|  |  |
| --- | --- |
|  | **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM** |

**ĐỒ ÁN THỰC HÀNH**

**TRÍ TUỆ NHÂN TẠO**

**CÂY QUYẾT ĐỊNH - I3**

***Ngành:*** **CÔNG NGHỆ THÔNG TIN**

Giảng viên hướng dẫn : Th.S Nguyễn Mạnh Hùng

Thành viên nhóm :

Bùi Phú Khuyên MSSV: 1611061839 Lớp: 16DTHA3

Phạm Quốc Sơn MSSV: 1611061362 Lớp: 16DTHA3

Võ Nhị Anh MSSV: 1611060551 Lớp: 16DTHA

TP. Hồ Chí Minh, 2018

**PHẦN 1: CHẠY TAY CÂY QUYẾT ĐỊNH**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 1 | 5.5 | 4.2 | 1.4 | 0.2 | Iris-Setosa |
| 2 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-Setosa |
| 3 | 5.0 | 3.2 | 1.2 | 0.2 | Iris-Setosa |
| 4 | 5.5 | 3.5 | 1.3 | 0.2 | Iris-Setosa |
| 5 | 6.8 | 2.8 | 4.8 | 1.4 | Iris-Versicolor |
| 6 | 6.7 | 3.0 | 5.0 | 1.7 | Iris-Versicolor |
| 7 | 6.0 | 2.9 | 4.5 | 1.5 | Iris-Versicolor |
| 8 | 5.7 | 2.6 | 3.5 | 1.0 | Iris-Versicolor |
| 9 | 6.4 | 2.7 | 5.3 | 1.9 | Iris-Virginica |
| 10 | 6.8 | 3.0 | 5.5 | 2.1 | Iris-Virginica |
| 11 | 5.7 | 2.5 | 5.0 | 2.0 | Iris-Virginica |
| 12 | 5.8 | 2.8 | 5.1 | 2.4 | Iris-Virginica |
| 13 | 6.4 | 3.2 | 5.3 | 2.3 | Iris-Virginica |
| 14 | 6.5 | 3.0 | 5.5 | 1.8 | Iris-Virginica |
| 15 | 7.7 | 3.8 | 6.7 | 2.2 | Iris-Virginica |
| 16 | 7.7 | 2.6 | 6.9 | 2.3 | Iris-Virginica |
| 17 | 6.0 | 2.2 | 5.0 | 1.5 | Iris-Virginica |
| 18 | 6.9 | 3.2 | 5.7 | 2.3 | Iris-Virginica |
| 19 | 5.6 | 2.8 | 4.9 | 2.0 | Iris-Virginica |
| 20 | 7.7 | 2.8 | 6.7 | 2.0 | Iris-Virginica |

***Bảng dữ liệu***

***Chuẩn hoá thuộc tính liên tục thành dữ liệu rời rạc:***

|  |  |  |  |
| --- | --- | --- | --- |
| **Giá trị thuộc tính**  **(Sepal length)** | **Setosa** | **Versicolor** | **Virginica** |
| (-∝, 5.6] | 4 | 0 | 1 |
| (5.6, 6.3] | 0 | 2 | 3 |
| (6.3, 7] | 0 | 2 | 5 |
| (7, +∝) | 0 | 0 | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Giá trị thuộc tính**  **(Sepal width)** | **Setosa** | **Versicolor** | **Virginica** |
| (-∝, 2.7] | 0 | 1 | 4 |
| (2.7, 3.2] | 2 | 3 | 7 |
| (3.2, 3.7] | 1 | 0 | 0 |
| (3.7, +∝) | 1 | 0 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Giá trị thuộc Tính**  **(Petal length)** | **Setosa** | **Versicolor** | **Virginica** |
| (-∝, 2.625] | 4 | 0 | 0 |
| (2.625, 4.05] | 0 | 1 | 0 |
| (4.05, 5.475] | 0 | 3 | 6 |
| (5.475, +∝) | 0 | 0 | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Giá trị thuộc tính**  **(Petal width)** | **Setosa** | **Versicolor** | **Virginica** |
| (-∝, 0.675] | 4 | 0 | 0 |
| (0.675, 1.25] | 0 | 1 | 0 |
| (1.25, 1.825] | 0 | 3 | 2 |
| (1.825, +∝) | 0 | 0 | 10 |

