International Journal of Software Innovation
Volume 11 • Issue 1

A Study on Prediction Performance Measurement of Automated Machine Learning:

Focusing on WiseProphet, a Korean Auto ML Service

Euntack Im, Soongsil University, South Korea
Jina Lee, Soongsil University, South Korea
Sungbyeong An, Soongsil University, South Korea
Gwangyong Gim, Soongsil University, South Korea*

ABSTRACT

In digital economics, where value creation using big data becomes important, the ability to analyze data using machine learning and deep learning technology is a key activity in corporate activities. Nevertheless, companies consider it difficult to introduce machine learning and artificial intelligence technologies because they need an understanding of the business as well as data and analysis algorithms. Accordingly, services such as automated machine learning have emerged for easy use of machine learning. In this study, the authors explored the automated machine learning service and compared the random forest and extreme gradient boosting analysis results using WiseProphet and Python. WiseProphet is used as a representative of automated machine learning solutions because it is a cloud-based service that anyone can easily access and can be used in various ways. It is contrasted with the model implemented by Python, which writes code with No coding. As a result of comparing the prediction performance, WiseProphet automatically outperformed the analysis result by parameter optimization.

KEYWORDS

Automated Machine Learning, Kaggle Data Set, Machine Learning, Wiseprophet, ML Performance Metrics

INTRODUCTION

The Fourth Industrial Revolution, which refers to digital transformation, created a digital economic system in which economic activities were carried out as a major factor in digital input such as digital technology, infrastructure, and data. As the COVID-19 pandemic promoted a non-face-to-face society centered on digital technology, the digital era came as digital technology was brought

DOI: 10.4018/IJSI.315656 *Corresponding Author

into a mainstream way of life, not an instrumental dimension for efficiency. Changes in the current economic environment require the use of strategic digital technology by companies (Jeong, 2019).

Through this, the corporate mindset is changing from product-oriented thinking to solving problems and inconveniences experienced by customers. Key accelerating factors include Connectivity and data collection (Lee et al, 2021). D.N.A (Data, Network, AI) is a key technology in the digital era and is rapidly growing around new businesses of companies using D.N.A. The existing resource and capital-oriented business model is not only changing to a data and artificial intelligence based business model, but also developing differentiated services using machine learning in existing industries such as finance and energy. In addition, D.N.A. is used in corporate activities in various ways such as reducing repetitive tasks or reducing existing costs by using it as a reference for important decisions(Jeong & Kim, 2021).

However, the Korea Institute for Information and Communication Policy suffers from a lack of quality data, difficulty in specifying tasks, understanding AI technology, and lack of internal talent in preparation, model development, and service operation after introduction. In a survey of 152 AI demand and supply companies in Korea, 34% of respondents chose a shortage of internal manpower as an obstacle to the introduction of AI technology (total %: 200%)(Lee&Kim, 2021). Also, Glue Coding, a simple task of putting multiple cords together, accounts for 90% of the total work in the stage of model development using machine learning. The version management of machine learning, the heterogeneity of the development environment, and the actual environment cause inefficiency and causes machine learning project failure(Valohai et al, 2021).

Automated Machine Learning means automating the steps from data preprocessing to model learning to build a model using machine learning. In automatic data processing, feature selection, machine learning algorithm selection, and parameter optimization, unnecessary repetitive tasks in place of human settings and coding can be reduced to enable efficient machine learning utilization projects(Simkek, 2019; Dawid, 2021).

Therefore, the introduction of Automated Machine Learning can contribute to the development of new business models through the universal data utilization of companies, as companies can reduce the probability of failure by reducing inefficiency in machine learning projects.

Thus, this study used python to evaluate the performance of automated machine learning that is being introduced. Random Forest and Extreme Gradient Boosting models were used to predict prediction accuracy. For comparison data, 'the German Credit Prediction' dataset(UCI machine learning, 2016), which is released on the big data platform Kaggle, was used.

