



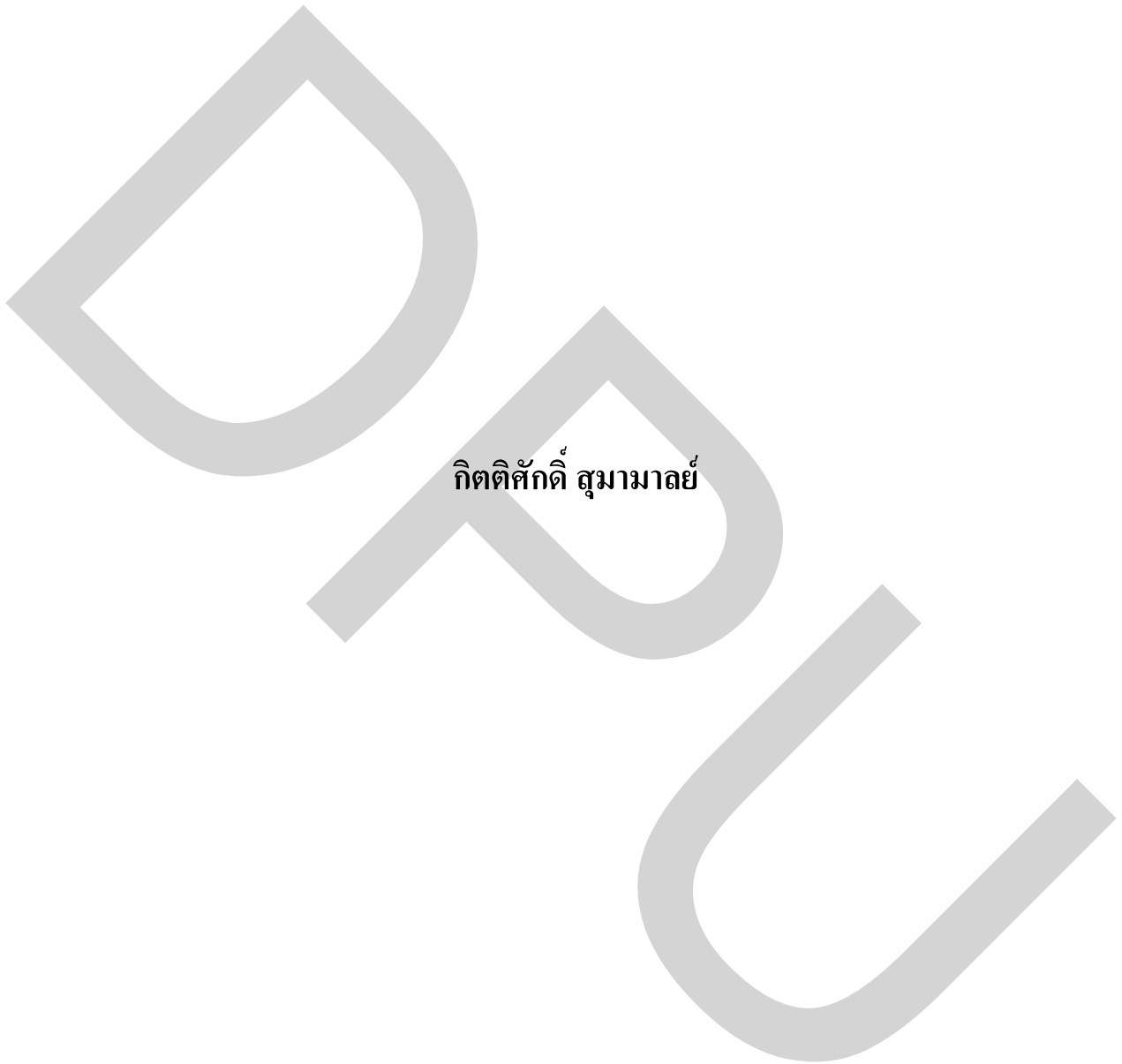
**การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล**  
**Basic Health Screening by Using Data Mining Techniques**

**กิตติศักดิ์ สุมามาลย์**

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมเว็บ  
คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

**พ.ศ. 2555**

## การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล



กิตติศักดิ์ สุมามาลย์

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2555

**Basic Health Screening by Using Data Mining Techniques**



**Kittisak Sumamal**

**Thematic Paper Submitted in Partial Fulfillment of  
the Requirements for the Degree of  
Master of Science in Web Engineering  
Faculty of Information Technology, Dhurakij Pundit University**

**2012**

## กิตติกรรมประกาศ

ผู้จัดทำสารนิพนธ์ขอขอบคุณอาจารย์ ดร.อรรณ อิ่มสมบัติ ที่เสียสละเวลาให้ความช่วยเหลือ อนุเคราะห์ และช่วยปรับแก้งานวิจัยนี้จนสำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณผู้ช่วยศาสตราจารย์ ดร.วรสิทธิ์ ชูชัยวัฒนา ผู้อำนวยการหลักสูตร สาขาวิศวกรรมเว็บ และคณาจารย์ทุกท่านที่ให้คำแนะนำ สนับสนุนและช่วยเหลือในการทำสารนิพนธ์

ขอขอบคุณศูนย์แพทย์ชุมชน จังหวัดบุรีรัมย์ ที่เอื้อเฟื้อข้อมูลที่ใช้ในการทำวิจัย

ขอกราบขอบพระคุณบิดามารดา ญาติพี่น้องทุกคน ที่เป็นกำลังใจ และให้คำปรึกษาในการทำสารนิพนธ์นี้

ผู้วิจัยหวังเป็นอย่างยิ่งว่า สารนิพนธ์ฉบับนี้จะเป็นประโยชน์กับผู้ที่ต้องการศึกษาการประยุกต์ใช้เทคนิคเหมืองข้อมูลในการคัดกรองสุขภาพ และเป็นประโยชน์ในการทำงานของหน่วยงานราชการทางด้านสาธารณสุข และหากมีข้อผิดพลาดประการใดในสารนิพนธ์ฉบับนี้ ผู้วิจัยต้องกราบขออภัยมา ณ ที่นี้ด้วย

กิตติศักดิ์ สุมามาลย์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	๗
บทคัดย่อภาษาอังกฤษ.....	๘
กิตติกรรมประกาศ.....	๑
สารบัญตาราง.....	๗
สารบัญภาพ.....	๘
บทที่	
1. บทนำ.....	1
1.1 ที่มาและความสำคัญของงาน.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	2
1.3 ขอบเขตของงานวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.5 นิยามศัพท์เฉพาะ.....	3
1.6 เครื่องมือที่ใช้ในการดำเนินการ.....	3
1.7 ระยะเวลาในการดำเนินการวิจัย.....	4
2. วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 นโยบายและทิศทางการพัฒนาระบบบริการสุขภาพของกระทรวงสาธารณสุข.....	5
2.2 ทฤษฎีการทำเหมืองข้อมูล (Data Mining).....	7
2.3 โปรแกรม Weka.....	15
2.4 งานวิจัยที่เกี่ยวข้อง.....	17
3. วิธีการดำเนินการวิจัย.....	20
3.1 การศึกษาและรวบรวมข้อมูล.....	20
3.2 การเตรียมข้อมูล (Data Pre-processing) และการวิเคราะห์ข้อมูล.....	21
3.3 การศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม.....	43
4. ผลดำเนินการวิจัย.....	51
4.1 การวิเคราะห์และสรุปผล.....	51
4.2 การพัฒนาระบบคัดกรองสุขภาพเบื้องต้น.....	59

## สารบัญ(ต่อ)

บทที่	หน้า
5. สรุปผลการศึกษา และข้อเสนอแนะการวิจัย.....	67
5.1 สรุปผลการวิจัย.....	67
5.2 ข้อเสนอแนะ.....	69
บรรณานุกรม.....	71
ภาคผนวก.....	74
ก. กฎการคัดกรองสุขภาพเบื้องต้นที่ได้จากการทำ Classification.....	75
ข. แบบสำรวจสภาวะสุขภาพประชาชน.....	78
ค. คู่มือลงรหัส การคัดกรองสุขภาพประชาชนปี 2555.....	82
ง. เอกสารงานประชุมวิชาการ การบริหารและการจัดการ ครั้งที่ 8.....	89
ประวัติผู้เขียน.....	103

## สารบัญตาราง

ตารางที่	หน้า
1.1 ระยะเวลาในการดำเนินการวิจัย.....	4
3.1 ผลการวิเคราะห์ค่าของข้อมูลเพศ.....	27
3.2 ผลการวิเคราะห์ค่าของข้อมูลสถานภาพสมรส.....	27
3.3 ผลการวิเคราะห์ค่าของข้อมูลการศึกษา.....	27
3.4 ผลการวิเคราะห์ค่าของข้อมูลดัชนีมวลกาย.....	28
3.5 ผลการวิเคราะห์ค่าของข้อมูลขนาดรอบเอว.....	28
3.6 ผลการวิเคราะห์ค่าของข้อมูลความดันโลหิต.....	28
3.7 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยของบิดามารดา.....	29
3.8 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยของพี่น้อง(สายตรง).....	29
3.9 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคเบาหวาน.....	30
3.10 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคความดันโลหิตสูง.....	30
3.11 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคตับ.....	30
3.12 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคอัมพาต.....	30
3.13 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคหัวใจ.....	31
3.14 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคไขข้อในเลือดผิดปกติ.....	31
3.15 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเป็นแผลที่เท้า/ตัดขา (จากเบาหวาน).....	31
3.16 ผลการวิเคราะห์ค่าของข้อมูลประวัติการคลอดบุตรน้ำหนักเกิน 4 กิโลกรัม.....	31
3.17 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการดื่มน้ำบ่อยและมาก.....	32
3.18 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการปัสสาวะกลางคืน 3 ครั้งขึ้นไป.....	32
3.19 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการกินจุแต่พอมลง.....	32
3.20 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการน้ำหนักลด/ อ่อนเพลีย.....	32
3.21 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการเป็นแผลที่ริมฝีปากบ่อยและหายยาก.....	33
3.22 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการคันตามผิวหนังและอวัยวะสืบพันธุ์.....	33
3.23 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการตาพร่ามัวต้องเปลี่ยนแว่นบ่อย.....	33
3.24 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการชาปลายมือปลายเท้าโดยไม่ทราบสาเหตุ.....	33
3.25 ผลการวิเคราะห์ค่าของข้อมูลการปฏิบัติตนถ้ามีประวัติการเจ็บป่วย.....	34

สารบัญตาราง(ต่อ)

ตารางที่	หน้า
3.26 ผลการวิเคราะห์ค่าของข้อมูลการสูบบุหรี่.....	34
3.27 ผลการวิเคราะห์ค่าของข้อมูลการดื่มเครื่องดื่มแอลกอฮอล์.....	34
3.28 ผลการวิเคราะห์ค่าของข้อมูลการออกกำลังกาย.....	35
3.29 ผลการวิเคราะห์ค่าของข้อมูลรสชาติของอาหารที่ชอบ.....	35
3.30 ผลการวิเคราะห์ค่าของข้อมูลการขับขี่หรือโดยสารรถจักรยานยนต์/รถยนต์.....	35
3.31 ผลการวิเคราะห์ค่าของข้อมูลการใช้ถุงยางอนามัยเมื่อมีเพศสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยา.....	36
3.32 ผลการวิเคราะห์ค่าของข้อมูลการตรวจเต้านมด้วยตนเอง.....	36
3.33 ผลการวิเคราะห์ค่าของข้อมูลความถี่ในการตรวจเต้านมด้วยตนเอง.....	36
3.34 ผลการวิเคราะห์ค่าของข้อมูลผลตรวจเต้านมด้วยตนเอง.....	37
3.35 ผลการวิเคราะห์ค่าของข้อมูลความถี่ในการตรวจมะเร็งปากมดลูก.....	37
3.36 ผลการวิเคราะห์ค่าของข้อมูลผลตรวจมะเร็งปากมดลูก.....	37
3.37 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารครบ 5 หมู่.....	37
3.38 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารมื้อเช้าเป็นหลัก.....	38
3.39 ผลการวิเคราะห์ค่าของข้อมูลการกินผักมากกว่าวันละ 3 ทัพพี.....	38
3.40 ผลการวิเคราะห์ค่าของข้อมูลการกินผลไม้วันละ 2-3 ส่วน.....	38
3.41 ผลการวิเคราะห์ค่าของข้อมูลการกินปลาอย่างน้อยวันละ 1 มื้อ.....	38
3.42 ผลการวิเคราะห์ค่าของข้อมูลการกินเนื้อสัตว์ไม่ติดมันสัปดาห์ละ 2-3 มื้อ.....	39
3.43 ผลการวิเคราะห์ค่าของข้อมูลการดื่มนมรสจืดหรือนมถั่วเหลืองผสมงาดำรสจืดวันละ 1-2 แก้ว.....	39
3.44 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารมื้อเย็นห่างจากเวลานอนไม่น้อยกว่า 4 ชั่วโมง.....	39
3.45 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารประเภทต้ม นึ่ง ลวก อบ ยำ หรือ หมก....	39
3.46 ผลการวิเคราะห์ค่าของข้อมูลการหลีกเลี่ยงอาหารไขมันสูง.....	40
3.47 ผลการวิเคราะห์ค่าของข้อมูลการหลีกเลี่ยงของหวานและขนมที่มีแป้งและน้ำตาล.....	40



สารบัญตาราง(ต่อ)

ตารางที่	หน้า
3.48 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารรสจัด.....	40
3.49 ผลการวิเคราะห์ค่าของข้อมูลการเลือกดื่มน้ำเปล่าแทนน้ำอัดลมหรือน้ำหวาน.....	40
3.50 ผลการวิเคราะห์ค่าของข้อมูลการหลีกเลี่ยงเครื่องดื่มที่มีแอลกอฮอล์.....	41
3.51 ผลการวิเคราะห์ค่าของข้อมูลอารมณ์ดีไม่เครียด.....	41
3.52 ผลการวิเคราะห์ค่าของข้อมูลการนอนหลับไม่น้อยกว่าวันละ 7-8 ชั่วโมง.....	41
3.53 ผลการวิเคราะห์ค่าของข้อมูลการออกกำลังกายสัปดาห์ละ 5 วันหรือ สัปดาห์ละ 5 ครั้ง.....	41
3.54 ผลการวิเคราะห์ค่าของข้อมูลขณะออกกำลังกายหายใจเร็วขึ้นกว่าปกติ และเหงื่อซึม.....	42
3.56 ผลการวิเคราะห์ค่าของข้อมูลทุกครั้งที่วัดรอบเอวได้เกณฑ์ปกติ คือ เพศหญิงไม่เกิน 80 ซม. และเพศชายไม่เกิน 90 ซม. ....	42
3.57 ผลการวิเคราะห์ค่าของข้อมูลสรุปพฤติกรรมสุขภาพ.....	42
3.58 ผลการวิเคราะห์ค่าของข้อมูลสรุปผลการคัดกรอง.....	43
4.1 เปรียบเทียบผลการทดสอบอัลกอริทึม C4.5 และ Partial Rules.....	56
4.2 Data Dictionary ของตาราง data.....	59
4.3 Data Dictionary ของตาราง rule.....	64

สารบัญภาพ

ภาพที่	หน้า
2.1 กระบวนการของData Mining.....	8
2.2 กระบวนการของData Integration.....	9
2.3 ตัวอย่างของNeural Network.....	12
2.4 แสดงหน้าจอการทำงาน โปรแกรมWeka.....	16
3.1 ข้อมูลที่นำมาใช้ในการศึกษา ซึ่งอยู่ในรูปแบบ Excel.....	21
3.2 การตัดแอทริบิวต์ที่ไม่เกี่ยวข้องกับการคัดกรองสุขภาพ.....	22
3.3 การตัดแอทริบิวต์ที่ถูกแปลงเป็นค่าไม่ต่อเนื่อง(Discretization).....	23
3.4 ค่าข้อมูลที่ผิดพลาดที่ต้องทำการแก้ไขในขั้นตอนการแก้ไขข้อมูลให้ถูกต้อง.....	24
3.5 ข้อมูลที่เก็บค่ามากกว่า1 ค่าในแอทริบิวต์เดียวกัน.....	25
3.6 การแปลงข้อมูลที่มีหลายค่าในหนึ่งแอทริบิวต์ให้กลายเป็นข้อมูลแบบ Binary.....	26
3.7 การบันทึกข้อมูลในรูปแบบนามสกุล.csv.....	45
3.8 การแปลงข้อมูลให้เป็นรูปแบบนามสกุล.arff.....	46
3.9 การแจกแจงข้อมูลในแต่ละแอทริบิวต์ในโปรแกรมWordPad.....	46
3.10 หน้าจอการวิเคราะห์ข้อมูล.....	47
3.11 การเลือกเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูล.....	47
3.12 หน้าจอแสดงผลฟังก์ชันการจำแนกกลุ่มข้อมูลด้วย เทคนิค Decision Rule: Partial Rules.....	48
3.13 หน้าจอแสดงผลฟังก์ชันความถูกต้องของโมเดล.....	48
3.14 การเลือกเทคนิคที่ใช้หากฎความสัมพันธ์ของข้อมูล.....	49
3.15 การตั้งค่ารายละเอียดในการหากฎความสัมพันธ์ของข้อมูล.....	49
3.16 หน้าจอแสดงผลฟังก์ชันหากฎความสัมพันธ์.....	50
4.1 ผลลัพธ์บางส่วนจากการหากฎความสัมพันธ์แบบApriori.....	52
4.2 ผลลัพธ์บางส่วนจากการหากฎความสัมพันธ์แบบFP Growth.....	53
4.3 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบDecision Tree: C4.5.....	54
4.4 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบDecision Rule: Partial Rules.....	55
4.5 ข้อมูลความผิดพลาดที่ได้จากผลลัพธ์การจำแนกกลุ่มของ Decision Rule: Partial Rules.....	57

สารบัญญภาพ(ต่อ)

ภาพที่	หน้า
4.6 ข้อมูลความผิดพลาดของการจำแนกกลุ่มปกติ.....	57
4.7 ผลลัพธ์การจำแนกกลุ่มหลังจากตัดแอทริบิวต์การปฏิบัติคนเมื่อเจ็บป่วย.....	58
4.8 ER-Diagram ของระบบคัดกรองสุขภาพเบื้องต้น.....	59
4.9 หน้าจอแบบสำรวจสภาวะสุขภาพประชาชน.....	64
4.10 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มปกติ.....	65
4.11 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มเสี่ยง.....	65
4.12 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มป่วย.....	66

หัวข้อสารนิพนธ์	การคัดกรองสุขภาพเบื้องต้น โดยใช้เทคนิคเหมืองข้อมูล
ชื่อผู้เขียน	กิตติศักดิ์ สุมาลย์
อาจารย์ที่ปรึกษา	ดร.อรวรรณ อิมสมบัติ
หลักสูตร	วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมเว็บ
ปีการศึกษา	2555

### บทคัดย่อ

งานวิจัยนี้ได้นำเสนอการวิเคราะห์ข้อมูลจากแบบสำรวจสภาวะสุขภาพประชาชนโดยใช้ข้อมูลของประชาชนในเขตเทศบาลเมืองจังหวัดบุรีรัมย์ ปี พ.ศ. 2555 จำนวน 1,071 เรคอร์ด ด้วยเทคนิคการทำเหมืองข้อมูล โดยแบ่งการศึกษาเป็น 2 ประเด็นคือ การศึกษาความสัมพันธ์ของการเป็นโรคต่างๆโดยใช้เทคนิค Association Rule และศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชนโดยแบ่งออกเป็น กลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย ด้วยเทคนิค Classification และนำผลลัพธ์ที่ได้ไปพัฒนาเป็นระบบคัดกรองสุขภาพเบื้องต้นเพื่อช่วยลดภาระงานของผู้เชี่ยวชาญในการวิเคราะห์ข้อมูล

ผลการทดลองพบว่า การศึกษาความสัมพันธ์ของการเป็นโรคต่างๆ อัลกอริทึม FP Growth ให้คำตอบที่เหมาะสมมากกว่าอัลกอริทึม Apriori และในการศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชน พบว่าอัลกอริทึม Partial Rules ให้ความถูกต้องมากกว่าอัลกอริทึมต้นไม้ตัดสินใจ C4.5 โดยมีค่าความถูกต้อง (precision) เท่ากับ 88.60% ค่าระลึก (recall) เท่ากับ 89.20% และค่าความเหวี่ยง (F-measure) เท่ากับ 88.80% เมื่อนำกฎที่ได้ไปให้ผู้เชี่ยวชาญตรวจสอบผู้เชี่ยวชาญได้ตัดข้อมูลบางแอทริบิวต์และกฎบางกฎที่ไม่เหมาะสมออกไป สุดท้ายจะได้ผลลัพธ์เป็นกฎจำนวน 42 กฎ และถูกนำไปใช้ในการสร้างระบบคัดกรองสุขภาพอัตโนมัติสำหรับให้ประชาชนทั่วไปเข้ามาตรวจสอบสุขภาพเบื้องต้นของตนเอง

Thematic Paper Title	Basic Health Screening by Using Data Mining Techniques
Author	Mr.Kittisak Sumamal
Thesis Advisor	Dr.Aurawan Imsombut
Academic Program	Master of Science in Web Engineering
Academic Year	2012

### ABSTRACT

This research proposed a method for 2012 health condition survey analysis of people from Buriram province, Thailand. The total number of obtained data from this survey is 1,071 records. A data mining technique was applied to the set of data. This study aims at finding a relationship between health conditions and diseases using Association Rule and developing a model of basic health screening to classify people as a normal group, a risky group, or a sick group using classification techniques. The result of this research will be applied to develop an automatic health screening system. This system could help an expert to perform health condition analysis task.

The result of experiment showed that FP Growth provided an appropriate solution comparing with Apriori for a task of finding a relationship between health conditions and diseases. For a task of basic health screening, Partial Rules algorithm outperformed Decision Tree. The precision and recall are 88.60% and 89.20% respectively and the F-measure is 88.80%. A group of experts were invited to examine the obtained rules. Some inappropriate attributes and rules were eliminated. The final outcome was 44 rules and these rules were embedded into the automatic health screening system that allows people to examine their health condition by themselves.

# บทที่ 1

## บทนำ

### 1.1 ที่มาและความสำคัญของงาน

จากข้อมูลของสำนักนโยบายและยุทธศาสตร์ กระทรวงสาธารณสุข (รายงานผู้ป่วยในช่วงปี พ.ศ. 2528-2552, 2554) พบว่าอัตราป่วยต่อประชากรแสนคนด้วยโรคหัวใจ โรคมะเร็ง โรคเบาหวาน โรคความดันโลหิตสูงและหลอดเลือดสมองมีแนวโน้มที่สูงขึ้นอย่างต่อเนื่อง โดยมีปริมาณที่เพิ่มขึ้นจากปี พ.ศ. 2548 จำนวน 4.7 ล้านคนเป็น 5.4 ล้านคนในปี พ.ศ. 2552 ซึ่งการเจ็บป่วยดังกล่าวเป็นโรคเรื้อรังที่สามารถป้องกันได้ในเบื้องต้น โดยประชาชนจะต้องดูแลสุขภาพตนเองด้วยการมีพฤติกรรมการบริโภคที่ถูกต้อง การควบคุมน้ำหนัก การออกกำลังกายอย่างสม่ำเสมอ ควรหลีกเลี่ยงการดื่มเครื่องดื่มที่มีแอลกอฮอล์ การพักผ่อนและการควบคุมความเครียด การผ่อนคลายในชีวิตประจำวัน รวมทั้งการตรวจสุขภาพอย่างสม่ำเสมอด้วยตนเองและสถานบริการสุขภาพ นอกจากนี้ประชาชนยังต้องมีส่วนร่วมในการกำหนดมาตรการทางสังคมในการอยู่ร่วมกัน เช่น เขตปลอดบุหรี่ การจำหน่ายอาหารปลอดภัย การรณรงค์ออกกำลังกาย การลดปัจจัยสาเหตุอย่างครอบคลุมทั้งเรื่องของพฤติกรรม การบริโภค ปัจจัยเสี่ยงจากสิ่งแวดล้อม การจราจร และวิถีการดำเนินชีวิตประจำวัน อันจะนำไปสู่การมีสังคมสุขภาพที่ดีร่วมกัน

กระทรวงสาธารณสุขจึงได้กำหนดนโยบายให้สำนักงานสาธารณสุขจังหวัดทุกจังหวัดเป็นแกนนำในการดำเนินกิจกรรมเพื่อให้ประชาชนตระหนักถึงความสำคัญของการมีพฤติกรรมสุขภาพที่เหมาะสม ซึ่งมีความจำเป็นต้องดำเนินการอย่างเร่งด่วนและต่อเนื่อง เพราะปัญหาที่เกิดขึ้นไม่เพียงแต่ส่งผลกระทบต่อสุขภาพเท่านั้น ยังมีผลต่อเนื่องถึงภาวะเศรษฐกิจและการพัฒนาประเทศด้วย

สำนักงานหลักประกันสุขภาพแห่งชาติจึงได้ส่งเสริมสนับสนุนงบประมาณในการเฝ้าระวังโรคเรื้อรังทุกจังหวัด โดยการใช้แบบสำรวจสภาวะสุขภาพของประชาชนอายุตั้งแต่ 15 ปีขึ้นไป ซึ่งใช้ในการคัดกรองสุขภาพเบื้องต้นเป็นประจำทุกปี ทั้งในเขตเทศบาลและนอกเขตเทศบาล ศูนย์แพทย์ชุมชน โรงพยาบาลบุรีรัมย์ ได้เล็งเห็นความสำคัญของนโยบายดังกล่าว จึงได้จัดทำการสำรวจสภาวะสุขภาพของประชาชนอายุตั้งแต่ 15 ปีในเขตเทศบาล ซึ่งข้อมูลประกอบด้วย 1) ข้อมูลทั่วไป 2) ข้อมูลครอบครัว 3) ข้อมูลการเจ็บป่วยและพฤติกรรมเมื่อเจ็บป่วย 4) ข้อมูลคะแนน

พฤติกรรมด้านสุขภาพ เพื่อนำข้อมูลที่ได้มาออกแบบกิจกรรมและงบประมาณให้สอดคล้องกับผลที่สำรวจได้ทุกปี ดังเช่น ผลการคัดกรองสุขภาพเบื้องต้นในชุมชนหนึ่งพบว่ามีคนที่อยู่ในกลุ่มเสี่ยงเป็นจำนวนมาก และเมื่อนำมาวิเคราะห์ทำให้ทราบว่าส่วนใหญ่เป็นคนอ้วนลงพุง จึงจัดให้มีการอบรมเพื่อปรับเปลี่ยนพฤติกรรมในกลุ่มเสี่ยงของชุมชนนั้นด้วย โครงการชุมชนลดหวาน มัน เค็ม ลดอ้วน ลดโรคใน ปี 2555 เป็นต้น

เทคนิคการทำเหมืองข้อมูล (Data Mining) เป็นเทคนิคหนึ่งที่ได้รับการนิยมนำมาใช้วิเคราะห์ข้อมูลที่มีปริมาณมาก (Han, J., et al., 2006) โดยสามารถนำมาใช้หากฎเกณฑ์ความสัมพันธ์ และรูปแบบที่แฝงอยู่ในฐานข้อมูล ได้เป็นความรู้หรือสารสนเทศที่น่าสนใจ โดยจะต้องเป็นความรู้ที่ไม่ธรรมดา ไม่สามารถค้นพบได้ด้วยวิธีการเรียกดูข้อมูลจากตาราง (Query) ต้องไม่เคยรู้มาก่อน และต้องสามารถนำไปใช้ประโยชน์ได้ เช่น การค้นพบความสัมพันธ์ของสินค้าที่ลูกค้าซื้อจากร้านสะดวกซื้อแห่งหนึ่งในประเทศสหรัฐอเมริกาพบว่าลูกค้ามักซื้อผ้าอ้อมกับเบียร์ไปด้วยกัน (Han, J., et al., 2006) ทำให้สามารถนำความรู้ที่ได้มาประยุกต์ในการวางกลยุทธ์เพื่อเพิ่มยอดขายหรือใช้ในการจัดชั้นสินค้าได้

ข้อมูลที่สำรวจได้จากแบบสำรวจสภาวะสุขภาพของประชาชนจะมีปริมาณเพิ่มสูงขึ้นทุกปี และยังขาดการจัดการข้อมูลเพื่อนำไปใช้ประโยชน์อย่างคุ้มค่า งานวิจัยนี้จึงนำเสนอการวิเคราะห์ข้อมูลจากแบบสำรวจสภาวะสุขภาพประชาชนด้วยเทคนิคการทำเหมืองข้อมูลโดยใช้เทคนิค Association Rule และ Classification เพื่อนำไปใช้ในการหาความสัมพันธ์ในการเกิดโรคต่างๆ และใช้ในการค้นหากฎในการจำแนกสุขภาพเบื้องต้นของประชาชนว่าอยู่ในกลุ่มปกติ กลุ่มเสี่ยงหรือกลุ่มป่วย และนำกฎที่ได้ไปสร้างระบบคัดกรองสุขภาพเบื้องต้นที่เปิดให้ประชาชนทั่วไปที่สนใจสามารถเข้ามาตรวจสอบสุขภาพได้ด้วยตนเอง ซึ่งช่วยลดภาระงานของผู้เชี่ยวชาญที่ต้องใช้เวลาวิเคราะห์ข้อมูลที่ละรายให้สามารถทำงานได้สะดวกและรวดเร็วมากยิ่งขึ้น

## 1.2 วัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาการเตรียมข้อมูลเพื่อนำมาใช้ในการทำเหมืองข้อมูล
2. เพื่อศึกษาเทคนิคการทำเหมืองข้อมูลแบบ Association Rule และ Classification
3. เพื่อศึกษาลักษณะความสัมพันธ์ของการเกิดโรคต่างๆ
4. เพื่อศึกษารูปแบบการคัดกรองผู้ป่วยจากข้อมูลแบบสำรวจสภาวะสุขภาพประชาชนแบบ

อัตโนมัติ

### 1.3 ขอบเขตของงานวิจัย

ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลของประชาชนในเขตเทศบาลเมืองจังหวัดบุรีรัมย์ ปี พ.ศ. 2555 จำนวนทั้งหมด 1,071 เรคอร์ด โดยเทคนิคที่ศึกษาประกอบด้วย

1. Association Rule ใช้ในการหาความสัมพันธ์ของการเกิดโรคต่างๆ ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง โรคตับ โรคอัมพาต โรคหัวใจ และโรคไขมันในเลือดผิดปกติ

2. Classification ใช้ในการหารูปแบบการพยากรณ์การคัดกรองสุขภาพเบื้องต้น โดยแบ่งออกเป็น กลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถกฎการจำแนกที่ได้ไปสร้างเป็นระบบคัดกรองสุขภาพเบื้องต้น เพื่อช่วยอำนวยความสะดวกในการทำงานของเจ้าหน้าที่ศูนย์แพทย์ชุมชน และสามารถให้ประชาชนทั่วไปที่สนใจสามารถเข้ามาสำรวจสุขภาพได้ด้วยตนเอง

2. สามารถนำความรู้ที่ได้ไปประยุกต์ใช้ในการหาทางป้องกันและลดจำนวนผู้ป่วยที่มีความเสี่ยงในการเกิดโรคที่ศึกษา

### 1.5 นิยามศัพท์เฉพาะ

1. Data Mining เป็นกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น (Han, J., et al., 2006)

2. Association Rule เป็นการค้นหากฎความสัมพันธ์ของข้อมูล โดยค้นหาความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าสองชุดขึ้นไป (Han, J., et al., 2006)

3. Classification เป็นการจัดแบ่งประเภทของข้อมูล โดยหาชุดต้นแบบหรือชุดของการทำงานที่อธิบายและแบ่งประเภทข้อมูล วัตถุประสงค์เพื่อให้สามารถใช้เป็นต้นแบบทำนายประเภทของวัตถุหรือข้อมูลที่ไม่มีการระบุประเภทหรือชนิดของข้อมูล ซึ่งต้นแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยอาจจะเป็นกลุ่มข้อมูลที่มีการระบุประเภทหรือกลุ่มเรียบร้อยแล้ว รูปแบบของต้นแบบแสดงได้หลายแบบเช่น Classification Rules, Decision Trees หรือ Neural Networks เป็นต้น (Han, J., et al., 2006)

### 1.6 เครื่องมือ (Tool) ที่ใช้ในการดำเนินการ

1. โปรแกรมสำเร็จรูป Weka เพื่อใช้ในการทำเหมืองข้อมูล
2. ระบบปฏิบัติการ: Windows



3. โปรแกรม Microsoft Excel ใช้ในการทำความสะอาดข้อมูล ลดขนาดของข้อมูลและเตรียมข้อมูลสำหรับนำเข้าทดสอบในโปรแกรม Weka

4. โปรแกรม WordPad ใช้ในแก้ไขเพิ่มข้อมูล .arff

5. โปรแกรม Adobe Dreamweaver CS5.5, PHP, Apache และฐานข้อมูล MySQL สำหรับพัฒนาระบบคัดกรองสุขภาพเบื้องต้น

### 1.7 ระยะเวลาในการดำเนินการวิจัย

ระยะเวลาในการดำเนินการวิจัย และแผนผังขั้นตอนในการดำเนินการวิจัย สามารถสรุปได้ตารางที่ 1.1

ตารางที่ 1.1 ระยะเวลาในการดำเนินการวิจัย

ลำดับ	ระยะเวลาดำเนินงาน (เดือน)	1	2	3	4	5	6	7	8	9	10
1	ศึกษาและรวบรวมข้อมูล	←→									
2	เตรียมข้อมูล		←→								
3	ศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม				←→						
4	วิเคราะห์และสรุปผล					←→					
5	พัฒนาระบบ						←→				
6	สรุปผลการวิจัยและข้อเสนอแนะ							←→			
7	เรียบเรียงรายงานค้นคว้าอิสระ							←→			

## บทที่ 2

### วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ ผู้วิจัย ได้ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องนำมาเสนอ ดังนี้

- 2.1 นโยบายและทิศทางการพัฒนาระบบบริการสุขภาพของกระทรวงสาธารณสุข
- 2.2 ทฤษฎีการทำเหมืองข้อมูล
- 2.3 โปรแกรม Weka
- 2.4 งานวิจัยที่เกี่ยวข้อง

#### 2.1 นโยบายและทิศทางการพัฒนาระบบบริการสุขภาพของกระทรวงสาธารณสุข

การพัฒนาประเทศไทยในระยะแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ ฉบับที่ 11 (พ.ศ.2555-2559) ยังคงยึดหลักปรัชญาของเศรษฐกิจพอเพียงที่มุ่งเน้นให้คนเป็นศูนย์กลางของการพัฒนา และสร้างสมดุลการพัฒนาในทุกมิติภายใต้วิสัยทัศน์ประเทศไทย ปี พ.ศ. 2570 “สังคมอยู่ร่วมกันอย่างมีความสุขด้วยความเสมอภาคเป็นธรรมและมีภูมิคุ้มกันต่อการเปลี่ยนแปลง” และทิศทางของแผนพัฒนาสุขภาพแห่งชาติ ฉบับที่ 11 (พ.ศ.2555-2559) มุ่งพัฒนาภายใต้หลักปรัชญาเศรษฐกิจพอเพียง เน้นการสร้างเอกภาพและธรรมาภิบาลในการอภิบาลระบบสุขภาพให้มีความสำคัญกับการสร้างกระบวนการมีส่วนร่วมของทุกภาคส่วนในสังคม มุ่งเน้นการสร้างหลักประกันและการจัดบริการที่ครอบคลุมเป็นธรรม โดยมีวิสัยทัศน์ “ประชาชนทุกคนมีสุขภาพดี ร่วมสร้างระบบสุขภาพพอเพียงเป็นธรรม นำสู่สังคมสุขภาวะ” ซึ่งมีความสอดคล้องกับรัฐธรรมนูญแห่งราชอาณาจักรไทย พ.ศ.2550 ได้บัญญัติไว้ว่า “ส่งเสริมสนับสนุนและพัฒนาระบบสุขภาพที่เน้นการสร้างเสริมสุขภาพ อันนำไปสู่สุขภาวะที่ยั่งยืนของประชาชน”

กระทรวงสาธารณสุขมีอำนาจหน้าที่เกี่ยวกับการสร้างเสริมสุขภาพอนามัย การป้องกันโรคและการรักษาโรค การฟื้นฟูสมรรถภาพของประชาชน เพื่อให้ประชาชนได้รับบริการอย่างครอบคลุม ท้าถึงเป็นธรรม และสร้างระบบบริการสุขภาพให้มีมาตรฐานในทุกระดับเพื่อตอบสนองต่อปัญหาในกลุ่มเป้าหมายและพัฒนาระบบส่งต่อที่ไร้รอยต่อ ตลอดจนสร้างกลไกกลางระดับชาติในการดูแลระบบบริการสุขภาพและพัฒนาระบบบริหารจัดการทรัพยากรให้มีประสิทธิภาพ ลดค่าใช้จ่ายด้านสุขภาพของประชาชน เพื่อเป้าหมายประชาชนมีคุณภาพชีวิตที่ดี จึงได้นำแนวคิดแบบ

