

บทที่ 4

การพัฒนาต้นแบบการตัดสินใจ

การศึกษานี้มุ่งเน้นการพัฒนาต้นแบบเพื่อช่วยสนับสนุนการตัดสินใจการให้สินเชื่อสำหรับ บริษัท พัฒนสินลิสซิ่ง (ซีพีแอล) จำกัด โดยใช้ฐานข้อมูลเดิมที่อยู่บนระบบจัดการฐานข้อมูลมายเอสคิวแอล ซึ่งประกอบด้วยตารางข้อมูลประวัติลูกค้า ตารางข้อมูลสัญญาเช่าซื้อ ตารางข้อมูลการชำระค่างวดตามสัญญาเช่าซื้อ ตารางอาชีพของลูกค้า และตารางอาชีพย่อยของลูกค้า การดำเนินการศึกษานี้ได้ใช้วิธีฟลอกซ์โปร 6.0 ในการพัฒนาโปรแกรมและใช้ระบบจัดการฐานข้อมูลมายเอสคิวแอลในการจัดเก็บข้อมูล เพื่อให้ได้เป้าหมายตามที่มุ่งหวังไว้ จึงได้กำหนดขั้นตอนการดำเนินการศึกษาในหัวข้อต่างๆ ดังนี้

- 4.1 หลักการทำเหมืองข้อมูลและวิธีการพยากรณ์ข้อมูล
- 4.2 ปัจจัยที่นำมาสร้างต้นแบบที่ต้องการ
- 4.3 รูปแบบของฐานข้อมูลในการทำนาย
- 4.4 ต้นแบบในการทำนาย
- 4.5 การตรวจสอบผลของต้นแบบในการทำนาย
- 4.6 ประเมินความถูกต้องของต้นแบบ

4.1 หลักการทำเหมืองข้อมูลและวิธีการพยากรณ์ข้อมูล

ผู้ศึกษาได้ทำการศึกษาและได้กล่าวไว้แล้วในบทที่ 3 จากวัตถุประสงค์ของการศึกษานี้ ผู้ศึกษาได้เลือกใช้เทคนิคต้นไม้ตัดสินใจ ขั้นตอนวิธีค่าอินฟอร์เมชันเอนทรีหรือ ID3 ในการสร้างต้นแบบ เพื่อช่วยสนับสนุนการตัดสินใจการให้สินเชื่อสำหรับ บริษัท พัฒนสินลิสซิ่ง (ซีพีแอล) จำกัด เนื่องด้วยเทคนิคดังกล่าว เป็นกระบวนการสร้างต้นแบบสำหรับการทำนาย โดยอาศัยการเรียนรู้จากฐานข้อมูลเดิมที่มีอยู่ หลังจากนั้นเมื่อมีลูกค้ารายใหม่ที่มาขอสินเชื่อ เรายังนำมาผ่านต้นแบบ ซึ่งต้นแบบจะสามารถทำนายออกมาได้ว่า ลูกค้ารายใหม่นี้จะ ดีหรือไม่ดีในระดับใด

4.2 ปัจจัยที่นำมาสร้างต้นแบบที่ต้องการ

จากการสำรวจความต้องการและความคิดเห็นจากเจ้าหน้าที่ที่เกี่ยวข้องในการพิจารณาการให้สินเชื่อของบริษัทฯ แล้วนั้นสามารถสรุปได้ว่า ปัจจัยที่ส่งผลกระทบต่อพิจารณาการให้สินเชื่อกับลูกค้าประกอบด้วย อายุขณะทำสัญญา สถานะภาพ สถานะในที่อยู่อาศัย ลักษณะของที่อยู่อาศัย

กรรมสิทธิ์ในที่อยู่อาศัย กลุ่มอาชีพ รายได้สุทธิเปรียบเทียบกับค่างวดที่ต้องผ่อนชำระ ยอดเงินกู้เปรียบเทียบกับยอดที่ให้กู้ได้ และผู้ค้ำประกัน

4.3 รูปแบบของฐานข้อมูลในการทำนาย

ขั้นตอนนี้ผู้ศึกษาได้ศึกษาลักษณะการจัดเก็บข้อมูลเดิมที่อยู่บนระบบจัดการฐานข้อมูลมายเอสคิวแอล จัดเตรียมข้อมูลและเขียนโปรแกรมสร้างฐานข้อมูลใหม่ขึ้นมา เพื่อจัดเก็บข้อมูลในรูปแบบที่เหมาะสม อีกทั้งจัดแบ่งข้อมูลสำหรับการนำไปใช้การสร้างและทดสอบต้นแบบในขั้นตอนต่อไป

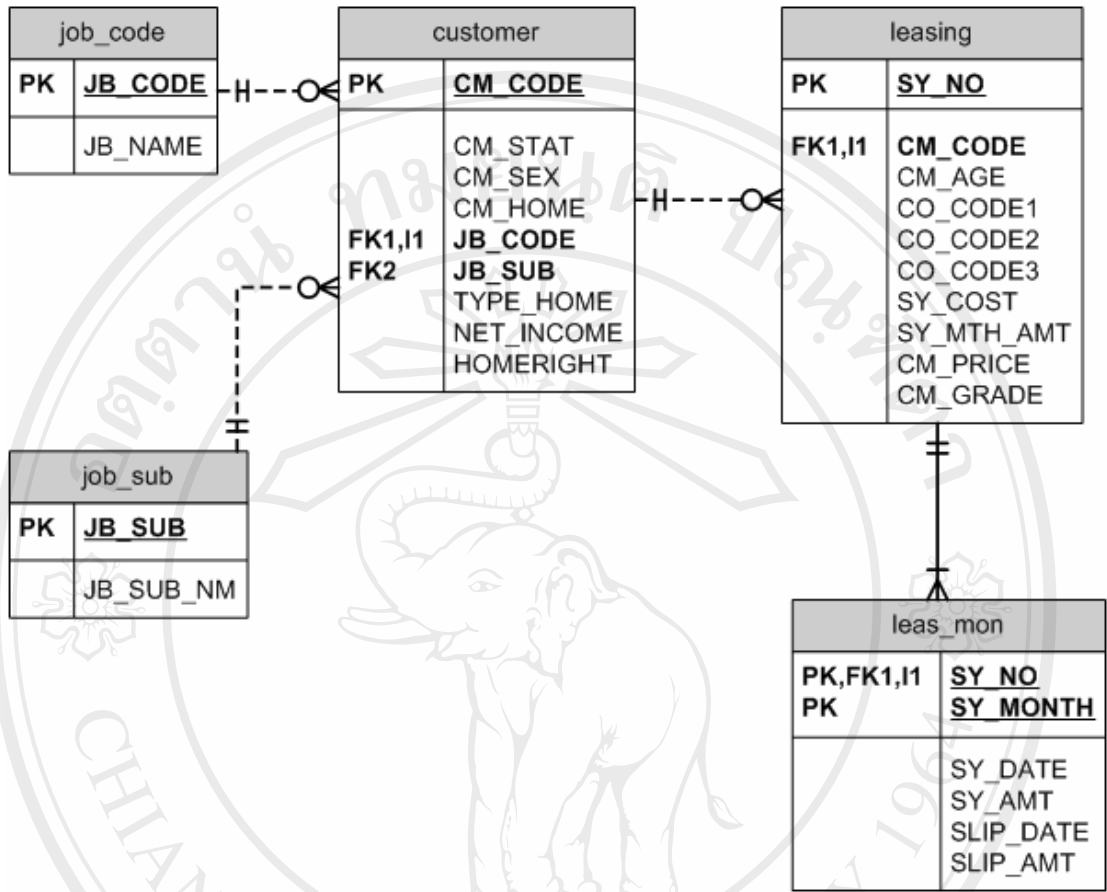
4.3.1 ศึกษาและจัดเตรียมข้อมูล

ข้อมูลที่นำมาศึกษาในครั้งนี้มาจากฐานข้อมูลเดิมที่อยู่บนระบบจัดการฐานข้อมูลมายเอสคิวแอล ซึ่งประกอบด้วยตารางข้อมูลประวัติลูกค้า ตารางข้อมูลสัญญาเช่าซื้อ ตารางข้อมูลการชำระค่างวดตามสัญญาเช่าซื้อ ตารางอาชีพของลูกค้า และตารางอาชีพย่อยของลูกค้า ดังตาราง 4.1

ตาราง 4.1 ตารางรายการตารางข้อมูลจากฐานข้อมูลเดิมที่นำมาศึกษา

ชื่อตารางข้อมูล	อธิบาย	จำนวนรายการข้อมูล
customer	ตารางข้อมูลประวัติลูกค้า	23,649
leasing	ตารางข้อมูลสัญญาเช่าซื้อ	11,708
leas_mon	ตารางข้อมูลการชำระค่างวดตามสัญญาเช่าซื้อ	427,234
job_code	ตารางข้อมูลอาชีพลูกค้า	23
job_sub	ตารางข้อมูลอาชีพย่อยลูกค้า	39

ตารางข้อมูลทั้ง 5 ตารางมีความสัมพันธ์กัน ซึ่งสามารถแสดงได้ดังรูป 4.1



รูป 4.1 รูปความสัมพันธ์ระหว่างตารางข้อมูลที่น่ามาศึกษา

รายละเอียดของโครงสร้างข้อมูลที่น่ามาศึกษาในครั้งนี้ เป็นดังตาราง 4.2 ถึง 4.6

ตาราง 4.2 โครงสร้างตารางข้อมูลอาชีพของลูกค้า

ชื่อตารางข้อมูล : job_code				
รายละเอียดตาราง : เก็บข้อมูลอาชีพของลูกค้า				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
JB_CODE	CHAR	2	รหัสอาชีพ	01
JB_NAME	VARCHAR	40	ชื่ออาชีพ	รับราชการ

ตาราง 4.3 โครงสร้างตารางข้อมูลอาชีพย่อยของลูกค้า

ชื่อตารางข้อมูล : job_sub				
รายละเอียดตาราง : เก็บข้อมูลอาชีพย่อยของลูกค้า				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
JB_SUB	CHAR	3	รหัสอาชีพย่อย	011
JB_SUB_NM	VARCHAR	40	ชื่ออาชีพย่อย	พนักงานของรัฐ

ตาราง 4.4 โครงสร้างตารางข้อมูลประวัติลูกค้า

ชื่อตารางข้อมูล : customer				
รายละเอียดตาราง : เก็บข้อมูลประวัติลูกค้า				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
CM_CODE	VARCHAR	6	รหัสลูกค้า	000001
CM_STAT	DECIMAL	4	สถานภาพ	1
CM_SEX	DECIMAL	4	เพศ	2
CM_HOME	DECIMAL	4	สถานะในที่อยู่อาศัย	1
JB_CODE	CHAR	2	รหัสอาชีพลูกค้า	01
JB_SUB	CHAR	3	รหัสอาชีพย่อยลูกค้า	011
TYPE_HOME	VARCHAR	60	ลักษณะที่อยู่อาศัย	บ้านเดี่ยว
NET_INCOME	DECIMAL	9	รายได้สุทธิ	12000
HOMERIGHT	VARCHAR	60	กรรมสิทธิ์ในที่อยู่อาศัย	เป็นของตนเอง

CM_STAT (สถานภาพ) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, ... หรือ 6 โดยที่

- | | |
|-----------------|----------------------------|
| 1 หมายถึง โสด | 4 หมายถึง หย่า |
| 2 หมายถึง สมรส | 5 หมายถึง แยกกันอยู่ |
| 3 หมายถึง หม้าย | 6 หมายถึง สมรสไม่จดทะเบียน |

CM_SEX (เพศ) ค่าของข้อมูลที่เป็นได้ คือ 1 หรือ 2 โดยที่

- | | |
|------------------|-------------------|
| 1 หมายถึง เพศชาย | 2 หมายถึง เพศหญิง |
|------------------|-------------------|

CM_HOME (สถานะในที่อยู่อาศัย) ค่าของข้อมูลที่เป็นได้ คือ 1, 2 หรือ 3 โดยที่

1 หมายถึง เจ้าบ้าน

3 หมายถึง ผู้อาศัย

2 หมายถึง เช่าอยู่

ตาราง 4.5 โครงสร้างตารางข้อมูลสัญญาเช่าซื้อ

ชื่อตารางข้อมูล : leasing				
รายละเอียดตาราง : เก็บข้อมูลสัญญาเช่าซื้อ				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
SY_NO	VARCHAR	15	เลขที่สัญญา	0050-02-00150-1
CM_CODE	VARCHAR	6	รหัสลูกค้า	000001
CM_AGE	DECIMAL	4	อายุ	30
CO_CODE1	VARCHAR	6	รหัสผู้ค้าคนที่ 1	000002
CO_CODE2	VARCHAR	6	รหัสผู้ค้าคนที่ 2	000003
CO_CODE3	VARCHAR	6	รหัสผู้ค้าคนที่ 3	000004
SY_COST	DECIMAL	9	ยอดกู้	150000
SY_MTH_AMT	DECIMAL	8	ค่างวดผ่อนชำระ/เดือน	3500
CM_PRICE	DECIMAL	9	ยอดที่ให้กู้ได้	180000
CM_GRADE *	CHAR	1	เกรด	A

ตาราง 4.6 โครงสร้างตารางข้อมูลกำหนดชำระและการชำระค่างวดตามสัญญาเช่าซื้อ

ชื่อตารางข้อมูล : leas_mon				
รายละเอียดตาราง : เก็บข้อมูลกำหนดชำระและการชำระค่างวดตามสัญญาเช่าซื้อ				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
SY_NO	VARCHAR	15	เลขที่สัญญา	0050-02-00150-1
SY_MONTH	DECIMAL	4	งวดที่	5
SY_DATE	DATE	-	วันที่ครบกำหนดชำระ	2007-05-15
SY_AMT	DECIMAL	8	ค่างวดที่ต้องชำระ	3500

ตาราง 4.6 โครงสร้างตารางข้อมูลกำหนดชำระและการชำระค่างวดตามสัญญาเช่าซื้อ(ต่อ)

ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
SLIP_DATE	DATE	-	วันที่มาชำระ	2007-05-10
SLIP_AMT	DECIMAL	5	ค่างวดที่มาชำระ	3500

จากฐานข้อมูลเดิมที่ได้ศึกษาแล้วนั้น ผู้ศึกษาได้เขียนโปรแกรมจัดเตรียมข้อมูลเพื่อให้ได้ข้อมูลที่ครบถ้วนสมบูรณ์พอเพียงกับการนำไปใช้ในขั้นตอนแปลงข้อมูลในขั้นตอนต่อไป โดยได้สร้างตารางข้อมูลใหม่ขึ้นมา ชื่อ step_1 โดยมีโครงสร้างดังตาราง 4.7

ตาราง 4.7 โครงสร้างตารางข้อมูลที่จัดเตรียมขึ้นต้น

ชื่อตารางข้อมูล : step_1				
รายละเอียดตาราง : เก็บข้อมูลขั้นต้นที่ใช้สำหรับใช้ในกระบวนการต่อไป				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
SY_NO	VARCHAR	15	เลขที่สัญญา	0050-02-00150-1
CM_CODE	VARCHAR	6	รหัสลูกค้า	000497
CM_STAT	DECIMAL	4	สถานภาพ	1
CM_SEX	DECIMAL	4	เพศ	1
CM_HOME	DECIMAL	4	สถานะในที่อยู่อาศัย	3
JB_CODE	CHAR	2	รหัสอาชีพ	21
CM_JOB	VARCHAR	30	รายละเอียดอาชีพ	บริษัทเอกชน
CM_SALARY	DECIMAL	8	เงินเดือน	8500
TYPE_HOME	VARCHAR	60	ลักษณะที่อยู่อาศัย	บ้านเดี่ยว
NET_INCOME	DECIMAL	8	รายได้สุทธิ	4800
HOMERIGHT	VARCHAR	60	กรรมสิทธิ์ในที่อยู่อาศัย	เป็นของตนเอง
JB_SUB	CHAR	3	รหัสอาชีพย่อย	023
CM_AGE	DECIMAL	4	อายุ	33
CO_CODE1	VARCHAR	6	รหัสผู้ค้าคนที่ 1	000498
CO_CODE2	VARCHAR	6	รหัสผู้ค้าคนที่ 2	000499

ตาราง 4.7 โครงสร้างตารางข้อมูลที่จัดเตรียมขั้นต้น(ต่อ)

ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
CO_CODE3	VARCHAR	6	รหัสผู้ค้าคนที่ 3	000500
SY_COST	DECIMAL	9	ยอดกู้	200000
SY_MTH_AMT	DECIMAL	9	ค่างวดผ่อนชำระ/เดือน	4904.17
CM_PRICE	DECIMAL	9	ยอดที่ให้กู้ได้	207000
CM_GRADE *	CHAR	1	เกรด	A

ตารางข้อมูลนี้ประกอบด้วยข้อมูลที่ได้ผ่านการคัดกรอง คัดเลือกและตรวจสอบข้อมูลขั้นต้นก่อนที่จะนำมาใช้ในการทำสร้างต้นแบบ เช่น ความสมบูรณ์และความถูกต้องของข้อมูลปรับเปลี่ยนรูปแบบของข้อมูล โดยการนำตารางทั้งสามในฐานข้อมูล ซึ่งได้แก่ตาราง leasing , customer และ leas_mon มาเชื่อมต่อกัน หลังจากนั้นได้ทำความสะอาดข้อมูลโดยตัดข้อมูลที่ไมสมบูรณ์ออกไป เช่น

- ลบข้อมูลที่รายการรหัสอาชีพ(JB_CODE) ที่มีค่าว่าง
- ลบข้อมูลที่รายการรหัสอาชีพย่อย(JB_SUB) ที่มีค่าว่าง
- ลบข้อมูลที่รายการลักษณะที่อยู่อาศัย(TYPE_HOME) ที่มีค่าว่าง
- ลบข้อมูลที่รายการรายได้สุทธิ(NET_INCOME) ที่มีค่าว่าง
- ลบข้อมูลที่รายการยอดที่ให้กู้ได้(CM_PRICE) ที่มีค่าว่าง
- ลบข้อมูลที่รายการเกรด(CM_GRADE) ที่มีค่าที่ไม่สามารถวัดผลได้ (เกรด I)

* ข้อมูลรายการ CM_GRADE เกิดจากการคำนวณ ดังนี้

กำหนดให้	<i>A_p</i>	หมายถึง จำนวนงวดที่ลูกค้ามาชำระเงินตามกำหนด
	<i>A</i>	หมายถึง จำนวนงวดที่ลูกค้ามาชำระเงินหลังกำหนดแต่ไม่เกิน 10 วัน
	<i>B</i>	หมายถึง จำนวนงวดที่ลูกค้ามาชำระเงินหลังกำหนดเกิน 10 วันแต่ยังอยู่ในงวดเดือนที่ครบกำหนดชำระนั้น
	<i>C</i>	หมายถึง จำนวนงวดที่ลูกค้ามาชำระเงินหลังกำหนดโดยข้ามไปยังงวดเดือนถัดไป แต่ยังไม่ถึงงวดที่ครบกำหนดชำระครั้งที่ 2
	<i>D</i>	หมายถึง จำนวนงวดที่ลูกค้ามาชำระเงินหลังกำหนดโดยข้ามไปยังงวดเดือนถัดไป 2 งวด แต่ยังไม่ถึงงวดที่ครบกำหนดชำระครั้งที่ 3
	<i>X</i>	หมายถึง คะแนนที่คำนวณได้จากสูตร ดังนี้

$$A_p + 0.9A + 0.8B + 0.6C + 0.3D$$

Y หมายถึง จำนวนงวดที่ถึงกำหนดชำระ ณ วันที่ใดที่ระบุ

Z หมายถึง อัตราส่วนของคะแนน X ต่อ Y คำนวณได้จากสูตร ดังนี้

$$\frac{100X}{Y}$$

กรณีที่ Y มีค่าน้อยกว่า 6 จะไม่คำนวณหาค่า Z และจะให้เกรดเป็น I

เมื่อได้ค่า Z แล้ว จะนำไปตัดเกรดดังตาราง 4.8

ตาราง 4.8 ตารางการให้เกรดลูกค้าตามช่วงคะแนน

ช่วงคะแนนของ Z	เกรด(CM_GRADE)
90 – 100	X (ดีเยี่ยม)
80 – 89	A (ดีมาก)
70 – 79	B (ดี)
60 – 69	C (พอใช้)
50 – 59	D (แย)
0 – 49	F (แย่มาก)

หลังจากผ่านโปรแกรมการจัดเตรียมข้อมูล สามารถสรุปได้ว่า ข้อมูลเดิม ณ วันที่ 30/04/2551 มีทั้งสิ้น 10,352 รายการที่ถูกนำมาประมวลผล และเมื่อประมวลผลจัดเตรียมข้อมูลแล้วเสร็จจะได้ข้อมูลที่ผ่านการคัดกรอง คัดเลือกและตรวจสอบข้อมูลขึ้นต้นจำนวน 1,823 รายการ ซึ่งจัดเก็บอยู่ในตารางข้อมูล step_1 ตัวอย่างข้อมูลในตารางข้อมูล step_1 เป็นดังตาราง 4.9

ตาราง 4.9 ตัวอย่างข้อมูลในตารางข้อมูล step_1

SY_NO	CM_CODE	CM_STAT	CM_SEX	CM_HOME	JB_CODE	CM_JOB	CM_SALARY	TYPE_HOME	NET_INCOME	HOMERIGHT	JB_SUB	CM_AGE	CO_CODE1	CO_CODE2	CO_CODE3	SY_COST	SY_MTH_AMT	CM_PRICE	CM_GRADE
0046-02-00972-2	000467	1	1	3/21	บริษัทเอกชน	6500	บ้านเดี่ยว	4800	สิทธิ์เช่า/เช่าเหมา	023	33	90000				90000	4654.5	207000	A
0046-02-00987-2	001511	2	2	3/05	ค้าขาย	30000	ทาวน์โฮต	15000	สิทธิ์เช่า	033	41	001680				295000	10689.35	324000	B
0046-02-00988-2	002227	2	2	3/21	บริษัทเอกชน	9000	บ้านเดี่ยว	5772	สิทธิ์เช่า	023	28	002228				135000	4082.75	135000	F
0046-02-00989-2	002229	2	1	3/21	บริษัทเอกชน	7174	บ้านเดี่ยว	5870	สิทธิ์เช่า	023	33	002230				100000	3031.67	135000	A
0046-02-00991-2	002234	6	2	3/05	ค้าขาย	20000	บ้านเดี่ยว	4000	สิทธิ์เช่า/เช่าเหมา	036	25					150000	3878.75	243000	F
0046-02-00993-2	002235	1	2	3/21	บริษัทเอกชน	6510	บ้านเดี่ยว	4604	สิทธิ์เช่า	023	42	002236	002237			150000	4547.5	153000	F
0046-02-00994-2	002238	2	2	3/05	ค้าขาย	7000	บ้านเดี่ยว	3600	สิทธิ์เช่า/เช่าเหมา	034	35	002239				70000	2954.39	67500	B
0046-02-00996-2	002244	1	1	3/22	สุภาพ ทั้งรับพนักงาน	7500	บ้านเดี่ยว	3800	สิทธิ์เช่า	045	24	002245				90000	3416.36	98000	A
0046-02-00998-2	000763	1	1	3/21	บริษัทเอกชน	5000	บ้านเดี่ยว	6000	สิทธิ์เช่า	022	31	000186				40000	2282.67	56000	A
0046-02-00999-2	002246	6	1	1/21	บริษัทเอกชน	15000	บ้านเดี่ยว	4500	สิทธิ์เช่า/เช่าเหมา	023	29	002247				120000	3638	153000	A
0046-02-01002-2	002232	1	1	2/21	บริษัทเอกชน	6000	ทาวน์โฮต	3800	สิทธิ์เช่า	023	25	002233				130000	3361.58	198000	C
0046-02-01001-2	002248	2	1	3/04	รับจ้าง	25000	บ้านเดี่ยว	7500	สิทธิ์เช่า	055	44	002249				150000	4547.5	207000	F
0046-02-01002-2	002250	1	1	3/22	สุภาพ ทั้งรับพนักงาน	8000	ทาวน์โฮต	4000	สิทธิ์เช่า/เช่าเหมา	042	22	002251				90000	3416.36	108000	B
0046-02-01004-2	002252	1	1	3/21	บริษัทเอกชน	6850	บ้านเดี่ยว	3340	สิทธิ์เช่า	022	24	002253				140000	3495.33	202500	X
0046-02-01005-2	002254	1	1	3/01	รับราชการ	8531	บ้านเดี่ยว	4212	สิทธิ์เช่า	013	28	002255				200000	5085	198000	X
0046-02-01006-2	002260	1	1	2/01	รับราชการ	9500	คอนโดมิเนียม	7000	สิทธิ์เช่า	012	31	002261				160000	5802.86	171000	A
0046-02-01007-2	002262	6	2	3/05	ค้าขาย	15000	บ้านเดี่ยว	4500	สิทธิ์เช่า	034	42	002263				120000	3638	128000	F
0046-02-01008-2	002258	2	2	3/21	บริษัทเอกชน	8000	บ้านเดี่ยว	7350	สิทธิ์เช่า	023	31	002259				250000	6241.67	297000	F
0046-02-01009-2	000766	2	1	1/05	ค้าขาย	25000	บ้านเดี่ยว	5000	สิทธิ์เช่า/เช่าเหมา	033	46	000767	002264			100000	3493.29	135000	A
0050-02-01010-2	002265	2	1	3/01	รับราชการ	7500	บ้านเดี่ยว	5550	สิทธิ์เช่า	013	32	002266	002267			135000	4082.75	144000	F
0050-02-01011-2	002268	4	2	3/05	ค้าขาย	12000	ทาวน์โฮต	4650	สิทธิ์เช่า	032	43	002269	002270			120000	3638	121500	F
0050-02-01012-2	000974	1	2	3/06	ธุรกิจส่วนตัว	20000	บ้านเดี่ยว	8000	สิทธิ์เช่า	033	47	000975				100000	3493.29	103500	A
0050-02-01013-2	002271	2	1	1/07	เกษตรกร	20000	บ้านเดี่ยว	8000	สิทธิ์เช่า/เช่าเหมา	051	29	002272				150000	7757.5	288000	X

4.3.2 แปลงข้อมูล

ขั้นตอนนี้ผู้ศึกษาได้เขียนโปรแกรมสร้างตารางข้อมูลใหม่ขึ้นมา เพื่อเก็บข้อมูลที่เหมาะสมสำหรับการนำไปใช้สร้างต้นแบบในขั้นตอนต่อไป โดยแหล่งข้อมูลที่น่ามาสร้างตารางข้อมูลใหม่นี้มาจากตาราง step_1 ที่ถูกสร้างมาจากขั้นตอน 4.3.1 ตารางข้อมูลใหม่ที่สร้างในขั้นตอนนี้ ชื่อ step_2 โดยมีโครงสร้างดังตาราง 4.10

ตาราง 4.10 โครงสร้างตารางข้อมูลสำหรับนำไปสร้างต้นแบบ

ชื่อตาราง : step_2				
รายละเอียดตาราง : เก็บข้อมูลสำหรับนำไปสร้างต้นแบบ				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
AGE	CHAR	1	ช่วงอายุลูกค้า	3
SEX	CHAR	1	เพศ	2
STATUS	CHAR	1	สถานภาพ	1
HOMESTAT	CHAR	1	สถานะในที่อยู่อาศัย	1
HOMETYPE	CHAR	1	ลักษณะที่อยู่อาศัย	2
HOMERIGHT	CHAR	1	กรรมสิทธิ์ในที่อยู่อาศัย	3
JOB	CHAR	1	กลุ่มอาชีพ	4
INC_RATE	CHAR	1	ช่วงอัตรารายได้สุทธิเทียบกับค่าแรงต่อเดือน	5
LOAN_RATE	CHAR	1	ช่วงอัตราดอกเบี้ยเทียบกับยอดที่ให้กู้ได้	1
GUARANTEE	CHAR	1	การมีผู้ค้ำประกัน	1
GRADE	CHAR	1	ระดับความน่าเชื่อถือ	X

AGE (ช่วงอายุลูกค้า) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, ... หรือ 6 โดยที่

1 หมายถึง ช่วงอายุ 1-19 ปี

4 หมายถึง ช่วงอายุ 40-49 ปี

2 หมายถึง ช่วงอายุ 20-29 ปี

5 หมายถึง ช่วงอายุ 50-59 ปี

3 หมายถึง ช่วงอายุ 30-39 ปี

6 หมายถึง ช่วงอายุตั้งแต่ 60 ปี ขึ้นไป

SEX (เพศ) ค่าของข้อมูลที่เป็นได้ คือ 1 หรือ 2 โดยที่

- | | |
|------------------|-------------------|
| 1 หมายถึง เพศชาย | 2 หมายถึง เพศหญิง |
|------------------|-------------------|

STATUS (สถานภาพ) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, ... หรือ 6 โดยที่

- | | |
|-----------------|----------------------------|
| 1 หมายถึง โสด | 4 หมายถึง หย่า |
| 2 หมายถึง สมรส | 5 หมายถึง แยกกันอยู่ |
| 3 หมายถึง หม้าย | 6 หมายถึง สมรสไม่จดทะเบียน |

HOMESTAT (สถานะในที่อยู่อาศัย) ค่าของข้อมูลที่เป็นได้ คือ 1, 2 หรือ 3 โดยที่

- | | |
|--------------------|--------------------|
| 1 หมายถึง เจ้าบ้าน | 3 หมายถึง ผู้อาศัย |
| 2 หมายถึง เช่าอยู่ | |

HOMETYPE (ลักษณะที่อยู่อาศัย) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, หรือ 9 โดยที่

- | | |
|------------------------|-------------------------|
| 1 หมายถึง บ้านเดี่ยว | 6 หมายถึง ตึกแถว |
| 2 หมายถึง ห้องเช่า | 7 หมายถึง อพาร์ทเมนต์ |
| 3 หมายถึง ทาวน์เฮาส์ | 8 หมายถึง บ้านแฝด |
| 4 หมายถึง อาคารพาณิชย์ | 9 หมายถึง หอพักหรือแฟลต |
| 5 หมายถึง คอนโดมิเนียม | |

HOMERIGHT (กรรมสิทธิ์ในที่อยู่อาศัย) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, หรือ 6 โดยที่

- 1 หมายถึง เป็นของตนเอง ไม่มีภาระผูกพัน
- 2 หมายถึง เป็นของตนเอง แต่มีภาระผูกพัน
- 3 หมายถึง เป็นของตนเอง แต่ที่ดินเป็นของคนอื่น(ญาติ)
- 4 หมายถึง อาศัยญาติ
- 5 หมายถึง บ้านพัก
- 6 หมายถึง เช่า

JOB (กลุ่มอาชีพ) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, หรือ 5 โดยที่

- 1 หมายถึง มั่นคงมาก เช่น ข้าราชการ รัฐวิสาหกิจ
- 2 หมายถึง มั่นคง เช่น สถาบันการเงิน บริษัทมหาชน ธุรกิจส่วนตัวขนาดใหญ่
- 3 หมายถึง ค่อนข้างมั่นคง เช่น ธุรกิจส่วนตัวขนาดกลาง ค้าขายมีหน้าร้าน ลูกจ้างประจำ พนักงานบริษัทเอกชน เกษตรกรที่มีที่ทำกิน 10 ไร่ ขึ้นไป
- 4 หมายถึง พิจารณาพิเศษ เช่น ธุรกิจส่วนตัวขนาดเล็ก เกษตรกรที่มีที่ทำกินน้อยกว่า 10 ไร่ พนักงานขาย ผู้รับเหมา ลูกจ้างชั่วคราว

5 หมายถึง ไม่รับพิจารณา เช่น ทหาร ตำรวจ ทนายความ นักการเมือง ตัวแทนขายประกัน อาชีพที่ไม่มีเงินเดือนประจำ

INC_RATE(ช่วงอัตรารายได้สุทธิเทียบกับค่างวดผ่อนชำระต่อเดือน) ค่าของข้อมูลที่เป็นได้ คือ

1, 2, 3, หรือ 5 โดยที่

1 หมายถึง $Z > 30\%$

4 หมายถึง $0\% \leq Z \leq 10\%$

2 หมายถึง $20\% < Z \leq 30\%$

5 หมายถึง $Z < 0\%$

3 หมายถึง $10\% < Z \leq 20\%$

$$\text{ทั้งนี้ } Z = \frac{(X - Y) \times 100}{Y}$$

X หมายถึง รายได้สุทธิ

Y หมายถึง ค่างวดผ่อนชำระต่อเดือน

LOAN_RATE(ช่วงอัตรายอดกู้เทียบกับยอดที่ให้กู้ได้) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, หรือ 6 โดยที่

1 หมายถึง $0 < Z \leq 60$

4 หมายถึง $80 < Z \leq 90$

2 หมายถึง $60 < Z \leq 70$

5 หมายถึง $90 < Z \leq 100$

3 หมายถึง $70 < Z \leq 80$

6 หมายถึง $Z > 100$

$$\text{ทั้งนี้ } Z = \frac{X \times 100}{Y}$$

X หมายถึง ยอดเงินกู้

Y หมายถึง ยอดที่ให้กู้ได้

GUARANTEE (การมีผู้ค้ำประกัน) ค่าของข้อมูลที่เป็นได้ คือ 1 หรือ 0 โดยที่

1 หมายถึง มีผู้ค้ำประกัน

0 หมายถึง ไม่มีผู้ค้ำประกัน

GRADE (ระดับความน่าเชื่อถือ) ค่าของข้อมูลที่เป็นได้ คือ X, Y หรือ Z โดยที่

X หมายถึง ดี ได้แก่ลูกค้าที่ได้เกรด X, A, หรือ B

Y หมายถึง ปานกลาง ได้แก่ลูกค้าที่ได้เกรด C, หรือ D

Z หมายถึง แย่ ได้แก่ลูกค้าที่ได้เกรด F

จากโครงสร้างของตารางข้อมูลทีพร้อมจะนำไปสร้างต้นแบบ step_2 นั้น ตารางข้อมูล step_1 ที่มีอยู่จะถูกแปลงและนำมาจัดเก็บไว้ในตาราง step_2 ดังนี้

รายการ CM_AGE จะต้องถูกแปลงค่าโดยเปรียบเทียบอายุ กับช่วงอายุที่ได้กำหนดไว้ แล้วแปลงค่าเพื่อนำไปจัดเก็บที่รายการ AGE ตัวอย่างเช่น CM_AGE มีค่า 45 ซึ่งจะอยู่ในช่วงอายุ 40-49 ปี นั่นคือ AGE จะมีค่าเป็น “4” เป็นต้น

รายการ CM_SEX จะต้องถูกแปลงค่าเพื่อนำไปจัดเก็บที่รายการ SEX ตัวอย่างเช่น CM_SEX มีค่า 1 นั่นคือ SEX จะมีค่าเป็น “1” เป็นต้น

รายการ CM_STAT จะต้องถูกแปลงค่าเพื่อนำไปจัดเก็บที่รายการ STATUS ตัวอย่างเช่น CM_STAT มีค่า 3 นั่นคือ STATUS จะมีค่าเป็น “3” เป็นต้น

รายการ CM_HOME จะต้องถูกแปลงค่าเพื่อนำไปจัดเก็บที่รายการ HOMESTAT ตัวอย่างเช่น CM_HOME มีค่า 1 นั่นคือ HOMESTAT จะมีค่าเป็น “1” เป็นต้น

รายการ TYPE_HOME จะต้องถูกแปลงค่าเพื่อนำไปจัดเก็บที่รายการ HOMETYPE ตัวอย่างเช่น TYPE_HOME มีค่า “บ้านเดี่ยว” นั่นคือ HOMETYPE จะมีค่าเป็น “1” เป็นต้น

รายการ HOMERIGHT จะต้องถูกแปลงค่าเพื่อนำไปจัดเก็บที่รายการ HOMERIGHT ตัวอย่างเช่น HOMERIGHT มีค่า “เป็นของตนเอง ไม่มีภาระผูกพัน” นั่นคือ HOMERIGHT จะมีค่าเป็น “1” เป็นต้น

รายการ NET_INCOME และ SY_MTH_AMT จะคำนวณเพื่อหาค่าและนำไปจัดเก็บที่รายการ INC_RATE ตามสูตรคำนวณที่ได้กล่าวไว้แล้วข้างต้น ตัวอย่างเช่น NET_INCOME มีค่า 4800 และ SY_MTH_AMT มีค่า 4904.17 สูตรคำนวณคือ $Z = \frac{(X - Y) \times 100}{Y}$

โดยที่ X หมายถึง รายได้สุทธิ ซึ่งก็คือรายการ NET_INCOME

Y หมายถึง ค่างวดผ่อนชำระต่อเดือน ซึ่งก็คือรายการ SY_MTH_AMT

เมื่อแทนค่าในสูตร จะสามารถคำนวณค่า Z ออกมาได้เท่ากับ -2.1241 และจากข้อกำหนดที่ว่า

1 หมายถึง $Z > 30\%$

4 หมายถึง $0\% \leq Z \leq 10\%$

2 หมายถึง $20\% < Z \leq 30\%$

5 หมายถึง $Z < 0\%$

3 หมายถึง $10\% < Z \leq 20\%$

นั่นคือ INC_RATE จะมีค่าเป็น “5” เป็นต้น

รายการ SY_COST และ CM_PRICE จะคำนวณเพื่อหาค่าและนำไปจัดเก็บที่รายการ LOAN_RATE ตามสูตรคำนวณที่ได้กล่าวไว้แล้วข้างต้น ตัวอย่างเช่น SY_COST มีค่า 200000

และ CM_PRICE มีค่า 207000 สูตรคำนวณคือ $Z = \frac{X \times 100}{Y}$

โดยที่ X หมายถึง ยอดเงินกู้ ซึ่งก็คือรายการ SY_COST

Y หมายถึง ยอดที่ให้ผู้กู้ได้ ซึ่งก็คือรายการ CM_PRICE

เมื่อแทนค่าในสูตร จะสามารถคำนวณค่า Z ออกมาได้เท่ากับ 96.6183 และจากข้อกำหนดที่ว่า

1 หมายถึง $0 < Z \leq 60$

4 หมายถึง $80 < Z \leq 90$

2 หมายถึง $60 < Z \leq 70$

5 หมายถึง $90 < Z \leq 100$

3 หมายถึง $70 < Z \leq 80$

6 หมายถึง $Z > 100$

นั่นคือ LOAN_RATE จะมีค่าเป็น “5” เป็นต้น

รายการ CO_CODE1, CO_CODE2 และ CO_CODE3 จะต้องถูกแปลงค่าเพื่อนำไปจัดเก็บที่รายการ GUARANTEE โดยมีเงื่อนไขว่าทั้งสามรายการคือ CO_CODE1, CO_CODE2 และ CO_CODE3 ถ้ามีค่าว่างทั้งหมดพร้อมกัน แสดงว่าไม่มีผู้ค้ำประกัน GUARANTEE จะมีค่าเป็น “1” หรือถ้ารายการใดรายการหนึ่งในสามรายการนั้นมีค่าที่ไม่ใช่ค่าว่าง แสดงว่ามีผู้ค้ำประกัน GUARANTEE จะมีค่าเป็น “0” เป็นต้น

รายการ CM_GRADE จะต้องถูกแปลงค่าโดยเปรียบเทียบเกรดกับระดับความน่าเชื่อถือที่ได้กำหนดไว้ แล้วแปลงค่าเพื่อนำไปจัดเก็บที่รายการ GRADE ตัวอย่างเช่น CM_GRADE มีค่า “A” นั่นคือ GRADE จะมีค่าเป็น “X” เป็นต้น

รายการ JB_CODE และ JB_SUB จะต้องถูกแปลงค่าเพื่อนำไปจัดเก็บที่รายการ JOB โดยนำค่าของ JB_CODE และ JB_SUB ไปอ้างอิงในตาราง job_map เพื่อให้ได้ค่าของรายการ JB_LEVEL แล้วไปจัดเก็บที่รายการ JOB สำหรับตาราง job_map ผู้ศึกษาได้ออกแบบโครงสร้างเพื่อจัดกลุ่มอาชีพ ให้เป็นหมวดหมู่เดียวกันเพื่อนำข้อมูล JB_CODE และ JB_SUB มาแปลงเป็นกลุ่มอาชีพ เพื่อนำไปสร้างต้นแบบในขั้นตอนต่อไป ผู้ศึกษาได้ออกแบบโครงสร้างตารางข้อมูล job_map ไว้ดังตาราง 4.11

ตาราง 4.11 โครงสร้างตารางข้อมูลอาชีพ, อาชีพย่อย และ กลุ่มอาชีพของลูกค้า

ชื่อตารางข้อมูล : job_map				
รายละเอียดตาราง : เก็บข้อมูลอาชีพ, อาชีพย่อย และ กลุ่มอาชีพของลูกค้า				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
JB_CODE	CHAR	2	รหัสอาชีพ	01
JB_NAME	VARCHAR	40	ชื่ออาชีพ	รับราชการ
JB_SUB	CHAR	3	รหัสอาชีพย่อย	011
JB_SUB_NM	VARCHAR	40	ชื่ออาชีพย่อย	พนักงานของรัฐ
JB_LEVEL	CHAR	1	รหัสกลุ่มอาชีพ	1

JB_LEVEL (รหัสกลุ่มอาชีพ) ค่าของข้อมูลที่เป็นได้ คือ 1, 2, 3, ... หรือ 5 โดยที่

- 1 หมายถึง มั่นคงมาก เช่น ข้าราชการ รัฐวิสาหกิจ
- 2 หมายถึง มั่นคง เช่น สถาบันการเงิน บริษัทมหาชน ธุรกิจส่วนตัวขนาดใหญ่
- 3 หมายถึง ค่อนข้างมั่นคง เช่น ธุรกิจส่วนตัวขนาดกลาง ค้าขายมีหน้าร้าน ลูกจ้างประจำ พนักงานบริษัทเอกชน เกษตรกรที่มีที่ทำกิน 10 ไร่ ขึ้นไป
- 4 หมายถึง พิจารณาพิเศษ เช่น ธุรกิจส่วนตัวขนาดเล็ก เกษตรกรที่มีที่ทำกินน้อยกว่า 10 ไร่ พนักงานขาย ผู้รับเหมา ลูกจ้างชั่วคราว
- 5 หมายถึง ไม่รับพิจารณา เช่น ทหาร ตำรวจ นายควม นักการเมือง ตัวแทนขายประกัน อาชีพที่ไม่มีเงินเดือนประจำ

ตัวอย่างข้อมูลในตารางข้อมูล job_map เป็นดังตาราง 4.12

ตาราง 4.12 ตารางตัวอย่างข้อมูลอาชีพ, อาชีพย่อย และ กลุ่มอาชีพของลูกค้า

JB_CODE ▾	JB_NAME	JB_SUB ▾	JB_SUB_NM	JB_LEVEL
01	รับราชการ	011	พนักงานของรัฐ	1
01	รับราชการ	012	ลูกจ้างประจำ	3
01	รับราชการ	013	ลูกจ้างชั่วคราว	4
01	รับราชการ	014	ทหาร/ตำรวจ	5
01	รับราชการ	015	ข้าราชการบำนาญ	1
02	รัฐวิสาหกิจ	011	พนักงานของรัฐ	1
02	รัฐวิสาหกิจ	012	ลูกจ้างประจำ	3
02	รัฐวิสาหกิจ	013	ลูกจ้างชั่วคราว	4
02	รัฐวิสาหกิจ	014	ทหาร/ตำรวจ	5
02	รัฐวิสาหกิจ	015	ข้าราชการบำนาญ	1
04	รับจ้าง	051	ที่ดินเกิน 10 ไร่	3
04	รับจ้าง	052	ที่ดินไม่เกิน 10 ไร่	4
04	รับจ้าง	053	ขายแรงงาน	4
04	รับจ้าง	054	ช่างฝีมือ	4
04	รับจ้าง	055	รับเหมางาน	4
04	รับจ้าง	056	ขับรถรับจ้าง	4
04	รับจ้าง	057	ควบคุมเครื่องจักร	4
05	ค้าขาย	031	ขนาดใหญ่	2
05	ค้าขาย	032	ขนาดกลาง	3
05	ค้าขาย	033	ขนาดเล็ก	4
05	ค้าขาย	034	ร้านค้าขนาดเล็ก	4

ตัวอย่างเช่น JB_CODE มีค่า “01” และ JB_SUB มีค่า “012” เมื่อนำไปอ้างอิงในตาราง job_map จะได้ JB_LEVEL มีค่า “3” นั่นคือ JOB จะมีค่าเป็น “3” เป็นต้น

หลังจากผ่านโปรแกรมการแปลงข้อมูล จะได้ข้อมูลที่พร้อมสำหรับนำไปสร้างต้นแบบจำนวน 1,816 รายการ ซึ่งจัดเก็บอยู่ในตารางข้อมูล step_2 ตัวอย่างข้อมูลในตาราง step_2 แสดงได้ดังตาราง 4.13

ตาราง 4.13 ตารางตัวอย่างข้อมูลหลังผ่านการแปลงข้อมูล

AGE	SEX	STATUS	HOMESTAT	HOMETYPE	HOMERIGHT	JOB	INC_RATE	LOAN_RATE	GUARANTEE	GRADE
1	1	1	3	1	4	3	4	1	0	X
4	2	2	3	3	4	4	1	5	1	X
2	2	2	3	1	4	3	1	5	1	Z
3	1	2	3	1	4	3	1	3	1	X
2	2	6	3	1	1	4	4	2	0	Z
4	2	1	3	1	4	3	4	5	1	Z
3	2	2	3	1	1	4	2	6	1	X
2	1	1	3	1	4	4	3	5	1	X
3	1	1	3	1	4	2	1	3	1	X
2	1	6	1	1	1	3	2	3	1	X
2	1	1	2	2	6	3	3	2	1	X
4	1	2	3	1	4	4	1	3	1	Z
2	1	1	3	2	5	4	3	4	1	Y
2	1	1	3	1	4	2	3	2	1	X
2	1	1	3	1	4	4	5	6	1	X
3	1	1	2	5	6	3	2	5	1	X
4	2	6	3	1	4	4	2	5	1	Z
3	2	2	3	1	4	3	3	4	1	Z
4	1	2	1	1	1	4	1	3	1	X
3	1	2	1	1	1	2	1	1	0	X

4.3.3 แบ่งข้อมูล

ขั้นตอนนี้ผู้ศึกษาได้เขียนโปรแกรมสร้างตารางข้อมูลใหม่ขึ้นมาอีก 2 ตารางคือ ตาราง step_31 และ ตาราง step_32 ซึ่งทั้งสองตารางนั้นมีโครงสร้างเหมือนกับตาราง step_2 เพียงแต่แยกข้อมูลของตาราง step_2 ออกมาเป็น 2 ชุด โดยตารางชุดแรกคือตาราง step_31 ใช้เก็บข้อมูลเพื่อการเรียนรู้ (Training Set) ซึ่งเป็นข้อมูลสำหรับการสร้างต้นแบบ และตารางชุดที่สองคือตาราง step_32 ใช้เก็บข้อมูลเพื่อการทดสอบ (Testing Data) ซึ่งเป็นข้อมูลสำหรับทดสอบต้นแบบที่ได้สร้างขึ้นมา

ผู้ศึกษาได้เลือกใช้วิธีแบ่งชุดข้อมูลโดยเทคนิคการสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) ซึ่งสุพจน์ บุญแรง (2551) กล่าวว่า การสุ่มวิธีนี้เมื่อทราบจำนวนประชากร (N) และจำนวนตัวอย่าง (n) แล้วจะทำให้สามารถคำนวณช่วงการสุ่ม (sampling interval) คือ $N/n = k$ และสัดส่วนของการสุ่ม (sampling fraction) = $n/N = 1/k$ คือทุกๆ k หน่วยจะเลือกมา 1 หน่วย

ตัวอย่าง จำนวนประชากร (N) = 50 ต้องการตัวอย่างจำนวน (n) = 10 ช่วงของการสุ่ม = $50/10 = 5$ หมายถึงต้องสุ่มในทุกๆ 5 หน่วย หรือ สุ่มทุก 1 ใน 5 หน่วย ซึ่งเรียกว่า sampling

fraction = $10/50 = 1/5$ ดังนั้นในช่วง 5 หน่วยแรกก็จะสุ่มอย่างธรรมดาค่าที่จะเริ่มต้นที่ใด เมื่อเลือกได้หน่วยที่ i เรียกว่า random start หน่วยถัดไปจะเลือกคือ $i+k, i+2k, +\dots$

จากตัวอย่างถ้าสุ่มได้หน่วยที่ 2 เป็น random start ก็จะได้หน่วยศึกษาดังนี้ 2, 2+5, 2+10,....., 2+45 นั่นคือหน่วยศึกษาที่ 2, 7, 12,....., 47

จักรกฤษณ์ สาราณใจ (2551) กล่าวว่า ในกรณีที่ N/n ไม่เป็นเลขจำนวนเต็มพอดี ให้เลือกเลขจำนวนเต็ม k ที่มีค่าใกล้เคียง N/n ให้มากที่สุด แล้วจัดเรียงลำดับหน่วยสมาชิกของประชากรเสมือนหนึ่งว่าสมาชิกเหล่านั้นเรียงกันเป็นวงกลม นั่นคือหมายเลข 1 จะไปต่อท้ายหมายเลข N จากนั้นให้สุ่มตัวเลขระหว่างเลข 1 ถึงเลข N ขึ้นมาหนึ่งตัว สมมติได้เลข p หน่วยตัวอย่างก็จะประกอบด้วยสมาชิกของประชากรลำดับที่ $p, p+k, p+2k, p+3k, \dots$ ไปเรื่อยๆ โดยไม่ต้องสนใจว่าจะข้ามลำดับที่สุดท้ายไปอย่างไร จนได้จำนวนตัวอย่างครบ n หน่วยตามต้องการ เช่น $N=70$, $n=12$ จะได้ $N/n = 70/12 = 5.83$ เลขจำนวนเต็มที่ใกล้เคียง 5.83 มากที่สุดคือ 6 จึงให้ $k=6$ แล้วสุ่มตัวเลขตั้งแต่ 1 ถึง 70 ขึ้นมาหนึ่งตัว สมมติได้เลข 42 จะได้หน่วยตัวอย่างคือสมาชิกประชากรลำดับที่ 42, 48, 54, 60, 66, 2, 8, 14, 20, 26, 32, 38 เป็นต้น

หลังจากผ่านโปรแกรมการแบ่งข้อมูล ซึ่งผู้ศึกษาได้ทดลองกำหนดอัตราส่วนของข้อมูลเพื่อการเรียนรู้ไว้ที่ร้อยละ 70 และข้อมูลเพื่อการทดสอบไว้ที่ร้อยละ 30 เมื่อประมวลผลแบ่งข้อมูลแล้วเสร็จจะได้ข้อมูลเพื่อการเรียนรู้ จำนวน 1,271 รายการ ซึ่งจัดเก็บอยู่ในตารางข้อมูล step_31 และได้ข้อมูลเพื่อการทดสอบ จำนวน 545 รายการ ซึ่งจัดเก็บอยู่ในตารางข้อมูล step_32

4.4 ต้นแบบในการทำนาย

ขั้นตอนนี้ผู้ศึกษาได้เขียนโปรแกรมเพื่อสร้างต้นแบบโดยใช้เทคนิคต้นไม้ตัดสินใจ ขั้นตอนวิธีค่าอินฟอร์เมชันแกน จากตารางข้อมูลที่เตรียมไว้เพื่อการเรียนรู้ของต้นแบบ นั่นคือใช้ข้อมูลจากตาราง step_31 ที่สร้างขึ้นในขั้นตอน 4.3.3 มาทำการสร้างต้นแบบ

จากการสำรวจความต้องการและความคิดเห็นจากเจ้าหน้าที่เกี่ยวข้องในการพิจารณาการให้สินเชื่อบริษัทฯ แล้วนั้นสามารถสรุปได้ว่า ปัจจัยที่ส่งผลต่อการพิจารณาการให้สินเชื่อกับลูกค้าประกอบด้วย อายุขณะทำสัญญา สถานะภาพ สถานะในที่อยู่อาศัย ลักษณะของที่อยู่อาศัย กรรมสิทธิ์ในที่อยู่อาศัย กลุ่มอาชีพ รายได้สุทธิเปรียบเทียบกับค่างวดที่ต้องผ่อนชำระ ยอดเงินกู้เปรียบเทียบกับยอดที่ให้กู้ได้ และผู้ค้ำประกัน เพื่อให้เห็นภาพชัดเจนว่าแต่ละปัจจัยที่กำหนดดังกล่าวมีค่าอินฟอร์เมชันแกนเป็นอย่างไร ผู้ศึกษาจึงเขียนโปรแกรมและเก็บรายละเอียดผลการคำนวณออกมา โดยกำหนดให้ฟิลด์ grade เป็นตัวแบ่งชุดข้อมูล ค่าของฟิลด์ grade ที่ใช้ในการแบ่งชุดข้อมูล คือค่า X, Y และ Z ได้ผลลัพธ์ออกมาดังตาราง 4.14 ถึง 4.23

ตาราง 4.14 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันแกนของปัจจัย อายุ

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
AGE	1	4	4	0	0
AGE	2	431	328	37	66
AGE	3	433	310	41	82
AGE	4	296	210	31	55
AGE	5	103	82	8	13
AGE	6	4	2	0	2

ตาราง 4.15 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันแกนของปัจจัย เพศ

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
SEX	1	776	562	73	141
SEX	2	495	374	44	77

ตาราง 4.16 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันแกนของปัจจัย สถานภาพ

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
STATUS	1	332	249	25	58
STATUS	2	577	424	59	94
STATUS	3	17	15	0	2
STATUS	4	70	43	11	16
STATUS	5	10	7	1	2
STATUS	6	265	198	21	46

ตาราง 4.17 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันแกนของปัจจัย สถานะในที่อยู่อาศัย

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
HOMESTAT	1	409	294	41	74
HOMESTAT	2	19	14	2	3
HOMESTAT	3	843	628	74	141

ตาราง 4.18 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันแกนของปัจจัย ลักษณะที่อยู่อาศัย

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
HOMETYPE	1	1123	822	103	198
HOMETYPE	2	25	22	2	1
HOMETYPE	3	24	18	1	5
HOMETYPE	4	33	24	2	7
HOMETYPE	5	8	5	1	2
HOMETYPE	6	11	10	0	1
HOMETYPE	7	2	2	0	0
HOMETYPE	8	14	9	2	3
HOMETYPE	9	31	24	6	1

ตาราง 4.19 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันของปัจจัย กรรมสิทธิ์ในที่อยู่อาศัย

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
HOMERIGHT	1	557	405	57	95
HOMERIGHT	4	601	447	46	108
HOMERIGHT	5	31	21	7	3
HOMERIGHT	6	82	63	7	12

ตาราง 4.20 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันของปัจจัย กลุ่มอาชีพ

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
JOB	1	113	86	11	16
JOB	2	173	142	12	19
JOB	3	333	254	27	52
JOB	4	648	451	67	130
JOB	5	4	3	0	1

ตาราง 4.21 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันของปัจจัย ช่วงอัตรารายได้สุทธิเทียบกับค่าแรงผ่อนชำระต่อเดือน

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
INC_RATE	1	807	606	73	128
INC_RATE	2	145	100	18	27
INC_RATE	3	145	102	13	30
INC_RATE	4	164	119	13	32
INC_RATE	5	10	9	0	1

ตาราง 4.22 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันของปัจจัย ช่วงอัตราดอกเบี้ยเทียบกับยอดที่ให้กู้ได้

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
LOAN_RATE	1	75	68	5	2
LOAN_RATE	2	62	48	1	13
LOAN_RATE	3	106	90	8	8
LOAN_RATE	4	217	168	24	25
LOAN_RATE	5	576	401	60	115
LOAN_RATE	6	235	161	19	55

ตาราง 4.23 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันของปัจจัย การมีผู้ค้ำประกัน

Attr_name	Attr_value	Count_all	Count_x	Count_y	Count_z
GUARANTEE	0	48	42	3	3
GUARANTEE	1	1223	894	114	215

คำอธิบายสำหรับฟิลด์ต่างๆ ตามตาราง 4.14 ถึง 4.23 มีดังนี้

Attr_name	หมายถึง	ปัจจัยที่นำมาคำนวณค่าสารสนเทศ
Attr_value	หมายถึง	ค่าของปัจจัยที่จำแนกได้
Count_all	หมายถึง	จำนวนรายการข้อมูล ที่เป็นไปตามค่าของปัจจัยที่จำแนก
Count_x	หมายถึง	จำนวนรายการข้อมูลที่ชุดข้อมูล(เกรด) มีค่าเป็น X ตามค่าของปัจจัยที่จำแนก
Count_y	หมายถึง	จำนวนรายการข้อมูลที่ชุดข้อมูล(เกรด) มีค่าเป็น Y ตามค่าของปัจจัยที่จำแนก
Count_z	หมายถึง	จำนวนรายการข้อมูลที่ชุดข้อมูล(เกรด) มีค่าเป็น Z ตามค่าของปัจจัยที่จำแนก

ยกตัวอย่างเช่นตาราง 4.23 ปัจจัยที่นำมาคำนวณค่าสารสนเทศคือ การมีผู้ค้ำประกัน (GUARANTEE) ค่าของปัจจัยที่จำแนกได้ มีสองค่าคือ 0 กับ 1 โดยที่ 0 หมายถึง ไม่มีผู้ค้ำ และ 1 หมายถึง มีผู้ค้ำ จำนวนรายการข้อมูลจากตารางข้อมูลเพื่อการเรียนรู้ของต้นแบบ step_31 มีทั้งสิ้น 1,271 รายการ สามารถจำแนกตามปัจจัยการมีผู้ค้ำประกันได้คือ ไม่มีผู้ค้ำ จำนวน 48 รายการ และมีผู้ค้ำ จำนวน 1,223 รายการ ในรายการที่ไม่มีผู้ค้ำ 48 รายการนั้น จำแนกตามชุดข้อมูลได้ว่ามีเกรด X จำนวน 42 รายการ, เกรด Y จำนวน 3 รายการ และเกรด Z จำนวน 3 รายการ ส่วนของมีผู้ค้ำจำนวน 1,223 รายการนั้นก็จำแนกได้ว่ามีเกรด X จำนวน 894 รายการ , เกรด Y จำนวน 114 รายการ และเกรด Z จำนวน 215 รายการ เป็นต้น

จากตารางข้อมูลการจำแนกตามปัจจัยต่างๆ ที่กล่าวไว้แล้วนั้น จะถูกนำมาคำนวณหาค่าอินฟอร์เมชันเอน ได้ดังตาราง 4.24

ตาราง 4.24 ตารางข้อมูลผลการคำนวณค่าอินฟอร์เมชันเอนของปัจจัยต่างๆ

Count_all	Count_x	Count_y	Count_z	Entropy_group	Attr_name	Entropy_attr	Gain_info	Sprit_info	Gain_ratio
1271	936	117	218	1.078107	HOMESTAT	1.077500	0.000607	1.009922	0.000601
					SEX	1.077137	0.000970	0.964448	0.001006
					GUARANTEE	1.074532	0.003575	0.231952	0.015412
					INC_RATE	1.073722	0.004384	1.566850	0.002798
					HOMERIGHT	1.073158	0.004949	1.418312	0.003490
					AGE	1.071934	0.006173	1.894019	0.003259
					STATUS	1.071323	0.006784	1.863306	0.003641
					JOB	1.069425	0.008682	1.729981	0.005019
					HOMETYPE	1.067021	0.011086	0.836488	0.013253
					LOAN_RATE	1.049807	0.028300	2.155497	0.013129

ตาราง 4.24 สามารถอธิบายข้อมูลได้ว่า จากตารางข้อมูลเพื่อการเรียนรู้ของต้นแบบ step_31 ซึ่งมีทั้งสิ้น 1,271 รายการนั้น เมื่อใช้ฟิลด์ grade เป็นตัวแบ่งชุดข้อมูลออกเป็น X, Y และ Z

ดังที่ได้กล่าวไว้แล้วนั้น ค่าสารสนเทศของกลุ่มข้อมูลกลุ่มนี้ มีค่าเท่ากับ 1.078107 โดยดูค่าจาก ฟิลด์ Entropy_group ซึ่งคำนวณได้จากสูตร

$$I(S) = \sum_{i=1}^n -\frac{p_i}{S} \log_2 \left(\frac{p_i}{S} \right)$$

$$\begin{aligned} I(936,117,218) &= -\frac{936}{1271} \log_2 \left(\frac{936}{1271} \right) - \frac{117}{1271} \log_2 \left(\frac{117}{1271} \right) - \frac{218}{1271} \log_2 \left(\frac{218}{1271} \right) \\ &= 1.078107 \end{aligned}$$

โดยที่ p_i มีค่าเท่ากับ 936, 117 และ 218 โดยดูได้จากฟิลด์ Count_x, Count_y และ Count_z ซึ่งเก็บจำนวนรายการที่ฟิลด์ grade มีค่า X, Y และ Z ตามลำดับ และ n มีค่าเท่ากับ 3 ซึ่งหมายถึงจำนวนชุดข้อมูลที่ต้องการแบ่ง นั่นคือ X, Y และ Z

สำหรับปัจจัยเพศ(SEX) มีค่าอินฟอร์เมชันเอนเท่ากับ 0.000970 โดยดูค่าจากฟิลด์ Gain_info ซึ่งคำนวณได้จากสูตร

$$Gain(S, A) = I(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} I(S_v)$$

โดยข้อมูล S_v นำมาจากตาราง 4.15 ตารางข้อมูลสำหรับคำนวณค่าอินฟอร์เมชันเอนของปัจจัยเพศ ซึ่งสามารถแสดงการคำนวณได้ตามลำดับดังนี้

$$Gain(S, Sex) = 1.078107 - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} I(S_v)$$

$$\sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} I(S_v) = \frac{776}{1271} I(Sex_1) + \frac{495}{1271} I(Sex_2)$$

$$\sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} I(S_v) = \frac{776}{1271} I(562,73,141) + \frac{495}{1271} I(374,44,77)$$

$$I(562,73,141) = -\frac{562}{776} \log_2 \left(\frac{562}{776} \right) - \frac{73}{776} \log_2 \left(\frac{73}{776} \right) - \frac{141}{776} \log_2 \left(\frac{141}{776} \right)$$

$$I(374,44,77) = -\frac{374}{495} \log_2 \left(\frac{374}{495} \right) - \frac{44}{495} \log_2 \left(\frac{44}{495} \right) - \frac{77}{495} \log_2 \left(\frac{77}{495} \right)$$

$$\sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} I(S_v) = \frac{776}{1271} 1.104962 + \frac{495}{1271} 1.033514$$

$$\sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} I(S_v) = 1.0771367$$

$$\text{Gain}(S, \text{Sex}) = 1.078107 - 1.0771367 = 0.000970$$

และสำหรับปัจจัยเพศมีค่าเกินเรโซเท่ากับ 0.001006 โดยดูค่าจากฟิลด์ Gain_ratio ซึ่งคำนวณได้จากสูตร

$$\text{GainRatio}(S, \text{Sex}) = \frac{\text{Gain}(S, \text{Sex})}{\text{SpritInfo}(S, \text{Sex})}$$

ซึ่งสามารถแสดงการคำนวณได้ตามลำดับดังนี้

$$\text{GainRatio}(S, \text{Sex}) = \frac{0.000970}{\text{SpritInfo}(S, \text{Sex})}$$

$$\text{SpritInfo}(S, \text{Sex}) = - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \log_2 \frac{|S_v|}{|S|}$$

$$\text{SpritInfo}(S, \text{Sex}) = - \frac{776}{1271} \log_2 \frac{776}{1271} - \frac{495}{1271} \log_2 \frac{495}{1271} = 0.964448$$

$$\text{GainRatio}(S, \text{Sex}) = \frac{0.000970}{0.964448} = 0.001006$$

จากตาราง 4.24 ทำให้เห็นว่าปัจจัยที่มีค่าอินฟอร์เมชันเกินสูงสุดคือปัจจัยช่วงอัตราดอกเบี้ยเทียบกับยอดที่กู้ได้ (LOAN_RATE) ซึ่งสามารถนำปัจจัยนี้ไปเป็น โหนดรากสำหรับการสร้างต้นไม้ตัดสินใจได้ในลำดับต่อไป

ผู้ศึกษาได้เขียนโปรแกรมสร้างต้นไม้ตัดสินใจ โดยสร้างตารางข้อมูลใหม่ขึ้นมา เพื่อเก็บข้อมูลต้นไม้ตัดสินใจ ตารางข้อมูลใหม่ที่สร้างในขั้นตอนนี้ชื่อ tree โดยมีโครงสร้างดังตาราง 4.25

ตาราง 4.25 โครงสร้างตารางข้อมูลต้นไม้ตัดสินใจ

ชื่อตารางข้อมูล : tree				
รายละเอียดตาราง : เก็บข้อมูลต้นไม้ตัดสินใจ				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
WHERE_TXT	VARCHAR	200	เงื่อนไขในการพิจารณา ข้อมูลที่สามารถตัดสินใจได้	LOAN_RATE='1' AND JOB='1'
RESULT	VARCHAR	10	ผลลัพธ์ของการตัดสินใจ	Same Class
CLASS_TXT	VARCHAR	1	ค่าของการตัดสินใจ	X

ตัวอย่างข้อมูลต้นไม้ตัดสินใจ จากตารางข้อมูล tree แสดงได้ดังตาราง 4.26

ตาราง 4.26 ตัวอย่างข้อมูลต้นไม้ตัดสินใจจาก ตาราง tree

WHERE_TXT	RESULT	CLASS_TXT
LOAN_RATE = '1' AND AGE = '2' AND HOMESTAT = '1' AND INC_RATE = '1' AND STATUS = '1'	Same Class	Y
LOAN_RATE = '1' AND AGE = '2' AND HOMESTAT = '1' AND INC_RATE = '1' AND STATUS = '3'	Same Class	X
LOAN_RATE = '1' AND AGE = '2' AND HOMESTAT = '1' AND INC_RATE = '2'	Same Class	X
LOAN_RATE = '1' AND AGE = '2' AND HOMESTAT = '1' AND INC_RATE = '4'	Same Class	X
LOAN_RATE = '1' AND AGE = '2' AND HOMESTAT = '3'	Same Class	X
LOAN_RATE = '1' AND AGE = '3'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '1'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '2' AND INC_RATE = '1'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '2' AND INC_RATE = '3'	Same Class	Z
LOAN_RATE = '1' AND AGE = '4' AND JOB = '3' AND GUARANTEE = '0' AND HOMERIGHT = '1'	Same Class	Y
LOAN_RATE = '1' AND AGE = '4' AND JOB = '3' AND GUARANTEE = '0' AND HOMERIGHT = '4'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '3' AND GUARANTEE = '1'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '1' AND GUARANTEE = '0'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '1' AND GUARANTEE = '1' AND INC_RATE = '1' AND HOMESTAT = '1'	Not Branch	
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '1' AND GUARANTEE = '1' AND INC_RATE = '1' AND HOMESTAT = '3'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '1' AND GUARANTEE = '1' AND INC_RATE = '2'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '2' AND GUARANTEE = '0' AND HOMERIGHT = '1' AND STATUS = '2'	Same Class	Z
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '2' AND GUARANTEE = '0' AND HOMERIGHT = '1' AND STATUS = '6'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '2' AND GUARANTEE = '0' AND HOMERIGHT = '4'	Same Class	X
LOAN_RATE = '1' AND AGE = '4' AND JOB = '4' AND SEX = '2' AND GUARANTEE = '1'	Same Class	X
LOAN_RATE = '1' AND AGE = '5' AND JOB = '1' AND GUARANTEE = '0'	Same Class	X
LOAN_RATE = '1' AND AGE = '5' AND JOB = '1' AND GUARANTEE = '1'	Not Branch	
LOAN_RATE = '1' AND AGE = '5' AND JOB = '3'	Same Class	X
LOAN_RATE = '1' AND AGE = '5' AND JOB = '4'	Same Class	X
LOAN_RATE = '2' AND INC_RATE = '1' AND STATUS = '1' AND JOB = '2'	Same Class	X

คำอธิบายสำหรับฟิลต์ต่างๆ ตามตาราง 4.26 มีดังนี้

Where_txt หมายถึง เงื่อนไขในการพิจารณาข้อมูลสำหรับการตัดสินใจ

Result หมายถึง ผลลัพธ์ของการตัดสินใจ มีได้ 2 กรณีคือ Same Class และ Not

Branch โดยที่ Same Class หมายถึง ข้อมูลผลลัพธ์ที่ได้จากเงื่อนไขของ Where_txt นั้น เกิดกลุ่มเดียวกันแล้วไม่จำเป็นต้องใช้ปัจจัยอื่นๆ มาแบ่งกลุ่มข้อมูลอีก และผลลัพธ์ของการตัดสินใจสามารถระบุได้ว่ามีค่าเป็นอย่างไรโดยดูได้จากฟิลต์ Class_txt ส่วน Not Branch หมายถึง ไม่สามารถช่วย

ตัดสินใจได้เนื่องจากข้อมูลผลลัพธ์ที่ได้จากเงื่อนไขของ Where_txt นั้น เกิดกลุ่มข้อมูลขึ้นอย่างน้อย 2 กลุ่มข้อมูล และไม่สามารถหาปัจจัยอื่นๆ มาแบ่งกลุ่มข้อมูลได้อีก

Class_txt หมายถึง ค่าของการตัดสินใจ มีได้ 2 กรณีคือ กรณีที่หนึ่ง ไม่มีค่า (ค่าของฟิลด์เป็นค่าว่าง) หมายถึงข้อมูลผลลัพธ์จากเงื่อนไขของ Where_txt เกิดกลุ่มข้อมูลขึ้น 2 กลุ่ม ทำให้ไม่สามารถช่วยตัดสินใจได้ กรณีที่สองมีค่าเป็น X, Y หรือ Z ตัวใดตัวหนึ่ง ซึ่งหมายถึงข้อมูลผลลัพธ์จากเงื่อนไขของ Where_txt เกิดกลุ่มข้อมูลเดียวกัน นั่นคือสามารถช่วยตัดสินใจได้ว่า ลูกค้าที่มีปัจจัยดังกล่าวตามเงื่อนไขของ Where_txt เป็นลูกค้าเกรด X, Y หรือ Z

จากตารางข้อมูลต้นไม้ตัดสินใจ tree ซึ่งมีทั้งหมด 571 รายการ เมื่อนำมาพิจารณาค่าของฟิลด์ Result เพื่อหาร้อยละของการช่วยตัดสินใจแบบทำนายได้ (ฟิลด์ Result มีค่าเป็น Same Class) กับสามารถช่วยตัดสินใจแบบทำนายคลาดเคลื่อน (ฟิลด์ Result มีค่าเป็น Not Branch) จะพบว่าร้อยละ 80.91 ของต้นไม้ตัดสินใจนี้ สามารถช่วยตัดสินใจแบบทำนายได้ และร้อยละ 19.09 สามารถช่วยตัดสินใจแบบทำนายคลาดเคลื่อน ประเด็นนี้พิจารณาได้ว่าข้อมูลเพื่อการเรียนรู้ที่นำมาสร้างต้นแบบนี้เป็นข้อมูลที่เกิดจากการพิจารณาการให้สินเชื่อโดยผู้เชี่ยวชาญและประสบการณ์ที่มีอยู่เดิมซึ่งมีหลายคน แต่ละคนก็มีหลักเกณฑ์ในการพิจารณาที่แตกต่างกัน นอกจากนี้ยังอาจเกิดกรณีลูกค้าที่ได้รับการอนุมัติเป็นพิเศษเนื่องจากการสนิทสนมกันเป็นพิเศษ ทำให้ละเลยกฎเกณฑ์ไม่พิจารณาการให้สินเชื่อตามหลักเกณฑ์ปกติ

จากตารางข้อมูลต้นไม้ตัดสินใจ tree สามารถนำมาแสดงในรูปต้นไม้เพื่อให้เห็นภาพที่ชัดเจนและเข้าใจง่ายๆ ดังจะแสดงไว้ในภาคผนวก

4.5 การตรวจสอบผลของต้นแบบในการทำนาย

ขั้นตอนนี้ผู้ศึกษาได้เขียนโปรแกรมเพื่อนำข้อมูลส่วนที่ใช้ทดสอบต้นแบบ นั่นคือข้อมูลจากตาราง step_32 มาทดสอบกับต้นแบบว่าเกรดที่ทำนายได้จากต้นแบบ เมื่อนำมาเปรียบเทียบกับเกรดที่เป็นอยู่จริงของข้อมูลรายการนั้น ถูกต้องตรงกันหรือไม่ มากน้อยเพียงใด

ผู้ศึกษาได้เขียนโปรแกรมทดสอบตัวแบบ และสร้างตารางข้อมูลใหม่ขึ้นมาเพื่อเก็บข้อมูลผลการทดสอบ ตารางข้อมูลใหม่ที่สร้างในขั้นตอนนี้ชื่อ test_tree โดยมีโครงสร้างดังตาราง 4.27

ตาราง 4.27 โครงสร้างตารางข้อมูลเก็บผลการทดสอบต้นแบบ

ชื่อตาราง : test_tree				
รายละเอียดตาราง : เก็บข้อมูลผลการทดสอบต้นแบบ				
ชื่อรายการ	ประเภท	ขนาด	คำอธิบาย	ตัวอย่างข้อมูล
AGE	CHAR	1	ช่วงอายุลูกค้า	3
SEX	CHAR	1	เพศ	2
STATUS	CHAR	1	สถานภาพ	1
HOMESTAT	CHAR	1	สถานะในที่อยู่อาศัย	1
HOMETYPE	CHAR	1	ลักษณะที่อยู่อาศัย	2
HOMERIGHT	CHAR	1	กรรมสิทธิ์ในที่อยู่อาศัย	3
JOB	CHAR	1	กลุ่มอาชีพ	4
INC_RATE	CHAR	1	ช่วงอัตรารายได้สุทธิเทียบกับค่างวดผ่อนชำระต่อเดือน	5
LOAN_RATE	CHAR	1	ช่วงอัตราดอกเบี้ยเทียบกับยอดที่ให้กู้ได้	1
GUARANTEE	CHAR	1	การมีผู้ค้ำประกัน	1
GRADE	CHAR	1	ระดับความน่าเชื่อถือ	X
WHERE_TXT	VARCHAR	200	เงื่อนไขในการพิจารณา ข้อมูลที่สามารถตัดสินใจได้	LOAN_RATE='1' AND JOB='1'
RESULT	VARCHAR	10	ผลลัพธ์ของการตัดสินใจ	Same Class
CLASS_TXT	VARCHAR	1	ค่าของการตัดสินใจ	X

จากตารางข้อมูลเพื่อทดสอบต้นแบบ step_32 ซึ่งมีทั้งหมด 545 รายการ เมื่อนำมาทดสอบกับต้นแบบว่าข้อมูลเข้าได้กับเงื่อนไขใดของต้นแบบ ผลลัพธ์ที่ได้คืออะไรและเกรดที่ทำนายได้จากต้นแบบเป็นอย่างไร แล้วนำค่าที่ได้มาจัดเก็บในตาราง test_tree ทั้งหมด

เมื่อนำข้อมูลจากตาราง test_tree มาพิจารณาค่าของฟิลด์ Result พบว่ามีค่าที่เป็นไปได้อยู่ 3 กรณีคือ กรณีที่หนึ่ง Result มีค่าว่าง ซึ่งหมายถึงข้อมูลเพื่อการทดสอบไม่สามารถเข้าเงื่อนไขใดๆ ได้เลยในตารางต้นแบบ กรณีที่สอง Result มีค่าเป็น Same Class ซึ่งหมายถึงข้อมูลเพื่อการทดสอบ

สามารถเข้าเงื่อนไขใดเงื่อนไขหนึ่งในตารางต้นแบบ ผลการทำนายเป็นแบบทำนายได้ จะเก็บเงื่อนไขที่เข้ากันได้และผลกรดในการทำนายไว้ที่ฟิลด์ where_txt และ class_txt ตามลำดับ กรณีที่สาม Result มีค่าเป็น Not Branch ซึ่งหมายถึงข้อมูลเพื่อการทดสอบสามารถเข้าเงื่อนไขใดเงื่อนไขหนึ่งในตารางต้นแบบ ผลการทำนายเป็นแบบทำนายคลาดเคลื่อน จะเก็บเงื่อนไขที่เข้ากันได้ไว้ที่ฟิลด์ where_txt ส่วนฟิลด์ class_txt จะเป็นค่าว่าง

ข้อมูลจากตาราง test_tree ทั้งสิ้นจำนวน 545 รายการนั้น เมื่อจำแนกตามกรณีทั้ง 3 ดังกล่าว จะพบว่า เป็นกรณีที่หนึ่งจำนวน 51 รายการคิดเป็นร้อยละ 9.36 กรณีที่สองจำนวน 377 รายการคิดเป็นร้อยละ 69.17 และกรณีที่สามจำนวน 117 รายการคิดเป็นร้อยละ 21.47 กรณีนี้สามารถอธิบายได้ว่าต้นแบบที่สร้างขึ้นมาสามารถช่วยตัดสินใจได้ทั้งสิ้นร้อยละ 90.64 โดยเป็นการช่วยตัดสินใจแบบทำนายได้ร้อยละ 69.17 และแบบทำนายคลาดเคลื่อนร้อยละ 21.47 และไม่สามารถช่วยตัดสินใจได้ร้อยละ 9.36

เมื่อพิจารณาลึกลงไปในกรณีที่ช่วยตัดสินใจแบบทำนายได้นั้นจะพบว่า จากทั้งหมด 377 รายการ ตัวแบบสามารถทำนายได้ตรงกับข้อมูลจริงที่เป็นอยู่ทั้งสิ้น 252 รายการคิดเป็นร้อยละ 66.84 และทำนายไม่ตรงทั้งสิ้น 125 รายการคิดเป็นร้อยละ 33.16

ในอีกมุมมองหนึ่งถ้าตัดจำนวนรายการที่ไม่ช่วยตัดสินใจจำนวน 51 รายการออกไป จะเหลือรายการทั้งสิ้น 494 รายการ และพิจารณาจำแนกกลุ่มออกเป็น 2 กรณี กรณีที่หนึ่งคือช่วยตัดสินใจแบบทำนายได้และถูกต้อง รวมกับช่วยตัดสินใจแบบทำนายคลาดเคลื่อน จะได้จำนวนรายการทั้งสิ้น 369 รายการคิดเป็นร้อยละ 74.70 กรณีที่สองคือช่วยตัดสินใจแบบทำนายได้แต่ไม่ถูกต้องจำนวน 125 รายการคิดเป็นร้อยละ 25.30

ในกรณีของรายการที่ไม่ช่วยตัดสินใจจำนวน 51 รายการนั้นไม่ได้เป็นเพราะต้นแบบไม่ถูกต้อง แต่เป็นเพราะข้อมูลเพื่อการเรียนรู้ในการนำมาสร้างตัวแบบนั้นมีความไม่สมบูรณ์อยู่ในตัวเองอยู่แล้วในระดับหนึ่ง เพราะเป็นข้อมูลที่เกิดจากการตัดสินใจโดยใช้ผู้เชี่ยวชาญหลายคน และธรรมชาติของการพิจารณาสินเชื่อที่มีระยะยาวนานไม่เท่ากัน ความสามารถในการผ่อนชำระของลูกค้าแปรเปลี่ยนไปตามสถานะเศรษฐกิจ ย่อมส่งผลให้ความน่าเชื่อถือที่พิจารณาจากการผ่อนชำระมีความผันแปรค่อนข้างสูง

4.6 ประเมินความถูกต้องของต้นแบบ

ขั้นตอนนี้ผู้ศึกษาได้ทำการประเมินความถูกต้องของต้นแบบ โดยกำหนดระดับความถูกต้องที่พึงพอใจของคำทำนายที่ได้จากต้นแบบ เพื่อนำมาเป็นเกณฑ์ในการเปรียบเทียบคำทำนายที่ได้จากต้นแบบ และปรับปรุงต้นแบบจนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจ

จากวัตถุประสงค์ของการศึกษาเพื่อพัฒนาระบบช่วยสนับสนุนการตัดสินใจการให้
สินเชื่อ และผลลัพธ์ที่ได้จากการตรวจสอบผลของต้นแบบในการทำงาน สามารถสรุปได้ว่า
ต้นแบบที่สร้างขึ้นมาเพื่อช่วยสนับสนุนการตัดสินใจในการศึกษาครั้งนี้อยู่ในเกณฑ์ที่สามารถ
ยอมรับได้เพราะสามารถช่วยงานได้ร้อยละ 74.70



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved