March - April 2020
ISSN: 0193-4120 Page No. 8083 - 8090

A Smart Building Energy Management using Internet of Things (IoT) and Machine Learning

NursyuraLaili Mazlan[1], Nor Azuana Ramli[2], LilikJamilatul Awalin[3], MohdBadrulhisham Ismail[4], Azizah Kassim[5], Ashwin Menon[6]
[1,2,3,4]Electrical Engineering Section, Universiti Kuala Lumpur British Malaysian Institute, Gombak, Selangor, Malaysia
[5]Green 3 SdnBhd, Petaling Jaya, Selangor, Malaysia
[6]Reneon Technologies Sdn. Bhd
nursyura.mazlan27@s.unikl.edu.my[1], norazuana@unikl.edu.my[2], lilik@unikl.edu.my[3], badrulhisham.ismail@unikl.edu.my[4], azizah.kassim@gmail.com[5], ashwin@reneontech.com[6]

Article Info
Volume 83
Page Number: 8083 - 8090
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 09 April 2020*

Abstract

Energy consumed by smart buildings in Malaysia accelerates continuously due to the growth of country's population. This occurrence creates awareness towards energy management team to organize and manage their energy consumption systematically. The main purpose of this research is to analyse and predict the energy consumptions in order to achieve a better energy management of a commercial smart building towards efficiency. In this paper, historical data of hourly consumption of maximum demand collected at the selected tenants of the smart commercial building implemented with Internet of Things (IoT) has been analysed using statistical method computed with the formula of mean, variance, skewness and kurtosis. k-nearestneighbour (k-NN) method has been applied on the data of consumptions for prediction process by using three different values of nearest neighbour. The predicted data has been separated into different training and testing ratios which are 70% and 30%. Root mean square error (RMSE) is proposed in this paper to evaluate the performance of predicted data. The results showed that the nearest neighbour with k = 5 is the most accurate since it provides the lowest average RMSE value with 5.73, 8.54 and 0.35 for each tenant of the building respectively. This model will be used as a reference to predict the energy usage for the upcoming period. In the future, data that has been predicted can be integrated into the available system for user monitoring and controlling purposes.

Keywords; *Internet of Things, smart building, maximum demand, machine learning, k-nearest neighbour*

I. INTRODUCTION

Energy consumptions in Malaysia increase gradually due to the growth of population which is in year of 2018, it hits 32 million people and will be estimated to increase more in the future [1]. The growth of population lead to the increasing of energy demand in this country and have been estimated to reach 116 million tons of oil equivalents (mtoe) by the year 2020. Energy provided in Malaysia is influenced by the main fossil fuel sources which included coal, natural gas and fuel oil. Buildings which including commercial, residential and industrial in our country utilizes a total of 48% of the electricity that have been created [2]. The increasing of energy consumptions towards buildings from day to day create enforcement to this country in managing and reducing the energy consumption as much as possible in order to improve energy efficiency.

Energy efficiency can be improved by applying a smart system management towards a building. Smart building is a cutting edge technology structures that can improve energy efficiency by simplifying the system connection in using computerized process [4]. All the electrical appliances can be controlled digitally and offer handiness towards consumers in managing their energy management. A smart building with the help of Internet of Things (IoT) is the most effective ways in supporting energy efficiency management and monitoring energy consumption [5]. IoT provides a digital platform to observe and supervise the energy consumption in a huge scale of data in the presence of internet. This technology also helps in reducing the energy costs and charges by enhancing the way in managing the energy consumption of building [6]. It is crucial for consumers especially energy management team to have a deep comprehension on the pattern of their smart building energy consumption. Statistical method is one of the analysis techniques that are used to obtain data regarding the energy consumption of building [7]. Analysing of energy consumption is really important in improving energy efficiency and cost saving for a building [8]. Hence, statistical method is proposed as the main objective in this research is to do statistical analysis for energy consumption of a smart building.

Prediction also brings convenience towards consumers and energy management team in estimating their usage and planning for upcoming period to improve the energy efficiency of a smart building. To achieve the next objective of the research, k-nearest neighbour (k-NN) method is proposed in this study in order to predict energy consumption for the upcoming period. This method known as one of the prediction technique that helps in energy management efficiency by minimize the energy consumption towards a building [9]. Moreover, this method is considered to be one of the simplest classification techniques used for prediction

This method also is a non-parametric and simple learning algorithm that uses a database concerning the energy consumption to predict and make a classification of new sample data [11]. Prediction method will encourage energy management team to supervise their energy consumption that will lead the management system to become more manageable and systematic towards achieving energy efficiency. Predicting their energy consumption also becomes substantial in order to control the energy usage and reduce the expenses on energy cost for billing purposes.

All the data used for analysis and prediction method were collected based on real time data from a smart commercial building located in KlangValley, Malaysia. This paper is separated into four sections. Comparison between present studies with literatures that has been reviewed will be discussed in the first section. Next, all methods for analysis and prediction that have been applied in this study will be deliberated in details. The results and findings obtained from both methods will be discussed in the last section.

II. LITERATURE REVIEW

There are many different approaches for energy consumption analysis and prediction that has been done from the previous studies. Energy forecasting is really important in improving the energy utilization rate and helps the energy management system of a building to make a better evaluation [12]. For this present study, analysis and prediction method using statistical and k-NN method give attention on maximum demand measured in kilowatt for selected tenants of a two smart commercial building, building A and B. Study done by [13] analysed energy consumption of electricity, oil and gas using time series combined with regression method. Data that has been analysed will be used for prediction method affected by seasonal factors. Result for this study demonstrates the accuracy of the method used. Building A shows that the error for predicted annual electricity, oil and gas are 6.326% and 1.753% meanwhile for building B the predicted consumption error are 1.0221% and 5.885%.

Analysis using statistical method has been proposed by [10] on energy consumption of apartment's

Published by: The Mattingley Publishing Co., Inc.

building in order to split the power consumption used into two categories which are lower power consumption and high power consumption. Data of hourly power consumption for all home equipment based on their daily used has been collected in order to undergo data processing similar with this research which are mean, variance, skewness and kurtosis. k-NN method employed with Euclidian distance has been applied in this study for prediction process. Furthermore, predicted data undergo training and testing ratio of 60%-40% with the help of 10-Fold and 5-Fold cross validation to evaluate the findings. Results for [10] concluded that training and testing ratio used giving 95.9615% accurate results.

There are some other methods that have been applied in order to do the analysis and energy prediction for different types of building done by previous studies but they were not included in this study since this paper only focuses on analysis and prediction method that undergoes analysis using statistical method and k-NN.

III. RESEARCH METHODOLOGY

This research starts with the study on the smart building system using Internet of Things (IoT). The smart building in this case study is a shopping complex building located in Klang Valley, Malaysia. Figure 1 shows the system architecture of the smart commercial building that demonstrated the flow of data that have been measured at the smart building. The smart building equipped with IoT meters which are connected to the power inlet socket at two major tenants of the building. Each tenant is divided into two area consists of two IoT meters named tenant A1, A2, B1 and B2. Collected data per minute that mapped into TNB requirements were saved in an open source web server.

User can observe and monitor the collected data internally through an online platform by giving an ID number and password.

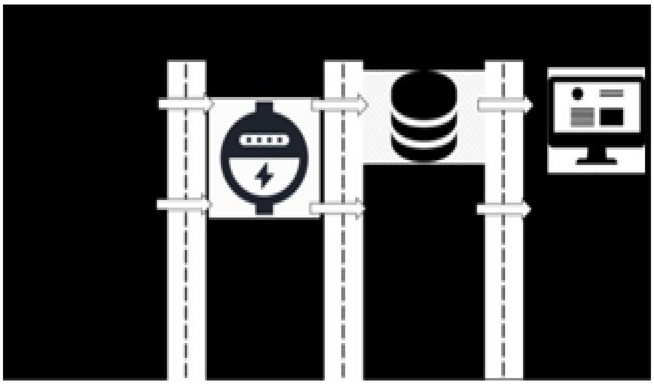

Fig. 1. System architecture of the smart commercial building

The online website really helps the users and team management in monitoring their building energy consumption data using systematic and manageable online system. It provides graph visualisation that shows the data collected for every hour in order to help the users in visualizing the behaviour or trend of the maximum demand of each tenant. For example, Figure 9 and 10 in the appendix shows the images that represented hourly data of maximum demand (MD) for tenant A1 and A2 collected by the smart meter reading that has been provided via the website in June 2018.

Collected data can be extracted manually from the online platform in the form of CSV file. Data of maximum demand consumption measured in kilowatt starting from June 2018 until December 2018 of each tenant were chosen as the main data used for analysis and prediction purposes that will be explained comprehensively in next subsection.

A. Data Analysis

Statistical method is applied to complete the analysis process computed with formula of mean, variance, skewness and kurtosis as equations below where N is the total number of hours. Total power consumption will determine the value of xi over ith hour of the day where i = 0, 1, 2,…., 23.

Highest values of collected maximum demand data for each month are retrieved in the form of CSV file. Afterward, all the data were imported into SPSS as

Published by: The Mattingley Publishing Co., Inc.

this analytic software already provides all the mathematical formulas. The purpose of doing the analysis is to know the normality of the maximum demand data itself before proceed to the next

method.

Mean, μ = (1)

Variance, σ = 2 (2)

Skewness, S = 3 (3)

Kurtosis, K = 4 (4)

B. Prediction by using k-Nearest Neighbour (k-NN) Method

In this study, k-nearest neighbour (k-NN) method has been chosen in predicting maximum demand. k-NN method is known as one of the machine learning algorithm as it is identified as a simple algorithm that able to do a classification and regression by storing all available cases and categorizes new cases based on a similarity measurement [10]. It is also known as lazy learning because of its training is held up to run time [14]. For classification, the output is determined by the majority of the value of its neighbours while regression's output is determined by the average of the k nearest neighbours. This method will be applied to predict the maximum demand based on the electrical usage measured in kilowatt (kW) at both tenants. Figure 2 shows the illustration in applying k-NN method in predicting maximum demand of the building.

Features of maximum demand used for k-NN method are voltage and current. This is because the multiplication of these two features will result the value of power which is resulting a maximum demand value. Voltage and current value are being selected from the CSV file that has been downloaded through the web browser. The numbers of neighbour is setting to 3, 4 and 5 which means, the classification of prediction will be included 3, 4 and 5 nearest neighbour of features to maximum demand. All the analysis will be done in SPSS software.

December 2018 by using statistical method.Descriptive statistics analysis table is the output provided by SPSS software that comprehends value of mean, variance, skewness and kurtosis. Figure 3 shows outcomes for analysis process on highest value of maximum demand data that has been completed for tenant A1, A2, B1 and B2.

Mean values provided in the table indicates the central values of maximum demand data. Value of 156.427, 58.766, 108.731 and 2.755 measured in kilowatt also indicates the peak position for graph distribution. As for variance, a large value of variance indicates that the numbers in the data set is far from the mean and each other while small variance indicates the opposite. For tenant A1, A2, and B2 the variance value calculated with 7.388, 7.805 and 1.366 shows that the data is closed to each other. Data for tenant B2 the data is far from mean and each other due to the large value of variance with 953.578.

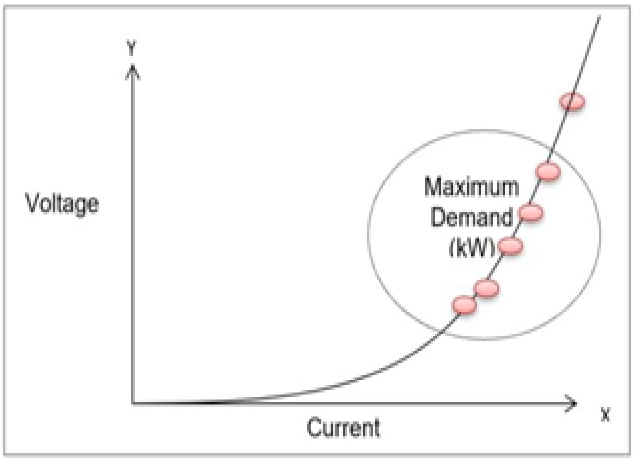

Fig. 2. The illustration of k-nearest neighbour method

The prediction process required historic data of the electrical consumption that has been collected at both tenants in a particular time to undergo prediction using nearest neighbour values which are 3, 4 and 5. The predicted data will be compared with the historical data for each values of nearest neighbour used. This prediction method used Euclidian distance function that is set as default in SPSS software.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8083 - 8090

C. Performance Evaluation

Root mean square error (RMSE) is used to evaluate the value of maximum demand that have been predicted using k-NN method. RMSE will calculates the difference between the predicted data and historical data value of maximum demand consumption for each month of every tenant using equation provided below where n is the total number of data, At is the historical values of maximum demand data and Ft is the predicted values. The accuracy of prediction value calculated by difference value of k was determined by RMSE calculation.

Root mean square (RMSE) = (5)

IV. RESULTS AND DISCUSSION

This section consists of results analysis of maximum demand consumption for each tenant from June 2018 until

Statistics

		demandA1	demandA2	demandB1	demandB2
N	Valid	7	7	7	7
	Missing	0	0	0	0
Mean		156.42714	58.76643	108.73071	2.75486
Variance		7.388	7.805	953.578	1.366
Skewness		-.911	.814	1.224	1.218
Std. Error of Skewness		.794	.794	.794	.794
Kurtosis		-.933	-1.041	-.844	-.831
Std. Error of Kurtosis		1.587	1.587	1.587	1.587
Minimum		152.183	55.826	88.628	1.991
Maximum		158.975	62.783	154.119	4.520

Fig. 3. Descriptive statisctis for energy demand consumption for each tenant

Skewness determines the symmetrical of the data distribution while kurtosis defines the shape of the distribution curves for graph distribution. Standard error of skewness and kurtosis for each tenant in the table has been multiplied with 3 and all the values are compared to the absolute value of skewness and kurtosis in order to know the normality of dataset of maximum demand consumption used to undergo analysis process. If the values of absolute skewness and kurtosis are lesser than the calculated standard error that has been multiplied earlier, which are known as score values, it shows that the data of hourly consumption of maximum demand used to undergo analysis using SPSS software is normal.

Figure 4 and 5 shows the skewness and kurtosis for all tenants. Bar graph of absolute values for skewness and kurtosis are lesser than bar graph of calculated score values. It can be concluded that all the data used for analysis is normal and can proceed to undergo prediction process.

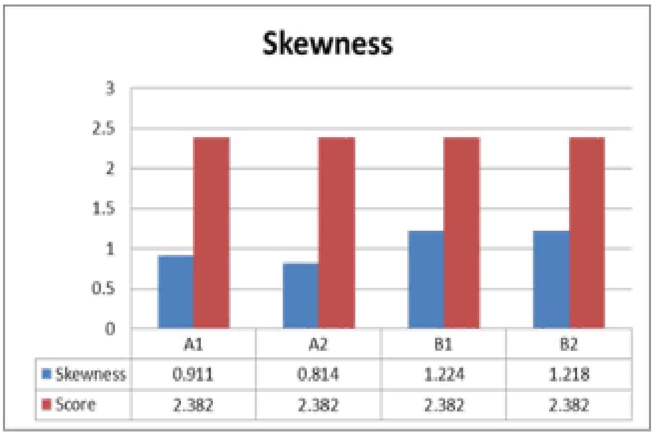

Fig. 4.Skewness graph for all tenants

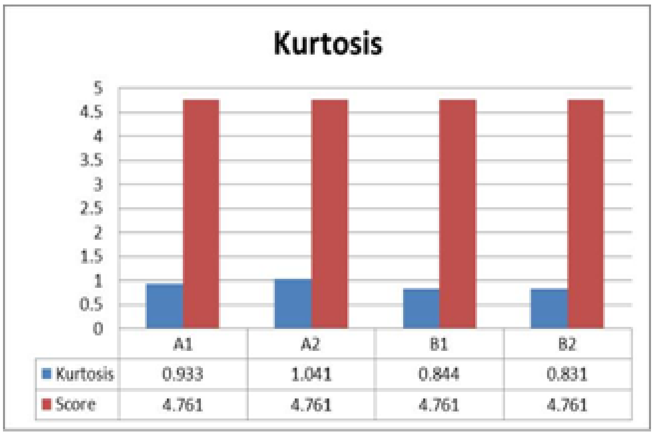

Fig. 5. Kurtosis graph for all tenants

For prediction process that used k-nearest neighbour (k-method, historical hourly consumption of maximum demand data for all tenants were computed into SPSS software with values of k = 3, 4 and 5. Training and test with the ratio of 70 to 30 was applied to avoid any overestimation or underestimation of the predicted data. Next, the most accurate value of k was selected by using RMSE calculation. Value of k that gave prediction value similar to the historical value of maximum

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8083 - 8090

demand data provides least value of RMSE. The comparison of average RMSE between values of k = 3, 4 and 5 for each tenant are shown in Figure 6. RMSE value for tenant A1, B1 and B2 have the lowest average value of RMSE for k = 5 with 5.73, 8.54 and 0.35. It can be concluded that k = 5 is the most accurate value used in this research compare to k = 3 and 4. The k value obtained in this process is used as a reference to predict maximum demand for upcoming period.

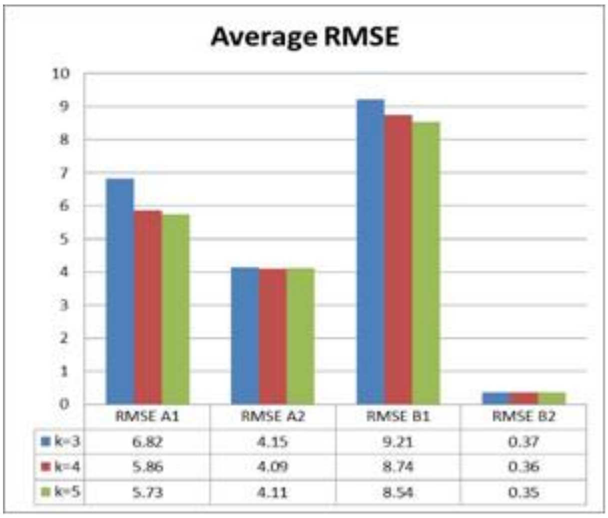

Fig. 6. Average RMSE value for different k tested on all tenants

Next, value of k = 5 is applied on the maximum demand value for each month of all tenants. Predicted values are affected by the average of power factor, single phase current and single L-N voltage. All these features are used to complete the prediction process with the help of SPSS software. Graphs in Figure 7 (a), 7 (b), 7 (c) and 7 (d) illustrate and compare the outline of line graph between historical and predicted maximum demand for tenant A1, A2, B1 and B2. From the graphs shown below, it can be concluded that predicted maximum demand are not much different compare to actual value.

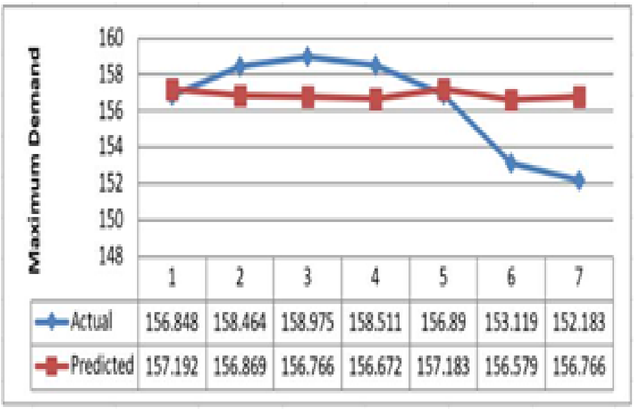

Fig. 7 (a). Comparison between actual and predicted maximum demand for tenant A1

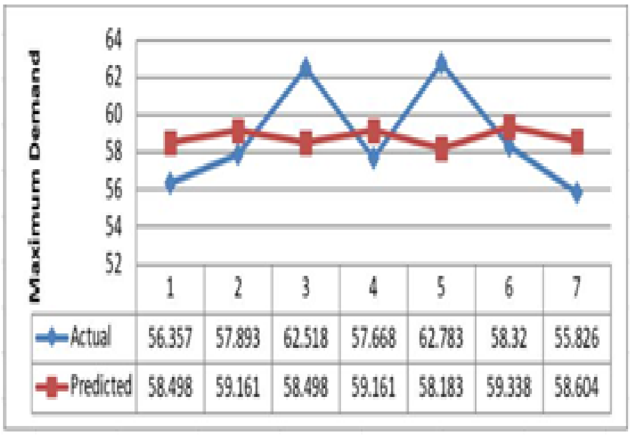

Fig. 7 (b). Comparison between actual and predicted maximum demand for tenant A2

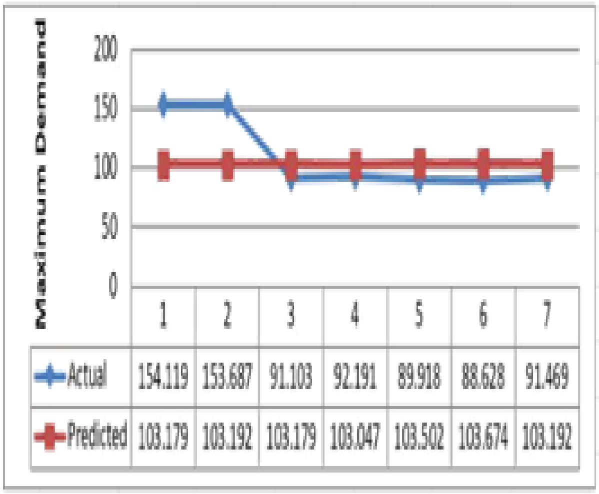

Fig. 7 (c). Comparison between actual and predicted maximum demand for tenant B1

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8083 - 8090

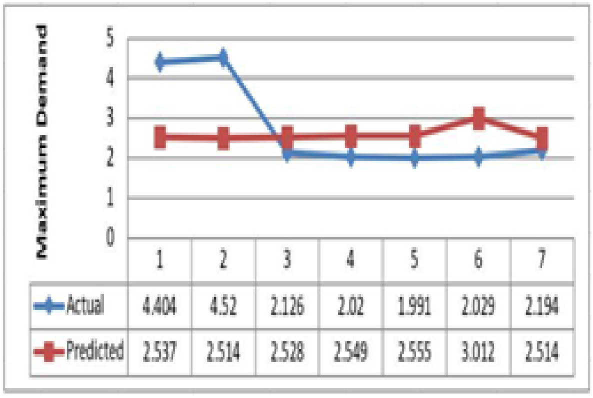

Fig. 7 (d). Comparison between actual and predicted maximum demand for tenant B2

Differences between prediction and actual values are calculated to see the difference between actual and predicted values when k-NN method with k = 5 is applied. Based on Figure 8 below, tenant A1, A2, B1 and B2 show slightest differences which are 0.43, 0.11, 5.45 and 0.15 respectively. Hence, based on the results above it can be concluded that the prediction method used for this research are able to make improvement in maximum demand values for each tenants.

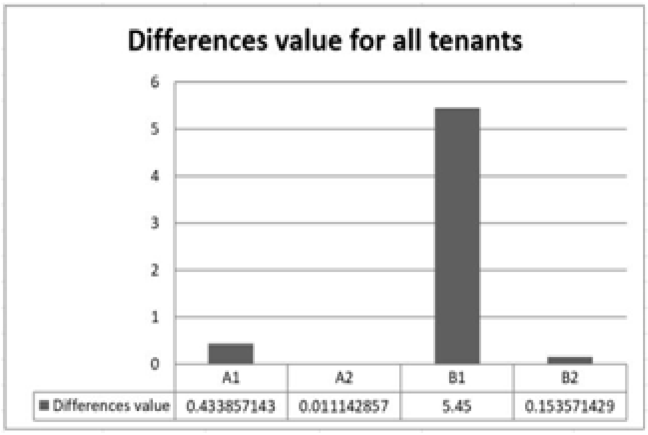

Fig. 8. Differences value for all tenants

V. CONCLUSION

In this paper, a research is conducted based on real data collection at the smart commercial building located in Klang Valley, Malaysia. Analysis on maximum demand consumption measured in kilowatt had been done concerning at major tenants of the building. Statistical method is proposed computed with several mathematical formulas consists of mean, variance, skewness and kurtosis in order to know the normality of collected dataset. Prediction in this research is completed by using k-nearest neighbour (k-NN) method with three different values of k. Predicted data undergo training and testing data in the ratio of 70 to 30 to avoid overestimating. Root mean square error (RMSE) is applied on the predicted data to determine the accuracy value of k. As a conclusion, value of k = 5 is the most accurate among other values due to the least value of RMSE calculated. Prediction on maximum demand for upcoming period also had been done by using k = 5 as a reference. To conclude, energy consumption analysis brings convenience towards user and energy management team to have a deep understanding on their energy behaviour. Prediction creates awareness and helps them to enhance their energy management system in order to reduce the energy bills. For future study, predicted data of maximum demand can be integrated into the available system for user monitoring and controlling purposes in achieving energy efficiency and sustainability.

VI. APPENDIX

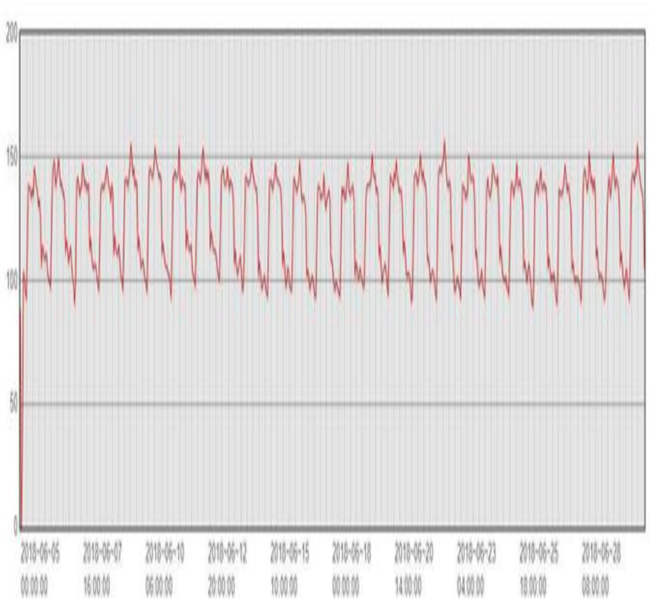

Fig. 9. Hourly consumption of maximum demand graph for Tenant A1 of June 2018

Published by: The Mattingley Publishing Co., Inc.

