

Assignment 5

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
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
3. Implement Snorkel's Label aggregation (Majority Label Voter) 5 Marks
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.Write your observations with experimental proof.

Upload the screen shots and code as a single PDF file.

Helpful links

1. <https://www.kdnuggets.com/weights-biases-a-kdnuggets-crash-course>
2. <https://snorkel.readthedocs.io/en/v0.9.3/packages/autosummary/labeling/snorkel.labeling.LabelingFunction.html>
3. <https://towardsdatascience.com/weak-supervision-with-snorkel-for-multilabel-classification-tasks-c7af4990ea45/>
4. <https://www.kdnuggets.com/2020/07/labelling-data-using-snorkel.html>
5. <https://snorkel.readthedocs.io/en/v0.9.3/packages/autosummary/labeling/snorkel.labeling.MajorityLabelVoter.html>
6. <https://www.youtube.com/playlist?list=PLZePYakcDhmgkczq3OTRXLbe1BTXRXF4I>