

บทที่ 4

ผลการวิจัย

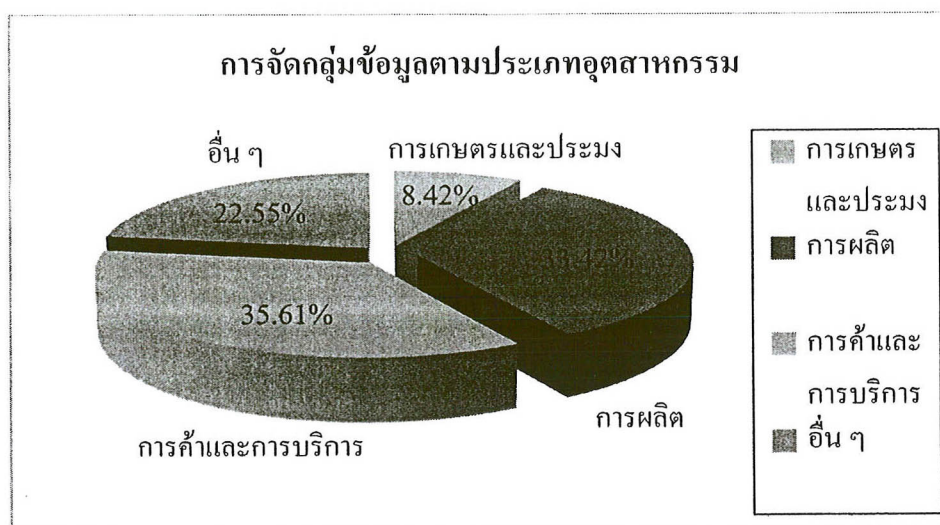
การทดลองเมื่อทำการคัดเลือกข้อมูลที่ต้องการและจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมในการทำเหมืองข้อมูลแล้ว จากนั้นนำข้อมูลที่จัดเตรียมมาทำการจัดกลุ่มข้อมูลตามประเภทอุตสาหกรรม แล้วนำผลที่ได้จากการจัดกลุ่มไปทำการสร้างแบบจำลองการจำแนกประเภทข้อมูล โดยใช้อัลกอริทึม C4.5 ซึ่งแสดงในรูปแบบต้นไม้การตัดสินใจ และทำการแปลงต้นไม้การตัดสินใจไปเป็นกฎ เพื่อศึกษาปัจจัยที่มีผลต่อรายได้ของประชากร สรุปผลการจัดกลุ่มข้อมูลและการจำแนกประเภทข้อมูลได้ดังต่อไปนี้

4.1 การจัดกลุ่มข้อมูลการสำรวจด้านการทำงานของประชากรตามประเภทอุตสาหกรรม

ขั้นตอนนี้จะทำการจัดกลุ่มข้อมูลการสำรวจด้านการทำงานของประชากรตามประเภทอุตสาหกรรม ซึ่งมีข้อมูลนำเข้าจำนวน 193,176 รายการ ในระยะเวลา 1 ปี มาจัดกลุ่ม โดยแบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 ประเภทเกษตรและประมง กลุ่มที่ 2 ประเภทการผลิต กลุ่มที่ 3 ประเภทการค้าและการบริการ กลุ่มที่ 4 ประเภทอื่น ๆ ได้ผลการจัดกลุ่มดังต่อไปนี้

ตารางที่ 4.1 ผลการจัดกลุ่มข้อมูลตามประเภทอุตสาหกรรม

ประเภทอุตสาหกรรม	จำนวนรายการ	ร้อยละ
การเกษตรและประมง	16,273	8.42
การผลิต	64,561	33.42
การค้าและการบริการ	68,790	35.61
อื่น ๆ (ข้าราชการ การศึกษา องค์กรระหว่างประเทศ เป็นต้น)	43,552	22.55



รูปที่ 4.1 แสดงผลการจัดกลุ่มตามประเภทอุตสาหกรรม

จากรูปที่ 4.1 แสดงให้เห็นว่ากลุ่มอุตสาหกรรมที่มีจำนวนรายการมากที่สุดคือ กลุ่มการค้าและการบริการ คิดเป็นร้อยละ 35.61 รองลงมาคือ กลุ่มการผลิต และกลุ่มอื่น ๆ คิดเป็นร้อยละ 33.42 และ 22.55 ตามลำดับ ส่วนกลุ่มการเกษตรและประมง มีจำนวนน้อยที่สุด คิดเป็นร้อยละ 8.42

ตารางที่ 4.2 แสดงเป็นจำนวนสมาชิกในการจัดกลุ่มข้อมูลตามประเภทอุตสาหกรรม

ปัจจัยที่ใช้วิเคราะห์		กลุ่มที่			
		1. การเกษตรและ ประมง	2. การผลิต	3. การค้าและ การบริการ	4. อื่น ๆ
		รวม	16,273	64,561	68,790
REG	กรุงเทพมหานคร	72	5,734	9,910	1,982
	ภาคกลาง	6,284	35,228	22,889	11,835
	ภาคเหนือ	5,116	9,683	12,343	9,509
	ภาคตะวันออกเฉียงเหนือ	3,298	7,629	11,629	12,781
	ภาคใต้	1,503	6,287	12,019	7,445

ตารางที่ 4.2 (ต่อ)

ปัจจัยที่ใช้วิเคราะห์		กลุ่มที่			
		1. การเกษตรและ ประมง	2. การผลิต	3. การค้าและ การบริการ	4. อื่น ๆ
		รวม 16,273	64,561	68,790	43,552
AREA	เทศบาล	5,439	37,907	53,023	33,804
	นอกเขตเทศบาล	10,834	26,654	15,767	9,748
SEX	เพศชาย	9,392	39,929	32,144	22,725
	เพศหญิง	6,881	24,632	36,646	20,827
AGE	15-24 ปี	1,835	9,237	10,399	1,417
	25-34 ปี	2,936	21,491	22,886	9,892
	35-49 ปี	6,562	25,576	26,055	19,296
	ตั้งแต่ 50 ปีขึ้นไป	4,940	8,257	9,450	12,947
MARITAL	โสด	2,875	15,863	21,423	9,362
	สมรส	11,244	44,012	40,934	31,246
	อื่น ๆ	2,154	4,686	6,433	2,944
EDUCATION	ต่ำกว่ามัธยมศึกษา	13,007	30,587	23,475	3,525
	ระดับมัธยมศึกษา	2,767	23,465	23,535	8,914
	ระดับอุดมศึกษา	499	10,509	21,780	31,113
OCCUP	กลุ่ม 1	43	977	2,161	4,199
	กลุ่ม 2	313	6,746	12,266	25,031
	กลุ่ม 3	830	21,746	33,932	10,816
	กลุ่ม 4	15,087	35,092	20,431	3,506
WAGE_TYPE	รายชั่วโมง รายวัน รายสัปดาห์	13,645	37,994	16,478	762
	รายเดือน	2,628	26,567	52,312	42,790
TOT_HR	1-34 ชั่วโมง	3,746	4,974	4,544	4,391
	35-45 ชั่วโมง	4,092	10,008	19,372	33,295
	มากกว่า 45 ชั่วโมง	8,435	49,579	44,874	5,866

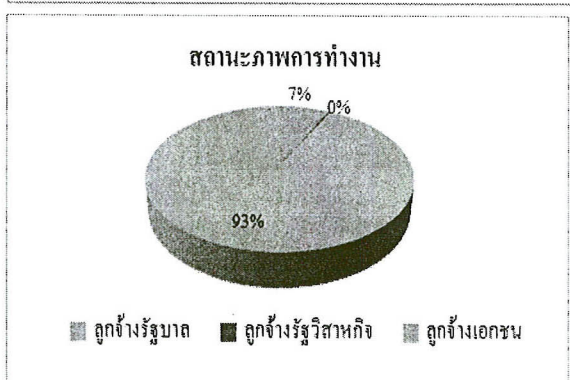
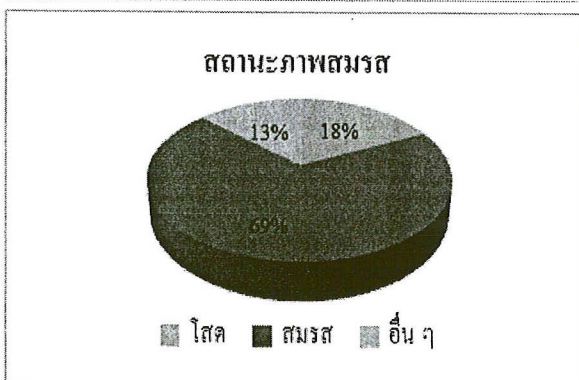
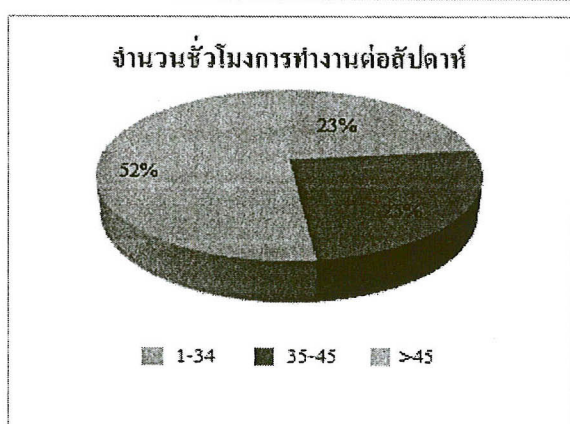
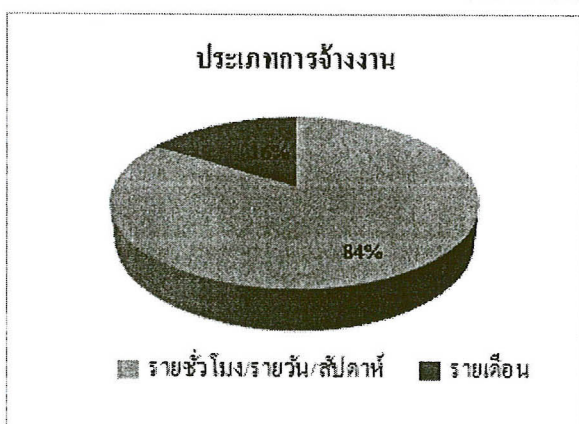
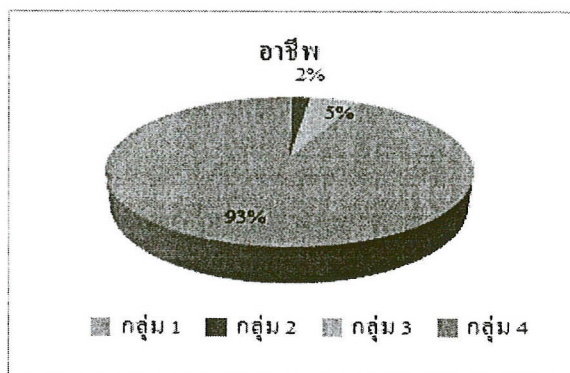
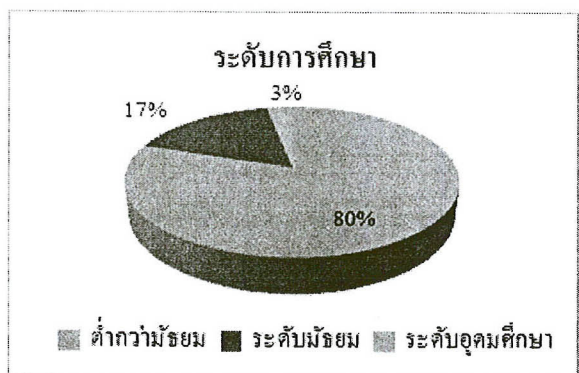
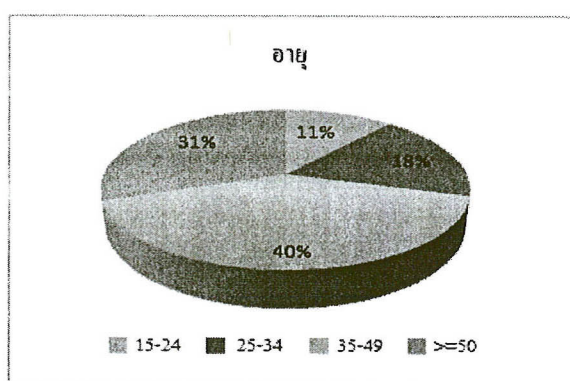
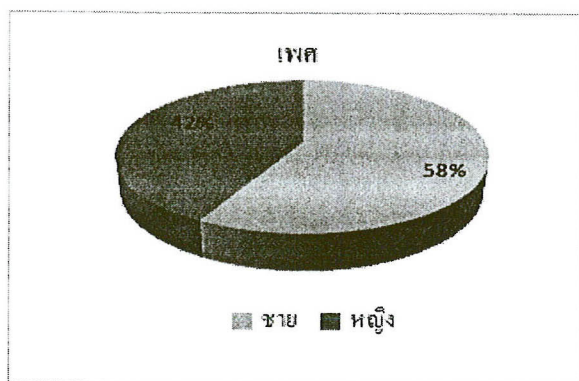
ตารางที่ 4.2 (ต่อ)