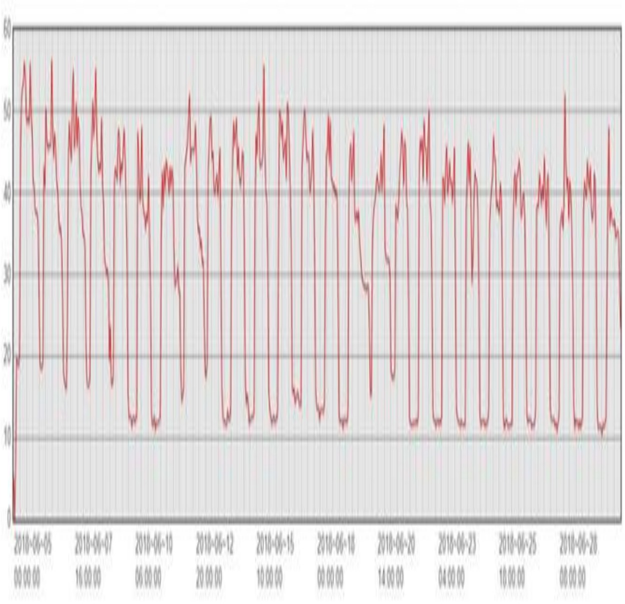

Fig. 10. Hourly consumption of maximum demand graph for Tenant A2 of June 2018

REFERENCES

[1] “Malaysia Population (LIVE).” Malaysia Population (2018) – Worldometers, 2018 [Online].

[2] J. S. Hassan, R. M. Zin, M. Z. A. Majid, S. Balubaid, and M. R. Hainin, “Building energy consumption in Malaysia: An overview,” J. Teknol., 2014.

[3] P. Tracy and P. Tracy, "What is a smart building and how can it benefit you?", RCR Wireless News, 2019 [Online]

[4] K. Akkaya, I. Guvenc, R. Aygun, N. Pala, and A. Kadri, “IoT-based occupancy monitoring techniques for energy-efficient smart buildings,” 2015 IEEE Wirel. Commun. Netw. Conf. Work. WCNCW, 2015

A. R. Al-Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, “A smart home energy management system using IoT and big data

[5] analytics approach,” IEEE Trans. Consum. Electron., 2017.

[6] V. V. Afanas’ev, V. G. Kovalev, V. V. Tarasova, and V. A. Tarasov,“The use of statistical methods to analyze electric-energy consumption by budget institutions,” Russ. Electr. Eng., 2017.

[7] M. N. .Salleh, A. Hariri, M. Z. .Yusof, A. . Leman, M. FarisMubin, and K. .Rahman, “Energy Consumption Analysis Based on Energy Efficiency Approach: A Case of Suburban Area,” 2016.

[8] F. Farahnakian, T. Pahikkala, P. Liljeberg, and J. Plosila, “Energy aware consolidation algorithm based on K-nearest neighbor regression for cloud data centers,” Proc. - 2013 IEEE/ACM 6th Int. Conf. Util. Cloud Comput. UCC 2013.

[9] F. Wahid and D. Kim, “A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings,” Int. J. Smart Home, 2016.

[10] "A Quick Introduction to K-Nearest Neighbors Algorithm", Noteworthy - The Journal Blog, 2019 [Online].

[11] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, “Building energy consumption prediction: An extreme deep learning approach,” Energies, 2017.

[12] R. Zhou, Y. Pan, Z. Huang, and Q. Wang, “Building Energy Use Prediction Using Time Series Analysis,” IEEE 6th Int. Conf. Serv. Comput. Appl, 2013.

[13] N. A. Ramli, M.F.A. Hamid, N. H. Azhan, and M. A. A. Ishak, “Solar Power Generation Prediction by using k- Nearest Neighbor Method,” unpublished.

Published by: The Mattingley Publishing Co., Inc.