

# TRƯỜNG ĐẠI HỌC BÁCH KHOA ĐÀ NẴNG

## ĐỀ THI KẾT THÚC HỌC PHẦN

### MÔN HỌC: KHOA HỌC DỮ LIỆU NÂNG CAO

KHÓA 2022-2023-Nhóm học phần: 19NH11

**Câu 1 ( 3 điểm ):** Anh/ chị được cung cấp một tập dữ liệu về các đặc điểm của khách hàng như tuổi, thu nhập và tình trạng độc thân cùng với quyết định mua hàng (Có/Không).

Khách hàng	Tuổi	Thu nhập	Tình trạng độc thân	Quyết định mua laptop
1	Trẻ	Thấp	Có	Không
2	Trẻ	Thấp	Không	Không
3	Trung niên	Trung bình	Có	Có
4	Cao tuổi	Cao	Không	Có
5	Cao tuổi	Cao	Có	Không
6	Trung niên	Cao	Có	Có
7	Trẻ	Trung bình	Không	Có
8	Trẻ	Cao	Có	Có
9	Trung niên	Thấp	Không	Không
10	Cao tuổi	Trung bình	Có	Có
11	Trung niên	Cao	Không	Có
12	Trẻ	Thấp	Có	Không
13	Cao tuổi	Thấp	Không	Không
14	Trung niên	Trung bình	Có	Không
15	Trẻ	Cao	Không	Có

Dùng thuật toán học máy Native Baye để xác định rằng một người có **độ tuổi trung niên có thu nhập cao và đang độc thân** sẽ mua laptop không ? Trình bày chi tiết các bước giải.

**Câu 2 ( 3 điểm ) :** Cho tập điểm

$$X1(4,1) ; X2(5,1) ; X3(5,2) ; X4(1,4) ; \\ X5(1,5) ; X6(2,4) ; X7(2,5)$$

Dùng K-Mean để gom nhóm (K=2). Dùng hàm tính khoảng cách Manhattan.

**Câu 3: ( 4 điểm):** Một siêu thị muốn tìm các quy tắc kết hợp trong giỏ hàng mua sắm của khách hàng.

Tập giao dịch được cho như sau:

Giao dịch	Mục hàng
T1	Sữa, Bánh mì, Bơ
T2	Bánh mì, Bơ
T3	Sữa, Trứng, Bánh mì
T4	Sữa, Bánh mì
T5	Trứng, Sữa, Bơ

Yêu cầu: • Vẽ cây **FP-Tree** với ngưỡng hỗ trợ tối thiểu là 2 (tức 40%). Từ đó, trích xuất **ít nhất 2 luật kết hợp** có:

- **Hỗ trợ  $\geq 50\%$**
- **Độ tin cậy  $\geq 70\%$ .**