

A Study on Classification Models for Predicting Bankruptcy Based on XAI

Jihong Kim[†] · Nammee Moon^{††}

ABSTRACT

Efficient prediction of corporate bankruptcy is an important part of making appropriate lending decisions for financial institutions and reducing loan default rates. In many studies, classification models using artificial intelligence technology have been used. In the financial industry, even if the performance of the new predictive models is excellent, it should be accompanied by an intuitive explanation of the basis on which the result was determined. Recently, the US, EU, and South Korea have commonly presented the right to request explanations of algorithms, so transparency in the use of AI in the financial sector must be secured. In this paper, an artificial intelligence-based interpretable classification prediction model was proposed using corporate bankruptcy data that was open to the outside world. First, data preprocessing, 5-fold cross-validation, etc. were performed, and classification performance was compared through optimization of 10 supervised learning classification models such as logistic regression, SVM, XGBoost, and LightGBM. As a result, LightGBM was confirmed as the best performance model, and SHAP, an explainable artificial intelligence technique, was applied to provide a post-explanation of the bankruptcy prediction process.

Keywords : Big Data, Classification Model, eXplainable AI, Bagging, Boosting, Ensemble

XAI 기반 기업부도예측 분류모델 연구

김지홍[†] · 문남미^{††}

요약

기업 부도의 효율적인 예측은 금융기관의 적절한 대출 결정과 여신 부실률 감소 측면에서 중요한 부분이다. 많은 연구에서 인공지능 기술을 활용한 분류모델 연구를 진행하였다. 금융 산업 특성상 새로운 예측 모델의 성능이 우수하더라도 어떤 근거로 결과를 출력했는지 직관적인 설명이 수반되어야 한다. 최근 미국, EU, 한국 등에서는 공통적으로 알고리즘의 설명요구권을 제시하고 있어 금융권 AI 활용에 투명성을 확보하여야 한다. 본 논문에서는 외부에 오픈된 기업부도 데이터를 활용하여 인공지능 기반의 해석 가능한 분류 예측 모델을 제안하였다. 먼저 데이터 전처리 작업, 5겹 교차검증 등을 수행하고 로지스틱 회귀, SVM, XGBoost, LightGBM 등 10가지 지도학습 분류모델 최적화를 통해 분류 성능을 비교하였다. 그 결과 LightGBM이 가장 우수한 모델로 확인되었고, 설명 가능한 인공지능 기법인 SHAP을 적용하여 부도예측 과정에 대한 사후 설명을 제공하였다.

키워드 : 빅데이터, 분류모델, eXplainable AI, 배깅, 부스팅, 앙상블

1. 서론

금융회사들은 빅데이터 분석에 기반한 시장분석 및 금융상품개발, 리스크관리, 업무자동화 등 활발한 연구를 진행하고 있다. 은행은 자산건전성 확보를 위해 여러 지표들을 관리하고 있는데 그 중에서 대출 리스크는 중요한 관리지표 중 하나이다. 기업의 부도 상황은 회사 전체에 영향을 미치며 대출 자금 손실이 발생할 경우 은행에 높은 비용을 초래하게 된다[1].

그래서 기업을 위한 부도 예측 모델 개발은 필수적이며 이의 분류 성능은 중요하다. 기업부도예측 분류 성능이 낮은 경

우, 특히 부도난 회사를 정상적인 회사로 모델이 예측할 때, 오분류로 인한 비용은 금융기관 및 투자자들의 비용으로 이어지기 때문이다.

연구자들과 실무자들은 초기에 제한적인 통계지표를 활용한 다중 판별 분석[2]과 로지스틱 분석[3]을 사용하였다. 이후 컴퓨팅 기술의 발전에 따라 인공지능 및 기계학습 모델을 사용한 방법[4]이 개발되었고 기계학습 모델과 결합된 앙상블 방법을 사용하여 더 나은 분류성능을 확인한 연구[5]도 나왔다.

부도 예측에서는 무엇보다도 정확하고 해석 가능한 인공지능 기술(XAI, Explainable AI)을 사용하는 것이 중요하다 [6]. 해석 가능하다는 의미는 예측을 위해 사용되는 회계변수들의 계산 방법을 사용자가 알 수 있다는 의미이다.

부도 예측에서 인공지능 분류모델 성능비교에 대한 연구 [14-19]들과 설명력을 제공하는 모델 개발 연구[7], 설명기법에 대한 연구와 조사[8-10]는 많이 찾아볼 수 있으나, 분류 모델 성능비교 결과이후 사후 설명을 제공하는 연구는 찾아보기 쉽지 않았다.

※ 이 논문은 2022년 한국정보처리학회 ACK 2022의 우수논문으로 "XAI를 활용한 기업 부도예측 분류모델 연구"의 제목으로 발표된 논문을 확장한 것이다.

† 중신회원 : 호서대학교 융합공학과 박사과정

†† 중신회원 : 호서대학교 컴퓨터학부 교수

Manuscript Received : December 26, 2022

First Revision : March 24, 2023

Accepted : April 24, 2023

* Corresponding Author : Nammee Moon(nammee.moon@gmail.com)

본 논문은 기업 부도예측을 위한 전통적인 통계분석 방법과 기계학습의 새로운 기술들을 포함한 분류 모델을 구성하여 성능을 분석한다. 또한 부도예측 해석을 위해 특정 예측에 대한 변수 중요도를 확인하는 SHAP(SHapley Additive exPlanations)[11] 모델을 적용하여 결과에 대한 설명력을 제공한다.

