Short Term Load Forecasting based on Machine Learning Algorithms using Smart Meter Data for Healthcare Applications

Sukanya Sonawane
Department of Computer Science and Engineering
Usha Mittal Institute of Technology
Mumbai, India
sukanyasonawane112@gmail.com

Kirti Thakur
Department of Computer Science and Engineering
Usha Mittal Institute of Technology
Mumbai, India
kirtithakur694@gmail.com

Prof. Narendra Gawai
Department of Computer Science and Engineering
Usha Mittal Institute of Technology
Mumbai, India
ngawai@rediffmail.com

***

Abstract - *These days, there is an ever-rising migration of people to urban areas. Energy consumption and health care facilities are the most challenging aspects that is greatly affected by the large inrush of people to urban areas. Also, urban communities around the globe are investing heavily with an end goal to provide better ecosystems to individuals [1]. In this transformation, many smart devices are equipped around the world, which provides a huge amount of refined and categorized data which can be analyzed to support smart city services. Many countries are rolling out smart electricity meters. Smart meters are advanced electricity meters for monitoring electric power consumption in homes that brings new benefits when compared to existing traditional meters. These include enhancing load forecasting for the electricity grid and minimize energy demand.*

For the assessment of the proposed system, we utilized the UK Domestic Appliance Level Electricity data-set (UK-Dale) time-series data of power consumption gathered from 2012 to 2015 for 5 houses [2]. In view of data points gathered from real time environment we conducted tests to show that smart meters data can be utilized as a solid spine of a short term load forecasting. We used four machine learning algorithms Long Short-term Memory (LSTM), Support Vector Machines Regression (SVR), Decision Forest Regression with Ada-Boost and Nearest Neighbors Regression to examine the performance of load forecasting system. Monitoring change in energy usage can allow us to indirectly assess an individual's well-being or state of mind. The ability to monitor the power consumption of individual appliances can thereafter be used for home health care.

Key Words: ***Forecasting, save electricity, LSTM, Machine Learning, SVM, kNN, Random Forest, Big data, Smart cities, Smart homes.***

1. INTRODUCTION

It is estimated that by the end of the year 2020, the number of smart meters in the world is expected to increase up to 780 million [3]. Nowadays, there are more than 50 million smart meters installed in the US. Electrical energy generation and distribution is a complex and expensive affair. To lower the cost of energy production, efficient grid management plays a major role. Grid management involves planning for load demand, maintenance of generation units, supply lines and efficient load distribution across the supply line. Therefore, an exact load forecast will increase the productivity of planning procedure of power generating companies. There's a solid commitment of power utilities with load disaggregation because of their enthusiasm to provide a better service to customers. Power generation companies do their plan on the basis of data collected manually. Hence real-time prediction is not possible. Real-time forecasting can be made possible if data can be collected in real-time. Enormous volumes of fine-grained and indexical data that is generated by smart devices can be analyzed to give real-time forecasting. In this paper, we have presented that real-time load forecasting is possible with the help of state of the art machine learning algorithms LSTM Network, Nearest Neighbors Regression, Support Vector Regression, and Decision tree Regression with AdaBoost. Smart grid technology can be used to provide useful information on the activities of daily living and can be used to monitor both the short-term and long-term health of individuals [4]. Detailed electricity usage patterns and trends can be identified to help understand daily consumer habits and routines [5]. Understanding how different factors can influence energy usage can provide a better understanding of people's habits and routines which can be used to identify normal and abnormal behaviors. Load forecasting can be used to monitor changes in energy usage that can allow us to indirectly assess a person's wellbeing or state of mind. Based on historical data, if we detect change while predicting electric power, it means that it is an abnormal behavior. For example, a consumer suddenly using little or no electricity could signify they are in difficulty or danger. The smart grid technology can therefore be used to identify reoccurring patterns which can be used for health care applications.

