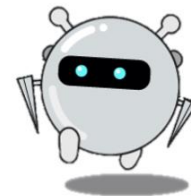


프로젝트 OT



일정 소개



일자	내용
8/1 (화)	프로젝트 OT
8/5 (토)	기획서 마감 (오전 11시 59분까지) / 팀별 프로젝트 선정 마감 (오전 11시 59분까지) 1차 프로젝트 스켈레톤 코드 업로드 (오후 12시)
8/8 (화)	2차 프로젝트 스켈레톤 코드 업로드 (오후 12시)
8/10 (목)	3차 프로젝트 스켈레톤 코드 업로드 (오후 12시)
8/12 (토)	프로젝트 관련 실시간 Q&A 세션 (게더타운, 오후 7-9시)
8/19 (토)	프로젝트 관련 실시간 Q&A 세션 (게더타운, 오후 7-9시)
8/24 (목)	팀 프로젝트 코드 제출 마감 (오전 11시 59분까지)



활동 소개



나만의 데이터 분석 end-to-end 프로젝트 기획서 작성

답겨야 하는 내용

1. 주제 의식 및 개요
2. 데이터 수집 방법 (데이터 출처 포함)
3. 예상되는 데이터 전처리 방법
4. 데이터 분석 방법
5. 기대 효과 및 예상되는 결과

마감 기한

8/5 (토) 오전 11시 59분까지

주의사항

팀별 제출이 아닌 개별 제출



나만의 데이터 분석 end-to-end 프로젝트 기획서 작성

예시: S&P 500 투자자를 위한 kNN 모델을 기반으로 한 주식 포트폴리오 빌더

1. 주제 의식 및 개요

- 코로나로 인한 주식의 급등락으로 사람들의 주식 투자에 대한 관심이 높아진 반면, 그만큼 주식에 대해서 이해도가 낮은 채 투자하는 사람들도 늘어났음. 따라서, 이들을 위해 가장 친숙한 대형주가 속해 있는 S&P 500 주식 데이터를 통해 주식 포트폴리오 빌더를 제공하여, 이들의 취향과 생각과 맞는 포트폴리오를 구성할 수 있도록 돕고자 함.

2. 데이터 수집 방법 (데이터 출처 포함)

- Yahoo Finance가 제공하는 yfinance 라이브러리를 이용하여 데이터 수집. Yfinance에 없는 데이터의 경우에는 SeekingAlpha 등 다른 주식 웹사이트에서 데이터를 수집.



나만의 데이터 분석 end-to-end 프로젝트 기획서 작성

예시: S&P 500 투자자를 위한 kNN 모델을 기반으로 한 주식 포트폴리오 빌더

3. 예상되는 데이터 전처리 방법

- 결측치가 많을 것으로 생각되어 결측치 처리가 주요할 것으로 보임. kNN 모델을 사용하기에 데이터 표준화도 필요할 것으로 예상되며, 수치형 데이터가 아닐 경우 One-Hot Encoding 등 적절한 인코딩 방법을 사용해야 할 것으로 생각됨.

4. 데이터 분석 방법

- kNN 모델을 이용하여 기본적인 추천 알고리즘을 구현하고 해당 알고리즘을 기반으로 input 기반 User-Interactive한 주식 포트폴리오 빌더를 구현.

5. 기대 효과 및 예상되는 결과

- 투자자들에게 조금 더 직관적인 방법으로 포트폴리오를 구성할 수 있게 도와줄 수 있음.



팀별 희망 프로젝트 정하기

마감 기한

8/5 (토) 오전 11시 59분까지

주의사항

1. 팀별 제출
2. 프로젝트별 난이도가 상이하고 성격도 다르기 때문에 팀 내에서의 충분한 상의를 통해 정해야 함.
3. 프로젝트마다 새로운 개념이 등장할 수 있으며, 필요 시 강의 자료가 제공될 예정임. (구글 클래스룸)
4. 프로젝트 신청이 마감되면 신청한 각 프로젝트 별 오픈채팅방에 입장.



프로젝트 주제 목록: ESG

2023 OUTTA 부트캠프 데이터반

한국 기업 ESG 데이터 크롤링 및 분석

기획 의도 | ESG 경영 평가에서 한국 기업들은 어떤 평가를 받고 있을까?

과정 설명 | 1. 웹 사이트 크롤링
2. 전처리 및 시각화, EDA
3. 클러스터링 및 의미도출

배울 점 | 1. 데이터 크롤링에서 발생하는 여러 에러들을 직접 해결할 수 있다.
2. 전처리 과정에서 데이터 특징에 대해 이해할 수 있다.
3. 클러스터링을 수행하면서 의미를 도출해볼 수 있다.

