

머신러닝을 이용한 창업기업의 대출부도 예측

장영민*

논문초록

최근 중소기업 부도 예측 연구에 머신러닝(Machine Learning)을 활발하게 적용하고 있으나, 창업기업 대출부도 예측에는 머신러닝을 활용한 연구는 드물다. 중소기업 부도 예측 연구의 미비점을 보완하기 위해 본 연구는 창업기업 대출부도 예측을 위한 머신러닝의 정확성을 평가하는 것을 목적으로 한다. 이를 위해 국내 창업기업과 일반기업의 대출부도 예측에 가장 정확성이 높은 머신러닝은 각각 무엇이고, 그것은 통계적 모형과 어떤 성능적 차이가 있는지 비교하였다. 실증적 결과로부터 발견된 내용을 정리하면 다음과 같다.

첫째, 창업기업의 장단기(1년~3년) 대출부도 예측에서 정확성과 안정성이 우수한 머신러닝은 아다부스트(AdaBoost)로 평가할 수 있고, 일반기업의 단기 예측에는 서포트벡터머신(Support Vector Machine)이 좋은 성능을 보여주었다. 둘째, 로지스틱 회귀모형(Logistic Regression) 등 통계적 모형의 창업기업 대출부도 예측 정확성은 머신러닝과 거의 대등한 수준이었다. 오히려 통계적 모형보다 예측 성과가 미흡한 머신러닝이 존재한다. 셋째, 재무적 특성이 창업기업의 대출부도 예측에도 유의미한 설명력을 제공한다. 창업기업과 일반기업의 장단기 대출부도 예측에 부채상환능력과 유동성 재무비율들이 유효하였다. 본 분석이 시사하는 바는 머신러닝이 중소기업 대출부도 예측의 정확성을 확실히 개선하는 것은 분명하나 예측 변수의 설명가능성이 저하되는 단점이 있어 이용 목적에 따라 유용성이 달라진다. 본 논문은 창업기업의 대출부도 예측을 위해 시점간 적합성(Intertemporal Validation) 검증 방법을 통한 다수의 머신러닝과 재무적 특성의 유용성을 평가한 점에서 공헌이 있다.

주제어

창업기업, 머신러닝, 대출부도 예측, 중소기업, 재무적 특성

Loan Default Prediction of Startups Using Machine Learning

Youngmin Jang*

Abstract

Although machine learning(hereafter ML) has been actively applied to recent research on predicting the default of small and medium-sized enterprises(hereafter SMEs), studies using ML to predict loan defaults of startups are rare. To bridge the gap in the field of SMEs default prediction research, this study aims to evaluate the accuracy of ML for predicting loan defaults of startups. To this end, we compares which ML model is most accurate for predicting loan defaults of domestic startups and established companies, and how its performance differs from statistical models. The main findings from the empirical results are summarized as follows:

Firstly, AdaBoost has been evaluated as highly accurate and stable for predicting short- to long-term (1 to 3 years) loan defaults in startups, while support vector machines showed good performance for 1-year predictions in established company. Secondly, the accuracy of loan default prediction for startups using statistical models, such as the logistic regression, was found to be nearly equivalent to that of ML. However, there were ML techniques that performed worse than statistical models. Thirdly, financial features provide significant explanatory power in predicting loan defaults of startups. Debt capacity ratio and liquidity ratios were effective for both short- and long-term loan default predictions in startups and established businesses. The analysis suggests that while machine learning significantly improves the accuracy of predicting loan default for SMEs, it has the drawback of lower interpretability of predictive variables. Therefore, its usefulness varies depending on the purpose of use. This paper contributes by evaluating the usefulness of multiple ML and the effectiveness of financial features for predicting the loan default of startups through an intertemporal validation method.

Keywords

Startups, Machine Learning, Loan Default Prediction, Small and Medium-sized Enterprises, Financial Features

* Corresponding Author, Korea Credit Guarantee Fund, Research Fellow, Ph.D. in Economics(jangym@kodit.co.kr)

I. 서론

금융권에서 최근 인공지능 기법을 업무 전반에 적극 활용하려는 노력이 확대되고 있다. 변화된 금융 시장 환경과 격화된 경쟁에 직면한 금융기관은 이자마진을 높이고 대출업무 효율을 개선하기 위해 인공지능을 도입하는 것이 이제 생존을 위한 필수 불가결한 요소가 되고 있다. 금융기관이 일선에서 우수 고객을 유치하고 고위험 고객을 선별하려면 신용평가의 정확성을 기하는 것이 매우 중요하다. 특히 중소기업 대출은 담보에 의존하는 개인 대출과 달라서 기업 신용도의 정교한 평가가 더 요구된다.

금융기관의 중소기업 신용평가는 바젤II 도입을 전후로 전문가 평가제도(Expert System) 중심에서 통계적 모형(Statistical Model) 기반 체제로 전환되었다. 통계적 모형은 기업의 대출부도(Loan Default)에 영향을 미치는 재무적 특성 등 제요소를 가중 결합하여 대출부도를 예측하는 모수적 모형(Parametric Model)을 일컫는다. 현재 국내 금융기관은 주로 로지스틱 회귀모형(Logistic Regression; LR)을 대출부도 예측 모형으로 사용한다. 로지스틱 회귀모형은 기업의 대출부도 가능성을 확률로 직접 추정하는 이점 외에도 대출승인 가부에 대한 설명이 가능하여 범용적으로 활용된다.

기업 신용평가에서 통계적 모형의 활용이 비교적 근래의 일이라면 학계의 기업 부도 예측 연구는 거의 1세기 전으로 거슬러 올라가고,¹⁾ 재무적 특성을 선형결합한 통계적 모형으로 부도를 예측한 것은 Altman(1968)의 판별분석(Linear Discriminant Analysis; LDA)이 효시이다.²⁾ 그런데 Odom and Sharda(1990)가 부도 예측에 신경망(Neural Networks; NN)을 최초로 적용한 이후 다양한 머신러닝(Machine Learning; ML) 기법들이 발전하면서 ML을 이용한 기업 부도 예측이 금융계에 큰 반향을 불러일으켰다.³⁾ 물론 이런 배경에는 빅데이터의 활용과 컴퓨터 기술의 눈부신 발전도 빼놓을 수 없다. ML 관점에서 부도 예측은 부도기업과 건전기업을 판별하기 위한 분류 문제로 일종의 패턴인식이다. ML의 두드러진 강점은 통계적 모형의 한계로 지적되는 재무적 특성과 부도와의 비선형적 관계와 재무적 특성 사이의 복잡한 상호작용을 고려할 수 있다는 것이다.

그래서 ML을 사용하면 기업 부도 예측의 정확성이 통계적 모형보다 향상되는지와 예측의 정확성이

1) 기업 부도 예측에 관한 연구는 부도의 정의에 따라 실패(Failure), 파산(Bankruptcy), 재무적 곤경(Financial Distress), 대출부도(Loan Default), 지급불능(Insolvency) 등의 예측으로 혼재된다. 본 논문에서는 용어의 혼동을 줄이기 위해 이러한 개념들을 부도로 통칭코자 한다. 대체로 사업실패(Business Failure)의 양상을 파산과 재무적 곤경으로 구분하며, 현금흐름의 일시적인 악화에서 발생하는 지급불능과 부도는 재무적 곤경에 해당한다. Altman and Hotchkiss(2006)에 의하면, 지급불능이란 기업의 부채가 자산보다 커서 기업이 해당 채무를 이행할 수 없게 되는 경우로 유동성 부족과 관련되고, 부도는 기업이 채무를 이행하지 못할 때 특히 대출 채무자가 금융기관과 맺은 계약조건을 이행하지 못할 때 발생한다. 실패는 투자자본의 실현수익률이 위험조정 경상수익률보다 연속으로 낮을 때를 의미한다.

2) Bellovary et al.(2007)은 1930년 이후 165개 부도 예측 문헌을 토대로 가장 보편적으로 활용되는 재무비율 42개의 목록을 제시하였다.

3) 금융부문에서 머신러닝의 이용에 관한 문헌조사는 Nazareth and Reddy(2023)가 있다.

가장 높은 ML 모형을 발견하는 것이 이 분야 연구의 주요 초점이었다. 실증적 결과에 따르면 대개 분석 당시에 새롭게 등장하였거나 유행했던 ML 알고리즘들이 예측 성능에서 두각을 드러냈다.⁴⁾ 그런데 기업 부도에 관해 폭넓게 수용되는 이론의 부재는 저자로 하여금 예측변수로서의 재무적 특성을 임의적으로 선택하게 한다. 이전 연구들은 보통 일군의 재무변수로부터 선택하는 과정을 거치거나 아예 미리 정의한 소수 재무비율을 사용하였다. 그런데도 재무적 특성이 부도 예측의 성능을 개선하는 상당히 의미 있는 실증적 증거들이 제시되었다. 그리고 ML의 취약점은 입력변수의 설명가능성(Interpretability)에 있는데 최신 방법들이 소개되어 연구자들은 이를 통해 재무적 특성의 영향력을 평가하였다.

만일 중소기업 재무 자료에 대해 제기되는 투명성과 신뢰성의 질적 수준이 문제가 된다면 재무비율을 이용한 중소기업 부도 예측은 그 정확성이 대기업에 비해 현저히 떨어져야 한다. 놀랍게도 국내외 선행연구들은 중소기업 부도 예측에 대한 재무비율의 유효성을 입증하고 있다. 그중 Son et al.(2019)의 연구는 재무비율을 주요 요소로 국내 비외감기업과 외감기업의 부도를 예측했을 때 비외감기업에서도 좋은 성과를 보여주었다.

사실 중소기업은 경영 양태의 다양성 측면에서 대기업보다 훨씬 이질적인 집단이다. 중소기업 부도 예측에서 예측의 정확도를 높이기 위해서는 이러한 이질성을 적절하게 다루어야 한다. 그 방법으로 기업이 영위하는 사업 분야를 세분화하여 부도를 예측하고 사업 분야간 부도 예측 성과를 비교한 연구가 몇 있다. 그런데 국내외를 막론하고 중소기업의 성장 단계를 구분하여 ML의 부도 예측 성과를 비교한 연구는 찾아보기 어렵다. 기업의 업력을 모형의 입력변수로 사용하는 정도에 그치고 있지 업력을 차등화하여 부도를 예측하려는 시도는 아직 이루어지지 않고 있다.

금융기관 입장에서 중소기업의 성장 단계에 적합한 대출부도 예측이 중요한 이유는 크게 두 가지가 있다. 첫째, 금융기관은 담보가 부족하고 신용위험이 높은 창업기업에 대출을 기피하는 경향이 있다. 그래서 창업기업의 금융 소외를 해소하기 위해 정책금융 지원의 초점이 창업기업에 맞춰져 있는데 한정된 재원의 효율적 배분과 재정 건전화의 위해서는 창업기업에 대해 정확한 신용위험 평가가 중요하기 때문이다. 둘째, 창업 초기에 대출의 신용위험은 상대적으로 높다. 다시 말해서 창업기업의 관찰된 대출부도율이 일반기업의 대출부도율을 상회한다. 따라서 창업기업의 매우 제한된 정보로 다수의 대출부도를 정확하게 예측하는 ML 모형을 탐색하는 것은 중소기업 포트폴리오의 대출부도 예측 정확성 향상과 위험관리를 위해 필요하다.

본 논문은 이에 중소기업 부도 예측 연구 분야의 미비점을 보완하고자 ML을 이용해 창업기업의 대

4) Alaka et al.(2018)은 부도 예측 모형을 개발할 때 전문가들이 통계적 모형과 ML에 대한 정확한 이해 없이 당시의 유행이나 직업적 배경에서 모형을 선택하는 점에 우려를 표했다. 그래서 2010~2015년 사이에 49개 논문을 체계적으로 검토하고 부도 예측 모형을 선택할 때 고려해야 할 13가지 기준을 제안하였다.

출부도를 예측하는데 목적이 있다. 국내 창업기업의 대출부도 예측에 뛰어난 성능을 보이는 ML은 무엇이며 그것이 일반 중소기업 또는 통계적 모형과 예측 정확성에서 실제로 어느 정도 격차가 존재하는지 고찰한다. 이는 동시에 대출부도 예측변수로서 재무적 특성이 창업기업에도 유효한지 평가하는 것이다.

본 논문은 중소기업 부도 예측 연구에 두 가지를 기여한다. 첫째, 창업기업의 대출부도 예측에 ML의 유용성을 평가하고 이를 일반기업과 비교한 점이다. 전통적 ML 기법부터 가장 최신의 ML 기법을 동원하되 응용 모형보다는 이전 연구에서 사용 빈도가 높은 기본 모형 위주로 10개의 ML 알고리즘과 2개의 통계적 모형을 사용하였다. 이용된 예측 모형은 의사결정트리(Decision Tree; DT), 나이브 베이즈(Naïve Bayes; NB), 다층퍼셉트론(Multi-Layer Perceptron; MLP), 서포트벡터머신(Support Vector Machine; SVM), 배깅(Bagging), 랜덤포레스트(Random Forest; RF), 아다부스트(Adaptive Boosting; AdaBoost), XGBoost(Extreme Gradient Boosting), LightGBM(Light Gradient Boosting Machine), CatBoost(Categorical Boosting), 판별분석, 로지스틱 회귀모형이다. 그리고 재무비율에만 국한하지 않고 업종과 지역의 기업체 특성을 사용하였고 특별히 단기(1년) 대출부도 예측에서 더 나아가 중장기(2년/3년) 대출부도 예측까지 ML의 유용성을 평가한 점에 의의가 있다. Liu et al.(2022)도 창업기업을 분석하였지만 창업기업에 대한 정의가 불분명하고 그 대상도 하이텍 기업으로 한정하여 유럽 7개국을 비교한 점에서 본 논문과 다르다.

둘째, 선행연구들은 대개 모형 추정과 검증을 위한 표본을 같은 기간 내에서 분리하였다. 본 논문은 시점간 적합성(Intertemporal Validation) 검증을 목적으로 학습용(Training)과 검증용(Validation)은 같은 기간, 시험용은 시간외(Out-of-Time) 표본을 구성해서 ML의 대출부도 예측 결과를 엄밀히 평가하였다. 그 이유는 추정 모형의 과적합(Overfitting)은 검증용 표본으로 점검하고, 시간외 표본에 재차 적용함으로써 시간 경과에 따른 대출부도 예측의 정확성을 확인할 수 있기 때문이다. Jones and Hensher(2004)는 같은 기간에서 표집한 시험용 표본으로는 모형의 예측 성능 평가에 한계가 있음을 피력하였다. 개발된 대출부도 예측 모형은 일정 기간이 지난 이후에나 신용평가에 활용되는 금융기관의 현실적 측면에서도 시점간 적합성 검증은 매우 중요하다. 본 분석의 시험용 표본은 특별히 팬데믹 시기에 해당하여 경제위기에도 ML 모형의 안정성이 유지되는지 살펴볼 수 있었다. Zhang et al.(2023)은 10-겹 교차타당성(10-Fold Cross Validation)의 한 표본을 시험용으로 사용하였고, Barboza et al.(2017)은 시간외 표본을 검증용으로 활용하여 상장기업을 분석한 점에서 본 분석과 구별된다.

본 논문은 다음과 같이 진행한다. 제2장은 ML을 활용한 중소기업 부도 예측과 관련한 이론적 배경으로서 선행연구를 검토하고 본 분석에 사용할 모형을 개괄한다. 제3장은 분석 방법, 제4장에서는 기

초 분석을 수행한다. 제5장은 창업기업과 일반기업에 대한 ML과 통계적 모형의 장단기 대출부도 예측 결과를 자세히 다룬다. 마지막 장은 실증적 결과로부터 시사점과 연구의 한계점을 제시한다.

II. 이론적 배경

1. 선행연구의 4가지 관점

머신러닝(ML)을 이용한 중소기업 부도 예측에 관한 연구는 ML 기법의 발전 역사와 컴퓨터 소프트웨어의 기술적 진보와 밀접히 연관되어 있다. 즉 연대기적으로 가장 최근 논문에서 사용한 최신의 ML은 좀 더 앞선 연구에서는 사용될 수 없었으며, 컴퓨터 기술의 발전은 꽤 오래전 이론적으로 완성된 ML 모형이 시간이 지난 후에 비교적 쉽게 활용되는 길을 열어 주었다.

최고의 기업 부도 예측 성능을 보이는 ML 모형을 찾으려는 목적에서 이 분야 연구들은 예측 요인으로 재무적 특성에 기반을 두고 일단의 분석 자료에 다수의 ML 알고리즘을 적용하는 방법을 채택하였다. 실험적인 방법론적 설정은 자의적으로 선택된 ML 모형의 개수, 표본의 성격, 자료의 크기에 따라 실증적 결과의 불일치를 초래하였다. 특히 비슷한 시기에 발표된 연구는 당시 새롭게 등장하였거나 대중적인 ML 기법에 집중하였으며, ML과 예측 성능을 비교하기 위한 준거모형(Benchmark)으로서 통계적 모형을 이용하거나 혹은 다수의 ML 모형간 예측 성과를 상호 비교하는 경향이 두드러졌다.⁵⁾

이 절에서는 예측 정확성이 높은 ML 기법, 통계적 모형의 부도 예측 성과, 재무적 특성의 유효성, 정확성 개선과 관련한 주요 쟁점의 4가지 관점으로 중소기업 부도 예측 선행연구를 정리하였다.

첫째, 부도 예측 성능이 가장 우수한 ML 모형에 대해서는 일치된 결론에 이르지 못하고 있다. 시기적으로 더 최근 연구일수록 최신 방법과 다수의 검증된 ML을 동원하여 분석한 점을 고려하면 대체로 그라디언트 부스팅(Gradient Boosting; GB)류의 모형이 뛰어난 예측 성능을 보였고, 서포트벡터머신과 신경망이 뒤를 이었다. 정확성 평가 계량지표는 AUC(Area Under the Curve)와 혼동행렬(Confusion Matrix)이 보편적으로 사용되었다.

Pápić and Pápićová(2023)은 슬로바키아 10만개 이상 중소기업에 대해 XGBoost, LightGBM, CatBoost를 비교한 결과 CatBoost가 비교적 좋은 성과를 보였다. 그런데 2020년 팬데믹 위기에서 부스팅의 정확성은 하락하였는데 CatBoost의 하락폭이 특히 컸다. CatBoost의 성과는 Jabeur et al.(2021)의 실증결과에서도 확인되었다. 저자들은 2017년 프랑스의 133개 부도기업을 대상으로 ML

5) 머신러닝 연구들은 판별분석과 로지스틱 회귀모형을 머신러닝 알고리즘 부류에 포함하고 있으나 후자의 두 모형은 모수를 추정하는 구조적 모형(Structural Model)으로서 여타 ML 모형과 구별되는 점에서 본 논문에서는 통계적 모형으로 구분하였다.

을 검증한 결과 CatBoost가 XGBoost보다 우수한 예측력을 보였고 심지어 딥러닝(Deep Learning)보다도 정확성이 높았다. 이 연구는 건전기업의 수를 제시하지 않았고 소용량 표본이라는 한계가 있다. Son et al.(2019)의 2011~2016년 국내 약 85만개 비외감기업의 2만2천여개 부도기업 예측 결과에 의하면, XGBoost, LightGBM, 신경망의 순으로 예측 정확성이 높았다. Zięba et al.(2016)은 폴란드 제조업 부문의 2000~2013년 건전기업 약 1만개, 부도기업 약 700개에 ML을 적용하여 XGBoost의 우수성을 실증하였다. Liu et al.(2022)은 2005~2014년에 7개 서유럽 국가의 하이텍 스타트업 자료-국가에 따라 400~5,820개 실패기업-를 이용했을 때 XGBoost의 예측 성능이 가장 우수하였다. 그런데 저자들은 하이텍 스타트업의 정의를 명확히 제시하지 않았다. Moscatelli et al.(2020)은 이탈리아의 2011~2017년 기간에서 매년 약 25만개 기업을 대상으로 검증했을 때 중소기업에서 랜덤포레스트와 그래디언트 부스팅이 좋은 부도 예측 성능을 보였다. 약 3만8천개 이탈리아 중소기업에 대한 Figini et al.(2017)의 분석도 그래디언트 부스팅과 랜덤포레스트의 예측 강건성을 확인해 주었다. Bacham and Zhao(2017)는 재무비율만을 이용할 때는 랜덤포레스트, 신용정보와 함께 재무정보를 이용할 때는 그래디언트 부스팅이 가장 좋은 성과를 보였다.

몇몇 연구는 모형 결합을 시도하였다. Mselmi et al.(2017)의 2010~2013년 프랑스 중소기업을 대상으로 한 부도 예측에서 편최소자승(Partial Least Squares; PLS)과 서포트벡터머신을 결합한 모형이 제일 좋은 예측력을 보였다. Sigrist and Hirnschall(2019)은 그래디언트 부스팅과 토빗(Tobit)을 결합한 Grabit모형을 제안하고 동 모형을 2016~2017년 스위스 소재 141개 중소기업 표본에 적용하였다. Grabit모형은 여타 ML보다 높은 예측 성과를 보였다. 이상훈·유동희(2022)는 국내에서 중소기업진흥공단의 정책자금지원을 받은 중소기업의 약 3천2백개 부도기업 예측에서 배깅과 부스팅으로 구성된 더블 앙상블이 기저 학습기(Base Learner)의 성능을 개선하는 결과를 발견하였다.

