

12. 세이지메이커 Ground Truth

3강. Ground Truth를 이용한 데이터 세트 구축 실습

학습목표

- Ground Truth 사용하여 데이터 세트 준비, 레이블링 지정 작업, 작업 결과 확인을 할 수 있다.

학습내용

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

1. 데이터 세트 준비하기

- 데이터 세트 업로드
 - ✓ AWS S3 콘솔 접속하여 버킷 생성
 - ✓ 버킷은 세이지메이커가 사용가능한 리전으로 생성

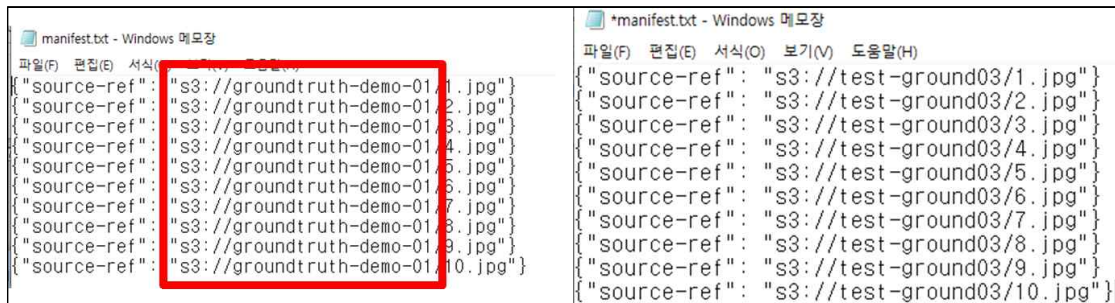
버킷 (2)

버킷은 S3에 저장되는 데이터의 컨테이너입니다. [자세히 알아보기](#)

🔍 이름으로 버킷 찾기

	이름 ▼	리전 ▼	액세스 ▼	생성 날짜 ▼
<input type="radio"/>	test-bdu-01	오류		2020. 8. 11. pm 3:06:11 PM KST
<input checked="" type="radio"/>	test-ground03	미국 동부(버지니아 북부) us-east-1	버킷 및 객체가 퍼블릭이 아님	2020. 8. 11. pm 8:28:24 PM KST

- ✓ 다운로드한 데이터 세트 중 manifest.txt 파일 열기
- ✓ S3 경로(버킷이름) 수정
- ✓ 생성한 버킷 이름 : test-ground03



- ✓ 생성한 버킷에 로컬에 다운로드 된 이미지 및 수정된 manifest.txt 파일 업로드

미국 동부(버지니아 북부)			
보기 1 대상 11			
<input type="checkbox"/> 이름 ▼	마지막 수정 ▼	크기 ▼	스토리지 클래스 ▼
<input type="checkbox"/> 1.jpg	8월 11, 2020 8:41:47 오후 GMT+0900	12.7 KB	스탠다드
<input type="checkbox"/> 10.jpg	8월 11, 2020 8:41:49 오후 GMT+0900	67.9 KB	스탠다드
<input type="checkbox"/> 2.jpg	8월 11, 2020 8:41:47 오후 GMT+0900	9.7 KB	스탠다드

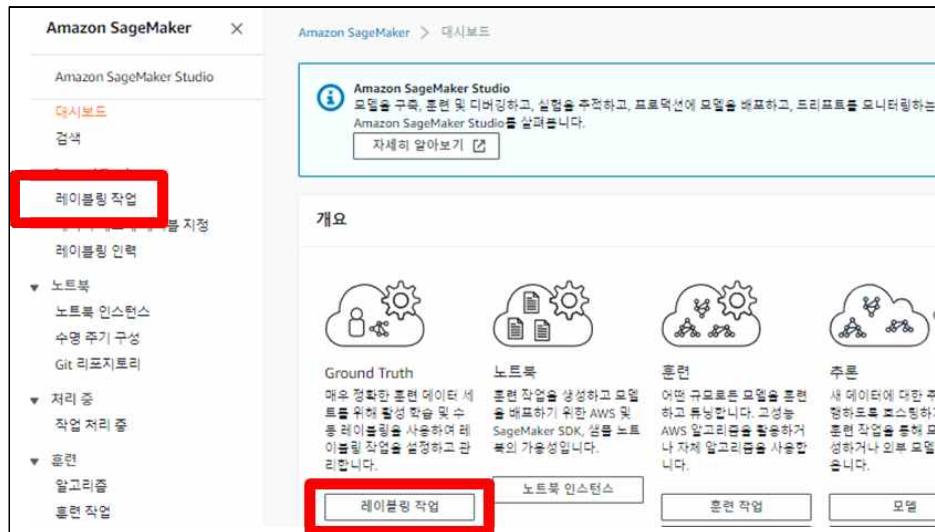
- ✓ Ground truth 출력 결과물을 구분하기 위하여 출력 전용 폴더생성

