บทที่ 3

วิธีการดำเนินงานโครงงาน

โครงงานเรื่องการวิเคราะห์ข้อมูลการใช้ไฟฟ้าในบ้านที่อยู่อาศัย ในปี 2564-2565 เพื่อใช้ในการเผยแพร่ข้อมูลบนเว็บไซต์ ในบทนี้จะเป็นการวิเคราะห์ข้อมูลด้วย เทคนิคทางดาต้ามายนิ่ง ซึ่งมีกระบวนการวิเคราะห์ที่สำคัญหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้ว จะเป็นการออกแบบเว็บไซต์ และออกแบบรูปแบบการแสดงผลและบทสรุปจากวิธีการดำเนินงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

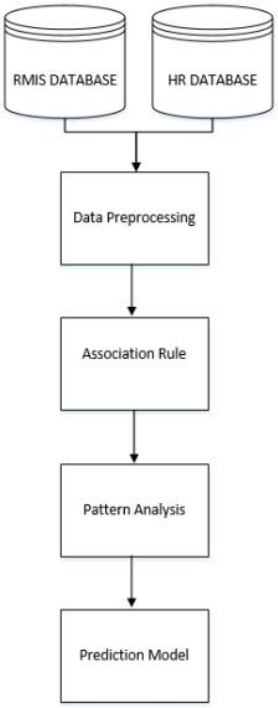
3.2 แผนภาพบริบท (Context Diagram)

3.3 แผนภาพกระแสข้อมูล Data Flow Diagram

3.4 ความสัมพันธ์ของข้อมูล (ER-Diagram)

3.5 การออกแบบเว็บไซต์

3.6 บทสรุป



ภาพที่ 38 บทที่ 3 กรอบแนวคิดของโครงงาน

3.1 การวิเคราะห์ข้อมูลด้วย CRISP-DM

ภาพที่ 39 บทที่ 3 กระบวนการวิเคราะห์ข้อมูล ด้วย CRISP-DM

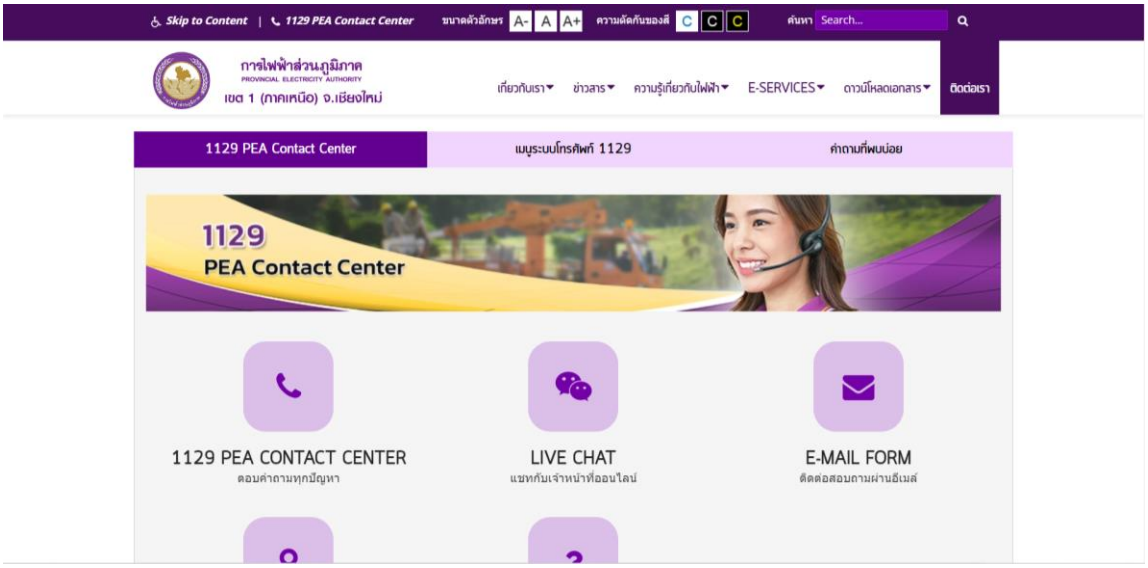
กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM หรือ Cross Industry Standard Process for Data Mining พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือของ 3 บริษัทคือ Daimler Chrysler, SPSS และ NCR ที่มีการพัฒนาเป็น Workflow มาตรฐานสำหรับการทำเหมืองข้อมูล ประกอบด้วย 6 ขั้นตอนหลัก ดังนี้

3.1.1 รู้จักและเข้าใจในธุรกิจ (Business understanding) เป็นขั้นตอนแรกของกระบวนการที่มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางธุรกิจโดยรวม

ผู้วิเคราะห์ข้อมูลทำความเข้าใจกับปัญหาให้อยู่ในรูปของการวิเคราะห์ข้อมูลทางดาต้าไมน์นิ่งโดยการวิเคราะห์ข้อมูล การใช้ไฟฟ้าในบ้านที่อยู่อาศัย ในปี 2564-2565 เพื่อใช้ในการเผยแพร่ข้อมูลบนเว็บไซต์ ซึ่งเป็นข้อมูลที่มีจำนวนมากและทำให้ซับซ้อน

3.1.2 จัดเก็บและรวบรวมข้อมูลให้ครบ (Data Understanding) ขั้นตอนการจัดเก็บ และรวบรวมข้อมูล ตลอดจนการพิจารณาตรวจสอบความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้สอดรับกับวัตถุประสงค์ที่กำหนดไว้

ผู้วิเคราะห์ข้อมูลได้ทำการเก็บรวบรวมข้อมูลจากเว็บไซต์ https://www.pea.co.th/จากนั้นจะทำการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้ เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าข้อมูลการใช้ไฟฟ้า จะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์



ภาพที่ 40 บทที่ 3 เว็บไซต์ https://www.pea.co.th ที่ให้ข้อมูลการใช้ไฟฟ้า

ซึ่งข้อมูลการใช้ไฟฟ้า มีจำนวน 3147 รายการ ประกอบด้วย 9 แอตทริบิวท์ ข้อมูลหลัก ๆ จะประกอบด้วย ปี จำนวนหน่วย จังหวัด ขนาดมิเตอร์ไฟฟ้า ประเภท เป็นต้น

	id	ปี	เดือน	อำเภอ	จังหวัด	ประเภท	จำนวนวัดต์	หน่วย	ขนาดมิเตอร์
2	1	2564	มกราคม	-	เชียงใหม่	บ้าน	1023	กิโลวัตต์/ชั่วโมง	30(100)A
3	2	2564	กุมภาพันธ์	-	เชียงใหม่	บ้าน	1030	กิโลวัตต์/ชั่วโมง	30(100)A
4	3	2564	มีนาคม	-	ลำปาง	บ้าน	1178	กิโลวัตต์/ชั่วโมง	30(100)A
5	4	2564	เมษายน	-	น่าน	บ้าน	1320	กิโลวัตต์/ชั่วโมง	30(100)A
6	5	2564	พฤษภาคม	-	น่าน	บ้าน	1301	กิโลวัตต์/ชั่วโมง	30(100)A
7	6	2564	มิถุนายน	-	น่าน	บ้าน	1236	กิโลวัตต์/ชั่วโมง	30(100)A
8	7	2564	กรกฎาคม	-	เชียงใหม่	บ้าน	1268	กิโลวัตต์/ชั่วโมง	30(100)A
9	8	2564	สิงหาคม	-	เชียงใหม่	บ้าน	1241	กิโลวัตต์/ชั่วโมง	30(100)A
10	9	2564	กันยายน	-	เชียงใหม่	บ้าน	1177	กิโลวัตต์/ชั่วโมง	30(100)A
11	10	2564	ตุลาคม	-	เชียงใหม่	บ้าน	1221	กิโลวัตต์/ชั่วโมง	30(100)A
12	11	2564	พฤศจิกายน	-	เชียงใหม่	บ้าน	1175	กิโลวัตต์/ชั่วโมง	30(100)A
13	12	2564	ธันวาคม	-	แพร่	บ้าน	1166	กิโลวัตต์/ชั่วโมง	30(100)A
14	13	2564	มกราคม	-	แพร่	บ้าน/กิจการ	448	กิโลวัตต์/ชั่วโมง	30(100)A
15	14	2564	กุมภาพันธ์	-	แพร่	บ้าน/กิจการ	453	กิโลวัตต์/ชั่วโมง	30(100)A
16	15	2564	มีนาคม	-	แพร่	บ้าน/กิจการ	518	กิโลวัตต์/ชั่วโมง	30(100)A
17	16	2564	เมษายน	-	แพร่	บ้าน/กิจการ	553	กิโลวัตต์/ชั่วโมง	30(100)A
18	17	2564	พฤษภาคม	-	เชียงใหม่	บ้าน/กิจการ	542	กิโลวัตต์/ชั่วโมง	30(100)A
19	18	2564	มิถุนายน	-	เชียงใหม่	บ้าน/กิจการ	528	กิโลวัตต์/ชั่วโมง	30(100)A
20	19	2564	กรกฎาคม	-	เชียงใหม่	บ้าน/กิจการ	553	กิโลวัตต์/ชั่วโมง	30(100)A
21	20	2564	สิงหาคม	-	เชียงใหม่	บ้าน/กิจการ	539	กิโลวัตต์/ชั่วโมง	30(100)A
22	21	2564	กันยายน	-	เชียงใหม่	บ้าน/กิจการ	516	กิโลวัตต์/ชั่วโมง	30(100)A
23	22	2564	ตุลาคม	-	เชียงใหม่	บ้าน/กิจการ	536	กิโลวัตต์/ชั่วโมง	30(100)A
24	23	2564	พฤศจิกายน	-	เชียงใหม่	บ้าน/กิจการ	515	กิโลวัตต์/ชั่วโมง	30(100)A
25	24	2564	ธันวาคม	-	น่าน	บ้าน/กิจการ	529	กิโลวัตต์/ชั่วโมง	30(100)A
26	25	2564	มกราคม	-	เชียงใหม่	อุตสาหกรรมขนาด	2854	กิโลวัตต์/ชั่วโมง	30(100)A
27	26	2564	กุมภาพันธ์	-	เชียงใหม่	อุตสาหกรรมขนาด	2798	กิโลวัตต์/ชั่วโมง	30(100)A
28	27	2564	มีนาคม	-	เชียงใหม่	อุตสาหกรรมขนาด	3219	กิโลวัตต์/ชั่วโมง	30(100)A
29	28	2564	เมษายน	-	เชียงใหม่	อุตสาหกรรมขนาด	3080	กิโลวัตต์/ชั่วโมง	30(100)A
30	29	2564	พฤษภาคม	-	อุตรดิตถ์	อุตสาหกรรมขนาด	3301	กิโลวัตต์/ชั่วโมง	30(100)A

ภาพที่ 41 บทที่ 3 ข้อมูลดิบการใช้ไฟฟ้า

3.1.3 เตรียมข้อมูลให้พร้อมใช้ (Data preparation) ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก โดยมีขั้นตอนดังนี้

3.1.3.1 ทำการคัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลที่เหมาะสมเพื่อนำมาใช้ในการวิเคราะห์ข้อมูล

ผู้วิเคราะห์ข้อมูลทำการคัดเลือกข้อมูลโดยการทำ Data Cleaning ข้อมูลรายงานการใช้ไฟฟ้า โดยแยกข้อมูลออกและตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะข้อมูลที่จำเป็นในการวิเคราะห์ในภาพรวม จำนวน 4 แอตทริบิวท์ ได้แก่ ปี จังหวัด ประเภท ขนาดมิเตอร์ ซึ่งเป็นข้อมูลที่จำเป็นในการนำไปวิเคราะห์ข้อมูล

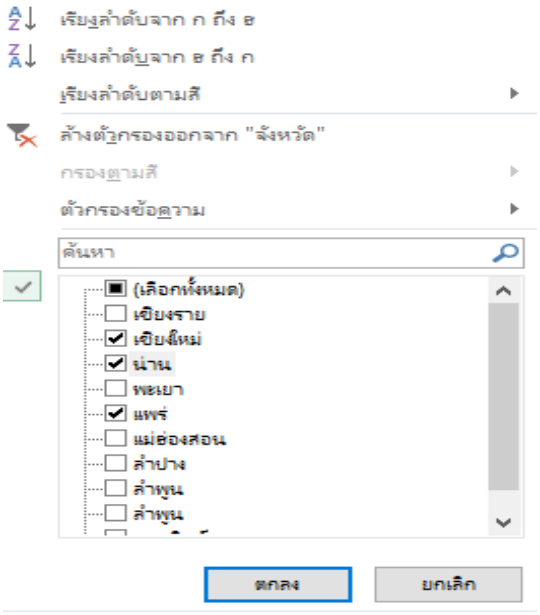
	meter	type	level
2	NO	house	M
3	NO	small industry	M
4	NO	agriculture	M
5	NO	agriculture	M
6	NO	house	M
7	NO	house	M
8	NO	small industry	M
9	NO	small industry	M
10	NO	agriculture	M
11	NO	agriculture	M
12	NO	agriculture	M
13	NO	agriculture	M
14	NO	small industry	S
15	NO	small industry	S
16	NO	house	S
17	NO	agriculture	S
18	NO	house	S
19	NO	small industry	S
20	NO	house	S
21	NO	small industry	S
22	NO	small industry	S
23	NO	house	S
24	NO	agriculture	S
25	NO	small industry	S
26	NO	agriculture	L
27	NO	small industry	L
28	NO	agriculture	XL
29	NO	except	XL
30	NO	small industry	XL

ภาพที่ 42 บทที่ 3 ข้อมูลรายงานการใช้ไฟฟ้าที่ทำการคัดเลือกข้อมูลแล้ว

3.1.3.2 ทำการกลั่นกรองข้อมูล (Data Cleaning) คือการทำความสะอาดข้อมูลเป็นกระบวนการตรวจสอบและการแก้ไข (หรือลบ) รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูลตารางหรือฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล ทางผู้วิเคราะห์ข้อมูลได้ดำเนินการดังนี้

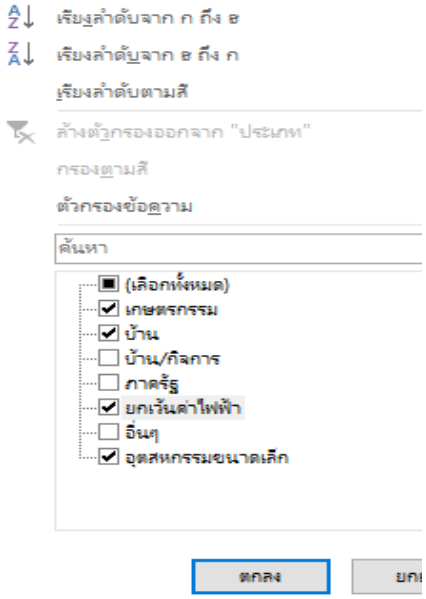
1) ข้อมูลรายงานการใช้ไฟฟ้า ผู้วิเคราะห์ข้อมูลได้ทำการแก้ไขและลบข้อมูล ซึ่งผู้วิเคราะห์ข้อมูลพบว่า ข้อมูลทั้งหมดนั้นมีจำนวนที่เยอะจึงเลือกส่วนหัวข้อย่อยในแต่ละแอตทริบิวท์มาเท่านั้น ดังนั้นผู้วิเคราะห์ข้อมูลได้ดำเนินการดังนี้

– จังหวัด (Province) มี 9 หัวข้อย่อย ผู้วิเคราะห์ พบว่าควรตัด หัวข้อย่อยที่มีรายการน้อยจึงทำการลบทิ้ง และมีหัวข้อย่อย 3 หัวข้อย่อยที่มีมาก คือ เชียงใหม่, น่าน, แพร่ ดังนี้



ภาพที่ 43 บทที่ 3 ทำการกลั่นกรองข้อมูล ปี เพื่อนำไปวิเคราะห์

– ประเภท (Type) มี 7 หัวข้อย่อย ผู้วิเคราะห์ พบว่าควรตัด หัวข้อย่อยที่มีรายการน้อย จึงทำการลบทิ้ง และมีหัวข้อย่อย 4 หัวข้อย่อยที่มีมาก คือ เกษตรกรรม, บ้าน, ยกเว้นค่าไฟฟ้า, อุคสหกรรมขนาดเล็ก ดังนี้



ภาพที่ 44 บทที่ 3 ทำการกลั่นกรองข้อมูล ประเภท เพื่อนำไปวิเคราะห์

3.1.3.3 แปลงรูปแบบของข้อมูล (Data Transformation) เป็นขั้นตอนการแปลงข้อมูลในรูปแบบตาราฐานข้อมูลให้อยู่ในรูป item set เพื่อใช้สำหรับการนำมาวิเคราะห์ด้วยวิธีการของ data mining ผู้วิเคราะห์ข้อมูลได้ดำเนินการกับข้อมูลการใช้ไฟฟ้า ดังนี้

1) ผู้วิเคราะห์ข้อมูลทำการแปลงรูปแบบข้อมูลด้วยการรวมกลุ่มของข้อมูลในแอตทริบิวต์ ขนาดมิเตอร์ (Meter) ได้แก่ ขนาดมิเตอร์ 15(45)A) เนื่องจากแอตทริบิวต์ ขนาดมิเตอร์ 15(45) ซึ่งทางผู้วิเคราะห์ข้อมูลจะขอรวมกลุ่มของขนาดมิเตอร์ ดังกล่าว ให้เหลือเพียง “15(45)A” อย่างเดียว แปลงข้อมูลให้อยู่ในรูป item set โดย “15(45)A” แทนคำว่า “yes”

2) ผู้วิเคราะห์ข้อมูลทำการแปลงรูปแบบข้อมูลด้วยการรวมกลุ่มของข้อมูลในแอตทริบิวต์ ขนาดมิเตอร์ (Meter) ขนาดมิเตอร์ 30(100)A แปลงข้อมูลให้อยู่ในรูป item set โดย “30(100)A” แทนคำว่า “no”

3.1.4 สร้างแบบจำลอง (Modeling) ขั้นตอนการสร้างตัวแบบทางคณิตศาสตร์ และสถิติเพื่อการวิเคราะห์ข้อมูล โดยสามารถใช้เทคนิควิธีการต่าง ๆ อาทิ การจำแนก (Classification) การแบ่งกลุ่ม (Clustering) และการสร้างความสัมพันธ์ (Association rule)

ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์นิ่ง แบบการจำแนกประเภทข้อมูล (Classification) โดยการใช้โมเดลการตัดสินใจแบบต้นไม้ (Decision Tree) ซึ่งในขั้นตอนนี้จะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือกดังนี้

	meter	type	level
1	meter	type	level
2	NO	house	M
3	NO	small industry	M
4	NO	agriculture	M
5	NO	agriculture	M
6	NO	house	M
7	NO	house	M
8	NO	small industry	M
9	NO	small industry	M
10	NO	agriculture	M
11	NO	agriculture	M
12	NO	agriculture	M
13	NO	agriculture	M
14	NO	small industry	S
15	NO	small industry	S
16	NO	house	S
17	NO	agriculture	S
18	NO	house	S
19	NO	small industry	S
20	NO	house	S
21	NO	small industry	S
22	NO	small industry	S
23	NO	house	S
24	NO	agriculture	S
25	NO	small industry	S
26	NO	agriculture	L
27	NO	small industry	L
28	NO	agriculture	XL
29	NO	except	XL
30	NO	small industry	XL

ภาพที่ 45 บทที่ 3 ข้อมูลแอตทริบิวต์ที่คัดเลือกมาวิเคราะห์ข้อมูลทั้งหมด

จากรูปภาพที่ ประกอบด้วย 3 แอตทริบิวต์ คือ

– ประเภท (type) ประกอบด้วย 4 ค่า คือ เกษตรกรรม, บ้าน, ยกเว้นค่าไฟฟ้า, อุตสหกรรมขนาดเล็ก

– ขนาดมิเตอร์ (meter) ประกอบด้วย 2 ค่า คือ ได้แก่ ขนาดมิเตอร์ 30(100)A (no) , ขนาดมิเตอร์ 15(45)A) (yes)

–ประเภทการใช้ไฟฟ้า ประกอบด้วย S การใช้ไฟฟ้าตั้งแต่ 1–1000 M การใช้ไฟฟ้าตั้งแต่ 1000–2000 L ใช้ไฟฟ้าตั้งแต่ 2000–3000 XL ใช้ไฟฟ้าตั้งแต่ 4000ขึ้นไป

การสร้างโมเดล Decision Tree จะทำการคัดเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของ Tree (root node) หลังจากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของแอตทริบิวต์นี้จะใช้ตัววัด ที่เรียกว่า Information Gain (IG) ค่านี้คำนวณได้จากสมการการคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุด

IG (parent, child) = entropy(parent) = – [p(c1) × entropy(c1) + p(c2) × entropy(c2) + …]
โดยที่ entropy(c1) = -p(c1) log p(c1) และ p(c2) คือ ค่าความน่าจะเป็นของ c1

