

산업 · 시장 경쟁구조 요인을 반영한 부도 예측: Explainable AI 접근

리우용상* · 김성환**

〈요 약〉

본 연구는 재무정보 중심의 전통적인 부도 예측모형이 과거 정보인 회계적인 관점에서 미래 관점에서 성장관점에서 산업 · 시장 구조적 경쟁요인을 충분히 반영하지 못한다는 한계를 보완하고, 예측력과 해석 가능성을 동시에 제고하는 분석 틀을 제시하는 것을 목적으로 한다. 이를 위해 부도위험을 연속형 지표인 부도거리(Distance to Default, DD)와 부도 여부(Default_D)로 구분하여 장기적 부도위험 수준과 단기 부도 발생 가능성을 병행 분석하였다. 주요 설명변수로는 산업집중도(HHI), 산업 성장률, 시장점유율, 기업 매출액 증가율을 포함하였다. 분석 방법론으로는 로지스틱 회귀, 고정효과 패널 회귀, 콕스 비례위험모형 등 전통 계량경제학 기법과 함께, Explainable AI 부스팅 머신(EBM)과 XGBoost를 구축하여 산업별 시장경쟁 구조(global) 및 기업별 재무적 특성(local) 수준의 해석을 수행하였다.

실증분석에는 CSMAR 데이터베이스에서 수집한 2016년부터 2023년까지의 중국 A주 상장기업 정보를 사용하였다. 분석 결과, DD 모형에서는 산업집중도가 부도 위험에 양(+)의 영향을, 산업의 매출액 증가율, 시장점유율과 기업의 매출액 증가율은 음(-)의 영향을 미치며 모두 통계적으로 유의하였다. 반면 부도여부를 종속변수로 한 모형에서는 기업의 매출액 증가율만이 유의한 음(-)의 효과를 보였다. 예측 성능은 XGBoost가 가장 우수하였고, EBM이 그 뒤를 이었으며, 로지스틱 회귀는 상대적으로 낮았다. 해석 결과, DD가 모든 모형에서 핵심 변수임이 확인되었고, 산업 수준 변수도 예측력 향상에 기여하였다. 특히 EBM은 산업 매출 증가율과 부도위험 간 비선형 관계를 규명하였고, XGBoost의 SHAP 분석 역시 정책적으로 유의미한 인사이트를 제공하였다.

본 연구는 산업 · 시장 경쟁 요인을 계량적으로 반영함으로써 부도 위험 분석의 범위를 확장하고, Explainable AI 기법을 통해 학문적 · 정책적 수준에서 예측 · 해석 통합형 모형을 제시했다는 점에서 의의가 있다.

주제어 : 기업 부도, 부도 거리, 산업 성장성, 기업 경쟁력, Explainable AI, EBM, XGBoost, SHAP

논문접수일 : 2025년 08월 16일 논문게재확정일: 2025년 08월 25일

* 제1저자, 경북대학교 경상대학 경영학부 박사과정, E-mail: tiwish@naver.com

** 교신저자, 경북대학교 경상대학 경영학부 교수, E-mail: indianak@knu.ac.kr

I. 서 론

기업 부도는 단일 기업의 경영 실패를 넘어 산업 전반과 지역 경제, 나아가 국가 경제 전반에 걸쳐 심각한 파급효과를 초래한다. 이러한 이유로 부도 위험 예측은 기업 경영진, 금융기관, 정책 당국 모두에게 핵심적인 과제로 왔다. 기존 부도 예측 연구의 상당수는 재무제표 기반 지표에 의존하여 기업의 부도 가능성을 평가해 왔다. Beaver(1966)와 Altman(1968)의 연구 이후, 재무비율과 통계모형(Logit, Probit, Survival Analysis 등)을 활용한 접근 방식은 그 시도만으로 예측정확도 향상을 보여주었으나, 기업이 속한 산업·시장 경쟁 구조의 특성을 충분히 반영하지 못한다는 한계가 있었다. 실제로 동일한 재무상태를 가진 기업이라도 경쟁이 치열한 산업에 속해 있거나 시장 점유율이 낮은 경우, 그렇지 않은 기업에 비해 부도 위험이 높을 수 있다. 반대로, 산업집중도가 높고 시장지배력이 강한 기업은 일시적 재무악화 상황에서도 생존 가능성이 크다. 그럼에도 불구하고, 기존 연구들은 기업의 재무정보에만 주로 의존하여 이러한 외부 산업 환경 요인을 정량적으로 분석한 사례가 부족하였다(Agarwal and Gort, 1999; Honjo, 2000; 김지홍, 문남미, 2023; Sinha and Kumar, 2024).

기업 부도 위험은 기업 자체의 재무적 요인뿐만 아니라, 해당 기업이 속한 산업의 구조적 특성과 시장 환경에 의해서도 크게 좌우된다. 산업조직이론(Industrial Organization Theory)과 Porter(1980)가 제시한 구조-행위-성과(Structure-Conduct-Performance, SCP) 패러다임에 따르면, 산업 내 경쟁 집중도, 산업 성장률, 개별 기업의 시장점유율은 가격 결정력, 진입장벽, 수익 안정성에 영향을 미쳐 부도 가능성을 변화시킬 수 있다. 예컨대 동일한 재무 상태를 가진 기업이라 하더라도 경쟁이 치열하고 진입장벽이 낮은 산업에 속해 있다면 부도 위험이 상대적으로 더 높아질 수 있다. 반대로, 산업집중도가 높고 시장지배력이 강한 기업은 일시적인 재무 악화 국면에서도 생존 가능성이 상대적으로 크다. 따라서 산업 구조와 시장 경쟁 구도를 부도 예측 모델에 반영하는 것은 중요한 의미를 가지며, 이는 단순히 재무제표 정보에 의존하는 기존 연구의 한계를 보완하고, 기업 외부 환경이 부도 위험에 미치는 체계적인 영향을 규명할 수 있다는 점에서 학문적·실무적 차별성을 확보하였다.

또한, 최근에는 기계학습 기반의 부도 예측모형이 활발히 활용되고 있지만, 대부분의 블랙박스형 모형(Random Forest, SVM, Gradient Boosting 등)은 예측력은 높더라도 변수별 영향력을 직관적으로 파악하기 어려워 정책 결정이나 실무 적용에서 해석 가능성이 부족하다는 한계가 지적된다. 특히 금융감독, 신용평가, 위험 관리와 같이 의사결정의 근거를

명확히 요구하는 분야에서는 예측 결과와 그 배경 요인을 함께 제시할 수 있는 모형이 필요하다(Lundberg and Lee, 2017; Nori et al., 2019).

본 연구는 이러한 한계를 보완하기 위해, 부도위험을 연속형 지표인 부도거리(Distance to Default, DD)와 이분형 지표인 부도 여부(Default_D)를 각각 종속변수로 설정하여, 장기적인 부도위험 수준과 단기 부도 발생 가능성을 병행 분석한다. 주요 설명변수로는 (1) 산업 내 경쟁집중도를 나타내는 허핀달-허쉬만 지수(Herfindahl-Hirschman Index, HHI), (2) 산업의 매출액 증가율, (3) 개별 기업의 산업 내 시장점유율, (4) 기업의 매출액 증가율(Firm Revenue Growth)을 포함하여, 기업의 재무상태뿐만 아니라 산업의 경쟁 구도, 성장단계, 시장 내 지위, 그리고 기업 자체의 성장성을 정량적으로 반영하였다.

방법론적으로는 전통적인 계량경제학 접근과 최신 설명 가능한 인공지능(Explainable AI) 기법을 결합하였다. 먼저 로지스틱 회귀(Logit), 고정효과 패널 회귀(Fixed Effects Panel Regression), 그리고 생존분석(Cox Proportional Hazards Regression)을 활용하여 변수 간 통계적 관계를 추정하였다. 이후, 설명 가능한 부스팅 머신(Explainable Boosting Machine, EBM)과 XGBoost를 구축하고, 각각 산업별 시장경쟁 구도(global) 및 기업별 재무적 특성(local) 수준에서 Explainable AI 분석을 수행하였다. 구체적으로, EBM은 변수의 산업별 시장경쟁 구도적 중요도와 기업별 재무적 특성적 민감도를 모두 파악할 수 있도록 설계하였으며, XGBoost에는 SHAP(Shapley Additive Explanations)를 결합하여 산업·시장 경쟁 구조 변수와 부도위험 간의 관계를 직관적으로 시각화하고, 각 변수의 기여도를 정량적으로 평가하였다.

분석에는 CSMAR 데이터베이스에서 수집한 2016년부터 2023년까지 중국 A주 상장기업 데이터를 사용하였다. 본 연구의 분석 결과, 부도거리(DD) 모형에서는 산업집중도가 부도위험에 양(+)의 영향을, 산업의 매출액 증가율, 시장점유율과 기업의 매출액 증가율이 음(-)의 영향을 미치며 모두 유의한 것으로 나타난다. 반면, 부도여부를 종속변수로 한 로지스틱 회귀와 콕스모형에서는 기업의 매출액 증가율만이 부도위험 완화와 관련된 유의한 음(-)의 효과를 보였다. 예측 성능 비교에서는 XGBoost가 가장 우수하였고, EBM이 뒤를 이었으며, 전통적 로지스틱 회귀는 상대적으로 낮았다. 해석 결과, DD가 모든 모형에서 핵심 변수임이 재확인되었고, 산업 수준 변수 또한 예측력 향상에 기여하였다. 특히 EBM의 상호작용 분석은 산업매출 증가율과 부도위험 간의 비선형 관계를 규명하였으며, XGBoost 역시 SHAP 해석을 통해 정책적으로 유의미한 인사이트를 제공할 수 있음을 확인하였다.

따라서 본 연구는 다음과 같은 결과를 나타낸다. 첫째, 기존의 재무정보 중심 부도 예측모형을 넘어 산업집중도, 산업의 매출액 증가율, 시장점유율 등 산업·시장 경쟁 구조

요인을 계량적으로 반영함으로써 부도 위험 분석의 범위를 확장하였다. 둘째, Explainable AI 기법을 활용하여 모형의 예측정확도와 해석 가능성을 동시에 제시함으로써 학계 연구와 정책·실무적 위험 관리에 모두 Explainable AI 분석 틀을 제시하였다.

본 논문의 구성은 다음과 같다. 제Ⅱ장은 선행연구를 검토하여 기업 부도 예측모형의 발전 과정과 기존 연구의 한계를 정리하고, 산업·시장 경쟁 구조 요인의 이론적 배경을 제시한다. 제Ⅲ장은 연구 가정과 분석 방법론을 설명하며, 변수 정의, 데이터 구성, 계량경제학 모형과 Explainable AI 기법의 적용 절차를 상세히 기술한다. 제Ⅳ장은 실증분석 결과를 제시하고, 전통 모형과 인공지능 모형의 예측력 비교, 변수별 기여도 분석 결과를 논의한다. 제Ⅴ장은 분석 결과의 학문적·정책적 시사점과 한계를 제시하고, 향후 연구 방향을 제안한다.

Ⅱ. 선행 연구

1. 기업 부도 예측 연구의 발전

기업 부도 예측 연구의 초기 고전으로는 Beaver(1966)와 Altman(1968)이 널리 알려져 있다. Beaver(1966)는 부도 기업과 건전 기업의 재무 특성을 비교하여 부도 가능성을 분석하였고, Altman(1968)은 다중판별분석(Multiple Discriminant Analysis, MDA)을 통해 두 집단 간 차이를 정량화한 Z-score 모형을 제시하였다. 이후 Ohlson(1980)은 로지스틱 회귀(Logit) 모형, Zmijewski(1984)는 프로빗(Probit) 모형을 도입하여 예측력을 높였으며, Odom and Sharda(1990)는 인공신경망(Artificial Neural Network, ANN)을 최초로 부도 예측에 적용하여 MDA보다 우수한 성능을 입증하였다. Baldwin and Gorecki(1990)와 Mata and Portugal(1994)은 생존분석(Survival Analysis)을 통해 부도의 지속성과 영향 요인을 규명하였다.

