Running notebook checklist FS25 (to be continued)

General criteria

Experiment tracking:

- Important metrics tracked (train loss, train accuracy, val loss, val accuracy)
- Report or view present and linked in introduction
- Report/view organized (important metrics shown, useful ordering, unimportant metrics hidden, useful smoothing settings)
- Useful naming of runs
- If many runs shown, useful ordering (e.g. by validation loss, not by date)

Notebook sections

Preprocessing: Data loading, preprocessing (includes converting to tensors, batching), input format

- Preprocessing decisions match the model and task
- Correct and justified decisions on:
    - Tokenization
    - Lowercasing, stemming, lemmatizing, stopword/punctuation removal
    - Removal of unknown/other words
    - Format cleaning (e.g. html-extracted text)
    - Truncation
    - Feature selection
    - Input format: how is data passed to the model?
    - Label format: what should the model predict?
    - Train/valid/test splits
    - Batching, padding
    - Vocabulary, embedding

Model: Model selection, model architecture, classifier

- Correct and justified model architecture:
    - Pretrained model checkpoint (if applicable)
    - Network architecture (layers, dims, nonlinearities, regularizations, normalizations, classifier)

- o Loss, optimizer (either here or in the training section)

Training: Training code, hyperparameters, training runs, selection of best model, early stopping

- Correct and justified decisions on:
  - o Loss, optimizer (either here or in the model section)
  - o Experiment design (model/optimizer/loss variants, hyperparameters)
    - ▪ What do we want to find out?
  - o Number of training runs
  - o Model checkpointing
  - o Early stopping
- Optional, but recommended: automated training runs (later: sweeps)

Evaluation: Code to compute results, presentation and description of results

- Metrics and confusion matrix present
- Good error analysis (e.g. analyze the examples where the model makes mistakes, try to identify patterns)
- Clean graphics, labeled axes (if used), no screenshots
- Objective description of results
- Subjective interpretation and speculation not present (belong in next section)

Tools used:

- Tools used are stated (includes AI tools)