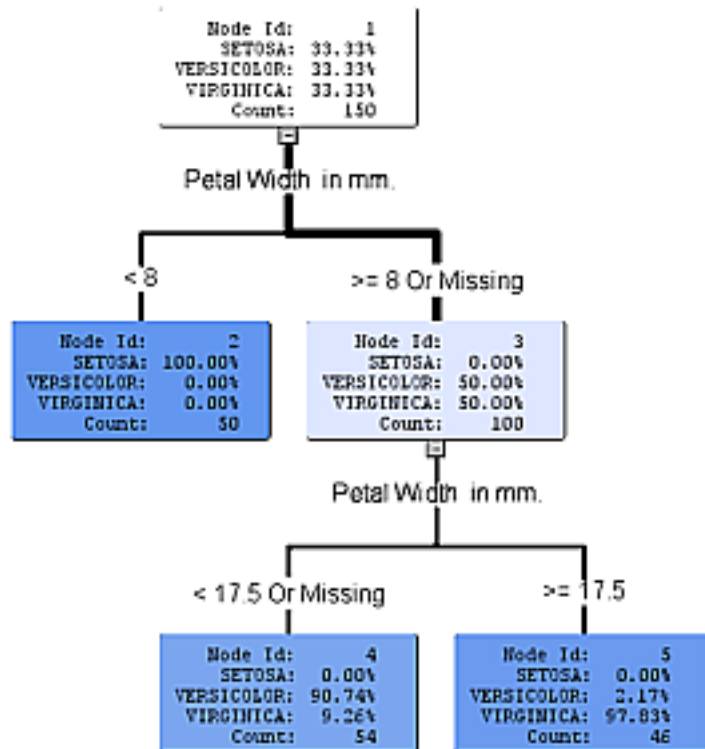Name:

1. **(1 pt.) True or False:** The logworth measure used for splitting in decision trees in SAS Enterprise Miner is a commonly used standard found in many popular open source packages.

False

Use the trained decision tree below to answer questions 2-5.

2. **(3 pts.)** Write out the English language rules defined by node 5.

If an input's petal width is greater than or equal to 17.5 and not missing THEN it is classified as 97.83% Virginica and 2.17% Versicolor

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

4. **(1 pt.)** What is the depth of this tree?
2

5.) **(1 pt.) (True or False)** This tree uses trinary splits to partition each internal node.
False

Name:

5.) **(3 pts.)** There are 54 observations in node 4. 49 are Versicolor irises and 5 are Virginica irises. Should this node be split again into 2 child nodes, one with 47 Versicolor and 1 Virginica iris, the other having 2 Versicolor and 4 Virginica irises? Calculate the Information Gain of this split and state whether the split is valid by this criterion.

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) = -(49/54)*log$_2$(49/54) - (5/54)*log$_2$(5/54) = 0.445

Entropy(child 1) = -(47/48)*log$_2$(47/48) - (1/48)*log$_2$(1/48) = 0.146

Entropy(child 2) = -(4/6)*log$_2$(4/6) - (2/6)*log$_2$(2/6) = 0.918

Information Gain = 0.445 – ((48/54)* 0.146 + (6/54)* 0.918) = 0.213

Yes

(entropy of split child nodes is less than the entropy of the parent node)