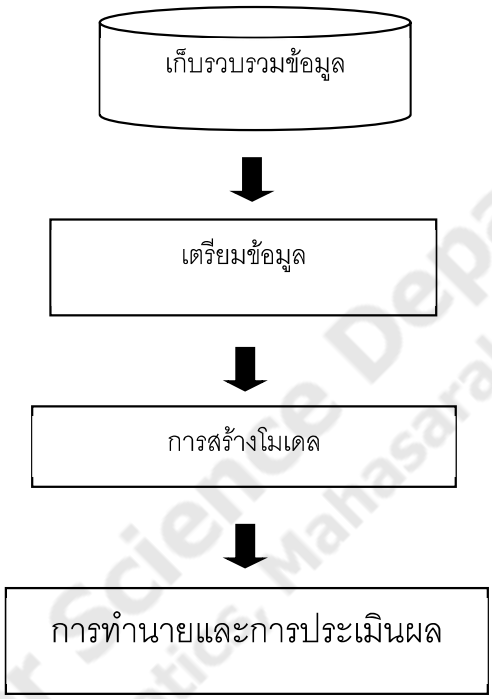
บทที่ 3
วิธีดำเนินงานวิจัย

งานวิจัยนี้เป็นงานวิจัยมีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลกระทบต่อโรคหลอดเลือดสมอง โดยขั้นตอนการดำเนินงานวิจัยแสดงได้ดังรูปที่ 3.1 โดยแต่ละขั้นตอนมีรายละเอียดดังหัวข้อต่อไปนี้



ภาพประกอบที่ 3.1 ขั้นตอนการดำเนินการวิจัย

3.1 การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในการวิจัยคือประวัติสุขภาพของผู้ป่วยที่เข้ารับบริการในโรงพยาบาลผ่านการบันทึกเวชระเบียน ซึ่งข้อมูลการรักษาถูกบันทึกโดยบุคลากรทางการแพทย์ สามารถนำไปใช้ในการศึกษาวิจัยทางวิทยาศาสตร์และด้านอื่นๆ ได้โดยการนำเทคโนโลยีและความรู้ทางวิทยาการคอมพิวเตอร์เข้ามาร่วมเพื่อช่วยวิเคราะห์ปัญหาสุขภาพของประชากรได้

โดยงานวิจัยนี้ได้ข้อมูลมาจากฐานข้อมูลผู้ป่วยที่เข้ารับบริการที่โรงพยาบาลมหาสารคามโดยประชากรกลุ่มตัวอย่างคือผู้ป่วยที่มีอายุ 60 ปีขึ้นไป ซึ่งข้อมูลที่เก็บรวบรวมจากผู้ป่วยแต่ละคนประกอบไปด้วย เพศ สถานภาพ การสูบบุหรี่ การดื่มสุรา การออกกำลังกาย อาชีพ ความดันโลหิต คอเลสเตอรอล และลำดับการเกิดโรคที่ถูกวินิจฉัยในการมารับบริการทุกครั้ง และการเป็นโรคหลอดเลือดสมอง

Computer Science Department
Faculty of Informatics Mahasarakham University

3.2 การเตรียมข้อมูล

ในขั้นตอนนี้จะทำรวบรวมผู้ป่วยให้อยู่ในรูปแบบของทรานแซคชั่น โดยในหนึ่งทรานแซคชั่น หมายถึงข้อมูลของผู้ป่วยหนึ่งคน ดังตารางที่ 3.1 ข้อมูลของผู้ป่วยแต่ละคนจะประกอบด้วย 2 ส่วน คือ

ส่วนที่ 1 เป็นส่วนของปัจจัยพื้นฐาน ซึ่งประกอบไปด้วย เพศ สถานภาพ การสูบบุหรี่ การดื่มสุรา การออกกำลังกาย อาชีพ ความดันโลหิต คอเลสเตอรอล

ส่วนที่ 2 เป็นส่วนของคลาสซึ่งประกอบด้วย 2 คลาส คือ เป็นโรคหลอดเลือดสมอง กับ ไม่เป็นโรคหลอดเลือดสมอง

ตารางที่ 3.1 ตัวอย่างข้อมูล

ID	เพศ	สถานภาพ	อาชีพ	สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	BP	คอเลสเตอรอล	คลาส
1	หญิง	สมรส	พนักงานบริษัท	ไม่สูบ	ไม่ดื่ม	ไม่ออก	สูง	สูง	เป็น
2	ชาย	สมรส	รับจ้าง	สูบ	ดื่ม	ไม่ออก	ปกติ	ปกติ	ไม่เป็น
3	ชาย	โสด	พนักงานบริษัท	ไม่สูบ	ไม่ดื่ม	ไม่ออก	สูง	สูง	เป็น
4	ชาย	สมรส	ค้าขาย	สูบ	ดื่ม	ไม่ออก	ปกติ	สูง	ไม่เป็น
5	หญิง	สมรส	รับจ้าง	ไม่สูบ	ไม่ดื่ม	ไม่ออก	ปกติ	ปกติ	ไม่เป็น
6	หญิง	สมรส	รับจ้าง	สูบ	ดื่ม	ออก	สูง	ปกติ	เป็น
7	ชาย	โสด	อาชีพอิสระ				สูง	ปกติ	ไม่เป็น

3.2.1 การทำความสะอาดข้อมูล (Data cleansing)

การทำความสะอาดข้อมูลเป็นกระบวนการตรวจสอบความไม่สมบูรณ์ ความไม่ถูกต้อง ความไม่สัมพันธ์กับข้อมูลอื่นๆ จึงต้องมีการแทนที่ การปรับปรุง หรือการลบข้อมูลที่ไม่ถูกต้องเหล่านี้ออกไป เพื่อให้เหลือเฉพาะข้อมูลที่จะนำไปวิเคราะห์ที่มีความถูกต้อง ครบถ้วนและมีคุณภาพ งานวิจัยนี้ทำความสะอาดข้อมูลโดยทำการตัดข้อมูลที่ไม่สมบูรณ์ออกไป ซึ่งเงื่อนไขการตัดข้อมูลออกดังนี้

Computer Science Department
Faculty of Informatics, Mahasarakham University

1. ถ้าผู้ป่วยมีข้อมูลพื้นฐานไม่ครบทั้ง 8 ปัจจัย ข้อมูลของผู้ป่วยทั้งทรานแซคชั่นจะถูกตัดทิ้งไป เช่น ในตารางที่ 1 จะตัดข้อมูลของผู้ป่วยคนที่ 7 ทิ้งไป เนื่องจากไม่พบข้อมูลการ การดื่มสุรา การสูบบุหรี่ การออกกำลังกาย เป็นต้น

เมื่อทำความสะอาดข้อมูลแล้วทำการสุ่มเลือกข้อมูลให้ผู้ป่วยที่เป็นโรคหลอดเลือดสมองมีจำนวนเท่ากับผู้ป่วยที่ไม่เป็นโรคหลอดเหลือดสมอง เพื่อให้การเรียนรู้การเป็นโรคหลอดเลือดสมองและการไม่เป็นโรคหลอดเลือดสมองเท่าเทียมกัน โดยสุ่มเลือกข้อมูลผู้ป่วยจำนวน 1,000 ราย โดยแบ่งเป็นผู้ป่วยที่เป็นโรคหลอดเลือดสมอง 500 ราย และผู้ป่วยที่ไม่เป็นโรคหลอดเลือดสมอง 500 ราย

เมื่อทำความสะอาดข้อมูลเรียบร้อยแล้ว ข้อมูลดังกล่าวจะถูกแปลงเป็นข้อมูลตัวเลขเพื่อให้ง่ายต่อการประมวลผลดังหัวข้อต่อไป

3.2.2 การแปลงข้อมูล (Data Transformation)

การแปลงข้อมูลเป็นขั้นตอนการแปลงให้ข้อมูลอยู่ในรูปแบบที่เหมาะสมเพื่อนำไปใช้กับขั้นตอนวิธีต่าง ๆ ได้ การแปลงข้อมูลในงานวิจัยนี้มีรายละเอียดดังนี้

1. ปัจจัยพื้นฐานแปลงเป็นตัวเลขดังตารางที่ 3.2 เช่น ในทรานแซคชั่นที่ 1 (ผู้ป่วยคนที่ 1) ข้อมูลปัจจัยพื้นฐานจะถูกแปลงเป็นตัวเลข 11 21 100 40 50 61 81 91 โดย 11 แทนเพศหญิง 21 แทน สถานภาพสมรส 100 แทนไม่ได้ทำงาน 40 แทนการไม่สูบหรี่ 50 แทนการไม่ดื่มสุรา 61 แทนการไม่ออกกำลังกาย 81 แทนความดันโลหิตผิกปกติ (สูง) 91 แทนระดับคอเลสเตอรอลผิดปกิต (สูง) เป็นต้น

2. ส่วนคลาสประกอบไปด้วย 2 คลาสแทนค่าดังตารางที่ 3.3 โดยกำหนดให้ Y แทนการไม่เป็นโรคหลอดเลือดสมอง ส่วน N แทนการเป็นโรคหลอดเลือดสมอง

เมื่อแปลงข้อมูลทั้งหมดเสร็จเรียบร้อยแล้ว ทำการแบ่งชุดข้อมูลออกเป็น 2 ชุดข้อมูล คือ ชุดข้อมูลปัจจัยพื้นฐาน และชุดข้อมูลโรค เพื่อใช้ในการทดสอบปัจจัยที่มีผลกระทบต่อการเป็นโรคหลอดเลือดสมอง โดยรายละเอียดของข้อมูลแต่ละชุดมีดังนี้

1.ชุดข้อมูลปัจจัยพื้นฐานประกอบไปด้วย เพศ สถานภาพ การสูบบุหรี่ การดื่มสุรา การออกกำลังกาย อาชีพ ความดันโลหิต คอเลสเตอรอล และคลาส ซึ่งทำการแปลงเป็นตัวเลขแสดงใน Error! Reference source not found.และมีจำนวนข้อมูลทั้งหมด 1,000 ทรานแซคชั่น

3.3 การสร้างโมเดล

งานวิจัยนี้ทำนายโรคหลอดเลือดสมองโดยใช้วิธีการ ต้นไม้ตัดสินใจ (Decision tree) เพื่อศึกษาปัจจัยที่นำไปสู่โรคหลอดเลือดสมอง ซึ่งสามารถอธิบายดังต่อไปนี้

จะทำการคำนวนข้อมูลตัวอย่างจากข้อมูลจริงตารางต่อไปนี้บอกปัจจัยในการทำนายโรคหลอดเลือดสมอง

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 3.2 ตารางแปลงค่าปัจจัยพื้นฐาน

ลำดับ	ปัจจัย	ความหมาย
1	เพศ	10=ชาย, 11=หญิง
2	สถานภาพ	20=โสด, 21=สมรส
3	สูบบุหรี่	40=ไม่สูบบุหรี่, 42=สูบบุหรี่
4	ดื่มสุรา	50=ไม่ดื่มสุรา, 52=ดื่มสุรา
5	ออกกำลังกาย	60=ออกกำลังกาย, 61=ไม่ออกกำลังกาย
6	อาชีพ	700=พนักงานบริษัท, 701=อาชีพอิสระ, 702=รับจ้าง
7	ความดันโลหิต	80=ปกติ, 81=สูง
8	คอเลสเตอรอล	90=ปกติ, 91=สูง

ตารางที่ 3.3 การแทนค่าคลาส

ค่า	แทนค่า
ไม่เป็นโรคหลอดเลือดสมอง	Y
เป็นโรคหลอดเลือดสมอง	N

ตารางที่ 3.4 ตัวอย่างข้อมูล

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y

Computer Science Department
Faculty of Informatics Mahasarakham University

ตารางที่ 3.4 ตัวอย่างข้อมูล(ต่อ)

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

สามารถสรุปอัลกอริทึม ID3 ได้ดังภาพประกอบด้านล่าง

Entropy(S) = ∑ – p(I) . log2p(I)

Gain (S, A) = Entropy(S) – ∑ [p(S|A) . Entropy(S|A)]

Entropy

เราต้องคำนวณ Entropyก่อน คอลัมน์การตัดสินใจประกอบด้วย 10 ตัวอย่าง และมีสองป้ายกำกับ: ใช่ และไม่ใช่ มีการตัดสินใจ 5 ข้อระบุว่าใช่ และการตัดสินใจ 5 ข้อระบุว่าไม่ใช่

Entropy (Decision) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy (Decision) = – (5/10) . log2(5/10) – (5/10) . log2(5/10) = 1

ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์แรก คือ sex

เราต้องคำนวณ (Decision| sex = หญิง) and (Decision| sex = ชาย) ก่อนเป็นอันดับแรก

ตารางที่ 3.5 ตัวอย่างข้อมูล sex = หญิง

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y

Computer Science Department
Faculty of Informatics Mahasarakham University

ตารางที่ 3.5 ตัวอย่างข้อมูล sex = หญิง(ต่อ)

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

มี 5 กรณีสำหรับ หญิง การตัดสินของ 3 ข้อคือไม่ใช่ และ 2 ข้อคือใช่

Entropy(Decision| sex = หญิง) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| sex = หญิง) = – (2/5) . log2(2/5) – (3/5) . log2(3/5) = 0.971

ตารางที่ 3.6 ตัวอย่างข้อมูล sex = ชาย

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

มี 5 กรณีสำหรับ ชายการตัดสินของ 2 ข้อคือไม่ใช่ และ 3 ข้อคือใช่

Entropy(Decision| sex = ชาย) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| sex = ชาย) = – (3/5) . log2(3/5) – (2/5) . log2(2/5) = 0.971

