International Journal of Engineering & Technology, 7 (4.6) (2018) -

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Power Consumption Prediction based on Infrastructure for Technical Educational Institutions

Dr. Poonkavithai Kalamegam[1]*, Dr. Dhaya C[2]

[1]*Hindustan Institute of Technology and Sciences*
[2]*Adhiparasakthi Engineering College*
**Corresponding author E-mail: a kpoonkavi@hindustanuniv.ac.in*

Abstract

Electricity plays a vital role in our daily routine activities. Particularly, educational institution power requirements may vary every day and the infrastructure of the building determines the amount of power required for a day. Hence advance prediction of power for every building in a university is very important for effective management of electricity during power crisis. In this paper, prediction of power consumption of each building in a university is carried out based on the infrastructure of a building. Feature engineering is carried out using ClassifierAttributeEval. This helps in reducing the size of the dataset as well as the computation time of prediction models. To forecast the energy production and usage in advance, Support Vector Machine (SVM), Neural Network (NN), Random Forest (RF) and Stochastic Gradient Descent (SGD) models are applied. From the experimental results and 10-fold cross validation sampling type , it is proved that SGD has better prediction accuracy when compared to SVM, NN and RF with Mean Square Error (MSE) of 1.102 , Root Mean Square Error (RMSE) of 1.050, Mean Absolute Error (MAE) of 0.613 and Coefficient of determination (R2) of 1.0000.

Keywords: *ClassifierAttributeEval feature selection; Support Vector Machine; Neural Network; Random Forest, Stochastic Gradient Descent;10-fold cross validation sampling*

1. Introduction

Electricity consumption and production takes place at the same time. Hence the knowledge of predicting how much power will be required in advance is very much essential in order to meet the electricity load demand according to our daily activities. The demand of electricity varies from building to building [1]. The pattern of power consumption in high residential areas may depend upon the region specific data as well as the temperature of the particular region. This can be achieved by several statistical and engineering approaches to analyze the electrical energy consumption in residential areas [2]. A university campus will be having several buildings. Each building will be requiring different load patterns based upon several factors like the semester, climate, holidays, working days etc. Many works has been done to determine the electricity usage of buildings in university which was based on external factors as mentioned above.
The following section gives the clear picture behind the experiments: Section 2 discusses about various machine learning methodologies used in predicting power consumption for various buildings with various factors. Section 3 describes about the dataset and the feature selection method used. Section 4 discusses about model selection, Section 5 portrays about performance evaluation and at last Section 6 concludes the paper.

2. Related Works

Jihoon et al [1] have predicted the power consumption across educational institutions based on time, weather, class schedule and work hour using ANN and SVR. April Morton et al [2] have proposed a hybrid approach for predicting the household level energy consumption in a high resolution pattern. Cyril et al [3] has found various solar irradiation forecasting methods as many countries are using renewable energy resources (wind power, solar power, photovoltaic etc) in their power systems. He has analyzed about various solar irradiation forecasting methods like Artificial Neural Networks (ANN), Autoregressive Integrated Moving Average (ARIMA), Support Vector Machine (SVM), Ensemble methods and regression trees.
Thanchanok et al [4] introduced shape based Dynamic Time Warping (DTW) method, to forecast the amount of energy usage across households in 24 hours. Based on DTW clusters, markov model is used to predict the usage of power one day before.. Mohammad et al [5] has found hybridization of Artificial Intelligence (AI) methods with Swarm Intelligence (SI) to improve the accuracy of prediction based on non-linear features like weather, hours, and day predictions. Yildiz et al [6] proposed single regression models to estimate the amount of power of consumption across commercial buildings as theatres, malls, hospitals, institutions etc with huge HVAC (Heating, Ventilating Air-Conditioning) systems. Luis et al [7] has presented a data driven approach with varying datasets and inside building and outside criteria for predicting the energy consumption across domestic appliances. Gradient Boosting Machines (GBM) has showed better results and accuracy rather than other regression models.
Zeyu et al [8] have proposed an Artificial Intelligence based approach, which uses statistical data for analysis, to building level energy usage. The buildings might include apartments, residents, small scale offices etc. Greig et al [9] has proposed an Artificial Neural Networks (ANN) based approach in predicting the energy consumption of a building during the design stage itself instead of analyzing the energy usage of building in operation since its evaluation cannot be correctly predicted. Pan et al [10] has proposed

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sparse autoregressive model to measure the power consumption of individual user. This method is based on LASSO (Least Absolute Shrinkage Selection) algorithm, in which selects the pattern of user consumption.

