บทที่ 4

การพัฒนาโปรแกรม C4.5 บนระบบปฏิบัติการวินโดว์

ในวิทยานิพนธ์ฉบับนี้จะเป็นการพัฒนาโปรแกรม C4.5 ซึ่งปกติจะทำงานอยู่บนระบบปฏิบัติการยูนิกซ์ (Unix) ให้สามารถทำงานกับเครื่องไมโครคอมพิวเตอร์ บนระบบปฏิบัติการไมโครซอฟท์วินโดว์ 95 (Microsoft Windows 95) ได้ โดยได้มีการเพิ่มส่วนของการจัดการและเตรียมข้อมูลที่จะใช้สอน โดยสามารถที่จะแก้ไข เพิ่มหรือลด ลักษณะหรือจำนวนข้อมูลที่ใช้สอน และได้มีการเปลี่ยนแปลงส่วนของการติดต่อกับผู้ใช้และแสดงผล จากเดิมที่เป็นแบบการสั่งรายบรรทัด (Command Line) มาเป็นแบบกราฟฟิก (Graphic User Interface) หลังจากนั้นจะได้มีการทดสอบความถูกต้องของโปรแกรม C4.5 ใหม่ ที่ทำงานบนระบบปฏิบัติการวินโดว์ เปรียบเทียบกับ C4.5 ที่ทำงานบนระบบปฏิบัติการยูนิกซ์ และทดสอบโปรแกรมที่ได้กับข้อมูลการพิจารณาอนุมัติสินเชื่อบุคคล เพื่อศึกษาความเป็นไปได้ในการนำไปใช้ต่อไป

โปรแกรม C4.5 ที่ทำงานบนระบบปฏิบัติการยูนิกซ์

โปรแกรม C4.5 เป็นโปรแกรมที่พัฒนาโดยใช้ภาษา C บนระบบปฏิบัติการยูนิกซ์ โดยโปรแกรมที่ใช้เป็นต้นแบบในการพัฒนาจะเป็นรุ่นที่ 8 (Release 8.0) ซึ่งมีโครงสร้างและการทำงานเหมือนกับโปรแกรม C4.5 ในรุ่นก่อนหน้า เพียงแต่ได้มีการปรับปรุงความสามารถในการเรียนรู้ให้ถูกต้องมากยิ่งขึ้น โดยจะประกอบด้วยโปรแกรมหลัก 4 โปรแกรม คือ

1. c4.5 เป็นโปรแกรมที่ใช้ในการเรียนรู้และสร้างต้นไม้ตัดสินใจจากตัวอย่างสอน
2. c4.5rules เป็นโปรแกรมที่ใช้ในการสร้างกฎการตัดสินใจจากต้นไม้ตัดสินใจที่ได้ในข้อ 1
3. consult เป็นโปรแกรมที่ตอบโต้กับผู้ใช้ในการตัดสินใจในตัวอย่างที่ผู้ใช้ป้อนให้ โดยใช้ต้นไม้ตัดสินใจ ที่ได้จากข้อ 1
4. consultr เป็นโปรแกรมที่ตอบโต้กับผู้ใช้ในการตัดสินใจในตัวอย่างที่ผู้ใช้ป้อนให้ โดยใช้กฎการตัดสินใจที่ได้จากข้อ 2

ในการทำงานของโปรแกรมจำเป็นต้องมีข้อมูลตัวอย่างที่ใช้ในการเรียนรู้ และทดสอบความถูกต้องของต้นไม้ตัดสินใจหรือกฎที่ได้ ซึ่งประกอบด้วยแฟ้มข้อความ (Text File) 3 แฟ้มด้วยกัน ซึ่งมีชื่อเดียวกัน แต่ต่างกันที่ส่วนขยายของแฟ้ม (File Extension) ดังนี้

1. แฟ้มชื่อ (Names File) อยู่ในรูปของ ชื่อแฟ้ม.names จะเก็บชื่อของพวกและชื่อของลักษณะต่าง รวมทั้งชนิดของข้อมูลและค่าที่เป็นไปได้ในแต่ละลักษณะ ดังรูปที่ 4.1 โดยมีรูปแบบที่

เฉพาะคือ ในบรรทัดแรกของแฟ้มจะเก็บค่าที่เป็นไปได้ของพวกเรียงกันไป โดยมีเครื่องหมาย "," เป็นตัวแบ่งแต่ละค่าออกจากกัน และจะปิดท้ายด้วยเครื่องหมาย "." เมื่อสิ้นสุดข้อมูลของพวก ต่อจากนั้นในบรรทัดต่อไปจะเป็นชื่อและค่าที่เป็นไปได้ในแต่ละลักษณะ โดยเริ่มด้วยชื่อของแต่ละลักษณะตามด้วยเครื่องหมาย ":" เพื่อบอกจุดสิ้นสุดของชื่อ ต่อจากนั้นจะตามด้วยค่าที่เป็นไปได้ในแต่ละลักษณะ ถ้าลักษณะใดเป็นค่าต่อเนื่องจะใช้คำว่า "continuous" แต่ถ้าลักษณะใดเป็นค่าไม่ต่อเนื่อง จะใส่ค่าที่เป็นไปได้ของลักษณะนั้นต่อเนื่องกันไป โดยมีเครื่องหมาย "," เป็นตัวแบ่งแยกแต่ละค่าออกจากกัน และปิดท้ายด้วยเครื่องหมาย "." เพื่อบอกจุดสิ้นสุดในแต่ละลักษณะ

```
good, bad.

duration:                          continuous.
wage increase first year:          continuous.
wage increase second year:         continuous.
wage increase third year:          continuous.
cost of living adjustment:         none, tcf, tc.
working hours:                     continuous.
pension:                           none, ret_allw, empl_contr.
standby pay:                       continuous.
shift differential:                continuous.
education allowance:               yes, no.
statutory holidays:                continuous.
vacation:                          below average, average, generous.
longterm disability assistance:    yes, no.
contribution to dental plan:       none, half, full.
bereavement assistance:            yes, no.
contribution to health plan:       none, half, full.
```

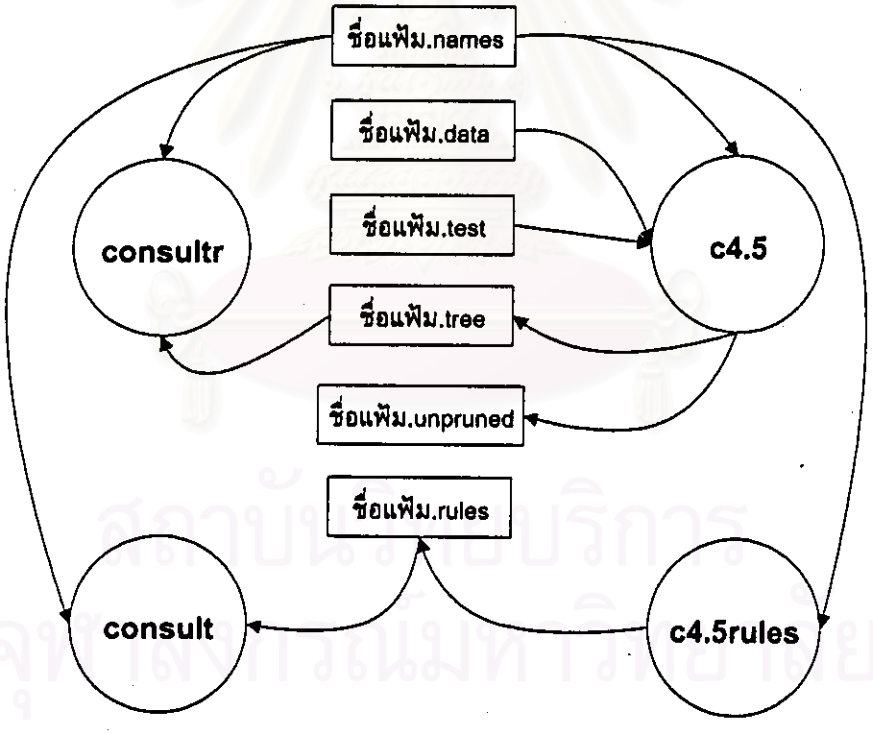
รูปที่ 4.1 ตัวอย่างของแฟ้มชื่อ labor-neg.names

2. แฟ้มตัวอย่างสอน (Data File) อยู่ในรูป "ชื่อแฟ้ม.data" จะเก็บรายละเอียดของข้อมูลแต่ละตัวอย่างที่ใช้ในการสอน ดังรูปที่ 4.3 โดยข้อมูลแต่ละตัวอย่างจะเก็บค่าของแต่ละลักษณะเรียงกันไปเป็นลำดับตามที่กำหนดไว้ในแฟ้มชื่อ เริ่มจากค่าของลักษณะไม่แบ่งพวกลำดับแรก แล้วปิดด้วยเครื่องหมาย "," เพื่อแยกค่าแต่ละลักษณะออกจากกัน จากนั้นจะตามด้วยค่าของลักษณะที่สอง ที่สาม ตามลำดับจนครบทุกลักษณะที่กำหนดไว้ในแฟ้มชื่อ ถ้าค่าของลักษณะใดไม่ทราบค่าจะใช้เครื่องหมาย "?" แทนค่านั้น โดยลักษณะสุดท้ายจะเป็นลักษณะแบ่งพวก แล้วตามด้วยเครื่องหมาย "." เพื่อบอกจุดสิ้นสุดของข้อมูล 1 ตัวอย่าง ข้อมูลตัวอย่างตัวต่อไปก็จะเก็บในลักษณะเช่นนี้จนจบแฟ้ม
3. แฟ้มตัวอย่างทดสอบ (Test File) อยู่ในรูป "ชื่อแฟ้ม.test" จะเก็บข้อมูลตัวอย่างที่ใช้ในการทดสอบความถูกต้องของต้นไม้ตัดสินใจหรือกฎที่ได้ดังรูปที่ 4.4 โดยมีโครงสร้างของแฟ้มเหมือนกับตัวอย่างสอน เพียงแต่เป็นคนละตัวอย่างกันเท่านั้น

ในการทำงานของโปรแกรม C4.5 จะสั่งงานโดยการส่งผ่านค่าตัวเลือก (Option) ให้กับโปรแกรมตามตัวเลือกที่กำหนดไว้ ดังตารางที่ 4.1 สำหรับโปรแกรม C4.5 และตารางที่ 4.2 สำหรับโปรแกรม

c4.5rules เมื่อโปรแกรมประมวลผลเสร็จจะส่งผลลัพธ์ออกทางจอภาพและทางแฟ้มข้อมูล โดยผลลัพธ์ที่ได้ทางจอภาพจะประกอบด้วยต้นไม้ตัดสินใจหรือกฎการตัดสินใจ พร้อมด้วยผลการทดสอบความถูกต้องกับตัวอย่างที่ใช้ในการเรียนรู้หรือตัวอย่างที่ใช้ในการทดสอบ ดังรูปที่ 4.5 สำหรับโปรแกรม c4.5 และรูปที่ 4.6 สำหรับโปรแกรม c4.5rules ส่วนผลลัพธ์ที่เก็บลงแฟ้มจะประกอบด้วยแฟ้มข้อมูล 3 แฟ้ม คือ

1. แฟ้มต้นไม้ตัดสินใจก่อนตัดแต่ง (Unpruned Decision Tree) สร้างจากโปรแกรม c4.5 อยู่ในรูป "ชื่อแฟ้ม.unpruned" เก็บต้นไม้ตัดสินใจก่อนการตัดแต่งกิ่ง เพื่อที่จะนำไปใช้ในการสร้างกฎการตัดสินใจต่อไป
2. แฟ้มต้นไม้ตัดสินใจหลังจากตัดแต่ง (Pruned Decistion Tree) สร้างจากโปรแกรม c4.5 อยู่ในรูป "ชื่อแฟ้ม.tree" เก็บต้นไม้ตัดสินใจหลังจากตัดแต่งกิ่งแล้ว เพื่อนำไปใช้กับโปรแกรม consult ในการตัดสินใจกับตัวอย่างที่ผู้ใช้ป้อนให้
3. แฟ้มกฎการตัดสินใจ (Production Rule) สร้างจากโปรแกรม c4.5rules อยู่ในรูป "ชื่อแฟ้ม.rules" เก็บกฎการตัดสินใจ เพื่อนำไปใช้กับโปรแกรม consultr ในการตัดสินใจกับตัวอย่างที่ผู้ใช้ป้อนให้



รูปที่ 4.2 แสดงการติดต่อกับแฟ้มข้อมูลต่าง ๆ ของโปรแกรม C4.5

การติดต่อกับแฟ้มข้อมูลของโปรแกรม C4.5 ได้แสดงไว้ดังรูปที่ 4.2 ส่วนการทำงานของโปรแกรม C4.5 เดิมจะมีหลักการทำงานเป็นดังที่ได้อธิบายไว้ในบทที่แล้ว จึงไม่ขอกล่าวถึงอีก

```
1,5.0,?,?,?,40,?,?,2,?,11,average,?,?,yes,?,good.
2,4.5,5.8,?,?,35,ret_allw,?,?,yes,11,below average,?,full,?,full,good.
?,?,?,?,?,38,empl_contr,?,5,?,11,generous,yes,half,yes,half,good.
3,3.7,4.0,5.0,tc,?,?,?,?,yes,?,?,?,?,yes,?,good.
3,4.5,4.5,5.0,?,40,?,?,?,?,12,average,?,half,yes,half,good.
2,2.0,2.5,?,?,35,?,?,6,yes,12,average,?,?,?,?,good.
3,4.0,5.0,5.0,tc,?,empl_contr,?,?,?,12,generous,yes,none,yes,half,good.
3,6.9,4.8,2.3,?,40,?,?,3,?,12,below average,?,?,?,?,good.
2,3.0,7.0,?,?,38,?,12,25,yes,11,below average,yes,half,yes,?,good.
1,5.7,?,?,none,40,empl_contr,?,4,?,11,generous,yes,full,?,?,good.
3,3.5,4.0,4.6,none,36,?,?,3,?,13,generous,?,?,yes,full,good.
2,6.4,6.4,?,?,38,?,?,4,?,15,?,?,full,?,?,good.
2,3.5,4.0,?,none,40,?,?,2,no,10,below average,no,half,?,half,bad.
3,3.5,4.0,5.1,tcf,37,?,?,4,?,13,generous,?,full,yes,full,good.
1,3.0,?,?,none,36,?,?,10,no,11,generous,?,?,?,?,good.
2,4.5,4.0,?,none,37,empl_contr,?,?,?,11,average,?,full,yes,?,good.
1,2.8,?,?,?,35,?,?,2,?,12,below average,?,?,?,?,good.
1,2.1,?,?,tc,40,ret_allw,2,3,no,9,below average,yes,half,?,none,bad.
1,2.0,?,?,none,38,none,?,?,yes,11,average,no,none,no,none,bad.
2,4.0,5.0,?,tcf,35,?,13,5,?,15,generous,?,?,?,?,good.
2,4.3,4.4,?,?,38,?,?,4,?,12,generous,?,full,?,full,good.
2,2.5,3.0,?,?,40,none,?,?,?,11,below average,?,?,?,?,bad.
3,3.5,4.0,4.6,tcf,27,?,?,?,?,?,?,?,?,?,?,good.
2,4.5,4.0,?,?,40,?,?,4,?,10,generous,?,half,?,full,good.
1,6.0,?,?,?,38,?,8,3,?,9,generous,?,?,?,?,good.
3,2.0,2.0,2.0,none,40,none,?,?,?,10,below average,?,half,yes,full,bad.
2,4.5,4.5,?,tcf,?,?,?,?,yes,10,below average,yes,none,?,half,good.
2,3.0,3.0,?,none,33,?,?,?,yes,12,generous,?,?,yes,full,good.
2,5.0,4.0,?,none,37,?,?,5,no,11,below average,yes,full,yes,full,good.
3,2.0,2.5,?,?,35,none,?,?,?,10,average,?,?,yes,full,bad.
3,4.5,4.5,5.0,none,40,?,?,?,no,11,average,?,half,?,?,good.
3,3.0,2.0,2.5,tc,40,none,?,5,no,10,below average,yes,half,yes,full,bad.
2,2.5,2.5,?,?,38,empl_contr,?,?,?,10,average,?,?,?,?,bad.
2,4.0,5.0,?,none,40,none,?,3,no,10,below average,no,none,?,none,bad.
3,2.0,2.5,2.1,tc,40,none,2,1,no,10,below average,no,half,yes,full,bad.
2,2.0,2.0,?,none,40,none,?,?,no,11,average,yes,none,yes,full,bad.
1,2.0,?,?,tc,40,ret_allw,4,0,no,11,generous,no,none,no,none,bad.
1,2.8,?,?,none,38,empl_contr,2,3,no,9,below average,yes,half,?,none,bad.
3,2.0,2.5,2.0,?,37,empl_contr,?,?,?,10,average,?,?,yes,none,bad.
2,4.5,4.0,?,none,40,?,?,4,?,12,average,yes,full,yes,half,good.
```

รูปที่ 4.3 ตัวอย่างของแฟ้มตัวอย่างสอน labor-neg.data

```
1,4.0,?,?,none,?,none,?,?,yes,11,average,no,none,no,none,bad.
2,2.0,3.0,?,none,38,empl_contr,?,?,yes,12,generous,yes,none,yes,full,bad.
2,2.5,2.5,?,tc,39,empl_contr,?,?,?,12,average,?,?,yes,?,bad.
2,2.5,3.0,?,tcf,40,none,?,?,?,11,below average,?,?,yes,?,bad.
2,4.0,4.0,?,none,40,none,?,3,?,10,below average,no,none,?,none,bad.
2,4.5,4.0,?,?,40,?,?,2,no,10,below average,no,half,?,half,bad.
2,4.5,4.0,?,none,40,?,?,5,?,11,average,?,full,yes,full,good.
2,4.6,4.6,?,tcf,38,?,?,?,?,?,?,yes,half,?,half,good.
2,5.0,4.5,?,none,38,?,14,5,?,11,below average,yes,?,?,full,good.
2,5.7,4.5,?,none,40,ret_allw,?,?,?,11,average,yes,full,yes,full,good.
2,7.0,5.3,?,?,?,?,?,?,?,11,?,yes,full,?,?,good.
3,2.0,3.0,?,tcf,?,empl_contr,?,?,yes,?,?,yes,half,yes,?,good.
3,3.5,4.0,4.5,tcf,35,?,?,?,?,13,generous,?,?,yes,full,good.
3,4.0,3.5,?,none,40,empl_contr,?,6,?,11,average,yes,full,?,full,good.
3,5.0,4.4,?,none,38,empl_contr,10,6,?,11,generous,yes,?,?,full,good.
3,5.0,5.0,5.0,?,40,?,?,?,?,12,average,?,half,yes,half,good.
3,6.0,6.0,4.0,?,35,?,?,14,?,9,generous,yes,full,yes,full,good.
```

รูปที่ 4.4 ตัวอย่างของแฟ้มตัวอย่างทดสอบ labor-neg.test

```
C4.5 [release 8] decision tree generator   Tue Feb 17 20:23:14 1998
----------------------------------------

    Options:
      File stem <labor-neg>

Read 40 cases (16 attributes) from labor-neg.data

Decision Tree:

wage increase first year <= 2.5 :
|   statutory holidays <= 10 : bad (6.0)
|   statutory holidays > 10 :
|   |   working hours <= 38 : good (2.3/1.0)
|   |   working hours > 38 : bad (3.0)
wage increase first year > 2.5 :
|   statutory holidays > 10 : good (21.2)
|   statutory holidays <= 10 :
|   |   wage increase first year <= 4 : bad (4.5/0.5)
|   |   wage increase first year > 4 : good (3.0)


Simplified Decision Tree:

wage increase first year <= 2.5 : bad (11.3/2.8)
wage increase first year > 2.5 :
|   statutory holidays > 10 : good (21.2/1.3)
|   statutory holidays <= 10 :
|   |   wage increase first year <= 4 : bad (4.5/1.7)
|   |   wage increase first year > 4 : good (3.0/1.1)


Tree saved


Evaluation on training data (40 items):

       Before Pruning          After Pruning
      ----------------   ---------------------------
      Size      Errors   Size      Errors    Estimate

        11    1( 2.5%)      7    1( 2.5%)     (17.4%)   <<
```

รูปที่ 4.5 ตัวอย่างผลลัพธ์จากโปรแกรม c4.5

```
C4.5 [release 8] rule generator      Tue Feb 17 20:23:47 1998
-------------------------------

    Options:
   File stem <labor-neg>

Read 40 cases (16 attributes) from labor-neg

------------------
Processing tree 0

Final rules from tree 0:

Rule 6:
          wage increase first year > 2.5
          statutory holidays > 10
   ->  class good  [93.0%]

Rule 5:
          wage increase first year > 4
   ->  class good  [90.6%]

Rule 4:
          wage increase first year <= 4
          statutory holidays <= 10
   ->  class bad  [87.1%]

Rule 3:
          wage increase first year <= 2.5
          working hours > 38
   ->  class bad  [79.4%]

Default class: good


Evaluation on training data (40 items):

Rule  Size  Error  Used  Wrong                 Advantage
----  ----  -----  ----  -----                 ---------
   6     2   7.0%    19      0 (0.0%)          0 (0|0)  good
   5     1   9.4%     3      0 (0.0%)          0 (0|0)  good
   4     2  12.9%    10      0 (0.0%)          7 (7|0)  bad
   3     2  20.6%     3      0 (0.0%)          3 (3|0)  bad

Tested 40, errors 1 (2.5%)   <<


   (a)   (b)    <-classified as
  ----  ----
    26          (a): class good
     1    13    (b): class bad
```

รูปที่ 4.6 ตัวอย่างผลลัพธ์จากโปรแกรม c4.5rules

การพัฒนาโปรแกรม C4.5 บนระบบปฏิบัติการไมโครซอฟท์วินโดว์

การพัฒนาโปรแกรม C4.5 บนระบบปฏิบัติการไมโครซอฟท์วินโดว์ จะใช้โปรแกรมไมโครซอฟท์วิชวล C++ (Microsoft Visual C++) เวอร์ชั่น 4.2 ซึ่งเป็นตัวแปลภาษา C++ (C++ Compiler) ชนิด 32 บิต เพื่อให้มีประสิทธิภาพในการทำงานที่เร็วขึ้น โดยโปรแกรมที่ได้จะสามารถทำงานได้บนระบบปฏิบัติการไมโครซอฟท์วินโดว์ 95

