ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

การพัฒนาระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเออัลกอริทึม

DEVELOPMENT OF ASSOCIATIVE CLASSIFICATION SYSTEM BY USING CBA ALGORITHM

H006388

โดย

วัลภรณ์ เอตะนาม

WANLAPORN ATANAM

อาจารย์ที่ปรึกษา

รศ. ดร. วรพจน์ กรีสุระเดช

เลขหมู่.....
เลขทะเบียน 06388
วัน,เดือน,ปี 14 มี.ค. 2554

.b.....
.i.....

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 2 ปีการศึกษา 2552

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

DEVELOPMENT OF ASSOCIATIVE CLASSIFICATION SYSTEM BY USING CBA ALGORITHM

WANLAPORN ATANAM

A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS OF THE COURSE

SYSTEM DEVELOPMENT PROJECT

MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2/ 2009

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2010

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใบรับรองโครงการพัฒนาระบบงาน (System Development Project)

เรื่อง

การพัฒนาระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ แบบ ซีบีเอ อัลกอริทึม

DEVELOPMENT OF ASSOCIATIVE CLASSIFICATION SYSTEM BY USING CBA ALGORITHM

นางสาววัลภรณ์ เอตะนาม

รหัสประจำตัว 49066711

ขอรับรองว่ารายงานฉบับนี้ ข้าพเจ้าไม่ได้คัดลอกมาจากที่ใด
รายงานฉบับนี้ได้รับการตรวจสอบและอนุมัติให้เป็นส่วนหนึ่งของ
การศึกษาวิชาโครงการพัฒนาระบบงาน หลักสูตรวิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศ)
ภาคเรียนที่ 2 ปีการศึกษา 2552

.....อาจารย์ที่ปรึกษา
(รศ.ดร. วรพจน์ กรีสุระเดช)

.....กรรมการสอบ
(รศ.ดร. อาริต ธรรมโน)

.....กรรมการสอบ
(ผศ.ดร.ภัทรชัย ลลิตโรจน์วงศ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การพัฒนาระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ อัลกอริทึม
นักศึกษา	นางสาววัลภรณ์ เอตะนาม
รหัสนักศึกษา	49066711
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2552
อาจารย์ที่ปรึกษา	รศ.ดร.วรพจน์ กรีสุระเดช

บทคัดย่อ

เทคนิคการจำแนกประเภทข้อมูลและเทคนิคการค้นหากฎความสัมพันธ์เป็นเทคนิคที่สำคัญของการทำเหมืองข้อมูล โครงการพัฒนาระบบฉบับนี้นำเสนอขั้นตอนและพัฒนาระบบงานการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ อัลกอริทึม โดยนำเทคนิคการจำแนกประเภทข้อมูลมาสร้างตัวแบบจำลองจำแนกประเภทข้อมูลจากกฎความความสัมพันธ์ที่ได้จากเทคนิคการค้นหากฎความสัมพันธ์อีกทีหนึ่ง ตัวแบบจำลองจำแนกประเภทข้อมูลที่ได้นี้จะสามารถนำไปใช้ช่วยในการตัดสินใจและนำไปใช้ประโยชน์ต่อไป ซึ่งตัวแบบจำลองจำแนกประเภทข้อมูลที่สร้างได้นี้มีประสิทธิภาพและถูกต้องแม่นยำกว่าเทคนิคการจำแนกประเภททั่วไป อีกทั้งเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ยังช่วยแก้ปัญหาต่างๆที่เกิดขึ้นจากการใช้เทคนิคการจำแนกประเภทข้อมูลทั่วไป

ในส่วนของการวิเคราะห์ ออกแบบและพัฒนาระบบ ผู้เขียนได้วิเคราะห์และออกแบบระบบด้วยแผนภาพแบบจำลอง ยูเอ็มแอล ทำการพัฒนาระบบด้วยภาษา Visual Basic .NET และจัดการข้อมูลด้วยระบบจัดการฐานข้อมูล Microsoft SQL Server 2005

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Development of Associative Classification System by Using CBA Algorithm
Student	Ms. Wanlaporn Atanam
Student ID.	49066711
Degree	Master of Science
Program	Information Technology
Major	Information Science
Academic Year	2009
Advisor	Assoc. Prof. Dr. Worapoj Kreesuradej

ABSTRACT

Data Classification and Association Rule Discovery are two important data mining techniques. This project is deal with the development of Associative Classification by using CBA Algorithm. The classifier, built this way, is more accurate than the one produced by typical classification systems. In addition, this associative classification technique helps to solve a number of problems that exist in the current classification system.