미국 동부(버지니아 북부)			
보기 1 대상 12			
<input type="checkbox"/> 이름 ▼	마지막 수정 ▼	크기 ▼	스토리지 클래스 ▼
<input type="checkbox"/> output	--	--	--
<input type="checkbox"/> 1.jpg	8월 11, 2020 8:41:47 오후 GMT+0900	12.7 KB	스탠다드
<input type="checkbox"/> 10.jpg	8월 11, 2020 8:41:49 오후 GMT+0900	67.9 KB	스탠다드

2. Ground Truth 레이블링 지정 작업

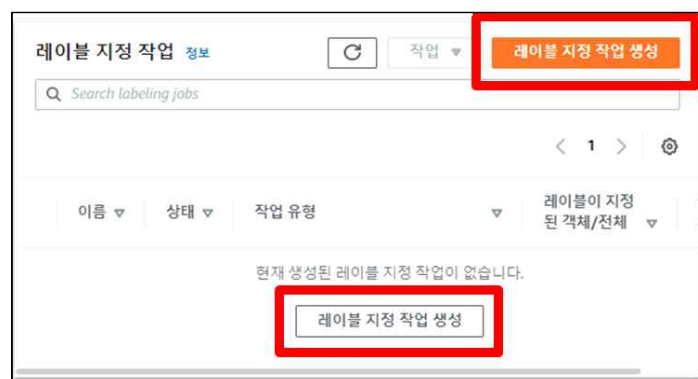
- 레이블링 작업 생성

✓ 세이지메이커 콘솔에 접속하여 왼쪽 대시보드에서 [레이블링 작업] 클릭



✓ 오른쪽 상단 및 하단에 [레이블 지정 작업 생성] 클릭

- 1번 이상 레이블 지정 작업을 수행할 경우 아래 레이블 지정 작업 생성 대신 수행한 목록이 뜬다



- 작업 세부 정보 지정
 - ✓ 원하는 작업명 입력

작업 개요

작업 이름

ground-demo-02

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.

- ✓ 입력 데이터 설정을 오른쪽 수동(Manual data set)으로 체크

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 선택
- ✓ s3에 업로드한 manifest.txt 선택

입력 데이터 세트 위치 정보

Provide a path to the S3 location where your manifest file is stored.

Browse S3

Select a manifest file

S3 buckets

Buckets (2)

Name	Creation date
test-bdu-01	Tue Aug 11 2020 15:06:11 GMT+0900 (대한민국 표준시)
test-ground02	Tue Aug 11 2020 20:28:24 GMT+0900 (대한민국 표준시)

Select a manifest file

S3 buckets > test-ground02

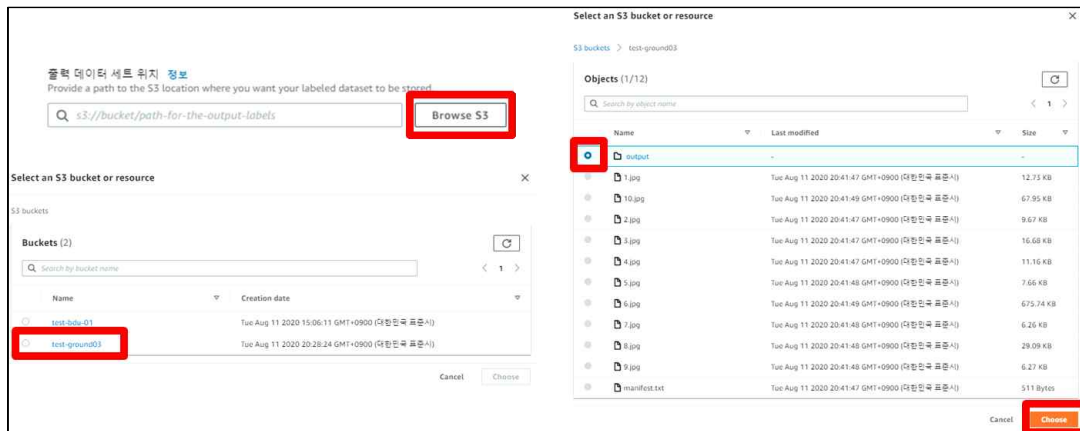
Objects (1/12)