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| sex) = 1 – (5/10)(0.971) – (5/10)(0.971) = 0.029

Computer Science Department
Faculty of Informatics Mahasarakham University

เราต้องใช้การคำนวณแบบเดียวกันสำหรับคอลัมน์อื่นๆ เพื่อค้นหาปัจจัยที่สำคัญที่สุดในการตัดสินใจ ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์ถัดไป คือ status โดยต้องคำนวณ (Decision| status = โสด) and (Decision| status = สมรส)

ตารางที่ 3.7 ตัวอย่างข้อมูล status = โสด

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

มี 4 กรณีสำหรับ โสด การตัดสินของ 1 ข้อคือไม่ใช่ และ 3 ข้อคือใช่

Entropy(Decision| **status** = โสด) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| **status** = โสด) = – (3/4) . log2(3/4) – (1/4) . log2(1/4) = 0.8113

ตารางที่ 3.8 ตัวอย่างข้อมูล status = สมรส

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y

Computer Science Department Faculty of Informatics Mahasarakham University

ตารางที่ 3.8 ตัวอย่างข้อมูล status = สมรส(ต่อ)

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y

มี 6 กรณีสำหรับ สมรสการตัดสินของ 4 ข้อคือไม่ใช่ และ 2 ข้อคือใช่

Entropy(Decision| status = สมรส) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| status = สมรส) = – (2/6) . log2(2/6) – (4/6) . log2(4/6) = 0.918

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| status) = 1 – (4/10)(0.8113) – (6/10)(0.918) = 0.125

ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์ถัดไป คือ occupation โดยต้องคำนวณ (Decision| occupation = พนักงานบริษัท), (Decision| occupation = อาชีพอิสระ) and (Decision| occupation = รับจ้าง)

ตารางที่ 3.9 ตัวอย่างข้อมูลพนักงานบริษัท

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

Computer Science Department
Faculty of Informatics, Mahasarakham University

มี 6 กรณีสำหรับ พนักงานบริษัท การตัดสินของ 4 ข้อคือไม่ใช่ และ 2 ข้อคือใช่

Entropy(Decision| occupation = พนักงานบริษัท) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| occupation = พนักงานบริษัท) = – (2/6) . log2(2/6) – (4/6) . log2(4/6) = 0.918

ตารางที่ 3.10 ตัวอย่างข้อมูลอาชีพอิสระ

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y

มี 2 กรณีสำหรับ อาชีพอิสระ การตัดสินของ 1 ข้อคือไม่ใช่ และ 1 ข้อคือใช่

Entropy(Decision| occupation = อาชีพอิสระ) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| occupation = อาชีพอิสระ) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

ตารางที่ 3.11 ตัวอย่างข้อมูลรับจ้าง

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

มี 2 กรณีสำหรับ รับจ้าง การตัดสินของ 0 ข้อคือไม่ใช่ และ 2 ข้อคือใช่

Entropy(Decision| occupation = รับจ้าง) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| occupation = รับจ้าง) = – (2/2) . log2(2/2) – (0/2) . log2(0/2) = 0

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| occupation) = 1 – (6/10)(0.918) – (2/10)(1) – (2/10)(0)= 0.2492

ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์ถัดไป คือ smoking โดยต้องคำนวณ (Decision| smoking = สูบบุหรี่บริษัท)and (Decision| smoking = ไม่สูบบุหรี่)

Computer Science Department Faculty of Informatics Mahasarakham University

ตารางที่ 3.12 ตัวอย่างข้อมูลสูบบุหรี่

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

มี 6 กรณีสำหรับ สูบบุหรี่ การตัดสินของ 2 ข้อคือไม่ใช่ และ 4 ข้อคือใช่

Entropy(Decision| smoking = สูบบุหรี่) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| smoking = สูบบุหรี่) = – (4/6) . log2(4/6) – (2/6) . log2(2/6) = 0.918

ตารางที่ 3.13 ตัวอย่างข้อมูลไม่สูบบุหรี่

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

Computer Science Department
Faculty of Informatics, Mahasarakham University

4 กรณีสำหรับ ไม่สูบบุหรี่ การตัดสินของ 3 ข้อคือไม่ใช่ และ 1 ข้อคือใช่

Entropy(Decision| smoking = ไม่สูบบุหรี่) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| smoking = ไม่สูบบุหรี่) = – (3/4) . log2(3/4) – (1/4) . log2(1/4) = 0.8113

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| smoking) = 1 – (6/10)(0.918) – (4/10)(0.8113) = 0.1247

ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์ถัดไป คือ alcohol โดยต้องคำนวณ (Decision| alcohol = ดื่มสุรา)and (Decision| alcohol = ไม่ดื่มสุรา)

ตารางที่ 3.14 ตัวอย่างข้อมูลดื่มสุรา

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y

มี 6 กรณีสำหรับ ดื่มสุรา การตัดสินของ 3 ข้อคือไม่ใช่ และ 3 ข้อคือใช่

Entropy(Decision| alcohol = ดื่มสุรา) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| alcohol = ดื่มสุรา) = – (3/6) . log2(3/6) – (3/6) . log2(3/6) = 1

Computer Science Department Faculty of Informatics, Mahasarakham University

ตารางที่ 3.15 ตัวอย่างข้อมูลไม่ดื่มสุรา

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

มี 4 กรณีสำหรับ ไม่ดื่มสุรา การตัดสินของ 2 ข้อคือไม่ใช่ และ 2 ข้อคือใช่

Entropy(Decision| alcohol = ไม่ดื่มสุรา) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| alcohol = ไม่ดื่มสุรา) = – (2/4) . log2(2/4) – (2/4) . log2(2/4) = 1

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| alcohol) = 1 – (6/10)(1) – (4/10)(1) = 0

ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์ถัดไป คือ exercise โดยต้องคำนวณ (Decision| exercise = ออกกำลังกาย)and (Decision| exercise = ไม่ออกกำลังกาย)

ตารางที่ 3.16 ตัวอย่างข้อมูลออกกำลังกาย

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

Computer Science Department
Faculty of Informatics, Mahasarakham University

มี 4 กรณีสำหรับ ออกกำลังกาย การตัดสินของ 2 ข้อคือไม่ใช่ และ 2 ข้อคือใช่

Entropy(Decision| exercise = ออกกำลังกาย) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| exercise = ออกกำลังกาย) = – (2/4) . log2(2/4) – (2/4) . log2(2/4) = 1

ตารางที่ 3.17 ตัวอย่างข้อมูลไม่ออกกำลังกาย

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y

มี 6 กรณีสำหรับ ไม่ออกกำลังกาย การตัดสินของ 3 ข้อคือไม่ใช่ และ 3 ข้อคือใช่

Entropy(Decision| exercise = ไม่ออกกำลังกาย) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| exercise = ไม่ออกกำลังกาย) = – (3/6) . log2(3/6) – (3/6) . log2(3/6) = 1

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| exercise) = 1 – (6/10)(1) – (4/10)(1) = 0

ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์ถัดไป คือ Blood pressure โดยต้องคำนวณ

(Decision| Blood pressure = สูง)and (Decision| Blood pressure = ปกติ)

ตารางที่ 3.18 ตัวอย่างข้อมูล Bloodpressure สูง

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y

Computer Science Department Faculty of Informatics Mahasarakham University

ตารางที่ 3.18 ตัวอย่างข้อมูล Bloodpressure สูง(ต่อ)

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N

มี 5 กรณีสำหรับ สูง การตัดสินของ 1 ข้อคือไม่ใช่ และ 4 ข้อคือใช่

Entropy(Decision| Blood pressure = สูง) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| Blood pressure = สูง) = – (4/5) . log2(4/5) – (1/5) . log2(1/5) = 0.7219

ตารางที่ 3.19 ตัวอย่างข้อมูล Blood pressure ปกติ

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y

มี 5 กรณีสำหรับ ปกติ การตัดสินของ 4 ข้อคือไม่ใช่ และ 1 ข้อคือใช่

Entropy(Decision| Blood pressure = ปกติ) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| Blood pressure = ปกติ) = – (1/5) . log2(1/5) – (4/5) . log2(4/5) = 0.7219

Computer Science Department
Faculty of Informatics, Mahasarakham University

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| Blood pressure) = 1 – (5/10)(0.7219) – (5/10)(0.7219) = 0.2781

ในขั้นตอนต่อไป จะทำการคำนวนหาปัจจัยของคอลัมน์ถัดไป คือ cholesterol โดยต้องคำนวณ (Decision| cholesterol = สูง)and (Decision cholesterol = ปกติ)

ตารางที่ 3.20 ตัวอย่างข้อมูล cholesterol สูง

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

มี 6 กรณีสำหรับ สูง การตัดสินของ 1 ข้อคือไม่ใช่ และ 5 ข้อคือใช่

Entropy(Decision| cholesterol = สูง) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| cholesterol = สูง) =– (5/6) . log2(5/6) – (1/6) . log2(1/6) = 0.65

ตารางที่ 3.21 ตัวอย่างข้อมูล cholesterol ปกติ

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y

Computer Science Department
Faculty of Informatics Mahasarakham University

ตารางที่ 3.21 ตัวอย่างข้อมูล cholesterol ปกติ(ต่อ)

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y

มี 4 กรณีสำหรับ ปกติ การตัดสินของ 4 ข้อคือไม่ใช่ และ 0 ข้อคือใช่

Entropy(Decision| cholesterol = ปกติ) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy(Decision| cholesterol = ปกติ) = – (0/4) . log2(0/4) – (4/4) . log2(4/4) = 0

โดยแทนค่าในสูตรดังต่อไปนี้

Gain(Decision| cholesterol) = 1 – (6/10)(0.65) – (4/10)(0) = 0.61

ดังที่เห็น ปัจจัยแนวโน้มในการตัดสินใจทำให้เกิดคะแนนสูงสุด นั่นเป็นเหตุผลที่การตัดสินใจของ cholesterol จะปรากฏในโหนดรูทของทรี

Gain(Decision| sex) = 0.029

Gain(Decision| status) = 0.125

Gain(Decision| occupation) = 0.2492

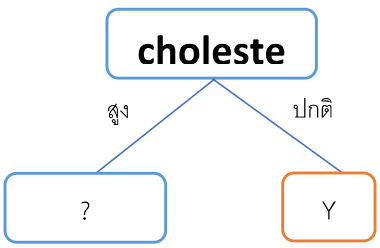
Gain(Decision| smoking) = 0.1247

Gain(Decision| alcohol) = 0

Gain(Decision| exercise) = 0

Gain(Decision| Blood pressure) = 0.2781

Gain(Decision| cholesterol) = 0.61



ภาพประกอบที่ 3.2 cholesterol

ทดสอบชุดข้อมูลสำหรับชุดย่อยที่กำหนดเองของแอตทริบิวต์ cholesterol ในรอบที่ 2 Normal cholesterol on decision

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 3.22 ตัวอย่างข้อมูล cholesterol ในรอบที่ 2

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	ปกติ	Y
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	ปกติ	ปกติ	Y
ชาย	โสด	อาชีพอิสระ	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	ปกติ	Y

พื้นฐานแล้ว การตัดสินใจมักจะไม่ใช่เสมอหากคอเรสเตอรอลปกติ High cholesterol on decision

ตารางที่ 3.23 ตัวอย่างข้อมูล คอเรสเตอรอลสูง

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	สูง	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	สูง	N

Computer Science Department Faculty of Informatics, Mahasarakham University

โดยเริ่มจากการหา Entropy ของ high

Entropy (Decision, cholesterol = สูง) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy (Decision, cholesterol = สูง) = – (5/6) . log2(5/6) – (1/6) . log2(1/6) = 0.65

ตารางที่ 3.24 ตัวอย่างข้อมูล คอเรสเตอรอลสูง sex = หญิง

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

Entropy (Decision, cholesterol = สูง|sex = หญิง) = – (0/2) . log2(0/2) – (2/2) . log2(2/2) = 0

Entropy (Decision, cholesterol = สูง|sex = ชาย) = – (3/4) . log2(3/4) – (1/4) . log2(1/4) = 0.8113

Gain(Decision, cholesterol = สูง| sex) = 0.65 – (2/6)(0) – (4/6)(0.8113) = 0.1091

ตารางที่ 3.25 ตัวอย่างข้อมูล cholesterol = สูง | status = สมรส

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

Entropy (Decision, cholesterol = สูง| status = สมรส) = – (2/3) . log2(2/3) – (1/3) . log2(1/3) = 0.9174

Entropy (Decision, cholesterol = สูง| status = โสด) = – (3/3) . log2(3/3) – (0/3) . log2(0/3) = 0

Computer Science Department Faculty of Informatics, Mahasarakham University

Gain(Decision, cholesterol =สูง| status) = 0.65 – (3/6)(0.9174) – (3/6)(0) = 0.1913

ตารางที่ 3.26 ตัวอย่างข้อมูล cholesterol = สูง | occupation = อาชีพอิสระ

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

Entropy (Decision, cholesterol = สูง| occupation = พนักงาน) = – (2/3) . log2(2/3) – (1/3) . log2(1/3) = 0.9174

Entropy (Decision, cholesterol = สูง| occupation = อาชีพอิสระ) = – (1/1) . log2(1/1) – (0/1) . log2(0/1) = 0

Entropy (Decision, cholesterol = สูง| occupation = รับจ้าง) = – (2/2) . log2(2/2) – (0/2) . log2(0/2) = 0

