ESG-Kor: A Korean Dataset for ESG-related Information Extraction and Practical Use Cases

Jaeyoung Lee **Geonyeong Son** **Misuk Kim***
Department of Data Science, Hanyang University, South Korea
{jaeylee, geonyeongson, misukkim}@hanyang.ac.kr

Abstract

With the expansion of pre-trained language model usage in recent years, the importance of datasets for performing tasks in specialized domains has significantly increased. Therefore, we have built a Korean dataset called `ESG-Kor` to automatically extract Environmental, Social, and Governance (ESG) information, which has recently gained importance. `ESG-Kor` is a dataset consisting of a total of 118,946 sentences that extracted information on each ESG component from Korean companies' sustainability reports and manually labeled it according to objective rules provided by ESG evaluation agencies. To verify the effectiveness and applicability of the `ESG-Kor` dataset, classification performance was confirmed using several Korean pre-trained language models, and significant performance was obtained. Additionally, by extending the ESG classification model to documents of small and medium enterprises and extracting information based on ESG key issues and in-depth analysis, we demonstrated potential and practical use cases in the ESG field.

1 Introduction

Pre-trained language models are exhibiting astonishing performances in various natural language processing (NLP) tasks, including classification, question answering (QA), machine translation, summarization, and conversation generation, and are currently being utilized in numerous fields. Moreover, pre-trained language models specialized domains, such as FinBERT (Yang et al., 2020), Med-BERT (Rasmy et al., 2021), and LEGAL-BERT (Chalkidis et al., 2020), which train the text of the respective domains, have emerged and shown superior performance in solving problems of each domain (Gururangan et al., 2020). However, these pre-trained language models show their limits when applied to more detailed subfields, leading to an increased significance of datasets from such detailed subfields (Hendrycks et al., 2021). Despite this, the construction of datasets from these detailed subfields necessitates the involvement of relevant experts, thereby requiring more time and costs and resulting in limited construction of related datasets. Moreover, most publicly available datasets and pre-trained language models are built based on English, limiting their applicability for performing specific tasks in non-English languages. While there have been pre-trained language models released for supporting multiple languages (Pires et al., 2019), they do not perform as well as the models based on a single language. Consequently, studies are being conducted to build datasets for specific language-based tasks by translating pre-constructed English data into an other language (Park et al., 2021). Nevertheless, these exhibit limited performance compared to English-based pre-trained language models and are especially difficult to use in highly specialized domains, prompting ongoing research into constructing datasets for specialized domains in diverse languages (Zadeh et al., 2020).

In this paper, we introduce `ESG-Kor`, a large-scale Korean dataset that encompasses Environmental, Social, and Governance (ESG) information. Considering the importance of ethical management, social issues such as racial and gender discrimination, and environmental problems like climate change, ESG is essential for a sustainable society. The importance of ESG is increasing globally as a measure of non-financial corporate activities. ESG information is mainly provided as text without a standardized format, making it very challenging to obtain only the necessary ESG information from a vast amount of data (Lee and Kim, 2023), and small-scale companies do not provide reports for the disclosure of ESG information. Therefore, the potential benefits are significant if ESG information can be extracted from various text data through

*Corresponding Author

Findings of the Association for Computational Linguistics: EACL 2024, pages 6627–6643

an ESG-based language model. For this, we used sustainability reports related to ESG in this study to build a Korean ESG dataset. We constructed a dataset of 118,946 sentences, manually labeled by three annotators possessing relevant knowledge. The annotators were conducted based on objective rules provided by ESG rating providers. We verified the performance by fine-tuning the existing pre-trained language models with the constructed `ESG-Kor` dataset and validated the effectiveness and applicability of `ESG-Kor`. Additionally, by extending our methodology and utilizing additional techniques, we demonstrate potential and practical use cases, such as the expansion of ESG into the small and medium enterprises (SMEs) environment and in-depth analysis of corporate ESG based on ESG key issues. Our contributions are as follows:

- We constructed `ESG-Kor`, a large-scale dataset for ESG sentence extraction and classification, comprised of 118,946 sentences by manually labeling sentences from sustainability reports, which are documents related to ESG. `ESG-Kor` is the first constructed large-scale dataset associated with ESG.

- We fine-tuned pre-trained language models using the `ESG-Kor`. Through experiments, we confirmed that the model appropriately performs the tasks of ESG sentence extraction and classification, verifying the dataset's usefulness.

- By extending our approach, we presented two use cases: the expansion of ESG into the SME environment and the extraction of ESG key issues. Through these use cases, we demonstrated the potential and usefulness of our dataset and model in the ESG field.

`ESG-Kor` is the only Korean-based professional NLP benchmark dataset manually labeled in the field of ESG. The entire construction process of `ESG-Kor` is illustrated in Figure 1. This dataset can be leveraged by diverse stakeholders, and the data construction process can be used as a benchmark for constructing ESG datasets in various languages. We release the `ESG-Kor` dataset using a data repository to advance ESG research. We have created temporary accounts with guaranteed anonymity to provide some dataset samples. The dataset can be accessed at `https://github.com/nowzer0/ESG-Kor`.

2 Related Work

For pre-trained language models to be effectively utilized in diverse nations and highly specialized domains, datasets and related research tailored to different languages and specialized domains are essential. Moreover, despite the increasing importance of ESG, research leverages machine learning techniques in the field of ESG is limited (Macpherson et al., 2021). We present research focused on specialized domains and minority languages, as well as research related to ESG.

2.1 Specialized Domain & Minority Language

Datasets tailored to specialized domains are being built. First, in the biomedical domain, a Chinese-based biomedical language understanding benchmark (Zhang et al., 2022), the PubMedQA (Jin et al., 2019), which is a biomedical QA set processed from data collected from the medical information search engine PubMed, the comprehensive benchmark dataset for PubMed-based Biomedical NLP known as BLURB (Gu et al., 2021), and the CORD-19 public research dataset provided for rapid access to COVID-19 researchers (Wang et al., 2020) have been constructed. In the legal domain, LexGLUE (Chalkidis et al., 2022), which standardizes various tasks for evaluation, and CUAD (Hendrycks et al., 2021), a dataset with annotations added by legal experts over a long period for reviewing legal contracts, have been released. In addition, Li et al. (2023) introduced the GeoGLUE dataset specialized for geographic understanding by collecting data from publicly available geographic resources for six natural language understanding tasks. The MultiHIERTT dataset (Zhao et al., 2022) has also been suggested for inferring answers to finance-related questions.

