



# 머신러닝을 통한 시장규모별 부실기업위험그룹 분류 및 투자전략 수립

HAY Dealer

# 목차

1. 주제선정이유
2. 선행연구 및 차별점
3. Workflow
4. 부실기업 정의 & Feature 선정
5. 최적모델선정 & 성능평가
6. 투자전략
7. 시사점
8. Timetable

<div><div> 통계청</div><div>보 도 자 료</div><div>다시 도약하는 대한민국 함께 잘사는 국민의 나라</div></div>			
보도 일시	2023. 3. 2.(목) 08:00	배포 일시	2023. 3. 2.(목) 08:00
담당 부서	경제동향통계심의관 산업동향과	책임자	
		담당자	

최근 경기순환기의 기준순환일 설정

- 제12순환기의 경기저점으로 『2020년 5월』을 잠정 설정

□ 통계청(청장 한훈)은 경기분석 시 참고할 수 있도록 제12순환기의 경기저점으로 '20년 5월을 잠정 설정하였음

2023년 3월 경제순환기 발표, 미 연방준비위원회의 금리상승으로 인한 국내 금리상승 압박으로 시장금리 상승중, 2023년 바젤3 최종적용, IFRS17/9 도입 등으로 인한 자본시장 급변으로 기업재무건전성이 투자결정에 에 대한 영향력 커지고 있다.



표 2    산업별 부도손실률의 기술통계량

구 분	관측치	평균	중위수	표준편차	최소값	최대값	왜 도	첨 도
건설업	1,374	0.653	0.754	0.329	0	1	-0.698	2.159
경공업	2,688	0.717	0.849	0.320	0	0.999	-1.038	2.771
도소매업	7,993	0.678	0.818	0.342	0	0.999	-0.831	2.272
서비스업	4,214	0.646	0.761	0.346	0	0.999	-0.668	2.007
건설업	2,663	0.674	0.791	0.330	0	0.999	-0.796	2.273
전 체	18,932	0.674	0.802	0.338	0	1	-0.806	2.252

차입금 부도손실률 (출처: 부도손실률의 특성에 관한 연구(장영민) Journal of SME Finance VOL.320)

- 차입금 기준으로 주식의 경우 부도손실률은 1에 수렴
- 청구권순위 1순위인 차입금 기준 부도손실률로 3순위 보통주주의 경우 부도로 인한 손실이 더 클것

# 주제선정이유

---

## 연구 진행방향

---

- KOSDAQ을 중심으로 연구를 진행한다.
  - KOSDAQ을 이용한 부실기업 예측과 KOSPI까지 고려한 부실기업 예측을 통해 시장규모의 차이가 부실기업 예측에 유의미한 영향을 미친다는 가정을 확인한다.
  - 기업의 업력, 현금흐름 등의 변수를 모형에 추가하여 이들이 부실기업에 미치는 영향을 확인한다.
-