ใหม่ (New management) มาใช้ในการบริหารจัดการเพื่อแก้ไขปัญหาสุขภาพและเป็นไปตามทิศทางของแผนพัฒนาสุขภาพแห่งชาติฉบับที่ 11 โดยการจัดทำแผนแม่บทในการพัฒนาระบบบริการสุขภาพทุกระดับเป็นแผน 5 ปี (Service Plan 2555-2559) ภายใต้ปรัชญาการจัดการทั่วถึงและเป็นธรรมไม่ว่าจะอยู่ ณ ภูมิภาคไหนของประเทศไทย โดยมีกลยุทธ์หลักที่สำคัญไว้ 3 ประเด็น (Key Strategic areas) ดังนี้ 1) การพัฒนาโรงพยาบาลระดับต่างๆ ให้เป็นส่วนหนึ่งของระบบเครือข่ายบริการสุขภาพ (Health service network system) เพื่อให้โรงพยาบาลแต่ละแห่งมีการเติบโตอย่างมีทิศทาง มีการดำเนินงานที่ชัดเจน และมีการเกี่ยวเนื่องซึ่งกันและกันภายในเครือข่ายบริการ 2) การพัฒนาเป็นศูนย์กลางเชี่ยวชาญระดับสูง 4 สาขาที่สำคัญคือ โรคหัวใจ โรคมะเร็ง อุบัติเหตุและทารกแรกเกิด 3) การพัฒนาระบบบริการปฐมภูมิ ซึ่งเป็นด้านหน้าใกล้ชิดประชาชนมากที่สุด การจัดการบริการในระดับปฐมภูมิที่สำคัญคืองานบริการเวชปฏิบัติครอบครัวและงานสุขภาพชุมชน ดูแลในมิติของสุขภาพองค์รวม การสร้างความเข้มแข็งให้ครอบครัวและชุมชน การสร้างเสริมสุขภาพเชิงรุกโดยให้ชุมชนและประชาชนมีบทบาทสำคัญในการจัดการสุขภาพตนเองในระดับต้น ดังนั้นการคัดกรองสุขภาพประชาชนจึงเป็นเครื่องมือที่สำคัญที่สามารถป้องกันโรค หยุดยั้งการลุกลามเข้าสู่ระยะรุนแรงและแก้ไขได้ในเบื้องต้น รวมทั้งนำข้อมูลที่ได้มาใช้ในการสะท้อนพฤติกรรมสุขภาพและสภาวะสุขภาพ สถานการณ์ใหม่ๆ ที่เป็นปัจจุบัน เพื่อวางแผนในการกำหนดทิศทางนโยบายและการแก้ไขที่ถูกต้องต่อไป

สำนักงานหลักประกันสุขภาพแห่งชาติจึงได้ส่งเสริมสนับสนุนงบประมาณในการเฝ้าระวังโรคเรื้อรังทุกจังหวัด โดยการใช้แบบสำรวจสภาวะสุขภาพของประชาชนอายุตั้งแต่ 15 ปีขึ้นไป ซึ่งใช้ในการคัดกรองสุขภาพเบื้องต้นเป็นประจำทุกปี ทั้งในเขตเทศบาลและนอกเขตเทศบาล ศูนย์แพทย์ชุมชน โรงพยาบาลบุรีรัมย์ ได้เล็งเห็นความสำคัญของนโยบายดังกล่าว จึงได้จัดทำการสำรวจสภาวะสุขภาพของประชาชนอายุตั้งแต่ 15 ปีในเขตเทศบาล ซึ่งข้อมูลประกอบด้วย 1) ข้อมูลทั่วไป 2) ข้อมูลครอบครัว 3) ข้อมูลการเจ็บป่วยและพฤติกรรมเมื่อเจ็บป่วย 4) ข้อมูลคะแนนพฤติกรรมด้านสุขภาพ เพื่อนำข้อมูลที่ได้มาออกแบบกิจกรรมและงบประมาณให้สอดคล้องกับผลที่สำรวจได้ทุกปี ดังเช่น ผลการคัดกรองสุขภาพเบื้องต้น ในชุมชนหนึ่งพบว่ามีคนที่อยู่ในกลุ่มเสี่ยงเป็นจำนวนมาก และเมื่อนำมาวิเคราะห์ทำให้ทราบว่าส่วนใหญ่เป็นคนอ้วนลงพุง จึงจัดให้มีการอบรมเพื่อปรับเปลี่ยนพฤติกรรมในกลุ่มเสี่ยงของชุมชนนั้นด้วย โครงการชุมชนลดหวาน มัน ไขมัน ลดอ้วน ลดโรคใน ปี 2555 เป็นต้น

## 2.2 ทฤษฎีการทำเหมืองข้อมูล (Data Mining)

Data Mining คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น (Han, J., et al., 2006) ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

### 2.2.1 วิวัฒนาการของ Data Mining

1. ปี ค.ศ 1960 Data Collection มีการนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่นำเชื่อถือ เพื่อป้องกันการสูญหายได้เป็นอย่างดี
2. ปี 1980 Data Access มีการนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์ระหว่างกัน เพื่อนำไปวิเคราะห์ และตัดสินใจอย่างมีประสิทธิภาพ
3. ปี 1990 Data Warehouse and Decision Support มีการนำข้อมูลมาเก็บลงในฐานข้อมูลขนาดใหญ่ ครอบคลุมการใช้งานทั้งหมดขององค์กร เพื่อช่วยสนับสนุนการตัดสินใจ
4. ปี 2000 Data Mining นำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผล โดยสร้างแบบจำลองและความสัมพันธ์ทางสถิติ

### 2.2.2 วัตถุประสงค์ในการใช้ Data Mining

1. เพื่อการค้นพบองค์ความรู้ใหม่ในฐานข้อมูล (Knowledge Discovery in Databases)
2. เพื่อการสกัดองค์ความรู้ที่ซ่อนเร้นอยู่ (Knowledge Extraction)
3. เพื่อจัดการกับข้อมูลในอดีต (Data Archeology)
4. เพื่อสำรวจข้อมูล (Data Exploration)
5. เพื่อค้นหา Pattern ของข้อมูลที่ซ่อนอยู่ (Data Pattern Processing)
6. เพื่อใช้ขุดเจาะข้อมูล (Data Dredging)
7. เพื่อให้ได้มาซึ่งสารสนเทศที่มีประโยชน์

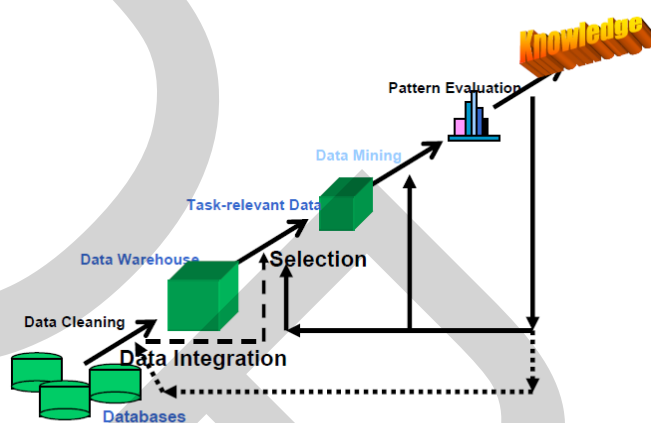
### 2.2.3 เป้าหมายหลักของ Data Mining

คุณลักษณะและเป้าหมายหลักของ Data Mining คือ ใช้ค้นหา Pattern ของข้อมูลที่ฝังลึก และซ่อนเร้นอยู่ภายในฐานข้อมูลขนาดใหญ่ โดยใช้สถาปัตยกรรม Client-Server ใช้เครื่องมือสมัยใหม่ที่สามารถแสดงผลแบบกราฟฟิก ผู้ใช้สามารถดูข้อมูลแบบเจาะลึก และสามารถใช้อุปกรณ์ในการสอบถามข้อมูลได้อย่างง่ายดาย โดยไม่ต้องอาศัยความชำนาญของ

ผู้พัฒนาโปรแกรม เพราะเครื่องมือถูกออกแบบมาให้ใช้งานได้ง่าย บ่อยครั้งอาจค้นพบผลลัพธ์ที่ไม่คาดหวังมาก่อน

#### 2.2.4 กระบวนการของการทำเหมืองข้อมูล

เป็นกระบวนการในการค้นหาลักษณะแฝงของข้อมูล (Pattern) ที่ซ่อนอยู่ในฐานข้อมูล โดยมีขั้นตอนดังภาพที่ 2.1

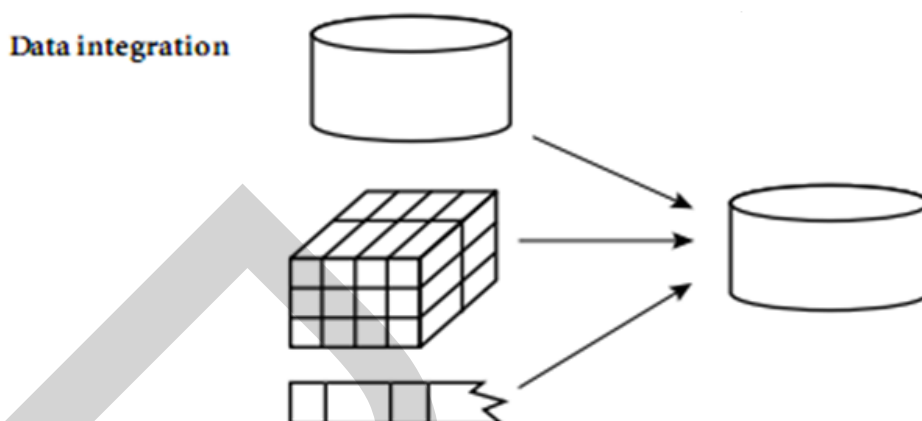


ภาพที่ 2.1 กระบวนการของ Data Mining

ขั้นตอนของการทำเหมืองข้อมูล มี 5 ขั้นตอนดังนี้

1. Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่มีเกี่ยวข้องออกไป โดยทั่วไปข้อมูลที่ถูกจัดเก็บอาจมีความผิดปกติต่างๆ ได้ เช่น ข้อมูลบางแอทริบิวต์ขาดหายไป (missing value) ขาดแอทริบิวต์ที่น่าสนใจ หรือขาดรายละเอียดของข้อมูล เป็นข้อมูลรบกวน (noisy data) เช่น ข้อมูลมีค่าผิดพลาด (error) หรือมีค่าผิดปกติ (Outliers)

2. Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน เช่น มีข้อมูลในคลังข้อมูล (Data Warehouse) ในรูปแบบของดาต้าคิวบ์ (Data Cube) และมีข้อมูลในรูปแบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) จำเป็นต้องทำการรวมข้อมูลให้เป็นข้อมูลชุดเดียวกันดังภาพที่ 2.2



ภาพที่ 2.2 กระบวนการของ Data Integration

3. Data Selection เป็นขั้นตอนระบุถึงแหล่งข้อมูลที่จะนำมาทำ mining รวมถึงการนำข้อมูลที่ต้องการออกจากฐานข้อมูล เพื่อสร้างกลุ่มข้อมูลสำหรับพิจารณาในเบื้องต้น

4. Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่ โดยใช้เทคนิคและอัลกอริทึมที่เหมาะสมกับข้อมูลที่มีอยู่

5. Pattern Evaluation เป็นขั้นตอนประเมินผล Pattern และนำเสนอองค์ความรู้ ในขั้นตอนนี้จะเป็นการวิเคราะห์ผลลัพธ์ที่ได้ แปลความหมาย และประเมินผลว่าผลลัพธ์นั้นเหมาะสมหรือตรงวัตถุประสงค์หรือไม่และนำเสนอ

#### 2.2.5 ประเภทของข้อมูลที่ใช้ในการทำเหมืองข้อมูล

1. ข้อมูลที่มาจากฐานข้อมูลเชิงสัมพันธ์ (Relational databases)
2. ข้อมูลจากคลังข้อมูล (Data warehouses)
3. ข้อมูลจากฐานข้อมูลรายการปรับปรุง (Transactional databases)
4. จากฐานข้อมูลพิเศษหรือที่เก็บข่าวสารพิเศษ ซึ่งได้แก่
  - ฐานข้อมูลเชิงวัตถุ
  - ฐานข้อมูลเกี่ยวกับเวลา
  - ฐานข้อมูลข้อความ (Text databases) และฐานข้อมูลมัลติมีเดีย
  - ฐานข้อมูลแบบเก่าในอดีตหรือข้อมูลที่มาจากต่างฐานข้อมูลกัน -
  - ข้อมูลจากแหล่ง WWW

#### 2.2.6 ชนิดขององค์ความรู้ที่ค้นพบ มีดังนี้

1. องค์ความรู้เกี่ยวกับคุณลักษณะของข้อมูล (Characterization) เช่น รู้ว่าคนที่สามารถเรียนต่อในระดับปริญญาเอกได้จะพิจารณาได้จากคุณลักษณะใด

2. องค์ความรู้เกี่ยวกับการจำแนกข้อมูล (Discrimination)

3. องค์ความรู้เกี่ยวกับความสัมพันธ์ของข้อมูล (Association) เช่น มีความสัมพันธ์ของการซื้อสินค้าพบว่า ถ้าลูกค้าซื้อป๊อปคอร์น จะต้องซื้อเปปซี่ตามมา

4. องค์ความรู้เกี่ยวกับการแยกประเภทข้อมูลและการพยากรณ์

(Classification/prediction)

5. องค์ความรู้เกี่ยวกับการจัดกลุ่มข้อมูล (Clustering)

6. องค์ความรู้เกี่ยวกับการวิเคราะห์ข้อมูลที่ผิดปกติ (Outlier analysis)

7. องค์ความรู้เกี่ยวกับข้อมูลอื่นๆ ในงานที่ค้นพบ (Other data mining tasks)

#### 2.2.7 งานของการทำเหมืองข้อมูล

การทำเหมืองข้อมูลสามารถทำงานในการขุดค้นข้อมูล ดังนี้

1. การวิเคราะห์คุณสมบัติและการแยกแยะข้อมูล (Characterization and Discrimination)

การวิเคราะห์คุณสมบัติ เช่น การพิจารณารับสมัครพนักงานของฝ่ายทรัพยากรบุคคล ต้องวิเคราะห์คุณสมบัติจากใบสมัครและการสัมภาษณ์

การแยกแยะข้อมูล เช่น การสูญเสียการได้ยินเป็นปัญหาสำคัญของคนงานที่ทำงานในโรงงานที่มีเสียงดัง ซึ่งการทราบถึงสมรรถภาพการได้ยินของตนเองจะช่วยให้คนงานสามารถป้องกันตนเองจากการสูญเสียการได้ยินได้

2. การหาความสัมพันธ์ของข้อมูล (Association) นิยมใช้ในการวิเคราะห์ตะกร้าสินค้า เพราะสามารถช่วยให้ค้นพบความสัมพันธ์ที่เป็นไปได้ทั้งหมดของการผสมกันของสินค้าที่น่าสนใจ เพื่อนำผลที่ได้ไปใช้ทำการตลาด

3. การจัดหมวดหมู่และการวิเคราะห์การถดถอย (Classification/ Regression)

การจัดหมวดหมู่ (Classification) ตัวอย่างของการจัดหมวดหมู่ ที่นำมาใช้กับงานด้านธุรกิจ เช่น การวิเคราะห์ลูกค้าที่ซื้อผลิตภัณฑ์จากรักภักดีต่อยี่ห้อสินค้า (Brand Loyalty) ขององค์กร

เทคนิคของ Data Mining ที่ใช้ในการแก้ปัญหาแบบ Classification ได้แก่ Decision Tree, Neural Networks, Naïve-Bayes และ K-nearest neighbor (K-NN)

ปัญหาแบบ Regression จะเหมือนกับแบบ Classification ต่างกันตรงที่ผลลัพธ์ที่ได้จาก Regression เป็นค่าแน่นอน ที่ไม่จำกัด จะเป็นค่าอะไรก็ได้ เช่น แบบจำลองทำนายว่า นาย B จะตอบรับข้อเสนอของบริษัท ถ้านาย B ได้รับผลกำไร 1,000 บาท (1,000 เป็นคำตอบเฉพาะที่แน่นอน แต่ไม่จำกัด ซึ่งตัวเลขอาจจะเป็นค่าอื่นไปได้เรื่อย ๆ ต่างจากคำตอบแบบ Yes, No )

#### 4. การวิเคราะห์การรวมกลุ่ม หรือ การแบ่งแยกข้อมูล (Cluster analysis/ Segmentation)

การวิเคราะห์การรวมกลุ่ม (Clustering) เป็นการรวมกลุ่มข้อมูลที่มีลักษณะเหมือนกัน รูปแบบหรือแนวโน้มที่จะเหมือนกัน การใช้เทคนิค Clustering จะไม่มีผลลัพธ์ (Output) ไม่มีตัวแปรอิสระ (Independent Variable) ไม่มีการจัดโครงสร้างของวัตถุ เราจะเรียกเทคนิคของ Clustering ว่าเป็นแบบเรียนรู้ข้อมูลโดยไม่ต้องอาศัยครูสอน(Unsupervised Learning) การทำ Clustering จะทำบนพื้นฐานของข้อมูลในอดีต

ตัวอย่างเช่น องค์กรต้องการทราบความเหมือนที่มีในกลุ่มของลูกค้าของตน เพื่อที่จะให้เข้าใจลักษณะเฉพาะของลูกค้ากลุ่มเป้าหมาย และสร้างกลุ่มของลูกค้าเพื่อที่องค์กรจะได้สามารถขายสินค้าได้ในอนาคต องค์กรจะทำการแยกกลุ่มของข้อมูลลูกค้าออกเป็นกลุ่ม ๆ (หาส่วนที่เป็น Intersection และ Union)

เทคนิคของ Data Mining เพื่อแก้ปัญหาแบบ Clustering คือวิธี Demographic Clustering กับ Neural Clustering

#### 5. การประเมินและการพยากรณ์ (Estimation/Prediction)

การประเมิน (Estimation) เป็นการประเมินที่ไม่สามารถกำหนดค่าหรือคุณสมบัติที่ชัดเจนได้ ใช้จัดการกับค่าที่มีผลแบบต่อเนื่อง เช่น ใช้ประเมินรายได้ของครอบครัว ประเมินความสูงของบุคคลในครอบครัว ประเมินจำนวนเด็ก ๆ ในครอบครัว

การพยากรณ์ (Prediction) จะเหมือนกับ Classification และ Estimation ต่างกันตรงที่ Record ถูกแยกจัดลำดับในการทำนายค่าในอนาคต และนำข้อมูลในอดีตมาสร้างเป็นแบบจำลอง ใช้ทำนายสิ่งที่จะเกิดขึ้นในอนาคต เช่น การทำนายว่าลูกค้ากลุ่มใด ที่องค์กรจะสูญเสียไปในอีก 6 เดือนข้างหน้า หรือ การทำนายยอดซื้อของลูกค้าจะเป็นเท่าใด ถ้าบริษัทลดราคาสินค้าลง 10%

#### 6. การบรรยายและการแสดงภาพของข้อมูล (Description / Visualization)

การบรรยาย (Description) เป็นการหาคำอธิบายถึงสิ่งที่จะเกิดขึ้น โดยอาศัยข้อมูลจากฐานข้อมูล เช่น กลุ่มคนที่มีการศึกษาหรือรายได้น้อย จะเลือกนักการเมืองที่มีนโยบายทุนนิยมมากกว่ากลุ่มคนชั้นกลาง

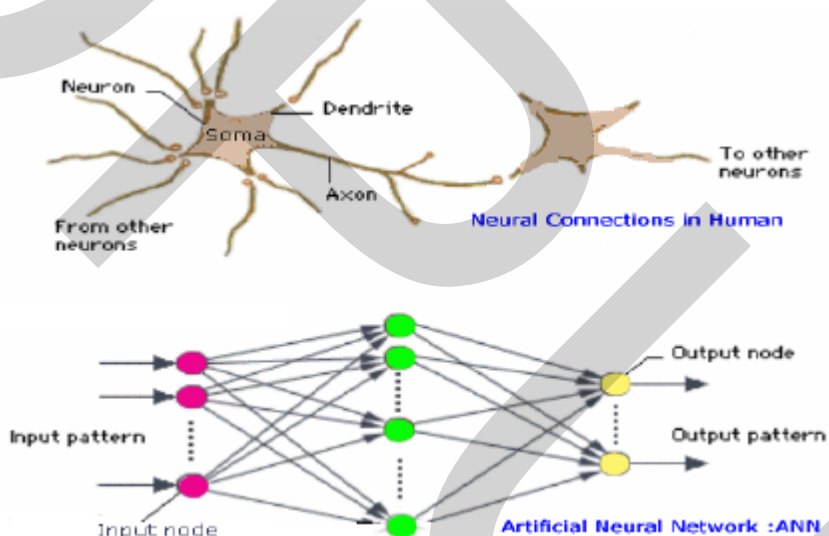
การแสดงผลภาพของข้อมูล (Visualization) เป็นการนำเสนอข้อมูลในรูปแบบกราฟฟิก หรืออาจนำเสนอในแบบ 2 มิติ สร้างรายละเอียดในการนำเสนอให้เข้าใจมากยิ่งขึ้น เช่น องค์กรต้องการหาสถานที่ในขยายสาขาใหม่ที่อยู่ในเขตพื้นที่ภาคเหนือของประเทศ ดังนั้นองค์กรจึงใช้แผนที่ Plot ที่ตั้งขององค์กรคู่แข่งที่มีสาขาอยู่ในเขตนั้น เพื่อพิจารณาสถานที่ตั้งที่เหมาะสมที่สุด



### 2.2.8 เทคนิคและอัลกอริทึมในการทำเหมืองข้อมูล

1. Neural Network เป็นแนวคิดให้คอมพิวเตอร์ทำงานเสมือนสมองของมนุษย์ เปลี่ยนตัวเองจากการประมวลผลตามลำดับ (Sequential Processing) ให้เป็นการประมวลผลแบบคู่ขนานได้ (Parallel Processing) มีลักษณะการทำงานโดยแต่ละ Process จะรับ Input เข้าไปคำนวณ และสร้าง Output ออกมาในลักษณะที่ไม่ใช่การทำงานแบบเชิงเส้นตรง เพราะ Input แต่ละตัวจะถูกให้ลำดับความสำคัญของค่าไม่เท่ากัน ค่าของ Output ที่ได้จากการเชื่อมโยงกันนี้ จะถูกนำมาเปรียบเทียบกับ Output ที่ได้ตั้งเอาไว้ ถ้าค่าที่ออกมาเกิดความคลาดเคลื่อน ก็จะนำไปสู่การปรับค่าหรือน้ำหนัก (weight) ของค่าที่ใส่ไว้ให้แต่ละ Input

Neural Network เป็นการสร้างแบบจำลอง ที่เลียนแบบการทำงานของสมองมนุษย์ มีโครงสร้างเป็นกลุ่มของ Node ที่เชื่อมโยงถึงกันในแต่ละ Layer คือ Input layer, Hidden layer, output layer



ภาพที่ 2.3 ตัวอย่างของ Neural Network

2. Decision Trees เป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างต้นไม้ (Decision Trees) ซึ่ง Decision Trees จะมีการทำงานแบบ Supervised Learning (คือการเรียนรู้แบบมีครูสอน) สามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ก่อนล่วงหน้าเรียกว่า Training set ได้อัตโนมัติ และพยากรกลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ ได้ด้วยรูปแบบของ Tree โครงสร้างประกอบด้วย Root Node, Child และ Leaf Node

3. Memory Based Reasoning (MBR) เปรียบเหมือนกับประสบการณ์ การเรียนรู้ของมนุษย์ ซึ่งอาศัยการสังเกตที่เกิดขึ้น แล้วสร้างรูปแบบของสิ่งนั้นขึ้นมา เราใช้ MBR เพื่อวิเคราะห์ฐานข้อมูลที่มีอยู่ และกำหนดลักษณะพิเศษของข้อมูลที่อยู่ therein ซึ่งข้อมูลจะต้องมีลักษณะที่สมบูรณ์ การสังเกตจึงจะสมบูรณ์และทำนายผลได้แม่นยำยิ่งขึ้น แบบจำลองจะถูกบอกคำตอบที่ถูกต้อง มีการเก็บคำตอบสำหรับแก้ปัญหาไว้ก่อนล่วงหน้าแล้ว (Supervised Learning)

4. Cluster Detection คือจะแบ่งฐานข้อมูลออกเป็น ส่วน ๆ เรียกว่า Segment (กลุ่ม Record ที่มีลักษณะคล้ายกัน) ส่วน Record ที่ต่างกันก็จะอยู่นอก Segment, Cluster Detection ถูกใช้เพื่อค้นหากลุ่มย่อย (Sub Group) ที่เหมือน ๆ กันในฐานข้อมูล เพื่อที่จะเพิ่มความถูกต้องในการวิเคราะห์ และสามารถมุ่งไปยังกลุ่มเป้าหมายได้ถูกต้อง

5. Link Analysis มุ่งเน้นทำงานบน Record ที่มีความสัมพันธ์กัน หรือเรียกว่า Association เทคนิคนี้จะมุ่งไปที่รูปแบบการซื้อหรือเหตุการณ์ที่เกิดขึ้นเป็นลำดับ มีอยู่ 3 เทคนิค คือ

5.1 Association Discovery ใช้วิเคราะห์การซื้อขายสินค้าในรายการเดียวกัน ศึกษาความสัมพันธ์อย่างใกล้ชิดที่ถูกปี ดซ่อนอยู่ของสินค้า ซึ่งสินค้าเหล่านั้นอาจมีแนวโน้มที่จะถูกซื้อควบคู่กัน ไปการวิเคราะห์แบบนี้เรียกว่า Market Basket Analysis คือ รายการทั้งหมดที่ถูกซื้อต่อครั้งที่ Super market การวิเคราะห์นี้สามารถนำมาใช้ประโยชน์ในการตัดสินใจ เช่น การเตรียมสินค้าคงเหลือ การวางแผนจัดชั้นวางสินค้า การทำ Mailing list สำหรับ Direct Mail การวางแผนเพื่อจัด Promotion สนับสนุนการขาย ตัวอย่างของ Association เช่น 75% ของผู้ซื้อน้ำอัดลมจะซื้อข้าวโพดคั่วด้วย

5.2 Sequential Pattern Discovery ถูกใช้ระบุความเกี่ยวเนื่องกันของการซื้อสินค้าของลูกค้า มีจุดหมายที่จะเข้าใจพฤติกรรมการซื้อสินค้าของลูกค้าในลักษณะ logn term เช่น ผู้ขายอาจพบว่าลูกค้าที่ซื้อ TV มีแนวโน้มที่จะซื้อ VDO ในเวลาต่อมา

5.3 Similar Time Sequence Discovery ค้นหาความเกี่ยวเนื่องกันระหว่างข้อมูล 2 กลุ่ม ซึ่งขึ้นต่อกันทางด้านเวลา โดยมีรูปแบบการเคลื่อนที่เหมือนกัน ผู้ขายสินค้ามักใช้เพื่อดูแนวโน้มเพื่อเตรียม Stock เช่น เมื่อไรก็ตามที่ยอดขายสินค้าน้ำอัดลมสูงขึ้น ยอดขายมันฝรั่งจะสูงขึ้นตาม

6. Genetic Algorithm (GA) เปรียบเสมือนเป็นการสร้างพันธุกรรมที่ดีที่สุด บนขั้นตอนของวิวัฒนาการทางชีวภาพ แนวคิดหลักคือ เมื่อเวลาผ่านไป วิวัฒนาการของเซลล์ชีวิตจะเลือกสายพันธุ์ที่ดีที่สุด “Fittest Species” GA มีความสามารถในการทำงานแบบรวมกลุ่มเข้าด้วยกัน เช่น มีการแบ่งกลุ่มและจัดรวมกลุ่มข้อมูลเป็น 3 ชุด ขั้นตอนการทำงานของ GA เริ่มจาก

จับกลุ่มข้อมูลเป็นกลุ่ม ๆ ด้วยการสุ่มเดา เปรียบเหมือนกลุ่ม 3 กลุ่มนี้เป็นเซลล์ของสิ่งมีชีวิต GA จะมี Fittest Function ที่จะบอกว่ากลุ่มข้อมูลใดเหมาะกับกลุ่ม ๆ ใด โดย Fittest Function จะเป็นตัวบ่งชี้ว่าข้อมูลเหมาะกับกลุ่มมากกว่าข้อมูลอื่น ๆ

GA จะมี Operator ซึ่งยอมให้มีการเลียนแบบและแก้ไขลักษณะของกลุ่มข้อมูล Operator จะจำลองหน้าที่ของชีวิตที่ถูกรบกวนในธรรมชาติ คือ มีการแพร่พันธุ์ จับคู่ผสมพันธุ์ และเปลี่ยนรูปร่างตามต้นแบบของพันธุกรรม เปรียบกับข้อมูลถ้ามีข้อมูลใดในกลุ่มถูกพบว่าตรงกับคุณสมบัติของ Fittest function แล้ว มันจะคงอยู่และถูกถ่ายเข้าไปในกลุ่มนั้น แต่ถ้าไม่ตรงกับคุณสมบัติ ก็ยังมีโอกาสที่จะถ่ายข้ามไปยังกลุ่มอื่นได้

7. Rule Induction ดึงเอาชุดกฎเกณฑ์ต่าง ๆ มาสร้างเป็นเงื่อนไขหรือกรณี วิธีการของ Rule Induction จะสร้างชุดของกฎที่เป็นอิสระ ซึ่งไม่จำเป็นต้องอยู่ในรูปแบบของโครงสร้างต้นไม้

8. K-nearest neighbor (K-NN) จะใช้วิธีในการจัดแบ่งคลาส โดยจะตัดสินใจว่าคลาสไหนที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวนของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไขหรือกรณีต่าง ๆ สำหรับแต่ละคลาสและกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกับมันมากที่สุด

K-NN ค่อนข้างใช้ปริมาณงานในการคำนวณสูงมากบนคอมพิวเตอร์ เพราะเวลาสำหรับการคำนวณจะเพิ่มขึ้นแบบแฟกทอเรียล ตามจำนวนจุดทั้งหมด เทคนิคของ K-NN จะมีการคำนวณเกิดขึ้นทุกครั้งที่มีการกรณีใหม่ๆ เกิดขึ้น ดังนั้นถ้าจะให้เทคนิคแบบ K-NN ทำงานได้เร็ว ข้อมูลที่ใช้อยู่ควรเก็บอยู่ใน MBR (Memory-Based Reasoning)

9. Logic Regression เป็นการวิเคราะห์ความถดถอยแบบเส้นตรงทั่วไป ใช้ในการพยากรณ์ผลลัพธ์ของ 2 ตัวแปร เช่น Yes/No , 0/1 แต่เนื่องจากตัวแปรตาม (Dependent Variable) มีค่าเพียง 2 อย่างเท่านั้น จึงไม่สามารถสร้างแบบจำลอง (Model) ได้สำหรับการวิเคราะห์แบบ Logic Regression

ดังนั้นแทนที่จะทำการพยากรณ์โดยอาศัยเพียงค่าของตัวแปรตามที่ได้ เราจะสร้าง Model โดยอาศัย Algorithm ของความน่าจะเป็นของการเกิดเหตุการณ์ เราเรียก Algorithm นี้ว่า Log Odds หรือ Logic Transromation

อัตราส่วนความน่าจะเป็น :  $\frac{\text{ความน่าจะเป็นที่จะเกิดเหตุการณ์}}{\text{ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์}}$

### 2.2.9 การประยุกต์ใช้ Data Mining กับงานด้านธุรกิจ

สามารถนำเทคนิคของ Data Mining ไปวิเคราะห์ข้อมูลในฐานข้อมูล เพื่อนำข้อมูลที่ได้ไปใช้ประโยชน์ในงานด้านต่าง ๆ ดังต่อไปนี้

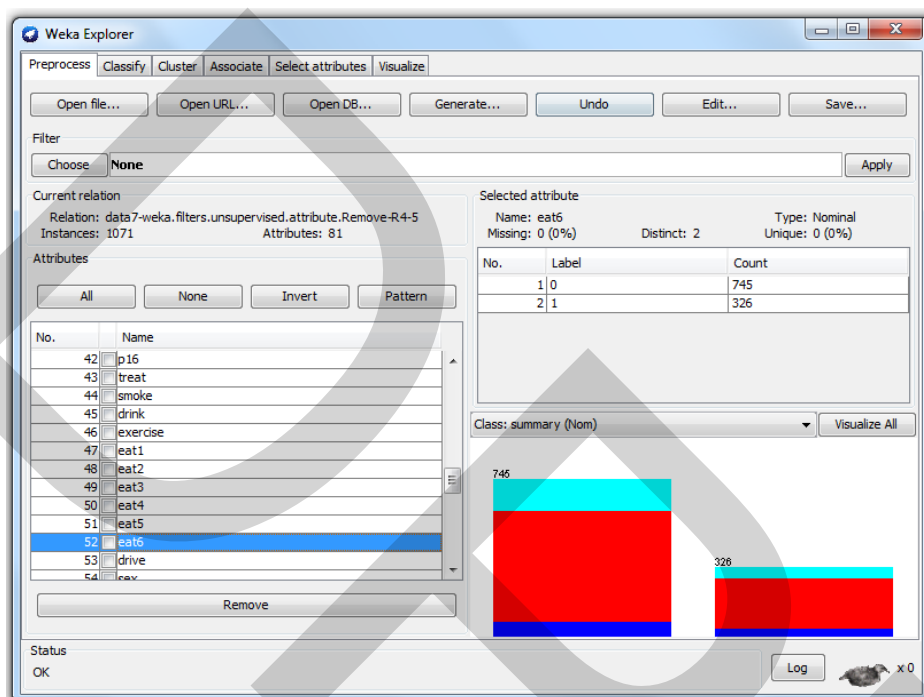
1. งานด้านการตลาด (Marketing) เช่น การทำ Promotion ส่งเสริมการขาย
2. งานด้านธนาคารและการเงิน (Banking / Financial Analysis) เช่น ใช้ในการวิเคราะห์การให้สินเชื่อแก่ลูกค้า การจัดทำ Package ในการกู้ยืม การทำนายอัตราดอกเบี้ย การแบ่งกลุ่มลูกค้าเพื่อหาเป้าหมายทางการตลาด (ลูกค้าชั้นดี)
3. งานด้านการขายปลีก (Retailing and sales) เป็นงานที่มีการเก็บข้อมูลจำนวนมาก ประยุกต์ใช้เพื่อหากลยุทธ์ ทำให้เกิดการได้เปรียบคู่แข่งทางการค้าในการหาลักษณะการซื้อของลูกค้า ความสัมพันธ์ของการซื้อกับช่วงเวลา ความสัมพันธ์ระหว่างตัวสินค้า และการวิเคราะห์ประสิทธิภาพของการโฆษณา เป็นต้น ช่วยให้เราสามารถหาวิธีการตอบสนองความต้องการของลูกค้าได้มากที่สุด และอาจหมายถึงส่วนแบ่งทางการตลาดที่เพิ่มขึ้นนั่นเอง
4. งานด้านการวางแผนในการผลิตสินค้า (Manufacturing and production) เช่น การพยากรณ์ยอดจำนวนการผลิตสินค้าเพื่อให้ได้กำไรมากที่สุด
5. งานด้านนายหน้าและความปลอดภัยด้านการค้า (Brokerage and securities trading) เช่น การพัฒนาวิธีการเพื่อสร้างความเชื่อมั่นในเรื่องความปลอดภัยของข้อมูล ในขณะที่มีการพัฒนาวิธีการเข้าถึงข้อมูล และการ Mining ให้สะดวกต่อการใช้งานมากขึ้น
6. งานด้านชีวการแพทย์และวิเคราะห์ DNA (Biomedical an DNA Analysis) เช่น การวิเคราะห์รูปแบบการจัดเรียงตัวของหน่วยพันธุกรรม เพื่อหาสาเหตุความผิดปกติที่ทำให้เกิดโรครวมไปถึงด้านการวินิจฉัยโรค การป้องกัน และการรักษา

นอกจากที่กล่าวมา ยังนำไปประยุกต์ใช้กับธุรกิจทางด้านประกันภัย (Insurance), Computer hardware และ software, หน่วยงานรัฐบาลและกระทรวงกลาโหม (Government and defense), สายการบิน (Airlines), งานด้านสุขภาพ (Health care), งานด้านการข่าว (Broadcasting) และงานด้านกฎหมาย (Law enforcement) ได้อีกด้วย

### 2.3 โปรแกรม Weka

Weka ย่อมาจาก Waikato Environment for Knowledge Analysis เป็นซอฟต์แวร์ฟรีที่แจกจ่ายภายใต้ GPL license โดยทั้งหมดนี้สร้างด้วยภาษาจาวาซึ่งประกอบไปด้วยอุปกรณ์การวิเคราะห์ข้อมูลและเพื่อนำไปใช้ทำเหมืองข้อมูลได้ เน้นการเรียนรู้ด้วยเครื่อง (Machine Learning)

กับการทำเหมืองข้อมูล (Data Mining) มีโมดูลย่อยสำหรับจัดการข้อมูล และตัวแสดงผลใช้ GUI และคำสั่ง ในการสั่งให้ซอฟต์แวร์ประมวลผล



ภาพที่ 2.4 แสดงหน้าจอการทำงาน โปรแกรม Weka

โปรแกรมหลักของซอฟต์แวร์ Weka

- Simple CLI (Command Line Interface) เป็น โปรแกรมรับคำสั่งการทำงานผ่านการพิมพ์
  - Explorer เป็น โปรแกรมที่ออกแบบในลักษณะ GUI
  - Experimenter เป็น โปรแกรมที่ออกแบบการทดลองและการทดสอบผล
  - Knowledge Flow เป็น โปรแกรมออกแบบผังการไหลของความรู้
  - Arff Viewer เป็น โปรแกรมที่ใช้สำหรับแก้ไขเพิ่มประเภท Arff
  - Log เป็น โปรแกรมที่ใช้อ่านข้อความบันทึกเก็บระหว่างการทำงาน
- ในการวิจัยนี้จะใช้เฉพาะ โปรแกรม Explorer ในการศึกษาวิเคราะห์และทำเหมืองข้อมูล ซึ่งมีหน้าจอการทำงานดังรูปที่ 2.4

เมนูหลักของ Explorer

- Preprocess การเตรียมข้อมูล

- Classify รวมโมดูลการทำเหมืองข้อมูลแบบจำแนกกลุ่ม
- Cluster รวมโมดูลการทำเหมืองข้อมูลแบบการจัดกลุ่ม
- Associate รวมโมดูลการทำเหมืองข้อมูลแบบกฎเชื่อมโยง
- Select attributes รวมโมดูลสำหรับการวิเคราะห์ความเกี่ยวข้องของลักษณะประจำ
- Visualize นำเสนอข้อมูลด้วยภาพนามธรรมสองมิติ

ประเภทของเพิ่มข้อมูลที่น่าเข้าข้อมูล ได้แก่

- เพิ่มข้อมูลในรูปแบบ arff, csv, C45
- ในกรณีเพิ่มข้อมูลอยู่ในเครือข่ายผู้ใช้สามารถเรียกใช้โดยอาศัย URL
- หรืออาจใช้ข้อมูลที่อยู่ในฐานข้อมูลที่เชื่อมโยงผ่าน JDBC

Weka เป็น โปรแกรมที่ได้พัฒนาขึ้นบนพื้นฐานของภาษาจาวา (Java) สามารถรัน (run) ได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทำงานในด้าน การทำเหมืองข้อมูล ที่รวบรวมแนวคิดอัลกอริทึมมากมาย ซึ่งอัลกอริทึมสามารถเลือกใช้งาน โดยตรงได้ 2 ทางคือ จากชุดเครื่องมือที่มีอัลกอริทึมมาให้ หรือเลือกใช้อัลกอริทึมที่ได้เขียนเป็น โปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล ได้แก่ Pre-Processing, Classification, Regression, Clustering, Association rules, Selection และ Visualization ข้อดีของโปรแกรม Weka คือ มีอัลกอริทึมที่รู้จักกันดีของการทำเหมืองข้อมูลให้ เลือกใช้อย่างครบถ้วน และสามารถเขียนฟังก์ชันเพิ่มเข้าไปในโปรแกรมได้

## 2.4 งานวิจัยที่เกี่ยวข้อง

ศุภกรใจ วุฒิกิจโกศล (2551,น.77) ใช้เทคนิคการทำเหมืองข้อมูลในผู้ป่วยข้อไหล่ติด โรงพยาบาลพระนั่งเกล้าเพื่อหาปัจจัยที่มีผลต่อการรักษาและสร้างกฎความสัมพันธ์ที่น่าสนใจเพื่อ เป็นแนวทางช่วยสนับสนุนการตัดสินใจเลือกเทคนิคการรักษาสำหรับนักกายภาพบำบัด โรงพยาบาลพระนั่งเกล้า โดยใช้ข้อมูลผู้ป่วยโรคข้อไหล่ติดที่ไม่มีโรคประจำตัวและไม่ได้เกิดจาก อุบัติเหตุมารับการรักษาทางกายภาพบำบัด โรงพยาบาลพระนั่งเกล้าที่รักษาหายในช่วงปี พ.ศ. 2548- 2550 จำนวน 550 ราย มีการเตรียมข้อมูลเพื่อให้ได้ข้อมูลที่ถูกต้อง และหาปัจจัยที่มีผลต่อการรักษา จากนั้นวิเคราะห์โดยการทำเหมืองข้อมูลด้วยเทคนิค Clustering และ Association Rule เลือกกฎ ความสัมพันธ์ที่น่าสนใจด้วยค่าสนับสนุนของกฎไม่ต่ำกว่าร้อยละ 20 และค่าความเชื่อมั่นไม่ต่ำกว่า ร้อยละ 90 ผลการศึกษาพบว่ามีปัจจัย อายุ ระยะเวลาที่เป็น องศาการยกแขน การไขว้หลัง ระดับ ความเจ็บปวด และเทคนิคการรักษาที่มีผลต่อการรักษา ตัวอย่างเช่น เมื่อใช้เทคนิค Association Rule หากกฎความสัมพันธ์ช่วงองศาการยกแขน 90-120 เลือกเทคนิคการรักษาด้วยแผ่นประคบความร้อน

บริเวณคอ ความร้อนคลื่นเหนือเสียงบริเวณไหล่ด้านหน้า ความร้อนคลื่นเหนือเสียงบริเวณ กล้ามเนื้อ Infraspinatus และกระตุ้นไฟฟ้าแบบ Surge บริเวณไหล่ด้านหน้า-หลังจะใช้จำนวนครั้ง ในการรักษาช่วง 19-24 ครั้งด้วยค่าสับสนุนของกฎ 43.48 เปอร์เซ็นต์และค่าความเชื่อมั่นที่ 100%

รักถีน เหลลาหา (2553,น.45) ศึกษาการพยากรณ์ความเสี่ยงการเกิดโรคมะเร็งปอด โดยใช้ทฤษฎีของการทำเหมืองข้อมูล สร้างระบบที่จัดกลุ่มของผู้ป่วยและพยากรณ์ความเสี่ยงการเกิด โรคมะเร็งปอด ทั้งนี้การจัดกลุ่มจะให้ทราบถึงค่าความเสี่ยงของปัจจัยเสี่ยงที่ทำให้เกิดโรคมะเร็ง ปอด จากนั้นนำค่าปัจจัยเสี่ยงมาวิเคราะห์และพยากรณ์ผู้ป่วยด้วยต้นไม้ตัดสินใจแบบ C4.5 โดยใช้ กลุ่มข้อมูลผู้ป่วยโรงพยาบาลมหาสารคาม ในเดือนสิงหาคม – ธันวาคม 2552 จำนวน 2,215 คน ผู้ป่วยมีสองกลุ่ม คือ ผู้ป่วยที่เป็น โรคมะเร็งปอดจำนวน 118 คน ไม่เป็นมะเร็งปอด 2,097 คนและได้ วัดค่าประสิทธิภาพการพยากรณ์จากค่าความแม่นยำและค่าความระลึก จากผลการวิจัยพบว่าปัจจัยที่ ส่งผลต่อค่าความเสี่ยงที่ทำให้เป็นมะเร็งปอดมากที่สุดคือ ปัจจัยประวัติด้านพันธุกรรม โดยมีค่า ความเสี่ยงเป็น 34.59 เท่าของคนที่ไม่ประวัติด้านพันธุกรรม รองลงมาคือ ปัจจัยเสี่ยง ด้านประวัติ การสูบบุหรี่ ประวัติการดื่มสุรา และอายุ ตามลำดับ ผลการพยากรณ์ให้ค่าความแม่นยำ (accuracy) ที่ 0.796 ค่าความระลึก (recall) 0.5 และค่าความถูกต้อง (precision) ของการจำแนกประเภทเฉลี่ยมี ค่าร้อยละ 96.8387

มาลิวรรณ บุญพลอย (2551,น.73) ได้ศึกษาเรื่องเหมืองข้อมูลและการค้นหาองค์ความรู้ สำหรับ โรคกล้ามเนื้ออ่อนแรง นำเสนอวิธีการใช้ขั้นตอนวิธีต้นไม้ช่วยตัดสินใจสองชั้น เพื่อจำแนกประเภทข้อมูลโรคกล้ามเนื้ออ่อนแรงเรื้อรัง ซึ่งกระบวนการทำงานของขั้นตอนวิธีนี้ สามารถแบ่งออกเป็นสองชั้นตอนใหญ่ๆคือ ชั้นตอนการจำแนกกลุ่มผู้ป่วยปกติ (Non-fatigued - NF) และชั้นตอนการจำแนกกลุ่มผู้ป่วยที่มีอาการผิดปกติแต่ไม่ใช่เป็นโรคกล้ามเนื้ออ่อนแรงเรื้อรัง (Insufficient number of symptoms or fatigue severity - ISF) และ กลุ่มผู้ป่วยที่เป็นโรคกล้ามเนื้อ อ่อนแรงเรื้อรัง (Chronic Fatigue Syndrome - CFS) วัตถุประสงค์ของงานวิจัยนี้เพื่อการพัฒนา อัลกอริทึมที่สามารถจำแนกประเภทข้อมูลของโรคกล้ามเนื้ออ่อนแรงเรื้อรังให้มีประสิทธิภาพมาก ขึ้นและทำการค้นหาองค์ความรู้ซึ่งจะทำการเปรียบเทียบประสิทธิภาพของวิธีต้นไม้ช่วยตัดสินใจ สองชั้น (Double Decision Trees) วิธีกลุ่มของต้นไม้ (Decision Tree Forests) และวิธีต้นไม้ช่วย ตัดสินใจ (Decision Trees) ข้อมูลที่นำมาใช้ในการทดลอง ประกอบด้วยข้อมูลของผู้ป่วยที่เป็นโรค กล้ามเนื้ออ่อนแรงเรื้อรัง ซึ่งข้อมูลของผู้ป่วยที่เป็นโรคกล้ามเนื้ออ่อนแรงนั้นประกอบไปด้วยข้อมูลผล เลือด และข้อมูลอาการป่วย ผลการวิจัยพบว่าประสิทธิภาพในการจำแนกประเภทข้อมูล โรค กล้ามเนื้ออ่อนแรงที่เรื้อรังโดยใช้ขั้นตอนวิธีต้นไม้ช่วยตัดสินใจสองชั้น นั้นสามารถจำแนกประเภท ของผู้ป่วยที่เป็นโรคกล้ามเนื้ออ่อนแรงเรื้อรังได้แม่นยำกว่าวิธีกลุ่มของต้นไม้และวิธีต้นไม้ช่วย

ตัดสินใจ โดยวิธีต้นไม้ช่วยตัดสินใจสองชั้นสามารถจำแนกข้อมูลกลุ่ม NF, CFS และ ISF มีค่าความแม่นยำเท่ากับ 83.60%, 83.30% และ 83.34% ตามลำดับ วิธีกลุ่มของต้นไม้สามารถจำแนกข้อมูลกลุ่ม NF, CFS และ ISF มีค่าความแม่นยำเท่ากับ 81.94%, 82.20% และ 81.79% ตามลำดับ วิธีต้นไม้ช่วยตัดสินใจสามารถจำแนกข้อมูลกลุ่ม NF, CFS และ ISF มีค่าความแม่นยำเท่ากับ 67.00%, 66.26% และ 67.00% ตามลำดับ เนื่องจากวิธีต้นไม้ช่วยตัดสินใจสองชั้นมีการแบ่งการจำแนกประเภทออกเป็น 2 ขั้นตอนโดยอาศัยหลักการคัดกรองข้อมูลผู้ป่วย ซึ่งทำให้ปัญหาการจำแนกประเภทข้อมูลลดความซับซ้อนลงจากเดิมเนื่องจากปัญหาในการจำแนกข้อมูลเป็นแบบหลายคลาส (Multi-class Problem) ได้ถูกลดรูปเป็นปัญหาแบบสองกลุ่ม (Two-class Problem) ซึ่งเป็นปัญหาที่ง่ายทำให้ขั้นตอนวิธีต้นไม้ช่วยตัดสินใจสองชั้นสามารถจำแนกข้อมูลได้อย่างมีประสิทธิภาพมากกว่า

อังคณา พิจาร โชติ (2552,น.73) ได้พัฒนาระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็นโรคเบาหวาน โดยใช้เทคนิคการทำเหมืองข้อมูล เพื่อช่วยในการวิเคราะห์หาความสัมพันธ์ของปัจจัยเสี่ยงต่างๆที่ก่อให้เกิดโรคเบาหวาน เช่น พ่อแม่มีประวัติเป็นโรคเบาหวาน มีประวัติการปัสสาวะบ่อย เป็นต้น ข้อมูลความสัมพันธ์ของปัจจัยเสี่ยงที่ได้จากการวิเคราะห์จะเป็นสารสนเทศสำคัญที่เป็นตัวช่วยให้หน่วยงานด้านการแพทย์นำไปใช้สำหรับวางแผนป้องกันโรค โดยระบบที่พัฒนานี้จะประกอบด้วยสามส่วนคือ ส่วนของคลังข้อมูลใช้สำหรับเก็บข้อมูลปัจจัยเสี่ยงที่ได้จากการคัดกรองกลุ่มเสี่ยง ส่วนของเหมืองข้อมูลเป็นตัวค้นหาความสัมพันธ์ของปัจจัยเสี่ยงต่างๆ และส่วนการแสดงผลข้อมูลรายงานซึ่งจะจัดทำในรูปแบบของเว็บแอปพลิเคชันเพื่อให้สะดวกต่อการเรียกใช้งาน ผลการค้นหาค่าความสัมพันธ์ด้วยค่าสนับสนุนที่ 0.5 และค่าความเชื่อมั่นที่ 0.3 พบว่าประวัติครอบครัวเป็นปัจจัยเสี่ยงที่มีอัตราการเสี่ยงมากที่สุดซึ่งพบคู่กับค่าดัชนีมวลกาย มีค่าความเชื่อมั่นที่ 0.94

งานวิจัยนี้ใช้เทคนิค Association Rules เช่นเดียวกับงานวิจัยของ สุกรัใจ วุฒิจิกโกศล (2551,น.77) และ อังคณา พิจาร โชติ (2552,น.73) สิ่งที่แตกต่างกันคือ งานวิจัยผู้ป่วยข้อไหล่ติดหาความสัมพันธ์ของเทคนิคการรักษากับผลการรักษา ส่วนงานวิจัยวิเคราะห์ปัจจัยเสี่ยงเบาหวานหาความสัมพันธ์ระหว่างปัจจัยเสี่ยงกับการเป็นโรค แต่งานวิจัยนี้หาความสัมพันธ์ของการเป็นโรคต่างๆ เทคนิค Classification ใช้ในงานวิจัยของ รักถิ่นเหลาหา (2553,น.45) ใช้อัลกอริทึม C4.5 ในการพยากรณ์ผู้ป่วยโรคมะเร็ง แต่งานวิจัยนี้ได้ทดลองอัลกอริทึม 2 ตัวคือ C4.5 และ Partial Rules แล้วจึงนำผลลัพธ์ที่ได้จากอัลกอริทึมที่มีค่า F-measure สูงกว่านั้นคือ Partial Rules นำไปพัฒนาระบบคัดกรองสุขภาพเบื้องต้น



## บทที่ 3

### วิธีการดำเนินการวิจัย

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูลมาใช้เพื่อช่วยพัฒนาระบบคัดกรองสุขภาพประชาชนแบบอัตโนมัติ โดยมีขั้นตอนการดำเนินการวิจัยดังนี้

1. ศึกษาและรวบรวมข้อมูล
2. เตรียมข้อมูลและวิเคราะห์ข้อมูล
3. การศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม
4. วิเคราะห์และสรุปผล
5. พัฒนาระบบคัดกรองผู้ป่วยจากข้อมูลแบบสำรวจสภาวะสุขภาพประชาชนแบบอัตโนมัติ
6. สรุปผลการวิจัยและข้อเสนอแนะ
7. เรียบเรียงงานค้นคว้าอิสระ

โดยในบทนี้จะกล่าวถึงขั้นตอน การศึกษาและรวบรวมข้อมูล การเตรียมข้อมูลและวิเคราะห์ข้อมูล และ การศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม

#### 3.1 การศึกษาและรวบรวมข้อมูล

แบบสำรวจสภาวะสุขภาพประชาชนจัดทำโดยศูนย์แพทยุมนชนสังกัด โรงพยาบาลบุรีรัมย์ ใช้สำหรับเก็บข้อมูลจากประชาชนทั่วไปโดยในการศึกษานี้เป็นการศึกษาข้อมูลการสำรวจสภาวะสุขภาพประชาชนจากประชาชน เขตเทศบาลเมืองบุรีรัมย์ โดยข้อมูลที่ได้มาจะอยู่ในรูปแบบของแฟ้มข้อมูล Excel ดังแสดงตัวอย่างได้ในภาพที่ 3.1 โดยประกอบด้วยแอทริบิวต์จำนวน 75 แอทริบิวต์ เช่น ลำดับที่ ชื่อ-สกุล รหัสบัตรประชาชน ที่อยู่ ชุมชน สิทธิการรักษา เพศ สถานภาพ การศึกษา น้ำหนัก ส่วนสูง ดัชนีมวลกาย ขนาดรอบเอว ความดัน ประวัติการเจ็บป่วยของครอบครัว ประวัติการเจ็บป่วย การรักษา การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การออกกำลังกาย รสชาติของอาหารที่ชอบ การขับขี้ เพศสัมพันธ์ การตรวจเต้านม การตรวจมะเร็งปากมดลูก พฤติกรรมสุขภาพด้านอาหาร อารมณ์ ออกกำลังกาย และสรุปผลการคัดกรองสุขภาพ เป็นต้น

ข้อมูลที่น่ามาใช้ในการศึกษาครั้งนี้เป็นข้อมูลที่ได้จากการเก็บแบบสำรวจสภาวะสุขภาพประชาชน ในเขตเทศบาลเมืองบุรีรัมย์ ปี พ.ศ. 2555 ซึ่งมีจำนวน 1,071 คน

ลำดับที่	ชื่อ-สกุล	ID	ที่อยู่	use	เพศ	อายุ	การศึกษา	น้ำหนัก	ส่วนสูง	BMI	sumBMI	hip	sumhip	ความดันโลหิต		
1	น.ส.กศิน	#####	44/4 จีระ	11	วัง	2	1	5	65	1.63	24.46	3	84	2	144	92
2	นางฉวีศรี	#####	86/7 อีสาน	11	ชรก	2	2	3	45	1.55	18.73	2	68	1	112	62
3	นายวิสิทธิ์	#####	86/3 อีสาน	11	บัตรทอง	1	2	5	85	1.78	26.83	4	97	2	145	89
4	นายอนุวิ	#####	86/3 อีสาน	11	บัตรทอง	1	1	4	54	1.74	17.84	1	77	1	113	65
5	น.ส.ณัฐก	#####	86/3 อีสาน	11	บัตรทอง	2	1	5	60	1.61	23.15	3	77	1	128	72
6	นางอมร	#####	86/3 อีสาน	11	บัตรทอง	2	2	3	46	1.55	19.15	2	77	1	122	91
7	นายเสกส	#####	86/3 อีสาน	11	บัตรทอง	1	1	3	65	1.75	21.22	2	82	1	120	76
8	นางสำรา	#####	10/3 จีระ	11	บัตรทอง	2	3	2	52	1.58	20.83	2	85	2	176	79
9	นายบุญ	#####	16 จีระ	11	บัตรทอง	1	2	2	82	1.7	28.37	4	99	2	147	88
10	นางอุษ	#####	27/2 นคร	11	บัตรทอง	2	3	2	32	1.5	14.22	1	77	1	189	92
11	นายสุกส	#####	16 จีระ	11	บัตรทอง	1	2	2	80	1.8	24.69	3	92	2	135	70
12	นายเส็ง	#####	27/1 นคร	11	ปกส	1	2	2	65	1.65	23.88	3	92	2	129	87
13	นางอจจร	#####	27/1 นคร	11	บัตรทอง	2	2	3	72	1.61	27.78	4	93	2	111	67
14	น.ส.อจจร	#####	110 อุนวรรณ์	11	บัตรทอง	2	1	3	58	1.59	22.94	2	78	1	131	92
15	นายบวร	#####	10/3 จีระ	11	บัตรทอง	1	2	4	68	1.75	22.20	2	88	1	127	84
16	นายท	#####	4/6 อีสาน	11	บัตรทอง	1	2	2	49	1.5	21.78	2	77	1	117	43

ภาพที่ 3.1 ข้อมูลที่น่ามาใช้ในการศึกษา ซึ่งอยู่ในรูปแบบ Excel

### 3.2 การเตรียมข้อมูล (Data Pre-processing) และการวิเคราะห์ข้อมูล

เนื่องจากข้อมูลที่ถูกจัดเก็บไว้ยังไม่อยู่ในรูปแบบที่สามารถนำมาใช้ในการทดสอบและวิเคราะห์ด้วยเครื่องมือการทำเหมืองข้อมูล จึงต้องมีการเตรียมข้อมูลก่อนนำไปใช้ในการทำเหมืองข้อมูลเพื่อให้ได้ผลการทดลองที่มีประสิทธิภาพสูงสุด โดยในขั้นตอนนี้จะประกอบด้วยขั้นตอนย่อย 3 ขั้นตอนคือ การลดขนาดของข้อมูล การแก้ไขข้อมูลให้ถูกต้อง และการแปลงข้อมูล

#### 3.2.1 การลดขนาดข้อมูล

เนื่องจากข้อมูลมีแอทริบิวต์ที่ไม่เกี่ยวข้องกับการคัดกรองอยู่เป็นจำนวนมากจึงจำเป็นต้องมีการลดแอทริบิวต์บางตัวออกไป ได้แก่ ลำดับที่ ชื่อ-สกุล รหัสบัตรประชาชน ที่อยู่ ชุมชน สิทธิการรักษา ดังแสดงในภาพที่ 3.2

ลำดับที่	ชื่อ-สกุล	ID	ที่อยู่	นามชน	use	เพศ	สถานภาพการศึกษ	น้ำหนัก	ส่วนสูง	BMI	sumBMhip	sumhip	ความดันครั้ง1	ความดันครั้ง2	สรุปค่า				
1	นางปรียะทอง	#####	58/2 นค	15	บัตรทอง	2	2	2	66	1.65	24.24	3	84	2	100	73	100	70	100
2	นางจตุรงค์	#####	66/1 นค	15	บัตรทอง	1	2	2	70	1.65	25.71	4	91	2	132	89	124	79	128
3	นางวิรัช เหว	#####	70 นคร	15	บัตรทอง	2	2	3	53	1.54	22.35	2	77	1	122	81	120	80	121
4	นส.ศุภา ศ	#####	70 นคร	15	บัตรทอง	2	1	3	45	1.6	17.58	1	66	1	100	77	97	69	99
5	นส.ศุภา อ	#####	70 นคร	15	บัตรทอง	2	3	2	50	1.5	22.22	2	72	1	126	73	120	70	123
6	นายพี ธิรม	#####	70 นคร	15	บัตรทอง	1	3	2	40	1.5	17.78	1	61	1	126	73	120	70	123
7	นายศักดิ์	#####	58/1 นค	15	ชรก	1	2	5	80	1.75	26.12	4	94	2	138	75	127	78	133
8	นส.ดวงพร	#####	58/1 นค	15	ชรก	2	1	5	63	1.6	24.61	3	76	1	110	70	110	70	110
9	นางสำราญ	#####	58/1 นค	15	ชรก	2	3	2	70	1.64	26.03	4	94	2	110	80	110	80	110
10	นส.วิรัช	#####	71/1 นค	15	ชรก	2	1	5	60	1.7	20.76	2	71	1	110	70	110	70	110
11	นายสุจินต์	#####	71/1 นค	15	ชรก	1	2	4	64	1.7	22.15	2	79	1	114	81	110	80	112
12	นายสมศักดิ์	#####	66/3 นค	15	บัตรทอง	2	2	2	85	1.72	28.73	4	102	2	129	77	120	70	125
13	นางจุฑารัตน	#####	79/3 นค	15	บัตรทอง	2	2	3	73	1.57	29.62	4	90	2	127	78	120	70	124
14	นายเปลื้อง	#####	79/3 นค	15	บัตรทอง	1	2	4	73	1.72	24.68	3	90	2	122	62	120	70	121
15	นายพิชญุต	#####	49/1 นค	15	ปกส	1	2	3	78	1.65	28.65	4	97	2	135	70	131	80	133
16	นางแสงโสม	#####	49/4 นค	15	ปกส	2	2	2	50	1.6	19.53	2	74	1	112	70	117	65	115
17	นส.โลหิตา	#####	49/4 นค	15	บัตรทอง	2	1	4	42	1.58	16.82	1	72	1	108	63	105	63	107
18	นส.ธนาภี ธิ	#####	508/19	16	บัตรทอง	2	1	5	45	1.6	17.58	1	64	1	94	60	88	59	91
19	นส.พรรณน	#####	508/19	16	ปกส	2	1	5	41	1.55	17.07	1	66	1	101	55	103	57	102
20	นายอิทธิพล	#####	508/19	16	ปกส	1	1	5	55	1.73	18.38	1	76.5	1	111	61	103	69	107
21	นางนพรัตน์	#####	508/19	16	ชรก	2	2	5	46	1.59	18.20	1	72	1	99	56	101	55	100
22	นางสุนทร	#####	40/2 นค	15	บัตรทอง	1	2	4	102	1.66	37.02	5	107	2	113	72	116	65	115
23	นส.ณัฐพร	#####	46/2 นค	15	ปกส	1	2	4	65	1.67	23.21	3	86	1	107	73	103	70	105

ภาพที่ 3.2 การตัดเอาทริวิตที่ไม่เกี่ยวข้องกับการคัดกรองสุขภาพ

นอกจากนี้แอทริวิตค่าดัชนีมวลกาย (BMI) เกิดจากการคำนวณของค่าน้ำหนัก (กิโลกรัม) หารด้วยส่วนสูง (เมตร) ยกกำลังสอง ทำให้สามารถตัดเอาทริวิตน้ำหนักและส่วนสูงออกไปได้

แอทริวิตบางส่วนเป็นค่าต่อเนื่องได้ถูกทำให้เป็นค่าไม่ต่อเนื่อง (Discretization) เพื่อประโยชน์ในการนำไปทดสอบกับบางอัลกอริทึมที่ไม่รองรับข้อมูลที่เป็นค่าต่อเนื่อง โดยเจ้าหน้าที่พยาบาลได้มีการให้ความหมายไว้อยู่แล้ว จากการบันทึกข้อมูลมาอยู่ในรูปแบบ Excel ได้แก่

- ค่าดัชนีมวลกาย (BMI) มีค่าที่เป็นไปได้ คือ 1, 2, 3 ซึ่งมีความหมายดังนี้ 1 หมายถึง ผอม (น้อยกว่า 18.5) 2 หมายถึง ปกติ (18.5-22.99) 3 หมายถึง ท้วม (23-24.99) 4 หมายถึง อ้วน (25-29.99) และ 5 หมายถึง อ้วนมาก (มากกว่าหรือเท่ากับ 30)

- ขนาดรอบเอว มีค่าที่เป็นไปได้ คือ 1, 2 ซึ่งมีความหมายดังนี้ 1 หมายถึง ผู้ชายที่มีขนาดรอบเอวน้อยกว่า 90 ซม. ผู้หญิงที่มีขนาดรอบเอวน้อยกว่า 80 ซม. 2 หมายถึงผู้ชายที่มีขนาดรอบเอวมากกว่าหรือเท่ากับ 90 ซม. ผู้หญิงที่มีขนาดรอบเอวมากกว่าหรือเท่ากับ 80 ซม.

- ความดันโลหิต มีค่าที่เป็นไปได้ คือ 1, 2, 3, 4 ซึ่งมีความหมายดังนี้ 1 หมายถึง น้อยกว่า 120/80 มิลลิเมตรปรอท 2 หมายถึง 120-139/80-89 มิลลิเมตรปรอท 3 หมายถึง 140/90 มิลลิเมตรปรอท และ 4 หมายถึง มากกว่าหรือเท่ากับ 180/110 มิลลิเมตรปรอท

- สรุปพฤติกรรมสุขภาพ มีค่าที่เป็นไปได้ คือ 1, 2, 3, 4 ซึ่งมีความหมายดังนี้ 1 หมายถึง คะแนนการประเมินพฤติกรรมสุขภาพน้อยกว่า 60 คะแนน, 2 หมายถึง 60-79 คะแนน, 3 หมายถึง 80-99 คะแนน และ 4 หมายถึง 100 คะแนน

จึงสามารถตัดแอทริบิวต์ที่เป็นค่าต่อเนื่องเหล่านี้ทิ้งไป คงไว้แต่ค่าที่มีการให้ความหมายแล้ว ดังแสดงในภาพที่ 3.3 เมื่อตัดแอทริบิวต์ต่างๆ เหล่านี้ออกไปจะทำให้ข้อมูลมีแอทริบิวต์ทั้งหมด 58 แอทริบิวต์

1	BM	sumBM	hip	sumhip	ความดันครั้ง1	ความดันครั้ง2	สรุปความดัน	sumBP	family1	family2	p1	p2	sumB	ResultBsummary				
2	2.24	3	84	2	100	73	100	70	100	2	1	9	2	2	85	3	2	
3	2.71	4	91	2	132	89	124	79	128	84	2	9	9	2	2	65	2	2
4	2.35	2	77	1	121	81	120	80	121	81	2	9	9	2	2	92	3	2
5	1.58	1	66	1	100	77	97	69	90	73	1	9	9	2	2	87	3	1
6	2.22	2	72	1	126	73	120	70	123	72	2	9	9	2	2	87	3	1
7	1.78	1	61	1	126	73	120	70	123	72	2	8	8	2	2	77	2	1
8	2.12	4	94	2	138	75	127	78	133	77	2	1+2	9	2	2	78	2	2
9	2.61	3	76	1	110	70	110	70	110	70	1	1	1+2	2	2	74	2	2
10	2.03	4	94	2	110	80	110	80	110	80	2	9	9	2	2	71	2	2
11	2.76	2	71	1	110	70	110	70	110	70	1	9	9	2	2	90	3	1
12	2.15	2	79	1	114	81	110	80	112	81	2	8	8	2	2	82	3	1
13	2.73	4	102	2	129	77	125	70	125	74	2	9	9	2	2	78	2	2
14	2.62	4	90	2	127	78	120	70	124	74	2	9	9	1	1	91	3	3
15	2.68	3	90	2	122	62	120	70	121	66	2	8	8	2	1	93	3	3
16	2.65	4	97	2	135	70	131	80	133	75	2	9	9	2	2	70	2	2
17	1.53	2	74	1	112	73	117	65	115	68	1	9	9	2	2	66	2	1
18	1.82	1	72	1	108	63	105	63	107	63	1	9	9	2	2	76	2	2
19	1.58	1	64	1	94	60	88	59	81	60	1	2+3	9	2	2	82	3	2
20	1.07	1	66	1	101	55	103	57	102	56	1	2+3	9	2	2	82	3	2
21	1.38	1	65	1	111	61	103	69	107	65	1	2+3	9	2	2	74	2	2
22	1.20	1	72	1	99	56	101	55	100	56	1	2+4	10	2	2	76	2	3
23	3.02	5	107	2	113	72	116	65	115	69	1	9	9	1	2	58	1	3
24	1.21	2	66	1	107	72	103	70	105	73	1	0	0	2	2	88	2	2

ภาพที่ 3.3 การตัดแอทริบิวต์ที่ถูกแปลงเป็นค่าไม่ต่อเนื่อง (Discretization)

### 3.2.2 การแก้ไขข้อมูลให้ถูกต้อง

โดยทั่วไปข้อมูลที่ถูกจัดเก็บอาจมีความผิดปกติต่างๆ ได้ เช่น ข้อมูลบางแอทริบิวต์ขาดหายไป (missing value) ขาดแอทริบิวต์ที่น่าสนใจ หรือขาดรายละเอียดของข้อมูล เป็นข้อมูลรบกวน (noisy data) เช่น ข้อมูลมีค่าผิดพลาด (error) หรือมีค่าผิดปกติ (Outliers) ซึ่งอาจเกิดจากการตอบแบบสอบถามที่ไม่ถูกต้อง การบันทึกข้อมูลลงคอมพิวเตอร์ผิดพลาดของเจ้าหน้าที่ เป็นต้น

จากการตรวจสอบข้อมูลที่ใช้ในการศึกษาครั้งนี้จากไฟล์ Excel โดยใช้คำสั่งเรียงลำดับจากมากไปน้อยในแต่ละแอทริบิวต์เพื่อให้เห็นข้อมูลที่มีค่าผิดพลาดได้ง่าย ไม่พบการขาดหายของข้อมูลแต่พบปัญหาข้อมูลมีค่าผิดพลาด โดยพบว่ามีอยู่ 3 แอทริบิวต์ที่มีค่าผิดพลาด คือ แอทริบิวต์การตรวจเต้านมด้วยตนเอง ผลตรวจเต้านมด้วยตนเอง และผลตรวจมะเร็งปากมดลูก ซึ่งมีค่าที่เป็นไปได้คือ 1, 2, 3 เท่านั้น แต่ปรากฏข้อมูลตัวเลข 5 และ 6 เป็นต้น ดังแสดงในภาพที่ 3.4 จึงแก้ไข

ให้เป็นค่าข้อมูลที่ถูกต้องโดยดูจากข้อมูลในเรคอร์ดอื่นที่มีลักษณะใกล้เคียงกัน เช่น ข้อมูลแอทริบิวต์การตรวจเต้านมด้วยตนเองมีค่าเท่ากับ 5 ต้องพิจารณาข้อมูลจากแอทริบิวต์ที่เกี่ยวข้องนั้นคือ ความถี่ในการตรวจเต้านมด้วยตนเอง กับ ผลตรวจเต้านมด้วยตนเอง มีค่าเท่ากับ 6 และ 3 เมื่อพิจารณาข้อมูลเรคอร์ดอื่นที่มีค่าความถี่ในการตรวจเต้านมด้วยตนเอง กับผลตรวจเต้านมด้วยตนเองที่มีค่าเท่ากับ 6 และ 3 พบว่าส่วนใหญ่ค่าการตรวจเต้านมด้วยตนเองมีค่าเท่ากับ 2 ดังนั้นจึงแก้ไขค่าการตรวจเต้านมด้วยตนเองจาก 5 เป็น 2 โดยในขั้นตอนนี้พบข้อมูลที่มีค่าผิดพลาดและแก้ไขให้ข้อมูลอยู่ในช่วงที่ถูกต้องได้ทั้งหมดจำนวน 4 เรคอร์ด

	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS
1	p15	p16	treat	smoke	drink	exercis	eat	drive	sex	chest	level	che:st	ch:cervic	test ce	b1	b2	b3	b4	b5	b6	b7
2	2	2	4	2	2	1	1	2	6	2	5	1	5	6	5	5	3	3	3	3	3
3	1	2	1	2	1	1	2	1	5	2	6	5	3	5	5	3	5	5	5	3	3
4	2	2	1	2	2	2	1+2+3	1	4	1	3	1	1	5	5	5	5	5	5	3	3
5	2	2	4	1	1	1	1+4	2	5	5	6	3	5	3	5	3	5	3	3	5	3
6	2	2	4	2	2	5	3	2	6	3	7	3	6	3	5	5	3	3	5	5	3
7	2	2	4	2	2	2	3	2	6	3	7	3	6	3	5	5	3	3	3	3	3
8	2	2	4	2	2	2	3	2	6	3	7	3	6	3	5	5	3	3	3	3	3
9	2	2	4	2	2	2	6	2	6	3	7	3	6	3	3	5	3	3	3	3	3
10	2	2	4	1	1	1	1	3	5	3	7	3	6	3	3	3	3	3	3	3	3
11	2	2	4	2	2	2	6	2	5	3	7	3	6	3	5	5	5	5	0	0	3
12	2	2	4	2	1	3	1+3	2	5	3	7	3	6	3	3	3	3	3	5	5	3
13	2	2	4	2	2	2	5	3	4	3	7	3	6	3	5	5	5	3	3	3	3

ภาพที่ 3.4 ค่าข้อมูลที่ผิดพลาดที่ต้องทำการแก้ไขในขั้นตอนการแก้ไขข้อมูลให้ถูกต้อง

นอกจากนี้ยังมีข้อมูลบางเรคอร์ดมีข้อมูลที่ไม่สอดคล้องกับตัวแบบสอบถาม เช่น แอทริบิวต์ Treat (การปฏิบัติตนหากมีประวัติเจ็บป่วย เช่น เป็นโรคเบาหวาน โรคความดันโลหิตสูง โรคตับ โรคอัมพาต โรคหัวใจ โรคไขมันในเลือดผิดปกติ เป็นต้น) มีค่าที่เป็นไปได้ ได้แก่ 1 หมายถึง รับประทานอยู่/ปฏิบัติตามที่แพทย์แนะนำ 2 หมายถึง รับประทานแต่ไม่สม่ำเสมอ 3 หมายถึง เคยรักษา ขณะนี้ไม่รักษา/หายาทานเอง และ 4 หมายถึง ไม่มีประวัติการเจ็บป่วย พบว่า

- มีข้อมูล 2 เรคอร์ดที่ค่า Treat ไม่ได้เป็น 4 แต่ไม่เคยมีประวัติการเจ็บป่วยเลย จึงทำการแก้ไขค่าแอทริบิวต์ Treat ให้เป็น 4

- มีข้อมูล 3 เรคอร์ดที่มีประวัติการเจ็บป่วยแต่ค่า Treat เป็น 4 จึงทำการแก้ไขแอทริบิวต์ Treat ให้เป็น 1 เพราะข้อมูลใกล้เคียงส่วนใหญ่มีค่า Treat เป็น 1



จากการสอบถามผู้เชี่ยวชาญถึงกฎเกณฑ์ในการสรุปผลการคัดกรองสุขภาพที่มีค่าที่เป็นได้ได้แก่ 1 หมายถึง กลุ่มปกติ 2 หมายถึง กลุ่มเสี่ยง และ 3 หมายถึง กลุ่มป่วย ผู้เชี่ยวชาญได้ให้ความหมายของกลุ่มป่วยคือ มีประวัติการเจ็บป่วยด้วยโรคเบาหวาน โรคความดันโลหิตสูง โรคตับ โรคอัมพาต โรคหัวใจ โรคไขมันในเลือดผิดปกติ หรือโรคใดโรคหนึ่ง พบว่ามีข้อมูล 2 เรคอร์ดที่มีประวัติการป่วยเป็นโรคดังกล่าวแต่เอทริบิวต์สรุปผลการคัดกรองสุขภาพไม่ได้เป็น 3 จึงทำการแก้ไขเป็น 3

### 3.2.3 การแปลงข้อมูล

เนื่องจากข้อมูลในบางเอทริบิวต์ผู้ตอบแบบสอบถามสามารถเลือกตอบได้มากกว่า 1 ค่า ได้แก่ ประวัติการเจ็บป่วยของบิดามารดา (family1) ประวัติการเจ็บป่วยของพี่น้องสายตรง (family2) และรสชาติของอาหารที่ชอบ (eat) ข้อมูลที่บันทึกอยู่ในไฟล์ Excel มีการเก็บข้อมูลเป็นตัวเลขคำตอบที่เชื่อมต่อกันด้วยเครื่องหมาย + ดังภาพที่ 3.5

	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	weight	height	BMI	sumBMI	hip	sumhip	sumBP	family1	family2	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13
2	70	1.68	24.8	3	90	2	1	2	9	2	1	2	2	2	2	2	2	2	2	2	2	2
3	80	1.65	29.38	4	92	2	1+2	3	1	1	1	2	2	2	2	2	2	2	2	2	2	2
4	50	1.49	22.52	2	82	2	3	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
5	63	1.56	25.89	4	86	2	1+2	1+2	2	1	1	2	2	2	2	2	2	2	2	2	2	2
6	70	1.65	25.71	4	89	2	2	5	2	2	2	2	2	2	1	2	2	2	2	2	2	2
7	57	1.6	22.27	2	79	1	2	5	9	2	2	2	2	2	1	2	2	2	2	2	2	2
8	66	1.55	27.47	4	83	2	2	5	2	2	2	2	2	2	2	2	2	2	2	2	2	2
9	68	1.6	26.56	4	81	2	2	5	2	2	2	2	2	2	1	2	2	2	2	2	2	2
10	47	1.48	21.46	2	78	1	2	9	1	2	1	2	2	2	2	1	2	2	2	2	2	2
11	56	1.58	22.43	2	82	1	3	8	2	2	2	1	2	2	2	2	2	2	2	2	2	2
12	69	1.53	29.48	4	89	2	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2
13	80	1.6	31.25	5	108	2	3	8	8	1	1	2	2	2	2	2	2	2	2	2	2	2
14	75	1.75	24.49	3	97	2	3	9	9	2	1	2	2	2	1	2	2	2	2	2	2	2
15	75	1.6	29.3	4	89	1	2	5	2	2	1	2	2	2	2	2	2	2	2	2	2	2
16	52	1.48	23.74	3	73	1	1	5	2	2	2	2	2	2	2	2	2	2	2	2	2	2
17	52	1.5	23.11	3	83	2	1+2	1+2	2	2	1	2	2	2	2	1	2	2	2	2	2	2
18	72	1.6	28.13	4	92	2	3	9	1	1	1	2	2	2	2	2	2	2	2	2	2	2
19	78	1.59	30.85	5	89	1	2	2	1+2	2	1	2	2	2	2	2	2	2	2	2	2	2
20	65	1.5	28.89	4	81	2	2	2	9	1	1	2	2	2	2	2	2	2	2	2	2	2
21	65	1.5	28.89	4	81	2	2	2	9	1	1	2	2	2	2	2	2	2	2	2	2	2
22	75	1.75	24.49	3	90	2	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2
23	66	1.59	26.11	4	87	1	3	2	9	2	2	1	2	2	2	2	2	2	2	2	2	2
24	52	1.51	22.81	2	79	1	2	8	2	2	1	2	2	2	1	1	2	2	2	2	2	2
25	70	1.62	26.67	4	89	2	2	9	1+2	2	1	2	2	2	2	2	2	2	2	2	2	2
26	64	1.71	21.89	2	95	2	3	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
27	76	1.67	27.25	4	87	2	1	9	3	2	2	2	2	2	2	1	2	2	2	2	2	2
28	77	1.66	27.94	4	85	2	1	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2
29	72	1.74	23.78	3	87	1	2	1+2	9	2	1	2	2	2	2	1	2	2	2	2	2	2
30	51	1.48	23.28	3	80	2	3	9	1	1	1	2	2	2	2	2	2	2	2	2	2	2
31	52	1.5	23.11	3	81	2	3	9	9	2	1	2	2	2	2	2	2	2	2	2	2	2
32	57	1.49	25.67	4	94	2	1	9	3	2	1	2	2	2	2	2	2	2	2	2	2	2
33	58	1.48	26.48	4	87	2	2	9	9	2	1	2	2	2	2	2	2	2	2	2	2	2
34	54	1.4	22.05	2	77	1	2	9	2	2	1	2	2	2	2	2	2	2	2	2	2	2

ภาพที่ 3.5 ข้อมูลที่เก็บค่ามากกว่า 1 ค่าในเอทริบิวต์เดียวกัน

เพื่อที่จะทำการวิเคราะห์ข้อมูลเหล่านี้ได้อย่างถูกต้อง จึงต้องทำการแปลงข้อมูลที่อยู่ในรูปแบบ String ให้กลายเป็นข้อมูลแบบ Binary (มีค่า 0 และ 1) โดยแตกออกเป็นหลายแอทริบิวต์ตามจำนวนค่าของข้อมูลที่เป็นไปได้ทั้งหมด เช่น แอทริบิวต์ประวัติการเจ็บป่วยของพี่น้องสายตรงผู้กรอกแบบสอบถามสามารถเลือกคำตอบได้ทั้งหมด 10 ตัวเลือก จึงได้นำข้อมูลในคอลัมน์ family2 มาแยกเป็น 10 คอลัมน์ย่อยตามจำนวนตัวเลือกที่ผู้ใช้สามารถเลือกตอบได้ดังภาพที่ 3.6 ส่งผลให้มีแอทริบิวต์ทั้งหมดเพิ่มขึ้นเป็น 81 แอทริบิวต์

	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	family2	fb1	fb2	fb3	fb4	fb5	fb6	fb7	fb8	fb9	fb10
2	2+6	0	1	0	0	0	1	0	0	0	0
3	2+5	0	1	0	0	1	0	0	0	0	0
4	2+5	0	1	0	0	1	0	0	0	0	0
5	2+5	0	1	0	0	1	0	0	0	0	0
6	2+3	0	1	1	0	0	0	0	0	0	0
7	1+7	1	0	0	0	0	0	1	0	0	0
8	1+4	1	0	0	1	0	0	0	0	0	0
9	1+2+6	1	1	0	0	0	1	0	0	0	0
10	1+2+4	1	1	0	1	0	0	0	0	0	0
11	1+2	1	1	0	0	0	0	0	0	0	0
12	1+2	1	1	0	0	0	0	0	0	0	0
13	1+2	1	1	0	0	0	0	0	0	0	0
14	1+2	1	1	0	0	0	0	0	0	0	0
15	1+2	1	1	0	0	0	0	0	0	0	0
16	1+2	1	1	0	0	0	0	0	0	0	0
17	1+2	1	1	0	0	0	0	0	0	0	0
18	1+2	1	1	0	0	0	0	0	0	0	0
19	1+2	1	1	0	0	0	0	0	0	0	0
20	1+2	1	1	0	0	0	0	0	0	0	0
21	1+2	1	1	0	0	0	0	0	0	0	0
22	1+2	1	1	0	0	0	0	0	0	0	0
23	1+2	1	1	0	0	0	0	0	0	0	0
24	1+2	1	1	0	0	0	0	0	0	0	0
25	1+2	1	1	0	0	0	0	0	0	0	0
26	1+2	1	1	0	0	0	0	0	0	0	0
27	1+2	1	1	0	0	0	0	0	0	0	0
28	1+2	1	1	0	0	0	0	0	0	0	0
29	1+2	1	1	0	0	0	0	0	0	0	0
30	1+2	1	1	0	0	0	0	0	0	0	0
31	1+2	1	1	0	0	0	0	0	0	0	0
32	1+2	1	1	0	0	0	0	0	0	0	0
33	1+2	1	1	0	0	0	0	0	0	0	0

ภาพที่ 3.6 การแปลงข้อมูลที่มีหลายค่าในหนึ่งแอทริบิวต์ให้กลายเป็นข้อมูลแบบ Binary

### 3.2.4 การวิเคราะห์ข้อมูล

ข้อมูลที่ใช้ในการศึกษาทั้งหมดสามารถวิเคราะห์แยกตามแต่ละแอทริบิวต์ได้ดังรายละเอียดในตารางที่ 3.1 - 3.58

ตารางที่ 3.1 ผลการวิเคราะห์ค่าของข้อมูลเพศ

ลำดับ	เพศ	จำนวน (เรคอร์ด)	ร้อยละ
1	เพศชาย	442	41.27
2	เพศหญิง	629	58.73
	รวม	1,071	100.00

ตารางที่ 3.2 ผลการวิเคราะห์ค่าของข้อมูลสถานภาพสมรส

ลำดับ	สถานภาพสมรส	จำนวน (เรคอร์ด)	ร้อยละ
1	โสด	354	33.05
2	สมรส	579	54.06
3	ม่าย/หย่า/แยก	138	12.89
	รวม	1,071	100.00

ตารางที่ 3.3 ผลการวิเคราะห์ค่าของข้อมูลการศึกษา

ลำดับ	การศึกษา	จำนวน (เรคอร์ด)	ร้อยละ
1	ไม่ได้เรียน	10	0.93
2	ประถมศึกษา	355	33.15
3	มัธยมศึกษา	360	33.61
4	อนุปริญญา	87	8.12
5	ปริญญาตรี	239	22.32
6	สูงกว่าปริญญาตรี	20	1.87
	รวม	1,071	100.00



ตารางที่ 3.4 ผลการวิเคราะห์ค่าของข้อมูลดัชนีมวลกาย

ลำดับ	ดัชนีมวลกาย	จำนวน (เรคอร์ด)	ร้อยละ
1	ผอม(น้อยกว่า 18.5)	73	6.81
2	ปกติ(18.5-22.99)	397	37.07
3	ท้วม(23-24.99)	296	27.64
4	อ้วน(25-29.99)	246	22.97
5	อ้วนมาก(มากกว่าหรือเท่ากับ 30)	59	5.51
	รวม	1,071	100.00

ตารางที่ 3.5 ผลการวิเคราะห์ค่าของข้อมูลขนาดรอบเอว

ลำดับ	ขนาดรอบเอว	จำนวน (เรคอร์ด)	ร้อยละ
1	ชายน้อยกว่า 90 ซม. หญิงน้อยกว่า 80 ซม.	626	58.45
2	ชายมากกว่าเท่ากับ 90 ซม. หญิง มากกว่าเท่ากับ 80 ซม.	445	41.55
	รวม	1,071	100.00

ตารางที่ 3.6 ผลการวิเคราะห์ค่าของข้อมูลความดันโลหิต

ลำดับ	ความดันโลหิต	จำนวน (เรคอร์ด)	ร้อยละ
1	น้อยกว่า 120/80	505	47.15
2	120-139/80-89	448	41.83
3	มากกว่าหรือเท่ากับ 140/90	115	10.74
4	มากกว่าหรือเท่ากับ 180/110	3	0.28
	รวม	1,071	100.00

ตารางที่ 3.7 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยของบิดามารดา

ลำดับ	ประวัติการเจ็บป่วยของบิดามารดา	จำนวน (เรคอร์ด)	ร้อยละ
1	เบาหวาน	168	13.82
2	ความดันโลหิตสูง	295	24.26
3	โรคเก๊าท์	21	1.73
4	ไตวายเรื้อรัง	9	0.74
5	กล้ามเนื้อหัวใจตาย	13	1.07
6	เส้นเลือดสมอง	14	1.15
7	ถุงลมโป่งพอง	5	0.41
8	ไม่ทราบ	162	13.32
9	ไม่มี	519	42.68
10	อื่นๆ	10	0.82
	รวม	1,216	100.00

ตารางที่ 3.8 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยของพี่น้อง (สายตรง)

ลำดับ	ประวัติการเจ็บป่วยของพี่น้อง (สายตรง)	จำนวน (เรคอร์ด)	ร้อยละ
1	เบาหวาน	66	5.91
2	ความดันโลหิตสูง	111	9.94
3	โรคเก๊าท์	9	0.81
4	ไตวายเรื้อรัง	2	0.18
5	กล้ามเนื้อหัวใจตาย	4	0.36
6	เส้นเลือดสมอง	3	0.27
7	ถุงลมโป่งพอง	1	0.09
8	ไม่ทราบ	190	17.01
9	ไม่มี	725	64.91
10	อื่นๆ	6	0.54
	รวม	1,117	100.00

ตารางที่ 3.9 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคเบาหวาน

ลำดับ	ประวัติการเจ็บป่วยโรคเบาหวาน	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	64	5.98
2	ไม่มี	901	84.12
3	ไม่เคยตรวจ	106	9.90
	รวม	1,071	100.00

ตารางที่ 3.10 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคความดันโลหิตสูง

ลำดับ	ประวัติการเจ็บป่วยโรคความดันโลหิตสูง	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	163	15.22
2	ไม่มี	837	78.15
3	ไม่เคยตรวจ	71	6.63
	รวม	1,071	100.00

ตารางที่ 3.11 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคตับ

ลำดับ	ประวัติการเจ็บป่วยโรคตับ	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	11	1.03
2	ไม่มี	941	87.86
3	ไม่เคยตรวจ	119	11.11
	รวม	1,071	100.00

ตารางที่ 3.12 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคอัมพาต

ลำดับ	ประวัติการเจ็บป่วยโรคเบาหวาน	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	3	0.28
2	ไม่มี	1,036	96.73
3	ไม่เคยตรวจ	32	2.99
	รวม	1,071	100.00

ตารางที่ 3.13 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยโรคหัวใจ

ลำดับ	ประวัติการเจ็บป่วยโรคหัวใจ	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	21	1.96
2	ไม่มี	943	88.05
3	ไม่เคยตรวจ	107	9.99
	รวม	1,071	100.00

ตารางที่ 3.14 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเจ็บป่วยไขมันในเลือดผิดปกติ

ลำดับ	ประวัติการเจ็บป่วยไขมันในเลือดผิดปกติ	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	43	4.01
2	ไม่มี	930	86.83
3	ไม่เคยตรวจ	98	9.15
	รวม	1,071	100.00

ตารางที่ 3.15 ผลการวิเคราะห์ค่าของข้อมูลประวัติการเป็นแผลที่เท้า/ตัดขา(จากเบาหวาน)

ลำดับ	ประวัติการเจ็บป่วยแผลที่เท้า/ตัดขา (จากเบาหวาน)	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	2	0.19
2	ไม่มี	1,064	99.35
3	ไม่เคยตรวจ	5	0.46
	รวม	1,071	100.00

ตารางที่ 3.16 ผลการวิเคราะห์ค่าของข้อมูลประวัติการคลอดบุตรน้ำหนักเกิน 4 กิโลกรัม

ลำดับ	ประวัติการคลอดบุตรน้ำหนักเกิน 4 กิโลกรัม	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	6	0.56
2	ไม่มี	1,061	99.07
3	ไม่เคยตรวจ	4	0.37
	รวม	1,071	100.00

ตารางที่ 3.17 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการคื่น้ำน้อยและมาก

ลำดับ	ประวัติอาการคื่น้ำน้อยและมาก	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	151	14.10
2	ไม่มี	917	85.62
3	ไม่เคยตรวจ	3	0.28
	รวม	1,071	100.00

ตารางที่ 3.18 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการปัสสาวะกลางคืน 3 ครั้งขึ้นไป

ลำดับ	ประวัติการอาการปัสสาวะกลางคืน 3 ครั้งขึ้นไป	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	98	9.15
2	ไม่มี	970	90.57
3	ไม่เคยตรวจ	3	0.28
	รวม	1,071	100.00

ตารางที่ 3.19 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการกินจุแต่พอมลง

ลำดับ	ประวัติอาการกินจุแต่พอมลง	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	19	1.77
2	ไม่มี	1,048	97.85
3	ไม่เคยตรวจ	4	0.37
	รวม	1,071	100.00

ตารางที่ 3.20 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการน้ำหนักลด / อ่อนเพลีย

ลำดับ	ประวัติการอาการน้ำหนักลด / อ่อนเพลีย	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	22	2.05
2	ไม่มี	1,045	97.57
3	ไม่เคยตรวจ	4	0.37
	รวม	1,071	100.00

ตารางที่ 3.21 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการเป็นแผลที่ริมฝีปากบ่อและหายยาก

ลำดับ	อาการเป็นแผลที่ริมฝีปากบ่อและหายยาก	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	1	0.09
2	ไม่มี	1,065	99.44
3	ไม่เคยตรวจ	5	0.47
	รวม	1,071	100.00

ตารางที่ 3.22 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการคันตามผิวหนังและอวัยวะสืบพันธุ์

ลำดับ	อาการคันตามผิวหนังและอวัยวะสืบพันธุ์	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	9	0.84
2	ไม่มี	1,057	98.69
3	ไม่เคยตรวจ	5	0.47
	รวม	1,071	100.00

ตารางที่ 3.23 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการตาพร่ามัวต้องเปลี่ยนแว่นบ่อ

ลำดับ	ประวัติอาการตาพร่ามัวต้องเปลี่ยนแว่นบ่อ	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	33	3.08
2	ไม่มี	1,033	96.45
3	ไม่เคยตรวจ	5	0.47
	รวม	1,071	100.00

ตารางที่ 3.24 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการชาปลายมือปลายเท้าโดยไม่ทราบสาเหตุ

ลำดับ	ประวัติการเจ็บป่วยโรคเบาหวาน	จำนวน (เรคอร์ด)	ร้อยละ
1	มี	50	4.67
2	ไม่มี	1,014	94.68
3	ไม่เคยตรวจ	7	0.65
	รวม	1,071	100.00

ตารางที่ 3.25 ผลการวิเคราะห์ข้อมูลการปฏิบัติตนถ้ามีประวัติการเจ็บป่วยในตารางที่ 3.11 ถึง 3.26

ลำดับ	การปฏิบัติตนเมื่อเจ็บป่วย	จำนวน (เรคอร์ด)	ร้อยละ
1	รับการรักษาอยู่/ปฏิบัติตามที่แพทย์แนะนำ	212	19.79
2	รับการรักษาแต่ไม่สม่ำเสมอ	0	0.00
3	เคยรักษา ขณะนี้ไม่รักษา/หายาทานเอง	0	0.00
4	ไม่มีประวัติการเจ็บป่วย	859	80.21
	รวม	1,071	100.00

ตารางที่ 3.26 ผลการวิเคราะห์ค่าของข้อมูลการสูบบุหรี่

ลำดับ	การสูบบุหรี่	จำนวน (เรคอร์ด)	ร้อยละ
1	สูบบุหรี่	102	9.52
2	ไม่สูบบุหรี่	958	89.45
3	เคยสูบบุหรี่แต่เลิกแล้ว	11	1.03
	รวม	1,071	100.00

ตารางที่ 3.27 ผลการวิเคราะห์ค่าของข้อมูลการดื่มเครื่องดื่มแอลกอฮอล์

ลำดับ	การดื่มเครื่องดื่มแอลกอฮอล์	จำนวน (เรคอร์ด)	ร้อยละ
1	ดื่ม	177	16.53
2	ไม่ดื่ม	844	78.80
3	เคยดื่มแต่เลิกแล้ว	50	4.67
	รวม	1,071	100.00

ตารางที่ 3.28 ผลการวิเคราะห์ค่าของข้อมูลการออกกำลังกาย

ลำดับ	การออกกำลังกาย	จำนวน (เรคอร์ด)	ร้อยละ
1	ไม่ออกกำลังกาย	192	17.93
2	ออกกำลังกายน้อยกว่าสัปดาห์ละ 3 ครั้ง	420	39.22
3	ออกกำลังกายสัปดาห์ละ 3 ครั้ง ครั้งละ 30 นาที สม่ำเสมอ	111	10.36
4	ออกกำลังกายมากกว่า 3 ครั้ง ครั้งละ 30 นาที สม่ำเสมอ	78	7.28
5	ออกกำลังกายทุกวัน ครั้งละ 30 นาที	270	25.21
	รวม	1,071	100.00

ตารางที่ 3.29 ผลการวิเคราะห์ค่าของข้อมูลรสชาติของอาหารที่ชอบ

ลำดับ	รสชาติของอาหารที่ชอบ	จำนวน (เรคอร์ด)	ร้อยละ
1	หวาน	238	17.32
2	เค็ม	199	14.48
3	มัน	162	11.79
4	เปรี้ยว	136	9.90
5	จืด	313	22.78
6	ไม่ชอบทุกข้อ	326	23.73
	รวม	1374	100.00

ตารางที่ 3.30 ผลการวิเคราะห์ค่าของข้อมูลการขับขี่หรือโดยสารรถจักรยานยนต์/รถยนต์

ลำดับ	การขับขี่หรือโดยสารรถจักรยานยนต์/รถยนต์	จำนวน (เรคอร์ด)	ร้อยละ
1	ไม่ขับขี่/ไม่โดยสาร	126	11.76
2	ขับขี่/โดยสารและใส่หมวกกันน็อค/คาดเข็มขัดนิรภัยทุกครั้ง	756	70.59
3	ขับขี่/โดยสารและใส่หมวกกันน็อค/คาดเข็มขัดนิรภัยบางครั้ง	178	16.62
4	ขับขี่/โดยสารและใส่หมวกกันน็อค/คาดเข็มขัดนิรภัยนานๆครั้ง (ใส่เฉพาะเมื่อมีด่านตรวจ)	11	1.03
	รวม	1,071	100.00



ตารางที่ 3.31 ผลการวิเคราะห์ค่าของข้อมูลการใช้ถุงยางอนามัยเมื่อมีเพศสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยา

ลำดับ	การใช้ถุงยางอนามัยเมื่อมีเพศสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยา	จำนวน (เรคอร์ด)	ร้อยละ
1	ใช้ทุกครั้ง	56	5.23
2	ใช้เมื่อถูกร้องขอ	0	0.00
3	ไม่ใช้	10	0.93
4	ไม่เคยมีสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยาของตนเอง	408	38.10
5	ไม่ตอบ	369	34.45
6	ไม่เคยมีเพศสัมพันธ์	228	21.29
	รวม	1,071	100.00

ตารางที่ 3.32 ผลการวิเคราะห์ค่าของข้อมูลการตรวจเต้านมด้วยตนเอง

ลำดับ	การตรวจเต้านมด้วยตนเอง	จำนวน (เรคอร์ด)	ร้อยละ
1	ได้	442	41.27
2	ไม่ได้	606	56.58
3	ไม่แน่ใจ	23	2.15
	รวม	1,071	100.00

ตารางที่ 3.33 ผลการวิเคราะห์ค่าของข้อมูลความถี่ในการตรวจเต้านมด้วยตนเอง

ลำดับ	ความถี่ในการตรวจเต้านมด้วยตนเอง	จำนวน (เรคอร์ด)	ร้อยละ
1	เดือนละครั้ง	204	19.05
2	เดือนเว้นเดือน	58	5.42
3	3-4 เดือนครั้ง	58	5.42
4	6 เดือนครั้ง	18	1.68
5	ปีละครั้ง	76	7.09
6	ไม่เคยตรวจ	629	58.73
7	จำไม่ได้	28	2.61
	รวม	1,071	100.00

ตารางที่ 3.34 ผลการวิเคราะห์ค่าของข้อมูลผลตรวจเต้านมด้วยตนเอง

ลำดับ	ผลตรวจเต้านมด้วยตนเอง	จำนวน (เรคอร์ด)	ร้อยละ
1	ปกติ	400	37.35
2	ผิดปกติ	2	0.19
3	จำไม่ได้	669	62.46
	รวม	1,071	100.00

ตารางที่ 3.35 ผลการวิเคราะห์ค่าของข้อมูลความถี่ในการตรวจมะเร็งปากมดลูก

ลำดับ	ความถี่ในการตรวจมะเร็งปากมดลูก	จำนวน (เรคอร์ด)	ร้อยละ
1	ทุกปี	161	15.03
2	ทุก 2 ปี	27	2.52
3	ทุก 3-5 ปี	17	1.59
4	มากกว่า 5 ปี	30	2.80
5	ไม่เคยตรวจ	836	78.06
	รวม	1,071	100.00

ตารางที่ 3.36 ผลการวิเคราะห์ค่าของข้อมูลผลตรวจมะเร็งปากมดลูก

ลำดับ	ผลตรวจมะเร็งปากมดลูก	จำนวน (เรคอร์ด)	ร้อยละ
1	ปกติ	226	21.10
2	ผิดปกติ	0	0.00
3	ไม่เคยตรวจ	845	78.90
	รวม	1,071	100.00

ตารางที่ 3.37 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารครบ 5 หมู่

ลำดับ	การกินอาหารครบ 5 หมู่	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	4	0.37
2	ครั้งคราว	207	19.33
3	ไม่เคยเลย	860	80.30
	รวม	1,071	100.00

ตารางที่ 3.38 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารมือเช้าเป็นมือหลัก

ลำดับ	การกินอาหารมือเช้า	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	15	1.40
2	ครั้งคราว	156	14.57
3	ไม่เคยเลย	900	84.03
	รวม	1,071	100.00

ตารางที่ 3.39 ผลการวิเคราะห์ค่าของข้อมูลการกินผักมากกว่าวันละ 3 ทัพพี

ลำดับ	การกินผัก	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	19	1.77
2	ครั้งคราว	440	41.08
3	ไม่เคยเลย	612	57.14
	รวม	1,071	100.00

ตารางที่ 3.40 ผลการวิเคราะห์ค่าของข้อมูลการกินผลไม้วันละ 2-3 ส่วน

ลำดับ	การกินผลไม้	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	18	1.68
2	ครั้งคราว	492	45.94
3	ไม่เคยเลย	561	52.38
	รวม	1,071	100.00

ตารางที่ 3.41 ผลการวิเคราะห์ค่าของข้อมูลการกินปลาอย่างน้อยวันละ 1 มื้อ

ลำดับ	การกินปลา	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	16	1.49
2	ครั้งคราว	435	40.62
3	ไม่เคยเลย	620	57.89
	รวม	1,071	100.00

ตารางที่ 3.42 ผลการวิเคราะห์ค่าของข้อมูลการกินเนื้อสัตว์ไม่ติดมันสัปดาห์ละ 2-3 มื้อ

ลำดับ	การกินผัก	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	46	4.30
2	ครั้งคราว	636	59.38
3	ไม่เคยเลย	389	36.32
	รวม	1,071	100.00

ตารางที่ 3.43 ผลการวิเคราะห์ค่าของข้อมูลการดื่มนมรสจืดหรือนมถั่วเหลืองผสมงาดำรสจืดวันละ 1-2 แก้ว

ลำดับ	การดื่มนมรสจืดหรือนมถั่วเหลือง	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	122	11.39
2	ครั้งคราว	516	48.18
3	ไม่เคยเลย	433	40.43
	รวม	1,071	100.00

ตารางที่ 3.44 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารมือเย็นห่างจากเวลานอนไม่น้อยกว่า 4 ชม.

ลำดับ	การกินอาหารมือเย็น	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	43	4.02
2	ครั้งคราว	502	46.87
3	ไม่เคยเลย	526	49.11
	รวม	1,071	100.00

ตารางที่ 3.45 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารประเภทต้ม นึ่ง ลวก อบ ยำหรือหมก

ลำดับ	กินอาหารประเภทต้ม นึ่ง ลวก อบ ยำหรือหมก	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	10	0.93
2	ครั้งคราว	486	45.38
3	ไม่เคยเลย	575	53.69
	รวม	1,071	100.00

ตารางที่ 3.46 ผลการวิเคราะห์ค่าของข้อมูลการหลีกเลี่ยงอาหารไขมันสูง

ลำดับ	การหลีกเลี่ยงอาหารไขมันสูง	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	37	3.45
2	ครั้งคราว	435	40.62
3	ไม่เคยเลย	599	55.93
	รวม	1,071	100.00

ตารางที่ 3.47 ผลการวิเคราะห์ค่าของข้อมูลการหลีกเลี่ยงของหวานและขนมที่มีแป้งและน้ำตาล

ลำดับ	การหลีกเลี่ยงของหวานและขนม	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	23	2.15
2	ครั้งคราว	482	45.00
3	ไม่เคยเลย	566	52.85
	รวม	1,071	100.00

ตารางที่ 3.48 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารรสจืด

ลำดับ	การกินอาหารรสจืด	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	56	5.23
2	ครั้งคราว	479	44.72
3	ไม่เคยเลย	536	50.05
	รวม	1,071	100.00

ตารางที่ 3.49 ผลการวิเคราะห์ค่าของข้อมูลการเลือกดื่มน้ำเปล่าแทนน้ำอัดลมหรือน้ำหวาน

ลำดับ	การดื่มน้ำเปล่า	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	27	2.52
2	ครั้งคราว	259	24.18
3	ไม่เคยเลย	785	73.30
	รวม	1,071	100.00

ตารางที่ 3.50 ผลการวิเคราะห์ค่าของข้อมูลการหลีกเลี่ยงเครื่องดื่มน้ำที่มีแอลกอฮอล์

ลำดับ	หลีกเลี่ยงเครื่องดื่มน้ำที่มีแอลกอฮอล์	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	89	8.31
2	ครั้งคราว	238	22.22
3	ไม่เคยเลย	744	69.47
	รวม	1,071	100.00

ตารางที่ 3.51 ผลการวิเคราะห์ค่าของข้อมูลอารมณ์ดีไม่เครียด

ลำดับ	อารมณ์ดีไม่เครียด	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	13	1.21
2	ครั้งคราว	206	19.23
3	ไม่เคยเลย	852	79.55
	รวม	1,071	100.00

ตารางที่ 3.52 ผลการวิเคราะห์ค่าของข้อมูลการนอนหลับไม่น้อยกว่าวันละ 7-8 ชั่วโมง

ลำดับ	การนอนหลับ	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	40	3.73
2	ครั้งคราว	212	19.79
3	ไม่เคยเลย	819	76.47
	รวม	1,071	100.00

ตารางที่ 3.53 ผลการวิเคราะห์ค่าของข้อมูลการออกกำลังกายสัปดาห์ละ 5 วันหรือสัปดาห์ละ 5 ครั้ง

ลำดับ	การออกกำลังกายสัปดาห์ละ 5 วัน	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	196	18.30
2	ครั้งคราว	359	33.52
3	ไม่เคยเลย	516	48.18
	รวม	1,071	100.00

ตารางที่ 3.54 ผลการวิเคราะห์ค่าของข้อมูลการออกกำลังกายวันละ 30 นาที

ลำดับ	การออกกำลังกายวันละ 30 นาที	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	190	17.74
2	ครั้งคราว	322	30.07
3	ไม่เคยเลย	559	52.19
	รวม	1,071	100.00

ตารางที่ 3.55 ผลการวิเคราะห์ค่าของข้อมูลขณะออกกำลังกายหัวใจเร็วขึ้นกว่าปกติและเหงื่อซึม

ลำดับ	ขณะออกกำลังกายหัวใจเร็วขึ้นกว่าปกติและเหงื่อซึม	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	253	23.62
2	ครั้งคราว	276	25.77
3	ไม่เคยเลย	542	50.61
	รวม	1,071	100.00

ตารางที่ 3.56 ผลการวิเคราะห์ค่าของข้อมูลทุกครั้งที่ทำรอบเอวได้เกณฑ์ปกติ คือ เพศหญิงไม่เกิน 80 ซม. และเพศชายไม่เกิน 90 ซม.

ลำดับ	การวัดรอบเอว	จำนวน (เรคอร์ด)	ร้อยละ
1	ประจำ	166	15.50
2	ครั้งคราว	199	18.58
3	ไม่เคยเลย	706	65.92
	รวม	1,071	100.00

ตารางที่ 3.57 ผลการวิเคราะห์ค่าของข้อมูลสรุปพฤติกรรมสุขภาพ

ลำดับ	สรุปพฤติกรรมสุขภาพ	จำนวน (เรคอร์ด)	ร้อยละ
1	ควรปรับเปลี่ยนพฤติกรรมให้ดีขึ้น	68	6.35
2	พฤติกรรมด้านสุขภาพดีปานกลาง	421	39.31
3	พฤติกรรมด้านสุขภาพดี	557	52.01
4	พฤติกรรมด้านสุขภาพดีมาก	25	2.33
	รวม	1,071	100.00

ตารางที่ 3.58 ผลการวิเคราะห์ค่าของข้อมูลสรุปผลการคัดกรอง

ลำดับ	สรุปผลการคัดกรอง	จำนวน (เรคอร์ด)	ร้อยละ
1	ปกติ	108	10.08
2	กลุ่มเสี่ยง	755	70.49
3	กลุ่มป่วย	208	19.42
	รวม	1,071	100.00

### 3.3 การศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม

ในงานนี้ได้ศึกษาเทคนิคการหากฎ Association Rule เพื่อหาความสัมพันธ์ของโรคกับโรค มีทั้งหมด 6 โรคได้แก่ โรคเบาหวาน (p1) โรคความดันโลหิตสูง (p2) โรคตับ (p3) โรคอัมพาต (p4) โรคหัวใจ (p5) และ โรคไขมันในเลือดผิดปกติ (p6) และเทคนิคการจำแนกกลุ่ม Classification เพื่อจำแนกกลุ่มประชากรออกเป็นกลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย โดยการศึกษาเทคนิคอัลกอริทึมต่างๆจากเอกสารและนำมาทดลองโดยใช้โปรแกรม Weka และได้เลือกเทคนิคที่เหมาะสมโดยมีรายละเอียดดังนี้

เทคนิคการหาความสัมพันธ์ (Association Rule) ได้แก่

- Apriori (Agrawal et al., 1993) ให้ผลลัพธ์ในรูปแบบกฎความสัมพันธ์ที่นำค่าทุกค่าที่เป็นไปได้มาสร้างเป็นกฎ ผลลัพธ์ที่ได้ส่วนใหญ่จะให้ผลเป็นค่าปฏิเสธยกตัวอย่างเช่น ไม่เป็นโรคที่ 1 และไม่เป็โรคที่ 2 แล้วจะไม่เป็นโรคที่ 4 ซึ่งกฎความสัมพันธ์ลักษณะนี้จะไม่ค่อยมีประโยชน์

- FP Growth (Kantardzic M., 2003) ให้ผลลัพธ์ในรูปแบบกฎความสัมพันธ์โดยในโปรแกรม Weka แต่ละแอทริบิวต์ที่นำมาพิจารณาจะต้องนำค่าที่สนใจศึกษาไปแจกแจงไว้ที่ตำแหน่งสุดท้ายในส่วนการประกาศแอทริบิวต์ในแฟ้ม .arff ยกตัวอย่างเช่น ค่าของโรคต่างๆ กำหนดให้ค่า 1 หมายถึงเป็นโรค, 0 หมายถึงไม่เป็นโรค ซึ่งหากต้องการจะหาความสัมพันธ์ของการเป็นโรคต่างๆ ควรแจกแจงค่าแอทริบิวต์ในแฟ้มข้อมูลในรูปแบบ @attribute p1 {0,1}

เทคนิคการจำแนกกลุ่มข้อมูล (Classification) มีหลากหลายชนิด และแต่ละเทคนิคก็ให้ผลลัพธ์ที่มีรูปแบบแตกต่างกัน เทคนิคที่ศึกษามีดังนี้

- Decision Tree เป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างต้นไม้ (Decision Trees) ซึ่ง Decision Trees สามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ก่อนล่วงหน้า เรียกว่า Training set ได้อัตโนมัติ และกลุ่มของ



รายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ด้วยรูปแบบของ Tree โครงสร้างประกอบด้วย Root Node, Child และ Leaf Node (<http://www.no-poor.com/dssandos/Chapter5-dss.htm>)

- SVM (Support Vector Machine) เป็นสมการที่ใช้ในการจำแนกค่าคุณลักษณะของสองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกัน โดยการสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้น เส้นตรงใดเป็นเส้นตรงที่ดีที่สุดโดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น จะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด แต่ในความเป็นจริงข้อมูลสองกลุ่มไม่ได้วางอยู่ในพื้นที่คุณลักษณะที่แยกกันชัดเจนและแบ่งได้โดยเส้นตรงแต่ข้อมูลอาจจะจับกลุ่มกันในตำแหน่งต่างๆ ดังนั้น จึงเป็นปัญหาทำให้ไม่สามารถที่จะใช้สมการซัพพอร์ตเวกเตอร์แมชชีนส์แบบเชิงเส้นได้ จึงต้องมีเครื่องมือมาช่วยให้ออกมาเหล่านี้นั้นเรียกว่าพื้นที่ที่เรียกว่า พื้นที่หลายมิติ (Higher Dimensional Space) และใช้หลักการการแบ่งข้อมูลโดยเส้นตรงและเส้นขอบมาแบ่งข้อมูลอีกครั้งหนึ่ง (ภัทรพงศ์ พ., 2009)

- Neural Network เป็นแนวคิดให้คอมพิวเตอร์ทำงานคล้ายสมองของมนุษย์ โดยเปลี่ยนจากการประมวลผลตามลำดับ (Sequential Processing) ให้เป็นการประมวลผลแบบคู่ขนานได้ (Parallel Processing) โครงสร้างของ Neural Network เป็นกลุ่มของ Node ที่เชื่อมโยงถึงกันในแต่ละ Layer คือ Input layer, Hidden layer และ output layer มีลักษณะการทำงานโดยแต่ละ Process จะรับ Input เข้าไปคำนวณ และสร้าง Output ออกมาในลักษณะที่ไม่ใช่การทำงานแบบเชิงเส้นตรง เพราะ Input แต่ละตัวจะถูกให้ลำดับความสำคัญของค่าไม่เท่ากัน ค่าของ Output ที่ได้ จากการเชื่อมโยงกันนี้จะถูกนำมาเปรียบเทียบกับ Output ที่ได้ตั้งเอาไว้ ถ้าค่าที่ออกมาเกิดความคลาดเคลื่อนจะนำไปสู่การปรับค่าน้ำหนัก (weight) ของค่าที่ใส่ไว้ให้แต่ละ Input

ในการศึกษาครั้งนี้ต้องการนำโมเดลที่ได้ไปพัฒนาระบบคัดกรองสุขภาพประชาชนแบบอัตโนมัติ ดังนั้นเทคนิคการจำแนกกลุ่มที่นำมาใช้ในการศึกษาครั้งนี้จึงเป็นเทคนิคที่ให้ผลลัพธ์เป็นกฎเพื่อให้สามารถทำความเข้าใจได้ง่ายและสามารถนำไปใช้พัฒนาระบบได้ เทคนิคที่ศึกษาแบบอัลกอริทึมในกลุ่ม Decision Tree ได้แก่

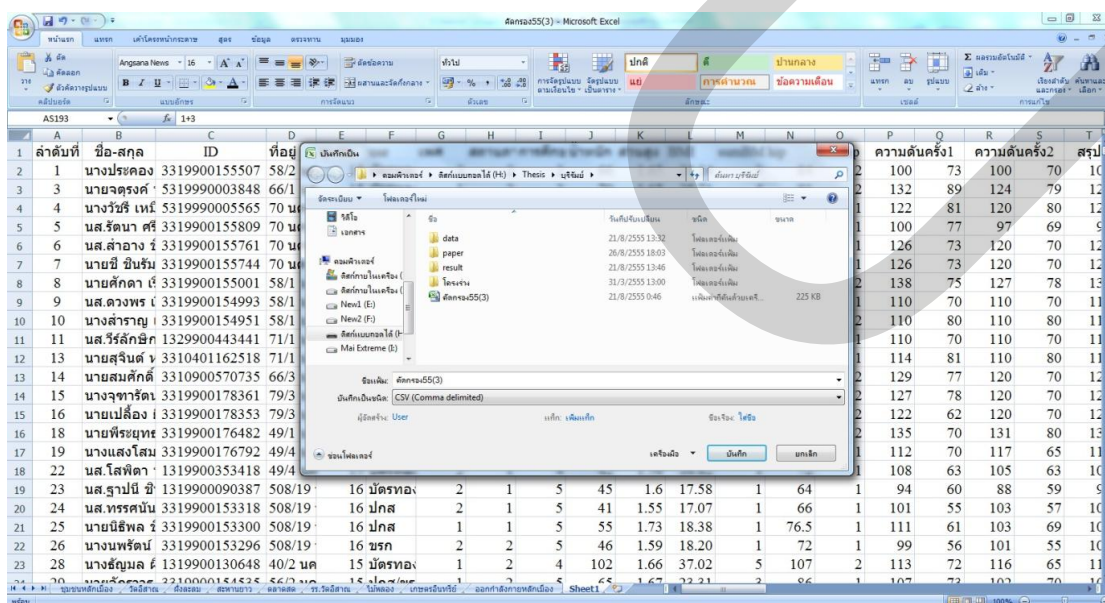
1. โมเดลการจำแนกกลุ่มแบบ Decision Tree: C4.5 (Kantardzic M., 2003) หรือในโปรแกรม Weka ใช้ชื่อว่า J48 เป็นการจัดกลุ่มโดยผลลัพธ์ที่ได้อยู่ในรูปของต้นไม้การตัดสินใจที่มีกิ่งต้นไม้จำนวนมาก โดยค่าของข้อมูลของแอทริบิวต์ที่มีค่าต่อเนื่องจะใช้เครื่องหมาย “<”, “<=”, “=”, “>” ในการพิจารณา ส่วนข้อมูลที่เป็นค่าไม่ต่อเนื่องจะใช้เครื่องหมาย “=” และใช้

เงื่อนไข “AND” ในการเชื่อมโยงระหว่างแอทริบิวต์ จะเชื่อมโยงไปจนถึงแอทริบิวต์สุดท้ายที่จะแบ่งกลุ่มได้

2. โมเดลการจำแนกกลุ่มแบบ Decision Rule: Partial Rules (Kantardzic M., 2003) หรือในโปรแกรม Weka ใช้ชื่อว่า PART เป็นการจัดกลุ่มโดยผลลัพธ์ที่ได้จะอยู่ในรูปของกฎที่นำเอาข้อมูลในแต่ละแอทริบิวต์มาเป็นเงื่อนไขในการตัดสินใจลักษณะของข้อมูลที่ใกล้เคียงกันให้มาอยู่ในกลุ่มเดียวกันโดยค่าของข้อมูลของแอทริบิวต์ที่มีค่าต่อเนื่องจะใช้เครื่องหมาย “<”, “<=”, “=”, “>” ในการพิจารณา ส่วนข้อมูลที่เป็นค่าไม่ต่อเนื่องจะใช้เครื่องหมาย “=” และใช้เงื่อนไข “AND” ในการเชื่อมโยงระหว่างแอทริบิวต์ จะเชื่อมโยงไปจนถึงแอทริบิวต์สุดท้ายที่จะแบ่งกลุ่มได้

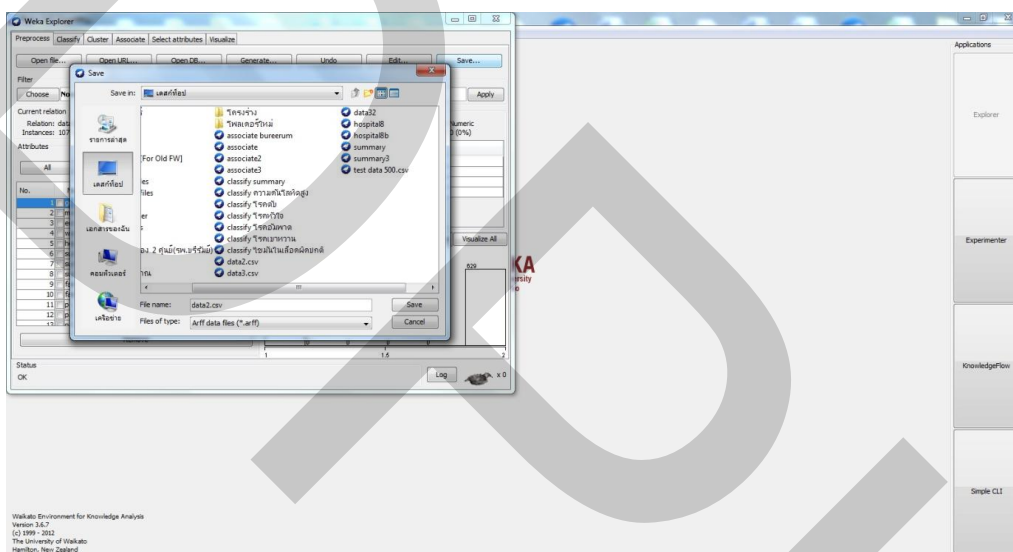
การศึกษาครั้งนี้ทำการทดสอบโมเดลโดยใช้โปรแกรมประยุกต์ Weka เวอร์ชัน 3.6.7 ข้อมูลที่ใช้ในการทดสอบมีจำนวน 1,071 เรคอร์ด ในการทดสอบได้ใช้การทดสอบแบบ 10 folds cross-validation ซึ่งเป็นการแบ่งกลุ่มข้อมูลออกเป็น 10 กลุ่ม ในแต่ละรอบจะนำกลุ่มข้อมูลจำนวน 9 กลุ่มเป็นกลุ่มศึกษา (Training Set) และกลุ่มข้อมูลที่เหลือเป็นกลุ่มทดสอบ (Testing Set) โดยจะทำซ้ำจำนวน 10 รอบ เพื่อเปลี่ยนกลุ่มทดสอบให้ครบทุกกลุ่ม และนำผลลัพธ์ที่ได้มาหาค่าเฉลี่ย โดยเกณฑ์ที่ใช้ในการพิจารณาคุณภาพของโมเดลที่ได้จากแต่ละอัลกอริทึม จะพิจารณาจากเปอร์เซ็นต์ความถูกต้อง (Precision), ค่าระลึก (Recall) และค่าความเหวี่ยง (F-measure) ที่มีค่ามากที่สุด

การนำเข้าข้อมูลของโปรแกรม Weka 3.6.7 ต้องทำการแปลงไฟล์ข้อมูลจาก .xls ให้เป็นนามสกุล csv ก่อนดังภาพที่ 3.7 จึงจะสามารถนำข้อมูลเข้าในโปรแกรม Weka ได้

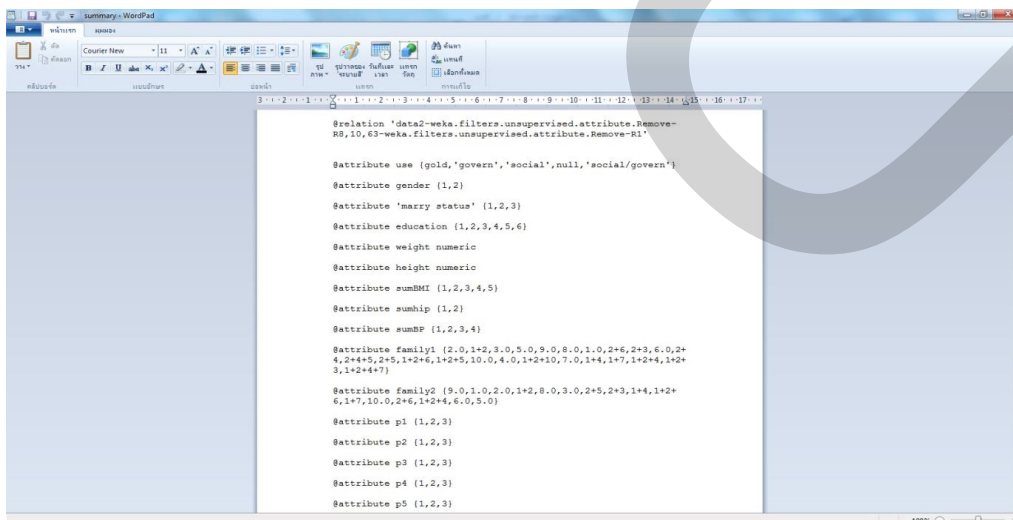


ภาพที่ 3.7 การบันทึกข้อมูลในรูปแบบนามสกุล .csv

ขั้นตอนต่อมาให้เปิดไฟล์ .csv ขึ้นมาในโปรแกรม Weka แล้วบันทึกเพิ่มข้อมูลเป็น .arff เพื่อให้เป็นรูปแบบมาตรฐานสำหรับโปรแกรม Weka ดังภาพที่ 3.8 เพิ่มข้อมูล .arff จะสามารถแจกแจงข้อมูลในแต่ละแอทริบิวต์ให้เป็นค่าไม่ต่อเนื่องได้ ซึ่งโดยปกติเมื่อทำการบันทึกเป็น .arff แล้ว ค่าปริยายของแต่ละแอทริบิวต์ที่เป็นตัวเลขจะถูกกำหนดให้เป็น numeric ให้เข้าไปแก้ไขโดยเปิดเพิ่มข้อมูลด้วยโปรแกรม WordPad ดังภาพที่ 3.9 โดยเปิดโปรแกรมเลือก Applications >> Explorer >> Open file... เลือกไฟล์ข้อมูลที่ต้องการนำเข้าข้อมูล

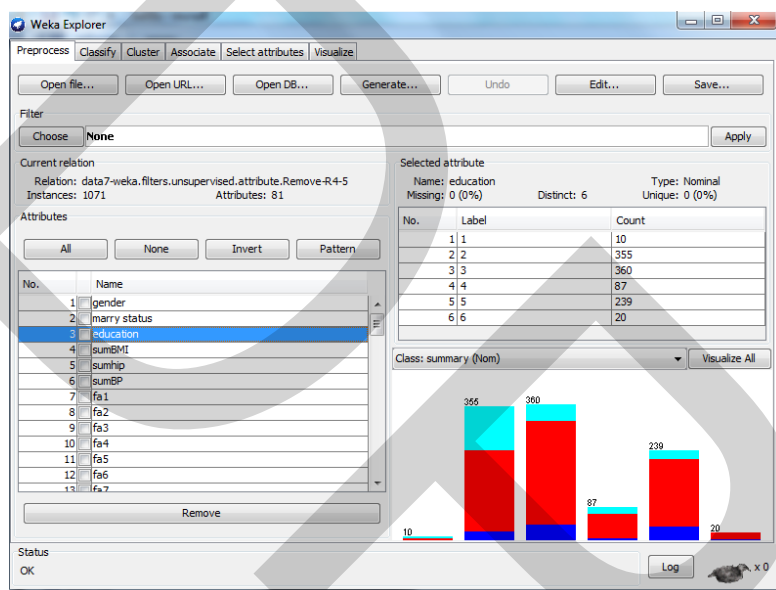


ภาพที่ 3.8 การแปลงข้อมูลให้เป็นรูปแบบนามสกุล .arff

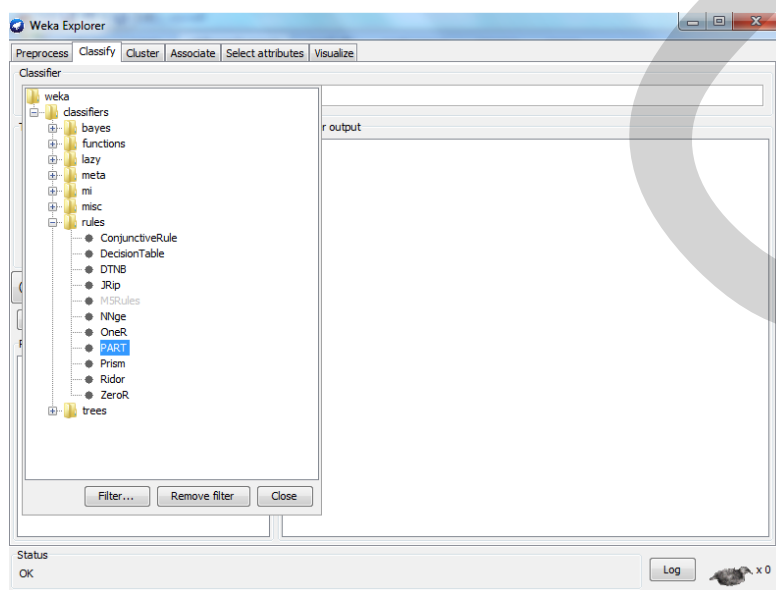


ภาพที่ 3.9 การแจกแจงข้อมูลในแต่ละแอทริบิวต์ใน โปรแกรม WordPad

เมื่อแปลงเพิ่มข้อมูล .arff แล้ว สามารถ double click เพื่อเข้าสู่โปรแกรม Weka ได้เลย หน้าจอแรกที่พบคือหน้าจอ Preprocess ดังภาพที่ 3.10 หน้าจอนี้จะทำการวิเคราะห์หาจำนวนข้อมูลของข้อมูลที่ใช้ในการวิจัยเพื่อให้ทราบจำนวนเปอร์เซ็นต์ต่อข้อมูลทั้งหมด และค่าเฉลี่ยของข้อมูลในแต่ละแอททริบิวต์

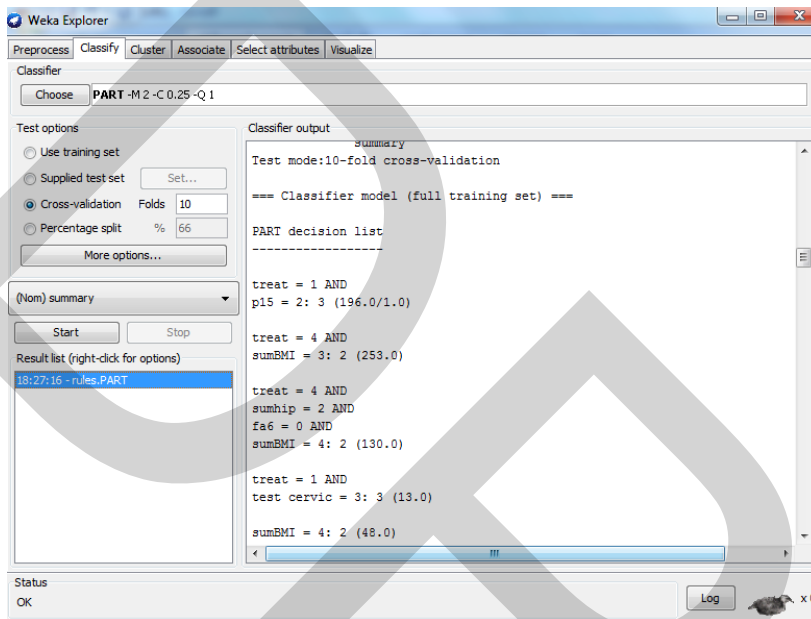


ภาพที่ 3.10 หน้าจอการวิเคราะห์ข้อมูล

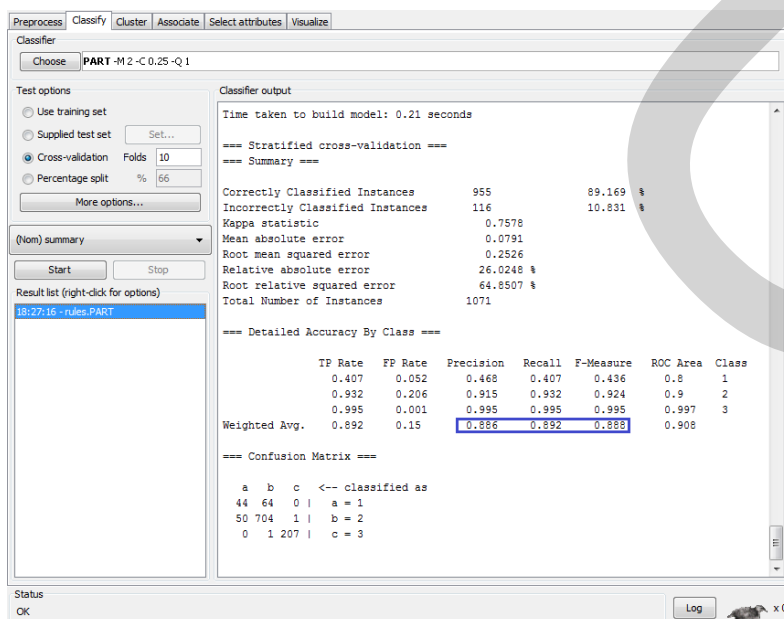


ภาพที่ 3.11 การเลือกเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูล

การจำแนกกลุ่มข้อมูลทำได้โดยเลือก Classify >> Choose >> rules >> เลือกเทคนิคการจำแนกข้อมูลที่ต้องการ เช่น PART ดังภาพที่ 3.11 เลือก Cross-validation Folds 10 แล้วกดปุ่ม Start จะได้ผลลัพธ์เป็นกฎดังภาพที่ 3.12



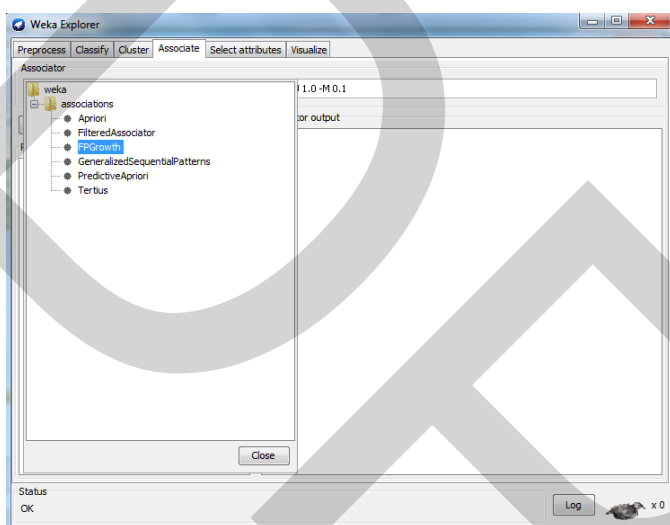
ภาพที่ 3.12 หน้าจอแสดงผลการจำแนกกลุ่มข้อมูลด้วยเทคนิค Decision Rule: Partial Rules



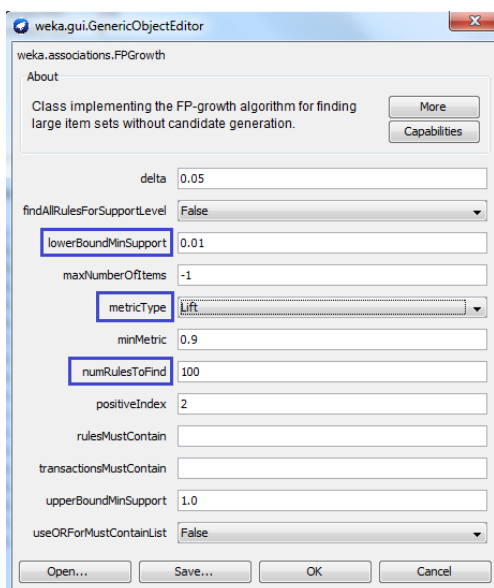
ภาพที่ 3.13 หน้าจอแสดงผลค่าความถูกต้องของโมเดล

ค่าที่ใช้เป็นเกณฑ์ในการเลือกเทคนิคที่เหมาะสมคือค่า Precision, Recall และ F-Measure โดยจะพิจารณาจากค่าเฉลี่ยถ่วงน้ำหนัก (Weighted Avg.) จากทุกๆคลาส ดังแสดงในภาพที่ 3.13

การหาความสัมพันธ์ของข้อมูลทำได้โดยเลือก Associate >> Choose >> เลือกเทคนิคการหาความสัมพันธ์ที่ต้องการ เช่น FPGrowth ดังภาพที่ 3.14



ภาพที่ 3.14 การเลือกเทคนิคที่ใช้หาความสัมพันธ์ของข้อมูล



ภาพที่ 3.15 การตั้งค่ารายละเอียดในการหาความสัมพันธ์ของข้อมูล

เมื่อเลือกอัลกอริทึมที่จะใช้ในการหากฎความสัมพันธ์แล้ว สามารถตั้งค่ารายละเอียดต่างๆ ได้ดังภาพที่ 3.15 โดยส่วนที่ต้องให้ความสำคัญในการตั้งค่ารายละเอียดก่อนการหาความสัมพันธ์ เช่น lowerBoundMinSupport คือ การกำหนด Minimum Support, metricType คือ การกำหนดการเรียงลำดับของผลลัพธ์ตามชนิดของตัววัดความถูกต้อง, numRulesToFind คือ การกำหนดจำนวนกฎที่จะแสดงผลเป็นต้น จากนั้นกดปุ่ม Start จะได้ผลลัพธ์กฎความสัมพันธ์ออกมาดังภาพที่ 3.16

