



# 중소형 상장 기업 대상 기업 정보의 동태적 분석을 통한 실질적 부실 기업 예측

“FBI(Financial Bankruptcy Investigation)”

---

# Contents

---

01 주제선정배경

02 Workflow

03 데이터 수집

04 데이터 전처리

05 모델링 및 성능 평가

06 Insight 및 한계점

# FBI



서선우

- 프로젝트 기획
- 메인코더



고원태

- 도메인
- ppt 및 발표

- 서브코더
- ppt 작성



김정우

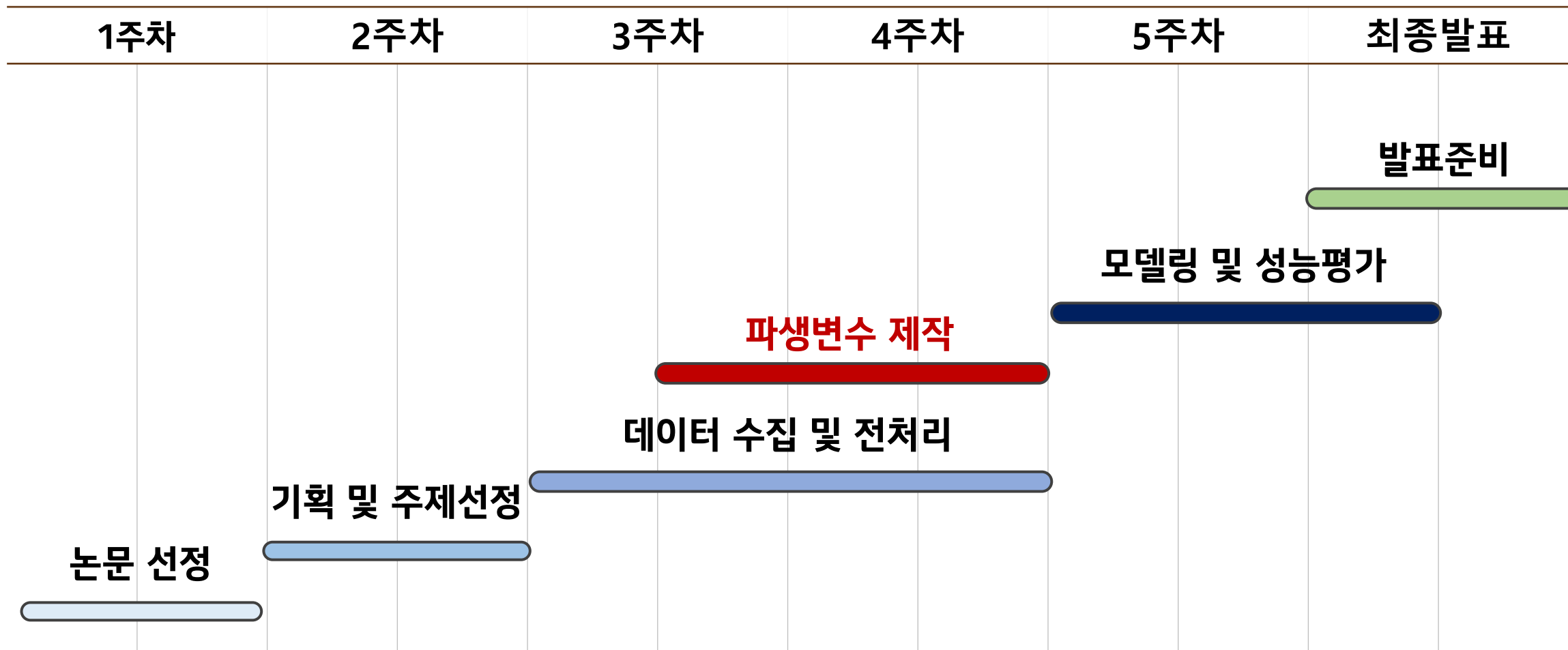
- 메인코더
- 도메인



서수아(리더)

# Time Table

프로젝트 기간 : 3/28 ~ 5/3



## 주제 선정 배경



## 뉴스 기사(23.02.15, 메트로신문)

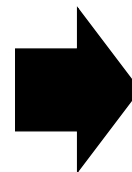
### 관리종목 지정·상장폐지 이슈 증가...투자 주의

[메트로신문] 지난해 실적 마감을 앞두고 관리종목 지정 또는 상장폐지 사유를 공시한 상장 업체들이 최근 증가하고 있다. 특히 실적이 나쁜 상장업체들은 고금리, 경기 침체 등으로 자금 조달 등에 어려움을 겪고 있어 관련 공시 종목들이 늘어날 것으로 예상된다. 이러한 종목들은 주가 급락은 물론 거래정지 및 상장폐지까지 이어질 수 있어 투자자들의 주의가 요구된다.



#### 기사 요약

기업들이 고금리, 경기침체 등의 사유로 관리종목 지정 또는 상장폐지 사유를 공시한 상장 업체들이 증가하는 추세이므로 투자자들의 주의가 요구된다.



**“실질적 부실 기업  
예측 필요성 증대”**

## 기존 논문들의 연구

- 재무정보의 정태적 분석 위주의 연구
- 시장정보 활용 한계(Merton 모형)



## 기존 논문들과의 차별점

- 재무정보의 변화율에 대한 파생변수 활용
- 시장정보를 가미한 파생변수 생성

# 키워드 정의

## 중소형 상장기업

- 시가총액 기준 100위 이내의 **대형주를 제외**
- 연구 기간(11~19년) 중 실질적 부실 사유로 부실 기업이 된 대형 상장 기업은 **2개**

## 동태적 분석

- 부실의 원인을 파악할 때 재무비율의 절대적인 값 보다 **재무비율의 변화 추세에 주목**
- **적시성**을 고려하여 **'시장정보'**를 포함

## 실질적 부실

- 상장폐지, 관리종목 지정 기업들 중 자진 상장폐지, 우회 상장, 인수 및 합병 등 **'자진 상장폐지 사유'는 제외**
- 자본 (부분/완전) 잠식, 감사의견거절은 포함



# 키워드 정리

## 중소형 상장기업

- 시가총액 기준 100위 이내의 **대형주를 제외**
- 분석 기간(11~19년) 중 실질적 부실 사유로 부실 기업이 된 대형 상장 기업은 **2개**

구분	대형주
2011년	0개
2012년	0개
2013년	1개 * STX팬오션(관리종목 지정)
2014년	0개
2015년	0개
2016년	1개 * 포스코플랜택(상장폐지)
2017년	0개
2018년	0개
2019년	0개

# 키워드 정의

## 중소형 상장기업

- 시가총액 기준 100위 이내의 **대형주를 제외**
- 연구 기간(11~19년) 중 실질적 부실 사유로 부실 기업이 된 대형 상장 기업은 **2개**

## 동태적 분석

- 부실의 원인을 파악할 때 재무비율의 절대적인 값 보다 **재무비율의 변화 추세에 주목**  
ex) 1. + 50% -> + 10%  
2. - 10% -> + 5%
- **적시성**을 고려한 '**시장정보**'를 포함

## 실질적 부실

상장폐지, 관리종목 지정 기업들 중 자진 상장폐지, 우회 상장, 인수 및 합병 등 '**자진 상장폐지 사유**'는 제외

자본 (부분/완전) 잠식, 감사의견거절은 포함

# 키워드 정의

## 중소형 상장기업

- 시가총액 기준 100위 이내의 **대형주를 제외**
- 연구 기간(11~19년) 중 실질적 부실 사유로 부실 기업이 된 대형 상장 기업은 **2개**

## 동태적 분석

- 부실의 원인을 파악할 때 재무비율의 절대적인 값 보다 **재무비율의 변화 추세에 주목**
- 적시성을 고려하여 **'시장정보'**를 포함

## 실질적 부실

- 상장폐지, 관리종목 지정 기업들 중 자진 상장폐지, 우회 상장, 인수 및 합병 등 **'자발적 상장폐지 사유'는 제외**
- **자본(부분/완전)잠식, 감사의견거절 포함**

## 연구 주제

중소형 상장 기업 대상  
기업 정보의 동태적 분석을 통한 실질적 부실 기업 예측

# Workflow



## 데이터 수집

### 분석 범위 설정

- 수집 기간: 2010 ~ 2019
- 종목 범위: KOSPI + KOSDAQ
- 제외 종목: 금융기업, 스펙, 외국계 기업, 시가총액 상위 100위 이내 기업
- 수집처: TS2000

### 수집 데이터

- 기업 수: 2,487사
- 데이터 개수: 20,505개
- 피쳐 개수: 134개  
-> 재무정보 + 시장정보

### 연구 키워드

- 중소형 상장 기업
- 동태적 분석
- 실질적 부실 기업

### 후보 Feature(27)

수익성(8), 안정성(6)  
생산성(2), 성장성(6),  
기타(2), Altman(3)

## 데이터 전처리

### 후보 파생변수(12)



### 데이터 확인

- EDA(기초통계량, 데이터 분포)
- 결측치 처리
- 이상치처리: IQR, Winsorizing

### Data Split

- Train Set : 2011 - 2017
- Test Set : 2018 - 2019

### Scaling

- Min-Max-Scaling
- Standard Scaling

### Resampling

- SMOTE
- BorderlineSMOTE
- 1:1, 2:1, 4:1, 10:1

### 피쳐선정

- T-test(24)
- LASSO(14)
- Step-wise(10)

