

## 마켓 타이밍과 유상증자\*

서성원  
건국대학교 경영학과 부교수

# Market Timing and Seasoned Equity Offering

Sung Won Seo<sup>a</sup>

<sup>a</sup>Department of Business Administration, Konkuk University, South Korea

Received 29 February 2024, Revised 20 March 2024, Accepted 24 March 2024

### Abstract

**Purpose** - In this study, we propose an empirical model for predicting seasoned equity offering (SEO here after) using machine learning methods.

**Design/methodology/approach** - The models utilize the random forest method based on decision trees that considers non-linear relationships, as well as the gradient boosting tree model. SEOs incur significant direct and indirect costs. Therefore, CEOs' decisions of seasoned equity issuances are made only when the benefits outweigh the costs, which leads to a non-linear relationship between SEOs and a determinant of them. Particularly, a variable related to market timing effectively exhibit such non-linear relations.

**Findings** - To account for these non-linear relationships, we hypothesize that decision tree-based random forest and gradient boosting tree models are more suitable than the linear methodologies due to the non-linear relations. The results of this study support this hypothesis.

**Research implications or Originality** - We expect that our findings can provide meaningful information to investors and policy makers by classifying companies to undergo SEOs.

**Keywords:** Issuance Costs, Machine Learning, Market Timing, Non-Linear Relations, SEO

**JEL Classifications:** G30, G32

## I. 서론

경영자의 주식 매도 결정이 주가에 유의미한 영향을 미치는 것을 종종 볼 수 있다. 신규 상장한 직후에 주가가 불안정할 때 경영자들이 주식을 매도하면 이러한 현상이 더욱 두드러지게 나타날 수 있다. 기업의 내부와 외부는 정보 비대칭이 있으므로 기업의 외부 투자자들은 기업 내부에서 발생하는 많은 정보를 알 수 없다. 따라서 외부에서는 기업 경영자들의 의사결정을 중요한 정보를 내포한 신호(signaling)로 보게 된다. 외부 투자자들은 기업의 내부자들이 개인 소유의 주식을 처분하는 경우에 기업이 고평가되어 있다는 신호로 보기 때문에 주가에 부담으로 작용하게 되는 것이다.

\* 본 연구는 미래에셋자산운용 신진학자 연구지원을 받았음.

<sup>a</sup> First Author, E-mail: seosw@konkuk.ac.kr

© 2024 The Institute of Management and Economy Research. All rights reserved.

경영자 개인 소유의 주식 매도와 마찬가지로 기업의 신규 주식을 발행하여 판매하는 유상증자도 같은 논리가 적용 가능하다. Masulis and Korwar (1986)과 Spiess and Afleck-Graves (1995)에서 보는 바와 같이 유상증자는 시장에서 악재로 보며 장기 및 단기 주가 수익률도 나쁘게 나타나는 것을 볼 수 있다. 유상증자도 기업의 입장에서 주식을 신규로 발행하여 판매하는 것이기 때문에 경영자 개인 소유의 주식 매도와 마찬가지로 외부 투자자들의 입장에서는 주가의 고평가를 이용하려는 것으로 볼 수 있다. 따라서, 본 연구에서는 머신 러닝 방법을 이용하여 유상증자를 예측하는 모형을 구축하였다. 기존에 주로 사용하던 선형 모형과는 다르게 비선형 관계를 잘 고려하는 머신 러닝 방법을 이용한다. 특히, 머신 러닝 방법론의 대표적인 의사결정 나무 모형에 기반한 랜덤 포레스트 방법과 그래디언트 부스팅 트리 모형을 이용하여 분석하였다.

Leary & Roberts (2005)와 Strebulaev (2007)에서 지적하는 바와 같이 기업의 부채비율 조정은 높은 비용이 발생하며, Lee, Lochhead, Ritter, and Zhao (1996)에서 보고하는 바와 같이 유상증자도 높은 비용이 발생한다. 따라서 필요한 비용보다 충분히 더 많은 이득일 발생할 때만 유상증자를 결정하게 된다. 예를 들어서 주가가 과대평가된 것을 이용하여 주식을 발행한다고 하면, 과대평가가 조금 되어있거나 과소평가일 때는 아무런 반응이 없다가 충분히 과대평가되어서 유상증자의 비용을 모두 지불 가능하다고 경영자가 판단할 때만 기업에서 유상증자를 결정한다. 이러한 비선형 관계를 고려하기 위해서는 기존에 사용하던 선형 모형 방법론은 적합하지 않으며, 의사결정 나무 모형 기반의 랜덤 포레스트와 그래디언트 부스팅 트리 모형이 적합하다.

본 연구에서는 마켓타이밍 관련 변수와 유상증자의 비선형적 관계에 주목한다. Baker and Wurgler (2002)는 경영자들이 주가가 높게 평가될 때 주식을 발행하는 현상을 보고한다. 시장가 대 장부가의 비율을 고평가의 지표로 보고, 시장가 대 장부가가 높을 때 외부에서 자본 조달을 많이 한 기업들은 주로 주식을 통하여 외부에서 자본 조달하는 현상을 찾았다. Baker and Wurgler (2002)에서는 이러한 현상을 마켓 타이밍이라고 부른다. 따라서 유상증자 의사결정에서 마켓 타이밍은 중요한 요인이다.

DeAngelo, DeAngelo, & Stulz (2010)에서는 기업의 연령이 유상증자에 미치는 영향에 대해서 연구했다. 기업들은 꼭 투자해야 하는 중요한 투자안이 있는데 내부 자금이 부족하여 주식을 발행하는 경우도 많다. DeAngelo, DeAngelo & Stulz (2010)에서는 기업의 연령이 증가하면 외부 자본 조달의 필요성이 점점 줄어들게 되어 유상증자를 적게 하게 된다는 것을 실증 연구를 통하여 확인했다. 이와 같은 기존의 문헌에서는 유상증자의 주요 결정 요인에 대해서 연구하는데, 주로 선형 회귀분석을 사용한다. 본 연구에서는 머신러닝 방법론을 이용하여 비선형적인 관계를 고려하면 이러한 모형의 설명력이 높아지는지 분석한다.

