



중소기업건강도지수를 활용한 중소기업 휴·폐업 예측

작성자: 권나연, 유영훈

작성 일: 2022년 11월 23일

1 분석 배경

분석 목표

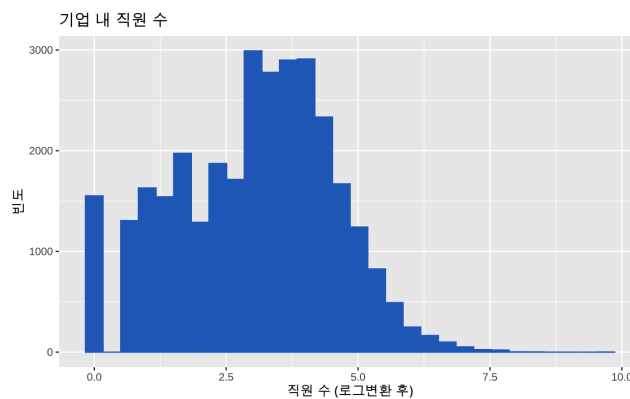
- SBHI 지수 (중소기업건강도지수)를 활용하여 중소기업의 휴·폐업 예측모델을 구축하는 것

중소기업 휴·폐업 예측에 비재무 데이터를 활용하는 이유

신용평가란

- 기업의 신용상태에 영향을 미치는 제반 경영적 사실을 조사하여 분석한 결과를 토대로 그 기업의 신용도에 관한 종합적인 판단을 내리는 것이다.
- 일반적으로 기업의 신용등급 평가 시 재무적 요인을 사용하지만, 비재무적 요인 또한 고려되어야 한다. 재무적 요인을 토대로 한 정보는 객관화하기 쉽지만 정보의 범위가 제한적이기 때문에 신용평가 및 기업도산의 예측에 한계를 가지고 있기 때문이다.
- 특히 중소기업의 경우 대기업에 비해 재무적 요인의 신뢰성이 낮을 뿐만 아니라 재무구조의 취약성과 여건변화에 민감한 사업구조 등으로 인해 대기업에 비해 상대적으로 비재무적 요인의 평가가 중요하다.

표본 내 기업의 규모



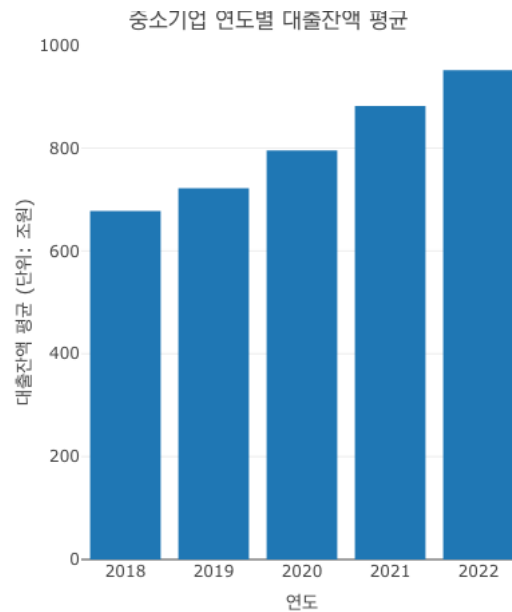
- 기업 내 직원 수 변수의 히스토그램과 사분위수를 통해 대부분의 기업 규모가 작다는 것을 알 수 있다. 히스토그램은 오른쪽 꼬리가 긴 분포를 보였으며, 기업의 75%는 직원 수가 58명 이하로 나타났기 때문이다 (사분위수는 [Appendix A](#)를 통해 확인 가능).
- 상대적으로 회계정보의 신뢰성이 낮은 일정 규모 이하의 중소벤처기업에 대해서는 비재무적 요인의 중요성이 더욱 커진다. 그러므로 예측 대상이 되는 데이터 내 기업의 휴·폐업 예측에는 비재무 데이터가 큰 영향을 미칠 수 있다.

Notes: 기업 내 직원 수 변수의 경우 0 이상이고, 범위 (range)가 넓다는 특징을 가지고 있으므로 로그 변환 후 그래프를 생성함

현 상황에서 비재무 데이터의 중요성

- 코로나19 이후 은행들은 공개된 정보가 한정적이고 상환 능력을 평가하기 까다로운 중소기업에 담보나 보증을 요구하는 경향이 강해졌다. 2020년 중소기업중앙회가 조사한 '2020년 중소기업 금융이용 및 애로실태 조사'에 따르면, 신규 대출을 받은 기업들은 '부동산 담보 부족'(24.1%)을 주요 애로사항으로 꼽았다. 이와 같이 은행의 담보 대출 비중이 늘어나면 담보가 부족하거나 외부기관 보증을 받지 못한 기업은 금융 접근성이 더 낮아질 수 있다.
- 하지만 담보가 부족한 기업에 보증부 대출을 늘리는 것은 지속 가능한 전략이 아니다. 기업이 대출상환을 하지 못하는 경우 등을 고려해야 하기 때문이다.

- 아래의 그래프를 통해 중소기업의 연도별 대출잔액 (2018~2022)이 증가 양상을 보인다는 것을 알 수 있다. 또한 올해 9월 한국은행의 금융 안정 상황 보고서에 따르면 한계기업의 비중이 높아지고 있으며 이들의 부실 위험도 커지고 있다.
 - 한계기업: 재무구조가 부실해 영업 활동으로 창출된 이익으로 이자비용조차 감당하지 못하는 기업을 뜻함



출처: 중소벤처기업부 통계자료

- 이러한 사회·경제적 상황에서의 지속 가능한 전략은 기업의 비재무 정보를 이용하여 중소기업의 신용평가를 다각적으로 하는 것이다.

다각적 신용평가 시 각 이해관계자의 이익

- 은행: 다양한 비재무 정보를 활용하여 대출 받는 기업들의 신용위험을 체계적으로 측정하고 관리할 수 있다.
- 중소기업: 담보대출의 대상이 되는 보유자산 규모가 작지만 장래 수익을 발생시킬 수 있는 기술력 등을 가진 중소기업들이 적극적으로 대출을 받을 수 있다.

2 활용 데이터 소개

외부 데이터

중소기업경기전망조사 데이터

- 중소벤처기업부는 중소기업의 경기 동향을 업종별로 조사하여 업계의 경영 계획 수립과 정부의 중소기업 지원 정책에 필요한 기초자료를 제공한다. 본 분석에서는 중소벤처기업부에서 제공하는 2018년 1월부터 2022년 6월 (상반기)까지의 월간 경기동향 실적 (경기전반, 생산, 내수 판매 등) 데이터를 활용하였다. 특히, 이 중 업종이 제조업으로 분류된 기업들의 SBHI 지수를 사용하였다.
- 활용 데이터 목록

데이터 명	활용 변수 명	설명
경기전반 실적	연평균 SBHI 지수	경기전반 실적의 SBHI가 100 이상이면 다음달 경기가 좋아질 것으로 예상하는 업체가 그렇지 않은 것으로 보는 업체보다 더 많음을 나타냄
고용수준 실적	연평균 SBHI 지수	역계열에 해당하므로, 경기 확장기에 SBHI 지수가 하락한다는 특징을 가짐
내수판매 실적	연평균 SBHI 지수	-
생산설비수준 실적	연평균 SBHI 지수	역계열에 해당함
영업이익 실적	연평균 SBHI 지수	-
원자재조달사정 실적	연평균 SBHI 지수	-
자금사정 실적	연평균 SBHI 지수	-
제품재고수준 실적	연평균 SBHI 지수	역계열에 해당함

데이터 명	활용 변수 명	설명
평균가동률	연평균 평균가동률	중소제조업 평균가동률은 보유 생산설비의 월간 생산능력대비 해당 월의 평균 생산량비율을 의미함

- 수집 대상: 매출액 5억원 이상의 중소기업
- 각 데이터의 관측값 수: 총 58개 (2018년 1월~2022년 6월)의 변수 포함하고 있으며 관측값의 수는 51개이다.
- SBHI (중소기업건강도지수) 지수
 - 산출방법: 경영부분 항목별로 5점 서열척도로 배분하여 가중치를 부여한 후 전체응답빈도수로 나누어 산출한다.
 - 변수 해석: SBHI가 100 이상이면 다음달 경기가 좋아질 것으로 전망한 업체가 그렇지 않은 것으로 보는 업체보다 더 많음을 나타낸다. 예외적으로 수준판단 항목인 생산설비, 제품재고수준, 고용수준의 SBHI는 일반적으로 경기 확장기에는 하락하고 경기 수축기에는 상승하는 역계열이다.
 - 또한 SBHI는 기존의 BSI 산출방식의 문제점을 보완한 지수로 본회 고유의 경기예측 지수이다.
 - $\text{기존 BSI} = \text{호전응답 업체비율}(\%) - \text{악화응답 업체비율}(\%) + 100$
 - 예측 모델에서 사용한 독립변수는 연평균 SBHI 지수를 의미하는 파생변수이다.

