

중소기업 휴폐업 예측 모델

NICE  dun & bradstreet

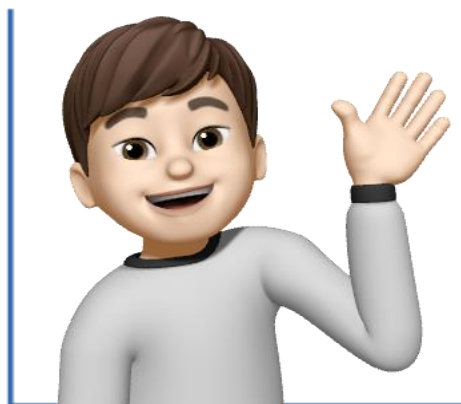
김홍범 이태영 하연진

NICE DNB 비재무 데이터를 활용한 중소기업 휴폐업 예측 챌린지
(2022.09.30 – 2022.11.24)

팀원 소개

✨ 모델링

김홍범(팀장)



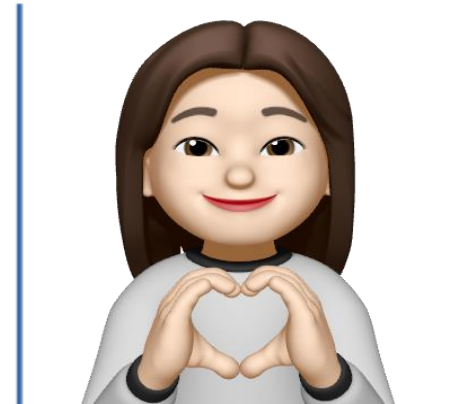
분류 모델링 개선

불균형 데이터 증강

XAI를 통한 변수 중요도 확인

✨ 데이터

하연진



인적자원지표 크롤링

거시경제지표 크롤링

외부파생변수 선행연구조사

이태영



베이스라인 전처리

재무제표 전처리

파생변수 생성

index

01 프로젝트 개요

- 01 프로젝트 목적
- 02 배경 및 필요성
- 03 대안 신용 평가 모델로서의 차별성 및 독창성

02 데이터 수집 및 전처리

- 01 미시경제 지표
- 02 거시경제 지표
- 03 인적자원 지표
- 04 Baseline 전처리 지표

03 분석 모형

- 01 데이터 불균형 및 정규화
- 02 모델 선정
- 03 모델 결과 해석 – feature importance
- 04 모델 결과 해석 – SHAP Value

04 참고 문헌

01

프로젝트 개요

01 프로젝트 목적

02 배경 및 필요성

03 대안 신용 평가 모델로서의 차별성 및 독창성

프로젝트 목적

프로젝트 개요

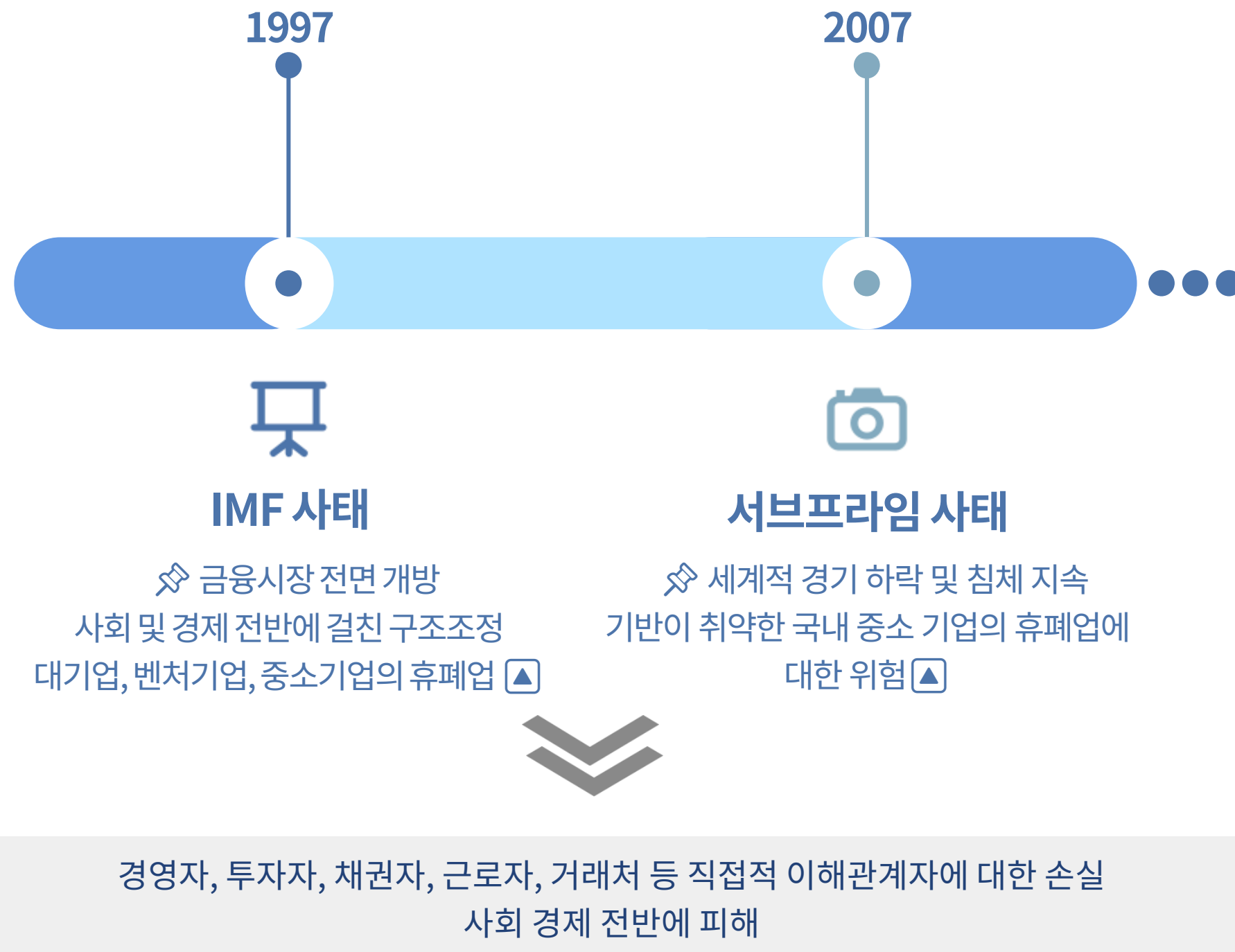
미시경제
거시경제
인적자원

관점의 다양한 파생변수를 기반으로 중소기업의 휴폐업을 예측하고
유의미한 지표를 발굴하며 최적화된 휴폐업 예측 모델을 개발하는
것을 목표로 함

- ❶ 첫째, NICE DNB 데이터, 논리적 근거에 기반한 다양한 파생변수로부터 휴폐업 결정에 유의미한 변수를 조합 & 선별
- ❷ 둘째, 휴폐업에 유의미한 영향을 미치는 요인들에 대한 요인분석을 진행
- ❸ 셋째, 유의미한 변수를 활용하고 성능을 향상시킨 예측 모델 및 인사이트를 도출

배경 및 필요성

프로젝트 개요



“

재무 · 비재무 · 인적자원 지표 활용 중소기업 휴폐업 예측 서비스

재무적 요인은 좋지만 비재무적 요인이 좋지 않아 휴폐업한 기업, 재무적
요인은 좋지 않지만 비재무적 요인이 양호해 지속되는 기업을 볼 수 있음

비재무 정보란?

재무제표에 나타나지 않는 기업 내부의 계량정보와 외부의 시장정보, 산업정보, 창업자에 대한 정보 등 기업과 관련된 모든 정성적인 정보를 포함하는 개념

비재무적 요인의 중요성

- ★ 재무자료 수집이 어려운 중소, 창업기업의 부실예측을 위해서는 비재무정보에 관한 검토가 선행되어야 하나, 비재무정보는 그 범위가 매우 다양하며 이를 수집하기가 어려움
- ★ 중소기업은 대기업에 비해 회계정보의 신뢰성이 낮을 뿐만 아니라 재무구조가 취약하고 경영환경 변화에 민감하기 때문에 비재무정보를 평가에 반영해야 함
- ★ 회계정보의 신뢰성이 낮은 중소기업에 대해서는 재무정보만을 이용한 모형보다는 비재무정보를 함께 분석한 모형이 더 신뢰성이 높은 것으로 나타남

윤대희 "금리 상승으로 폐업 리스크 커 템 확대할 것"

[서경이 만난 사람] 윤대희 신용보증기금 이사장
상환유예 재연장으로 곳곳 부실 징후...시나리오별 대응 강화
40년 쌓인 중기 데이터 활용해 '한국형 페이덱스' 구축 주력
AI기반 기업 분석·비대면 보증 등 신보 디지털화 적극 추진

김지영 기자 2022-02-27 15:35:53 금융정책



윤대희 신용보증기금 이사장이 지난 21일 서울 마포프론트원에서 중소기업 지원 장

"금리가 상승하고 소상공인 폐업이 늘고 있는 상황 등을 감안하면 리스크가
정적인 운용배수로 운용하고 있지만 산업별·지역별 부실 모니터링을 강화
고 합니다."

기업신용평가서 비재무 데이터도 활용...신한은행, 객관성 확보한 시스템 개선

차진형 | 승인 2022.07.11 15:39

다양한 대안정보 활용해 비재무평가 경쟁력 확보



사진=신한은행

신한은행은 11일 빠르게 변화하는 금융환경 속 기업신용평가 부분에서 '비재무 객관화' 모형을 전체 항목에 적용했다고 밝혔다.

비재무 객관화 가이드 제공 시스템은 기업신용평가 시 정성적으로 평가했던 사업, 경영위험 등 비재무적 요인을 최신의 다양한 데이터를 활용해 평가할 수 있도록 가이드 정보를 제공한다.

'데이터' 활용 대안 신용정보 제공

업 관련 대안 신용정보를 제공합니다.

용한 기업 관련 대안 신용정보 제공 서비스를 개시했다

같은 금융결제원이 처리하는 방대한 지급 결제 관련 내

비재무 자료를 기반으로 한 대안 신용평가 모델을 공동

기업 관련 신용평가를 하거나 대출 심사에 활용할 수 있
보를 제공할 예정입니다.

