ICCEPM 2022

The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

Energy Use Prediction Model in Digital Twin

Jihwan Wang[1]*, Chengquan Jin[2], Yeongchan Lee[3], Sanghoon Lee[4], Changtaek Hyun[5]

[1, 2, 3, 4, 5] *University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, Seoul, Republic of Korea,*
[1]E-mail address: jihwan1173@gmil.com, [2] E-mail address: yuchenghe22@naver.com,
[3]E-mail address: lyc56895@gmail.com, [4]E-mail address: sanghoon.lee@uos.ac.kr,
[5]E-mail address: cthyun@uos.ac.kr

Abstract: With the advent of the Fourth Industrial Revolution, the amount of energy used in buildings has been increasing due to changes in the energy use structure caused by the massive spread of information-oriented equipment, climate change and greenhouse gas emissions. For the efficient use of energy, it is necessary to have a plan that can predict and reduce the amount of energy use according to the type of energy source and the use of buildings. To address such issues, this study presents a model embedded in a digital twin that predicts energy use in buildings. The digital twin is a system that can support a solution of urban problems through the process of simulations and analyses based on the data collected via sensors in real-time. To develop the energy use prediction model, energy-related data such as actual room use, power use and gas use were collected. Factors that significantly affect energy use were identified through a correlation analysis and multiple regression analysis based on the collected data. The proof-of-concept prototype was developed with an exhibition facility for performance evaluation and validation. The test results confirm that the error rate of the energy consumption prediction model decreases, and the prediction performance improves as the data is accumulated by comparing the error rates of the model. The energy use prediction model thus predicts future energy use and supports formulating a systematic energy management plan in consideration of characteristics of building spaces such as the purpose and the occupancy time of each room. It is suggested to collect and analyze data from other facilities in the future to develop a general-purpose energy use prediction model.

Key words: Energy use prediction model, Digital twin, Correlation analysis, Multiple regression analysis

1. INTRODUCTION

Cities are responsible for about 60-80% of the world's energy use [1], of which buildings account for more than one-third of all energy use [2]. With the advent of the 4th Industrial Revolution, the energy efficiency of each component of a building has increased with the development of technology. However, the total energy use in the building sector is still on the rise as residential spaces and products used by residents become larger and hours of their use become longer [3]. Furthermore, energy use in buildings is expected to continue to increase due to economic development, global warming, urbanization in developing countries, and continuous growth [4]. In order to use energy efficiently, a plan is necessary to predict and manage energy use according to the types of energy sources and the purposes of buildings.

Meanwhile, the digitalization of each industry in the 4th Industrial Revolution has led to the rapid development of communication and information technology [5]. Digital twin technology is in the spotlight as one of the technologies to realize this development [6]. Digital twin is a system that can help to solve urban problems by visualizing the status of real-life objects on their identical counterparts in a virtual space through simulations and analyses using various sensing data collected in real-time [7]. By utilizing these advantages of the digital twin, real-time information such as facility information, energy use, and climate conditions can be easily obtained, thereby realizing efficient energy management through identifying energy use status and predicting future energy use. Therefore, this study aims to develop a digital twin-based energy use prediction model using a correlation analysis and multiple regression analysis for efficient energy management.

2. LITERATURE REVIEW

2.1. Digital Twin

The digital twin began to be introduced into the manufacturing industry in the 2000s when General Electric (GE), an American home appliance maker, spread the concept. It has been used in various fields such as aviation, construction, energy, and cities since then [9]. As the utilization of the digital twin becomes greater, studies to adapt the digital twin technology in the energy field have also been conducted. Some of the studies closely related to this research are summarized as follows. Kaewunruen(2018) evaluated the technical and the financial feasibilities of Net Zero Energy Buildings (NZEB) on existing buildings for the energy management of buildings, facilitated visualization and collaboration between stakeholders using the digital twin, and accurately estimated costs and related technical problems incurred when producing NZEB [10]. Shin (2020) simulated the installation direction and angle of solar panels in the digital twin of buildings in consideration of future demands for solar power generation, producing the amount of solar power generation and the amount of sunlight occurring in such places as the apartment balcony, rooftop and roof of buildings [11]. In addition, Ravi (2020) developed a new Urban Building Energy CPS (UBE-CPS) framework that connects the digital world to manage energy use of time-varying renewable energy sources such as solar heat and wind power [12]. Although there have been some efforts on utilizing the digital twin to optimize energy management and production as aforementioned studies did, few studies are related to predicting energy use for energy management, and thus systematic research is still required.

