# **Đề tài: Viết chương trình phân nhóm người theo thu nhập và chi tiêu hàng tháng của họ. Xuất ra khoảng trị thu nhập và chi tiêu hàng tháng của mỗi nhóm.**

Thành viên:

Trần Ngọc Khoa 15110064

Đào Văn Toản 15110142

# **Usecase diagram:**

k-means

Nhập file input, #k

# **Giải thuật k-means clustering:**

**Đầu vào:** Dữ liệu X và số lượng cluster cần tìm K.

**Đầu ra:** Các center M và label vector cho từng điểm dữ liệu Y.

**Các bước giải thuật:**

1. Chọn K điểm bất kỳ làm các center ban đầu.
2. Phân mỗi điểm dữ liệu vào cluster có center gần nó nhất.
3. Nếu việc gán dữ liệu vào từng cluster ở bước 2 không thay đổi so với vòng lặp trước nó (center không thay đổi) thì ta dừng thuật toán.
4. Cập nhật center cho từng cluster bằng cách lấy trung bình cộng của tất các các điểm dữ liệu đã được gán vào cluster đó sau bước 2.
5. Quay lại bước 2.

# **Tính tích vô hướng của 2 vector:**

Python

def dot\_product(a, b):  
 return sum([a[i]\*b[i] for i in range(len(a))])

# **Cấu trúc dữ liệu sử dụng:**

**Input data:** Ma trận m\*n với m là số dòng sữ liệu đầu vào và n là số cột của một dòng dữ liệu.

*Hiện thực trong python: list m phần tử, mỗi phần tử là một list n phần tử:*

*[[x00, x01, x02],*

*[x10, x11, x12],*

*[x20, x21, x22]]*

**Tính khoảng cách:** Vector có số phần tử bằng số cột của một dòng dữ liệu.

*Hiện thực trong python: list n phần tử.*