ISSN: 2508-7894 © 2021 KODISA & KAIA.
KJAI website: http://www.kjai.or.kr
doi: http://dx.doi.org/10.24225/kjai.2021.9.1.29

Adoption Factor Prediction to Prevent Euthanasia Based on Artificial Intelligence

Song-Eun KIM [1], Jeong-Hyun CHOI [2], Minsoo KANG [3]

Received: February 15, 2021. Revised: March 24, 2021. Accepted: June 05, 2021.

Abstract

In this paper, we analyzed the factors of adoption and implemented a predictive model to activate the adoption of animals. Recently, animal shelters are saturated due to the abandonment and loss of companion animals. To address this, we need to find a way to encourage adoption. In this paper, a study was conducted using two data from an open data portal provided by Austin, Texas. First, a correlation analysis was conducted to identify the attributes that affect the result value, and it was found that Animal Type Intake, Intake Type, and Age upon Outcome influence the Outcome Type with correlation coefficients of 0.4, 0.26, and -0.2, respectively. For these attributes, the analysis was conducted using Multiclass Logistic Regression. As a result, dogs had a higher probability of Adoption than cats, and animals subjected to euthanasia were more likely to adopt. In the case of Public Assist and Stray, it was found that the Missing rate was high. Also, the length of stay for cats increased to 12.5 years of age, while dogs generally adopted smoothly at all ages. These results showed an overall accuracy of 62.7% and an average accuracy of 91.7%, showing a fairly reliable result. Therefore, it seems that it can be used to develop a plan to promote the adoption of animals according to various factors. Also, it can be expanded to various services by interlocking with the webserver.

Keywords : Pet, Animal Shelter, Adoption, Transfer, Euthanasia

Major Classification Code : Artificial Intelligence, Supervised Learning, Logistic Regression

1. Introduction

Abandoned animals refer to animals or companion animals that are abandoned due to the mistake of the owner or intentional purpose. It is also defined as an animal that has been thrown away in a paper box or wandering around in public places such as roads and parks without the owner. (Hong, 2015) Recently, as the number of single-person households has increased, the number of households raising pets is also increasing. Accordingly, various social problems have arisen. The biggest problem among them is the abandonment and loss of companion animals. Many abandoned or lost companion animals are either protected or euthanized in shelters. (Yu, 2019) Besides, even if they are not euthanized, the management cost of the shelter is limited, so the environment is poor and it is difficult to keep the animals in a healthy state. (Protopopova & Wynne, 2014) Since animals in animal shelters are very unlikely to return to their original owners, most animal shelters are currently saturated and management costs are limited, so the

1 First Author, Student, Department of Medical IT, Eulji University, Korea, Email: songeun4939@naver.com.
2 Co-Author, Director, LG Uplus corporation, Korea, Email: nevergiveup@lguplus.co.kr.
3 Corresponding Author, Professor, Department of Medical IT, Eulji University, Korea, Email: mskang@eulji.ac.kr.

environment is poor and it is difficult for animals to remain healthy. Therefore, to prevent unnecessary euthanasia of animals, death from disease or natural deaths due to poor conditions, it is necessary to find a way to encourage the adoption of animals that cannot return to their owners within the deadline. (Choi et al., 2019) In particular, more and more places are violating euthanasia regulations for animals that are euthanized. Also, most of them are profit-structured animal shelters. (Animal People, 2020) So, if we can analyze the patterns of the characteristics of the animals being adopted in shelters, it can help us to be in a more favorable position for adoption. (Jang, 2013)

Therefore, in this paper, we intend to activate the adoption of animals to prevent unnecessary euthanasia while solving the saturation of animal shelters. To this end, we will analyze the characteristics that affect adoption and implement a predictive model to proceed with this study. Also, the study results can be used to more effectively expose animals with high adoption probability to shorten the time spent in shelters. The economic benefits obtained through this can be used to provide a better environment for unadopted animals, and to expand adoption opportunities by delaying euthanasia.

For this, data from the open data portal provided by Austin, Texas, is used. The data used are data on the condition when the animal arrives at the animal shelter and when it leaves. After deriving significant factors through correlation analysis using the given data, adoption was predicted according to various factors using Multiclass Logistic Regression.

2. Literature Review

In the paper of Hyunsun Yoo's “A Study on the Adoption Prediction Model according to the Characteristics of Abandoned and Lost Dogs” written in 2019, it is said that abandonment accidents are increasing as the number of households with companion animals increases. Also, the animal protection management system managed by the Animal and Plant Quarantine Service publishes a notice, and after the announcement period, organic animals are adopted or euthanized according to domestic law. To prevent such unnecessary euthanasia of animals, it is suggested that measures to encourage the adoption of animals that have not been returned to their owners within the deadline are needed. Studies conducted overseas show that the adoption of animals includes age, sex, skin color, breed, obedience status, injury status, and reasons for abandonment. Following this abstract, the study conducted in this paper attempted to estimate the possibility of adoption by using information such as city, breed, skin color, sex, neutralization status, age, and weight obtained through the announcement of the animal protection management system, and clustering analysis using decision tree and k Using the -prototype method, three variables, such as breed, age, and weight, identified the main determinants of adoption. And It was also possible to predict the adoption probability by learning the logistic regression model and naïve Bayes model for clusters. (Yoo, 2019) In the abstract of the paper “Mobile app design to promote adoption of abandoned animals” by Jayoung Hong written in 2015, as interest in companion animals is increasing recently, the number of discarded animals is increasing, and social problems related to this are also occurring. Besides, since the rate of such discarded animals becoming companion animals is low, it is another problem, so the research intends to suggest a plan to promote adoption. The research was conducted by looking for the definition of abandoned companion animals, investigating the actual state and state of adoption perception, and the causes and problems of abandonment. Besides, pre-and in-depth interviews were conducted for those involved in adoption. These results were used to derive requirements, improvement plans, and functions through Persona using affinity diagrams. Through this, the concept of the “Adopt Pet” service was derived and the main functions were defined. Finally, an experiment was performed on the manufactured prototype. In other words, in this paper, through an in-depth investigation of behavior patterns, reasons, and requirements of people interested in adopting abandoned animals, a method for revitalizing adoption was proposed by providing observation information of abandoned animals. (Hong, 2015)

"Applying Artificial Intelligence for Diagnostic Classification" by Eun Soo Choi, Hee Jeong Yoo, Min Soo Kang, and Soon Ae Kim of Korean Autism Spectrum Disorder", Azure ML using a multiclass decision forest algorithm was applied, and the diagnostic algorithm score value of 1,269 Korean ADI-R test data was used for prediction. In the second experiment, we used 539 Korean ADI-R case data (over 48 months with verbal language) to apply mutual information to rank items used in the ADI diagnostic algorithm. The experiments were conducted using Microsoft Azure Machine Learning Studio. (Choi et al., 2020)

Seo Won Song, Young Sang Kwak, Min Soo Kang are studied by “A Study on Graph-based Weighted KNN based on Machine Learning”. This paper was that the KNN algorithm is one of the classification techniques of machine learning and is used because it is simple and easy to understand. It is a method that classifies objects without labels into the category of the most similar of the labeled objects. This method finds K-nearest-neighbors according to the similarity between samples in the training data set and classifies them in an intuitive way. (Song et al., 2020) “Determining factors for successful adoption of dogs from

