TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CHUYÊN ĐỀ HỌC PHẦN**

**MÔN NHẬP MÔN TRÍ TUỆ NHÂN TẠO**

**ĐỀ TÀI:**

**XÂY DỰNG HỆ CHUYÊN GIA TƯ VẤN MUA LAPTOP**

**Sinh viên thực hiện: VŨ CHUNG HIẾU - 19810310232**

**ĐỖ THỊ MAI HƯƠNG – 19810310349**

**NGUYỄN THỊ LOAN - 19810310070**

**Giảng viên hướng dẫn: PHẠM THỊ KIM DUNG**

**Ngành: CÔNG NGHỆ THÔNG TIN**

**Chuyên ngành: CÔNG NGHỆ PHẦN MỀM**

**Lớp: D14CNPM3**

**Khóa: 2019 – 2024**

**PHIẾU CHẤM ĐIỂM**

Sinh viên thực hiện:

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Họ và tên** | **Điểm** | **Chữ ký** |
| **1** | VŨ CHUNG HIẾU |  |  |
| **2** | ĐỖ THỊ MAI HƯƠNG |  |  |
| **3** | NGUYỄN THỊ LOAN |  |  |

Giảng viên chấm thi:

|  |  |  |
| --- | --- | --- |
| **Họ và tên giảng viên** | **Chữ ký** | **Ghi chú** |
| Giảng viên chấm 1: |  |  |
| Giảng viên chấm 2: |  |  |

**MỤC LỤC**

[LỜI NÓI ĐẦU 1](#_Toc87119375)

[CHƯƠNG 1: TỔNG QUAN VỀ HỆ CHUYÊN GIA 2](#_Toc87119376)

[1.1.Khái niệm 2](#_Toc87119377)

[1.2.Thành phần của hệ chuyên gia 2](#_Toc87119378)

[1.3.Đặc trưng và ưu nhược điểm của hệ chuyên gia 2](#_Toc87119379)

[1.4.Ứng dụng thực tiễn 4](#_Toc87119380)

[CHƯƠNG 2: ỨNG DỤNG CÂY QUYẾT ĐỊNH VÀO XÂY DỰNG HỆ CHUYÊN GIA TƯ VẤN MUA LAPTOP 6](#_Toc87119382)

[2.1.Khái niệm 6](#_Toc87119383)

[2.2.Thuật toán tạo cây quyết định 7](#_Toc87119386)

[2.3.Các kiểu cây quyết định 8](#_Toc87119387)

[2.4.Thuật toán ID3 8](#_Toc87119388)

[2.5.Áp dụng thuật toán ID3 vào chương trình 11](#_Toc87119389)

[CHƯƠNG 3: CÀI ĐẶT 27](#_Toc87119392)

[3.1.Công nghệ sử dụng 27](#_Toc87119393)

[3.2.Cơ sở dữ liệu 27](#_Toc87119394)

[3.3.Cài đặt 29](#_Toc87119397)

[KẾT LUẬN 33](#_Toc87119401)

[TÀI LIỆU THAM KHẢO 34](#_Toc87119402)

**DANH MỤC HÌNH ẢNH**

[Hình 2.1 Mô tả cây quyết định 7](#_Toc87119217)

[*Hình 2.2 Bảng dữ liệu* 11](#_Toc87119222)

[*Hình 2.2 Nhánh Windows* 25](#_Toc87119223)

[*Hình 3.1 dbo Bảng DuLieu* 28](#_Toc87119227)

[*Hình 3.2 dbo Bảng Luat* 28](#_Toc87119228)

[*Hình 3.3 Form chính* 31](#_Toc87119230)

[*Hình 3.4 Kết quả sau khi được tư vấn* 31](#_Toc87119231)

[*Hình 3.5 Bảng quản lý dữ liệu* 32](#_Toc87119232)

# 

# **LỜI NÓI ĐẦU**

Thế giới ngày nay phát triển mạnh mẽ với các hoạt động vô cùng đa dạng và phức tạp đòi hỏi khả năng giải quyết vấn đề ở mức độ trí tuệ nhân tạo ngày càng cao. Lĩnh vực trí tuệ nhận tạo nói chung và hệ chuyên gia nói riêng góp phần tạo ra các hệ thống có khả năng trí tuệ của con người, có được tri thức tiên tiến của các hệ chuyên gia để giải quyết các vấn đề phức tạp trong cuộc sống .Hệ chuyên gia được thu hút mãnh mẽ vì những ưu điềm sau: Các chương trình hệ chuyên gia ngày càng tỏ ra hữu hiệu và tiện lợi đáp ứng nhu cầu thực tế, Các chương trình hệ chuyên gia ngày càng tỏ ra có tính khả thi cao, Hệ chuyên gia không có tính đơn lẻ ,phù hợp với nhiều cá nhân.

Đặc biệt các sản phẩm của công nghệ thông tin ngày nay còn đóng vai trò như một chuyên gia trong nhiều lĩnh vực như y học, hóa học, dự báo,..có thể mã hóa các tri thức cũng như lập luận của những chuyên gia này và đưa vào máy tính nhằm hỗ trợ người sử dụng. Chính vì vậy mà chúng em chọn đề tài “*Xây dựng hệ chuyên gia tư vấn mua laptop”* cho môn học này. Chúng em rất mong nhận được sự ủng hộ của quý thầy, cô để chúng em đạt được kết quả tốt nhất. Chúng em xin chân thành cảm ơn cô giáo Phạm Thị Kim Dung Tạo đã hướng dẫn, giúp đỡ chúng em trong quá trình học tập. Cô đã giúp trang bị cho chúng em kiến thức môn học và hơn cả là động lực để tiếp tục trên con đường học tập.

# **CHƯƠNG 1: TỔNG QUAN VỀ HỆ CHUYÊN GIA**

## **1.1.Khái niệm**

Hệ chuyên gia, còn gọi là hệ thống dựa tri thức, là một chương trình máy tính chứa một số tri thức đặc thù của một hoặc nhiều chuyên gia con người về một chủ đề cụ thể nào đó.

Một chương trình ứng dụng được xây dựng dựa trên cơ sở tri thức và mô tơ suy diễn. Trong đó cơ sở tri thức được lấy từ nguồn tri thức. Có hai loại là xin ý kiến từ các chuyên gia trong lĩnh vực đó, cũng có thể lấy theo cách thứ hai đó là tổng hợp từ các tài liệu chuyên môn. Còn mô tơ suy diễn phụ thuộc vào người dùng do người dùng đưa ra.

## **1.2.Thành phần của hệ chuyên gia**

+ Cơ sở tri thức: Là nơi lưu trữ biểu diễn các tri thức mà hệ đảm nhận, làm cơ sở cho các hoạt động của hệ.

+ Động cơ suy diễn: Là quá trình sử dụng các tri thức từ cơ sở tri thức để khớp hoặc suy diễn các tri thức đó để rút ra kết luận cho vấn đề cần giải quyết, sử dụng các đối tượng trong cơ sở tri thức để giải quyết các vấn đề cần suy diễn.

+ Giao diện người – máy: Là nơi người sử dụng và hệ chuyên gia trao đổi với nhau.

+ Bộ thu nạp tri thức: Lấy tri thức từ hai nguồn (chuyên gia và kho dữ liệu tri thức) để bổ sung cho cơ sở tri thức. Đồng thời bộ thu nạp tri thức còn giúp loại bỏ các tri thức lỗi thời, mẫu thuẫn.

## **1.3.Đặc trưng và ưu nhược điểm của hệ chuyên gia**

**- Đặc trưng:**

* Hiệu quả cao (high performance): Khả năng trả lời với mức độ tinh thông bằng hoặc cao hơn so với hệ chuyên gia (người) trong cùng lĩnh vực.
* Thời gian trả lời thỏa đáng (adequate response time): Thời gian trả lời hợp lý, bằng hoặc nhanh hơn so với hệ chuyên gia (người) để đi đến cùng một quyết định. Hệ chuyên gia là một hệ thống thời gian thực.
* Độ tin cậy cao (good reliability): Không thể xảy ra sự cố hoặc giảm sút độ tin cậy khi sử dụng.
* Dễ hiễu (understandable): Hệ chuyên gia giải thích các bước suy luận một cách dễ hiểu và nhất quán , không giống như cách trả lời bí ẩn của các hộp đen

**- Ưu điểm:**

+ Phổ cập (increased availability): Là sản phẩm chuyên gia, được phát triển không ngừng với hiệu quả sử dụng không thể phủ nhận.

