

รายงาน Decision tree

LED Display Domain Data Set([Link](#))

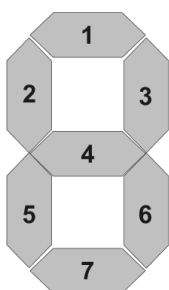
- ชื่อ รหัส นักศึกษา (2 คะแนน)
ชื่อ นาย นนทวัฒน์ กองสีใจ 610610547
- อธิบายรายละเอียดของ attribute และคลาสของข้อมูลที่ได้รับมอบหมาย (3 คะแนน)

เป็นเรื่องเกี่ยวกับ seven segment ที่ใน attribute จะมี {0,1} คือ ติดและดับ ซึ่งมีอยู่ 7 attribute ด้วยกัน และ lable ของเราคือ {1,2,3,4,5,6,7,8,9}

Data Dictionary

| Attribute | Data type | Nullable | Description |
|-------------|-----------|----------|-------------------------------------|
| Attribute_1 | integer | not null | 1 - A 0 - B |
| Attribute_2 | integer | not null | 1 - C 0 - D |
| Attribute_3 | integer | not null | 1 - E 0 - F |
| Attribute_4 | integer | not null | 1 - G 0 - H |
| Attribute_5 | integer | not null | 1 - I 0 - J |
| Attribute_6 | integer | not null | 1 - K 0 - L |
| Attribute_7 | integer | not null | 1 - M 0 - N |
| Attribute_8 | integer | not null | range of number {1,2,3,4,5,6,7,8,9} |

ที่ต้องทำให้ แต่ละตัว Uni แบบนี้เพราะต้องการใช้ Alg จากตัวอย่างในการนับจำนวน



Attribute_1 คือหมายเลข 1 และ Attribute_2 คือหมายเลข 2 ตามลำดับ



3. การทำ data preprocessing ไม่ว่านักศึกษาเห็นว่าชุดข้อมูลที่ได้รับมอบหมายควรทำหรือไม่ให้อธิบาย และแสดงผลการ pre processing (ถ้าได้ทำ) (10 คะแนน)

เขียนโปรแกรมโดยใช้ Algorithm ID3 ที่ต่อยอดจากตัวอย่างที่กำหนดให้ เพื่อใช้กับชุดข้อมูลของตน

ไม่ได้ทำ data preprocessing เพราะ เป็น **Data-Generator** ซึ่งใน Code สามารถใส่ ค่า noisy ผมเลยไม่ใส่ เราจึงได้ข้อมูลที่ดี

4. แสดง Code ที่พัฒนา (15 คะแนน) พร้อมทั้งคำอธิบาย (คำอธิบาย 10 คะแนน)และผลการทำงาน(10 คะแนน)

Code

<https://github.com/NonthawatNK/FindGin>

Link

Root

```
***Gain results of all dataset***
Ginn (Ind) is 3.319
Gain (Att_1) is 0.720
Gain (Att_2) is 0.974
Gain (Att_3) is 0.720
Gain (Att_4) is 0.895
Gain (Att_5) is 0.969
Gain (Att_6) is 0.472
Gain (Att_7) is 0.898
max gain of attb is 0.974 position is 1
```

Att_2

Level 2

Left

```
***Gain results of left dataset***
Ginn (Ind) is 2.584
Gain (Att_1) is 0.641
Gain (Att_3) is 0.919
Gain (Att_4) is 0.634
Gain (Att_5) is 1.000
Gain (Att_6) is 0.000
Gain (Att_7) is 0.641
max gain of attb is 1.000 position is 3
```

Att_5

Right

```
***Gain results of right dataset***
Ginn (Ind) is 1.993
Gain (Att_1) is 0.814
Gain (Att_3) is 0.000
Gain (Att_4) is 0.996
Gain (Att_5) is 0.810
Gain (Att_6) is 0.810
Gain (Att_7) is 0.996
max gain of attb is 0.996 position is 2
```

Att_4

Level 3

Left Left

```
***Gain results of left left dataset***
Ginn (Ind) is 1.584
Gain (Att_1) is 0.000
Gain (Att_3) is 0.929
Gain (Att_6) is 0.000
Gain (Att_7) is 0.000
max gain of attb is 0.929 position is 0
```

Att_3

Left Right

```
***Gain results of left right dataset***
Ginn (Ind) is 1.584
Gain (Att_1) is 0.909
Gain (Att_3) is 0.909
Gain (Att_6) is 0.000
Gain (Att_7) is 0.909
max gain of attb is 0.909 position is 0
```

Att_1

Right Left

```
***Gain results of right left dataset***
Ginn (Ind) is 0.996
Gain (Att_1) is 0.000
Gain (Att_3) is 0.000
Gain (Att_6) is 0.996
Gain (Att_7) is 0.000
max gain of attb is 0.996 position is 2
```

Att_6

Right Right

Level 4

Left Left Left

```
***Gain results of left left left dataset***
Ginn (Ind) is 1.000
Gain (Att_4) is 1.000
Gain (Att_7) is 0.000
max gain of attb is 1.000 position is 1
```

Att_4

Left Left Left

Number 6

Left Right Left

Att_3

Left Right Right

Number 4

Right Left Left

Number 3

Right Left Right

Number 2

Level 5

Left Left Left Left

Number 6

Left Left Left Right

Number 0

Left Right Left Left

Number 9

Left Right Left Right

Number 5

```

***Gain results of right right dataset***
Ginn (Ind) is 0.997
Gain (Att_1) is 0.997
Gain (Att_3) is 0.000
Gain (Att_6) is 0.000
Gain (Att_7) is 0.000
max gain of attb is 0.997 position is 1

```

Right Right Left

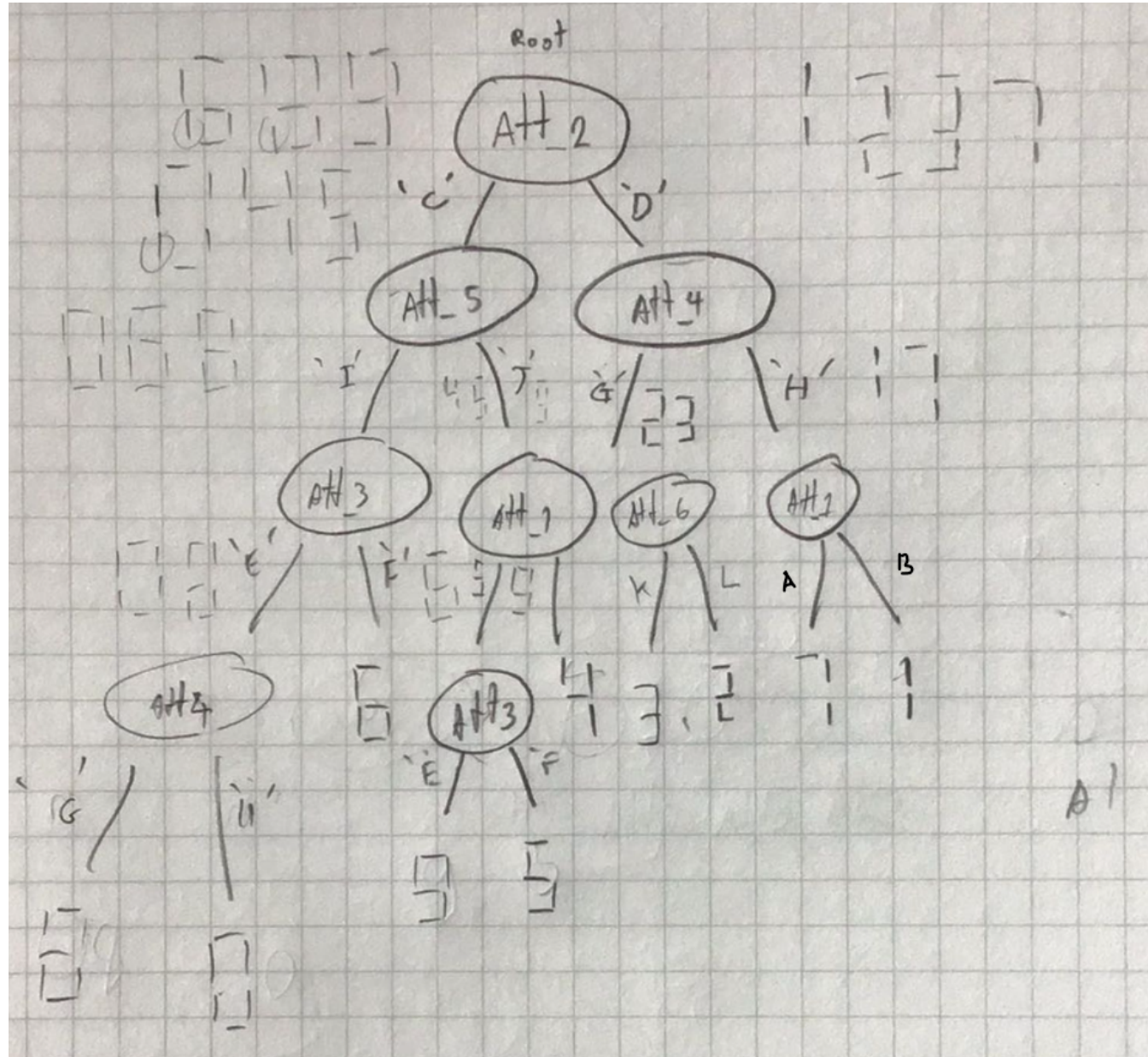
Number 7

Right Right Right

Number 1

Att_1

5. แสดง Decision tree ที่ได้จากการทำงานในข้อ 3(ไม่ต้องเขียนโปรแกรม สามารถ วาดโดยใช้ MS word หรือเครื่องมืออื่นได้) (5 คะแนน)



6. แสดง Code ที่พัฒนาเพื่อสร้าง classification rule จาก Decision tree ที่ได้ และผลการทำงาน(10 คะแนน)


```
1
2 filename = "out.csv"
3 f=open(filename,"r",encoding="utf-8")
4 X=f.readlines()
5 n = len(X)
6 for i in range(0,n):
7     if((X[i].count('A')==1) and (X[i].count('H')==1) and (X[i].count('M')==1) ):
8         print("0")
9     if((X[i].count('B')==1) and (X[i].count('D')==1) and (X[i].count('I')==1)):
10        print("1")
11    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('I')==1) and (X[i].count('L')==1) ):
12        print("2")
13    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('J')==1) and (X[i].count('D')==1) ):
14        print("3")
15    if((X[i].count('B')==1) and (X[i].count('C')==1) ):
16        print("4")
17    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('F')==1) and (X[i].count('J')==1) ):
18        print("5")
19    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('F')==1) and (X[i].count('I')==1) ):
20        print("6")
21    if((X[i].count('A')==1) and (X[i].count('H')==1) and (X[i].count('N')==1) ):
22        print("7")
23    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('I')==1) and (X[i].count('K')==1) ):
24        print("8")
25    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('J')==1) and (X[i].count('C')==1) ):
26        print("9")
```

```
1 A,C,E,H,I,K,M,0
2 B,D,E,H,J,K,N,1
3 A,D,E,G,I,L,M,2
4 A,D,E,G,J,K,M,3
5 B,C,E,G,J,K,N,4
6 A,C,F,G,J,K,M,5
7 A,C,F,G,I,K,M,6
8 A,D,E,H,J,K,N,7
9 A,C,E,G,I,K,M,8
10 A,C,E,G,J,K,M,9
11
```

Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

D:\data_mining\Dessiontree\FindGin>python -u "d:\data_mining\Dessiontree\FindGin\Check.py"

```
0
1
2
3
4
5
6
7
8
9
```

7. แสดง Code ที่พัฒนาเพื่อทดสอบ ชุดข้อมูลกับโครงสร้าง Tree ผ่าน Rule ที่สร้างขึ้นและวิเคราะห์ผลการทำงาน อธิบายความสอดคล้องกับโครงสร้าง tree ที่สร้างขึ้นกับผลการจำแนก(10 คะแนน)

```
1
2 filename = "test_relace.csv"
3 f=open(filename,"r",encoding="utf-8")
4 X=f.readlines()
5 n = len(X)
6 for i in range(0,n):
7     if((X[i].count('A')==1) and (X[i].count('H')==1) and (X[i].count('M')==1) ):
8         print("0")
9     if((X[i].count('B')==1) and (X[i].count('D')==1) and (X[i].count('I')==1)):
10        print("1")
11    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('I')==1) and (X[i].count('L')==1) ):
12        print("2")
13    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('J')==1) and (X[i].count('D')==1) ):
14        print("3")
15    if((X[i].count('B')==1) and (X[i].count('C')==1) ):
16        print("4")
17    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('F')==1) and (X[i].count('J')==1) ):
18        print("5")
19    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('F')==1) and (X[i].count('I')==1) ):
20        print("6")
21    if((X[i].count('A')==1) and (X[i].count('H')==1) and (X[i].count('N')==1) ):
22        print("7")
23    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('I')==1) and (X[i].count('K')==1) ):
24        print("8")
25    if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('E')==1) and (X[i].count('J')==1) and (X[i].count('C')==1) ):
26        print("9")
```

```
1 A,C,E,G,J,K,M,9
2 A,D,E,H,J,K,N,7
3 B,D,E,H,J,K,N,1
4 A,C,E,G,J,K,M,9
5 B,C,E,G,J,K,N,4
6 A,C,F,G,J,K,M,5
7 A,C,E,G,J,K,M,9
8 B,D,E,H,J,K,N,1
9 A,C,F,G,J,K,M,5
10 A,C,E,H,I,K,M,0
11 A,C,E,G,J,K,M,9
12 A,C,E,G,I,K,M,8
13 A,C,E,H,I,K,M,0
14 A,C,E,G,J,K,M,9
15 A,C,E,G,J,K,M,9
16 A,D,E,G,I,L,M,2
17 A,C,F,G,I,K,M,6
18 A,D,E,H,J,K,N,7
19 A,C,E,G,J,K,M,9
20 A,C,F,G,I,K,M,6
21
```

Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

D:\data_mining\Dessiontree\FindGin>python -u "d:\data_mining\Dessiontree\FindGin\Check.py"

```
9
7
1
9
4
5
9
1
5
0
9
8
0
9
2
6
7
9
6
```

จากกฎที่สร้างขึ้นเองเมื่อเอาไปเทียบกับข้อมูลสำหรับหรับการสร้าง Tree แล้วผลลัพธ์ออกมาแม่นยำเท่ากัน