Name : Niranjan Tiwari

Roll. NO: 142502019

Question 1 : Dataset loading + stats

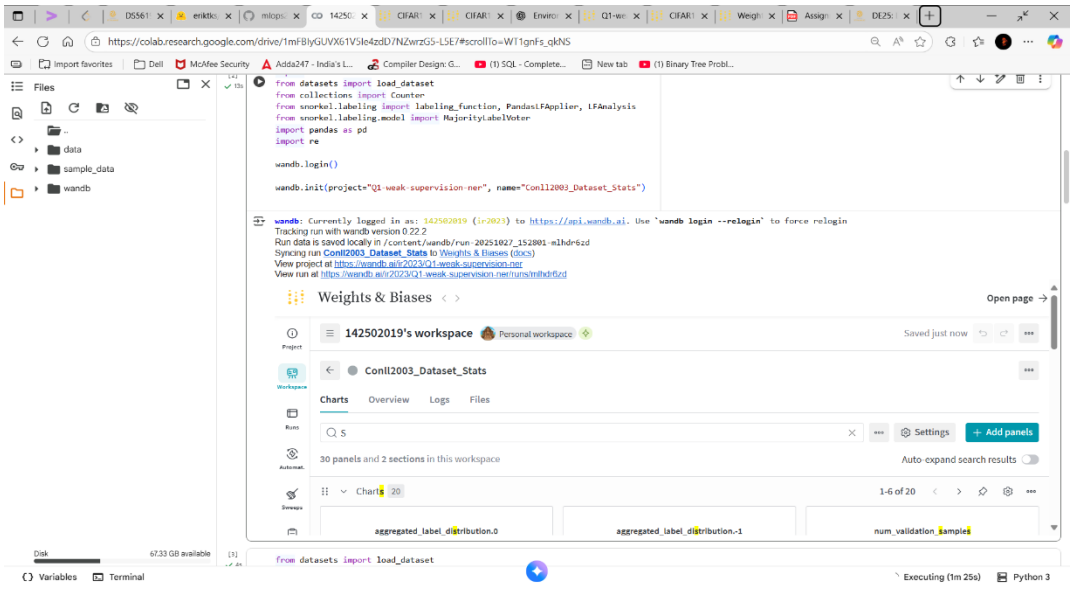
[1] ✓ 21s

```
import wandb
import torch
import torchvision
import torch.nn as nn
import torch.optim as optim
import torchvision.transforms as transforms
from datasets import load_dataset
from collections import Counter
from snorkel.labeling import labeling_function, PandasLFApplier, LFAnalysis
from snorkel.labeling.model import MajorityLabelVoter
import pandas as pd
import re

wandb.login()

wandb.init(project="Q1-weak-supervision-ner", name="Conll2003_Dataset_Stats")
```

```
wandb: Currently logged in as: 142502019 (ir2023) to https://api.wandb.ai. Use `wandb login --relogin` to force relogin
Tracking run with wandb version 0.22.2
Run data is saved locally in /content/wandb/run-20251027_124230-hzkigkhz
Syncing run Conll2003_Dataset_Stats to Weights & Biases (docs)
View project at https://wandb.ai/ir2023/Q1-weak-supervision-ner
View run at https://wandb.ai/ir2023/Q1-weak-supervision-ner/runs/hzkigkhz
Display W&B run
```



Question 2: Labeling functions + W&B logs

[3] ✓ 4s

```python
from datasets import load_dataset
dataset = load_dataset("conll2003")


# Dataset statistics
num_train = len(dataset['train'])
num_valid = len(dataset['validation'])
num_test  = len(dataset['test'])

# Count entity tags across all splits
all_entities = []
for split in ['train','validation','test']:
    for sample in dataset[split]['ner_tags']:
        all_entities.extend(sample)
entity_counts = Counter(all_entities)

# Log to W&B
wandb.log({
    "num_train_samples": num_train,
    "num_validation_samples": num_valid,
    "num_test_samples": num_test,
    "entity_distribution": dict(entity_counts)
})
print(" Dataset statistics logged to W&B.")
```

[3] ✓ 1s

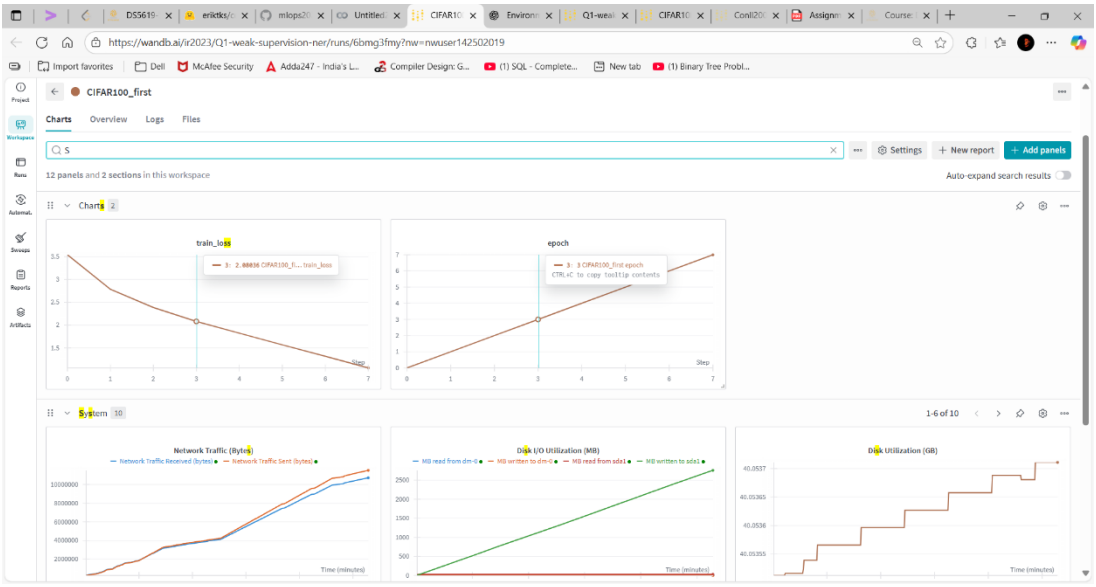
```python
# Convert a small subset to Pandas DataFrame for Snorkel demo
train_df = pd.DataFrame({
    "tokens": [" ".join(tokens) for tokens in dataset['train']['tokens'][:2000]],  # use subset for speed
    "ner_tags": dataset['train']['ner_tags'][:2000]
})
train_df.head()
```

	tokens	ner_tags
0	EU rejects German call to boycott British lamb .	[3, 0, 7, 0, 0, 0, 7, 0, 0]
1	Peter Blackburn	[1, 2]
2	BRUSSELS 1996-08-22	[5, 0]
3	The European Commission said on Thursday it di...	[0, 3, 4, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, ...
4	Germany 's representative to the European Unio...	[5, 0, 0, 0, 0, 3, 4, 0, 0, 0, 1, 2, 0, 0, 0, ...

Next steps: Generate code with train_df | New interactive sheet

```
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
The repository for conll2003 contains custom code which must be executed to correctly load the dataset. You can inspect the repository conten
You can avoid this prompt in future by passing the argument `trust_remote_code=True`.

Do you wish to run the custom code? [y/N] y
Downloading data: 100% 983k/983k [00:00<00:00, 5.79MB/s]
Generating train split: 100% 14041/14041 [00:02<00:00, 7148.81 examples/s]
Generating validation split: 100% 3250/3250 [00:00<00:00, 6727.53 examples/s]
Generating test split: 100% 3453/3453 [00:00<00:00, 8086.24 examples/s]
✅ Dataset statistics logged to W&B.
```



Question 3: Label aggregation

[10] ✓ 0s

```
from collections import Counter
from snorkel.labeling.model import MajorityLabelVoter

# Aggregate labels using MajorityLabelVoter
majority_model = MajorityLabelVoter()
majority_labels = majority_model.predict(L=L_train)

# Convert NumPy int64 keys to str for wandb
label_counts = Counter(majority_labels)
label_counts_clean = {str(int(k)): int(v) for k, v in label_counts.items()}

# Log cleaned counts to W&B
wandb.log({
    "aggregated_label_distribution": label_counts_clean
})

print(" Aggregated label distribution logged to W&B successfully.")
```

```
Aggregated label distribution logged to W&B successfully.
```

Question 4: CIFAR training + experiments

```
from torchvision.models import resnet18

# 1 CIFAR100 → CIFAR10
train100, test100, n100 = get_loaders("CIFAR100")
model = resnet18(num_classes=n100).to(device)
train_model(model, train100, test100, epochs=10, run_name="CIFAR100_first")

train10, test10, n10 = get_loaders("CIFAR10")
model.fc = nn.Linear(model.fc.in_features, n10).to(device)
train_model(model, train10, test10, epochs=10, run_name="CIFAR100_then_CIFAR10")

# 2 CIFAR10 → CIFAR100
train10, test10, n10 = get_loaders("CIFAR10")
model = resnet18(num_classes=n10).to(device)
train_model(model, train10, test10, epochs=10, run_name="CIFAR10_first")

train100, test100, n100 = get_loaders("CIFAR100")
model.fc = nn.Linear(model.fc.in_features, n100).to(device)
train_model(model, train100, test100, epochs=10, run_name="CIFAR10_then_CIFAR100")
```

```
Finishing previous runs because reinit is set to 'default'.
```

```
100%|████████| 169M/169M [00:03<00:00, 42.9MB/s]
Finishing previous runs because reinit is set to 'default'.
```

Run history:

num_test_samples _
num_train_samples _
num_validation_samples _

Run summary:

num_test_samples 3453
num_train_samples 14041
num_validation_samples 3250

View run Conll2003_Dataset_Stats at: https://wandb.ai/ir2023/Q1-weak-supervision-ner/runs/hzkigkhz
View project at: https://wandb.ai/ir2023/Q1-weak-supervision-ner
Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)
Find logs at: `./wandb/run-20251027_124230-hzkigkhz/logs`
Tracking run with wandb version 0.22.2
Run data is saved locally in `/content/wandb/run-20251027_125112-6bmg3fmy`
Syncing run **CIFAR100_first** to Weights & Biases (docs)
View project at https://wandb.ai/ir2023/Q1-weak-supervision-ner
View run at https://wandb.ai/ir2023/Q1-weak-supervision-ner/runs/6bmg3fmy

Code cell output actions

Run history:

epoch ▁▃▅▆▇▇▇▇▇▇
train_loss ▇▆▅▄▃▃▂▂▂▁

Run summary:

epoch 9
train_loss 0.6386

View run CIFAR100_first at: https://wandb.ai/ir2023/Q1-weak-supervision-ner/runs/6bmg3fmy
View project at: https://wandb.ai/ir2023/Q1-weak-supervision-ner
Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)
Find logs at: `./wandb/run-20251027_125112-6bmg3fmy/logs`
100%|████████| 170M/170M [00:02<00:00, 75.6MB/s]
Tracking run with wandb version 0.22.2
Run data is saved locally in `/content/wandb/run-20251027_145020-m1a3rdz6`
Syncing run **CIFAR100_then_CIFAR10** to Weights & Biases (docs)
View project at https://wandb.ai/ir2023/Q1-weak-supervision-ner
View run at https://wandb.ai/ir2023/Q1-weak-supervision-ner/runs/m1a3rdz6

```
[ ]  Start coding or generate with AI.
```

Final WnB Components: