**Câu hỏi 1**

Cho mẫu m1 = (4,12,6), m2 = (2,15,9). Tính khoảng cách Manhattan từ mẫu m1 đến m2?

a. 6

b. 4.690

c. 8

d. 4.960

**Câu hỏi 2**

Cho tập dữ liệu S như sau:

Bảng dữ liệu

Entropy thuộc tính Fever trong tập dữ liệu S có giá trị bằng bao nhiêu?

Chọn một:

a. 0.92

b. 0

c. 0.39

d. 0.53

**Câu hỏi 3**

Giải thuật nào sau đây không phải giải thuật phân cụm?

Chọn một:

a. K-Means

b. Naive Bayes

c. K-Medians

d. Expectation Maximisation

**Câu hỏi 4**

Hoạt động nào sau đây là một ứng dụng trong lĩnh vực khai phá dữ liệu?

Chọn một:

a. Phân chia khách hàng trong một công ty theo lợi nhuận của họ

b. Dự đoán kết quả cuả việc tung một cặp xúc xắc

c. Phân chia khách hàng trong một công ty theo giới tính của họ

d. Dự đoán giá cổ phiếu trong tương lai của một công ty bằng cách sử dụng các hồ sơ trong quá khứ

**Câu hỏi 5**

Cho tập dữ liệu như sau:

Bảng dữ liệu

Có bao nhiêu mẫu dữ liệu trong tập dữ liệu trên?

Chọn một:

a. 8

b. 5

c. 6

d. 7

**Câu hỏi 6**

Cho mẫu m1 = (4,12,6), m2 = (2,15,9). Tính khoảng cách Euclidean từ mẫu m1 đến m2?

Chọn một:

a. 4.213

b. 4.123

c. 4.960

d. 4.690

**Câu hỏi 7**

Giải thuật nào sau đây thuộc giải thuật phân lớp?

Chọn một:

a. Logistic Regression

b. Linear Regression

c. K-Means

d. K-Nearest Neighbors

**Câu hỏi 8**

Cho dãy số D = {25, 31, 40, 46, 72, 79, 86, 98, 101}. Median(D) có giá trị bằng bao nhiêu?

Chọn một:

a. 59

b. 59.625

c. 72

d. 46

**Câu hỏi 9**

Trong một ứng dụng AI, con người và máy được kết nối với nhau bằng cách nào?

Chọn một:

a. Quá trình xử lý tín hiệu nhiễu

b. Quá trình xử lý ngôn ngữ tự nhiên

c. Quá trình xử lý sóng cao tần

d. Quá trình xử lý sóng hạ tần

**Câu hỏi 10**

Cho tập dữ liệu D như sau:

Bảng dữ liệu

Trong giải thuật KNN với K = 3 (sử dụng khoảng cách Euclidean), hãy dự đoán nhãn lớp cho mẫu dữ liệu sau:

X = (SepalLengthCm = 5.6, SepalWidthCm = 2.5, PetalLengthCm = 3.9, PetalWidthCm = 1.1)

Chọn một:

a. Outlier

b. Iris-versicolor

c. Noise

d. Iris-setosa

**Câu hỏi 11**

Trong giải thuật cây quyết định, thuộc tính được chọn làm nút gốc cần có đặc điểm gì?

Chọn một:

a. Độ lợi thông tin lớn nhất

b. Độ lợi thông tin bằng 1

c. Độ lợi thông tin bằng 0

d. Độ lợi thông tin nhỏ nhất

**Câu hỏi 12**

Cho tập dữ liệu D như sau:

Bảng dữ liệu

Áp dụng giải thuật K-means với k = 2 (áp dụng khoảng cách Euclidean). Giả sử chọn tâm cụm C1, C2 lần lượt là x1 = (1,1) thuộc C1 và x3 = (4,3) thuộc C2.

Kết quả phân cụm cuối cùng của thuật toán là gì?

Chọn một:

a. C1 = {x1, x4} và C2 = {x3, x2}

b. C1 = {x1, x2} và C2 = {x3, x4}

c. C1 = {x3, x4} và C2 = {x1, x2}

d. C1 = {x1, x3} và C2 = {x2, x4}

**Câu hỏi 13**

Trong giải thuật KNN (K-nearest neighbors), K được hiểu là gì?

Chọn một:

a. Là số lượng điểm dữ liệu gần nhất của điểm dữ liệu được xem xét

b. Là số lượng mẫu dữ liệu được dùng để huấn luyện mô hình

c. Là số lượng cụm được chia dựa trên cấu trúc có sẵn có tập dữ liệu

d. Là số lượng mẫu dữ liệu được dùng để kiểm tra mô hình

**Câu hỏi 14**

Cho dãy số D = {25, 31, 45, 73, 40, 29, 56, 58}. Mean(D) có giá trị bằng bao nhiêu?

Chọn một:

a. 44.265

b. 44.625

c. 44.562

d. 44.526

**Câu hỏi 15**

Đâu là nhược điểm của giải thuật K-means?

Chọn một:

a. Đơn giản

b. Tất cả các đối tượng đều phải gán vào các nhóm, kể cả noise và outlier

c. Các đối tượng tự động gán vào các nhóm

d. Thường đạt tối ưu cực độ

**Câu hỏi 16**

Thuật toán nào sau đây thuộc phương pháp "Ensemble Learning"?

Chọn một:

a. Decision Tree

b. K-Nearest neighbors

c. K-Means

d. Random Forest

**Câu hỏi 17**

Cho tập dữ liệu D như sau

Bảng dữ liệu

Giả sử có một đối tượng mới với các giá trị thuộc tính như sau:

B = (Outlook = Rainy, Temperature = Hot, Humidity = High, Windy = False)

Áp dụng thuật toán Naïve Bayes, hãy dự đoán với điều kiện thời tiết của đối tượng B thì sẽ thuộc nhãn lớp nào và với xác suất bao nhiêu?

Chọn một:

a. Play Golf = Yes và P(B|Play Golf = Yes)\*P(Play Golf = Yes) = 0.008

b. Play Golf = No và P(B|Play Golf = No)\*P(Play Golf = No) = 0.056

c. Play Golf = No và P(B|Play Golf = No)\*P(Play Golf = No) = 0.5

d. Play Golf = Yes và P(B|Play Golf = Yes)\*P(Play Golf = Yes) = 0.012

**Câu hỏi 18**

Trong giải thuật K-Means, yếu tố nào đại diện cho mỗi cụm?

Chọn một:

a. Số lượng đối tượng thuộc cụm

b. Khoảng cách từ các đối tượng đến tâm cụm

c. Số lượng cụm của tập dữ liệu

d. Trọng tâm cụm

**Câu hỏi 19**

Đâu không phải là một kỹ thuật để thu thập những dữ liệu Web?

Chọn một:

a. Logging

b. Scraping

c. Interviews

d. Crawling

**Câu hỏi 20**

Cho dãy số D = {20, 12, 32, 23, 40, 45, 55, 54, 35, 21, 56, 27}. Midrange(D) có giá trị bằng bao nhiêu?

Chọn một:

a. 12

b. 56

c. 34

d. 35

**Câu hỏi 21**

Giả sử thuộc tính A có các giá trị như sau A = {24, 15, 36, 40, 48, 50, 72, 80}. Dùng phương pháp chuẩn hóa z-score, hãy chuẩn hóa cho giá trị 72 của thuộc tính A?

Chọn một:

a. 1.235

b. 1.273

c. 1.565

d. 3.372

**Câu hỏi 22**

Đâu không phải là một ví dụ về AI?

Chọn một:

a. Trợ lý ảo Siri

b. Hợp đồng thông minh Etherium

c. Robot Sophia

d. Trợ lý giải đáp Watson

**Câu hỏi 23**

Giả sử thuộc tính A có các giá trị như sau A = {26, 31, 40, 48, 72, 79, 86, 98}. Khi ánh xạ các giá trị của thuộc tính A thuộc [-1,1] sử dụng phương pháp chuẩn hóa min-max, thì giá trị 98 của thuộc tính A sau khi chuẩn hóa là bao nhiêu?

Chọn một:

a. -1

b. 0.5

c. 0

d. 1

**Câu hỏi 24**

Những lĩnh vực nào góp phần thúc đẩy sự phát triển của AI trong giai đoạn hiện tại?

Chọn một:

a. Deep Learning

b. Machine Learning

c. Data Science

d. Cả 3 đáp án đều đúng

**Câu hỏi 25**

Tiến bộ giáo dục nào dưới đây đến từ công nghệ AI?

Chọn một:

a. Bài giảng số

b. Cá nhân hóa việc học

c. Kết nối học tập từ bất cứ đâu

d. Thiết lập thí nghiệm ảo

**Câu hỏi 26**

Cho dãy số D = {46, 36, 40, 46, 76, 79, 46, 40, 96}. Mode(D) có giá trị bằng bao nhiêu?

Chọn một:

a. 76

b. 46

c. 40

d. 36

**Câu hỏi 27**

Thứ tự các bước trong tiến trình khai phá dữ liệu:

Chọn một:

a. Data collection; Data vizualization &amp; Grasping; Analysis, hypothesis testing &amp; Machine learning; Data processing; Insight &amp; policy decision

b. Data collection; Data processing; Data vizualization &amp; Grasping; Analysis, hypothesis testing &amp; Machine learning; Insight &amp; policy decision

c. Data collection; Analysis, hypothesis testing &amp; Machine learning; Data processing; Data vizualization &amp; Grasping; Insight &amp; policy decision

d. Data collection; Data processing; Analysis, hypothesis testing &amp; Machine learning; Data vizualization &amp; Grasping; Insight &amp; policy decision

**Câu hỏi 28**

Giải thuật nào sau đây không thuộc phương pháp học có giám sát?

Chọn một:

a. Naive Bayes

b. Linear Regression

c. Decision Tree

d. Principal Component Analysis (PCA)

**Câu hỏi 29**

Cho tập dữ liệu S như sau:

Bảng dữ liệu

Trong tập dữ liệu S ở trên, thuộc tính nào sẽ được chọn làm nút gốc (root)?

Chọn một:

a. Không có thuộc tính nào được chọn

b. Breathing issues

c. Fever

d. Cough

**Câu hỏi 30**

Trong học máy, kết quả của quá trình "Training" là gì?

Chọn một:

a. Các mẫu dữ liệu cũ bị thay đổi giá trị

b. Các mẫu dữ liệu mới được sinh ra

c. Mô hình học máy đã được huấn luyện

d. Độ chính xác của mô hình học máy