

# [기후금융] 재무정보를 이용한 상장폐지 - XGB

# ✔ 과제

- 상장폐지 사유 필터 → 비슷한 사유끼리 카테고리화
- 칼럼 설명 → 공유문서 docs
- 금융회사 제거 → 금융기업 대분류를 기준으로 제거
- XGBoost

### 고민해보기

- Nan값이 하나라도 있는 결측치 row를 모두 drop하면, row개수: 111903 → 36084
  - 모든 columns에서 Nan값인 row는 drop하되, 일부 columns에서 Nan값인 row는 중앙값이나 평균값으로 fillna하는 것이 필요한가
  - ∘ 데이터 개수가 크게 줄어들긴 하지만, 36084개로도 성능이 충분함
- {상장유지를 상장폐지로 예측}인 예측오류가 예측정답보다 압도적으로 많음
  - 상장폐지 사유 중 '상장폐지 조건을 2년 계속 미충족할 경우'가 많으므로 이 부분을 이용한 파생변수 적절한 파 생변수를 추가할 것
    - 상장연도~폐지연도 변화율 기울기 (1차함수 또는 지수함수 기울기값)
- 한번(연도)이라도 상장폐지로 분류이 되면, 해당 상장폐지 기업으로 정확도 다시 측정

#### **▼** XGBoost parameters

- n\_estimators : 학습 모델의 수, 많아질수록 성능 향상의 가능성이 있으나, 속도가 느려짐
- learning rate : 학습률, 너무 크면 gradient 발산의 가능성이 있으며, 너무 작으면 학습이 느림
- max\_depth : 최대 탐색 깊이, 너무 크면 과적합의 가능성, 너무 작으면 학습 성능 저하
- min\_samples\_split : 분할 종료 최소 샘플 수, 큰 수면 과적합을 막지만 학습 성능 저하 가능성
- min samples leaf: leaf node가 되기 위한 최소 샘플 수, min samples split과 비슷한 용도
- random state : 특정 숫자로 고정 시에 실행 시마다 고정된 결과 반환

#### ▼ 1. 금융회사 제외



#### 대분류에서 '금융 및 보험업'에 해당되는 모든 기업 drop

• 산업분류카테고리 별 고유값 개수

산업분류카테고리	분류 업종 개수	금융회사 관련 업종	drop 여부
대분류	18	금융 및 보험업	0
중분류	65	금융 및 보험 관련 서비스업, 금융업, 보험 및 연금업	0

산업분류카테고리	분류 업종 개수	금융회사 관련 업종	drop 여부
소분류	169	금융 지원 서비스업 보험 및 연금관련 서비 스업 기타금융업 신탁업 및 집합 투자업 은 행 및 저축기관 보험업 재 보험업	0
세분류	380	금융 지원서비스업	0
세세분류	623	금융 지원서비스업	0

### • 산업 분류 카테고리

。 스팩(SPAC)은 Special Purpose Acquisition Company의 약어로 우리말로는 기업인수목적회사라 하며, 오로지 기업 인수만을 목적으로 하는 페이퍼 컴퍼니

중분류	소분류	세분류	세세분류
금융 및 보험 관련 서비스업	금융 지원 서비스업	금융 지원 서비스업	금융 지원 서비스업
	: 스팩(SPAC) + etc	기타 금융 지원 서비스업	그 외 기타 금융 지원 서비스업
	etc에는 카카오페이 등	기업 포함	기타 금융 지원 서비스업
		증권 및 선물 중개업	증권 중개업
	보험 및 연금관련 서비스업	보험 및 연금관련 서비스업	보험 대리 및 중개업
금융업	기타 금융업	그 외 기타 금융업	그 외 기타 금융업
			그 외 기타 분류 안된 금융업
			지주회사
		기타 금융업	기타 금융업
		여신금융업	그 외 기타 여신금융업
			금융리스업
			신용카드 및 할부금융업
			여신금융업
	신탁업 및 집합투자업	신탁업 및 집합투자업	기타 금융 투자업
			신탁업 및 집합투자업
	은행 및 저축기관	신용조합 및 저축기관	상호저축은행 및 기타 저축기관
			신용조합
		일반은행	국내은행
보험 및 연금업	및 연금업 보험업	생명 보험업	생명 보험업
		손해 및 보증 보험업	손해 보험업
	재 보험업	재 보험업	재 보험업