거시·산업적 환경을 반영한 분석도 등장하였다. Agarwal and Gort(1999)는 경기 변동, 산업 성장률, 시장 점유율이 부도 위험에 미치는 영향을 실증 분석하였고, Honjo(2000)는 실패 원인을 내부 요인(재무 건전성, 경영 역량)과 외부 요인(시장 경쟁 구조, 경쟁 압력)으로 구분하여 평가하였다.

21세기 이후에는 모형의 정밀화와 예측정확도 제고로 연구 초점이 이동하였다. Campbell et al.(2008)은 다항 로짓(Multinomial Logit) 모형으로 다기간 부도 확률을 산출했고, Aretz et al.(2018)은 이를 비미국 기업에 적용하여 부도 위험 프리미엄을 확인하였다. Tian et al.(2015)은 LASSO를 활용한 변수 선택으로 모형 성능을 개선하였으며, Traczynski(2017)는 베이지안(Bayesian) 접근이 단일 모형보다 더 높은 예측력을 제공함을 보고하였다. Kukuk

and Rönnberg(2013)는 혼합 로지트(Mixed Logit)로 부도 위험을 분석하였다.

최근에는 딥러닝과 머신러닝 기반 접근이 활발히 이루어지고 있다. Mai et al.(2019)은 재무 및 공시 텍스트 데이터를 활용하여 전통 모형과 비교한 결과, 딥러닝 모형의 우수성을 입증하였다. Alexandropoulos et al.(2019) 역시 로지스틱 회귀 대비 높은 성능을 보고하였다. Arratia and Sepúlveda(2019)는 재무 데이터를 이미지화하여 학습하는 합성곱 신경망(Convolutional Neural Network, CNN)을 제시하였고, Vochozka et al.(2020)은 LSTM(Long Short-Term Memory) 네트워크로 시계열 부도 예측 가능성을 검증하였다.

이와 같이, 기초적 통계 분석에서 최신 인공지능 기법에 이르기까지 다양한 방법론이 축적되며 부도 예측 연구는 정밀성과 신뢰성을 높여 왔다.

2. 구조적 모형: Merton 모형과 KMV 모형

Merton(1974)은 조건부 청구권 모형(contingent claims model)을 제시하여, 기업의 주식을 자산에 대한 콜옵션으로 해석하는 구조적 신용위험 모형의 기반을 마련하였다. 이후 Black and Cox(1976)는 재융자 제한(refinancing limitations)을, Geske(1977)는 복합 옵션(compound options) 구조를 도입하여 다중 채무 만기 문제를 설명하였다. Turnbull(1979)은 법인세와 부도 비용을 반영하여 현실성을 높였고, Marcus and Shaked(1984)는 부도 거리(Distance to Default, DD)를 활용해 FDIC 예금보험료의 공정성을 평가하였다.

KMV 모형은 Kealhofer, McQuown, Vasicek이 설립한 KMV사에서 개발되었으며, 2002년 Moody's에 의해 상업화되었다(Kealhofer and Kurbat, 2001). Duffie et al.(2007)은 Merton DD가 부도 기간 구조 예측에 유효하다고 보고하였고, Gropp et al.(2006)은 DD가 은행 등급 하락을 6~18개월 선행 예측할 수 있음을 실증하였다. 반면 Bharath and Shumway(2008)는 DD의 충분통계(sufficient statistic)로서의 한계를 지적했지만, 여전히 부도 예측에서 유용하다고 평가하였다. Harada et al.(2010)은 은행 부도 직전 DD 급등 현상을 발견해 조기경보 지표로서의 가능성을 제시했고, Coffinet et al.(2010), Castrén and Kavonius(2009)는 금융위기 시 DD 상승 시점이 상대적으로 늦게 나타나는 경향을 보고하였다. 최근 연구에서는 Jessen and Lando(2015)가 DD의 순위화(rank ordering) 성능과 변동성 조정 방안을 제안하였다.

3. 내재적 해석 모형과 블랙박스 모형

해석 가능성과 예측정확도 간의 균형은 기업 부도 예측모형에서 중요한 연구 과제이다.

내재적 해석 가능 모형은 변수와 예측 결과 간의 관계를 명확하게 파악할 수 있어, 특히 금융 규제 환경에서 유리하게 작용한다. Beaver(1966)와 Altman(1968) 이후, 로지스틱 회귀(Ohlson, 1980), 프로빗(Zmijewski, 1984), Cox 비례위험모형(Cox, 1972) 등은 계수 해석을 통해 예측변수의 주변 효과를 직관적으로 이해할 수 있는 대표적인 방법으로 활용되어 왔다. Hastie and Tibshirani(1986)는 비선형 관계를 가법 구조로 반영하면서도 해석 가능성을 유지하는 일반화 가법모형(Generalized Additive Models, GAMs)을 제안하였으며, Nori et al.(2019)은 이를 확장하여 얇은 트리에 기반한 그라디언트 부스팅(Gradient Boosting)을 적용한 설명 가능한 부스팅 머신(Explainable Boosting Machine, EBM)을 개발하였다.

한편, 딥러닝과 앙상블 기반의 블랙박스 모형은 높은 예측력을 제공지만 내재적 해석 가능성이 부족하여, 금융 분야에서의 활용에 제약이 따른다. Mai et al.(2019)은 재무지표와 공시 텍스트를 활용하여 딥러닝과 전통 알고리즘을 비교한 결과, 딥러닝의 예측력이 우수함을 입증하였다. Arratia and Sepúlveda(2019)는 CNN 기반의 이미지 변환 기법을, Vochozka et al.(2020)은 LSTM 기반 시계열 학습을 기업 부도 예측에 적용하였다. 이러한 블랙박스 모형을 보완하기 위해, 사후 해석 기법(post-hoc explanation methods)을 나타낸다. Goldstein et al.(2015)의 ICE(Individual Conditional Expectation) 플롯은 단일 관측치 기준으로 특정 변수를 변화시켰을 때 예측값의 민감도를 시각화함으로써, 변수의 비선형 효과나 임계값을 식별하는 데 유용한 도구로 활용된다. 또한, Lundberg and Lee(2017)의 SHAP(SHapley Additive Explanations)는 협동 게임이론을 기반으로, 각 변수의 개별 예측 기여도를 일관성 있게 산출할 수 있다.

Ⅲ. 연구가설 및 모형

1. 연구가설

기업 부도 위험은 재무적 요인뿐 아니라, 기업이 속한 산업의 구조적 특성과 시장 환경에 의해 결정된다. 산업조직 이론(Industrial Organization Theory)과 Porter(1980)의 구조-행위-성과(Structure-Conduct-Performance) 패러다임에 따르면, 산업 내 경쟁집중도, 산업 성장률, 개별 기업의 시장점유율은 가격 결정력, 진입장벽, 수익 안정성에 영향을 미쳐 부도 가능성을 변화시킬 수 있다. 중국 상장기업은 2016년 이전까지는 사실상 법적 부도이나 구조조정 사례가 존재하지 않았으나, 2016년 이후 지속적인 탈레버리지 정책 속에서 점차 부도 기업이 나타나기 시작하였다. 기존 연구들이 주로 미국, 일본 등 선진국 사례를 대상으로

한 반면, 본 연구는 2016-2024년 중국 상장기업 데이터를 활용하여 이러한 산업 · 시장 요인이 부도 위험에 미치는 영향을 분석한다. 특히 HHI(산업집중도), 산업 매출 증가율, 시장점유율, 기업별 매출 증가율 등 보다 정교한 지표를 도입하고, 설명가능한 AI(XAI) 기법을 적용함으로써 기존 연구와 차별성을 갖는다. 본 연구에서는 다음과 같이 네 가지 가설을 설정한다.

산업집중도는 시장 내 경쟁 강도를 나타내는 핵심 지표로, 높은 산업집중도는 일부 대형기업의 시장지배력을 강화하는 반면, 중소기업의 시장 접근성을 제한하여 산업 내 경쟁구조를 왜곡시킬 수 있다(Sutton, 1991). 이러한 구조는 대기업 중심의 과점시장에서 공급자 위험 집중을 유발하고, 특정 경기 · 정책 충격 시 전체 산업의 부도위험을 높일 수 있다. 일부 연구에서는 산업집중도가 높을수록 산업 전반의 취약성이 증가한다는 결과를 보고하였다(Borenstein and Rose, 1994; Choi and Weiss, 2005).

H1: 산업의 시장 집중도는 부도위험에 양(+)의 영향을 미친다.

산업별 매출액 증가율은 해당 산업의 경기 상황과 성장성을 반영한다(Porter, 1980). 경기 확장 국면의 산업에서는 기업이 외부 충격을 흡수할 수 있는 여력이 커지고, 전반적인 매출 확대가 기업 부도위험을 완화한다. 선행연구에 따르면 산업 성장률이 높을수록 부도율이 낮아진다는 결과가 보고되었다(Opler and Titman, 1994; Fama and French, 2004).

H2: 산업의 매출액 증가율은 부도위험에 음(-)의 영향을 미친다.

시장점유율은 기업의 시장지배력과 경쟁우위를 나타내는 대표적 지표이다(Scherer and Ross, 1990). 높은 시장점유율을 보유한 기업은 가격 경쟁에서 우위를 점하고 안정적인 매출을 확보할 가능성이 크다. 선행연구에서는 높은 시장점유율이 기업의 재무성과를 향상하고 파산위험을 낮춘다는 결과를 제시하였다(Bain, 1951; Lie, 2005).

H3: 시장점유율은 부도위험에 음(-)의 영향을 미친다.

기업 차원의 매출액 증가율은 영업 성과와 성장성을 나타내는 핵심 지표로, 매출이 증가하면 현금흐름과 수익성이 개선되어 부도위험이 감소한다(Altman, 1968). 여러 연구에서 기업의 매출 증가가 부도 가능성을 유의하게 낮춘다는 결과가 보고되었다(Shumway, 2001).

H4: 개별 기업의 매출액증가율은 부도위험에 음(-)의 영향을 미친다.

2. 모형 소개

1) KMV 모형

KMV 모형은 Merton(1974)의 구조적 신용위험 모형에 기반하며, 기업 자본을 만기 시점 부채 상환액을 행사가격으로 하는 유럽형 콜옵션으로 본다.

기업 자산가치 V_A 가 부채 상환액 D 보다 작으면 주주는 아무런 가치를 얻지 못하고 ($V_A < D$), 크면 부채를 상환한 후 잔여 가치를 보유한다($V_A > D$). 시점 T 에서의 자본 가치는 다음과 같다.

$$E_T = \max(V_A - D, 0) \quad (1)$$

Black-Scholes-Merton 모형에 따라 현재 시점 자본가치 E_0 는 다음과 같다.

$$E_0 = V_A N(d_1) - D e^{-r_f T} N(d_2) \quad (2)$$

$$d_1 = \frac{\ln(V_A/D) + (r + \frac{1}{2}\sigma_V^2)T}{\sigma_V \sqrt{T}}, \quad d_2 = d_1 - \sigma_V \sqrt{T}$$

여기서, $N(\cdot)$ 는 누적 정규 분포, 투자 자산의 기대 수익률(r)은 무위험 이자율(r_f)과 일치하고, σ_V 는 자산 변동성, σ_E 는 자본 변동성이다.

자산가치 V_A 와 변동성 σ_V 는 직접 관측 불가능하므로, 관측 가능한 E_0 와 σ_E 를 이용해 다음 두 식을 동시에 풀어 추정한다.

(1) 자본가치 방정식(식 (2))

(2) 변동성 관계식 (Ito의 Lemma 기반):

$$\sigma_E E_0 = \frac{\partial E}{\partial V} \sigma_V V_0 = N(d_1) \sigma_V V_0 \quad (3)$$

부도 거리(Distance to Default, DD)는 현재 자산가치가 부도 임계값(Default Point, DP)에서

떨어진 정도를 나타내며,

$$DD = \frac{\ln(V_A/DP) + (r - \frac{1}{2}\sigma_A^2)T}{\sigma_V\sqrt{T}} = d_2 \quad (4)$$

부도확률(Probability of Default, PD)은 위험 중립 가정하에 $PD = N(-DD)$ 로 계산된다.

