

AI\_11\_CP1

# 감정분석 Updown게임 (영화편)

*If you do nothing, nothing happens*

# Index

1. | 프로젝트 배경 및 내용 소개

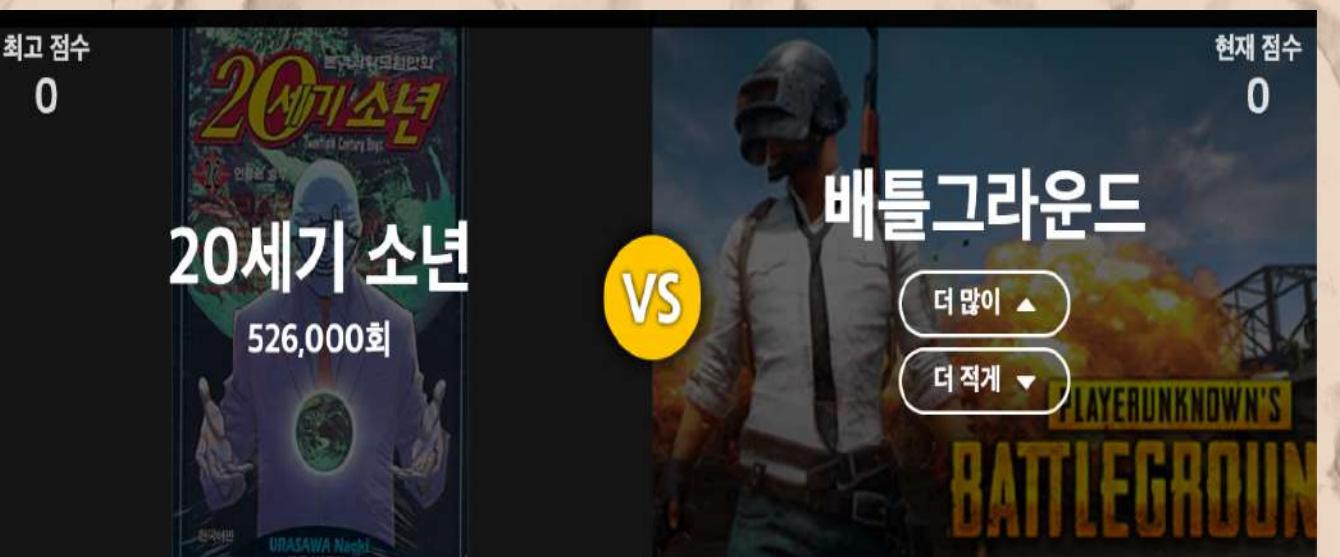
2. | 파이프라인 설명

3. | 인원별 업무 소개

4. | CNN 모델 / KoBERT 모델

5. | 모델 시연

# 더많이 더적게



## 개발 개요:

- CV 보다 NLP 컨텐츠 프로젝트가 상대적으로 적음
- 검색량 조회 게임 벤치마킹

## 개발 목표:

- Youtube 영화 예고편 댓글을 KoBERT로 감정 분석
- 두 영화 중 감정 높은 쪽 맞추는 게임 사이트 구축

# 프로젝트 개요

필요한 기간과 작업내용을 설정하고, 협업에 필요한 플랫폼을 적절히 활용하였습니다.

