

# COMPSCI 4NL3: Competition Launch

Winter 2025  
Due: March 16th

## 1 Overview

In this part of the project, you will prepare your data and a simple baseline model for condabench. The deliverables will be a single slide advertising your project, and a codabench page.

## 2 Data Preparation

You need ground truth labels for your dataset. You might already have these if you are using an existing dataset. If you are not, you need decide the ground truth from what you have annotated. These will be the final labels used for training and evaluating your models. Some possible methods of determining ground truth include:

1. Majority Vote: The most common label is taken as the ground truth. This requires having more than 2 annotations per data point. If there are only 2 annotators and you need a third vote, you could label some of the data points yourself to provide the tiebreaker.
2. Weighted Voting: Assign weights to annotators based on experience or reliability.
3. Adjudication: Manually review data points where annotators disagree and discuss between the annotators until a resolution is reached on the correct label.

You will need to explain how you computed ground truth or if you used the existing dataset as is. Next, you should analyze the distribution of ground truth labels. Create a visualization to show the amount of data with each label. Are the classes balanced or imbalanced? This information should be described in the Codabench page (probably in a tab named data, but this is up to you).

## 3 Codabench

Create a Codabench page to benchmark models for your task. This platform will allow you and others to submit models and compare their performance. Steps include:

1. Read the Guide: Learn how Codabench works
2. Task Definition: Describe your task and provide a clear description of the dataset.
3. Split Data: Choose a training, validation, and test split for your dataset.
4. Data Upload: Upload your dataset for participants to download or use via an API.
5. Evaluation Metric: Specify which metrics (e.g., accuracy, F1-score) will be used to evaluate submissions.
6. Baseline Model: Set up a baseline model that participants can use as a reference point.
7. Configuration: Configure the task in the way that you see fit. You do not need to use the compute queue, you can allow participants to directly upload a submission file with their model outputs and you compute the score.

Also check the Codabench References and Codabench Guidelines in the Project Steps folder on Avenue. If you need more help, use the final projects channel on Teams.

### 3.1 Create Baselines

Develop baseline models to compare against more advanced models. This will include:

1. Simple Baselines: These are simple models that provide a baseline performance. There are a few options, and you should choose whichever of these two gives the best performance: (1) Random Baseline: Randomly assign labels to the data. (2) Majority Baseline: Always predicts the most common label in the dataset.
2. Trained Model: Train at least one model (e.g. Logistic Regression, Random Forest, Feedforward Neural Network) on your dataset. Evaluate it's performance using a validation set and compare it to your simple baselines.

These will set the lower bound for performance on your task. Make sure to include these with the initial task launch on Codabench. Your trained model should outperform the simple baseline. If you have trouble achieving this, feel free to post on Teams and we can help you with this.

Note that you do NOT have to do anything complicated here. We want a simple rule baseline and a simple trained baseline. Logistic regression with bag-of-words is good. We just want something that others can compare to. Later on, in a different step, you can make a more sophisticated model.

## 4 Slide Advertisement

You should make one slide with Google slides to advertise your task to the class. This should serve as a quick description of your task. You may want to include an example and show what data you have and what the model should predict. Why should others work on your task? How does your Codabench page work? What is the baseline performance? We don't have time to have everyone give a presentation but this will be distributed to the rest of the class.

## 5 Deliverables

On Avenue, your submission should contain a link to your Codabench page with the required information and a link to the Google Slide (make sure it is viewable to anyone who has the link). Make sure that your Codabench page includes the necessary information about your task, your dataset (remember points from §2), and a description of the baselines.