

บทที่ 4

ผลดำเนินการวิจัย

ในบทนี้จะวิเคราะห์และสรุปผลการศึกษาทั้งหมดได้แก่ Association Rule และ Classification

4.1 การวิเคราะห์และสรุปผล

ในการศึกษาการใช้เทคนิค Association Rule เพื่อหาความสัมพันธ์ของโรคกับโรค มีทั้งหมด 6 โรคได้แก่ โรคเบาหวาน (p1) โรคความดันโลหิตสูง (p2) โรคตับ (p3) โรคอัมพาต (p4) โรคหัวใจ (p5) และ โรคไขมันในเลือดผิดปกติ (p6) พบว่าข้อมูลที่มี มีจำนวนผู้ป่วยเป็นโรคดังกล่าวจำนวนน้อย คือ 192 เรคอร์ด ดังนั้นในการทดลองจะกำหนดค่าสนับสนุนขั้นต่ำไว้ค่อนข้างต่ำเพื่อให้โอกาสในเกิดกฎเกณฑ์ที่น่าสนใจเพิ่มมากขึ้น โดยค่าที่เป็นไปได้ของโรคต่างๆเหล่านี้มี 2 ค่า คือ 1 หมายถึงมีประวัติการเป็นโรค และ 0 หมายถึง ไม่มีประวัติการเป็นโรค

4.1.1 ผลการหาความสัมพันธ์แบบ Apriori

จากการทดลองพบว่าเทคนิค Apriori ด้วยค่าสนับสนุนขั้นต่ำของกฎ 0.01 และค่า lift ที่ 0.9 ให้ผลลัพธ์กฎความสัมพันธ์ทั้งหมด 3,573 กฎ และผลลัพธ์ที่ได้ส่วนใหญ่จะเป็นกฎที่มีค่าเป็นปฏิเสธ (ไม่มีประวัติการเป็นโรค) ผสมอยู่ในกฎด้วย ซึ่งกฎในลักษณะดังกล่าวจะไม่ค่อยมีประโยชน์มากนักเนื่องจากคนที่มีความสุขภาพปกติทั่วไปก็จะเป็นโรคต่างๆอยู่แล้ว ยกตัวอย่างเช่น $(p2=1 \ p5=0 \implies p3=0 \ 138)$ หมายถึง ถ้ามีประวัติการเป็นโรคความดันโลหิตสูง (p2)และไม่มีประวัติการเป็นโรคอัมพาต (p5) จะไม่เป็นโรคตับ (p3) ซึ่งเกิดขึ้นจำนวน 138 เรคอร์ด เป็นต้น กฎที่ไม่มีค่าเป็นปฏิเสธผสมอยู่จะมีจำนวนน้อยมาก ในขณะที่อัลกอริทึม FP Growth จะได้กฎที่แสดงเฉพาะค่า positive (มีประวัติการเป็นโรค) เท่านั้นและมีจำนวนค่อนข้างมาก ภาพที่ 4.1 แสดงผลลัพธ์บางส่วนที่ได้จากการหาความสัมพันธ์แบบ Apriori

Best rules found:

```

1. p2=1 p4=0 149 ==> p1=0 p3=0 p5=0 p6=0 80   conf:(0.54) < lift:(1.27)> lev:(0.09) [17] conv:(1.23)
2. p1=0 p3=0 p5=0 p6=0 81 ==> p2=1 p4=0 80   conf:(0.99) < lift:(1.27)> lev:(0.09) [17] conv:(9.07)
3. p2=1 151 ==> p1=0 p3=0 p5=0 p6=0 81   conf:(0.54) < lift:(1.27)> lev:(0.09) [17] conv:(1.23)
4. p1=0 p3=0 p5=0 p6=0 81 ==> p2=1 81   conf:(1) < lift:(1.27)> lev:(0.09) [17] conv:(17.3)
5. p2=1 151 ==> p1=0 p3=0 p4=0 p5=0 p6=0 80   conf:(0.53) < lift:(1.27)> lev:(0.09) [17] conv:(1.22)
6. p1=0 p3=0 p4=0 p5=0 p6=0 80 ==> p2=1 80   conf:(1) < lift:(1.27)> lev:(0.09) [17] conv:(17.08)
7. p2=1 p5=0 138 ==> p1=0 p3=0 p4=0 p6=0 80   conf:(0.58) < lift:(1.22)> lev:(0.08) [14] conv:(1.23)
8. p1=0 p3=0 p4=0 p6=0 91 ==> p2=1 p5=0 80   conf:(0.88) < lift:(1.22)> lev:(0.08) [14] conv:(2.13)
9. p1=0 p3=0 p4=0 120 ==> p6=1 29   conf:(0.24) < lift:(1.22)> lev:(0.03) [5] conv:(1.05)
10. p6=1 38 ==> p1=0 p3=0 p4=0 29   conf:(0.76) < lift:(1.22)> lev:(0.03) [5] conv:(1.43)
11. p2=1 p5=0 138 ==> p1=0 p3=0 p6=0 81   conf:(0.59) < lift:(1.21)> lev:(0.07) [14] conv:(1.23)
12. p1=0 p3=0 p6=0 93 ==> p2=1 p5=0 81   conf:(0.87) < lift:(1.21)> lev:(0.07) [14] conv:(2.01)
13. p2=1 p4=0 p5=0 137 ==> p1=0 p3=0 p6=0 80   conf:(0.58) < lift:(1.21)> lev:(0.07) [13] conv:(1.22)
14. p1=0 p3=0 p6=0 93 ==> p2=1 p4=0 p5=0 80   conf:(0.86) < lift:(1.21)> lev:(0.07) [13] conv:(1.9)
15. p1=0 p3=0 122 ==> p6=1 29   conf:(0.24) < lift:(1.2)> lev:(0.03) [4] conv:(1.04)
16. p6=1 38 ==> p1=0 p3=0 29   conf:(0.76) < lift:(1.2)> lev:(0.03) [4] conv:(1.39)
17. p1=0 p3=0 122 ==> p4=0 p6=1 29   conf:(0.24) < lift:(1.2)> lev:(0.03) [4] conv:(1.04)
18. p1=0 p6=1 38 ==> p1=0 p3=0 29   conf:(0.76) < lift:(1.2)> lev:(0.03) [4] conv:(1.39)
19. p2=1 p3=0 p4=0 148 ==> p1=0 p5=0 p6=0 80   conf:(0.54) < lift:(1.19)> lev:(0.07) [12] conv:(1.17)
20. p1=0 p5=0 p6=0 87 ==> p2=1 p3=0 p4=0 80   conf:(0.92) < lift:(1.19)> lev:(0.07) [12] conv:(2.49)
21. p2=1 p3=0 150 ==> p1=0 p5=0 p6=0 81   conf:(0.54) < lift:(1.19)> lev:(0.07) [13] conv:(1.17)
22. p1=0 p5=0 p6=0 87 ==> p2=1 p3=0 81   conf:(0.93) < lift:(1.19)> lev:(0.07) [13] conv:(2.72)
23. p2=1 p3=0 150 ==> p1=0 p4=0 p5=0 p6=0 80   conf:(0.53) < lift:(1.19)> lev:(0.07) [12] conv:(1.17)
24. p1=0 p4=0 p5=0 p6=0 86 ==> p2=1 p3=0 80   conf:(0.93) < lift:(1.19)> lev:(0.07) [12] conv:(2.69)
25. p2=1 p4=0 149 ==> p1=0 p5=0 p6=0 80   conf:(0.54) < lift:(1.18)> lev:(0.07) [12] conv:(1.16)
26. p1=0 p5=0 p6=0 87 ==> p2=1 p4=0 80   conf:(0.92) < lift:(1.18)> lev:(0.07) [12] conv:(2.44)
27. p2=1 151 ==> p1=0 p5=0 p6=0 81   conf:(0.54) < lift:(1.18)> lev:(0.07) [12] conv:(1.16)
28. p1=0 p5=0 p6=0 87 ==> p2=1 81   conf:(0.93) < lift:(1.18)> lev:(0.07) [12] conv:(2.65)
29. p2=1 151 ==> p1=0 p4=0 p5=0 p6=0 80   conf:(0.53) < lift:(1.18)> lev:(0.06) [12] conv:(1.16)
30. p1=0 p4=0 p5=0 p6=0 86 ==> p2=1 80   conf:(0.93) < lift:(1.18)> lev:(0.06) [12] conv:(2.62)
31. p1=0 p3=0 p5=0 106 ==> p2=1 p4=0 p6=0 80   conf:(0.75) < lift:(1.18)> lev:(0.06) [12] conv:(1.41)
32. p2=1 p4=0 p6=0 123 ==> p1=0 p3=0 p5=0 80   conf:(0.65) < lift:(1.18)> lev:(0.06) [12] conv:(1.25)
33. p1=0 p3=0 p5=0 106 ==> p2=1 p6=0 81   conf:(0.76) < lift:(1.17)> lev:(0.06) [11] conv:(1.42)
34. p2=1 p6=0 125 ==> p1=0 p3=0 p5=0 81   conf:(0.65) < lift:(1.17)> lev:(0.06) [11] conv:(1.24)
35. p1=0 p3=0 p4=0 p5=0 105 ==> p2=1 p6=0 80   conf:(0.76) < lift:(1.17)> lev:(0.06) [11] conv:(1.41)
36. p2=1 p6=0 125 ==> p1=0 p3=0 p4=0 p5=0 80   conf:(0.64) < lift:(1.17)> lev:(0.06) [11] conv:(1.23)
37. p1=0 p4=0 126 ==> p6=1 29   conf:(0.23) < lift:(1.16)> lev:(0.02) [4] conv:(1.03)