2) 실증분석 모형

본 연구에서는 기업의 부도위험에 영향을 미치는 요인을 분석하기 위해 로지스틱 회귀모형, 고정효과 패널 회귀모형, Cox 생존분석 모형의 세 가지 실증모형을 사용하였다. 각 모형은 종속변수의 형태 및 연구 목적에 따라 선택되었으며, 아래와 같이 수식으로 표현된다.

(1) 고정효과 패널 회귀모형

부도거리(Distance to Default, $DD_{i,t}$)와 같은 연속형 종속변수에 대해서는, 다음과 같은 고정효과 패널 회귀모형을 적용하였다:

$$DD_{i,t} = \beta_0 + \beta_1 X_{i,t-1} + \gamma_i + \delta_t + \varepsilon_{i,t} \quad (6)$$

여기서

$X_{i,t-1}$ 는 기업 i 의 시점 $t-1$ 에서의 설명변수,

γ_i 는 기업 고정효과,

δ_t 는 연도 고정효과,

$\varepsilon_{i,t}$ 는 오차항이다.

이 모형은 기업의 관측 불가능한 특성을 통제함으로써, 설명변수가 부도 위험에 미치는 순효과를 추정할 수 있다.

(2) 로지스틱 회귀모형(Logistic Regression)

부도 여부를 나타내는 이진 종속변수 Default_D (부도더미)에 대해, 로지스틱 회귀모형은 다음과 같이 설정된다.

$$\Pr(Default_{i,t} = 1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 X_{i,t-1} + \gamma_i + \delta_t))} \quad (5)$$

여기서,

$X_{i,t-1}$ 는 기업 i 의 시점 $t-1$ 에서의 설명변수,

γ_i 는 기업 고정효과,

δ_t 는 연도 고정효과이다.

해당 모형은 기업의 부도 확률에 영향을 미치는 재무 및 산업 구조적 요인을 식별하는데 적합하다.

(3) Cox 비례위험 모형 (Cox Proportional Hazards Model)

부도 사건의 발생 시점(시간)을 고려한 분석을 위해 Cox 생존분석 모형을 사용하며, 기본 수식은 다음과 같다.

$$h_i(t) = h_0(t) \exp(\beta_i X_{i,t-1}) \quad (7)$$

여기서,

$h_i(t)$ 는 기업 i 의 시점에서의 위험률,

$h_0(t)$ 는 기준 위험함수,

$X_{i,t-1}$ 는 설명변수이다.

Cox 모형과 같이 다양한 회귀모형을 병행하여 분석함으로써, 부도위험 결정요인에 대한 보다 견고하고 다각적인 실증분석이 가능하였다.

3) 기계학습 예측모형

(1) 내재적 Explainable AI 모형 (EBM)

모형의 투명성을 확보하고 예측 결과에 대한 설명하기 위해, 본 연구에서는 내재적으로 해석 가능한(Intrinsically Interpretable) 모형을 포함하였다. Hastie and Tibshirani(1986)가 제안한 일반화 가법 모형(Generalized Additive Models, GAMs)은 비선형 관계를 유연하게 가법(additive) 구조로 하면서도 해석 가능성을 유지할 수 있게 한다. 각 변수는 독립적인

평활 함수(smooth function)를 통해 결과에 기여하므로, 변수의 효과를 시각적으로 쉽게 확인할 수 있다.

이를 바탕으로, Nori et al.(2019)은 GAM의 가법 구조를 유지하면서 얇은 의사결정나무(shallow tree)에 대한 그래디언트 부스팅(Gradient Boosting)을 적용하여 Explainable Boosting Machine(EBM)을 제안하였다. EBM에서는 목표 변수의 로그 오즈(log-odds)를 개별 특성과 선택된 쌍별 상호작용의 가법 함수로 한다.

$$\text{logit}(P(y=1|x)) = \beta_0 + \sum f_i(x_i) + \sum f_{i,j}(x_i, x_j) \quad (5)$$

여기서 $f_i(x_i)$ 는 단일 변수의 기여를 나타내는 학습된 비선형 형상 함수(shape function)이며, $f_{i,j}(x_i, x_j)$ 는 중요한 변수 쌍 간의 상호작용 효과를 나타낸다. 모형 출력은 로지스틱(시그모이드) 함수를 거쳐 확률로 변환한다.

EBM의 핵심 강점은 산업별 시장경쟁 구도(global) 및 기업 특성(local) 수준 모두에서 해석 가능성을 제공한다는 점이다. 산업별 시장경쟁 구도적으로는 학습된 형상 함수를 시각화하여 전체 데이터에서 각 변수가 모형 예측에 미치는 영향을 보여줄 수 있다. 이를 통해 주요 변수, 비선형 효과, 임계값 등을 식별할 수 있다. 기업 특성 관점으로는 개별 예측값을 각 변수의 가법적 기여도를 기준으로 분해함으로써 인스턴스 단위의 설명이 가능하다.

(2) 블랙박스 모형 (XGBoost)

그래디언트 부스팅 머신(Gradient Boosting Machines, GBM) 기법은 순차적으로 의사결정나무를 생성하는 가법 모형(Additive Model)으로, 각 새로운 나무는 이전 앙상블의 잔차 오차를 보정하도록 학습된다. Chen and Guestrin(2016)이 제안한 XGBoost는 정규화기업별 재무적 특성효율적인 트리 생성 기법을 도입하여 성능과 확장성을 한다.

그래디언트 부스팅(Gradient Boosting)의 예측 함수는 다음과 같다.

$$\hat{y} = \sum_{t=1}^T f_t(x), \quad f_t \in F \quad (6)$$

여기서 F 는 회귀나무(regression tree)의 공간이고, T 는 트리의 개수이며, 각 $f_t(x)$ 는 입력 x 를 점수로 매핑한다.

목표는 다음과 같은 정규화된 손실 함수를 최소화하는 것이다.

$$L(\hat{y}, y) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{t=1}^T ohm(f_t) \tag{7}$$

여기서 l 은 미분 가능한 손실 함수이고, $ohm(f)$ 트리 복잡도에 대한 패널티를 부여하는 정규화 항이다.

(3) 사후 해석 기법(Post-hoc Explanation Methods)

Lundberg and Lee(2017)가 제안한 Shapley Additive Explanations(SHAP)는 XGBoost, 랜덤 포레스트, 서포트 벡터 머신, 심층 신경망 등 복잡한 블랙박스 머신러닝 모형의 출력을 해석하기 위한 통합 분석 모형(framework)이다. 협동 게임이론(cooperative game theory)에 기반하여, SHAP은 각 변수의 공정한 기여도를 샤플리 값(Shapley value)으로 계산하여 개별 예측에 대한 특성별 기여를 정량화한다.

훈련된 모형 $f(x)$ 와 입력 인스턴스 $x = (x_1, x_2, \dots, x_p)$ 가 주어졌을 때, SHAP은 모형 출력을 다음과 같은 가법 특성 기여 모형으로 한다.

$$f(x) = \phi_0 + \sum_{i=1}^p \phi_i \tag{8}$$

여기서 ϕ_0 는 데이터셋 전체의 평균 예측값(base value), ϕ_i 는 변수 x_i 의 주변(marginal) 기여도를 나타낸다.

샤플리 값은 다음과 같이 정의된다.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(p - |S| - 1)!}{p!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \tag{9}$$

여기서 N 은 모든 변수의 집합, S 는 변수 i 를 포함하지 않는 부분집합, $f_S(x_S)$ 는 변수 집합 S 로 학습된 모형이다. 가중치 항은 가능한 모든 변수 순서에 대해 평균을 공정하게 반영한다.

SHAP은 해석 가능성과 이론적 타당성을 높이는 여러 속성을 만족한다. 첫째, 기업 특성변수

기준의 정확성(local accuracy)을 보장하여, 모든 변수의 SHAP 값을 합하면 모형의 예측값과 정확히 일치한다. 둘째, 일관성(consistency)을 보장하여, 모형이 특정 변수의 기여도를 높이는 방향으로 변화하면 해당 SHAP 값이 감소하지 않는다. 셋째, 결측성(missingness) 원리를 준수하여, 모형에 포함되지 않은 변수에는 0의 값을 부여함으로써 불필요한 기여할당을 방지한다.

금융 분야에서는 SHAP이 신용평가, 투자 위험, 사기 탐지, 부도 예측모형의 해석에 널리 활용되고 있다. 특히 규제와 컴플라이언스가 중요한 환경에서 산업별 시장경쟁 구도 및 기업 특성변수 기준의 수준의 설명을 모두 제공할 수 있다는 점에서 가치가 크다.

3. 변수정의

<표 1>은 본 연구에서 사용된 주요 변수의 정의를 정리한 것이다. 본 연구의 종속변수는 두 가지로 구성된다. 첫째는 부도 발생 여부를 나타내는 이진 변수인 Default_D로, 기업이 해당 연도에 실제로 부도 상태에 진입한 경우에는 1, 그렇지 않은 경우에는 0으로 설정하였다. 둘째는 구조적 신용 위험 측정치인 DD(Distance to Default)로, 이는 KMV 모형에 기반하여 기업 자산가치가 부도경계(default point)에 도달하기까지의 표준화된 거리를 나타내며, 기업의 잠재적 부도 위험 수준을 연속형 변수로 측정한다.

주요 설명변수에는 산업 내 경쟁 정도를 나타내는 허핀달-허쉬만 지수(HHI), 산업 및 기업의 매출 증가율(MktRev_Growth, FirmRev_Growth), 기업의 시장점유율(MktShare) 등이 포함된다.

또한, Altman Z-score를 구성하는 재무 비율 변수들(NWCR, RER, EBITR, MVTLR, SaleR)은 각각 운전자본 비율, 유보이익 비율, 영업이익 비율, 시가총액 대비 부채비율, 매출자산비율로 정의되며, 기업의 유동성, 수익성, 시장평가 등을 반영한다. 기업 특성 변수로는 정부 소속 여부(SOE_D), 특허 수(Ln_Patent), 연구개발비 비율(RnD_Ratio), 최대주주 지분율(LJholder_Rate)이 포함되며, 이는 지배구조 및 혁신역량을 통제하기 위한 변수들이다. 마지막으로, 자산 규모(Ln_Asset), 은행 예금 대비 매출 비율(LDR), 지방정부의 조세수입 비중(TaxRev_Ratio) 등의 외부 환경 변수도 함께 고려하여, 기업의 부도 위험에 영향을 미치는 다양한 요인을 종합적으로 분석하고자 하였다. 본 연구에서 특허 수와 자산 규모 변수는 기업 간 규모 격차가 크고 분포의 왜도가 높아 원자료를 그대로 사용할 경우 극단치의 영향이 과도하게 반영될 수 있다. 이에 로그 변환을 적용하여 분포를 안정화하고 추정 계수의 경제적 해석(비율 변화 및 탄력성)을 용이하게 조치하였다.

