

상장기업 ESG 평가 모델 개발 계획서

2022 미래에셋증권 빅데이터 페스티벌

» 데이터 안내

- 01 데이터 정의
- 02 사용데이터 구분
- 03 데이터 수집 및 전처리 계획

» ESG 평가모델 개발 구현 계획

- 01 선정 모델 및 설명
- 02 평가 시뮬레이션

» ESG 평가 모델 Flowchart

» 평가 모델을 활용한 비즈니스 제안

쿠빅 팀

김지호

박지우

이병주

가. 데이터 안내

01 데이터 정의

- 비재무 데이터

ESG 각 부문에 따라 평가에 필요한 항목들을 지정하여 다음의 수치 데이터들을 선정하였습니다.

ESG 평가에 활용할 비재무 데이터 목록은 다음과 같습니다.

기본평가			심화평가
E 환경 평가 요소	S 사회 평가 요소	G 지배구조 평가 요소	ESG 통합 등급 (가산점)
온실가스 배출량	정규직비율	현금배당성향	사회적 기업 인증 여부
에너지 사용량	제재 등과 관련된 사항(건수)	현금배당수익률	지속가능경영보고서 발행여부
폐기물 배출량	직원 성비	대주주 지분율	지배구조보고서 공시 여부
환경법규현황 (건수)	여성급여 비율	소액주주지분율	
재생에너지사용량	산업재해율	사외이사 비율	
원료 사용량		이사 보수	
용수 사용량		감사 보수	
		임원 성비	
		직원수	
		직원 성비	

- 재무 데이터

기업의 ESG 경영이 가시적인 성과를 냈는지 확인하기 위해 재무데이터도 사용할 예정입니다.

ESG 등급과 재무비율의 선형적인 관계를 연구한 논문을 참고하여 9가지 재무 비율을 선정하였고 기존 대회에서 제공된 재무 데이터를 활용하여 다음의 9가지 지표들을 산출하였습니다.

각 등급과 유의미하다고 증명된 feature들을 구분하여 사용할 예정입니다.

변수명	변수 설명	산출식	E	S	G
OPOS	매출액영업이익률	영업이익/매출액		○	○
NOE	자기자본순이익률	당기순이익/자본금		○	○
CR	유동비율	유동자산/유동부채	○	○	○
DR	부채비율	부채총액/자본금	○	○	○
PER	주가수익비율	한국거래소 krx 크롤링	○	○	○
PBR	주가순자산비율	한국거래소 krx 크롤링	○		○
AGE	기업연령	log (총자산)			○
SIZE	자산규모	2022-상장연도	○	○	○
KIS	신용등급	한국거래소 krx 크롤링	○	○	○

참고 논문 : 백상미, 최정미 (2021). ESG 평가등급과 기업특성에 관한 연구

02 사용데이터 구분

업종마다 ESG 경영에 대해 이해관계자들이 요구하는 바가 다릅니다. 가령, IT기업은 화학기업에 비해 개인정보보안 이슈에 특히 민감하게 대처해야하며 평가 기준 또한 더 엄격하게 적용받아야할 것입니다.

이처럼 업종별 핵심 이슈를 파악하는 것은 기업 입장에서는 효율적인 경영전략을 세우는데 도움이 되고, 평가자의 입장에서는 보다 공정한 평가를 가능하게 합니다. 저희는 성격이 다른 산업들의 ESG 등급을 일관된 잣대로 평가하는 것은 바람직하지 않다고 판단하여 우선 기업들을 업종별로 구분하여 접근하고, 각 산업의 특성에 따라 평가 가중치를 달리한 모델을 적용하기로 하였습니다.

- 업종을 어떻게 구분 할 것인가?

ESG 핵심 리스크를 바탕으로 업종을 분류하고자 합니다. 회사의 리스크는 그 실체가 모호하기 때문에 실질적인 평가를 하려면 이를 구체화 시켜야합니다. 저희는 모건스탠리(MSCI 지수)가 발표한 10개의 테마와 35가지 핵심이슈와 한국거래소가 발표하는 지수들을 참고하여 ESG 관점에서 중요하게 다뤄지는 대토픽과, 이해관계자들에게 중요도가 높았던 토픽을 포함하도록 산업을 구분 하였습니다.

TRAIN SET과 TEST SET의 기업을 한국거래소에 명시된 산업으로 분류한 결과 총 39개의 분류가 있었는데, 중복되거나 포함관계에 있는 경우가 많았습니다. 업종별로 모델 학습에 필요한 데이터 수를 확보하기 위해 다음과 같이 9개의 업종으로 묶었습니다.

- 업종 분류 : 산업별 key 이슈에 따라

에너지 화학	화학, 전기가스업
금융	은행, 증권, 보험
헬스케어	제약, 의약품, 의료,정밀기기
커뮤니케이션 서비스	컴퓨터서비스, 디지털 콘텐츠, 오락.문화, 소프트웨어, 방송서비스, 통신업, 인터넷 , 서비스업 ,기타서비스
반도체 및 소재 부품	전기전자, 반도체, it부품, 일반전기전자, 통신장비, 기계장비, 기계
운송	유통업, 운수창고업, 유통, 운송장비
소비재	숙박.음식, 음식료품,음식료.담배, 농업, 임업 및 어업, 섬유의복, 기타제조업
산업재	철강금속, 비금속광물, 종이목재, 종이. 목재, 비금속, 금속
건설	건설업, 건설

에너지 화학 "탄소 감축"

지속적으로 증가하는 에너지 수요와 더불어 화석연료 사용에 따른 환경오염의 영향으로, 친환경적이면서도 경제적인 새로운 에너지원에 대한 연구가 중요해지고 있습니다. 특히 탄소 감축을 위해 새로운 물질 개발부터 최첨단 엔지니어링에 이르는 연구를 중점적으로 다루는 산업들은 환경적 요소를 더 민감하게 반영해야 한다는 점에서 에너지 화학산업으로 구분하였습니다.

금융 "지속가능한 금융"

금융회사는 수익성을 추구하는 영리기업인 동시에, 자금 순환의 중개자로서 실물 경제를 뒷받침하는 역할을 수행합니다. 개별 금융기업의 경영은 국가 전체의 금융시스템의 안전성과 직결된다는 측면에서 다른 기업들과는 본질적으로 다르다고 판단하였고 은행, 보험, 증권 등을 금융업으로 분류하고 따로 다루기로 하였습니다.

헬스케어 "디지털 헬스케어"

코로나의 확산으로 오프라인 의료기관과 인력 부족 문제가 대두되면서, 디지털 헬스케어 산업의 수요가 크게 증가하였습니다. 특히 빅데이터와 시가 의료산업과 융합되면서 디지털 헬스케어 분야는 더욱 성장하고 있는데, 이러한 산업 변화를 반영하여 의약품, 제약, 의료기기 관련 산업을 '헬스케어'로 분류하였습니다.

커뮤니케이션 서비스 "플랫폼 비즈니스"

전세계적으로 온라인으로 시장과 소통하는 형태의 플랫폼 비즈니스가 각광 받고 있습니다. 산업적 가치가 변화에 따라 인터넷 플랫폼, SNS 등 디지털 콘텐츠를 제공하는 기업들을 커뮤니케이션 서비스 분야로 분류했는데, 다른 산업에 비해 성장성이 높고 인프라, 정보 보안, 데이터 관리 및 분석에 관한 리스크에 민감하다는 특징이 있습니다.

반도체 및 소재·부품 산업 "산업 간 연계성"

반도체, 소재산업은 여러 기술과 소재를 필요로 하기 때문에 국내외 기업과의 연계가 중요합니다. 이 산업은 소자를 직접 생산하는 소자산업과 이 소자산업을 가능하게 하는 후방산업으로 이루어지는데, 최근, 외교적 이유로 후방산업이 위기를 맞아 반도체 산업이 큰 타격을 입는 사례가 있었습니다. 이러한 산업적 특성에 주목하여 반도체 산업을 타산업들과 구분하고자 합니다.