한편, Véganzones and Séverin(2018)의 무작위 표집된 1,500개의 프랑스 기업 분석은 서포트벡터머신의 우수성을 입증하였다. Liang et al.(2016)의 1999~2009년 기간 대만의 259개 부도기업과 건전기업의 균형 표본에서도 서포트벡터머신의 예측 성과가 가장 좋았다. 송현준 등(2021)은 2012~2016년 국내 약 62만개 비외감기업의 2만개 이상 부도기업을 대상으로 ML의 예측 성능을 시험했을 때 표집 방법에 따른 차이는 있었으나 대체로 신경망의 예측력이 우수한 편이었다.⁶⁾

둘째, 통계적 모형은 ML과 견주어 대체로 낮은 예측 정확성을 보여주었으나 그 차이는 크지 않았고, 통계적 모형의 성능이 더 좋은 증거도 있었다. Mselmi et al.(2017)의 연구는 1년 부도 예측의 경우에

6) 김소정·이군희(2022)는 신용정보회사로부터 약 176만개의 매우 방대한 자료로 심층신경망(Deep Neural Network)을 이용하여 국내 중소기업 부도를 예측하였다. 분석 결과 심층신경망이 로지스틱 회귀모형보다 높은 예측 성과를 보였다. 이 결과는 불균형이 매우 큰 자료에서 나온 것이다.

편최소자승(PLS)을 서포트벡터머신과 결합했을 때보다 판별분석과 결합했을 때 예측 성과가 우수하였다.

통계적 모형의 예측 성능을 개선하기 위한 조건에 관해, Veganzones and Séverin(2018)은 자료 균형화(Data Balancing)의 중요성을 강조하였다. 또한 Moscatelli et al.(2020)에 의하면, 소용량 자료의 경우 통계적 모형이 ML과 대등한 정확성을 시현하였고, 재무정보 외에 은행대출액, 업력 등의 질적 신용정보를 활용할 때 로지스틱 회귀모형의 예측력은 더 향상되었다. 또한 보통 통계적 모형들이 정규성 가정에 기초하므로 극단값의 적절한 처리가 요구된다. Jones et al.(2015)은 Box-Cox 변환 입력변수를 사용하였으며, Figini et al.(2017)은 극단값 제거 방식을 통해 로지스틱 회귀모형의 예측 성능을 향상시켰다. 그러면서 부도 사건이 드문 상황에서는 부도확률을 과소추정할 수 있는 로짓(Logit)의 대칭 연계함수 대신 일반화 극단값(Generalized Extreme Value) 분포를 사용하면 좋은 결과를 얻을 수 있다고 주장하였다.

세 번째로 이전 연구에서는 부도 예측에 유효한 변수가 이용된 표본의 성격에 좌우되는 것으로 나타났다. 재무적 특성 중심의 부도 예측에 다양한 기업체 고유 속성을 보완하면 예측 모형의 성능은 개선되었다. 최근에는 업력, 산업, 등 기업체 특성, 그리고 대출액 수준, 연체 등 기업 신용정보, 심지어 온라인 기업활동 자료까지 포괄하여 기업의 신용도를 다면적으로 평가하고 있다. 하지만 더 많은 변수를 동원하여 부도 예측의 성과에만 집중하면 차원의 저주(Curse of Dimensionality)와 변수의 해석 문제를 심화시킬 수 있다. 입력변수의 추가에 따른 정확성의 개선 효과보다 모형 복잡도의 증가라는 부정적 효과가 더 클 수 있다. 그래서 예측변수의 부도에 대한 영향을 설명하기 어려운 점이 ML의 약점 중 하나로 꼽힌다. Dastile et al.(2020)에 의하면, 조사 문헌 중 겨우 8%만이 변수의 설명가능성을 다루고 있었다. 이 결과는 개인금융 부문 등의 연구까지 포괄한 결과이긴 하지만 본 논문에서 고찰한 연구들도 대부분 입력변수의 설명가능성을 보고하지 않았다.⁷⁾

일반적으로 선행연구들은 다수의 후보 입력변수 집단을 형성하고 몇 단계 특성 선택(Feature Selection) 과정을 거쳐 최종 예측변수를 결정하였다. 그러나 Barboza et al.(2017)은 특성 선택 과정 없이 처음부터 임의로 선택한 입력변수를 그대로 사용하였다. 선행연구에서 정의한 다양한 재무비율은 통상 성장성, 유동성, 수익성, 레버리지(Leverage), 부채상환능력, 현금흐름, 활동성으로 범주화할 수 있다. Pápić and Pápićová(2023)은 경제위기시에 레버리지, 평상시에는 유동성, 수익성, Son et al.(2019)의 국내 비외감기업에는 유동성, 부채상환능력이 중요하게 나타났다. Pápić and Pápićová(2023)에 의하면, 업종, 기업규모, 지역, 기업형태, 소유구조, 업력의 질적 특성이 중소기업

7) Dastile et al.(2020)은 머신러닝을 사용한 부도 예측 및 금융기관의 신용점수모형 개발과 관련하여 2010~2018년 동안 74개 논문의 체계적 분석(Systematic Analysis)을 수행한 결과 가장 많이 활용된 모형은 로지스틱 회귀모형, 서포트벡터머신, 신경망, 부스팅으로 나타났다.

에 긴요할 수 있다. Moscatelli et al.(2020)의 분석에서는 재무비율과 신용정보를 결합할 때 부도 예측 능력이 현저히 개선되었다. 특히 매출액성장률은 부도와 비선형적 관계를 보이는데 앙상블에서 이 변수가 중요한 변수로 부각된 것은 두 변수간 비선형적 관계를 ML이 잘 포착하는 증거이다.

마지막으로, ML 부도 예측 과정에서 부도 예측 성능과 연관된 주요 쟁점은 불균형 자료의 균형화, 결측값 처리와 변수의 변환, 초매개변수(Hyperparameter)의 조정 방법으로 요약된다. 자료 불균형과 예측 정확성의 관계에 대해 Veganzones and Séverin (2018)은 부도기업의 수가 전체 자료의 20%를 밑돌면 예측 정확성이 하락하는 실험적 결과를 보고하였다. 그리고 서포트벡터머신과 랜덤포레스트가 불균형 자료에도 강건함을 보여주었다. Brown and Mues(2012)의 분석에서는 자료 불균형이 매우 심할 때 랜덤포레스트와 그래디언트 부스팅이 비교적 좋은 성과를 보였다. 자료 균형화 방식으로 과소표집(Undersampling)과 과대표집(Oversampling)이 있다.⁸⁾ Dastile et al.(2020)의 조사 결과를 보면 조사 문헌의 18%만이 자료 균형을 시도했으며 그중에는 과소표집이 주로 활용되었다. Moscatelli et al.(2020)는 무작위 과소표집 방법을 써서 신용정보 대용량 자료를 소용량으로 축소하였다. 과대표집 방법으로는 SMOTE(Synthetic Minority Oversampling Technique)가 대중을 이루었다. Liu et al.(2022)의 분석은 SMOTE의 우수성을 입증하였다.⁹⁾ Kim et al.(2016)은 자료 불균형을 줄이면서 분류의 정확성을 개선하기 위해 최적의 학습용 표본을 구성하는 과소표집 방법으로 군집(Clustering)과 유전자 알고리즘(Genetic Algorithm)의 결합 방법을 제안하였다.

한편, Jones et al.(2015)은 Box-Cox 변수 변환과 특이값 분해(Singular Value Decomposition)를 사용한 결측값 대체가 ML 부도 예측의 정확성에 미치는 효과를 분석하였다. 비록 중소기업에 대한 검증은 아니더라도 종단면과 횡단면 시험용 표본에서 아다부스트와 랜덤포레스트의 예측 성과가 우수한 것으로 나타났다. 초매개변수에 관해서 기존 문헌들은 격자탐색(Grid Search)과 무작위탐색(Random Search) 방법을 선호하지만 Zhang et al.(2023)은 베이지안 최적화(Bayesian Optimization)의 활용을 권고하였다.

8) 과소표집은 표본의 크기가 작아져 모형의 학습시간이 감축되는 장점이 있지만 다수 계급의 본래 분포를 왜곡할 수 있고 자료의 축소로 정보가 손실되는 문제가 있다. 이 방법의 정확성 개선 정도는 표본을 제거하거나 선택하는 규칙이 적절한가에 달려 있다. 과대표집은 부도 예측에 유익한 자료가 사라지는 위험은 없으나 일반화의 오류와 추정 시간 소요라는 문제가 있다.

9) Garcia(2022)는 2010.1월~2018.12월 공익사업과 금융기업을 제외한 1,824개 상장기업(47개 부도기업) 분석에서 SMOTE 방법을 사용할 때 예측 성과가 개선되는 것을 보여주었다.

2. 머신러닝 모형의 개요

1) 통계적 모형

판별분석(LDA)는 같은 집단에 속하는 관측치 사이에 분산은 최소화하고, 다른 집단과의 관측치 사이에는 분산을 최대화하는 방법론이다. 판별분석은 변수들이 정규분포되어야 하고, 변수들의 공분산 행렬이 부도기업과 건전기업 집단에서 같아야 하는 통계적 조건을 만족해야 한다.

Ohlson(1980)이 부도 예측에 최초로 적용한 로지스틱 회귀모형(LR)은 기업 특성과 부도 사이에 선형적 관계를 가정하지만 부도확률과 변수들의 분포에 대한 어떤 사전적 가정도 하지 않는다. 로지스틱 회귀모형의 추정결과는 부도확률로 해석할 수 있는 장점이 있다. 그러나 기업 특성과 부도 사이의 비선형적 관계와 복잡한 상호작용을 고려하기 어려운 한계가 있다.

2) 머신러닝

Quinlan(1986)이 개발한 의사결정트리(DT)는 기업 특성들과 부도 여부에 따라 축차적으로 기업들을 동질의 집단으로 분류하는 규칙의 모음이다. 특성 집합에서 하나의 특성과 그 특성의 값이 결정되면 하위 나뭇가지로 기업들이 분할된다. 마지막 나뭇잎까지 나뭇가지가 뻗어나가면 분리된 각 나뭇잎에 속한 기업 중 부도기업 개수가 부도확률의 추정치로 이용된다. 이 알고리즘은 어떤 분류 변수가 이전 분류 변수에 의해 그 영향력이 결정되기 때문에 변수간 상호작용을 고려할 수 있다. 또한 같은 변수가 여러 번 이용될 수 있어 부도와 비선형관계를 다룰 수 있다. 그러나 바로 이 점 때문에 변수의 작은 변화가 부도확률의 큰 변화를 초래할 수도 있다. 의사결정트리는 과적합과 국부 최적화 가능성이 높아 표본외 예측에서 한계를 드러낸다.

나이브 베이즈(NB)는 베이즈 정리(Bayes' Theorem)를 이용해 분류한다. 나이브 베이즈는 특성들이 서로 독립이라는 가정에서 출발한다. 이 알고리즘은 작업속도가 빠르고 대용량 자료에서 좋은 예측 성능을 보인다. 하지만 특성들이 서로 독립이라는 가정 때문에 특성이 많고 상관성이 높으면 예측 성능의 악화를 초래할 수 있다.

신경망(NN)의 가장 일반적인 형태가 다층퍼셉트론(MLP)으로 모든 층(Layer)과 노드(Node)가 순방향(Feed-Forward)으로 정렬된다. 다층퍼셉트론은 기업 특성 정보를 받는 입력층(Input Layer), 입력된 자료로부터 복잡한 패턴을 찾고 중요한 특징을 추출하는 은닉층(Hidden Layer), 부도와 비부도를 예측하는 출력층(Output Layer)으로 구성된다. 다층퍼셉트론은 보통 하나의 은닉층과 하나의 출력층을 갖는다. 이 알고리즘의 높은 예측 정확성에도 불구하고 복잡한 상호 연결성으로 인해 예측 결과를 해석하기 어려운 단점이 있다.

서포트벡터머신(SVM)은 Cortes and Vapnik (1995)이 제안하였다. 이 방법은 두 집단을 분리하는 초평면(Hyperplane)과 초평면에 가장 가까운 관측치 사이의 거리를 가장 크게 하는 이른바 마진(Margin)을 극대화하는 초평면을 찾는 것이다. 이 방법은 예측함수의 형태를 명시적으로 설정하지 않고 커널함수를 통해 건전기업과 부도기업을 분리한다. 서포트벡터머신은 전역적 유일 최적해를 찾기에 용이하고 차원의 저주를 피할 뿐만 아니라 일반화에 우수한 능력을 지니고 있다. 다만 모형 추정 시간이 오래 걸리는 문제가 있다.

배깅(Bagging)은 전체 자료 중 무작위 붓스트랩(Bootstrap)으로 표집된 자료에 기저 학습기를 적용하고 이 절차를 여러 번 반복한 후 각 학습기의 결과들을 투표하여 가장 많은 득표를 받은 계급(Class)을 예측으로 사용한다. 기저 학습기로는 의사결정트리가 주로 사용된다. 배깅은 대다수 트리의 맨 처음 분류에서 가장 예측력이 높은 변수가 우선 이용된다. 그러면 개별 트리들이 서로 비슷할 것이고 결국 예측의 상관성이 높다.

Breiman(2001)은 배깅의 이와 같은 문제를 해결하기 위해 랜덤포레스트(RF)를 개발하였다. 랜덤포레스트는 부표본에 대해 분류 함수를 반복적으로 생성시키는 것은 배깅과 같지만 트리의 각 노드에서 특성의 집합도 무작위로 선택하는 점에서 배깅과 다르다.

Freund and Schapire (1997)가 이진(Binary) 분류 문제에 아다부스트(AdaBoost)를 최초로 적용하였다. 이 방법은 앞에서 예측한 오차에 가중치를 더 높이면서 순차적으로 오차를 보정해 나가면서 규칙을 보완한다. 아다부스트는 최종 노드의 수가 2개인 가장 간단한 의사결정트리인 그루터기(Stump)를 약한 학습기(Weak Learner)로 사용한다. 약한 학습기란 무작위 예측보다 약간 좋은 예측 능력을 지닌 모형을 뜻한다. 이 방법은 간단하고 효과적인 방법으로 매우 높은 정확성을 달성할 수 있고 과적합이 상대적으로 적다. 하지만 대용량 자료에서 학습 시간이 많이 소요될 수 있다.

XGBoost는 그래디언트 부스팅(GB) 기반의 알고리즘으로 Chen and Guestrin (2016)이 개발하였다.¹⁰⁾ 이 방법은 과적합 방지를 위해 규제(Regularization)를 추가하고 속도를 높이기 위해 여러 CPU 코어에서 병렬처리를 지원한다. XGBoost는 트리가 최대 깊이에 도달할 때까지 성장시킨 후 가지치기(Pruning)를 통해 트리를 최적화한다. XGBoost는 대용량 자료에서 효율적이다. 알고리즘은 시작 단계에서 모든 관측값에 대해 같은 예측값을 할당하고, 매 부스팅 단계에서 잔차에 대해 학습하고 이전 단계의 오류를 줄이는 방향으로 트리가 추가된다.

LightGBM은 마이크로소프트에서 개발한 GB 알고리즘으로 대용량 자료에서 XGBoost보다 더 빠

10) 그래디언트 부스팅(GB)은 직전 학습의 예측오차에 새로운 학습기를 학습시킨다. GB는 제로원(0-1) 손실함수를 최소화하기 위해 기울기 강하(Gradient Descent)를 이용하여 최적의 해를 찾는다. GB는 너무 많이 반복하면 과적합이 발생하기 때문에 수렴할 때까지 반복하지 않고 일정한 수로 반복하는 축소추정을 수행한다. GB는 예측 정확성이 높고, DT를 기저 학습기로 이용하기 때문에 범주형 변수와 연속형 변수를 모두 고려할 수 있는 유연성이 높은 장점이 있다.

른 학습 속도와 효율적인 메모리 사용을 제공하는 것이 특징이다. LightGBM은 깊이 중심 트리 분리 방식이 아닌 전체 손실을 최대한 감소시키는 나뭇잎 중심 우선 분할 방식을 채택한다. 이 때문에 모형의 해석이 복잡할 수 있지만 대용량 자료 처리 능력에 강점을 지닌다.

CatBoost는 Yandex가 개발한 GB 알고리즘이다. 이 알고리즘은 특히 범주형 자료에 대해 별도의 전처리나 더미변수 처리를 하지 않아도 직접 처리할 수 있다. CatBoost는 대칭 트리(Symmetric Trees)를 사용하여 모형을 구축하기 때문에 계산 효율성이 높고 모형의 복잡성을 줄일 수 있다.

III. 분석 방법

본 분석은 대출부도 예측 결과의 일반화(Generalization)를 제고하기 위해 방법론적으로 네 가지에 주안점을 두었다. 첫째, 실증분석의 포괄성이다. 본 분석은 균형 표본과 불균형 표본에 대해 AUC 등 다수의 분류 정확성 계량지표로 총 12개 모형의 성능을 평가하였다. 특히 선행연구들이 예측 성능적 차이에 대한 통계적 검증이 미흡하여 본 분석에서는 이에 대한 가설검증을 수행하였다. 둘째, 대용량 분석 표본을 활용한다. 본 분석은 신용보증기금이 보유한 방대한 양의 중소기업 재무 자료와 대출부도 자료를 활용하였다.¹¹⁾ 대출부도는 보증부 대출 기업이 은행에 이자지급과 원금상환 등 채무계약을 제때 이행하지 못할 때 발생한다. 대용량 표본을 사용하면 ML이 더 많은 자료를 학습할 수 있어 부도 예측의 정확성을 높이고 변동성은 줄임으로써 예측 결과의 신뢰성을 높일 수 있다. 셋째, 분석 변수의 다양성이다. 소수의 재무적 특성으로 미리 한정하지 않고 재무 범주별 주요 재무비율과 기업체 고유 속성 변수를 예측변수로 사용하였다. 창업기업의 경우 소유와 경영의 분리가 분명치 않아 경영자의 신용정보가 기업 부도와 연관성이 있겠으나 본 분석에서는 기업 속성에 집중하여 개인 신용정보는 포함하지 않았다. 넷째, 결측값과 극단값 처리 등 모형 추정을 위한 전처리 과정은 각각의 사안마다 여러 가지 선택 대안이 있지만 선행연구에서 대중을 이루는 방법이거나 필자가 적당하다고 판단한 방법을 이용하였다. 그 중에서 특히 ML마다 초매개변수값은 교차타당성 검증으로 과적합을 철저히 통제하여 결정하였다.

본 연구의 기본 자료는 신용보증기금에서 입수한 2012~2020년말 보증기업 중 결산기가 12월인 기업의 약 82만개 재무제표이고, 이중 자산총액 600억원 미만은 중소기업으로 정의하였다. 이질적인 재무제표를 갖는 금융보험업은 분석에 포함하지 않았다. 창업기업을 구분하는 기준은 기업 설립일 이후 상기 연도 말일까지의 기간이 5년 미만이면 창업기업, 그 이상은 일반기업으로 정의하였다. 국내 법령

11) 신용보증기금은 중소기업이 은행으로부터 대출을 받을 때 기업의 신용도를 평가하고, 이를 기초로 신용보강을 통해 중소기업의 자금 융통을 원활하게 해주는 역할을 한다. 중소기업이 은행에 대출채무를 상환하지 못할 때 신용보증기금이 대신 채무를 변제한다. 중소기업 금융지원의 오랜 전통을 지닌 신용보증기금은 방대한 중소기업의 재무 자료와 부도 자료를 보유한 중소기업 전문 금융지원 기관이다.

은 창업 후 7년까지 창업기업으로 정의하고 있으나,¹²⁾ 이는 창업기업을 우대하려는 정책적 측면에서 정의된 것으로 보는 것이 합당하다. 본시 창업기업에 대한 정책금융 지원의 논리적 근거는 기업 생애주기에서 죽음의 계곡(Death Valley)이라 불리는 창업 초기에 일시적 자금경색으로 어려움을 겪는 기업의 대출부도를 사전에 예방하여 지속적인 성장을 도모하는 데 있다. 김정숙 등(2014)에 의하면, 창업기업의 부도확률이 감소세로 전환되는 시점이 창업 후 3~5년이였다. 이 논리적 연장선에서 죽음의 계곡을 지나면 안정적 성장 단계로 진입하기 때문에 설립 후 5년까지를 창업기업으로 정의하였다.

