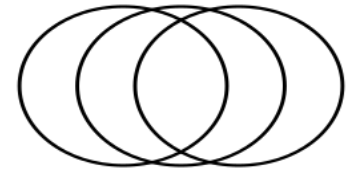
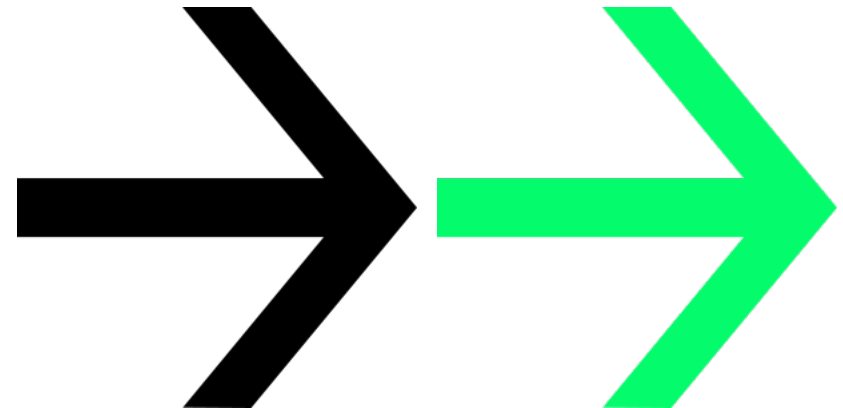


# 머신러닝 기반 비상장외감기업의

## 상장 가능 예측 모델



IPickO 윤도형, 정현지, 박찬미



# 팀 소개



리더  
윤도형



메인 코더  
정현지



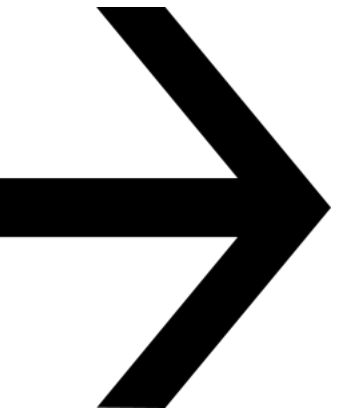
서브 코더  
박찬미

나(=투자자)는  
**Pick**고른다  
☐ 성공할 기업만을!



비상장기업의 상장 성공 여부를 예측하여,  
비상장기업 투자를 고려하는 개인투자자에게  
**효율적인 투자지표**를 제공

# CON- TENTS



01

서론  
- 연구주제  
- 주제선정배경

04

Feature Selection  
& EDA

02

WORKFLOW

05

Modeling  
& 성능평가

03

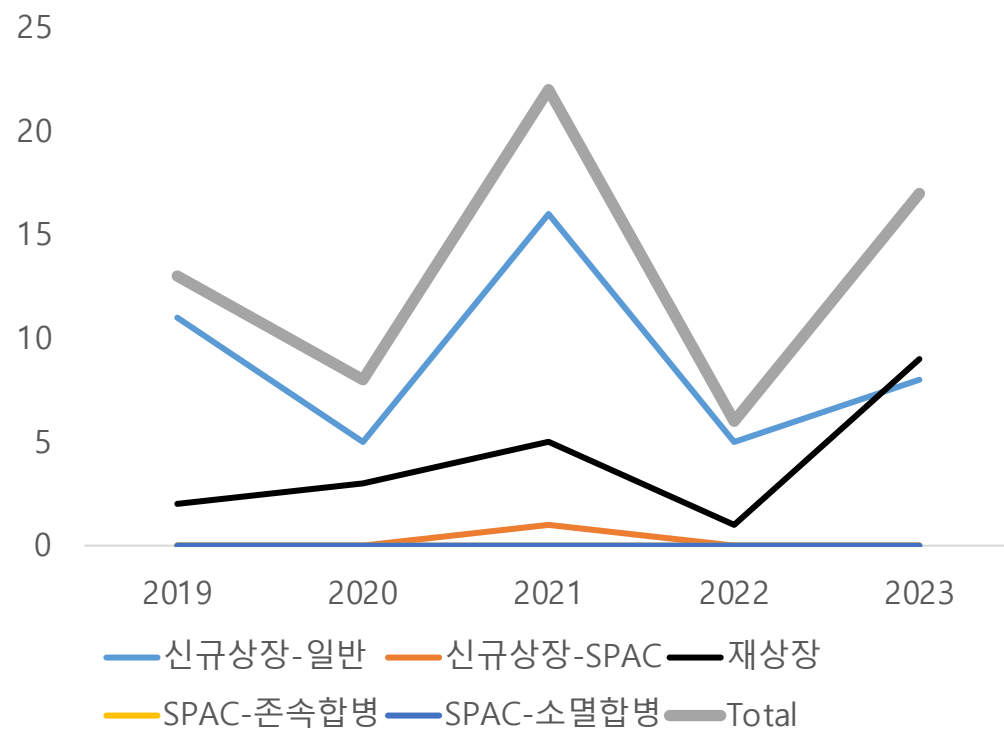
데이터 수집  
& 전처리

06

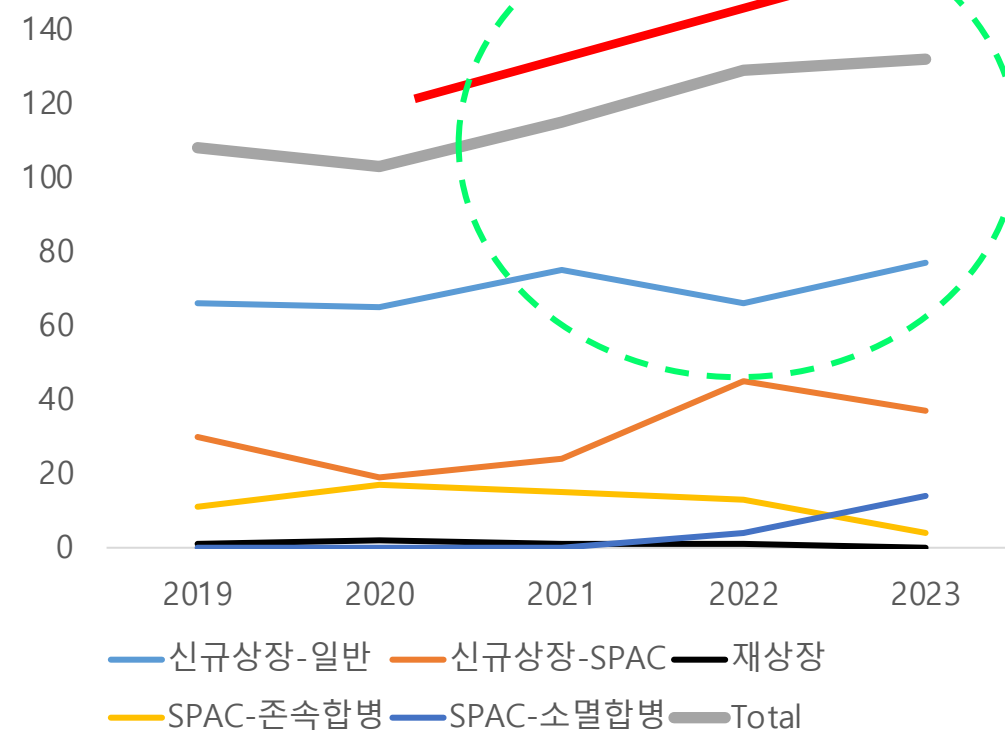
연구의의 및 한계

# 비상장기업 투자 현황

## 유가증권



## 코스닥



# 비상장기업 투자 요인

- ✓ **성장가능성이 높은 기업 발굴**  
→ 성장을 지원하여 미래의 수익률 기대
- ✓ **저평가된 기업 발굴**  
→ 실제 가치 미리 발견하고 투자하여 미래의 수익률 기대
- ✓ **투자 포트폴리오 다양성 UP**  
→ 장내시장에 비해 경쟁이 적은 비상장기업에 투자해  
포트폴리오를 다양화시켜 리스크 분산 기대