운송 "인권"

코로나 이후 폭증한 물류량에 노동자들이 과로에 시달리면서 택배나 물류파업의 사례가 증가하고 있습니다. 운송업의 경우, 타산업 대비 자동화가 덜 진행된 상태이기 때문에 인권, 노동, 안전, 보건 등 사람과 관련된 이슈가 특히 중요하게 여겨집니다. 따라서 노동 환경 및 직원 안전 리스크가 특히 중요한 운송업은 따로 분류하는 것이 적절하다고 판단하였습니다.

건설업 "산업재해"

산업재해가 가장 많이 발생하는 산업은 건설분야입니다. 주로 부적절한 노동관행이 산업재해의 원인으로 꼽히는데, 따라서 건설업의 사회부문이나 지배구조는 타 업종들과 다른 비중으로 평가되어야 할 것입니다.

산업재 "분쟁광물"

전세계적으로 천연광물을 둘러싼 분쟁이 많이 일어나고 있습니다. 미국과 유럽은 법으로써 광물 원산지 표기를 의무화하였고 국내 기업 LG에너지솔루션은 원재료의 구매단계에서 분쟁광물은 배제시키는 등, 산업재 기업들은 분쟁지역의 노동력 착취와 인권침해를 방지하기 위한 ESG 경영을 실천하고 있습니다. 따라서 원자재 거래와 직접 관련된 업종은 산업재 산업으로 따로 분류하고자 합니다.

소비재 "그린슈머"

최근 친환경 의식을 갖춘 소비자들이 늘면서 이러한 '그린슈머'들을 겨냥한 마케팅도 활발하게 이루어지고 있습니다. 기업들은 생산 과정에서의 환경오염을 최소화시키는 한 편, 가습기 살균제 사고와 비슷한 일이 반복되지 않도록 소비자 안전 또한 중요시 여겨야 합니다. 이처럼 소비자의 영향을 많이 받는 특정 산업들은 소비재 산업으로 분류하여 평가하고자 합니다.

03 세부수집 방법 및 전처리 - feature 별 정리

- E 평가 항목

feature	type	수집 방법	전처리
온실가스 배출량	연속	NGMS(국가온실가스 종합관리시스템)	기업마다 scale차이 큼 → standard scaler
에너지 사용량	연속	NGMS(국가온실가스 종합관리시스템)	표준화 (standard scaler)
폐기물 배출량	연속	환경정보공개시스템	표준화 (standard scaler)
환경법규현황 (건수)	연속	지속가능경영보고서 (기업별)-selenium	이상치가 많음 → robust scaler
재생에너지사용량	연속	지속가능경영보고서 (기업별)-selenium	표준화 (standard scaler)
원료 사용량	연속	환경빅데이터플랫폼	표준화 (standard scaler)
용수 사용량	연속	환경빅데이터플랫폼	표준화 (standard scaler)

- S 평가 항목

feature	type	수집 방법	전처리
정규직비율	비율	dart.fss open API	
제재 등과 관련된 사항(건수)	연속	dart.fss open API	표준화 (standard scaler)
직원 성비	비율	dart.fss open API	
여성급여 비율	비율	dart.fss open API	정규화 (MinMax Scaler)
산업재해율	비율	고용노동부 제공 데이터	정규화 (MinMax Scaler)