In the analysis design and development part, we have analyzed and designed the system by using UML diagram, Visual Basic.NET language and Microsoft SQL Server 2005.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงการพัฒนาระบบสำเร็จได้อย่างดี ด้วยคำแนะนำ ความช่วยเหลือ และคำปรึกษาจากบุคคลต่างๆเหล่านี้

ขอขอบคุณ รศ.ดร. วรพจน์ กรีสุระเดช ซึ่งเป็นอาจารย์ที่ปรึกษา ที่กรุณาให้คำปรึกษา และให้แนะนำต่างๆในการพัฒนาระบบงาน

ขอกราบพระคุณคณาจารย์ภาควิชาวิทยาการสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า

ขอขอบคุณเพื่อนๆ พี่ๆ ในภาควิชาวิทยาการสารสนเทศ สถาบันเทคโนโลยี พระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกคนที่ให้คำแนะนำต่างๆ และคอยให้กำลังใจเสมอมา

ขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกเรื่องๆ ทำให้ข้าพเจ้าสามารถทำโครงการฉบับนี้สำเร็จลุล่วงด้วยดี

วัลภรณ์ เอตะนาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....I

บทคัดย่อภาษาอังกฤษ.....II

กิตติกรรมประกาศ.....III

สารบัญ..... IV

สารบัญตาราง..... VI

สารบัญรูป..... VII

บทที่ 1 บทนำ.....1

1.1 ความเป็นมาและความสำคัญของปัญหา.....1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 ทฤษฎีและแนวคิดที่ใช้ในการวิจัย.....2

1.4 ขอบเขตของการพัฒนาระบบ.....3

1.5 ขั้นตอนของการศึกษา.....3

บทที่ 2 การทำเหมืองข้อมูล.....4

2.1 ความหมายของการทำเหมืองข้อมูล.....4

2.2 กระบวนการทำเหมืองข้อมูล.....4

2.3 ทำเหมืองข้อมูล.....6

2.4 การจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ 12

2.5 CBA อัลกอริทึม (Classification Based on Associations) 13

บทที่ 3 วิเคราะห์และออกแบบระบบ..... 24

3.1 ความต้องการและข้อจำกัดของโปรแกรม..... .24

3.2 การวิเคราะห์ระบบโดยใช้ยูสเคสไดอะแกรม และ แอคทิวิตีไดอะแกรม..... .25

3.3 การออกแบบระบบโดยใช้คลาสไดอะแกรมและซีเควนไดอะแกรม.....43

3.4 การออกแบบฐานข้อมูล..... 48

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

บทที่ 4 การออกแบบและพัฒนาระบบ..... ……...50
4.1 เครื่องมือที่ใช้ในการออกแบบและพัฒนาระบบ..... 50
4.2 การออกแบบหน้าจอการรับข้อมูล.....51
4.3 การออกแบบหน้าจอรายงานผลลัพธ์.....51
4.4 การใช้งานโปรแกรม..... ……..52

บทที่ 5 สรุปผลการวิจัย และข้อเสนอแนะ..... …..67
5.1 สรุปผลการพัฒนาระบบ..... ……...67
5.2 ปัญหาที่พบในการพัฒนา..... 68
5.3 ข้อเสนอแนะ.....69

