

12. 세이지메이커 Ground Truth

2강. Ground Truth를 이용한 데이터 세트 구축

학습목표

- Ground Truth를 사용하기 위하여 데이터 세트 준비, 레이블링 지정 작업, 작업 결과 확인 단계를 설명할 수 있다.

학습내용

- 데이터 세트 준비하기
- Ground Truth 레이블링 지정 작업
- Ground Truth 작업 결과 확인

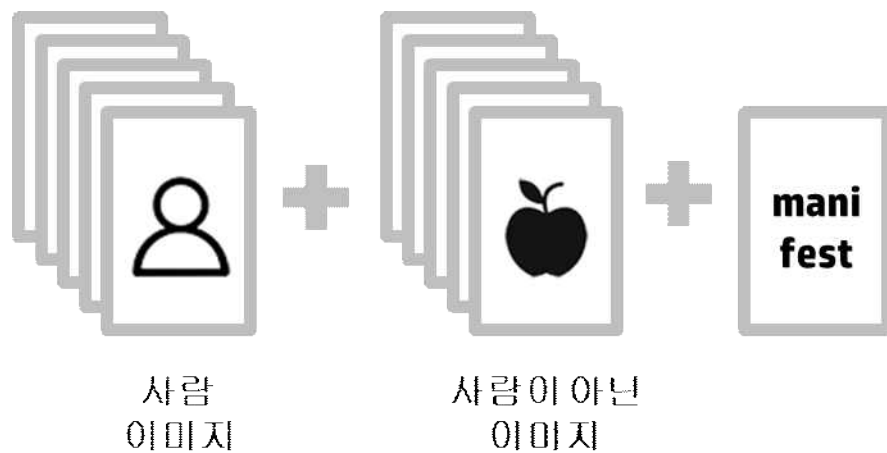
1. 데이터 세트 준비하기

- 데이터 세트 준비 단계

- ① 데이터 준비
- ② S3에 데이터 업로드

✓ 데이터 준비

- 사람과 사람이 아닌 이미지 10장을 준비



✓ S3에 데이터 업로드

- 데이터 세트 업로드를 위하여 S3 버킷 생성



- 버킷은 세이지메이커가 사용가능한 리전으로 생성



- 버킷에 로컬에 다운로드된 이미지 데이터 파일만 업로드



미국 동부(버지니아 북부)				
보기 1 대상 10				
<input type="checkbox"/> 이름 ▼	마지막 수정 ▼	크기 ▼	스토리지 클래스 ▼	
<input type="checkbox"/> 1.jpg	8월 11, 2020 2:56:42 오후 GMT+0900	12.7 KB	스탠다드	
<input type="checkbox"/> 10.jpg	8월 11, 2020 2:56:43 오후 GMT+0900	67.9 KB	스탠다드	
<input type="checkbox"/> 2.jpg	8월 11, 2020 2:56:41 오후 GMT+0900	9.7 KB	스탠다드	
<input type="checkbox"/> 3.jpg	8월 11, 2020 2:56:42 오후 GMT+0900	16.7 KB	스탠다드	

- 로컬에 manifest.txt 생성
- 파일을 열어 이미지 파일을 업로드한 s3 경로로 위치 수정 후 s3에 업로드

{ "source-ref": "s3 이미지 파일 경로 위치/파일이름.확장자" }

manifest.txt - Windows 메모장

```

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
{ "source-ref" "s3://ground-true-test-0 /1. jpg" }
{ "source-ref" "s3://ground-true-test-0 /2. jpg" }
{ "source-ref" "s3://ground-true-test-0 /3. jpg" }
{ "source-ref" "s3://ground-true-test-0 /4. jpg" }
{ "source-ref" "s3://ground-true-test-0 /5. jpg" }
{ "source-ref" "s3://ground-true-test-0 /6. jpg" }
{ "source-ref" "s3://ground-true-test-0 /7. jpg" }
{ "source-ref" "s3://ground-true-test-0 /8. jpg" }
{ "source-ref" "s3://ground-true-test-0 /9. jpg" }
{ "source-ref" "s3://ground-true-test-0 /10. jpg" }