2. RELATED WORK

Load forecasting prediction was done by many others using different algorithms. Short-Term Residential Load Forecasting Based on Resident Behavior Learning is studied in [6]. Kong, W. proposed an LSTM based deep learning forecasting framework with appliance consumption was compared with the Feed Forward Neural Network (FFNN) and K-nearest neighbors (KNN). According to the paper, if a lifestyle pattern of a resident can be learned better meter forecasting can be achieved. They showed that forecasting accuracy can be improved using appliance measurements. The work in [7] proposes Short term power load forecasting using Deep Neural Networks. The authors proposes the combination of Discrete Wavelet Transform (DWT) and Artificial Neural Network/ Support Vector Machine (STLF) to accurately forecast the short-term loads. Hybrid methodology for short-term load forecasting is presented in [8]. The authors propose the combination of Discrete Wavelet Transform (DWT) and Artificial Neural Network/Support Vector Machine(STLF) taking into consideration temperature, humidity dew point and load consumed a particular day at particular hour to accurately forecast the short-term loads. Smart meters data is also used in [9] where K-means clustering algorithm was utilised to cluster similar individual consumers and fit distinct models per cluster. Public holidays were taken into consideration for changing customer behaviour,as was periodicity of the day, week and year. In [10], an experimental demonstration based on physical-statistical approach to improving forecast accuracy is presented. It is discussed to some currently popular machine learning models such as Support Vector Machine (SVM) and ANN A new method is proposed to address the heterogeneity challenge. The authors in [11] use Recurrent Neural Networks (RNN) to capture time dependencies and propose a novel energy load forecasting methodology based on sample generation and Sequence-to-Sequence (S2S) deep learning algorithm. In this paper first physical statistical/heterogeneous building energy modeling approach is proposed and validated. Taking in consideration of the above mentioned work we have implemented an smart meter data based load forecasting system. The core algorithm of the forecasting system is a machine learning algorithm. To select best performing algorithm we have tested performance of several machine learning algorithm with a new dataset called "The UKDALE dataset" [2].

3. PROPOSED MODEL

Following are the steps to give Short Term Load Forecasting.
- Data collection from Smart meters
- Preprocessing and Filtering
- Training Machine Learning Models
- Load Forecasting

Smart meter devices are configured to upload power usage data to the server. The central processing unit is configured to do the calculation based on a selected machine learning algorithm. Figure 1 shows the workflow of the system. The main idea of this paper is to forecast total power consumption based on the data collected from smart meter devices. To determine which prediction algorithm works best for our system we have implemented four machine learning algorithms and compared their performance.

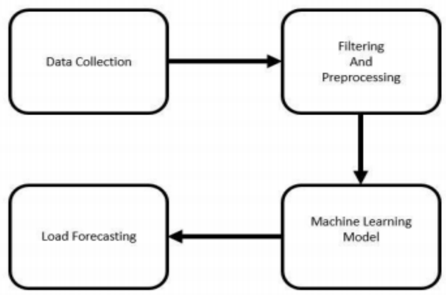

Figure -1: Proposed Model

3.1 Data Collection and Preprocessing

The dataset includes 400 million raw records at a time resolution of 6 seconds. In the preprocessing stage of the cleaning process, we developed customized procedures to remove noises from the data and prepare it for mining. After preprocessing, the dataset is reduced to 20 million. Smart meters records time-series raw data, which is a high time resolution data and it is transformed into a 1-minute resolution load data. It is then further subsequently translated into a 30 minutes time-resolution source data, 48 readings per day per appliance while recording start time and end time for each active appliance.

