

Câu 1. Khảo sát nhu cầu X (C = có, K = không) đối với một sản phẩm theo vùng miền Y (T = thành thị, N = nông thôn, M = miền núi), ta có mẫu cỡ 200 như sau:

$X \backslash Y$	T	N	M
C	26	48	24
K	51	43	8

Sử dụng tiêu chuẩn χ^2 , với mức ý nghĩa 5%, hãy kiểm định xem X và Y độc lập nhau không?

Câu 2. Cho cơ sở dữ liệu giao dịch đơn giản X :

X :	TID	Items
	T01	A, B, C, D
	T02	A, C, D, F
	T03	C, D, E, G, A
	T04	A, D, F, B
	T05	B, C, G
	T06	D, F, G
	T07	A, B, G
	T08	C, D, F, G

Sử dụng ngưỡng giá 25% và ngưỡng tin cậy 60%, tìm:

- Tất cả các tập mục thường xuyên trong cơ sở dữ liệu X
- Các luật kết hợp mạnh cho cơ sở dữ liệu X

Câu 3. Tìm cây quyết định cho tập dữ liệu huấn luyện Y :

Y :	A	B	C	Lớp
	15	1	y	C_1
	20	3	n	C_2
	25	2	y	C_1
	30	4	y	C_1
	35	2	n	C_2
	25	4	y	C_1
	15	2	n	C_2
	20	3	n	C_2