

일 일 보 고 서

프로젝트	무신사 온라인 소비자 리뷰 기반 데이터 분석		
작 성 자	team3	일 시	2025-11-03

금 일 실 시 사 항	비 고
[곽동원] – 팀장	
- 프로젝트 개요 및 가설 재설정	
- 데이터셋 선정	
[강지연]	
- 모델 선정 및 구현 개요	
- 데이터셋 선정	
[안호용]	
- 모델 선정 및 구현 개요	
- 데이터셋 선정	
- 데이터셋 수집	
[정수아]	
- 프로젝트 개요 및 가설 재설정	
- 모델 선정 및 구현 개요	
- 데이터셋 선정	

상세 실시 사항

1. 프로젝트 개요 및 가설 재설정

1.1. 초기 문제 정의

- 프로젝트 선정 및 목적 정의

1.2 가설 재설정

- 타겟 세분화
- 가설 근거 보완

2. 모델 선정 및 구현 개요

2.1 BERT 기반 한국어 리뷰에 특화된 모델 구현 방법 확인

- BERT의 구조 및 학습 과정 파악
- 토큰화 및 사전 학습에 사용되는 함수 및 라이브러리 정리

2.2 ELECTRA 기반 한국어 리뷰에 특화된 모델 구현 방법 확인

- ELECTRA 이론 및 BERT와의 차이점 비교

3. 데이터셋 수집

3.1 사전 학습용 데이터셋 195만개 수집

3.2 파인 투닝용 데이터셋 50만개 수집

3.3 학습용 데이터셋 12만개 수집 예정

명 일 예 정 사 항	비 고
<ul style="list-style-type: none">● 데이터셋 전처리● DAPT(도메인 적응 사전 학습) 학습<ul style="list-style-type: none">- label이 없는 데이터가 존재할 때, 도메인에 특화된 사전 학습 수행● 크롤링 학습	

이 슈 사 항
