Câu 1: Kỹ thuật phân cụm nào yêu cầu phương pháp tiếp cận hợp nhất

A. Phân vùng (Partitional)

B. Naive Bayes

C. phân cấp (Hierarchical)

D. Tát cá déu sai

Câu 3: Thư viện nào được sử dụng để kiểm tra tính chính xác của một mô hình tiên đoán trong

sklearn?

A. validate

B. check_score

C. test accuracy

D. metrics

Câu 4: Trong một số trường hợp, các công ty viễn thông mong muốn phân khúc khách hàng của họ thành các nhóm riêng biệt để gin vụ đải đăng ký phù hợp và có liên quan. Đây có thể được coi là một ví dụ về phương pháp nào sau đây

A. Học có giám sát (Supervised learning)

B. Học không giám sát (Unsupervised learning)

C. Duong rang cua (Serration)

D. Khai thác dữ liệu (Data extraction)

Câu 5. Anh chị được cung cấp dữ liệu về hoạt động địa chấn ở Nhật Bản và anh chị muốn dự đoán cường độ của trận động đất sắp tới. Đây có thể được coi là một ví dụ về phương pháp nào sau đây?

A. Duong rang cura (Serration)

B. Học không giám sát (Unsupervised learning)

C. Học có giám sát (Supervised leaming)

Đ. Giảm kích thước (Dimensionality reduction)

Câu 6: Một lớp thuật toán học tập cố gắng tìm cách phân loại tối ưu của một tập hợp các ví dụ bằng cách sử dụng lý thuyết xác suất được đặt tên là

Câu 7: Đặc tính nào của "Big data" liên quan nhiều nhất đến lĩnh vực khoa học dữ liệu?

A. Dung lượng lớn (Volume)

C. Tính xác thực (Veracity)

B. Chuyển động không ngừng (Velocity)

D. Su da dang (Variety)

Câu 8: Phương pháp nào sau đây đề cập đến bài toán tìm các mẫu (hoặc cấu trúc) trừu tượng trong dữ liệu chưa được gần nhân?

Câu 10: Anh (chị) hãy cho biết thuật toán Apriori có nhược điểm chính là gì?

A. Kết quả của thuật toàn không ứng dụng được trong các bài toán thực tế.

B. Không tìm được các tập thường xuyên.

C. Tốn nhiều bộ nhớ và thời gian. Không thích hợp với các mẫu lớn. Chi phí để duyệt CSDL nhiều.

D. Thuật toán quá phức tạp, khó hiểu

Câu 12: Thuật ngữ nào sau đây để cấp đến việc truy vấn dữ liệu văn bản phi cấu trúc

A. Truy vấn thông tin (Information retrieval)

B. Truy cấp thông tin (Information access)

C. Theo từng thông tin (Information manipulation)

Đ. Cập nhật thông tin (Information update)

Câu 13: Ngay sau khi thu thập được dữ liệu, nhà khoa học dữ liệu cần phải làm gì?

A. Tích hợp dữ liệu (Data Integration)

C. Làm sạch dữ liệu (Data Cleansing)

B. Mô phỏng dữ liệu (Data Replication)

D. Tất cả đều sai

Câu 15: Trong thuật toán K-mean, sau khi gần các đối tượng vào k cụm ta cần làm gi

A. Trên các cụm hại với nhau để số cụm sinh ra là ít nhất

B. Tìm một số phân tử đại diện của cụm

C. Tinh lại tâm của các cụm

D. Tỉnh khoảng cách giữa các phân tử trong cụm

Câu 16: Giả sử chúng ta đang dự đoàn số con trai mới sinh dựa trên quy mô dân số của nước ta,

số con trai mới sinh là

A. Observation (quan sát)

C. Attrbute (thuộc tỉnh)

B. Feature (tính năng)

D. Outcome (kết quả)

Câu 18: Anh (chỉ) 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?

A. Trích chọn dữ liệu, tiền xử lý dữ liệu, biến đổi dữ liệu, khai phá dữ liệu, đánh giả và biểu diễn tri thức.

B. Tiền xử lý dữ liệu, biến đổi dữ liệu, khai phá dữ liệu, đánh giả và biểu diễn tri thức

C. Trích chọn dữ liệu, tiền xử lý dữ liệu, biến đổi dữ liệu, khai phá dữ liệu, khai phá luật kết

hop.

Đ. Tiền xử lý dữ liệu, phân lớp, phân cụm, đánh giá và biểu diễn tri thức

Câu 19: Anh (chị) hãy cho biết trong thuật toán phân cụm k-mean, sau khi chọn được k điểm làm tâm, phân tử x sẽ được gắn vào cụm C nếu thỏa mãn điều kiện nào

A. Khoảng cách tử x đến tìm cụm C là nhỏ nhất

B. Khoảng cách từ x đến tắm cụm C là lớn nhất

C. Khoảng cách từ x đến tâm cụm C bằng k

D. Khoảng cách từ x đến tâm cụm C = 0