3.2 Load Forecasting

In particular, the following machine learning algorithms for forecasting were used:

3.2.1 Nearest Neighbor Regression: KNN regression is used to calculate the average of the targeted value of k nearest neighbors. Using the inverse distance weighted average of the k nearest neighbors can also be calculated. KNN classification and KNN regression use the same distance functions. KNN regression is used to calculate the average of the targeted value of k nearest neighbors. Using the inverse distance weighted average of the k nearest neighbors can also be calculated. KNN classification and KNN regression use the same distance functions. With the help of the following functions distance between neighbors are measured[12]:

Euclidean-

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (1)$$

Manhattan-

$$\sum_{i=1}^{k} |x_i - y_i| \tag{2}$$

Minkowski

$$\left(\sum_{i=1}^{k} (|x_i - y_i|)^q\right)^{1/q} \tag{3}$$

The above equation can only be used for continuous variables. While inspecting the data the ideal value for K is chosen. With a large K value, the noise is reduced but it becomes harder to detect the distinct features. To determine good K value using independent data set to validate K values, cross-validation is an ideal way. The ideal K for most datasets is 10 or more which produces better results than 1-NN.

3.2.2 Support Vector Machine: Support Vector Machine can also be used as a method of regression, keeping all the fundamental elements intact that designate the algorithm (maximal margin). With scarcely trivial distinction, the Support Vector Regression utilizes the same postulates as the SVM for categorization. Assuming a set of training points $(x_1, y_1), \dots, (x_n, y_n)$ where x_i a feature vector and $x_i \in R^n$ and $y_i \in R^1$ the target output. If the given parameters $C > 0$ and $\epsilon > 0$ then according to Vapnik[13]. Support Vector Regression is

$$\sum_{i=0}^{n} (-\alpha_i + \alpha_i^*)K(x_i, x) + b \tag{4}$$

3.2.3 Decision Tree Regression with AdaBoost: The decision tree establishes regression models in the form of a tree structure. It simultaneously takes down dataset onto smaller and smaller subsets and correct decision is incrementally established. The final result is a tree with respective decision nodes and leaf nodes. A numerical target is illustrated by a leaf node. The highest decision node in a tree that becomes equivalent to the leading predictor is called the root node. Both categorical and numerical data can be conducted by decision trees[14]. To increase the accuracy of the model we boosted the model using AdaBoost.

Assuming training vectors $x_i \in R^l$, $i = 1, \dots, n$ and targeted values containing vector $y \in R^n$ a partition of the space is made recursively such that the samples with the same labels are in a group. Suppose the data at node k be represented by P. Splitting is done using $\theta = (j, t_k)$ an attribute j and threshold t_m. After partitioning the data is kept into $P_{left}(\theta)$ and $P_{right}(\theta)$ subsets $P_{left}(\theta) = (x, y)|x_j \leq t_k$ and $P_{right}(\theta) = P \setminus P_{left}(\theta)$. Noise at k is calculated using a function H() which calculates impurity. The choice of function depends on the method of solving (classification or regression).

$$G(P, \theta) = \frac{n_{left}}{N_k} H\left(P_{left}(\theta)\right) + \frac{n_{right}}{N_k} H\left(P_{right}(\theta)\right) \tag{5}$$

3.2.4 Long Short Term Memory: LSTM networks are renowned for their ability to remember patterns and sequences [15]. Human behavior tends to be repetitive. From this intuition, we used LSTM Network to learn the behavior pattern of power usages. LSTM network can give prediction based on the calculation of cell state, forget gate. The input given to an LSTM (Long Short Term Network) is a 3D matrix. The dataset we have used in this project is a 2D matrix which consists of columns and rows. The number of columns corresponds to the dimension of a feature vector and the number of rows corresponds to the number of data points. The input of the LSTM Network consists of another dimension of the matrix, which corresponds to time steps. LSTM network uses this time steps to keep track of the previous occurrence. To prevent over fitting of the model a threshold was given. A program always keeps track of the difference between loss of training set and loss of test set. When the loss of training set is decreasing and loss of test set is increasing and the difference is bigger than the threshold, it stops training the model. LSTM network can give prediction based on the calculation of cell state; forget gate Equation 7, input gate Equation 8 and output gate Equation 6[16]

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{6}$$

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{7}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{8}$$

4. Experimental Work

5.1 Experimental Protocol

In this paper, we have used "The UK-DALE dataset" created by Jack Kelly and William Knottenbelt . This dataset contains the appliance level disaggregated power consumption record as well as aggregated whole house power consumption record. In Figure 2 whole house power consumption of the last year is given. We assumed that all the appliances in these datasets are smart meter devices that record the power consumption of appliances and uploads in the database. For training and testing we have used data of house 1 because it contains a maximum number of appliances. Also, they have given more emphasis on recording house 1 data.