Furthermore, datasets based on diverse languages are being constructed. Yum et al. (2021) have developed a word embedding dataset that captures semantic similarity and relatedness specifically for Korean medical terms. For Japanese-based datasets, NWJC2Vec (Asahara, 2018), a Japanese word embedding dataset, JaQuAD (So et al., 2022), a QA dataset using Japanese Wikipedia data, and Japanese clinical domain semantic textual similarity (Mutinda et al., 2021) have been created. Datasets constructed in Russian include RuBQ (Korablinov and Braslavski, 2020), a knowledge-based QA dataset through Wikidata, SberQuAD (Efimov et al., 2020), a large-scale Reading Comprehension

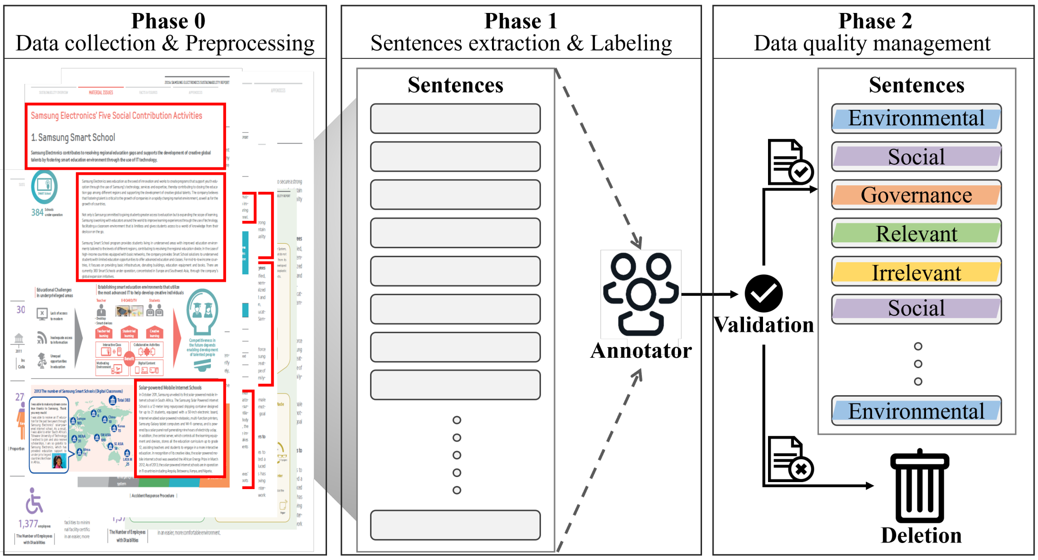

Figure 1: ESG-Kor construction process. Phase 0: Collect sustainability reports data and undergo preprocessing. Phase 1: Extract only the text from the preprocessed data, and proceed with labeling through annotators equipped with ESG-relevant knowledge. Phase 2: For the data quality, go through a verification process, and sentences of low quality are deleted. The example of the sustainability report on the left side of the figure is from the 2014 sustainability report published by Samsung Electronics.

(RC) dataset created similarly to English-based SQuAD (Rajpurkar et al., 2016), and NEREL-BIO (Loukachevitch et al., 2023), a PubMed abstract dataset. For Spanish datasets, there is the Biomedical dataset CoWeSe (Carrino et al., 2021) and a word embedding dataset in the Medical domain (Soares et al., 2019). In the realm of Arabic datasets, notable examples include the dialect dataset introduced by Abdul-Mageed et al. (2018) and the aspect-based sentiment analysis dataset presented by Obiedat et al. (2021). In addition, research efforts focus on creating publicly available resources in various areas. For instance, CMU-MOSEAS (Zadeh et al., 2020) dataset, a multi-modal sentiment dataset, has been made accessible, which encompasses not only a single language but also provides broader multilingual coverage, including languages such as Spanish, Portuguese, German, and French. Hence, constructing datasets not just for specific domains but for diverse languages is a crucial task.

2.2 ESG Research using Machine Learning

Unlike the traditional approach of evaluating companies based on their financial performance, the focus on non-financial performance based on sustainability has increased, leading to increasing interest in ESG. Accordingly, machine learning research is being conducted in the field of ESG to solve various problems, but it is limited due to the lack of expertise or relevant ESG datasets. De Lucia et al. (2020) used ESG indicators and economic indicators to predict the correlation between the two indicators and key financial indicators through various machine learning algorithms. Lanza et al. (2020) identified ESG indicators effective for building an investment portfolio through tree algorithms. Recently, research using ESG ratings provided by ESG evaluation companies is also being conducted (De Franco et al., 2020; Serafeim and Yoon, 2022). With the emergence of high-performance pre-trained language models, research has also been conducted to classify sentences as sustainable or unsustainable using pre-trained language models (Pontes et al., 2022), build an ESG index through media data (Sokolov et al., 2020), profile ESG-related language in discourse, and construct an ESG relevance dataset through a similarity-based unsupervised method (Raman et al., 2020). Machine learning in the ESG field can solve various ESG problems and is expected to reduce considerable environmental and social costs, so constructing ESG-related datasets is necessary. In the environmental aspect of ESG, recent research has been published on constructing datasets for detecting the presence of Net zero and reduction targets in texts provided by various organizations (Schimanski et al., 2023). In the ESG field, machine learning is needed to address multiple ESG issues. It can be expected to contribute to solving sustainable society challenges, offering significant environmental and social cost reduction benefits. Therefore, the construction of ESG-related datasets is essential.

3 ESG-Kor dataset

3.1 Sustainability Reports

We used Korean companies' sustainability reports as raw data for constructing ESG-Kor dataset. A sustainability report is a report that discloses a company's significant sustainability impacts on stakeholders based on economic, social, and environmental activities, as well as the company's performance over the past year. It reports various non-financial performances and includes various ESG-related information. Many diverse companies are publishing sustainability reports due to the increased demand for disclosure of ESG-related information from various company stakeholders and the institutionalization of the sustainability report disclosure system (Siew, 2015).

We have collected sustainability reports from 53 sectors of Korean companies published from 2002 to 2021. Since most sectors have fewer than 20 reports or have a small amount of data in the reports, we initially selected the automobile, aviation, electronics, finance, and heavy industry sectors, which had relatively more data among the sectors with more than 20 published reports, to use as raw data. For the purpose of evaluating data utility, we also built a dataset from the sustainability reports of POSCO in the steel sector and NAVER in the IT sector, which are significant companies in Korea. The sustainability reports used to build the dataset consist of 173 volumes, and the number of reports per sector is presented in Table 1. For NAVER, only one report had been published as of 2021, and only this report was used.

3.2 Preprocessing

The collected 173 volumes of sustainability reports are provided in portable document format (PDF) and contain various forms of data, including text, tables, and images. We extracted only the text from the PDF reports for dataset construction. To label each sentence, we used the Korean sentence splitter (KSS) (Ko and Park, 2021), a Korean sentence separation library, to distinguish the extracted text data at the sentence level. For ease of analysis, we performed preprocessing tasks such as unifying the numerical data format, deleting special characters, and removing duplicate sentences, and used the Korean spell-check API, py-hanspell (Han, 2015), to perform spell checks. Finally, we manually check the noise data generated by automatically not properly divided sentences, or too long or too short sentences, and perform segmentation or deletion.

