

2023 통계데이터 활용대회

비재무데이터를 활용한 중소기업 휴폐업예측 모델 제안 및 결정요인 분석



INDEX

1. 연구 소개

연구 배경

연구 목적

2. 재무데이터 수집 및 전처리

데이터 소개

변수 선택

확정 재무데이터

파생변수 생성

재무데이터 결측치 처리

3. 재무데이터 모델링

재무데이터 모델 성능 비교 및 해석

4. 비재무데이터 수집 및 전처리

비재무데이터 수집

금융 감성분석

비재무데이터 결측치 처리

5. 재무+비재무데이터 모델링

재무+비재무 모델 성능

모델 변수 중요도 파악

재무+비재무 모델 Feature Importances

폐업 예측에 대한 SHAP

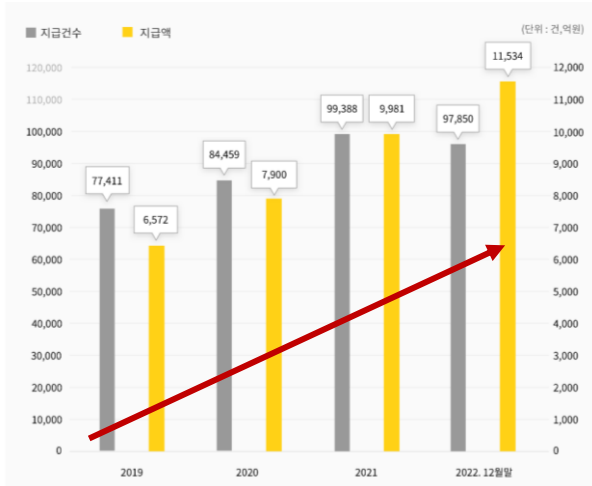
6. 결론

정책 활용방안 및 기대효과

연구 요약

고금리, 고물가, 고환율에 수출과 내수 침체가 겹친 복합 위기가 길어지면서 많은 중소기업이 재정위기에 놓였다. 이러한 상황에서 대기업 및 중견기업에 비해 투자유치에 한계가 있는 중소기업 대상 여신 리스크 관리의 필요성이 대두되고 있다. 따라서 대출 전 **건전성이 높은 기업을 선별하기 위해 중소기업 휴폐업 요인을 파악하는 것은** 중요한 과제이다. 이에 본 연구는 중소기업의 휴폐업 예측을 위해 사용되는 전통적 지표인 재무변수와 더불어, 데이터의 특성 상 다루기 어려워 활용성이 낮았던 **비재무데이터에 주목했다**. 이에 통계청에서 이미 수집하고 있는 비재무데이터를 이용하고 **Feature Importances와 SHAP**을 적용해 어떤 변수가 휴폐업에 주요한 영향을 미치는지를 제시하였다. 나아가 중요성이 확인된 변수를 통해 정책에 **활용할 수 있는 방안**을 제언하였다.

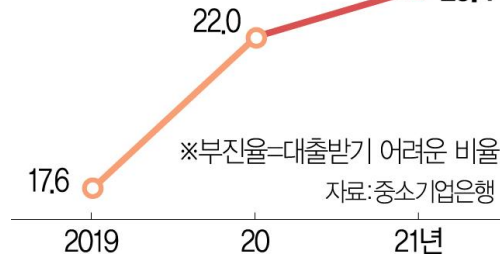
연구 배경



출처: 노란우산공제

중소기업의 은행 차입여건 부진율

(단위: %)



