Application of Stacking Ensemble Machine in Big Data: Analyze the Determinants for Vitalization of the Multicultural Support Center

Raeho Lee[1]
Department of Korean Language & Literature
Kangwon National University
Chuncheon, 24341, Republic of Korea

Haewon Byeon[2]*
Department of Digital Anti-aging Healthcare (BK-21)
Inje University
Gimhae 50834, Republic of Korea

***Abstract*—For multicultural families to successfully promote social adaptation and achieve desirable social integration, the role of the multicultural family support center (Multi-FSC) is crucial. In addition, it's important to examine the factors that will contribute to the multicultural support center's vitality from the standpoint of the customers. In this study, machine learning models based on a single machine learning model and stacking ensemble using survey data from all multicultural families were used to examine the determinants for the utilization of multicultural family support centers for multicultural families. Additionally, based on the constructed prediction model, this study offered the foundational data for the revitalization of the multicultural support center. In this study, 281,606 adults (19 years or older), 56,273 of whom were married immigrants or naturalized citizens as of 2012, were examined. The stacking ensemble method was employed in this work to forecast the use of multicultural family support centers. In the base stage (model) of this model, logistic regression was employed, along with Classification and Regression Tree (CART), Radial Basis Function Neural Network (RBF-NN), and Random Forest (RF) model. The RBF-NN-Logit reg model had the best prediction performance, according to the study's findings (RMSE = 0.20, Ev = 0.45, and IA = 0.68). The findings of this study suggested that the prediction performance of the stacking ensemble can be improved when creating classification or prediction models using epidemiological data from a community.**

Keywords*—Stacking ensemble machine; radial basis function neural network; random forest; multicultural family support centers; prediction model

I. Introduction

In South Korea, the number of married immigrants and foreign workers has increased rapidly since the 1990s. Consequently, South Korea has also transformed into a multi-cultural society starting in the 21st century. Multicultural Demographic Statistics [1] reported that the number of foreigners residing in South Korea was 1.13 million as of 2011, exceeding 2.3% of the resident registration population in South Korea. Furthermore, multicultural marriages increased by 4.0% (24,721 cases), and the proportion of multicultural marriages in total marriages in South Korea was 10.3%, up 1.1% from 2019 [1]. If current trends continue, the number of married immigrant families will exceed 2 million, accounting for 5% of the total population, by 2050 [2]. In addition, the Ministry of Education, Science and Technology [3] estimated that one out of five citizens in their 20s will be from multicultural families and one out of three newborns will be born in multicultural families in 2030. South Korean society must prepare for measures to cope with various issues, which may arise during the paradigm shift to multiculturalism (e.g., early adaptation of marriage immigrants to South Korean society), as well as these rapid changes in the demographic structure.

*Corresponding Author.

As the number of multicultural families has increased, support for multicultural families has also been diversifying over the past 20 years [4]. Currently, the South Korean government has established and operated multicultural family support centers nationwide as a part of national policy for stably supporting a multicultural society [5]. The multicultural family support center (Multi-FSC) is a central management that has been operated since 2006 for the purpose of helping marriage immigrants to adapt to South Korean society quickly and supporting for stable family life of multicultural families through providing services including family education, counseling, and cultural programs for multicultural families. Twenty-one centers were established in 2006, and the number of centers increased to 201 by 2012, following the passage of the Multicultural Families Support Act (No. 8937) in 2008 [6,7].

In one aspect, as the number of multicultural family support centers was increasing, the necessity of careful management for participants was emphasized in addition to the effectiveness of services [8]. Nevertheless, studies on the Multi-FSC have focused on the results of one region or just analyzed the achievement or satisfaction of multicultural support projects [9, 10, 11, 12].

The role of Multi-FSC is very important for the successful social adaptation support and desirable social integration of multicultural families. In addition, it's important to examine the factors that will contribute to the multicultural support center's vitality from the standpoint of the customers. This study explored the predictors for the use of Multi-FSCs for multicultural families using machine learning models based on a single machine learning model and stacking ensemble using the survey data of all multicultural families. Furthermore, based on the developed prediction model, this study provided foundational data for the revitalization of the multicultural support center.

