MLOPS ASSIGNMENT – 5

Name : Nelli Uma Maheswar Reddy

Roll no:142502018

Github_link:https://github.com/UmaMaheswarReddy-IIT-Pkd/mlops2025_142502018/tree/main/mlops_Assignments

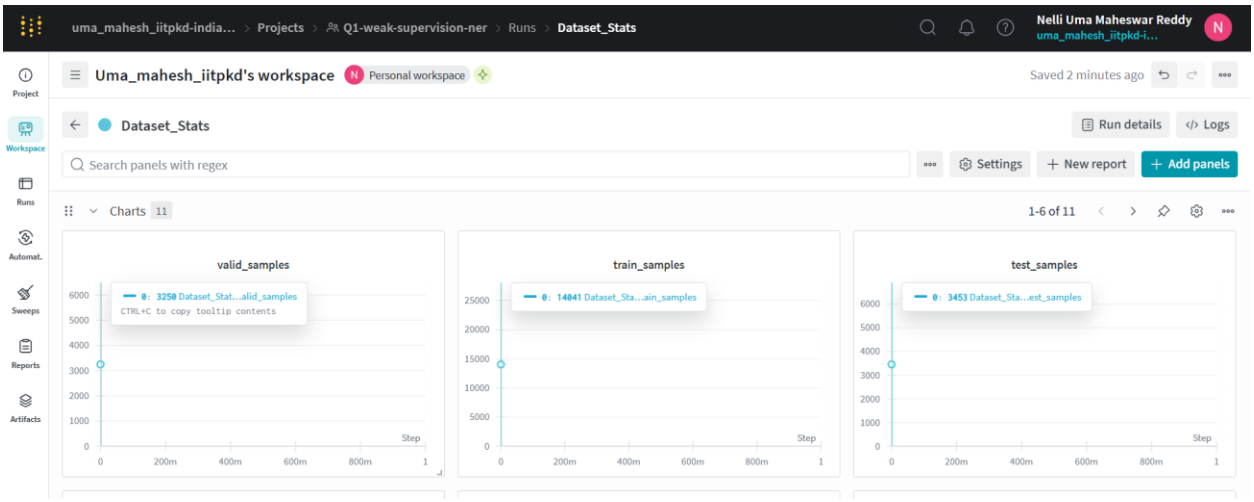
Question 1:

Use the CoNLL-2003 Named Entity Recognition dataset which contains four entity types: PER (Person names) LOC (Locations) ORG (Organizations) MISC (Miscellaneous entities) Load the CoNLL-2003 dataset using HuggingFace datasets (https://huggingface.co/datasets/eriktks/conll2003) and initialize a Weights & Biases project called "Q1-weak-supervision-ner". Log the dataset statistics (number of samples, entity distribution) to W&B as summary metrics. 5 Marks

Output :

Run summary:

test_samples	3453
train_samples	14041
valid_samples	3250


uma_mahesh_iitpkd-india... > Projects > Q1-weak-supervision-ner > Runs > Dataset_Stats
Nelli Uma Maheswar Reddy
uma_mahesh_iitpkd-i...
Project
Workspace
Runs
Automat.
Sweeps
Reports
Artifacts
Uma_mahesh_iitpkd's workspace
Personal workspace
Saved 2 minutes ago
Dataset_Stats
Run details
Logs
Search panels with regex
Settings
New report
Add panels
Charts 11
1-6 of 11
valid_samples
0: 3250 Dataset_Stat...alid_samples
CTRL+C to copy tooltip contents
train_samples
0: 14041 Dataset_Sta...ain_samples
test_samples
0: 3453 Dataset_Sta...est_samples
Step


entity_distribution_train.I-PER
0: 4528 Dataset_Stat..._train.I-PER
CTRL+C to copy tooltip contents
entity_distribution_train.I-ORG
0: 3704 Dataset_Sta...train.I-ORG
entity_distribution_train.I-MISC
0: 1155 Dataset_Sta...rain.I-MISC
Step

Question 2:

