# 문제해결 빅데이터 활용 프로젝트 기획안

기획안 작성일자 : 2023-06-27

| 조 명          | 3조 : 3조 투어  |
|--------------|---|
| 조 원          | 조장: 최전석 조원: 한진서, 양미숙, 양정아, 이종우, 홍중기   |
| 프로젝트 주제 및 개요 | 개인 맞춤형 여행지 추천 서비스   |
| 프로젝트 수행 방향   | 1. 요구사항 명세화 a. 기능 요구사항 b. 비기능 요구사항 b. 비기능 요구사항 2. 테이터 수집 및 가공 a. 네이버 블로그 본문 크롤링 b. 국민여행조사 데이터 c. 인기관광지 데이터 3. 머신러닝 알고리즘 개발 a. 여행지에 대한 정보를 제공하는 알고리즘 i. 키워드 시각화, 감성분석을 통한 긍/부정 키워드 추출 b. 사용자의 특성에 맞는 여행지를 추천하는 알고리즘 i. 유사집단 기반 여행지 추천 ii. 사용자가 관심있는 관광지와 유사한 여행지 추천 4. 웹 서비스 개발 a. 직관적이고 가시성 높은 UI 개발 b. sqlite3 Database를 통한 데이터 연동 및 검색 c. 분류 결과별 URL 분리를 통한 쉬운 탐색 5. 클라우드 환경 구축 a. AWS를 통한 웹 서비스 제공 환경 구축 b. VM(가상머신)과 Hadoop을 통한 DB 구축 6. 서비스 테스트 a. 여행지 정보 제공 및 여행지 추천 알고리즘 테스트 b. DB 테스트 7. 서비스 배포 및 개선  ◆ 수행도구 -Front end HTML, CSS, python -Back end Django, hadoop, spark, python, SQL, AWS, VM, Open API |

# ● 데이터소개 국민여행조사 데이터 - 2019, 2022년도 여행객 4만명의 다양한 설문조사 항목 수록 네이버 블로그 링크 - 각 지역의 관광지, 명소 등의 키워드로 검색하여 나온 블로그의 링크. - Naver search API 사용 네이버 블로그 본문 - 네이버 블로그 링크를 통해 크롤링 - 이후 전처리한 본문 데이터 인기관광지 데이터 - 한국관광 데이터랩에서 수집한 인기관광지 데이터 - 각 시군구당 3개의 인기관광지 리스트 수집 ● 역할분담 최전석(조장, 데이터 엔지니어, 데이터 분석가): - 프로젝트 진행상황 관리(WBS) - 프로젝트 수행일지 작성 - 프로젝트 회의록 작성 - 요구사항 작성, 데이터 수집 환경 구축, VM 데이터베이스 구성, 외부 시스템 연동, DB 테스트, 머신러닝 알고리즘 개발 이종우(프론트 엔지니어): - 요구사항 작성, 웹 서비스 개발, 클라우드 환경 구축 프로젝트 조직 (구성원 및 역할) 양미숙(백엔드 엔지니어): - 요구사항 작성, 외부 시스템 연동, 서비스 배포 한진서(데이터 분석가): - 데이터 가공, 머신러닝 알고리즘 개발, 알고리즘 테스트, PPT 제작 양정아(데이터 분석가): - 데이터 가공, 머신러닝 알고리즘 개발, 알고리즘 테스트, PPT 제작 홍중기(데이터 분석가):

- 데이터 가공, 머신러닝 알고리즘 개발, 알고리즘 테스트

#### ● 일정

#### 6월 4~5주(6/20~6/30)

6/20~ 6/28: 주제 선정 및 일정 수립

6/24~ 6/24: 주제 검토

6/28~ 6/29: 요구사항 작성

6/29~ 6/30: 데이터 조사 및 수집, api 신청

### 7월 1주(7/03~7/07, 멘토링 7/08)

# [공통]

7/07~ 7/07: 2차 멘토링 발표 준비

#### [데이터 분석]

7/03~ 7/05: 도메인 지식 학습 및 문서화

7/05~ 7/06: 전처리 관련 소스코드 조사

7/07: 데이터 조사

7/08: 머신러닝 관련 레퍼런스 및 기법 조사

# [웹 서비스]

조사

7/03~ 7/03: AWS 계정 설정, 웹 디자인 기획(1)

