**2019 동계 계절학기 오픈소스 소프트웨어 프로젝트 (OSSP) 계획서**

컴퓨터공학과 20144800 천승현

**1. 주제**

-kaggle에서 제공하는 데이터셋을 이용한 kernel을 개선하여 영화 추천 프로그램 구현

**2. 구현 목표 및 달성효과**

-데이터셋에서 제공하는 여러가지 정보를 종합적으로 분석하여 최종적으로 유저가 선택한 영화와 유사한 영화들을 추천해줌. 데이터셋의 요소에는 각각 영화별 장르, 제목, 점수, ID넘버등의 속성을 가지고 있음. 이러한 요소들을 사용하여 파이썬 라이브러리를 이용하여 데이터를 병합 하고 유사속성 추천 알고리즘을 구현하여 함수로 정의하고 프로그램에 사용함. 영화 추천 프로그램을 구현함으로서 더 효율적이고 빠르게 유저가 원하는 영화를 찾을 수 있고, 어떠한 영화들이 높은 평가를 받는지 분석할 수 있음.

**3. 데이터 셋**

-<https://www.kaggle.com/tmdb/tmdb-movie-metadata/home>

TMDB 영화 데이터셋

-<https://www.kaggle.com/ibtesama/getting-started-with-a-movie-recommendation-system>

개선시킬 kernel

kaggle에서 제공하는 데이터셋을 이용. 해당 데이터는 IMDb라는 인터넷 사이트의 자료를 가공하여 데이터 분석에 용이하게 따로 편집한 유사 데이터셋 TMDb(The Movie Database)이다. 5000개가 넘는 영화들의 제목, 개봉일, 배우, 인기도등의 여러가지 정보를 제공하고 있다.

**4. 작업 환경**

-사용 언어: 파이썬 3.x

-사용 소프트웨어: Anaconda Navigator - Jupyter Notebook , PyCharm