<표 1> 변수 정의

변수명	변수 정의
Default_D(부도더미)	해당 연도에 기업이 부도이면 1, 아니면 0
DD(부도거리)	자산가치가 부도장벽까지 도달하는 표준화 거리(KMV 모형 기반)
HHI(산업집중도)	산업 내 각 기업의 시장점유율 제곱을 합한 값
MktRev_G(산업의 매출액 증가율)	전년 대비 산업 전체 매출의 증가율
MktShare(시장점유율)	기업이 산업에서 차지하는 매출 비중; 기업매출 / 산업 총매출
FirmRev_G(기업의 매출액 증가율)	전년 대비 기업 매출의 증가율
NWCR(운전자본 비율)	운전자본이 총자산에서 차지하는 비율; 운전자본 / 총자산 (Altman Z-Score의 X_1)
RER(유보이익 비율)	유보이익이 총자산에서 차지하는 비율; 유보이익 / 총자산 (Altman Z-Score의 X_2)
EBITR(영업이익 비율)	영업이익이 총자산에서 차지하는 비율; 영업이익 / 총자산 (Altman Z-Score의 X_3)
MBR(시가총액 대비 부채비율)	주식 시가총액이 총부채에서 차지하는 비율; 시가총액 / 총부채 (Altman Z-Score의 X_4)
SaleR(매출자산비율)	매출액이 총자산에서 차지하는 비율; 매출액 / 총자산 (Altman Z-Score의 X_5)
SOE_D(정부지배기업더미)	정부 지분 보유 또는 통제 기업이면 1, 아니면 0
Ln_Patent(특허로그)	특허 수에 1을 더한 값의 자연로그
RnD_R(연구개발강도)	연구개발 지출이 총자산에서 차지하는 비율; 연구개발지출 / 총자산
LHolder_R(최대주주지분율)	최대주주가 보유한 주식의 비율; 최대주주 보유주식 / 총발행주식
Ln_Asset(자산로그)	총자산의 자연로그
LDR(대출예금비율)	은행 대출액이 예금액에서 차지하는 비율; 대출액 / 예금액
TaxRev_R(조세수입비중)	정부 조세수입이 총재정수입에서 차지하는 비율; 조세수입 / 총재정수입

IV. 실증분석

1. 표본 및 데이터

<표 2>는 2016년부터 2023년까지 중국 상장기업을 대상으로 한 연도별 부도 기업과 생존기업의 현황을 제시한다. 재무정보는 CSMAR 데이터베이스에서 수집하였으며, 금융업종에 속하는 기업과 재무정보에 이상치(극단값)가 존재하는 기업은 분석에서

제외하였다. 따라서 표에 제시된 수치는 이러한 기업들을 제외한 나머지 표본을 기준으로 산출되었다.

본 연구의 원자료는 31,957개 관측치로 구성되어 있다. 일부 변수에서 음수 자산, 음수 특허 등 극단적 이상치가 존재하여 그대로 활용할 경우 추정결과가 왜곡될 수 있다. 따라서 비 부도 기업에서 나타나는 비현실적인 값(예: 음수 자산, 음수 특허 등)만 제외하였으며, 산업구조의 이질성을 고려하여 금융업종도 제외하였다. 최종적으로 30,744개의 유효 표본을 확보하여 실증 분석을 수행하였다.

<표 2> 연도별 부도 기업과 생존기업 현황

연도	전체 기업 수 (A)	부도 기업 수 (B)	생존기업 수 (C)	비율(F = B/A, %)
2016	2,900	1	2,899	0.0345
2017	3,306	5	3,301	0.1512
2018	3,366	16	3,350	0.4753
2019	3,528	19	3,509	0.5385
2020	3,892	25	3,867	0.6423
2021	4,313	22	4,291	0.5101
2022	4,617	25	4,592	0.5415
2023	4,822	26	4,796	0.5392
전체	30,744	139	30,605	0.4521

표의 A 열은 각 연도의 전체 기업 수를, B 열은 부도 상태에 진입한 기업 수를, C 열은 해당 연도에 부도 없이 생존한 기업 수를 나타낸다. 마지막 열(F)은 전체 기업 대비 부도 기업의 비율을 의미하며, 이를 통해 연도별 부도율 변화를 확인할 수 있다.

2. 기초 통계량 분석

<표 3>은 본 연구에서 사용된 종속변수와 주요 설명변수의 기초 통계량을 제시한다. 종속변수는 부도 발생 여부(Default_D)와 부도거리(DD)이며, 각각 기업의 해당 연도 부도 상태 여부와 KMV 모형 기반의 신용위험 수준을 나타낸다. 주요 설명변수로는 산업집중도(HHI), 산업매출 증가율(MktRev_G), 시장점유율(MktShare), 기업매출 증가율(FirmRev_G)이 포함되며, 이는 산업 구조와 시장 내 기업의 경쟁력 및 성장성을 측정하기 위한 변수들이다. 평균값을 살펴보면, HHI는 0.082로 전반적으로 낮은 산업 집중도를 나타내며, MktRev_G는

0.211로 완만한 산업 성장세를 보인다. MktShare의 평균은 0.017로, 개별 기업의 산업 내 매출 비중이 상대적으로 낮음을 시사하며, FirmRev_G는 평균 0.180으로 기업 매출이 전반적으로 완만하게 증가하고 있음을 보여준다. 나머지 재무비율 및 기업 특성 등 통제변수에 대한 통계치는 생략하였다.

<표 3> 기초 통계량

변수	관측치 수	평균	중위수	표준편차	최소값	최대값
Default_D	30,744	0.005	0.000	0.067	0.000	1.000
DD	30,744	1.912	1.895	0.332	0.000	2.619
HHI	30,744	0.082	0.079	0.081	0.011	0.497
MktRev_G	30,744	0.211	0.003	0.525	-1.425	2.115
MktShare	30,361	0.017	0.026	0.049	0.000	0.673
FirmRev_G	30,744	0.180	5.651	1.003	-3.420	5.392
NWCR	30,744	0.253	0.253	0.256	-1.739	0.815
RER	30,744	0.116	0.163	0.300	-7.584	0.535
EBITR	30,744	0.042	0.048	0.084	-1.942	0.234
MBR	30,744	1.758	1.619	1.003	0.077	4.515
SaleR	30,744	0.578	0.496	0.386	0.000	2.370
SOE_D	30,744	0.287	0.000	0.452	0.000	1.000
Ln_Patent	30,744	1.428	0.000	2.347	0.000	7.406
RnD_R	30,744	0.022	0.017	0.026	0.000	0.134
LHolder_R	30,744	0.327	0.303	0.145	0.052	0.734
Ln_Asset	30,744	22.269	22.069	1.334	16.412	28.697
LDR	30,744	0.686	0.714	0.101	0.318	0.863
TaxRev_R	30,744	0.766	0.765	0.076	0.551	0.883

3. 상관관계 분석

<표 4>는 본 연구에서 사용된 주요 변수들의 상관관계를 제시한다. 주요 변수 간의 상관관계를 살펴보면, 부도 발생 여부(Default_D)와 부도거리(DD)는 음의 상관관계(-0.112)를 보여 부도 위험이 높을수록 DD가 낮아지는 경향이 나타난다. 대부분의 변수 간 상관계수는 유의미하게 나타낸다. 따라서, 다중공선성 문제를 피하기 위해 VIF(분산팽창요인) 검증을 하였다.

<표 4> 변수간 상관관계

여기서, (1)은 Default_D, (2)는 DD, (3)은 HHI, (4)는 MktRev_G를 나타내며, 이하 동일한 방식으로 표기한다. * ** *** 은 각각 1%, 5%, 10% 수준에서 통계적으로 유의함을 나타낸다.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
DD	-0.112***	1.000															
HHI	0.007	0.009	1.000														
MktRev_G	-0.005	0.008	0.053***	1.000													
MktShare	-0.004	0.041***	0.341***	-0.027***	1.000												
FirmRev_G	-0.042***	-0.013**	0.002	0.313***	0.125***	1.000											
NWCR	-0.160***	0.276***	-0.192***	-0.042**	-0.139***	-0.018***	1.000										
RER	-0.211***	0.246***	-0.066***	-0.015***	0.033***	0.044***	0.337***	1.000									
EBITR	-0.267***	0.155***	-0.025***	-0.010*	0.048**	0.100***	0.287***	0.559***	1.000								
MBR	-0.078***	0.221***	-0.156***	-0.058***	-0.162***	-0.043***	0.660***	0.150***	0.243***	1.000							
SaleR	-0.031***	-0.112***	0.050***	0.021***	0.126***	0.074***	-0.053***	0.028***	0.167***	-0.127***	1.000						
SOE_D	-0.021***	0.068***	0.147***	0.023***	0.128***	0.044***	-0.233***	-0.033***	-0.013**	-0.322***	0.036***	1.000					
Ln_Patent	-0.017***	-0.011*	-0.158***	0.001	-0.066***	-0.010*	0.161***	0.079***	0.036***	0.123***	-0.046***	-0.147***	1.000				
Rat_R	-0.032***	-0.032***	-0.296***	0.082***	-0.103***	0.050***	0.159***	0.095***	0.049***	0.163***	0.090***	-0.159***	0.185***	1.000			
LHolder_R	-0.038***	0.082***	0.100***	-0.021***	0.102***	0.012**	0.018***	0.201***	0.177***	-0.048***	0.070***	0.232***	-0.016***	-0.121***	1.000		
Ln_Asset	0.002	0.025***	0.148***	0.050***	0.323***	0.099***	-0.367***	0.165***	0.093***	-0.601***	0.063***	0.367***	-0.068***	-0.086***	0.130***	1.000	
LDR	0.006	0.039***	-0.023***	-0.076***	0.000	-0.022***	0.088***	0.023***	-0.043***	-0.001	-0.012**	-0.028***	0.082***	0.198***	-0.010**	0.035***	1.000
TaxRev_R	-0.026***	0.002	-0.020***	-0.127***	0.010*	-0.063***	0.123***	0.073***	0.028***	0.081***	0.001	-0.116***	0.028***	0.078***	0.020***	-0.033***	0.400***

4. 집단 비교분석

<표 5>는 부도기업과 생존기업 간의 평균값 차이를 t-검정을 통해 비교한 결과를 제시한다. 부도기업 집단과 생존기업 집단의 분산이 동일하지 않다고 가정하였다. 분석 결과, 부도기업은 생존기업에 비해 부도거리(DD)가 유의하게 낮고, 운전자본 비율(NWCR), 유보이익 비율(RER), 영업이익 비율(EBITR), 매출자산비율(SaleR) 등 주요 재무비율에서 전반적으로 낮은 값을 보였다. 또한 최대주주지분율(LHolder_R)과 기업의 매출액 증가율(FirmRev_G) 역시 부도기업에서 유의하게 낮게 나타났다. 반면, 산업집중도(HHI)와 총자산(Ln_Asset)은 두 집단 간 유의한 차이가 없었다. 이러한 결과는 부도기업이 전반적으로 재무건전성과 성장성이 취약함을 시사한다.

<표 5> 부도기업과 생존기업의 평균 비교 검정

부도기업 그룹과 생존기업 그룹 간 분산이 같지 않다고 가정하여 그룹 평균을 비교 검증하였다. *, **, ***은 각각 1%, 5%, 10% 수준에서 통계적으로 유의함을 나타낸다.

변수	부도기업 (A)	생존기업 (B)	차이 (A-B)	t-통계량
DD	0.866	1.916	-1.051***	38.09
HHI	0.090	0.082	0.008	1.16
MktRev_G	0.176	0.211	-0.036	-0.80
MktShare	0.014	0.017	-0.003	-0.77
FirmRev_G	-0.439	0.182	-0.621***	-7.29
NWCR	-0.356	0.255	-0.612***	-28.47
RER	-0.826	0.121	-0.947***	-37.92
EBITR	-0.291	0.043	-0.334***	-48.60
MBR	0.593	1.764	-1.170***	-13.78
SaleR	0.398	0.579	-0.181***	-5.50
SOE_D	0.144	0.288	-0.144***	-3.74
Ln_Patent	0.830	1.430	-0.601***	-3.01
RnD_R	0.010	0.022	-0.012***	-5.63
LHolder_R	0.246	0.328	-0.081***	-6.60
Ln_Asset	22.300	22.269	0.031	0.27
LDR	0.695	0.686	0.009	1.05
TaxRev_R	0.737	0.766	-0.029***	4.53

5. 회귀분석

<표 6>은 종속변수(Default_D 또는 DD)를 대상으로 한 세 가지 회귀모형의 추정결과를 제시한다. 첫째, (1) 열은 Breusch and Pagan의 Lagrange 승수검정(LM 검정)을 통한

시간효과 검정과 Hausman 검정을 통한 기업효과 검정 결과, 가장 적합하다고 판정된 고정효과 패널모형(Fixed Effects Panel Model)의 계수와 t-통계량을 보고한다.

둘째, (2) 열은 로지스틱 회귀모형(Logistic Regression)을 이용하여 종속변수가 부도 여부(이항변수)일 때의 계수와 z-통계량을 제시한다.

셋째, (3) 열은 콕스 비례위험모형(Cox Proportional Hazards Model)을 이용하여 종속변수가 부도 여부(이항변수)일 때의 위험비율(Hazard ratio)과 z-통계량을 나타낸다. 여기서 위험비율은 특정 변수의 변화가 사건 발생 위험에 미치는 상대적 크기를 의미하며, 1보다 크면 위험이 증가하고 1보다 작으면 위험이 감소함을 나타낸다.