- 최종 피쳐  
-> 2번 중복 : 15개  
-> 3번 중복 : 9개

Pipeline

### 통계검정

- 등분산성 : Bartlett
- 정규성 : Shapiro-Wilks

## 모델링

### 사용 모델

- Logistic Regression
- SVM
- Decision Tree
- Random Forest
- XGBoost
- Stacking
- Tabnet

### 하이퍼파라미터 조정

- GridSearchCV
- RandomizedSearchCV

### 성능 평가

- Accuracy
- Precision
- Recall
- F1 Score
- ROC-AUC

## 데이터 수집



# 데이터 수집

대상	KOSPI & KOSDAQ 상장 기업
종류	재무정보 + 시장정보
제외	금융, 외국계, SPAC 기업, <b>시가총액 상위 100위 내 기업(대형주)</b>
기간	<b>수집 : 2010~2019 / 분석 2011~2019</b>
개수	2,478사 / 20,505개 / 134개 피쳐
수집처	TS2000



# 데이터 전처리



# Raw\_data 작업

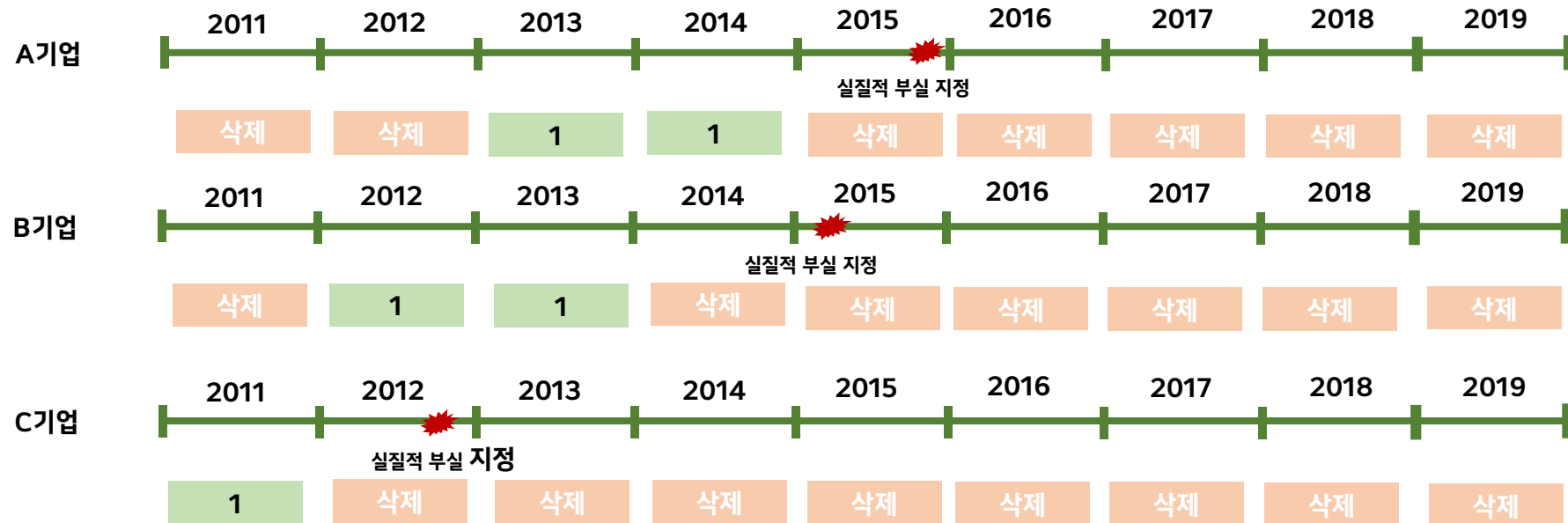
결측치 처리

타겟 설정

논문 공통 변수

이상치 제거

대형주 삭제



Y = 0(정상)

✓ 상장폐지 전년도(t-1) 및 그 전전년도(t-2) 데이터를 부실로 간주

Y = 1(부실)

✓ 상장폐지 해당 연도(t) 이후 및 그 전년도(t-1) 또는 전전년도(t-2) 이전의 데이터는 삭제

\* 상장폐지가 4월 전에 이루어졌다면 전년도의 데이터가 없기에 t-2기, t-3기의 데이터 사용

\* 관리종목의 경우 최초 지정된 년도를 부실 발생 년도로 판단)

# Raw\_data 작업

결측치 처리

타겟 설정

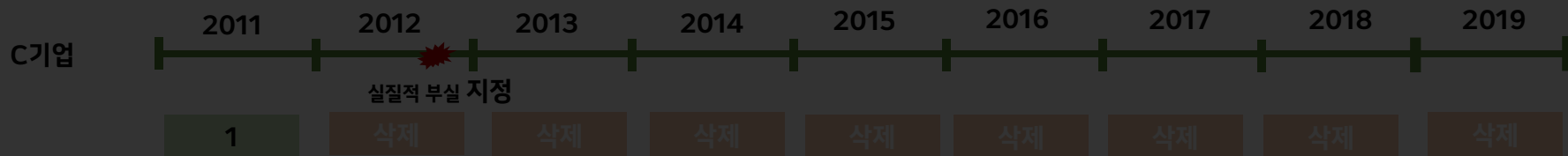
논문 공통 변수

이상치 제거

대형주 삭제



- \* 연결 재무데이터 결측치를 개별 재무 데이터 값으로 대체
- \* 대체 할 수 없는 값들은 삭제



Y = 0(정상)

✓ 상장폐지 전년도( $t-1$ ) 및 그 전전년도( $t-2$ ) 데이터를 부실로 간주

Y = 1(부실)

✓ 상장폐지 해당 연도( $t$ ) 이후 및 그 전년도( $t-1$ ) 또는 전전년도( $t-2$ ) 이전의 데이터는 삭제

\* 상장폐지가 4월 전에 이루어졌다면 전년도의 데이터가 없기에  $t-2$ 기,  $t-3$ 기의 데이터 사용

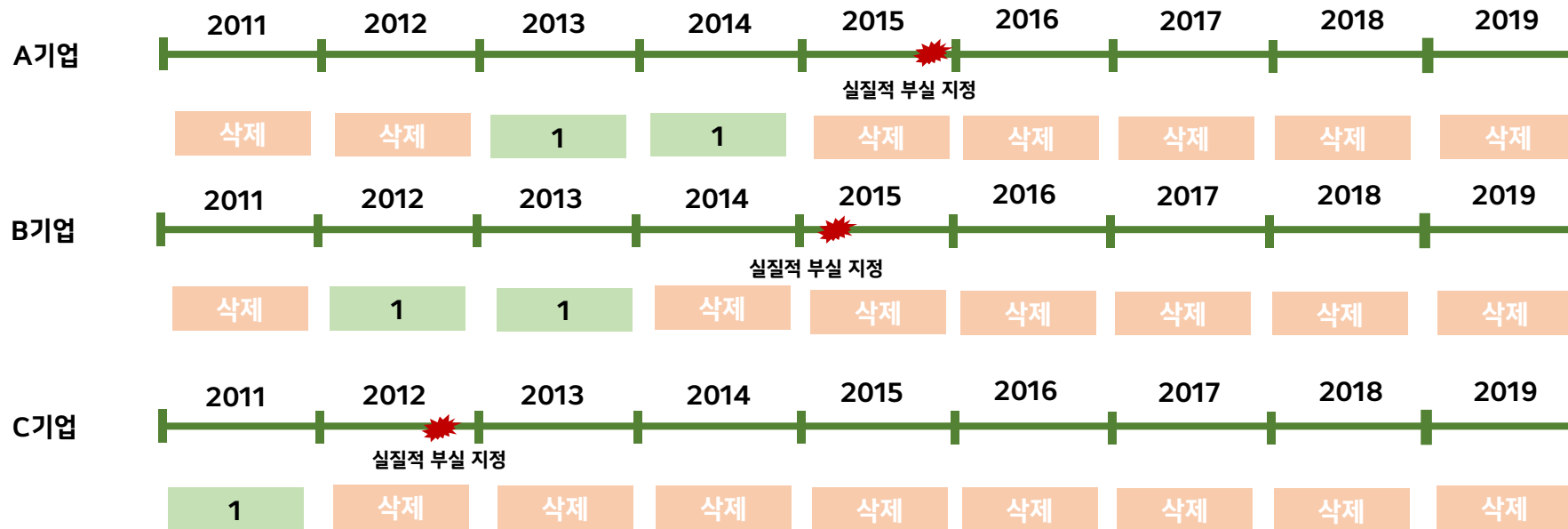
\* 관리종목의 경우 최초 지정된 년도를 부실 발생 년도로 판단)

# Raw\_data 작업

결측치 처리

타겟 설정

논문 공통 변수



Y = 0(정상)

✓ 상장폐지 전년도(t-1) 및 그 전전년도(t-2) 데이터를 부실로 간주

Y = 1(부실)

✓ 상장폐지 해당 연도(t) 이후 및 그 전년도(t-1) 또는 전전년도(t-2) 이전의 데이터는 삭제

\* 상장폐지가 4월 전에 이루어졌다면 전년도의 데이터가 없기에 t-2기, t-3기의 데이터 사용

\* 관리종목의 경우 최초 지정된 년도를 부실 발생 년도로 판단

\* 관리종목의 경우 최초 지정된 년도를 부실 발생 년도로 판단)