Gọi tập S = [Se, Ve, Vi]

**Entropy (S)** = -PSelog2PSe -PVelog2PVe -PVilog2PVi

=

**ĐỘ ĐO (INFORMATION GAIN):**

***Công thức chung:***

* Gain(S, Sepal length):

|  |  |  |
| --- | --- | --- |
| **SSepal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 5.6] | [4, 0, 1] | 0.72 |
| (5.6, 6.3] | [0, 2, 3] | 0.97 |
| (6.3, 7] | [0, 2, 5] | 0.86 |
| (7, +∝) | [0, 0, 3] | 0 |

**→** Gain(S, Sepal length):

* Gain(S, Sepal width):

|  |  |  |
| --- | --- | --- |
| **SSepal width = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.7] | [0, 1, 4] | 0.72 |
| (2.7, 3.2] | [2, 3, 7] | 1.38 |
| (3.2, 3.7] | [1, 0, 0] | 0 |
| (3.7, +∝) | [1, 0, 1] | 1 |

**→** Gain(S, Sepal length):

* Gain(S, Petal length):

|  |  |  |
| --- | --- | --- |
| **SPetal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.625] | [4, 0, 0] | 0 |
| (2.625, 4.05] | [0, 1, 0] | 0 |
| (4.05, 5.475] | [0, 3, 6] | 0.92 |
| (5.475, +∝) | [0, 0, 6] | 0 |

**→** Gain(S, Sepal length):

* Gain(S, Petal width):

|  |  |  |
| --- | --- | --- |
| **SPetal width = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 0.675] | [4, 0, 0] | 0 |
| (0.675, 1.25] | [0, 1, 0] | 0 |
| (1.25, 1.825] | [0, 3, 2] | 0.97 |
| (1.825, +∝) | [0, 0, 10] | 0 |

**→** Gain(S, Sepal width):

**Ta có:**

Gain(S, Sepal length) = 0.64

Gain(S, Sepal width) =0.26

Gain(S, Petal length) = 0.96

Gain(S, Petal width) = **1.13**

**Nhận xét:** Từ đây ta thấy độ đo của S đối với thuộc tính **"Petal width"** là lớn nhất trong 4 thuộc tính. Như vậy ta quyết định chọn **"Petal width"** làm thuộc tính đầu tiên để khai triển cây.

**Pental width**

**?**

**?**

**?**

**?**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

*Hình 1: Khai triển cây theo thuộc tính "Petal width"*

Tương tự như vậy, ta có thể tiến hành triển khai các nút ở bước tiếp theo:

* **Petal width** (S(-∝, 0.675]) S = [4, 0, 0] **Entropy = 0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 1 | 5.5 | 4.2 | 1.4 | **0.2** | Iris-Setosa |
| 2 | 4.9 | 3.1 | 1.5 | **0.1** | Iris-Setosa |
| 3 | 5.0 | 3.2 | 1.2 | **0.2** | Iris-Setosa |
| 4 | 5.5 | 3.5 | 1.3 | **0.2** | Iris-Setosa |

|  |  |  |
| --- | --- | --- |
| **SSepal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 5.6] | [**4**, 0, 0] | 0 |
| (5.6, 6.3] | [0, 0, 0] | 0 |
| (6.3, 7] | [0, 0, 0] | 0 |
| (7, +∝) | [0, 0, 0] | 0 |

|  |  |  |
| --- | --- | --- |
| **SSepal width = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.7] | [0, 0, 0] | 0 |
| (2.7, 3.2] | [**2**, 0, 0] | 0 |
| (3.2, 3.7] | [**1**, 0, 0] | 0 |
| (3.7, +∝) | [**1**, 0, 0] | 0 |

|  |  |  |
| --- | --- | --- |
| **SPetal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.625] | [**4**, 0, 0] | 0 |
| (2.625, 4.05] | [0, 0, 0] | 0 |
| (4.05, 5.475] | [0, 0, 0] | 0 |
| (5.475, +∝) | [0, 0, 0] | 0 |