# 선행연구 및 차별점

	경기순환국면을 반영한 기업부실화 예측모델	재무비율을 이용한 부도예측에 대한 연구	감사의견, 감사법인 및 기업 부실리스크의 예측
연구목적	경기순환국면을 반영한 기업부실화 예측모형을 만드는데 있어 거시경제위험요소나	재무비율을 이용한 부도예측모형을 개발 평가하고 외감기업 의 부도위험 관리	재무비율 변수 중 통계적으로 부실기업과 건전기업간 유의가 있는 재무비율을 선정, 이들을 기초로 부실예측모형 구축
대상기업	KOSPI + KOSDAQ 제조업 및 비제조업 1,691개 기업 대상	외감기업	<ul style="list-style-type: none"><li>• KOSPI 상장기업 + KOSDAQ 기업</li><li>• 금융업 제외</li><li>• 12월 결산법인</li></ul>
대상기간	2000-2011년	2001-2007	2005-2009
Feature 선정방법	기존 재무보고서에서의 재무비율 + <ul style="list-style-type: none"><li>• 재무상태변동성</li><li>• 산업수준상태변동성(Top 10)</li><li>• 산업수준상태변동성(Average)</li></ul> Information Gain, Gain Ratio 이용	<ul style="list-style-type: none"><li>• T검정</li><li>• 단일변량 로짓</li><li>• Accuracy Ratio 등</li></ul>	선행연구를 바탕으로 유의한 변수 사용 <ul style="list-style-type: none"><li>• T-test 분석을 사용하여 Feature 선정</li></ul>
최적 결정모형	AdaBoost	다변량 LOGIT 분석	로짓회귀분석 T-test, Pearson-Cor, Kai 검정
차별점 / 한계점	<ul style="list-style-type: none"><li>• 다음의 문제를 해결하고자 함<ul style="list-style-type: none"><li>• 데이터셋 구성하는 과정에서의 Bias문제</li><li>• 거시경제위험 요소의 미반영문제</li><li>• 데이터 불균형문제</li><li>• 서술적 Bias</li></ul></li><li>• 경기순환국면을 데이터셋에 반영</li><li>• 산업수준에 따른 Feature를 파생변수로 만들</li></ul>	<ul style="list-style-type: none"><li>• 외감기업 전체를 표본분석</li><li>• 금융결제원의 당좌거래 정지정보에 기초하여 포괄적인 부도기업표본을 구축하여 분석 평가</li></ul>	<ul style="list-style-type: none"><li>• 메이저 회계법인의 감사여부를 감사의견의 질에 반영함 – 감사주체에 따라 도산율의 차이가 존재하는가? 에 대한 가설을 검정</li><li>• 쌍대표본추출법으로 연구자 판단에 의해 통제기업 선정 : 연구자의 자의성이 개입될 가능성</li></ul>

# 선행연구 및 차별점

	부실예측모형의 예측력 비교 : 판별분석모형 및 로짓분석모형 중심으로	기업부실예측 데이터의 불균형 문제 해결을 위한 앙상블 학습
연구목적	서울 및 제주지역 특1급 호텔의 14개 재무비율을 변수로 활용한 부실예측모형 개발	해당 연구에서는 GM-Boost를 데이터 불균형이 존재하는 2범주 분류 문제인 기업부실 예측에 적용하여 제안하여 알고리즘의 성과를 검증
대상기업	특1급 호텔 21개(서울호텔 16개, 제주호텔 5개)	(주)한신평정보 기업
대상기간	2015년	2002-2005년
Feature 선정방법	t-test, 판별분석, 로짓모형	ML 이용한 피쳐선정 후 벤치마크끼리 정확도(Accuracy)를 T검정 수행(SVM, AdaBoost)
최적 결정모형	-	XGB
차별점 / 한계점	정상기업과 부실기업간 변수의 영향력 차이를 T-test를 통해 비교 재무비율을 표준재무비율로 환산하여 변수에 사용	데이터 불균형 별 오버샘플링 비율에 따른 데이터셋을 나눠 검증

I. 부실 정의 및 Feature 선정

부실기업 정의

- 실적부진사유로 상장폐지가 일어난 경우

대상기간, 대상기업

- 2011년 – 2019년 / KOSDAQ + KOSPI

Dataset 구성

- Dataset1 : KOSDAQ
- Dataset2 : KOSDAQ + KOSPI

가져온 Features

- 선행연구논문 등에서 참고한 재무변수
- 업력을 고려한 Feature 고려

EDA / 데이터 전처리

- 이상치, 결측치 확인 및 조정
- 통계분석 수행

변수선정방법

- MDA(다변량 판별분석)
- Logit(Logistic Regression)
- Lasso 등

Feature 선정

II. Modeling & 성능평가

Train

- 2011-2017

Test

- 2018-2019

분류모형 :

- 판별분석
- Logistic Regression
- Decision Tree
- Random Forest
- KNN
- XGboost
- MLP

