

CPE 393: Machine Learning Operations

Data Labeling Project using Label Studio (<https://labelstud.io/>)

Label Studio Installation - <https://labelstud.io/guide/install>

Installation

Use virtual environment OR default python environment

Install using pip

To install Label Studio with pip and a virtual environment, you need Python version 3.8 or later. Run the following:

```
python3 -m venv env
source env/bin/activate
python -m pip install label-studio
```

To install Label Studio with pip, you need Python version 3.8 or later. Run the following:

```
pip install label-studio
```

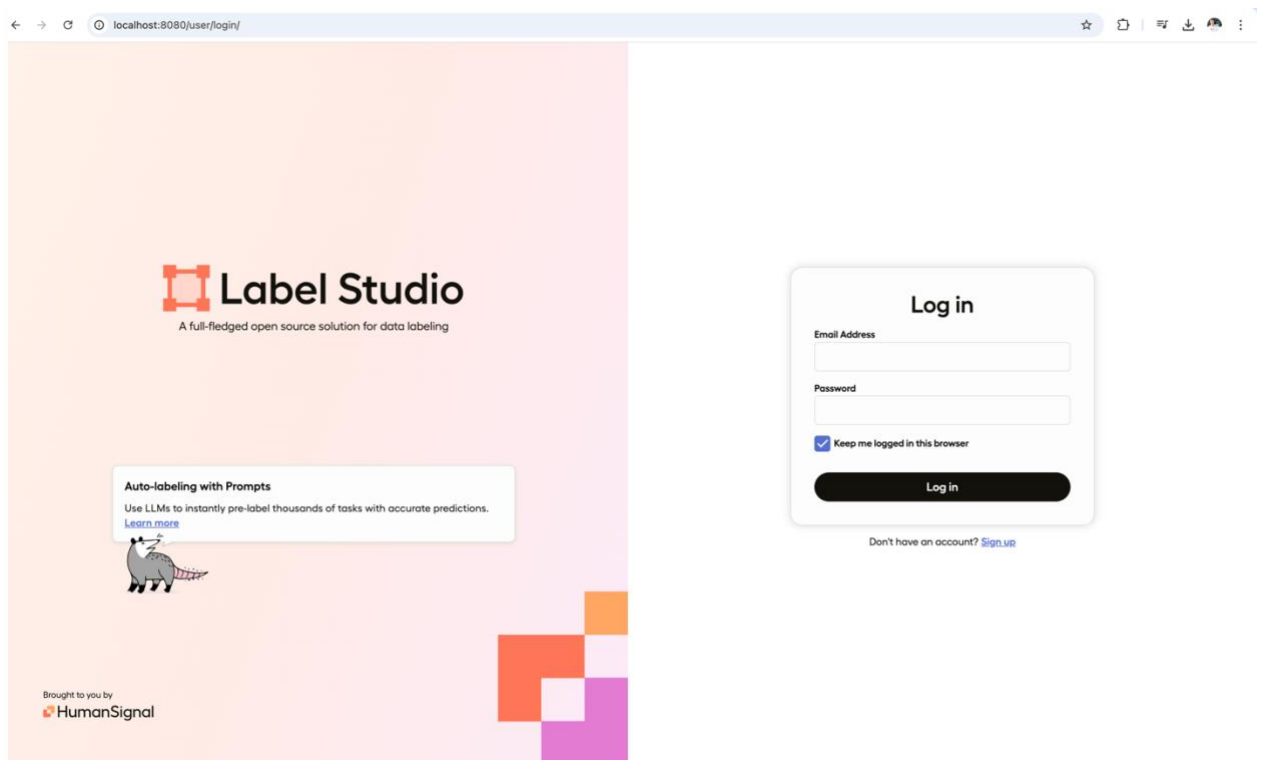
After you install Label Studio, start the server with the following command:

```
label-studio
```