การคำนวณค่าแต่ละแอตทริบิวต์เทียบกับคลาสเพื่อหาแอตทริบิวต์ที่มีค่า IG มากที่สุดมาเป็น Root ของ Decision tree กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็น meter (ขนาดมิเตอร์) yes และ no ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ year จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)
= – [0.6167 × log2(0.6167) + 0.3833 × log2(0.3833)]
= – [0.6167 × –0.6974 + 0.3833 × –1.3835]
= – [–0.4321 + –0.5303]
= 0.9623

entropy (ผล = 2564) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)
= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]
= – [0.5932 × –0.7534 + 0.4068 × –1.2976]
= – [–0.4469 + –0.5279]
= 0.9748

entropy (ผล = 2565) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

IG (parent, child) = entropy(parent) – [p (ผล = 2564) * entropy (ผล = 2564) + p (ผล = 2565) * entropy (ผล = 2565)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315]

= 0.9603 – [0.6064 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

2) คำนวณค่า IG ของแอตทริบิวต์ province จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6167 × log2(0.6167) + 0.3833 × log2(0.3833)]

= – [0.6167 × –0.6974 + 0.3833 × –1.3835]

= – [–0.4321 + –0.5303]

= 0.9623

entropy (ผล = Chiang Mai) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = Nan) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

entropy (ผล = Phrae) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6434 × log2(0.6434) + 0.3566 × log2(0.3566)]

= – [0.6434 × –0.6362 + 0.3566 × –1.4876]

= – [–0.4093 + –0.5305]

= 0.9398

IG (parent, child) = entropy(parent) – [p (ผล = Chiang Mai) * entropy (ผล = Chiang Mai) + p (ผล = Nan) * entropy (ผล = Nan) + p (ผล = Phrae) * entropy (ผล = Phrae)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315+ 0.1213 * 0.9398]

= 0.9603 – [0.6064 + 0.1114 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

3) คำนวณค่า IG ของแอตทริบิวต์ type จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6167 × log2(0.6167) + 0.3833 × log2(0.3833)]

= – [0.6167 × –0.6974 + 0.3833 × –1.3835]

= – [–0.4321 + –0.5303]

= 0.9623

entropy (ผล = house) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = small industry) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

entropy (ผล = agriculture) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6434 × log2(0.6434) + 0.3566 × log2(0.3566)]

= – [0.6434 × –0.6362 + 0.3566 × –1.4876]

= – [–0.4093 + –0.5305]

= 0.9398

entropy (ผล = except) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6434 × log2(0.6434) + 0.3566 × log2(0.3566)]

= – [0.6434 × –0.6362 + 0.3566 × –1.4876]

= – [–0.4093 + –0.5305]

= 0.9398

IG (parent, child) = entropy(parent) – [p (ผล = house) * entropy (ผล = house) + p (ผล = small industry) * entropy (ผล = small industry) + p (ผล = agriculture) * entropy (ผล = agriculture) + p (ผล = except) * entropy (ผล = except)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315 + 0.1213 * 0.9398 + 0.1667 * 0.8810]

= 0.9603 – [0.6064 + 0.1114 +0.1114 + 0.1469]

= 0.9603 – 0.9571

= 0.1172

จากการคำนวณค่า IG ของทุกแอตทริบิวต์พบว่าค่า IG ของแอตทริบิวต์ type มีค่ามากที่สุด (0.1172) ดังนั้นจึงเลือกแอตทริบิวต์ type ขึ้นมาเป็นโหนด root และจะต้องทำการแตกกิ่งจากโหนด root ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน และผู้วิเคราะห์ข้อมูลพบว่าการคำนวณแอตทริบิวต์ type (house) และ type (small industry) ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์ ขนาดมิเตอร์ที่ได้รับมาตรฐาน กับ มิเตอร์ขนาดเล็ก ดังนั้นผู้วิเคราะห์ข้อมูลจึงสร้างโหนดในระดับถัดไปของแอตทริบิวต์ type agriculture และ except

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 2 ต่อจากโหนด root เพื่อหาค่า IG ที่มากที่สุดของแอตทริบิวต์ type กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นขนาดมิเตอร์ที่ได้รับมาตรฐาน และ มิเตอร์ขนาดเล็ก ดังนี้

1) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ province ของ agriculture จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – $p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

= – [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= – [0.3333 × –1.5851 + 0.6667 × –0.5849]

= – [–0.5283 + –0.3899]

= 0.9183

entropy (ผล = Chiang Mai) = – $p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = Nan) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

entropy (ผล = Phrae) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6434 × log2(0.6434) + 0.3566 × log2(0.3566)]

= – [0.6434 × –0.6362 + 0.3566 × –1.4876]

= – [–0.4093 + –0.5305]

= 0.9398

IG (parent, child) = entropy(parent) – [p (ผล = Chiang Mai) * entropy (ผล = Chiang Mai) + p (ผล = Nan) * entropy (ผล = Nan) + p (ผล = Phrae) * entropy (ผล = Phrae)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315+ 0.1213 * 0.9398]

= 0.9603 – [0.6064 + 0.1114 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

1.2) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ province ของ small industry จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= – [0.3333 × –1.5851 + 0.6667 × –0.5849]

= – [–0.5283 + –0.3899]

= 0.9183

entropy (ผล = Chiang Mai) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = Nan) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

entropy (ผล = Phrae) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6434 × log2(0.6434) + 0.3566 × log2(0.3566)]

= – [0.6434 × –0.6362 + 0.3566 × –1.4876]

= – [–0.4093 + –0.5305]

= 0.9398

IG (parent, child) = entropy(parent) – [p (ผล = Chiang Mai) * entropy (ผล = Chiang Mai) + p (ผล = Nan) * entropy (ผล = Nan) + p (ผล = Phrae) * entropy (ผล = Phrae)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 *
0.9315+ 0.1213 * 0.9398]
= 0.9603 – [0.6064 + 0.1114 + 0.1114]
= 0.9603 – 0.9571
= 0.0029

1.3) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ province ของ house จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × log2p(y) + p(n) × log2p(n)
= – [0.3333 × log2(0.3333) + 0.6667 ×
log2(0.6667)]
= – [0.3333 × –1.5851 + 0.6667 × –0.5849]
= – [–0.5283 + –0.3899]
= 0.9183

entropy (ผล = Chiang Mai) = – p(y) × log2p(y) + p(n) × log2p(n)
= – [0.5932 × log2(0.5932) + 0.4068 ×
log2(0.4068)]
= – [0.5932 × –0.7534 + 0.4068 × –1.2976]
= – [–0.4469 + –0.5279]
= 0.9748

entropy (ผล = Nan) = – p(y) × log2p(y) + p(n) × log2p(n)
= – [0.6529 × log2(0.6529) + 0.3471 ×
log2(0.3471)]
= – [0.6529 × –0.6151 + 0.3471 × –1.5266]
= – [–0.4016 + –0.5299]
= 0.9315

entropy (ผล = Phrae) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6434 × log2(0.6434) + 0.3566 × log2(0.3566)]

= – [0.6434 × –0.6362 + 0.3566 × –1.4876]

= – [–0.4093 + –0.5305]

= 0.9398

IG (parent, child) = entropy(parent) – [p (ผล = Chiang Mai) * entropy (ผล = Chiang Mai) + p (ผล = Nan) * entropy (ผล = Nan) + p (ผล = Phrae) * entropy (ผล = Phrae)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315+ 0.1213 * 0.9398]

= 0.9603 – [0.6064 + 0.1114 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

1.4) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ province ของ except จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= – [0.3333 × –1.5851 + 0.6667 × –0.5849]

= – [–0.5283 + –0.3899]

= 0.9183

entropy (ผล = Chiang Mai) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = Nan) $= - p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

$= - [0.6529 \times \log2(0.6529) + 0.3471 \times \log2(0.3471)]$

$= - [0.6529 \times -0.6151 + 0.3471 \times -1.5266]$

$= - [-0.4016 + -0.5299]$

$= 0.9315$

entropy (ผล = Phrae) $= - p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

$= - [0.6434 \times \log2(0.6434) + 0.3566 \times \log2(0.3566)]$

$= - [0.6434 \times -0.6362 + 0.3566 \times -1.4876]$

$= - [-0.4093 + -0.5305]$

$= 0.9398$

IG (parent, child) = entropy(parent) – [p (ผล = Chiang Mai) * entropy (ผล = Chiang Mai) + p (ผล = Nan) * entropy (ผล = Nan) + p (ผล = Phrae) * entropy (ผล = Phrae)]

$= 0.9603 - [0.6221 * 0.9748 + 0.1196 * 0.9315 + 0.1213 * 0.9398]$

$= 0.9603 - [0.6064 + 0.1114 + 0.1114]$

$= 0.9603 - 0.9571$

$= 0.0029$

2) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ year ของ agriculture จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) $= - p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

$= - [0.3333 \times \log2(0.3333) + 0.6667 \times \log2(0.6667)]$

$= - [0.3333 \times -1.5851 + 0.6667 \times -0.5849]$

$= - [-0.5283 + -0.3899]$

= 0.9183

entropy (ผล = 2564) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = 2565) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

IG (parent, child) = entropy(parent) – [p (ผล = 2564) * entropy (ผล = 2564) + p (ผล = 2565) * entropy (ผล = 2565)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315]

= 0.9603 – [0.6064 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

2.1) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ year ของ small industry จากข้อมูล สามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= – [0.3333 × –1.5851 + 0.6667 × –0.5849]

= – [–0.5283 + –0.3899]

= 0.9183

entropy (ผล = 2564) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = 2565) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

IG (parent, child) = entropy(parent) – [p (ผล = 2564) * entropy (ผล = 2564) + p (ผล = 2565) * entropy (ผล = 2565)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315]

= 0.9603 – [0.6064 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

2.2) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ year ของ house จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × log2p(y) + p(n) × log2p(n)

= – [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= – [0.3333 × –1.5851 + 0.6667 × –0.5849]

= – [–0.5283 + –0.3899]

= 0.9183

entropy (ผล = 2564) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = 2565) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

IG (parent, child) = entropy(parent) – [p (ผล = 2564) * entropy (ผล = 2564) + p (ผล = 2565) * entropy (ผล = 2565)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315]

= 0.9603 – [0.6064 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

2.3) คำนวณค่า IG ของแอตทริบิวต์ Type และแอตทริบิวต์ year ของ except จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.3333 × log2(0.3333) + 0.6667 × log2(0.6667)]

= – [0.3333 × –1.5851 + 0.6667 × –0.5849]

= – [–0.5283 + –0.3899]

= 0.9183

entropy (ผล = 2564) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.5932 × log2(0.5932) + 0.4068 × log2(0.4068)]

= – [0.5932 × –0.7534 + 0.4068 × –1.2976]

= – [–0.4469 + –0.5279]

= 0.9748

entropy (ผล = 2565) = – p(y) × $\log_2$p(y) + p(n) × $\log_2$p(n)

= – [0.6529 × log2(0.6529) + 0.3471 × log2(0.3471)]

= – [0.6529 × –0.6151 + 0.3471 × –1.5266]

= – [–0.4016 + –0.5299]

= 0.9315

IG (parent, child) = entropy(parent) – [p (ผล = 2564) * entropy (ผล = 2564) + p (ผล = 2565) * entropy (ผล = 2565)]

= 0.9603 – [0.6221 * 0.9748 + 0.1196 * 0.9315]

= 0.9603 – [0.6064 + 0.1114]

= 0.9603 – 0.9571

= 0.0029

จากการคำนวณค่า IG ของแอตทริบิวต์ Type ต่อแอตทริบิวต์ province และแอทริบิวต์ year พบว่าค่า IG ของแอตทริบิวต์ Type (agriculture) ต่อแอตทริบิวต์ province มีค่ามากที่สุด (0.2516) และแอตทริบิวต์ Type (except) ต่อแอตทริบิวต์ year มีค่ามากรองลงมาเป็น (0.2040) ดังนั้นจึงเลือกแอตทริบิวต์ province และแอตทริบิวต์ year ขึ้นมาเป็นโหนดในระดับที่ 2 ต่อจาก

โหนด Root และผู้วิเคราะห์ข้อมูลพบว่าการคำนวณแอตทริบิวต์ Type (house, small industry) ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์ขนาดมิเตอร์ที่ได้รับมาตรฐาน และ มิเตอร์ขนาดเล็ก ดังนั้นผู้วิเคราะห์ข้อมูลจึงทำการแตกกิ่งจากโหนดในระดับที่ 2 ของแอตทริบิวต์ province, year ออกไปจนข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกัน

การคำนวณค่าแต่ละแอตทริบิวต์ในระดับที่ 3 กับจำนวนข้อมูลทั้งหมดโดยใช้ผลลัพธ์เป็นขนาดมิเตอร์ที่ได้รับมาตรฐาน และ มิเตอร์ขนาดเล็ก ดังนี้

1)คำนวณค่า IG ของแอตทริบิวต์ Type (agriculture) และแอตทริบิวต์ province (Chiang Mai) ไปแอตทริบิวต์ year (2564, 2565) จากข้อมูลสามารถคำนวณค่า IG ได้ดังนี้

entropy (parent) $= - p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

$= - [0.5000 \times \log2(0.5000) + 0.5000 \times \log2(0.5000)]$

$= - [0.5000 \times -1 + 0.5000 \times -1]$

$= - [-0.5000 + -0.5000]$

$= 1$

entropy (ผล = 2564) $= - p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

$= - [0.5932 \times \log2(0.5932) + 0.4068 \times \log2(0.4068)]$

$= - [0.5932 \times -0.7534 + 0.4068 \times -1.2976]$

$= - [-0.4469 + -0.5279]$

$= 1$

entropy (ผล = 2565) $= - p(y) \times \log_2 p(y) + p(n) \times \log_2 p(n)$

$= - [0.6529 \times \log2(0.6529) + 0.3471 \times \log2(0.3471)]$

$= - [0.6529 \times -0.6151 + 0.3471 \times -1.5266]$

$= - [-0.4016 + -0.5299]$

= 0

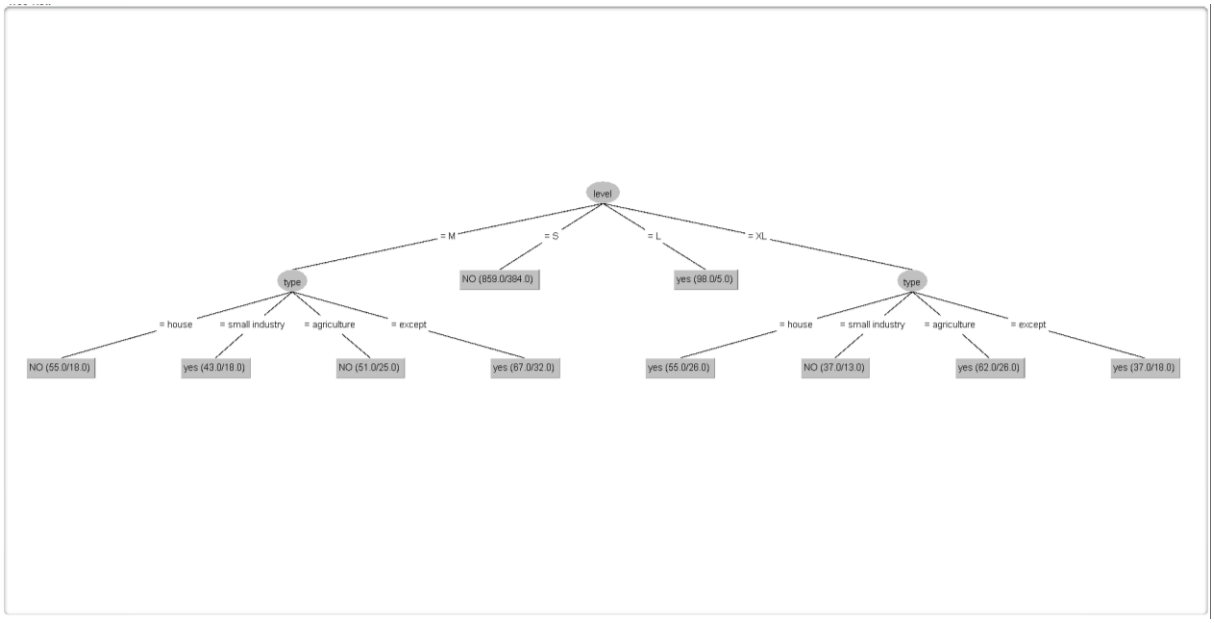
IG (parent, child) = entropy(parent) – [p (ผล = 2564) * entropy (ผล = 2564) + p (ผล = 2565) * entropy (ผล = 2565)] = 1 – [1 * 1 + 0 * 0]

= 1 – [1 + 0]

= 1 – 1

= 0

จากการคำนวณค่า IG ของแอตทริบิวต์ Type (agriculture) ต่อแอตทริบิวต์ province ไปแอตทริบิวต์ year พบว่าค่า IG ของแอตทริบิวต์ Type ต่อแอตทริบิวต์ province ไปแอตทริบิวต์ year มีค่ามากที่สุด (0.0796) ดังนั้นจึงเลือกแอตทริบิวต์ year ขึ้นมาเป็นโหนดในระดับที่ 3 ต่อจากโหนด Root ต่อจากโหนดระดับที่ 2 และผู้วิเคราะห์ข้อมูลพบว่าการคำนวณแอตทริบิวต์ year (2564, 2565) ไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด จึงสรุปข้อมูลได้เป็นผลลัพธ์ ขนาดมิเตอร์ที่ได้รับมาตรฐาน และ มิเตอร์ขนาดเล็ก และแอตทริบิวต์ Type (except) ต่อแอตทริบิวต์ year เป็นแอตทริบิวต์สุดท้าย พบว่าแอตทริบิวต์ province มีความสัมพันธ์กับแอตทริบิวต์ year มากที่สุด ซึ่งพบว่าข้อมูลในแต่ละโหนดมีคลาสคำตอบเดียวกันแล้ว คือ ขนาดมิเตอร์ที่ได้รับมาตรฐาน และ มิเตอร์ขนาดเล็ก ตามภาพที่ 44



ภาพที่ 46 บทที่ 3 เส้นแสดงความสัมพันธ์

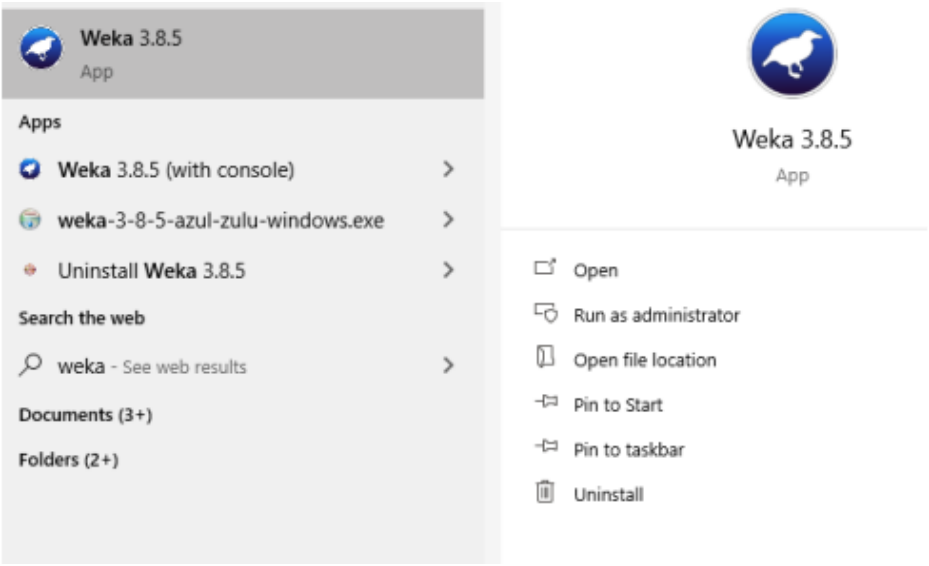
จากภาพที่ 45 โมเดลต้นไม้ตัดสินใจ จากการคำนวณด้วยมือนี้ ผู้วิเคราะห์ข้อมูลได้ผลลัพธ์ว่า โมเดลต้นไม้ตัดสินใจ Root node ที่คือ แอตทริบิวต์ Type และได้ interior node คือ แอตทริบิวต์ province และ leaf node คือ แอตทริบิวต์ year ซึ่งไม่สามารถสร้างกิ่งแต่ละโหนดต่อไปได้ เนื่องจากไม่มีความสัมพันธ์กับแอตทริบิวต์ใด ก็จะได้ผลลัพธ์ที่ แอตทริบิวต์ year 2564 เป็นขนาดมิเตอร์ที่ได้รับมาตรฐาน year 2565 เป็น มิเตอร์ขนาดเล็ก แอตทริบิวต์ province house เป็นขนาดมิเตอร์ที่ได้รับมาตรฐาน และ แอตทริบิวต์ province small industry เป็นมิเตอร์ขนาดเล็ก

3.1.5 การประเมินผล (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 3.1.4 ไปใช้งาน ด้วยการวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรมว่าถูกต้องหรือไม่