3. Dataset Descriptions and Feature Selection

3.1. Dataset Information

The dataset consists of university building power consumption information. In this research, five department buildings namely, Computer Science and Engineering (CSE), Electrical and Communication Engineering (ECE), Electrical and Electronics Engineering (EEE), Civil and Mechanical department were taken into account. The dataset consists of what is the date and time, semester, whether it is working or non-working day for each department, at which time every department is conducting special class/lab, evening class/lab, how many number of classes and labs are there for each department, how many fans, tube lights are used in each department's class, how many AC, Desktops are there in every lab of each department , how many machineries are there in each department, for how many hours all these appliances are used, temperature of the day and finally the total power consumption per day for every department. In this dataset, the following wattage levels for every appliance were considered [23].

Table-1: Dataset description

Dataset characteristics	**Multivariate**	**Number of instances**	**5605**
Attribute characteristics	Categorical, Integer, Real	**Number of attributes**	33
Associated tasks	Classification	**Missing values**	No

For effective prediction and computation, feature extraction method performs the transformation of data from high-dimensional space to low-dimensional space. It gives ranking for all attributes based on the correlation with the target attribute. The low-ranked feature, that is attribute having less significance on the expected value, can be skipped from the dataset, thereby decreasing the size of the attribute, which results in the less computation time and good performance measure.

Table-2: Wattage levels of appliances

APPLIANCE	WATTAGE LEVEL
Fan	40
Tube light	25
Desktop	100
AC	3000
Machinery	2500-4000

3.2. Classifier Attribute Eval

ClassifierAttributeEval is feature selection method applied to the above dataset, which evaluates the worth of each and every attribute by using the training dataset. This feature selection method supports, unary, Binary, Empty, Missing, Numeric, Date and Nominal attributes. When the ClassifierAttributeEval is applied to the dataset, Date, no.of.hours.l, AC in use (3000), Light in use (25watts), Fans in use(40watts), no.of.hours.f, Temperature in Celsius, no.of.hours.a.c, class count, lab count, AC count, light count and fan count had least ranking, when compared to other attributes as the total power consumption had no correlation with these least ranked attributes. Hence these attributes can be removed from the dataset which gives faster computation time and more accuracy.

4. Model Selection

In this paper, four models namely SVM, ANN, Random Forest, Stochastic Gradient Descent, has been applied to predict the power consumption among educational institutions. By applying the dataset to the ClassifierAttributeEval feature selection method, low ranked attributes were removed from the dataset, thereby reducing the size of the attribute as well preserving that the removed attributes has no consequence on the target value. This reduced dataset now takes the role of training dataset, in which total power consumption feature is selected as target variable. This training dataset is then applied to SVM, ANN, Random Forest and Stochastic Gradient Descent machine models, in order to analyze which model predicts more accurately against the target attribute. From the prediction, it is found that Stochastic Gradient Descent machine learning model prediction had better accuracy when compared to that of SVM, ANN and Random Forest. The prediction results are shown in the below report:

Data & Predictions

	Random Forest	SVM	Neural Network	SGD	total energy consumption in kwph per day
1	154.680	113.931	139.315	163.235	163.640
2	0.000	101.195	0.348	0.221	0.000
3	0.000	101.530	-0.783	0.182	0.000
4	188.368	115.594	154.259	188.263	188.320
5	187.810	115.248	161.477	187.363	187.920
6	154.680	114.057	139.693	163.045	163.960
7	189.044	115.552	159.278	188.889	188.920
8	145.135	112.935	119.888	138.964	139.920
9	0.000	101.328	0.324	0.256	0.000
10	0.000	101.469	-0.794	0.146	0.000
11	187.770	115.314	157.604	187.634	187.720
12	187.717	115.160	162.700	187.153	187.720
13	188.017	115.093	160.838	187.230	187.720
14	145.135	112.261	135.438	137.840	138.960
15	187.475	114.921	163.186	187.344	187.320
16	75.720	113.060	21.585	76.415	75.720
17	0.000	101.459	-0.814	0.075	0.000
18	75.720	112.896	38.656	76.400	75.720
19	75.720	112.964	39.976	75.955	75.720
20	75.720	112.773	39.425	76.032	75.720
21	75.971	112.823	42.264	76.468	75.720
22	275.762	124.646	246.944	271.649	272.550
23	54.052	115.948	36.765	47.778	46.630
24	0.000	101.601	-0.842	0.004	0.000
25	358.344	125.175	295.602	359.590	360.250
26	360.641	125.028	308.523	361.371	362.200
27	343.631	124.001	280.172	359.487	360.550
28	360.641	124.311	299.244	360.180	360.250
29	360.641	123.967	319.873	364.986	364.650
30	189.220	115.778	118.439	189.152	189.120

Figure 1: Predcition Report

From the visualizations it was found that Stochastic Gradient Descent model, predicts nearly to the target value when compared to SVM, ANN and Random Forest. Hence from prediction report and visualizations, it found that SGD and Random Forest prediction was approximately and nearly accurate to the target value. To be more precise about the model, a 10-fold cross validation sampling technique is performed, to calculate and compare the performance of the models.

5. Performance Evaluation Metrics

From the prediction report and visualizations, Stochastic Gradient Descent (SGD) and Random Forest (RF) showed better accuracy when compared to that of SVM and ANN. To produce much more precise model, a 10-fold cross validation sampling method is applied with Root-Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE) and Coefficient of determination (R2) as evaluation metrics.

5.1 k-fold cross validation

In k-fold cross validation technique, the dataset is divided into k-folds and at each iteration; different folds are reserved for testing and remaining as training sets. [30].

5.1.1 Root-Mean Square Error (RMSE)[30][31]

The Root-Mean Square Error used for evaluating regression based datasets as well as probability based dataset. It calculates the square of the error rate between actual values and target values. The formula for computing RMSE is as follows:

RMSE (f) = sqrt (1/m

Where m denotes number of test sample, f (xi) is the output on xi and yi is the actual value.

5.1.2 Mean Square Error (MSE) [31]

The Mean Square Error estimates the squares of the errors between the predicted and actual value and takes average result from the calculation. MSE is always non-negative and models having values closer to zero is a better model.

MSE = 1/n

5.1.3 Mean Absolute Error [31]

Mean absolute error is the average absolute variance between the actual and predicted values. Suppose there are n points in scatter plot and point i has (xi,yi) as co-ordinate, mean absolute error is the vertical and horizontal distance between the points where Y=X.

MAE= ()/n

5.1.4 Co-efficient of determination (R2) [31]

The co-efficient of determination R2 computes square of the correlation(r) between predicted y values and actual y values. It ranges from 0 to 1. If R2 is zero target prediction will be wrong and if it is 1 target can be predicted without errors. If R2 is 0.20 it indicates that only 20 percent of target value can be predicted from known values.

R2 = {(1/N) *

where N indicates number of interpretations used to fit the model, Σ is the summation symbol, xi is the x value for remark i, x is the mean x value, yi is the y value for observation i, y is the mean y value, σx is the standard deviation of x, and σy is the standard deviation of y.

5.1.5 Report on evaluation metrics

From the above sections four evaluation metrics namely RMSE, MSE, MAE and R2 were castoff to evaluate the performance of the machine learning models used for the experiment. The report is drawn for the dataset with feature selection namely ClassifierAttributeEval and without feature selection step. It is clearly seen that, the accuracy of the models is better after performing dimensionality reduction, when compared to the models executed without feature selection. The report clearly states that, the Stochastic Gradient Descent method has less error rates with MSE as 1.102, RMSE as 1.050, and MAE as 0.613 and R2 1.000. Hence from the above predictions and evaluation metrics it is found that SGD has good prediction and better performance evaluation criteria when compared to that of SVM, ANN and Random Forest. A 10-fold cross validation technique along with above mentioned evaluation metrics were performed on the feature selected dataset and the report is as follows:

Test & Score

Settings

Sampling type: 10-fold Cross validation

Scores

Method	MSE	RMSE	MAE	R2
SVM	13363.266	115.600	95.465	0.165
SGD	0.034	0.183	0.071	1.000
Random Forest	60.461	7.776	2.095	0.996
Neural Network	2569.688	50.692	27.081	0.839

Figure 2: Epoch 1

Test & Score

Settings

Sampling type: 10-fold Cross validation

Scores

Method	MSE	RMSE	MAE	R2
SVM	12257.411	110.713	89.851	0.234
SGD	1.102	1.050	0.613	1.000
Random Forest	155.054	12.452	2.698	0.990
Neural Network	1187.357	34.458	21.384	0.926

Figure 3: Epoch 2

6. Conclusion

Power predictions in educational institutions are very necessary as the power requirements may vary from each department based on even semester, odd semester, working day, non-working day, special classes, evening classes, internal exam, and university / model exam. In order to predict the power requirement in advance, four machine learning models namely Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF) and Stochastic Gradient Descent models were applied. Based on the prediction report and 10-fold cross validation sampling technique with Root-Mean Square Error (RMSE), Mean-Square Error (MSE), Mean Absolute Error (MAE) and Co-efficient of determination (R2) as the evaluation metrics for the models, Stochastic Gradient Descent prediction was accurate with Mean Square Error (MSE) of 1.102, Root Mean Square Error (RMSE) of 1.050, Mean

Absolute Error (MAE) of 0.613 and Coefficient of determination (R2) of 1.0000.