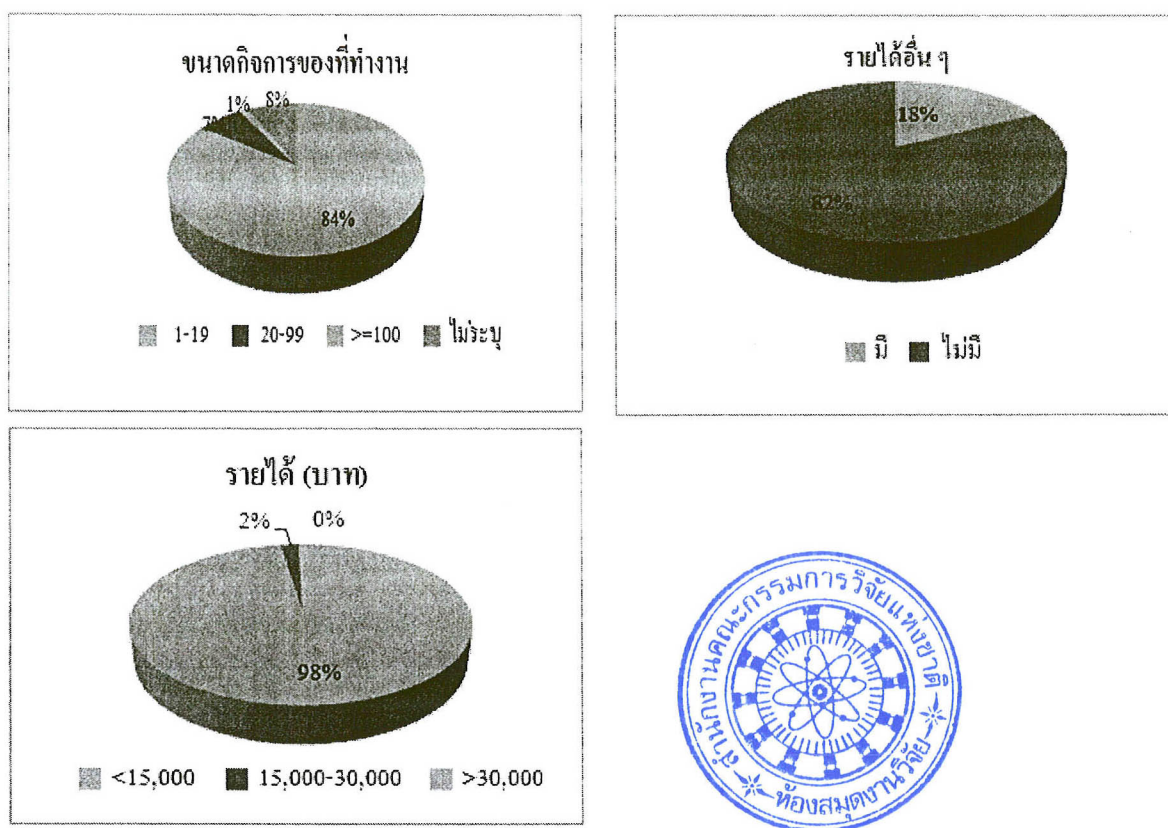
ปัจจัยที่ใช้วิเคราะห์		กลุ่มที่				
		1.การเกษตรและ	2.การผลิต	3.การค้าและ	4.อื่น ๆ	
		ประมง		การบริการ		
		รวม	16,273	64,561	68,790	43,552
STATUS	ลูกจ้างรัฐบาล	1,160	1,081	11,780	40,365	
	ลูกจ้างรัฐวิสาหกิจ	46	1,784	3,018	289	
	ลูกจ้างเอกชน	15,067	61,696	53,992	2,898	
SIZE	1-19 คน	13,737	23,940	33,495	816	
	20-99 คน	1,112	8,538	9,931	1,629	
	ตั้งแต่ 100 คนขึ้นไป	214	29,210	10,453	453	
	ไม่ระบุ	1,210	2,873	14,811	40,654	
OTHER_INCOME	มี	2,864	23,359	27,606	23,234	
	ไม่มี	13,409	41,202	41,184	20,318	
INCOME	น้อยกว่า 15,000 บาท	15,895	57,787	55,230	21,074	
	15,000-30,000 บาท	344	5,126	10,107	13,548	
	มากกว่า 30,000 บาท	34	1,648	3,453	8,930	

จากตารางที่ 4.2 แสดงให้เห็นถึงข้อมูลการสำรวจด้านการทำงานของประชากร ในแต่ละปัจจัยต่างๆ ทั้งหมด 13 ปัจจัย โดยในแต่ละกลุ่มนั้นจะมีลักษณะแตกต่างกันไปดังต่อไปนี้

4.1.1 แอตทริบิวต์ของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 1 การเกษตรและประมง





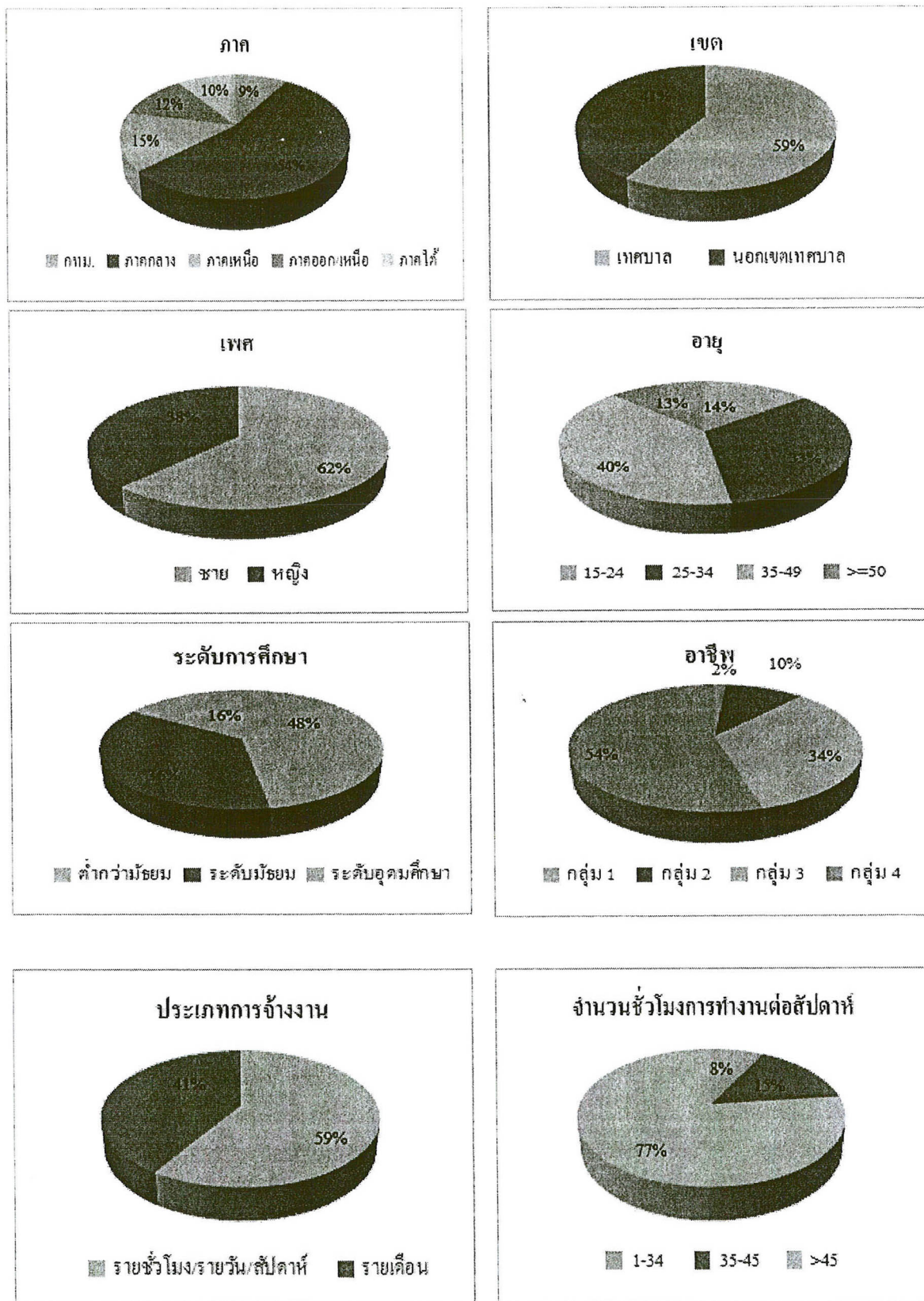


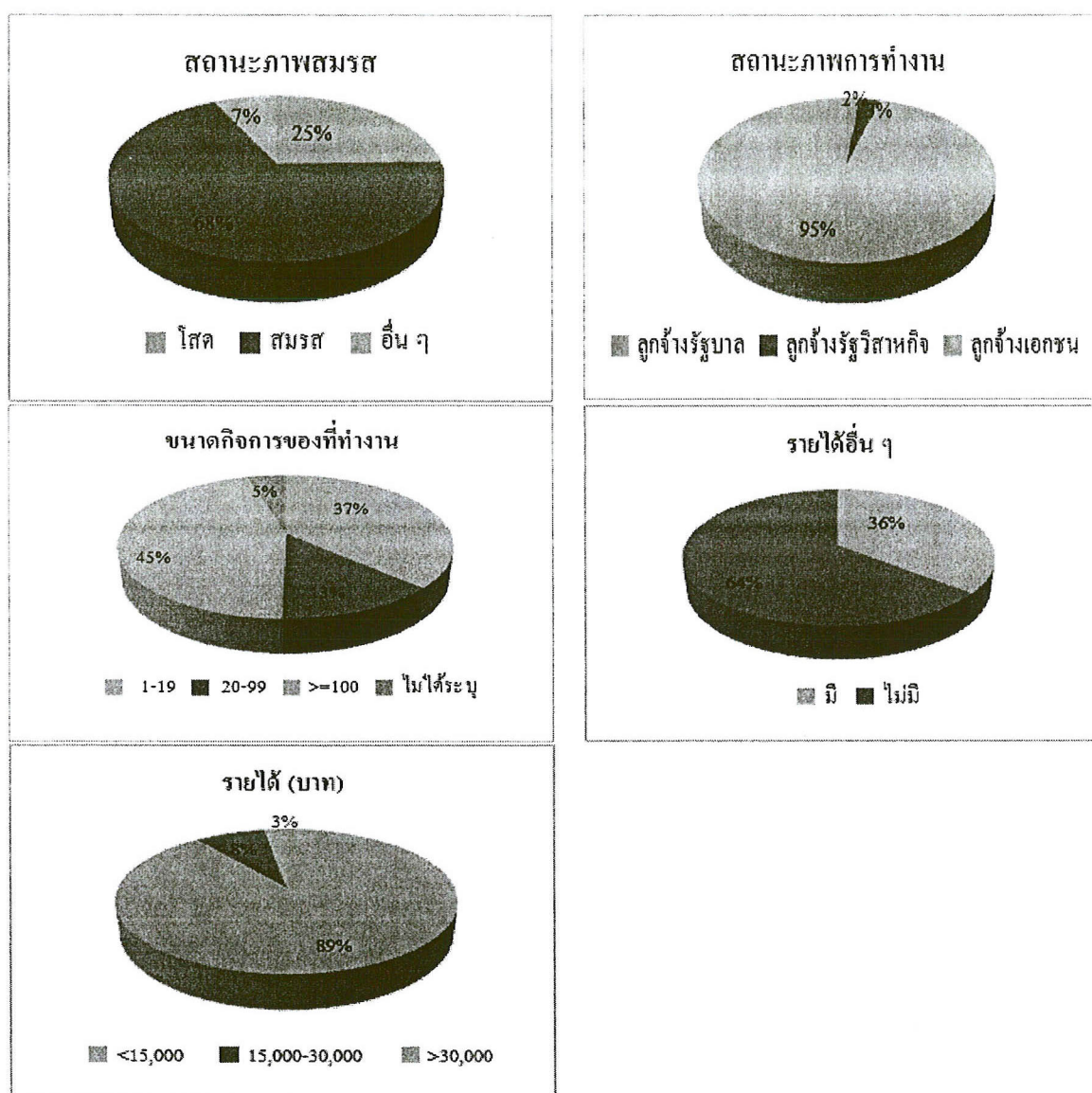
รูปที่ 4.2 แสดงแอตทริบิวต์ของข้อมูลในกลุ่มที่ 1 การเกษตรและประมง

จำนวนรายการทั้งหมดในกลุ่มที่ 1 นี้ มีจำนวนทั้งสิ้น 16,273 รายการ หรือคิดเป็นร้อยละ 8.42 ของจำนวนรายการทั้งหมด ลักษณะโดยทั่วไปของข้อมูลในกลุ่มมีประชากรอาศัยอยู่นอกเขตเทศบาล คิดเป็นร้อยละ 67 ส่วนใหญ่จะเป็นเพศชายคิดเป็นร้อยละ 58 มีอายุอยู่ในช่วง 35-49 ปีขึ้นไป คิดเป็นร้อยละ 40 มีการศึกษาอยู่ในต่ำกว่าระดับมัธยมศึกษาคิดเป็นร้อยละ 80 รองลงมาคือระดับมัธยมศึกษาคิดเป็นร้อยละ 17 มีอาชีพอยู่ในกลุ่มที่ 4 คือผู้ปฏิบัติงานที่มีฝีมือในด้านการเกษตรและการประมง ผู้ปฏิบัติงานในธุรกิจด้านความสามารถทางฝีมือและธุรกิจอื่นๆที่เกี่ยวข้อง ผู้ปฏิบัติการเครื่องจักรโรงงานและเครื่องจักร และผู้ปฏิบัติงานด้านการประกอบคิดเป็นร้อยละ 93 เป็นการจ้างงานแบบรายชั่วโมง รายวัน หรือ รายสัปดาห์ คิดเป็นร้อยละ 84 มีสถานะภาพการสมรสร้อยละ 69 เป็นลูกจ้างเอกชน ร้อยละ 93 ทำงานในบริษัทที่มีพนักงาน 1-19 คน คิดเป็นร้อยละ 84 มีจำนวนชั่วโมงการทำงานต่อสัปดาห์ค่อนข้างสูงคือมากกว่า 45 ชั่วโมงคิดเป็นร้อยละ 52 รองลงมาคือทำงานอยู่ระหว่าง 35-45 ชั่วโมงคิดเป็นร้อยละ 25 กลุ่มที่ 1 ไม่ได้มีรายได้อื่น ๆ นอกจากรายได้ประจำคิดเป็นร้อยละ 82 ช่วงรายได้ที่มีรายได้ได้น้อยกว่า 15,000 บาทต่อเดือนคิดเป็นร้อยละ 98

รายการในกลุ่มที่ 1 อยู่ในภาคกลางมากที่สุด มีจำนวน 6,284 รายการ คิดเป็นร้อยละ 39 รองลงมา คือ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ คิดเป็นร้อยละ 31 ร้อยละ 20 และร้อยละ 9 ตามลำดับ โดยพื้นที่ที่มีรายการน้อยที่สุดคือ กรุงเทพมหานคร คิดเป็นร้อยละ 1

4.1.2 แอตรวิพวคัของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 2 การผลิต





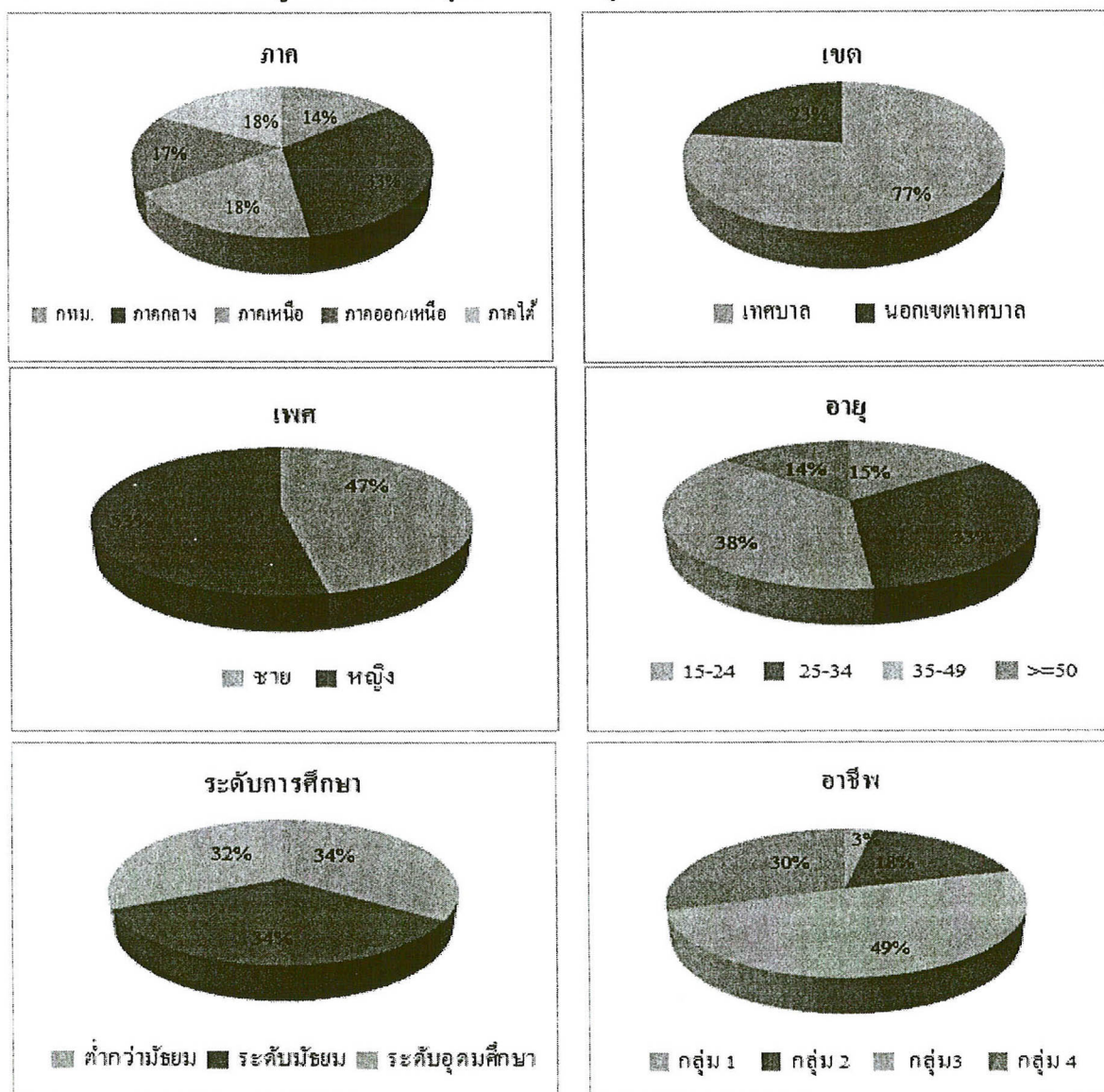
รูปที่ 4.3 แสดงแอตทริบิวต์ของข้อมูลในกลุ่มที่ 2 การผลิต

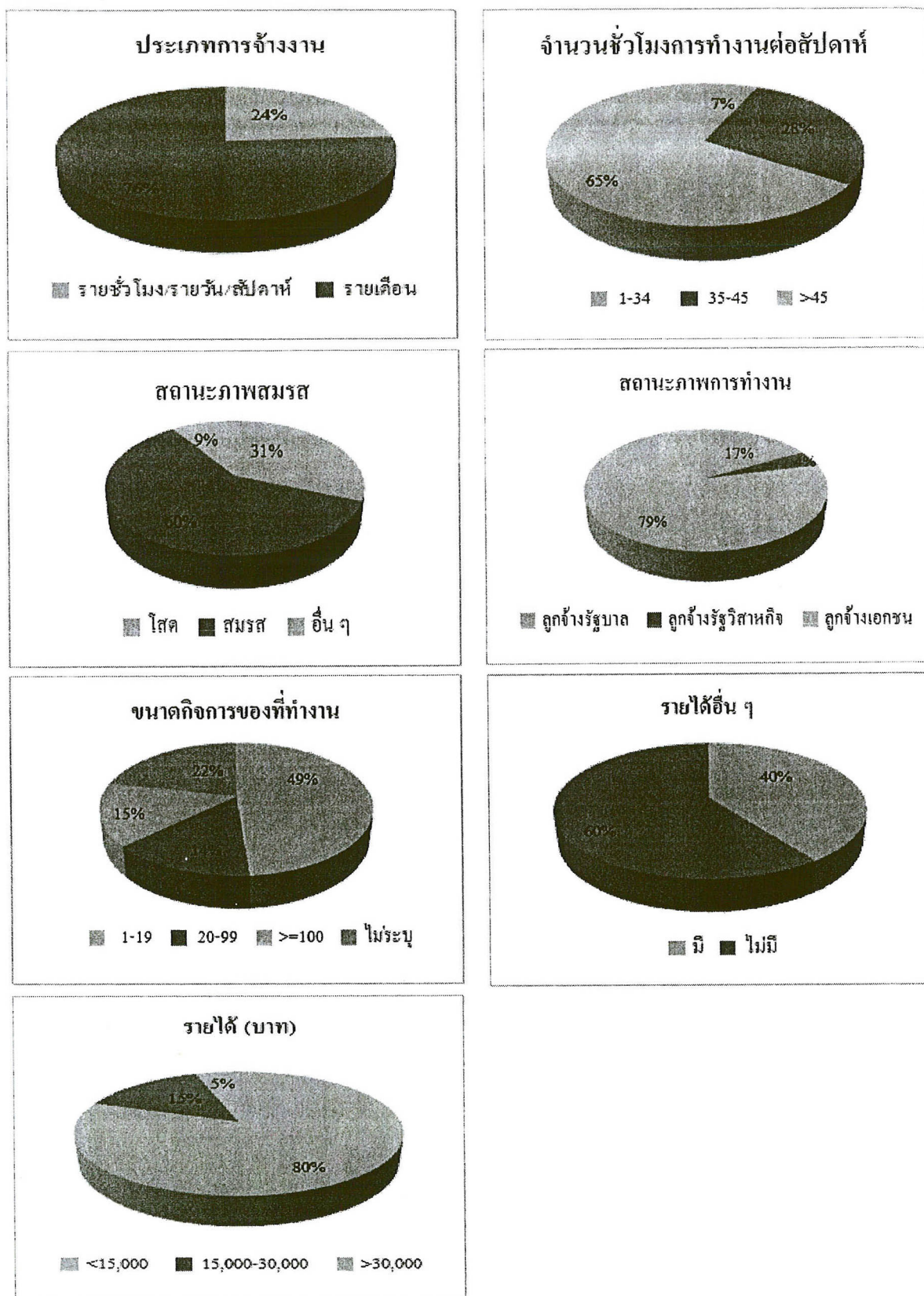
จำนวนรายการทั้งหมดในกลุ่มที่ 2 นี้ มีจำนวนทั้งสิ้น 64,561 รายการ หรือคิดเป็นร้อยละ 33.42 ของจำนวนรายการทั้งหมด ลักษณะโดยทั่วไปของข้อมูลในกลุ่มมีประชากรอาศัยอยู่ในเขตเทศบาลคิดเป็นร้อยละ 59 เป็นเพศชายคิดเป็นร้อยละ 62 มีอายุอยู่ในช่วง 35-49 ปี คิดเป็นร้อยละ 40 รองลงมาคือ 25-34 ปี คิดเป็นร้อยละ 33 สมรสแล้วคิดเป็นร้อยละ 68 ทำงานเป็นลูกจ้างเอกชนคิดเป็นร้อยละ 95 ทำงานในบริษัทที่มีพนักงานจำนวนมากกว่า 100 คนขึ้นไปร้อยละ 45 มีการศึกษาอยู่ในระดับต่ำกว่ามัธยมศึกษาคิดเป็นร้อยละ 48 รองลงมาคือระดับมัธยมศึกษาคิดเป็นร้อยละ 36 มีอาชีพอยู่ในกลุ่มที่ 4 คือ ผู้ปฏิบัติงานที่มีฝีมือในการเกษตรและการประมง ผู้ปฏิบัติงานในธุรกิจด้านความสามารถทางฝีมือ และธุรกิจอื่นๆที่เกี่ยวข้อง ผู้ปฏิบัติการเครื่องจักรโรงงานและเครื่องจักรและผู้ปฏิบัติงานด้านการประกอบ คิดเป็นร้อยละ 54 รองลงมาคืออาชีพกลุ่มที่ 3 คือ เสมียน พนักงานบริการ พนักงานขายใน

ร้านค้าและตลาด อาชีพชั้นพื้นฐานต่างๆ คิดเป็นร้อยละ 34 เป็นการจ้างงานแบบรายชั่วโมง รายวัน หรือรายสัปดาห์ คิดเป็นร้อยละ 59 มีจำนวนชั่วโมงการทำงานต่อสัปดาห์มากกว่า 45 ชั่วโมงคิดเป็นร้อยละ 77 รองลงมาคือทำงานอยู่ระหว่าง 35-45 ชั่วโมงคิดเป็นร้อยละ 15 กลุ่มที่ 2 ไม่ได้มีรายได้อื่น ๆ นอกจากรายได้ประจำคิดเป็นร้อยละ 64 ช่วงรายได้ที่มีรายได้ไม่น้อยกว่า 15,000 บาทต่อเดือนคิดเป็นร้อยละ 89

