

Ground Truth

Ground Truth

Annotation is the process of labeling data that has been previously unlabeled.

An example of this, maybe reading a text document from 1-5, or classifying the sentiment of text as positive or negative.

Ground Truth provides the tools necessary to manage annotation for a wide range of activities.

Labeled and unlabeled data

- Labeled data is what you ordinarily see with a dataset, where you have a number of data points and with each one a target value denoting the outcome of the data point.
- Unlabeled data is similar to labeled data but without the target value.

Labeled vs. Unlabeled Data

Labeled

Data	Label
"This is a good day"	positive
"This is a bad day"	negative

"This is a good day","positive"

"This is a bad day","negative"

Unlabeled

"This text is missing a label!"

"Is this text positive or negative?"

JSON Lines (JSONL)

- Contains JSON objects on each line
- Each line separated by a newline

Example

```
{"source": "Super great, 5 stars!"}  
{"source": "1 stars. it was bad!"}  
{"source": "5 stars! Awesome!!!"}  

```

Annotation Creation

Creating an annotation job in Ground Truth starts by uploading your data file, in `jsonl`, to S3.

When creating an annotation job, the key things to focus on are:

- Job name
- Manual data setup with your dataset as the input selection and the same bucket for output
- Select the IAM role you created when creating your Sagemaker Studio instance
- Task type is `Text`, with `Text Classification (Single Label)` as the sub option
- Choose `Private` workforce so only you annotate the data
- Create option labels and necessary descriptions for annotation

Inviting Workers

Once your job is created, you'll need to invite workers to annotate it. Create a workforce but creating a team, adding workers to the team, and then adding the team to the annotation job.

Annotation Job

After receiving an email about the annotation job, you can open the invite in a browser and begin the annotation.

Once the annotation is done, complete the annotation job in order to have the data exported to S3.

Additional Resources

- If you want to learn more about Ground Truth, visit [AWS's main page](#) for the service.
- Additional information and documentation on Ground Truth can be found [here](#).