an animal shelter", written in 1998 by JM Posage and two others, aims to determine which characteristics are related to successful adoption through logistic regression analysis. As a result, Terrier, hound, toy, and nonsporting breeds were found to be significantly associated with a successful adoption. And Certain coat colors (gold, gray, and white), small size, and history of an indoor environment were also significant predictors of a successful adoption. The correlation coefficient (0.048) indicated that only a small percentage of variance in adoption success could be explained by the multiple logistic regression model. (Posage et al, 1998) Also, "Prediction of Adoption Versus Euthanasia Among Dogs and Cats in a California Animal Shelter" by Merry Lepper and two others, written in 2002, investigates the factors determining the adoption of cats and dogs. The study was conducted on animals confiscated by the Sacramento County Department of Animal Care and Regulation, and a model to predict conditional adoption probability was constructed using logistic regression and a final multiple logistic regression. As a result of these studies, age, sex, hair color, and reason for giving up were the main determinants of cat adoption. Besides, age, sex, hair color, the reason for giving up, breed, obedience status, and injury status were the main factors in determining dog adoption. Therefore, the study concludes that shelter staff can use this information to increase the adoption of animals. (Merry et al, 2002) Looking at the "Prediction Model for Adoption Probability of Abandoned and Lost Dogs in Local Government Animal Shelters," by Sung-sung Choi and five others written in 2019, it was abandoned on the website of the Animal Protection Management System operated by the Agriculture, Forestry and Livestock Quarantine Division Announcements on lost companion animals are crawled, pre-processed, and utilized. Through this, the factors affecting the success of the dog's adoption are identified and the adoption probability prediction model is established. Also, through this, the animal protection management system website or other dog adoption-related organizations suggest ways to promote adoption. (Choi et al, 2019) Therefore, in this paper, a correlation analysis was conducted using two data provided by the Austin Animal Shelter to identify major factors that influence animal adoption, and logistic regression analysis was performed using these factors. Also, based on these results, it is proposed to expand it to web or application services in the future. Azure is Microsoft's cloud computing platform that has been in service since 2010. This platform provides more than 600 services, and in this paper, Azure Machine Learning Studio is used. At this time, this platform can collect and manage data in the cloud, create a model, and easily build a web service and apply it to various devices. Also, unlike existing tools, it provides an easy-to-access GUI environment in consideration of user convenience. Also, you can easily create a model by dragging and dropping blocks. And scripts written in R and Python languages can be inserted and utilized in block form, and the results can be checked through visualization. (Kang et al., 2018) Supervised Learning is a method of machine learning to infer a function from Training Data. Training data generally includes the properties of the input object in the form of vectors, and the desired result for each vector is displayed. Among these inferred functions, outputting continuous values is called regression, and marking what kind of value a given input vector is called classification. What the Supervised Learner does is to correctly guess the value it wants to predict for a given data from the training data. To achieve this goal, the learner must be able to generalize and process situations that did not appear from existing training data through appropriate methods. (Wikipedia, 2020) Logistic regression analysis is a regression model used to statistically estimate categorical variables in which the targets to be analyzed are divided into two or more groups. Such logistic regression analysis includes binary logistic regression analysis and polynomial logistic regression analysis. When there are three or more selectable attributes, polynomial logistic regression analysis is used. (Lee, 2020) In the abstract of the paper "A Study on Comparison of Lung Cancer Prediction Using Ensemble Machine Learning" by Yujin NAM and Wonji SHIN written in 2019, we review the existing research on artificial intelligence technology that can be used for judgment on lung cancer, a chronic disease. It is said that the study was conducted through comparative analysis using Azure ML provided by Microsoft. The results of this study show different predictions calculated by three algorithms: SVM, Two-Class Support Decision Jungle, and Multiclass Decision Jungle. The results of the experiment show that the Multi-Class Decision Jungle algorithm has the highest accuracy. There is a limit to the size of big data. However, it is said that it is possible to make more accurate sorting machines by using the cooperation of the institution in the future. (NAM & SHIN, 2019) In the abstract of the paper "A Study on Methods to Prevent the Spread of COVID-19 Based on Machine Learning" by Youngsang KWAK and Min Soo KANG written in 2020, it is said that currently no specific vaccine or treatment for COVID-19. Therefore, it is said that a self-diagnosis or self-diagnosis questionnaire is essential if possible due to the lack of a safe spread method and vaccine. However, at this time, if there are too many questions or ambiguous questions in self-diagnosis, it takes a lot of time, so it is said that the SVM, Decision Tree and correlation analysis are used to discover two important factors and show an accuracy of 80%. As a result of these studies, it is said that it is effective to check drowsiness, chest pain, and travel history. (KWAK & KANG, 2020) Also in the abstract of the paper "A Study on Methods to Prevent

Pima Indians Diabetes using SVM" by Sanghyuck YOU and Min Soo KANG written in 2020, It is said that the study was conducted. Diabetes causes a variety of problems and affects about 422 million people worldwide, with both cases and prevalence of diabetes increasing steadily. Therefore, through SVM, Decision Tree, and correlation analysis, three important factors for predicting diabetes were found, and the accuracy was 70%. As a result of these studies, doctors can quickly diagnose and prevent potential patients by testing a patient's glucose, BMI, and age. (YOU & KANG, 2020)Therefore, in this paper, correlation analysis was conducted using data from Austin's animal shelters. Based on the factors derived through this correlation analysis, a model was easily created and an experiment was conducted using Azure, which can be easily handled by users. Also, logistic regression analysis is used among Supervised Learning in which the desired result is displayed for each vector. At this time, the logistic regression analysis used in this experiment uses multinomial logistic regression analysis to analyze the outcome types of animals in the animal shelter.