Name	Last modified	Size
output	-	-
1.jpg	Tue Aug 11 2020 20:41:47 GMT+0900 (대한민국 표준시)	12.73 KB
10.jpg	Tue Aug 11 2020 20:41:49 GMT+0900 (대한민국 표준시)	67.95 KB
2.jpg	Tue Aug 11 2020 20:41:47 GMT+0900 (대한민국 표준시)	9.67 KB
3.jpg	Tue Aug 11 2020 20:41:47 GMT+0900 (대한민국 표준시)	16.68 KB
4.jpg	Tue Aug 11 2020 20:41:47 GMT+0900 (대한민국 표준시)	11.10 KB
5.jpg	Tue Aug 11 2020 20:41:48 GMT+0900 (대한민국 표준시)	7.66 KB
6.jpg	Tue Aug 11 2020 20:41:48 GMT+0900 (대한민국 표준시)	6.26 KB
7.jpg	Tue Aug 11 2020 20:41:48 GMT+0900 (대한민국 표준시)	29.09 KB
8.jpg	Tue Aug 11 2020 20:41:48 GMT+0900 (대한민국 표준시)	6.27 KB
manifest.txt	Tue Aug 11 2020 20:41:47 GMT+0900 (대한민국 표준시)	317 Bytes

manifest.txt

Cancel **Choose**

- ✓ 출력 데이터 세트 위치 선택
- ✓ 이전 버킷에 생성한 output 폴더 선택



✓ 최종 입력, 출력 데이터 세트 위치

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

Q s3://test-ground03/manifest.txt X Browse S3

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

Q s3://test-ground03/output/ X Browse S3

✓ IAM 역할은 기존에 생성된 역할 선택

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

옵션 선택 ▲

Create a new role

Enter a custom IAM role ARN

Use existing role

AmazonSageMaker-ExecutionRole-20200810T205407

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

AmazonSageMaker-ExecutionRole-20200810T205407 ▼

- ✓ 작업 유형에서 카테고리를 이미지 선택
- ✓ 작업 선택에서 이미지 분류(단일 이미지) 선택

작업 유형 정보

Task category
Select the type of data being labeled to view available task templates for it or select 'Custom' to create a custom task.

Image

Image

Text

Video

Point cloud

사용자 지정

작업 선택
작업자가 데이터 세트의 각체에 레이블을 지정하기 위해 수행할 작업을 선택합니다.

☒ 이미지 분류(단일 레이블)
작업자가 이미지를 개별 클래스로 분류하도록 합니다. [정보](#)

☐ 이미지 분류(다중 레이블)
작업자가 이미지를 하나 이상의 클래스로 분류하도록 합니다. [정보](#)

☒ Basketball

☐ Soccer

☒ Human

☒ Vehicle

☐ Animal

- ✓ 향상된 이미지 액세스 활성화 체크 후 다음 클릭

향상된 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

취소 **다음**

- 작업자와 구성 도구 선택
 - ✓ 작업 유형 [퍼블릭] 선택

작업자 정보

작업자 유형

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

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

☐ 공급업체 관리
AWS Marketplace를 통해 제공되는 데이터 레이블 지정 서비스 제 3자 전문으로 하는 타사 공급업체의 쿼리스트를 사용합니다.

- ✓ 테스크 제한시간/ 작업당 가격은 기본값을 사용

태스크 제한 시간
작업자가 단일 태스크에서 작업할 수 있는 최대 시간

0 시간 5 분

작업당 가격
지불하는 가격은 작업자가 작업을 완료하는 데 걸리는 시간에 따라 결정됩니다.

\$0.012
완료하는 데 5초 미만 걸림 ▼

- ✓ 아래 두 항목 체크

☒ 이 데이터 세트는 성인 콘텐츠를 포함하지 않습니다. [정보](#)

☒ 귀하는 Amazon Mechanical Turk 인력이 세계 각지의 독립 계약업체로 구성되며, 이들 인력과 기밀 정보, 개인 정보 또는 보호 대상 의료 정보를 공유해서는 안 된다는 점을 이해하고 이에 동의합니다. [정보](#)

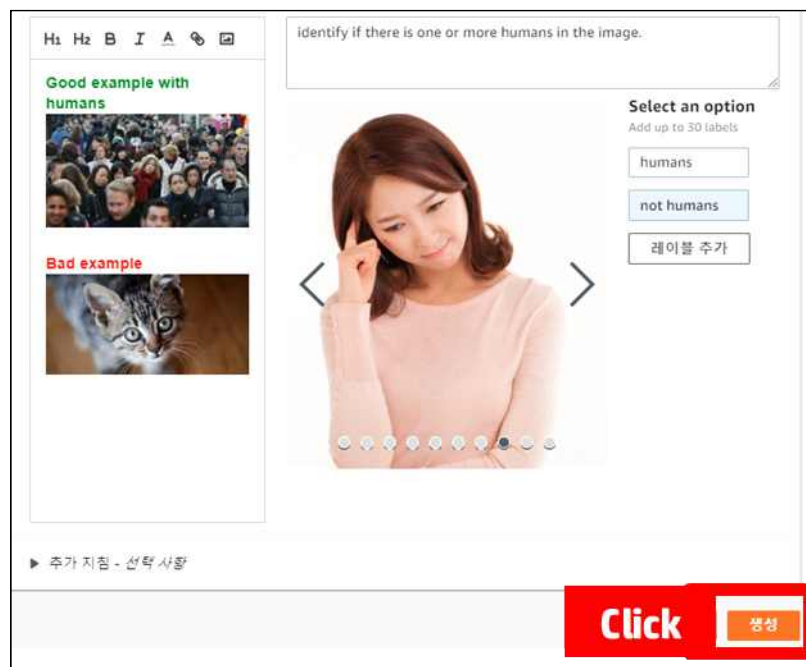
☐ 데이터 레이블 지정 자동화 [정보](#)
Amazon SageMaker will automatically label a portion of your dataset. It will train a model in your AWS account using Built-in Algorithm and your dataset. When you enable this, training jobs use new computing resources on your behalf. For cost information, See SageMaker [pricing](#) [🔗](#)

▶ 추가 구성 - 선택 사항
데이터 세트 각체당 작업자

✓ 구성 도구 선택에서 작업 지침과 레이블 설정



✓ 구성 도구 선택에서 작업 지침과 레이블 설정 후 생성 클릭



3. Ground Truth 작업 결과 확인

- 작업자와 구성 도구 선택
 - ✓ 생성 후 레이블링 작업이 진행 중

Amazon SageMaker > 레이블 지정 작업

레이블 지정 작업 정보

Search labeling jobs

	이름	상태	작업 유형
<input type="radio"/>	ground-demo-02	완료	이미지 분류(단일 레이블)
<input type="radio"/>	ground-demo-01	완료	이미지 분류(단일 레이블)

