

บทที่ 3

วิธีการดำเนินการวิจัย

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูลมาใช้เพื่อช่วยพัฒนาระบบคัดกรองสุขภาพประชาชนแบบอัตโนมัติ โดยมีขั้นตอนการดำเนินการวิจัยดังนี้

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

โดยในบทนี้จะกล่าวถึงขั้นตอน การศึกษาและรวบรวมข้อมูล การเตรียมข้อมูลและวิเคราะห์ข้อมูล และ การศึกษา ทดสอบและเลือกเทคนิคที่เหมาะสม

3.1 การศึกษาและรวบรวมข้อมูล

แบบสำรวจสถานะสุขภาพประชาชนจัดทำโดยศูนย์แพทยุมนชนสังกัดโรงพยาบาลบุรีรัมย์ ใช้สำหรับเก็บข้อมูลจากประชาชนทั่วไปโดยในการศึกษานี้เป็นการศึกษาข้อมูลการสำรวจสถานะสุขภาพประชาชนจากประชาชน 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	ชุมชน	use	เพศ	อายุ	การศึกษา	น้ำหนัก	ส่วนสูง	BMI	sumBMI	hip	sumhip	ความดันโลหิต		
1	นส.เกศินี	#####	44/4	จ๊ะจ๊ะ	11	ว่าง	2	1	5	65	1.63	24.46	3	84	2	144	92
2	นางฉวีรัตน์	#####	86/7	อิตาณ	11	ชรัก	2	2	3	45	1.55	18.73	2	68	1	112	62
3	นายเวียงศักดิ์	#####	86/3	อิตาณ	11	บัตรทอง	1	2	5	85	1.78	26.83	4	97	2	145	89
4	นายอนุวัฒน์	#####	86/3	อิตาณ	11	บัตรทอง	1	1	4	54	1.74	17.84	1	77	1	113	65
5	นส.ณัฐกานต์	#####	86/3	อิตาณ	11	บัตรทอง	2	1	5	60	1.61	23.15	3	77	1	128	72
6	นางอมร	#####	86/3	อิตาณ	11	บัตรทอง	2	2	3	46	1.55	19.15	2	77	1	122	91
7	นายเสกสรรค์	#####	86/3	อิตาณ	11	บัตรทอง	1	1	3	65	1.75	21.22	2	82	1	120	76
8	นางสาวราวี	#####	10/3	จ๊ะจ๊ะ	11	บัตรทอง	2	3	2	52	1.58	20.83	2	85	2	176	79
9	นายบุญชู	#####	16	จ๊ะจ๊ะ	11	บัตรทอง	1	2	2	82	1.7	28.37	4	99	2	147	88
10	นางอนุชญา	#####	27/2	นคร	11	บัตรทอง	2	3	2	32	1.5	14.22	1	77	1	189	92
11	นายศุภสิทธิ์	#####	16	จ๊ะจ๊ะ	11	บัตรทอง	1	2	2	80	1.8	24.69	3	92	2	135	70
12	นายเส็ง	#####	27/1	นคร	11	ปกติ	1	2	2	65	1.65	23.88	3	92	2	129	87
13	นางอัจฉรา	#####	27/1	นคร	11	บัตรทอง	2	2	3	72	1.61	27.78	4	93	2	111	67
14	นส.อัจฉรา	#####	110	อนุวรรณ์	11	บัตรทอง	2	1	3	58	1.59	22.94	2	78	1	131	92
15	นายบรรณ	#####	10/3	จ๊ะจ๊ะ	11	บัตรทอง	1	2	4	68	1.75	22.20	2	88	1	127	84
16	นายไพโรจน์	#####	4/6	อิตาณ	11	บัตรทอง	1	2	7	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 เหม	#####	70 นคร	15	บัตรทอง	2	2	3	53	1.54	22.35	2	77	1	122	81	120	80	121
4	นส.สุธาดา ส	#####	70 นคร	15	บัตรทอง	2	1	3	45	1.6	17.58	1	66	1	100	77	97	69	99
5	นส.สุธาดา	#####	70 นคร	15	บัตรทอง	2	3	2	50	1.5	22.22	2	72	1	126	73	120	70	123
6	นายชัย ชื่นชม	#####	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.วีรลักษณ์	#####	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	นางจุฑารัตน์	#####	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.สุปณี ช	#####	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	นายนิธิพล	#####	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	นางสุกมล	#####	40/2 นค	15	บัตรทอง	1	2	4	102	1.66	37.02	5	107	2	113	72	116	65	115
23	นางสุกมล	#####	46/2 นค	15	ปกส	1	2	5	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 หมายถึงผู้ชายที่มีขนาดรอบเอวมมากกว่าหรือเท่ากับ 90 ซม. ผู้หญิงที่มีขนาดรอบเอวมมากกว่าหรือเท่ากับ 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

	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	BV	BW	BX	BY	BZ
1	BM	sumBM	hip	sumhip	ความดันครั้ง1	ความดันครั้ง2	สรุปความดัน	sumBP	family1	family2	p1	p2	sumB	ResultBsummary						
2	21.24	3	84	2	100	73	100	70	100	2	1	9	9	2	2	85	3	2		
3	21.71	4	91	2	102	89	124	79	128	84	2	9	9	2	2	65	2	2		
4	22.35	2	77	1	122	81	120	80	121	81	2	9	9	2	2	92	3	2		
5	17.58	1	66	1	100	77	97	69	90	73	1	9	9	2	2	87	3	1		
6	22.22	2	72	1	126	73	120	70	123	72	2	9	9	2	2	87	3	1		
7	17.78	1	61	1	126	73	120	70	123	72	2	8	8	2	2	77	2	1		
8	25.12	4	94	2	138	75	127	78	133	77	2	1+2	9	2	2	78	2	2		
9	21.61	3	76	1	110	70	110	70	110	70	1	1	1+2	2	2	74	2	2		
10	25.03	4	94	2	110	80	110	80	110	80	2	9	9	2	2	71	2	2		
11	20.76	2	71	1	110	70	110	70	110	70	1	9	9	2	2	90	3	1		
12	22.15	2	79	1	114	81	110	80	112	81	2	8	8	2	2	82	3	1		
13	23.73	4	102	2	129	77	120	70	125	74	2	9	9	2	2	78	2	2		
14	22.62	4	90	2	127	78	120	70	124	74	2	9	9	1	1	91	3	3		
15	21.68	3	90	2	122	62	120	70	121	66	2	8	8	2	1	93	3	3		
16	23.65	4	97	2	135	70	131	80	133	75	2	9	9	2	2	70	2	2		
17	19.53	2	74	1	112	70	117	65	115	68	1	9	9	2	2	66	2	1		
18	15.82	1	72	1	108	63	105	63	107	63	1	9	9	2	2	76	2	2		
19	17.58	1	64	1	94	60	88	59	81	60	1	2+3	9	2	2	82	3	2		
20	17.07	1	66	1	101	55	103	57	102	56	1	2+3	9	2	2	82	3	2		
21	15.38	1	65	1	111	61	103	69	107	65	1	2+3	9	2	2	74	2	2		
22	18.20	1	72	1	99	56	101	55	100	56	1	2+4	10	2	2	76	2	3		
23	37.02	5	107	2	113	72	116	65	115	69	1	9	9	1	2	58	1	3		
24	16.21	2	66	1	107	72	102	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 โดยใช้คำสั่งเรียงลำดับจากมากไปน้อยในแต่ละแอทริบิวต์เพื่อให้เห็นค่าผิดปกติได้ง่าย ไม่พบการขาดหายของข้อมูลแต่พบปัญหาข้อมูลมีค่าผิดพลาด โดยพบว่ามีอยู่ 3 แอทริบิวต์ที่มีค่าผิดพลาด คือ แอทริบิวต์การตรวจด้านมด้วยตนเอง ผลตรวจด้านมด้วยตนเอง 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	test 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
3	1	2	1	2	1	1	2	1	5	2	6	5	3	5	5	3	5	3	5	3	5	3
4	2	2	1	2	2	2	1+2+3	1	4	1	3	1	1	5	5	5	5	5	5	5	3	5
5	2	2	4	1	1	1	1+4	2	5	5	6	3	5	3	5	3	5	3	3	3	5	3
6	2	2	4	2	2	5	3	2	6	3	7	3	6	3	5	5	3	3	3	5	5	5
7	2	2	4	2	2	2	3	2	6	3	7	3	6	3	5	5	3	3	3	3	3	5
8	2	2	4	2	2	2	3	2	6	3	7	3	6	3	5	5	3	3	3	3	3	5
9	2	2	4	2	2	2	6	2	6	3	7	3	6	3	3	5	3	3	3	3	3	5
10	2	2	4	1	1	1	1	3	5	3	7	3	6	3	3	3	3	3	3	3	3	5
11	2	2	4	2	2	2	6	2	5	3	7	3	6	3	5	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	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	5

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

ข้อมูลที่ใช้ในการศึกษาทั้งหมดสามารถวิเคราะห์แยกตามแต่ละเอทริบิวต์ได้ดังรายละเอียดในตารางที่ 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 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการค้ำน้ำบ่อยและมาก

ลำดับ	ประวัติอาการค้ำน้ำบ่อยและมาก	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 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการเป็นแผลที่ริมฝีปากบ่นและหายยาก

ลำดับ	อาการเป็นแผลที่ริมฝีปากบ่นและหายยาก	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 ผลการวิเคราะห์ค่าของข้อมูลประวัติอาการตาพร่ามัวต้องเปลี่ยนแว่นบ่น

ลำดับ	ประวัติอาการตาพร่ามัวต้องเปลี่ยนแว่นบ่น	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 ผลการวิเคราะห์ค่าของข้อมูลการกินอาหารรสจัด

ลำดับ	การกินอาหารรสจัด	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	หลีกเลี่ยงเครื่องดื่มที่มีแอลกอฮอล์	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	การทำรอบเอว	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 และไม่เป็นโรคที่ 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 จึงต้องมีเครื่องมือมาช่วยให้ออกมาเหล่านี้นเรียกว่าตัวเสียใหม่ในพื้นที่ ที่เรียกว่า พื้นที่หลายมิติ (Higher Dimensional Space) และใช้หลักการการแบ่งข้อมูลโดยเส้นตรงและเส้นขอบมาแบ่งข้อมูลอีกครั้งหนึ่ง (ภัทร์พงศ์ 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 ถ้าค่าที่ออกมาเกิดความคลาดเคลื่อนจะนำไปสู่การปรับค่าหรือน้ำหนัก (weight) ของค่าที่ใส่ไว้ให้แต่ละ Input

ในการศึกษาครั้งนี้ต้องการนำโมเดลที่ได้ไปพัฒนาระบบคัดกรองสุขภาพประชาชนแบบอัตโนมัติ ดังนั้นเทคนิคการจำแนกกลุ่มที่นำมาใช้ในการศึกษาครั้งนี้จึงเป็นเทคนิคที่ให้ผลลัพธ์เป็นกฎเพื่อให้สามารถทำความเข้าใจได้ง่ายและสามารถนำไปใช้พัฒนาระบบได้ เทคนิคที่ศึกษาแบบอัลกอริทึมในกลุ่ม Decision Tree ได้แก่

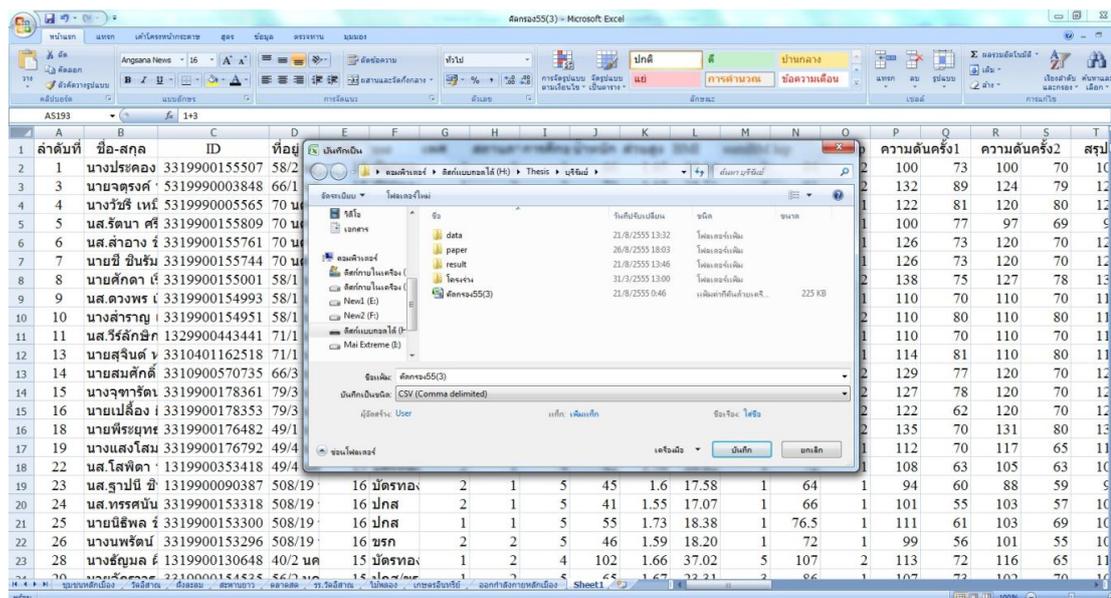
1. โมเดลการจำแนกกลุ่มแบบ Decision Tree: C4.5 (Kantardzic M., 2003) หรือในโปรแกรม Weka ใช้ชื่อว่า J48 เป็นการจัดกลุ่มโดยผลลัพธ์ที่ได้อยู่ในรูปของต้นไม้การตัดสินใจที่มีกิ่งต้นไม้จำนวนมาก โดยค่าของข้อมูลของแอทริบิวต์ที่มีค่าต่อเนื่องจะใช้เครื่องหมาย “<”, “<=”, “=”, “>” ในการพิจารณา ส่วนข้อมูลที่เป็นค่าไม่ต่อเนื่องจะใช้เครื่องหมาย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

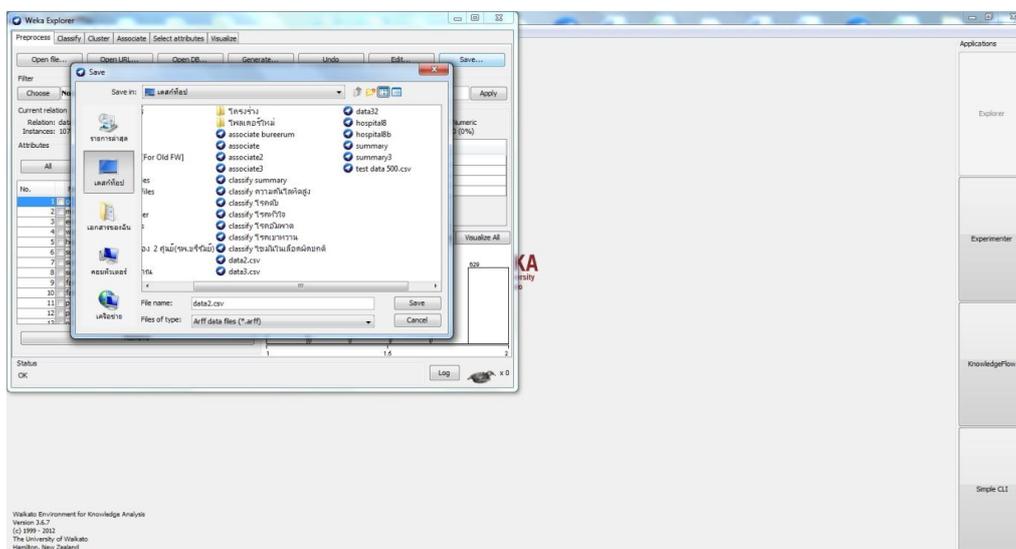
การศึกษาครั้งนี้ทำการทดสอบโมเดลโดยใช้โปรแกรมประยุกต์ Weka เวอร์ชัน 3.6.7 ข้อมูลที่ใช้ในการทดสอบมีจำนวน 1,071 เรคอร์ด ในการทดสอบได้ใช้การทดสอบแบบ 10 folds cross-validation ซึ่งเป็นการแบ่งกลุ่มข้อมูลออกเป็น 10 กลุ่ม ในแต่ละรอบจะนำกลุ่มข้อมูลจำนวน 9 กลุ่มเป็นกลุ่มศึกษา (Training Set) และกลุ่มข้อมูลที่เหลือเป็นกลุ่มทดสอบ (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

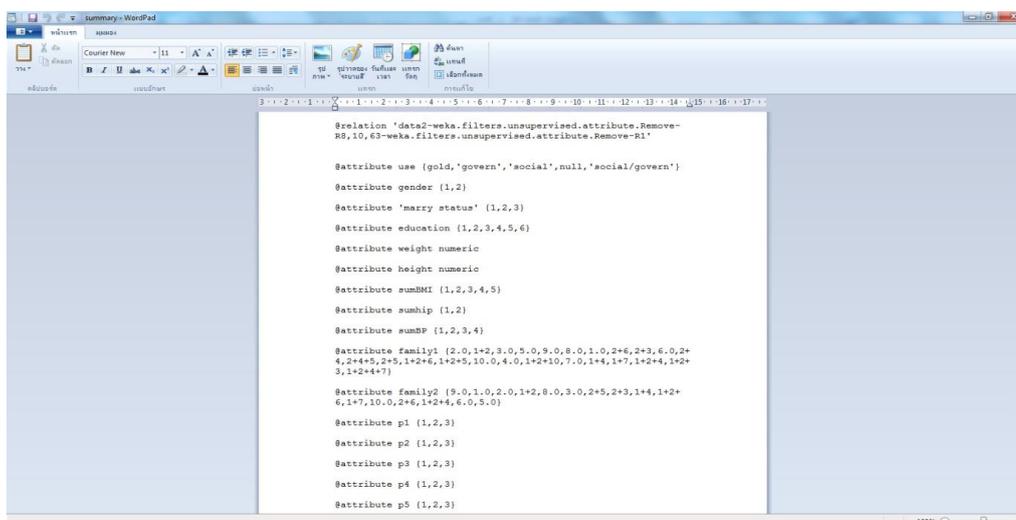


ภาพที่ 3.7 การบันทึกข้อมูลในรูปแบบนามสกุล .csv

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

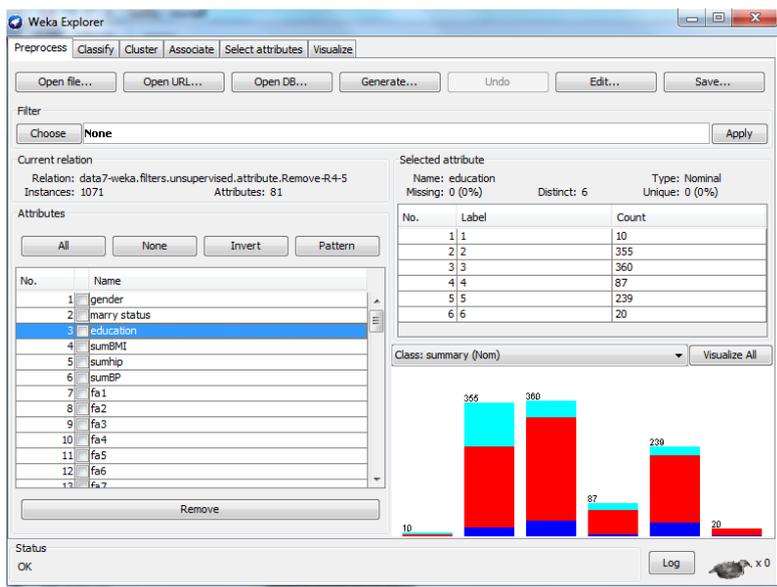


ภาพที่ 3.8 การแปลงข้อมูลให้เป็นรูปแบบนามสกุล .arff

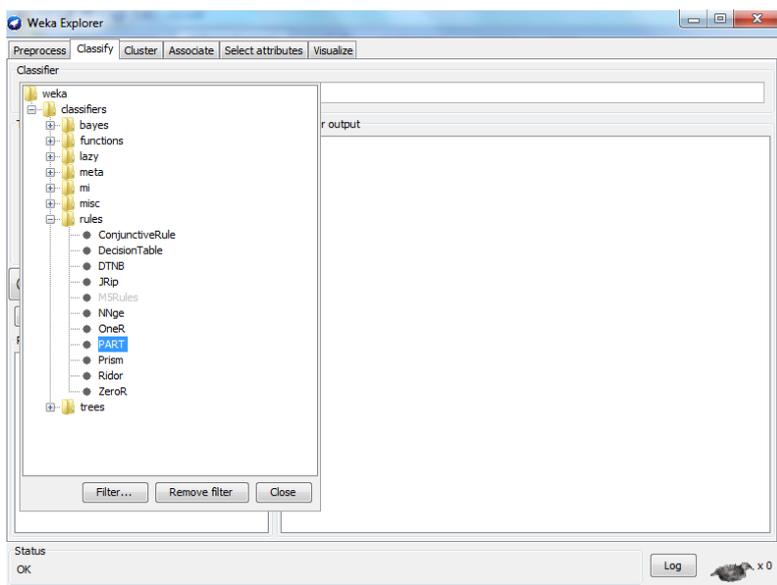


ภาพที่ 3.9 การแจกแจงข้อมูลในแต่ละแอทริบิวต์ใน โปรแกรม WordPad

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

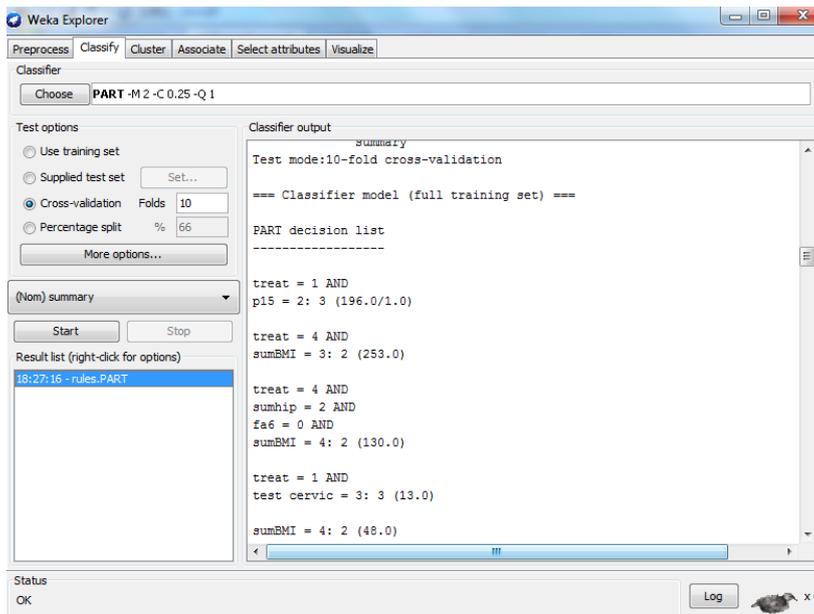


ภาพที่ 3.10 หน้าจอการวิเคราะห์ข้อมูล

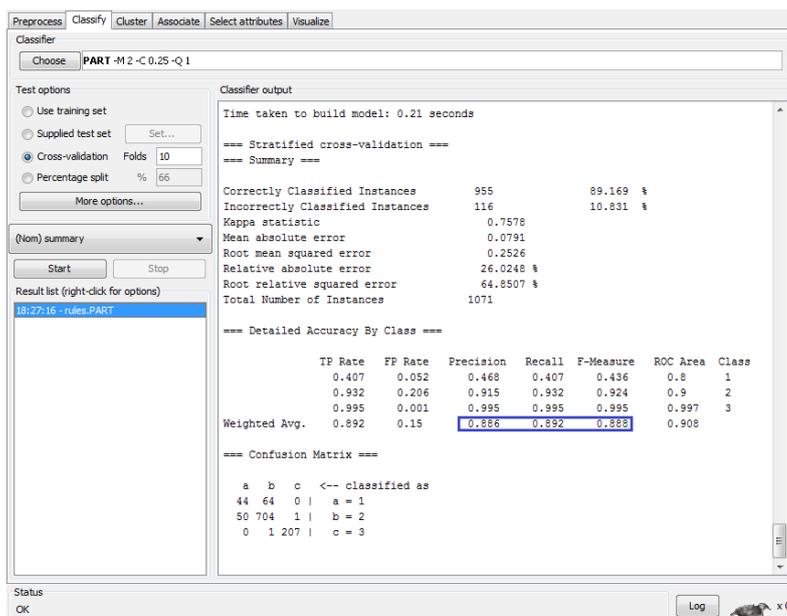


ภาพที่ 3.11 การเลือกเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูล

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



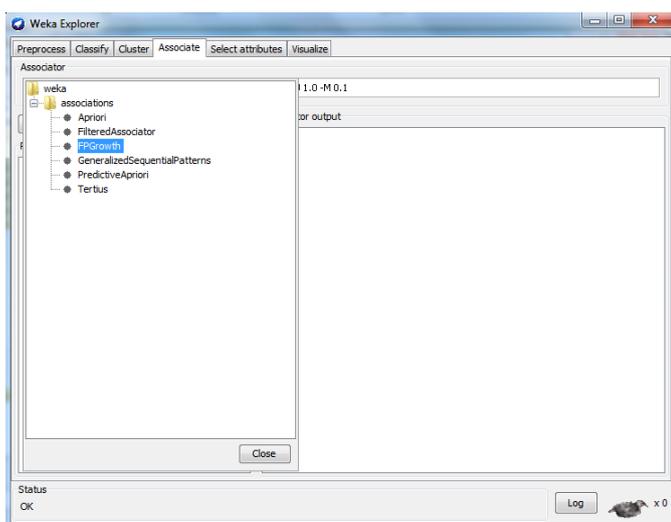
ภาพที่ 3.12 หน้าจอแสดงผลการจำแนกกลุ่มข้อมูลด้วยเทคนิค Decision Rule: Partial Rules



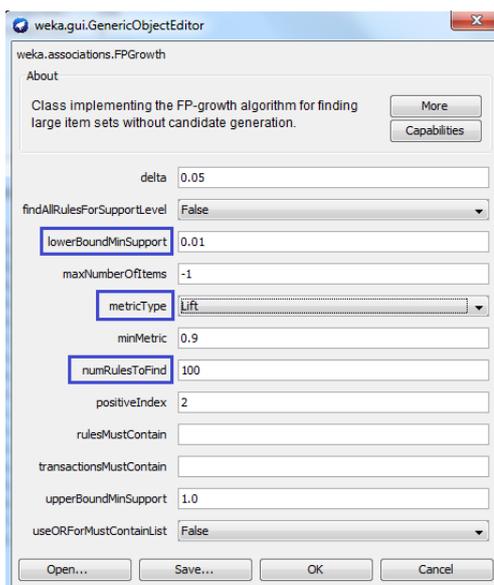
ภาพที่ 3.13 หน้าจอแสดงผลค่าความถูกต้องของโมเดล

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

เมื่อเลือกอัลกอริทึมที่จะใช้ในการหาความสัมพันธ์แล้ว 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 'Associator output' window is open, displaying 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