Sector	Company	Number of reports	Number of sentences
Automobile	Hyundai motor company	17	15,154
	KIA corporation	14	9,921
Aviation	Korean airlines	17	7,022
	Asiana airlines	8	2,302
Electronics	Samsung electronics	14	12,491
	LG electronics	14	11,107
Finance	BNK financial group	7	4,591
	Hana financial group	10	5,080
	Industrial bank of Korea	6	3,435
	KB financial group	10	5,837
	Shinhan financial group	14	8,091
	Woori financial group	1	707
Heavy industry	Hyundai heavy industries	6	3,420
	Samsung heavy industries	10	5,219
	Doosan heavy industries	8	4,970
Steel industry	POSCO	16	18,886
IT	NAVER	1	713
	Total	173	118,946

Table 1: Sustainability reports data

3.3 Labeling

We manually labeled the preprocessed sustainability reports to construct ESG-Kor for extracting and classifying ESG information. The labeling was referenced from the criteria of leading global ESG rating providers and Korean ESG evaluation agencies such as Morgan Stanley Capital International (MSCI), Refinitiv, and the Korean Institute of Corporate Governance and Sustainability (KCGS). Three annotators, equipped with relevant knowledge based on the standards of each institution, familiarized and organized the labeling rules. The labeling rules are as follows:

- `E(Environmental)` Environmental factors such as green management, climate change, carbon emissions, natural resources, waste, resource use, energy conservation, etc.
- `S(Social)` Social factors such as human capital, community service, workers, working environment, suppliers, community contribution, human rights, gender and diversity, customer satisfaction, information protection, etc.
- `G(Governance)` Governance factors such as ethical management, shareholder rights, board

of directors, audit organization, stakeholders, dividends, regulatory compliance, taxation, etc.

- R(Relevant) Containing ESG-related information but not clearly categorized into individual ESG elements or including contents of both ESG elements simultaneously,

- I(Irrelevant): Not including ESG information.

The labeling was divided into two stages to ensure the dataset's quality. First, three annotators well-acquainted with the labeling rules annotated sentences clearly classified into each class. Subsequently, hard sentences to distinguish were separately discussed among the three annotators. They annotated if all agreed on labeling it to a specific class. However, if even one annotator had a different opinion, the sentence was deleted to prevent potential degradation of dataset quality. Through discussions among annotators, the majority of sentences were annotated into respective classes, and sentences deleted due to differing opinions among annotators, comprising less than 0.1% of the total dataset. The dataset construction involved labeling for 2 hours a day, five days a week, with an average of 500-600 sentences labeled per 2 hours. Therefore, each annotator labeled approximately 2,500 sentences per week, and the process first focused on labeling clear sentences over about four months. In the following two months, sentences that were difficult to categorize were discussed for labeling or deletion, and a period of cross-checking was conducted over another two months, totaling approximately 8-9 months of labeling time. Examples of sentences for each class in the ESG-Kor dataset are shown in Table 2. Also, the number of sentences by sector for each ESG element in the dataset is shown in Table 3.

3.4 Exploratory Data Analysis

After data preprocessing and labeling, the ESG-Kor dataset consists of a total of 118,946 sentences and classes, and it provides information about the year of issue and issuing company. Figure 2 shows the distribution of labeled classes for each sector and the entire dataset. As shown in Figure 2, most sectors show a similar distribution to the entire dataset. However, the finance sector has a significantly higher proportion of S class and fewer sentences labeled as E class than G class. The finance sector has less environmentally impactful activities and emphasizes social activities more than other industrial sectors. This confirms that the ESG-Kor dataset has been appropriately labeled according to real-world situations. Since pre-trained language models are affected by the maximum token length of text data, the distribution of token lengths for each sentence in the ESG-Kor dataset is shown in Figure 3. The maximum token length of ESG-Kor is 233, the minimum token length is 3, the average length is 42, and the standard deviation is 20. The shortest sentences in the dataset consist of only

Class	Sentences
E	We consider the environmental impact of new products during their whole development cycle.
S	The company also has created open communication channels to maintain strong relationships with suppliers.
G	Manages a website that offers information on ethical business management and provides an external reporting mechanism since 2002.
R	Analyzing the social and environmental impact of their manufacturing process and communicate the company's efforts with major suppliers in a transparent manner - something the company hopes to accomplish through sustainability reporting.
I	The company has also introduced a new focus on the business to business (B2B) market, going beyond the business-to-consumer (B2C) market

Table 2: Examples of labeled sentences

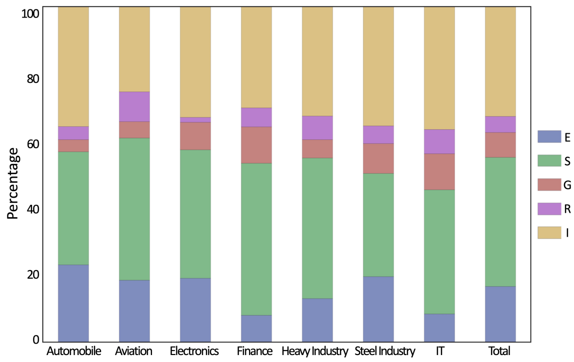

Figure 2: Class distribution of classes by sector

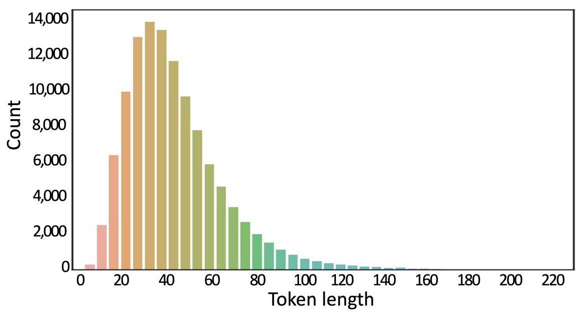

Figure 3: Distribution of token length

Class	Automobile	Aviation	Electronics	Finance	Heavy industry	Steel industy	IT	Total
E	5,783	1,727	4,511	2,228	1,779	3,701	60	19,789
S	8,466	3,946	9,033	12,578	5,693	5,802	264	45,782
G	896	463	1,928	3,014	751	1,690	77	8,819
R	972	825	337	1,566	957	996	51	5,704
I	8,958	2,363	7,789	8,355	4,429	6,697	261	38,852
Total	25,075	9,324	23,598	27,741	13,609	18,886	713	118,946