# 선행 연구

KCI 등재

## 비재무적 지표를 이용한 벤처기업의 신규상장(IPO)성공요인에 관한 연구

A Study on Success Factors of Initial Public Offerings for Venture Corporation Using Non-Financial Factors

중소기업연구

2007, vol.29, no.2, pp. 1-22 (22 pages)

UCI : G704-SER000001590.2007.29.2.008

발행기관 : 한국중소기업학회

연구분야 : 사회과학 > 경영학

나종호<sup>1</sup>, 박성규<sup>2</sup>, 김창수 / KimChangSoo<sup>3</sup>

<sup>1</sup>보람기업자문주

<sup>2</sup>가톨릭대학교

<sup>3</sup>중앙대학교

우수등재

## 비재무 정보를 활용한 IPO 주식의 상장일가격 등락 예측에 관한 연구

A study on initial price change prediction of IPO shares using non-financial information

한국데이터정보과학회지

약어 : JKDIS

2018, vol.29, no.2, pp. 425-439 (15 pages)

DOI : 10.7465/jkdi.2018.29.2.425

발행기관 : 한국데이터정보과학회

연구분야 : 자연과학 > 통계학

신상훈<sup>1</sup>, 이현준 / Hyun Jun Lee<sup>2</sup>, 안재준 / Jaejoon Ahn<sup>3</sup>

<sup>1</sup>연세대학교 투자정보공학과

<sup>2</sup>연세대학교 산업공학과

<sup>3</sup>연세대학교

→ 기존 연구: **재무적요인 OR 비재무적요인**을 통한 비상장기업의 상장, 가격 등락 예측  
 ✓ **재무적지표& 비재무적지표 모두 이용**하여 상장 성공 예측하는 모델 제안

# 연구 주제

## ✓ 목표

외부 감사를 받는 모든 기업 대상  
상장승인 여부 예측 모델 구현

## ✓ 의의

투자 결정 내리는 데 도움되는 효율적인 정보 제공  
투자 성공률↑ & 수익률 ↑

# WorkFlow

## 1. 데이터수집

### 대상

- 비상장  
외부감사대상법인

### 기간

- 1999년~2022년

### 수집처

- TS2000
- KIND

## 2. 데이터전처리 & EDA

### 결측치

- 공식 계산 후 대체
- ### 이상치
- Winsorizing
  - (상위 1%, 하위 1%)

### Train\_Test\_Split

- Train : 70%
- Test : 30%

### 스케일링

- Min-Max Scaler

### Feature Selection

- Lasso

### EDA

- 상관계수
- 기초 통계량
- 정규 분포

## 3.모델링

### 트리모델

- DecisionTree
- RandomForest
- XGBoost
- CATBoost
- LightGBM

### 회귀모델

- LogisticRegression

### 리샘플링

- SMOTE

## 4.평가 및 검증

### 성능평가

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

### 검증

- StratifiedKFold



# DATA 수집

# DataSet

## 예비심사기업(2000~2023)

기업	심사결과	상장 종류
기업A	상장승인	이전상장
기업B	상장승인	신규상장
기업C	상장철회	신규상장
기업D	상장승인	신규상장
기업E	심사미승인	SPAC
기업F	상장승인	신규상장
기업G	공모철회	SPAC
⋮	⋮	⋮



기업	심사결과	상장 종류
기업B	상장승인	신규상장
기업D	상장승인	신규상장
기업F	상장승인	신규상장
⋮	⋮	⋮



상장승인 기업 중  
신규상장 기업



그 외 나머지

이전상장, 재상장, 심사미승인, SPAC존속합병,  
SPAC소멸합병, 공모철회, 상장철회

⋮

# DataSet



	상장승인 외감기업	비상장 외감기업
수집기간 수집대상	2000~ 2023년 Kospi & Kosdaq에 상장한 기업 (~2023)	1999~ 2022년 Kospi & Kosdaq에 상장하지 않은 외부감사기업(~2023)
수집 데이터 수집처	각 기업별 <b>심사청구 연도 기준 전</b> 5개년 사업보고서 KIND, TS2000	각 기업별 <b>최근</b> 2년 사업보고서 TS2000
기업 수	<b>2,108 개</b>	<b>75,088 개</b>

# DATA 전처리

# 상장승인 외감 기업 Dataset

기업수		
2,108		
1,796	-312	✓ 리츠(21개) , 스펙 (291개) 기업 제거
1,696	-100	✓ 심사청구 이전 사업보고서 없는 기업 제거
1,285	-411	✓ 재무비율 없는 기업 제거
1,118	-167	✓ 사업보고서가 2개년~3개년 없는 기업 제거
1,273	+155	✓ 결측치를 추정치로 대체하며 제거된 기업 다시 추가
1,150	-123	✓ 기술상장특례 기업 기업 제거

심사 청구년도 별 기업 수

청구년도	종목코드의 COL
2000	176
2001	224
2002	118
2003	54
2004	51
2005	81
2006	61
2007	74
2008	53
2009	60
2010	83
2011	61
2012	31
2013	30
2014	43
2015	75
2016	63
2017	55
2018	71
2019	67
2020	83
2021	78
2022	64
2023	40
총계	1796

# 비상장 외감기업 Dataset

기업수		
75,088	-7,487	✓ 재무비율 없는 기업 제거
67,601	-34,362	✓ 2022년 사업보고서 없는 기업 제거
33,239	-587	✓ 2021 or 2022년 사업보고서 없는 기업 제거
✓ 32,652		

# 기술 성장 특례 기업 제거

기술력이 우수한 기업에 대해 외부 검증기관을 통해 심사한 뒤  
수익성 요건을 충족하지 못하더라도 상장 기회를 주는 제도(2005~)

✓ 상장 시 소속에 따라 요건 상이  
→ **기술성장특례기업** 제외

[12020] 상장회사 상세검색

[12020] 상장회사 상세검색

통계 > 기본통계 > 주식 > 세부안내 > 상장회사 상세검색

시장구분: ☒ 전체 ☐ KOSPI ☐ KOSDAQ ☐ KONEX

종목명: 전체

정렬방법: ☐ 종목명 ☒ 상세검색

업종명: 전체

액면가: 전체

결산월: 전체

상장주식수: 이상

자본금: 이상

☒ 소속부 기술성장기업부

☐ 시가총액 전체

전체  
우량기업부  
벤처기업부  
중견기업부  
✓ 기술성장기업부  
일반기업부  
크라우드펀딩기업부

KRX 정보데이터시스템 > 기본통계 > 주식 > 세부안내 > 상장회사 상세검색

KRX 기준 기술성장 특례 상장 기업 **204개 제거**  
(그 중 스펙, 이전상장 42개 기업 리스트 제외 시, **162개 신규상장 제외됨**)

