International Journal of Engineering & Technology, 7 (2.12) (2018) 161-166

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Application of artificial neural network analysis and decision tree analysis to develop a model for predicting life satisfaction of the elderly in south korea

Haewon Byeon *

[1] *Department of Speech Language Pathology, College of Health Science,Honam University, 417, Eodeung-daero, Gwangju, Republic of Korea*
**Corresponding author E-mail: bhwpuma@naver.com*

Abstract

Background/Objectives: This study developed a prediction model with taking into account various factors that could affect the life satisfaction of the elderly in South Korea by using data mining techniques.
Methods/Statistical analysis: This study analyzed the data of 2,111 elderly (879 males and 1, 232 females) who were equal to or older than 60 among 7,761 people completed the Seoul Welfare Pane Study 2010. The life satisfaction, a result variable, was classified as 'satisfactory', 'normal', and 'dissatisfactory' based on the question of 'how are you satisfied with your current life?'The latent factors of the life satisfaction of the elderly were explored by using the neural network. The decision tree model was constructed by using the classification and regression tree (CART) algorithm.
Findings: Subjective friendship, subjective health status, subjective family relationship, and the highest level of education were significant classification variables. The most predominant predictive variable was subjective friendship. Moreover, it was predicted that 'the elderly with good subjective friendship and subjective health' and 'the elderly with good subjective friendship, subjective health, and family relationship and whose highest level of education was higher than middle school graduate' would be groups with high life satisfaction.
Improvements/Applications: It is necessary to expand the perceived social network support for promoting the family relationship and friendship as well as the health enhancement in order to improve the life satisfaction of the elderly
.

Keywords: *Datamining; Neural Network; Decision Tree; Risk Factors; Life Satisfaction.*

1. Introduction

The South Korean society faces a rapid aging problem in recent years. The proportion of the elderly, over 65 years old, was only 3.8% of the population in 1980 but it was reported as 13.0% in 2015, which was more than triple one.
Particularly, it is known that South Korea has the fastest aging rate in the world. It has been reported that it has entered an aging society, indicating that more than 7% of the population is over 65 years old, since 2000 [2]. Moreover, it is expected that it will enter a super-aging society, more than 21% of the population is over 65 years old, in 2027, just in 27 years from it. It is faster than China (34 years) and Thailand (35 years), which are experiencing rapid aging. The US Bureau of Statistics forecasted that, if this aging trend continues, South Korea will be the country with the second highest aging population ratio in the world in 2050 [3].
Due to rapid aging, the factors constituting the quality of the life of the elderly and affecting the life satisfaction has been diversified and changed rapidly. The economic problem was the most serious problem of the elderly in 2005 but the health problem was found more critical to the elderly than the economic problem in 2007 [4]. Therefore, in order to improve the quality of the life of the elderly, it is necessary to identify various factors influencing the quality of life in addition to the economic aspect.
Particularly, the "new old generation", who were baby boomers and born in the 1950s after the Korean War, is self-reliance, productive, and actively participating in social activities [5], [6]. This generation is entering the aged community and has a different lifestyle than the previous generations [5], [6]. This generation is interested in the quality and satisfaction of life than previous generations [5].
Life satisfaction means the subjective assessment of the own quality of life [7]. Therefore, the life satisfaction of the elderly is determined by the expectation of the present situation and the actual satisfaction in the real life. It is known that the elderly of South Korea has low life satisfaction. The report of Organization for Economic Cooperation and Development (OECD) showed that the life satisfaction of South Korean was 5.8 out of 10, which was lower than the OECD average (6.58) [8]. It revealed that the life satisfaction abruptly decreased with older respondents. Consequently, it is necessary to identify the factors affecting the life satisfaction of the elderly and improve them.
Previous studies have reported that sociodemographic characteristics, physical and mental health status, economic activities and economic level, and social relationship are factors affecting the life satisfaction of the elderly 9-14. These studies revealed that the male elderly had higher life satisfaction than the female elderly, the religious belief positively influenced the life satisfaction of the elderly, and the age adversely affected it.
However, a number of previous studies just explored the individual factors in order to predict the life satisfaction of the elderly [15] and only a few studies considered the interactions of factors. Especially, the previous studies mostly focused on the generation before the

Copyright © 2018 Haewon Byeon. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

baby boomer generation so studies reflecting the characteristics of "new old generation" are needed.
This study developed a prediction model with taking into account various factors that could affect the life satisfaction of the elderly in South Korea by using data mining techniques.

