

과제 제안서

과제 개요

제목	설명
과제 이름	데이터 분석 및 예측 모델 훈련
과제 기간	2024년 01월 30일 ~ 2024년 01월 31일 (진행 상황에 따라 조정)
과제 발표	2024년 02월 01일 (진행 상황에 따라 조정)
과제 설명	개별적으로 학습한 데이터 수집, 저장, 분석, 전처리, 예측 모델 개발 기법을 통합하여 전체 프로세스로 수행

과제 팀 구성

팀	인원
1팀	강건, 김효정
2팀	김민중, 정윤서
3팀	권도윤, 박다영, 오상민

과제 수행 요구사항

- 과제는 팀 단위로 진행되며 팀 별로 데이터 분석 및 1개 이상의 예측 모델을 개발해야 한다

- 과제는 다음의 단계에 따라 작업을 수행해야 한다
 - 주제 선정 및 가설 설정
 - 데이터 수집
 - 데이터 저장
 - 탐색적 데이터 분석 및 데이터 전처리
 - 통계기반 데이터 분석 및 검증
 - 예측 모델 개발

과제 수행 요구사항

- 데이터 수집은 파일, OpenAPI, 웹크롤링 등 필요한 방법을 선택하여 수행한다
- 데이터 저장은 파일 및 데이터베이스에 저장한다
- 탐색적 데이터 분석 및 전처리 과정은 다음의 작업을 포함해야 한다
 - 각 속성별 결측치 탐지 및 처리 (삭제, 대체 등)
 - 각 속성별 기초 통계량 및 분포 특성 파악
 - 각 속성별 이상값 탐지 및 처리
 - 속성간 상관관계 분석
 - 그 외 필요한 분석
- 통계 분석 과정은 다음의 작업을 포함해야 한다
 - 수치형 속성에 t 검정 수행
 - 범주형 속성에 카이제곱 검정 수행

과제 수행 요구사항

- 예측 모델링 과정은 다음의 작업을 포함해야 한다
 - 피처와 타겟 결정
 - 타겟의 데이터 특성에 따라 회귀/분류 모델링 방법 결정
 - 훈련 데이터 세트와 테스트 데이터 세트 분할
 - 적용 가능한 모든 알고리즘을 적용하여 예측 모델 수행
 - 알고리즘을 적용할 때 모델 파라미터(하이퍼 파라미터)를 변경하면서 과대 적합 및 과소 적합 상태 조정
- 개발된 모델은 강사가 제공하는 웹 애플리케이션에 통합하여 서비스한다
 - 이를 위해 강사와 협의를 통해 함수 또는 클래스 구현과 관련된 사항을 결정해야 한다

과제 수행 산출물

- 데이터 분석 기획서 (설명으로 대체)
- 분석 데이터 아키텍처 명세 (설명으로 대체)
- 데이터 분석 수행 노트북 소스 파일