# Raw\_data 작업

결측치 처리

타겟 설정

논문 공통 변수(27개)

이상치 제거

대형주 삭제

안정성(6)	수익성(8)	성장성(6)	생산성(2)	기타(2)	Altman(3)
유동비율	총자본경상이익률	매출액증가율	설비투자효율	영업활동현금흐름	EBTA
당좌비율	매출액영업이익률	총자산증가율	총자본투자효율	EV/EBITDA	RETA
부채비율	매출액순이익률	순이익증가율			OM
자기자본비율	자기자본순이익률	영업이익증가율			
이자보상배율	매출채권회전율	유형자산증가율			
순운전자본비율	재고자산회전율	유동자산증가율			
	총자본회전율				
	영업활동현금흐름				

Y = 1(부실)

\* 상장폐지가 4월 전에 이루어졌다면 전년도에 데이터가 없기에 t-2기, t-3기의 데이터 사용

\* 관리종목의 경우 최초 지정된 년도를 부실 발생 년도로 판단)

# Raw\_data 작업

결측치 처리

타겟 설정

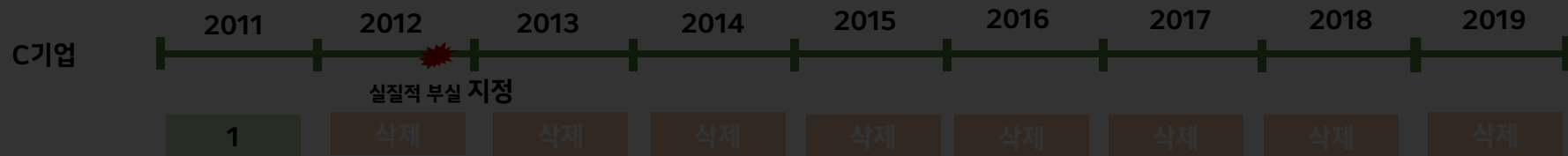
논문 공통 변수

이상치 제거

대형주 삭제



\* 연도별 시가총액 상위 100위 이내 기업 삭제



Y = 0(정상)

✓ 상장폐지 전년도(t-1) 및 그 전전년도(t-2) 데이터를 부실로 간주

Y = 1(부실)

✓ 상장폐지 해당 연도(t) 이후 및 그 전년도(t-1) 또는 전전년도(t-2) 이전의 데이터는 삭제

\* 상장폐지가 4월 전에 이루어졌다면 전년도의 데이터가 없기에 t-2기, t-3기의 데이터 사용

\* 관리종목의 경우 최초 지정된 년도를 부실 발생 년도로 판단)

# Raw\_data 작업

결측치 처리

타겟 설정

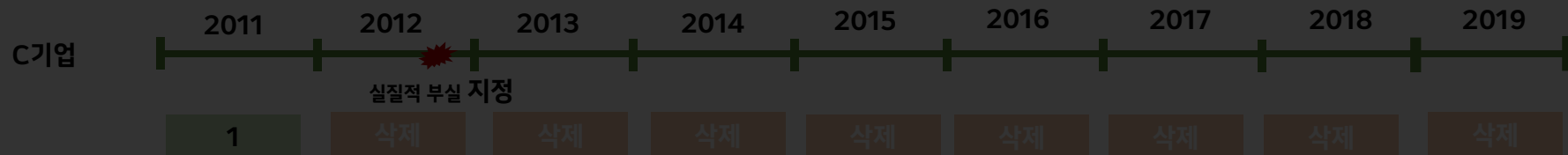
논문 공통 변수

이상치 제거

대형주 삭제



\* IQR 0.5이하 99.5이상 총 1% 제거



Y = 0(정상)

✓ 상장폐지 전년도(t-1) 및 그 전전년도(t-2) 데이터를 부실로 간주

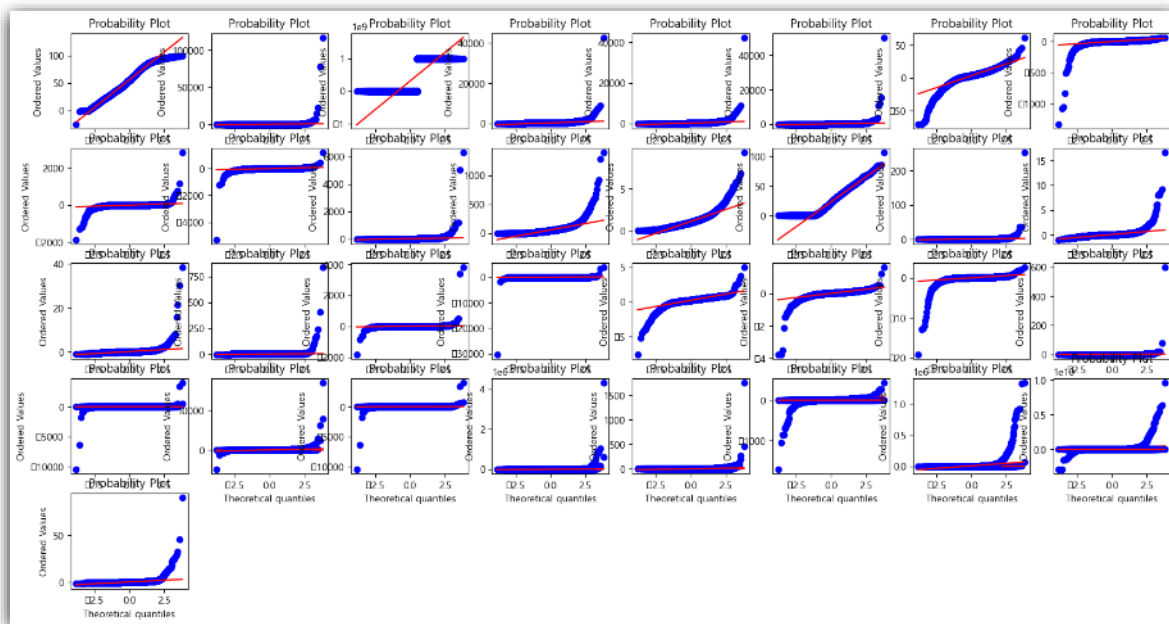
Y = 1(부실)

✓ 상장폐지 해당 연도(t) 이후 및 그 전년도(t-1) 또는 전전년도(t-2) 이전의 데이터는 삭제

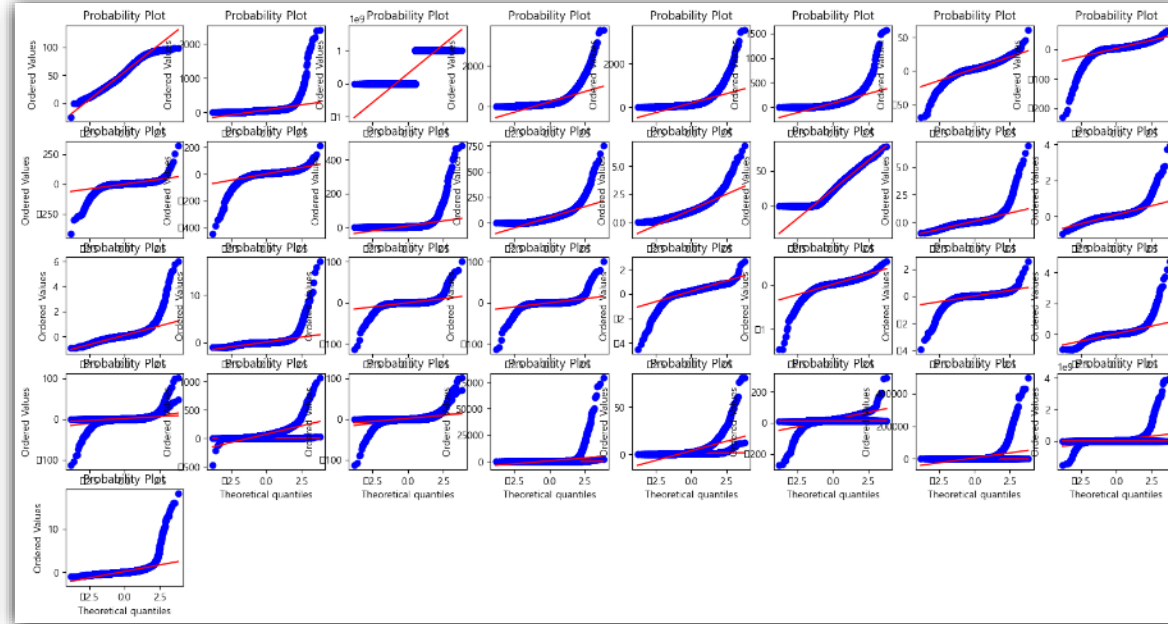
\* 상장폐지가 4월 전에 이루어졌다면 전년도의 데이터가 없기에 t-2기, t-3기의 데이터 사용

\* 관리종목의 경우 최초 지정된 년도를 부실 발생 년도로 판단)

