| 빅데이터 분석 정의서 |
| --- |

| 주제 : | 사과당도 품질데이터를 통한  사과 당도등급 예측 인공지능 서비스 |
| --- | --- |

2022.08.26

Ⅰ. 개요

1. 아이디어 주제

: 사과당도 품질데이터를 통한 사과 당도등급 예측 인공지능 서비스

1. 개발 목표

: 농장에서 직접 사용가능한 사과 당도 예측 서비스

1. 개발 내용

: 환경데이터를 통한 당도등급 예측 서비스

: 사과 이미지를 통한 당도등급 예측 서비스

Ⅱ. 기능별 빅데이터 분석 명세서

| **기능명** | 환경데이터를 통한 당도등급 예측 서비스 |
| --- | --- |
| **1.데이터준비** | |
| **데이터 정의** | 총 4가지 사과종류에 대한 환경데이터  (홍로,아리수,시나노골드,후지) |
| **데이터 획득 방법** | AI HUB 데이터 활용(전북장수 사과당도 품질데이터)  <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100> |
| **2. 전처리** | |
| **전처리 과정** | ⚬ 불필요 특성 및 결측값 제거  (촬영농가위치, 파괴당도, 비파괴당도, 파일종류, 라이센스관련 내용 등)  ⚬ 원핫인코딩, 라벨 인코딩 |
| **3. 모델 생성 및 학습** | |
| **모델링 목표** | 당도 등급 예측 |
| **모델링 가능 알고리즘** | 머신러닝 앙상블 모델(BaggingClassifier) |
| **학습** | 훈련용 데이터셋 활용 훈련 실시(약 4만개) |
| 1. **검증** | |
| **모델링 검증 방안** | 검증용 데이터셋 활용 검증 실시(약 3만개) |
| **모델링 평가 결과** | 당도등급 예측 결과(정확도) = 0.9515 |

| **기능명** | 사과이미지를 통한 당도등급 예측 서비스 |
| --- | --- |
| **1.데이터준비** | |
| **데이터 정의** | 총 4가지 사과종류에 대한 환경데이터 및 이미지 데이터  (홍로,아리수,시나노골드,후지) |
| **데이터 획득 방법** | AI HUB 데이터 활용(전북장수 사과당도 품질데이터)  <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100> |
| **2. 전처리** | |
| **전처리 과정** | ⚬ MASK R-CNN기법을 통한 이미지에서 사과부분 추출 후 segmentation 좌표 획득  ⚬ segmentation 좌표, img\_file\_name을 특성으로한 데이터프레임 생성후 저장  ⚬ 환경데이터셋의 당도등급과 통합후 전처리(불필요 특성제거, 원핫인코딩 등)  ⚬ 불필요 특성 및 결측값 제거  (촬영농가위치, 파괴당도, 비파괴당도, 파일종류, 라이센스관련 내용 등) |
| **3. 모델 생성 및 학습** | |
| **모델링 목표** | 이미지를 통한 당도 등급 예측 |
| **모델링 가능 알고리즘** | MASK R-CNN, 머신러닝 랜덤포레스트 모델(RandomForestClassifier) |
| **학습** | 일부 훈련용 데이터셋 활용 훈련 실시(이미지 600개) |
| 1. **검증** | |
| **모델링 검증 방안** | 일부 검증용 데이터셋 활용 검증 실시(이미지 480개) |
| **모델링 평가 결과** | 당도등급 예측 결과(정확도) = 0.3908 |