

## รายงาน Decision tree

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

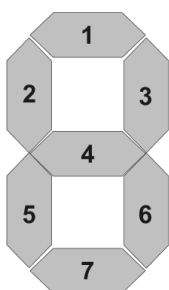
- ชื่อ รหัส นักศึกษา (2 คะแนน)  
ชื่อ นาย นนทวัฒน์ กองสีใจ 610610547
- อธิบายรายละเอียดของ attribute และคลาสของข้อมูลที่ได้รับมอบหมาย (3 คะแนน)

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 ไม่ว่านักศึกษาเห็นว่าชุดข้อมูลที่รับมอบหมายควรทำหรือไม่ให้อธิบาย 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 0.413
Gain (Att_2) is 0.535
Gain (Att_3) is 0.402
Gain (Att_4) is 0.473
Gain (Att_5) is 0.517
Gain (Att_6) is 0.240
Gain (Att_7) is 0.489
max gain of attb is 0.535 position is 1
```

Att\_1

#### Level 2

Left

```
***Gain results of left dataset***
Ginn (Ind) is 3.078
Gain (Att_2) is 2.089
Gain (Att_3) is 1.095
Gain (Att_4) is 2.368
Gain (Att_5) is 1.744
Gain (Att_6) is 0.652
Gain (Att_7) is 0.320
max gain of attb is 2.368 position is 2
```

Att\_4

Right

```
***Gain results of right dataset***
Ginn (Ind) is 2.424
Gain (Att_2) is 2.424
Gain (Att_3) is 2.424
Gain (Att_4) is 2.424
Gain (Att_5) is 2.424
Gain (Att_6) is 2.424
Gain (Att_7) is 1.417
max gain of attb is 2.424 position is 0
```

Att\_2

#### Level 3

Left Left

```
***Gain results of left dataset***
Ginn (Ind) is 2.812
Gain (Att_3) is 0.514
Gain (Att_5) is 0.513
Gain (Att_6) is 0.330
Gain (Att_7) is 0.095
max gain of attb is 0.514 position is 0
```

Att\_3

Left Right

```
***Gain results of left right dataset***
Ginn (Ind) is 2.273
Gain (Att_3) is 0.186
Gain (Att_5) is 0.562
Gain (Att_6) is 0.125
Gain (Att_7) is 0.576
max gain of attb is 0.576 position is 3
```

Att\_7

Right Left

```
***Gain results of right left dataset***
Ginn (Ind) is 1.955
Gain (Att_3) is 0.340
Gain (Att_5) is 0.420
Gain (Att_6) is 0.064
Gain (Att_7) is 0.542
max gain of attb is 0.542 position is 3
```

Att\_7

Right Right

#### Level 4

Left Left Left

```
***Gain results of left left left dataset***
Ginn (Ind) is 2.513
Gain (Att_5) is 0.489
Gain (Att_6) is 0.464
max gain of attb is 0.489 position is 0
```

Att\_5

Left Left Right

```
***Gain results of left left right dataset***
Ginn (Ind) is 1.966
Gain (Att_5) is 0.566
Gain (Att_6) is 0.121
max gain of attb is 0.566 position is 0
```

Att\_5

#### Level 5

Left Left Left Left

```
***Gain results of left left left left dataset***
Ginn (Ind) is 1.876
Gain (Att_6) is 0.634
max gain of attb is 0.634 position is 0
```

Left Left Left Right

