# 감정 분류를 위한 대화 음성 데이터셋 4, 5차

7가지(happiness, angry, disgust, fear, neutral, sadness, surprise)의 감정으로 5명이 라벨링한 자료. 말은 7가지인데 angry, disgust, sadness가 압도적으로 많음

4차년도 13,912건, 5차년도 10,011건

# 단발성 대화 데이터셋

7가지 ( 기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립 ) 감정으로 라벨링

단발성 38,594문장

깃허브: <https://github.com/afnf33/myNLPproject/tree/181a11dda57cf325a75de2a6941a45ea76f511f1/Korean_multisentiment_classification>

# 연속성 대화 데이터셋

7가지 ( 기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립 ) 감정으로 라벨링

중립이 가장 많음. 중립 제외 가끔 분노, 놀람 있는 편

연속적 10,000개 대화 (단발성 55,627문장)

대략 슬픔 1973, 행복 1124, 공포 115, 혐오 224, 분노 3633, 놀람 4867

감정을 판단할 수 있는 문장 = 11936

# 트위터 대화 시나리오

감정 분류X 시나리오만 존재

#영화 리뷰

데이터셋: <https://raw.githubusercontent.com/e9t/nsmc/master/ratings.txt>

깃허브: <https://github.com/e9t/nsmc/>

참고 블로그: <https://devtimes.com/nlp-korea-movie-review>

20만개 리뷰. 0번 부정, 1번 긍정

* 텍스트 파일 다운이 안되는건지 못하는 건지 모르겠음..자꾸 웹 페이지 형태로 열림
* 이걸 엑셀 형태로 바꾸는게 가능할까..? 아님 저 웹페이지로 이동해서 컨트롤a 해서 다 복붙하고 메모장에서 바꿔야 하냐? 20만개 양 존나 많아서 이거 어케해야할지를 모르겠음

#뉴스 댓글 감정 분류 모델

* 기쁨, 놀람, 슬픔, 분노, 혐오, 공포 6가지 감정

## https://github.com/woodongk/news-comments\_emotion\_classification-CNN-tensorflow