```

- Ground truth 실습의 출력을 위한 폴더 생성

The top screenshot shows the 'Create Folder' dialog in the AWS S3 console. The 'Name' field contains 'output'. Below the field, there is explanatory text in Korean: '폴더를 생성하면 S3 콘솔에서 위 이름에 접두사 "/"가 붙은 객체를 생성하며, 해당 객체는 S3 콘솔에 폴더로 표시됩니다. 해당 객체에 대한 암호화 설정을 선택합니다.' (When you create a folder, S3 creates an object with the prefix "/" in the console, which is displayed as a folder. Select the encryption settings for this object). There are three radio button options: '없음(버킷 설정 사용)' (Selected), 'AES-256' (Amazon S3 관리 키(SSE-S3)로 서버 측 암호화 사용), and 'AWS-KMS' (AWS KMS 관리 키(SSE-KMS)로 서버 측 암호화 사용). At the bottom are '저장' (Save) and '취소' (Cancel) buttons.

The bottom screenshot shows the S3 bucket contents. At the top are buttons: '업로드' (Upload), '+ 폴더 만들기' (Create Folder), '다운로드' (Download), and '작업' (Actions). Below is a table of objects:

<input type="checkbox"/>	이름 ▼	마지막 수정 ▼	크기 ▼
<input type="checkbox"/>	output	--	--
<input type="checkbox"/>	1.jpg	8월 11, 2020 2:56:42 오후 GMT+0900	12.7 K

2. Ground Truth 레이블링 지정 작업

- Ground Truth 레이블링 지정 작업단계

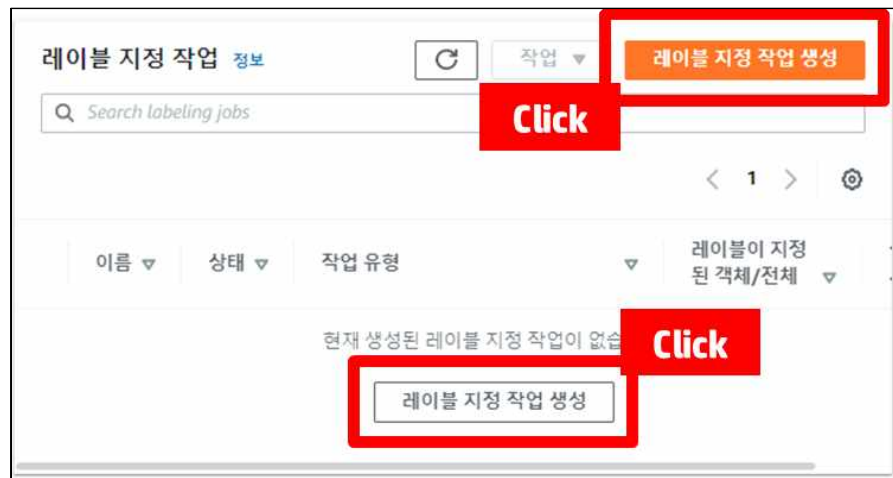
- ① 레이블링 작업 생성
- ② 작업 세부 정보 지정
- ③ 작업자와 구성 도구 선택

✓ 레이블링 작업 생성

- AWS 세이지메이커 콘솔 접속
- 대시보드에서 Ground Truth - [레이블링 작업 클릭]



- 오른쪽 상단 및 하단에 [레이블 지정 작업 생성] 클릭



✓ 작업 세부 정보 지정

- 작업 세부 정보 지정은 작업개요, 작업유형, 향상된 이미지 액세스 활성화, 태그로 구성되어 있음



- 작업개요

작업이름	레이블링 작업에 작업을 설명하는이름을 지정
입력데이터설정	입력 데이터를 식별하는 방법 설정(자동/수동)
입력데이터세트 위치	생성한 manifest 파일의 s3 위치
출력데이터세트 위치	출력 데이터가 기록되는 위치
IAM역할	세이지메이커FullAccess IAM 정책이 연결되어 있는 IAM 역할을 생성하거나 선택