이를 통해 금융 산업에서 요구되는 투명성과 신뢰성에 대응하고 인공지능 시스템 의사결정을 신뢰하면서 업무를 진행할 수 있도록 금융 전문가들을 지원하고자 한다.

2. 관련 연구

부도 예측 모델 연구분야에서는 위험을 나타내는 여러 회계비율 데이터를 사용하여 통계 또는 AI 기반으로 여러 머신러닝 알고리즘과 중요 기준들을 활용한 연구가 진행되고 있다. Barboza[12]는 배킹, 부스팅, Random Forest(RF) 모델이 다중판별분석(MDA), Logistic Regression(LR) 보다 성능이 우수하고 예측 정확도가 높다는 연구결과를 보여주었다.

Alaka[13]는 MDA, LR, Artificial Neural Network(ANN), Support Vector Machine(SVM), Rough Sets(RS), 사례기반 추론(CBR), Decision Tree(DT) 및 유전자 알고리즘(GA)의 8가지 분류모델을 비교 검토하였으며, 우수한 단일 모델은 없다는 결과를 도출하였다.

Jones[14]는 Gradient Boosting Model(GBM)이 MDA, LR모델의 문제점을 해결하고 우수한 예측성능을 보인다는 결과를 보여주었다.

최근 연구[15]에서는 앙상블 기법은 다변량 정규성, 다중공선성, 편향된 분류비율과 같은 기존 모델의 약점을 줄이기 위한 접근 방법이며, Gradient Boosting(GB)은 성능이 우수한 기계학습 기술이라는 결과를 보여주었다. Table 1은 부도예측에서 GB를 사용한 주요 연구들이다.

Table 1. Summary of Studies Using Gradient Boosting in Bankruptcy Prediction

Ref.	Training Models	Data	Best model
[14]	LR, GBM	U.S. public company bankruptcies	GBM
[15]	LR, SVM, NN, RF, GBM, XGBoost, CatBoost	financial data of French firms	CatBoost
[16]	SVM, HACT, XGBoost, DBN	financial data of Taiwanese firms	XGBoost
[17]	LR, RF, XGBoost	U.S. national commercial banks	XGBoost
[18]	LR, NN, GBM	Danish stock-based firms	NN
[19]	LR, RF, XGBoost, LightGBM, ANN	financial data of Korean companies	XGBoost, LightGBM

* 참고 : NN(Neural network), HACT(Hybrid associative classifier with translation), DBN(Deep Belief Network), ANN(Artificial neural network)

예측력이 우수한 분류 기계학습 모델을 신뢰할 수 있으면 높은 예측 성능을 유지하면서, 설명가능성이 필수적입니다. Hanif[1]는 은행과 금융권 시계열 데이터에 XAI를 적용한 대화형 증거기반 접근방식을 연구하였다.

3. 방법론 및 이론적 배경

본 논문에서는 잘 알려진 분류 기계학습 10개 기술(LR, SVM, KNN(K-Nearest Neighbor), NB(Naive Bayes), RF, DT(Decision Tree), XGBoost, LightGBM, CatBoost, NGBoost)을 사용하여 대만기업 부도 예측 성능을 비교 측정하였으며, 성능이 우수한 모델을 선정하여 분류 결과에 대한 충분한 사후 설명을 제공하기 위해 SHAP 모델을 사용하였다.

3.1 Logistic Regression

로지스틱 회귀분석(LR)[20]은 단순한 다변량 통계기법으로 여러 독립변수들과 범주형 종속변수 간의 관계를 학습시키고 로짓변환을 통해 분석하며, 이진 분류에 활용된다. 하지만 다중공선성에 민감한 상관관계가 높은 변수를 포함하는 것은 피해야 한다. 재무비율 변수들의 경우 일반적으로 변수들 간의 상관관계가 있다.

3.2 Support Vector Machine

SVM[21]은 데이터를 두 개의 클래스로 올바르게 분리하기 위해 여백(margin)을 최대화하는 초평면(hyperplane)을 구성하는 기술로 회귀와 분류작업에 사용한다. 적은 데이터로도 좋은 성능을 보이나 알고리즘 복잡도가 높아 메모리가 많이 필요하며, 테스트 단계에서 느린 단점이 있다. 최근 신용등급, 시계열 예측 및 사기 탐지와 같은 금융 애플리케이션에 사용되었다.

3.3 K-Nearest Neighbor

K-최근접 이웃(K-Nearest Neighbor, KNN)[22]은 다수결 규칙에 따라 테스트 데이터의 레이블을 예측한다. 즉, 두 데이터 간의 유사성을 계산하여 분류한다. 이때 유사성 측정을 위한 거리측정 방식과 k값 선택이 중요하다. KNN은 단순성과 효율성 때문에 분류, 회귀 등에서 많이 사용된다.

3.4 Naive Bayes

나이브 베이즈(NB)[23] 이론은 설명변수들 사이의 독립을 가정하는 베이즈 정리를 적용한 조건부 확률모델이다. 의사결정나무 및 신경망과 유사한 분류기술을 포함하며 빅 데이터를 적용할 때 뛰어난 정확도와 속도를 제공한다.

3.5 Random Forest

랜덤 포레스트[24]는 배깅(Bagging)과 랜덤노드 최적화의 앙상블 학습법과 의사결정 트리를 결합하여 데이터를 모델링하고 변수 중요도를 측정한다. 가장 가능성 있는 클래스를 다수의 투표로 정의하므로 더 정확한 예측을 제공하고 가장 중

요한 것은 데이터 과적합을 방지하는 부분이다. 신용평가 및 부도예측에 널리 사용되고 있다.