자료의 처리 결과 창업기업은 195,916개, 일반기업은 618,171개였다. 대출부도는 신용보증기금의 부도 사유를 준용하여 원금 및 이자 연체, 휴폐업 등으로 정의하였다. 결산이 공표되는 4월부터 이듬해 3월말까지를 부도 관측기간 1년으로 정의하고, 이 기간 내에 부도 사유가 발생한 기업을 1년 부도기업으로 식별하였다. 그리고 Párik and Papíková(2023), Mselmi et al.(2017)에 따라 재무정보의 장기 부도 예측 효과를 평가하고자 부도 관측기간을 1년씩 연장하여 2년 부도, 3년 부도로 정의하였다.

〈표 1〉에 분석 기초자료를 제시하였다. 총 부도기업 수는 51,210개로 이 중 창업기업이 15,110개, 일반기업이 36,100개이다. 그리고 1년 부도기업은 창업기업 4,445개, 일반기업 11,916개, 2년 부도기업은 창업기업 5,913개, 일반기업 13,522개, 3년 부도기업은 창업기업 4,752개, 일반기업 10,662개이다. 부도율 수준은 창업기업이 일반기업보다 높게 나타났다.

〈표 1〉 분석 자료

구분	전체 기업	1년 부도기업		2년 부도기업		3년 부도기업		(계)	
		개수	부도율 ¹⁾	개수	부도율	개수	부도율	개수	부도율
창업기업	195,916	4,445	2.27%	5,913	3.02%	4,752	2.43%	15,110	7.71%
일반기업	618,171	11,916	1.93%	13,522	2.19%	10,662	1.72%	36,100	5.84%
(계)	814,087	16,361	2.01%	19,435	2.39%	15,414	1.89%	51,210	6.29%

주1) 부도율은 부도기업의 개수를 전체기업의 개수로 나눈 값이다.

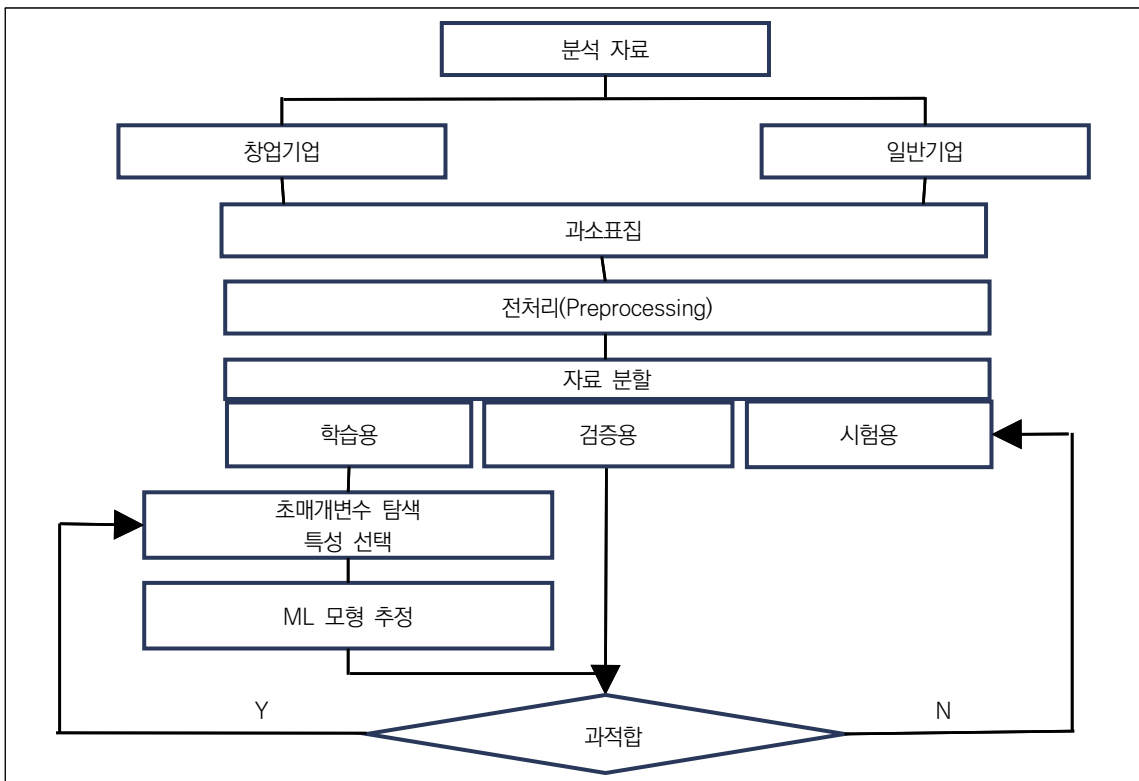
본 논문은 〈그림 1〉과 같은 절차로 분석을 진행하였다. 〈표 1〉의 분석 자료에서 볼 수 있듯이 연차별 부도기업의 수가 전체에서 약 2%~3%를 차지할 정도로 자료가 매우 불균형적이다. 본 분석은 이에 Moscatelli et al.(2020), Barboza et al.(2017), Mselmi et al.(2017) 등의 분석에서와 같이 건전기업의 수를 부도기업의 수와 같게 하는 과소표집 방식으로 균형 표본을 생성하여 분석의 기본자료로 이

12) 중소기업창업원 지원법 제2조는 중소기업을 창업하여 사업을 개시한 날부터 7년이 지나지 아니한 기업(법인과 개인사업자를 포함한다)으로 창업기업을 정의한다.

용하였다. 본 논문에서는 ML 모형 학습 과정에서 특정 계급에 치우치지 않고 부도와 비부도의 두 계급에 대해 같은 중요도를 할당하여 모형의 예측 편향을 줄이고자 하였다. 다만 과소표집은 모형의 일반화 문제가 제기될 수 있다. 그래서 건전기업을 부도기업의 3배수로 표집하는 불균형 표본을 별도로 구성하여 균형 표본에서의 ML 모형의 예측 성과가 불균형 표본에서도 강건하게 유지되는지를 확인하였다.

건전기업을 표집할 때 여러 개의 재무제표를 보유한 부도기업은 부도로 인식된 결산 재무제표 외 모든 재무제표는 건전기업으로 표집되지 않도록 대상에서 제외하였다. 그리고 한국표준산업분류 대분류 기준으로 부도기업의 산업별 분포와 동일한 표본분포를 형성하기 위해 층화임의표집(Stratified Random Sampling) 방식으로 건전기업을 표집하였다.

〈그림 1〉 대출부도 예측 절차



〈표 2〉에 창업기업과 일반기업의 부도기업 개수를 표시하였다. 〈표 2〉를 통해서 보면 창업기업 1,820개(1년 473개, 2년 686개, 3년 661개), 일반기업 3,706개(1년 1,181개, 2년 1,372개, 3년 1,153개)가 2021년에 부도로 식별된 기업이고, 여기에 동일 연도의 건전기업 자료를 표집한 것이 2020년 시험용 표본에 해당한다. 학습용과 검증용 표본은 Jones et al.(2015)의 분석에 따라 학습용 70%, 검증용 30%로 분리하였다.

대출부도 예측을 위한 재무적 특성은 Mselmi et al.(2017), Jones et al.(2015)의 분석과 같이 재무 범주별로 기본적 재무비율을 선택하는 방식으로 정의하였다. 〈표 3〉에 표시된 바와 같이 재무비율은 성장성 5개, 수익성 7개, 안정성 4개, 부채상환능력 6개, 효율성 3개, 활동성 4개, 유동성 4개, 현금흐름 2개로 총 35개이다. 재무변수 외에도 기업의 질적 특성 변수들로 Moscatelli et al.(2020)이 사용한 지역과 산업, 기업규모-총자산과 매출액-을 입력변수로 고려하였다. 산업은 17개 대분류와 이를 제조업/건설업/도소매서비스업으로 재정의한 변수, 기업체 소재 지역은 17개 광역 행정 구역으로 구분하였다. 그리고 총자산과 매출액은 8개 범주로 구간화하였다.¹³⁾ 변수의 변환은 이상치의 영향을 줄이며, ML 모형이 재무변수의 특정 값 주변에서 지나치게 적합되는 것을 방지하고, 비선형 관계의 포착에 긍정적인 효과가 있다. 또한 ML 모형과 예측 성과를 비교하기 위한 준거모형으로서 통계적 모형을 추정할 때 이상치의 영향을 감소시킨다. 이 변수들은 Pápić and Papíková(2023)의 분석에 따라 더미변수화(One-Hot Encoding) 하였다. 외부감사 더미변수와 법인기업 더미변수를 포함해서 총 90개가 입력변수로 이용되었다.

분석자료의 전처리 단계에서 재무비율이 결측값인 경우 Jones et al.(2015)이 사용한 방식에 의거 특이값 분해를 통해 결측값을 대체하였다. 특이값 분해는 창업기업과 일반기업 각각 표본(3개)/총자산(8개)/산업(3개) 범주 조합에 대해 수행하였다.¹⁴⁾ 재무비율의 극단값들은 창업기업과 일반기업 각 3개 부도연차에 대해 Son et al.(2019)의 극단값 처리 방법에 따라 상하 5%에서 조정하였다(Winsorizing). 그리고 재무비율은 창업기업과 일반기업의 학습용, 검증용, 시험용 자료에 대해 최소-최대 정규화(Normalization)를 수행하였다. 이 방법은 재무비율의 최소값과 최대값을 사용하여 자료를 구간 [0,1] 범위로 조정하는 것이다.¹⁵⁾ 초매개변수는 ML 모형의 일반적 구조를 정의하는 것으로 그것의 결정에 따라 ML 모형의 성능은 큰 영향을 받는다. 그래서 Moscatelli et al.(2020)의 분석처럼

13) 총자산과 매출액은 5억원미만/10억원미만/20억원미만/30억원미만/50억원미만/100억원미만/300억원미만/300억원이상의 8개로 구간화하였다.

14) 특이값 분해는 직교(Orthogonal) 선형변환으로 자료의 분산을 최적으로 유지한다. Son et al.(2019)은 결측값을 0으로, Pápić and Papíková(2023)은 평균으로 대체하였다.

15) Liang et al.(2016)은 정규화, Pápić and Papíková(2023)은 표준화(Standardization), Jones et al.(2015)과 Son et al.(2019)은 Box-Cox 변환 방식을 사용하였다.

학습용 자료를 이용해 격자탐색 방법으로 5-겹 교차타당성을 수행하여 최적의 초매개변수를 탐색하였다.

〈표 2〉 창업기업과 일반기업의 연도별 부도기업수

연도	창업기업				일반기업			
	건전기업 개수	부도기업 ¹⁾ 개수			건전기업 개수	부도기업 개수		
		1년	2년	3년		1년	2년	3년
2012년	15,433	168	617	579	49,811	677	1,640	1,475
2013년	16,172	391	568	576	50,805	1,249	1,537	1,598
2014년	17,878	394	673	632	55,895	1,345	1,787	1,546
2015년	20,341	527	766	754	64,598	1,535	1,820	1,857
2016년	21,561	587	857	838	65,416	1,483	1,933	1,696
2017년	24,173	708	983	712	70,480	1,572	1,902	1,337
2018년	26,047	641	763	661	78,566	1,494	1,531	1,153
2019년	27,338	556 ²⁾	686	-	88,523	1,380	1,372	-
2020년	26,973	473 ²⁾	-	-	94,077	1,181	-	-
(계)	195,916	4,445	5,913	4,752	618,171	11,916	13,522	10,662

주1) 부도기업 1년은 연도+1년, 부도기업 2년은 연도+2년, 부도기업 3년은 연도+3년으로 예컨대 창업기업 2018년 3년 661개는 2018년말 건전기업 중 2021년에 부도가 발생한 기업수이다.

주2) 굵은 글씨는 2020년 시험용 자료, 기울인 글씨는 2019년 시험용 자료로 이용된 부도기업 개수이다.

〈표 3〉 분석 변수

범주	변수명	산출식 ¹⁾
성장성	Gro1 매출액증가율	$(\text{매출액}_t - \text{매출액}_{t-1}) / \text{매출액}_{t-1}$
	Gro2 총자산증가율	$(\text{자산총계}_t - \text{자산총계}_{t-1}) / \text{자산총계}_{t-1}$
	Gro3 매출채권증가율	$((\text{매출채권}_t - \text{매출채권대손충당금}_t + \text{공사미수금}_t + \text{분양미수금}_t) - (\text{매출채권}_{t-1} - \text{매출채권대손충당금}_{t-1} + \text{공사미수금}_{t-1} + \text{분양미수금}_{t-1})) / (\text{매출채권}_{t-1} - \text{매출채권대손충당금}_{t-1} + \text{공사미수금}_{t-1} + \text{분양미수금}_{t-1})$
	Gro4 재고자산증가율	$(\text{재고자산}_t - \text{재고자산}_{t-1}) / \text{재고자산}_{t-1}$
	Gro5 영업이익증가율	$(\text{영업이익}_t - \text{영업이익}_{t-1}) / \text{영업이익}_{t-1}$
수익성	Pro1 총자산순이익율	$\text{당기순이익}_t / (\text{자산총계}_t + \text{자산총계}_{t-1}) / 2$
	Pro2 총자산영업이익율	$\text{영업이익}_t / (\text{자산총계}_t + \text{자산총계}_{t-1}) / 2$
	Pro3 자기자본순이익율	$\text{당기순이익}_t / (\text{자본총계}_t + \text{자본총계}_{t-1}) / 2$
	Pro4 자기자본영업이익율	$\text{영업이익}_t / (\text{자본총계}_t + \text{자본총계}_{t-1}) / 2$
	Pro5 매출액총이익율	$\text{매출총이익}_t / \text{매출액}_t$
	Pro6 매출액영업이익율	$\text{영업이익}_t / \text{매출액}_t$
	Pro7 매출액순이익율	$\text{당기순이익}_t / \text{매출액}_t$
안정성	Lev1 부채비율	$\text{부채총계}_t / \text{자본총계}_t$
	Lev2 자기자본비율	$\text{자본총계}_t / \text{부채와자본총계}_t$
	Lev3 차입금의존도	$\text{차입금}_t / \text{부채와자본총계}_t$
	Lev4 단기차입금의존도	$(\text{단기차입금}_t + \text{단기금융리스부채}_t + \text{유동성장기부채}_t) / \text{자산총계}_t$
부채상환능력	Dcr1 EBITDA이자보상비율	$(\text{영업이익}_t + \text{감가상각비}_t + \text{무형자산상각비}_t) / (\text{이자비용}_t + \text{매출채권처분손실}_t)$
	Dcr2 이자보상비율	$\text{영업이익}_t / (\text{이자비용}_t + \text{매출채권처분손실}_t)$
	Dcr3 금융비용/매출액	$(\text{이자비용}_t + \text{매출채권처분손실}_t) / \text{매출액}_t$
	Dcr4 부채상환계수	$(\text{영업이익}_t + \text{감가상각비}_t + \text{무형자산상각비}_t) / (\text{차입금}_t + \text{이자비용}_t + \text{매출채권처분손실}_t)$
	Dcr5 차입금/매출액	$\text{차입금}_t / \text{매출액}_t$
	Dcr6 영업현금흐름/차입금	$\text{영업현금흐름}_t / \text{차입금}_t$
효율성	Eff1 총자본투자효율	$\text{부가가치}_t / (\text{부채와자본총계}_t + \text{부채와자본총계}_{t-1}) / 2$ *부가가치=법인세비용차감전계속사업이익+급여+퇴직급여+복리후생비+임차료+감가상각비+세금과공과+이자비용+매출채권처분손실-이자수익+노무비+중기임차료
	Eff2 부가가치율	$\text{부가가치}_t / \text{매출액}_t$
	Eff3 설비투자효율	$\text{부가가치}_t / ((\text{유형자산}_t - \text{건설중인자산}_t) + (\text{유형자산}_{t-1} - \text{건설중인자산}_{t-1})) / 2$
활동성	Tov1 총자산회전율	$\text{매출액}_t / (\text{자산총계}_t + \text{자산총계}_{t-1}) / 2$
	Tov2 매출채권회전율	$\text{매출액}_t / ((\text{매출채권}_t - \text{매출채권대손충당금}_t + \text{공사미수금}_t + \text{분양미수금}_t) + (\text{매출채권}_{t-1} - \text{매출채권대손충당금}_{t-1} + \text{공사미수금}_{t-1} + \text{분양미수금}_{t-1})) / 2$
	Tov3 재고자산회전율	$\text{매출액}_t / (\text{재고자산}_t + \text{재고자산}_{t-1}) / 2$
	Tov4 매입채무회전율	$\text{매출액}_t / ((\text{매입채무}_t + \text{공사선수금}_t + \text{분양선수금}_t) + (\text{매입채무}_{t-1} + \text{공사선수금}_{t-1} + \text{분양선수금}_{t-1})) / 2$
유동성	Liq1 유동비율	$\text{유동자산}_t / \text{유동부채}_t$
	Liq2 운전자본비율	$(\text{유동자산}_t - \text{유동부채}_t) / \text{자산총계}_t$
	Liq3 당좌비율	$\text{당좌자산}_t / \text{유동부채}_t$
	Liq4 현금비율	$(\text{현금및현금성자산}_t + \text{단기금융상품}_t) / \text{유동부채}_t$
현금흐름	Cfr1 EBITDA/자기자본	$(\text{영업이익}_t + \text{감가상각비}_t + \text{무형자산상각비}_t) / \text{자본총계}_t$
	Cfr2 EBITDA/매출액	$(\text{영업이익}_t + \text{감가상각비}_t + \text{무형자산상각비}_t) / \text{매출액}_t$

주1) 부채상환계수와 활동성 변수들을 제외한 모든 재무비율은 백분율(%)로 계산하였다.

〈표 4〉에 ML 모형의 특성 선택 방법을 표시하였다. 과적합을 예방하고 ML의 성능을 제고하기 위해 특성 선택 과정이 선행되어야 한다. 특성 선택은 필터(Filter), 래퍼(Wrapper), 임베드(Embedded) 방법이 있다.¹⁶⁾ 본 분석은 12개 모형마다 적절한 특성 선택 방법을 적용하여 30개의 입력변수를 선택하였다.¹⁷⁾

12개 모형 추정 및 검증은 파이썬(Python) 버전 3.9.17에서 사이킷런(Scikit-Learn) 등의 패키지를 활용하고 필요한 경우 프로그래밍하였고, AUC와 혼동행렬 계량지표의 통계적 가설검증은 R 버전 4.4.0의 관련 패키지를 사용하였다.

〈표 4〉 특성 선택 방법

모형	방법	세부 방법	추정방법 ¹⁾
LDA	래퍼	축차특성제거(Recursive Feature Elimination)	LDA
LR	래퍼	축차특성제거	LR
DT	임베드	특성 중요도(Feature Importance)	-
NB	래퍼	축차특성제거	DT
MLP	래퍼	축차특성제거	DT
SVM	필터	ANOVA F검정	-
Bagging	래퍼	축차특성제거	DT
RF	임베드	특성 중요도	-
AdaBoost	임베드	특성 중요도	-
XGBoost	임베드	특성 중요도	-
LightGBM	임베드	특성 중요도	-
CatBoost	임베드	특성 중요도	-

주1) 추정방법은 축차특성제거에 사용된 방법을 의미한다.

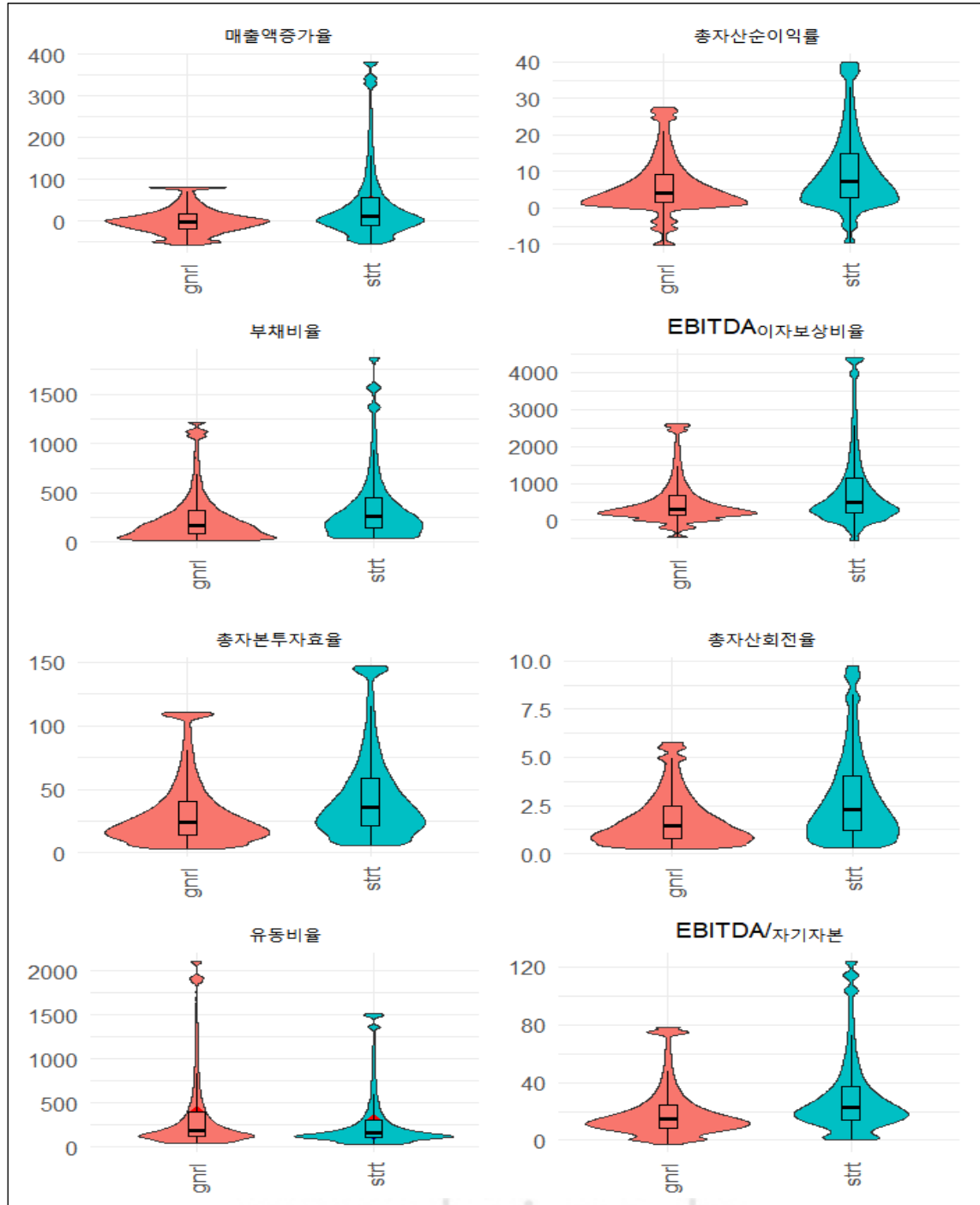
16) 필터는 특성의 단변량 분석에 기반하고, 래퍼는 사전에 정의한 정확성 평가지표를 이용하는 방식으로 단계적(Stepwise) 선택법이 있다. 임베드는 모형에 내재된 방식을 사용하는데 특성을 규제하는 LASSO(Least Absolute Shrinkage and Selection Operator)가 한 예이다.

17) Moscatelli et al.(2020)은 24개 재무비율을 포함한 38개 변수 중에서 3단계 절차를 거쳐 24개 변수를 선택하였다. Mselmi et al.(2017)은 41개 재무비율 중 단계적 회귀방식으로 10개 이내의 비율을 사용하였다.

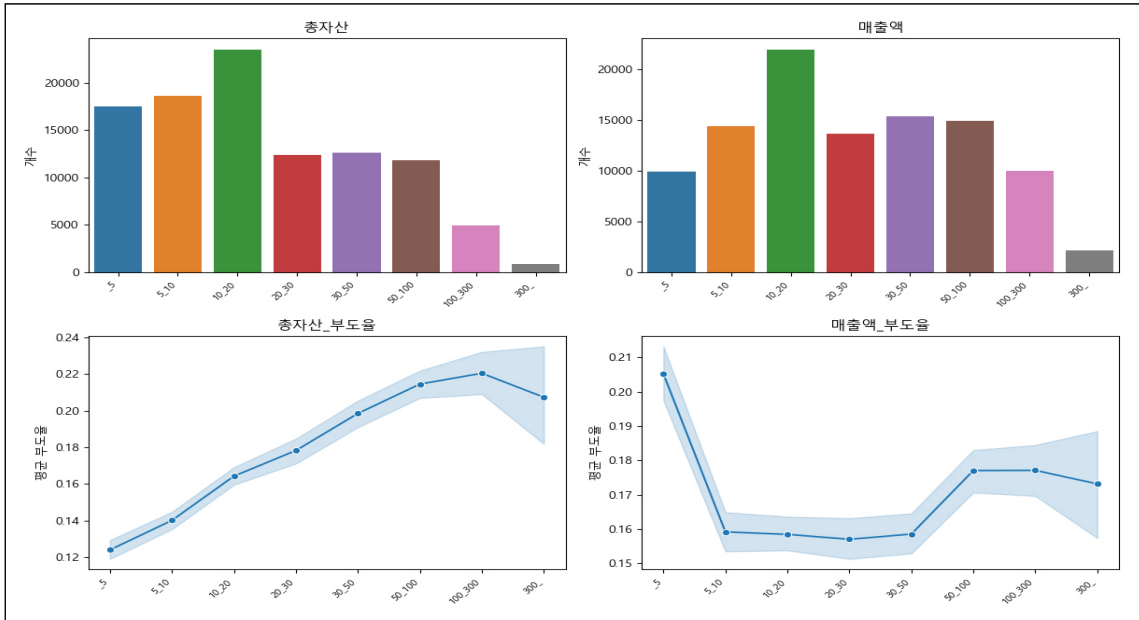
IV. 기초 분석

부도기업과 건전기업의 균형 표본 102,420개 자료로 재무비율의 분포를 살펴보기 위한 바이올린(Violin) 그래프는 <그림 2>에 제시되었다. 이 도표는 상자도표(Boxplot)와 밀도함수를 결합한 그래프이다. 범주별 대표 재무비율을 중앙값으로 살펴보면, 창업기업이 유동비율(Liq1)을 예외로 하고 매출액증가율(Gro1), 총자산순이익률(Pro1), 부채비율(Lev1), EBITDA이자보상비율(Dcr1), 총자본투자효율(Eff1), 총자산회전율(Tov1), EBITDA/자기자본(Cfr1)이 일반기업보다 높게 나타났다. 창업기업의 재무비율 분포가 일반기업의 재무비율 분포에 비해 더 편향된 분포를 보이고 극단값들이 훨씬 더 많이 존재하는 것을 알 수 있다. 이는 창업기업 경영활동의 역동성이 반영된 결과로 보인다. 예컨대 창업기업의 매출액증가율 범위가 일반기업의 4배 이상이며 극단값들도 다수 발견된다.

<그림 3>은 바이올린 그래프에서 이용된 전체 균형 표본으로 총자산과 매출액의 각 8개 범주별 1년 부도율을 추정한 결과이다. 총자산과 매출액은 일반적으로 기업규모의 대용변수(Proxy)이다. 두 변수의 상관계수는 0.651로 높은 편이나 완전한 선형관계에 가깝지는 않다. <그림 3>에서 총자산과 매출액의 부도율은 다른 패턴을 보인다. 총자산은 커질수록 부도율이 증가하고, 매출액은 부도율과 비선형적 관계가 나타났다. 총자산과 부도율은 음(-)의 상관을 기대할 수 있다. 그러나 <그림 2>에서 사용된 8개 주요 재무비율의 총자산과 매출액 범주별 중앙값을 살펴본 결과 자산규모가 커질수록 재무적 성과가 저조한 것으로 확인되었다. 기업경영 성과가 좋지 않은 기업군은 부실기업이 다수 분포하여 부도율이 높다. 매출액의 경우 총자산순이익률(Pro1), EBITDA이자보상비율(Dcr1), 총자본투자효율(Eff1)이 매출액 수준과 역의 U자 형태를 보였다. 특히 매출액이 5억원을 초과하면 세 변수값이 큰 폭으로 증가하였다. 매출액 중위구간(5억원~50억원)의 낮은 부도율은 기업들의 양호한 경영 성과에 기인한다. 또한 5억원미만 기업군에서 매출액증가율(Gro1)이 약 -15%로 나타나 매출액이 적은 기업들의 판매부진이 재무적 성과를 악화시키는 원인이 되고 있다.

〈그림 2〉 주요 재무비율의 바이올린 그래프¹⁾

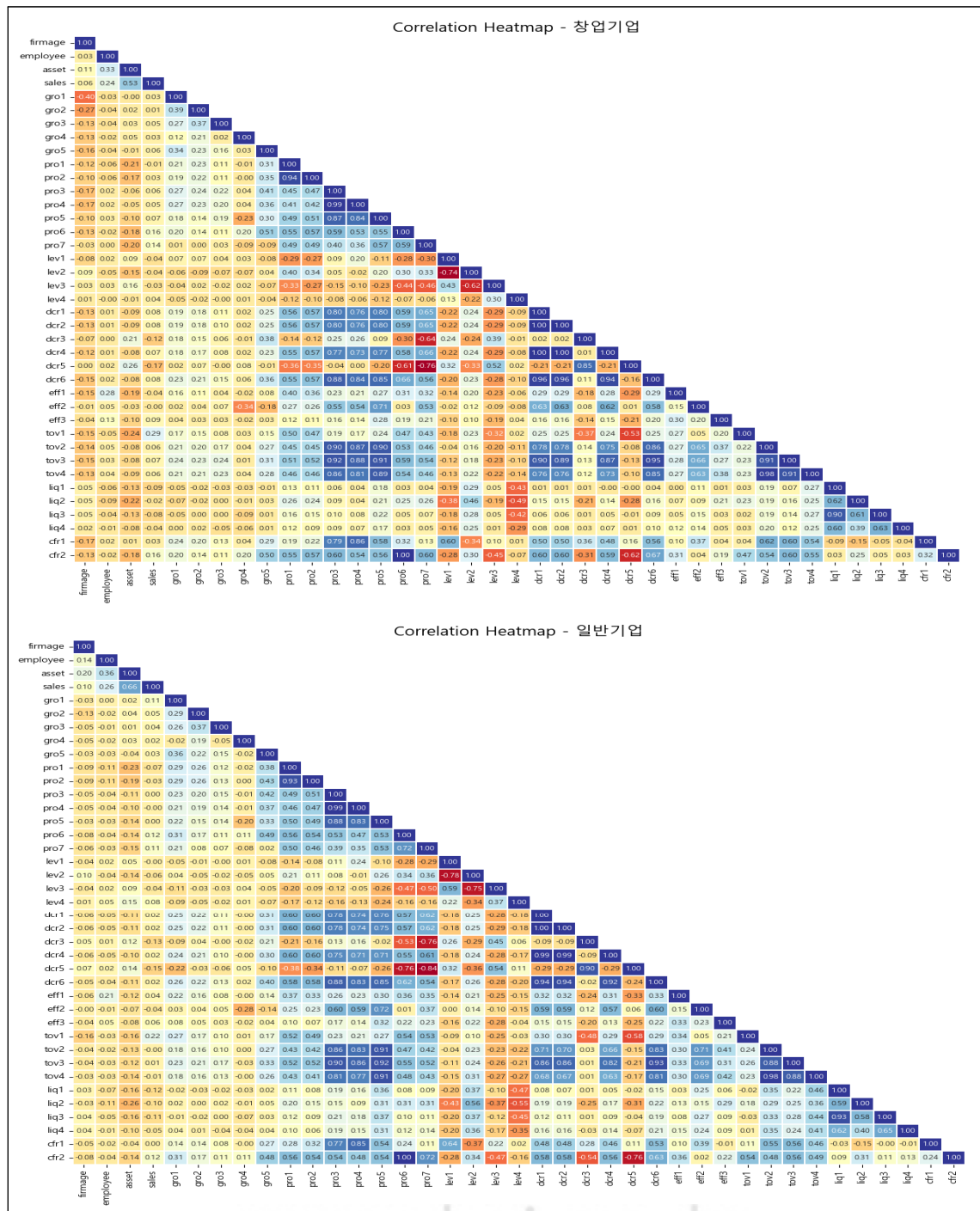
주1) gnrl은 일반기업, strt는 창업기업을 나타낸다. 단위는 총자산회전율을 제외하고 모두 백분율(%)이다.

〈그림 3〉 총자산과 매출액의 부도율¹⁾

주1) 총자산과 매출액 구간의 단위는 억원이다. 부도율은 구간별 1년 추정 부도율을 표시한 것이다. 평균 부도율은 붓스트랩(Bootstrap) 평균을 의미하고 음영은 95% 신뢰구간이다.

〈그림 4〉는 창업기업과 일반기업에 대한 재무변수들의 상관계수 열지도(Heatmap)이다. 창업기업과 일반기업의 상관계수는 대체로 비슷한 양상을 보이지만 상관관계의 강도는 일반기업이 약간 높다. 같은 범주내 재무비율 사이의 상관성이 비교적 높다. 성장성 변수들은 여타 범주의 변수와의 상관관계가 낮게 나타난다. 특히 수익성, 부채상환능력, 현금흐름, 활동성 변수끼리의 상관관계가 매우 높은 편이다. 이 범주의 변수들이 주로 수익과 현금흐름으로 측정되는 데서 비롯된다. 창업기업에서 업력(Firmage)과 매출액증가율(Gro1)의 상관계수가 -0.40으로 비교적 높게 나타난 것은 창업 이후 시간이 지나면서 성장률이 하락하는 매출액 성장세의 기저효과이다.

한편 부도기업과 건전기업간 재무변수의 평균에 대한 t-검증 결과는 〈표 5〉에 표시하였다. t-검증에는 1년 부도 균형 표본이 사용되었다. 일반기업의 경우 설비투자효율만 예외로 하고 예상대로 성장성, 수익성, 안정성, 부채상환능력, 효율성, 활동성, 현금흐름에서 건전기업이 부도기업보다 양호한 재무적 상태를 보이면서 통계적으로 유의적인 평균적 차이를 나타냈다. 하지만 창업기업은 현금비율을 제외한 유동성 변수들이 부도기업과 건전기업간 평균의 차이가 없다는 귀무가설을 5% 유의수준에서 기각하지 못하였다. 다른 재무변수들의 평균에 비추어 대체로 부도기업의 재무적 성과는 건전기업의 재무적 성과보다 낮았다.

〈그림 4〉 상관계수 열지도¹⁾

〈표 5〉 재무변수의 t-검증¹⁾

변수명	창업기업				일반기업			
	건전	부도	t값	유의확률	건전	부도	t값	유의확률
매출액증가율(Gro1)	45.3	35.4	5.048	0.000	4.3	-4.3	20.250	0.000
총자산증가율(Gro2)	50.4	51.5	-0.724	0.469 [§]	9.8	7.0	9.637	0.000
매출채권증가율(Gro3)	51.4	57.5	-2.217	0.027	14.0	12.1	2.365	0.018
재고자산증가율(Gro4)	59.9	62.6	-0.819	0.413 [§]	14.3	12.4	2.061	0.039
영업이익증가율(Gro5)	35.7	16.6	5.753	0.000	-1.0	-25.8	18.056	0.000
총자산순이익율(Pro1)	11.2	7.0	18.198	0.000	7.1	3.4	37.097	0.000
총자산영업이익율(Pro2)	12.5	8.7	15.566	0.000	8.5	5.2	31.174	0.000
자기자본순이익율(Pro3)	8.5	6.5	11.296	0.000	5.7	3.8	26.843	0.000
자기자본영업이익율(Pro4)	11.8	9.5	9.761	0.000	7.9	5.6	23.900	0.000
매출액총이익율(Pro5)	2.6	1.6	14.018	0.000	2.0	1.2	27.640	0.000
매출액영업이익율(Pro6)	-1.7	-8.1	15.865	0.000	-5.2	-11.6	32.311	0.000
매출액순이익율(Pro7)	-0.9	-4.6	17.299	0.000	-1.8	-4.9	31.572	0.000
부채비율(Lev1)	378.4	464.7	-9.067	0.000	243.6	311.2	-17.540	0.000
자기자본비율(Lev2)	32.8	28.0	12.621	0.000	42.7	36.1	24.673	0.000
차입금의존도(Lev3)	43.3	49.4	-12.930	0.000	39.5	45.6	-22.008	0.000
단기차입금의존도(Lev4)	16.8	18.5	-3.941	0.000	15.1	19.7	-18.621	0.000
EBITDA이자보상비율(Dcr1)	1009.2	585.1	19.061	0.000	657.5	305.3	42.092	0.000
이자보상비율(Dcr2)	933.3	538.4	19.109	0.000	607.0	279.3	42.146	0.000
금융비용/매출액(Dcr3)	1.5	1.9	-9.032	0.000	1.3	1.7	-15.783	0.000
부채상환계수(Dcr4)	2.5	1.4	19.217	0.000	1.6	0.7	42.110	0.000
차입금/매출액(Dcr5)	36.7	51.3	-13.789	0.000	37.2	51.1	-24.261	0.000
영업현금흐름/차입금(Dcr6)	22.5	14.1	17.147	0.000	15.1	7.7	37.981	0.000
총자본투자효율(Eff1)	48.6	39.8	11.966	0.000	34.8	27.8	20.604	0.000
부가가치율(Eff2)	4.1	3.3	7.407	0.000	3.7	3.1	13.504	0.000
설비투자효율(Eff3)	1263.0	1199.9	1.329	0.184 [§]	1049.2	1122.0	-2.707	0.007
총자산회전율(Tov1)	2.9	2.4	10.351	0.000 [§]	1.9	1.5	22.776	0.000
매출채권회전율(Tov2)	2.9	2.3	11.588	0.000	2.2	1.5	25.090	0.000
재고자산회전율(Tov3)	14.3	10.0	14.327	0.000	10.0	6.1	30.483	0.000
매입채무회전율(Tov4)	13.5	10.4	11.527	0.000	10.6	7.3	24.753	0.000
유동비율(Liq1)	304.2	307.2	-0.380	0.704 [§]	427.5	382.6	6.745	0.000
운전자본비율(Liq2)	26.8	26.4	0.662	0.508 [§]	31.9	30.2	4.369	0.000
당좌비율(Liq3)	219.6	210.0	1.731	0.084 [§]	324.5	283.7	7.786	0.000
현금비율(Liq4)	43.7	25.4	15.369	0.000	60.8	32.6	26.980	0.000
EBITDA/자기자본(Cfr1)	32.8	30.7	3.242	0.001 [§]	21.1	18.7	9.875	0.000
EBITDA/매출액(Cfr2)	-1.4	-7.5	15.965	0.000	-4.8	-10.8	32.424	0.000

주1) § 표시가 없는 비율은 건전과 부도의 이분산(Heteroscedastic)으로 Welch t-검증을 수행하였다.

V. 대출부도 예측 결과

금융기관의 신용위험 관리를 위한 시계(Time Horizon)는 일반적으로 1년이다. 예상손실의 측정, 대출금리의 결정, 대손준비금의 계상, 그리고 자기자본비율의 산출 등에 가장 기본적 요소가 1년 부도 확률(Probability of Default)이다. 이에 본 분석은 1년 대출부도 예측에 일차적 관심이 있다. <표 6>~<표 9>는 중소기업 균형 표본의 1년 대출부도 예측 결과이다. <표 6>과 <표 7>에는 창업기업과 일반기업의 학습용/검증용/시험용 자료에 대해 통계적 모형과 ML 모형의 예측 정확성 계량지표값이 기록되어 있다. <표 6>은 ML의 1년 대출부도 예측으로 2020년을 시험용으로 사용한 결과이다. <표 7>은 <표 6>의 대출부도 예측 결과가 시험용 표본의 선택편의(Selection Bias)에 따른 것인지 검증하기 위해 2019년을 시험용 표본으로 예측한 결과이다. <표 8>과 <표 9>는 1년 대출부도 예측 계량지표값의 통계적 검증 결과이다.

계량지표는 이진 분류에서 주로 사용하는 AUC¹⁸⁾와 혼동행렬로 측정한 5개가 사용되었다.¹⁹⁾ AUC는 ML 모형의 부도확률 추정치와 실제 부도여부를 이용해 계산하며 계급 불균형에 영향을 받지 않는 계량지표이다. 혼동행렬은 ML 모형이 어떤 계급 예측으로 편향되는지 확인하는 장점은 있으나 판별점(Cut-Off Point)의 결정이 관건이다. 본 분석에서는 AUC와 F1스코어의 평균을 극대화하는 판별점을 찾아 혼동행렬을 구성하였다.²⁰⁾

12개 모형 중에서 창업기업과 일반기업의 대출부도 예측 성능이 가장 뛰어난 ML 모형을 결정하는 기준은 다음과 같이 설정하였다. 먼저 시험용 자료에서 AUC값이 높은 경쟁적 모형을 찾는다. 그리고 학습용과 비교한 AUC 증감률이 더 작고, 세 표본 AUC값 분산이 더 작은 모형을 과적합과 안정성이 가장 우수한 모형으로 평가하였다. 그 이유는 시험용에서의 대출부도 예측 성과를 기본적으로 중시하더라도 그것이 학습용 표본에서의 대출부도 예측 성과와 큰 괴리를 보이지 않아야 하기 때문이다. 기존

18) AUC는 ROC곡선(Receiver Operating Characteristic) 이하의 면적으로 1에 근접할수록 예측 정확성이 높고 0.5는 예측력이 없는 모형으로 본다.

19) 혼동행렬은 정해진 판별점 하에서 실제값과 예측값의 조합을 2×2 행렬표로 나타내고, 이로부터 정확도(Accuracy), 민감도(Sensitivity) 또는 재현율(Recall), 정밀도(Precision), 특이도(Specificity), F1스코어, 기하평균(Geometric Mean)을 측정한다.