The main dataset contains five folders. Each folder corresponds to each house. Under each house numbers of CSV files according to the number of devices are given. Each CSV file contains records of power consumption with time. In each CSV file time is give in the format of the UNIX time epoch. Interval data recording is six seconds.

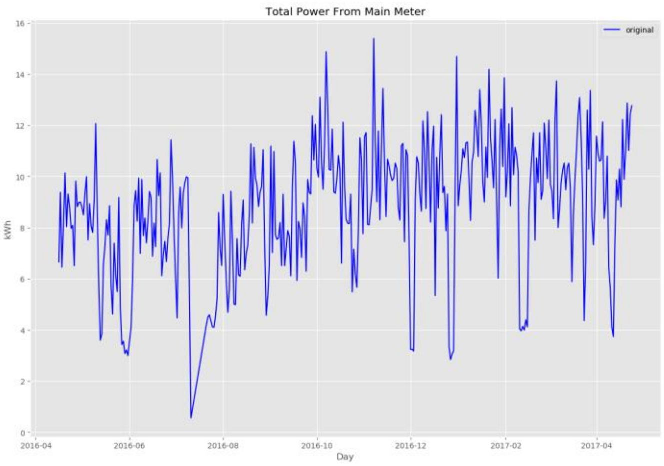

Figure 2. Total Power Consumption of last 1 year in KWh

Noise and misleading data in any dataset are bad for any model to train on. A misleading dataset will eventually produce a hypothesis which will not do well in unseen data. Therefore noise cancellation has done with great care. The dataset contains UML configuration file for every houses. The file has details description of meter devices and appliances. Each appliance has upper bound and lower bound of power consumption. Any power consumption beyond that limit is considered as noise or bad reading. The most frequently used measures of the differences between predicted values by an estimator and the actual values are the Root Mean Square Error (RMSE). The RMSE of predicted data y^t for survey of i, for variables y_i calculated for n numbers of cases using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

The parameters for building the predictive models of the algorithms are:

1. Nearest Neighbors Regression: The total number of n neighbors is 15.
2. Support Vector Regression: Penalty parameter C=1.0, epsilon=0.2 (specifies the epsilon-tube within which no penalty is associated in the training loss function with points predicted within a distance epsilon from the actual value).
3. Decision Tree Regression with AdaBoost: max depth=16, n estimators=300.
4. LSTM: the network contains on LSTM cell. Input of every time step contains a sequence of 7 samples.

5.2. Results

As our goal is to predict the power consumption of the house the next day depending on the power usage of the appliances of the present day. Model is evaluated on test set. The error result is obtained from test set. For preprocessing we scaled the data using the library function scale (). We used Kfold splits for splitting the dataset as it's a time series dataset. Later we trained our model on train set and the error rate is acquired by evaluating the models against our test set. We used all the default settings of the library and few changes in parameters of few algorithms. We used an API out of 3 APIs of the library to measure our trained model's performance. The API scoring parameter contains model-evaluating tools using cross-validation depends on an internal scoring strategy. We used this API to find the MSE of our models then used square root to find out RMSE and we compared the algorithms based on the result of the API.

Nearest Neighbors has RMSE of 1.9331727. As we are to predict the total power consumption of a house of the next day depending on the usage of the today's power consumption of the appliances, for which we are getting this much higher RMSE value. In Figure 3 the comparison of the real value and the predicted value of the trained model Nearest Neighbors Regression is given where x-axis is the date and y-axis is power consumption of the house in KWh.

