

일 일 보 고 서

프로젝트	무신사 온라인 소비자 리뷰 기반 데이터 분석		
작성자	team3	일시	2025-11-05

금 일 실 시 사 항	비고
[곽동원] - 데이터 정제 규칙 기반 전처리 코드 확정 및 공유 - BERT 모델에 파인튜닝 할 데이터 Kc-BERT 로 분류 - BERT 모델 학습 및 BERT 모델 사전학습 실시 - ELECTRA 모델 학습	
[강지연] - 데이터 정제 규칙 기반 전처리 코드 확정 및 공유 - BERT 모델에 파인튜닝 할 데이터 Kc-BERT 로 분류 - BERT 모델 학습 및 BERT 모델 사전학습 실시 - ELECTRA 모델 학습	
[안호용] - 데이터 정제 규칙 기반 전처리 코드 확정 및 공유 - BERT 모델에 파인튜닝 할 데이터 Kc-BERT 로 분류 - BERT 모델 학습 및 BERT 모델 사전학습 실시 - ELECTRA 모델 학습	
[정수아] - 데이터 정제 규칙 기반 전처리 코드 확정 및 공유 - BERT 모델에 파인튜닝 할 데이터 Kc-BERT 로 분류 - BERT 모델 학습 및 BERT 모델 사전학습 실시 - ELECTRA 모델 학습	

상세 실시 사항

1. 데이터 정제 규칙 기반 전처리 코드 확정 및 공유

1.1 사전학습 데이터셋 : 구어체, 뉴스 원문 및 뉴스 리뷰, 영화 및 게임 리뷰

-> Pre-training_dataset.txt

1.2 파인튜닝 데이터셋 : 네이버 쇼핑 리뷰

-> Fine-tuning_shopping.txt

2. BERT 모델에 파인튜닝 할 데이터 Kc-BERT로 분류

2.1 Kc-BERT 제작자 Git-hub 참고

2.2 “KcBERT: 한국어 댓글로 학습한 BERT” 논문 참고

2.3 파인튜닝 할 데이터 Kc-BERT로 분류

3. BERT 모델 학습 및 BERT 모델 사전 학습 실시

3.1 Kc-bert 제작자 Git-hub 사전학습 방법 참고

3.2 사전학습 test 진행

명일 예정 사항

비고

-
- BERT 사전 학습 실시
 - 파인튜닝 방법 학습
 - ELECTRA 모델 학습, 논문 리뷰
 - 크롤링 데이터 재정의
-

이 슈 사 항