2.2. Energy use prediction

This study aims at developing an energy use prediction model based on the collected building components and environmental information. Thus, this section examines preceding studies related to energy use prediction.

For the purpose of reducing greenhouse gas emissions, Kim (2017) developed algorithms to analyze the correlation of seasonal cooling and heating energy use depending on climate conditions and their energy use patterns, and reduce energy use, by employing monthly energy use data from the building energy management system. In the study of Ku (2018), for the efficient utilization of the building energy management system (BEMS), an electric energy prediction model was developed using the Kriging method only with meteorological factors such as monthly temperature, humidity, and wind speed [13]. Kim (2019) analyzed the influencing factors on annual and seasonal energy use by integrating physical and regional characteristics in residential buildings, and the data used were based on the physical characteristics and monthly energy use of buildings in 2016. In addition, Kang (2020) used OpenStudio and EnergyPlus to derive the annual energy use of buildings and produced the basis data for effective greenhouse gas reductions [14].

While the studies predicted energy use using various methods, most of the factors used to predict energy use have employed only environmental information, such as monthly average values including temperature, humidity, and general building information. Moreover, changes in the predictive model as data accumulated were not considered. Thus, this study intends to present a digital twin-based energy use prediction model that can predict energy use by selecting major variables and real-time calculations in light of detailed building components and additional environmental information such as difference in preferred temperature by season (Spring, Autumn: 18~20˚C, Summer: 24~26˚C, Winter: 20~22˚C).

3. ENERGY USE PREDICTION MODEL

3.1. Model building methodology

This section describes the methodology to build an energy use prediction model using correlation analysis and multiple regression analysis based on past building performance data. Correlation analysis is a method used to explain the relationship between independent and dependent variables measured on an equal or ratio scale and to verify significance [15]. In this study, Pearson correlation coefficient, a correlation coefficient applied with the assumption that variables follow a normal distribution, is used. A pair of variables can be said to be more relevant when the correlation coefficient value is closer to -1 or +1.

For the correlation analysis, practical data that is easily available to users are used. The data can be grouped into the following three categories: 1) building components that include room area, room height, designed room capacity, electric heater (W/), referring to the building specifications; 2) environmental data provided by the Korea Meteorological Administration such as daily temperature difference, average wind speed, and average humidity; 3) difference in preferred temperature that calculates differences between daily outdoor temperature and temperature to maintain seasonal comfort. Since the statistical significance probability accepted in social science is usually 5% (0.05), the significance probability is usually determined by applying a 95% confidence interval, and in more conservative cases, the significance probability can be set to 1% [16]. In this study, a criterion of 1% significance probability is used to extract more explanatory variables. Among the variables, statistical variables with a significance probability of 0.01 or more are judged to be insignificant and removed, and only significant variables with a significance probability of 0.01 or less are used. In addition, if the independent variables have an excessively close relationship with each other, the regression analysis results can be distorted. Therefore, they must be removed in advance to prevent any statistical distortion.

In this study, in order to establish an energy use prediction model, multiple regression analysis is conducted based on the significant variables selected through the above correlation analysis. Multiple regression analysis is used in the case where there are two or more independent variables. Since the pure effect of individual independent variables on the dependent variable can be analyzed separately, it is not difficult to determine the contribution of each independent variable to the variation of the dependent variable. In this study, energy use is selected as the dependent variable, and the independent variables are selected by using the results of correlation analysis. In the results of the multiple regression analysis, it is possible to derive a regression equation in the following format through the value of the non-standardization coefficient B.

(1)

is the dependent variable to be predicted through the regression equation, is the Y-intercept, is the non-standardization coefficient, and is the independent variable. From the equation, the dependent variable (i.e., energy use) is calculated when values are entered into the Y-intercept, the non-standardization coefficient and each independent variable.

3.2. Automated energy use prediction model

In this section, an automated executable file (.exe) capable of automatically generating an energy use prediction model in the digital twin is developed. Microsoft Office Excel 2020 is used for correlation and regression analyses, and an executable to automate the process is developed in Python. The components and the process of the executable file are shown in Figure 1.

