**기업요구사항 기반의 문제해결 프로젝트 기획안**

**2022년 4월 15일**

| **프로젝트 조** | **4조** | |
| --- | --- | --- |
| **프로젝트 팀원** | 팀명: 미안하다 4랑한다  팀장: 김채원  팀원: (사이언스) 윤형석, 이상동 (엔지니어링) 명재정, 전동준, 허정윤 | |
| **프로젝트 주제** | 서울시 맛집을 찾고 공유하는 플랫폼 ‘먹잘알’ | |
| **프로젝트 수행 방향**  **(주요 기능 설명)** | **< 벤치마킹 >**   * 1. 벤치마크할 사이트 (식신,다이닝 코드,망고 플레이트)를 선정해 시장 조사 진행.      1. 망고플레이트 : 리뷰페이지는 존재하지만 유저별 마이페이지나 키워드별 검색이 안 됨      2. 식신 : 리뷰페이지, 키워드별 검색은 가능하지만 유저별 마이페이지나 날씨를 이용한 테마별 검색은 불가능      3. 다이닝 코드 : 리뷰 기능은 잘 되어있지만 날씨를 이용한 테마별 검색은 불가능하고 지도에서 전체 검색은 불가능   2. 각 사이트마다 존재하는 기능과 존재하지 않는 기능을 분류한 뒤 우리가 사용할 수 있는 기능을 취합해 구현할 기능을 추가함 | |
| **사이언스** | **엔지니어링** |
| 음식점을 검색할 수 있는 대표 테마 및 키워드 선정   1. KoNLPy로 형태소 분류 및 불용어(stopwords) 제거 2. 자연어 처리(NLP),   군집화 (K-means)   1. LDA (토픽 모델링) | 1. 사용자 위치 중심으로 1km내 음식점을 가까운 순서대로 리스트화해서 추천  2. 검색기능   1. 음식 검색(ex 김치 -> 김치전, 김치찌개) 2. 대표 키워드 검색   3. 음식점들의 상세 정보를 가져올 때, 다양한 곳에서 정보들을 가져오기 때문에, 데이터가 중복될 가능성이 매우 높음. 따라서, 검색 키워드를 ‘구군 + 상호명’ 형태로 크롤링을 진행할 계획  4. 관리자,회원과 비회원에 따른 권한부여  5. 리뷰 작성 시 사용자가 업로드할 사진의 위치(메타)데이터와 해당 음식점의 위치데이터를 비교해 오차범위 내에서 일치한다면, 리뷰 업로드를 가능하게 함 |
| **프로젝트 수행 도구** | **사이언스** | **엔지니어링** |
| Python, Jupyter notebook, Pandas, Numpy, Tensorflow, Django | Python, Hadoop, Spark, MySQL, MongoDB, Airflow, ElasticSearch, Django, AWS |

| **프로젝트 목적** | **사용자 친화적인 SNS 서비스를 만들어 음식과 맛집에 대해서 적극적으로 소통할 수 있도록 서비스하는 것** | | |
| --- | --- | --- | --- |
| **필수 기능**  **(구현 기능)** | **사이언스** | **엔지니어링** |  |
| 1. 리뷰 데이터에서 자연어 처리/ 토픽 모델링/군집화를 활용하여 키워드 추출 | 1. 회원 가입 서비스 2. 검색기능 3. 위치 탐색 4. 사용자 위치기반 음식점 검색 5. 사용자 정보 조회 |  |
| **포함 기술**  **(수행 도구에 대한 세부 내용)** | **사이언스** | **엔지니어링** |  |
| 1. 엔지니어링에서 넘겨받은 데이터로 Jupyter Notebook에서 Tensorflow를 이용하여 Word2Vec등의 단어 임베딩(문서 유사도 비교)하거나 LDA등의 토픽 모델링 작업 2. 유클리드 거리/코사인 유사도 등을 활용하여 서비스를 위한 최종 키워드 도출 | 1. API와 크롤링으로 수집한 데이터를 Hadoop에 저장 2. Hadoop에 저장한 데이터 Spark에서 처리 3. Spark에서 처리한 데이터를 MySQL의 테이블로 저장 4. Spark에서 처리한 데이터   (위치정보를 필수로 포함)한 데이터 MongoDB에 저장   1. MongoDB의 Geojson을 사용해 사용자 기준 1km 반경 내 특정 음식점들의 리스트를 생성 2. ElasticSearch로 검색 기능 구현 3. AirFlow를 사용해서 정기적으로 파이프라인 관리 |  |

**□ WBS 첨부 ( OR 별도로 폴더 업로드)**