2. Materials

2.1. Subject

This study analyzed the raw data of Seoul Welfare Pane Study, conducted by Seoul Welfare Foundation, targeting the citizens of Seoul from Jun 1 to Aug 31, 2010. Seoul Welfare Pane Study was carried out to identify the welfare status of households residing in Seoul, recognize the actual condition of welfare vulnerable class, and estimate the demand of welfare services after being approved by Statistics Korea (No. 20113) in 2009 [16]. This study used the households residing in Seoul at the time of 2005 Population and Housing Census. This study sampled by using the stratified cluster extraction method for the 25 city districts of Seoul. This study employed the computer-assisted personal interviewing, which an interviewer visited a target household and entered the responses regarding the structured questionnaire to a portable computer. This study analyzed the data of 2,111 elderly (879 males and 1,232 females) who were equal to or older than 60 among 7,761 people completed the survey.

2.2. Measurement and definition of variables

The life satisfaction, a dependent variable, was classified as 'satisfactory', 'normal', and 'dissatisfactory' based on the question of 'how are you satisfied with your current life?' Explanatory variables included age (60-69 and 70 years and over), sex (male and female), the highest level of education (below elementary school, middle school, high school, and college graduate and above), current employment status (employed and unemployed), total monthly average income per household (between 2million KRW, 2-4million KRW, and above 4million KRW), the presence of a spouse (living with a spouse, not living with a spouse, and no spouse), current smoking status (yes and no), high-risk drinking (yes and no), subjective health status (good, normal, and bad), current chronic illness (yes and no), regular exercise (yes and no), volunteer activity experience within the past one year (yes and no), subjective family relationship (good, normal, and bad), friendship (good, normal, and bad), and depression symptom within the past one month (yes and no). Chronic illnesses included circulatory illness (e.g., hypertension and stroke), endocrine illness (e.g., diabetes), musculoskeletal illness (e.g., osteoporosis), respiratory illness (e.g., chronic bronchitis), ENT illness, cancer, and urology illness (e.g., chronic renal failure). They were binary variables (yes and no). High-risk drinking was defined as drinking more than seven shots of soju per day (alcohol 60g) for a male and five shots of soju per day (alcohol 40g) for a female.

3. Methods

3.1. The latent factors of the life satisfaction of the elderly

The latent factors of the life satisfaction of the elderly were explored by using the neural network. The neural network is a data mining modeling technique [17] identifying the hidden pattern from the actual data through repetitive learning processes imitating the neural network of the human brain. It is a non-linear model used to solve a prediction question from the data with complex structure [18]. This study used the RBF neural network, which utilizes the radial basis function (RBF) as the combining function of hidden layers [19]. When the relative importance of inputs of a variable was equal to or higher than 0.05 from the artificial neural network analysis, it was considered as a major explanatory variable affecting the decision of the dependent variable and the variable was included in the decision tree model.

3.2. Decision tree model

The decision tree model was constructed by using the classification and regression tree (CART) algorithm. CART measures the impurity by using Gini Index and it is an algorithm based on a binary split forming only two child nodes from a parent node [20].
This study set the separation and merger threshold value of the decision rule for the CART algorithm to 0.05. The number of parent nodes was 200 and that of child nodes was 100. The number of split branches was limited to 5 [21]. The validity evaluation of the final model was determined by using the 10-fold cross-validation method. The analysis was conducted by Decision Tree version 20.0(IBM Inc., Chicago, Illinois, USA) and statistical significance was determined at alpha=0.05.