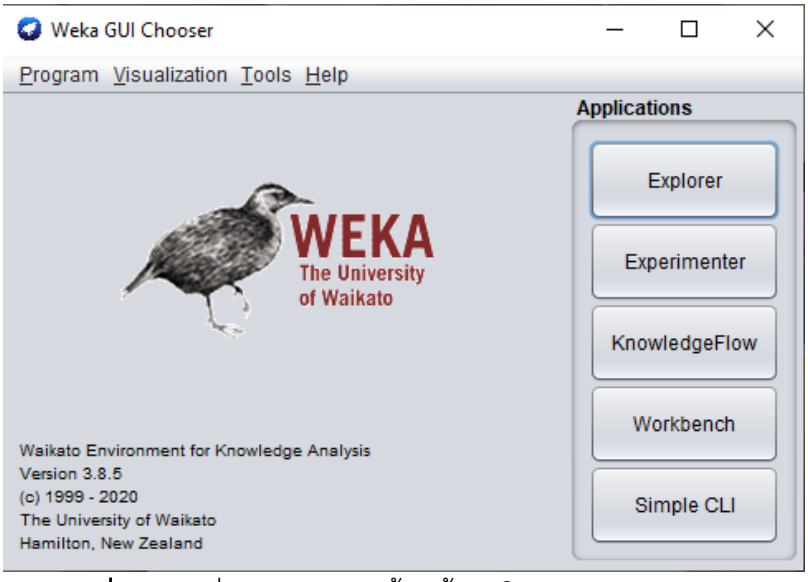
ผู้วิเคราะห์ข้อมูลได้ทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพที่ตรงกับความต้องการ ซึ่งการวัดประสิทธิภาพด้วยวิธี Self-Consistency Test เหมาะสำหรับใช้ในการทดสอบประสิทธิภาพเพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น และเมื่อนำข้อมูลมาทดสอบ (Testing data) กับโปรแกรมที่ผู้วิเคราะห์เลือกมาทดสอบกับข้อมูลที่ผ่านการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining จากการ

สร้างโมเดล Decision Tree จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka 3.8.5 ซึ่งมีขั้นตอนการทำงาน ดังนี้

ขั้นตอนที่ 1 เปิดโปรแกรม Weka 3.8.5 ขึ้นมา

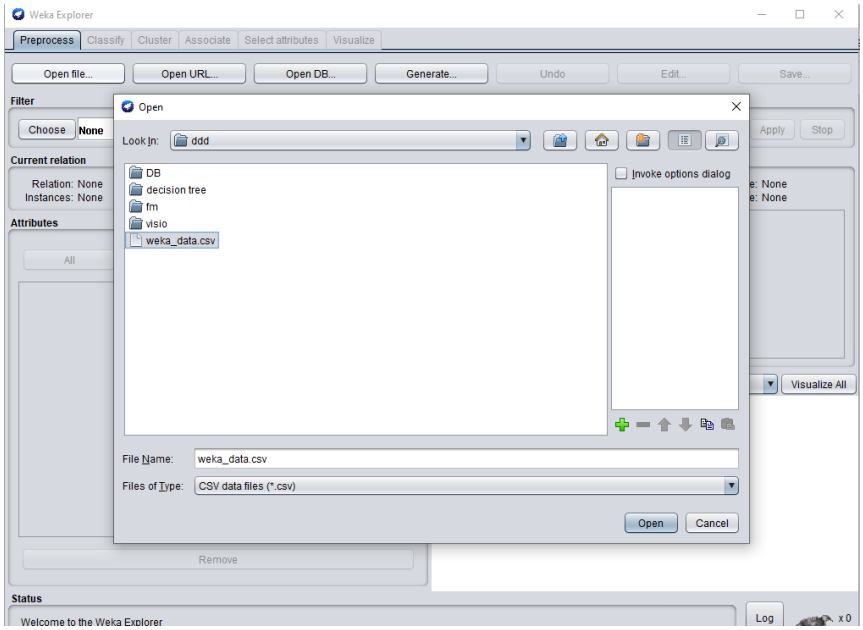


ภาพที่ 47 บทที่ 3 แสดงการเปิดโปรแกรม weka 3.8.5

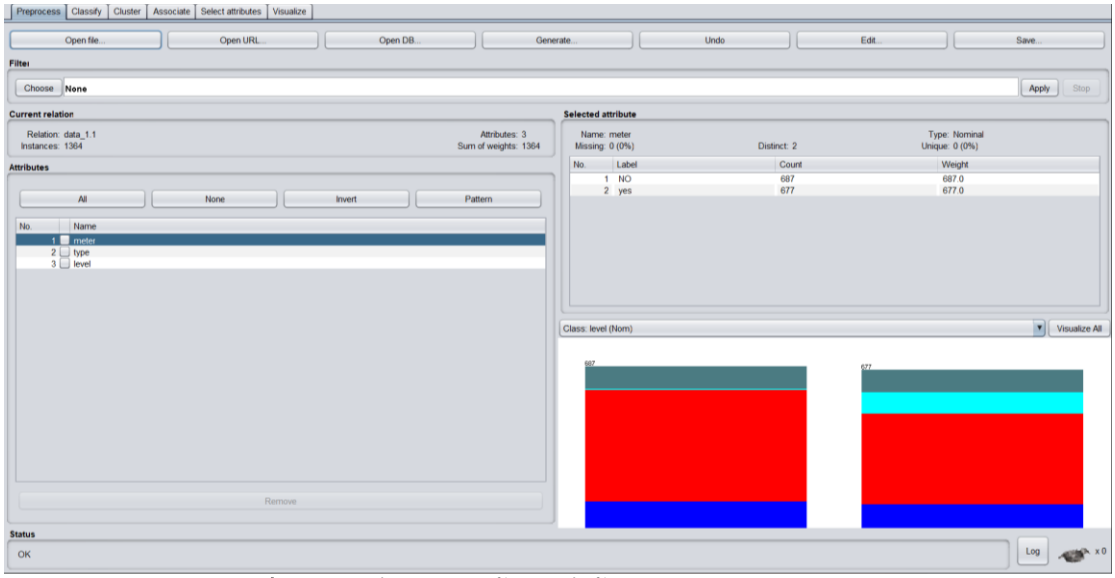


ภาพที่ 48 บทที่ 3 แสดงการเข้าหน้าจอโปรแกรม Weka 3.8.5

ขั้นตอนที่ 2 นำเข้าข้อมูลที่ได้จัดเตรียมไว้ โดยเลือกที่ Application>>Explorer>>Open file เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบตามภาพที่ 3.11 และหลังจากนั้นโปรแกรมแสดงหน้าจอข้อมูล ตามภาพที่ 3.12

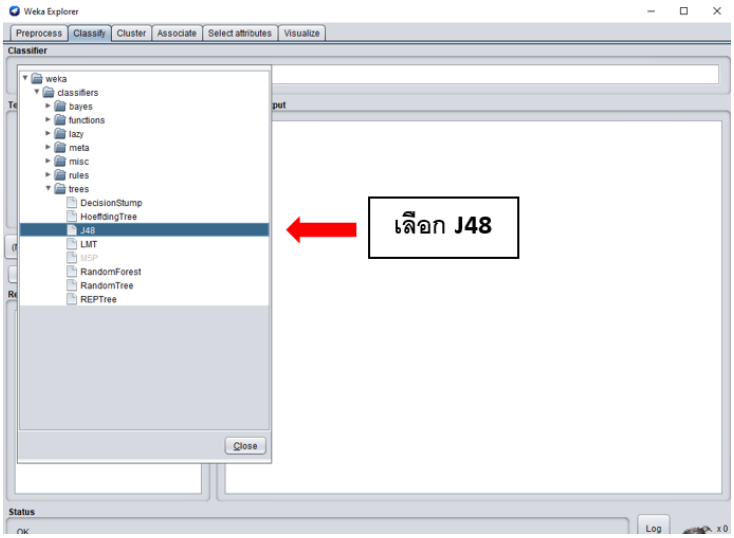


ภาพที่ 49 บทที่ 3 แสดงการนำเข้าข้อมูลเข้าในโปรแกรม Weka 3.8.5

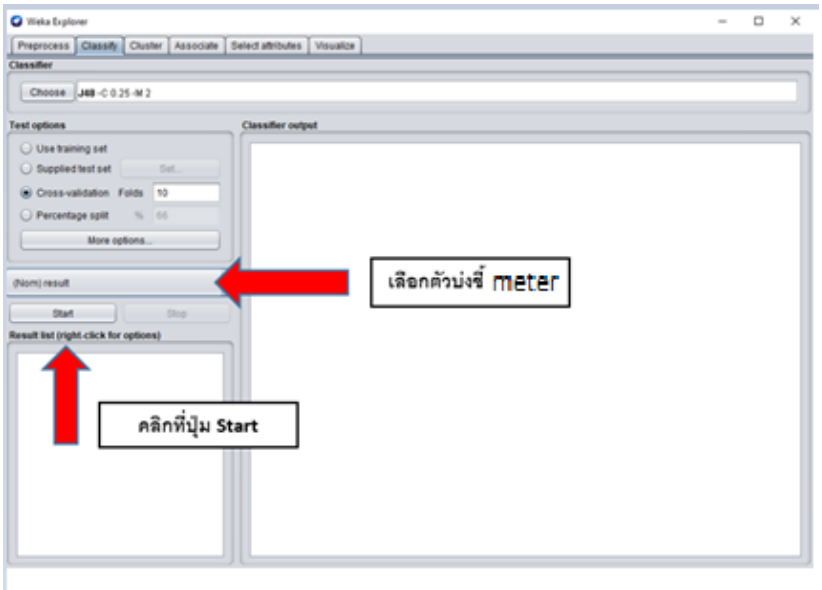


ภาพที่ 50 บทที่ 3 แสดงข้อมูลที่เข้าในโปรแกรม Weka 3.8.5

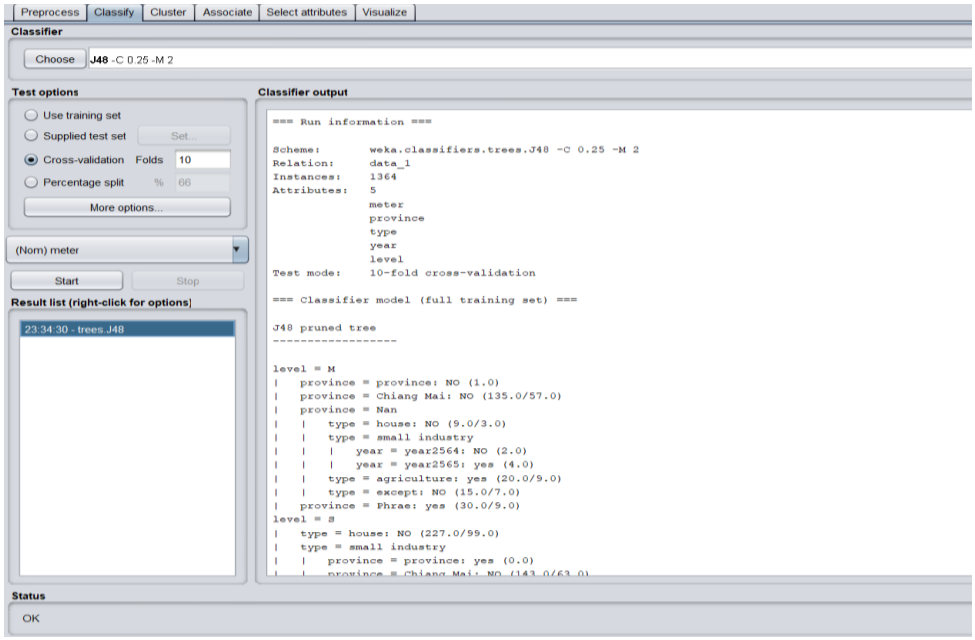
ขั้นตอนที่ 3 ดำเนินการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree โดยเลือกที่ Classification>>Choose>>tree และเลือกรูปแบบเป็น J48 ตามภาพที่ 3.14 จากนั้นเลือกตัวบ่งชี้ใน ที่นี้ใช้เป็นขนาดมิเตอร์ (meter) ส่วนที่ใช้ในโปรแกรม ใช่ชื่อว่า (meter) จากนั้นกดปุ่ม Start ตาม ภาพที่ 3.15 จะแสดงผลลัพธ์ที่ได้ตามภาพที่ 3.16



ภาพที่ 51 บทที่ 3 แสดงการเลือกเทคนิคการจัดกลุ่มข้อมูลแบบ Decision Tree รูปแบบเป็น J48

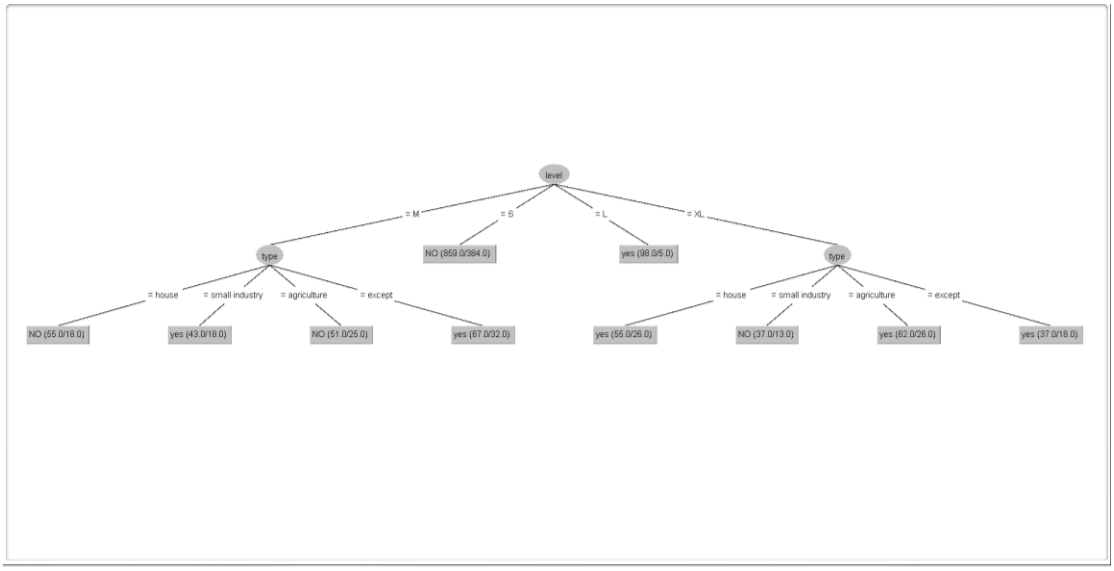


ภาพที่ 52 บทที่ 3 แสดงการเลือกตัวบ่งชี้เป็น meter แล้ว คลิกที่ start

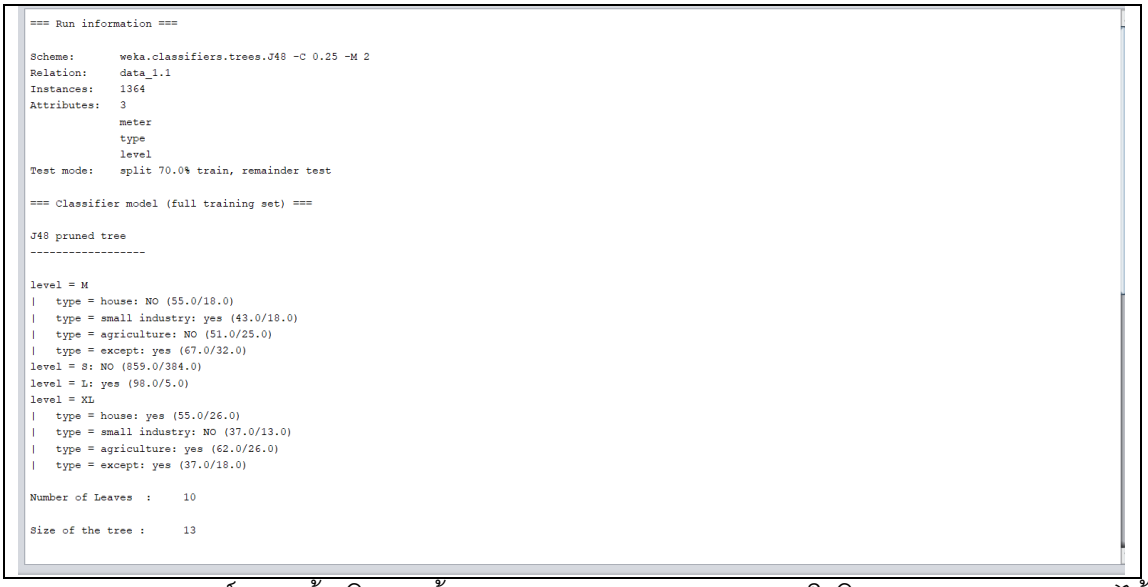


ภาพที่ 53 บทที่ 3 แสดงหน้าจอผลลัพธ์ของโมเดลการจัดกลุ่มข้อมูลแบบ Decision Tree : J48

จากผลลัพธ์การทดลองพบว่าเทคนิค Decision Tree : J48 ให้ผลลัพธ์การจำแนกประเภทขนาดมิเตอร์ที่ได้รับมาตรฐาน (yes) และ มิเตอร์ขนาดเล็ก (no) มีความถูกต้องถึง 58.06% แสดงผลลัพธ์แผนภาพโมเดลต้นไม้ตัดสินใจที่มีกิ่งแตกออกมา ดังภาพที่ 3.17



ภาพที่ 54 บทที่ 3 แสดงผลลัพธ์แผนภาพโมเดลต้นไม้ตัดสินใจ ในโปรแกรม Weka 3.8.5



```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     data_1.1
Instances:    1364
Attributes:   3
              meter
              type
              level
Test mode:    split 70.0% train, remainder test

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

J48 pruned tree
------------------

level = M
|   type = house: NO (55.0/18.0)
|   type = small industry: yes (43.0/18.0)
|   type = agriculture: NO (51.0/25.0)
|   type = except: yes (67.0/32.0)
level = S: NO (859.0/384.0)
level = L: yes (98.0/5.0)
level = XL
|   type = house: yes (55.0/26.0)
|   type = small industry: NO (37.0/13.0)
|   type = agriculture: yes (62.0/26.0)
|   type = except: yes (37.0/18.0)

Number of Leaves  :     10

Size of the tree :      13
```

จากผลลัพธ์การสร้างโมเดลด้วยเทคนิค Decision Tree: J48 ในโปรแกรม Weka 3.8.5 ได้สร้างกฎจากการจำแนกกลุ่มต้นไม้การตัดสินใจแบบ Decision Tree มีกิ่งแตกออกมา ดังภาพที่ 3.20

ภาพที่ 55 บทที่ 3 แสดงผลลัพธ์กฎต้นไม้การตัดสินใจ ของโปรแกรม Weka 3.8.5

ดังนั้นผู้วิเคราะห์ข้อมูลจะใช้เทคนิคของการจำแนกกลุ่มแบบ Decision Tree: J48 มาใช้ในการศึกษา เนื่องจากให้ผลลัพธ์ของกฎที่สามารถทำนายได้จำนวน 8 กฎ ซึ่งสามารถนำไปใช้ในการแบ่งกลุ่มได้ตามเงื่อนไขได้ชัดเจน และสามารถนำกฎที่ได้ไปวิเคราะห์กฎต่อไปได้ โดยสามารถจำแนกกฎได้ดังนี้

กฎข้อที่ 1 IF level M = type = house THEN meter NO การใช้ไฟฟ้าในระดับปานกลางประเภทบ้านใช้มิเตอร์ไฟฟ้า 30(100)

กฎข้อที่ 2 IF level M = type = small industey THEN meter yes การใช้ไฟฟ้าในระดับปานกลางในประเภทธุรกิจขนาดเล็กใช้มิเตอร์ไฟฟ้า 15(45)

กฎข้อที่ 3 IF level M = type = agriculture THEN meter NO การใช้ไฟฟ้าระดับปานกลางในประเภทเกษตรกรรมใช้มิเตอร์ไฟฟ้า 30(100)

กฎข้อที่ 4 IF level M = type = except THEN meter yes การใช้ไฟฟ้าระดับปานกลางในประเภทยกเว้นค่าไฟฟ้าใช้มิเตอร์ไฟฟ้า 15(45)

กฎข้อที่ 5 IF level S THEN meter NO แสดงถึงการใช้ไฟฟ้าน้อยแต่มิเตอร์ไฟฟ้ามีขนาด 30(100) การใช้ไฟฟ้าในปีนี้อาจจะน้อย

กฎข้อที่ 6 IF level L THEN meter yes การใช้ไฟฟ้าระดับสูงมีการใช้มิเตอร์ขนาด 15(45)

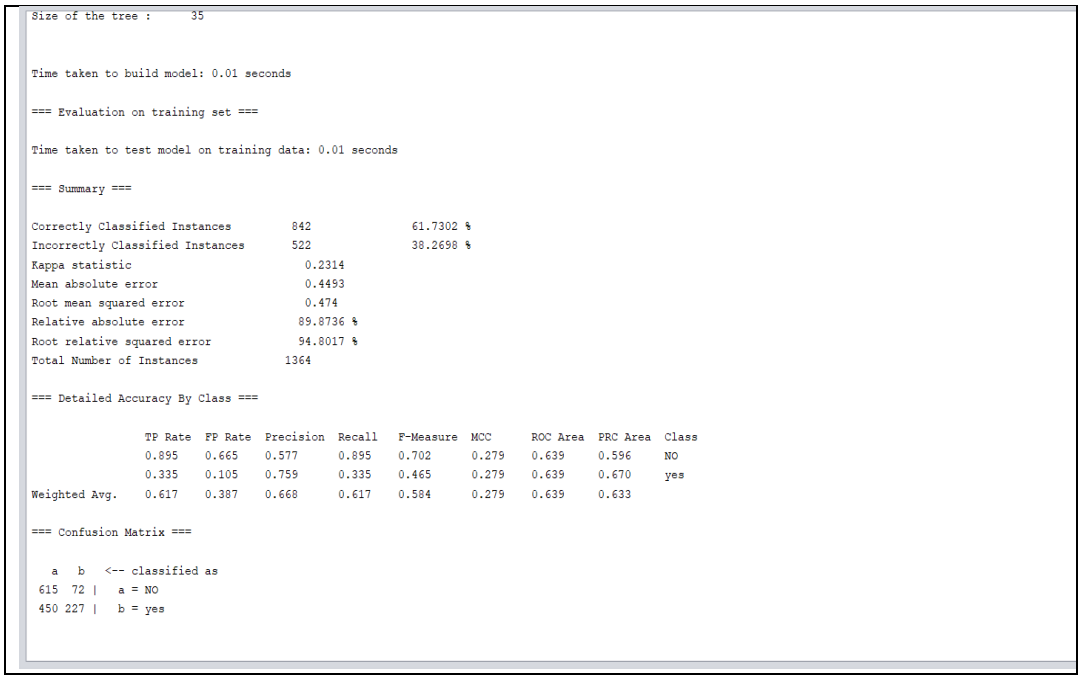
กฎข้อที่ 7 IF level XL = type= house THEN meter NO การใช้ไฟฟ้าระดับปานกลางในจังหวัดน่านประเภทยกเว้นค่าไฟฟ้า มิเตอร์ไฟฟ้าไม่เหมาะสม

กฎข้อที่ 8 IF level XL = type = small industey THEN meter NO การใช้ไฟฟ้าระดับสูงพิเศษประเภทธุรกิจขนาดเล็กใช้มิเตอร์ไฟฟ้า 30(100)

กฎข้อที่ 9 IF level XL = type = agriculture THEN meter yes การใช้ไฟฟ้าระดับสูงพิเศษในประเภทเกษตรกรรมใช้มิเตอร์ไฟฟ้า 15(45)

กฎข้อที่ 10 IF level XL = type = except THEN meter yes การใช้ไฟฟ้าระดับสูงพิเศษในประเภทยกเว้นค่าไฟฟ้าใช้มิเตอร์ไฟฟ้า 15(45)

หลังจากผู้วิเคราะห์ข้อมูลเลือกการทดสอบประสิทธิภาพของ Model ด้วยวิธี Self-Consistency Test หรือเรียกว่า Use Training Set เป็นวิธีการที่นำข้อมูลที่ใช้ในการสร้างโมเดล (model) และข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน คือข้อมูลการโจมตีที่เกิดขึ้นบนเว็บไซต์ ในปี 2019 ถึง ปี 2021 ที่ได้ทำการคัดเลือกมาทั้งหมด ซึ่งผู้วิเคราะห์ข้อมูลเลือกโปรแกรมที่ใช้นำเสนอ คือ โปรแกรม Weka 3.8.5 พบว่าการทดสอบประสิทธิภาพโมเดล Decision Tree (J48) พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้น มีค่าความถูกต้องเฉลี่ยในทุกโมเดลเท่ากับ 48.17% มีค่าการทำนายข้อมูลไม่ถูกต้องเท่ากับ 51.83% และมีค่าความคลาดเคลื่อนเท่ากับ 0.502 และเมื่อพิจารณาส่วนค่า Confusion Matrix ในภาพที่ 3.22 พบว่าการหาค่าของข้อมูลค่าจริง กับจำนวนข้อมูลจากการทำนาย แบ่งตามประเภทของขนาดของมิเตอร์ ขนาดมิเตอร์ที่ได้รับมาตรฐาน (yes) และ มิเตอร์ขนาดเล็ก (no) และนำมาหาค่าเฉลี่ยรวมของทุก class ได้ค่าเฉลี่ยรวมเท่ากับ 0.48 มีผลลัพธ์ตรงกันอยู่ในระดับค่อนข้างดี สามารถนำโมเดลไปใช้งานได้



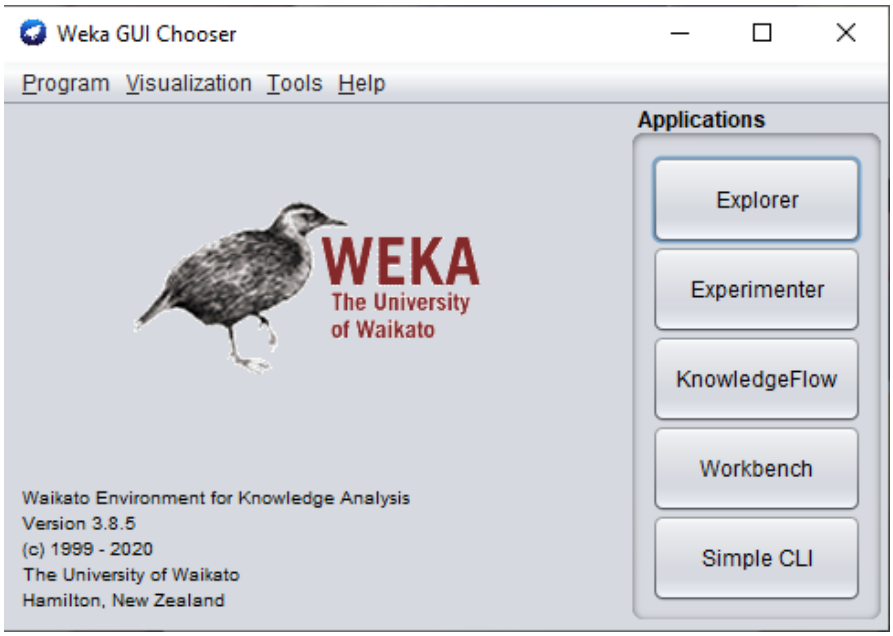
```
Size of the tree :      35


Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances         842               61.7302 %
Incorrectly Classified Instances       522               38.2698 %
Kappa statistic                          0.2314
Mean absolute error                      0.4493
Root mean squared error                  0.474
Relative absolute error                 89.8736 %
Root relative squared error             94.8017 %
Total Number of Instances             1364

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.895    0.665    0.577      0.895    0.702      0.279    0.639     0.596     NO
                 0.335    0.105    0.759      0.335    0.465      0.279    0.639     0.670     yes
Weighted Avg.    0.617    0.387    0.668      0.617    0.584      0.279    0.639     0.633

=== Confusion Matrix ===

   a   b   <-- classified as
 615  72 |   a = NO
 450 227 |   b = yes
```

ภาพที่ 1 แสดงผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree : J48 ในโปรแกรม Weka 3.8.5

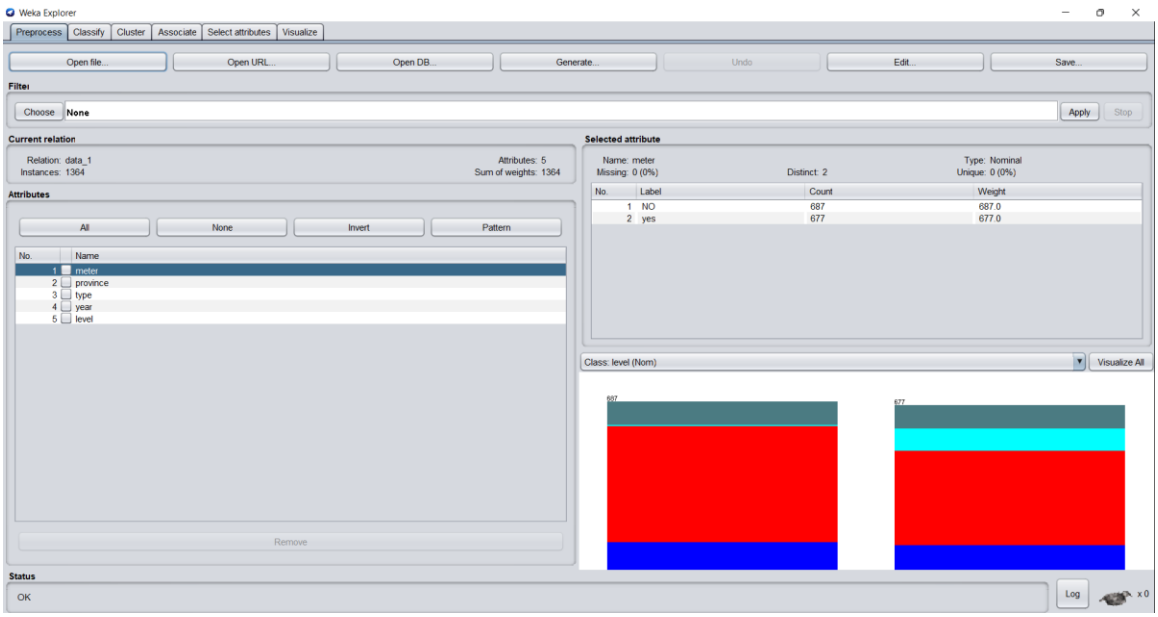
3.1.6 เผยแพร่ผลวิเคราะห์ (Deployment) ขั้นตอนการนำผลลัพธ์ที่ได้ไปใช้งานเป็นการทั่วไป อาจจัดทำเป็นรูปแบบของรายงาน (Report) หรือแผนภาพ (Dashboard) ที่พร้อมให้ฝ่ายต่าง ๆ นำไปใช้ประโยชน์ในการวางแผน กำหนดกลยุทธ์ และดำเนินการต่าง ๆ ในทางธุรกิจ

ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์มาแสดงผลข้อมูลบนหน้าเว็บไซต์ ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม

ขั้นตอนที่ 1 คลิกเปิดโปรแกรม Weka

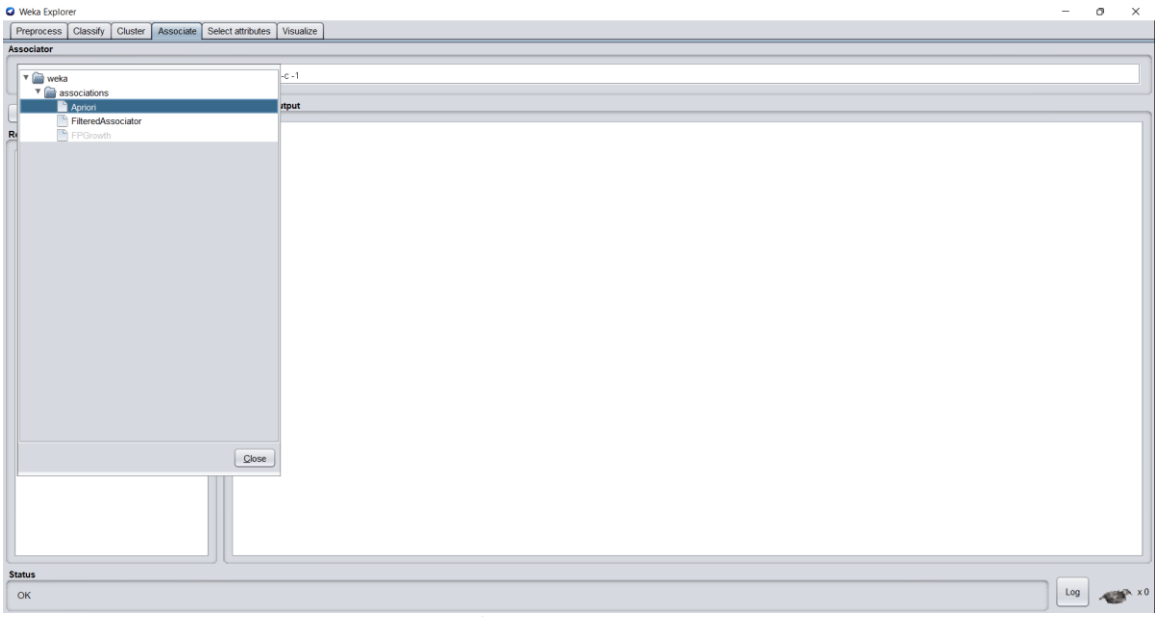


ภาพที่ 2 บทที่ 3 แสดงการเข้าหน้าจอโปรแกรม Weka 3.8.5



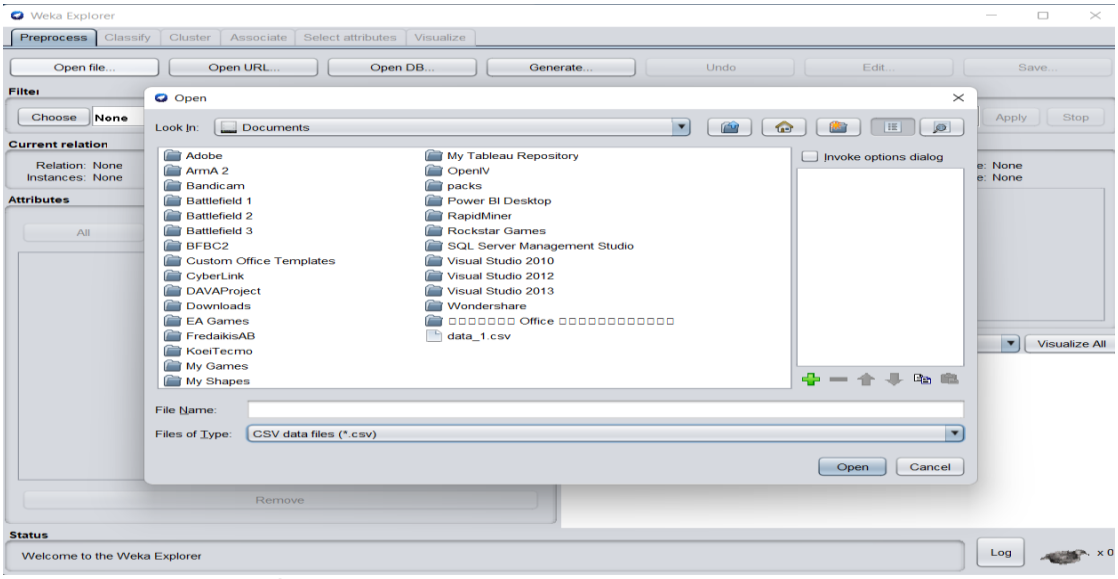
ภาพที่ 58 บทที่ 3 หน้าจอแสดงข้อมูลที่นำเข้าโปรแกรม

ขั้นตอนที่ 3 คลิกที่แท็บ Associate หน้าตาของ workspace จะเปลี่ยนไปเป็นส่วน ของแท็บ Associate แล้วคลิกที่ปุ่ม Choose จะมีลิสต์ (list) แสดงเทคนิคต่างๆ ของการ Associate ให้คลิกเลือกที่เมนูAssociations หลังจากนั้นให้เลือกที่เทคนิคการจ าแนกข้อมูลด้วย Apriori โดยคลิกที่เมนู Apriori



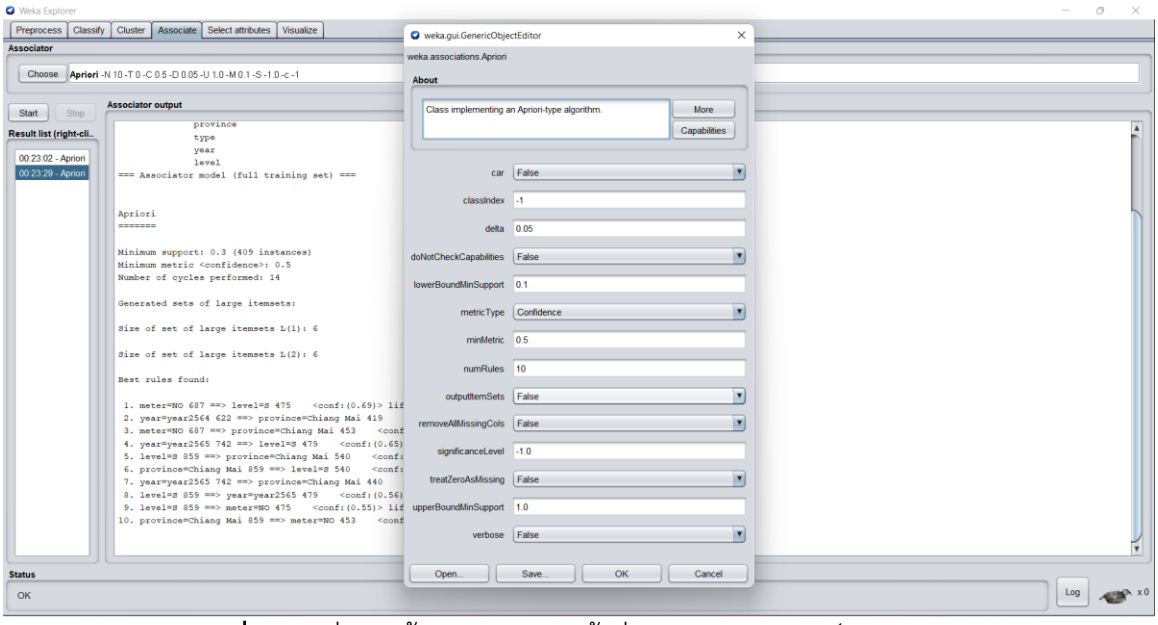
ภาพที่ 59 บทที่ 3 เลือกไฟล์

ขั้นตอนที่ 2 นำข้อมูลที่ได้จัดเตรียมไว้ โดยเลือกที่ Explore >> Openfile เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบ



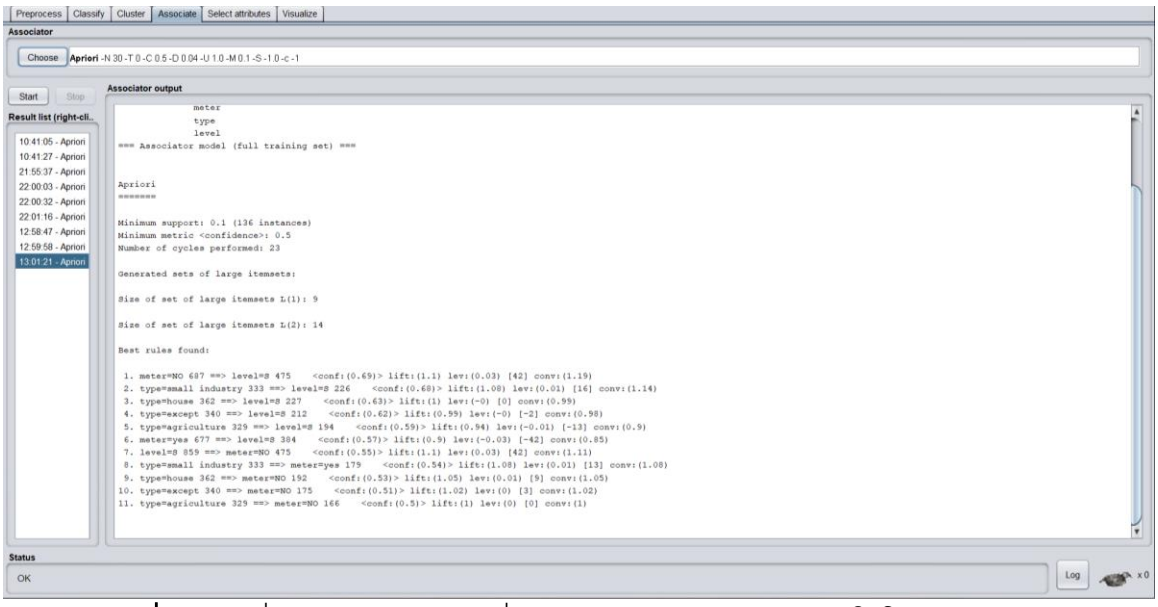
ภาพที่ 60 บทที่ 3 แสดงการนำข้อมูลเข้าในโปรแกรม Weka

ขั้นตอนที่ 4 เลือก ในการใช้เทคนิคการคำนวณแบบ Association บนแถบ Associateจำเป็นต้องกำหนดการตั้งค่าในค่า lowerBoundMinSupport คือ การกำหนดค่า MinimumSupport ค่า metricType คือ การกำหนด Type ของกฎความสัมพันธ์ ค่า minMetric คือ การกำหนดค่า minimum ของ Type กฎความสัมพันธ์ ค่า numRules คือ การระบุจำนวนกฎ 10ความสัมพันธ์ที่ต้องการกำหนดให้ค่า Minimum Support เท่ากับ 0.01 เนื่องจากข้อมูลมีการกระจายกลุ่มข้อมูลอย่างมาก และ มีปริมาณที่สูงเกินไป ทำให้การตั้งค่า Minimum Support ที่มีค่ามากกว่า 0.01 จะไม่สามารถหากลุ่ม ItemSet และ Best Rules กฎความสัมพันธ์ที่ดีที่สุดคือค่า Minimum Support = 0.1 ค่า Minimum Confident = 0.80 ค่าที่แสดงกฎความสัมพันธ์NumRules เป็น 10 กฎความสัมพันธ์ที่ดีที่สุด



ภาพที่ 3 บทที่ 3 หน้าจอแสดงการตั้งค่ากฎความสัมพันธ์แบบ Apriori

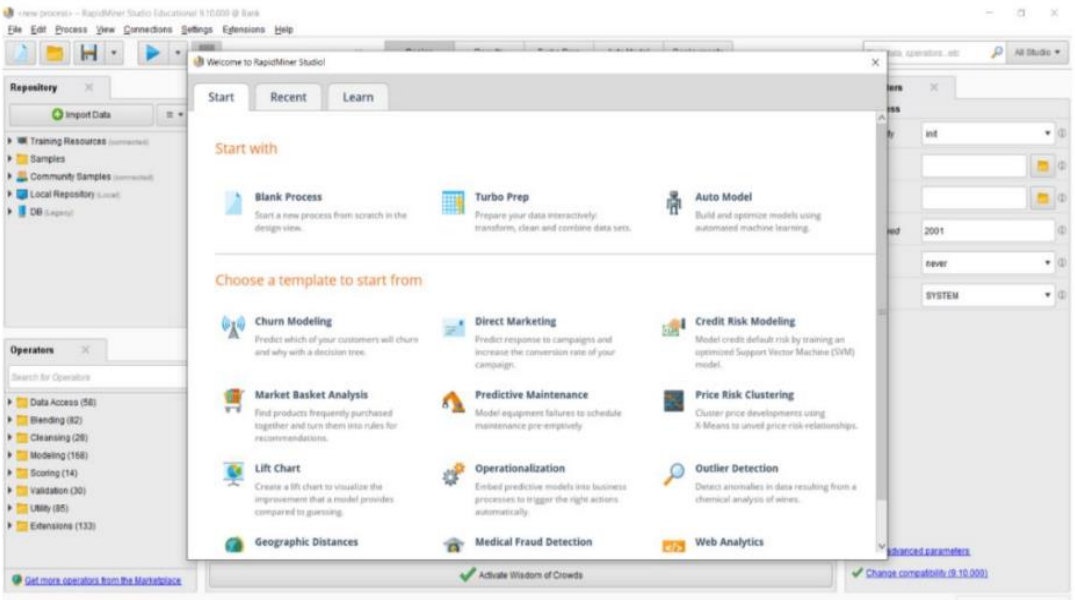
ขั้นตอนที่ 5 หลังจากนั้นคลิกที่ปุ่ม Start Weka จะทำงานและแสดงผลลัพธ์การทำงาน



ภาพที่ 62 บทที่ 3 ผลการจำแนกกลุ่มแบบ Associations : Apriori ในโปรแกรม Weka

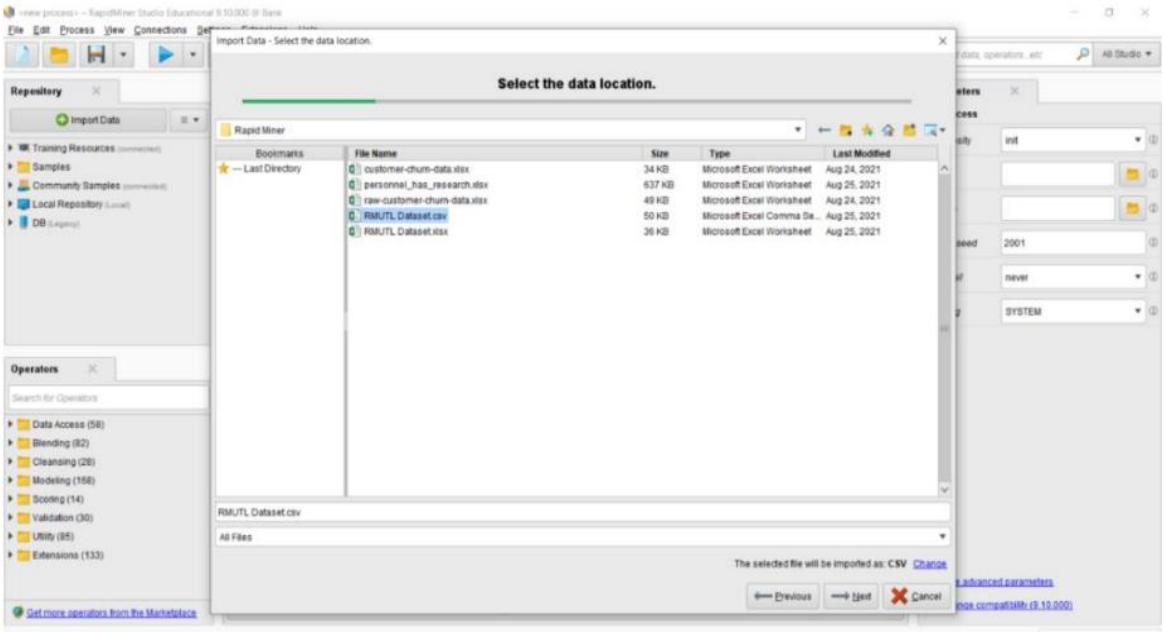
3.1.6 สร้างโมเดล Association Rule ด้วยโปรแกรม RapidMiner Studio

ขั้นตอนที่ 1 เปิดโปรแกรม RapidMiner Studio

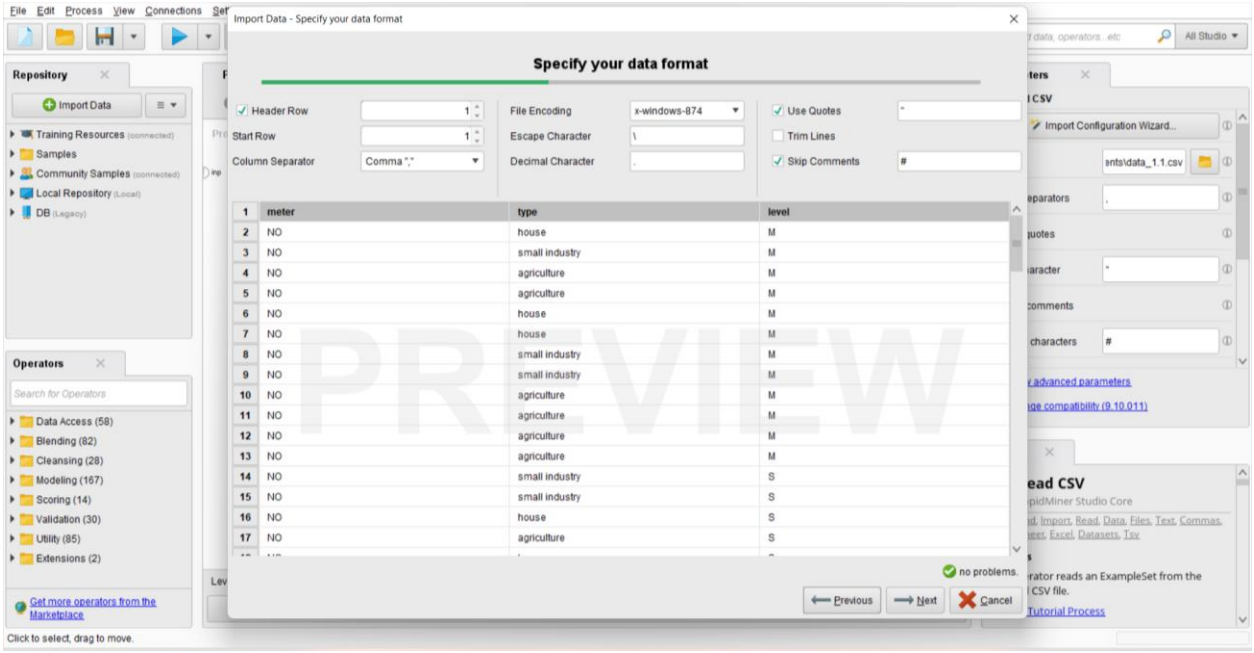


ภาพที่ 63 บทที่ 3 แสดงการเข้าหน้าจอโปรแกรม RapidMiner Studio

ขั้นตอนที่ 2 เลือก Blank Process เพื่อเปิดหน้าต่างโปรแกรมขึ้นมา

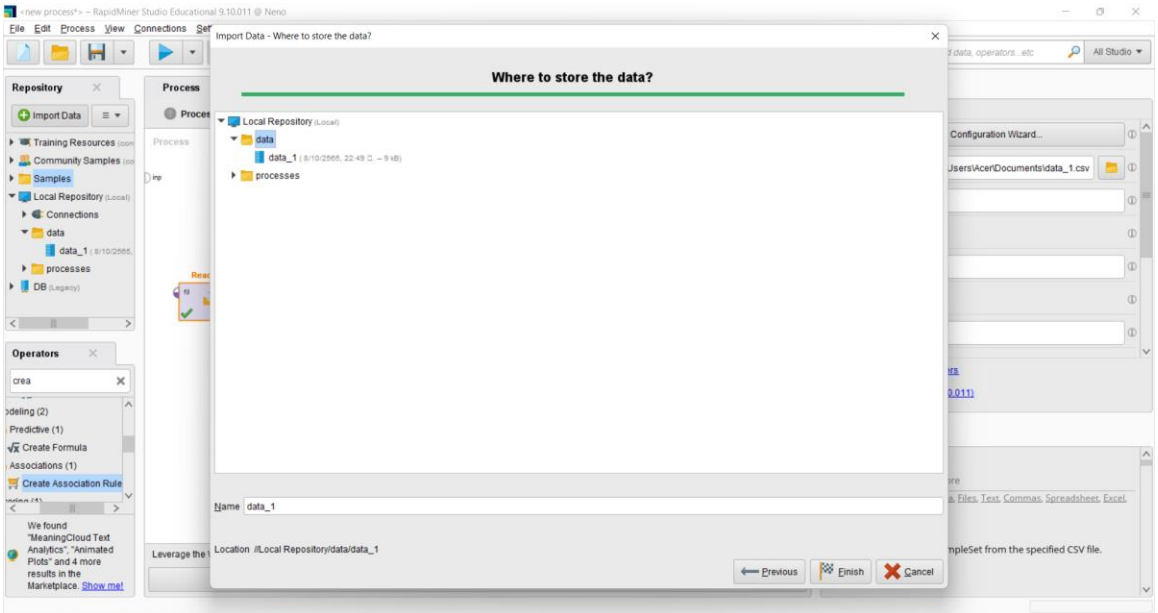


ภาพที่ 64 บทที่ 3 แสดงการเลือกข้อมูลที่จะนำมาวิเคราะห์



ภาพที่ 4 บทที่ 3 แสดงข้อมูลทั้งหมดของไฟล์ที่นำเข้าข้อมูล

ขั้นตอนที่ 4 โปรแกรมจะแสดงผลของข้อมูลออกมาทั้งหมด โดย Row1 จะแสดงชื่อของแอตทริบิวต์ (Attribute) ในแต่ละแถว และ Row2 เป็นต้นไปจะแสดงข้อมูลที่นำเข้า



ภาพที่ 5 บทที่ 3 แสดงพื้นที่ในการจัดเก็บข้อมูล

ขั้นตอนที่ 5 ทำการเลือกพื้นที่ในการจัดเก็บข้อมูลเพื่อนำไปใช้ในการวิเคราะห์โปรแกรมRapidMiner Studio

ภาพที่ 67 บทที่ 3 แสดงโมเดล Association Rule ด้วยโปรแกรม RapidMIner Studio

ขั้นตอนที่ 6 สร้างโมเดลเพื่อเข้าสู่กระบวนการวิเคราะห์ข้อมูล Retrieve โหนดนี้หลักการทำงานอ่านไฟล์ข้อมูลที่จัดเตรียมไว้ สามารถอ่านไฟล์ได้แค่ไฟล์เดียว

	meter	type	level
2	NO	house	M
3	NO	small industry	M
4	NO	agriculture	M
5	NO	agriculture	M
6	NO	house	M
7	NO	house	M
8	NO	small industry	M
9	NO	small industry	M
10	NO	agriculture	M
11	NO	agriculture	M
12	NO	agriculture	M
13	NO	agriculture	M
14	NO	small industry	S
15	NO	small industry	S
16	NO	house	S
17	NO	agriculture	S

PREVIEW

ภาพที่ 68 บทที่ 3 วิเคราะห์ข้อมูลการใช้ไฟฟ้า Project โดยโปรแกรม Rapid Miner กำหนดค่า Confident 60%

ผู้วิเคราะห์ข้อมูลนำผลข้อมูลที่ทำการวิเคราะห์มาแสดงผลข้อมูลบนหน้าเว็บไซต์ ร่วมกับการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม TableauPublic

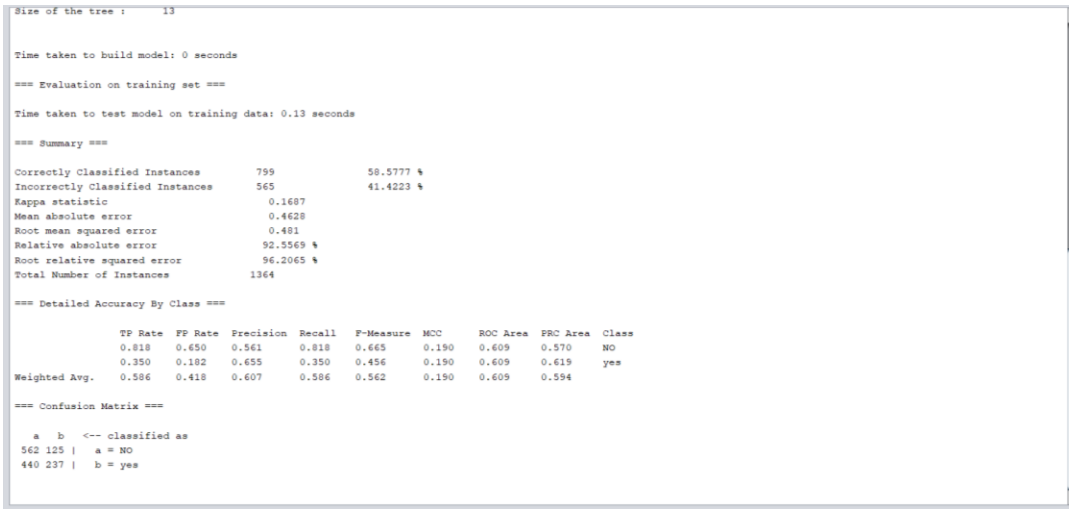
```
            meter
            type
            level
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.1 (136 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 23

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 14

Best rules found:

 1. meter=NO 687 ==> level=S 475    <conf:(0.69)> lift:(1.1) lev:(0.03) [42] conv:(1.19)
 2. type=small industry 333 ==> level=S 226    <conf:(0.68)> lift:(1.08) lev:(0.01) [16] conv:(1.14)
 3. type=house 362 ==> level=S 227    <conf:(0.63)> lift:(1) lev:(-0) [0] conv:(0.99)
 4. type=except 340 ==> level=S 212    <conf:(0.62)> lift:(0.99) lev:(-0) [-2] conv:(0.98)
 5. type=agriculture 329 ==> level=S 194    <conf:(0.59)> lift:(0.94) lev:(-0.01) [-13] conv:(0.9)
 6. meter=yes 677 ==> level=S 384    <conf:(0.57)> lift:(0.9) lev:(-0.03) [-42] conv:(0.85)
 7. level=S 859 ==> meter=NO 475    <conf:(0.55)> lift:(1.1) lev:(0.03) [42] conv:(1.11)
 8. type=small industry 333 ==> meter=yes 179    <conf:(0.54)> lift:(1.08) lev:(0.01) [13] conv:(1.08)
 9. type=house 362 ==> meter=NO 192    <conf:(0.53)> lift:(1.05) lev:(0.01) [9] conv:(1.05)
10. type=except 340 ==> meter=NO 175    <conf:(0.51)> lift:(1.02) lev:(0) [3] conv:(1.02)
11. type=agriculture 329 ==> meter=NO 166    <conf:(0.5)> lift:(1) lev:(0) [0] conv:(1)
```

ภาพที่ 69 บทที่ 3 วิเคราะห์ข้อมูลการใช้ไฟฟ้า Project โดยโปรแกรม Weka กำหนดค่า Confident 60%

3.1.7 การประเมินผล (Evaluation)เป็นขั้นตอนก่อนนำผลไปวิเคราะห์ข้อมูลด้วยการวัดประสิทธิผลที่ได้กับกับวัตถุประสงค์ที่ตั้งไว้ในตอนแรกว่ามีความสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด ด้วยการประเมินผลจากโปรแกรม RapidMiner Studio และ Wekaค่า confidence แสดงความเชื่อมั่นของกฎความสัมพันธ์ที่เมื่อรูปแบบ A เกิดขึ้นแล้วรูปแบบ B จะเกิดขึ้นด้วยเป็นจำนวนกี่เปอร์เซ็นต์ การคำนวณค่า confidence หาได้จาก โดยที่support (A,B) คือ ค่า support ที่รูปแบบ A และ B เกิดขึ้นพร้อม ๆ กัน

Confidence (A => B) = Support (A and B) Support (A)ถ้าค่า Confidence เท่ากับ 1 (100%) หรือใกล้เคียง ก็แสดงว่ามีค่าความเชื่อมั่นที่สูงค่า Lift จะเป็นมาตรวัดที่ใช้วัดประสิทธิภาพกฎความสัมพันธ์โดยจะวัดอิทธิพลของกฎความสัมพันธ์ที่เกิดขึ้นLift (A => B) = Support (A and

B)Support (A) x Support (B) ถ้าค่า Lift มากกว่า 1 เท่า แสดงว่าถ้าเกิด A แล้วมีความน่าจะเป็นที่จะมีโอกาสเกิด B มากขึ้น



```
Size of the tree :      13


Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.13 seconds

=== Summary ===

Correctly Classified Instances         799               58.5777 %
Incorrectly Classified Instances       565               41.4223 %
Kappa statistic                          0.1687
Mean absolute error                      0.4628
Root mean squared error                  0.481
Relative absolute error                 92.5569 %
Root relative squared error             96.2065 %
Total Number of Instances             1364

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.818    0.650    0.561      0.818    0.665      0.190    0.609     0.570     NO
               0.350    0.182    0.655      0.350    0.456      0.190    0.609     0.619     yes
Weighted Avg.  0.586    0.418    0.607      0.586    0.562      0.190    0.609     0.594

=== Confusion Matrix ===

   a   b   <-- classified as
 562 125 |   a = NO
 440 237 |   b = yes
```

ภาพที่ 6 บทที่ 3 ผล Accuracy หรือค่าความแม่นยำของโมเดล Decision Tree

```
            meter
            type
            level
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.1 (136 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 23

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 14

Best rules found:

 1. meter=NO 687 ==> level=S 475    <conf:(0.69)> lift:(1.1) lev:(0.03) [42] conv:(1.19)
 2. type=small industry 333 ==> level=S 226    <conf:(0.68)> lift:(1.08) lev:(0.01) [16] conv:(1.14)
 3. type=house 362 ==> level=S 227    <conf:(0.63)> lift:(1) lev:(-0) [0] conv:(0.99)
 4. type=except 340 ==> level=S 212    <conf:(0.62)> lift:(0.99) lev:(-0) [-2] conv:(0.98)
 5. type=agriculture 329 ==> level=S 194    <conf:(0.59)> lift:(0.94) lev:(-0.01) [-13] conv:(0.9)
 6. meter=yes 677 ==> level=S 384    <conf:(0.57)> lift:(0.9) lev:(-0.03) [-42] conv:(0.85)
 7. level=S 859 ==> meter=NO 475    <conf:(0.55)> lift:(1.1) lev:(0.03) [42] conv:(1.11)
 8. type=small industry 333 ==> meter=yes 179    <conf:(0.54)> lift:(1.08) lev:(0.01) [13] conv:(1.08)
 9. type=house 362 ==> meter=NO 192    <conf:(0.53)> lift:(1.05) lev:(0.01) [9] conv:(1.05)
10. type=except 340 ==> meter=NO 175    <conf:(0.51)> lift:(1.02) lev:(0) [3] conv:(1.02)
11. type=agriculture 329 ==> meter=NO 166    <conf:(0.5)> lift:(1) lev:(0) [0] conv:(1)
```

ภาพที่ 71 บทที่ 3 ผล Confidence ของโมเดล Association Rule ผ่านโปรแกรม weka

...	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Co
16	NO	S	0.348	0.691	0.897	-0.659	0.031	1.098	1.2
15	small industry	S	0.166	0.679	0.937	-0.323	0.012	1.078	1.1
14	house	S	0.166	0.627	0.922	-0.364	-0.001	0.996	0.9
13	except	S	0.155	0.624	0.925	-0.343	-0.002	0.990	0.9
12	agriculture	S	0.142	0.590	0.920	-0.340	-0.010	0.936	0.9
11	yes	S	0.282	0.567	0.856	-0.711	-0.031	0.901	0.8
10	S	NO	0.348	0.553	0.827	-0.911	0.031	1.098	1.1
9	small industry	yes	0.131	0.538	0.909	-0.357	0.010	1.083	1.0
8	house	NO	0.141	0.530	0.902	-0.390	0.007	1.053	1.0
7	except	NO	0.128	0.515	0.903	-0.370	0.003	1.022	1.0
6	agriculture	NO	0.122	0.505	0.904	-0.361	0.000	1.002	1.0
5	agriculture	yes	0.120	0.495	0.902	-0.363	-0.000	0.998	0.9
4	except	yes	0.121	0.485	0.897	-0.378	-0.003	0.978	0.9

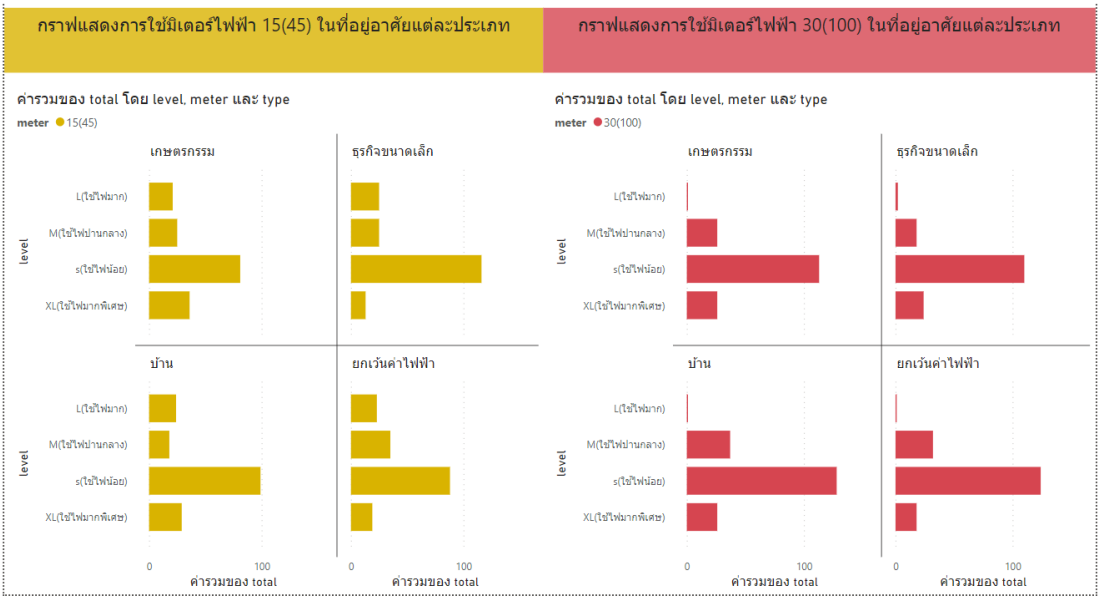
ภาพที่ 72 บทที่ 3 ผล Confidence ของโมเดล Association Rule ผ่านโปรแกรม RapidMiner Studio

3.1.8 สรุปผลการเปรียบเทียบ

ทางผู้จัดทำได้ทำการสร้างโมเดลขึ้นมาทั้งหมด 2 โมเดลเพื่อทำการเปรียบเทียบข้อมูลผลการวิเคราะห์จากทั้งโมเดล Decision Tree และ Association Rule ทั้งสองโมเดลได้อัพโหลดชุดข้อมูลผ่านโปรแกรม weka แต่ในโมเดล Association Rule การสร้างโมเดลในอีกหนึ่งโปรแกรมคือ RapidMiner Studio เพื่อสร้างแบบจำลองชุดข้อมูล ทั้งสองโมเดลได้มีข้อมูลที่ตรงกันและสามารถนำข้อมูลที่ได้มาใช้ประโยชน์ในด้านต่างๆที่สนใจ ทางผู้จัดทำได้ทำการเปรียบเทียบทั้งสองโมเดล จากโมเดล Decision Tree ค่าความแม่นยำจะอยู่ที่ 58.57% และโมเดล Association Rule ค่า Confidence จะอยู่ที่ 0.69 จะเห็นได้ว่าโมเดล Association Rule จะมีประสิทธิภาพในการสร้างกฎความสัมพันธ์ที่มากกว่า

TableauPublic

ภาพที่ 73 บทที่ 3 กราฟแผนภูมิวงกลมของระดับการใช้ไฟฟ้าและการใช้มิเตอร์ขนาด 15(45) และ 30(100) โดยโปรแกรม power bi



ภาพที่ 74 บทที่ 3 กราฟแผนภูมิวงกลมของการใช้ไฟฟ้าใน3จังหวัด โดยโปรแกรม power bi

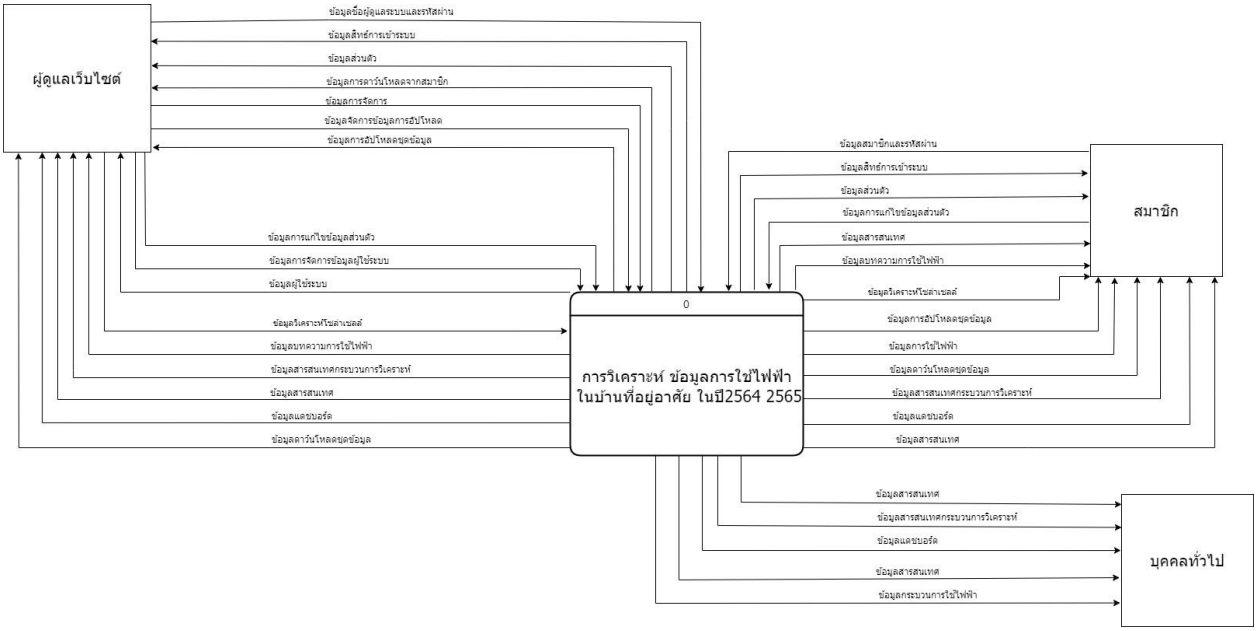
3.2 แผนภาพบริบท (Context Diagram)

แผนภาพบริบทเป็นแผนภาพที่แสดงถึงภาพรวมของระบบ และความสัมพันธ์ระบบกับสิ่งแวดล้อมที่เกี่ยวข้องกับระบบรวมถึงเหตุการณ์ต่าง ๆ ที่ใช้ในระบบการพัฒนาเว็บไซต์สำหรับวิเคราะห์ข้อมูลการใช้ไฟฟ้าในบ้านที่อยู่อาศัย ในปี 2564-2565 ซึ่งสามารถแบ่งออกมาได้ ดังนี้

ตารางที่ 3.1 ตารางรายละเอียดของเอ็กเทอร์นัลเอ็นทิตี้ และโปรเซสที่เกี่ยวข้อง

ผู้ใช้	รายการข้อมูล	รายการโปรเซส
1.) บุคคลทั่วไป 2.) สมาชิก 3.) ผู้ดูแลเว็บไซต์	1.) แฟ้มข้อมูลผู้ใช้งาน 2.) แฟ้มข้อมูลการใช้ไฟฟ้า 3.) แฟ้มข้อมูลการดาวน์โหลดไฟล์ 4.) แฟ้มข้อมูลการจัดเก็บไฟล์	1.) ตรวจสอบการเข้าสู่ระบบ 2.) จัดการข้อมูลส่วนตัว 3.) จัดการข้อมูลหน้าเว็บไซต์ 4.) รายงานข้อมูลสารสนเทศของเว็บไซต์

จากการกำหนดผู้ใช้และความต้องการ ที่ใช้ระบบ คือ บุคคลทั่วไป, สมาชิก และผู้ดูแลเว็บไซต์ สามารถแสดงความสัมพันธ์ด้วยแผนผังบริบทดังนี้



ภาพที่ 75 บทที่ 3 แผนภาพบริบท (Context Diagram)

จากรูปภาพที่ 3.34 เป็นแผนภาพบริบทระบบของการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลการใช้ไฟฟ้าในบ้านที่อยู่อาศัย ในปี 2564–2565 โดยสามารถแบ่งผู้ใช้ออกเป็น 3 ประเภท ดังนี้

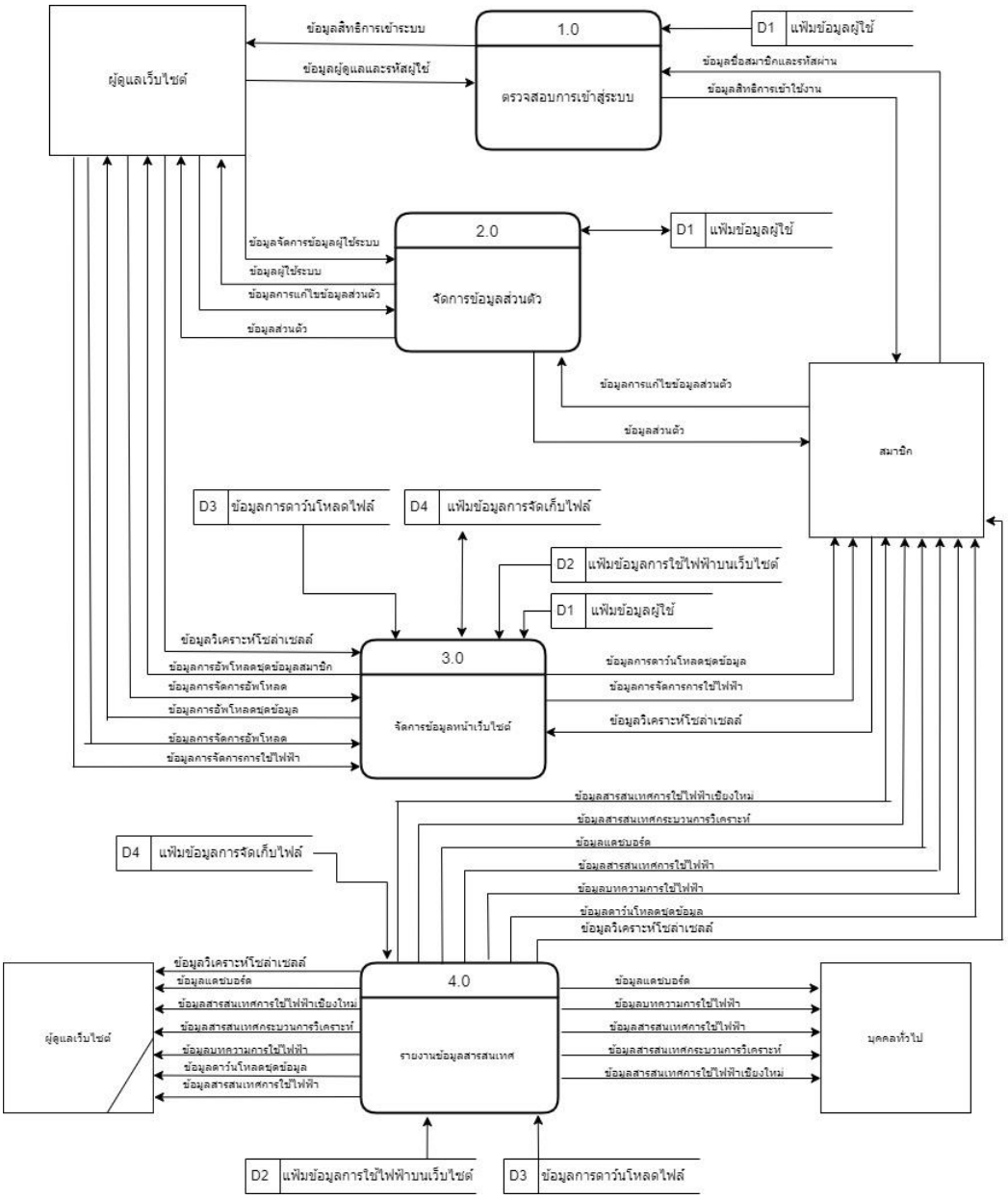
1.) บุคคลทั่วไป สามารถดูข้อมูลบทความเกี่ยวกับเว็บไซต์และการใช้ไฟฟ้า และเกี่ยวกับผู้จัดทำเว็บไซต์ได้ สามารถดูสารสนเทศกระบวนการวิเคราะห์ข้อมูลได้ สามารถดูสารสนเทศการวิเคราะห์ข้อมูล decision tree ได้ สามารถดูสารสนเทศรายงานการวิเคราะห์ข้อมูลต่างๆ ข แสดงผลในรูปแบบของ data Visualization ได้ ดูสารสนเทศรายงานการวิเคราะห์ข้อมูลอื่นๆ แสดงผลในรูปแบบของ data Visualization ได้ และดูสารสนเทศแผนภูมิรูปภาพต่างๆ

2) สมาชิก สามารถกรอกข้อมูลการใช้ไฟฟ้าบนเว็บไซต์ กรณีที่ ข้อมูลมีจำนวนน้อยได้ สามารถอัปโหลดชุดข้อมูลเกี่ยวกับการใช้ไฟฟ้าที่มีจำนวนมากๆ เพื่อให้ผู้ดูแลเว็บไซต์นำข้อมูลไปจัดการข้อมูลที่อัปเดตลงบนเว็บไซต์ สามารถดาวน์โหลดไฟล์ชุดข้อมูลที่ผู้ดูแลเว็บไซต์เปิดให้ดาวน์โหลดเพื่อนำไปศึกษาหรือใช้งานต่อได้ สามารถดูสารสนเทศกระบวนการวิเคราะห์ข้อมูลได้ สามารถดูสารสนเทศการวิเคราะห์ข้อมูล decision tree ได้ สามารถดูสารสนเทศรายงานการวิเคราะห์ข้อมูลต่างๆ แสดงผลในรูปแบบของ data Visualization ได้ ดูสารสนเทศรายงานการวิเคราะห์ข้อมูลอื่นๆ แสดงผลในรูปแบบของ data Visualization ได้ และดูสารสนเทศแผนภูมิรูปภาพต่างๆ

3.) ผู้ดูแลเว็บไซต์ สามารถเพิ่ม ลบ และแก้ไขข้อมูลสารสนเทศการวิเคราะห์ข้อมูลการใช้ไฟฟ้าได้ สามารถจัดการสมาชิก ลบ แก้ไข ข้อมูล และดูข้อมูลสมาชิกได้ สามารถดาวน์โหลดไฟล์ชุดข้อมูลที่ ผู้ใช้งานอัปโหลดเข้ามา เพื่อนำไปทำ data cleaning ให้ข้อมูลนั้นพร้อมนำมาอัปเดตบนเว็บไซต์ได้ สามารถกรอกข้อมูลการใช้ไฟฟ้า กรณีที่ ข้อมูลมีจำนวนน้อยได้ สามารถอัปเดตชุดข้อมูลการใช้ไฟฟ้าที่มีจำนวนมากๆ เพื่อให้ข้อมูลที่อัปโหลดมานั้นนำใช้แสดงผลในรูปแบบของ data Visualization บนเว็บไซต์ได้ สามารถดูสารสนเทศกระบวนการวิเคราะห์ข้อมูลได้ สามารถดูสารสนเทศการวิเคราะห์ข้อมูล decision tree ได้ สามารถดูสารสนเทศรายงานการวิเคราะห์ข้อมูลต่างๆ แสดงผลในรูปแบบของ data Visualization ได้ ดูสารสนเทศรายงานการวิเคราะห์ข้อมูลอื่นๆ แสดงผลในรูปแบบของ data Visualization ได้ และดูสารสนเทศแผนภูมิรูปภาพต่างๆ

3.5 แผนภาพกระแสข้อมูล Data Flow Diagram

แผนภาพกระแสข้อมูลเป็นแผนภาพที่แสดงถึงกระบวนการทำงานต่าง ๆ ของระบบว่ามีผู้ใช้งานเกี่ยวข้องกับกระบวนการทำงานในด้านใดบ้าง และแสดงการไหลของข้อมูลในกระบวนการรวมถึงการจัดเก็บ



ภาพที่ 76 บทที่ 3 แผนภาพ Data Flow Diagram

พจนานุกรมข้อมูล (Data Dictionary)

การวิเคราะห์เพื่อให้ได้มาซึ่งแผนภาพอีอาร์ หรืออีอาร์ไดอะแกรมนั้นจะให้พื้นฐานหลักอยู่ 3 ประการด้วยกัน ได้แก่

3.3.1 เอ็นติตี้ (Entity) คือ บุคคล วัตถุ สถานที่ และรวมถึงเหตุการณ์ที่ทำให้เกิดกลุ่มของข้อมูลที่ต้องการจัดเก็บ ซึ่งบ่งชี้ถึงความเป็นเอกลักษณ์เฉพาะตัวได้ (Uniquely identifiable)

3.3.2 ความสัมพันธ์(Relation) คือ ค่าความสัมพันธ์ระหว่างเอ็นติตี้

3.3.3 แอททริบิวต์(Attribute) คือ คุณสมบัติของเอ็นติต

ตารางที่ 10 บทที่ 3 แสดงเอ็นติตี้ทั้งหมดภายในกระบวนการของการพัฒนาเว็บไซต์สำหรับการวิเคราะห์การใช้ไฟฟ้าในบ้านที่อยู่อาศัยในปี 2564-2565

ลำดับ	ชื่อตาราง	ประเภท	รายละเอียด
D1	Users	Master	แฟ้มข้อมูลผู้ใช้งาน
D2	Data	Transaction	แฟ้มข้อมูลการใช้ไฟฟ้า
D3	downloadfile	Transaction	แฟ้มข้อมูลการดาวน์โหลดไฟล์
D4	uploadfile	Transaction	แฟ้มข้อมูลการจัดเก็บไฟล์

คำอธิบาย: ประเภทของตาราง ได้แก่

master หมายถึง ตารางข้อมูลหลัก

transaction หมายถึง ตารางที่มีการเปลี่ยนแปลงของข้อมูล

reference หมายถึง ตารางที่มีการอ้างอิง

คำอธิบาย: ประเภทของตาราง ได้แก่ master หมายถึง ตารางข้อมูลหลัก transaction หมายถึง ตารางที่มีการเปลี่ยนแปลงของข้อมูล reference หมายถึง ตารางที่มีการอ้างอิง

ตารางที่ 11 บทที่ 3 แสดงรายละเอียดของตาราง user

ชื่อตาราง : Users ประเภทตาราง : Master คำอธิบาย : แฟ้มข้อมูล คีย์หลัก : id			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
id	varchar(6)	รหัสผู้ใช้งาน	1
username	varchar(100)	ผู้สมาชิก	admin
e-mail	varchar(100)	อีเมล์	benceo@gmail.com
password	varchar(50)	รหัสผ่าน	123456
address	varchar(255)	ที่อยู่	84 ม.7 ต.สันทราย อ.เมือง จ.เชียงใหม่
Status	tiny int(1)	สิทธิ์	1
tel	varchar(10)	เบอร์โทร	0805157370
profile	varchar(100)	รูปภาพผู้ใช้งาน	35456.jpg

คำอธิบาย: สิทธิ์ ได้แก่

หมายเลข 1 หมายถึง admin ผู้ดูแลเว็บไซต์

หมายเลข 2 หมายถึง member ผู้สมาชิกเว็บไซต์

ตารางที่ 12 บทที่ 3 แสดงรายละเอียดของตาราง Data

ชื่อตาราง : Data ประเภทตาราง : Master คำอธิบาย : เก็บข้อมูลผู้ใช้งาน คีย์หลัก : data_id			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
data_id	int(11)	ลำดับข้อมูล	5
Province	varchar(100)	จังหวัด	น่าน
Type	varchar(100)	ประเภท	บ้าน
Watt	varchar(100)	จำนวนวัตต์	1023
Meter	varchar(100)	ขนาดมิเตอร์	30(100)A
Date	timestamp	วัน/เดือน/ปี	2022-09-05 20:59:14

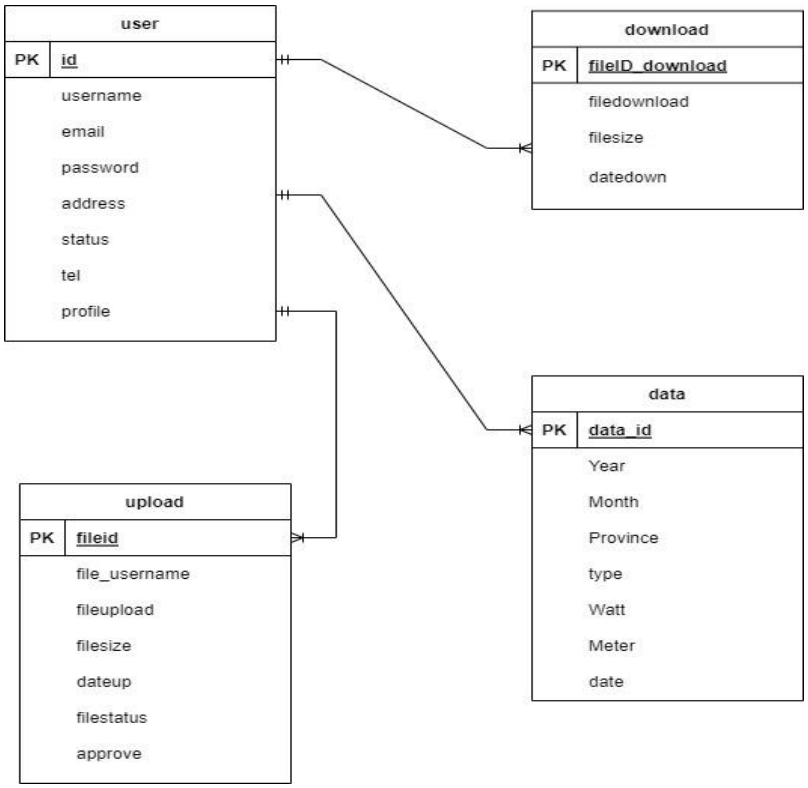
ตารางที่ 13 บทที่ 3 แสดงรายละเอียดของตาราง downloadfile

ชื่อตาราง : downloadfile ประเภทตาราง : Master คำอธิบาย : แฟ้มข้อมูลการจัดเก็บไฟล์ คีย์หลัก : fileID_download			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
fileID_download	int(11)	ลำดับไฟล์	98
filedownload	varchar(100)	ชื่อชุดข้อมูล	แบบฟอร์ม.csv
filesize	varchar(100)	ขนาดไฟล์	68 B
datedown	timestamp	วัน/เดือน/ปี	2022-09-05 20:59:14

ตารางที่ 14 บทที่ 3 แสดงรายละเอียดของตาราง uploadfile

ชื่อตาราง : uploadfile ประเภทตาราง : Master คำอธิบาย : แฟ้มข้อมูลการจัดเก็บไฟล์ คีย์หลัก : fileID			
เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
fileID	int(11)	ลำดับไฟล์	98
file_username	varchar(100)	ชื่อผู้อัพโหลดไฟล์	Admin
fileupload	varchar(100)	ชื่อชุดข้อมูล	แบบฟอร์ม.csv
filesize	varchar(100)	ขนาดไฟล์	68 B
dateup	timestamp	วัน/เดือน/ปี	2022-09-05 20:59:14
approve	bit(1)	ไฟล์อนุมัติ	1

3.4 ความสัมพันธ์ของข้อมูล (ER-Diagram)



ภาพที่ 7 บทที่ 3 แสดงภาพความสัมพันธ์ของข้อมูลการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลการใช้ไฟฟ้าในบ้านที่อยู่อาศัย ในปี 2564-2565

1.)ตาราง users ความสัมพันธ์ของข้อมูลแบบ One – to – Many กับตาราง Data คือ ผู้ใช้งานหนึ่งคนสามารถเพิ่มข้อมูลการใช้ไฟฟ้าได้หลายข้อมูลการใช้ไฟฟ้า และแต่ละข้อมูลการใช้ไฟฟ้าจะมีผู้ใช้งานจัดการข้อมูลเพียงคนเดียวเท่านั้น

2.)ตาราง users ความสัมพันธ์ของข้อมูลแบบ One – to – Many กับตาราง downloadfile คือ ผู้ใช้งานหนึ่งคนสามารถดาวน์โหลดชุดข้อมูลได้หลายไฟล์ และแต่ละไฟล์ชุดมูลจะมีผู้ใช้งานดาวน์โหลดชุดข้อมูลเพียงคนเดียวเท่านั้น

3.)ตาราง users ความสัมพันธ์ของข้อมูลแบบ One – to – Many กับตาราง uploadfile คือ ผู้ใช้งานหนึ่งคนสามารถอัปโหลดชุดข้อมูลได้หลายไฟล์ และแต่ละไฟล์ชุดมูลจะมีผู้ใช้งานอัปโหลดชุดข้อมูลเพียงคนเดียวเท่านั้น

คำอธิบายกระบวนการ

ตารางที่ 15 บทที่ 3 DFD Number 1 ตรวจสอบการเข้าสู่ระบบ

Process Description	
System	ระบบการข้อมูลสารสนเทศการใช้ไฟฟ้า
DFD Number	1.0
Process Name	ตรวจสอบการเข้าสู่ระบบ
Input Data Flow	ข้อมูลชื่อผู้ดูแลระบบและรหัสผ่าน, ข้อมูลชื่อผู้สมาชิกและรหัสผ่าน
Output Data Flow	ข้อมูลสิทธิการเข้าระบบ
Data Store Used	แฟ้มข้อมูลผู้ใช้งาน
Description	เป็นกระบวนการตรวจสอบ ข้อมูลและรหัสผ่านข้อผู้ใช้ สำหรับ ผู้ดูแลเว็บไซต์ และสมาชิก

ตารางที่ 16 บทที่ 3 DFD Number 2 จัดการข้อมูลส่วนตัว

Process Description	
System	ระบบการข้อมูลสารสนเทศการใช้ไฟฟ้า
DFD Number	2.0
Process Name	จัดการข้อมูลส่วนตัว
Input Data Flow	ข้อมูลจัดการข้อมูลผู้ใช้ระบบ, ข้อมูลการแก้ไขข้อมูลส่วนตัว
Output Data Flow	ข้อมูลผู้ใช้ระบบ, ข้อมูลส่วนตัว
Data Store Used	แฟ้มข้อมูลผู้ใช้งาน
Description	เป็นการจัดการข้อมูลส่วนตัวของผู้ใช้ในระบบ

ตารางที่ 17 บทที่ 3 DFD Number 3 จัดการข้อมูลหน้าเว็บไซต์

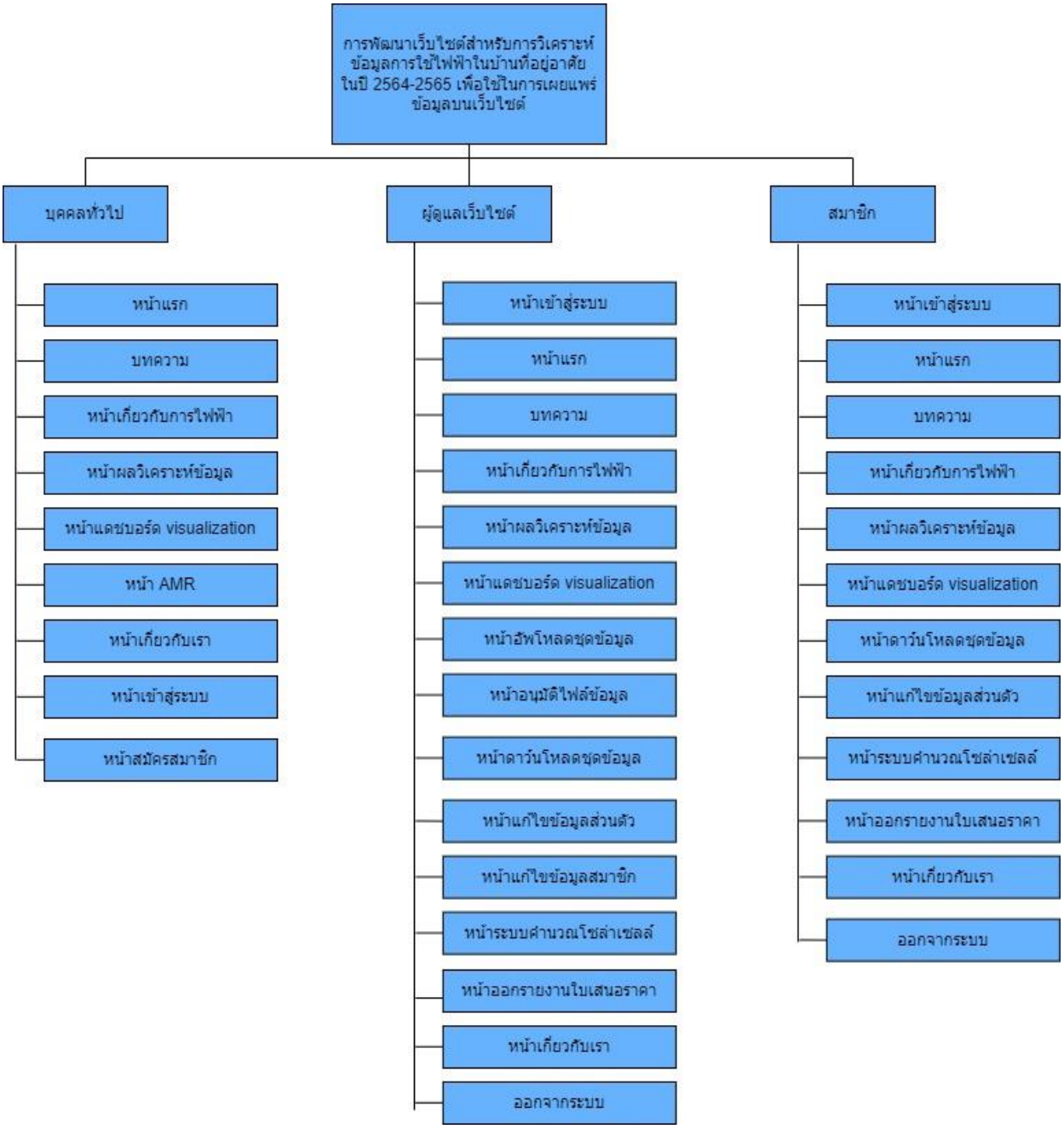
Process Description	
System	ระบบการข้อมูลสารสนเทศการใช้ไฟฟ้า
DFD Number	3.0
Process Name	จัดการข้อมูลหน้าเว็บไซต์
Input Data Flow	ข้อมูลจัดการข้อมูลการอัปโหลด, ข้อมูลจัดการข้อมูลการอัปโหลด, ข้อมูลการ จัดการใช้ไฟฟ้า, ข้อมูลการอัปโหลดชุดข้อมูล
Output Data Flow	ข้อมูลการดาวน์โหลดชุดข้อมูลของสมาชิก, ข้อมูลการอัปโหลดชุดข้อมูล
Data Store Used	แฟ้มข้อมูลการดาวน์โหลดไฟล์, แฟ้มข้อมูลการจัดเก็บไฟล์
Description	เป็นกระบวนการจัดการหน้าเว็บไซต์ ของผู้ดูแลเว็บไซต์ และสมาชิกเข้าใช้งานเว็บไซต์ โดยจะมีผู้ดูแลเว็บไซต์ที่สามารถแก้ไขอัพเดตข้อมูลสารสนเทศภายในเว็บไซต์ได้

ตารางที่ 18 บทที่ 3 DFD Number 4 รายงานข้อมูลสารสนเทศ ของเว็บไซต์

Process Description	
System	ระบบการข้อมูลสารสนเทศการใช้ไฟฟ้า
DFD Number	4.0
Process Name	แฟ้มข้อมูลการใช้ไฟฟ้าบนเว็บไซต์, แฟ้มข้อมูลการดาวน์โหลดไฟล์
Input Data Flow	ข้อมูลสารสนเทศการไฟฟ้าประเทศไทย, ข้อมูลสารสนเทศกระบวนการวิเคราะห์ข้อมูล,ข้อมูลดาวน์โหลดชุดข้อมูล, ข้อมูลสารสนเทศการไฟฟ้าบนเว็บไซต์, ข้อมูลแดชบอร์ด
Output Data Flow	ข้อมูลการดาวน์โหลดชุดข้อมูลของสมาชิก, ข้อมูลการอัปโหลดชุดข้อมูล
Data Store Used	แฟ้มข้อมูลการใช้ไฟฟ้าบนเว็บไซต์, แฟ้มข้อมูลการดาวน์โหลดไฟล์
Description	เป็นกระบวนการแสดงผลลัพธ์แดชบอร์ดเกี่ยวกับไฟฟ้าให้กับเว็บไซต์ โดยแสดงผลในแบบ data visualization โดยสิทธิสมาชิกสามารถดาวน์โหลดข้อมูลได้

3.5 ออกแบบหน้าเว็บไซต์

โครงสร้างระบบการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลการใช้ไฟฟ้าในบ้านที่อยู่อาศัย ในปี 2564-2565 เพื่อใช้ในการเผยแพร่ข้อมูลบนเว็บไซต์



ภาพที่ 78 บทที่ 3 แสดงโครงสร้างระบบการพัฒนาเว็บไซต์สำหรับการวิเคราะห์ข้อมูลการใช้ไฟฟ้าในบ้านที่อยู่อาศัยในปี 2564-2565 เพื่อใช้ในการเผยแพร่ข้อมูลบนเว็บไซต์

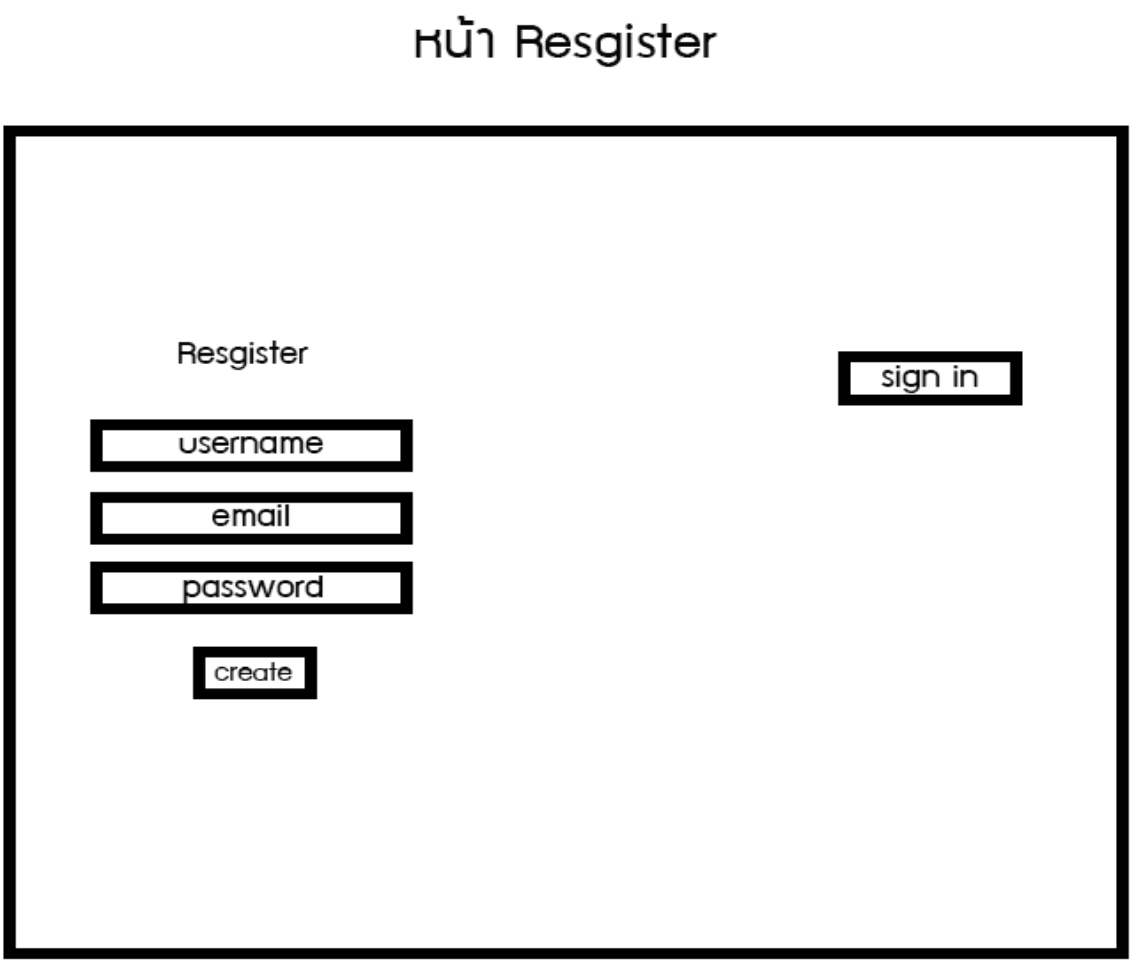
การออกแบบ Wireframe หน้าจอเว็บไซต์มีการออกแบบโดยแบ่งตามประเภทของผู้ใช้งานซึ่งประกอบด้วย 3 ผู้ใช้งาน ดังนี้