김수인·변진호 (2016)에서는 대주주 거래와 유상증자의 관계에 관해서 연구했다. 유상증자 전에 대주주의 지분이 증가한 경우에는 유상증자 이후의 주가 수익률이 높게 나타났다. 또한 임병권·윤평식 (2017)에서는 유상증자 공시 전 정보거래에 대해서 분석한다. 그 결과, 유상증자 공시일 이전에 기관 투자자와 외국인 투자자들에게서 비정상적인 거래량이 발생하는 것을 발견했다. 신현한·윤평식 (2023)에서는 유상증자 전 내부자 거래에 대해서 분석하여 유상증자 전에 내부자의 순매도가 증가하는 것을 보여준다. 박형진·홍종운 (2019)에서는 유상증자 방식과 장기성과에 관해서 연구한다. 주주배정의 방식이 일반공모나 제3자 배정에 비해서 장기성과가 좋다는 결과는 보여준다. 이는 회사 내부자들이라고 볼 수 있는 주주들이 회사에 자금을 투입하는 경우에 장기성과가 좋아진다는 것을 의미한다.

주요 설명 변수로는 *기업의 연령*, *과대 평가를 나타내는 시장가 대 장부가*, 그 외에 기업 특성 및 거시 경제 상황을 고려하는 변수들을 포함하였다. 그 결과, 선형 모형에 비해서 랜덤 포레스트 모형과 그래디언트 부스팅 트리 모형이 더 높은 설명력을 나타냈다. 이러한 관계는 종속 변수를 유상증자의 규모로 했을 때도 잘 나타났으며 유상증자를 했는지를 나타내는 지시변수를 사용했을 때도 비슷하게 나타났다. 또한 코스피와 코스닥에 상장되어 있는 기업들을 각각 나누어 분석했을 때도 각각의 그룹에 대해서 모두 비선형 모형이 선형 모형보다 높은 설명력을 가지는 것으로 나타났다. 분석 결과, 상장된 지 얼마 안된 기업, 같은 산업의 다른 기업들에 비해서 시장가 대 장부가 비율이 높은 기업들이 유상증자를 할 확률이 높았다.

Hovakimian (2006)와 Baker and Wurgler (2002)에서와 같이 자기 자본을 총자산 - 총부채 - 우선주 + 이연된 세금으로 정의했다. 유상증자 의사 결정을 설명하기 위한 주요 변수는 *기업의 연령, 시장가 대 장부가 비율, 영업이익률, 순이익률*과 같은 수익성 변수들과 그 외의 기업의 특성을 반영하는 변수들, 그리고 주식 시장의 상황을 알려주는 *경제 심리지수*와 같은 여러 변수를 고려하였다. 특히, *경제 심리 지수*는センチ멘트 투자자들이 얼마나 주식 시장에 활발하게 참여하는지 측정하는 변수로 사용하였다.センチ멘트 투자자들이 주식 시장에 활발히 참여하면, 그만큼 주식을 발행하는 것이 수월하기 때문에 기업의 입장에서 유상증자를 수월하게 할 수 있다.

최근에 머신러닝 방법론을 이용하여 예측 모형을 만드는 연구들이 생겨나고 있지만, 경영자의 의사결정 예측에 이러한 방법론을 이용하려는 논문은 아직 충분하지 못하다. 경영자의 의사결정은 개인의 의사결정 이라기 보다는 기업의 장단기 명운을 결정하는 중요한 의사결정이기 때문에 여러 가지 요인이 복합적으로 영향을 미친다. 따라서 설명 요인과 의사결정이 단순히 선형의 관계가 있다고 보기 어렵고 비선형 관계를 고려해야 한다. 본 연구는 기존 연구에서 사용한 선형 관계의 방법론에서 벗어나 설명 변수와 유상증자 결정의 비선형 관계를 고려하여 설명력을 높였다는 점에서 기여도가 있다. 또한 투자론적인 측면에서도 외부 투자자들이 어떤 기업이 유상증자를 할 확률이 높은지를 미리 파악한다면, 악재를 피할 수 있는 좋은 정보로 사용할 수 있을 것으로 보인다.

본 연구의 2장 이후의 내용은 다음과 같다. 2장에서는 본 연구에서 사용한 랜덤 포레스트와 그라디언트 부스팅 방법론에 관해서 설명한다. 3장에서는 주요 연구 결과를 소개하고, 4장에서는 추가 분석 결과를 보여준다. 추가 분석에서는 세부적으로 코스피와 코스닥 기업들을 각각 나누어서 분석하고, 표본 외 분석 결과와 변수의 중요도를 보여준다. 마지막으로 5장의 결론에서는 본 연구의 결과를 요약 및 정리하고 시사점에 관해서 서술한다.

## II. 연구 방법론

랜덤 포레스트 방법론과 그라디언트 부스팅 분석은 모두 의사결정 트리 모형(Decision tree model)에 기반하고 있다. 의사결정 트리 모형은 전체 데이터를 작은 세부 그룹으로 분류하여 비슷한 그룹끼리 같은 의사결정을 한다는 전제로 방법론이 구성되어 있다. Hastie et al (2009)에서는 임계점을 기준으로 각각의 세부 그룹을 만들고 예측값과 실제값의 차이를 최소화하는 방향으로 점점 조정하여 최적의 기준점을 찾는 방법이라고 설명하고 있다. 유상증자할 때는 많은 비용이 소요되기 때문에 기업의 입장에서 소폭 과대평가가 되어있다면, 과대평가로 인한 이득이 유상증자로 인한 비용보다 작다. 이런 상황에서는 유상증자를 실행할 수 없으므로, 어떠한 기준점을 넘어서야 유상증자를 결정하게 된다. 이러한 과대 평가와 유상증자 확률간의 비선형 관계를 분석하기 위해서는 의사결정 트리 모형에 기반한 랜덤 포레스트 방법과 그라디언트 부스팅 방법이 적합하다.

### 1. 랜덤 포레스트 방법

랜덤 포레스트 방법은 의사결정 트리 모형에 기반하지만, 자료 분석의 에러를 줄이고 과적합(overfitting)을 최소화하도록 구성되어 있다. 의사결정 트리 모형은 전체 샘플에 대해서 분석을 하지만, 랜덤 포레스트는 전체 샘플에서 랜덤하게 여러 개의 트리를 만들어서 분석하고 분석된 여러 개의 의사결정 트리에서 도출되는 값을 평균하여 의사결정을 하게 된다. 따라서 자료에 포함된 에러를 줄이고 머신러닝 방법론에서 흔히 발견되는 과적합 문제를 완화하는데 도움이 된다.

## 2. 그라디언트 부스팅 방법

그라디언트 부스팅 방법은 여러 가지 모형의 앙상블로 새로운 예측 모형을 만들어 낸다. 예를 들어서 하나의 트리 모형으로 예측 모형을 구성하고 남은 잔차를 예측하기 위해서 다른 트리 모형을 만들어서 다시 예측하는 형태이다. 다만, 이 과정에서 과적합(overfitting)되는 경우가 있다는 우려가 있다. 따라서 본 연구에서는 과적합을 완화하기 위하여 교차 검증, 무작위 추출 검증, 표본의 분석과 같은 여러 가지 검증 방법을 사용한다.

## III. 연구 결과

### 1. 자료 및 주요 변수 설명

주요 변수는 장부가 자본의 변화량( $netequity = (\text{장부가 자본량}[t] - \text{장부가 자본량}[t-1]) / \text{자산}$ ), 산업 평균을 고려한 시장가 대 장부가 비율 ( $mbadj$ , 시장가 대 장부가 비율 / 산업 평균 시장가 대 장부가 비율), 기업의 연령( $age$ , 상장 이후 경과 연도), 회사의 크기 ( $size$ ,  $\log$  (총자산)), 유동자산 ( $caa$ , 유동자산 / 총자산), 부채비율 ( $lev$ , 총부채/총자산), 영업이익률( $oi$ , EBIT/매출액)과 같은 기업 특성 측정치를 고려했다. 또한, 시장 상황 관련 변수는  $senti$ (소비자 심리 지수)을 통제 변수로 추가했다. 샘플 분석 기간은 2003년부터 2019년까지이다. 그리고 극단적인 값을 가지는 수치(outlier)에 의해서 결과가 달라지는 것을 막기 위하여 1%와 99%에서 윈저라이징(winsorization)을 했다. 한국의 KOSPI와 KOSDAQ에 상장된 모든 기업에 대해서 분석했다.

표1에서는 각각의 변수에 대한 주요 통계량을 보여준다. 산업 평균 대비 시장가대 장부가 비율은 평균이 0.943이며 하위 1%는 0.097으로 산업 대비 약 10분의 1정도의 시장가대 장부가 비율을 가지는 기업이었 고, 상위 1%는 7.095으로 산업 평균대비 7배 높은 시장가대 장부가 비율을 가지는 기업이었다. 기업이 상장된 연령은 평균적으로 14.03년 정도였으며, 99%에 해당하는 기업의 연령은 47년이었다. 시장 센티멘트 지수는 약 99.7의 평균값을 가지며, 상위 1%의 값은 약 112.7이고 하위 1%의 값은 약 77.6을 나타내고 있다.

**Table 1.** Summary Statistics

Variable	Obs	Mean	Std. Dev.	1%	99%
$netequity[t]$	23,086	0.033	0.133	-0.843	0.776
$Mb\ adj\ [t-1]$	23,086	0.943	0.887	0.097	7.095
$Age\ [t-1]$	23,086	14.034	10.912	0.000	47.000
$Oi\ [t-1]$	23,086	0.019	0.372	-9.523	0.670
$Size\ [t-1]$	23,086	25.775	1.390	20.258	30.418
$Caa\ [t-1]$	23,086	0.465	0.194	0.020	0.991
$Lev\ [t-1]$	23,086	0.394	0.205	0.022	1.098
$Senti\ [t-1]$	23,086	99.686	8.698	77.600	112.700

아래의 표2에서는 주요 변수들 간의 상관관계를 보여준다. 주식 발행량은 직전 연도의 기업의 *연령*과 음의 상관관계를 가지고 있어서 기업의 *연령*이 적을수록 유상증자를 많이 하는 것으로 나타났다. 이는 기업의 *연령*이 적을수록 상대적으로 투자 기회가 많기 때문인 것으로 보인다. 또한, 기업이 산업 평균보다 높은 시장가 대 장부가 비율을 가지고 있는 기업은 유상증자를 많이 하는 것으로 나타났다. 시장가 대 장부가 비율을 보통 고평가의 측정치로 사용하기 때문에 이는 기업의 주가가 고평가되면 주식 발행을 많이 하는 것으로 풀이된다. 부채비율이 높은 기업들이 유상증자를 많이 하는 것으로 나타나서 기업이 부채비율이 높아서 부채를 추가로 일으키기 어려운 경우에는 주식 발행량을 늘리는 것으로 보인다.

**Table 2. Correlation Table**

	netequity[t]	Age [t-1]	Mb adj [t-1]	Oi [t-1]	Size [t-1]	Caa [t-1]	Lev [t-1]
<i>netequity[t]</i>	1.000						
<i>Age [t-1]</i>	-0.060 ( <.0001)	1.000					
<i>Mb adj [t-1]</i>	0.171 ( <.0001)	-0.159 ( <.0001)	1.000				
<i>Oi [t-1]</i>	0.041 ( <.0001)	0.061 ( <.0001)	-0.178 ( <.0001)	1.000			
<i>Size [t-1]</i>	-0.068 ( <.0001)	0.392 ( <.0001)	-0.122 ( <.0001)	0.156 ( <.0001)	1.000		
<i>Caa [t-1]</i>	0.049 ( <.0001)	0.075 ( <.0001)	0.129 ( <.0001)	-0.083 ( <.0001)	0.159 ( <.0001)	1.000	
<i>Lev [t-1]</i>	0.060 ( <.0001)	-0.254 ( <.0001)	0.114 ( <.0001)	-0.055 ( <.0001)	-0.307 ( <.0001)	-0.048 ( <.0001)	1.000

Leary & Roberts (2005)와 Strebulaev (2007)에서 보고하는 바와 같이 기업의 부채비율을 조정하려면 높은 비용이 들어간다. 따라서 기업들은 부채비율을 자주 조정하지 않는다. 이는 유상증자도 마찬가지이다. 본 연구에서는 고평가된 기업의 경우에는 유상증자에 따른 비용보다 충분히 큰 이익이 발생할 때만 고평가를 이용하여 주식을 발행할 것으로 가설을 세우고 검증했다. 아래의 <Fig 1>은 *산업 평균 대비 시장가대 장부가 비율*(이하, *조정 시장가 대 장부가 비율*)을 연도별로 20개의 그룹으로 나누어서 그룹 0은 가장 *조정 시장가 대 장부가 비율*이 낮은 기업이고 그룹 19는 *조정 시장가대 장부가 비율*이 높은 기업이다. Fig 2는 각각의 그룹에서 장부가 자본의 변화량이 전체 자산 대비 20% 이상 증가하여 유의미한 양의 유상증자를 한 기업의 비율을 보여주고 있다. 각 그룹별 유상 증자 비율은 가장 *조정 시장가 대 장부가 비율*이 낮은 그룹에서부터 14번째 그룹까지는 크게 변동이 없다. 하지만, 15번째 그룹부터 가장 *조정 시장가 대 장부가 비율*이 높은 마지막 그룹까지는 대규모 유상증자를 한 기업의 비율이 급격히 높아지는 것을 볼 수 있다. 이는 Leary & Roberts (2005)와 Strebulaev (2007)와 일맥상통하는 결과로 볼 수 있다. 기업들은 약한 정도의 고평가는 직접적으로 활용하지 않는 반면에 발행 비용을 넘어서는 정도로 고평가되는 경우에는 적극적으로 이를 유상증자 의사결정에 반영하는 것으로 나타난다. 가장 *조정 시장가 대 장부가 비율*이 높은 그룹의 경우에는 약 23%에 달하는 기업들이 전체 자산 대비 20%이상 대규모 유상증자를 하는 것으로 나타났다.

**Fig 1.** The Ratio of SEO Firms and The Industry Adjusted Market To Book Ratio

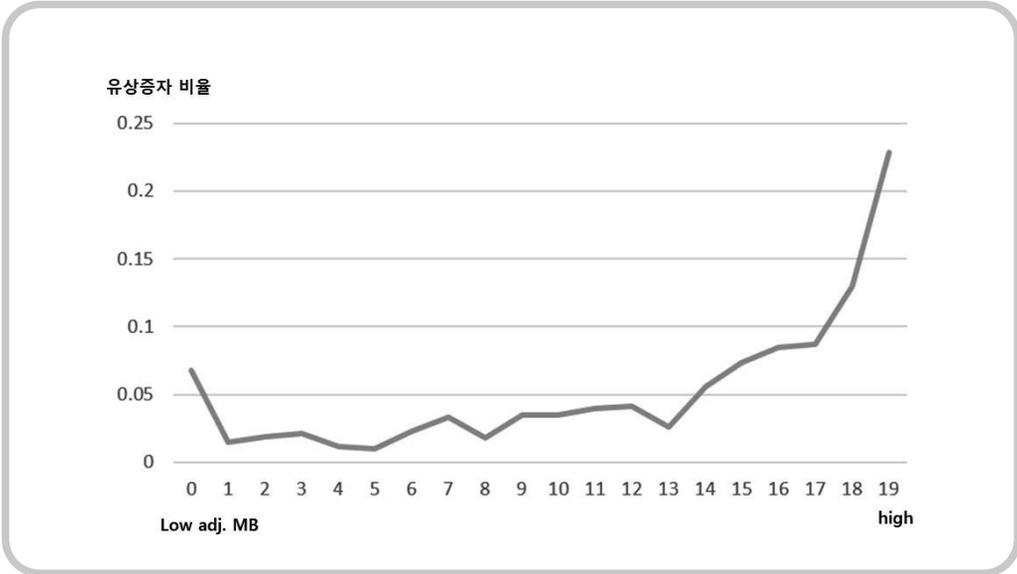
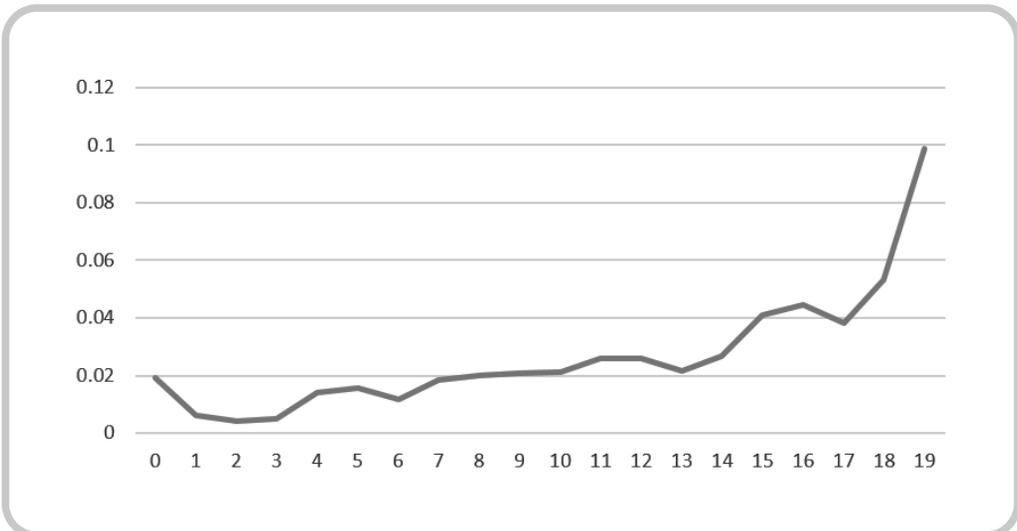


Fig 2는 산업 평균 대비 시장가 대 장부가 비율을 20개의 그룹으로 나누어 각각의 그룹에서 평균 유상증자 규모를 도출한다. 산업 평균 대비 시장가 대 장부가 비율이 가장 낮은 그룹에서부터 13번째 그룹까지는 전체 자산대비 평균적으로 2%보다 낮은 유상증자 규모를 보이지만, 그 이후로 가파른 상승세를 보이며 가장 조정 시장가 대 장부가 비율이 높은 그룹은 전체 자산대비 평균 10% 정도의 높은 유상증자 규모를 나타내고 있다. 이는 높은 유상증자 비용에 따라서 충분히 주가가 고평가된 경우에 경영자가 이를 유상증자 의사결정에 반영한다는 기존 문헌과 일치하는 결과이다.

**Fig. 2.** The Issuing Amount of SEO and The Industry Adjusted Market to Book Ratio



## 2. 유상증자 규모에 대한 분석

검증 방법은 교차 검증 (Cross Validation)과 무작위 추출(Random sampling)을 사용하였다. 교차 검증은 전체 자료를 10개의 그룹으로 나누어 하나의 그룹을 테스트 그룹으로 하고 나머지를 학습 그룹으로 하는 것을 각각의 그룹 10개에 대해서 반복하는 방법이다. 무작위 추출은 전체 자료 중에서 66%를 뽑아서 학습하고 나머지에 대해서 테스트하는 것을 10번 반복하여 평균을 산출한 결과이다.

모형의 예측력을 측정하기 위해서는 RMSE, MAE와 R-square를 사용했다. RMSE는 평균 제곱근 편차라고도 불리며 실제값과 예측치의 차이를 제공한 것의 평균에 제곱근을 취한 형태이다. MAE는 평균 절대값 오차로 실제값과 예측치의 차이의 절대값의 평균이다. 표3에서 나타나는 바와 같이 RMSE와 MAE는 선형 모형이 가장 커서 세 모형 중에서 오차가 가장 큰 것으로 나타났으며, 그 다음이 랜덤 포레스트 방법이고, 가장 오차가 적은 모형이 그래디언트 부스팅 방법으로 나타났다.

R-square 역시 비슷한 결과를 나타내는데, 선형 모형이 가장 R-square가 작았으며, 랜덤 포레스트, 그래디언트 부스팅의 순서로 값이 커진다. 교차 검증의 경우에 선형 모형의 R-square는 4%에 불과하지만, 랜덤 포레스트 모형은 7.4%, 그래디언트 부스팅 모형은 9.6%로 커지는 것을 볼 수 있다. 따라서 그래디언트 부스팅 모형의 R-square는 선형 모형에 비해서 5.6%p가 높다. 무작위 추출 검증 방식에서도 이와 크게 다르지 않은 패턴을 보여주고 있다. 이와 같은 결과는 유상증자의 규모를 예측할 때 유상증자 결정 요인과 유상증자 규모의 비선형 관계를 고려한다면 더 높은 설명력을 달성할 수 있다는 것을 의미한다.

**Table 3.** RMSE, MAE, and R-square

### Panel A. Cross Validation

Model	RMSE	MAE	R-square
Linear model	0.131	0.076	0.040
Random Forest	0.128	0.075	0.074
Gradient Boosting	0.127	0.073	0.096

### Panel B. Random Sampling

Model	RMSE	MAE	R-square
Linear model	0.130	0.076	0.040
Random Forest	0.128	0.074	0.072
Gradient Boosting	0.127	0.073	0.092

## 3. 유상증자 지시변수에 대한 분석

표4의 결과에서는 유상증자 지시변수를 사용하여 분석했다. Hovakimian (2006)에서와 일치하는 방식으로 전체 자산대비 자기자본의 변화량이 5% 이상의 기업을 유상증자한 것으로 보고 유상증자 지시변수를 1로 정의한다. 그리고 나머지 기업들에게 0을 부여하는 지시변수이다. 주요 설명 변수는 앞에서와 마찬가지로 *산업 평균 대비 시장가 대 장부가 비율, 영업이익률, 기업 연령, 회사 크기, 부채비율* 등이 있다.

각각의 모형의 예측력을 측정하기 위해서는 정확도(CA), 정밀도(precision), 재현율, F1을 이용했다. 정확도는 전체 분석 대상에서 유상증자를 한 기업(True: 변수값이 1로 정의)과 유상증자를 하지 않은 기업(False: 변수값이 0으로 정의)을 정확히 맞춘 비율이다. 정밀도(precision)은 모델이 유상증자를 한 기업로 분류한 것 중에서 실제로 유상증자를 한 비율이며, 재현율(Recall)은 실제 유상증자를 한 기업

중에서 모델이 유상증자할 것이라고 예측한 기업의 비율이다. F1은 정밀도와 재현율의 조화 평균이다. 따라서 정확도, 정밀도, 재현율, F1이 높을수록 더 설명력이 높은 실증 모형으로 볼 수 있다.

아래의 표4에서는 로지스틱 모델에 비해서 랜덤 포레스트 모델은 정확도가 약 4.2%p가 높으며, 그래디언트 부스팅 모형도 4.4%p가 높게 나타난다. 또한, 정밀도, 재현율, F1도 로지스틱 모형에 비해서 랜덤포레스트와 그래디언트 방법이 더 높은 것을 볼 수 있다. 따라서 로지스틱 모형에 비해서 랜덤 포레스트와 그래디언트 부스팅과 같은 머신 러닝 방법론을 도입한다면 더 높은 정확도를 얻을 수 있다.

**Table 4.** CA, F1, Precision & Recall

**Panel A.** Cross Validation

Model	CA	F1	Precision	Recall
Logistic	0.652	0.597	0.637	0.652
Random Forest	0.694	0.679	0.684	0.694
Gradient Boosting	0.696	0.679	0.688	0.696

**Panel B.** Random Sampling

Model	CA	F1	Precision	Recall
Logistic	0.655	0.600	0.644	0.655
Random Forest	0.695	0.679	0.685	0.695
Gradient Boosting	0.696	0.679	0.687	0.696

## IV. 추가 분석

III장의 결과가 코스피와 코스닥에 상장된 기업들에게 각각 다르게 적용되는지 알아보기 위하여 코스피 기업과 코스닥 기업들을 분리해서 분석했다. 분석 결과, 코스피와 코스닥에서 모두 비선형 관계를 고려한 랜덤포레스트와 그래디언트 부스팅 트리 방법론이 선형 방법론과 비교해서 설명력이 좋은 것으로 나타났다.

### 1. 코스피 기업과 코스닥 기업

다음의 표5와 표6은 코스피 기업들만을 따로 분리하여 분석한 결과이다. 표5에서는 유상증자 규모에 대하여 분석했다. RMSE와 MAE 모두 선형 모형보다는 랜덤 포레스트와 그래디언트 부스팅 모형이 더 낮게 나타나는 것을 볼 수 있다. 이는 교차 검증과 무작위 추출 방식 모두에서 일관되게 나타난다. R-sqaure는 선형 모형과 비교하여 랜덤 포레스트와 그래디언트 모형이 모두 높게 나타나는 것을 볼 수 있다. 이는 표3과 비슷한 결과이다. 이와 같은 결과는 분석 대상을 무작위 추출하여 검증한 panel B에서도 비슷하게 나타난다.

**Table 5. RMSE, MAE, and R-square**

**Panel A. Cross Validation**

Model	RMSE	MAE	R-square
Linear model	0.107	0.060	0.039
Random Forest	0.105	0.059	0.066
Gradient Boosting	0.104	0.058	0.085

**Panel B. Random Sampling**

Model	RMSE	MAE	R-square
Linear model	0.106	0.059	0.033
Random Forest	0.104	0.058	0.068
Gradient Boosting	0.103	0.057	0.082

아래의 표6에서는 유상증자 지시변수를 이용하여 전체 자산대비 자본의 변화량이 5%가 넘는 경우에 1을 가지고 그렇지 않은 경우에는 0을 가지는 유상증자 지시변수에 대해서 분석한 결과이다. 표4와 마찬가지로 코스피 기업들의 경우에도 선형 모형인 로지스틱 모형에 비해서 랜덤포레스트와 그래디언트 부스팅 트리 모형이 CA, F1, precision, recall가 모두 높은 것으로 나타났다. 이와 같은 결과는 교차 검증과 무작위 추출에서 모두 같다.

**Table 6. CA, F1, Precision & Recall**

**Panel A. Cross Validation**

Model	CA	F1	Precision	Recall
Logistic	0.691	0.618	0.661	0.691
Random Forest	0.730	0.703	0.716	0.730
Gradient Boosting	0.729	0.705	0.713	0.729

**Panel B. Random Sampling**

Model	CA	F1	Precision	Recall
Logistic	0.690	0.617	0.660	0.690
Random Forest	0.725	0.698	0.708	0.725
Gradient Boosting	0.725	0.701	0.708	0.725

표7과 표8은 코스닥 기업들만을 대상으로 따로 분류하여 분석한 결과이다. 표7은 유상증자 규모에 대해서 분석했으며, 표8은 유상증자 지시변수를 사용하여 결과를 도출했다. 코스닥 기업에 대해서 분석했을 때에도 코스피 기업들과 마찬가지로 선형 모형보다 비선형 관계를 고려하는 랜덤 포레스트와 그래디언트 모형이 설명력이 높게 나타났다. 이러한 관계는 교차 검증과 무작위 추출 검증에 대해서도 마찬가지로 나타났다. 표7은 코스닥 기업을 대상으로 유상증자 규모에 대해서 분석하였다. 표7은 대체로 표5의 코스피 기업을 대상으로 분석한 결과와 비슷하게 선형 모형보다는 비선형 모형인 랜덤 포레스트와 그래디언트 부스팅 트리 모형의 설명력이 높게 나타났다. 표8에서는 유상증자 지시변수를 이용한 결과이다. 표8에서도

위의 표와 비슷하게 나타나므로 본 연구에서 제시하는 결과가 강건함을 알 수 있다.

