

# QF632-2025-W6

Number of participants: 28



## 1. What is FALSE about a decision tree?

**12 correct answers**  
out of 17 respondents

A root node is the topmost node of the tree



2 votes

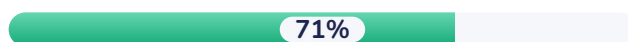
A parent node can have multiple child nodes



2 votes



A leaf node can be further split



12 votes

Decision tree can be interpreted as rules



1 vote



**CART can handle both classification  
2. (y is categorical) and regression (y  
is continuous)**

**15 correct answers**  
out of 16 respondents

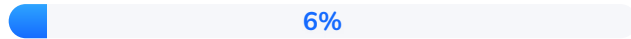


True



15 votes

False



1 vote



### Which of the following statements 3. about the test set in the 3-way split is FALSE?

**7 correct answers**  
out of 14 respondents

It remains unseen until after the optimal, pruned tree is finalized.



5 votes

It provides an unbiased estimate of generalization error.



0 votes



It influences the choice of pruning level through early stopping.



7 votes

No further model modifications are allowed once test-set results are inspected.



2 votes



#### 4. Which node has the largest Gini impurity index?

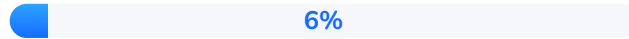
**12 correct answers**  
out of 17 respondents

Node has 5 rows (y value: 3 Yes, 2 No)



4 votes

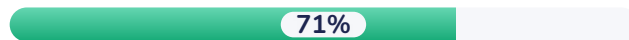
Node has 4 rows (y value: 4 Yes, 0 No)



1 vote



Node has 4 rows (y value: 2 Yes, 2 No)



12 votes

Node has 5 rows (y value: 2 Yes, 3 No)



0 votes



#### 5. What is the entropy of a node with 10 rows, and the y value has 5 Yes, and 5 No?

**5 correct answers**  
out of 13 respondents

$[(5/10)^2 + (5/10)^2]$



0 votes

$1 - [(5/10)^2 + (5/10)^2]$



8 votes

$5/10 * \log_2(5/10) + 5/10 * \log_2(5/10)$



0 votes



$-5/10 * \log_2(5/10) - 5/10 * \log_2(5/10)$



5 votes



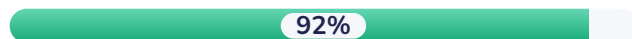
6.

**Empirically, CART splits chosen via entropy tend to be slightly more “pure” than those via Gini when classes are imbalanced. The most defensible explanation is that entropy**

**11 correct answers**  
out of 12 respondents



penalizes rare classes more heavily due to the log term's curvature



11 votes

is always larger than Gini, so its maximization yields deeper trees



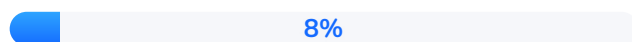
0 votes

forces equal class proportions at every split



0 votes

avoids the bias toward multi-level categorical predictors



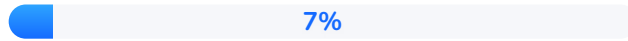
1 vote



## 7. The purpose of pruning for CART is to prevent underfitting of the tree

**13 correct answers**  
out of 14 respondents

True



1 vote



False



13 votes



## 8. Which of the following is a primary advantage of decision trees?

**12 correct answers**  
out of 12 respondents

They can handle both numerical and categorical data.



0 votes

They require little data preparation.



0 votes

They can handle multi-output problems.



0 votes



All of the above.



12 votes



9.

**If a decision tree is grown to its maximum depth without any restrictions, what potential issue might arise?**

**12 correct answers**  
out of 12 respondents

The tree might not be deep enough to capture important patterns.

0%

0 votes



The tree could become too complex and overfit the training data.

100%

12 votes

The tree could become biased towards the majority class.

0%

0 votes

The tree might not be able to handle categorical variables.

0%

0 votes