```

ภาพที่ 4.1 ผลลัพธ์บางส่วนจากการหาความสัมพันธ์แบบ Apriori

4.1.2 ผลการหาความสัมพันธ์แบบ FP Growth

จากการทดลองพบว่าเทคนิค FP Growth ด้วยค่าสนับสนุนขั้นต่ำของกฎ 1 เปอร์เซ็นต์ และค่า lift ที่ 0.9 ให้ผลลัพธ์กฎความสัมพันธ์ทั้งหมด 18 กฎ และกฎที่ได้จะเป็นความสัมพันธ์ของการเป็นโรคต่างๆ (ไม่มีค่าที่เป็นปฏิเสธ) ภาพที่ 4.2 แสดงผลลัพธ์บางส่วนที่ได้จากการหาความสัมพันธ์แบบ FP Growth

Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 1000 -T 1 -C 0.9 -D 0.05 -U 1.0 -M 0.01
 Relation: associate bureereum
 Instances: 192
 Attributes: 6

=== Associator model (full training set) ===

FPGrowth found 18 rules (displaying top 18)

1. [p6=1]: 38 ==> [p2=1, p5=1]: 5 conf:(0.13) <lift:(1.94)> lev:(0.01) conv:(1.04)
2. [p2=1, p5=1]: 13 ==> [p6=1]: 5 conf:(0.38) <lift:(1.94)> lev:(0.01) conv:(1.16)
3. [p2=1, p6=1]: 26 ==> [p5=1]: 5 conf:(0.19) <lift:(1.76)> lev:(0.01) conv:(1.05)
4. [p5=1]: 21 ==> [p2=1, p6=1]: 5 conf:(0.24) <lift:(1.76)> lev:(0.01) conv:(1.07)
5. [p2=1]: 151 ==> [p1=1, p5=1]: 5 conf:(0.03) <lift:(1.27)> lev:(0.01) conv:(1)
6. [p2=1]: 151 ==> [p6=1, p5=1]: 5 conf:(0.03) <lift:(1.27)> lev:(0.01) conv:(1)
7. [p1=1, p5=1]: 5 ==> [p2=1]: 5 conf:(1) <lift:(1.27)> lev:(0.01) conv:(1.07)
8. [p6=1, p5=1]: 5 ==> [p2=1]: 5 conf:(1) <lift:(1.27)> lev:(0.01) conv:(1.07)
9. [p6=1]: 38 ==> [p5=1]: 5 conf:(0.13) <lift:(1.2)> lev:(0) conv:(1)
10. [p5=1]: 21 ==> [p6=1]: 5 conf:(0.24) <lift:(1.2)> lev:(0) conv:(0.99)
11. [p1=1]: 64 ==> [p2=1, p5=1]: 5 conf:(0.08) <lift:(1.15)> lev:(0) conv:(0.99)
12. [p2=1, p5=1]: 13 ==> [p1=1]: 5 conf:(0.38) <lift:(1.15)> lev:(0) conv:(0.96)
13. [p2=1]: 151 ==> [p1=1, p6=1]: 7 conf:(0.05) <lift:(0.99)> lev:(0) conv:(0.99)
14. [p1=1, p6=1]: 9 ==> [p2=1]: 7 conf:(0.78) <lift:(0.99)> lev:(0) conv:(0.64)
15. [p2=1, p1=1]: 47 ==> [p5=1]: 5 conf:(0.11) <lift:(0.97)> lev:(0) conv:(0.97)
16. [p5=1]: 21 ==> [p2=1, p1=1]: 5 conf:(0.24) <lift:(0.97)> lev:(0) conv:(0.93)
17. [p2=1]: 151 ==> [p1=1]: 47 conf:(0.31) <lift:(0.93)> lev:(-0.02) conv:(0.96)
18. [p1=1]: 64 ==> [p2=1]: 47 conf:(0.73) <lift:(0.93)> lev:(-0.02) conv:(0.76)

ภาพที่ 4.2 ผลลัพธ์บางส่วนจากการหากฎความสัมพันธ์แบบ FP Growth

วิเคราะห์ผลลัพธ์ที่ได้โดยพิจารณาจากค่า lift พบกลุ่มโรคที่มีความสัมพันธ์กัน 2 กลุ่ม กลุ่มแรกคือกฎที่ 1-4, 6, 8 คือโรคความดันโลหิตสูง (p2) โรคหัวใจ (p5) และโรคไขมันในเลือดผิดปกติ(p6) กลุ่มที่สองคือกฎที่ 5, 7, 11, 12 คือ โรคเบาหวาน (p1) โรคความดันโลหิตสูง (p2) และโรคหัวใจ (p5) จากความสัมพันธ์นี้ทำให้ทราบว่าเมื่อเป็นโรสดังกล่าวแล้วมักจะมีโรคแทรกซ้อนตามมา เช่น ถ้าคนไข้เป็นโรคหัวใจและโรคความดันโลหิตสูงอยู่แล้ว ก็ต้องระวังจะเป็นโรคเบาหวานและโรคไขมันในเลือดผิดปกติซึ่งสามารถประยุกต์ใช้ความรู้นี้ในการประชาสัมพันธ์ผู้ป่วยเพื่อป้องกันการเกิดโรคแทรกซ้อน เช่น พยายามลดพฤติกรรมเสี่ยงที่จะทำให้เกิดโรคเบาหวานและไขมันในเลือดผิดปกติเสีย เพื่อลดความเสี่ยงที่จะเป็นโรสดังกล่าว

Classification จะเป็นการจำแนกกลุ่มประชากรออกเป็น กลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย โดยเทคนิคที่ใช้ในการศึกษาคั้งนี้คือ Decion Tree: C4.5 และ Decision Rule: Partial Rules แล้วเลือกเทคนิคที่มีค่าความถูกต้องสูงที่สุดนำไปพัฒนาระบบต่อไป

4.1.3 ผลการจำแนกกลุ่มแบบ Decision Tree: C4.5

จากการทดลองพบว่าเทคนิค Decision Tree: C4.5 ที่กำหนดค่า confidenceFactor = 0.25 ให้ผลลัพธ์การจำแนกกลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วยมีความถูกต้อง (Precision) เท่ากับ 87.70% ค่าระลึก (Recall) เท่ากับ 88.70% และ ค่าความเหวี่ยง (F-measure) เท่ากับ 88.10% ภาพที่ 4.3 แสดงผลลัพธ์บางส่วนที่ได้จากการจำแนกกลุ่มแบบ Decision Tree: C4.5 ซึ่งจะแสดงผลเป็นกฎในลักษณะของต้นไม้การตัดสินใจที่มีกิ่งต้นไม้จำนวนมาก

```

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: data7-weka.filters.unsupervised.attribute.Remove-R4-5
Instances: 1071
Attributes: 81

Test mode:10-fold cross-validation

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

J48 pruned tree
-----

treat = 1
| p15 = 1
| | test cervic = 1: 2 (3.0)
| | test cervic = 2: 3 (0.0)
| | test cervic = 3: 3 (13.0)
| p15 = 2: 3 (196.0/1.0)
| p15 = 3: 3 (0.0)
treat = 2: 2 (0.0)
treat = 3: 2 (0.0)
treat = 4
| sumBMI = 1
| | fb9 = 0: 2 (11.0/1.0)
| | | fb9 = 1
| | | | drink = 1: 2 (8.0)
| | | | drink = 2
| | | | | b20 = 0: 1 (3.0)
| | | | | b20 = 3: 1 (2.0/1.0)
| | | | | b20 = 5
| | | | | cervic = 1: 2 (5.0)
| | | | | cervic = 2: 2 (0.0)
| | | | | cervic = 3: 2 (0.0)
| | | | | cervic = 4: 2 (0.0)
| | | | | cervic = 5
| | | | | fa8 = 0
| | | | | | marry status = 1
| | | | | | | fa1 = 0
| | | | | | | | eat3 = 0: 2 (19.0/4.0)
| | | | | | | | eat3 = 1: 1 (3.0/1.0)
| | | | | | | | fa1 = 1: 1 (2.0)
| | | | | | | | marry status = 2: 1 (5.0)
| | | | | | | | marry status = 3: 2 (1.0)
| | | | | | | | fa8 = 1: 2 (2.0)
| | | | | | | | drink = 3: 2 (1.0)
| | | | | | | | sumBMI = 2
| | | | | | | | | eat5 = 0
| | | | | | | | | fa6 = 0
| | | | | | | | | | b10 = 0: 2 (7.0)

```

ภาพที่ 4.3 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Decision Tree: C4.5

ผลลัพธ์ที่ได้จะแสดงเงื่อนไขของกฎ โดยค่าของข้อมูลของแอทริบิวต์ที่มีค่าต่อเนื่องจะใช้เครื่องหมาย “<” “<=” และ “>” “>=” ในการพิจารณา ส่วนข้อมูลที่เป็นค่าไม่ต่อเนื่องจะใช้เครื่องหมาย “=” และใช้เงื่อนไข “AND” ในการเชื่อมโยงระหว่างแอทริบิวต์โดยจะเชื่อมโยงไปจนถึงกลุ่มที่ต้องการจำแนก และจะมีข้อมูลความถูกต้องของกฎ ซึ่งได้แก่จำนวนเรคอร์ดที่สอดคล้องกับกฎและจำนวนเรคอร์ดที่ไม่สอดคล้องกับกฎ เช่น

treat = 4

sumBMI = 1

fb9 = 1

drink = 1: 2 (8.0) หมายถึง หากบุคคลนั้น ไม่มีประวัติการเจ็บป่วย (treat = 4) และมีค่าดัชนีมวลกาย = 1 (น้อยกว่า 18.5: ผอม) และพี่น้องสายตรงไม่เคยมีประวัติเจ็บป่วย (fb9 = 1) และดื่มเครื่องดื่มที่มีแอลกอฮอล์ (drink = 1) จะมีสุขภาพอยู่ในกลุ่มเสี่ยง (2) โดยมีข้อมูลที่สอดคล้องกับกฎนี้จำนวน 8 ตัวอย่าง และไม่มีข้อมูลที่ขัดแย้งกับกฎนี้

4.1.4 ผลการจำแนกกลุ่มแบบ Decision Tree: Partial Rules

เทคนิค Decision Tree: Partial Rules โดยกำหนดค่า Confidence Factor = 0.25 ให้ผลลัพธ์การจำแนกกลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย มีความถูกต้อง (Precision) เท่ากับ 88.60% ค่าระลึก (Recall) เท่ากับ 89.20% และ ค่าความเหวี่ยง (F-measure) เท่ากับ 88.80% ภาพที่ 4.4 แสดงผลลัพธ์บางส่วนที่ได้จากการจำแนกกลุ่มแบบ Decision Tree: Partial Rules ซึ่งแสดงผลเป็นกฎไม่ได้อยู่ในรูปแบบต้นไม้ที่มีขนาดใหญ่ ทำให้เข้าใจง่าย

```
Scheme:weka.classifiers.rules.PART -M 2 -C 0.25 -Q 1
Relation: data7-weka.filters.unsupervised.attribute.Remove-R4-5
Instances: 1071
Attributes: 81
```

Test mode:10-fold cross-validation

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

PART decision list

```
treat = 1 AND
p15 = 2: 3 (196.0/1.0)
```

```
treat = 4 AND
sumBMI = 3: 2 (253.0)
```

```
treat = 4 AND
sumhip = 2 AND
fa6 = 0 AND
sumBMI = 4: 2 (130.0)
```

```
treat = 1 AND
test cervic = 3: 3 (13.0)
sumBMI = 4: 2 (48.0)
```

```
p11 = 2 AND
sumhip = 2 AND
sumBMI = 5: 2 (32.0)
```

```
p11 = 2 AND
p1 = 3: 2 (33.0/2.0)
```

```
p11 = 2 AND
eat5 = 0 AND
fa6 = 0 AND
b10 = 3 AND
education = 2: 2 (31.0)
```

```
p11 = 2 AND
smoke = 1 AND
b11 = 5: 2 (12.0)
```

```
p11 = 2 AND
sumhip = 2 AND
sex = 4: 2 (19.0)
```

ภาพที่ 4.4 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Decision Rule: Partial Rules

ตัวอย่างเช่น $treat = 4$ AND $sumBMI = 3$: 2 (253.0)

หมายถึง หากบุคคลไม่มีประวัติการเจ็บป่วย ($treat = 4$) และมีค่าดัชนีมวลกาย = 3 (23-24.99: ท้วม) จะมีสุขภาพอยู่ในกลุ่มเสี่ยง (2) และมีข้อมูลที่สอดคล้องกับกฎนี้หรือให้ผลลัพธ์ตรงกันจำนวน 253 ตัวอย่าง

4.1.5 การวิเคราะห์และการคัดเลือกกฎ

จากผลการทดลองพบว่า การจำแนกกลุ่มทั้ง 2 แบบ ให้ผลลัพธ์ที่มีค่าใกล้เคียงกันดังตารางที่ 4.1 ดังนั้นจึงเลือกผลลัพธ์ที่ได้จากเทคนิค Decision Tree: Partial rules นำไปพัฒนาระบบเนื่องจากมีค่า F-measure สูงกว่าอีกวิธีหนึ่ง โดยมีจำนวนกฎทั้งสิ้น 39 กฎ เช่น $p11 = 2$ AND $sumhip = 2$ AND $sumBMI = 5$: 2 (32.0) หมายถึง หากบุคคลนั้นไม่เคยอาการกินจุแต่พอมลง และมีขนาดรอบเอวมากกว่า 90 เซนติเมตร (ชาย) หรือมากกว่า 80 เซนติเมตร(หญิง) และมีค่าดัชนีมวลกายมากกว่าหรือเท่ากับ 30 (อ้วนมาก) จะมีสุขภาพอยู่ในกลุ่มเสี่ยง เป็นต้น

ตารางที่ 4.1 เปรียบเทียบผลการทดสอบอัลกอริทึม C4.5 และ Partial Rules

เทคนิค	Precision	Recall	F-measure
C4.5	87.70%	88.70%	88.10%
Partial Rules	88.60%	89.20%	88.80%

โดยเมื่อนำตัวอย่างที่โมเดลจำแนกกลุ่มผิดพลาดมาวิเคราะห์พบว่าข้อมูลที่อยู่ในกลุ่มเสี่ยงแต่โมเดลทำนายว่าเป็นกลุ่มป่วยมีจำนวน 1 ตัวอย่างดังภาพที่ 4.5 จากกฎ $treat = 1$ AND $p15 = 2$: 3 (196.0/1.0) ซึ่งกฎที่เกิดขึ้นจากมีข้อมูลที่เหมือนกันจำนวนมากที่ผู้ป่วยตาไม่พรั่มัว ($p15 = 2$) และอยู่ในกลุ่มป่วย เมื่อพิจารณาข้อมูลตัวอย่างรายนี้พบว่า มีลักษณะค่าเอทริบิวต์บางตัวที่เหมือนกับคนกลุ่มใหญ่ที่ป่วยทำให้ถูกมองว่าป่วยไปด้วย ซึ่งข้อมูลนี้อาจถือได้ว่าเป็นข้อมูลที่ผิดปกติ (outlier) เนื่องจากมีคุณสมบัติแตกต่างไปจากกลุ่มคนทั่วไป

```

Time taken to build model: 0.16 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      955      89.169 %
Incorrectly Classified Instances    116      10.831 %
Kappa statistic                     0.7578
Mean absolute error                 0.0791
Root mean squared error             0.2526
Relative absolute error             26.0248 %
Root relative squared error        64.8507 %
Total Number of Instances          1071

=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
0.407    0.052    0.468     0.407   0.436     0.8       1
0.932    0.206    0.915     0.932   0.924     0.9       2
0.995    0.001    0.995     0.995   0.995     0.997     3
Weighted Avg. 0.892  0.15     0.886     0.892   0.888     0.908

=== Confusion Matrix ===
 a  b  c  <-- classified as
44 64 0 | a = 1
50 704 1 | b = 2
0 1 207 | c = 3

PART decision list
-----
treat = 1 AND
p15 = 2: 3 (196.0/1.0)

treat = 4 AND
sumBMI = 3: 2 (253.0)

treat = 4 AND
sumhip = 2 AND
fa6 = 0 AND
sumBMI = 4: 2 (130.0)

treat = 1 AND
test cervic = 3: 3 (13.0)

sumBMI = 4: 2 (48.0)

p11 = 2 AND
sumhip = 2 AND
sumBMI = 5: 2 (32.0)

p11 = 2 AND
p1 = 3: 2 (33.0/2.0)

p11 = 2 AND
eat5 = 0 AND
fa6 = 0 AND
b10 = 3 AND
education = 2: 2 (31.0)

```

ภาพที่ 4.5 ข้อมูลความผิดพลาดที่ได้จากผลลัพธ์การจำแนกกลุ่มของ Decision Rule: Partial Rules

นอกจากนี้ยังพบว่าโมเดลจำแนกกลุ่ม Class 1 (กลุ่มปกติ) มีความผิดพลาดสูงคือ 50 จาก 94 ตัวอย่าง คิดเป็น 53.2% เนื่องจากมีจำนวนตัวอย่างให้ระบบเรียนรู้มีอยู่เพียง 108 ตัวอย่างเท่านั้น ทำให้มีค่า Precision กับ Recall ที่ต่ำคือ 46.8% และ 40.7% ตามลำดับ ดังภาพที่ 4.6 วิธีแก้ไขคือต้องใส่ข้อมูลที่เป็นกลุ่มปกติเข้าไปให้ระบบได้เรียนรู้เพิ่มอีก เพื่อที่จะได้รูปแบบการจำแนกกลุ่มที่ถูกต้องมากยิ่งขึ้น

```

=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
0.407    0.052    0.468     0.407   0.436     0.8       1
0.932    0.206    0.915     0.932   0.924     0.9       2
0.995    0.001    0.995     0.995   0.995     0.997     3
Weighted Avg. 0.892  0.15     0.886     0.892   0.888     0.908

=== Confusion Matrix ===
 a  b  c  <-- classified as
44 64 0 | a = 1
50 704 1 | b = 2
0 1 207 | c = 3

```

ภาพที่ 4.6 ข้อมูลความผิดพลาดของการจำแนกกลุ่มปกติ

เมื่อให้ผู้เชี่ยวชาญช่วยตรวจสอบผลลัพธ์ของกฎที่ได้พบว่ากฎส่วนใหญ่ที่พิจารณา แอทริบิวต์การปฏิบัติคนเมื่อเจ็บป่วยจะเป็นกฎที่ไม่ถูกต้องเนื่องจากแอทริบิวต์นี้เป็นแอทริบิวต์ที่มองภาพรวมของประวัติการเจ็บป่วยและอาการต่างๆทั้งหมด 16 แอทริบิวต์ แต่มีเพียง 6 แอทริบิวต์ที่ถ้าเป็นโรคแล้วผู้เชี่ยวชาญจะถือว่าเป็นกลุ่มป่วยทันที ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง โรคตับ โรคอัมพาต โรคหัวใจ และ โรคไขข้อในเลือดผิดปกติ ดังนั้นแอทริบิวต์การปฏิบัติคนเมื่อเจ็บป่วยจึงเป็นสาเหตุที่ทำให้เกิดการสร้างกฎที่ผิดพลาดจึงทำการทดลองตัดแอทริบิวต์การปฏิบัติคนเมื่อเจ็บป่วยออก และกำหนดค่า Confidence Factor = 1 พบว่าการจำแนกกลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วยให้ผลลัพธ์ที่ดีขึ้นโดยมีความถูกต้อง (Precision) เท่ากับ 89.20% ค่าระลึก (Recall) เท่ากับ 89.40% และ ค่าความเหวี่ยง (F-measure) เท่ากับ 89.30% ตามลำดับ ดังภาพที่ 4.7

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.463	0.052	0.5	0.463	0.481	0.766	1
	0.932	0.196	0.919	0.932	0.926	0.893	2
	0.981	0.001	0.995	0.981	0.988	0.99	3
Weighted Avg.	0.894	0.144	0.892	0.894	0.893	0.899	

=== Confusion Matrix ===

a	b	c	<-- classified as
50	58	0	a = 1
50	704	1	b = 2
0	4	204	c = 3

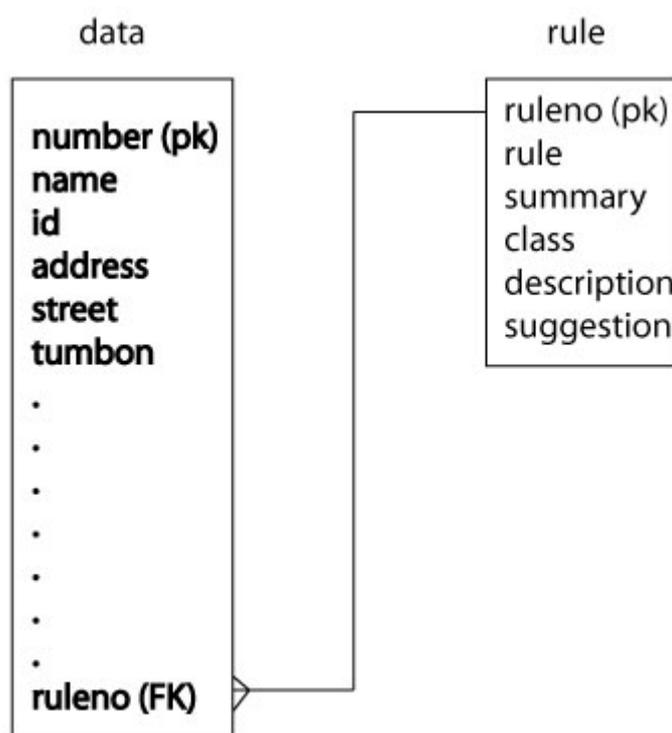
ภาพที่ 4.7 ผลลัพธ์การจำแนกกลุ่มหลังจากตัดแอทริบิวต์การปฏิบัติคนเมื่อเจ็บป่วย

สรุปกฎที่ได้จากผลการทดสอบการจำแนกกลุ่มข้อมูลแบบ Decision Rule: Partial Rules เป็นการจับกลุ่มที่เหมาะสมที่สุด ได้จำนวนกฎทั้งสิ้น 55 กฎ และเมื่อผู้เชี่ยวชาญได้ทำการตรวจสอบและลบบางกฎทิ้งไป คงเหลือจำนวนทั้งสิ้น 42 กฎ ผู้วิจัยจึงได้นำกฎที่ได้ไปพัฒนาเป็นระบบคัดกรองสุขภาพเบื้องต้นต่อไป

4.2 การพัฒนาระบบคัดกรองสุขภาพเบื้องต้น

ในขั้นตอนนี้เป็นการนำกฎที่ได้จากการทำเหมืองข้อมูลมาเป็นเงื่อนไขในการพัฒนาโปรแกรมระบบคัดกรองสุขภาพเบื้องต้นให้เป็นเครื่องมือในการคัดกรองสุขภาพประชาชน ภาษาที่ใช้ในการพัฒนาเป็นเป็นภาษา PHP และใช้ฐานข้อมูล MySQL

4.2.1 การออกแบบฐานข้อมูล ฐานข้อมูลสำหรับเก็บข้อมูลที่เกี่ยวข้องในระบบคัดกรองสุขภาพเบื้องต้น แสดงได้ในรูปแบบ ER-Diagram ดังภาพที่ 4.8 และรายละเอียดของแต่ละตารางแสดงในรูปแบบ Data Dictionary ดังภาพที่ 4.1 และ 4.2



ภาพที่ 4.8 ER-Diagram ของระบบคัดกรองสุขภาพเบื้องต้น

ตารางที่ 4.1 Data Dictionary ของตาราง data

ลำดับ	ชื่อฟิลด์	PK	ชนิดข้อมูล	คำอธิบาย
1	number	PK	int(10)	ใช้แสดงลำดับกำหนดเป็น auto_increment
2	name		varchar(30)	ชื่อ นามสกุล
3	id		varchar(13)	เลขบัตรประจำตัวประชาชน
4	address		varchar(50)	ที่อยู่

ตารางที่ 4.1 (ต่อ)

5	street		varchar(30)	ถนน
6	tumbon		varchar(20)	ตำบล
7	amphur		varchar(20)	อำเภอ
8	province		varchar(20)	จังหวัด
9	choomchon		varchar(20)	ชุมชน
10	telephone		varchar(15)	เบอร์โทรศัพท์
11	use		varchar(30)	สิทธิการรักษา
12	gender		char(1)	เพศ
13	sDay		varchar(2)	วันเกิด
14	sMonth		varchar(2)	เดือนเกิด
15	sYear		varchar(4)	ปีเกิด
16	status		char(1)	สถานภาพสมรส
17	education		varchar(20)	การศึกษา
18	weight		int(3)	น้ำหนัก
19	height		int(3)	ส่วนสูง
20	bmi		float(4,2)	ค่าดัชนีมวลกาย (ค่าต่อเนื่อง)
21	sumbmi		char(1)	ค่าดัชนีมวลกาย (ค่าไม่ต่อเนื่อง)
22	hip		int(3)	ขนาดรอบเอว (ค่าต่อเนื่อง)
23	sumhip		char(1)	ขนาดรอบเอว (ค่าไม่ต่อเนื่อง)
24	pressure1a		int(3)	ความดันตัวบนครั้งที่ 1
25	pressure1b		int(3)	ความดันตัวล่างครั้งที่ 1
26	pressure2a		int(3)	ความดันตัวบนครั้งที่ 2
27	pressure2b		int(3)	ความดันตัวล่างครั้งที่ 2
28	meanpa		int(3)	ค่าเฉลี่ยความดันตัวบน
29	meanpb		int(3)	ค่าเฉลี่ยความดันตัวล่าง
30	sumbp		char(1)	สรุปความดัน (ค่าไม่ต่อเนื่อง)
31	fa1		char(1)	ประวัติโรคเบาหวานของบิดามารดา
32	fa2		char(1)	ประวัติโรคความดันโลหิตสูงของบิดามารดา

ตารางที่ 4.1 (ต่อ)

33	fa3		char(1)	ประวัติโรคเก๊าท์ของบิดามารดา
34	fa4		char(1)	ประวัติโรคไตวายเรื้อรังของบิดามารดา
35	fa5		char(1)	ประวัติโรคกล้ามเนื้อหัวใจตายของบิดามารดา
36	fa6		char(1)	ประวัติโรคเส้นเลือดสมองของบิดามารดา
37	fa7		char(1)	ประวัติโรคถุงลมโป่งพองของบิดามารดา
38	fa8		char(1)	ประวัติโรคของบิดามารดา (ไม่ทราบ)
39	fa9		char(1)	ประวัติโรคของบิดามารดา (ไม่มี)
40	fa10		char(1)	ประวัติโรคของบิดามารดา (อื่นๆ)
41	fa100		varchar(20)	ประวัติโรคของบิดามารดา (อื่นๆ ระบุ)
42	fb1		char(1)	ประวัติโรคเบาหวานของพี่น้องสายตรง
43	fb2		char(1)	ประวัติโรคความดันโลหิตสูงของพี่น้องสายตรง
44	fb3		char(1)	ประวัติโรคเก๊าท์ของพี่น้องสายตรง
45	fb4		char(1)	ประวัติโรคไตวายเรื้อรังของพี่น้องสายตรง
46	fb5		char(1)	ประวัติโรคกล้ามเนื้อหัวใจตายของพี่น้องสายตรง
47	fb6		char(1)	ประวัติโรคเส้นเลือดสมองของพี่น้องสายตรง
48	fb7		char(1)	ประวัติโรคถุงลมโป่งพองของพี่น้องสายตรง
49	fb8		char(1)	ประวัติโรคของพี่น้องสายตรง (ไม่ทราบ)
50	fb9		char(1)	ประวัติโรคของพี่น้องสายตรง (ไม่มี)
51	fb10		char(1)	ประวัติโรคของพี่น้องสายตรง (อื่นๆ)
52	fb100		varchar(20)	ประวัติโรคของพี่น้องสายตรง (อื่นๆ ระบุ)
53	p1		char(1)	ประวัติการเจ็บป่วยด้วยโรคเบาหวาน
54	p2		char(1)	ประวัติการเจ็บป่วยด้วยโรคความดันโลหิตสูง
55	p3		char(1)	ประวัติการเจ็บป่วยด้วยโรคตับ
56	p4		char(1)	ประวัติการเจ็บป่วยด้วยโรคอัมพาต
57	p5		char(1)	ประวัติการเจ็บป่วยด้วยโรคหัวใจ
58	p6		char(1)	ประวัติการเจ็บป่วยด้วยโรคไขมันในเลือดผิดปกติ
59	p7		char(1)	ประวัติการเป็นแผลที่เท้า/ตัดขา จากเบาหวาน
60	p8		char(1)	ประวัติการคลอดบุตรน้ำหนักเกิด 4 กิโลกรัม

ตารางที่ 4.1 (ต่อ)

61	p9		char(1)	ประวัติการค้ำน้ำบ่อยและมาก
62	p10		char(1)	ประวัติการปัสสาวะกลางคืน 3 ครั้งขึ้นไป
63	p11		char(1)	ประวัติการกินจุแต่พอมลง
64	p12		char(1)	ประวัติน้ำหนักลด / อ่อนเพลีย
65	p13		char(1)	ประวัติการเป็นแผลที่ริมฝีปากบ่อยและหายยาก
66	p14		char(1)	ประวัติการคันตามผิวหนังและอวัยวะสืบพันธุ์
67	p15		char(1)	ประวัติตาพร่ามัว ต้องเปลี่ยนแว่นบ่อย
68	p16		char(1)	ประวัติการชาปลายมือปลายเท้าโดยไม่ทราบสาเหตุ
69	treat		char(1)	การปฏิบัติตนเมื่อเจ็บป่วย
70	smoke		char(1)	การสูบบุหรี่
71	drink		char(1)	การดื่มเครื่องดื่มที่มีแอลกอฮอล์
72	exercise		char(1)	การออกกำลังกาย
73	eat1		char(1)	ชอบรสหวาน
74	eat2		char(1)	ชอบรสเค็ม
75	eat3		char(1)	ชอบรสมัน
76	eat4		char(1)	ชอบรสเปรี้ยว
77	eat5		char(1)	ชอบรสจืด
78	eat6		char(1)	ไม่ชอบทุกรส
79	drive		char(1)	การขับขี่หรือโดยสาร
80	sex		char(1)	การใช้ถุงยางอนามัย
81	chest		char(1)	ความสามารถในการตรวจเต้านมด้วยตนเอง
82	levelchest		char(1)	ความถี่ในการตรวจเต้านมด้วยตนเอง
83	testchest		char(1)	ผลตรวจเต้านมด้วยตนเอง
84	cervic		char(1)	ความถี่ในการตรวจมะเร็งปากมดลูก
85	testcervic		char(1)	ผลตรวจมะเร็งปากมดลูก
86	b1		int(1)	กินอาหารครบ 5 หมู่
87	b2		int(1)	กินอาหารมือเช้าเป็นมือหลัก

ตารางที่ 4.1 (ต่อ)

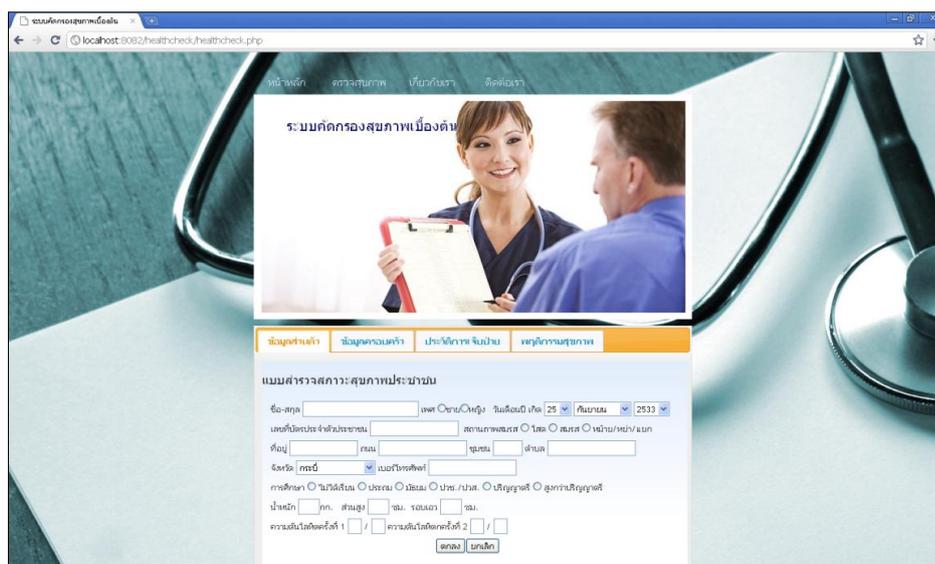
88	b3		int(1)	กินผักมากกว่าวันละ 3 ท็อปพี
89	b4		int(1)	กินผลไม้วันละ 2-3 ส่วน(1ส่วนเท่ากับ 6-8 คำ)
90	b5		int(1)	กินปลาอย่างน้อยวันละ 1 มื้อ
91	b6		int(1)	กินเนื้อสัตว์ไม่ติดมัน สัปดาห์ละ 2-3 มื้อ
92	b7		int(1)	ดื่มนมรสจืดหรือนมถั่วเหลืองผสมงาคาร์สจืดวันละ 1-2 แก้ว
93	b8		int(1)	กินอาหารมื้อเย็นห่างจากเวลานอนไม่น้อยกว่า 4 ชั่วโมง
94	b9		int(1)	กินอาหารประเภทต้ม นึ่ง ลวก อบ ยำหรือหมก
95	b10		int(1)	หลีกเลี่ยงอาหารไขมันสูง
96	b11		int(1)	หลีกเลี่ยงของหวานและขนมที่มีแป้งและน้ำตาลมาก
97	b12		int(1)	กินอาหารรสจืด
98	b13		int(1)	เลือกดื่มน้ำเปล่าแทนน้ำอัดลมหรือน้ำหวาน
99	b14		int(1)	หลีกเลี่ยงเครื่องดื่มที่มีแอลกอฮอล์
100	b15		int(1)	อารมณ์ดี ไม่เครียด
101	b16		int(1)	นอนหลับไม่น้อยกว่าวันละ 7-8 ชั่วโมง
102	b17		int(1)	ออกกำลังกายสัปดาห์ละ 5 วันหรือ สัปดาห์ละ 5 ครั้ง
103	b18		int(1)	ออกกำลังกายวันละ 30 นาที
104	b19		int(1)	ขณะออกกำลังกายหายใจเร็วกว่าปกติและเหงื่อซึม
105	b20		int(1)	ทุกครั้งวิ่งรอบเอดได้เกณฑ์ปกติ
106	sumb		int(3)	คะแนนรวมพฤติกรรมสุขภาพด้านอาหาร อารมณ์ การออกกำลังกาย (ค่าต่อเนื่อง)
107	resultb		char(1)	สรุปผลพฤติกรรมสุขภาพด้านอาหาร อารมณ์ การออกกำลังกาย (ค่าไม่ต่อเนื่อง)
108	summary		char(1)	สรุปผลการคัดกรองสุขภาพเบื้องต้น
109	ruleno	FK	char(3)	ลำดับของกฎ
110	datetime		datetime	วันที่ทำการบันทึก

ตารางที่ 4.2 Data Dictionary ของตาราง rule

ลำดับ	ชื่อฟิลด์	PK	ชนิดข้อมูล	คำอธิบาย
1	rueno	PK	int(3)	ใช้แสดงลำดับของกฎ
2	rule		varchar(200)	ใช้บันทึกกฎ
3	summary		char(1)	สรุปผลการคัดกรอง (1,2,3)
4	class		varchar(15)	สรุปผลการคัดกรอง(กลุ่มปกติ,กลุ่มป่วย,กลุ่มเสี่ยง)
5	description		varchar(100)	รายละเอียดของกฎ
6	suggestion		varchar(100)	คำแนะนำในการปฏิบัติตน