```
***Gain results of left left left right dataset***
Ginn (Ind) is 2.175
Gain (Att_2) is 0.635
Gain (Att_6) is 0.178
max gain of attb is 0.635 position is 0
```

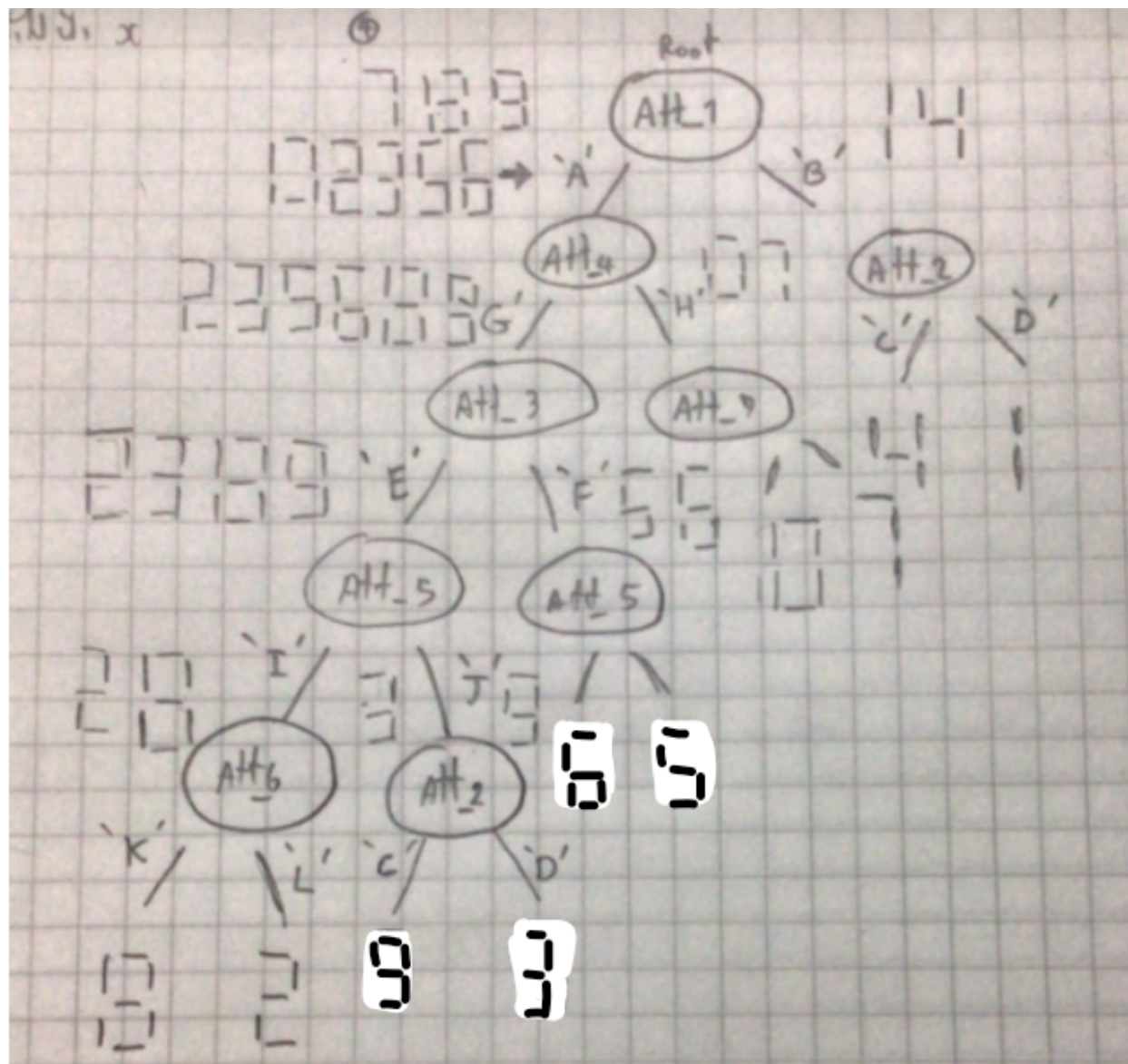
```

***Gain results of right right dataset***
Ginn (Ind) is 1.799
Gain (Att_3) is 0.047
Gain (Att_5) is 0.256
Gain (Att_6) is 0.262
Gain (Att_7) is 0.393
max gain of attb is 0.393 position is 3

```

Att\_7

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



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

```

1  filename = "out.csv"
2  f=open(filename,"r",encoding="utf-8")
3  X=f.readlines()
4  n = len(X)
5  for i in range(0,n):
6      if((X[i].count('A')==1) and (X[i].count('H')==1) and (X[i].count('M')==1) ):
7          print("0")
8      if((X[i].count('B')==1) and (X[i].count('D')==1) and (X[i].count('I')==1)):
9          print("1")
10     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) ):
11         print("2")
12     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) ):
13         print("3")
14     if((X[i].count('B')==1) and (X[i].count('C')==1) ):
15         print("4")
16     if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('F')==1) and (X[i].count('J')==1) ):
17         print("5")
18     if((X[i].count('A')==1) and (X[i].count('G')==1) and (X[i].count('F')==1) and (X[i].count('I')==1) ):
19         print("6")
20     if((X[i].count('A')==1) and (X[i].count('H')==1) and (X[i].count('N')==1) ):
21         print("7")
22     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) ):
23         print("8")
24     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) ):
25         print("9")
26
Microsoft Windows [Version 10.0.22621.1105]
(c) Microsoft Corporation. All rights reserved.

D:\data_mining\Decisiontree\FindGin>python -u "d:\data_mining\Decisiontree\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)

1234567891011121314151617181920212223242526

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

test\_relace.csv

```
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
```

4

OUTPUT

TERMINAL

COMMENTS

DEBUG CONSOLE

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

D:\data_mining\Deessiontree\FindGin>python -u "d:\data_mining\Deessiontree\FindGin\Check.py"
9
7
1
9
4
5
9
1
5
0
9
9
8
0
9
9
9
2
6
7
9
6
```

จากที่เราสร้างโดยกฎสอดคล้องกันมากตาม Tree ที่เราได้ทำมาจากการหาค่า Gain ถูกต้องทั้งหมด