4. Results

4.1. The general characteristics of the subjects by their life satisfaction

The general characteristics of the subjects by their life satisfaction are shown in Table 1. Among 2,111 subjects, 414 subjects (19.6%) responded that they were satisfied with their lives. The results of chi-square test revealed that life satisfaction was significantly affected by the highest level of education, household income, the presence of a spouse, current smoking, high risk drinking, subjective health status, the present of current chronic illness, regular exercise, volunteer activity experience in the past one year, subjective family relationship, subjective friendship, and the depression symptom within the past one month($p<0.05$).

4.2. Exploring potential factors of life satisfaction in elderly using artificial neural network

The artificial neural network analysis was conducted for the 1,151 people of the learning sample (60.9%), the 547 people of the verification sample (29.0%), and the 191 people of the validation sample (10.1%). The analysis withdrew 9 hidden layers generating the minimum error from the verification sample. The classification accuracies were 63.9, 62.4, and 63.6% for the training, verification, and validation samples, respectively. The area under receiver operating characteristic (AUROC) was 0.879 (Figure 1) and the fitness and explanatory power of the classification model were excellent (Figure 2).

Table 1: Characteristics of the Subjects Based Life Satisfaction, N (%)

Variables	Life satisfaction			p
	Dissatisfactory (n=752)	Normal (n=945)	Satisfactory (n=414)	
Age				0.062
60-69	382 (33.8)	508 (45.0)	240 (21.2)	
70 years and over	370 (37.7)	437 (44.5)	174 (17.7)	
sex				0.069
Male	289 (32.9)	405 (46.1)	185 (21.0)	
Female	463 (37.6)	540 (43.8)	229 (18.6)	
Highest level of education				<0.001
Below elementary school	417 (45.6)	401 (43.9)	96 (10.5)	
Middle school	128 (34.3)	173 (46.4)	72 (19.3)	
High school	145 (29.2)	228 (46.0)	123 (24.8)	
College graduate and above	62 (18.9)	143 (43.6)	123 (37.5)	

Total monthly average income per household				<0.001
Between 2million KRW	619 (45.3)	571 (41.8)	176 (12.9)	
2-4million KRW	105 (21.6)	237 (48.8)	144 (29.6)	
Above 4million KRW	8 (8.6)	60 (64.5)	25 (26.9)	
Presence of a spouse				<0.001
Living with a spouse	442 (31.1)	658 (46.4)	319 (22.5)	
Not living with a spouse	22 (52.4)	15 (35.7)	5 (11.9)	
No spouse	288 (44.3)	272 (41.8)	90 (13.8)	
Current employment status				0.059
Employed	107 (30.2)	175 (49.4)	72 (20.3)	
Unemployed	645 (36.7)	770 (43.8)	342 (19.5)	
Current smoking status				<0.001
No	650 (34.3)	858 (45.2)	389 (20.5)	
Yes	102 (47.7)	87 (40.7)	25 (11.7)	
High-risk drinking				0.025
No	697 (36.1)	868 (45.0)	365 (18.9)	
Yes	55 (30.4)	77 (42.5)	49 (27.1)	
Subjective health status				<0.001
Good	91 (15.6)	229 (39.2)	264 (45.2)	
Normal,	173 (24.8)	440 (63.0)	85 (12.2)	
Bad	488 (58.9)	276 (33.3)	65 (7.8)	
Current chronic illness				<0.001
Yes				
No	134 (26.3)	237 (46.6)	138 (27.1)	
Regular exercise				<0.001
No	497 (42.5)	520 (44.4)	153 (13.1)	
Yes	255 (27.1)	425 (45.2)	261 (27.7)	
Volunteer activity experience within the past one year				<0.001
Yes	29 (20.4)	60 (42.3)	53 (37.3)	
No	723 (36.7)	885 (44.9)	361 (18.3)	
Subjective family relationship				<0.001
Good	295 (24.3)	542 (44.6)	377 (31.1)	
Normal,	286 (43.2)	348 (52.6)	28 (4.2)	
Bad	138 (77.1)	37 (20.7)	4 (2.2)	
Friendship				<0.001
Good	137 (19.5)	224 (31.9)	342 (48.6)	
Normal,	357 (32.7)	674 (61.7)	61 (5.6)	
Bad	258 (81.6)	47 (14.9)	11 (3.5)	
Depression symptom within the past one month				<0.001
No	426 (27.1)	773 (49.2)	371 (23.6)	
Yes	326 (60.3)	172 (31.8)	43 (7.9)	