주요 설명변수 중 산업집중도(HHI)는 고정효과 패널모형에서 1% 유의수준 하에서 음(-)의

<표 6> 고정효과 패널모형, 로지스틱 회귀, 콕스 비례위험모형 추정결과

모형 (1) 시간 효과에 대한 Breusch and Pagan의 Lagrange 승수 검정(LM 검정)과 기업 효과에 대한 Hausman 검정 등 모형 선택 검정을 거쳐, 가장 적합한 모형으로 선정된 고정효과모형(FEM)의 결과를 보고하였다. ***, **, * 표시는 각각 1%, 5%, 10% 유의수준에서의 단측검정 통계적 유의성을 의미한다.

종속변수: DD/ Default_D	(1) 고정효과 패널모형		(2) 로지스틱 회귀모형		(3) 콕스 비례위험모형	
	계수	t-통계량	계수	z-통계량	위험비율 (Hazard ratio)	z-통계량
HHI	-2.497***	-9.57	-0.489	-0.22	0.582	-0.29
MktRev_G	0.705***	29.45	-0.026	-0.13	0.977	-0.13
MktShare	0.249	0.57	3.220	0.89	26.558	1.06
FirmRev_G	0.076***	6.38	-0.235***	-2.83	0.807***	-2.85
DD			-0.392***	-3.98	0.691***	-4.05
NWCR	3.279***	24.32	-0.833	-1.24	0.492	-1.13
RER	1.486***	15.39	-1.513***	-4.90	0.267***	-4.88
EBITR	-0.859***	-4.09	-1.515*	-1.36	0.303	-1.19
MBR	0.150***	3.72	-2.113***	-7.53	0.146***	-7.61
SaleR	-0.792***	-9.91	-0.956***	-2.75	0.425***	-2.62
SOE_D	0.273***	2.96	-1.587***	-5.14	0.231***	-5.00
Ln_Patent	0.017**	1.97	0.010	0.20	1.009	0.18
RnD_R	-11.719***	-8.88	-13.688**	-1.77	0.000**	-1.76
LHolder_R	-0.462*	-1.58	-1.637**	-1.83	0.197**	-1.94
Ln_Asset	-0.197***	-4.19	0.018	0.15	1.023	0.21
LDR	-1.436***	-8.59	0.708	0.57	1.786	0.52
TaxRev_R	-0.518	-1.27	-3.826***	-2.36	0.027***	-2.45
상수항	12.086***	6.71	2.686	0.96		
산업 고정효과		YES		YES		YES
기업 고정효과		YES		YES		YES
연도 고정효과		YES		YES		YES
관측치(기업 수)	25,190(4,825)		21,123(4,110)		25,193(4,827)	
VIF 검정	1.08~2.59		1.07~2.78		1.07~2.78	

영향을 미쳤다. 이는 시장집중도가 높을수록 기업의 부도 거리(DD)가 작아져 부도 가능성이 증가함을 의미한다. 반면, 로지스틱 회귀모형과 콕스 비례위험모형에서는 통계적으로 유의하지 않았다. 또한 산업 매출액 증가율(MktRev_G), 기업의 시장점유율(MktShare), 기업의 매출액 증가율(FirmRev_G)은 고정효과 패널모형에서 모두 양(+)의 영향을 보였다. 이는 산업의 전체 시장 매출액이 증가하는 경우 기업의 부도 거리가 확대되어 부도 위험이 감소함을 시사한다. 그러나 산업의 매출액 증가율(MktRev_G)과 시장점유율(MktShare)은 로지스틱 회귀모형과 콕스 비례위험모형에서 유의하지 않았다. 반면, 기업의 매출액 증가율(FirmRev_G)은 세 모형 모두에서 1% 유의수준 하에서 유의하게 부도 위험을 감소시키는 것으로 나타나, 기업 매출의 증가가 부도 가능성을 유의하게 낮추는 중요한 요인임을 확인할 수 있다. 통제변수에 대한 상세한 설명은 이하에서 생략한다.

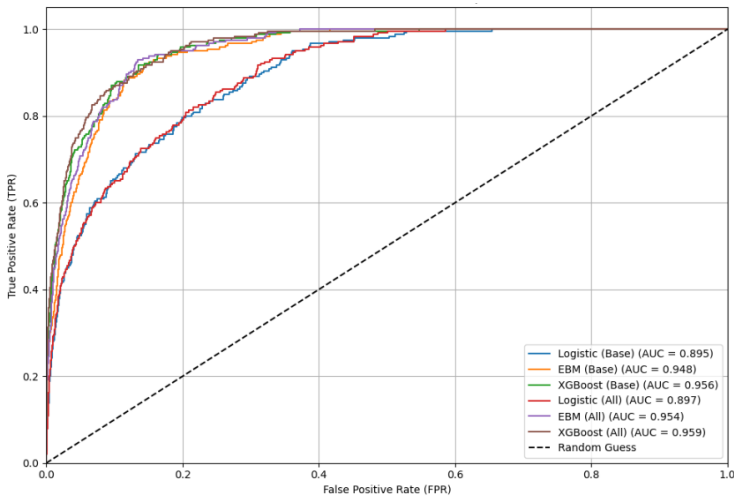
6. 기계학습 모형의 예측 결과

본 절에서는 세 가지 기계학습 예측 모형(Logistic 회귀, 설명 가능한 부스팅 모형(EBM), XGBoost)을 활용하여 기업 부도 여부에 대한 분류 성능을 비교하였다. 특히, 모형별로 기초 재무 변수(base variables)만을 사용한 경우와, 산업·시장 경쟁구조 관련 변수(market variables)를 추가하여 시장경쟁 변수(all variables)를 도입한 활용한 경우로 나누어 모형의 예측력이 어떻게 변화하는지를 분석하였다.

사용된 기본 모형(base variables)에는 부도거리(DD), Altman Z-Score 항목들(X1~X5), 국유기업 여부(SOE_D), 특허 보유(Log(특허수)), 연구개발 비율(RnD_R), 최대주주 보유비율(LHolder_R), 자산규모(Ln_Asset), 부채비율(LDR), 세수 비율(TaxRev_R) 등이 포함된다. 한편, 시장 관련 변수(market variables)로는 산업 집중도(HHI), 산업 매출 성장률(MktRev_G), 시장점유율(MktShare), 기업 매출 증가율(FirmRev_G) 등이 사용되었다.

[그림 1]은 로지스틱 회귀(Logistic Regression), 설명 가능한 부스팅 모형(Explainable Boosting Machine, EBM), 그리고 XGBoost 세 가지 예측 모형을 대상으로, 기초 재무 변수(Base variables)만을 사용한 경우와 시장 변수(Market variables)를 추가하여 시장경쟁 모형(All variables)를 사용한 경우 각각에 대한 ROC 곡선을 제시한다. 각 곡선의 AUC 값은 범례에 병기되어 있으며, 이는 모형의 판별 성능을 정량적으로 비교하는 지표로 활용된다. 그림에서 확인할 수 있듯이, 모든 모형에서 변수 확장(All variables) 시 AUC 값이 소폭 향상되는 경향을 보이며, 특히 XGBoost 모형은 두 변수 조합 모두에서 가장 높은 AUC (기존: 0.956 및 수정: 0.959)를 기록하였으며, EBM(기존: 0.948 및 수정: 0.954) 보다는 상당히

개선된 결과라 할 수 있다. 특히 전통적인 재무적인 모형(기준: 0.895 및 수정: 0.897) 보다는 획기적인 개선을 이룸으로써, 기본적으로 방법론의 개선과 함께, 연구의 핵심대상이 되는 기업의 근본적인 경쟁력이 모형에 반영됨으로써 고위험 · 기술지향적 · 혁신 중심 산업이나 신생기업 등에서는 설명력을 개선할 수 있다고 판단한다.



[그림 1] 모형 및 변수 조합별 ROC 곡선 비교

<표 7> 성능 비교: 로지스틱 회귀 vs EBM vs XGBoost

모형	정확도	정밀도	재현율	F1 점수
Logistic(기본 모형)	0.975	0.600	0.087	0.153
EBM(기본 모형)	0.978	0.781	0.208	0.329
XGBoost(기본 모형)	0.979	0.818	0.263	0.397
Logistic(시장경쟁 모형)	0.975	0.579	0.092	0.158
EBM(시장경쟁 모형)	0.978	0.773	0.212	0.333
XGBoost(시장경쟁 모형)	0.979	0.808	0.263	0.396

<표 7>은 로지스틱 회귀(Logistic Regression), 설명 가능한 부스팅 모형(Explainable Boosting Machine, EBM), 그리고 XGBoost 세 가지 분류 모형의 성능을 기본 모형(Base variables)만을 사용한 경우와 시장 변수(Market variables)를 포함한 시장경쟁 모형(All variables)를 사용한 경우로 구분하여 비교한 결과를 제시한다. 성능 평가는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), 그리고 F1 점수(F1 Score) 네 가지 지표를 기준으로 수행되었다.

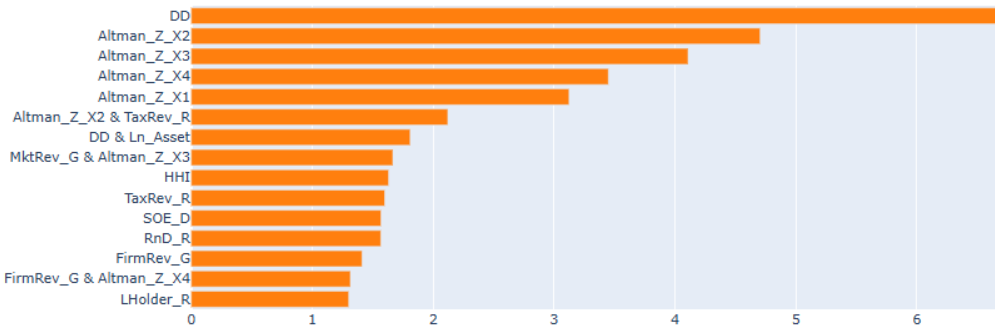
분석 결과, 시장경쟁 모형을 사용한 경우가 기본 모형만을 사용한 경우에 비해 전반적으로 성능이 다소 향상되었으나, 그 폭은 제한적이었다. 특히 모든 모형에서 재현율(Recall) 값이 매우 낮게 나타났는데, 이는 표본의 불균형(class imbalance)으로 인한 결과로 해석된다. 한편, XGBoost 모형은 두 변수 조합 모두에서 가장 우수한 성능을 보였으며, 이는 블랙박스(black-box) 모형 즉, XGBoost 기법이 예측 성능 측면에서 가지는 우월성을 입증하는 결과라 할 수 있다. 이러한 분석은 변수 선택 및 표본 균형 조정이 부도 예측 모형의 성능 향상에 중요한 역할을 할 수 있음을 시사한다.

7. 모형 결과에 대한 해석적 분석

본 절에서는 부도 예측 모형의 예측 결과를 해석하기 위해 두 가지 방법을 활용하였다. 첫째, 내재적으로 Explainable AI 모형인 설명 가능한 부스팅 모형(Explainable Boosting Machine, EBM)을 사용하여 산업별 시장경쟁 구도(Global) 및 기업 특성(Local) 수준에서 변수 중요도와 영향 방향을 분석하였다. 둘째, 예측 성능이 우수한 블랙박스(black-box) 계열 모형인 XGBoost에 대하여 SHAP(Shapley Additive Explanations) 값을 산출함으로써, 개별 예측에 기여한 변수와 그 기여 정도를 정량적으로 파악하였다.

1) EMB 해석적 분석

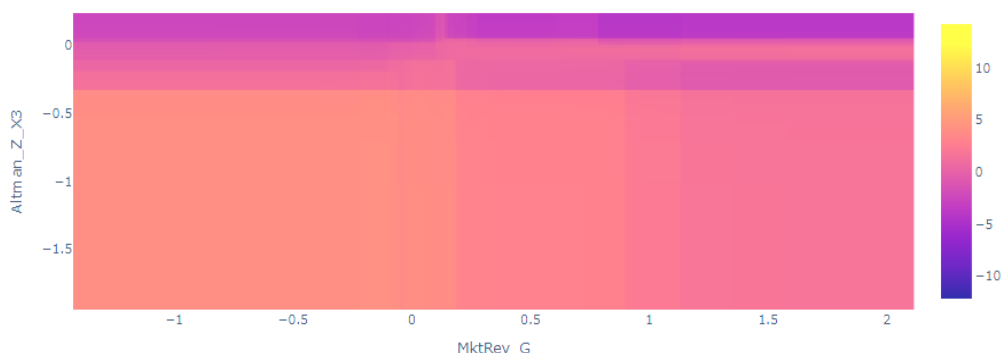
[그림 2]는 설명 가능한 부스팅 모형(Explainable Boosting Machine, EBM)을 활용하여 산출된 산업별 시장경쟁 구도 변수 중요도 결과를 나타낸 것이다. 변수 중요도는 각 특성이 전체 예측 결과에 기여하는 평균 절대값(Mean Absolute Score, Weighted)을 기준으로 측정되었으며, 값이 클수록 해당 변수가 부도 예측에 미치는 영향력이 크다는 것을 의미한다.



[그림 2] EBM 기반 산업별 시장경쟁 구도 변수 중요도 분석 결과

분석 결과, 부도거리(Distance to Default, DD)가 가장 높은 중요도를 보였으며, Altman Z-Score 구성 변수들(특히 Altman_Z_X2, X3, X4) 또한 상위권에 위치하였다. 이는 전통적인 재무 건전성 지표들이 여전히 부도 예측에 있어 핵심적인 역할을 하고 있음을 시사한다. 또한 시장매출성장률(MktRev_G), 산업집중도(HHI), 국유기업 여부(SOE_D)와 같은 시장 및 기업 특성 변수들도 상당한 수준의 설명력을 보이며, 단일 재무지표 외에도 구조적 변수의 통합이 예측 성능 향상에 기여할 수 있음을 보여준다.

[그림 3]은 설명 가능한 부스팅 모형(Explainable Boosting Machine, EBM)을 통해 도출된 이차 상호작용 결과로, 산업의 매출액 증가율(MktRev_G)과 Altman_Z_X3 변수 간의 상호작용이 기업 부도 예측에 미치는 영향을 시각화한 것이다. 색상은 두 변수의 조합에 따라 예측값(부도 위험도)의 변화를 나타내며, 색상 스케일은 예측값의 증감 정도를 나타낸다.

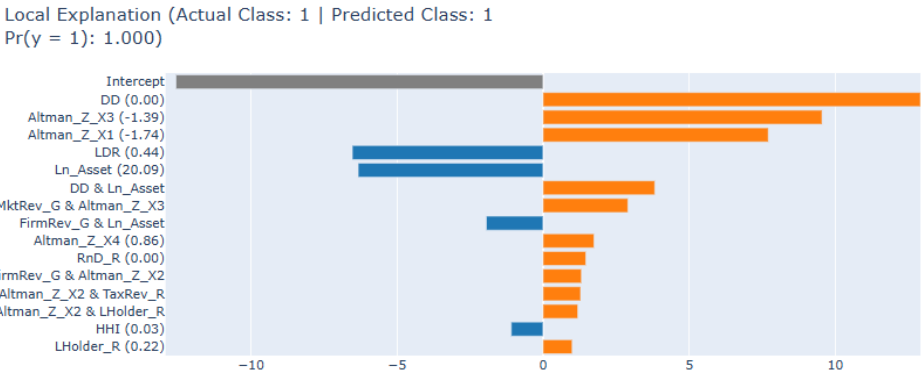


[그림 3] EBM 이차원적 상호작용

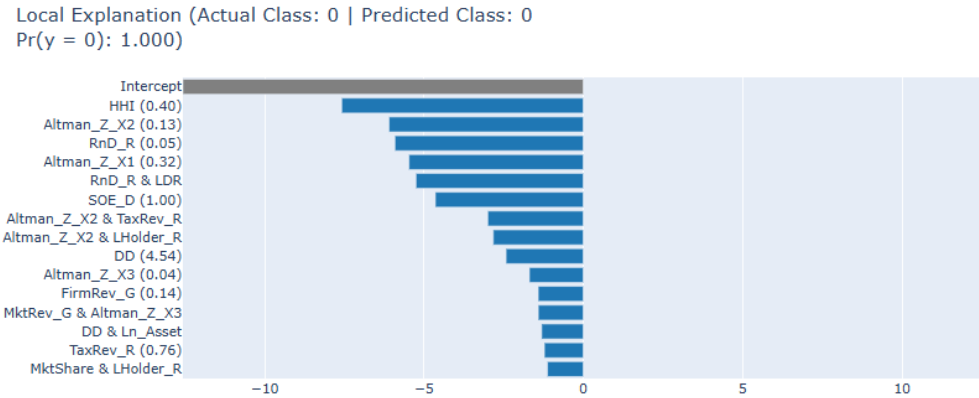
산업의 매출액 증가율(MktRev_G)이 높은 경우에는 개별 기업의 수익성(Altman_Z_X3)이 다소 낮더라도 부도 확률이 상대적으로 낮게 유지되는 경향을 보였다. 반면, 산업 성장률과 기업 수익성이 동시에 매우 낮은 구간에서는 부도 확률이 비선형적으로 급격히 증가하는 양상이 확인되었다. 이는 거시적 시장 환경과 개별 기업 재무 성과 간의 상호작용이 부도 위험에 중대한 영향을 미침을 시사하며, 단일 변수 분석으로는 포착하기 어려운 복잡한 위험 구조를 EBM을 통해 보다 명확히 해석할 수 있음을 보여준다.

[그림 4]는 EBM 모형을 활용하여 실제 부도 기업(Actual Class: 1)에 대한 기업별 재무적 특성적(local) 예측 결과를 시각화한 것이다. 구조적 신용 위험 측정치인 부도거리(DD)는 해당 기업의 부도 확률에 가장 큰 영향을 미친 요인으로 나타났으며, 유보이익 비율(Altman_Z_X2), 영업이익 비율(Altman_Z_X3)은 부도 가능성을 높이는 방향으로

작용하였다. 반면, 기업 특성 변수 기준인 은행 부채예금비율(LDR), 총자산 규모의 자연로그(Ln_Asset) 등은 부도 위험을 완화하는 요인으로 기여하였다. 이와 같이 EBM 모형을 통해 개별 관측치 수준에서 주요 변수들의 예측기여도를 정량적으로 파악함으로써, 단일 기업에 대한 부도 판단의 논리적 근거를 확인할 수 있다.



[그림 4] EBM 기반 기업별 재무적 특성적(local) 변수 기여도 분석 결과



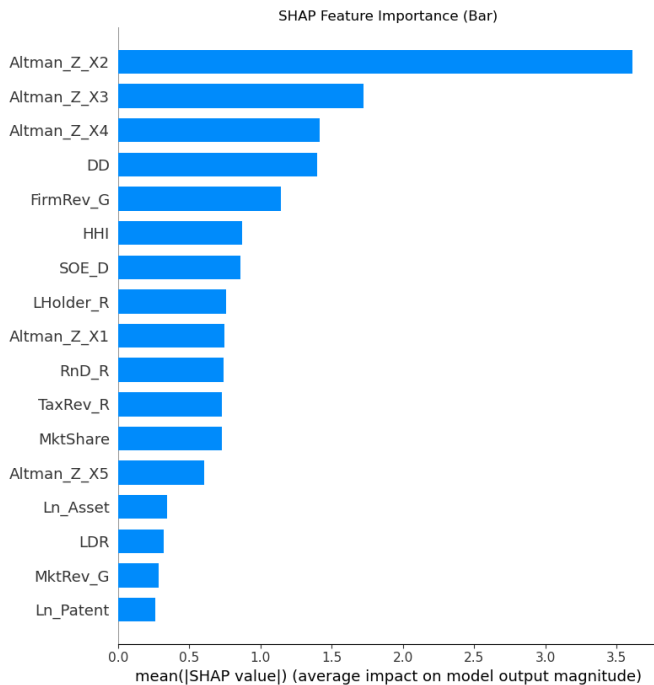
[그림 5] EBM 기반 기업별 재무적 특성적(local) 변수 기여도 분석 결과

[그림 5]는 EBM 모형을 활용하여 실제 비부도 기업(Actual Class: 0)에 대한 기업별 재무적 특성적(local) 예측 결과를 시각화한 것이다. 산업집중도 지수(HHI), 유보이익 비율(Altman_Z_X2), 연구개발비 비율(RnD_R), 운전자본 비율(Altman_Z_X1) 등은 해당 기업의 부도 확률을 낮추는 방향으로 기여하였다. 이처럼 EBM 모형을 활용한 기업별 재무적 특성적 해석을 통해, 개별 기업의 부도 가능성에 대한 예측 결과의 논리적 근거를 정량적으로

파악할 수 있다.

2) SHAP 해석적 분석

[그림 6]은 SHAP(Shapley Additive exPlanations) 값을 기반으로 산출한 변수 중요도를 평균 절댓값 기준으로 정렬하여 시각화한 바 차트(Bar chart)이다. 가장 영향력이 큰 변수는 유보이익 비율(Altman_Z_X2)로, 해당 변수는 평균적으로 모형 예측에 가장 큰 영향을 미친다. 영업이익 비율(Altman_Z_X3), 시가총액 대비 부채비율(Altman_Z_X4), 부도거리(DD), 기업 매출액 증가율(FirmRev_Growth) 등의 순으로 나타났으며, 이는 기업의 수익성 및 재무 건전성이 부도 위험 예측에 중요한 역할을 함을 시사한다. 또한, 산업 집중도(HHI), 정부 소속 여부(SOE_D), 최대주주 지분율(LHolder_Rate) 등도 유의한 영향력을 보이며, 구조적 변수와 지배구조 변수 역시 무시할 수 없는 설명력을 지닌다.

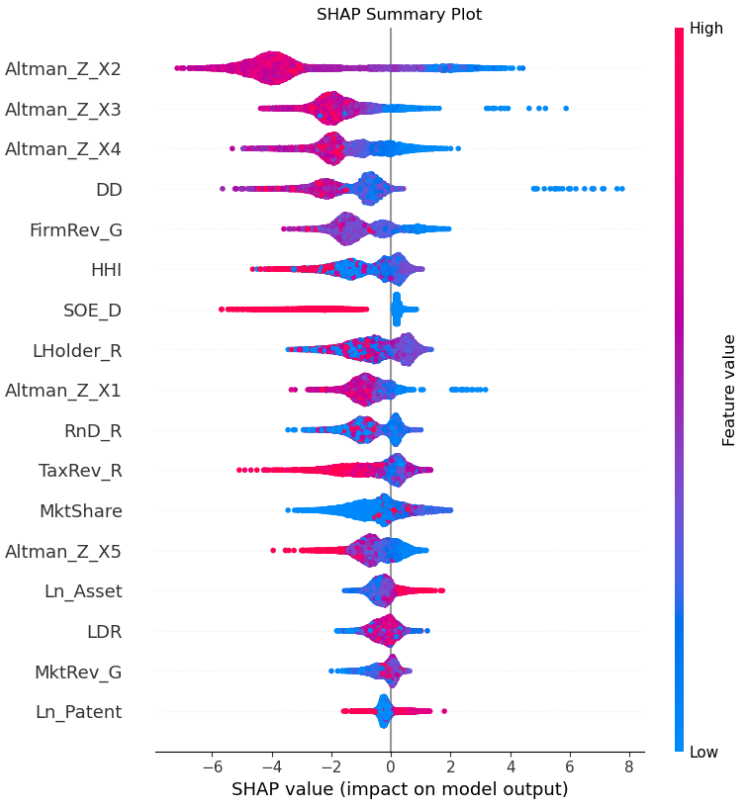


[그림 6] SHAP 변수 중요도 (평균 절댓값 기준)

[그림 7]은 SHAP 값과 각 변수의 실제 값(feature value)을 결합하여 변수들의 산업별 시장경쟁 구도적 영향력을 시각화한 그래프이다.

색상은 변수 값의 크기를 나타내며, 붉은색은 높은 값, 푸른색은 낮은 값을 의미한다. 가로축은 SHAP 값으로, 각 변수 값이 모형 예측에 미치는 방향성과 강도를 보여준다.

예를 들어, 유보이익 비율(Altman_Z_X2)의 값이 낮을수록(파란색) SHAP 값이 양(+)의 방향으로 나타나며, 이는 해당 변수의 감소가 부도 위험 증가에 기여함을 의미한다. 이러한 결과는 유동성 부족이 기업의 부도 가능성을 높인다는 기존 채무 이론과 연구결과와 부합한다.

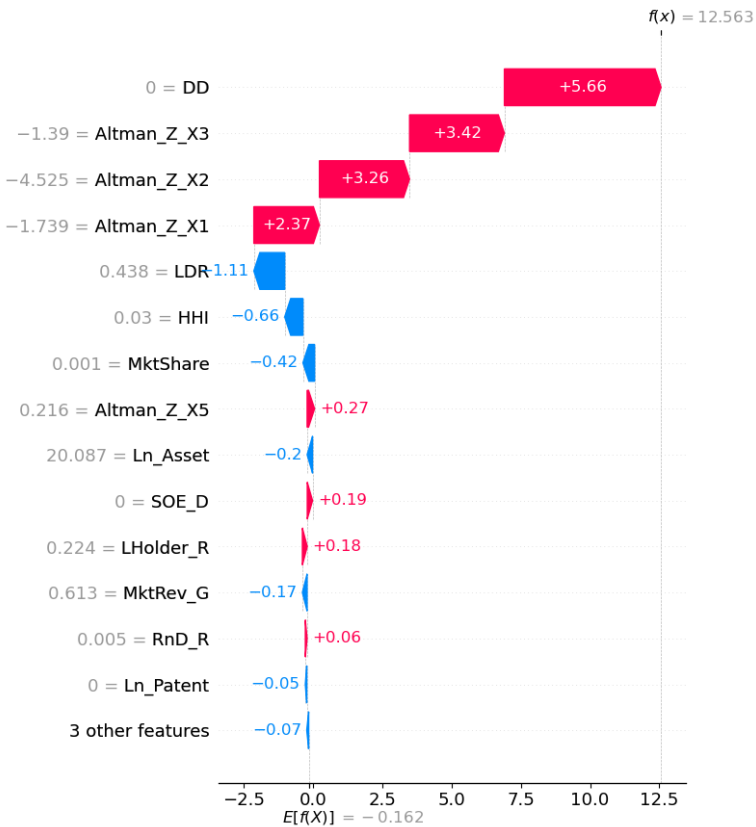


[그림 7] SHAP 기반 산업별 시장경쟁 구도 변수 영향력

또한, 기업 매출액 증가율(FirmRev_Growth)이나 연구개발비 비율(RnD_Ratio)과 같은 변수들 역시 값이 낮을수록 부도 가능성이 높아지는 방향으로 작용함을 보여주며, 이는 높은 성장성과 혁신 역량이 기업의 재무건전성에 긍정적인 영향을 미친다는 점을 시사한다.

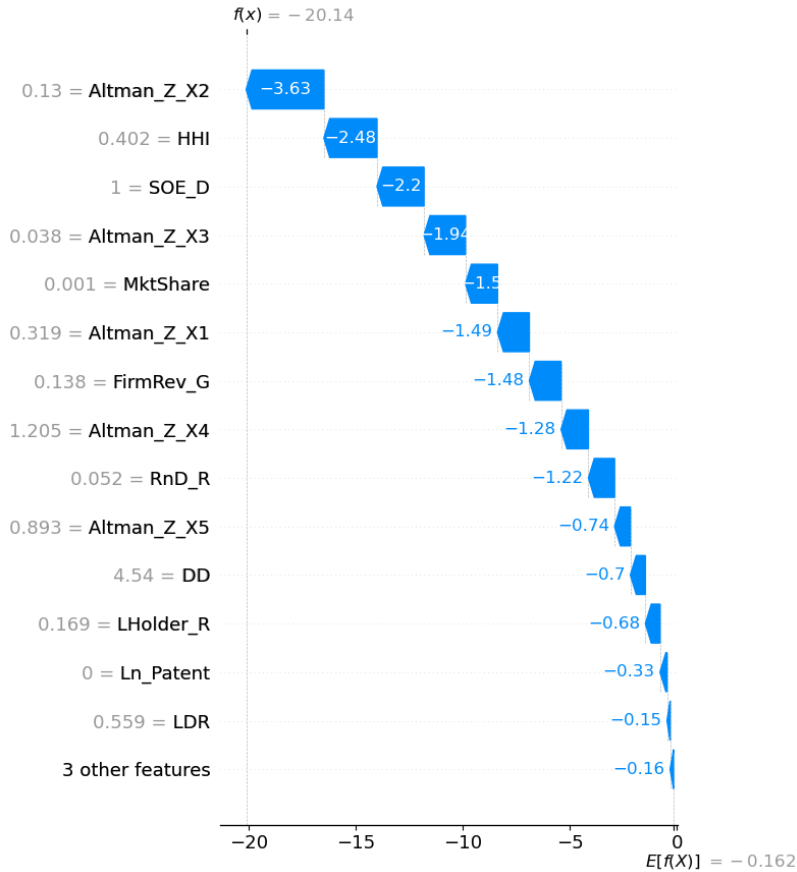
[그림 8]은 실제 부도 기업에 대한 SHAP 기반 기업별 재무적 특성적(local) 변수 기여도 분석 결과를 나타낸 것이다. 전체 예측값 중에서, 부도거리(DD), 영업이익 비율(Altman_Z_X3),

유보이익 비율(Altman_Z_X2), 운전자본 비율(Altman_Z_X1)은 각각 +5.66, +3.42, +3.26, +2.37의 SHAP 값을 보이며 부도 위험을 증가시키는 주요 변수로 작용하였다. 반면, 기업 특성 변수 기준의 은행 부채예금비율(LDR), 산업집중도(HHI), 시장점유율(Market Share, MktShare) 등은 음의 SHAP 값을 보여 부도 가능성을 낮추는 방향으로 기여한다.



[그림 8] SHAP 기반 기업별 특성적(local) 변수 기여도 분석 결과: 부도기업 사례

[그림 9]는 실제 비부도(non-bankruptcy) 기업 사례에 대한 SHAP 기반 기업별 재무적 특성적(local) 변수 기여도 분석 결과를 보여준다. 총 예측값 중에서, 유보이익 비율(Altman_Z_X2), 산업 집중도(HHI), 국유기업 여부(SOE_D), 영업이익 비율(Altman_Z_X3)은 각각 -3.63, -2.48, -2.20, -1.94의 SHAP 값을 나타내며 부도 위험 감소에 크게 기여한 변수들이고 시장점유율(MktShare), 운전자본 비율(Altman_Z_X1), 기업매출성장률(FirmRev_G), R&D 투자비율(RnD_R)은 음(-)의 방향으로 작용하여 부도 가능성을 낮추는 데 영향을 미쳤다.



[그림 9] SHAP 기반 기업별 특성적(local) 변수 기여도 분석 결과: 비 부도기업 사례

두 방법론의 결과는 미세하게 차이가 있으며, 이는 모형의 특성과 계산 방식의 차이에서 비롯된다. EBM 모형은 본질적으로 자체적으로 설명 가능한(interpretable) 구조를 가진 가산적(additive) 모형으로, 변수의 전역적(global) 중요도를 직접 산출한다. 반면 SHAP은 모형에 독립적으로 적용되는 사후(post-hoc) 해석 기법으로, 모형 복잡도에 의존하지 않고 각 변수의 한계적 기여도를 반복 계산하여 예측 결과에 대한 평균적 중요도를 도출한다. 따라서 SHAP은 특정 상황(예: 부도 기업의 이질적 패턴)에서 변수의 영향력을 더 강조하는 반면, EBM은 전체 표본에서 설명력이 큰 변수들을 상대적으로 부각시키는 경향이 있다.

그러나 두 방법 모두 공통적으로 기업의 재무적 특성(예: 부도거리 DD, Altman Z-scores)을 중요한 변수로 도출하였으며, 이는 기업의 기초적 재무건전성이 부도 예측에서 핵심적 역할을 한다는 점을 시사한다. 또한 산업집중도(HHI)와 같은 산업 특성 역시 높은 중요도를

보여, 외부 산업·시장 구조가 기업의 부도 위험에 실질적인 영향을 미친다는 점을 확인할 수 있다. 즉, 두 방법론이 도출한 결과에는 일부 이질성이 존재하지만, 상호 보완적으로 해석할 때 재무 요인과 산업 요인의 중요성이 일관되게 확인할 수 있는 것이다.

V. 결 론

본 연구는 2016년부터 2023년까지의 중국 상장기업 데이터를 바탕으로, 기업 부도위험에 영향을 미치는 결정요인을 전통적인 재무비율 보다는 기업의 산업내 경쟁력과 산업구조 중심으로 분석하고, 이를 바탕으로 다양한 예측모형을 비교함으로써 부도위험 예측의 정확도와 해석력을 동시에 제고를 하였다. 연구는 크게 두 부분으로 구성되며, 첫째는 고정효과 패널 회귀모형(Fixed Effects Panel Regression), 로지스틱 회귀(Logistic Regression), 생존분석(Cox Proportional Hazards Model)을 활용한 가설 검증, 둘째는 로지스틱 회귀, 설명 가능한 기계학습(Explainable Boosting Machine, EBM), 그리고 블랙박스 모형인 XGBoost를 활용한 예측 분석으로 구성된다. 또한, EBM을 통해 산업별 시장경쟁 구도 및 기업 특성 수준의 해석을 수행하였고, XGBoost 모형은 SHAP(SHapley Additive exPlanations)를 통해 예측 결과에 대한 전후방 해석을 시도하였다. 연구결과를 요약하면 다음과 같다.

첫째, 회귀분석 결과에 따르면, 부도거리(Distance to Default, DD)를 종속변수로 사용할 경우 산업집중도(HHI)는 부도위험에 양(+)의 영향을, 산업의 매출액 증가율, 시장점유율, 기업의 매출액 증가율은 음(-)의 영향을 미치는 것으로 나타났으며, 네 변수 모두 통계적으로 유의하였다. 그러나 종속변수를 부도여부를 전환한 로지스틱 회귀 및 콕스모형에서는 네 변수 중 기업의 매출액 증가율만이 유의한 영향을 보였고, 그 영향 방향은 부도위험을 완화하는 음(-)의 관계였다.

둘째, 예측모형 비교 결과, XGBoost는 가장 뛰어난 예측정확도를 보였으며, EBM이 그 뒤를 이었고, 전통적 로지스틱 회귀모형은 상대적으로 낮은 성능을 보였다. XGBoost와 EBM 간의 예측력 차이는 크지 않으나, EBM은 높은 예측 성능과 동시에 우수한 해석력을 제공한다는 점에서 정책적 활용한다.

셋째, 산업별 시장경쟁 구도 및 기업 특성 해석 결과에 따르면, 기업의 재무 건전성 지표인 부도거리(DD)가 모든 모형에서 가장 높은 예측기여도를 보였으며, 여전히 기업 부도위험 예측의 핵심 변수임이 확인되었다. 그러나 산업집중도, 산업의 매출액 증가율 등 산업 수준의 변수 역시 상당한 수준에서 예측정확도를 향상하는 것으로 나타났다.

넷째, EBM의 변수 간 상호작용 분석 결과, 산업 전체 매출 증가율이 높은 경우에는 기업의 수익성이 일시적으로 악화더라도 부도확률이 급격히 증가하지 않는 경향이 발견되었다. 반면, 산업 전체 매출액이 감소하는 상황에서는 기업의 재무상태 악화가 부도위험을 급격히 증가시키는 것으로 나타나, 이러한 효과가 비선형적으로 작용함을 보여준다.

