

머신러닝 기반 KOSDAQ 시장의 관리종목 지정 예측 연구*

윤양현**

광운대학교 경영학부, 학부

김태경***

광운대학교 경영학부, 부교수

김수영****

광운대학교 수학과, 학부

박용균*****

광운대학교 수학과, 학부

국 문 요 약

관리종목 지정 제도는 상장 기업 내 기업의 부실화를 경고하여 기업에게는 회생 기회를 주고, 투자자들에게는 투자 위험을 경고하기 위한 시장규제 제도이다. 본 연구는 관리종목과 비관리종목의 기업의 재무 데이터를 표본으로 하여 관리종목 지정 예측에 대한 연구를 진행하였다. 분석에 쓰인 분석 방법은 로지스틱 회귀분석, 의사결정나무, 서포트 벡터 머신, 소프트 보팅, 랜덤 포레스트, LightGBM이며 분류 정확도가 82.73%인 LightGBM이 가장 우수한 예측 모형이었으며 분류 정확도가 가장 낮은 예측 모형은 정확도가 71.94%인 의사결정나무였다. 대체적으로 앙상블을 이용한 학습 모형이 단일 학습 모형보다 예측 성능이 높았다.

핵심어: KOSDAQ, 관리종목, 머신러닝, 앙상블

1. 서론

일반적으로 벤처기업은 기업의 지속적인 발전을 위한 단계로 상장기업이 되기를 희망한다. 그러나 실제로 많은 기업들이 오랜 기간 존속되지 못하고 부실의 단계로 이어지는 경우가 있다. 한국 벤처 환경에서는 부실화 가능성이 충분한 벤처기업의 경우 부실기업으로 곧바로 지정되기보다 관리종목으로 지정함으로써 채도약의 기회를 부여하는 제도가 마련되어 있다. KRX(Korea Exchange, 한국거래소)에서는 상장폐지기준에 해당하는 상장기업들을 관리종목으로 지정하고, 기업이 경영을 지속 혹은 존속하기 힘든 기업들을 관리종목 지정 제도를 통하여 투자자들에게 공시한다. 관리종목으로 지정이 되면 해당 기업은 신용거래 대상에서 제외되거나 증권시장에서 일정기간 매매거래가

정지되는 등의 불이익을 받게 되고, 투자자들 입장에서는 관리종목 지정 이후 주가가 하락하면서 투자 손실을 입게 된다. 따라서 관리종목 지정은 기업과 투자자 입장에서 많은 손실과 부정적인 결과를 일으키는 요인이 될 수 있기 때문에 관리종목 지정 예측 연구에 대해 중요시할 필요가 있다. 국내, 해외 모두 기업의 부도예측에 관한 연구는 많지만, 상장기업 내 부실화 가능성이 높은 기업에 대한 경고를 하는 국내 시장규제 제도인 관리종목 지정 예측에 대한 연구는 단적으로 찾아보기 어렵다(신동인·곽기영, 2018).

본 연구는 이러한 점을 고려하여 기업의 재무 데이터를 통해 관리종목 지정 예측을 할 수 있는 모델을 제안하고, 다양한 머신러닝 기법을 활용하여 예측 모델의 비교 분석을 진행한다. 관리종목 지정 경향성이 있는 기업을 조기에 파악하고 부정적인 결과를 미연에 예방하는 정책 수립 등

* 이 논문은 2020년도 광운대학교 교내학술연구비 지원에 의해 연구되었음(2020-0323)

** eb3434@naver.com,

*** kimtk@kw.ac.kr,

**** sooyoung6262@naver.com,

***** az8928@naver.com,

에 기여하고자 다양한 예측 모델들을 검증하였다.

II. 연구 배경

당해 법인이 상장폐지기준에 해당되었을 경우 KRX는 상장된 기업을 폐지할 수 있게 된다. 그러나 해당 기업이 일시적으로 상장폐지기준에 해당할 가능성이 있기에 즉각 상장폐지조치를 취하면 기업이 회생할 기회도 없이 증권 시장에서 퇴출 당하게 되는 결과를 초래하게 된다. 또한, 해당기업의 투자자들에게 투자액 회수의 기회도 상실할 수 있는 문제점도 가지게 된다. 관리종목 지정은 바로 이러한 문제점들을 해결하기 위해 내리지는 조치이다(표영인·김일, 2002). 과거의 선행연구에서 관리종목 공시 이전부터 주가가 하락함을 알 수 있으며(김태혁·엄철준, 1997), 관리종목 지정 자체가 시장에서 부정적인 정보로 인식됨을 알 수 있다(손성규·오명진, 2008).