II. Materials and Methods

A. Source Data

This study analyzed survey data from a national survey of multicultural families (NS-mulfam) that the Ministries of Health, Welfare and Family Affairs, Justice, and Gender Equality jointly conducted in 2017 on multicultural families living in South Korea. The purpose of the NS-nulfam was to create tailored policies by identifying the living circumstances and welfare requirements of multicultural families [13]. The general characteristics of multicultural families, employment, economic status, marriage, health, and medical care were among the survey's questions. The NS-nulfam targeted 154,333 married immigrants for its survey, all of whom were living in South Korea. A survey on married immigrants, a survey on their spouses, and a survey on their kids were all independently conducted as part of this study. The survey subjects were sampled by using the status of foreign residents in 16 cities and provinces in South Korea and the basic status data of multicultural families organized by the Ministry of Public Administration and Security. The survey was carried out from July 20 to October 31, 2017, and a separate sample design was not taken into consideration because this study used a complete enumeration. According to the Multicultural Families Support Act (No. 8937), the multicultural families used in this research were defined as either 1) families made up of marriage immigrants and South Korean citizens who became citizens by birth, acknowledgment, or naturalization, or 2) families made up of foreigners who became citizens of South Korea by acknowledgment or naturalization and South Korean citizens who became citizens by birth, acknowledgment, or naturalization. After removing 1,618 individuals who were under the age of 19 from the study, 281,606 adults (56,273 men and 225,333 females) who were married immigrants or naturalized as of 2012 were examined.

B. Definition of Variables

The outcome (label) was defined as the use experience of a Multi-FSC (not aware of or never used it OR used it before). Features included gender, marital status (single, having a spouse, bereavement / widowed, or divorcement / divorced / separation / separated), highest education level (elementary school or below, junior high school, high school, or college or higher), occupation (regular worker, temporary worker, day-to-day worker, self-employed with employees, self-employed without employees, or unpaid family worker), age (19-29, 30-39, 40-49, 50-, 60-69, or 70 years old or older), family relationship satisfaction (not applicable, satisfactory, not satisfied or dissatisfied, or dissatisfied), relationship satisfaction with a spouse's siblings (not applicable, satisfactory, not satisfied or dissatisfied, or dissatisfied), relationship satisfaction with children (not applicable, satisfactory, not satisfied or dissatisfied, or dissatisfied), life satisfaction (good, moderate, or bad), relationship satisfaction with a spouse's parents (not applicable, satisfactory, not satisfied or dissatisfied, or dissatisfied), heritage // nationality of origin (Chinese, Korean-Chinese, Taiwanese or Hong Kong, Vietnam, Philippines, Japan, Mongolia, Uzbekistan, Russia, Thailand, Cambodia, North America, other southeast Asian countries, Western Europe, or Southern Asia), Korean citizenship (yes or no), social discrimination experiences (yes or no), and subjective state of health (good, moderate, or bad).

C. Base Model (Single Machine Learner): Classification and Regression Tree (CART)

One of the statistical decision classification methods, CART, uses the Gini Index to quantify impurity [14]. It is a binary split-based technique that only creates two child nodes from a parent node [14]. Advantages of CART include its ability to handle both continuous and categorical data as well as the ease with which its rules may be understood [15]. The likelihood that two randomly selected items from n elements belong to separate groups is known as the Gini coefficient [16]. The classifier and ideal classification that reduce the Gini coefficient the most are chosen as a child node once the algorithm has determined how much the Gini coefficient has decreased.

D. Radial basis Function Neural Network (RBF-NN)

RBF-NN is a data mining modeling technique that finds hidden patterns in data by repeating learning from real data and mimicking the human brain's neural network [17, 18]. It is a nonlinear flexible model used to forecast using data with complicated structures [17,18]. The neural network is a hierarchical algorithm made of several processing components [19]. The relationship between input and output is learned while weights are repeatedly changed using historical input data and corresponding output values [19]. The neural network is made up of an input layer made up of nodes that correspond to input variables and a hidden layer (or layers) made up of numerous hidden nodes. The hidden node uses a nonlinear function to handle the linear combination of variable values given from the input layer and sends it to the output layer or another hidden layer. The coupling function for the hidden layer in this work was RBF-NN, which employs the Radial Basis function. Fig. 1 illustrates the RBF-NN idea.