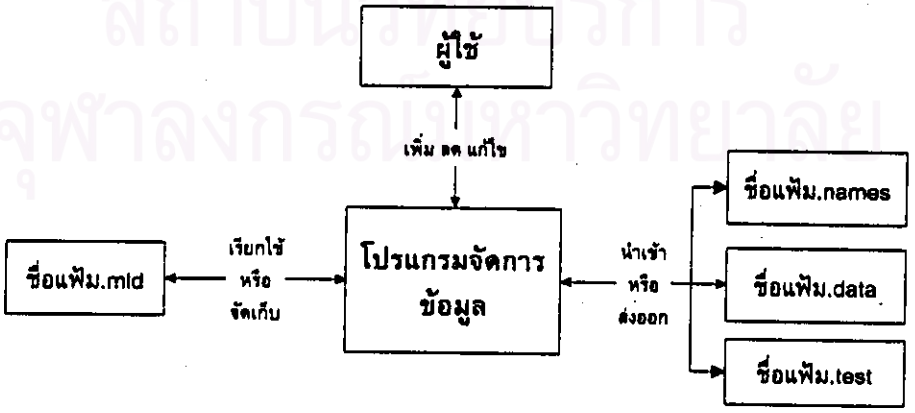
ในการพัฒนาได้แยกโปรแกรมออกเป็น 3 โปรแกรมด้วยกัน คือ

1. โปรแกรมจัดการข้อมูล
2. โปรแกรมสร้างต้นไม้ตัดสินใจ และการใช้งานต้นไม้ตัดสินใจ
3. โปรแกรมสร้างกฎการตัดสินใจ และการใช้งานกฎการตัดสินใจ

โปรแกรมจัดการข้อมูล

โปรแกรมนี้จะจัดการในส่วนของการเตรียมตัวอย่างเพื่อใช้ในการเรียนรู้และทดสอบ โดยมีความสามารถในการจัดการข้อมูลของลักษณะต่าง ๆ ที่จะใช้ในการเรียนรู้ สามารถที่จะเพิ่ม ลดหรือแก้ไข ชื่อ ชนิดของข้อมูลที่เก็บ และค่าที่เป็นไปได้ในแต่ละลักษณะ สามารถที่จะเพิ่ม ลดหรือแก้ไข ข้อมูลของตัวอย่างที่จะใช้สอนหรือทดสอบ สามารถที่จะเปิดแฟ้มชื่อ แฟ้มตัวอย่าง และแฟ้มตัวอย่างทดสอบของ C4.5 เดิมขึ้นมาแก้ไขเพิ่มเติม และเก็บลงแฟ้มเพื่อใช้ในการเรียนรู้และทดสอบได้ ซึ่งปกติในโปรแกรม C4.5 เดิมจะไม่มี ต้องใช้โปรแกรมบรรณาธิกรณ์ข้อความ (Text Editor) แก้ไขโดยตรงกับแฟ้มข้อมูล ซึ่งไม่สะดวกและอาจทำให้แฟ้มข้อมูลมีความผิดพลาดได้

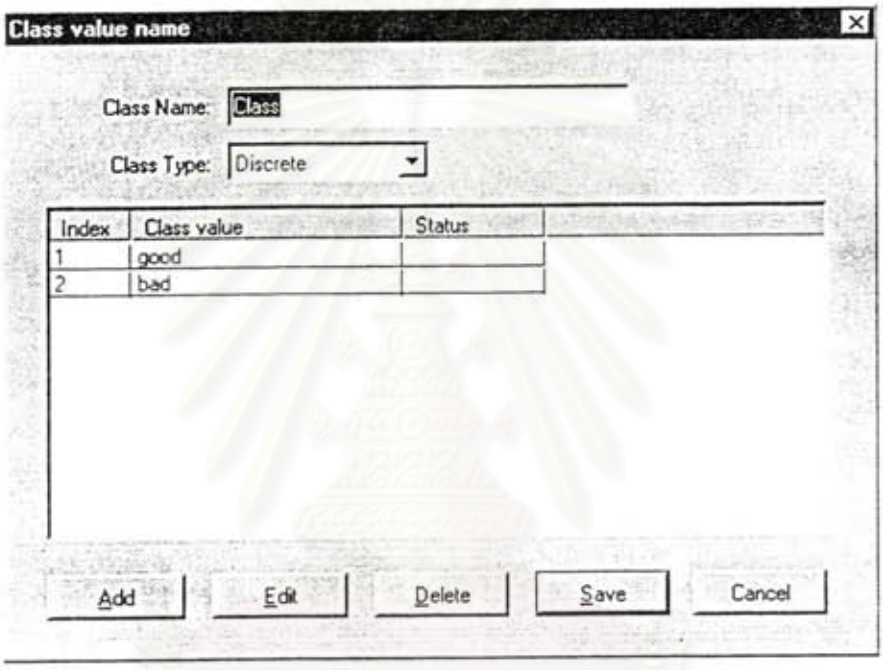
การทำงานของโปรแกรมนี้จะเกี่ยวข้องกับแฟ้มข้อมูล 4 แฟ้ม ซึ่ง 3 แฟ้มจะเป็นของ C4.5 เดิม คือ แฟ้มชื่อ (ชื่อแฟ้ม.names) แฟ้มตัวอย่างสอน (ชื่อแฟ้ม.data) แฟ้มตัวอย่างทดสอบ (ชื่อแฟ้ม.test) และที่เพิ่มมาใหม่อีกหนึ่งแฟ้ม คือ แฟ้มที่เก็บข้อมูลตัวอย่างของโปรแกรมจัดการข้อมูลนี้เอง ซึ่งจะเก็บทั้งส่วนของชื่อและข้อมูลตัวอย่างไว้ด้วยกัน โดยชื่อแฟ้มจะอยู่ในรูป "ชื่อแฟ้ม.mld" และแฟ้มต่าง ๆ จะมีการติดต่อกับโปรแกรม ดังแสดงในรูปที่ 4.7



รูปที่ 4.7 แสดงการติดต่อกับแฟ้มข้อมูลของโปรแกรมจัดการข้อมูล

ฟังก์ชั่นและความสามารถในการทำงานหลัก ๆ ของโปรแกรมจัดการข้อมูลมีดังนี้

1. การจัดการลักษณะแบ่งพวก ในตัวอย่างที่ใช้เรียนรู้หรือทดสอบจะต้องมีลักษณะนี้เพียงหนึ่งลักษณะและมีชนิดของข้อมูลเป็นค่าไม่ต่อเนื่องเท่านั้น เพื่อใช้ในการจำแนกจัดกลุ่มตัวอย่าง โปรแกรมนี้จะไม่ยอมให้มีการสร้างลักษณะนี้เพิ่ม หรือลบลักษณะนี้ออกจากแฟ้มข้อมูล แต่จะยอมให้มีการเพิ่ม ลดหรือแก้ไขเปลี่ยนแปลงค่าที่เป็นไปได้ของคลาสหรือพวกได้ตามต้องการ ดังรูปที่ 4.8

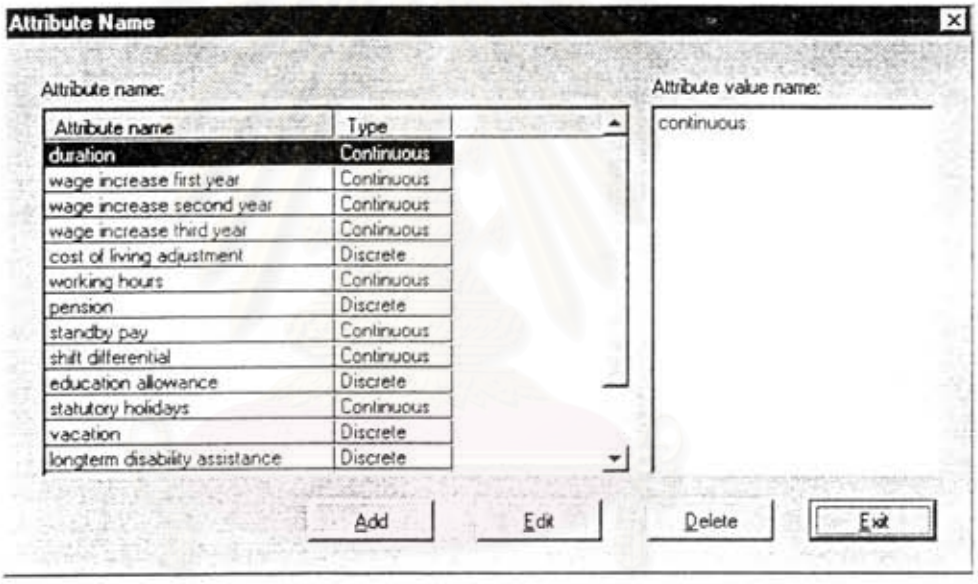


รูปที่ 4.8 แสดงจอภาพการจัดการข้อมูลของลักษณะแบ่งพวก

จากรูปที่ 4.8 ซึ่งเป็นการจัดการในส่วนของลักษณะแบ่งพวก จะประกอบด้วยชื่อของพวก (Class Name) ชนิดของข้อมูลที่เก็บ (Class Type) ซึ่งจะสามารถเลือกได้เพียงค่าไม่ต่อเนื่องเท่านั้น และค่าที่เป็นไปได้ในพวกนี้ (Class value name) ในการทำงานของโปรแกรมจะมีปุ่มคำสั่ง (Command Button) อยู่ 5 ปุ่มที่มีหน้าที่ต่าง ๆ กัน คือ ปุ่มเพิ่ม (Add) ใช้ในการเพิ่มค่าของพวก ใหม่ให้กับระบบ ปุ่มแก้ไข (Edit) ใช้ในการแก้ไขค่าของพวกที่มีอยู่แล้ว ปุ่มลบ (Delete) ใช้ในการลบค่าของพวกที่ไม่ต้องการ ปุ่มจัดเก็บ (Save) ใช้ในการจัดเก็บข้อมูลที่แก้ไขแล้วลงหน่วยความจำ แล้วออกจากจอภาพการจัดการลักษณะแบ่งพวก ส่วนปุ่มสุดท้าย คือ ปุ่มยกเลิก (Cancel) ใช้ในการออกจากจอภาพการแก้ไขลักษณะแบ่งพวก โดยไม่เก็บข้อมูลที่มีการเพิ่มหรือแก้ไขลงหน่วยความจำ ในส่วนของปุ่มเพิ่มและปุ่มแก้ไข เมื่อกดจะปรากฏหน้าต่างเพื่อรับค่าของลักษณะไม่แบ่งพวก (Class Value Name) ที่จะเพิ่มหรือแก้ไข ส่วนปุ่มลบเมื่อกดจะปรากฏคำว่า "Deleted" ในส่วนของสถานะ (Status) ของค่า

ของลักษณะที่มีแถบแสง และถ้ากดซ้ำอีกครั้งคำว่า "Deleted" ก็จะหายไป เพื่อเป็นการกำหนดว่าต้องการจะลบค่าไหนก่อนที่จะลบจริงเมื่อกดปุ่มจัดเก็บ

2. การจัดการลักษณะไม่แบ่งพวก ในตัวอย่างที่ใช้เรียนรู้หรือทดสอบจะมีลักษณะนี้ได้หลายลักษณะ โปรแกรมจะยอมให้มีการเพิ่มหรือลดจำนวนลักษณะ เปลี่ยนแปลงชื่อของแต่ละลักษณะที่ใช้ในการเรียนรู้ได้ แต่สามารถจะเลือกชนิดข้อมูลของแต่ละลักษณะว่าจะเป็นชนิดค่าต่อเนื่อง (Continuous Value) หรือค่าไม่ต่อเนื่อง (Discrete Value) ได้เฉพาะในตอนเพิ่มลักษณะใหม่เท่านั้น หลังจากนั้นจะไม่สามารถเปลี่ยนชนิดของข้อมูลได้อีก ถ้าต้องการเปลี่ยนจะต้องลบลักษณะนี้ออกก่อนแล้วค่อยเพิ่มเข้าไปใหม่ ในลักษณะที่มีชนิดของข้อมูลเป็นค่าไม่ต่อเนื่องจะต้องสร้างชุดของค่าที่เป็นไปได้ในแต่ละลักษณะ (Attribute value name) เพื่อใช้เป็นตัวเลือกในตอนใส่ค่าให้กับตัวอย่าง ซึ่งค่าที่เป็นไปได้นี้สามารถเพิ่ม ลด หรือแก้ไขได้ตลอดเวลา ดังรูปที่ 4.9

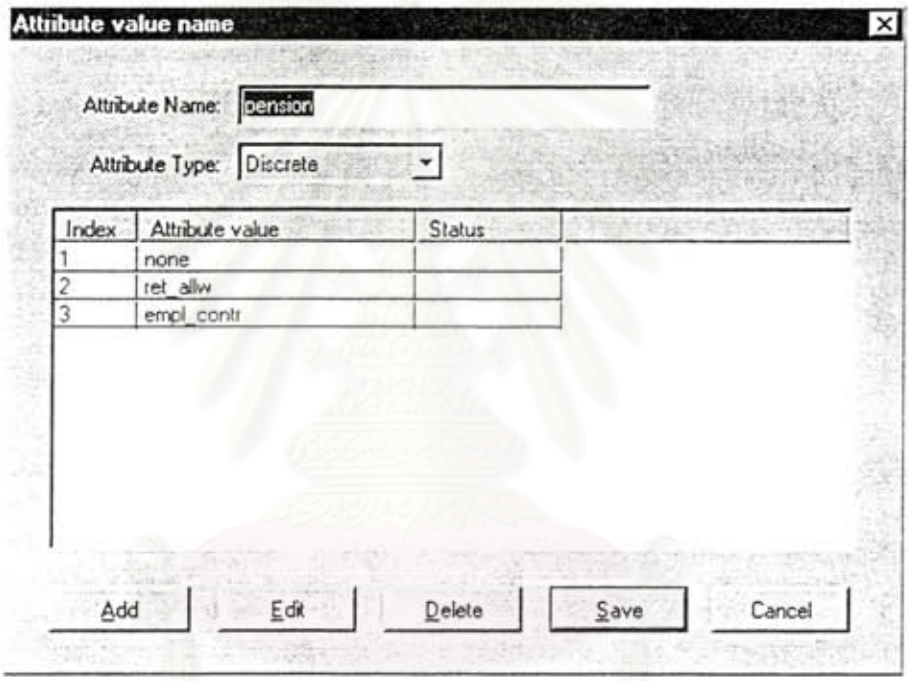


Attribute name	Type
duration	Continuous
wage increase first year	Continuous
wage increase second year	Continuous
wage increase third year	Continuous
cost of living adjustment	Discrete
working hours	Continuous
pension	Discrete
standby pay	Continuous
shift differential	Continuous
education allowance	Discrete
statutory holidays	Continuous
vacation	Discrete
longterm disability assistance	Discrete

รูปที่ 4.9 แสดงจอภาพการจัดการข้อมูลของลักษณะไม่แบ่งพวก

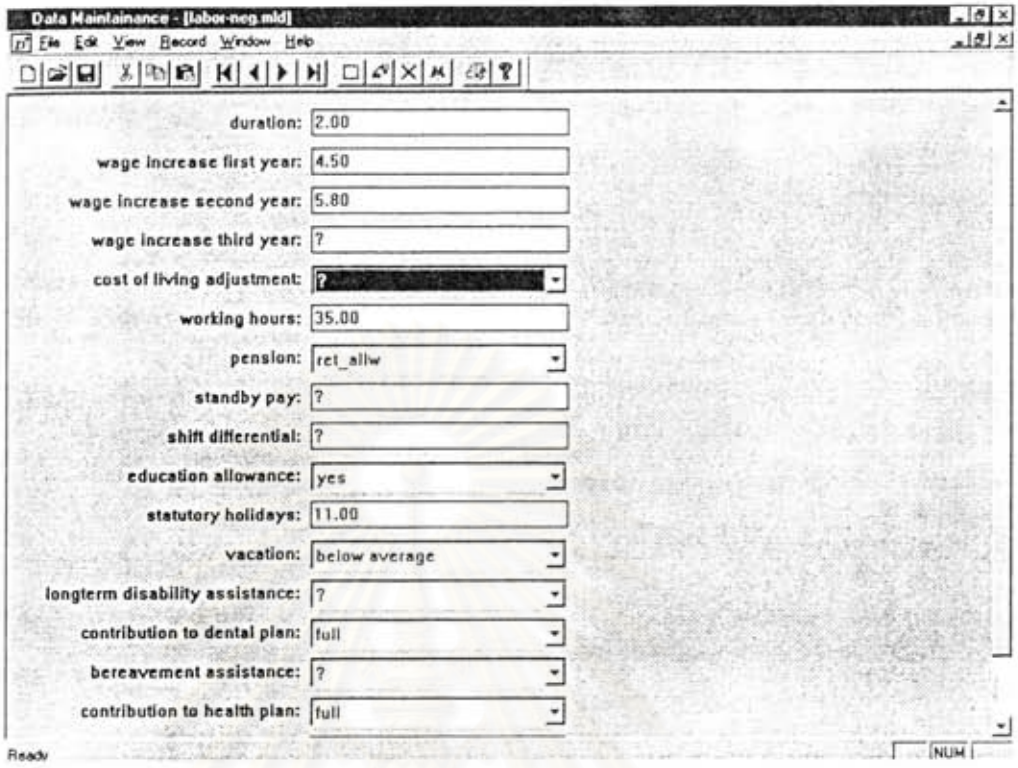
จากรูปที่ 4.9 ในจอภาพจะประกอบด้วย ชื่อของลักษณะไม่แบ่งพวก (Attribute Name) แต่ละลักษณะพร้อมทั้งชนิดของค่าที่เก็บ (Type) โดยด้านขวามือจะแสดงค่าที่เป็นไปได้ในแต่ละลักษณะ เมื่อผู้ใช้เลือกลักษณะใดลักษณะหนึ่งในตารางด้านซ้ายมือ ในการทำงานของโปรแกรมจะมีปุ่มอยู่ 4 ปุ่ม คือ ปุ่มเพิ่ม (Add) ใช้ในการเพิ่มลักษณะไม่แบ่งพวกลักษณะใหม่ให้กับระบบ เมื่อกดปุ่มนี้จะปรากฏหน้าต่างดังรูปที่ 4.10 เพื่อป้อนชื่อและชนิดของค่าที่เก็บ โดยถ้าเป็นค่าต่อเนื่องโปรแกรมจะไม่ยอมให้เพิ่มค่าที่เป็นไปได้ของลักษณะ แต่ถ้าเป็นค่าไม่ต่อเนื่องโปรแกรมจะยอมให้เพิ่มหรือแก้ไขค่าที่เป็นไปได้ของลักษณะนี้จนกว่าจะเก็บข้อมูลลงหน่วยความจำ หลังจากนั้นจะไม่สามารถเปลี่ยนแปลงชนิดของค่าที่เก็บได้ แต่จะสามารถแก้ไขเปลี่ยนแปลงในส่วนของค่าที่เป็นไปได้ของลักษณะนี้ ปุ่มแก้ไข (Edit) ใช้ใน

การแก้ไขค่าที่เป็นไปได้ในแต่ละลักษณะ เมื่อกดปุ่มนี้จะปรากฏหน้าต่างคล้ายกับหน้าต่างของการเพิ่มลักษณะ เพียงแต่ว่าไม่สามารถแก้ไขชนิดของข้อมูลที่เก็บได้ แต่จะสามารถเพิ่ม ลดหรือแก้ไขค่าได้ โดยถ้าเป็นการลบค่าบางค่าในลักษณะนี้ที่ถูกใช้อยู่ในตัวอย่าง จะทำให้ค่าของลักษณะนี้ในตัวอย่างเปลี่ยนเป็นลักษณะที่ไม่ทราบค่า แต่ถ้าเป็นการแก้ไขค่าของลักษณะก็จะทำให้ค่าในตัวอย่างที่ใช้อยู่เปลี่ยนแปลงตามค่าใหม่ที่แก้ไข ปุ่มลบ (Delete) ใช้ในการลบลักษณะที่ไม่ต้องการออกจากหน่วยความจำ เมื่อกดปุ่มนี้โปรแกรมจะถามยืนยันการลบ เพื่อป้องกันการผิดพลาดจากการลบโดยไม่ตั้งใจ เพราะเมื่อมีการลบแล้วจะไม่สามารถเรียกลักษณะที่ลบไปแล้วกลับคืนมาได้ ปุ่มออกจากโปรแกรม (Exit) ใช้ในการออกจากโปรแกรมการจัดการลักษณะไม่แบ่งพวก



รูปที่ 4.10 แสดงจอภาพการจัดการค่าที่เป็นไปได้ของลักษณะไม่แบ่งพวก

3. การจัดการข้อมูลตัวอย่าง โปรแกรมในส่วนนี้จะเป็นการเพิ่ม ลดหรือแก้ไขรายละเอียดของแต่ละตัวอย่างที่จะใช้ในการเรียนรู้หรือทดสอบ โดยในลักษณะที่เป็นค่าต่อเนื่องการแก้ไขจะกระทำในกล่องตัวอักษรที่ให้ผู้ใช้ใส่ข้อมูลตัวเลขได้เท่านั้น ส่วนในลักษณะที่เป็นค่าไม่ต่อเนื่องผู้ใช้ต้องเลือกจากชุดของค่าที่เป็นไปได้ภายในกล่องตัวเลือกเท่านั้น เพื่อเป็นการตรวจสอบข้อมูลให้สอดคล้องกัน ดังรูปที่ 4.11

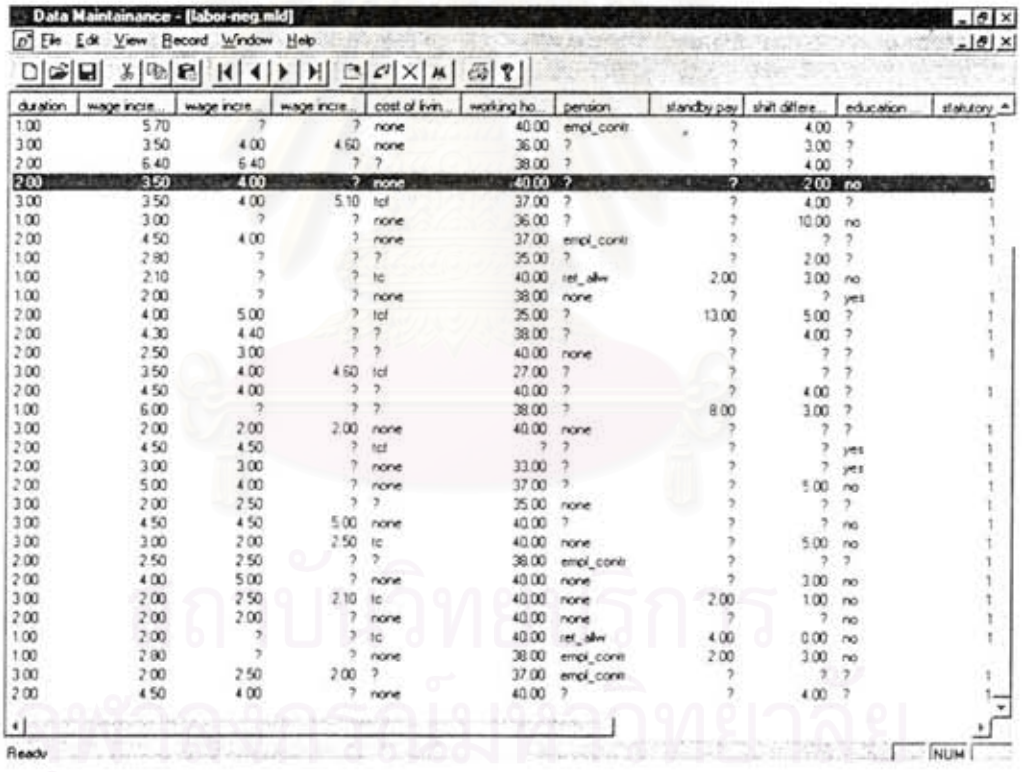


รูปที่ 4.11 แสดงจอภาพการจัดการข้อมูลตัวอย่าง

จากรูปที่ 4.11 การจัดการตัวอย่างจะประกอบด้วยส่วนของการป้อนข้อมูลและส่วนของปุ่มคำสั่งซึ่งจะอยู่บนแถบคำสั่ง (Tools Bar) ในส่วนของการป้อนข้อมูลจะประกอบด้วยส่วนรับข้อมูลที่เป็นค่าต่อเนื่องในรูปของกล่องตัวอักษร ที่จะยอมรับเฉพาะข้อมูลที่เป็นตัวเลขและเครื่องหมาย "?" ซึ่งแทนการไม่ทราบค่าของลักษณะนั้น ๆ อีกส่วนหนึ่งจะเป็นส่วนของการรับข้อมูลที่เป็นค่าไม่ต่อเนื่องอยู่ในรูปของกล่องตัวเลือกที่จะแสดงค่าที่เป็นไปได้ของลักษณะนี้พร้อมกับเครื่องหมาย "?" ที่แทนการไม่ทราบค่า เพื่อให้ผู้ใช้เลือก ในส่วนของปุ่มคำสั่งจะประกอบด้วยปุ่มนำทาง (Navigator) ซึ่งมีด้วยกัน 4 ปุ่ม คือ ปุ่ม |◀| ใช้ในการเคลื่อนที่ไปยังตัวอย่างแรก ปุ่ม |▶| ใช้ในการเคลื่อนที่ไปยังตัวอย่างถัดไป ปุ่ม |◀| ใช้ในการเคลื่อนที่ไปยังตัวอย่างก่อนหน้า และปุ่ม |▶| ใช้ในการ เคลื่อนที่ไปยังตัวอย่างสุดท้ายของข้อมูลได้ โดยถ้าข้อมูลของตัวอย่างมีการเปลี่ยนแปลงแล้วมีการเคลื่อนไปยังตัวอย่างอื่นจะมีการถามยืนยันการเก็บข้อมูลก่อนที่จะเลื่อนไปยังตัวอย่างที่ต้องการ และยังประกอบด้วยปุ่มคำสั่งในการจัดการข้อมูลได้แก่ ปุ่ม |□| ใช้ใช้ในการเพิ่มตัวอย่างใหม่ให้กับระบบ ปุ่ม |💾| ใช้ในการจัดเก็บตัวอย่างที่เพิ่มใหม่หรือแก้ไขลงสู่หน่วยความจำ โดยปกติปุ่มนี้จะเรียกใช้งานได้เมื่อมีการเปลี่ยนแปลงในส่วนของข้อมูลในแต่ละลักษณะของตัวอย่าง ปุ่ม |✕| ใช้ในการลบตัวอย่างที่ไม่ต้องการออกจากหน่วยความจำ โดยเมื่อกดปุ่มนี้จะมีการถามยืนยันการลบ เพื่อป้องกันการผิดพลาดจากการลบโดยไม่ได้ตั้งใจ

4. การแสดงผลของตัวอย่าง จะมีการแสดงผล 2 แบบ คือ แบบที่หนึ่งจะเป็นการแสดงผลทีตัวอย่าง (Record View) ดังรูปที่ 4.11 ซึ่งจะเป็นการจัดการข้อมูลตัวอย่างดังได้อธิบายไว้ใน

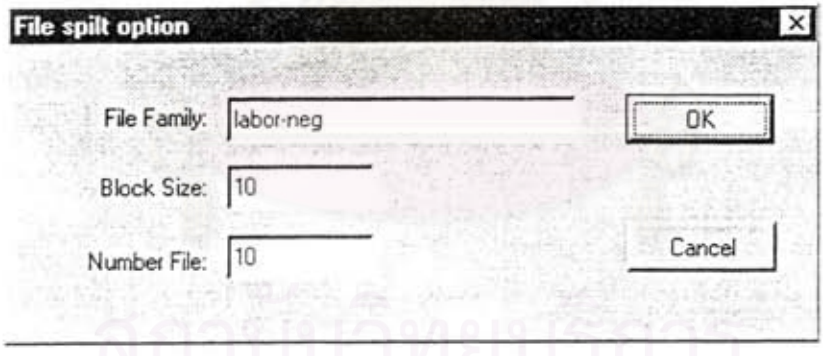
หัวข้อที่แล้ว และอีกแบบหนึ่งจะเป็นการแสดงผลทั้งหมด (List View) ดังรูปที่ 4.12 ซึ่งจะแสดงตัวอย่างทั้งหมดในหน่วยความจำ โดยแสดงผล 1 บรรทัดต่อ 1 ตัวอย่าง ในการผลแบบนี้ผู้ใช้ไม่สามารถที่จะแก้ไขเปลี่ยนแปลงข้อมูลของตัวอย่างได้ แต่จะสามารถดูข้อมูลเพื่อเปรียบเทียบกันระหว่างตัวอย่างได้ ผู้ใช้สามารถที่จะเคลื่อนที่ไปบนตัวอย่างโดยใช้ปุ่มบนแป้นพิมพ์ดังนี้ ปุ่มลูกศรลง (↓) ใช้ในการเคลื่อนที่ไปยังตัวอย่างถัดไปที่อยู่ด้านล่างของจอภาพ ปุ่มลูกศรขึ้น (↑) ใช้ในการเคลื่อนที่ไปยังตัวอย่างก่อนหน้าที่อยู่ด้านบนของจอภาพ ปุ่มลูกศรขวา (→) ใช้ในการเคลื่อนที่ไปทางด้านขวาเพื่อดูลักษณะที่ต้องการเมื่อมีลักษณะต่าง ๆ หลายลักษณะจนไม่สามารถแสดงในจอภาพเดียวกันได้ ปุ่มลูกศรซ้าย (←) ใช้ในการเคลื่อนที่ไปกลับมาดูลักษณะทางด้านซ้าย ปุ่มบ้าน (Home) ใช้ในการเคลื่อนที่ไปยังตัวอย่างแรกของข้อมูล ปุ่มจบ (End) ใช้ในการเคลื่อนที่ไปยังตัวอย่างสุดท้ายของข้อมูล ปุ่มหน้าบน (Page Up) ใช้ในการเลื่อนการแสดงผลขึ้นหนึ่งจอภาพ เพื่อดูตัวอย่างในส่วนบนของจอภาพ ปุ่มหน้าล่าง (Page Down) ใช้ในการเลื่อนการแสดงผลลงหนึ่งจอภาพ เพื่อดูตัวอย่างในส่วนล่างของจอภาพ



<u>รูปที่ 4.12</u> แสดงจอภาพการแสดงผลแบบทั้งหมดของโปรแกรมการจัดการข้อมูล

5. การเรียกใช้ (Open) จัดเก็บข้อมูล (Save) และการสร้างแฟ้มข้อมูลใหม่ (New) โปรแกรมจะจัดเก็บข้อมูลทั้งในส่วนของชื่อและข้อมูลของแต่ละตัวอย่างลงสู่แฟ้มข้อมูลแฟ้มเดียวกัน ซึ่งแฟ้มนี้จะมีส่วนขยายเป็น mld สามารถที่จะเรียกกลับขึ้นมาใช้งานและจัดเก็บลงแฟ้มข้อมูลได้ตามต้องการโดยผ่านเมนู (Menu) หรือปุ่มบนแถบเครื่องมือ

6. การนำข้อมูลเข้า (Import) และส่งข้อมูลออก (Export) เป็นการจัดเก็บหรือเรียกใช้แฟ้มข้อมูลตัวอย่างสอนหรือตัวอย่างที่ใช้ทดสอบในรูปแบบของแฟ้มตัวอักษรของ C4.5 เดิม ซึ่งจะประกอบด้วยแฟ้มข้อมูล 3 แฟ้มดังที่ได้กล่าวมาแล้ว คือ แฟ้มชื่อ (ชื่อแฟ้ม.names), แฟ้มตัวอย่างสอน (ชื่อแฟ้ม.data) และ แฟ้มตัวอย่างทดสอบ (ชื่อแฟ้ม.test) ในการนำข้อมูลเข้าต้องการแฟ้มข้อมูล 2 แฟ้มคือ แฟ้มชื่อกับแฟ้มตัวอย่างสอน หรือแฟ้มชื่อกับแฟ้มตัวอย่างทดสอบอย่างใดอย่างหนึ่ง ส่วนในการส่งข้อมูลออกจะสามารถกระทำได้ 2 แบบ คือ แบบที่หนึ่งจะเป็นการส่งข้อมูลทั้งหมดออกสู่แฟ้มชื่อกับแฟ้มตัวอย่างสอน หรือแฟ้มชื่อกับแฟ้มตัวอย่างทดสอบแล้วแต่ผู้ใช้เลือกแบบไหน แบบที่สองจะเป็นการส่งข้อมูลแบบเป็นชุด คือ ใน 1 ชุดจะประกอบด้วยแฟ้มชื่อ แฟ้มตัวอย่างสอน และแฟ้มตัวอย่างทดสอบ ที่เกิดจากการแบ่งตัวอย่างทั้งหมดออกเป็น n ชุด ซึ่งมีจำนวนตัวอย่างและการกระจายตัวของลักษณะแบ่งพวกเท่ากัน แล้วนำมาสร้างเป็นแฟ้มตัวอย่างสอนโดยการรวมตัวอย่าง n-1 ชุดเข้าด้วยกัน ส่วนอีก 1 ชุดที่เหลือนำไปสร้างเป็นแฟ้มข้อมูลทดสอบ การสร้างแฟ้มจะสลับชุดกันไป โดยแฟ้มตัวอย่างทดสอบที่ได้จะไม่มีตัวอย่างซ้ำกันเลยในแต่ละชุด ซึ่งสามารถจะสร้างชุดของแฟ้มข้อมูลได้มากที่สุด n ชุด เมื่อมีการเรียกโปรแกรมนี้ขึ้นมาทำงานจะปรากฏหน้าต่างดังรูปที่ 4.13 อันประกอบด้วยส่วนของชื่อแฟ้มที่จะเป็นพื้นฐานในการตั้งชื่อแฟ้มในแต่ละชุด (File Family) จำนวนชุดที่จะแบ่ง (Block Size) หรือค่า n และจำนวนชุดของแฟ้มที่จะจัดเก็บ (Number File) โดยชื่อแฟ้มที่จะตั้งให้กับแต่ละชุดจะใช้ชื่อแฟ้มพื้นฐานแล้วตามด้วยหมายเลขลำดับในการสร้างแฟ้มข้อมูล ซึ่งมีค่าเริ่มต้นตั้งแต่ 0 จนถึงจำนวนชุดลบหนึ่ง



รูปที่ 4.13 แสดงจอภาพการตั้งค่าตัวเลือกในการส่งข้อมูลออก

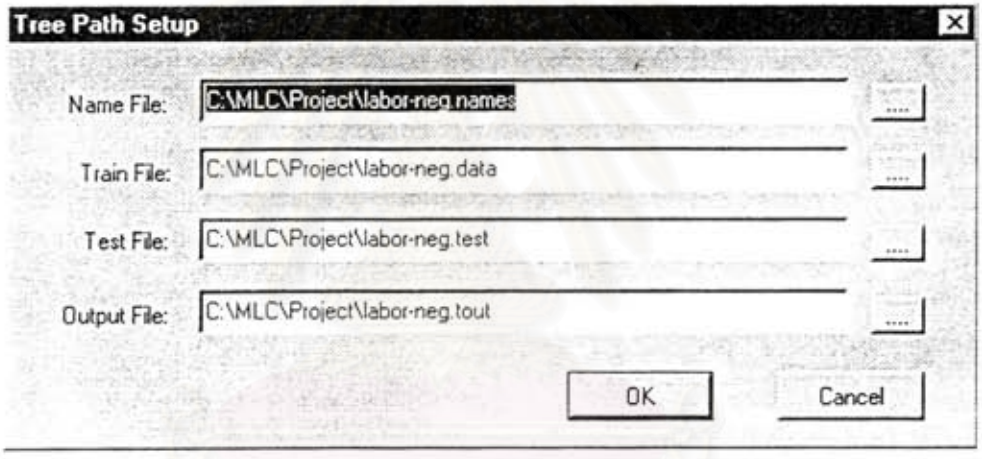
โปรแกรมสร้างต้นไม้ตัดสินใจและการใช้งานต้นไม้ตัดสินใจ

โปรแกรมนี้จะเป็นการเรียนรู้จากข้อมูลตัวอย่างที่ได้จากโปรแกรมแรก เพื่อสร้างเป็นต้นไม้ตัดสินใจ โดยจะเป็นการนำโปรแกรมเดิมซึ่งอยู่ในรูปภาษา C มาดัดแปลงให้อยู่ในรูปภาษา C++ ที่สามารถทำงานได้บนระบบปฏิบัติการวินโดว์ และเพิ่มเติมในส่วนของการติดต่อกับผู้ใช้และแสดงผลให้อยู่ในรูปกราฟฟิก จากเดิมที่แสดงอยู่ในรูปแบบของตัวอักษร

การดัดแปลงจะเริ่มจากการนำโปรแกรม c4.5 และ consult บนภาษา C เดิม มาสร้างเป็นคลาสใหม่บนภาษา C++ ซึ่งมีฟังก์ชั่นการทำงานเหมือนกับโปรแกรมเดิม โดยได้มีเปลี่ยนแปลงในส่วนของการตั้งค่าตัวเลือกที่ใช้ควบคุมโปรแกรมในการเรียนรู้จากเดิมที่เป็นคำสั่งแบบรายบรรทัด (Command Line) ให้เป็นการใส่ค่าผ่านกล่องตัวอักษร (Text Box) หรือเลือกค่าจากกล่องตัวเลือก (Combo Box) ซึ่งสามารถจะตั้งค่าตัวเลือกและเรียนรู้กี่ครั้งก็ได้ และได้เพิ่มส่วนการแสดงผลของต้นไม้ตัดสินใจก่อนและหลังตัดแต่งให้อยู่ในรูปการแสดงผลแบบต้นไม้ (Tree View) ซึ่งสามารถจะขยายหรือลดรูปต้นไม้เพื่อดูบางส่วนได้

<u>การทำงานของโปรแกรมนี้จะประกอบด้วยฟังก์ชั่นหลัก ๆ คือ</u>

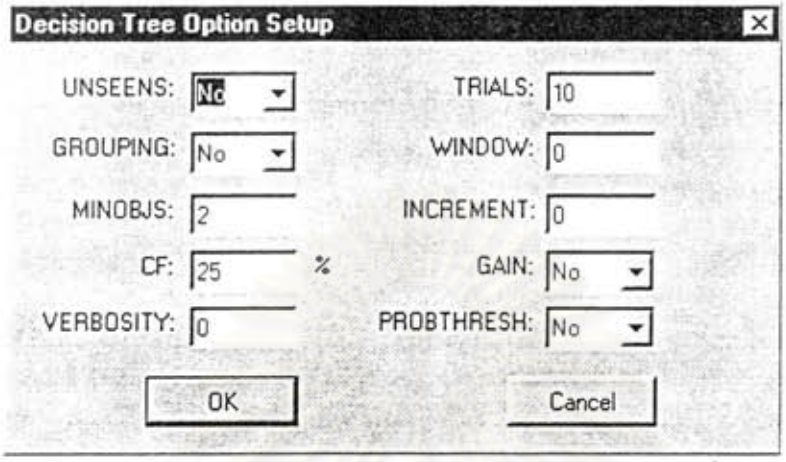
1. การตั้งค่าตัวเลือกในการเรียนรู้ เป็นการกำหนดค่าต่าง ๆ ที่ใช้ในการควบคุมโปรแกรม เพื่อใช้ในการเรียนรู้ โดยค่าที่สามารถจะกำหนดได้มีดังรูปที่ 4.14 และรูปที่ 4.15 ซึ่งจะประกอบด้วย



<u>รูปที่ 4.14</u> แสดงจอภาพการตั้งค่าแฟ้มข้อมูลที่ใช้ในการสร้างต้นไม้ตัดสินใจ

- NAME FILE เป็นชื่อแฟ้มของ แฟ้มชื่อ ของตัวอย่างที่จะใช้ในการเรียนรู้
- TRAIN FILE เป็นชื่อแฟ้มของ แฟ้มตัวอย่างสอน
- TEST FILE เป็นชื่อแฟ้มของ แฟ้มตัวอย่างทดสอบ
- OUTPUT FILE เป็นชื่อแฟ้มที่เก็บผลของการเรียนรู้ ซึ่งมีส่วนขยายเป็น "tout"
- UNSEENS เป็นตัวเลือกที่บอกให้โปรแกรมทราบว่า จะให้โปรแกรมทดสอบต้นไม้ตัดสินใจที่ได้กับชุดตัวอย่างทดสอบหรือไม่ เปรียบเทียบได้กับตัวเลือก -u ในโปรแกรมเดิม ซึ่งค่าโดยปริยายจะเป็น No คือ ไม่ต้องทดสอบต้นไม้ตัดสินใจที่ได้บนตัวอย่างทดสอบ
- GROUPING เป็นตัวเลือกที่บอกว่าในลักษณะที่มีค่าเป็นชนิดไม่ต่อเนื่อง ค่าที่เป็นไปได้จะถูกรวมกันเป็นชุดเพื่อทดสอบและสร้างต้นไม้ตัดสินใจหรือไม่ ค่าโดยปริยายจะเป็น

No คือไม่มีการรวมกัน โดยแต่ละค่าที่เป็นไปได้ในลักษณะนี้จะแยกกิ่งกันออกไปไม่รวมเป็นกิ่งเดียวกัน ซึ่งเปรียบได้กับ -s ใน c4.5 เดิม



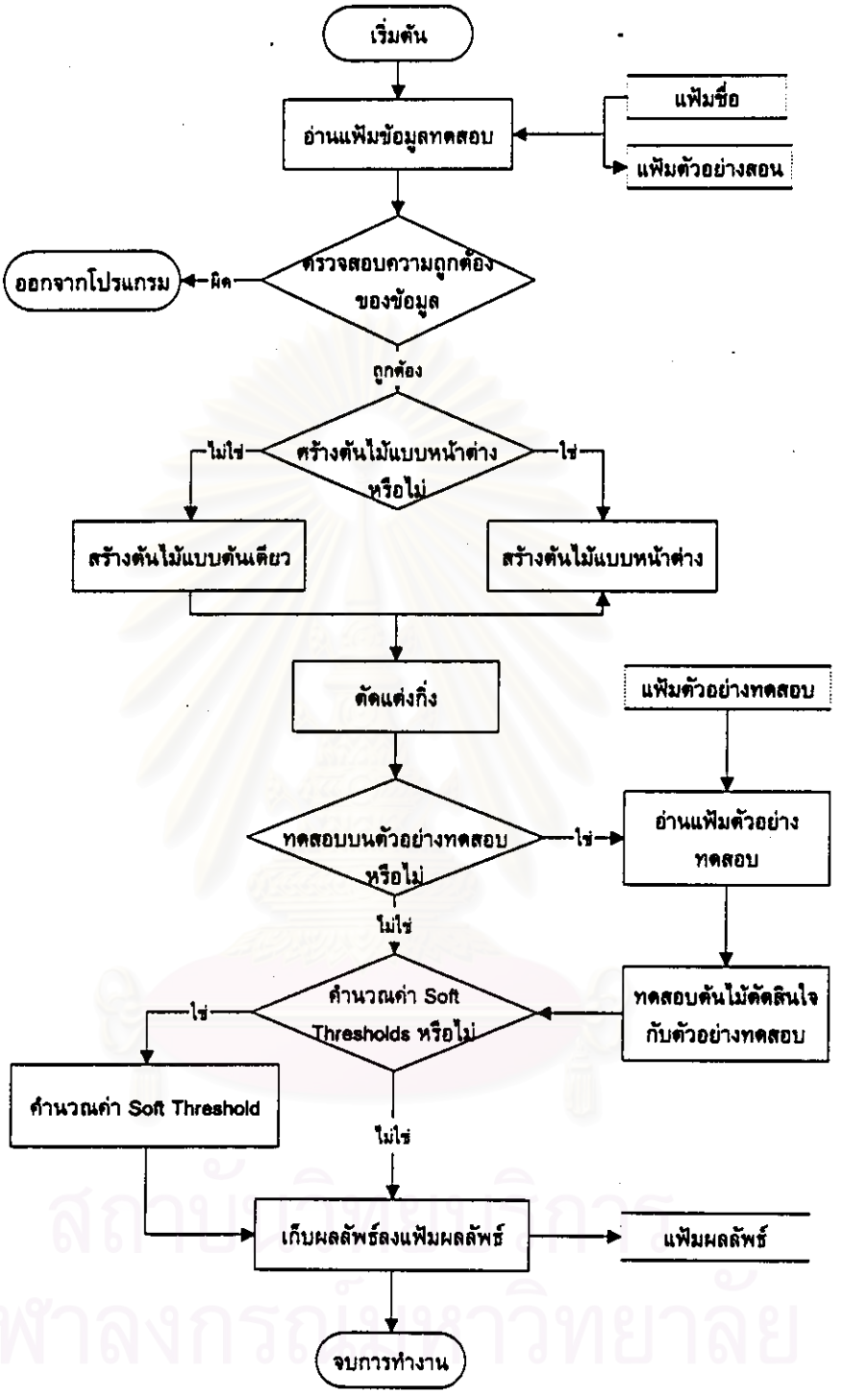
รูปที่ 4.15 แสดงจอภาพการตั้งค่าตัวเลือกที่ใช้ในการสร้างต้นไม้ตัดสินใจ

