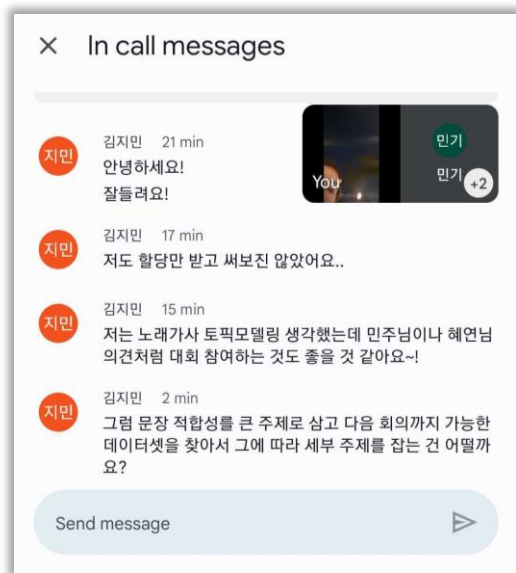


# CUAI 동계 컨퍼런스 NLP팀 주제 발표

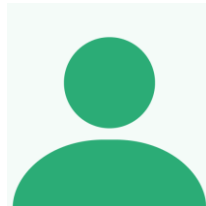
2022.12.27

발표자 : 김민기

## 팀 소개 & 1차 회의 진행



12월 26일 오전 10시  
주제 회의 진행



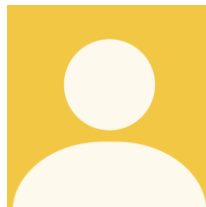
김민기  
경영학과 17학번  
졸업 못함ㅠ.ㅠ



김지민  
응용통계학 & 소프트웨어학부 20학번  
종강했지만 근로때문에 매일 학교가는 중..



김민주  
경영학 & 응용통계학 18학번  
한국에 없음



이혜연  
응용통계학 21학번  
종강해서 매우 행복한 상태

## 주제 선정

과제	벤치마크	예제
감정 분석 Sentiment Analysis	SST, IMDB, NSMC	입력 문장: 신만이 이 영화를 용서할수 있다
		출력 긍정 · 부정
유사도 예측 Similarity Prediction	STS, MRPC, QQP, PAWS-X	입력 문장1: 파이썬과 자바 중 뭐부터 배워야 하나요? 문장2: Java나 Python 중 하나를 배워야 한다면, 뭐부터 시작해야 할까요?
		출력 유사 · 비유사
자연어 추론 Natural Language Inference	SNLI, MNLI, XNLI, aNLI	입력 전제: 회색 개가 숲에서 쓰러진 나무를 알고 있다. 가설: 개가 밖에 있다.
		출력 참 · 거짓 · 모름
언어적 용인 가능성 Linguistic Acceptability	CoLA	입력 문장: 그는 한 번도 프랑스에 갔다.
		출력 가능 · 불가능
기계 독해 Reading Comprehension	SQuAD, RACE, MS MARCO, KorQuAD	입력 지문: 시카고 대학의 학자들은 경제학, 사회학, 법학 분석, 문학 비평, 신학, 행동주의 정치학 등 다양한 학문 발전에 중요한 역할을 담당해왔다. [...] 이 대학은 미국 최대 대학 출판사인 시카고 대학 프레스(University Press of Chicago Press)의 본거지이기도 하다. 오는 2020년 완공될 버락 오바마 대통령 센터에는 오바마 대통령 도서관과 오바마 재단 사무소가 세워질 예정이다.
		출력 2020년
의도 분류 Intent Classification	ATIS, SNIPS, AskUbuntu	입력 질문: 현재 위치 주변에 있는 맛집 두 곳만 알려줘
		출력 장소 검색 (식당, 가까운, 평점 좋은, 2)

### 3. 결 론

개인 맞춤형 서비스에 대한 수요가 증가함에 따라 개인의 취향, 감성에 적합한 서비스 제공의 필요성이 증대되었다. 이에 본 연구는 가사의 감성을 분석하여, 사용자가 원하는 감성의 가사를 지닌 노래를 추천하는 것을 목표로 하였다.

연구를 진행하는 과정에서 예상보다 상당히 낮은 수치의 정확도를 얻게되어, 문제점에 대한 고찰을 통해 데이터의 결함을 발견하였다. 데이터의 레이블 오류와 학습 및 테스트 데이터 특성의 간극으로 인하여 모델 성능이 저하되는 문제가 발생하였다. 문제 해결을 위해 데이터 라벨링 작업을 추가로 진행하거나, 알맞게 라벨링이 이루어진 데이터 수집이 필요할 것으로 보인다. 분류하고자 하는 감정에 대한 특징을 잘 포함하고 있는 문장들을 잘 선별하거나 수집한 후, 적절하게 오버샘플링 기법을 적용하여 양질의 데이터를 증가시키는 것이 학습에 큰 도움을 줄 것으로 예상된다. 또한, 본 연구의 목표가 노래 '가사'에 내재된 감성을 분석하는 것이기에, 노래 가사와 유사한 특성을 지닌 데이터(e.g. 문학 작품 구절)를 테스트 데이터로 활용해야 할 것으로 보인다. 이러한 점을 보완한다면 본 연구의 모델 성능이 높아져 최종 목표인 감성 분류에 따른 노래 추천이 용이할 것으로 기대한다.

### 노래가사 감성분석, 토픽 분류

- CUIAI 과거 유사한 프로젝트 有
- 보완 및 성능 향상 목표로 진행

### 언어적 용인 가능성(Linguistic Acceptability)

- 문맥적으로 적합한 문장을 판단
- ➔ 사회에 긍정적인 영향 미치는 SW개발 ex. 조현병 징후 파악
- 적합한 DATA 찾기에 난관 예상

## 추가 진행

목표 : 발제한 아이디어 바탕으로 인공지능 인문학 경연대회 논문 제출

문제 : 대회 마감까지 10일 정도 남음



주최주관

중앙대학교 인문콘텐츠연구소

지원기간

2022년 11월 15일 ~ 2023년 01월 6일 07:00

활동분야

IT/학술/논문

IT/학술/논문

지원하기



## 계획

~ ASAP : 주제에 적합한 데이터 발굴하기

23/01/06 : 논문 제출

23/01/17 : 중간 발표

23/02/07 : 최종 제출

23/02/14 : 본선 발표

...

43/01/01 : 억만장자 되어서 요트타고 일출 보기

감사합니다