- G 평가 항목

feature	type	수집 방법	전처리
현금배당성향	비율	dart.fss open API	정규화 (MinMax Scaler)
현금배당수익률	비율	dart.fss open API	정규화 (MinMax Scaler)
대주주 지분율	비율	dart.fss open API	정규화 (MinMax Scaler)
소액주주지분율	비율	dart.fss open API	정규화 (MinMax Scaler)
사외이사 비율	비율	dart.fss open API	
이사 보수	연속	dart.fss open API	정규화 (MinMax Scaler)
감사 보수	연속	dart.fss open API	정규화 (MinMax Scaler)
임원 성비	비율	dart.fss open API	
직원수	연속	dart.fss open API	정규화 (MinMax Scaler)
직원 성비	비율	dart.fss open API	

표준화 : 주로 데이터가 가우시안 분포를 띠는 경우 많이 사용 → standard scaler, robust scaler 등
 정규화 : 데이터를 0~1 사이 값으로 매핑함 데이터의 분포를 모를 때 사용 → minmaxscaler 가 대표적

- 재무 데이터

feature	type	수집 방법	전처리
매출액영업이익률	비율	대회 제공 데이터	영업이익/매출액
자기자본순이익률	비율	대회 제공 데이터	당기순이익/자본금
유동비율	비율	대회 제공 데이터	유동자산/유동부채
부채비율	비율	대회 제공 데이터	부채총액/자본금
자산규모	연속	대회 제공 데이터	한국거래소 krx 크롤링
기업연령	연속	dart.fss open API - 기업개황	한국거래소 krx 크롤링
신용등급	범주 (AAA ~)	dart.fss open API	log (총자산)
주가수익비율	비율	한국거래소 krx - selenium	2022-상장연도
주가순자산비율	비율	한국거래소 krx - selenium	한국거래소 krx 크롤링

- 심화 평가

feature	type	수집 방법	전처리
사회적 기업 인증 여부	범주	한국사회적기업진흥원 공시자료 - selenium	one-hot encoding
지속가능경영보고서 발행여부	범주	krx esg 포털 - selenium	one-hot encoding
지배구조보고서 공시 여부	범주	krx esg 포털 - selenium	one-hot encoding

위의 3가지 항목은 종합 등급 산출 단계에서 가산점 형태로 활용 예정인 심화 평가 요소입니다.

기업과 투자자간 소통채널의 기능을 한다는 점에서 기업에게 공시 활동은 매우 중요합니다. ESG 경영과 관련한 내용은 기업이 자체적으로 공개하는 지속가능경영보고서나 지배구조보고서에서 확인할 수 있는데, 아직까지는 이러한 보고서들의 공시가 선택 사항으로 규정되어 있습니다.

따라서 저희는 공시 의무가 아직 없음에도 불구하고 대외적으로 투자자들이 열람할 수 있도록 노력하는 기업들에게 가산점을 부여하고자 합니다. 이 두 가지 보고서(지속가능경영보고서/ 지배구조 보고서)의 공시여부와 더불어 한국사회적기업진흥원에서 고용노동부와 함께 인증서를 발급해주는 사회적 기업 인증 여부 또한 가산점 항목으로 활용할 예정입니다.

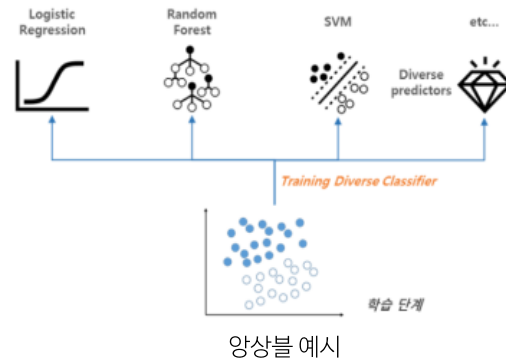
사용 예정인 python library
 dart.fss open API : dart 공시 정보 크롤링
 selenium : 보고서 정보 크롤링
 pandas : 데이터 프레임 활용
 scikit-learn : 머신러닝 모델 사용
 tensorflow, keras : 딥러닝 모델 사용

나. ESG 평가모델 개발 구현 계획

01 선정 모델 및 설명

E, S, G 각 등급을 산출하는 과정에서는 1부터 9까지의 등급 label을 내주는 다음의 다중 분류 모델들을 사용할 예정입니다.

- 로지스틱회귀 분석
- K-최소 근접 (K-Nearest Neighbor) 알고리즘
- 심층 신경망 (Deep Neural Network)
- 의사결정트리
- 앙상블 기법 (배깅, 랜덤포레스트, 부스팅)

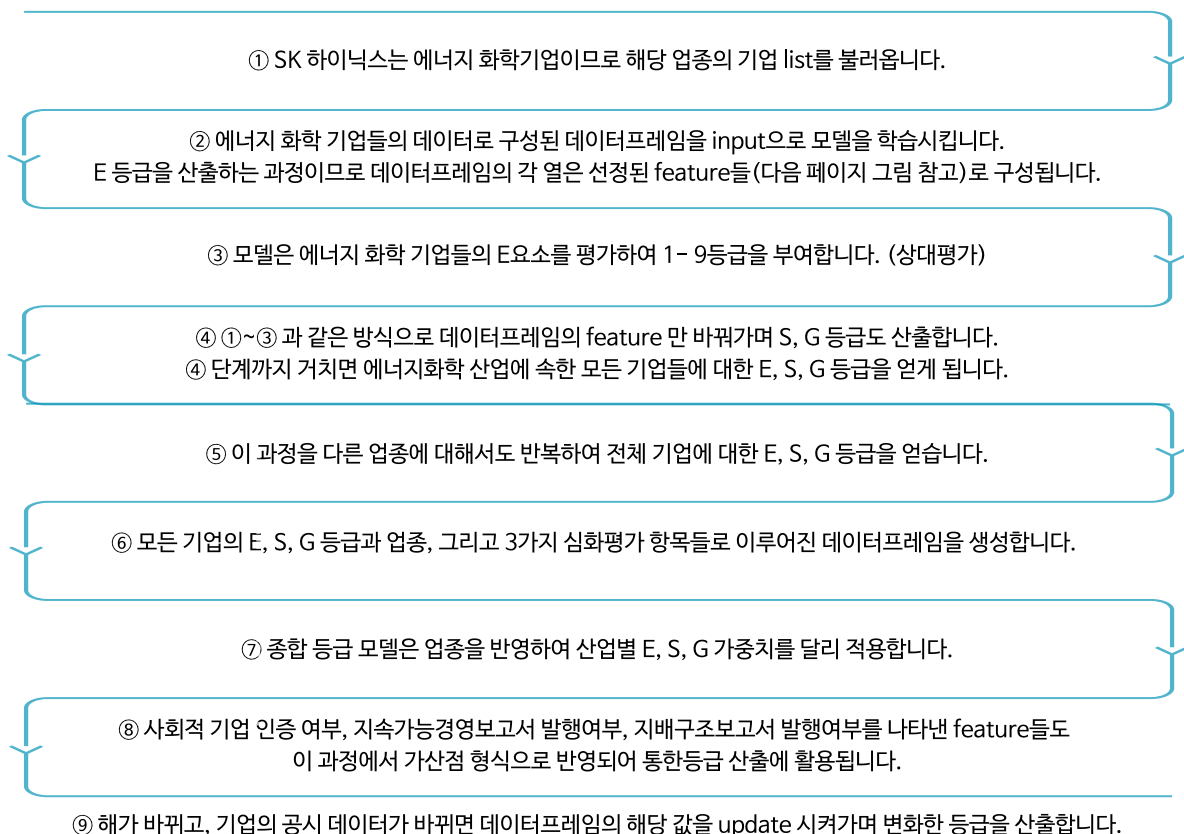


E, S, G 부문에 대한 각각의 feature들을 입력 받은 후 9개의 업종별로 나눠 학습을 시키면 분류 모델을 통해 E, S, G 등급이 산출됩니다. 기업들의 업종을 구분하여 학습시키기 때문에 각 기업의 E, S, G 등급은 동일한 업종 내에서 상대평가 방식으로 계산됩니다. 기업 input 값은 업종에 따른 가중치로 계산되기 때문에 업종의 특수성을 고려할 수 있게 됩니다.

예를 들어 에너지 화학산업은 다른 업종에 비해서 절대적인 에너지 사용량이 많아 모두 낮은 E 등급을 받는 것이 아니라 같은 업종 내에서 상대적으로 에너지 사용량이 적다면 높은 E 등급을 받을 수 있을 것입니다.

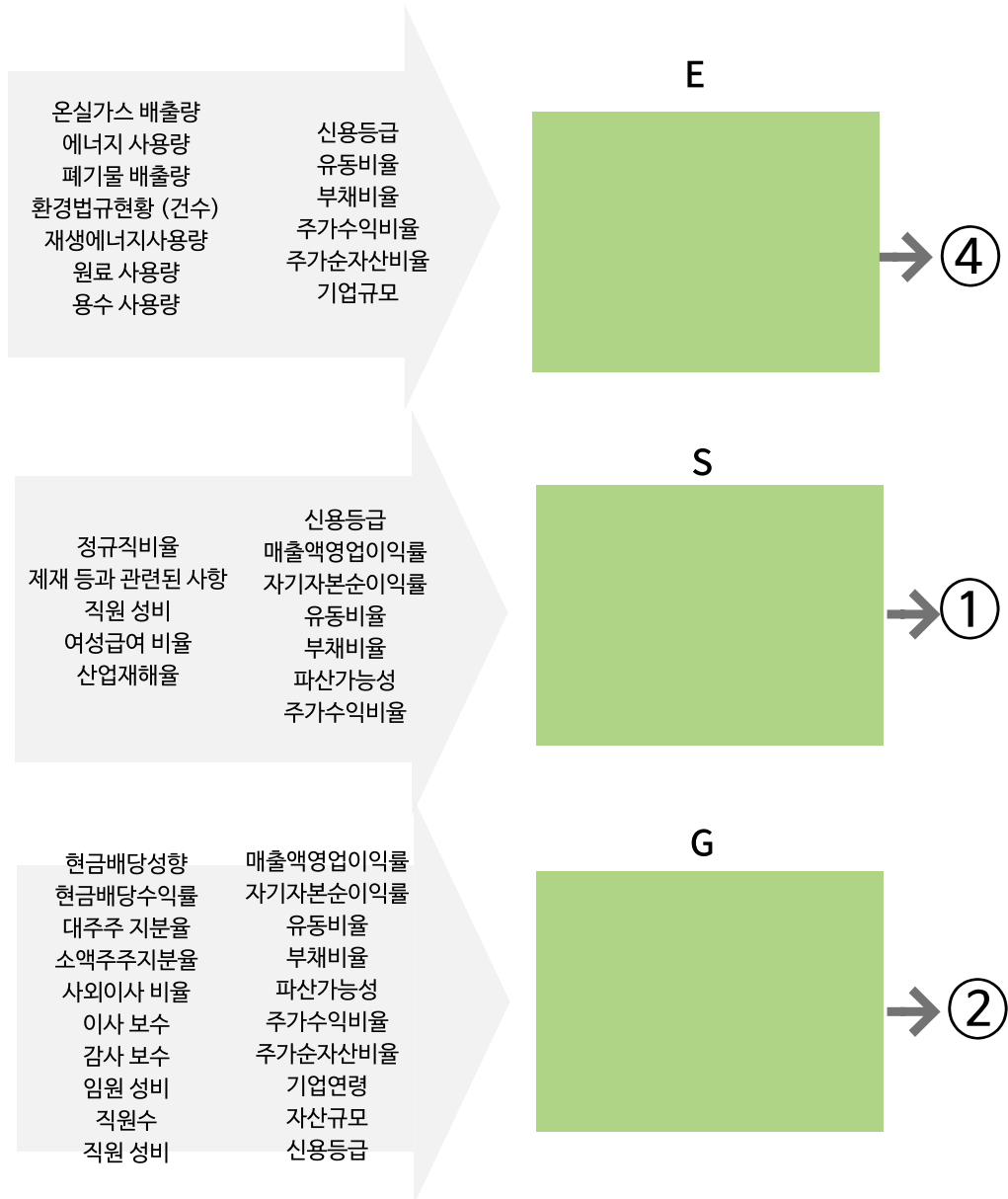
E, S, G 3가지 등급이 산출된 다음에는 업종에 따라 서로 다른 가중치를 적용한 다음 사회적 기업 인증 여부, 지속가능경영보고서 발행여부, 지배구조보고서 공시 여부를 가산점으로 주어 종합 ESG등급을 계산할 것입니다.

