**분석 설계서**

2022.04.22.

|  |  |
| --- | --- |
| **분석 목표 및**  **환경 정의** | **1. 분석 목표**   * 이미지 학습을 통한 식물과 식물병을 인식 * 해당 식물을 처음 접하는 사람들에게 식물에 관한 상세 정보 제공 * 반려식물의 관리를 위해 식물병 인식 후 관련 정보 제공   **2. 분석 환경 및 도구**   * 분석 도구: Python 3.7.13, Keras 2.8.0, Tensorflow 2.8.0, Pytorch 1.10.0+cu111 * 분석 환경: Colab, AWS   **3. 평가 기준 (BaseLine Model)**   * ResNet50 정확도 59% / 손실 값 * VGG16 정확도 79% / 손실 값 1.236 * YOLOv5 아직 진행 중   → 지금까지 정확도가 가장 높은 VGG16 모델을 베이스라인 모델로 선정 |
| **데이터 탐색**  **(EDA, 전처리)** | **1. 데이터 양**   * 구글, 네이버 크롤링한 이미지 사용 * 식물 종류: 금전수(1508개), 스투키(362개), 테이블야자(990개), 여인초(565개), 몬스테라(425개)   + 추후 식물 추가 예정 / 학습 데이터 편차 줄이기 위해 추가 크롤링 진행 중   * 식물병 종류: roboflow 데이터 사용: 흰가루병, 탄저병   **2. 데이터 정제 및 변환:** 결측치, 이상치 처리   * 이미지 데이터 중 학습과 관련 없는 사진 삭제 및 자르기 |
| **분석 피처 및**  **데이터 선정** | **1. 학습/검증/테스트용 데이터 선정**  (1) 식물 인식   * ResNet, VGG16: 전체 데이터 중 학습 50%, 검증 25%, 테스트 25%로 분할 * YOLOv5: train\_test\_split을 사용하여 학습 80%, 검증 10%, 테스트 10%로 분할   (2) 식물병 인식 (예정)   * ResNet, VGG16: 전체 데이터 중 학습 50%, 검증 25%, 테스트 25%로 분할 * YOLOv5: train\_test\_split을 사용하여 학습 80%, 검증 10%, 테스트 10%로 분할 |
| **기본 분석 모델 구현** | **1. 분석 모델 선정**   * + - 식물인식, 식물병인식 모두 ResNet50, VGG16, YOLOv5 모델링 진행     - 모델링 결과가 가장 좋은 모델로 추후 선택 예정   **2. 훈련/예측/검증 방안**   * 훈련   - ResNet50 : 이미지 사이즈 224X224, epochs 50, batch\_size 24로 설정  - VGG16 : 이미지 사이즈 224X224, epochs 20, batch\_size 20으로 설정  - YOLOv5 : 이미지 사이즈 224x224, epochs 3, batch\_size 1로 설정   * 검증   + 정확도와 loss 값을 통해 학습되지 않은 이미지에서 얼마나 정확하게 식물 종류를 예측하는지를 검토함으로써 성능을 검증 * 예측 * 학습과 검증에 사용하지 않은 test 데이터를 학습된 모델을 통해 정확도와 손실 값으로 성능 확인 |
| **모델 성능 개선** | **1. 모델 최적화를 위한 하이퍼파라미터 및 최적화 기법 선정**   * ImageDataGenerator를 통해 이미지 증식 * Optimizers(옵티마이저) 변경 : Adam, SGD, RMSProp, Adagrad 등 * Dropout(드롭아웃) : 임의의 뉴런을 골라 삭제하여 신호가 전달되지 않게 함 (과대적합 방지)   **2. 예측 성능 검증 및 평가 방안**   * 학습과 검증에 사용하지 않은 test 데이터를 학습된 모델을 통해 정확도와 손실 값으로 성능 확인 |
| **서비스 적용을 위한 분석 모델 도입 방안** | 1. 식물 이미지 인식 모델링을 통해 식물명 예측  2. 식물병 이미지 인식 모델링을 통해 식물병 예측 |
| **분석 일정 및 담당자** | **1. 식물 인식 모델링**   * 기간: 04월 13일~04월 24일 * 담당: ResNet50(강교철), VGG16(박혜인), YOLOv5(이유진)   **2. 식물병 인식 모델링**   * 기간: 04월 24일~04월 26일 * 담당: ResNet50(이유진), VGG16(강교철), YOLOv5(박혜인) |