**231116 미팅 자료**

**<To do list>**

1. 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)

1. 네이버 뉴스 → 카테고리(ex- 정치 / 사회 등등) 구분 가능 유무 점검
2. Raw Data 참고해서 데이터 필러링 기준 세팅

* 정치 쪽 기사는 다루지 않는 게 좋을 듯 한데..
  + 오히려 중립 / 보수 / 진보를 나눌 수 있다면 오히려 깔끔하려나?
  + ⇒ Echo chamber 효과를 멈추기 위해서
* **1) 뉴스 ⇒ ‘팩트체크’ 란 활용 ⇒ 이걸로!**
  + 장점 : 그대로 사용 가능
  + 단점 : 프로젝트 목표 다운 그레이드
* **2) 인스타 - 인사이트**
  + 문제점 : 어그로성 제목 ↔ 실제 컨텐츠 사이의 갭
  + 장점 : 적절 / 어그로성이 크다보니, 프로젝트 방향과 맞음
  + 단점 : 데이터 확보 — 어떻게?
* **3) 프로젝트 방향 변경**
  + True / False ⇒ 다른 Classification
  + 장점 : 융튱성이 넓어진다
  + 단점 : 너무 넓어져서 탈이다.

1. 필터링 된 데이터로 모델 성능 체크
2. 모델의 Input/Output CSV
   1. Raw output CSV :
      1. 기사 1 / 언론사 / Prediction : False
      2. 기사 100
   2. Raw output CSV - Filtering / Summarization : ⇒ Test.ipynb 임시 코드를 작성해서 공유 가능
      1. 조선 일보 : 75.5%

⇒ 성능을 어떻게 좋게 만들 것인가?

* 성능을 더 좋게 할 것인가? 해야되나..
* 모델 업그레이드 ⇒ 제한적. ⇒ 비현실적..
  + 단점 : 추가 데이터 필요함.
    - Content
    - 언론 매체(뉴스 / 인스타 / 페이스북)
    - 기자의 이름
    - 시간
    - 사람들 좋아요 유무
  + ⇒ 데이터 확보 How?
  + 편법..? Chat GPT API 를 연결해야 한다.
  + ChatGPT에 타이틀 넣고, 이거 팩트야? 라고 물어보면
  + ⇒ 최근 기사를 얼추 반영해서 판단을 해줘요
  + ⇒ 애를 Test data로 넣기
  1. 해야하나…

**⇒ 판단 보류.**

1. React 로 기능 구현 가능한지 점검
2. [핵심] 웹 서비스 → 어떻게 구현할지 ⇒ 병렬 필요

* HTML - 서버 / React ⇒ 주현님
* Back 엔드

필요 자료

* 기사 링크
* 작성 일자
* 출처
* 카테고리 추출 가능 유무 확인 요청
  + 사회 / 경제 ⇒

⇒ 구현에 얼마나 걸릴까.

* GIthub 올리기

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

**<논문 발표 준비>**

* Introduction
  + Knowledge Graphs
  + Semantic Proximity from Transitive closure
* **Result**
  + **Calibration**
  + **Value of Indirect connections**
  + **Validation**
* Discussion
* **Materials and Method ⇒ Light**
  + Wikipedia Knowledge graph
  + Ideaolocial classifications of the US congress
  + Simple factual statements
  + Independent corpus of statements
* Introduction 1 ⇒ 현규님
  + Knowledge Graphs
  + **Materials and Method ⇒ Light**

2 ⇒ 현구

* + Semantic Proximity from Transitive closure
* **Result**
  + **Calibration**

3 ⇒ 주현님

* + **Value of Indirect connections**
  + **Validation**
* Discussion

⇒ 일요일 저녁

→ 따로 미팅