# 금융업 제거

✓ 업종 별 특성에 따라 재무제표 상이  
→ **금융사**(증권사, 은행, 대부업) 제외

산업코드	산업명
116502	금융지주업
117101	여신금융업
117201	기타금융업



# Final Data Set

TOTAL

Before

77,196개



After

33,802개  
(100%)



상장 승인 외감기업

2,108 개



1,150개  
(약 3%)



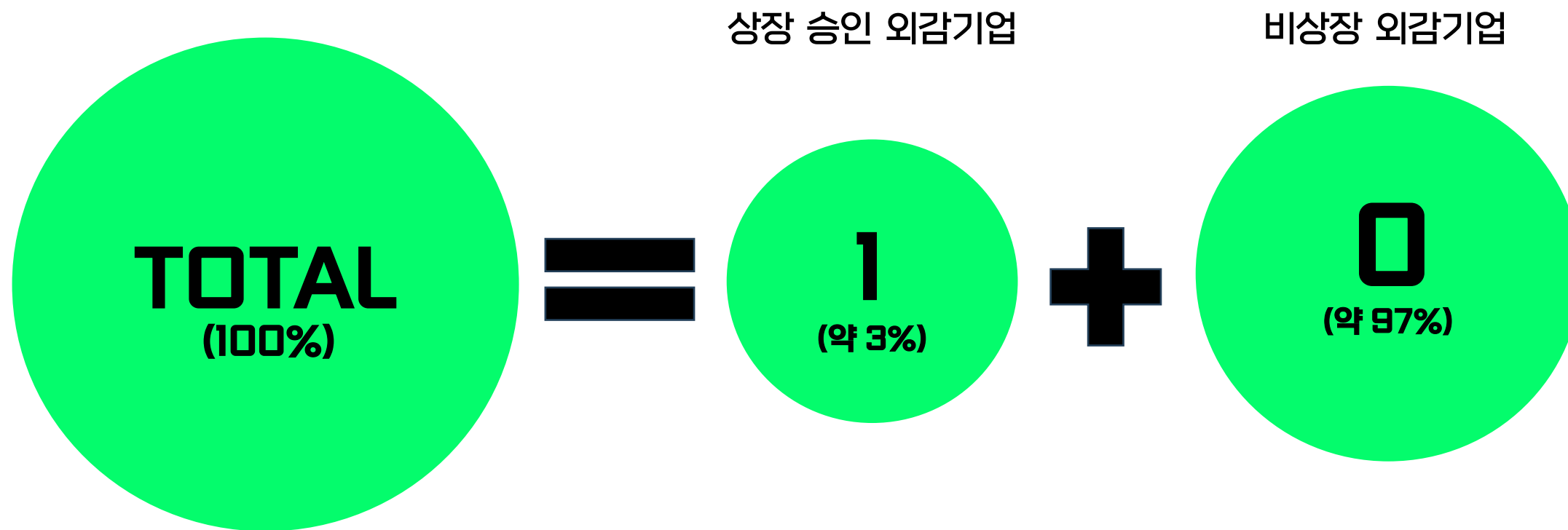
비상장 외감기업

75,088 개



32,652 개  
(약 97%)

# Lable



# 데이터 탐색

# Winsorizing :이상치 처리



상위 1%와 하위 1% data → 99% 이내로 맞추기

## 자산총계\_1:

원본 데이터 : 상위 1%: 733879258.00000054, 하위 1%: 3026209.0

원저라이징 데이터 : 상위 1%: 733879258.00000054, 하위 1%: 3150000.0

## 부채총계\_1:

원본 데이터 : 상위 1%: 452784590.00000018, 하위 1%: 987209.0

원저라이징 데이터 : 상위 1%: 452784590.00000018, 하위 1%: 1080000.0

## 자본총계\_1:

원본 데이터 : 상위 1%: 368636544.00000098, 하위 1%: -34106772.0

원저라이징 데이터 : 상위 1%: 368636544.00000098, 하위 1%: -29902000.0

## 법인세비용차감전(계속사업)손익\_1:

원본 데이터 : 상위 1%: 62656292.000000164, 하위 1%: -23537997.0

원저라이징 데이터 : 상위 1%: 62656292.000000164, 하위 1%: -21642000.0

## 영업이익\_1:

원본 데이터 : 상위 1%: 52232189.00000001, 하위 1%: -14329582.0

원저라이징 데이터 : 상위 1%: 52232189.00000001, 하위 1%: -13387000.0

# 데이터 스케일링 : Min-Max Scaler

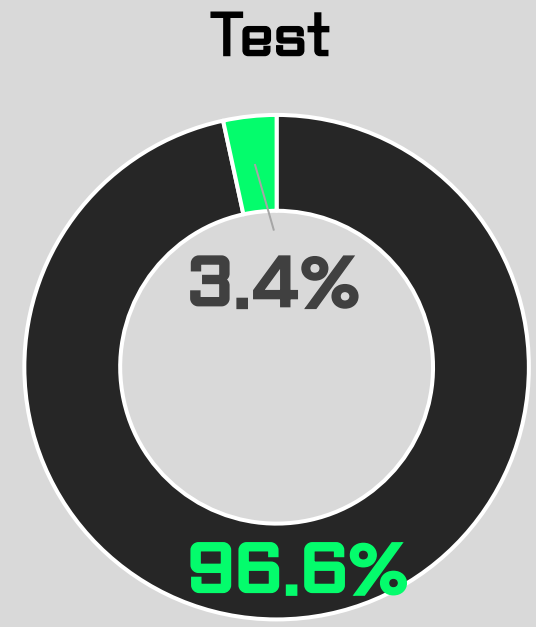
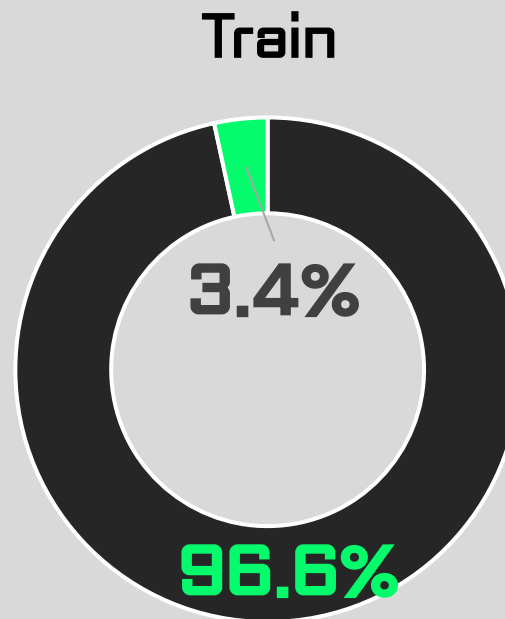
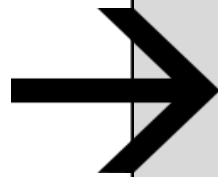
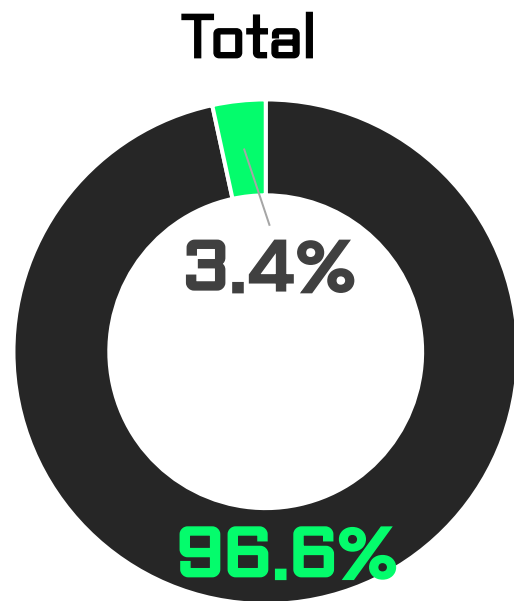
## Before

