

연구보고서 25-04

RESEARCH
REPORT

머신러닝을 활용한 재무제표 정보의 유용성 평가: 부실 징후의 조기 탐색을 중심으로

노성호 | 이상호

머신러닝을 활용한 재무제표 정보의 유용성 평가: 부실 징후의 조기 탐색을 중심으로

2025. 2

연구위원 노성호

연구위원 이상호

* 본고의 견해와 주장은 필자 개인의 것이며, 자본시장연구원의 공식적인 견해가 아님을 밝힙니다.
금융산업실 연구위원 노성호(nohs@kcmi.re.kr), 자본시장실 연구위원 이상호(sangholy@kcmi.re.kr)

Contents

I. 서론	3
1. 연구배경 및 목적	3
2. 주요 연구결과	9

II. 선행연구	15
1. 재무제표 주석 정보의 유용성에 관한 선행연구	15
2. 부실징후 예측에 관한 선행연구	17

III. 연구방법론	23
1. 멀티모달 신경망 모형	23
2. 모형의 학습 과정	25

IV. 연구 결과	37
1. 재무제표 및 주석 정보의 학습 효율성 비교	37
2. 부실 평가 기업의 특성 분석	46
3. 공시 유용성 기여 요인 평가	49

V. 결론 및 시사점	55
--------------------	-----------

Executive Summary

본 연구보고서는 국내 상장기업의 부실 위험을 조기에 탐지하는 것을 목표로 한다. 구체적으로, 상장폐지에 앞서 발생하는 거래정지 시점을 기준으로 부실 징후를 예측하여 실효성 있는 조기 경보 모형을 설계하고자 한다. 또한, 한국채택국제회계기준(K-IFRS) 도입 이후 급증한 주식 정보의 양과 중요성을 반영하여, 재무제표 본문의 정량적 정보뿐만 아니라 주식에 포함된 비정형 정보까지 활용한 예측을 시도한다. 이를 위해 멀티모달 신경망(multimodal neural network) 기반의 머신러닝(machine learning) 방법론을 적용하였다.

최근 경제 불확실성의 확대와 한계기업 문제의 만성화로 본 연구에서 고안하는 부실 예측 모형은 투자자와 금융 당국에 유의미한 정보를 제공할 것으로 기대된다. 특히, 기업의 부실 사건에 대한 주식 정보의 예측력을 평가하는 점은 악재(bad news)를 더욱 적시에 인식하라는 회계의 보수주의 원칙이 주식 정보에도 충실히 반영되고 있는지를 검토하는 것으로, 학술적으로도 중요한 공헌이 예상된다. 아울러, 주식 정보의 작성 실태에 대한 광범위한 실증결과를 바탕으로, 주식 공시의 개선 방향에 대해 정책적 제언을 제시한다.

주요 분석을 위해 2005년부터 2019년까지 유가증권 및 코스닥시장에 상장된 12월 말 결산 비금융업종 16,815 기업-연도를 대상으로 표본을 구성하였다. 기업의 한계 요인과 외부 원천별 정보를 포함한 43개의 정량적 지표와 37개의 주요 계정과목, 그리고 재무제표 주식에 포함된 모든 비정형 정보를 학습에 활용하였으며, 이를 바탕으로 모형의 예측력을 검증하기 위한 평가를 수행하였다. 특히, 한국 시장의 특성상 기업 부실로 인한 상장폐지 시 장기간의 거래정지가 수반되는 점을 고려하여, 거래정지 이전 시점에서 가용한 정보만을 모형에 반영하였다.

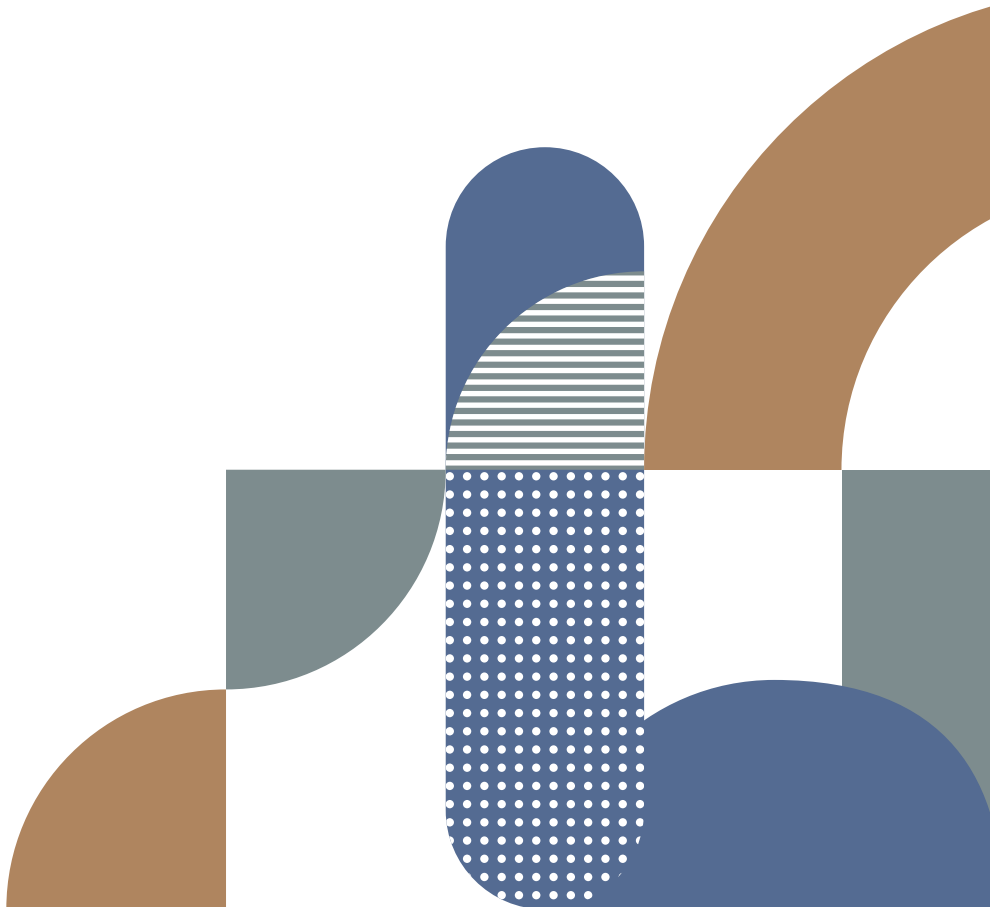
그 결과, 상장폐지 위험을 거래정지 이전에 높은 정확도로 예측할 수 있음을 확인하였다. 이는 투자자들이 실제 의사결정을 내릴 수 있는 시점에 부실을 예측한 결과라는 점에

서 그 의미가 깊다. 다만, 우수한 예측력은 상대적으로 부실 가능성이 낮은 기업을 정확하게 식별한 결과에서 기인하므로, 모형의 개선 필요성도 함께 제기된다.

또한, 부실화 가능성이 높은 기업을 예측하는 데 주식 정보의 기여도가 제한적이었는데, 이는 주식 작성 시 회계의 보수주의 원칙이 충분히 고려되고 있는지를 면밀히 검토할 필요가 있음을 시사한다. 나아가, 부실 징후 예측에서 주식 정보의 유용성을 높이기 위해서는 재무제표 본문과 주식 정보 간의 연계성을 강화하고, 주식 정보의 양과 이를 전달하는 양태 측면에서도 개선이 필요한 것으로 나타났다. 주식 정보는 핵심 내용을 강조해야 하며, 기간 간 및 유사 기업 간의 비교가능성을 높이는 방향으로 작성되어야 한다. 향후 재무제표 주식 정보의 품질을 개선하고, 이를 바탕으로 예측의 정확도와 유용성을 개선하는 것은 중요한 후속 과제로 남겨둔다.

I. 서론

1. 연구배경 및 목적
2. 주요 연구결과



I. 서론

1. 연구배경 및 목적

코로나바이러스감염증-19 이후 자금조달 비용 증가, 경기 둔화 우려 등의 영향이 본격화되면서 연체 발생 기업이 증가하고 있다.⁰¹ 최근 급격한 기준 금리 인상을 유발한 물가 상승 압력은 다소 완화되었으나, 취약·한계기업의 저하된 수익성이 구조적으로 개선되지 않고 있어 신용위험의 확대 가능성에 예의주시해야 한다는 지적이 있다(IMF, 2023). 금융 안정을 책임지는 감독 당국으로서 기업 부실의 징후를 조기에 탐지하고 선제적으로 관리하는 것이 중요한 과제일 것이다. 또한, 우리나라 자본시장은 상장폐지 사유가 발생하면 장기간 거래가 정지되는 경향이 있어(이상호, 2022), 투자자들 역시 기업의 부실 징후를 사전에 예측할 유인이 강하다.

이에 본 연구보고서에서는 우리나라 상장기업을 대상으로 부실 징후를 조기에 탐지할 가능성을 분석하고자 한다. 기업 부실은 기초체력(fundamental)의 저하로 인해 발생하는 상장폐지 사건으로 정의하며(Campbell et al., 2008; 이인로·김동철, 2015), 이를 예측하기 위해 멀티모달 신경망(multimodal neural network) 기반 머신러닝(machine learning) 기법을 활용한다. 이는 재무제표 본문에 담긴 정량적 정보뿐만 아니라, 주식에 기재된 방대한 비정형 정보까지 통합하여 예측 모형을 구축하려는 시도로, 기존 연구와 차별화된다. 이러한 접근을 통해 재무제표의 정량적·정성적 정보가 기업의 부정적 경제 상황을 예측하는 데 유용한 질적 특성을 내포하고 있는지를 검증한다.

재무제표는 크게 본문과 주석으로 구성된다. 재무제표 본문은 재무상태표, 포괄손익 계산서, 현금흐름표, 자본변동표를 포함하며, 기업의 재무 현황과 경영 성과 등을 정량적 수치로 요약한다. 주석은 본문의 수치 정보를 보완하며, 기업의 경제적 실질을 보다 심층적으로 이해할 수 있는 다양한 정성적 정보를 제공한다.⁰² 그러나 기존 연구는 주로 재무제표 본문에 포함된 정량적 정보만을 예측 변수로 활용해 왔다(예: Dechow et al.(2011), Bao et al.(2020), 이인로·김동철(2015), 오세경 외(2017), 나현중·정태진(2022) 등).⁰³

01 금융감독원(2023. 12. 18)

02 경제적 실질에 대한 보충적 설명을 제공하는 주석 정보의 예로는 계류중인 소송 사건, 담보제공 자산, 부채 조달 원천의 질적 특성, 회계정책 변경 등 매우 다양하며, 구체적 항목은 <표 I-1>을 참조한다.

03 오세경 외(2017)는 기업의 부도 위험 예측에 텍스트 정보를 통합 활용하였으나, 이는 뉴스 키워드 정보를 반영한

이는 방대한 비정형 정보를 다루기 어려운 기존 연구모형의 한계와 무관하지 않을 것이나, 엄밀하게는 주석을 포함한 전체 재무제표 정보가 기업의 부실 징후를 예측하는 데 얼마나 유용한지에 대한 실증적 검토는 부족한 상황이라 평가할 수 있다.

원칙에 따라 작성된 재무제표 정보는 본질적으로 기업의 부실 징후를 예측하는 데 유용하여야 한다. 회계의 신중성(prudence) 개념은 기업에 불리한 경제적 상황이 확정되지 않았더라도(예: 자산손상, 우발채무, 소송, 담보제공 등), 이를 왜곡 없이 충실히 표현하도록 상당한 주의를 요구하기 때문이다. 보수주의(conservatism) 원칙 역시 불확실한 상황에서 손실을 이익보다 더 신속히 인식하도록 하여, 불리한 정보를 은폐하려는 경영진의 기회주의적 성향을 억제하고자 한다(Hutton et al., 2009; Kim & Zhang, 2016). 이러한 개념적 기반을 토대로 한국채택국제회계기준(이하 K-IFRS)은 주석에 기재해야 할 정보 항목을 구체적으로 규정하고 있다(<표 I -1> 참조). 따라서 주석 정보는 기업의 경제적 실질을 이해하고 부실 징후를 파악하는 데 중요한 속성을 가질 것으로 예상할 수 있다.

구체적으로, <표 I -1>의 (3) 영업부문에 관한 주석 정보는 내부 경영진에 보고되는 관리회계 수준에서 제공되어야 한다(K-IFRS 제1108호 문단 8). 기업 내부자의 시각에서 파악된 부문별 성과 정보가 주석에 공시되어야 하므로, 영업적으로 중요한 부문의 부실 징후가 은폐될 가능성은 줄어든다. 또한, (32) 우발채무나 (36) 자산손상에 관한 주석 정보는 미래의 불확실성에 관한 다양한 비계량적 정보와 함께 자원의 유출 가능성, 회수가능액 등의 추정치를 포함하므로(K-IFRS 제1036호 문단 130, 제1037호 문단 86), 정보이용자는 이를 통해 부정적인 사건의 예상 규모와 실현 가능성을 보다 정확히 평가할 수 있을 것이다.

결과로 재무제표의 주석 정보는 고려되지 않았다.

<표 I-1> K-IFRS 도입 후 최초로 공표된 주식 항목

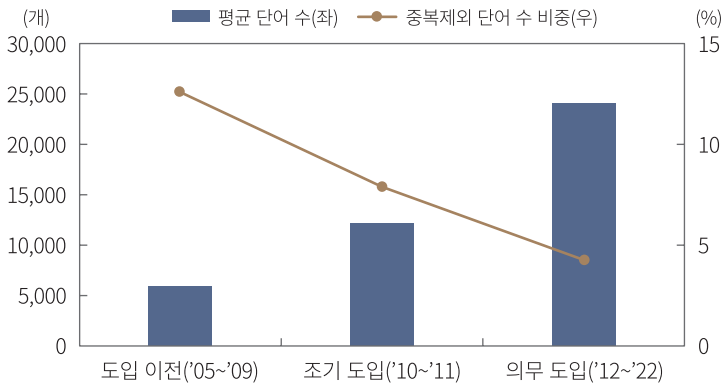
번호	내용	번호	내용
1	일반적 사항	25	종업원급여
2	중요한 회계처리 방침	26	주식기준보상약정
3	영업부문	27	기타충당부채
4	중단영업	28	차입금
5	사업결합	29	신주인수권부사채
6	초인플레이션 경제에서의 재무보고	30	전환사채
7	보고기간후 사건	31	상환전환우선주
8	중간재무보고	32	약정사항, 담보제공, 지급보증내역 및 우발채무
9	특수관계자	33	자산, 부채, 자본의 하위분류
10	IFRS 최초 적용	34	진행중인 계약
11	유형자산	35	수익과 비용의 분석
12	탐사평가자산	36	손상차손과 손상차손환입
13	무형자산	37	금융상품
14	생물자산과 농림어업활동을 위한 정부보조금	38	금융부채와 자본 사이의 재분류를 가져오는 상환금지대상
15	투자부동산	39	리스
16	조인트벤처 투자지분	40	리스의 법적 형식을 포함한 거래
17	관계기업에 대한 투자	41	민간투자사업
18	사용이 제한된 금융상품	42	정부보조금
19	재고자산	43	법인세
20	자본금, 적립금, 기타지분	44	기금 지분의 성격과 사용에 대한 제약사항
21	이익잉여금처분계산서	45	현금흐름표
22	자기주식	46	보험계약
23	배당금	47	연결재무제표와 별도재무제표
24	주당이익		

자료: 금융감독원(2010. 12. 21)

그러나 상술한 개념적 당위성에도 불구하고, 주석을 포함한 총체적인 재무제표 정보가 기업 부실 징후에 대한 예측력을 반드시 제고할지는 확실치 않다. 일부 선행연구에

따르면, 과도한 주식 공시는 오히려 정보이용자의 이해를 저해할 수 있으며(Bloomfield, 2012), 주식은 기업 고유의 정보를 중심으로 간결하고 명확하게 작성되어야 한다는 지적이 있다(이재경 · 한봉희, 2019). 실제로, 우리나라는 K-IFRS 도입 이후 주식 정보량이 급격히 증가하였으나(<그림 I-1> 참조)⁰⁴, 공시된 주식 정보의 복잡성, 중복성도 함께 증가한 것으로 나타났다(<그림 I-2> 참조). 전년도 주식 정보와 단어의 교집합 크기로 추정된 시계열적 유사도뿐만 아니라, 동종업계 내에서의 유사도 또한 증가한 것으로 확인되어(<그림 I-3> 참조), 전반적으로 중요한 정보가 비교할 수 있는 형태로 강조된 영향인지, 아니면, 틀에 박힌(boilerplate) 형식적 공시가 늘어난 영향인지 불분명하다.

