

중소형 상장 기업 대상

기업 정보의 동태적 분석을 활용한 부실 기업 예측

데이터 수집

- 수집 기간: 2010 ~ 2019
- 분석 기간: 2011 ~ 2019
- 종목범위: KOSPI + KOSDAQ
- 제외종목: 금융기업, 스펙, 외국계 기업, 시가총액 상위 100위 이내 기업
- 출처: TS2000

재무 데이터

- 기업 수: 2,487사
- 데이터 개수: 20,505개
- 피쳐 개수: 134개

주가 데이터

- 거래량, 액면가, 종가

데이터 확인

- 기초 통계량 확인

부실의 정의

- 상장폐지, 관리 종목

독립변수

수익성(8), 안정성(7)
생산성(2), 성장성(6)

기타(2)

- 영업활동현금흐름, EV/EBITDA

데이터 전처리

파생 변수

거래량회전율, 수정거래량, 거래량증가율, 로그시가총액, 시가총액증가율, DOL, DFL, 종업원수증가율, 금융비용부담률, 고정비율, R&D비율, 채무부담비율

EDA 및 결측치 · 이상치 처리

EDA

결측치처리

이상치처리: IQR, Winsorizing

Data Split

Train: 2011 ~ 2017

Test: 2018 ~ 2019

Resampling

SMOTE

Random-over Sampling

Scaling

Min-Max Scaling

Standard Scaling

피쳐 선정

Lasso

Step-wise

T-test(회귀가정)

선정된 피쳐

2번 중복: 13개

3번 중복: 9개

통계 검정

등분산성: Bartlett

정규성: Shapiro-Wilks

독립성: Durbin-Watson

평가 및 검증

사용 모델

Decision Tree

Random Forest

XGBoost

KNN

SVM

Logistic Regression

Stacking

Tabnet

하이퍼파라미터 조정

GridSearchCV

RandomizedSearchCV

성능 평가

Accuracy

Precision

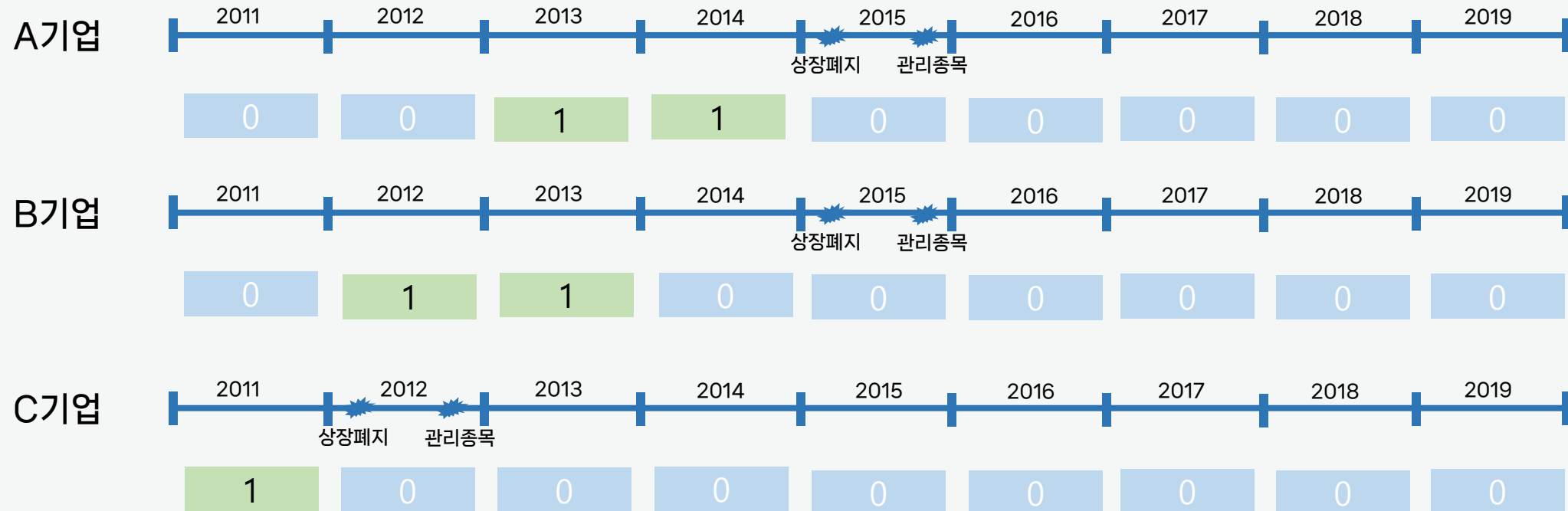
Recall

F1 score

ROC-AUC

Robust-Test

타겟 설정



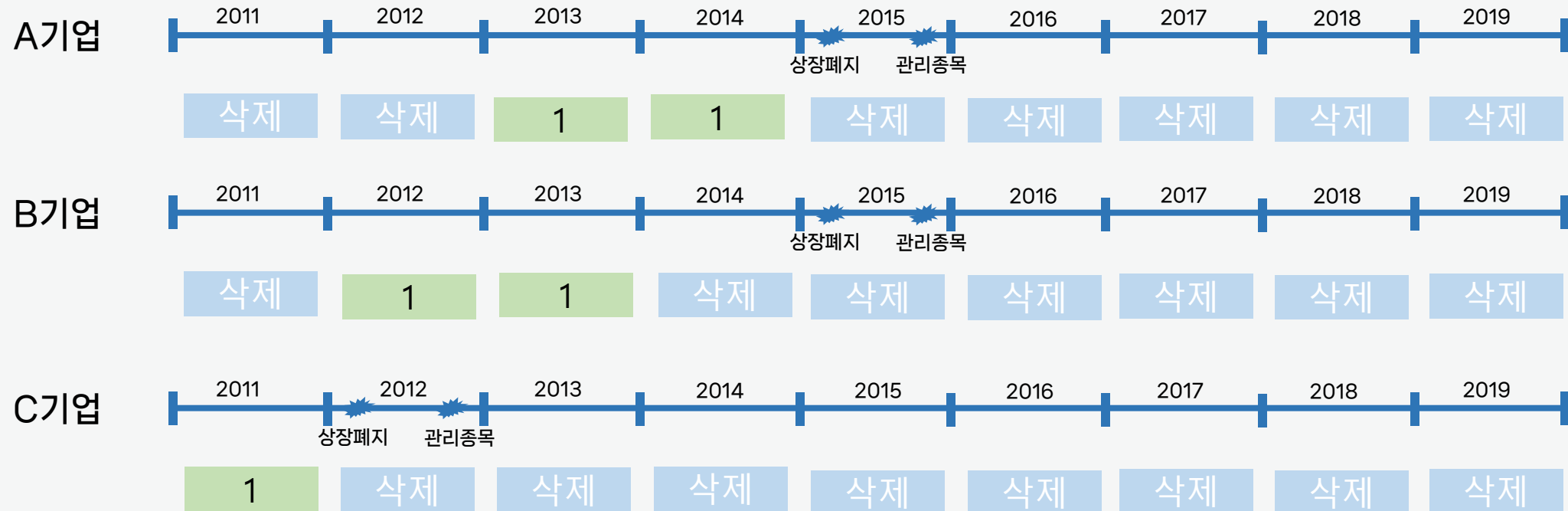
Y = 0(정상)

Y = 1(부실)

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

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

타겟 설정



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