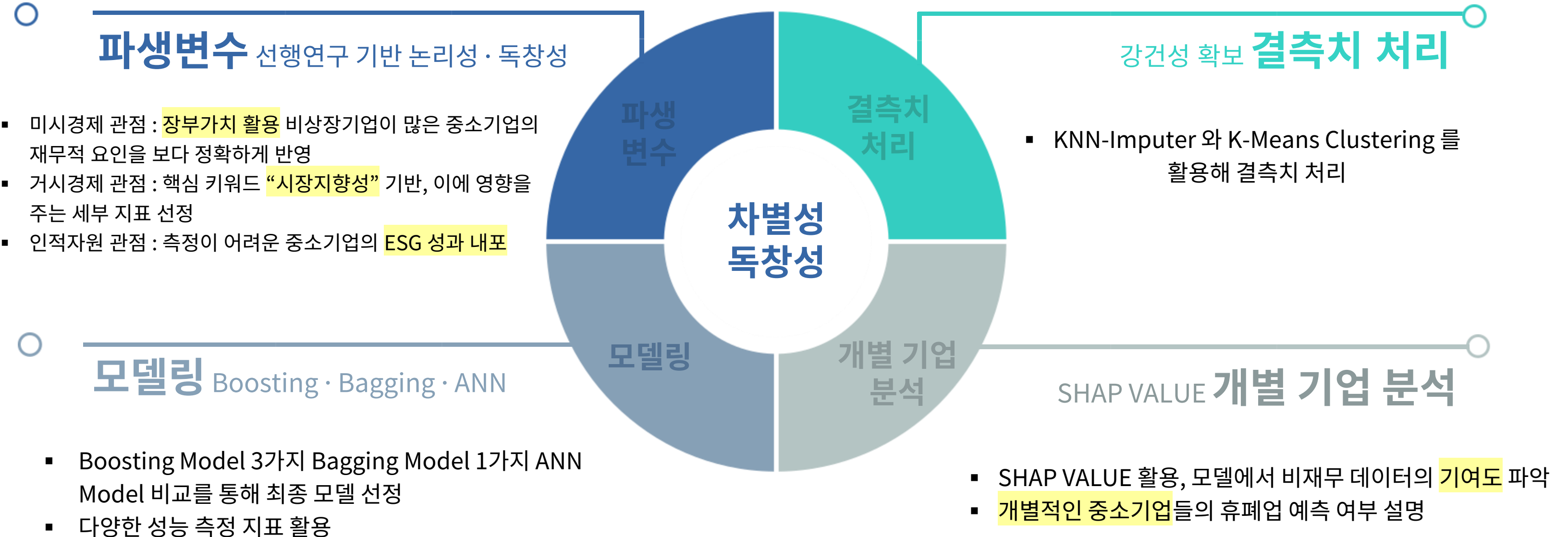
평가데이터를 통해 금융사에 유상으로 제공할 계획입

통해 재무정보가 부족해 전통적인 신용평가가 어려운 기
대한다"고 말했습니다.

대안 신용 평가 모델로서의 차별성 및 독창성

프로젝트 개요

기존의 대안 신용 평가 모델은 주로 대기업이나 중견기업들과 관련된 재무적 요인 중심으로 이루어져, 기업 환경이 열악하고 재무적 요인이 상대적으로 미약한 중소기업에 대해서는 정확성이 확보되지 못함



02

데이터 수집 및 전처리

- 01 미시경제 지표
 - 02 거시경제 지표
 - 03 인적자원 지표
 - 04 Baseline 전처리 지표
-

01 미시경제 지표

(2018.01.01~2022.06.30)

01-1) 다변량 판별 분석모형 – Altman K Score

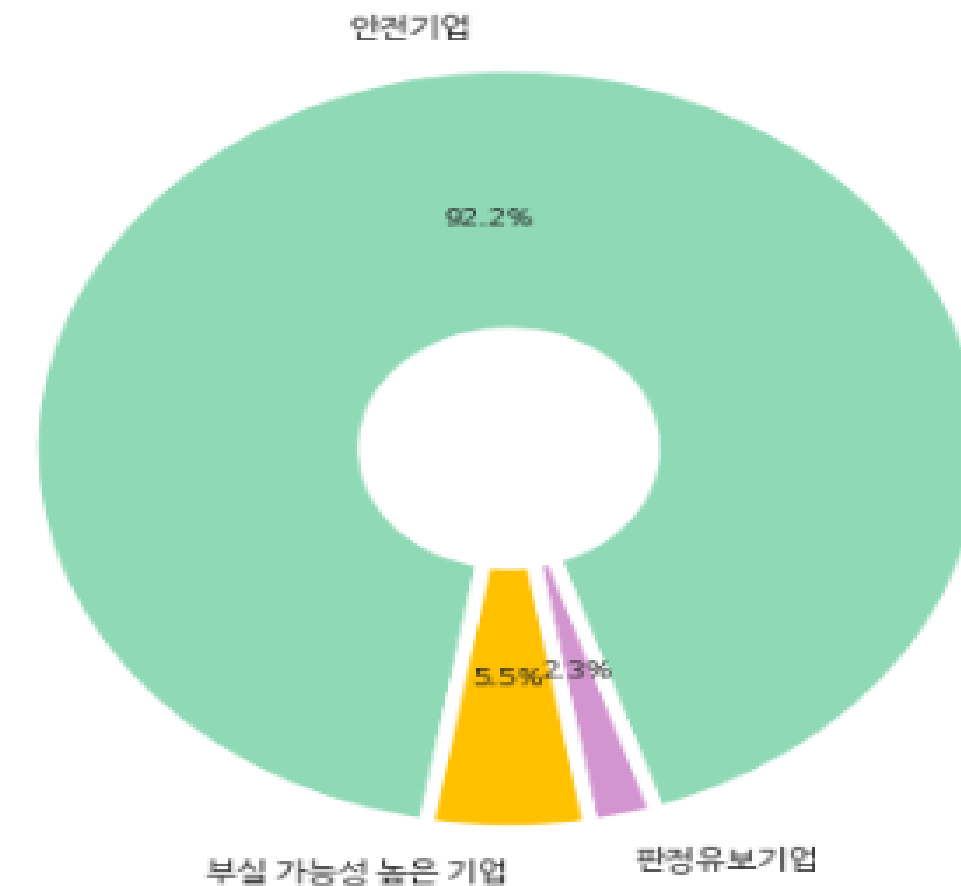
- Z-Score의 창시자 알트만이 1996년에 한국기업의 도산을 예측할 수 있도록 개발한 판별모형으로 Z-Score보다 재무비율이 더 적지만 정확도가 높다는 장점이 있음
- 특히 IMF 당시 놀라운 적중률을 보여줌으로써 상당한 신뢰를 얻었음

$$K\text{-Score} = -17.862 + 1.472X1 + 3.041X2 + 14.839X3 + 1.516X4$$

X1 : Log(자산총계), X2:매출액/자산총계, X3 : 이익잉여금/자산총계, X4: 자본총계/자산총계

$0.75 \leq K$	부실가능성이 없다.	2
$-2.00 \leq K < 0.75$	판정유보	1
$K < -2.00$	부실가능성 심각하다.	0