02 평가 시뮬레이션

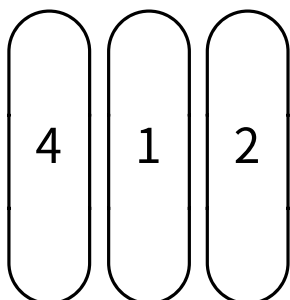




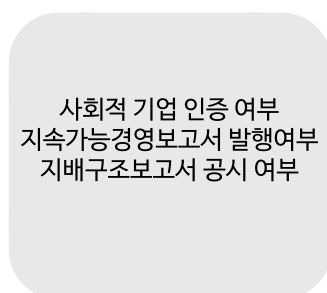
SK 하이닉스
에너지 화학



E S G



+



종합 ESG 등급

②

업종에 따라 다른 가중치 부여

가산점

03 Flow Chart



STEP 1 데이터 수집

ESG등급을 영역별로 평가하는 데 있어 필요한 데이터를 수집하는 단계입니다. 사용할 데이터는 업종분류 데이터, 환경데이터(온실가스, 에너지 등), 재무데이터, DART 데이터, 사회적 기업 인증여부 등이 있습니다. 저희는 산업에서의 핵심 키워드를 가지고 업종을 분류했기 때문에 기존에 있던 업종 분류 데이터를 참고자료로 사용할 것입니다.

STEP2 데이터 전처리

1.업종 분류 처리

저희는 업종 분류를 자체적으로 진행했습니다. ESG 평가 기관들이 발표한 핵심 key 이슈를 산업과 연관시켜 이를 토대로 에너지 화학 /금융 /헬스케어 /커뮤니케이션 서비스 /반도체 및 소재부품 /운송 /소비재 /산업재 /건설의 9개 업종으로 분류했습니다.

2.데이터 분포에 따른 전처리

신용등급 데이터는 범주형 변수로, AAA와 같은 등급으로 되어 있기 때문에 이를 순서형 변수로 변환하여 이용할 것입니다. 예를 들어 AAA - 1등급과 같이 바꾸고자 합니다. 그리고 사회적 기업인증 여부, 지배구조보고서 공시여부, 지속가능경영보고서 발행여부는 여부에 따라 0, 1 로 나누어서 범주형 변수처리를 합니다.

STEP 3 영역별 평가

3가지 모델로 E, S, G 등급을 각각 산출하는 단계입니다. 모델 별로 입력 받을 feature가 다르므로 전체 데이터셋에서 모델에 필요한 데이터만을 추출하여 학습합니다. 모델은 로지스틱회귀 분석, K-최소 근접 알고리즘, 심층 신경망, 의사결정트리와 다양한 앙상블 기법을 시도하여 예측 성능이 가장 높은 방법을 선택합니다.

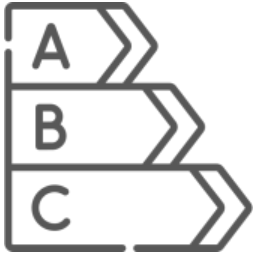
STEP 4 종합등급 평가

기업별로 E,S,G 등급이 각각 산출되면 이를 이용해서 통합 등급을 매깁니다.

이 단계에서 심화 평가 요소로 사회적 기업 인증 여부 / 지속가능경영보고서 발행여부/ 지배구조보고서 공시 여부 가 가산점 요소로 활용됩니다. 회귀 모델에서 기업별 E/S/G 등급과 3가지 공시 여부 그리고 업종에 대한 데이터를 입력 받고, 모델에서는 통합 등급을 산출하는 과정에서 업종에 따른 가중치를 학습시켜 최종 등급을 산출합니다.

다. 평가 모델을 활용한 비즈니스 제안

01 ESG 등급 표시제



제안 배경

평가기관에서는 ESG 등급을 매기고 있지만 정작 소비자들은 이에 대해 알 수 있는 방법이 없습니다. 이러한 등급의 중요성은 기업들과 투자자들만 실감할 수 있는 것이고 정작 소비자들이 소비를 할 때는 ESG 경영에 대한 정보를 참고하기가 어려운 것입니다. 결국 ESG 등급은 소비자들은 신경쓰지 않는 등급제로 될 수도 있습니다.

내용

저희가 제안하고자 하는 서비스는 에너지 효율등급표시제와 유사한 형태의 "ESG 등급 표시제"입니다. 에너지 효율등급 표시제는 에너지 사용 증가로 인해 제품마다 에너지 효율 등급을 붙여서 소비자가 제품을 살 때 참고할 수 있도록 했습니다.

이와 같이 제품을 만든 기업의 ESG 등급을 표시하여 소비자가 이를 참고해서 소비를 할 수 있게 만드는 서비스를 제공할 수 있습니다.

기대 효과

이 서비스를 통해 기업들이 보여주기가 아니라 실제로 ESG 경영에 신경쓰고 노력하는 분위기를 조성할 수 있습니다.

소비자들은 에너지효율등급표시제를 보고 제품을 구매하듯 ESG등급을 보고 제품을 사면서 '그린슈머'를 넘어선 'ESG슈머'가 될 수 있습니다. 따라서 이 서비스는 단순한 서비스만이 아니라 소비자들에게서 현명한 소비를 이끌어 수 있고 기업들의 현명한 경영을 이끌 수 있습니다.

02 ESG 등급 모의 예측 서비스



제안 배경

ESG 등급은 기업이 국내외 투자를 유치할 때 핵심적인 정보로 작용하기 때문에 이 등급 평가는 기업의 미래와 직결되어 있다고 볼 수 있습니다. 그러나 평가기관마다 가중치가 달라 하나의 기업에 대해서도 평가 점수가 달라지기 때문에 기업의 입장에서는 중요한 평가를 앞두고 경영에 있어서 어떤 부분을 점검해야 할 지 알아낼 방법이 없습니다.

내용

서비스를 의뢰한 기업이 전년도 재무 성과와 함께 올해 기대 성과 혹은 목표 성과를 제출한다면 해당 기업은 올해 기대 등급과 목표에 도달했을 때의 예상 등급을 안내 받게 됩니다. 이렇게 예측된 등급을 참고하면 기업은 현재 목표의 방향성을 점검할 수 있고, 평가기관의 공식 평가를 받기 전 보완해야 할 부분을 구체적으로 파악할 수 있는 기회가 됩니다.

기대 효과

ESG 경영에 대한 방향성을 못 잡는 기업들에게 몇 가지 재무 데이터를 입력받고 가벼운 버전의 모델을 돌려서 등급을 예측해주는 서비스를 제공한다면, 기업입장에서는 공식적인 평가 전 사전 점검을 할 수 있는 기회가 생기는 것이기 때문에 해당 서비스에 대한 수요가 높을 것이라 예상합니다. 나아가 많은 기업들이 투자와 직결된 등급평가를 잘 받기 위해서라도 ESG 경영에 힘쓸 것이므로 이는 사회 전체적인 발전에도 이바지하는 결과를 불러올 것이라 기대합니다.