성능평가 : Confusion Matrix 기반

- Accuracy, Precision, Recall, ROC AUC 등

부실기업 예측  
최적 Model 선정

III. Investment Strategies

상관분석

- 최적모델의 Feature와 주식 수익률간 상관관계

주가와 유의미한 Feature 선정

- 주식선정기준, 제외기준 수립

투자 Universe 구축

- 위 기준에서 필터링 된 기업 대상
- Quant 투자, 기술적 분석 등의 방법으로 투자

Backtesting

- SHARPE RATIO 등 확인
- Feature 변경해가며 결과 확인

포트폴리오  
수익률 확인



# 부실기업의 정의 & Feature의 선정

## I. 부실 정의 및 Feature 선정

**부실기업 정의**  
•실적부진사유로 상장폐지가 일어난 경우

**대상기간, 대상기업**  
•2011년 – 2019년 / KOSDAQ + KOSPI

□ **부실기업의 정의** : 실적부진사유로 상장폐지가 일어난 경우

### 부실기업에 해당하는 사유

- 부도발생
- 화의절차개시신청
- 회사정리절차개시신청
- 감사인의 의견거절
- 은행거래정지

### 부실기업에 해당하지 않는 사유

- 신규/변경상장
- 특수목적에 의한 상장폐지
- 자진등록취소 등

□ **대상기간** : 2011년 - 2019년

□ **대상기업** : KOSDAQ + KOSPI  
□ 금융업 제외 / 12월 결산법인만 고려

□ **자료출처** : DART, TS2000, Naver금융

주권매매거래정지해제		
1.대상종목	(주)지아이바이오	보통주
2.해제사유	상장폐지에 따른 정리매매 개시	
3.해제일시	2013-07-29	-
4.근거규정	코스닥시장상장규정 제29조 및 동규정시행세칙 제29조	
5.기타	* 상장폐지 내역 - 상장폐지사유 : 감사의견거절(감사범위 제한 및 계속기업으로서의 존속능력에 대한 불확실성) - 정리매매기간 : 2013.07.29~2013.08.06(7매매일) - 상장폐지일 : 2013.08.07	

상장폐지 결정		
1. 상장폐지신청 예정일자	2019-05-23	
2. 상장폐지신청 사유	한국거래소 유가증권시장 이전상장에 따른 코스닥시장 상장폐지	

**\*\* 데이터 상의 상장폐지사유를 DART에서 개별기업 별로 확인해보는 과정을 진행하였음**

# 부실기업의 정의 & Feature의 선정

## I. 부실 정의 및 Feature 선정

### 부실기업 정의

•실적부진사유로 상장폐지가 일어난 경우

### 대상기간, 대상기업

•2011년 – 2019년 / KOSDAQ + KOSPI

### Dataset 구성

- Dataset1 : KOSDAQ
- Dataset2 : KOSDAQ + KOSPI

## □ Dataset을 두개로 구성

### Dataset 1

KOSDAQ 기업으로만 구성된 DATASET 1

KOSDAQ

Train : 2011 - 2017

KOSDAQ

Test : 2018 - 2019

### Dataset 2

KOSDAQ + KOSPI 기업으로 구성된 DATASET 2

KOSDAQ

KOSPI

Train :  
2011 - 2017

KOSDAQ

KOSPI

Test :  
2018 - 2019

# 부실기업의 정의 & Feature의 선정

## I. 부실 정의 및 Feature 선정

**부실기업 정의**  
• 실적부진사유로 상장폐지가 일어난 경우

**대상기간, 대상기업**  
•2011년 – 2019년 / KOSDAQ + KOSPI

**Dataset 구성**  
•Dataset1 : KOSDAQ  
•Dataset2 : KOSDAQ + KOSPI

**가져온 Features**  
•선행연구논문 등에서 참고한 재무변수  
•업력을 고려한 Feature 고려

### 재무비율

### 기타

#### 수익성

- 매출액총이익률
- 매출액순이익률
- 총자산사업이익률
- 총자산순이익률
- 자기자본순이익률

...