3.5.1.) บุคคลทั่วไป

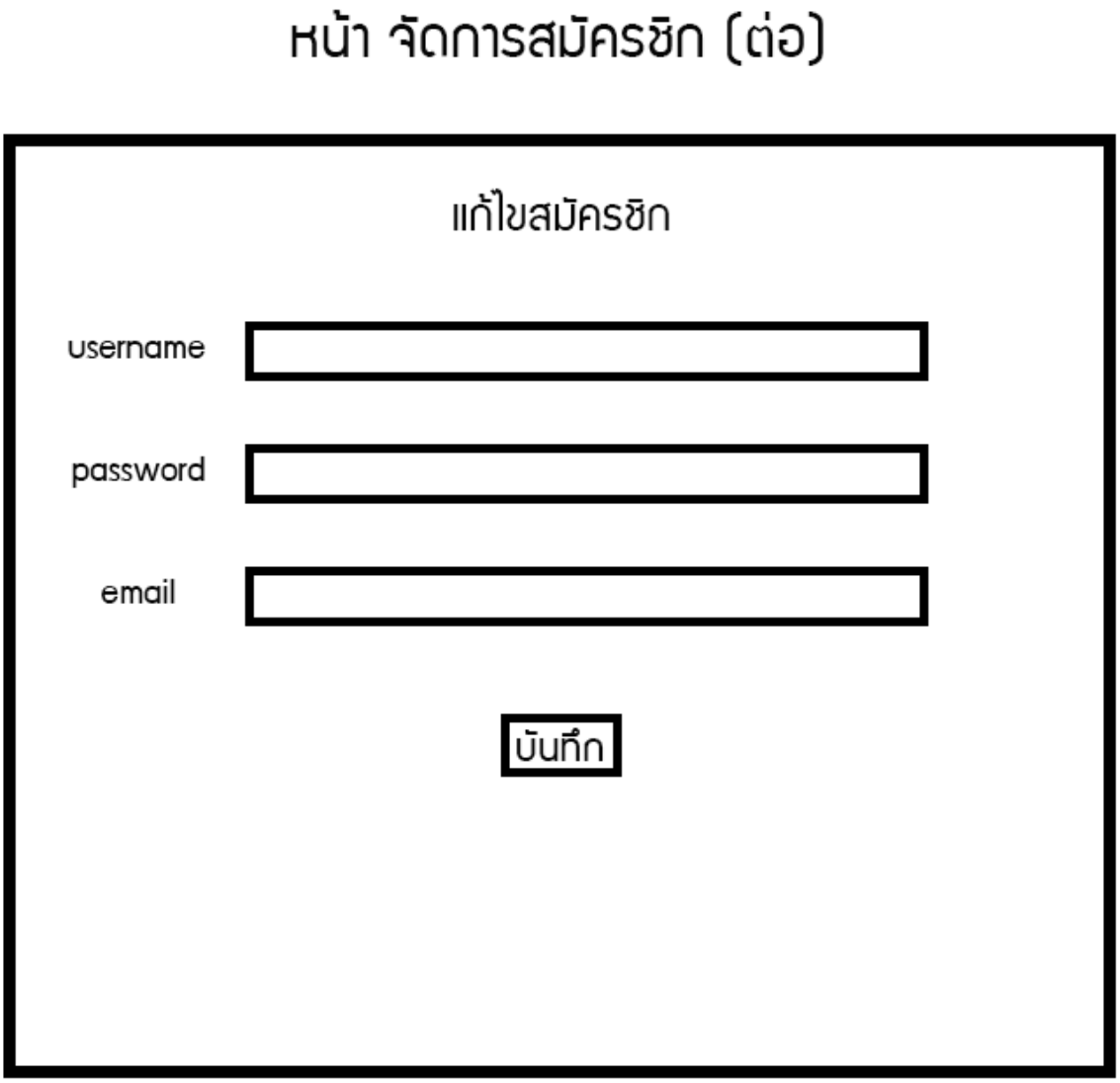
3.5.2.) สมาชิก

3.5.3.) ผู้ดูแลเว็บไซต์

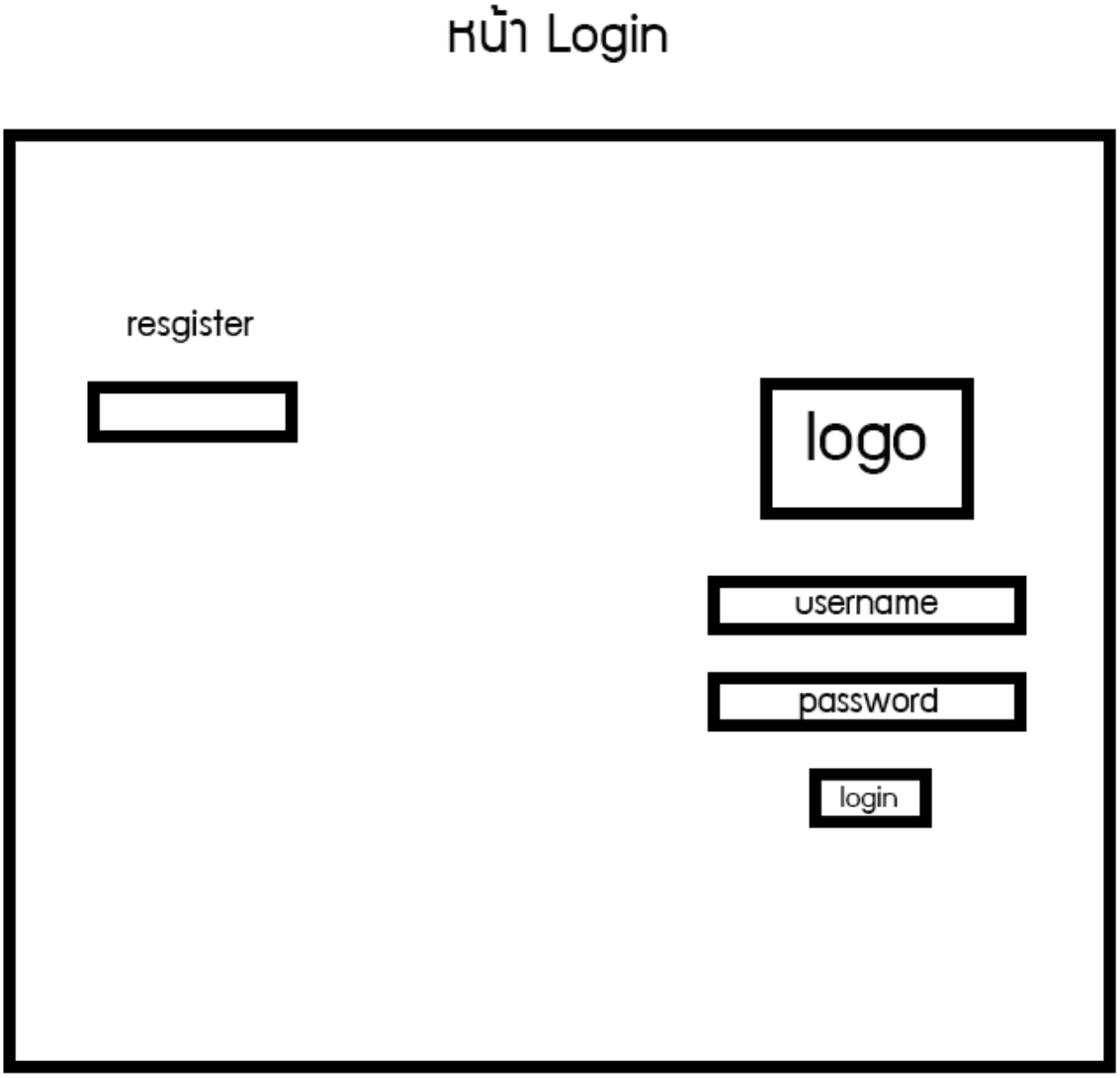
ภาพที่ 8 บทที่ 3 หน้าจอโฮมแพจ



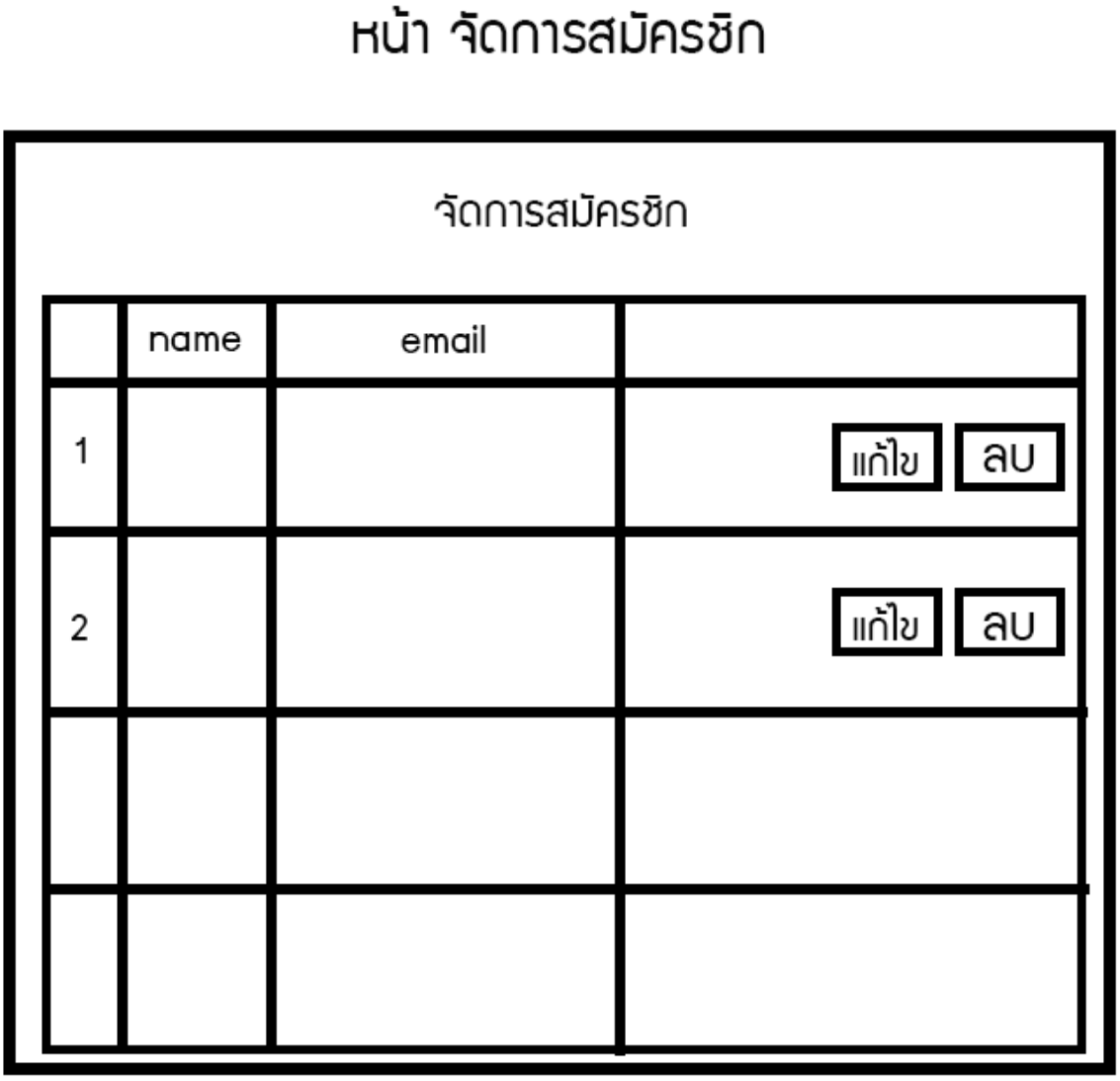
ภาพที่ 80บทที่ 3 หน้าจอสมัครสมาชิก



ภาพที่ 9 บทที่ 3 หน้าจอสมัครสมาชิก

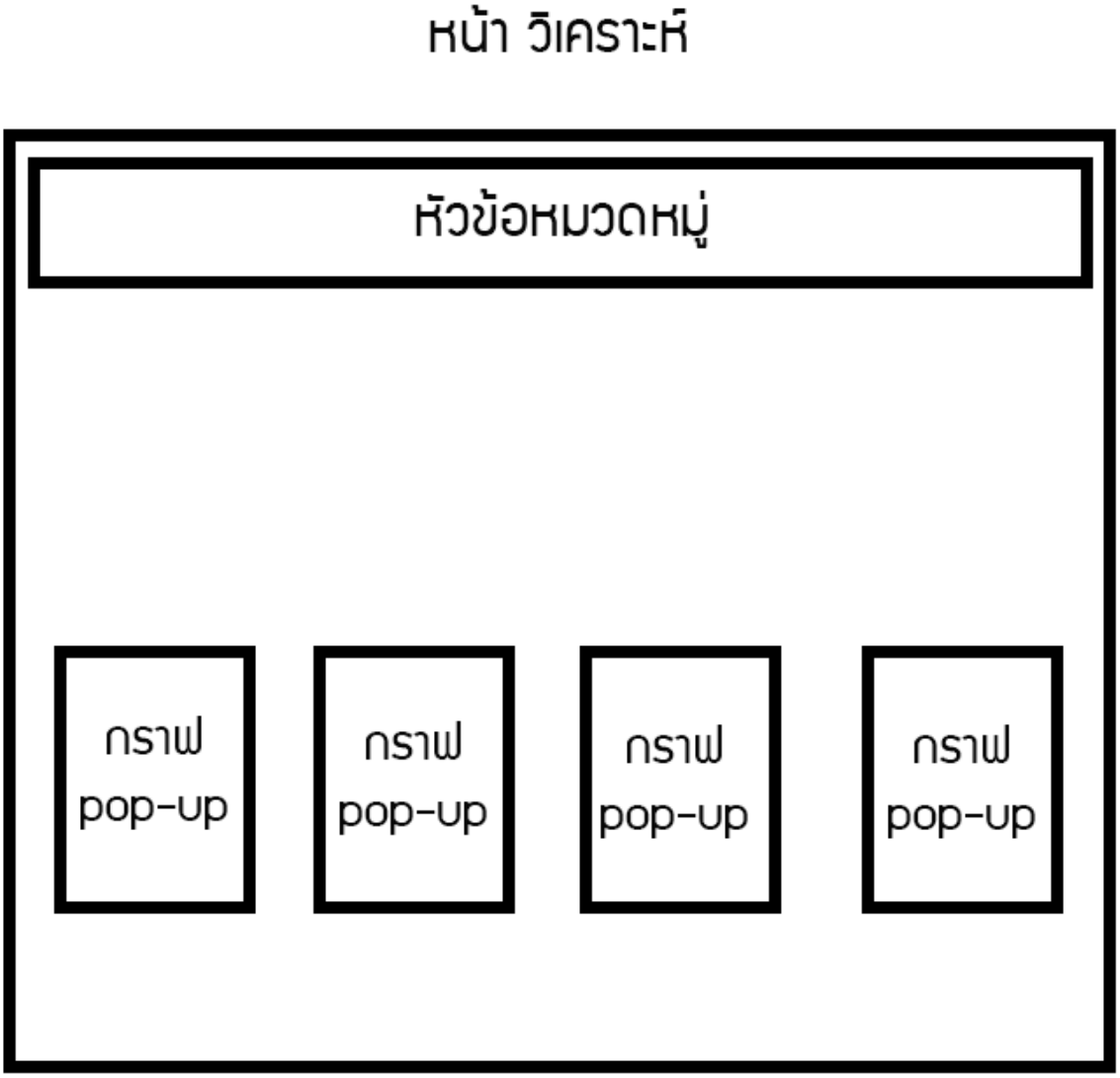


ภาพที่ 10 บทที่ 3 หน้าจอล็อกอิน

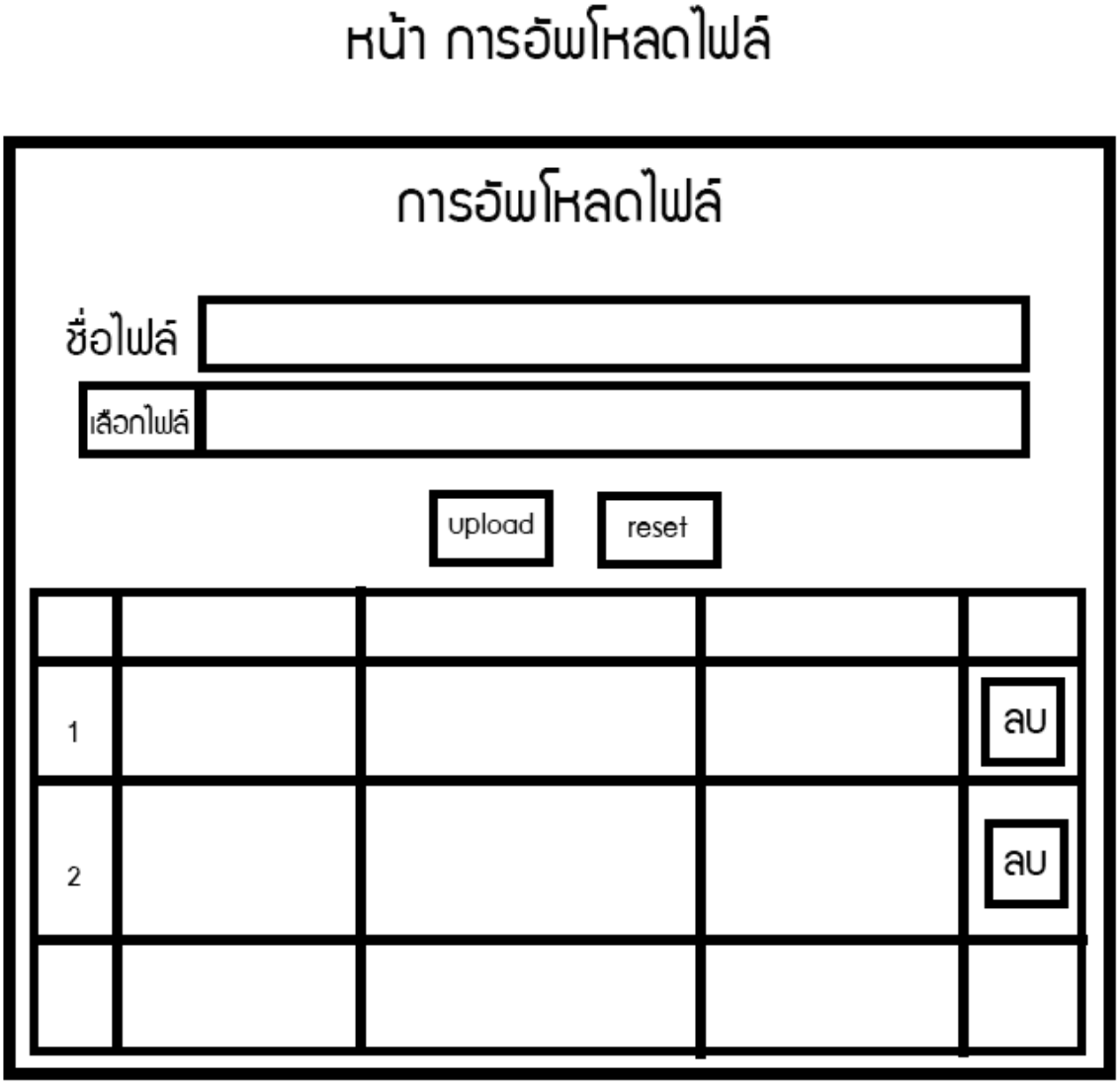


ภาพที่ 83 บทที่ 3 หน้าจอจัดการสมาชิก

ภาพที่ 84 บทที่ 3 หน้าจอแสดงภาพ



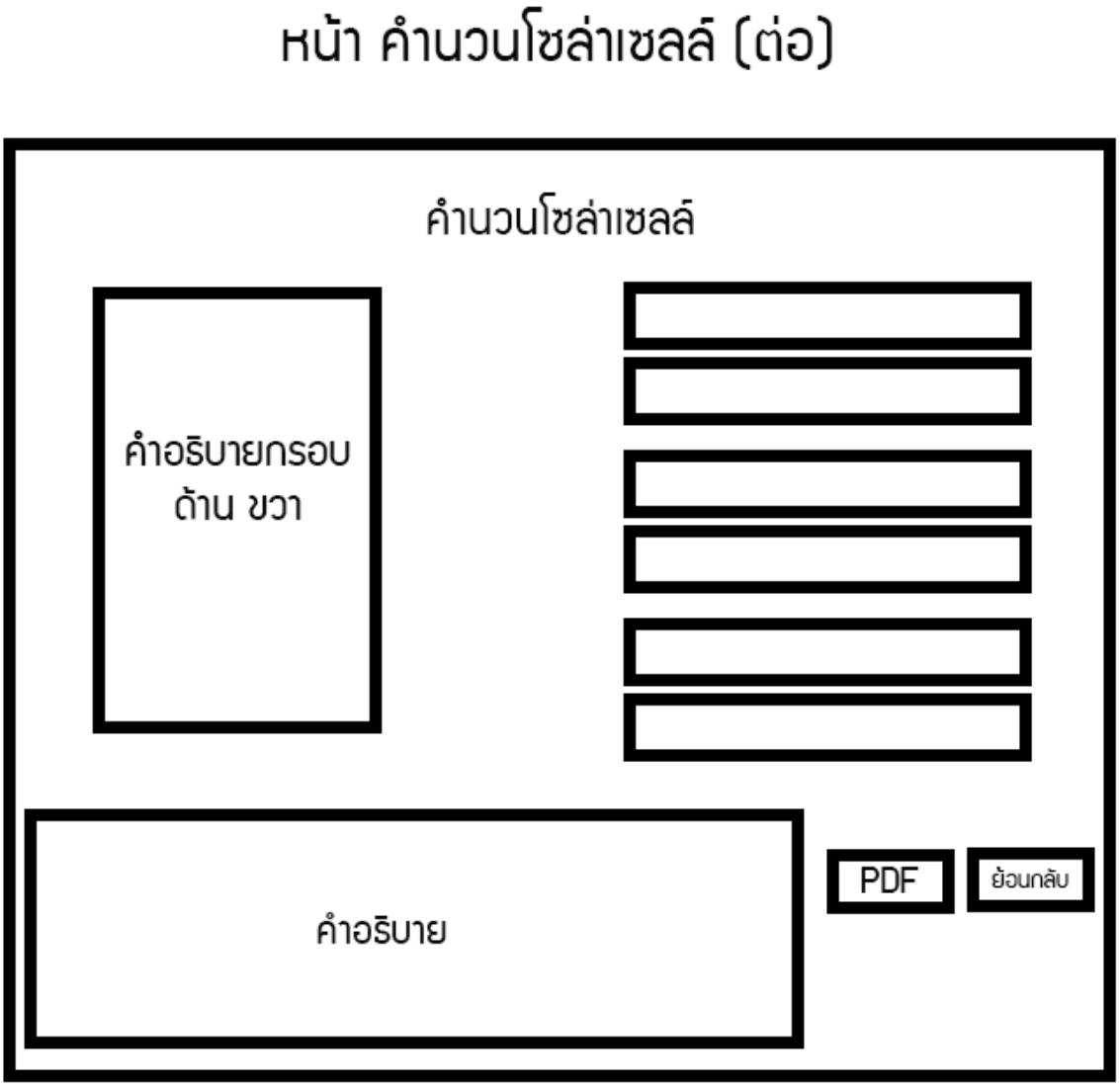
ภาพที่ 85 บทที่ 3 หน้าจอวิเคราะห์ข้อมูล



ภาพที่ 86 บทที่ 3 หน้าจอการอัพโหลดข้อมูล

ภาพที่ 87 บทที่ 3 หน้าจอดาว์นโหลดข้อมูล

ภาพที่ 88 บทที่ 3 หน้าจอระบบคำนวณโซล่าเซลล์



ภาพที่ 89 บทที่ 3 หน้าจอแสดงข้อมูลการคำนวณโซล่าเซลล์

3.6 บทสรุป

ระบบของเว็บไซต์วิเคราะห์ข้อมูลการใช้ไฟฟ้าในที่อยู่อาศัยในปี 2564-2565 ทางผู้จัดทำได้ทำการแสดงข้อมูลการใช้ไฟฟ้าโดยนำจังหวัดที่มีความแตกต่างในการใช้ไฟฟ้ามาทั้งหมด 3 จังหวัด 1.จังหวัดเชียงใหม่ 2.แพร่ 3.น่าน และมีปะเภทของที่อยู่อาศัยที่แตกต่างประกอบด้วย 4 ประเภท 1.บ้าน 2.ที่อยู่อาศัยที่ได้รับการยกเว้นค่าไฟ 3.เกษตรกรรม 4.ธุรกิจขนาดเล็ก เพื่อได้ทราบถึงการใช้ไฟฟ้าของแต่ละประเภท เพื่อนำไปใช้ประโยชน์ในด้านต่างๆ เช่น การวิเคราะห์ขนาดมิเตอร์ไฟฟ้าให้เหมาะสมกับการใช้งานทางผู้จัดทำได้เพิ่มระบบการวิเคราะห์โซล่าเซลล์ลงไปด้วย นั้นเพื่อได้รับประโยชน์ให้ได้มากที่สุดระบบโซล่าเซลล์จะสามารถคำนวณผ่านบิลค่าไฟฟ้าที่ได้รับในแต่ละเดือน และยังมีหน้า AMR คือการอ่านค่าผ่านระบบมิเตอร์อัตโนมัติและนำค่าหน่วยไฟฟ้ามาคำนวณโซล่าเซลล์โดยที่ไม่ต้องรอบิลค่าไฟฟ้าในแต่ละเดือนยังทราบถึงพฤติกรรมการใช้ไฟฟ้าในแต่ละเดือน วัน ปี สามารถนำข้อมูลที่ได้มาปรับเปลี่ยนวิธีการใช้ไฟฟ้าเพื่อให้เกิดประโยชน์สูงสุด ทางผู้จัดทำได้นำข้อมูลสารสนเทศมาทำการแสดงผลออกทางหน้าจอ ของเว็บไซต์ที่จะเผยแพร่บน Web browser ใช้ Adobe XD ในการออกแบบหน้าต่างรายละเอียดของเว็บไซต์ และใช้ Visual Studio Code ในการเขียนพัฒนาเว็บไซต์