<그림 I-1> K-IFRS 도입 전후 재무제표 공시 분량

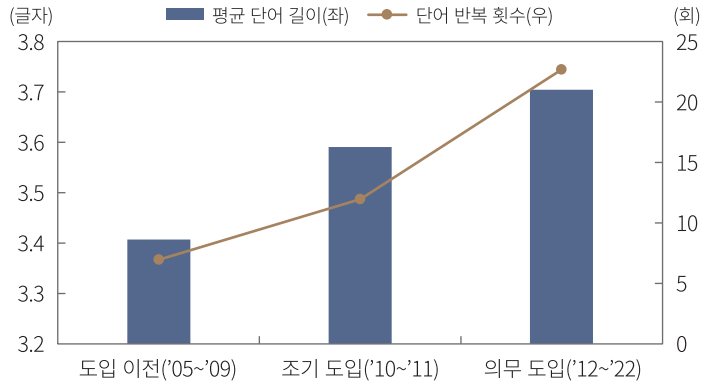


주 : 가용한 모든 사업보고서의 재무제표 본문 및 주식 공시에서 숫자와 특수문자를 제외하고 추출된 단어 수의 평균 및 이와 대비하여 중복되는 단어를 제외한 수치의 비중

자료: 금융감독원 DART를 바탕으로 저자 계산

04 K-IFRS는 2010 사업연도부터 조기 도입이 허용되었으며, 2012 사업연도에 전면 도입된 이후, 2014 사업연도부터 분·반기보고서로 확대 적용되었다.

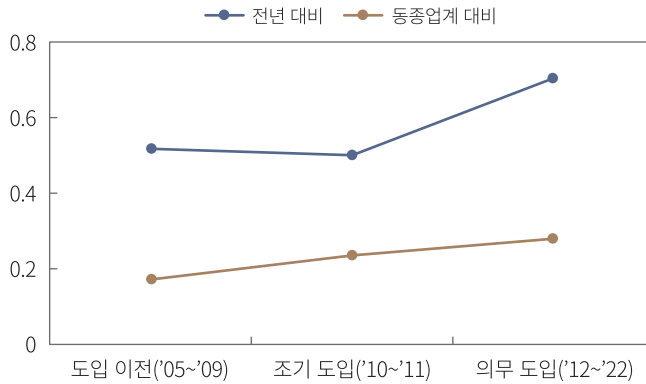
<그림 1-2> K-IFRS 도입 전후 재무제표 내 단어 길이 및 반복 횟수



주 : 재무제표의 본문 및 주석에서 사용된 단어의 평균 글자 수와 각 단어가 반복 사용되는 횟수의 평균

자료: 금융감독원 DART를 바탕으로 저자 계산

<그림 1-3> K-IFRS 도입 전후 재무제표 자카드 유사도



주 : 재무제표의 본문 및 주석에서 사용된 단어의 분포를 바탕으로 전년 및 동종업계(FnGuide 산업분류 기준) 대비 자카드 유사도의 평균

자료: 금융감독원 DART를 바탕으로 저자 계산

결과적으로, 늘어난 주식의 정보량이 투자자의 시각에서 얼마나 유용한지에 대해서는 다각도의 실증적 검증이 필요한 상황으로 평가된다(현지원 외, 2022). 특히, 모에린·서윤석(2019)의 연구에 따르면, 전년 대비 새로운 주식 정보가 많이 공시될수록 자기자본비용이 낮아지는데, 이는 주식에 부정적 정보가 투명하게 공개되지 않거나, 긍정적 정보 위주의 비대칭적인 공시가 이루어질 가능성을 시사한다. 경영진의 기회주의적인 재무보고

성향이 주식의 비정형 정보 측면에서도 효과적으로 억제될 수 있는지는 학술적으로 엄밀한 검증이 필요해 보이며, 이는 주식에 대한 재무제표의 감독체계 측면에서도 중요한 시사점을 제공할 것이다.

세부적인 분석을 위해 본 연구에서는 멀티모달 신경망 모형을 기반으로 실제 기업 부실 사건의 특성을 파악한다. 이는 재무제표 본문의 정량적 정보와 주식의 정량적·정성적 정보를 동시에 학습하는 머신러닝 기법이다. 해당 모형의 학습 및 예측 과정이 인간의 의사결정 방식과 완전히 동일하지는 않을 것이나, 멀티모달 신경망 모형은 다수의 정형화된 수치뿐만 아니라 다양한 비정형 정보를 종합적으로 학습하여 예측 결과를 도출한다는 점에서 현실 투자자의 의사결정 방식과 상당히 유사하다고 평가할 수 있다.

이와 같은 머신러닝 기법의 장점은 기존 선형회귀모형과 비교할 때 더욱 두드러진다. 머신러닝 모형은 기본적으로 이용할 수 있는 정보의 모든 비선형적 조합을 고려하므로, 복수의 계정에서 비정상적인 증감으로 나타나는 부실 징후를 포착하는 데 기존 선형회귀 분석 모형이 가지는 한계를 극복한다(Cao & You, 2024).⁰⁵ 또한, 인간은 처리해야 하는 정보량이 증가할수록 주의력(attention)의 한계로 가격 민감 정보에 효율적으로 반응하지 못하는 경향이 있는데(예: Hirshleifer & Teoh(2003), Chen et al.(2023) 등), 기계는 설비 규모에 제약을 두지 않는 한 이러한 한계로부터 자유롭다. 따라서 주식 정보가 방대하고 중복될수록 머신러닝 모형이 인간보다 더 효율적으로 유용한 정보를 습득할 가능성이 있다. 실제로 K-IFRS 도입 이후 급격히 증가한 주식 정보는 머신러닝 기법을 비롯한 기계의 도움 없이는 효율적인 분석이 어려운 양적 수준에 이르렀다.

이상의 논의를 바탕으로 본 보고서에서는 첫째, 머신러닝 기법을 활용하여 재무제표의 정량적 정보가 기업 부실 징후 예측에 얼마나 유용한지를 분석하고, 상장폐지와 같은 기업의 극단적 부실 사건을 거래정지 이전 시점에 조기에 탐지할 수 있는지를 평가한다. 둘째, 재무제표의 주식 정보를 본문 정보와 결합하였을 때, 부실 징후에 대한 예측력이 향상되는지를 확인한다. 이를 통해 감독 당국에는 금융 안정을 위한 조기 경보 모형을, 투자자에게는 퇴출 이전 장기간 거래정지를 동반하는 시장 제도적 특성에 대응할 수 있는

05 예를 들어, 선형회귀모형에서 매출채권회전율, 재고자산회전율 및 매출총이익률 변수를 부실 사건의 예측 인자로 고려하였다면, 이는 각 변수가 독립적으로 부실 사건의 발생 확률을 설명하는 정도를 추정할 것이다. 그러나 실제 재무 전문가들은 세 변수 간 유기적 관계 또한 부실 징후의 판단에 중요하게 고려한다. 만약, 기업의 활동성 지표인 매출채권회전율과 재고자산회전율이 비정상적으로 감소한 상황에서 수익성 지표인 매출총이익률이 증가하였다면, 이는 경영진이 매출이나 재고자산을 과대계상하여 기업 부실을 은폐하고 있을 가능성을 암시한다. 합리적인 재무 전문가라면 매출채권의 허위 계상 여부나 재고자산의 실재성 등에 대해 추가적인 확인을 얻고자 할 것이다.

위험 관리 모형을 제공하고자 한다. 아울러, 재무제표 주석에 기업의 부정적인 경제 현상에 대한 보충적 설명이 충실히 제공되고 있는지를 검토하여, 회계의 보수주의 원칙이 주석 정보에도 적용되고 있는지 학술적으로 평가한다.

본 보고서의 구성은 다음과 같다. 이어지는 제 I 장 2절에서는 본 연구의 핵심 결과를 요약하고 시사점을 논의한다. 제 II 장에서는 재무제표 주석 정보의 유용성과 부실 징후 예측에 관한 선행연구를 개관하며, 제 III 장에서는 비정형 정보를 포함한 재무제표 자료의 구축 방법과 신경망 기반 학습 모형에 기반한 부실 징후의 예측방법론을 설계한다. 제 IV 장에서는 이를 활용한 분석 결과를 제시하고, 마지막으로 제 V 장에서 결론을 맺고, 연구의 한계점과 향후 과제를 고찰한다.

2. 주요 연구결과

신경망 모형에 기반하여 부실 징후를 예측하기 위해서는 먼저 부실기업 특성에 대한 충분한 학습이 필요하다. 본 연구에서는 예측 모형의 학습에 활용할 정보를 선별하는 과정에서, 한계 상황에 처한 기업의 특성을 고려하였다. 구체적으로 이자보상배율로 측정되는 한계 요인을 수익성, 안정성, 활동성, 신용위험, 시장위험으로 분해하여, 30개의 지표를 선정하였다. 또한, 기업 부실 정보를 보다 포괄적으로 반영하기 위해, 외부 모니터링 원천별 정보도 고려하였으며, 이에 는 감독 당국의 시장 조치, 투자자의 기업 가치평가, 그리고 기타 무형경제적 특성을 반영하는 13개 지표가 포함된다. 나아가, 재무제표 내 주요 계정과목 37개를 추가하여 고유 금액 정보를 통해 포착할 수 있는 기업 부실 특성을 종합적으로 학습할 수 있게 설계하였다. 마지막으로, 재무제표에 포함된 방대한 정성적 정보를 동시에 학습하기 위해 멀티모달 신경망 모형에 기반한 머신러닝 방법론을 활용하였다.

한편, 기존 연구에서는 상장폐지 기업을 대상으로 부실 예측 모형을 구성하면서 거래정지 여부를 명시적으로 고려하지 않고 있는데, 본 연구에서는 상장폐지에 앞서 장기간 거래정지를 동반하게 되는 심사 절차 및 개선 기간 부여 이전의 시점을 기준으로 예측 모형을 설계하였다. 이는 상장폐지 절차가 길어질수록 적시성이 높은 정보의 활용을 제한하게 되어 모형의 예측력 제고에는 불리한 요소로 작용할 것이나, 부실 징후를 사전적으로 예측하여 실제 투자자의 결정에 활용할 수 있는 실효적 모형을 고안한다는 점에서 중요한 함의를 가진다.

2005년부터 2019년까지 유가증권 및 코스닥시장에 상장된 12월 말 결산 비금융업종 16,815 기업-연도를 대상으로 분석한 주요 연구결과는 다음과 같다. 첫째, 재무제표의 정량적 정보만을 활용한 멀티모달 신경망 모형의 학습 결과, 기업 부실화로 인한 상장폐지 가능성을 매우 높은 정확도로 예측할 수 있었다. 구체적으로, 개별재무제표 기준 99.4% 이상, 연결재무제표 기준 99.5% 이상의 정확도를 기록했다. 특히, 이는 상장폐지 실질 심사와 같은 거래정지 사유가 발생하기 이전의 시점에서 이루어진 예측 결과라는 점에 주목할 만하다. 기존 연구는 거래가 장기간 정지된 상황에서 공시된 재무제표 정보를 포함하여 예측 모형을 구축한 것으로 보인다.⁰⁶ 반면, 본 연구에서는 실효적인 조기 경보 체계로서의 활용성을 담보하기 위해, 정보이용자가 실질적으로 의사결정이 가능한 시점에 예측을 수행한 가운데, 높은 예측 정확도를 달성하였다는 점에서 차별성을 가진다.

그러나 위의 예측 결과는 상당 부분 상장이 유지될 기업을 정확하게 식별하여 도출된 결과라는 점을 유의하여 해석할 필요가 있다. 매수(long) 포지션을 선호하는 투자자에게는 극단적인 부실화 가능성이 있는 기업을 사전에 높은 확률로 투자 대상에서 제외할 수 있어 상당한 효용 가치가 있을 것이나, 실제로 상장폐지 가능성이 높은 기업을 조기에 탐지하여 관리할 목적이 강한 감독 당국으로서는 효용성이 떨어질 수 있다. 요컨대, 정량적 정보만을 학습한 신경망 모형은 조기 경보 체계로서 높은 예측 정확성을 보이나, 극단적인 부실이 발생하지 않을 기업과 발생할 기업 간 예측 정확도의 균형성 측면에서는 명확한 한계가 존재한다.

둘째, 주식의 비정형 정보를 포함한 재무제표의 총체적 정보를 모형 학습에 활용한 결과, 개별재무제표 기준 98.4% 이상, 연결재무제표 기준 99.3% 이상의 예측 정확도를 기록했다.⁰⁷ 다만, 본 연구에서 선별한 정량적 정보만으로도 예측 정확도가 99%에 이르는 것으로 확인된 만큼, 주식의 비정형 정보를 추가 학습하더라도 정량적 정보만을 활용한 경우에 비해 경제적으로 유의미한 예측력 향상은 제한적임을 시사한다. 이에선 정량적 정보 학습을 위해 설계된 신경망 모형에 실질적인 개선 여지가 적은 영향뿐만 아니라, 상장폐지 가능성이 높은 기업을 정확히 판별하는 정확도의 균형성 측면에서 예측 품질의 개선

06 거래정지 상황에서 공시된 정보를 포함하여 추정된 결과는 실질 심사 단계 및 개선 기간에 드러난 악재성 정보가 추가로 고려되는 만큼 모형의 예측력은 제고되었을 것이나, 실제 투자자의사결정에 활용이 어려워 조기 경보로서의 효용성은 제한된다.

07 기존 선형모형과 비교했을 때, 정확도가 11% 증가한 것으로 나타났으나, 이는 주로 정량적 정보의 기계학습 효과에 의한 결과로 해석된다. 학습 효과에 대하여 보다 구체적인 논의는 IV장에서 다루기로 한다.

이 소폭에 그친 점도 영향을 미친 것으로 보인다. 이러한 결과는 재무제표 주식 정보에 회계의 보수주의 원칙이 충분히 반영되지 않고 있음을 시사할 수 있다.

추가적으로, 부실 예측의 질적 수준을 결정하는 요인을 분석한 결과, 우리나라 시장에서 기업의 부정적 경제 현상에 대한 주식의 정보성을 높이기 위해서는 정보량 측면에서 i) 새로운 정보가 충분히 공시될 필요가 있으며, 공시의 양태 측면에서도 ii) 핵심 정보를 강조하되, iii) 기간 간, iv) 유사 기업 간 비교가능한 방식으로 공시될 필요가 있음을 확인할 수 있었다.

결론적으로, 머신러닝 기반 예측 모형의 강점을 극대화하려면, 첫째, 주식에 기업의 고유 정보가 충실히 표현되어야 한다. 둘째, 그러한 고유 정보를 기계가 이해하는 데 어려움이 없어야 한다. 학습에 활용되는 정보의 품질과 가독성(readability)이 담보되지 않으면, 예측 결과의 정확성과 질적 수준을 제고하기 어렵기 때문이다(Allee et al., 2018). 이에 최근 글로벌 공시 제도 개선의 화두는 인간과 기계의 협력을 어떻게 강화할 것인지에 있다.⁰⁸ 핵심 과제는 정보성이 있는 내용을 기계의 가독성을 고려한 형태로 제공하는 것으로, 정보이용자들이 더욱 효율적으로 의사결정을 할 수 있도록 기계의 활용 기반을 고도화할 것을 강조한다(Li et al., 2024).

이러한 흐름은 우리나라 정책당국에도 중요한 시사점을 제공한다. 우리나라는 높은 수준의 전자공시시스템을 구축하는 데 주목할 만한 성과를 이뤘지만, 여전히 일부 기업의 재무제표는 기계가 읽기에 적합한 방식으로 제공되지 않고 있다. 아울러, 기존의 선행 연구는 주식의 정보적 유용성을 제한하는 대표적 요인으로 주식 내 반복적 표현을 지적하고 있는데(이재경·한봉희, 2019), 이는 주의력에 제약이 있는 인간의 시각에서 분석된 결과에 해당한다. 본 연구의 실증결과, 기계의 관점에서 반복적 표현은 중요한 정보의 명확한 강조로 인지될 수 있다. 이에 비추어 앞으로는 근원 정보의 유용성을 제고하되, 우리나라 또한 인간과 기계의 협력성을 강화하는 관점에서 공시 체계의 장기적인 발전 방향을 모색할 필요가 있다.

본 연구보고서는 우리나라 상장기업의 부실 징후 예측에서 재무제표의 본문과 주식 정보의 유용성을 종합적으로 평가한 선도적 연구로, 중요한 학술적 기여가 기대된다. 단, 예측의 균형성을 고려할 때, 본 연구모형을 실무적으로 광범위하게 적용하기에는 그 한계점도 명확해 보인다. 예측의 근간이 되는 주식 정보의 품질을 근본적으로 개선하는

08 Bauguess(2018. 5. 3)

머신러닝을 활용한 재무제표 정보의 유용성 평가: 부실 징후의 조기 탐색을 중심으로

가운데, 모형의 정밀도와 재현율을 함께 제고하기 위한 다각적인 후속 연구가 필요할 것으로 판단된다.

II. 선행연구

1. 재무제표 주식 정보의 유용성에 관한 선행연구
2. 부실징후 예측에 관한 선행연구



II. 선행연구

본 장에서는 재무제표 주석 정보의 유용성에 관한 학계의 축적된 논의를 검토하고, 기업의 부실 징후를 예측한 선행연구의 방법론을 조사한다. 이를 통해 주석 정보를 포함하는 완전한 조합으로서의 재무제표 정보가 기업의 경제적 부실 징후를 포착할 가능성이 있는지를 탐색하고, 나아가 부실 징후 예측에 관한 기존 연구모형이 조기 경보 체계로서 효용성이 있는지를 검토한다. 이와 같은 논의를 종합하여 기업의 부정적 경제 현상을 실효적으로 예측하기 위한 재무제표 정보의 활용 방안을 고찰한다.

1. 재무제표 주석 정보의 유용성에 관한 선행연구

K-IFRS 도입 이후 재무제표에서 주석 정보의 양과 중요성이 크게 증가하였다(이재경·한봉희, 2019; 현지원 외, 2022). 이는 원칙중심(principle-based) 기준에 따라 경영진의 회계 선택(accounting choice)에 대한 재량권이 증가하였을 뿐만 아니라, 공정가치 평가 확대와 회계 기준의 복잡성 증가 등 다양한 요인에 의해 주석에 공시해야 하는 정보의 범위가 확장되었기 때문이다. 특히, 재무제표 본문에는 핵심적인 요약 정보만을 제시하고, 세부적인 내용은 주석에 기재하도록 요구되면서 주석의 보충적 정보 제공 기능이 강화되었다.

이처럼 재무제표에서 주석이 담당하는 역할이 크게 확대되고 있으나, 주석 정보의 유용성에 관한 선행연구는 혼재된 결론을 제시하고 있다. Imhoff et al.(1993)은 효율적인 자본시장이라면 가치 관련 정보는 가격에 즉시 반영되므로, 재무제표 본문과 주석 정보는 본질적으로 동등한 유용성을 가진다고 주장하였다. 반면, 동일한 정보를 주석에 표시할 경우, 본문에 기재할 때보다 정보 전달력이 감소한다고 보고한 실증 연구들도 존재한다.⁰⁹ 이는 기업과 계약 관계를 맺는 여러 이해관계자가 재무제표 본문 정보를 주석 정보보다 더욱 공인된 정보로 계약에 반영할 가능성이 높다는 계약이론(contracting theory) 관점으로 설명되기도 하나(Barth et al., 2003), 실무적으로는 정보이용자의 제한된 주의

09 예를 들어, 자산손상(Aboody 1996), 퇴직급여부채(Davis-Friday et al., 1999), 파생상품 공정가치(Ahmed et al., 2006), 주식기준보상(Frederickson et al., 2006), 리스회계(Libby et al., 2006) 등의 항목에서 재무제표 본문과 주석 간 정보 전달력의 차이가 있음이 실증된 바 있다.

력(limited attention) 내지는 부주의(inattention)로 인해 가치관련 정보가 효율적으로 가격에 반영되지 못하는 현상과 밀접한 관련이 있는 것으로 설명된다(예: Hirshleifer & Teoh(2003), Chen et al.(2023) 등).