2022 폐업공제금 역대 최대

재무지표가 부족한 중소기업은
대출이 어려운 환경중소기업도 대출심사 시 정밀하게
평가받을 수 있는 평가모형이 필요

연구 목적

1

비재무데이터를 활용하여
재무지표만 사용한 모델보다
개선된 성능의 휴폐업 예측 모델을 개발

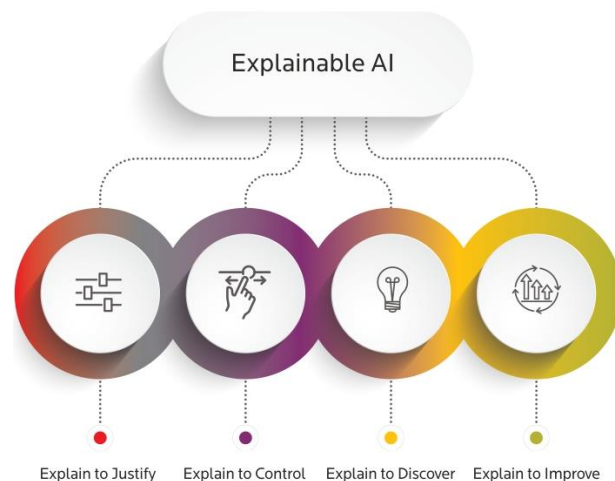


비재무데이터란?

재무관련 지표 외 기업 관련 다양한 자료
Ex. 평균 연봉, 직원 수, 퇴사율, 관련 뉴스기사

2

개선된 모델에서 각 비재무적 요인이
모델에 미치는 영향을 해석하고 중요도를
파악하여 활용 방안 제시



2

재무데이터 수집 및 전처리

데이터 소개

외부감사를 받은 중소기업 중
휴폐업 여부 데이터가 존재하는
 중소기업의 **자산, 부채, 자본** 데이터

유동자산	매출채권	비유동자산	...
24,965,602	NA	219,000	...
615,988	185,615	42,385	...
1,596,577	3,802	353,082	...
30,077,990	6,034,264	15,251,977	...
28,431,311	15,948,763	21,334,735	...
...

변수 선택

t-검정

단일 변량 기준 집단 간 평균의 차이가 존재하는지
 통계적으로 검정하는 방법



분류 성능 개선을 위해 사전적으로
 각각의 재무 변수에 대해 t-검정을 통해
 생존기업과 부도기업 간 평균의 차이가 있는지를 검토
 $P\text{-value} < 0.05$ 만족하는 유의한 변수만 선택

김용환(2022), 비정형 데이터를 활용한 머신러닝 기반 기업신용평가모형에 관한 연구



총 88개의 재무 변수에 대해
 통계적으로 유의한 47개의 재무 변수 선택

확정 재무데이터

* 재무데이터는 보유 자료 중 가장 최신 자료를 기준으로 작성

변수명	변수명	변수명
유동자산	법인세비용	유동부채비율
매출채권	당기순이익손실	비유동부채비율
비유동자산	기업순이익률	부채총계대매출액
유형자산	유보액총자산	총자본회전율
자산총계	유보액납입자본	재고자산회전율
유동부채	매출액총이익률	매출채권회전율
비유동부채	매출액영업이익률	매입채무회전율
부채총계	매출액순이익률	미수금
자본금	경상수지비율	매출원가
이익잉여금결손금	금융비용대매출액비율	재고자산
자본총계	금융비용대부채비율	순운전자본비율
매출액	금융비용대총비용비율	
판매비와관리비	부채비율	
영업이익손실	차입금의존도	
법인세비용차감전 순손익	자기자본비율	

파생변수 생성

파생변수명	설명
재무결측치개수	해당 기업의 재무관련변수 결측치 개수
매출액총이익률 변화율	보유한 자료의 결산기간 중 매출액총이익률변화율
금융비용대매출액 비율변화율	보유한 자료의 결산기간 중 금융비용대매출액비율 변화율
금융비용대총비용 비율변화율	보유한 자료의 결산기간 중 금융비용대총비용비율 변화율
변화율결측	변화율이 결측치인 재무관련변수의 개수
변화율이상치	변화율이 이상치인 재무관련변수의 개수
Target	휴폐업여부(binary) 0: 생존기업 1: 휴폐업기업

