



DecisionTree  
11 Questions

NAME : \_\_\_\_\_

CLASS : \_\_\_\_\_

DATE : \_\_\_\_\_

1. Thuật toán CART viết tắt của

☐ A Classification and Regression Trees

☐ B Customer Analysis and Research Tool

☐ C Communication Access Real-time Translation

☐ D Computerized Automatic Rating Technique

2. Cây quyết định là phương pháp \_\_\_\_\_.

☐ A học có giám sát

☐ B học không giám sát

☐ C cả hai đều đúng

☐ D cả hai đều sai

3. Khi đang ở một nút trong đó tất cả dữ liệu của nó có cùng nhãn thì?

☐ A Tìm feature tốt nhất để tách

☐ B Tạo node lá chứa kết quả dự đoán

☐ C Dừng lại quá trình đệ quy

☐ D Trở lại node PARENT và chọn feature khác để tiếp tục

4. 

x1	x2	y
0	1	+1
1	0	-1
0	1	+1
1	1	+1

 Chọn feature nào thì phân tách tốt nhất

☐ A x2

☐ B x1

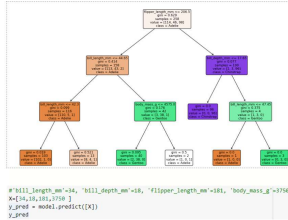
5. `DecisionTreeClassifier(criterion="entropy", max_depth=x)` max\_depth càng tăng thì hiện tượng gì xảy ra ?

☐ A Overfitting

☐ B underfitting

☐ C good fitting

6.



Giá trị y\_pred

A

Gentoo

B

Adelie

C

Chinstrap

7.

Ưu điểm của cây quyết định?

A

Cần ít dữ liệu để huấn luyện

B

Xử lý tốt một lượng dữ liệu lớn trong thời gian ngắn

C

Đơn giản để hiểu và để giải thích, dễ hình dung

D

Chắc chắn tạo được cây tối ưu nhất

8.

Giả sử bạn cần dự đoán giá của của Bitcoin. Bạn cần loại cây gì để dự đoán?

A

Cây phân loại

B

Cây hồi quy

C

Cả hai

9.

Nhược điểm của cây quyết định là gì?

A

Có thể có quá nhiều dữ liệu

B

Sự thay đổi nhỏ trong dữ liệu đầu vào có thể dẫn đến một cây quyết định khác

C

Chúng ta phải cân bằng tập dữ liệu trước khi đào tạo mô hình

D

Tất cả đều đúng

10.

Giả sử trong một bài toán phân loại, bạn đang sử dụng cây quyết định và bạn sử dụng chỉ số Gini làm tiêu chí cho thuật toán chọn đặc trưng cho nút gốc. Tính năng có chỉ số Gini \_\_\_\_ sẽ được chọn.

A

Tối đa

B

Cao nhất

C

Ít nhất

D

Tất cả đều sai

11. Cây quyết định có mấy thành phần chính

A

2

B

3

C

4

D

5

Answer Key			
1. a	2. a	3. b	4. a
5. a	6. b	7. a, b, c	8. b
9. d	10. c	11. b	