### ▼ 2. 상장폐지 사유

### 상장폐지 사유

Aa 상장폐지 사유		:= 부실기 업
해산 사유 발생	3200	0
감사의견 의견거절	1696	0
<u>피흡수합병</u>	1504	X

. LITHRITI II.O	를 개	:= 부실기
Aa 상장폐지 사유	수	업
<u> 자본전액잠식</u>	1504	0
기업의 계속성, 경영의 투명성 또는 기타 코스닥시장의 건전성 등을 종합적으로 고려하여 상장폐지가 필요하다고 인정	896	0
<u>감사의견거절(감사범위제한)</u>	864	0
<u>주식분산기준 미달</u>	800	0
주식분산기준 미달후 6월 경과	704	0
회사정리절차 폐지 결정	576	0
기업의 계속성, 경영의 투명성 또는 기타 코스닥시장의 건전성 등을 종합적으로 고려하여 상장폐지 기준에 해당한다고 결정	544	0
<u>감사의견거절(감사범위 제한)</u>	512	0
기업의 계속성 및 경영의 투명성 등을 종합적으로 고려하여 상장폐지기준에 해당한다고 결정	512	0
감사의견 거절(감사범위 제한 및 계속기업 존속능력에 대한 불확실성)	416	0
신청에 의한 상장폐지	416	X
기업의 계속성 및 경영의 투명성 등을 종합적으로 고려하여 상장폐지 기준에 해당한다고 결정	384	0
증권거래소 상장	352	X
영업활동정지 6월 계속	352	0
<u>상장폐지신청</u>	352	X
감사의견거절(감사범위 제한 및 계속기업으로서의 존속능력에 대한 불확실성)	352	0
<u>상</u> 장폐지유예기간종료	320	0
감사의견 거절(감사범위 제한)	288	0
자본전액잠식 등	256	0
최종부도발생	256	0
최종부도	256	0
주식분산기분 미달	256	0
최종부도 발생	224	0
지주회사(최대주주등)의 완전자회사화 등	224	X
상장폐지 신청	224	X
감사의견 부적정	192	0
최종부도 및 은행거래정지	192	0
주식분산기준 미달후 1년 경과	192	0
<u>발행회사 취소요청</u>	160	X
<u>주식분산기준미달</u>	160	0
<u> 자본전액잠식(2년)</u>	160	0
자진 등록취소	160	X
기타 등록취소	160	X
<u>부도</u>	160	0
<u>기업의 계속성 및 경영의 투명성 등을 종합적으로 고려하여 상장폐지기준에 해당</u>	160	0

Aa 상장폐지 사유		∷ 부실기 업
발행한 어음 또는 수표가 주거래은행에 의하여 최종부도로 결정되거나 거래은행에 의한 거래정지	160	0
감사의견거절(감사범위 제한 및 계속기업 존속능력에 대한 불확실성)	128	0
감사의견거절(감사범위제한 및 계속기업 존속 불확실)	128	0
감사의견 의견거절 2년 계속	128	0
등록법인의 취소신청	128	X
당해 코스닥시장상장법인의 계속성 및 경영의 투명성 등을 종합적으로 고려한 결과 상장폐지기준에 해당한다고 결정	128	0
자본전액잠식 2년 계속	128	0
<u>당좌거래정지</u>	128	0
<u>자진등록취소</u>	128	X
<u> 주된영업의 양도</u>	128	X
회사정리절차 개시 신청 기각 결정	128	0
사업보고서 법정제출기한 익일부터 10일이내 미제출	128	0
계약이전	32	X
<u>유가증권시장 상장</u>	32	X
액면가액일정비율 미달	32	0
보통주 시가총액미달	32	0
시가총액 미달	32	0
<u>회생절차폐지결정사유 및 당해 코스닥시장상장법인의 계속기업 가능성 유무를 고려한 결과 상장폐지기준에 해당한다고 결정</u>	32	0

### ▼ 3. 재상장

• 과거 최초 상장했을 적 기업정보X

### ▼ 4. 변수 선정 실험 결과

<결과비교>	2차	3차	4차	5차	6차
상장폐지를 상장유지로 예측	248	10	10	10	8
상장유지를 상장폐지로 예측	719	85	85	85	85

### ▼ 1차



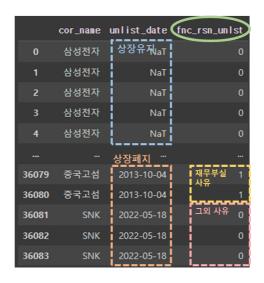
## 🍁 변수 선정 및 모델

- 시계열 고려 X, <u>상장폐지 연도가 아니어도 상장폐지 기업이면 '1'로 라벨</u>
- **독립변수**: 재무데이터 + 재상장 라벨링 변수 ['relist'](bool)
- **종속변수**: 상장폐지사유 라벨링 ['fnc\_rsn\_unlst'](object) → [0, 1, 2]