+ Giảm giá thành (reduced cost).

+ Giảm rủi ro (reduced dangers): Giúp con người tránh được trong các môi trường rủi ro, nguy hiểm.

+ Tính thường trực (Permanance): Bất kể lúc nào cũng có thể khai thác sử dụng, trong khi con người có thể mệt mỏi, nghỉ ngơi hay vắng mặt.

+ Đa lĩnh vực (multiple expertise): Chuyên gia về nhiều lĩnh vực khác nhau và được khai thác đồng thời bất kể thời gian sử dụng.

+ Độ tin cậy (increased relialility): Luôn đảm bảo độ tin cậy khi khai thác.

+ Khả năng giảng giải (explanation): Câu trả lời với mức độ tinh thông được giảng giải rõ ràng chi tiết, dễ hiểu.

+ Khả năng trả lời (fast reponse): Trả lời theo thời gian thực, khách quan.

+ Tính ổn định, suy luận có lý và đầy đủ mọi lúc mọi nơi (steady, une motional, and complete response at all times).

+ Trợ giúp thông minh như một người hướng dẫn (intelligent -tutor)

+ Có thể truy cập như là một cơ sở dữ liệu thông minh (intelligent database).

**- Nhược điểm:**

+Hệ chuyên gia có thể hoạt động tốt trong một lĩnh vực hẹp với một độ phức tạp nhất định. Việc mô hình hóa một hệ chuyên gia cho nhiều lĩnh vực có thể gây tác dụng ngược lại.

+Hệ chuyên gia không hoàn toàn đảm bảo hay tính đúng đắn của các luật trong quá trình suy luận.

+Việc thêm các luật, các tri thức mới có thể sẽ xung đột với những luật hiện tại và có thể giảm độ tin cậy của hệ thống.

## **1.4.Ứng dụng thực tiễn**

Cho đến nay, hàng trăm hệ chuyên gia đã được xây dựng và đã được báo cáo thường xuyên trong các tạp chí, sách, báo và hội thảo khoa học. Ngoài ra còn các hệ chuyên gia được sử dụng trong các công ty, các tổ chức quân sự mà không được công bố vì lý do bảo mật. Bảng dưới đây liệt kê một số lĩnh vực ứng dụng diện rộng của các hệ chuyên gia.

|  |  |
| --- | --- |
| **Lĩnh vực** | **Ứng dụng** |
| Cấu hình | Tập hợp thích đáng những thành phần của một hệ thống theo cách riêng |
| Chuẩn đoán | Lập luận dựa trên những chứng cứ quan sát được |
| Truyền đạt | Dạy học kiểu thông minh sao cho sinh viên có thể hỏi vì sao (why?), như thế nào (how?) và cái gì nếu (what if?) giống như hỏi một người thầy giáo |
| Giải thích | Giải thích những dữ liệu thu nhận được |
| Kiểm tra | So sánh dữ liệu thu lượm được với dữ liệu chuyên môn để đánh giá kết quả |
| Lập kế hoạch | Lập kế hoạch sản xuất theo yêu cầu |
| Dự đoán | Dự đoán hậu quả từ một tình huống xảy ra |
| Chữa trị | Chỉ định cách thụ lý một vấn đề |
| Điều khiển | Điều khiển một quá trình, đòi hỏi diễn giải, chẩn đoán, kiểm tra, lập kế hoạch, dự đoán và chữa trị |

### *Bảng 1.1 Ứng dụng hệ chuyên gia trong thực tiễn*

# **CHƯƠNG 2: ỨNG DỤNG CÂY QUYẾT ĐỊNH VÀO XÂY DỰNG HỆ CHUYÊN GIA TƯ VẤN MUA LAPTOP**

## **2.1. Khái niệm**

Việc học trên cây quyết định dựa trên dữ liệu huấn luyện có dạng bảng, trong đó các cột là thuộc tính diều kiện và thuộc tính mục tiêu (quyết định), mỗi dòng sẽ được dán nhãn là giá trị của thuộc tính mục tiêu. Chẳng hạn với bảng dữ liệu dưới đây có các thuộc tính “Age” và “Salary” là thuộc tính điều kiện còn thuộc tính “Class” là thuộc tính mục tiêu.

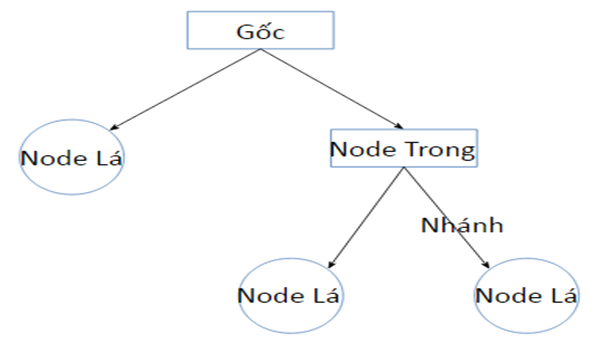
|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Age** | **Salary** | **Class** |
| 1 | 30 | 65 | Good |
| 2 | 23 | 15 | Bad |
| 3 | 40 | 75 | Good |
| 4 | 55 | 40 | Bad |
| 5 | 55 | 100 | Good |
| 6 | 45 | 60 | Good |

### *Bảng 2.1 Ví dụ dữ liệu huấn luyện*

Cây quyết định là một công cụ khá phổ biến trong nhiều dạng ứng dụng, với cơ chế rút trích các dạng luật if- then dựa trên tập liệu mẫu.

Cây quyết định là dạng đặc biệt của cáu trúc cây nhằm hỗ trợ cho tiến trình ra quyết định. Cây gồm các thành phần:

* Nút trong: Biểu diễn thuộc tính.
* Nhánh: Biểu diễn giá trị cụ thể của thuộc tính đó.
* Nút lá: Đại diện cho giá trị dự đoán của thuộc tính mục tiêu.



### Hình 2.1 Mô tả cây quyết định

## **2.2. Thuật toán tạo cây quyết định**

* Ý tưởng chính của thuật toán
* Phương pháp “tham lam” (greedy).
* Phân chia tập mẫu dựa trên thuộc tính cho kết quả tối ưu hóa tiêu chuẩn.
* Vấn đề:
* Xác định cách phân chia các mẫu.
* Dựa trên độ đo sự đồng nhất của dữ liệu.
* Điều kiện dừng:
* Tất cả các mẫu rơi vào một nút thuộc về cùng một lớp.
* Không còn thuộc tính nào có thể dùng để phân chia mẫu nữa.
* Không còn lại mẫu nào tại nút.
* Cách phân chia mẫu:
* Tiêu chuẩn phân chia: tạo ra các nhóm sao cho một lớp chiếm ưu thế trong từng nhóm.
* Thuộc tính được chọn là thuộc tính cho độ đo tốt nhất, có lợi cho quá trình phân lớp.
* Độ đo để đánh giá chất lượng phân chia là độ đo sự đồng nhất:
* Entropy (Information Gain)
* InformationGainRatio
* GiniIndex

## **2.3 Các kiểu cây quyết định**

Cây hồi quy (Regresstion tree) :Ước lượng các hàm giá có giá trị là số thực thay vì được sử dụng thay vì được sử dụng cho các nhiệm vụ phan loại.

Cây phân loại (Classification tree): Nếu biến chúng ta cần tìm hiểu là biến phân loại như :giới (tính nam hay nữ ), kết quả của một trận đấu( thắng hay thua).

