Smart Media Journal / Vol.11,No.5 / ISSN:2287-1322
https://dx.doi.org/10.30693/SMJ.2022.11.5.38

A Comparative Study Between Linear Regression and Support Vector Regression Model Based on Environmental Factors of a Smart Bee Farm

A. B. M. Salman Rahman, MyeongBae Lee, Saravanakumar Venkatesan, JongHyun Lim, ChangSun Shin*

Abstract

Honey is one of the most significant ingredients in conventional food production in different regions of the world. Honey is commonly used as an ingredient in ethnic food. Beekeeping is performed in various locations as part of the local food culture and an occupation related to pollinator production. It is important to conduct beekeeping so that it generates food culture and helps regulate the regional environment in an integrated manner in preserving and improving local food culture. This study analyzes different types of environmental factors of a smart bee farm. The major goal of this study is to determine the best prediction model between the linear regression model (LM) and the support vector regression model (SVR) based on the environmental factors of a smart bee farm. The performance of prediction models is measured by R^2 value, root mean squared error (RMSE), and mean absolute error (MAE). From all analysis reports, the best prediction model is the support vector regression model (SVR) with a low coefficient of variation, and the R^2 values for Farm inside temperature, bee box inside temperature, and Farm inside humidity are 0.97, 0.96, and 0.44.

Keywords : Smart Farm | beekeeping | environmental factor | Farm inside temperature | bee-box inside temperature

I. INTRODUCTION

Historically, beekeeping is performed in different locations worldwide as a part of local food culture and an activity related to pollinator production. The primary beekeeping product is honey, which is the natural sweet substance produced by honey bees from the nectar of plants that the bees collect, transform, deposit, store, and leave in the honeycomb to mature and mature with specific substances of their own. In many countries around the world, beekeeping is so popular. For example, honey is an essential component of food culture in many countries like Japan, Korea, Bangladesh, India, and so on [1,2].

Smart farming is a growing way of controlling farms with IoT, robotics, drones, and AI to boost the quantity and efficiency of commodities while reducing the amount of human labor required by production[3]. The world's population is expected to reach 9.1 billion people by 2050. According to the United Nations' Food and Agriculture Organization (FAO), food availability would need to be boosted by roughly 70% to sustain this increasing population. According to the Environmental Change Research Program, climate change poses several hazards to crop output, animal health, and rural economies[4].

In South Korea, native beekeeping has been practiced since ancient times, and it is an

* This work was supported by a Research promotion program of SCNU

Manuscript : 2022. 05. 06
Revised : 2022. 06. 13
Confirmation of Publication : 2022.06.29
Corresponding Author : Changsun Shin e-mail : csshin@scnu.ac.kr

activity that connects forest communities and ecosystems. Beekeeping began in Korea at least 2000 years ago, during King Dongmyeonseng of the Kokuryo Kingdom[5,6]. In South Korea, there are two sorts of species. 1) Korean native bees and 2) Western honey bees[7]. During the Chosun dynasty (1392- 1910), beekeeping was practiced in roughly half of the Korean peninsula counties. Exotic bees already have a stronghold in South Korea. Western bees accounted for 83 percent of beehives in South Korea in 2002[8]. Beehive numbers in South Korea are estimated to be approximately 2,000,000, with honey production ranging between 20,000 and 27,000 tons recently[9].

This study aims to analyze different environmental parameters in a Smart Bee Farm and then find out the best prediction model among the linear regression model (LM) and the support vector regression model (SVR) based on different environmental factors.

II. RELATED WORK

The existence of an apiculture ecosystem depends on consistent access to high-quality bee harvest materials in order to create stable and profitable colonies. Many beekeepers, particularly those in the United States and Europe, relocate their larvae to gather fodder supplies after honey flows through public and private land [10].

A. B. M. Salman Rahman, and et al, has published a comparative study based on SVR for the change of strawberry production by the variation of nutrient water flow[11]. Spatio-temporal variations in plant flora and pollinator population dynamics significantly impact environmental connections between plants and pollinators. Plant-pollinator relationships across the community environment, blooming phenology of plant species, and population dynamics of essential pollinators should all be researched along relevant geographical and temporal gradients[12].

In statistics, linear regression is a linear approach for identifying the relationship between a scalar response and one or more explanatory variables [13,14].

Statistical learning consists of a collection of approaches to modeling techniques and understanding complex datasets. It is a newly developed field in statistics that integrates with similar developments in computer science as machine learning. The field includes other approaches, such as lasso and sparse regression, classification and regression trees, and the boosting and support of vector regression[15]. In statistics, support vector regression is supervised learning models with associated learning algorithms that analyze data [16]. Often, ecologists and environmental scientists need to track and predict how populations will react to environmental disruption, change over gradients of the landscape, or vary between different ecosystems. For several reasons, this mission is fraught with confusion[17].

III. METHODOLOGY

The process of creating, processing, and validating a model that may be used to generate future predictions using known outcomes is known as predictive modeling. Regression and neural networks are two of the most extensively used predictive modeling approaches. This study uses two different regression models to find the best prediction model for Smart Bee Farm. Two methods are 1) the Linear Regression Model and 2) the Support Vector Regression Model. The linear regression model (LM) and the support vector regression (SVR) model are the most widely used machine learning (ML) methods for predicting active compounds and molecular properties. This part describes these two regression models in detail step by step.

Linear Regression:

Linear regression is a linear method in Statistics to model the relationship between a scalar dependent variable Y and one or more explanatory variables denoted as X. The case of one explaining variable is called a simple linear regression, and the procedure is called multiple linear regression for more than one explanatory variable [18,19].

We assume a linear regression multivariate model,

$$Y_i = \beta_0 + \beta_1 X_i + \ldots . + e_i \tag{1}$$

Here, Y_i is the dependent variable, X_i is the explanatory variable, β_0 and β_1 are two unknown constants that represent the intercept and slope, also known as coefficients or parameters, and e_i is the error term.

Support Vector Regression:

Support vector machine (SVMs, also support vector networks) is a supervised learning model in machine learning with associated learning algorithms that analyze data. Support Vector Machine can be used as a regression method, retaining all the key characteristics (maximum margin) that define the algorithm[20,21]. For classification, the Support Vector Regression (SVR) follows the same concepts as the SVM, with just a few slight variations[22]. In the case of regression, a margin of tolerance (epsilon) is about in approximation to the SVM, which might have already been requested from the matter, but besides this fact, there is also an additional sophisticated reason, the formula is more complicated therefore to be taken in consideration.

Equation of Support vector regression

$$f(x) = \sum_{i=1}^{m} (a_i^* - a_i) k(x_i, x) + b \tag{2}$$

Here, the value of $f(x)$ is equal to the inner product of two vectors xi and x_j in the feature space $\phi(xi)$ and $\phi(x_j)$, that is, $K(x_i,x_j) = \phi(x_i) \cdot \phi(x_j)$. All necessary computations can be done directly in the input space without having to compute the map ϕ(x), by the use of kernels. Some of the popular kernel functions are the linear kernel $K(x_i,x_j) = x_i \cdot x_j$, polynomial kernel $K(x_i,x_j) = (x_i \cdot x_j + 1)^d$.

IV. DATA DESCRIPTION

Our studies used smart bee farm environmental data for every minute from Marh2020 to April 2020, and the data sets were collected from an ecological smart bee farm at Gwangyang, a town in South Jeolla Province, South Korea. The data set consists of a Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity. In this study, the data sets have a total of 53824 entries with three variables. Our study aims to find out the best prediction model between the linear regression model (LM) and the support vector regression model (SVR) based on the environmental factors of a smart bee farm.

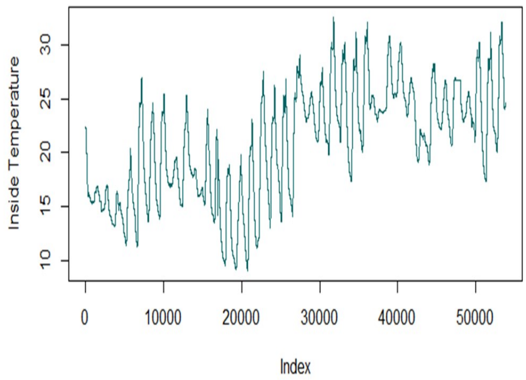

Fig. 1. Farm Inside Temperature for Every Second

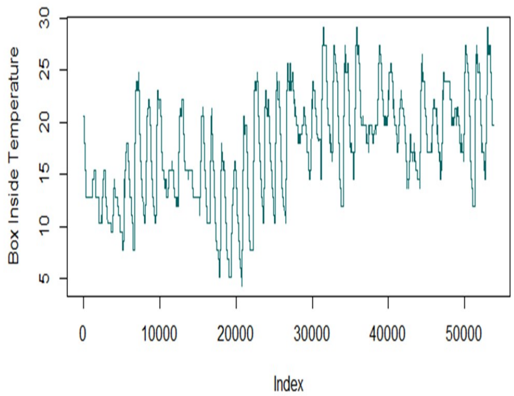

Fig. 2. Bee_Box Inside Temperature for Every Second

Figures 1&2 show Farm Inside and Bee_Box inside temperature plot by original data for every second. Every plot presents the temperature changes in a Smart Bee Farm for every second. The figure X-axis shows the number of seconds, and the Y-axis shows the temperature in Celsius. Figure 3 shows the Farm inside humidity plot by original data for every second. From the plot, we can easily find out the humidity changes in Smart Bee Farm every second. The X-axis shows the number of seconds in the figure, and the Y-axis shows the humidity.

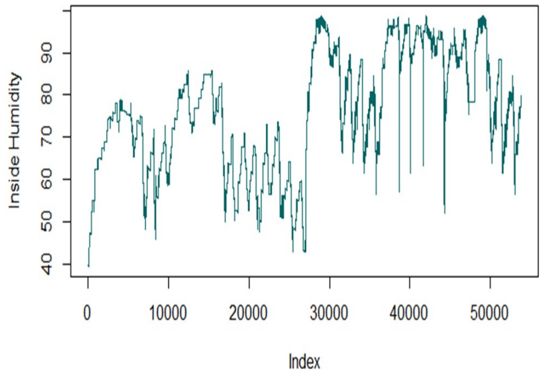

Fig. 3. Farm Inside Humidity for Every Second.

Figure 4 shows the relationship among three variables for the Smart Bee data set. This diagram shows how the directional histogram plots the bivariate scatter plots around the directional and the spearman connection above. This is the estimate of the two variables' monotonic relations. The correlation value of 1 is the positive overall correlation, −1 is the overall negative correlation, and 0 does not reflect a correlation between variables. For each pair, the linear regression redline is seen.

The figure shows the positive correlation among Farm inside Temperature & Bee_Box inside Temperature is (.95), Farm inside Temperature & Farm inside Humidity is (.27), and Bee_Box inside Temperature & Farm inside Humidity is (.09). From the figure, we can easily find out that Farm inside Temperature is highly correlated with Bee_Box inside Temperature.

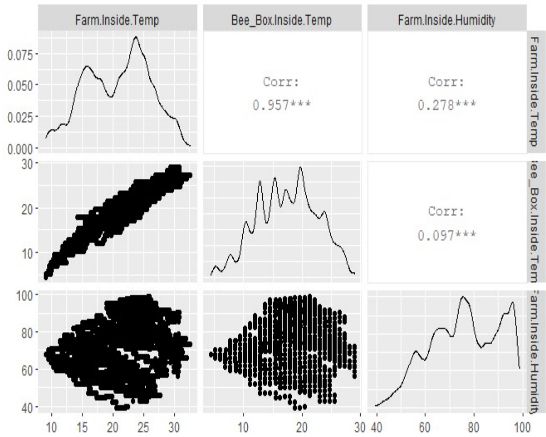

Fig. 4. The relation among three environmental factors.

V. PERFORMANCE EVALUATION INDICES

We examine R^2 values, mean absolute error (MAE), and root mean square error (RMSE) values to get the best prediction outcomes.

Evaluating criteria are used to calculate the performance of prediction models. The Root Mean Squared Error (RMSE) is used to find the square error relative to actual values, as well as the square root of the summation factor and the prediction's square error. The Root Mean square error (RMSE) is a level−dependent variable made up of values from the same measurement units.

R^2 values between 0.0 and 1.0, and larger values, indicate a better bargain [23]. So, until the R^2 value is almost equal to one, it is unlikely that the prediction results will be accurate enough.

The equation of R− square value is,

$$R^2 = 1 - \frac{AA_{regression}}{AA_{Total}} \quad (3)$$

In this formula, $AA_{regression}$ is the variance of the total squared regression, and AA_{Total} is the cumulative squared error number.

The average model prediction error is expressed by both MAE and RMSE in units of the interest variable. The two metrics will vary from 0 to ∞ and are oblivious to the error direction. They are negatively driven reviews, which suggests that lower values are good[24].