Aktman K - score에 따른 부실기업 분포

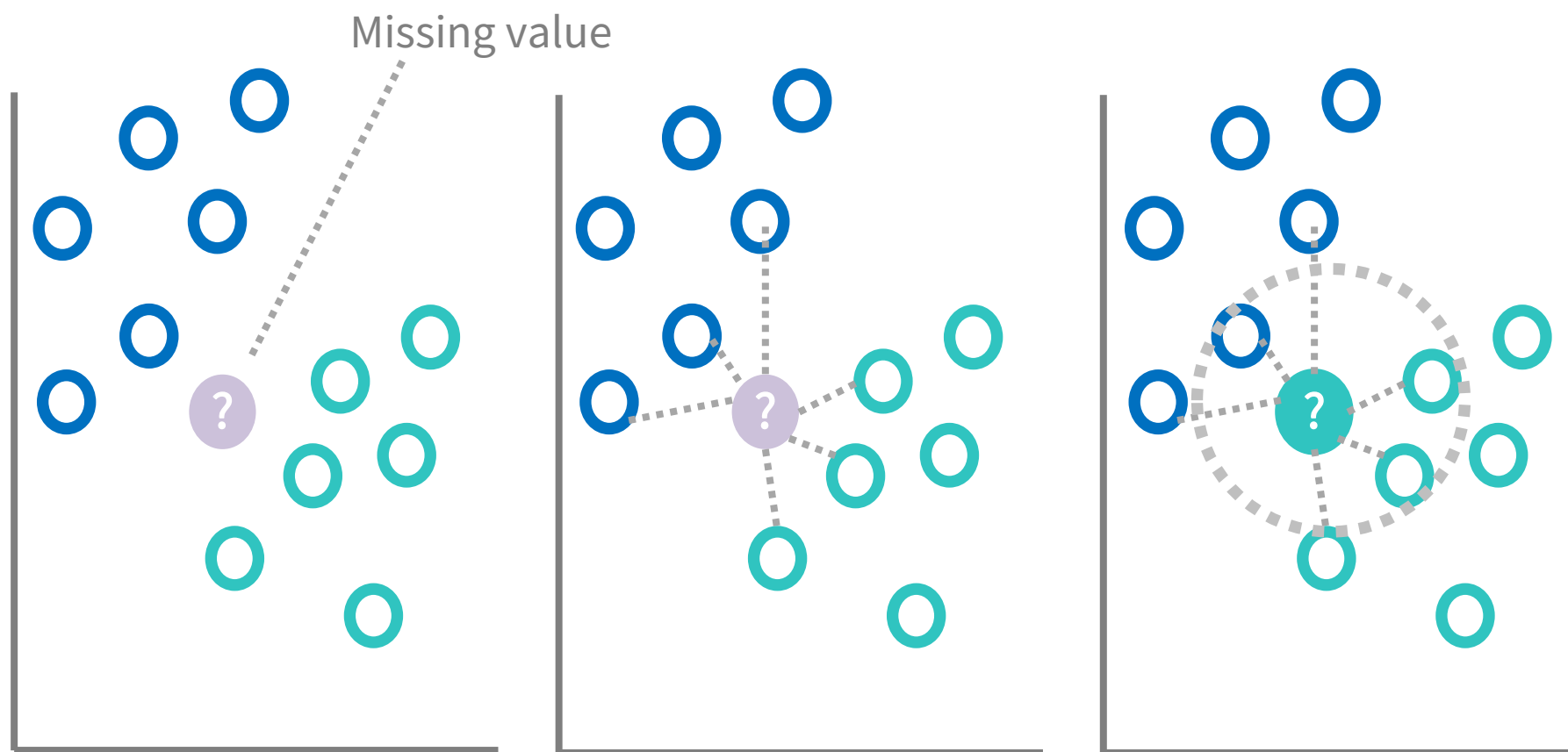


01 미시경제 지표

01-2) 재무 데이터 - 결측치 처리

Step1. KNN Imputer

사업자번호별 재무제표 가중치 평균값



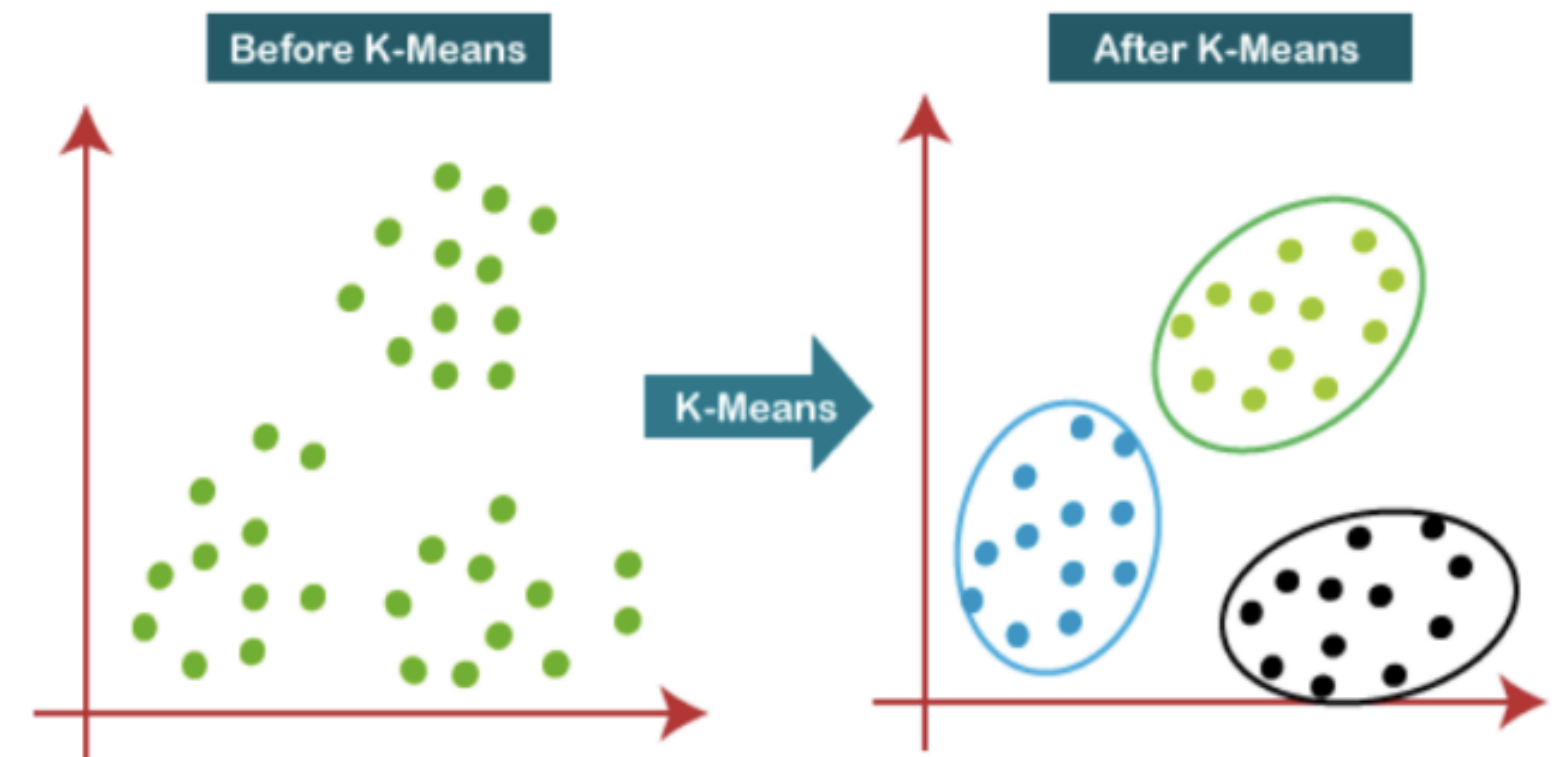
Step2. K-Means 군집별 평균값

2-1. 표준화

- 변수간 유사도 비교를 위해 표준화 진행

2-2. K-Means 군집별 평균값

- 기업 규모를 나타내는 변수를 선정해 클러스터링
- [유동자산, 비유동자산, 매출채권]



01 미시경제 지표

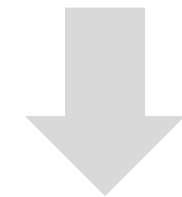
01-2) 재무 데이터 - 하나의 결산년월 추출

- 사업자 번호별로 결산년월이 각자 다르고 제공된 재무제표 개수도 각각 다름
- 통일된 기준이 필요함

사업자번호별 가장 최근 결산년월 재무 데이터만 반환

사업자 번호	결산년월	유동자산	...	매출채권	비유동자산
1018100340	20181231	2667731.0	...	127320.0	32346444.0
1018100340	20191231	2168837.0	...	94085.0	34151633.0
1018100340	20201231	2272260.0	...	78612.0	38329680.0
1018100340	20211231	1175030.0	...	273522.0	37135811.0
...

107214
rows



35086
rows

02 거시경제 지표

활용배경 - 수출시장 지향성



(출처 : 중소기업부.관세청)

시장 지향성이란?

- 시장지향성(Market Orientation)이란 기업의 지속적인 경쟁우위와 높은 수익을 실현하기 위해서 시장요구에 대해 경쟁자들보다 더 잘 대응해야 하며, 시장의 변화를 잘 예측해야 한다는 것을 의미
- 한국 수출업체 중 상당수가 중소기업
- 중소기업의 지속가능을 위한 경쟁역량 창출과정에서 수출시장지향성은 핵심적인 요소로 대두

“수출 시장지향성”은
중소기업에게 주요 경쟁우위 요소

따라서, 해외 시장을 겨냥한 “수출시장지향성”에 영향을 미치는 대표적인 거시경제 지표인
환율, 금리, 발틱 운임 지수, 곡물식량가격, 국제 유가, 원자재 지표를 선정

02 거시경제 지표 환율(Exchange)

(2018.01.01~2022.06.30)

Subject	정의	배경
환율	서로 다른 통화의 교환비율로, 기업의 수출 활동에 가장 직접적인 영향을 미치는 요인	규모가 큰 수출기업에 비해서 작은 수출기업이 환율 변화에 대해서 보다 민감하게 반응
금리	빌려준 돈이나 예금 따위에 붙는 이자. 또는 그 비율	금리가 인상될 경우, 기업의 부채 상환 부담이 가중
발틱 운임 지수(BDI)	석탄, 철광석, 곡물 등을 싣고 세계 26개 주요 항로를 오가는 선박의 운임지수	운송비에 대한 주요 지표이며 전세계 무역과 제조 활동량에 대해 알려주어 세계 경기의 대표 선행지표
곡물식량가격지수	유엔식량농업기구(FAO)가 대표 곡물식량 국제거래 가격을 종합해 매달 발표하는 지수	한국은 곡물 해외 의존도가 높아 국제 곡물 가격변동은 가공식품 및 배합사료 산업의 성과에 영향
국제 유가	WTI(미국의 서부 텍사스) Brent(북해의 브렌트 유전) Dubai(중동 아랍에미리트)	국제유가 변동은 기업의 생산비용 증대, 인플레이션 유발, 가계 구매력 감소 야기. 제조업 영향↑
원자재 가격	공업 생산의 원료가 되는 자재	경쟁력이 낮은 기업들은 생산비용 증가를 제품 가격에 전가하지 못해 채산성이 악화, 가격경쟁력 상실



적용 데이터	데이터 출처
원달러 환율	Finance DataReader

02 거시경제 지표 금리(Interest rate)

(2018.01.01~2022.06.30)

Subject	정의	배경
환율	서로 다른 통화의 교환비율로, 기업의 수출 활동에 가장 직접적인 영향을 미치는 요인	규모가 큰 수출기업에 비해서 작은 수출기업이 환율 변화에 대해서 보다 민감하게 반응
금리	빌려준 돈이나 예금 따위에 붙는 이자. 또는 그 비율	금리가 인상될 경우, 기업의 부채 상환 부담이 가중
발틱 운임 지수(BDI)	석탄, 철광석, 곡물 등을 싣고 세계 26개 주요 항로를 오가는 선박의 운임지수	운송비에 대한 주요 지표이며 전세계 무역과 제조 활동량에 대해 알려주어 세계 경기의 대표 선행지표
곡물식량가격지수	유엔식량농업기구(FAO)가 대표 곡물식량 국제거래 가격을 종합해 매달 발표하는 지수	한국은 곡물 해외 의존도가 높아 국제 곡물 가격변동은 가공식품 및 배합사료 산업의 성과에 영향
국제 유가	WTI(미국의 서부 텍사스) Brent(북해의 브렌트 유전) Dubai(중동 아랍에미리트)	국제유가 변동은 기업의 생산비용 증대, 인플레이션 유발, 가계 구매력 감소 야기. 제조업 영향 ↑
원자재 가격	공업 생산의 원료가 되는 자재	경쟁력이 낮은 기업들은 생산비용 증가를 제품 가격에 전가하지 못해 채산성이 악화, 가격경쟁력 상실



적용 데이터

CD91일, 콜금리,
국고채3년, 회사채3년

데이터 출처

Naver Finance

02 거시경제 지표 발틱운임지수(BDI)

(2018.01.01~2022.06.30)

Subject	정의	배경
환율	서로 다른 통화의 교환비율로, 기업의 수출 활동에 가장 직접적인 영향을 미치는 요인	규모가 큰 수출기업에 비해서 작은 수출기업이 환율 변화에 대해서 보다 민감하게 반응
금리	빌려준 돈이나 예금 따위에 붙는 이자. 또는 그 비율	금리가 인상될 경우, 기업의 부채 상환 부담이 가중
발틱 운임 지수(BDI)	석탄, 철광석, 곡물 등을 싣고 세계 26개 주요 항로를 오가는 선박의 운임지수	운송비에 대한 주요 지표이며 전세계 무역과 제조 활동량에 대해 알려주어 세계 경기의 대표 선행지표
곡물식량가격지수	유엔식량농업기구(FAO)가 대표 곡물식량 국제거래 가격을 종합해 매달 발표하는 지수	한국은 곡물 해외 의존도가 높아 국제 곡물 가격변동은 가공식품 및 배합사료 산업의 성과에 영향
국제 유가	WTI(미국의 서부 텍사스) Brent(북해의 브렌트 유전) Dubai(중동 아랍에미리트)	국제유가 변동은 기업의 생산비용 증대, 인플레이션 유발, 가계 구매력 감소 야기. 제조업 영향 ↑
원자재 가격	공업 생산의 원료가 되는 자재	경쟁력이 낮은 기업들은 생산비용 증가를 제품 가격에 전가하지 못해 채산성이 악화, 가격경쟁력 상실



적용 데이터

발틱운임지수(BDI)