บรรณานุกรม.....71

ประวัติผู้เขียน.....72

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

2.1 ตัวอย่างข้อมูลสภาพอากาศ และการตัดสินใจว่าจะออกไปนอกบ้านหรือไม่.....9

2.2 ตัวอย่างข้อมูลรายการซื้อสินค้า.....11

2.3 ตัวอย่างกฎความสัมพันธ์ที่หาได้จากเทคนิคการค้นหากฎความสัมพันธ์.....11

2.4 ตัวอย่างข้อมูลคนไข้.....20

2.5 อัลกอริทึม CBA-RG ทำการค้นหา Frequent ruleitems รอบแรก.....20

2.6 อัลกอริทึม CBA-RG ทำการค้นหา Frequent ruleitems รอบที่สอง.....21

2.7 CBA-RG สร้าง Class-Association Rule(CARs) ที่ผ่านค่าสนับสนุน.....22

2.8 CBA-RG สร้าง Class-Association Rules (CARs) ที่ผ่านค่าความเชื่อมั่น.....22

2.9 CBA-CB เรียงกฎตามค่าความเชื่อมั่น.....23

2.10 CBA-CB สร้างแบบจำลองในการทำนาย.....23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

2.1 ขั้นตอนของเทคนิคการจำแนกประเภทข้อมูล....8
2.2 ตัวอย่างแบบจำลองที่ได้จากเทคนิคการจำแนกประเภทข้อมูล....9
2.3 ตัวอย่างกฎความสัมพันธ์....10
2.4 ตัวอย่างรูปแบบกฎความสัมพันธ์แบบมีคลาส (CARs)....12
2.5 ขั้นตอนของเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์....13
2.6 ตัวอย่างรูปแบบของ ruleitem....14
2.7 ตัวอย่างรูปแบบ ruleitem....14
2.8 ตัวอย่างรูปแบบกฎความสัมพันธ์แบบ CARs....16
2.9 ตัวอย่างรูปแบบ Frequent ruleitem....16
2.10 อัลกอริทึม CBA-RG....17
2.11 ตัวอย่างรูปแบบ Classifier....18
2.12 อัลกอริทึม CBA-CB....19
3.1 ภาพรวมระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์....26
3.2 คำอธิบายยูสเคส UC1: selectData....28
3.3 แอคทิวิตีไดอะแกรม select Data....29
3.4 คำอธิบายยูสเคส UC2: clean Data....30
3.5 แอคทิวิตีไดอะแกรม clean Data....30
3.6 คำอธิบายยูสเคส UC3: transform Data....31
3.7 แอคทิวิตีไดอะแกรม transform Data....32
3.8 คำอธิบายยูสเคส UC4: buildClassifier....33
3.9 แอคทิวิตีไดอะแกรม buildClassifier....34
3.10 คำอธิบายยูสเคส UC5: saveClassifierModel....35
3.11 แอคทิวิตีไดอะแกรม saveClassifierModel....36
3.12 คำอธิบายยูสเคส UC6: testClassifierModel....37
3.13 แอคทิวิตีไดอะแกรม testClassifierMode....38
3.14 คำอธิบายยูสเคส UC7: showClassifierModel....39
3.15 แอคทิวิตีไดอะแกรม showClassifierModel....40
3.16 คำอธิบายยูสเคส UC8: predictData....41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

3.17 แอคทิวิตีไดอะแกรม predictData 42

3.18 คลาสไดอะแกรมการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ 43

3.19 ซีเควนไดอะแกรมการสร้างแบบจำลอง.....45

3.20 ซีเควนไดอะแกรมการทำนายประเภทข้อมูล.....47

3.21 อีอาร์ไดอะแกรมการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ.....48

4.1 หน้าจอการรับข้อมูล.....51

4.2 หน้าจอแสดงผลลัพธ์การประมวลผล.....51

4.3 เมนูการทำงานของระบบในส่วนขั้นตอนของการเตรียมข้อมูล.....52

4.4 เมนูการทำงานของระบบในส่วนขั้นตอนการทำเหมืองข้อมูล.....52

4.5 ขั้นตอนการติดต่อฐานข้อมูล.....52

4.6 กล่องข้อความเมื่อระบบเชื่อมต่อกับฐานข้อมูลเรียบร้อยแล้ว.....52

4.7 การเลือกข้อมูลจากตารางและแอตทริบิวท์.....53

4.8 การเลือกข้อมูลแอตทริบิวท์สำหรับการทำเหมืองข้อมูล.....53

4.9 การคลีนข้อมูล.....54

4.10 กล่องข้อความเมื่อทำขั้นตอนการคลีนข้อมูลเรียบร้อย.....54

4.11 ขั้นตอนการ Discrete ข้อมูล.....55

4.12 กล่องข้อความเมื่อทำขั้นตอนการแปลงข้อมูลเรียบร้อยแล้ว.....56

4.13 แสดงคอลัมน์ที่แปลงข้อมูลแล้วเป็นข้อมูลประเภท Discrete.....56

4.14 การขั้นตอนการสร้างแบบจำลองแบบซีบีเอ อัลกอริทึม.....56

4.15 ขั้นตอนการสร้างแบบจำลองแบบซีบีเอ เลือกข้อมูลสำหรับทำเหมืองข้อมูล.....57

4.16 แสดงข้อความในกล่องข้อความสร้างแบบจำลองเรียบร้อย.....58

4.17 แสดงผลลัพธ์แบบจำลองที่ถูกสร้างขึ้น.....58

4.18 แสดงผลแบบจำลองที่ถูกจัดเรียงตามค่าความเชื่อมั่น.....59

4.19 แสดงผลแบบจำลองที่ถูกจัดเรียงตามค่าสนับสนุน.....60

4.20 ฟอร์มการบันทึกชื่อ Classifier Model Name.....60

4.21 แสดงข้อความยืนยันว่าบันทึกแบบจำลองเรียบร้อย.....60

4.22 หน้าจอการเลือกข้อมูลสำหรับการทดสอบระบบ.....61

4.23 แสดงแบบจำลองโดยพิจารณาจากค่าสนับสนุนและค่าความเชื่อมั่นต่ำสุด.....62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.24 แสดงผลลัพธ์การทดสอบแบบจำลองจากข้อมูลทดสอบระบบ..... 62
4.25 เปรียบเทียบการประเมินความถูกต้องของแบบจำลองและการทดสอบความแม่นยำ... 62
4.26 ขั้นตอนการเลือกข้อมูลจากฐานข้อมูลสำหรับการทำนายกลุ่มข้อมูล..... 63
4.27 หน้าจอ ClassifierModel แสดงผลแบบจำลองที่เลือก..... 64
4.28 หน้าจอการทำนายข้อมูล..... 64
4.29 แสดงผลลัพธ์การทำนายข้อมูล..... 65
4.30 แสดงแบบจำลองโดยพิจารณาค่าจากความถูกต้องของแบบจำลอง..... 66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในโลกยุคปัจจุบันมีการแข่งขันทางด้านธุรกิจสูงมาก องค์กรต่างๆ จึงต้องคิดค้นกลยุทธ์เพื่อให้สามารถสร้างโอกาสและสร้างความได้เปรียบทางธุรกิจเหนือคู่แข่ง ซึ่งเทคโนโลยีสารสนเทศนับว่าเป็นหัวใจสำคัญในการสร้างความได้เปรียบทางธุรกิจดังกล่าว เกือบทุกองค์กรทั่วโลกได้เล็งเห็นถึงความสำคัญและนำเทคโนโลยีสารสนเทศเข้ามาประยุกต์ใช้ทั้งในด้านการปฏิบัติงานและงานทางด้านการบริหารองค์กร ซึ่งแน่นอนแต่ละองค์กรจะต้องมีการจัดเก็บจัดการข้อมูลให้อยู่ในรูปแบบที่เหมาะสม เช่น ระบบฐานข้อมูล ซึ่งเมื่อผู้ใช้งานต้องการเรียกดู(Query) ข้อมูลสามารถทำการสืบค้นข้อมูลได้จากข้อมูลที่ได้ถูกจัดเก็บไว้ในรูปแบบของระบบฐานข้อมูลขนาดใหญ่ อีกทั้งผู้ใช้ยังสามารถจัดเก็บ และจัดการกับข้อมูลเหล่านั้นได้ด้วย ซึ่งนอกเหนือจากเทคนิคการจัดการข้อมูลด้วยระบบฐานข้อมูลแล้ว ยังมีเทคนิคการทำเหมืองข้อมูล(Data Mining) ซึ่งมีประโยชน์และมีประสิทธิภาพสูงสุดช่วยในการดำเนินงานขององค์การ เทคนิคการทำเหมืองข้อมูลมีประโยชน์ คือ ช่วยการสืบค้นความรู้ ค้นหาความสัมพันธ์และรูปแบบทั้งหมดที่ถูกซ่อนอยู่ เมื่อได้ความรู้นั้นมาแล้วจึงนำมาใช้เพื่อช่วยในการตัดสินใจ และนอกจากนี้แล้วอาจจะได้รับข้อมูลใหม่ๆที่เป็นประโยชน์ทำให้การวิเคราะห์ข้อมูลมีประสิทธิภาพมากขึ้น

ในการสืบค้นความรู้บนฐานข้อมูลขนาดใหญ่ด้วยเทคนิคการจำแนกประเภทข้อมูลทั่วไป (Classification System) เช่น เทคนิคของ C4.5 นั้นยังพบปัญหาที่เทคนิคการจำแนกประเภทข้อมูลทั่วไปเกิดขึ้น เช่น การที่เทคนิคนี้ไม่สามารถค้นหา (Discovered) เซตของกฎความสัมพันธ์ขนาดเล็ก ผู้ใช้ไม่สามารถเข้าใจถึงเหตุผลที่มาของกฎความสัมพันธ์ ไม่สามารถบอกความสัมพันธ์ของกฎทางด้านซ้ายได้ (no sense) เป็นต้น โดยเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative Classification) สามารถช่วยแก้ปัญหาที่เกิดขึ้นได้โดยใช้ขั้นตอนวิธีแบบ Classification Based on Associations เข้ามาใช้ในการสร้างตัวแบบจำลองเพื่อใช้ในการทำนายกลุ่มข้อมูล โดยข้อดีของเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์เมื่อเปรียบเทียบกับเทคนิคการจำแนกประเภทข้อมูลทั่วไป เช่น เทคนิคของ C4.5 นั่นก็คือ เทคนิคนี้สามารถค้นหากฎความสัมพันธ์ขนาดเล็กได้(Small Set of Rules) ทำให้มีความแม่นยำในการทำนายสูง ถึงแม้ว่าข้อมูลที่จะนำมาใช้จะมีความซับซ้อนและมีความหลากหลายของคลาสก็ตาม รวมถึงสามารถที่จะให้เหตุผลในการทำนายข้อมูลได้โดยดูจากกฎความสัมพันธ์ที่ใช้ในการทำนาย ดังนั้นผู้ใช้สามารถทำความเข้าใจกับผลของการทำนายได้ง่าย ส่วนเหตุผลที่ทำให้เทคนิคดังกล่าวมีประสิทธิภาพและมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความแม่นยำในการทำนายสูงนั้น ขึ้นอยู่กับปัจจัยทั้ง 2 ส่วนคือ ในส่วนของการสร้างกฎความสัมพันธ์(Rule Generator) และในส่วนของการสร้างตัวแบบจำลองในการทำนายข้อมูล (Classifier Builder)

