

**ĐỀ 1**

(Sinh viên được sử dụng 01 tờ A4 tài liệu, đề thi gồm 02 trang)

HỌ VÀ TÊN SV: ..... MSSV:.....STT: ..... PHÒNG THI:.....	<b><u>CÁN BỘ COI THI</u></b>
--	------------------------------

**Câu 1 (2.0 điểm)** Sinh viên chọn **MỘT** trong các câu sau:

1. Trình bày tóm tắt mục tiêu và hoạt động của các bước trong quá trình Khám phá tri thức (Knowledge Discovery from Data – KDD).
2. Phân biệt giữa bài toán phân lớp nhị phân (binary classification), phân lớp đa lớp (multi-class classification). Cho ví dụ minh họa.
3. Trình bày một phương pháp để đánh giá mô hình phân lớp. Cho ví dụ minh họa.

**Câu 2 (6.0 điểm)** Cho CSDL về *Thông tin bệnh lý về tim mạch*, chi tiết trong bảng sau.

	<b>Độ tuổi (ĐT)</b>	<b>Giới tính (GT)</b>	<b>Điện tâm đồ (ECG)</b>	<b>Mức độ đau ngực (ĐN)</b>	<b>Kết quả (KQ)</b>
1	Trên 60 tuổi	Nam	Bất thường	Đau vừa	Có
2	41-60 tuổi	Nữ	Bình thường	Đau nhẹ	Không
3	Trên 60 tuổi	Nam	Bất thường	Đau nặng	Có
4	Dưới 40 tuổi	Nữ	Bình thường	Đau nhẹ	Không
5	Dưới 40 tuổi	Nam	Phì đại	Không đau	Có
6	Trên 60 tuổi	Nữ	Bất thường	Đau vừa	Có
7	41-60 tuổi	Nam	Bất thường	Đau nhẹ	Không
8	41-60 tuổi	Nữ	Phì đại	Đau vừa	Có
9	Trên 60 tuổi	Nam	Bất thường	Đau nặng	Có
10	41-60 tuổi	Nữ	Bình thường	Đau vừa	Không

**Lưu ý:**

- *Kết quả (KQ)* là thuộc tính quyết định. Các thuộc tính còn lại là thuộc tính điều kiện.
- Sinh viên có thể dùng từ viết tắt của thuộc tính trong khi làm bài.
- Các kết quả tính toán làm tròn đến **03** chữ số thập phân.

1. Áp dụng thuật toán Apriori, tìm tập phổ biến thỏa ngưỡng  $minsup = 35\%$ . Chọn 1 tập phổ biến tối đại, liệt kê một luật kết hợp có dạng  $X \wedge Y \rightarrow KQ$  thỏa  $minconf = 90\%$ , với X và Y là các thuộc tính điều kiện của CSDL trên **(2.0đ)**
2. Cho  $B = \{Giới tính, Điện tâm đồ\}$ ,  $X = \{1, 3, 5, 6, 8, 9\}$  (tập các mẫu có giá trị *Kết quả* = “Có”). Sử dụng tập thô tính: xấp xỉ trên, xấp xỉ dưới và hệ số xấp xỉ. **(0.75đ)**
3. Xác định nút gốc của cây quyết định sử dụng *Chỉ số Gini*. **(2.0đ)**
4. Sử dụng công thức *Naïve Bayes* có làm tròn *Laplace* để phân lớp mẫu sau: **(1.25đ)**  
 $X = \{Độ tuổi = “Trên 60 tuổi”, Giới tính = “Nữ”, Điện tâm đồ = “Phì đại”, Mức độ đau ngực = “Đau vừa”\}$

**Câu 3 (2.0 điểm)**

1. Hãy biến đổi miền giá trị của 3 thuộc tính *Độ tuổi (ĐT)*, *Điện tâm đồ (ECG)*, *Mức độ đau ngực (ĐN)* ở câu 2 về dạng số theo nguyên tắc sau: **(0.5đ)**
  - *Độ tuổi (ĐT)* có các giá trị: Dưới 40 tuổi: 1; 41-60 tuổi: 2; Trên 60 tuổi: 3
  - *Điện tâm đồ (ECG)* có các giá trị: Bình thường: 1; Bất thường: 2; Phì đại: 3
  - *Mức độ đau ngực (ĐN)* có các giá trị: Không đau: 1; Đau nhẹ: 2; Đau vừa: 3; Đau nặng: 4
2. Dựa vào 03 thuộc tính trên (sau khi biến đổi miền giá trị), áp dụng thuật toán K-means với khoảng cách Euclide để gom cụm 10 bệnh nhân ở câu 2 vào **02 cụm**, biết rằng trọng tâm ban đầu của 02 cụm là C1 (2.8; 2.2; 3.4) và C2 (1.6; 1.6; 2.0). **(1.5đ)**

**Lưu ý:**

- Tọa độ trọng tâm theo thứ tự: *Độ tuổi*, *Điện tâm đồ*, *Mức độ đau ngực*.
- Thực hiện các bước để tìm được cụm cho các bệnh nhân.
- Các kết quả tính toán làm tròn đến **02** chữ số thập phân.