รายการที่อยู่ในกลุ่มที่ 2 อยู่ในภาคกลางมากที่สุดมีจำนวน 35,228 รายการ คิดเป็นร้อยละ 54 รองลงมาคือภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคใต้ คิดเป็นร้อยละ 15 ร้อยละ 12 และร้อยละ 10 ตามลำดับ โดยพื้นที่ที่มีรายการน้อยที่สุดคือกรุงเทพมหานคร คิดเป็นร้อยละ 9

4.1.3 แอตทริบิวต์ของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 3 การค้าและการบริการ



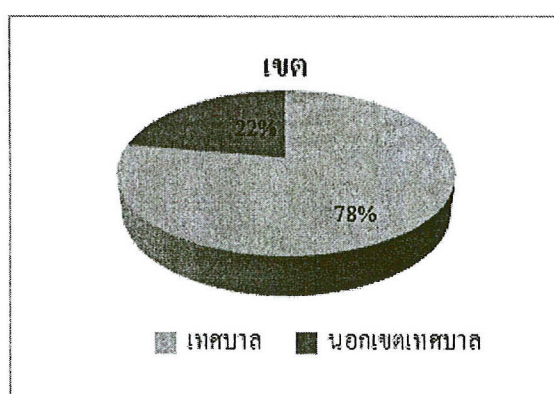
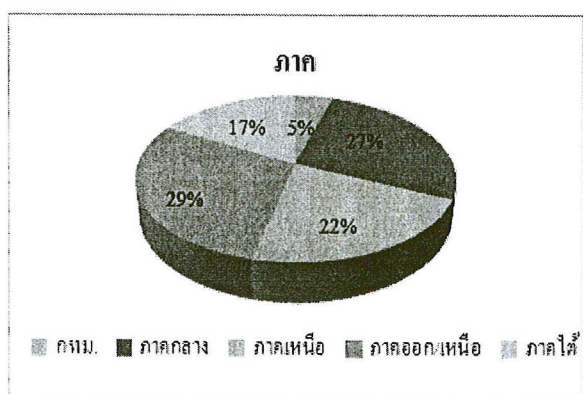


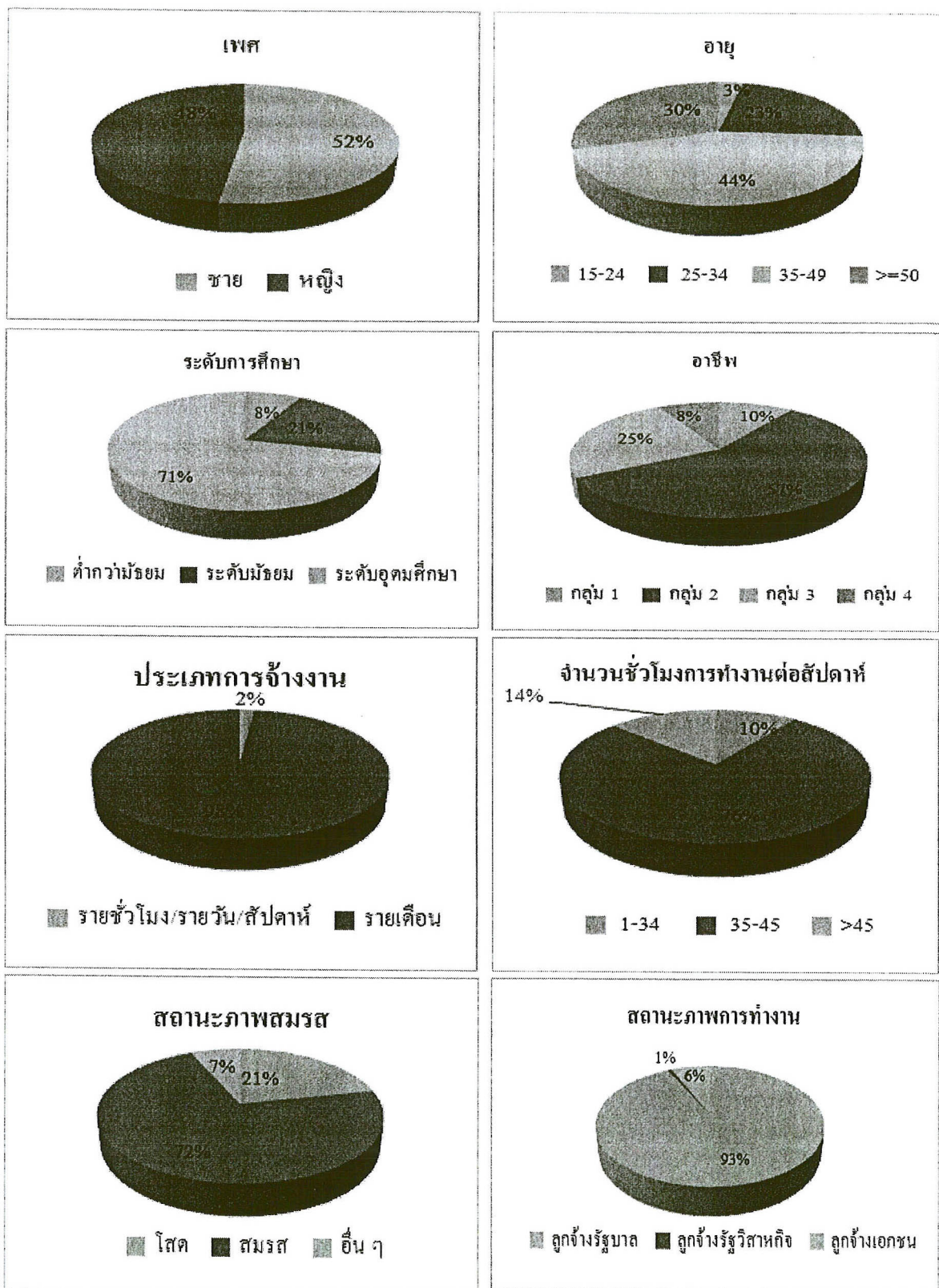
รูปที่ 4.4 แสดงแอตทริบิวต์ของข้อมูลในกลุ่มที่ 3 การค้าและการบริการ

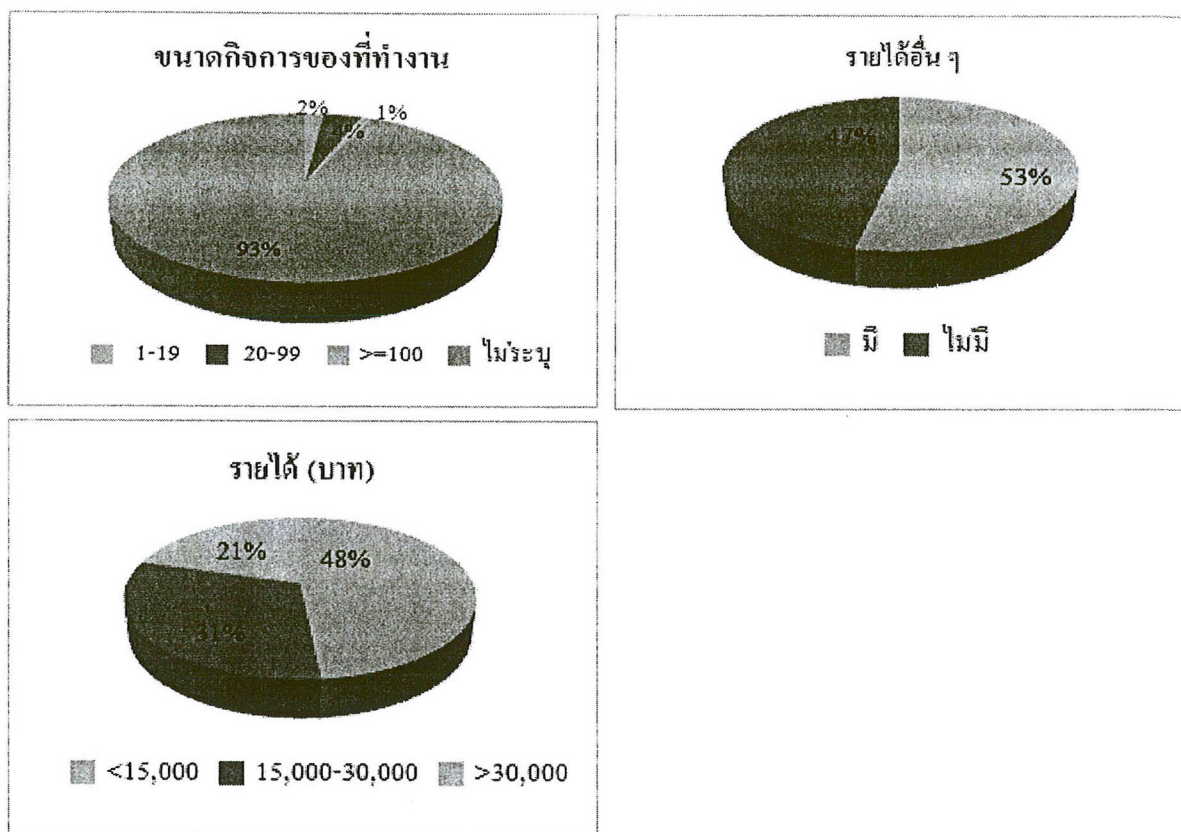
จำนวนรายการทั้งหมดในกลุ่มที่ 3 นี้ มีจำนวนทั้งสิ้น 68,790 รายการ หรือคิดเป็นร้อยละ 35.61 ของจำนวนรายการทั้งหมด ลักษณะโดยทั่วไปของข้อมูลในกลุ่มมีประชากรอาศัยอยู่ในเขตเทศบาล ร้อยละ 77 เป็นเพศหญิงคิดเป็นร้อยละ 53 มีอายุอยู่ในช่วง 35-49 ปีคิดเป็นร้อยละ 38 รองลงมาคือ 25-34 ปี คิดเป็นร้อยละ 33 สมรสแล้วคิดเป็นร้อยละ 60 เป็นลูกจ้างเอกชนคิดเป็นร้อยละ 79 ทำงานในบริษัทที่มีจำนวนพนักงาน 1-19 คน คิดเป็นร้อยละ 49 มีการศึกษาอยู่ในระดับมัธยมศึกษาเท่ากับระดับต่ำกว่าระดับมัธยมศึกษาคือคิดเป็นร้อยละ 34 รองลงคือระดับอุดมศึกษาคิดเป็นร้อยละ 32 มีอาชีพอยู่ในกลุ่มที่ 3 คือ เสมียน พนักงานบริการ และพนักงานขายในร้านค้าและตลาด อาชีพขั้นพื้นฐานต่างๆ คิดเป็นร้อยละ 49 รองลงมาคือ อาชีพกลุ่มที่ 4 คือ ผู้ปฏิบัติงานที่มีฝีมือในด้านการเกษตรและการประมง ผู้ปฏิบัติงานในธุรกิจด้านความสามารถทางฝีมือและธุรกิจอื่นๆ ที่เกี่ยวข้อง ผู้ปฏิบัติการเครื่องจักร โรงงานและเครื่องจักร และผู้ปฏิบัติงานด้านการประกอบคิดเป็นร้อยละ 30 เป็นการจ้างงานแบบรายเดือนคิดเป็นร้อยละ 76 มีจำนวนชั่วโมงการทำงานต่อสัปดาห์มากกว่า 45 ชั่วโมง คิดเป็นร้อยละ 65 รองลงมาคือ ทำงานอยู่ระหว่าง 35-45 ชั่วโมง คิดเป็นร้อยละ 28 กลุ่มที่ 3 ไม่ได้มีรายได้อื่นๆ นอกจากรายได้ประจำคิดเป็นร้อยละ 60 ช่วงรายได้ที่มีรายได้น้อยกว่า 15,000 บาทต่อเดือนคิดเป็นร้อยละ 80

รายการที่อยู่ในกลุ่มที่ 3 อยู่ในภาคกลางมากที่สุดมีจำนวน 22,889 รายการ คิดเป็นร้อยละ 33 รองลงมาคือภาคเหนือ ภาคใต้ และภาคตะวันออกเฉียงเหนือ คิดเป็นร้อยละ 18 ร้อยละ 18 และร้อยละ 17 ตามลำดับ โดยพื้นที่ที่มีรายการน้อยที่สุดคือกรุงเทพมหานครคิดเป็นร้อยละ 14

4.1.4 แอตทริบิวต์ของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 4 อื่น ๆ (ข้าราชการ การศึกษา องค์กรระหว่างประเทศ เป็นต้น)