재무데이터 결측치 처리

- 결측치 포함 기업 단순 제거 시
너무 많은 데이터가 소실
- 평균 등 대표값으로 대체 시
업종 별 차이 반영이 어려움



재무데이터의 특성 상
0으로 대체할 수 있는 변수가 다수



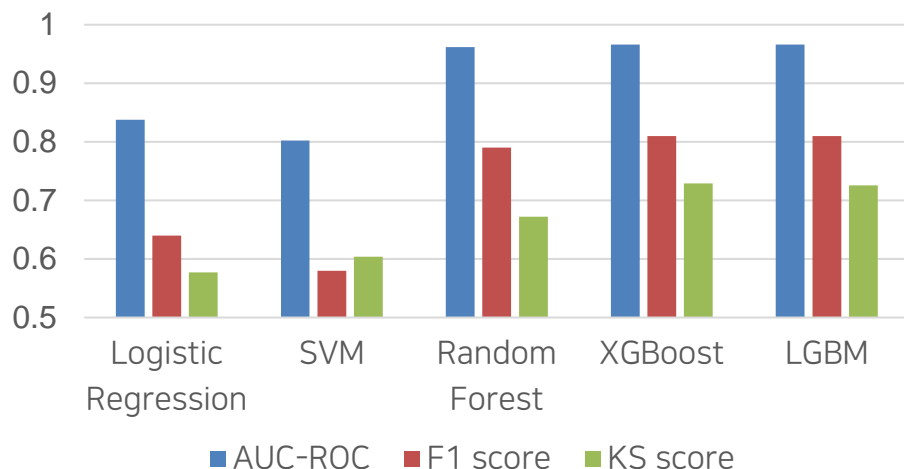
특성을 고려하여 결측치는 0으로 대체

유동자산	...	미수금	...	당기순이익손실	...	변화율 이상치
615,988	...	0	...	9,373,300	...	5
30,077,990	...	968,621	...	-10,703	...	0
28,431,311	...	0	...	0	...	0
...

[결측치 대체 후 데이터셋]

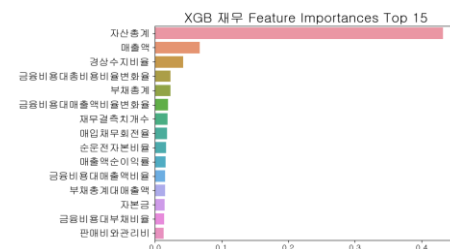
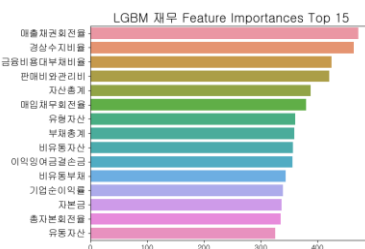
재무데이터 모델 성능 비교 및 해석

재무데이터 활용 분류모델 성능 비교



LGBM, XGBoost 등

트리 기반 머신러닝 모델의 성능이 우수



XGBoost

트리 기반 Gradient Boosting 모형
Level-wise(균형 트리 분할) 확장 방식
과적합 방지

LightGBM

트리 기반 Gradient Boosting 모형
Leaf-wise(리프 중심 트리 분할) 확장 방식으로 빠른 속도
대용량 데이터에 적합

* 부스팅(Boosting): 약한 예측 모형들의 학습 에러에 가중치를 두고, 순차적으로 다음 학습 모델에 반영하여 강한 예측모형을 만드는 것

비재무데이터 수집

N 뉴스

네이버 뉴스

‘뉴스의 감성분석이 부도예측 연구에서 예측 성능을 향상시킨다’는 선행연구

네이버 뉴스 API 사용해 대상 기업명과 정확히 일치하는 뉴스의 제목과 본문을 수집

→ 기업명 기준 **뉴스 기사** 크롤링



크레딧잡

‘인적자원관리가 기업성과에 영향을 준다’는 선행연구

크레딧잡에 검색해 해당 기업의 연봉(만 원), 입사율, 퇴사율 수집

→ **입사율, 퇴사율, 연봉**을 크롤링

기타 비재무 데이터

상용근로자(남), 상용근로자(여), 임시근로자(남), 임시근로자(여), 조직형태코드, 조직구분, 본점지점구분, 국외투자법인여부, 벤처기업여부, 중견기업보호여부, 설립구분, 직원수, 한국표준산업분류

