A Study on the Development of Future Corporate Value Forecasting Classifier Reflecting ESG Information
ESG 정보를 반영한 미래 기업가치 예측 분류기 개발에 관한 연구

Kyung Gu Kang(First Author)
Research Professor, The Institute for Industrial Policy Studies, IPS
(*thomas3877@naver.com*)
Hyung Jong Na(Corresponding Author)
Professor, Department of Accounting and Taxation at Semyung University
(*freshna77@semyung.ac.kr*)
Cheong Yeul Park(Co-Author)
Associate Professor, Seoul Business School, aSSIST University
(*cypark@assist.ac.kr*)

Companies above a certain size that operate globally are showing increasing commitment to ESG (environmental, social, and governance) activities. The main goal of this study is to design a model that can predict future corporate value based on ESG score data. To this end, this study compares the predictions of the basic future corporate value prediction model on which previous studies have been based and those of the future corporate value prediction model proposed herein that includes ESG ratings. For a more rigorous analysis that obtains more comprehensive results, the current study presents results using five machine learning methods: CatBoost, Extra Trees, LGBM, Random Forest, and Gradient Boost. These results indicate that models that encompass ESG data consistently outperform models that do not encompass ESG data in terms of predicting future corporate value. This paper is characterized by its use of an interdisciplinary research methodology that uniquely introduces machine learning techniques, which are rarely used for empirical analysis in the financial and accounting fields. This innovative and future-oriented research method is expected to inspire subsequent scholars in these domains and others in which machine learning techniques are not typically used.

Key Words: Future Corporate Value, Tobin's Q, ESG rating, Machine Learning, Classifier

Submission Date: 08. 23. 2023
Revised Date: (1st: 12. 31. 2023, 2nd: 01. 25. 2024)
Accepted Date: 02. 08. 2024

Ⅰ. Introduction

In the field of economics, businesses have traditionally been conceptualized as companies that should be mainly focused on profit maximization and enhancing their financial worth. However, in recent years, there has been an increasing focus on corporate social responsibility (CSR). Therefore, there is a growing sentiment that businesses—as integral parts of the community—have a responsibility to contribute back to society, along with the idea that such contributions are essential for a business's long-term viability(Clarkson et al., 2019). In particular, the increasing global attention on CSR has led to an increasing emphasis on Environmental, Social, and Governance (ESG) facets(Abdi et al., 2022). These additional dimensions of corporate responsibility focus specifically on environmental conservation and sound governance practices.

When companies issue reports on their sustainable operations, which are intended to provide comprehensive insights into their undertakings, the most important revelations come in the form of their ESG commitments (Abdul and Alsayegh, 2021). A universally accepted and UN-endorsed format for these reports encompasses three core areas: environmental initiatives, societal contributions, and governance practices(Alareeni and Hamdan, 2020). Numerous governmental bodies, financial organizations, and leading businesses are actively creating and disseminating such reports; the current global trend mandates that major publicly-traded enterprises disclose their ESG activities(Aboud and Diab, 2018). In this context, forward-thinking companies are not only generating these sustainable management reports but also proactively engaging in ESG initiatives(Aouadi and Marsat, 2018).

Today's consumers are increasingly attentive to the ESG commitments of the businesses they support(Li et al., 2018). A similar trend is evident among investors, who now tend to consider a company's ESG activities when making investment decisions(Huang, 2021). To thrive in this landscape, it is no longer enough to merely offer high-quality products or services, as companies aiming to bolster their reputation must now prioritize their ESG efforts, enhance their ratings in this arena, and cultivate a positive public image that resonates with both consumers and investors (Feng and Wu, 2021). The global emphasis on ESG assessments and the significant role they play in sustainable business practices has led to intensifying academic research in this area(Arvidsson and Dumay, 2022). Such research has shown that, in the short term, successful companies aim for outstanding performance, while in the long run, their focus shifts toward increasing corporate value. In other words, to achieve continuous success and future value growth, it is important for companies to make ESG-related efforts to manage their image and improve their ESG-related indicators.

The importance of ESG has already been recognized in the finance and accounting research sectors, where a number of studies have examined the relationship between ESG activities and corporate value(Fatemi et al., 2018; Behl et al., 2022; Chouaibi and Chouaibi, 2021). However, most such studies have only used regression models. Due to the nature of regression models, these studies mainly examined which variables have a significant effect on other variables, and there have not been many studies focusing on predictions.

Therefore, our study aims to break away from the research method trend of using only these regression analyses by developing predictive models using machine learning techniques. The introduction of this innovative methodology into the field of finance and accounting research is expected to help advance such research and create opportunities to study new research topics from new perspectives.

There have been a few recent studies in the field of financial research that have developed predictive models using machine learning or deep learning models in addition to regression analysis models. Two such studies developed stock price prediction models(Patel et al., 2015; Chen et al., 2022), two different studies developed corporate default prediction models(Kezelj and Gruenbichler, 2021; Sills et al., 2021), and another study developed a sales prediction model(Ranjitha and Spandana, 2021). However, unlike in the field of financial research, there has yet to be such research in the field of accounting. In particular, there have yet to be any papers describing the development of a future corporate value prediction model. With this background, the current paper aims to conduct a study focused on the development of a future corporate value prediction model that considers ESG ratings.

Corporate data can be largely classified into financial and non-financial data. Both types of corporate data—i.e., financial and non-financial data—are important for grasping information about a company(Kureljusic and Karger, 2023). To date, due to data access constraints, researchers evaluating corporate value have mainly relied on financial data that are readily available in financial statements. However, a more holistic approach that integrates non-financial elements such as ESG is more appropriate for predicting future corporate value, which can determine whether a company will continue to succeed in the long run.

In the present study, we develop a machine learning model that predicts future corporate value by considering the company's ESG rating, which is a representative non-financial indicator. In particular, the goal of this study is to demonstrate that predictions of future corporate values that are made by considering ESG ratings are relatively more accurate than such predictions that are made without considering ESG ratings. The firm value prediction model is designed using financial variables that have been used in previous financial and accounting studies, along with data on ESG ratings, which have not been used as much. The machine learning analysis methodology used in this study is derived from ideas in prior studies in the field of management information systems. Specifically, although the present research focuses on the fields of finance and accounting, the research methodology is inspired by research in management information systems. This interdisciplinary integrated research is expected to allow us to explore new areas of research.

For Korean companies to actively engage in ESG activities and thrive in the global economy, it is necessary for them to engage in transparent and fair ESG evaluations. To this point, companies have evaluated their ESG activities according to their own procedures, including self-ratings, which has not led to ESG evaluations that are consistently fair and objective. To prevent ESG washing, wherein companies promote ESG activities that they are not actually engaging in, companies must be expected to evaluate their ESG activities according to more strict guidelines. The research results of this paper are expected to provide objective and useful information on the importance of corporate ESG evaluation.

The rest of our paper is structured as follows: Section 2 details the theoretical background and the development of our research question on forecasting future corporate value while considering ESG rating. The research methods are described in Section 3, while the results are presented in Section 4. Finally, Section 5 concludes the paper and provides the implications of our research results that can help guide future researchers and business practitioners.

Ⅱ. Theoretical background and research question development

In terms of research scope, most previous studies focused on the association between ESG and firm value using regression analysis alone(Wong et al., 2021; Feng and Wu, 2021 Behl et al., 2022). These studies reported that ESG has a significantly positive impact on future firm value and found some differences in the relationship between ESG and future corporate value depending on factors such as firm size, corporate age, and ESG disclosure level. To more deeply elucidate the relationship between ESG activities and future corporate value, there is a need for research that introduces new methods rather than relying on regression analysis along.

Our study differs from previous studies in the following three ways: First, few studies predict future corporate value using machine learning. Existing papers using machine learning or deep learning methods have mainly focused on predicting stock prices, corporate bankruptcy, and sales (Chen, et al., 2022; Sills et al., 2021; Ranjitha and Spandana, 2021). Second, this study demonstrates that predictions of future corporate value that consider information on ESG ratings are more accurate and meaningful than predictions that do not include ESG information. In other words, the results of this study suggest that it is desirable to consider ESG ratings when predicting future corporate value, as ESG ratings are related to long-term corporate sustainability. Third, this paper considered prior research in the accounting and finance fields when designing the machine learning model and presented the results of an empirical analysis using five classifiers to enhance the reliability of the research results.

2.1 Previous literature related to ESG and future company value

For a company to grow in the long term, it should not only pursue profits—as was the main focus in the past—but also make contributions to society(Abdi et al., 2022). In other words, companies should take the lead in activities promoting environmental protection or social service for sustainable management from a long-term perspective(Aboud and Diab, 2018). Corporations also have a trustee responsibility, which means they should have a transparent governance structure through which they manage shareholder wealth(Huang, 2021). ESG is an indicator that encompasses these concepts, which are necessary for companies to achieve success in modern society(Chouaibi and Chouaibi, 2021).

As corporate ESG activities have come to be considered increasingly important worldwide, there has been a growing body of research examining the relationship between ESG activities and corporate value. Many previous studies have reported that corporate ESG activities improve future corporate value (Arvidsson and Dumay, 2022; Feng and Wu, 2021). Companies need to actively disclose information on ESG activities to inform the market about such activities, which improves the company image and also helps improve future corporate value(Fatemi et al., 2018; Wong et al., 2021). People perceive companies that actively engage in ESG activities as being more ethical, which is expected to improve the corporate value of such companies in the long run(Aouadi and Marsat, 2018). Therefore, companies need to be certified for ESG, and it is important to establish strategies for engaging in ESG activities that companies can follow to sustainably manage their ESG ratings(Alareeni and Hamdan, 2020).

There has also been a number of studies that have examined factors that significantly affect the positive (+) relationship between corporate ESG activities and future corporate value. For example, Abdi et al.(2022) found that ESG has a significant relationship between corporate size and corporate age in the relationship between corporate value. Li et al.(2018) emphasized the important role played by the CEO in increasing corporate ESG activities with the aim of increasing future corporate value. Meanwhile, Chouaibi and Chouaibi (2021) explained that, the more ethical actions that are helpful to society that a company performs, the more its future corporate value increases. In particular, their results emphasized the importance of innovative actions that consider the environment.

Although there has been research examining the relationship between ESG and future corporate value, most of these were empirical analysis studies that used regression analysis models(Feng and Wu, 2021; Aboud and Diab, 2018; Behl et al., 2022), as the research was conducted in the field of finance and accounting. Therefore, the present study aims to conduct an empirical analysis investigating the relationship between ESG and future corporate value using deep learning, an artificial intelligence technique, thus differentiating the current work from previous studies.

2.2 Previous studies on forecasting in accounting and finance fields

There have to this point not been many studies using machine learning and deep learning in the accounting and finance fields. The previous studies using machine learning and deep learning can be largely divided into studies examining stock price prediction, corporate bankruptcy prediction, and sales prediction in the accounting and financial fields.

To begin, the prior studies on stock price prediction using machine learning and deep learning are listed in the following. Chen et al.(2021) used machine learning techniques to develop a stock price prediction model that could be used to construct an optimal portfolio. Jiang et al.(2022) presented a stock price prediction model that was developed using the two-stage machine learning ensemble model. Long et al. (2020) and Yu and Yan(2020) conducted stock price prediction studies using various deep learning techniques. Jing et al.(2021) developed a stock price prediction model using deep learning techniques while considering emotional analysis of investors' opinions. Rezaei et al.(2021) presented a deep learning stock price prediction model that considered the waves of the stock price graph.

Next, prior studies on the prediction of corporate bankruptcy using machine learning and deep learning are as follows. Traczynski(2017) developed a corporate bankruptcy prediction model using the Bayesian model, while Abedin et al.(2020) presented a corporate bankruptcy prediction model that they prepared using machine learning techniques. Kim et al.(2022) used machine learning techniques to develop a corporate bankruptcy prediction model whereas Yu et al.(2022) used machine learning techniques to develop a corporate bankruptcy risk and credit rating prediction model. Mai et al.(2019) used a deep learning method to develop a corporate bankruptcy prediction model.

Lastly, the prior studies on the prediction of sales of companies using machine learning and deep learning are as follows. Tsoumakas (2019) presented a model for predicting sales in the food industry that was developed using machine learning techniques. Wisesa et al. (2020) developed a company's sales prediction model using the gradient boost algorithm. Pavlyschenko(2019) used machine learning techniques to design a model for predicting corporate sales and sales volume. Islam and Amin(2020) included supply chain data in the development of a sales prediction model for which Random Forest and Gradient Boosting were used as machine learning techniques. Lastly, Kilimci et al.(2019) used deep learning techniques to predict corporate sales.

Among the preceding studies using machine learning and deep learning, the research that predicted corporate value is as follows. Lee et al. (2021) developed a machine learning-based corporate value prediction model using online corporate text reviews. Furthermore, Kang et al. (2023) developed a future corporate value prediction model, including ESG evaluation ratings through machine learning regression methods.

Moreover, previous studies on technology valuation and value investment are as follows. Kim et al. (2021) examined deep learning-based intelligent technology valuation. Sung et al. (2021) estimated evaluation variables through deep learning-based technology value evaluation. Choi et al. (2021) conducted a study to estimate sales by evaluating technology value for the marine and fisheries industry through deep learning. Park et al. (2022) investigated an analysis of value investment by industry through a deep learning model.

Meanwhile, research has also been done using deep learning to predict corporate performance, intangible asset value, and asset price. Pechlivanidis et al. (2022) developed a model for predicting the value of intangible assets based on deep learning. Lee et al. (2017) developed a deep learning-based corporate performance prediction model considering technical capabilities. Chen et al. (2023) investigated the asset price prediction model through deep learning analysis.

A study by Kang et al. (2023) is a little similar to the current study, but there are some apparent differences between the two. Both studies investigate the prediction of future corporate value, including ESG ratings, using different machine learning analysis methods. The regression and classification methods are two distinct techniques with different analysis structures and algorithms. This means that this study found ESG rating information to be significant in predicting future corporate value through a different method than Kang et al. (2023). Additionally, this study analyzed a sample period in 2022 to increase objectivity. To derive more sophisticated research results, corporate governance variables such as the largest shareholders' and foreign investors' ratios were included as control variables when designing the future corporate value prediction model.

This study examines how a company's ESG rating can predict its future corporate value in the long term. To make this prediction, the study uses machine learning techniques. The paper focuses on a unique approach by using the classification method of machine learning techniques to create a model for predicting future corporate value. The study uses several machine learning classifiers (CatBoost, XGBoost, LightGBM, Random Forest, Gradient Boost) to obtain objective results. It also provides helpful information on classifiers suitable for predicting future corporate value. Finally, this research aims to offer valuable guidelines for other researchers who perform machine learning analysis using corporate financial data.

2.3 Need for ESG future firm value prediction study

Various scholars have argued that it is time for artificial intelligence research techniques to be actively introduced into research in the accounting and finance fields(Kureljusic and Karger, 2023). Hamid and Habib(2014) argued that research on financial information prediction should be actively conducted using machine learning or deep learning techniques in the fields of accounting and finance. Krylov (2018) explained that research in the accounting and finance fields should be conducted in a more diverse manner using machine learning and deep learning. Polak et al. (2020) suggested that artificial intelligence research methodologies should be introduced in the accounting and finance fields, which are perceived as being conversative in terms of their adoption of new methodologies, to allow for research on new topics that have yet to be studied.

Although there has been research examining the relationship between ESG and future corporate value, most of those studies were empirical analysis studies using regression analysis models, because the research topic was in the field of finance and accounting(Behl et al., 2022; Huang, 2021; Feng and Wu, 2021; Aboud and Diab, 2018). Although some studies have predicted stock prices and corporate bankruptcies using machine learning or deep learning(Jing et al., 2021; Abedin et al., 2020; Tsoumakas, 2019; Wisesa et al., 2020; Pavlyschenko, 2019; Chen et al., 2022), there have been no studies predicting future corporate values. Therefore, the current study aims to use deep learning techniques to develop a model that can predict future corporate value. In particular, this study demonstrates that predictions of future corporate value that include ESG ratings are more accurate method than such predictions made without including ESG ratings.

Research question: Is the model predicting future corporate value including ESG ratings more accurate than model predicting future corporate value without including ESG ratings?

In predicting future corporate value by considering ESG ratings, this study is expected to contribute to expanding the academic scope of related research fields while also providing companies with useful actionable information.

Ⅲ. Data

3.1 Sample

The financial data utilized in this research has been collected from companies that are listed on KOSPI(Korea Composite Stock Price Index) and KOSDAQ(Korea Securities Dealers Automated Quotation) in Korea, with a sample period from 2011 to 2022. To increase the comparability of corporate financial data, since IFRS(International Financial Reporting Standards) were introduced in Korea in earnest from 2011, the first year of the sample period in this study was set to 2011. IFRS is an international standard for preparing financial statements that is maintained by IASB(International Accounting Standards Board) to enable cross-border understanding of financial statements of a company's operating performance and financial status. The financial data of the companies considered in this study were downloaded from and then processed and used through KIS-VALUE provided by NICE Credit Rating Co., Ltd(www.kisvalue.com).

The financial data was excluded from the sample data because the accounting standards differed from those of general companies. Companies that were not settled in December were also excluded from the sample. Moreover, companies for which financial data could not be obtained from KIS-VALUE were excluded from the sample, as were those that did not disclose the number of ESG ratings. In total, 5,625 corporate samples were ultimately used in this study. Information on the sample composition in this study is presented in Table 1.

The ESG rating data utilized in this research spans from 2011 to 2022. This data-

〈Table 1〉 Sample composition of this study

Contents	The number of samples
Number of non-financial companies among KOSPI and KODAQ listed companies between 2011 and 2022	25,853
Excluding the number of companies that are not settled in December	-2,753
Excluding the number of companies without financial data	-1,831
Exclude the number of companies that did not disclose ESG rating	-15,644
Number of Final Samples	5,625

set focuses on companies listed on KOSPI and KOSDAQ that have publicly disclosed their ESG ratings through the Korea ESG Standards Institute(www.cgs.or.kr). The Korea ESG Standards Institute classifies companies' ESG ratings into S grade, A+ grade, A grade, B+ grade, B grade, C+ grade, and D grade every year, but no company has yet been evaluated as S grade. For this research, we have numerically represented the ESG ratings along a score range from 0 to 7, with such a score assigned to each of three categories: environment(E), social activity(S), and governance (G). The aggregate ESG rating was then determined by summing the scores of these three elements. This scoring methodology was used to differentiate between varying ESG evaluation grades, thus positioning them as ordinal variables. In this framework, firm assessments lacking data for any of these three ESG metrics were omitted from our sample.

3.2 Variable definitions and descriptive statistics

Table 2 lists the definitions of the variables used in our analysis. The following variables are selected for the empirical investigation. In this paper, Tobin's Q is used as a measure of future corporate value. Two proxies are used for corporate value(D'Amato and Palivena, 2020). ESG metrics are divided into environmental activity rating(ENV), social responsibility activity rating(SOC), and governance activity rating(GOV). The aggregate score of these three ratings is denoted as ESG(Alareni and Hamdan, 2020).

Moreover, drawing from various studies, we have also included factors that are known to significantly influence corporate value. These are integrated into our foundational machine learning model for predicting future firm value. SIZE represents firm size, drawn from Kalbuana et al.(2021). DEBT represents the debt ratio, drawn from Asim and Ismail(2019). SGR signifies the sales growth rate, drawn from Gu and Kim(2002). ROA indicates return on total assets, drawn from Morgan et al.(2009). OCR indicates the operating cash ratio, drawn from Barua et al.(2010). AGE indicates the age of the company, drawn from

〈Table 2〉 Variable definition

Variables	Definition
TOQ	Firm value measures = Tobin's Q; (the market value of equity + the book value of debt) / the book value of total assets(D'Amato and Falivena, 2020)
ENV	Environmental activity rating(Alareeni and Hamdan, 2020)
SOC	Social Responsibility activity rating(Alareeni and Hamdan, 2020)
GOV	Governance activity rating(Alareeni and Hamdan, 2020)
TESG	Total ESG activity rating of the firm = Environmental activity rating(ENV) + Social Responsibility activity rating(SOC) + Governance activity rating(GOV) (Alareeni and Hamdan, 2020)
SIZE	Firm Size = Natural log value of total assets(Kalbuana et al., 2021)
DEBT	Debt ratio = total liabilities / total assets(Asim and Ismail, 2019)
SGR	Sales growth rate = (sales for the current year - sales for the previous year) / sales for the previous year(Gu and Kim, 2002)
ROA	Return on total assets = net income / total assets(Morgan et al., 2009)
OCR	Operating cash ratio = operating cash flow / total assets(Barua et al., 2010)
AGE	Corporate age = natural log value of corporate age(D'Amato and Falivena, 2020)
LOS	Dummy variable capturing whether current loss is reported = 1 if current loss has occurred, 0 otherwise(Qiu et al., 2021)
LAR	The largest shareholder's share ratio = the number of shares owned by the largest shareholder / the total number of shares owned by the shareholder(Li et al., 2019)
FOR	Foreign investor share ratio = the number of shares owned by foreign investors / the total number of shares owned by the shareholder(Husna and Satria, 2019)

D'Amato and Falivena(2020). Finally, LOS is a dummy variable that indicates if a current loss has been reported, drawn from Qiu et al.(2021).

To consider corporate governance, we established the following variables to predict future corporate value: LAR denotes the largest shareholder's stake, drawn from Li et al.(2019). FOR signifies the foreign investor share ratio, drawn from Husna and Satria(2019).

Table 3 presents the descriptive statistics of the variables used in this research. For TOQ, which is a variable that is intended to capture future corporate value, the average value is 1.342, the standard deviation is 1.079, and the median is 1.017. Among the ESG grades, ENV, which is a variable that capture the environmental grade, the average value is 2.268, the standard deviation is 1.447, and the median is 2.000. Among the ESG grades, SOC, which is a variable that is related to social activities, the average value is 2.597, the standard deviation is 1.462, and the median is 3.000. Among the ESG grades, GOV, which is a variable that means the grade related to governance, the average value is 2.947, 〈Table 3〉 Descriptive statistics (N=5,625)

Variables	N	Mean	Std	Min	Q1	Median	Q3	Max
TOQ	5,625	1.342	1.079	0.240	0.783	1.017	1.481	7.240
ENV	5,625	2.268	1.447	0.000	1.000	2.000	3.000	6.000
SOC	5,625	2.597	1.462	0.000	2.000	3.000	3.000	6.000
GOV	5,625	2.947	0.908	0.000	2.000	3.000	3.000	6.000
TESG	5,625	7.812	3.226	1.000	6.000	8.000	9.000	18.000
SIZE	5,625	26.546	1.467	23.786	25.549	26.343	27.279	30.971
DEBT	5,625	0.354	0.227	0.000	0.172	0.357	0.529	0.872
SGR	5,625	0.051	0.304	-0.748	-0.077	0.022	0.128	1.715
ROA	5,625	0.021	0.088	-0.341	-0.001	0.025	0.060	0.269
OCR	5,625	0.050	0.081	-0.200	0.007	0.046	0.092	0.301
AGE	5,625	3.405	0.669	0.693	2.944	3.584	3.892	4.828
LOS	5,625	0.241	0.428	0.000	0.000	0.000	0.000	1.000
LAR	5,625	0.424	0.167	0.000	0.302	0.427	0.538	1.000
FOR	5,625	0.091	0.125	0.000	0.011	0.040	0.122	0.897

the standard deviation is 0.908, and the median is 3.000. These values are combined in TESG(total ESG ratings), which consists of ENV(a variable representing environmental ratings among ESG ratings), SOC(a variable representing social activity ratings among ESG ratings), and GOV(a variable representing governance ratings among ESG ratings): TESG has an average value of 7.812, a standard deviation of 3.226, and a median of 8.000.

Ⅳ. Methods

4.1 Research model

In the fields of accounting and finance, almost all studies that have studied future corporate value have used regression analysis. By contrast, the present paper establishes a machine learning model while referring to previous studies with the aim of predicting future corporate value.

Above all, company size is the factor that most significantly increases future corporate value(Kalbuana et al., 2021). The larger the firm size, the greater the capital and potential, which makes it highly likely that the corporate value will increase in the future. However, some studies have reported that the debt ratio has a contradictory effect on future corporate value(Asim and Ismail, 2019). A low debt ratio can be interpreted as reflecting a stable financial position and helping improve future corporate value, while a high debt ratio is associated with an increased risk of bankruptcy, but it can also be interpreted as indicative of active investments to improve future corporate value.

Another important factor is sales growth rate: The higher the sales growth rate, the faster the company grows, which is a positive sign for the company's future corporate value (Gu and Kim, 2002). As the current corporate profit ratio increases, the company's financial situation improves and it can use extra funds to reinvest for the future, which has a positive effect on the company's future corporate value(Morgan et al., 2009). If the operating cash flow increases, cash inflows to companies also increase through operating activities, which can naturally be seen as a good sign for future corporate value(Barua et al., 2010). A high corporate age also means that companies have been doing business for a long time, which might suggest that a company has the ability to overcome various risks, thus making it a positive factor for future corporate value (D'Amato and Falivena, 2020).

The stock ratio of the largest shareholder and the corresponding ratio of foreign investors are both representative variables of corporate governance that are also factors that significantly impact future corporate value. First of all, the principal-agent theory posits that a high stake by the largest shareholder increases the ownership awareness of the largest shareholder in the company, which can positively affect future corporate value (Alexander, 2006). By contrast, if the largest shareholder's stake is high, the corporate governance structure may be interpreted as poor, which may negatively affect future corporate value(Li et al., 2019). Meanwhile, a high stake among foreign investors strengthens the monitoring effect of foreign investors and increases corporate accounting transparency, which has a positive effect on future corporate value (Husna and Satria, 2019). Finally, if a company reports losses without reporting profits, it will have a negative impact on future corporate value and seriously damage the sustainability of the company(Qiu et al., 2021). The current study designed a basic future corporate value prediction model using the aforementioned variables that were used in prior research. Information on the financial variables used in the basic model for predicting future corporate value is summarized in Equation (1) below. Moreover, information on the model, including ESG rating information in the basic model for predicting future corporate value, is expressed as a function in Equation (2). For an explanation of each variable, refer to the variable definition presented in Table 2.

Basic Model:

$$F(SIZE, DEBT, SGR, ROA, OCR, AGE, LOS, LAR, FOR) = TOQ \quad \text{Equation (1)}$$

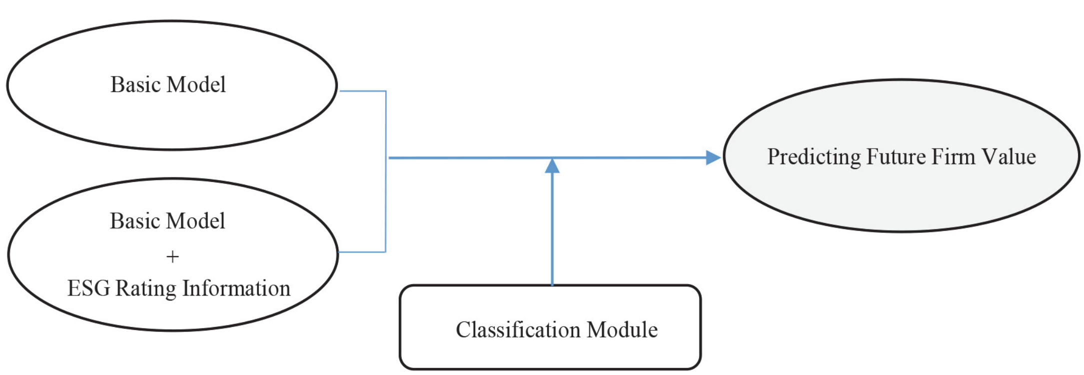
Basic Model + ESG rating information:
F(SIZE, DEBT, SGR, ROA, OCR, AGE, LOS, LAR, FOR, ENV, SOC, GOV, TESG) = TOQ

Equation (2)

The primary aim of this research is to integrate ESG rating details into the foundational prediction framework to ultimately create a model that can be used to forecast future corporate value. Specifically, this research contrasts the efficacy of the foundational corporate value forecasting model, which has been derived from past studies, with a version that encompasses ESG-related insights. Figure 1 shows a visual representation of the research approach.

The detailed description of the machine learning analysis method is as follows. In this study, Python version 3.10.2 was used for machine learning analysis. The data purification process of this study is as follows. First, our researchers confirmed that there is a null value using the is null function, and the null value was removed using the dropna function. In addition, it was visually confirmed that there was an outlier using the boxplot function of the seaborn library. To eliminate outliers, IQR (Q3-Q1) was multiplied by a coefficient of 1.5 to remove the data at the top/bottom, and a total of 5,625 data were used for the analysis. In order to utilize the classification algorithm, it is necessary to change the continuous variable into a categorical variable, and for this purpose, the median value of TQ was calculated using the describe function, and binary classification was performed with a value above the median value as 1 and a value below 0.

The train data and test data were separated using the train_test_split function in the sklearn library, and the entire data was separated by setting the train_size to 0.7 and the test_size to 0.3. In addition, a process of creating dif-



〈Figure 1〉 Research Model

ferent data sets was needed to verify the effectiveness of variables, and train and test data generated through split were generated, respectively. There are a total of five machine learning models used for the analysis, and predictions were extracted using the CatBoost classifier, ExtraTrees classifier, Random Forest classifier, Gradient Boosting classifier, and LGBM classifier functions, respectively. Moreover, to evaluate this, accuracy was evaluated as an accuracy function, precision was a precision function, reproduction was a recall function, and the harmonious average of precision and reproduction was an F1_score function, and the lower area of the ROC curve was evaluated using the AUC function.

4.2 Classification analysis

The machine learning classification module is designed to sort input data into one or multiple distinct classes. Models that are rooted in machine learning for categorization typically use the softmax function in their concluding output layer, as highlighted in research by Zhao et al.(2017). The essence of the softmax function is to render the likelihood of each class, ensuring that their combined probabilities equate to 1, as suggested in research by Zhu et al.(2019).

When making predictions, the module typically opts for the class with the peak probability. Moreover, to address classification challenges, the module leverages the cross-entropy loss function, which gauges the variance between the model's class probability predictions and the actual class labels. It then adjusts the model's weights to diminish this discrepancy (Zhu et al., 2019). The key metrics that serve as benchmarks to assess the efficacy of the classification model are accuracy, precision, recall, AUC(Area Under the ROC Curve), and F1-score, as suggested by Manikandan and Bhuvaneswari (2022).

Accuracy is a metric that is often used in the realm of deep learning evaluations. It denotes how often a model's predictions align with the true class within a classification framework, as explained by Kiran et al.(2020). Accuracy is the quotient of correctly predicted samples to the overall sample count, as detailed by Litjens et al.(2016). A higher quotient suggests that the model exhibits commendable predictive capabilities. Accuracy represents the percentage of samples that the model correctly predicted among the total samples. Accuracy is an indicator of a model's overall performance and is useful in the absence of class imbalances. However, it can be limited when class imbalances exist.

$$\text{Accuracy} = (\text{correct number of predictions}) / (\text{total number of predictions})$$

Precision is another essential metric in machine learning evaluations. Essentially, precision calculates the fraction of samples that are genuinely positive out of those predicted as such, as highlighted by Ghorbani et al. (2020). It gauges the model's aptitude in pinpointing a positive class accurately, as emphasized by Alakus and Turkoglu (2020). While precision values range between 0 and 1, a value nearing 1 indicates the model's adeptness at making positive predictions, as stated by Kiran et al.(2020). However, precision alone might not give a comprehensive view of model performance, as noted by Rai et al.(2023). Researchers are instead encouraged to pair precision with recall for a more rounded assessment, as suggested by Ghorbani et al. (2020). And precision represents the percentage of samples that are actually positive among the positive classes predicted by the model. pPrecision measures how accurately the model predicts the positive class and is important when it is important to reduce the number of false positives.

Precision = (number of positives correctly predicted among actual positives) / (number of positives predicted)

Recall, which is another pivotal metric in machine learning evaluations, computes the fraction of genuinely positive samples among those that the model has deemed positive, as elucidated by Kiran et al.(2020). It assesses the model's proficiency in capturing the positive class without oversight(Rai et al., 2023). Recall values are between 0 and 1, with values closer to 1 signifying the model's precision in positive predictions, as mentioned by Sitaula and Shahi(2022). Recall, which is often synonymous with sensitivity or reproducibility, sheds light on the model's acumen in detecting positive instances, as pointed out by Litjens et al.(2016). Recall represents the percentage of positive samples that the model correctly predicted among the actual positive classes. Recall represents how well we detect real positive samples without missing them, and is important when it is important to reduce the number of false negatives.

Recall = (number of positives correctly predicted among actual positives) / (number of actual positives)

AUC is another significant metric in the realm of machine learning evaluations. AUC gauges the efficacy of a binary classification model, primarily based on how the model discerns between positive and negative classifications, as highlighted by Aggarwal et al. (2021). The ROC curve illustrates the model's True Positive Rate(TPR) juxtaposed with its False Positive Rate(FPR) across varied classification benchmarks, as noted by Gupta et al.(2016). The AUC—or the space below this ROC curve—encapsulates the model's classification prowess. AUC spans between 0 and 1, with values nearer to 1 signifying superior model classification capabilities, as elucidated by Kiran et al.(2020). AUC is a valuable comparative tool for use in multiple binary classification models or in establishing classification parameters for specific models (Hijab et al., 2019). Moreover, in instances of pronounced class imbalances, AUC offers a more unbiased metric than mere accuracy, as explained by Gupta et al.(2016). AUC is a visualized graph of the model's classification performance, with 1-specificity on the x-axis and sensitivity on the y-axis. AUC represents the area under the ROC curve and how well the model has classification performance. Closer to 1 AUC is better for the model, and closer to 0.5 it is similar to random predictions.

F1-score is another pivotal metric in machine learning evaluations. This score, which is articulated as the harmonic mean of precision and recall, offers a holistic view of a model's classification capabilities, as observed by Sadhukhan et al.(2023). As it is derive from both precision and recall, a commendable F1-score indicates high scores in both of these metrics(Wang et al., 2021). Therefore, F1-score is perceived as a metric that harmoniously melds precision and recall (Syed et al., 2021). This score's values oscillate between 0 and 1, with values nearing 1 indicating optimal model classification, according to Christopher et al.(2018). In data sets marked by class imbalances, or in scenarios where it is pivotal to strike a balance between precision and recall, the F1-score emerges as a trusted metric, as underscored by Ghorbani et al.(2020). The F1-score is calculated as a harmonized mean of precision and reproducibility, and takes into account both the precision and reproducibility of the model. The F1-score is useful for severe class imbalances and is an indicator of the balance between precision and reproducibility.

$$\text{F1-score} = 2 \times (\text{precision} \times \text{reproducibility}) / (\text{precision} + \text{reproducibility})$$

These metrics are used to evaluate a model's performance and to select models, tune hyperparameters, or compare them to other models. Which metrics are considered important may depend on the nature of the problem and the business requirements, and it is often important to consider other metrics as well as accuracy alone.

4.3 Machine learning classifier

In this paper, among the various machine learning analysis techniques available, the analysis is performed using five techniques: CatBoost, Extra Trees, LGBM (LightGBM), Random Forest, and Gradient Boosting. The reliability of the research results obtained through this work is improved by presenting the analysis results of various machine learning methods. Each of these classifiers has its strengths and is suited for specific types of tasks and data. The specific choice of classifier often depends on the nature of the data and the problem at hand.

4.3.1 CatBoost

CatBoost is a gradient boosting library that is particularly effective for categorical data (Hancock and Khoshgoftaar, 2020). CatBoost stands for Category Boosting, and it is an open-source library that is especially known for its ability to handle categorical features directly(Wang, 2022; Joo et al., 2023). One of the main advantages of CatBoost is its ability to naturally handle categorical features (Sanjeetha et al., 2021). Most algorithms require explicit preprocessing, such as one-hot encoding or label encoding, to manage categorical data(Luo et al., 2021). CatBoost automates this process, leading to gains in both efficiency and accuracy(Hancock and Khoshgoftaar, 2020). Due to its default settings and the way it is constructed, CatBoost is less prone to overfitting compared to some other algorithms, especially when the dataset is small(Sanjeetha et al., 2021). CatBoost is efficient in terms of both its runtime and memory usage(Luo et al., 2021). Its speed is comparable to—or sometimes even faster than—those of other gradient boosting libraries, particularly when working with categorical features(Sanjeetha et al., 2021). CatBoost provides built-in tools to visualize the training process, relative feature importance, and other essential insights that are valuable during model development(Wang, 2022). CatBoost can handle missing data without requiring any explicit imputation(Luo et al., 2021; Chelgani et al., 2023). CatBoost is particularly useful for use with datasets that contain many categorical features or when aiming to avoid the need for a tedious preprocessing step for those features(Wei et al., 2023). In conclusion, CatBoost is a powerful and efficient gradient boosting library that is optimized for categorical features. For many applications, it can simplify the modeling process while providing performance that is competitive with or even superior to those of traditional models.

4.3.2 Extra Trees

Extra Trees, which stands for Extremely Randomized Trees, is an ensemble learning method that is fundamentally similar to Random Forest(Saeed et al., 2021; Wahid et al., 2023). It is designed to fit a number of randomized decision trees on various subsamples of the dataset(Alsariera et al., 2020). However, there are differences in the way that splits in these trees are made, and these differences are what give Extra Trees its extremely randomized nature(Abbas et al., 2021). To understand Extra Trees, it is useful to start by considering the basic idea behind decision trees. A decision tree splits the data into subsets based on feature thresholds. These splits are made while aiming to improve the purity of the data in each subset, by aiming to segregate different classes or reducing regression error(Eslami et al., 2020). Extra Trees adds another layer of randomness as follows: Instead of computing the best threshold for each feature to split on, a random threshold for each feature is chosen, and the best of these random splits is used(Manavalan et al., 2019). This makes Extra Trees extremely randomized, as both the feature and the threshold for the split are selected randomly(Saeed et al., 2021). Due to this extra randomness, Extra Trees can sometimes produce trees that are more independent of each other than those produced by Random Forest, which can reduce variance(Abbas et al., 2021; El Bilali et al., 2023). Since we are picking random thresholds instead of searching for the best ones, Extra Trees can be faster to train than Random Forest(Alsariera et al., 2020). The additional randomness can also make Extra Trees less likely to overfit to noise in the data(Eslami et al., 2020). However, this is a trade-off, as this can sometimes also lead to slightly increased bias(Saeed et al., 2021; Wang et al., 2023).

4.3.3 LGBM (LightGBM)

LGBM, or LightGBM, stands for Light Gradient Boosting Machine(Sharma and Singh, 2020). It is an open-source, distributed, high-performance implementation of the gradient boosting framework that is specifically designed for speed and efficiency(Csizmadia et al., 2022). Developed by Microsoft, LightGBM is particularly popular because of its efficiency at scale and the fact that it can handle large datasets without using excessive memory (Gong and Liu, 2022; Mishra and Paliwal, 2023; Chen et al., 2023; Xi et al., 2023). Its histogram-based algorithm speeds up the training process and reduces memory usage (Massaoudi et al., 2021). While many gradient boosting algorithms struggle with memory and speed issues as the dataset size grows(Csizmadia et al., 2022), LightGBM is designed to scale efficiently, thus making it particularly suitable for large datasets that other algorithms cannot efficiently handle. LightGBM can handle categorical features directly without the need for manual one-hot encoding(Chen et al., 2023). The framework supports distributed training, thus allowing it to handle even bigger datasets(Gong and Liu, 2022). It also supports GPU acceleration, which can further speed up the training process(Gong and Liu, 2022). Unlike other boosting algorithms that grow trees either depth-wise or level-wise, LightGBM uses a best-first approach(Massaoudi et al., 2021). Specifically, it chooses the leaf with the maximum delta loss to grow, which can lead to deeper trees, but this also tends to lead to more errors(Csizmadia et al., 2022). While LightGBM is powerful and efficient, it is essential to tune its hyperparameters for optimal performance, just like any other machine learning algorithm(Sharma and Singh, 2020; Wang et al., 2023).

4.3.4 Random Forest

Random Forest is a popular and versatile machine learning algorithm that can be used for both classification and regression tasks (Speiser et al., 2019). As an ensemble learning method, it combines multiple algorithms to obtain better predictive performance than could be obtained from any of the individual algorithms alone(Sheykhmousa et al., 2020). Random Forest builds multiple decision trees and merges their outputs(Tyralis et al., 2019). For each tree, a random sample of the data is drawn with replacement; this process is known as bootstrapping(Iwendi et al., 2020). This means that some data points might be sampled multiple times whereas others might not be sampled at all. At each split in the decision tree, instead of finding the best split among all features, Random Forest picks a random subset of features and then finds the best split among those(Chen, et al., 2021). This introduces more diversity among the trees and helps prevent overfitting. For classification tasks, each tree in the forest votes for a class, and the class with the most votes is the Random Forest's final prediction(Tyralis et al., 2019). This culminates in a model that, thanks to the wisdom of the crowd principle, is typically more accurate and less prone to overfitting than an individual decision tree(Speiser et al., 2019). Random Forest has the following advantages: Random Forest is often more accurate than individual trees. By averaging out multiple trees, it tends to avoid overfitting, which can be a problem with individual decision trees(Tyralis et al., 2019). Random Forest can handle missing values without requiring imputation, and it also provides insights into which features are important in making predictions(Speiser et al., 2019; Wang et al., 2023). However, a Random Forest model can become complex, especially with a large number of trees(Iwendi et al., 2020). This can make the model slower for predictions than some other algorithms. In conclusion, Random Forest is a robust and versatile algorithm that offers high accuracy across a wide range of tasks (Iwendi et al., 2020; Nhat-Duc and Van-Duc, 2023). Its ensemble nature, which aggregates multiple decision trees, helps it deliver a balanced trade-off between bias and variance, ultimately making it one of the go-to algorithms for many machine learning practitioners (Chen, et al., 2021).

4.3.5 Gradient Boosting

Gradient Boosting is a machine learning

technique that can be used for both regression and classification problems(Bentéjac et al., 2021; Zhang et al., 2023). It is an ensemble method, which means it combines the predictions of several models to improve accuracy and reduce overfitting(Duan et al., 2020). Gradient boosting involves building a series of weak learners—typically decision trees—in a sequential manner, where each tree tries to correct the errors of its predecessor(Yoon, 2021; Louk et al., 2023). Gradient Boosting could be the mean of the target variable for regression tasks or the log-odds for a classification task(Taha and Malebary, 2020). Unlike Random Forest, which builds trees in parallel, Gradient Boosting builds trees one at a time(Nasiboglu and Nasibov, 2023). Each new tree is fit to either the negative gradient or the residual errors of the combined ensemble of existing trees(Bentéjac et al., 2021; Velthoen et al., 2023). After each tree is built, the predictions are updated to incorporate the new tree's output(Duan et al., 2020). This is done by adding a fraction learning rate of the new tree's predictions to the ensemble's accumulated predictions. Regularization techniques like tree pruning, learning rate shrinkage, and the addition of randomization can be applied to reduce overfitting(Yoon, 2021). Several steps are repeated until a predetermined number of trees are constructed, or until the model's performance ceases to improve on an out-of-sample test set(Bentéjac et al., 2021). Gradient Boosting often provides high predictive accuracy that can rarely be beat by other algorithms(Duan et al., 2020). It can be used for both regression and classification tasks, and it is amenable to handling different types of predictor variables(Duan et al., 2020). Like most tree-based algorithms, gradient boosting provides insights into feature importances, which can be a valuable tool for feature selection and elucidating a model; however, due to its sequential nature, it can also be computationally expensive(Bentéjac et al., 2021). In conclusion, Gradient Boosting is a powerful ensemble technique that can consistently provide high predictive accuracy across a range of applications. However, to maximize its benefits, practitioners need to understand its hyperparameters and the trade-offs involved in tuning them(Yoon, 2021; Douiba et al., 2023).

The advantages, disadvantages, and differences of various machine learning techniques (CatBoost, XGBoost, LightGBM, Random Forest, and GradientBoost) that may occur when analyzing a company's financial data are described in detail as follows. First, since CatBoost is specialized in categorical data processing, CatBoost provides the ability to automatically process and encode categorical variables. And since CatBoost has a function to prevent overfitting, generalization ability can be improved through the built-in overfitting prevention function. In addition, parameter tuning may not be required because CatBoost automatically tunes parameters to provide good performance even with the default setting. However, CatBoost can have longer processing time for large financial data. In other words, CatBoost can take some time to learn and predict for large financial data sets.

Second, XGBoost (extreme gradient boosting) is a model with excellent predictive performance and can show high accuracy in research using financial data. XGBoost can handle all different financial data types. In other words, it can handle both categorical and numerical variables. And XGBoost has fast learning and prediction speed in terms of speed and efficiency. However, unlike CatBoost, XGBoost may require parameter tuning for optimal performance and sometimes it can be difficult to process categorical data.

Third, LightGBM (Light Gradient Boosting Machine) is effective in processing large amounts of data because it has fast learning and prediction speed when analyzing corporate financial data. In addition, LightGBM can be used even in small memory and shows excellent performance when processing large datasets. And LightGBM is optimized for categorical data processing. However, caution is required because overfitting problems may occur when the amount of financial data in the business is small, and parameter tuning may be required, such as XGBoost.

Fourth, Random Forest provides excellent predictive performance for various data types. Random forest reduces overfitting because it averages several trees to prevent overfitting. And Random Forest calculates the importance of characteristics that tell which characteristics of financial data are important for prediction. However, processing large amounts of data can be difficult when analyzing a company's financial data. Processing of large financial data sets can be slow and memory usage can be high. And parameter tuning may be required, such as LightGBM and XGBoost.

Fifth, Gradient Boost can show relatively higher accuracy in the analysis of corporate financial data. This is because Gradient Boost improves predictive performance with an ensemble effect. And Gradient Boost can handle both categorical and numerical data among financial data. However, because Gradient Boost learns trees sequentially, it may be slow to learn large amounts of financial data and may require parameter tuning such as Random Forest and LightGBM and XGBoost.

Each machine learning technique may be suitable for a different situation in the study of a company's financial data, and should be selected according to the characteristics of the data and the research objectives. Therefore, in this study, the results of machine learning analysis are presented through five classes.

Ⅴ. Results

Taken together, the findings of the machine learning classification analysis in this study demonstrate that future corporate value prediction models that consider ESG ratings are superior to such models that do not include ESG ratings. In particular, to improve the reliability and objectivity of the research results, this paper presents the results of five machine learning classifiers (CatBoost, Extra Trees, LGBM, Random Forest, and Gradient Boosting).

Table 4 presents the accuracy results obtained by each classifier and the accuracy differences between models. When predicting future firm value using CatBoost, the accuracy value by the basic model is 0.8047. On the other hand, when the basic model includes grade information on the environment, the accuracy value is 0.8991, it is 0.8998 when the ESG rating information includes social activity, and it is 0.8991 when total ESG rating information is included.

When forecasting future corporate value using Extra Trees, the accuracy value by the basic model is 0.8907. However, when the basic model includes grade information on the environment, the accuracy value is 0.8944, it is 0.8957 when the ESG rating information includes social activity, it is 0.8967 when the ESG rating information includes governance, and it is 0.8957 when total ESG rating information is included.

When predicting future company value using LGBM, the accuracy value obtained by the basic model is 0.8920. When the basic model includes grade information on the environment, the accuracy value is 0.8931, it is 0.8948 when the ESG rating information includes social activity, it is 0.8931 when the ESG rating information includes governance, and it is 0.8940 when total ESG rating information is included.

When forecasting future corporate value using Random Forest, the accuracy value obtained by the basic model is 0.8839. Meanwhile, when the basic model includes grade information on the environment, the accuracy value is

〈Table 4〉 Tobin's Q Accuracy results

Model	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy
Basic model	CatBoost	0.8047	Extra Trees	0.8907	LGBM	0.8920	Random Forest	0.8839	Gradient Boosting	0.8951
ENV		0.8991		0.8944		0.8931		0.8965		0.8894
SOC		0.8998		0.8957		0.8948		0.8925		0.8953
GOV		0.8945		0.8967		0.8931		0.8940		0.8894
TESG		0.8991		0.8957		0.8940		0.8946		0.8967

0.8965, and it is 0.8946 when total ESG rating information is included.

When predicting future company value using Gradient Boosting, the accuracy value obtained by the basic model is 0.8951. However, when the basic model includes grade information on the environment, the accuracy value is 0.8894, it is 0.8953 when the ESG rating information includes social activity, it is 0.8894 when the ESG rating information includes governance, and it is 0.8967 when total ESG rating information is included.

Table 5 shows the precision results obtained by each classifier and compares precision differences between models. When predicting future firm value using CatBoost, the precision value obtained by the basic model is 0.8009. When the basic model includes grade information on the environment, the precision value is 0.8213, it is 0.8208 when the ESG rating information includes social activity, and it is 0.8229 when total ESG rating information is included.

When forecasting future corporate value using Extra Trees, the precision value obtained by the basic model is 0.8310. Meanwhile, when the basic model includes grade information on the environment, the precision value is 0.8264, it is 0.8323 when the ESG rating information includes social activity, it is 0.8341 when the ESG rating information includes by governance, and it is 0.8299 when total ESG rating information is included.

When predicting future company value using LGBM, the precision value obtained by the basic model is 0.8002. However, when the basic model includes grade information on the environment, the precision value is 0.8047, it is 0.8093 when the ESG rating information includes by social activity, it is 0.8034 when the ESG rating information includes by governance, and it is 0.8006 when total ESG rating information is included.

When forecasting future corporate value using Random Forest, the precision value obtained by the basic model is 0.8201. When the basic model includes grade information on the environment, the precision value is 0.8287, and it is 0.8291 when the ESG rating information includes social activity.

〈Table 5〉 Tobin's Q Precision results

Model	Classifier	Precision	Classifier	Precision	Classifier	Precision	Classifier	Precision	Classifier	Precision
Basic model	CatBoost	0.8009	Extra Trees	0.8310	LGBM	0.8002	Random Forest	0.8201	Gradient Boosting	0.8014
ENV		0.8213		0.8264		0.8047		0.8287		0.8033
SOC		0.8208		0.8323		0.8093		0.8291		0.8085
GOV		0.8010		0.8341		0.8034		0.8202		0.8017
TESG		0.8229		0.8299		0.8006		0.8114		0.8057

When predicting future company value using Gradient Boosting, the precision value obtained by the basic model is 0.8014. However, when the basic model includes grade information on the environment, the precision value is 0.8033, it is 0.8085 when the ESG rating information includes by social activity, it is 0.8017 when the ESG rating information includes governance, and it is 0.8057 when total ESG rating information is included.

Table 6 lists the recall results obtained by each classifier and the recall differences between models. When predicting future company value using CatBoost, the recall value obtained by the basic model is 0.7403. On the other hand, when the basic model includes grade information on the environment, the recall value is 0.7471, and it is 0.7428 when the ESG rating information includes social activity.

When forecasting future corporate value using Extra Trees, the recall value obtained by the basic model is 0.7138. Meanwhile, when the basic model includes grade information on the environment, the recall value is 0.7218, it is 0.7225 when the ESG rating information includes social activity, and it is 0.7252 when total ESG rating information is included.

When predicting future company value using LGBM, the recall value obtained by the basic model is 0.7229. When the basic model includes grade information on the environment, the recall value is 0.7362, it is 0.7393 when the ESG rating information includes social activity, it is 0.7289 when the ESG rating information includes governance, and it is 0.7271 when total ESG rating information is included.

When forecasting future corporate value using Random Forest, the recall value obtained by the basic model is 0.7101. Meanwhile, when the basic model includes grade information on the environment, the recall value is 0.7209, and it is 0.7269 when the ESG rating information includes social activity.

When predicting future company value using Gradient Boosting, the recall value obtained by the basic model is 0.7071. When the basic model includes grade information on the environment, the recall value is 0.7151, it is

〈Table 6〉 Tobin's Q Recall results

Model	Classifier	Recall	Classifier	Recall	Classifier	Recall	Classifier	Recall	Classifier	Recall
Basic model	CatBoost	0.7403	Extra Trees	0.7138	LGBM	0.7229	Random Forest	0.7101	Gradient Boosting	0.7071
ENV		**0.7471**		**0.7218**		**0.7362**		**0.7209**		**0.7151**
SOC		**0.7428**		**0.7225**		**0.7393**		**0.7103**		**0.7162**
GOV		0.7369		**0.7237**		**0.7289**		**0.7092**		**0.7146**
TESG		0.7345		**0.7252**		**0.7271**		**0.7269**		**0.7169**

0.7162 when the ESG rating information includes social activity, it is 0.7146 when the ESG rating information includes governance, and it is 0.7169 when total ESG rating information is included.

Table 7 presents AUC results by each classifier and AUC differences between models. When predicting future company value using CatBoost, the AUC value obtained by the basic model is 0.8105. Meanwhile, when the basic model includes grade information on the environment, the AUC value is 0.8213, it is 0.8208 when the ESG rating information includes social activity, it is 0.8201 when the ESG rating information includes governance, and it is 0.8229 when total ESG rating information is included.

When forecasting future corporate value using Extra Trees, the AUC value obtained by the basic model is 0.8170. When the basic model includes grade information on the social activity, the AUC value is 0.8323, it is 0.8341 when the ESG rating information includes governance, and it is 0.8399 when total ESG rating information is included.

When predicting future company value using LGBM, the AUC value obtained by the basic model is 0.8086. However, when the basic model includes grade information on the environment, the AUC value is 0.8087, it is 0.8093 when the ESG rating information includes social activity, it is 0.8088 when the ESG rating information includes governance, and it is 0.8092 when total ESG rating information is included.

When forecasting future corporate value using Random Forest, the AUC value obtained by the basic model is 0.8160. Meanwhile, when the basic model includes grade information on the environment, the AUC value is 0.8187, it is 0.8191 when the ESG rating information includes social activity, and it is 0.8183 when total ESG rating information is included.

When predicting future company value using Gradient Boosting, the AUC value obtained by the basic model is 0.8033. When the basic model includes grade information on the environment, the AUC value is 0.8044, it is 0.8085 when the ESG rating information includes social activity, it is 0.8057 when the

〈Table 7〉 Tobin's Q AUC results

Model	Classifier	AUC	Classifier	AUC	Classifier	AUC	Classifier	AUC	Classifier	AUC
Basic model	CatBoost	0.8105	Extra Trees	0.8170	LGBM	0.8086	Random Forest	0.8160	Gradient Boosting	0.8033
ENV		**0.8213**		**0.8264**		**0.8087**		**0.8187**		**0.8044**
SOC		**0.8208**		**0.8323**		**0.8093**		**0.8191**		**0.8085**
GOV		**0.8201**		**0.8341**		**0.8088**		0.8072		**0.8057**
TESG		**0.8229**		**0.8399**		**0.8092**		**0.8183**		**0.8057**

ESG rating information includes governance, and it is 0.8057 when total ESG rating information is included.

Table 8 lists the F1-Score results by each classifier and the F1-Score differences between models. When predicting future company value using CatBoost, the F1-Score value obtained by the basic model is 0.7628. Meanwhile, when the basic model includes grade information on the environment, the F1-Score value is 0.7694, it is 0.7735 when the ESG rating information includes social activity, it is 0.7689 when the ESG rating information includes governance, and it is 0.7687 when total ESG rating information is included.

When forecasting future corporate value using Extra Trees, the F1-Score value obtained by the basic model is 0.7350. When the basic model includes grade information on the environment, the F1-Score value is 0.7423, it is 0.7442 when the ESG rating information includes social activity, it is 0.7474 when the ESG rating information includes governance, and it is 0.7460 when total ESG rating information is included.

When predicting future company value using LGBM, the F1-Score value obtained by the basic model is 0.7519. However, when the basic model includes grade information on the environment, the F1-Score value is 0.7564, it is 0.7601 when the ESG rating information includes social activity, it is 0.7580 when the ESG rating information includes governance, and it is 0.7557 when total ESG rating information is included.

When forecasting future corporate value using Random Forest, the F1-Score value obtained by the basic model is 0.7545. Meanwhile, when the basic model includes grade information on the environment, the F1-Score value is 0.7556, and it is 0.7556 when total ESG rating information is included.

When predicting future company value using Gradient Boosting, the F1-Score value obtained by the basic model is 0.7413. When the basic model includes grade information on the environment, the F1-Score value is 0.7492, it is 0.7419 when the ESG rating information includes social activity, it is 0.7489 when the ESG rating information in-

〈Table 8〉 Tobin's Q F1-Score results

Model	Classifier	F1-Score	Classifier	F1-Score	Classifier	F1-Score	Classifier	F1-Score	Classifier	F1-Score
Basic model	CatBoost	0.7628	Extra Trees	0.7350	LGBM	0.7519	Random Forest	0.7545	Gradiint Boosting	0.7413
ENV		0.7694		0.7423		0.7564		0.7556		0.7492
SOC		0.7735		0.7442		0.7601		0.7503		0.7419
GOV		0.7689		0.7474		0.7580		0.7475		0.7489
TESG		0.7687		0.7460		0.7557		0.7556		0.7418

cludes governance, and it is 0.7418 when total ESG rating information is included.

As a result of forecasting future firm value, in terms of accuracy, we found that prediction performance was relatively better when forecasted including environmental and total grades among ESG ratings. And in terms of accuracy in forecasting future corporate value, Extra Trees, LGBM, and Gradint Boost have higher accuracy than CatBoost and Random Forest. In aspect of precision, the results presented that prediction performance was relatively better when forecasted, including environmental and social activity ratings among ESG ratings. And in aspect of precision in forecasting future corporate value, Extra Trees, LGBM, and Gradint Boosting outperformed CatBoost and Random Forest. In terms of recall, we found that prediction performance was relatively better when forecasted including environmental ratings among ESG ratings, whereas prediction performance did not improve significantly when forecasted including governance ratings. And in terms of recall in future corporate value forecasts, LGBM and Gradint Boost have higher accuracy than Extra Trees, CatBoost, and Random Forest. In aspect of AUC, the results indicated that prediction performance was relatively better when forecasted, including environmental, social activity, and total ratings among ESG ratings. And in aspect of AUC in forecasting future corporate value, CatBoost, LGBM, and Gradint Boosting have higher accuracy in their forecasts than Extra Trees and Random Forest. In terms of F1-Score, we found that prediction performance was relatively better when forecasted, including environmental and total ratings, among ESG ratings. And in terms of F1-Score forecasting future corporate value, CatBoost, Extra Trees, LGBM, and Gradeint Boosting were more accurate than Random Forest. The research results of this paper show that predicting future corporate value, including information on environmental ratings among ESG ratings, helps improve predictive performance.

Ⅵ. Discussion

6.1 Main findings

With the ultimate aim of obtaining more reliable and objective results, the present study used five types of machine learning classifiers (CatBoost, Extra Trees, LGBM, Random Forest, and Gradient Boosting) and compared the predictions made by a basic model for future firm value forecasting with those predicted by adding ESG ratings to the basic model. This research also included a classification module to demonstrate the excellence of the model that predicts future corporate value while including ESG information.

This paper is valuable as a convergence paper that uses corporate financial data to predict future corporate value by machine learning analysis. In particular, it has proven to perform better when predicting future corporate value, including the company's ESG rating information. In this study, a classification module was used instead of a regression module similar to the existing regression analysis among machine learning techniques, and five classes (CatBoost, Extrudes, LGBM, Random Forest, and Gradient Boost) were used to present objective and reliable analysis results.

The difference between the machine learning analysis method using financial data and the analysis through the existing regression analysis is as follows. Analyzing financial data using a machine learning method is similar to analyzing other data using machine learning. You can use the form of the company's financial variables that were previously used for regression analysis. For example, the company size variable uses the natural logarithm of total assets, and the debt ratio uses the variable generated by dividing total liabilities by total assets.

However, there is a difference in how many classes the parameter value is set, how many classes the target variable should be divided for classification analysis, and how much ratio the training set and the test set are set. If machine learning classification analysis is used, the performance of existing and newly developed machine learning models is compared using accuracy, precision, recall, F1-score, and AUC values as the result values. Furthermore, because machine learning analysis is not a statistical analysis, there is no significance level, and it is judged only by the size of the result value.

Summarizing the analysis results of the machine learning regression module, we found that the model in which ESG rating information was added performed better than the model that did not reflect ESG rating information in terms of predicting future corporate value. In this study, the analysis results of five machine learning classifiers(CatBoost, Extra Trees, LGBM, Random Forest, and Gradient Boosting) consistently confirmed that future corporate value prediction models that include ESG ratings make superior predictions to future corporate value prediction models that do not include ESG ratings. These results mean that, in predicting future corporate value, making predictions while including information about ESG ratings helps predict future corporate value more accurately. In other words, the results of this paper suggest that information on corporate ESG is an essential factor to consider when predicting corporate value from a long-term perspective, as it is perceived by many as meaningful information that is related to corporate value.

6.2 Theoretical implications

This study aims to offer several academic contributions. First, the research approach adopted here may help develop a novel method within the domain of future firm valuation studies. While numerous papers have delved into future company valuation within accounting and finance, most such studies have used regression analysis techniques. Therefore, this study's approach might offer valuable insights for future scholars aiming to delve into future corporate value predictions.

Secondly, this study aims to broaden the horizons of ESG research. At this juncture, when global corporate ESG initiatives are gaining momentum in terms of their relation to a company's sustained growth, this study showcases analytical outcomes that underline ESG's significance through advanced machine learning methods. The techniques and findings in this paper are poised to enrich the breadth of research related to ESG.

Lastly, in the realm of future firm valuation studies, particularly within the accounting and finance sectors, this research pioneers the application of machine learning methods for predictions. This paper stands out as an interdisciplinary endeavor, as it ingeniously infuses machine learning techniques into empirical scrutiny within finance and accounting. Such groundbreaking and forward-thinking research strategies are set to inspire subsequent scholars that are well-versed in their primary investigative techniques.

6.3 Managerial implications

This research provides valuable practical insights in several areas. First, the results of this study can supply investors with relevant data about the importance of the ESG score information. The findings highlight that considering ESG scores is likely to predict more accurate future corporate value. Hence, these results suggest the importance of establishing sustainable investment strategies. Namely, the empirical analysis results of this study remind investors that using ESG information to predict future corporate value more accurately can help such investors make more successful investment decisions.

Second, the results of this research indicate significant information that strengthens companies' awareness of ESG management in practice. The results of the machine learning analysis in this paper objectively showed that ESG-related information can be an important factor in forecasting future company valuation estimation. Most people may predict future firm value through ESG indicators, and the empirical results of this paper are expected to be considered as more apparent evidence for corporations to establish and implement sustainable management strategies in terms of the environment, society, and governance.

This research demonstrated that the prediction rate is improved when a model for forecasting future corporate value includes ESG ratings, which is non-financial information. These results indicate that non-financial information about corporations is also important in identifying future corporate value, particularly in modern society in which company sustainable management is emphasized; firm ESG evaluation information can be a helpful indicator of future firm value.

To help Korean companies actively engage in ESG activities and enter the global market successfully, it is necessary first to make ESG evaluations transparent and fair. Until now, companies have been evaluating ESG activities in their own ways, such as through self-ratings, so fair and objective ESG evaluations have not been consistently made. To prevent ESG washing, wherein companies promote ESG activities that they are not actually engaging in, it is necessary to evaluate companies' ESG activities more strictly in the future.

References

Abbas, S., Z. Jalil, A. R. Javed, I. Batool, M. Z. Khan, A. Noorwali, and A. Akbar(2021), "BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm," *PeerJ Computer Science*, 7, e390.

Abdi, Y., X. Li, and X. Càmara-Turull(2022), "Exploring the impact of sustainability (ESG) disclosure on firm value and financial performance (FP) in airline industry: the moderating role of size and age," *Environment, Development and Sustainability*, 24(4), pp. 5052-5079.

Abdul Rahman, R. and M. F. Alsayegh(2021), "Determinants of corporate environment, social and governance (ESG) reporting among Asian firms," *Journal of Risk and Financial Management*, 14(4), p.167.

Abedin, M. Z., G. Chi, M. M. Uddin, M. S. Satu, M. I. Khan, and P. Hajek(2020), "Tax default prediction using feature transformation-based machine learning," *IEEE Access*, 9, pp. 19864-19881.

Aboud, A. and A. Diab(2018), "The impact of social, environmental and corporate governance disclosures on firm value: Evidence from Egypt," *Journal of Accounting in Emerging Economies*, 8(4), pp.442-458.

Aggarwal, R., V. Sounderajah, G. Martin, D. S. Ting, A. Karthikesalingam, D. King, and A. Darzi, (2021), "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis," *NPJ Digital Medicine*, 4(1), p.65.

Alakus, T. B. and I. Turkoglu(2020), "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons and Fractals*, 140, p.110120.

Alareeni, B. A. and A. Hamdan(2020), "ESG impact on performance of US SandP 500-listed firms," *Corporate Governance: The International Journal of Business in Society*, 20(7), pp.

1409-1428.

Alsariera, Y. A., V. E. Adeyemo, A. O. Balogun, and A. K. Alazzawi(2020), "Ai meta-learners and extra-trees algorithm for the detection of phishing websites," *IEEE Access*, 8, pp. 142532-142542.

Aouadi, A. and S. Marsat(2018), "Do ESG controversies matter for firm value? Evidence from international data," *Journal of Business Ethics*, 151, pp.1027-1047.

Arvidsson, S. and J. Dumay(2022), "Corporate ESG reporting quantity, quality and performance: Where to now for environmental policy and practice?," *Business Strategy and the Environment*, 31(3), pp.1091-1110.

Asim, A. and A. Ismail(2019), "Impact of leverage on earning management: Empirical evidence from the manufacturing sector of Pakistan," *Journal of Finance and Accounting Research*, 1(1), pp.70-91.

Barua, A., L. F. Davidson, D. V. Rama, and S. Thiruvadi(2010), "CFO gender and accruals quality," *Accounting Horizons*, 24(1), pp. 25-39.

Behl, A., P. R. Kumari, H. Makhija, and D. Sharma (2022), "Exploring the relationship of ESG score and firm value using cross-lagged panel analyses: Case of the Indian energy sector," *Annals of Operations Research*, 313(1), pp. 231-256.

Bentéjac, C., A. Csörgő, and G. Martínez-Muñoz (2021), "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, 54, pp.1937-1967.

Chelgani, S. C., H. Nasiri, A. Tohry, and H. R. Heidari(2023), "Modeling industrial hydrocyclone operational variables by SHAP-CatBoost-A "conscious lab" approach," *Powder Technology*, 420, p.118416.

Chen, H., X. Li, Z. Feng, L. Wang, Y. Qin, M. J. Skibniewski, and Y. Liu(2023), "Shield attitude prediction based on Bayesian-LGBM machine learning," *Information Sciences*, 632, pp.105-129.

Chen, L., M. Pelger, and J. Zhu, (2023), "Deep learning in asset pricing," *Management Science*, 70(2), pp.714-750.

Chen, Y., J. Wu, and Z. Wu(2022), "China's commercial bank stock price prediction using a novel K-means-LSTM hybrid approach," *Expert Systems with Applications*, 202, p.117370.

Chen, Y., W. Zheng, W. Li, and Y. Huang(2021), "Large group activity security risk assessment and risk early warning based on random forest algorithm," *Pattern Recognition Letters*, 144, pp.1-5.

Choi, J. H., M. H. Ahn, C. H. Lee, M. S. Kim, Y. J. Jang, J. H. Lee, and T. E. Sung, (2021), "Deep learning based sales estimation research for technology valuation: focusing on marine fisheries," *Journal of the Society of Technological Innovation*, 24(5), pp.951-965.

Chouaibi, S. and J. Chouaibi(2021), "Social and ethical practices and firm value: The moderating effect of green innovation: Evidence from international ESG data," *International Journal of Ethics and Systems*, 37(3), pp.442-465.

Christopher, M., A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, and L. M. Zangwill(2018), "Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs," *Scientific Reports*, 8 (1), p.16685.

Clarkson, P., Y. Li, G. Richardson, and A. Tsang (2019), "Causes and consequences of voluntary assurance of CSR reports: International evidence involving Dow Jones Sustainability Index Inclusion and Firm Valuation," *Accounting, Auditing and Accountability Journa,* 32(8), pp.2451-2474.

Csizmadia, G., K. Liszkai-Peres, B. Ferdinandy, Á. Miklósi, and V. Konok(2022), "Human activity recognition of children with wearable devices using LightGBM machine learning," *Scientific Reports*, 12(1), p.5472.

D'Amato, A. and C. Falivena(2020), "Corporate social responsibility and firm value: Do firm size and age matter? Empirical evidence from European listed companies," *Corporate Social Responsibility and Environmental Management*, 27(2), pp.909-924.

D'Amato, A. and C. Falivena(2020), "Corporate social responsibility and firm value: Do firm size and age matter? Empirical evidence from European listed companies," *Corporate Social Responsibility and Environmental Management*, 27(2), pp.909-924.

Douiba, M., S. Benkirane, A. Guezzaz, and M. Azrour (2023), "An improved anomaly detection model for IoT security using decision tree and gradient boosting," *The Journal of Supercomputing*, 79(3), pp.3392-3411.

Duan, T., A. Anand, D. Y. Ding, K. K. Thai, S. Basu, A. Ng, and A. Schuler(2020), "Ngboost: Natural gradient boosting for probabilistic prediction," I*n International Conference on Machine Learning,* PMLR, pp.2690-2700.

El Bilali, A., T. Abdeslam, N. Ayoub, H. Lamane, M. A. Ezzaouini, and A. Elbeltagi(2023), "An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation," *Journal of Environmental Management*, 327, p.116890.

Eslami, E., A. K. Salman, Y. Choi, A. Sayeed, and Y. Lops(2020), "A data ensemble approach for real-time air quality forecasting using extremely randomized trees and deep neural networks," *Neural Computing and Applications,* 32, pp.7563-7579.

Fatemi, A., M. Glaum, and S. Kaiser(2018), "ESG performance and firm value: The moderating role of disclosure," *Global Finance Journal*, 38, pp.45-64.

Feng, Z. and Z. Wu(2021), "ESG disclosure, REIT debt financing and firm value," *The Journal of Real Estate Finance and Economics*, pp. 1-35.

Ghorbani, A., D. Ouyang, A. Abid, B. He, J. H. Chen, R. A. Harrington, and J. Y. Zou(2020), "Deep learning interpretation of echocardiograms," *NPJ Digital Medicine*, 3(1), p.10.

Gong, D. and Y. Liu(2022), "A Mechine Learning Approach for Botnet Detection Using Light GBM," *2022 3rd International Conference on Computer Vision, Image and Deep Learning and International Conference on Computer Engineering and Applications (CVIDL and ICCEA),* IEEE, pp.829-833.

Gu, Z. and H. Kim(2002), "Determinants of restaurant systematic risk: A reexamination," *The Journal of Hospitality Financial Management*, 10(1), pp.1-13.

Gupta, S., W. Zhang, and F. Wang(2016), "Model accuracy and runtime tradeoff in distributed deep learning: A systematic study," *2016 IEEE 16th International Conference on Data*

Mining (ICDM), IEEE, pp.171-180.

Hamid, S. A. and A. Habib(2014), "Financial forecasting with neural networks," *Academy of Accounting and Financial Studies Journal*, 18(4), p.37.

Hancock, J. T. and T. M. Khoshgoftaar(2020), "CatBoost for big data: an interdisciplinary review," *Journal of Big Data*, 7(1), pp.1-45.

Hijab, A., M. A. Rushdi, M. M. Gomaa, and A. Eldeib (2019), "Breast cancer classification in ultrasound images using transfer learning," *2019 Fifth international conference on advances in biomedical engineering (ICABME)*, IEEE, pp.1-4.

HU, W., Y. Y. CHAN, J. Huang, W. Zhou, and X. Li (2023), "Innovation Novelty and Firm Value: Deep Learning based Text Understanding," International Conference on Information Systems 2023 (ICIS23).

Huang, D. Z.(2021), "Environmental, social and governance (ESG) activity and firm performance: A review and consolidation," *Accounting and Finance*, 61(1), pp.335-360.

Husna, A. and I. Satria(2019), "Effects of return on asset, debt to asset ratio, current ratio, firm size, and dividend payout ratio on firm value," *International Journal of Economics and Financial Issues*, 9(5), pp.50-54.

Islam, S. and S. H. Amin(2020), "Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques," *Journal of Big Data*, 7, pp.1-22.

Iwendi, C., A. K. Bashir, A. Peshkar, R.Sujatha, J. M. Chatterjee, S. Pasupuleti, and O. Jo (2020), "COVID-19 patient health prediction using boosted random forest algorithm," *Frontiers in Public Health*, 8, p.357.

Jiang, M., L. Jia, Z. Chen, and W. Chen(2022), "The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm," *Annals of Operations Research*, pp.1-33.

Jing, N., Z. Wu, and H. Wang(2021), "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, 178, p. 115019.

Joo, C., H. Park, J. Lim, H. Cho, and J. Kim(2023), "Machine learning-based heat deflection temperature prediction and effect analysis in polypropylene composites using catboost and shapley additive explanations," *Engineering Applications of Artificial Intelligence*, 126, p.106873.

Kalbuana, N., B. Prasetyo, P. Asih, Y. Arnas, S. L. Simbolon, A. Abdusshomad, and F. M. Mahdi (2021), "Earnings management is affected by firm size, leverage and roa: evidence from Indonesia," *Academy of Strategic Management Journal*, 20, pp.1-12.

Kang K. G., J. Y. Park, and H. J. Na(2023), "A comparative study of machine learning-based future corporate value prediction models: the impact of including ESG ratings," *Journal of the Korean Society of Management*, 36 (9), pp.1515-1537.

Kezelj, T. and R. Gruenbichler(2021), "A Systematic Literature Review on Corporate Insolvency Prevention Using Artificial Intelligence Algorithms," *Journal of Strategic Innovation and Sustainability*, 16(4).

Kilimci, Z. H., A. O. Akyuz, M. Uysal, S. Akyokus,

M. O. Uysal, B. Atak Bulbul, and M. A. Ekmis(2019), "An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain," *Complexity*, 2019.

Kim M. S., J. S. Lee, E. S. Oh, C. H. Lee, J. H. Choi, Y. J. Jang, and T. E. Sung(2021), "A study on deep learning-based intelligent technology valuation: a model for predicting qualitative evaluation indicators through deep neural network learning," *Journal of Technological Innovation*, 24(6), pp.1141-1162.

Kim, H., H. Cho, and D. Ryu(2022), "Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data," *Computational Economics*, 59(3), pp. 1231-1249.

Kiran, R., P. Kumar, and B. Bhasker(2020), "DNNRec: A novel deep learning based hybrid recommender system," *Expert Systems with Applications*, 144, p.113054

Krylov, S.(2018), "Target financial forecasting as an instrument to improve company financial health," *Cogent Business and Management*, 5(1), p.1540074.

Kureljusic, M. and E. Karger(2023), "Forecasting in financial accounting with artificial intelligence - A systematic literature review and future research agenda," *Journal of Applied Accounting Research*.

Lee H. J., D. W. Shin, and H. E. Kim(2021), "Machine Learning-based enterprise value prediction model: utilizing online enterprise reviews," *Journal of Korean Society for Internet Information*, 22(5).

Lee, J., D. Jang, and S. Park(2017), "Deep learning-based corporate performance prediction model considering technical capability," *Sustainability*, 9(6), 899.

Li, M., C. Liu, and T. Scott(2019), "Share pledges and firm value," *Pacific-Basin Finance Journal*, 55, pp.192-205.

Li, Y., M. Gong, X. Y. Zhang, and L. Koh(2018), "The impact of environmental, social, and governance disclosure on firm value: The role of CEO power," *The British Accounting Review*, 50(1), pp.60-75.

Litjens, G., C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, and J. Van Der Laak(2016), "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Scientific reports*, 6(1), p.26286.

Long, J., Z. Chen, W. He, T. Wu, and J. Ren(2020), "An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market," *Applied Soft Computing*, 91, p.106205.

Louk, M. H. L. and B. A. Tama(2023), "Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system," *Expert Systems with Applications*, 213, p.119030.

Luo, M., Y. Wang, Y. Xie, L. Zhou, J. Qiao, S. Qiu, and Y. Sun(2021), "Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass," *Forests*, 12(2), p.216.

Mai, F., S.Tian, C. Lee, and L. Ma(2019), "Deep learning models for bankruptcy prediction using textual disclosures," *European Journal of Operational Research*, 274(2), pp.743-758.

Manavalan, B., S. Basith, T. H. Shin, L. Wei, and G. Lee(2019), "AtbPpred: a robust sequence-based prediction of anti-tubercular peptides using extremely randomized trees," *Computational and Structural Biotechnology Journal*, 17, pp.972-981.

Manikandan, G. and G. Bhuvaneswari(2022), "Knowledge discovery in data of prostate cancer by applying ensemble learning," *Indian Journal of Computer Science and Engineering (IJCSE), e-ISSN*, 0976-5166.

Massaoudi, M., S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub(2021), "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," *Energy*, 214, p. 118874.

Mishra, A. K. and S. Paliwal(2023), "Mitigating cyber threats through integration of feature selection and stacking ensemble learning: the LGBM and random forest intrusion detection perspective." *Cluster Computing*, 26(4), pp. 2339-2350.

Morgan, N. A., D. W. Vorhies, and C. H. Mason (2009), "Market orientation, marketing capabilities, and firm performance," *Strategic Management Journal*, 30(8), pp.909-920.

Nasiboglu, R. and E. Nasibov(2023), "WABL method as a universal defuzzifier in the fuzzy gradient boosting regression model," *Expert Systems with Applications*, 212, p.118771.

Nhat-Duc, H. and T. Van-Duc(2023), "Comparison of histogram-based gradient boosting classification machine, random Forest, and deep convolutional neural network for pavement raveling severity classification," *Automation in Construction*, 148, p.104767.

Park W. J., S. S. Shin, H. J. Kim, J. N. Yu, and J. H. Kim(2022), "An analysis of value investment by industry through deep learning models," *Thesis book for the academic conference of the Korean Society of Communications*, pp.914-915.

Patel, J., S. Shah, P. Thakkar, and K. Kotecha (2015), "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), pp.259-268.

Pavlyshenko, B. M.(2019), "Machine-learning models for sales time series forecasting," *Data*, 4 (1), p.15.

Pechlivanidis, E., D. Ginoglou, and P. Barmpoutis (2022), "Can intangible assets predict future performance? A deep learning approach," *International Journal of Accounting & Information Management*, 30(1), pp.61-72.

Polak, P., C. Nelischer, H. Guo, and D. C. Robertson (2020), "Intelligent finance and treasury management: what we can expect," *Ai and Society*, 35, pp.715-726.

Qiu, S. C., J. Jiang, X. Liu, M. H. Chen, and X. Yuan(2021), "Can corporate social responsibility protect firm value during the COVID-19 pandemic?," *International Journal of Hospitality Management*, 93, p.102759.

Rai, N., Y. Zhang, B. G. Ram, L. Schumacher, R. K. Yellavajjala, S. Bajwa, and X. Sun(2023), "Applications of deep learning in precision weed management: A review," *Computers and Electronics in Agriculture*, 206, p.107698.

Ranjitha, P. and M. Spandana(2021), "Predictive analysis for big mart sales using machine learning algorithms," *2021 5th International*

Conference on Intelligent Computing and Control Systems (ICICCS), pp.1416-1421.

Rezaei, H., H. Faaljou, and G. Mansourfar(2021), "Stock price prediction using deep learning and frequency decomposition," *Expert Systems with Applications*, 169, p.114332.

Sadhukhan, B., S. Chakraborty, and S. Mukherjee (2023), "Predicting the magnitude of an impending earthquake using deep learning techniques," *Earth Science Informatics*, 16 (1), pp.803-823.

Saeed, U., S. U. Jan, Y. D. Lee, and I. Koo(2021), "Fault diagnosis based on extremely randomized trees in wireless sensor networks," *Reliability Engineering and System Safety*, 205, p.107284.

Sanjeetha, R., A. Raj, K. Saivenu, M. I. Ahmed, B. Sathvik, and A. Kanavalli(2021), "Detection and mitigation of botnet based DDoS attacks using catboost machine learning algorithm in SDN environment," *International Journal of Advanced Technology and Engineering Exploration*, 8(76), p.445.

Sharma, A., and B. Singh(2020), "AE-LGBM: Sequence-based novel approach to detect interacting protein pairs via ensemble of autoencoder and LightGBM," *Computers in Biology and Medicine*, 125, p.103964.

Sheykhmousa, M., M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni(2020), "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, pp.6308-6325.

Sills, M. R., M. Ozkaynak, and H. Jang(2021), "Predicting hospitalization of pediatric asthma patients in emergency departments using machine learning," I*nternational Journal of Medical Informatics*, 151, p.104468.

Sitaula, C. and T. B. Shahi(2022), "Monkeypox virus detection using pre-trained deep learning-based approaches," *Journal of Medical Systems*, 46(11), p.78.

Speiser, J. L., M. E. Miller, J. Tooze, and E. Ip (2019), "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, 134, pp.93-101.

Sung T. E., M. S. Kim, C. H. Lee, J. H. Choi, Y. J. Jang, and J. H. Lee, (2021), "Technology valuation and estimation of evaluation variables based on deep learning," *Paper of the Korean Society of Contents*, 21(10), pp. 48-58.

Syed, D., H. Abu-Rub, A. Ghrayeb, and S. S. Refaat (2021), "Household-level energy forecasting in smart buildings using a novel hybrid deep learning model," *IEEE Access*, 9, pp. 33498-33511.

Taha, A. A., and S. J. Malebary(2020), "An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine," *IEEE Access*, 8, pp.25579-25587.

Traczynski, J.(2017). "Firm default prediction: A Bayesian model-averaging approach.," *Journal of Financial and Quantitative Analysis*, 52 (3), pp.1211-1245

Tsoumakas, G(2019), "A survey of machine learning techniques for food sales prediction," *Artificial Intelligence Review*, 52(1), pp.441-447.

Tyralis, H., G. Papacharalampous, and A. Langousis (2019), "A brief review of random forests for water scientists and practitioners and their recent history in water resources," *Water*, 11(5), p.910.

Velthoen, J., C. Dombry, J. J. Cai, and S. Engelke (2023), "Gradient boosting for extreme quantile regression," *Extremes*, pp.1-29.

Wahid, N., A. Zaidi, G. Dhiman, M. Manwal, D. Soni, and R. R. Maaliw(2023), "Identification of Coronary Artery Disease using Extra Tree Classification," *2023 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, pp.787-792.

Wang, J., C. Rao, M. Goh, and X. Xiao(2023), "Risk assessment of coronary heart disease based on cloud-random forest," *Artificial Intelligence Review*, 56(1), pp.203-232.

Wang, S., J. Wang, H. Lu, and W. Zhao(2021), "A novel combined model for wind speed prediction - Combination of linear model, shallow neural networks, and deep learning approaches," *Energy*, 234, p.121275.

Wang, X., L. Tan, and J. Fan(2023), "Performance Evaluation of Mangrove Species Classification Based on Multi-Source Remote Sensing Data Using Extremely Randomized Trees in Fucheng Town, Leizhou City, Guangdong Province," *Remote Sensing*, 15(5), p.1386.

Wang, Y(2022), "Personality type prediction using decision tree, gbdt, and cat boost," *2022 International Conference on Big Data, Information and Computer Network* (BDICN), IEEE. pp.552-558.

Wang, Y., Y. Liu, J. Zhao, and Q. Zhang(2023), "Low-Complexity Fast CU Classification Decision Method Based on LGBM Classifier," *Electronics*, 12(11), p.2488.

Wei, X., C. Rao, X. Xiao, L. Chen, and M. Goh(2023), "Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model," *Expert Systems with Applications*, 219, p.119648.

Wisesa, O., A. Adriansyah, and O. I. Khalaf(2020), "Prediction analysis sales for corporate services telecommunications company using gradient boost algorithm," *2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering* (BCWSP), IEEE, pp.101-106.

Wong, W. C., J. A. Batten, S. B. Mohamed-Arshad, S. Nordin, and A. A. Adzis(2021), " Does ESG certification add firm value?. *Finance Research Letters*, 39, p.101593.

Xi, B., E. Li, Y. Fissha, J. Zhou, and P. Segarra (2023), "LGBM-based modeling scenarios to compressive strength of recycled aggregate concrete with SHAP analysis," *Mechanics of Advanced Materials and Structures*, pp. 1-16.

Yoon, J.(2021), "Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach," *Computational Economics*, 57(1), pp.247-265.

Yu, B., C. Li, N. Mirza, and M. Umar(2022), "Forecasting credit ratings of decarbonized firms: Comparative assessment of machine learning models," *Technological Forecasting and Social Change*, 174, p.121255.

Yu, P., and X. Yan(2020), "Stock price prediction based on deep neural networks," *Neural Computing and Applications*, 32, pp.1609-1628.

Zhang, W., Y. He, L. Wang, S. Liu, and X. Meng, (2023), "Landslide Susceptibility mapping using random forest and extreme gradient boosting: A case study of Fengjie," *Chongqing. Geological Journal*, 58(6), pp.2372-2387.

Zhao, B., J. Feng, X. Wu, and S. Yan(2017), "A survey on deep learning-based fine-grained object classification and semantic segmentation," *International Journal of Automation and Computing*, 14(2), pp.119-135.

Zhu, H., W. Ge, and Z. Liu(2019), "Deep learning-based classification of weld surface defects," *Applied Sciences*, 9(16), p.3312.

• The author Kyung Gu Kang is currently the chairman of Commerce Holdings Inc. and SUNGJIN OCEAN & ENERGY Inc. and a research professor at the Institute of Industrial Policy. He obtained a master's degree in international finance from New York State University, a master's degree in economics from Korea University, and a doctorate in business administration from Seoul School of Integrated Sciences & Technologies, aSSIST. His primary research areas are ESG management, corporate value, management strategy, and machine learning/deep learning.

• The author Hyung Jong Na is currently a professor of accounting and taxation at Semyung University. He obtained bachelor's, master's, and doctorate degrees from Kyung Hee University. After that, he worked as a research professor at Kyung Hee University and Sungkyunkwan University. His main research fields are tax policy, corporate value, and ESG, and recently, the focus has been on prosperous research using text mining and machine learning/deep learning.

• The author Cheong Yeul Park is an Associate Professor of Seoul Business School at aSSIST University. He obtained his bachelor's, master's, and doctoral degrees in psychology from Chung-Ang University. After completing his doctoral degree, he worked as a postdoctoral researcher at Hankuk University of Foreign Studies. He later served as a research professor in the Department of Psychology at Korea University. His primary research interests include positive psychology, work-life balance, and organizational engagement.