รูปที่ 4.5 แสดงแอตทริบิวต์ของข้อมูลในกลุ่มที่ 4 อื่น ๆ

จำนวนรายการทั้งหมดในกลุ่มที่ 4 นี้ มีจำนวนทั้งสิ้น 43,552 รายการ หรือคิดเป็นร้อยละ 22.55 ของจำนวนรายการทั้งหมด ลักษณะโดยทั่วไปของข้อมูลในกลุ่มมีประชากรอาศัยอยู่ในเขตเทศบาลคิดเป็นร้อยละ 78 เป็นเพศชายคิดเป็นร้อยละ 52 มีอายุอยู่ในช่วง 35-49 ปีคิดเป็นร้อยละ 44 รองลงมาคือตั้งแต่ 50 ปีขึ้นไปคิดเป็นร้อยละ 30 สมรสแล้วร้อยละ 72 เป็นลูกจ้างรัฐบาลคิดเป็นร้อยละ 93 ขนาดกิจการของที่ทำงานไม่ได้มีการระบุจำนวนพนักงานคิดเป็นร้อยละ 93 มีการศึกษาอยู่ในระดับอุดมศึกษาคิดเป็นร้อยละ 71 รองลงคือระดับมัธยมศึกษาคิดเป็นร้อยละ 21 มีอาชีพอยู่ในกลุ่มที่ 2 คือ ผู้ประกอบวิชาชีพด้านต่างๆ ช่างเทคนิคสาขาต่างๆ และผู้ประกอบวิชาชีพอื่นๆ ที่เกี่ยวข้อง คิดเป็นร้อยละ 57 รองลงมาคือ อาชีพกลุ่มที่ 3 คือ เสมียน พนักงานบริการ พนักงานขายในร้านค้าและตลาด อาชีพขั้นพื้นฐานต่างๆ คิดเป็นร้อยละ 25 เป็นการจ้างงานแบบรายเดือนคิดเป็นร้อยละ 98 มีจำนวนชั่วโมงการทำงานต่อสัปดาห์อยู่ในช่วง 35-45 ชั่วโมงคิดเป็นร้อยละ 76 กลุ่มที่ 4 มีรายได้อื่น ๆ นอกจากรายได้ประจำคิดเป็นร้อยละ 53 ช่วงรายได้ที่มีรายได้น้อยกว่า 15,000 บาทต่อเดือนคิดเป็นร้อยละ 48

รายการที่อยู่ในกลุ่มที่ 4 อยู่ในภาคตะวันออกเฉียงเหนือมากที่สุดมีจำนวน 12,781 รายการ คิดเป็นร้อยละ 29 รองลงมาคือภาคกลาง ภาคเหนือ และได้ คิดเป็นร้อยละ 27 ร้อยละ 22 และร้อยละ 17 ตามลำดับ โดยพื้นที่ที่มีรายการน้อยที่สุดคือกรุงเทพมหานครคิดเป็นร้อยละ 5

4.2 การสร้างแบบจำลองการจำแนกประเภทข้อมูลการสำรวจด้านการทำงานของประชากร

ขั้นตอนนี้จะทำการสร้างแบบจำลองการจำแนกประเภทข้อมูลการสำรวจด้านการทำงานของประชากร ตามที่ได้จัดกลุ่มไว้แล้วในขั้นตอนที่ 4.1 ซึ่งแบ่งการทำงานออกเป็น 2 ขั้นตอน คือ

ขั้นตอนที่ 1) การสร้างแบบจำลองจากข้อมูลฝึก โดยทดลองกับค่า Confidence Factor 4 ค่า คือ 0.01, 0.1, 0.25, และ 0.5 เพื่อหาค่า CF ที่เหมาะสม ผลลัพธ์ที่ได้คือ แบบจำลองต้นไม้การตัดสินใจ

ขั้นตอนที่ 2) การทดสอบประสิทธิภาพของแบบจำลองโดยใช้ข้อมูลทดสอบ

ตารางที่ 4.3 สัดส่วนของข้อมูลในการทดลอง

ประเภทอุตสาหกรรม	จำนวนข้อมูลที่ใช้ฝึก : ร้อยละ 80	จำนวนข้อมูลที่ใช้ทดสอบ : ร้อยละ 20
กลุ่มที่ 1 การเกษตรและประมง	13,018	3,255
กลุ่มที่ 2 การผลิต	51,649	12,912
กลุ่มที่ 3 การค้าและการบริการ	55,032	13,758
กลุ่มที่ 4 อื่น ๆ	34,842	8,710

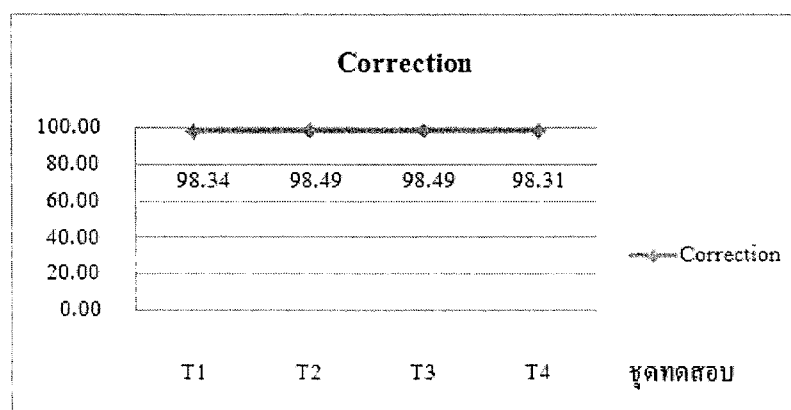
4.2.1 ผลการสร้างแบบจำลองต้นไม้ตัดสินใจของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 1 การเกษตร และประมง

ตารางที่ 4.4 ผลการสร้างแบบจำลองต้นไม้ตัดสินใจของกลุ่มที่ 1 การเกษตร และประมง

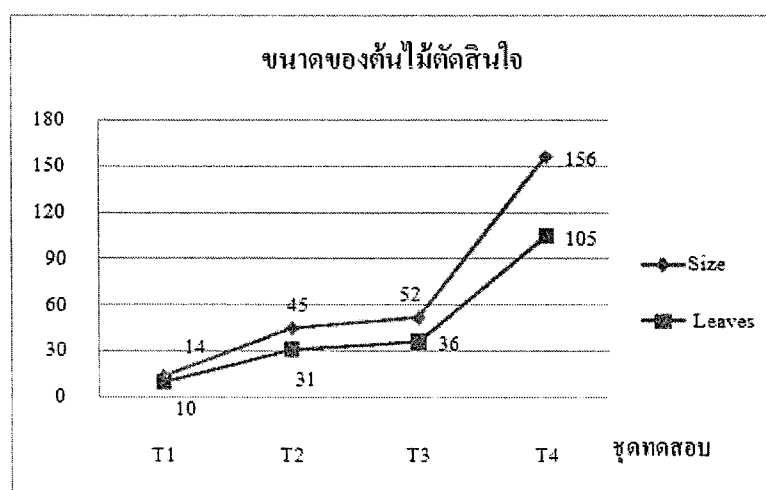
ชุด ทดสอบ	Correction	Precision	Recall	F-Measure	Size of the tree	Number of Leaves
T1	98.34	0.98	0.98	0.98	14	10
T2	98.49	0.98	0.99	0.98	45	31
T3	98.49	0.98	0.99	0.98	52	36
T4	98.31	0.98	0.98	0.98	156	105

จากตารางที่ 4.4 พบว่าเมื่อทำการทดลองโดยการกำหนดค่า Confidence Factor (CF) ที่แตกต่างกัน ผลลัพธ์ที่ได้จะให้ค่า Correction ใกล้เคียงกันมาก ชุดทดสอบที่ให้ค่า Correction สูงที่สุด คือ T2 และ T3 มีค่า Correction คือร้อยละ 98.49 รองลงมาคือ T1 และ T4 มีค่า Correction เท่ากับร้อยละ 98.34 และ ร้อยละ 98.31 ตามลำดับ โดยทั้ง 4 ชุดทดสอบมีค่า Precision, Recall, และ F-Measure ใกล้เคียงกันมาก คือ 0.98, 0.98-0.99, และ 0.98 ตามลำดับ ส่วนขนาดของต้นไม้การตัดสินใจและจำนวนของโหนดใบจะเพิ่มขึ้นเมื่อมีการกำหนดค่า CF เพิ่มขึ้น ทำให้ต้นไม้มีขนาดใหญ่ขึ้นตามลำดับ

จากผลดังกล่าวนักวิจัยเลือกชุดทดลอง T2 มาใช้เพื่อสร้างแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 1 การเกษตรและประมง เนื่องจากให้ค่า Correction สูงที่สุด และขนาดของต้นไม้ก็ไม่ซับซ้อนเกินไป



รูปที่ 4.6 แสดงการเปรียบเทียบค่า Correction ของแต่ละชุดทดสอบกลุ่มที่ 2



รูปที่ 4.7 แสดงการเปรียบเทียบขนาดของต้นไม้การตัดสินใจ และจำนวนโหนดใบของแต่ละชุดทดสอบของกลุ่มที่ 1

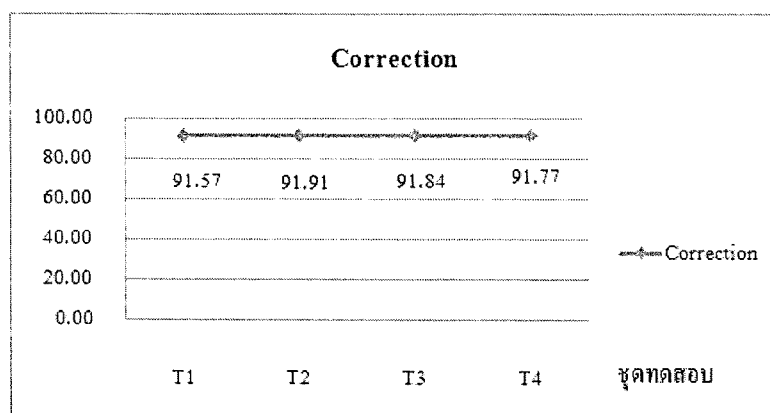
4.2.2 ผลการสร้างแบบจำลองต้นไม้สินของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 2 การผลิต

ตารางที่ 4.5 ผลการสร้างแบบจำลองต้นไม้สินใจของกลุ่มที่ 2 การผลิต

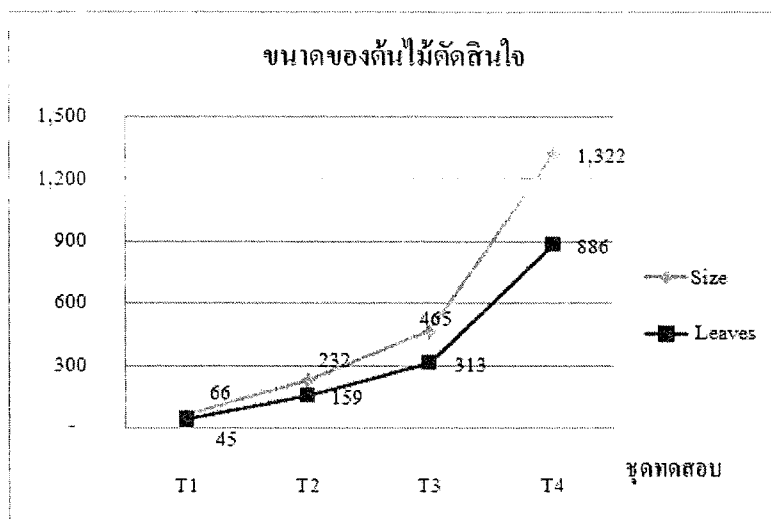
ชุดทดสอบ	Correction	Precision	Recall	F-Measure	Size of the tree	Number of Leaves
T1	91.57	0.90	0.92	0.90	66	45
T2	91.91	0.91	0.92	0.91	232	159
T3	91.84	0.90	0.92	0.91	465	313
T4	91.77	0.91	0.92	0.91	1,322	886

จากตารางที่ 4.5 พบว่าเมื่อทำการทดลองโดยการกำหนดค่า Confidence Factor (CF) ที่แตกต่างกัน ผลลัพธ์ที่ได้จะให้ค่า Correction ใกล้เคียงกันมาก ชุดทดสอบที่ให้ค่า Correction สูงที่สุด คือ T2 มีค่า Correction คือร้อยละ 91.91 รองลงมาคือ T3, T4 และ T1 มีค่า Correction เท่ากับร้อยละ 91.84 ร้อยละ 91.77 และร้อยละ 91.57 ตามลำดับ โดยทั้ง 4 ชุดทดสอบมีค่า Precision, Recall, และ F-Measure ใกล้เคียงกันมาก คือ 0.90-0.91, 0.92, และ 0.90-0.91 ตามลำดับ ส่วนขนาดของต้นไม้การตัดสินใจและจำนวนของโหนดใบ (Leaves) จะเพิ่มขึ้นเมื่อมีการกำหนดค่า CF เพิ่มขึ้น ทำให้ต้นไม้มีขนาดใหญ่ขึ้นตามลำดับ