4

비재무데이터 수집 및 전처리

금융 감성분석

금융감성분석



파생변수 생성

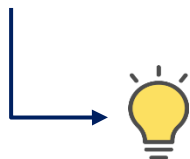
KLUE-BERT 모델에
금융감성 label 사전학습



사전학습된 모델에
수집한 텍스트데이터를
문장 단위로 입력



반환된 문장 별 label의 비율과
기사 수를 데이터셋에 추가



KLUE-BERT : 벤치마크 데이터인 KLUE에서 베이스라인으로 사용되었던 모델
KoBERT에 비해 더 최신화되고 다양한 출처의 데이터로 학습되어 성능이 우수

파생변수 생성
전체 기사 수,
긍정·부정·중립 기사 비율

기업명	기사 내 문장 예시
울**설	23개사 모임으로 시작한 차세대 클럽은 10여년만인 올해 회원사가 72개로 늘었고...
울**설	제 월급은 대기업 다닐 때보다 반으로 줄었고요

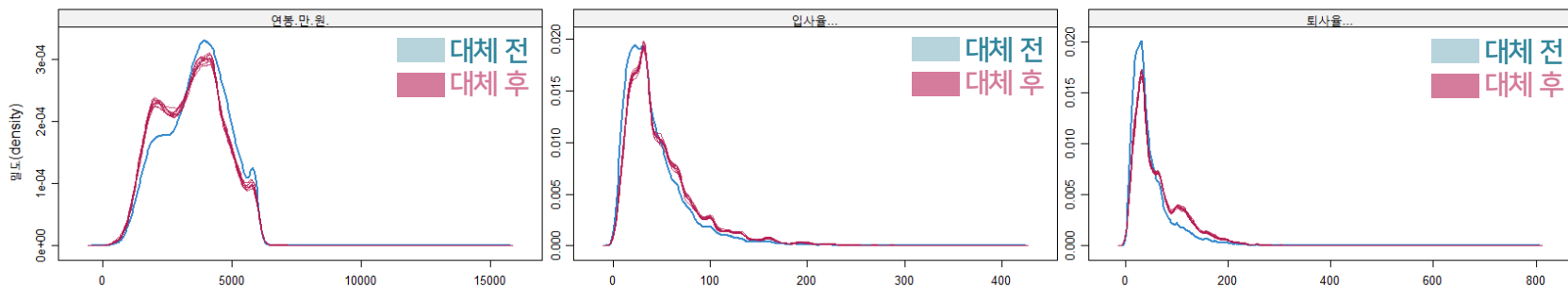
기업명	기사 수	긍정	부정	중립
울**설	40	0.82	0.10	0.07

비재무데이터 결측치 처리

연봉, 입사율, 퇴사율

MICE 사용해 다중 대체 (Method = 'rf')

* MICE : 조건부 분포를 활용해 여러 개의 독립적인 모델을 사용하여 반복적으로 결측치를 예측하고, 결측치의 불확실성과 분산을 반영해 최종 데이터셋을 구성하는 방법



뉴스 기사 파생변수

기사 수, 긍정 비율, 부정 비율, 중립 비율

모두 0으로 대체

기타 비재무 변수

직원수 : 동일 산업군 내 평균

설립구분, 한국표준산업분류 : 추가 범주('미분류') 생성

재무+비재무 모델 성능

모델	AUC-ROC	F1 score	KS score
Logistic Regression	0.838	0.64	0.577
SVM	0.802	0.58	0.604
Random Forest	0.962	0.79	0.672
XGBoost	0.966	0.81	0.719
LGBM	0.966	0.81	0.726