Using the kernel RBF performs better than other kernels in SVR. RBF kernel performs better in this context because of the data. The higher the degree, the performance of other kernels are worse than RBF. SVR using kernel polynomial and Gaussian Radial Basis function with degree of 3 has RMSE of 2.1229618 and 1.8341087 respectively. But kernel linear performs slightly lesser than RBF but better than the polynomial as it has a degree of 1. It has RMSE of 1.8474361. For having the lowest RMSE among the kernel function, we preferred the kernel function RBF for SVR. In Figure 4 the performance of Support Vector regression (kernel RBF) is shown by comparing the true value with predicted value where x-axis is date and y-axis is the power consumption of the appliances in KWh.

The predicted power consumption is the output of the model. The RMSE of Decision Tree Regression with the AdaBoost algorithm-based model is 1.9202281. As we restricted the depth to 16 and we selected all the features to train the model we are getting much high RMSE. The deeper the tree is the model tends to over fit. To avoid over fitting we specified the maximum depth of the tree. In Figure 5 Evaluation of the model Decision Forest Regression with AdaBoost is shown by comparing predicted and real value. In Figure 5 x-axis is the date and the y-axis is the power consumption of the house in KWh.

In this experiment, we used a single-cell LSTM network. We have tested with the LSTM network with up to 3 LSTM cells stacked top of one another. Stacking more than one LSTM cell made computation heavier but did not give better result. In some cases, it went bad. We have also experimented with the length of look back. Here look back is how many samples is given as an input in each time step. We have tested the variable length of look back. Most significant were 7 for 7 days, 15 for 15 days, 30 for 1 month. Length of look back (between 7 and 15), have given better results than longer look back like 30.

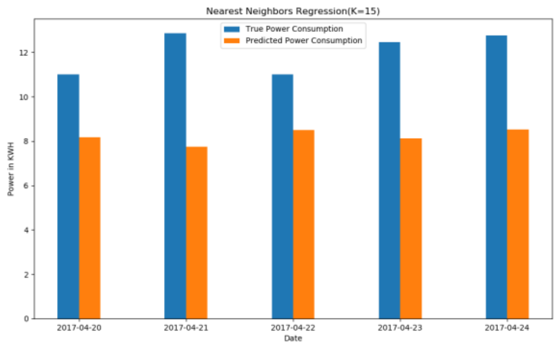

Figure 3 - Nearest Neighbors Regression, empirical comparison

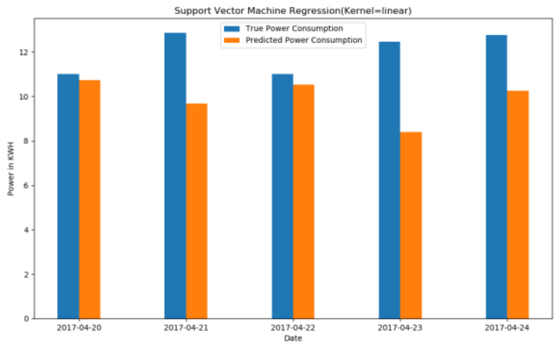

Figure 4- Support Vector Regression, empirical comparison

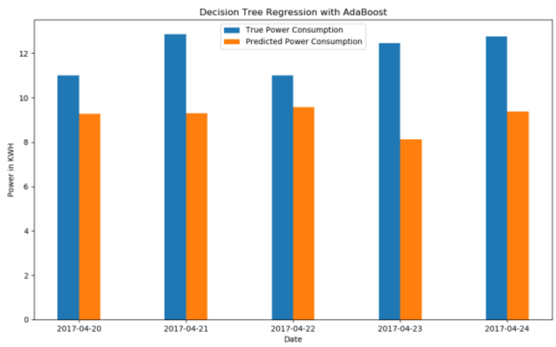

Figure 5-Decision Tree Regression with AdaBoost, empirical comparison

We have found best result in look back length 7. In Table II we can see LSTM has given lowest RMSE score. Reason of the lowest score is ability of a LSTM to process sequence of samples rather than a single sample.

