

프로젝트 리스트와 대략적인 계획

해당 파일은 12 월 19 일에 시작할 프로젝트들의 리스트들의 후보와 대략적인 진행 방향에 대해 설명합니다

• 영화 추천

목표: 유저가 좋아하는 영화를 제공하였을 때 유저에게 좋아할만한 영화 추천하기

- 1) 데이터 전처리 (처리할 내용이 없기에 생략) - 0 시간

사용할 데이터: MovieLens(<https://grouplens.org/datasets/movielens/>)

데이터 타입: movieId(int), userId(int), rating(float: 0.1f)

- ## 2) 훈련 데이터로 만들기 - 4 시간

Input: [userId|movieId]

movieId 와 userId 를 원핫인코딩하기

movieId 와 userId 를 하나로 합치기

Output: rating (원핫인코딩 여부 결정 필요)

- ### 3) 모델 만들기 및 학습 (DNN/ML) - 8 시간

- #### 4) 학습한 모델을 이용하여 영화 추천하기 - 16 시간

Output 직전 모델 레이어(Layer)에 나오는 행렬과 다른 영화들의 행렬을 이용하여 사용자가 좋아할만한 영화 추천하기

Flask 를 이용하여 영화 추천 웹사이트 만들기

- ## 5) 추가적으로 해 볼만한 것

User 들을 K-Means 를 이용하여 군집화 해보기

- ## 6) 장단점

장점:

- 모델을 만드는 것보다 활용하는 것에 중점을 두었기에 재미 요소가 높다
 - 전처리 할 데이터가 없어서 할 게 적다
 - 분석이 아닌 보여줄만한 웹사이트가 있다

단점:

- 생각보다 금방 끝날 가능성이 있다
- 할게 많이 없어 이력서에 넣기 애매하다
- 모델을 이용해서 추천하는 방식을 이해하기 힘들 수 있다 (수업 외적인 요소)

• Spotify 음악 추천

목표: 유저가 자주 듣는 곡으로 다른 곡 추천하기

위의 영화 추천과 비슷하나 전처리 과정이 있음

사용할 데이터: MillionSongData, lastfmDAta

(<https://www.kaggle.com/datasets/undefinednull/million-song-dataset-spotify-lastfm?select=Music+Info.csv>)

• 날씨데이터로 전력량 분석 및 예측

목표: 날씨와 전력량 사이의 관계를 분석하고 앞으로의 전력 소모량을 예측해보기

1) 데이터 전처리 - 8 시간

사용할 데이터:

법전동시간별 전력량

(<https://data.seoul.go.kr/dataList/OA-22835/F/1/datasetView.do>)

서울 날씨 데이터

(<https://data.kma.go.kr/data/rmt/rmtList.do?code=410&pgmNo=571>)

전처리 과정:

전력량 데이터의 날짜와 시간 포맷 변경하기

전력량 데이터와 서울날씨 데이터를 법전동으로 합치기

2) 각 데이터 분석하기 - 2 시간

전력량과의 연관관계 분석 후 혼동행렬로 표현하기

3) 훈련 데이터로 만들기 - 2 시간

Input: 연관관계가 높은 데이터 5~6 개

Output: 전력량

데이터 표준화나 필요할 경우 하기

5) 모델 만들기 및 학습 (DNN/ML/LSTM) - 12 시간

LSTM은 시간이 남으면 하기

6) 학습한 모델 사용하기 - 16 시간

학습한 모델로 앞으로 3개월간의 전력 사용량을 예측해보기

동일한 환경에서 기온 또는 다른 요소가 조금 변경되었을 때 전력량의 변화 분석

시간이 날 경우 Notion 페이지를 만들어 결과 보여주기

7) 장단점

장점:

- 수업에서 해본 것들이기에 따라하기 쉽다
- 결과값이 풍부해서 보여줄게 많아 이력서에 적합하다
- 활용방식이 다양하고 모델들 간의 비교가 가능하다

단점:

- 보기보다 할게 많다
- 주제가 흥미에 맞지 않을 수 있다

- 날씨 데이터로 따릉이 사용량 분석 및 예측

- 데이터만 바꿔뿐 위와 동일하 과정이 될것으로 보임
- 앞서 비슷한 프로젝트를 한 팀이 있다는 단점이 있음

- 지역별 카페/맛집 지도 서비스

해당 아이디어는 너무 좋으나 이를 모델로 만들어 사용할 방법이 생각이 안나 밑의

2 개를 추천

- 음식점을 개점했을 때 나올 별점 예측

목표: 특정 지역의 나이와 성별 인구로 특정 음식점의 업종만으로 별점 예측해보기

1) 데이터 전처리 - 16 시간

사용할 데이터:

행정동 단위 서울 생활인구

(<https://data.seoul.go.kr/dataList/OA-14991/S/1/datasetView.do>)

서울시 일반음식점 인허가 정보

<https://www.data.go.kr/data/15071829/fileData.do?recommendDataYn=Y#>

별점: 네이버 지도 웹크롤링

전처리 과정:

별점 웹크롤링 하기

인허가 정보에서 폐업한 곳과 필요없는 데이터 지우기

해당 지역의 음식점 수와 동일 업종 수 구하기

3 개의 데이터 합치기

2) 훈련 데이터로 만들기 - 2 시간

Input: (나이,성별) 데이터, 동일 업종 수, 음식점 수

Output: 별점

데이터 표준화나 필요할 경우 하기

3) 모델 만들기 및 학습 (DNN/ML) - 8 시간

4) 학습한 모델 사용하기 - 6 시간

시간이 날 경우 Notion 페이지를 만들어 결과 보여주기

5) 추가적으로 해 볼만한 것

Input으로 해볼만한 다른 데이터가 있으면 넣어보기

6) 장단점

장점:

- 수업에서 해본 것들이기에 따라하기 쉽다
- 주제가 우리 일상에 맞아 재미있어 보인다
- 실제로 회사에서 해볼만한 주제이다

단점:

- 모델보다는 전처리를 하는데 오래 걸릴 것으로 보인다
- 보기보다 할게 많다

• 음식점 추천

목표: 유저가 좋아하는 음식점을 넣었을때 다른 음식점을 추천하기

데이터 : Yelp Open Data

<https://business.yelp.com/data/resources/open-dataset/>

위의 추천 프로젝트와 유사

장점: 다른 추천들과 다르게 종목에 따라 상위 항목을 지도 api에 표시 가능

단점: 다른 추천들과 다르게 할게 많을 예정

외국 데이터

• 좋은 아이디어지만 제외한 목록

날씨 기반 코디 추천 서비스

아이디어가 너무 마음에 들었으나 코디 관련 데이터를 찾을수가 없었음

환율 변환기 + 시각화

환율을 변환하는 것은 모델을 만들지 않아도 할 수 있고 환율과 주식 같은 것들은 모델을 만들어도 결과값이 안 좋을 가능성이 크기에 제외하였습니다.

배달 앱 주문량 패턴 변화 분석

해당 역시 데이터를 찾을수가 없었어요

가맹점 매출 기반 폐업 위험 예측

다른 조가 폐업관련 주제를 해서 최대한 안 겹치기 위해 제외하였습니다

부동산 매매가 예측, 한국 영화 리뷰 감성 분류

수업 시간에 했던 내용과 크게 다르지 않을 것 같아 제외하였습니다.

주식 시장 업종별 수익률 변화 분석

환율 분석기와 동일하게 주식은 결과값이 안 좋을 가능성이 커서 제외하였습니다.