[재무 모델]



모델	AUC-ROC	F1 score	KS score
Logistic Regression	0.893	0.70	0.652
SVM	0.851	0.69	0.702
Random Forest	0.973	0.79	0.667
XGBoost	0.975	0.84	0.761
LGBM	0.980	0.85	0.778

[재무+비재무 모델]

LGBM, XGBoost, Random Forest 등 트리 기반 머신러닝 모델의 성능이 우수

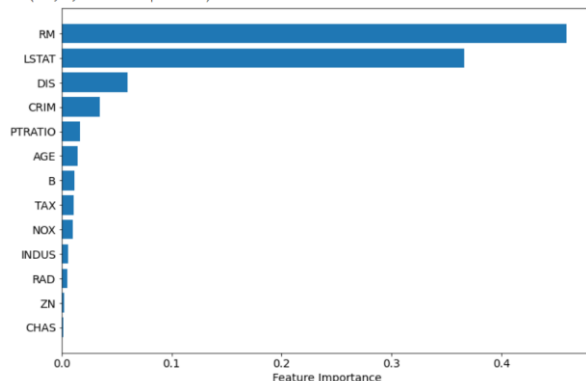
비재무 변수를 추가했을 때 전반적으로 성능이 개선

모델 변수 중요도 파악

Feature importances

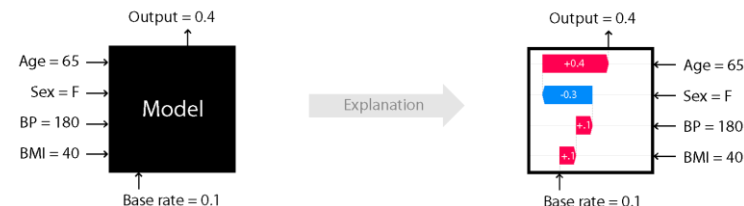
- 모든 Feature에서 하나의 Feature를 제거하고 학습해 성능을 측정했을 때 성능이 하락하는 만큼 제거한 **Feature가 Prediction에 중요한 영향을 끼쳤다고** 보는 것
- 단, 변수가 어떤 방향으로 영향을 미치는지 확인할 수 없음
- 중요도에 따라 모델 수정 시 해당 변수의 중요도가 **inconsistent**함

Text(0.5, 0, 'Feature Importance')

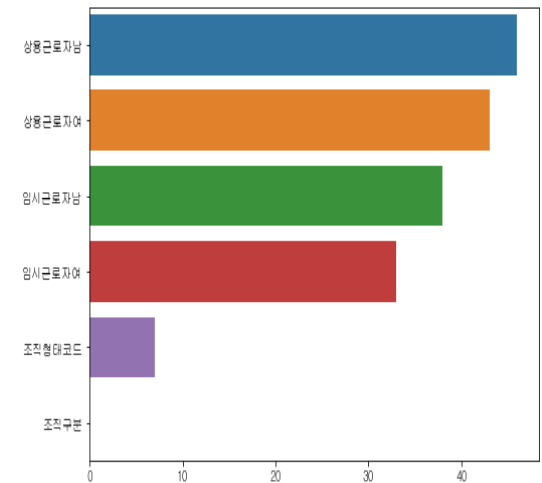
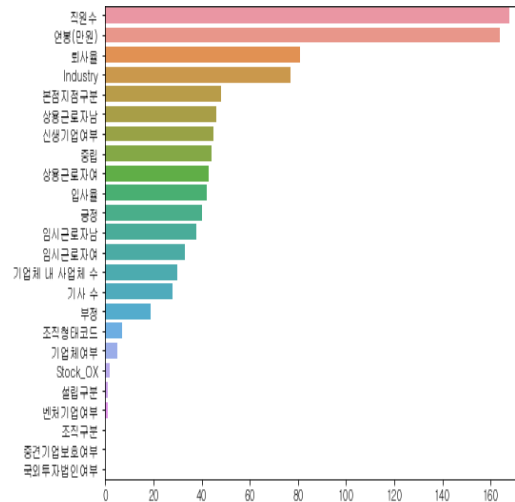
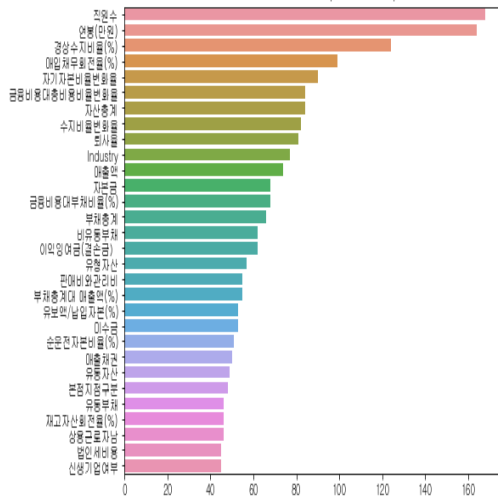