## **2.4 Thuật toán ID3**

Với bảng dữ liệu huấn luyện có nhiều thuộc tính điều kiện. Vấn đề đặt ra là lựa chọn thuộc tính nào làm gốc của cây. Để lựa chọn thuộc tính làm gốc của cây ta đưa ra tiêu chuẩn để đánh giá, một trong các tiêu chuẩn đó là độ lợi thông tin (information Gain) của thuộc tính. Thuộc tính có độ lợi thông tin lớn nhất sẽ được chọn làm gốc của cây.

Độ lợi thông tin của thuộc tính A, ký hiệu là Gain( S, A) và được xác định qua công thức:

**Gain (S, A) =Entropy(S) -∑((|Sv | / |S|)\* Entropy(Sv) )**

**Thuật toán ID3:**

Cho X là tập vectơ đặc trưng( không gian đặc trưng):

X={x1, x2,...,xn}

C là tập các giá trị phân lớp

C: X ─> C thực hiện gán nhãn cho các phần tử trong X

D= { D = {(x1, c(x1)), (x2, c(x2)),…, (xn, c(xn))} ⊆ X × C}

Nhiệm vụ: Dựa trên D, hãy xây dựng cây quyết định T xấp xỉ c.

**ID3( D, Atributes, Target)**

{

* Tạo nút t cho cây.
* Nếu tất cả các mẫu trong D đều thuộc lớp dương thì trả về cây có nút gốc t và được dán nhãn “+”.
* Nếu tất cả các mẫu trong D đều thuộc lớp âm thì trả về cây có nút gốc t và được dán nhãn “-”.
* Đặt t là nhãn phổ biến nhất của thuộc tính mục tiêu trong D.
* Nếu Atributes rỗng thì trả về cây có nút gốc trỏ bởi t.
* Trường hợp khác:
* Đặt A\* là thuộc tính phân lớp tốt nhất trong D.
* Thuộc tính quyết định của t là A\*.
* Với mỗi giá trị a của A\*  thực hiện:
  + - Bổ sung nhánh mới dưới t ứng với A\* = “a”.
    - Đặt D\_a là tập con của D chứa các mẫu mà A\* = “a”.
    - Nếu D\_a rỗng thì nhánh mới này bổ sung nút lá với nhãn phổ biến nhất trong D.
    - Ngược lại, dưới nhánh mới này bổ sung cây con ID3( D\_a, Atributes – { A\*}, Target).
* Trả về t.

}

**Sau đây là mã giả cho thuật giải ID3:**

**ID3( D, Atributes, Target)**

**{**

1. t= createNode().
2. IF IF ∀(x, c(x)) ∈ D: c(x) = 1 THEN label(t) = ‘+’ return(t) ENDIF.
3. IF ∀(x, c(x)) ∈ D: c(x) = 0 THEN label(t) = ‘-’ return(t) ENDIF.
4. label(t) = mostCommonClass(D, Target).
5. IF Attributes = ∅ THEN return(t) ENDIF.
6. A\* = argmaxA∈Attributes(Gain(D, A)).
7. FOREACH a∈A\* DO

Da = {(x, c(x)) ∈ D: x| A\* = a}

IF Da = ∅ THEN

t’ = createNode()

label(t’) = mostCommonClass(D, Target)

createEdge(t, a, t’)

ELSE

createEdge(t, a, ID3(Da, Attributes – {A\*}, Target))

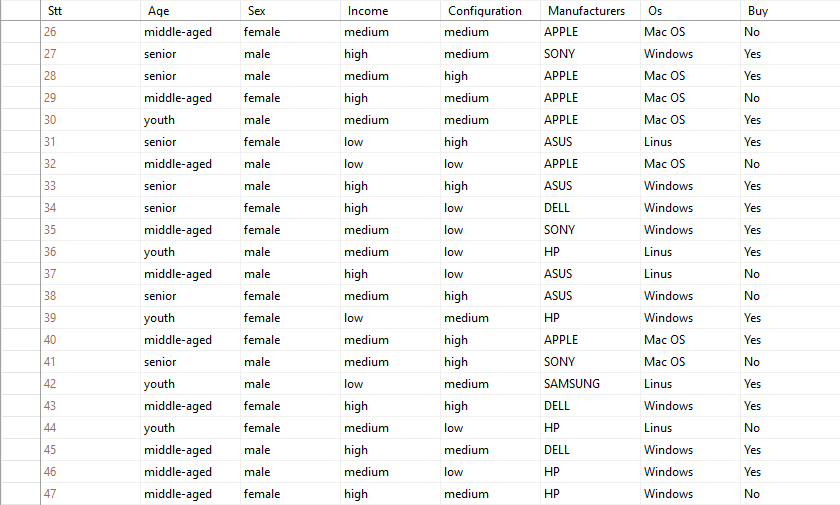
ENDIF

ENDDO

1. return(t)

**}**

## **2.5. Áp dụng thuật toán ID3 vào chương trình**



### *Hình 2.2 Bảng dữ liệu*

Để tối giản nhất có thể, chúng ta lựa chọn ba hệ điều hành phổ biến nhất hiện nay được nhiều người sử dụng là Windows, Mac OS, Linux.

**Sử dụng giải thuật ID3 để tạo cây quyết định**

Tập hợp S là tập có 47 mẫu, trong đó có 27 mẫu thuộc lớp “Yes” và 20 mẫu thuộc lớp “ No”.

Vậy ta có : Entropy( S )= -(27/47)\* log2(27/47) - (20/47)\* log2(20/47 )

* Entropy( S ) = 0.984

Tính độ lợi thông tin của các thuộc tính trong tập dữ liệu trong mẫu dựa vào thuật giải ID3 :

* **Xét thuộc tính “Age”:**

Entropy(Syouth) = - (9/16) log2(9/16) – (7/16) log2(7/16) = 0.989

Entropy(Smiddle-aged) = - (7/15) log2(7/15) - (8/15) log2(8/15) = 0.997

Entropy(Ssenior) = - (11/16) log2(11/16) - (5/16) log2(5/16) = 0.896

→ Gain( S, Age)= Entropy( S) – (16/47)\* Entropy(Syouth) – (15/47)\* Entropy(Smiddle-aged) – (16/47)\* Entropy(Ssenior)

= 0.984 – (16/47) \* 0.989 – (15/47) \* 0.997 – (16/47) \* 0.896 **= 0.024**

* **Xét thuộc tính “Sex”:**

Entropy(Smale) = - (17/26)\* log2(17/26) – (9/26)\* log2(9/26) = 0.931

Entropy(Sfemale) = - (10/21)\* log2(10/21) – (11/21)\* log2(11/21) = 0.998

→ Gain(S, Sex) = Entropy(S) – (26/47) \* Entropy(Smale) – (21/47) \* Entropy(Sfemale)

= 0.984 – (26/47) \* 0.931 – (21/47) \* 0.998 **= 0.023**

* **Xét thuộc tính “Income”:**

Entropy(Shigh) = - (9/15)\* log2(9/15) – (6/15)\* log2(6/15) = 0.971

Entropy(Smedium) = - (12/18)\* log2(12/18) – (6/18)\* log2(6/18) = 0.918

Entropy(Slow) = - (6/14)\* log2(6/14) – (8/14)\* log2(8/14) = 0.985

→ Gain(S, Income) = Entropy(S) – (15/47) \* Entropy(Shigh) – (18/47) \* Entropy(Smedium) – (14/47) \* Entropy(Slow)

= 0.984 – (15/47) \* 0.971 – (18/47) \* 0.918 – (14/47) \* 0.985 **= 0.029**

* **Xét thuộc tính “Configuration”:**

Entropy(Shigh) = - (8/17)\* log2(8/17) – (9/17)\* log2(9/17) = 0.998

Entropy(Smedium) = - (10/17)\* log2(10/17) – (7/17)\* log2(7/17) = 0.977

Entropy(Slow) = - (9/13)\* log2(9/13) – (4/13)\* log2(4/13) = 0.890

→ Gain(S, Configuration) = Entropy(S) – (17/47) \* Entropy(Shigh) – (17/47) \* Entropy(Smedium) – (13/47) \* Entropy(Slow)

