|  |
| --- |
| **빅데이터 분석 정의서** |

|  |  |
| --- | --- |
| **과제명 :** | **원하는 재료로 레시피를 추천해주는 서비스** |

**2023.04.20**

**Ⅰ. 개요**

1. 아이디어 주제

: 원하는 재료로 레시피를 추천해주는 서비스

2. 개발 목표

: 식품 이미지 분석 후 분석된 항목과 직접 입력 받은 항목들을 재료로 레시피를 추천 서비스

3. 개발 내용

: 사진으로 재료 판별 서비스

: 원하는 재료들로 레시피를 추천해 주는 서비스

**Ⅱ. 기능별 빅데이터 분석 명세서**

\*기능별로 정의되어야 함

|  |  |
| --- | --- |
| **기능명** | 사진으로 재료 판별 기능 |
| **데이터 준비** | |
| **데이터 정의** | 9종의 요리재료 사진데이터  (마늘, 양파, 대파, 고추, 무, 배추, 당근, 감자, 계란) |
| **데이터 획득 방법** | 1. 식재료 사진 구글 이미지에서 각 재료별 약 500장씩 크롤링  2. kaggle에서 데이터 수집  (<https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset>  https://www.kaggle.com/datasets/moltean/fruits/code  https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition) |
| **2. 전처리** | |
| **전처리 과정** | 크롤링 이미지중 맞지않는 데이터 제거,  크롤링, kaggle 이미지 크기 224x224로 변경  Keras의 Imagegenerator를 사용해서 이미지 Rotatingm, shift, fliping등을 통해 이미지 데이터 증강 |
| **3. 모델 생성 및 학습** | |
| **모델링 목표** | 요리재료 종류 예측 |
| **모델링 가능 알고리즘** | 요리재료 종류 예측 : VGG-16, ResNet, AlexNet |
| **학습** | 요리재료 종류 예측 : 존재하는 모델들을 우리 기준에 맞게 약간 변형하여 학습 |
| **4. 검증** | |
| **모델링 검증 방안** | 검증용 이미지데이터 확보 (약 90장)  확보한 이미지로 모델 검증 진행 |
| **모델링 평가 결과** | VGG-16(약 85%) > resnet(약 20%) > alexnet(약 10%) |

|  |  |
| --- | --- |
| **기능명** | 레시피 추천 기능 |
| **데이터 준비** | |
| **데이터 정의** | 약 11만개의 레시피 데이터 |
| **데이터 획득 방법** | 1. 무료 레시피 데이터 - KADX 농식품 빅데이터 거래소 에서 만개의레시피 csv파일 획득  (https://kadx.co.kr/product/detail/0c5ec800-4fc2-11eb-8b6e-e776ccea3964) |
| **2. 전처리** | |
| **전처리 과정** | 빈항목이 있는 레시피 제거,  레시피에서 저장할 데이터만 분리,  Countervectorizer를 사용하기위해 재료명 변환한 테이블 추가  재료명 테이블 기존 레시피 테이블과 분할  재료명 | 를 기준으로 분할,  대표 재료명으로 재료명 변환  재료단위 통합 |
| **3. 목표 및 수행내역** | |
| **목표** | 레시피 추천 |
| **알고리즘** | Content\_based\_filtering(CountVectorizer, cosine\_similarity) |
| **수행 내용** | Content\_based\_filtering 으로 먼저 재료명 일치 데이터를 가져온뒤,  가져온 레시피와 보유중인 레시피 재료 양을 비교하여 순위를 설정하고,  추가적으로 필요한 금액을 산출해준다. |