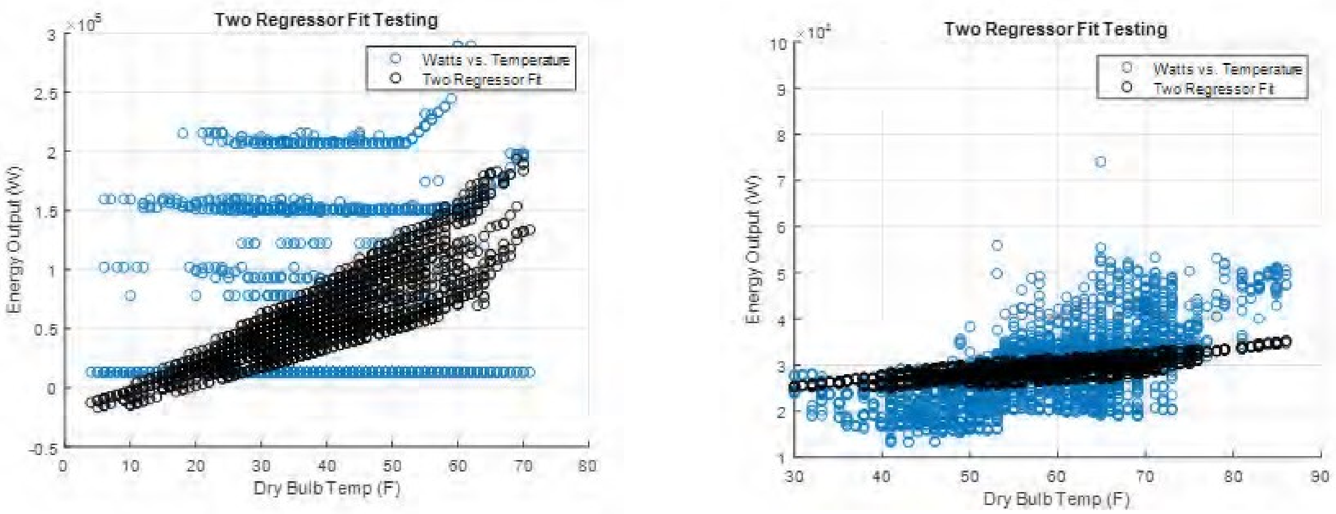

Figure 7. Building 2 and 4 with Model 3 fit onto the data.

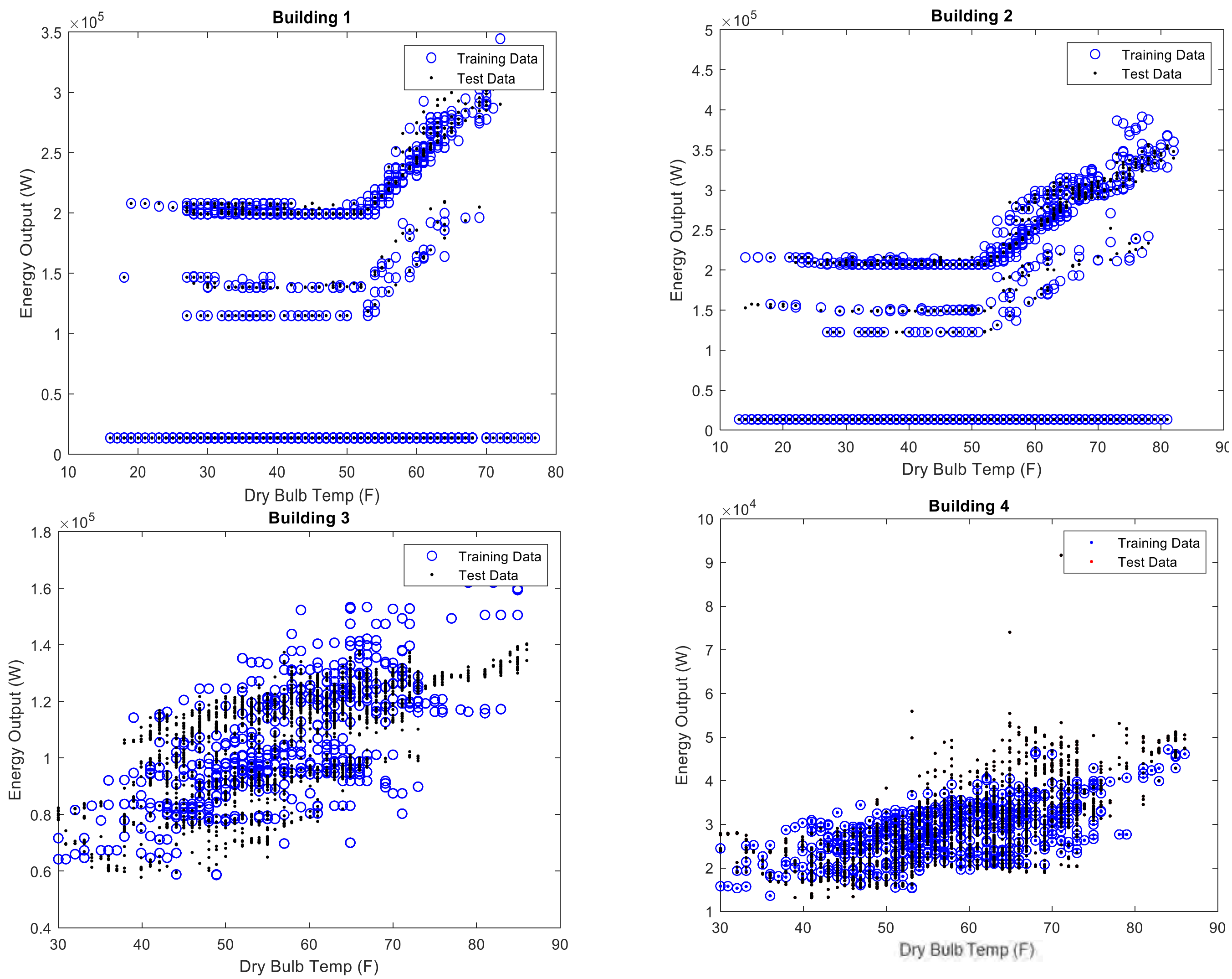

Figure 8. All proposed model results on Buildings 1-4.