중소기업 대출잔액 데이터

- 중소벤처기업부에서 제공하는 2018년 1월부터 2022년 8월까지의 중소기업 대출잔액 데이터
- 연도별 대출잔액 평균을 나타내는 파생변수 (단위: 조원)를 생성하여 중소기업 신용평가 시 비재무 정보의 필요성을 제기하였다.
 - 2022년의 경우 1월부터 8월까지의 평균을 이용함

재무 데이터

- 정상적인 기업의 재무제표를 보여주는 변수와 재무제표 안정성 관련지표를 사용하였다.
 - 재무제표의 안정성 관련지표는 한국은행의 연간 기업분석결과 보도자료 (2021)를 참고하여 선택함
- 분석에 이용한 독립변수들의 목록은 다음의 표를 통해 확인할 수 있다.

변수 한글 명	변수 영문 명	설명
		정상적인 기업의 재무제표를 보여주는 변수
유동자산	CUR_AST	현금·예금·주식·사채 등의 유가증권
유형자산	TAN_AST	토지와 1년 이상의 내구성을 가진 건물·구축물, 기계장치, 선박·차량운반구, 공구·기구·비품, 건설 중인 자산 등 형체가 있는 자산
자산총계	TOTAL_AST	유형자산+유동자산
매출액	REVENUE	제품이나 상품 등을 판매하고 얻은 대가
판매비와 관리비	OPR_EXP	제품, 상품 등의 판매활동과 기업의 관리활동에서 발생하는 비용, 매출원가에 속하지 않는 모든 영업비용을 포함함
매출액영업이익률	OPR_GAIN	$(\text{영업이익}/\text{매출액}) \times 100\%$
매출원가	CST_SALES	매출을 실현하기 위한 생산이나 구매과정에서 발생한 상품과 서비스의 소비액과 기타 경비
		재무제표의 안정성 관련지표
부채비율	DEBT	$(\text{부채}/\text{자기자본}) \times 100$
차입금의존도	TOTAL_DEBT	$[(\text{차입금} + \text{회사채}) / \text{총자본}] \times 100$
자기자본비율	EQUITY	$(\text{자기자본} / \text{총자본}) \times 100$

3 탐색적 데이터 분석

요약 통계량

- Appendix B를 통해 최종 데이터 내 변수들의 요약 통계량을 확인할 수 있다.
- 요약 통계량 결과 요약
 - 업종 분류가 제조업인 기업이 99%이다. 그러므로 휴·폐업 예측 모델의 대상을 '제조업' 업종에 속한 기업으로 한정하였다. 또한 이 변수는 휴·폐업 예측에 유의미한 변수가 아니므로 예측 모델에 포함하지 않았다.

- 휴·폐업 구분 변수 (반응변수) 내 휴·폐업 기업의 비율은 18%이며, 액티브 기업의 비율은 82%이다. 그러므로 휴·폐업 기업과 액티브 기업의 비율 차이를 고려하여 데이터 셋 (train, valid, test)을 분리하였다.
 - 데이터 셋 분리 방법: 데이터를 휴·폐업 기업과 액티브 기업으로 나누어 8:1:1 로 데이터를 분리한 후 데이터를 다시 합치는 방법을 택함
- 공공기관 유형 변수는 '해당사항 없음'에 해당하는 기업의 비율이 100%이며, 중견기업 보호여부, 공기업 여부, 상장 여부 변수의 경우에도 '아니오'에 속한 기업이 100%이다. 또한 기업 규모는 모두 '중소기업'이며 개인법인구분 변수는 모두 '법인'에 해당한다. 이 변수들은 휴·폐업 예측에 유의미한 영향을 미치지 않을 것이라고 예상할 수 있으므로 분석에서 제외하였다.
- 연속형 변수의 경우 재무 데이터와 외부 데이터의 단위 차이가 크게 나타났다. 그러므로 예측 모델에 넣기 전 min-max scaling을 진행하였다.