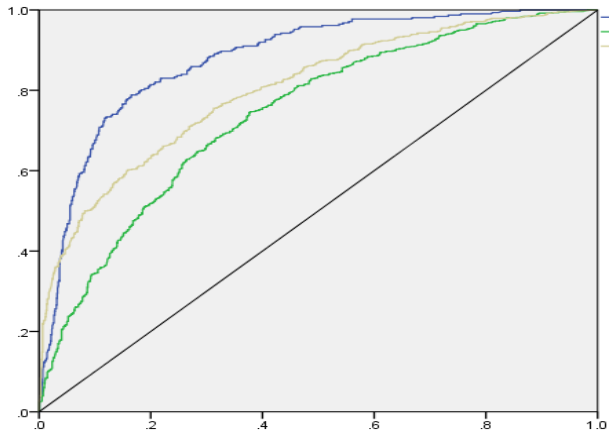


Fig. 1: Arear under Receiver Operating Characteristic of Neural Network Model.

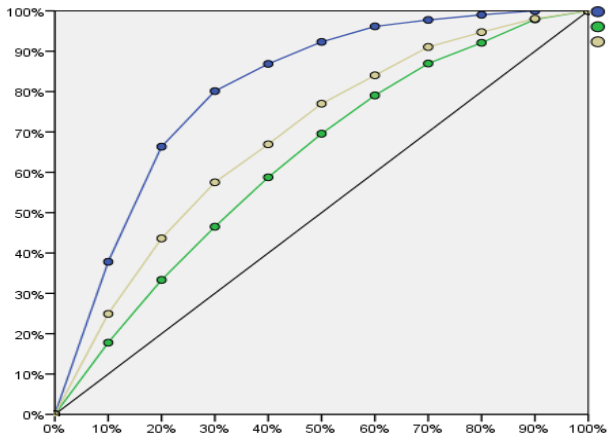


Fig. 2: Gain Index Ratio of Artificial Neural Network Model**.**

A synaptic weighted network diagram was constructed by using a multi-layer neural network model by inputting all explanatory variables as input variables and the results of it are shown in Figure 3. In the network diagram, the synapse weights visually present the relationship between a given layer and the next layer and the thickness of the line indicates the magnitude of the connection weight. This model assumed that sex, the highest level of education, current smoking status, subjective health status, current illness status, regular exercise, subjective family relationship, subjective friendship, and depression symptom within the last one month as major variables of life satisfaction holding high weights. The relative importance of inputs is presented in Table 2. The normalized importance normalized the units of the magnitude of each variable's influence by the influence category and it is shown in Figure 4.

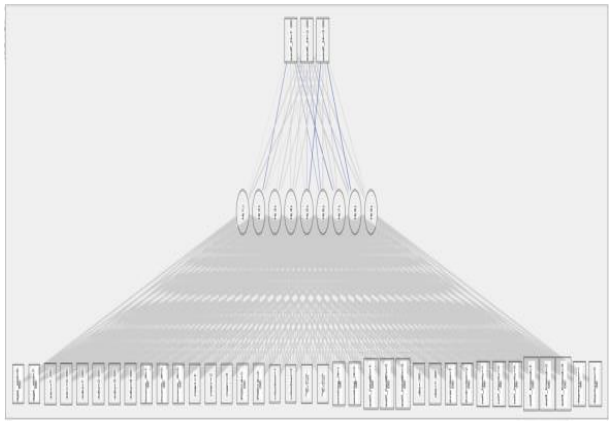


Fig. 3: Synaptic Weighted Network Diagram.