Table 3: Number of sentences by class and sector

Task	Model	Automobile	Aviation	Electronics	Finance	Heavy industry	Steel industry	ESG-Kor
Four-class classification	KoBERT	84.16	91.62	82.68	87.03	90.60	85.84	84.70
	DistilKoBERT	83.53	90.75	81.98	86.33	90.64	85.33	84.19
	KB-ALBERT	83.88	91.15	82.28	86.57	91.17	84.69	84.59
	KLUE-RoBERTa-base	<u>84.35</u>	<u>92.16</u>	<u>82.97</u>	<u>87.83</u>	<u>91.43</u>	<u>86.14</u>	<u>85.27</u>
	KLUE-RoBERTa-large	**84.59**	**92.64**	**83.82**	**88.28**	**91.56**	**87.30**	**85.87**
Binary classification	KoBERT	84.70	91.42	84.57	<u>88.71</u>	90.95	<u>85.88</u>	86.11
	DistilKoBERT	84.36	90.56	84.24	87.53	92.67	84.73	85.20
	KB-ALBERT	84.20	90.38	84.08	87.83	92.08	83.84	85.69
	KLUE-RoBERTa-base	<u>85.29</u>	<u>92.07</u>	<u>85.48</u>	88.60	<u>93.16</u>	85.23	<u>86.30</u>
	KLUE-RoBERTa-large	**85.96**	**92.51**	**86.12**	**88.78**	**93.54**	**86.10**	**86.67**

Table 4: Classification accuracy (%) by sector and model

three tokens and represent financial items such as *current bad debt allowance*, *cash and cash equivalents*, and *current depreciation allowance*. These should be labeled as `Irrelevant` to clearly distinguish them from ESG information. After this sentence labeling, we manually reviewed and removed sentences that were too short or too long. As most PLM models set the maximum token length to 512 or more, all sentences in `ESG-Kor` can be used as input without any tokens being cut off.

4 Experiments

4.1 Experimental Settings

To validate the effectiveness and applicability of `ESG-Kor`, we used models based on BERT (Devlin et al., 2019). Since `ESG-Kor` is a Korean dataset, we utilized popular Korean pre-trained language models such as KoBERT (SKTBrain, 2019), DistilKoBERT (Park, 2019), KB-ALBERT (KB-AI-Research, 2021), and KLUE-RoBERTa (Park et al., 2021). We fine-tuned these pre-trained models using the `ESG-Kor` dataset. The sentence x of the `ESG-Kor` dataset is represented by the [CLS] token through the layers of the model, and the model calculates the probability of belonging to classes c using [CLS] token C and classification layer W, as shown in Equation 1. The model is trained to minimize the cross-entropy loss, as shown in Equation 2.

$$p(y = c) = softmax(CW^T) \tag{1}$$

$$\mathcal{L} = -\sum_c y_c \log p(y = c) \tag{2}$$

We set the batch size to `64`, maximum length to `256`, optimizer to `Adamw`, epsilon to `1e-08`, and the learning rate of KoBERT, DistilKoBERT, KB-ALBERT, and KLUE-RoBERTa to `1e-04`, `5e-04`, `5e-05`, and `5e-05`, respectively. The dataset was split into `70%` training, `15%` validation, and `15%` test data. All experiments were conducted identically on `NVIDIA GeForce RTX 3090` 4ea. The model parameters and the time taken for each experiment are provided in the Appendix A.1.

4.2 Classification Result

The `ESG-Kor` dataset can be utilized for multiclass and binary classification tasks. The multiclass classification task classifies sentences into each E, S, and G class, and the binary classification task determines whether the sentence is related to ESG or not. The E, S, G, and R classes were consolidated into a `relevant` class for the binary classification dataset. At the same time, the `irrelevant` data remains the same as in the multiclass classification dataset. We didn't conduct sector-specific experiments for the IT sector due to the small number of collected data, but it was included in the dataset for the overall performance evaluation. The performance was evaluated based on classification accuracy, and the results for each sector and the overall experiment are shown in Table 4. The `ESG-Kor` dataset comprises five classes, but we experimented with a four-class classification task by excluding the R class, which contains both ESG information simultaneously. In the experimental results, KLUE-RoBERTa-large showed the highest accuracy of 85.87%, and even the DistilKoBERT, which recorded the lowest performance, confirmed meaningful performance with an accuracy of 84.19%. In the binary classification experiment results, KLUE-RoBERTa-large showed the best performance with an accuracy of 86.67%. In the experiments, models with more parameters performed better. However, with fewer parameters, KB-ALBERT performed as well as or better than KoBERT and DistilKoBERT. This can be interpreted as KB-ALBERT, a financial specialized model pre-trained on financial documents, understanding ESG documents better despite having fewer parameters. The number of parameters for the models is provided in the Appendix A.1. These experimental results confirmed that if fine-tuning is performed using the ESG-Kor dataset on Korean pre-trained language models, classification tasks can be appropriately performed regardless of the type of pre-trained language models. In other words, the experiment confirmed that when a dataset is specialized in a specific domain within a particular language, the applicability of pre-trained language models can be extended.

5 Use Cases

We propose two potential and practical cases to demonstrate additional applications of the ESG-Kor dataset and model. The first shows the scalability of ESG in the SME environment using a model trained on the ESG-Kor dataset. The second case shows a method that utilizes the trained model and additional methodologies for in-depth analysis of corporate ESG based on ESG key issues.

5.1 SMEs ESG Information Classification

Korean SMEs are classified based on their revenue or asset size. SMEs are smaller companies in terms of income or assets. They account for approximately 99% of all Korean enterprises. The importance of ESG management, previously dominated by large enterprises, is now increasing in SMEs due to reasons such as supplier selection and government procurement. However, most SMEs face difficulties with ESG management. We extend and validate ESG research into the SME environment through a model trained on the ESG-Kor dataset. Using ESG reports published by SMEs, we applied our methodology in an environment different from large enterprises. Since SMEs typically face challenges in publishing ESG reports, and their number is limited, we crawled the 2022 ESG reports of three SMEs through searches. We constructed an SME ESG data classification validation dataset using the same preprocessing and labeling process as the ESG-Kor dataset. The dataset consisted of 403 sentences and, considering the amount of data, was set up for binary classification related to ESG relevance rather than multiclass. The test results and class ratios using the model trained with the ESG-Kor dataset are as shown in Table 5. The experimental results showed that the model trained with the ESG-Kor dataset achieved an accuracy of 86.72% in classifying ESG reports from SMEs. This confirms that a model trained on data from large corporations can be applicable in the SME environment, validating our methodology's ability to be extended to SME settings. Utilizing our methodology, SMEs can extract and classify ESG data, serving as an auxiliary tool for SMEs facing difficulties in ESG management, information disclosure, and report writing. Furthermore, it can be utilized across various ESG areas by all stakeholders related to SMEs.