# EDA(Q-Q plot)



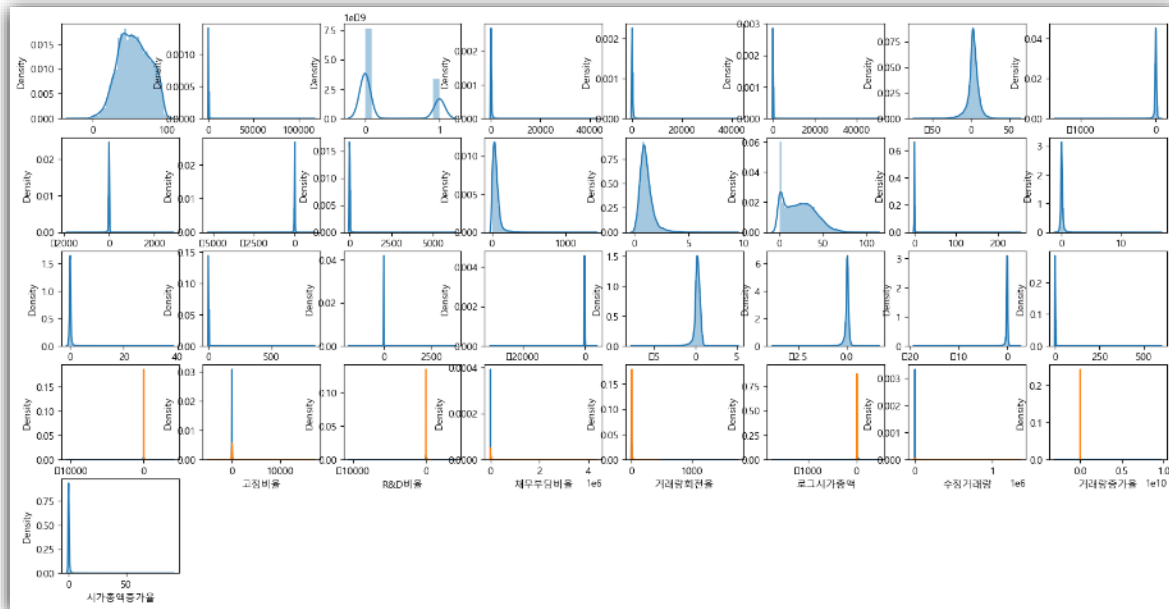
이상치 제거 전



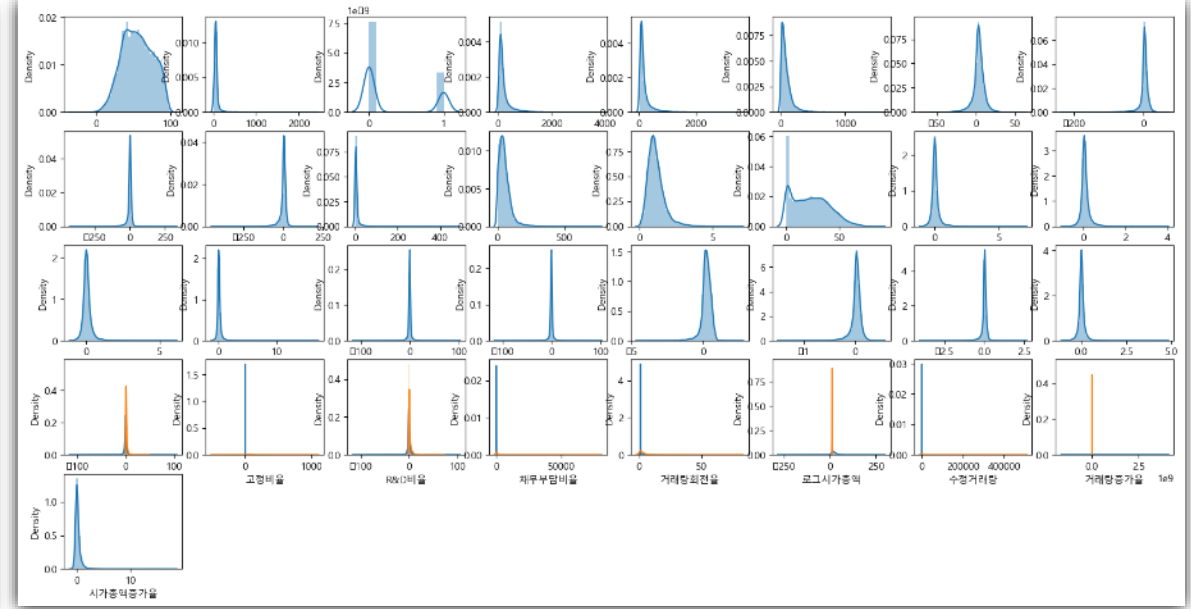
이상치 제거 후



# EDA(Distplot)



이상치 제거 전



이상치 제거 후

## 전체 데이터 비율



정상기업 수 : 1,354사

데이터 개수 : 9,428개

부실기업 수 : 231사

데이터 개수 : 367개

데이터 비율

**25 : 1**

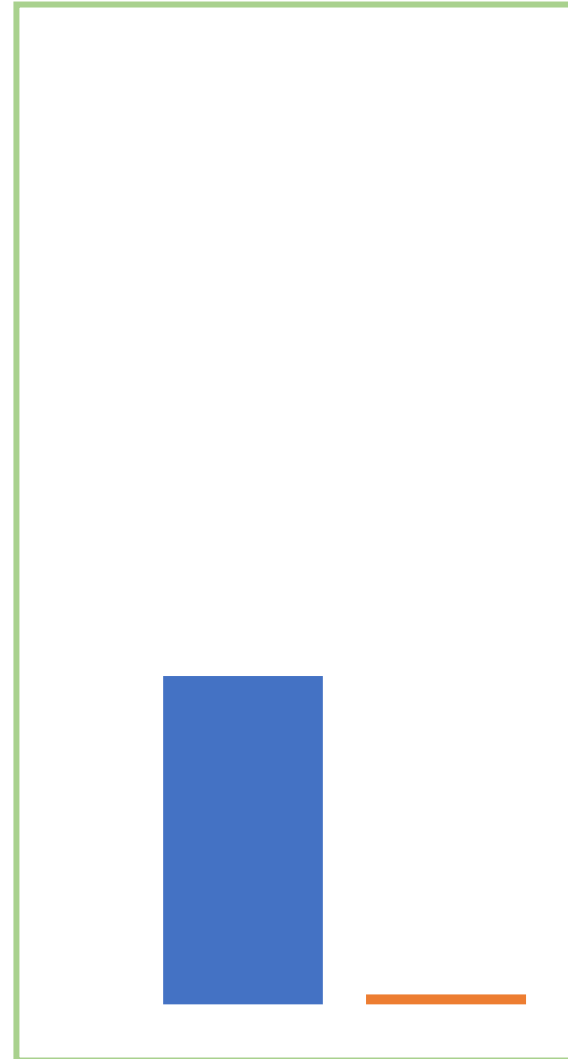
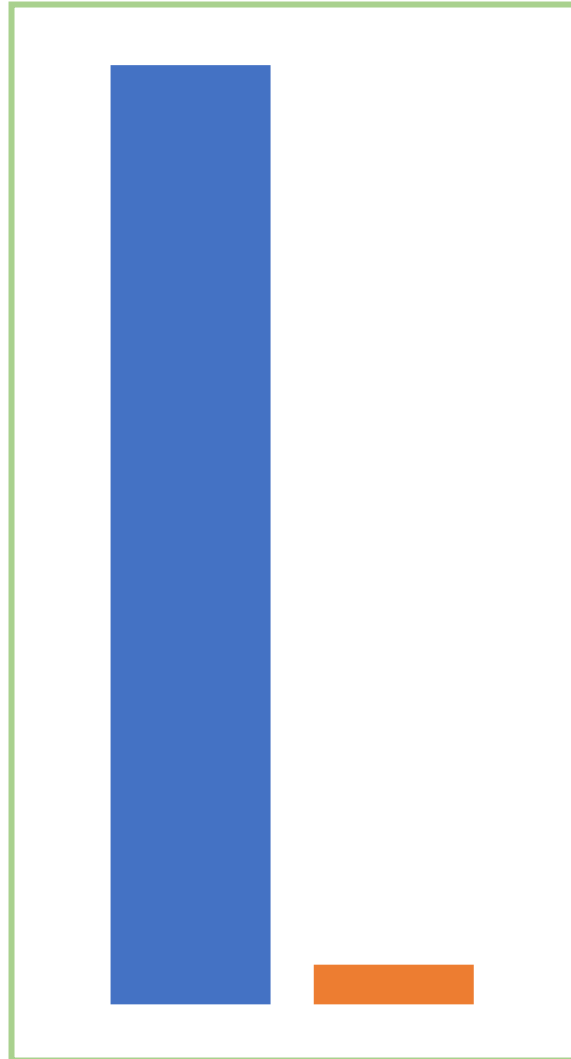
# Data Split



## Data Split 이후 데이터 비율

Train set  
정상 : 7,242개  
부실 : 336개

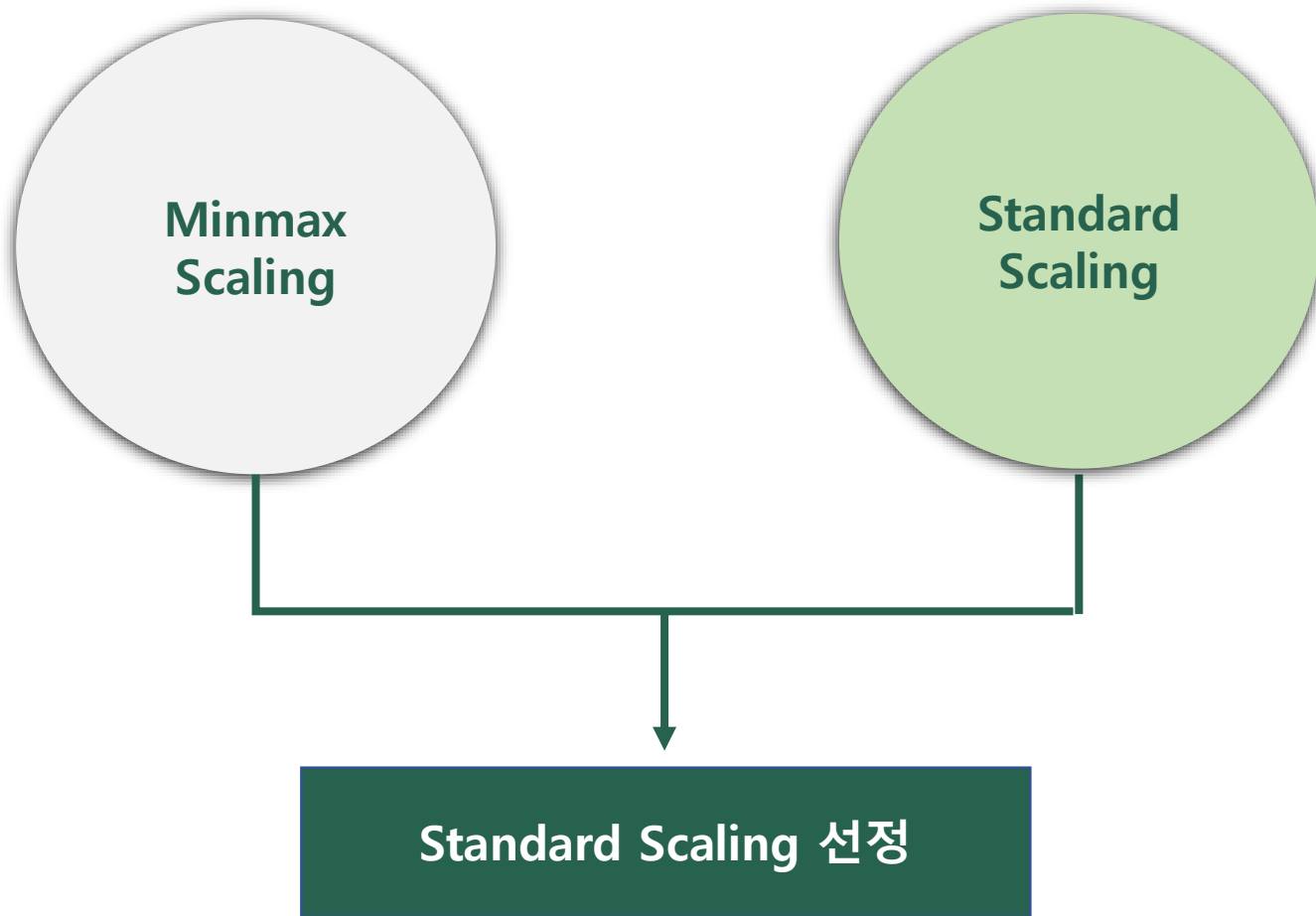
데이터 비율  
**22 : 1**



Test set  
정상 : 2,515개  
부실 : 87개

데이터 비율  
**29 : 1**

# Scaling(2가지로 각각 진행)



# Resampling(Oversampling)

방법

SMOTE

BorderlineSMOTE

비율

1:1

2:1

4:1

10:1

## 파생변수 생성 & Feature Selection



## 재무정보의 변화율을 이용한 파생변수(7)

1. DOL : 영업이익변화율 / 매출액변화율
2. DFL : 주당순이익변화율 / 영업이익변화율
3. 고정비율 : (유형자산/총자산) X 100(%)
4. R&D비율 : (연구개발비/매출액) X 100(%)
5. 종업원(합계기말인원) 수 증가율  
: (T-1기 종업원수 - T기 종업원수 / T-1기 종업원수) X 100(%)
6. 금융비용 부담률 : (이자비용/매출액) X 100(%)
7. 채무부담비율 : (재고자산회전율/매입채무회전율) X 100(%)

## 시장정보를 이용한 파생변수(5)

1. 로그 시가총액 :  $\text{np.log10}(\text{시가총액})$
2. 시가총액증가율  
: (T-1기 시가총액 - T기 시가총액 / T-1기 시가총액) X 100(%)
3. 거래량 증가율  
: (T-1기 1년거래량 - T기 1년거래량 / T-1기 1년거래량) X 100(%)
4. 거래량 회전율 : 1년거래량 / 발행주식 총수
5. 수정 종가 거래량 : 1년거래량 / 수정종가



## 재무정보의 변화율을 이용한 파생변수

\* DOL : 영업 레버리지(Degree of Operation Leverage)

$$= \text{영업이익변화율} / \text{매출액변화율} \rightarrow \text{민감도}$$

\* DFL : 재무 레버리지(Degree of Financial Leverage)

$$= \text{주당순이익변화율} / \text{영업이익변화율} \rightarrow \text{민감도}$$

\* 고정비율 : (유형자산/총자산) X 100(%)

\* R&D비율 : (연구개발비/매출액) X 100(%)

\* 종업원수 증가율

$$: (\text{T-1기 종업원수} - \text{T기 종업원수} / \text{T-1기 종업원수}) \times 100(\%)$$

\* 금융비용 부담률 : (이자비용/매출액) X 100(%)

\* 채무부담비율 : (재고자산회전율/매입채무회전율) X 100(%)

$$= ((\text{매출액} / \text{재고자산}) / (\text{매출액} / \text{매입채무})) \times 100(\%)$$

$$= (\text{매입채무} / \text{재고자산}) \times 100(\%)$$

## 시장정보를 이용한 파생변수

\* 로그 시가총액 :  $\text{np.log10}(\text{시가총액})$

\* 시가총액 증가율

:  $(T-1\text{기 시가총액} - T\text{기 시가총액} / T-1\text{기 시가총액}) \times 100(\%)$

\* 거래량 증가율

:  $(T-1\text{기 1년 거래량} - T\text{기 1년 거래량} / T-1\text{기 1년 거래량}) \times 100(\%)$

\* 거래량 회전율 : 1년 거래량 / 발행주식 총수

\* 수정 종가 거래량 : 1년 거래량 / 수정 종가

- 수정 종가 =  $\frac{(\text{기초 종가} \times \text{기초가중치}) + (\text{기말 종가} \times \text{기말가중치})}{2}$

- 가중치 => 액면가 고려(100원, 1,000원, 5,000원 등)

- 기초 가중치 => 5,000 / 기초 액면가

- 기말 가중치 => 5,000 / 기말 액면가

# 피처 선정 결과

## T-test(24)

DOL, EBTA, OM, RETA,  
거래량회전율, 고정비율, R&D비율,  
금융비용부담률, 당좌비율,  
로그시가총액, 매출액순이익률,  
매출액정상영업이익률, 부채비율  
매출액증가율, 수정거래량,  
순운전자본비율, 유동비율, 총자본투자효율,  
자기자본구성비율, 재고자산회전률,  
자기자본순이익률, 총자본회전률  
총자본정상영업이익률, 총자본증가율,

## Stepwise(10)

EBTA, RETA, 부채비율,  
수정거래량, 총자본증가율,  
자기자본순이익률,  
총자본회전률, 로그시가총액,  
총자본투자효율,  
재고자산회전률

## LASSO(14)

부채비율, RETA, EBTA, OM,  
총자본정상영업이익률,  
매출액정상영업이익률,  
자기자본순이익률, 고정비율  
재고자산회전률, 수정거래량,  
총자본회전률, 금융비용부담률,  
거래량회전률, 로그시가총액

## 최종 피처(9)

[안정성] 부채비율, [수익성] 자기자본순이익률, 총자본회전율, RETA, EBTA,  
[파생변수](4) 금융비용부담률, 로그시가총액, 수정증가거래량, 거래량회전율

# 피쳐 선정 결과

T-test(24)

Stepwise(10)

LASSO(14)

DOL, EBTA, OM, RETA,  
거래량회전은, 고정비용, D&D비용

개별 피쳐 선정 모델링과 공통 피쳐 모델링 간의 성능차이가 크지 않아  
최종 피쳐를 공통 피쳐로 선정

자기자본구성비율, 재고자산회전률,  
자기자본순이익률, 총자본회전률  
총자본정상영업이익률, 총자본증가율,

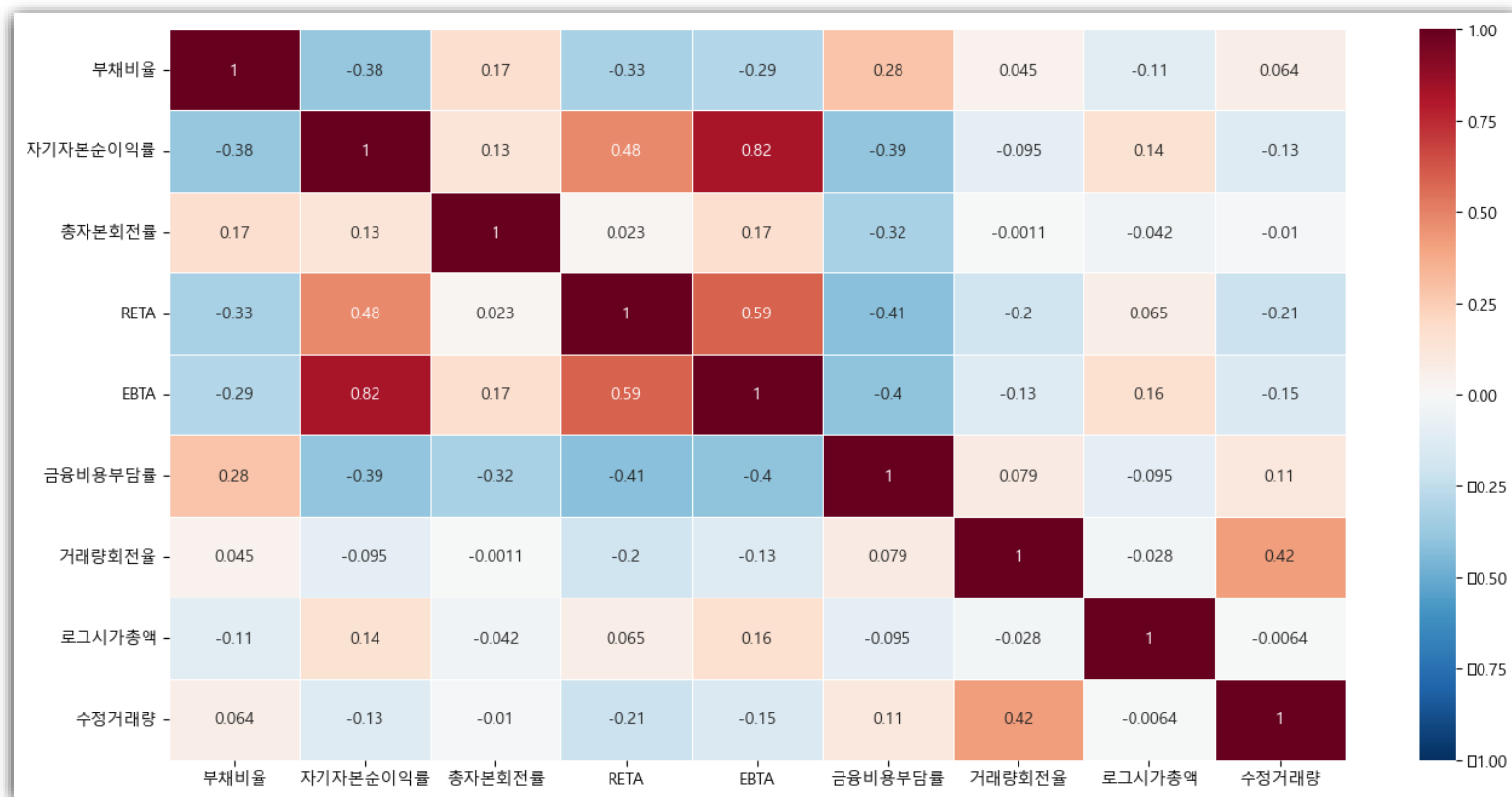
재고자산회전률

거래량회전률, 로그시가총액

최종 피쳐(9)

[안정성] 부채비율, [수익성] 자기자본순이익률, 총자본회전율, RETA, EBTA,  
[파생변수](4) 금융비용부담률, 로그시가총액, 수정증가거래량, 거래량회전율

# Heatmap 및 다중공선성



VIF	
EBTA	3.728471
자기자본순이익률	3.362430
RETA	1.755549
금융비용부담률	1.501646
부채비율	1.350181
총자본회전률	1.256378
수정거래량	1.242278
거래량회전율	1.232342
로그시가총액	1.000070

## 정상, 부실기업의 기초통계량

정상 (9716개)	부채비율	ROE	재고자산 회전률	총자본 회전률	RETA	EBTA	금융비용 부담률	거래량 회전율	로그 시가총액	수정 거래량(주)
평균	87.78	1.74	52.54	1.12	23.46	2.37	161.13	364.42	11.93	6862
표준편차	532.79	29.08	51.91	0.59	35.76	9.66	544.73	718.51	0.49	36564
최솟값	0	-1224.59	0	0	-372.29	-146.26	0	0	10.01	0
25%	27.78	-1.06	23.05	0.75	9.06	-0.24	25.93	81.08	11.65	150
50%	57.39	4.72	43.36	1.02	25.02	3.07	83.69	188.99	11.94	619
75%	105.75	9.95	69.66	1.37	43.41	6.96	180.04	416.54	12.25	2661
최댓값	50259.95	1191.98	1394.72	6.36	493.88	161.01	44818.34	45703.56	13.01	1345120
부실 (420개)	부채비율	ROE	재고자산 회전률	총자본 회전률	RETA	EBTA	금융비용 부담률	거래량 회전율	로그 시가총액	수정 거래량(주)
평균	306.48	-72.83	71.59	0.92	-33.62	-20.69	539.98	718.66	11.75	28895
표준편차	988.69	273.96	90.08	0.74	83.79	38.58	778.93	1378.02	0.49	96237
최솟값	0	-5300.39	0.02	0.04	-761.82	-376.92	0	0	10.11	0
25%	52.94	-83.03	22.17	0.53	-44.16	-26.94	134.03	177.90	11.47	860
50%	116.00	-34.66	47.33	0.77	-13.55	-12.31	346.37	409.95	11.80	4081
75%	263.60	-5.01	81.28	1.15	6.80	-2.30	623.14	750.98	12.05	20610
최댓값	15340.84	57.66	860.79	9.24	96.34	32.72	6514.09	20208.36	12.93	1329816

# 모델링 및 성능 평가



# 하이퍼파라미터 조정 전(default)

	Train set					Test set				
	Accuracy	Precision	Recall	F1-Score	AUC	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.96	0.63	0.19	0.30	0.59	0.97	0.43	0.12	0.19	0.56
SVM	0.97	0.99	0.35	0.51	0.68	0.97	0.53	0.14	0.22	0.57
Decision Tree	1.0	1.0	1.0	1.0	1.0	0.94	0.20	0.30	0.24	0.63
Random Forest	1.0	1.0	1.0	1.0	1.0	0.97	0.46	0.16	0.24	0.58
XGBoost	1.0	1.0	1.0	1.0	1.0	0.97	0.36	0.16	0.23	0.58
Stacking	0.96	0.65	0.22	0.33	0.61	0.97	0.5	0.16	0.25	0.58
TabNet	0.82	0.76	0.95	0.84	0.89	0.70	0.09	0.85	0.17	0.79



# 모델별 Hyperparameter

**Logistic  
Regression**

penalty=l1, solver=liblinear, C=0.000695, max\_iter=200,  
class\_weight=balanced

**SVM**

kernel=rbf, gamma=1, C=0.1, class\_weight=balanced

**Decision Tree**

criterion=log\_loss, max\_depth=30, max\_features=0.1,  
class\_weight=balanced

**Random  
Forest**

n\_estimators : 80, criterion : gini, max\_depth : 5, max\_features : sqrt,  
min\_samples\_split : 5

**XGBoost**

learning\_rate=0.05, n\_estimators=300,  
scale\_pos\_weight=scale\_pos\_weight\_ss, min\_child\_weight=4, subsample=0.5,  
colsample\_bytree=0.6, max\_depth=7,  
gamma=1, reg\_alpha=0.8, reg\_lambda=0

## 하이퍼파라미터 조정 후

	Train set					Test set				
	Accuracy	Precision	Recall	F1-Score	AUC	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.87	0.20	0.76	0.31	0.81	0.84	0.12	0.73	0.21	0.79
SVM	0.83	0.19	0.98	0.31	0.90	0.71	0.08	0.88	0.15	0.79
Decision Tree	0.97	0.49	0.40	0.51	0.70	0.96	0.31	0.22	0.26	0.60
Random Forest	0.97	0.93	0.28	0.43	0.64	0.97	0.41	0.10	0.16	0.55
XGBoost	0.99	0.86	1.0	0.92	0.99	0.96	0.38	0.53	0.44	0.75
Stacking	0.96	0.53	0.10	0.17	0.55	0.83	0.12	0.77	0.21	0.80
TabNet	0.82	0.76	0.95	0.84	0.89	0.70	0.09	0.85	0.17	0.79

## 오버샘플링(1:1)

	Train set					Test set				
	Accuracy	Precision	Recall	F1-Score	AUC	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.80	0.85	0.74	0.79	0.80	0.83	0.12	0.73	0.20	0.78
SVM	0.88	0.81	0.99	0.89	0.88	0.67	0.08	0.90	0.14	0.79
Decision Tree	0.83	0.87	0.79	0.82	0.83	0.83	0.11	0.68	0.19	0.76
Random Forest	0.84	0.87	0.80	0.83	0.84	0.83	0.12	0.78	0.21	0.81
XGBoost	0.95	0.91	1.0	0.95	0.95	0.65	0.07	0.92	0.13	0.77
Stacking	0.93	0.89	0.99	0.94	0.93	0.86	0.15	0.79	0.25	0.83
TabNet	0.82	0.76	0.95	0.84	0.89	0.70	0.09	0.85	0.17	0.79

## 오버샘플링(1:1)

	Train set					Test set				
	Accuracy	Precision	Recall	F1-Score	AUC	Accuracy	Precision	Recall	F1-Score	AUC
Logistic	0.80	0.85	0.74	0.79	0.80	0.83	0.12	0.73	0.20	0.78

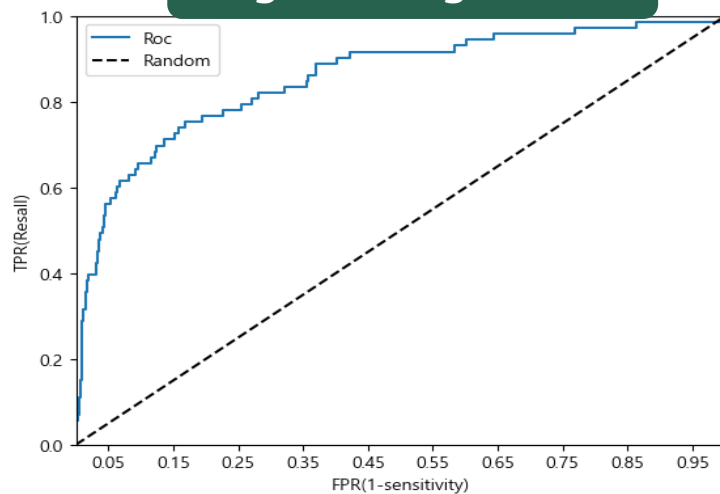
### Recall(재현율) 주목 이유

**: 실제 부실 기업 중에서 모델이 부실 기업이라고 예측하는 것이 중요**  
(2종 오류를 피하기 위해서)

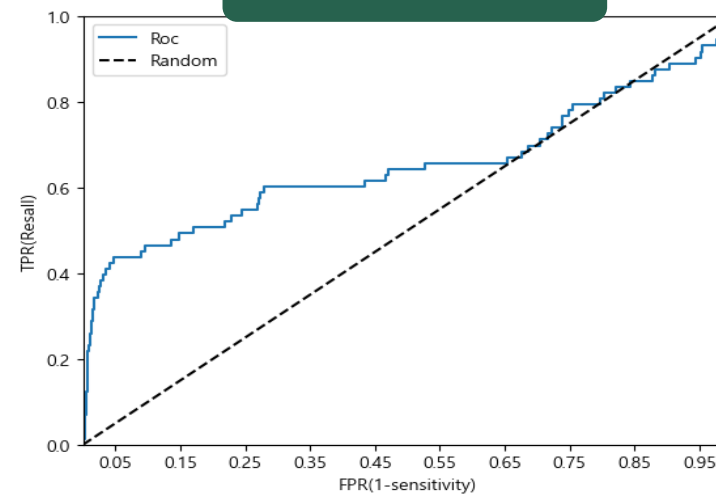
Random Forest	0.84	0.87	0.80	0.83	0.84	0.83	0.12	0.78	0.21	0.81
XGBoost	0.95	0.91	1.0	0.95	0.95	0.65	0.07	0.92	0.13	0.77
Stacking	0.93	0.89	0.99	0.94	0.93	0.86	0.15	0.79	0.25	0.83
TabNet	0.82	0.76	0.95	0.84	0.89	0.70	0.09	0.85	0.17	0.79

# ROC - AUC 커브

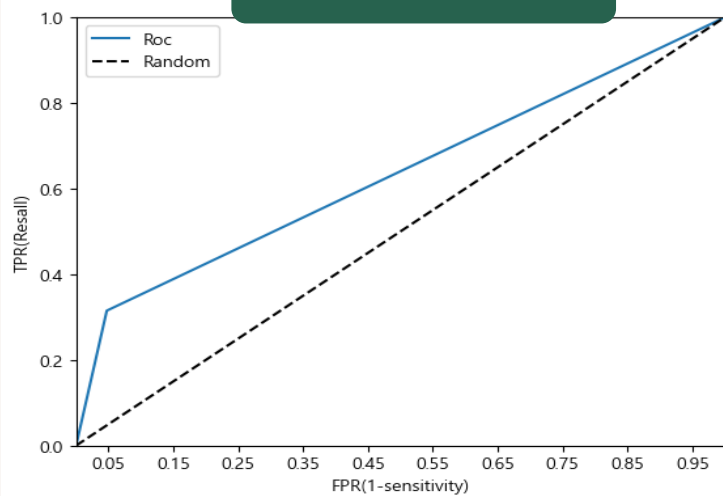
## Logistic Regression



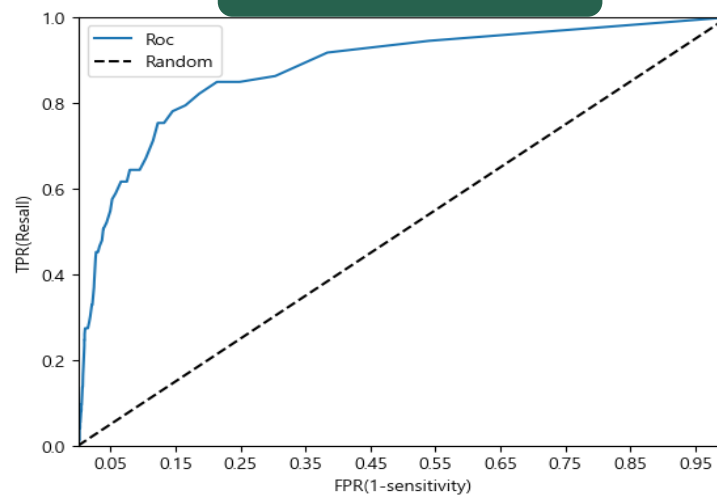
## SVC



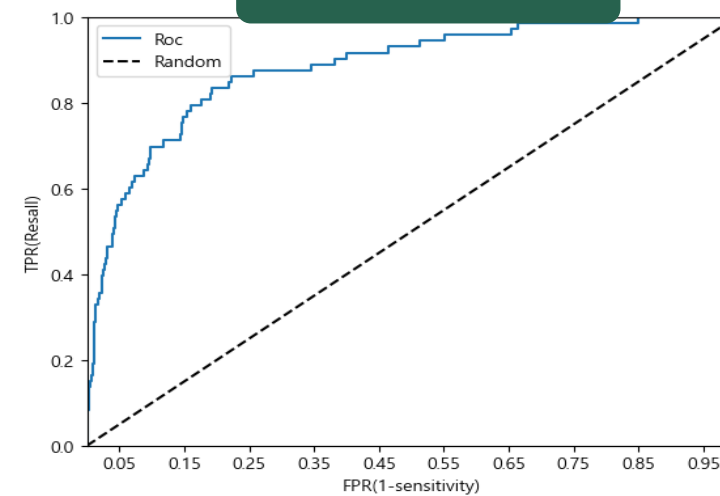
## Decision Tree



## Random Forest

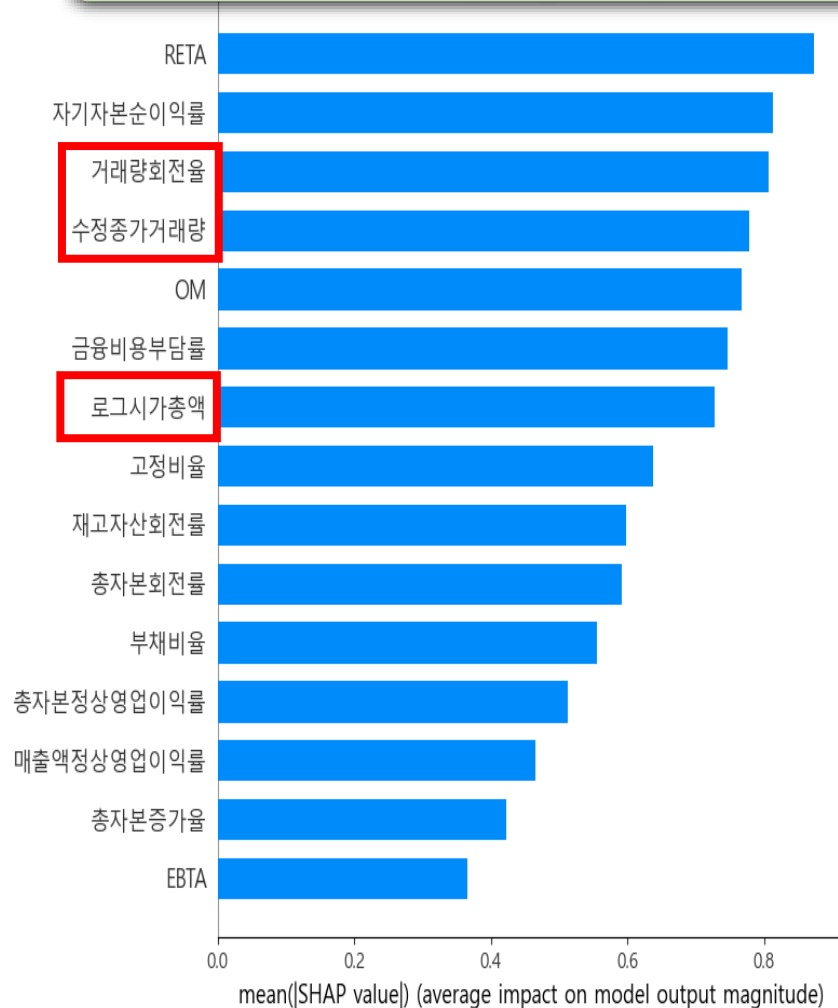


## XGBoost



# Feature 중요도

## XGBoost : Shapley Importance



## TabNet

자기자본순이익률, 금융비용부담률,  
EBTA, 수정증가거래량(3),  
RETA, 매출액순이익률(2),  
부채비율, 총자본투자효율, 총자본회전율, 거래량회전  
율, 로그시가총액, EV/EBITDA, 유동비율, 당좌비율,  
OM,  
이자보상배율, 순운전자본비율,  
당기순이익변화율, 총자본정상영업이익률, 당기순이익  
변화율, 종업원수증가율(1)