  - 1: 재무부실사유 상장폐지 (폐지 전 연도 + 폐지 후 마지막 결산연도)
  - 2: 그외 사유 상장폐지
- 모델
  - XGBClassifier(n\_estimators=500, early\_stopping\_rounds=1, max\_depth=3)

### • 데이터

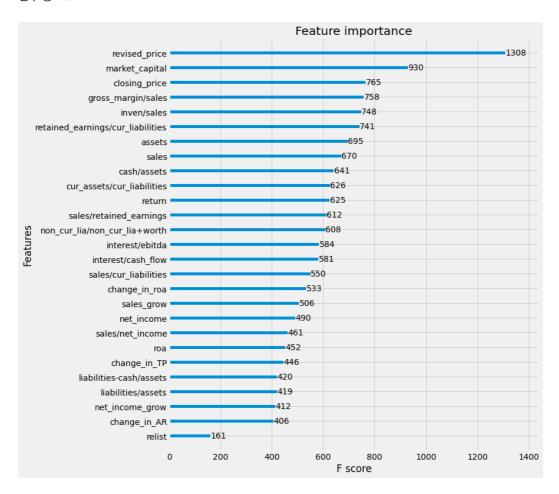
○ 종속변수 ['fnc\_rsn\_unlst']



### • 결과

• Stopping. Best iteration: [725] <u>0.164611</u>

。 변수 중요도



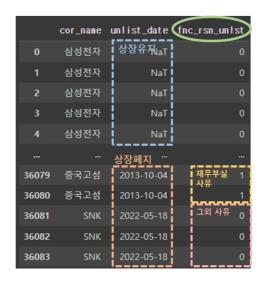
▼ **2차 :** 정확도 <u>0.8628</u>

### 👍 변수 선정 및 모델

- 시계열 고려 X, 상장폐지 연도이 아니어도 상장폐지 기업이면 '1'로 라벨
- **독립변수**: 재무데이터, 재상장 라벨링 변수 ['relist'](bool) 추가
- **종속변수**: 상장폐지사유 라벨링 ['fnc\_rsn\_unlst'](object) → [0, 1]
  - 0: 상장폐지 X, 그외 사유 상장폐지
  - 1: 재무부실사유 상장폐지(폐지 전 연도 + 폐지 후 마지막 결산연도)

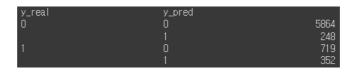
#### • 데이터

o 종속변수 ['fnc\_rsn\_unlst']

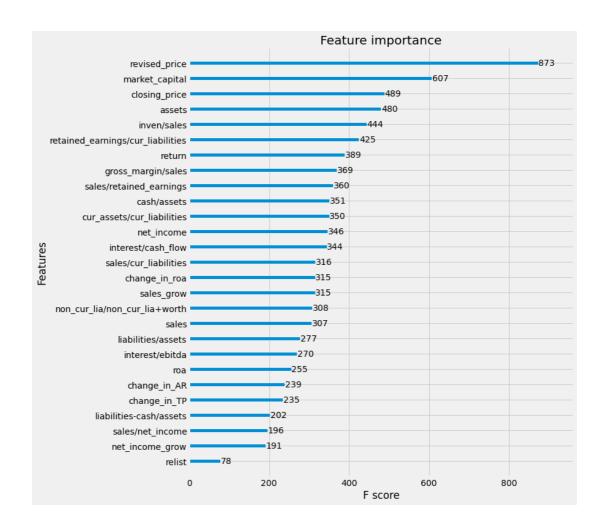


#### • 결과

- o validation: [499] <u>0.13773</u>
- 。 예측 오류 개수



- 상장폐지를 상장유지로 예측: 248개
- 상장유지를 상장폐지로 예측: 719개
- 。 변수 중요도



#### ▼ 3차 : 정확도 0.9868

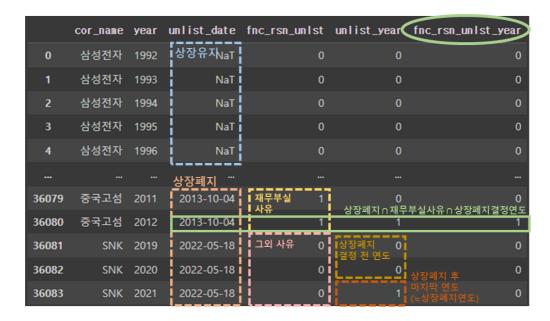


### 👍 변수 선정 및 모델

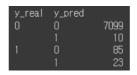
- 시계열 고려 X, 상장폐지 후 마지막 결산재무정보 row에만 '1'로 라벨링
- 독립변수: 재무데이터, 재상장 라벨링 변수 ['relist'](bool) 추가
- **종속변수**: 상장폐지사유 라벨링 ['fnc\_rsn\_unlst\_year'](object) → [0, 1]
  - 0: 상장폐지 X, 그외 사유 상장폐지, 재무부실사유 상장폐지(폐지 전 연도)
  - 1: 재무부실사유 상장폐지 연도(폐지 후 마지막 결산연도)

