



PREDICTION

딥러닝을 활용한 기업 위험 예측

2019. 4

목 차

왜 기업 부실 예측 데이터가 필요한가?

01

예측모형 설계 및 검증 결과

02

산업별 부실 위험 모니터링

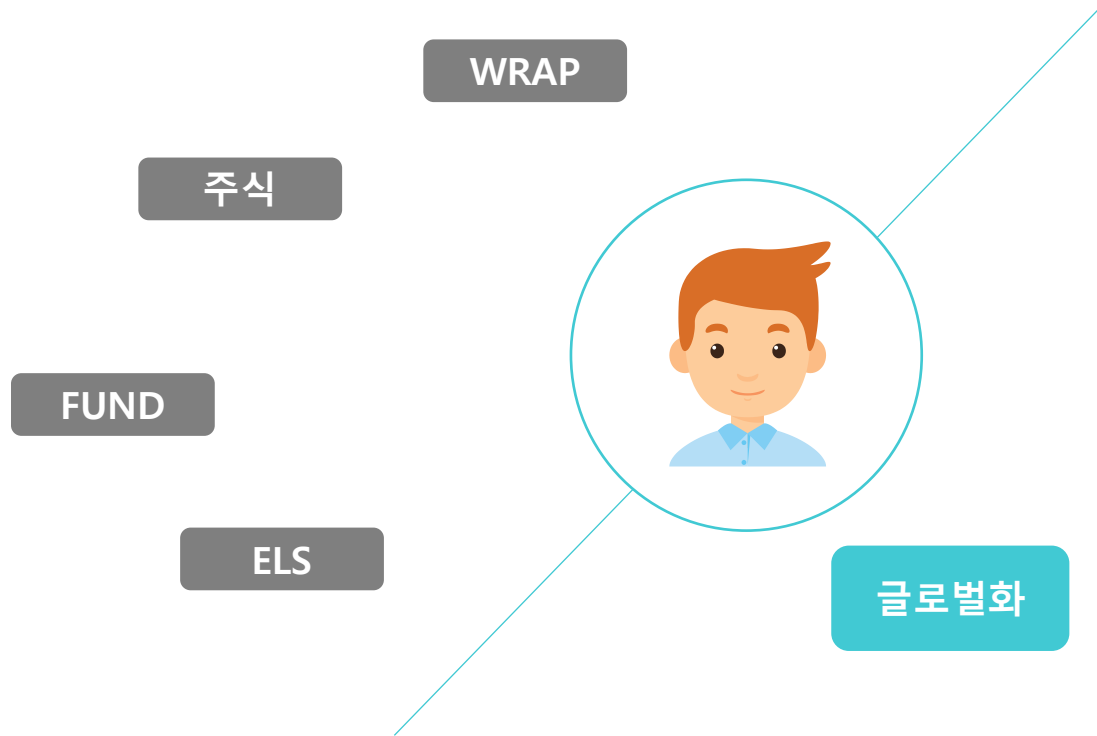
03

기업위험 예측 변수 선정

04

1-1. 고객 금융 경험의 변화

고객의 변화

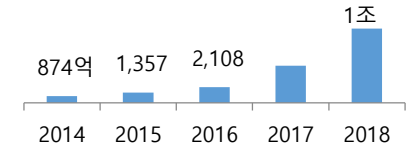


1. 왜 기업부실 예측 데이터가 필요한가?

* 2017년 은행권 자산관리 서비스 도입 이후 급격히 증가
 * 2018년부터 확장기 진입하여 2020년 5조, 2025년 35조 예상
 (출처: 2018년 한국로보어드바이저 보고서, 하나은행)
 * 2016년 한국과학기술정보연구원의 2018년 5,000억 예상규모를 2배 초과 달성함

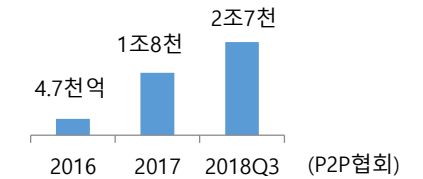
Plus

로보어드바이저



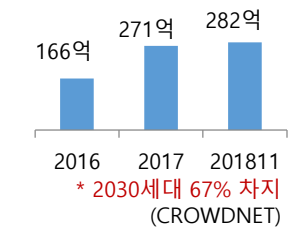
Plus

P2P대출¹



Plus

증권형클라우드펀딩



- ✓ 기성세대의 새로운 상품 경험의 증대 및 금융 경험이 없던 영 밀레니얼세대²의 금융에 대한 인식 패러다임 전환
- ✓ 당사 고객의 데이터 분석으로 확인될 수 없는 미래 잠재고객들의 금융경험이 급격히 변화하고 있음

주1) P2P대출(peer to peer platform) : 주로 부동산담보/PF 플랫폼이 다수이나, 부동산 외에도 '기업어음', '개인신용', '소상공인법인', 'ART' 등 다양한 종류의 P2P 대출이 확대되는 추세임

주2) 영 밀레니얼 세대 : 1980년~2000년대 초반에 태어난 밀레니얼 세대 중 18~24세

33곳 상폐사유·69곳 관리종목... '회계감사 실패' 휩쓴 3월

파이낸셜뉴스 입력 : 2019.04.02 16:58 수정 : 2019.04.03 10:54

코스피 5개사 상폐절차에 돌입 동부제철 등 3곳 관리종목 지정
아시아나·금호산업 등도 충격파. 코스피도 상폐사유 50%나 늘어

■ 코스피, 5개사 상폐 사유 발생

올해는 코스피 상장사도 회계감사 이슈에서 자유롭지 못했다. 2일 한국거래소에 따르면 12월 결산법인 가운데 5개사는 상장폐지 절차가 진행되고, 3개사는 관리종목으로 지정됐다.

■ 코스닥, 상폐 사유 발생 '급증'

코스닥시장에서는 상폐 사유 발생기업이 급증했다. 당초 시장 예상치(40~50개)만 증가율이 높았다. 상장폐지 사유가 발생한 코스닥 상장사는 28개사에 이른다. 다 50% 이상 늘어난 수치다

현실된 '新외감법 피바람'...상폐위기 상장사 '급증'

매일경제 김규리 기자 입력 : 2019.04.02 16:19:05

상폐 사유 기업이 증가한 배경에는 올해부터 강화된 '외부감사에 관한 법률(외감법)' 때문이라는 분석이 나온다. 금융당국은 지난 2015년 대우조선해양의 분식회계 사건을 계기로 개정된 신외감법은 감사인의 독립성을 강화하면서 책임 역시 대폭 강화했다. 이에 따라 외부감사인에서도 기업 회계내용에 대해 보다 깐깐한 감사 기준을 적용하면서 높은 회계기준을 적용한 것으로 보인다.

올해 관리종목으로 34개사가 신규 지정된 반면, 12개사가 관리종목에서 해제되면서 관리종목 순증 규모는 23개로 집계됐다. 이 역시 전년(13곳)대비 10곳이나 늘어난 것이다. 현재 관리종목으로 지정된 법인은 총 66개사다.

투자주의환기종목 순증은 23개사로 전년(17개사) 대비 6개사 증가했다. 신규 지정된 30개사에서 해제된 7개사를 제한 수치다. 투자주의환기종목 지정법인은 총 52개사다.

