**프로젝트 기획안**

**2021년 10 월 11 일**

**과정명 : 데이터 사이언스/엔지니어링 전문가**

|  |  |
| --- | --- |
| 팀 명 | 도라짱  혼저옵서예  제주어디가 |
| 팀 원 | 팀장: 김남경  팀원:김용호, 문준영, 여정문, 이종민, 이주남 |
| 프로젝트 주제 | 제주도 관광객(40대 이하)을 대상으로 카테고리별 여행지 추천 |
| 가상 시나리오 | 제주도를 방문하는 관광객을 대상으로 여행지 추천 서비스 사업을 추진하고자 함   * 사이언스 팀에는 다양한 데이터를 분석하여 사업 타당성을 검토하고 구축된 학습 데이터를 이용, 학습하여 모형화 작업을 진행 후 서비스로의 연계를 위한 모델 결과 도출 지시 * 엔지니어링 팀에는 제주도 관광지별 이용객 현황, 이동경로 데이터, 결제금액 데이터, 리뷰 데이터 등 모형화에 필요한 데이터를 수집, 가공하여 학습용 데이터 구축을 지시 * 서비스개발팀에는 효율적인 서비스 운영방안, 혁신적인 기술이 반영된 데모 서비스 개발 지시 |
| 프로젝트 수행 방향 | ● 데이터 수집 및 구축 개요  - 제주데이터허브([www.jejudatahub.net](http://www.jejudatahub.net))에 공개된 다양한 공공데이터 수집  - 비짓 제주 리뷰 텍스트, 트립 어드바이저 리뷰 텍스트 크롤링하여 수집  - AWS에 데이터 파이프라인 구축 (Hadoop 및 Spark 설치, 연결)  ● 주요 분석 내용  - 제주도 내 관광지별 방문객 추이 분석  - 이동경로 데이터, 결제금액 데이터, 리뷰 데이터 등 여러 데이터 간 상관관계 분석  - 적합한 데이터를 활용하여 추천 알고리즘 모형화  ● 기대 효과  - 사용자에게 적합한 제주도 내 관광지 추천 |
| 역할 분담 및 일정 | ● 역할분담  김남경 : 데이터 전처리, 데이터 모형화 작업  김용호 : 데이터 전처리, 데이터 모형화 작업  이종민 : 데이터 전처리, 데이터 모형화 작업  문준영 : 데이터 수집 및 전처리, 가공 및 시각화, 웹 배포  여정문 : 데이터 수집 및 전처리, 가공 및 시각화, 웹 배포  이주남 : 데이터 수집 및 전처리, 가공 및 시각화, 웹 배포  ● 일정  10/5 ~ 10/11 : 주제 선정 및 일정 수립  10/12 ~ 10/16 : 데이터 수집  10/18 : AWS 파이프라인 구축  10/19 ~ 10/22 : 데이터 모형화  10/23 ~ 10/25 : 데이터 시각화  10/26 ~ 11/7 : 웹 배포  11/8 ~ 11/11 : 발표 포트폴리오 작성 및 발표 준비  11/12 : 프로젝트 경진대회 |
| 프로젝트 수행 도구 | ● 빅데이터 수집, 저장 시스템  - Hadoop  - Spark  ● 분석기술  - pandas  - numpy  - BeautifulSoup  - request  - selenium  ● 시각화 기술  - matplotlib  - seaborn  - plotly  - visual analyst  - zepplin  - folium  ● 웹서비스 기술  - Django |