E. Random Forest (RF)

The RF algorithm randomly selects which decision trees to learn [21]. Multiple bootstrap samples are used in this technique to create a decision tree for each sample, and the results showing the highest frequency among the decision tree results are used to forecast the outcome of a new dataset [22]. In Fig. 2, the RF structure is shown.

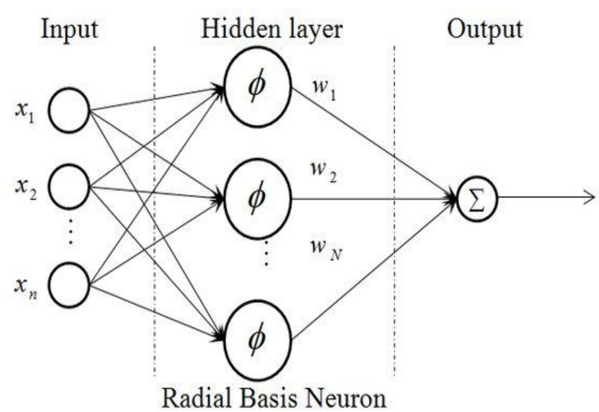

Fig. 1. Structure of the RBF-NN Modeling [20].

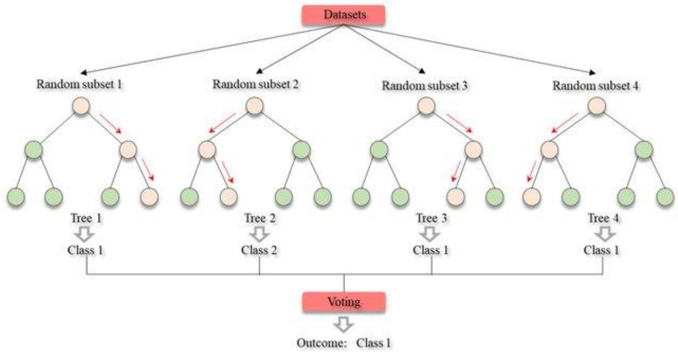

Fig. 2. Structure of the Random Forest Modeling [23].

F. Meta Model (Stacking Ensemble)

The stacking ensemble technique was used in this study to predict the use of Multi-FSCs. In recent years, stacking ensemble outperforms a single machine learning model in terms of generalization and robustness, and it has been used for classification and prediction in a variety of fields [24, 25, 26, 27, 28]. This technique generates a new model by stacking several different models [29]. It improves the final model's performance by combining the strengths of each model and supplementing the weaknesses of each model through two stages (base and meta) [29].

In the base stage (model), this model used RBF-NN, CART, and RF. In the meta stage, the model used the logistic regression algorithm (model). The regression algorithm is the most straightforward way to maximize generality and stability while increasing the reliability of the base model, and it is unlikely to overfit the training data [30]. Due to this, the regression algorithm has been employed in numerous recent publications [30, 31] as the meta model for the stacking ensemble algorithm. For the same reason, the regression algorithm for a meta model was used in this study. Fig. 3 illustrates the final stacking ensemble model's structural layout.

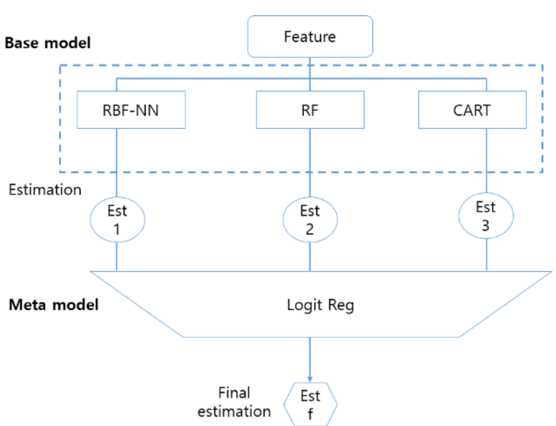

Fig. 3. Structure of the Prediction Model (Stacking Ensemble).