Table 2: Relative Importance of Inputs

Inputs	Relative importance	Normalized importance, %
Age	0.033	19.1
Sex	0.056	32.1
Level of education	0.051	29.0
Monthly average income per household	0.042	23.8
Presence of a spouse	0.040	23.1
Current employment status	0.043	24.6
Volunteer activity experience within the past one year	0.020	11.7
High-risk drinking	0.031	17.5
Current smoking status	0.076	43.4
Subjective health status	0.143	81.9
Current chronic illness	0.053	30.5
Regular exercise	0.055	31.5
Subjective family relationship	0.095	54.5
Friendship	0.175	100.0
Depression symptom within the past one month	0.087	49.7

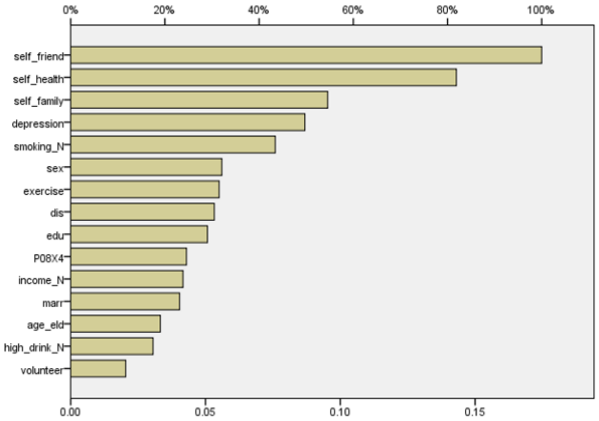


Fig. 4: Normalized Importance of Artificial Neural Network Model.

4.2. The prediction model of life satisfaction of the elderly based on a decision tree model

The prediction model of life satisfaction of the elderly based on a decision tree model is presented in Figure 5. Subjective friendship, subjective health status, subjective family relationship, and the highest level of education were significant classification variables. The most predominant predictive variable was subjective friendship.Table 3 shows the profitability index of the decision tree model, which presents the path predicting the life satisfaction of the elderly in the decreasing order of magnitude. Two paths were found to be significant for predicting the life satisfaction of the elderly among total 10 paths. The first path had the highest profitability index value (349.7%) and it was 'the elderly with good subjective friendship and subjective health'. The second path was 'the elderly with good subjective friendship, subjective health, and family relationship and whose highest level of education was higher than middle school graduate' and the profitability index value was 277.4%. The 10-fold cross-validation test was conducted to compare the stability of the derived models. The results showed that the risk index and error classification rate of the model were 0.345 and 34%, respectively. They were similar with 0.358 and 35% of the prediction model, respectively.

Table 3: Gains Chart of Predictor Variable by CART Algorithm

Node no	Node n (%)[1]	Gain n (%)[2]	Response %[3]	Gain Index %[4]	Characteristics
3	347 (16.4)	238 (57.5)	68.6	349.7	Elderly with good subjective friendship and subjective health
14	155 (7.3)	66 (15.9)	42.6	217.1	Elderly with good subjective friendship, subjective health, and family relationship and whose highest level of education was higher than middle school graduate

1 Node n(%); node number, % to 2,111
2 Gain n(%); gain number, % to 414
3 Response (%): The fraction of the life satisfaction
4 Gain index (%):=349.7 in total 10 node

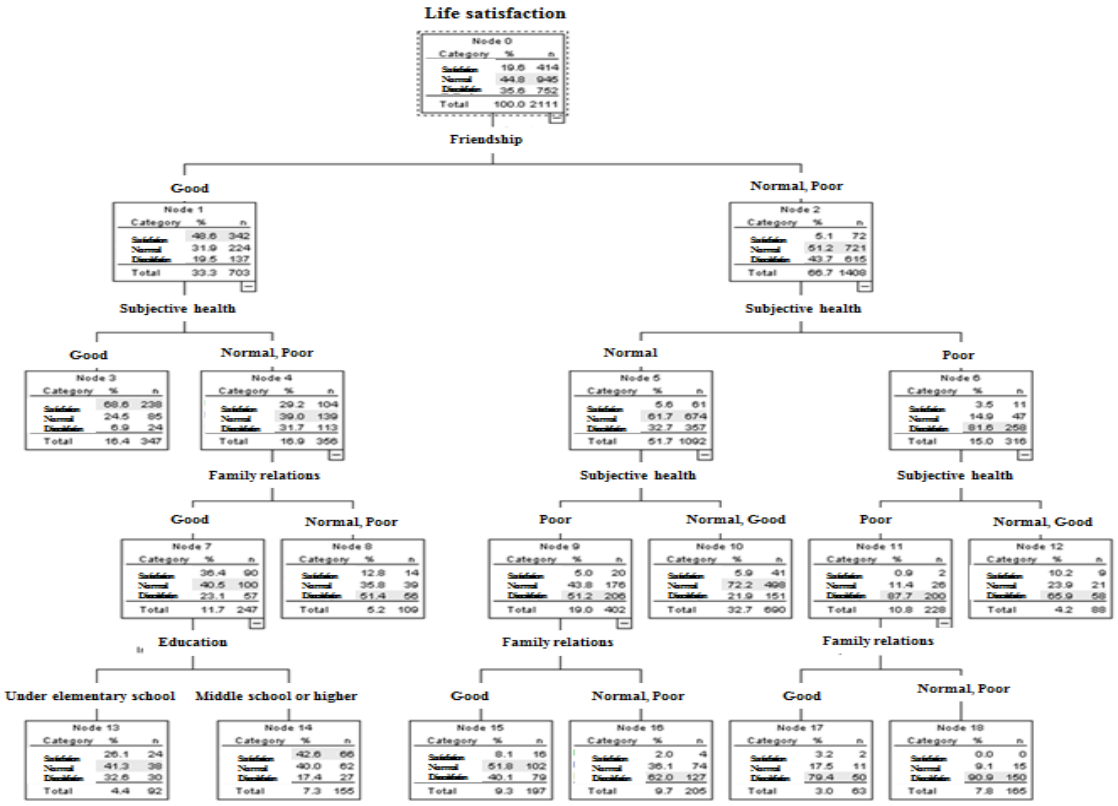


Fig. 5: A Decision Tree Model for Life Satisfaction of Korean Elderly.

5. Discussion

This study predicted the life satisfaction of the elderly in South Korea by using a data mining technique and it showed that subjective friendship, subjective health, subjective family relationship, and the highest level of education were significant classification variables. Moreover, it was predicted that 'the elderly with good subjective friendship and subjective health' and 'the elderly with good subjective friendship, subjective health, and family relationship and whose highest level of education was higher than middle school graduate' would be groups with high life satisfaction. Numerous previous studies reported that the elderly with higher satisfaction in the parent-children relationship including the communication with children had higher life satisfaction 22. The results of the previous studies also showed that the elderly who had good physical health and abilities to do daily activities had higher life satisfaction 23.

Particularly, the subjective friendship was the most dominant predictor in this study and the results implied that the participation in the social activities posed a great impact on life satisfaction. The results of the previous study also indicated that the elderly who participated in social activities had high life satisfaction 22. Additionally, the frequency of participation in leisure activities and life satisfaction had a positive relationship and, on the contrary to this, the elderly who had a lower social interaction frequency was reported to have lower life satisfaction 24.

Nevertheless, the community connection of the elderly in South Korea is yet insufficient. The survey of OECD revealed that 'the perceived social network support', reflecting the presence of friends or relatives to rely on, was the lowest among 34 member countries 3. Especially, the level of perceived social network support was very low with the elderly. Therefore, it is necessary to expand the perceived social network support for promoting the family relationship and friendship as well as the health enhancement in order to improve the life satisfaction of the elderly. Furthermore, the results of this study suggested that multidimensional support is necessary to boost the life satisfaction of the elder.

References

[1] National Statistical Office, Estimates and Projections of the Population of the Korea, National Statistical Office, Seoul, 2016.

[2] Statistics Korea, Statistics on the Elderly, National Statistical Office, Seoul, 2013.

[3] He W, Goodkind D, Kowal P. R, An aging world: 2015, United States Census Bureau, Washington, DC, 2016.

[4] Chung K. H, Lee Y. K, Park B. M, Lee S. J, Lee Y, Analysis of the Survey of Living Conditions and Welfare Needs of Korean Older Persons 2011, Korea Institute for Health and Social Affairs, Seoul, 2012.

