**231110 미팅 자료**

**<To do list 공유>**

* 현규님 : 팩트 체크 -> 크롤링 -> train dataset 구비
  + ⇒ 크롤링 전부 됨. 생각보다 소요시간이 걸리고 있음.
    - 링크 안으로 들어가서 본문을 긁어오는 건 어려움
    - 제목 / 라벨링까지는 가능함.
* 주현님 : 서비스 기획 / 웹 틀 구현
  + 최종 서비스 Output 형태 / 웹 상 구현 ⇒ 기획 완료
* 현구 : Fake news AI 모델 고정 => Train / Test 셋만 넣을 수 있게

[test.ipynb](https://prod-files-secure.s3.us-west-2.amazonaws.com/80e822f5-8452-4555-8390-64aeee836f3a/953f9342-9b6b-4552-8daf-bea833860ac3/test.ipynb)

* + 모델 구현 완료

=============================================================

**<논의사항>**

* + 모델 디벨롭을 더 할 것인가?
    - 구할 수 있는 데이터로 모델 성능 체크
      * 안 좋으면 ⇒ 디벨롭
      * 좋으면 ⇒ 디벨롭 불 필요
  + 모든 기사에 대해서 fake detection을 적용할 것인가?
    - Fake detection을 안해도 되는 기사들 존재 ⇒ 필요성 합의
    - 단, Filtering 할 기준이 모호하여, 현재로선 “정치/사회/이슈” 등 기사 카테고리에 따라 정리할 것
      * Train data(Factcheck) : 카테고리 O
      * Test data(네이버뉴스) : 카테고리 유무 점검 필요
  + 크롤링은 언제할 것인가?
  + 각 언론사별 신뢰성을 보이기 위해 최소 1달치의 자료는 필요하다 판단함
    - ⇒ 사전에 CSV를 만들어둔 후 시각화하기
  + 현재 기획안을 기반으로 구현할 수 있는가?
  + React 모듈에 있을 것으로 기대됨
  + 자세한 설명이 나와있는 곳이 있음.

=========================================================

**<To do list>**

* + Raw data(Train 용)
  + 최대한 많이 긁어보기

[20231110\_1346\_전체 라벨.csv](https://prod-files-secure.s3.us-west-2.amazonaws.com/80e822f5-8452-4555-8390-64aeee836f3a/3dc1598b-e784-4bad-b16a-6c26560b7b18/20231110_1346_%EC%A0%84%EC%B2%B4_%EB%9D%BC%EB%B2%A8.csv)

* + 네이버 뉴스 → 카테고리(ex- 정치 / 사회 등등) 구분 가능 유무 점검
  + Raw Data 참고해서 데이터 필러링 기준 세팅
  + 필터링 된 데이터로 모델 성능 체크
  + 모델의 Input/Output CSV
    - Raw output CSV :
      * 기사 1 / 언론사 / Prediction : False
      * 기사 100
    - Raw output CSV - Filtering / Summarization : ⇒ Test.ipynb 임시 코드를 작성해서 공유 가능
      * 조선 일보 : 75.5%
  + React 로 기능 구현 가능한지 점검
  + [핵심] 웹 서비스 → 어떻게 구현할지 ⇒ 병렬 필요
  + HTML - 서버 / React ⇒ 주현님
  + Back 엔드

============

추가 논의 / 해야할 일

* + 교수님 질문에 대한 답변 ⇒ 현구
  + 11월 20일 논문 발표 준비 ⇒ 다음 미팅까지 해당 논문을 1회독

⇒ 다음주 미팅 언제할까요?

* + 목요일 4시 Zoom 으로
  + 최종 output 발표 : 12월 4일 / 6일
    - 구현에 집중!