파생변수들의 중요도가 높음을 확인

## Insight 및 한계점



# 인사이트 및 기여점

1. 재무비율의 변화율과 시장정보를 이용한 동태적 분석 진행

2. 거래량을 이용한 파생변수

1) 각각의 Feature Selection 과정에서 공통적으로 선택

2) TabNet과 Shapley Importance에서 중요 Feature로 선정

3. 파생변수를 이용한 실질적 부실기업과 거래량의 관계 파악

-> 투자자들에게 간접적인 기여가 가능한 파생변수임을 확인



# 정상기업과 부실기업의 거래량 비교분석

정상기업	수정증가거래량 (주)	거래량증가율 (%)	거래량회전율 (%)
개수	9693	9693	9693
평균	5908.76	80.69	353.74
표준편차	23290.97	339.60	514.59
최솟값	0	-100.00	0
25%	149.65	-41.61	81.01
50%	615.01	-4.80	188.56
75%	2628.37	68.19	413.81
최댓값	542114.18	6100.63	8103.75

부실기업	수정증가거래량 (주)	거래량증가율 (%)	거래량회전율 (%)
개수	414	414	414
평균	21045.52	153.42	637.48
표준편차	50161.86	494.15	872.59
최솟값	0	-100	0
25%	782.50	-42.65	177.79
50%	3974.97	6.67	399.89
75%	18221.83	139.57	728.69
최댓값	498022.25	5158.33	7600.48

# 정상 및 부실기업의 거래량 파생변수 비교분석

정상기업	수정증가거래량 (주)	거래량증가율 (%)	거래량회전율 (%)	부실기업	수정증가거래량 (주)	거래량증가율 (%)	거래량회전율 (%)
개수	9693	9693	9693	개수	414	414	414
평균	5908.76	80.69	353.74	평균	21045.52	153.42	637.48

\* 정상기업이 거래량이 많고 부실기업이 거래량이 적을 것이라 예상

\* 실제 결과는 부실기업의 거래량이 더 많음을 데이터로 확인

이유는..?

50%	615.01	-4.80	188.56	50%	3974.97	6.67	399.89
75%	2628.37	68.19	413.81	75%	18221.83	139.57	728.69
최댓값	542114.18	6100.63	8103.75	최댓값	498022.25	5158.33	7600.48

# 정리매매기간에서 실질적 부실기업과 거래량의 관계

<표 III-2> 상장폐지 전후 유동성 및 변동성 비교 출처 : 강소현, 2020

		상장폐지 전 1년 (정리매매 제외)	정리매매	상장폐지 후 3개월
종목 수(개)		52		
실거래일수(일)	평균값	123(52%)	7(100%)	28(46%)
	중앙값	127(53%)	7(100%)	31(50%)
주가(원)	평균값	3,854***	1,616***	1,299
	중앙값	1,286***	153***	66
거래량(주)	평균값	1,432,237**	22,934,435***	35,666
	중앙값	12,926***	14,947,459***	967
거래대금(만원)	평균값	198,855***	383,628***	192
	중앙값	8,255***	250,737***	4.01
일중 변동성 <sup>1)</sup> (최저가 대비)	평균값	8.71%***	50.57%***	8.22%
	중앙값	6.40%***	39.91%***	0.00%
일간 변동성 <sup>1)</sup>	평균값	-2.55%	-21.39%***	-0.04%
	중앙값	-0.50%***	-25.20%***	0.00%

상장폐지 전 1년간의 기간에 비하여 정리매매기간 동안 주가는 1/8 수준으로 하락하였다. 정리매매기간 동안 거래규모와 변동성은 급등하였다. 이전 기간에 비하여 정리매매기간 동안 거래대금은 30배, 거래량은 1천배 이상 증가하였으며 일중 변동성은 6.4%에서 40%까지 치솟았다. 그러나 장외시장으로 이전한 이후에는 주가 급락은 안정되었으며 변동성 또한 극히 낮아졌다.

뉴 로스 78 0 (0.00%)



**상장폐지 직전 거래량 증가 확인**

# 정리매매기간에서 실질적 부실기업과 거래량의 관계

<표 III-2> 상장폐지 전후 유동성 및 변동성 비교 출처 : 강소현, 2020

		상장폐지 전 1년 (정리매매 제외)	정리매매	상장폐지 후 3개월
종목 수(개)		52		
실거래일수(일)	평균값	123(52%)	7(100%)	28(46%)
	중앙값	127(53%)	7(100%)	31(50%)
주가(원)	평균값	3,854***	1,616***	1,299

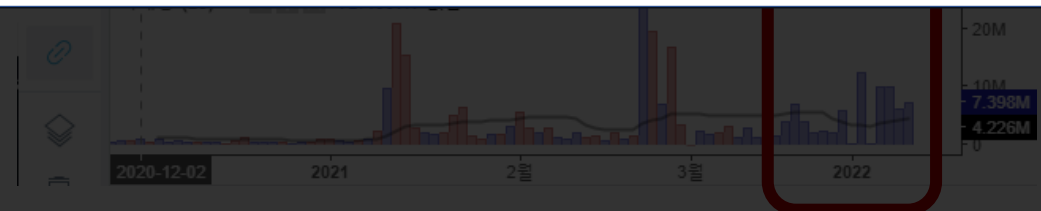
뉴 로스 78 0 (0.00%)



**정리매매기간 동안 거래량이 급증한다고 해서 요행을 바라고 투자하지 말자**

일간 변동성 <sup>1)</sup>	평균값	-2.55%	-21.39%***	-0.04%
	중앙값	-0.50%***	-25.20%***	0.00%

상장폐지 전 1년간의 기간에 비하여 정리매매기간 동안 주가는 1/8 수준으로 하락하였다. 정리매매기간 동안 거래규모와 변동성은 급등하였다. 이전 기간에 비하여 정리매매기간 동안 거래대금은 30배, 거래량은 1천배 이상 증가하였으며 일중 변동성은 6.4%에서 40%까지 치솟았다. 그러나 장외시장으로 이전한 이후에는 주가 급락은 안정되었으며 변동성 또한 극히 낮아졌다.



**상장폐지 직전 거래량 증가 확인**

# 한계점

1. 대형주를 포함한 분석 결과와 인사이트를 비교하지 못 함
2. KOSPI, KOSDAQ 대상 정상기업과 부실기업의 데이터 불균형
3. 일반화의 어려움 (코로나 기간, 해외기업, 외감기업 제외)

## 후속 연구방향 제안

1. 대형주를 포함한 분석결과와의 비교 분석
2. 경기충격 상황을 포함한 기간 설정
3. 해외기업 분석에서의 시장정보 활용

# Reference

- 강소현, 상장폐지종목의 장외거래 특성 분석, 자본시장연구원 이슈보고서, 2020년 5월
- 김남호, "중소기업 부실예측에 관한 분석모형 비교연구: 로짓분석, 판별분석, k-최소근방분석을 중심으로" 保證月報, 3-41, 2004
- 김량형, 유동희, 김건우, 경기순환국면을 반영한 기업부실화 예측 모델, Information Systems Review Vol. 18, No. 2 (2016)
- 배재권, 재무비율을 이용한 기업부도. 예측모형의 실증적 연구, 서강대학교 대학. 원 석사학위 논문 (2006).
- 성성민, CTGAN 및 TabNet 기법을 활용한 불균형 정형 데이터 이진분류 모델링 개발, 아주대학교 일반대학원 인공지능학, 2022년 2월
- 오우석, 김진화, 인공지능기법을 이용한 기업부도 예측, 대한산업경영학회지 제 15권, 2017
- 이인로, 김동철, 회계정보와 시장정보를 이용한 부도 예측모형의 평가 연구, 재무연구 제28권 제4호, 2015
- 이현상, 오세환, 시계열 예측 을 위한 LSTM 기반 딥러 닝 기업 신 용평점 예측 사례 , 정보시스템연구 제29권 제1호, 2020
- 조재혁, 안은주, 김성수. 딥러닝 기반 부실기업 예측모형에 관한 연구, Journal of Business Research Vol. 36 No. 1 (2021) pp.99~113
- Altman, E. I., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," The Journal of Finance, Vol. 23, No. 4, 1968, pp. 589-609
- S.Ö. Arik, T. Pfister. TabNet: Attentive Interpretable Tabular Learning, Proceedings of the AAAI Conference on Artificial Intelligence, 35 (8) (2021), pp. 6679-6687



**감사합니다**