[5] Jung K. H, Son C. K, Park B. M, Policy Challenges posed by Emerging 'New Class of Older Persons', Korea Institute for Health and Social Affairs, Seoul, 2010.

[6] Schiffman L. G, Sherman E, Value orientations of new-age elderly: the coming of an ageless market. Journal of Business Research, 1991, 22(2), pp. 187-94.

[7] Corrigan J. D, Kolakowsky-Hayner S, Wright J, Bellon K, Carufel P, The satisfaction with life scale. The Journal of Head Trauma Rehabilitation, 2013, 28(6), pp. 489-91.

[8] Durand M, The OECD Better Life Initiative: How's Life? and the Measurement of Well-Being. Review of Income and Wealth, 2015, 61(1), pp. 4-17.

[9] Banting K, Sharpe A, France, St-Hilaire, The Review of Performance and Social Progress. The Institute for Research on Public Policy, Montreal, 2001.

[10] Blanchflower D. G, Oswald A. J, Well-being over time in Britain and the USA. Journal of Public Economics, 2004, 88(7), pp. 1359-86.

[11] Kirchengast S, Haslinger B, Gender differences in health-related quality of life among healthy aged and old-aged Austrians: cross-sectional analysis. Gender Medicine, 2008, 5(3), pp. 270-78.

[12] Breeze E, Jones D. A, Wilkinson P, Bulpitt C. J, Grundy C, Latif A. M, Fletcher A. E, Area deprivation, social class, and quality of life among people aged 75 years and over in Britain. International Journal of Epidemiology, 2005, 34(2), pp. 276-83.

[13] Sun F, Norman I. J, While A. E, Physical activity in older people: a systematic review. BMC Public Health, 2013, 13(1), pp. 449.

[14] Enkvist Å, Ekström H, Elmståhl S, What factors affect life satisfaction (LS) among the oldest-old?. Archives of Gerontology and Geriatrics, 2012, 54(1), 140-45.

[15] Depp C. A, Jeste D. V, Definitions and predictors of successful aging: a comprehensive review of larger quantitative studies. The American Journal of Geriatric Psychiatry, 2006, 14(1), pp. 6-20.
[16] Seoul Welfare Foundation, Seoul Welfare Panel Study 2010, Seoul Welfare Foundation, Seoul, 2010.
[17] Rafiq M. Y, Bugmann G, Easterbrook D. J, Neural network design for engineering applications. Computers & Structures, 2001, 79(17), pp. 1541-52.
[18] Byeon H, Cho S, The Factors of Subjective Voice Disorder Using Integrated Method of Decision Tree and Multi-Layer Perceptron Artificial Neural Network Algorithm. International Journal of Advanced Computer Science and Applications, 2016, 7(5), pp. 112-16.
[19] Cho S, Yu S, Byeon H, A Prediction Model for Benign Laryngeal Disease Using Supervised Learning Techniques. International Journal of Bio-Science and Bio-Technology, 2016, 8(4), pp. 105-10.
[20] Byeon H, The risk factors of laryngeal pathology in Korean adults using a decision tree model. Journal of Voice, 2015, 29(1), pp. 59-64.
[21] Lee R, Byeon H, The factors that affects the needs of the Korean language education in children in multi-cultural families: results of a national cross sectional survey. International Journal of Applied Engineering Research, 2015, 10(10), pp. 26657-67.
[22] Lee J, Predictors of life satisfaction among older adults in s. Korea: differences by education level. Journal of the Korea Gerontological Society, 2010, 30(3), pp. 709-26.
[23] Hu S, Kim J, Analysis of multi-level effectiveness on life satisfaction in old age at KLIPS 2006. Journal of the Korea Gerontological Society, 2011, 31(2), pp. 407-18.
[24] Back J. U, Kim S. W, Kim M. Y, Factors that influence satisfaction level towards the life by the leisure activity of the users of the senior citizens' leisure welfare facilities. Korean Journal of Clinical Social Work, 2010, 7(1), pp. 37-58.