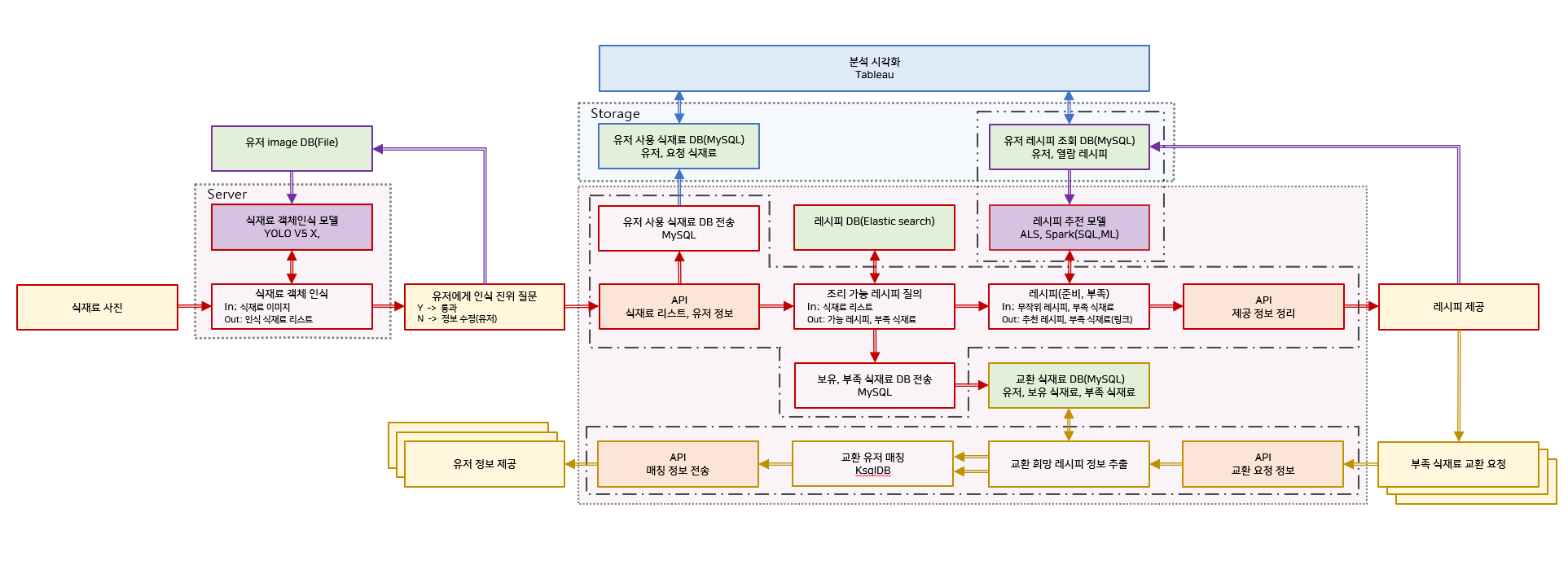
**기업요구사항 기반의 문제해결 프로젝트 기획안**

2022년 07월 22 일

| 프로젝트 조 | 1조 | |  |
| --- | --- | --- | --- |
| 프로젝트 팀원 | 팀명: 오늘은뭐먹조 팀장: 김건우  팀원: (DS) 고기원, 백지헌, 이은평, 이지원  (DE) 계은서, 정현식 | | |
| 프로젝트 주제 | 도마를 부탁해 | | |
| 프로젝트 수행 방향 | 사이언스 | 엔지니어링 | 앱, 웹 개발 |
| - 식재료 객체 인식 모델 개발 - 레시피 추천 시스템 구축(컨텐츠기반) - 서비스 관련 사회 조사 분석 - 사용자 데이터 시각화(태블로) | - 레시피 DB 구축 - 전체 데이터파이프 구축 - 레시피 추천 시스템 구축(협업필터링) | - 웹브라우저 개발 - 서버 통신 |
| 프로젝트 수행 도구 | 사이언스 | 엔지니어링 | 웹 개발 |
| - Python, Pytorch - Labeling(bounding Tool)  - Pandas - Tableau | - Python - Hadoop, Spark - Kafka, KSQLl, Airflow - Elastic Search, MySQL - AWS, Docker | - Flask, HTML, CSS, JS |

