Câu hỏi 1: Hãy cho biết “Data Mining” được định nghĩa là gì?

1. Là khai khoáng dữ liệu.
2. Là tìm kiếm thông tin trên Internet.
3. **Là một quy trình tìm kiếm, phát hiện các tri thức mới, tiềm ẩn, hữu dụng trong cơ sở dữ liệu lớn.**
4. Là khai phá dữ liệu.

Câu hỏi 2: Hãy cho biết thuật ngữ “Knowledge Discovery from Databases -KDD” có nghĩa là gì?

1. Là khai thác dữ liệu.
2. Là tìm kiếm dữ liệu.
3. **Là trích chọn các mẫu hoặc tri thức hấp dẫn (không tầm thường, ẩn, chưa biết và hữu dụng tiềm năng) từ tập dữ liệu lớn.**
4. Là khai phá dữ liệu.

Câu 3: Trong thuật toán k-mean, sau khi gán các đối tượng vào k cụm ta cần làm gì?

1. Tính khoảng cách giữa các phần tử trong cụm
2. Tìm một số phần tử đại diện của cụm
3. Trộn các cụm lại với nhau để số cụm sinh ra là ít nhất
4. **Tính lại tâm của các cụm**

Câu hỏi 4: Hãy cho biết phát biểu nào sau đây là đúng ?

1. **Mining là một bước trong quá trình khai phá tri thức-KDD.**
2. Tiền xử lí dữ liệu là qua trình tìm kiếm thông tin có ích từ cơ sở dữ liệu lớn.
3. Data Mining là quá trình tìm kiếm thông tin có ích trên Internet.
4. Thuật ngữ Data Mining đồng nghĩa với thuật ngữ Knowledge Discovery from Databases.

Câu hỏi 5: Hãy cho biết thuật ngữ Tiền xử lí dữ liệu bằng tiếng Anh là?

1. Data Processing
2. **Data Preprocessing**
3. Preprocessing in Database
4. Data Process

Câu hỏi 6: Cho CSDL Giao dịch như hình vẽ, hãy cho biết số lượng giao dịch trong cơ sở dữ liệu là bao nhiêu?



1. 10
2. **5**
3. 16
4. 6

Câu hỏi 7: Cho CSDL giao dịch như hình vẽ, Anh (chị) hãy cho biết độ hỗ trợ của tập mục X={A, M} là bao nhiêu?

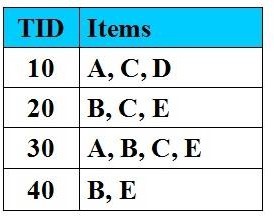


1. 5 (100%)
2. **3 (60%)**
3. 4 (80%)
4. 2 (40%)

Câu hỏi 8: Anh (chị) hãy cho biết thuật toán Apriori có nhược điểm chính là gì?

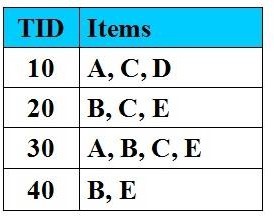
1. Kết quả của thuật toán không ứng dụng được trong các bài toán thực tế.
2. Thuật toán quá phức tạp, khó hiểu.
3. **Tốn nhiều bộ nhớ và thời gian. Không thích hợp với các mẫu lớn. Chi phí để duyệt CSDL nhiều.**
4. Không tìm được các tập thường xuyên.

Câu hỏi 9: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%). Anh (chị) hãy cho biết tập nào là tập mục thường xuyên thỏa Min\_support?



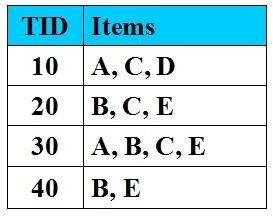
1. {A,D}
2. **{A,C}**
3. {B, C, D}
4. {D}

Câu hỏi 10: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%), nếu sử dụng thuật toán Apriori thì sau lần duyệt thứ nhất, tập mục chứa 1-item nào sẽ bị loại bỏ?



1. **{A}, {D}**
2. {D}
3. {A}
4. {B}

Câu hỏi 11: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%), Min\_Cofidence = 50%. Anh (chị) hãy cho biết luật kết hợp nào thỏa mãn các điều kiện đã cho?

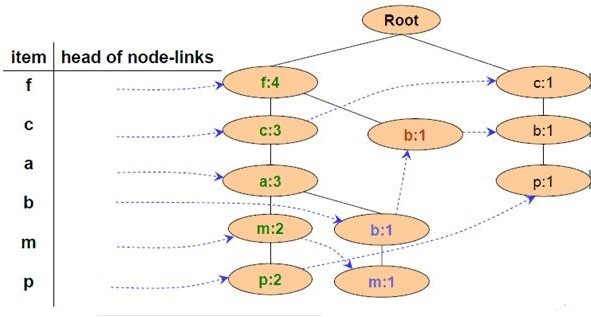


1. A-->D
2. AB-->C
3. A-->C
4. **A--> E**

Câu hỏi 12: Cho tập mục thường xuyên X={A, B}. Anh (chị) hãy cho biết từ tập X có thể sinh ra các luật kết hợp nào?

1. A--> B, A-->A
2. **A--> B, B--> A, không tính luật AB --> Æ và Æ --> AB**
3. B--> A, B-->B
4. A-->B, B--> A, A--> Æ và Æ --> B

Câu hỏi 13: Cho FP-Tree như hình vẽ, Anh (chị) hãy cho biết có mấy đường đi kết thúc ở nút m?



1. 1 đường đi
2. 4 đường đi
3. 3 đường đi
4. **2 đường đi**

Câu hỏi 14: Đường kính của cụm được định nghĩa là khoảng cách giữa 2 phần tử xa nhau nhất trong cùng 1 cụm. Cho 1 cụm gồm các phần tử C={x1, x2, x3, x4} với x1(0,0), x2(1,0), x3(6,0), x4(10, 0). Anh (chị) hãy cho biết đường kính d của cụm có giá trị là bao nhiêu?

1. **d=10**
2. d=6
3. d=1
4. d=4

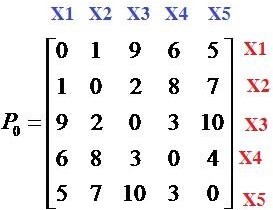
Câu hỏi 15: Hãy cho biết thuật ngữ tiếng Anh nào sau đây có nghĩa là Khai phá dữ liệu?

1. Association Rule
2. Data Classification
3. **Data Mining**
4. Data Clustering

Câu hỏi 16: Hãy cho biết thuật ngữ tiếng Anh nào sau đây có nghĩa là Phân lớp dữ liệu?

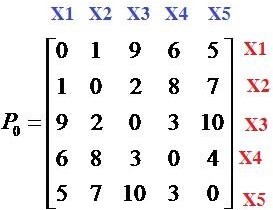
1. Association Rule
2. Data Clustering
3. **Data Classification**
4. Data Mining

Câu hỏi 17: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình. Anh (chị) hãy cho biết khoảng cách giữa 2 phần tử x1 và x2 bằng bao nhiêu?



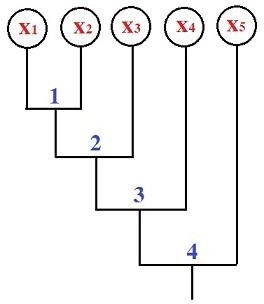
1. bằng 0
2. bằng 9
3. bằng 2
4. **bằng 1**

Câu hỏi 18: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình vẽ. Sử dụng thuật toán liên kết đơn (Single Linkage), bước đầu tiên 2 phần tử nào được chọn để gom thành 1 cụm?



1. x1 và x3
2. **x1 và x2**
3. x2 và x3
4. x3 và x5

Câu hỏi 19: Cho sơ đồ ngưỡng không tương tự như hình vẽ. Cắt sơ đồ tại ngưỡng bằng 2.5 hỏi có mấy cụm được sinh ra ?



1. 2 cụm
2. 1 cụm
3. 4 cụm
4. **3 cụm**

Câu hỏi 20: Anh (chị) hãy cho biết quá trình khai phá tri thức trong CSDL (KDD) có thể phân chia thành các giai đoạn nào trong các giai đoạn sau?

1. Trích chọn dữ liệu, tiền xử lý dữ liệu, biến đổi dữ liệu, khai phá dữ liệu, khai phá luật kết hợp.
2. **Trích chọn dữ liệu, tiền xử lý dữ liệu, biến đổi dữ liệu, khai phá dữ liệu, đánh giá và biểu diễn tri thức.**
3. Tiền xử lý dữ liệu, phân lớp, phân cụm, đánh giá và biểu diễn tri thức.
4. Tiền xử lý dữ liệu, biến đổi dữ liệu, khai phá dữ liệu, đánh giá và biểu diễn tri thức.

Câu hỏi 21: Cho cơ sở dữ liệu giao dịch gồm N giao dịch (bản ghi). I là tập chứa tất cả các mục (item) trong CSDL. X là một tập chứa các mục thuộc I. Giao dịch hỗ trợ X là giao dịch chứa tất cả các mục có trong X. Anh (chị) hãy cho biết độ hỗ trợ của tập mục X được định nghĩa là gì?

1. **Support(X)=Số lượng giao dịch hỗ trợ X / N**
2. Support(X)=Số lượng giao dịch hỗ trợ X
3. Support(X)=Số lượng giao dịch hỗ trợ X \*100%
4. Support(X)=Số lượng giao dịch hỗ trợ X / N \* |I|, trong đó |I| là tổng số mục trong CSDL.

Câu hỏi 22: Cho cơ sở dữ liệu giao dịch gồm N giao dịch (bản ghi). I là tập chứa tất cả các mục (item) trong CSDL. X, Y là tập chứa các mục thuộc I. Anh (chị) hãy cho biết độ tin cậy của luật kết hợp X→ Y được định nghĩa là gì?

1. **Confidence(X→Y)=Số lượng giao dịch hỗ trợ cả X và Y / Số lượng giao dịch hỗ trợ X.**
2. Confidence(X→Y)=Số lượng giao dịch hỗ trợ cả X và Y / Số lượng giao dịch hỗ trợ Y.
3. Confidence(X→Y)=Số lượng giao dịch hỗ trợ X / Số lượng giao dịch hỗi trợ Y.
4. Confidence(X→Y)=Số lượng giao dịch hỗ trợ cả X và Y /N.

Câu hỏi 23: Cho cơ sở dữ liệu giao dịch gồm N giao dịch (bản ghi). I là tập chứa tất cả các mục (item) trong CSDL. X, Y là tập chứa các mục thuộc I. Anh (chị) hãy cho biết độ hỗ trợ của luật kết hợp X→ Y được định nghĩa là gì?

1. Support(X→Y)=Số lượng giao dịch hỗ trợ cả X / Số lượng giao dịch hỗ trợ Y.
2. Support(X→Y)=Số lượng giao dịch hỗ trợ cả X và Y / Số lượng giao dịch hỗ trợ Y.
3. **Support(X→Y)=Số lượng giao dịch hỗ trợ cả X và Y / N.**
4. Support(X→Y)=Số lượng giao dịch hỗ trợ cả X và Y / Số lượng giao dịch hỗ trợ X.

Câu hỏi 24: Cho cơ sở dữ liệu giao dịch gồm N giao dịch (bản ghi). I là tập chứa tất cả các mục (item) trong CSDL. Min\_Supp là độ hỗ trợ tối thiểu, Min\_Conf là độ tin cậy tối thiểu. X, Y là tập chứa các mục thuộc I. Luật kết hợp X→Y được chọn nếu thỏa mãn điều kiện nào?

1. Support(X→Y)<Min\_Supp, Confidence(X→Y)<Min\_Conf
2. Support(X→Y)=Min\_Supp, Confidence(X→Y)=Min\_Conf
3. **Support(X→Y)>=Min\_Supp, Confidence(X→Y)>=Min\_Conf**
4. Support(X→Y)>Min\_Supp, Confidence(X→Y)=Min\_Conf

Câu hỏi 25: Cho CDSL giao dịch như hình vẽ, Độ hỗ trợ tối thiểu Min\_Support = 3 (60%) và độ tin cậy tối thiểu Min\_Confidence = 100%.

Anh (chị) hãy cho biết tập mục thường xuyên có 4 mục thỏa mãn Min\_Supp là tập nào trong các tập mục sau?



1. FCAM:4
2. FCAE:4
3. **FCAM:3**
4. FCAN:2

Câu hỏi 26: Cho CSDL giao dịch như hình vẽ. Độ hỗ trợ tối thiểu Min\_Support = 3 (60%) và độ tin cậy tối thiểu Min\_Confidence = 100%.

Anh (chị) hãy cho biết cơ sở điều kiện của nút M là gì?



1. **{F:2, C:2, A:2} và { F:1, C:1, A:1, B:1}**
2. {F: 3, C:3, A:3}
3. {F:2, C:2, A:2}
4. {F:1, C:1, A:1}

Câu hỏi 27: Cho CDSL giao dịch như hình vẽ. Độ hỗ trợ tối thiểu Min\_Support = 3 (60%) và độ tin cậy tối thiểu Min\_Confidence = 100%.

Anh (chị) hãy cho biết cây điều kiện FP của P là gì?



1. {C:4}|p
2. {CF:3}|p
3. Cây điều kiện là rỗng
4. **{C:3}| p**

Câu hỏi 28: Giả sử ta có các tập mục thường xuyên {A,B}, {A,C}, {B,D} chứa 2-item. Sử dụng thuật toán Apriori để ghép các tập mục có 2-item thành các tập mục có 3-item. Anh (chị) hãy cho biết các ứng viên sinh ra có 3-item là các tập mục nào?

1. {A, B, C}, {C, B, D}
2. **{A, B, C}, {A, B, D}**
3. {A, B, C}, {B, C, D}
4. {A, B, C}, {A, B, D}, {A, B, C, D}

Câu hỏi 29: Trong thuật toán Apriori, tập mục chứa k-item được tạo ra bằng cách nào trong các cách sau?

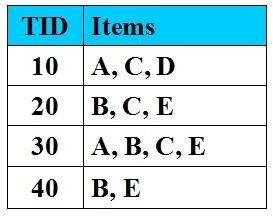
1. Lấy ngẫu nhiên k item sau đó ghép lại với nhau.
2. **Tạo ra từ tập chứa k-1 item bằng cách ghép 2 tập k-1 item với nhau với điều kiện là 2 tập k-1 item này phải có chung nhau k-2 item.**
3. Tổ hợp k item từ các item có trong cơ sở dữ liệu giao dịch.
4. Sinh mọi tập con có k item từ các item có trong cơ sở dữ liệu giao dịch.