구분		예측	
		건전기업	부도기업
실제	건전기업	a	b
	부도기업	c	d

정확도=(a+d)/(a+b+c+d), 민감도=d/(c+d), 특이도=a/(a+b), 정밀도=d/(b+d), F1스코어=(민감도×정밀도×2)/(민감도+정밀도), 기하평균=(민감도×특이도)^{0.5}

20) 본 논문이 고찰한 기존 연구들은 혼동행렬 계량지표 산출시 판별점의 선택에 관한 구체적 언급이 없다. 판별점의 기본값으로서 보통 0.5를 사용한 것으로 여겨진다. 본 논문이 사용한 방식 외에 판별점의 결정 방식으로 민감도와 특이도의 합 또는 평균의 극대화, F1스코어의 극대화 등을 사용할 수 있다.

연구들은 경쟁적 모형 사이에 AUC의 차이가 미세한데도 불구하고 AUC의 대소만 비교하여 예측 모형의 성과를 판단하는 경향이 있었다. 본 논문에서는 이에 모형간 AUC의 차이가 통계적으로 유의미한지 살펴보기 위해 Veganzones and Séverin(2018), Sigrist and Hirnschall(2019)이 사용한 DeLong et al.(1988)의 비모수적(Nonparametric) 검증을 수행하였고 그 결과를 <표 8>에 제시하였다.

<표 6>에서 볼 수 있듯이 모든 모형에서 AUC는 시험용이 학습용보다 낮다. 부도 예측 성과의 우수성을 평가하는 최소 기준은 경험법칙에 따르면 AUC 0.7이다. 시험용 자료에서 창업기업의 AUC가 0.7을 초과하는 모형은 다층퍼셉트론(MLP), 서포트벡터머신(SVM), AdaBoost, XGBoost, LightGBM이다. 가장 높은 AUC는 다층퍼셉트론으로 0.719이었고, 다음으로 서포트벡터머신이 0.716으로 약간 낮았다.

그런데 시험용 표본에서 아다부스트(AdaBoost)의 AUC는 0.711로 다층퍼셉트론과 서포트벡터머신보다 약간 낮지만 학습용과 검증용의 AUC를 함께 고려한 AUC 감소율과 분산은 각각 2.5%와 0.015로 두 모형보다 낮았다. 학습용에서 최고의 예측 성능은 배깅(Bagging)이었으나 시험용에서 AUC가 34.3% 하락하여 과적합이 나타났고, 다음으로 높은 AUC는 XGBoost, 다층퍼셉트론, 서포트벡터머신, LightGBM 순서를 보이지만 검증용과 시험용 사이에 AUC의 감소율이 아다부스트의 감소율보다 컸다. 결국 시험용에서 비슷한 정확성을 보이는 경쟁적 모형 중에 일부는 과적합을 나타냈다. 예컨대 XGBoost의 경우 학습용 AUC는 0.779, 시험용 AUC는 0.715로 두 표본에서 아다부스트의 AUC보다 높았다. 그러나 XGBoost는 시험용에서 AUC가 8.2% 하락하였고 표본간 분산도 0.161로 아다부스트의 하락폭과 분산보다 더 컸다. Jones et al.(2015)의 분석에서도 시험용 표본에 대한 예측 정확성은 다수의 ML이 학습용에 미치지 못하였으며 아다부스트와 랜덤포레스트(RF)가 좋은 성과를 보여주었다. 창업기업에서 판별분석(LDA)과 로지스틱 회귀모형(LR)은 AUC가 0.7에 매우 근접하는 예측 성능을 보여주면서 과적합도 발생하지 않았다. 더군다나 두 모형은 랜덤포레스트와 나이브 베이즈(NB)를 능가하는 예측 성능을 시험하였다.

본 예측이 팬데믹 경제위기의 정점에 있었던 시기를 시험용으로 사용한 선택편의적 결과인지를 확인하기 위해 2019년을 시험용으로 예측하였고 그 결과는 <표 7>에 표시하였다. 전체 자료에서 2020년을 제거한 이후 자료의 표집과 전처리는 전술한 방법대로 진행하였다. 표를 통해서 보면 다층퍼셉트론이 시험용 표본에서 가장 높은 AUC를 보이나, AUC 증감률과 분산을 고려할 때 아다부스트가 더 안정적인 예측 능력을 보유하는 것을 확인할 수 있다. 이것은 아다부스트가 창업기업의 대출부도 예측에 강건한 증거이다. 따라서 본 논문은 창업기업의 대출부도 예측 정확성과 안정성 면에서 아다부스트가 우수한 것으로 평가하였다.²¹⁾

그런데 창업기업(일반기업)의 시험용 AUC 증감률 평균은 2020년 -6.7%(-3.6%), 2019년은 -

7.5%(-5.6%)였다. 또한 창업기업에서 아다부스트를 비롯하여 다수의 ML 모형이 2019년 예측보다 2020년 예측이 좀 더 좋은 성과를 보였다. 만약 경제 상황과 예측 정확성이 연관된다면 2020년 예측 정확도가 2019년보다 더 하락할 것으로 기대할 수 있다. 하지만 실증결과는 경제 상황과 예측 정확성 사이에 어떤 연관성이 없음을 보여주었다.

〈표 8〉의 DeLong 검증은 AUC가 확연한 차이를 보이는 모형들보다는 비슷한 AUC를 갖는 모형 사이에 AUC가 유의적으로 다른지를 고찰하기에 유용하다. 창업기업의 2020년 시험용 자료에 대해 DeLong 검증을 수행한 결과 아다부스트와 서포트벡터머신의 AUC가 같다는 귀무가설을 5% 유의수준에서 기각하여 두 모형의 예측 성과가 통계적으로 유의적인 차이가 있는 것으로 나타났다. DeLong 검증은 시험용 표본에 대한 정태 분석으로서 서포트벡터머신이 가장 우수한 예측 모형을 의미하는 것은 아니다.

분석 결과를 정리하면, 창업기업의 대출부도 예측은 ML 모형이 통계적 모형보다 예측 정확성에서 우수하고, 아다부스트가 경쟁적 모형보다 안정적인 성능을 보여주었다. 이는 한편으로 창업기업의 대출부도 예측에 재무적 특성이 유효한 증거이다.

일반기업의 대출부도 예측 결과를 살펴보면, 〈표 6〉의 2020년 시험용 표본에서 의사결정트리, 나이브 베이즈, 배깅을 제외한 모든 모형이 AUC 0.7을 상회하였다. 그리고 학습용과 비교한 시험용 AUC의 감소폭은 창업기업보다 낮았다. 이것은 오랜 기간 사업을 영위한 기업의 대출부도 예측에는 재무적 특성이 창업기업에 비해 더 효과적임을 확인해 준다. 가장 좋은 예측 정확성과 양호한 안정성을 보인 모형은 서포트벡터머신이다. 이 모형의 시험용 AUC는 0.744이고, 다음으로 XGBoost와 아다부스트가 각각 AUC 0.736, 0.735로 우수한 성능을 시험하였다. 후자의 두 모형 중에는 아다부스트가 분산이 작고 증감률도 낮아 비교적 안정적이었다. 판별분석과 로지스틱 회귀모형은 AUC 0.727로 부스팅보다 약간 낮은 예측 정확성을 보여주고 있다. 그리고 일반기업 대출부도 예측에 대한 서포트벡터머신의 강건성은 〈표 7〉의 2019년 시험용 표본에서도 확인할 수 있다.

〈표 8〉의 일반기업 DeLong 검증은 창업기업과 마찬가지로 서포트벡터머신과 아다부스트가 유의적으로 다른 결과를 보여주었다. 통계적 모형과도 유의미한 차이가 있었다. 아다부스트와 XGBoost의 예측 성능은 최근 문헌들에서 증명된 LightGBM과 CatBoost보다 유의적으로 미미한 우위를 보였다.

21) 계량지표에 따른 결과의 강건성을 검증하기 위해 Hand(2009)의 H-측도(Measure)를 살펴본 결과 창업기업에서 아다부스트(AdaBoost)가 상대적으로 우수한 예측 성과를 갖는 것을 확인할 수 있다. 아래 표는 창업기업과 일반기업의 H-측도이다.

구분	LDA	LR	DT	NB	MLP	SVM	Bagging	RF	AdaBoost	XGBoost	LightGBM	CatBoost
창업기업	0.151	0.156	0.083	0.098	0.190	0.144	0.042	0.143	0.178	0.178	0.165	0.157
일반기업	0.190	0.194	0.138	0.128	0.189	0.156	0.085	0.170	0.200	0.204	0.189	0.178

일반기업의 1년 대출부도 예측 결과를 종합해보면, 일반기업의 대출부도 예측은 서포트벡터머신이 가장 우수하고 통계적 모형과 유의적인 성능적 차이가 있다. 다만, 통계적 모형과 ML 모형간 예측 성능의 차이만으로는 금융기관이 일반 중소기업의 신용평가를 위해 굳이 복잡한 ML을 이용해야 하는지를 확신할 수 없다. 왜냐하면 ML은 예측 정확성을 뚜렷하게 개선시키지 않으면서 대출승인 가부의 원인에 대한 설명만 어렵게 만들기 때문이다.

균형 표본에서 대출부도 예측 정확성은 불균형 표본의 대출부도 예측 결과에서도 나타났다. <부표 3>에서 볼 수 있듯이 창업기업과 일반기업의 1년 대출부도 예측 정확성은 대체로 균형 표본의 결과와 비슷하였다. 창업기업은 시험용 표본에서 아다부스트가 AUC 0.722로 측정되었다. 이 결과는 불균형 표본에서 아다부스트의 강건성을 시사한다.

혼동행렬의 계량지표는 분류의 목적과 계급 불균형 정도에 따라 다양하게 선택될 수 있다. 신용평가는 부도기업의 정확한 예측, 즉 부도기업에 대출을 승인할 제1종 오류(Type 1 Error)를 최소화하는 데에 집중한다. 이 경우 민감도가 중요한 계량지표이다. 하지만 계급 불균형이 심할수록 보통 F1스코어나 기하평균이 좋은 지표이다. <표 6>에서 민감도로 대출부도 예측 결과를 평가하면 일반기업의 대출부도 예측은 최신 부스팅 기법들이 우수하였다. 그러나 창업기업에서 과적합이 나타났다. ML 모형 중에는 창업기업은 아다부스트, 일반기업은 CatBoost가 세 표본에서 일관적으로 좋은 성능을 보였다.

〈표 6〉 1년 대출부도 예측 결과: 2020년 시험용 표본¹⁾

모형 ²⁾	대상	시험용						학습용						검증용						AUC ³⁾	
		AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	증감률(%)	분산 × 100
LDA	창업	.695	.648	.683	.638	.660	.647	.703	.645	.707	.629	.666	.642	.694	.632	.708	.615	.658	.628	-1.1	0.002
	일반	.727	.679	.752	.656	.701	.675	.733	.674	.732	.656	.692	.671	.729	.669	.733	.650	.689	.666	-0.9	0.001
LR	창업	.696	.641	.691	.628	.658	.639	.703	.651	.703	.637	.668	.649	.690	.630	.700	.614	.654	.626	-1.0	0.004
	일반	.727	.682	.730	.666	.696	.680	.732	.672	.729	.654	.690	.670	.729	.670	.737	.650	.690	.666	-0.7	0.001
DT	창업	.641	.608	.666	.597	.629	.605	.717	.652	.630	.658	.644	.651	.625	.588	.585	.589	.587	.588	-10.6	0.241
	일반	.692	.622	.772	.593	.671	.603	.722	.663	.619	.679	.648	.662	.687	.635	.594	.647	.619	.634	-4.2	0.036
NB	창업	.636	.597	.651	.588	.618	.595	.641	.600	.692	.585	.634	.593	.646	.606	.712	.587	.644	.596	-0.7	0.003
	일반	.667	.623	.754	.598	.667	.609	.665	.611	.772	.584	.665	.590	.670	.609	.777	.581	.665	.585	0.3	0.001
MLP	창업	.719	.666	.683	.661	.672	.666	.753	.685	.588	.729	.651	.678	.712	.660	.575	.693	.628	.655	-4.6	0.048
	일반	.726	.663	.677	.659	.668	.663	.729	.660	.731	.640	.682	.656	.723	.657	.737	.636	.682	.652	-0.4	0.001
SVM	창업	.716	.646	.698	.632	.663	.644	.747	.679	.708	.670	.688	.679	.699	.650	.713	.633	.671	.647	-4.3	0.060
	일반	.744	.674	.780	.643	.705	.665	.762	.693	.733	.678	.705	.692	.742	.672	.710	.660	.684	.671	-2.3	0.012
Bagging	창업	.586	.560	.662	.550	.601	.551	.892	.806	.771	.829	.799	.805	.584	.560	.497	.569	.530	.556	-34.3	3.147
	일반	.648	.616	.688	.601	.642	.612	.892	.808	.776	.829	.802	.807	.644	.603	.551	.614	.581	.600	-27.3	2.016
RF	창업	.688	.623	.719	.603	.656	.615	.728	.656	.676	.649	.662	.655	.688	.630	.671	.621	.645	.629	-5.5	0.053
	일반	.714	.649	.738	.626	.678	.642	.731	.669	.715	.655	.684	.668	.722	.658	.708	.643	.674	.656	-2.4	0.008
AdaBoost	창업	.711	.667	.732	.648	.687	.664	.730	.672	.755	.647	.697	.667	.706	.636	.736	.613	.669	.628	-2.5	0.015
	일반	.735	.671	.730	.653	.689	.668	.741	.676	.735	.657	.694	.673	.732	.667	.729	.648	.686	.664	-0.8	0.002
XGBoost	창업	.715	.643	.717	.624	.667	.638	.779	.703	.788	.673	.726	.698	.704	.642	.742	.618	.675	.634	-8.2	0.161
	일반	.736	.664	.759	.638	.693	.658	.755	.690	.778	.662	.715	.684	.734	.671	.759	.645	.697	.665	-2.6	0.013
LightGBM	창업	.702	.641	.700	.626	.661	.638	.747	.685	.764	.659	.708	.680	.701	.637	.728	.616	.667	.630	-6.1	0.070
	일반	.727	.659	.760	.632	.690	.651	.742	.680	.784	.649	.710	.672	.730	.671	.775	.641	.702	.663	-2.0	0.006
CatBoost	창업	.696	.636	.717	.617	.663	.631	.711	.654	.766	.626	.689	.645	.686	.625	.750	.600	.667	.613	-2.1	0.016
	일반	.721	.655	.782	.623	.693	.642	.723	.656	.779	.625	.694	.645	.721	.654	.787	.621	.694	.640	-0.3	0.000

주1) 이용된 자료는 창업기업에서 학습용 5,560개, 검증용 2,384개, 시험용 946개, 일반기업에서 학습용 15,029개, 검증용 6,441개, 시험용 2,362개이다. AUC값이 가장 높은 모형을 굵은 글씨로 표시하였다.

주2) LDA는 선형판별분석, LR은 로지스틱 회귀모형, DT는 의사결정트리, NB는 나이브 베이즈, MLP는 다층퍼셉트론, SVM은 서포트벡터머신, RF는 랜덤포레스트이다.

주3) 증감률은 학습용 대비 시험용의 AUC 변동률이다. 분산은 세 표본 AUC의 분산이다.

〈표 7〉 1년 대출부도 예측 결과: 2019년 시험용 표본¹⁾

모형 ²⁾	대상	시험용						학습용						검증용						AUC ³⁾	
		AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	증감률(%)	분산 × 100
LDA	창업	.700	.647	.714	.629	.669	.643	.706	.648	.707	.632	.668	.645	.675	.627	.674	.616	.644	.625	-0.9	0.026
	일반	.712	.650	.750	.625	.682	.642	.733	.675	.741	.654	.695	.671	.730	.675	.738	.655	.694	.671	-2.9	0.013
LR	창업	.698	.639	.576	.660	.615	.636	.708	.650	.701	.637	.667	.648	.683	.632	.678	.621	.648	.631	-1.5	0.016
	일반	.713	.657	.736	.635	.682	.652	.734	.680	.735	.662	.697	.678	.732	.678	.732	.660	.694	.676	-2.9	0.014
DT	창업	.643	.591	.615	.587	.601	.590	.712	.658	.668	.655	.662	.658	.619	.588	.598	.586	.592	.588	-9.8	0.234
	일반	.687	.629	.696	.614	.653	.626	.727	.672	.700	.663	.681	.672	.693	.647	.680	.638	.658	.647	-5.6	0.047
NB	창업	.626	.599	.647	.590	.617	.597	.638	.599	.703	.582	.637	.590	.626	.590	.684	.576	.625	.583	-1.9	0.005
	일반	.646	.595	.754	.572	.650	.573	.669	.613	.773	.585	.666	.591	.672	.612	.779	.584	.667	.588	-3.4	0.019
MLP	창업	.720	.665	.683	.659	.671	.664	.765	.699	.656	.717	.685	.697	.696	.654	.624	.663	.643	.653	-5.9	0.120
	일반	.712	.650	.749	.625	.681	.642	.727	.668	.601	.695	.644	.665	.729	.671	.603	.698	.647	.668	-2.0	0.008
SVM	창업	.697	.632	.626	.634	.630	.632	.749	.685	.731	.669	.699	.683	.680	.630	.663	.622	.642	.629	-7.0	0.132
	일반	.738	.668	.775	.638	.700	.659	.758	.690	.740	.672	.705	.688	.743	.680	.726	.665	.694	.678	-2.7	0.011
Bagging	창업	.594	.559	.626	.552	.587	.555	.891	.813	.780	.835	.807	.812	.590	.567	.518	.574	.545	.565	-33.3	2.980
	일반	.624	.583	.671	.571	.617	.577	.896	.810	.758	.847	.800	.809	.633	.598	.544	.609	.575	.595	-30.4	2.393
RF	창업	.670	.616	.606	.618	.612	.616	.736	.673	.700	.664	.682	.672	.663	.612	.635	.607	.621	.611	-8.9	0.159
	일반	.704	.625	.817	.591	.686	.595	.732	.657	.729	.638	.680	.654	.721	.656	.736	.635	.682	.651	-3.8	0.019
AdaBoost	창업	.705	.660	.709	.646	.676	.658	.733	.673	.744	.652	.695	.670	.677	.625	.705	.608	.653	.620	-3.8	0.079
	일반	.718	.656	.751	.631	.686	.649	.739	.670	.750	.647	.694	.665	.736	.670	.750	.646	.694	.665	-2.8	0.013
XGBoost	창업	.694	.635	.644	.633	.638	.635	.747	.689	.711	.681	.696	.689	.673	.622	.651	.616	.633	.622	-7.2	0.149
	일반	.724	.661	.764	.633	.692	.652	.763	.692	.782	.663	.717	.686	.733	.668	.762	.641	.696	.661	-5.0	0.040
LightGBM	창업	.695	.645	.696	.631	.662	.643	.753	.693	.752	.673	.710	.691	.675	.621	.695	.605	.647	.617	-7.6	0.162
	일반	.717	.659	.760	.632	.690	.651	.744	.676	.773	.648	.705	.669	.729	.666	.761	.640	.695	.659	-3.6	0.018
CatBoost	창업	.690	.648	.714	.631	.670	.645	.709	.659	.747	.635	.687	.653	.665	.620	.716	.601	.653	.613	-2.7	0.049
	일반	.706	.640	.756	.614	.677	.630	.721	.655	.766	.627	.690	.646	.724	.652	.764	.624	.687	.642	-2.1	0.009

주1) 이용된 자료는 창업기업에서 학습용 4,782개, 검증용 2,050개, 시험용 1,112개, 일반기업에서 학습용 13,097개, 검증용 5,613개, 시험용 2,760개이다. AUC값이 가장 높은 모형을 굵은 글씨로 표시하였다.

주2) LDA는 선형판별분석, LR은 로지스틱 회귀모형, DT는 의사결정트리, NB는 나이브 베이즈, MLP는 다층퍼셉트론, SVM은 서포트벡터머신, RF는 랜덤포레스트이다.

주3) 증감률은 학습용 대비 시험용의 AUC 변동률이다. 분산은 세 표본 AUC의 분산이다.

〈표 8〉 DeLong 검증: 2020년 시험용 표본

Panel A: 창업기업(1년 대출부도 예측)¹⁾

	LDA	LR	DT	NB	MLP	SVM	Bagging	RF	AdaBoost	XGBoost	Light GBM
LR	0.425 (0.671)										
DT	-3.103 (0.002)	-3.166 (0.002)									
NB	-3.910 (0.000)	-4.027 (0.000)	0.462 (0.644)								
MLP	1.837 (0.066)	1.686 (0.092)	4.361 (0.000)	4.612 (0.000)							
SVM	-0.561 (0.575)	-0.674 (0.500)	2.899 (0.004)	3.353 (0.001)	-2.233 (0.026)						
Bagging	-5.329 (0.000)	-5.395 (0.000)	-2.607 (0.009)	-3.403 (0.001)	-6.701 (0.000)	-5.197 (0.000)					
RF	-0.531 (0.595)	-0.634 (0.526)	3.305 (0.001)	3.958 (0.000)	-2.335 (0.020)	-0.008 (0.994)	5.750 (0.000)				
AdaBoost	1.379 (0.168)	1.258 (0.208)	4.513 (0.000)	4.941 (0.000)	-0.571 (0.568)	2.181 (0.029)	6.569 (0.000)	2.943 (0.003)			
XGBoost	1.437 (0.151)	1.308 (0.191)	4.590 (0.000)	4.334 (0.000)	-0.320 (0.749)	2.120 (0.034)	6.525 (0.000)	2.715 (0.007)	0.354 (0.723)		
LightGBM	0.479 (0.632)	0.373 (0.709)	3.795 (0.000)	3.557 (0.000)	-1.342 (0.180)	1.068 (0.286)	5.864 (0.000)	1.430 (0.153)	-1.155 (0.248)	-2.562 (0.010)	
CatBoost	0.077 (0.938)	-0.019 (0.985)	3.535 (0.000)	3.273 (0.001)	-1.659 (0.097)	0.694 (0.488)	5.659 (0.000)	0.970 (0.332)	-2.238 (0.025)	-2.325 (0.020)	-0.816 (0.414)

주1) 통계량은 DeLong의 z-통계량, ()는 유의확률

Panel B: 일반기업(1년 대출부도 예측)¹⁾

	LDA	LR	DT	NB	MLP	SVM	Bagging	RF	AdaBoost	XGBoost	Light GBM
LR	0.114 (0.909)										
DT	-3.663 (0.000)	-3.818 (0.000)									
NB	-5.405 (0.000)	-5.763 (0.000)	-0.936 (0.349)								
MLP	-0.031 (0.975)	-0.072 (0.942)	3.860 (0.000)	5.008 (0.000)							
SVM	-3.011 (0.003)	-3.112 (0.002)	1.640 (0.101)	3.050 (0.002)	-3.666 (0.000)						
Bagging	-7.391 (0.000)	-7.633 (0.000)	-4.130 (0.000)	-3.756 (0.000)	-7.239 (0.000)	-5.629 (0.000)					

	LDA	LR	DT	NB	MLP	SVM	Bagging	RF	AdaBoost	XGBoost	Light GBM
RF	-1.721 (0.085)	-1.863 (0.062)	3.317 (0.001)	5.003 (0.000)	-1.834 (0.067)	1.414 (0.157)	7.491 (0.000)				
AdaBoost	1.152 (0.249)	1.168 (0.243)	6.149 (0.000)	7.291 (0.000)	1.418 (0.156)	4.935 (0.000)	9.357 (0.000)	5.535 (0.000)			
XGBoost	1.1444 (0.252)	1.1445 (0.252)	5.698 (0.000)	6.215 (0.000)	1.349 (0.177)	4.325 (0.000)	8.635 (0.000)	4.478 (0.000)	0.162 (0.871)		
LightGBM	0.057 (0.955)	0.027 (0.978)	4.736 (0.000)	5.413 (0.000)	0.095 (0.924)	3.242 (0.001)	7.935 (0.000)	3.136 (0.002)	-2.011 (0.044)	-3.325 (0.001)	
CatBoost	-0.703 (0.482)	-0.765 (0.444)	3.943 (0.000)	5.183 (0.000)	-0.822 (0.411)	2.554 (0.011)	7.592 (0.000)	1.949 (0.051)	-4.599 (0.000)	-3.367 (0.001)	-1.694 (0.090)

주1) 통계량은 DeLong의 z-통계량, ()는 유의확률

〈표 9〉 Nemenyi 검증: 2020년 시험용 표본

Panel A: 창업기업(1년 대출부도 예측)¹⁾

	LR	DT	NB	MLP	SVM	Bagging	RF	AdaBoost	XGBoost	Light GBM	CatBoost
LDA	0.793 (1.000)	2.887 (0.665)	3.567 (0.326)	1.132 (1.000)	0.170 (1.000)	3.906 (0.196)	1.529 (0.995)	1.981 (0.964)	0.000 (1.000)	0.679 (1.000)	0.793 (1.000)
LR		2.095 (0.946)	2.774 (0.720)	1.925 (0.971)	0.962 (1.000)	3.114 (0.548)	0.736 (1.000)	2.774 (0.720)	0.793 (1.000)	0.113 (1.000)	0.000 (1.000)
DT			0.679 (1.000)	4.020 (0.162)	3.057 (0.578)	1.019 (1.000)	1.359 (0.998)	4.869 (0.029)	2.887 (0.665)	2.208 (0.923)	2.095 (0.946)
NB				4.699 (0.042)	3.737 (0.256)	0.340 (1.000)	2.038 (0.955)	5.548 (0.005)	3.567 (0.326)	2.887 (0.665)	2.774 (0.720)
MLP					0.962 (1.000)	5.039 (0.019)	2.661 (0.771)	0.849 (1.000)	1.132 (1.000)	1.812 (0.982)	1.925 (0.971)
SVM						4.076 (0.147)	1.698 (0.989)	1.812 (0.982)	0.170 (1.000)	0.849 (1.000)	0.962 (1.000)
Bagging							2.378 (0.877)	5.888 (0.002)	3.906 (0.196)	3.227 (0.490)	3.114 (0.548)
RF								3.510 (0.351)	1.529 (0.995)	0.849 (1.000)	0.736 (1.000)
AdaBoost									1.981 (0.964)	2.661 (0.771)	2.774 (0.720)
XGBoost										0.679 (1.000)	0.793 (1.000)
LightGBM											0.113 (1.000)

주1) 평균순위 $q_a \sqrt{k(k+1)/6N}$ 보다 크거나 같은지 검증(q_a : 임계값, k : 모형 수, N : 지표 수), () 유의확률

Panel B: 일반기업(1년 대출부도 예측)¹⁾

	LR	DT	NB	MLP	SVM	Bagging	RF	AdaBoost	XGBoost	Light GBM	CatBoost
LDA	0.057 (1.000)	3.510 (0.351)	3.737 (0.256)	2.151 (0.935)	0.113 (1.000)	4.416 (0.077)	2.831 (0.693)	1.302 (0.999)	1.246 (0.999)	1.698 (0.989)	1.698 (0.989)
LR		3.567 (0.326)	3.793 (0.235)	2.208 (0.923)	0.057 (1.000)	4.472 (0.069)	2.887 (0.665)	1.359 (0.998)	1.302 (0.999)	1.755 (0.986)	1.755 (0.986)
DT			0.226 (1.000)	1.359 (0.998)	3.623 (0.302)	0.906 (1.000)	0.679 (1.000)	2.208 (0.923)	2.265 (0.909)	1.812 (0.982)	1.812 (0.982)
NB				1.585 (0.994)	3.85 (0.215)	0.679 (1.000)	0.906 (1.000)	2.434 (0.859)	2.491 (0.839)	2.038 (0.955)	2.038 (0.955)
MLP					2.265 (0.909)	2.265 (0.909)	0.679 (1.000)	0.849 (1.000)	0.906 (1.000)	0.453 (1.000)	0.453 (1.000)
SVM						4.529 (0.061)	2.944 (0.636)	1.415 (0.998)	1.359 (0.998)	1.812 (0.982)	1.812 (0.982)
Bagging							1.585 (0.994)	3.114 (0.548)	3.17 (0.519)	2.717 (0.746)	2.717 (0.746)
RF								1.529 (0.995)	1.585 (0.994)	1.132 (1.000)	1.132 (1.000)
AdaBoost									0.057 (1.000)	0.396 (1.000)	0.396 (1.000)
XGBoost										0.453 (1.000)	0.453 (1.000)
LightGBM											0.000 (1.000)

주1) 평균순위가 $q_{\alpha}\sqrt{k(k+1)/6N}$ 보다 크거나 같은지 검증(q_{α} : 임계값, k: 모형 수, N: 지표 수), () 유의확률

혼동행렬의 5개 계량지표 평균 순위로 모형의 예측 성과를 검증한 Nemenyi(1963) 결과는 <표 9>에 있다.²²⁾ 표를 통해서 보면 창업기업과 일반기업 모두 대체로 모형간 대출부도 예측 성과가 유의적으로 다르지 않았다. 이는 많은 모형과 계량지표를 이용하면서 ML 모형마다 판별점을 최적화한 것이 영향을 미친 것으로 판단된다.

한편, 균형 표본의 2년 대출부도 예측 성과는 <부표 1>에 표시하였다. 2년 대출부도 예측의 성과는 AUC 지표로 평가할 때 1년 대출부도 예측 성과보다 낮았다. 표를 통해서 볼 때, 창업기업의 2년 대출부도 예측 정확성이 가장 높은 ML 모형은 XGBoost였다. 하지만 AUC의 분산과 증감률을 고려하면 LightGBM이 우수한 예측력을 보여주었다. 일반기업의 경우 다층퍼셉트론이 시험용에서 가장 높았으나 과적합 경향이 나타났고, 결과를 종합하면 아다부스트가 우수한 성능을 보여주었다.

22) Demšar(2006)의 권고에 따라 Nemenyi 검증을 수행한다. Friedman(1940)의 카이제곱 통계량의 F검증 결과 12개 모형의 성과가 유사하다는 귀무가설을 1% 유의수준에서 기각하였다. 이에 모형들의 모든 쌍에 대해 Nemenyi 검증을 수행하였다.

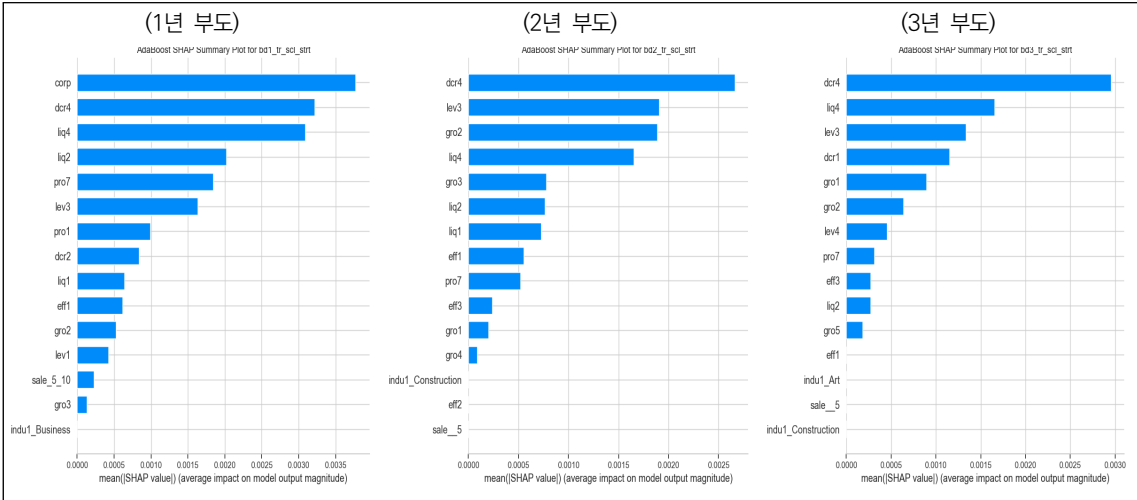
〈부표 2〉의 3년 대출부도 예측에서 AUC의 증감률과 분산을 고려할 때 아다부스트가 창업기업과 일반기업 모두 좋은 예측 성과를 보여주었다. 표를 보면 3년 대출부도 예측의 AUC는 2년 대출부도 예측의 AUC보다 낮았다. 이것은 현재의 예측변수로 장기적 예측이 어려운 당연한 결과로 풀이된다. 시간이 경과하면서 현재의 정보적 유효성은 점진적으로 감소한다. 그에 따른 예측 정확성의 하락폭은 창업기업이 일반기업보다 크게 나타났다. 〈부표 3〉의 불균형 표본 예측 결과도 이 현상을 확인해 준다. 전술한 바대로 현재의 재무적 성과지표 위주로 역동적인 창업기업의 장기적 신용도 평가는 한계가 있다.

〈그림 5〉와 〈그림 6〉은 2020년 시험용 예측에서 안정적인 예측 성과가 확인된 아다부스트의 예측변수를 살펴보기 위해 SHAP(Shapley Additive Explanations)의 변수중요도 요약도표(Summary Plot)와 점도표(Dot Plot)를 표시한 도표이다. SHAP는 각 특성이 예측에 어떻게 기여하는지를 측정하는 것으로 특히 복잡한 ML 모형에서 각 입력변수가 예측 결과에 미치는 영향을 설명하기에 효과적인 지표이다. 〈그림 5〉의 변수중요도 요약도표는 아다부스트 모형에서 영향력이 큰 15개 변수를 가장 높은 SHAP값을 갖는 변수부터 내림차순으로 표시하였다.

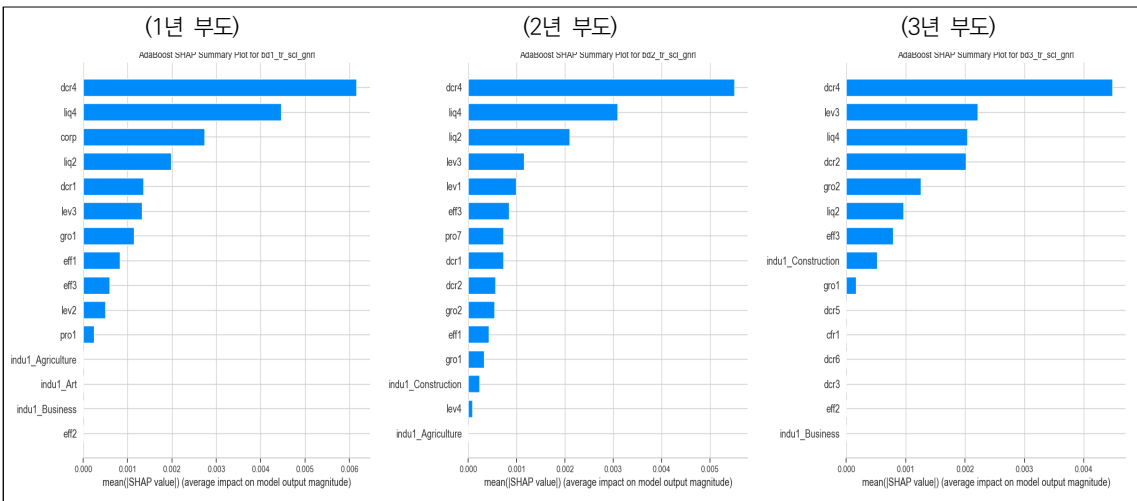
창업기업의 1년 대출부도 예측의 경우 법인여부(Corp), 부채상환계수(Dcr4), 현금비율(Liq4), 운전자본비율(Liq2), 매출액순이익률(Pro7), 차입금의존도(Lev3)가 대출부도 예측에 영향력이 높은 것으로 나타났다. 그런데 일반기업 모형의 예측변수는 이 6개 변수 중 창업기업의 매출액순이익률이 EBITDA이자보상비율(Dcr1)로 대체될 뿐 나머지 5개 변수가 같았다. 또한 창업기업과 일반기업 모두 부채상환계수(Dcr4), 현금비율(Liq4), 차입금의존도(Lev3)가 중장기 예측에 공통으로 사용되었다. 이 결과는 부채상환능력과 유동성을 대리하는 변수들이 중소기업의 대출부도 예측에 매우 유효한 실증적 증거로서 Son et al.(2019)의 분석 결과를 지지해준다. 자금조달을 금융기관 차입에 의존하는 중소기업은 채무상환을 위해 충분한 수익과 현금을 창출하지 못하면 대출부도로 이어질 수 있다. 다른 한편으로, 창업기업의 중장기 예측에는 총자산증가율(Gro2) 등 성장성 범주 변수들이 중요한 예측요소로 나타났다. 창업기업의 성장 지체가 장기적 대출부도의 원인이 될 수 있다. 한편, 산업, 지역, 기업규모 요소는 창업기업과 일반기업 대출부도 예측에 대해 재무적 특성만큼 유효한 설명력을 보유하지 않았다. 예컨대, 〈그림 5〉에서 창업기업 1년 부도 예측의 경우 매출액 5~10억원(Sale_5_10)과 사업서비스 산업(Indu1_Business) 변수 정도가 낮은 변수중요도 순위를 나타내고 있다.

〈그림 5〉 AdaBoost SHAP 변수중요도 요약 도표¹⁾

Panel A: 창업기업



Panel B: 일반기업



주1) 횡축은 SHAP 절대값의 평균을 의미하고, 종축의 변수들은 <표 3>에 정의되었다.

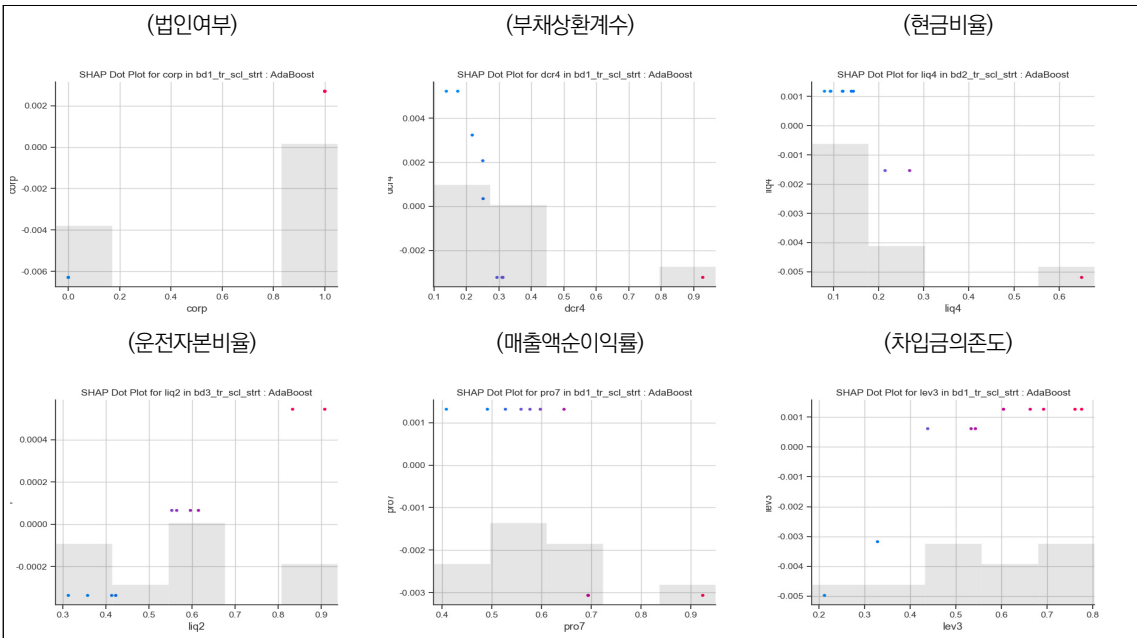
변수중요도는 예측에 영향을 미치는 변수들의 평균적인 중요도를 의미할 뿐 대출부도에 양(+) 또는 음(-)의 영향을 미치는지에 관한 정보는 제공하지 않는다. SHAP 방법으로 작성한 점도표는 예측변수의 방향성을 확인해 준다.²³⁾ <그림 6>은 아다부스트 1년차 대출부도 예측에서 변수중요도 상위 6개 변

23) SHAP 방법은 표본 관측치가 많을 때 계산시간이 기하급수적으로 증가하는 문제가 있다. 그래서 전체 자료의 특성을 잘 반영하면서 계산을 효율적으로 수행하는 k-평균 군집화(K-means Clustering) 방식으로 도표를 작성하였다. 그림에서 점도표에서 가로축은 변수값이고, 세로축은 SHAP 값, 음영 막대는 변수의 확률밀도이다. 세로축의 SHAP 값이 양수이면 해당 변수값이 부도확률을 증가시

수의 SHAP 점도표이다. 창업기업의 주요 변수를 살펴보면 부채상환계수, 현금비율, 매출액순이익률은 부도확률과 음(-)의 관계, 운전자본비율과 차입금의존도는 부도확률과 양(+) 관계를 시현하였다. 법인여부 변수는 법인의 부도확률이 더 높고, 운전자본비율은 일반적 기대와 달리 부도확률과 양(+)의 상관관을 보여주고 있다. 일반기업의 경우 주요 변수들의 부도확률과의 관계는 창업기업의 결과와 유사하였다.

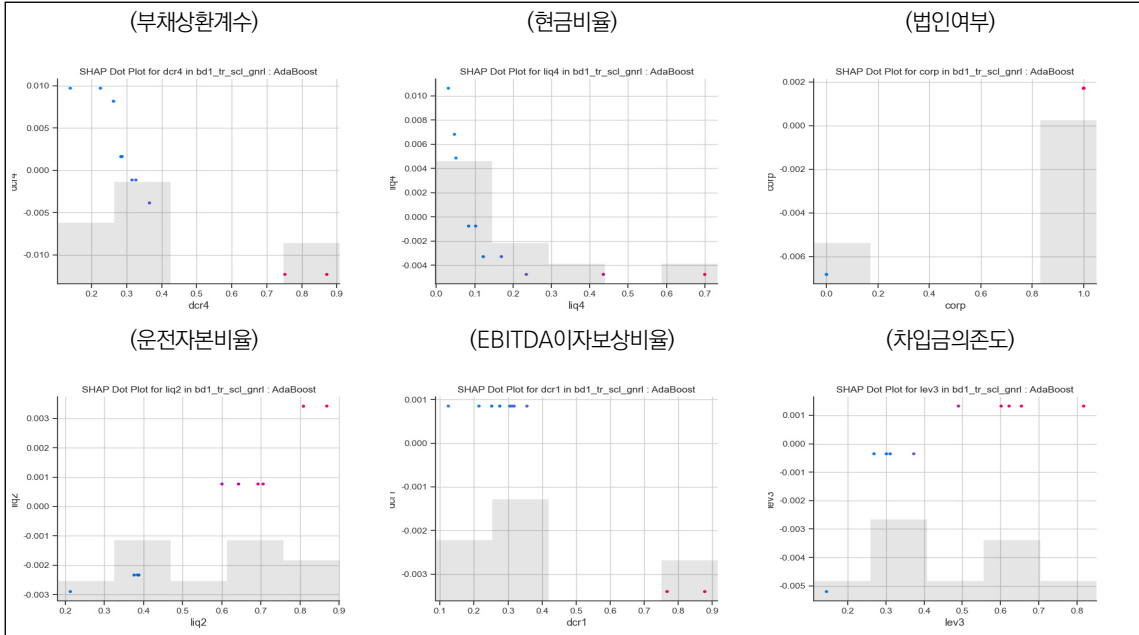
〈그림 6〉 AdaBoost SHAP 점도표: 1년 대출부도 예측¹⁾

Panel A: 창업기업



키는 쪽으로 작용했음을 의미하고, 음수이면 반대로 부도확률을 감소시키는 쪽으로 작용했음을 뜻한다. 변수값이 낮은 영역에서 양(+)의 SHAP값은 변수값이 작을 때 부도확률을 증가시키고, 변수값이 높은 영역에서 SHAP값이 음(-)의 값이면 변수값이 클 때 부도확률을 감소시키는 것으로 해석한다.

Panel B: 일반기업



주1) 점도표에서 가로축은 변수값이고, 세로축은 SHAP 값, 음영 막대는 확률밀도이다.

본 실증분석으로부터 발견된 주요 내용은 다음과 같이 정리할 수 있다.

첫째, 머신러닝 모형을 이용한 창업기업의 대출부도 예측은 일반기업의 대출부도 예측보다 예측 정확도가 낮았다. 하지만 창업기업과 일반기업의 예측 정확도의 차이는 평균적으로 크게 다르지 않았다. 중소기업의 1년 단기 대출부도 예측에서 정확성과 안정성이 우수한 머신러닝은 창업기업의 경우 아다부스트이고 일반기업은 서포트벡터머신이다. 다만 장기 대출부도 예측은 이용한 머신러닝 모형 모두에서 정확성이 하락하였고 특히 창업기업의 머신러닝 모형의 정확성 하락 폭이 현저하였다.

둘째, 판별분석과 로지스틱 회귀모형의 예측 정확도는 일부 머신러닝과 거의 대등하였다. 오히려 대출부도 예측 성과가 통계적 모형보다 저조한 머신러닝도 있다. 과적합을 통제하기 위해 머신러닝의 초매개변수를 조정한 결과라는 점을 고려하더라도 로지스틱 회귀모형과 판별분석이 창업기업의 대출부도 예측에 유용하다는 증거이다. 머신러닝 모형을 적용할 때 예측 결과의 일반화를 제고하기 위해서 요구되는 과정 중 하나는 분석 변수의 전처리이다. 전처리 과정 중 본 논문이 사용한 정규화와 이상치의 변수 변환은 변수의 특정값 주위에서 머신러닝 알고리즘이 과도하게 적합되는 것을 예방한다. 이 변환 과정은 통계적 모형의 추정시 극단값의 영향력을 줄여준다. 또한 앞서 기초 분석에서 살펴본 매출액과 부도율의 비선형적 관계를 통계적 모형 같은 선형모형은 포착하기 어려운데, 변수의 터미변수화는 동관계를 통계적 모형이 반영할 수 있게 해준다. 이와 같은 변수 변환 과정을 통해 통계적 모형의 대출부

도 예측 성능은 동일한 조건에서 머신러닝 모형과 비교될 수 있다. 따라서 본 분석에서는 적절한 변수 변환 과정이 창업기업 대출부도 예측 정확성을 높이는데 기여한 것으로 판단된다.

셋째, 재무적 특성은 창업기업의 대출부도 예측에 유효한 요소로 나타났다. 아다부스트의 예측변수들로 평가할 때 부채상환능력과 유동성 범주의 재무비율들이 창업기업과 일반기업의 장단기 대출부도 예측에 큰 영향을 미쳤다. 이윤과 현금창출 능력이 중소기업의 영속성을 결정하는 매우 중요한 요소이다. 금융기관이 중소기업에 대한 대출심사에서 기업의 신용도를 평가하는 제일의 요소는 기업경영 성과의 측정이다. 대출의 가치가 기업의 영업에서 창출되는 수익에 달려 있기 때문에 재무제표로부터 산출된 재무적 특성은 채무상환을 예측하는 정보로서 가치가 있다. 그런데 재무제표의 신뢰성과 투명성의 이슈가 창업기업에 제기됨에도 불구하고 창업기업의 대출부도 예측에 재무적 특성이 유효한 실증적 결과는 다른 예측 요소를 고려하지 않더라도 일정 수준의 예측 정확성을 확보할 수 있음을 입증해 준다.

VI. 결론

본 논문은 특별히 창업기업을 대상으로 머신러닝 방법을 적용하여 대출부도를 예측한 점에 큰 의의가 있다. 실증분석은 단순히 대출부도 예측 정확도 개선에 치중하지 않고 대용량 자료의 활용, 통계적 모형을 포함한 다수의 머신러닝 알고리즘 방법의 적용, 시점간 적합성 검증을 토대로 예측 결과의 일반화에 집중하였다. 본 장에서는 실증적 결과로부터 시사점을 도출하고 분석의 한계점을 제시하면서 연구의 결론을 갈음한다. 제시한 연구 한계점은 또한 머신러닝을 이용한 중소기업 대출부도 예측 연구의 도전 과제이다.

1. 시사점

첫째, 본 실증적 결과가 시사하는 바는 최신의 방법이 최고의 예측 성능을 꼭 보장하지 않는다는 것이다. 즉 그라디언트 부스트 류의 최신 알고리즘이 최근에 활용되고 있지만 본 분석에서 아다부스트와 서포트벡터머신의 전통적 기법이 우수한 대출부도 예측 성능을 보여주었다. 금융기관들은 보통 금융 관련 신기술의 빠른 추격자(Fast Follower)이다. 이것이 가능한 이유는 사업모형이 종단적(End-To-End)이기 때문이다. 게다가 금융환경의 변화와 상업적 유행이 가세하면 신기술이라는 이름으로 무턱대고 금융업무에 도입하려는 경향이 있다. 본 머신러닝의 실증적 결과들은 표본의 성질이 예측에 중대한 영향을 끼칠 수 있음을 시사한다. 따라서 금융인들이 머신러닝 모형을 도입할 때는 여러 비교 가능한 기법들의 철저한 검증 과정이 필요하다.

둘째, 창업기업 대출부도 예측에 로지스틱 회귀모형이 비교적 좋은 예측 성과를 보인 것은 금융기관의 창업기업 신용평가를 위한 기본적 모형으로서 로지스틱 회귀모형이 실무적으로 활용가치가 높음을 시사한다. 로지스틱 회귀모형은 구조적 모형의 특성상 고객에게 대출의 가부 결정을 투명하게 설명할 수 있는 장점이 있다. 머신러닝의 예측 성능적 우위에도 불구하고 금융기관이 중소기업 신용평가에 로지스틱 회귀모형의 통계적 모형을 선호하는 근본적 이유가 여기에 있다. 다만, 대출부도 예측 모형 개발자는 자료의 전처리 과정에서 재무적 특성 등 입력변수들의 통계적 특성을 주도면밀하게 분석 검토하는 것이 예측 정확성 향상을 위해 필요한 과제임을 유념해야 한다.

셋째, 기업 경영자 관련 특성을 입력변수로 이용하지 않고 재무적 특성 위주로도 창업기업의 대출부도 예측에 소기의 성과를 창출할 수 있음을 본 분석은 시사한다. 일반기업에 비해 창업기업은 규모가 영세하고 소유와 경영의 분리가 모호하여 기업 경영자의 특성이 기업 대출부도 예측에 중요한 정보로서 역할을 한다. 그래서 금융기관이 창업기업 대출부도 예측 정확도에 대한 절대적 목표치를 상향시키면 재무적 특성 외에 신용정보, 공공정보 등 다른 추가 요소들이 요구된다. 또한 기업의 부도에 영향을 미칠 수 있는 기업의 기술 수준, 정부 정책적 요인, 기타 기업체 속성도 고려할 필요가 있다. 이러한 자료들은 법률적 제약뿐만 아니라 비용이 수반되기 때문에 모형 개발자는 예측 정확도와 운영 효율성을 종합적으로 검토해야 한다.

2. 한계점

첫째, 창업기업의 개념적 모호성이다. 국내에서 창업기업의 범위에 대한 학술적 정의는 없고 업력을 기준으로 하는 경제정책적 측면에서의 논의만 있을 뿐이다. 본 분석이 중소기업창업 지원법에서 규정한 업력 7년까지로 창업기업을 정의하지 않았던 것은 창업의 범위가 매우 넓다는 판단 때문이었다. 물론 성장 속도가 빠른 산업은 업력 5년도 넓은 기준일 수 있다. 이렇듯 창업기업의 범위를 어떻게 정의하느냐는 실증적으로 중요한 쟁점이다. 업력을 세분화하여 대출부도 예측 결과를 비교해 보는 방법이 있으나 다수의 알고리즘을 사용하기 때문에 예측 결과의 해석이 복잡해질 개연성이 크고 무엇보다도 자료 분할에 따른 부도 표본의 축소가 모형 추정 결과의 일반화 문제를 야기할 수 있다.

둘째, 본 분석에서 이용된 AUC 등의 계량지표는 예측의 정확도를 평가할 뿐 예측 오류로 인한 비용을 고려하지 못하는 한계가 있다. 가장 정확한 예측 모형이 금융기관에게 반드시 최고의 수익을 보증하지 않기 때문이다. 금융기관의 신용평가는 실제 부도기업이 건전기업으로 잘못 분류된 비율인 제1종 오류에 초점을 맞추고, 실제 건전기업이 부도기업으로 잘못 분류된 비율인 제2종 오류에는 관심을 덜 두는 경향이 있다. 즉 잠재고객의 상실로써 발생하는 기회이익의 손실보다 부실대출의 승인으로 발생하는 부도손실(Loss Given Default)이 더욱 크기 때문에 금융기관은 부도기업의 식별에 더욱 집중한

다. 그래서 금융기관이 단순 예측 정확도 개선이 아니라 금융수익 극대화에 목적을 둔 신용평가체계를 구축하려 한다면 대출부도 예측 모형을 개발할 때 두 오분류 비용(Misclassification Cost)을 함께 고려할 수 있는 Hand(2009)의 H-측도 같은 계량지표를 활용할 필요가 있다. 다만, H-측도는 AUC의 대안으로 사용할 수 있으나 비용함수를 정의해야 하는 까다로운 과제가 있다.

셋째, 머신러닝을 이용한 대출부도 예측은 임의성의 한계를 피하기 어렵다. 머신러닝 알고리즘, 부도의 정의, 표본의 구성, 변수의 변환, 초매개변수 등 예측을 위한 방법과 절차에 다양한 선택 대안이 있으나 합의된 단일 표준은 존재하지 않는다. 이러한 요소들에 대해서 무수히 많은 경우의 수를 시험하는 것은 복잡하고 해석도 어려워 좋은 방법은 아닐 것이다. 선택의 저주는 본 분석만의 문제를 넘어 머신러닝 부도 예측 분야 공통의 한계이다.

[부록]

〈부표 1〉 2년 대출부도 예측 결과: 2020년 시험용 표본¹⁾

모형 ²⁾	대상	시험용						학습용						검증용						AUC ³⁾	
		AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	증감률(%)	분산 × 100
LDA	창업	.635	.587	.758	.565	.648	.562	.650	.614	.651	.607	.628	.613	.649	.609	.648	.601	.624	.608	-2.3	0.007
	일반	.661	.624	.698	.608	.650	.620	.694	.644	.711	.627	.666	.641	.686	.644	.684	.633	.657	.643	-4.9	0.031
LR	창업	.635	.595	.719	.576	.639	.582	.652	.614	.650	.606	.627	.612	.649	.608	.637	.602	.619	.607	-2.6	0.008
	일반	.659	.620	.703	.603	.649	.614	.693	.644	.708	.627	.665	.640	.688	.641	.682	.631	.656	.640	-4.9	0.034
DT	창업	.624	.585	.735	.565	.639	.565	.661	.610	.622	.608	.615	.610	.587	.553	.559	.552	.555	.553	-5.6	0.135
	일반	.629	.593	.733	.572	.643	.576	.689	.638	.680	.628	.653	.637	.656	.616	.642	.610	.626	.616	-8.7	0.090
NB	창업	.580	.544	.882	.527	.659	.427	.593	.555	.758	.539	.630	.517	.589	.557	.754	.541	.630	.521	-2.3	0.005
	일반	.581	.571	.681	.558	.614	.560	.619	.575	.789	.553	.650	.534	.633	.590	.785	.565	.657	.557	-6.1	0.070
MLP	창업	.635	.603	.797	.575	.668	.572	.657	.601	.835	.569	.676	.553	.651	.594	.826	.564	.670	.546	-3.4	0.013
	일반	.678	.631	.710	.613	.658	.626	.750	.684	.792	.651	.714	.675	.704	.651	.731	.630	.677	.646	-9.5	0.130
SVM	창업	.629	.589	.666	.577	.618	.584	.691	.638	.698	.624	.659	.636	.642	.599	.658	.588	.621	.596	-9.0	0.108
	일반	.669	.616	.735	.594	.657	.604	.718	.659	.710	.644	.675	.657	.703	.646	.679	.637	.657	.645	-6.9	0.064
Bagging	창업	.555	.550	.621	.543	.580	.545	.755	.683	.603	.718	.655	.678	.564	.540	.476	.546	.509	.537	-26.4	1.271
	일반	.577	.548	.614	.543	.576	.544	.771	.699	.636	.727	.678	.696	.601	.571	.497	.584	.537	.566	-25.3	1.125
RF	창업	.633	.597	.707	.579	.637	.587	.684	.633	.682	.621	.650	.631	.635	.589	.634	.581	.606	.587	-7.5	0.085
	일반	.645	.614	.706	.596	.646	.607	.692	.638	.704	.622	.660	.634	.690	.636	.695	.621	.656	.633	-6.8	0.071
AdaBoost	창업	.636	.596	.685	.582	.629	.590	.673	.628	.722	.607	.660	.620	.649	.607	.696	.591	.639	.600	-5.5	0.035
	일반	.664	.615	.706	.597	.647	.608	.698	.642	.754	.616	.678	.633	.697	.639	.738	.616	.672	.632	-5.0	0.039
XGBoost	창업	.662	.613	.774	.585	.667	.591	.724	.655	.810	.618	.701	.636	.644	.595	.750	.573	.649	.575	-8.6	0.179
	일반	.663	.610	.733	.588	.653	.597	.723	.655	.801	.620	.699	.639	.694	.629	.766	.602	.674	.614	-8.3	0.090
LightGBM	창업	.651	.614	.790	.584	.672	.588	.687	.632	.800	.599	.685	.609	.643	.597	.763	.572	.654	.573	-5.3	0.054
	일반	.657	.605	.711	.587	.643	.596	.704	.644	.771	.614	.684	.631	.691	.635	.750	.610	.673	.625	-6.7	0.060
CatBoost	창업	.628	.585	.701	.569	.628	.573	.649	.609	.729	.588	.651	.598	.632	.594	.714	.576	.637	.582	-3.3	0.012
	일반	.652	.611	.714	.592	.648	.602	.679	.628	.756	.602	.671	.615	.690	.633	.748	.608	.671	.622	-4.0	0.038

주1) 이용된 자료는 창업기업에서 학습용 7,317개, 검증용 3,137개, 시험용 1,372개, 일반기업에서 학습용 17,010개, 검증용 7,290개, 시험용 2,744개이다. AUC값이 가장 높은 모형을 굵은 글씨로 표시하였다.

주2) LDA는 선형판별분석, LR은 로지스틱 회귀모형, DT는 의사결정트리, NB는 나이브 베이즈, MLP는 다층퍼셉트론, SVM은 서포트벡터머신, RF는 랜덤포레스트이다.

주3) 증감률은 학습용 대비 시험용의 AUC 변동률이다. 분산은 세 표본 AUC의 분산이다.

〈부표 2〉 3년 대출부도 예측 결과: 2020년 시험용 표본¹⁾

모형 ²⁾	대상	시험용						학습용						검증용						AUC ³⁾	
		AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	AUC	정확도	민감도	정밀도	F1스코어	기하평균	증감률(%)	분산×100
LDA	창업	.595	.554	.737	.539	.623	.523	.632	.598	.633	.592	.612	.597	.601	.574	.606	.569	.587	.573	-5.8	0.039
	일반	.650	.608	.672	.596	.632	.605	.662	.620	.688	.606	.644	.616	.649	.618	.691	.603	.644	.613	-1.7	0.005
LR	창업	.603	.570	.666	.558	.607	.561	.637	.599	.623	.595	.608	.599	.604	.569	.592	.566	.578	.569	-5.4	0.037
	일반	.650	.608	.728	.587	.650	.596	.661	.618	.680	.605	.640	.615	.649	.617	.686	.603	.642	.614	-1.6	0.004
DT	창업	.561	.549	.823	.532	.646	.476	.624	.594	.569	.598	.583	.593	.543	.536	.527	.537	.532	.536	-10.1	0.182
	일반	.594	.575	.766	.555	.643	.543	.646	.610	.711	.592	.646	.602	.604	.579	.680	.566	.618	.570	-8.2	0.078
NB	창업	.545	.542	.738	.530	.617	.505	.584	.551	.803	.534	.642	.490	.577	.540	.794	.526	.633	.476	-6.6	0.042
	일반	.582	.556	.770	.539	.634	.513	.599	.565	.785	.545	.643	.520	.605	.568	.790	.547	.646	.523	-2.9	0.015
MLP	창업	.500	.500	1.00	.500	.667	.000	.500	.500	1.00	.500	.667	.000	.500	.500	1.00	.500	.666	.000	0.0	0
	일반	.645	.598	.794	.571	.664	.565	.650	.611	.689	.596	.639	.606	.641	.608	.694	.593	.639	.602	-0.7	0.002
SVM	창업	.586	.571	.622	.565	.592	.569	.675	.621	.709	.603	.652	.615	.602	.575	.659	.564	.608	.569	-13.2	0.225
	일반	.659	.610	.717	.591	.648	.601	.689	.633	.706	.616	.658	.628	.662	.619	.692	.604	.645	.614	-4.3	0.026
Bagging	창업	.554	.539	.622	.533	.574	.532	.742	.674	.594	.707	.646	.669	.524	.514	.432	.517	.471	.508	-25.3	1.395
	일반	.553	.535	.617	.530	.570	.529	.886	.803	.762	.829	.794	.802	.565	.548	.492	.554	.521	.545	-37.6	3.562
RF	창업	.602	.565	.582	.563	.572	.565	.693	.632	.742	.609	.669	.623	.597	.552	.663	.542	.597	.540	-13.1	0.290
	일반	.646	.602	.636	.596	.615	.601	.676	.615	.738	.592	.657	.603	.646	.601	.729	.581	.647	.588	-4.4	0.030
AdaBoost	창업	.617	.577	.610	.572	.590	.576	.658	.617	.694	.602	.645	.612	.613	.579	.652	.569	.607	.574	-6.2	0.063
	일반	.658	.610	.715	.591	.647	.601	.670	.619	.761	.593	.667	.603	.651	.614	.754	.589	.662	.598	-1.9	0.010
XGBoost	창업	.610	.579	.716	.562	.629	.562	.680	.629	.788	.598	.680	.608	.597	.565	.715	.550	.622	.545	-10.3	0.200
	일반	.654	.606	.711	.587	.643	.597	.704	.649	.796	.615	.694	.632	.652	.613	.769	.586	.665	.593	-7.1	0.085
LightGBM	창업	.613	.577	.664	.566	.611	.571	.684	.638	.749	.613	.674	.628	.599	.565	.668	.554	.606	.556	-10.4	0.209
	일반	.652	.607	.721	.587	.647	.596	.678	.625	.773	.597	.673	.608	.649	.606	.759	.581	.658	.586	-3.9	0.027
CatBoost	창업	.598	.564	.772	.545	.639	.525	.642	.590	.833	.561	.670	.537	.604	.568	.803	.546	.650	.517	-6.8	0.055
	일반	.638	.595	.700	.578	.633	.585	.651	.605	.735	.583	.651	.591	.636	.602	.735	.581	.649	.587	-2.0	0.006

주1) 이용된 자료는 창업기업에서 학습용 5,727개, 검증용 2,455개, 시험용 1,322개, 일반기업에서 학습용 13,312개, 검증용 5,706개, 시험용 2,306개이다. AUC값이 가장 큰 모형을 굵은 글씨로 표시하였다.

주2) LDA는 선형판별분석, LR은 로지스틱 회귀모형, DT는 의사결정트리, NB는 나이브 베이즈, MLP는 다층퍼셉트론, SVM은 서포트벡터머신, RF는 랜덤포레스트이다.

주3) 증감률은 학습용 대비 시험용의 AUC 변동률이다. 분산은 세 표본 AUC의 분산이다.

〈부표 3〉 불균형 표본의 대출부도 예측 결과: 2020년 시험용 표본

모형	대상	자료		시험용					학습용					검증용				
		학습용	검증용	AUC	정확도	민감도	정밀도	F1스코어	AUC	정확도	민감도	정밀도	F1스코어	AUC	정확도	민감도	정밀도	F1스코어
판별분석 (LDA)	부도1년	11,121	4,767	.701	.665	.647	.396	.492	.659	.754	.073	.558	.129	.695	.755	.085	.564	.147
	부도2년	14,635	6,273	.641	.570	.660	.324	.434	.597	.751	.026	.522	.049	.630	.747	.027	.413	.051
	부도3년	11,454	4,910	.594	.523	.666	.297	.411	.563	.749	.003	.313	.007	.604	.752	.009	.733	.018
	부도1년	30,058	12,882	.734	.662	.722	.402	.517	.681	.760	.126	.598	.208	.736	.762	.127	.618	.211
	부도2년	34,020	14,580	.657	.606	.647	.346	.451	.619	.749	.035	.467	.065	.685	.748	.038	.463	.070
	부도3년	26,625	11,411	.647	.542	.732	.319	.444	.592	.749	.007	.412	.014	.646	.750	.011	.526	.021
로지스틱 회귀모형 (LR)	부도1년	11,121	4,767	.706	.684	.598	.410	.487	.653	.755	.052	.614	.096	.695	.754	.055	.591	.100
	부도2년	14,635	6,273	.639	.554	.676	.317	.431	.589	.750	.016	.475	.031	.629	.747	.015	.353	.029
	부도3년	11,454	4,910	.583	.510	.678	.293	.409	.555	.750	.000	.000	.000	.604	.750	.000	.000	.000
	부도1년	30,058	12,882	.734	.679	.688	.415	.518	.682	.763	.149	.609	.240	.740	.762	.146	.601	.235
	부도2년	34,020	14,580	.662	.621	.626	.354	.452	.623	.750	.034	.487	.063	.691	.750	.040	.500	.074
	부도3년	26,625	11,411	.641	.577	.669	.329	.441	.604	.750	.000	.500	.001	.643	.750	.001	1.00	.001
의사결정 트리 (DT)	부도1년	11,121	4,767	.650	.612	.596	.341	.434	.606	.750	.074	.502	.130	.642	.750	.081	.503	.140
	부도2년	14,635	6,273	.613	.474	.783	.293	.427	.539	.752	.037	.547	.069	.623	.749	.035	.482	.065
	부도3년	11,454	4,910	.555	.474	.649	.270	.382	.520	.750	.007	.583	.014	.540	.748	.002	.200	.005
	부도1년	30,058	12,882	.683	.603	.672	.348	.459	.624	.756	.078	.599	.138	.708	.756	.074	.590	.132
	부도2년	34,020	14,580	.629	.560	.667	.318	.431	.591	.750	.000	.000	.000	.662	.750	.000	.000	.000
	부도3년	26,625	11,411	.602	.546	.633	.304	.411	.572	.750	.015	.526	.030	.619	.750	.008	.451	.016
나이브 베이지스 (NB)	부도1년	11,121	4,767	.651	.594	.649	.338	.444	.611	.747	.146	.477	.224	.654	.749	.161	.495	.243
	부도2년	14,635	6,273	.595	.491	.698	.287	.407	.543	.738	.087	.390	.143	.600	.737	.093	.388	.150
	부도3년	11,454	4,910	.561	.477	.725	.285	.409	.534	.747	.015	.364	.029	.585	.750	.020	.500	.039
	부도1년	30,058	12,882	.678	.589	.711	.344	.464	.625	.704	.434	.412	.422	.682	.709	.434	.420	.427
	부도2년	34,020	14,580	.605	.524	.690	.302	.420	.569	.709	.278	.386	.323	.646	.711	.286	.394	.332
	부도3년	26,625	11,411	.600	.536	.657	.303	.415	.571	.706	.220	.357	.272	.612	.705	.212	.352	.265

모형	대상	자료		시험용					학습용					검증용							
		학습용	검증용	A U C	정확도	민감도	정밀도	F1 스코어	기하평균	A U C	정확도	민감도	정밀도	F1 스코어	기하평균	A U C	정확도	민감도	정밀도	F1 스코어	기하평균
다층 퍼셉트론 (MLP)	부도1년	11,121	4,767	.720	.673	.672	.407	.507	.673	.718	.765	.185	.594	.282	.420	.699	.766	.199	.594	.298	.436
	부도2년	14,635	6,273	.648	.588	.653	.334	.442	.608	.655	.750	.000	.000	.000	.000	.641	.750	.000	.000	.000	
	부도3년	11,454	4,910	.604	.547	.652	.308	.418	.577	.626	.750	.000	.000	.000	.000	.601	.750	.000	.000	.000	
	부도1년	30,058	12,882	.737	.707	.608	.438	.509	.671	.750	.769	.224	.603	.327	.462	.748	.770	.223	.608	.326	.461
	부도2년	34,020	14,580	.667	.641	.589	.365	.451	.623	.698	.750	.000	.667	.000	.015	.695	.750	.001	1.00	.001	.023
	부도3년	26,625	11,411	.657	.610	.624	.345	.444	.614	.658	.750	.000	.000	.000	.000	.652	.750	.000	.000	.000	
서포트 벡터머신 (SVM)	부도1년	11,121	4,767	.674	.633	.630	.364	.462	.632	.772	.758	.046	.749	.087	.214	.679	.752	.039	.575	.072	.196
	부도2년	14,635	6,273	.589	.595	.499	.309	.381	.559	.764	.750	.000	.000	.000	.000	.597	.750	.000	.000	.000	
	부도3년	11,454	4,910	.567	.586	.464	.293	.359	.539	.775	.750	.000	.000	.000	.000	.545	.750	.000	.000	.000	
	부도1년	30,058	12,882	.726	.700	.660	.434	.524	.686	.770	.766	.124	.672	.209	.348	.732	.765	.120	.669	.204	.344
	부도2년	34,020	14,580	.658	.659	.552	.376	.448	.619	.760	.750	.000	.000	.000	.000	.667	.750	.000	.000	.000	
	부도3년	26,625	11,411	.588	.620	.460	.319	.377	.556	.753	.750	.000	.000	.000	.000	.603	.750	.000	.000	.000	
배깅 (Bagging)	부도1년	11,121	4,767	.608	.571	.575	.308	.401	.573	.794	.769	.103	.786	.182	.319	.595	.742	.037	.355	.067	.190
	부도2년	14,635	6,273	.559	.631	.331	.291	.309	.492	.901	.790	.167	.965	.285	.409	.553	.748	.015	.390	.028	.121
	부도3년	11,454	4,910	.527	.627	.284	.268	.276	.459	.902	.786	.150	.964	.260	.387	.542	.745	.010	.255	.019	.098
	부도1년	30,058	12,882	.654	.663	.499	.370	.425	.598	.914	.804	.237	.927	.377	.485	.638	.756	.073	.589	.130	.268
	부도2년	34,020	14,580	.577	.643	.348	.310	.328	.508	.909	.793	.181	.959	.304	.425	.588	.748	.021	.411	.041	.146
	부도3년	26,625	11,411	.567	.655	.330	.317	.323	.502	.902	.786	.148	.959	.257	.385	.563	.746	.015	.339	.029	.122
랜덤 포레스트 (RF)	부도1년	11,121	4,767	.704	.671	.588	.394	.472	.641	.727	.755	.035	.727	.066	.185	.703	.755	.038	.692	.072	.194
	부도2년	14,635	6,273	.626	.556	.644	.312	.420	.582	.657	.750	.000	.000	.000	.000	.624	.750	.000	.000	.000	
	부도3년	11,454	4,910	.594	.531	.646	.298	.408	.564	.647	.750	.000	.000	.000	.000	.613	.750	.000	.000	.000	
	부도1년	30,058	12,882	.719	.657	.671	.392	.495	.662	.734	.758	.059	.691	.109	.242	.732	.758	.061	.679	.111	.245
	부도2년	34,020	14,580	.640	.609	.592	.338	.431	.603	.677	.750	.000	.000	.000	.000	.677	.750	.000	.000	.000	
	부도3년	26,625	11,411	.629	.549	.642	.308	.416	.577	.647	.750	.000	.000	.000	.000	.638	.750	.000	.000	.000	

모형	대상	자료		시험용					학습용					검증용							
		학습용	검증용	A U C	정 확 도	민 감 도	정 밀 도	F1 스 코 어	기하 평균	A U C	정 확 도	민 감 도	정 밀 도	F1 스 코 어	기하 평균	A U C	정 확 도	민 감 도	정 밀 도	F1 스 코 어	기하 평균
아다부스트 (AdaBoost)	부도1년	11,121	4,767	.722	.672	.647	.403	.496	.663	.730	.668	.690	.404	.510	.675	.714	.657	.659	.390	.490	.658
	부도2년	14,635	6,273	.611	.484	.778	.297	.430	.548	.619	.499	.757	.301	.430	.559	.604	.488	.735	.292	.418	.546
	부도3년	11,454	4,910	.570	.556	.545	.292	.380	.552	.592	.539	.642	.301	.410	.569	.580	.534	.616	.294	.398	.559
	부도1년	30,058	12,882	.739	.678	.681	.412	.514	.679	.748	.677	.697	.413	.519	.683	.750	.677	.700	.414	.520	.685
	부도2년	34,020	14,580	.669	.630	.619	.360	.455	.626	.701	.639	.667	.375	.480	.648	.700	.640	.663	.375	.479	.647
	부도3년	26,625	11,411	.608	.556	.638	.311	.418	.581	.623	.548	.700	.317	.436	.590	.618	.548	.688	.315	.432	.587
XGBoost	부도1년	11,121	4,767	.698	.642	.651	.375	.476	.645	.709	.621	.723	.369	.488	.652	.690	.614	.697	.360	.475	.639
	부도2년	14,635	6,273	.631	.569	.637	.319	.425	.590	.649	.584	.666	.334	.445	.609	.627	.566	.643	.318	.425	.589
	부도3년	11,454	4,910	.596	.500	.691	.290	.409	.549	.632	.501	.762	.302	.433	.562	.591	.492	.711	.290	.412	.546
	부도1년	30,058	12,882	.699	.611	.710	.359	.477	.641	.717	.608	.754	.363	.490	.649	.717	.608	.762	.364	.493	.651
	부도2년	34,020	14,580	.642	.598	.620	.335	.435	.605	.679	.612	.687	.357	.469	.635	.677	.609	.671	.352	.462	.628
	부도3년	26,625	11,411	.631	.546	.683	.313	.429	.585	.645	.533	.759	.316	.445	.587	.635	.534	.736	.315	.442	.586
LightGBM	부도1년	11,121	4,767	.696	.645	.651	.378	.479	.647	.710	.624	.725	.371	.491	.655	.688	.615	.694	.360	.474	.639
	부도2년	14,635	6,273	.629	.593	.599	.328	.424	.595	.651	.609	.618	.344	.442	.612	.628	.591	.592	.325	.420	.591
	부도3년	11,454	4,910	.587	.509	.637	.285	.393	.545	.629	.508	.741	.302	.429	.564	.582	.485	.684	.282	.399	.535
	부도1년	30,058	12,882	.707	.666	.655	.398	.496	.663	.723	.666	.692	.402	.509	.674	.718	.659	.683	.394	.500	.667
	부도2년	34,020	14,580	.644	.612	.596	.342	.435	.607	.684	.628	.670	.367	.474	.641	.678	.627	.660	.364	.469	.637
	부도3년	26,625	11,411	.631	.603	.598	.336	.430	.602	.648	.594	.656	.339	.447	.613	.640	.591	.640	.334	.439	.607
CatBoost	부도1년	11,121	4,767	.695	.649	.643	.380	.478	.647	.699	.614	.703	.361	.477	.641	.686	.599	.684	.347	.460	.625
	부도2년	14,635	6,273	.622	.542	.656	.306	.417	.575	.632	.564	.657	.319	.430	.592	.611	.552	.631	.307	.413	.576
	부도3년	11,454	4,910	.578	.482	.705	.284	.405	.536	.605	.481	.771	.294	.426	.544	.594	.484	.733	.289	.415	.542
	부도1년	30,058	12,882	.722	.664	.660	.397	.495	.663	.725	.668	.660	.400	.498	.665	.729	.668	.672	.402	.503	.669
	부도2년	34,020	14,580	.639	.626	.536	.342	.418	.593	.665	.642	.587	.366	.451	.623	.672	.639	.587	.363	.449	.621
	부도3년	26,625	11,411	.620	.560	.624	.311	.415	.580	.631	.563	.669	.321	.434	.594	.630	.560	.653	.316	.426	.588

참고문헌

- 김경숙·도영호·장영민(2014), “신생 중소기업의 헤저드모형을 이용한 산업별 생존요인에 관한 연구,” 「경영학연구」, 제43권 제1호, pp. 121-144.
- 김소정·이균희(2022), “심층신경망의 설명가능성과 하이퍼파라미터 특성에 관한 연구,” 「중소기업금융연구」, 제42권 제1호, pp. 3-36.
- 송현준·박도준·이준기(2021), “머신러닝을 이용한 외감기업 및 비외감기업의 부도예측에 관한 연구,” 「한국IT정책경영학회 논문지」, 제13권 제3호, pp. 2521-2527.
- 이상훈·유동희(2022), “더블 앙상블 기법을 이용한 기술사업성평가 기반의 중소기업 부실 예측 연구,” 「중소기업금융연구」, 제42권 제2호, pp. 3-28.
- Alaka, H.A., L.O. Oyedele, H.A. Owolabi, V. Kumar, S.O. Ajayi, O.O. Akinade, and M. Bilal(2018), “Systematic Review of Bankruptcy Prediction Models: Towards a Framework for Tool Selection,” *Expert Systems With Applications*, 94, pp. 164-184.
- Altman, E.(1968), “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *Journal of Finance*, 23, pp. 589-609.
- Altman, E. and E. Hotchkiss(2006), *Corporate Financial Distress and Bankruptcy*. Hoboken, New Jersey: Wiley.
- Bacham, D. and J. Zhao(2017), “Machine Learning: Challenges, Lessons, and Opportunities in Credit Risk Modeling,” *Moody’s Analytics Risk Perspectives*, 9, pp. 30-35.
- Barboza, F., H. Kimura, and E. Altman(2017), “Machine Learning Models and Bankruptcy Prediction,” *Expert Systems With Application*, 83, pp. 405-417.
- Bellovary, J.L., D.E. Giacomino, and M.D. Akers(2007), “A Review of Bankruptcy Prediction Studies: 1930 to Present,” *Journal of Financial Education*, 33, pp. 1-42.
- Breiman, L.(2001), “Random Forests,” *Machine Learning*, 45, pp. 5-32.
- Brown, I. and C. Mues(2012), “An Experimental Comparison of Classification Algorithms for Imbalanced Credit Scoring Data Sets,” *Expert System with Application*, 39, pp. 3446-3453.
- Chen, T. and C. Guestrin(2016), “XGBoost: A Scalable Tree Boosting System,” *Conference: the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.
- Cortes, C. and V. Vapnik(1995), “Support-Vector Networks,” *Machine Learning*, 20, pp. 273-297.
- Dastile, X., T. Celik, and M. Postane(2020), “Statistical and Machine Learning Models in Credit

- Scoring: A Systematic Literature Survey,” *Applied Soft Computing Journal*, 91, 106263, available at ScienceDirect.
- DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson(1988), “Comparing Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach,” *Biometrics*, 44(3), pp. 837–847.
- Demšar, J.(2006), “Statistical Comparisons of Classifiers over Multiple Data Sets,” *Journal of Machine Learning Research*, 7, pp. 1–30.
- Figini, S., F. Bonelli, and E. Giovannini(2017), “Solvency Prediction for Small and Medium Enterprises in Banking,” *Decision Support Systems*, 102, pp. 91–97.
- Freund, Y. and R.E. Schapire(1997), “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” *Journal of Computer and System Science*, 55, pp. 119–139.
- Friedman, M.(1940), “A Comparison of Alternative Tests of Significance for the Problem of m Rankings,” *Annals of Mathematical Statistics*, 11, pp. 86–92.
- Garcia, J.(2022), “Bankruptcy Prediction Using Synthetic Sampling,” *Machine Learning with Applications*, 9, 100343, available at ScienceDirect.
- Hand, D.J.(2009), “Measuring Classifier Performance: A Coherent Alternative to the Area Under the ROC Curve,” *Machine Learning*, 77, pp. 103–123.
- Jabeur, S.B., C. Gharib, S. Mefteh-Wali, and W.B. Arfi(2021), “CatBoost Model and Artificial Intelligence Techniques for Corporate Failure Prediction,” *Technological Forecasting and Social Change*, 166, 120658, available at ScienceDirect.
- Jones, S. and D.A. Hensher(2004), “Predicting Firm Financial Distress: A Mixed Logit Model,” *The Accounting Review*, 79, pp. 1011–1038.
- Jones, S., D. Johnstone, and R. Wilson(2015), “An Empirical Evaluation of the Performance of Binary Classifiers in the Prediction of Credit Ratings Changes,” *Journal of Banking and Finance*, 56, pp. 72–85.
- Kim, H-J., N-O. Jo, and K-S. Shin(2016), “Optimization of Cluster-based Evolutionary Undersampling for the Artificial Neural Networks in Corporate Bankruptcy Prediction,” *Expert Systems with Applications*, 59, pp. 226–234.
- Liang, D., C-C. Lu, C-F. Tsai, and G-A. Shih(2016), “Financial Ratios and Corporate Governance Indicators in Bankruptcy Prediction: A Comprehensive Study,” *European Journal of Operational Research*, 252, pp. 561–572.
- Liu, Y., Q. Zeng, B. Li, L. Ma, and J. Ordieres-Meré(2022), “Anticipating Financial Distress of High-tech Startups in the European Union: A Machine Learning Approach for Imbalanced

- Samples,” *Journal of Forecasting*, 41, pp. 1131–1155.
- Moscatelli, M., F. Parlapiano, S. Narizzano, and G. Viggiano(2020), “Corporate Default Forecasting with Machine Learning,” *Expert Systems With Applications*, 161, 113567, available at ScienceDirect.
- Mselmi, N., A. Lahiani, and T. Hamza(2017), “Financial Distress Prediction: The Case of French Small and Medium-sized Firms,” *International Review of Financial Analysis*, 50, pp. 67–80.
- Nazareth, N. and Y.V.R. Reddy(2023), “Financial Applications of Machine Learning: A Literature Review,” *Expert Systems With Applications*, 219, 119640, available at ScienceDirect.
- Nemenyi, P.B(1963)., *Distribution-free Multiple Comparisons*, PhD Thesis, Princeton University.
- Odom, M. and R. Sharda(1990), “A Neural Network for Bankruptcy Prediction,” *International Joint Conference on Neural Networks*, 2, pp. 163–168.
- Ohlson, J.A.(1980), “Financial Ratios and the Probabilistic Prediction of Bankruptcy,” *Journal of Accounting Research*, 18, pp. 109–131.
- Párik, M. and L. Papíková(2023), “Impacts of Crisis on SME Bankruptcy Prediction Models’ Performance,” *Expert Systems With Applications*, 214, 119072 available at ScienceDirect.
- Quinlan, J.R.(1986), “Introduction of Decision Tree,” *Machine Learning*, 1, pp. 81–106.
- Sigrist, F. and C. Hirnschall(2019), “Gabbit: Gradient Tree-based Tobit Models for Default Prediction,” *Journal of Banking and Finance*, 102, pp. 177–192.
- Son, H., C. Hyun, D. Phan, and H.J. Hwang(2019), “Data Analytic Approach for Bankruptcy Prediction,” *Expert System With Application*, 138, 112816, available at ScienceDirect.
- Veganzones, D. and E. Séverin(2018), “An Investigation of Bankruptcy Prediction in Imbalanced Datasets,” *Decision Support Systems*, 112, pp. 111–124.
- Zhang, L, J. Wang, and Z. Liu(2023), “What Should Lenders be More Concerned About? Developing a Profit-driven Loan Default Prediction Model,” *Expert Systems With Applications*, 213, 118938, available at ScienceDirect.
- Zięba, M., S.K. Tomczak, and J.M. Tomczak(2016), “Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction,” *Expert Systems With Applications*, 58, pp. 93–101.