TABLE II. RMSE OF ML MODELS

Algorithms	Error
	Root Mean Square Error
Nearest Neighbors Regression	1.93
Support Vector Regression	1.83
Decision Tree Regression with AdaBoost	1.86
Long Short Term Memory	1.82

In Figure 6 we can see that predicted data points by a LSTM almost catches the pattern of electricity usages.

In Figure 8 an overall summary of the outputs of four algorithm is given.

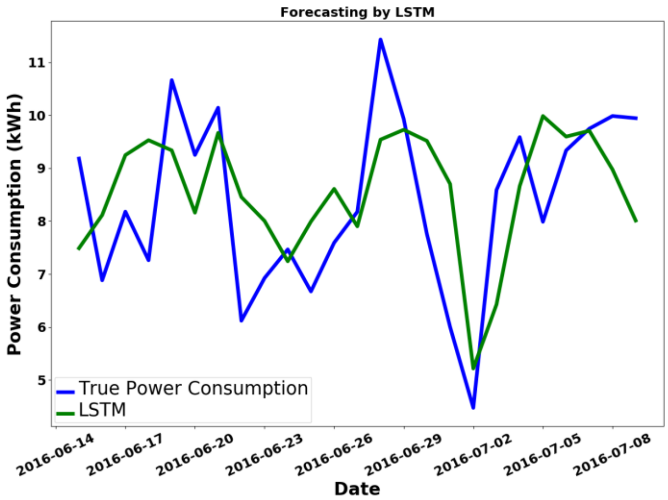

Figure 6. Load forecasting by LSTM

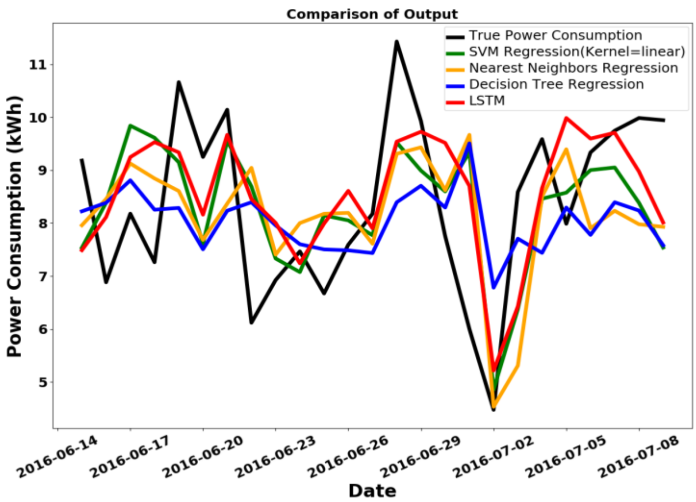

Figure 8. Comparison of four ML Algorithms based on their outputs on same test set.

4. CONCLUSION

Day by day our processors are getting stronger and less power hungry. Likewise processors are accompanied with dedicated core for neural net and artificial intelligence. Therefore cost and computational power required for preparing a neural system won't be issue later on. Neural networks like LSTM has the ability to adopt with a great variety of patterns and the ability of recognize those pattern. In our evaluation, LSTM has given better result with compared to Support Vector Machines Regression, Decision Forest Regression with AdaBoost and Nearest Neighbors Regression. In this paper we have introduced a framework which can give prediction based on data collected from smart meters data. To demonstrate the dependability of the framework we have tested the framework with real world datasets. We have conducted several experiments to estimate the performance of four machine learning algorithms and concluded the experiment with a comparison of RMSE loss score. Long Short Term Memory network has given lowest RMSE in the experiment.

