

Name		ID		Section	
------	--	----	--	---------	--

1. Suppose you are using decision trees to classify loan applications and observe the following across multiple models. Model A: Tree depth = 2, training accuracy = 65%, test accuracy = 64%, Model B: Tree depth = 10, training accuracy = 92%, test accuracy = 78% and Model C: Tree depth = 20, training accuracy = 99%, test accuracy = 70%. For each model, analyze the bias and variance levels. Which model best balances the bias-variance tradeoff? Explain how you would choose the best tree depth. [5]

2. Consider a dataset with 4 samples, where the class labels are Yes and No. [1+1+1+2]

Class	Frequency
Yes	3
No	1

- Calculate the *Gini Index* for this dataset.
- Calculate the *Entropy* for this dataset.
- Which measure (Gini or Entropy) gives a higher value in this case? Explain why.
- Suppose this dataset is split into two groups based on some feature, and each group contains exactly the same class distribution as the original dataset. Would the Gini Index and Entropy for each group be the same? Justify your answer.