## 프로젝트 기간별 진행사항

**6/23 ~ 26** | 사전 기획 / 아이디어 선정

**6/27 ~ 29** | 필요 데이터 수집 / API 활용

**6/29 ~ 30** | 데이터 정제 및 정규화

**6/30 ~ 7/2** | 모델 구현

**7/2 ~ 5** | 웹 서비스 플랫폼 구현

### Notion

| 기획서 및 아이디어 공유



### Google Drive

| 파일 공유



AI\_11\_CP1\_1팀



### Gather Town

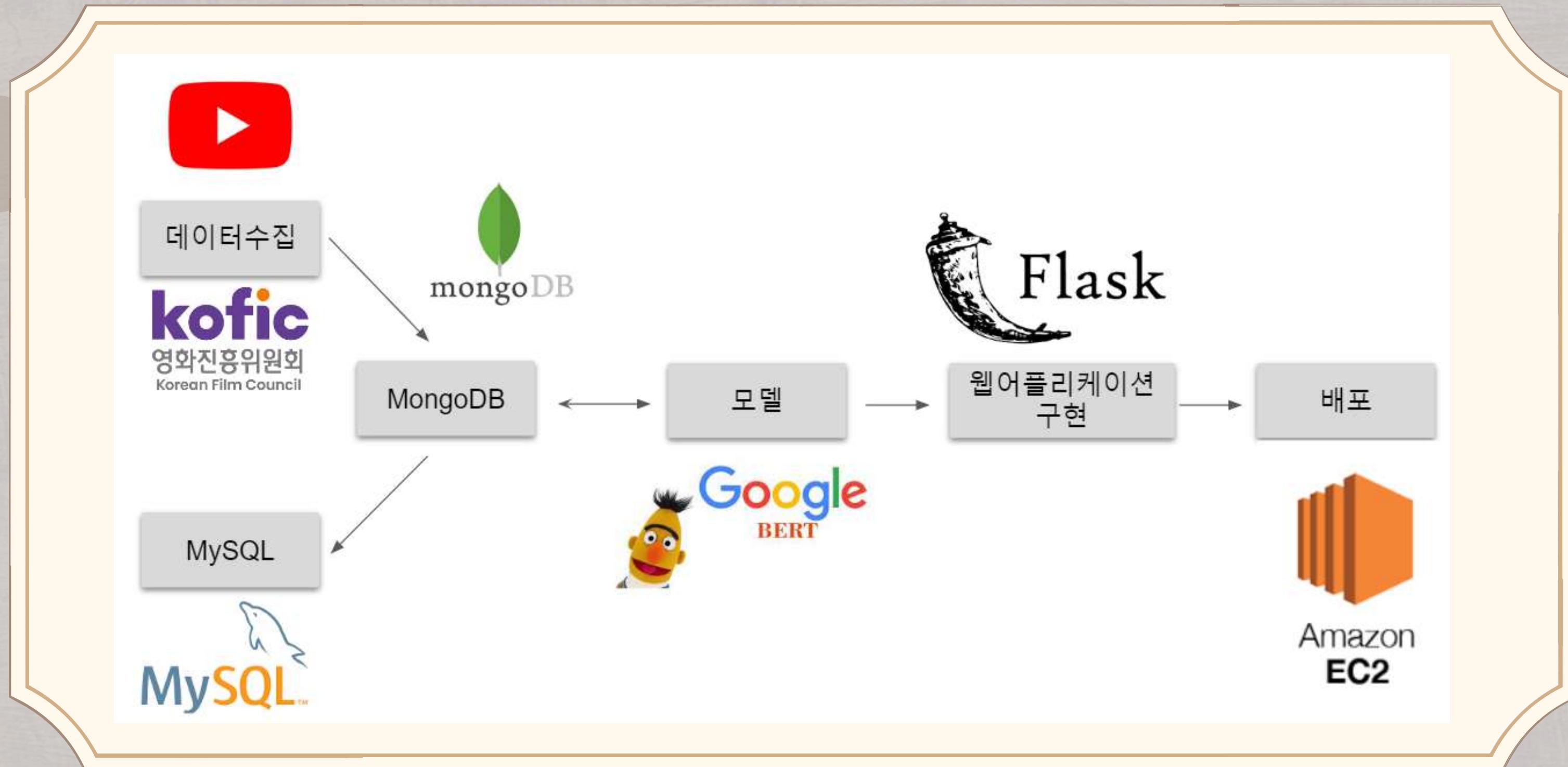
| 미팅룸



### Discord

| 진행사항 공유

# 파이프 라인



# 팀원별 업무 분담

CP1 짧은 기간동안 최대의 효율을 내기 위해 각 파트를 나누어 진행하였습니다.

