

# Data 명세서

작성자	안호용	작성일	2025-11-12
-----	-----	-----	------------

출처	데이터명	크기	확장자	비고
<a href="https://github.com/bab2min/corpus/tree/master/sentiment">https://github.com/bab2min/corpus/tree/master/sentiment</a>	steam, naver_shopping	11359K B, 20141K B	txt, txt	스팀 게 임 리뷰 데이터 10만개, 네이버 쇼핑 리 뷰 20만 개
<a href="https://github.com/Beomi/KcBERT/releases/tag/v2022.3_Q">https://github.com/Beomi/KcBERT/releases/tag/v2022.3_Q</a>	2022_3q	1GB	zip(txt 압축파 일)	뉴스 댓 글 데이 터 150 만개 발 췌
<a href="https://github.com/e9t/nsmc">https://github.com/e9t/nsmc</a>	ratings.txt	19058K B	txt	네이버 영화 리 뷰 댓글 20만개
<a href="https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&amp;topMenu=100&amp;aihubDataSe=dataset&amp;dataSetSn=126">https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&amp;topMenu=100&amp;aihubDataSe=dataset&amp;dataSetSn=126</a>	ratings.txt	19058K B	txt	네이버 영화 리 뷰 댓글 20만개

https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=71603	3_문어체_뉴스(1)_200226, 3_문어체_뉴스(2)_200226, 3_문어체_뉴스(3)_200226, 3_문어체_뉴스(4)_200226, 4_문어체_한국문화_200226, 6_문어체_지자체웹사이트 _200226, 1_구어체(1)_200226	zip, xlsx, xlsx, xlsx, xlsx, xlsx, xlsx	49MB, 40.7MB, 51.71KB, 44.06KB, 15.24KB, 16.65KB, 15.08KB	문어체 100만 개, 일상 구어체 20만개
https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=71603	Training₩01.원천데이터, Validation₩01.원천데이터 데이터 전부	zip (내부 xlsx)	24.8MB, 3.35MB	쇼핑몰 리뷰 22 만 5천 개, sns 리뷰 2 만 5천 개