✓ 레이블 지정 작업 요약

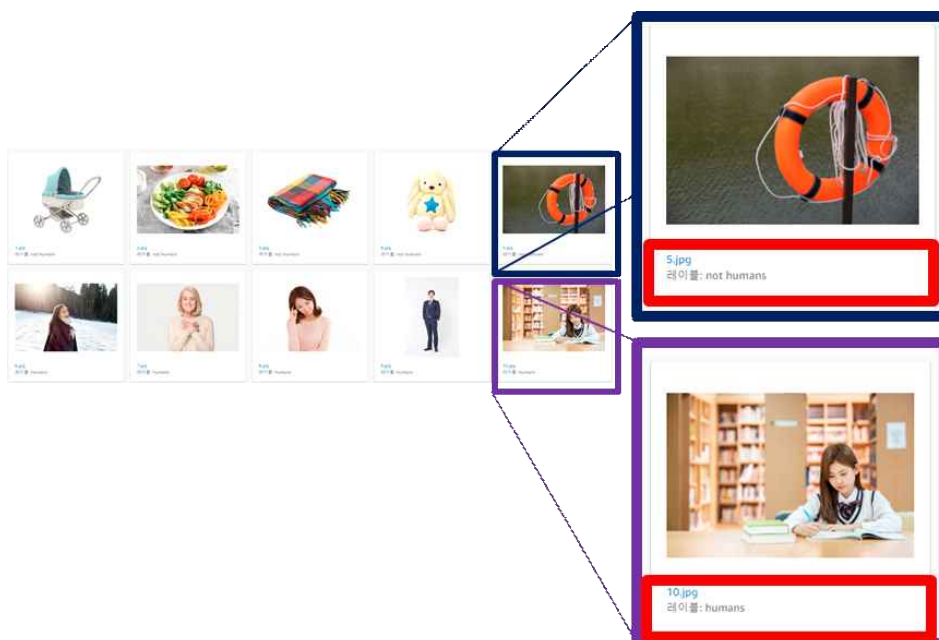
ground-demo-02

레이블 지정 작업 요약

레이블 지정 도구 보기

상태 완료	입력 데이터 세트 위치 s3://test-ground03/manifest.txt	ARN arn:aws:sagemaker:us-east-1:402345010737:labeling-job/ground-demo-02
작업 유형 이미지 분류(단일 레이블)	레이블 지정됨/전체 데이터 세트 객체 10 / 10	작업자 유형 퍼블릭
생성 시간 2020년 8월 11일 오후 12:53 UTC	출력 데이터 세트 위치 s3://test-ground03/output/ground-demo-02/	작업팀 ARN arn:aws:sagemaker:us-east-1:394669845002:workteam/public-crowd/default

✓ 출력 결과



✓ 출력 결과 검토

- 레이블 지정 작업 요약 세션에서 출력 데이터 세트 위치 클릭

ground-demo-02 작업 ▼

레이블 지정 작업 요약 레이블 지정 도구 보기

상태 완료	입력 데이터 세트 위치 s3://test-ground03/manifest.txt	ARN arn:aws:sagemaker:us-east-1:402345010737:labeling-job/ground-demo-02
작업 유형 이미지 분류(단일 레이블)	레이블 지정됨/전체 데이터 세트 객체 10 / 10	작업자 유형 퍼블릭
생성 시간 2020년 8월 11일 오후 12:53 UTC	출력 데이터 세트 위치 s3://test-ground03/output/ground-demo-02/	작업팀 ARN arn:aws:sagemaker:us-east-1:394669845002:workteam/public-crowd/default

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

Amazon S3 > test-ground03 > output > ground-demo-02

test-ground03

개요

Q 검색하려면 접두사를 입력하고 Enter 키를 누릅니다. 지우려면 Esc 키를 누릅니다.

업로드
+ 폴더 만들기
다운로드
작업 ▼
미국 동부(버지니아 북부) ↻

<input type="checkbox"/> 이름 ▼	마지막 수정 ▼	크기 ▼	보기 1 대상 4 스토리지 클래스 ▼
<input type="checkbox"/> annotation-tool	--	--	--
<input type="checkbox"/> annotations	--	--	--
<input type="checkbox"/> intermediate	--	--	--
<input type="checkbox"/> manifests	--	--	--

보기 1 대상 4

- ✓ manifests 디렉터리로 이동하여 output. manifest 다운로드 후 열기
(text 확장자 사용)
- manifests파일 안에서 사람으로 레이블링 된 데이터

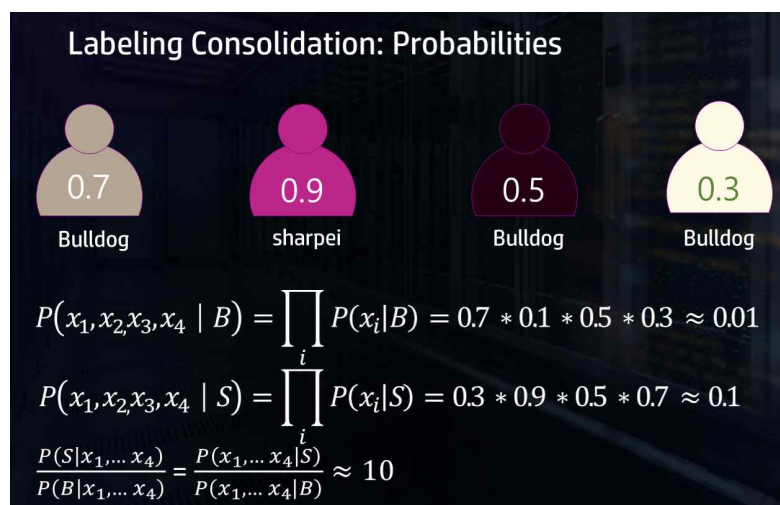
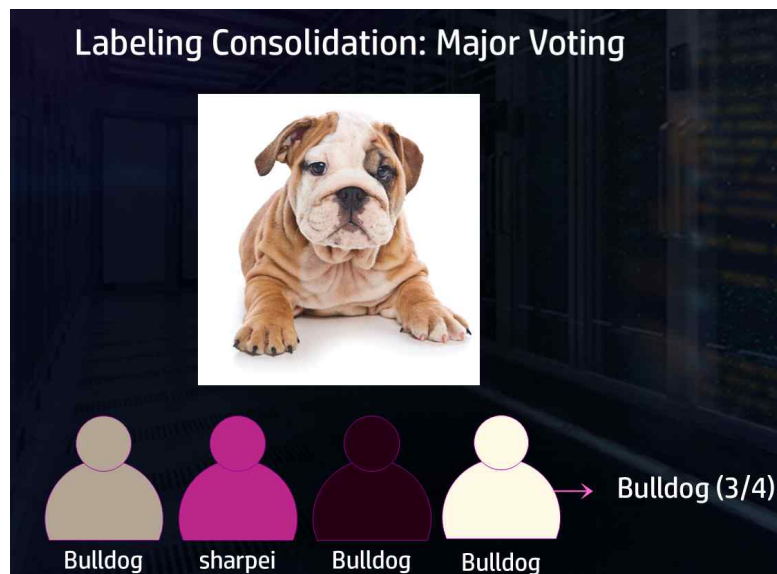
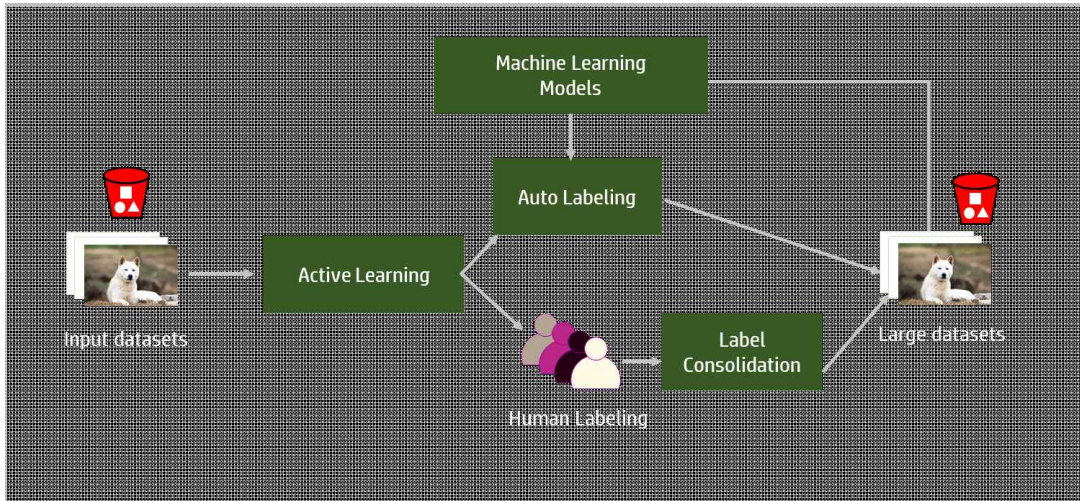
```
{
  "source-ref": "s3://groundtruth-demo-01/1.jpg",
  "ground-demo-01": 0,
  "ground-demo-01-metadata": {
    "confidence": 0.95,
    "job-name": "labeling-job/ground-demo-01",
    "class-name": "humans",
    "human-annotated": "yes",
    "creation-date": "2020-08-11T09:08:13.312152",
    "type":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

※ 나도 전문가다

- 레이블링 퀄리티를 위한 Labeling consolidation



평가하기

1. Ground Truth 레이블링 작업 종료 후 레이블 지정 작업 요약에서 확인할 수 없는 정보를 고르시오.

- ① 작업 유형
- ② 생성 시간
- ③ 작업자 유형
- ④ manifest 파일

- 정답 : ④ 번

해설 : Ground Truth 레이블 지정 작업 요약에서는 manifest의 위치(경로)만 확인할 수 있으면 실제 manifest는 저장된 S3에서 확인 가능합니다.

2. Ground Truth 레이블링 지정 작업에서 작업유형과 작업 선택의 관계가 옳지 않은 것을 고르시오.

- ① 이미지 - 경계 상자
- ② 텍스트 - 의미체계 세분화
- ③ 비디오 - 동영상 클립 분류
- ④ 포인트 클라우드 - 객체 추적

- 정답 : ② 번

해설 : 텍스트 작업 유형은 텍스트 데이터에 대한 레이블링 워크플로를 제공하며 텍스트 분류, 명명된 엔터티 인식이 있습니다.

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