

프로젝트 제안서

프로젝트 개요

제목	설명
프로젝트 이름	공공 데이터 활용 빅데이터 분석 프로젝트
프로젝트 기간	2024년 03월 06일 ~ 2024년 04월 03일
프로젝트 발표	2024년 04월 03일
프로젝트 설명	국내/해외 공공 데이터 및 다양한 데이터 서비스를 활용하는 데이터 분석 서비스를 설계하고 데이터 수집, 가공, 전처리, 분석 수행 및 예측 모델 개발

프로젝트 목표

- 프로젝트 수행을 통해 데이터 분석 및 예측 모델 개발 프로세스에 필요한 종합적인 역량 향상
 - ▷ 학습자는 실제 프로젝트 수행을 통해
 - ▷ 다양한 문제 상황을 경험하고 이를 해결하는 과정에서
 - ▷ 교육을 통해 학습한 기술 및 역량에 대한 성취도가 종합적으로 향상될 수 있으며
 - ▷ 데이터 분석 기획, 데이터 수집, 전처리, 분석, 예측 모델 개발 등의 업무 프로세스에서 실질적인 역할 수행이 가능한 수준으로 역량 향상

프로젝트 팀 구성

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

적용 기술 및 도구

- 적용 기술

분류	기술
개발 언어	Python (, HTML, CSS)
데이터 수집	File, Open API, Web Crawling
데이터베이스	DB Modeling, SQL, DBMS
데이터 전처리	Python
탐색적 분석	데이터 시각화, 통계 분석
예측 모델링	머신러닝, 딥러닝

적용 기술 및 도구

- 도구

분류	도구
코드 개발	Visual Studio Code
데이터베이스	OracleXE, SQLDeveloper, eXERD
데이터베이스 연동	pymysql
형상관리	github, git-client
탐색적 분석	numpy, pandas, matplotlib, seaborn 등
머신러닝	scikit-learn, xgboost, tensorflow

프로젝트 수행 요구사항

- 프로젝트는 팀 단위로 진행되며 팀 별로 데이터 분석 및 2개 이상의 예측 모델을 개발해야 한다
 - ▷ 예측 모델은 마이크로 데이터 기반 머신러닝 모델과 딥러닝 기반 모델을 각각 1개 이상 개발
 - ▷ 전체 팀원이 참여해서 역할을 수행할 수 있도록 프로젝트의 주제와 범위 결정
 - ▷ 각 팀 구성은 전체 팀 주제에 포함되는 개별 분석 작업을 수행하되 각자 데이터 수집, 전처리, 분석, 예측 모델 전체 과정을 수행해 볼 수 있도록 역할 분담
- 과제는 다음의 단계에 따라 작업을 수행해야 한다
 - ▷ 주제 선정 및 가설 설정 (서비스를 중심으로 주제 선택)
 - ▷ 데이터 수집
 - ▷ 데이터 저장
 - ▷ 탐색적 데이터 분석 및 데이터 전처리
 - ▷ 통계기반 데이터 분석 및 검증
 - ▷ 머신러닝, 딥러닝 기반 예측 모델 개발

과제 수행 요구사항

- 데이터 수집은 파일, OpenAPI, 웹크롤링 등 필요한 방법을 선택하여 수행한다

- 데이터 저장은 파일 또는 데이터베이스에 저장한다

- 데이터의 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
 - › 비정형 데이터의 경우 가능한 범위에서 분석 및 전처리 수행

과제 수행 요구사항

- 데이터의 통계 분석 과정은 다음의 작업을 포함해야 한다
 - 수치형 속성에 t 검정 수행
 - 범주형 속성에 카이제곱 검정 수행
 - 그 외 필요한 통계 분석
 - 비정형 데이터의 경우 가능한 범위에서 통계 분석 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
- 라이브러리 활용
 - 신경망 이외의 모델을 사용하는 경우 scikit-learn을 기본으로 사용하고 필요한 경우 다른 라이브러리 사용
 - 신경망 모델을 사용하는 경우 tensorflow + keras 또는 pytorch 사용

과제 수행 요구사항

- 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 (프로젝트 발표 문서에 포함)
- 데이터 분석 수행 일정 명세 (프로젝트 발표 문서에 포함)
- 프로젝트 발표 문서
- 프로젝트 시연 영상
- 데이터 분석 수행 노트북 소스 파일
- 데이터 분석 수행 노트북 HTML 출력 파일