	Number of data		Accuracy
	Relevant	Irrelevant	
SME's binary classification	254	149	86.72%

Table 5: Classification results for SME's

5.2 Extraction of ESG Key Issues

MSCI is a global financial services company. In the ESG sector, MSCI evaluates and analyzes companies and investment products and is one of the most representative ESG rating agencies. MSCI, a leading ESG rating agency, proposes key issues for ESG evaluation. Within the environmental domain, there are four main issues and 13 detailed issues; in the social domain, there are four main issues and 14 detailed issues; and in the governance domain, there are two main issues and six detailed issues, making a total of 10 main categories and 33 detailed issues. More details are in Appendix A.2. MSCI utilizes these elements to evaluate and score companies' ESG performances. As companies generate vast amounts of information, directly scoring or extracting relevant information based on ESG key issues from this bulk is challenging and inefficient. Therefore, we demonstrate an efficient method to extract and analyze ESG information based on ESG key issues by combining a model trained with the ESG-Kor and NLP methodologies.

The process of extracting ESG information based on ESG key issues proceeds as follows. First, a model trained with the ESG-Kor dataset classifies ESG-related sentences from sustainability reports. Using a sentence embedding model, these sentences are mapped into the same vector space as ESG key issues. This step first extracts ESG-related data, saves resources and costs for embedding and retrieval, and filters out irrelevant noise information. Next, semantic search based on vector cosine similarity is employed to extract sentences most similar to each ESG key issue. For experimentation, we used sustainability reports from Hyundai Construction, a company in the construction sector not included in the ESG-Kor dataset, spanning from 2010 to 2020. The data from Hyundai Construction's sustainability reports consisted of 6,403 sentences, from which 4,588 ESG-related sentences were extracted using the model trained on the ESG-Kor dataset. As the sentence embedding model, we utilized KoSimCSE with RoBERTa as its backbone, the Korean version of SimCSE (Gao et al., 2021), which effectively learns sentence representations through contrastive learning. Using KoSimCSE, we extracted the top 10 sentences with the highest similarity to each detailed ESG key issue from the extracted ESG-related sentences in Hyundai Construction's sustainability reports. Next, the extracted sentences were combined based on the main key issue and reconstructed into the top 10 sentences. We included at least one sentence with the highest cosine similarity for each detail issue. This was because sentences extracted related to a particular detail issue could show high cosine similarity, leading to the possibility of extracting sentences pertaining only to one detail issue. For example, one sentence with the highest similarity to Carbon Emissions, Product Carbon Footprint, Financing Environmental Impact, and Climate Change Vulnerability within the Climate Change issue was included in the results, and a total of 10 sentences were compiled based on similarity within the same main category issue. Examples of the extracted sentences are shown in Appendix A.3. To evaluate this approach, we conducted a human evaluation with 10 participants, and the evaluation scale and results are presented in Table 6.

Issues		Score
Environmental	Climate Change	4.9
	Natural Capital	4.6
	Pollution & Waste	4.7
	Env. opportunities	4.8
Social	Human Capital	4.8
	Product Liability	4.5
	Stakeholder Opposition	4.6
	Social Opportunities	4.6
Governance	Corporate Governance	4.8
	Corporate Behavior	4.4

Table 6: The results of human evaluation. The evaluation scale is set as follows: 1 point = Not included, 2 points = Barely included, 3 points = Partially included, 4 points = Mostly included, and 5 points = Fully included.

The human evaluation results indicated that the Climate Change issue within the Environmental category scored the highest, with a score of 4.9. This confirms that the extracted sentences appropriately included content related to Carbon Emissions, Product Carbon Footprint, Financing Environmental Impact, and Climate Change Vulnerability within Hyundai Construction's sustainability reports. The Governance Behavior issue scored the lowest, yet it still achieved a score of 4.4, indicating that the sentences extracted mostly contained content relevant to the issue. The average score across all evaluations was 4.67, validating our approach in appropriately extracting sentences related to all ESG key issues. Utilizing the presented method, it becomes feasible to easily access information related to required ESG issues amidst the vast and diverse information generated daily. It significantly facilitates extracting information related to ESG key issues from large text volumes, enabling more accurate and detailed analysis. Furthermore, as many ESG rating agencies base their ESG scores on key issues, extracting information relevant to these key issues can provide valuable insights for ESG scoring. In the future, the information extracted could contribute to the automation of ESG scoring for companies, enhancing the efficiency and accuracy of ESG evaluations.

5.3 Discussion

ESG-Kor can be utilized more diversely beyond the use cases presented in Section 5.1 and 5.2. Companies can use it to monitor their external ESG performance and compile sustainability reports by extracting or classifying ESG information from various sources. It can also be used internally to evaluate data for sustainability reports prepared for external disclosure. Individual investors or investment firms can extract and classify ESG information of specific companies for investment decision-making. Government and regulatory bodies can use the extracted and classified information to supervise companies and develop relevant policies. ESG-Kor can also be used for ESG scoring. Many agencies evaluate corporate ESG, but each has different criteria, leading to inconsistent ESG ratings for the same company. Additionally, these agencies do not disclose precise ESG evaluation criteria. Therefore, an automatic and objective ESG scoring system is needed, and ESG-Kor can be used to extract objective ESG information for scoring. Besides practical applications, ESG-Kor is also academically significant. Many researchers are currently conducting studies on ESG data analysis, ESG rating prediction, and the correlation between ESG and financial performance. However, available ESG-related datasets are often small or limited in application. Our dataset is expected to benefit ESG researchers in various fields.

6 Conclusions

We manually constructed the high-quality, large-scale ESG-related ESG-Kor dataset to effectively utilize pre-trained models in the ESG field to extract and classify ESG information. We collected and preprocessed sustainability reports containing ESG information to construct the dataset and annotated them manually according to various ESG criteria. The ESG-Kor dataset consists of 118,946 sentences, focusing on the automobile, aviation, electronics, finance, heavy industry, steel industry, and IT sectors. We conducted experiments using Korean pre-trained language models to verify the effectiveness and applicability of the ESG-Kor dataset. As a result of the experiments, the KLUE-RoBERTa-large model recorded accuracies of 85.87% in multiclass classification and 86.67% in binary classification, confirming that the model appropriately learned and performed the task. This suggests that applying the ESG-Kor dataset to existing pre-trained language models allows for proper extraction and classification of relevant information in the field of ESG of Korea, a specialized domain in a particular country. We also show use cases of how our methodology can be extended to the SME environment and the extraction of ESG key issues. By collecting SME ESG reports and labeling them, we evaluated the performance through a model trained with the ESG-Kor dataset, achieving an accuracy of 86.72%. Furthermore, by combining our model with NLP techniques to extract ESG information based on ESG key issues, we validated its usefulness and potential in the ESG field through human evaluation.

In countries that use minority languages, research in specialized domains like ESG is limited due to the lack of datasets. Many datasets for minority languages are constructed by translating major languages, but translating specialized domains is challenging, and its performance is limited. Various countries are currently publishing documents such as sustainability reports that contain ESG information. Therefore, the entire construction process of the ESG-Kor dataset can serve as a benchmark for building ESG datasets in various countries. Furthermore, our approach can contribute to various ESG stakeholders, including researchers and practitioners, and can be utilized in multiple countries.

Limitation

In preparation for future work, we analyzed the limitations of our research. First, our study focuses on classifying ESG information from text data. Table or graph data within sustainability reports also contain crucial information about each company's ESG. ESG-Kor was built using only the text within sustainability reports. We plan to expand the dataset to a broader scope by adding more diverse information, such as table and graph data. Second, our studies are based on BERT models. While it would have been possible to use large language models (LLMs) like Llama (Touvron et al., 2023), which show excellent performance across various NLP tasks, we opted for models that were more accessible in terms of resources. Small language models can be more advantageous and easier to use in specific domain tasks or situations than large language models (Bosley et al., 2023). Additionally, our research primarily deals with corporate data, where LLMs may pose limitations regarding data privacy when used internally by corporations. Lastly, our work is currently limited to extracting and classifying ESG information. There might be a need for automatic scoring of ESG levels using the extracted ESG-related information. However, scoring requires the prior step of extracting and classifying ESG information from large texts, and the extracted ESG information can be directly utilized in various ways. The task of ESG information extraction is essential and meaningful, and we plan to extend it into a framework that can use the extracted ESG information to assign scores.

Ethics Statement

The primary objective of this study is to enable ESG stakeholders to effectively extract and classify ESG-related information, thereby contributing positively to the ESG field and society. The dataset we have constructed does not contain any socially controversial or ethically inappropriate data, and there are no concerns related to the ethics statement in this study.

Acknowledgments

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2020-II201373, Artificial Intelligence Graduate School Program (Hanyang University)) and the research fund of Hanyang University (HY-202400000001018)

References

Muhammad Abdul-Mageed, Hassan Alhuzali, and Mohamed Elaraby. 2018. You tweet what you speak: A city-level dataset of arabic dialects. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Masayuki Asahara. 2018. Nwjc2vec: word embedding dataset from 'ninjal web japanese corpus'. *Terminology*, 24(1):7–22.

Mitchell Bosley, Musashi Jacobs-Harukawa, Hauke Licht, and Alexander Hoyle. 2023. Do we still need bert in the age of gpt? comparing the benefits of domain-adaptation and in-context-learning approaches to using llms for political science research.

Casimiro Pio Carrino, Jordi Armengol-Estapé, Ona de Gibert Bonet, Asier Gutiérrez-Fandiño, Aitor Gonzalez-Agirre, Martin Krallinger, and Marta Villegas. 2021. Spanish biomedical crawled corpus: A large, diverse dataset for spanish biomedical language models. *arXiv preprint arXiv:2109.07765*.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2898–2904, Online. Association for Computational Linguistics.

Ilias Chalkidis, Abhik Jana, Dirk Hartung, Michael Bommarito, Ion Androutsopoulos, Daniel Katz, and Nikolaos Aletras. 2022. LexGLUE: A benchmark dataset for legal language understanding in English. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4310–4330, Dublin, Ireland. Association for Computational Linguistics.

Carmine De Franco, Christophe Geissler, Vincent Margot, and Bruno Monnier. 2020. Esg investments: Filtering versus machine learning approaches. *arXiv preprint arXiv:2002.07477*.

Caterina De Lucia, Pasquale Pazienza, and Mark Bartlett. 2020. Does good esg lead to better financial performances by firms? machine learning and logistic regression models of public enterprises in europe. *Sustainability*, 12(13):5317.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Pavel Efimov, Andrey Chertok, Leonid Boytsov, and Pavel Braslavski. 2020. Sberquad–russian reading comprehension dataset: Description and analysis. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11*, pages 3–15. Springer.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.

Suhun Han. 2015. py-hanspell. https://github.com/ssut/py-hanspell.

Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. CUAD: An expert-annotated NLP dataset for legal contract review. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William Cohen, and Xinghua Lu. 2019. PubMedQA: A dataset for biomedical research question answering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2567–2577, Hong Kong, China. Association for Computational Linguistics.

KB-AI-Research. 2021. Kb-albert. https://github.com/KB-AI-Research/KB-ALBERT.

Hyunwoong Ko and Sang-kil Park. 2021. Kss: A toolkit for korean sentence segmentation. https://github.com/hyunwoongko/kss.

Vladislav Korablinov and Pavel Braslavski. 2020. Rubq: A russian dataset for question answering over wikidata. In *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II 19*, pages 97–110. Springer.

Ariel Lanza, Enrico Bernardini, and Ivan Faiella. 2020. Mind the gap! machine learning, esg metrics and sustainable investment. *Machine Learning, ESG Metrics and Sustainable Investment (June 26, 2020). Bank of Italy Occasional Paper*, (561).

Jaeyoung Lee and Misuk Kim. 2023. Esg information extraction with cross-sectoral and multi-source adaptation based on domain-tuned language models. *Expert Systems with Applications*, 221:119726.

Dongyang Li, Ruixue Ding, Qiang Zhang, Zheng Li, Boli Chen, Pengjun Xie, Yao Xu, Xin Li, Ning Guo, Fei Huang, et al. 2023. Geoglue: A geographic language understanding evaluation benchmark. *arXiv preprint arXiv:2305.06545*.

Natalia Loukachevitch, Suresh Manandhar, Elina Baral, Igor Rozhkov, Pavel Braslavski, Vladimir Ivanov, Tatiana Batura, and Elena Tutubalina. 2023. Nerel-bio: a dataset of biomedical abstracts annotated with nested named entities. *Bioinformatics*, 39(4):btad161.

Martina Macpherson, Andrea Gasperini, and Matteo Bosco. 2021. Implications for artificial intelligence and esg data. *Available at SSRN 3863599*.

Faith Wavinya Mutinda, Shuntaro Yada, Shoko Wakamiya, and Eiji Aramaki. 2021. Semantic textual similarity in japanese clinical domain texts using bert. *Methods of Information in Medicine*, 60(S 01):e56–e64.

Ruba Obiedat, Duha Al-Darras, Esra Alzaghoul, and Osama Harfoushi. 2021. Arabic aspect-based sentiment analysis: A systematic literature review. *IEEE Access*, 9:152628–152645.