= 0.984 – (17/47) \* 0.998 – (17/47) \* 0.977 – (13/47) \* 0.890 **= 0.023**

* **Xét thuộc tính “Manufacturers”:**

Entropy(SSONY) = - (3/5)\* log2(3/5) – (2/5)\* log2(2/5) = 0.971

Entropy(SAPPLE) = - (7/13)\* log2(7/13) – (6/13)\* log2(6/13) = 0.996

Entropy(SDELL) = - (6/9)\* log2(6/9) – (3/9)\* log2(3/9) = 0.918

Entropy(SASUS) = - (4/8)\* log2(4/8) – (4/8)\* log2(4/8) = 1

Entropy(SSAMSUNG) = - (2/3)\* log2(2/3) – (1/3)\* log2(1/3) = 0.918

Entropy(SHP) = - (5/9)\* log2(5/9) – (4/9)\* log2(4/9) = 0.991

→ Gain(S, Manufacturers) = Entropy(S) – (5/47) \* Entropy(SSONY) – (13/47) \* Entropy(SAPPLE) – (9/47) \* Entropy(SDELL) – (8/47) \* Entropy(SASUS) – (3/47) \* Entropy(SSAMSUNG) – (9/47) \* Entropy(SHP)

= 0.984 – (5/47) \* 0.971 – (13/47) \* 0.996 – (9/47) \* 0.918 – (8/47) \* 1– (3/47) \* 0.918 – (9/47) \* 0.991 **= 0.011**

* **Xét thuộc tính “OS”:**

Entropy(SWindows) = - (15/22)\* log2(15/22) – (7/22)\* log2(7/22) = 0.902

Entropy(SMac OS) = - (7/14)\* log2(7/14) – (7/14)\* log2(7/14) = 1

Entropy(SLinux) = - (5/11)\* log2(5/11) – (6/11)\* log2(6/11) = 0.994

→ Gain(S, OS) = Entropy(S) – (22/47) \* Entropy(SWindows)

– (14/47) \* Entropy(SMac OS) – (11/47) \* Entropy(SLinux)

= 0.984 – (22/47) \* 0.902 – (14/47) \* 1 – (11/47) \* 0.994 **= 0.031**

Trong các thuộc tính điều kiện trên ta thấy “OS” có độ lợi thông tin lớn nhất nên thuộc tính này sẽ được chọn làm gốc của cây.

Nhánh OS = Windows

Trong nhánh **OS= Windows**, ta có:

SWindows = {1, 2, 6, 11, 12, 13, 14, 16, 17, 22, 23, 24, 27, 33, 34, 35, 38, 39, 43, 45, 46, 47}

**Ta có: Entropy(SWindows) = 0.902**

* **Xét thuộc tính Age:**

Entropy(Syouth) = - (5/8)\* log2(5/8) – (3/8)\* log2(3/8) = 0.954

Entropy(Smiddle-aged) = - (4/5)\*log2(4/5) – (1/5)\* log2(1/5) = 0.722

Entropy(Ssenior) = - (6/9)\* log2(6/9) – (3/9)\* log2(3/9) = 0.918

→ Gain(SWindows, Age) = Entropy(SWindows) – (8/22) \* Entropy(Syouth) – (5/22) \* Entropy(Smiddle-aged) – (9/22) \* Entropy(Ssenior)

= 0.902 – (8/22) \* 0.954 – (5/22) \* 0.722 – (9/22) \* 0.918 **= 0.015**

* **Xét thuộc tính Sex:**

Entropy(Smale) = - (9/10)\* log2(9/10) – (1/10)\* log2(1/10) = 0.469

Entropy(Sfemale) = - (6/12)\* log2(6/12) – (6/12)\* log2(6/12) = 1

→ Gain(SWindows, Sex) = Entropy(SWindows) – (10/22) \* Entropy(Smale) - (12/22) \* Entropy(Sfemale)

= 0.902 – (10/22) \* 0.469 – (12/22) \* 1 **= 0.131**

* **Xét thuộc tính Income:**

Entropy(Shigh) = - (8/10)\* log2(8/10) – (2/10)\* log2(2/10) = 0.722

Entropy(Smedium) = - (5/7)\* log2(5/7) – (2/7)\* log2(2/7) = 0.863

Entropy(Slow) = - (2/5)\* log2(2/5) – (3/5)\* log2(3/5) = 0.971

→ Gain(SWindows, Income) = Entropy(SWindows) – (10/22) \* Entropy(Shigh) – (7/22) \* Entropy(Smedium) – (5/22) \* Entropy(Slow)

= 0.902 – (10/22) \* 0.722 – (7/22) \* 0.863 – (5/22) \* 0.971

**= 0.079**

* **Xét thuộc tính Configuration:**

Entropy(Shigh) = - (3/8)\* log2(3/8) – (5/8)\* log2(5/8) = 0.954

Entropy(Smedium) = - (6/8)\* log2(6/8) – (2/8)\* log2(2/8) = 0.811

Entropy(Slow) = - (6/6)\* log2(6/6) = 0

→ Gain(SWindows, Configuration) = Entropy(SWindows) – (8/22) \* Entropy(Shigh) – (8/22) \* Entropy(Smedium) – (6/22) \* Entropy(Slow)

= 0.902 – (8/22) \* 0.954 – (8/22) \* 0.811 – (6/22) \* 0

= 0.260

* **Xét thuộc tính Manufacturers:**

Entropy(SSONY) = - (3/4)\* log2(3/4) – (1/4)\* log2(1/4) = 0.811

Entropy(SHP) = - (3/4)\* log2(3/4) – (1/4)\* log2(1/4) = 0.811

Entropy(SDELL) = - (6/8)\* log2(6/8) – (2/8)\* log2(2/8) = 0.811

Entropy(SASUS) = - (3/5)\* log2(3/5) – (2/5)\* log2(2/5) = 0.971

Entropy(SSAMSUNG) = - (1/1)\* log2(1/1) = 0

→ Gain(SWindows, Manufacturers) = Entropy(SWindows) – (4/22) \* Entropy(SSONY) – (4/22) \* Entropy(SHP) – (8/22) \* Entropy(SDELL) – (5/22) \* Entropy(SASUS) – (1/22) \* Entropy(SSAMSUNG)

= 0.902 – (4/22) \* 0.811 – (4/22) \* 0.811 – (8/22) \* 0.811 – (5/22) \* 0.971 – (1/22) \* 0 **= 0.092**

Trong các thuộc tính điều kiện trên ta thấy **“Configuration”** có độ lợi thông tin lớn nhất nên thuộc tính này sẽ được chọn làm gốc.

Bây giờ ta tiếp tục xét lần lượt từng nhánh trong **Configuration.**

Nhánh Configuration = high

Trong nhánh **Configuration = high**, ta có:

Shigh = {1, 6, 11, 14, 23, 33, 38, 43}.