	자산 총 계_1	부채 총 계_1	자본 총 계_1	매출액 _1	법인세비 용차감전 (계속사 업)손익 _1	영업이 익_1	당기순 이익_1	투자활동 으로 인한 현금흐름 _1	매출액 영업이 익률_1	매출액 순이익 률_1
25137	7994000	6591000	1403000	70871000	-486000	-226000	-486000	705000	-0.32	-0.69
33423	36927000	13151000	23776000	15273000	2120000	2538000	2196000	-8307000	16.62	14.38
24689	28803000	14322000	14481000	7798000	726000	407000	591000	-1704000	5.22	7.58
8221	21503000	2826000	18677000	492000	2690000	-507000	2133000	4772000	-103.03	215.07
1356	14716000	14313000	403000	0	-737000	-337000	-737000	7000	0.00	0.00
...	...	...	...	...	...	...	...	...	...	...
49	22235000	15255000	6980000	24254000	659000	821000	620000	481000	3.39	2.56
21208	43575000	18947000	24628000	47072000	658000	746000	656000	-2268000	1.59	1.39
30110	53549000	46193000	7356000	1920000	-149000	1297000	-149000	5525000	67.55	-7.77
14614	15978000	10402000	5576000	2572000	773000	981000	667000	-4000	38.14	25.94
23354	3318000	3108000	210000	22775000	46000	567000	35000	390000	2.49	0.15

## After

	자산 총계 _1	부채 총계 _1	자본 총계 _1	매출 액_1	법인세비 용차감전(계속 사업)손익_1	영업 이익 _1	당기 순이 익_1	투자활동으 로 인한 현 금흐름_1	매출액 영업이 익률_1	매출액 순이익 률_1
25137	0.01	0.01	0.07	0.11	0.24	0.19	0.27	0.80	0.96	0.92
33423	0.04	0.03	0.13	0.02	0.27	0.23	0.30	0.71	0.97	0.92
24689	0.03	0.03	0.11	0.01	0.25	0.20	0.28	0.77	0.96	0.92
8221	0.02	0.00	0.12	0.00	0.27	0.18	0.30	0.84	0.89	1.00
1356	0.01	0.03	0.07	0.00	0.23	0.19	0.26	0.79	0.96	0.92
...	...	...	...	...	...	...	...	...	...	...
49	0.02	0.03	0.09	0.04	0.25	0.20	0.28	0.80	0.96	0.92
21208	0.05	0.04	0.13	0.08	0.25	0.20	0.28	0.77	0.96	0.92
30110	0.06	0.09	0.09	0.00	0.24	0.21	0.27	0.85	1.00	0.91
14614	0.02	0.02	0.08	0.00	0.25	0.20	0.28	0.79	0.98	0.93
23354	0.00	0.00	0.07	0.04	0.24	0.20	0.27	0.80	0.96	0.92

# Train\_Test\_Split



# Feature Selection

# Feature

## ✓ 상장 승인 외감 기업

사용 사업보고서:심사청구일 전년도와 전전년도

EX) 총자본증가율

→총자본증가율 (심사청구일 기준 1년전)

→총자본증가율\_1 (심사청구일 기준 2년전)

## ✓ 비상장 외감 기업

사용 사업보고서:최근 사업보고서(2022, 2021)

EX) R & D 투자효율

→R & D 투자효율 (심사청구일 기준 1년전)

→R & D 투자효율\_1 (심사청구일 기준 2년전)



# 후보 feature

#재무적 지표

#비재무적 지표

재무데이터 수집기간  
→ T-2년, T-1년(2y)

T-1	T-1 , T-2			
1. 총자본증가율_1 2. 영업이익증가율_1 3. 순이익증가율_1 4. 재고자산증가율_1 5. 자기자본증가율_1 6. 매출액증가율_1 7. 유동자산증가율_1 8. 비유동자산증가율_1	9. 매출액영업이익률 10. 매출액영업이익률_1 11. 매출액순이익률 12. 매출액순이익률_1 13. 총자본순이익률 14. 총자본순이익률_1 15. 자기자본영업이익률 16. 자기자본영업이익률_1 17. 자기자본순이익률 18. 자기자본순이익률_1 19. 기업순이익률 20. 기업순이익률_1 21. 금융비용부담률 22. 금융비용부담률_1 23. R & D 투자효율 24. R & D 투자효율_1 25. 유동자산구성비율 26. 유동자산구성비율_1 27. 자기자본구성비율 28. 자기자본구성비율_1 29. 타인자본구성비율 30. 타인자본구성비율_1	31. 비유동비율 32. 비유동비율_1 33. 유동비율 34. 유동비율_1 35. 당좌비율 36. 당좌비율_1 37. 부채비율 38. 부채비율_1 39. 차입금비율 40. 차입금비율_1 41. 이자보상배율(이자비용) 42. 이자보상배율(이자비 용)_1 43. 총자본회전률 44. 총자본회전률_1 45. 매출채권회전률 46. 매출채권회전률_1 47. 순운전자본회전률 48. 순운전자본회전률_1 49. 현금비율 50. 현금비율_1	51. 순운전자본비율 52. 순운전자본비율_1 53. 차입금의존도 54. 차입금의존도_1 55. 투자집중도 56. 투자집중도_1 57. CASH FLOW 대 부채비율 58. CASH FLOW 대 부채비율_1 59. CASH FLOW 대 매출액비율 60. CASH FLOW 대 매출액비율_1 61. 자기자본회전률 62. 자기자본회전률_1 63. 유동자산회전률 64. 유동자산회전률_1 65. 비유동자산회전률 66. 비유동자산회전률_1 67. 총자본투자효율 68. 총자본투자효율_1 69. 재고자산 대 유동자산비율 70. 재고자산 대 유동자산비율_1 71. 자산총계 72. 자산총계_1	73. 자본총계 74. 자본총계_1 75. 매출액 76. 매출액_1 77. 법인세비용차감전(계속 사업)손익 78. 법인세비용차감전(계속 사업)손익_1 79. 영업이익 80. 영업이익_1 81. 당기순이익 82. 당기순이익_1 83. 투자활동으로 인한 현금 흐름 84. 투자활동으로 인한 현금 흐름_1 85. 종업원 86. 종업원_1

# Lasso

후보 피쳐(86) → 최종 피쳐(18)

변수명	회귀계수	변수명	회귀계수	변수명	회귀계수
기업순이익률	0.12900371740760339	자기자본회전률	0.04576370150141524	유동자산구성비율	0.013069328645978115
총자본증가율	0.12852457548534274	자기자본구성비율	0.03417375113745475	이자보상배율 (이자비용)	-0.0054686068078293
R & D 투자효율	0.12017454658706715	자기자본회전률_1	0.029369907786108877	자본총계_1	-0.0053399695707242
R & D 투자효율_1	0.08912827982104714	차입금비율_1	0.027141851671905266	순운전자본회전률_1	-0.00305151204465026
총자본회전률_1:	-0.06683350857569192	총자본순이익률	0.014567316617256932	종업원	0.000469492316731853
유동자산회전률	0.0573932956492206	이자보상배율 (이자비용)_1	-0.01408339249895183	종업원_1	-0.00030267751214034

---

# EDA

---

# Feature 별 기초통계량

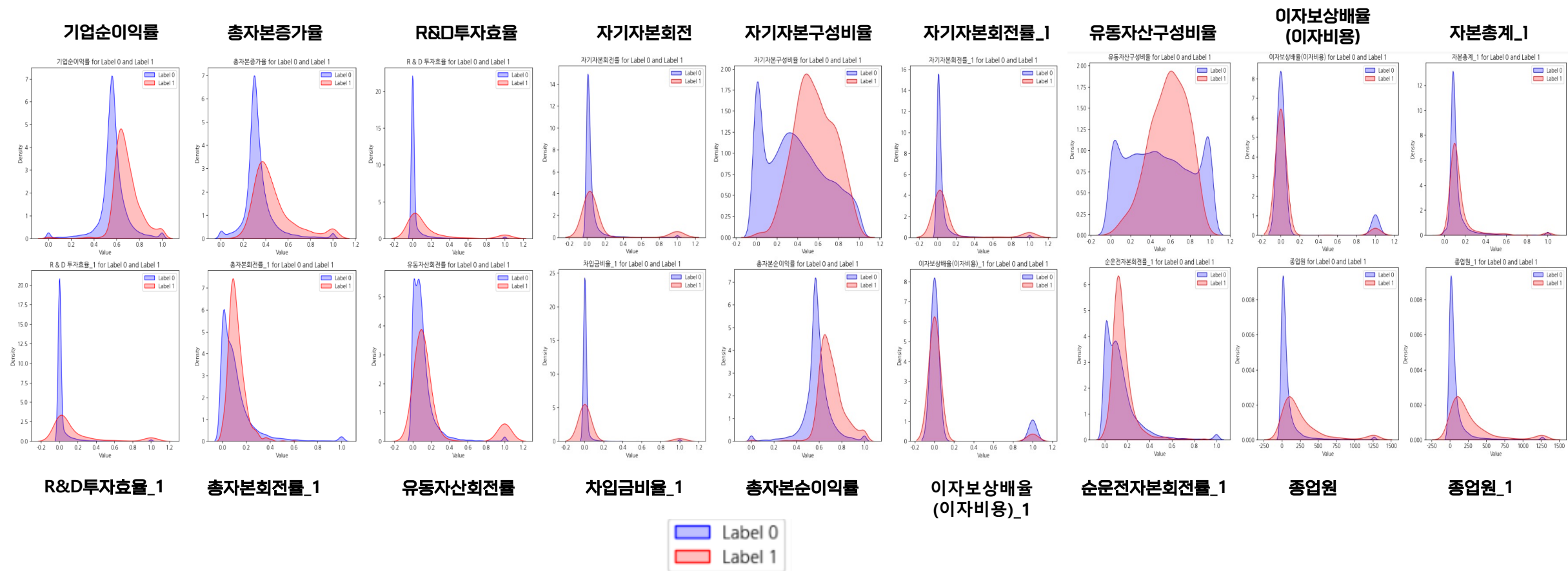
Lable 1

label1	기업순이익 률	총자본증가 율	R & D 투자 효율	R & D 투자 효율_1	총자본회전 률_1	유동자산회 전률	자기자본회 전률	자기자본구 성비율	자기자본회 전률_1	차입금비율 _1	총자본순이 익률	이자보상배 율(이자비용) _1	유동자산구 성비율	이자보상배 율(이자비용)	자본총계_1	순운전자본 회전률_1	종업원	종업원_1
count	805	805	805	805	805	805	805	805	805	805	805	805	805	805	805	805	805	805
mean	0.69	0.47	0.17	0.17	0.12	0.21	0.14	0.57	0.13	0.07	0.7	0.06	0.58	0.05	0.15	0.16	271.51	270.53
std	0.12	0.18	0.29	0.3	0.07	0.29	0.31	0.19	0.29	0.26	0.11	0.23	0.19	0.22	0.15	0.1	310.49	308.56
min	0	0.12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25%	0.62	0.34	0	0	0.07	0.07	0.02	0.43	0.01	0	0.63	0	0.45	0	0.09	0.1	74	74
50%	0.67	0.41	0.04	0.03	0.1	0.1	0.03	0.55	0.02	0	0.69	0	0.6	0	0.1	0.14	159	159
75%	0.74	0.53	0.16	0.18	0.15	0.15	0.05	0.71	0.04	0	0.76	0	0.73	0	0.13	0.18	326	326
max	1	1	1	1	0.6	1	1	1	1	1	1	1	0.99	1	1	0.89	1267	1267