- 작업이름 : ground-demo-01
- 입력 데이터 설정 : 수동
- 입력데이터위치 : s3 버킷안에 manifest.txt파일 선택
- 출력데이터위치: s3 버킷안에 output 폴더 선택
- IAM 역할 : 이전에 생성한 역할 선택

작업 이름

ground-demo-01

최대 63자의 영숫자입니다. 하이픈(-)을 포함할 수 있지만 공백은 포함할 수 없습니다. AWS 리전의 계정에서 고유해야 합니다.

☐ I want to specify a label attribute name different from the labeling job name.
Label attribute name is the key where your labels are stored in the augmented manifest. Ground Truth uses the labeling job name as the default label attribute name.

Input data setup 정보
 Ground Truth needs a way to identify your input data. Use the automated setup to have Ground Truth automatically identify your dataset in S3. Use the manual setup if you have an input manifest file.

☐ Automated data setup
 Provide the S3 location of the dataset you want labeled and let Ground Truth automatically connect to and use this dataset for your job.

☒ Manual data setup
 Provide the S3 location of a file (an input manifest file) that identifies the data objects you want labeled

입력 데이터 세트 위치 정보
 Provide a path to the S3 location where your manifest file is stored.

s3://groundtruth-demo-01/manifest.txt

출력 데이터 세트 위치 정보
 Provide a path to the S3 location where you want your labeled dataset to be stored.

s3://groundtruth-demo-01/output/

IAM 역할 정보
 Amazon SageMaker에 사용자를 대신하여 다른 서비스를 호출할 권한이 필요합니다. 역할을 선택하거나, 연결된 AmazonSageMakerFullAccess IAM 정책을 사용하여 역할을 생성합니다.

AmazonSageMaker-ExecutionRole-20200810T205407 ▼

- 작업유형



- 작업 범주 : 이미지 선택

- 작업 선택 : 이미지 분류(단일 레이블) 선택

The screenshot shows a web interface for task selection. At the top, 'Task category' is set to 'Image'. Below this, the '작업 선택' (Task Selection) section shows two options: '이미지 분류(단일 레이블)' (Image Classification (Single Label)) which is selected, and '이미지 분류(다중 레이블)' (Image Classification (Multiple Label)). The selected option shows a preview of a basketball game with labels 'Basketball' and 'Soccer'.

- 향상된 이미지 액세스 활성화 [체크 표시] 후 [다음] 선택

- 교차 원본 리소스 공유(CORS) 구성 활성화 ON/OFF 가능
- CORS는 여러 서버의 콘텐츠가 있는 웹 페이지에서 해당 콘텐츠가 사용되는 방법을 제어하는 브라우저 보안 정책
- 올바른 CORS 정책이 없는 특정 이미지 작업의 경우 작업자가 확인 또는 수정사항을 제출하려고 할 때 오류가 발생

항상된 이미지 액세스 활성화

☒ 교차 원본 리소스 공유(CORS) 구성 활성화 정보

CORS는 리소스 공유를 위한 브라우저 보안 정책입니다. S3 버킷에 CORS가 적절하게 설정되어 있지 않으면 작업자가 이미지 주석 작업을 완료하지 못할 수 있습니다.

▶ 태그 - 선택 사항

Click

취소 다음

✓ 작업자와 구성 도구 선택

- 작업자, 분류 레이블 지정도구로 구성됨
- 작업자 유형은 퍼블릭, 프라이빗, 공급업체 관리로 나뉘어져 있음

작업자와 구성 도구 선택

작업자 정보

작업자 유형

☒ 퍼블릭
Amazon Mechanical Turk에서 지원하는 전 세계 50만 이상의 독립적 계약업체의 24x7 온디맨드 작업 인력입니다.

☐ 프라이빗
조직 내에서 보관해야 하는 데이터를 처리하기 위해 자체 직원 또는 계약업체를 포함하여 작업 조달한 작업자 팀입니다.

☐ 공급업체 관리
AWS Marketplace를 통해 제공되는 데이터 레이블 지정 서비스 제공을 전문으로 하는 타사 공급업체의 큐레이트 목록입니다.

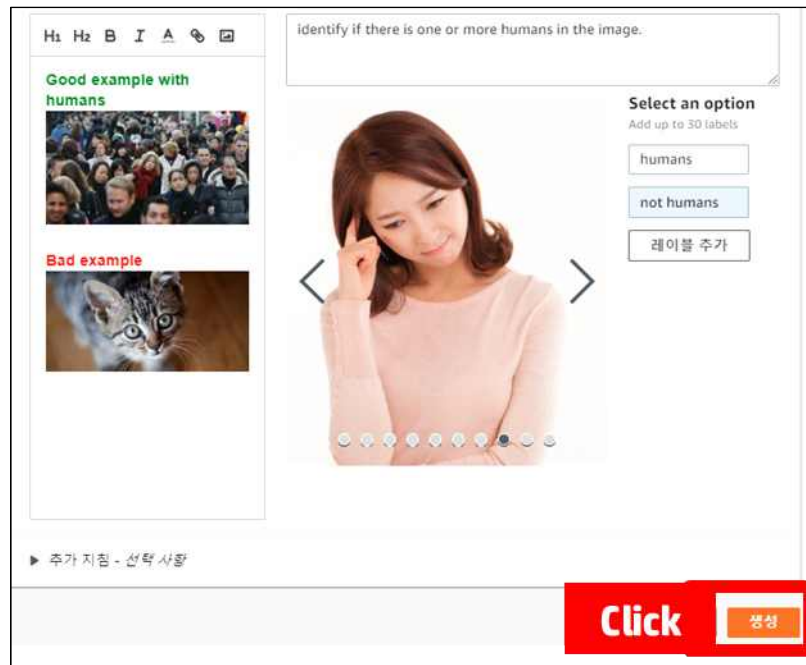
- 작업자 유형 [퍼블릭] 선택하기

퍼블릭	<ul style="list-style-type: none"> • 전 세계 500,000명을 상회하는 독립 계약자의 인력에 온디맨드 24/7의 방식으로 액세스할 수 있음 • 민감하지 않은 데이터에 권장
프라이빗	<ul style="list-style-type: none"> • 자체 직원 또는 계약자로 구성된 팀에 액세스할 수 있음 • 민감한 데이터인 경우 또는 레이블 지정 작업에 도메인 전문성이 필요한 경우에 권장
공급업체관리	<ul style="list-style-type: none"> • Amazon이 승인하고 AWS Marketplace를 통해 전문적인 데이터 레이블 지정 서비스를 제공하는 제 3자 공급업체 목록에 액세스할 수 있음

- 작업자 : 퍼블릭
- 태스크 제한 시간 : 5분
- 작업당 가격 : 0.012\$ 선택
- 데이터 세트에 성인 콘텐츠를 포함하지 않는 항목과 퍼블릭 작업 인력에 게 데이터 세트가 표시된다는 내용을 읽고 수락(체크)
- 구성 도구 선택



- 구성 도구 선택에서 작업의 지침과 레이블을 설정
 - ① Good example, Bad example
 - ② 레이블 지정(최대 30개)
 - ③ 작업에 대한 설명
- 작업의 지침과 레이블을 설정한 후 오른쪽 하단의 생성 클릭



3. Ground Truth 작업 결과 확인

- 레이블링 지정 작업 결과 확인 단계
 - ✓ 레이블링 지정 작업이 시작 되면 이미지 분류 작업이 시작됨



- ✓ 레이블링 지정 작업이 완료될 경우 레이블 지정 작업 이름을 선택하여 레이블 지정 작업 요약 및 결과 확인 가능



✓ 레이블링 지정 작업 요약

ground-demo-01

레이블 지정 작업 요약

레이블 지정 도구 보기

상태 완료	입력 데이터 세트 위치 s3://groundtruth-demo-01/manifest.txt	ARN arn:aws:sagemaker:us-east-1:402345010737:labeling-job/ground-demo-01
작업 유형 이미지 분류(단일 레이블)	레이블 지정됨/전체 데이터 세트 객체 10 / 10	작업자 유형 퍼플릭
생성 시간 2020년 8월 11일 오전 9:01 UTC	출력 데이터 세트 위치 s3://groundtruth-demo-01/output/ground-demo-01/	작업팀 ARN arn:aws:sagemaker:us-east-1:394669845002:workteam/public-crowd/default

✓ 레이블링 출력 결과

- 레이블 지정된 데이터 세트 객체 섹션에서 데이터 세트의 이미지 썸네일이 아래에 해당 레이블이 부착된 상태로 표시

5.jpg
레이블: not humans

10.jpg
레이블: humans

✓ 출력결과 검토

- 레이블 지정 작업 요약 세션에서 출력 데이터 세트 위치 클릭



- s3://버킷이름/output/레이블링 지정 작업 이름 폴더 클릭



✓ manifests 디렉터리로 이동하여 output.manifest 다운로드 후 열기(text 확장자 사용)



- ✓ manifests파일 안에서 사람으로 레이블링 된 데이터

```
{
  "source-ref": "s3://groundtruth-demo-01/1.jpg",
  "ground-demo-01": 0,
  "ground-demo-01-metadata": {
    "confidence": 0.95,
    "job-name": "labeling-job/ground-demo-01",
    "class-name": "humans",
    "human-annotated": "yes",
    "creation-date": "2020-08-11T09:08:13.312152",
    "type": "groundtruth/image-classification"
  }
}
```

- ✓ manifests파일 안에서 사람이 아닌 것으로 레이블링 된 데이터

```
{
  "source-ref": "s3://groundtruth-demo-01/10.jpg",
  "ground-demo-01": 1,
  "ground-demo-01-metadata": {
    "confidence": 0.95,
    "job-name": "labeling-job/ground-demo-01",
    "class-name": "not humans",
    "human-annotated": "yes",
    "creation-date": "2020-08-11T09:06:27.127950",
    "type": "groundtruth/image-classification"
  }
}
```

평가하기

1. Ground Truth 레이블링 지정 작업 단계에서 작업 세부 정보로 입력해야 할 정보가 아닌 것을 고르시오.

- ① 작업이름
- ② 입력 데이터 설정
- ③ 작업자 유형
- ④ IAM 역할

- 정답 : ③ 번

해설 : 작업자 세부 정보로 입력해야 할 정보로는 작업 이름, 입력 데이터 설정, 입·출력 데이터 위치, IAM 역할, 작업 유형이 있습니다. 작업자 유형은 작업자와 구성도구 선택에서 입력해야 합니다.

2. 작업자와 구성 도구 선택에서 “퍼블릭”작업자 유형은 자체 직원 또는 계약자로 구성된 팀에 액세스 할 수 있다. (O/X)

- 정답 : X

해설 : 자체 직원 또는 계약자로 구성된 팀에 액세스할 수 있는 작업자 유형은 “프라이빗” 입니다.

학습정리

1. 데이터 세트 준비하기

- 데이터 준비
- S3에 데이터 세트 업로드

2. Ground Truth 레이블링 지정 작업

- 레이블링 작업 생성
- 작업 세부 정보 지정
- 작업자와 구성 도구 선택

3. Ground Truth 작업 결과 확인

- 레이블링 지정 작업 요약
- 레이블링 출력 결과
- 출력 결과 검토