5. FUTURE WORK

The system is based on data collected from smart home meters. This framework can be implemented for home administrations and grid management. Models can be improved if the data clustering is based on the time interval of hours. This might lower the error of the models. In the future, we are anticipating to analyze models utilizing different regression-based ML algorithms. Privacy is a major worry here. In the future, we additionally need to take a shot at the security side of this framework. Predictions can be improved by selection of features and changing the parameters. Also implementation of current and future grid produces extensive datasets which can also be used by researchers, customers and energy providers to build an accurate representation of user behaviour. This representation can be used in the field of health, where support from healthcare providers directly involved in independent living care provisions can be seeked.

ACKNOWLEDGEMENT

We are indebted to our guide, Mr. Narendra Gawai, Assistant Professor in the Department of Computer Science Technology Engineering who guided us in the research pro- cess. We also want to acknowledge the contributions of Dr. Sanjay Pawar, Head of the Department, Computer Science Technology Engineering. We are very much obliged to our project coordinator, Ms. Sonali Bodekar, Assistant Professor in the Department of Computer Science Technology Engineering who was instrumental in familiarizing us with the technologies.

REFERENCES

[1] M. S. Hossain, "Cloud-supported cyber-physical localization framework for patients monitoring," IEEE Systems Journal, vol. 11, no. 1, pp. 118–127, March 2017.

[2] K. Jack and K. William, "The UK-DALE dataset, domestic appliancelevel electricity demand and whole-house demand from five UK homes," Scientific Data, vol. 2, no. 150007, 2015.

[3] Telefnica. 2014. "The Smart Meter Revolution. Towards a smarter future." (2014),11.

[4] Thomas, T., Cashen, C., Russ, S. (2013). "Leveraging smart grid technology for home health care." 2013 IEEE International Conference on Consumer Electronics (ICCE). doi:10.1109/icce.2013.6486892

[5] Chalmers, C., Hurst, W., Mackay, M., Fergus, P. (2015). "Smart meter profiling for health applications". 2015 International Joint Conference on Neural Networks (IJCNN). doi:10.1109/ijcnn.2015.7280836

[6] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "ShortTerm Residential Load Forecasting Based on Resident Behaviour Learning," IEEE Trans. Power Syst., vol. 33, no. 1, pp. 1087–1088, Jan. 2018.

[7] G. M. U. Din and A. K. Marnerides, "Short term power load forecasting using Deep Neural Networks," in 2017 International Conference on Computing, Networking and Communications (ICNC), 2017, pp. 594598.

[8] P. Ray, S. Sen, and A. K. Barisal, "Hybrid methodology for short-term load forecasting," in 2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), 2014, pp. 16.

[9] Kell, A., McGough, A. S., and Forshaw, M. (2018). "Segmenting Residential Smart Meter Data for Short-Term Load Forecasting." Proceedings of the Ninth International Conference on Future Energy Systems - e-Energy 18.doi:10.1145/3208903.3208923

[10] L, X.; Lu, T.; Kibert, C.J.; Viljanen, M. Modeling and forecasting energy consumption for heterogeneous buildings using a physicalstatistical approach. Appl. Energy 2015, 144, 261275.

[11]Sehovac, L., Nesen, C., & Grolinger, K. (2019). "Forecasting Building Energy Consumption with Deep Learning: A Sequence to Sequence Approach," 2019 IEEE International Congress on Internet of Things (ICIOT). doi:10.1109/iciot.2019.00029

[12] N. S. Altman, "An Introduction to Kernel and Nearest Neighbor Nonparametric Regression," Am. Stat., vol. 46, no. 3, pp. 175—185, Aug. 1992.

[13] V. N. Vapnik, The Nature of Statistical Learning Theory. New York, NY: Springer New York, 2000.

[14] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees'' J. Anim. Ecol., vol. 77, no. 4, pp. 802—813, Jul. 2008.

[15] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neu- ral Comput., vol. 9, no. 8, pp. 1735—1780, Nov. 1997.

[16] F. A. Gers, F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM," NEURAL Comput., vol. 12, pp. 2451—2471, 1999.