#### • 데이터

○ 종속변수 ['fnc\_rsn\_unlst\_year']



- o Stopping. Best iteration: [98] 0.013163
- 예측 오류 개수



- 상장폐지를 상장유지로 예측: 10개
- 상장유지를 상장폐지로 예측: 85개

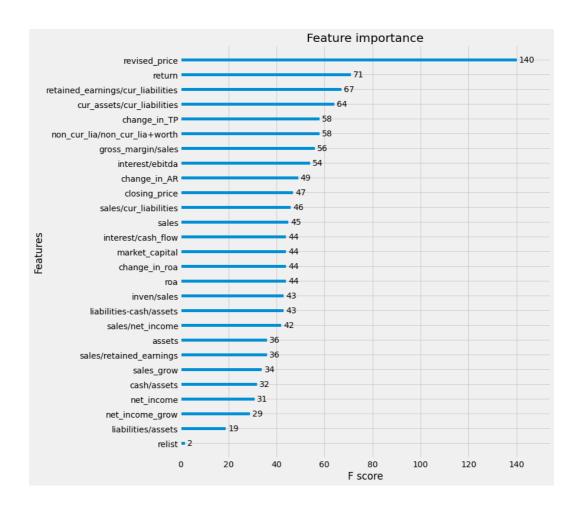
#### ▼ 예측에러 row 원인 분석

- '감사의견거절'사유 카테고리가 압도적으로 많은 예측오류 비중을 차지하지만, 단순히 원래 기존 데이터에 서 비중이 큰 사유이므로 이와 같은 결과가 나옴
- 왜냐하면, 비중이 높은 다른 사유인 '해산사유발생', '기타 등록취소'도 원래 비중이 컸으며, '감사의견거 절'카테고리에서 상세사유를 보았을 때도 눈에 띄는 점은 없었음





○ 변수 중요도



#### ▼ 4**차** : 정확도 <u>0.9868</u>



## 👍 변수 선정 및 모델

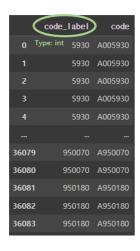
- 시계열 고려 X, 상장폐지 후 마지막 결산재무정보 row에만 '1'로 라벨링
- 독립변수: 재무데이터, <u>기업코드 레이블 ['code\_label']</u> 변수 추가
- **종속변수**: 상장폐지사유 라벨링 ['fnc\_rsn\_unlst\_year'](object) → {0, 1}
  - [0]: 상장폐지 X, 그외 사유 상장폐지, 재무부실사유 상장폐지(폐지 전 연도)
  - 1: 재무부실사유 상장폐지 연도(폐지 후 마지막 결산연도)

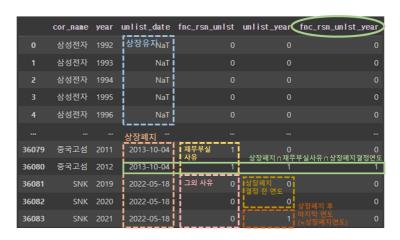
#### • 데이터

독립변수

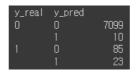
종속변수 ['fnc\_rsn\_unlst\_year']

['code\_label']





- o Stopping. Best iteration: [4] 0.013163
- 예측 오류 개수



- 상장폐지를 상장유지로 예측: 10개
- 상장유지를 상장폐지로 예측: 85개

#### ▼ 5차 : 정확도 0.9868



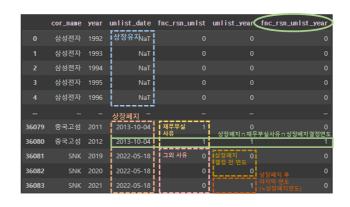
### 🧼 변수 선정 및 모델

- 시계열 고려 X, 상장폐지 후 마지막 결산재무정보 row에만 '1'로 라벨링
- 독립변수: 재무데이터, 연도 ['year'] 변수 추가
- **종속변수**: 상장폐지사유 라벨링 ['fnc\_rsn\_unlst\_year'](object) → {0, 1}
  - [0]: 상장폐지 X, 그외 사유 상장폐지, 재무부실사유 상장폐지(폐지 전 연도)
  - 1: 재무부실사유 상장폐지 연도(폐지 후 마지막 결산연도)