3.6 Decision Tree

의사결정나무(DT)[25]는 순서도 같은 나무 구조와 유사한 분류기로 인간 추론과 유사한 특성을 가지고 있으며 의사결정 정보를 얻기 위해 사용된다. DT는 루트 노드에서 시작하여 학습 알고리즘에 따라 각 노드를 재귀적으로 분할한다. 많이 사용하는 알고리즘은 ID3, C4.5, CART 이다.

3.7 XGBoost

XGBoost[26]는 분산 환경에서 효과적인 데이터 처리를 위해 과적합을 제어하고 병렬학습을 수행하여 높은 정확도를 보여주는 확장가능한 트리 부스팅 시스템이다. XGBoost는 여러 기계 학습 및 데이터 마이닝 대회에서 사용되었으며 뛰어난 성능으로 널리 알려져 있다.

3.8 LightGBM

Gradient Boosting Decision Tree(GBDT)[27]는 널리 알려진 기계학습 알고리즘이다. 하지만 Feature 차원이 높고 데이터 크기가 큰 경우 정보획득을 위해 모든 데이터를 검색함으로써 많은 작업시간이 필요하다. 이런 문제들을 해결하기 위해 LightGBM[28]은 정보이득(Information gain)을 효율적으로 추정하는 기울기 기반 무작위 샘플링(Gradient-based One-Side Sampling)과 대규모 Feature 수를 다루기 위해 상호배타적 Feature 번들(Exclusive Feature Bundling)을 사용하여 빠른 계산속도와 메모리 측면에서 높은 효율성과 정확도를 보여주고 있다.

3.9 CatBoost

CatBoost[29]는 타겟누수(target leakage)를 방지하기 위해 Ordered boosting 구현과 범주형 Feature를 처리하기 위한 새로운 알고리즘이다. 트리의 다형성과 오버피팅 문제를 Ordered Boosting, 무작위 순열 등 내부 알고리즘으로 해결하고 있어서 매개변수 조정을 통한 결과와는 크게 차이가 없다. 최근 금융과 같은 다양한 분야에서 적용되고 있다.

3.10 NGBoost

NGBoost[30]는 2019년 발표되었으며 확률적 예측을 위한 모듈식 부스팅 알고리즘이다. 기본 학습자, 연속 매개변수 분포계열, 채점 규칙으로 구성되어 있으며, 확률적 예측을 위한 기존 방법에서 유연성과 확장성을 제공한다.

3.11 SHAP

XAI[10]는 인공지능의 최종 결과를 인간이 더 이해하기 쉽게 만드는 것을 목표로 하는 연구분야로서 XAI라는 용어는 형식적인 기술 개념이라기 보다는 AI 투명성과 신뢰문제에 대응하기 위한 노력을 의미하기도 한다.

유럽연합[31]은 2018년 5월, 데이터 주체에게 알고리즘의 투명성을 제공하는 일반 개인정보보호 규정(General Data

Protection Regulation)을 제정하였고, 우리나라는 2018년 5월 I-Korea 4.0 실현을 위한 인공지능 R&D전략[32]에서 설명가능 학습, 추론을 핵심기술로 포함하여 정책을 추진하고 있다.

XAI는 여러 가지 형태로 구현할 수 있다. 모델 자체가 설명력을 가지고 있지 않은 경우 사후 해석을 통해 해결하는 방법이 보편적이다. 사후 해석모델 중의 하나인 SHAP[11]은 모델 불가지론(model agnostic)의 로컬 대리(local surrogate)기법으로 모델 유형에 상관없이 예측에 대한 변수 중요도(개별 입력변수 기여도)를 Shapley Value로 측정하여 종속변수와 상관계수를 해석한다. 즉, 설명하고자 하는 분류모델의 조건부 예측함수의 Shapley Value이다. 이를 기반으로 제시된 SHAP 값은 각 독립변수들이 분류 모델에 얼마나 기여하는지 측정한다.

$$\phi_i = \sum_{S \subset F_i} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (1)$$

Shapley Value(ϕ_i)는 Equation (1)로 표현된다. 여기서 F 는 모든 독립변수들의 집합이며, S 는 F 의 부분집합이다. $f_{S \cup \{i\}}$ 는 독립변수 i 가 포함된 학습모델이며, f_S 는 독립변수 i 를 포함하지 않은 학습모델이다. 두 모델 예측은 현재 입력 $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ 에서 비교하여 독립변수 i 의 기여도를 산출하고, 각 부분집합 S 에 대해 가중평균하여 계산하게 한다.

SHAP은 해석범위에 따라 개별 입력 값에 대한 결과를 설명하는 지역적(Local) 해석과 데이터 전체에 대한 예측 결과를 설명하는 전역적(Global) 해석으로 구분한다.

4. 실험 및 결과 분석

4.1 데이터 전처리

기업부도 데이터[33]는 대만경제저널에 제공된 6,819개 기업데이터를 사용하였으며, 영업총이익, 영업이익률, 영업비율, 주당순가치 등 95개 독립변수로 구성되어 있고 종속변수는 부도여부(bankruptcy value 0/1)이다. 데이터에 포함된 기업들의 사업 분야는 제조, 서비스 및 기타 산업들로 구성되어 있다.

기계학습 알고리즘들은 데이터 분포의 정규성을 가정하므로 이를 위해서 데이터 표준화 작업 및 독립변수들의 결측치를 확인하고 정제하는 과정을 진행하였다.

전체 데이터 중 부도기업은 220개로 3.2% 비율을 나타내고 있다. 데이터 셋이 불균형인 경우[34] 대부분의 분류 알고리즘 성능은 크게 떨어진다. 데이터 분류 비율의 차이가 크면 우세 분류를 선택하는 모델의 정확도가 높아지기 때문이다. 이를 해결하기 위해 오버 샘플링, 언더 샘플링 및 SMOTE 같은 데이터 수준의 접근 방법을 적용(Fig. 1)하였으며, 복합샘플링을 통해 생성된 데이터는 Table 2에서 볼 수 있듯이 부도 6,256개(53.6%), 정상 5,426개이다.

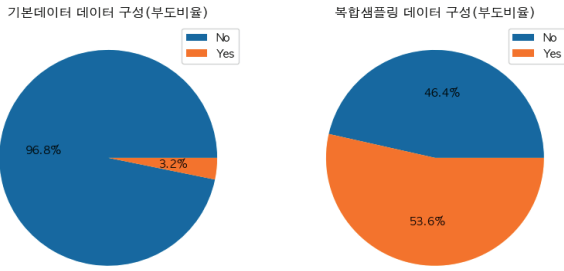


Fig. 1. Complex Sampling

Table 2. Data Configuration

		Normal	Bankrupt	Total
Raw Data		6,599	220	6,819
Complex Sampling Data	Train	4,355	4,990	9,345
	Test	1,071	1,266	2,337
	Sum	5,426	6,256	11,682

4.2 분류 모델

일반적인 기계학습 알고리즘은 매개변수와 하이퍼파라미터를 사용한다. LR, SVM, NB, DT의 경우 대부분의 매개변수를 기본값으로 설정하여도 분류성능에 영향을 미치지 않았다. 그러나 KNN, RF, XGBoost, LightGBM, CatBoost, NGBoost의 경우 기본적인 매개변수로는 높은 성능을 보이는데 한계가 있었다. 분류성능을 개선하고 과적합 문제를 피하기 위해 그리드 검색을 사용하여 최적 매개변수를 조정하였다. 그리드 검색[35]은 하이퍼파라미터를 조정하는데 사용하는 최적화 기술유형으로 가능한 모든 매개변수를 결합하여 각 분류 모델에 대한 최상의 조합을 찾는 방법이다.

4.3 평가 방법

모델 성능에 영향을 미칠 수 있는 데이터의 변동성과 편향을 최소화하기 위해서 5겹 교차검증[36]을 적용(Fig. 2)하였다.

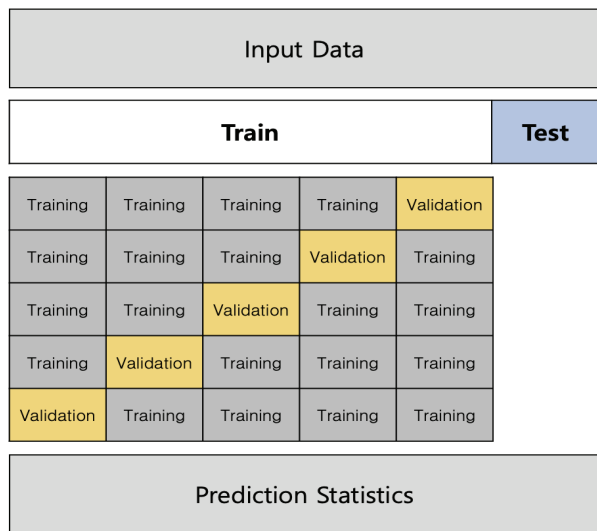


Fig. 2. 5-Fold Cross Validation

Table 3. The Parameters of Classification Models

Model	Parameters
LR	max_iter=5000
SVM	Kernel=linear
KNN	n_neighbors=5, metric=minkowski
NB	var_smoothing=0.14796880626863962
DT	criterion=entropy
RF	criterion=entropy, n_estimators=10
XGBoost	n_estimators=500, colsample_bytree=0.5, max_depth=7, min_child_weight=1
LightGBM	boosting=gbdt, max_bin=512, num_leaves=30, learning_rate=0.03
CatBoost	learning_rate=0.1
NGBoost	n_estimators=1000, learning_rate=0.03, Base=default_tree_learner, Score=MLE

먼저 전체 데이터를 학습데이터와 테스트데이터로 분리한다. 여기서 학습데이터는 5개 폴드로 무작위로 균등하게 나누고, 1개의 폴드는 다시 5개로 쪼갠 다음, 검증데이터 셋 1개를 제외한 나머지 4개의 훈련데이터 셋으로 모델을 학습시킨다. 다음 폴드에서는 검증데이터 셋을 바꿔서 지정하고 첫 번째 폴드와 동일한 방법으로 학습을 진행하게 된다. 5개 폴드에 대한 테스트 결과를 평균 계산하여 최적의 매개변수를 선정한다. Table 3은 분류모델별로 선정된 최적의 매개변수들이다.

마지막으로 모델별 예측 성능비교를 위해 전체 학습데이터를 이용하여 최적의 매개변수를 적용한 모델들을 학습시킨 후 테스트데이터에 대한 모델의 예측 성능을 측정하고 비교 작업을 수행하였다.

4.4 평가 지표

다양한 분류 모델의 예측성능을 평가하기 위해 정확도(Accuracy), F1-Score와 ROC(Receiver Operating Characteristic)곡선 아래 면적(Area Under the Curve -AUC) 세 가지 평가지표를 사용하였다[37].