Lable 0

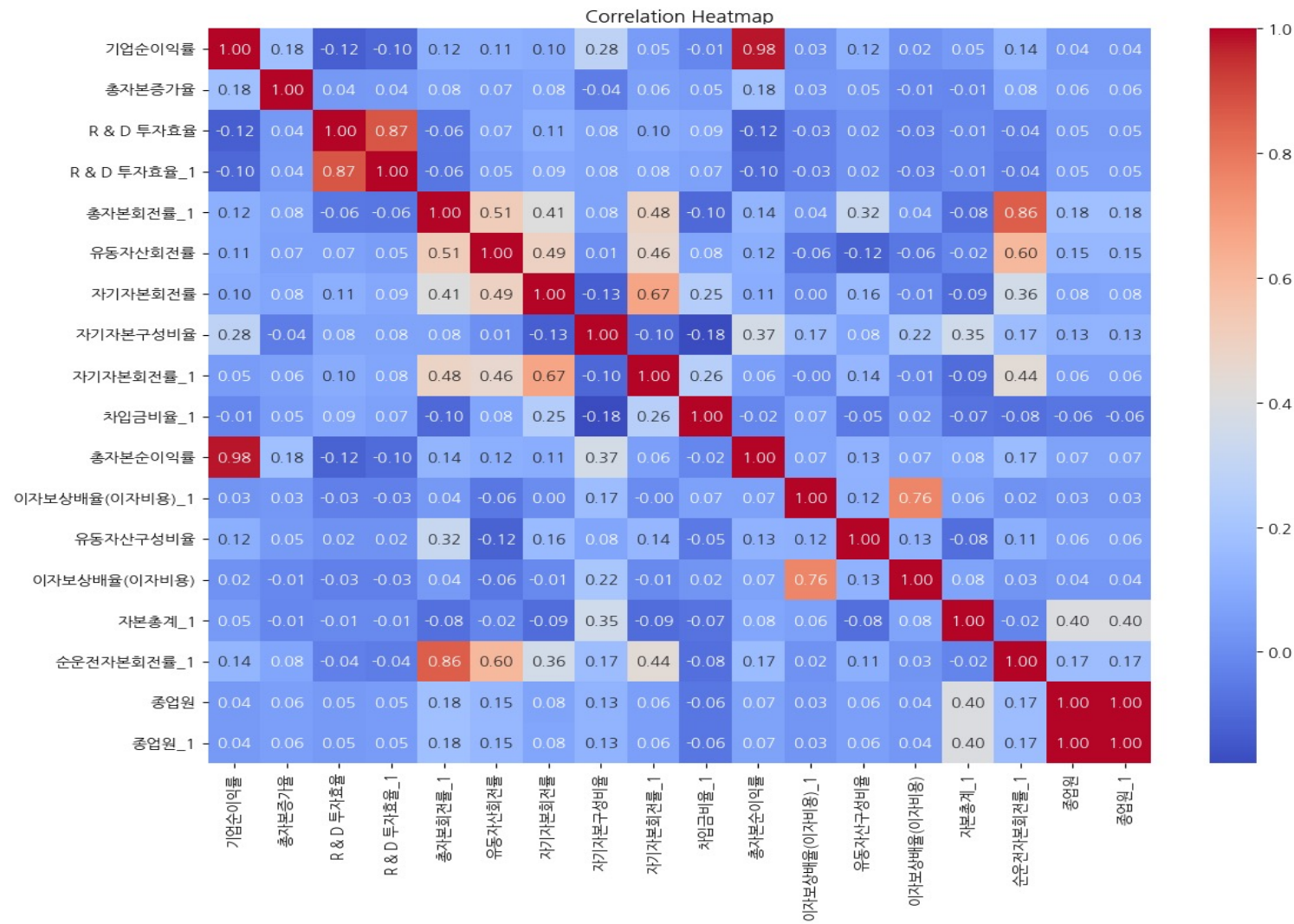
label0	기업순이익 률	총자본증가 율	R & D 투자 효율	R & D 투자 효율_1	총자본회전 률_1	유동자산회 전률	자기자본회 전률	자기자본구 성비율	자기자본회 전률_1	차입금비율 _1	총자본순이 익률	이자보상배 율(이자비용) _1	유동자산구 성비율	이자보상배 율(이자비용)	자본총계_1	순운전자본 회전률_1	종업원	종업원_1
count	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856	22856
mean	0.57	0.33	0.03	0.03	0.12	0.11	0.05	0.39	0.05	0.02	0.58	0.12	0.49	0.11	0.12	0.14	83.19	83.19
std	0.13	0.14	0.11	0.12	0.16	0.13	0.12	0.29	0.12	0.11	0.13	0.32	0.32	0.31	0.13	0.16	171.23	171.23
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25%	0.53	0.28	0	0	0.02	0.03	0	0.14	0	0	0.54	0	0.22	0	0.08	0.03	6	6
50%	0.56	0.31	0	0	0.07	0.08	0.02	0.36	0.02	0	0.57	0	0.48	0	0.09	0.1	32	32
75%	0.61	0.36	0	0	0.14	0.13	0.04	0.6	0.04	0.01	0.62	0	0.76	0	0.12	0.18	79	79
max	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1267	1267

# Lable 별 Feature 분포(정규분포)



# Feature간 상관관계

#Heatmap



# 주요 평가지표



## Precision

$$\frac{TP}{TP + FP}$$

$$= \frac{\text{상장성공한 기업}}{\text{상장 성공할거라고 예측한 기업}}$$

-> 제 1종오류: 잘못된 긍정  
(상장하지 않을 기업을 상장할거라고 예측)

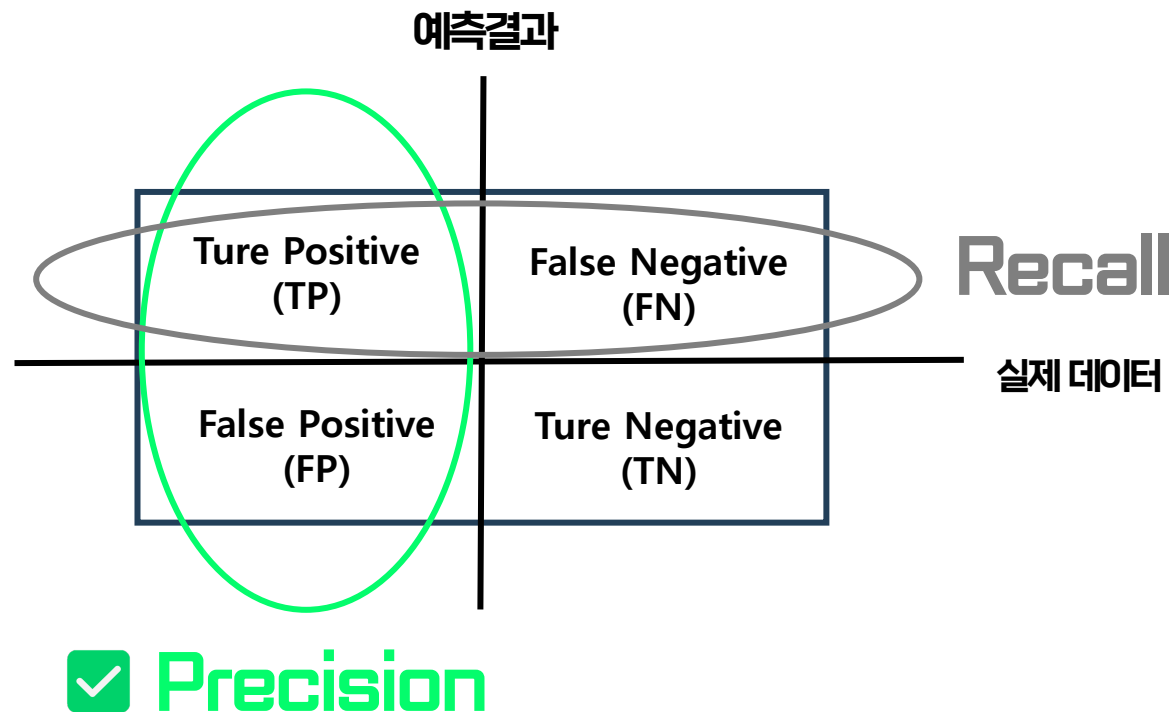


## Recall

$$\frac{TP}{TP + FN}$$

$$= \frac{\text{상장성공한 기업}}{\text{실제 상장 성공 기업}}$$

-> 제 2종오류: 잘못된 부정  
(상장할 기업을 상장 안할거라고 예측)



# 모델링 & 성능평가



# 모델링 및 성능평가

모델 종류	평가지표					Threshold (임계값)
	Accuracy	Precision	Recall	F1-Score	Roc-Auc	
Decision Tree	0.96	0.43	0.45	0.44	0.71	0.5
Random Forest	0.97	0.71	0.44	0.55	0.94	0.4
Logistic Regression	0.97	0.71	0.22	0.33	0.89	0.5
✓ Catboost	0.97	0.77	0.54	0.63	0.96	0.4
LightGBM	0.97	0.71	0.55	0.62	0.95	0.35
XGboost	0.97	0.73	0.48	0.58	0.95	0.5

# 모델링 및 성능평가 - 하이퍼파라미터



모델 종류	평가지표					Threas hold (임계값)	하이퍼파라미터
	Accuracy	Precision	Recall	F1-Score	Roc-Auc		
Decision Tree	0.96	1.0	0.09	0.17	0.88	0.5	'max_depth': 4, 'min_samples_split': 4, 'min_samples_leaf': 2, 'max_features': 0.31051534652224516
Random Forest	0.97	0.85	0.34	0.49	0.94	0.4	'n_estimators': 69, 'max_depth': 7, 'min_samples_split': 4, 'min_samples_leaf': 2, 'max_features': 0.9586217235283597
Logistic Regression	0.97	0.70	0.22	0.34	0.89	0.5	'C': 8.642102971873866, 'max_iter': 989
✓ Catboost	0.97	0.78	0.48	0.59	0.96	0.4	'learning_rate': 0.1997510992059283, 'iterations': 199, 'depth': 6, 'min_child_samples': 4, 'subsample': 0.9016308407360691, 'colsample_bylevel': 0.8537563845424978
LightGBM	0.97	0.70	0.51	0.59	0.95	0.35	'learning_rate': 0.15339748268313785, 'n_estimators': 126, 'max_depth': 4, 'min_child_samples': 2, 'subsample': 0.8182752890301975, 'colsample_bytree': 0.9035874689513481
XGboost	0.97	0.75	0.50	0.60	0.9	0.5	'learning_rate': 0.19779818283580508, 'n_estimators': 153, 'max_depth': 4, 'min_child_weight': 4, 'subsample': 0.8841848400135726, 'colsample_bytree': 0.9745108017963585

# 모델링 및 성능평가: 오버샘플링

모델 종류	평가지표					Threshold (임계값)
	Accuracy	Precision	Recall	F1-Score	Roc-Auc	
Decision Tree	0.93	0.26	0.57	0.75	0.36	0.8
Random Forest	0.96	0.44	0.48	0.46	0.94	0.7
Logistic Regression	0.92	0.25	0.60	0.35	0.89	0.8
Catboost	0.96	0.5	0.61	0.55	0.94	0.7
LightGBM	0.96	0.48	0.60	0.53	0.94	0.8
✓ XGboost	0.96	0.54	0.60	0.57	0.95	0.8

# 모델링 및 성능평가: 오버샘플링 하이퍼파라미터 수정



모델 종류	평가지표					Threshold (임계값)	하이퍼파라미터
	Accuracy	Precision	Recall	F1-Score	Roc-Auc		
Decision Tree	0.95	0.38	0.55	0.44	0.89	0.8	'max_depth': 7, 'min_samples_split': 5, 'min_samples_leaf': 2, 'max_features': 0.8328335839391356
Random Forest	0.95	0.42	0.53	0.47	0.93	0.7	'n_estimators': 182, 'max_depth': 7, 'min_samples_split': 5, 'min_samples_leaf': 2, 'max_features': 0.8644900875455609
Catboost	0.97	0.60	0.53	0.57	0.95	0.7	'learning_rate': 0.19251176157858185, 'iterations': 200, 'depth': 7, 'min_child_samples': 5, 'subsample': 0.8026166712632623, 'colsample_bylevel': 0.9730062168597597
LightGBM	0.97	0.64	0.53	0.58	0.95	0.8	'learning_rate': 0.19923246095672892, 'n_estimators': 192, 'max_depth': 7, 'min_child_samples': 4, 'subsample': 0.9013867731950866, 'colsample_bytree': 0.8161215080084934
XGboost	0.96	0.64	0.50	0.57	0.95	0.8	'learning_rate': 0.1998361843261015, 'n_estimators': 182, 'max_depth': 7, 'min_child_weight': 1, 'subsample': 0.8152297901852169, 'colsample_bytree': 0.973058106891242
Logistic Regression	0.95	0.38	0.38	0.89	0.38	0.8	'C': 8.689002207900677, 'max_iter': 323

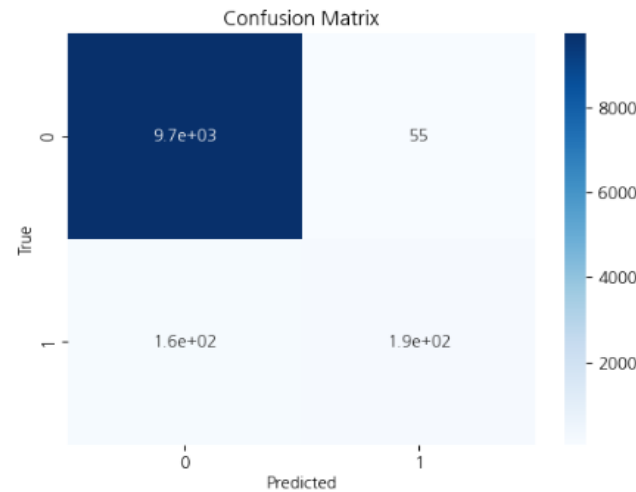
# StratifiedKFold

최종 선정 모델 : **CatBoost**

평가지표

평가지표					Threas hold (임계값)
Accura cy	Precisio n	Recall	F1-Scor e	Roc-Auc	
0.97	0.77	0.54	0.63	0.96	0.4

Confusion Matrix

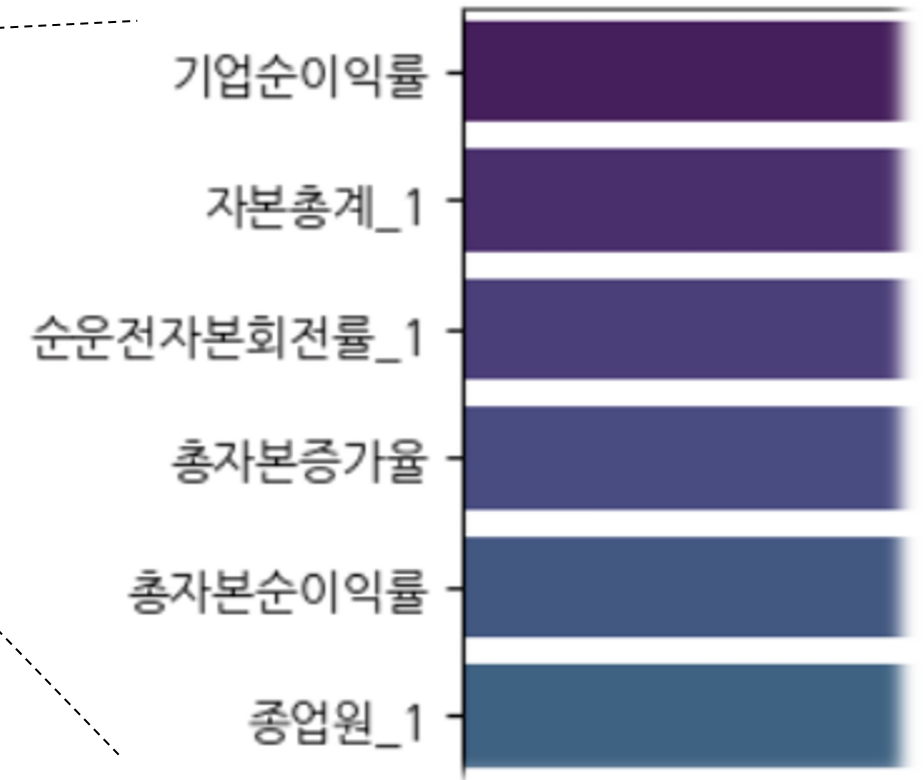
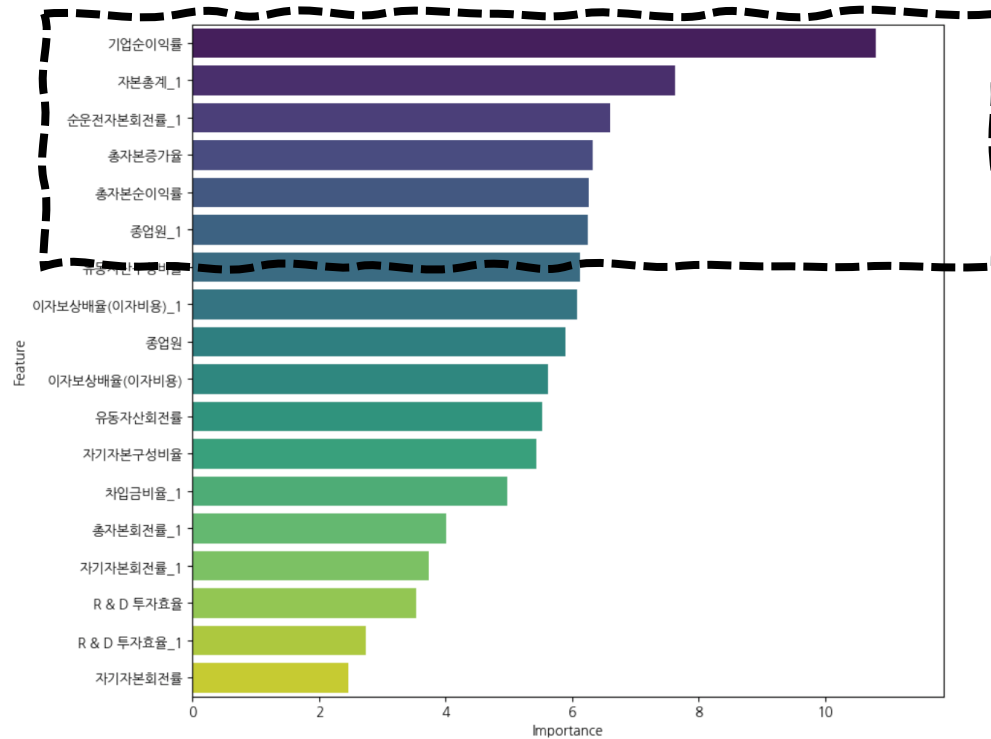


StratifiedKFold 결과

평균 정밀도 : 0,721927458593827  
 평균 재현율 : 0,49054012345679016  
 평균 정확도 : 0,9761634063184225  
 평균 f1 Score : 0,5811400810929128  
 평균 ROC AUC Score : 0,9644796622521998

# Feature Importance

Feature Importance using Catboost



# 연구 의의 및 한계점

## 의의

- ✓ 외부 투자자들에게 투자 지표로 활용할 수 있는 모델 제공
- ✓ 비상장회사가 성장을 위해 필요한 요인을 파악하여 한정된 자원을 효율적 배분하는 데 활용

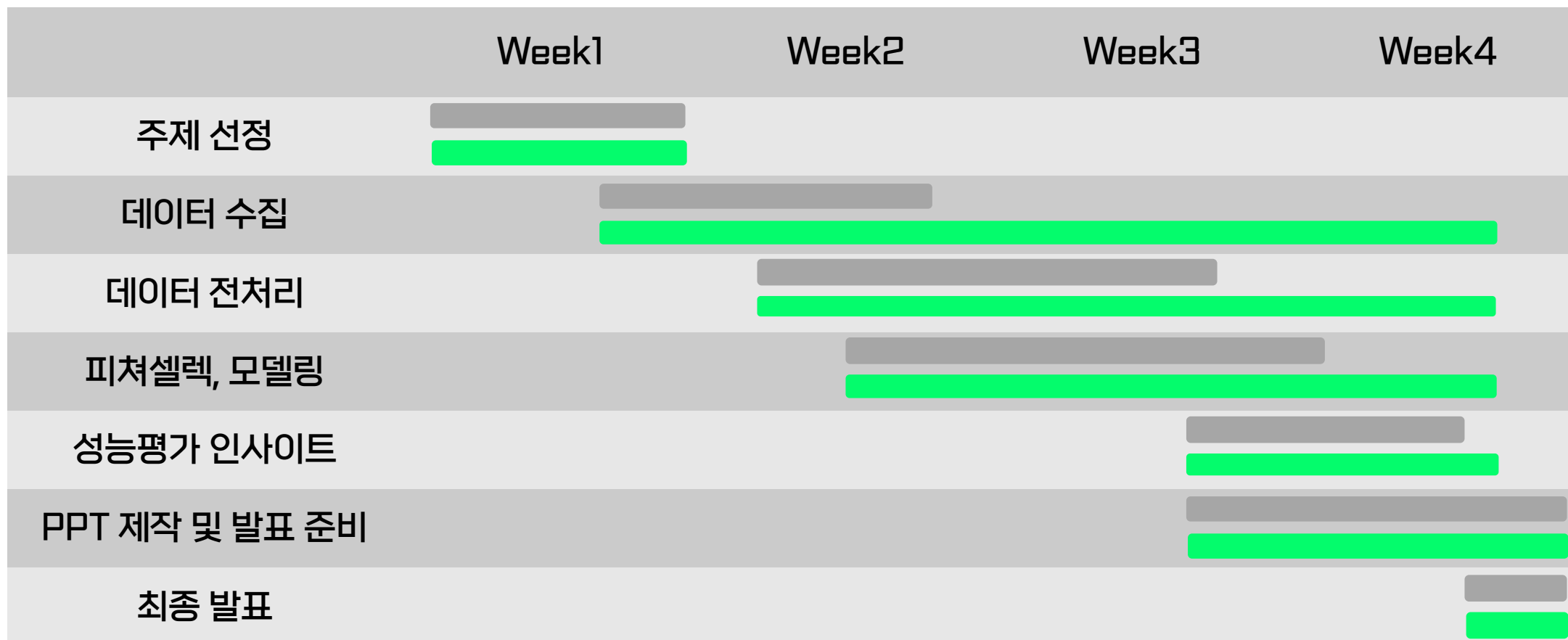
## 한계점

- ✓ 기술특례 기업의 경우 상장 기준 상이함으로 인해 예측 어려움
- ✓ 변수간 상관관계에 따른 모델 성능 변화를 고려하지 않음
- ✓ 데이터 불균형으로 인해 Precision과 Recall 모두 높은 모델 구현하는 데 어려움이 있었음
- ✓ Train과 Test set을 Random하게 구분

## 추후 연구 목표

- ✓ KOSPI & KOSDAQ 시장 구분하여 별도 모델 구현
- ✓ 상관관계가 높은 변수를 제거하고 모델 구현

# Time Table





# 참고문헌

- 나종호, 박성규 and 김창수. (2007). 비재무적 지표를 이용한 벤처기업의 신규상장(IP0)성공요인에 관한 연구. 중소기업연구, 29(2), 1-22.
- 우원석. (2015). 경영자 지분율이 IPO 공모비율과 저가발행현상에 미치는 영향. Journal of The Korean Data Analysis Society, 17(4), 2035-2054.
- 하나은행 하나금융경영연구소, 보고서 달아오른 비상장기업 투자  
<http://www.hanaif.re.kr/boardDetail.do?hmpeSeqNo=35232>

The background features a light grey 'X' in the top-left corner. On the right side, there are two overlapping green circles, with a darker green 'X' inside the upper circle. The bottom of the slide is a solid black bar.

**감사합니다**