**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN BÁO CÁO THỰC HÀNH**

KHOA HỆ THỐNG THÔNG TIN Môn: Khai thác dữ liệu

**BÀI THỰC HÀNH SỐ 5**

Giảng viên hướng dẫn: ThS. Nguyễn Hồ Duy Trí

Sinh viên thực hiện: Phạm Thiện Bảo

MSSV: 20521107

Lớp: IS252.N22.HTCL

**Bài làm**

**IV. Thực hành:**

# Câu 1: Trong quá trình thống kê doanh thu, một công ty kinh doanh chuỗi cửa hàng pizza phát hiện những vị trí có nhu cầu cao nhưng chưa có cửa hàng trong khu vực. Tọa độ tương đối của những vị trí tiềm năng đó được cho trong bảng sau:

A picture containing text, screenshot, number, font

Description automatically generated

Tìm ra vị trí đặt 3 cửa hàng sao cho thuận lợi nhất và chỉ ra cụm khách hàng của từng cửa hàng với những yêu cầu cụ thể sau:

## a) Sử dụng thuật toán k-Means

Khởi tạo ma trận phân hoạch:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | VT 1 | VT 2 | VT 3 | VT 4 | VT 5 | VT 6 | VT 7 | VT 8 | VT 9 |
| Cụm 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cụm 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cụm 3 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Vector trọng tâm của các cụm lúc này là:

+ Cụm 1: V1 = (8, 4, 200)

+ Cụm 2: V2 = (8, 6, 350)

+ Cụm 3: V3 = (11.28, 6, 388.57)

Khoảng cách Euclide từ các bức tranh đến lần lượt các cụm là:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cụm 1 | Cụm 2 | Cụm 3 |
| Vị trí 1 | 0 | 150.01 | 188.61 |
| Vị trí 2 | 150.01 | 0 | 38.71 |
| Vị trí 3 | 450.01 | 300 | 261.44 |
| Vị trí 4 | 200.01 | 50.05 | 11.54 |
| Vị trí 5 | 120.03 | 30.21 | 68.6 |
| Vị trí 6 | 50.25 | 100.06 | 138.58 |
| Vị trí 7 | 400.02 | 250.03 | 211.43 |
| Vị trí 8 | 100.12 | 50.16 | 88.58 |
| Vị trí 9 | 6.08 | 150.12 | 188.59 |

Ma trận phân hoạch các điểm thuộc cụm:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | VT 1 | VT 2 | VT 3 | VT 4 | VT 5 | VT 6 | VT 7 | VT 8 | VT 9 |
| Cụm 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Cụm 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| Cụm 3 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |

Vector trọng tâm của các cụm lúc này là:

+ Cụm 1: V1 = (11, 5.67, 216.67)

+ Cụm 2: V2 = (10.33, 5.67, 323.33)

+ Cụm 3: V3 = (10.33, 6, 550)

Khoảng cách Euclide từ các bức tranh đến lần lượt các cụm là:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cụm 1 | Cụm 2 | Cụm 3 |
| Vị trí 1 | 17.02 | 123.36 | 350.01 |
| Vị trí 2 | 133.36 | 26.77 | 200.01 |
| Vị trí 3 | 433.34 | 326.67 | 100.01 |
| Vị trí 4 | 188.33 | 76.67 | 150 |
| Vị trí 5 | 103.34 | 3.78 | 230 |
| Vị trí 6 | 33.4 | 73.37 | 300 |
| Vị trí 7 | 383.33 | 276.67 | 50.02 |
| Vị trí 8 | 83.35 | 23.43 | 250 |
| Vị trí 9 | 16.95 | 123.39 | 350.02 |

Ma trận phân hoạch các điểm thuộc cụm:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | VT 1 | VT 2 | VT 3 | VT 4 | VT 5 | VT 6 | VT 7 | VT 8 | VT 9 |
| Cụm 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Cụm 2 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| Cụm 3 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |

Vector trọng tâm của các cụm lúc này là:

