

1) 문제 구성하고, 특징 척도 식별, 성능 측정 → **5000명 선호 상품 선정**

[교차 검증]

2) 데이터 전처리 후 검증

- 금융 데이터 경우, 시계열 특성 가진

3) 데이터 이해하고, 정보성 특징 생성

↳ 그걸로 쓸 경우, 미래 데이터를 가지고 과거 데이터 예측 가능성 있음

4) 데이터에 적합한 알고리즘 선택

⇒ **Sklearn.model.selection 모듈의 TimeSeriesSplit**은 선행 데이터 예측 split

5) 훈련. 테스트 조정

# 백테스팅 Tool인 Quantopian 서버에 올리...

Alpha factor와 features to predict Stock Returns

↳ 사용자를 정하기 위해 트레이닝에는 거의 가치 X → 과회전화, 정량 X,  
데이터 아키텍처만 가독

\* ta-lib : 주가 차트 기술적 지표 계산 라이브러리 (RSI, BBANDS, MACD, ATR)

불안정성

market cap 시가총액  
ipo year 기업공개 연도 (상장시기)  
ticker : 미국주식에 각 종목 구별 번호

np.quantile 백분위수

price.resample('M') : price에서 달단위로 값 뽑기

→ **1월/2월/3월/4월/5월/6월/7월/8월/9월/10월/11월/12월로** **target도!!**  
기간 나눠서 나눠서 계산 (lag, clip, min, max)

group by 이용해서 (같은 값(종목) 묶은 값제일과 확인 가능 method 이용) 120 일만 index만 keep

Rolling OLS 이용해 25일 기준, drop('const')

Momentum 고려

Noise 제거 alpha한 사용 (불가)