Implement two basic labeling functions using Snorkel AI: a. A heuristic function detecting years (1900-2099) as potential DATE/MISC entities b. A pattern-matching function identifying organizations by common suffixes ("Inc.", "Corp.", "Ltd.") Log each labeling function's coverage and accuracy to W&B using wandb.log() 10 Marks

```
              j Polarity  Coverage  Overlaps  Conflicts  Correct  Incorrect  \
lf_year       0      [7]  0.073322   0.00195    0.00195      279      14651
lf_org_suffix 1       []  0.000000   0.00000    0.00000        0          0
lf_loc_suffix 2      [5]  0.027846   0.00195    0.00195      266       5404

              Emp. Acc.
lf_year        0.018687
lf_org_suffix  0.000000
lf_loc_suffix  0.046914
```

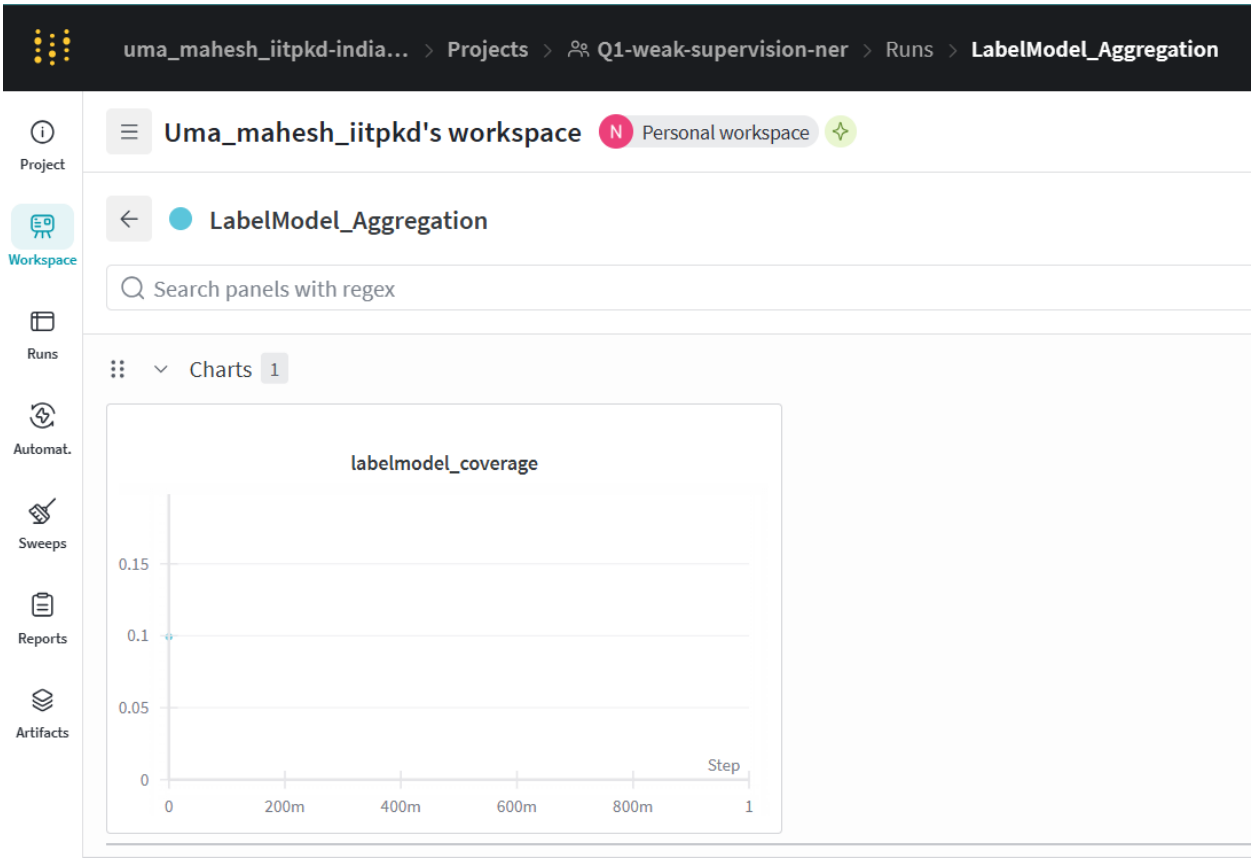
Run summary:

lf_loc_suffix_accuracy	0.04691
lf_loc_suffix_conflicts	0.00195
lf_loc_suffix_coverage	0.02785
lf_loc_suffix_overlaps	0.00195
lf_org_suffix_accuracy	0
lf_org_suffix_conflicts	0
lf_org_suffix_coverage	0
lf_org_suffix_overlaps	0
lf_year_accuracy	0.01869
lf_year_conflicts	0.00195
+2	...

