## **1월 5주차 회의록 (21.01.30)**

4조 김태희, 안세훈, 정영훈, 고준호

**모델링 계획 변경**

**신뢰도 분석 모델 적용 취소**  
- 현재 프로젝트는 네이버 플레이스의 짧은 리뷰(1-3줄) 이므로 가짜리뷰 판별이 쉽지 않음  
- 현존하는 가짜 리뷰 시스템은 대부분 블로그 크롤링을 통해 광고성 포스팅을 판별  
- 따라서 신뢰도 지표는 이용하기 어려움

=> 신뢰도 지수를 없애는 대신 기존의 긍부정 분석과 주제 분석을 좀 더 강화하기로 결정

**긍부정 분석**  
- 현재 네이버 플레이스는 별점 제도를 폐지함 (오래된 리뷰만 별점이 있음)  
- 따라서 긍부정 분석의 활용도는 더욱 높아질 것으로 전망 (별점 예측도 가능)

**주제 분석**  
- 사전 주제 선정 후 그에 맞게 태깅된 데이터를 파인 튜닝 예정  
- 추가로 사용자가 입력한 주제로 분석하는 기능도 추가 고려  
=> 사례) 고를 수 있는 주제가 가격, 청결 밖에 없는데 사용자는 분위기를 분석하길 원할 때  
=> 사용자가 입력란에 분위기를 입력하면 분위기와 관련된 리뷰 우선순위 정렬 가능  
=> 다만 해당 분석은 파인튜닝이 되지않은 사전학습 모델만 이용가능

**도메인 변경**

성남시 스터디 카페 -> 성남시 카페로 확장  
- 선정이유: 스터디 카페와 속성이 비슷하면서 리뷰가 훨씬 많음  
- 이후 크롤링 툴이 더 보완되면 음식점도 추가 가능

**2월 계획**  
- 변경된 도메인으로 크롤링 툴 완성 & 크롤링  
- 기존 모델에 적용  
- 데이터 태깅  
- 파인튜닝  
- 개선 모델 적용

**1월 5주차 진행사항**

**[변경 도메인 조사]**

**음식점**

음식점 주제: 맛 / 친절 / 주차 / 가성비 / 청결 / 재료 신선 / 매장 크기 / 분위기/ 특별한 날 / 특별한 메뉴 / 양 / 단체모임 / 인테리어/ 화장실 / 뷰 / 혼밥

**수정구 음식점**

6페이지 ( 1페이지당 50개 )

리뷰개수 : 40 ~ 4800개 ( 대략 100~500가 제일 많이 분포하는 것 같음 )

**중원구 음식점**

6페이지 ( 1페이지당 50개 )

20~ 1500개 ( 대략 50~500 가 제일 많이 분포하는 것 같음 )

**분당구 음식점**

6페이지 ( 1페이지당 50개 )

20~ 4324개

**카페**

카페 주제: 커피 맛/ 음료 / 디저트 / 뷰 / 주차/ 메뉴 / 인테리어 / 사진/ 청결/ 친절 / 좌석 / 화장실 / 집중 / 가성비 / 대화

**수정구 카페**

6페이지 ( 1페이지당 50개 )

9~ 4511

**중원구 카페**

6페이지 ( 1페이지당 50개 )

1~ 3183

**분당구 카페**

6페이지 ( 1페이지당 50개 )

34~3265