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาและทำความเข้าใจเกี่ยวกับขั้นตอนกระบวนการทำเหมืองข้อมูล
2. เพื่อศึกษาการนำหลักการทำงานทางด้านการทำเหมืองข้อมูล โดยใช้เทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์
3. เพื่อนำความรู้ ความเข้าใจที่ได้จากการศึกษา ค้นคว้ามาประยุกต์ใช้ให้เกิดประโยชน์และช่วยแก้ปัญหาในกระบวนการทำงานต่างของแต่ละธุรกิจ
4. สร้างซอฟต์แวร์ที่สามารถทำนายประเภทข้อมูลจากแบบจำลองต้นแบบ
5. เพื่อให้ได้รูปแบบของตัวแบบจำลองเพื่อใช้ในการวิเคราะห์ วางแผน และช่วยในการตัดสินใจได้อย่างถูกต้องแม่นยำ

1.3 ทฤษฎีและแนวคิดที่ใช้ในการวิจัย

ในการพัฒนาระบบ การพัฒนาระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเออัลกอริทึม(Classification Based on Associations) ซึ่งเป็นอัลกอริทึมพื้นฐานของเทคนิคของเทคนิคการจำแนกประเภทโดยใช้กฎความสัมพันธ์(Associative Classification) เป็นเทคนิคที่เกิดจากการรวมกันระหว่าง 2 เทคนิค คือ เทคนิคการจำแนกประเภทข้อมูล (Data Classification) และ เทคนิคการสืบค้นกฎความสัมพันธ์ (Association Rule Discovery) โดยจุดประสงค์ของเทคนิคการจำแนกประเภทข้อมูลคือ เพื่อสืบค้นความรู้เพื่อสรุปหาแบบจำลองของฐานข้อมูลนั้นๆ(Quinlan, 1993) เพื่อใช้ในการทำนายข้อมูลใหม่ และจุดประสงค์ของเทคนิคการสืบค้นกฎความสัมพันธ์คือเพื่อค้นหากฎความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ที่มีอยู่ เพื่อนำกฎที่ได้เหล่านั้นไปวิเคราะห์เพื่อช่วยในการตัดสินใจ โดยกฎความสัมพันธ์ทั้งหมดมีความสำคัญและบ่งบอกถึงคุณลักษณะของฐานข้อมูล โดยกฎเหล่านี้จะต้องผ่านค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำด้วย โดยเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์นั้นได้แบ่งออกเป็น 2 ส่วนหลักๆ คือ ส่วนที่ใช้ในการสร้างกฎความสัมพันธ์(Rule generator phase) และส่วนที่นำกฎความสัมพันธ์ไปสร้างแบบจำลองเพื่อใช้ในการทำนายข้อมูล(classifier builder phase)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ขอบเขตของการพัฒนาระบบ

การทำงานของระบบ การพัฒนาระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเอ อัลกอริทึม มีขอบเขตการศึกษาและพัฒนาดังนี้

1. นำข้อมูลเข้าสู่กระบวนการเตรียมข้อมูล(Preparation Data) ในกรณีที่ข้อมูลคลาสปลายทางเป็นข้อมูลที่มีค่าต่อเนื่อง(Continuous) จะต้องทำการแปลงประเภทข้อมูล (Discretizing data)ให้เป็นประเภทข้อมูลแบบช่วงข้อมูล(Categorical) เพื่อสร้างเป็นคลาสปลายทาง พร้อมสำหรับการทำนายกลุ่มข้อมูล
2. นำข้อมูลเข้าสู่กระบวนการทำเหมืองข้อมูล โดยทำงานในส่วนแรกของการสร้างกฎความสัมพันธ์แบบมีคลาส(Class Association Rules)
3. ส่วนที่ 2 คือสร้างตัวแบบจำลองจากกฎความสัมพันธ์แบบมีคลาสเพื่อเป็นแบบจำลองต้นแบบสำหรับใช้ในการทำนายข้อมูล

1.5 ขั้นตอนของการศึกษา

โครงการพัฒนาระบบงานฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ

บทที่ 1 กล่าวถึงความเป็นมาและความสัมพันธ์ของปัญหาของการพัฒนาระบบงาน ความมุ่งหมายและวัตถุประสงค์ของการศึกษา ทฤษฎีและแนวคิดที่ใช้ในการวิจัย ขอบเขตของการพัฒนาระบบงานและขั้นตอนการศึกษา

บทที่ 2 กล่าวถึงทฤษฎีที่เกี่ยวข้องกับการทำเหมืองข้อมูล และเทคนิคที่สำคัญของการทำเหมืองข้อมูล

บทที่ 3 กล่าวถึงการวิเคราะห์และออกแบบระบบ

บทที่ 4 กล่าวถึงการออกแบบและพัฒนาระบบ

บทที่ 5 กล่าวถึงสรุปและข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
การทำเหมืองข้อมูล

2.1 ความหมายของการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data mining) คือ วิธีการสืบค้นความรู้ ความสัมพันธ์ และรูปแบบทั้งหมดที่ถูกซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ การทำเหมืองข้อมูลจะทำการสำรวจและวิเคราะห์อย่างอัตโนมัติจากข้อมูลที่มีอยู่แล้วให้อยู่ในรูปแบบที่เต็มไปด้วยความหมายและอยู่ในรูปของกฎความสัมพันธ์ โดยที่ความสัมพันธ์เหล่านี้แสดงให้เห็นถึงความรู้ต่างๆที่เป็นประโยชน์และน่าสนใจ ประโยชน์หลักของการทำเหมืองข้อมูลคือ การสืบค้นความรู้ในฐานข้อมูลขนาดใหญ่เมื่อได้ความรู้นั้นมาแล้วจึงนำไปใช้ในการตัดสินใจ

2.2 กระบวนการทำเหมืองข้อมูล

กระบวนการทำเหมืองข้อมูลหรือเรียกว่าอีกชื่อหนึ่งว่าการทำ Knowledge Discovery in Database (KDD) เป็นกระบวนการทำงานเพื่อค้นหาความรู้ที่ค้นพบในฐานข้อมูลขนาดใหญ่มาก เป็นกระบวนการสร้างแบบจำลองหรือรูปแบบจากกลุ่มของข้อมูล ทำให้เกิดความเข้าใจในรูปแบบของความเกี่ยวข้องกันของกลุ่มข้อมูลจำนวนมาก และแนวโน้มเพื่อใช้ในการทำนายข้อมูลนั้นๆ การทำเหมืองข้อมูลก็มีวัตถุประสงค์ต่างๆกัน ขึ้นอยู่กับผลลัพธ์ของกระบวนการโดยรวมที่ต้องการ ดังนั้นจึงควรมีการนำเสนอวิธีการที่หลากหลายสำหรับเป้าหมายที่แตกต่างกัน เพื่อให้ได้ผลลัพธ์ที่เหมาะสมตามที่ต้องการ หลังจากนำไปใช้งานแล้ว

กระบวนการทำเหมืองข้อมูลประกอบด้วยขั้นตอนต่างๆ โดยที่แต่ละขั้นตอนสามารถทำซ้ำหรือวนกลับมาทำใหม่ได้ โดยมีขั้นตอนการทำงานดังนี้

1. การกำหนดวัตถุประสงค์(Business Objective Determination)

การกำหนดวัตถุประสงค์จะต้องกำหนดถึงปัญหาและความต้องการ เพราะการทำงานในส่วนนี้จะต้องทำการวิเคราะห์ถึงปัญหาและวิเคราะห์ข้อมูลเบื้องต้นว่า เรามีข้อมูลอะไรและต้องการข้อมูลอะไร รวมไปถึงการวิเคราะห์ไปถึงการกำหนดอัลกอริทึมและฐานข้อมูลที่ใช้ด้วย เพื่อให้สามารถกำหนดแบบจำลองที่เหมาะสมเป็นไปตามวัตถุประสงค์การที่กำหนด

2. การเลือกข้อมูล (Select Target Data)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเลือกข้อมูลคือการระบุแหล่งของข้อมูลที่มี และดึงข้อมูลออกมาใช้สำหรับขั้นตอนการวิเคราะห์ข้อมูลเบื้องต้นคือขั้นตอนการเตรียมข้อมูลสำหรับการทำเหมืองข้อมูล การเลือกข้อมูลจะแตกต่างไปตามวัตถุประสงค์ของแต่ละธุรกิจหรือจะแตกต่างไปตามแต่ละอัลกอริทึมที่เหมาะสมสำหรับการทำเหมืองข้อมูล ตามที่ได้กำหนดไว้ตั้งแต่ต้น และการเลือกข้อมูลเพื่อเตรียมตัวสำหรับทำเหมืองข้อมูลนั้นจะต้องทำความเข้าใจถึงความหมายของตัวแปรแต่ละตัว ประกอบด้วยอะไรบ้าง ต้องมีการอธิบายอย่างชัดเจนเกี่ยวกับชนิดของข้อมูล ค่าที่เป็นไปได้ แหล่งกำเนิดของข้อมูล และรูปแบบของข้อมูล ประเภทของข้อมูลแบ่งได้ 2 ประเภทคือ

- Categorical แบ่งได้เป็น 2 ประเภทคือ Nominal Variable คือ ตัวแปรที่ลำดับไม่มีความสำคัญ เช่น สถานการแต่งงาน และ Ordinal Variable คือ ตัวแปรที่ลำดับมีความสำคัญ เช่น ลูกค้า (ดี ปานกลาง ไม่ดี)
- Quantitative มีการวัดความแตกต่างระหว่างค่าที่เป็นไปได้ แบ่งได้เป็น 2 ประเภทคือ Continuous เก็บค่าตัวเลขที่เป็นจริง เช่น รายได้ และ Discrete เก็บค่าตัวเลขที่เป็นจำนวนเต็ม เช่น จำนวนพนักงานในบริษัท

ตัวแปรของข้อมูลที่ถูกเลือกมาสำหรับทำเหมืองข้อมูลเรียกว่า "Active Variable" เพราะว่ามันถูกใช้สำหรับสร้างความแตกต่างของกลุ่มย่อยๆต่างๆ และสามารถถูกนำมาทำนายผลได้

3. การกลั่นกรองข้อมูล(Data Preprocessing)

การกลั่นกรองข้อมูลเพื่อทำให้มั่นใจว่าข้อมูลที่ถูกเลือกมาสำหรับการทำเหมืองข้อมูลนั้นมีความเหมาะสมและเป็นข้อมูลที่มีคุณภาพ สามารถนำไปใช้งานในขั้นตอนต่อๆไปได้

ในการกลั่นกรองข้อมูลจะมีปัญหาบ่อยๆที่มักพบได้แก่

Noisy Data คือ ตัวแปรหนึ่งหรือมากกว่า ซึ่งมีค่าเกินกว่าที่เราคาดเอาไว้ อาจเป็นข้อมูลที่ไม่สมบูรณ์ สาเหตุเกิดจากความผิดพลาดของมนุษย์ เช่น พนักงานอายุ 250 ปี

Missing Value คือค่าที่ไม่ได้ถูกเลือกมาจากขั้นตอนการเลือกข้อมูล เป็นค่าว่างหรือเป็นค่าที่ไม่สมบูรณ์ที่ถูกลบออกไป ระหว่างการทำ Noise Detection ทำให้ข้อมูลบางส่วนอาจหายไป ดังนั้นมีวิธีที่ช่วยแก้ปัญหาดังกล่าว คือ Data Cleaning เป็นวิธีการทำให้ข้อมูลมีความสมบูรณ์และสอดคล้องกัน ด้วยวิธีการเพิ่มค่าที่ขาดหายไป ป้อนข้อมูลเข้าสู่ระบบ การรับส่งข้อมูลความไม่สอดคล้องในการตั้งชื่อ แล้วทำการปรับปรุงให้ค่าของข้อมูลมีความสอดคล้องกัน โดยอาจจะตัดฟิลด์ที่ข้อมูลขาดหายมากๆทิ้งไป หรืออาจแทนค่าในช่องที่ขาดหายไปด้วย "Unknown"

4. การแปลงข้อมูล(Data Transform)

การแปลงข้อมูลเป็นการนำข้อมูลที่ผ่านการกลั่นกรองแล้วมาทำการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลที่พร้อมจะถูกวิเคราะห์ คือรูปแบบข้อมูลไม่มีความขัดแย้ง มีความเหมาะสมสำหรับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อัลกอริทึมที่จะใช้ในขั้นตอนการทำเหมืองข้อมูล เช่นการแปลงข้อมูลตัวเลขให้มีลักษณะเป็นช่วงๆ ใช้สำหรับอัลกอริทึมใน Decision Tree

5. การทำเหมืองข้อมูล(Data Mining)

การทำเหมืองข้อมูลเป็นการประมวลผลตามอัลกอริทึมที่ได้กำหนดไว้มีความสัมพันธ์กับการวิเคราะห์ข้อมูลในขั้นตอนที่ผ่าน ซึ่งในการทำเหมืองข้อมูลจะใช้อัลกอริทึมหลายๆแบบด้วยกัน โดยผู้ใช้จะต้องเลือกอัลกอริทึมให้เหมาะสมกับลักษณะของปัญหา ผลลัพธ์ที่ได้จะอยู่ในรูปแบบของแบบจำลองหรือกฎความสัมพันธ์ใช้ในการวิเคราะห์หรือพยากรณ์ข้อมูล

6. การวิเคราะห์ผลลัพธ์(Interpretation and Evaluation)

การวิเคราะห์ผลลัพธ์เป็นการนำผลลัพธ์ที่ได้จากขั้นตอนการทำเหมืองข้อมูล อาจอยู่ในรูปแบบของแบบจำลองหรือกฎความสัมพันธ์ของข้อมูล มาทำการตรวจสอบ ตีความหมาย และตรวจสอบความถูกต้องของผลลัพธ์ที่ได้มา จากนั้นทำการวิเคราะห์ว่าผลลัพธ์ที่ได้นั้นมีประโยชน์บรรลุวัตถุประสงค์ตามที่กำหนดไว้หรือไม่ และประเมินความถูกต้องของผลลัพธ์ที่ได้จากการทำเหมืองข้อมูล ถ้าผลการวิเคราะห์ไม่เป็นที่พอใจ หรือไม่ผ่านการประเมินตามที่กำหนดไว้จะต้องกลับไปทำกระบวนการทำเหมืองข้อมูลใหม่จนกว่าผลการวิเคราะห์จะผ่านเกณฑ์การประเมินเป็นที่น่าพอใจ