SHAP(SHapley Additive exPlanation)

- 변수중요도 뿐만 아니라 개별 예측값에 대한 각 변수들의 영향력을 모형 클래스에 상관없이 누적으로 배분하는 방식
- 인간의 생각과 유사한 해석을 제공하며, **consistent**하고 변수중요도의 **방향성**을 알 수 있음



재무+비재무 모델 Feature Importances



모델 중요변수 해석

- 모델링에 사용한 모든 변수들의 변수 중요도를 살펴본 결과,

직원수, 연봉, 경상수지비율, 매입채무회전율 등이 중소기업 휴폐업 예측에 영향을 많이 미치는 변수로 판명됨

- 비재무 변수들 간의 변수 중요도를 살펴본 결과,

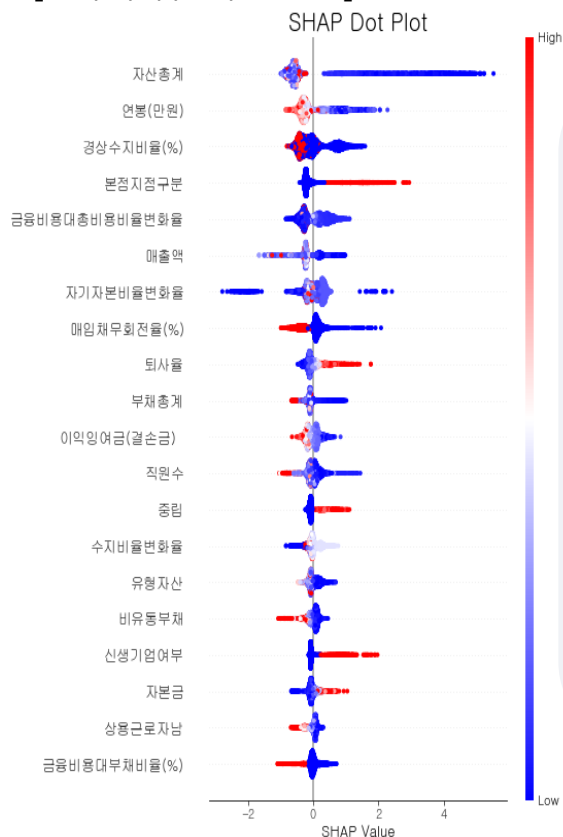
직원수, 연봉, 퇴사율 등이 중소기업 휴폐업 예측에 영향을 많이 미치는 변수로 판명됨