Câu hỏi 30: Cho tập L3={abc, abd, ade, ace} là các tập mục thường xuyên chứa 3-item. Để tạo các ứng viên chứa 4-item abcd, ta cần ghép các tập chứa 3-item nào với nhau?

1. **abc và abd**
2. abc và ace
3. abd và ade
4. abc và ade

Câu hỏi 31: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%), Min\_Cofidence = 50%.

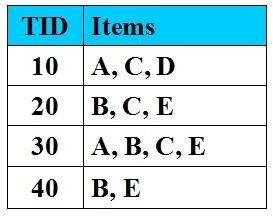
Anh (chị) hãy cho biết luật kết hợp nào có độ tin cậy = 100%?



1. A-->D
2. C-->A
3. C-->
4. **B-->E**

Câu hỏi 32: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%).

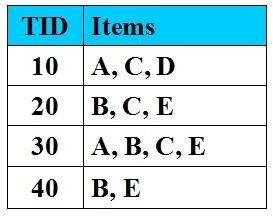
Anh (chị) hãy cho biết tập nào là tập mục thường xuyên có độ hỗ trợ cao nhất?



1. {B, C, D}
2. **{B, E}**
3. {A, E}
4. {A, C, D}

Câu hỏi 33: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%).

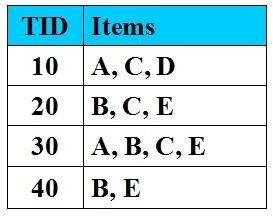
Sử dụng thuật toán Apriori, sau lần duyệt thứ nhất, các danh sách L1 chứa các tập mục thường xuyên có 1-item được tạo ra là gì?



1. L1={{A}, {B}, {D}, {E}}
2. L1={{A}, {D}, {C}}
3. L1={{A}, {B}, {C}, {D}}
4. **L1={{A}, {B}, {C}, {E}}**

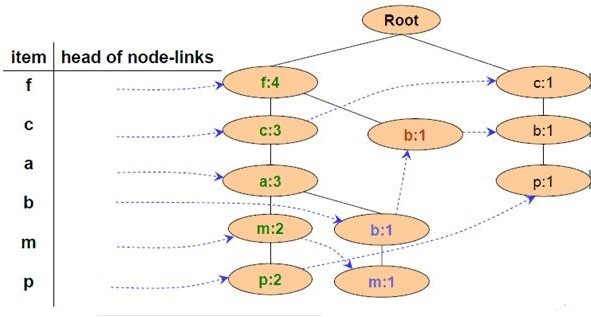
Câu hỏi 34: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%).

Nếu sử dụng thuật toán Apriori để tìm các tập mục thường xuyên thì số lần duyệt CSDL là bao nhiêu?



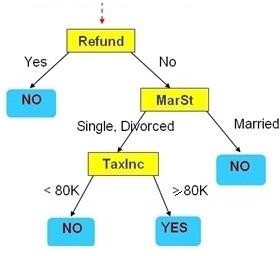
1. 4 lần
2. **3 lần**
3. 2 lần
4. 1 lần

Câu hỏi 35: Cho FP-Tree như hình vẽ, Anh (chị) hãy cho biết mũi tên nét đứt biểu thị cho điều gì?



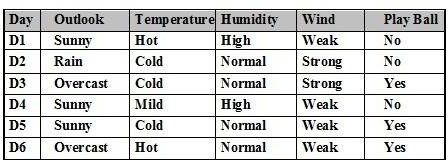
1. Hướng để duyệt cây.
2. **Con trỏ xuất phát từ bảng đầu mục, trỏ vào nút sinh ra đầu tiên có cùng tên. Nút sinh ra sau có con trỏ từ nút cùng tên sinh ra ngay trước đó trỏ vào.**
3. Nút sinh ra sau trỏ vào nút cùng tên sinh ra trước.
4. Đường đi trên cây.

Câu hỏi 36: Cho cây quyết định như hình vẽ. Anh (chị) hãy cho biết có bao nhiêu luật sinh ra từ cây quyết định trên?



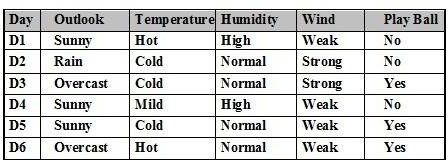
1. **4 luật.**
2. 2 luật.
3. 1 luật.
4. Nhiều luật.

Câu hỏi 37: Cho tập ví dụ học như bảng. Anh (chị) hãy cho biết Entropy của kết luận C= Play Ball là gì?



1. **Entropy(C)=1**
2. Entropy(C)=3
3. Entropy(C)=0.5
4. Entropy(C)=6

Câu hỏi 38: Cho tập ví dụ học như bảng. Anh (chị) hãy cho biết P(Play Ball= ‘Yes’ | Outlook=’Overcast’) có giá trị bằng bao nhiêu?



1. **0/3**
2. 1/3
3. 2/3
4. 3/3

Câu hỏi 39: Anh (chị) hãy cho biết độ phân biệt (độ lộn xộn) của kết luận C với thuộc tính A được tính theo công thức nào?

1. **Gain(C,A)=Entropy(C)-Entropy(A)**
2. Gain(C,A)=Entropy(C)+Entropy(A)
3. Gain(C,A)=Entropy(C)\*Entropy(A)
4. Gain(C,A)=Entropy(A)-Entropy(C)

Câu hỏi 40: Kết luận C gồm 2 giá trị Yes và No. Anh (chị) hãy cho biết Entropy(C)= 1 nói nên điều gì ?

1. **Số kết luận ‘Yes’=Số kết luận ‘No’.**
2. Số kết luận ‘Yes’ =0.
3. Không kết luận được điều gì.
4. Số kết luận ‘No’ =0.

Câu hỏi 41: Kết luận C gồm 2 giá trị Yes và No. Anh (chị) hãy cho biết Entropy(C)= 0 nói nên điều gì:

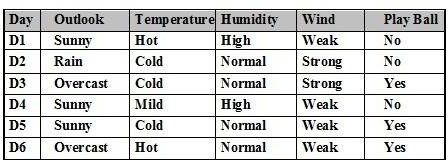
1. **Số kết luận ‘Yes’ bằng 0 hoặc Số kết luận ‘No’ bằng 0**
2. Số kết luận ‘Yes’ bằng Số kết luận ‘No’
3. Không đủ thông tin để phân lớp được tập dữ liệu.

Số kết luận ‘No’ bằng 1 và Số kết luận ‘Yes’ bằng 1

Câu hỏi 42: Anh (chị) hãy cho biết Entropy là một đại lượng có miền giá trị là gì?

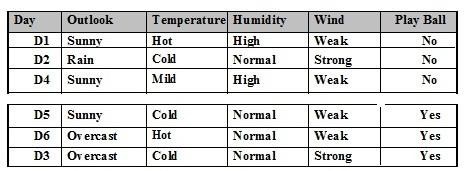
1. Miền giá trị là tập số thực dương.
2. **[0 ; 1].**
3. Miền giá trị là tập số nguyên dương.
4. (0 ; 1).

Câu hỏi 43: Cho tập ví dụ học như bảng. Sử dụng thuật toán ILA. Anh (chị) hãy cho biết có bao nhiêu tổ hợp gồm có 2 thuộc tính phân biệt?