The equation for RMSE is,

$$RMSE = \sqrt{\Sigma \frac{(Y_{Pred} - Y_{act})^2}{N}} \quad (4)$$

Here, Y_{pred} symbolizes the predicted value, Y_{act} symbolizes the real value, and N symbolizes the sample size.

The equation for MAE is,

$$MAE = \frac{\Sigma_{i=1}^{n} |Y_{pred} - Y_{act}|}{N} \quad (5)$$

Here, Y_{pred} symbolizes the predicted value, Y_{act} symbolizes the real value, and N represents the total number.

VI. RESULTS AND DISCUSSION

In the result part, we analyze different environmental factors like Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity using the linear regression model and the support vector regression model. Here, we compare two regression models to find out the best prediction model for Smart Bee Farm. Nowadays, the linear regression model (LM) and the support vector regression (SVR) model are the most widely used machine learning (ML) methods for prediction. From this part, we can easily find out the best prediction model for this study.

Figures 5, 6, and 7 show the actual curve of Farm Inside Temperature, Bee Box Inside Temperature, and Farm Inside Humidity with Linear Regression Model (LM) fitted line.

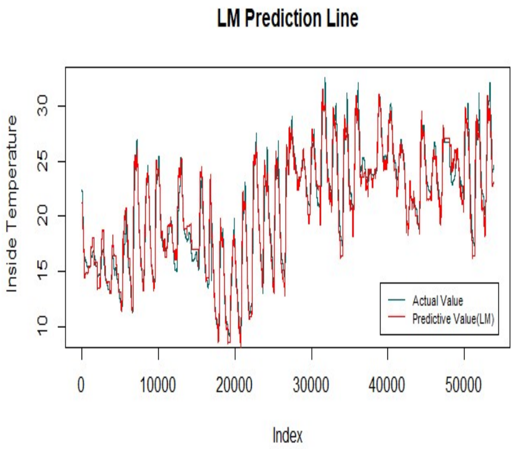

Fig. 5. Farm Inside Temperature for Every Second with Linear Regression Prediction Line.

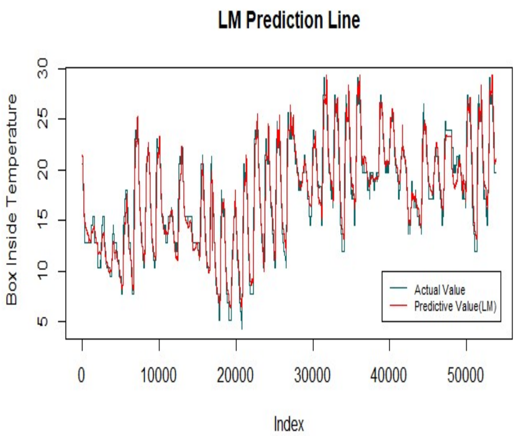

Fig. 6. Bee_Box Inside Temperature for Every Second with Linear Regression Prediction Line.

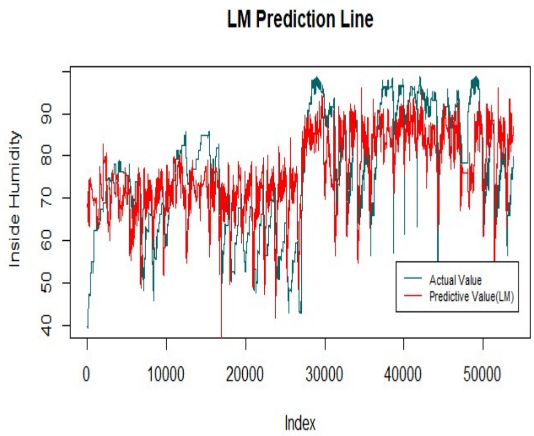

Fig. 7. Inside Humidity for Every Second with Linear Regression Prediction Line.

In figures 5, 6, and 7 green line shows the actual values of Farm Inside Temperature, Bee_Box inside temperature, and Farm Inside Humidity. On the other hand, the red line shows the linear regression model (LM) predictive results of Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity. From the figure, we can say that the prediction lines are very close to the actual line. Figures 8, 9, and 10 show the actual curve of Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity with the Support Vector Regression Model (SVR).