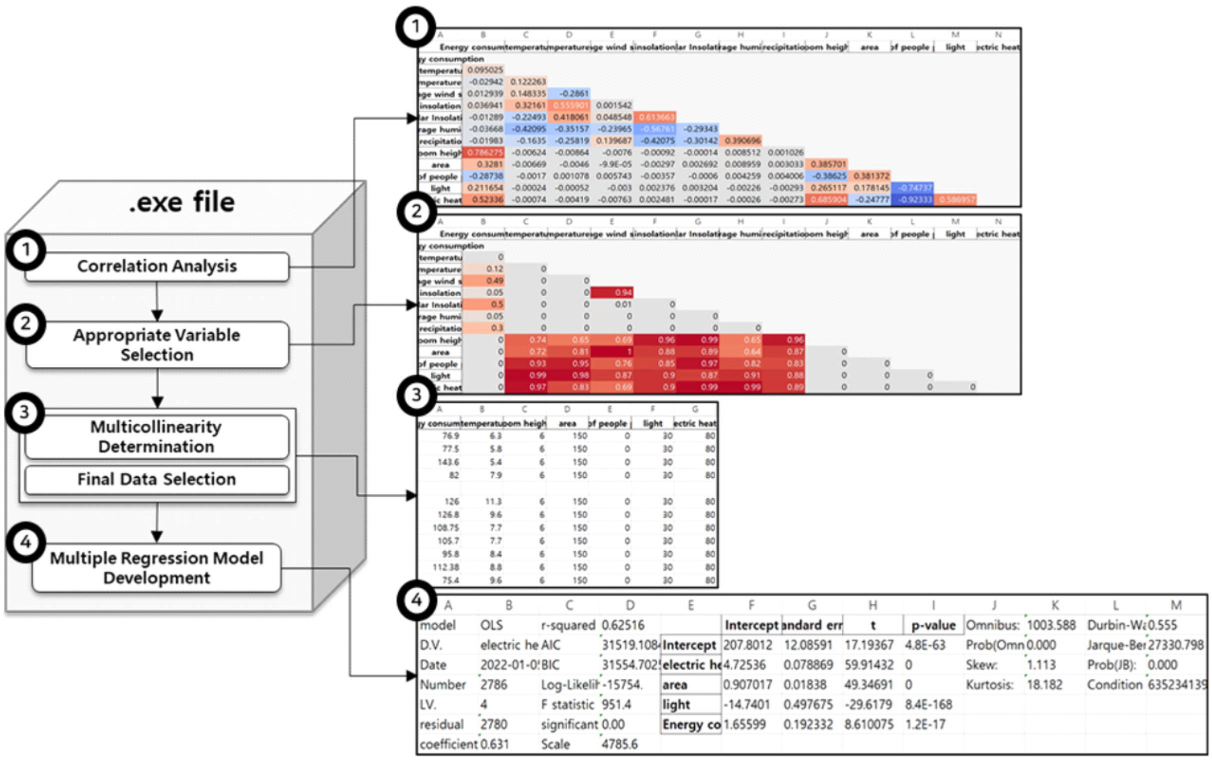

Figure 1. Procedures of automated energy use prediction

As shown in Figure 1, the step-by-step description is as follows:

① when the correlation analysis is performed with loaded data, correlation coefficients for each variable are displayed in the Excel sheet. Variable pairs with positive (+) correlation are displayed in red, and those with negative (-) correlation in blue. The higher the correlation coefficient, the higher the saturation of each color; ② after identifying the correlation coefficient, an appropriate variable for regression analysis is selected. In this study, in order to automatically remove insignificant variables, variables with a significance probability exceeding 0.01 are removed; ③ variables with high correlation with each other should be excluded because multicollinearity can cause errors in regression analysis results. Among the variables with a correlation coefficient of 0.9 or more, the variable with the highest correlation coefficient with energy use is selected. The final data is selected through the previous process; ④ when the final data for multiple regression analysis is selected, multiple regression analysis is performed using the data. The results of multiple regression analysis are generated on the last sheet. The results derived in a series of steps described above are generated in each sheet (sheet ①, ②, ③, and ④). The energy use prediction based on the derived regression model is described in Figure 3 of 4.2.

4. PREDICTION MODEL IN DIGITAL TWIN

4.1. Digital Twin-based prediction process

For the development of the digital twin, Unity, a game engine, is used in this study. Unity is a software development environment for the development of interactive applications with real-time graphic visualization functions. The digital twin-based energy use prediction process is shown in Figure 2 below.

