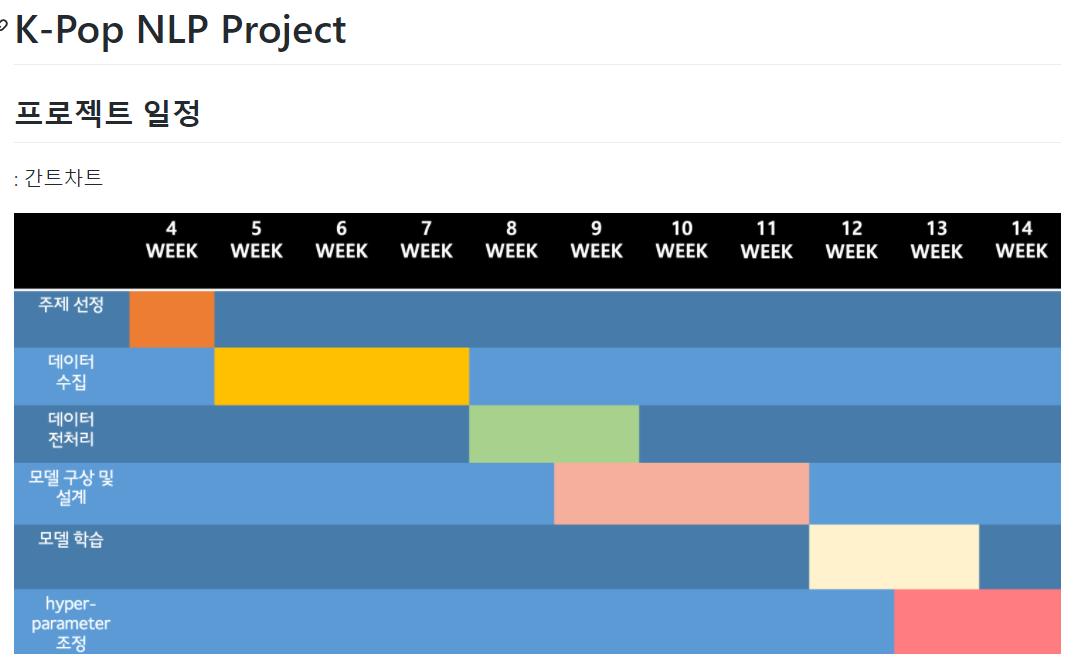
**2020 11 02 회의록**

**회의 인원 : 팀원 모두 참가**



역할분담(알고리즘 별 분담)

LDA (지민, 준영) K-means(은후, 선우)

(시간이 된다면 웹/앱으로 구현 | 시간이 부족할 경우 모델 학습한 결과만 보여줘야 할 듯(팀 반반))

* 일정은 한 주만 타이트하게
* 11.27일 까지 모델 학습 완료
* 발표 및 ppt제작은 나중에 정해보기

**공통적으로 전처리 할 것 생각하기(준영님) : 11/8일 자정까지 – 최종 발표,ppt 제외**

* 제거 속성(작사가/작곡가/소속사/최고순위/성별)
* 필요 속성이 EMPTY인 경우 드랍시키기
* 특수문자 제거 (개행/느낌표 등) : 한글, 영문만 남기고 모두 제거
* 가사 일본어 드랍시키기
* 토큰화(나/는/밥/을/먹/었/다) 후 불용어 제거 (밥/먹) : 이 과정에서 한글/영어 둘 다 처리 가능한 multilingual model 필요

**LDA 모델 전처리 과정 추가**

* 표제어 찾기
* TF-IDF 행렬 만들기
* 참고 사이트 <https://medium.com/@omicro03/%EC%9E%90%EC%97%B0%EC%96%B4%EC%B2%98%EB%A6%AC-nlp-8%EC%9D%BC%EC%B0%A8-lda-f571b4da9d04>
* 이후 모델 학습하면 됨

**다음 회의(11/13 금요일)까지 각자 해올 것**

1. 공통적으로 데이터 전처리 마친 데이터를 가지고 각자 알고리즘 별 추가 전처리 과정 거치기
2. 각자 알고리즘에 맞게 모델 학습해오기(완벽하지 않아도 되나 각 알고리즘 별로 학습 가능한 모델은 만들어오기)