**TRƯỜNG ĐẠI HỌC SƯ PHẠM HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN**

**---\*\*\*---**

****

**HỌC PHẦN KHAI PHÁ DỮ LIỆU**

**BÁO CÁO BÀI TẬP LỚN**

**DỰ ĐOÁN BỆNH TIỂU ĐƯỜNG**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn:** | **TS. Lê Thị Tú Kiên** |
| **Sinh viên thực hiện:** | **715105184\_Đỗ Đức Quang Phương** |

***Hà Nội, 2024***

# MỤC LỤC

[MỤC LỤC 1](#_Toc174907985)

[1. Khảo sát dữ liệu 2](#_Toc174907986)

[1.1. Tổng quan về dữ liệu 2](#_Toc174907987)

[1.2. Thống kê: 4](#_Toc174907988)

[1.3. Biểu đồ cột: 4](#_Toc174907989)

[1.4. Biểu đồ ploxbot: 7](#_Toc174907990)

[1.5. Biểu đồ histogram 9](#_Toc174907991)

[2. Phân lớp dữ liệu 13](#_Toc174907992)

[2.1. Mô tả cách thực hiện 13](#_Toc174907993)

[2.1.1. Decision tree 16](#_Toc174907994)

[2.1.2. Naive bayes 46](#_Toc174907995)

[2.1.3. Support Vector Machines 49](#_Toc174907996)

[2.2. Phân tích và so sánh 58](#_Toc174907997)

[3. Phân cụm dữ liệu 58](#_Toc174907998)

[3.1. Mô tả cách thực hiện 58](#_Toc174907999)

[3.1.1. Kmeans 61](#_Toc174908000)

[3.1.2. Kmedoids 67](#_Toc174908001)

[3.2. Phân tích và so sánh 77](#_Toc174908002)

[3.2.1. Kmeans 77](#_Toc174908003)

[3.2.2. Kmedoids 78](#_Toc174908004)

# Khảo sát dữ liệu

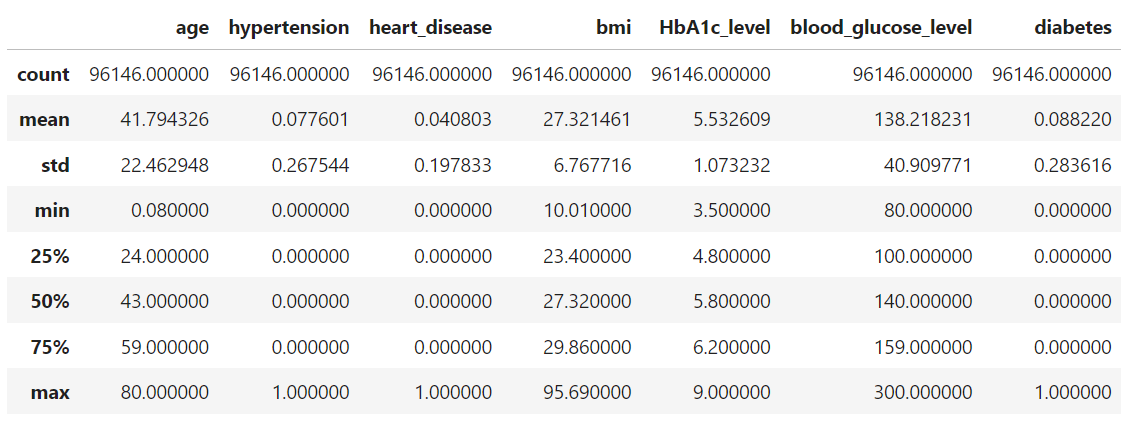
Diabetes prediction dataset là một bộ sưu tập dữ liệu y tế và dân số học từ bệnh nhân, cùng với tình trạng tiểu đường của họ (dương tính hoặc âm tính). Dữ liệu bao gồm các đặc điểm như tuổi, giới tính, chỉ số khối cơ thể (BMI), huyết áp cao, bệnh tim mạch, lịch sử hút thuốc, mức độ HbA1c và mức độ glucose trong máu. Tập dữ liệu này có thể được sử dụng để xây dựng các mô hình học máy để dự đoán tiểu đường ở bệnh nhân dựa trên lịch sử y tế và thông tin dân số của họ. Điều này có thể hữu ích cho các chuyên gia y tế trong việc xác định bệnh nhân có thể mắc bệnh tiểu đường và trong việc phát triển kế hoạch điều trị cá nhân hóa. Ngoài ra, tập dữ liệu cũng có thể được sử dụng bởi các nhà nghiên cứu để khám phá mối quan hệ giữa các yếu tố y tế và dân số khác nhau và khả năng phát triển tiểu đường.

## Tổng quan về dữ liệu

Tập tin: [diabetes\_prediction\_dataset.csv](https://docs.google.com/spreadsheets/d/1hhOZUT1ap0mKHLHsqMzLZl6KnTgwnt_D-FfZFVy5J0c/edit?usp=sharing)

* Tập tin gồm 100000 bản ghi và 9 cột
* Tập tin gồm 9 thuộc tính sau:
  + **gender:** (Giới tính) đề cập đến giới tính sinh học của đối tượng. Gồm 3 giá trị:
    - Female: Nữ: 0
    - Male: Nam: 1
    - Other: Khác: 2
  + **age:** (Tuổi) là giá trị số thực, miền giá trị từ 0.08 – 80 tuổi.
    - 0 – 14: nhóm tuổi dưới lao động: 1
    - 15 – 64: nhóm tuổi lao động: 2
    - 65 trở lên: nhóm tuổi ngoài lao động: 3
  + **hypertension:** (Tăng huyết áp) là một tình trạng y tế trong đó áp lực máu trong các động mạch được duy trì ở mức cao. Có giá trị là 0 hoặc 1:
    - 0: không mắc huyết áp cao
    - 1: mắc huyết áp cao
  + **heart\_disease:** (Bệnh tim mạch) là một tình trạng y tế khác liên quan đến nguy cơ tăng cao của việc phát triển bệnh tiểu đường. Có giá trị là 0 hoặc 1
    - 0 chỉ rằng họ không mắc bệnh tim mạch và 1 chỉ rằng họ mắc bệnh tim mạch.
  + **smoking\_history: (**Lịch sử hút thuốc) là một yếu tố nguy cơ cho bệnh tiểu đường và có thể làm trầm trọng thêm các biến chứng liên quan đến tiểu đường. Có 6 hạng mục:
    - No Info: Không có thông tin: 0
    - Current: hút thuốc hiện tại: 1
    - Ever: từng hút thuốc: 2
    - former: từng hút thuốc nhưng không hiện tại: 3
    - Never: không bao giờ: 4
    - not current: không hút thuốc hiện tại: 5
  + **bmi:** Chỉ số khối cơ thể (BMI) là một chỉ số đo lường lượng mỡ trong cơ thể dựa trên cân nặng và chiều cao. Các giá trị BMI cao liên kết với nguy cơ tiểu đường cao hơn. Phạm vi của BMI trong tập dữ liệu là từ 10 đến 95.7.
    - <18,5: thiếu cân: 0
    - 18,5 - 24,9: cân đối: 1
    - 25 – 19.9: Thừa cân: 2
    - 30 – 34.9: Béo phì: 3
    - >= 35: Béo phì nguy hiểm: 4
  + **HbA1c\_level:** Mức độ HbA1c (Hemoglobin A1c) là một chỉ số đo lường trung bình của mức đường huyết của một người trong vòng 2-3 tháng gần đây. Các mức độ cao hơn cho thấy nguy cơ phát triển bệnh tiểu đường cao hơn. Đa số các mức độ HbA1c cao hơn 6,5% chỉ ra sự hiện diện của tiểu đường. Phạm vi của hba1c trong tập dữ liệu từ 3.5 – 9:
    - < 5.7%: Bình thường: 0
    - 5.7% - 6.4%: Tiền đái tháo đường: 1
    - >= 6.5%: Đái tháo đường: 2
  + **blood\_glucose\_level**: Mức đường glucose trong máu đề cập đến lượng glucose trong tuần hoàn máu tại một thời điểm nhất định. Các mức đường glucose trong máu cao là một chỉ báo chính của tiểu đường. Chỉ số trong tập dữ liệu có miền giá trị từ 80 đến 300
    - < 100: Bình thường: 0
    - 100 – 125: Tiền tiểu đường: 1
    - >= 126: Tiểu đường: 2
  + **diabetes:** Tiểu đường là biến mục tiêu được dự đoán, gồm 2 giá trị 0 và 1, với
    - 1: tiểu đường
    - 0: không tiểu đường.

## Thống kê:



## Biểu đồ cột:

A bar graph with numbers and a number

Description automatically generated

A graph of a number of people

Description automatically generated with medium confidence

A graph of a heart disease

Description automatically generated

A graph of smoking and smoking history

Description automatically generated with medium confidence

A blue and orange rectangular bar graph

Description automatically generated

## Biểu đồ ploxbot:

A blue rectangular object with black lines

Description automatically generated

A graph of a blue rectangular object with black dots

Description automatically generated

A blue rectangular object with black lines

Description automatically generated

A diagram of a blood glucose level

Description automatically generated

## Biểu đồ histogram

A graph with blue bars

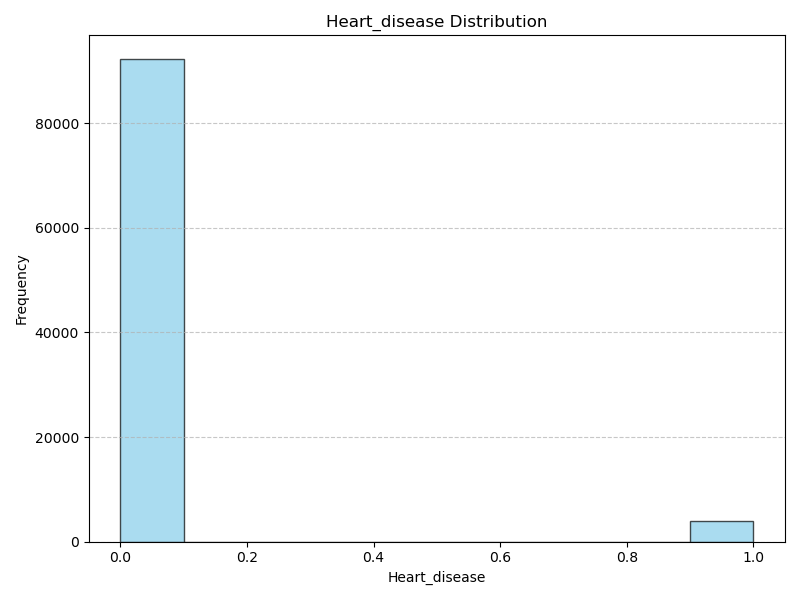
Description automatically generated with medium confidence

A graph of age distribution

Description automatically generated

A graph with a bar

Description automatically generated

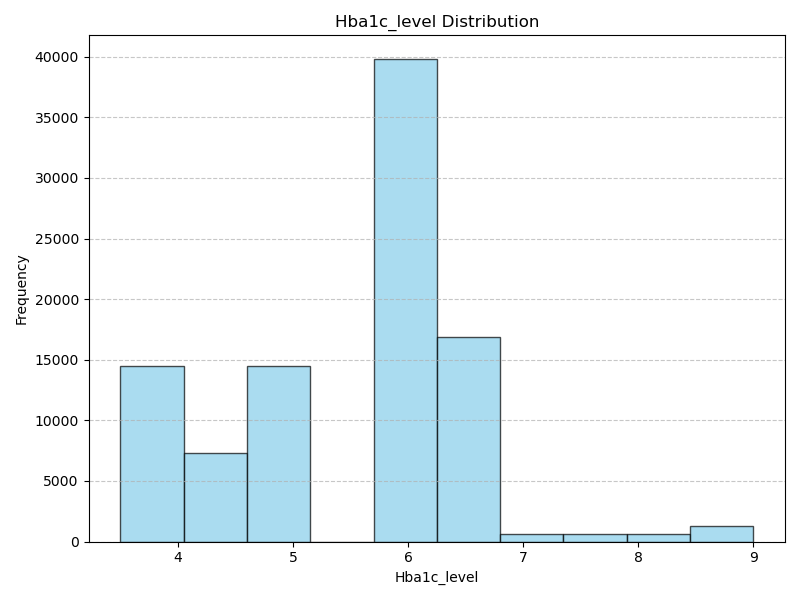


A graph of smoking and smoking

Description automatically generated with medium confidence

A graph with a bar graph

Description automatically generated with medium confidence



A graph of blood glucose level distribution

Description automatically generated

A graph of diabetes distribution

Description automatically generated

# Phân lớp dữ liệu

## Mô tả cách thực hiện

* Dữ liệu gốc:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| **D1** | Male | 34 | 1 | 0 | No Info | 35.41 | 6.6 | 100 | 0 |
| **D2** | Male | 53 | 0 | 0 | current | 30.8 | 6.6 | 280 | 1 |
| **D3** | Male | 55 | 0 | 0 | current | 27.32 | 6 | 80 | 0 |
| **D4** | Other | 28 | 0 | 0 | ever | 27.32 | 4.5 | 126 | 0 |
| **D5** | Female | 62 | 1 | 1 | never | 43.16 | 8.8 | 280 | 1 |
| **D6** | Female | 55 | 0 | 0 | ever | 18.6 | 6 | 160 | 0 |
| **D7** | Male | 6 | 0 | 1 | ever | 17.17 | 6.5 | 240 | 1 |
| **D8** | Male | 73 | 0 | 0 | former | 25.91 | 9 | 160 | 1 |
| **D9** | Female | 12 | 0 | 0 | No Info | 30.58 | 6.2 | 240 | 1 |
| **D10** | Female | 80 | 0 | 0 | former | 27.32 | 6.1 | 100 | 0 |
| **D11** | Male | 31 | 0 | 0 | former | 25.25 | 4.8 | 140 | 0 |
| **D12** | Other | 45 | 0 | 0 | never | 27.32 | 4 | 159 | 0 |
| **D13** | Female | 64 | 0 | 0 | never | 36.71 | 5 | 100 | 0 |
| **D14** | Female | 67 | 0 | 0 | never | 27.32 | 6.2 | 159 | 1 |
| **D15** | Other | 23 | 0 | 0 | No Info | 24.23 | 6.1 | 140 | 0 |
| **D16** | Other | 53 | 0 | 0 | No Info | 27.32 | 6.6 | 160 | 0 |
| **D17** | Female | 68 | 0 | 0 | No Info | 40.31 | 7.5 | 260 | 1 |
| **D18** | Other | 10 | 0 | 0 | not current | 14.09 | 5 | 140 | 0 |
| **D19** | Other | 52 | 0 | 0 | not current | 31.63 | 6.6 | 159 | 0 |
| **D20** | Male | 50 | 1 | 0 | not current | 25.49 | 6.1 | 260 | 1 |

* Dữ liệu sau khi xử lý:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| **D1** | 1 | 1 | 1 | 0 | 0 | 4 | 2 | 0 | 0 |
| **D2** | 1 | 1 | 0 | 0 | 1 | 3 | 2 | 2 | 1 |
| **D3** | 1 | 1 | 0 | 0 | 1 | 2 | 1 | 0 | 0 |
| **D4** | 2 | 1 | 0 | 0 | 2 | 2 | 0 | 1 | 0 |
| **D5** | 0 | 1 | 1 | 1 | 4 | 4 | 2 | 2 | 1 |
| **D6** | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 2 | 0 |
| **D7** | 1 | 0 | 0 | 1 | 2 | 0 | 1 | 2 | 1 |
| **D8** | 1 | 2 | 0 | 0 | 3 | 2 | 2 | 2 | 1 |
| **D9** | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 2 | 1 |
| **D10** | 0 | 2 | 0 | 0 | 3 | 2 | 1 | 0 | 0 |
| **D11** | 1 | 1 | 0 | 0 | 3 | 2 | 0 | 2 | 0 |
| **D12** | 2 | 1 | 0 | 0 | 4 | 2 | 0 | 2 | 0 |
| **D13** | 0 | 1 | 0 | 0 | 4 | 4 | 0 | 0 | 0 |
| **D14** | 0 | 2 | 0 | 0 | 4 | 2 | 1 | 2 | 1 |
| **D15** | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 2 | 0 |
| **D16** | 2 | 1 | 0 | 0 | 0 | 2 | 2 | 2 | 0 |
| **D17** | 0 | 2 | 0 | 0 | 0 | 4 | 2 | 2 | 1 |
| **D18** | 2 | 0 | 0 | 0 | 5 | 0 | 0 | 2 | 0 |
| **D19** | 2 | 1 | 0 | 0 | 5 | 3 | 2 | 2 | 0 |
| **D20** | 1 | 1 | 1 | 0 | 5 | 2 | 1 | 2 | 1 |

### Decision tree

**Lần 1:**

|  |  |
| --- | --- |
| 1 | 8 |
| 0 | 12 |

* 
* Gender:

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 4 | 3 |
| 1 | 4 | 3 |
| 2 | 0 | 6 |

* + 
  + 
  + 
  + 
* Age:

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 2 | 1 |
| 1 | 3 | 10 |
| 2 | 3 | 1 |

* + 
  + 
  + 
  + 
* Hypertension

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 6 | 11 |
| 1 | 2 | 1 |

* + 
  + 
  + 
* heart\_disease

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 6 | 12 |
| 1 | 2 | 0 |

* + 
  + 
  + 
* smoking\_history

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 2 | 3 |
| 1 | 1 | 1 |
| 2 | 1 | 2 |
| 3 | 1 | 2 |
| 4 | 2 | 2 |
| 5 | 1 | 2 |

* + 
  + 
  + 
  + 
  + 
  + 
  + 
* bmi

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 2 |
| 2 | 3 | 6 |
| 3 | 2 | 1 |
| 4 | 2 | 2 |

* + 
  + 
  + 
  + 
  + 
  + 
* HbA1c\_level

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 5 |
| 1 | 4 | 4 |
| 2 | 4 | 3 |

* + 
  + 
  + 
  + 
* blood\_glucose\_level

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 4 |
| 1 | 0 | 1 |
| 2 | 8 | 7 |

* + 
  + 
  + 
  + 

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **gender** | age | hypertension | heart\_disease | smoking\_history | bmi | HbA1c\_level | blood\_glucose\_level |
| **0,2815** | 0,165 | 0,037 | 0,1448 | 0,015 | 0,12 | 0,226 | 0,223 |

A diagram of a gender

Description automatically generated

**Lần 2**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| D5 | 1 | 1 | 1 | 4 | 4 | 2 | 2 | 1 |
| D6 | 1 | 0 | 0 | 2 | 1 | 1 | 2 | 0 |
| D9 | 0 | 0 | 0 | 0 | 3 | 1 | 2 | 1 |
| D10 | 2 | 0 | 0 | 3 | 2 | 1 | 0 | 0 |
| D13 | 1 | 0 | 0 | 4 | 4 | 0 | 0 | 0 |
| D14 | 2 | 0 | 0 | 4 | 2 | 1 | 2 | 1 |
| D17 | 2 | 0 | 0 | 0 | 4 | 2 | 2 | 1 |

|  |  |
| --- | --- |
| 1 | 4 |
| 0 | 3 |

* 
* Age:

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 0 |
| 1 | 1 | 2 |
| 2 | 2 | 1 |

* + 
  + 
  + 
  + 
* Hypertension

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 3 | 3 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* heart\_disease

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 3 | 3 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* smoking\_history

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 2 | 0 |
| 1 | 0 | 0 |
| 2 | 0 | 1 |
| 3 | 0 | 1 |
| 4 | 2 | 1 |
| 5 | 0 | 0 |

* + 
  + 
  + 
  + 
  + 
  + 
  + 
* bmi

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 0 |
| 1 | 0 | 1 |
| 2 | 1 | 1 |
| 3 | 1 | 0 |
| 4 | 2 | 1 |

* + 
  + 
  + 
  + 
  + 
  + 
* HbA1c\_level

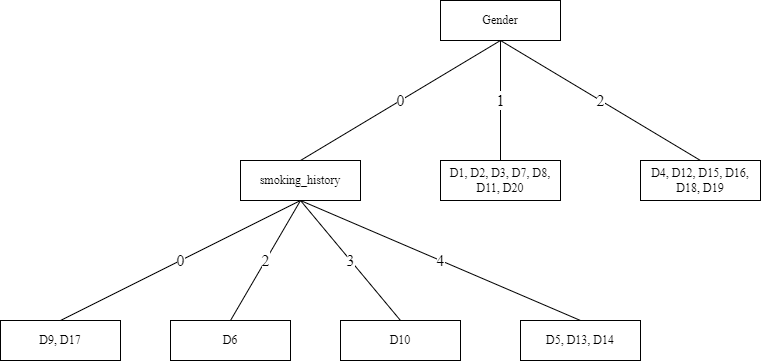
|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 1 |
| 1 | 2 | 2 |
| 2 | 2 | 0 |

* + 
  + 
  + 
  + 
* blood\_glucose\_level

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 2 |
| 1 | 0 | 0 |
| 2 | 4 | 1 |

* + 
  + 
  + 
  + 

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| age | hypertension | heart\_disease | **smoking\_history** | bmi | HbA1c\_level | blood\_glucose\_level |
| 0,918 | 0,128 | 0,128 | **0,596** | 0,306 | 0,416 | 0,469 |

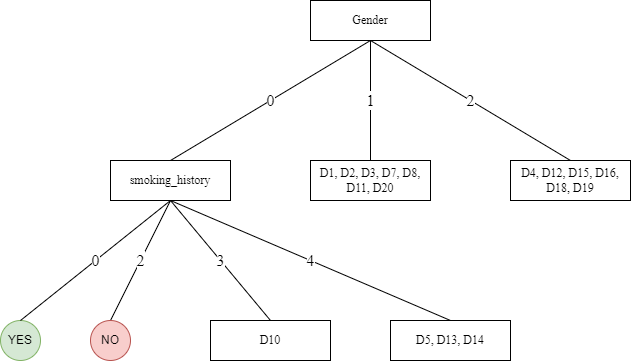


**Lần 3**

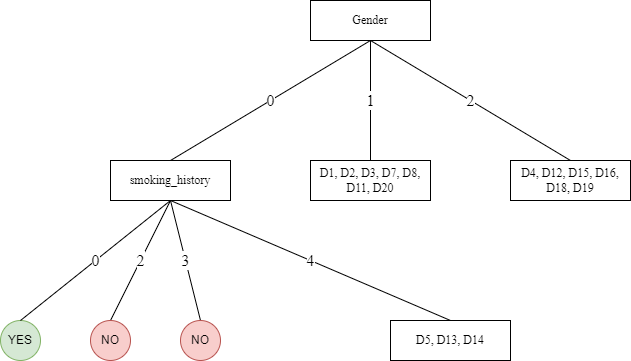
A diagram of a structure

Description automatically generated

**Lần 4**



**Lần 5**



**Lần 6**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | **age** | **hypertension** | **heart\_disease** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| D5 | 1 | 1 | 1 | 4 | 2 | 2 | 1 |
| D13 | 1 | 0 | 0 | 4 | 0 | 0 | 0 |
| D14 | 2 | 0 | 0 | 2 | 1 | 2 | 1 |

|  |  |
| --- | --- |
| 1 | 2 |
| 0 | 1 |

* 
* Age:

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 0 |
| 1 | 1 | 1 |
| 2 | 1 | 0 |

* + 
  + 
  + 
  + 
* Hypertension

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 1 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* heart\_disease

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 1 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* bmi

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 2 | 1 | 0 |
| 3 | 0 | 0 |
| 4 | 1 | 1 |

* + 
  + 
  + 
  + 
  + 
  + 
* HbA1c\_level

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 1 |
| 1 | 1 | 0 |
| 2 | 1 | 0 |

* + 
  + 
  + 
  + 
* blood\_glucose\_level

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 1 |
| 1 | 0 | 0 |
| 2 | 2 | 0 |

* + 
  + 
  + 
  + 

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| age | hypertension | heart\_disease | bmi | **HbA1c\_level** | blood\_glucose\_level |
| 0,251 | 0,251 | 0,251 | 0,251 | **0,918** | 0,918 |

A diagram of a person's relationship

Description automatically generated

**Lần 7**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **STT** | **age** | **hypertension** | **heart\_disease** | **bmi** | **blood\_glucose\_level** | **diabetes** |
| D1 | 1 | 1 | 0 | 4 | 0 | 0 |
| D2 | 1 | 0 | 0 | 3 | 2 | 1 |
| D3 | 1 | 0 | 0 | 2 | 0 | 0 |
| D7 | 0 | 0 | 1 | 0 | 2 | 1 |
| D8 | 2 | 0 | 0 | 2 | 2 | 1 |
| D11 | 1 | 0 | 0 | 2 | 2 | 0 |
| D20 | 1 | 1 | 0 | 2 | 2 | 1 |

|  |  |
| --- | --- |
| 1 | 4 |
| 0 | 3 |

* 
* Age:

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 0 |
| 1 | 2 | 3 |
| 2 | 1 | 0 |

* + 
  + 
  + 
  + 
* Hypertension

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 3 | 2 |
| 1 | 1 | 1 |

* + 
  + 
  + 
* heart\_disease

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 3 | 3 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* bmi

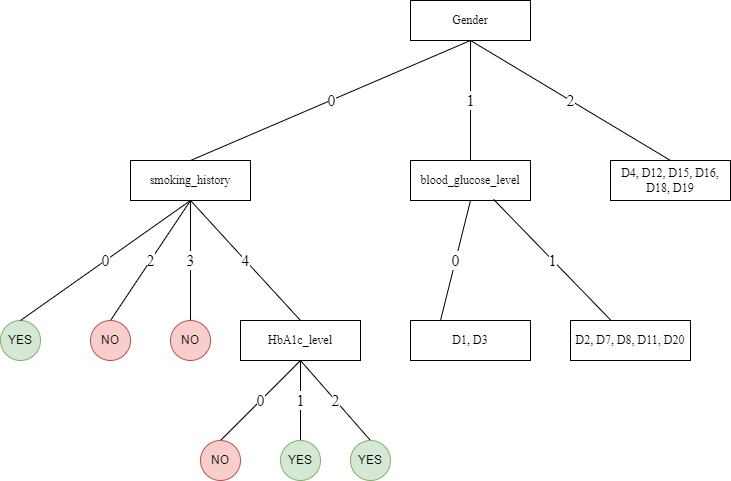
|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 2 | 2 | 2 |
| 3 | 1 | 0 |
| 4 | 0 | 1 |

* + 
  + 
  + 
  + 
  + 
  + 
* blood\_glucose\_level

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 2 |
| 1 | 0 | 0 |
| 2 | 4 | 1 |

* + 
  + 
  + 
  + 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| age | hypertension | heart\_disease | bmi | **blood\_glucose\_level** |
| 0,291 | 0,291 | 0,128 | 0,414 | **0,469** |



**Lần 8**

A diagram of a flowchart

Description automatically generated

**Lần 9**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STT | **age** | **hypertension** | **heart\_disease** | **bmi** | **diabetes** |
| D2 | 1 | 0 | 0 | 3 | 1 |
| D7 | 0 | 0 | 1 | 0 | 1 |
| D8 | 2 | 0 | 0 | 2 | 1 |
| D11 | 1 | 0 | 0 | 2 | 0 |
| D20 | 1 | 1 | 0 | 2 | 1 |

|  |  |
| --- | --- |
| 1 | 4 |
| 0 | 1 |

* 
* Age:

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 0 |
| 1 | 2 | 1 |
| 2 | 1 | 0 |

* + 
  + 
  + 
  + 
* Hypertension

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 3 | 1 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* heart\_disease

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 3 | 1 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* bmi

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 2 | 2 | 1 |
| 3 | 1 | 0 |
| 4 | 0 | 0 |

* + 
  + 
  + 
  + 
  + 
  + 

|  |  |  |  |
| --- | --- | --- | --- |
| age | hypertension | heart\_disease | **bmi** |
| 0,221 | 0,123 | 0,123 | **0,221** |

A diagram of a tree

Description automatically generated

**Lần 10**

A diagram of a tree

Description automatically generated

**Lần 11**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | **age** | **hypertension** | **heart\_disease** | **diabetes** |
| D8 | 2 | 0 | 0 | 1 |
| D11 | 1 | 0 | 0 | 0 |
| D20 | 1 | 1 | 0 | 1 |

|  |  |
| --- | --- |
| 1 | 2 |
| 0 | 1 |

* 
* Age:

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 0 |
| 1 | 1 | 1 |
| 2 | 1 | 0 |

* + 
  + 
  + 
  + 
* Hypertension

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 1 |
| 1 | 1 | 0 |

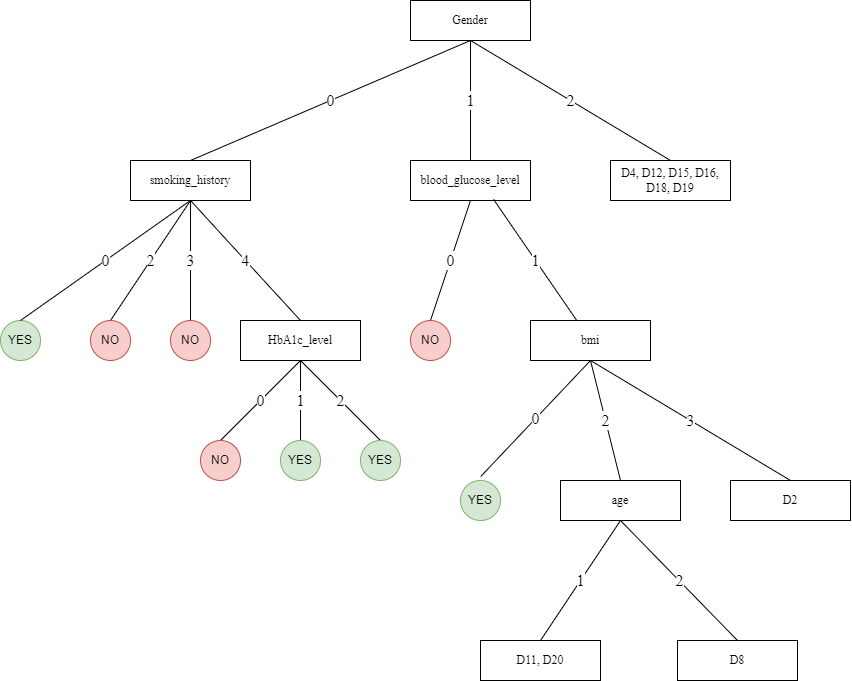
* + 
  + 
  + 
* heart\_disease

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 2 | 1 |
| 1 | 0 | 0 |

* + 
  + 



|  |  |  |
| --- | --- | --- |
| **age** | hypertension | heart\_disease |
| **0,251** | 0,251 | 0 |



