Câu hỏi **1**

Cho dãy số D = {15, 30, 20, 22, 45, 50, 73, 90, 105}. Mean(D) có giá trị bằng bao nhiêu?

Chọn một:

a. 105

b. 60

c. 50

d. 15

Câu hỏi **2**

Giải thuật nào sau đây thuộc giải thuật phân lớp?

Chọn một:

a. Logistic Regression

b. K-Nearest Neighbors

c. K-Means

d. Linear Regression

Câu hỏi **3**

Cho mẫu m1 = (3,4,6,10,14), m2 = (5,1,9,15,26). Tính khoảng cách Euclidean từ mẫu m1 đến m2?

Chọn một:

a. 5

b. 25

c. 3.72

d. 13.82

Câu hỏi **4**

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

Chọn một:

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

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

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

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

Câu hỏi **5**

Cho mẫu m1 = (2,5,16,30,25), m2 = (4,9,10,35,30). Tính khoảng cách Manhattan từ mẫu m1 đến m2?

Chọn một:

a. 4.69

b. 10.30

c. 22

d. 3.21

Câu hỏi **6**

Trong giai đoạn tiền xử lý dữ liệu, có thể sử dụng phương pháp nào sau đây?

Chọn một:

a. Transforming

b. Integrating

c. Tất cả các đáp án đều đúng

d. Cleaning

Câu hỏi **7**

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

Chọn một:

a. Data collection; Analysis, hypothesis testing &amp; Machine learning; Data processing; Data vizualization &amp; Grasping; 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; Data vizualization &amp; Grasping; Analysis, hypothesis testing &amp; Machine learning; Data processing; 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 **8**

Giả sử giá trị trung bình và độ lệch chuẩn của thuộc tính A tương ứng là 45000 và 20000. Sử dụng phương pháp chuẩn hóa z-score, giá trị 79500 của thuộc tính A sẽ có giá trị là?

Chọn một:

a. 3.975

b. 1.767

c. 1.725

d. 1.322

Câu hỏi **9**

Giả sử giá trị lớn nhất và nhỏ nhất của thuộc tính A tương ứng là 2000 và 250. Khi ánh xạ các giá trị của thuộc tính A thuộc [-1,1] sử dụng phương pháp chuẩn hóa min-max, thì giá trị 1500 của thuộc tính A sau khi chuẩn hóa là bao nhiêu?

Chọn một:

a. 0.143

b. 0.357

c. 0.429

d. 0.714

Câu hỏi **10**

Phát biểu sau đây đúng hay sai: Học máy là xây dựng những hệ thống có khả năng tự cải thiện bản thân bằng cách học từ dữ liệu

Chọn một:

a. Sai

b. Đúng

Câu hỏi **11**

Phát biểu sau đây đúng hay sai: Học máy là một lĩnh vực khoa học riêng, độc lập với trí tuệ nhân tạo

Chọn một:

a. Sai

b. Đúng

Câu hỏi **12**

Giả sử thuộc tính A có các giá trị như sau A = {24, 15, 36, 40, 48, 50, 72, 80}. Sử dụng phương pháp chuẩn hóa z-score, giá trị 79500 của thuộc tính A sẽ có giá trị là?

Chọn một:

a. 1.218

b. 3.372

c. 1.565

d. 1.235

Câu hỏi **13**

Đâu không phải là một ví dụ về AI?

Chọn một:

a. Robot Sophia

b. Trợ lý giải đáp Watson

c. Hợp đồng thông minh Etherium

d. Trợ lý ảo Siri

Câu hỏi **14**

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 giới tính của họ

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

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

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

Câu hỏi **15**

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. Trọng tâm cụm

d. Số lượng cụm của tập dữ liệu

Câu hỏi **16**

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

Chọn một:

a. Decision Tree

b. K-Means

c. Random Forest

d. K-Nearest neighbors

Câu hỏi **17**

Phát biểu sau đây đúng hay sai: Khai phá dữ liệu là tìm ra những tri thức mới và hữu dụng từ các tập dữ liệu lớn

Chọn một:

a. Đúng

b. Sai

Câu hỏi **18**

Bài toán nào sau đây điển hình cho phương pháp học máy không giám sát?

Chọn một:

a. Bài toán hồi quy

b. Cả 3 đáp án đều đúng

c. Bài toán phân cụm

d. Bài toán phân lớp

Câu hỏi **19**

Đoạn văn câu hỏi

Cho tập dữ liệu D như sau:

|  |  |  |
| --- | --- | --- |
|  | **A1** | **A2** |
| **x1** | **1** | **1** |
| **x2** | **2** | **1** |
| **x3** | **4** | **3** |
| **x4** | **5** | **4** |

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.

Tâm cụm C1 và C2 được cập nhật sau lần lặp thứ nhất có giá trị bao nhiêu?

a.C1 = (2, 1) và C2 = (5, 4)

b. C1 = (1.5, 1) và C2 = (4.5, 3.5)

c. C1 = (1, 1) và C2 = (4, 3)

d. C1 = (2.5, 2) và C2 = (3.5, 2)

Câu hỏi **20**

Trong hệ thống phân loại email (normal/spam), khi ánh xạ sang bộ (P,T,E) thì P được hiểu là gì?

Chọn một:

a. Tỷ lệ phân loại chính xác các email vào đúng nhóm của nó

b. Số lượng các email được phân loại đúng vào nhóm của nó

c. Số lượng các email được phân loại sai vào nhóm của nó

d. Tỷ lệ phân loại sai các email vào nhóm của nó

Câu hỏi **21**

AI chưa được ứng dụng vào lĩnh vực nào?

Chọn một:

a. Chế tạo máy

b. Sáng tác

c. Trị liệu tâm lý

d. Đánh cờ

Câu hỏi **22**

Cho dãy số D = {7, 9, 10, 25, 45, 59, 65, 100}. Median(D) có giá trị bằng bao nhiêu?

Chọn một:

a. 45

b. 35

c. 25

d. 40

Câu hỏi **23**

Trí tuệ nhân tạo là gì?

a. Đặt bộ nhớ thông minh vào máy tính

b. Huấn luyện để máy tính trở nên thông minh

c. Tăng tốc độ đường truyền cho máy tính

d. Tăng bộ nhớ cho máy tính

Câu hỏi **24**

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

a. Độ chính xác của mô hình học máy

b. Các mẫu dữ liệu mới được sinh ra

c. Mô hình học máy đã được huấn luyện

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

Câu hỏi **25**

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

a. Crawling

b. Interviews

c. Scraping

d. Logging

Câu hỏi **26**

Phương pháp học máy có giám sát là gì?

a. Tất cả dữ liệu huấn luyện đã được gán nhãn, mô hình học máy sau khi huấn luyện có khả năng dự đoán nhãn đầu ra cho các mẫu dữ liệu mới

b. Tất cả dữ liệu không được gán nhãn và các giải thuật học máy học các cấu trúc vốn có từ dữ liệu đầu vào

c. Cả 3 đáp án đều sai

d. Một số dữ liệu được gán nhãn nhưng phần lớn dữ liệu còn lại không có nhãn được sử dụng kết hợp cho các giải thuật học máy

Câu hỏi **27**

Điều kiện dừng của giải thuật K-means là gì?

a. Tổng khoảng cách từ các đối tượng thuộc cụm đến trọng tâm cụm giảm dần

b. Sau K vòng lặp

c. Trọng tâm các cụm không thay đổi

d. Sau 10 vòng lặp

Câu hỏi **28**

Dữ liệu nào sau đây được xem là dữ liệu có cấu trúc?

a. Dạng bảng

b. URL

c. Image

d. Voice

Câu hỏi **29**

Cho tập dữ liệu D như hình trên. Trong giải thuật KNN với K = 3 (sử dụng khoảng cách Euclidean), hãy dự đoán nhãn lớp cho mẫu dữ liệu sau:

X = (Pregnancies = 101, Glucose = 76, BloodPressure = 48, SkinThickness =180, Insulin = 32.9)

a. 1

b. Noise

c. Outlier

d. 0

Câu hỏi **30**

Cho dãy số D = {5, 15, 30, 15, 45, 40, 30, 5, 25, 50, 15, 35}. Mode(D) có giá trị bằng bao nhiêu?

a. 45

b. 5

c. 30

d. 15

Câu hỏi **31**

**Cho tập dữ liệu S như sau:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Fever** | **Cough** | **Breathing issues** | **Infected** |
| **1** | **No** | **No** | **No** | **No** |
| **2** | **Yes** | **Yes** | **Yes** | **Yes** |
| **3** | **Yes** | **Yes** | **No** | **No** |
| **4** | **Yes** | **No** | **Yes** | **Yes** |
| **5** | **No** | **Yes** | **No** | **No** |

**Entropy (S) có giá trị bằng bao nhiêu?**

a. 0.52

b. 0.48

c. 0.84

d. 0.97

Câu hỏi **32**

Một hệ thống học máy được định nghĩa thông qua bộ (P,T,E). Trong đó E được hiểu là gì?

a. Là những kinh nghiệm hoặc những mẫu dữ liệu đã được nhận biết hoặc được gán nhãn trước

b. Là những công việc được xác định khi xây dựng mô hình học máy

c. Là hiệu suất của mô hình học máy

d. Là những mẫu dữ liệu mới đưa vào để kiểm tra mô hình

Câu hỏi **33**

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

a. Thiết lập thí nghiệm ảo

b. Cá nhân hóa việc học

c. Bài giảng số

d. Kết nối học tập từ bất cứ đâu