THEORETICAL BACKGROUND

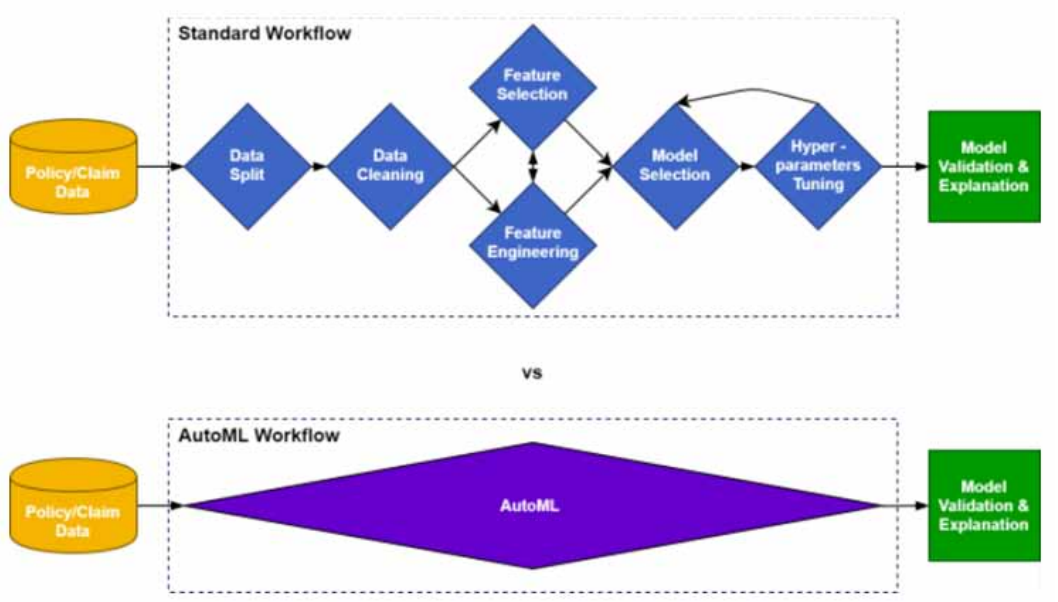
Machine Learning Project and Automated Machine Learning

The Machine Learning project goes through the process of Data collection, Data Cleaning, Feature selection and Engineering, Model selection, Hyperparameters Tuning, and Model validation and Exploration, as shown in Figure 1 (Simsek, 2019).

Data cleaning is not perfect, so machine learning models can learn properly by processing noise present in collected data with errors (Chai, 2020). Feature Selection is a task necessary to eliminate overfitting and shorten learning time from real data mixed with variables necessary and unnecessary to predict dependent variables (Jovic et al., 2015). If Feature selection is to find input variables, Feature Engineering is to transform input variables into ranges and shapes so that modeling can perform better (Zheng & Casari, 2018).

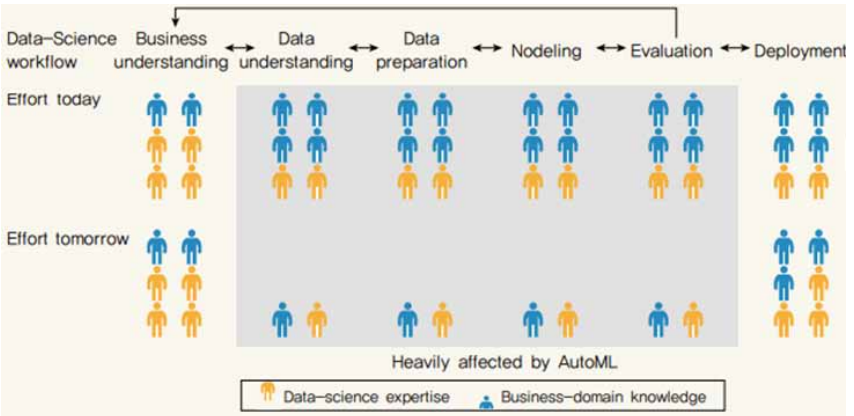
Hyper parameter refers to the value reflected in the model learning process. It refers to values such as a learning rate, a loss function, and a batch size that humans can adjust in advance before starting learning. Since it is a value reflected in the learning process, optimizing the Hyper-parameter has a great influence on machine learning performance. Hence, it is a very important process to explore the values that can improve learning outcomes the most using methods such as Bayesian.

Figure 1. Standard machine learning workflow and AutoML workflow(simsek, 2019)



Thus, for a successful machine learning project, not only domain knowledge but also human resources with an understanding of datasets and variables, statistics, and algorithms are needed. A data science expert with this capability is required, but the supply of manpower is insufficient compared to the demand, causing competition for manpower rotation and recruitment (Lee, 2019). Not only do companies have difficulty carrying out machine learning projects, but they also need to spend a lot of costs and effort to manage their manpower even if they adopt machine learning. However, by automating the processes required for machine learning, the company's universal machine learning project is becoming possible through AutoML which can prevent companies from needing Data science experts. The process is shown in Figure 2(Feuer & Hutter, 2019).

Figure 2. Reducing data science effort by AutoML (Feuer & Hutter, 2019)



Benefits of Automated Machine Learning

Automated machine learning can be defined as a tool that can increase the efficiency of data analysis by automating the stage of machine learning model development. Time-consuming and repetitive tasks can be resolved through automation of the data preprocessing process. It also has the advantage of efficiently preventing potential errors by automating major processes such as hyper-parameter and feature selection (Han, 2020).

The most time-consuming task in data science is cleaning and organizing data such as data labeling, data transformation, missing, and outlier data cleaning (60% of all tasks), but the least playable parts are also cleaning and organizing data (CrowdFlower, 2016). If AutoML is used, it is possible to label and clean each attribute of raw data based on the existing learned information using machine learning. Also, the tasks that humans had to write code can be done easily by simply clicking.

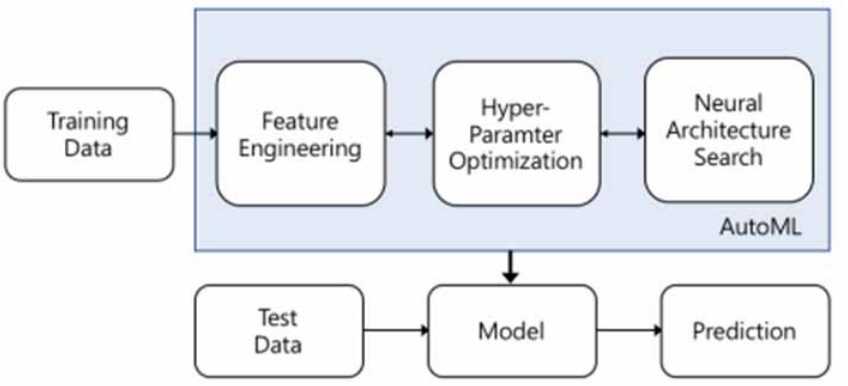
Automated machine learning increases efficiency through automation in the process of affecting machine learning performance, such as feature engineering, hyper-parameter optimization, and architecture search, in the modeling process through machine learning training as well as data management. The process is shown in Figure 3. Feature engineering supports model selective decision making by providing information on the impact on the target to be predicted through machine learning or statistical methods in the process of selecting variables without separate coding tasks. In addition, the model is repeatedly performed, and optimized values such as learning rate and placement that can eliminate overfitting and improve the performance of the model are derived by grid, Bayesian, and gradient methods. Not only that, the model architecture according to the shape of the target variable is recommended, or weights and structures of the model are explored in the course of learning using Evolutionary algorithms and Reinforcement Learning(Mun et al, 2016).

Automated Machine Learning Service

Currently, AWS, Microsoft, and Google, which are based on big data, cloud, and artificial intelligence technologies they own, are representative companies that provide automated services in the development of global Machine Learning and Artificial Intelligence models. Representative characteristics of global leading companies are that they provide computing resources such as GPU and TPU and development frameworks such as Tensorflow and Pythorch based on cloud computing for developers. Not only that, it provides an algorithm-based API held in natural language processing, voice recognition, and image recognition (Yoon, 2021).

Automated machine learning service refers to a service that allows users to conveniently build and use machine learning models.

Figure 3. AutoML main technical trends(Mun et al, 2016)



AWS's SageMaker Autopilot, Microsoft's Azure AutoML, and Google's AI platform provide predictive explanations (feature importance, dashboard, etc.) of machine learning models from data cleansing to regression and classification tasks(Yoon, 2021)., automated in Chapter 2.2. Although machine learning provides functions supported by these companies, the most representative competitiveness of these companies will be global developer infrastructure, data management based on existing technology, application linkage, and open dataset provision and utilization.

In Korea, Naver's Clova supports the use of voice recognition, image recognition, and natural language processing applications based on Naver's artificial intelligence technology. AutoML companies can use machine learning through services that include WiseDQ and WiseIntelligence along with Wiseprophet, which is Wiseitech's autoML service. Through these services, data quality control, visualization, and label-ing services are available(Wiseitech, 2020). In addition, there are AutoML solutions such as AI the.IO and AccuTuning. They support data clearing, feature engineering, parameter optimization, and machine learning model selection.

In this study, automated machine learning solution and python code were written using WiseProphet. Also, the performance of the machine learning model created through this was compared. The purpose is that not only can it be used for corporate businesses, but it also has the advantage that anyone can access and use it easily through the cloud-based web without installation. This is because it contrasts with the Python machine learning model, which uses 'No coding' simply by clicking.

RESEARCH DESIGN

Research Framework

This study compares the performance of Random Forest and Extreme Gradient Boosting machine learning models built with Python code and WiseProphet, an automated machine learning service in Korea. To do this, a dataset(UCI machine learning, 2016) was prepared on the big data platform Kaggle. For an accurate comparison, the same train and test dataset was used. WiseProphet automatically splits data, not allowing users to manually set it. Hence, WiseProphet first completed data preprocing such as missing value processing and data scaling. Thereafter, a data set of 50/50 and 70/30 ratios automatically split by WiseProphet was prepared for each algorithm to generate a Separation variable. After that, the study tried to secure the reliability using the target data labeling used in WiseProphet when build Python code machine learning model.

After completing missing value processing and data scaling as in WiseProphet, the study chose Random Forest and Extreme Gradient Boost, which are representative methods of bagging and boosting methods, as machine learning models for classification.

Machine Learning Algorithm Using in the Research

Random Forest

Random Forest (Breiman, 2019) is a machine learning method in which an individual classification and regression tree (CART) combines several decision trees(Lee et al, 2020). It is one of the representative techniques of bagging (Bootstrap Aggregation), which performs aggregation by resampling multiple data sets through boosting. In building a model, some feature variables and some data of the total data are used (Bühlmann, 2012).

Random Forest resamples the entire data to create a model that considers variables that are not relatively considered to construct a tree. Hence, it has the advantage of effectively reducing the variance with the prediction star by reflecting various variable situations (Lee et al, 2020). Not only is it excellent in creating a general model that can be predicted stably even if actual data is introduced, but it is also effective for missing values because it resamples some of the data through restoration and extraction.

Figure 4. Research framework

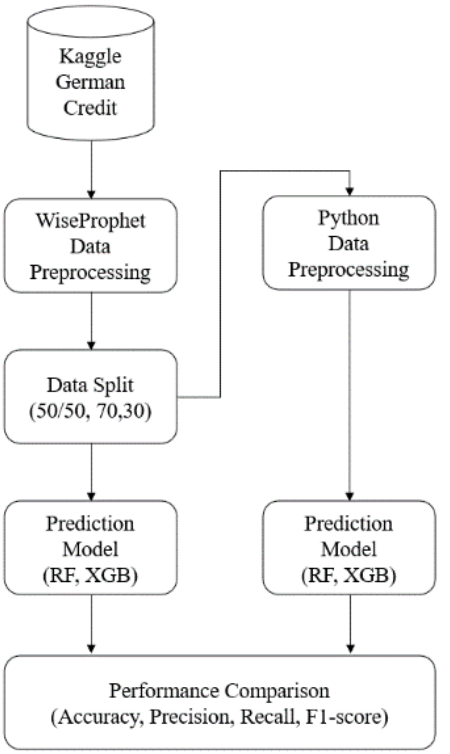
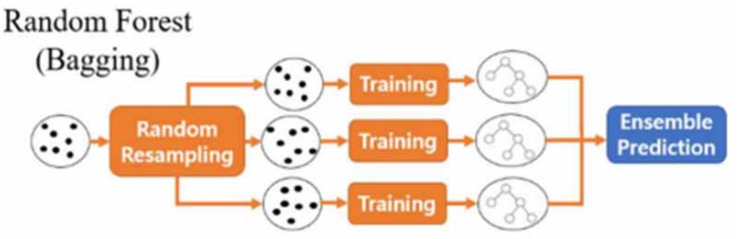


Figure 5. Random Forest model(Lee et al, 2020)



EXTREEM GRADIENT BOOSTING

The extreme gradient boosting (XGB) (Chen et al., 2015) is a representative boosting machine learning model. Bagging has the advantage of creating a general model with less variance through voting. On the other hand, boosting shows excellent performance to solve a specific problem by increasing the weight of the wrong answer (Bühlmann, 2012)

XGB is a sequential machine learning model that weights the residual of the training model and reflects it in the training of the next generated model. While the gradient boosting model (Friedman, 2002) can reduce the vias of the machine learning model, the problem of overfitting that can occur in the learning structure occurs. To solve this problem, XGB tried to overcome the problems of GBM by applying a penalty to the weights (Lee et al., 2020) by reflecting regularization and subsampling.

Figure 6. Gradient Boosting model (Lee et al., 2020)

MACHINE LEARNING PERFOEMANCE

To evaluate the predictive power of machine learning in the binary classification problem, a confusion matrix can be generated to indicate how real positives and negatives are predicted and separated. Also, Accuracy, precision, recall, and F1-score can be measured based on the separated values. The confusion matrix is shown in Table 1.

Accuracy means the degree of classification in real positive and negative among all predictions. Of all predicted cases, it is calculated as the ratio of the number classified in real positive and negative. Precision means calculating the ratio of the actual positive among the cases predicted as positive. On the contrary, Recall can be calculated by calculating the ratio of real positives among the predicted positives. Finally, the F1-score, which is the harmonic mean of Precision and Recall, was developed by considering that the performance of machine learning cannot be properly specified if Accuracy is not uniform in the real positive and negative of the data(Sokolova,2006). Performance measures are shown in Table 2.

Table 1. confusion matrix in the binary classification problem

	Positive predict	**Negative predict**
Real Positive	True positive(tp)	False negative(fn)
Real Negative	False positive(fp)	True negative(tn)

Table 2. machine learning performance measures and calculation

Performance measures	**Calculation**
$accuracy$	$\frac{tp+tn}{tp+fp+fn+tn}$
$Precision$	$\frac{tp}{tp+fp}$
$Recall$	$\frac{tp}{tp+fn}$
$F1-score$	$2\times\frac{Precision\times Recall}{Precision+Recall}$

RESEARCH RESULT

Data set

Kaggle's 'German Credit Prediction' dataset, which was used to compare the performance of machine learning models built with WiseProphet, Automated Machine Learning service in Korea, and Python code, has a total of 1,000 rows of bank loan customer information. Also, a dataset with 10 variables 'Age', 'Sex', 'Job', 'Housing', 'Checking accounts', 'Credit account', 'Duration', 'Purpose', and 'Risk' was used in the study for compare the perfo. The dependent variable to be predicted in the study, 'Risk', is two categorical variables, 'good' and 'bad', which mean an individual's credit status (UCI machine learning, 2016).

Data Preprocessing

Among the 10 variables, 'Saving accounts' and 'Checking accounts' each have a missing value of 183 (18.3%) and 394 (39.4%) of the total 1,000 data. When the missing values of the 'Checking account' are deleted, the data is drastically deleted. Because the data is drastically deleted, the variable was deleted and 18.3% of the missing values of the 'Saving account' are removed.

'Job', which consists of numerical data from 0 to 3, is actually a categorical variable, meaning 'unskilled and non-resident', 'unskilled and resident', 'skilled', and 'highly skilled'. Thus, the data type was changed from numerical to categorical.

Data scaling was performed to reduce the possibility of performance degradation of machine learning that may occur due to differences in the minimum and maximum differences between variables due to different units of continuous variables of 'Age', 'Duration', and 'Credit account'. Min-Max scale was used because all continuous variables were greater than 0.

In the data preprocessing process, WiseProphet, a machine learning analysis tool, was clicked to work. In Python, missing values treatment, variable shape change, and data scaling were treated in the same way through coding.

Data Split

In learning and model validation of predictive models, the ratio of training and verification data is universally 50:50, 70:30, and 80:20. In this study, model learning and verification were conducted by dividing it into training and verification data at a ratio of 50:50 and 70:30. In this process, data split is automatically performed in WiseProphet. Thus, to compare the performance of the machine learning model in Wiseprophet and python, the model was first made in Wiseprophet to use the same data, downloaded the data used at that time, compared it with the entire data, and created a new segregation variable to distinguish it from Python.

Machine Learning Prediction Model

The target looked at Accuracy, Precision, Recall, and F1-score with 'good' meaning good credit status in the target variable 'Risk'. The results can be found in [Table 3].

Accuracy refers to the degree to which the actual 'good' and 'bad' are predicted among the total prediction data. When Random Forest was used, performance evaluation using WiseProphet was 71% and 67% at 50:50 and 70:30, respectively. Python showed similar performance with 68% and 69%. On the XGB, and on the WiseProphet, 71%, and 72% respectively. When data of 71% and 66% of python and 70:30 of python were used, similar performance was shown in addition to the 6% difference present.

The results of Precession, meaning the ratio of predicted 'good' to actual 'good', are as follows. Random forest analysis using WiseProphet showed precession of 71% and 68%. The precession performance of the random forest algorithm using Python was 71% and 70%. This was the prediction precession performance similar to the result using WiseProphet. The prediction performance using

Table 3. prediction performance results

	Random Forest				Extreme Gradient Boosting			
	wise		python		wise		python	
	50	70	50	70	50	70	50	70
Accuracy	0.71	0.67	0.68	0.69	0.71	0.72	0.71	0.66
Precision	0.71	0.68	0.71	0.70	0.74	0.72	0.74	0.71
Recall	0.96	0.96	0.89	0.92	0.88	0.93	0.81	0.81
F1-score	0.82	0.79	0.79	0.79	0.81	0.81	0.77	0.75

the Extreme Gradient Boosting algorithm was 74% and 72% precession when WiseProphet was used, and 74% and 71% when Python was used, showing similar performance.

Among the prediction performance, recall refers to the degree of prediction as 'good' among the actual 'good' as the target. The random forest recall performance performed by WiseProphet showed 96% performance in both the 50:50 and 70:30 datasets. The Ranfom Forest using python was 89% and 92%, respectively, and overall recall performance was lower than when WiseProphet was used. The recall performance of XGB was 88% and 93% when analyzed by WiseProphet, and 88% and 81% in the model using Python.

Finally, the result of F1-score, which can judge the overall machine learning performance with the combination of procession and recall, is as follows. As a result of random forest analysis using WiseProphet, the F1-scores in the 50:50 and 70:30 datasets were 82% and 79%, respectively. When Python was used, f1-socre of 79% and 79% were analyzed, respectively. This shows similar performance in WiseProphet and Python results. Although similar results were obtained, the prediction performance of WiseProphet was better on the 50:50 dataset.

CONCLUSION

Securing competitiveness by utilizing big data has become extremely important in modern businesses. As a result, various industries are striving to utilize machine learning and artistic intelligence. However, using big data utilization technology requires experts who understand all the fields of data management, algorithms, statistics, domains, and computing. However, securing it is one of the difficult tasks.

It is very natural to use automated machine learning, which automatically enables data preprocessing, feature engineering, modeling, and visualization to solve these difficulties. Automated Machine Learning is one of the important solutions that can reduce the probability of failure of a machine learning project.

Thus, this study attempted to compare the performance of automated machine learning, which is increasing in need, with the machine learning model implemented in python and examine its effectiveness. Wiseprophet was used as the automated ma-chine learning solution, and Random forest and Extreme Gradient Boosting, which are models that can represent bagging and boosting among the representative assemble models, were used as algorithms.

As a result of the analysis, Wiseprophet, an automated machine learning, showed similar or excellent performance when compared to python code. Also, the ad-vantage of 'no coding' was that data preprocessing, feature selection, and model were possible with just clicking.

This study not only had academic and practical significance in evaluating the performance of automated machine learning but also left implications for the practical use of automated machine learning. This study has a limitation in evaluating performance only with Random Forest and XGB in classification. Future research should be done to evaluate performance in various algorithms and environments. Also, future research should be conducted from various perspectives including re-sources invested in the analysis.

REFERENCES

Breiman, L. (2001). Random Forest. *Machine learning (Springer)*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. In Handbook of computational statistics. Springer, 985-1022.

Chai, C. P. (2020). The importance of data cleaning: Three visualization examples. *Chance*, *33*(1), 4–9. doi:10.1080/09332480.2020.1726112

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1-4.

CrowdFlower. (2016). *Data Science Report*. CrowdFlower.

Dawid, K. (2021). Automated Machine Learning. *Quantee(online)*, https://www.quantee.ai/resources/automated-machine-learning.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In Automated machine learning. Springer, 3-33

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. doi:10.1016/S0167-9473(01)00065-2

Han, E. Y. (2020). Implications of artificial intelligence technology development for talent nurturing policy: A case of AutoML. Korea Informaion Society Development Institute (KISDI).

Jeong, D.H.(2019). The strategic value of tacit knowledge for digital transformation. *KDB Development Bank Future Strategy Research Institute Issue Analysis*, 766

Jeong, J.H. & Park, S.Y. (2021). Current status and tasks of D.N.A. (Data, Network, AI) policies for the digital age. *National Assembly Research Service*, no. 1828.

Jović, A., Brkić, K., & Bogunović, N. (2015). A review of feature selection methods with applications. *In 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, 1200-1205. IEEE.

Lee, K. M., Hong, C. I., Lee, E. H., & Yang, W. H. (2020). Comparison of Artificial Intelligence Methods for Prediction of Mechanical Properties. *IOP Conference Series. Materials Science and Engineering*, *967*(1), 12–31. doi:10.1088/1757-899X/967/1/012031

Lee, K.S.& Kim, S.O.(2021). Analysis of factors hindering the introduction and spread of AI and policy implications. *KSDI premium report*, *21*(3).

Lee, M. W. (2019). [Above the battle for AI talent] Even in blocking and scouting… Insufficient 10,000 AI talents. [newspaper]. *Asiaeconomy*, *2019*(10), 7.

Lee, W.B., Lee, S.W.& Jeong, J.S.(2021). Business model innovation mechanism according to digital transformation. *Mechanism Journal*, *1(*1), 1-22.

Mun, S. E., Jang, S. B., Lee, J. H., & Lee, J. S. (2016). Technical Trends of Machine Learning and Deep Learning. *Information and Communications Magazine*, *33*(10), 49–56.

Simsek, G. (2019). Introduction to Automated Machine Learning(AutoML). *SOFTWARE ENGINEEING DAILY.* https://softwareengineeringdaily.com/2019/05/15/introduction-to-automated-machine-learning-automl/.

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006, December). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *In Australasian joint conference on artificial intelligence,Springer,Berlin, Heide*, 1015-1021.

UCI machine learning. (2016). German Credit Risk. *Kaggle*. https://www.kaggle.com/datasets/uciml/german-credit

Valohai.(2021). MLOps in Silicon Valley: A Practical MLOps Guide to Building Machine Learning Services. *Valohai.*

Wiseitech. (2020).machine learning solutions of wiseitech. *Wiseitech*. http://wise.co.kr/eng/

Yoon,C.H..(2021). Analysis of global leading companies in artificial intelligence platform. *IT&Futrure Strategy(NIA)*, *5.*

Zheng, A., & Casari, A.(2018). *Feature engineering for machine learning: principles and techniques for data scientists.* O'Reilly Media, Inc.