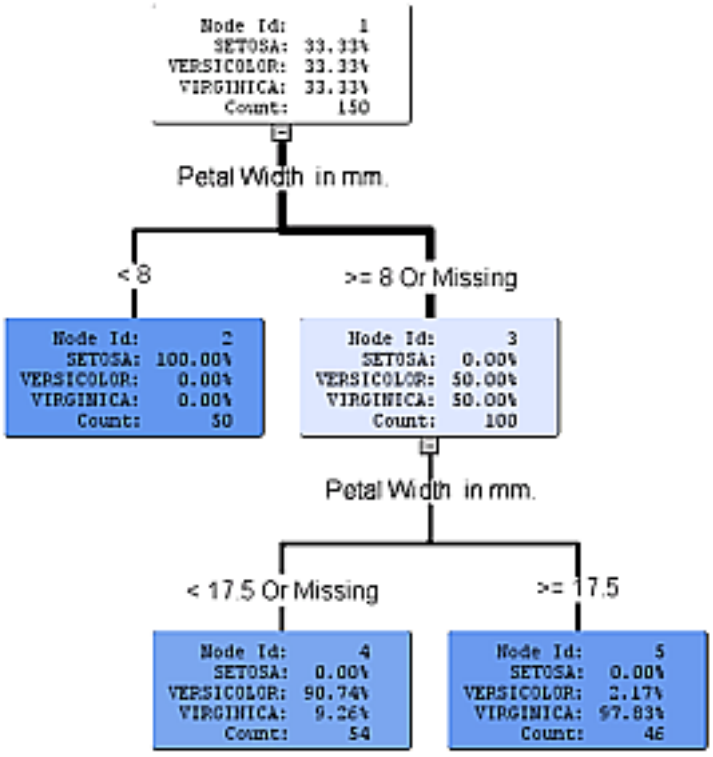
DNSC 6279
Quiz 04

Name:

Use the trained decision tree below to answer the following questions.



1. **(1 pt.)** Write out the English language rules for node 2.

If Petal Width is less than 8 mm, then the observation is predicted to be SETOSA with probability of 1.

2. **(1 pt.)** How many leaf nodes does this tree contain?
3

3. **(1 pt.)** What is the depth of this tree?
2

4. **(2 pts.)** Name two obvious hyperparameters that could have been used when training this decision tree.

Max. depth = 2 (1 pt.), number splits = 2 (1 pt.).

5. **(1 pt.)** If a new record of data has the following attributes:

Petal Length (mm)	Petal Width (mm)	Sepal Length (mm)	Sepal Width (mm)
20	.	14	18

How would it be classified by this tree?
Node 4 OR Versicolor 90.74% and Virginica 9.26%

DNSC 6279 Name:
Quiz 04

6. **(2 pts.)** There are 46 observations in node 5. 1 is a Versicolor iris and 45 are Virginica irises. If a split could be found that split node 5 into 2 child nodes, one node with 1 Versicolor and 0 Virginica irises, the other node having 0 Versicolor and 45 Virginica irises, would this be an advantageous split? Calculate the Information Gain of this split and say whether the split should be performed based on that value.

Information Gain is defined by:

$$GAIN_{split} = Entropy(p) - \left(\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right)$$

where,

$$Entropy(t) = -\sum_j p(j \mid t) \log_2 p(j \mid t)$$

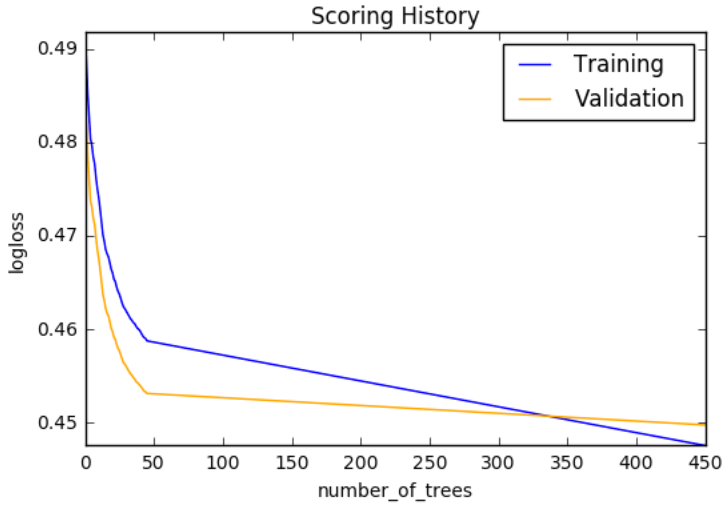
Entropy(parent) = -(1/46)*log$_2$(1/46) - (45/46)*log$_2$(45/46) = 0.151

Entropy(child 1) = -(1/1)*log$_2$(1/1) - (0/1)*log$_2$(0/1) = 0

Entropy(child 2) = -(45/45)*log$_2$(45/45) - (0/45)*log$_2$(0/45) = 0

Information Gain = 0.151 – ((1/46)* 0 + (45/46)* 0) = 0.151 (1 pt.)**, Information Gain is positive so split should occur** (1 pt.)**.**

7. **(1 pt.) True or False:** The iteration plot for the gradient boosting machine displayed below indicates adding additional decision trees into the ensemble could likely decrease validation error.



True (validation error is still decreasing)

8. **(1 pt.) True or False:** In a random forest, numerous decision trees are trained sequentially, each tree attempting to improve the predictions of all previous trees.

False