G. Meta Model (Stacking Ensemble)

Through the use of seven-fold cross-validation, each machine learning model was verified. With this approach, ten groups of equal size are randomly selected from the entire sample, and one group is used for validation while the other is used for learning. Seven times are run through this procedure to ensure that the knowledge is retained. The root-mean-square error (RMSE), error variance (Ev), and index of agreement (IA) were used in this study to assess how well the developed models performed as predictors. A lower RMSE among these indices denotes a prediction model with greater accuracy, whereas a lower Ev denotes a model with greater stability. A model becomes more stable as IA gets closer to 1.

III. Results

A. General Characteristics of Subjects by the Experience of using the Multi-FSC

Table I depicts the general characteristics of the research subject based on their experience with the Multi-FSC. Among the 281,606 subjects, 95,826 subjects (34.0%) had used the Multi-FSC at least once, while 185,780 subjects (66.0%) had not used the Multi-FSC. The results of chi-square test showed that the group with experience of using the Multi-FSC and the group without experience of using the Multi-FSC had significantly ($p<0.05$) different age, gender, relationship satisfaction with children, the highest level of education, family relationship satisfaction, relationship satisfaction with a spouse's parents, siblings, subjective health status, life satisfaction, occupation, nationality of origin, South Korean citizenship, and social discrimination experience.

TABLE I. General Characteristics According to the Experience of using the using the Multi-FSC, n (%)

Variables	**Experience of using the Multi-FSC**		**p**
	Yes (n=95,826)	***No (n=185,780)***	
Age			<0.001
19-29	42,122 (55.9)	33,258 (44.1)	
30-39	32,219 (37.5)	53,794 (62.5)	
40-49	16,905 (23.8)	54,090 (76.2)	
50-59	3,511 (11.3)	27,586 (88.7)	
60-69	827 (6.6)	11,614 (93.4)	
70+	244 (4.3)	5,438 (95.7)	
Gender			
Male	4,079 (7.2)	52,194 (92.8)	
Female	91,747 (40.7)	133,586 (59.3)	
Marital status			<0.001
Single	481 (5.9)	7,639 (94.1)	
Having a spouse	92,910 (36.4)	162,535 (63.6)	
Bereavement/widowed	872 (16.5)	4,421 (83.5)	
Divocement/divorced/separation/separated	1,564 (12.3)	11,185 (87.7)	
Residence			<0.001
Urban	61,866 (28.0)	159,150 (72.0)	
Countryside	33,961 (56.0)	26,630 (44.0)	
Level of education			<0.001

