# 0710 모델 개선

- 고정

  click_time_hour, click_time_minute　(+2)

  - 기존에는 second 변수도 넣었지만 제외하는걸로

시계열 데이터에서

- oofp(K) vs 8:2 split(C) vs 전체 다 씀

훈련 데이터 사이즈

- sample(S) vs full(L)

**기초 변수**

- ip, app, os, device, channel

  다 남겨두기 (+5)　vs 중에서 app만 남기기 (+1)

**시간 카운트 변수**

- 옵션 1: 1시간, 6시간 사이 클릭 횟수, ip를 기준으로 (2개)
- 옵션 2: 31개 조합에 대해서 진행
- 옵션 1 (+2) vs 옵션 2 (+62)

**시간 차이 변수(train_new_click_final.csv)**

- 고려 안함 (+0) vs 앞뒤로 31개 조합에 대해 (+62)

**LDA(train_new_final, test_new_final)**

- 안 넣냐 (+0) vs 넣냐 (+ 100)

다음 미션: 어떤 룰로 답지를 만들 것인지

- 정답지 만들기
  - 훈련 데이터
- 앞의 실험을 반복. rule click 평균 확률
  - click 확률 계산. 얼마 이하인지를 경계값
  - user를 어떻게 정의할 것인가
    - 기초 변수 5개 조합으로 블락 여부 결정
    - 더 적은 개수의 조합도 고려

- 그래프의 개형(Flaud)
  - 룰이 있을거고
  - 룰에 있어서 click 전체 중 사기의 비율
  - 유저 2~3% (목표: 3~5%)
    - 테스트 해봐야 실제로 결과가 나옴

코로나 논문:

유저 → 나라

- 어떤 나라에서 어떤 나라로 이동했는지? 관찰 필요

click 이벤트

정답지 까보기(의사 결정)

- 감염률이 높지 않아도 샘플로 exploration

        - 예측할 때의 불확실성
        - 시간 당 까볼 수 있는 정답지 개수 제한

      click Flaud →

        - 밝혀졌을 때 할 수 있는 액션
        - bandit

      딜레이 피드백(리소스 alocation)

        - 6시간 변수

- 매뉴얼한 룰
  - 