머신러닝 기법을 적용하여 기업의 부도를 예측하려는 기업부도예측모델(Bankruptcy Prediction Model: BPM)에 관한 연구를 수행한 결과를 찾아보기 어렵지는 않지만(Barboza et al., 2017; Devi & Radhika, 2018; 김형준 외, 2019; 엄하늘 외, 2020), 관리종목 지정을 예측하는 연구는 상대적으로 적은 관심을 받고 있다. 본 연구는 선행연구(신동인·곽기영, 2018)로 진행되었던 방식들, 가령 로지스틱 회귀분석이나 의사결정나무와 같은 것과 함께 서포트 벡터 머신(support vector machine: SVM), 소프트 보팅, 랜덤 포레스트, LightGBM 분석과 같은 다양한 접근을 통해 결과를 도출하고 상호 비교하였다.

III. 데이터 및 결과

에프앤가이드의 DataGuide 5.0을 활용하여 K-IFRS가 적용된 2011년부터 2020년까지의 코스닥 기업의 재무 데이터를 수집하였다. 관리종목으로 편입된 시점의 해당 년도 재무 데이터를 기준으로 결측치 혹은 극단치가 없는 관리종목 표본 347개, 1:1 쌍대표본으로 추출한 비관리종목 표본 347개, 도합 692개 기업을 표본으로 선정하였다. 관리종목 지정 예측을 하기 위해 활용된 독립변수는 기업의 수익성, 안정성, 활동성, 성장성을 보여주는 지표인 해당 재무비율 21개를 선정하여 분석을 진행하였다.

Python을 활용하여 분석을 하였으며, 오픈소스 라이브러리인 Scikit-learn을 주로 참조하였다. 클라우드 기반 주피터 노트북 개발환경인 Google Colab을 사용하였다. 학습 데이터 분류와 모델 학습기에 관련된 내용은 <표 1>에 제시하

였다.

전체 모형에 대한 종합 결과는 <표 2>에 나와 있으며, 6가지 예측 모형들의 분류 정확도는 70% 초반부터 80% 초반 사이로 의사결정나무 모형이 71.94%로 가장 낮았으며 LightGBM 모형이 82.73%로 가장 높은 정확도를 보였다. 모형의 평가기준은 정확도만으로 평가하면 편향된 결과를 초래할 수 있기에 다양하게 고려를 해야 하는데, 특히 기업이 관리종목으로 지정되면 야기될 부정적인 효과를 고려해봤을 때 비관리종목을 관리종목으로 예측하는 손실이 상당히 클 수 있다. 따라서 모형 비교 및 평가 시 전체 정확도 확인뿐만 아니라 정밀도(Precision), AUC의 비교를 통해 모형들의 예측 성능을 평가할 수 있다.

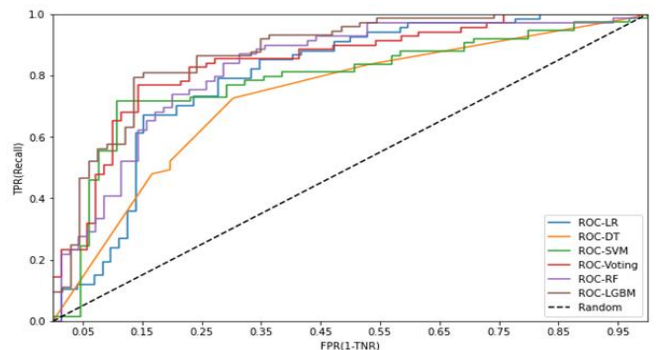
<표 1> 머신러닝 모형 비교

분류	모델1	모델2	모델3	모델4	모델5	모델6
	로지스틱 회귀분석	의사결정 나무	서포트벡터 머신	소프트 보팅	랜덤 포레스트	부스팅
분류기	LR	DT	SVM	LR, RF, SVM	RF	LightGBM
학습 데이터*	(555,21)	(555,21)	(555,21)	(555,21)	(555,21)	(555,21)
테스트 데이터	(139,21)	(139,21)	(139,21)	(139,21)	(139,21)	(139,21)
정규화 (normalization)	정규화	비정규화	정규화	정규화	비정규화	비정규화

*데이터셋은 각 행과 열에 해당하는 수치이다. 데이터셋을 전체 692개 기업 중 80%를 학습 데이터로, 나머지 20%는 테스트 데이터로 활용하였다.

<그림 1>은 모형의 ROC 곡선을 비교한 결과이며, 대체적으로 앙상블 학습 모형이 단일 학습 모형보다 우수하다는 것을 입증할 수 있었다. 관리종목이 지정되는 요인인 재무변수에 대해 세세하게 알아보기 위해, 의사결정나무 기반 모형으로 산출된 변수 중요도 상위 3개를 <표 3>에 제시하였다.

재무변수 중요도 확인 결과, ROA(영업이익), ROE(당기순이익), 자본금회전율, 자기자본현금흐름률, 매출액증가율이 관리종목 지정에 중요한 변수로 선정되었다.



<그림 1> 전체 모형 ROC 곡선 종합 비교

<표 2> 모형 종합 결과

분류	Accuracy	5-fold acc. mean	Precision	Recall	F1-score	Kappa	ROC-AUC
로지스틱 회귀분석	0.7338	0.7802	0.6923	0.8060	0.7448	0.4700	0.7956
의사결정 나무	0.7194	0.7784	0.7297	0.7397	0.7347	0.4370	0.7389
서포트 벡터 머신	0.7410	0.7658	0.7794	0.7162	0.7465	0.4827	0.7848
소프트 보팅	0.7842	0.7838	0.7600	0.8261	0.7917	0.5686	0.8435
랜덤 포레스트	0.7482	0.7928	0.8036	0.6522	0.7200	0.4957	0.8259
LightGBM	0.8273	0.7712	0.8657	0.7945	0.8286	0.6553	0.8719

<표 3> 변수 중요도 상위 3개

의사결정나무	랜덤 포레스트	LightGBM
P7*-ROE(Net income)	P7-ROE(Net income)	A2-Capital stock turnover ratio
A2-Capital stock turnover ratio	S2-Cashflow(comm) to total equity	G3-Sales growth rate
P1-ROA(Operating income)	A2-Capital stock turnover ratio	P7-ROE(Net income)

*P는 수익성(Profitability), S는 안정성(Stability), A는 활동성(Activity), G는 성장성(Growth)으로 재무비율에 대한 약어를 의미한다.

IV. 마무리

기업에 회생 기회를 부여하고, 투자자가 투자위험에 대한 인지를 조기에 할 수 있도록 마련된 관리종목 지정 제도는 중요하다. 그럼에도 불구하고 충분한 연구가 부족하다는 점, 선행연구의 결과가 다소 제한된 머신러닝 방법에 의존한다는 한계가 있다. 이러한 점들을 고려했을 때 본 연구는 연구의 필요성과 현실적 요구를 충족시키는 결과들을 제시했다는 점에서 의의가 있다. 투자자들이 관리종목 지정을 예측하도록 하고 투자 포트폴리오 리스크 관리 혹은 리밸런싱에 대한 참고자료를 제공한다는 측면에서 실무적 활용 가능성도 높다고 본다. 한편 연구의 결과들은 한계점과 함께 이해될 필요가 있다. 가령, 기업의 재무제표를 토대로 한 재무비율 등을 주요 변수들로 사용하였기에 관리종목 지정 사유인 불성실공시, 자본잠식, 감사의견 등에 대한 비재무적 요인에 대한 변수를 충분히 고려하지 못했다. 후속 연구를 통해 비재무적 요인을 추가로 반영하여 분석이 진행되기를 기대한다.

참고문헌

김태혁·엄철준(1997). 관리대상종목의 수익률과 위험 속성에 관한 연구. *증권 금융연구*, 3(1), 93-133.
 김형준·류두잔·조훈(2019). 기업부도예측과 기계학습. *金融工學*

研究 18(3), 131-152.
 손성규·오명천(2008). 관리종목 기업의 회계정보 효과. *연세경영연구*, 45(2), 127-146.
 신동인·곽기영(2018). KOSDAQ 시장의 관리종목 지정 탐지 모형 개발. *지능정보연구*, 24(3), 157-176.
 엄하늘·김재성·최상욱(2020). 머신러닝 기반 기업부도위험 예측 모델 검증 및 정책적 제언: 스택킹 앙상블 모델을 통한 개선을 중심으로. *지능정보연구*, 26(2), 105-129.
 표영인·김일(2002). 관리종목지정 시점에 발생하는 산업내 정보 전도효과. *경영학연구*, 31(3), 751-767.
 Barboza, F., Kimura, H., & Altman, E.(2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417.
 Devi, S. S., & Radhika, Y.(2018). A survey on machine learning and statistical techniques in bankruptcy prediction. *International Journal of Machine Learning and Computing*, 8(2), 133-139.