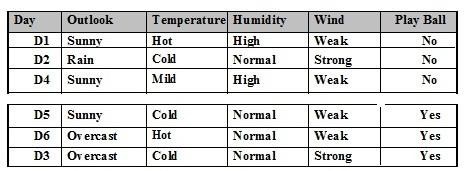
1. 4 tổ hợp.
2. 5 tổ hợp.
3. **6 tổ hợp.**
4. 2 tổ hợp.

Câu hỏi 44: Với bảng Play Ball=’No’, anh (chị) hãy cho biết tổ hợp một thuộc tính nào là tổ hợp lớn nhất?



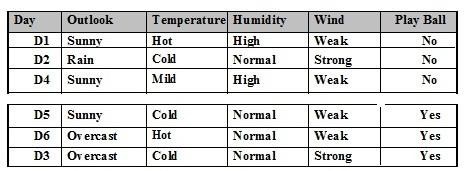
1. **Humidity=’High’**
2. Humidity=’Normal
3. Outlook=’Sunny’
4. Wind=’Weak’

Câu hỏi 45: Cho tập ví dụ học như bảng. Sử dụng thuật toán ILA. Anh (chị) hãy cho biết với bảng Play Ball=’No’ ta tìm được bao nhiêu luật?



1. 3 luật.
2. 4 luật.
3. **2 luật.**
4. 1 luật.

Câu hỏi 46: Cho tập ví dụ học như bảng. Sử dụng thuật toán ILA. Anh (chị) hãy cho biết với bảng Play Ball=’No’ ta tìm được các luật sau nào?



1. If Humidity=’High’ Then Play Ball = ‘No’
2. If Humidity=’Normal’ Then Play Ball = ‘No’và If Outlook = ‘Rain’ Then Play Ball = ‘No’
3. If Outlook=’Sunny’ Then Play Ball = ‘No’và If Wind = ‘Strong’ Then Play Ball = ‘No’
4. **If Humidity=’High’ Then Play Ball = ‘No’ và If Outlook = ‘Rain’ Then Play Ball = ‘No’**

Câu hỏi 47: Trong mặt phẳng tọa độ Oxy, cho 2 điểm A(x1, y1), B(x2, y2). Anh (chị) hãy cho biết khoảng cách Ơclit giữa 2 điểm này được tính bằng công thức nào?

1. a=x1\*y1+x2\*y2.
2. d=x1\*x2+y1\*y2.
3. **d=sqr(sqrt(x1-x2)+sqrt(y1-y2)) trong đó sqr là hàm bình phương, sqrt là hàm lấy căn bậc 2.**
4. d=sqr(sqrt(x1+x2)+sqrt(y1+y2)) trong đó sqr là hàm bình phương, sqrt là hàm lấy căn bậc 2.

Câu hỏi 48: Trong mặt phẳng tọa độ Oxy, cho hai điểm A(0,1), B(4, 4). Anh (chị) hãy cho biết nếu sử dụng độ đo khoảng cách Ơclit thì khoảng cách giữa 2 điểm có giá trị là bao nhiêu ?

1. **d(A,B)=5**
2. d(A,B)=1
3. d(A,B)=4
4. d(A,B)=3

Câu hỏi 49: Cho tập C={x1, x2,. ..xk} gồm k phần tử, mỗi phần tử là một vector trong không gian N chiều. Vector trung bình mC của tập C là một vector trong không gian N chiều. Theo anh chị, vector mC được định nghĩa là gì?

1. mC= (x1+x2+...+xk)/k\*N
2. mC= (x1+x2+...+xk)
3. mC= (x1+x2+...+xk)/N
4. **mC= (x1+x2+...+xk)/k**

Câu hỏi 50: Anh (chị) hãy cho biết trong thuật toán phân cụm k-mean, sau khi chọn được k điểm làm tâm, phần tử x sẽ được gán vào cụm C nếu thỏa mãn điều kiện nào?

1. Khoảng cách từ x đến tâm cụm C bằng 0.
2. **Khoảng cách từ x đến tâm cụm C là nhỏ nhất.**
3. Khoảng cách từ x đến tâm cụm C bằng k.
4. Khoảng cách từ x đến tâm cụm C là lớn nhất.

Câu hỏi 51: Anh (chị) hãy cho biết trong thuật toán k-mean, sau khi gán các đối tượng vào k cụm cần phải là gì?

1. **Tính lại tâm của các cụm.**
2. Trộn các cụm lại với nhau để số cụm sinh ra là ít nhất.
3. Tính khoảng cách giữa các phần tử trong cụm.
4. Tìm một số phần tử đại diện của cụm.

Câu hỏi 52: Cho các điểm A(1, 1), B(2, 1), C(4, 3), D(5, 4), E( 1, 0). Sử dụng thuật toán phân cụm k-mean để chia 5 điểm vào 2 cụm. Anh (chị) hãy cho biết kết quả phân cụm là gì?

1. C1={A, C, E} ; C2={B, D}
2. C1={A, B, D} ; C2={C, E}
3. C1={A, B, C} ; C2={D, E}
4. **C1={A, B, E} ; C2={C, D}**

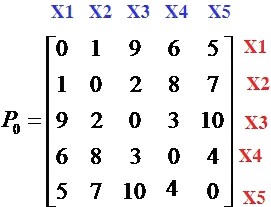
Câu hỏi 53: Cho cụm C gồm các điểm A(1, 1), B(2, 1), C(3,1). Giả sử đại diện của cụm là một điểm (vector trung bình). Anh (chị) hãy cho biết vector trung bình của cụm là có tọa độ là gì?

1. mC = (2 ; 0)
2. mC=(0 ; 0)
3. mC = (2.5 ;1.5)
4. **mC = (2 ; 1)**

Câu hỏi 54: Anh (chị) cho biết thuật toán phân cụm k-mean dừng khi nào?

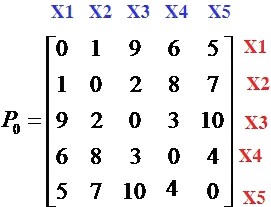
1. Tùy theo yêu cầu của người dùng.
2. Khi tất cả các phần tử đã được gán vào k cụm.
3. **Không thể gán (hoặc gán lại) từng điểm vào cụm khác.**
4. Số cụm sinh ra là k.

Câu hỏi 55: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình. Sử dụng thuật toán liên kết đơn (Single Linkage), sau khi gom x1 và x2 thành cụm C={x1, x2} thì khoảng cách giữa cụm C và x3 bằng bao nhiêu ?



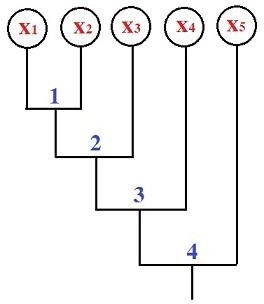
1. bằng 3.
2. bằng 0.
3. **bằng 2.**
4. bằng 9.

Câu hỏi 56: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình. Sử dụng thuật toán liên kết đơn (Single Linkage), sau khi gom x1 và x2 thành cụm C={x1, x2} thì khoảng cách giữa x3 và x4 bằng bao nhiêu?



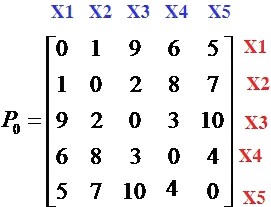
1. bằng 0.
2. bằng 9.
3. **bằng 3.**
4. bằng 2.

Câu hỏi 57: Cho sơ đồ ngưỡng không tương tự như hình vẽ. Cắt sơ đồ tại ngưỡng bằng 3.5. Anh (chị) hãy cho biết các cụm sinh ra là gì?



1. C1={x1, x2} ; C2={x4, x5} ; C3={x3}
2. C1={x1, x2, x3} ; C2={x4, x5}
3. C1={x1, x2} ; C2={x3, x4, x5}
4. **C1={x1, x2, x3, x4} ; C2={x5}**

Câu hỏi 58: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình. Sử dụng thuật toán liên kết đầy đủ (Complete Linkage). Anh (chị) hãy cho biết sau khi gom x1 và x2 thành cụm C={x1, x2} thì khoảng cách giữa x3 và x4 bằng bao nhiêu?



1. bằng 1.
2. bằng 2.
3. **bằng 3.**
4. bằng 9.

Câu hỏi 59: Thuật toán Apriori : Ghép 2 tập mục có 3-tems với nhau để được tập mục có 4-item. Anh (chị) hãy cho biết các tập mục nào có thể ghép được với nhau?

1. L1={gbc}, L2={abd}
2. L1={abc}, L2={ebd}
3. **L1={abc}, L2={abd}**
4. L1={abc}, L2={aed}

Câu hỏi 60: Thuật toán Apriori : Cho 2 tập mục L1={abcd}, L2={abck}. Ghép L1 và L2 với nhau thành tập mục L, ta được tập mục nào sau đây?

1. L={abc}
2. L={dk}
3. L={abck}
4. **L={abcdk}**

Câu hỏi 61: Cho tập mục thường xuyên L={ABCDE}, giả sử tất cả các luật kết hợp sinh ra đều thỏa mãn điều kiện. Anh (chị) hãy cho biết có bao nhiêu luật kế hợp được sinh ra?

1. **Có 30 luật.**
2. Có 5 luật.
3. Có 32 luật.
4. Có 20 luật.

Câu hỏi 62: Cho các điểm A(1, 1), B(2, 1), C(4, 3), D(5, 4), E(0,0). Sử dụng thuật toán phân cụm k-mean để chia 5 điểm vào 2 cụm. Anh (chị) hãy cho biết kết quả phân cụm là gì?

1. C1={A, B, E} ; C2={C, D, E}
2. C1={A, E} ; C2={B, C, D}
3. **C1={A, B, E} ; C2={C, D}**
4. C1={A, B} ; C2={C, D}

Câu hỏi 63: Cho các điểm A(1, 1), B(2, 1), C(4, 3), D(5, 4), E(0,0). Sử dụng thuật toán phân cụm k-mean để chia 5 điểm vào 2 cụm. Cụm C1 được hình thành gồm 3 phần tử A, B, E. Anh (chị) hãy cho biết vector trung bình (trọng tâm) của cụm là gì?

1. mC1=(0, 2/3)
2. mC1=(1, 2)
3. mC1=(1, 3)
4. **mC1=(1, 2/3)**

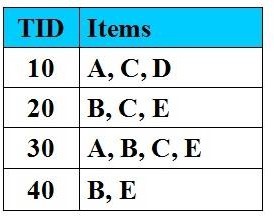
Câu hỏi 64 : Anh (chị) cho biết trong quá trình .Tiền xử lí dữ liệu người ta thường dùng một số phương pháp chuẩn hóa dữ liệu nào trong nhóm các phương pháp sau?

1. Đưa về hệ đếm thập phân, Hệ nhị phân, hệ Hecxa.
2. **Min-Max, z-Score, Tỷ lệ thập phân – decimal scale**.
3. 2NF, 3NF, BCNF.
4. Chuẩn hóa về dữ liệu văn bản, hình ảnh, âm thanh.

Câu hỏi 65: Cho X ={A, B} không là tập mục thường xuyên, Y = {A, B, C}. Anh (chị) hãy cho biết kết luận nào sau đây là đúng?

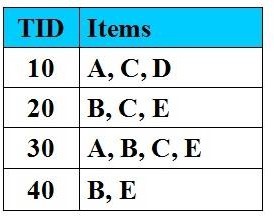
1. C không là tập mục thường xuyên.
2. **Y không là tập mục thường xuyên.**
3. Y là tập mục thường xuyên.
4. X là tập mục thường xuyên.

Câu hỏi 66: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%).Sử dụng thuật toán Apriori. Anh (chị) hãy cho biết sau lần duyệt thứ hai, danh sách L2 chứa các tập mục thường xuyên có 2-item được tạo ra là gì?



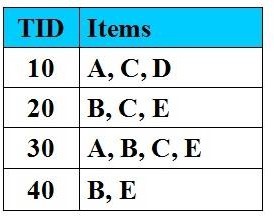
1. L2={{ A,D}, {B,D}, {B,E}, {C,E}}
2. L2= {{A,C}, {C,E}}
3. L2= {{B,C}, {B,E}, {C,E}}
4. **L2={{A,C}, {B,C}, {B,E}, {C,E}}**

Câu hỏi 67: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%).Sử dụng thuật toán Apriori. Anh (chị) hãy cho biết sau lần duyệt thứ hai, các tập mục có 2-item không thỏa mãn độ hỗ trợ tối thiểu là các tập nào?



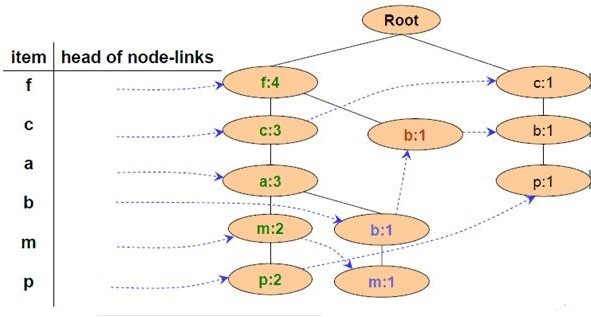
1. **{A,B} và {A,E}.**
2. {A,C} và {B,C }.
3. {A,B} và {A,D}.
4. {A,C}, {B,E}.

Câu hỏi 68: Cho CSDL giao dịch như hình vẽ với Min\_Support = 2 (50%).Sử dụng thuật toán Apriori, cho L2={{A,C}, {B,C}, {B,E}, {C,E}} là danh sách các tập mục thường xuyên có 2-item. Giả sử tập mục {A,B} và {A,E} không là tập mục thường xuyên. Sau khi ghép các tập mục thường xuyên 2-item với nhau để được danh sách L3 chứa các tập mục thường xuyên có 3-item. Anh (chị) hãy chọn danh sách L3 đúng trong số các danh sách sau?



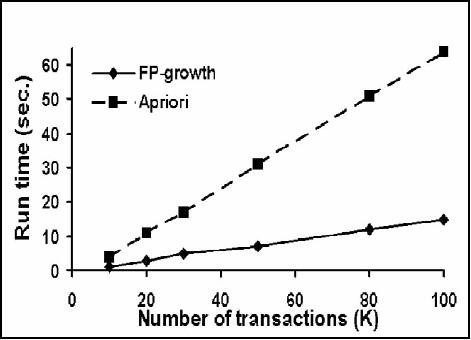
1. L3={{A,B,C,}}
2. **L3={{B, C, E}}**
3. L3={{A,B, C} và {A,C,E}}
4. L3={{A,C,E} và {B,C,E }}

Câu hỏi 69: Cho FP-Tree như hình vẽ, Anh (chị) hãy cho biết cây điều kiện FP của nút a là cây nào?



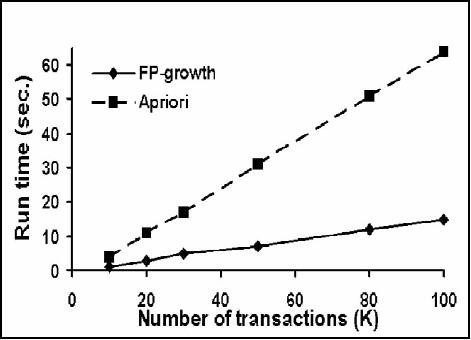
1. {f:4, c:3}
2. **{f:3, c:3}**
3. {f:3, c:3, a:3
4. {f:4, c:3, a:3}

Câu hỏi 70: Anh (chị) hãy cho biết đồ thị sau biểu diễn điều gì?



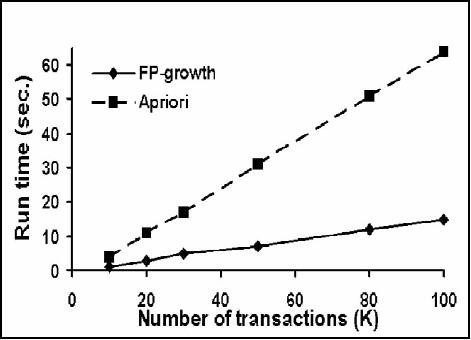
1. **So sánh giữa Thời gian thực thi (tính bằng giây) của 2 thuật toán FP-Growth và Apriori theo số lượng giao dịch (nghìn giao dịch).**
2. So sánh Thời gian thực thi với số lượng giao dịch.
3. So sánh 2 thuật toán FP-Growth và Apriori.
4. Mối quan hệ giữa 2 thuật toán FP-Growth và Apriori.

Câu hỏi 71: Cho đồ thị như hình vẽ. Từ đồ thị anh chị có nhận xét gì?



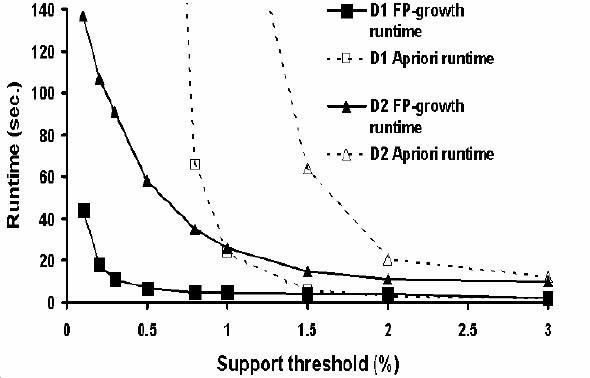
1. **Với cùng số lượng giao dịch như nhau, thời gian thức thi thuật toán FP-Growth luôn ít hơn thời gian thực thi thuật toán Apriori.**
2. Thuật toán Apriori thực hiện nhanh hơn thuật toán FP-Growth.
3. Hai thuật toán FP-Growth và Apriori đều thức thi với thời gian rất nhỏ.
4. Với cùng số lượng giao dịch như nhau, thời gian thực thi của thuật toán FP-Growth luôn nhiều hơn thời gian thực thi của thuật toán Apriori.

Câu hỏi 72: Cho đồ thị như hình vẽ. Anh (chị) hãy cho biết nhận xét nào sau đây là sai?



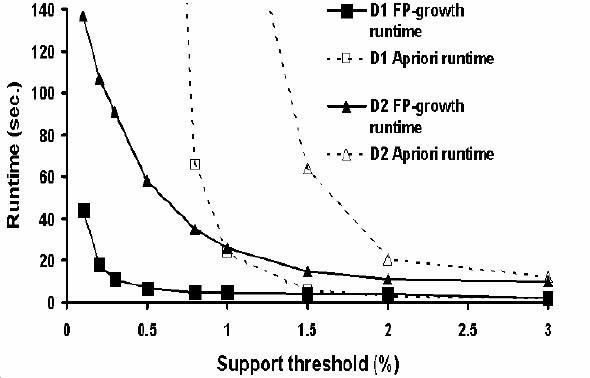
1. Với cùng số lượng giao dịch như nhau, thời gian thức thi thuật toán FP-Growth luôn ít hơn thời gian thực thi thuật toán Apriori.
2. **Thuật toán Apriori thực hiện nhanh hơn thuật toán FP-Growth**.
3. Khi số lượng giao dịch rất nhỏ, thời gian thực thi của 2 thuật toán FP-Growth và Apriori là tương đương.
4. Thuật toán FP-Growth thực hiện nhanh hơn thuật toán Apriori.

Câu hỏi 73: Cho đồ thị như hình vẽ, Anh (chị) hãy cho biết đồ thị trên biểu diễn gì ?



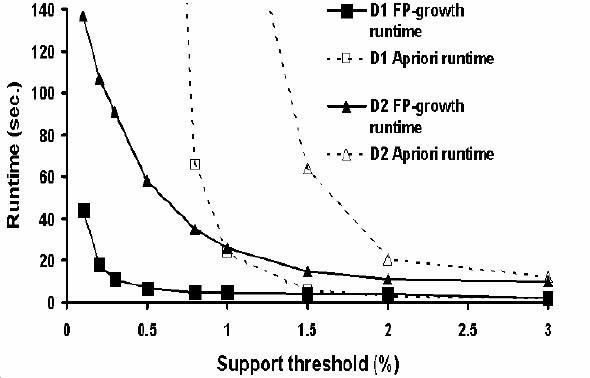
1. So sánh giữa Thời gian thực thi (tính bằng giây) của 2 thuật toán FP-Growth và Apriori trên 2 Database D1 và Database D2.
2. **So sánh giữa Thời gian thực thi (tính bằng giây) của 2 thuật toán FP-Growth và Apriori theo ngưỡng của độ hỗ trợ trên 2 Database D1 và Database D2.**
3. So sánh 2 thuật toán FP-Growth và Apriori theo ngưỡng độ tin cậy.
4. Mối quan hệ giữa 2 thuật toán FP-Growth và Apriori.

Câu hỏi 74: Cho đồ thị như hình vẽ. Từ đồ thị anh chị thấy điều gì?



1. **Với cùng ngưỡng của độ hỗ trợ, thời gian thức thi thuật toán FP-Growth luôn ít hơn thời gian thực thi thuật toán Apriori.**
2. Với cùng ngưỡng của độ hỗ trợ, thời gian thực thi của thuật toán FP-Growth luôn nhiều hơn thời gian thực thi của thuật toán Apriori.
3. Hai thuật toán FP-Growth và Apriori đều thức thi với thời gian rất nhỏ.
4. Thuật toán Apriori thực hiện nhanh hơn thuật toán FP-Growth.

Câu hỏi 75: Cho đồ thị như hình vẽ. Anh (chị) hãy cho biết nhận xét nào sau đây là sai?



1. Với cùng ngưỡng độ hỗ trợ, thời gian thức thi thuật toán FP-Growth luôn ít hơn thời gian thực thi thuật toán Apriori.
2. Thuật toán FP-Growth thực hiện nhanh hơn thuật toán Apriori.
3. **Thuật toán Apriori thực hiện nhanh hơn thuật toán FP-Growth.**
4. Khi số ngưỡng của độ hỗ trợ rất lớn, thời gian thực thi của 2 thuật toán FP-Growth và Apriori là tương đương.

Câu hỏi 76: Cho 3 điểm x, y, z. Anh (chị) hãy cho biết đ ộ đo khoảng cách d phải thỏa mãn các điều kiện nào?

1. d(x,y)>=0 ; d(x,y)<=d(x,z)+d(z,y)
2. **d(x,y)>=0 ; d(x,y)=d(y,x) ; d(x,x) =0 ; d(x,y)<=d(x,z)+d(z,y)**
3. d(x,y)>0 ; d(x,y)=d(y,x) ; d(x,y) =0 ; d(x,y)<=d(x,z)+d(z,y)
4. d(x,y)>=0 ; d(x,x) =0 ; d(x,y)<=d(x,z)+d(z,y)

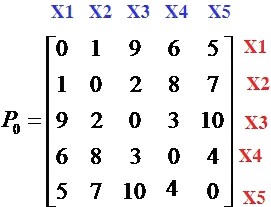
Câu hỏi 77: Anh (chị) hãy cho biết có những cách nào để chọn đại diện cho các cụm gì?

1. Đại diện siêu phẳng, đại diện điểm.
2. **Đại diện điểm, đại diện siêu phẳng và đại diện siêu cầu.**
3. Đại diện điểm, đại diện siêu cầu.
4. Đại diện siêu cầu, đại siêu phẳng.

Câu hỏi 78: Có 4 phần tử cần chia thành 2 cụm, mỗi cụm có ít nhất 1 phần tử. Anh (chị) hãy cho biết có bao nhiêu cách chia cụm?

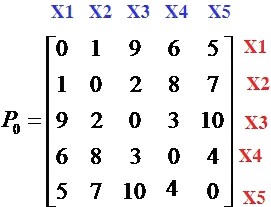
1. 15 cách.
2. **7 cách.**
3. 16 cách.
4. 1 cách.

Câu hỏi 79: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như sau:Sử dụng thuật toán liên kết đầy đủ (Complete Linkage). Bước đầu tiên ta gom x1, x2 hình vào cụm C1={x1, x2}. Anh (chị) hãy cho biết bước thứ 2 ta sẽ thực hiện thế nào?



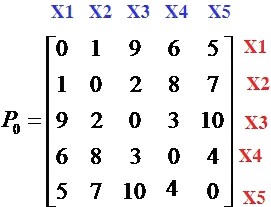
1. Gom x4 với C1.
2. **Gom x3 với x4.**
3. Gom x3 với C1.
4. Gom x3 với x5.

Câu hỏi 80: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình.Sử dụng thuật toán liên kết đơn (Single Linkage). Bước đầu tiên ta gom x1, x2, vào cụm C1, Anh (chị) hãy cho biết bước thứ 2 ta sẽ gom hai cụm nào với nhau?



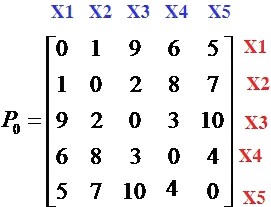
1. Gom x3 với x4
2. Gom x4 với C1
3. **Gom x3 với C1**
4. Gom x3 với x5

Câu hỏi 81: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình.Sử dụng thuật toán liên kết đơn (Single Linkage). Bước đầu tiên ta gom x1, x2 vào cụm C1, Anh (chị) hãy cho biết ma trận không tương tự P1 được sinh ra có dòng thứ nhất liệt kê từ trái sang phải là gì?



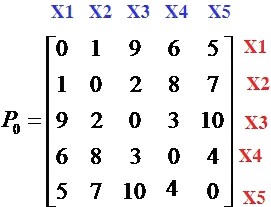
1. **0 ; 2 ; 6 ; 5**
2. 0 ; 9 ; 8 ; 7
3. 1 ; 9 ; 6 ; 5
4. 1 ; 2 ; 6 ; 6

Câu hỏi 82: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình.Sử dụng thuật toán liên kết đầy đủ (Complete Linkage). Bước đầu tiên ta gom x1, x2 vào cụm C1, Anh (chị) hãy cho biết ma trận không tương tự P1 được sinh ra có dòng thứ nhất liệt kê từ trái sang phải là gì?



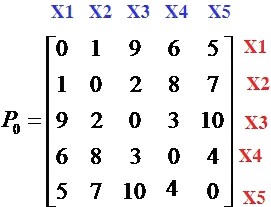
1. **0 ; 9 ; 8 ; 7**
2. 1 ; 9 ; 6 ; 5
3. 1 ; 2 ; 6 ; 5
4. 0 ; 2 ; 6 ; 5

Câu hỏi 83: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình.Sử dụng thuật toán liên kết đơn (Single Linkage). Bước đầu tiên gom x1 và x2 hình thành cụm C1={x1, x2}; bước 2 ta gom C1 và x3 hình thành cụm C2={x1, x2, x3}. Anh (chị) hãy cho biết bước thứ 3, ta sẽ gom thế nào?



1. Gom x4 với C1.
2. **Gom x4 với C2.**
3. Gom x4 với x5.
4. Gom x5 với C1.

Câu hỏi 84: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như hình sau:Sử dụng thuật toán liên kết đầy đủ (Complete Linkage). Bước đầu tiên gom x1 và x2 hình thành cụm C1={x1, x2}; bước 2 ta gom x3 và x4 hình thành cụm C2={x3, x4}. Anh (chị) hãy cho biết bước thứ 3, ta sẽ gom thế nào?



1. **Gom C1 với x5.**
2. Gom x5 với C2.
3. Gom x3 với x5.
4. Gom x3 với C2.

Câu hỏi 85: Cho CSDL giao dịch như bảng sau (Min\_Support = 3 (60%)):Sau khi loại bỏ các item không thỏa mãn Min\_Support, ta xây dựng lại bảng các item chỉ chứa các mục thường xuyên, trong mỗi giao dịch được sắp xếp theo độ giảm dần của độ thường xuyên. Anh (chị) hãy cho biết giao dịch thứ 2 sau khi được sắp xếp là gì?



1. a, b, c, d, e
2. c, a, b, m
3. f, c, a, b, m, o
4. **f, c, a, b, m, p**

Câu hỏi 86: Cho CSDL giao dịch như bảng sau (Min\_Support = 3 (60%)):Anh (chị) hãy cho biết nhóm item bị loại bỏ do không thỏa mãn Min\_Support là nhóm item nào?



1. **d, e, g, h, i, j, k, l, n, o, s**
2. c, a, b, m
3. a, b, c, d, e
4. f, c, a, b, m, p

Câu hỏi 87: Cho CSDL giao dịch như bảng sau (Min\_Support = 3 (60%)), Anh (Chị) hãy cho biết các item nào có độ hỗ trợ là 80%?



1. f, a và c
2. **f và c**
3. c và a
4. a, b và m

Câu hỏi 88. Cho CSDL giao dịch như bảng sau (Min\_Support = 3 (60%)), Anh (chị) hãy cho biết độ hỗ trợ của tập mục {f, c} là bao nhiêu?



1. 80%
2. **60%**
3. Giá trị khác
4. 100%

Câu hỏi 89: Cho CSDL giao dịch như bảng sau (Min\_Support = 3 (60%), Min\_Conference = 70%). Anh (chị) hãy cho biết luật kết hợp nào thỏa mãn độ hỗ trợ và độ tin cậy tối thiểu?



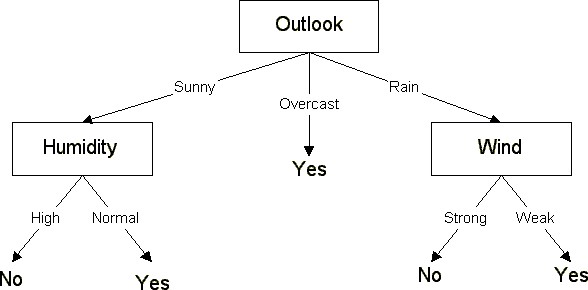
1. a--> p
2. **c--> f**
3. b-->c
4. m --> p

Câu hỏi 90: Cho CSDL giao dịch như bảng sau (Min\_Support = 3 (60%), Min\_Conference = 100%). Anh (chị) hãy cho biết trong các luật kết hợp sau, luật nào thỏa mãn độ hỗ trợ và độ tin cậy tối thiểu?



1. c--> a
2. f --> c
3. b-->c
4. **a--> c**

Câu hỏi 91: Cho cây quyết định như hình sau, Anh (chị) hãy cho biết có bao nhiêu luật sinh ra từ cây quyết định trên?



1. 2 luật
2. **5 luật**
3. 3 luật
4. 4 luật

Câu hỏi 92: Khoảng cách Ơclit giữa 2 cụm C1, C2 được định nghĩa là khoảng cách lớn nhất giữa phần tử x thuộc C1 và y thuộc C2 (x, y bất kỳ). Cho C1 gồm các phần tử x1=(1,0,0), x2=(2, 0, 0), x3=(0, 0, 0); C2 gồm các phần tử y1=(0, 2, 1), y2=(3, 4, 0 ). Anh (chị) hãy cho biết khoảng cách d giữa cụm C1 và C2 có giá trị là gì?

1. d=3
2. **d=5**
3. Giá trị khác
4. d=4

Câu hỏi 93: Khoảng cách Ơclit giữa 2 cụm C1, C2 được định nghĩa là khoảng cách lớn nhất giữa phần tử x thuộc C1 và y thuộc C2 (x, y bất kỳ). Cho C1 gồm các phần tử x1=(1, 0, 0), x2=(2, 0, 0), x3=(0, 0, 0); C2 gồm các phần tử y1=(0, 2, 1), y2=(3, 4, 0 ). Anh (chị) hãy cho biết khoảng cách giữa cụm C1 và C2 là khoảng cách giữa 2 phần tử nào sau đây?

1. x3 và y2
2. x2 và y2
3. x1 và y1
4. **x3 và y2**

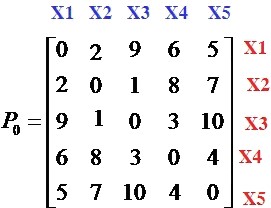
Câu hỏi 94: Khoảng cách Ơclit giữa 2 cụm C1, C2 được định nghĩa là khoảng cách bé nhất giữa phần tử x thuộc C1 và y thuộc C2 (x, y bất kỳ). Cho C1 gồm các phần tử x1=(1,1,0), x2=(2, 0, 0), x3=(0, 0, 0); C2 gồm các phần tử y1=(0, 2, 1), y2=(2, 2, 0 ). Anh (chị) hãy cho biết khoảng cách d giữa cụm C1 và C2 có giá trị là bao nhiêu?

1. **d=1.414**
2. d=2.000
3. d=4.353
4. d=1.021

Câu hỏi 95: Khoảng cách Ơclit giữa 2 cụm C1, C2 được định nghĩa là khoảng cách giữa các vector đại diện của cụm. Giả sử vector đại diện của cụm là vector trung bình của cụm. Cho C1 gồm x1=(1,0), x2=(3,2) ; C2 gồm y1=(3, 2) , y2=(7, 8). Anh (chị) hãy cho biết khoảng cách d giữa cụm C1 và C2 có giá trị là bao nhiêu ?

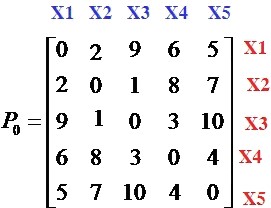
1. d = 1
2. d=2
3. d=6
4. **d=5**

Câu hỏi 96: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như sau:Sử dụng thuật toán liên kết đơn (Single Linkage). Bước đầu tiên ta gom x2, x3 vào cụm C1={x2, x3}, Anh (chị) hãy cho biết dòng thứ 2 của ma trận P1 là gì?



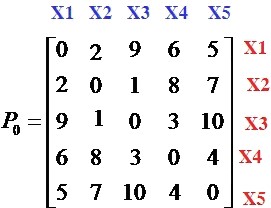
1. 0; 2; 5; 6
2. **2; 0; 3; 7**
3. 9 ;1; 0 ;3
4. 2; 0; 1; 8

Câu hỏi 97: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như sau:Sử dụng thuật toán liên kết đơn (Single Linkage). Bước đầu tiên ta gom x2, x3 vào cụm C1={x2, x3}, bước 2 ta sẽ gom x1 và cụm C1 được C2={x1, x2, x3}, bước 3 ta sẽ gom x4 vào cụm C2 hình thành cụm C3= {x1, x2, x3, x4}. Anh (chị) hãy cho biết ngưỡng để hình thành cụm C3 là bao nhiêu?



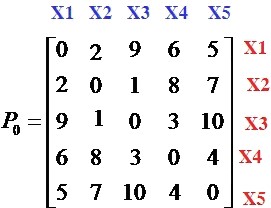
1. **3**
2. 5
3. 0
4. 10

Câu hỏi 98: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như sau:Sử dụng thuật toán liên kết đơn (Single Linkage) để phân cụm tập dữ liệu trên. Anh (chị) hãy cho biết tất cả các phần tử được gom vào 1 cụm ở ngưỡng bằng bao nhiêu?



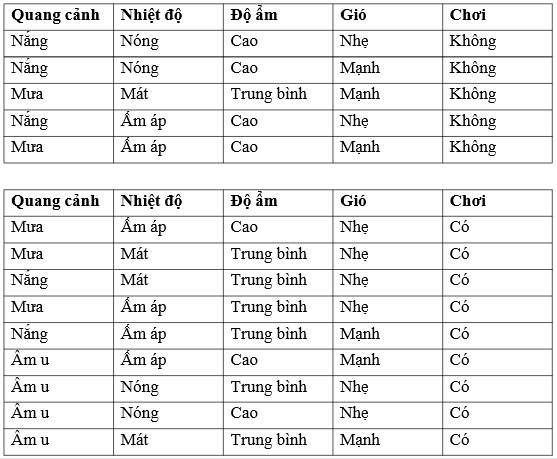
1. 10
2. 5
3. 3
4. **4**

Câu hỏi 99: Cho tập dữ liệu X={x1, x2, x3, x4, x5} và ma trận không tương tự như sau:Sử dụng thuật toán liên kết đơn (Single Linkage) để xây dựng sơ đồ ngưỡng. Anh (chị) hãy cho biết cắt sơ đồ ngưỡng tại ngưỡng 3.5 thì hình thành bao nhiêu cụm?



1. 1 cụm
2. **2 cụm**
3. 3 cụm
4. 4 cụm

Câu hỏi 100:Cho tập mẫu X = và phương pháp phân lớp theo mô hình xác suất Bayes.Tập mẫu X được gán vào lớp C1 (Chơi = ‘Không’) với xác suất là bao nhiêu?



1. **P(C1)\*P(X,C1) = 0.005.**
2. P(C1)\*P(X,C1) = 0.021.
3. P(C1)\*P(X,C1) = 0.01.
4. Giá trị khác.