Ta có: **Entropy(Shigh) = 0.954**

* + - **Xét thuộc tính Age:**

Entropy(Syouth) = - (1/4)\* log2(1/4) – (3/4)\* log2(3/4) = 0.811

Entropy(Smiddle-aged) = - (1/2)\* log2(1/2) – (1/2)\* log2(1/2) = 1

Entropy(Ssenior) = - (1/2)\* log2(1/2) – (1/2)\* log2(1/2) = 1

→ Gain(Shigh, Age) = Entropy(Shigh) – (4/8) \* Entropy(Syouth) – (2/8) \* Entropy(Smiddle-aged) – (2/8) \* Entropy(Ssenior)

= 0.954 – (4/8) \* 0.811 – (2/8) \* 1 – (2/8) \* 1 **= 0.049**

* + - **Xét thuộc tính Sex:**

Entropy(Smale) = - (2/3)\* log2(2/3) – (1/3)\* log2(1/3) = 0.918

Entropy(Sfemale) = - (1/5)\* log2(1/5) – (4/5)\* log2(4/5) = 0.722

→ Gain(Shigh, Sex) = Entropy(Shigh) – (3/8) \* Entropy(Smale) –

(5/8) \* Entropy(Sfemale)

= 0.954 – (3/8) \* 0.918 – (5/8) \* 0.722 **= 0.159**

* + - **Xét thuộc tính Income:**

Entropy(Shigh) = - (3/4)\* log2(3/4) – (1/4)\* log2(1/4) = 0.811

Entropy(Smedium) = - (2/2)\* log2(2/2) = 0

Entropy(Slow) = - (2/2)\* log2(2/2) = 0

→ Gain(Shigh, Income) = Entropy(Shigh) – (4/8) \* Entropy(Shigh) – (2/8) \* Entropy(Smedium) – (2/8) \* Entropy(Slow)

= 0.954 – (4/8) \* 0.811 – (2/8) \* 0 – (2/8) \* 0 **= 0.549**

* + - **Xét thuộc tính Manufacturers:**

Entropy(SSONY) = - (1/1)\* log2(1/1) = 0

Entropy(SDELL) = - (1/3)\* log2(1/3) - (2/3)\* log2(2/3) = 0.918

Entropy(SASUS) = - (1/3)\* log2(1/3) - (2/3)\* log2(2/3) = 0.918

Entropy(SSAMSUNG) = - (1/1)\* log2(1/1) = 0

→ Gain(Shigh, Manufacturers) = Entropy(Shigh) –

(1/8) \* Entropy(SSONY) – (3/8) \* Entropy(SDELL) – (3/8) \* Entropy(SASUS) – (1/8) \* Entropy(SSAMSUNG)

= 0.954 – (1/8) \* 0 – (3/8) \* 0.918 – (3/8) \* 0.918 – (1/8) \* 0

**= 0.266**

Trong các thuộc tính điều kiện trên ta thấy **Income** có độ lợi thông tin lớn nhất nên thuộc tính này được chọn làm gốc.

Bây giờ ta tiếp tục xét từng nhánh trong Income.

Nhánh Income = high

Trong nhánh Incom= High, Shigh = {1, 6, 33, 43}.

Entropy(Shigh) = 0.811

* + - **Xét thuộc tính Age:**

Entropy(Syouth) = - (1/2)\* log2(1/2) – (1/2)\* log2(1/2) = 1

Entropy(Smiddle-aged) = - (1/1)\* log2(1/1) = 0(YES)

Entropy(Ssenior) = - (1/1)\* log2(1/1) = 0(YES)

→ Gain(Shigh, Age) = Entropy(Shigh) – (2/4) \* Entropy(Syouth) – (1/4) \* Entropy(Smiddle-aged) – (1/4) \* Entropy(Ssenior)

= 0.811 – (2/4) \* 1 – (1/4) \* 0 – (1/4) \* 0 **= 0.311**

* + - **Xét thuộc tính Sex:**

Entropy(Smale) = - (2/2)\* log2(2/2) = 0(yES)

Entropy(Sfemale) = - (1/2)\* log2(1/2) - (1/2) log2(1/2) = 1

→ Gain(Shigh, Sex) = Entropy(Shigh) – (2/4) \* Entropy(Smale) - (2/4) \* Entropy(Sfemale) = 0.811 – (2/4) \* 0 – (2/4) \* 1 **= 0.311**

* + - **Xét thuộc tính Manufacturers:**

Entropy(SSONY) = - (1/1)\* log2(1/1) = 0

Entropy(SDELL) = - (1/2)\* log2(1/2) - (1/2)\* log2(1/2) = 1

Entropy(SASUS) = - (1/1)\* log2(1/1) = 0

→ Gain(Shigh, Manufacturers) = Entropy(Shigh) – (1/4) \* Entropy(SSONY) – (2/4) \* Entropy(SDELL) – (1/4) \* Entropy(SASUS)

= 0.811 – (1/4) \* 0 – (2/4) \* 1 – (1/4) \* 0 **= 0.311**

Cả 3 thuộc tính trên đều có độ lợi thông tin bằng nhau nên ta chọn 1 trong 3 thuộc tính để làm gốc. Chọn thuộc tính **“Age”** làm gốc.

Bây giờ ta tiếp tục xét từng nhánh trong Age:

Nhánh Age = middle- age

Smiddle-age = {43} YES

Do Entropy(Smiddle-age) = 0 nên nhánh dừng.

Nhánh Age = senior

Ssenior = {33} YES

Do Entropy(Ssenior) = 0 nên nhánh dừng.

Nhánh Age = youth

Syouth= {1, 6}

Entropy(Syouth) = 1

* + - * **Xét thuộc tính Sex:**

Entropy(Sfemale) = - (1/1)\* log2(1/1) = 0

Entropy(Smale) = - (1/1)\* log2(1/1) = 0

→Gain(Syouth, Sex) = Entropy(Syouth) – (1/2)\*Entropy(Sfemale)

– (1/2)\*Entropy(Sfemale) = 1 – (1/2) \* 0 – (1/2) \* 0 **= 1**

* + - * **Xét thuộc tính Manufacturers:**

Entropy(SSONY) = - (1/1)\* log2(1/1)= 0

Entropy(SDELL) = - (1/1)\* log2(1/1)= 0

→Gain(Syouth, Manufacturers) = Entropy(Syouth) –

(1/2)\* Entropy(SSONY) – (1/2)\*Entropy(SDELL)

= 1 – (1/2) \* 0– (1/2) \* 0**= 1**

Do thuộc tính **“Sex”** và **“Manufacturers”**có độ lợi thông tin bằng nhau nên chọn Sex làm gốc.

Ta tiếp tục xét từng nhánh trong **Sex:**

Nhánh Sex= male

Smale = {1} YES

Do Entropy(Smale) = 0 nên nhánh dừng.

Nhánh Sex= female

Sfemale= {6} No

Do Entropy(Sfemale) = 0 nên nhánh dừng.

Nhánh Income = medium

Smedium = {11, 38} NO

Do Entropy(Smedium) = 0 nên nhánh dừng

Nhánh Income = low

Slow = {14, 23} NO

Do Entropy(Slow) = 0 nên nhánh dừng

Nhánh Configuration= medium

Smedium = {13, 17, 22, 24, 27, 39, 45, 47}

Ta có: Entropy(Smedium) = 0.811

* + - **Xét thuộc tính Age:**

Entropy(Syouth) = - (3/3)\* log2(3/3) = 0

Entropy(Smiddle-aged) = - (1/1)\* log2(1/1) = 0

Entropy(Ssenior) = - (2/4)\* log2(2/4) – (2/4)\* log2(2/4) = 1

→ Gain(Smedium, Age) = Entropy(Smedium) – (3/8) \* Entropy(Syouth) – (1/8) \* Entropy(Smiddle-aged) – (4/8) \* Entropy(Ssenior)

= 0.811 – (3/8) \* 0 – (1/8) \* 0 – (4/8) \* 1 **= 0.311**

* + - **Xét thuộc tính Sex:**

Entropy(Smale) = - (3/3)\* log2(3/3) = 0

Entropy(Sfemale) = - (3/5)\* log2(3/5) – (2/5)\* log2(2/5) = 0.971

→ Gain(Smedium, Sex) = Entropy(Smedium) – (3/8) \* Entropy(Smale) – (5/8) \* Entropy(Sfemale)

= 0.811 – (3/8) \* 0 – (5/8) \* 0.971 **= 0.204**

* + - **Xét thuộc tính Income:**

Entropy(Shigh) = - (3/4)\* log2(3/4) – (1/4)\* log2(1/4) = 0.811

Entropy(Smedium) = - (2/2)\* log2(2/2) = 0

Entropy(Slow) = - (1/2)\* log2(1/2) – (1/2)\* log2(1/2) = 1

→Gain(Smedium, Income)=Entropy(Smedium)– (4/8) \* Entropy(Shigh) – (2/8) \* Entropy(Smedium) – (2/8) \* Entropy(Slow)

= 0.811 – (4/8) \* 0.811 – (2/8) \* 0 – (2/8) \* 1 **= 0.156**

* + - **Xét thuộc tính Manufacturers:**

Entropy(SSONY) = - (1/2)\* log2(1/2) – (1/2)\* log2(1/2) = 1

Entropy(SDELL) = - (2/2)\* log2(2/2) = 0

Entropy(SASUS) = - (1/1)\* log2(1/1) = 0

Entropy(SHP) = - (2/3)\* log2(2/3) – (1/3)\* log2(1/3) = 0.918

→ Gain(Smedium, Manufacturers) = Entropy(Smedium) –

(2/8) \* Entropy(SSONY) – (2/8) \* Entropy(SDELL) – (1/8) \* Entropy(SASUS) – (3/8) \* Entropy(SHP)

= 0.811 – (2/8) \* 1 – (2/8) \* 0 – (1/8) \* 0 – (3/8) \* 0.918 **= 0.217**

Trong các thuộc tính điều kiện trên ta thấy **“Age”** có độ lợi thông tin lớn nhất nên thuộc tính này được chọn làm gốc.

