

Pattern Recognition

จัดทำโดย

นาย วีระกิจ เกษพอง

รหัสนิสิต 6440206420

งานมอบหมายนี้เป็นส่วนหนึ่งของวิชา

01418384 การวิเคราะห์รูปภาพและคอมพิวเตอร์วิทัศน์

ภาคปลาย ปีการศึกษา 2566

1. ชุดข้อมูลภาพที่ใช้ในงานนี้ได้แก่

ชุดข้อมูลที่ประกอบด้วยภาพชื่อ NutRais-01.jpg - NutRais-30.jpg

วัตถุในภาพประกอบด้วย อัลมอนต์, ลูกเกดแดง, มะม่วงหิมพานต์

โดยมีการแบ่งข้อมูลดังนี้

1.1 Training set ได้แก่ภาพ NutRais-01.jpg - NutRais-30.jpg

1.2 Test set ได้แก่ภาพ NutRais-_31.jpg - NutRais-_80.jpg

2. ฟีเจอร์ที่ถูกสกัดจากชุดข้อมูลภาพประกอบด้วย

2.1 Area

2.2 Eccentricity

2.3 MajorAxisLength

2.4 MinorAxisLength

2.5 Perimeter

2.6 Mean

2.7 Std

2.8 Skewness

2.9 Energy

2.10 Snr

2.11 Kurtosis

2.12 Powerbw

2.13 Obw

2.14 Enbw

3. ผลลัพธ์จากการสร้างโมเดล

3.1 C4.5

3.1.1 Decision Tree

W-J48

J48 pruned tree

```
OBW <= 0.094229
| Perimeter <= 1638.772
| | ENBW <= 1.015742: Almond (9.0)
| | ENBW > 1.015742: Red-raisin (3.0)
| Perimeter > 1638.772: Cashew (76.0/1.0)
OBW > 0.094229
| Eccentricity <= 0.828044: Almond (106.0)
| Eccentricity > 0.828044
| | Energy <= 110563.451048: Almond (21.0)
| | Energy > 110563.451048
| | | Mean <= 0.715814: Almond (3.0)
| | | Mean > 0.715814
| | | Eccentricity <= 0.830888
| | | Area <= 175454: Red-raisin (2.0)
| | | Area > 175454: Almond (3.0)
| | Eccentricity > 0.830888: Red-raisin (59.0/1.0)
```

Number of Leaves : 9

Size of the tree : 17

3.1.2 ผลความแม่นยำที่ได้รับจากการฝึกข้อมูล

a) Confusion matrix

accuracy: 96.07% +/- 4.28% (micro average: 96.10%)

	true Red-raisin	true Almond	true Cashew	class precision
pred. Red-raisin	61	5	0	92.42%
pred. Almond	1	138	3	97.18%
pred. Cashew	1	1	72	97.30%
class recall	96.83%	95.83%	96.00%	

b) Accuracy ที่ได้รับจากการฝึกข้อมูล 96.07%

3.2 C5.0

3.2.1 Decision Tree

Decision tree:

```
OBW <= 0.09422864:
...Std > 0.1080094: Red-raisin (3)
: Std <= 0.1080094:
: ...MajorAxisLength <= 627.0718: Almond (9)
: ...MajorAxisLength > 627.0718: Cashew (76/1)
OBW > 0.09422864:
...Std > 0.1430105: Red-raisin (30)
: Std <= 0.1430105:
: ...Mean <= 0.8259368: Almond (134)
: ...Mean > 0.8259368: Red-raisin (30)
```

Evaluation on training data (282 cases):

```
Decision Tree
-----
Size      Errors
6      1( 0.4%)  <<

(a) (b) (c)  <-classified as
-----
143    1      (a): class Almond
      75      (b): class Cashew
           63  (c): class Red-raisin
```

Attribute usage:

```
100.00% Std
100.00% OBW
58.16% Mean
30.14% MajorAxisLength
```

Time: 0.0 secs

3.2.2 ผลความแม่นยำที่ได้รับจากการฝึกข้อมูล

a) Confusion matrix

	True Almond	True Cashew	True Red-raisin	Class precision
Pred.Almond	140	2	1	99.30%
Pred.Cashew	0	75	0	100%
Pred.Red-raisin	1	0	63	100%
Class recall	100%	98.68%	100%	

b) Accuracy ที่ได้รับจากการฝึกข้อมูล $9281/282 = 99.64\%$

3.3 RapidMiner Decision tree

3.3.1 Decision Tree

Tree

```
OBW > 0.085
|   MinorAxisLength > 467.881: Cashew {Red-raisin=0, Almond=0, Cashew=2}
|   MinorAxisLength ≤ 467.881
|   |   Mean > 0.862: Red-raisin {Red-raisin=30, Almond=0, Cashew=0}
|   |   Mean ≤ 0.862: Almond {Red-raisin=30, Almond=138, Cashew=0}
OBW ≤ 0.085
|   Mean > 0.865: Red-raisin {Red-raisin=3, Almond=0, Cashew=0}
|   Mean ≤ 0.865
|   |   MajorAxisLength > 627.144: Cashew {Red-raisin=0, Almond=1, Cashew=73}
|   |   MajorAxisLength ≤ 627.144: Almond {Red-raisin=0, Almond=5, Cashew=0}
```

3.3.2 ผลความแม่นยำที่ได้รับจากการฝึกข้อมูล

a) Confusion matrix

accuracy: 87.20% +/- 7.71% (micro average: 87.23%)

	true Red-raisin	true Almond	true Cashew	class precision
pred. Red-raisin	30	0	1	96.77%
pred. Almond	33	143	1	80.79%
pred. Cashew	0	1	73	98.65%
class recall	47.62%	99.31%	97.33%	

b) Accuracy ที่ได้รับจากการฝึกข้อมูล 87.20%

3.4 CART

3.4.1 Decision Tree

W-SimpleCart

CART Decision Tree

```
OBW < 0.08496052658058237
| Perimeter < 1639.6034999999997
| | Area < 174811.0: Almond(5.0/0.0)
| | Area >= 174811.0: Red-raisin(3.0/0.0)
| Perimeter >= 1639.6034999999997: Cashew(73.0/1.0)
OBW >= 0.08496052658058237
| Eccentricity < 0.8281199580370893
| | Area < 240920.0: Almond(110.0/0.0)
| | Area >= 240920.0: Cashew(2.0/0.0)
| Eccentricity >= 0.8281199580370893
| | Energy < 110639.49342865852: Almond(21.0/0.0)
| | Energy >= 110639.49342865852
| | | Mean < 0.7162335699339308: Almond(3.0/0.0)
| | | Mean >= 0.7162335699339308
| | | Std < 0.11448967228987592: Almond(2.0/0.0)
| | | Std >= 0.11448967228987592: Red-raisin(60.0/2.0)
```

Number of Leaf Nodes: 9

Size of the Tree: 17

3.4.2 ผลความแม่นยำที่ได้รับจากการฝึกข้อมูล

a) Confusion matrix

accuracy: 93.57% +/- 6.25% (micro average: 93.62%)

	true Red-raisin	true Almond	true Cashew	class precision
pred. Red-raisin	58	8	0	87.88%
pred. Almond	4	135	4	94.41%
pred. Cashew	1	1	71	97.26%
class recall	92.06%	93.75%	94.67%	

b) Accuracy ที่ได้รับจากการฝึกข้อมูล 93.57%

4. การเลือกโมเดลมาใช้ในการระบุชนิดวัตถุ

โมเดลที่เลือกคือ C5.0 เพราะ ได้ค่าความแม่นยำ 99.64%

5. ผลการทดสอบความแม่นยำ

5.1 ผลการทดสอบความแม่นยำเมื่อทดสอบด้วย Training set(30)

a) ให้ค่าความแม่นยำที่ 99.64%

5.2 ผลการทดสอบความแม่นยำเมื่อทดสอบด้วย Test set(50)

	True Almond	True Cashew	True Red-raisin	Class precision
Pred.Almond	148	0	4	97.36%
Pred.Cashew	0	215	7	96.84%
Pred.Red-raisin	0	0	91	100%
Class recall	100%	100%	90.09%	

b) ให้ค่าความแม่นยำที่ 97.63%

6. นิสิตได้รับความรู้ะไรจากการเรียนและการทำงานมอบหมายนี้

6.1 การจะสกัดฟีเจอร์ของข้อมูลจะต้องคำนึงถึงปริมาณของข้อมูลด้วยเพราะจะมีผลต่อการทำนาย

โมเดล

6.2 ได้ความแตกต่างของข้อมูล เช่นต่อให้ใช้ข้อมูลเดียวกัน แต่คนละโมเดลก็จะให้ผลการทำนายที่

ต่างกัน

6.3 ได้ทราบถึงการทำข้อมูลเริ่มต้นหรือการเก็บข้อมูลนั้นมีความสำคัญมาก เพราะถ้าเราเก็บหรือจัดการ
ได้ไม่ดีจะส่งผลต่อโมเดลของเราได้

7. สรุปปัญหาและอุปสรรคที่พบจากการทำงานมอบหมาย

7.1 โปรแกรม R Studio มีการอ่านค่า cvs เกินมา ทำให้ไม่สามารถทำ decision tree ได้ ต้องทำการ
แก้ไขไฟล์ csv ใหม่

7.2 การทำ decision tree ที่ c 5.0 ครั้งแรก ไม่สามารถทำการทำนายผลได้ เพราะ โค้ดในการสกัด
ฟีเจอร์ผิดพลาดทำให้ ต้องทำการแก้ไขโค้ดในการสกัดฟีเจอร์ใหม่

7.3 การตรวจสอบความแม่นยำของโปรแกรมนั้นใช้วิธีตรวจแบบรันทีละรูปแล้วตรวจสอบทำให้เกิด
ความล่าช้าในการทำผลสรุป