Jangwon Park. 2019. Distilkobert: Distillation of kobert. https://github.com/monologg/DistilKoBERT.

Sungjoon Park, Jihyung Moon, Sungdong Kim, Won Ik Cho, Jiyoon Han, Jangwon Park, Chisung Song, Junseong Kim, Yongsook Song, Taehwan Oh, Joohong Lee, Juhyun Oh, Sungwon Lyu, Younghoon Jeong, Inkwon Lee, Sangwoo Seo, Dongjun Lee, Hyunwoo Kim, Myeonghwa Lee, Seongbo Jang, Seungwon Do, Sunkyoung Kim, Kyungtae Lim, Jongwon Lee, Kyumin Park, Jamin Shin, Seonghyun Kim, Lucy Park, Alice Oh, Jungwoo Ha, and Kyunghyun Cho. 2021. Klue: Korean language understanding evaluation. *Preprint*, arXiv:2105.09680.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Elvys Linhares Pontes, Mohamed Benjannet, Jose G Moreno, and Antoine Doucet. 2022. Using contextual sentence analysis models to recognize esg concepts. *arXiv preprint arXiv:2207.01402*.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.

Natraj Raman, Grace Bang, and Armineh Nourbakhsh. 2020. Mapping esg trends by distant supervision of neural language models. *Machine Learning and Knowledge Extraction*, 2(4):453–468.

Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. 2021. Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, 4(1):86.

Schimanski, Bingler, Hyslop, Kraus, and Leippold. 2023. Climatebert-netzero: Detecting and assessing net zero and reduction targets. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*.

George Serafeim and Aaron Yoon. 2022. Which corporate esg news does the market react to? *Financial Analysts Journal*, 78(1):59–78.

Renard YJ Siew. 2015. A review of corporate sustainability reporting tools (srts). *Journal of environmental management*, 164:180–195.

SKTBrain. 2019. Korean bert pre-trained cased. https://github.com/SKTBrain/KoBERT.

ByungHoon So, Kyuhong Byun, Kyungwon Kang, and Seongjin Cho. 2022. Jaquad: Japanese question answering dataset for machine reading comprehension. *arXiv preprint arXiv:2202.01764*.

Felipe Soares, Marta Villegas, Aitor Gonzalez-Agirre, Martin Krallinger, and Jordi Armengol-Estapé. 2019. Medical word embeddings for spanish: Development and evaluation. In *Proceedings of the 2nd clinical natural language processing workshop*, pages 124–133.

Alik Sokolov, Jonathan Mostovoy, Jack Ding, and Luis Seco. 2020. Building machine learning systems to automate esg index construction.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darrin Eide, Kathryn Funk, Yannis Katsis, Rodney Michael Kinney, Yunyao Li, Ziyang Liu, William Merrill, Paul Mooney, Dewey A. Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex D. Wade, Kuansan Wang, Nancy Xin Ru Wang, Christopher Wilhelm, Boya Xie, Douglas M. Raymond, Daniel S. Weld, Oren Etzioni, and Sebastian Kohlmeier. 2020. CORD-19: The COVID-19 open research dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Online. Association for Computational Linguistics.

Yi Yang, Mark Christopher Siy Uy, and Allen Huang. 2020. Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.

Yunjin Yum, Jeong Moon Lee, Moon Joung Jang, Yoojoong Kim, Jong-Ho Kim, Seongtae Kim, Unsub Shin, Sanghoun Song, and Hyung Joon Joo. 2021. A word pair dataset for semantic similarity and relatedness in korean medical vocabulary: Reference development and validation. *JMIR Medical Informatics*, 9(6):e29667.

Amir Zadeh, Yan Sheng Cao, Simon Hessner, Paul Pu Liang, Soujanya Poria, and Louis-Philippe Morency. 2020. Cmu-moseas: A multimodal language dataset for spanish, portuguese, german and french. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2020, page 1801. NIH Public Access.

Ningyu Zhang, Mosha Chen, Zhen Bi, Xiaozhuan Liang, Lei Li, Xin Shang, Kangping Yin, Chuanqi Tan, Jian Xu, Fei Huang, Luo Si, Yuan Ni, Guotong Xie, Zhifang Sui, Baobao Chang, Hui Zong, Zheng Yuan, Linfeng Li, Jun Yan, Hongying Zan, Kunli Zhang, Buzhou Tang, and Qingcai Chen. 2022. CBLUE: A Chinese biomedical language understanding evaluation benchmark. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7888–7915, Dublin, Ireland. Association for Computational Linguistics.

Yilun Zhao, Yunxiang Li, Chenying Li, and Rui Zhang. 2022. MultiHiertt: Numerical reasoning over multi hierarchical tabular and textual data. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6588–6600, Dublin, Ireland. Association for Computational Linguistics.

A Appendix

A.1 Model Details

Table A.1 provides the number of parameters for each model, along with the elapsed time for both the 4-class and 2-class classification tasks.

Model	Number of Parameters	4-class Elapsed Time (s)	2-class Elapsed Time (s)
KoBERT	92M	2812	2944
DistilKoBERT	28M	809	839
KB-ALBERT	9M	1409	1424
KLUE-RoBERTa	111M	3133	3286
KLUE-RoBERTa-large	337M	9492	9965

Table A.1: Model parameters and elapsed time for experiments

A.2 MSCI's ESG key issues

We utilized the ESG key issues from MSCI, a leading ESG rating agency, to automatically extract information related to them. MSCI presents key issues for evaluating each domain of ESG. MSCI's ESG key issues are first categorized into environmental, social, and governance domains and then further into related issues. These are subsequently broken down into detailed issues. The key issues for each ESG domain are shown in Table A.2.

Issue		
Environmental	Climate Change	Carbon Emissions Product Carbon Footprint Financing Environmental Impact Climate Change Vulnerability
	Natural Capital	Water Stress Biodiversity & Land Use Raw Material Sourcing
	Pollution & Waste	Toxic Emissions & Waste Packaging Material & Waste Electronic Waste
	Env. opportunities	Clean Tech Green Building Renewable Energy
Social	Human Capital	Labor Management Health & Safe Human Capital Development Supply Chain Labor Standards
	Product Liability	Product safety & Quality Consumer Financial Protection Privacy & Data Security Responsible Investment Chemical safety
	Stakeholder Opposition	Controversial Sourcing Community Relations
	Social Opportunities	Access to Finance Access to Health Care Opportunities in Nutrition & Health
Governance	Corporate Governance	Board Pay Ownership Accounting
	Corporate Behavior	Business Ethics Tax Transparency

Table A.2: MSCI's ESG key issues

A.3 Human Evaluation

To evaluate each piece of information extracted based on ESG key issues, we conducted a human evaluation with 10 participants. To ensure accurate assessment, we scored the presence of detailed items under ESG key issues within a total of 10 pieces of extracted information in each domain on a scale of 1 (Not included) to 5 (Fully included). Examples of actual templates for each domain containing the ten pieces of extracted information are shown in Tables A.3, A.4, and A.5. The extracted sentences in the examples were taken from the sustainability report of Hyundai Construction, a South Korean construction company, and translated from Korean to English.

