
저자 (Authors)	이혜림, 황진호, 유도열, 김종우
출처 (Source)	한국지능정보시스템학회 학술대회논문집 , 2021.6, 102-103 (2 pages)
발행처 (Publisher)	한국지능정보시스템학회 Korea Intelligent Information Systems Society
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10588120
APA Style	이혜림, 황진호, 유도열, 김종우 (2021). BERT와 다중회귀분석법을 이용한 ESG등급 예측. 한국지능정보시스템학회 학술대회논문집, 102-103.
이용정보 (Accessed)	한양대학교 166.***.182.218 2022/02/28 22:23 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

D3.5 BERT와 다중회귀분석법을 이용한 ESG등급 예측

이혜림
한양대학교 경영대학
경영학부
easyoffbang@
hanyang.ac.kr

황진호
한양대학교 경영대학
경영학부
appler@hanyang.ac.kr

유도열
한양대학교 일반대학원
비즈니스인포매틱스학과
tnfuswk@hanyang.ac.kr

김종우
한양대학교 경영대학
경영학부
kjjw@hanyang.ac.kr

Abstract - 비재무적 성과를 나타내는 ESG 등급에 대한 투자자들의 관심이 증대되는 상황에서, 기업이 ESG 등급을 효율적으로 향상하는 것이 중요해졌다. 따라서 본 연구에서는 ESG 등급을 예측하는 모델을 제안하고 평가함과 동시에 등급 결정에 큰 영향을 미치는 요소들을 식별하고자 했다. 한국기업지배구조원(KCGS)의 연도별, 분기별의 두 가지로 나뉘는 KCGS 등급산정 방식에 따라 다중 회귀분석과 BERT 모델을 사용하여 ESG 등급을 예측하였다. 다중 회귀분석은 KCGS에서 공표한 ESG 핵심지표를 토대로 데이터를 수집하였고, BERT 모델은 ESG 관련 뉴스 데이터를 수집한 후 예측을 진행하였다. 본 연구 결과를 통해 기업들은 예측된 ESG 등급을 자기 평가 지표로 삼음과 동시에 보다 유의한 영향을 미치는 Key Index에 중점을 두어 효율적인 등급 향상이 가능할 것으로 보인다.

Key Terms - ESG등급예측, 다중 회귀분석, BERT, 다중라벨 분류

이 논문은 ‘대학혁신지원사업 비즈니스 랩 기반의 프로젝트 학기제’의 지원을 받아서 수행된 연구임

I. 서론

기업의 비재무적 가치에 대한 관심이 높아지면서 기업의 ESG 등급과 재무적 성과의 상관관계를 연구하는 시도들이 다양하게 이루어지고 있다. 이러한 연구들은 다양한 결과를 보이나 최근에는 양(+)의 관계를 보고하는 연구들이 늘어나고 있다. 이처럼 ESG 등급의 중요성이 커지는 추세에 따라 본 연구는 ESG등급을 예측하는 모델을 제안하고 평가함과 동시에 ESG 등급 결정에 큰 영향을 미치는 요소들을 식별하고자 했다.

예측 모델의 종속변수는 한국기업지배구조원(이하 KCGS)에서 공표하는 등급을 기준으로 하였다. KCGS의 ESG등급 산정방식은 크게 두 가지로 나뉜다. 핵심지표(Key Index) [표1]에 따라 매년

정기적으로 산정하거나 기업의 ESG관련 이슈에 따라 분기마다 조정한다. 이 평가 과정에서 전문가가 분석하는 등 다소 수동적인 방법을 사용하고 있다.

본 연구에서는 이런 한계점을 개선하고자 Key Index 데이터를 기반으로 하는 다중회귀 분석모델과 ESG 관련 뉴스데이터를 최신 자연어처리 모델인 BERT에 입력하는 모델로 나누어 두 가지 방식으로 ESG등급 예측을 자동화하고자 한다. 나아가, 본 연구는 기업의 효율적인 등급 향상을 위한 기초 자료로써 활용될 것으로 기대된다.