- 근로형태 관련 비재무 변수를 별도 추출해 변수 중요도를 살펴본 결과,

상용근로자 수가 임시근로자 수에 비해 예측에 영향을 많이 미치는 변수라는 것이 판명됨

폐업 예측에 대한 SHAP

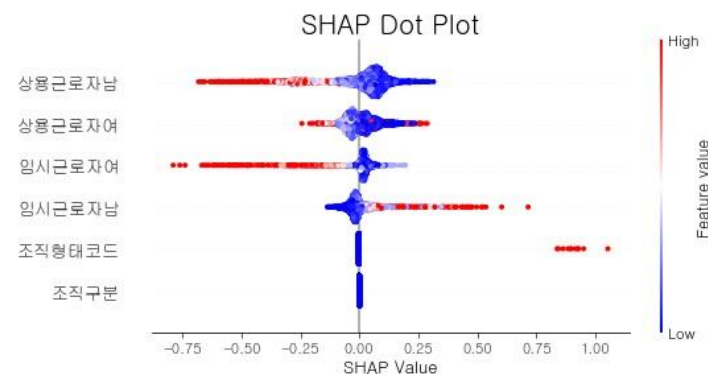
[전체 비재무변수 SHAP]



SHAP 기반 중요도 해석

연봉: 폐업률과 **음**의 관계
 직원수: 폐업률과 **음**의 관계
 퇴사율: 폐업률과 **양**의 관계
 기사 수: 영향력 파악이 어려움
 신생기업: 폐업률과 **양**의 관계
 종립기사비율: 폐업률과 **양**의 관계

[근로형태 비재무변수 SHAP]



SHAP 기반 중요도 해석

상용근로자(남): 폐업률과 **음**의 관계
 상용근로자(여): 영향력 파악이 어려움
 임시근로자(여): 폐업률과 **음**의 관계
 임시근로자(남): 폐업률과 **양**의 관계

* 중소기업의 기사형 광고는 기업의 전망에 부정적 영향을 주는 것으로 해석

정책 활용방안 및 기대 효과 | ① 공공기관 및 공기업

창업지원금의 선별적 제공

- 정부의 창업 지원 혜택만 이용하는 **체리피커형 스타트업** 多
- 설립 5년 이상 된 청년창업기업 10곳 중 6곳은 고용 전무
- 정부 지원금을 받고 재창업한 10곳 중 7곳이 지원 3년 후 고용인원이 1명 이하로 떨어짐
- **정책 활용방안 : 재무 지표가 부족한 초기 창업 기업이라도 비재무 지표를 통해 도산 여부 예측**
→ 유망 기업 위주의 지원 가능

정책금융기관 대출 심사 강화

- 중소기업에 대한 지원이 많은 정책금융기관 특성상 신용위험이 높은 기업 여신이 多
- **정책 활용방안: 비재무지표를 통합적으로 고려한 대출 심사를 통해 건전성이 더욱 높은 중소기업 선별**

주요 시중은행 NPL 비율 (단위: %)
(6월 말 기준) 자료: 각 사



주요 시중은행 NPL 비율 / 그래픽=김은실 디자이너

기대효과

- 창업지원 정책의 실효성 강화
- 정책금융기관의 부실채권비율 완화

정책 활용방안 및 기대 효과 | ② 중소기업

재무적 요소와 비재무적 요소의 균형성장 독려

- 기업 경영자가 비재무적 요소가 기업 전망 예측에 있어 중요한 역할을 한다는 점을 인식
→ 비재무적 요소와 재무적 요소의 균형 있는 발전을 위한 노력
- **정책 활용방안: 중소기업의 균형성장을 독려하는 조직문화 조성 지원 확대**

신생기업의 비재무지표 평가

- 업력이 짧아 자본이 부족한 신생기업일지라도 비재무적 요인이 우수하다면 신용도 제고 효과를 누릴 수 있음
- **정책 활용방안: 휴폐업 예측에 있어 유의미한 요소로 작용한 인적관리 관련 변수(연봉, 입사율 및 퇴사율 등)를 기업 신용평가 보조 지표로 사용**