다섯째, XGBoost는 블랙박스 모형임에도 불구하고 SHAP 기반의 사후 해석 도구를 통해 산업별 시장경쟁 구도 및 기업 특성 차원에서 예측 근거를 확인할 수 있었다. 이를 통해, Explainable AI 기계학습뿐만 아니라 블랙박스 모형 또한 적절한 해석 도구를 활용할 경우, 설명력 있는 정책적 통찰력을 제공할 수 있음을 시사한다.

본 연구는 기업 재무정보뿐만 아니라 산업 수준의 경쟁구조적 특성변수들이 부도위험 예측에 기여함을 실증적으로 입증하였다. 이에 따라 정책 및 금융 감독기관은 개별 기업의 재무 건전성 평가뿐 아니라, 기업이 속한 산업의 성장성, 집중도, 경쟁구조 등의 거시적 조건을 종합적으로 고려한 위험 관리 체계를 마련할 필요가 있다. 특히 산업매출이 감소하는 시기에는 개별 기업의 재무 취약성이 더 쉽게 부도위험으로 전이될 수 있으므로, 산업별 경기 변동에 대한 조기경보 시스템 구축과 차등적인 금융완화 정책이 요구된다. 또한, 설명 가능한 기계학습 기반의 모형(EBM)을 활용함으로써, 규제기관이 부도예측 결과를 정량적으로 해석하고 정책 대응에 활용할 수 있다.

그러나 본 연구는 몇 가지 제한점을 내포하고 있다. 첫째, 본 연구는 상장기업 표본과 정형화된 재무·산업지표에 한정되어 비상장·중소기업이나 비정형 데이터의 부도 징후를 반영하지 못했고, SHAP 기반 해석은 변수 간 인과 추론에 한계가 있다. 둘째, 중국 A주 시장은 국유기업 비중, 정부 개입 등 제도적 특수성이 크기 때문에, 이러한 요인이 분석 결과에 영향을 미쳤을 가능성이 있으며, 연구 결과가 다른 국가나 시장에도 일반화될 수 있는지에 대해서는 신중한 논의가 필요하다. 따라서 향후 연구에서는 본 모형을 중국에 한정하지 않고 다양한 국가와 시장에 적용하여 그 보편성과 적합성을 검증할 필요가 있다.

참 고 문 헌

- 김지홍, 문남미, “XAI 기반 기업부도예측 분류모델 연구”, 정보처리학회 논문지(KTSDE), 제12권 제8호, 2023, 333-340.
- Agarwal, R. and M. Gort, “Firm and Industry Dynamics in a Model of Adjustment Costs,” *International Journal of Industrial Organization*, 17(8), (1999), 1115-1144.
- Alexandropoulos, D., G. Patsakis, and E. Vergidis, “Deep Learning for Bankruptcy Prediction: A Comprehensive Analysis,” *Expert Systems with Applications*, 128, (2019), 135-150.
- Altman, E. I., “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *Journal of Finance*, 23(4), (1968), 589-609.
- Aretz, K., D. Arora, and S. Bartram, “Corporate Default Risk and the Macroeconomy: International Evidence,” *Journal of Financial Economics*, 128(3), (2018), 633-651.
- Arratia, A. and M. Sepúlveda, “Visualizing Financial Data for Default Prediction Using Convolutional Neural Networks,” *Finance Research Letters*, 28, (2019), 73-80.
- Bain, J. S., “Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940,” *The Quarterly Journal of Economics*, 65(3), (1951), 293-324.
- Baldwin, J. R. and P. Gorecki, “Firm Entry and Exit in the Canadian Manufacturing Sector,” *Canadian Journal of Economics*, 23(4), (1990), 651-676.
- Beaver, W. H., “Financial Ratios as Predictors of Failure,” *Journal of Accounting Research*, 4, (1966), 71-111.
- Bharath, S. T. and T. Shumway, “Forecasting Default with the Merton Distance to Default Model,” *Review of Financial Studies*, 21(3), (2008), 1339-1369.
- Black, F. and J. C. Cox, “Valuing Corporate Securities: Some Effects of Bond Indenture Provisions,” *Journal of Finance*, 31(2), (1976), 351-367.
- Borenstein, S. and N. L. Rose, “Competition and Price Dispersion in the U.S. Airline Industry,” *Journal of Political Economy*, 102(4), (1994), 653-683.
- Campbell, J. Y., J. Hilscher, and J. Szilagyi, “In Search of Distress Risk,” *Journal of Finance*, 63(6), (2008), 2899-2939.
- Castrén, O. and I. K. Kavonius, “Balance Sheet Interlinkages and Macro-Financial Risk Analysis in the Euro Area,” *ECB Working Paper*, (2009).

- Chen, T. and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), 785–794.
- Choi, B. P. and M. A. Weiss, "An Empirical Investigation of Market Structure, Efficiency, and Performance in Property–Liability Insurance," *Journal of Risk and Insurance*, 72(4), (2005), 635–673.
- Coffinet, J., A. Pop, and A. Tiesset, "Monitoring Financial Distress in a High-Frequency Environment," *Banque de France Working Paper*, (2010).
- Cox, D. R., "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), (1972), 187–202.
- Duffie, D., L. Saita, and K. Wang, "Multi-Period Corporate Default Prediction with Stochastic Covariates," *Journal of Financial Economics*, 83(3), (2007), 635–665.
- Fama, E. F. and K. R. French, "New Lists: Fundamentals and Survival Rates," *Journal of Financial Economics*, 73(2), (2004), 229–269.
- Geske, R., "The Valuation of Corporate Liabilities as Compound Options," *Journal of Financial and Quantitative Analysis*, 12(4), (1977), 541–552.
- Goldstein, A., A. Kapelner, J. Bleich, and E. Pitkin, "Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation," *Journal of Computational and Graphical Statistics*, 24(1), (2015), 44–65.
- Gropp, R., J. Vesala, and G. Vulpes, "Equity and Bond Market Signals as Leading Indicators of Bank Fragility," *Journal of Money, Credit and Banking*, 38(2), (2006), 399–428.
- Harada, K., T. Ito, and T. Takahashi, "Is the Distance to Default a Good Measure in Predicting Bank Failures? Case Studies," *IMF Working Paper*, (2010).
- Hastie, T. and R. Tibshirani, "Generalized Additive Models," *Statistical Science*, 1(3), (1986), 297–310.
- Honjo, Y., "Business Failure of New Firms: An Empirical Analysis Using a Multivariate Model," *International Journal of Industrial Organization*, 18(4), (2000), 557–574.
- Jessen, C. and D. Lando, "Robustness of Distance-to-Default," *Journal of Banking & Finance*, 50, (2015), 493–505.
- Kealhofer, S. and M. Kurbat, "The Default Prediction Power of the KMV Model," *KMV Corporation Technical Document*, (2001).

- Kukuk, M. and C. Rönnberg, "Mixed Logit Models in Corporate Bankruptcy Prediction," *Journal of Forecasting*, 32(6), (2013), 518-530.
- Lie, E., "Operating Performance Following Open Market Share Repurchase Announcements," *Journal of Accounting and Economics*, 39(3), (2005), 411-436.
- Lundberg, S. M. and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, 30, (2017), 4765-4774.
- Mai, F., D. Schmalz, and B. Sun, "Deep Learning for Corporate Bankruptcy Prediction: An Empirical Study of Financial Variables," *European Journal of Operational Research*, 272(2), (2019), 632-646.
- Marcus, A. J. and I. J. Shaked, "The Valuation of FDIC Deposit Insurance Using Option-Pricing Estimates," *Journal of Money, Credit and Banking*, 16(4), (1984), 446-460.
- Mata, J. and P. Portugal, "Life Duration of New Firms," *Journal of Industrial Economics*, 42(3), (1994), 227-245.
- Merton, R. C., "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, 29(2), (1974), 449-470.
- Nori, H., S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A Unified Framework for Machine Learning Interpretability," arXiv preprint arXiv:1909.09223, (2019).
- Odom, M. D. and R. Sharda, "A Neural Network Model for Bankruptcy Prediction," *Proceedings of the IEEE International Joint Conference on Neural Networks*, 2, (1990), 163-168.
- Ohlson, J. A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, 18(1), (1980), 109-131.
- Opler, T. C. and S. Titman, "Financial Distress and Corporate Performance," *The Journal of Finance*, 49(3), (1994), 1015-1040.
- Porter, M. E., *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, Free Press, (1980).
- Scherer, F. M. and D. Ross, *Industrial Market Structure and Economic Performance*, University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship, (1990).
- Shumway, T., "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *The*

- Journal of Business*, 74(1), (2001), 101–124.
- Sinha, R. K., and Kumar, P., “Can Explainable AI Build Trust in the Financial Domain: Study of an Explainable AI Model for Bankruptcy Prediction”, In *Advanced Computing and Communications Conference*, Cham: Springer Nature Switzerland, 2024, 404–417.
- Sutton, J., “Sunk Costs and Market Structure: Price Competition”, *Advertising, and the Evolution of Concentration*, MIT Press, (1991).
- Tian, S., T. Yu, and L. Hu, “Bankruptcy Prediction Using LASSO Variable Selection,” *Finance Research Letters*, 13, (2015), 191–198.
- Traczynski, J., “Bayesian Corporate Default Risk Models,” *Journal of Financial Econometrics*, 15(2), (2017), 372–397.
- Turnbull, S. M., “The Valuation of Options on Corporate Debt with Taxes and Default,” *Journal of Finance*, 34(2), (1979), 551–562.
- Vochozka, M., J. Klietik, and P. Krulicky, “Predicting Financial Distress Using LSTM Networks: Evidence from European Firms,” *Sustainability*, 12(3), (2020), 1257.
- Zmijewski, M. E., “Methodological Issues Related to the Estimation of Financial Distress Prediction Models,” *Journal of Accounting Research*, 22, (1984), 59–82.

THE KOREAN JOURNAL OF FINANCIAL MANAGEMENT
Volume 42, Number 4, August 2025

Industry and Market Structure in Bankruptcy Prediction: An Explainable AI Approach

Yong Shang Liu* · Sunghwan Kim**

〈Abstract〉

This study aims to address the limitations of conventional bankruptcy prediction models that primarily rely on firm-level financial data and fail to adequately incorporate industry and market structure competition structure factors, while simultaneously enhancing both predictive accuracy and interpretability. We measure bankruptcy risk using two dependent variables: the continuous indicator of Distance to Default (DD) and the binary indicator of default occurrence (Default_D). The key explanatory variables include the Herfindahl–Hirschman Index (HHI), industry growth rate, market share, and firm revenue growth rate.

Methodologically, we combine traditional econometric techniques, including logistic regression, fixed effects panel regression, and Cox proportional hazards regression, with explainable AI models. Specifically, we apply the Explainable Boosting Machine (EBM) and XGBoost with SHAP, conducting both global and local interpretability analyses.

The empirical analysis is based on data for Chinese A-share listed firms from 2016 to 2023 obtained from the CSMAR database. The results show that, in the DD models, industry concentration positively affects bankruptcy risk, whereas industry growth rate, market share, and firm revenue growth rate have negative effects; all are statistically significant. In contrast, when using the binary default indicator, only firm revenue growth rate has a statistically significant negative effect. In terms of predictive performance, XGBoost outperforms all other models, followed by EBM, while logistic regression shows relatively lower accuracy. Interpretation results confirm DD as the most influential variable across all models, with industry-level variables also contributing to predictive improvements. EBM further reveals a nonlinear relationship between industry sales growth and bankruptcy risk, while SHAP-based interpretation of XGBoost provides policy-relevant insights. This study contributes to the literature by extending bankruptcy risk analysis beyond firm-level financials to incorporate quantitative industry and market competition structure factors, and by proposing an integrated predictive-interpretive framework that offers high academic and policy relevance through explainable AI techniques.

Keywords : Corporate Default, Distance to Default, Industry Competition, Firm Competence, Explainable Artificial Intelligence, EBM, XGBoost, SHAP

* First Author, Ph.D. Student, School of Business Administration, Kyungpook National University,
E-mail: tiwish@naver.com

** Corresponding Author, Professor, School of Business Administration, Kyungpook National University,
E-mail: indianak@knu.ac.kr