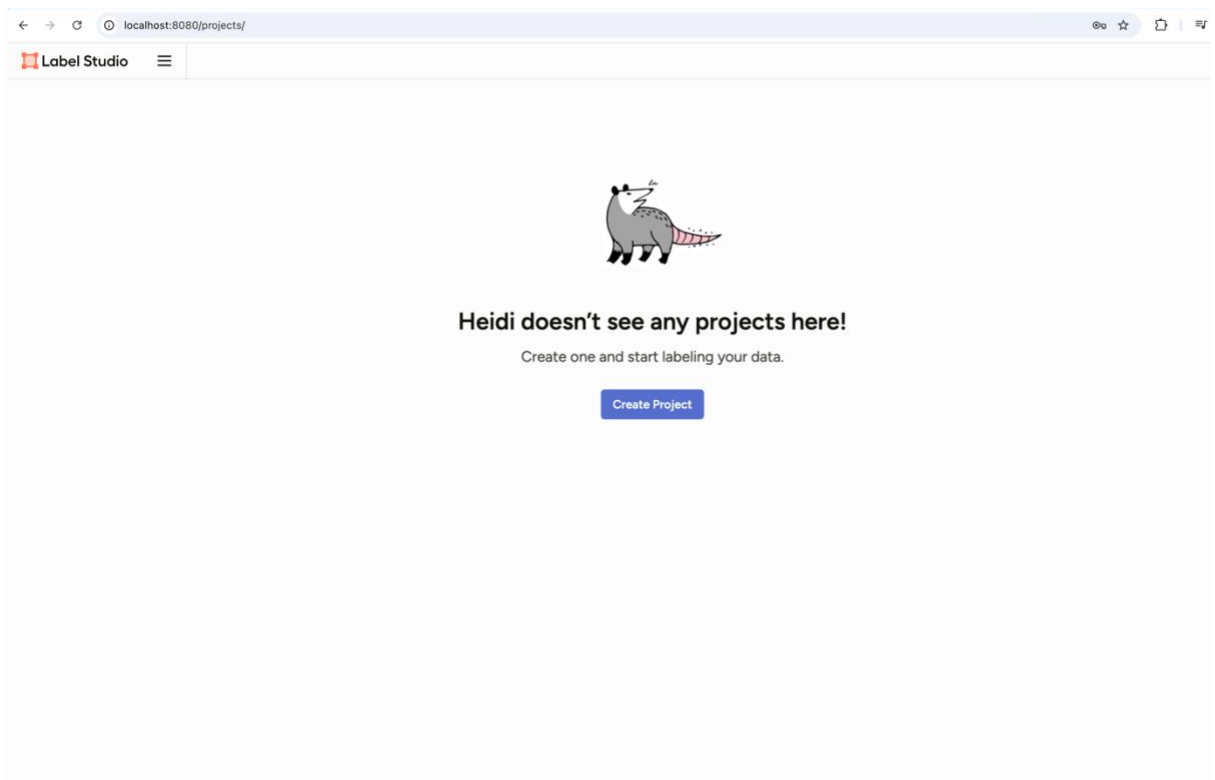
Lab 1: Creating a simple data annotation project. (5 points)

Submission: Screenshot of your project home page and your labeled data as CSV file. Make a zip folder.

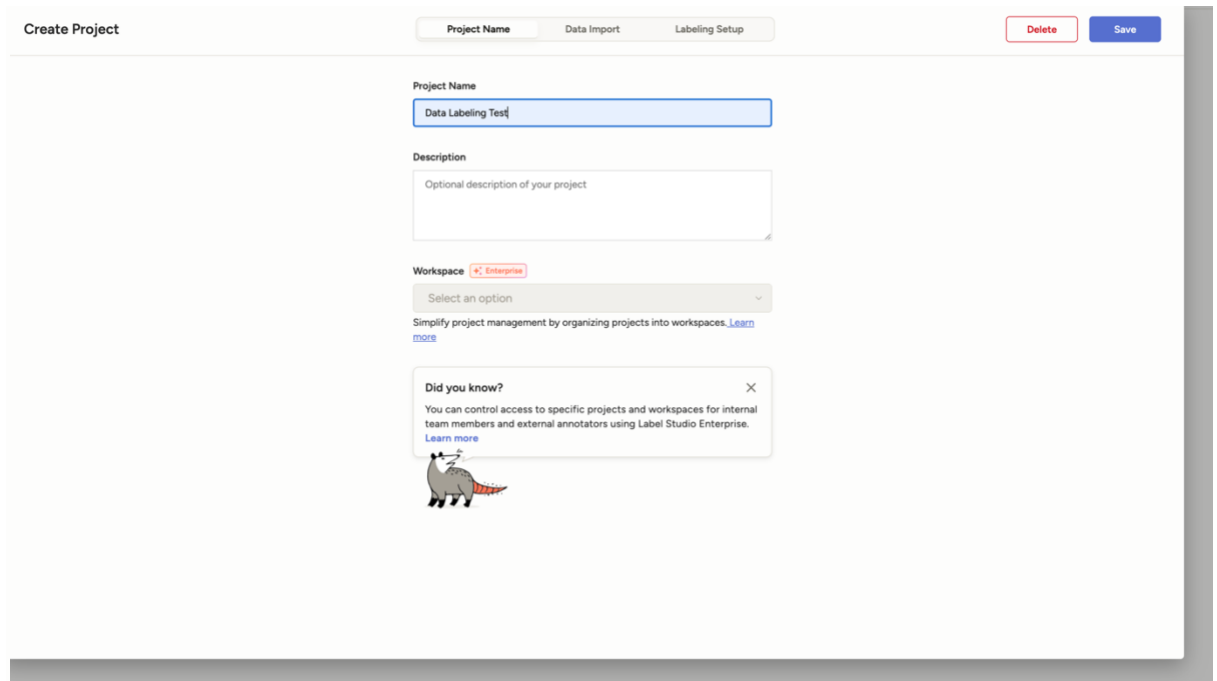
1. Open Label Studio at **http://localhost:8080**.



2. Sign up with an email address and password that you create.
3. Click **Create** to create a project and start labeling data.



4. Name the project and optionally enter a description and select a color. (Project Name – Data Labeling Test)



Create Project

Project Name Data Import Labeling Setup

Project Name

Data Labeling Test

Description

Optional description of your project

Workspace Enterprise

Select an option

Simplify project management by organizing projects into workspaces. [Learn more](#)

Did you know?

You can control access to specific projects and workspaces for internal team members and external annotators using Label Studio Enterprise. [Learn more](#)

Save

5. Click **Data Import** and upload the data files that you want to use (Coronoa-TweetNLP.csv) .

6. Click **Labeling Setup** and choose a template (Natural Language Processing→ Text Classification) and customize the label names for your use case. Set “OriginalTweet” in text field. You can change the label names according to your label guidelines. For this project, we will use

Configure data

Use text from field <set manually> \$OriginalTweet

Add choices

Use new line as a separator to add multiple labels

Add

Choices (3)

Positive

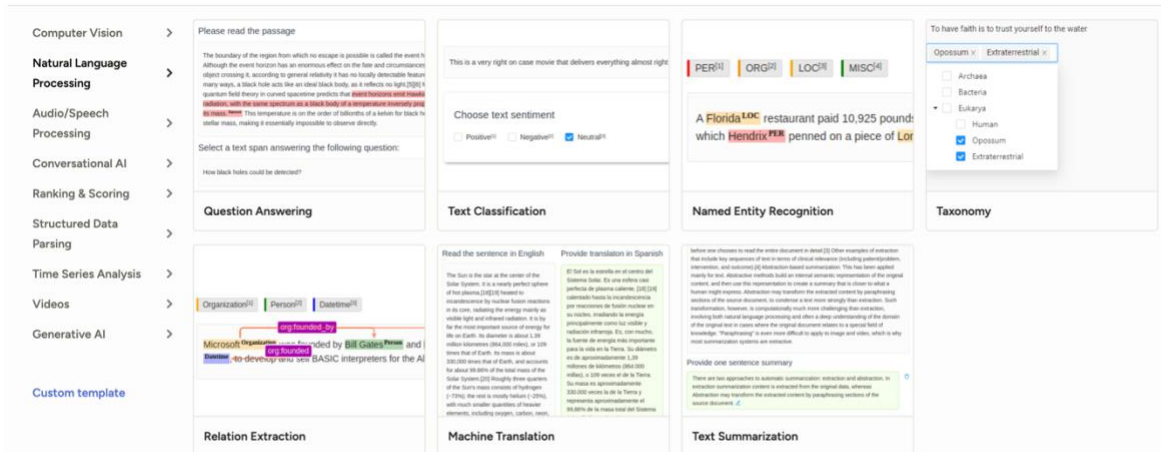
Negative

Neutral

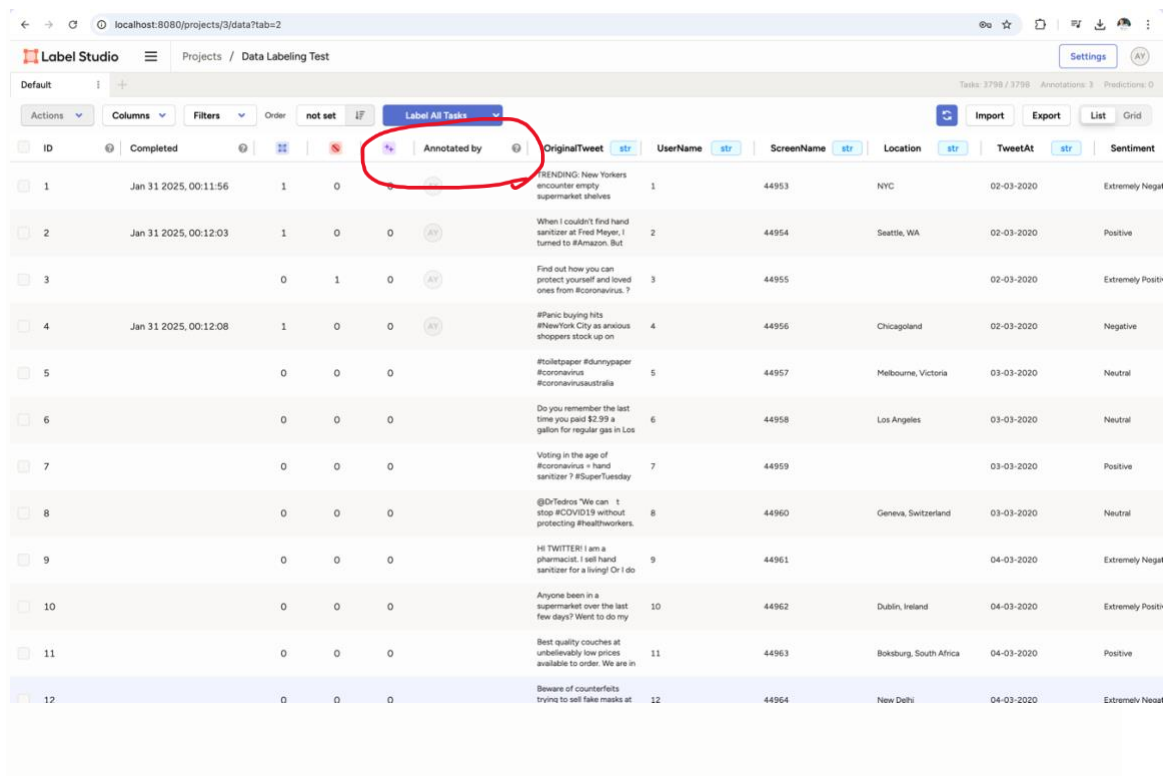
×

×

×



7. Click **Save** to save your project. Then, you can start label your data.
8. Click on the bright blue `Label All Tasks` button at the center of the toolbar to enter the labeling interface. You'll then be presented with a screen with the label interface you defined during setup. You'll notice that the text is drawn from the data set we imported. There will be a directive to choose from the sentiment options. To the right of the screen, there are additional buttons to "Skip" or "Submit" an annotation.
9. Using your judgment, you can select the sentiment of the text and choose "Submit" to move on to the next labeling task. Within this particular interface, you can speed up your annotations with the keyboard shortcuts listed next to the annotation options, in this case '1' for 'Positive' and '2' for 'Negative.' If you're unsure of the sentiment to apply to the text, just select "Skip" to move to the next task.