실제 많은 양의 주식 정보가 공시될수록, 투자자들은 중요 정보를 식별하고 파악하기가 어려워져, 정보의 전달 효과가 감소하는 것으로 보고된다(Iannaconi, 2012). 따라서, 주식 공시는 핵심적이고 새로운 정보를 강조하는 방식으로 효율화될 필요가 있다는 의견이 제기된다(Bloomfield, 2012). 특히, Henderson(2016)의 질적 연구에 따르면, 주식 정보가 늘어나고 복잡해지면서 상당수 정보이용자는 주석을 재무제표의 일부로 인식하면서도, 이를 충실히 읽고 의사결정에 활용하지는 않고 있는 것으로 응답하였다.

K-IFRS 도입 이후 주식 정보의 작성 실태를 조사한 국내 선행연구 역시 주식 정보의 유용성에 대해 유사한 우려를 제기한다. 이재경·한봉희(2019)는 국내 상장기업 20개의 주식 작성 현황을 조사하여, 주식 정보의 유용성을 저하시키는 요인으로 i) 방대한 정보량, ii) 유기적 정보의 산발적 제시, iii) 회계정책 관련 기준서 조문의 형식적 기술, iv) 기업 간 정형화된 서술 등을 도출하였다. 현지원 외(2022)도 기업 간, 그리고 기간 간 주식 내용의 코사인(cosine) 유사도가 높다는 점을 지적하며, 각종 참고 서식¹⁰에 지나치게 의존하여 작성된 주석이 기업의 고유한 상황을 충분히 반영하지 못하고 있다고 비판하였다.

한편, 이준일 외(2023)는 2018년 11월 「주식회사 등의 외부감사에 관한 법률(이하 외부감사법)」이 전면적으로 개정 시행된 이후, 참고 서식에 의존한 정형화된 주식 공시가 줄어들었을 가능성을 제기하였다. 일반적으로 기업은 감사인이 제공하는 템플릿을 활용하여 주석을 작성하는데, 동일한 외부감사인을 선임한 기업 간에는 동일한 템플릿이 공유되므로, 주식 정보의 유사도가 높게 나타나는 경향이 있다(현지원 외, 2022). 그러나 개정 외부감사법 시행 이후 이러한 유사도가 감소한 것으로 확인되었으며, 이는 주석을 포함한 재무제표 작성에 관한 기업의 책임이 강화되고, 외부감사 환경도 더욱 엄격해지면서 주석 작성 실태가 개선된 결과로 해석된다.

모예린·서윤석(2019)은 기업 내 기간 간 주석 유사도 변화가 자기자본비용에 미치는 영향을 조사하였으며, 전년 대비 주석 내용의 변화가 클수록 자기자본비용이 감소한다는 결과를 도출하였다. 동 연구는 이를 주석에 새로운 정보가 반영됨에 따라 정보비대칭성이

10 예를 들어, 금융감독원이 제공하는 주석 공시 모범사례 및 공시서식 작성기준, 대형 회계법인이 제공하는 템플릿 등이 이에 해당한다.

완화된 결과로 해석하고 있으나, 이는 효율적 시장을 전제로 할 때만 가능한 해석임을 명확히 할 필요가 있다. 또한, 유사도 측정치는 주석 내용의 변화가 긍정적 정보에 기반하는지, 아니면 부정적 정보에 기반하는지를 구분하지 못하므로 그 해석과 판단에 제약이 뒤따른다. 만약, 기업이 긍정적 정보는 주석에 상세히 기재하고, 부정적 정보는 은폐한다면, 정보비대칭성이 심화하는 가운데 자기자본비용은 감소할 수 있다(Hutton et al., 2009).

결국, 주석 정보가 기준서의 요구에 따라 핵심적이고 중요한 사항을 위주로 충실히 작성된다면, 본문의 수치 정보를 보완하여 재무제표 전반의 정보성을 높이는 것은 자명하다(IASB, 2017). 반면, 주석에 가치 관련 정보가 충실히 기재되지 않거나, 기재된 정보가 비효율적인 방식으로 제공되어 정보이용자가 이를 읽고 이해하는 데 많은 주의력이 요구된다면, 주석 정보의 유용성은 크게 저하될 수 있다. 다만, 본 연구에서는 머신러닝 기법을 활용하여 방대한 주석 정보를 효율적으로 학습하므로, 정보이용자의 제한된 주의력에 따른 유용성 저하 영향은 상당 부분 배제한 분석이 가능하며, 궁극적으로 기업이 주석에 부정적인 경제 상황을 얼마나 충실히 표현하였는지에 따라 주석 정보의 유용성 수준을 평가할 수 있을 것이다.

2. 부실징후 예측에 관한 선행연구

부실기업의 퇴출은 장기적으로 자본의 효율적 배분을 달성하여 경제 전반의 생산성을 높이는 경로로 작용한다(Banerjee & Hoffmann, 2022). 그러나 단기적으로는 투자자 피해가 불가피하며, 금융 안정의 위협 요인으로 이어질 가능성이 있어 기업 부실 징후를 조기에 예측하기 위한 노력은 학계와 실무계에서 지속되었다. 특히, 글로벌 금융위기 이후 장기간 이어진 저금리 기조로 저비용 차입자본에 의존하는 한계기업의 좀비화(zombification) 현상이 심화하는 가운데, 차입의존도가 높은 기업의 시장 퇴출 우려가 점증하면서(Acharya et al., 2024), 부실 징후를 탐지하기 위한 모형도 발전을 거듭해 왔다.

관련 연구의 초기에는 재무제표상 정량적 회계정보만을 이용하는 회계 모형(accounting-based model)이 주를 이루었다. Beaver(1966)가 기업 부도 예측에 재무비율이 유용함을 주장한 이후 다수 연구는 재무제표 본문의 계정과목을 이용하여 계산한 여러 재무비율 정보를 기반으로 기업의 부도 위험을 예측하였다. 대표적으로 Altman(1968)은 다변량 판별 분석(multiple discriminant analysis)에 기초한 Z-점수 모형을, Ohlson(1980)

은 로짓 분석(logit analysis)에 기초한 O-점수 모형을 기업 부도 위험에 대한 예측 모형으로 제안하였다. 해당 모형에서는 기업의 재무적 안정성을 나타내는 지표들과 함께 잉여현금흐름 특성을 대리하는 수익성 지표 등을 주요 변수로 고려하였다.

이후 시장의 가격 정보를 활용하는 시장 모형(market-based model)이 등장하였는데(Merton, 1974; Bharath & Shumway, 2008), 이는 계속기업(going-concern) 가정이 불확실한 상황에서 채무 상환의 재원이 되는 기업의 청산가치를 더욱 정확하게 추정하기 위함이다.¹¹ 대표적으로 Merton(1974)은 기업이 상환해야 할 부채 규모가 총자산의 시장가치 대비 얼마나 초과하는지를 거리(distance)로 측정하는 모형을 고안하였다. 이때, Bharath & Shumway(2008)는 주가 자료에 기초한 반복(iteration) 추정 방식을 통해 총자산의 시장가치를 추정하고, 부도 거리 및 확률 등을 계산하였다.

상술한 회계 모형과 시장 모형 간 예측의 우월성에 대해서는 한동안 엇갈린 평가가 지속되었는데, 그러한 과정에서 두 원천별 정보를 통합 활용하는 Hazard 모형이 등장하였다(Shumway, 2001). 재무제표의 정량적 정보로 산출하는 재무비율뿐만 아니라 시장에서 거래되는 가격과 변동성에 내재한 정보를 함께 활용하는 경우, 기업의 부도 위험을 더욱 정확하게 예측하는 것으로 나타났다. 특히, Campbell et al.(2008)은 재무비율 계산 시장부상 자산총액을 활용하지 않고, 시장 총자산(market total asset)을 활용하는 Hazard 모형을 제안하였으며, 수익성 비율 역시 최근 성과에 더 높은 가중치를 두어 회계정보가 현재 상황을 적시에 반영하지 못하는 문제를 보완하였다. 국내에서도 이인로·김동철(2015)에 따르면, 회계·시장 정보를 통합 활용한 Hazard 모형이 각각의 모형 대비 우수한 예측력을 보이는 것으로 나타났다.

이와 같은 모형들은 주로 선형 회귀모형에 기반하여 기업의 부도 위험을 예측하는데, 최근에는 머신러닝 기법을 통해 기업의 부도 위험을 학습·예측하는 모형 또한 발전하고 있다. 머신러닝 기법은 기존 선형 회귀모형과 비교하여 더 유연하게 수치 정보의 비선형적 조합이 내포하는 함의를 반영할 수 있어 복수의 계정에서 비정상적인 증감으로 나타나는 부실 위험을 포착하는 데 장점이 있을 뿐만 아니라, 다양한 형태의 정성적 정보까지 분석의 범위를 확장할 수 있다는 이점이 있다(Cao & You, 2024). 국내에서도 오세경 외(2017)는 회계·시장 지표에 뉴스 기사의 키워드 정보를 종합적으로 활용하면 매우 높은

11 일반적으로 재무상태표상 자산총액 정보는 활성 시장(active market)에서 거래되는 가격 정보 대비 적시성이 떨어지며, 역사적 원가(historical cost)로 측정된 자산이 많을수록 공정가치(fair value)와의 괴리가 커진다.

정확도로 부실 위험 예측이 가능함을 실증하였다.

한편, 회계 부정의 관점에서 기업 부실 징후를 탐지하는 모형을 고안하는 연구도 존재한다. 기업의 진행 중인 부실은 경영진이 공격적인 회계처리를 할 동기(incentive)와 압박감(pressure)으로 작용하기에(Cressey, 1986), 회계 부정과 기업 부실 징후는 매우 밀접한 관련성을 보인다. 본 영역에서도 초기에는 Beneish(1999)의 M-점수 모형, Dechow et al.(2011)의 F-점수 모형 등 로짓 분석에 기초하여 재무제표의 중대한 왜곡표시 가능성을 예측하는 모형이 주를 이루었으나, 최근에는 머신러닝을 활용한 연구도 활발하다. 초기에는 머신러닝 기반 모형이 기존의 로짓 모형 대비 예측력에 큰 우월성이 없는 것으로 보고되기도 하였으나(Perols, 2011), 최근에는 머신러닝의 알고리즘이 개선되고 학습 범위도 확장되면서 머신러닝의 예측력이 우월하다는 실증결과가 보고된다. 대표적으로, Bao et al.(2020), Bertomeu et al.(2021) 등은 미국 시장에서 앙상블(ensemble) 머신러닝 기법¹²이 기존 로짓 모형보다 부정 징후에 대한 예측력이 우수함을 실증하였는데, 이러한 결과는 국내 기업을 대상으로도 일관된 것으로 확인된다(나현중 · 정태진, 2022).

이처럼 국내 · 외 선행연구는 기업의 부도, 상장폐지, 회계부정으로 인한 감리 지적 등 다양한 부실 사건에 대한 예측 모형을 개발하고 발전시켜 왔다. 아울러, 부실 징후 예측을 위해 활용하는 정보의 범위 또한 회계 · 시장정보에 더하여 뉴스 기사를 비롯한 비정형 정보까지 확장하고 있다. 단, 기존 연구에서는 재무제표 본문을 중심으로 한 정량적 정보에 한정하여 회계정보를 활용한 경향이 있으며, 주식 정보까지 종합적으로 예측에 반영하여 모형을 개선하고자 한 시도는 드문 상황이다.

더욱이, 상당수 연구는 기업의 부실 사건을 상장폐지 여부로 정의하는데, 국내의 경우 상장폐지에 앞서 거래정지가 빈번한 제도적 환경이 명확히 고려되었는지가 불분명하다.¹³ 한국거래소는 거래정지와 관련한 시장 조치가 발동하는 경우 원인 사유의 해소를 거래 재개 요건으로 둔다. 따라서, 상장폐지 심사 사유와 같이 단기에 해결이 어려운 부정적 사건이 발생하면 상당 기간 거래가 불가능한 상황에 놓이게 된다. 그뿐만 아니라, 우리나라는 장외시장이 발달하지 않아 정규 시장 퇴출 이후 거래의 재개 가능성이 낮은 제한 환경적 특성이 있다. 이에 본 연구에서는 상장폐지 사유의 발생으로 거래가 정지되기 이전에

12 앙상블 기법은 부실기업과 같이 정상기업 대비 표본 수가 현저히 적은 반면 부실 징후와 연관된 지표들이 사례별로 다양하고 이질적일 경우 표본의 불균형성을 완화하여 예측력을 강화할 수 있다.

13 대다수 연구는 상장폐지, 회계 부정 등 기업 부실 사건의 발생 여부를 파악할 때, 시장에서 거래가 가능한 상황인지 여부를 명시적으로 고려하지 않고 있다.

가용한 정보만을 예측 정보로 활용하고자 한다. 이를 통해, 예측력의 과대평가 가능성을 완화하고, 투자자 및 감독 당국의 관점에서 실제 효용 가치가 있는 예측 모형을 구축하고자 하였다. 상세한 방법론은 이어지는 제Ⅲ장에서 후술한다.

III. 연구방법론

1. 멀티모달 신경망 모형
2. 모형의 학습 과정



III. 연구방법론

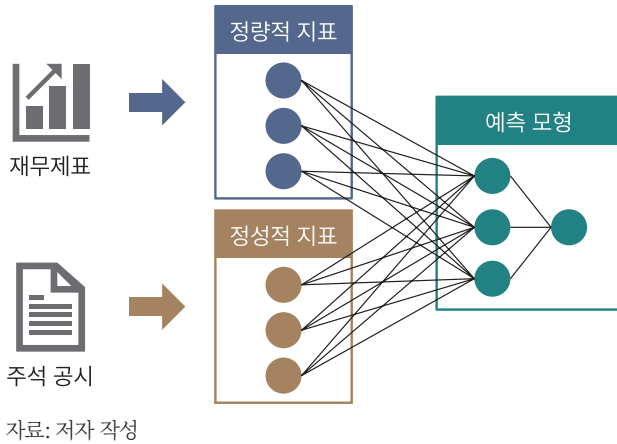
본 장에서는 재무제표 등 공시 정보를 학습시킨 머신러닝 모형을 바탕으로 기업의 부실 징후를 판단하는 실증분석 방법론을 소개한다. 기업의 재무적 특성을 포괄적으로 활용하기 위하여 재무비율 등 정량적인 지표와 주식 공시자료 등 정성적인 지표를 일괄적으로 학습하는 멀티모달(multi-modal) 신경망(neural network) 모형을 바탕으로 개별 기업의 상장폐지 확률을 예측하여 이를 기업의 부실 징후에 대한 지표로 사용한다. 이때, 첫 거래정지 시점으로부터 상장폐지까지 상당한 기간이 소요되는 점을 고려하여 거래정지 시점에 가용한 최신의 재무 정보만을 활용해서 기업의 부실 징후를 조기에 탐지하도록 모형을 설계하였다.

1. 멀티모달 신경망 모형

본 고에서는 재무 공시 정보의 분석에 특화된 멀티모달 신경망 모형을 구축하여 공시 정보의 유용성을 평가하였다. 멀티모달 모형은 서로 다른 형태를 가지는 데이터를 동시에 학습할 수 있도록 유연성이 높은 모형으로서 현실에서 사람이 정보를 처리하는 방식에 보다 가깝도록 진화한 머신러닝 기법이라고 할 수 있다.

멀티모달 모형에서 말하는 모달리티(modality: 양식, 양상)는 데이터가 기록된 형식을 의미한다. 예를 들어, 수치 정보와 문자, 사진, 영상 등은 서로 모달리티가 다른 학습 데이터라고 볼 수 있다. 따라서 멀티모달 모형이란 서로 다른 형태로 기록된 학습 데이터를 포괄적으로 사용하여 추론 및 예측을 수행할 수 있는 단일 모형 구조를 의미한다. 이처럼 서로 다른 형태의 데이터를 종합적으로 처리하고 이들 사이의 연관성을 추론하는 사고방식은 인간이 다양한 기관을 통해 감지한 정보를 분석하고 이를 바탕으로 추론하는 과정을 머신러닝으로 구현하는 방법론으로서 비교적 최근에 급속하게 발전하고 있다.

<그림 III-1> 멀티모달 신경망 모형의 학습 구조



재무 공시를 분석하는 데 있어 멀티모달 모형이 가지는 장점은 인간이 재무 정보를 이해하고 이를 처리하여 판단하는 과정에 가장 가까운 수리통계적인 방식이라는 점이다. 일반적으로 재무 공시 정보는 재무제표로 제공되는 수치적인 정보뿐만 아니라 이를 해석 및 보완하는 주식 공시를 포함하여 일컫는다. 멀티모달 신경망 모형은 재무제표에서 추출한 정량적인 특징(feature)과 주식 공시에서 추출한 정성적인 특징을 동시에 입력 정보로 활용하여 예측 모형을 구축한다(<그림 III-1> 참조). 따라서 수치로 표현된 전자와 문자로 표현된 후자를 연결 지어 포괄적으로 분석하고 의미를 추론하는 과정이 수반되는데, 이는 멀티모달 모형의 구조와 매우 유사함을 알 수 있다.

멀티모달 모형은 서로 다른 모달리티를 가진 학습 자료를 통합(fusion)하는 방식에 따라서 크게 세 가지 형식으로 나뉜다. 첫 번째는 early fusion으로 서로 다른 모달리티를 가진 학습 자료를 동일하거나 비교적 유사한 차원의 자료로 수치화한 후 예측 모형에 일괄적으로 입력하는 방식이다. 두 번째로 late fusion은 자료의 모달리티 개수만큼의 신경망 모형을 독립적으로 학습시킨 후 그 결과를 다시 포괄하는 모형을 학습시킨 후 이를 종합적으로 평가하는 방식이다. 마지막으로 intermediate fusion은 서로 다른 모달리티 개수만큼의 잠재변수를 상정하고 이를 은닉층(hidden layer)으로 포함하는 하나의 신경망 모형을 학습시키는 방식이다.

본 연구에서는 재무비율 및 시장지표로 이루어진 정량적 자료와 재무제표 본문 및 주석으로 이루어진 정성적 자료의 두 가지 모달리티를 가진 데이터를 학습하는 신경망

모형을 통하여 포괄적인 재무제표 분석 및 평가 방식을 머신러닝 모형으로 구현하였다. 이때, 정량적 정보와 정성적 정보에 맞는 각각의 전처리(pre-processing) 과정을 거쳐 두 가지 정보의 차원을 유사하게 표현하여 동일선상에서 예측 지표로 활용하는 early fusion 방식을 사용하였다. 이와 같은 방식은 입력 자료의 전처리와 학습을 거쳐 예측으로 이루어지는 모형의 추론 과정을 직관적으로 이해할 수 있다는 장점이 있다.

2. 모형의 학습 과정

가. 학습 데이터의 전처리

모형의 학습에 사용된 데이터로는 주요 재무비율 및 시장지표를 사용하여 구축한 정량적 변수들과 더불어 사업보고서의 재무제표 개별 계정 및 주식 공시에서 추출한 정성적 지표를 종합적으로 사용하였다. 학습에 포함된 전체 표본의 수는 회계연도 기준으로 2005년에서 2015년 사이 10,361개 연간 사업보고서가 대상이며 이를 바탕으로 2016년에서 2019년 사이 6,454개의 사업보고서를 대상으로 학습된 모형을 평가하였다(<표 III-1> 참조). 수집된 표본에서 2020년 이후 사업보고서는 학습 및 평가에서 제외하였는데 이는 팬데믹 기간이 산업 전반에 미친 영향으로 인하여 기업의 재무 성과와 건전성 지표가 이전 기간과는 현저하게 다른 양상을 보이는 것을 감안하였다. 학습은 시계열적으로 표본의 수를 점진적으로 증가시키는 순차적 예측을 사용하였는데 이는 본 절의 마지막에서 보다 자세히 설명한다.

<표 III-1> 모형의 학습을 위한 데이터 선택 과정

	학습 데이터(Training)	평가 데이터(Testing)
회계연도	2005 ~ 2015	2016 ~ 2019
전체 표본 수	24,477	10,009
재무 변수만 있는 경우	(12,888)	(3,159)
공시 문서만 확인된 경우	(1,228)	(396)
가용 표본	10,361	6,454

주 : 표본 수는 서로 다른 기업-연도 쌍의 개수를 의미하며 재무 변수만 계산 가능하거나 공시 문서만 확인된 경우는 연구에서 제외하였다.

자료: DataGuide, 금융감독원 DART

학습에 사용된 기업-연도 표본에 대응하는 변수 자료는 금융감독원 공시시스템(DART)과 DataGuide를 활용하여 수집하였다. DataGuide에서는 정량적 변수인 재무비율 및 시장지표 등을 수집하였으며 DART에서는 공시 자료 원문을 확보하여 언어적 정보에 기반한 변수를 구축하는 데 활용하였다. 두 가지 다른 원천 자료의 연결은 기업의 종목코드와 회계연도의 일치 여부로 판단하였다. 이 과정에서 필연적으로 적지 않은 자료가 연결되지 않을 수 있으며 <표 III-1>에서도 동일한 문제를 확인할 수 있다.¹⁴ 학습 결과의 강건성 및 예측 결과의 일관성을 보장하기 위하여 본 연구에서는 연결되지 않은 자료를 제외하고 실증분석을 수행하였다.

정량적 변수로는 DataGuide에서 수집한 상장기업의 연도별 재무비율 및 시장지표를 바탕으로 변수 목록을 구축하였다(부록 참조). 정량적 변수는 종류별로 크게 세 가지 군으로 분류할 수 있다. 첫 번째는 재무 건전성과 연관되는 다양한 요인 변수들로 이는 다시 수익성, 안정성, 활동성 측면과 더불어 신용위험과 시장위험을 반영하는 지표를 포함한다. 두 번째로 외부의 정보 중 재무제표의 특성을 파악하는 데 도움이 되는 변수들을 함께 고려하였다. 정보의 원천에 따라 감독 당국의 재량적 판단에 사용되는 지표와 더불어 투자자의 기업 가치 평가를 반영하는 변수 그리고 무형자산 비율과 주식발행 척도가 학습 변수에 포함된다. 마지막으로 고유 계정항목을 학습 변수로 사용하였는데 이는 재무제표 및 본문으로부터 문자 정보를 추출하는 과정에서 생략되는 정량적인 수치를 학습에서 누락시키지 않으면서 동시에 모형의 예측력을 높일 수 있기 때문이다.¹⁵

정성적 지표로는 금융감독원 DART에서 가용한 사업보고서 원문에서 재무제표 관련 공시를 추출한 후, 재무제표 본문 및 주석에 포함된 문자 정보를 자연어 처리(Natural Language Processing: NLP) 기법을 활용하여 전처리한 변수를 사용하였다. 멀티모달 모형을 학습시키기 위한 과정으로서 NLP의 핵심 개념은 문자로 표현된 정보를 수치로 변환하여 앞서 설명한 재무비율 등 정량적 학습 자료와 차원을 일관되도록 변환하는 데 있다.

14 <표 III-1>에서도 확인할 수 있듯이 연결되지 않은 자료의 빈도는 점차 감소하고 있는데, 이는 디지털 공시의 확산과 공시 자료에 대한 기계적인 접근에 대한 빈도가 증가하면서 선진국의 시장 감독 기관을 중심으로 기업 공시 아카이브(archive)를 고도화하는 노력으로 이어지고 있는 것(홍지연, 2023; 노성호, 2024)과 무관하지 않은 것으로 보인다.

15 Bao et al.(2020)에 따르면 회계부정을 통하여 부실 징후를 예측하는 머신러닝 모형의 성과는 재무비율 외에도 해당 비율을 계산하는 데 사용된 원자료인 계정항목을 직접 학습할 경우 더욱 향상된다는 점을 실증적으로 확인하였다.

본 연구에서는 재무제표 본문 및 주석 공시에서 숫자와 특수기호를 제외하고 한글로 작성된 문서 부분을 TF-IDF(Term Frequency-Inverse Document Frequency) 방식으로 전처리하여 정량적 지표와 함께 학습시키는 방식을 사용하였다. TF-IDF 방식으로 문서를 변환하는 과정은 다음과 같다. 우선 표본 내 문서에서 조사, 관계사, 대명사 등 재무적인 의미를 포함하지 않은 표현을 제거한 단어들의 집합¹⁶을 구한 다음 단어별로 문서 안에서 사용된 횟수를 계산한다. 이처럼 문서마다 포함된 단어의 빈도를 나타내는 행렬을 Term Frequency 행렬이라고 한다. 다음으로 모든 문서 집합에서 각 단어가 사용된 빈도의 역수로 단어의 정보량을 계산하는데 이를 Inverse Document Frequency라고 한다. 표본 내 문서 전체의 집합을 D 라고 할 때, 단어 t 의 inverse document frequency는 다음과 같이 계산된다.

$$IDF(t,D) = \ln\left(\frac{|D|}{|d \in D: t \in d|}\right) + 1$$

이때, $|A|$ 는 집합 A 에 포함된 원소의 개수를 의미한다. 따라서 단어 t 를 포함하는 문서가 많을수록 $IDF(t,D)$ 는 1에 가까운 값을 가지게 되어 해당 단어가 특정 문서의 정보량에 기여하는 바가 낮음을 의미한다. 마지막으로 각 문서 안에 포함된 단어에 중요도를 곱하여서 문서-단어 행렬(Document-Term Matrix: DTM)을 계산할 수 있다.

<표 III-2> TF-IDF를 활용한 자연어 처리 예시

Term Frequency				Inverse Document Frequency	Document-Term Matrix						
	채권	대손	전액			채권	대손	전액			
문서1	9	3	1	×	채권	1.00	=	문서1	9.00	3.87	1.69
문서2	6	2	0		대손	1.29		문서2	6.00	2.58	0.00
문서3	5	0	1		전액	1.69		문서3	5.00	0.00	1.69
문서4	8	1	0					문서4	8.00	1.29	0.00

주 : 위 예시는 개념 설명을 위해 가공한 자료이며, 실제 주석 사례에 기재된 단어의 빈도와는 일치하지 않을 수 있다.
 자료: 저자 작성

16 문장안에서 특수한 의미를 가지지 않는 조사, 관계사, 대명사 등을 흔히 불용어(stopwords)라고 하는데 이는 전처리 과정에서 대체로 삭제된다. 더불어 본 연구에서는 전체 표본에서 극히 낮은 빈도로 등장하거나 지나치게 많이 등장하는 단어도 삭제하였다.

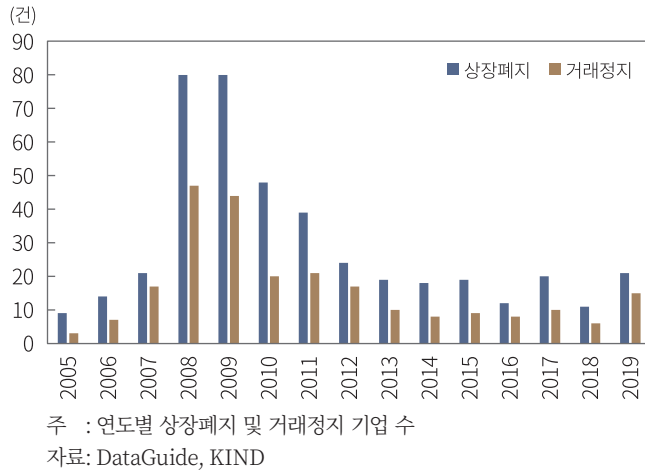
<표 III-2>는 4개의 문서와 3개의 단어를 포함한 표본에서 TF-IDF 방식을 통해 문자 정보를 수치화하는 예시를 보여주고 있다. 예시로 사용된 단어 중 ‘채권’의 경우 모든 문서에 사용되고 있어 특정 문서가 내포하고 있는 정보를 다른 문서와 구분하는데 유용하지 않은 단어로 볼 수 있다. 따라서 해당 단어는 DTM에서 정확히 단어의 빈도와 같은 가중치를 얻고 있다. 한편 ‘대손’은 문서3에서는 등장하지 않기 때문에 ‘채권’보다 문서들 사이의 비교에 유용한 정보를 제공하며, 따라서 IDF가 1보다 큰 값을 가진다. 이를 반영하여 DTM에서는 ‘대손’이 등장하는 문서에서 해당 단어가 단순한 빈도보다 큰 가중치를 부여받고 있다. 마지막으로 ‘전액’은 가장 낮은 빈도로 사용되어 IDF는 앞선 두 개의 예시 단어보다 큰 값을 가진다. 하지만 ‘전액’이 등장하는 문서1에서도 ‘채권’과 ‘대손’보다 낮은 빈도로 사용되어 DTM에서 낮은 가중치로 반영되고 있음을 확인할 수 있다.

정성적 공시자료를 DTM으로 변환하는 장점으로 정량적 학습 자료와 연결하여 통일된 구조를 가지는 하나의 모형을 구축할 수 있다는 점을 들 수 있다. 앞 절에서 설명한 바와 같이 본 연구에서는 서로 다른 양식을 가진 재무 변수와 주식 공시를 함께 학습하는 멀티모달 모형에 기반한 분석을 수행하는데 이 중 문자로 구성된 공시 정보를 DTM 형태의 수치 행렬로 변환할 경우 재무 변수만 고려하는 모형과 구조적으로 동일한 형태의 신경망 모형을 학습시킬 수 있게 된다. 따라서 모형의 구조를 유지하면서 입력하는 정보의 양을 변화시키는 실험을 통해 주식 공시가 학습 효율성을 제고할 수 있는지에 대한 일관된 비교 분석이 가능해진다.

나. 학습 목표의 설정: 시차를 고려한 상장폐지 지표

기업의 부실 정도를 예측하기 위한 학습 목표로는 상장폐지 여부를 나타내는 과거 자료를 활용하였다. 구체적으로 개별 기업의 연도별 사업보고서에 포함된 재무제표 공시와 해당 사업보고서 이후 기업의 상장폐지 여부를 연결하여 모형이 재무제표 공시로부터 상장폐지로 이어지는 연결고리를 학습할 수 있도록 하였다.

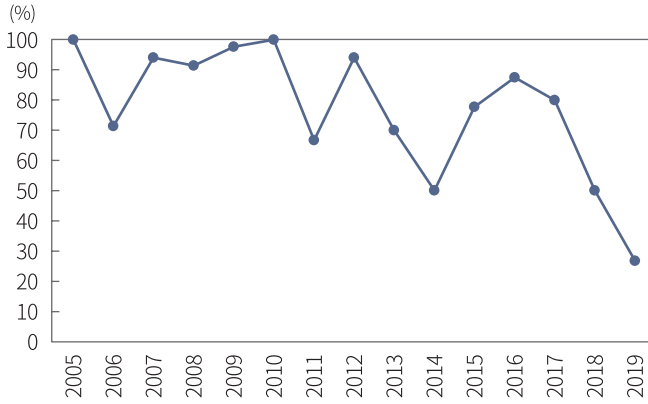
<그림 III-2> 상장폐지 및 거래정지 기업 수



상장폐지를 지표로 사용하여 모형을 학습시키는 데 있어서 한계점은 크게 두 가지로 요약할 수 있다. 첫 번째는 학습에 사용되는 데이터의 총량 대비 상장폐지 건수가 매우 적은 비율로 관측된다는 점이다. 이는 흔히 머신러닝 이론에서 불균형표본(unbalanced sample)의 문제로 언급되는 현상으로서 예를 들어 두 가지 결과(본 연구의 경우에는 상장폐지와 유지)의 비율이 한쪽으로 크게 치우쳐 분포할 때(상장폐지된 표본이 유지된 표본에 비해 현저하게 적을 경우), 모형의 학습 효율이 저하될 수 있다.¹⁷ <그림 III-2>에서 볼 수 있듯이 표본 기간 내 상장폐지 건수는 연도별로 최대 80건을 넘지 않으며 이는 전체 표본수(<표 III-1> 참조) 대비 매우 적은 비율만이 상장폐지로 이어졌음을 알 수 있다.

17 머신러닝 이론에서 학습 데이터가 두 가지 결과 중 하나에 지나치게 편중되어 분포하는 경우를 클래스 불균형(class imbalance) 문제라고 한다. 머신러닝 모형의 학습 알고리즘은 모든 클래스에서 고르게 예측력을 높이도록 구조화되어 있기 때문에 클래스의 불균형이 심할 경우 데이터의 대부분을 차지하는 결과(본 연구의 경우는 상장이 유지되는 기업)에 대한 정확도에 높은 비중을 두어 소수의 표본(상장폐지 사례)에 대한 학습 효율이 저하되는 단점이 있으며(Guo et al., 2008), 이는 표본의 수가 증가하거나 학습 알고리즘을 변경하는 경우에도 대체로 유지된다(Japkowicz & Stephen, 2002). 이를 보완하기 위한 방법론으로는 표본 가중치를 조정하는 등의 방법이 제안되었다(Krawczyk, 2016).

<그림 III-3> 거래정지 이후 상장폐지 비율



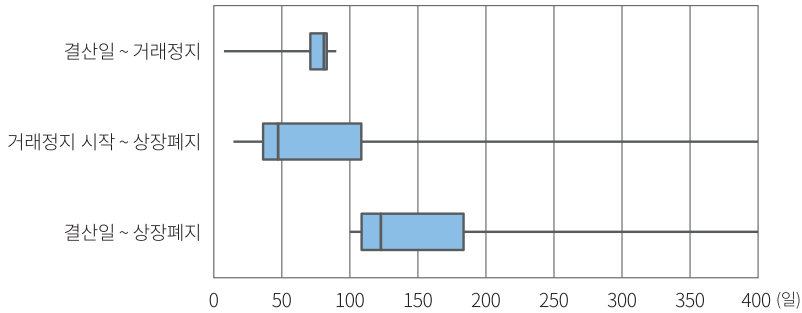
주 : 거래정지건 중 최종적으로 상장폐지된 기업의 비중을 거래정지 발생 직전 결산연도를 기준으로 나타내었다.

자료: DataGuide, KIND

이와 더불어 상장폐지를 학습 목표변수로 정할 경우의 추가적인 단점은 상장폐지에 앞서 거래정지가 먼저 발생하는 경우가 많다는 점이다. 거래정지는 상장폐지에 앞서 발생하는 경우가 많아서(<그림 III-3> 참조) 거래정지 사건의 발생은 해당 기업의 재무상태에 대한 평가에 있어 거래정지가 발생하지 않은 기업과 질적으로 다른 경향을 보일 가능성이 크다.¹⁸ 따라서 상장폐지 직전에 공시된 재무제표를 학습하여 직후에 발생할 상장폐지 사건을 예측할 경우 이미 상장폐지를 유발한 재무적 특성이 학습 대상이 되는 변수에 반영되어 예측력을 과도하게 향상시킬 우려가 있다.

18 <그림 III-3>에서 특히 2016년 이후 거래정지된 기업 중 상장폐지로 이어지는 비중이 줄어드는 것으로 보이는 현상은 상장폐지 실질심사 기간이 길어지는 경향과 연관된 것으로 볼 수 있다. 이에 대하여 보다 자세한 논의는 <그림 III-4>와 관련된 설명을 참조하기 바란다.

<그림 III-4> 거래정지 및 상장폐지 경과 기간



주 : 표본기간(회계연도 기준 2005~2019)내 거래정지와 이어서 상장폐지가 발생한 모든 기업에 대하여 결산일로부터 거래정지 시작일, 거래정지 시작일로부터 상장폐지일, 결산일로부터 상장폐지일까지의 기간의 분포를 상자그림(box plot)으로 나타내었다. 상자의 좌끝단, 중간, 우끝단은 각각 1분위수(25%), 중위값, 3분위수(75%)를 의미하며 막대는 최소값과 최대값을 표시하였으나 지면을 고려하여 최대값은 400일에서 제한하였다.

자료: DataGuide를 바탕으로 저자 계산

거래정지로부터 상장폐지까지의 기간이 평균적으로 상당한 점은 이와 같은 기업의 재무상태를 평가하는 모형이 학습하는 과정에서 상당한 편의를 발생시킬 수 있다. 학습에 사용된 2005~2019년 자료에서 확인된 200건의 거래정지 이후 상장폐지 사례의 분포를 보았을 때, 거래정지는 결산일로부터 비교적 단기간에 발생하는 반면 상장폐지까지는 상당한 시간이 걸리는 것을 알 수 있다(<그림 III-4> 참조). 결산일로부터 거래정지까지 걸리는 기간은 최대 90일을 초과하지 않았으나 거래정지 시작일로부터 상장폐지일까지는 최소 14일에서 최대 581일(중위값 47일)이 소요되어 이를 합산하면 거래정지 직전 결산일로부터 최소 100일에서 최대 622일(중위값 123일)이 걸리는 것으로 확인되었다.

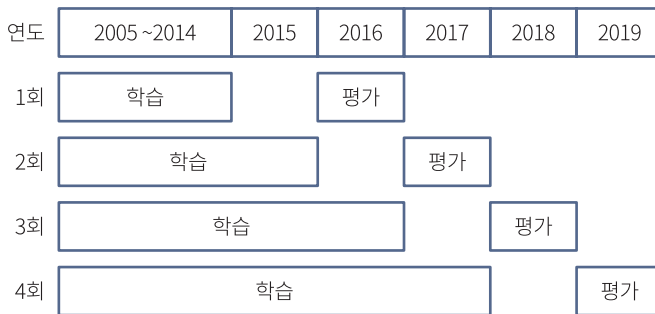
상장폐지 이전 거래정지 기간이 연장되면 해당 기간 안에 최소 한 번의 결산보고서가 공시될 수 있다는 점을 의미한다. 따라서 거래정지 기간 동안에 공시된 재무제표는 상장폐지로 이어지게 되는 재무적인 특성이 미리 반영되어 미래 참조 편향(look-ahead bias)에 노출되었을 뿐만 아니라 해당 재무 공시가 이루어지는 시점에서 투자자는 자산의 매도 등 손실을 회피하기 위한 실질적인 행동이 제한될 수밖에 없다. 이와 같은 문제점에 대한 인식을 바탕으로 본 연구에서는 결산일 기준으로 1년 이내에 거래정지 사건이 발생하고 이후 상장폐지로 이어지는 결과만을 식별하여 해당 결산일의 재무제표 등 공시자료로부터 부실 징후를 추론하도록 모형을 학습시키는 방법을 사용하였다.

다. 순차적 학습 및 평가

모형의 학습 과정은 데이터의 시계열적 특성을 고려하여 학습에 활용되는 데이터의 양을 단계적으로 증가시키면서 학습 정확도를 추정하는 rolling-window 방식을 사용하였다. 예를 들어, 최초 10개년(FY2005~2014) 사업보고서를 학습하여 (FY2015를 제외하고) FY2016 재무제표의 거래정지 가능성을 예측, 이후 1개 연도를 추가하여(FY2005~2015) FY2017 재무제표의 거래정지 가능성을 예측하는 방식이다. 이는 머신러닝에서 일반적으로 사용되는 K-fold 교차검증(cross validation) 방법을 사용하는 경우 미래 정보를 활용하여 과거 예측값을 평가할 수 있으므로 시계열 또는 패널 데이터의 특성에 적합하지 않을 수 있다는 점을 감안하였다.¹⁹

학습 데이터의 양을 점진적으로 증가시키는 과정을 반복하여 사업연도 기준 2016~2019년 재무제표(총 6,454개)에 대한 거래정지 확률 예측치를 계산하고 예측된 확률값과 실제 거래정지된 경우를 비교하여 예측 오차를 계산할 수 있으며 이를 바탕으로 모형의 정확도를 평가하여 모형의 적합도 및 학습 효율성을 검증하였다. 2020년 이후 사업연도에 대한 재무제표는 모형의 평가 과정에서 제외하였는데 이는 팬데믹으로 인하여 기업의 재무 성과가 전 산업에 걸쳐 영향을 받았기 때문에 이전의 재무제표를 학습하여 일관된 예측을 하기 어려울 것으로 판단하였다.

<그림 III-5> 순차적 학습 방식

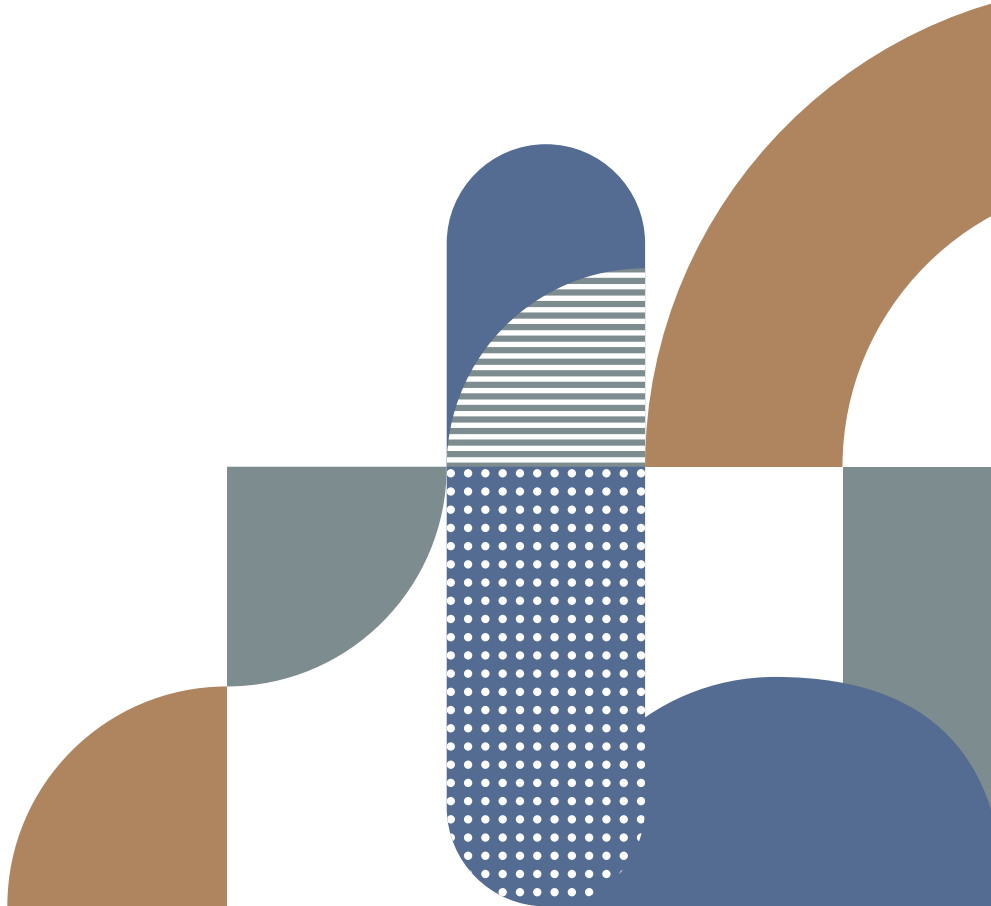


19 다만, 회계 부정을 탐지하는 머신러닝 방법론에 관한 기존 연구에서는 K-fold 교차검증이 널리 사용되고 있다는 점을 밝힌다(Perols, 2011; Perols et al., 2017; Bertomeu et al., 2021).

상장기업 전반의 재무 성과가 연도별로 편차를 가질 수 있다는 점을 고려하여 모형의 학습과 평가는 순차적인 방식으로 진행하였다(<그림 III-5> 참조). 우선 최초 학습 단계에서는 2005~2014년의 10년 치 자료를 학습하여 2016년 표본 기업의 재무 건전성을 예측하고 그 성과를 평가하였다. 이때, 2015년 표본을 의도적으로 생략함으로써 직전 연도의 재무 성과에 대한 학습 결과가 다음 연도의 예측 가능성을 과대평가하는 것을 방지하여 정확도를 보수적으로 평가하였다. 다음 회차에서는 2015년을 학습 자료에 추가하는 한편 2016년을 제외하고 2017년 표본에 대하여 예측치를 계산하고 성과를 평가하였다. 이와 같은 과정을 2019년까지 반복하여 2016~2019년 표본에 대한 성과를 종합적으로 평가하여 모형의 학습 성과를 최종적으로 판단하였다.

IV. 연구 결과

1. 재무제표 및 주식 정보의 학습 효율성 비교
2. 부실 평가 기업의 특성 분석
3. 공시 유용성 기여 요인 평가



IV. 연구 결과

공시 정보를 학습한 멀티모달 모형을 바탕으로 재무제표가 기업의 부실 징후에 대한 정보를 효과적으로 반영하고 있는지를 실증적으로 분석하고자 한다. 실증 분석은 크게 세 가지 방식으로 구성된다. 첫 번째는 앞 장에서 소개한 멀티모달 신경망 모형의 학습 성과를 예측 정확도의 측면에서 평가하고, 특히 재무제표 주식에서 추출한 정성적 정보가 예측력의 향상에 기여하고 있는지 여부를 판단하였다. 두 번째로 학습된 모형을 바탕으로 추정된 기업의 부실 확률 지표를 통계적으로 분석하고 부실 확률이 높게 측정된 표본의 재무적 특성을 분석하였다. 마지막으로 부실 확률에 대한 모형의 예측 결과와 실제 부실기업 간의 괴리도를 바탕으로 재무제표의 실질적 유용성에 기여하는 요소를 회귀분석을 통해 판별하였다.

1. 재무제표 및 주식 정보의 학습 효율성 비교

가. 모형의 학습 효율성 평가 지표

기업의 상장폐지 여부에 대하여 실제와 예측 결과가 다르게 나타날 수 있다는 점을 감안할 때 일반적으로 다음 네 가지 상황을 상정할 수 있다. 참긍정(True Positive: TP)과 참부정(True Negative: TN)은 실제와 예측 결과가 부합하는 두 가지 경우로서 상장폐지가 될 것으로 예측된 기업이 실제로 상장이 폐지되었거나 상장폐지가 되지 않을 것으로 예측된 기업이 실제로 상장을 유지하는 경우이다. 반대로 거짓긍정(False Positive: FP)과 거짓부정(False Negative: FN)은 실제와 예측 결과가 상충하는 경우를 말한다. 이처럼 머신러닝 기반 분류분석(classification analysis)에서 발생가능한 네 가지 결과를 <표 IV-1>과 같은 혼동행렬(confusion matrix)로 흔히 묘사할 수 있다.

<표 IV-1> 혼동행렬(confusion matrix)

		예측 결과	
		상장폐지(PP)	유지(PN)
실제	상장폐지(P)	True Positive(TP)	False Negative(FN)
	유지(N)	False Positive(FP)	True Negative(TN)

주: P와 N은 각각 Positive와 Negative를 의미하며 실제 상장폐지 여부를 나타낸다. PP와 PN은 각각 Predicted Positive와 Predicted Negative를 의미하며 모형이 예측한 확률값이 0.5 이상이면 PP, 반대의 경우 PN을 부여한다.

혼동행렬의 네 가지 요소에 해당하는 표본의 분포를 조합하여 모형의 정확도를 평가하는 지표를 다양한 방식으로 계산할 수 있다. 첫 번째로 ACC(accuracy)는 모형의 평가에 사용된 전체 표본 중 실제와 예측이 부합하는 표본의 비율을 의미하며 계산식은 다음과 같다.

$$ACC = \frac{TP + TN}{P + N}$$

이는 정확도를 판단하는 매우 직관적인 지표이나 앞 절에서 설명하였듯이 전체 표본 중 상장폐지된 경우의 비율이 매우 낮은 특성으로 인하여 상장이 유지된 경우에 대하여 정확하게 예측한 경우(즉, true negative)가 많이 나타날 수 있다는 점에서 실질적인 정확도보다 과대추정될 우려가 있다(Guo et al., 2008). 이와 같은 문제점을 보완한 값이 BA(Balanced Accuracy)로서 상장폐지와 유지 각각의 경우에 대하여 예측이 맞은 표본의 비중을 계산하고 이에 대한 산술평균값으로 아래의 식과 같이 정의한다.

$$BA = \frac{1}{2} \times \left(\frac{TP}{P} + \frac{TN}{N} \right) = \frac{1}{2} \times (TPR + TNR)$$

이때, 상장폐지된 표본 중에서 예측이 맞은 경우의 비중을 TPR(True Positive Rate), 유지된 표본 중에서 예측이 맞은 경우의 비중을 TNR(True Negative Rate)이라고 한다. TPR은 recall, sensitivity와 같으며 TNR은 흔히 specificity라고도 불린다.

부실기업 여부 예측과 같은 분류분석 모형의 정확도를 다각도로 측정하기 위하여 정밀도(precision)와 재현율(recall) 지표를 앞서 소개한 ACC, BA와 동시에 고려할 수 있다.

우선, precision은 상장폐지로 예측된 모든 표본 중 실제로 상장폐지가 있었던 표본의 비율을, recall은 전체 상장폐지된 표본 중에서 실제로 그와 같이 예측된 표본의 비율을 의미하며 다음과 같이 계산된다.

$$Precision = \frac{TP}{TP+FP}, Recall = Sensitivity = \frac{TP}{TP+FN}$$

두 가지 지표는 예측의 효율성을 서로 다른 관점에서 평가하는 것으로 볼 수 있다. Precision의 경우 모형의 예측값에 기반하여 부실기업으로 판정된 경우 중 실제로 부실기업인 표본의 비율을 의미한다. 예를 들어, 부실기업을 평가하고 매도 여부를 정하는 투자자의 입장에서 precision 값이 클수록 모형의 예측에 대한 신뢰도가 높을 것이다. 한편 recall은 전체 부실기업 중에서 모형의 예측 결과와 부합하는 경우의 비중을 의미한다. 이와 같은 지표는 시장 감독기관의 관점에서 모형의 분석 결과가 효율적으로 부실기업을 선별할 수 있는지에 대한 지표가 될 수 있다.

Precision과 recall이 정확도의 서로 다른 면을 상징하기 때문에 이를 종합적으로 판단하기 위하여 다음과 같은 F1 지표를 계산할 수 있다.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1 지표는 precision과 recall 값의 조화평균으로 정의되어 0에서 1 사이의 값을 가진다. F1이 큰 값을 가질수록 모형이 실제 부실기업을 식별하는데 높은 정확도를 나타내고 있는 것으로 해석할 수 있다.

모형의 학습 효율성을 평가하기 위하여 정확도와 더불어 모형의 적합도(fitness)와 관련된 다음 두 가지 지표를 추가로 분석하였다.

$$Brier = \frac{1}{n} \sum_{i=1}^n (y_i - \Pr(y_i = 1))^2$$

$$Log-loss = -\sum_{i=1}^n (y_i \log \Pr(y_i = 1) + (1 - y_i) \log (1 - \Pr(y_i = 1)))$$

여기서 y_i 는 표본 i 의 상장폐지 여부를 나타내는 더미변수로 $y_i = 1$ 일 경우 기업이 당해 결산일로부터 1년 이내에 거래정지 이후 상장폐지된 경우를 의미하며 $\Pr(y_i = 1)$ 은 상장폐지될 확률을 모형이 예측한 값을 나타낸다. 첫 번째로 Brier 지표는 실제와 예측 확률 사이 편차의 제곱합을 계산한 것으로 평균제곱오차(mean squared error)와 같은 의미이다. 한편 log-loss는 실제 상장폐지 여부와 예측 확률을 이용하여 베르누이 분포(Bernoulli distribution)에 근거한 가능도(likelihood function)를 계산한 것이다. 두 지표 모두 작은 값을 가질수록 모형의 적합도가 높음을 의미한다.

나. 학습 성과의 비교

모형의 학습성과를 보다 세부적으로 평가하기 위하여 <표 IV-2>에서는 앞서 설명한 세 가지 정확도 지표(ACC, BA, F1)와 두 가지 적합도 지표(Brier, Log-loss)를 개별재무제표 학습 모형을 바탕으로 분석하였다. 머신러닝 모형의 학습 성과를 평가하기 위한 기준으로 동일한 재무변수를 사용한 로짓(logit) 회귀분석 모형의 정확도 및 적합도 지표를 함께 나타내었다.

<표 IV-2> 개별재무제표 학습 모형의 정확도 및 적합도 평가

		재무변수만 학습		재무변수 및 주석공시 학습
		로짓 모형	머신러닝 모형	머신러닝 모형
예측 정확도	ACC	0.9966	0.9961	0.9961
	BA	0.5289	0.5874	0.5580
	F1	0.0833	0.1935	0.1379
모형 적합도	Brier	0.0026	0.0030	0.0029
	Log-loss	0.0101	0.0116	0.0114

주: 2016~2019 기업 표본을 바탕으로 평가한 결과를 나타낸다. 모형의 정확도 지표로 사용된 ACC, BA, F1과 적합도 지표 Brier, Log-loss의 정의는 본문을 참조하기 바란다.

정확도 지표를 우선적으로 볼 때, 세 가지 모형에서 모두 ACC가 매우 높게 나타나는 반면 BA는 상대적으로 낮게 평가되는 것을 확인할 수 있다. 이는 학습에 사용된 데이터에서 상장이 유지된 사례가 불균형하게 많이 분포하기 때문에 이를 정확하게 예측한

경우(TN)가 많아 정확도(ACC)를 과대추정한 것으로 볼 수 있다. 더불어 앞서 설명하였듯이 부실기업이 매우 적게 분포하는 표본의 불균형성으로 인하여 일반적인 정확도 지표는 모형 사이에 유의미한 차이를 보이지 않아 학습 효율성을 평가하는 데 적합하지 않은 지표임을 확인할 수 있다.

한편, BA와 F1을 기준으로 평가할 경우 재무제표를 학습한 머신러닝 모형이 로짓 모형보다 높은 정확도를 보이는 것을 확인할 수 있다. 정량적 정보만을 학습한 경우, 머신러닝 모형은 로짓 모형 대비 BA 지표가 약 11%(=0.5874/0.5289-1) 상승하였고 F1 지표는 약 2.3배(=0.1935/0.0833) 증가하였음을 <표 IV-2>에서 확인할 수 있다. 이는 고차원 변수를 동시에 학습하여 종합적인 판단을 요구하는 문제에서 로짓 모형 대비 머신러닝 모형의 접근이 신뢰도가 높은 예측값을 제공할 수 있다는 점을 실증적으로 보여주는 결과라고 하겠다.

이와 같이 재무변수만을 학습한 모형이 로짓 모형 대비 정확도가 가시적으로 증가한 반면 주식 공시에 포함된 정성적인 정보까지 포괄적으로 학습한 멀티모달 모형의 경우 재무변수만 학습한 모형 대비 오히려 정확도가 다소 감소하는 양상을 보인다. <표 IV-2>에서 볼 수 있듯이 재무 공시를 포괄적으로 학습한 모형은 정량적 정보만을 학습한 모형 대비 BA가 약 5%(=0.5580/0.5874-1) 감소했고 F1 지표는 약 29%(=0.1379/0.1935-1) 감소한 것으로 나타났다. 다만, 이 경우에도 로짓 모형보다 높은 BA와 F1 값을 보여 여전히 머신러닝 모형의 학습 효율이 우수하다고 평가할 수 있으나, 동시에 정성적 공시 정보가 학습 효율성을 제고하는 데 기여하는 바가 제한적임을 시사하고 있다.

모형의 적합도 지표는 모든 경우에 있어서 머신러닝 모형에서 다소 증가함을 확인할 수 있다. 재무변수만을 학습하는 경우 로짓 모형 대비 머신러닝 모형의 Brier 및 log-loss 지표가 모두 증가하여 적합도가 다소 감소하지만 이는 정성적인 공시 정보를 추가적으로 학습하는 경우 다소 개선되는 것으로 보인다. 따라서 머신러닝 모형의 복잡성에 비추어 볼 때, 다차원적인 학습 정보를 추가적으로 활용함으로써 적합도를 개선할 수 있지만 그와 같은 노력이 반드시 예측 정확도를 높이는 방향으로 나타나지는 않을 수 있음을 시사한다.

<표 IV-3> 연결재무제표 학습 모형의 정확도 및 적합도 평가

		재무변수만 학습		재무변수 및 주석공시 학습
		로짓 모형	머신러닝 모형	머신러닝 모형
예측 정확도	ACC	0.9964	0.9966	0.9960
	BA	0.5289	0.5876	0.5873
	F1	0.0800	0.2143	0.1875
모형 적합도	Brier	0.0028	0.0028	0.0031
	Log-loss	0.0109	0.0104	0.0120

주: 2016~2019 기업 표본을 바탕으로 평가한 결과를 나타낸다. 모형의 정확도 지표로 사용된 ACC, BA, F1과 적합도 지표 Brier, Log-loss의 정의는 본문을 참조하기 바란다.

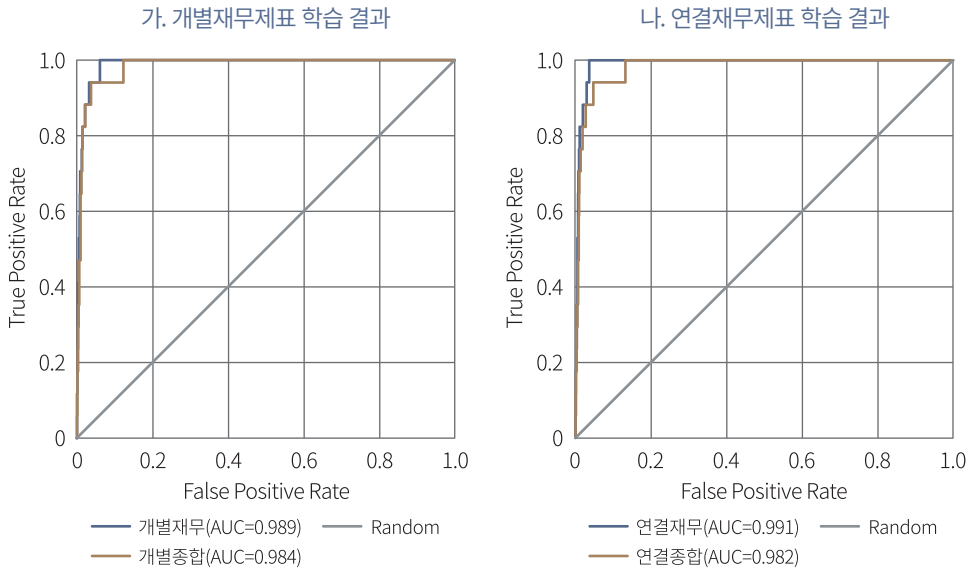
이어서 <표 IV-3>에서는 연결재무제표 학습 모형을 앞서 소개한 것과 동일한 정확도 및 적합도 지표를 활용하여 분석하였다. 개별재무제표 모형의 경우와 같이 연결재무제표 모형에서도 단순한 정확도 지표(ACC)는 모형 사이의 우열을 판별하기 어려우므로 BA와 F1 지표가 비교분석에 용이한 것으로 보인다. 연결재무제표 학습 모형의 경우에서도 머신러닝 모형의 예측 정확도가 로짓 모형 대비 높은 것으로 나타나고 있는 것을 확인할 수 있다. <표 IV-3>의 결과를 볼 때, BA 지표를 기준으로 볼 경우 머신러닝 모형을 사용하여 로짓 모형 대비 약 11%(=0.5876/0.5289-1)의 정확도 상승을 기대할 수 있는 것으로 나타났다. 특히, 정성적 공시자료를 함께 학습한 모형에서도 앞서 <표 IV-2>에서 보인 결과와 다르게 BA 지표가 크게 감소하지 않아 개별재무제표 모형과 차이를 보이고 있다. 이는 연결재무제표 분석에서 주석 정보가 학습 효율을 높이는 데 있어서 개별재무제표보다 상대적으로 더 효과적임을 시사하고 있다.

다. 주석 공시의 유용성 평가: ROC 곡선과 PR 곡선 분석

모형의 학습 효율성을 종합적으로 판단하기 위하여 ROC(Receiver Operating Characteristic) 곡선과 PR(Precision-Recall) 곡선을 분석하였다. ROC 곡선은 분류분석 모형을 평가하는 데 흔히 사용되는 기법으로서 예측 확률에 대한 기준값(threshold)에 따라서 모형의 민감도(sensitivity)와 오류 확률을 비교할 수 있으며 특히 분류 분석(classification analysis) 문제에 있어서 통상적인 정확도 비율보다 직관적으로 선호되는 장점

이 있다(Bradley, 1997; Fawcett, 2006). <그림 IV-1>에서 볼 수 있듯이 ROC곡선의 가로축은 FPR(False Positive Rate)을 의미하는데 이는 $1-TNR(=specificity)$ 과 같은 값으로서 1종 오류-상장이 유지되었음에도 상장폐지 확률이 높은 것으로 예측되는 경우-의 비중을 의미한다. 한편, 세로축은 TPR(True Positive Rate)을 나타내며 이는 실제 상장폐지된 기업 중 모형이 참으로 예측한 경우의 비중을 의미한다. 일반적으로 예측 모형은 1종 오류를 점진적으로 허용하여(false positive rate의 증가) 예측 확률을 높일 수 있으므로(TPR의 증가) ROC 곡선은 단조증가함수의 형태로 나타난다.

<그림 IV-1> ROC 곡선



주: 파란색 선은 정량적 변수만을 학습한 경우, 갈색 선은 정량적 변수와 정성적 지표를 동시에 학습한 경우에 대한 ROC 곡선을 나타낸다. 회색 선은 False Positive Rate과 True Positive Rate이 같은 값을 가지는 점을 표시하였으며 이는 무작위로 예측하였을 경우에 해당한다. ROC 곡선이 회색선의 좌상단에 있을 경우 모형이 무작위 예측보다 유의미한 성과를 보이고 있음을 의미한다.

ROC 곡선의 형태를 볼 때, 재무제표를 학습한 모형이 상장폐지를 매우 높은 정확도로 조기에 예측하고 있음을 알 수 있다. ROC 곡선 하부의 면적(Area Under the Curve: AUC)을 기준으로 평가하면 개별재무제표를 학습한 모형(<그림 IV-1>의 가.)과 연결재무제표를 학습한 모형(<그림 IV-1>의 나.) 모두 높은 정확도를 보이는 것을 알 수 있다. 한편, 재무변수(정량적 지표)만을 학습한 모형과 정성적 공시까지 종합적으로 고려한 모형의

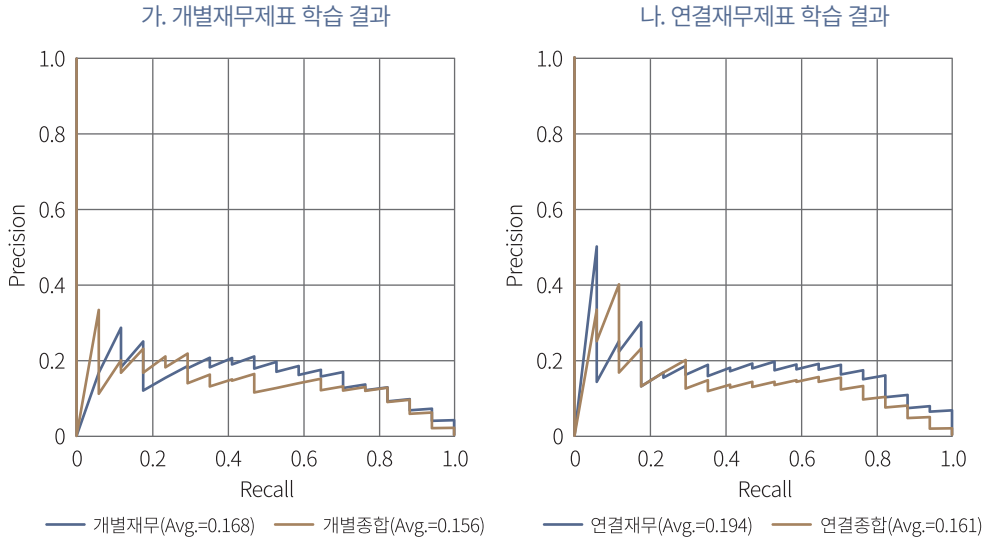
정확도를 비교할 때, 개별재무제표와 연결재무제표 모형 모두 전자가 상대적으로 높은 정확도로 부실기업을 조기에 예측하는 것으로 나타났다. 특히, 정확도의 격차는 연결재무제표 모형($0.009=0.991-0.982$)이 개별재무제표 모형($0.005=0.989-0.984$)보다 큰 것으로 나타나 연결재무제표를 바탕으로 부실기업을 평가하는 데 있어서 정성적 공시의 유용성이 정량적 지표와 비교하여 상대적으로 낮음을 시사하고 있다.

상장폐지 표본의 특수성을 감안하여 추가적으로 PR(Precision-Recall) 곡선을 분석하여 모형의 학습 효율성을 평가하였다. 일반적으로 표본 분포의 불균형이 심할 경우 ROC 곡선은 모형의 예측력을 과대평가할 위험이 있는 것으로 알려져 있다(Davis & Goadrich, 2006; Cook & Ramadas, 2020). 따라서 앞서 분석한 ROC 곡선과 더불어서 PR 곡선을 기준으로 모형의 학습 성과를 재평가할 필요가 있다.

PR 곡선의 가로축은 recall(또는 sensitivity, TPR), 세로축은 precision을 나타낸다(<그림 IV-2> 참조). 상장폐지의 예측 기준점을 높게 설정하면(예측 확률이 1에 가까운 경우만을 상장폐지될 것으로 평가하면) 극소수의 표본만 상장폐지될 것으로 예측되어 낮은 recall 값에서 비교적 높은 precision을 얻을 수 있다.²⁰ 반대로 예측 기준점을 매우 낮게 설정할 경우에는 실제 상장폐지된 표본의 상당수가 맞게 예측되어 recall이 상승하지만 동시에 상장폐지될 것으로 예측된 표본 중 실제로 상장폐지된 경우의 비중이 줄어 precision은 감소하게 된다. 따라서 recall이 증가함에 따라 precision의 감소폭이 작은 경우 모형의 정확도가 높은 것으로 해석할 수 있다.

20 극단적으로 기준점을 1로 정한다면 상장폐지될 것으로 예측된 표본의 수는 0일 것이고 이 경우 recall(전체 상장폐지 기업 중 실제로 그렇게 예측된 경우)은 0, precision(상장폐지될 것으로 예측된 기업 중 실제로 상장폐지된 경우)은 무한대의 값을 가진다. 이는 <그림 IV-2>에서 가장 좌측이 1에서 시작되는 것으로 표시하였다.

<그림 IV-2> PR 곡선



주: 파란색 선은 정량적 변수만을 학습한 경우, 갈색 선은 정량적 변수와 정성적 지표를 동시에 학습한 경우에 대한 PR 곡선을 나타낸다. Recall 값이 1보다 작은 영역에서 PR 곡선이 평균 recall보다 높은 지점에 위치할 때, 모형이 무작위 예측보다 유의미한 성과를 보임을 의미한다.

모든 recall 값에 대응하는 precision 값을 해석하였을 때, 재무제표를 학습한 모형이 무작위로 부실기업을 예측하는 경우보다 유의미하게 높은 분별력을 나타내고 있음을 알 수 있다. 예를 들어, 정량적 지표만 학습한 개별재무제표 모형은 평균 precision이 0.168로 나타나는데 이는 표본기간(2016~2019)내 거래정지 후 상장폐지 건수가 22건(전체 표본 대비 약 0.2%)에 불과하다는 점에서 매우 효율적인 예측 결과이다. 이 수치는 연결재무제표 학습 모형에서 더욱 높다(0.194)는 점에서 연결재무제표가 부실기업을 선별하는데 있어서 의미 있는 정보를 제공할 수 있음을 시사한다. 반면, 주식 공시 정보를 포함하여 학습시킨 모형의 경우 개별재무제표와 연결재무제표에서 일관적으로 precision이 다소 감소하는 양상을 보이고 있다. 이는 앞서 ROC 곡선의 해석과 함께 정성적 공시에 포함된 정보가 부실기업을 식별하는데 추가적으로 유용한 정보를 제공하는지 여부에 대한 재평가가 필요하다는 점을 의미한다.

2. 부실 평가 기업의 특성 분석

가. 부실 징후 예측값의 분포

학습이 완료된 신경망 모형을 바탕으로 상장기업의 부실 징후를 예측할 수 있다. 이때, 예측된 값은 0에서 1 사이의 확률값으로 모형의 구조적 특징에 근거하여 해석하면 해당 사업연도의 결산일로부터 1년 이내에 향후 상장폐지를 전제로 한 거래정지 사건이 발생할 확률로 볼 수 있다. 실제 표본에서 거래정지 및 상장폐지가 일어나는 빈도가 전체 상장기업 수 대비 극소수인 점을 감안할 때, 예측값 또한 대부분의 경우에 극히 작은 값을 가지고 극소수의 표본에서만 큰 값을 가지는 형태로 분포할 것을 예상할 수 있다.

<표 IV-4> 부실 위험도 예측값 분포

		연도	표본수	평균	표준편차	중위값	상위 5%	상위 1%	최대값
개별 재무 제표	재무	2016	1,498	0.0030	0.0237	0.0004	0.0052	0.0609	0.7345
		2017	1,574	0.0027	0.0292	0.0004	0.0048	0.0356	0.9855
		2018	1,653	0.0041	0.0339	0.0004	0.0058	0.0683	0.7638
		2019	1,729	0.0079	0.0567	0.0005	0.0081	0.2178	0.9031
	종합	2016	1,498	0.0017	0.0159	0.0001	0.0035	0.0344	0.4919
		2017	1,574	0.0013	0.0081	0.0002	0.0035	0.0265	0.2036
		2018	1,653	0.0027	0.0294	0.0001	0.0025	0.0373	0.8028
		2019	1,729	0.0077	0.0572	0.0004	0.0081	0.2219	0.9213
연결 재무 제표	재무	2016	1,498	0.0019	0.0226	0.0001	0.0029	0.0303	0.7974
		2017	1,574	0.0025	0.0183	0.0004	0.0056	0.0375	0.5831
		2018	1,653	0.0032	0.0316	0.0002	0.0037	0.0420	0.7856
		2019	1,729	0.0073	0.0533	0.0004	0.0079	0.2025	0.8037
	종합	2016	1,498	0.0027	0.0259	0.0001	0.0046	0.0508	0.7796
		2017	1,574	0.0012	0.0074	0.0001	0.0033	0.0232	0.1856
		2018	1,653	0.0035	0.0372	0.0001	0.0030	0.0614	0.8734
		2019	1,729	0.0082	0.0617	0.0002	0.0069	0.2763	0.9121

주: 학습된 모형을 바탕으로 연도별로 표본에 속한 모든 재무제표에 대하여 부실 확률을 예측하여 그 분포를 나타내었다.

<표 IV-4>는 평가 대상 표본(2016~2019년)에 포함된 사업보고서 전체를 대상으로 모형이 예측한 부실 징후의 예측치 분포를 나타낸 것이다. 앞서 설명한 바와 같이 예측값의 평균과 중위값이 매우 낮으며 이는 대부분의 경우 상장폐지로 이어질 정도의 큰 부실 징후는 나타나지 않았음을 시사한다. 반면 표준편차는 평균보다 약 10배 정도 크게 추정되는데, 이는 극소수의 사업보고서에서 예측값이 상당히 크게 추정된 것에 기인한다. 예를 들어, 2019년 표본 기업을 대상으로 정량적 지표만을 활용하여 개별재무제표를 평가한 결과 부실 확률 기준 상위 1%의 기업의 약 21.8%의 확률로 상장폐지의 가능성이 예측되었으며 해당 연도의 표본에서 부실 확률 예측치의 최대값은 90.3%로 추정되었다. 이는 표본 전체의 평균값인 0.79%와 비교하여 현저히 높은 값으로 표준편차를 과대추정하는 통계적 요인이 되고 있다.

나. 부실기업의 특성 비교

모형이 비교적 정확하게 기업의 부실 징후를 예측할 수 있다고 한다면 예측된 부실 확률값을 기준으로 부실 가능성이 높은 기업과 그렇지 않은 기업을 구분할 수 있을 것이다. 구체적으로 모형의 부실 확률 예측치가 50%를 넘는 경우를 상장폐지 예상 기업(PP), 반대의 경우를 유지 예상 기업(PN)으로 구분하였을 때, 각각에 속한 기업들의 재무적 특성을 비교해 볼 수 있다.

<표 IV-5> t-test 결과

		PN	PP	t-stat	p-value
부실 확률 예측값		0.0016	0.6937	-135.3100	0.0000
재무적 특성	InASSET	19.2324	18.2897	2.0669	0.0388
	DtoA	0.3877	0.6945	-4.6691	0.0000
	BM3	1.0773	1.3415	-0.9029	0.3666
	SALEG	0.1353	-0.1703	0.2292	0.8188
	OIG	-0.6976	-0.2933	-0.0727	0.9421
	ROE	-0.0242	-0.7015	6.6287	0.0000
	ROEStd	0.1474	0.5725	-3.8367	0.0001
	OITASD	0.0434	0.1171	-5.1499	0.0000
	InvTURN	207.9033	9.6385	0.0953	0.9241
	RecTURN	9.3183	5.6944	0.1994	0.8420
	LossOI3yr	0.0678	0.6667	-7.1188	0.0000
	absPMDA	0.0977	0.2401	-4.2633	0.0000
공시 정보 특성	fs_length	11.5514	11.2245	2.1073	0.0351
	word_reps	28.1418	20.8001	2.7041	0.0069
	Δ unique_words	-0.0191	-0.0175	-0.0206	0.9835
	cosine_1	0.8243	0.8480	-0.5796	0.5622
	cosine_p	0.5385	0.5209	1.2987	0.1941

주: 부실 확률 예측값이 0.5를 초과하는 표본을 PP, 그렇지 않은 경우를 PN으로 분류하여 각 변수들의 표본평균의 차이를 t-test로 비교하였으며 5% 유의수준에서 통계적으로 유의미한 차이를 보이는 변수는 굵은 글씨로 표시하였다.