Gain(Decision, cholesterol =สูง| occupation) = 0.65 – (3/6)(0.9174) – (1/6)(0) - (2/6)(0) = 0.1913

ตารางที่ 3.27 ตัวอย่างข้อมูล cholesterol = สูง | smoking = สูบบุหรี่

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

Entropy (Decision, cholesterol = สูง| smoking = สูบบุหรี่) = – (4/5) . log2(4/5) – (1/5) . log2(1/5) = 0.7219

Computer Science Department Faculty of Informatics, Mahasarakham University

Entropy (Decision, cholesterol = สูง| smoking = ไม่สูบบุหรี่) = – (0/1) . log2(0/1) – (1/1) . log2(1/1) = 0

Gain(Decision, cholesterol =สูง| smoking) = 0.65 – (5/6)(0. 7219) – (1/6)(0) = 0.0484

ตารางที่ 3.28 ตัวอย่างข้อมูล cholesterol = สูง | alcohol = ไม่ดื่มสุรา

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	cholesterol
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

Entropy (Decision, cholesterol = สูง| alcohol = ดื่มสุรา) = – (3/4) . log2(3/4) – (1/4) . log2(1/4) = 0.8113

Entropy (Decision, cholesterol = สูง| alcohol = ไม่ดื่มสุรา) = – (2/2) . log2(2/2) – (0/2) . log2(0/2) = 0

Gain(Decision, cholesterol =สูง| alcohol) = 0.65 – (4/6)(0.8113) – (2/6)(0) = 0.1091

ตารางที่ 3.29 ตัวอย่างข้อมูล cholesterol = สูง | exercise = ไม่ออก

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

Entropy (Decision, cholesterol = สูง| exercise = ไม่ออก) = – (3/4) . log2(3/4) – (1/4) . log2(1/4) = 0.8113

Computer Science Department Faculty of Informatics, Mahasarakham University

Entropy (Decision, cholesterol = สูง| exercise = ออก) = – (2/2) . log2(2/2) – (0/2) . log2(0/2) = 0

Gain(Decision, cholesterol =สูง| exercise) = 0.65 – (4/6)(0.8113) – (2/6)(0) = 0.1091

ตารางที่ 3.30 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

Entropy (Decision, cholesterol = สูง|BP = ปกติ) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

Entropy (Decision, cholesterol = สูง|BP = สูง) = – (4/4) . log2(4/4) – (0/4) . log2(0/4) = 0

Gain(Decision, cholesterol =สูง| BP) = 0.65 – (2/6)(1) – (4/6)(0) = 0.3167

ปัจจัยแนวโน้มในการตัดสินใจทำให้เกิดคะแนนสูงสุด

Gain(Decision, cholesterol = สูง| sex) = 0.1091

Gain(Decision, cholesterol = สูง| status) = 0.1913

Gain(Decision, cholesterol = สูง| occupation) = 0.1913

Gain(Decision, cholesterol = สูง| smoking) = 0.0484

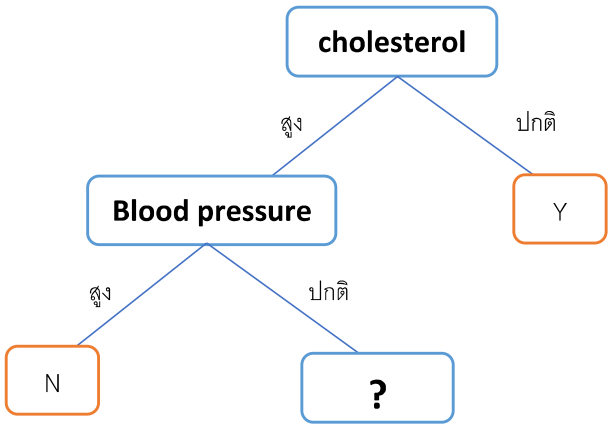
Gain(Decision, cholesterol = สูง| alcohol) = 0.1091

Gain(Decision, cholesterol = สูง| exercise) = 0.1091

Gain(Decision, cholesterol = สูง| BP) = 0.3167

Blood pressure คือแนวโน้มในการเป็นโรคหลอดสมอง เมื่อ cholesterol สูง เมื่อ cholesterol สูง และ Blood pressureสูง จะเป็นโรคหลอกเลือดสมอง

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 3.3 cholesterol สูง

ทดสอบชุดข้อมูลสำหรับชุดย่อยที่กำหนดเองของแอตทริบิวต์ Blood pressure ในรอบที่ 3 high cholesterol on decision

ตารางที่ 3.31 ตัวอย่างข้อมูล Blood pressure ในรอบที่ 3

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	value
หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออกกำลังกาย	สูง	N
หญิง	โสด	รับจ้าง	ไม่สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	สูง	N
ชาย	โสด	พนักงานบริษัท	สูบบุหรี่	ไม่ดื่มสุรา	ออกกำลังกาย	สูง	N

โดยพื้นฐานแล้ว การตัดสินใจมักจะใช่เสมอหากความดันโลหิตสูง normal cholesterol on decision

ตารางที่ 3.32 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ

sex	status	occupation	smoking	alcohol	exercise	Blood pressure	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	ปกติ	Y

โดยเริ่มจากการหา Entropy ของ normal

Entropy (Decision, cholesterol = สูง| BP = ปกติ) = – p(Yes) . log2p(Yes) – p(No) . log2p(No)

Entropy (Decision, cholesterol = สูง| BP = ปกติ) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 3.33 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ | sex = ชาย

sex	status	occupation	smoking	alcohol	exercise	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	Y

Entropy (Decision, cholesterol = สูง| BP = ปกติ | sex = ชาย) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

Entropy (Decision, cholesterol = สูง| BP = ปกติ | sex = หญิง) = – (0/0) . log2(0/0) – (0/0) . log2(0/0) = 0

Gain(Decision, cholesterol =สูง| | BP = ปกติ | sex) = 1 – (2/2)(1) – (0/2)(0) = 0

ตารางที่ 3.34 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ | status = สมรส

sex	status	occupation	smoking	alcohol	exercise	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	Y

Entropy (Decision, cholesterol = สูง| BP = ปกติ | status = สมรส) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

Entropy (Decision, cholesterol = สูง| BP = ปกติ | status = โสด) = – (0/0) . log2(0/0) – (0/0) . log2(0/0) = 0

Gain(Decision, cholesterol =สูง| | BP = ปกติ | status) = 1 – (2/2)(1) – (0/2)(0) = 0

ตารางที่ 3.35 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ | occu = พนักงาน

sex	status	occupation	smoking	alcohol	exercise	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	Y

Entropy (Decision, cholesterol = สูง| BP = ปกติ | occu = พนักงาน) = – (0/1) . log2(0/1) – (1/1) . log2(1/1) = 0

Entropy (Decision, cholesterol = สูง| BP = ปกติ | occu = อาชีพอิสระ) = – (1/1) . log2(1/1) – (0/1) . log2(0/1) = 0

Computer Science Department
Faculty of Informatics, Mahasarakham University

Entropy (Decision, cholesterol = สูง| BP = ปกติ | occu = รับจ้าง) = – (0/0) . log2(0/0) – (0/0) . log2(0/0) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | occu) = 1 – (1/2)(0) – (1/2)(0) – (0/2)(0) = 1

ตารางที่ 3.36 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ | smoking = สูบบุหรี่

sex	status	occupation	smoking	alcohol	exercise	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	Y

Entropy (Decision, cholesterol = สูง| BP = ปกติ | smoking = สูบบุหรี่) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

Entropy (Decision, cholesterol = สูง| BP = ปกติ | smoking = ไม่สูบ) = – (0/0) . log2(0/0) – (0/0) . log2(0/0) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | smoking) = 1 – (2/2)(1) – (0/2)(0) = 0

ตารางที่ 3.37 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ | alcohol = ดื่มสุรา

sex	status	occupation	smoking	alcohol	exercise	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	Y

Entropy (Decision, cholesterol = สูง| BP = ปกติ | alcohol = ดื่มสุรา) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

Entropy (Decision, cholesterol = สูง| BP = ปกติ | alcohol = ไม่ดื่ม) = – (0/0) . log2(0/0) – (0/0) . log2(0/0) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | alcohol) = 1 – (2/2)(1) – (0/2)(0) = 0

ตารางที่ 3.38 ตัวอย่างข้อมูล cholesterol = สูง | BP = ปกติ | exercise

sex	status	occupation	smoking	alcohol	exercise	value
ชาย	สมรส	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	N
ชาย	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ไม่ออกกำลังกาย	Y

Entropy (Decision, cholesterol = สูง| BP = ปกติ | exercise = ไม่ออก) = – (1/2) . log2(1/2) – (1/2) . log2(1/2) = 1

Computer Science Department
Faculty of Informatics Mahasarakham University

Entropy (Decision, cholesterol = สูง| BP = ปกติ | exercise = ออก) = – (0/0) . log2(0/0) – (0/0) . log2(0/0) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | exercise) = 1 – (2/2)(1) – (0/2)(0) = 0

ดังนั้น ปัจจัยที่เป็นแนวโน้มในการเป็นโรคหลอดเลือดสมอง ทำให้เกิดคะแนนสูงสุด จะปรากฏใน รูทโหนดของทรีดังนี้

Gain(Decision, cholesterol =สูง| | BP = ปกติ | sex) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | status) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | occupation) = 1

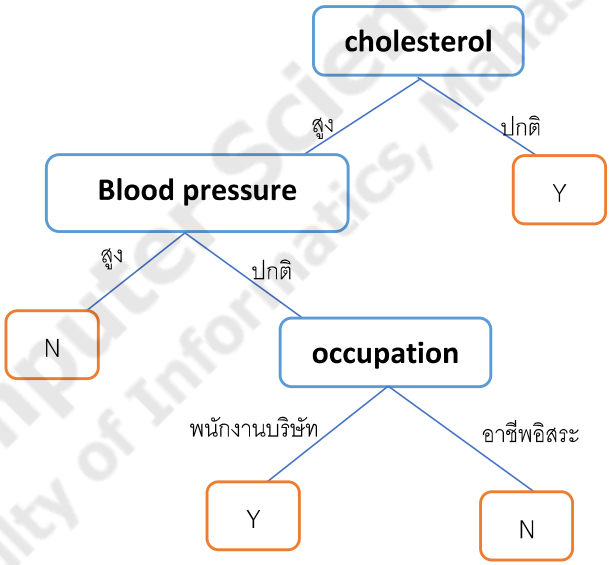
Gain(Decision, cholesterol =สูง| | | BP = ปกติ | smoking) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | alcohol) = 0

Gain(Decision, cholesterol =สูง| | | BP = ปกติ | exercise) = 0

คือการตัดสินใจ เพราะมันให้คะแนนสูงสุดหากแนวโน้มมี cholesterol สูง และ Blood pressure ปกติ ณ จุดนี้ การทำนายจะเกิดขึ้นหากมี Blood pressure สูง

ดังนั้น การสร้างต้นไม้เพื่อการตัดสินใจจึงสิ้นสุดลง เราสามารถใช้กฎต่อไปนี้ในการตัดสินใจ



ภาพประกอบที่ 3.4 ตัดสินใจ cholesterol สูง และ Blood pressure ปกติ occupation

กฎที่ได้จากต้นไม้ตัดสินใจ

If (cholesterol==high & Blood pressure==high) {การทำนายโรคหลอดเลือดสมอง = yes}

If (cholesterol == high & Blood pressure==normal & occupation == company employee) {การทำนายโรคหลอดเลือดสมอง = No}

If (cholesterol == high & Blood pressure==normal & occupation == freelance)

Computer Science Department
Faculty of Informatics, Mahasarakham University

{การทำนายโรคหลอดเลือดสมอง = yes}

If (cholesterol== normal) {การทำนายโรคหลอดเลือดสมอง = No}

ภาพประกอบที่ 3.5 การทำนายโรคหลอดเลือดสมอง

3.4 การทำนายและการประเมินผล

แบ่งข้อมูลชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบด้วยวิธีการแบบ Hold-out validation คือ การแบ่งข้อมูลออกเป็น 2 ส่วนตามอัตราส่วนที่กำหนด โดยจะต้องทำการสุ่มข้อมูลก่อนทำการแบ่ง เช่น จำนวนข้อมูลมีทั้งหมด 1000 คน แบ่งชุดข้อมูลเรียนรู้เป็น 70% จากข้อมูลทั้งหมดและแบ่งข้อมูลทดสอบเป็น 30% จากข้อมูลทั้งหมด หลังจากนั้นจะนำข้อมูลทดสอบไปทดสอบกับกฎที่สร้างขึ้นโดยต้นไม้ตัดสินใจ สมมติข้อมูลดังตารางต่อไปนี้ ตารางที่ 3.5 ตัวอย่างชุดข้อมูลทดสอบ และกฎที่ได้มีดังนี้

กฎข้อที่ 1 :: If (cholesterol==high) and (Blood pressure==high) then yes

กฎข้อที่ 2 :: If (cholesterol == high) and (Blood pressure==normal) and (occupation == company employee) then No

กฎข้อที่ 3 :: If (cholesterol == high) and (Blood pressure==normal) and (occupation== freelance) then yes