References

[1] Jihoon Moon, Jinwoong Park, Sanghoon Jun, Forecasting power consumption for higher educational institutions based on machine learning,Sprnger (2017).
[2] Morton A., Nagle N., Piburn J., Stewart R.N., McManamay R, A Hybrid Dasymetric and Machine Learning Approach to High-Resolution Residential Electricity Consumption Modeling, Advances in Geographic Information Science, Springe(2017).
[3] Cyril Voyant, Gilles Notton, Soteris Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte, Alexis Fouilloy, Machine Learning methods for solar radiation forecasting: a review, Renewable Energy (2016).
[4] Thanchanok Teeraratkul, Daniel O'Neill, and Sanjay Lal, Shape-Based Approach to Household Electric Load Curve Clustering and Prediction, IEEE Transactions on Smart Grid (2016).
[5] Daut, M.A.M., Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review, Renewable and Sustainable Energy Reviews, Elsevier (2016).
[6] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renewable and Sustainable Energy Reviews (2017).
[7] Luis M. Candanedo, Véronique Feldheim, Dominique Deramaix, Data driven prediction models of energy use of appliances in a low-energy house, Energy and Buildings ,Elsevier (2017).
[8] Wang, Z., Renewable and Sustainable Energy Reviews, Elsevier (2017)
[9] Greig Paterson, Dejan Mumovic, Payel Das, and Judit Kimpian, Energy use predictions with machine learning during architectural concept design, Science and Technology for the Built Environment (2017).
[10] Pan Li, Baosen Zhang, Yang Weng, and Ram Rajagopal, A Sparse Linear Model and Significance Test for Individual Consumption Prediction, IEEE Transactions on Power Systems (2017).
[11] Chun-Nam Yu, Piotr Mirowski, Tin Kam Ho, A Sparse Coding Approach to Household Electricity Demand Forecasting in Smart Grids, IEEE transactions on smart grid(2016).
[12] Cyril Voyant, Fabrice Motte, Alexis Fouilloy, Gilles Notton, Christophe Paoli, Marie-Laure Nivet, Forecasting method for global radiation time series without training phase: comparison with other well-known prediction methodologies, Energy (2016).
[13] Oğuz KAYNAR, Halil ÖZEKİCİOĞLU, Ferhan DEMİRKOPARAN, Forecasting of Turkey's Electricity Consumption with Support Vector Regression and Chaotic Particle Swarm Algorithm, Yönetim Bilimleri Dergisi/Journal of Administrative Sciences Cilt (2017).
[14] Oğuz KAYNAR, Halil ÖZEKİCİOĞLU, Ferhan DEMİRKOPARAN, Forecasting of Turkey's Electricity Consumption with Support Vector Regression and Chaotic Particle Swarm Algorithm, Yönetim Bilimleri Dergisi/Journal of Administrative Sciences Cilt (2017).
[15] Wi YM, Kong S, Lee J, Joo SK ,Demand-side management program planning using stochastic, J Electr Eng Technol (2016)
[16] Son H, Kim C, Short-term forecasting of electricity demand for the residential sector using weather and social variables,Resources, conservation and recycling (2016).
[17] Grolinger K, L'Heureux A, Capretz MA, Seewald L, Energy forecasting for event venues: Big data and prediction accuracy. Energy Build 112:222–233(2016).
[18] Kyung-Bin Song, Young-sik Baek, Dug-Hun Hong,G.Jang, Short-term load forecasting for the holidays using fuzzy linear regression method,IEEE transactions on power systems(2005).
[19] A.D. Papalexopoulos, T.C.HesterBerg, A regression-based approach to short-term system load forecasting, IEEE transactions on power systems(1990).
[20] Seunghyeon Wang ,Hyeonyong Hae and Juhyung Kim, Development of Easily Accessible Electricity Consumption Model Using Open Data and GA-SVR,Energies (2018)
[21] Wu J, Wang J, Lu H, Dong Y, Lu X, Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model, Energy Convers Manag (2013).
[22] Aowabin Rahman, Vivek Srikumar, Amanda D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Applied Energy (2018).
[23] A.S.Ahmad , M.Y.Hassan , M.P.Abdullah , H.A.Rahman , F.Hussin , H.Abdullah , R.Saidur, A review on applications of ANN and SVM for building electrical energy consumption forecasting, Renewable and sustianable Energy reviews (2015)
[24] M.E. Rinker, Sr. School of Construction Management, University of Florida, Gainesville, A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models, Renewable and sustianable Energy reviews (2016)
[25] WILLIAM HEDÉN, Predicting Hourly Residential Energy Consumption using Random Forest and Support Vector Regression, TRITA-MAT-E 2016:20
[26] Selvam Nallathambi, Prediction of electricity consumption based on DT and RF: An application on USA country power consumption, IEEE Explore (2017).
[27] Muhammad WaseemAhmad, MonjurMourshed, YacineRezgui, Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption, Energy and buildings (2017).
[28] David Rosenberg,Gradient and Stochastic Gradient Descent, New York University(2017).
[29] J, ustin Sirignano, Konstantinos SpiliopoulosStochastic Gradient Descent in Continuous Time: A Central Limit Theorem, Machine Learning (stat.ML),(2017).
[30] Ming Jin, Lin Zhang, Costas J. Spanos, Power Prediction through Energy Consumption Pattern Recognition for Smart Buildings, IEEE Xplore(2015).
[31] Khuram Pervez Amber, Muhammad Waqar Aslam 2, Anzar Mahmood, Anila Kousar, Muhammad Yamin Younis , Bilal Akbar, Ghulam Qadar Chaudhary and Syed Kashif HussainEnergy Consumption Forecasting for University Sector Buildings, Energies (2017).
[32] Evlauating Machine learning models, Jose Hernadas Orello, universitat politecnice vaencia