**Nhận xét:** Vì các thuộc tính **Sepal length, Sepal width và Petal length** đều nhận duy nhất Class **Iris-Setosa** trong quá trình khai triển nên với **Petal width** (S(-∝, 0.675]) nhận kết quả trả về là **Iris-Setosa**

**Pental width**

**Iris Setosa**

**?**

**?**

**?**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

*Hình 2: Khai triển cây theo thuộc tính "Petal width" nhận kết quả: Iris-Setosa* (S(-∝, 0.675])

* **Petal width** (S(0.675, 1.25])S = [0, 1, 0] **Entropy = 0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 8 | 5.7 | 2.6 | 3.5 | **1.0** | Iris-Versicolor |

|  |  |  |
| --- | --- | --- |
| **SSepal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 5.6] | [0, 0, 0] | 0 |
| (5.6, 6.3] | [0, **1**, 0] | 0 |
| (6.3, 7] | [0, 0, 0] | 0 |
| (7, +∝) | [0, 0, 0] | 0 |

|  |  |  |
| --- | --- | --- |
| **SSepal width = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.7] | [0, **1**, 0] | 0 |
| (2.7, 3.2] | [0, 0, 0] | 0 |
| (3.2, 3.7] | [0, 0, 0] | 0 |
| (3.7, +∝) | [0, 0, 0] | 0 |

|  |  |  |
| --- | --- | --- |
| **SPetal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.625] | [0, 0, 0] | 0 |
| (2.625, 4.05] | [0, **1**, 0] | 0 |
| (4.05, 5.475] | [0, 0, 0] | 0 |
| (5.475, +∝) | [0, 0, 0] | 0 |

**Nhận xét:** Vì các thuộc tính **Sepal length, Sepal width và Petal length** đều nhận duy nhất Class **Iris-Versicolor** trong quá trình khai triển nên với **Petal width** (S(0.675, 1.25]) nhận kết quả trả về là **Iris-Versicolor**

**Pental width**

**Iris-Setosa**

**?**

**?**

**Iris-Versicolor**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , -∝)

*Hình 3: Khai triển cây theo thuộc tính "Petal width" nhận kết quả: Iris-Versicolor* (S(0.675, 1.25])

* **Petal width:** (S(1.25, 1.825])S = [0, 3, 2] Entropy = 0.97

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 5 | 6.8 | 2.8 | 4.8 | **1.4** | Iris-Versicolor |
| 6 | 6.7 | 3.0 | 5.0 | **1.7** | Iris-Versicolor |
| 7 | 6.0 | 2.9 | 4.5 | **1.5** | Iris-Versicolor |
| 14 | 6.5 | 3.0 | 5.5 | **1.8** | Iris-Virginica |
| 17 | 6.0 | 2.2 | 5.0 | **1.5** | Iris-Virginica |

* Gain(S(1.25, 1.825], Sepal length):

|  |  |  |
| --- | --- | --- |
| **SSepal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 5.6] | [0, 0, 0] | 0 |
| (5.6, 6.3] | [0, 1, 1] | 1 |
| (6.3, 7] | [0, 2, 1] | 0.92 |
| (7, +∝) | [0, 0, 0] | 0 |

**→** Gain(S(1.25, 1.825], Sepal length):

* Gain(S(1.25, 1.825], Sepal width):

|  |  |  |
| --- | --- | --- |
| **SSepal width = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.7] | [0, 0, 1] | 0 |
| (2.7, 3.2] | [0, 3, 1] | 0.81 |
| (3.2, 3.7] | [0, 0, 0] | 0 |
| (3.7, +∝) | [0, 0, 0] | 0 |

**→** Gain(S(1.25, 1.825], Sepal width):

* Gain(S(1.25, 1.825], Petal length):

|  |  |  |
| --- | --- | --- |
| **SPetal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.625] | [0, 0, 0] | 0 |
| (2.625, 4.05] | [0, 0, 0] | 0 |
| (4.05, 5.475] | [0, 3, 1] | 0.81 |
| (5.475, +∝) | [0, 0, 1] | 0 |

**→** Gain(S(1.25, 1.825], Petal length):

**Ta có:**

Gain(S(1.25, 1.825], Sepal length) = 0.018

Gain(S(1.25, 1.825], Sepal width) =**0.32**

Gain(S(1.25, 1.825], Petal length) = **0.32**

**Nhận xét:** Từ đây ta thấy độ đo của S(1.25, 1.825] đối với thuộc tính **"Sepal width"** và **"Petal length"** là lớn nhất. Như vậy ta có thể quyết định chọn **"Sepal width"** làm thuộc tính tiếp theo để khai triển cây.

