

# ESG 보고서 키워드 분석



아메드 사이드 알리

# 목차



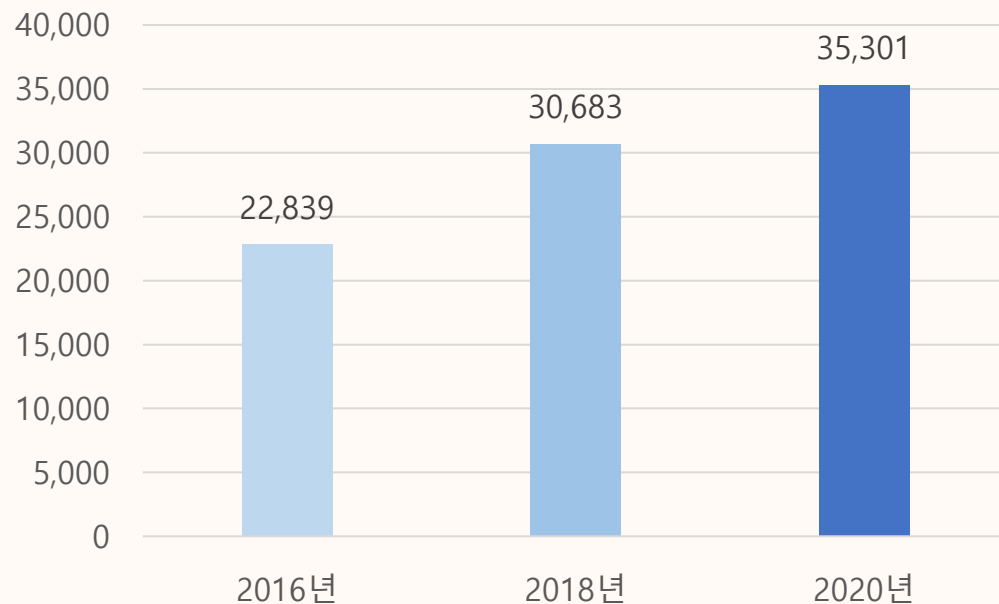
01. 분석의 필요성

02. 분석 방법

03. 분석 결과 및 정리

# 01. 분석의 필요성

■ 기업 경영에서 ESG가 중요해짐에 따라 기업의 ESG 활동과 그 평가를 파악하고자 함



자료: GSIA(2020)

매년 기업의 ESG를 고려하여 투자하는 글로벌 ESG 투자 규모는 증가 중

KBS | 6일 전 | 네이버뉴스

[ET] SPC에서 또 근로자 사고... '불매운동'에 파리바게트 가맹점주...  
SPC 파리바게트의 점주들이 공포에 떨고 있습니다. 최근 SPC 계열사에서 발생한 20대 여성 노동자 사고에 이어... 사망 사고 원인 규명을 위한 합동 현장 감식을 진...



국민일보 PICK | 2022.10.22. | 네이버뉴스

거세지는 '불매운동'...파리바게트 가맹점주들 "분노 공감"

온라인상에서는 파리바게트 뿐만 아니라 배스킨라빈스, 던킨, 샤니, 삼립식품 등 SPC 그룹 계열사 리스트를 공유하며 불매운동 동참을 호소하고 있다. 대학가에서...



"시중에 파는 빵 죄다 SPC, 완전 '...' 데일리안 PICK | 2022.10.22. | 네이버뉴스

레디앙 | 2022.10.18.

파리바게트 20대 여성노동자 죽음 2인1조, 안전교육, 안전센서도...

파리바게트에 반죽과 완제품 빵을 생산해 납품하는 SPC그룹의 계열사 SPL에서 20대 여성 노동자가 끼임 사고로 사망하는 참사가 벌어진 가운데, 업무 강도가 높고...