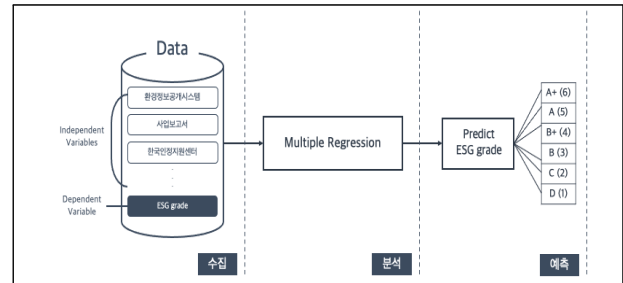
<표 1> ESG 평가 핵심지표 중 일부

E(환경) 핵심지표	S(사회) 핵심지표	G(지배구조) 핵심지표
환경경영인증 보유 여부	기간제 근로자 비중	이사회 운영 실적
국제 이니셔티브 참여 여부	제품 및 서비스 안전성 인증	이사회 독립성
온실가스 배출량	공정거래 프로그램	이사회 내 전문 위원회

II. 연구 방법

2-1. Key Index 기반 다중회귀 분석법

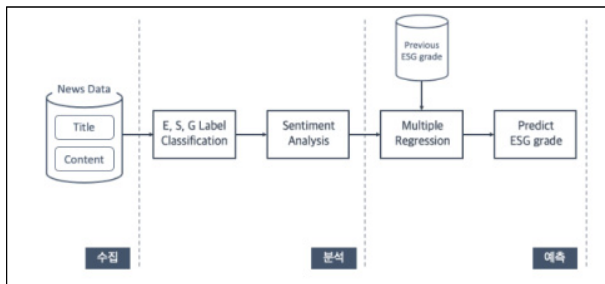
Key Index 기반 다중회귀 분석법은 <그림 1>과 같은 절차를 통해 진행되었다. 사업보고서 등의 문서에서 텍스트마이닝(Text mining)을 통해 핵심지표(Key Index)에 해당하는 데이터를 수집하였다. 2016년부터 2019년까지 4개년도의 데이터를 포함하며, 수치형 변수는 산업별 특성을 반영하고자 산업별 평균값으로 나누었다. 종속변수로는 범주형 변수인 KCGS에서 공개한 4개년도의 등급을 연속된 숫자 형으로 바꾸어 설정하였다.



<그림 1> Key Index 기반 다중 회귀분석 연구 절차

2-2. 뉴스 데이터 기반 BERT 모델

뉴스 데이터 기반 BERT모델은 <그림 2>와 같은 절차를 통해 진행되었다. 국내 뉴스 플랫폼 네이버 뉴스에서 환경, 노동, 사회 일반, 인권복지 섹션으로 범위를 한정한 후, 2016년부터 2019년까지 4개년도의 ESG 관련 기사의 제목, 본문, 네티즌 감성표현을 수집하였다. 수집된 섹션에 맞게 E, S, G의 라벨을 부여하고 네티즌 감성표현의 긍정·부정 비율에 따라 극성을 부여하였다. 수집한 네이버 뉴스 데이터에 더하여, KCGS에서 공표한 ESG 모범규준 데이터와 통계청에서 공개한 기사 감성 라벨링 데이터를 학습데이터에 추가하였다.



<그림 2> 뉴스데이터 기반 BERT 모델 연구 절차

본 연구의 BERT 모델은 SKTBrain에서 공개한 한국어 특화 사전 학습 모델인 KoBERT를 활용하였다. ESG 다중 라벨 분류 모델은 2016년 1월부터 2018년 12월까지의 뉴스 데이터와 ESG 모범규준 데이터를 합한 2,000개의 데이터, 감성분류 모델은 2017년 1월부터 2018년 12월까지의 뉴스 데이터와 통계청의 감성 데이터를 합한 약 8,400개의 데이터로 Fine-Tuning을 진행하였다. 검증 데이터는 2019년의 기업 뉴스데이터를 사용하였으며 모델의 최종 결과값으로 기사에 E, S, G라벨과 0(부정), 1(중립), 2(긍정)의 감성점수가 부여된 기사가 산출된다. 이후 기업마다 E, S, G 별 평균 감성 점수를 구한 후, 전년도 ESG 등급과 함께 독립변수로 적용해 ESG 등급을 예측하는 회귀식을 도출하였다.

