

Department of Software Convergence

Data Analysis Capstone Design Project   
Weekly Progress Report

| Project Title | 딥러닝을 활용한 노래 가사 분석 및 추천 시스템 |
| --- | --- |
| Student / ID | 김은비 / 2019100858  유창현 / 2018102121  이은경 / 2019100896  이인석 / 2018110659 |
| Reporting Week | 2022-04-22 ~ 2022-04-28 |
| Project Manager of Reporting Week | 이인석 |
| Faculty Supervisor | 유창현 |

# **Tasks Outlined in Previous Weekly Progress Report** (Provide detailed information on the tasks to be completed in this week)

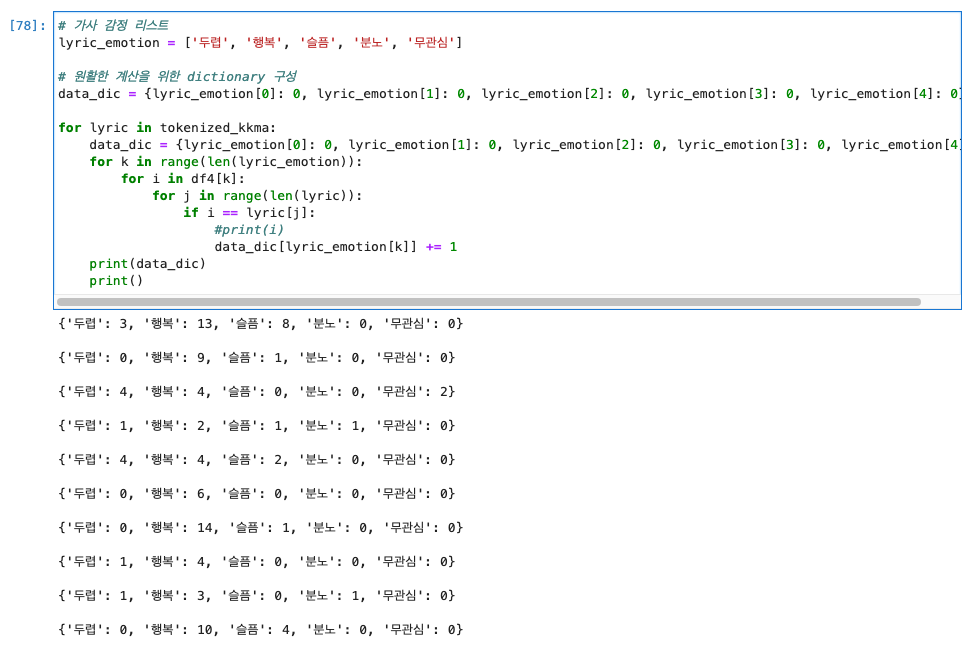
* 감정 키워드와 연관성이 없는 단어 직접 제거
* 5가지 감정에 대한 감성사전 구축

# **Progress Made in Reporting Week** (Provide detailed information on the progress that you made in the reporting week. Limit your write-up to no more than two page)

* 각자 담당한 키워드별로 해당 키워드와 거리가 멀다고 생각하는 단어를 직접 제거해 주었다. 제거하기 전 키워드별로 구성하고 있 감정사전의 평균 개수는 200개 정도 되었고 제거한 후 (‘두렵', 행복', ‘슬픔', ‘분노', ‘무관심')이 각 (70, 50, 50, 80, 110)개가 되었다. 이렇게 최종 감정사전을 구축했다.



* 구축한 감성사전을 기반으로 곡별로 감성사전에 포함된 단어가 있을 시 해당 단어의 유사도를 그 감정의 감정점수에 더하여 가장 큰 점수를 가지는 감정을 매핑하는 방식으로 라벨링을 진행하였다. 하지만 결과를 확인해보니 제대로된 라벨링이 진행되었는가에 대한 의구심이 들었고, 그 원인이 감성사전의 단어 수가 적기 때문인지, 추출한 데이터의 근본적인 특성인지, 이것을 앉고 가야 하는 부분인지 고민해보고 이에 대한 해답을 도출하는것이 급선무라는 생각이 들었다. 그리고 감성사전의 보완과 수정이 필요하다는 결론을 내렸다.



# **Difficulties Encountered in Reporting Week** (Provide detailed information on the difficulties and issues that you encountered in the reporting week. Limit your write-up to no more than one page)

# 기존에 Word2vec의 most\_similar 함수를 통해 감정별 키워드와 유사도가 0.5이상인 단어를 추출하고 연관성이 없는 단어를 삭제하다 보니 감성사전에 포함된 단어의 수가 크게 줄어들고 감정들 간에 단어 수의 차이도 커져 곡별 감정 라벨링의 정확도가 감소하는 문제가 생겼다. 이러한 이유로 감성사전의 수정 필요성을 느꼈고, 키워드를 추가 등의 방식을 통해서 감성사전의 단어수를 늘리고 정확도를 높일 수 있는 수정과정을 진행할 계획을 세웠다.

# **Tasks to Be Completed in Next Week** (Outline the tasks to be completed in the following week)

* 감성사전 최종 수정 및 이를 기반으로 곡별 감정 라벨링 정확도 향상
* 감정 분류에 활용할 딥러닝 모델 결정 및 설계 진행