시각화

- 개방 데이터의 주요사업내용 변수를 이용하여 워드 클라우드를 생성하였다. 그 결과, 중요도가 높은 단어에는 제조, 부동산, 판매업, 임대업, 개발, 도소매업 등이 있었다. 데이터 내 제조업 업종에 해당하는 기업이 대부분을 차지하는 것과 동일한 맥락의 결과이다.



4 데이터 전처리

예측모델 적용 대상

- 제조업 업종에 속한 중소기업

반응변수 정의

- 휴·폐업 구분 변수 (CLSBZ_GB)를 이용하여 반응변수를 정의하였다. 현재 휴·폐업 중소기업의 경우 1, 액티브 중소기업의 경우 0을 부여하여 binary classification 문제로 변환하였다.
- 현재 휴·폐업 상태인지에 초점을 맞추어 휴·폐업에서 액티브로 전환된 기업들의 경우에도 액티브 중소기업으로 분류하였다.

독립변수 생성

분석 시 활용 변수 전처리

- 재무 데이터와 외부 데이터의 연속형 변수들은 min-max scaling 하였다.
- 예측모델에 독립변수를 넣을 때 상태 발생일자 (STAT_OCR_DATE)의 연도 (2018~2022)에 대응하는 SBHI 변수를 사용하였다.
 - 예를 들어 예측모델에 경기전반 실적 SBHI 변수 이용 시, 관측값의 상태 발생일자가 20220518 이라면 2022년의 SBHI 평균을 이용하도록 했다.

SBHI1_SORT3	SBHI1_MEAN18	SBHI1_MEAN19	SBHI1_MEAN20	SBHI1_MEAN21	SBHI1_MEAN22
식료품	87.2	87.7	73.3	85.5	82.5
음료	95.2	96.8	82.8	86.2	88.5
섬유제품;의복제외	73.3	65.4	51.1	73.4	70.8
의복,의복악세서리및모피제품	77.9	80.4	65.5	73.1	77.1
가죽,가방및신발	78.0	74.7	51.2	59.9	66.9
목재및나무제품;가구제외	75.5	72.4	66.2	75.8	75.6
펄프,종이및종이제품	79.4	77.3	67.6	83.3	81.1

- 데이터 내 결측치 처리 방법은 다음의 표와 같다.

변수 한글 명	변수 영문 명	결측치 수	처리 방법
직원 수	EMP_CNT	1,434	중앙값으로 채움
유동자산	CUR_AST	180	0으로 채움
유형자산	TAN_AST	2228	0으로 채움
자산총계	TOTAL_AST	174	0으로 채움
매출액	REVENUE	1539	0으로 채움
판매비와 관리비	OPR_EXP	313	0으로 채움
매출액영업이익률	OPR_GAIN	91	0으로 채움
매출원가	CST_SALES	5322	0으로 채움
부채비율	DEBT	9	0으로 채움
차입금의존도	TOTAL_DEBT	77	0으로 채움
자기자본비율	EQUITY	9	0으로 채움

분석 시 제외한 변수

- 개방 데이터 내 모든 관측치가 동일한 값을 가진 변수들은 분석에서 제외하였다.
 - 제외 변수 명: CMP_SCL (기업 규모), PSN_CORP_GB (개인법인구분), MDSCO_PRTC_YN (중견기업 보호여부), PBCO_GB (공기업 여부), LIST_GB (상장여부), BZ_TYP (업종)
- 개방 데이터 내 이름과 관련된 변수들은 분석에서 제외하였다.
 - 제외 변수 명: CMP_PFIX_NM (기업접두명), CMP_NM (기업명), CMP_SFIX_NM (기업접미명), CMP_ENM (기업영문명), NATN_NM (국가명), HOMEPAGE_URL (홈페이지URL), CEO_NM (대표자명)

데이터 병합

- 개방 데이터와 재무 데이터 병합 시 사업자등록번호(BIZ_NO) 변수를 이용하여 개방 데이터를 기준으로 `left_join` 하였다.
- 개방 데이터와 외부 데이터 병합 시 10차 한국표준산업분류 개정 분류체계 (KSIC)의 중분류 코드를 이용하여 개방 데이터 기준 `left_join` 하였다.
- 다음의 표를 통해 외부 데이터 병합 시 사용한 KSIC 코드와 이에 대응하는 중분류 목록을 확인할 수 있다.