- MINOBJS เป็นค่าที่กำหนดว่าในแต่ละใบของต้นไม้ตัดสินใจ จะต้องมีจำนวนตัวอย่างที่เป็นพวกเดียวกันอย่างน้อยเท่ากับค่าที่กำหนดในตัวเลือกนี้ โดยปกติจะมีค่าเป็น 2 เปรียบได้กับตัวเลือก -m ใน c4.5 เดิม
- CF เป็นค่าที่มีผลต่อการตัดแต่งกิ่ง โดยค่าที่น้อยจะทำให้มีการตัดแต่งกิ่งมากกว่าค่าที่มาก ค่านี้จะอยู่ระหว่าง 0 ถึง 100 % ซึ่งค่าโดยปริยายจะเป็น 25 % เปรียบเทียบได้กับตัวเลือก -c ใน c4.5 เดิม
- VERBOSITY เป็นค่าที่อยู่ระหว่าง 0 ถึง 5 ซึ่งจะบอกถึงระดับการแสดงผลลัพธ์ของการเรียนรู้ โดยค่า 0 ซึ่งเป็นค่าโดยปริยายจะแสดงผลลัพธ์น้อยที่สุด ส่วนค่าที่มากขึ้นผลลัพธ์หรือค่าจากการคำนวณจะแสดงออกมามากขึ้นตามลำดับ เปรียบเทียบได้กับตัวเลือก -v ใน c4.5 เดิม
- TRIALS เป็นการเรียกใช้การสร้างต้นไม้แบบหน้าต่าง (Windowing) โดยกำหนดว่าจะสร้างต้นไม้เท่ากับจำนวนที่ป้อนให้ในตัวเลือกนี้ ก่อนที่จะเลือกต้นไม้ที่ดีที่สุดเพียงต้นเดียวมาใช้งาน ค่าโดยปริยายจะเป็น 10 แต่เมื่อค่านี้เป็น 0 จะหมายถึงว่าไม่ใช้การสร้างต้นไม้แบบหน้าต่าง แต่จะสร้างต้นไม้เพียงต้นเดียว เปรียบได้กับตัวเลือก -t ใน c4.5 เดิม
- WINDOW เป็นการเรียกใช้การสร้างต้นไม้แบบหน้าต่าง โดยกำหนดจำนวนตัวอย่างที่จะใช้เริ่มต้นในแต่ละหน้าต่าง ค่าโดยปริยายจะเป็นค่าที่สูงที่สุดระหว่างค่า 20% ของจำนวนตัวอย่างกับค่าสองเท่าของการถอดรากที่สอง (Square Root) ของจำนวนตัวอย่างสอนทั้งหมด ถ้าค่านี้เป็น 0 จะไม่ใช้การสร้างต้นไม้แบบหน้าต่าง แต่จะสร้างต้นไม้เพียงต้นเดียว เปรียบได้กับตัวเลือก -w ของ c4.5 เดิม

- INCREMENT เป็นการเรียกใช้การสร้างต้นไม้แบบหน้าต่าง โดยการกำหนดจำนวนตัวอย่างสูงที่สุดที่สามารถเพิ่มเข้าไปในหน้าต่างของแต่ละรอบการทดลองของวิธีการสร้างต้นไม้แบบหน้าต่าง ค่าโดยปริยายจะมีค่าเป็น 20% ของจำนวนตัวอย่างที่กำหนดให้ในหน้าต่างแรก ถ้าค่านี้เป็น 0 จะไม่ใช้การสร้างต้นไม้แบบหน้าต่าง เปรียบเทียบได้กับตัวเลือก -i ใน c4.5 เดิม
- GAIN เป็นตัวเลือกที่จะระบุให้โปรแกรมใช้ค่ามาตรฐานเกนหรือค่าอัตราส่วนมาตรฐานเกนในการสร้างต้นไม้ตัดสินใจ ค่าโดยปริยายจะเป็น No คือไม่ใช้ค่ามาตรฐานเกน แต่ใช้ค่าอัตราส่วนมาตรฐานเกน เปรียบได้กับตัวเลือก -g ของ c4.5 เดิม
- PROBTHRESH เป็นตัวเลือกว่าจะใช้จุดแบ่ง Z+ และ Z- แบบแข็ง (Hard Thresholds) หรือแบบอ่อน (Soft Thresholds) ซึ่งมีผลในการแปลค่าการตัดสินใจของต้นไม้ที่ได้ เมื่อใช้ตัดสินใจกับตัวอย่างใหม่ โดยปกติค่าจะเป็น No คือ ใช้จุดแบ่งแบบแข็ง เปรียบได้กับตัวเลือก -p ของ c4.5 เดิม

c4.5 เดิม	โปรแกรมสร้างต้นไม้ตัดสินใจ	ความหมายของตัวเลือก
-f ชื่อแฟ้ม	ใช้การตั้งค่าแฟ้มข้อมูล • NAME FILE • TRAIN FILE • TEST FILE • OUTPUT FILE	ชื่อแฟ้มข้อมูลในการเรียนรู้ เพื่อสร้างต้นไม้ตัดสินใจ
-u	UNSEENS	ทดสอบบนตัวอย่างทดสอบ
-s	GROUPING	การรวมกลุ่มของลักษณะที่มีค่าไม่ต่อเนื่อง
-m จำนวนตัวอย่าง	MINOBJS	จำนวนตัวอย่างน้อยที่สุดที่เป็นพวกเดียวกันที่ใบ
-c ค่า CF	CF	ระดับการตัดแต่งกิ่ง
-v ระดับ	VERBOSITY	ระดับของผลลัพธ์
-t จำนวนต้นไม้	TRIALS	จำนวนต้นไม้ที่ใช้ในการสร้างแบบหน้าต่าง
-w จำนวนตัวอย่าง	WINDOW	จำนวนตัวอย่างเริ่มต้นในแต่ละหน้าต่าง
-I จำนวนตัวอย่าง	INCREMENT	จำนวนตัวอย่างมากที่สุดที่จะเพิ่มในแต่ละรอบของหน้าต่าง
-g	GAIN	การเรียนรู้แบบใช้ค่ามาตรฐานเกน
-p	PROBTHRESH	การเลือกใช้จุดแบ่งแบบอ่อน

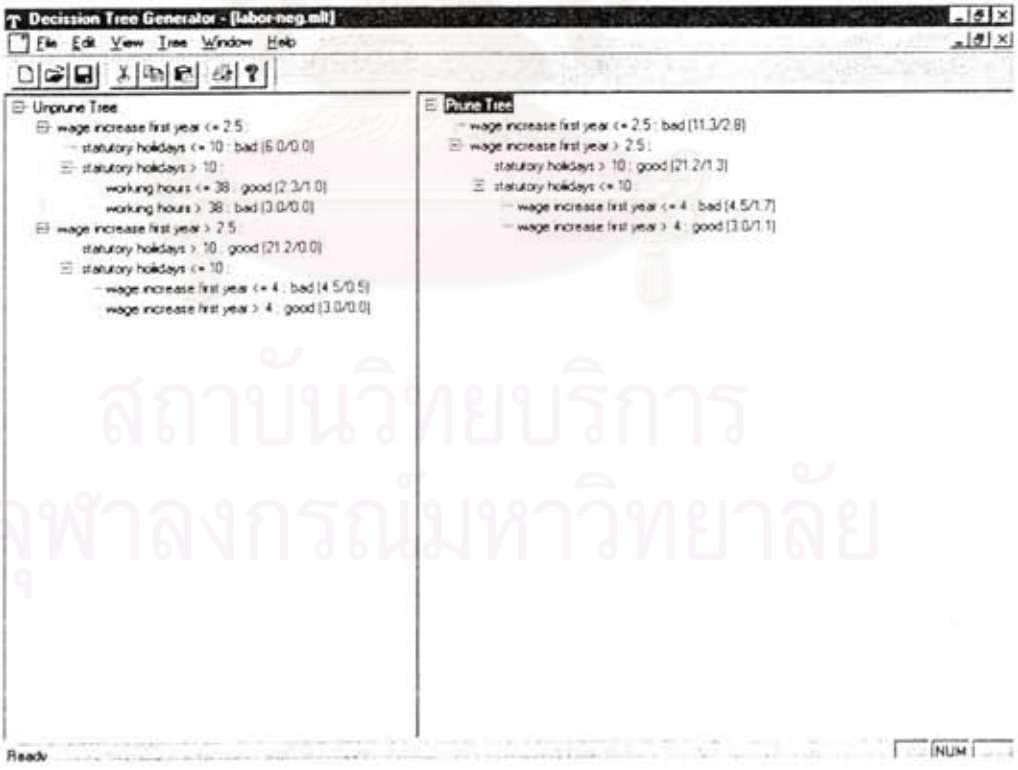
ตารางที่ 4.1 แสดงตัวเลือกของโปรแกรม c4.5 เดิมเปรียบเทียบกับโปรแกรมการสร้างต้นไม้ตัดสินใจ



<u>รูปที่ 4.16</u> แสดงขั้นตอนการทำงานในการสร้างต้นไม้ตัดสินใจ

2. การเรียนรู้และสร้างต้นไม้ตัดสินใจ (Learning) การทำงานจะเริ่มจากการอ่านข้อมูลตัวอย่างที่ใช้ในการสอน ซึ่งประกอบด้วยแฟ้มชื่อ และแฟ้มตัวอย่างสอน จากนั้นจะตรวจสอบความถูกต้องของข้อมูลที่อ่าน ถ้าไม่ถูกต้องจะแจ้งความผิดพลาดให้ผู้ใช้ทราบแล้วหยุดการทำงาน

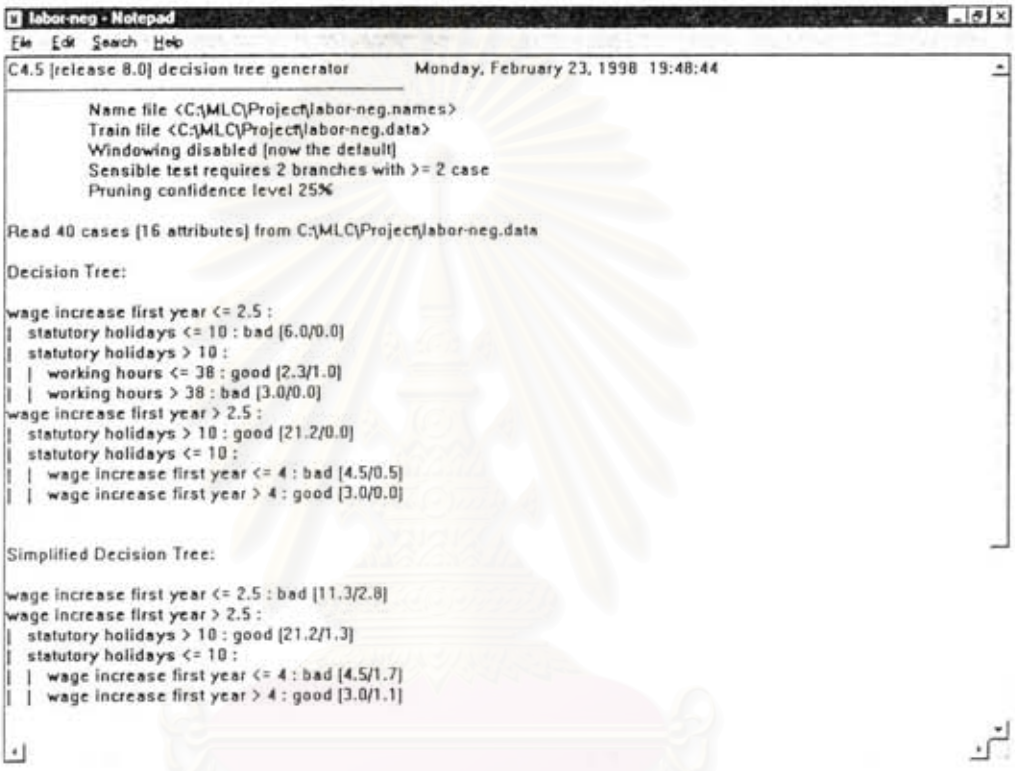
แต่ถ้าข้อมูลที่อ่านได้ถูกต้องก็จะเริ่มต้นกระบวนการเรียนรู้ตามค่าตัวเลือกที่ได้ตั้งไว้ให้ โดยในขั้นแรกโปรแกรมจะตัดสินใจว่าจะใช้การเรียนรู้แบบหน้าต่างหรือไม่ จากตัวเลือก TRIALS, WINDOW และ INCREMENT ถ้าตัวเลือกใดตัวเลือกหนึ่งมีค่าไม่เท่ากับ 0 ก็จะใช้การสร้างต้นไม้แบบหน้าต่างโดยใช้ค่าตามที่ป้อนให้หรือค่าโดยปริยาย แต่ถ้าตัวเลือกทั้งหมดเท่ากับ 0 โปรแกรมจะใช้การสร้างต้นไม้ที่ดีที่สุดต้นเดียว ในขั้นตอนของการสร้างต้นไม้ตัดสินใจก่อนการตัดแต่งกิ่ง จะคำนึงถึงตัวเลือก MINOBJS, GROUPING และ GAIN ตามที่ผู้ใช้ได้กำหนดไว้ให้ หลังจากนั้นจะทำการตัดแต่งต้นไม้ตัดสินใจตามค่าตัวเลือก CF ที่กำหนด ได้เป็นต้นไม้ที่ตัดแต่งกิ่งเรียบร้อยแล้ว จากนั้นโปรแกรมจะตรวจสอบว่าจะต้องทดสอบต้นไม้หลังจากตัดแต่งแล้วบนตัวอย่างทดสอบหรือไม่จากตัวเลือก UNSEENS ถ้าตัวเลือกเป็น Yes โปรแกรมจะทำการอ่านแฟ้มข้อมูลทดสอบ ตรวจสอบความถูกต้อง และนำไปทดสอบกับต้นไม้ตัดสินใจหลังจากตัดแต่งแล้ว แต่ถ้าตัวเลือกนี้เป็น No ก็จะไม่มีการทดสอบบนตัวอย่างทดสอบ หลังจากนั้นโปรแกรมจะตรวจสอบว่ามีการใช้จุดแบ่งแบบใดจากตัวเลือก PROBTHRESH ถ้าตัวเลือกเป็น Yes ก็จะมีการคำนวณค่าจุดแบ่งแบบอ่อน (Soft Thresholds) ให้กับต้นไม้ตัดสินใจหลังจากตัดแต่ง แต่ถ้าตัวเลือกเป็น No ก็จะไม่มีการคำนวณจุดแบ่งแบบอ่อน เมื่อเสร็จสิ้นกระบวนการเรียนรู้ก็จะมีการแสดงต้นไม้ทั้งก่อนการตัดแต่งและหลังจากการตัดแต่งออกทางจอภาพ และมีการเก็บผลของการเรียนรู้ตามระดับที่กำหนดไว้ในตัวเลือก VERBOSITY ลงสู่แฟ้มผลลัพธ์ ดังได้สรุปไว้ในรูปที่ 4.16



รูปที่ 4.17 แสดงต้นไม้ตัดสินใจก่อนและหลังการตัดแต่ง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

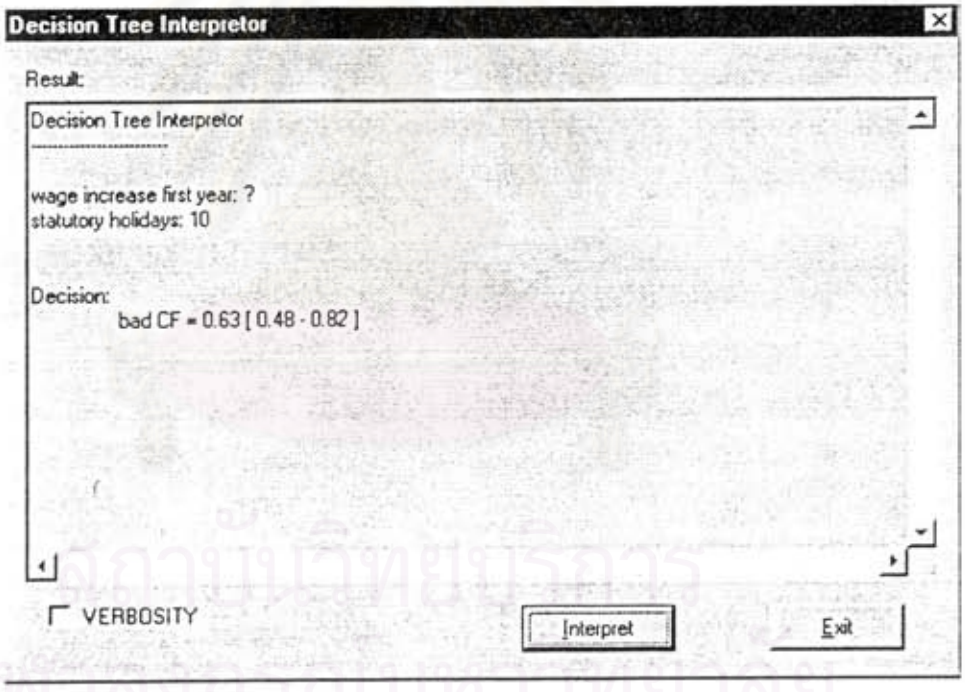
3. การแสดงผลลัพธ์ (View Output) หลังจากการเรียนรู้เสร็จสิ้น โปรแกรมก็จะแสดงต้นไม้ตัดสินใจก่อนการตัดแต่ง (Unpruned Tree) และต้นไม้ตัดสินใจหลังจากตัดแต่งกิ่ง (Pruned Tree) ออกทางจอภาพดังรูปที่ 4.17 และเก็บผลลัพธ์ลงแฟ้มผลลัพธ์ (ชื่อแฟ้ม.tout) ซึ่งอยู่ในรูปของแฟ้มตัวอักษร การนำผลลัพธ์ขึ้นมาแสดงทางจอภาพจะใช้โปรแกรม Notepad ที่มีอยู่ในระบบปฏิบัติการไมโครซอฟท์วินโดว์ 95 ช่วยในการแสดงผลหลังจากที่เลือกเมนูแสดงผลลัพธ์ ดังรูปที่ 4.18



<u>รูปที่ 4.18</u> แสดงผลลัพธ์ของการสร้างต้นไม้ตัดสินใจโดยใช้ Notepad

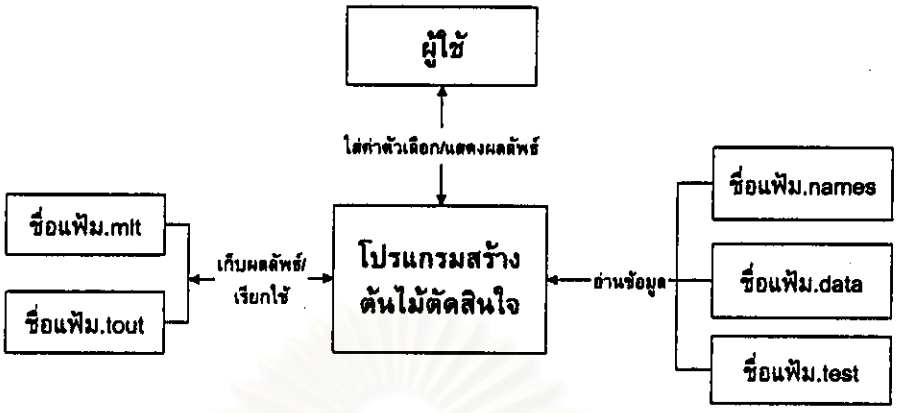
4. การใช้งานต้นไม้ตัดสินใจ (Interpret) จะเป็นการนำต้นไม้ตัดสินใจหลังจากตัดแต่งกิ่งแล้วไปใช้งานในการตัดสินใจกับตัวอย่างที่ผู้ใช้ป้อนให้ โดยเริ่มต้นจากจอภาพในรูปที่ 4.19 ซึ่งประกอบด้วยส่วนของการแสดงผลลัพธ์จากการตัดสินใจ (Result) ที่จะแสดงผลการตัดสินใจหลังจากผู้ใช้ป้อนข้อมูลลักษณะต่าง ๆ ของตัวอย่างที่โปรแกรมถาม ผลการตัดสินใจจะอยู่ในรูปของค่าความน่าจะเป็น เช่น 0.99 [0.97 – 1.00] จะหมายถึงว่าจะตัดสินใจให้เป็นพวกนี้ด้วยความเชื่อมั่น 99% จากช่วงของความเชื่อมั่น 97% ถึง 100% ส่วนของตัวเลือกระดับการแสดงผลลัพธ์ (VERBOSITY) จะใช้กำหนดให้โปรแกรมแสดงรายละเอียดของผลการตัดสินใจมากขึ้น ปุ่มเริ่มต้นการตัดสินใจ (Interpret) จะใช้เริ่มต้นคำถามที่โปรแกรมจะถามผู้ใช้และให้ผู้ใช้ใส่ค่าในแต่ละลักษณะที่ถาม โดยเริ่มจากรากของต้นไม้ตัดสินใจหลังจากตัดแต่งแล้วไล่ลงไปยังกิ่งต่าง ๆ จนกระทั่งถึงใบซึ่งเป็นผลของการตัดสินใจ ในการถามของโปรแกรมจะมีด้วยกันสองแบบ คือ การถามบนลักษณะที่เป็นค่าต่อเนื่องซึ่งจะยอมให้ผู้ใช้

ป้อนข้อมูลเฉพาะที่เป็นตัวเลข หรือช่วงของจำนวนโดยใช้เครื่องหมาย "," เป็นตัวคั่นระหว่างช่วงเมื่อผู้ใช้ไม่ทราบค่าที่แน่นอนหรือต้องการทราบผลการตัดสินใจที่เป็นช่วง และเครื่องหมาย "?" เมื่อลักษณะนั้นไม่ทราบค่า ส่วนในการถามบนลักษณะที่เป็นค่าไม่ต่อเนื่องโปรแกรมจะให้ผู้ใช้เลือกค่าที่เป็นไปได้ของลักษณะนั้นจากตัวเลือกที่มีอยู่ในกล่องตัวเลือก หรือบางครั้งผู้ใช้ไม่แน่ใจว่าค่าที่จะป้อนให้ถูก 100% ผู้ใช้ก็สามารถที่จะกำหนดความน่าจะเป็นให้กับค่าที่ป้อนให้ได้ โดยใช้เครื่องหมาย ":" แล้วตามด้วยค่าความน่าจะเป็นซึ่งต้องน้อยกว่าหรือเท่ากับ 1 หรือในบางครั้งผู้ใช้ไม่ทราบค่าที่แน่นอนแต่อาจจะเกิดจากหลายค่าที่เป็นไปได้รวมกัน ผู้ใช้ก็สามารถจะป้อนค่าที่เป็นไปได้หลาย ๆ ค่ารวมกัน โดยมีเครื่องหมาย "," คั่นระหว่างค่า และในแต่ละค่าที่ป้อนจะต้องมีค่าความน่าจะเป็นโดยใช้เครื่องหมาย ":" คั่นระหว่างค่าที่ป้อนกับค่าความน่าจะเป็น แต่ทั้งนี้ผลรวมของความน่าจะเป็นทั้งหมดรวมกันจะต้องเท่ากับ 1 ส่วนปุ่มสุดท้ายก็คือปุ่มออกจากการใช้งานต้นไม้ตัดสินใจ (Exit) ซึ่งจะใช้ออกจากโปรแกรมนี้เมื่อต้องการ และในระหว่างการถามผู้ใช้ก็สามารถยกเลิกการถามได้ทันทีที่กดปุ่มยกเลิก (Cancel) ในแต่ละหน้าต่างคำถาม แล้วกลับมาสู่การเริ่มต้นคำถามใหม่ได้



รูปที่ 4.19 แสดงจอภาพการใช้งานต้นไม้ตัดสินใจ

5. การเรียกใช้ (Open) และการจัดเก็บ (Save) ในส่วนของการเรียกใช้จะเป็นการนำข้อมูลต้นไม้ตัดสินใจทั้งก่อนและหลังจากการตัดแต่งจากแฟ้มต้นไม้ตัดสินใจ ซึ่งมีส่วนขยายของแฟ้มเป็น mit ขึ้นมาแสดงผลหรือใช้งาน ส่วนการจัดเก็บจะเป็นการจัดเก็บข้อมูลของต้นไม้ตัดสินใจลงสู่แฟ้มต้นไม้ เพื่อใช้งานในภายหลังและเพื่อนำไปใช้ในการสร้างกฎการตัดสินใจต่อไป ดังรูปที่ 4.20



รูปที่ 4.20 แสดงการติดต่อกับแฟ้มต่าง ๆ ของการสร้างต้นไม้ตัดสินใจ

โปรแกรมการสร้างกฎการตัดสินใจและการใช้งานกฎการตัดสินใจ

โปรแกรมนี้จะเป็นการนำต้นไม้ตัดสินใจก่อนการตัดแต่งมาสร้างเป็นกฎการตัดสินใจ โดยจะเป็นการนำโปรแกรมเดิมซึ่งอยู่ในรูปภาษา C มาดัดแปลงให้อยู่ในรูปภาษา C++ ที่สามารถทำงานได้บนระบบปฏิบัติการวินโดว์ 95 และได้เพิ่มเติมในส่วนการติดต่อกับผู้ใช้และการแสดงผลให้อยู่ในรูปกราฟฟิก จากเดิมที่อยู่ในรูปตัวอักษร

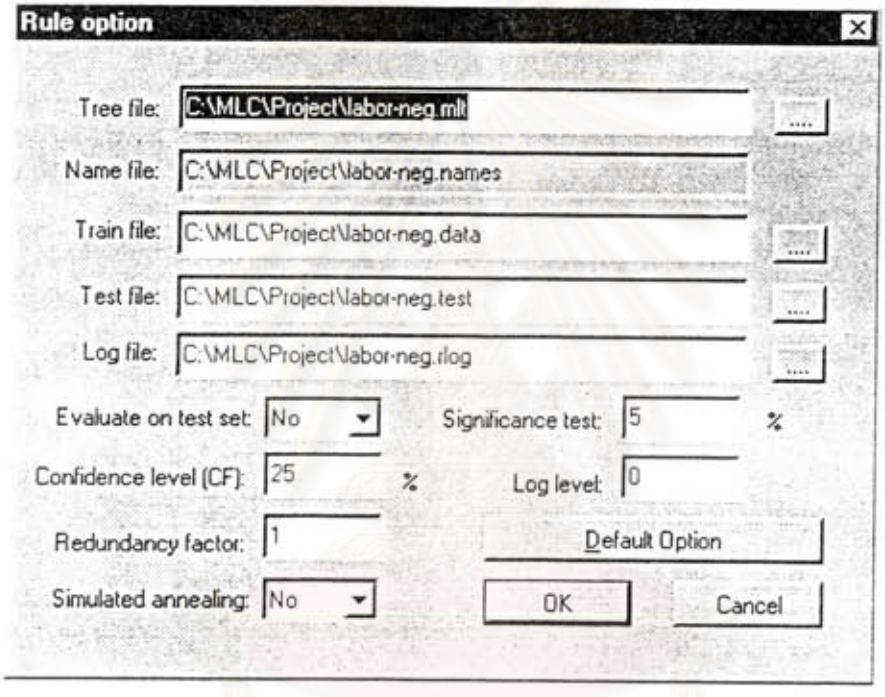
การดัดแปลงจะเริ่มจากการนำโปรแกรม c4.5rules และ consultr บนภาษา C เดิม มาสร้างเป็นคลาสใหม่บนภาษา C++ ซึ่งมีฟังก์ชั่นการทำงานเหมือนกับโปรแกรมเดิม โดยได้มีเปลี่ยนแปลงในส่วนของการตั้งค่าตัวเลือกที่ใช้ควบคุมโปรแกรมในการสร้างกฎการตัดสินใจจากเดิมที่เป็นคำสั่งแบบรายบรรทัด (Command Line) ให้เป็นการใส่ค่าผ่านกล่องตัวอักษร (Text Box) หรือเลือกค่าจากกล่องตัวเลือก (Combo Box) และได้เพิ่มส่วนการแสดงผลของกฎการตัดสินใจให้อยู่ในรูปการแสดงผลแบบต้นไม้ (Tree View) ซึ่งสามารถจะขยายหรือลดรูปต้นไม้เพื่อดูกฎบางส่วนได้

การทำงานของโปรแกรมนี้จะประกอบด้วยหน้าที่หลัก คือ

1. การตั้งค่าตัวเลือกในการสร้างกฎการตัดสินใจ เป็นการกำหนดค่าต่าง ๆ ที่ใช้ในการควบคุมโปรแกรม ซึ่งประกอบด้วยค่าที่สามารถกำหนดได้ดังรูปที่ 4.21

   - Tree file เป็นชื่อของแฟ้มข้อมูลต้นไม้ตัดสินใจที่ได้จากโปรแกรมการสร้างต้นไม้ตัดสินใจ ซึ่งมีส่วนขยายของแฟ้มเป็น mlt
   - Name file เป็นชื่อแฟ้มของ แฟ้มชื่อ ของตัวอย่างที่ใช้ในการสร้างกฎการตัดสินใจ
   - Train file เป็นชื่อแฟ้มของ แฟ้มตัวอย่างสอน
   - Test file เป็นชื่อแฟ้มของ แฟ้มตัวอย่างทดสอบ

- Log file เป็นชื่อแฟ้มที่เก็บส่วนของผลลัพธ์จากการสร้างกฎการตัดสินใจ ซึ่งมีส่วนขยายเป็น rlog
- Evaluate on test เป็นตัวเลือกที่กำหนดให้ทดสอบกฎการตัดสินใจที่ได้กับตัวอย่างทดสอบหรือไม่ โดยถ้ามีค่าเป็น No จะไม่มีการทดสอบกับตัวอย่างทดสอบ แต่ถ้ามีค่าเป็น Yes จะมีการทดสอบกฎที่ได้กับตัวอย่างทดสอบ เทียบได้กับตัวเลือก –u ในโปรแกรม c4.5rules
- Confidence level (CF) เป็นค่าที่กำหนดระดับการตัดแต่งกฎคล้ายกับการตัดแต่งต้นไม้ ซึ่งโดยปกติจะมีค่า 25% เทียบได้กับตัวเลือก –c ในโปรแกรม c4.5rules



รูปที่ 4.21 แสดงจอภาพของการตั้งค่าตัวเลือกในการสร้างกฎการตัดสินใจ

- Redundancy factor เป็นค่าที่กำหนดระดับความซ้ำซ้อนของลักษณะ ซึ่งจะมีผลต่อจำนวนของบิตที่ใช้ในการเข้ารหัสของกฎ ค่าที่มากแสดงว่าข้อมูลในแต่ละลักษณะมีการซ้ำซ้อนกันมากกว่าค่าที่น้อย โดยค่านี้จะมีประโยชน์ต่อผู้ที่เข้าใจลักษณะของข้อมูลเป็นอย่างดี ปกติค่านี้จะมีค่าเท่ากับ 1 เทียบได้กับตัวเลือก –r ในโปรแกรม c4.5rules
- Simulated annealing เป็นการกำหนดว่าจะใช้วิธีนี้ในการหาชุดของกฎที่ดีที่สุดหรือไม่ โดยถ้ามีค่าเป็น Yes ก็จะใช้วิธีนี้ แต่ถ้ามีค่าเป็น No ก็ไม่ใช้วิธีนี้ เทียบได้กับตัวเลือก –a ในโปรแกรม c4.5rules
- Significance test เป็นการทดสอบนัยสำคัญของแต่ละเงื่อนไขของกฎ โดยใช้วิธี Fisher's exact test แต่ละเงื่อนไขจะต้องมีนัยสำคัญภายในระดับความเชื่อมั่นที่กำหนดให้โดยตัวเลือกนี้ ซึ่งจะมีผลทำให้ได้กฎที่สั้นกว่าการตัดแต่งโดยทั่วไป ถ้าค่า

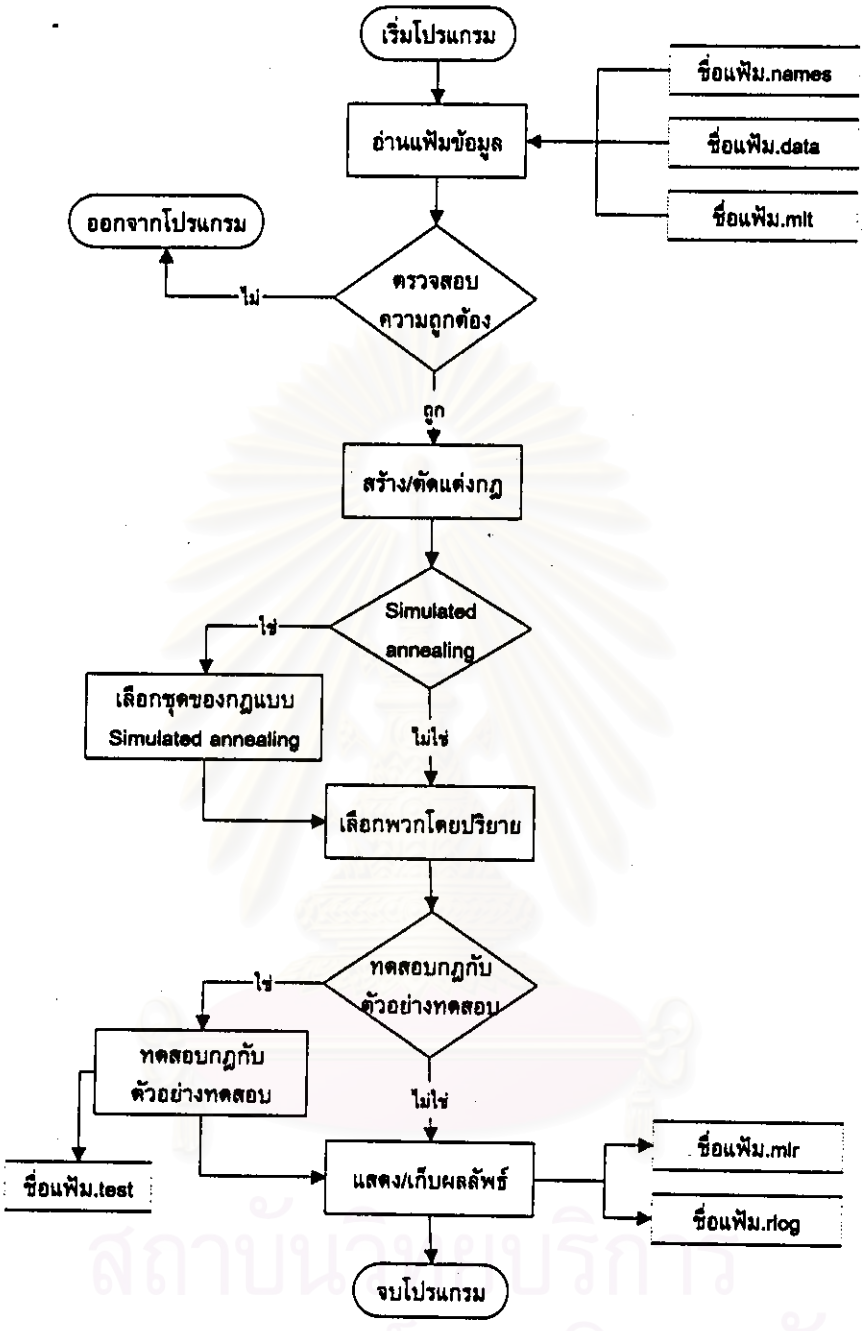
เท่ากับ 0 แสดงว่าไม่ต้องการใช้วิธีนี้ในการตัดแต่งกฎ เทียบได้กับตัวเลือก –F ของโปรแกรม c4.5rules

- Log level เป็นการกำหนดระดับของผลลัพธ์ที่จะแสดงออกมาในแฟ้มผลลัพธ์ โดยปกติจะมีค่าเป็น 0 เทียบได้กับตัวเลือก –v ในโปรแกรม c4.5rules

c4.5rules	โปรแกรมสร้างกฎการตัดสินใจ	ความหมายของตัวเลือก
-f ชื่อแฟ้ม	กำหนดชื่อแฟ้ม • Tree file • Name file • Train file • Test file	การกำหนดชื่อแฟ้มที่ใช้ในการสร้างกฎการตัดสินใจ
-u	Evaluated on test set	กำหนดการทดสอบบนตัวอย่างทดสอบ
-v ระดับผลลัพธ์	Log level	กำหนดระดับการแสดงผลลัพธ์
-c ค่า CF	Confidence level	กำหนดระดับความชื่อมั่นในการตัดแต่งกฎ
-F ค่า CF	Significance test	กำหนดการทดสอบนัยสำคัญ
-r ค่าความซ้ำ	Redundancy factor	กำหนดค่าการซ้ำของลักษณะ
-a	Simulated annealing	กำหนดวิธีการเลือกชุดของกฎ

ตารางที่ 4.2 แสดงตัวเลือกของโปรแกรม c4.5rules เดิมเปรียบเทียบกับโปรแกรมการสร้างกฎการตัดสินใจ

2. การสร้างกฎการตัดสินใจ (Make Rule) เป็นการสร้างกฎการตัดสินใจจากต้นไม้ตัดสินใจ โดยการทำงานจะเริ่มจากการอ่านแฟ้มชื่อและแฟ้มตัวอย่างสอน ตรวจสอบความถูกต้องของข้อมูลที่อ่านเข้ามา จากนั้นจะเริ่มขั้นตอนของการสร้างกฎโดยการอ่านแฟ้มข้อมูลต้นไม้ตัดสินใจ เพื่อนำต้นไม้ตัดสินใจก่อนตัดแต่งเข้าสู่หน่วยความจำ สร้างกฎที่เป็นไปได้จากต้นไม้ต้นนี้ซึ่งก็คือเส้นทางจากรากจนถึงใบของแต่ละใบ ตัดแต่งกฎเหล่านี้ตามค่า Significance test หรือค่า Confidence level และค่า Redundancy factor ที่กำหนด คัดเลือกชุดของกฎที่ดีที่สุด โดยถ้าตัวเลือก Simulated annealing เป็น Yes ก็จะใช้วิธีการเลือกชุดของกฎแบบ Simulated annealing ขั้นตอนสุดท้ายของการสร้างกฎจะเป็นการเลือกค่าพวกโดยปริยาย (Default Class) ให้กับตัวอย่างที่ไม่เข้ากฎใดเลย จากนั้นจะมีการทดสอบกฎที่ได้กับตัวอย่างสอน และถ้าตัวเลือก Evaluate on test set เป็น Yes ก็จะมีการทดสอบกฎที่ได้นี้บนตัวอย่างทดสอบ และเก็บผลลัพธ์ลงสู่แฟ้มผลลัพธ์ตามระดับผลลัพธ์ที่กำหนด จากนั้นจะแสดงชุดของกฎที่ได้ออกทางจอภาพ ดังได้สรุปขั้นตอนการทำงานของการสร้างกฎการตัดสินใจไว้ในรูปที่ 4.22

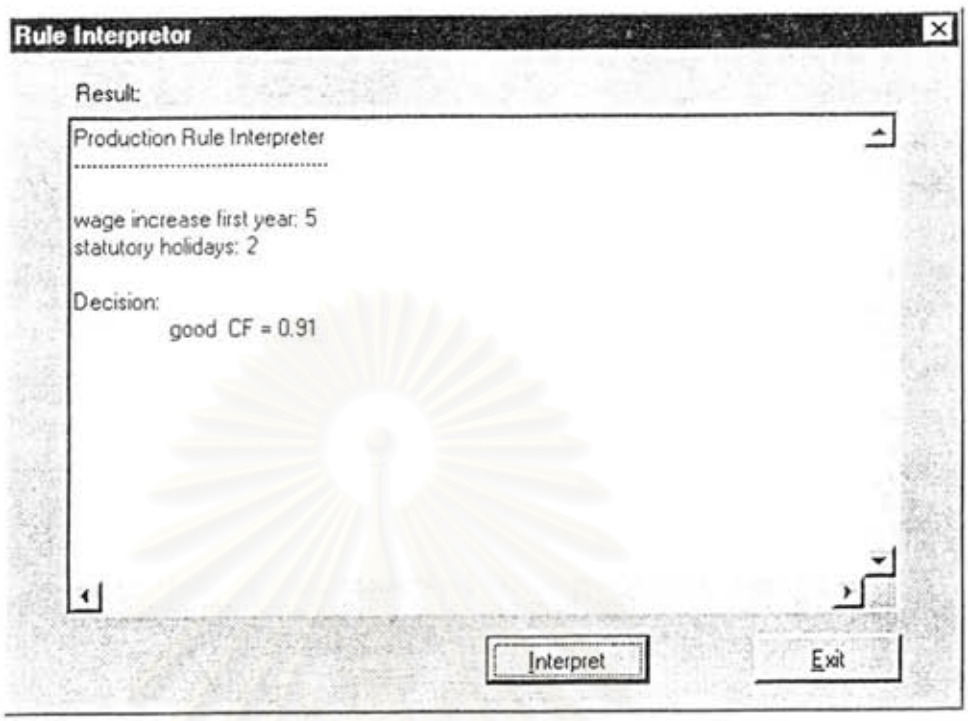


<u>รูปที่ 4.22</u> แสดงขั้นตอนการทำงานในการสร้างกฎการตัดสินใจ

3. การแสดงผลลัพธ์ (View Output) จะมีการแสดงผลลัพธ์ออกทางจอภาพดังรูปที่ 4.23 หลังจากการสร้างกฎสำเร็จลง และผู้ใช้สามารถจะดูผลลัพธ์จากแฟ้มผลลัพธ์ (ชื่อแฟ้ม.rlog) ซึ่งเป็นแฟ้มแบบตัวอักษรได้ โดยผ่านเมนูที่จะเรียกโปรแกรม Notepad ขึ้นมาเปิดแฟ้มผลลัพธ์ให้ทันทีเหมือนกับการแสดงผลลัพธ์จากการสร้างต้นไม้ตัดสินใจในรูปที่ 4.18

รูปที่ 4.23 การแสดงกฎการตัดสินใจออกทางจอภาพ

4. การใช้งานกฎการตัดสินใจ จะคล้ายกับการใช้งานต้นไม้ตัดสินใจ เพียงแต่เปลี่ยนคำถามในต้นไม้มาเป็นคำถามในกฎแทน การทำงานจะเริ่มจากหน้าต่างในรูปที่ 4.24 ซึ่งจะประกอบด้วยส่วนการแสดงผลการตัดสินใจ (Result) ที่จะแสดงผลการตัดสินใจเมื่อผู้ใช้ป้อนข้อมูลลักษณะของตัวอย่างที่โปรแกรมต้องการจนครบแล้ว ปุ่มเริ่มใช้งาน (Interpret) เป็นปุ่มที่จะเริ่มต้นการตัดสินใจ โดยโปรแกรมจะถามค่าของลักษณะของตัวอย่างเริ่มจากกฎแรกแล้วไล่ลงไปตามเงื่อนไขในกฎนี้ ถ้าไม่เข้ากฎก็จะถามลักษณะที่ยังไม่ได้ถามในกฎถัดไปเรื่อย ๆ ถ้ายังไม่เข้ากฎใดก็จะใช้ค่าโดยปริยายของพวกที่กำหนดเป็นค่าการตัดสินใจ โดยคำถามที่ถามจะเหมือนกับของคำถามบนต้นไม้ตัดสินใจที่มีทั้งลักษณะที่เป็นค่าต่อเนื่องและไม่ต่อเนื่อง การป้อนข้อมูลให้กับระบบก็จะเหมือนกัน สามารถที่จะป้อนเป็นช่วงได้ในลักษณะที่เป็นค่าต่อเนื่อง สามารถจะป้อนค่าของลักษณะไม่ต่อเนื่องแบบหลายค่าได้เช่นกัน ส่วนผลการตัดสินใจจะเป็นระดับความเชื่อมั่นเท่านั้นไม่มีช่วงของความเชื่อมั่น เมื่อผู้ใช้ต้องการออกจากการตัดสินใจกลางครัน ก็สามารถใช้ปุ่ม Cancel ได้เช่นเดียวกัน และเมื่อต้องการออกจากการใช้งานกฎการตัดสินใจ ก็สามารถใช้ปุ่ม Exit ได้



รูปที่ 4.24 แสดงจอภาพการใช้งานกฎการตัดสินใจ

5. การจัดเก็บ และเรียกใช้ข้อมูลกฎการตัดสินใจ จะเป็นการจัดเก็บกฎการตัดสินใจที่ได้ลงสู่แฟ้มกฎการตัดสินใจที่มีส่วนขยายเป็น mlr เพื่อนำกลับมาใช้งานได้ตามต้องการในภายหลัง โดยไม่ต้องสร้างกฎใหม่ ดังสรุปไว้ในการติดต่อกับแฟ้มข้อมูลการสร้างกฎการตัดสินใจดังรูปที่ 4.25

รูปที่ 4.25 แสดงการติดต่อกับแฟ้มข้อมูลต่าง ๆ ของโปรแกรมสร้างกฎการตัดสินใจ

การทดสอบความถูกต้องของโปรแกรม C4.5 บนระบบปฏิบัติการวินโดว์

การทดสอบโปรแกรม C4.5 ที่ปรับปรุงให้ทำงานบนระบบปฏิบัติการวินโดว์ จะใช้วิธีเปรียบเทียบผลลัพธ์ที่ได้จากการเรียนรู้จากโปรแกรม C4.5 ที่ทำงานบนระบบปฏิบัติการยูนิกซ์กับโปรแกรม C4.5 ที่ทำงานบนระบบปฏิบัติการวินโดว์ ในการเรียนรู้บนตัวอย่างที่ใช้ทดสอบทั้งหมด 8 ตัวอย่าง ดังแสดงไว้ในตารางที่ 4.3 การทดสอบจะเริ่มจากโปรแกรมการสร้างต้นไม้ตัดสินใจ โดยใช้ค่าตัวเลือกที่เป็นค่าโดยปริยายดังนี้ คือ การเรียนรู้จะใช้ค่ามาตรฐานอัตราส่วนเกน โดยมีค่า CF เท่ากับ 25% ค่า MINOBJS เท่ากับ 2 ไม่ใช้การสร้างต้นไม้แบบหน้าต่าง และมีการทดสอบต้นไม้ที่ได้กับตัวอย่างทดสอบ ซึ่งจะเปรียบเทียบความถูกต้องในด้านขนาดของต้นไม้ จำนวนตัวอย่างที่ผิดพลาด และอัตราส่วนความผิดพลาด ทั้งก่อนและหลังการตัดแต่งต้นไม้ รวมทั้งค่าคาดการณ์ความผิดพลาด (Expected Error) ของต้นไม้ตัดสินใจ ส่วนโปรแกรมการสร้างกฎการตัดสินใจจะใช้ค่าตัวเลือกที่เป็นค่าโดยปริยายดังนี้ คือ ค่าความเชื่อมั่น (Confidence level) เท่ากับ 25% ค่าความซ้ำ (Redundancy factor) เท่ากับ 1 ไม่มีการทดสอบนัยสำคัญ (Significance test) ไม่ใช้การเลือกชุดของกฎแบบ Simulated annealing และมีการทดสอบกฎที่ได้บนตัวอย่างทดสอบ จำนวนกฎที่ได้ จำนวนตัวอย่างที่ผิดพลาด และอัตราส่วนความผิดพลาด

ชื่อตัวอย่าง	จำนวนลักษณะ		จำนวนตัวอย่าง	
	ค่าต่อเนื่อง	ค่าไม่ต่อเนื่อง	สอน	ทดสอบ
Auto	15	10	136	69
Breast-cancer	-	9	191	95
Glass	9	-	142	72
Heart	13	-	180	90
Hepatitis	6	13	103	52
Iris	4	-	100	50
Liver-disorder	6	-	230	115
Credit data	12	15	3,600	400

ตารางที่ 4.3 แสดงรายละเอียดของตัวอย่างที่ใช้ในการทดสอบโปรแกรม C4.5

จากการทดสอบโปรแกรมการสร้างต้นไม้ตัดสินใจจะได้ผลลัพธ์ดังตารางที่ 4.4 และ 4.5 โดยผลที่ได้ในแต่ละหัวข้อที่ใช้เปรียบเทียบระหว่างโปรแกรมการสร้างต้นไม้ตัดสินใจที่ทำงานบนระบบปฏิบัติการวินโดว์กับโปรแกรมที่ทำงานบนระบบปฏิบัติการยูนิกซ์ จะให้ผลที่ไม่แตกต่างกันในแต่ละตัวอย่างที่ใช้ในการเปรียบเทียบ ส่วนในการทดสอบโปรแกรมการสร้างกฎการตัดสินใจจะให้ผลการทดสอบดังตารางที่ 4.6 ซึ่งให้ผลการเปรียบเทียบที่ไม่แตกต่างในแต่ละหัวข้อที่ใช้ทดสอบเช่นกัน

ชื่อตัวอย่าง	ชนิดตัวอย่าง	ระบบปฏิบัติการ	ต้นไม้ตัดสินใจก่อนการตัดแต่ง			ค่าคาดการณ์ความผิดพลาด
			ขนาดต้นไม้	ตัวอย่างผิดพลาด	อัตราส่วนความผิดพลาด	
Auto	สอน	ยูนิกซ์	123	2	1.5	27.1
		วินโดว์	123	2	1.5	27.1
	ทดสอบ	ยูนิกซ์	123	20	29.0	27.1
		วินโดว์	123	20	29.1	27.1
Breast-cancer	สอน	ยูนิกซ์	120	24	12.6	30.6
		วินโดว์	120	24	12.6	30.6
	ทดสอบ	ยูนิกซ์	120	36	37.9	30.6
		วินโดว์	120	36	37.9	30.6
Glass	สอน	ยูนิกซ์	39	10	7.0	24.2
		วินโดว์	39	10	7.0	24.2
	ทดสอบ	ยูนิกซ์	39	27	37.5	24.2
		วินโดว์	39	27	37.5	24.2
Heart	สอน	ยูนิกซ์	41	7	3.9	17.4
		วินโดว์	41	7	3.9	17.4
	ทดสอบ	ยูนิกซ์	41	16	17.8	17.4
		วินโดว์	41	16	17.8	17.4
Hepatitis	สอน	ยูนิกซ์	17	5	4.9	15.6
		วินโดว์	17	5	4.9	15.6
	ทดสอบ	ยูนิกซ์	17	9	17.3	15.6
		วินโดว์	17	9	17.3	15.6
Iris	สอน	ยูนิกซ์	7	1	1.0	6.3
		วินโดว์	7	1	1.0	6.3
	ทดสอบ	ยูนิกซ์	7	4	8.0	6.3
		วินโดว์	7	4	8.0	6.3
Liver-disorder	สอน	ยูนิกซ์	15	77	33.5	39.3
		วินโดว์	15	77	33.5	39.3
	ทดสอบ	ยูนิกซ์	15	45	39.1	39.3
		วินโดว์	15	45	39.1	39.3
Credit data	สอน	ยูนิกซ์	1,697	234	6.5	23.6
		วินโดว์	1,697	234	6.5	23.6
	ทดสอบ	ยูนิกซ์	1,697	115	28.8	23.6
		วินโดว์	1,697	115	28.8	23.6

ตารางที่ 4.4 แสดงผลการเปรียบเทียบโปรแกรมการสร้างต้นไม้ตัดสินใจ ระหว่างโปรแกรมที่ทำงานบนระบบปฏิบัติการยูนิกซ์กับวินโดว์ของต้นไม้ตัดสินใจก่อนการตัดแต่ง

ชื่อตัวอย่าง	ชนิดตัวอย่าง	ระบบปฏิบัติการ	ต้นไม้ตัดสินใจหลังการตัดแต่ง			ค่าคาดการณ์ความผิดพลาด
			ขนาดต้นไม้	ตัวอย่างผิดพลาด	อัตราส่วนความผิดพลาด	
Auto	สอน	ยูนิกซ์	67	11	8.1	27.1
		วินโดว์	67	11	8.1	27.1
	ทดสอบ	ยูนิกซ์	67	26	37.7	27.1
		วินโดว์	67	26	37.7	27.1
Breast-cancer	สอน	ยูนิกซ์	41	38	19.9	30.6
		วินโดว์	41	38	19.9	30.6
	ทดสอบ	ยูนิกซ์	41	24	25.3	30.6
		วินโดว์	41	24	25.3	30.6
Glass	สอน	ยูนิกซ์	39	10	7.0	24.2
		วินโดว์	39	10	7.0	24.2
	ทดสอบ	ยูนิกซ์	39	27	37.5	24.2
		วินโดว์	39	27	37.5	24.2
Heart	สอน	ยูนิกซ์	33	10	5.6	17.4
		วินโดว์	33	10	5.6	17.4
	ทดสอบ	ยูนิกซ์	33	15	16.7	17.4
		วินโดว์	33	15	16.7	17.4
Hepatitis	สอน	ยูนิกซ์	11	5	4.9	15.6
		วินโดว์	11	5	4.9	15.6
	ทดสอบ	ยูนิกซ์	11	10	19.2	15.6
		วินโดว์	11	10	19.2	15.6
Iris	สอน	ยูนิกซ์	7	1	1.0	6.3
		วินโดว์	7	1	1.0	6.3
	ทดสอบ	ยูนิกซ์	7	4	8.0	6.3
		วินโดว์	7	4	8.0	6.3
Liver-disorder	สอน	ยูนิกซ์	11	78	33.9	39.3
		วินโดว์	11	78	33.9	39.3
	ทดสอบ	ยูนิกซ์	11	43	37.4	39.3
		วินโดว์	11	43	37.4	39.3
Credit data	สอน	ยูนิกซ์	568	494	13.7	23.6
		วินโดว์	568	494	13.7	23.6
	ทดสอบ	ยูนิกซ์	568	108	27.0	23.6
		วินโดว์	568	108	27.0	23.6

ตารางที่ 4.5 แสดงผลการเปรียบเทียบโปรแกรมการสร้างต้นไม้ตัดสินใจ ระหว่างโปรแกรมที่ทำงานบนระบบปฏิบัติการยูนิกซ์กับวินโดว์ของต้นไม้ตัดสินใจหลังการตัดแต่ง

ชื่อตัวอย่าง	ชนิดตัวอย่าง	ระบบปฏิบัติการ	กฎการตัดสินใจ		
			จำนวนกฎ	ตัวอย่างผิดพลาด	อัตราส่วนความผิดพลาด
Auto	สอน	ยูนิกซ์	19	8	5.9
		วินโดว์	19	8	5.9
	ทดสอบ	ยูนิกซ์	17	27	39.1
		วินโดว์	17	27	39.1
Breast-cancer	สอน	ยูนิกซ์	6	46	24.1
		วินโดว์	6	46	24.1
	ทดสอบ	ยูนิกซ์	5	29	30.5
		วินโดว์	5	29	30.5
Glass	สอน	ยูนิกซ์	16	11	7.7
		วินโดว์	16	11	7.7
	ทดสอบ	ยูนิกซ์	14	24	33.3
		วินโดว์	14	24	33.3
Heart	สอน	ยูนิกซ์	9	16	8.9
		วินโดว์	9	16	8.9
	ทดสอบ	ยูนิกซ์	9	14	15.6
		วินโดว์	9	14	15.6
Hepatitis	สอน	ยูนิกซ์	7	5	4.9
		วินโดว์	7	5	4.9
	ทดสอบ	ยูนิกซ์	7	10	19.2
		วินโดว์	7	10	19.2
Iris	สอน	ยูนิกซ์	4	1	1.0
		วินโดว์	4	1	1.0
	ทดสอบ	ยูนิกซ์	3	4	8.0
		วินโดว์	3	4	8.0
Liver-disorder	สอน	ยูนิกซ์	4	68	29.6
		วินโดว์	4	68	29.6
	ทดสอบ	ยูนิกซ์	3	39	33.9
		วินโดว์	3	39	33.9
Credit data	สอน	ยูนิกซ์	4	68	29.6
		วินโดว์	4	68	29.6
	ทดสอบ	ยูนิกซ์	3	39	33.9
		วินโดว์	3	39	33.9

ตารางที่ 4.6 แสดงผลการเปรียบเทียบโปรแกรมการสร้างกฎการตัดสินใจ ระหว่างโปรแกรมที่ทำงานบนระบบปฏิบัติการยูนิกซ์กับวินโดว์

ในการทดสอบโปรแกรม C4.5 กับข้อมูลการอนุมัติสินเชื่อจะใช้วิธี เท็น-เวย์คลอสวาลิเดชั่น (ten-way cross validation) ซึ่งจะเป็นการแบ่งตัวอย่างสอนออกเป็น 10 ชุดเท่า ๆ กันอย่างสุ่ม ที่มีการกระจายของลักษณะแบ่งพวกเท่ากัน จากนั้นจะมีการจัดกลุ่มตัวอย่างแต่ละชุดเป็นตัวอย่างสอนและตัวอย่างทดสอบใหม่ โดยเลือก 1 ชุดตัวอย่างเป็นตัวอย่างทดสอบ และที่เหลืออีก 9 ชุดนำมารวมกันเป็นตัวอย่างสอน แล้วจึงนำตัวอย่างที่ได้ไปเรียนรู้สร้างเป็นต้นไม้ตัดสินใจและกฎการตัดสินใจ จากนั้นจึงเลือกชุดตัวอย่างใหม่ที่ไม่ซ้ำกับชุดตัวอย่างที่เคยเลือกไปเป็นตัวอย่างทดสอบแล้ว 1 ชุด เพื่อนำมาเป็นตัวอย่างทดสอบใหม่ และรวมชุดตัวอย่างที่เหลือเป็นตัวอย่างสอนอีกครั้ง นำไปเรียนรู้สร้างต้นไม้ตัดสินใจและกฎการตัดสินใจ กระทำเช่นนี้จนครบ 10 ครั้งหรือครบจำนวนชุดที่แบ่ง จากนั้นหาค่าเฉลี่ยจากการเรียนรู้ทั้ง 10 ครั้ง เพื่อเป็นตัวแทนของการเรียนรู้ในแต่ละค่าตัวเลือกที่กำหนดให้ โดยค่าที่ใช้เปรียบเทียบในโปรแกรมการสร้างต้นไม้ตัดสินใจจะประกอบด้วย ขนาดของต้นไม้ จำนวนตัวอย่างที่ผิดพลาด และอัตราส่วนความผิดพลาด ทั้งก่อนและหลังการตัดแต่งต้นไม้ รวมทั้งค่าคาดการณ์ความผิดพลาดของต้นไม้ตัดสินใจในการทดสอบกับตัวอย่างสอนและตัวอย่างทดสอบ ส่วนค่าที่ใช้เปรียบเทียบในโปรแกรมการสร้างกฎการตัดสินใจจะประกอบด้วย จำนวนตัวอย่างที่ผิดพลาด และอัตราส่วนความผิดพลาด ในการทดสอบโดยใช้ตัวอย่างสอนและตัวอย่างทดสอบ

การเปรียบเทียบจะทดลองตั้งค่าตัวเลือกแบบต่าง ๆ เปรียบเทียบกับผลที่ได้โดยการใช้ค่าตัวเลือกที่เป็นค่าโดยปริยาย เพื่อหาตัวเลือกที่เหมาะสมกับข้อมูลชุดนี้ ซึ่งตัวเลือกที่เหมาะสมจะทำให้ค่าอัตราส่วนความผิดพลาดมีค่าน้อย โดยหลังจากที่ได้ทดลองกับข้อมูลการอนุมัติสินเชื่อจำนวน 4,000 ตัวอย่าง แบ่งเป็นตัวอย่างสอน 3,600 ตัวอย่าง และตัวอย่างทดสอบ 400 ตัวอย่าง ซึ่งประกอบด้วยลักษณะไม่แบ่งพวก 27 ลักษณะและลักษณะแบ่งพวก 3 พวก ได้ผลการทดสอบการสร้างต้นไม้ตัดสินใจออกมาดังข้อมูลในตารางที่ 4.7 และ 4.8 และรูปที่ 4.26 ถึง รูปที่ 4.31 ส่วนการทดสอบการสร้างกฎการตัดสินใจได้ผลดังข้อมูลในตารางที่ 4.9 และ รูปที่ 4.32 ถึง รูปที่ 4.34

จากรูปที่ 4.26 และ 4.28 พบว่าเมื่อค่าตัวเลือก MINOBJS มีค่าสูงขึ้นจะทำให้ขนาดของต้นไม้ตัดสินใจมีขนาดที่เล็กลงทั้งก่อนและหลังการตัดแต่ง แต่จะทำให้อัตราส่วนความผิดพลาดเพิ่มขึ้นเมื่อทดสอบกับตัวอย่างสอนหรือตัวอย่างทดสอบดังรูปที่ 4.27 และ 4.29 สำหรับค่า CF จะมีผลต่อการตัดแต่งกิ่งโดยค่าที่น้อยจะทำให้มีการตัดแต่งกิ่งมากทำให้ได้ต้นไม้ตัดสินใจที่มีขนาดเล็กกว่าค่า CF ที่มาก ดังรูปที่ 4.30 ซึ่งก็จะมีผลต่ออัตราส่วนความผิดพลาด โดยถ้าต้นไม้ยิ่งเล็กก็จะมีอัตราส่วนความผิดพลาดสูงกว่าต้นไม้ที่มีขนาดใหญ่สำหรับการทดสอบบนตัวอย่างสอน แต่จะให้ผลไม่แตกต่างกันมากนักเมื่อทดสอบบนตัวอย่างทดสอบ ดังรูปที่ 4.31 ส่วนการใช้ค่ามาตรฐานเกนในการสร้างต้นไม้ตัดสินใจแทนการใช้ค่าอัตราส่วนมาตรฐานเกนจะให้ค่าอัตราส่วนความผิดพลาดที่สูงกว่า

ส่วนผลการทดลองสร้างกฎการตัดสินใจพบว่า เมื่อมีการตั้งค่า CF ที่ระดับต่าง ๆ กัน ดังรูปที่ 4.32 หรือใช้วิธี Simulated annealing หรือการตั้งค่า Significance ที่ระดับต่าง ๆ กัน ดังรูปที่ 4.34 จะไม่ทำให้อัตราส่วนความผิดพลาดแตกต่างกันมากนัก ต่างกับการตั้งค่าความซ้ำที่จะทำให้อัตราส่วนความผิดพลาดมีค่าต่ำลงเมื่อตั้งค่าความซ้ำเพิ่มขึ้นในการทดสอบกับตัวอย่างที่ใช้ในการเรียนรู้ แต่จะไม่แตกต่างกันเมื่อทดสอบกับตัวอย่างทดสอบ ดังรูปที่ 4.33

ในการวิจัยนี้ผลของต้นไม้ตัดสินใจและกฎการตัดสินใจที่ได้จากข้อมูลตัวอย่างการอนุมัติสินเชื่อไม่สามารถจะแสดงให้เห็นได้ เนื่องจากข้อมูลเป็นความลับของทางธนาคารและข้อมูลยังอยู่ในรูปของรหัสที่ไม่สื่อความหมายทำให้ไม่สามารถเข้าใจได้ จึงมีเพียงแต่ผลการทดสอบความถูกต้องของต้นไม้ตัดสินใจหรือกฎการตัดสินใจเท่านั้นที่ใช้เป็นตัวประเมินผล

ค่าตัวเลือก			ผลการสร้างต้นไม้ตัดสินใจ						
MIN OBJS	CF (%)	GAIN	ต้นไม้ตัดสินใจก่อนการตัดแต่ง			ต้นไม้ตัดสินใจหลังการตัดแต่ง			ค่าคาดการณ์ผิดพลาด
			ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	
1	1	ไม่	2802.0	22.2	0.6	126.3	837.7	23.3	33.3
3	1	ไม่	1308.3	328.2	9.1	103.4	842.9	23.4	33.2
5	1	ไม่	782.1	488.4	13.6	79.6	864.7	24.0	33.2
10	1	ไม่	407.7	661.2	18.4	62.4	892.7	24.8	33.3
15	1	ไม่	270.2	744.5	20.7	58.5	903.1	25.1	33.4
1	25	ไม่	2802.0	22.2	0.6	931.3	356.2	9.9	23.0
3	25	ไม่	1308.3	328.2	9.1	508.1	512.1	14.2	23.9
5	25	ไม่	782.1	488.4	13.6	361.4	598.5	16.6	24.6
10	25	ไม่	407.7	661.2	18.4	208.7	723.3	20.1	26.0
15	25	ไม่	270.2	744.5	20.7	145.4	787.0	21.9	26.8
1	50	ไม่	2802.0	22.2	0.6	1523.9	179.0	5.0	16.1
3	50	ไม่	1308.3	328.2	9.1	756.2	409.9	11.4	18.8
5	50	ไม่	782.1	488.4	13.6	494.5	537.5	14.9	20.4
10	50	ไม่	407.7	661.2	18.4	291.4	678.4	18.9	22.7
15	50	ไม่	270.2	744.5	20.7	196.4	760.1	21.1	24.0
1	75	ไม่	2802.0	22.2	0.6	2281.2	43.9	1.2	9.3
3	75	ไม่	1308.3	328.2	9.1	941.9	367.2	10.2	16.1
5	75	ไม่	782.1	488.4	13.6	579.4	512.2	14.2	18.7
10	75	ไม่	407.7	661.2	18.4	316.0	671.1	18.7	21.7
15	75	ไม่	270.2	744.5	20.7	213.4	753.7	20.9	23.2
1	99	ไม่	2802.0	22.2	0.6	2613.9	23.1	0.7	1.9
3	99	ไม่	1308.3	328.2	9.1	1146.1	339.3	9.4	13.2
5	99	ไม่	782.1	488.4	13.6	646.3	500.8	13.9	17.1
10	99	ไม่	407.7	661.2	18.4	341.3	666.5	18.5	20.7
15	99	ไม่	270.2	744.5	20.7	233.7	747.9	20.8	22.5
1	1	ใช่	3206.2	38.9	1.1	57.8	970.2	26.9	34.4
3	1	ใช่	1311.6	401.3	11.1	58.4	966.6	26.8	34.4
5	1	ใช่	789.3	573.4	15.9	51.0	971.1	27.0	34.3
10	1	ใช่	409.4	727.2	20.2	46.1	976.6	27.1	34.4
15	1	ใช่	267.3	792.3	22.0	42.6	987.1	27.4	34.5
1	25	ใช่	3206.2	38.9	1.1	839.3	475.4	13.2	26.1
3	25	ใช่	1311.6	401.3	11.1	461.7	613.0	17.0	26.8
5	25	ใช่	789.3	573.4	15.9	291.3	720.5	20.0	27.5
10	25	ใช่	409.4	727.2	20.2	175.6	804.8	22.4	28.2
15	25	ใช่	267.3	792.3	22.0	143.1	839.0	23.3	28.4

ตารางที่ 4.7 แสดงผลการทดลองสร้างต้นไม้ตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อ เมื่อทดสอบบนตัวอย่างสอน

ค่าตัวเลือก			ผลการสร้างต้นไม้ตัดสินใจ						
MIN OBJS	CF (%)	GAIN	ต้นไม้ตัดสินใจก่อนการตัดแต่ง			ต้นไม้ตัดสินใจหลังการตัดแต่ง			ค่าคาดการณ์ผิดพลาด
			ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	
1	50	ใช่	3206.2	38.9	1.1	1600.4	248.6	6.9	19.3
3	50	ใช่	1311.6	401.3	11.1	712.4	502.3	13.9	21.8
5	50	ใช่	789.3	573.4	15.9	455.8	634.0	17.6	23.3
10	50	ใช่	409.4	727.2	20.2	270.1	751.4	20.9	25.0
15	50	ใช่	267.3	792.3	22.0	192.3	805.3	22.4	25.7
1	75	ใช่	3206.2	38.9	1.1	2565.6	79.2	2.2	12.1
3	75	ใช่	1311.6	401.3	11.1	934.8	445.4	12.4	19.1
5	75	ใช่	789.3	573.4	15.9	562.4	603.1	16.8	21.7
10	75	ใช่	409.4	727.2	20.2	295.4	743.2	20.6	24.0
15	75	ใช่	267.3	792.3	22.0	206.4	799.6	22.2	24.9
1	99	ใช่	3206.2	38.9	1.1	3064.1	39.5	1.1	3.4
3	99	ใช่	1311.6	401.3	11.1	1145.2	414.5	11.5	16.2
5	99	ใช่	789.3	573.4	15.9	642.6	588.2	16.3	20.1
10	99	ใช่	409.4	727.2	20.2	324.9	736.5	20.5	23.0
15	99	ใช่	267.3	792.3	22.0	219.2	795.9	22.1	24.1

ตารางที่ 4.7 แสดงผลการทดลองสร้างต้นไม้ตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อ เมื่อทดสอบบนตัวอย่างสอน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