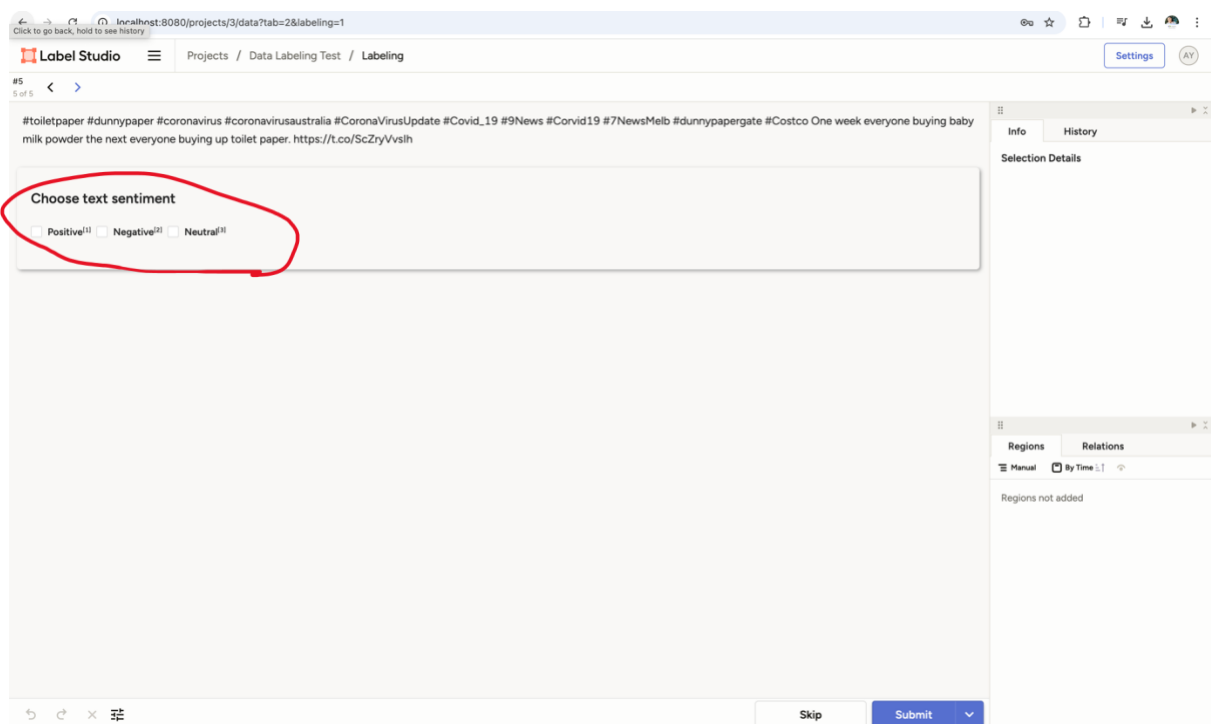


Label Studio Projects / Data Labeling Test

Tasks: 3798 / 3798 Annotations: 3 Predictions: 0

Actions Columns Filters Order not set IF Label All Tasks Import Export List Grid

ID	Completed	Order	IF	Label	Annotated by	OriginalTweet	str	UserName	str	ScreenName	str	Location	str	TweetAt	str	Sentiment
1	Jan 31 2025, 00:11:56	1	0	0	AY	TRENDING: New Yorkers encounter empty supermarket shelves	1	44953		NYC		02-03-2020				Extremely Negat
2	Jan 31 2025, 00:12:03	1	0	0	AY	When I couldn't find hand sanitizer at Fred Meyer, I turned to #Amazon. But	2	44954		Seattle, WA		02-03-2020				Positive
3		0	1	0	AY	Find out how you can protect yourself and loved ones from #coronavirus. ?	3	44955				02-03-2020				Extremely Positi
4	Jan 31 2025, 00:12:08	1	0	0	AY	#panic buying hits #NewYork City as anxious shoppers stock up on	4	44956		Chicagoland		02-03-2020				Negative
5		0	0	0		#toiletpaper #dunnypaper #coronavirus #coronavirusaustralia	5	44957		Melbourne, Victoria		03-03-2020				Neutral
6		0	0	0		Do you remember the last time you paid \$2.99 a gallon for regular gas in Los	6	44958		Los Angeles		03-03-2020				Neutral
7		0	0	0		Voting in the age of #coronavirus = hand sanitizer ? #SuperTuesday	7	44959				03-03-2020				Positive
8		0	0	0		@DrTedros "We can't stop #COVID19 without protecting #healthworkers."	8	44960		Geneva, Switzerland		03-03-2020				Neutral
9		0	0	0		Hi TWITTER! I am a pharmacist. I sell hand sanitizer for a living! Oh I do	9	44961				04-03-2020				Extremely Negat
10		0	0	0		Anyone been in a supermarket over the last few days? Went to do my	10	44962		Dublin, Ireland		04-03-2020				Extremely Positi
11		0	0	0		Best quality couches at unbelievably low prices available to order. We are in	11	44963		Boksburg, South Africa		04-03-2020				Positive
12		0	0	0		Beware of counterfeits trying to sell fake masks at	12	44964		New Delhi		04-03-2020				Extremely Neut



Label Studio Projects / Data Labeling Test / Labeling

Click to go back, hold to see history

#5 5 of 5

#toiletpaper #dunnypaper #coronavirus #coronavirusaustralia #CoronaVirusUpdate #Covid_19 #9News #Corvid19 #7NewsMelb #dunnypapergate #Costco One week everyone buying baby milk powder the next everyone buying up toilet paper. <https://t.co/ScZryVvslh>

Choose text sentiment

☒ Positive ☐ Negative ☐ Neutral

Info History

Selection Details

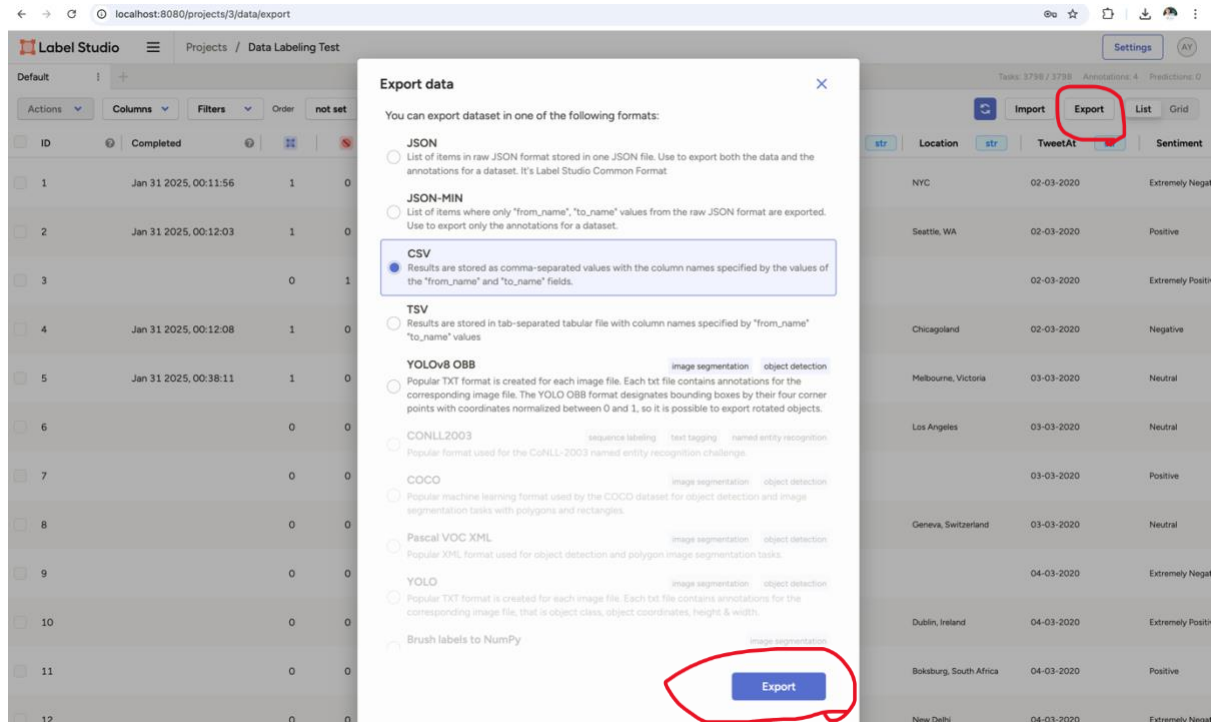
Regions Relations

Manual By Time

Regions not added

Skip Submit

10. In the interface, select the “Export” button. Then select “CSV” as the export format. Select the “Export” button at the bottom of the dialogue, and you will be prompted to save the exported data to your local disk. You can label around 5 to 10 records. You can use this file in your machine learning model training.



Lab 2: Develop a data labeling project for image classification. (5 points)

1. Create a project in label studio and name “fruit image classification labeling”
2. Choose image classification template from Computer Vision.
3. Create three labels: “apple”, “orange”, “strawberry”.
4. Download three images for each fruit and import to your project.
5. Label the images and export your labeled data as CSV file.

Submission – Screenshot of your project page and CSV file.

Optional Lab (Intermediate Level) – Machine Learning Integration

Using Machine Learning Model as a backend service in label studio (<https://labelstud.io/guide/ml>)

In this lab, you can practice integration of machine learning model to your data labeling pipeline. You can use an ML backend to integrate your model development pipeline with your data labeling workflow. There are several use cases, including:

- **Pre-annotate/autolabel data:** Let ML/AI models predict labels autonomously, which can then be reviewed by human annotators.
- **Interactive labeling:** Integrate ML models into the platform to help humans label or annotate large datasets more efficiently and accurately.
- **Model evaluation and fine-tuning:** Labelers review and analyze the model outputs to assess model accuracy and optimize performance.

Tutorial – using Segment Anything (SAM) model in your data labeling pipeline

<https://labelstud.io/blog/get-started-using-segment-anything/>