Elementary school or below	9,522 (35.4)	17,407 (64.6)	
Junior high school	23,060 (37.1)	39,103 (62.9)	
High school	41,661 (33.1)	84,172 (66.9)	
College or higher	21,583 (32.4)	45,098 (67.6)	
Family relationship satisfaction			<0.001
Good	68,356 (33.5)	135,699 (66.5)	
Moderate	23,033 (37.9)	37,757 (62.1)	
Bad	3,956 (45.8)	4,685 (54.2)	
Relationship satisfaction with a spouse's parent			<0.001
Not applicable	14,336 (22.1)	50,554 (77.9)	
Satisfactory	44,747 (37.4)	74,819 (62.6)	
Not satisfied or dissatisfied	29,032 (39.9)	43,670 (60.1)	
Dissatisfied	7,230 (44.3)	9,098 (55.7)	
Relationship satisfaction with a spouse’s siblings			<0.001
Not applicable	4,794 (16.5)	24,236 (83.5)	
Satisfactory	45,496 (34.5)	86,465 (65.5)	
Not satisfied or dissatisfied	36,005 (39.3)	555,691 (60.7)	
Dissatisfied	9,049 (43.5)	11,749 (65.1)	
Relationship satisfaction with children			<0.001
Not applicable	16,878 (19.1)	71,568 (80.9)	
Satisfactory	68,229 (44.3)	85,703 (55.7)	
Not satisfied or dissatisfied	8,851 (32.6)	18,272 (67.4)	
Dissatisfied	1,387 (34.8)	2,598 (65.2)	
Subjective health status			<0.001
Good	64,320 (37.0)	109,636 (63.0)	
Moderate	24,853 (32.2)	52,363 (67.8)	
Bad	6,653 (21.9)	23,781 (78.1)	
Life satisfaction			<0.001
Good	51,458 (35.9)	91,996 (64.1)	
Moderate	37,758 (32.7)	77,859 (67.3)	
Bad	6,611 (29.3)	15,925 (70.7)	
Heritage / nationality of origin			<0.001
Chinese	18,294 (28.6)	45,602 (71.4)	
Korean-Chinese	14,865 (15.6)	80,600 (84.4)	
Taiwanese or Hong Kong	453 (11.3)	3,572 (88.7)	
Japan	6,240 (37.4)	10,431 (62.6)	
Mongolia	1,769 (55.5)	1,420 (44.5)	
Vietnam	32,898 (66.1)	16,900 (33.9)	
Philippines	10,145 (69.3)	4,502 (30.7)	
Thailand	1,930 (57.5)	1,426 (42.5)	
Cambodia	3,805 (73.4)	1,381 (26.6)	
Uzbekistan	1,300 (57.0)	981 (43.0)	
Russia	570 (30.5)	1,299 (69.5)	
North America	466 (4.7)	9,514 (95.3)	
Other southeast Asian countries	832 (35.7)	1,499 (64.3)	
Southern Asia	792 (36.4)	1,384 (63.6)	
Western Europe	111 (4.8)	2,213 (95.2)	
Occupation			<0.001
Regular worker	13,441 (24.6)	41,230 (75.4)	
Temporary worker	16,681 (33.0)	33,880 (67.0)	
Day-to-day worker	7,324 (21.2)	27,158 (78.8)	
Self-employed with employees	401 (9.8)	3,672 (90.2)	
Self-employed without employees	2,365 (24.1)	7,430 (75.9)	
Unpaid family worker	72,410 (56.6)	72,410 (56.6)	
Korean citizenship			<0.001
Yes	35,709 (27.0)	96,491 (73.0)	
No	60,117 (40.2)	89,288 (59.8)	
Social discrimination experiences			0.001
Yes	40,162 (34.4)	76,705 (65.6)	
No	55,665 (33.8)	109,075 (66.2)	

B. A Comparison of the Prediction Performance of Models Anticipating the Utilization of the Multi-FSC

Fig. 4 to 6 depict the prediction performance (RMSE, Ev, and IA respectively) of six machine learning models for forecasting the use of Multi-FSCs. The findings revealed that the RBF-NN-Logit reg model had the best prediction performance (RMSE=0.20, Ev=0.45, and IA=0.68).

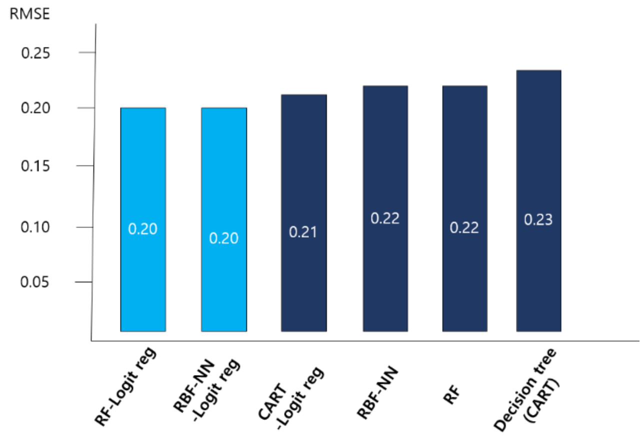

Fig. 4. Models for Machine Learning Compared using the Root-mean-Square Error.