**Pental width**

**Iris-Setosa**

**Sepal width**

**?**

**Iris-Versicolor**

**?**

**?**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

(-∝, 2.7]

(2.7 , 3.2]

*Hình 4: Khai triển cây theo thuộc tính "Sepal width"*

Tương tự như vậy, trước khi ta qua các thuộc tính cuối cùng của "Pental length" (S(1.825 , +∝)). Ta có thể tiến hành triển khai các nút của "Sepal width" trước:

* **Sepal width:** (S(-∝, 2.7])S = [0, 0, 1] **Entropy = 0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 17 | 6.0 | **2.2** | 5.0 | **1.5** | Iris-Virginica |

|  |  |  |
| --- | --- | --- |
| **SSepal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 5.6] | [0, 0, 0] | 0 |
| (5.6, 6.3] | [0, 0, **1**] | 0 |
| (6.3, 7] | [0, 0, 0] | 0 |
| (7, +∝) | [0, 0, 0] | 0 |

|  |  |  |
| --- | --- | --- |
| **SPetal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.625] | [0, 0, 0] | 0 |
| (2.625, 4.05] | [0, 0, 0] | 0 |
| (4.05, 5.475] | [0, 0, **1**] | 0 |
| (5.475, +∝) | [0, 0, 0] | 0 |

**Nhận xét:** Vì các thuộc tính **Sepal length và Petal length** đều nhận duy nhất Class **Iris-Virginica** trong quá trình khai triển nên với **Petal width** (S(-∝, 2.7]) nhận kết quả trả về là **Iris-Virginica**

**Pental width**

**Iris-Setosa**

**Sepal width**

**?**

**Iris-Versicolor**

**?**

**Iris-Virginica**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

(-∝, 2.7]

(2.7 , 3.2]

*Hình 5: Khai triển cây theo thuộc tính "Sepal width" nhận kết quả: Iris-Virginica* (S(-∝, 2.7])

* **Sepal width:** (S(2.7, 3.2])S = [0, 3, 1] Entropy = 0.81

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 5 | 6.8 | **2.8** | 4.8 | **1.4** | Iris-Versicolor |
| 6 | 6.7 | **3.0** | 5.0 | **1.7** | Iris-Versicolor |
| 7 | 6.0 | **2.9** | 4.5 | **1.5** | Iris-Versicolor |
| 14 | 6.5 | **3.0** | 5.5 | **1.8** | Iris-Virginica |

* Gain(S(2.7, 3.2], Sepal length):

|  |  |  |
| --- | --- | --- |
| **SSepal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 5.6] | [0, 0, 0] | 0 |
| (5.6, 6.3] | [0, 1, 0] | 0 |
| (6.3, 7] | [0, 2, 1] | 0.92 |
| (7, +∝) | [0, 0, 0] | 0 |

→ Gain(S(2.7, 3.2], Sepal length):

* Gain(S(2.7, 3.2], Petal length):

|  |  |  |
| --- | --- | --- |
| **SPetal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.625] | [0, 0, 0] | 0 |
| (2.625, 4.05] | [0, 0, 0] | 0 |
| (4.05, 5.475] | [0, 3, 0] | 0 |
| (5.475, +∝) | [0, 0, 1] | 0 |

→ Gain(S(2.7, 3.2], Petal length) =

**Ta có:**

Gain(S(2.7, 3.2], Sepal length) = 0.12

Gain(S(2.7, 3.2], Petal length) = **0.81**

**Nhận xét:** Từ đây ta thấy độ đo của S(2.7, 3.2] đối với thuộc tính **"Petal length"** là lớn nhất. Như vậy ta có thể quyết định chọn **"Petal length"** làm thuộc tính tiếp theo để khai triển cây.

**Pental width**

**Iris-Setosa**

**Sepal width**

**?**

**Iris-Versicolor**

**Petal length**

**Iris-Virginica**

**?**

**?**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

(-∝, 2.7]

(2.7 , 3.2]

(4.05 , 5.475]

(5.475 , +∝)

*Hình 6: Khai triển cây theo thuộc tính "Petal length"*

* **Petal lengh:** (S(4.05, 5.475])S = [0, 3, 0] **Entropy = 0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 5 | 6.8 | **2.8** | **4.8** | **1.4** | Iris-Versicolor |
| 6 | 6.7 | **3.0** | **5.0** | **1.7** | Iris-Versicolor |
| 7 | 6.0 | **2.9** | **4.5** | **1.5** | Iris-Versicolor |

**Nhận xét:** Vì các thuộc tính **Sepal length** nhận duy nhất Class **Iris-Versicolor** trong quá trình khai triển nên với **Petal lenght** (S(4.05, 5.475]) nhận kết quả trả về là **Iris-Versicolor**

**Pental width**

**Iris-Setosa**

**Sepal width**

**?**

**Iris-Versicolor**

**Petal length**

**Iris-Virginica**

**Iris-Versicolor**

**?**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

(-∝, 2.7]

(2.7 , 3.2]

(4.05 , 5.475]

(5.475 , +∝)

*Hình 7: Khai triển cây theo thuộc tính "Petal length" nhận kết quả: Iris-Versicolor* (S(4.05, 5.475])

* **Petal lengh:** (S(5.475, +∝))S = [0, 0, 1] **Entropy = 0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 14 | 6.5 | **3.0** | **5.5** | **1.8** | Iris-Virginica |

**Nhận xét:** Vì các thuộc tính **Sepal length** nhận duy nhất Class **Iris-Vinginica** trong quá trình khai triển nên với **Petal lenght** (S(5.475, +∝)) nhận kết quả trả về là **Iris-Vinginica**

**Pental width**

**Iris-Setosa**

**Sepal width**

**?**

**Iris-Versicolor**

**Petal length**

**Iris-Virginica**

**Iris-Versicolor**

**Iris-Viginica**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

(-∝, 2.7]

(2.7 , 3.2]

(4.05 , 5.475]

(5.475 , +∝)

*Hình 8: Khai triển cây theo thuộc tính "Petal length" nhận kết quả: Iris-Viginica* (S(5.475, +∝))

Nhận thấy, để hoàn thành cây ID3 vẫn còn thuộc tính Pental width S(1.825, +∝) chưa được khai triển nên ta tiến hành khai triển:

* **Petal width (S**(1.825, +∝)**)**  S = [0, 0, 10] **Entropy = 0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Sepal length** | **Sepal width** | **Petal length** | **Petal width** | **Class** |
| 9 | 6.4 | 2.7 | 5.3 | **1.9** | Iris-Virginica |
| 10 | 6.8 | 3.0 | 5.5 | **2.1** | Iris-Virginica |
| 11 | 5.7 | 2.5 | 5.0 | **2.0** | Iris-Virginica |
| 12 | 5.8 | 2.8 | 5.1 | **2.4** | Iris-Virginica |
| 13 | 6.4 | 3.2 | 5.3 | **2.3** | Iris-Virginica |
| 15 | 7.7 | 3.8 | 6.7 | **2.2** | Iris-Virginica |
| 16 | 7.7 | 2.6 | 6.9 | **2.3** | Iris-Virginica |
| 18 | 6.9 | 3.2 | 5.7 | **2.3** | Iris-Virginica |
| 19 | 5.6 | 2.8 | 4.9 | **2.0** | Iris-Virginica |
| 20 | 7.7 | 2.8 | 6.7 | **2.0** | Iris-Virginica |

|  |  |  |
| --- | --- | --- |
| **SSepal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 5.6] | [0, 0, **1**] | 0 |
| (5.6, 6.3] | [0, 0, **2**] | 0 |
| (6.3, 7] | [0, 0, **4**] | 0 |
| (7, +∝) | [0, 0, **3**] | 0 |

|  |  |  |
| --- | --- | --- |
| **SSepal width = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.7] | [0, 0, **3**] | 0 |
| (2.7, 3.2] | [0, 0, **6**] | 0 |
| (3.2, 3.7] | [0, 0, **0**] | 0 |
| (3.7, +∝) | [0, 0, **1**] | 0 |

|  |  |  |
| --- | --- | --- |
| **SPetal length = [Se, Ve, Vi]** | | **Entropy** |
| (-∝, 2.625] | [0, 0, 0] | 0 |
| (2.625, 4.05] | [0, 0, 0] | 0 |
| (4.05, 5.475] | [0, 0, **5**] | 0 |
| (5.475, +∝) | [0, 0, **5**] | 0 |

**Nhận xét:** Vì các thuộc tính **Sepal length, Sepal width và Petal length** đều nhận duy nhất Class **Iris-Virginica** trong quá trình khai triển nên với **Petal width** (S(1.825, +∝))nhận kết quả trả về là **Iris-Virginica**

**Pental width**

**Iris-Setosa**

**Sepal width**

**Iris-Versicolor**

**Petal length**

**Iris-Virginica**

**Iris-Versicolor**

**Iris-Viginica**

**Iris-Viginica**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

(-∝, 2.7]

(2.7 , 3.2]

(4.05 , 5.475]

(5.475 , +∝)

*Hình 9: Khai triển cây theo thuộc tính "Petal width" nhận kết quả: Iris-Viginica* (S(1.825, +∝))

**PHẦN 2: CHẠY CÂY QUYẾT ĐỊNH TRÊN WEKA**

**1. Nội dung File \*arff:**

**@Relation** IRIS

**@Attribute** Sepallength real

**@Attribute** Sepalwidth real

**@Attribute** Petallength real

**@Attribute** Petalwidth real

**@Attribute** Class {Iris-Setosa, Iris-Versicolor, Iris-Virginica}

**@Data**

5.5, 4.2, 1.4, 0.2, Iris-Setosa

4.9, 3.1, 1.5, 0.1, Iris-Setosa

5.0, 3.2, 1.2, 0.2, Iris-Setosa

5.5, 3.5, 1.3, 0.2, Iris-Setosa

6.8, 2.8, 4.8, 1.4, Iris-Versicolor

6.7, 3.0, 5.0, 1.7, Iris-Versicolor

6.0, 2.9, 4.5, 1.5, Iris-Versicolor

5.7, 2.6, 3.5, 1.0, Iris-Versicolor

6.4, 2.7, 5.3, 1.9, Iris-Virginica

6.8, 3.0, 5.5, 2.1, Iris-Virginica

5.7, 2.5, 5.0, 2.0, Iris-Virginica

5.8, 2.8, 5.1, 2.4, Iris-Virginica

6.4, 3.2, 5.3, 2.3, Iris-Virginica

6.5, 3.0, 5.5, 1.8, Iris-Virginica

7.7, 3.8, 6.7, 2.2, Iris-Virginica

7.7, 2.6, 6.9, 2.3, Iris-Virginica

6.0, 2.2, 5.0, 1.5, Iris-Virginica

6.9, 3.2, 5.7, 2.3, Iris-Virginica

5.6, 2.8, 4.9, 2.0, Iris-Virginica

7.7, 2.8, 6.7, 2.0, Iris-Virginica

**2. Thực nghiệm cây quyết định: Chọn thuật toán ID3, Vẽ lại cây quyết định:**

**a. Kết quả chạy thuật toán ID3 trên Weka:**

=== Classifier model (full training set) ===

Id3 *//Cây bắt đầu từ Petalwidth và có 3 nhánh*

Petalwidth = '(-inf-0.675]': Iris-Setosa *//Là nhánh 1 - Khi Petalwidth = (-inf-0.675] là hoa Setosa*

Petalwidth = '(0.675-1.25]': Iris-Versicolor *//Là nhánh 2 - Khi Petalwidth =* (0.675-1.25] *là hoa Vericolor*

Petalwidth = '(1.25-1.825]' *//Là nhánh 3:*

| Sepalwidth = '(-inf-2.7]': Iris-Virginica *// 3.1: Khi Sepalwidth = (-inf-2.7] là hoa Virginica*

| Sepalwidth = '(2.7-3.2]' *// 3.2: Khi Sepalwidth = (2.7-3.2]:*

| | Petallength = '(-inf-2.625]': null *//3.2.1: Không có nhánh Petallength =* (-inf-2.625]

| | Petallength = '(2.625-4.05]': null *//3.2.2: Không có nhánh Petallength =* (2.625-4.05]

| | Petallength = '(4.05-5.475]': Iris-Versicolor //3.2.3 *Khi Petallength =* (4.05-5.475] *là hoa Versicolor*

| | Petallength = '(5.475-inf)': Iris-Virginica //3.2.4 *Khi Petallength =* (5.475-inf) *là hoa Virginica*

| Sepalwidth = '(3.2-3.7]': null //3.3: Không có nhánh Sepalwidth = (3.2-3.7]

| Sepalwidth = '(3.7-inf)': null *//3.4: Không có nhánh Sepalwidth = (3.7-inf)*

Petalwidth = '(1.825-inf)': Iris-Virginica *//Là nhánh 4: Khi Petalwidth = (1.825-inf) là hoa Virginica*

**b. Vẽ cây quyết định:**

**Pental width**

**Iris-Setosa**

**Sepal width**

**Iris-Versicolor**

**Petal length**

**Iris-Virginica**

**Iris-Versicolor**

**Iris-Viginica**

**Iris-Viginica**

**Null**

**Null**

**Null**

**Null**

(0.675 , 1.25]

(-∝ , 0.675]

(1.25 , 1.825]

(1.825 , +∝)

(-∝, 2.7]

(2.7 , 3.2]

(4.05 , 5.475]

(5.475 , +∝)

(3.2, 3.7]

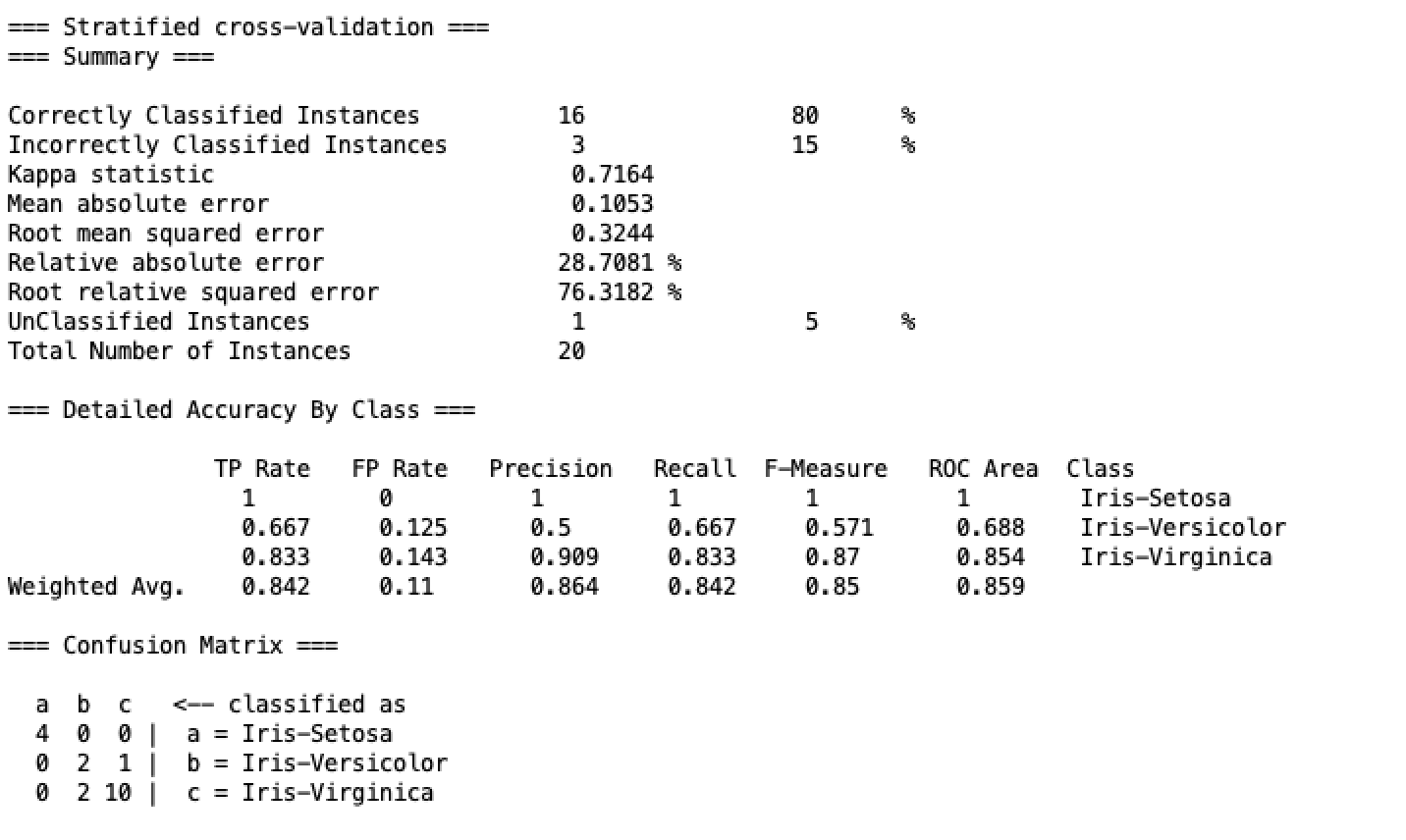
(3.7 , +∝)