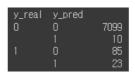
### 데이터

독립변수 ['year'] 종속변수 ['fnc\_rsn\_unlst\_year']





- o Stopping. Best iteration: [4] 0.013163
- 예측 오류 개수



- 상장폐지를 상장유지로 예측: 10개
- 상장유지를 상장폐지로 예측: 85개

#### ▼ 6차 : 정확도 0.9871



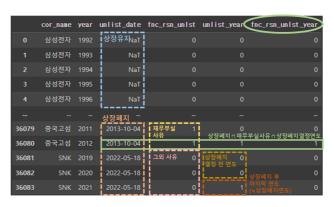
### 🧼 변수 선정 및 모델

- 시계열 고려 X, 상장폐지 후 마지막 결산재무정보 row에만 '1'로 라벨링
- 독립변수: 재무데이터, 기업코드 레이블 ['cor\_label'] & 연도 ['year'] 추가
- **종속변수**: 상장폐지사유 라벨링 ['fnc\_rsn\_unlst\_year'](object) → {0, 1}
  - 。 ◎: 상장폐지 X, 그외 사유 상장폐지, 재무부실사유 상장폐지(폐지 전 연도)
  - 1: 재무부실사유 상장폐지 연도(폐지 후 마지막 결산연도)

### • 데이터

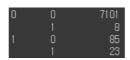
독립변수 ['year'] ['code\_label']

### 종속변수 ['fnc\_rsn\_unlst\_year']

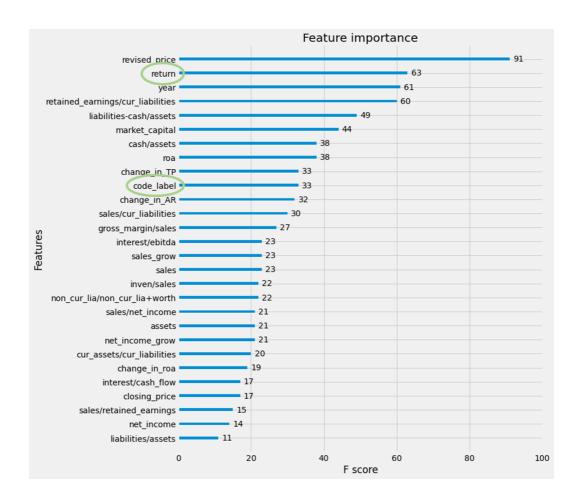




- $\circ~$  Stopping. Best iteration: [21] 0.012886
- 예측 오류 개수



- 상장폐지를 상장유지로 예측: 8개
- 상장유지를 상장폐지로 예측: 85개
- 변수 중요도



### ▼ 5. 시계열 XGBoost

▼ 1차 : 정확도 0.9927

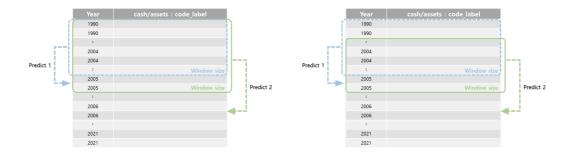


#### 변수 선정 및 모델

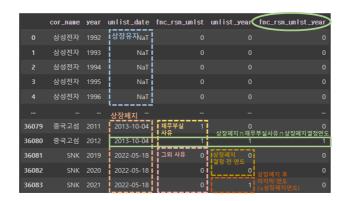
- 시계열 고려 X, 상장폐지 후 마지막 결산재무정보 row에만 '1'로 라벨링
- 독립변수: 재무데이터, <u>기업코드 레이블</u> ['cor\_label']
- Index: 연도 ['year']
- **종속변수**: 상장폐지사유 라벨링 ['fnc\_rsn\_unlst\_year'](object) → {0, 1}
  - 0: 상장폐지 X, 그외 사유 상장폐지, 재무부실사유 상장폐지(폐지 전 연도)
  - 1: 재무부실사유 상장폐지 연도(폐지 후 마지막 결산연도)
- 데이터
  - o 독립변수 ['year'], ['code\_label']

**Recursive window** 

**Rolling window** 



### o 종속변수 ['fnc\_rsn\_unlst\_year']



### • 결과

#### ○ 예측 오류 개수

Window size	2005년 이전 (14)	2013년 이전 (22)
상장폐지를 상장유지로 예측	16	12
상장유지를 상장폐지로 예측	108	100