| 프로젝트 목적 | 식재료를 통한 요리 추천, 추가 식재료 추천 | | |
| --- | --- | --- | --- |
| 필수 기능 | 사이언스 | 엔지니어링 | 앱 개발 |
| -Yolo v5 x를 통한 객체 인식 및 식재료 리스트 추출 -컨텐츠 기반 레시피 추천시스템 설계 및 구축 -서비스 이용 데이터 분석 시각화 설계 및 구축 | - Kafka를 활용한 서비스 데이터 파이프라인 설계 및 구축  -Elastic Search를 활용한 레시피 DB 구축 및 질의기능 구현 -사용자 기반 협업필터링 추천 시스템 설계 및 구축  -서비스 이용 데이터 DB구축 -KSQL을 활용한식재료교환 매칭 시스템 설계 및 구축 | - 로그인(보류) - 파일업로드 - 식재료 인식 요청 - 레시피 추천 요청 - 식재료 교환 요청 |
| 포함 기술 | 사이언스 | 엔지니어링 | 앱 개발 |
| - Yolo v5 x 모델 개발 - Image augmentation (OpenCV, PIL, ImageDataGenrator:Keras) | Elastic Search: \_search KSQL: Stream-Straem join Spark: ALS Googel Image Crawler 개발 | - 로그인(보류) - 파일업로드 - Yolo API - main service API - matching service API |

# Service PIPELINE



**# 단계별 기획**

▷ 1-0. 프로토타입 테스트

| 소주제 | 소스 코드 정상 작동 확인 및 성능 확인 |
| --- | --- |
| 분석개요 | ● 분석 방식 별 성능 확인 및 세부 모델 선정 |
| 필요데이터 및 도구 | 1. 데이터  * 피스톨 사진 (2,970장) * 참고자료   + [분석 방식 성능 비교 논문](https://www.koreascience.or.kr/article/JAKO202105653790745.pdf)   + [yolov5 X 소스 코드](https://github.com/ultralytics/yolov5/releases)   + [efficientdet 소스코드](https://github.com/xuannianz/EfficientDet)  1. 도구  * Yolo V5 (X) |
| 프로젝트  시나리오 | 1. 객체인식 소스코드 확보 2. 작동여부 확인 3. Output 정확도 확인함 4. 모델 성능 비교 및 모델 선정 |
| 활용 방안 | 향후 모델 프레임 기반에 활용 |

▷ 1-1. 데이터 수집(수집, 저장, 탐색, 라벨링, 전처리)

| 소주제 | 객체인식 모델 정보 검색 및 이미지 수집 |
| --- | --- |
| 분석개요 | ● 모델에 대한 정보 수집  ● 모델 성능테스트를 위한 데이터(sample dataset) 수집 및 전처리 |
| 필요데이터 및 도구 | 1. 도구  * Labeling(bounding Tool)  1. 식재료 이미지  * 식재료 리스트 * 바운딩처리 필요 데이터 수: 식재료당 1,000개 |
| 프로젝트  시나리오 | 1. 식재료 이미지 데이터 수집 및 저장  * 수집 : Kaggle Data set, 구글 이미지 크롤링 등 (크롤러 개발) * 저장 : 구글드라이브 path: Drive/DS/dataset * Image augmentation   + OpenCV   + PIL   + ImageDataGenrator(Keras)  1. 데이터 전처리  * Resize * Bounding   + Labeling(bounding Tool)  1. EDA 실행  * 데이터 편향성 및 적합성 검사 * 모델 성능 확인 |
| 활용 방안 | Yolo v5 x 모델 훈련 데이터 확보 및 전처리 |

▷ 1-2. 모델링

| 소주제 | Yolo v5 x 모델링 |
| --- | --- |
| 분석개요 | 프로토 타입 모델 프레임을 활용한 모델링 및 성능향상 |
| 필요데이터 및 도구 | * 데이터 * 1-1에서 수집한 이미지 데이터 활용 * 모델 개선 방안 수집 * 도구 * Yolo v5 x |
| 프로젝트  시나리오 | 1. 모델 학습  * 전처리 데이터(이미지, 라벨링) 학습 * 서비스에 활용할 수 있는 가중치 확보  1. 모델별 성능 및 개선점 확인  * 백그라운드이미지 활용, 최적의 하이퍼파라미터 적용 |
| 활용 방안 | * 객체 인식 서비스 모델로 활용 |

▷ 2-1.