60%가 넘는 소비자들은 제품 구매 시 ESG 활동을 고려(대한상의, 2021)

# 01. 분석의 필요성

## 분석목적

기업들의 ESG 보고서의 텍스트 분석을 통해 기업들이 ESG에 어떠한 노력을 하고 있는지 파악하고, 어떠한 노력이 실제 ESG 등급에 좋은 영향을 미치는지 파악

## 분석방법

Wordcloud를 활용하여 ESG 등급이 높은 기업과 낮은 기업의 ESG 보고서를 비교 분석하고자 함

## 02. 분석 방법

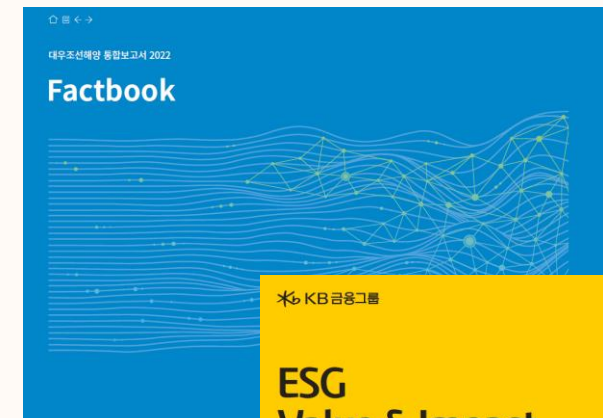
### 데이터



기업의 ESG 보고서 중 ESG 등급이 높은 기업과 낮은 기업의 키워드를 비교 및 분석 하고자 함

기업	기관	평가년도	종합점수	환경	사회	지배구조	요약보고서
KB금융	KCGS	2021	A+	A+	A+	A+	
	MSCI	2022	AA	-	-	-	-

기업	기관	평가년도	종합점수	환경	사회	지배구조	요약보고서
대우조선해양	KCGS	2021	A	B	A+	A	
	MSCI	2020	CCC	-	-	-	-



ESG 보고서가 공개된 기업 중 ESG 등급이 높은 기업(KB 금융), 낮은 기업(대우조선해양)을 선정

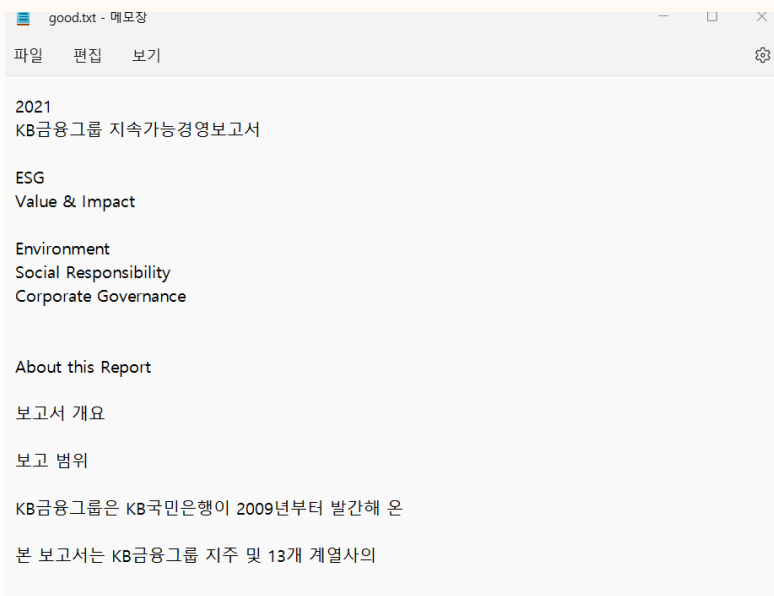
선정한 기업의 ESG 보고서를 다운로드  
(pdf 파일)

## 02.

## 분석 방법 데이터



기업의 ESG 보고서 중 ESG 등급이 높은 기업과 낮은 기업의 키워드를 비교 및 분석 하고자 함



```
from konlpy.tag import Okt
okt=Okt()
from wordcloud import WordCloud, STOPWORDS # word cloud 내 wordcloud, stopwords 설치
from PIL import Image # pillow 내 image 불러오기
import matplotlib.pyplot as plt # matplotlib 라이브러리 불러오기
import pandas as pd # pandas 불러오기
import numpy as np # numpy 불러오기
mask= np.array(Image.open('C:\\python\\star.jpg')) # 마스크를 numpy의 array 행렬 형태로 불러오기

#reading csv text file
df = open("C:\\python\\bad.txt", 'rt', encoding='UTF-8') #한글 파일 인코딩(utf-8)하여 텍스트파일 불러오기
line= df.readlines() # 텍스트 파일을 한 줄 씩 읽기
comment_words='' #아래에서 모든 단어들을 한줄로 연결하기 위해 표시하는 띄어쓰기
stopwords=set(STOPWORDS) #stopwords, 삭제하고 싶은 의미없는 단어 설정할 수 있는 변수 생성
nouns=[]
stopwords=['또한', '위해', '대한', '통해', '단위', '구분', '기준', '관련']
# iterate through the csv file
for val in line: # 텍스트 파일에서 한줄 씩 불러온 val에 접근
    val= str(val)# val을 문자열 형태로 변환
    val=okt.nouns(val)
    nouns.append(val)

comment_words = ' '
for t in nouns:
    for tokens in t:
        comment_words = comment_words + tokens + ' '
```