정확도는 분류 성능을 측정하기 위해 필요한 지표 중 하나이며, 다음과 같이 전체 샘플 수에 대한 올바르게 분류된 샘플 간 비율로 정의한다.

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \tag{2}$$

여기서 P와 N은 Positive와 Negative 샘플 수를 나타내며, TP는 True Positive, TN은 True Negative, FP는 False Positive, FN는 False Negative를 나타낸다.

F1-Score는 Equation (3)처럼 Precision과 Recall의 조화평균으로 표현된다. F1-Score 값은 0에서 1까지이며 F1-Score 값이 클수록 분류성능이 높다는 것을 나타낸다.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

ROC곡선은 이진분류 진단 능력을 나타내는 그래프이다. 임계값이 다양할 때 예측 모델을 설명해준다. AUC는 분류모델 평가시 널리 사용되며 다음과 같이 정의된다.

$$AUC = \int_0^1 TRP(t_i) dFPR(t_i) \quad (4)$$

여기서 $TPR(t_i)$ 및 $FPR(t_i)$ 는 각각 True Positive Rate, False Positive Rate를 나타낸다.

4.5 결과 분석

분류모델별 예측 성능은 5점 교차검증을 통해 선정된 매개변수를 적용하여 분류모델을 훈련한 후 테스트 데이터에 대한 모델의 예측성능을 평가하였다. 최종 예측결과는 Table 4에 나타내었다.

본 연구에서 사용하는 세 가지 평가지표 중 하나의 지표가 우수하다고 그 분류모델 성능이 좋다고 판단하기는 어렵다. 하지만 본 연구에서는 정확도가 높은 경우 F1-Score도 높게 나왔으며, AUC도 1에 가까운 결과를 보여주었다.

부스팅 분류모델들은 전반적으로 느린 학습속도 결과를 보여주었으며, 그 결과, 최적 매개변수 조정을 위해 더 많은 시간이 필요했다. 그러나 부스팅 모델들의 예측성능은 우수하였다.

CatBoost는 매개변수 조정시 오히려 분류성능이 저하되는 결과를 보여주었다. NGBoost는 분류성능은 우수하지만 다른 분류모델에 비해 약 13분에 이르는 매우 긴 수행시간을 기록하였으며, 매개변수 조정 및 학습데이터 규모에 따라 수행시간이 급격하게 증가하는 단점을 보여주었다. LightGBM은 매개변수 조정에 따라 성능결과에서 많은 차이를 보여주었다.

모델별 분류성능 비교시 LightGBM은 세 가지 평가지표에서 다른 분류모델 보다 높은 예측성능을 보였다. 그 다음으로는 CatBoost, XGBoost, RF, DT, NGBoost 순으로 높은 예측성능을 보이고 있다. 시각적 이해를 돕기 위해 모델별 예측성능 비교 결과를 Fig. 3으로 구성하였다. 가장 높은 성능을 보여준 LightGBM은 정확도 0.9914, F1-Score 0.9922, AUC 0.9907이다.

Table 4. Comparison of Test Set Performance by Classification Models

Model	Accuracy	F1-score	AUC	Execution Time (Sec)
LR	0.9247	0.9316	0.9227	0.3
SVM	0.9285	0.9351	0.9265	4.2
KNN	0.9521	0.9576	0.9477	0.5
NB	0.8716	0.8881	0.8653	0.1
DT	0.9581	0.9617	0.9568	0.8
RF	0.9799	0.9817	0.9784	4.1
XGBoost	0.9863	0.9875	0.9851	5.2
LightGBM	0.9914	0.9922	0.9907	12.4
CatBoost	0.9876	0.9887	0.9865	7.1
NGBoost	0.9576	0.9617	0.9555	744.5

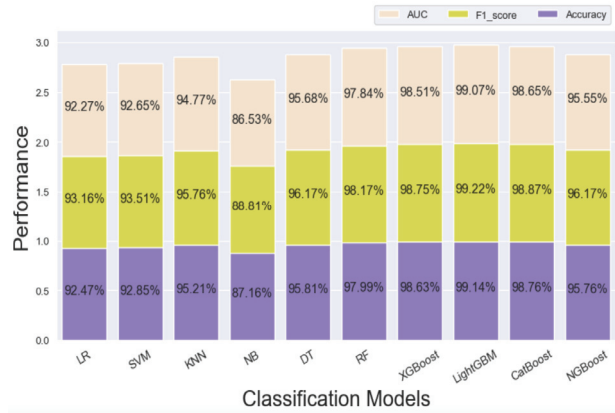


Fig. 3. Prediction of Classification Models

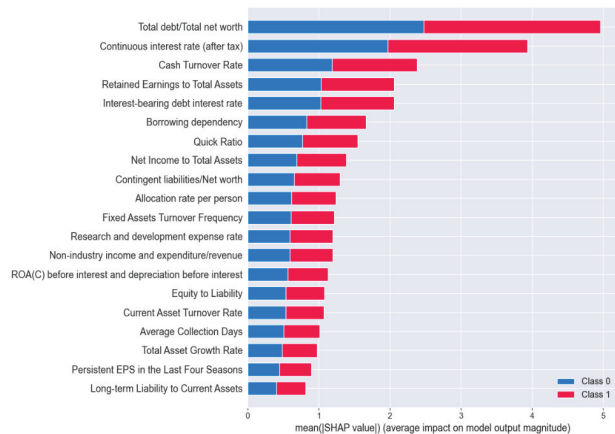


Fig. 4. Feature Importance by Mean(|SHAP Value|)

본 연구는 분류모델 사후 설명력 강화를 위해 우수한 분류성능을 기록한 LightGBM 모델에 SHAP을 적용하여 모델 예측에 대한 시각화 및 설명을 추가로 진행하였다.

