Week 3 Lab: Data Labeling with Label Studio

CS 203: Software Tools and Techniques for AI

Duration: 3 hours

Lab Overview

Goal: Master the labeling workflow from setup to export and quality analysis.

Structure:

1. **Setup**: Install and run Label Studio.
2. **Configuration**: Define labeling interfaces (UI).
3. **Annotation**: Perform text and image labeling.
4. **Quality Analysis**: Calculate Inter-Annotator Agreement (Cohen's Kappa).

Exercise 1: Label Studio Setup (20 min)

Task: Get the environment running.

```
# 1. Create environment
python -m venv label_env
source label_env/bin/activate

# 2. Install
pip install label-studio

# 3. Launch
label-studio start
```

Action:

- Open browser at `http://localhost:8080`.
- Create an account (local only).

Exercise 2: Text Classification Project (30 min)

Task: Create a "Sentiment Analysis" project.

1. **Create Project**: Name it "Movie Sentiment".
2. **Import Data**: Create a file `reviews.txt` with 5 lines:

   ```
   The movie was fantastic!
   I slept through the whole thing.
   Acting was okay, but plot was weak.
   Worst movie ever.
   Masterpiece.
   ```

   Upload this file.
3. **Setup Labeling Interface**:
   - Go to `Settings` -> `Labeling Interface`.
   - Select `Text Classification`.
   - Edit labels: Add `Positive`, `Negative`, `Neutral`.

Exercise 3: Image Annotation Project (30 min)

Task: Create an "Object Detection" project.

1. **Create Project**: "Car Detection".
2. **Import Data**: Download 3 random car images from the web and upload them.
3. **Setup Interface**:
    - Select `Object Detection with Bounding Boxes`.
    - Labels: `Car`, `Wheel`, `Lights`.
4. **Annotate**: Draw boxes around cars and wheels in your images.

Deliverable: Export the result as JSON and inspect the coordinate format.

Exercise 4: Inter-Annotator Agreement (60 min)

Task: Calculate Cohen's Kappa for reliability.

Since we are individuals, we will *simulate* two annotators.

Scenario:

- **Annotator A**: `[1, 1, 0, 0, 1, 1, 0, 1, 0, 0]`
- **Annotator B**: `[1, 1, 0, 1, 1, 0, 0, 1, 0, 0]`
- (1 = Positive, 0 = Negative)

Code: Write a Python script `calc_kappa.py`.

```
from sklearn.metrics import cohen_kappa_score

annotator_a = [1, 1, 0, 0, 1, 1, 0, 1, 0, 0]
annotator_b = [1, 1, 0, 1, 1, 0, 0, 1, 0, 0]

kappa = cohen_kappa_score(annotator_a, annotator_b)
```

Exercise 5: Export and Integration (20 min)

Task: Load labeled data into Pandas.

1. Export your "Movie Sentiment" project as **JSON**.
2. Write `load_labels.py`:

```
import pandas as pd
import json

with open('project-1-at-2023...json') as f:
    data = json.load(f)

# Extract text and label
records = []
for item in data:
    text = item['data']['text']
    # Label Studio can have multiple annotations per item
    # We take the first one
    label = item['annotations'][0]['result'][0]['value']['choices'][0]
    records.append({'text': text, 'label': label})
```

Submission

Deliverables:

1. Screenshots of your Label Studio Text & Image projects.
2. `calc_kappa.py` with manual calculation logic.
3. `load_labels.py` showing successful parsing of your export.

Resources:

- Label Studio Config Tags: `https://labelstud.io/tags/`
- Sklearn Metrics: `sklearn.metrics.cohen_kappa_score`