+ Cụm 1: V1 = (11, 5.67, 216.67)

+ Cụm 2: V2 = (10.25, 5.5, 342.5)

+ Cụm 3: V3 = (10.5, 6.5, 625)

Khoảng cách Euclide từ các bức tranh đến lần lượt các cụm là:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cụm 1 | Cụm 2 | Cụm 3 |
| Vị trí 1 | 17.02 | 142.52 | 425.01 |
| Vị trí 2 | 133.36 | 7.85 | 275.01 |
| Vị trí 3 | 433.34 | 307.5 | 25.05 |
| Vị trí 4 | 188.33 | 57.5 | 225 |
| Vị trí 5 | 103.34 | 22.56 | 305.01 |
| Vị trí 6 | 33.4 | 92.54 | 375 |
| Vị trí 7 | 383.33 | 257.5 | 25.05 |
| Vị trí 8 | 83.35 | 42.57 | 325 |
| Vị trí 9 | 16.95 | 142.55 | 425.02 |

Như vậy các điểm thuộc cụm không thay đổi, thuật toán dừng lại, với kết quả sau:

+ Cụm 1: Vị trí 1, Vị trí 6, Vị trí 9

+ Cụm 2: Vị trí 2, Vị trí 4, Vị trí 5, Vị trí 8

+ Cụm 3: Vị trí 3, Vị trí 7

## b) Sử dụng mạng Kohonen với các thông số: 𝑒𝑝𝑜𝑐ℎ𝑠 = 10, 𝑅 = 0 và 𝛼 = 0,8.

## c) So sánh kết quả thu được từ thuật toán k-Means và mạng Kohonen với nhau.

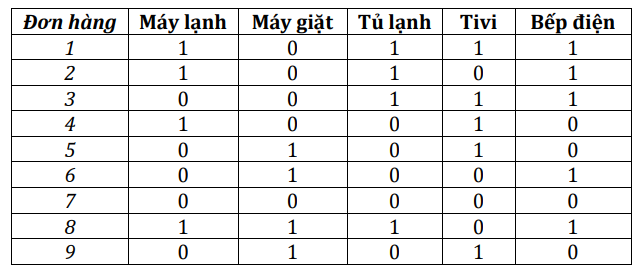
## d) Một vị trí tiềm năng mới xuất hiện tại tọa độ:

A picture containing text, screenshot, font, number

Description automatically generated

Sinh viên hãy giúp công ty xác định cửa hàng nào trong 3 cửa hàng trên sẽ phục vụ cho vị trí này tốt nhất.

# Câu 2: (Cơ bản) Một website thương mại điện tử chuyên kinh doanh thiết bị gia dụng thu thập được những đơn hàng như sau:



Công ty dự định xây dựng một hệ thống khuyến nghị dựa trên việc gom cụm những đơn hàng trên thành 3 nhóm có độ tương đồng với nhau. Từ đó gợi ý cho việc mua hàng của khách hàng dựa trên sự tương đồng này. Sinh viên hãy tham gia xây dựng hệ thống khuyến nghị bằng cách thực hiện những yêu cầu sau:

## a) Sử dụng thuật toán k-Means.

Khởi tạo ma trận phân hoạch:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | VT 1 | VT 2 | VT 3 | VT 4 | VT 5 | VT 6 | VT 7 | VT 8 | VT 9 |
| Cụm 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cụm 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cụm 3 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Vector trọng tâm của các cụm lúc này là:

## b) Sử dụng mạng Kohonen với các thông số: 𝑒𝑝𝑜𝑐ℎ𝑠 = 10, 𝑅 = 0 và 𝛼 = 0,8.

## c) So sánh kết quả thu được từ thuật toán k-Means và mạng Kohonen với nhau.

## d) Một khách đang thực hiện đặt hàng trên hệ thống đã mua Máy lạnh, tủ lạnh. Theo em những mặt hàng nào nên được khuyến nghị dựa theo kết quả đã tính được từ các câu trên.