파이썬의 PDFminer 모듈을 활용하여  
pdf의 텍스트를 추출하여 txt 파일로 저장

ESG 등급이 높은 기업, 낮은 기업 각  
각을 대상으로 형태소 분석 및 word  
cloud  
분석 실행

## 02.

## 분석 방법

분석 코드

```
!pip install wordcloud
from wordcloud import WordCloud, STOPWORDS
from PIL import Image
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

```
mask= np.array(Image.open('c:/python/star.jpg'))
df = open("C:\\python\\good.txt",'rt', encoding='UTF-8')
line=df.readlines()
comment_words = ''
```

```
nouns=[]
stopwords=['또한', '위해', '대한', '통해','단위','구분','기준','관
련',
'금융','그룹','기업','가능','그룹','국민은행','보고서']
```

-> Wordcloud, matplotlib 등 import

-> 마스크를 array 형태로 import  
-> 데이터 파일 오픈  
-> 파일 한줄씩 읽기  
-> 키워드를 모으기 위한 변수 선언

-> 명사를 추출하기 위한 리스트 선언  
불용어(stopwords) 설정

## 02.

## 분석 방법

분석 코드

```
from konlpy.tag import Okt  
okt=Okt()
```

```
for val in line:  
    val = str(val)  
    val=okt.nouns(val)  
    nouns.append(val)
```

```
comment_words = ''  
for t in nouns:  
    for tokens in t:  
        comment_words = comment_words + tokens +  
        ''
```

-> konlpy의 Okt import

-> 텍스트 파일에서 불러온 line을 문자열로 변환

-> val에 있는 단어 okt를 활용하여 명사로 변환

-> comment+words 변수 설정

-> 모든 명사 한줄로 연결



## 02.

## 분석 방법

분석 코드

```
wordcloud = WordCloud(font_path  
    ='c:/python/NanumSquare.ttf  
    width = 1600, height = 800,  
    background_color = 'white',  
    stopwords = stopwords,  
    min_font_size = 1,  
    mask=mask).generate(comment_words)
```