Table A.3 is the human evaluation template for Natural Capital in the environmental domain. Natural Capital consists of three detailed issues: water stress, biodiversity & land use, and raw material sourcing. The extracted information includes content such as water usage at construction sites, rainwater utilization systems, analysis of the impact on the ecosystem around construction sites, and wastewater recycling, confirming the appropriate extraction of relevant information for Natural Capital. Table A.4 is the evaluation template for Human Capital in the social domain. Human Capital comprises four detailed issues: labor management, health & safety, human capital development, and supply chain labor standards. This result includes content on building a talent development system, nurturing talents, institutional improvement through labor-management councils, improving the working environment, and reflecting special terms when selecting suppliers, confirming the appropriate extraction of Human Capital issues. Table A.5 provides an example of Corporate Behavior in the governance domain. Corporate Behavior includes two detailed issues: business ethics and tax transparency. The extracted information contains content on ethical management, compliance with fair trade, faithful tax payment, transparent disclosure of tax-related matters, and establishment of an ethical corporate culture. Corporate Behavior received the lowest human evaluation score of 4.4. This is likely because the content related to business ethics was relatively abstract, making it difficult for human evaluators to determine if it adequately addressed business ethics. However, it was confirmed that the extracted sentences included some content on ethical management. Thus, it can be interpreted that the current report does not deal with Business Ethics information in detail. This result can be extended for use in developing future ESG scoring systems.

Question	Do the following sentences include issues such as Natural Capital (Water Stress / Biodiversity & Land Use / Raw Material Sourcing)?
Extracted sentences	• The necessary water is obtained on construction sites through surface water, groundwater, and tap water. • The impacts on the ecosystem surrounding the construction site, including biodiversity, are scientifically analyzed and incorporated into the biodiversity management strategy's planning and implementation process. • Green purchasing involves buying products that contribute to resource savings and reduce environmental pollution compared to other products, services, or raw materials for the same use. • Additionally, purchasing materials directly on-site reduces costs, enhances cost competitiveness, and increases the sales of local partners. • A large amount of the water used at the site is for spraying to reduce fugitive dust, managing water usage and wastewater treatment. • Environmentally friendly water use has become significant as the global water shortage intensifies, alongside energy, highlighting the reduction of water consumption in building operations. • Rainwater utilization systems are applied to save water consumed during building operations by actively using rainwater for landscaping purposes. • From the construction planning stage, the impacts on the ecosystem surrounding the construction site are scientifically analyzed to preserve biodiversity as much as possible and minimize changes in topography and vegetation damage by applying appropriate design and construction methods. • On construction sites, water consumption is reduced by using rainwater storage facilities, groundwater dewatering systems, and sprinklers instead of water tank trucks for dust control. • Additionally, the application of middle water recycling systems that recycle domestic sewage and rainwater utilization systems is being expanded on-site, and facilities that purify river water, seawater, and groundwater for environmentally friendly use are being established to optimize water resource usage.
Inclusion level	◯ Not included ◯ Barely included ◯ Partially included ◯ Mostly included ◯ Fully included

Table A.3: Example of an evaluation template for the environmental domain for human evaluation

Question	Do the following sentences include issues such as Human Capital (Labor Management / Health & Safe / Human Capital Development / Supply Chain Labor Standard)?
Extracted sentences	• Additionally, we are developing experts by establishing an advanced talent development system through the eHRD (Human Resource Development) system. • Continuous talent cultivation is necessary for sustained growth, requiring the nurturing of excellent personnel. • The labor union discusses system improvements, productivity enhancement, welfare, and working environment improvements through the labor-management council. • By establishing safety and health policies aimed at instilling a safety culture through education and improving worker welfare, we regularly track and manage risks related to occupational health and safety, providing employees with a safe working environment. • Based on this, we include environmental, labor, and health and safety standards as special conditions in supplier selection contracts, making them mandatory for all contracts. • To this end, we are implementing talent development strategies, presenting a vision for personal growth, and supporting the development of global competencies. • Additionally, we conduct leadership training and operate global advanced study programs to develop our employees into global leaders. • We are expanding investments in employee competency development, striving to cultivate multifaceted talents with rich creativity and imagination, beyond simple problem-solving abilities. • Hyundai Engineering & Construction is making continuous efforts for talent development through domestic and international training programs and in-house online education. • We plan to strengthen immediately applicable training programs, such as CPD (Continuing Professional Development) for engineers, competency enhancement courses for site managers, and leadership training for those promoted, while also expanding education for employees' spouses and children.
Inclusion level	◯ Not included ◯ Barely included ◯ Partially included ◯ Mostly included ◯ Fully included

Table A.4: Example of an evaluation template for the social domain for human evaluation

Question	Do the following sentences include issues such as Corporate Behavior (Business Ethics / Tax Transparency)?
Extracted sentences	• Practicing ethical management is not a choice but a necessity for a company's sustainable growth. • Hyundai Engineering & Construction maintains an open, honest, and positive relationship with tax authorities, faithfully pays taxes annually according to each country's policies, and is committed to transparently disclosing all tax-related matters. • We are devoting our entire capabilities to realizing the values of sustainable management, such as transparent ethical management, environmentally friendly business operations, win-win development with partners, and fulfilling social responsibilities. • Expanding the scope of application, Hyundai Engineering & Construction practices ethical management together with stakeholders related to its business. • Hyundai Engineering & Construction maintains an open, honest, and positive relationship with tax authorities, including the National Tax Service, and faithfully pays taxes annually according to each country's tax policies. • Hyundai Engineering & Construction is leading the establishment of an ethical corporate culture not only among its employees but also among its affiliates, partners, and business associates. • All employees and partners of Hyundai Engineering & Construction share and internalize the company's comprehensive ethical management system, practicing the code of ethics that reflects long-term corporate value. • As a responsible corporate citizen, Hyundai Engineering & Construction strives to become a respected company among stakeholders through ethical business activities and fair trade compliance. • Through this, the company has demonstrated its strong commitment to not only ethical management but also fulfilling its social responsibilities. • The global market, where top-tier companies compete, is our stage of activity.
Inclusion level	◯ Not included ◯ Barely included ◯ Partially included ◯ Mostly included ◯ Fully included

Table A.5: Example of an evaluation template for the governance domain for human evaluation