

Estimated schedule

Time	Task	Leads
9:30 - 9:45	CS ML - Intro (15 min)	All (GL leads)
9:45 - 10:00	CS ML - Raising questions (15 min)	GR
10:00 - 10:10	CS ML - Thinking tool (10 min)	GR
10:10 - 10:15	CS ML - Team formation (5 min)	GL
10:15 - 10:45	CS ML - Investigation 1 - toy problem (30 min)	All (walk around)
10:45 - 11:00	CS ML - Thinking tool - accuracy calculation (15 min)	GL
11:00 - 11:30	CS ML - Investigation 1 - cont. (30 min)	All (walk around)
11:30 - 11:45	CS ML - Thinking tool - Introduce real world data and Python jupyter notebook with Google Collab, sklearn, etc. (15 min)	LA
11:45 - 12:45	Lunch for 38 interns + 12 staff (50 total) - Medical + Registration	All
12:45 - 1:30	CS ML - Recap + Investigation 2 - real world problem (30 min)	All (LA leads; walk around)
1:30 - 1:40	CS ML - [Facilitate as necessary] Checkpoint - train/test, model complexity tradeoff (10 min)	All (walk around)
1:40 - 2:30	CS ML - Investigation 2 - cont. (50 min)	All (walk around)
2:30 - 2:40	CS ML - [Facilitation as necessary] Checkpoint - predict new data (10 min)	All (walk around)
2:40 - 3:00	CS ML - Investigation 3 - predict and reflect on new outcome (20 min)	All (walk around)
3:00 - 3:15	CS ML - Prep for jigsaw (15 min)	LA
3:15 - 3:45	CS ML - Jigsaw presentation (3 groups) - CAT (30 min)	All (separate group)
3:45 - 4:00	CS ML - Synthesis + practice reflection (15 min)	All (GL leads, but all will talk)

Content rubric: decision trees	Things to notice
A decision tree predicts a desired attribute of a dataset.	<ul style="list-style-type: none"> - What is the input and the output of the model? - What are you trying to predict?
The process of choosing the “best” feature to split the tree on at each level.	<ul style="list-style-type: none"> - How did you decide which feature to use for the root node? <ul style="list-style-type: none"> - Intuition? Random? - How many features did you pick? - Did you try them all or only a subset? - How do you grow the rest of the tree? - How good is the model once you decided on a feature for a node? - When do you stop growing the tree?
A metric, accuracy, is used to determine model goodness.	<ul style="list-style-type: none"> - How to calculate it? - Can you calculate it for your model?
Tradeoffs between accuracy and complexity of the model.	<ul style="list-style-type: none"> - Plot the two properties and observe the relationship. - Can you pick a best model based on trade-off plot? - Why is it the best model?

STEM practice rubric: optimization	Things to notice
Describe and use a metric to determine model goodness.	<ul style="list-style-type: none"> - Describe the metric - what does it mean, how to calculate. - Apply to the model and calculate the results.
Identify and justify important features in the model.	<ul style="list-style-type: none"> - How to decide whether a feature is important? (Do this at least for the root node for our content)
Perform trade-offs between two desirable but incompatible properties of the model to optimize for best model.	<ul style="list-style-type: none"> - What are the two properties used for trade-off in this case? - What is the relationship between them? - How to optimize for best model using this trade-off?

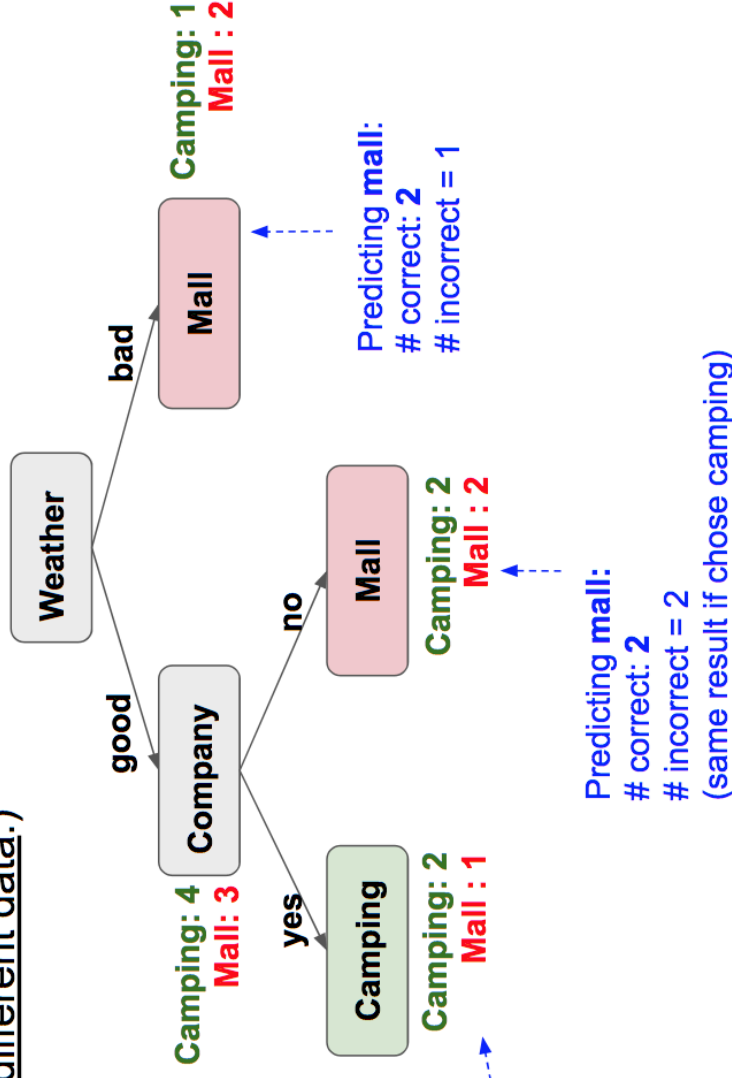
(Same features as before, but different data.)

Predicting: **plan**

Features: **weather, company**

Total samples = **10**

- 5 mall
- 5 camping



Predicting mall:
correct: 2
incorrect = 2
(same result if chose camping)

Overall accuracy

$$= \frac{\text{\# correctly predicted samples}}{\text{total samples}} = \frac{2 + 2 + 2}{10} = 0.6 = \mathbf{60\%}$$

Acceptable? Can we improve the model to get a better accuracy?

Preparing for Jigsaw presentation

Content Prompt:

Build a decision tree to accurately predict a desired attribute of a dataset. Explain or justify how your decision tree solves this predictive problem and maximizes the prediction accuracy.

- Present your final model accuracy and supporting artifacts
 - Visualization of the final decision tree model
 - Trade-off plot that supports your decision on such pick
- Present your new prediction on the new dataset and summarize your thoughts.