กฎข้อที่ 4 :: If cholesterol== normal then No

ตารางที่ 3.39 ตัวอย่างชุดข้อมูลทดสอบ

ID	เพศ	สถานภาพ	อาชีพ	สูบบุหรี่	ดื่มสุรา	ออกกำลังกาย	BP	คอเลสเตอรอล	คลาส
1	หญิง	สมรส	พนักงานบริษัท	สูบบุหรี่	ดื่มสุรา	ออก	สูง	สูง	เป็น
2	หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ดื่มสุรา	ไม่ออก	ปกติ	ปกติ	ไม่เป็น
3	ชาย	โสด	รับจ้าง	สูบบุหรี่	ไม่ดื่มสุรา	ออก	สูง	สูง	เป็น
4	ชาย	โสด	อาชีพอิสระ	สูบบุหรี่	ดื่มสุรา	ไม่ออก	ปกติ	สูง	ไม่เป็น
5	หญิง	สมรส	พนักงานบริษัท	ไม่สูบบุหรี่	ไม่ดื่มสุรา	ไม่ออก	ปกติ	ปกติ	ไม่เป็น

นำข้อมูลแถวแรกของตารางนี้ไปพิจารณากับกฎ โดยพิจารณาจากค่าคอเลสเตอรอลที่มีค่าเท่ากับสูง และ ค่าBPมีค่าเท่ากับสูง จะได้ผลการทำนายว่าเป็นโรคหลอดเลือดสมองซึ่งตรงกับกฎข้อที่ 2 โดยเขียนลง

Computer Science Department Faculty of Informatics, Mahasarakham University

ตารางช่องที่ P11 ส่วนข้อมูลแถวที่ 2 ของตารางไปพิจารณากับกฎ โดยพิจารณาจากค่าคอเลสเตอรอลที่มีค่าเท่ากับปกติ จะได้ผลการทำนายว่าไม่เป็นโรคหลอดเลือดสมองซึ่งตรงกับกฎข้อที่ 1 โดยเขียนลงตารางช่องที่ P22 ส่วนข้อมูลแถวที่ 3 ของตารางไปพิจารณากับกฎ โดยพิจารณาจากค่าคอเลสเตอรอลที่มีค่าเท่ากับสูง และค่าBPมีค่าเท่ากับสูง จะได้ผลการทำนายว่าเป็นโรคหลอดเลือดสมองซึ่งตรงกับกฎข้อที่ 2 โดยเขียนลงตารางช่องที่ P11 ส่วนข้อมูลแถวที่ 4 ของตารางไปพิจารณากับกฎ โดยพิจารณาจากค่าคอเลสเตอรอลที่มีค่าเท่ากับสูง ค่าBPมีค่าเท่ากับปกติ และ ทำอาชีพค้าขาย จะได้ผลการทำนายว่าเป็นไม่โรคหลอดเลือดสมองซึ่งไม่ตรงกับกฎข้อใด โดยเขียนลงตารางช่องที่ P21 ส่วนข้อมูลแถวที่ 5 ของตารางไปพิจารณากับกฎ โดยพิจารณาจากค่าคอเลสเตอรอลที่มีค่าเท่ากับปกติ จะได้ผลการทำนายว่าไม่เป็นโรคหลอดเลือดสมองซึ่งตรงกับกฎข้อที่ 1 โดยเขียนลงตารางช่องที่ P22 แสดงดังตารางที่ 3.40 ผลการทำนาย

ตารางที่ 3.40 ตารางผลการทำนาย

		ผลการทำนาย	
		c_1	c_2
ค่าความจริง	c_1	$p_{11} = 2$	$p_{12} = 0$
	c_2	$p_{21} = 1$	$p_{22} = 2$

โดย c1 คือ เป็นโรคหลอดเลือดสมอง

C2 คือ ไม่เป็นโรคหลอดเลือกสมอง

P11 คือ จำนวนข้อมูลทดสอบที่ทำนายถูกต้องว่าเป็นโรคหลอดเลือดสมอง

P12 คือ จำนวนข้อมูลทดสอบที่ทำนายว่าไม่เป็นโรคหลอดเลือดสมองแต่ความจริงเป็นโรคหลอดเลือดสมอง

P21 คือ จำนวนข้อมูลทดสอบที่ทำนายว่าเป็นโรคหลอดเลือดสมองแต่ความจริงไม่เป็นโรคหลอดเลือดสมอง

P22 คือ จำนวนข้อมูลทดสอบที่ทำนายถูกต้องว่าไม่เป็นโรคหลอดเลือดสมอง

การหาค่าความถูกต้อง

$$AC = \frac{p_{11} + p_{22}}{p_{11} + p_{12} + p_{21} + p_{22}}$$

$$AC = \frac{2 + 2}{2 + 0 + 1 + 2}$$

$$AC = \frac{4}{5} = 0.8$$

สรุปได้ว่าค่าความถูกต้องในการทำนายของตัวจำแนก คือ 0.8 หรือ 80% นั้นเอง

Computer Science Department
Faculty of Informatics, Mahasarakham University

การหาค่าความแม่นยำ

$$precision_{c_1} = \frac{p_{11}}{p_{11} + p_{21}}$$

$$precision_{c_1} = \frac{2}{2+1} = \frac{2}{3} = 0.67$$

$$precision_{c_2} = \frac{p_{22}}{p_{12} + p_{22}}$$

$$precision_{c_2} = \frac{2}{0+2} = \frac{2}{2} = 1$$

สามารถสรุปได้ว่า

- ค่าความแม่นยำในการทำนายคลาส C1 มีค่าเท่ากับ 0.67 หรือ 67%
- ค่าความแม่นยำในการทำนายคลาส C2 มีค่าเท่ากับ 1.00 หรือ 100%
- แสดงให้เห็นว่าตัวจำแนกมีความแม่นยำในการทำนายคลาส C_2 มากที่สุด

การหาค่าระลึก

$$recall_{c_1} = \frac{p_{11}}{p_{11} + p_{12}}$$

$$recall_{c_1} = \frac{2}{2+0} = \frac{2}{2} = 1$$

$$recall_{c_2} = \frac{p_{22}}{p_{21} + p_{22}}$$

$$recall_{c_2} = \frac{2}{1+2} = \frac{2}{3} = 0.67$$

สามารถสรุปได้ว่า

- ค่าระลึกในการทำนายคลาส C1 มีค่าเท่ากับ 1.00 หรือ 100%
- ค่าระลึกในการทำนายคลาส C2 มีค่าเท่ากับ 0.67 หรือ 67%
- แสดงให้เห็นว่าตัวจำแนกมีความสามารถในการทำนายคลาส C_1 ได้ดีที่สุด

การหาค่าประสิทธิภาพ

$$F - Measure_{c_1} = \frac{2 \times precision_{c_1} \times recall_{c_1}}{(precision_{c_1} + recall_{c_1})}$$

$$F - Measure_{c_1} = \frac{2 \times 0.67 \times 1}{(0.67 + 1)} = \frac{1.34}{1.67} = 0.8$$

Computer Science Department
Faculty of Informatics Mahasarakham University

$$F - Measure_{c_2} = \frac{2 \times precision_{c_2} \times recall_{c_2}}{(precision_{c_2} \times recall_{c_2})}$$

$$F - Measure_{c_2} = \frac{2 \times 1 \times 0.67}{(1 + 0.67)} = \frac{1.34}{1.67} = 0.8$$

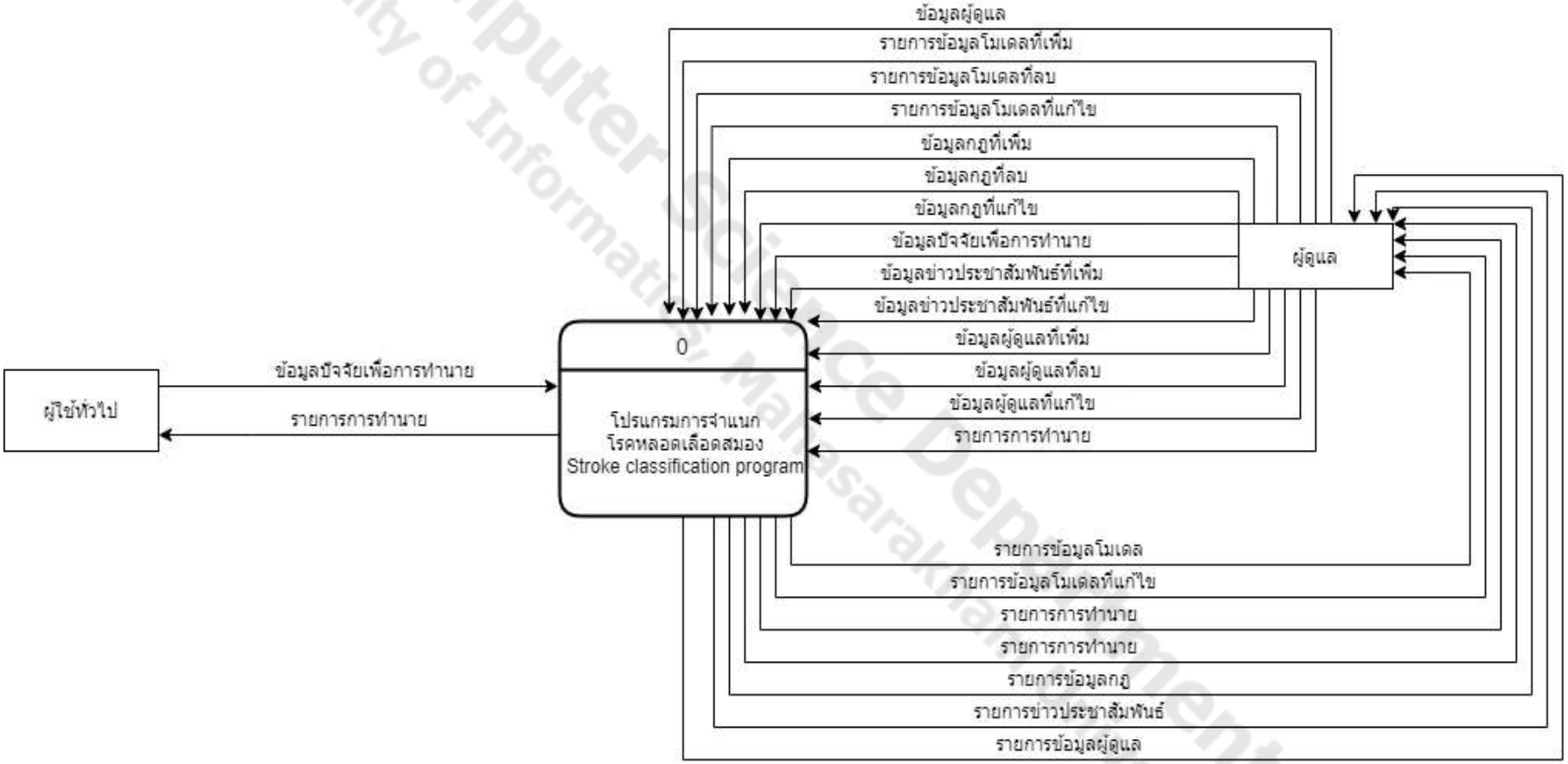
สามารถสรุปได้ว่า

- ค่าประสิทธิภาพโดยรวม ในการทำนายคลาส C1 มีค่าเท่ากับ 0.8 หรือ 80%
- ค่าประสิทธิภาพโดยรวมในการทำนายคลาส C2 มีค่าเท่ากับ 0.8 หรือ 80%
- แสดงให้เห็นว่าตัวจำแนกมีประสิทธิภาพโดยรวมในการทำนายคลาส C_1 และคลาส C_2 มีค่าเท่ากัน

Computer Science Department
Faculty of Informatics, Mahasarakham University

3.5 แผนภาพบริบท (Context Diagram)

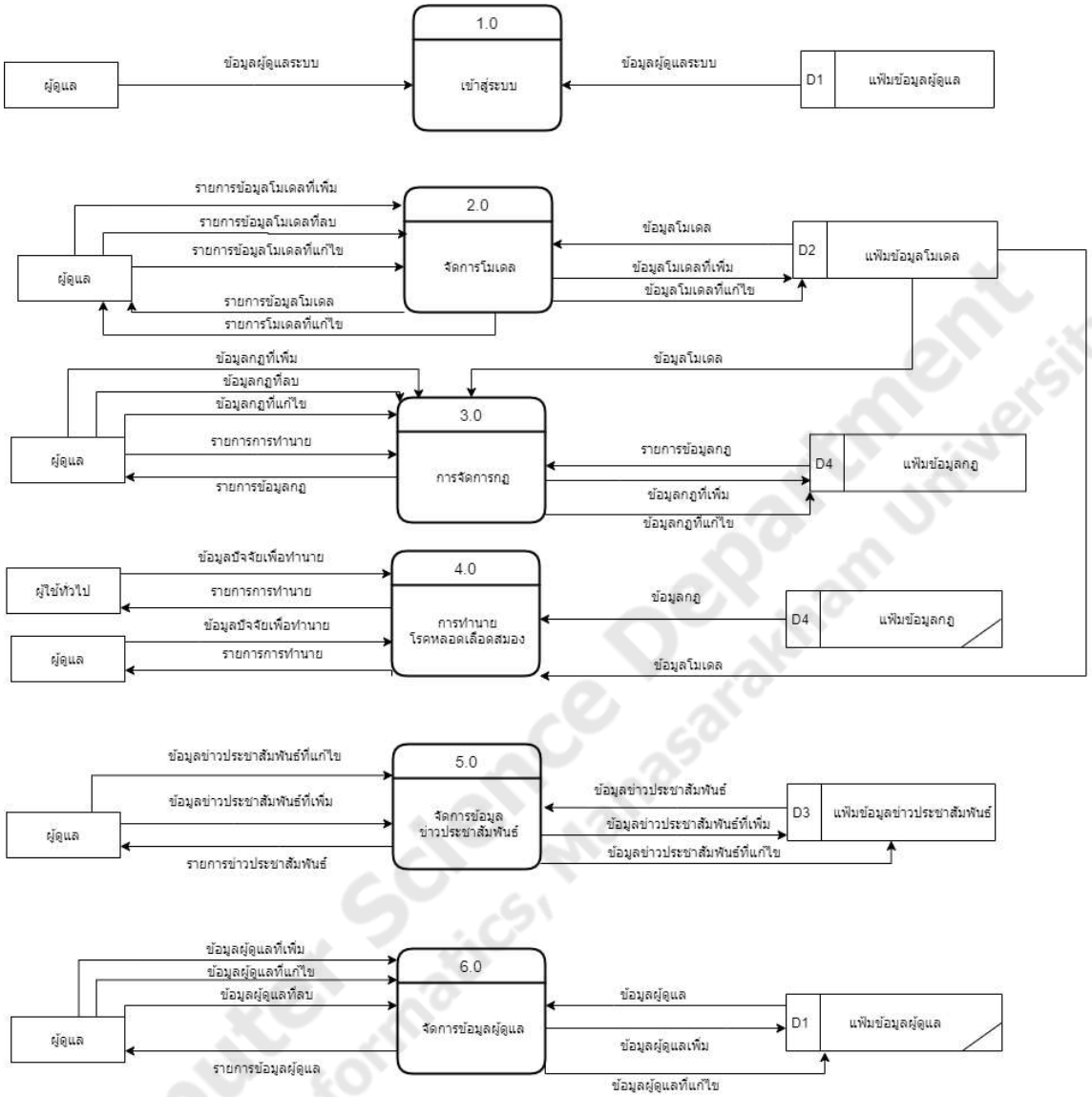
Context Diagram คือ แผนภาพกระแสข้อมูลที่แสดงภาพรวมของการทำงานระบบที่มีความสัมพันธ์ของผู้ใช้งาน แสดงถึงขอบเขตของระบบบริหารจัดการหน่วยวิจัย : กรณีศึกษาคณะสิ่งแวดล้อมและทรัพยากรศาสตร์ดังภาพประกอบที่ 3-2 แผนภาพบริบทของระบบ



ภาพประกอบที่ 3.6 แผนภาพบริบทของระบบ

Computer Science Department
Faculty of Informatics, Mahasarakham University

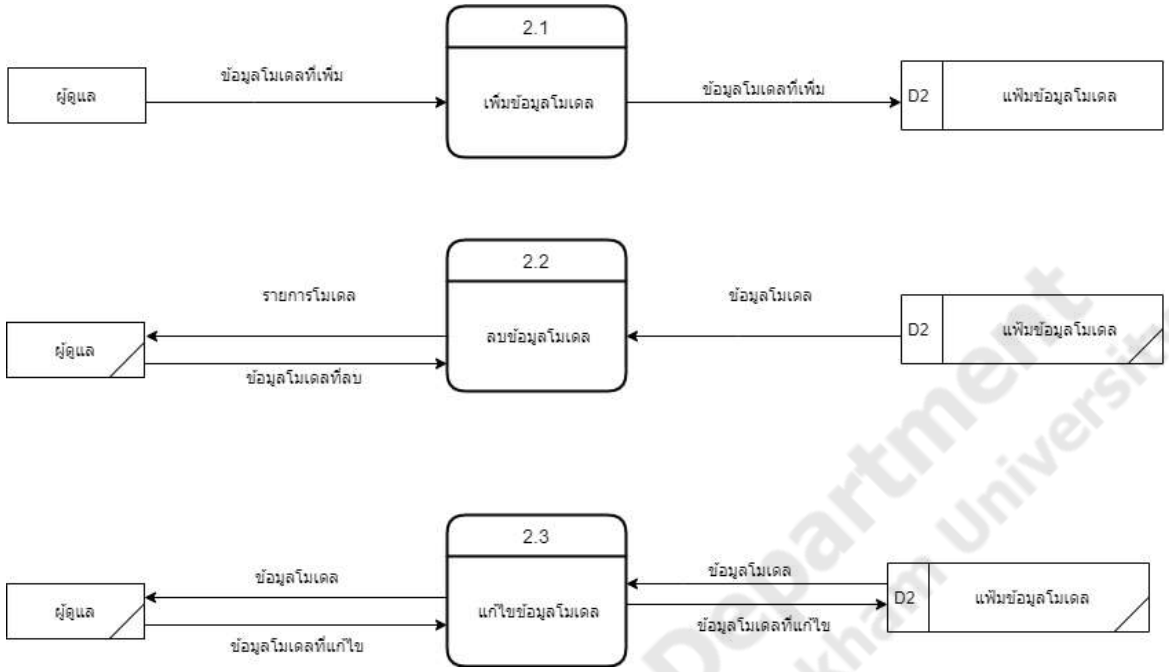
3.6 แผนภาพการไหลของข้อมูล (Data Flow Diagram)



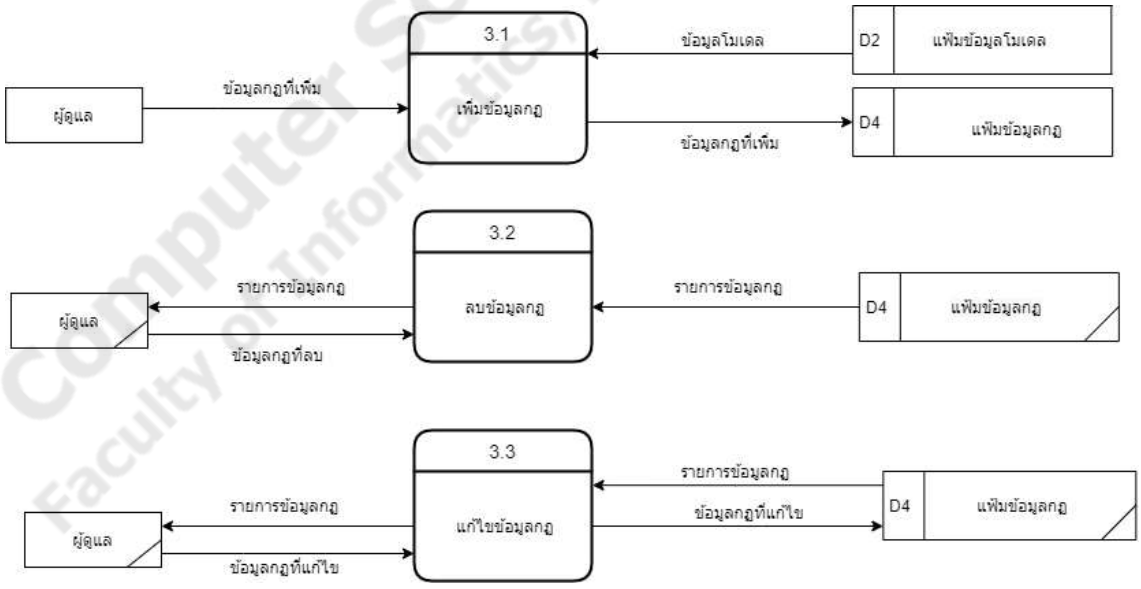
ภาพประกอบที่ 3.7 Data Flow Diagram Level 1

Computer Science Department
Faculty of Informatics, Mahasarakham University

Level 2

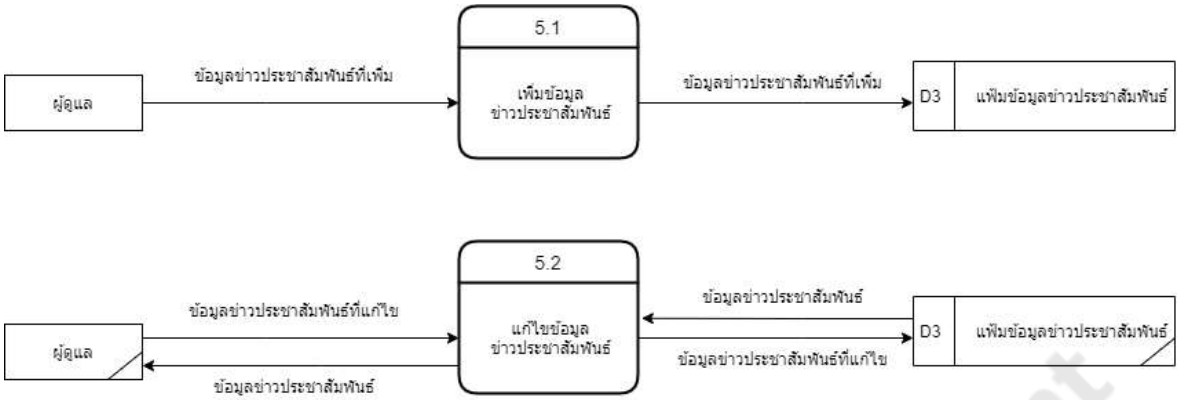


ภาพประกอบที่ 3.8 Data Flow Diagram Level 2 process 2.0 จัดการโมเดล

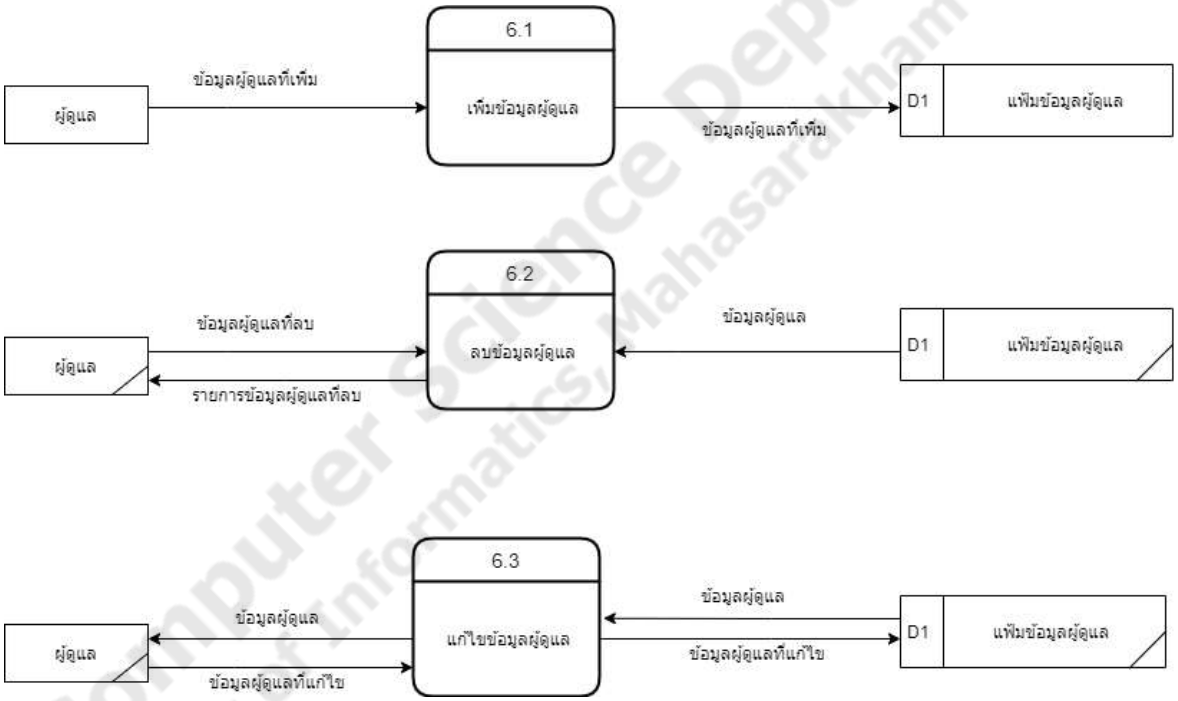


ภาพประกอบที่ 3.9 Data Flow Diagram Level 2 process 3.0 จัดการข้อมูลกฎ

Computer Science Department
Faculty of Informatics, Mahasarakham University



ภาพประกอบที่ 3.10 Data Flow Diagram Level 2 process 5.0 จัดการข่าวประชาสัมพันธ์



ภาพประกอบที่ 3.11 Data Flow Diagram Level 2 process 6.0 จัดการข้อมูลผู้ดูแล

Computer Science Department
Faculty of Informatics, Mahasarakham University

3.7 External Entity Description

พจนานุกรมข้อมูลช่วยอธิบายการทำงานของแผนภาพการไหลของข้อมูล (Data Flow Dictionaries) ดัง ตารางที่ 3.7 External Entity Description

ตารางที่ 3.41 External Entity Description

Name	Description	Input Data Flow	Output Data Flow
ผู้ใช้ทั่วไป	บุคคลที่ไม่ได้เป็นเจ้าของระบบ	- ข้อมูลปัจจัยเพื่อการทำนาย	- รายการการทำนาย
ผู้ดูแล	เป็นผู้ดูแลระบบ	- ข้อมูลผู้ดูแล - รายการข้อมูลโมเดลที่เพิ่ม - รายการข้อมูลโมเดลที่ลบ - รายการข้อมูลโมเดลที่แก้ไข - ข้อมูลกฎที่เพิ่ม - ข้อมูลกฎที่ลบ - ข้อมูลกฎที่แก้ไข - รายการการทำนาย - ข้อมูลปัจจัยเพื่อการทำนาย ข้อมูลข่าวประชาสัมพันธ์ที่เพิ่ม - ข้อมูลข่าวประชาสัมพันธ์ที่แก้ไข - ข้อมูลผู้ดูแลที่เพิ่ม - ข้อมูลผู้ดูแลที่ลบ - ข้อมูลผู้ดูแลที่แก้ไข	- รายการข้อมูลโมเดล - รายการข้อมูลโมเดลที่แก้ไข - รายการการทำนาย - รายการข้อมูลกฎ - รายการข่าวประชาสัมพันธ์ - รายการข้อมูลผู้ดูแล

Computer Science Department
Faculty of Informatics, Mahasarakham University

3.8 Data Flow (Data Flow Description and Data Structure of Data Flow)

เป็นขั้นตอนการทำงานของระบบซึ่งทำให้เราทราบถึงการรับ-ส่งข้อมูลแสดงถึงการไหลของข้อมูลทั้งข้อมูลเข้า (Input) และข้อมูลส่งออก (Output) ระหว่างข้อมูลต้นทางถึงข้อมูลปลายทางโดยอธิบายข้อมูลและขั้นตอนในการดำเนินงาน ดังตารางที่ 3.8 Data Flow

ตารางที่ 3.42 Data Flow

Name	Description	Source	Destination	Data Structure
ข้อมูลผู้ดูแล	ข้อมูลผู้ดูแล	ผู้ดูแล	Process 1.0	ชื่อผู้ใช้ + รหัสผ่าน
		D1 แฟ้มข้อมูลผู้ดูแล	Process 1.0	
ข้อมูลโมเดลที่เพิ่ม	ข้อมูลโมเดลที่ต้องการเพิ่มในแฟ้มจัดการโมเดล	ผู้ดูแลระบบ	Process 2.1 เพิ่มข้อมูลโมเดล	สถานะการใช้งาน + วันที่สร้าง + ชื่อโมเดล + รหัสโมเดล + ค่าความถูกต้อง + ค่า recall C1 + ค่า recall C2 + Precision C1 + Precision C2 + F-Measure 1 + F-Measure 1
		Process 2.1 เพิ่มข้อมูลโมเดล	D2 แฟ้มข้อมูลโมเดล	
ข้อมูลโมเดลที่ลบ	โมเดลที่จะลบ	ผู้ดูแล	Process 2.2 ลบข้อมูลโมเดล	รหัสโมเดล
ข้อมูลโมเดล , รายการโมเดล	ข้อมูลโมเดล , รายการโมเดล	D2 แฟ้มข้อมูลโมเดล	Process 2.2 ลบข้อมูลโมเดล	สถานะการใช้งาน + วันที่สร้าง + ชื่อโมเดล + รหัสโมเดล + ค่าความถูกต้อง + ค่า recall C1 + ค่า recall C2 + Precision C1 + Precision C2 + F-Measure 1 + F-Measure 1
		Process 2.2 ลบข้อมูลโมเดล	ผู้ดูแล	

Computer Science Department Faculty of Informatics, Mahasarakham University

ตารางที่ 3.42 Data Flow(ต่อ)

Name	Description	Source	Destination	Data Structure
ข้อมูลโมเดลที่แก้ไข	ข้อมูลโมเดลที่แก้ไขแล้ว	ผู้ดูแล	Process 2.3 แก้ไขข้อมูลโมเดล	สถานะการใช้งาน + วันที่สร้าง + ชื่อโมเดล + รหัสโมเดล + ค่าความถูกต้อง + ค่า recall C1 + ค่า recall C2 + Precision C1 + Precision C2 + F-Measure 1 + F-Measure 1
		Process 2.3 แก้ไขข้อมูลโมเดล	D2 แฟ้มข้อมูลโมเดล	
ข้อมูลโมเดล	ข้อมูลโมเดล	D2 แฟ้มข้อมูลโมเดล	Process 2.3 แก้ไขข้อมูลโมเดล	สถานะการใช้งาน + วันที่สร้าง + ชื่อโมเดล + รหัสโมเดล + ค่าความถูกต้อง + ค่า recall C1 + ค่า recall C2 + Precision C1 + Precision C2 + F-Measure 1 + F-Measure 1
		Process 2.3 แก้ไขข้อมูลโมเดล	ผู้ดูแล	
ข้อมูลกฎที่เพิ่ม	ข้อมูลกฎ	ผู้ดูแล	Process 3.1 เพิ่มข้อมูลกฎ	รหัสโมเดล + รหัสกฎ + ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา + เพศ + สถานภาพ + คอเลสเตอรอล + Class
		Process 3.1 เพิ่มข้อมูลกฎ	D4 แฟ้มข้อมูลกฎ	
ข้อมูลโมเดล	ข้อมูลโมเดล	D2แฟ้มข้อมูลโมเดล	Process 3.1เพิ่มข้อมูลกฎ	
ข้อมูลกฎที่ลบ	ข้อมูลกฎ	ผู้ดูแล	Process 3.2 ลบข้อมูลกฎ	รหัสกฎ
รายการข้อมูลกฎ	รายการข้อมูลกฎ	Process 3.2 ลบข้อมูลกฎ	ผู้ดูแล	รหัสกฎ + รหัสโมเดล
		D4 แฟ้มข้อมูลกฎ	Process 3.2 ลบข้อมูลกฎ	

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 3.42 Data Flow(ต่อ)

Name	Description	Source	Destination	Data Structure
ข้อมูลกฎ	ข้อมูลกฎ	D4 แฟ้มข้อมูลกฎ	Process 3.2 ลบข้อมูลกฎ	รหัสโมเดล + รหัสกฎ + ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา+เพศ + สถานภาพ + คอเลสเตอรอล + Class
รายการข้อมูลกฎ	รายการข้อมูลกฎ	Process 3.2 ลบข้อมูลกฎ	ผู้ดูแล	รหัสกฎ + รหัสโมเดล
ข้อมูลกฎที่แก้ไข	ข้อมูลกฎที่ต้องการแก้ไข	ผู้ดูแล	Process 3.3 ลบข้อมูลกฎ	รหัสโมเดล + รหัสกฎ + ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา+เพศ + สถานภาพ + คอเลสเตอรอล + Class
		Process 3.2 ลบข้อมูลกฎ	D4 แฟ้มข้อมูลกฎ	
ข้อมูลปัจจัยเพื่อทำนาย	ข้อมูลปัจจัยในการทำนาย	ผู้ใช้ทั่วไป	Process 4.0 การทำนายโรคหลอดเลือดสมอง	ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา +เพศ+ สถานภาพ + คอเลสเตอรอล + Class
รายการการทำนาย	ผลการทำนาย	Process 4.0 การทำนายโรคหลอดเลือดสมอง	ผู้ใช้ทั่วไป	ผลการทำนาย
ข้อมูลปัจจัยเพื่อการทำนาย	ข้อมูลปัจจัยในการทำนาย	ผู้ดูแล	Process 4.0 การทำนายโรคหลอดเลือดสมอง	ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา +เพศ+ สถานภาพ + คอเลสเตอรอล + Class
รายการการทำนาย	ผลการทำนาย	Process 4.0 การทำนายโรคหลอดเลือดสมอง	ผู้ดูแล	ผลการทำนาย
ข้อมูลกฎ	ข้อมูลกฎ	D4 แฟ้มข้อมูลกฎ	Process 4.0 การทำนายโรคหลอดเลือดสมอง	ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา +เพศ+ สถานภาพ + คอเลสเตอรอล + Class

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 3.42 Data Flow(ต่อ)

Name	Description	Source	Destination	Data Structure
ข้อมูลกฎ	ข้อมูลกฎ	D4 แฟ้มข้อมูลกฎ	Process 4.0 การทำนายโรคหลอดเลือดสมอง	ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา +เพศ+ สถานภาพ + คอเลสเตอรอล + Class
ข้อมูลโมเดล	ข้อมูลโมเดล	D2แฟ้มข้อมูลโมเดล	Process 4.0 การทำนายโรคหลอดเลือดสมอง	ความดันโลหิต+อาชีพ+ ออกกำลังกาย+สูบุหรี่ + ดื่มสุรา+เพศ+สถานภาพ + คอเลสเตอรอล + Class
ข้อมูลกฎ	ข้อมูลกฎ	D4 แฟ้มข้อมูลกฎ	Process 3.2 ลบข้อมูลกฎ	รหัสโมเดล+รหัสกฎ+ ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา+เพศ + สถานภาพ + คอเลสเตอรอล + Class
ข้อมูลโมเดล	ข้อมูลโมเดล	D2แฟ้มข้อมูลโมเดล	Process 4.0 เพิ่มข้อมูลกฎ	ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา + เพศ + สถานภาพ + คอเลสเตอรอล + Class
ข้อมูลข่าวประชาสัมพันธ์ที่เพิ่ม	ข้อมูลประชาสัมพันธ์ที่ต้องการเพิ่มในแฟ้มประชาสัมพันธ์	ผู้ดูแล	Process 5.1 เพิ่มข้อมูลข่าวประชาสัมพันธ์	รหัส + รายละเอียด + วันที่
		Process 5.1 เพิ่มข้อมูลข่าวประชาสัมพันธ์	D3 แฟ้มข้อมูลข่าวประชาสัมพันธ์	
ข้อมูลข่าวประชาสัมพันธ์	ข้อมูลประชาสัมพันธ์	D3 แฟ้มข้อมูลข่าวประชาสัมพันธ์	Process 5.2 แก้ไขข้อมูลข่าวประชาสัมพันธ์	ข้อมูลข่าวประชาสัมพันธ์

Computer Science Department Faculty of Informatics, Mahasarakham University

ตารางที่ 3.42 Data Flow(ต่อ)

Name	Description	Source	Destination	Data Structure
		Process 5.2 แก้ไขข้อมูลข่าวประชาสัมพันธ์	ผู้ดูแล	
ข้อมูลข่าวประชาสัมพันธ์ที่แก้ไข	ข้อมูลข่าวประชาสัมพันธ์ที่ต้องการแก้ไข	ผู้ดูแล	Process 5.2 แก้ไขข้อมูลข่าวประชาสัมพันธ์	
		Process 5.2 แก้ไขข้อมูลข่าวประชาสัมพันธ์	D3 แฟ้มข้อมูลข่าวประชาสัมพันธ์	
ข้อมูลผู้ดูแลที่เพิ่ม	ข้อมูลผู้ดูแลที่เพิ่มและเก็บลงแฟ้มข้อมูลผู้ดูแล	ผู้ดูแล	Process 6.1 เพิ่มข้อมูลผู้ดูแล	รหัส User + Username + Password + ชื่อ + นามสกุล + ที่อยู่ + เบอร์ติดต่อ + สภานะการใช้งาน
		Process 6.1 เพิ่มข้อมูลผู้ดูแล	ผู้ดูแล	
ข้อมูลผู้ดูแล	ข้อมูลผู้ดูแล	D1 แฟ้มข้อมูลผู้ดูแล	Process 6.2 ลบข้อมูลผู้ดูแล	รหัส User + Username + Password + ชื่อ + นามสกุล + ที่อยู่ + เบอร์ติดต่อ + สภานะการใช้งาน
รายการผู้ดูแล	รายการผู้ดูแล	Process 6.2 ลบข้อมูลผู้ดูแล	ผู้ดูแล	
ข้อมูลผู้ดูแลที่ลบ	ผู้ดูแลที่จะลบ	ผู้ดูแล	Process 6.2 ลบข้อมูลผู้ดูแล	รหัสUser
ข้อมูลผู้ดูแล	ข้อมูลผู้ดูแล	D1 แฟ้มข้อมูลผู้ดูแล	Process 6.3 แก้ไขข้อมูลผู้ดูแล	รหัส User + Username + Password + ชื่อ + นามสกุล + ที่อยู่ + เบอร์ติดต่อ + สภานะการใช้งาน

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 3.42 Data Flow(ต่อ)

Name	Description	Source	Destination	Data Structure
		Process 6.3 แก้ไขข้อมูลผู้ดูแล	ผู้ดูแล	
ข้อมูลผู้ดูแลที่แก้ไข	ข้อมูลผู้ดูแลที่ต้องการแก้ไข	ผู้ดูแล	Process 6.3 แก้ไขข้อมูลผู้ดูแล	
		Process 6.3 แก้ไขข้อมูลผู้ดูแล	D1 แฟ้มข้อมูลผู้ดูแล	

3.9 Data Store (Data Flow Description and Data Structure of Data Store)

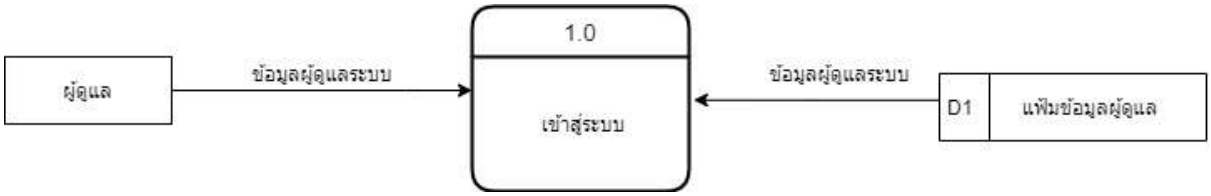
Data Store คือ การนำข้อมูลเข้าไปจัดการในฐานข้อมูล โดยมีการแยกออกเป็นแฟ้มข้อมูล ดังตารางที่ 3.9 Data Store

ตารางที่ 3.43 Data Store

Data Store Description		
Id	**Data store**	**Structure**
D1	แฟ้มข้อมูลผู้ดูแล	รหัส User + Username + Password + ชื่อ + นามสกุล + ที่อยู่ + เบอร์ติดต่อ + สภานะการใช้งาน
D2	แฟ้มข้อมูลโมเดล	รหัสโมเดล + สถานะการใช้งาน + วันที่สร้าง + ชื่อโมเดล + ค่าความถูกต้อง + ค่า recall C1 + ค่า recall C2 + Precision C1 + Precision C2 + F-Measure 1 + F-Measure 1
D3	แฟ้มข้อมูลประชาสัมพันธ์	รหัส + หัวข้อข่าว + วันที่ + รายละเอียด
D4	แฟ้มข้อมูลกฎ	รหัสกฎ + ความดันโลหิต + อาชีพ + ออกกำลังกาย + สูบุหรี่ + ดื่มสุรา + เพศ + สถานภาพ + คอเลสเตอรอล + Class

Computer Science Department Faculty of Informatics, Mahasarakham University

3.10 คำอธิบายการประมวลผล (Process Description Specifications)

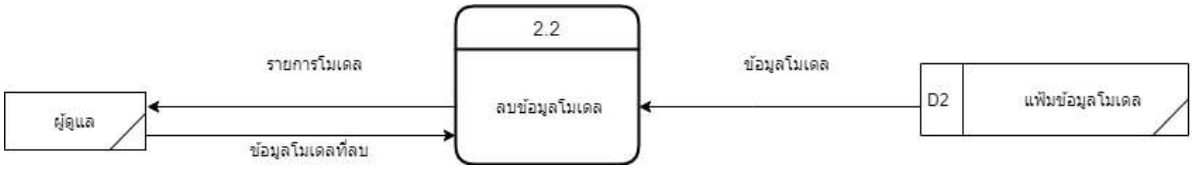


ID	Process 1.0
NAME	เข้าสู่ระบบ
DESCRIPTION	ตรวจสอบว่าชื่อสมาชิกเป็นสมาชิกของระบบเข้าสู่ระบบสำเร็จ ไม่มี ข้อความแจ้ง
INPUT DATA FLOWS	- ข้อมูลผู้ดูแลระบบ - ข้อมูลผู้ดูแลระบบ
OUTPUT DATA FLOWS	
PROCESS DESCRIPTION	เริ่มต้น 1. รับข้อมูลผู้ดูแลระบบ(ชื่อผู้ใช้+รหัสผ่าน) 2. ตรวจสอบข้อมูลผู้ใช้ 2.1 ถ้าการป้อนข้อมูลผู้ดูแลครบถ้วน และถูกต้องให้ไป 3. 2.2 ถ้าป้อนข้อมูลผู้ดูแลไม่ถูก “แสดง ป้อนข้อมูล ผู้ดูแลให้ถูกต้อง ” ไปที่ 1 3. ตรวจสอบข้อมูลผู้ดูแล ในแฟ้มข้อมูลผู้ดูแล 3.1 ถ้ามีข้อมูลในแฟ้มข้อมูลผู้ดูแลให้แสดง “เข้าสู่ระบบแล้ว” 3.2 ถ้าไม่มีข้อมูลให้แสดง “กรุณาตรวจสอบ รหัสผู้ใช้ หรือ รหัสผ่านให้ถูกต้อง” จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University

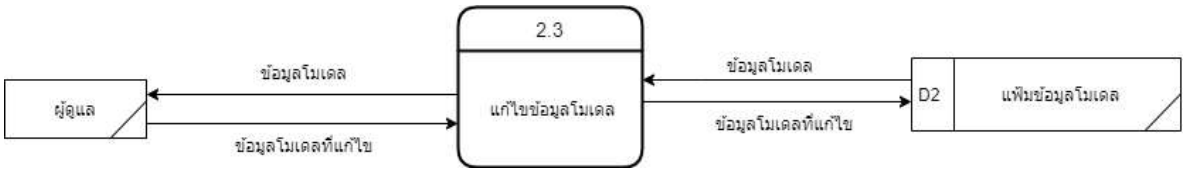
ID	Process 2.1
NAME	เพิ่มข้อมูลโมเดล
DESCRIPTION	เพิ่มข้อมูลไฟล์เพื่อสร้างโมเดล
INPUT DATA FLOWS	- ข้อมูลโมเดลที่เพิ่ม
OUTPUT DATA FLOWS	-ข้อมูลโมเดลที่เพิ่ม
PROCESS DESCRIPTION	เริ่มต้น 1. รับข้อมูลโมเดลที่ต้องการเพิ่ม 2. เพิ่มข้อมูลโมเดลลงใน แฟ้มข้อมูลโมเดล 3. ตรวจสอบว่ามีชื่อและรหัสโมเดลอยู่ใน แฟ้มข้อมูลโมเดล แล้วหรือไม่ 3.1 ถ้ามีชื่อและรหัสโมเดลใน แฟ้มข้อมูลโมเดล แสดง “มีชื่อหรือรหัสโมเดลอยู่แล้ว”ไปข้อ 1. 3.2 ถ้าไม่มีชื่อและรหัสใน แฟ้มข้อมูล ให้บันทึกลงใน แฟ้มข้อมูลโมเดล จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University



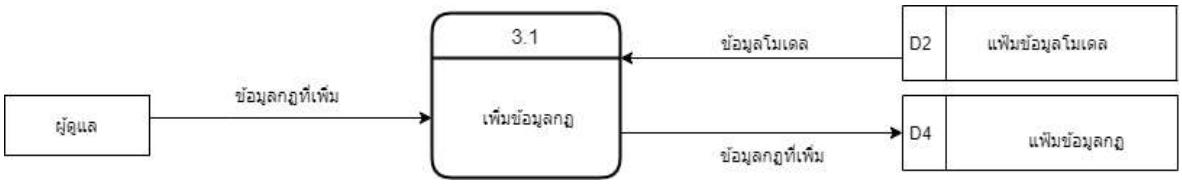
ID	Process 2.2
NAME	ลบข้อมูลโมเดล
DESCRIPTION	ผู้ดูแลลบข้อมูลโมเดล
INPUT DATA FLOWS	-ข้อมูลโมเดลที่ลบ -ข้อมูลโมเดล
OUTPUT DATA FLOWS	-รายการโมเดล
PROCESS DESCRIPTION	เริ่มต้น 1. แสดงรายการโมเดลที่จะลบ 2. ผู้ดูแลเลือกรายการโมเดลที่ต้องการลบ แสดง “ยืนยันการลบ” 3. ทำการลบ ข้อมูลโมเดลใน แฟ้มข้อมูลจัดการโมเดล 4. แสดงข้อมูล “ลบข้อมูลโมเดลสำเร็จ” จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University



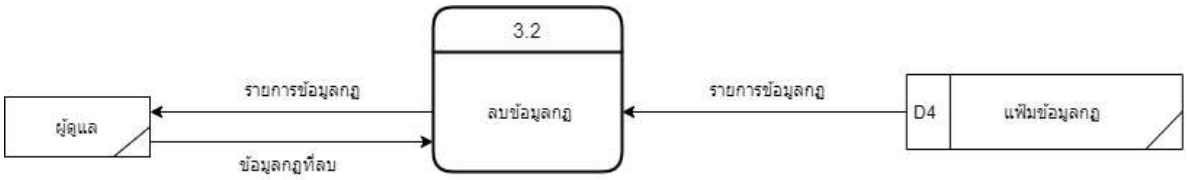
ID	Process 2.3
NAME	แก้ไขข้อมูลโมเดล
DESCRIPTION	ผู้ดูแลแก้ไขข้อมูลโมเดล
INPUT DATA FLOWS	ข้อมูลโมเดลที่แก้ไข ข้อมูลโมเดล
OUTPUT DATA FLOWS	ข้อมูลโมเดล ข้อมูลโมเดลที่แก้ไข
PROCESS DESCRIPTION	เริ่มต้น 1. แสดงโมเดลที่ต้องการแก้ไข 2. ผู้ดูแลเลือกข้อมูลโมเดลที่แก้ไข 2.1 ถ้าผู้ดูแลแก้ไขข้อมูลโมเดลครบถ้วนและถูกต้องทุกรายการให้บันทึกลง แฟ้มข้อมูลจัดการโมเดล 2.2 ถ้าไม่ครบถ้วน แสดง “ข้อมูลไม่ถูกต้อง” ไปที่ข้อ 2 จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University



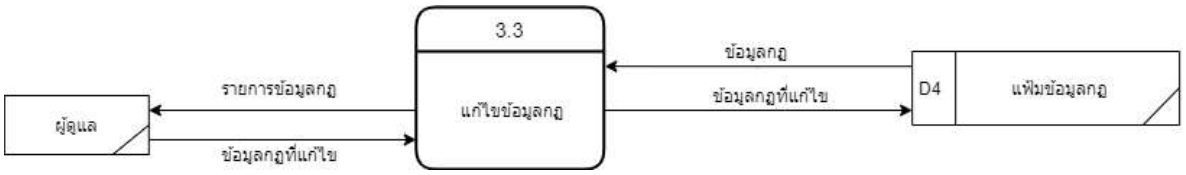
ID	Process 3.1
NAME	เพิ่มข้อมูลกฎ
DESCRIPTION	ผู้ดูแลเพิ่มข้อมูลกฎ
INPUT DATA FLOWS	ข้อมูลกฎที่เพิ่ม ข้อมูลโมเดล
OUTPUT DATA FLOWS	ข้อมูลกฎที่เพิ่ม
PROCESS DESCRIPTION	เริ่มต้น 1. แสดงข้อมูลโมเดล 2. รับข้อมูลกฎที่ต้องการเพิ่ม 3. เพิ่มข้อมูลกฎลงใน แฟ้มข้อมูลกฎ 4. ตรวจสอบว่ามีรหัสกฎอยู่ใน แฟ้มข้อมูลจัดการกฎแล้วหรือไม่ 4.1 ถ้ามีรหัสกฎใน แฟ้มข้อมูลจัดการกฎ แสดง “มีรหัสกฎอยู่แล้ว”ไปข้อ 1. 4.2 ถ้าไม่รหัสกฎใน แฟ้มข้อมูล ให้บันทึกลงใน แฟ้มข้อมูลกฎ จบการทำงาน

Computer Science Department
Faculty of Informatics Mahasarakham University



ID	Process 3.2
NAME	ลบข้อมูลกฎ
DESCRIPTION	ผู้ดูแลลบข้อมูลกฎ
INPUT DATA FLOWS	รายการข้อมูลกฎ ข้อมูลกฎที่ลบ
OUTPUT DATA FLOWS	รายการข้อมูลกฎ
PROCESS DESCRIPTION	เริ่มต้น 1. แสดงรายการข้อมูลกฎที่จะลบ 2. ผู้ดูแลเลือกรายการข้อมูลกฎที่ต้องการลบ แสดง “ยืนยันการลบ” 2.1 ถ้าตกลง ไปข้อ 3 2.2 ถ้ายกเลิก ไปข้อ 1 3. ทำการลบ ข้อมูลกฎใน แฟ้มข้อมูลกฎ 4. แสดงข้อมูล “ลบข้อมูลกฎสำเร็จ” จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University



ID	Process 3.3
NAME	แก้ข้อมูลกฎ
DESCRIPTION	ผู้ดูแลแก้ไขข้อมูลกฎ
INPUT DATA FLOWS	ข้อมูลกฎ ข้อมูลกฎที่แก้ไข
OUTPUT DATA FLOWS	ข้อมูลกฎที่แก้ไข รายการข้อมูลกฎ
PROCESS DESCRIPTION	เริ่มต้น 1. แสดงกฎที่ต้องการแก้ไข 2. ผู้ดูแลเลือกข้อมูลกฎที่ต้องการแก้ไข 2.1 ถ้าผู้ดูแลแก้ไขข้อมูลกฎครบถ้วนและถูกต้องทุกรายการให้บันทึกลง แฟ้มข้อมูลกฎ 2.2 ถ้าข้อมูลไม่ครบถ้วน ให้แสดงข้อความ “กรุณาป้อนข้อมูลให้ครบ “ กลับไป 1. จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University

ID	Process 5.1
NAME	เพิ่มข้อมูลข่าวประชาสัมพันธ์
DESCRIPTION	ผู้ดูแลเพิ่มข่าวประชาสัมพันธ์
INPUT DATA FLOWS	ข้อมูลข่าวประชาสัมพันธ์ที่เพิ่ม
OUTPUT DATA FLOWS	ข้อมูลข่าวประชาสัมพันธ์ที่เพิ่ม
PROCESS DESCRIPTION	เริ่มต้น 1. รับข้อมูลข่าวประชาสัมพันธ์ที่ต้องการเพิ่ม 2. เพิ่มข้อมูลข่าวประชาสัมพันธ์ลงใน แฟ้มข้อมูลข่าวประชาสัมพันธ์ 3. ตรวจสอบว่ามีรหัสอยู่ใน แฟ้มข้อมูลข่าวประชาสัมพันธ์แล้วหรือไม่ 3.1 ถ้ามีรหัสใน แฟ้มข้อมูลข่าวประชาสัมพันธ์แสดง “มีรหัสอยู่แล้ว”ไปข้อ 1. 3.2 ถ้าไม่มีรหัสใน แฟ้มข้อมูล ให้บันทึกลงใน แฟ้มข้อมูลข่าวประชาสัมพันธ์ จบการทำงาน

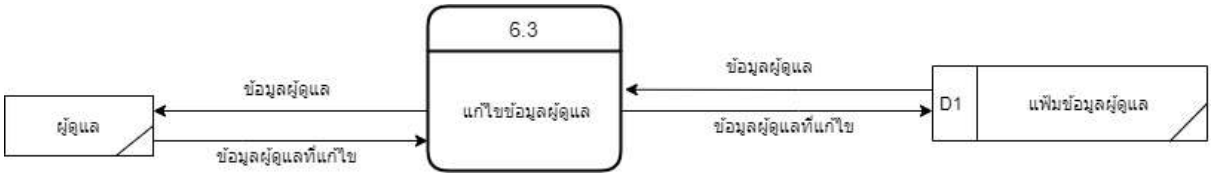
Computer Science Department
Faculty of Informatics Mahasarakham University

ID	Process 5.2
NAME	แก้ไขข้อมูลข่าวประชาสัมพันธ์
DESCRIPTION	ก่อนการแก้ไขข่าวประชาสัมพันธ์ แสดงข้อมูลของประชาสัมพันธ์ จากนั้นผู้ดูแลสามารถแก้ไขข่าวประชาสัมพันธ์ตามที่ต้องการ
INPUT DATA FLOWS	- ข้อมูลข่าวประชาสัมพันธ์ที่แก้ไข - ข้อมูลข่าวประชาสัมพันธ์
OUTPUT DATA FLOWS	- ข้อมูลข่าวประชาสัมพันธ์แก้ไข - ข้อมูลข่าวประชาสัมพันธ์
PROCESS DESCRIPTION	เริ่มต้น 1. แสดงข้อมูลข่าวประชาสัมพันธ์ 2. ผู้ดูแลเลือกข้อมูลข่าวประชาสัมพันธ์ที่ต้องการแก้ไข 3. ตรวจสอบข้อมูลว่าครบถ้วนและถูกต้องหรือไม่ 3.1 ถ้าป้อนข้อมูลข่าวประชาสัมพันธ์ครบถ้วน และถูกต้องทุกรายการให้ บันทึกข้อมูลลง แฟ้มข้อมูลประชาสัมพันธ์ 3.2 ถ้าข้อมูลไม่ครบถ้วน ให้แสดงข้อความ “กรุณาป้อนข้อมูลให้ครบ “ กลับไป 1. จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University

ID	Process 6.1
NAME	เพิ่มผู้ดูแลระบบ
DESCRIPTION	Adminสามารถเพิ่มผู้ดูแลระบบเข้าสู่ฐานข้อมูล
INPUT DATA FLOWS	- ข้อมูลผู้ดูแลที่เพิ่ม
OUTPUT DATA FLOWS	- ข้อมูลผู้ดูแลที่เพิ่ม
PROCESS DESCRIPTION	เริ่มต้น 1. รับข้อมูลผู้ดูแลที่ต้องการเพิ่ม 2. ตรวจสอบว่ามีข้อมูลครบถ้วนและถูกต้องหรือไม่ 2.1 ถ้าข้อมูลครบถ้วน และถูกต้องต้องทุกรายการให้ไปทำ 3. 2.2 ถ้าข้อมูลไม่ครบถ้วน ให้แสดงข้อความ “กรุณา ป้อนข้อมูลให้ครบ “กับไป 1. 3. ตรวจสอบใน แฟ้มข้อมูลผู้ดูแล ว่ามีข้อมูลผู้ดูแลแล้วหรือไม่ 3.1 ถ้ามี ชื่อผู้ดูแลแล้ว แสดงข้อมูลเตือน “พบข้อมูลซ้ำในระบบ “ กลับ 1. 3.2 ถ้าไม่มีชื่อผู้ดูแลในแฟ้มข้อมูล ให้บันทึกข้อมูลผู้ดูแลลงในแฟ้มข้อมูลผู้ดูแล จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University

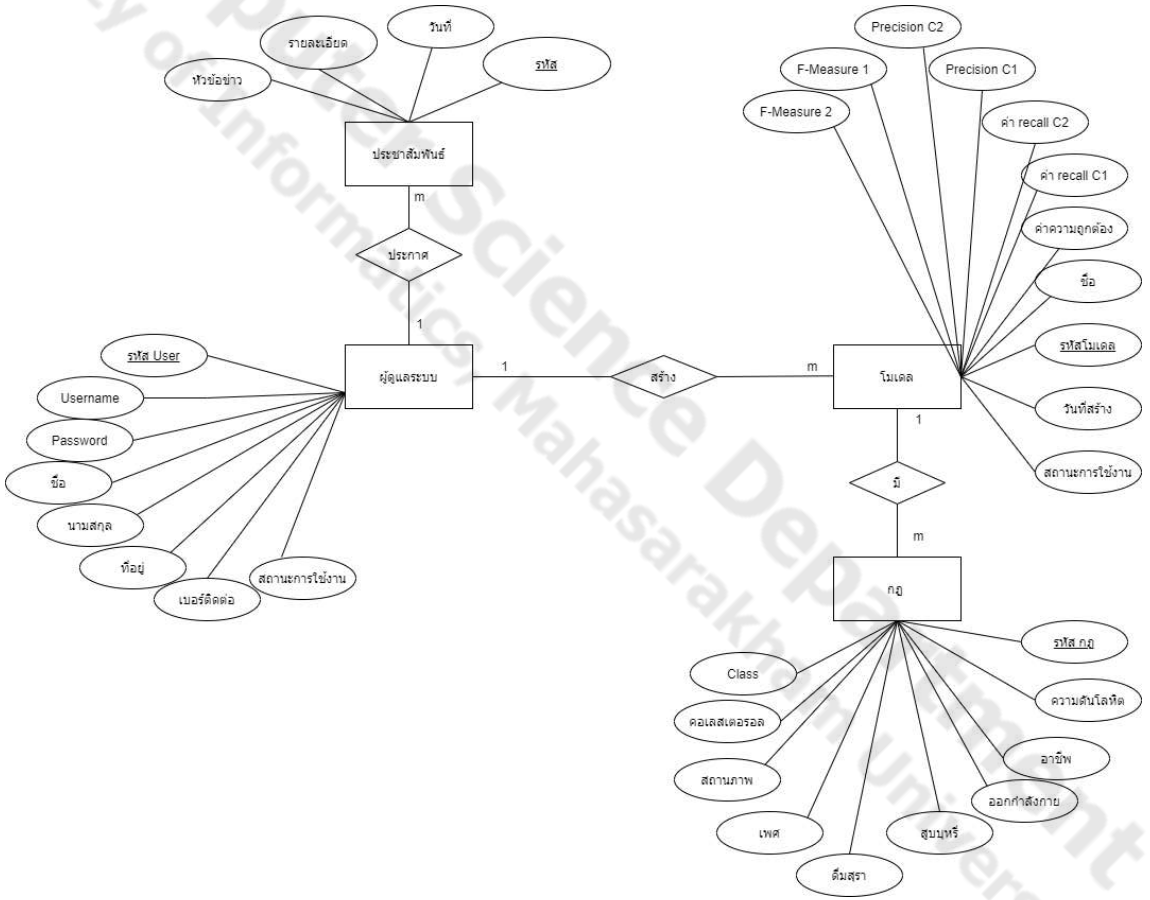


ID	Process 6.3
NAME	แก้ไขข้อมูลผู้ดูแล
DESCRIPTION	ผู้ดูแลสามารถแก้ไขข้อมูลผู้ดูแลได้
INPUT DATA FLOWS	ข้อมูลผู้ดูแลที่แก้ไข ข้อมูลผู้ดูแล
OUTPUT DATA FLOWS	ข้อมูลผู้ดูแล ข้อมูลผู้ดูแลที่แก้ไข
PROCESS DESCRIPTION	เริ่มต้น 1. แสดงข้อมูลผู้ดูแล 2. ผู้ดูแลป้อนข้อมูล ผู้ดูแลที่แก้ไข 3. ตรวจสอบว่ามีข้อมูลครบถ้วนและถูกต้องหรือไม่ 3.1 ถ้าข้อมูลที่ต้องการแก้ไขครบถ้วน และถูกต้องต้องทุกรายการให้บันทึกข้อมูลผู้ดูแลที่ต้องการแก้ไข ลงในบันทึกลงแฟ้มข้อมูลผู้ดูแล 3.2 ถ้าข้อมูลไม่ครบถ้วน ให้แสดงข้อความ “กรุณาป้อนข้อมูลให้ครบ “ กลับไป 1. จบการทำงาน

Computer Science Department
Faculty of Informatics, Mahasarakham University

3.11 แผนภาพ Entity Relationship Diagram (ER-Diagram)

แบบจำลองที่ใช้อธิบายโครงสร้างของฐานข้อมูลที่ใช้ในเว็บไซต์ซึ่งเขียนออกมาใน ลักษณะของรูปภาพการอธิบายโครงสร้างและความสัมพันธ์ของข้อมูลดัง ภาพประกอบที่ 3.8 แผนภาพ ER Diagram



ภาพประกอบที่ 3.12 แผนภาพ ER Diagram

Computer Science Department
Faculty of Informatics Mahasarakham University

3.12 การออกแบบฐานข้อมูล (Database Design)

ตารางที่ 3.44 ผู้ดูแลระบบ

Attribute	type	Description	Example Data	Constraint
ID	int(11)	รหัส user	1	pk
Username	varchar(100)	ชื่อผู้ใช้	pumipat12345	Not null
password	varchar(30)	รหัสผ่าน	123456789	Not null
Name	varchar(100)	ชื่อ	จักพงษ์	Not null
Last Name	varchar(100)	นามสกุล	หรรษา	Not null
Address	varchar(100)	ที่อยู่ผู้ดูแล	56,หมู่ 8 บ้านดอนเวียง, ต.ท่าขอนยาง อ.กันทรวิชัย จ.มหาสารคาม 44150	Not null
Phone	varchar(10)	เบอร์ติดต่อ	0911828291	Not null
Status	varchar(50)	สถานะ	(0:ยกเลิกการใช้งาน, 1: ใช้งาน)	Not null

ตารางที่ 3.45 ประชาสัมพันธ์

Attribute	type	Description	Example Data	Constraint
ID	int(50)	รหัสประชาสัมพันธ์	1	Pk(auto run)
Adminid	Int(50)	รหัสผู้ดูแล	1	fk
Date	Date	วันที่ประชาสัมพันธ์	10/02/2554	Not noll
Topic	Varchar(100)	หัวข้อข่าวประชาสัมพันธ์	โรคหลอดเลือดสมองหรือโรคอัมพฤกษ์ อัมพาต	Not null

Computer Science Department
Faculty of Informatics, Mahasarakham University

ตารางที่ 3.45 ประชาสัมพันธ์(ต่อ)

Attribute	type	Description	Example Data	Constraint
Detail	text()	รายละเอียดประชาสัมพันธ์	โรคหลอดเลือดสมองหรือโรคอัมพฤกษ์อัมพาต คือภาวะที่สมองขาดเลือดไปเลี้ยง ทำให้เกิดอาการชาที่ใบหน้า ปากเบี้ยว พูดไม่ชัด แขน ขา ข้างใดข้างหนึ่ง อ่อนแอ เคลื่อนไหวไม่ได้หรือเคลื่อนไหวลำบาก อย่างทันทีทันใด เป็นนาทีหรือเป็นชั่วโมง	Not null

ตารางที่ 3.46 โมเดล

Attribute	type	Description	Example Data	Constraint
ID	Int(50)	รหัสโมเดล	1	pk
AdminID	Int(50)	รหัสผู้ดูแล	1	fk
Date	Date	วันที่สร้างโมเดล	10/02/2554	Not noll
Name	varchar(100)	ชื่อโมเดล	Decision tree	Not null
Status	varchar(50)	สถานะ	(0:ไม่ใช้งาน, 1:	Status
Accuracy	float	ค่าความถูกต้อง	0.8	Not null
recallC1	float	ค่า recall C1	0.8	Not null
recall C2	float	ค่า recall C2	0.74	Not null
PrecisionC1	float	ค่า Precision C1	0.82	Not null

Computer Science Department Faculty of Informatics Mahasarakham University

ตารางที่ 3.46 โมเดล(ต่อ)

Attribute	type	Description	Example Data	Constraint
PrecisionC2	float	ค่า Precision C	0.98	Not null
F-Measure1	float	ค่า F-Measure 1	0.83	Not null
F-Measure2	float	ค่า F-Measure 1	0.1	Not null

ตารางที่ 3.47 กฎ

Attribute	type	Description	Example Data	Constraint
ID	char(50)	รหัสกฎ	1	pk
sex	int (5)	เพศ	(10:ชาย, 11:หญิง)	Not null
status	int (5)	สถานพ	(20:โสด, 21:แต่งงานแล้ว)	Not null
occupation	int(5)	ชื่ออาชีพ	700 พนักงานบริษัท, 701อาชีพอิสระ,702 รับจ้าง	Not null
Smoking	int (5)	สูบบุหรี่	(40:ไม่สูบบุหรี่, 41:สูบบุหรี่)	Not null
alcohol	int (5)	ดื่มสุรา	(50:ไม่ดื่มสุรา, 51:ดื่มสุรา)	Not null
Exercise	int (5)	ออกกำลังกาย	(60:ไม่ออกกำลังกาย, 61:ออกกำลังกาย)	Not null
cholesterol	int (5)	ค่าคอเลสเตอรอล	(90:ปกติ, 91:ไม่ปกติ)	Not null
Blood pressure	int (5)	ค่า BP	(80:ต่ำ, 81: สูง)	Not null
Class	int (5)	ค่าบอกว่า เป็นกับไม่เป็น	(y:เป็น, n:ไม่เป็น)	Not null

Computer Science Department
Faculty of Informatics Mahasarakham University