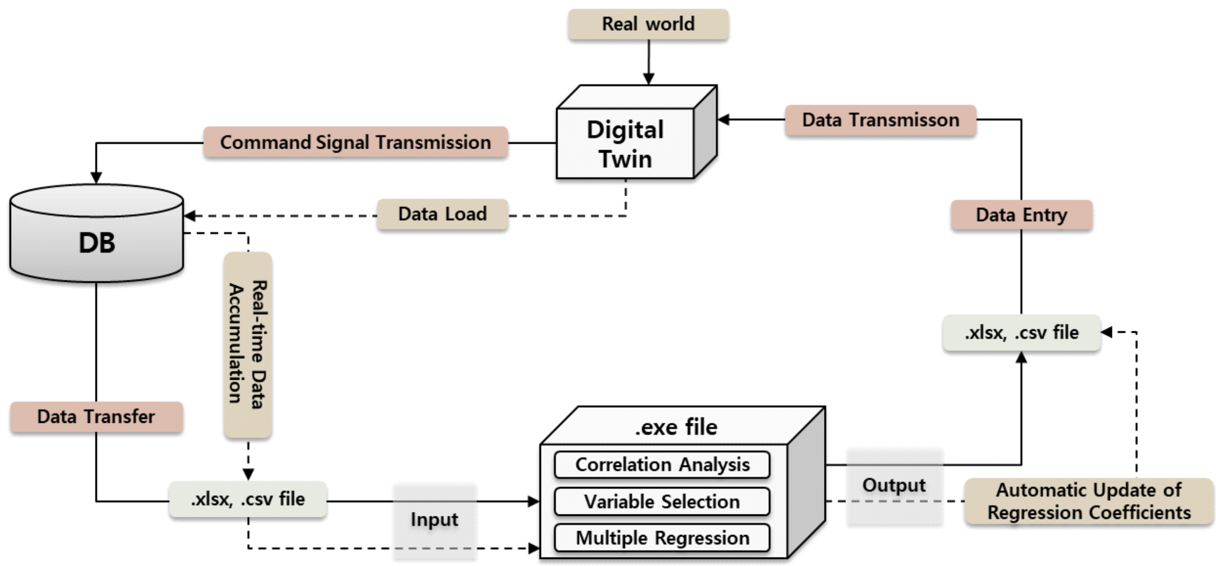

Figure 2. Digital twin-based energy use prediction process

The actual data of the building is stored in the digital twin DB, and the executable file uses the data in the format of .xlsx and .csv files in the DB through the transmission from the digital twin by the user's command. The regression models derived through execution are stored in the format of .xlsx and .csv. These are then, in turn, transmitted back to the digital twin. The predicted value of energy use for the present or future is displayed to the user in the digital twin. Actual data of energy use is also accumulated in real time in DB. As data is added to additional energy use prediction, the regression model is automatically updated through the executable file when predicting energy use.

4.2. Visualization in Unity

In this study, as shown in Figure 3, energy use is calculated and checked at the user interface according to the user's requirements within the digital twin. By pressing the ″RUN″ button at the bottom right of the left UI in the digital twin, the data variables used for correlation analysis and variables selected for multiple regression analysis are derived. The variables to be used for multiple regression analysis are displayed through significance probability verification and multicollinearity analysis by variable. In the results of multiple regression analysis, the analysis date is indicated in real time, and the name of the dependent variable and the number of independent variables used in the regression analysis are also displayed. The p-value for the F value is also displayed so that the user can verify whether the regression model derived through the executable file is appropriate. Moreover, by visualizing the regression model, the relative influence of each independent variable on energy use can be compared. In the UI on the right, the regression equation derived through the previous process is displayed, and the user can directly input the value of each independent variable to predict energy use. As building data accumulates, the regression model also varies flexibly, and accordingly, the items of independent variables that users can input vary.