(2.625, 4.05]

(-∝, 2.625]

*Hình 10: Khai triển cây theo kết quả chạy Weka*

**3. Test cây đã tạo sử dụng cross-validation (=20) . Hãy cho biết tỉ lệ chính xác của bộ phân lớp, số mẫu phân lớp đúng/ sai/ không phân lớp được. Diễn giải lại phân bố dữ liệu thông tin từ Confusion Matrix.** *(Chụp ảnh màn hình)*



***Trong đó:***

Correctly Classified Instances 16 80 % *// Là tỷ lệ % độ chính xác của Phân lớp*

Incorrectly Classified Instances 3 15 % *// Là tỷ lệ % độ không chính xác của Phân lớp*

Kappa statistic 0.7164

Mean absolute error 0.1053 *// Là tỷ lệ % lỗi tuyệt đối của Phân lớp*

Root mean squared error 0.3244

Relative absolute error 28.7081 % // *Lỗi tương đối tuyệt đối*

Root relative squared error 76.3182 %

UnClassified Instances 1 5 % *// Là tỷ lệ % không Phân lớp được*

Total Number of Instances 20 *// Tổng số thuộc tính đã được phân lớp*

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 0 1 1 1 1 Iris-Setosa

0.667 0.125 0.5 0.667 0.571 0.688 Iris-Versicolor

0.833 0.143 0.909 0.833 0.87 0.854 Iris-Virginica

Weighted Avg. 0.842 0.11 0.864 0.842 0.85 0.859

=== Confusion Matrix ===

a b c <-- classified as

4 0 0 | a = Iris-Setosa

0 2 1 | b = Iris-Versicolor

0 2 10 | c = Iris-Virginica

***Ý nghĩa:*** Tổng cộng có **20** trường hợp phân lớp được trong đó có **16** phân lớp ***đúng*** và **3** phân lớp ***sai*** và **1** trường hợp ***không phân lớp***được.

- Trong **16** trường hợp phân lớp đúng thì:

+ Có **4** trường là hợp hoa Iris Setosa

+ Có **2** trường hợp là hoa Iris Versicolor

+ Có **10** trường hợp là hoa Iris Virginica

- Trong **3** trường hợp phân lớp sai thì:

+ Có **1** trường hợp là hoa Iris Versicolor (mà Weka trả lời là hoa Iris Virginica)

+ Có **2** trường hợp là hoa Iris Virginica (mà Weka trả lời là hoa Iris Vericolor)

- Còn **1** trường hợp không phân lớp được là của hoa Iris Versicolor