**Table 7. RMSE, MAE, and R-square**

**Panel A. Cross Validation**

Model	RMSE	MAE	R-square
Linear model	0.146	0.089	0.046
Random Forest	0.143	0.086	0.082
Gradient Boosting	0.142	0.085	0.099

**Panel B. Random Sampling**

Model	RMSE	MAE	R-square
Linear model	0.146	0.089	0.048
Random Forest	0.143	0.086	0.084
Gradient Boosting	0.142	0.085	0.097

**Table 8. CA, F1, Precision & Recall**

**Panel A. Cross Validation**

Model	CA	F1	Precision	Recall
Logistic	0.621	0.580	0.614	0.621
Random Forest	0.675	0.663	0.670	0.675
Gradient Boosting	0.669	0.659	0.664	0.669

**Panel B. Random Sampling**

Model	CA	F1	Precision	Recall
Logistic	0.623	0.583	0.616	0.623
Random Forest	0.670	0.658	0.665	0.670
Gradient Boosting	0.668	0.658	0.662	0.668

## 2. 표본 외 실험(out-of-sample test)

교차 검증과 무작위 추출 검증에 추가하여 표본외 테스트를 진행하였다. 표본외 위하여 전체 분석 기간을 반으로 나누었다. 2003년에서 2011년까지는 학습 기간이고, 2012년에서 2019년까지는 테스트 기간이다. 아래의 표9와 표10에서는 학습 기간의 자료를 이용하여 추정하고 추정된 모형을 테스트 기간에 적용하여 모형의 설명력을 도출하였다.

아래의 표9에서는 *유상증자 규모*에 대해서 분석했으며 표10에서는 *유상증자 지시변수*에 대해서 분석했다. 그 결과, 앞서 제시한 표와 마찬가지로 선형 모형보다는 비선형 관계를 고려하는 랜덤 포레스트와 그래디언트 부스팅 트리 모형이 설명력이 높은 것으로 나타났다.

**Table 9.** RMSE, MAE, and R-square

Model	RMSE	MAE	R-square
Linear model	0.129	0.075	-0.004
Random Forest	0.127	0.073	0.030
Gradient Boosting	0.125	0.072	0.068

**Table 10.** CA, F1, Precision & Recall

Model	CA	F1	Precision	Recall
Logistic	0.680	0.673	0.669	0.680
Random Forest	0.702	0.704	0.706	0.702
Gradient Boosting	0.695	0.698	0.703	0.695

### 3. 변수의 중요도

유상증자에 대해서 어떤 변수들이 설명력이 높은지 알아보기 위해서 중요도 분석을 했다. 종속 변수는 유상증자 규모와 유상증자 지시변수이며, 주요 설명 변수는 영업 이익, 산업 평균 대비 장부 대 시장가 비율, 부채비율, 회사 규모 등이 있다.

유상증자의 규모를 설명하기 위하여 랜덤 포레스트와 그래디언트 부스팅 방법론을 사용했다. 아래의 결과에서는 각각의 설명 변수를 제외하고 분석했을 때 R-square가 얼마나 감소하는지 보여주고 있다. 따라서 각각의 설명 변수가 얼마나 유상증자 규모를 설명하는 데 중요한지 알 수 있다. 랜덤 포레스트와 그래디언트 부스팅 모두 영업 이익이 가장 중요도가 높은 것으로 나타났으며, 그 다음이 산업 평균 대비 장부대 시장가 비율이다. 앞서 서술한 바와 같이 산업 평균 대비 장부대 시장가 비율은 마켓타이밍과 연관이 있다. 그러므로 마켓 타이밍이 유상증자의 규모를 설명하는데 상당히 중요한 것을 알 수 있다.

**Fig. 3.** Analysis of the Importance of Variables on the Issuing Amount of SEO

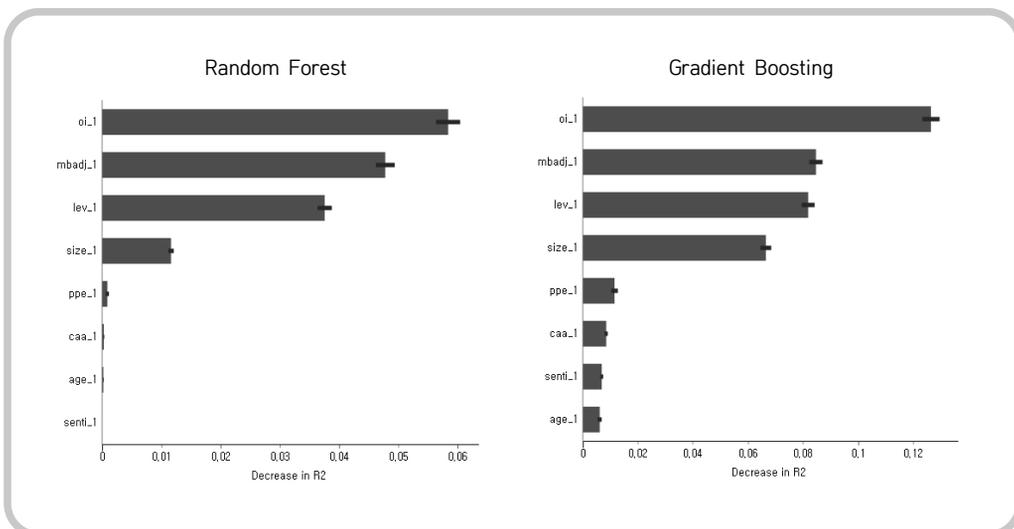
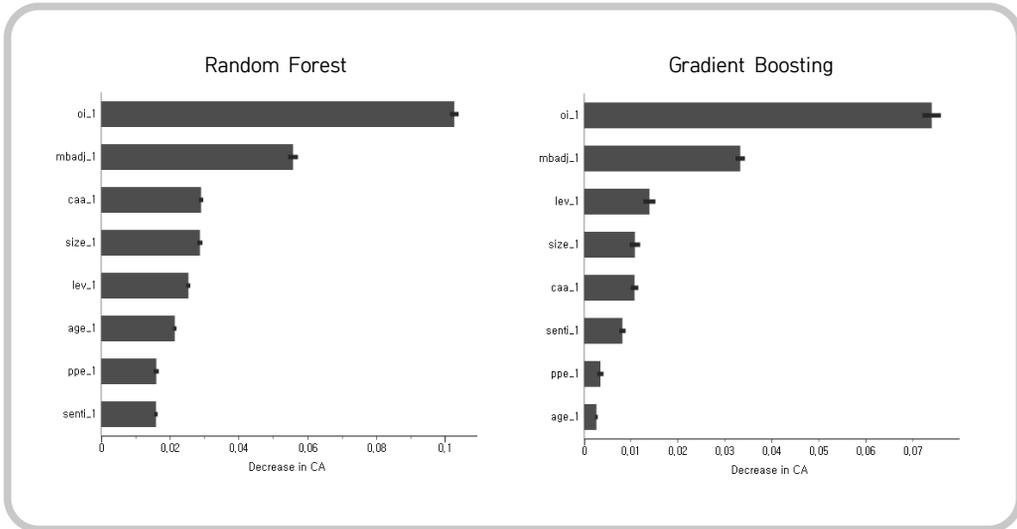


Fig 4.는 유상증자 지시변수에 대해서 분석한 결과이다. 정확도(CA)에 대해서 분석했는데, 각각의 설명 변수를 제외하면 정확도가 얼마나 감소하는지 보여준다. 결과는 Fig 3.와 비슷하게 나타나서 영업 이익이 가장 중요한 것으로 나타났고, 그 다음이 산업 평균 대비 장부 대 시장가 비율인 것으로 나타난다. 따라서 마켓타이밍이 중요한 것임을 다시 한 번 확인할 수 있다.

Fig. 4. Analysis of the Importance of Variables on an SEO Indicator



## V. 결론

본 연구에서는 유상증자를 예측하기 위하여 기존의 선형 모형과 함께 비선형 관계를 고려할 수 있는 랜덤 포레스트와 그래디언트 부스팅 트리 모형을 이용하여 분석한다. 교차 검증과 무작위 추출 방법을 이용하여 검증하며, 추가 분석을 위하여 전체 분석 대상을 코스피와 코스닥 상장 기업들을 각각 분석했다. 또한 표본의 검증 결과와 변수의 중요도 분석도 실시하였다. 그 결과, 선형 모형보다는 랜덤 포레스트와 그래디언트 부스팅 트리 모형이 설명력이 높은 것으로 나타났으며, 이와 같은 결과는 코스피와 코스닥 기업들에 대해서 각각 분석했을 때에도 모두 같게 나타났다.

정보 비대칭이 없는 이상적인 경제 상황에서 기본적으로 주식 발행의 이유는 좋은 투자안에 투자하기 위한 자본 조달이다. 따라서 이상적인 상황에서는 유상증자가 미래 주가 수익률에 부정적인 영향을 미치지 않는다. 하지만, 실제로 기업 내부와 외부 사이에는 정보 비대칭이 존재하기 때문에 유상증자를 하면, 기업에서 단기적인 고평가를 이용하는 것인지에 대한 의혹이 생길 수 있다. 본 연구에서는 유상증자를 할 것으로 예측되는 기업들을 분류하여 의미있는 정보를 전달하고 정보의 비대칭을 해소하는데 도움이 되고자 한다.

투자자들은 유상증자의 가능성이 높을 것으로 예측되는 기업들에 투자할 때는 더욱 조심할 수 있고, 금융 당국은 이러한 기업들을 면밀히 관찰할 수 있다. 결과적으로 본 연구는 여러 이해 관계자들이 참고할 수 있는 정보를 제공하여 긍정적인 목적으로 주식을 발행하는 기업들이 피해를 보지 않도록 정보비대칭을 줄이는데 기여한다.

## References

- 김수인·변진호 (2016), “유상증자와 대주주 거래 연구,” *재무연구*, 29(1), 1-36.
- 임병권·윤평식 (2017), “유상증자 공시 전 정보거래에 관한 연구,” *한국증권학회지*, 46(1), 133-157.
- 신현한·윤평식 (2023), “유상증자 전 내부자거래 부재가 시장에 정보를 제공하는가?,” *한국증권학회지*, 52(6), 947-978.
- 박형진·홍종운 (2019), “유상증자 방식과 시장상황에 따른 유상증자후 장기성과 분석,” *대한경영학회지*, 32(4), 703-719.
- Baker, M. and J. Wurgler (2002), “Market timing and capital structure”, *The Journal of Finance*, 57(1), 1-32.
- DeAngelo, H., L. DeAngelo and R. M. Stulz (2010), “Seasoned equity offerings, market timing, and the corporate lifecycle”, *Journal of Financial Economics*, 95(3), 275-295.
- Hastie, T., R. Tibshirani and J. Friedman (2009), “The elements of statistical learning: data mining, inference, and prediction”, Springer Science & Business Media.
- Gu, S., B. Kelly and D. Xiu (2020), “Empirical asset pricing via machine learning”, *The Review of Financial Studies*, 33(5), 2223-2273.
- Leary, M. T. and M. R. Roberts (2005), “Do firms rebalance their capital structures?”, *The Journal of Finance*, 60(6), 2575-2619.
- Lee, I., S. Lochhead, J. Ritter and Q. Zhao (1996), “The costs of raising capital”, *Journal of Financial Research*, 19(1), 59-74.
- Masulis, R. W. and A. N. Korwar (1986), “Seasoned equity offerings: An empirical investigation”, *Journal of Financial Economics*, 15(1-2), 91-118.
- Spiess, D. K. and J. Affleck-Graves (1995), “Underperformance in long-run stock returns following seasoned equity offerings”, *Journal of Financial Economics*, 38(3), 243-267.
- Strebulaev, I. A. (2007), “Do tests of capital structure theory mean what they say?”, *The Journal of Finance*, 62(4), 1747-1787.