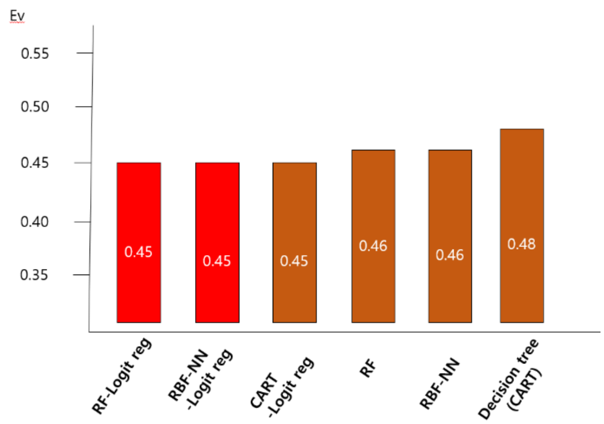

Fig. 5. Models for Machine Learning Compared using Error Variance.

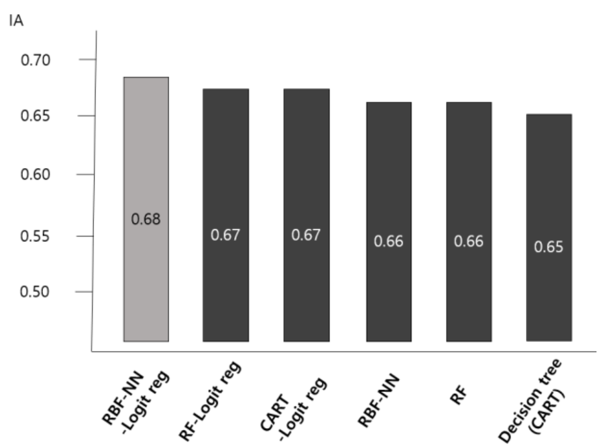

Fig. 6. Models for Machine Learning Compared using Index of Agreement.

C. Predictors for the use of Multi-FSCs in South Korea

The normalized importance of the variables of the RBF-NN-Logit reg model, the final model, is presented in Fig. 7. The results confirmed that gender (100%), relationship satisfaction with children (93%), age (85%), country of origin (69%), relationship satisfaction with a spouse's siblings (61%), and relationship satisfaction with a spouse's parents (57%) were a significant factor that carried a significant amount of weight in the South Korean experience of using Multi-FSCs. The gender (female) of these was the one that had the greatest impact on the final model.

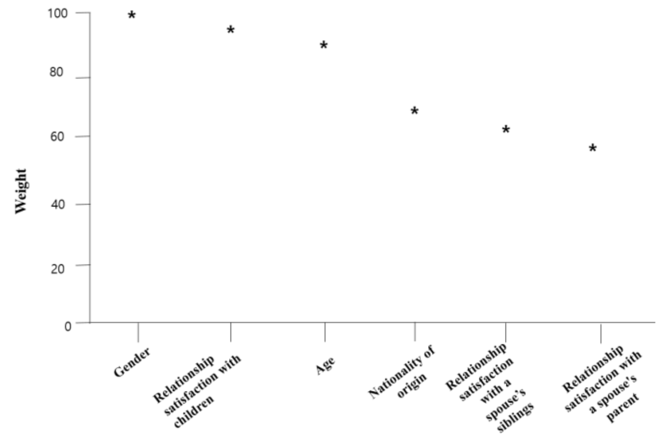

Fig. 7. Variable's Importance for use of Multi-FSCs (only top 6).

IV. Conclusions

In our study, the RBF-NN-Logitreg model predicted gender, relationship satisfaction with children, age, country of origin, relationship satisfaction with spouse's siblings, and relationship satisfaction with spouse's parents for the application of Multi-FSC. The South Korean government can use these forecast results as policy data to further promote Multi-FSC.

Using six machine learning techniques—three base models and three stacking ensemble models—this study created prediction models for factors linked to the use of the Multi-FSC in South Korea and evaluated the effectiveness of the models. The RBF-NN-Logit reg model based on the stacking ensemble algorithm had the best prediction performance, according to the results. Given that their RMSE was 0.03 less than that of the base models, prediction models based on the stacking ensemble technique in particular demonstrated improved accuracy (single machine learning model). Furthermore, given that their EV was 0.03 higher than the EV of base models, they had better stability. The findings of this study suggested that the prediction performance of the stacking ensemble may be superior to that of the single machine learning model when developing classification or prediction models using epidemiology data from a community.