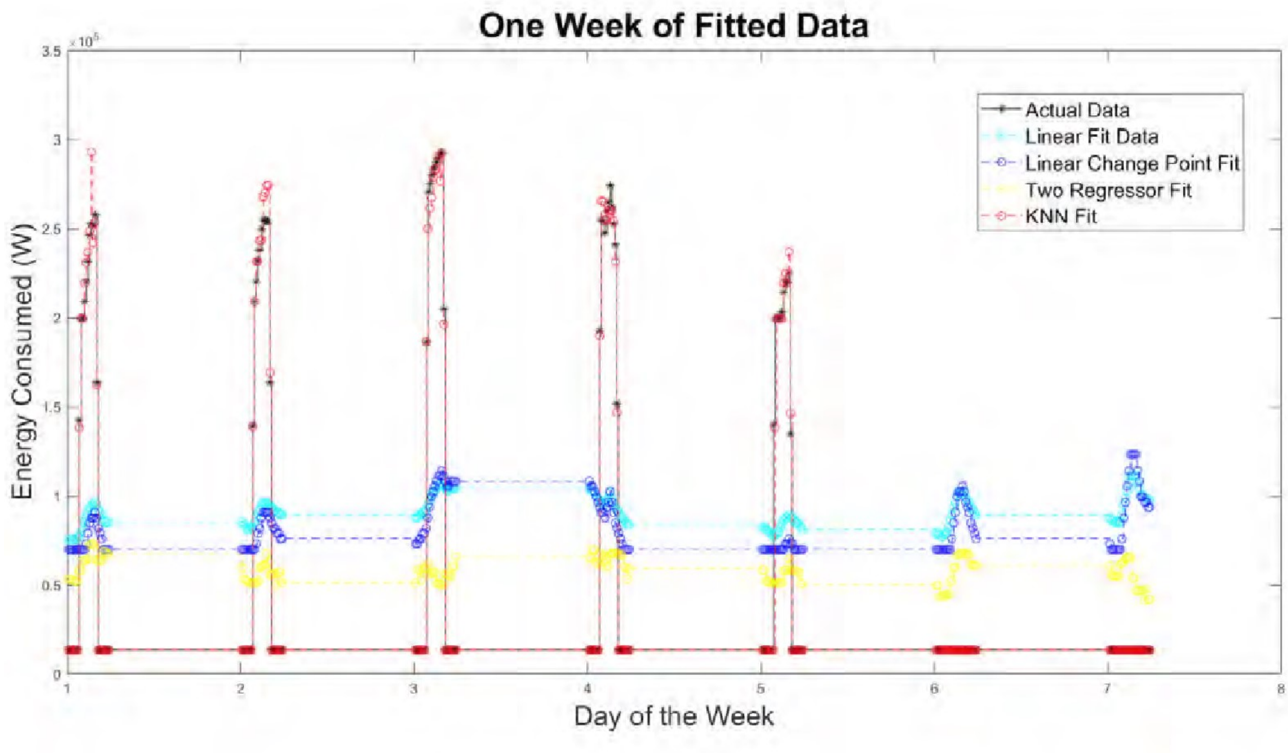

Figure 9. One week of testing data for Building 1

Overall, Model 4 does the best when compared to the other three models. By training the Classification Learner with more than one regressor as well as using neural network to train create the model, the R-Squared values were greater than the linear model. Figure 7 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 linear fit models, as seen in Figure 4.

Figure 9 shows how the models fit as compared to the actual data per one week of data. On the x-axis, one being Monday and seven being Sunday. The proposed model, model 4, fits very closely to the actual data unlike model 1, 2, and 3. This demonstrates how unwell these models fit the actual data as compared to Model 4. This can also be seen in Figure 4 with Model 4 being much closer of a fit with the R-Squared values

5 SAVING CALCULATIONS

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

Table 2. Estimated Saving Percent for each model.

Annual saving calculations were done on the simulation buildings. 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 optimized data using nine months of training and three months of testing. As seen in Table 3, 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. The estimated data in Table 3 is within two percent of the actual data, showing that using the classification learner is a good fit 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 4 also is the most consistent with its errors as compared to the other three models because it is close by similar percentages, a little less than two percent, unlike the variations of the other two models.

6 CONCLUSION

Four different data-based models or estimating energy consumptions were tested on four buildings. The models are two single regressor models, a two regressor model, and one classification learning model. The data was collected from a simulation and from real building data. The models were evaluated using these data sets. 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 which had an R-Squared testing value of 0.97. In Building 4, this pattern is seen again with Model 1,2, and 3 having R-Squared values of 0.17, 0.35, and 0.46. Model 4 was only slightly greater at 0.55. This result comes from issues with how noisy the data is in the actual buildings and affects the r-squared value. Overall, the proposed model performed better than any of the other three models. While in building 4 the r-squared value of model 3 gets close to model 4, model 4 shows its benefit by being more consistent with its results always above 0.5 r-squared value. By using a Nearest Neighbor model to predict how the energy consumption of the building, it is effective in predicting models. The Fine KNN model is also effective because of the short training period it takes. If used for saving calculation, it can be effective for all businesses to implement this prediction model and predict and optimize their building for saving 0.39 for test values, respectively. Model 4 has an R-Squared value of 0.53, higher than the other models. Overall, using the classification learner compared to the linear models is 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 3 demonstrates that predicting the saving with the Classification Learner is close, within 2%, of the actual predicted savings when used with optimization data within Buildings 1 and 2.

7 ACKNOWLEDGMENTS

The authors would like to thank all the volunteers, and all publications support and staff, who wrote and provided helpful comments on previous versions of this document.

8 REFERENCES

1. ASHRAE. 2015. ASHRAE Handbook-Applications. Chapter 41. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.
2. eQuest. **QU**ick **E**nergy **S**imulation **T**ool, eQUEST Version 3.65. http://www.doe2.com/equest/
3. Nassif, N. (2018). Regression and Artificial Neural Network Models with Data Classifications for Building Energy Predictions, ASHRAE Transactions 124 (2).
4. Seem, J.E. (2007). Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy and Buildings 39: 52–58.*
5. Wang, S.W., Q. Zhou, and F. Xiao. 2010. A System-level Fault Detection and Diagnosis Strategy for HVAC Systems Involving Sensor Faults. *Energy and Buildings, 42: 477-490*.
6. Yong Xu, Qi Zhu, Zizhu Fan, Minna Qiu, Yan Chen, and Hong Liu. 2013. Coarse to fine K nearest neighbor classifier. *Pattern Recognition Letters, 34: 980-986.*