In figures 8, 9, and 10, the green line shows the actual values of Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity. On the other hand, the red line shows the support vector regression (SVR) predictive results of Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity. From the figure, we can easily find out that the prediction lines are very close to the actual line.

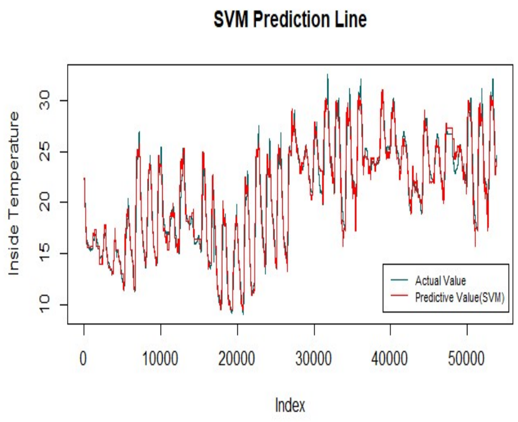

Fig. 8. Farm Inside Temperature for Every Second with Support Vector Regression Prediction Line.

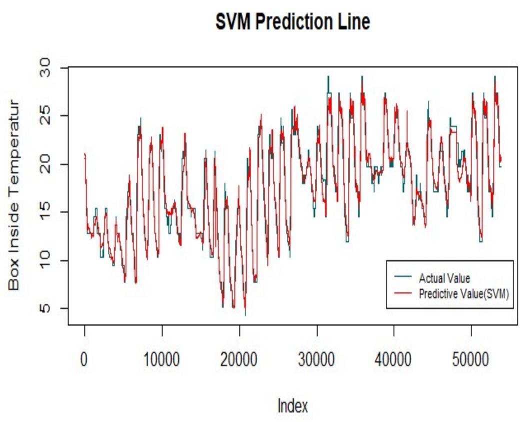

Fig. 9. Bee_Box Inside Temperature for Every Second with Support Vector Regression Prediction Line.

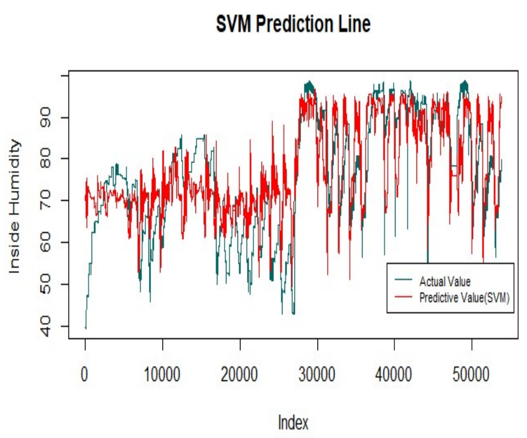

Fig. 10. Farm Inside Humidity for Every Second with Support Vector Regression Prediction Line.

We can easily find out the best prediction model between the linear regression model (LM) and the support vector regression model (SVR) in table 1. To find out the best prediction results, we consider R^2 value, root mean squared error (RMSE), and mean absolute error (MAE). The ranges of R^2 values from 0.0 to 1.0 and higher values indicate a better deal.

Table1: Models of Performance analysis results for Linear Regression Model and Support Vector Regression Model

Smart Bee Farm	Linear Regression Model			Support Vector Regression Model		
	R-Square Value	RMSE	MAE	R-Square Value	RMSE	MAE
Farm Inside Temperature	0.96	1.01	0.77	0.97	0.84	0.59
Bee_Box Inside Temperature	0.95	1.14	0.87	0.96	0.88	0.64
Farm Inside Humidity	0.41	8.16	6.69	0.44	7.05	9.77

Table 1 shows the R^2 value, RMSE, MAE for Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity for smart bee farm using the linear regression model (LM) and support vector regression model (SVR). From the table, we can easily find out the best prediction model between the two. After analyzing the Farm Inside Temperature, Bee_Box Inside Temperature, and Farm Inside Humidity of Smart Bee Farm, we find out the best prediction model. The best prediction model is the support vector regression model (SVR) based on R^2 value 0.97 for Farm Inside Temperature, 0.96 for Bee_Box Inside Temperature, and 0.44 for Inside Humidity.

Ⅶ. CONCLUSION

The bee lives in a variety of natural and man-made environments. It is necessary as a pollinator for agricultural crops, and the beekeeping sector produces honey, bee venom, pollen, and wax as a source of income. In conclusion, our study analyzed and identified the best prediction model among the linear regression models (LM) and the support vector regression model (SVR). All results and analyses give us acuteness between two models, and we find out the support vector regression model (SVR) gives us the best prediction results.

REFERENCES

[1] A. B. M. Salman Rahman, Meyongbae Lee, Jonghyun Lim, Yongyun Cho, Changsun Shin, "Identifying Best Fit Ecological Smart Bee Farm Based on Environmental Issues by Linear Regression Model," *Journal of Green Engineering*, Vol. 11, No. 2, pp. 1951-1963, Feb. 2021.

[2] Hae-KyungChung, Hye JeongYang, DayeonShin, Kyung RhanChung, "Aesthetics of Korean foods: The symbol of Korean culture," *Journal of Ethnic Foods*, Vol. 3, No. 3, pp. 178-188, Sep. 2016.

[3] Dieisson Pivoto, Paulo Dabdab Waquilb, EdsonTalamini, Caroline Pauletto Spanhol Finocchio, Vitor Francisco Dalla Corte, Gianade Vargas Mores, "Scientific Development Of Smart Farming Technologies And Their Application In Brazil," *Information Processing In Agriculture*, Vol. 5, pp. 21-32, 2018.

[4] Mobasshir Mhabub, "A Smart Farming Concept Based On Smart Embedded Electronics, Internet Of Things And Wireless Sensor Network," *Internet Of Things*, Vol. 9, Mar. 2020.

[5] M.S. Park, Y.C. Youn, "Traditional knowledge of Korean native beekeeping and sustainable forest management," *Forest Policy Economics*, Vol. 15, pp. 37-45, Feb. 2012.

[6] A. B. M. Salman Rahman, Meyongbae Lee, Jonghyun Lim, Yongyun Cho, Changsun Shin, "Systematic Analysis of Environmental Issues on Ecological Smart Bee Farm by Linear Regression Model," *International Journal of Hybrid Information Technology*, Vol. 14, No. 1, pp. 61-68, 2021.

[7] Ryo Kohsaka, Mi Sun Park, Yuta Uchiyama, "Beekeeping and honey production in Japan and South Korea: past and present," *Journal of Ethnic Foods*, Vol. 4, No. 2, pp. 72-79, Jun. 2017.

[8] D.H. Oh, S.H. Choi, "Farming status and honey quality of Korean native oriental honeybee (Apis cerana Fab.) in Jiri mountain area," *Korean J Apic*, Vol. 19, No.1, pp. 11-16, 2004.

[9] Chuleui Jung, Sang-kyun Cho, "Relationship Between Honeybee Population and Honey Production in Korea: A Historical Trend Analysis," *Journal of Apiculture*, vol. 30, pp. 7-12, 22 Apr. 2015.

[10] Patel, Vidushi, Eloise Biggs, Natasha Pauli, Bryan Boruff, "Using a social-ecological system approach to enhance understanding of structural interconnectivities within the beekeeping industry for sustainable decision making," *Ecology and Society*, Vol. 25, no. 2, 2020.

[11] A. B. M. Salman Rahman, Meyongbae Lee, Vasanth Ragu, Yongyun Cho, Jangwoo Park, Changsun Shin, "A Comparative Study Based on SVR for the changes Strawberry Productions by the Variation of Nutrient Water Flow," *Journal of Knowledge Information Technology and Systems(JKITS)*, Vol. 14, No. 3, pp. 291~303, Jun. 2019.

[12] Naoki Inari, Tsutomu Hiura, Masanori J. Toda, Gaku Kudo, "Pollination linkage between canopy flowering, bumble bee abundance and seed production of understorey plants in a cool temperate forest," *Journal of Ecology*, Vol. 100, pp. 1534-1543, 2012.

[13] Linear Regression, https://en.wikipedia.org/wiki/Linear_regression (accessed on Feb., 19, 2022).

[14] A.B.M.Salman Rahman, Vasanth Ragu, Myeongbae Lee, Jangwoo Park, Yongyun Cho, Meonghun Lee, Changsun Shin, "An Analysis Study Based on Linear Regression Model for Changes of Fruit Size over Plum Diseases," *Journal of Knowledge Information Technology and Systems*, Vol. 12, No. 5, pp. 509-519, 2018.

[15] G. James, D. Witten, T. Hastie, R.Tibshirani, *An introduction to statistical learning with applications in R*, Springer New York Heidelberg Dordrecht London, 2013.

[16] Support-vector machine, https://en.wikipedia.org/wiki/Support-vector_machine (accessed on Jan., 12 2022).

[17] Neal M. Williams, Elizabeth E. Crone, T'ai H. Roulston, Robert L. Minckley, Laurence Packer, Simon G. Potts, "Ecological and life-history traits predict bee species responses to environmental disturbances," *Biological Conservation*, Vol. 143, pp. 2280-2291, 24 Apr. 2010.

[18] A. B. M. Salman Rahman, Lee Myeongbae, Lim Jonghyun, Yongyun Cho, Shin Changsun, "A Comparative Study of Energy Big Data Analysis for Product Management in a Smart Factory," *Journal of Organizational and End User Computing*, Vol. 34, No. 2, pp. 1-17, 2022.

[19] Mu Moung Cho Han, Yangsok Kim, Choong Kwon Lee, "Analysis of the Relations between Social Issues and Prices Using Text Mining - Avian Influenza and Egg Prices -," *Smart Media Journal*, Vol. 7, No. 1, pp.45-51, 2018.

[20] A. B. M. Salman Rahman, Meyongbae Lee, Jonghyun Lim, Yongyun Cho, Changsun Shin, "A PREDICTION MODEL FOR STEEL FACTORY MANUFACTURING PRODUCT BASED ON ENERGY CONSUMPTION USING DATA MINING TECHNIQUE," *Journal of Science and Engineering Management*, Vol. 1, No. 2, pp. 9-16, Dec. 2020.

[21] Hyoung Ju Kim, Moon Jong Min, Pan Koo Kim, "An analysis study on the quality of article to improve the performance of hate comments discrimination," *Smart Media Journal*, Vol. 10, No. 4, pp. 71-79, 2021.

[22] Yongju Choi, Jeeyoung Oh, Daihee Park, Yongwha Chung, Hee-Young Kim, "Replacement Condition Detection of Railway Point Machines Using Data Cube and SVM," *Smart Media Journal*, Vol. 6, No. 2, pp. 33-41, 2017.

[23] David R. Legates, "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation," *water resources research*, Vol. 35, No. 1, pp. 233-241, Jan. 1999.

[24] A. B. M. Salman Rahman, Meyongbae Lee, Jonghyun Lim, Yongyun Cho, Changsun Shin, "Modeling the Smart Factory Manufacturing Products Characteristics: The Perspective of Energy Consumption," *Discrete Dynamics in Nature and Society*, Vol. 2021, pp. 1-15, Dec. 2021.

Authors

A. B. M. Salman Rahman

He completed his Bachelor degree in ETE from Southeast University, Dhaka, Bangladesh. He received his Integrated Masters with Ph.D. in Information and Communication Engineering from Sunchon National University, Suncheon, South Korea. Currently, work as an Academic Research Professor in the Department of Information and Communication Engineering at Sunchon National University, Suncheon, South Korea. His area of interest includes Forecast Model, Ubiquitous Computing, and Big Data Processing.

MyeongBae Lee

He received his Bachelor degree in Computer Engineering from Korea. He received Master degree on Computer Science in South Korea. And currently pursuing Doctorate degree in Information and Communication Engineering. He is working as a Researcher in the Department of Information and Communication Engineering at Sunchon National University Suncheon, South Korea. His area of interest includes Advanced Agriculture Technology, IT Convergence, Cloud and Ubiquitous Computing.

Saravanakumar Venkatesan

He received his Bachelor degree in Mathematics from Madras University and a Master of Science in Information and Communication Engineering at Sunchon National University in South Korea. Currently pursuing Ph.D. in the Department of Information and Communication Engineering, Sunchon National University. His current research interests include Big Data Analytics, Data Mining.

JongHyun Lim

He Completed Bachelor degree in Information and Communication Engineering in Korea. He received a Master degree in Information and Communication Engineering at Sunchun University in South Korea. He currently studying for a Doctorate degree in Information and Communication Engineering at Sunchon University. His area of interest includes System Software, Ubiquitou.

ChangSun Shin

Received the PhD degree in Computer Engineering at Wonkwang University. Currently, he is a professor of the Dept. of Information &Communication Engineering in Sunchon National University. His researching interests include Distributed Computing, IoT, Machine Learning, and Agriculture/ICT Convergence.