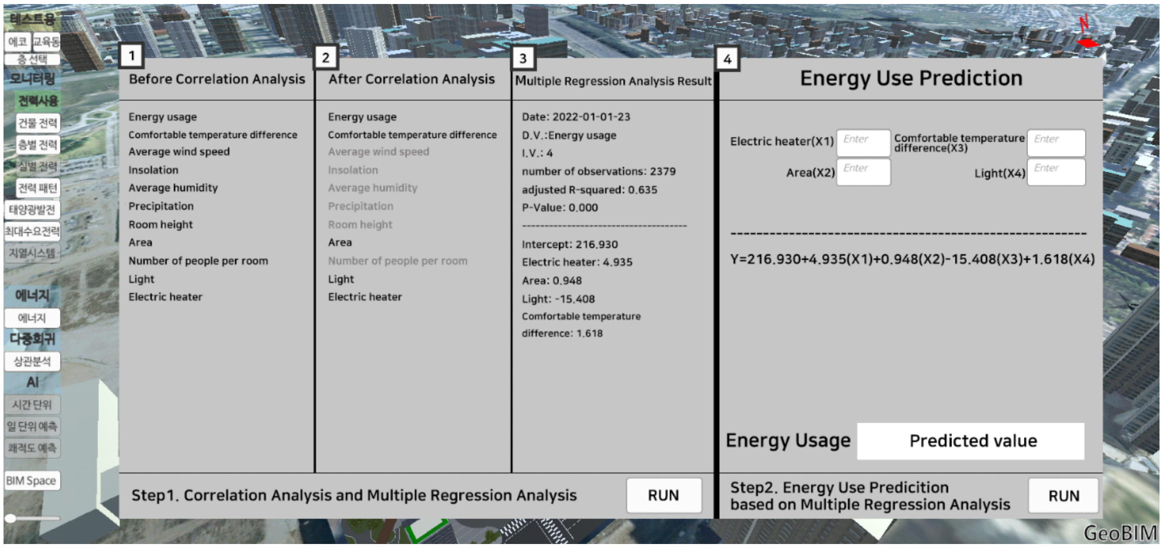

Figure 3. Visualization in Unity

4.3. Validation

In order to verify the accuracy of the energy use prediction model, the error rate between the actual value and the predicted value is compared based on the data of each room of the exhibition facility building. The initial energy use prediction model utilizes data collected from October 2020 to May 2021, and then the model is updated by additionally collecting data of each room of the building from June to September 2021. As the data accumulated in this way, error rate verification is performed with the two models (the first and the second models) to check the improvement of the accuracy of the prediction model. The cafeteria operated in the exhibition facility is arbitrarily selected as a verification target, and verification is conducted by way of predicting energy use from October to November 2021 and comparing them with actual uses.

In order to develop an energy use prediction model, ten items of data such as differences in preferred temperature, average wind speed, solar radiation, actual area, lighting, and electric heater are used. As a result of conducting a correlation analysis and identifying the significance probability and multicollinearity, five items are selected as variables for the first model: differences in preferred temperature, solar radiation, actual area, lighting, and electric heater. Four items are chosen as variables for the second model: differences in preferred temperature, actual area, lighting, and electric heater. Then, a multiple regression analysis is performed to derive a regression equation (Table. 1), and the p-value of the two models is 0.000, confirming that the regression model is statistically significant.

Table 1. Error rate according to actual electric energy use (Cafeteria)

Regression model (2020.10.~2021.5.)	2021.10.	2021.11.

	Actual Value	Predicted Value	Error rate	Actual Value	Predicted Value	Error rate
Y=174.770+4.662+0.886-14.312 +3.497-0.710 (Y= Energy use, =Heater, =Area, =Light, = comfortable temperature difference, =solar radiation)	1508.9829	815.6249	46%	881.6135	1397.2213	58%
Regression model (2020.10.~2021.9.)		Predicted Value	Error rate		Predicted Value	Error rate
Y=216.93+4.935+0.948-15.408 +1.618- (Y= Energy use, =Heater, =Area, =Light,= comfortable temperature difference)		1280.2542	15%		1103.6617	25%

As shown in Table 1, the composition of the independent variables of the first model and the second model modified through data accumulation is different. The solar radiation variable is removed as the data accumulated, and the regression model changes through the correlation analysis accordingly. The predicted error rate of the first model compared to the actual value is approximately 52% on average, and the error rate of the second model is about 20% on average. As a result, it can be found that as data accumulates, the error rate of the energy use prediction model decreases, and the prediction performance is improved.

5. CONCLUSION

Energy use of buildings increases due to changes in energy use structures with the spread of information devices, climate change, and greenhouse gas emissions as well as the advent of the 4th Industrial Revolution. Although the efficiency of each component of a building has increased, total energy use is still increasing. Meanwhile, digital twin technology has been in the spotlight, which helps to address urban problems through simulation and analysis based on data collected in real time. The need for energy management through digital twin-based energy use prediction is on the rise.