Bây giờ ta tiếp tục xét từng nhánh trong Age.

Nhánh Age= youth

Syouth = {22, 24, 39} YES

Do Entropy(Syouth) = 0 nên nhánh dừng

Nhánh Age = middle-aged

Smiddle-aged = {45} YES

Do Entropy(Smiddle-aged) = 0 nên nhánh dừng

Nhánh Age = senior

Ssenior = {13, 17, 27, 47}

Entropy(Ssenior) = 1

* + - * **Xét thuộc tính Sex:**

Entropy(Smale) = - (2/2)\* log2(2/2) = 0

Entropy(Sfemale) = - (2/2)\* log2(2/2) = 0

→ Gain(Ssenior, Sex) = Entropy(Ssenior) – (2/4) \* Entropy(Smale) – (2/4) \* Entropy(Sfemale)

= 1 – (2/4) \* 0 – (2/4) \* 0 **= 1**

* + - * + **Xét thuộc tính Income:**

Entropy(Shigh) = - (1/2)\* log2(1/2) – (1/2) log2(1/2) = 1

Entropy(Smedium) = - (1/1)\* log2(1/1) = 0

Entropy(Slow) = - (1/1)\* log2(1/1) = 0

→ Gain(Ssenior, Income) = Entropy(Ssenior) – (2/4) \* Entropy(Shigh) – (1/4) \* Entropy(Smedium) – (1/4) \* Entropy(Slow)

= 1 – (2/4) \* 1 – (1/4) \* 0 – (1/4) \* 0 **= 0.5**

* + - * + **Xét thuộc tính Manufacturers:**

Entropy(SSONY) = - (1/2)\*log2(1/2) – (1/2)log2(1/2) = 1

Entropy(SHP) = - (1/2)\* log2(1/2) – (1/2)\* log2(1/2) = 1

→ Gain(Ssenior, Manufacturers) = Entropy(Ssenior) – (2/4) \* Entropy(SSONY) – (2/4) \* Entropy(SHP)

= 1 – (2/4) \* 1 – (2/4) \* 1 **= 0**

Trong các thuộc tính điều kiện trên ta thấy **“Sex”** có độ lợi thông tin lớn nhất nên thuộc tính này được chọn làm gốc.

Nhánh Sex= male

Smale = {13, 27} YES

Do Entropy(Smale) = 0 nên nhánh dừng

Nhánh Sex= female

Sfemale = {17, 47} NO

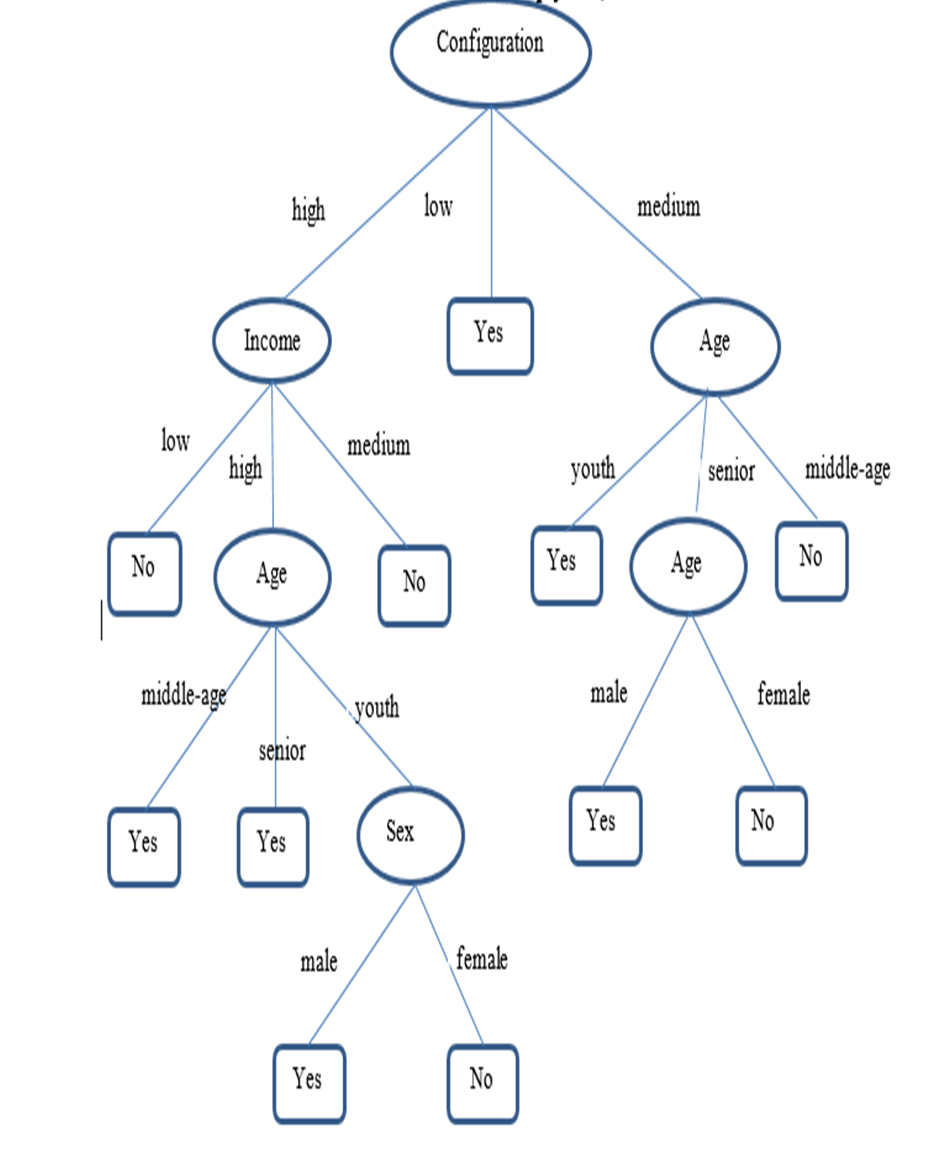
Do Entropy(Sfemale) = 0 nên nhánh dừng

Nhánh Configuration= low

Slow = {2, 12, 16, 34 , 35, 46} YES

Do Entropy(Slow) = 0 nên nhánh dừng.

Nhánh Windows đã hoàn thành và sinh ra cây quyết định:



### *Hình 2.2 Nhánh Windows*

**Tập luật sinh cho nhánh Windows:**

* Rule 1: IF OS = Windows AND Configuration = high AND Income = low THEN Buy = No
* Rule 2: IF OS = Windows AND Configuration = high AND Income = medium THEN Buy = No
* Rule 3: IF OS = Windows AND Configuration = high AND Income = high AND Age = middle-age THEN Buy = Yes
* Rule 4: IF OS = Windows AND Configuration = high AND Income = high AND Age = senior THEN Buy = Yes
* Rule 5: IF OS = Windows AND Configuration = high AND Income = high AND Age = youth AND Sex = male THEN Buy = Yes
* Rule 6: IF OS = Windows AND Configuration = high AND Income = high AND Age = youth AND Sex = female THEN Buy = No
* Rule 7: IF OS = Windows AND Configuration = low THEN Buy = Yes
* Rule 8: IF OS = Windows AND Configuration = medium AND Age = youth THEN Buy = Yes
* Rule 9: IF OS = Windows AND Configuration = medium AND Age = middle –age THEN Buy = Yes
* Rule 10: IF OS = Windows AND Configuration = medium AND Age = senior AND Sex = male THEN Buy = Yes
* Rule 11: IF OS = Windows AND Configuration = medium AND Age = senior AND Sex = female THEN Buy = No

# **CHƯƠNG 3: CÀI ĐẶT**

## **3.1.Công nghệ sử dụng**

Chương trình TuVanMuaMayTinh được viết trên nền tảng ngôn ngữ C#. Sử dụng SQL server để lưu trữ dữ liệu (Tập dữ liệu huấn luyện và các tập luật sinh ra).

**+ Ưu điểm:**

* Có khả năng tự động tính toán và sinh ra cây quyết định, từ đó tự động sinh ra tập luật để tự đưa ra câu trả lời.
* Linh hoạt khi thay đổi tập dữ liệu huấn luyện.
* Đáp án chính xác.
* Gọn nhẹ, dễ dàng sử dụng.
* Dễ hiểu, cây quyết định sẽ tự mở nhánh theo câu trả lời của người dùng.

**+ Nhược điểm:**

Máy tính cần cài đặt Net Framework và SQL server để có thể chạy chương trình.

## **3.2.Cơ sở dữ liệu**

Dữ liệu gồm 2 bảng chính:

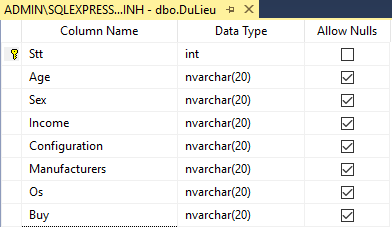
- Bảng dữ liệu:

+ Stt: Số thứ tự + Configuration: Cấu hình

+ Age: Tuổi + Manufacturers: Nhà sản xuất

+ Sex: Giới tính + Os: Hệ điều hành

+Income: Thu nhập + Buy: Mua



### *Hình 3.1 dbo Bảng DuLieu*

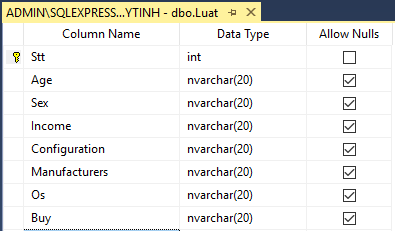
- Bảng luật:

+ Stt: Số thứ tự + Configuration: Cấu hình

+ Age: Tuổi + Manufacturers: Nhà sản xuất

+ Sex: Giới tính + Os: Hệ điều hành

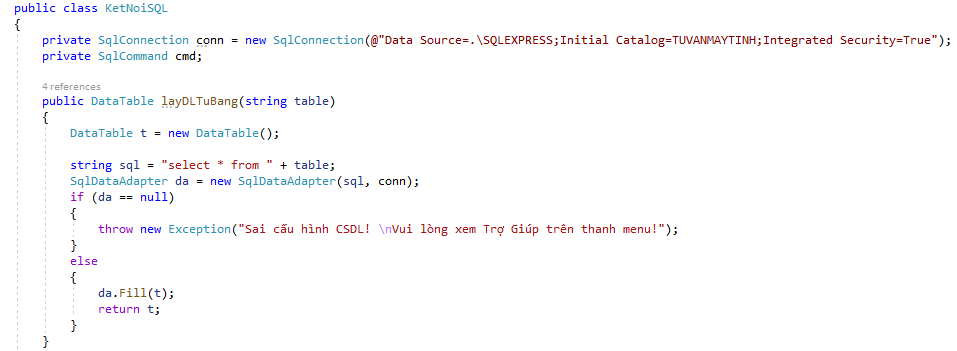
+Income: Thu nhập + Buy: Mua



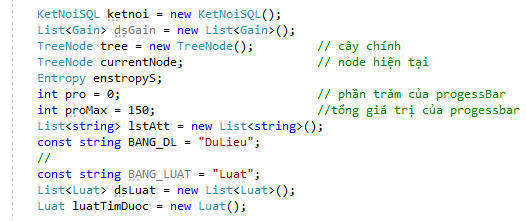
### *Hình 3.2 dbo Bảng Luat*

## **3.3.Cài đặt**

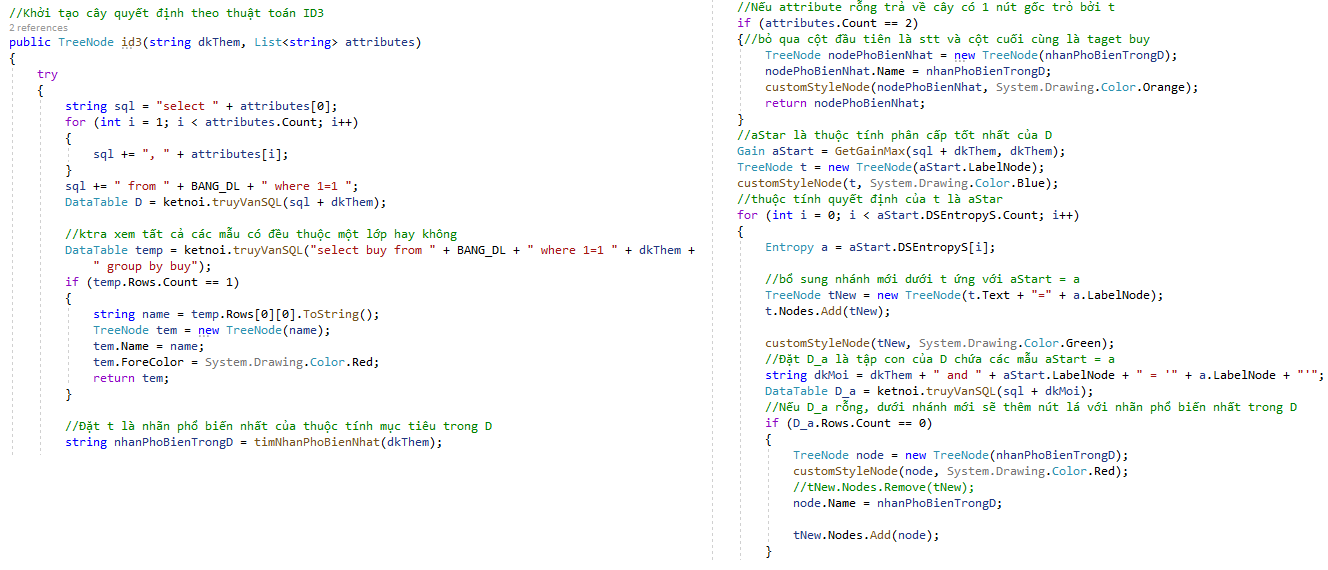
Kết nối với cơ sở dữ liệu



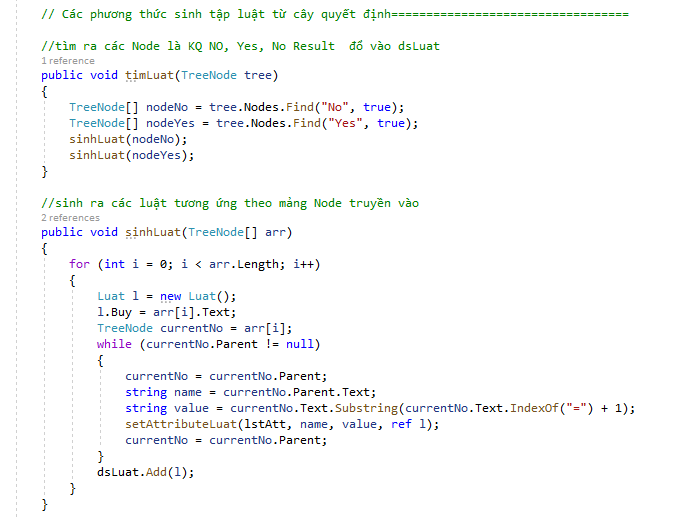
Khởi tạo cây chính, node và các giá trị ban đầu