Fig. 4는 LightGBM에 SHAP을 적용하여 분류예측 과정에 대한 설명을 부여한 것으로, 테스트데이터 2,337개에서 95개 독립변수들의 SHAP value 절댓값의 평균, 즉 변수 중요도를 내림차순으로 상위 20개를 표현한 것이다. 절댓값 평균이 가장 높은 변수로는 총 부채/총 순자산(Total debt/Total net worth), 세후연속이자율(Continuous interest rate-after tax), 현금회전율(Cash Turnover Rate), 총자산 대비 이익잉여금(Retained Earnings to Total Assets), 이자부 부채이자율(Interest-bearing debt interest rate) 순이며, 이 변수들은 기업부도 분류에 많은 영향을 미쳤음을 설명하고 있다.

기업부도에 영향을 미친 변수들은 SHAP summary plot에서 좀 더 세부적으로 확인할 수 있다(Fig. 5). 색상대비가 명확한 변수들을 대상으로 분석하면 총자산 대비 이익잉여금(Retained Earnings to Total Assets)과 총자산대비 당기순이익(Net Income to Total Assets)이 높을수록, 차용 의존성(Borrowing dependency)과 이자부 부채이자율이 낮을수록 기업부도 가능성이 낮으며, SHAP value가 낮게 나타났다. 총자산 대비 이익잉여금이 낮을수록, 차용 의존성(Bor-

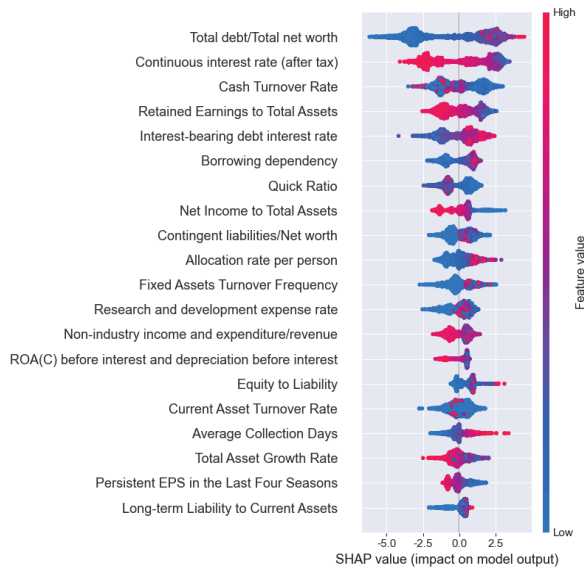


Fig. 5. Summary Plot of SHAP Values

rowing dependency)이 높을수록 기업부도 가능성이 높으며 SHAP value는 높게 나타났다. 색상이 고르게 분포되어 있는 변수들은 다른 변수들과의 연관성에 따라 결과가 다르게 나옴을 예측할 수 있다.

Fig. 6에서 독립변수들의 SHAP 의존도와 다른 변수와의 상호작용을 확인할 수 있다. dependence_plot은 먼저 특정 독립변수의 분포와 SHAP 추정값을 분포한 후 추가로 연결된 다른 독립변수와의 관계를 보여주는 산점도이다. x축은 총자산 대비 이익잉여금이고, y축은 총자산 대비 이익잉여금의 SHAP 추정값이다. 색상은 오른쪽 y축에 있는 세후연속이자율 값을 표현한 것으로, 이 값이 클수록 붉은색으로 표시된다. 총자산 대비 이익잉여금이 0.5~2 인 경우 SHAP 추정값은 대부분 음수(-)이며 정상기업으로 예측한다. 이때 세후연속이자율은 0.02~0.06 구간의 색상으로 표시되고 있다. Fig. 6은 거의 뚜렷한 수직 색상패턴으로 두 변수(총자산 대비 이익잉여금과 세후연속이자율) 간 상호작용 효과가 있음을 알 수 있다.

또한, SHAP은 특정 기업데이터에 중요한 영향을 준 독립변수들을 분석할 수 있다. 이를 위해 SHAP value를 분해하고 시각화하였다(Fig. 7). force_plot에서 기준값(base value: -0.04832)은 모든 예측의 평균이며, SHAP 추정값은 -21.48이다. 각 독립변수들은 부도 예측가능성을 높이거나 낮추는 데 영향을 준다. 해당 데이터 A 기업은 예측값과 실제값이 모두 정상 기업이다. 기업부도에 긍정적인 영향을 준 요인은 자기 자본대비 부채(Equity to Liability)이며, 부정적인 영향을 준 요인은 총 부채/총 순자산, 세후연속이자율, 현금회전을 순으로 분석되었다.

다음은 부도기업에 대한 시각화 분석이다(Fig. 8). B 기업의 예측값과 실제값은 모두 부도 상태로 나타났으며, SHAP 추정값은 10.72 이다. 부도에 긍정적인 영향을 준 요인은 총 부채/총 순자산, 세후연속이자율, 차용 의존성 순이며, 부정적인 영향을 준 요인은 유동자산 장기부채(Long-term Liability to Current Assets)로 분석되었다.

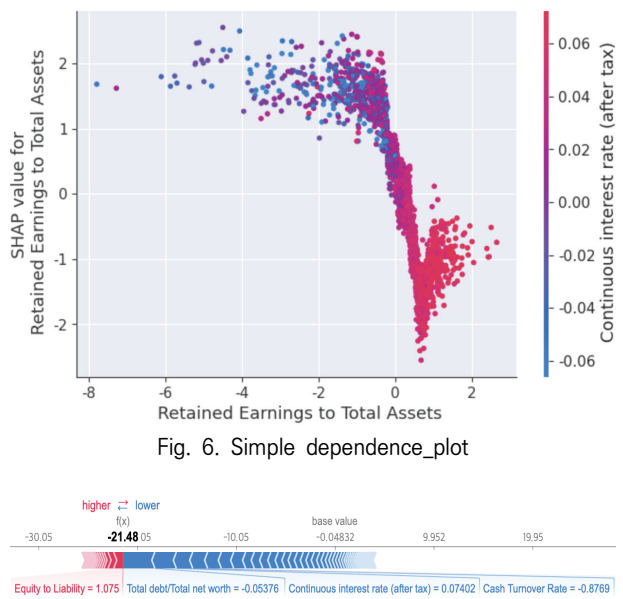


Fig. 6. Simple dependence_plot

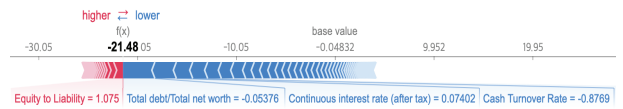


Fig. 7. force_plot of Individual Company(A)

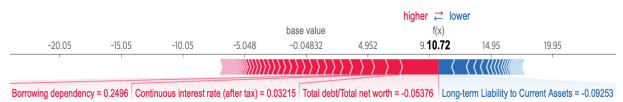


Fig. 8. force_plot of Individual Company(B)

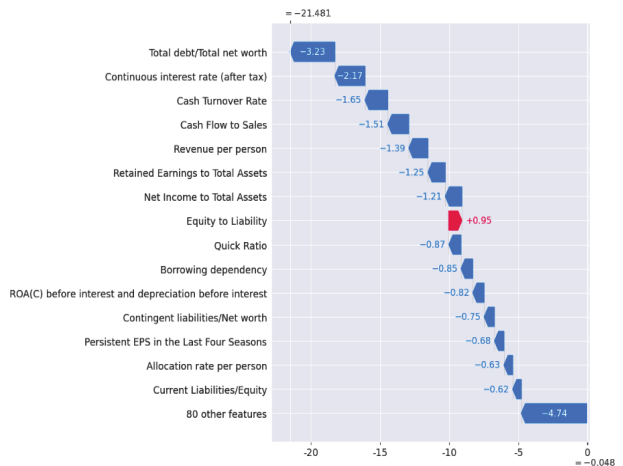


Fig. 9. waterfall_plot of Individual Company(A)

A, B 기업의 부도 예측 결과에 대한 판단 근거를 보여주기 위해 추가 작업을 진행하였다(Fig. 9, Fig. 10). waterfall_plot은 각 독립변수들의 기여도와 SHAP 추정값을 직관적으로 표현할 수 있다. 맨 아래 기준값(-0.048)에서 시작하여 각 변수들의 기여도를 긍정(+), 부정(-)으로 표시하여 SHAP 추정값(-21.481, 10.716)에 이르는 과정을 시각적으로 보여준다. Fig. 9의 A기업은 총 부채/총 순자산, 세후연속이자율, 현금회전을 항목이 기업부도에 부정적으로 기여하고 있음을 확인할 수 있다. Fig. 10의 B기업은 총 부채/총 순자산, 세후연속이자율, 차용 의존성 항목이 기업부도에 긍정적으로 기여하고 있음을 알 수 있다.



Fig. 10. waterfall_plot of Individual Company(B)

5. 결론 및 향후 과제

본 논문에서는 계량적 재무지표를 기반으로 기업부도 예측 분류 알고리즘들을 정리하였으며, 그 중에서도 의사결정나무, 배깅, 부스팅 계열의 알고리즘을 사용하였다. 기업부도 예측에 사용한 데이터는 1999년~2009년 동안 수집된 대만기업의 부도 데이터를 이용하였으며, 5겹 교차검증을 적용하여 훈련, 검증 및 테스트 과정을 진행하였다. 그 결과 모델예측 성능은 LightGBM이 가장 우수한 것으로 나타났다. 그리고 우수 모델의 사후 해석을 위해 로컬 대리(local surrogate)기법 중의 하나인 SHAP을 적용하여 개별 독립변수들의 모델 예측에 대한 중요도와 독립변수들 간 상관관계 등을 분석하고, 시각화 작업을 통해 부도 예측에 대한 설명력을 제공하였다.

향후 분류모델 성능 고도화와 예측결과에 대한 설명가능 요약 자동화 연구를 통해 데이터 전처리에서부터 모델 예측 결과 설명에 이르는 분류예측모델 전주기를 자동화하는 시스템을 제시하고자 한다.

References

- [1] A. Hanif, "Towards explainable artificial intelligence in banking and financial services," *arXiv preprint arXiv: 2112.08441*, 2021.
- [2] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, Vol.23, No.4, pp.589-609, 1968.
- [3] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, pp.109-131, 1980.
- [4] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, Vol.20, pp.273-297, 1995.
- [5] G. Wang, J. Hao, J. Ma, and H. Jiang, "A comparative assessment of ensemble learning for credit scoring," *Expert Systems with Applications*, Vol.38, No.1, pp.223-230, 2011.
- [6] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, Vol.2, No.1, pp.56-67, 2020.
- [7] T. N. Chou, "An explainable hybrid model for bankruptcy prediction based on the decision tree and deep neural network," In *2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII)*, IEEE, pp. 122-125, 2019.
- [8] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you? Explaining the predictions of any classifier," In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135-1144, 2016.
- [9] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM computing surveys (CSUR)*, Vol.51, No.5, pp.1-42, 2018.
- [10] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, Vol.6, pp.52138-52160, 2018.
- [11] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, Vol.30, 2017.
- [12] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, Vol.83, pp.405-417, 2017.
- [13] H. A. Alaka, L. O. Oyedele, H. A. Owolabi, V. Kumar, S. O. Ajayi, O. O. Akinade and M. Bilal, "Systematic review of bankruptcy prediction models: Towards a framework for tool selection," *Expert Systems with Applications*, Vol.94, pp.164-184, 2018.
- [14] S. Jones, "Corporate bankruptcy prediction: A high dimensional analysis," *Review of Accounting Studies*, Vol.22, pp.1366-1422, 2017.
- [15] S. B. Jabeur, C. Gharib, S. Mefteh-Wali, and W. B. Arfi, "CatBoost model and artificial intelligence techniques for corporate failure prediction," *Technological Forecasting and Social Change*, Vol.166, pp.120658, 2021.
- [16] Y. P. Huang and M. F. Yen, "A new perspective of performance comparison among machine learning algorithms for financial distress prediction," *Applied Soft Computing*, Vol.83, pp.105663, 2019.
- [17] P. Carmona, F. Climent, and A. Momparler, "Predicting failure in the US banking sector: An extreme gradient boosting approach," *International Review of Economics & Finance*, Vol.61, pp.304-323, 2019.
- [18] R. Matin, C. Hansen, C. Hansen, and P. Mølgaard, "Predicting distresses using deep learning of text segments in annual reports," *Expert Systems with Applications*, Vol.132, pp.199-208, 2019.

- [19] H. Son, C. Hyun, D. Phan, and H. J. Hwang, "Data analytic approach for bankruptcy prediction," *Expert Systems with Applications*, Vol.138, pp.112816, 2019.
- [20] S. B. Jabeur, "Bankruptcy prediction using partial least squares logistic regression," *Journal of Retailing and Consumer Services*, Vol.36, pp.197-202, 2017.
- [21] Y. Wu, Y. Xu, and J. Li, "Feature construction for fraudulent credit card cash-out detection," *Decision Support Systems*, Vol.127, pp.113155, 2019.
- [22] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for knn classification," *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol.8, No.3, pp.1-19, 2017.
- [23] M. I. Habibie and N. Nurda, "Performance analysis and classification using naive bayes and logistic regression on big data," In *2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS)*, IEEE, pp.48-52, 2022.
- [24] L. Breiman, "Random forests," *Machine Learning*, Vol.45, pp.5-32, 2001.
- [25] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," *arXiv preprint arXiv:1203.3832*, 2012.
- [26] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785-794, 2016.
- [27] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, pp.1189-1232, 2001.
- [28] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, Vol.30, 2017.
- [29] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," *Advances in Neural Information Processing Systems*, Vol.31, 2018.
- [30] T. Duan, A. Anand, D. Y. Ding, K. K. Thai, S. Basu, A. Ng, and A. Schuler, "Ngboost: Natural gradient boosting for probabilistic prediction," In *International Conference on Machine Learning*, PMLR, pp.2690-2700, 2020.
- [31] G. D. P. Regulation, "General data protection regulation (GDPR)," Intersoft Consulting, Accessed in October, 24.1, 2018.
- [32] Ministry of Science and ICT, "Artificial Intelligence (AI) R&D Strategy for the Realization of I-Korea 4.0," 2018.
- [33] D. Liang, C. C. Lu, C. F. Tsai, and G. A. Shih, "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study," *European Journal of Operational Research*, Vol.252, No.2, pp.561-572, 2016.
- [34] J. Sun, H. Li, H. Fujita, B. Fu, and W. Ai, "Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting," *Information Fusion*, Vol.54, pp.128-144, 2020.
- [35] J. Fan, X. Wang, F. Zhang, X. Ma, and L. Wu, "Predicting daily diffuse horizontal solar radiation in various climatic regions of China using support vector machine and tree-based soft computing models with local and extrinsic climatic data," *Journal of Cleaner Production*, Vol.248, pp.119264, 2020.
- [36] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.32, pp.3, pp.569-575, 2009.
- [37] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, Vol.17, pp.1, pp.168-192, 2021.



김 지 흥

<https://orcid.org/0000-0002-1380-5700>

e-mail : jh.wisard@gmail.com

1994년 경희대학교 수학과(학사)

2016년 한양사이버대학교 IT MBA(석사)

2021년 ~ 현 재 호서대학교 융합공학과 박사과정

2009년 ~ 현 재 IBK시스템 혁신사업본부

관심분야: 빅데이터, 딥러닝, XAI, NLP, 금융공학



문 남 미

<https://orcid.org/0000-0003-2229-4217>

e-mail : nammee.moon@gmail.com

mnm@hoseo.edu

1985년 이화여자대학교 컴퓨터공학(학사)

1987년 이화여자대학교 컴퓨터공학(석사)

1998년 이화여자대학교 컴퓨터공학(박사)

1999년 ~ 2000년 아주대학교 디지털미디어학과 조교수

2000년 ~ 2003년 이화여자대학교 조교수

2003년 ~ 2008년 서울벤처정보대학원대학교 디지털미디어학과 부교수

2008년 ~ 현 재 호서대학교 컴퓨터학부 교수

관심분야: 빅데이터 처리 및 분석, 딥러닝, HCI, Social Learning