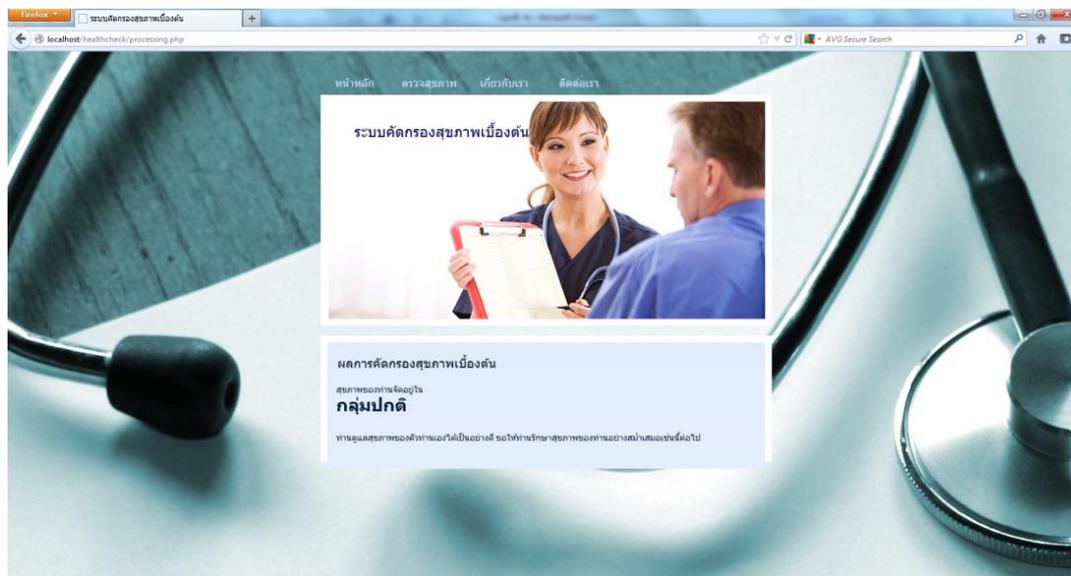
4.2.2 การออกแบบหน้านำเข้าข้อมูล

หน้าจอหลักประกอบด้วยเมนู หน้าหลักสำหรับแนะนำให้ผู้ใช้รู้จักระบบคัดกรองสุขภาพเบื้องต้น วิธีการใช้งานและประโยชน์ที่จะได้รับ หน้าตรวจสอบสุขภาพสำหรับให้ผู้ใช้บันทึกข้อมูลสุขภาพดังภาพที่ 4.10 เมื่อทำการบันทึกข้อมูลเสร็จระบบจะประมวลผลแล้วแจ้งผลการคัดกรองสุขภาพเบื้องต้นให้ทราบทันทีดังภาพที่ 4.11-4.14 หน้าเกี่ยวกับเราสำหรับให้ข้อมูลรายละเอียดต่างๆเกี่ยวกับที่มาที่ไปของระบบ ติดต่อเราสำหรับติดต่อผู้ดูแลระบบ

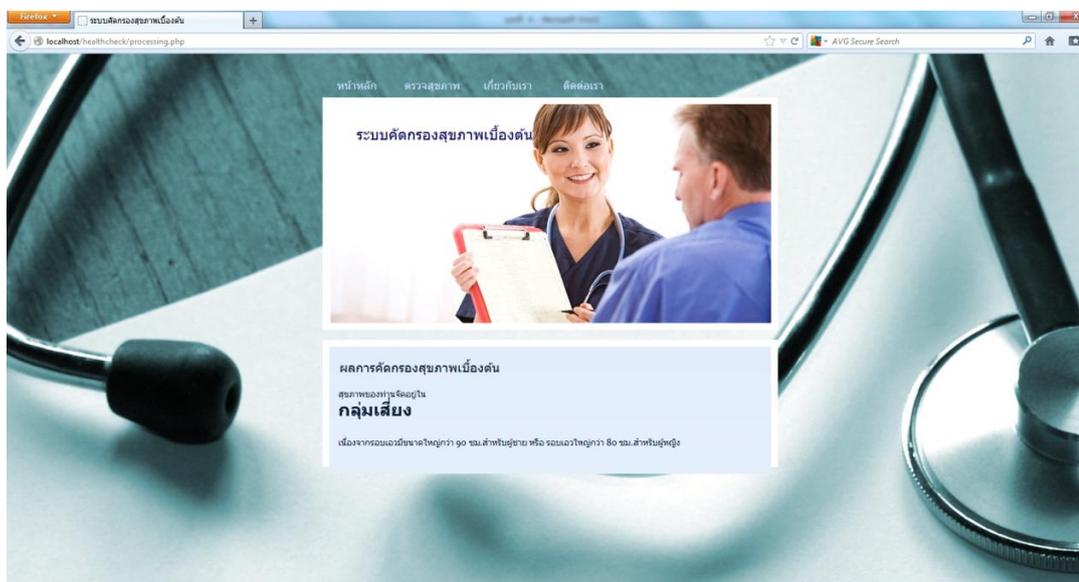


ภาพที่ 4.9 หน้าจอแบบสำรวจสภาวะสุขภาพประชาชน

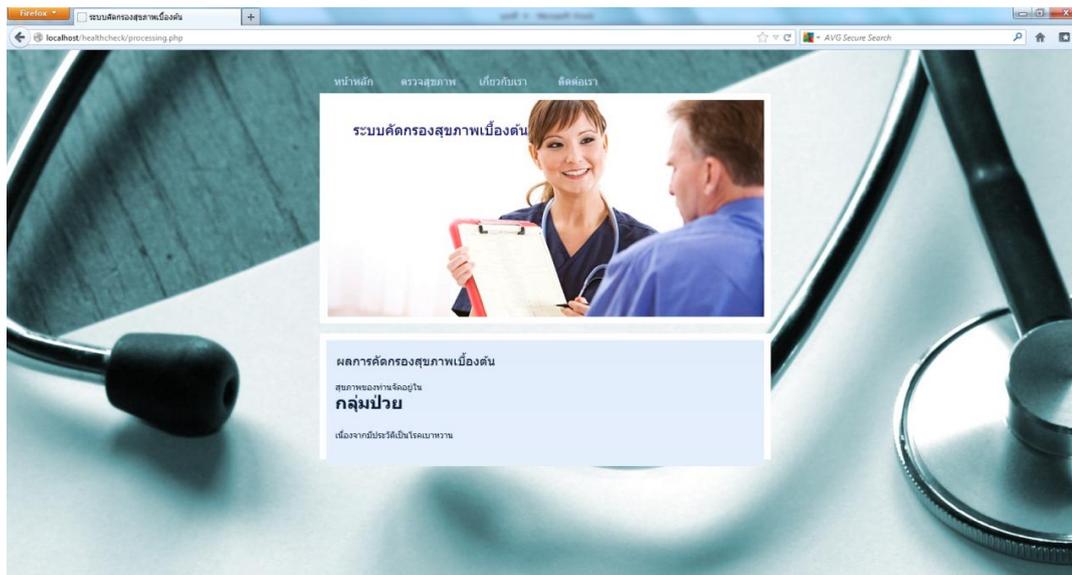
เมื่อผู้ใช้ทำการกรอกข้อมูลสุขภาพตามรายการต่างๆ แล้วทำการคลิกที่ปุ่มบันทึกและประมวลผล โปรแกรมจะทำการประมวลผลและจำแนกกลุ่มว่าเป็นกลุ่มปกติ กลุ่มเสี่ยง หรือ กลุ่มป่วย



ภาพที่ 4.10 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มปกติ



ภาพที่ 4.11 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มเสี่ยง



ภาพที่ 4.12 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มป่วย