The screenshot shows the Weka software interface with the 'Associate' tab selected. The 'FPGrowth' algorithm is chosen, and the 'Start' button has been pressed. The 'Associator output' window displays the following information:

```

=== Run information ===
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
             p1
             p2
             p3
             p4
             p5
             p6

=== 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 3.16 หน้าจอแสดงผลลัพธ์กฎความสัมพันธ์



## 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 ค่าคือ 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. p4=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 ผลลัพธ์บางส่วนจากการหาความสัมพันธ์แบบ 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) จากความสัมพันธ์นี้ทำให้ทราบว่าเมื่อเป็นโรคดังกล่าวแล้วมักจะมีโรคแทรกซ้อนตามมา เช่น ถ้าคนไข้เป็นโรคหัวใจและโรคความดันโลหิตสูงอยู่แล้ว ก็ต้องระวังจะเป็นโรคเบาหวานและโรคไขมันในเลือดผิดปกติซึ่งสามารถประยุกต์ใช้ความรู้นี้ในการประชาสัมพันธ์ผู้ป่วยเพื่อป้องกันการเกิดโรคแทรกซ้อน เช่น พยายามลดพฤติกรรมเสี่ยงที่จะทำให้เกิดโรคเบาหวานและไขมันในเลือดผิดปกติเสีย เพื่อลดความเสี่ยงที่จะเป็นโรคดังกล่าว

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

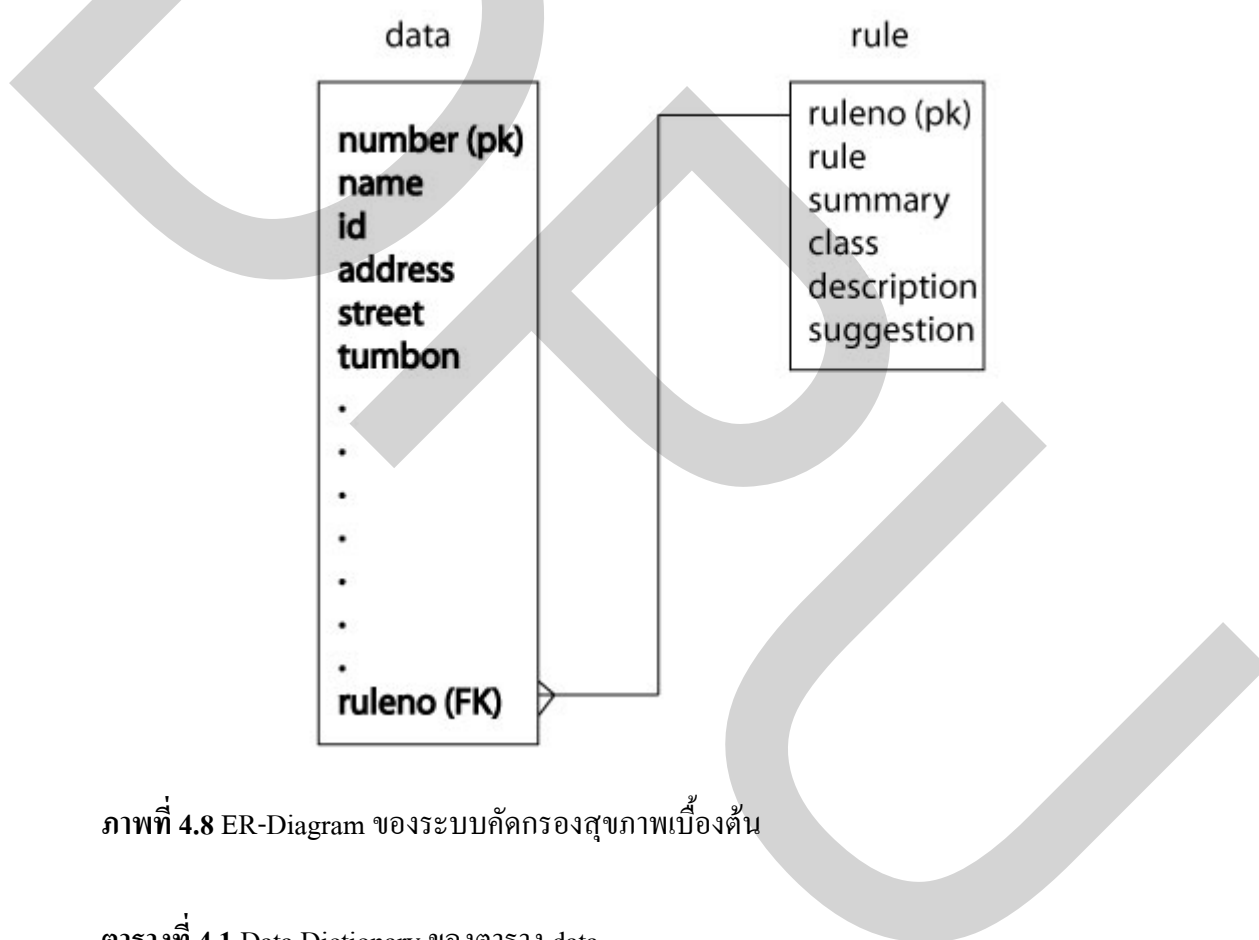
ภาพที่ 4.7 ผลลัพธ์การจำแนกกลุ่มหลังจากตัดแอทริบิวต์การปฏิบัติคนเมื่อเจ็บป่วย

สรุปกฎที่ได้จากผลการทดสอบการจำแนกกลุ่มข้อมูลแบบ Decision Rule: Partial Rules เป็นการจัดกลุ่มที่เหมาะสมที่สุด 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)	การขับขีหรือโดยสาร
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)

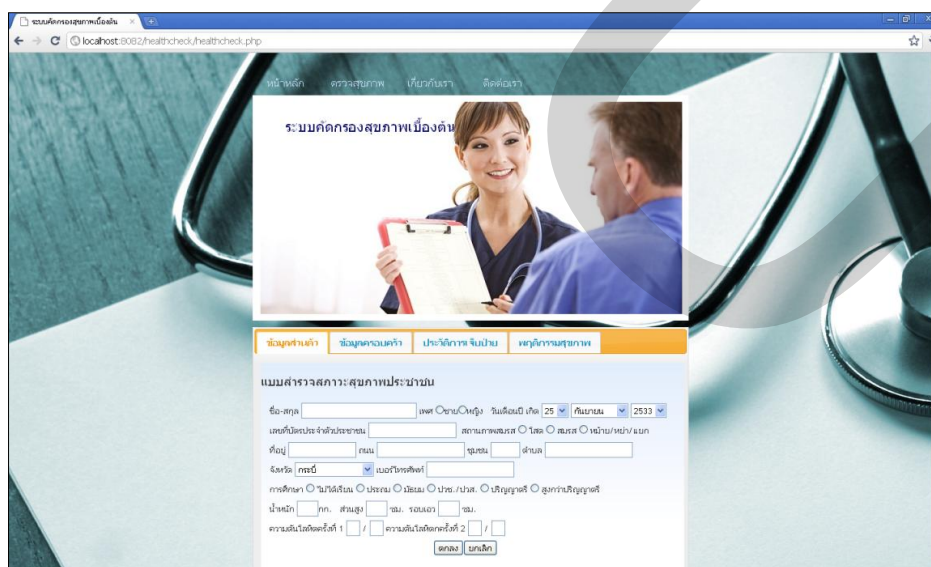
88	b3		int(1)	กินผักมากกว่าวันละ 3 ท็อปพี
89	b4		int(1)	กินผลไม้วันละ 2-3 ส่วน(1ส่วนเท่ากับ 6-8 คำ)
90	b5		int(1)	กินปลาอย่างน้อยวันละ 1 มื้อ
91	b6		int(1)	กินเนื้อสัตว์ไม่ติดมัน สัปดาห์ละ 2-3 มื้อ
92	b7		int(1)	ดื่มนมรสจืดหรือนมถั่วเหลืองผสมงาดำรสจืดวันละ 1-2 แก้ว
93	b8		int(1)	กินอาหารมื้อเย็นห่างจากเวลานอนไม่น้อยกว่า 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	ruleno	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

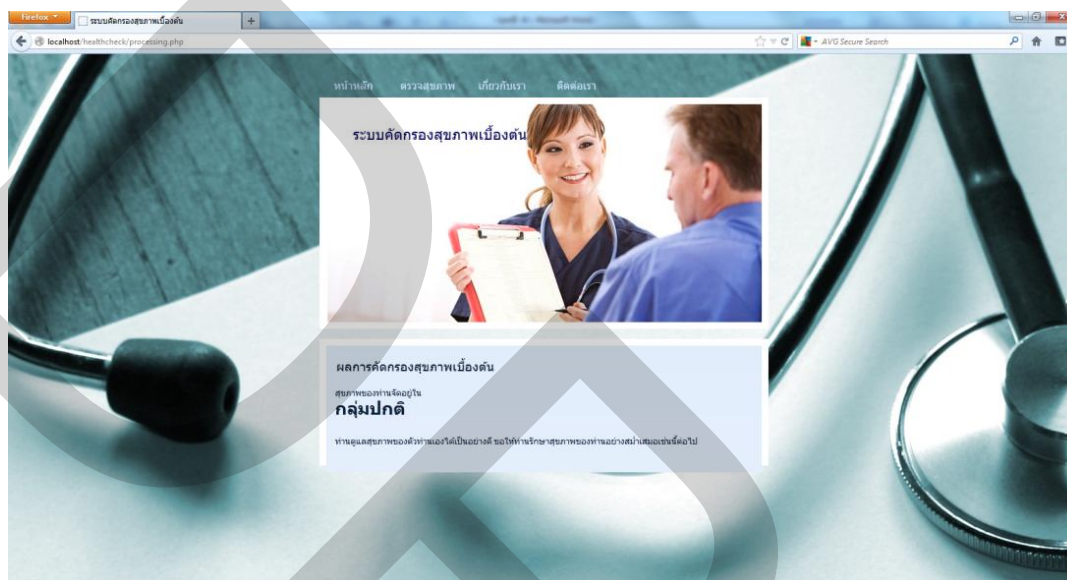
#### 4.2.2 การออกแบบหน้านำเข้าสู่ข้อมูล

หน้าจอหลักประกอบด้วยเมนู หน้าหลักสำหรับแนะนำให้ผู้ใช้งานรู้จักระบบคัดกรองสุขภาพเบื้องต้น วิธีการใช้งานและประโยชน์ที่จะได้รับ หน้าตรวจสอบสุขภาพสำหรับผู้ใช้งานที่กรอกข้อมูลสุขภาพดังภาพที่ 4.10 เมื่อทำการบันทึกข้อมูลเสร็จระบบจะประมวลผลแล้วแจ้งผลการคัดกรองสุขภาพเบื้องต้นให้ทราบทันทีดังภาพที่ 4.11-4.14 หน้าเกี่ยวกับเราสำหรับให้ข้อมูลรายละเอียดต่างๆเกี่ยวกับที่มาที่ไปของระบบ ติดต่อเราสำหรับติดต่อผู้ดูแลระบบ

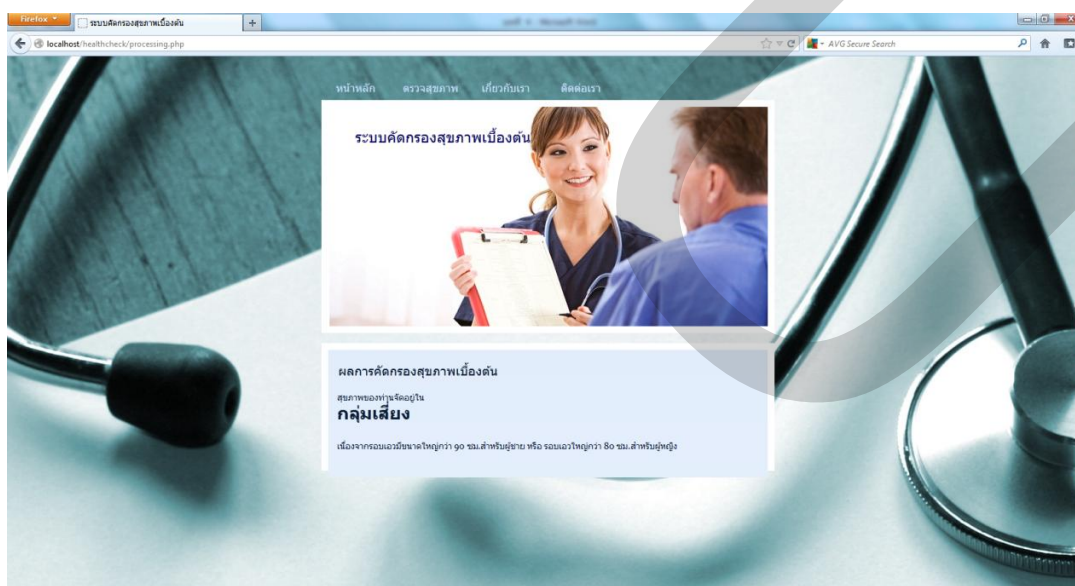


ภาพที่ 4.9 หน้าจอแบบสำรวจสภาวะสุขภาพประชาชน

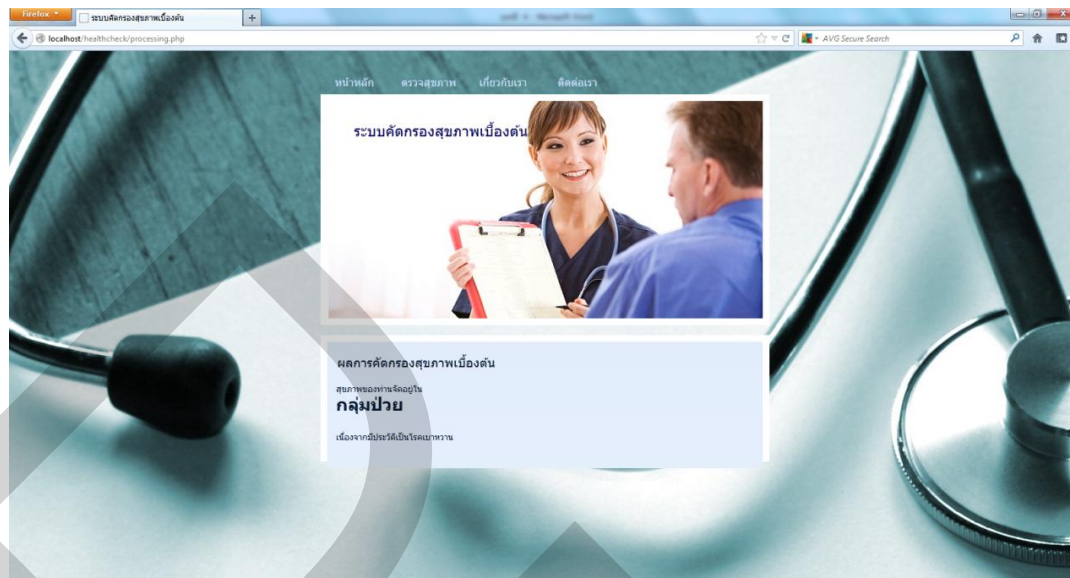
เมื่อผู้ใช้ทำการกรอกข้อมูลสุขภาพตามรายการต่างๆ แล้วทำการคลิกที่ปุ่มบันทึกและประมวลผล โปรแกรมจะทำการประมวลผลและจำแนกกลุ่มว่าเป็นกลุ่มปกติ กลุ่มเสี่ยง หรือ กลุ่มป่วย



ภาพที่ 4.10 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มปกติ



ภาพที่ 4.11 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มเสี่ยง



ภาพที่ 4.12 หน้าจอแสดงผลการคัดกรองสุขภาพกลุ่มป่วย

## บทที่ 5

### สรุปผลการศึกษา และข้อเสนอแนะการวิจัย

#### 5.1 สรุปผลการวิจัย

จากการดำเนินการวิจัยผู้วิจัยได้ดำเนินการวิจัยตามวัตถุประสงค์การวิจัยคือ การศึกษา การเตรียมข้อมูลเพื่อใช้ในการทำเหมืองข้อมูล ได้แก่ การลดขนาดข้อมูล (Data Reduction), การทำความสะอาดข้อมูล (Data Cleaning), การแปลงข้อมูล (Data Transformation) การศึกษาและเลือกโมเดลที่เหมาะสมในการหาความสัมพันธ์ของการเกิดโรคต่างๆ และการแบ่งกลุ่มของข้อมูลเพื่อเป็นแนวทางในการคัดกรองสุขภาพเบื้องต้น โดยเสนอขั้นตอนวิธีการทำเหมืองข้อมูล (Data Mining) เพื่อจำแนกกลุ่มประชากร (Classification) และนำตัวแบบที่ได้ไปประยุกต์ใช้เป็นแนวทางในการคัดกรองสุขภาพเบื้องต้น ซึ่งการทำเหมืองข้อมูลครั้งนี้มีขั้นตอนหลักๆ 5 ขั้นตอนดังนี้

ขั้นตอน 1 การศึกษาและรวบรวมข้อมูล แบบสำรวจสภาวะสุขภาพประชาชนจัดทำโดยศูนย์แพทย์ชุมชน สังกัดโรงพยาบาลบุรีรัมย์ ใช้สำหรับเก็บข้อมูลจากประชาชนทั่วไป โดยในการศึกษานี้เป็นการศึกษาข้อมูลการสำรวจสภาวะสุขภาพประชาชนจากประชาชน เขตเทศบาลเมืองบุรีรัมย์ ปี พ.ศ. 2555 ซึ่งมีจำนวน 1,071 คน โดยข้อมูลที่ได้มาจะอยู่ในรูปแบบของแฟ้มข้อมูล Excel โดยประกอบด้วยแอทริบิวต์จำนวน 75 แอทริบิวต์ เช่น ลำดับที่ ชื่อ-สกุล รหัสบัตรประชาชน ที่อยู่ ชุมชน สิทธิการรักษา เพศ สถานภาพ การศึกษา น้ำหนัก ส่วนสูง ดัชนีมวลกาย ขนาดรอบเอว ความดัน ประวัติการเจ็บป่วยของครอบครัว ประวัติการเจ็บป่วย การรักษา การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การออกกำลังกาย รสชาติของอาหารที่ชอบ การจับชีพจร สัมผัสพันซ์ การตรวจเต้านม การตรวจมะเร็งปากมดลูก พฤติกรรมสุขภาพด้านอาหาร อารมณ์ ออกกำลังกาย และสรุปผลการคัดกรองสุขภาพ เป็นต้น

ขั้นตอน 2 การเตรียมข้อมูล ทำการลดขนาดข้อมูล เนื่องจากข้อมูลมีแอทริบิวต์อยู่เป็นจำนวนมากจึงจำเป็นต้องมีการลดแอทริบิวต์บางตัวที่ไม่เกี่ยวข้องกับการคัดกรองออกไป เช่น ชื่อ-สกุล รหัสบัตรประชาชน เป็นต้น แอทริบิวต์บางส่วนเป็นค่าต่อเนื่อง เช่น ค่าดัชนีมวลกาย (BMI) ได้ถูกทำให้เป็นค่าไม่ต่อเนื่อง (Discretization) มีค่าที่เป็นไปได้ ได้แก่ 1, 2, 3 ซึ่งมีความหมายดังนี้ 1 หมายถึงผอม (น้อยกว่า 18.5) 2 หมายถึง ปกติ (18.5-22.99) 3 หมายถึง ท้วม (23-24.99), 4 หมายถึง อ้วน (25-29.99) และ 5 หมายถึง อ้วนมาก (มากกว่าหรือเท่ากับ 30) จึงสามารถ



ตัดแอทริบิวต์ที่เป็นค่าต่อเนื่องเหล่านี้ทิ้งไป คงไว้แต่ค่าที่มีการให้ความหมายแล้ว ทำให้สามารถลดจำนวนแอทริบิวต์ลงไปได้จาก 73 แอทริบิวต์เป็น 58 แอทริบิวต์ ต่อมาข้อมูลที่ไดยังมีความผิดปกติอยู่ต้องมีการแก้ไขข้อมูลให้ถูกต้อง โดยมีข้อมูลที่ได้รับการแก้ไขให้ถูกต้องทั้งหมด 11 แถว และเนื่องจากข้อมูลในบางแอทริบิวต์ผู้ตอบแบบสอบถามสามารถเลือกได้มากกว่า 1 คำตอบ ได้แก่ ประวัติการเจ็บป่วยของบิดามารดา ประวัติการเจ็บป่วยของพี่น้องสายตรง และรสชาติของอาหารที่ชอบ จึงต้องทำการแปลงข้อมูล (Data Transformation) ที่อยู่ในรูปแบบ String ให้กลายเป็นข้อมูลแบบ Binary (มีค่า 0 และ 1) โดยแตกออกเป็นหลายๆ แอทริบิวต์ตามจำนวนค่าของข้อมูลที่เ็นไป ได้ทั้งหมด ส่งผลให้มีแอทริบิวต์ทั้งหมดเพิ่มขึ้นเป็น 81 แอทริบิวต์

ขั้นตอน 3 การทดสอบและเลือกเทคนิคที่เหมาะสม ในการหากฎความสัมพันธ์ (Association Rule) ได้ทดสอบเทคนิค 2 เทคนิค คือ Apriori และ FP Growth จากการทดสอบพบว่าข้อมูลที่เกี่ยวข้องกับเรื่องของสุขภาพจะเหมาะสมกับเทคนิค FP Growth มากกว่า Apriori เพราะจะให้ผลลัพธ์เป็นกฎที่มีแต่ค่า positive (มีประวัติการเป็นโรค) เท่านั้น ในขณะที่ Apriori จะให้ผลลัพธ์ที่เป็นกฎที่มีทั้งค่า positive และ negative (ไม่มีประวัติการเป็นโรค) ผสมกัน ซึ่งไม่มีประโยชน์มากนักในการนำมาใช้เพื่อทำนายการเกิดโรคใหม่ในอนาคตของผู้ป่วย ส่วนการจำแนกกลุ่มข้อมูล (Classification) ได้ทดสอบเทคนิคในการจำแนกกลุ่ม 2 เทคนิคที่ให้ผลลัพธ์เป็นกฎ เนื่องจากต้องนำกฎไปพัฒนาเป็นระบบคัดกรองคือ โมเดลการจำแนกกลุ่มแบบ Decision Tree: C4.5 หรือ ในโปรแกรม Weka เรียกว่า โมเดล Classifiers tree J48 และ โมเดลการจำแนกกลุ่มแบบ Decision Tree: Partial Rules หรือในโปรแกรม Weka เรียกว่า โมเดล Classifiers PART จากการทดสอบการจำแนกกลุ่มทั้ง 3 เทคนิคพบว่า เทคนิค Decision Tree: Partial Rules ให้ผลลัพธ์การจำแนกกลุ่มที่ได้ค่า F-measure ที่มากที่สุดคือ 88.80% จึงเลือกเทคนิคนี้ไปพัฒนาระบบ

ขั้นตอน 4 การวิเคราะห์ผลลัพธ์ที่ได้และความผิดพลาดที่เกิดขึ้น จากผลการทดสอบหาความสัมพันธ์ด้วยเทคนิค FP Growth โดยพิจารณาจากค่า lift พบกลุ่มโรคที่มีความสัมพันธ์กัน 2 กลุ่ม กลุ่มแรกคือกฎที่เกี่ยวข้องกับโรคความดันโลหิตสูง (p2) โรคหัวใจ (p5) และโรคไขมันในเลือดผิดปกติ (p6) กลุ่มที่สองคือกฎที่เกี่ยวข้องกับโรคเบาหวาน (p1) โรคความดันโลหิตสูง (p2) และโรคหัวใจ (p5) จากความสัมพันธ์นี้ทำให้ทราบว่าเมื่อเป็นโรคดังกล่าวแล้วมักจะมีโรคแทรกซ้อนตามมา เช่น ถ้าคนไข้เป็นโรคหัวใจและโรคความดันโลหิตสูงอยู่แล้ว ก็ต้องระวังจะเป็นโรคเบาหวานและโรคไขมันในเลือดผิดปกติซึ่งสามารถประยุกต์ใช้ความรู้นี้ในการประชาสัมพันธ์ผู้ป่วยเพื่อป้องกันการเกิดโรคแทรกซ้อน เช่น พยายามลดพฤติกรรมเสี่ยงที่จะทำให้เกิดโรคเบาหวานและไขมันในเลือดผิดปกติเสียแต่เนิ่นๆ เพื่อลดความเสี่ยงที่จะเป็นโรคดังกล่าว และจากผลการทดสอบการจำแนกกลุ่มข้อมูลแบบ Decision Rule: Partial Rules พบความผิดพลาดของ

ผลลัพธ์ที่ได้เช่น พบว่าข้อมูลที่อยู่ในกลุ่มเสี่ยงแต่โมเดลทำนายว่าเป็นกลุ่มป่วยมีจำนวน 1 ตัวอย่าง จากกฎ หากบุคคลนั้นมีประวัติการเจ็บป่วยกำลังทำการรักษาอยู่ ปฏิบัติตนตามที่แพทย์แนะนำ และ ไม่มีอาการตาพร่ามัวจะจำแนกว่าเป็นกลุ่มป่วย (196.0/1.0) ซึ่งกฎที่ได้เกิดจากมีข้อมูลที่เหมือนกัน จำนวนมากที่ผู้ป่วยตาไม่พร่ามัว ( $p15 = 2$ ) และอยู่ในกลุ่มป่วย เมื่อพิจารณาข้อมูลตัวอย่างรายนี้ พบว่ามีลักษณะค่าแอทริบิวต์บางตัวที่เหมือนกับคนกลุ่มใหญ่ที่ป่วยทำให้ถูกมองว่าป่วยไปด้วย ซึ่ง ข้อมูลนี้อาจถือได้ว่าเป็นข้อมูลที่ผิดปกติ (outlier) เนื่องจากมีคุณสมบัติแตกต่างไปจากกลุ่มคนทั่วไป นอกจากนี้ยังพบว่าโมเดลจำแนกกลุ่ม Class 1 (กลุ่มปกติ) มีความผิดพลาดสูงคือผิดพลาด 50 ตัวอย่างจาก 94 ตัวอย่าง คิดเป็น 53.2% เนื่องจากมีจำนวนตัวอย่างให้ระบบเรียนรู้มีอยู่เพียง 108 ตัวอย่างเท่านั้น ทำให้มีค่า Precision กับ Recall ที่ต่ำคือ 46.8% และ 40.7% ตามลำดับ วิธีแก้ไขคือควรใส่ข้อมูลที่เป็นกลุ่มปกติเข้าไปให้ระบบได้เรียนรู้เพิ่มอีก เพื่อที่จะได้รูปแบบการจำแนกกลุ่ม ที่ถูกต้องมากยิ่งขึ้น เมื่อให้ผู้เชี่ยวชาญช่วยตรวจสอบพบว่าแอทริบิวต์การปฏิบัติตนเมื่อเจ็บป่วยเป็น ข้อมูลที่ไม่เหมาะสมที่จะใช้จำแนกกฎเนื่องจากเป็นแอทริบิวต์ที่มองภาพรวมของประวัติการ เจ็บป่วยและอาการต่างๆทั้งหมด 16 แอทริบิวต์ แต่มีเพียง 6 แอทริบิวต์ที่ถ้าเป็นโรคแล้วผู้เชี่ยวชาญ จะถือว่าเป็นกลุ่มป่วยทันที ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง โรคตับ โรคอัมพาต โรคหัวใจ และ โรคไขมันในเลือดผิดปกติ ดังนั้นแอทริบิวต์การปฏิบัติตนเมื่อเจ็บป่วยจึงเป็นสาเหตุ ที่ทำให้เกิดการสร้างกฎที่ผิดพลาดจึงทำการทดลองตัดแอทริบิวต์การปฏิบัติตนเมื่อเจ็บป่วยออก พบว่าการจำแนกกลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วยให้ผลลัพธ์ที่ดีขึ้น โดยมีความถูกต้อง (Precision) เท่ากับ 89.20% ค่าระลึก (Recall) เท่ากับ 89.40% และ ค่าความเหวี่ยง (F-measure) เท่ากับ 89.30% ตามลำดับ

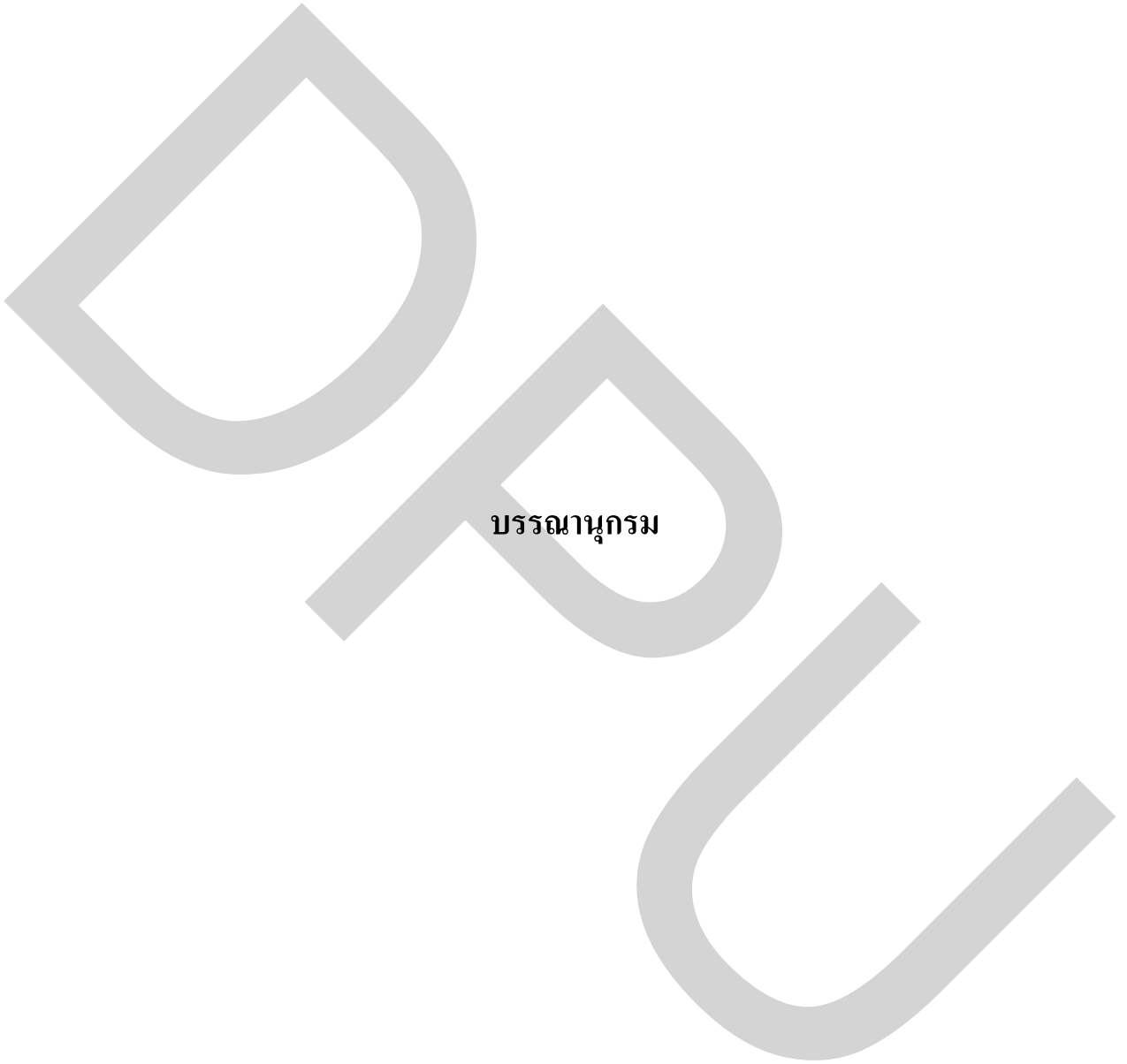
ขั้นตอน 5 การพัฒนาระบบคัดกรองสุขภาพเบื้องต้น เป็นการนำกฎที่ได้จากการจำแนก กลุ่มด้วยเทคนิค Decision Rule: Partial Rules ไปพัฒนาเป็นเงื่อนไขในการคัดกรองสุขภาพ ประชาชนเพื่อช่วยลดภาระงานของเจ้าหน้าที่พยาบาลผู้มีหน้าที่เก็บบันทึกและวิเคราะห์ข้อมูลของ ประชาชน และเปิดให้ประชาชนทั่วไปที่สนใจในเรื่องสุขภาพสามารถเข้ามาตรวจคัดกรองสุขภาพ เบื้องต้นของตัวเองได้

## 5.2 ข้อเสนอแนะ

1. ค่าความถูกต้องของการจำแนกข้อมูลบางกลุ่ม เช่น กลุ่มปกติ มีความถูกต้องน้อยเพราะว่ามี จำนวนตัวอย่างที่เป็นกลุ่มปกติอยู่น้อย ดังนั้นควรจัดเตรียมข้อมูลสำหรับให้ระบบได้เรียนรู้ที่มี จำนวนในแต่ละกลุ่มใกล้เคียงกัน

2. หากมีการปรับปรุงแก้ไขแบบฟอร์มการสอบถามพฤติกรรมสุขภาพของประชาชนในอนาคต ยังสามารถประยุกต์ใช้เทคนิควิธีการต่างๆของงานวิจัยนี้ได้ หรือสามารถใช้กับเอกสารแบบฟอร์มอื่นๆได้

3. ในอนาคตเมื่อระบบคัดกรองสุขภาพเบื้องต้นสามารถเก็บบันทึกข้อมูลของประชาชนทั้งประเทศได้เป็นจำนวนมากแล้ว และต้องการที่จะนำข้อมูลดังกล่าวมาหารูปแบบการจำแนกใหม่ด้วยโปรแกรม Weka อาจจำเป็นต้องแยกข้อมูลออกเป็นภูมิภาคต่างๆเช่น ภาคเหนือ ภาคกลาง ภาคอีสาน ภาคใต้ เป็นต้น เพราะลักษณะของคนในแต่ละภูมิภาคมีความแตกต่างกัน ถ้านำมาจำแนกกลุ่มร่วมกันอาจได้ผลลัพธ์ที่มีความถูกต้องน้อย หรือถ้าต้องการความถูกต้องมากยิ่งขึ้นไปอีกก็อาจจำแนกกลุ่ม โดยใช้ข้อมูลแยกเป็นแต่ละจังหวัด



**บรรณานุกรม**

## บรรณานุกรม

### ภาษาไทย

#### วิทยานิพนธ์

- มาติวรรณ บุญพลอย (2551) *เหมืองข้อมูลและการค้นหาคำที่มีความรู้สำหรับโรคกล้ามเนื้ออ่อนแรง*  
 ปริญญาวิทยาศาสตรมหาบัณฑิต สาขา วิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการ  
 คอมพิวเตอร์
- ศุภกรใจ วุฒิจิกโกศล (2551) *การใช้เทคนิคการทำเหมืองข้อมูลในผู้ป่วยข้อไหล่ติด* โรงพยาบาลพระ  
 นังเกล้า ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาการจัดการเทคโนโลยีสารสนเทศ
- รักถิ่น เหลลหา (2553) *การพยากรณ์ความเสี่ยงการเกิดโรคมะเร็งปอด โดยใช้ทฤษฎีของการทำ*  
*เหมืองข้อมูล* ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ  
 คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
- อังคณา พิจาร โชติ (2552) *ระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็น*  
*โรคเบาหวาน* ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะ  
 วิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

#### เอกสาร

- เอกสารการจัดทำแผนแม่บทในการพัฒนาระบบบริการสุขภาพ (Service Plan 2555- 2559)*  
 สำนักบริหารการสาธารณสุข สำนักงานปลัดกระทรวงสาธารณสุข (2554)
- แผนพัฒนาสุขภาพแห่งชาติ ในช่วงแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ ฉบับที่ 11*  
 (พ.ศ. 2555 - 2559) สำนักนโยบายและยุทธศาสตร์ กระทรวงสาธารณสุข (2554)

### ภาษาต่างประเทศ

#### BOOKS

- Agrawal et al. (1993) *Mining Association Rules Between Sets of Items in Lard Databases. In P.*  
*Buneman and S.Jajordai, eds. New York:ACM.*

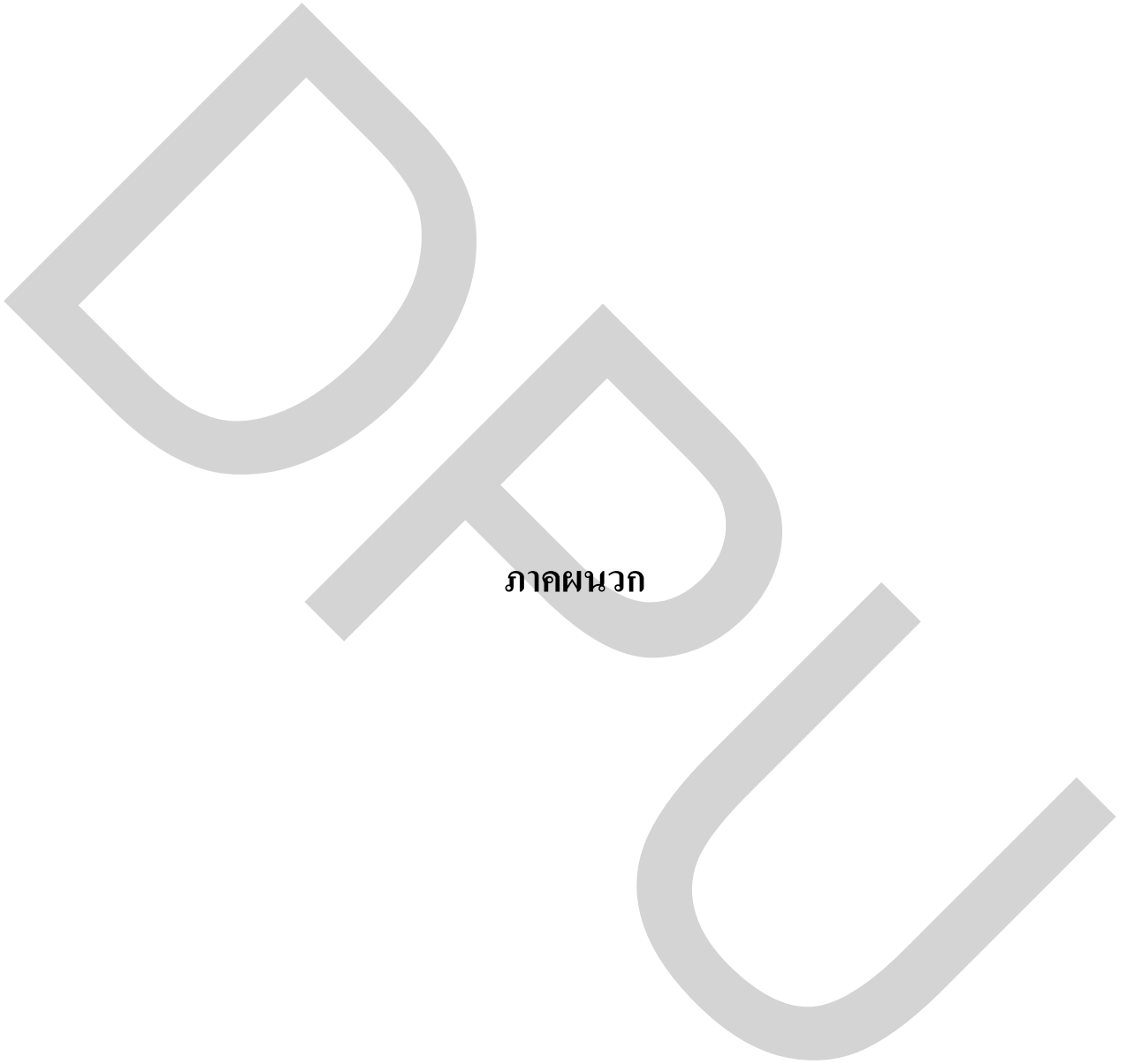
Kantardzic M. (2003) *Data mining Concepts, Models, Methods, and Algorithms*. United States of America: IEEE Press.

Han, J., et al. (2006) *Data mining concepts and techniques* (2nd ed.). United States of America: Morgan Kaufman Publishers.

### **ELECTRONIC SOURCE**

WEKA Machine Learning Group at University of Waikato, from

<http://www.cs.waikato.ac.nz/ml/weka/>



ภาคผนวก



ภาคผนวก ก

กฎการคัดกรองสุขภาพเบื้องต้นที่ได้จากการทำ Classification



กฎการคัดกรองสุขภาพเบื้องต้นที่ได้จากการทำ Classification

rule no	rule	class	description
1	p2 = 1	กลุ่มป่วย	มีประวัติเป็นโรคความดันโลหิตสูง
2	p1 = 3 AND p3 = 3 AND p6 = 3	กลุ่มเสี่ยง	ไม่เคยตรวจโรคเบาหวาน โรคตับ และโรคไขมันในเลือดผิดปกติ
3	fb5 = 0 AND p1 = 1	กลุ่มป่วย	มีประวัติเป็นโรคเบาหวาน
4	fb5 = 0 AND p6 = 2 AND p5 = 2 AND p3 = 3 AND b12 = 3	กลุ่มเสี่ยง	ไม่เคยตรวจโรคตับ
5	fb5 = 0 AND p6 = 2 AND p5 = 2 AND p3 = 2 AND sumbmi = 3	กลุ่มเสี่ยง	ค่าดัชนีมวลกายเท่ากับ 23-24.99 (ท้วม)
6	fb5 = 0 AND p6 = 2 AND p5 = 2 AND p3 = 2 AND sumhip = 2 AND fa6 = 0 AND sumbmi = 4	กลุ่มเสี่ยง	รอบเอวมีขนาดใหญ่กว่า 90 ซม.สำหรับผู้ชาย หรือรอบเอวใหญ่กว่า 80 ซม.สำหรับผู้หญิง และมีค่าดัชนีมวลกาย
7	fb5 = 0 AND p14 = 2 AND p6 = 2 AND p5 = 2 AND p3 = 2 AND sumbmi = 4	กลุ่มเสี่ยง	ค่าดัชนีมวลกายเท่ากับ 25-29.99 (อ้วน)
8	fb5 = 0 AND p14 = 2 AND p6 = 2 AND p5 = 2 AND p3 = 2 AND sumhip = 2 AND sumbmi = 2 AND b9 = 5	กลุ่มเสี่ยง	รอบเอวมีขนาดใหญ่กว่า 90 ซม.สำหรับผู้ชาย หรือรอบเอวใหญ่กว่า 80 ซม.สำหรับผู้หญิง
9	fb5 = 0 AND p6 = 3	กลุ่มเสี่ยง	ไม่เคยตรวจโรคไขมันในเลือดผิดปกติ
10	p6 = 1	กลุ่มป่วย	มีประวัติเป็นโรคไขมันในเลือดผิดปกติ
11	p5 = 3	กลุ่มเสี่ยง	ไม่เคยตรวจโรคหัวใจ
12	p5 = 1	กลุ่มป่วย	มีประวัติเป็นโรคหัวใจ
13	p3 = 2 AND p16 = 1 AND b7 = 3	กลุ่มเสี่ยง	เคยมีอาการชาปลายมือปลายเท้าโดยไม่ทราบสาเหตุ
14	p3 = 2 AND smoke = 1 AND b11 = 3 AND b12 = 3	กลุ่มเสี่ยง	สูบบุหรี่
15	p3 = 2 AND sumbmi = 5	กลุ่มเสี่ยง	ค่าดัชนีมวลกายมากกว่าหรือเท่ากับ 30 (อ้วนมาก)
16	p3 = 2 AND eat2 = 1 AND p10 = 2 AND eat4 = 0	กลุ่มเสี่ยง	ชอบอาหารรสเค็ม
17	p3 = 2 AND p11 = 2 AND exercise = 1 AND levelchest = 6 AND b8 = 5	กลุ่มเสี่ยง	ไม่ออกกำลังกาย และไม่เคยตรวจเต้านมด้วยตนเอง
18	p3 = 2 AND p11 = 2 AND testchest = 1 AND fb2 = 1	กลุ่มเสี่ยง	พี่น้องสายตรงของท่านมีประวัติเป็นโรคความดันโลหิตสูง
19	p3 = 2 AND fa6 = 0 AND p11 = 2 AND b13 = 3 AND cervic = 5 AND fb9 = 1 AND b14 = 5 AND sex = 6 AND gender = 2	กลุ่มเสี่ยง	ไม่เคยตรวจมะเร็งปากมดลูก
20	p3 = 2 AND fa6 = 0 AND p11 = 2 AND eat5 = 1 AND b10 = 3 AND b1 = 5	กลุ่มปกติ	สุขภาพปกติ
21	p3 = 2 AND fa6 = 0 AND p11 = 2 AND b10 = 3 AND eat5 = 0 AND eat1 = 1	กลุ่มเสี่ยง	ชอบอาหารรสหวาน
22	p3 = 1	กลุ่มป่วย	มีประวัติเป็นโรคตับ
23	fa6 = 0 AND fb2 = 0 AND p2 = 2 AND p11 = 2 AND smoke = 1 AND b8 = 3	กลุ่มเสี่ยง	สูบบุหรี่
24	fa6 = 0 AND fb2 = 0 AND p2 = 2 AND p11 = 2 AND fb3 = 0 AND fa3 = 0 AND exercise = 4 AND b20 = 5	กลุ่มปกติ	สุขภาพปกติ
25	fa6 = 0 AND fb2 = 0 AND p1 = 2 AND p11 = 2 & p10 = 1 AND status = 2	กลุ่มปกติ	สุขภาพปกติ
26	fa6 = 0 AND fb2 = 0 AND p1 = 3	กลุ่มเสี่ยง	ไม่เคยตรวจโรคเบาหวาน

27	fb2 = 0 AND p11 = 1	กลุ่มเสี่ยง	เคยมีอาการกินจุแต่ผอมลง
28	fb2 = 0 AND smoke = 2 AND fa8 = 1 AND b20 = 5	กลุ่มเสี่ยง	ไม่ทราบประวัติการเจ็บป่วยของบิดาหรือมารดา
29	smoke = 2 AND fb2 = 0 AND sumBMI = 1 AND eat2 = 0 & testchest = 3	กลุ่มปกติ	สุขภาพปกติ
30	fb2 = 0 AND smoke = 2 AND b6 = 5 AND b1 = 5 & sex = 5 & education = 3	กลุ่มปกติ	สุขภาพปกติ
31	fb2 = 0 AND smoke = 2 AND drink = 1	กลุ่มเสี่ยง	ดื่มเครื่องดื่มที่มีแอลกอฮอล์
32	fb2 = 0 AND smoke = 2 & sumhip = 1 AND b14 = 3 AND b18 = 3	กลุ่มปกติ	สุขภาพปกติ
33	smoke = 1	กลุ่มเสี่ยง	สูบบุหรี่
34	fa3 = 0 AND sumhip = 2	กลุ่มเสี่ยง	รอบเอวมีขนาดใหญ่กว่า 90 ซม. สำหรับผู้ชาย หรือรอบเอวใหญ่กว่า 80 ซม. สำหรับผู้หญิง
35	fa3 = 0 AND exercise = 3 AND b3 = 3	กลุ่มปกติ	สุขภาพปกติ
36	eat3 = 0 AND fa3 = 0 AND b15 = 5 AND b19 = 0	กลุ่มปกติ	สุขภาพปกติ
37	eat3 = 1	กลุ่มเสี่ยง	ชอบอาหารรสมัน
38	sumbmi = 1	กลุ่มเสี่ยง	ค่าดัชนีมวลกายน้อยกว่า 18.5 (ผอม)
39	resultb = 3 AND cervic = 1	กลุ่มปกติ	สุขภาพปกติ
40	resultb = 3 AND eat6 = 1	กลุ่มปกติ	สุขภาพปกติ
41	b12 = 5 AND resultb = 3 AND levelchest = 1	กลุ่มปกติ	สุขภาพปกติ
42	resultb = 3	กลุ่มปกติ	สุขภาพปกติ



ภาคผนวก ข

แบบสำรวจสถานะสุขภาพประชาชน



## แบบสำรวจสภาวะสุขภาพประชาชน

วันที่สำรวจ..... หน้าที่ 1

## แบบคัดกรองสุขภาพเบื้องต้น เขตเทศบาลเมืองบุรีรัมย์

 ประกันสังคม  ข้าราชการ  สิทธิหลักประกันสุขภาพถ้วนหน้าชื่อ-สกุล..... เพศ  หญิง  ชาย วัน เดือน ปี เกิด.....เลขที่บัตรประจำตัวประชาชน..... สถานภาพสมรส  โสด  สมรส  หม้าย/หย่า/แยก

ที่อยู่..... ถนน..... ชุมชน..... ต.ในเมือง อ.เมือง จ.บุรีรัมย์ เบอร์โทรศัพท์.....

การศึกษา  ไม่ได้เรียน  ประถม  มัธยม  ปวช./ปวศ.  ป.ตรี  สูงกว่าป.ตรี น้ำหนัก..... กก. ส่วนสูง..... ซม.

ดัชนีมวลกาย..... รอบเอว..... ซม. ความดันโลหิต ครั้งที่ 1 ...../..... ความดันโลหิต ครั้งที่ 2 ...../.....

## 1. ข้อมูลครอบครัว

## 1.1 บิดาหรือมารดาของท่านมีประวัติการเจ็บป่วยด้วย

 เบาหวาน  ความดันโลหิตสูง  โรคเก๊าท์  ไตวายเรื้อรัง  กล้ามเนื้อหัวใจตาย  
 เส้นเลือดสมอง  ออกรับโป่งพอง  ไม่ทราบ  ไม่มี  อื่นๆระบุ .....

## 1.2 พี่น้อง (สายตรง) ของท่านมีประวัติการเจ็บป่วยด้วยโรค

 เบาหวาน  ความดันโลหิตสูง  โรคเก๊าท์  ไตวายเรื้อรัง  กล้ามเนื้อหัวใจตาย  
 เส้นเลือดสมอง  ออกรับโป่งพอง  ไม่ทราบ  ไม่มี  อื่นๆระบุ.....

## 2. ท่านมีประวัติการเจ็บป่วย หรือต้องพบแพทย์ ด้วยโรคหรืออาการดังต่อไปนี้

โรค	มี	ไม่มี	ไม่เคย มี	ไม่คอย ตรวจ	สถานที่รับยา
2.1 โรคเบาหวาน					<input type="checkbox"/> รพ.บร. <input type="checkbox"/> ศูนย์แพทย์ฯ <input type="checkbox"/> ศูนย์เทศบาล <input type="checkbox"/> คลินิก <input type="checkbox"/> อื่นๆ.....
2.2 โรคความดันโลหิตสูง					<input type="checkbox"/> รพ.บร. <input type="checkbox"/> ศูนย์แพทย์ฯ <input type="checkbox"/> ศูนย์เทศบาล <input type="checkbox"/> คลินิก <input type="checkbox"/> อื่นๆ.....
2.3 โรคตับ					<input type="checkbox"/> รพ.บร. <input type="checkbox"/> ศูนย์แพทย์ฯ <input type="checkbox"/> ศูนย์เทศบาล <input type="checkbox"/> คลินิก <input type="checkbox"/> อื่นๆ.....
2.4 โรคอัมพาต					<input type="checkbox"/> รพ.บร. <input type="checkbox"/> ศูนย์แพทย์ฯ <input type="checkbox"/> ศูนย์เทศบาล <input type="checkbox"/> คลินิก <input type="checkbox"/> อื่นๆ.....
2.5 โรคหัวใจ					<input type="checkbox"/> รพ.บร. <input type="checkbox"/> ศูนย์แพทย์ฯ <input type="checkbox"/> ศูนย์เทศบาล <input type="checkbox"/> คลินิก <input type="checkbox"/> อื่นๆ.....
2.6 ไขมันในเลือดผิดปกติ					<input type="checkbox"/> รพ.บร. <input type="checkbox"/> ศูนย์แพทย์ฯ <input type="checkbox"/> ศูนย์เทศบาล <input type="checkbox"/> คลินิก <input type="checkbox"/> อื่นๆ.....
2.7 แผลที่เท้า/ค้ำขา (จากเบาหวาน)					
2.8 คลอคุนดรีน้ำหนักเกิน 4 กิโลกรัม					
2.9 คิมน้ำบ่อและมาก					
2.10 ปีศาจกลางคืน 3 ครั้งขึ้นไป					
2.11 กินจุแต่ผอมลง					
2.12 น้ำหนักลด / อ่อนเพลีย					
2.13 เป็นแผลที่ริมฝีปากบ่อและหายยาก					
2.14 ค้นตามผิวหนังและอวัยวะสืบพันธุ์					
2.15 ตาพร่ามัว ต้องเปลี่ยนแว่นบ่อย					
2.16 ขาปลายมือปลายเท้าโดยไม่ทราบสาเหตุ					

3. กรณีที่ท่านมีประวัติเจ็บป่วย ตามข้อ 2 ท่านปฏิบัติตนอย่างไร

- รับการรักษาอยู่/ปฏิบัติตามที่แพทย์แนะนำ  รับการรักษาแต่ไม่สม่ำเสมอ  เคยรักษา ขณะนี้ไม่รักษา/หายาตนเอง

4. ท่านสูบบุหรี่หรือไม่

- สูบ.....มวน/วัน ชนิดของบุหรี่  ยาเส้น  ก้นกรอง ระยะเวลา.....ปี (ตั้งแต่เริ่มสูบบุหรี่จนถึงปัจจุบัน)

ไม่สูบ

- เคยสูบแต่เลิกแล้วชนิดของบุหรี่ที่เคยสูบ  ยาเส้น  ก้นกรอง ระยะเวลา.....ปี (ตั้งแต่เริ่มสูบจนถึงเลิก)จำนวน....ซอง/ปี

5. ท่านดื่มเครื่องดื่มที่มีแอลกอฮอล์หรือไม่

- ดื่ม.....ครั้ง/สัปดาห์ (ดื่มเหล้า > 45 cc. ต่อวัน, ดื่มเบียร์ > 240 cc. ต่อวัน, ดื่มไวน์ > 120 cc. ต่อวัน)

- ไม่ดื่ม  เคยดื่มแต่เลิกแล้ว

6. ท่านออกกำลังกาย/เล่นกีฬา

- ไม่ออกกำลังกายเลย  ออกกำลังกายน้อยกว่าสัปดาห์ละ 3 ครั้ง

- ออกกำลังกายสัปดาห์ละ 3 ครั้งๆละ 30 นาที สม่ำเสมอ  ออกกำลังกายสัปดาห์ละมากกว่า 3 ครั้งๆละ 30 นาที สม่ำเสมอ

- ออกกำลังกายทุกวัน ครั้งละ 30 นาที

7. ท่านชอบอาหารรสใด (ตอบได้มากกว่า 1 ข้อ)

- หวาน  เค็ม  มัน เปรี้ยว  จืด  ไม่ชอบทุกข้อ

8. ท่านขี้ขี้หรือโดยสารรดจักรยานยนต์/รถยนต์

- ไม่ขี้ขี้/ไม่โดยสาร  ขี้ขี้/โดยสาร และใส่หมวกกันน็อก/คาดเข็มขัดนิรภัยทุกครั้ง

- ขี้ขี้/โดยสาร และใส่หมวกกันน็อก/คาดเข็มขัดนิรภัยบางครั้ง

- ขี้ขี้/โดยสาร และใส่หมวกกันน็อก/คาดเข็มขัดนิรภัยนานๆ ครั้ง (ใส่เฉพาะเมื่อมีด่านตรวจ)

9. เมื่อมีเพศสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยาของท่านหรือคู่ของท่าน ใช้ถุงยางอนามัยหรือไม่

- ใช้ทุกครั้ง  ใช้เมื่อถูกร้องขอ

- ไม่ใช่  ไม่เคยมีเพศสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยาของตนเอง

- ไม่ตอบ  ไม่เคยมีเพศสัมพันธ์

10. ท่านสามารถตรวจเต้านมด้วยตนเอง  ได้  ไม่ได้

11. ความถี่ในการตรวจเต้านมด้วยตนเอง

- เดือนละครั้ง  เดือนเว้นเดือน  3-4 เดือนครั้ง  6 เดือนครั้ง  ปีละครั้ง  ไม่เคยตรวจ

12. ผลตรวจเต้านมด้วยตนเอง

- ปกติ  ผิดปกติ รักษาที่.....วันเริ่มรักษา.....

13. ความถี่ในการตรวจมะเร็งปากมดลูก  ทุกปี  ทุก 2 ปี  ทุก 3 - 5 ปี  มากกว่า 5 ปี  ไม่เคยตรวจ

13. ผลตรวจมะเร็งปากมดลูก

- ปกติ ตรวจครั้งสุดท้ายเมื่อ.....

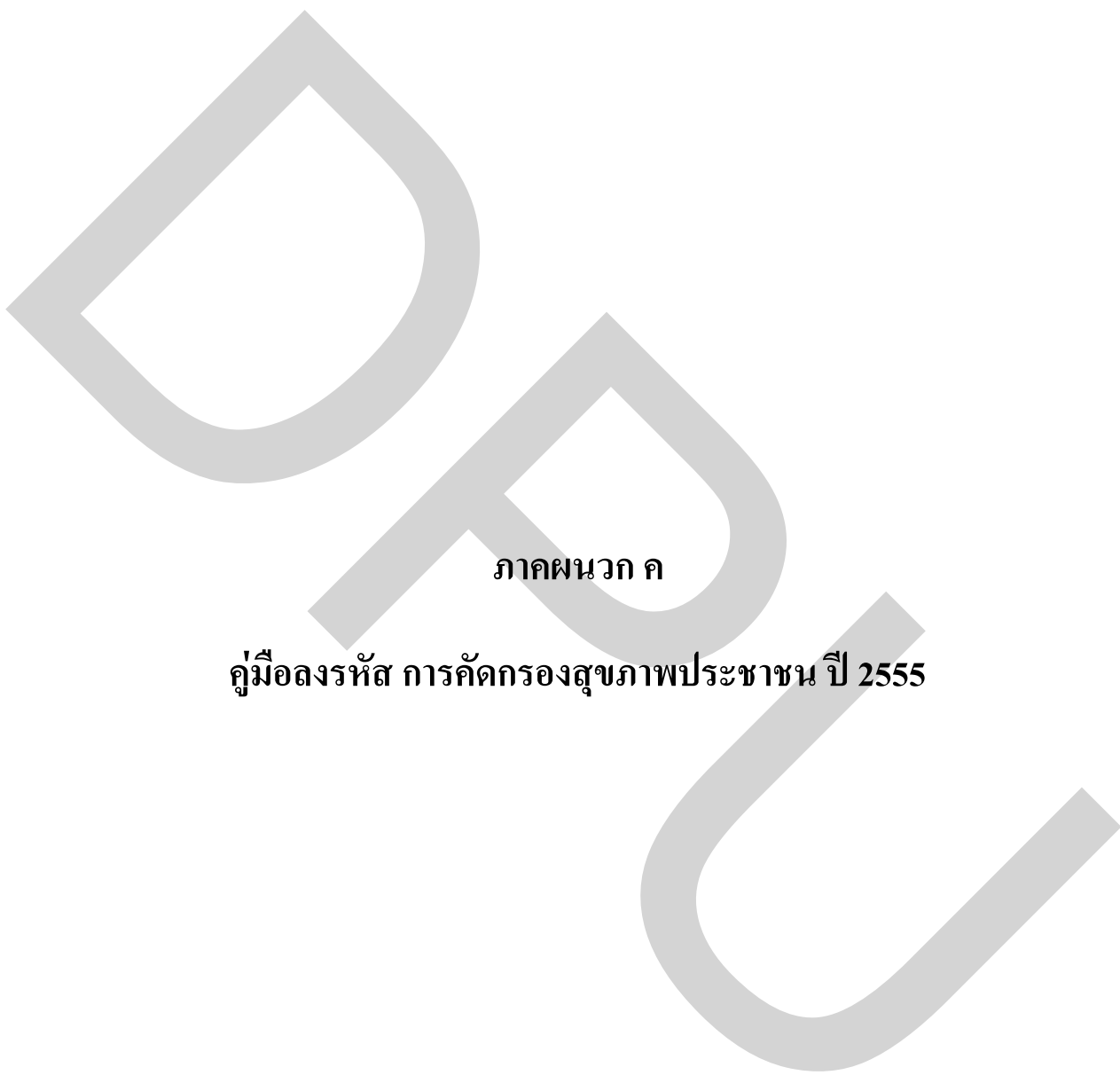
- ผิดปกติ รักษาที่.....วันเริ่มรักษา.....

**แบบประเมินพฤติกรรมสุขภาพด้านอาหาร อารมณ์ ออกกำลังกาย**

ขอให้ท่านขีดเครื่องหมาย ✓ ลงในช่องที่ท่านได้ปฏิบัติเป็นส่วนใหญ่ ตามความจริง

พฤติกรรมที่ปฏิบัติ	ความถี่ในการปฏิบัติ		
	ประจำ(5) (5-7 วัน/สัปดาห์)	ครั้งคราว(3) (1-4วัน/สัปดาห์)	ไม่เคยเลย (0)
1.กินอาหารครบ 5 หมู่ (ข้าว ผัก ผลไม้ เนื้อสัตว์ นม) 51			
2.กินอาหารมื้อเช้า เป็นมื้อหลัก 52			
3.กินผักมากกว่าวันละ 3 ท็อปปี 53			
4.กินผลไม้วันละ 2-3 ส่วน(1 ส่วนเท่ากับ 6-8 คำ) 54			
5.กินปลาอย่างน้อยวันละ 1 มื้อ 55			
6.กินเนื้อสัตว์ไม่ติดมัน สัปดาห์ละ 2-3 มื้อ 56			
7.ดื่มนมรสจืดหรือนมถั่วเหลืองผสมงาคาร์โบไฮเดรตวันละ 1-2 แก้ว 57			
8.กินอาหารมื้อเย็นห่างจากเวลานอนไม่น้อยกว่า 4 ชั่วโมง 58			
9.กินอาหารประเภทต้ม นึ่ง ลวก อบ ยำหรือหมก 59			
10.หลีกเลี่ยงอาหารไขมันสูง 60			
11.หลีกเลี่ยงของหวานและขนมที่มีแป้งและน้ำตาลมาก 61			
12.กินอาหารรสจืด 62			
13.เลือกดื่มน้ำเปล่าแทนน้ำอัดลมหรือน้ำหวาน 63			
14.หลีกเลี่ยงเครื่องดื่มที่มีแอลกอฮอล์ 64			
15.อารมณ์ดี ไม่เครียด 65			
16.นอนหลับไม่น้อยกว่า วันละ 7-8 ชั่วโมง 66			
17.ออกกำลังกายสัปดาห์ละ 5 วันหรือ สัปดาห์ละ 5 ครั้ง 67			
18.ออกกำลังกายวันละ 30 นาที 68			
19.ขณะออกกำลังกายหายใจเรื้อรังกว่าปกติและเหงื่อซึม 69			
20.ทุกครั้งวัดรอบเอวได้เกณฑ์ปกติ คือ คือเพศหญิงไม่เกิน 80 ซม. และเพศชายไม่เกิน 90 ซม. 70			
รวมคะแนนเท่ากับ..... คะแนน 71 = 24 จาก 100 คะแนน			

คะแนนรวมเท่ากับ	100	คะแนน พฤติกรรมด้านสุขภาพของท่านดีมาก
คะแนนรวมเท่ากับ	80-99	คะแนน พฤติกรรมด้านสุขภาพดี
คะแนนรวมเท่ากับ	60-79	คะแนน พฤติกรรมด้านสุขภาพดีปานกลาง
คะแนนรวมน้อยกว่า	60	คะแนน ควรปรับเปลี่ยนพฤติกรรมให้ดีขึ้น



**ภาคผนวก ค**

**คู่มือลงรหัส การคัดกรองสุขภาพประชาชน ปี 2555**

## คู่มือลรห้ส การค้ดกรองสุขภาพประชชน ปี 2555

## คู่มือลรห้ส การค้ดกรองสุขภาพประชชน ปี 2555

ตัวแปร	รหัส	ความหมาย
เพศ	1-2	1.เพศชาย 2.เพศหญิง
สถานภาพ	1-3	1.โสด 2.สมรส 3.ม่าย/หย่า/แยก
การศึกษา	1-6	1.ไม่ได้เรียน 2.ประถมศึกษา 3.มัธยมศึกษา 4.อนุปริญญา 5.ปริญญาตรี 6.สูงกว่าปริญญาตรี
ดัชนีมวลกาย	1-5	1.ผอม(น้อยกว่า 18.5) 2.ปกติ(18.5-22.99) 3.ท้วม(23-24.99) 4.อ้วน(25-29.99) 5.อ้วนมาก(มากกว่าหรือเท่ากับ 30)
รอบเอว	1-2	1.ชายน้อยกว่า 90 ซม. หญิงน้อยกว่า 80 ซม. 2.ชาย $\geq$ 90 ซม. หญิง $\geq$ 80 ซม.
ความดันโลหิต	1-4	1.น้อยกว่า 120/80 2.120-139/80-89 3.140/90 4.มากกว่าหรือเท่ากับ 180/110
บิดามารดามีประวัติการเจ็บป่วย	1-10	1.เบาหวาน 2.ความดันโลหิตสูง 3.โรคเก๊าท์ 4.ไตวายเรื้อรัง 5.กล้ามเนื้อหัวใจตาย 6.เส้นเลือดสมอง 7.ถุงลมโป่งพอง 8.ไม่ทราบ 9.ไม่มี 10.อื่น ๆ ระบุ



ตัวแปร	รหัส	ความหมาย
พี่น้อง(สายตรง)มีประวัติการเจ็บป่วย	1-10	1.เบาหวาน 2.ความดันโลหิตสูง 3.โรคเก๊าท์ 4.ไตวายเรื้อรัง 5.กล้ามเนื้อหัวใจตาย 6.เส้นเลือดสมอง 7.ถุงลมโป่งพอง 8.ไม่ทราบ 9.ไม่มี 10.อื่น ๆ ระบุ
P1(โรคเบาหวาน)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P2(โรคความดันโลหิตสูง)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P3(โรคตับ)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P4(โรคอัมพาต)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P5(โรคหัวใจ)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P6(ไขมันในเลือดสูง)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P7(แผลที่เท้า/ตัดขา จากเบาหวาน)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P8(คลอسترโรลน้ำหนักรวม 4 กิโลกรัม)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P9(ดื่มน้ำบ่อยและมาก)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ

ตัวแปร	รหัส	ความหมาย
P10(ปีสภาวะกลางคืน 3 ครั้งขึ้นไป)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P11(กินजूแต่ผสมลง)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P12(น้ำหนักลด/อ่อนเพลีย)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P13(เป็นแผลที่ริมฝีปากบ่อยและหายยาก)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P14(คันตามผิวหนังและอวัยวะสืบพันธุ์)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P15(ตาพร่ามัว ต้องเปลี่ยนแว่นบ่อย)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
P16(ขาปลายมือปลายเท้าโดยไม่ทราบสาเหตุ)	1-3	1.มี 2.ไม่มี 3.ไม่เคยตรวจ
Treat(กรณีที่มีประวัติการเจ็บป่วย ท่านปฏิบัติตัวอย่างไร)	1-4	1.รับการรักษาอยู่/ปฏิบัติตามที่แพทย์แนะนำ 2.รับการรักษาแต่ไม่สม่ำเสมอ 3.เคยรักษา ขณะนี้ไม่รักษา/หายาทานเอง 4.ไม่มี
Smoke(ท่านสูบบุหรี่หรือไม่)	1-3	1.สูบบุหรี่ 2.ไม่สูบบุหรี่ 3.เคยสูบบุหรี่แต่เลิกแล้ว
Drink(ท่านดื่มเครื่องดื่มแอลกอฮอล์หรือไม่)	1-3	1.ดื่ม 2.ไม่ดื่ม 3.เคยดื่มแต่เลิกแล้ว
Excercise (ท่านออกกำลังกาย/เล่นกีฬาหรือไม่)	1-3	1.ไม่ออกกำลังกาย 2.ออกกำลังกายน้อยกว่าสัปดาห์ละ 3 ครั้ง 3.ออกกำลังกายสัปดาห์ละ 3 ครั้งๆละ 30 นาที สม่ำเสมอ 4.ออกกำลังกายมากกว่า 3 ครั้งๆละ 30 นาที สม่ำเสมอ 5.ออกกำลังกายทุกวัน ครั้งละ 30 นาที

ตัวแปร	รหัส	ความหมาย
Eat(ท่านชอบอาหารรสใด)	1-6	1.หวาน 2.เค็ม 3.มัน 4.เปรี้ยว 5.จืด 6.ไม่ชอบทุกข้อ
Eat(ท่านขบขี้หรือโดยสารรถจักรยานยนต์/รถยนต์) <i>Alive</i>	1-4	1.ไม่ขบขี้/ไม่โดยสาร 2.ขบขี้/โดยสารและใส่หมวกกันน็อค/คาดเข็มขัดนิรภัย ทุกครั้ง 3.ขบขี้/โดยสารและใส่หมวกกันน็อค/คาดเข็มขัดนิรภัย บางครั้ง 4.ขบขี้/โดยสารและใส่หมวกกันน็อค/คาดเข็มขัดนิรภัย นาน ๆ ครั้ง(ใส่เฉพาะเมื่อมีด่านตรวจ)
sex(เมื่อมีเพศสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยา ของท่านหรือคู่ของท่าน ใช้ถุงยางอนามัยหรือไม่)	1-6	1.ใช้ทุกครั้ง 2.ใช้เมื่อถูกร้องขอ 3.ไม่ใช้ 4.ไม่เคยมีสัมพันธ์กับผู้ที่ไม่ใช่สามีหรือภรรยาของ ตนเอง 5.ไม่ตอบ 6.ไม่เคยมีเพศสัมพันธ์
Chest(ท่านสามารถตรวจเต้านมด้วยตนเอง)	1-2	1.ได้ <i>3.8 = ไม่แน่</i> 2.ไม่ได้
Level chest(ความถี่ในการตรวจเต้านมด้วยตนเอง)	1-6	1.เดือนละครั้ง <i>7 = จำไม่ได้</i> 2.เดือนเว้นเดือน 3.3-4 เดือนครั้ง 4.6 เดือนครั้ง 5.ปีละครั้ง 6.ไม่เคยตรวจ
Test Chest(ผลตรวจเต้านมด้วยตนเอง)	1-2	1.ปกติ <i>3. จำไม่ได้</i> 2.ผิดปกติ
Cervic(ความถี่ในการตรวจมะเร็งปากมดลูก)	1-5	1.ทุกปี 2.ทุก 2 ปี 3.ทุก 3-5 ปี 4.มากกว่า 5 ปี 5.ไม่เคยตรวจ
Test Cervic(ผลตรวจมะเร็งปากมดลูก)	1-2	1.ปกติ 2.ผิดปกติ

ตัวแปร	รหัส	ความหมาย
B1(กินอาหารครบ 5 หมู่)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B2(กินอาหารเข้าเป็นมื้อหลัก)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B3(กินผักมากกว่าวันละ 3 ทัพพี)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B4(กินผลไม้วันละ 2-3 ส่วน)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B5(กินปลาอย่างน้อยวันละ 1 มื้อ)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B6(กินเนื้อสัตว์ไม่ติดมันสัปดาห์ละ 2-3 มื้อ)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B7(ดื่มนมรสจืดหรือนมถั่วเหลืองผสมงาดำรสจืดวันละ 1-2 แก้ว)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B8(กินอาหารมือเย็นห่างจากเวลานอนไม่น้อยกว่า 4 ชั่วโมง)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B9(กินอาหารประเภทต้ม นึ่ง ลวก อบ ยำหรือหมก)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B10(หลีกเลี่ยงอาหารไขมันสูง)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B11(หลีกเลี่ยงของหวานและขนมที่มีแป้งและน้ำตาลมาก)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B12(กินอาหารรสจืด)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย

ตัวแปร	รหัส	ความหมาย
B13(เลือกดื่มน้ำเปล่าแทนน้ำอัดลมหรือน้ำหวาน)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B14(หลีกเลี่ยงเครื่องดื่มที่มีแอลกอฮอล์)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B15(อารมณ์ดี ไม่เครียด)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B16(นอนหลับไม่น้อยกว่า วันละ 7-8 ชั่วโมง)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B17(ออกกำลังกายสัปดาห์ละ 5 วันหรือสัปดาห์ละ 5 ครั้ง)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B18(ออกกำลังกายวันละ 30 นาที)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B19(ขณะออกกำลังกายหายใจเร็วขึ้นกว่าปกติและเหงื่อซึม)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
B20(ทุกครั้งวัดรอบเอวได้เกณฑ์ปกติ คือ เพศหญิงไม่เกิน 80 ซม.และเพศชายไม่เกิน 90 ซม.)	1-3	1.ประจำ 2.ครั้งคราว 3.ไม่เคยเลย
Sum B(สรุปผลคะแนนการประเมินพฤติกรรมสุขภาพ)	1-4	1.น้อยกว่า 60 คะแนน 2.60-79คะแนน 3.80-99คะแนน 4.100 คะแนน
Result B (สรุปพฤติกรรมสุขภาพ)	1-4	1.ควรปรับเปลี่ยนพฤติกรรมให้ดีขึ้น 2.พฤติกรรมด้านสุขภาพดีปานกลาง 3.พฤติกรรมด้านสุขภาพดี 4.พฤติกรรมด้านสุขภาพของท่านดีมาก
Summary(สรุปผลการคัดกรอง)	1-3	1.ปกติ 2.กลุ่มเสี่ยง 3.กลุ่มป่วย

ภาคผนวก ง

เอกสารงานประชุมวิชาการ การบริหารและการจัดการ ครั้งที่ 8

## เอกสารงานประชุมวิชาการ การบริหารและการจัดการ ครั้งที่ 8



**DPU** 2511

มหาวิทยาลัยธุรกิจบัณฑิต Dhurakij Pundit University

มุ่งนำความรู้สู่การปฏิบัติ ๑๑๐/๑-๔ ถนนประชาชื่น หลักสี่ กรุงเทพฯ ๑๐๒๑๐ โทรศัพท์ (๖๖๒) ๙๕๔-๗๓๐๐ โทรสาร (๖๖๒) ๕๘๙-๙๖๐๕-๖  
110/1-4 Prachachuen Rd., Laksi, Bangkok 10210 Tel. (662) 954-7300 Fax. (662) 589-9605-6 www.dpu.ac.th

ที่ มธบ 0501(1) / 09357

24 กันยายน 2555

เรื่อง แจ้งผลการพิจารณาผลงานวิชาการและยื่นรับการให้นำเสนอผลงาน การประชุมวิชาการ  
การบริหารและการจัดการ ครั้งที่ 8

เรียน คุณกิตติศักดิ์ สุวมาลัย

ด้วยมหาวิทยาลัยธุรกิจบัณฑิต ร่วมกับ เครือข่ายวิจัย “ประชาชื่น” จัดการประชุมวิชาการ  
การบริหารและการจัดการ ครั้งที่ 8 “การบริหารจัดการแห่งโลกพลวัต” ในวันศุกร์ที่ 12 ตุลาคม 2555  
ณ มหาวิทยาลัยธุรกิจบัณฑิต โดยมีวัตถุประสงค์เพื่อการกระตุ้นให้เกิดการศึกษาค้นคว้า พัฒนาองค์ความรู้ใหม่ๆ  
ทางการบริหารและการจัดการในโลกแห่งการเปลี่ยนแปลง และเป็นการเปิดโอกาสให้นักวิชาการ นักวิจัย นักศึกษา  
และบุคคลที่สนใจนำเสนอผลงานทางวิชาการ และร่วมประชุมเพื่อให้ความรู้ความก้าวหน้าทางด้านการบริหารและการ  
จัดการ

ในกรณีนี้ ขอขอบคุณที่ท่านให้ความสนใจ และได้ส่งผลงานวิชาการเข้าร่วมนำเสนอในการประชุมวิชาการฯ  
ครั้งนี้ ซึ่งผลงานวิชาการของท่านได้รับการประเมินจากผู้ทรงคุณวุฒิที่มีความรู้ความเชี่ยวชาญในศาสตร์ที่เกี่ยวข้อง  
จากการพิจารณาผลงานวิชาการของท่านในเรื่อง “การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล” นั้น เห็น  
ควรให้นำเสนอผลงานวิชาการในการประชุมวิชาการ การบริหารจัดการ ครั้งที่ 8 ซึ่งมีเวลานำเสนอ 15 นาที  
และถาม-ตอบ 5 นาที

จึงเรียนมาเพื่อเข้าร่วมนำเสนอผลงานวิชาการในวันเวลาดังกล่าวข้างต้น ขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(ศาสตราจารย์ ดร.ไพฑูรย์ สินารัตน์)

รองอธิการบดีฝ่ายวิจัย ปฏิบัติการแทน

อธิการบดี

ศูนย์วิจัย มหาวิทยาลัยธุรกิจบัณฑิต

โทร. 02-9547300 ต่อ 528, 128

โทรสาร 02-5800064



**ตารางนำเสนอผลงานวิชาการ**  
**ประชุมวิชาการ การบริหารและการจัดการ ครั้งที่ 8 “การบริหารจัดการแห่งโลกพลวัต”**  
**๘ ธันวาคม ๒๕๕๕**

ห้องกลุ่ม	13.00-13.20 น.	13.20-13.40 น.	13.40-14.00 น.	14.00-14.20 น.	14.20-14.40 น.
ห้อง 7305	การศึกษาพฤติกรรมและสาเหตุของการคิดนอกกรอบออนไลน์ กรณีศึกษานักเรียนชั้นมัธยมศึกษาตอนต้น โรงเรียนปทุมวิไล	การพัฒนากระบวนการประเมินผลการจัดการเรียนการสอนผ่านระบบเครือข่าย อินเทอร์เน็ต	การศึกษารองสาขาเพื่อป้องกันโดยใช้เทคนิคหนึ่งข้อมูล	การพัฒนากระบวนการแนะแนววิถีไอ การแข่งขันเรือกาประเพณี โดยการใช้โปรแกรมสำเร็จของผู้ใช้ เปรียบเทียบกับค่าเฉลี่ยของวิถีไอ	ระบบจัดการองค์ความรู้เกี่ยวกับแมงป่องด้วยเว็บ
เทคโนโลยีสารสนเทศ เทคโนโลยีคอมพิวเตอร์	โดย อ.ธันวาคม ฐิติเดชอย่าง สถาบันการจัดการปัญญาภิวัฒน์	โดย อ.ศักดิ์ชาย เหลืองมณีโรจน์ ม.เทคโนโลยีราชมงคลธัญชัง	โดย คุณกิตติศักดิ์ สุนามาศย์ ม.ธุรกิจบัณฑิตย์	โดย คุณปริญญา ร่มแสง ม.ธุรกิจบัณฑิตย์	โดย คุณสุนศักดิ์ จิระวัชร ม.นครสวรรค์
การศึกษา	14.40-15.00 น.	15.00-15.20 น.	15.20-15.40 น.	15.40-16.00 น.	16.00-16.20 น.
อ.วีรพล สวรรค์พิทักษ์ (ผู้ดำเนินรายการ) & สุพรรณิ นวลนาท (ผู้ช่วย)	ระบบการจัดการองค์ความรู้ พรชัย ไม้วงศ์ในประเทศไทย ผ่านเว็บ	รูปแบบการจัดการเรียนการสอน ภาษาอังกฤษของนักเรียนผ่านเทคโนโลยี สารสนเทศบนอุปกรณ์โทรศัพท์เคลื่อนที่ เพื่อรองรับประชาคมอาเซียน ตามแนว ทฤษฎีคอนสตรัคติวิสต์ฯ	รูปแบบการเรียนการสอนผ่านเว็บร่วมกับ สื่อสังคมเครือข่าย เรื่องการสร้างเว็บไซต์ โปรแกรมบุคลิกภาพของนักศึกษา ประกาศนียบัตรวิชาชีพวิทยาลัย เทคโนโลยีสยาม บริหารธุรกิจ นนทบุรี	ยุทธศาสตร์การบริหารกิจการ นักเรียนเพื่อประสิทธิภาพ พัฒนาผู้เรียนในยุคสารสนเทศ	รูปแบบการบริหารงานวิชาการ ของโรงเรียนมัธยมศึกษาของ ประเทศไทยเพื่อสร้างความ ผู้นำด้านการศึกษาในกลุ่ม ประเทศอาเซียน
	โดย คุณวิศุทธิ์ พุฒ่า ม.นครสวรรค์	โดย คุณกมลวรรณ กุสันทะยะ ม.เทคโนโลยีพระจอมเกล้าพระนครเหนือ	โดย คุณนพรัตน์ วนิชคม ม.เทคโนโลยีพระจอมเกล้าพระนครเหนือ	โดย คุณเสาวภาคย์ เหลืองมณีโรจน์ ม.ธุรกิจบัณฑิตย์	โดย คุณมานิช หัวตระกูล ม.ธุรกิจบัณฑิตย์



## การคัดกรองสุขภาพเบื้องต้นโดยใช้เทคนิคเหมืองข้อมูล

กิตติศักดิ์ สุมามาลัย

อรรวรรณ อิ่มสมบัติ

### บทคัดย่อ

งานวิจัยนี้ได้นำเสนอการวิเคราะห์ข้อมูลจากแบบสำรวจสภาวะสุขภาพประชาชนโดยใช้ข้อมูลของประชาชนในเขตเทศบาลเมืองจังหวัดบุรีรัมย์ ปี พ.ศ. 2555 จำนวน 1 ,071 เรคอร์ด ด้วยเทคนิคการทำเหมืองข้อมูล โดยแบ่งการศึกษาเป็น 2 ประเด็นคือ การศึกษาความสัมพันธ์ของการเป็นโรคต่างๆโดยใช้เทคนิค Association Rule และศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชนโดยแบ่งออกเป็น กลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย ด้วยเทคนิค Classification และนำผลลัพธ์ที่ได้ไปพัฒนาเป็นระบบคัดกรองสุขภาพเบื้องต้นเพื่อช่วยลดภาระงานของผู้เชี่ยวชาญในการวิเคราะห์ข้อมูล

ผลการทดลองพบว่า การศึกษาความสัมพันธ์ของการเป็นโรคต่างๆอัลกอริทึม FP Growth ให้คำตอบที่เหมาะสมมากกว่าอัลกอริทึม Aprioriและในการศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชน พบว่าอัลกอริทึม Partial Rules ให้ความถูกต้องมากกว่าอัลกอริทึมต้นไม้ตัดสินใจ C4.5 โดยมีค่าความถูกต้อง (precision) เท่ากับ 88.60% ค่าระลึก (recall) เท่ากับ 89.20% และค่าความเหวี่ยง (F-measure) เท่ากับ 88.80% เมื่อนำกฎที่ได้ไปให้ผู้เชี่ยวชาญตรวจสอบผู้เชี่ยวชาญได้ตัดข้อมูลบางแอทริบิวต์และกฎบางกฎที่ไม่เหมาะสมออกไป สุดท้ายจะได้ผลลัพธ์เป็นกฎจำนวน 44 กฎ และถูกนำไปในการสร้างระบบคัดกรองสุขภาพอัตโนมัติสำหรับให้ประชาชนทั่วไปเข้ามาตรวจสอบสุขภาพเบื้องต้นของตนเอง

### บทนำ

สำนักงานหลักประกันสุขภาพแห่งชาติได้มีนโยบายส่งเสริมสนับสนุนการเฝ้าระวังโรคเรื้อรังในทุกจังหวัด โดยได้มีการแจกแบบสำรวจสภาวะสุขภาพของประชาชนอายุตั้งแต่ 15 ปีเพื่อใช้ในการคัดกรองสุขภาพเบื้องต้นเป็นประจำทุกปี ทั้งนี้ศูนย์แพทย์ชุมชนสังกัดโรงพยาบาลบุรีรัมย์ได้เล็งเห็นความสำคัญของนโยบายดังกล่าว จึงได้จัดทำการสำรวจสภาวะสุขภาพของประชาชนอายุตั้งแต่ 15 ปีในเขตเทศบาล ซึ่งข้อมูลประกอบด้วย 1) ข้อมูลทั่วไป 2) ข้อมูลครอบครัว 3) ข้อมูลการเจ็บป่วยและพฤติกรรมเมื่อเจ็บป่วย 4) ข้อมูลคะแนนพฤติกรรมด้านสุขภาพ เพื่อนำข้อมูลที่ได้ออกมาวิเคราะห์ และออกแบบกิจกรรมและงบประมาณให้สอดคล้องกับผลที่สำรวจได้เพื่อแก้ไข

ปัญหาสุขภาพให้ตรงจุด เช่น ผลการคัดกรองสุขภาพเบื้องต้นในชุมชนหนึ่งพบว่ามีคนที่อยู่ในกลุ่มเสี่ยงเป็นจำนวนมาก และเมื่อนำมาวิเคราะห์ทำให้ทราบว่าส่วนใหญ่เป็นคนอ้วนลงพุง จึงจัดให้มีการอบรมเพื่อปรับเปลี่ยนพฤติกรรมในกลุ่มเสี่ยงของชุมชนนั้นด้วยโครงการชุมชนลดหวาน มัน เค็ม ลดอ้วน ลดโรคใน ปี 2555 เป็นต้น

ข้อมูลที่สำรวจได้จากแบบสำรวจสภาวะสุขภาพของประชาชนจะมีปริมาณเพิ่มสูงขึ้นทุกปี

และยังขาดการจัดการข้อมูลเพื่อนำไปใช้ประโยชน์อย่างคุ้มค่าจากการทบทวนวรรณกรรมพบว่า มีงานวิจัยหลายชิ้นที่ได้ประยุกต์เทคนิคการทำเหมืองข้อมูลมาช่วยในการวิเคราะห์ข้อมูลด้าน

การแพทย์และสาธารณสุข เช่น มาลีวรรณ บุญพลอย (2551) ศึกษาเรื่องเหมืองข้อมูลและการค้นหาองค์ความรู้สำหรับโรคกล้ามเนื้ออ่อนแรงแรงนำเสนอวิธีการใช้ขั้นตอนวิธีต้นไม้ช่วยตัดสินใจสองชั้น

เพื่อจำแนกประเภทข้อมูลโรคกล้ามเนื้ออ่อนแรงแรงเรื้อรัง , รักถัน เหลาหา (2553) ศึกษาการพยากรณ์

ความเสี่ยงการเกิดโรคมะเร็งปอดโดยใช้ทฤษฎีต้นไม้ตัดสินใจแบบ C4.5 ในการจัดกลุ่มผู้ป่วยและ

พยากรณ์ความเสี่ยงการเกิดโรคมะเร็งปอด , ศุภกรใจ วุฒิกิจโกศล (2551) ศึกษาการใช้เทคนิคการทำ

เหมืองข้อมูลในผู้ป่วยข้อไหล่นิด โรงพยาบาลพระนั่งเกล้าเป็นแนวทางช่วยสนับสนุนการตัดสินใจ

เลือกเทคนิคการรักษาสำหรับนักกายภาพบำบัดด้วยเทคนิค Cluster และ Association Rule, อังคณา

พิจารโชติ (2552) พัฒนาระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็น

โรคเบาหวาน ประกอบด้วยสามส่วนคือ ส่วนของคลังข้อมูลใช้สำหรับเก็บข้อมูลปัจจัยเสี่ยงที่ได้จาก

การคัดกรองกลุ่มเสี่ยง ส่วนของเหมืองข้อมูลเป็นตัวค้นหาความสัมพันธ์ของปัจจัยเสี่ยงต่างๆ และ

ส่วนการแสดงผลข้อมูลรายงาน ผลการค้นหาความสัมพันธ์ปัจจัยเสี่ยงที่ทำให้เกิดโรคเบาหวานพบว่า

ประวัติครอบครัวเป็นปัจจัยเสี่ยงที่มีอัตราการเสี่ยงมากที่สุดซึ่งพบคู่กับค่าดัชนีมวลกาย

งานวิจัยนี้จึงนำเสนอการวิเคราะห์ข้อมูลจากแบบสำรวจสภาวะสุขภาพประชาชนด้วย

เทคนิคการทำเหมืองข้อมูลโดยใช้เทคนิค Association Rule และ Classification เพื่อนำไปสร้าง

ระบบเว็บแอปพลิเคชันเพื่อช่วยคัดกรองสุขภาพเบื้องต้น ซึ่งเปิดให้ประชาชนทั่วไปที่สนใจสามารถ

เข้ามาสำรวจสุขภาพได้ด้วยตนเอง และช่วยลดภาระงานของผู้เชี่ยวชาญ

### วิธีดำเนินการวิจัย

ขั้นตอนในดำเนินการวิจัยการคัดกรองสุขภาพเบื้องต้น โดยใช้เทคนิคเหมืองข้อมูล มี

ขั้นตอนดังนี้คือ

1. การศึกษาและรวบรวมข้อมูล
  2. การเตรียมข้อมูล
  3. การศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม
- โดยแต่ละขั้นตอนมีรายละเอียดดังนี้

#### การศึกษาและรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลที่ได้จากแบบสำรวจสภาวะสุขภาพประชาชนซึ่งจัดทำโดยศูนย์แพทย์ชุมชนสังกัดโรงพยาบาลบุรีรัมย์ โดยข้อมูลที่ใช้ในการศึกษารั้งนี้ เป็นข้อมูลจากการสำรวจสภาวะสุขภาพประชาชนในเขตเทศบาลเมืองบุรีรัมย์ จำนวน 1,071 แถว ซึ่งประกอบด้วย แอทธิบิวต์ จำนวน 75 แอทธิบิวต์ เช่น ลำดับที่ ชื่อ-สกุลรหัสบัตรประชาชนที่อยู่ชุมชนสิทธิการรักษา เพศสถานภาพ การศึกษาน้ำหนักส่วนสูงดัชนีมวลกายขนาดรอบเอวความดันประวัติการเจ็บป่วยของครอบครัว ประวัติการเจ็บป่วยการรักษาการสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การออกกำลังกาย รสชาติของอาหารที่ชอบ การขับถ่ายเพศสัมพันธ์ การตรวจเต้านม การตรวจมะเร็งปากมดลูก พฤติกรรมสุขภาพด้านอาหาร อารมณ์ ออกกำลังกายและสรุปผลการคัดกรองสุขภาพ เป็นต้น

#### การเตรียมข้อมูล (Data Pre-processing)

เนื่องจากข้อมูลที่ถูกจัดเก็บไว้ยังไม่อยู่ในรูปแบบที่สามารถนำมาใช้ในการทดสอบและวิเคราะห์ด้วยเครื่องมือการทำเหมืองข้อมูล จึงต้องมีการเตรียมข้อมูลก่อนนำไปใช้ในการทำเหมืองข้อมูลเพื่อให้ได้ผลการทดลองที่มีประสิทธิภาพสูงสุด โดยในขั้นตอนนี้จะประกอบด้วยขั้นตอนย่อย 3 ขั้นตอนคือ การลดขนาดของข้อมูล (Data Reduction) การทำให้เป็นค่าไม่ต่อเนื่อง (Discretization) การทำความสะอาดข้อมูล (Data Cleaning) และการแปลงข้อมูล (Data Transformation)

การลดขนาดข้อมูล (Data Reduction): เนื่องจากข้อมูลมีแอทธิบิวต์อยู่เป็นจำนวนมาก จึงจำเป็นต้องมีการลดแอทธิบิวต์บางตัวที่ไม่เกี่ยวข้องกับการคัดกรองออกไป ได้แก่ ลำดับที่ชื่อ-สกุลรหัสบัตรประชาชนที่อยู่ชุมชนสิทธิการรักษา นอกจากนี้ยังตัดแอทธิบิวต์น้ำหนักและส่วนสูงออกไป เนื่องจากมีการข้อมูลค่าดัชนีมวลกาย (BMI) ซึ่งคำนวณมาจากค่าน้ำหนัก (กิโลกรัม) หารด้วยส่วนสูง (เมตร) ยกกำลังสอง

การทำให้เป็นค่าไม่ต่อเนื่อง (Discretization): เนื่องจากบางอัลกอริทึมที่ใช้ในการศึกษารั้งนี้ ไม่รองรับข้อมูลที่เป็นค่าไม่ต่อเนื่อง จึงจำเป็นต้องมีการทำให้เป็นค่าไม่ต่อเนื่อง เช่น ค่าดัชนีมวลกาย (BMI) จะถูกทำให้เป็นค่าไม่ต่อเนื่องดังนี้ 1 หมายถึงผอม(น้อยกว่า 18.5) 2 หมายถึง ปกติ

(18.5-22.99) 3 หมายถึง ท้วม (23-24.99) 4 หมายถึง อ้วน (25-29.99) และ 5 หมายถึง อ้วนมาก (มากกว่าหรือเท่ากับ 30)

การทำความสะอาดข้อมูล (Data Cleaning) จากการวิเคราะห์ข้อมูลที่น่ามาศึกษาพบว่า มีข้อมูลบางแอทริบิวต์มีค่าข้อมูลที่ผิดพลาด เช่น แอทริบิวต์การตรวจเต้านมด้วยตนเอง ซึ่งมีค่าที่เป็นไปได้คือ 1 (หมายถึง ตรวจเต้านมด้วยตนเองได้) 2 (หมายถึง ตรวจเต้านมด้วยตนเองไม่ได้) 3 (หมายถึง ไม่แน่ใจ) เท่านั้น แต่ปรากฏมีข้อมูลตัวเลข 5 และ 6 เป็นต้น จึงทำการแก้ไขโดยพิจารณาข้อมูลเรคอร์ดอื่นที่มีลักษณะใกล้เคียงกัน หรือมีค่าแอทริบิวต์อื่นๆที่เกี่ยวข้องเป็นค่าเดียวกัน เช่น พิจารณาจากแอทริบิวต์ความถี่ในการตรวจเต้านมด้วยตนเองและผลตรวจเต้านมด้วยตนเองหากมีค่าเท่ากับ 6 (หมายถึง ไม่เคยตรวจ) และ 3 (หมายถึง จำไม่ได้) ค่าแอทริบิวต์การตรวจเต้านมด้วยตนเองจะเป็น 2 โดยในขั้นตอนนี้พบข้อมูลที่มีค่าผิดพลาดและแก้ไขให้ข้อมูลอยู่ในช่วงที่ถูกต้องได้ทั้งหมดจำนวน 4 เรคอร์ด

การแปลงข้อมูล (Data Transformation): เนื่องจากข้อความบางข้อในแบบสำรวจ ผู้ตอบคำถามสามารถเลือกตอบได้มากกว่า 1 คำตอบ เช่น ประวัติการเจ็บป่วยของบิดามารดา รสชาติของอาหารที่ชอบ และเจ้าหน้าที่พยาบาล ได้บันทึกข้อมูลในรูปแบบข้อความที่เป็นหมายเลข คำตอบต่างๆเชื่อมต่อกันด้วยเครื่องหมาย + เช่น 1+3 หมายถึงเป็นโรคลำดับที่ 1 และโรคลำดับที่ 3 ซึ่งข้อมูลในรูปแบบนี้ไม่เหมาะสมในการนำไปใช้ ผู้วิจัยจึงได้ทำการแปลงข้อมูลให้อยู่ในรูปแบบ Binary (มีค่า 0 และ 1) โดย 1 ตัวเลือกของคำถามในลักษณะนี้จะถูกนำไปเป็น 1 แอทริบิวต์ และหากผู้ตอบแบบสอบถามเลือกตอบตัวเลือกนี้จะมีเป็น 1 และเป็น 0 เมื่อไม่เลือกตัวเลือกนี้ เช่น แอทริบิวต์ประวัติการเจ็บป่วยของพี่น้องสายตรง มีตัวเลือกเป็นโรคต่างๆจำนวน 10 ตัวเลือก ดังนั้นทุกตัวเลือกจะถูกนำมาสร้างเป็นแอทริบิวต์ 10 แอทริบิวต์ และกำหนดค่าของข้อมูลเป็น 1 หรือ 0 ตามคำตอบของผู้ตอบแบบสอบถาม จากขั้นตอนการเตรียมข้อมูลนี้ทำให้มีแอทริบิวต์ที่ถูกนำมาใช้ในการทดลองทั้งหมด 81 แอทริบิวต์

### การศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม

งานวิจัยครั้งนี้ได้ศึกษาการใช้เทคนิค Association Rule เพื่อหาความสัมพันธ์ของการเป็นโรค และศึกษาการใช้เทคนิค Classification เพื่อคัดกรองสุขภาพเบื้องต้นของประชาชน ซึ่งมีรายละเอียดในการศึกษาอัลกอริทึมที่เกี่ยวข้องดังนี้

เทคนิคการหาความสัมพันธ์ (Association Rule) ใช้ในการหาความสัมพันธ์ของการเป็นโรค 6 โรคได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง โรคตับ โรคอัมพาต โรคหัวใจ และโรคไขมันในเลือดผิดปกติอัลกอริทึมที่ใช้ในการทดลองครั้งนี้มี 2 อัลกอริทึม คือ Apriori

(Agrawalet al., 1993) และ FP Growth (Kantardzic M., 2003) ซึ่งผลที่ได้จากการศึกษาวิเคราะห์การทำงานแต่ละอัลกอริทึมมีดังนี้

1. Apriori เป็นอัลกอริทึมที่ได้รับความนิยมในการนำมาวิเคราะห์ข้อมูลเพื่อสร้างกฎความสัมพันธ์ และพบว่ากฎความสัมพันธ์ที่ได้เป็นการนำค่าของข้อมูลทุกค่าที่เป็นไปได้มาสร้างเป็นกฎ ทำให้เงื่อนไขในกฎส่วนใหญ่มีค่าเป็น Negative ตัวอย่างเช่น ไม่เป็นโรคที่ 1 และไม่เป็นโรคที่ 2 แล้วจะไม่เป็นโรคที่ 4 เป็นต้น

2. FP Growth ให้ผลลัพธ์ในรูปแบบกฎความสัมพันธ์และจะสร้างกฎที่มีเงื่อนไขเป็นค่าเช่น หากเป็นโรคที่ 1 และเป็นโรคที่ 4 แล้วจะเป็นโรคที่ 5 เป็นต้น

การศึกษาคำใช้เทคนิค Classification เพื่อคัดกรองสุขภาพเบื้องต้นของประชาชนในครั้งนี้มีวัตถุประสงค์เพื่อนำโมเดลที่ได้ไปพัฒนาเป็นระบบคัดกรองสุขภาพประชาชนแบบอัตโนมัติ ดังนั้นเทคนิคการจำแนกกลุ่มที่นำมาใช้ในการศึกษาครั้งนี้จึงเป็นเทคนิคที่ให้ผลลัพธ์ในรูปแบบกฎเพื่อให้สามารถทำความเข้าใจได้ง่ายและสามารถนำไปใช้พัฒนาระบบได้ อัลกอริทึมที่ศึกษาได้แก่

1. Decision Tree: C4.5 (Kantardzic M., 2003) เป็นการจัดกลุ่มโดยผลลัพธ์ที่ได้อยู่ในรูปของต้นไม้การตัดสินใจที่มีกิ่งต้นไม้จำนวนมาก

2. Decision Rules: Partial Rules (Kantardzic M., 2003) เป็นการจัดกลุ่มโดยผลลัพธ์ที่ได้จะอยู่ในรูปของกฎที่นำเอาข้อมูลในแต่ละแอทริบิวต์มาเป็นเงื่อนไขในการตัดสินใจ

การศึกษานี้ใช้โปรแกรม Weka (<http://www.cs.waikato.ac.nz/ml/weka/>) เวอร์ชัน 3.6.7 เป็นเครื่องมือในการทดลอง และข้อมูลที่ใช้ในการทดลองด้วยเทคนิคการหาความสัมพันธ์ (Association Rule) มีจำนวน 192 เรคอร์ด (ใช้เฉพาะข้อมูลบุคคลที่เป็นโรคมามากกว่า 1 โรค) และใช้ข้อมูลจำนวน 1,071 เรคอร์ดในการทดลองด้วยเทคนิค Classification โดยได้ทำการทดลองแบบ 10 folds cross-validation ซึ่งเป็นการแบ่งกลุ่มข้อมูลออกเป็น 10 กลุ่ม ในแต่ละรอบจะนำกลุ่มข้อมูลจำนวน 9 กลุ่มเป็นกลุ่มศึกษา (Training Set) และกลุ่มข้อมูลที่เหลือเป็นกลุ่มทดสอบ (Testing Set) โดยจะทำซ้ำจำนวน 10 รอบ เพื่อเปลี่ยนกลุ่มทดสอบให้ครบทุกกลุ่ม และนำผลลัพธ์ที่ได้มาหาค่าเฉลี่ย โดยเกณฑ์ที่ใช้ในการพิจารณาคุณภาพของโมเดลที่ได้จากแต่ละอัลกอริทึม จะพิจารณาจากเปอร์เซ็นต์ความถูกต้อง (Precision) ค่าระลึก (Recall) และค่าความเหวี่ยง (F-measure) ที่มีค่ามากที่สุด

## ผลการวิจัย

การศึกษาความสัมพันธ์ของการเป็นโรคต่างๆ

ในการศึกษาการใช้เทคนิค Association Rule เพื่อหาความสัมพันธ์ของโรค พบว่า ข้อมูลที่มี มีจำนวนผู้ป่วยเป็นโรคดังกล่าวจำนวนน้อย คือ 192 เรคอร์ด ดังนั้นในการทดลองจะ กำหนดค่าสนับสนุนขั้นต่ำไว้ค่อนข้างต่ำ เพื่อให้โอกาสในเกิดกฎเกณฑ์ที่น่าสนใจเพิ่มมากขึ้น โดย ค่าที่เป็นไปได้ของโรคต่างๆเหล่านี้มี 2 ค่าคือ 1 หมายถึงมีประวัติการเป็นโรค และ 0 หมายถึง ไม่มีประวัติการเป็นโรค

ผลการหาความสัมพันธ์แบบ Apriori โดยการทดลองด้วยการตั้งค่าสนับสนุนขั้นต่ำ ของกฎ 0.01 และค่า lift ที่ 0.90 พบว่าได้ผลลัพธ์เป็นกฎความสัมพันธ์ทั้งหมด 1,824 กฎ และผลลัพธ์ ที่ได้ส่วนใหญ่จะเป็นกฎที่มีค่าเป็นปฏิเสธ (ไม่มีประวัติการเป็นโรค) ผสมอยู่ในกฎด้วย ซึ่งกฎใน ลักษณะดังกล่าวจะไม่ค่อยมีประโยชน์มากนักเนื่องจากคนที่มีสุขภาพปกติทั่วไปก็จะเป็นโรค ต่างๆอยู่แล้ว ยกตัวอย่างเช่น ถ้ามีประวัติการเป็นโรคความดันโลหิตสูงและไม่มีประวัติการเป็นโรค อัมพาต จะไม่เป็นโรคตับ ซึ่งเกิดขึ้นจำนวน 138 เรคอร์ด เป็นต้น กฎที่ไม่มีค่าเป็นปฏิเสธผสมอยู่จะ มีจำนวนน้อยมาก ในขณะที่อัลกอริทึม FP Growth จะได้กฎที่แสดงเฉพาะค่า positive (มีประวัติ การเป็นโรค) เท่านั้นและมีจำนวนค่อนข้างมาก

ผลการหาความสัมพันธ์แบบ FP Growth จากการทดลองด้วยค่าสนับสนุนขั้นต่ำของ กฎ 0.01 และค่า lift ที่ 0.9 พบว่าได้ผลลัพธ์ทั้งหมด 18 กฎ และกฎที่ได้จะเป็นความสัมพันธ์ของการ เป็นโรคต่างๆ (ไม่มีค่าที่เป็นปฏิเสธ) จากวิเคราะห์ผลลัพธ์ที่ได้พบว่าได้กฎที่มีกลุ่มโรคที่ต่างกัน กำหนดซึ่งกัน โดยได้แบ่งได้ 2 กลุ่ม กลุ่มแรกคือกลุ่มของโรคความดันโลหิตสูง โรคหัวใจ และโรค ไชมันในเลือดผิดปกติ กลุ่มที่สองคือกลุ่ม โรคเบาหวาน โรคความดันโลหิตสูง และโรคหัวใจ จาก ความสัมพันธ์นี้สรุปได้ว่าเมื่อเป็นโรคใดๆในกลุ่มแล้วมักจะมีโรคอื่นๆในกลุ่มแทรกซ้อนตามมา เช่น ถ้าคนไข้เป็นโรคหัวใจและโรคความดันโลหิตสูงอยู่แล้ว ก็มักจะเป็นโรคเบาหวานและโรค ไชมันในเลือดผิดปกติซึ่งสามารถประยุกต์ใช้ความรู้นี้ในการประชาสัมพันธ์ผู้ป่วยเพื่อดูแลตนเอง และป้องกันการเกิดโรคแทรกซ้อนต่างๆตามกฎที่ค้นพบ

การศึกษารายการพยากรณ์การคัดกรองสุขภาพเบื้องต้นของประชาชน

ในการศึกษาการใช้เทคนิค Classification เพื่อจำแนกกลุ่มประชากรออกเป็น 3 กลุ่ม คือ กลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วย ด้วยเทคนิค Decision Tree: C4.5 และ Decision Rules: Partial Rules โดยจะเลือกเทคนิคที่มีค่าความถูกต้องสูงที่สุดเพื่อนำไปพัฒนาระบบเว็บแอปพลิเคชันเพื่อช่วย

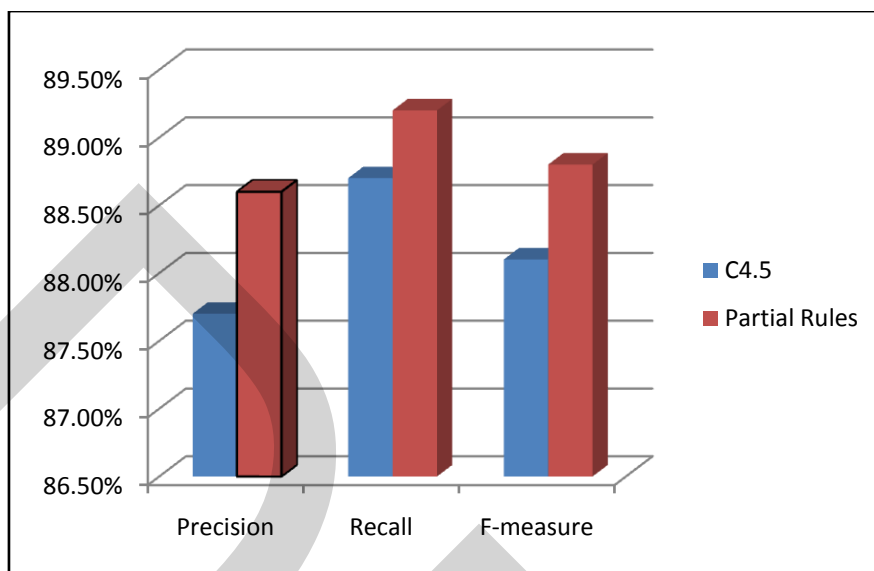
คัดกรองสุขภาพเบื้องต้นสำหรับประชาชนทั่วไป ตารางที่ 1 แสดงผลการทดสอบการจำแนกข้อมูล การคัดกรองสุขภาพเบื้องต้นของประชาชนด้วยเทคนิค C4.5 และ Partial Rules

ตารางที่ 1 ผลการทดสอบการจำแนกข้อมูลด้วยอัลกอริทึม Decision Tree: C4.5 และ Decision Rules: Partial Rules

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

การทดลองจำแนกกลุ่มด้วยอัลกอริทึม Decision Tree: C4.5 พบว่าเทคนิค Decision Tree: C4.5 ให้ผลลัพธ์การจำแนกที่มีความถูกต้อง (Precision) เท่ากับ 87.70% ค่าระลึก (Recall) เท่ากับ 88.70% และค่าความเหวี่ยง (F-measure) เท่ากับ 88.10% ตัวอย่างกฎที่ได้ เช่น หากบุคคลนั้นไม่มีประวัติการเจ็บป่วย และมีค่าดัชนีมวลกายน้อยกว่า 18.5 (ผอม) และพื่น้องสายตรงไม่เคยมีประวัติเจ็บป่วย และดื่มเครื่องดื่มที่มีแอลกอฮอล์ จะมีสุขภาพอยู่ในกลุ่มเสี่ยง โดยมีข้อมูลที่สอดคล้องกับกฎนี้จำนวน 8 ตัวอย่าง และไม่มีข้อมูลที่ขัดแย้งกับกฎนี้

ผลการจำแนกกลุ่มแบบ Decision Rules: Partial Rules พบว่าผลลัพธ์การจำแนกมีความถูกต้อง (Precision) เท่ากับ 88.60% ค่าระลึก (Recall) เท่ากับ 89.20% และค่าความเหวี่ยง (F-measure) เท่ากับ 88.80% ตัวอย่างกฎที่ได้ เช่น หากบุคคลไม่มีประวัติการเจ็บป่วยและมีค่าดัชนีมวลกายเท่ากับ 23-24.99 (ท้วม) จะมีสุขภาพอยู่ในกลุ่มเสี่ยง โดยมีข้อมูลที่สอดคล้องกับกฎนี้จำนวน 253 ตัวอย่าง และไม่มีข้อมูลที่ขัดแย้งกับกฎนี้



ภาพที่ 1 ผลการทดลองของอัลกอริทึม C4.5 และ Partial Rules

จากผลการทดลองการจำแนกกลุ่มข้อมูลด้วยอัลกอริทึมทั้ง 2 แบบ พบว่าเทคนิค Decision Tree: Partial rules ให้ผลลัพธ์ที่มีค่าความถูกต้อง (Precision) ค่าระลึก (Recall) และค่าความเฉลี่ย (F-measure) สูงกว่าอัลกอริทึม Decision Tree: C4.5 จึงได้เลือกกฎที่ได้จากอัลกอริทึม Decision Tree: Partial rules เพื่อนำไปพัฒนาระบบ โดยเมื่อนำผลลัพธ์ที่ได้จากการทดลองด้วยอัลกอริทึม Decision Tree: Partial rules มาพิจารณาความผิดพลาดในการจำแนกกลุ่มต่างๆ ดังแสดงในตารางที่ 2

ตารางที่ 2 รายละเอียดผลการจำแนกกลุ่มด้วยอัลกอริทึม Decision Tree: Partial rules

การจำแนกกลุ่ม		ผลการทำนาย		
		กลุ่มปกติ	กลุ่มเสี่ยง	กลุ่มป่วย
ข้อมูลจริง	กลุ่มปกติ	44	64	0
	กลุ่มเสี่ยง	50	704	1
	กลุ่มป่วย	0	1	207



พบว่าผลการจำแนกกลุ่ม Class 1 (กลุ่มปกติ) มีความผิดพลาดสูงคือ 50 ตัวอย่างจากทั้งหมด 94 ตัวอย่าง คิดเป็น 53.2 % เนื่องจากมีจำนวนตัวอย่างให้ระบบเรียนรู้น้อย โดยมีอยู่เพียง 108 ตัวอย่างเท่านั้น ทำให้ค่า Precision กับ Recall ของการจำแนกกลุ่ม Class 1 (กลุ่มปกติ) ที่ได้มีค่าต่ำคือ 46.8% และ 40.7% ตามลำดับ วิธีแก้ไขคือต้องใส่ข้อมูลที่เป็นกลุ่มปกติเข้าไปให้ระบบได้เรียนรู้เพิ่มอีก เพื่อที่จะได้รูปแบบการจำแนกกลุ่มที่ถูกต้องมากยิ่งขึ้น

เมื่อนำกฎให้ผู้เชี่ยวชาญช่วยตรวจสอบพบว่าแอทริบิวต์การปฏิบัติคนเมื่อเจ็บป่วยเป็นข้อมูลที่ไม่เหมาะสมที่จะใช้จำแนกกฎเนื่องจากเป็นแอทริบิวต์ที่มองภาพรวมของประวัติการเจ็บป่วยและอาการต่างๆทั้งหมด 16 แอทริบิวต์ แต่มีเพียง 6 แอทริบิวต์ที่ถ้าเป็นโรคแล้วผู้เชี่ยวชาญจะถือว่าเป็นกลุ่มป่วยทันที ได้แก่ โรคเบาหวาน โรคความดันโลหิตสูง โรคตับ โรคอัมพาต โรคหัวใจ และ โรคไขมันในเลือดผิดปกติ ดังนั้นแอทริบิวต์จึงเป็นสาเหตุที่ทำให้เกิดการสร้างกฎที่ผิดพลาดจึงทำทดลองตัดแอทริบิวต์ treat ออก พบว่าการจำแนกกลุ่มปกติ กลุ่มเสี่ยง และกลุ่มป่วยให้ผลลัพธ์ที่ดีขึ้น โดยมีความถูกต้อง (Precision) เท่ากับ 89.20% ค่าระลึก (Recall) เท่ากับ 89.40% และ ค่าความเหวี่ยง (F-measure) เท่ากับ 89.30%ตามลำดับ ได้จำนวนกฎทั้งสิ้น 55 กฎ แต่เนื่องจากกฎที่ได้มีจำนวนมาก และกฎบางกฎมีเงื่อนไขไม่เหมาะสม ดังนั้นผู้เชี่ยวชาญจึงได้ลบกฎบางกฎทิ้งไปเหลือเพียง 44 กฎ เช่น หากบุคคลนั้นมีพี่น้องสายตรงที่มีประวัติเป็นโรคความดันโลหิตสูง ผลการคัดกรองสุขภาพจะอยู่ในกลุ่มปกติ เป็นกฎไม่เหมาะสมเนื่องจากโรคความดันโลหิตสูงเป็นโรคที่มีความเกี่ยวข้องกับพันธุกรรม การที่มีพี่น้องสายตรงเป็นโรสดังกล่าว บุคคลนั้นจะมีโอกาสสูงที่จะเป็นโรคความดันโลหิตสูง

#### การพัฒนาบบคัดกรองสุขภาพเบื้องต้นแบบอัตโนมัติ

หน้าจอหลักประกอบด้วยเมนู หน้าหลักสำหรับเกริ่นนำให้ผู้ใช้งานระบบคัดกรองสุขภาพเบื้องต้น วิธีการใช้งานและประโยชน์ที่จะได้รับ หน้าตรวจสอบสุขภาพสำหรับให้ผู้ใช้บันทึกข้อมูลสุขภาพ เมื่อทำการบันทึกข้อมูลเสร็จระบบจะประมวลผลแล้วแจ้งผลการคัดกรองสุขภาพเบื้องต้นให้ทราบทันที หน้าเกี่ยวกับเราสำหรับให้ข้อมูลรายละเอียดต่างๆเกี่ยวกับที่มาที่ไปของระบบ ติดต่อเราสำหรับติดต่อผู้ดูแลระบบ

ภาพที่ 2 หน้าจอแบบสำรวจสภาวะสุขภาพประชาชน

### อภิปรายผล

ในการหาความสัมพันธ์ (Association Rule) ของการเป็นโรคได้ทดสอบกับ อัลกอริทึม Apriori และอัลกอริทึม FP Growth พบว่าอัลกอริทึม FP Growth ให้ผลลัพธ์ที่เหมาะสมมากกว่า อัลกอริทึม Apriori เพราะจะให้ผลลัพธ์เป็นกฎที่มีแต่ค่า positive (มีประวัติการเป็นโรค) เท่านั้น ในขณะที่ Apriori จะให้ผลลัพธ์ที่เป็นกฎที่มีทั้งค่า positive และ negative (ไม่มีประวัติการเป็นโรค) ผสมกัน ซึ่งไม่ค่อยมีประโยชน์มากนักในการนำมาใช้เพื่อทำนายการเกิดโรคใหม่ในอนาคตของผู้ป่วย ส่วนการศึกษาคัดกรองสุขภาพเบื้องต้นด้วยเทคนิคการจำแนกกลุ่ม (Classification) ได้ทำการศึกษาเฉพาะอัลกอริทึมที่ให้ผลลัพธ์เป็นกฎเนื่องจากต้องนำผลลัพธ์ที่ได้ในรูปแบบกฎไปพัฒนาเป็นระบบคัดกรองอัตโนมัติต่อไป ซึ่งได้แก่ อัลกอริทึม Decision Tree: C4.5 และอัลกอริทึม Decision Tree: Partial Rules จากการทดสอบพบว่าอัลกอริทึม Decision Tree: Partial Rules ให้ผลลัพธ์ที่ดีกว่าด้วยค่าความถูกต้อง (Precision) เท่ากับ 88.60% ค่าระลึก (Recall) เท่ากับ 89.20% และค่าความเหวี่ยง (F-measure) เท่ากับ 88.80% และเมื่อนำกฎที่ได้ไปให้ผู้เชี่ยวชาญตรวจสอบพบว่าแอทริบิวต์การปฏิบัติตนเมื่อเจ็บป่วยเป็นข้อมูลที่ไมเหมาะสมที่จะใช้จำแนกกฎเนื่องจากเป็นแอทริบิวต์ที่มองภาพรวมของประวัติการเจ็บป่วยและอาการต่างๆ และเป็นสาเหตุที่ทำให้เกิดการสร้างกฎที่ผิดพลาด ผู้เชี่ยวชาญจึงแนะนำให้ตัดแอทริบิวต์นี้ออก และเมื่อนำข้อมูลมาทดลองใหม่ พบว่าการจำแนกกลุ่มให้ผลลัพธ์ที่ดีขึ้น โดยมีค่าความถูกต้อง (Precision) เท่ากับ 89.20% ค่าระลึก

(Recall) เท่ากับ 89.40% และ ค่าความเหวี่ยง (F-measure) เท่ากับ 89.30% ตามลำดับ นอกจากนี้ ผู้เชี่ยวชาญยังได้ตัดกฎที่ไม่เหมาะสมออกอีกคงเหลือกฎเพียง 44 กฎ และนำไปพัฒนาเป็นระบบเว็บ แอปพลิเคชันเพื่อช่วยคัดกรองสุขภาพเบื้องต้น ซึ่งเปิดให้ประชาชนทั่วไปที่สนใจได้เข้ามาสำรวจ สุขภาพด้วยตนเอง

### บรรณานุกรม

มาลิวรรณ บุญพลอย (2551) เหมืองข้อมูลและการค้นหาคำความรู้สำหรับโรคกล้ามเนื้ออ่อนแรง  
รักถิ่น เหลลหา (2553) การพยากรณ์ความเสี่ยงการเกิดโรคมะเร็งปอดโดยใช้ทฤษฎีของการทำ

เหมืองข้อมูล

ศุภกรใจ วุฒิกิจ โสศล (2551) การใช้เทคนิคการทำเหมืองข้อมูลในผู้ป่วยข้อไหล่ติด โรงพยาบาลพระ

นั่งเกล้า

อังคณา พิจารโชติ (2552) ระบบสนับสนุนการตัดสินใจสำหรับวิเคราะห์ปัจจัยเสี่ยงการเป็น

โรคเบาหวาน

Agrawal et al. (1993) Mining Association Rules Between Sets of Items in Large Databases. In P. Buneman and S. Jajordai, eds. New York: ACM.

Kantardzic M. (2003) Data mining Concepts, Models, Methods, and Algorithms. United States of America: IEEE Press.

Han, J., et al. (2006) Data mining concepts and techniques (2nd ed.). United States of America: Morgan Kaufman Publishers.

Roiger, J., et al. (2003) Data Mining: a tutorial-based primer. United States of America: Pearson Education, Inc.

WEKA Machine Learning Group at University of Waikato, from

<http://www.cs.waikato.ac.nz/ml/weka/>

## ประวัติผู้เขียน

ชื่อ-นามสกุล

นาย กิตติศักดิ์ สุมามาลย์

ประวัติการศึกษา

วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมสิ่งแวดล้อม

มหาวิทยาลัยเกษตรศาสตร์ พ.ศ. 2547

บริหารธุรกิจมหาบัณฑิต สาขาการเงิน

มหาวิทยาลัยหอการค้าไทย พ.ศ. 2550

ประสบการณ์ทำงาน

- วิศวกร บริษัท โซคอนเอนจิเนียริงจำกัด

- Web Programmer บริษัท ดิจิมิเดียเอเชียจำกัด