The stacking ensemble takes longer to run than the single machine learning model because its algorithm is more complex than the base model [32, 33, 34, 35, 36, 37]. Furthermore, depending on the combination of the base model and meta model, it has been reported that it is more likely to cause overfitting than a single machine learning model in some cases [32, 33, 34, 35, 36, 37]. As a result, more research is required to evaluate the prediction performance of the stacking ensemble. Future research is required to investigate the stacking ensemble model with the best performance using more diverse combinations of base models and meta models in order to find a model that can minimize overfitting while maximizing accuracy.

ACKNOWLEDGMENT

This study was supported by 2021 Research Grant from Kangwon National University.

REFERENCES

[1] Statistics Korea, 2019 Multicultural Demographic Statistics. Statistics Korea, 2021. http://kostat.go.kr.

[2] Ministry of Public Administration and Security, Statistical foreign migrants 2019. Ministry of Public Administration and Security, Seoul, 2019.

[3] Ministry of Education, Science and Technology, Status of multicultural families and support measures, Ministry of Education. Science and Technology, Seoul, 2018.

[4] D. H. Lee, A study on multicultural families support act. The Journal of the Korea Contents Association, vol. 19, no. 7, pp. 650-658, 2019.

[5] Y. J. Oh, Role of multicultural family support center for school adaptation of elementary school children of multicultural families. Cultural Exchange and Multicultural Education, vol. 7, no. 3, pp. 79-99, 2018.

[6] J. H. Song, and T. Y. Lee, A study on the legislative process of the support for multicultural families act. Social Policy, vol. 39, no.3, pp.151-179, 2012.

[7] A. Kim, Social exclusion of multicultural families in Korea. Social Sciences, vol. 74, p. 63. 2018. https://doi.org/10.3390/socsci7040063

[8] K. J. Seon, The relationship between job-stress and psychological-exhaustion of counselors at multicultural family support center. Journal of The Korea Society of Computer and Information. vol. 18, no. 7, pp. 157-164, 2013.

[9] H. A. Lee, Analysis of satisfaction with the fatherhood programs provided in healthy family and multicultural family support center. Korean Family Resource Management Association, vol. 22, no. 3, pp. 61-76, 2018.

[10] M. K. Park, Y. J. Cha, and H. J. Lee, The effects of using the multicultural family support centers on the bicultural identity of the multicultural youths: focusing on propensity score matching. The Journal of Multicultural Society, vol. 12, no. 3, pp. 107-140, 2019.

[11] M. Chin, S. Noh, and H. So, Awareness of healthy family support center and multicultural family support center and parenting stress and family healthiness. Journal of Korean Home Management Association, vol. 35, no. 3, pp. 113-126, 2017.

[12] I. Kwon, B. Lee, and S. Kim, The experience of alienation of marriage migrant women in the healthy family and multicultural family support center. Welfare & Cultural Diversity Studies, vol. 22, no. 2, pp. 41-73, 2020.

[13] Ministry of Gender Equality & Family, A study on the national survey of multicultural families. Ministry of Gender Equality & Family, Seoul, 2012.

[14] W. Chen, X. Xie, J. Wang, B. Pradhan, H. Hong, D. T. Bui, Z. Duan, and J. Ma, A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena, vol. 151, pp. 147-160, 2017.

[15] H. Byeon, Development of prediction model for endocrine disorders in the Korean elderly using CART algorithm. International Journal of Advanced Computer Science and Applications, vol. 6, no. 9, pp. 125-129, 2015.

[16] H. Byeon, and. R. Lee, Prediction model for the smoking in Korean adolescent using CART algorism. Information, vol. 17, no. 12A, pp. 6273-6728, 2014.

[17] F. Bonanno, G. Capizzi, G. Graditi, C. Napoli, and G. M. Tina, A radial basis function neural network based approach for the electrical characteristics estimation of a photovoltaic module. Applied Energy, vol. 97, pp. 956-961, 2012.