```
plt.figure(figsize = (8, 8), facecolor = None)  
plt.imshow(wordcloud)  
plt.axis("off")  
plt.tight_layout(pad = 0)  
plt.show()
```

-> Wordcloud 함수 : 폰트 : 나눔스퀘어, 넓이 : 1600, 높이 : 800, 배경색 : 흰색, 불용어 및 마스크 설정, 최소 폰트 사이즈 : 1

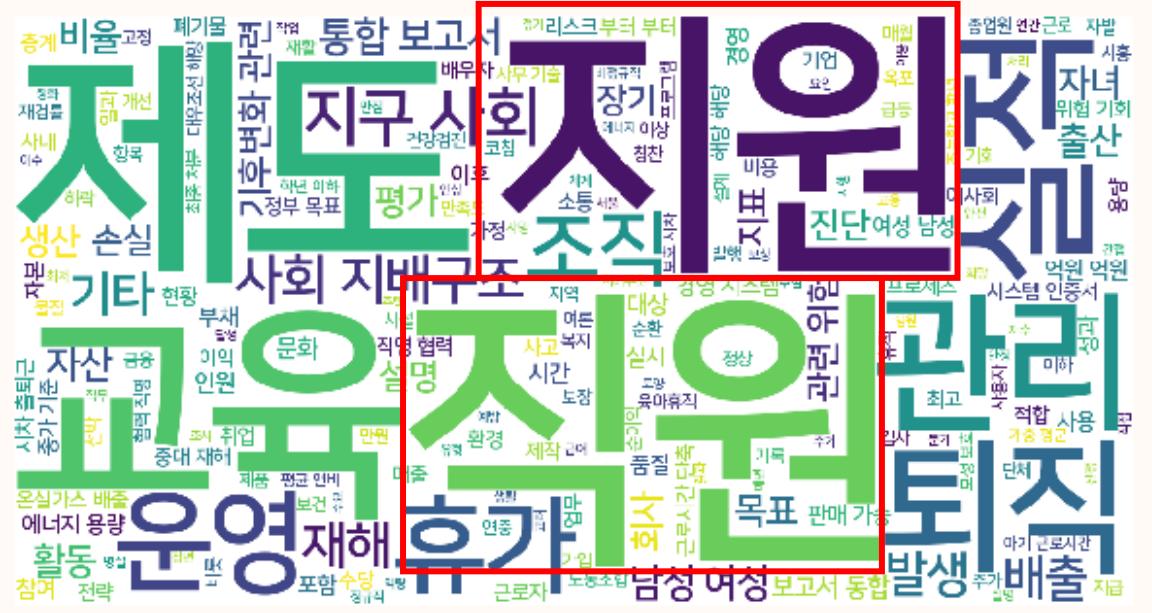
-> figure 사이즈 : 8,8 , 테두리색 없음  
워드클라우드 표시  
-> 축없음  
-> 공백정도 설정  
-> 워드 클라우드 보이기

## 03. 분석 결과

| ESG 등급이 높은 기업(KB 금융)|



| ESG 등급이 낮은 기업(KB 금융)|



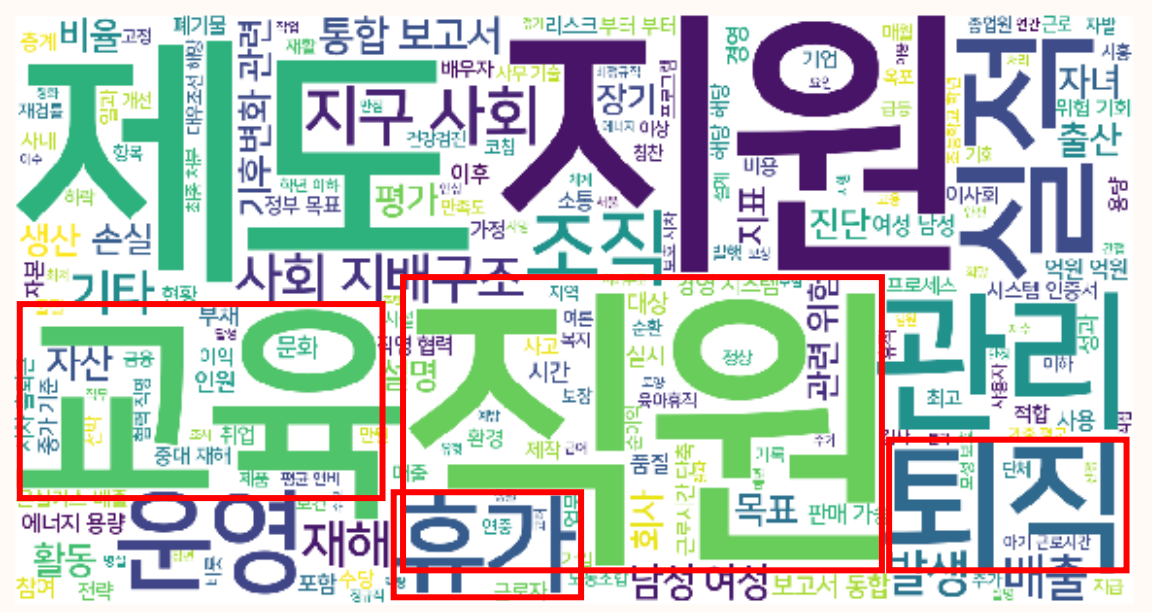
- 두 기업 모두 '직원', '지원'이 높은 빈도로 나타남
- > ESG에 힘쓰는 기업은 대체적으로 직원과 관련된 활동을 많이 함
  - > ESG에 힘쓰는 기업은 다양한 지원 사업을 진행하는 것으로 보임

## 03. 분석 결과

| ESG 등급이 높은 기업(KB 금융)|



| ESG 등급이 낮은 기업(KB 금융)|

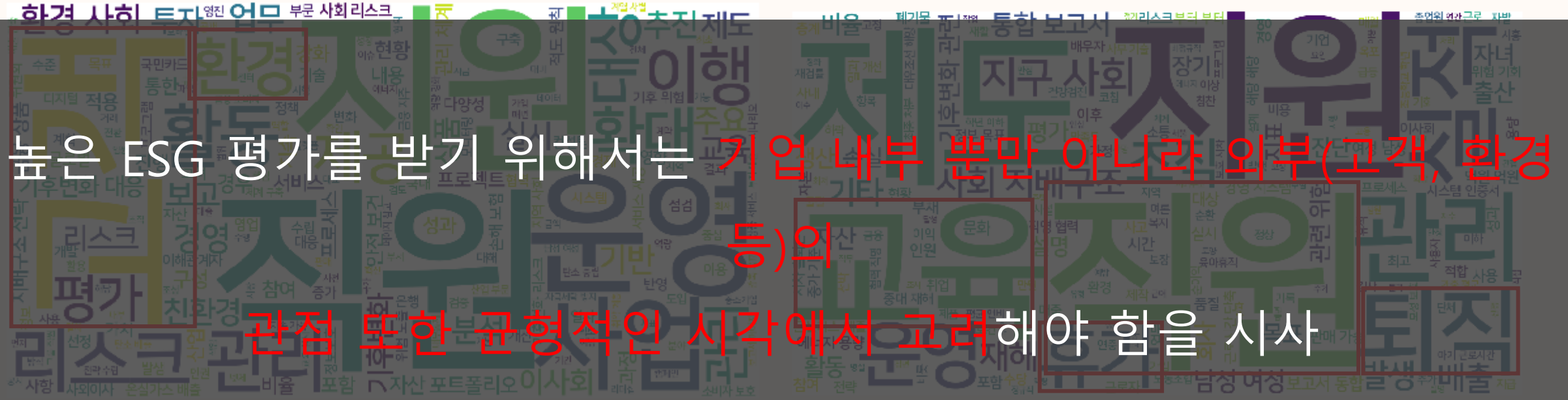


- ESG 등급이 높은 기업은 사회, 환경적 측면과 같이 기업 외부에 초점을 두고 있으며, 낮은 기업은 지배구조 측면 등 기업 내부에 초점을 두고 있음
  - > ESG 등급이 높은 기업은 고객, 환경과 같이 기업 외부의 키워드도 다수 보였으나,
  - > ESG 등급이 낮은 기업은 교육, 직원, 휴가, 퇴직 등 기업 내부의 지배구조 측면의 키워드가 다수 확인

### 03. 분석 결과

| ESG 등급이 높은 기업(KB 금융)|

| ESG 등급이 낮은 기업(KB 금융)|



- ▶ ESG 등급이 높은 기업은 사회, 환경적 측면과 같이 기업 외부에 초점을 두고 있으며, 낮은 기업은 가버넌스 측면 등 기업 내부에 초점을 두고 있음
