

코드, 데이터 파일 설명

코드 파일

| 파일명 | 설명 |
|---|--|
| Crawling_comments_from_videos_in_videolist(new).ipynb | 유튜브 API를 통한 크롤링 재생목록에 있는 모든 영상 추출 |
| Ens_te.ipynb | weighted voting 앙상블 코드 |
| kcbert_performance_test.ipynb | kcbert 성능 종합하기 위한 코드 |
| mj_sentiment_model_with_kcbert의_최종의_사본.ipynb | kcbert 모델 제작 코드 |
| mj_sentiment_RNN(GRU).ipynb | GRU 모델 제작 코드 |
| mj_sentiment_RNN(LSTM).ipynb | LSTM 모델 제작 코드 |
| TF-IDF_revised_2.ipynb | decision tree 모델 제작 코드 |
| 모델로 라벨 교정.ipynb | 학습 데이터 제작을 위한 코드 일부 데이터로 전체 데이터 라벨 예측 |
| 모델로 라벨 교정2.ipynb | 학습 데이터 제작을 위한 코드 일부 데이터로 전체 데이터 라벨 예측 |
| corpus.R | 학습 데이터 제작을 위한 코드 학습 데이터 전처리 방법 연구 |
| corpus2.R | 학습 데이터 제작을 위한 코드 학습 데이터 전처리 방법 연구 |
| score_infer.R | 학습 데이터 제작을 위한 코드 학습 데이터 전처리, 라벨 교정 및 완성을 위한 코드 |

데이터 파일

| 파일명 | 설명 |
|-----------------------------|--|
| pr2_revised4.xlsx | 22만개의 유튜브 댓글 (전처리 완료, 댓글 라벨 교정 완료) |
| mj_sentiment_test_big.xlsx | pr2_revised4.xlsx를 train_test_split으로 쪼개서 얻은 test data |
| mj_sentiment_train_big.xlsx | pr2_revised4.xlsx를 train_test_split으로 쪼개서 얻은 train data |
| last_train_data.xlsx | 16만개의 유튜브 댓글 파일 긍정, 부정 댓글 비율 1:1 pr2_revised4.xlsx에서 추출함 |

| 파일명 | 설명 |
|-------------------------------|---|
| mj_sentiment_train.xlsx | last_train_data_small.xlsx를 train_test_split으로 쪼개서 얻은 train data |
| mj_sentiment_test.xlsx | pr2_revised4.xlsx를 train_test_split으로 쪼개서 얻은 test data |
| last_train_data_small.xlsx | 10만개의 유튜브 댓글 파일 긍정, 부정 댓글 비율 1:1 |
| mj_sentiment_train_small.xlsx | last_train_data_small.xlsx를 train_test_split으로 쪼개서 얻은 train data |
| mj_sentiment_test_small.xlsx | last_train_data_small.xlsx를 train_test_split으로 쪼개서 얻은 test data |
| 선거결과(te).xlsx | weighted voting을 위한 파일 decision tree, lstm, gru, kcbert의 예측한 라벨과 댓글이 담겨 있음. |