기대효과

- 인식의 전환을 통해 비재무적인 요소의 성장을 위해 노력
→ 지속가능한 기업(=ESG 기업)으로 성장
- 비재무적 측면이 우수한 신생기업이 가치를 저평가 받는 것을 예방

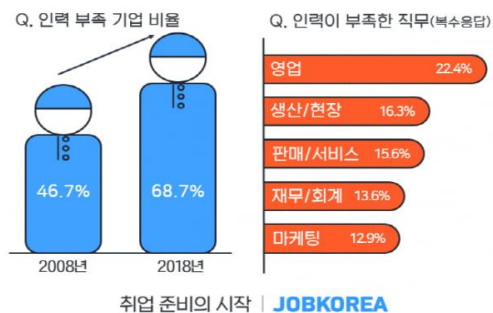
정책 활용방안 및 기대 효과 | ③ 사회

중소기업 업무환경 개선

- 중소기업의 10곳 중 7곳은 인력난을 겪고 있음
- 중소기업의 직장 만족도가 전체 기업군 중 가장 낮음
- **정책 활용방안:** 인적자원관리 능력과 기업 전망의 매우 높은 상관성을 입증한 본 연구 결과를 통해, 중소기업의 전반적 업무환경 개선을 독려할 수 있음

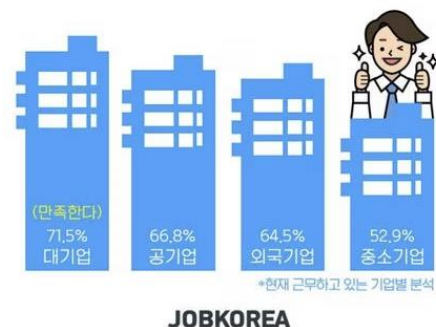
중소기업 인력부족 현황

※중소기업 214개사 대상 조사, 자료: 잡코리아



‘근로자의 날’ 직장 만족도 현황

※ 남녀 직장인 2,258명 대상 조사, 자료: 잡코리아



기대효과

- 업무환경의 개선으로 중소기업의 인력난 해소 및 우수 인재 유입 효과
- 직원 퇴사율 감소로 인적 자원 손실 예방

참고문헌

- 현지원, 이준일, 조현권(2022). KoBERT를 이용한 기업관련 신문기사 감성 분류 연구. 회계학연구. 47권(4호)
- 정승호(2022). TabNet을 활용한 딥러닝 성능 비교와 설명가능한 AI 활용성에 대한 연구-기업 신용평가 모형을 중심으로. 서강대학교 대학원.
- 남윤미(2017). 국내 자영업의 폐업률 결정요인 분석. BOK 경제연구. 5
- 장제훈(2021). 머신러닝 기법을 활용한 자영업자 폐업 예측 모형 연구 서울시 25개 자치구를 중심으로. 인문사회. 12권(1호)
- 박준식, 최진, 이성호(2023). 머신러닝을 이용한 소상공인 창업기업의 폐업 예측 모형 개발-도소매산업을 중심으로. 한국창업학회지. 18권(1호)
- 유원규, 이철규(2015). 비재무적 요인이 중소벤처기업의 신용평가에 미치는 영향. 대한경영학회지. 28권(12호)
- 김용환(2022). 비정형 데이터를 활용한 머신러닝 기반 기업 신용평가 모형에 관한 연구. 송실대학교 대학원.
- 방준아, 손광민, 이소정, 이현근, 조수빈(2018). 서울 치킨집 폐업 예측 모형 개발 연구. 한국빅데이터학회지. 3권(2호)
- 김용환, 김도형, 허재혁, 김광용(2022). 오차 앙상블모형을 활용한 기업 신용평가모형의 비교 연구. The Journal of Korean Institute of Communications and Information Sciences. 47권(1호)

감사합니다