각 팀원들은 맡은 부분들에 집중했을 뿐만 아니라, 다른 팀원들이 필요한 부분들을 피드백 해주었습니다.

01

데이터 수집 &  
모델링



02

Front-end &  
ppt 작성



03

Back-end &  
프로젝트 발표



최종호 / 최하림

우진성

박성우

# 데이터 수집 / 모델링

유튜브 API를 통해 '영화 제목', '썸네일 이미지', '동영상 댓글'을 크롤링 하여 MongoDB에 적재하였습니다.

데이터 전처리 및 분석을 실시하여 모델의 성능 및 인사이트를 도출하였습니다.

(영화 선택 기준: 관객수 순위 1위 ~ 50위)



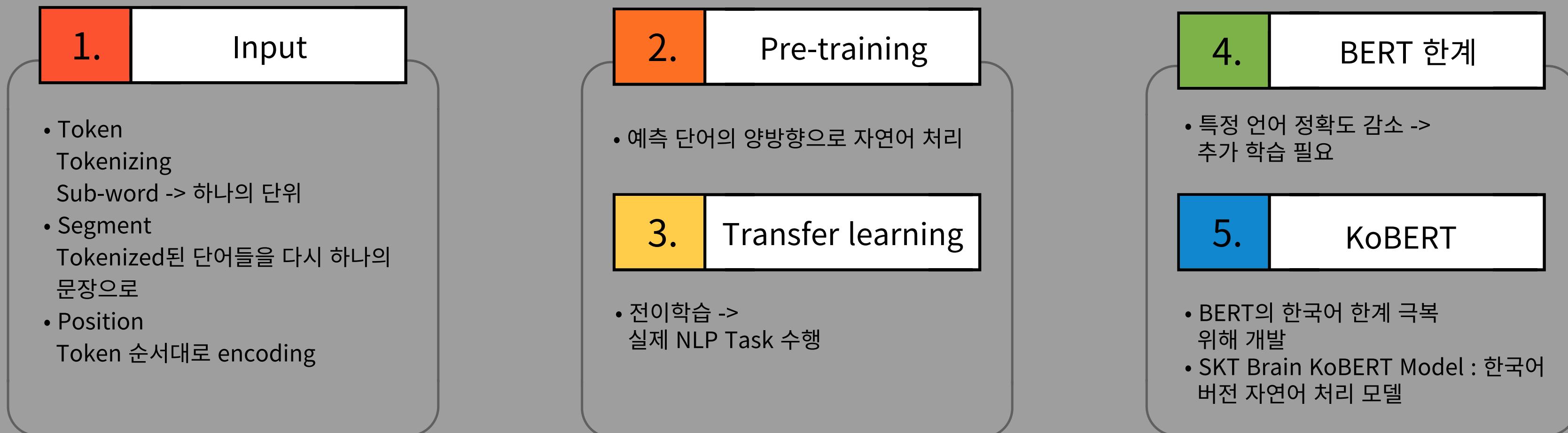
| 차단된 댓글 삭제 (광고, 욕설 등)

| 크롤링된 데이터 저장 (영상 1개당 약 1,000개)

| KoBERT 사용 (BERT의 한국어 버전)

| AI hub 데이터셋 사용 (단발성대화, 감정분류)