---

**Khoa/ Bộ môn duyệt đề**

**TM. Giảng viên ra đề**

**Bảng ma trận đáp ứng chuẩn đầu ra.**

Câu hỏi	CĐRMH (*)
1	G2.1, G3.1, G4.1, G4.2
2	G3.1, G3.2, G4.1
3	G3.1, G3.2, G4.1

(\*) Chi tiết các chuẩn đầu ra môn học (CĐRMH) xem trong Đề cương môn học

**ĐỀ 2**

(Sinh viên được sử dụng 01 tờ A4 tài liệu, đề thi gồm 02 trang)

HỌ VÀ TÊN SV: ..... MSSV:.....STT: ..... PHÒNG THI:.....	<b><u>CÁN BỘ COI THI</u></b>
--	------------------------------

**Câu 1 (2.0 điểm) Sinh viên chọn MỘT trong các câu sau:**

1. Trình bày tóm tắt mục tiêu và hoạt động của các bước trong quá trình Khám phá tri thức (Knowledge Discovery from Data – KDD).
2. Phân biệt giữa bài toán phân lớp nhị phân (binary classification), phân lớp đa lớp (multi-class classification). Cho ví dụ minh họa.
3. Trình bày một phương pháp để đánh giá mô hình phân lớp. Cho ví dụ minh họa.

**Câu 2 (6.0 điểm) Cho CSDL về Thông tin bệnh lý về tim mạch, chi tiết trong bảng sau.**

	<b>Độ tuổi (ĐT)</b>	<b>Giới tính (GT)</b>	<b>Điện tâm đồ (ECG)</b>	<b>Mức độ đau ngực (ĐN)</b>	<b>Kết quả (KQ)</b>
1	Dưới 40 tuổi	Nam	Phì đại	Không đau	Có
2	Trên 60 tuổi	Nữ	Bất thường	Đau vừa	Có
3	Trên 60 tuổi	Nam	Bất thường	Đau vừa	Có
4	41-60 tuổi	Nữ	Phì đại	Đau vừa	Có
5	41-60 tuổi	Nam	Bất thường	Đau nhẹ	Không
6	Dưới 40 tuổi	Nữ	Bình thường	Đau nhẹ	Không
7	Trên 60 tuổi	Nam	Bất thường	Đau nặng	Có
8	41-60 tuổi	Nữ	Bình thường	Đau vừa	Không
9	Trên 60 tuổi	Nam	Bất thường	Đau nặng	Có
10	41-60 tuổi	Nữ	Bình thường	Đau nhẹ	Không

**Lưu ý:**

- *Kết quả (KQ)* là thuộc tính quyết định. Các thuộc tính còn lại là thuộc tính điều kiện.
- Sinh viên có thể dùng từ viết tắt của thuộc tính trong khi làm bài.
- Các kết quả tính toán làm tròn đến **03** chữ số thập phân.

1. Áp dụng thuật toán Apriori, tìm tập phổ biến thỏa ngưỡng  $minsup = 40\%$ . Chọn 1 tập phổ biến tối đại, liệt kê một luật kết hợp có dạng  $X \wedge Y \rightarrow KQ$  thỏa  $minconf = 85\%$ , với X và Y là các thuộc tính điều kiện của CSDL trên **(2.0đ)**
2. Cho  $B = \{Giới tính, Điện tâm đồ\}$ ,  $X = \{5, 6, 8, 10\}$  (tập các mẫu có giá trị Kết quả = “Không”). Sử dụng tập thô tính: xấp xỉ trên, xấp xỉ dưới và hệ số xấp xỉ. **(0.75đ)**
3. Xác định nút gốc của cây quyết định sử dụng Chỉ số Gini. **(2.0đ)**
4. Sử dụng công thức Naïve Bayes có làm tròn Laplace để phân lớp mẫu sau: **(1.25đ)**  
 $X = \{Độ tuổi = “41-60 tuổi”, Giới tính = “Nam”, Điện tâm đồ = “Bình thường”, Mức độ đau ngực = “Đau nặng”\}$

**Câu 3 (2.0 điểm)**

1. Hãy biến đổi miền giá trị của 3 thuộc tính *Độ tuổi (ĐT)*, *Điện tâm đồ (ECG)*, *Mức độ đau ngực (ĐN)* ở câu 2 về dạng số theo nguyên tắc sau: **(0.5đ)**
  - *Độ tuổi (ĐT)* có các giá trị: Dưới 40 tuổi: 1; 41-60 tuổi: 2; Trên 60 tuổi: 3
  - *Điện tâm đồ (ECG)* có các giá trị: Bình thường: 1; Bất thường: 2; Phi đại: 3
  - *Mức độ đau ngực (ĐN)* có các giá trị: Không đau: 1; Đau nhẹ: 2; Đau vừa: 3; Đau nặng: 4
2. Dựa vào 03 thuộc tính trên (sau khi biến đổi miền giá trị), áp dụng thuật toán K-means với khoảng cách Euclide để gom cụm 10 bệnh nhân ở câu 2 vào **02 cụm**, biết rằng trọng tâm ban đầu của 02 cụm là C1 (1.6, 1.6, 2.0) và C2 (2.8, 2.2, 3.4). **(1.5đ)**

**Lưu ý:**

- Tọa độ trọng tâm theo thứ tự: *Độ tuổi*, *Điện tâm đồ*, *Mức độ đau ngực*.
- Thực hiện các bước để tìm được cụm cho các bệnh nhân.
- Các kết quả tính toán làm tròn đến **02** chữ số thập phân.

---

**Khoa/ Bộ môn duyệt đề**

**TM. Giảng viên ra đề**

**Bảng ma trận đáp ứng chuẩn đầu ra.**

Câu hỏi	CĐRMH (*)
1	G2.1, G3.1, G4.1, G4.2
2	G3.1, G3.2, G4.1
3	G3.1, G3.2, G4.1

(\*) Chi tiết các chuẩn đầu ra môn học (CĐRMH) xem trong Đề cương môn học