# 1. 기업 관련

# 임의 선별

# 삼성전자

# NAVER

# 카카오

# 랜덤 선별

# rnd.seed(48)

# stock\_list.iloc[rnd.randint(len(stock\_list))]

# 금호석유

# 티움바이오

# 테크윙

# 제테마

# 주성엔지니어링

# 고바이오랩

# 고영

# 2. 학습 기간 관련

# 훈련 기간

2021/01/04 ~ 2021/09/03

# 검증 기간 (2회 검증)

2021/09/06 ~ 2021/09/10

2021/09/13 ~ 2021/09/17

# 테스트 기간

2021/09/27 ~ 2021/10/01

<분석 프로세스>

1. 데이터 로드 2021/01/04 ~ 2021/10/01

2. 영업일로 정렬된 데이터프레임 생성

3. 피처 추가

4. 시각화

5. 불필요한 피처 드랍 및 정규성을 확보하기 위한 transformation

-> 선택 피처

date, close, fast\_d, obv, fore\_mv20, inst\_mv20, kospi, trading\_amount\_mv20

(date 피처는 학습 전 드랍)

-> log1p 변환 피처

close, fast\_d, kospi, trading\_amount\_mv20

6. train validation split 1차

train 기간 :2021/01/04 ~ 2021/09/03

validation 기간 : 2021/09/06 ~ 2021/09/10

7. 모델링 및 성능평가

평가지표 : 모든 기업 평균 Normalise MAE

8. train validation split 2차

train 기간 : 2021/01/04 ~ 2021/09/10

validation 기간 : 2021/09/13 ~ 2021/09/17

9. 모델링 및 성능평가

평가지표 : 모든 기업 평균 Normalise MAE

10. test 데이터셋 예측

train : 2021/01/04 ~ 2021/09/24

test : 2021/09/27 ~ 2021/10/01