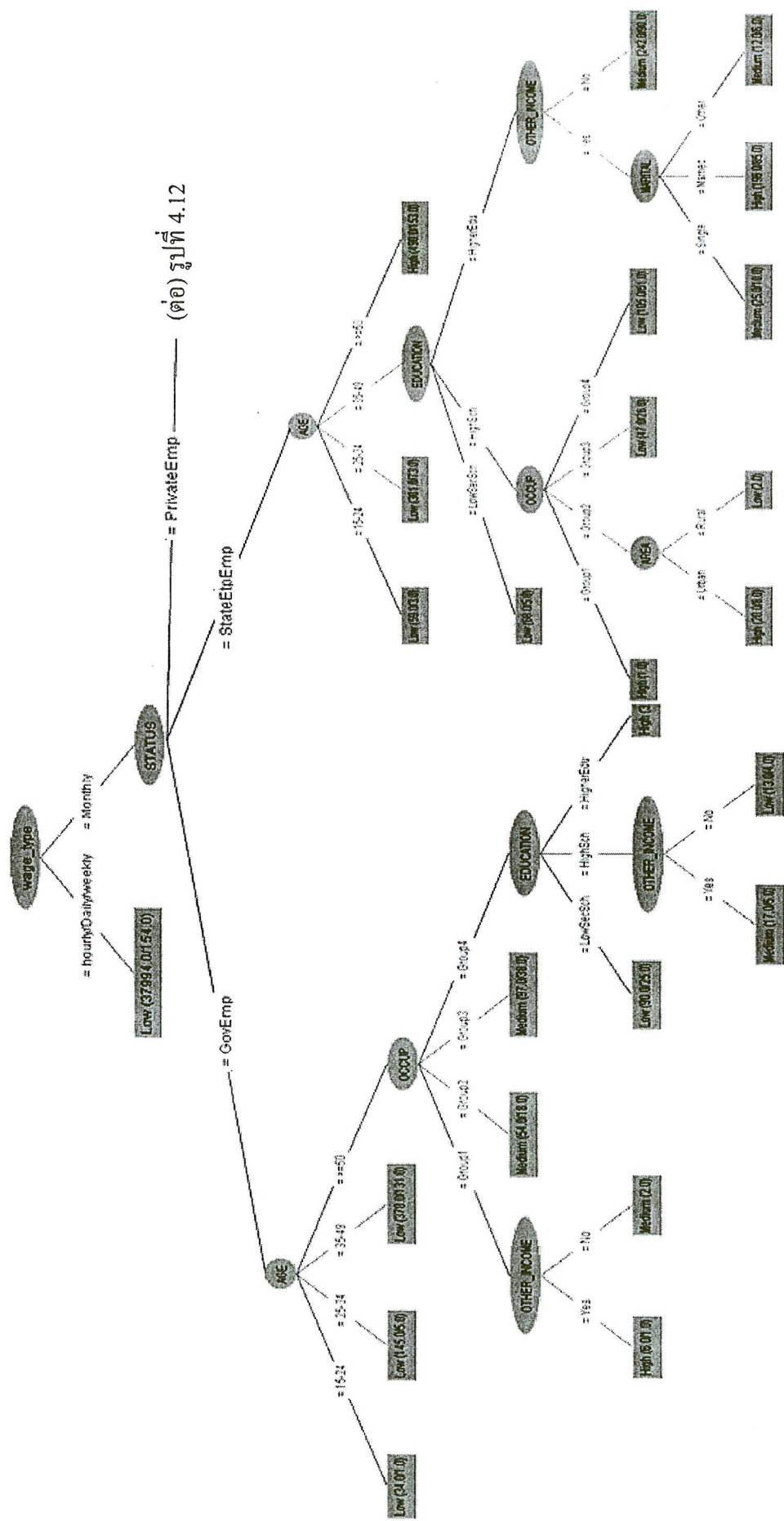
จากผลดังกล่าวนักวิจัยเลือกชุดทดลอง T1 มาใช้เพื่อสร้างแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 2 การเกษตรและประมง เนื่องจากจะเห็นได้ว่า ชุดทดสอบอื่น ๆ ถึงแม้จะมีค่า Correction สูงกว่า T1 แต่ขนาดของต้นไม้มีขนาดใหญ่เกินไป ทำให้กฎที่ได้มีความซับซ้อนสูง แต่ชุดทดสอบ T1 ถึงแม้จะให้ค่า CF น้อยที่สุด (ค่า CF ของแต่ละชุดทดสอบใกล้เคียงกัน) แต่มีขนาดของต้นไม้ไม่ซับซ้อนเกินไป



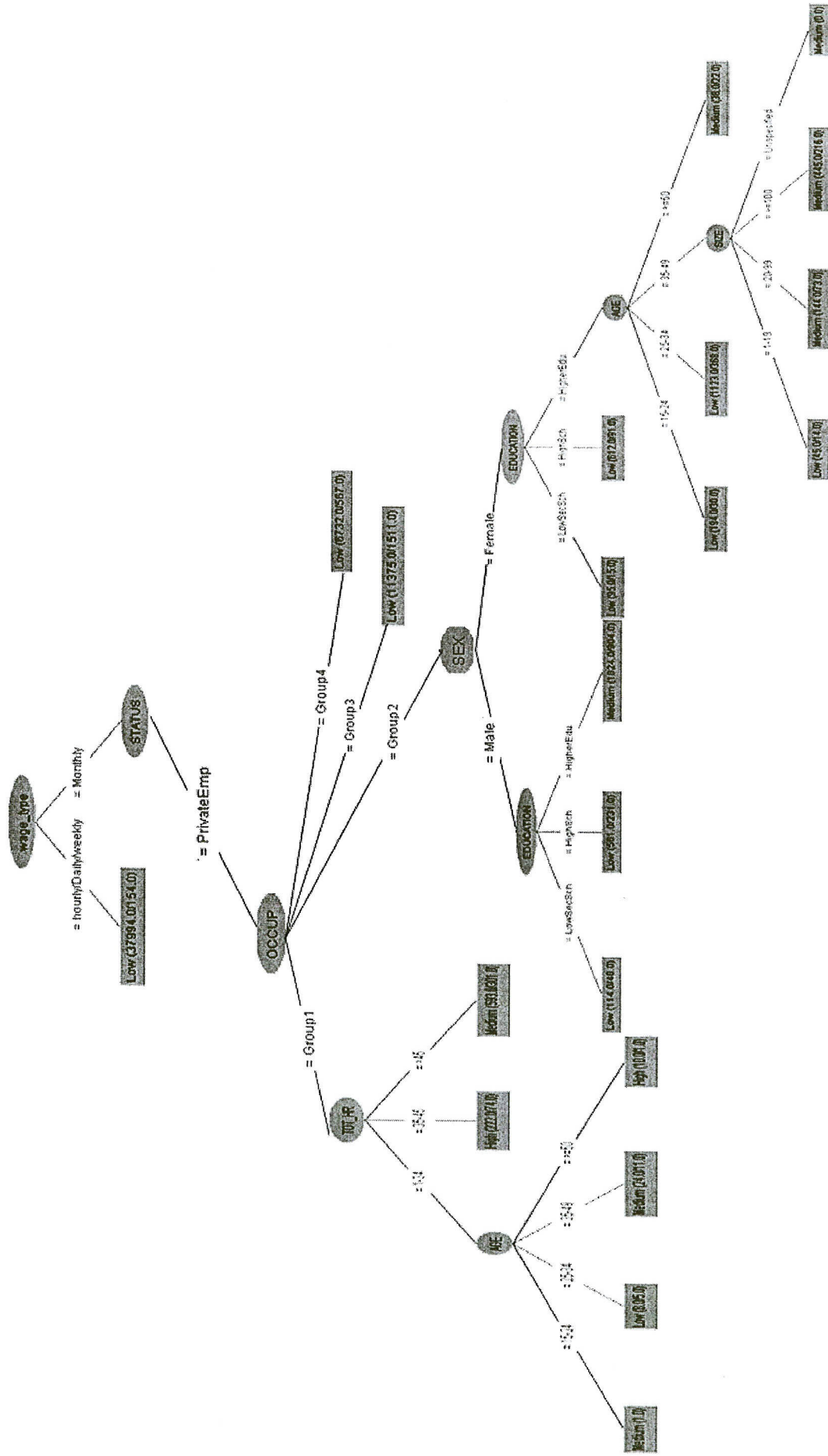
รูปที่ 4.9 แสดงการเปรียบเทียบค่า Correction ของแต่ละชุดทดสอบของกลุ่มที่ 2



รูปที่ 4.10 แสดงการเปรียบเทียบขนาดของต้นไม้การตัดสินใจ และจำนวนโหนดใบของแต่ละชุดทดสอบของกลุ่มที่ 2



รูปที่ 4.11 แสดงต้นไม้การตัดสินใจของกลุ่มที่ 2 การผลิต



รูปที่ 4.12 แสดงต้นไม้การตัดสินใจของกลุ่มที่ 2 การผลิต (ต่อ)

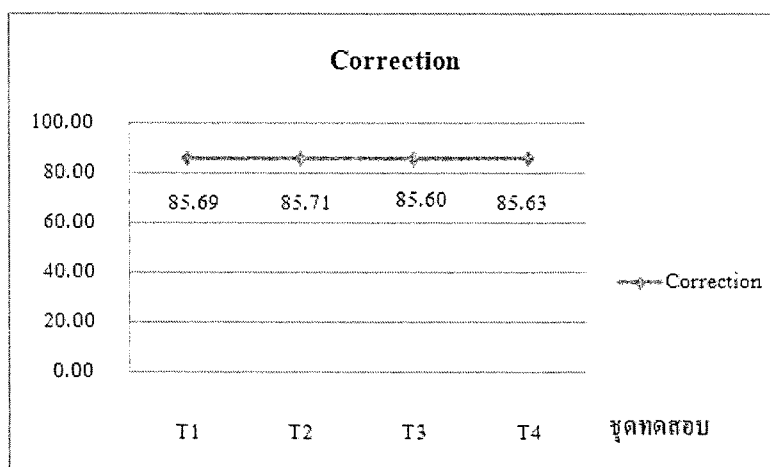
4.2.3 ผลการสร้างแบบจำลองต้นไม้สินของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 3 การค้าและการบริการ

ตารางที่ 4.6 ผลการสร้างแบบจำลองต้นไม้สินใจของกลุ่มที่ 3 การค้าและการบริการ

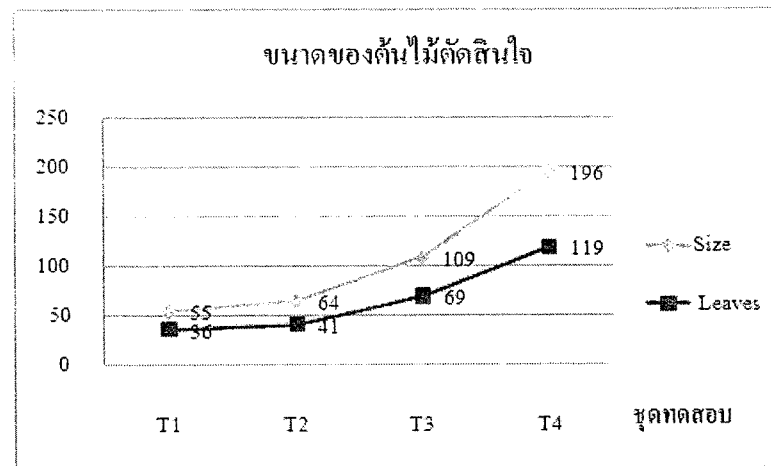
ชุดทดสอบ	Correction	Precision	Recall	F-Measure	Size of the tree	Number of Leaves
T1	85.69	0.85	0.86	0.85	55	36
T2	85.71	0.85	0.86	0.85	64	41
T3	85.60	0.85	0.86	0.85	109	69
T4	85.63	0.85	0.86	0.85	196	119

จากตารางที่ 4.6 พบว่าเมื่อทำการทดลองโดยการกำหนดค่า Confidence Factor (CF) ที่แตกต่างกัน ผลลัพธ์ที่ได้จะให้ค่า Correction ใกล้เคียงกันมาก ชุดทดสอบที่ให้ค่า Correction สูงที่สุด คือ T2 มีค่า Correction คือร้อยละ 85.71 รองลงมาคือ T1, T4 และ T3 มีค่า Correction เท่ากับร้อยละ 85.69 ร้อยละ 85.63 และร้อยละ 85.60 ตามลำดับ โดยทั้ง 4 ชุดทดสอบมีค่า Precision, Recall, และ F-Measure เท่ากัน คือ 0.85, 0.86, และ 0.85 ตามลำดับ ส่วนขนาดของต้นไม้การตัดสินใจและจำนวนของโหนดใบจะเพิ่มขึ้นเมื่อมีการกำหนดค่า CF เพิ่มขึ้นทำให้ต้นไม้มีขนาดใหญ่ขึ้นตามลำดับ