3. Data Set

In this project, two data sets are joined and used. One data is data on the condition of the animal when it arrives at the shelter, and the other is data on the condition when the animal leaves the shelter. These data were collected and used from the open data portal provided by Austin, Texas, USA. The contents of each data set are as follows.

Table 1: Austin Animal Center Intakes.csv Data Description

Variable name	Variable contents
Animal ID	Unique ID is given to each animal
Name	Animal's name
Date Time	Date and time of entry into the animal shelter
Month Year	Year and month of entering the animal shelter
Found Location	Where animals were found
Intake Type	Type of animal shelter inflow
Intake Condition	The condition when entering the animal shelter
Animal Type	Type of animal
Sex upon Intake	Sex when entering animal shelters
Age upon Intake	Age at the time of entry into the animal shelter
Breed	Animal's breed
Color	Animal's color

The Austin Animal Center Intakes data set is one of the datasets applied to the experiment and shows the condition of the animals when they arrive at the animal center. It consists of a total of 12 columns and 63,328 rows. Table1 above is the data structure of the data set.

Table 2: Austin Animal Center Outcomes.csv Data Description

Variable name	Variable contents
Animal ID	Unique ID is given to each animal
Name	Animal's name
Date Time	Date and time of occurrence of the outcome at the animal shelter
Month Year	Year and month of occurrence of results in the animal shelter
Date of Birth	Date of birth
Outcome Type	Result type
Outcome Subtype	Sub result type
Animal Type	Type of animal
Sex upon Outcome	Sex at the time of outcome at the animal shelter
Age upon Outcome	Sex at the time of outcome at the animal shelter
Breed	Animal's breed
Color	Animal's color

Austin Animal Center Outcomes is another dataset that has been applied to the experiment, which shows the condition of animals as they leave the animal center. It consists of a total of 12 columns and 63,643 rows. Table2 above is the data structure of the data set.

4. Experiments and Results

4.1. Experiment Process

First, we went through the process of joining Austin Animal Center Intakes.csv and Austin Animal Center Outcomes.csv to construct the dataset required for the experiment. To this end, each data was pre-processed by deleting the corresponding row when there was a missing value of each data, and Left Outer Join was performed by setting 'Animal ID, Animal Type, Breed, Color' as a join key.

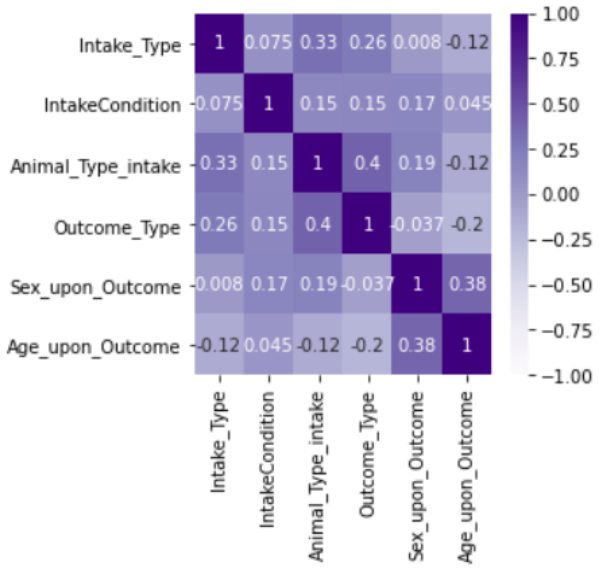

Figure 1: Correlation Analysis Results Using Austin Animal Center Data

Figure 1 is the result of analyzing the correlation through the fields of 'Intake Type, Intake Condition, Animal Type, Sex upon Outcome, Age upon Outcome, Outcome Type' in the dataset created by joining as above. As a result of this correlation analysis, it was found that Animal Type Intake, Intake Type, and Age upon Outcome affect the Outcome Type with correlation coefficients of 0.4, 0.26, and -0.2, respectively. Therefore, the experiment was conducted with three attributes: Animal Type Intake, Intake Type, and Age upon Outcome. At this time, the analysis algorithm used multiple logistic regression analysis to analyze the training data as 0.7 and test data as 0.3. Also, when creating a learning model, it was executed with 'Outcome Type' as a label. Therefore, through experiments like this one can predict the type of animal adoption.

4.2. Experiment Results

As a result of prediction through the experiment, according to Animal Type Intake, the probability of adoption was higher in dogs than cats. Next, according to the Intake Type, the probability of Adoption was high in the Euthanasia Request type. Also, in the case of Public Assist and Stray, it was found that the ratio of Missing was high. Finally, according to Age upon Outcome, the length of stay for cats increased to 12.5 years, whereas for dogs, it can be seen that generally smooth adoption is achieved for all ages.

Table 3: Accuracy Evaluation

Overall accuracy	0.626857
Average accuracy	0.917079
Micro-averaged precision	0.626857
Macro0averaged precision	NaN
Micro-averaged recall	0.626857
Macro-averaged recall	0.254945

Table 3 evaluates the accuracy of the derived model. The overall accuracy is 62.7%, and the average accuracy is 91.7%, showing a fairly reliable result. On the other hand, it can be seen that the accuracy is high only for specific result items, which can be seen in Figure 2.

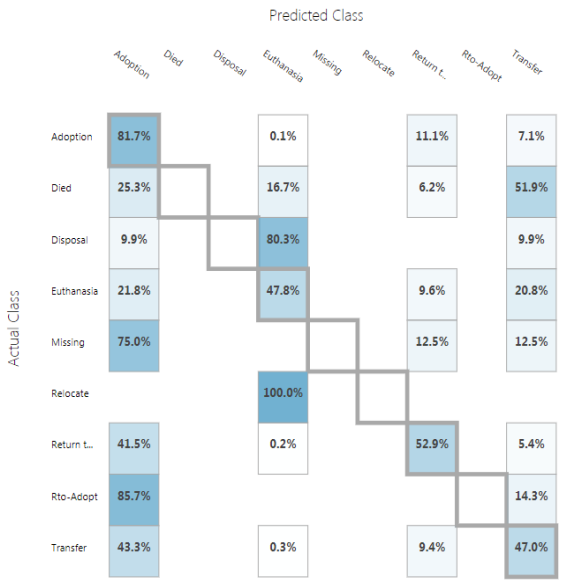

Figure 2: Rate of the evaluation result

Figure 2 shows the results of visualizing the predicted and actual values. When looking at these results, the accuracy is high because it is predicted with an 81.7% probability of adoption. However, other items are also predicted as adoption. Also, as a result of prediction, it is not possible to predict Died, Disposal, Missing, and Relocate items. Also, Euthanasia, Return to Owner, and Transfer are predicted with a probability of about 50%.

5. Conclusions

In this paper, the adoption outcome was predicted based on public data provided by the Austin Animal Shelter in Texas, USA. With the recent increase in single-person households, the number of households with companion animals is also increasing. Accordingly, to solve the problem of commercialization of companion animals, the adoption of companion animals is gaining a lot of popularity with a slogan such as “Do not buy, adopt,”. The problem of animal saturation at animal shelters has been alleviated as influencers also appealed through media such as SNS. However, despite these positive changes, it is still struggling because there are significantly fewer animals to be adopted than animals that are introduced to animal shelters. This also leads to problems such as euthanizing animals to prevent saturation. Also, this euthanasia can be a more serious problem as it is progressing undermining the dignity of animals. To address this, more and more active adoption must be promoted. There should also be opportunities for more animals to go to a good environment.

Therefore, in this paper, we analyzed the factors affecting the adoption of animals. Based on the results of the investigation, it seems that it is possible to come up with a plan to promote the adoption of animals according to various factors. In particular, the results of the experiment showed a significant correlation in Animal Type Intake, Intake Type, and Age upon Outcome, and in particular, the highest correlation was shown in Animal Type Intake attribute. Also, as a result of predicting the outcome of animals in animal shelters based on the significant factors derived through correlation analysis, the probability of prediction was high in 'Adoption'. Therefore, based on these results, we intend to expand to various services by linking with the webserver. For example, this data can be used when creating an animal shelter application or web page to promote safe and fast adoption of animals. However, since the project was to use standardized data excluding unstructured data, the data provided in real-time could not be utilized and various factors could not be actively utilized. Therefore, we will make up for these regrets and improve them so that they can clearly grasp the preferred factors affecting the adoption and use them in various places.

Acknowledgment

This work was supported by the research grant of the KODISA Scholarship Foundation in 2021.

References

Hong, J. Y. (2015). Mobile app design to promote adoption of abandoned animals. Domestic Master's Thesis Ewha Womans University Graduate School of Design, Seoul.

Yu, H. S. (2019). A Study on the Adoption Prediction Model according to the Characteristics of Abandoned and Lost Dogs. Domestic Master's Thesis Graduate School, Dongguk University, Seoul.

Protopopova, A., & Wynne, C. D. L. (2014). Adopter-dog interactions at the shelter: Behavioral and contextual predictors of adoption. *Applied Animal Behaviour Science, 157,* 109-116.

Choi, S. E., Yu, H. S., Jung, H. W., Park, Y. M., & Lee, G. J. (2019). Prediction Model for Adoption Probability of Abandoned or Lost Dogs at Animal Shelters of Local Governments. *Journal of the Korean Data Analysis Society, 21*(5), 2365-2378.

Animal People. (2020). "Pain Death" without anesthesia… Abandoned animal protection system collapses. Retrieved November 12, 2020 from http://www.hani.co.kr/arti/animalpeople/companion_animal/969697.html.

Jang, E. H. (2013). Protection Systems of ‘Living Trash’, Abandoned Animals: A Case of Daegu City. Domestic Master's Thesis Graduate School, Kyungpook National University, Daegu

Posage, J. M., Bartlett, P. C., Thomas, D.K. (1998). Determining factors for successful adoption of dogs from an animal shelter. *J Am Vet Med Assoc, 213*(4), 478-82.

Merry, L. P., Kass, H., & Hart, L. A. (2002). Prediction of Adoption Versus Euthanasia Among Dogs and Cats in a California Animal Shelter*, Journal of Applied Animal Welfare Science*, *5*(1), 29-42.

Kang, M. S., Kang, H. J., Yoo, K. B., Ihm, C. H., & Choi, E. S. (2018). *Getting started with Machine Learning using Azure Machine Learning studio*. Seoul, Korea: Hanti media.

Wikipedia. (2020). Retrieved August 16, 2020 from https://ko.wikipedia.org/wiki/%EC%A7%80%EB%8F%84_%ED%95%99%EC%8A%B5.

Lee, M. W. (2020). Analysis of errors by language area of Korean learners using multinomial logistic regression analysis. *Bilingualism*, *79*, 135-160.

Nam, Y. J., & Shin, W. J. (2019). A Study on Comparison of Lung Cancer Prediction Using Ensemble Machine Learning, *Korean Journal of Artificial Intelligence*, *7*(2), 19-24

Kwak, Y. S., & Kang, M. S. (2020). A Study on Methods to Prevent the Spread of COVID-19 Based on Machine Learning, *Korean Journal of Artificial Intelligence*, *8*(1), 7-9.

You, S. H., & Kang, M. S. (2020). A Study on Methods to Prevent Pima Indians Diabetes using SVM, *Korean Journal of Artificial Intelligence*, *8*(2), 7-10.

Lee, J. G., Cho, I. P., & Lee, S. Y. (2020). A Study on Data Collection Environment and Analysis Using Virtual Server Hosting of Azure Cloud Platform. *Journal of the Korean Society of Computer Information*, *28*(2), 329-330.

Seo, M. J., Yoo, H. J., Jang, & S. W. (2020). Azure cloud-based environment monitoring responsive web app development. *Journal of the Korean Society of Mechanical Engineers*, *1,* 257-257.

Cho, J. J., Jang, M. W., & Beak, S. H. (2020). Research on sign recognition techniques based on Azure IoT Edge. *Journal of the Korean Institute of Communication Sciences, 12,* 828-829.

Yoo, K. Y. (2017). Introduction of the Seoul Companion Animal Center. Policy Report, (222), 1-24.

Moon, Y. J., Lee, E. J., Jang, W. Y., Han, J. Y., & Hong, D. S. (2018). Virtual adoption service experience design for sponsoring abandoned animals. Proceedings of the Korean Society of Design Science326-327.

Park, H. M., & Park, S. Y. (2019). Issues and Alternatives to Companion Animal Policy. *Issue & Diagnosis, 1,* 1-25.

Bae, J. E. & Kim, S. I. (2015). Development of digital contents for organic seedlings. *Digital Design Studies, 15*(1), 155-164.

Choi, E. S., Yoo, H. J., Kang, M. S., & Kim, S. A. (2020). Applying Artificial Intelligence for Diagnostic Classification of Korean Autism Spectrum Disorder, *ORIGINAL ARTICLE*, *17(*11), 1090-1095.

Song, S. W., Kwak, Y. S., & Kang, M. S. (2020). A Study on Graph-based Weighted KNN based on Machine Learning, *Test Engineering & Management*, *83*, 12264-12271.