난이도 | 데이터 크롤링: ★★★★★
데이터 전처리: ★★★
데이터 시각화: ★★★
모델링: ★★★

신규 기술 | 클러스터링 기술



프로젝트 주제 목록: 이커머스

2023 OUTTA 부트캠프 데이터반

이커머스 상품 별점 예측 및 추천시스템 구현

기획의도 | 이커머스 데이터를 기반으로 상품 별점 예측과 User-Based CF를 구현하자

과정 설명 | 1. 이커머스 스킨케어 제품의 다양한 정보를 크롤링
2. 데이터 분포 불균형 해결 및 최적의 하이퍼 파라미터를 찾기 위한 전처리
3. 제품 및 별점 데이터를 기반으로 User-Based CF 추천시스템 구현

배울 점 | 1. 크롤링에서 발생하는 다양한 에러를 해결하는 방법론에 대해 이해할 수 있다.
2. 수집한 데이터를 원하는 형태로 가공하면서 전처리 실력을 향상시킬 수 있다.
3. 유사한 고객이 만족한 제품을 제시하는 추천시스템의 알고리즘을 이해할 수 있다.

난이도 | 데이터 크롤링: ★★★★★
데이터 전처리: ★★★★★
데이터 시각화: ★★
모델링: ★★★★★

신규 기술 | SMOTE 오버샘플링, Grid Search, User-Based CF



프로젝트 주제 목록: 교육

2023 OUTTA 부트캠프 데이터반

한국 대학 입결 결정 양상 분석

기획 의도 | 한국 대학의 입결은 어떠한 요인에 의해 결정되며, 어떠한 요인에 영향을 받을까?

과정 설명 | 1. 다양한 웹 사이트에서 크롤링하고 데이터 병합
2. Feature Selection 및 Feature Engineering 중심의 데이터 전처리
3. 선형회귀분석을 이용한 변수간의 관계 및 조절효과 분석

배울 점 | 1. 한국의 입시과열과 그에 따른 행동패턴에 대해서 해석하고 이해할 수 있다.
2. 다양한 소스에서 수집된 데이터를 적절히 병합하는 과정에 대해 이해할 수 있다.
3. 고정효과, 조절효과 등 선형회귀의 해석 방법론에 대해 심층적으로 이해할 수 있다.

난이도 | 데이터 크롤링: ★★★★★
데이터 전처리: ★★★★★
데이터 시각화: ★★★★★
모델링: ★★★★★

신규 기술 | 고정 효과, 조절 효과



프로젝트 스켈레톤 코드 (구글 클래스룸)

프로젝트마다 3개로 분할된 스켈레톤 코드를 기간에 걸쳐 제공할 예정.

프로세스를 분할해서 이해하도록 하기 위함임.

제출 마감일자는 모두 합쳐서 8월 24일 (목) 오전 11시 59분까지 제출.

프로젝트 관련 실시간 Q&A 세션 (게더타운)

일자: 8월 12일 (토) 오후 7-9시, 8월 19일 (토) 오후 7-9시

진행방식: 이전에 진행되었던 게더타운 실시간 Q&A 세션과 동일.



수료 기준 Recap

- ✓ 제공되는 퀴즈 및 과제 모두 완수
- ✓ 퀴즈 & 과제 평균 50점 이상
- ✓ 최종 프로젝트 완수 (기획서 및 프로젝트)



우수자 시상 기준 Recap

우수 참가자

- ✓ 퀴즈&과제 평균 점수
- ✓ 최종 프로젝트 성과

우수 활동 팀

- ✓ 팀 프로젝트 성과
- ✓ 주제별 한 팀씩 시상

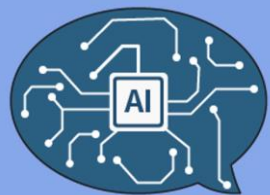


일정 Recap



일자	내용
8/1 (화)	프로젝트 OT
8/5 (토)	기획서 마감 (오전 11시 59분까지) / 팀별 프로젝트 선정 마감 (오전 11시 59분까지) 1차 프로젝트 스켈레톤 코드 업로드 (오후 12시)
8/8 (화)	2차 프로젝트 스켈레톤 코드 업로드 (오후 12시)
8/10 (목)	3차 프로젝트 스켈레톤 코드 업로드 (오후 12시)
8/12 (토)	프로젝트 관련 실시간 Q&A 세션 (게더타운, 오후 7-9시)
8/19 (토)	프로젝트 관련 실시간 Q&A 세션 (게더타운, 오후 7-9시)
8/24 (목)	팀 프로젝트 코드 제출 마감 (오전 11시 59분까지)





감사합니다