#### 활동성

- 총자본회전률
- 자기자본회전률
- 순운전자본회전률
- 매출채권회전률
- 재고자산회전률

...

#### 현금흐름표

- 영업현금흐름
- 재무현금흐름
- 투자현금흐름

#### 성장성

- 매출액증가율
- 총자산증가율
- 자기자본증가율
- 순이익증가율
- 유형자산증가율

...

#### 안정성

- 유동비율
- 당좌비율
- 순운전자본비율
- 부채비율
- 이자보상배율

...

**업력 등  
기타 Feature**

위의  
Feature들을  
기반으로  
Feature  
선정을 수행함

# 부실기업의 정의 & Feature의 선정

## I. 부실 정의 및 Feature 선정

### 부실기업 정의

•실적부진사유로 상장폐지가 일어난 경우

### 대상기간, 대상기업

•2011년 – 2019년 / KOSDAQ + KOSPI

### Dataset 구성

•Dataset1 : KOSDAQ  
•Dataset2 : KOSDAQ + KOSPI

### 가져온 Features

•선행연구논문 등에서 참고한 재무변수  
•업력을 고려한 Feature 고려

### EDA / 데이터 전처리

•이상치, 결측치 확인 및 조정  
•통계분석 수행

## □ Target Labeling :

상장폐지 직전연도:1	0	1	
이 외 : 0	정상	정상	부도

- **EDA** : QQ Plot, Boxplot, 등을 이용한 Feature별 데이터 분포를 확인
  - 상관관계 분석 : Heatmap, VIF를 이용한 다중공선성 및 상관관계 확인
- **결측치** : TS2000의 데이터와 Dart의 사업보고서 수치를 Cross Check
  - 연결재무제표에 누락된 데이터 개별재무제표의 수치로 대체 혹은 제거
  - 중앙값, 평균값 등으로 대체 혹은 제거
- **이상치** : Winsorizing을 통한 양 극단값 대체

# 부실기업의 정의 & Feature의 선정

## I. 부실 정의 및 Feature 선정

### 부실기업 정의

•실적부진사유로 상장폐지가 일어난 경우

### 대상기간, 대상기업

•2011년 – 2019년 / KOSDAQ + KOSPI

### Dataset 구성

- Dataset1 : KOSDAQ
- Dataset2 : KOSDAQ + KOSPI

### 가져온 Features

- 선행연구논문 등에서 참고한 재무변수
- 업력을 고려한 Feature 고려

### 데이터 전처리

- 이상치, 결측치 확인 및 조정
- 통계분석 수행

### 변수선정방법

- MDA(다변량 판별분석)
- Logit(Logistic Regression)
- Lasso 등

## Dataset1 : KOSDAQ

## Dataset2 : KOSDAQ + KOSPI

### MDA / LOGIT / LASSO / T-test / Stepwise / KBest

MDA	LOGIT	...	Kbest
매출액증가율	부채비율	...	부채비율
부채비율	유동자산증가율	...	당좌비율
재고자산회전율	매출액증가율	...	순운전자본회전율
유동비율	영업현금흐름	...	순이익증가율
당좌비율	자기자본순이익률	...	유동자산증가율

MDA	LOGIT	...	Kbest
부채비율	부채비율	...	부채비율
비유동비율	당좌비율	...	당좌비율
유형자산증가율	매출액증가율	...	순운전자본회전율
총자본회전율	영업현금흐름	...	순이익증가율
자기자본순이익률	자기자본순이익률	...	유동자산증가율