III. 분석 결과

3-1. Key Index 기반 다중회귀 분석법

E(환경)에 해당하는 Key Index 15개 중 변수 선택법을 통해 8개를 선택하였으며, 단계별 설명력이 향상되어 모형설계에 문제가 없다고 보인다. 또한 분산팽창지수(VIF) 값이 모두 5 이하로 다중공산성이 없다고 판단된다. 분석 결과, E(환경) 등

급에 대한 단계별 다중 회귀분석의 회귀모형은 유의하였으며 (Prob(F-statistics)<.001) 모형의 설명력은 76.7%로 나타났다. <표2>

<표 2> E(환경) 다중 회귀분석 결과

Model	Coefficients	Standard error	t	p-value
(constant)	0.5685	.036	15.877	.000***
온실가스 배출량	-.0639	.014	-4.642	.000***
ISO14001보유 여부	.0539	.012	4.537	.000***
환경정보 공시 여부	.1478	.031	4.820	.000***
UNGC참여 여부	.1433	.054	2.629	.009**
CDP참여 여부	.2908	.050	5.838	.000***
GMS보유 여부	.0029	.001	2.778	.006**
환경평가 체계 구축 여부	.1990	.036	5.574	.000***
지난 연도 E등급	.7308	.015	49.254	.000***

$R^2 = .769$, Adjusted $R^2 = .768$

S(사회)에 해당하는 Key Index 8개 중 변수 선택법을 통해 5개를 선택하였으며, 단계별 설명력이 향상되어 모형설계에 문제가 없다고 보인다. 또한 분산팽창지수(VIF)값이 모두 3 이하로 다중공산성이 없다고 판단된다. 분석 결과, S 등급에 대한 단계별 다중 회귀분석의 회귀모형은 유의하였으며 (Prob(F-statistics)<.001) 모형의 설명력은 55.3%로 나타났다. <표3>

<표 3> S(사회) 다중 회귀분석 결과

Model	Coefficients	Standard error	t	p-value
(constant)	0.9424	.051	18.350	.000***
동반성장지수	.1189	.014	8.372	.000***
공정거래프로그램 운영 여부	.4587	.059	7.809	.000***
매출액 대비 기부금 비율	.0752	.015	4.960	.000***
정보보호 인증 여부	.0054	.051	2.016	.044*
지난 연도 S등급	.6763	.018	37.623	.000***

$R^2 = .554$, Adjusted $R^2 = .553$

3-2. 뉴스 데이터 기반 BERT 모델

ESG 다중라벨 분류모델은 0.87의 Accuracy와 0.34의 Loss 값을 보였으며 감성분류모델은 0.88의 Accuracy와 0.30의 Loss값을 보였다. 위 모델로 추출한 뉴스 기사의 감성 평균점수와 지난 연도 등급을 독립변수로 하여, Ridge Regression (alpha=0.05)으로 E, S, G 각각 회귀식을 도출하였다. 각 회귀 모델의 결과는 다음과[표4] 같다.

<표 4> 뉴스 데이터 기반 BERT 모델 회귀분석 결과

Label	Independent variables	Coefficients	RMSE	MAE
E	지난 연도 E등급	0.132	0.549	0.307
	E관련 뉴스 기사 감성 점수	0.949		
S	지난 연도 S등급	0.247	0.669	0.527
	S관련 뉴스 기사 감성 점수	0.819		
G	지난 연도 G등급	0.207	0.675	0.522
	G관련 뉴스 기사 감성 점수	0.758		