코드	중분류	코드	중분류
10	식료품	23	비금속광물제품
11	음료	24	1차금속
13	섬유제품;의복제외	25	금속가공제품;기계및가구제외
14	의복,의복악세서리및모피제품	26	전자부품,컴퓨터,영상,음향및통신장비
15	가죽,가방및신발	27	의료,정밀,광학기기및시계
16	목재및나무제품;가구제외	28	전기장비
17	펄프,종이및종이제품	29	기타기계및장비
18	인쇄및기록매체복제업	30	자동차및트레일러
20	화학물질및화학제품;의약품제외	31	기타운송장비
21	의료용물질및의약품	32	가구
22	고무제품및플라스틱제품	33	기타제품

총 관측값 수

- 데이터 전처리 후 총 관측값 수는 32,392개이며 총 변수는 28개이다.
- [Appendix B](#)를 통해 예측 시 사용한 변수 목록을 확인할 수 있다.

5 모델링

예측모델 개발

Random Forest

- 랜덤 포레스트는 다수의 결정 트리의 조합으로 예측하는 모델로써, 병렬로 트리를 작성한다. 그러므로 결정 트리의 개수가 지나치게 증가하여 모델 성능이 낮아지지 않는다는 장점이 있다.
- 각 결정 트리의 학습에서 행 데이터나 특징을 샘플링해 전달함으로써 다양한 결정 트리를 작성하고, 이들을 앙상블하여 일반화 성능이 높은 예측을 실시한다.
- 분류 문제에서는 지니 불순도가 가장 감소하도록 분기를 시행한다.

CatBoost

- 예측 시 범주형 변수가 중요 변수일 때 우수한 성능을 보이는 부스팅 기반의 모델이다.
- 훈련 (training) 속도는 다소 길게 나타날 수 있지만, 예측 (prediction) 속도는 다른 그래디언트 부스팅 결정 트리 (GBDT) 모델보다 13~16배 빠르다.
- CatBoost는 다른 GBDT 모델과 달리 하이퍼파라미터 튜닝에 따른 성능 차이가 크지 않다. 그러므로 결정 트리 앙상블에 익숙하지 않은 사람들이 사용하기 용이하다.

BART (Bayesian additive regression trees)

- 베이저안 확률 모형을 기반으로 하는 예측모델이며, 반응변수가 연속형, 범주형, 그리고 이진 (binary) 변수일 때 모두 사용 가능하다.
- 결정 트리 모델이며, 독립변수 정보를 이용하여 만들어진 의사결정규칙에 따라 관측값을 분류한다. BART 모델의 최종 목적은 나무의 구조 (가지의 개수, 의사결정규칙 등)와 마지막에 대푯값으로 설정되는 모수 μ 를 예측하는 것이다.
- BART 모델은 강한 모수적 가정을 피하면서 다양한 회귀 모델을 적합시키는 유연한 접근 방식을 제공한다는 장점이 있다.

예측모델 평가지표

AUROC (Area under ROC)

- ROC curve: 임계값 (threshold)을 0에서 1까지 변화시켜 가면서 x 축에는 거짓 긍정률 (False Positive Rate; FPR)을, y 축에는 참 긍정률 (True positive rate; TPR)을 표시해서 그린 곡선이다.
- AUROC: ROC curve 아래 면적을 의미하며, 이 값이 1에 가까울 수록 분류 모델의 성능이 좋다고 평가한다.

Accuracy

- 전체 데이터 중 예측을 정확하게 한 데이터의 비율이다.

F1 score

- F1 score는 정밀도 (precision)와 재현율 (recall)의 조화평균으로 다음과 같이 표현할 수 있다.

$$2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

- 정밀도 (precision): Positive로 예측한 데이터 중 실제 positive인 데이터의 비율이다.
- 재현율 (recall): 실제 positive인 데이터 중 모형이 positive로 예측한 데이터의 비율이다.

평가지표 선택 이유

- Binary classification 문제에서 대중적으로 활용되는 지표를 사용하였다.
- 모델을 종합적으로 평가하기 위하여 다양한 지표를 활용하였다.