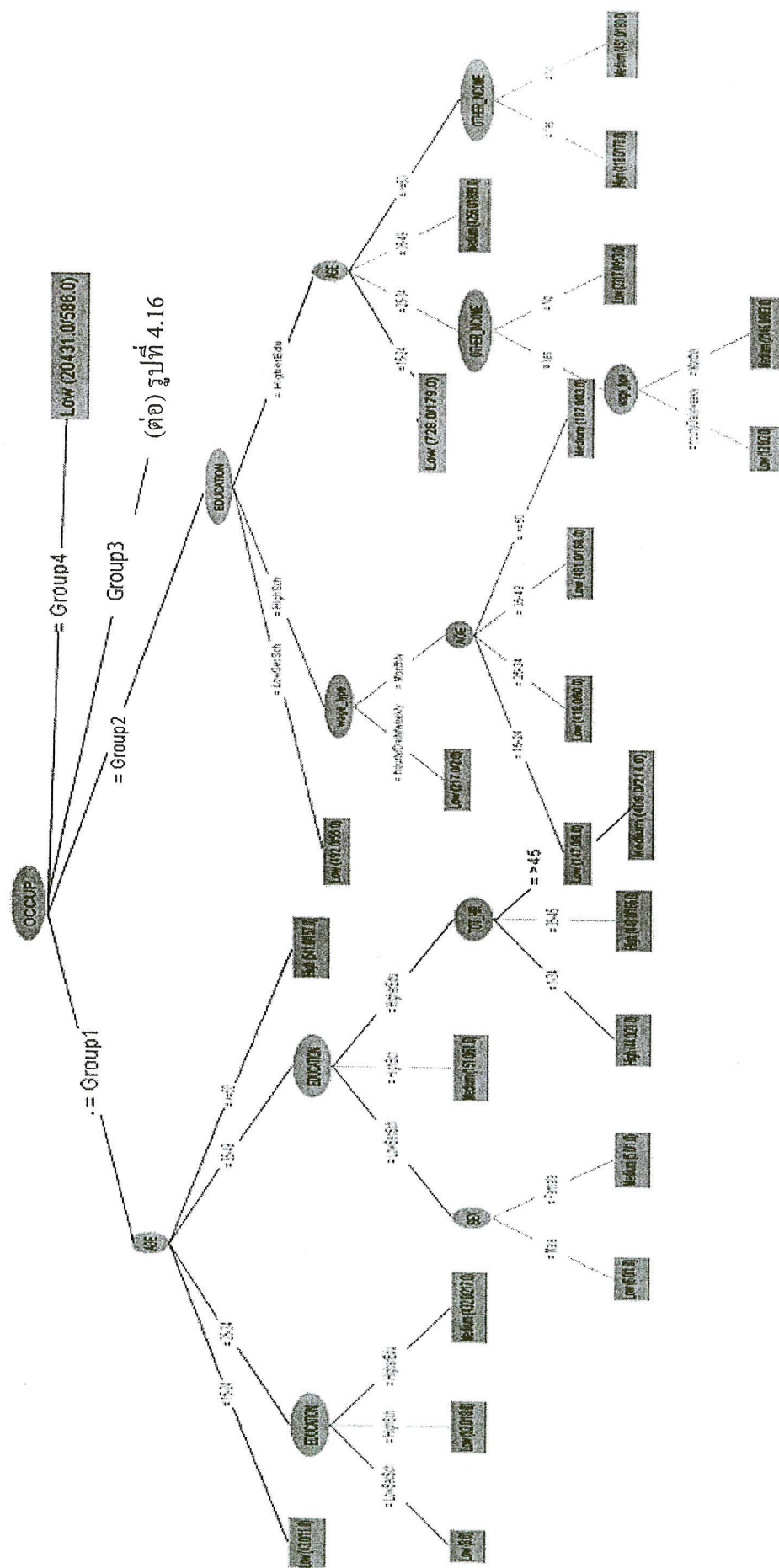
จากผลดังกล่าว นักวิจัยเลือกชุดทดสอบ T2 มาใช้เพื่อสร้างแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 3 การค้าและการบริการ เนื่องจากให้ค่า Correction สูงที่สุด และขนาดของต้นไม้ก็ไม่ซับซ้อนเกินไป



รูปที่ 4.13 แสดงการเปรียบเทียบค่า Correction ของแต่ละชุดทดสอบของกลุ่มที่ 3



รูปที่ 4.14 แสดงการเปรียบเทียบขนาดของต้นไม้การตัดสินใจ และจำนวนโหนดใบของแต่ละชุดทดสอบของกลุ่มที่ 3



รูปที่ 4.15 แสดงต้นไม้การตัดสินใจของกลุ่มที่ 3 การค้าและการบริการ



รูปที่ 4.16 แสดงต้นไม้การตัดสินใจของกลุ่มที่ 3 การก้าและการบริการ(ต่อ)

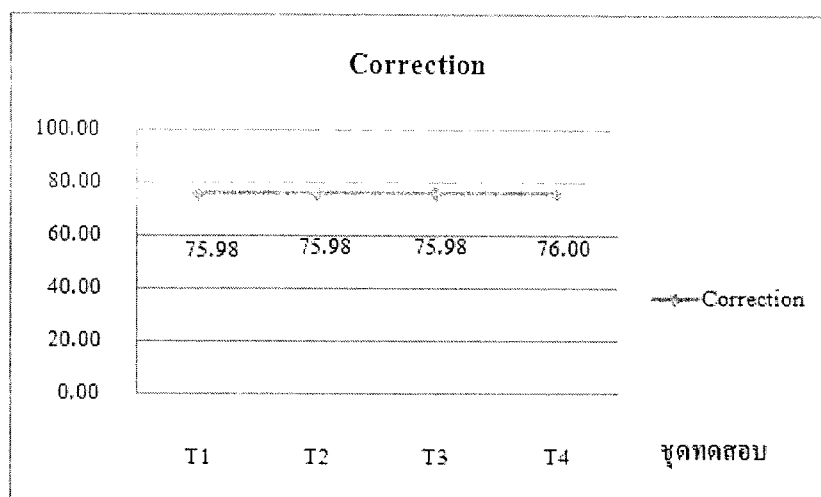
4.2.4 ผลการสร้างแบบจำลองต้นไม้สินของข้อมูลของประเภทอุตสาหกรรม กลุ่มที่ 4 อื่น ๆ (ข้าราชการ การศึกษา องค์กรระหว่างประเทศ เป็นต้น)

ตารางที่ 4.7 ผลการสร้างแบบจำลองต้นไม้สินใจของกลุ่มที่ 4 อื่น ๆ (ข้าราชการ การศึกษา องค์กรระหว่างประเทศ เป็นต้น)

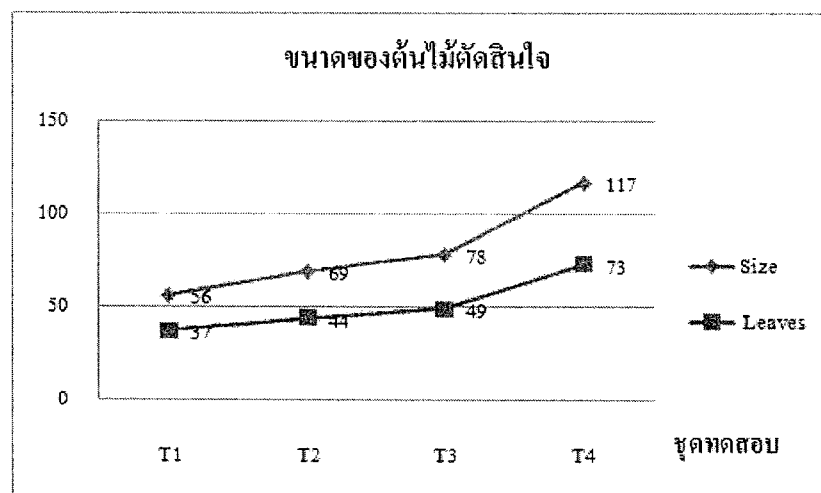
ชุดทดสอบ	Correction	Precision	Recall	F-Measure	Size of the tree	Number of Leaves
T1	75.98	0.77	0.76	0.76	56	37
T2	75.98	0.77	0.76	0.76	69	44
T3	75.98	0.77	0.76	0.76	78	49
T4	76.00	0.77	0.76	0.76	117	73

จากตารางที่ 4.7 พบว่าเมื่อทำการทดลองโดยการกำหนดค่า Confidence Factor (CF) ที่แตกต่างกัน ผลลัพธ์ที่ได้จะให้ค่า Correction ใกล้เคียงกันมาก ชุดทดสอบที่ให้ค่า Correction สูงที่สุด คือ T4 มีค่า Correction คือร้อยละ 76 รองลงมาคือ T1, T2 และ T3 มีค่า Correction เท่ากันคือร้อยละ 75.98 โดยทั้ง 4 ชุดทดสอบมีค่า Precision, Recall, และ F-Measure เท่ากัน คือ 0.77, 0.76 และ 0.76 ตามลำดับ ส่วนขนาดของต้นไม้การตัดสินใจและจำนวนของโหนดใบจะเพิ่มขึ้นเมื่อมีการกำหนดค่า CF เพิ่มขึ้น ทำให้ต้นไม้มีขนาดใหญ่ขึ้นตามลำดับ

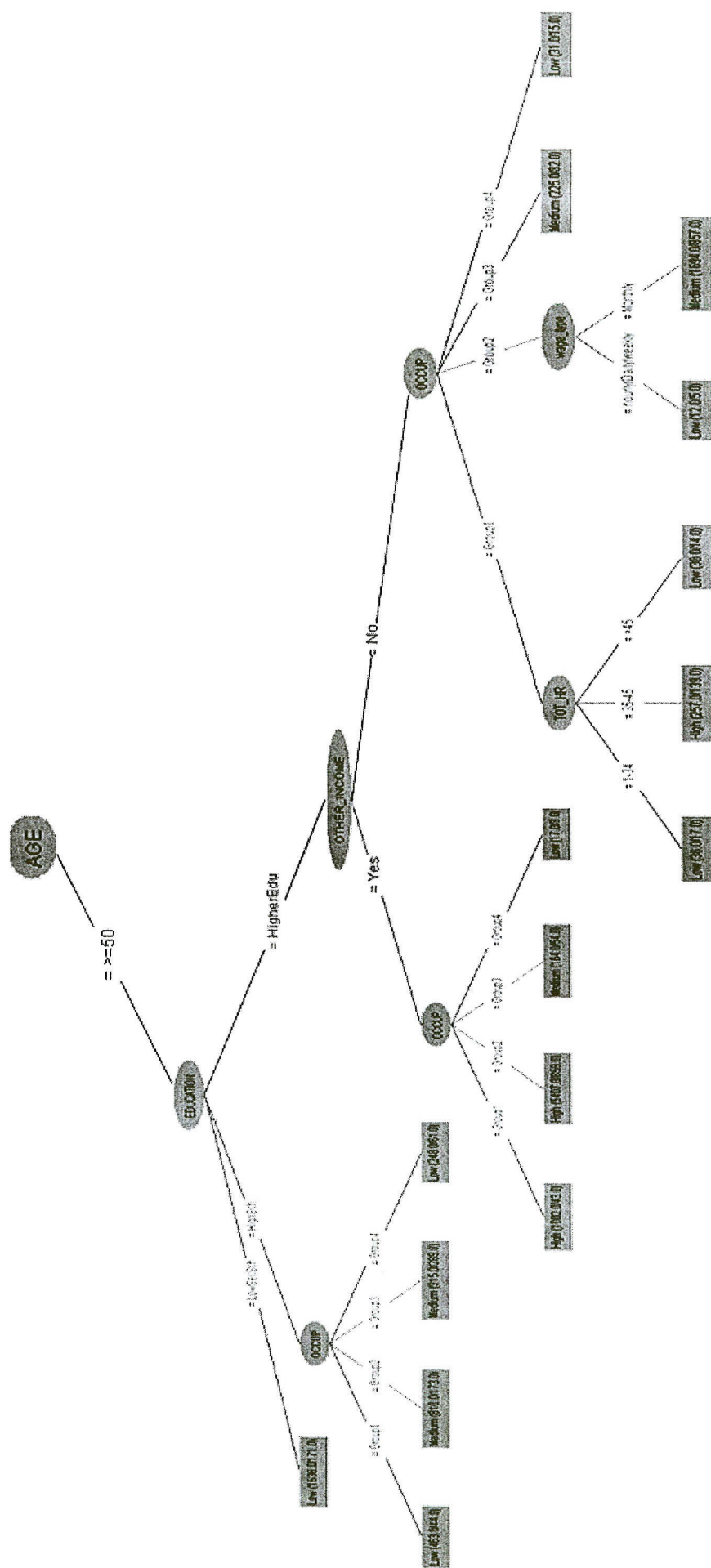
จากผลดังกล่าว นักวิจัยเลือกชุดทดลอง T1 มาใช้เพื่อสร้างแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 4 อื่น ๆ เนื่องจากจะเห็นได้ว่าชุดทดสอบอื่น ๆ ถึงแม้จะมีค่า Correction สูงกว่า T1 แต่ขนาดของต้นไม้มีขนาดใหญ่มาก ทำให้กฎที่ได้มีความซับซ้อนสูง แต่ชุดทดสอบ T1 ถึงแม้จะให้ค่า Correction น้อยที่สุด (ค่า CF ของแต่ละชุดทดสอบใกล้เคียงกัน) แต่มีขนาดของต้นไม้ไม่ซับซ้อนเกินไป