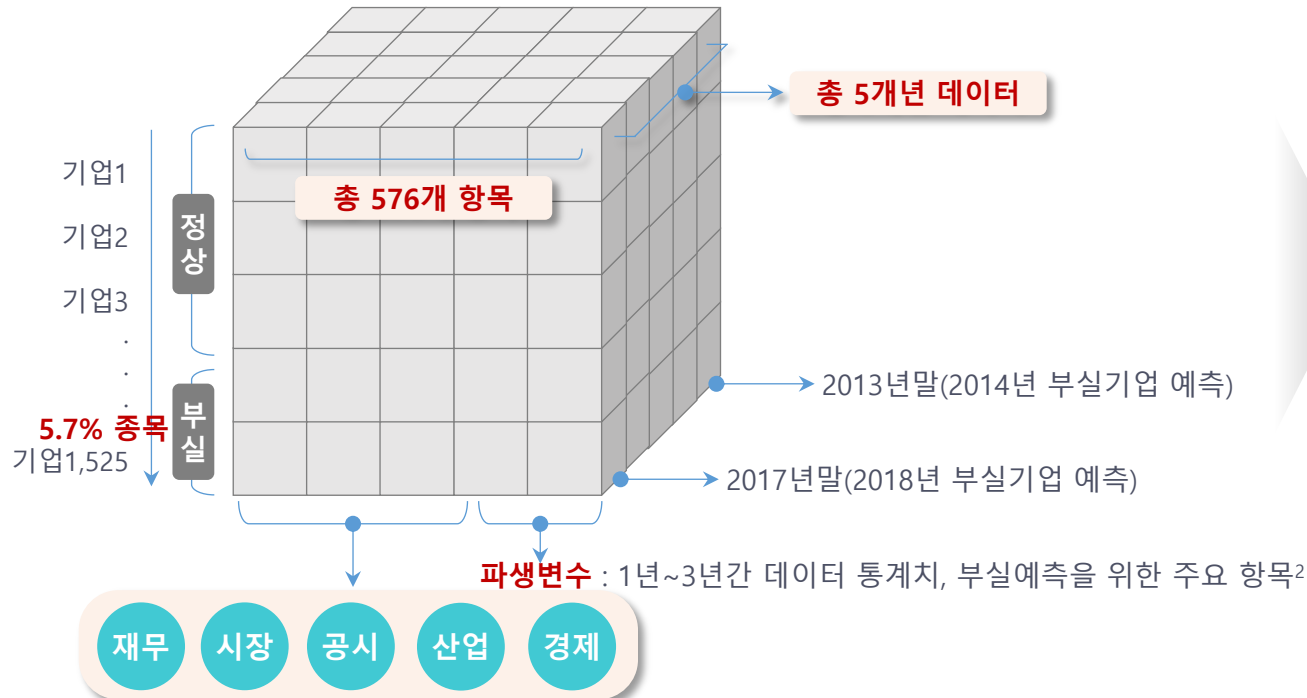
2-1. 예측모형 모델링 절차(1/2)

2. 예측모형 설계 및 검증 결과

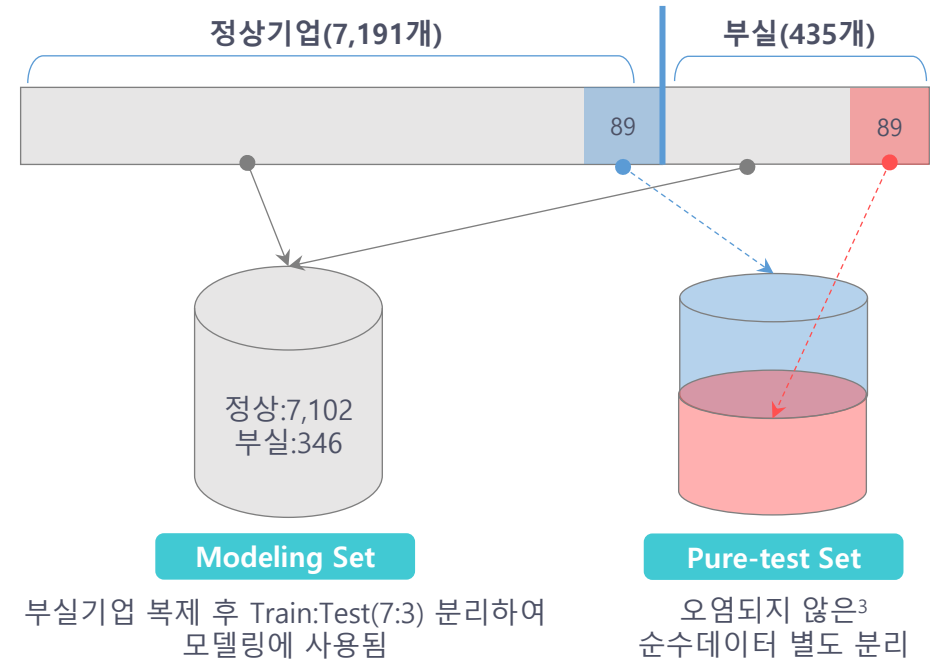
데이터셋 확보

- ✓ [대상기업] 전체기업 : **KRX 상장기업**(단, 금융업 제외, 상장 후 최소 2년 경과 기업. 연평균 1,525개 → 총 7,626개)
부실기업 : **상장폐지, 관리종목, 투자주의환기종목, 한계기업¹** (연평균 87개 → 총 435개(5.7%))
- ✓ [대상기간] 분석기준 : **2013년~2017년말 기준. 총 5개년**
수집데이터 : 2011년 이후. 기준연도별 최근 3년간 데이터 포함
- ✓ [데이터 수집 항목] : **총 576개. 재무/시장,공시(Dart),산업(통계청),경제(BOK)**
- ✓ [모델링 결과 해석] : **부실 기업 가능성**

[횡단면] 동시점 데이터 수집 및 파생변수 생성, [종단면] 시계열 데이터 확보



순수검증용 데이터셋과 모델링셋 분리 2차 검증 수행

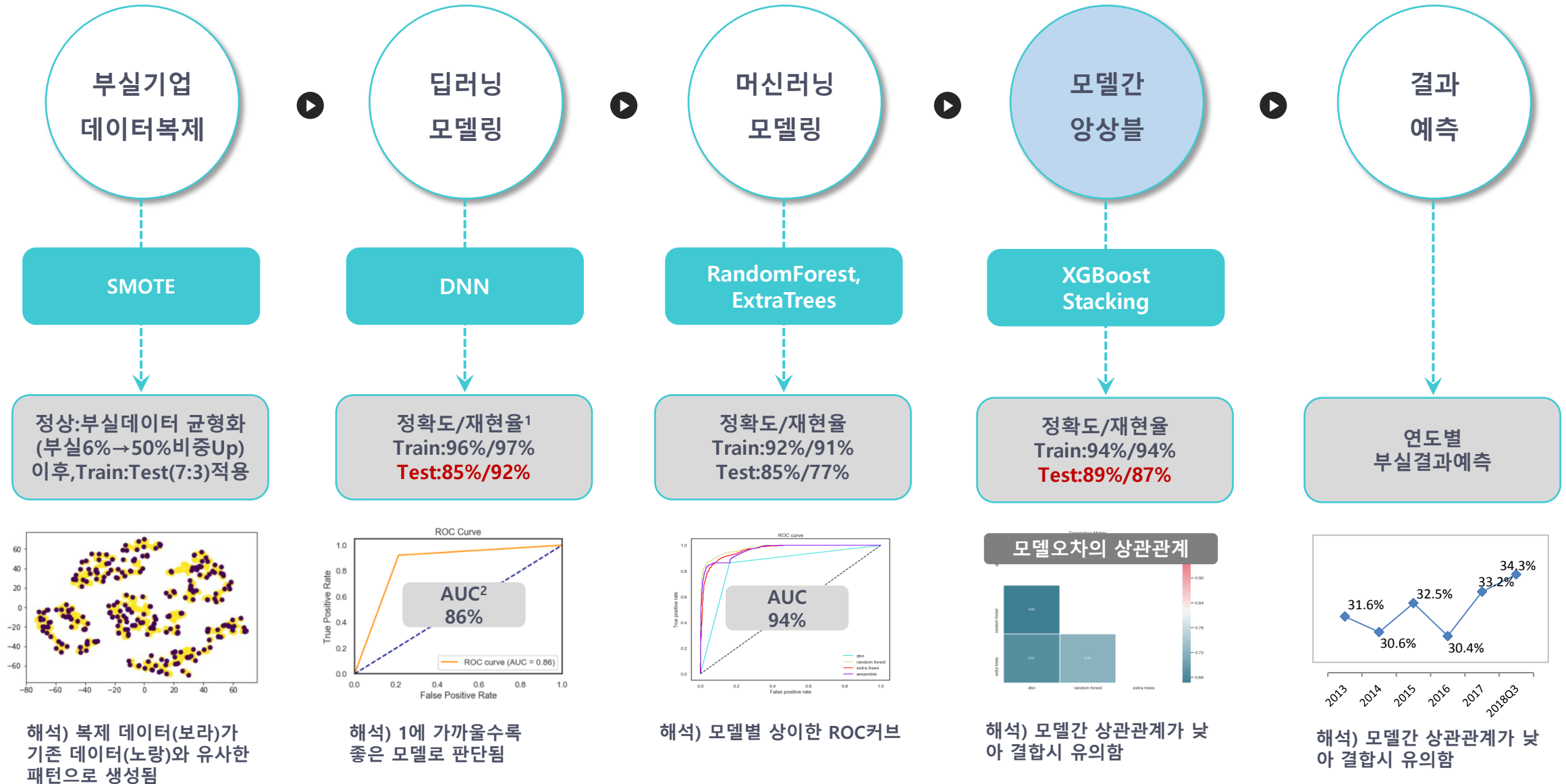


주1) 한계기업이란, 3년간 이자보상비율 1미만을 의미함. 단, 0.7 이상의 경우 이듬해 회복될 가능성이 있는 것으로 판단되어 직전 해 기준은 0.7로 변경하여 대상군 산출함.