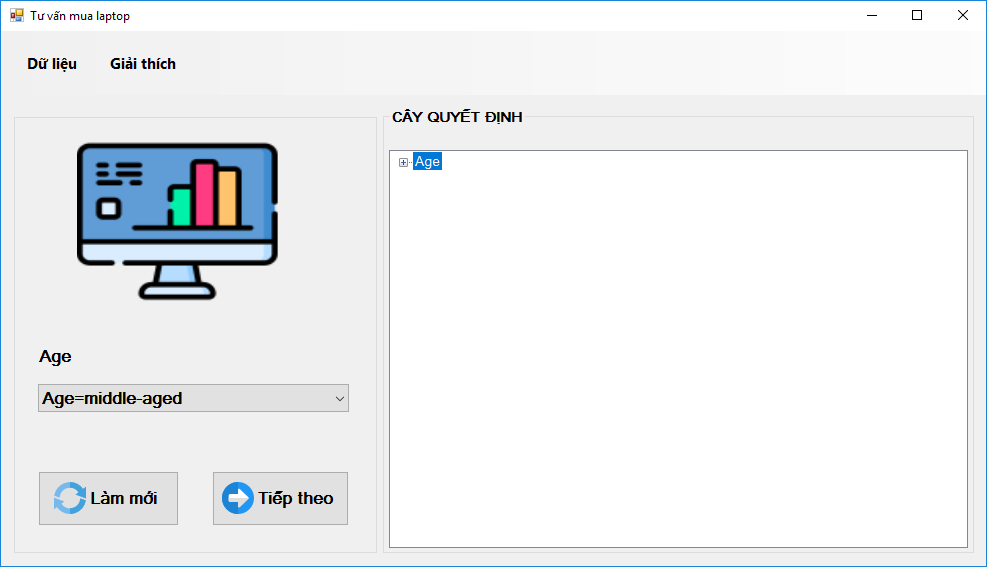
Khởi tạo cây quyết định theo thuật toán ID3



Tìm ra các Node là NO, YES để đổ vào danh sách Luật, sinh ra các luật tương ứng theo mảng Node truyền vào

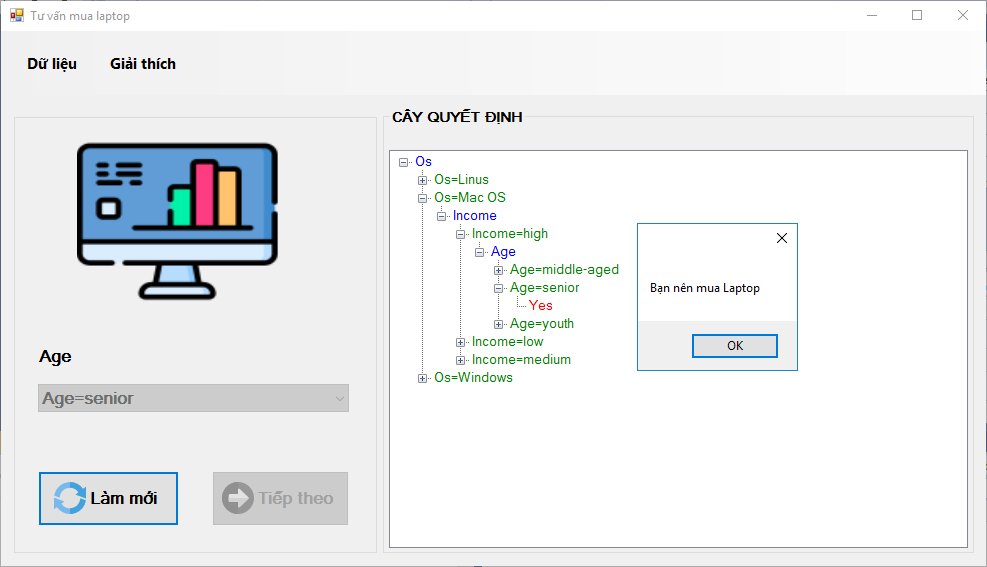


**Kết quả**



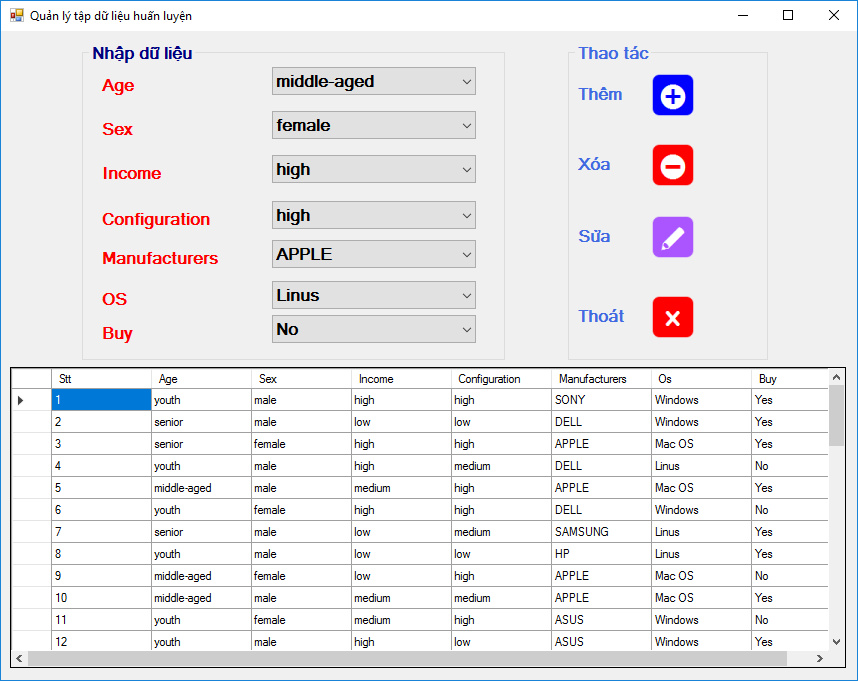
### *Hình 3.3 Form chính*

Đây là form chính sau khi chạy chương trình, bên trái là các lựa chọn để tiến hành tư vấn mua laptop, khi người dùng chọn xong nhấn tiếp theo để chuyển sang câu hỏi kế tiếp, bên phải là minh họa cho cây quyết định



### *Hình 3.4 Kết quả sau khi được tư vấn*

Sau khi nhận được các câu trả lời, chuyên gia sẽ tư vấn cho người dử dụng có nên mua hay không.



### *Hình 3.5 Bảng quản lý dữ liệu*

Ở bảng này chúng ta có thể xem thông tin về các dữ liệu, ngoài ra có thể thêm, sửa hoặc xóa

# **KẾT LUẬN**

Sau thời gian nghiên cứu và bắt tay vào thực hiện, với mong muốn hiểu được bản chất và nắm rõ được cây quyết định và đặc biệt là thuật toán ID3, chúng em đã hoàn thành đề tài “*Xây dựng hệ chuyên gia tư vấn mua laptop”.*

Kết quả đạt được:

* Hiểu được thuật toán ID3
* Xây dựng được chương trình tư vấn mua laptop bằng thuật toán ID3

Mặc dù đã rất cố gắng hoàn thiện báo cáo với tất cả sự nỗ lực, tuy nhiên do bước đầu tìm hiểu và xây dựng báo cáo môn học trong thời gian có hạn, với lượng kiến thức còn hạn chế, chúng em rất mong nhận được sự quan tâm, thông cảm và những đóng góp quý báu của các thầy cô và các bạn để báo cáo này ngày càng hoàn thiện hơn.

Hướng phát triển:

* Xây dựng chương trình với nhiều chức năng hơn
* Kết quả tư vấn đạt độ chính xác cao hơn