รูปที่ 4.17 แสดงการเปรียบเทียบค่า Correction ของแต่ละชุดทดสอบของกลุ่มที่ 4



รูปที่ 4.18 แสดงการเปรียบเทียบขนาดของต้นไม้การตัดสโนใจ และจำนวนหน่อใบของแต่ละชุดทดสอบของกลุ่มที่ 4



รูปที่ 4.20 แสดงต้นไม้การตัดสินใจของกลุ่มที่ 4 อื่น ๆ (ต่อ)



4.3 การสร้างกฎความสัมพันธ์จากต้นไม้การตัดสินใจ

ขั้นตอนนี้ จะทำการแปลงต้นไม้การตัดสินใจที่ได้จากขั้นตอนที่ 4.2 ของแต่ละกลุ่ม เป็นกฎความสัมพันธ์ เพื่อง่ายต่อการศึกษาปัจจัยที่มีผลต่อรายได้ของประชากร

4.3.1 ผลการสร้างกฎความสัมพันธ์ของประเภทอุตสาหกรรม กลุ่มที่ 1 การเกษตรและประมง

ตารางที่ 4.8 แสดงตัวอย่างกฎความสัมพันธ์ที่สร้างจากแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 1

เหตุการณ์/เงื่อนไข (Antecedent หรือ IF)	สิ่งที่เกิดขึ้น/ผลลัพธ์ (Consequence/THEN)	ค่าความเชื่อมั่น (Confidence Factor)
1 OCCUP=1, EDUCATION=ระดับอุดมศึกษา, REG=กลาง	INCOME=มาก (HIGH)	0.8
2 OCCUP=1, EDUCATION=ระดับอุดมศึกษา, REG=ได้, AGE=35-49 ปี	INCOME=มาก (HIGH)	0.9

ตารางที่ 4.8 นี้ แสดงให้เห็นถึงตัวอย่างกฎที่ได้จากต้นไม้การตัดสินใจของกลุ่มที่ 1 จำนวน 2 กฎ จากจำนวนกฎทั้งสิ้น 22 กฎ

กฎความสัมพันธ์ที่ 1 คือเมื่อประชากรมีอาชีพอยู่ในกลุ่ม 1 คือข้าราชการระดับอาวุโส และผู้จัดการ มีระดับการศึกษาในระดับอุดมศึกษา และอาศัยอยู่ในภาคกลาง แล้วจะมีรายได้ในช่วงมาก คือรายได้มากกว่า 30,000 บาทขึ้นไป ตามกฎความสัมพันธ์ที่ 1 นี้ โอกาสที่ประชากรจะมีรายได้ในช่วงมากคิดเป็นร้อยละ 80 สืบเนื่องจากค่าความเชื่อมั่นมีค่าเท่ากับ 0.8 สามารถแสดงกฎความสัมพันธ์ในรูปของ IF-THEN ได้ดังรูปที่ 4.21

กฎความสัมพันธ์ที่ 2 คือเมื่อประชากรมีอาชีพอยู่ในกลุ่ม 1 คือข้าราชการระดับอาวุโส และผู้จัดการ มีระดับการศึกษาในระดับอุดมศึกษา อาศัยอยู่ในภาคใต้ และมีอายุอยู่ในช่วง 35-49 ปี แล้วจะมีรายได้ในช่วงมาก คือรายได้มากกว่า 30,000 บาทขึ้นไป ตามกฎความสัมพันธ์ที่ 2 นี้ โอกาสที่ประชากรจะมีรายได้ในช่วงมากคิดเป็นร้อยละ 90 สืบเนื่องจากค่าความเชื่อมั่นมีค่าเท่ากับ 0.9

IF OCCUP=1 AND EDUCATION =ระดับอุดมศึกษา AND REG = ภาคกลาง **THEN** INCOME = HIGH

รูปที่ 4.21 แสดงการเขียนกฎความสัมพันธ์ในลักษณะ IF-THEN (1)

4.3.2 ผลการสร้างกฎความสัมพันธ์ของประเภทอุตสาหกรรม กลุ่มที่ 2 การผลิต

ตารางที่ 4.9 แสดงตัวอย่างกฎความสัมพันธ์ที่สร้างจากแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 2

	เหตุการณ์/เงื่อนไข (Antecedent หรือ IF)	สิ่งที่เกิดขึ้น/ผลลัพธ์ (Consequence/THEN)	ค่าความเชื่อมั่น (Confidence Factor)
1	WAGE_TYPE=รายเดือน, STATUS=ลูกจ้างรัฐบาล, AGE>=50, OCCUP=4, EDUCATION= ระดับอุดมศึกษา	INCOME=มาก (HIGH)	1
2	WAGE_TYPE=รายเดือน, STATUS=พนักงานรัฐวิสาหกิจ, AGE=35-49, EDUCATION=ระดับ มัธยมศึกษา, OCCUP=1	INCOME=มาก (HIGH)	1
3	WAGE_TYPE=รายเดือน, STATUS=ลูกจ้างเอกชน, OCCUP=1, TOT_HR=1-34, AGE>=5	INCOME=มาก (HIGH)	0.9
4	WAGE_TYPE=รายเดือน, STATUS=ลูกจ้างรัฐบาล, AGE>=50, OCCUP=1, OTHER_INCOME=มี	INCOME=มาก (HIGH)	0.83
5	WAGE_TYPE=รายเดือน, STATUS=พนักงานรัฐวิสาหกิจ, AGE>= 50	INCOME=มาก (HIGH)	0.69
6	WAGE_TYPE=รายเดือน, STATUS=ลูกจ้างเอกชน, OCCUP=1, TOT_HR=35-45	INCOME=มาก (HIGH)	0.67

ตารางที่ 4.9 นี้ แสดงให้เห็นถึงตัวอย่างกฎที่ได้จากต้นไม้การตัดสินใจของกลุ่มที่ 2 จำนวน 6 กฎ จากจำนวนกฎทั้งสิ้น 33 กฎ

กฎความสัมพันธ์ที่ 1 คือเมื่อประชากรมีประเภทการจ้างงานแบบรายเดือน มีสถานะภาพเป็นลูกจ้างรัฐบาล มีอายุมากกว่าหรือเท่ากับ 50 ปีขึ้นไป อาศัยอยู่ในกลุ่มที่ 4 คือผู้ปฏิบัติงานที่มีฝีมือในด้านการเกษตรและการประมง ผู้ปฏิบัติงานในธุรกิจด้านความสามารถทางฝีมือและธุรกิจอื่นๆ ที่เกี่ยวข้อง ผู้ปฏิบัติการเครื่องจักรโรงงานและเครื่องจักร และผู้ปฏิบัติงานด้านการประกอบ ซึ่งเป็นผู้ที่มีระดับ

การศึกษาในระดับอุดมศึกษา แล้วจะมีรายได้ในช่วงมาก คือรายได้มากกว่า 30,000 บาทขึ้นไป ตามกฎความสัมพันธ์ที่ 1 นี้ โอกาสที่ประชากรจะมีรายได้ในช่วงมากด้วยทุกครั้ง สังเกตค่าความเชื่อมั่นที่เท่ากับ 1 สามารถแสดงกฎความสัมพันธ์ในรูปของ IF-THEN ได้ดังรูปที่ 4.22

IF WAGE_TYPE = รายเดือน AND STATUS = ลูกจ้างรัฐบาล AND AGE >=50 AND OCCUP=4 AND EDUCATION=ระดับอุดมศึกษา **THEN** INCOME = HIGH

รูปที่ 4.22 แสดงการเขียนกฎความสัมพันธ์ในลักษณะ IF-THEN (2)

4.3.3 ผลการสร้างกฎความสัมพันธ์ของประเภทอุตสาหกรรม กลุ่มที่ 3 การค้าและการบริการ

ตารางที่ 4.10 แสดงตัวอย่างกฎความสัมพันธ์ที่สร้างจากแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 3

	เหตุการณ์/เงื่อนไข (Antecedent หรือ IF)	สิ่งที่เกิดขึ้น/ผลลัพธ์ (Consequence/THEN)	ค่าความเชื่อมั่น (Confidence Factor)
1	OCCUP=1, AGE>=50	INCOME=มาก (HIGH)	0.71
2	OCCUP=1, AGE=35-49, EDUCATION=ระดับอุดมศึกษา, TOT_HR=35-49	INCOME=มาก (HIGH)	0.62
3	OCCUP=2, EDUCATION= ระดับอุดมศึกษา, AGE>=50, OTHER_INCOME=มี	INCOME=มาก (HIGH)	0.59
4	OCCUP=1, AGE=35-49, EDUCATION=ระดับอุดมศึกษา, TOT_HR=1-34	INCOME=มาก (HIGH)	0.52

ตารางที่ 4.10 นี้ แสดงให้เห็นถึงตัวอย่างกฎที่ได้จากต้นไม้การตัดสินใจของกลุ่มที่ 3 จำนวน 4 กฎ จากจำนวนกฎทั้งสิ้น 35 กฎ

กฎความสัมพันธ์ที่ 1 คือเมื่อประชากรมีอาชีพอยู่ในกลุ่มที่ 1 คือข้าราชการระดับอาวุโส และผู้จัดการ และมีอายุตั้งแต่ 50 ปีขึ้นไปจะมีรายได้ในช่วงมากคือรายได้มากกว่า 30,000 บาทขึ้นไป ตามกฎความสัมพันธ์ที่ 1 นี้ โอกาสที่ประชากรจะมีรายได้ในช่วงมากคิดเป็นร้อยละ 71 สังเกตค่าความเชื่อมั่นที่เท่ากับ 0.71 สามารถแสดงกฎความสัมพันธ์ในรูปของ IF-THEN ได้ดังรูปที่ 4.23

IF OCCUP = 1 AND AGE >= 50 THEN INCOME = HIGH

รูปที่ 4.23 แสดงการเขียนกฎความสัมพันธ์ในลักษณะ IF-THEN (3)

ตารางที่ 4.11 แสดงตัวอย่างกฎความสัมพันธ์ที่สร้างจากแบบจำลองต้นไม้การตัดสินใจของกลุ่มที่ 4

เหตุการณ์/เงื่อนไข (Antecedent หรือ IF)	สิ่งที่เกิดขึ้น/ผลลัพธ์ (Consequence/THEN)	ค่าความเชื่อมั่น (Confidence Factor)
1 AGE>=50, EDUCATION=ระดับอุดมศึกษา, OTHER_INCOME=มี, OCCUP=1	INCOME=มาก (HIGH)	0.96
2 AGE>=50, EDUCATION=ระดับอุดมศึกษา, OTHER_INCOME=มี, OCCUP=2	INCOME=มาก (HIGH)	0.84
3 AGE=35-49, EDUCATION=ระดับอุดมศึกษา, OTHER_INCOME=มี, OCCUP=1	INCOME=มาก (HIGH)	0.7
4 AGE>=50, EDUCATION=ระดับอุดมศึกษา, OTHER_INCOME=ไม่มี, OCCUP=1, TOT_HR=35-45	INCOME=มาก (HIGH)	0.46

ตารางที่ 4.11 นี้ แสดงให้เห็นถึงตัวอย่างกฎที่ได้จากต้นไม้การตัดสินใจของกลุ่มที่ 4 จำนวน 4 กฎ จากจำนวนกฎทั้งสิ้น 33 กฎ

กฎความสัมพันธ์ที่ 1 คือเมื่อประชากรมีอายุมากกว่าหรือเท่ากับ 50 ปีขึ้นไป ระดับการศึกษาอยู่ในระดับอุดมศึกษา มีรายได้อื่น ๆ และมีอาชีพอยู่ในกลุ่มที่ 1 คือข้าราชการระดับอาวุโส และผู้จัดการ จะมีรายได้ในช่วงมากคือรายได้มากกว่า 30,000 บาทขึ้นไป ตามกฎความสัมพันธ์ที่ 1 นี้ มีโอกาสที่ประชากรจะมีรายได้ในช่วงมากถึงร้อยละ 96 สังเกตได้จากค่าความเชื่อมั่นที่เท่ากับ 0.96 สามารถแสดงกฎความสัมพันธ์ในรูปของ IF-THEN ได้ดังรูปที่ 4.24

IF AGE $\geq$ 50 AND EDUCATION = ระดับอุดมศึกษา AND OTHER_INCOME = มี
AND OCCUP = 1 **THEN** INCOME = HIGH

รูปที่ 4.24 แสดงการเขียนกฎความสัมพันธ์ในลักษณะ IF-THEN (4)