## KOSDAQ 대상 Feature 선정

## KOSDAQ + KOSPI대상 Feature 선정



# 최적 Model 선정 & 성능평가

## II. Modeling & 성능평가

Train

•2011-2017

Test

•2018-2019

Test

•2020-2021

□ **Data Split** : IFRS 적용을 시작한 2011년부터 2019년까지의 데이터

Dataset 1

Train / Validation

•2011-2017

Test

•2018-2019

Dataset 2

Test

•2020-2021

- 2011년부터 2017년까지 데이터를 Split 하여 학습에 사용
- 불황에도 모델이 유효한지 확인을 위해 코로나 시기의 테스트셋에서도 모델의 성능을 평가해보려고 함
- 데이터의 불균형 문제는 Sampling을 통해 해결하려고 함 :  
Over-Sampling, Under-Sampling, SMOTE
- Scaling은 MinMaxScale, StandardScale, RobustScale을  
사용하려고 함

# 최적 Model 선정 & 성능평가

## II. Modeling & 성능평가

Train

•2011-2017

Test

•2018-2019

### 분류모형 :

- 판별분석
- Logistic Regression
- Decision Tree
- Random Forest
- KNN
- XGboost
- MLP

### 성능평가 : Confusion Matrix 기반

•Accuracy, Precision, Recall, ROC AUC 등

선정된  
Features

판별분석

Logistic Regression

Decision Tree

Random Forest

KNN

XGboost

MLP

True  
Positive

False  
Negative

False  
Positive

True  
Negative

- ❑ Confusion Matrix 기반
- ❑ Recall, Precision, F1 Score, ROC AUC 등을 확인

Hyper-Parameter 조정 등  
최적화 과정 수행

부실기업 예측  
최적 Model 선정

## III. Investment Strategies

### 상관분석

- 최적모델의 Feature와 주식 수익률간 상관관계

### 주가와 유의미한 Feature 선정

- 선정기준, 제외기준 수립

□ **상관분석:** 선정된 Feature들과 주식 수익률간의 상관관계를 파악

#### 선정된 Features(예)

- 부채비율
- 당좌비율
- 매출액증가율
- 주당이익
- 자기자본회전율

주식 수익률

□ **상관분석을 통해 유의미한 Feature 선정**

□ **유의미한 Feature에서 투자에 고려할 기준을 선정**

- 예) 부채비율이 1.5 이하
- 예) 매출액증가율이 n% 이상

# 투자전략

## III. Investment Strategies

### 상관분석

- 최적모델의 Feature와 주식 수익률간 상관관계

### 주가와 유의미한 Feature 선정

- 선정기준, 제외기준 수립

### 투자 Universe 구축

- 위 기준에서 필터링 된 기업 대상
- Quant 투자, 기술적 분석 등의 방법으로 투자

## 투자 전략

### ■ 기술적 분석을 통한 투자

- 이동평균선 분석 및 Buy and Hold
- Bollinger Band를 이용한 매매전략

### ■ Quant투자 방법론 이용 고려

- Benjamin Graham : NCAV(순유동자산)과 시가총액을 비교, 두 조건을 만족하는 포트폴리오 구성
  - 순유동자산이 시가총액의 1.5배 이상
  - 흑자기업
- Peter Lynch
  - $PEG = .PER / 3\text{년평균 EPS증가율}$
- John B. Neff
  - $(\text{당기순이익증가율} + \text{배당수익률}) / PER$
- Joel Greenblatt의 마법공식(Magic Formula)
  - ... 등 퀀트투자전략을 이용하여 포트폴리오 구성

# 투자전략

## III. Investment Strategies

### 상관분석

- 최적모델의 Feature와 주식 수익률간 상관관계

### 주가와 유의미한 Feature 선정

- 선정기준, 제외기준 수립

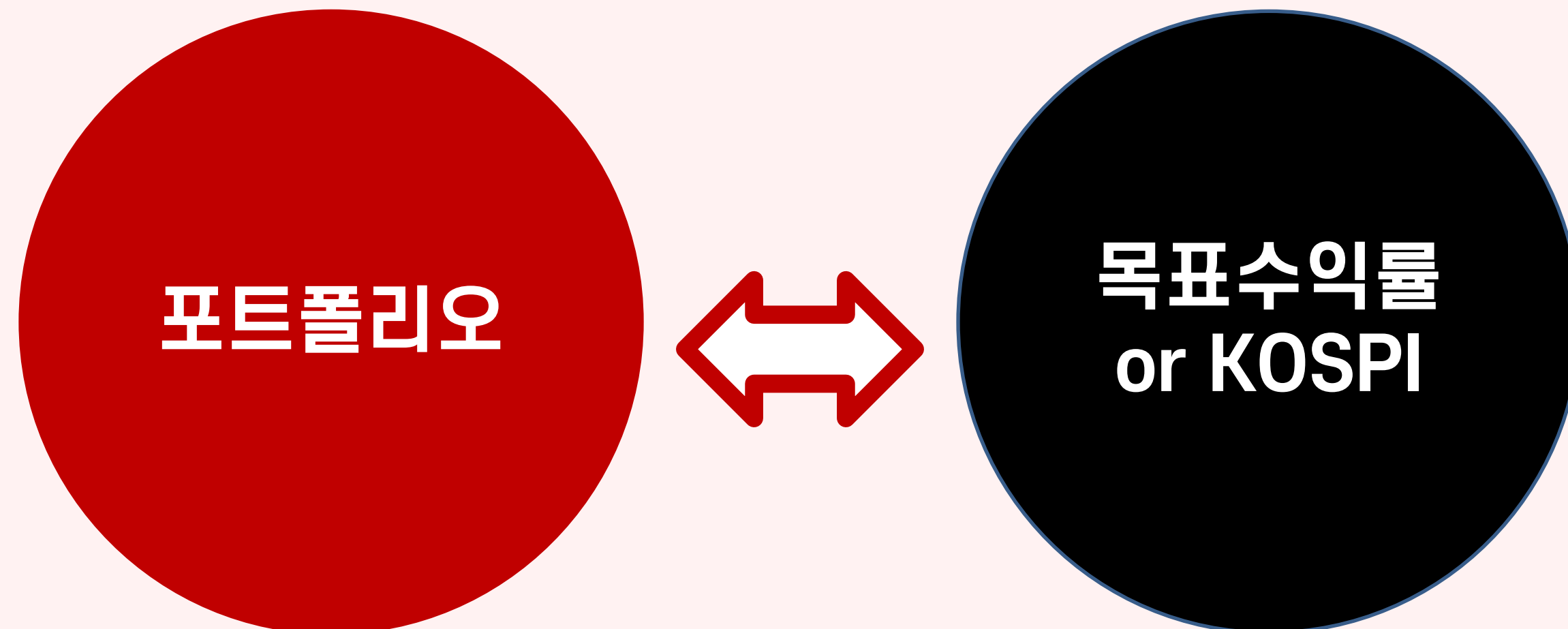
### 투자 Universe 구축

- 위 기준에서 필터링 된 기업 대상
- Quant 투자, 기술적 분석 등의 방법으로 투자

### Backtesting

- SHARPE RATIO 등 확인
- Feature 변경해가며 결과 확인

포트폴리오 수익률 확인



- Sharpe Ratio 확인 및 포트폴리오 재조정
- Feature 변경 혹은 주식 비율 조정을 통한 포트폴리오 재조정, 최적 포트폴리오 수익률 확인



