

# 프로젝트 평가가이드

## 미션 정보

### 평가 미션

프로젝트 ID	프로젝트 이름	미션 ID	미션 이름
3	계량기 읽기	5	수도계량기 숫자인식 머신러닝을 위한 데이터 분류
코드 자동 평가	퀴즈	동료 평가	멘토 평가
	√	√	

### 미션 설명

- 이전까지의 미션에서는 개별 숫자 이미지를 crop 을 통해 얻을 수 있었고, Google tesseract 를 통해 개별 숫자 이미지를 인식할 수 있었다.
- 그 결과 모든 이미지들에 대해서 개별 낱자를 crop 할 수 있었고, 개별 낱자들을 인식하여 labeling 을 붙임으로써 dataset 을 구성할 수 있게 되었다.
- Google tesseract 를 이용하여 자동 레이블링을 하고, 값의 오류가 있는 부분들은 수작업으로 정확한 레이블링을 한다.
- 정확한 레이블링이 된 낱자 숫자 파일들을 Numpy ndarray 를 이용 머신러닝의 training data 의 container 로 사용하는 포맷으로 packing 하고 저장한다.

### 결과물

- 낱글자 숫자 이미지화일들에 대해 각 파일에 대한 정확한 레이블링 정보 및 이를 처리하는 파이썬 프로그램

### 평가 항목

구분	평가 항목
Data Classification	데이터 분류 과정에서 발견된 결측치를 제거하였는가?
	훈련 데이터 셋과 테스트 데이터 셋을 적절한 비율로 분류하였는가?

### 퀴즈 문항

분류	문제	정답
Data Classification	Google Cloud 문자인식 서비스의 정확도는?	프로그래밍 편의성 면, 인식의 정확도 면
	글자 낱자로 인식하는 것과 몇개 글자를 붙여서 인식하는 것이 어떤 차이가 있는가?	인식의 정확도 면

어떤 관점으로 학생의 미션을 평가해야 하는가?

해당 미션을 통해 학생은 매우 많은 수의 이미지 파일을 다루어 보는 것을 경험해볼 수 있다. 따라서 실제 머신러닝을 진행할 때 느낄 수 있는 레이블링 과정에서의 고충을 이해할 수 있고, 그러한 고충을 해결하기 위해 효과적인 방법을 모색할 수 있다.

해당 미션에서는 학생의 기술적인 역량도 중요하지만 문제를 해결하고자 하는 학생의 의지가 중요하다. 따라서 다음의 항목들을 중점적으로 평가한다.

- 다음 단계로 나가고자 하는 즉, 문제를 풀고자 하는 집요함
- 자신이 택한 방법이 최선인지를 잘 모르는 상태에서 갖는 스스로에 대한 믿음
- 그러한 믿음을 위한 가지기위한 각자의 해결 방법들

출제자의 의도

필자는 머신러닝에서 가장 중요한 부분이 data의 확보라고 생각한다. 하지만 현존하는 대부분의 머신러닝 교육 프로그램은 이 과정을 한 줄로 처리하여 어딘가에서 데이터를 다운로드 받도록 하고 있는데, 현실세계에서 머신러닝 업무를 진행할 때 필요한 부분은 바로 '데이터의 취득', '데이터의 전처리', '레이블링' 및 'NumPy ndarray 형태의 패키징 작업'일 것이다.

따라서 해당 미션을 통해 학생들에게 머신러닝 업무 과정의 from A To Z 전체 경험을 경험하게 한다. 또한 현실세계에서 진행되어야 하는 일련의 작업에 대한 시간, 노동력, 난이도, 괴로움의 정도, 단순한 반복작업을 또다른 프로그램, 혹은 google 등 플랫폼 서비스를 이용하여 효율화 시킬 수 있는지 등에 대한 열린 사고 등을 경험해 보도록 한다.