  - > ESG 등급이 높은 기업은 고객, 환경과 같이 기업 외부의 키워드도 다수 보였으나,
  - > ESG 등급이 낮은 기업은 교육, 직원, 휴가, 퇴직 등 기업 내부의 가버넌스 측면의 키워드가 다수 확인

# 00.

## 부록

### Pdf 텍스트 추출 코드

```
from io import StringIO
from pdfminer.converter import TextConverter
from pdfminer.layout import LAParams
from pdfminer.pdfdocument import PDFDocument
from pdfminer.pdfinterp import PDFResourceManager, PDFPageInterpreter
from pdfminer.pdfpage import PDFPage
from pdfminer.pdfparser import PDFParser
```

```
def read_pdf_PDFMINER(pdf_file_path):
    output_string = StringIO()
    with open(pdf_file_path, 'rb') as f:
        parser = PDFParser(f)
        doc = PDFDocument(parser)
        rsrcmgr = PDFResourceManager()
        device = TextConverter(rsrcmgr, output_string, laparams=LAParams())
        interpreter = PDFPageInterpreter(rsrcmgr, device)
        for page in PDFPage.create_pages(doc):
            interpreter.process_page(page)
    return str(output_string.getvalue())
```

```
v=read_pdf_PDFMINER('D:/Data/esg report/대우조선해양.pdf')
```

감사합니다.