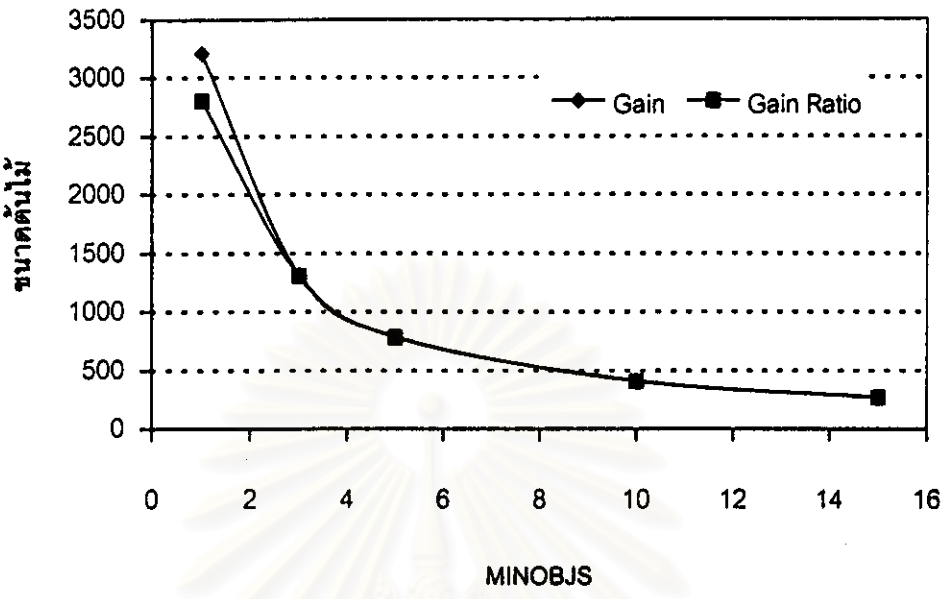
ค่าตัวเลือก			ผลการสร้างต้นไม้ตัดสินใจ						
MIN OBJS	CF (%)	GAIN	ต้นไม้ตัดสินใจก่อนการตัดแต่ง			ต้นไม้ตัดสินใจหลังการตัดแต่ง			ค่าคาดการณ์ผิดพลาด
			ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	
1	1	ไม่	2802.0	115.4	28.9	126.3	108.0	27.0	33.3
3	1	ไม่	1308.3	114.2	28.6	103.4	109.0	27.3	33.2
5	1	ไม่	782.1	112.8	28.2	79.6	107.9	27.0	33.2
10	1	ไม่	407.7	109.4	27.3	62.4	106.4	26.6	33.3
15	1	ไม่	270.2	112.2	28.1	58.5	108.5	27.1	33.4
1	25	ไม่	2802.0	115.4	28.9	931.3	106.3	26.6	23.0
3	25	ไม่	1308.3	114.2	28.6	508.1	106.9	26.7	23.9
5	25	ไม่	782.1	112.8	28.2	361.4	108.2	27.1	24.6
10	25	ไม่	407.7	109.4	27.3	208.7	105.8	26.5	26.0
15	25	ไม่	270.2	112.2	28.1	145.4	106.9	26.7	26.8
1	50	ไม่	2802.0	115.4	28.9	1523.9	107.3	26.8	16.1
3	50	ไม่	1308.3	114.2	28.6	756.2	107.6	26.9	18.8
5	50	ไม่	782.1	112.8	28.2	494.5	109.2	27.3	20.4
10	50	ไม่	407.7	109.4	27.3	291.4	105.9	26.5	22.7
15	50	ไม่	270.2	112.2	28.1	196.4	110.2	27.8	24.0
1	75	ไม่	2802.0	115.4	28.9	2281.2	114.8	28.7	9.3
3	75	ไม่	1308.3	114.2	28.6	941.9	109.3	27.3	16.1
5	75	ไม่	782.1	112.8	28.2	579.4	110.4	27.6	18.7
10	75	ไม่	407.7	109.4	27.3	316.0	106.2	26.5	21.7
15	75	ไม่	270.2	112.2	28.1	213.4	111.4	27.9	23.2
1	99	ไม่	2802.0	115.4	28.9	2613.9	115.0	28.8	1.9
3	99	ไม่	1308.3	114.2	28.6	1146.1	112.7	28.2	13.2
5	99	ไม่	782.1	112.8	28.2	646.3	111.5	27.9	17.1
10	99	ไม่	407.7	109.4	27.3	341.3	107.1	26.8	20.7
15	99	ไม่	270.2	112.2	28.1	233.7	111.0	27.7	22.5
1	1	ใช่	3206.2	121.2	30.3	57.8	118.8	29.7	34.4
3	1	ใช่	1311.6	120.3	30.1	58.4	118.6	29.7	34.4
5	1	ใช่	789.3	124.9	31.2	51.0	118.2	29.5	34.3
10	1	ใช่	409.4	119.5	29.9	46.1	117.8	29.4	34.4
15	1	ใช่	267.3	114.1	28.5	42.6	117.4	29.4	34.5
1	25	ใช่	3206.2	121.2	30.3	839.3	111.2	27.8	26.1
3	25	ใช่	1311.6	120.3	30.1	461.7	115.3	28.8	26.8
5	25	ใช่	789.3	124.9	31.2	291.3	117.5	29.4	27.5
10	25	ใช่	409.4	119.5	29.9	175.6	117.1	29.3	28.2
15	25	ใช่	267.3	114.1	28.5	143.1	117.6	29.4	28.4

ตารางที่ 4.8 แสดงผลการทดลองสร้างต้นไม้ตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อ เมื่อทดสอบกับข้อมูลตัวอย่างทดสอบ

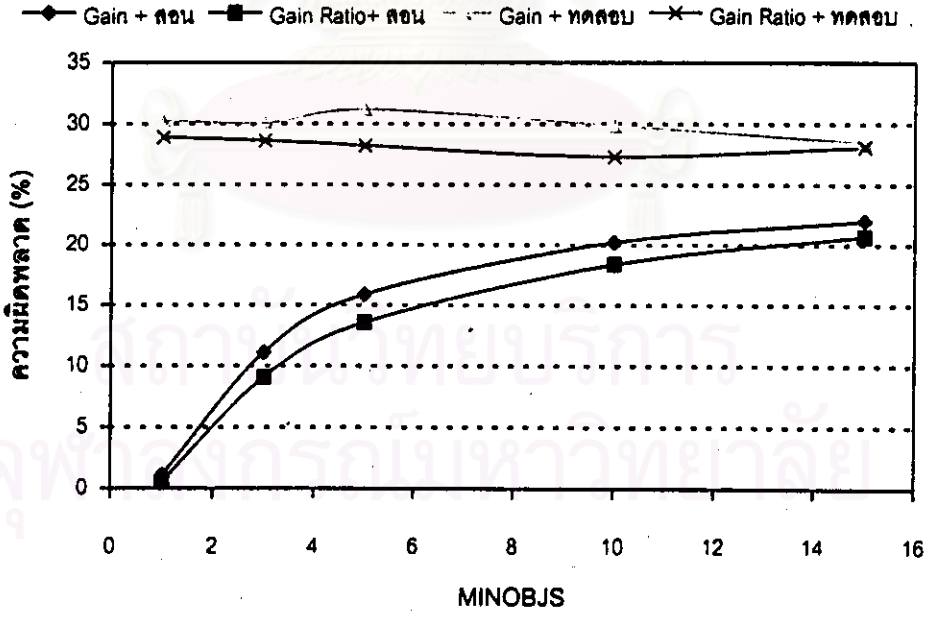
ค่าตัวเลือก			ผลการสร้างต้นไม้ตัดสินใจ						
MIN OBJS	CF (%)	GAIN	ต้นไม้ตัดสินใจก่อนการตัดแต่ง			ต้นไม้ตัดสินใจหลังการตัดแต่ง			ค่าคาดการณ์ผิดพลาด
			ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	ขนาดต้นไม้	ตัวอย่างผิดพลาด	% ผิดพลาด	
1	50	ใช่	3206.2	121.2	30.3	1600.4	114.3	28.6	19.3
3	50	ใช่	1311.6	120.3	30.1	712.4	113.3	28.3	21.8
5	50	ใช่	789.3	124.9	31.2	455.8	117.2	29.3	23.3
10	50	ใช่	409.4	119.5	29.9	270.1	118.5	29.6	25.0
15	50	ใช่	267.3	114.1	28.5	192.3	114.8	28.7	25.7
1	75	ใช่	3206.2	121.2	30.3	2565.6	119.0	29.7	12.1
3	75	ใช่	1311.6	120.3	30.1	934.8	116.4	29.1	19.1
5	75	ใช่	789.3	124.9	31.2	562.4	119.0	29.8	21.7
10	75	ใช่	409.4	119.5	29.9	295.4	118.4	29.6	24.0
15	75	ใช่	267.3	114.1	28.5	206.4	115.1	28.8	24.9
1	99	ใช่	3206.2	121.2	30.3	3064.1	120.6	30.1	3.4
3	99	ใช่	1311.6	120.3	30.1	1145.2	118.5	29.6	16.2
5	99	ใช่	789.3	124.9	31.2	642.6	121.7	30.4	20.1
10	99	ใช่	409.4	119.5	29.9	324.9	118.1	29.5	23.0
15	99	ใช่	267.3	114.1	28.5	219.2	115.2	28.8	24.1

ตารางที่ 4.8 แสดงผลการทดลองสร้างต้นไม้ตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อ เมื่อทดสอบกับข้อมูลตัวอย่างทดสอบ

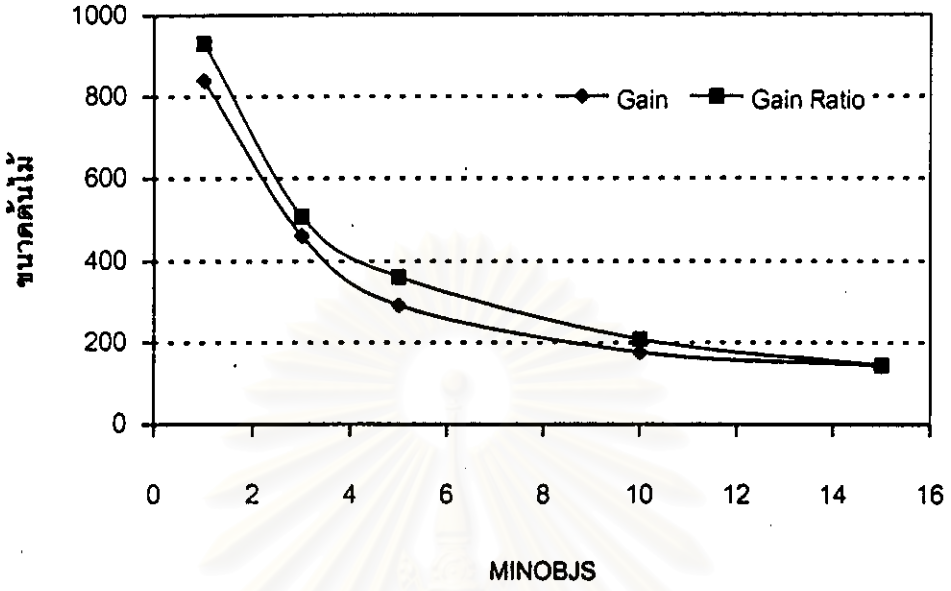
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



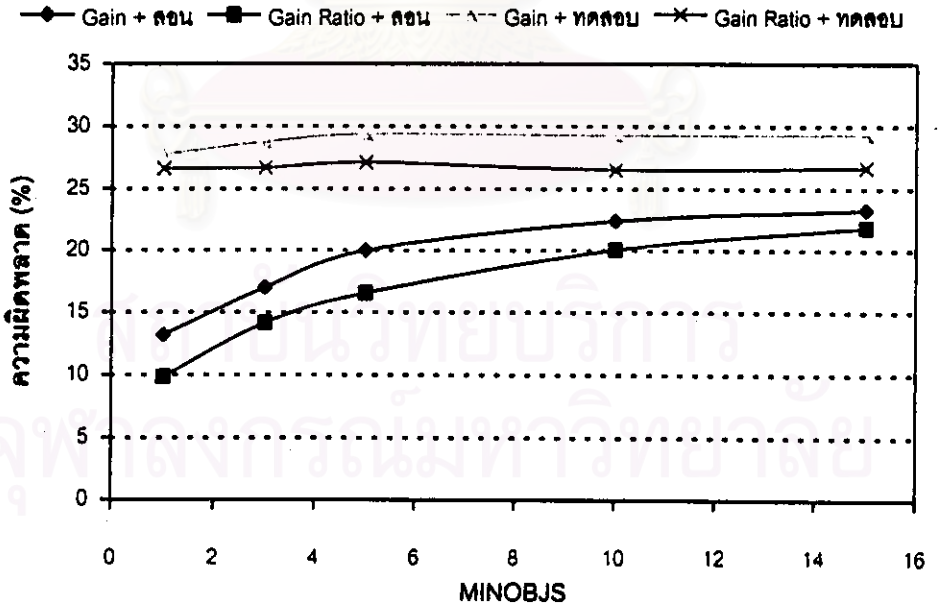
รูปที่ 4.26 กราฟแสดงขนาดของต้นไม้ตัดสินใจก่อนการตัดแต่งเมื่อตั้งค่า MINOBJS ที่ระดับต่าง ๆ และค่า CF เท่ากับ 25 % กับการใช้ค่ามาตรฐานเกนหรือค่ามาตรฐานอัตราส่วนเกน



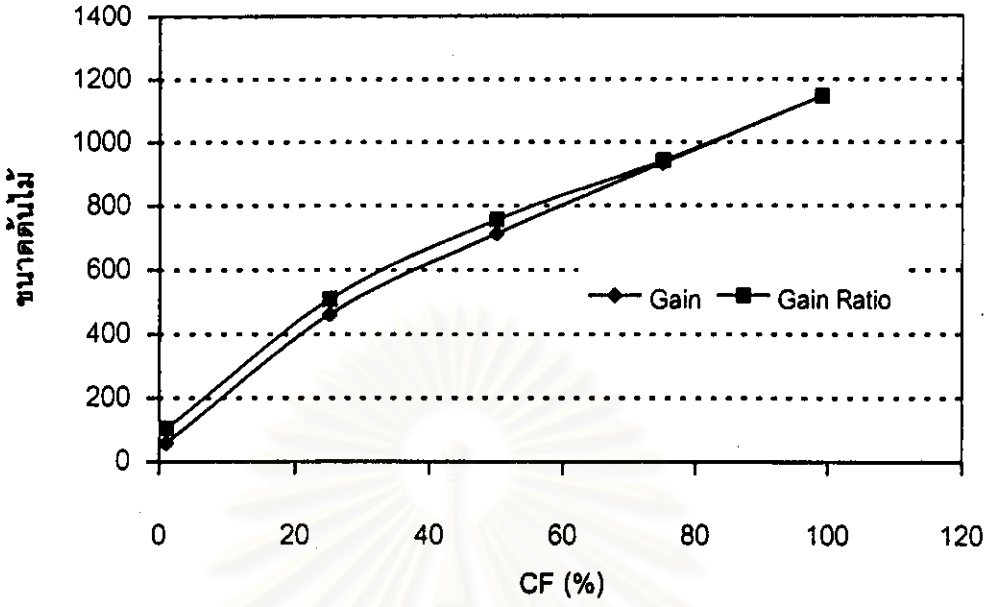
รูปที่ 4.27 กราฟแสดงอัตราส่วนความผิดพลาดของต้นไม้ตัดสินใจก่อนการตัดแต่งเมื่อตั้งค่า MINOBJS ที่ระดับต่าง ๆ และค่า CF เท่ากับ 25 % กับการใช้ค่ามาตรฐานเกนหรือค่ามาตรฐานอัตราส่วนเกน โดยทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ



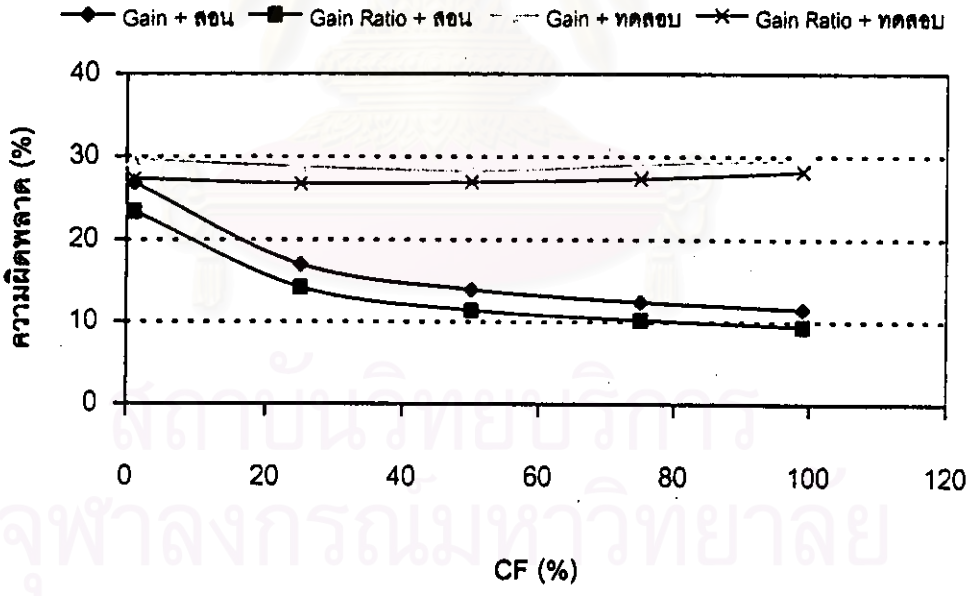
รูปที่ 4.28 กราฟแสดงขนาดของต้นไม้ตัดสินใจหลังการตัดแต่งเมื่อตั้งค่า MINOBJS ที่ระดับต่าง ๆ และค่า CF เท่ากับ 25 % กับการใช้ค่ามาตรฐานเกนหรือค่ามาตรฐานอัตราส่วนเกน



รูปที่ 4.29 กราฟแสดงอัตราส่วนความผิดพลาดของต้นไม้ตัดสินใจหลังการตัดแต่งเมื่อตั้งค่า MINOBJS ที่ระดับต่าง ๆ และค่า CF เท่ากับ 25 % กับการใช้ค่ามาตรฐานเกนหรือค่ามาตรฐานอัตราส่วนเกน โดยทดสอบบนตัวอย่างทดสอนและตัวอย่างทดสอบ



รูปที่ 4.30 กราฟแสดงขนาดของต้นไม้ตัดสินใจหลังการตัดแต่งเมื่อตั้งค่า CF ที่ระดับต่าง ๆ และค่า MINOBJS เท่ากับ 3 กับการใช้ค่ามาตรฐานเกนหรือค่ามาตรฐานอัตราส่วนเกน



รูปที่ 4.31 กราฟแสดงอัตราส่วนความผิดพลาดของต้นไม้ตัดสินใจหลังการตัดแต่งเมื่อตั้งค่า CF ที่ระดับต่าง ๆ และค่า MINOBJS เท่ากับ 3 กับการใช้ค่ามาตรฐานเกนหรือค่ามาตรฐานอัตราส่วนเกน โดยทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
1	1	ไม่	ไม่	756.4	21.0	106.9	26.7
25	1	ไม่	ไม่	707.3	19.6	106.9	26.7
50	1	ไม่	ไม่	697.8	19.4	110.7	27.6
75	1	ไม่	ไม่	716.5	19.9	108.6	27.1
99	1	ไม่	ไม่	747.2	20.8	107.7	26.9
1	5	ไม่	ไม่	633.2	17.6	105.9	26.5
25	5	ไม่	ไม่	400.9	11.1	110.8	27.7
50	5	ไม่	ไม่	344.7	9.6	112.7	28.2
75	5	ไม่	ไม่	315.5	8.8	116.1	29.0
99	5	ไม่	ไม่	320.6	8.9	116.1	29.0
1	10	ไม่	ไม่	626.1	17.4	107.0	26.7
25	10	ไม่	ไม่	370.6	10.3	111.5	27.9
50	10	ไม่	ไม่	313.8	8.7	113.9	28.5
75	10	ไม่	ไม่	297.9	8.3	116.5	29.1
99	10	ไม่	ไม่	299.5	8.3	115.7	28.9
1	15	ไม่	ไม่	619.3	17.2	106.9	26.7
25	15	ไม่	ไม่	361.9	10.1	111.5	27.9
50	15	ไม่	ไม่	313.8	8.7	114.5	28.6
75	15	ไม่	ไม่	295.6	8.2	115.4	28.9
99	15	ไม่	ไม่	295.9	8.2	115.6	28.9
1	20	ไม่	ไม่	618.7	17.2	107.3	26.8
25	20	ไม่	ไม่	361.9	10.1	111.7	27.9
50	20	ไม่	ไม่	313.6	8.7	114.5	28.6
75	20	ไม่	ไม่	291.6	8.1	115.5	28.9
99	20	ไม่	ไม่	294.0	8.2	115.6	28.9
1	1	ใช่	ไม่	779.4	21.7	108.4	27.1
25	1	ใช่	ไม่	710.7	19.7	110.2	27.6
50	1	ใช่	ไม่	687.0	19.1	107.7	26.9
75	1	ใช่	ไม่	706.4	19.6	110.1	27.5
99	1	ใช่	ไม่	750.8	20.9	111.0	27.8
1	5	ใช่	ไม่	671.6	18.6	105.9	26.5
25	5	ใช่	ไม่	470.4	13.1	105.5	26.4
50	5	ใช่	ไม่	378.6	10.5	112.7	28.2
75	5	ใช่	ไม่	347.6	9.7	112.6	28.1
99	5	ใช่	ไม่	349.8	9.7	113.9	28.5

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
1	10	ใช่	ไม่	646.4	18.0	109.0	27.3
25	10	ใช่	ไม่	412.1	11.4	108.8	27.2
50	10	ใช่	ไม่	356.5	9.9	112.4	28.1
75	10	ใช่	ไม่	321.2	8.9	114.0	28.5
99	10	ใช่	ไม่	314.9	8.8	113.2	28.3
1	15	ใช่	ไม่	641.0	17.8	107.4	26.8
25	15	ใช่	ไม่	422.1	11.7	110.3	27.6
50	15	ใช่	ไม่	351.6	9.8	113.8	28.5
75	15	ใช่	ไม่	314.3	8.7	114.7	28.7
99	15	ใช่	ไม่	319.3	8.9	112.6	28.1
1	20	ใช่	ไม่	654.3	18.2	107.6	26.9
25	20	ใช่	ไม่	395.8	11.0	111.3	27.8
50	20	ใช่	ไม่	353.5	9.8	111.6	27.9
75	20	ใช่	ไม่	309.7	8.6	116.2	29.0
99	20	ใช่	ไม่	314.7	8.7	112.5	28.1
1	1	ไม่	1	776.5	21.6	102.9	25.7
1	1	ไม่	25	762.2	21.2	107.7	26.9
1	1	ไม่	50	759.8	21.1	106.6	26.7
1	1	ไม่	75	759.8	21.1	106.6	26.7
1	1	ไม่	99	759.8	21.1	106.6	26.7
25	1	ไม่	1	762.2	21.2	105.1	26.3
25	1	ไม่	25	696.3	19.3	106.8	26.7
25	1	ไม่	50	704.1	19.6	106.8	26.7
25	1	ไม่	75	704.1	19.6	106.8	26.7
25	1	ไม่	99	704.1	19.6	106.8	26.7
50	1	ไม่	1	776.9	21.6	108.7	27.2
50	1	ไม่	25	670.4	18.6	108.4	27.1
50	1	ไม่	50	619.3	19.2	112.2	26.0
50	1	ไม่	75	696.4	19.4	110.9	27.7
50	1	ไม่	99	696.4	19.4	110.9	27.7
75	1	ไม่	1	770.1	21.4	109.1	27.3
75	1	ไม่	25	678.2	18.8	108.1	27.0
75	1	ไม่	50	694.5	19.3	106.4	26.6
75	1	ไม่	75	718.7	20.2	108.5	27.1
75	1	ไม่	99	716.9	19.9	109.2	27.3

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
99	1	ไม่	1	762.7	21.2	106.1	26.5
99	1	ไม่	25	677.4	18.8	109.3	27.4
99	1	ไม่	50	704.3	19.6	108.3	27.1
99	1	ไม่	75	739.9	20.6	107.5	26.9
99	1	ไม่	99	747.2	20.8	107.7	26.9
1	5	ไม่	1	720.9	20.0	107.6	26.9
1	5	ไม่	25	629.6	17.5	107.1	26.8
1	5	ไม่	50	633.1	17.6	106.6	26.6
1	5	ไม่	75	633.1	17.6	106.6	26.6
1	5	ไม่	99	633.1	17.6	106.6	26.6
25	5	ไม่	1	695.2	19.3	107.9	26.9
25	5	ไม่	25	399.8	11.1	110.6	27.7
25	5	ไม่	50	400.3	11.1	111.3	27.8
25	5	ไม่	75	400.3	11.1	111.3	27.8
25	5	ไม่	99	400.3	11.1	111.3	27.8
50	5	ไม่	1	715.0	19.9	109.8	27.5
50	5	ไม่	25	362.8	10.1	111.5	27.9
50	5	ไม่	50	344.8	9.6	112.9	28.2
50	5	ไม่	75	344.5	9.6	112.9	28.2
50	5	ไม่	99	344.5	9.6	112.9	28.2
75	5	ไม่	1	697.9	19.4	109.7	27.4
75	5	ไม่	25	357.8	9.9	114.6	28.6
75	5	ไม่	50	328.1	9.1	116.4	29.1
75	5	ไม่	75	318.8	8.8	115.9	29.0
75	5	ไม่	99	315.7	8.8	116.2	29.1
99	5	ไม่	1	713.4	19.8	109.3	27.3
99	5	ไม่	25	368.9	10.3	114.5	28.6
99	5	ไม่	50	320.6	8.9	117.0	29.2
99	5	ไม่	75	321.6	8.9	119.1	29.8
99	5	ไม่	99	320.6	8.9	116.1	29.0
1	10	ไม่	1	721.0	20.0	106.1	26.5
1	10	ไม่	25	628.0	17.4	107.9	27.0
1	10	ไม่	50	625.3	17.4	107.5	26.9
1	10	ไม่	75	625.3	17.4	107.5	26.9
1	10	ไม่	99	625.3	17.4	107.5	26.9

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
25	10	ไม่	1	699.7	19.4	106.4	26.4
25	10	ไม่	25	386.0	10.7	110.4	27.6
25	10	ไม่	50	371.5	10.3	111.5	27.9
25	10	ไม่	75	371.5	10.3	111.5	27.9
25	10	ไม่	99	371.5	10.3	111.5	27.9
50	10	ไม่	1	704.4	19.6	110.8	27.7
50	10	ไม่	25	342.9	9.5	111.8	28.0
50	10	ไม่	50	314.3	8.7	114.2	28.6
50	10	ไม่	75	313.4	8.7	114.0	28.5
50	10	ไม่	99	313.4	8,7	114.0	28.5
75	10	ไม่	1	696.7	19.4	109.0	27.2
75	10	ไม่	25	340.3	9.4	114.8	28.7
75	10	ไม่	50	300.8	8.4	116.1	29.0
75	10	ไม่	75	297.9	8.3	116.6	29.1
75	10	ไม่	99	297.8	8.3	116.4	29.1
99	10	ไม่	1	697.5	19.4	109.0	27.2
99	10	ไม่	25	341.6	9.5	114.1	28.5
99	10	ไม่	50	305.8	8.5	116.5	29.1
99	10	ไม่	75	297.6	8.3	117.2	29.3
99	10	ไม่	99	299.5	8.3	115.7	28.9
1	15	ไม่	1	720.0	20.0	106.1	26.5
1	15	ไม่	25	613.0	17.0	106.6	26.6
1	15	ไม่	50	619.0	17.2	107.3	26.8
1	15	ไม่	75	619.0	17.2	107.3	26.8
1	15	ไม่	99	619.0	17.2	107.3	26.8
25	15	ไม่	1	688.7	19.1	106.1	26.5
25	15	ไม่	25	387.8	10.8	109.2	27.3
25	15	ไม่	50	362.0	10.1	111.7	27.9
25	15	ไม่	75	362.0	10.1	111.7	27.9
25	15	ไม่	99	362.0	10.1	111.7	27.9
50	15	ไม่	1	696.2	19.3	111.1	27.8
50	15	ไม่	25	335.9	9.3	113.0	28.2
50	15	ไม่	50	313.4	8.7	114.6	28.7
50	15	ไม่	75	312.1	8.7	114.4	28.6
50	15	ไม่	99	312.1	8.7	114.4	28.6

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
75	15	ไม่	1	699.8	19.4	108.9	27.2
75	15	ไม่	25	336.8	9.3	114.5	28.6
75	15	ไม่	50	297.6	8.3	115.7	28.9
75	15	ไม่	75	295.4	8.2	115.5	28.9
75	15	ไม่	99	295.3	8.2	115.3	28.8
99	15	ไม่	1	696.2	19.3	108.5	27.1
99	15	ไม่	25	340.0	9.4	114.7	28.7
99	15	ไม่	50	302.4	8.4	116.0	29.0
99	15	ไม่	75	291.9	8.1	115.2	28.8
99	15	ไม่	99	295.9	8.2	115.6	28.9
1	20	ไม่	1	722.4	20.1	106.3	26.6
1	20	ไม่	25	611.9	17.0	106.7	26.7
1	20	ไม่	50	615.9	17.1	108.4	27.1
1	20	ไม่	75	615.9	17.1	108.4	27.1
1	20	ไม่	99	615.9	17.1	108.4	27.1
25	20	ไม่	1	689.0	19.1	106.2	26.6
25	20	ไม่	25	378.7	10.5	109.6	27.4
25	20	ไม่	50	362.9	10.1	111.9	28.0
25	20	ไม่	75	362.9	10.1	111.9	28.0
25	20	ไม่	99	362.9	10.1	111.9	28.0
50	20	ไม่	1	702.8	19.5	111.6	27.9
50	20	ไม่	25	337.2	9.4	111.2	27.8
50	20	ไม่	50	313.3	8.7	114.4	28.6
50	20	ไม่	75	312,5	8.7	114.4	28.6
50	20	ไม่	99	312.5	8.7	114.4	28.6
75	20	ไม่	1	688.1	19.1	110.2	27.5
75	20	ไม่	25	337.8	9.4	112.8	28.2
75	20	ไม่	50	297.5	8.2	115.4	28.8
75	20	ไม่	75	287.4	8.0	115.9	29.0
75	20	ไม่	99	291.3	8.1	115.4	28.9
99	20	ไม่	1	696.5	19.4	109.1	27.3
99	20	ไม่	25	338.0	9.4	114.9	28.7
99	20	ไม่	50	298.3	8.3	116.4	29.1
99	20	ไม่	75	290.8	8.1	115.7	28.9
99	20	ไม่	99	294.0	8.2	115.6	28.9

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
1	1	ใช่	1	800.9	22.2	109.8	27.4
1	1	ใช่	25	785.8	21.8	111.8	28.0
1	1	ใช่	50	775.3	21.5	109.0	27.2
1	1	ใช่	75	775.3	21.5	109.0	27.2
1	1	ใช่	99	775.3	21.5	109.0	27.2
25	1	ใช่	1	812.5	22.6	110.9	27.8
25	1	ใช่	25	711.7	19.8	108.3	27.1
25	1	ใช่	50	708.3	19.7	110.0	27.5
25	1	ใช่	75	708.3	19.7	110.0	27.5
25	1	ใช่	99	708.3	19.7	110.0	27.5
50	1	ใช่	1	810.5	22.5	113.0	28.2
50	1	ใช่	25	669.4	18.6	107.5	26.9
50	1	ใช่	50	696.5	19.4	107.5	26.9
50	1	ใช่	75	686.7	19.1	108.4	27.1
50	1	ใช่	99	686.7	19.1	108.4	27.1
75	1	ใช่	1	842.6	23.4	112.3	28.1
75	1	ใช่	25	697.9	19.4	108.5	27.1
75	1	ใช่	50	693.2	19.3	109.3	27.3
75	1	ใช่	75	713.7	19.8	110.1	27.5
75	1	ใช่	99	708.2	19.7	111.1	27.8
99	1	ใช่	1	819.2	22.8	109.8	27.4
99	1	ใช่	25	690.0	19.2	109.1	27.3
99	1	ใช่	50	695.6	19.3	111.3	27.8
99	1	ใช่	75	718.7	20.0	112.2	28.1
99	1	ใช่	99	750.8	20.9	111.0	27.8
1	5	ใช่	1	759.5	21.1	106.3	26.6
1	5	ใช่	25	673.5	18.7	109.6	27.4
1	5	ใช่	50	678.1	18.8	105.1	26.3
1	5	ใช่	75	678.1	18.8	105.1	26.3
1	5	ใช่	99	678.1	18.8	105.1	26.3
25	5	ใช่	1	739.9	20.6	108.3	27.1
25	5	ใช่	25	481.8	13.4	109.6	27.4
25	5	ใช่	50	453.8	12.6	105.1	26.3
25	5	ใช่	75	453.8	12.6	105.1	26.3
25	5	ใช่	99	453.3	12.6	105.1	26.3

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
50	5	ใช่	1	727.5	20.2	109.0	27.2
50	5	ใช่	25	406.2	11.3	110.2	27.6
50	5	ใช่	50	375.1	10.4	113.1	28.3
50	5	ใช่	75	378.3	10.5	112.3	28.1
50	5	ใช่	99	378.3	10.5	112.3	28.1
75	5	ใช่	1	739.9	20.6	108.2	27.1
75	5	ใช่	25	413.9	11.5	112.3	28.1
75	5	ใช่	50	363.4	10.1	112.3	28.1
75	5	ใช่	75	349.5	9.7	113.8	28.4
75	5	ใช่	99	352.8	9.8	112.9	28.2
99	5	ใช่	1	739.9	20.6	110.9	27.7
99	5	ใช่	25	392.2	10.9	111.9	28.0
99	5	ใช่	50	356.9	9.9	113.5	28.4
99	5	ใช่	75	337.8	9.4	113.4	28.4
99	5	ใช่	99	349.8	9.7	113.9	28.5
1	10	ใช่	1	755.6	21.0	109.6	27.4
1	10	ใช่	25	669.0	18.6	106.7	26.7
1	10	ใช่	50	652.0	18.1	108.9	27.2
1	10	ใช่	75	652.0	18.1	108.9	27.2
1	10	ใช่	99	652.0	18.1	108.9	27.2
25	10	ใช่	1	736.2	20.4	111.8	27.9
25	10	ใช่	25	419.5	11.7	112.0	28.0
25	10	ใช่	50	423.1	11.7	109.1	27.3
25	10	ใช่	75	423.1	11.7	109.1	27.3
25	10	ใช่	99	423.1	11.7	109.1	27.3
50	10	ใช่	1	723.0	20.1	110.8	27.7
50	10	ใช่	25	377.0	10.5	111.3	27.8
50	10	ใช่	50	354.1	9.8	111.2	27.8
50	10	ใช่	75	354.6	9.8	112.0	28.0
50	10	ใช่	99	354.6	9.8	112.0	28.0
75	10	ใช่	1	740.7	20.6	110.7	27.7
75	10	ใช่	25	372.7	10.3	114.7	28.7
75	10	ใช่	50	338.0	9.4	113.5	28.4
75	10	ใช่	75	321.4	8.9	113.9	28.5
75	10	ใช่	99	320.1	8.9	113.8	28.4

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

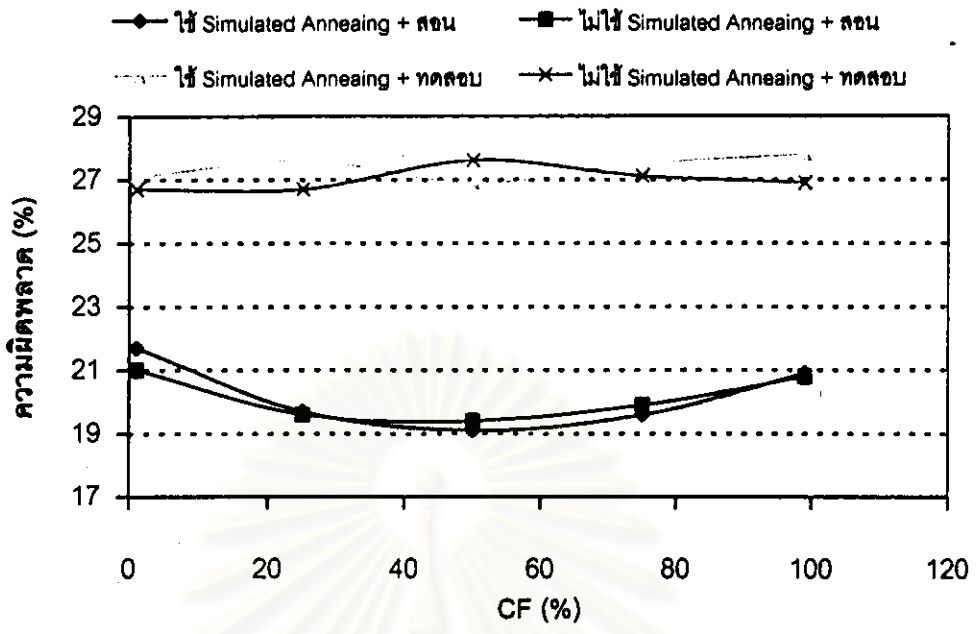
ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
99	10	ใช่	1	755.0	21.0	112.9	28.2
99	10	ใช่	25	386.7	10.8	113.5	28.4
99	10	ใช่	50	325.4	9.0	113.6	28.4
99	10	ใช่	75	321.9	8.9	114.9	28.7
99	10	ใช่	99	314.9	8.8	113.2	28.3
1	15	ใช่	1	753.9	20.9	107.8	26.9
1	15	ใช่	25	659.8	18.3	109.0	27.2
1	15	ใช่	50	641.5	17.8	107.5	26.9
1	15	ใช่	75	641.5	17.8	107.5	26.9
1	15	ใช่	99	641.5	17.8	107.5	26.9
25	15	ใช่	1	741.1	20.6	109.9	27.5
25	15	ใช่	25	404.9	11.2	110.3	27.6
25	15	ใช่	50	409.9	11.4	110.6	27.6
25	15	ใช่	75	409.9	11.4	110.6	27.6
25	15	ใช่	99	409.9	11.4	110.6	27.6
50	15	ใช่	1	752.6	20.9	113.0	28.3
50	15	ใช่	25	370.6	10.3	110.5	27.6
50	15	ใช่	50	351.2	9.8	112.9	28.2
50	15	ใช่	75	349.5	9.7	113.8	28.5
50	15	ใช่	99	349.5	9.7	113.8	28.5
75	15	ใช่	1	745.8	20.7	109.8	27.4
75	15	ใช่	25	371.3	10.3	112.9	28.2
75	15	ใช่	50	331.2	9.2	113.2	28.3
75	15	ใช่	75	306.1	8.5	115.6	28.9
75	15	ใช่	99	341.5	8.7	114.6	28.6
99	15	ใช่	1	730.7	20.3	110.3	27.6
99	15	ใช่	25	398.5	11.1	114.2	28.5
99	15	ใช่	50	317.4	8.8	116.0	29.0
99	15	ใช่	75	309.3	8.6	112.2	28.1
99	15	ใช่	99	319.3	8.9	112.6	28.1
1	20	ใช่	1	758.2	21.1	108.4	27.1
1	20	ใช่	25	645.7	17.9	108.1	27.0
1	20	ใช่	50	652.8	18.1	107.9	27.0
1	20	ใช่	75	652.8	18.1	107.9	27.0
1	20	ใช่	99	652.8	18.1	107.9	27.0

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

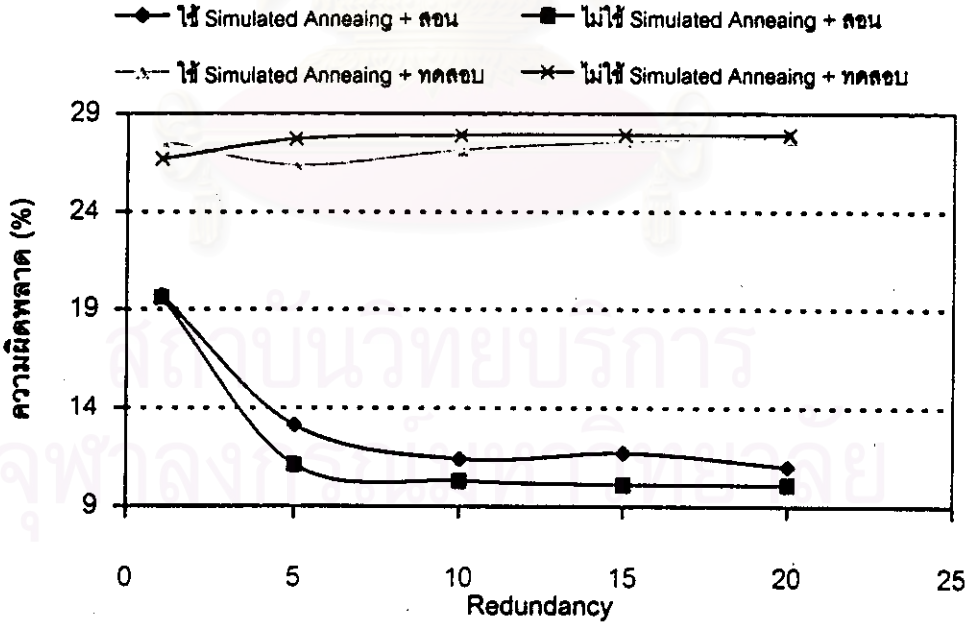
ค่าตัวเลือก				ทดสอบกับตัวอย่างสอน		ทดสอบกับตัวอย่างทดสอบ	
CF (%)	ค่าความซ้ำ	Simulated annealing	Significance (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)	จำนวนตัวอย่างผิดพลาด	ความผิดพลาด (%)
25	20	ใช่	1	732.9	20.3	111.3	27.8
25	20	ใช่	25	414.2	11.5	107.5	26.9
25	20	ใช่	50	401.7	11.2	112.5	28.1
25	20	ใช่	75	401.7	11.2	112.5	28.1
25	20	ใช่	99	401.7	11.2	112.5	28.1
50	20	ใช่	1	734.8	20.4	112.8	28.2
50	20	ใช่	25	357.0	9.9	111.1	27.8
50	20	ใช่	50	348.0	9.7	113.6	28.4
50	20	ใช่	75	350.0	9.7	112.0	28.0
50	20	ใช่	99	350.0	9.7	112.0	28.0
75	20	ใช่	1	726.9	20.2	109.3	27.3
75	20	ใช่	25	378.6	10.5	113.3	28.3
75	20	ใช่	50	327.7	9.1	113.4	28.4
75	20	ใช่	75	308.3	8.6	113.7	28.4
75	20	ใช่	99	305.7	8.5	115.9	29.0
99	20	ใช่	1	746.5	20.7	110.8	27.7
99	20	ใช่	25	387.3	10.8	114.8	28.7
99	20	ใช่	50	321.4	8.9	114.7	28.7
99	20	ใช่	75	310.8	8.6	112.9	28.2
99	20	ใช่	99	314.7	8.7	112.5	28.1

ตารางที่ 4.9 แสดงผลการทดลองสร้างกฎการตัดสินใจ ในแต่ละค่าตัวเลือกของข้อมูลการอนุมัติสินเชื่อเมื่อทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

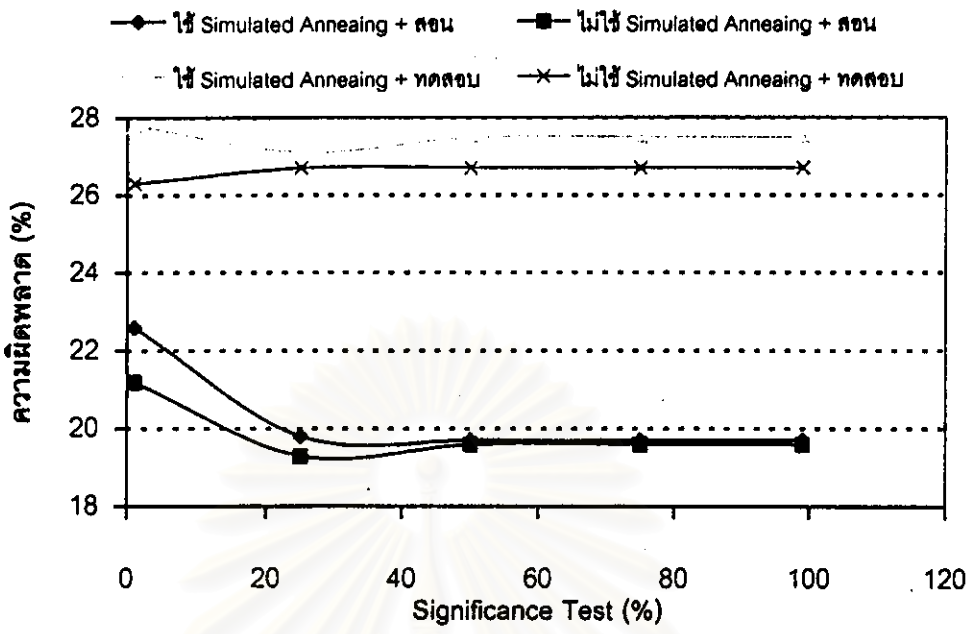
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.32 กราฟแสดงอัตราส่วนความผิดพลาดของกฎการตัดสินใจเมื่อตั้งค่า CF ที่ระดับต่าง ๆ และค่า Redundancy เท่ากับ 1 และไม่มีการทดสอบนัยสำคัญ กับการใช้และไม่ใช้วิธี Simulated Annealing โดยทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ



รูปที่ 4.33 กราฟแสดงอัตราส่วนความผิดพลาดของกฎการตัดสินใจเมื่อตั้งค่า Redundancy ที่ระดับต่าง ๆ และค่า CF เท่ากับ 25 % และไม่มีการทดสอบนัยสำคัญ กับการใช้และไม่ใช้วิธี Simulated Annealing โดยทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ



รูปที่ 4.34 กราฟแสดงอัตราส่วนความผิดพลาดของกฎการตัดสินใจเมื่อตั้งค่า Significance Test ที่ระดับต่าง ๆ และค่า CF เท่ากับ 25 % และค่า Redundancy เท่ากับ 1 กับการใช้และไม่ใช้วิธี Simulated Annealing โดยทดสอบบนตัวอย่างสอนและตัวอย่างทดสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย