**기업요구사항 기반의 문제해결 프로젝트 기획안**

**2022년 8 월 26일**

| **프로젝트 조** | **1조 (연봉일조)** | |
| --- | --- | --- |
| **프로젝트 팀원** | 팀장: (사이언스) 우성균 (엔지니어링) 유서원  팀원: (사이언스) 송창화, 안세원 (엔지니어링) 문영호, 최유진 | |
| **프로젝트 주제** | 채식 레시피 추천 서비스 | |
| **프로젝트 수행 방향**  **(주요 기능 설명)** | **사이언스** | **엔지니어링** |
| 사용자 리뷰 기록을 바탕으로 사용자 맞춤 , 중요도 가중치 선택 시 채식 레시피 추천   1. 레시피 재료 기준 클러스터링하여, 레시피를 분류 2. 사용자의 레시피에 대한 ‘좋아요’ 로그를 활용하여 콘텐츠 기반 필터링 3. 맛,영양,가격등의 관심도를 바탕으로 고객집단을 세그먼트화, 각 집단에 맞는 추천을 하는 협업 필터링 4. 트위터 API를 활용하여, 비건 레시피 트렌드 시각화 5. 농산물 가격 API를 활용하여, 가격 급락폭이 큰 재료 및 관련 레시피 추천 | \* Spark Streaming을 이용한 실시간 트윗 / 채소 가격 수집  \* Airflow를 통한 Hadoop, Spark, MySQL, MongoDB 사용 자동화  - Airflow: 스케줄러 및 오케스트레이션  - Hadoop: 원본 데이터 적재  - Spark: 분석 활용을 위한 데이터 가공 및 처리  - MySQL: 가공된 데이터 적재  - MongoDB: 로그 데이터 저장  \* Elastic Search를 이용한 검색엔진 구현  \* Plotly/Kibana를 이용한 데이터 시각화  \* Django를 이용한 서비스 페이지 구현 |
| **프로젝트 수행 도구** | **사이언스** | **엔지니어링** |
| \* 텍스트 전처리  \* Clustering  \* Doc2Vec(for CBF)  \* Collaborative Filtering(CF)  \* Matrix Factorization(MF)  \* Surprise  \* 딥러닝 | \* Airflow  - Spark Streaming  - Hadoop  - Spark  - MySQL  - MongoDB  \* ELK  - Elastic Search  - Logstash  - Kibana  \* Django |

| **프로젝트 목적** | 채식 메뉴 순위별 추천 | | |
| --- | --- | --- | --- |
| **필수 기능** | **사이언스** | **엔지니어링** |  |
| \* 클러스터링을 활용한 레시피 카테고리 분류  \* 콘텐츠 기반 추천  \* 협업 필터링 추천  \* 재료 가격 변동에 따른 레시피 추천 | \* 실시간 채소 가격 수집 및 적재  \* 채식 식단 및 식당 데이터 수집  \* 데이터 전처리  \* 데이터 적재  \* 위치 데이터 활용  \* 검색엔진을 통한 검색기능 |  |
| **포함 기술** | **사이언스** | **엔지니어링** |  |
| \* LDA(Latent Dirichlet Allocation)를 통한 토픽 모델링  \* CBF을 위하여 레시피-재료 데이터를 Doc2Vec을 사용  \* Collaborative Filtering(CF), Matrix Factorization(MF), Surprise, 딥러닝을 통한 추천 알고리즘 설계  \* [공통] Django를 이용한 서비스 페이지 구현 | \* spark streaming/kafka를 통한 실시간 채소 가격 수집  \* Kibana를 이용한 시각화  \* Airflow를 통한 Hadoop, Spark, MySQL, MongoDB 사용 자동화  \* Elastic Search를 이용한 검색엔진 구현  \* [공통] Django를 이용한 서비스 페이지 구현 |  |

**□ WBS 첨부 ( OR 별도로 폴더 업로드 )**