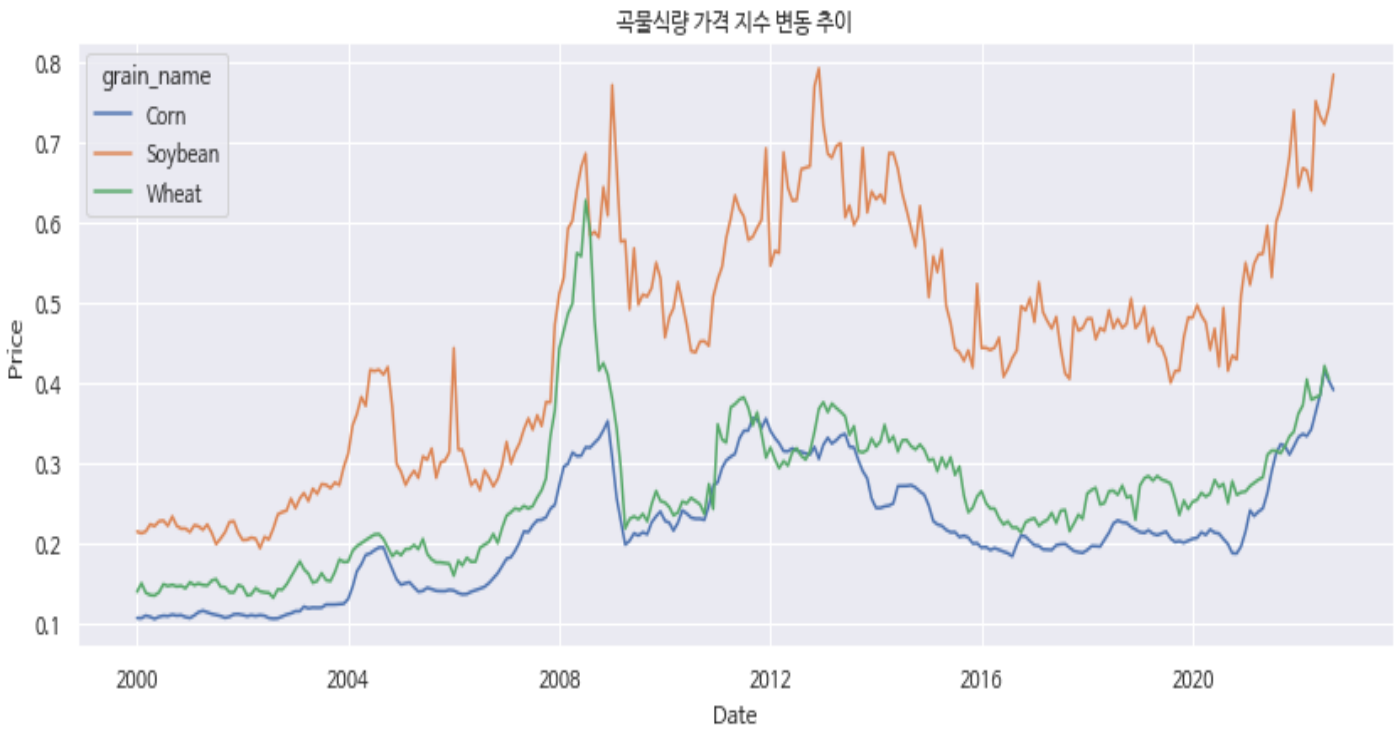
데이터 출처

Investing.com

02 거시경제 지표 곡물가격지수 & 식량가격지수

(2018.01.01~2022.06.30)

Subject	정의	배경
환율	서로 다른 통화의 교환비율로, 기업의 수출 활동에 가장 직접적인 영향을 미치는 요인	규모가 큰 수출기업에 비해서 작은 수출기업이 환율 변화에 대해서 보다 민감하게 반응
금리	빌려준 돈이나 예금 따위에 붙는 이자. 또는 그 비율	금리가 인상될 경우, 기업의 부채 상환 부담이 가중
발틱 운임 지수(BDI)	석탄, 철광석, 곡물 등을 싣고 세계 26개 주요 항로를 오가는 선박의 운임지수	운송비에 대한 주요 지표이며 전세계 무역과 제조 활동량에 대해 알려주어 세계 경기의 대표 선행지표
곡물식량가격지수	유엔식량농업기구(FAO)가 대표 곡물식량 국제거래 가격을 종합해 매달 발표하는 지수	한국은 곡물 해외 의존도가 높아 국제 곡물 가격변동은 가공식품 및 배합사료 산업의 성과에 영향
국제 유가	WTI(미국의 서부 텍사스) Brent(북해의 브렌트 유전) Dubai(중동 아랍에미리트)	국제유가 변동은 기업의 생산비용 증대, 인플레이션 유발, 가계 구매력 감소 야기. 제조업 영향↑
원자재 가격	공업 생산의 원료가 되는 자재	경쟁력이 낮은 기업들은 생산비용 증가를 제품 가격에 전가하지 못해 채산성이 악화, 가격경쟁력 상실

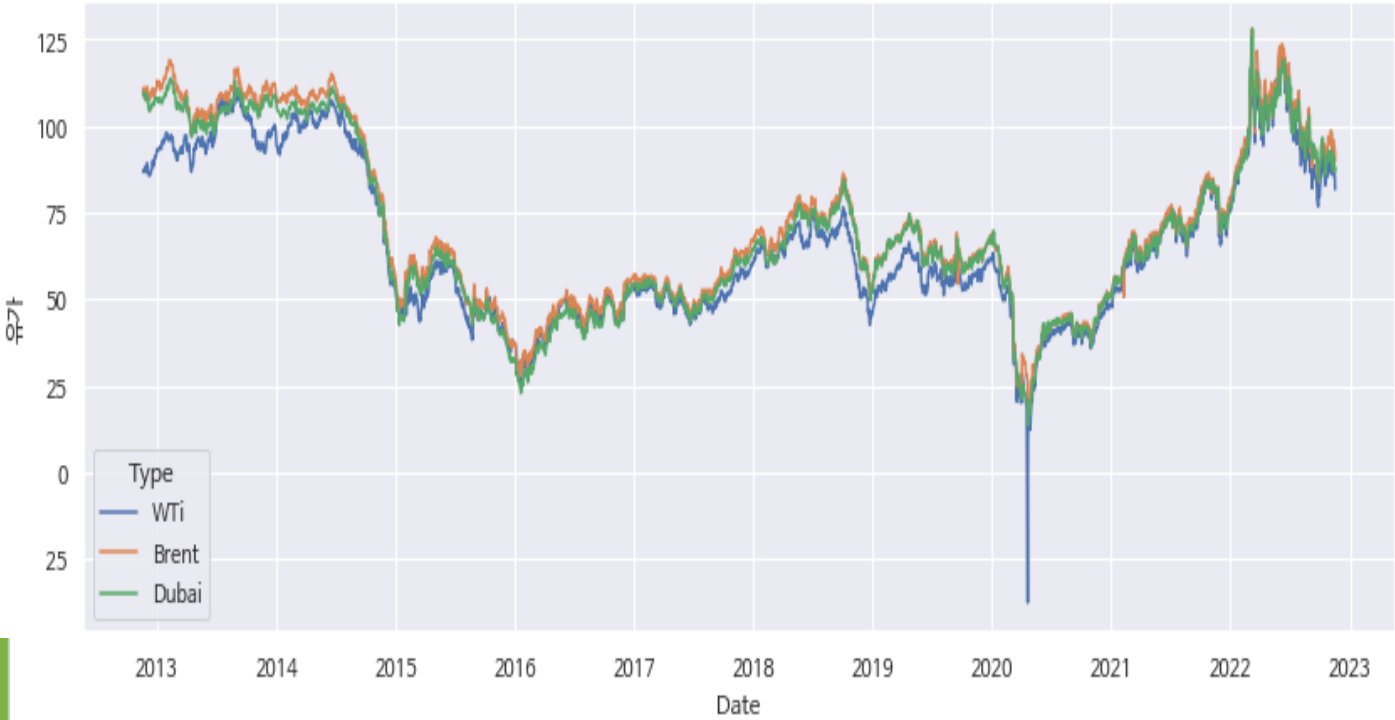


적용 데이터	데이터 출처
옥수수, 밀, 콩, 식량가격지수	국제연합식량기구

02 거시경제 지표 국제 유가

(2018.01.01~2022.06.30)

Subject	정의	배경
환율	서로 다른 통화의 교환비율로, 기업의 수출 활동에 가장 직접적인 영향을 미치는 요인	규모가 큰 수출기업에 비해서 작은 수출기업이 환율 변화에 대해서 보다 민감하게 반응
금리	빌려준 돈이나 예금 따위에 붙는 이자. 또는 그 비율	금리가 인상될 경우, 기업의 부채 상환 부담이 가중
발틱 운임 지수(BDI)	석탄, 철광석, 곡물 등을 싣고 세계 26개 주요 항로를 오가는 선박의 운임지수	운송비에 대한 주요 지표이며 전세계 무역과 제조 활동량에 대해 알려주어 세계 경기의 대표 선행지표
곡물식량가격지수	유엔식량농업기구(FAO)가 대표 곡물식량 국제거래 가격을 종합해 매달 발표하는 지수	한국은 곡물 해외 의존도가 높아 국제 곡물 가격변동은 가공식품 및 배합사료 산업의 성과에 영향
국제 유가	WTI(미국의 서부 텍사스) Brent(북해의 브렌트 유전) Dubai(중동 아랍에미리트)	국제유가 변동은 기업의 생산비용 증대, 인플레이션 유발, 가계 구매력 감소 야기. 제조업 영향 ↑
원자재 가격	공업 생산의 원료가 되는 자재	경쟁력이 낮은 기업들은 생산비용 증가를 제품 가격에 전가하지 못해 채산성이 악화, 가격경쟁력 상실