| 소주제 | 데이터 파이프라인 개요 |
| --- | --- |
| 분석개요 | ● 전과정 데이터 파이프라인 소개 |
| 필요데이터 및 도구 | ㅇ 데이터 및 수집 데이터   * 레시피 DB : 만개의 레시피 * 서비스 이용 데이터: 가데이터 생성   ㅇ 도구   * 파이프라인 : Kafka * 레시피 DB 구축 및 검색 도구 : Elastic Search * DB : Elastic Search, MySQL * 사용 데이터 가공 : SparkSQL * 추천 머신러닝 : SparkML(ALS) * 추천 모델 업데이트 스케줄링 : Airflow * 식재료 교환 매칭 기능 : KSQL |
| 프로젝트  시나리오 | ㅇ 서버   1. 요청정보 Kafka 파이프에 전송 후 데이터 리턴  * 요청 정보 : 서비스이용 서비스 이용 정보(식재료)  1. 레시피DB에서 요청정보로 요리 가능한 레시피 검색 및 추출  * 바로 요리 가능한 레시피와 식재료가 부족한 레시피로 구분 * 부족한 식재료 구매링크 첨부 * 사용자 교환 식재료 DB에 데이터  1. 추천시스템으로 레시피 랭크  * 콘텐츠 기반 * 사용자 기반 협업필터링  1. 추천 레시피 전송 |
| 활용 방안 | 메인서비스 파이프 구축 |

2-2.

| 소주제 | DB 관리 |
| --- | --- |
| 분석개요 | ● DataBase 설계 및 구축 |
| 필요데이터 및 도구 | ㅇ 데이터   * 만개의 레시피 데이터(크롤링) * 서비스 이용 데이터   + 사용자 서비스 요청 데이터   + 사용자 레시피 열람 데이터   ㅇ 도구   * Elasticsearch * MySQL |
| 프로젝트  시나리오 | ㅇ Server DB   1. 레시피  * 레시피 DB - elastic search   + recipe\_code, recipe\_nm, ingredients\_nm, qnt, cooking\_time,level\_nm, recipe\_url, img\_src * 레시피 meta data - mysql   + recipe\_nm, recipe\_code * 식재료 meta data - mysql   + ingredients, ingredients\_code  1. 교환 요청 DB  * 사용자 교환 식재료 DB 구축 - mysql   + userid, offer\_ingredientes\_code, lack\_recipe, lack\_ingredients\_code, timestamp, latitude, longitude   ㅇ Storage   1. 사용자 서비스 데이터  * 사용자 서비스 요청 데이터 - mysql   + userid, ingredientes, basic)ingredients lack\_recipe, lack\_ingredients\_code, timestamp, latitude, longitude * 사용자 레시피 열람 데이터 - mysql   + userid, recipe\_code, recipe, timestamp, latitude, longitude |
| 활용 방안 | 검색에 용이한 레시피 DB 구축 |

2-3.

| 소주제 | 레시피 추천 시스템 |
| --- | --- |
| 분석개요 | ● 식재료에 의한 요리 서치 이후, 사용자에 적합한 요리 추천 |
| 필요데이터 및 도구 | 1. 콘텐츠 기반 추천 시스템  * 레시피 타이틀 활용  1. 사용자 기반 협업 필터링 추천 시스템  * ALS(SparkML)  1. 랭킹 기반 추천 시스템  * 누적 사용자 열람 데이터 |
| 프로젝트  시나리오 | 1. 콘텐츠 기반 추천 시스템(보류)  * 레시피 타이틀에서 특정 키워드 추출 * 레시피 별 추출 키워드 태킹 * 레시피 열람시 태킹 키워드 사용자에 누적 * 사용자 누적 태킹에 따른 레시피 추천  1. 사용자 기반 협업 필터링 추천 시스템  * 사용자 누적 열람 데이터를 활용하여 ALS 모델 훈련 * 사용자 별 ALS 모델에 의한 레시피 추천 |
| 활용 방안 | 1. 다양한 방식의 추천 시스템 구축 |

2-4.

| 소주제 | 부족한 식재료 구매 및 교환 서비스 |
| --- | --- |
| 분석개요 | ● 1, 2개의 식재료 부족으로 인해 요리 불가능 시, 식재료 확보 여건 제공 |
| 필요데이터 및 도구 | 1. 식재료 구매링크 2. 식재료 교환 사용자 요청  * KsqlDB |
| 프로젝트  시나리오 | ㅇ 구매   1. 사용자의 부족 식재료별 구매링크 첨부   ㅇ 교환   1. 사용자 식재료 교환 요청정보로 ‘교환 식재료 DB’에서 매칭데이터 검색 및 추출 2. KSQL DB에 매칭데이터 전송 및 Stream-Stream Join 3. 매칭시 매칭 유저 정보 전송 |
| 활용 방안 | 1. 부족한 식재료의 확보안을 제공하여 서비스 이용 자유도 향상 |

3-1.

| 소주제 | 웹 브라우저 개발 |
| --- | --- |
| 분석개요 | ● 사용자 편의를 위한 서비스 개발 |
| 필요데이터 및 도구 | Flask, HTML, CSS, JS |
| 프로젝트  시나리오 | 1. 로그인 기능(보류)  * 개인 정보를 담아 서비스 이용 편의성 확보 * 매칭 시스템 활용  1. 이미지 업로드 기능 2. 업로드 이미지 객체 인식 모델 서비스 요청 3. 서비스 레시피 추천 서비스 요청 4. 부족 식재료 매칭 서비스 요청 |
| 활용 방안 | 1. 사용자 편의 성을 고려한 서비스화 |

4-1.

| 소주제 | 서비스 이용 데이터 분석 시각화 |
| --- | --- |
| 분석개요 | ● 마케팅시점으로 이용가능한 서비스 분석 |
| 필요데이터 및 도구 | 1. 데이터  * 사용자 서비스 요청 데이터 * 사용자 레시피 열람 데이터  1. 도구  * Tableau |
| 프로젝트  시나리오 | 1. 유저(타겟) 측면   * 특정 성별 유저수 * 특정 연령대 유저수 * 특정 지역 유저수(시/도단위 행정구역 기준) * 일4구간(아침/점심/저녁/야식)별 유저수 * 유저별 선택한 레시피 리스트 * 날짜/시간별 선택한 레시피 리스트   2. 서비스(시스템) 측면   * 일일 이용자 수 증감(전일 대비), 누계 * 유저특성별 일일 이용자 수(연령대,성별,지역별) * 월간 일평균 이용자 수 (DAU) * 연간 월평균 이용자 수 (MAU) * 식재료 인식 후 레시피 선택까지의 경과시간 * 유저 이탈율 (유지/이탈 분석)   3. 컨텐츠 측면   * 주간/월간 인기 레시피(조회빈도수) * 주간/월간 인기 식재료(인식빈도수) * 유저특성(연령대/성별/지역)별 인기 레시피 * 계절별(분기별/월별) 인기 레시피 * 일4구간(아침/점심/저녁/야식)별 인기 레시피 * 레시피별 조회수 변화 비교 |
| 활용 방안 | 향후 서비스 개선에 활용 될 수 있는 인사이트 제공 |

5-1.

| 소주제 | 서비스 사회 조사 및 분석 |
| --- | --- |
| 분석개요 | ● 서비스 시장성 확보 및 서비스 고객, 수요 예측 |
| 필요데이터 및 도구 | ㅇ KOSIS(통계청), 환경부, RISS, Miri Canvas, |
| 프로젝트  시나리오 | <홈쿡의 관심도가 증가하게 된 요인 파악>  ㅇ 통계 자료 조사   * 통계자료 중 ‘집에서 직접 요리를 해서 끼니를 해결한다’가 비율이 높았는데 왜 이런 결과가 나왔는지 요인들을 파악하고 그와 관련된 시장이 어떻게 발전되었는가 조사   ㅇ 코로나19  - 재택 근무, 원격 수업 등으로 주거 공간 위주의 생활  ㅇ 배달비 인상  - 배달 서비스에 대한 부담 증가  ㅇ 물가 상승  - 외식비 부담 증가  ㅇ 요리 관련 크리에이터  - 요리 블로그 방문자 수 증가  - 요리 영상 조회수 증가  <경쟁사 조사>  ㅇ 만개의레시피  - 사용자 리뷰로 단점 확인, 단점 개선 가능 여부 조사  <시장 조사>  ㅇ 레시피 공유 서비스를 하는 어플들의 이용자 수 확인  ㅇ 요리 블로그 방문자 수 추이 확인  ㅇ 요리 영상 조회수 추이 확인 |
| 활용 방안 | 사회 조사 및 시장 조사를 통해 ‘건우네 반찬’ 서비스의 타당성과 경쟁성 확인 |
|  |  |