**Lần 12**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | **hypertension** | **heart\_disease** | **diabetes** |
| D11 | 0 | 0 | 0 |
| D20 | 1 | 0 | 1 |

|  |  |
| --- | --- |
| 1 | 1 |
| 0 | 1 |

* 
* Hypertension

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 0 | 1 |
| 1 | 1 | 0 |

* + 
  + 
  + 
* heart\_disease

|  |  |  |
| --- | --- | --- |
| Giá trị | 1 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 0 |

* + 
  + 



|  |  |
| --- | --- |
| **hypertension** | heart\_disease |
| **1** | 0 |

A diagram of a tree

Description automatically generated

**Lần 13**

A diagram of a company

Description automatically generated

**Lần 14**

A diagram of a tree

Description automatically generated

**Lần 15**

A diagram of a tree

Description automatically generated

### Naive bayes

* **Tập 1:**
* Dữ liệu vào:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| Female | 65 | 0 | 0 | current | 27.32 | 7.5 | 145 | 1 |

* Dữ liệu sau khi xử lý:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| 0 | 1 | 0 | 0 | 1 | 2 | 2 | 2 | 1 |

* Tính
* H1: Tiểu đường =1
* H2: Tiểu đường = 0
* X1: Gender = Female=0
* X2: age = 65=1
* X3: Hypertension = 0
* X4:Heart-Disease =0
* X5: Moking-history = Curent=1
* X6: Bmi = 27,32 =2
* X7 HbA1c-Level = 7,5=2
* X8: blood-glucose-level=145=2
* P(H1) = 8/20 = 0,4
  + - P(X1/H1) = 4/8 = 0,5
    - P(X2/H1) = 3/8 = 0,375
    - P(X3/H1) = 6/8 = 0,75
    - P(X4/H1) = 6/8 = 0,75
    - P(X5/H1) =1/8 = 0,125
    - P(X6/H1) = 3/8 = 0,375
    - P(X7/H1) = 4/8 = 0,5
    - P(X8/H1) = 8/8 =1

=> P(X/H1) = 0,5.0,375.0,75.0,75.0,125.0,375.0,5.1=0,002471923828125

* P(H2) = 12/20 = 0,6
  + P(X1/H2) =3/12 =0,25
  + P(X2/H2) = 10/12 = 0,83
  + P(X3/H2) = 11/12 = 0,916
  + P(X4/H2) = 12/12 = 1
  + P(X5/H2) = 1/12 = 0,083
  + P(X6/H2) = 6/12 = 0,5
  + P(X7/H2) = 3/12 = 0,25
  + P(X8/H2) = 7/12 = 0,583

=>P(X/H2) = 0,25.0,83.0,916.1.0,083.0,5.0,25.0,583=0.00114966215375

Do đó, tập dữ liệu thuộc nhãn 1.

* **Tập 2:**
  + Dữ liệu vào:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| Female | 26 | 0 | 0 | not current | 38.95 | 3.5 | 100 | 0 |

* + Dữ liệu sau khi xử lý:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| 0 | 1 | 0 | 0 | 4 | 4 | 0 | 0 | 0 |

* Tính
* H1: Tiểu đường =1
* H2: Tiểu đường =0
* X1: Gender = Female=0
* X2 :age = 26=1
* X3: Hypertension =0
* X4: heart-disease=0
* X5: Moking-history = not current=4
* X6: Bmi = 38,95=4
* X7: HbA1c-level = 3,5=0
* X8:blood-glucose-level = 100=0
* P(H1) = 8/20 = 0,4
  + - P(X1/H1) = 4/8 = 0,5
    - P(X2/H1) = 3/8 = 0,375
    - P(X3/H1) = 6/8= 0,75
    - P(X4/H1) = 6/8 = 0,75
    - P(X5/H1) =2/8=0,25
    - P(X6/H1) =2/8= 0,25
    - P(X7/H1) =0/8=0
    - P(X8/H1) = 0/8=0
    - =>P(X/H1) = 0
* P(H2) = 12/20 = 0,6
  + - P(X1/H2) = 3/12 =0,25
    - P(X2/H2) = 10/12 = 0,83
    - P(X3/H2) = 11/12 = 0,916
    - P(X4/H2) = 12/12 = 1
    - P(X5/H2) =1/12=0,083
    - P(X6/H2) =6/12= 0,5
    - P(X7/H2) =3/12 = 0,25
    - P(X8/H2) = 7/12=0,583

=>P(X/H2) = 0.25\*0.83\*0.916\*1\*0.083\*0.5\*0.25\*0.583=0.00114966215375

Do đó, tập dữ liệu thuộc nhãn 1

### Support Vector Machines

* Tập dữ liệu thực hiện:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| **D1** | Female | 80 | 0 | 0 | never | 24.01 | 9 | 130 | 1 |
| **D2** | Female | 46 | 0 | 0 | No Info | 27.32 | 7.5 | 130 | 1 |
| **D3** | Female | 62 | 0 | 0 | never | 21.52 | 7.5 | 260 | 1 |
| **D4** | Female | 52 | 0 | 1 | former | 27.32 | 6.1 | 126 | 1 |
| **D5** | Female | 43 | 0 | 0 | never | 34.76 | 8.8 | 126 | 1 |
| **D6** | Female | 51 | 0 | 0 | never | 24.76 | 9 | 300 | 1 |
| **D7** | Male | 62 | 1 | 0 | No Info | 28.03 | 9 | 240 | 1 |
| **D8** | Male | 71 | 1 | 0 | never | 32.97 | 5.7 | 300 | 1 |
| **D9** | Male | 62 | 0 | 0 | ever | 34.72 | 9 | 140 | 1 |
| **D10** | Female | 49 | 0 | 0 | former | 57.7 | 6.1 | 130 | 1 |
| **D11** | Female | 58 | 0 | 0 | current | 20.16 | 8.2 | 155 | 1 |
| **D12** | Female | 61 | 0 | 1 | ever | 27.32 | 6.5 | 260 | 1 |
| **D13** | Male | 50 | 1 | 0 | never | 55.92 | 5.8 | 155 | 1 |
| **D14** | Male | 64 | 0 | 0 | never | 39.17 | 6 | 80 | 0 |
| **D15** | Female | 16 | 0 | 0 | No Info | 21.49 | 6.2 | 90 | 0 |
| **D16** | Female | 53 | 0 | 0 | never | 29.4 | 3.5 | 160 | 0 |
| **D17** | Male | 61 | 1 | 0 | never | 27.32 | 6.2 | 145 | 0 |
| **D18** | Female | 55 | 0 | 0 | never | 40.73 | 4.8 | 80 | 0 |
| **D19** | Female | 42 | 0 | 0 | never | 44.44 | 5.7 | 158 | 0 |
| **D20** | Female | 71 | 0 | 0 | not current | 30.55 | 5.7 | 85 | 0 |

* Tập dữ liệu sau khi sử lý:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| **D1** | 0 | 1 | 0 | 0 | 0,8 | 0,1025572722 | 1 | 0,2272727273 | 1 |
| **D2** | 0 | 0,46875 | 0 | 0 | 0 | 0,1907298881 | 0,7272727273 | 0,2272727273 | 1 |
| **D3** | 0 | 0,71875 | 0 | 0 | 0,8 | 0,03622802344 | 0,7272727273 | 0,8181818182 | 1 |
| **D4** | 0 | 0,5625 | 0 | 1 | 0,6 | 0,1907298881 | 0,4727272727 | 0,2090909091 | 1 |
| **D5** | 0 | 0,421875 | 0 | 0 | 0,8 | 0,3889184869 | 0,9636363636 | 0,2090909091 | 1 |
| **D6** | 0 | 0,546875 | 0 | 0 | 0,8 | 0,1225359616 | 1 | 1 | 1 |
| **D7** | 1 | 0,71875 | 1 | 0 | 0 | 0,2096430474 | 1 | 0,7272727273 | 1 |
| **D8** | 1 | 0,859375 | 1 | 0 | 0,8 | 0,3412360149 | 0,4 | 1 | 1 |
| **D9** | 1 | 0,71875 | 0 | 0 | 0,4 | 0,3878529568 | 1 | 0,2727272727 | 1 |
| **D10** | 0 | 0,515625 | 0 | 0 | 0,6 | 1 | 0,4727272727 | 0,2272727273 | 1 |
| **D11** | 0 | 0,65625 | 0 | 0 | 0,2 | 0 | 0,8545454545 | 0,3409090909 | 1 |
| **D12** | 0 | 0,703125 | 0 | 1 | 0,4 | 0,1907298881 | 0,5454545455 | 0,8181818182 | 1 |
| **D13** | 1 | 0,53125 | 1 | 0 | 0,8 | 0,9525839105 | 0,4181818182 | 0,3409090909 | 1 |
| **D14** | 1 | 0,75 | 0 | 0 | 0,8 | 0,5063931806 | 0,4545454545 | 0 | 0 |
| **D15** | 0 | 0 | 0 | 0 | 0 | 0,03542887587 | 0,4909090909 | 0,04545454545 | 0 |
| **D16** | 0 | 0,578125 | 0 | 0 | 0,8 | 0,2461374534 | 0 | 0,3636363636 | 0 |
| **D17** | 1 | 0,703125 | 1 | 0 | 0,8 | 0,1907298881 | 0,4909090909 | 0,2954545455 | 0 |
| **D18** | 0 | 0,609375 | 0 | 0 | 0,8 | 0,5479488546 | 0,2363636364 | 0 | 0 |
| **D19** | 0 | 0,40625 | 0 | 0 | 0,8 | 0,6467767714 | 0,4 | 0,3545454545 | 0 |
| **D20** | 0 | 0,859375 | 0 | 0 | 1 | 0,2767714438 | 0,4 | 0,02272727273 | 0 |