[18] J. D. Wu, and J. C. Liu, A forecasting system for car fuel consumption using a radial basis function neural network. Expert Systems with Applications, vol. 39, no, 2, pp. 1883-1888, 2012.

[19] A. D. Dongare, R. R. Kharde, and A. D. Kachare, Introduction to artificial neural network. International Journal of Engineering and Innovative Technology, vol. 2, no. 1, pp. 189-194, 2012.

[20] H. He, Y. Yan, T. Chen, and P. Cheng, Tree height estimation of forest plantation in mountainous terrain from bare-earth points using a DoG-coupled radial basis function neural network. Remote Sensing, vol. 11, no, 11, pp. 1271, 2019.

[21] M. Belgiu, and L. Drăguţ, Random forest in remote sensing: a review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 114, pp. 24-31, 2016.

[22] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp. 93-104, 2012.

[23] J. Yang, J. Gong, W. Tang, Y. Shen, C. Liu, and J. Gao, Delineation of urban growth boundaries using a patch-based cellular automata model under multiple spatial and socio-economic scenarios. Sustainability, vol. 11, no. 21, pp. 6159, 2019.

[24] S. Cui, Y. Yin, D. Wang, Z. Li, Y. Wang, A stacking-based ensemble learning method for earthquake casualty prediction. Applied Soft Computing, vol. 101, pp. 107038. 2021.

[25] M. Akour, I. Alsmadi, and I. Alazzam, Software fault proneness prediction: a comparative study between bagging, boosting, and stacking ensemble and base learner methods. International Journal of Data Analysis Techniques and Strategies, vol. 9, no. 1, pp. 1-16, 2017.

[26] J. Lee, J. Kim, and W. Ko, Day-ahead electric load forecasting for the residential building with a small-size dataset based on a self-organizing map and a stacking ensemble learning method. Applied Sciences, vol. 9, no. 6, pp. 1231, 2019.

[27] T. P. Williams, and J. Gong, Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. Automation in Construction, vol. 43, pp. 23-29, 2019.

[28] Y. Xiong, M. Ye, and C. Wu, Cancer classification with a cost-sensitive naive bayes stacking ensemble. Computational and Mathematical Methods in Medicine, vol. 2021, pp. 5556992, 2021.

[29] G. Wang, J. Hao, J. Ma, and H. Jiang, A comparative assessment of ensemble learning for credit scoring. Expert Systems with Applications, vol. 38, no. 1, pp. 223-230, 2011.

[30] L. Feng, Y. Li, Y. Wang, and Q. Du, Estimating hourly and continuous ground-level PM2.5 concentrations using an ensemble learning algorithm: the ST-stacking model. Atmospheric Environment, vol. 223, pp. 117242, 2020.

[31] J. Chen, J. Yin, L. Zang, T. Zhang, and M. Zhao, Stacking machine learning model for estimating hourly PM2. 5 in China based on Himawari 8 aerosol optical depth data. Science of The Total Environment, vol. 697, pp. 134021, 2019.

[32] F. Divina, A. Gilson, F. Goméz-Vela, M. García Torres, and J. F. Torres, Stacking ensemble learning for short-term electricity consumption forecasting. Energies, vol. 11, no, 4, pp. 949, 2018.

[33] E. Menahem, L. Rokach, and Y. Elovici, Troika–an improved stacking schema for classification tasks. Information Sciences, vol. 179, no, 24, pp. 4097-4122, 2009.

[34] M. Surdeanu, and C. D. Manning, Ensemble models for dependency parsing: cheap and good?. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 649-652, 2010.

[35] J. Yan, Y. Qi, and Q. Rao, Detecting malware with an ensemble method based on deep neural network. Security and Communication Networks, vol. 2018, p. 7247095, 2018. https://doi.org/10.1155/2018/7247095.

[36] S. Young, T. Abdou, and A. Bener, Deep super learner: a deep ensemble for classification problems. In Canadian Conference on Artificial Intelligence, pp. 84-95, 2018.

[37] J. Moon, S. Jung, J. Rew, S. Rho, and E. Hwang, Combination of short-term load forecasting models based on a stacking ensemble approach. Energy and Buildings, vol. 216, pp. 109921, 2020.