Question 3:

3. Implement Snorkell's Label aggregation (Majority Label Voter) 5 Marks

Output:




```
INFO:root:Computing O...
INFO:root:Estimating \mu...
  0%|                                                                                                | 0/500 [00:00<?, ?epoch/s]INFO:roo
t:[0 epochs]: TRAIN:[loss=0.003]
 16%|████████████                                                                                   | 78/500 [00:00<00:01, 362.54epoch/s]INFO:roo
t:[100 epochs]: TRAIN:[loss=0.000]
 37%|████████████████████████████                                                                   | 185/500 [00:00<00:00, 341.74epoch/s]INFO:roo
t:[200 epochs]: TRAIN:[loss=0.000]
 58%|████████████████████████████████████████████                                                   | 289/500 [00:00<00:00, 328.78epoch/s]INFO:roo
t:[300 epochs]: TRAIN:[loss=0.000]
 79%|████████████████████████████████████████████████████████████                                   | 394/500 [00:01<00:00, 341.41epoch/s]INFO:roo
t:[400 epochs]: TRAIN:[loss=0.000]
100%|████████████████████████████████████████████████████████████████████████████████████████████| 500/500 [00:01<00:00, 332.38epoch/s]
INFO:root:Finished Training
```

```
Aggregated label coverage: 0.0992
```

Run history:

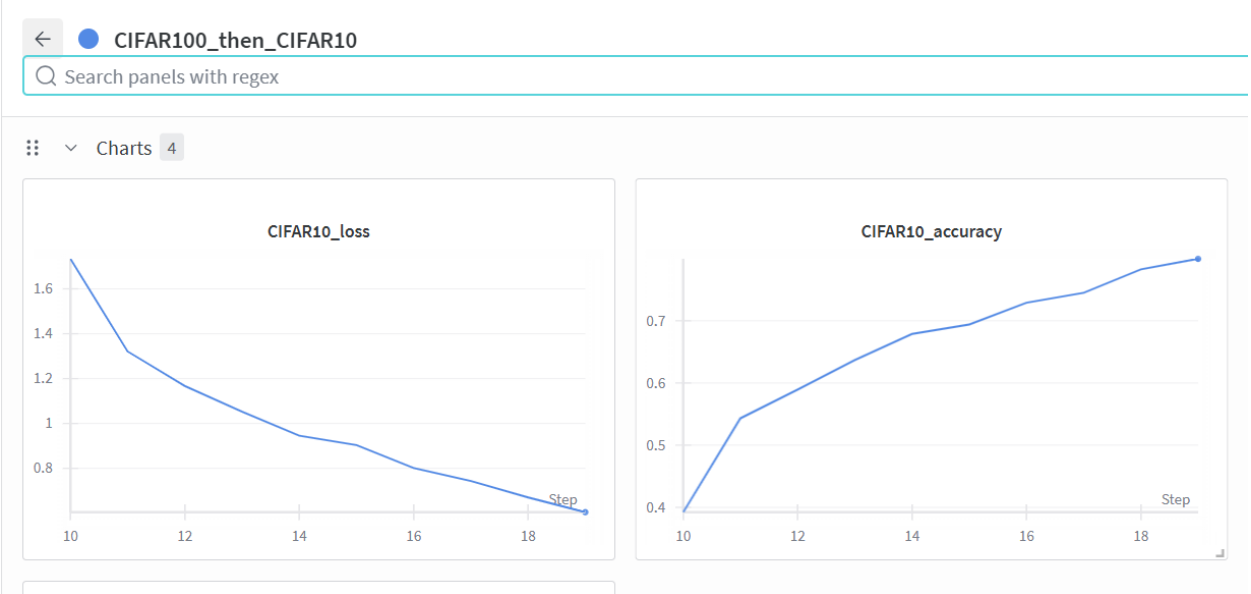
labelmodel_coverage	_

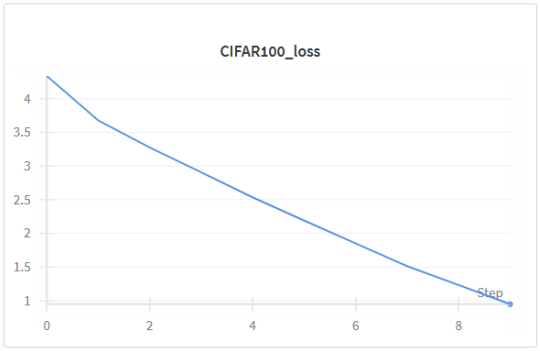
Run summary:

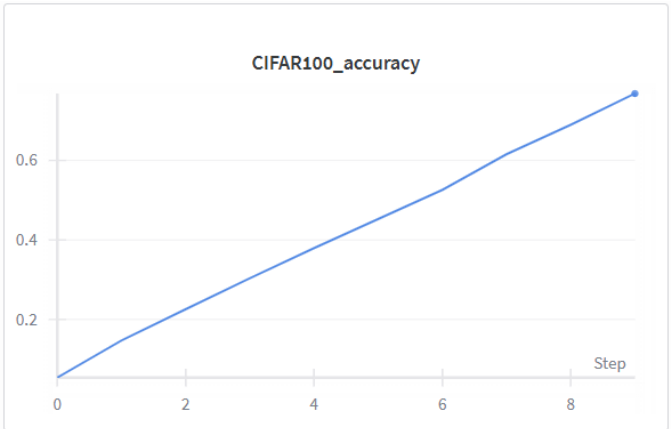
labelmodel_coverage	0.09922

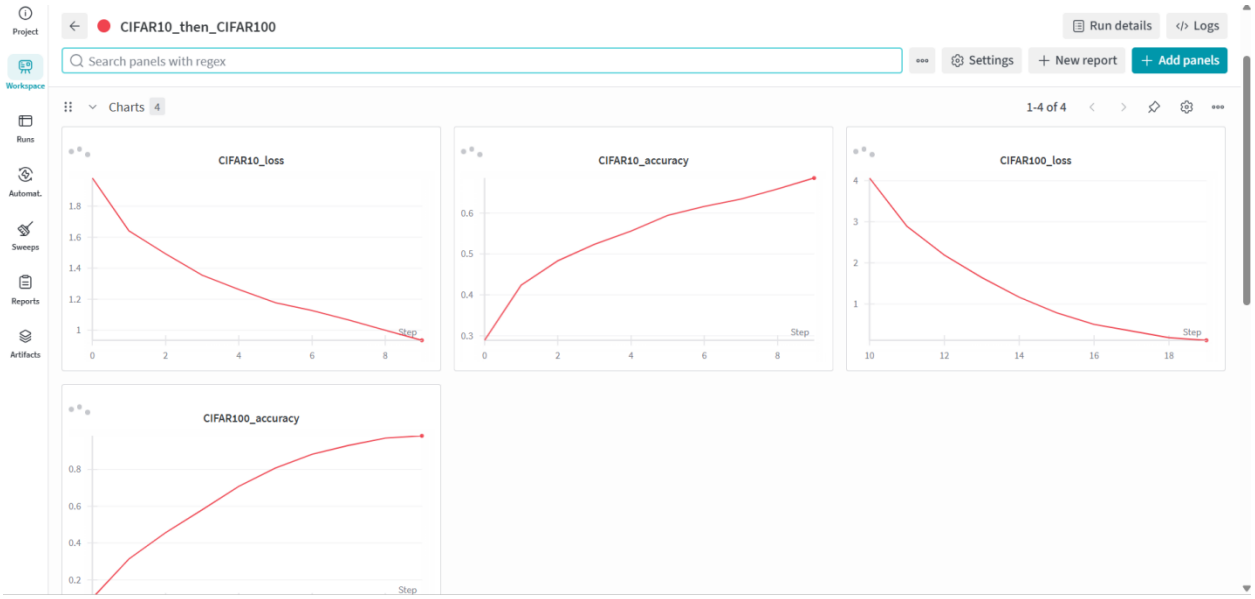
Question 4:

Implement the following in Weights and Bias: a. Train CIFAR 100 and CIFAR 10 sequentially for 100 epochs b. Train CIFAR 10 and CIFAR 100 sequentially for 100 epochs.








Project
Workspace
Runs
Automat.
Sweeps
Reports
Artifacts
CIFAR10_then_CIFAR100
Run details
Logs
Search panels with regex
Settings
New report
Add panels
Charts 4
1-4 of 4
CIFAR10_loss
CIFAR10_accuracy
CIFAR100_loss
CIFAR100_accuracy
Step