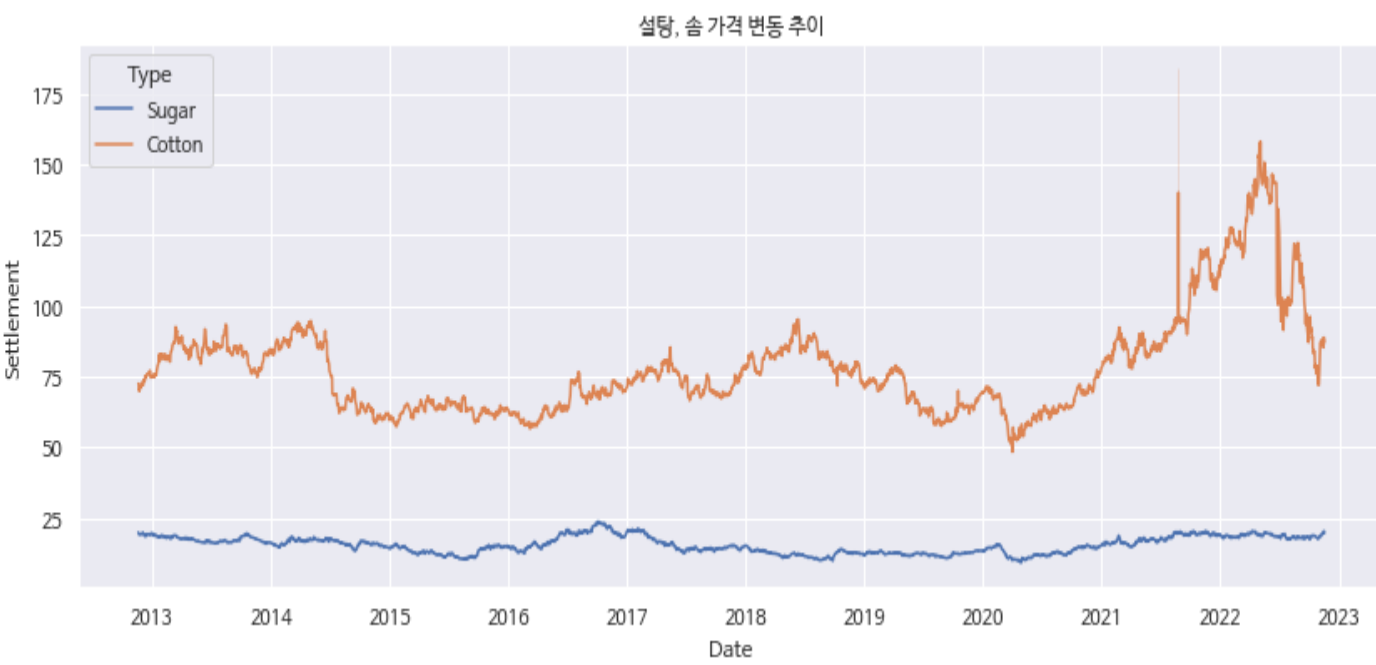


적용 데이터	데이터 출처
WTI, Brent, Dubai	한국석유공사

02 거시경제 지표 원자재

(2018.01.01~2022.06.30)

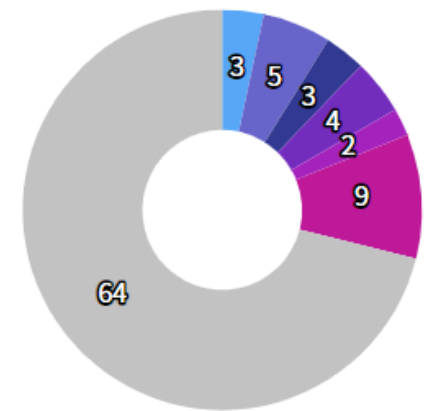
Subject	정의	배경
환율	서로 다른 통화의 교환비율로, 기업의 수출 활동에 가장 직접적인 영향을 미치는 요인	규모가 큰 수출기업에 비해서 작은 수출기업이 환율 변화에 대해서 보다 민감하게 반응
금리	빌려준 돈이나 예금 따위에 붙는 이자. 또는 그 비율	금리가 인상될 경우, 기업의 부채 상환 부담이 가중
발틱 운임 지수(BDI)	석탄, 철광석, 곡물 등을 싣고 세계 26개 주요 항로를 오가는 선박의 운임지수	운송비에 대한 주요 지표이며 전세계 무역과 제조 활동량에 대해 알려주어 세계 경기의 대표 선행지표
곡물식량가격지수	유엔식량농업기구(FAO)가 대표 곡물식량 국제거래 가격을 종합해 매달 발표하는 지수	한국은 곡물 해외 의존도가 높아 국제 곡물 가격변동은 가공식품 및 배합사료 산업의 성과에 영향
국제 유가	WTI(미국의 서부 텍사스) Brent(북해의 브렌트 유전) Dubai(중동 아랍에미리트)	국제유가 변동은 기업의 생산비용 증대, 인플레이션 유발, 가계 구매력 감소 야기. 제조업 영향↑
원자재 가격	공업 생산의 원료가 되는 자재	경쟁력이 낮은 기업들은 생산비용 증가를 제품 가격에 전가하지 못해 채산성이 악화, 가격경쟁력 상실



적용 데이터	데이터 출처
금, 은, 구리, 설탕, 솜	퀀들(Quandle), KREI 해외곡물시장정보

03 인적자원 지표

다우존스지속가능지수 (DJSI, Dow Jones Sustainability Indices)

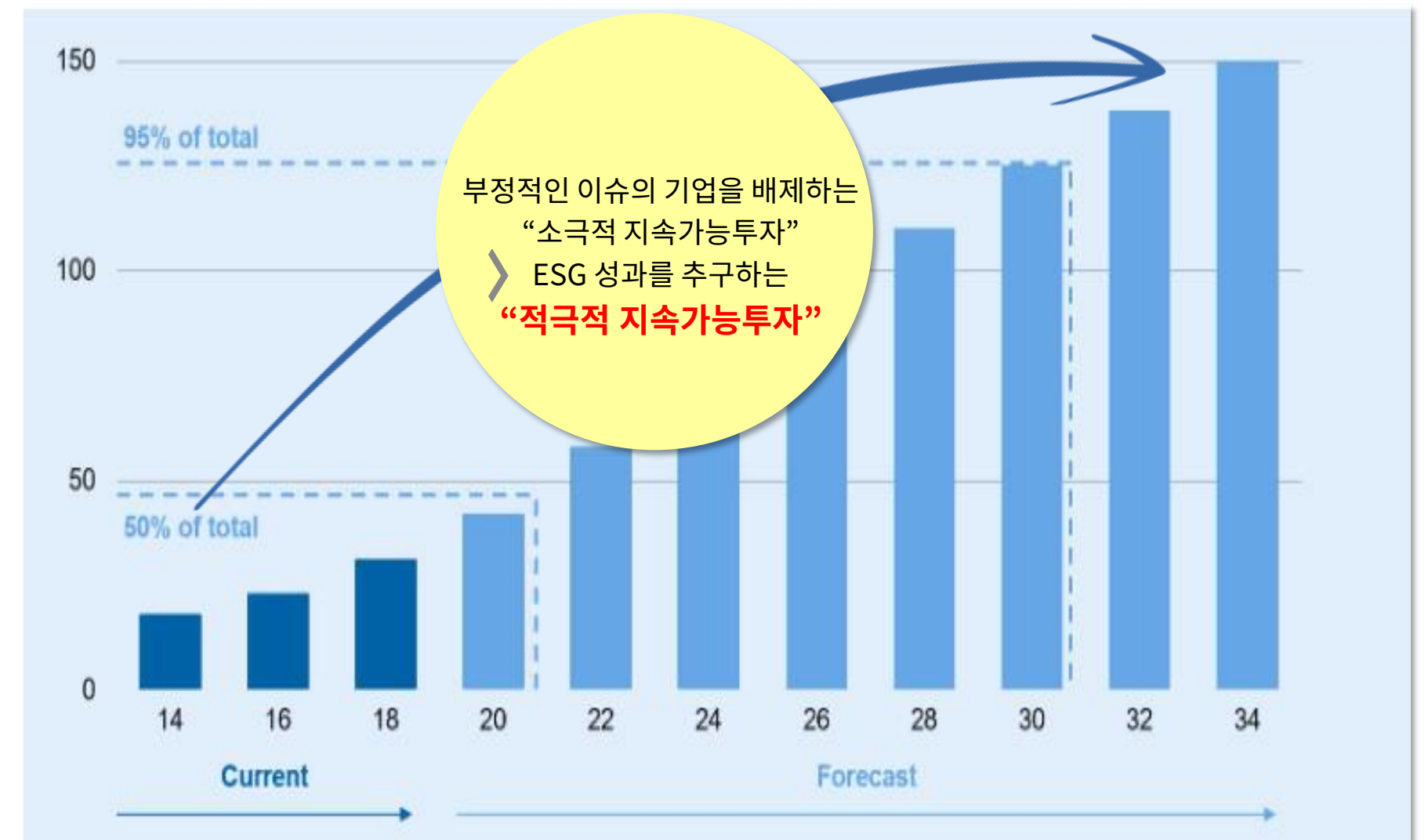


대표적인 ESG 평가지표인
다우존스지속가능지수에 따르면, HR 관련
정책을 잘 펼치느냐가 기업의 투자매력도를
결정하는 데 1/3 정도의 높은 영향력을 가지고
있음

- 중소기업은 중견기업이나 대기업에 비해 ESG에 대한 인식이 부족하고 대비가 되어있지 않기 때문에, ESG 확산에 따른 손실리스크가 큼
- 지속 가능성 프로그램(ESG)이 강력한 회사의 사기가 열악한 회사에 비해 55% / 직원 충성도 38% / 생산성 16%가 더 높음, 연간 퇴직률은 3% 이상 하락하여 이직으로 인한 교체 비용도 절감할 수 있음. *By. Michael E. Porter*

전 세계 ESG 투자 총 운용자산(AUM) 규모 및 추이 예측

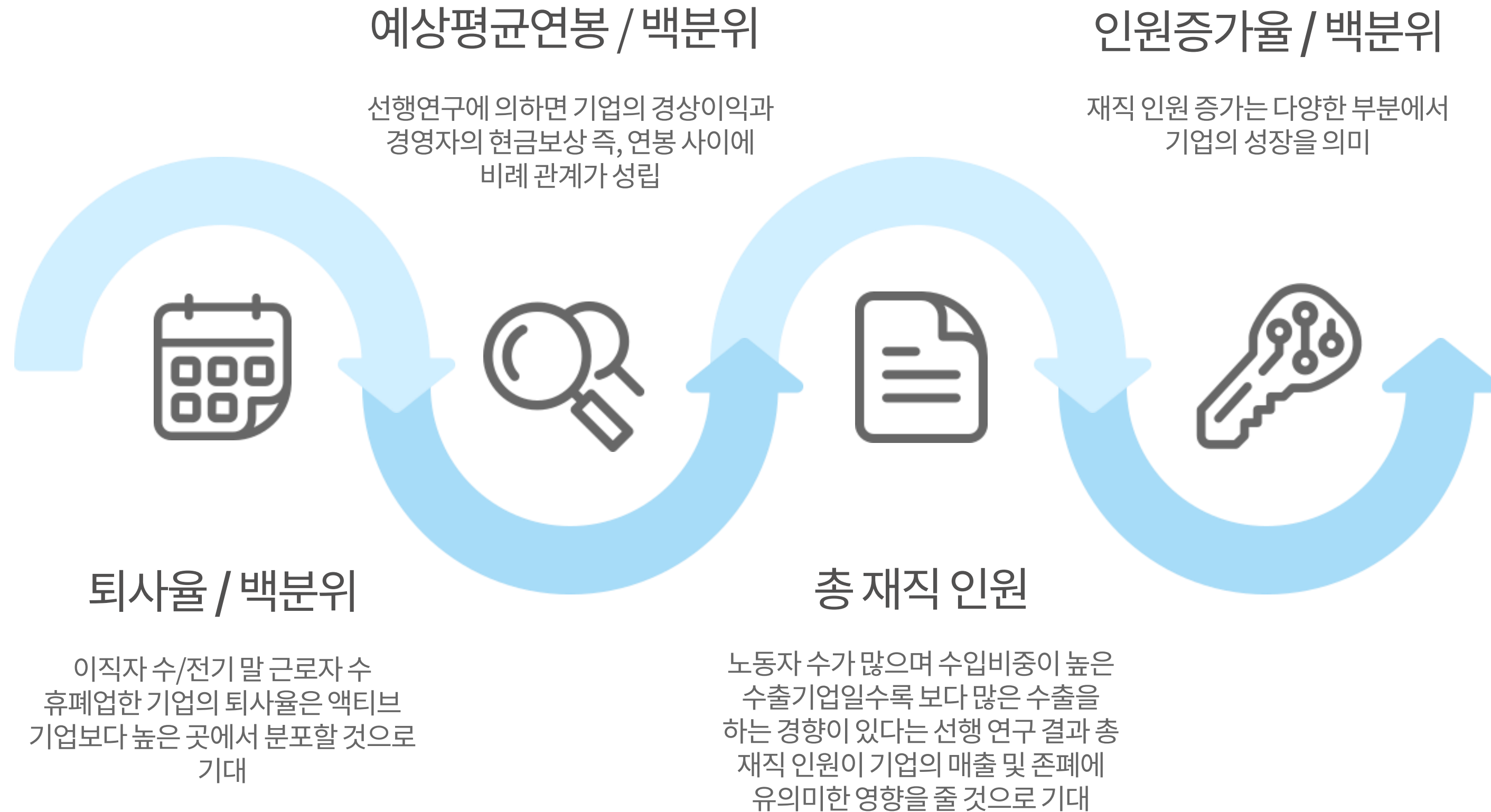
(단위 : 1조 달러)



자료 : Deutsche Bank, Global Sustainable Alliance, zeb research.

인적 자원 지표를 통해 측정이 상대적으로 어려운 중소기업의 ESG 실천 정도 및 기업 경쟁력, 지속가능성을 역으로 유추할 수 있을 것이다.

03 인적자원 지표



03 인적자원 지표

예상평균연봉 / 백분위

선행연구에 의하면 기업의
경상이익과 경영자의 현금보상 즉,
연봉

인원증가율 / 백분위

재직 인원 증가는 다양한 부분에서
기업의 성장을 의미한다.

“크레딧잡” 크롤링을 통한 데이터 수집

국민연금에서 제공받은 데이터를 기반으로 전국 42만개의 기업에 대한
업력/사원수 /올해 입사율/올해 퇴사율/연봉상위율/월별 평균 급여에 대한 정보를
제공하고 있으며, 타 플랫폼보다 정량적 지표의 정확도가 높은 것으로 알려져 있음

퇴사율 / 백분위

이직자 수/전기 말 근로자 수
휴폐업한 기업의 퇴사율은 액티브
기업보다 높은 곳에서 분포할
것으로 기대한다.

총 재직 인원

노동자 수가 많으며 수입비중이
높은 수출기업일수록 보다 많은
수출을 하는 경향이 있다는 선행
연구 결과 총 재직 인원이 기업의
매출 및 존폐에 유의미한 영향을 줄
것으로 기대한다.

04 Baseline 데이터

전처리 & 새로운 파생변수 생성

- 액티브 & 휴폐업 중소기업 데이터 가공을 통한 Baseline 구축

BIZ_NO	(주)여부	HDOF_C MP_CD	LIST_CD	HOMEPA GE_URL	휴폐업_여부
3148517348	1	1	0	0	0
1308505171	1	1	0	1	0
...
1138502432	0	1	0	0	0
2218512338	1	1	0	1	0

휴폐업 여부

(주) 여부

본점기업코드

상장코드

홈페이지 URL

휴폐업 여부 타겟 컬럼 생성

1) 이진 분류 타겟컬럼 생성

-> 액티브 기업:0

-> 휴폐업 기업:1

2) 액티브 기업의 경우 외부감사대상 기업만 선정

04 Baseline 데이터

전처리 & 새로운 파생변수 생성

- 액티브 & 휴폐업 중소기업 데이터 가공을 통한 Baseline 구축

BIZ_NO	(주)여부	HDOF_C MP_CD	LIST_CD	HOMEPA GE_URL	휴폐업_여부
3148517348	1	1	0	0	0
1308505171	1	1	0	1	0
...
1138502432	0	1	0	0	0
2218512338	1	1	0	1	0

휴폐업 여부

(주) 여부

본점기업코드

상장코드

홈페이지 URL

(주) 여부 & 상장코드 & 홈페이지 URL

- 1) 기업명 앞뒤에 (주) 여부에 따른 파생 변수 생성
'(주)'가 있으면 1, 없으면 0
- 2) 상장코드 결측치 여부에 따라 파생 변수 생성
상장코드 있으면 1, 없으면 0
- 3) 홈페이지 URL 결측치 여부에 따라 파생 변수 생성
홈페이지 URL 있으면 1, 없으면 0

04 Baseline 데이터

전처리 & 새로운 파생변수 생성

- 액티브 & 휴폐업 중소기업 데이터 가공을 통한 Baseline 구축

BIZ_NO	(주)여부	HDOF_C MP_CD	LIST_CD	HOMEPA GE_URL	휴폐업_ 여부
31485173 48	1	1	0	0	0
13085051 71	1	1	0	1	0
...
11385024 32	0	1	0	0	0
22185123 38	1	1	0	1	0

휴폐업 여부

(주) 여부

본점기업코드

상장코드

홈페이지 URL

본점기업코드

1) value_counts를 통해 해당하는 개수를 대입
-> 숫자가 높을수록 체인점을 많이 보유한
업종임을 판단

2) 결측치는 etc로 대체한 후 value_counts 진행

최종 Feature

(2018.01.01~ 2021.12.31)

미시경제 지표
재무데이터, Altman K score

거시경제 지표
환율, 금리, BDI, 곡물식량가격지수, 국제유가, 원자재 국내 선물가격

인적자원 지표
퇴사율, 예상평균연봉, 총 재직 인원, 최근 1년 인원증가율

BIZ_NO	CMP_NM	BZ_TYP	CMP_SCL	HDOF_BR_GB	...	turnover_r ate	turnover_r ate_rankin g	employee _growth_r ate	employee _growth_r ate_rankin g
113811173 9	한일가스산 업	M	2	1	...	26.000000	17.000000	5.000000	100.00000 0
207813219 3	엠펙아이	M	2	1	...	59.120483	55.080538	47.626773	59.519504
131816756 5	도영운수	M	2	1	...	59.190483	55.080538	47.626773	59.519504
504812825 1	중원산업	M	2	1	...	59.190483	55.080538	47.626773	59.519504
...

34883 중소기업 X 92 Col = 3,209,236

03

분석 모형

01 데이터 불균형 및 정규화

02 모델 선정

03 모델 결과 해석 – feature importance

04 모델 결과 해석 – SHAP Value

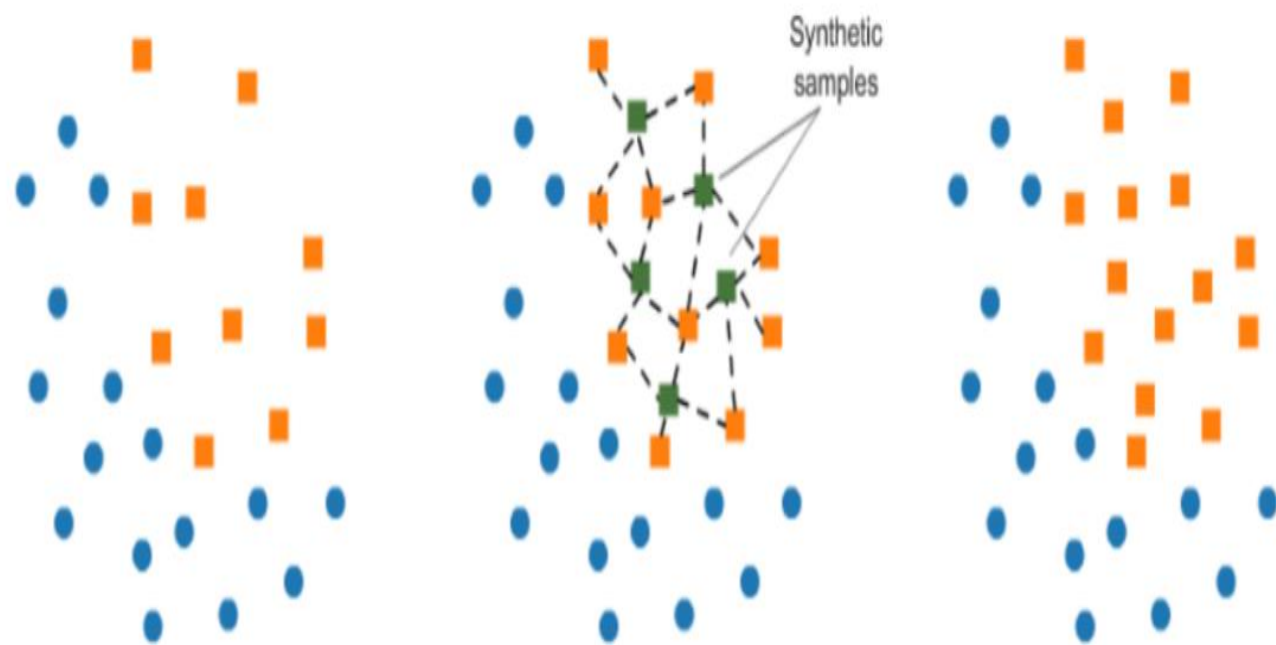
분석 모형 및 방법론

데이터 불균형 및 정규화

Step1. 데이터 불균형

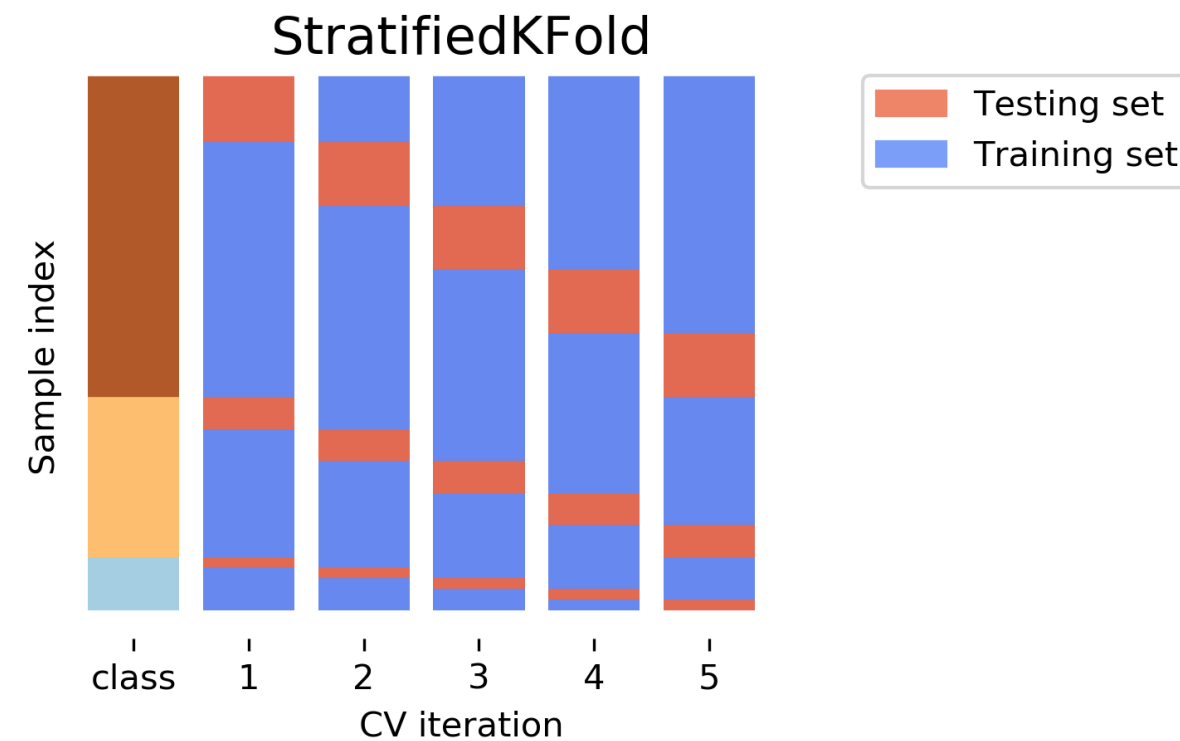
- 오버 샘플링(SMOTE)을 적용하여 데이터의 불균형(5%)을 맞추어 모델 예측 성능을 높임
- Stratified - kfold Cross Validation 을 진행함으로써 특정 데이터가 한쪽으로 치우쳐 학습 및 검증되는 것을 방지함

SMOTE(Synthetic Minority Over-sampling Techniques)



Step2. Scaling

- MinMaxScaler (정규화)를 진행하여 0~1 사이 값으로 scaling을 진행하였고 해당 방법론을 통해 모델 성능 향상을 기대하였음



분석 모형 및 방법론

모델 선정

- 한가지 평가지표보다는 다양한 평가지표를 활용하여 모델 결과의 신뢰성을 높임
- 1가지 bagging 모델 및 3가지 boosting 이진 분류 앙상블 모델을 활용함
- 또한 대표적인 ANN 모델을 학습하여 앙상블 모델과의 성능 비교를 진행하였음
- 평가지표 선정 : Accuracy, Recall, Precision, AUPRC, F1_Score, AUROC

	RandomForest	XGBoost	LightGBM	CatBoost	ANN(Artificial Neural Network)
Accuracy	0.966	0.966	0.965	0.964	0.940
Recall	0.974	0.974	0.978	0.974	0.961
Precision	0.961	0.961	0.956	0.958	0.929
AUPRC	0.948	0.948	0.945	0.946	0.911
F1_score	0.966	0.966	0.965	0.964	0.940
AUROC	0.998	0.998	0.965	0.998	0.992



성능 및 범용성이 높은 최종모델 선정

Model	Accuracy	Recall	Precision	AUPRC	F1_score	AUROC
XGBoost	0.966	0.974	0.961	0.948	0.966	0.998

분석 모형 및 방법론

모델 결과 해석 – Feature importance

변수 중요도(feature importance)

- 변수 중요도란 학습된 모형에 대하여 반응 변수와의 관련성 또는 예측 관점에서 각 변수들의 영향력을 수치화 함
- 특정 변수가 모형의 예측에 어떻게 영향을 끼치는지 만을 보여주는 지표

단점

- 해석이 불가능하고 변수 중요도 측정 기준별로 중요 변수가 달라짐
- 변수의 중요도만 알 수 있고, 얼마나 어떻게 영향을 미치는지 (양수/음수, 크기) 알 수 없음
- “feature importance가 높다고 중요한 지표다” 라고 말할 수 없음 (inconsistency – 모형마다 다르게 산출됨)

변수 중요도 해석

- 오른쪽 그림 결과 비재무 지표의 중요도를 파악할 수 없음
- 직원수를 제외하고 모두 재무 데이터가 높은 중요도를 가지는 것으로 나타남

Top 10 feature importance

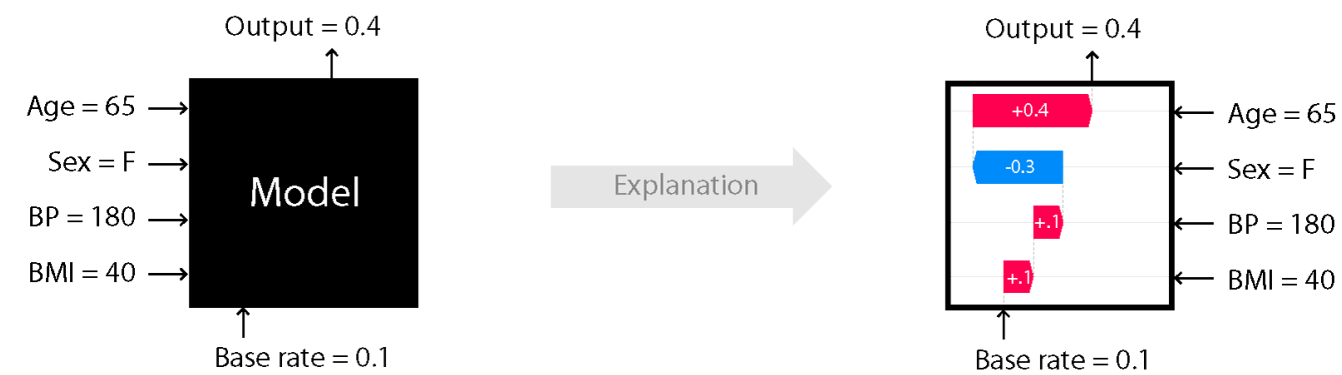
변수명	Feature importance
직원수	170
유형자산	138
자산총계	132.0
무형자산	131.0
자본금	122
매출액	115
매출채권	108
비유동자산	105
비유동부채	101

분석 모형 및 방법론

모델 결과 해석 - SHAP VALUE

SHAP VALUE란?

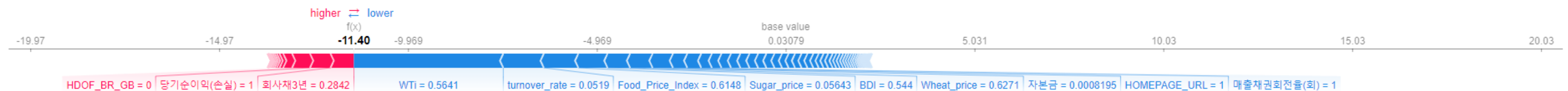
- 목적 : 예측에 대한 각 형상의 기여도를 계산하여 인스턴스(데이터) X의 예측을 설명하는 것



SHAP Value의 특징

- 개별 예측값에 대한 각 변수의 중요도를 파악 가능
- 예측 모형 클래스의 관계없이 적용할 수 있어 유연함
- 개별 예측값(Local)에 대한 변수 중요도 뿐만 아니라 예측에 대한 전반적인 변수 중요도 역시 계산 가능

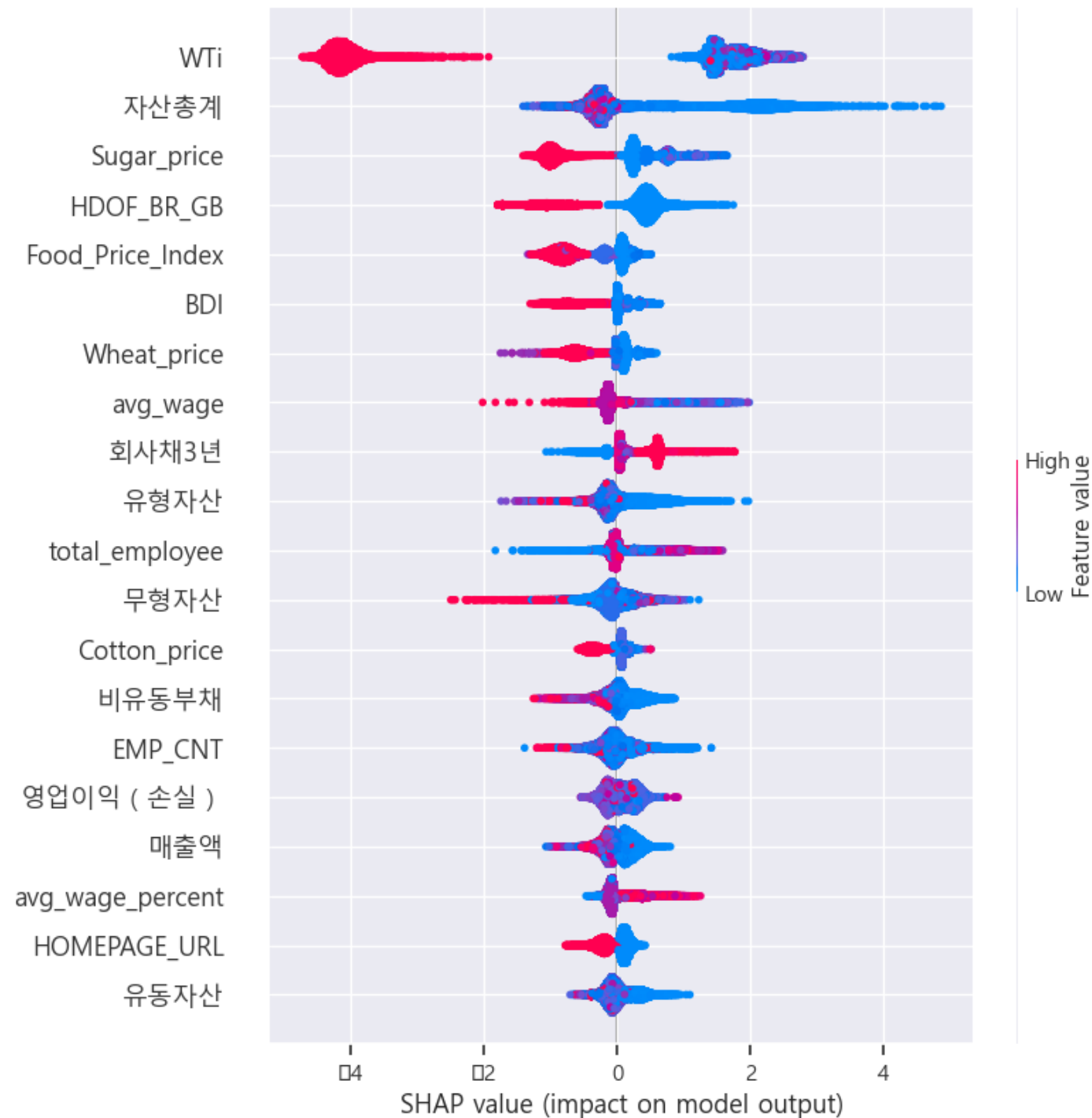
- 특정 중소기업의 휴, 폐업 예측 기여도 판단 가능(force plot)



분석 모형 및 방법론

모델 결과 해석 - SHAP VALUE

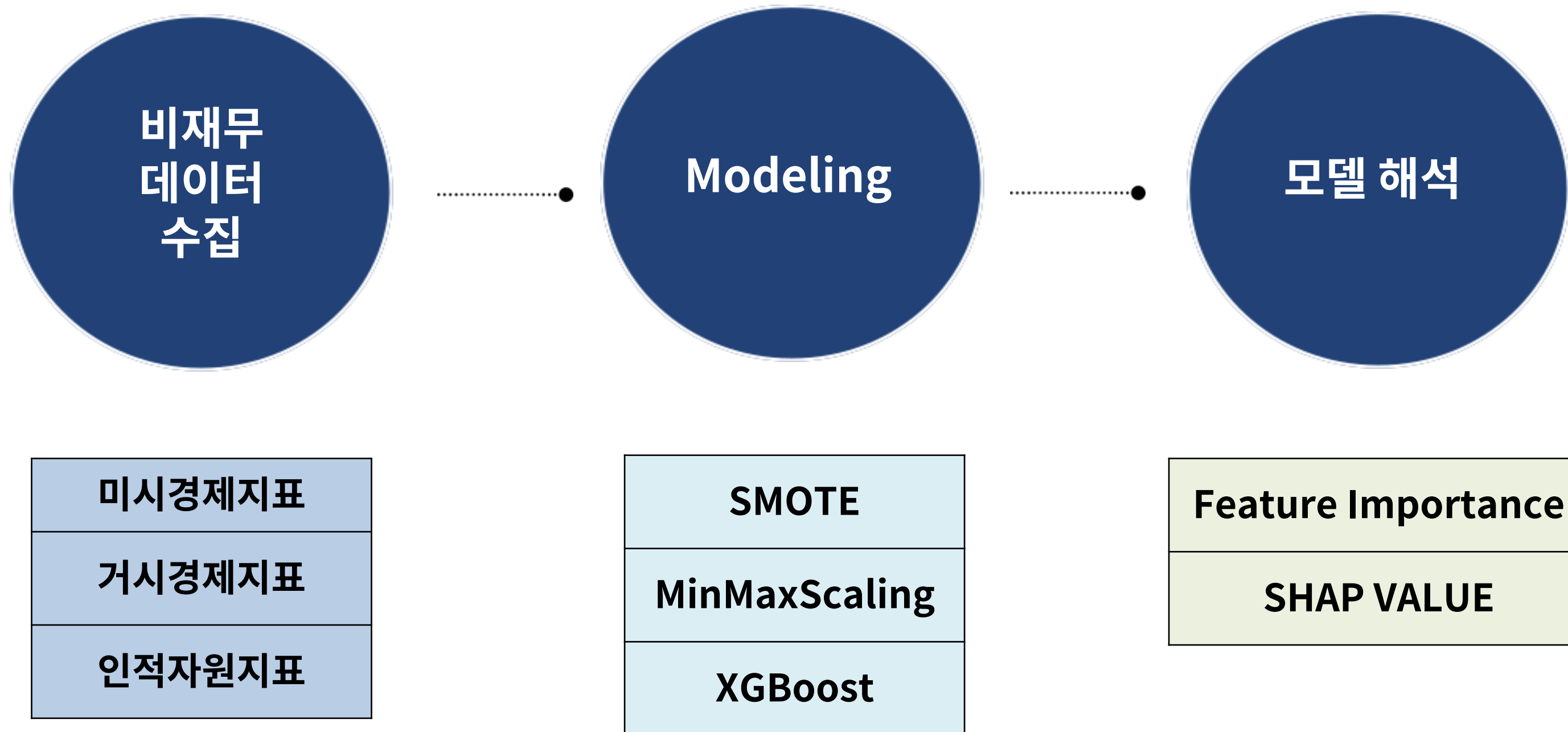
SHAP summary_plot



SHAP VALUE 결과 해석

- 가장 큰 SHAP VALUE를 가지는 지표는 WTi(서부텍사스산원유)
- 특성값이 작을수록(파란색) 예측에 긍정적인 영향을, 클수록(빨간색) 부정적인 영향을 미치고 있음을 의미함
- 즉 유가가 낮을수록 휴폐업을 할 확률이 높았음을 의미하고, 반대로 유가가 높을수록 휴,폐업을 할 확률이 상대적으로 낮음을 의미함
- 해당 결과 단순 재무지표인 자산총계 뿐만 아니라 다양한 비재무데이터가 모델 예측에서 중요하게 사용되었음을 확인할 수 있음

결론



대안신용평가 모델에서의 “비재무 데이터” 중요성 파악

04

참고문헌

참고문헌

[1] 남기정, 이동명 and 진로. (2019). 비재무정보를 이용한 창업기업의 부실요인에 관한 실증연구. 벤처창업연구, 14(1), 139-149.*

[2] 조옥래(1998), 『중소기업 도산예측과정에서 비재무제표 항목의 유용성검증과 기업 분석시스템 개발에 관한 연구』,경희대학교,석사학위논문.

[3] 또 다른 발자국, 2017년 02월 08일, https://nicksstory.tistory.com/182

[4] 정대용,유봉호,“기업가 지향성과 시장지향성이 조직유효성과 사업성과에 미치는 영향”,벤처경영연구,제10권,제2호,pp.69-93,2007.

[5] A.KohliandB.Jaworski,“MarketOrientation:The Construct, Research Proposition, and ManagerialImplication,” JournalofMarketing, Vol.54,No.2,pp.1-18,1990

[6]J.C.NarverandS.F.Slater,“TheEffectofaMarketOrientationonBusinessProfitability,” JournalofMarketing,Vol.54,No.4,pp.20-35,1990

[7] Chung, H.(2012), “Export Market Orientation, Managerial Ties, and Performance,” International Marketing Review, 29(4), 403-423.

[8] 윤덕룡 · 김효상. (2017). 환율변화가 한국기업에 미치는 영향분석과 정책적 시사점: 기업데이터 분석을 중심으로. 대외경제정책연구원.

[9] 한민수 and 편주현. (2021). 환율 변화가 기업 규모별 우리나라 수출기업에 미치는 영향. 시장경제연구, 50(1), 99-127.

[10] 최기홍, 김동윤. (2019). 발틱운임지수가 한국 주가 변동성에 미치는 영향. 한국항만경제학회지, 35(2), 61-75.

*[11] 서지영.서현원/KDI 경제정보센터, 2010.08.02,
https://eiec.kdi.re.kr/material/clickView.do?click_yymm=201512&cid=1267*

[12] 김종진, 박성진 and 박지원. (2020). 수입곡물 가격변동이 국내 소비자물가에 미치는 영향 분석. 농촌경제, 43(2), 29-58.

*[13] 임현우, “*WTI·브렌트유·두바이유···'3대 原油'는 고향이 모두 달라요”, 한국경제, 2021.03.01,
[*https://sgsg.hankyung.com/article/2021022654681*](https://sgsg.hankyung.com/article/2021022654681)

[14] 김상배. (2021). 유가 변화가 우리나라 주식시장 유동성에 미치는 영향. 에너지경제연구, 20(1), 79-103.

[15] 김바우, 김정현 and 강성우, 최근 원자재 가격 상승의 배경과 국내 제조업에 미치는 영향, KIET 산업경제, 2021.11.09

*[16] 금융위원회 증권제도과, 1999-07-06, [해설] 선물거래의 기본구조,
https://www.fsc.go.kr/po010101/63344?srchCtgry=&curPage=1160&srchKey=&srchText=&srchBeginDt=&srchEndDt=*

[17] 강내영 and 강성은. (2021). 국제원자재 가격의 변동요인 및 우리 수출에의 영향 분석. 한국무역협회, 17.

*[18] 천성현. “왜 HR이 ESG에 주목해야 하는가”, HR insight, 2020.12.31,
https://www.hrinsight.co.kr/view/view.asp?in_cate=112&bi_pid=31723*

*[19] 통계청, 통계용어및지표,
https://kostat.go.kr/understand/info/info_lge/1/detail_lang.action?bmode=detail_lang&pageNo=7&keyWord=0&cd=SL4175&sTt=*

[20] 지성권 and 김태수. (2001). 경영자 보상과 회계이익특성의 관계에 관한 실증연구. 회계학연구, 26(1), 25-54.

[21] 한민수, 편주현. (2021). 환율 변화가 기업 규모별 우리나라 수출기업에 미치는 영향. 시장경제연구, 50(1), 99-127. Minsoo Han, Ju Hyun Pyun. (2021). Effects of exchange rate changes on Korean export firms by firm size. SOGANG ECONOMIC PAPERS, 50(1), 99-127.

[22] *오낙교, 윤성수 and 박원구. (2014). 알트만 Z-스코어를 이용한 신생 중소기업의 지속가능성 분석: 신재생에너지산업을 중심으로. 기술혁신연구, 22(2), 185-220.*

[23] 분석가 꿈냥이, 2022.08.25, https://zephyrus1111.tistory.com/271

[그림1] KOSI 중소벤처기업연구원. 중소기업 포커스 제 21-14호 2021.08.09 : ESG 확산이 중소기업에 미치는 영향 및 지원 방향. 나수미

The background features a grayscale city skyline at the top, a set of stairs leading upwards in the center, and several upward-pointing arrows of varying sizes on the left and right sides. A large blue rectangle with diagonal lines at its corners is positioned in the middle of the slide.

Thank You