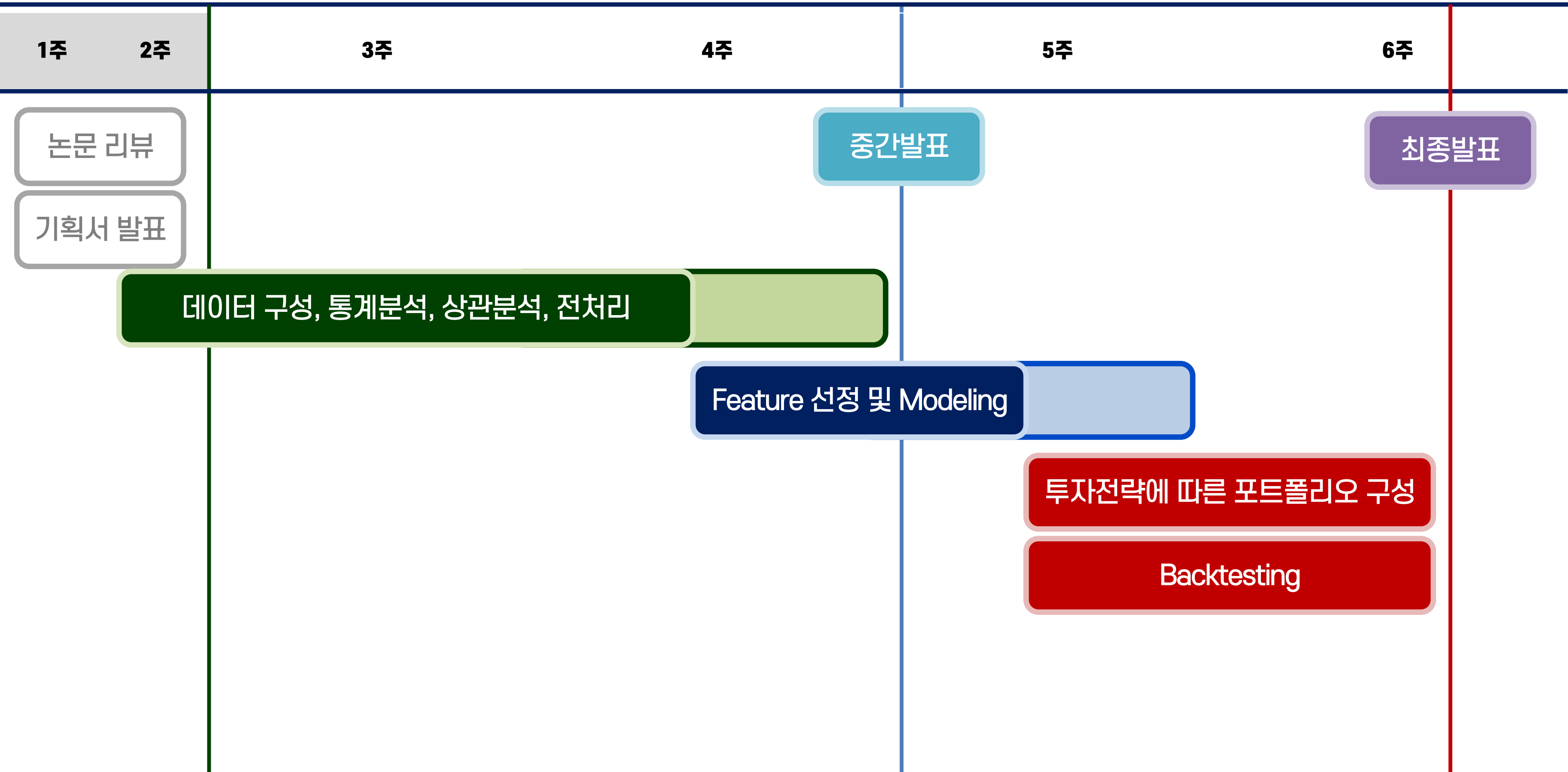
## 의의

- 시장의 차이에 따라 부실기업을 판단하는 재무비율의 유의미한 차이를 확인
- 다양한 Feature 선정 방식과 분류모델을 사용하여 그 값을 비교하고 최적의 모델을 선택하였다는 점
- 부실기업에 유의미한 재무비율을 이용한 투자를 수행해 보았다는 점

## 한계

- 부실기업에 대한 데이터 수의 한계
- Sampling을 사용
- 산업별 기업의 부실에 대한 요인을 고려하지 못한 점
- 거시경기국면에 대한 요인을 반영하지 못한 점

# Timetable







감사합니다

HAY Dealer