In this study, a digital twin-based energy use prediction model using a correlation analysis and multiple regression analysis is developed. Unity, a game engine, is used to implement the model in the digital twin. In addition, a series of processes such as correlation analysis and multiple regression analysis for predicting energy use in the digital twin are automated. Through the model, an efficient energy management plan can be established by identifying and intensively managing items that affect energy use.

The model is verified by checking the error rate of one room according to the accumulation of target data, which decreases as data accumulates. However, due to the small amount of analyzed data, it is necessary to improve the accuracy of the model by continuously collecting and analyzing data from various buildings in the future, and also to consider other data items that affect energy use such as building insulation in addition to the data items used in this study.

ACKNOWLEGEMENTS

The support of Korea Agency for Infrastructure Technology Advancement (No. 1615012359) is gratefully acknowledged.

REFERENCES

[1] www. ec.europa.eu/clima/policies/international/paris_protocol/cities_en. [Website], Retrieved from url, 2021.

[2] F. Ascione, N. Bianco, C. Stasico, G. Mauro, G. Vanoli, ″CASA, cost-optimal analysis by multi-objective optimization and artificial neural networks: A new framework for the robust assessment of cost-optimal energy retrofit, feasible for any building″, Energy and Buildings, pp. 200-219, 2017

[3] S. Lee, J. Lee, ″Practical Application of System for the Minimization of Energy Use: Building Sector″, Daejeon Development Institute, 2011.

[4] I. Cho, S. Lee, T. Kim, K. Hahm, ″A Renewable Energy Digital Twin Platform Architecture using 3D Authoring Tools″, Korean Society for Energy, p.114, 2020.

[5] V. Parida, D. Sjödin, W. Reim, ″Reviewing Literature on Digitalization, Business Model Innovation and Sustainable Industry: Past Achievements and Future Promises″, Sustainability, vol. 11, no. 2, pp.391-408, 2019.

[6] F. Tao, H. Zhang, A. Liu, ″Digital Twin in Industry: State-of-the-Art″, IEEE Transactions on Industrial Information, vol. 15, no. 4, pp.2405-2415, 2019.

[7] J. Jeong, Y. Chae, ″Optimization for Distributed Energy Resource of City-Scale based on Digital Twin″, Architectural Institute of Korea, vol. 39, no. 1, p.304, 2019.

[8] Y. Lee, S. Lee, C. Hyun, C. Jin, ″Development of Light-weight Smart City Digital Twin Model and System for Web Service″, Korean Journal of Computational Design and Engineering, vo1. 26, no. 3, pp.193-207, 2021.

[9] W. Kim, Y. Park, Y. Jeong, Y. Choi, S. Jeon, ″Trends of the 4th Industrial Revolution ′Digital Twin′ Technology and Plans for Military″, pp.132-143, 2021.

[10] S. Kaewunruen, P. Rungskunroch, J. Welsh, ″A Digital-Twin Evaluation of Net Zero Energy Building for Existing Buildings″, Sustainability, vol. 11, no. 1, pp.159-171, 2018.

[11] R. Srinivasan, R. Issa, ″Urban Building Energy CPS (UBE-CPS): Real-Time Demand Response Using Digital Twin″, Cyber-Physical Systems in the Built Environment, pp.309-322, 2020.

[12] H. Shin, S. Lee, ″Development of 3-dimensional Solar Photovoltaic Simulation System based on Digital Twin in the Smart City″, The transactions of The Korean Institute of Electrical Engineers, vol. 69, no. 12, pp.1834-1841, 2020.

[13] K. Ku, S. Jeong, ″Building electric energy prediction modeling for BEMS using easily obtainable weather factors with Kriging model and data mining, Building Simulation″, Building Simulation, vol. 11, no. 4, pp.739-751, 2018.

[14] J. Kang, Y. Kwak, H. Shin, S. Mun, J. Huh, ″Definition of Reference Buildings in Non-residential Buildings for Estimating Energy Consumption″, Proceeding of Annual Conference of the Architectural Institute of Korea, vol. 40, no. 1, pp.310-311, 2020.

[15] S. Song, Y. Jeon, ″Understanding and Application of Modern Basic Statistics″, Hakjisa, 2019.

[16] I. Min, ″Basic Statistics and Regression Analysis″, Jiphil Media, 2020.