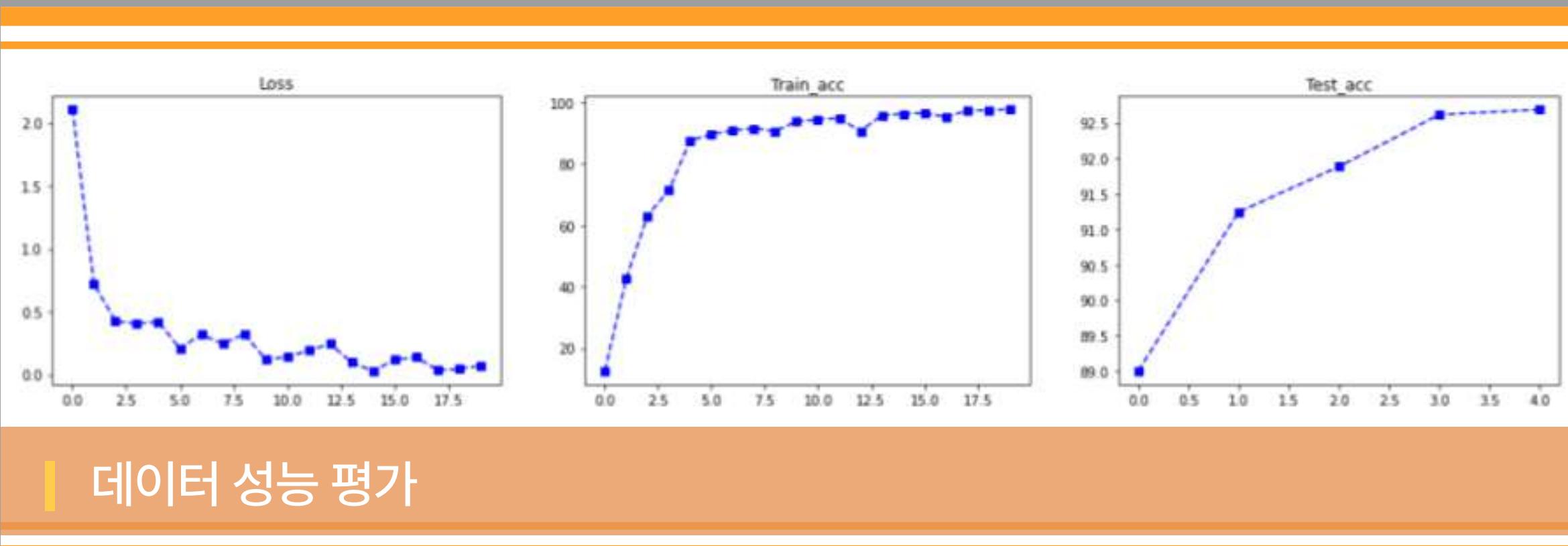
# BERT / KoBERT



# BERT / KoBERT

## 구축내용

- 감성대화 어플리케이션을 이용한 수집
- 7가지 감정에 대해 라벨링
- 총 데이터 갯수: 43,991문장



데이터 성능 평가

## 성능개선

- 기존 단발성 대화 데이터셋 (~ 55%)
- 감정분류를 위한 대화 데이터셋 (~ 93%)

	Sentence	Emotion
1	어. 청소 니가 대신 해 줘!	분노
2	둘 다 청소 하기 싫어. 귀찮아.	분노
3	둘 다 빠가 실어서 화내.	분노
4	그럼 방세는 어떡해.	분노
5	권태간을 알았는데 다른 사람이 생겼나보다라고.	슬픔
6	그냥 걷고 있어.	슬픔
7	어. 고등학교 통학인데 이렇게 더럽게 삐뚤빼뚤 왔어.	분노
8	처음 학원에서 만났다가 서로 좋아해서 사귀게 되었.	슬픔
9	내가 예전 표현을 잠 못해서 자주 싸우긴 했어.	슬픔
10	오늘 헤어졌어.	슬픔
11	톰메이트와 너무 자주 싸우게 돼.	분노
12	그리고 신촌에 번화한 데문에 그런 수가 있다.	분노

데이터셋 예시

# Front-End

웹 서비스에 필요한 라이브러리를 적절하게 사용하였습니다.  
아이디어 회의에 활발히 참여하고 의견을 공유하였습니다.

## 다양한 라이브러리

여러개의 라이브러리를 사용하면서 각각의 버전을 일치화 시켜주었고, requirement.txt를 통해 저장하였습니다.