(즉, 이자보상배율 3년전 1미만, 2년전 1미만, 1년전 0.7 미만 대상)

주2) 부실예측을 위한 주요 항목 : 논문 및 연구자료 참고

주3) 데이터 복제 및 모델링에 영향을 받지 않은 순수데이터 검증을 위해 사전에 별도 분리함(Pure기업은 ModelingSet에서 제외함)



주1) 재현율(Recall) : 실제 부실 기업을 부실로 예측한 확률(30회 반복 수행 평균)

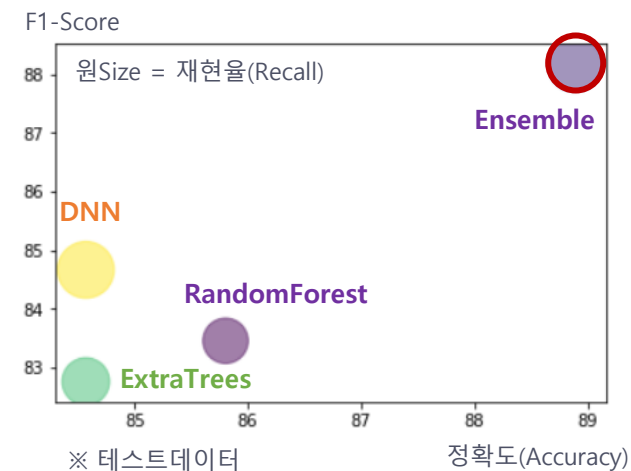
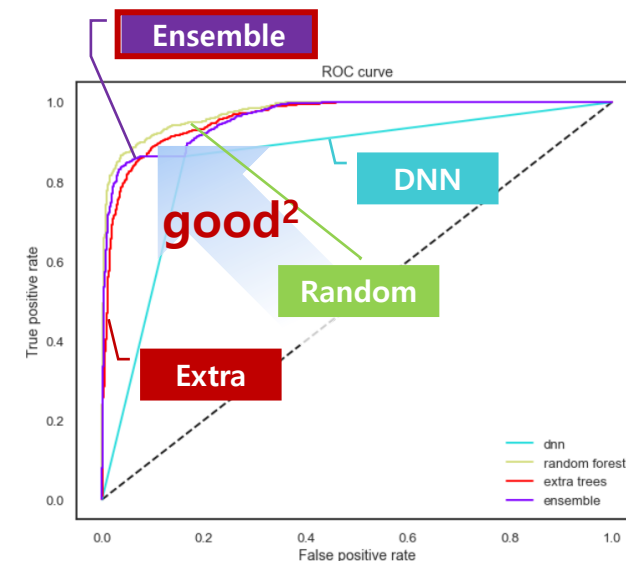
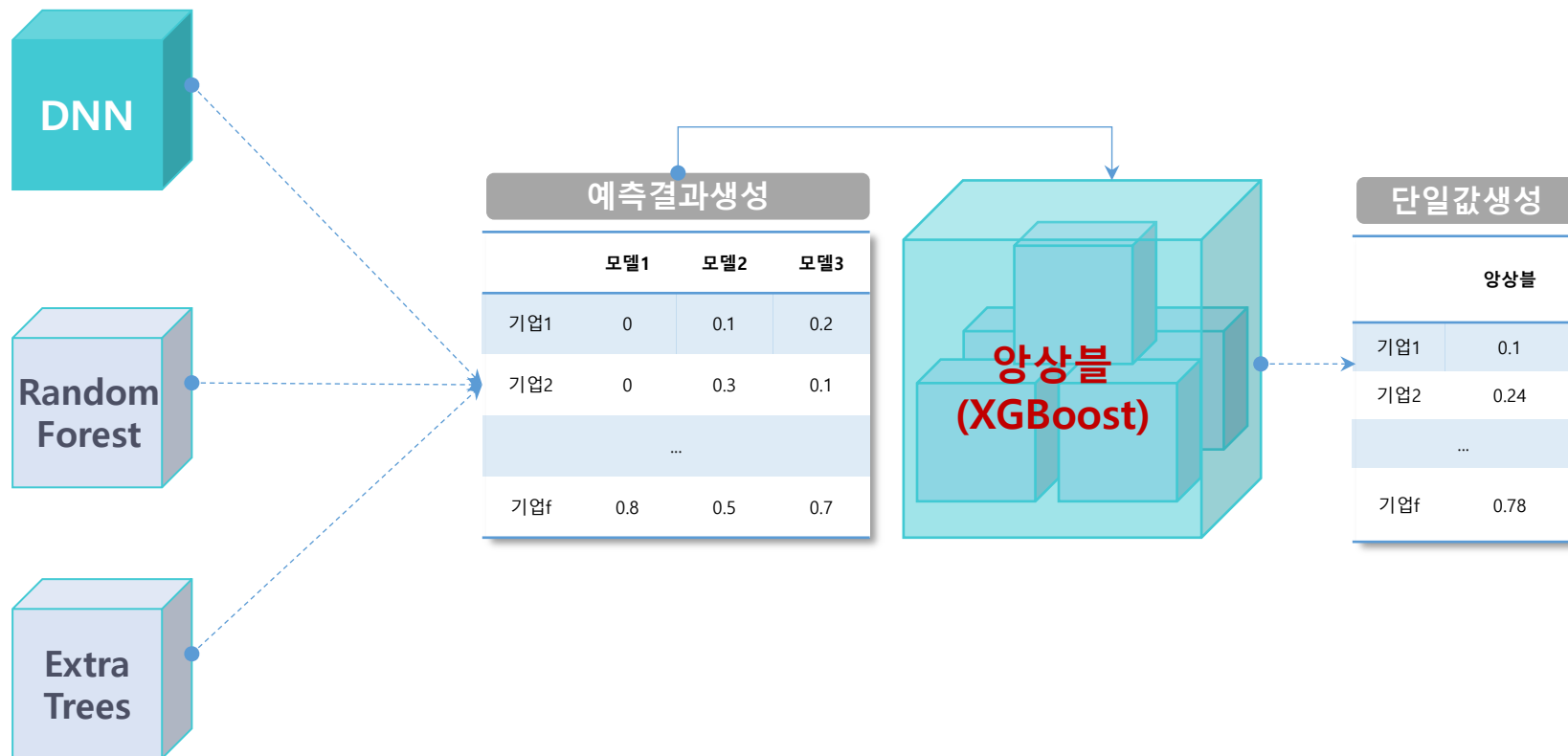
주2) AUC : ROC 커브의 면적으로 민감도와 특이도를 고려해 우수한 모델을 판단하는 기준(30회 반복 수행 평균)

2-2. 앙상블 모형 성능 검증

2. 예측모형 설계 및 검증 결과

특성이 다른 3개 모델 선정¹

앙상블 모델링을 통해 성능 극대화



주1) 모델별 기본적 튜닝 적용. 로지스틱회귀모형/Adaboost/KNN 등은 테스트 재현율 저하 및 트레인 오버피팅 우려로 제외함

주2) AUC : ROC 커브의 면적으로 민감도와 특이도를 고려해 우수한 모델을 판단하는 기준(30회 반복 수행 평균)

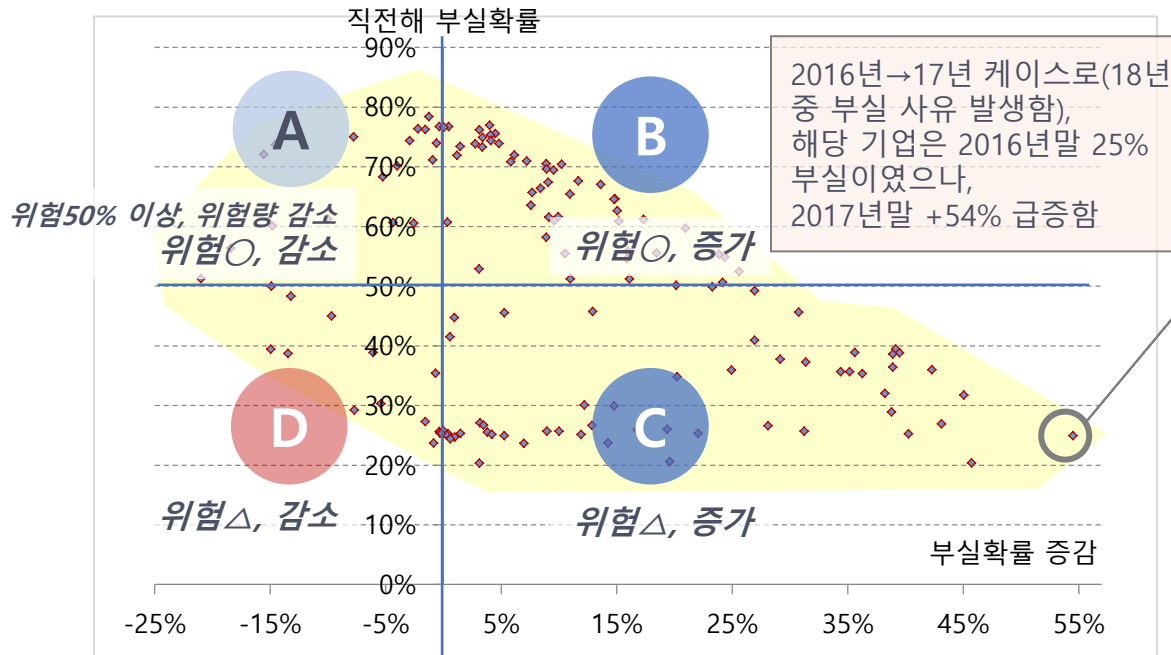
2-3. 결과 데이터 검증 및 분석(부실기업1/2)

2. 예측모형 설계 및 검증 결과

부실기업으로 최초 전환된 기업의 '직전 해 위험량 대비 부실 증감량' 비교

주1) 확보 데이터 기준으로 2014년 이후 최초 부실 기업 전환 기준

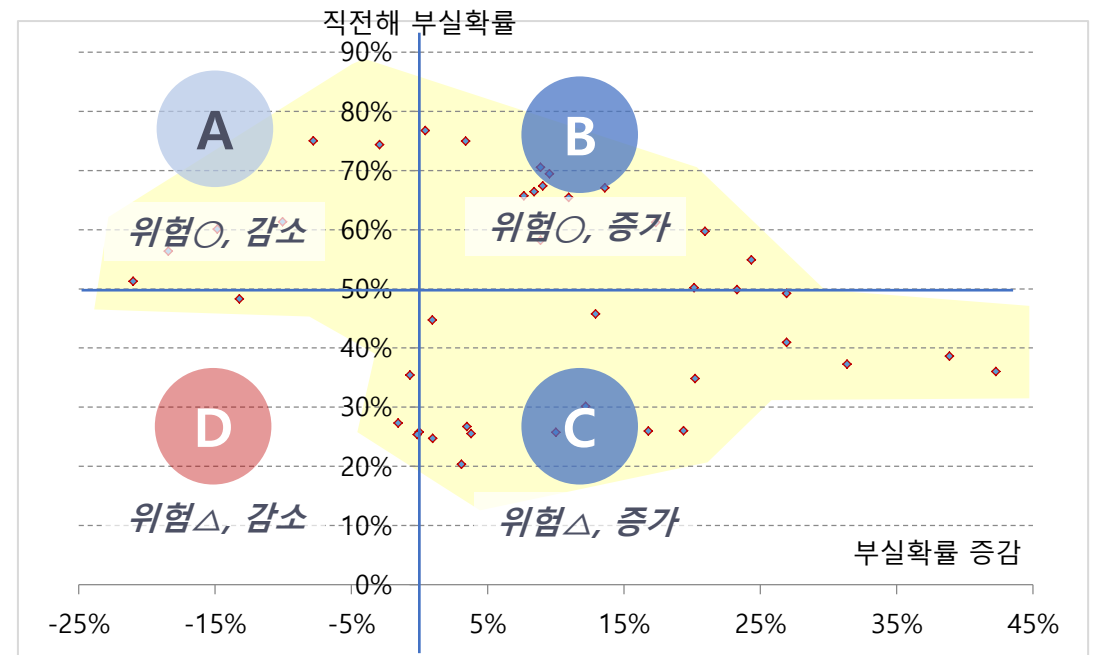
전체 데이터 위험량 증감 분포



| 분류 | 기업수 | 비중 | 직전해 | 기준해 | 증감 |
|-----|-----|--------|-------|-------|-------|
| A | 18 | 13.1% | 68.7% | 61.6% | -7.1% |
| B | 50 | 36.5% | 64.9% | 75.1% | 10.2% |
| C | 54 | 39.4% | 31.4% | 52.4% | 21.0% |
| D | 15 | 10.9% | 33.9% | 27.9% | -6.0% |
| 총합계 | 137 | 100.0% | 48.8% | 59.2% | 10.4% |

실제 부실 전환 기업을 '위험량 증가'로 예측한 비중(B+C) **76%**, 위험량 감소하였으나 '이미 위험 수준'에 머물러 있는 비중(A) **13%**

테스트 데이터 위험량 증감 분포



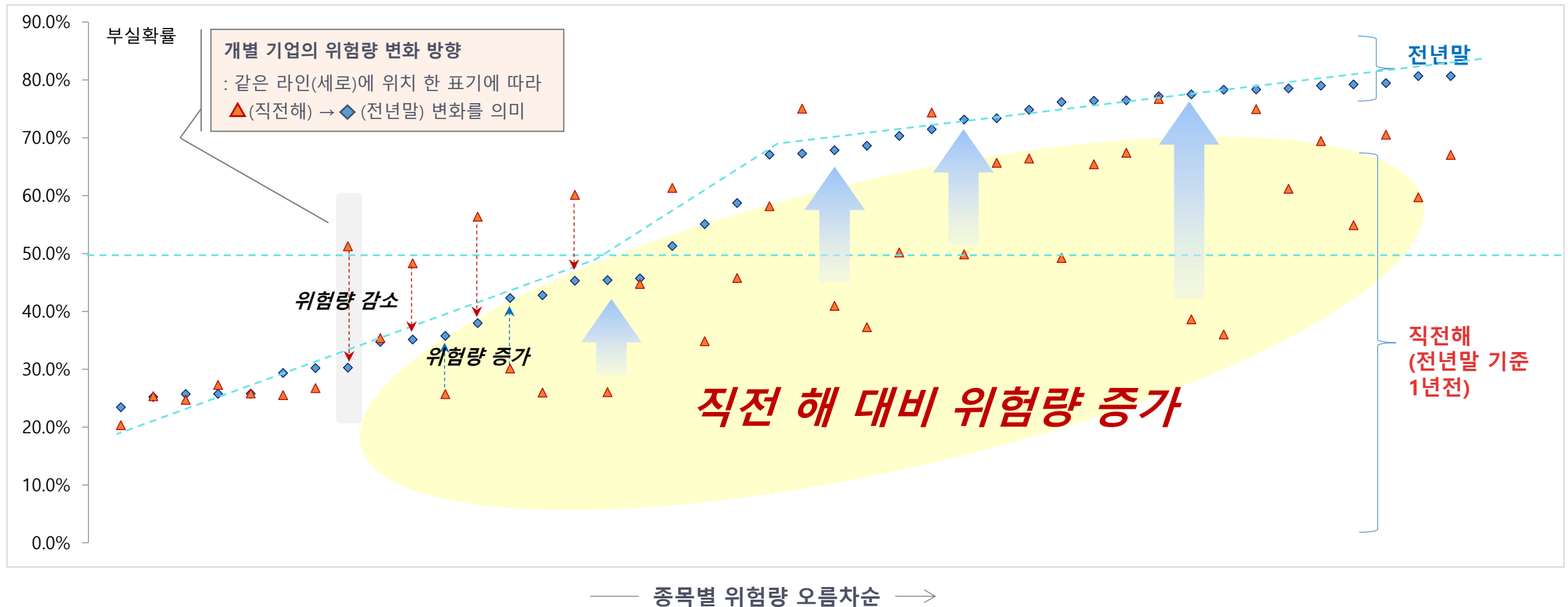
| 분류 | 기업수 | 비중 | 직전해 | 기준해 | 증감 |
|-----|-----|--------|-------|-------|--------|
| A | 6 | 14.3% | 63.1% | 50.6% | -12.5% |
| B | 14 | 33.3% | 64.8% | 76.5% | 11.7% |
| C | 17 | 40.5% | 34.2% | 51.5% | 17.3% |
| D | 5 | 11.9% | 32.4% | 29.3% | -3.1% |
| 총합계 | 42 | 100.0% | 48.3% | 57.1% | 8.7% |

실제 부실 전환 기업을 '위험량 증가'로 예측한 비중(B+C) **74%**, 위험량은 감소하였으나 '이미 위험 수준'에 머물러 있는 비중(A) **14%**

2-3. 결과 데이터 검증 및 분석(부실기업2/2)

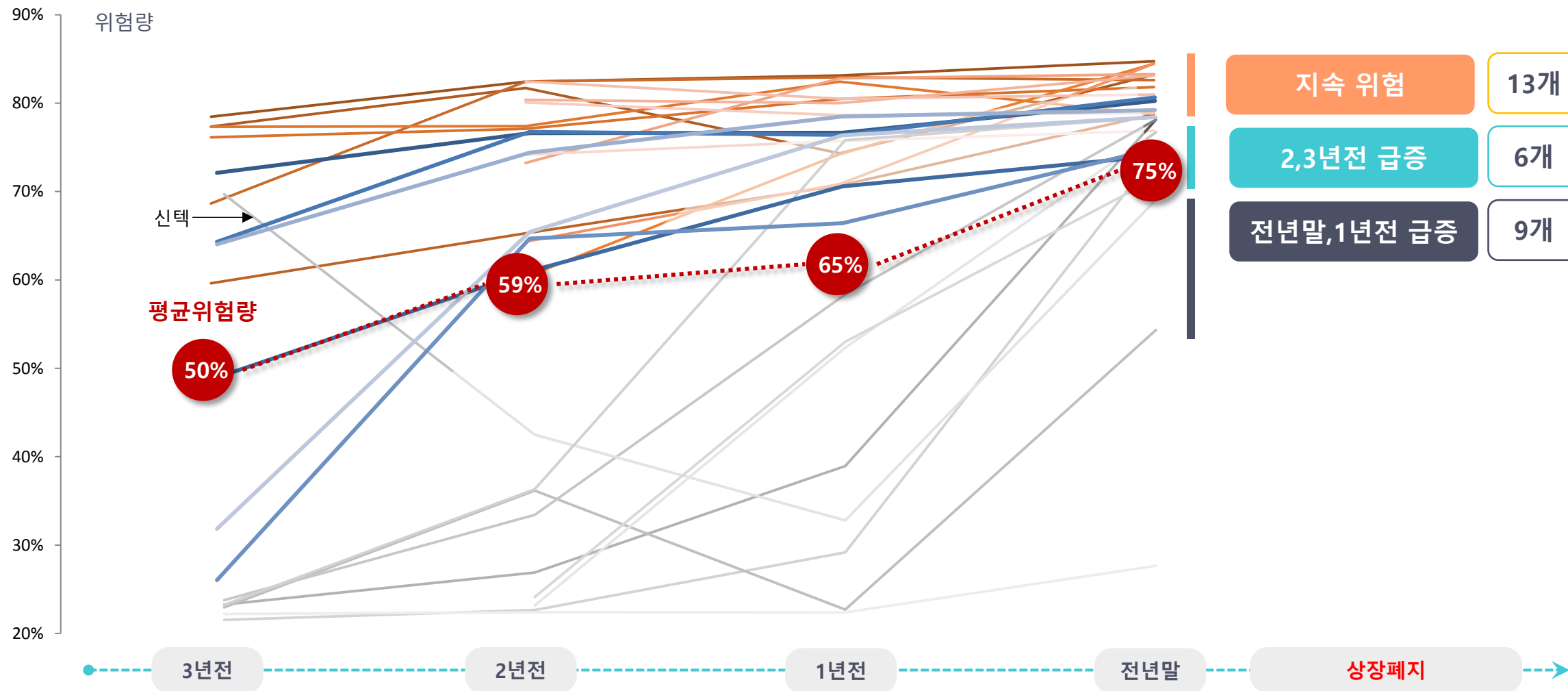
부실기업으로 최초 전환된 기업의 '직전 해 위험량 대비 부실 증감량' 분포

테스트 데이터 부실 전환 기업 분포



2-3. 결과 데이터 검증 및 분석(상폐기업)

부실기업 대상 중 **상장폐지 기업 위험량 추이**



※ 2016~2018 상장폐지 기업 중 데이터 확보 기업 28개 대상

2018년 상장폐지 사유발생 기업 부실확률 추이(1/2)

정상
30%미만

주의
30~50%

위험
50~70%

고위험
70%이상

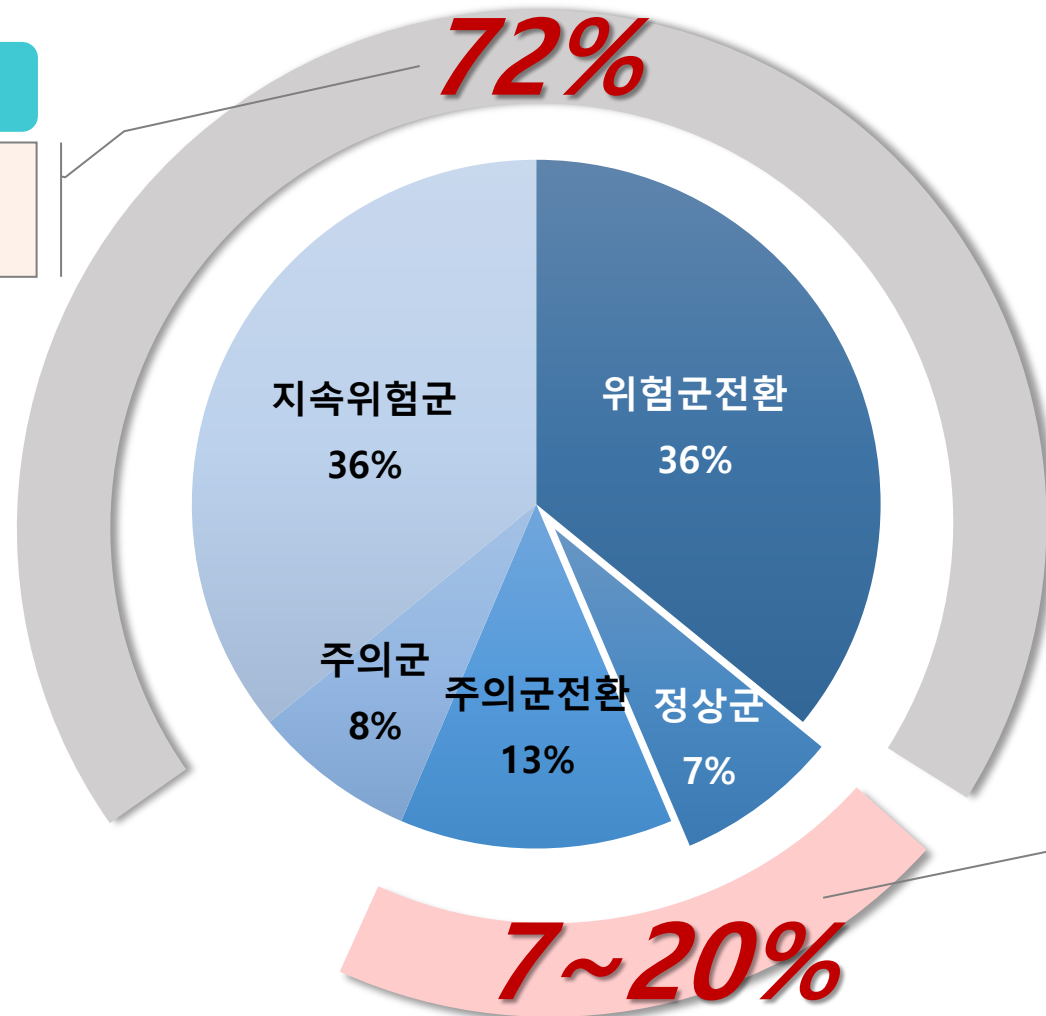
첨부

※ 감사보고서 미제출 이슈기업 포함

| no | 시장 | 기업 | 상장일 | 감사의견 | 201512 | 201612 | 201712 | 201809 | 201812 | 위험량변화그룹 (~18년9월) |
|----|-----|----|-----|----------|--------|--------|--------|--------|--------|---------------------|
| 1 | 코스닥 | | | 의견거절 | 74% | 67% | 74% | 78% | 81% | 지속위험군 |
| 2 | 코스닥 | | | 의견거절 | 76% | 77% | 81% | 81% | 78% | 지속위험군 |
| 3 | 거래소 | | | 의견거절 | 71% | 77% | 81% | 81% | 77% | 지속위험군 |
| 4 | 코스닥 | | | 의견거절 | 51% | 50% | 73% | 77% | 77% | 지속위험군 |
| 5 | 코스닥 | | | 의견거절 | 79% | 79% | 79% | 71% | 77% | 지속위험군 |
| 6 | 코스닥 | | | 의견거절 | 66% | 75% | 81% | 78% | 77% | 지속위험군 |
| 7 | 거래소 | | | 의견거절 | | | 66% | 76% | 76% | 지속위험군 |
| 8 | 코스닥 | | | 의견거절 | 49% | 64% | 80% | 81% | 76% | 지속위험군 |
| 9 | 코스닥 | | | 의견거절 | 77% | 39% | 79% | 71% | 74% | 지속위험군 |
| 10 | 코스닥 | | | 의견거절 | | | 49% | 71% | 74% | 지속위험군 |
| 11 | 거래소 | | | 의견거절 | 60% | 56% | 36% | 57% | 73% | 지속위험군 |
| 12 | 코스닥 | | | 의견거절 | | | 70% | 70% | 41% | 지속위험군 |
| 13 | 코스닥 | | | 미제출법인 | 61% | 79% | 77% | 77% | | 지속위험군 |
| 14 | 코스닥 | | | 미제출법인 | 75% | 75% | 79% | 81% | | 지속위험군 |
| 15 | 코스닥 | | | 의견거절 | 41% | 36% | 75% | 81% | 81% | 위험군 전환 |
| 16 | 코스닥 | | | 의견거절 | 35% | 39% | 54% | 54% | 81% | 위험군 전환 |
| 17 | 코스닥 | | | 의견거절 | 36% | 37% | 56% | 74% | 78% | 위험군 전환 |
| 18 | 코스닥 | | | 의견거절 | 45% | 39% | 77% | 81% | 77% | 위험군 전환 |
| 19 | 코스닥 | | | 의견거절 | 67% | 36% | 70% | 71% | 77% | 위험군 전환 |
| 20 | 코스닥 | | | 의견거절 | 25% | 32% | 31% | 67% | 74% | 위험군 전환 |
| 21 | 코스닥 | | | 의견거절 | 25% | 25% | 34% | 68% | 71% | 위험군 전환 |
| 22 | 코스닥 | | | 의견거절 | 26% | 29% | 36% | 61% | 71% | 위험군 전환 |
| 23 | 코스닥 | | | 범위제한 한정 | 25% | 25% | 55% | 69% | 69% | 위험군 전환 |
| 24 | 코스닥 | | | 범위제한 한정 | 31% | 36% | 38% | 55% | 60% | 위험군 전환 |
| 25 | 거래소 | | | 의견거절 | 31% | 31% | 56% | 57% | 58% | 위험군 전환 |
| 26 | 거래소 | | | 미제출법인 | 26% | 25% | 53% | 56% | | 위험군 전환 |
| 27 | 코스닥 | | | 미제출법인 | 35% | 32% | 70% | 69% | | 위험군 전환 |
| 28 | 코스닥 | | | 미제출법인 | 26% | 25% | 28% | 52% | | 위험군 전환 |
| 29 | 코스닥 | | | 의견거절 | | | 42% | 39% | 69% | 주의군 |
| 30 | 코스닥 | | | 의견거절 | 70% | 29% | 29% | 38% | 66% | 주의군 |
| 31 | 코스닥 | | | 범위제한 한정 | 46% | 30% | 35% | 38% | 61% | 주의군 |
| 32 | 코스닥 | | | 의견거절 | 35% | 25% | 30% | 42% | 75% | 주의군 전환 |
| 33 | 코스닥 | | | 의견거절 | 25% | 26% | 26% | 45% | 72% | 주의군 전환 |
| 34 | 코스닥 | | | 의견거절 | 25% | 24% | 31% | 30% | 68% | 주의군 전환 |
| 35 | 코스닥 | | | 범위제한 한정 | 25% | 25% | 25% | 31% | 67% | 주의군 전환 |
| 36 | 코스닥 | | | 미제출법인 | 26% | 25% | 25% | 37% | | 주의군 전환 |
| 37 | 코스닥 | | | 의견거절 | 26% | 26% | 25% | 25% | 54% | 정상군 |
| 38 | 거래소 | | | 주식분산요건미달 | 26% | 25% | 26% | 26% | 27% | 정상군 |
| 39 | 코스닥 | | | 의견거절 | | | 23% | 23% | 25% | 정상군 |

사전 위험 대응 가능 영역

지속 모니터링 프로세스 및 위기 대응을 위한
액션 플랜 수립 필요



| 그룹 | 종목수 | 비중 |
|-------|-----|------|
| 지속위험군 | 14 | 36% |
| 위험군전환 | 14 | 36% |
| 주의군 | 3 | 8% |
| 주의군전환 | 5 | 13% |
| 정상군 | 3 | 8% |
| 총합계 | 39 | 100% |

지속 위험 노출 우려 영역

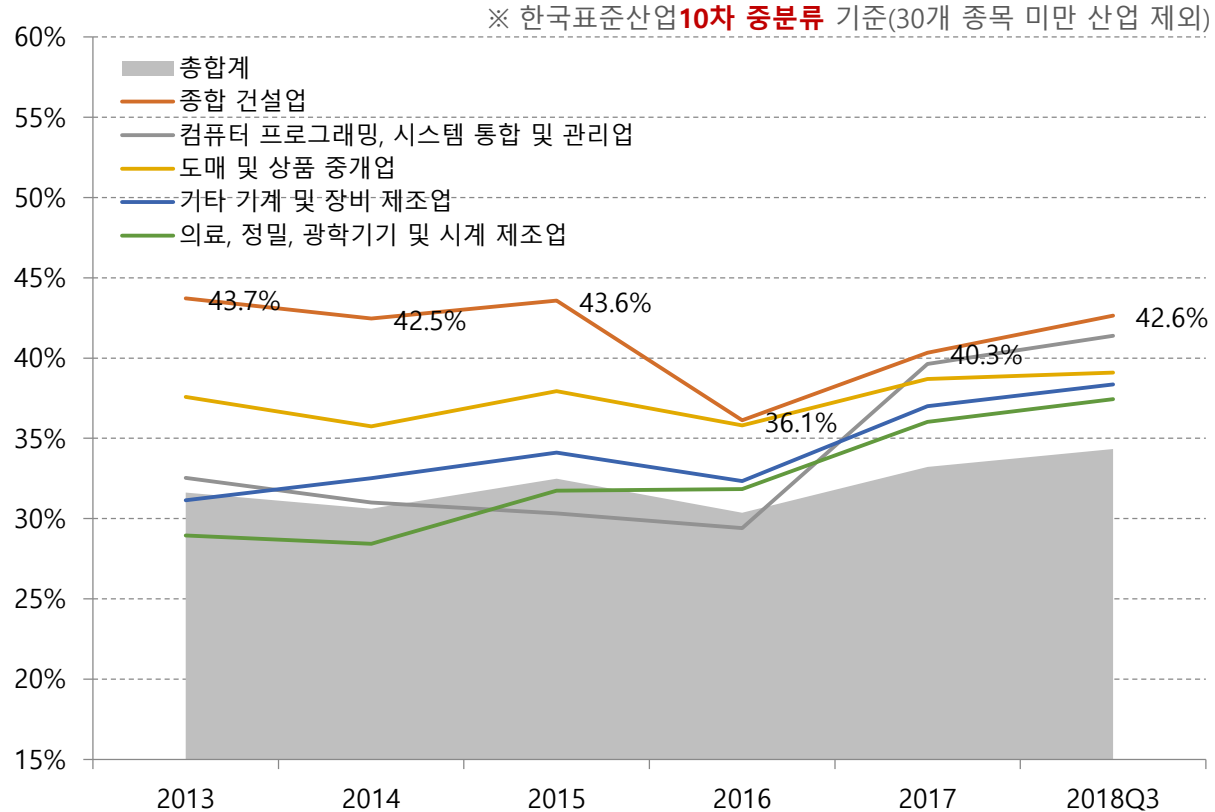
- 상시 기업의 부정 이슈 파악을 위한 지속적인 노력이 필요함.
- 향후 뉴스, SNS 등의 텍스트 분석을 통한 '부실 예측 모형 고도화' 지속 추진

3. 산업별 부실 위험 모니터링(1/2)

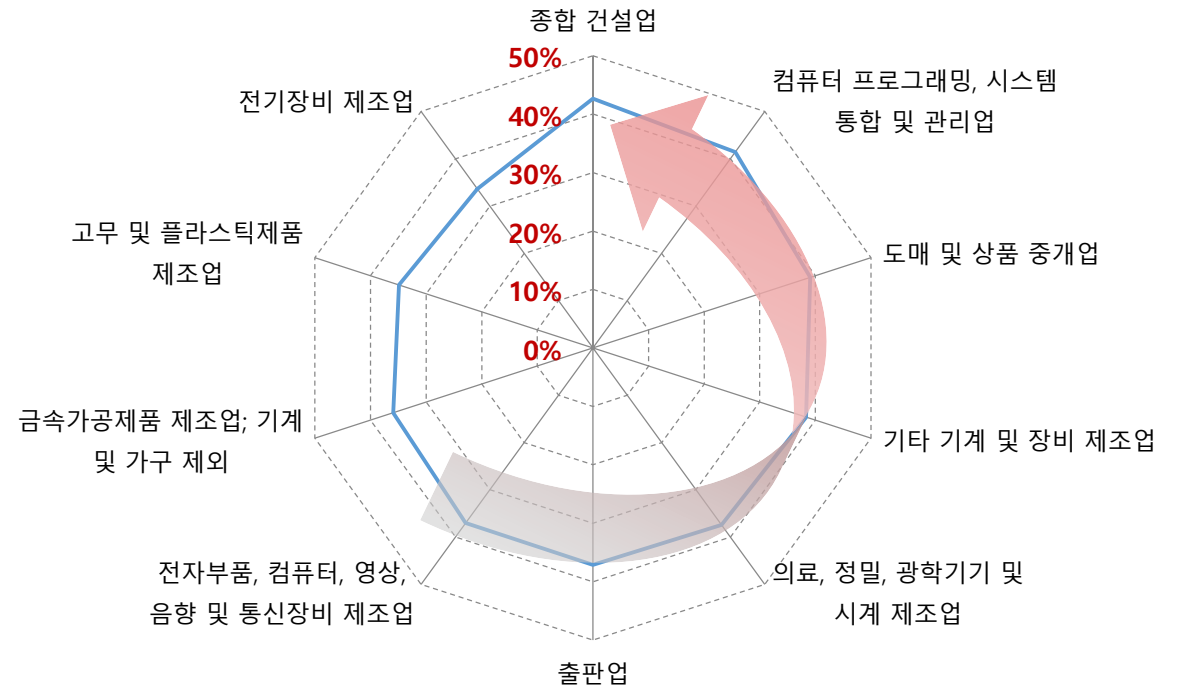
3. 산업별 부실 위험 모니터링

산업별 위험량 변화 감지

산업별 부실위험 추이(Top5)



산업 중분류 위험량 (Top10)



2018년 3분기 기준. 반시계 방향으로 부실위험이 높음

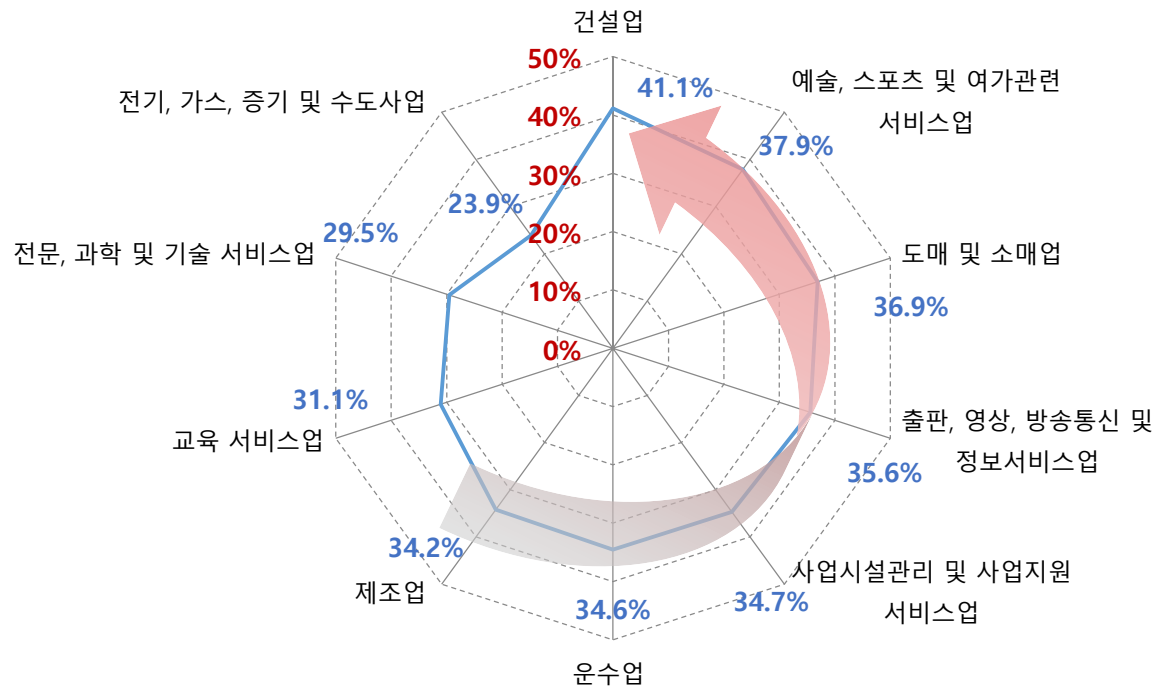
- ✓ 대다수 산업의 위험이 2015~2016년 저점을 기록한 이후, 최근 2년간 증가 추세에 있음
- ✓ 중소형사 중심의 컴퓨터/시스템 관리업과 건설업 부실화 증가폭이 두드러짐

3. 산업별 부실 위험 모니터링(2/2)

3. 산업별 부실 위험 모니터링

산업별 위험량 변화 감지

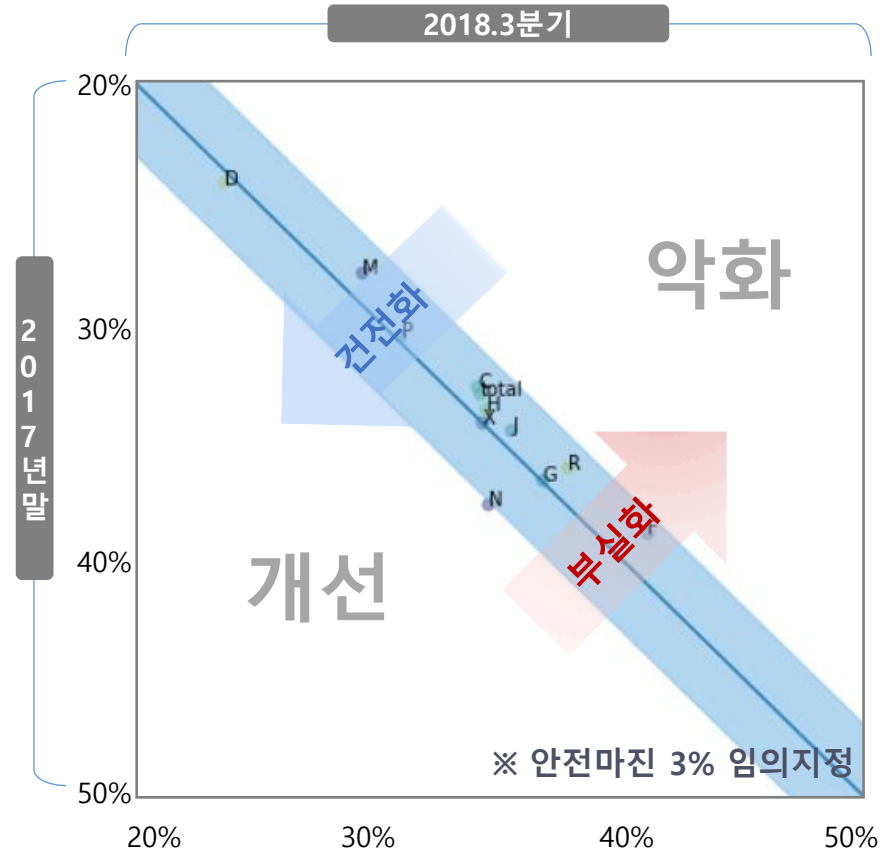
산업 **대분류** 위험량 그래프



2018년 3분기 기준. 반시계방향으로 부실위험이 높음

※ 한국표준산업10차 **대분류** 기준. 10개 종목 미만 산업분류 제외

산업별 부실위험 전환 Matrix



| 코드 | 산업명 | 증가율 |
|-------|-----------|-----|
| Total | 총합계 | 1% |
| F | 건설업 | 2% |
| R | 여가관련 서비스 | 2% |
| M | 기술 서비스업 | 1% |
| C | 제조업 | 1% |
| J | 정보서비스업 | 1% |
| H | 운수업 | 1% |
| P | 교육 서비스업 | 0% |
| G | 도매 및 소매업 | 0% |
| X | 기타 | 0% |
| D | 전기, 가스 등 | -1% |
| N | 사업지원 서비스업 | -3% |

※ 대분류 기준. 전체 종목

4-1. 데이터 구성

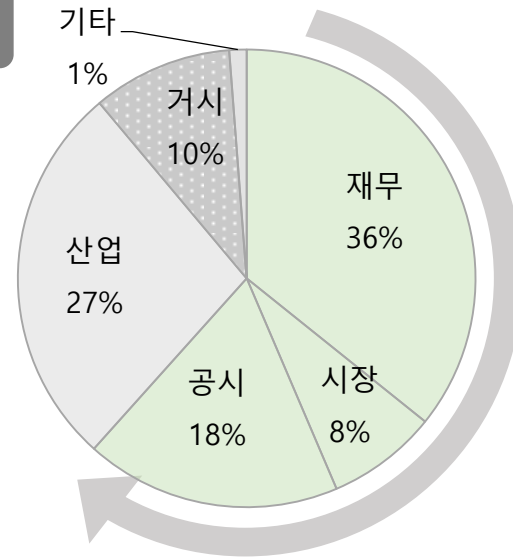
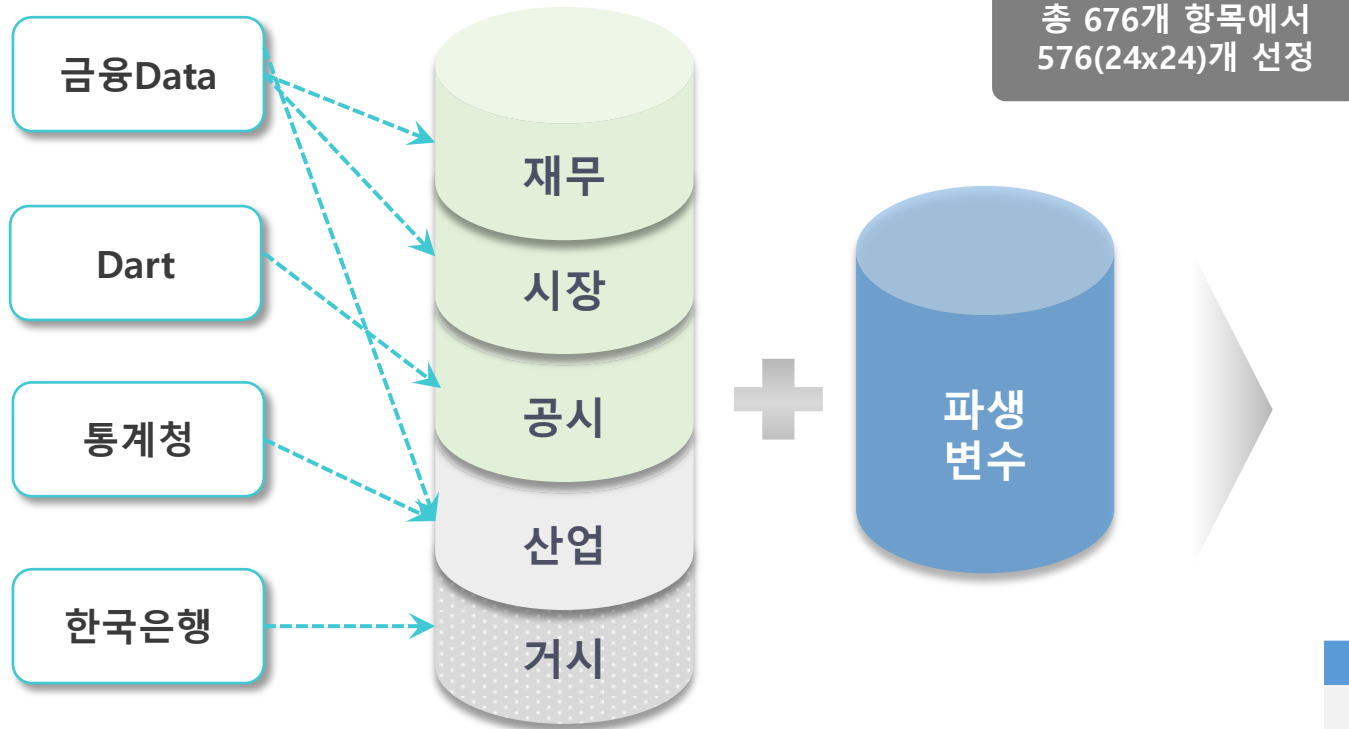
4. 기업 위험 예측 변수 선정

데이터 수집

파생변수 생성

변수 선정 작업

최종 변수 확정



총 576개 항목

기업 관련 : 355개 61%

산업 : 157개 27%

거시 : 57개 10%

기타 : 7개 1%

(재별구분 및 부도기업수 등)

데이터 수집 항목 및 파생변수 생성 근거

- 일반 기업 재무 및 시장 정보
- 각종 논문 및 학술연구자료 참고
- 금융기관 보도자료 참고

주요 검토 항목으로 반영

| 구분 | 원본 | 파생 | 합계 |
|----|------------|------------|------------|
| 재무 | 97 | 109 | 206 |
| 시장 | 35 | 10 | 45 |
| 공시 | 104 | 0 | 104 |
| 산업 | 153 | 4 | 157 |
| 거시 | 28 | 29 | 57 |
| 기타 | 3 | 4 | 7 |
| | 420 | 156 | 576 |

※ 단, 수집 사이트 내 일반 재무/시장 정보의 파생항목은 원본에 포함됨.

※ 공시데이터의 경우, 원본에 항목별 1,2,3개년 누적 포함. 최대 1개 공시당 3항목에 포함 가능

기존 사례 연구를 기반으로 유의 항목 데이터 수집

주요항목 검토

논문/연구

보도자료

기사 등

기타수집

주요항목
검토



파생변수 생성

전년대비

3년평균

3년변동성

파생
변수



누적 공시변수 생성

3년간

2년간

1년간

출처 및 방법론

결과

유의할 것으로 추정되는

91개 항목 검토

최종 86개 항목 모델 반영

특수 변수를 제외하고

항목별 3가지 파생변수 기본 생성 후

모델링을 통해 유의변수 도출

공시 'Title' 기준, 카테고리화 후

최종 54개 카테고리 변수 생성

1,2,3년간 누적 횟수 반영

모델을 100% 표현가능한 수준의 항목 선정

딥러닝 모델(DNN)의 경우 변수별 가중치 확인이 어려우나, 트리유형의 머신러닝 기법 혹은 기타 알고리즘으로 변수 중요도 파악이 가능함.
단, DNN에 미치는 영향을 보장할 수 없으므로 **다양한 방법론을 적용해 포괄적 범위로 선정함.**

※ 핵심변수 추출보다 불필요한 변수를 제거하는 방식으로 접근