7/04~ 7/04: 요구사항 분석, 웹페이지 구조도 수정 및 영문명 설정

7/05~ 7/05: 화면구상, UI/UX 디자인, csv파일 활용 코드 작성

7/06~ 7/08: AWS 환경설정, 프로그램 설치, django 코드 작성에 필요한 자료

# 프로젝트 추진 일정

7/08~ 7/10: 테이블정의서, 데이터 정의서 작성

#### [데이터 엔지니어]

7/03~ 7/03: 데이터 수집 경로 조사

7/04~ 7/04: Naver Search API 호출 및 호출된 데이터 저장

7/05~ 7/10: 데이터 처리 작업 로직 개발

# 7월 2주(7/10~7/14)

# [데이터 분석]

7/10: 데이터 조사

7/11~ 7/12: 국민여행조사 데이터 전처리

7/11~ 7/14: 블로그 데이터 전처리

7/12~ 7/12: 데이터 컬럼 정의서 작성

7/13~ 7/14: 키워드 추출 및 벡터화, 협업필터링 관련 자료 조사, 여행지 군집

화

7/14~ 7/14: 여행지별 블로그 데이터 감성분석

#### [웹 서비스]

7/10~ 7/11 : HTML 구조 작성 1차(HTML 문서 작성, 섹션 분리)

- 프로젝트 생성, 회원가입 모델 생성

7/12~ 7/12 :

CSS 스타일링 1차(스타일, 레이아웃 지정)

- 로그인, 로그아웃 구현, 테이블 설계

7/12~ 7/12: 웹페이지 구동 테스트(AWS), 테이블 설계

7/13~ 7/14: 데이터베이스 저장 테스트, 코드 작성 - 화면 입력 상세페이지

#### [데이터 엔지니어]

7/11~ 7/12: 데이터 처리 작업 로직 최적화 및 테스트

7/13~ 7/14: 크롤링 데이터 파이프라인 테스트

7/13~ 7/20: 크롤링 데이터 수집

#### 7월 3주(7/18~7/21, 멘토링 7/22)

#### [공통]

7/21~ 7/21: 3차 멘토링 발표 준비

#### [데이터 분석]

7/18~ 7/21: 협업 필터링 관련 자료 조사

7/18~ 7/22: 블로그 데이터 전처리, 키워드 추출 및 벡터화, 감성분석 및 여행

지별 블로그데이터 기반 긍/부정 키워드 추출 및 보완

7/19~ 7/22: 여행지 군집화 수행

7/20~ 7/22: 협업필터링을 이용한 여행지 추천 모델 개발

#### [웹 서비스]

7/18~ 7/19 : HTML 구조 작성 2차(상하좌우 배치 설정), 코드 작성 - 화면 입력, 상세페이지(1)

7/19~ 7/19 : CSS 스타일링 2차(세부스타일 지정)

7/20~ 7/20 : JavaScript 개발(동적 기능 구현, HTML에 코드 삽입), 코드 작성 -상세페이지(2)

7/21~ 7/21 : 웹페이지 프로토타입 개발 마무리

7/21~ 7/22 : 프로토타입 테스트

7/22~ 7/22 : 웹페이지에 실제 데이터 연동 및 테스트

#### [데이터 엔지니어]

7/18~ 7/18: Hadoop 환경 구축 및 설정, 데이터 다운로드 받는 방법 학습

7/19~ 7/19: Hadoop 데이터 적재 방법 학습, 데이터 적재

7/20~ 7/20: 데이터 품질 체크 및 데이터베이스 구축

7/21~ 7/21: 크롤링 데이터 파이프라인 다이어그램 제작, 네이버 API 호출, 크롤링 데이터 hadoop 업로드 내용 정리.

# 7월 4주(7/24~7/28)

#### [공통]

7/26~ 7/28: 프로젝트 발표 준비(PPT, 발표 스크립트 작성)

#### [데이터 분석]

7/24~ 7/24: 여행지 군집화 수행, 군집화 결과 및 유사도 기반 여행지 추천 알고리즘 개발, 감성분석을 통한 궁/부정 키워드 추출 보완, 협업 필터링을 통한 여행지 추천 알고리즘 개발

#### [웹 서비스]

7/24~ 7/26: 분석 코드 삽입과 오류 해결 및 테스트, 로컬 웹 호스팅
7/25~ 7/28: 웹페이지에 실제 데이터 연동 및 배포
[데이터 엔지니어]
- 데이터 분석 및 오류 해결 지원

7월 5주(7/31~7/31)
7/31~ 7/31: 프로젝트 발표