* Biên cứng

dexpr float cost = (x1 \* x1 + x2 \* x2 + x3 \* x3 + x4 \* x4 + x5 \* x5 + x6 \* x6 + x7 \* x7 + x8 \* x8 ) \* 0.5;

minimize cost;

subject to {

0.0 \* x1 + 1.0 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.10255727224294087 \* x6 + 1.0 \* x7 + 0.22727272727272718 \* x8 + b>= 1;

0.0 \* x1 + 0.46875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.0 \* x5 + 0.19072988811933933 \* x6 + 0.7272727272727274 \* x7 + 0.22727272727272718 \* x8 + b>= 1;

0.0 \* x1 + 0.71875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.036228023441662205 \* x6 + 0.7272727272727274 \* x7 + 0.818181818181818 \* x8 + b>= 1;

0.0 \* x1 + 0.5625 \* x2 + 0.0 \* x3 + 1.0 \* x4 + 0.6 \* x5 + 0.19072988811933933 \* x6 + 0.4727272727272728 \* x7 + 0.2090909090909091 \* x8 + b>= 1;

0.0 \* x1 + 0.421875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.38891848694725617 \* x6 + 0.9636363636363637 \* x7 + 0.2090909090909091 \* x8 + b>= 1;

0.0 \* x1 + 0.546875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.1225359616409164 \* x6 + 1.0 \* x7 + 0.9999999999999999 \* x8 + b>= 1;

1.0 \* x1 + 0.71875 \* x2 + 1.0 \* x3 + 0.0 \* x4 + 0.0 \* x5 + 0.20964304741608952 \* x6 + 1.0 \* x7 + 0.7272727272727272 \* x8 + b>= 1;

1.0 \* x1 + 0.859375 \* x2 + 1.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.3412360149174214 \* x6 + 0.4 \* x7 + 0.9999999999999999 \* x8 + b>= 1;

1.0 \* x1 + 0.71875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.4 \* x5 + 0.3878529568460308 \* x6 + 1.0 \* x7 + 0.2727272727272727 \* x8 + b>= 1;

0.0 \* x1 + 0.515625 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.6 \* x5 + 1.0 \* x6 + 0.4727272727272728 \* x7 + 0.22727272727272718 \* x8 + b>= 1;

0.0 \* x1 + 0.65625 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.2 \* x5 + 0.0 \* x6 + 0.8545454545454544 \* x7 + 0.34090909090909083 \* x8 + b>= 1;

0.0 \* x1 + 0.703125 \* x2 + 0.0 \* x3 + 1.0 \* x4 + 0.4 \* x5 + 0.19072988811933933 \* x6 + 0.5454545454545455 \* x7 + 0.818181818181818 \* x8 + b>= 1;

1.0 \* x1 + 0.53125 \* x2 + 1.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.9525839104954714 \* x6 + 0.4181818181818181 \* x7 + 0.34090909090909083 \* x8 + b>= 1;

-1.0 \* x1 -0.75 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.5063931806073521 \* x6 -0.4545454545454545 \* x7 -0.0 \* x8 - b>= 1;

-0.0 \* x1 -0.0 \* x2 -0.0 \* x3 -0.0 \* x4 -0.0 \* x5 -0.03542887586574317 \* x6 -0.4909090909090911 \* x7 -0.045454545454545414 \* x8 - b>= 1;

-0.0 \* x1 -0.578125 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.24613745338305804 \* x6 -0.0 \* x7 -0.36363636363636365 \* x8 - b>= 1;

-1.0 \* x1 -0.703125 \* x2 -1.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.19072988811933933 \* x6 -0.4909090909090911 \* x7 -0.2954545454545454 \* x8 - b>= 1;

-0.0 \* x1 -0.609375 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.547948854555141 \* x6 -0.23636363636363633 \* x7 -0.0 \* x8 - b>= 1;

-0.0 \* x1 -0.40625 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.6467767714437931 \* x6 -0.4 \* x7 -0.3545454545454545 \* x8 - b>= 1;

-0.0 \* x1 -0.859375 \* x2 -0.0 \* x3 -0.0 \* x4 -1.0 \* x5 -0.2767714437932871 \* x6 -0.4 \* x7 -0.022727272727272707 \* x8 - b>= 1;

}

* x1 = 0.28582;
* x2 = 2.3061;
* x3 = 0.28582;
* x4 = 2.7983;
* x5 = -2.4138;
* x6 = 3.5192;
* x7 = 5.9235;
* x8 = 3.2125;
* b = -5.7903;
* Biên mềm:

dexpr float cost = (x1 \* x1 + x2 \* x2 + x3 \* x3 + x4 \* x4 + x5 \* x5 + x6 \* x6 + x7 \* x7 + x8 \* x8 ) \* 0.5 + 5 \* (c1 +c2 +c3 +c4 +c5 +c6 +c7 +c8 +c9 +c10 +c11 +c12 +c13 +c14 +c15 +c16 +c17 +c18 +c19 +c20);

minimize cost;

subject to {

0.0 \* x1 + 1.0 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.10255727224294087 \* x6 + 1.0 \* x7 + 0.22727272727272718 \* x8 + b + c1>= 1;

0.0 \* x1 + 0.46875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.0 \* x5 + 0.19072988811933933 \* x6 + 0.7272727272727274 \* x7 + 0.22727272727272718 \* x8 + b + c2>= 1;

0.0 \* x1 + 0.71875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.036228023441662205 \* x6 + 0.7272727272727274 \* x7 + 0.818181818181818 \* x8 + b + c3>= 1;

0.0 \* x1 + 0.5625 \* x2 + 0.0 \* x3 + 1.0 \* x4 + 0.6 \* x5 + 0.19072988811933933 \* x6 + 0.4727272727272728 \* x7 + 0.2090909090909091 \* x8 + b + c4>= 1;

0.0 \* x1 + 0.421875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.38891848694725617 \* x6 + 0.9636363636363637 \* x7 + 0.2090909090909091 \* x8 + b + c5>= 1;

0.0 \* x1 + 0.546875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.1225359616409164 \* x6 + 1.0 \* x7 + 0.9999999999999999 \* x8 + b + c6>= 1;

1.0 \* x1 + 0.71875 \* x2 + 1.0 \* x3 + 0.0 \* x4 + 0.0 \* x5 + 0.20964304741608952 \* x6 + 1.0 \* x7 + 0.7272727272727272 \* x8 + b + c7>= 1;

1.0 \* x1 + 0.859375 \* x2 + 1.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.3412360149174214 \* x6 + 0.4 \* x7 + 0.9999999999999999 \* x8 + b + c8>= 1;

1.0 \* x1 + 0.71875 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.4 \* x5 + 0.3878529568460308 \* x6 + 1.0 \* x7 + 0.2727272727272727 \* x8 + b + c9>= 1;

0.0 \* x1 + 0.515625 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.6000000000000001 \* x5 + 1.0 \* x6 + 0.4727272727272728 \* x7 + 0.22727272727272718 \* x8 + b + c10>= 1;

0.0 \* x1 + 0.65625 \* x2 + 0.0 \* x3 + 0.0 \* x4 + 0.2 \* x5 + 0.0 \* x6 + 0.8545454545454544 \* x7 + 0.34090909090909083 \* x8 + b + c11>= 1;

0.0 \* x1 + 0.703125 \* x2 + 0.0 \* x3 + 1.0 \* x4 + 0.4 \* x5 + 0.19072988811933933 \* x6 + 0.5454545454545455 \* x7 + 0.818181818181818 \* x8 + b + c12>= 1;

1.0 \* x1 + 0.53125 \* x2 + 1.0 \* x3 + 0.0 \* x4 + 0.8 \* x5 + 0.9525839104954714 \* x6 + 0.4181818181818181 \* x7 + 0.34090909090909083 \* x8 + b + c13>= 1;

-1.0 \* x1 -0.75 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.5063931806073521 \* x6 -0.4545454545454545 \* x7 -0.0 \* x8 - b + c14>= 1;

-0.0 \* x1 -0.0 \* x2 -0.0 \* x3 -0.0 \* x4 -0.0 \* x5 -0.03542887586574317 \* x6 -0.4909090909090911 \* x7 -0.045454545454545414 \* x8 - b + c15>= 1;

-0.0 \* x1 -0.578125 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.24613745338305804 \* x6 -0.0 \* x7 -0.36363636363636365 \* x8 - b + c16>= 1;

-1.0 \* x1 -0.703125 \* x2 -1.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.19072988811933933 \* x6 -0.4909090909090911 \* x7 -0.2954545454545454 \* x8 - b + c17>= 1;

-0.0 \* x1 -0.609375 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.547948854555141 \* x6 -0.23636363636363633 \* x7 -0.0 \* x8 - b + c18>= 1;

-0.0 \* x1 -0.40625 \* x2 -0.0 \* x3 -0.0 \* x4 -0.8 \* x5 -0.6467767714437931 \* x6 -0.4 \* x7 -0.3545454545454545 \* x8 - b + c19>= 1;

-0.0 \* x1 -0.859375 \* x2 -0.0 \* x3 -0.0 \* x4 -1.0 \* x5 -0.2767714437932871 \* x6 -0.4 \* x7 -0.022727272727272707 \* x8 - b + c20>= 1;

}

* x1 = -0.73105;
* x2 = 1.1825;
* x3 = 0.55621;
* x4 = 1.4879;
* x5 = -1.2214;
* x6 = 2.1523;
* x7 = 2.9984;
* x8 = 1.8335;
* c1 = 3.6385e-10;
* c2 = 0.069359;
* c3 = 4.7541e-10;
* c4 = 5.1799e-10;
* c5 = 8.8777e-10;
* c6 = 3.6385e-10;
* c7 = 3.6385e-10;
* c8 = 4.0939e-10;
* c9 = 3.6385e-10;
* c10 = 3.6385e-10;
* c11 = 3.6385e-10;
* c12 = 3.6385e-10;
* c13 = 0.22611;
* c14 = 4.9e-10;
* c15 = 1.9001e-9;
* c16 = 3.6385e-10;
* c17 = 0.47211;
* c18 = 3.6745e-10;
* c19 = 1.1132;
* c20 = 4.3207e-10;
* b = -2.6315;
* Kết quả:
  + Dữ liệu test:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| **D21** | Female | 65 | 0 | 0 | current | 27.32 | 7.5 | 145 | 1 |
| **D22** | Female | 56 | 0 | 0 | No Info | 27.32 | 8.2 | 130 | 1 |
| **D23** | Male | 64 | 0 | 0 | current | 27.32 | 8.8 | 160 | 1 |
| **D24** | Female | 79 | 1 | 0 | never | 30.51 | 6.6 | 155 | 1 |
| **D25** | Male | 74 | 1 | 0 | former | 33.23 | 5.7 | 280 | 1 |
| **D26** | Female | 15 | 0 | 0 | No Info | 21.63 | 4.5 | 100 | 0 |
| **D27** | Male | 49 | 0 | 0 | never | 27.32 | 6 | 200 | 0 |
| **D28** | Female | 26 | 0 | 0 | not current | 38.95 | 3.5 | 100 | 0 |
| **D29** | Female | 62 | 0 | 0 | never | 27.32 | 6.2 | 158 | 0 |
| **D30** | Male | 17 | 0 | 0 | No Info | 27.32 | 5 | 145 | 0 |

* + Dữ liệu sau khi xử lý:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** | **Biên cứng** | **Lớp** | **Biên mềm** | **Lớp** |
| **D21** | 0 | 0,78125 | 0 | 0 | 0,25 | 0,32852194 | 0,7547169811 | 0,25 | 1 | 1,837716074 | 1 | 1,415374293 | 1 |
| **D22** | 0 | 0,640625 | 0 | 0 | 0 | 0,32852194 | 0,8867924528 | 0,1666666667 | 1 | 2,631511485 | 1 | 1,797658658 | 1 |
| **D23** | 1 | 0,765625 | 0 | 0 | 0,25 | 0,32852194 | 1 | 0,3333333333 | 1 | 3,808145557 | 1 | 1,554096001 | 1 |
| **D24** | 0 | 1 | 1 | 0 | 0,75 | 0,5127020785 | 0,5849056604 | 0,3055555556 | 1 | 1,241857056 | 1 | 1,608665927 | 1 |
| **D25** | 1 | 0,921875 | 1 | 0 | 0,5 | 0,6697459584 | 0,4150943396 | 1 | 1 | 3,728657235 | 1 | 2,192690282 | 1 |
| **D26** | 0 | 0 | 0 | 0 | 0 | 0 | 0,1886792453 | 0 | 0 | -4,67265849 | 0 | -2,065764151 | 0 |
| **D27** | 1 | 0,53125 | 0 | 0 | 0,75 | 0,32852194 | 0,4716981132 | 0,5555555556 | 0 | -0,3547539678 | 0 | -0,5103683697 | 0 |
| **D28** | 0 | 0,171875 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | -4,288539063 | 0 | -1,497357813 | 0 |
| **D29** | 0 | 0,734375 | 0 | 0 | 0,75 | 0,32852194 | 0,5094339623 | 0,3222222222 | 0 | -0,6982024368 | 0 | 0,1462074459 | 1 |
| **D30** | 1 | 0,03125 | 0 | 0 | 0 | 0,32852194 | 0,2830188679 | 0,25 | 0 | -1,7966927 | 0 | -1,31154033 | 0 |

## Phân tích và so sánh

A graph of different colored rectangular shapes

Description automatically generated

# Phân cụm dữ liệu

## Mô tả cách thực hiện

* Dữ liệu gốc:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| **D1** | Male | 34 | 1 | 0 | No Info | 35.41 | 6.6 | 100 | 0 |
| **D2** | Male | 53 | 0 | 0 | current | 30.8 | 6.6 | 280 | 1 |
| **D3** | Male | 55 | 0 | 0 | current | 27.32 | 6 | 80 | 0 |
| **D4** | Other | 28 | 0 | 0 | ever | 27.32 | 4.5 | 126 | 0 |
| **D5** | Female | 62 | 1 | 1 | never | 43.16 | 8.8 | 280 | 1 |
| **D6** | Female | 55 | 0 | 0 | ever | 18.6 | 6 | 160 | 0 |
| **D7** | Male | 6 | 0 | 1 | ever | 17.17 | 6.5 | 240 | 1 |
| **D8** | Male | 73 | 0 | 0 | former | 25.91 | 9 | 160 | 1 |
| **D9** | Female | 12 | 0 | 0 | No Info | 30.58 | 6.2 | 240 | 1 |
| **D10** | Female | 80 | 0 | 0 | former | 27.32 | 6.1 | 100 | 0 |
| **D11** | Male | 31 | 0 | 0 | former | 25.25 | 4.8 | 140 | 0 |
| **D12** | Other | 45 | 0 | 0 | never | 27.32 | 4 | 159 | 0 |
| **D13** | Female | 64 | 0 | 0 | never | 36.71 | 5 | 100 | 0 |
| **D14** | Female | 67 | 0 | 0 | never | 27.32 | 6.2 | 159 | 1 |
| **D15** | Other | 23 | 0 | 0 | No Info | 24.23 | 6.1 | 140 | 0 |
| **D16** | Other | 53 | 0 | 0 | No Info | 27.32 | 6.6 | 160 | 0 |
| **D17** | Female | 68 | 0 | 0 | No Info | 40.31 | 7.5 | 260 | 1 |
| **D18** | Other | 10 | 0 | 0 | not current | 14.09 | 5 | 140 | 0 |
| **D19** | Other | 52 | 0 | 0 | not current | 31.63 | 6.6 | 159 | 0 |
| **D20** | Male | 50 | 1 | 0 | not current | 25.49 | 6.1 | 260 | 1 |

* Dữ liệu sau khi xử lý:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STT** | **gender** | **age** | **hypertension** | **heart\_disease** | **smoking\_history** | **bmi** | **HbA1c\_level** | **blood\_glucose\_level** | **diabetes** |
| **D1** | 1 | 34 | 1 | 0 | 0 | 35,41 | 6,6 | 100 | 0 |
| **D2** | 1 | 53 | 0 | 0 | 1 | 30,8 | 6,6 | 280 | 1 |
| **D3** | 1 | 55 | 0 | 0 | 1 | 27,32 | 6 | 80 | 0 |
| **D4** | 2 | 28 | 0 | 0 | 2 | 27,32 | 4,5 | 126 | 0 |
| **D5** | 0 | 62 | 1 | 1 | 4 | 43,16 | 8,8 | 280 | 1 |
| **D6** | 0 | 55 | 0 | 0 | 2 | 18,6 | 6 | 160 | 0 |
| **D7** | 1 | 6 | 0 | 1 | 2 | 17,17 | 6,5 | 240 | 1 |
| **D8** | 1 | 73 | 0 | 0 | 3 | 25,91 | 9 | 160 | 1 |
| **D9** | 0 | 12 | 0 | 0 | 0 | 30,58 | 6,2 | 240 | 1 |
| **D10** | 0 | 80 | 0 | 0 | 3 | 27,32 | 6,1 | 100 | 0 |
| **D11** | 1 | 31 | 0 | 0 | 3 | 25,25 | 4,8 | 140 | 0 |
| **D12** | 2 | 45 | 0 | 0 | 4 | 27,32 | 4 | 159 | 0 |
| **D13** | 0 | 64 | 0 | 0 | 4 | 36,71 | 5 | 100 | 0 |
| **D14** | 0 | 67 | 0 | 0 | 4 | 27,32 | 6,2 | 159 | 1 |
| **D15** | 2 | 23 | 0 | 0 | 0 | 24,23 | 6,1 | 140 | 0 |
| **D16** | 2 | 53 | 0 | 0 | 0 | 27,32 | 6,6 | 160 | 0 |
| **D17** | 0 | 68 | 0 | 0 | 0 | 40,31 | 7,5 | 260 | 1 |
| **D18** | 2 | 10 | 0 | 0 | 5 | 14,09 | 5 | 140 | 0 |
| **D19** | 2 | 52 | 0 | 0 | 5 | 31,63 | 6,6 | 159 | 0 |
| **D20** | 1 | 50 | 1 | 0 | 5 | 25,49 | 6,1 | 260 | 1 |

### Kmeans

* Lần 1:
  + Tâm cụm:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **C1** | 6 | 103 | 0 | 55 | 0 | 0 | 2 | 18,6 | 6 | 160 | 0 |
| **C2** | 14 | 96 | 0 | 67 | 0 | 0 | 4 | 27,32 | 6,2 | 159 | 1 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 1** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 65,80224996 | 68,21735923 | C1 |
| D2 | 120,6449336 | 121,898607 | C1 |
| D3 | 80,48626218 | 79,96899399 | C2 |
| D4 | 44,35412495 | 51,19462862 | C1 |
| D5 | 122,7437721 | 122,1706413 | C2 |
| D6 | 0 | 15,00261311 | C1 |
| D7 | 93,83653286 | 101,9368064 | C1 |
| D8 | 19,70878231 | 6,987710641 | C2 |
| D9 | 91,63274742 | 98,04400849 | C1 |
| D10 | 65,59000229 | 60,42358811 | C2 |
| D11 | 31,99472613 | 40,80741232 | C1 |
| D12 | 13,74912361 | 22,2 | C1 |
| D13 | 63,35591606 | 59,8298596 | C2 |
| D14 | 15,00261311 | 0 | C2 |
| D15 | 38,2584226 | 48,23440784 | C1 |
| D16 | 9,402042331 | 14,73634962 | C1 |
| D17 | 103,1822373 | 101,9236484 | C2 |
| D18 | 49,59173419 | 61,57493727 | C1 |
| D19 | 13,89751417 | 15,77136963 | C1 |
| D20 | 100,416543 | 102,4517394 | C1 |

* Lần 2:
  + Tâm cụm:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **C1** | 1,307692308 | 34,76923077 | 0,1538461538 | 0,07692307692 | 2,230769231 | 25,78538462 | 5,815384615 | 177,2307692 |
| **C2** | 0,2857142857 | 67 | 0,1428571429 | 0,1428571429 | 2,714285714 | 32,57857143 | 6,942857143 | 162,7142857 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 2** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 77,87313512 | 70,98484926 | C2 |
| D2 | 104,5049127 | 118,1469824 | C1 |
| D3 | 99,33345202 | 83,77162055 | C2 |
| D4 | 51,72099779 | 53,90777955 | C1 |
| D5 | 107,7971385 | 117,8963887 | C1 |
| D6 | 27,56157046 | 18,66251427 | C2 |
| D7 | 69,59426837 | 99,66683169 | C1 |
| D8 | 42,06381348 | 9,628171251 | C2 |
| D9 | 66,99452547 | 94,92163901 | C1 |
| D10 | 89,52754667 | 64,27012289 | C2 |
| D11 | 37,44824291 | 43,25356892 | C1 |
| D12 | 21,12627049 | 23,21080682 | C1 |
| D13 | 83,3301355 | 62,96586805 | C2 |
| D14 | 37,12891364 | 6,616331217 | C2 |
| D15 | 39,1488238 | 50,32588566 | C1 |
| D16 | 25,25304393 | 15,53976867 | C2 |
| D17 | 90,41873668 | 97,63752043 | C1 |
| D18 | 46,31705257 | 64,17744794 | C1 |
| D19 | 25,92707118 | 15,74856625 | C2 |
| D20 | 84,21030435 | 99,05028144 | C1 |

* Lần 3:
  + Tâm cụm:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | 1,090909091 | 35,27272727 | 0,1818181818 | 0,1818181818 | 2,363636364 | 27,79272727 | 6,009090909 | 205,9090909 |
| C2 | 0,7777777778 | 59,22222222 | 0,1111111111 | 0 | 2,444444444 | 28,61555556 | 6,455555556 | 130,8888889 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 3** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 106,221584 | 40,53744078 | C2 |
| D2 | 76,2564473 | 149,2641294 | C1 |
| D3 | 127,4536069 | 51,10323195 | C2 |
| D4 | 80,26132937 | 31,71647459 | C2 |
| D5 | 80,33031043 | 149,8789232 | C1 |
| D6 | 50,82031744 | 31,09047522 | C2 |
| D7 | 46,18450981 | 121,9430256 | C1 |
| D8 | 59,531179 | 32,42604739 | C2 |
| D9 | 41,45429083 | 118,935654 | C1 |
| D10 | 114,9745944 | 37,26352528 | C2 |
| D11 | 66,11105176 | 29,89890252 | C2 |
| D12 | 47,98867377 | 31,68817313 | C2 |
| D13 | 110,1202145 | 32,3670562 | C2 |
| D14 | 56,66812549 | 29,24908896 | C2 |
| D15 | 67,18488467 | 37,70812548 | C2 |
| D16 | 49,28440737 | 29,92246062 | C2 |
| D17 | 64,51859211 | 129,9660413 | C1 |
| D18 | 71,96766623 | 52,2205606 | C2 |
| D19 | 50,03180262 | 29,31790986 | C2 |
| D20 | 56,17553723 | 129,5067345 | C1 |

* Lần 4:
  + Tâm cụm:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **C1** | 0,5 | 41,83333333 | 0,3333333333 | 0,3333333333 | 2 | 31,25166667 | 6,95 | 260 |
| **C2** | 1,142857143 | 47,85714286 | 0,07142857143 | 0 | 2,571428571 | 26,83928571 | 5,892857143 | 134,5 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 4** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 160,2609775 | 38,25862648 | C2 |
| D2 | 22,94543897 | 145,6550074 | C1 |
| D3 | 180,5303128 | 54,99097991 | C2 |
| D4 | 134,8009415 | 21,67472229 | C2 |
| D5 | 30,93645481 | 147,1403982 | C1 |
| D6 | 101,6602208 | 27,76338365 | C2 |
| D7 | 43,39734058 | 113,91877 | C1 |
| D8 | 104,9074317 | 35,96026081 | C2 |
| D9 | 35,99324431 | 111,5257613 | C1 |
| D10 | 164,542843 | 47,17183649 | C2 |
| D11 | 120,662653 | 17,84215836 | C2 |
| D12 | 101,2010894 | 24,79935225 | C2 |
| D13 | 161,6459998 | 39,40078771 | C2 |
| D14 | 104,1866298 | 31,15086585 | C2 |
| D15 | 121,7012447 | 25,73578971 | C2 |
| D16 | 100,7310636 | 26,16834266 | C2 |
| D17 | 27,7762951 | 127,8591069 | C1 |
| D18 | 125,3920198 | 40,41525084 | C2 |
| D19 | 101,5681986 | 25,44606089 | C2 |
| D20 | 10,50805894 | 125,5527217 | C1 |

### Kmedoids

* Lần 1:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | 3 | 150 | 1 | 55 | 0 | 0 | 1 | 27,32 | 6 | 80 | 0 |
| C2 | 14 | 96 | 0 | 67 | 0 | 0 | 4 | 27,32 | 6,2 | 159 | 1 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 1** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 51,69 | 106,49 | C1 |
| D2 | 206,08 | 142,88 | C2 |
| D3 | 0 | 95,2 | C1 |
| D4 | 76,5 | 77,7 | C1 |
| D5 | 231,64 | 146,44 | C2 |
| D6 | 90,72 | 23,92 | C2 |
| D7 | 221,65 | 156,45 | C2 |
| D8 | 104,41 | 13,21 | C2 |
| D9 | 208,46 | 143,26 | C2 |
| D10 | 48,1 | 73,1 | C1 |
| D11 | 89,27 | 60,47 | C2 |
| D12 | 95 | 26,2 | C2 |
| D13 | 43,39 | 72,59 | C1 |
| D14 | 95,2 | 0 | C2 |
| D15 | 97,19 | 72,19 | C2 |
| D16 | 84,6 | 21,4 | C2 |
| D17 | 209,49 | 120,29 | C2 |
| D18 | 124,23 | 93,43 | C2 |
| D19 | 91,91 | 22,71 | C2 |
| D20 | 191,93 | 122,93 | C2 |

* + Chi phí: 1385,46
* Lần 2:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | 14 | 96 | 0 | 67 | 0 | 0 | 4 | 27,32 | 6,2 | 159 | 1 |
| C2 | D19 | 68484 | 2 | 52 | 0 | 0 | 5 | 31,63 | 6,6 | 159 | 0 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 2** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 106,49 | 87,78 | C2 |
| D2 | 142,88 | 127,83 | C2 |
| D3 | 95,2 | 91,91 | C2 |
| D4 | 77,7 | 66,41 | C2 |
| D5 | 146,44 | 149,73 | C1 |
| D6 | 23,92 | 22,63 | C2 |
| D7 | 156,45 | 146,56 | C2 |
| D8 | 13,21 | 33,12 | C1 |
| D9 | 143,26 | 129,45 | C2 |
| D10 | 73,1 | 95,81 | C1 |
| D11 | 60,47 | 51,18 | C2 |
| D12 | 26,2 | 14,91 | C2 |
| D13 | 72,59 | 80,68 | C1 |
| D14 | 0 | 22,71 | C1 |
| D15 | 72,19 | 60,9 | C2 |
| D16 | 21,4 | 11,31 | C2 |
| D17 | 120,29 | 133,58 | C1 |
| D18 | 93,43 | 80,14 | C2 |
| D19 | 22,71 | 0 | C2 |
| D20 | 122,93 | 111,64 | C2 |

* + Chi phí: 1428,28
* Lần 3:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | D1 | 65885 | 1 | 34 | 1 | 0 | 0 | 35,41 | 6,6 | 100 | 0 |
| C2 | D2 | 146 | 1 | 53 | 0 | 0 | 1 | 30,8 | 6,6 | 280 | 1 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 3** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 0 | 205,61 | C1 |
| D2 | 205,61 | 0 | C2 |
| D3 | 51,69 | 206,08 | C1 |
| D4 | 46,19 | 186,58 | C1 |
| D5 | 223,95 | 29,56 | C2 |
| D6 | 102,41 | 136,8 | C1 |
| D7 | 190,34 | 102,73 | C2 |
| D8 | 114,9 | 149,29 | C1 |
| D9 | 169,23 | 83,62 | C2 |
| D10 | 59,59 | 213,98 | C1 |
| D11 | 58,96 | 171,35 | C1 |
| D12 | 86,69 | 139,08 | C1 |
| D13 | 38,9 | 202,51 | C1 |
| D14 | 106,49 | 142,88 | C1 |
| D15 | 64,68 | 179,07 | C1 |
| D16 | 89,09 | 125,48 | C1 |
| D17 | 201,8 | 47,41 | C2 |
| D18 | 93,92 | 206,31 | C1 |
| D19 | 87,78 | 127,83 | C1 |
| D20 | 191,42 | 33,81 | C2 |

* + Chi phí: 1298,42
* Lần 4:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C1 | D3 | 150 | 1 | 55 | 0 | 0 | 1 | 27,32 | 6 | 80 | 0 |
| C2 | D4 | 71139 | 2 | 28 | 0 | 0 | 2 | 27,32 | 4,5 | 126 | 0 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 4** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 51,69 | 46,19 | C2 |
| D2 | 206,08 | 186,58 | C2 |
| D3 | 0 | 76,5 | C1 |
| D4 | 76,5 | 0 | C2 |
| D5 | 231,64 | 214,14 | C2 |
| D6 | 90,72 | 73,22 | C2 |
| D7 | 221,65 | 150,15 | C2 |
| D8 | 104,41 | 86,91 | C2 |
| D9 | 208,46 | 138,96 | C2 |
| D10 | 48,1 | 82,6 | C1 |
| D11 | 89,27 | 21,37 | C2 |
| D12 | 95 | 52,5 | C2 |
| D13 | 43,39 | 75,89 | C1 |
| D14 | 95,2 | 77,7 | C2 |
| D15 | 97,19 | 25,69 | C2 |
| D16 | 84,6 | 63,1 | C2 |
| D17 | 209,49 | 193,99 | C2 |
| D18 | 124,23 | 48,73 | C2 |
| D19 | 91,91 | 66,41 | C2 |
| D20 | 191,93 | 164,43 | C2 |

* + Chi phí: 1701,56
* Lần 5:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D5 | 1135 | 0 | 62 | 1 | 1 | 4 | 43,16 | 8,8 | 280 | 1 |
| D6 | 103 | 0 | 55 | 0 | 0 | 2 | 18,6 | 6 | 160 | 0 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 5** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 223,95 | 102,41 | C2 |
| D2 | 29,56 | 136,8 | C1 |
| D3 | 231,64 | 90,72 | C2 |
| D4 | 214,14 | 73,22 | C2 |
| D5 | 0 | 158,36 | C1 |
| D6 | 158,36 | 0 | C2 |
| D7 | 128,29 | 132,93 | C1 |
| D8 | 152,45 | 30,31 | C2 |
| D9 | 111,18 | 137,18 | C1 |
| D10 | 219,54 | 94,82 | C2 |
| D11 | 196,91 | 53,85 | C2 |
| D12 | 162,64 | 25,72 | C2 |
| D13 | 194,25 | 90,11 | C2 |
| D14 | 146,44 | 23,92 | C2 |
| D15 | 208,63 | 61,73 | C2 |
| D16 | 155,04 | 15,32 | C2 |
| D17 | 36,15 | 138,21 | C1 |
| D18 | 229,87 | 75,51 | C2 |
| D19 | 149,73 | 22,63 | C2 |
| D20 | 55,37 | 116,99 | C1 |

* + Chi phí: 1120,82
* Lần 6:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D7 | 60553 | 1 | 6 | 0 | 1 | 2 | 17,17 | 6,5 | 240 | 1 |
| D8 | 40 | 1 | 73 | 0 | 0 | 3 | 25,91 | 9 | 160 | 1 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 6** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 190,34 | 114,9 | C2 |
| D2 | 102,73 | 149,29 | C1 |
| D3 | 221,65 | 104,41 | C2 |
| D4 | 150,15 | 86,91 | C2 |
| D5 | 128,29 | 152,45 | C1 |
| D6 | 132,93 | 30,31 | C2 |
| D7 | 0 | 160,24 | C1 |
| D8 | 160,24 | 0 | C2 |
| D9 | 23,71 | 152,47 | C1 |
| D10 | 227,55 | 72,31 | C2 |
| D11 | 136,78 | 66,86 | C2 |
| D12 | 136,65 | 37,41 | C2 |
| D13 | 223,04 | 85,8 | C2 |
| D14 | 156,45 | 13,21 | C2 |
| D15 | 128,46 | 78,58 | C2 |
| D16 | 141,25 | 27,81 | C2 |
| D17 | 110,14 | 124,9 | C1 |
| D18 | 113,58 | 101,82 | C2 |
| D19 | 146,56 | 33,12 | C2 |
| D20 | 77,72 | 129,32 | C1 |

* + Chi phí: 1296,04
* Lần 7:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D9 | 46859 | 0 | 12 | 0 | 0 | 0 | 30,58 | 6,2 | 240 | 1 |
| D10 | 1367 | 0 | 80 | 0 | 0 | 3 | 27,32 | 6,1 | 100 | 0 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 7** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 169,23 | 59,59 | C2 |
| D2 | 83,62 | 213,98 | C1 |
| D3 | 208,46 | 48,1 | C2 |
| D4 | 138,96 | 82,6 | C2 |
| D5 | 111,18 | 219,54 | C1 |
| D6 | 137,18 | 94,82 | C2 |
| D7 | 23,71 | 227,55 | C1 |
| D8 | 152,47 | 72,31 | C2 |
| D9 | 0 | 214,36 | C1 |
| D10 | 214,36 | 0 | C2 |
| D11 | 129,73 | 93,37 | C2 |
| D12 | 125,46 | 99,1 | C2 |
| D13 | 203,33 | 27,49 | C2 |
| D14 | 143,26 | 73,1 | C2 |
| D15 | 119,45 | 105,09 | C2 |
| D16 | 126,66 | 92,5 | C2 |
| D17 | 87,03 | 189,39 | C1 |
| D18 | 126,69 | 128,33 | C1 |
| D19 | 129,45 | 95,81 | C2 |
| D20 | 70,19 | 195,83 | C1 |

* + Chi phí: 1446,3
* Lần 8:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D11 | 307 | 1 | 31 | 0 | 0 | 3 | 25,25 | 4,8 | 140 | 0 |
| D12 | 33805 | 2 | 45 | 0 | 0 | 4 | 27,32 | 4 | 159 | 0 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 8** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 58,96 | 86,69 | C1 |
| D2 | 171,35 | 139,08 | C2 |
| D3 | 89,27 | 95 | C1 |
| D4 | 21,37 | 52,5 | C1 |
| D5 | 196,91 | 162,64 | C2 |
| D6 | 53,85 | 25,72 | C2 |
| D7 | 136,78 | 136,65 | C2 |
| D8 | 66,86 | 37,41 | C2 |
| D9 | 129,73 | 125,46 | C2 |
| D10 | 93,37 | 99,1 | C1 |
| D11 | 0 | 37,87 | C1 |
| D12 | 37,87 | 0 | C2 |
| D13 | 86,66 | 90,39 | C1 |
| D14 | 60,47 | 26,2 | C2 |
| D15 | 14,32 | 50,19 | C1 |
| D16 | 49,87 | 15,6 | C2 |
| D17 | 178,76 | 146,49 | C2 |
| D18 | 35,36 | 69,23 | C1 |
| D19 | 51,18 | 14,91 | C2 |
| D20 | 143,54 | 112,93 | C2 |

* + Chi phí: 1342,4
* Lần 9:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D13 | 1407 | 0 | 64 | 0 | 0 | 4 | 36,71 | 5 | 100 | 0 |
| D14 | 96 | 0 | 67 | 0 | 0 | 4 | 27,32 | 6,2 | 159 | 1 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 9** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 38,9 | 106,49 | C1 |
| D2 | 202,51 | 142,88 | C2 |
| D3 | 43,39 | 95,2 | C1 |
| D4 | 75,89 | 77,7 | C1 |
| D5 | 194,25 | 146,44 | C2 |
| D6 | 90,11 | 23,92 | C2 |
| D7 | 223,04 | 156,45 | C2 |
| D8 | 85,8 | 13,21 | C2 |
| D9 | 203,33 | 143,26 | C2 |
| D10 | 27,49 | 73,1 | C1 |
| D11 | 86,66 | 60,47 | C2 |
| D12 | 90,39 | 26,2 | C2 |
| D13 | 0 | 72,59 | C1 |
| D14 | 72,59 | 0 | C2 |
| D15 | 100,58 | 72,19 | C2 |
| D16 | 87,99 | 21,4 | C2 |
| D17 | 174,1 | 120,29 | C2 |
| D18 | 119,62 | 93,43 | C2 |
| D19 | 80,68 | 22,71 | C2 |
| D20 | 189,32 | 122,93 | C2 |

* + Chi phí: 1351,45
* Lần 10:
  + Tâm:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D15 | 23266 | 2 | 23 | 0 | 0 | 0 | 24,23 | 6,1 | 140 | 0 |
| D16 | 31985 | 2 | 53 | 0 | 0 | 0 | 27,32 | 6,6 | 160 | 0 |

* + Tính:

|  |  |  |  |
| --- | --- | --- | --- |
| **Lần 10** | **Distance to** | | **Cluster** |
| **Data** | **C1** | **C2** |
| D1 | 64,68 | 89,09 | C1 |
| D2 | 179,07 | 125,48 | C2 |
| D3 | 97,19 | 84,6 | C2 |
| D4 | 25,69 | 63,1 | C1 |
| D5 | 208,63 | 155,04 | C2 |
| D6 | 61,73 | 15,32 | C2 |
| D7 | 128,46 | 141,25 | C1 |
| D8 | 78,58 | 27,81 | C2 |
| D9 | 119,45 | 126,66 | C1 |
| D10 | 105,09 | 92,5 | C2 |
| D11 | 14,32 | 49,87 | C1 |
| D12 | 50,19 | 15,6 | C2 |
| D13 | 100,58 | 87,99 | C2 |
| D14 | 72,19 | 21,4 | C2 |
| D15 | 0 | 53,59 | C1 |
| D16 | 53,59 | 0 | C2 |
| D17 | 184,48 | 130,89 | C2 |
| D18 | 29,24 | 82,83 | C1 |
| D19 | 60,9 | 11,31 | C2 |
| D20 | 155,26 | 112,33 | C2 |

* + Chi phí: 1262,11
* Lần 5 là tốt nhất

## Phân tích và so sánh

### Kmeans

* Với 3 thuộc tính: bmi, HbA1c\_level, blood\_glucose\_level

A graph of a graph

Description automatically generated with medium confidence

* Với tất cả thuộc tính:
  + Total square error: 120544666.60129748

### Kmedoids

* Với tất cả thuộc tính:
  + Total square error: 30381.780068442953
* Sử dụng 1000 bản ghi ngẫu nhiên và 3 cột thuộc tính: age, HbA1c\_level, blood\_glucose\_level:

A graph of blue and red triangles

Description automatically generated