## FRONT-END

### 추가 작업물



- | 카카오톡 공유하기
- | 카카오 애드핏 광고 설치
- | 코드스테이츠 아이콘 사용
- | 반응형 웹페이지

### 여러 문제점



- | 버전 충돌
- | 로컬 / 배포용 통합
- | PC / 모바일 반응형
- | 디자인 다듬기

# Back-End

AWS EC2, MongoDB, MySQL 등 전반적인 서버 관리 및 도메인을 제공하였습니다.

팀장으로서 팀원들과 소통하며 필요한 부분을 채워주었습니다.

전체적인 흐름을 이해하며 프로젝트 기획서와 발표를 맡았습니다.

01 | DB적재 및  
운용



## [ Data Engineering ]

생산된 수많은 데이터를 모을 수 있도록 거대한 데이터 파이프라인을 설계, 구축합니다. 모두가 쉽고 안전하게 다룰 수 있도록 가공 처리를 하며, 데이터의 성격에 따라 스트리밍 혹은 배치 처리를 합니다.

02 | 웹 배포 및  
도메인 연결



## [ CD & CI ]

팀원들이 수정한 변경 사항에 대해 버그를 신속하게 찾아 해결하고, 소프트웨어의 품질을 개선하고. 새로운 업데이트의 검증 및 릴리즈의 시간을 단축시켰습니다. 개인서버에 업로드하여 애플리케이션을 실시간으로 배포할 수 있었습니다.

03 | 기획서 작성 및  
프로젝트 발표



## [ CPI\_1팀\_팀장 ]

팀장으로서 진행상황을 파악하여 남은 기간에 맞는 계획을 수립하여 프로젝트를 총괄, 지휘하였습니다. 기획서를 제작하고 프로젝트에 대한 발표를 진행하였습니다.

# 시연하기

<http://colacan.kro.kr:5000/>

# 자체 평가 의견

어려운 모델과 구현에 시간과 정성을 많이 쓴 결과, 좋은 결과물을 얻을 수 있었습니다.

지속적인 개발을 위해 좋았던 점과 아쉬운 점을 복기하였습니다.

## 프로젝트를 하면서 들었던 생각들

### 좋았던 점

- 전체적인 서비스 기획의도대로 구현 (퀴즈 알고리즘, 외부 API 서비스)
- 높은 모델 정확도 (92.5% 이상)
- 카카오톡 광고, 공유하기

### 아쉬운 점

- 댓글 전처리 필터링 (광고, 욕설)
- 모바일 어플리케이션 최적화 일부 미적용
- mysql 시각화

# 피드백

정말 고생많이 하셨습니다. 잘 안풀리던 부분도 있었고 소통에 힘들었던 부분도 있었을텐데  
서로 양보하고 배려했던 기억이 많이 남아있습니다.

## 팀원들에게 하고 싶은 말

느낀 점	박성우	우진성	최종호	최하림
	다음 팀프로젝트에서는 깃을 사용하여 버전관리를 더 철저히 해보고 싶습니다.	프론트엔드의 다양한 문법과 구문들을 공부할 수 있었던 시간이었습니다.	데이터 수집과 모델링하는 과정이 배울수록 흥미로웠고, 특히 NLP를 학습할 수 있게 되어 재밌었습니다.	문제해결 과 함께 동료와의 커뮤니케이션으로부터 팀 프로젝트의 장점을 느낄 수 있었습니다.
아쉬운 점	서버에 부하가 걸려 자주 꺼지곤 했는데 배포, 연동에 있어 더 세심하게 봐야 할 것 같습니다.	각 라이브러리마다의 버전이 상이하여 맞추는데 시간이 많이 소요되었습니다.	익숙치 않은 라이브러리들과 문법들에 대해 더 해봐야겠다 생각했습니다.	모델링을 마친 데이터를 DB에 더 빠르고 쉽게 적재할 수 있는 방법에 대해 공부하고 싶습니다.

AI\_11\_CP1\_1팀

---

AI\_11기 화이팅!

THANK YOU.

COMMENT

마무리까지 1달!

끝까지 열심히 달려봅시다!