<표 IV-5>에서는 개별재무제표를 종합적으로 학습한 모형을 바탕으로 평가 표본내 기업의 연도별 부실 확률을 예측한 후, 50%를 기준으로 부실 예상 기업(PP)과 그렇지 않은 기업(PN)을 나누어 재무 및 공시정보 특성의 평균값을 비교하였다.²¹ 재무적인 특성을 살펴볼 때, PP에 속한 기업들이 평균적으로 규모(InASSET)가 작고, 차입금의존도(DtoA)가 높으며, 자기자본수익률(ROE)은 낮은 반면 수익률변동성(ROEStd)은 높은 것으로 나타났다. 더불어 지난 3년간 연속으로 영업손실이 발생(LossOI3yr)하였을 가능성이 10배 가량 높은 것으로 나타났는데 이는 앞서 수익률이 낮고 수익률의 변동성이 높다는 실증분석

21 연결재무제표를 학습한 모형에 기반하여 예측한 부실 예상 기업과 그렇지 않은 기업의 특성 또한 비교하였으며 그 결과는 부록에 포함하였다.

결과와 부합한다고 볼 수 있다. 더불어 재량적 발생액의 절댓값(absPMDA)도 부실 징후가 있는 기업의 평균이 더 높게 나타났는데 이는 모형이 이익 조정 수준이 높은 기업을 부정적으로 평가하고 있음을 시사한다.

공시 정보의 특성과 관련해서는 부실 확률이 높은 기업들이 대체로 재무제표 공시의 길이(fs_length)가 짧고 단어의 반복적 사용(word_reps)이 적은 것으로 나타났다. 다만, 해당 변수들은 공시 정보의 특성을 요약한 지표로서 재무제표의 정량적인 요소의 차이와 밀접하게 연관되어 있을 것으로 예상되기 때문에 이와 같은 통계적인 차이를 해석하는 데 어려움이 있다. 따라서 공시 자료의 정성적인 특성이 기업의 부실 징후를 예측하는데 있어서 어떠한 방식으로 기여할 수 있는지에 대하여 보다 정교한 실증 분석을 수행할 필요가 있다.

3. 공시 유용성 기여 요인 평가

부실기업을 예측하는 문제에 있어 공시정보의 실질적인 유용성이 높다면 해당 기업, 연도의 재무제표를 학습시킨 모형이 판별한 부실 확률과 실제 거래정지 및 상장폐지 사건의 발생 여부 사이의 괴리가 적을 것이라고 예상할 수 있다. 따라서 본 절에서는 개별 사업보고서별로 예측 결과와 실제 부실기업 사이의 차이를 계산하고 이에 기여하는 요소들을 회귀분석을 통해 평가하였다.

모형의 평가에 사용된 2016~2019년 동안의 상장기업 재무제표 표본에 대하여 정보의 유용성 정도(*INFORM*)를 다음과 같은 식으로 계산하였다.

$$INFORM_i = y_i \log \Pr(y_i = 1) + (1 - y_i) \log(1 - \Pr(y_i = 1))$$

INFORM_i 값을 정의할 때 (log-loss와는 반대로) 음의 값을 갖도록 하여 값이 클수록 (0에 가까울수록) 기업의 실제 부실 여부와 학습된 모형에 기반한 예측이 일치하며 값이 작을수록 반대가 되도록 하였다. 이는 다시 말하면, 실제 거래정지가 된 경우($y_i = 1$) 예측 확률($\Pr(y_i = 1)$)이 1에 가까울수록, 반대의 경우 0에 가까울수록 괴리도는 큰 값을 가지게 됨을 의미한다. 따라서 *INFORM_i* 값이 클수록 공시 정보를 학습한 모형이 기업의 실제 부실 여부를 일관되게 판별할 수 있는 것으로 해석할 수 있다.

공시정보의 유용성을 종속변수로 하여 아래와 같은 회귀분석 모형을 가정하였다.

$$INFORM_i = \alpha + \beta X_i + Controls_i + FE + \epsilon_i$$

종속변수인 $INFORM_i$ 을 설명하는 공시자료의 정성적인 요소로는 다섯 가지 변수를 평가하였다. 우선, 공시의 정보량 측면을 반영하기 위하여 중복 사용된 단어를 제외한 단어 수의 전년도 공시자료 대비 변화량(Δ unique_words)을 고려하였다. 더불어 공시의 양태 측면에서 재무제표 공시의 길이(fs_length), 단어의 평균 반복 횟수(word_reps), 직전 회계연도 대비 코사인 유사도(cosine_1), 동일업종 평균 대비 코사인 유사도(cosine_p)를 평가하였다.

공시 정보의 유용성은 기업의 재무적 특성에 따라서 차이를 보일 수 있다. 이를 통제하기 위하여 해당 기업-연도의 재무적인 특성을 반영하는 변수를 선정하여 통제변수(controls)로 사용하였다. 통제변수로 사용된 지표는 구체적으로 기업규모(lnASSET), 차입금의존도(DtoA), 장부가치 대비 시장가치 비율(BM3), 총매출성장률(SALEG), 영업이익성장률(OIG), 순이익변동성(ROEStd), 영업이익변동성(OITASD), 재고자산회전률(InvTURN), 매출채권회전률(RecTURN), 3년연속 영업현금흐름 손실 여부(LossOI3yr), 재량적 발생액의 절댓값(absPMDA)이다.

<표 IV-6> 회귀분석 결과

	정보량 측면		공시 양태 측면		종합	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ unique_words	0.1569*** (0.0333)	0.1444*** (0.0341)			0.2386** (0.0921)	0.2032* (0.0940)
fs_length			-0.2994** (0.0948)	-0.3381*** (0.1024)	-0.2572** (0.0953)	-0.3088** (0.1038)
word_reps			0.0142*** (0.0040)	0.0154*** (0.0042)	0.0125** (0.0040)	0.0144*** (0.0043)
cosine_1			0.0903 (0.0577)	0.1255* (0.0632)	0.1224 (0.0813)	0.2019* (0.0936)
cosine_p			1.3152* (0.6120)	1.6634* (0.6639)	1.4011* (0.6112)	1.7273** (0.6639)
표본수	4966	4966	4995	4995	4873	4873
R^2	0.143	0.151	0.147	0.156	0.150	0.160
$\overline{R^2}$	0.141	0.147	0.145	0.152	0.147	0.155
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year-Ind FE	No	Yes	No	Yes	No	Yes
Cluster SE	firm	firm	firm	firm	firm	firm

중: 괄호 안의 수치는 표준오차를 의미하며 추정된 회귀계수가 5%, 1%, 0.1% 유의수준에서 통계적으로 유의한 값을 가지는 경우 *, **, ***을 추정값 우측에 표시하였다. 모든 결과는 동일한 통제변수를 포함하고 있으며 연도·산업군별 고정효과(Fixed Effect: FE)는 (2), (4), (6)에만 포함하였다. 표준오차는 기업 수준에서 군집표준오차(clustered standard error)로 산출하였다.

<표 IV-6>의 결과를 통하여 개별재무제표 평가 모형 중 재무변수와 주식 공시를 종합적으로 학습한 결과를 바탕으로 재무제표의 유용성 지표를 계산한 후, 공시 문서의 특성 중 유용성에 유의미한 영향을 미치는 요소를 확인할 수 있다.²² (1)과 (2)를 볼 때, 정보의 양적인 측면에서 전년도 재무제표 공시와 비교하여 (중복을 제외하고) 새롭게 사용된 단어가 많은 경우 해당 공시가 유용성이 높은(학습을 통해 예측한 결과와 실제 간의 괴리가 적은) 것으로 나타났다. 이는 전년도와 비교하여 새롭게 제공된 정보가 많은 공시일수록 부실기업(또는 부실하지 않은 기업)을 평가하는 데 있어 공시의 유용성이 증가한 것으로 볼 수 있다.

22 연결재무제표를 학습한 모형을 바탕으로 동일한 분석을 수행한 결과는 부록에 포함하였다.

한편 공시의 양태 측면에서는 평가에 사용된 지표별로 유용성 측면에서 다른 영향을 보이는 것을 알 수 있다. 특히, 재무제표 공시의 절대적인 분량이 증가하는 것은 유용성 측면에서 오히려 부정적인 영향을 주고 있음을 결과표 (3)과 (4)에서 재무공시의 길이(fs_length)의 회귀계수가 유의하게 음수로 추정된 사실에서 확인할 수 있다. 이는 앞서 설명한 바와 같이 정보량의 증가(Δ unique_words)의 회귀계수가 양인 점과 대비되어 새롭게 추가된 정보가 아닌 단순한 양적 공시의 증가는 오히려 유용성에 악영향을 줄 수 있다는 점을 시사하고 있다. 반면 단어의 반복적 사용(word_reps)은 유용성을 증가시키는 것으로 나타났다. 이는 주석 내 반복적 표현이 인간의 관점에서는 주의력의 제약을 심화하는 요인으로 작용할 가능성이 클지라도(이재경 · 한봉희, 2019), 기계의 관점에서는 중요 정보의 명확한 강조로 인지될 수 있음을 의미한다.

공시 정보가 전년 대비(cosine_1) 그리고 동일업종 대비(cosine_p) 유사성이 높은 경우는 유용성에 대체로 긍정적인 영향을 미치는 것으로 나타났다. 이 중, 전년도 대비 유사성의 영향은 통계적으로 유의하게 나타나지는 않았는데 앞서 설명한 바와 같이 전년 대비 새롭게 사용된 단어가 많을수록 유용성이 증가하는 결과와 연관 지어 볼 수 있다. 동일 기업의 재무제표는 시계열적으로 유사도가 높으며 이는 대체로 정보의 습득을 쉽게 하지만 전년 대비 추가적인 정보가 부족할 경우 그 유용성은 제한적일 것이다. 한편, 동일 업종의 평균적인 재무제표 양식과 유사한 형태의 공시는 정보의 전달력을 증가시키는 것으로 나타났다.

이와 같은 결과는 모든 변수를 동시에 고려하였을 경우인 (5), (6)에서도 유사하게 관측되나 각 변수의 유의성에서는 다소 차이를 보이고 있다. 우선, 동일 기업의 전년도 대비 재무제표의 정보량이 유용성에 미치는 영향은 그 통계적 유의성이 다소 감소하는 경향을 보인다. 반면, 중요한 정보의 반복과 동일 업종 내 비교가능성은 통계적으로 강한 유의성을 보인다는 점에서 재무제표 공시의 정보전달력을 제고할 수 있는 방향에 대한 시사점을 제시할 수 있을 것이다.

V. 결론 및 시사점



V. 결론 및 시사점

본 연구에서는 머신러닝 기반 예측 모형을 활용하여 재무제표 본문 및 주식 정보가 상장기업의 부실 징후를 조기에 탐지하는 데 유용한지를 평가하였다. 특히, 멀티모달 신경망 모형을 적용하여 재무제표 본문 정보와 주식 정보를 통합 학습함으로써, 정량적·정성적 정보를 포괄적으로 예측에 반영하는 접근법을 제시하였다. 이는 수익성, 건전성 등 전통적인 재무비율 지표와 더불어 주식의 방대한 비정형 정보를 활용하여 부실 징후를 예측하고자 한 것으로, 기존 부실 예측 모형과 차별화되는 의의를 지닌다.

2005년부터 2019년까지 유가증권 및 코스닥시장에 상장된 12월 말 결산 비금융업종 16,815 기업-연도를 대상으로 분석을 수행한 결과, 본 연구에서 구축한 머신러닝 모형은 상장폐지 위험을 우수한 정확도로 예측하였다. 특히, 상장폐지로 인한 거래정지 이전 단계에서 가용한 정보를 바탕으로 부실 징후를 조기 탐지하여, 거래가 가능한 시점에 부실 신호를 전달할 수 있는 활용 가능성을 확인하였다. 다만, 높은 예측력은 주로 상장유지 기업의 정확한 식별에서 기인한 것으로, 결과 해석과 활용 시 유의할 필요가 있다.

본 연구는 다음과 같은 시사점을 제시한다. 첫째, 재무제표 정보를 활용한 부실 예측 모형의 유용성을 높이기 위해 본문의 정량적 정보와 주식의 정성적 정보 간 유기성을 제고할 필요가 있다. 예측 결과, 정량적 지표만을 학습한 모형에 비해 주식 정보를 포함한 모형의 효율성이 상대적으로 낮은 것으로 나타나 주식 공시의 정보 품질 개선이 필요할 수 있음을 시사한다.

둘째, 주식 정보의 이용자 특성을 고려하는 사고의 틀을 확장할 필요가 있다. 기존 문헌에서는 주식의 과도한 양이 정보이용자의 주의력을 분산시키므로 비교적 간결한 작성을 권고하였다. 그러나 주의력에 제약이 없는 기계의 관점에서 실증 분석을 수행한 결과, 정보의 반복적 제시는 핵심 정보의 강조로 인지될 수 있는 것으로 나타났다. 정보성이 없는 상투적 기술은 지양하되, 중요한 정보는 반복적이고 명확히 강조하는 구조가 유용할 것으로 판단된다.

셋째, 주식 공시는 기간 간 및 유사 기업 간 비교가능성을 고려한 형태로 작성될 필요가 있다. 재무제표 주식 정보는 본문 수치의 맥락을 보완하는 역할을 해야 하며, 동일 기업 내 기간 간 일관성을 유지하고 동종 산업 내 기업 간 유사성을 확보할 때 부실 징후의 예측 오차가 완화되는 것으로 나타났다.

마지막으로, 기계 가독성(machine readability)을 고려한 공시 체계 구축이 필요하다. 자본시장의 발전과 거래 구조의 복잡화에 따라 공시 정보의 양은 필연적으로 증가하고 있으며, 이에 웹 스크래핑 등 정보 처리 알고리즘을 통한 공시 정보 분석 시도가 확대되고 있다(SEC, 2009; Felo et al., 2018; 노성호, 2024). 재무제표 정보가 인간과 기계 모두에게 효과적으로 전달될 수 있도록 XBRL 등 구조화된 공시를 강화하여 접근성과 가독성을 제고할 필요가 있다.

본 연구는 국내 상장기업의 부실 징후 예측에서 재무제표 본문과 주석 정보를 종합적으로 활용한 선도적 연구로서 학술적 기여가 기대된다. 다만, 부실 사건의 희소성으로 인한 표본의 클래스 불균형 문제는 본 연구의 한계로, 예측의 균형성과 재현율을 제고하기 위한 후속 연구가 필요하다. 예를 들어, 방대한 주석 정보 중 주요 부실 사례별로 특화된 학습 모형을 개발할 수 있을 것이다. 이는 결국, 현 주석 정보의 질적 수준을 고려할 때, 회계, 금융, 경제학적 지식을 갖춘 인간 연구자가 기계 모형을 보완함으로써 정보의 유용성을 높이는 협력 과정이 필요함을 시사한다.

Abstract

This study aims to predict corporate failures in the Korean stock market. In particular, it seeks to design an effective early warning model that forecasts delisting prior to trading suspensions. To achieve this, the model incorporates not only quantitative data from the main financial statements, but also unstructured information from footnotes which has significantly increased in volume and importance since the adoption of K-IFRS. A machine learning methodology based on multimodal neural networks was employed to integrate these diverse data sources.

Amid growing economic uncertainties and the ongoing issue of corporate zombieification, the insolvency prediction model proposed in this study offers valuable insights for both investors and financial authorities. By investigating the predictive power of footnotes information in identifying signs of insolvency, this study also offers a unique contribution to the field of academia. It examines whether accounting conservatism, the core accounting principle to disclose bad news in a timely manner, is adequately reflected in footnotes. In addition, based on extensive empirical analysis of current footnote disclosures, this study offers policy recommendations for enhancing the quality and effectiveness of footnote disclosures.

The empirical analysis draws on a dataset comprising 16,815 firm-year observations from non-financial industries with fiscal year-ends in December, listed on the KOSPI and KOSDAQ markets between 2005 and 2019. The model utilizes 43 quantitative indicators, including a firm's marginal factors and external information, 37 major accounts, and all unstructured data contained in financial

statement footnotes. Given the specific characteristics of the Korean market, where delisting due to insolvency is frequently preceded by prolonged trading suspensions, the model integrates solely on information available prior to the suspension.

The results demonstrate that the risk of delisting can be predicted with high accuracy before trading suspension occurs. This result is particularly meaningful as it would enable investors to incorporate insolvency risk at the time of investment decision making and not after. However, this high predictive accuracy is largely attributed to the model's proficiency in correctly identifying firms with a relatively low risk of failure, suggesting the need for further refinement to improve its capacity to predict firms at high risk of insolvency.

Furthermore, the study found that the contribution of footnote information to predicting firms with a high likelihood of distress was limited. This finding calls for a re-evaluation of whether accounting conservatism is sufficiently incorporated in the preparation of footnote disclosures. To improve the usefulness of footnotes in predicting financial distress, it is essential to strengthen the linkage between the main body of financial statements and footnote information, as well as to enhance both the quantity and presentation of footnote disclosures. Footnotes should be concise yet comprehensive, and structured to facilitate comparability across time periods and between peer companies. Enhancing the quality of footnote disclosures, and by extension improving the balance of predictive models, remains a key area for future research.

참고문헌

- 금융감독원, 2010. 12. 21, 『2010 사업연도 회계현안설명회』, 보도자료.
- 금융감독원, 2023. 12. 18, 『2023 정기 신용위험평가 결과 및 향후 계획』, 보도자료.
- 나현중·정태진, 2022, 머신러닝을 활용한 회계부정 탐지에 관한 탐색적 연구, 『회계학 연구』 47(1), 177-205.
- 노성호, 2024, AI의 시대에 기업 재무 공시의 구조화가 가지는 의미, 자본시장연구원 『자본시장포커스』 2024-06.
- 모예린·서윤석, 2019, 주식 내용의 변동과 주식시장: 주식 내용의 변동이 자기자본비용과 주식거래량 및 이익반응계수에 미치는 영향, 『회계학연구』 44(4), 215-249.
- 오세경·최정원·장재원, 2017, 빅데이터를 이용한 딥러닝 기반의 기업 부도예측 연구. KIF working paper 2017(8), 1-113.
- 이상호, 2022, 주식 거래정지 장기화와 투자자 보호 관점의 전환 필요성, 자본시장연구원 『자본시장포커스』 2022-16.
- 이인로·김동철, 2015, 회계정보와 시장정보를 이용한 부도예측모형의 평가 연구, 『재무 연구』 28(4), 626-666.
- 이재경·한봉희, 2019, 재무제표 주식 유용성의 실태조사와 개선방안, 『회계저널』 28(3), 151-188.
- 이준일·현지원·조현권, 2023, 주식 작성에 미치는 감사인의 영향과 신뢰감법 도입 효과: 주식유사도를 이용한 검증, 『회계저널』 32(3), 193-219.
- 현지원·이준일·조현권, 2022, 재무제표 주식은 서로 얼마나 유사한가, 『회계저널』 31(5), 283-304.
- 홍지연, 2023, 국내 기업의 XBRL 재무공시 의무화 확대 의의와 과제, 자본시장연구원 『자본시장포커스』 2023-10.

- Aboody, D., 1996, Recognition versus disclosure in the oil and gas industry, *Journal of Accounting Research* 34, 21-32.
- Acharya, V.V., Crosignani, M., Eisert, T., Eufinger, C., 2024, Zombie credit and (dis-) Inflation: Evidence from Europe, *Journal of Finance* 79(3), 1883-1929.
- Ahmed, A.S., Kilic, E., Lobo, G.J, 2006, Does recognition versus disclosure matter? Evidence from value -relevance of banks' recognized and disclosed derivative financial instruments, *Accounting Review* 81(3), 567-588.
- Allee, K.D., DeAngelis, M.D., Moon Jr, J.R., 2018, Disclosure “scriptability”, *Journal of Accounting Research* 56(2), 363-430.
- Altman, E.I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance* 23(4), 589-609.
- Banerjee, R., Hofmann, B., 2022, Corporate zombies: Anatomy and life cycle, *Economic Policy* 37(112), 757-803.
- Bao, Y., Ke, B., Li, B., Yu, Y.J., Zhang, J., 2020, Detecting accounting fraud in publicly traded US firms using a machine learning approach, *Journal of Accounting Research* 58(1), 199-235.
- Barth, M.E., Clinch, G., Shibano, T., 2003, Market effects of recognition and disclosure, *Journal of Accounting Research* 41(4), 581-609.
- Bauguess, S.W., 2018. 5. 3, The Role of Machine Readability in an AI World.
- Beaver, W.H., 1966, Financial ratios as predictors of failure, *Journal of Accounting Research* 71-111.
- Beneish, M.D., 1999, The detection of earnings manipulation, *Financial Analysts Journal* 55(5), 24-36.
- Bertomeu, J., Cheynel, E., Floyd, E., Pan, W., 2021, Using machine learning to detect misstatements, *Review of Accounting Studies* 26, 468-519.
- Bharath, S.T., Shumway, T., 2008, Forecasting default with the Merton distance to default model, *Review of Financial Studies* 21(3), 1339-1369.

- Bloomfield, R.J., 2012, A pragmatic approach to more efficient corporate disclosure, *Accounting Horizons* 26(2), 357-370.
- Bradley, A.P., 1997, The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition* 30(7), 1145-1159.
- Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008, In search of distress risk, *Journal of Finance* 63(6), 2899-2939.
- Cao, K., You, H., 2024, Fundamental analysis via machine learning, *Financial Analysts Journal* 80(2), 74-98.
- Chen, X., He, W., Tao, L., Yu, J., 2023, Attention and underreaction-related anomalies, *Management Science* 69(1), 636-659.
- Cook, J., Ramadas, V., 2020, When to consult precision-recall curves, *Stata Journal* 20(1), 131-148.
- Cressey, D.R., 1986, Why managers commit fraud, *Australian & New Zealand Journal of Criminology* 19(4), 195-209.
- Davis, J., Goadrich, M., 2006, The relationship between Precision-Recall and ROC curves, *Proceedings of the 23rd international conference on Machine learning*, 233-240.
- Davis-Friday, P.Y., Folami, L.B., Liu, C.S., Mittelstaedt, H.F, 1999, The value relevance of financial statement recognition vs. disclosure: Evidence from SFAS No. 106, *Accounting Review* 74(4), 403-423.
- Dechow, P.M., Ge, W., Larson, C.R., Sloan, R.G, 2011, Predicting material accounting misstatements, *Contemporary Accounting Research* 28(1), 17-82.
- Fawcett, T., 2006, An introduction to ROC analysis, *Pattern Recognition Letters* 27(8), 861-874.

- Felo, A.J., Kim, J.W., Lim, J.H., 2018, Can XBRL detailed tagging of footnotes improve financial analysts' information environment? *International Journal of Accounting Information Systems* 28, 45-58.
- Frederickson, J.R., Hodge, F.D., Pratt, J.H., 2006, The evolution of stock option accounting: Disclosure, voluntary recognition, mandated recognition, and management disavowals, *Accounting Review* 81(5), 1073-1093.
- Guo, X., Yin, Y., Dong, C., Yang, G., Zhou, G., 2008, On the class imbalance problem, In 2008 Fourth International Conference on Natural Computation Vol. 4, 192-201.
- Henderson, E., 2016, Users' Perceptions of Financial Statement Note Disclosure and the Theory of Information Overload, Northcentral University.
- Hirshleifer, D., Teoh, S.H., 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36(1-3), 337-386.
- Hutton, A.P., Marcus, A.J., Tehranian, H., 2009, Opaque financial reports, R2, and crash risk, *Journal of Financial Economics* 94(1), 67-86.
- Iannaconi, T., 2012, Disclosure overload: Gaining control of the process, *Financial Executive* 28(2).
- International Monetary Fund(IMF), 2023, *Regional Economic Outlook for Asia and Pacific*.
- Imhoff Jr, E.A., Lipe Jr, R., Wright Jr, D.W., 1993, The effects of recognition versus disclosure on shareholder risk and executive compensation, *Journal of Accounting, Auditing & Finance* 8(4), 335-368.
- International Accounting Standards Board, 2017, Disclosure Initiative: Principles of Disclosure, IFRS Discussion paper.
- Japkowicz, N., Stephen, S., 2002, The class imbalance problem: A systematic study, *Intelligent data analysis* 6(5), 429-449.

- Kim, J.B., Zhang, L., 2016, Accounting conservatism and stock price crash risk: Firm-level evidence, *Contemporary Accounting Research* 33(1), 412-441.
- Krawczyk, B., 2016, Learning from imbalanced data: open challenges and future directions, *Progress in Artificial Intelligence* 5(4), 221-232.
- Li, W., Shi, Y., Zeng, S., 2024, The hidden cost of clarity: Market and real efficiency in the age of machine reading, SSRN Working paper.
- Libby, R., Nelson, M.W., Hunton, J.E., 2006, Retracted: Recognition v. disclosure, auditor tolerance for misstatement, and the reliability of stock-compensation and lease information, *Journal of Accounting Research* 44(3), 533-560.
- Merton, R.C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29(2), 449-470.
- Ohlson, J.A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18(1), 109-131.
- Perols, J., 2011, Financial statement fraud detection: An analysis of statistical and machine learning algorithms, *Auditing: A Journal of Practice & Theory* 30(2), 19-50.
- Perols, J.L., Bowen, R.M., Zimmermann, C., Samba, B., 2017, Finding needles in a haystack: Using data analytics to improve fraud prediction, *Accounting Review* 92(2), 221-245.
- Shumway, T., 2001, Forecasting bankruptcy more accurately: A simple hazard model, *Journal of Business* 74(1), 101-124.
- U.S. Securities and Exchange Commission(SEC), 2009, Interactive Data to Improve Financial Reporting, Release Nos. 33-9002; 34-59324; 39-2461.

<부록>

1. 학습에 사용된 정량적 지표

본문에서 소개한 머신러닝 모형의 학습에 사용된 지표는 재무적 특성을 나타내는 요인 30가지, 외부 원천별 지표 13가지, 고유계정 항목 37가지를 포함하여 총 80가지의 변수를 포함한다(<부록 표1> 참조). 지표를 선정하는 데 있어서 재무정보의 다면적인 특성을 최대한 반영하기 위하여 다양한 변수를 포괄적으로 고려하였다. 특히 전통적인 모형에서 정보의 중복성 또는 다중공선성(multicollinearity)에 대한 우려로 재무적으로 유사한 정보를 다른 식으로 표현한 변수 중 일부만 고려하였던 데 비하여 머신러닝 모형에서는 이를 모두 포함하여 학습하였다. 나아가 재무비율을 계산하는 데 있어서 기초가 되는 계정 항목들을 해당 비율과 동일선상에서 학습시켰는데, 이는 머신러닝 학습 알고리즘의 특성에 근거하여 모형이 최적화 과정에서 스스로 가장 유용한 변수를 선택하도록 유도하기 위함이다.

<부록 표1> 재무제표 정량적 지표 목록

기업 특성	번호	세부 지표	변수명		
			개별재무제표	연결재무제표	
한계 요인의 분해	수익성	1	매출총이익률	GPM	GPM_m
		2	매출액영업이익률	OPM	OPM_m
		3	매출액순이익률	NPM	NPM_m
		4	총자산순이익률	ROA	ROA_m
		5	자기자본순이익률	ROE	ROE_m
		6	회계보수성	C_Score	
		7	성과대응재량적발생액	PMDA	
	안정성	8	유동비율	CACL	CACL_m
		9	당좌비율	LIQ	LIQ_m
		10	현금보유비중	CHOLD	CHOLD_m
		11	차입금의존도	DtoA	DtoA_m
		12	자본잠식률	ImpCapt	ImpCapt_m
	활동성	13	총자산회전율	ATTURN	ATTURN_m
		14	재고자산회전율	InvTURN	InvTURN_m
		15	매출채권회전율	RecTURN	RecTURN_m
		16	매입채무회전율	PayTURN	PayTURN_m

<부록 표1> 재무제표 정량적 지표 목록(계속)

기업 특성		번호	세부 지표	변수명	
				개별재무제표	연결재무제표
한계 요인의 분해	신용 위험	17	신용등급	InRATE	
		18	부도거리	DD_iterative	
		19	부도확률	PD_iterative	
		20	차입이자율	INTRATE	
		21	영업이익변동성	OITAStd	OITA_mStd
		22	순이익변동성	ROEStd	ROE_mStd
		23	감사의견	OPN	
		24	기업규모	lnASSET	lnASSET_m
	시장 위험	25	Beta	BETA	
		26	과거수익률(3, 6, 12개월)	L0_R3m	
		27		L0_R6m	
		28		L0_R12m	
		29	고유변동성	IDVOL	
		30	비유동성	ILLIQUIDITY	
외부 원천별 추가 정보	감독 당국	31	3년연속 영업손실 여부	LossOI3yr	
		32	3년연속 영업현금흐름손실 여부	LossCFO3yr	
		33	4년연속 영업손실 여부	LossOI4yr	
		34	2년연속 자본잠식률 50% 이상 여부	HugeImp	
		35	최근 3사업연도 내 2회이상 자기자본 50%초과 세전손실 여부	HugeLoss	
	투자자 밸류 에이션	36	장부가치 대비 시장가치 비율	BM3	BM3_m
		37	순이익 대비 시장가치 비율	EP3	EP3_m
		38	잉여현금흐름 대비 시장가치 비율	CP3	CP3_m
		39	현금배당 대비 시장가치 비율	DP3	DP3_m
		40	시장가치	LnMV	
	기타	41	무형자산 집중도	RDSALE	RDSALE_m
		42	주식발행 척도(3년, 5년)	EISSUE3yr	
		43		EISSUE5yr	
	계정항목	44~ 80	고유계정 항목	37개 항목	

2. 연결재무제표 학습 모형 분석 결과

본 장에서는 앞서 IV-2.절과 3.절에서 논의한 실증분석을 연결재무제표를 학습한 모형으로 확장하여 결과의 강건성을 확인하였다. 머신러닝 모형의 학습 효율성 및 예측 정확도는 개별재무제표를 학습한 경우와 연결재무제표를 학습한 경우에서 비슷하게 높은 수치를 보였으나 연결재무제표의 경우 정성적 정보를 포함한 모형과 그렇지 않은 모형의 편차가 개별재무제표의 경우보다 다소 크게 관측된 점(IV-1.절 참조)으로 미루어 볼 때, 실증분석 일관성을 가지는지 여부를 추가로 확인할 필요성이 있다. 따라서 본문의 내용과 대응하여 연결재무제표 모형의 예측 결과를 바탕으로 부실 징후가 있는 기업과 그렇지 않은 기업의 재무·공시 지표를 t-test를 통해 비교하였으며 연결재무제표의 예측 오차에 기여하는 재무공시의 특성 요인을 회귀분석을 통해 확인하였다.

가. 부실 기업의 특성 비교

<부록 표2>에서는 연결재무제표 학습 모형을 바탕으로 부실 징후가 있는 기업과 그렇지 않은 기업을 구분한 후, 두 집단 사이의 평균적인 특성을 t-test를 통해 비교하였다. 연결재무제표 학습 모형에서도 부실 징후가 발견되는 기업과 그렇지 않은 기업의 특성 사이에 유사한 관계성을 확인할 수 있다. 즉, 부실 가능성이 높게 예측된 기업은 상대적으로 작은 규모($\ln ASSET_m$)이며, 차입금의존도($DtoA_m$)가 높고, 자기자본순이익률(ROE_m)은 낮은 반면 그 변동성(ROE_mStd)은 높은 경향을 보였다. 더불어 지난 3년간 영업손실($LossOI3yr$)이 있었을 확률이 높고 재량적발생액의 절댓값($absPMDA$)이 큰 것으로 나타났다. 공시의 특징적인 측면에서도 분량(\log_fs_length)이 다소 짧고 정보의 반복적 강조($word_reps$)가 적은 것으로 확인되어, 전체적으로 개별재무제표 학습 모형의 결과(<표 IV-4>)와 해석이 일치하는 것을 확인할 수 있다.

<부록 표2> 연결재무제표 모형 예측 t-test 결과

		PN	PP	t-stat	p-value
부실확률 예측값		0.0017	0.7072	-118.65	0.0000
재무적 특성	lnASSET_m	19.4240	18.1284	2.6830	0.0073
	DtoA_m	0.4394	0.7014	-3.9165	0.0001
	BM3_m	1.1201	1.4143	-0.9607	0.3368
	SALEG_m	0.1111	-0.1974	0.8915	0.3727
	OIG_m	-0.2184	-0.4761	0.0560	0.9553
	ROE_m	-0.0223	-0.6413	6.4666	0.0000
	ROE_mStd	0.1585	0.5683	-2.9491	0.0032
	OITA_mStd	0.0445	0.1308	-6.0382	0.0000
	InvTURN_m	74.0454	8.5290	0.1270	0.8990
	RecTURN_m	9.3999	6.2033	0.1774	0.8592
	LossOI3yr	0.0678	0.6667	-7.1188	0.0000
	absPMDA	0.0977	0.2400	-4.2622	0.0000
공시 정보 특성	fs_length	11.5514	11.2112	2.1933	0.0283
	word_reps	28.1420	20.6876	2.7457	0.0061
	Δunique_words	-0.0191	-0.0092	-0.1286	0.8977
	cosine_l	0.8243	0.8497	-0.6229	0.5334
	cosine_p	0.5385	0.5186	1.4666	0.1425

주: 부실 확률 예측값이 0.5를 초과하는 표본을 PP, 그렇지 않은 경우를 PN으로 분류하여 각 변수들의 표본평균의 차이를 t-test로 비교하였으며 5% 유의수준에서 통계적으로 유의미한 차이를 보이는 변수는 굵은 글씨로 표시하였다.

나. 공시 정보의 유용성 기여 요인 평가

<부록 표3>은 본문에서 논의한 회귀분석을 연결재무제표를 학습한 모형에도 동일하게 적용하여 예측 오차에 기여하는 요인을 분석하였다. 앞서 본문에서 개별재무제표 모형을 바탕으로 분석한 결과(<표 IV-5> 참조)와 유사한 값을 나타냄을 확인할 수 있다. 우선, 전년도 대비 공시의 정보량이 증가할수록 부실 징후를 예측하는 데 공시 자료가 유용한 것으로 나타났다(아래 결과 (1), (2) 참조). 반면 재무 공시의 양적인 면은 예측 오차를 줄이는데 기여하지 않으며(결과 (3)) 중요한 정보의 반복적인 강조는 유용한 것으로 나타났다(결과 (4)). 종합적으로 볼 때, 연결재무제표에서도 증분적인 정보를 간결하게 전달하는 것이 유용성 측면에서 가장 바람직한 것으로 볼 수 있다.

<부록 표3> 연결재무제표 모형 오차 회귀분석 결과

	정보량 측면		공시 양태 측면		종합	
	(1)	(2)	(3)	(4)	(5)	(6)
Δunique_words	0.1939*** (0.0454)	0.1845*** (0.0461)			0.3040* (0.1262)	0.2820* (0.1271)
fs_length			-0.3793** (0.1242)	-0.4376** (0.1351)	-0.3281** (0.1253)	-0.3989** (0.1372)
word_reps			0.0181*** (0.0052)	0.0201*** (0.0056)	0.0160** (0.0052)	0.0185** (0.0056)
cosine_1			0.1216 (0.0811)	0.1634 (0.0884)	0.1403 (0.1084)	0.2343 (0.1283)
cosine_p			1.7066* (0.8112)	2.1765* (0.8852)	1.8737* (0.8029)	2.3320** (0.8800)
표본수	4966	4966	4995	4995	4873	4873
R^2	0.129	0.133	0.134	0.140	0.136	0.142
$\overline{R^2}$	0.127	0.129	0.132	0.135	0.133	0.137
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year-Ind FE	No	Yes	No	Yes	No	Yes
Cluster SE	firm	firm	firm	firm	firm	firm

주: 괄호 안의 수치는 표준오차를 의미하며 추정된 회귀계수가 5%, 1%, 0.1% 유의수준에서 통계적으로 유의한 값을 가지는 경우 *, **, ***을 추정값 우측에 표시하였음. 모든 결과는 동일한 통제변수를 포함하고 있으며 연도·산업군별 고정효과(Fixed Effect: FE)는 (2), (4), (6)에만 포함하였다. 표준오차는 기업 수준에서 군집표준오차 (clustered standard error)로 산출하였다.