- 특히 F1 score의 경우 데이터 레이블이 불균형 구조일 때 모델의 성능을 정확히 평가할 수 있는 지표이다.
 - 반응변수의 휴·폐업 기업과 액티브 기업 비율은 각각 18%, 82%의 불균형 구조이므로 모델의 평가지표로 적합하다고 할 수 있음

예측모델 평가 결과

사용 예측모델과 데이터

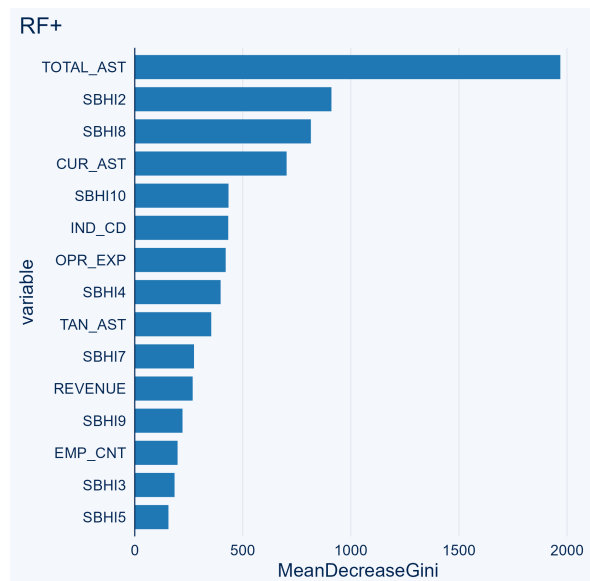
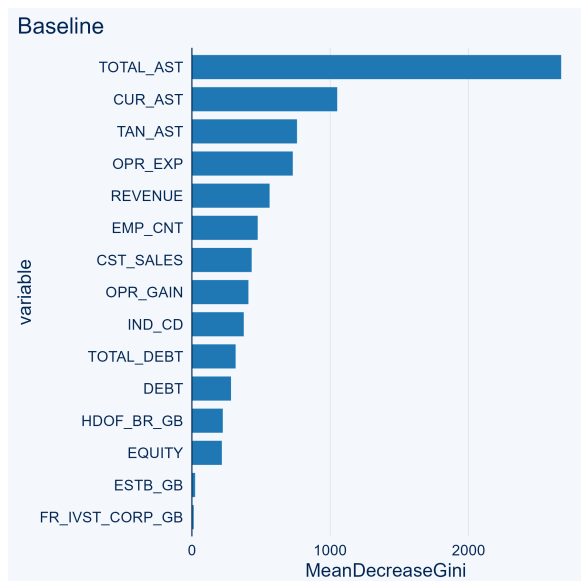
- 베이스라인: 분석 시 개방 데이터와 재무 데이터만을 이용하였으며, Random Forest 모델을 통해 분석함
- RF+: 분석 시 추가적으로 외부데이터를 활용하였으며, Random Forest 모델을 이용함
- CatBoost: 분석 시 모든 변수들을 포함하였으며, CatBoost 모델을 이용함
- BART: 분석 시 모든 변수들을 활용하였으며, BART 모델을 이용하여 분석을 진행함

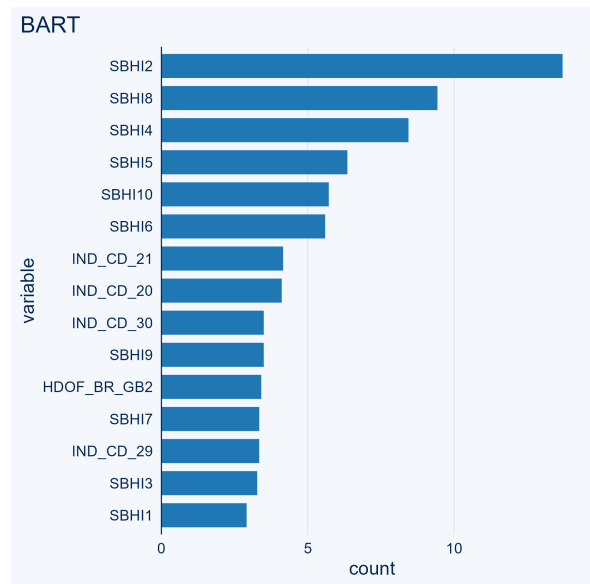
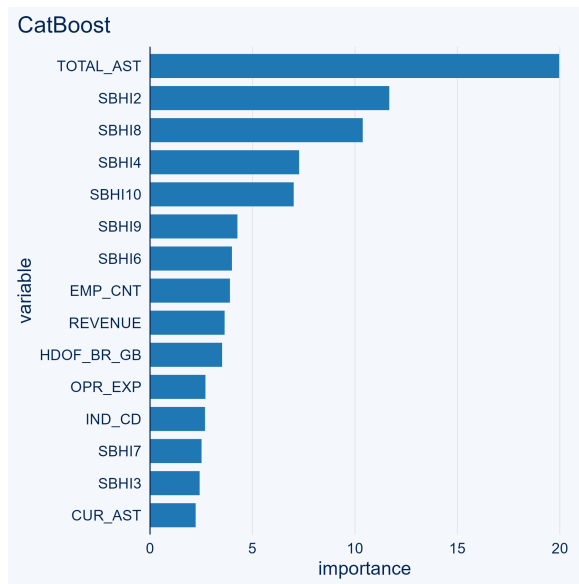
방법	사용 데이터
베이스라인	개방 데이터, 재무 데이터
RF+	개방 데이터, 재무 데이터, 외부 데이터
CatBoost	개방 데이터, 재무 데이터, 외부 데이터
BART	개방 데이터, 재무 데이터, 외부 데이터

예측모델 평가 결과

- AUROC: 모든 예측모델의 AUROC 값이 0.5~0.6 사이였으며 베이스라인 모델의 AUROC 값 (0.556)이 가장 컸다. AUROC를 기준으로 했을 때 모델의 성능이 높다고 평가할 수 없지만, AUROC 값은 데이터와 과업 (task)에 따라 다르게 나타날 수 있다.
- F1 score: CatBoost 모델의 F1 score가 0.992로 가장 높았으며, 모든 모델의 F1 score가 1에 가까운 값을 가진다. 그러므로 F1 score를 기준으로 예측모델을 평가하였을 때 우수한 성능을 보인다고 할 수 있다.
- Accuracy: F1 score와 마찬가지로 CatBoost 모델이 가장 높은 값 (0.987)을 가졌다. 또한 모든 모델의 accuracy가 0.9 이상이었다. 이는 모든 모델에서 전체 데이터 중 예측을 정확하게 한 데이터의 비율이 0.9 이상임을 뜻한다.
- 각 모델을 통해 선정된 중요도가 높은 변수들의 목록은 [Appendix C](#)를 통해 확인할 수 있다.

방법	AUROC	F1 score	Accuracy
베이스라인	0.556	0.964	0.940
RF+	0.511	0.991	0.985
CatBoost	0.504	0.992	0.987
BART	0.536	0.977	0.961





5 결과 및 인사이트

결과 요약 및 인사이트

- 개방 데이터의 변수들보다 재무 데이터의 변수들이 중소기업의 휴·폐업 예측에 중요한 변수였다.
- 외부 데이터인 SBHI (중소기업건강도지수)를 사용한 경우, 사용하지 않았을 때보다 F1 score와 Accuracy의 개선이 이루어졌다. 특히 F1 score는 본 데이터와 같이 데이터 레이블이 불균형 구조일 때 유용한 지표이므로, F1 score의 향상은 유의미한 결과라고 할 수 있다.
- BART를 제외한 모델에서는 자산총계가 가장 중요한 변수로 나타났지만 BART 모델에서는 고용수준 실적(SBHI2)이 가장 중요한 변수로 나타났다.
- 모든 모델 (Random Forest, CatBoost, BART)에서 최소 8개 이상의 SBHI 변수들이 중요한 변수로 나타났다.
- 고용수준 실적 (SBHI2), 내수판매 실적(SBHI3), 생산설비수준 실적 (SBHI4), 영업이익 실적 (SBHI7), 원자재조달사정 실적 (SBHI8), 자금사정 실적 (SBHI9), 제품재고수준 실적(SBHI10)은 외부데이터를 활용한 모든 모델에서 중요한 변수로 나타났다.
- SBHI (중소기업건강도지수)는 중소기업 종사자의 다음달 경기에 대한 전망을 나타내는 지수이다. 그러므로 중소기업 종사자의 전망이 중소기업 휴·폐업 예측에 중요 변수로써 이용될 수 있다고 할 수 있다.
- 이 결과는 중소벤처기업부가 진행하는 중소기업의 경기 동향에 대한 조사가 중소기업 신용평가 모델에도 활용될 수 있다는 가능성을 보여준다.

한계점

- 표본의 99%가 제조업 업종에 해당하는 기업이기 때문에 분석 결과를 다른 업종 (e.g., 서비스업)에 적용하기 어렵다.
- 외부 데이터 병합 시 10차 한국표준산업분류 개정 분류체계 (KSIC)를 이용하였는데, 개방 데이터 내에 이 기준을 따르지 않는 정보가 존재할 가능성이 있다.
- 외부 데이터 수집 시 기업의 경영자 요인 (경영자의 전문성, 리더십 등), 기술력 등의 세부적인 정보를 활용하는데 한계가 있었다.

References

가도와키 다이ске, 사카타 류지, 호사카 게이스케, & 히라마쓰 유지. (n.d.). 데이터가 뛰어노는 AI 놀이터, 캐글 (pp. 286-288). n.p.: 한빛 미디어.

유원중, & 이철규. (2015). 비재무적 요인이 중소벤처기업의 신용평가에 미치는 영향 (pp. 3191-3210). 대한경영학회지: 대한경영학회.

한국은행. (2021). 2021년 연간 기업경영분석 결과. 기업경영분석, 2022-10-14호, pp. 1-12.

갈수록 느는 중소기업 담보 대출...“은행 자체 여신 관리 강화해야”. (2021).

https://www.hani.co.kr/arti/economy/economy_general/982388.html.

[기업대출 적신호] 이자도 못내는 좀비 중소기업...금리충격에 줄도산하나. (2022).

<https://www.ajunews.com/view/20221016074233604>.

CatBoost – A New Game Of Machine Learning . (n.d.). <https://affine.ai/catboost-a-new-game-of-machine-learning/>.

Appendix

Appendix A: 기업 내 직원 수 변수의 사분위수

0%	25%	50%	75%	100%
0	6	23	58	16310

Appendix B: 요약 통계량

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/52888d37-86a4-454a-aed8-d59d9d2463ba/Summary_statistics.pdf

Appendix C: 예측모델 평가 결과 (변수 중요도 순으로 정렬)

베이스라인		RF+		CatBoost		BART	
TOTAL_AST	자산총계	TOTAL_AST	자산총계	TOTAL_AST	자산총계	SBHI2	고용
CUR_AST	유동자산	SBHI2	고용수준 실적	SBHI2	고용수준 실적	SBHI8	원자 실적
TAN_AST	유형자산	SBHI8	원자재조달사정 실적	SBHI8	원자재조달사정 실적	SBHI4	생산 적
OPR_EXP	판매비와 관리비	CUR_AST	유동자산	SBHI4	생산설비수준 실 적	SBHI5	생산
REVENUE	매출액	SBHI10	제품재고수준 실 적	SBHI10	제품재고수준 실 적	SBHI10	자금
EMP_CNT	직원수	IND_CD	산업코드	SBHI9	자금사정 실적	SBHI6	생산 적
CST_SALES	매출원가	OPR_EXP	판매비와 관리비	SBHI6	평균가동률	IND_CD_21	의료 품
OPR_GAIN	매출액영업이익률	SBHI4	생산설비수준 실 적	EMP_CNT	직원수	IND_CD_20	화학 품(으
IND_CD	산업코드	TAN_AST	유형자산	REVENUE	매출액	IND_CD_30	자동
TOTAL_DEBT	차입금의존도	SBHI7	영업이익 실적	HDOF_BR_GB	본점지점구분	SBHI9	자금
DEBT	부채비율	REVENUE	매출액	OPR_EXP	판매비와 관리비	HDOF_BR_GB	본점
HDOF_BR_GB	본점지점구분	SBHI9	자금사정 실적	IND_CD	산업코드	SBHI7	영업
EQUITY	자기자본비율	EMP_CNT	직원수	SBHI7	영업이익 실적	IND_CD_29	기타
ESTB_GB	설립구분	SBHI3	내수판매 실적	SBHI3	내수판매 실적	SBHI3	내수
FR_IVST_CORP_GB	판매비와 관리비	SBHI5	생산실적	CUR_AST	유동자산	SBHI1	경기