7. การประยุกต์ใช้งาน(Taking Action of Knowledge)

การประยุกต์ใช้งาน เป็นการประยุกต์นำผลลัพธ์ที่ได้ อาจอยู่ในรูปแบบของแบบจำลองต้นแบบที่สร้างได้จากกระบวนการทำเหมืองข้อมูลมาประยุกต์ใช้ให้เกิดประโยชน์ต่อองค์กร เช่น ช่วยในการตัดสินใจ

จากกระบวนการในการทำเหมืองข้อมูลทั้งหมดนี้ ขั้นตอนที่สำคัญมากในการทำเหมืองข้อมูลคือ ขั้นตอนการเตรียมข้อมูลจะประกอบด้วยการเลือกข้อมูล การกลั่นกรองข้อมูล การแปลงข้อมูล ซึ่งขั้นตอนนี้ใช้เวลาในการทำงานนานที่สุด เพราะว่าถ้าหากเรานำข้อมูลที่ไม่มีคุณภาพมาทำเหมืองข้อมูล
ผลลัพธ์ก็จะ ไม่มีคุณภาพและไม่สามารถนำไปใช้งานได้(Wasilewska)

2.3 ทำเหมืองข้อมูล

ในการทำเหมืองข้อมูลประกอบด้วยเทคนิคย่อยหลายๆเทคนิคได้แก่ เทคนิคการจำแนกประเภทข้อมูล(Data Classification) เทคนิคการแบ่งกลุ่มข้อมูล (Data Segmentation) เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule Discovery) และ เทคนิค Time-Series Analysis เป็นต้น ซึ่งแต่ละเทคนิคมีความเหมาะสมสำหรับงานต่างประเภทกันไป ขึ้นอยู่กับว่าจะใช้เพื่อวัตถุประสงค์อะไร หรือช่วยแก้ไขปัญหาใดบ้างที่เกิดขึ้น เทคนิคการจำแนกประเภทข้อมูล และเทคนิคการค้นหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

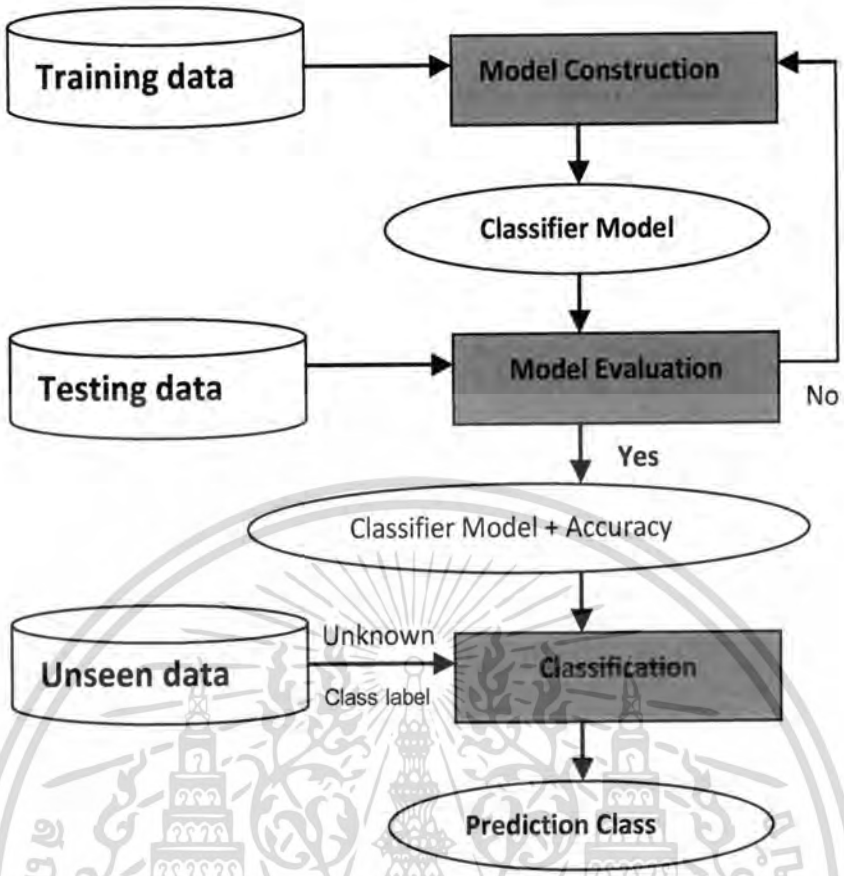
อะไร หรือช่วยแก้ไขปัญหาใดบ้างที่เกิดขึ้น เทคนิคการจำแนกประเภทข้อมูล และเทคนิคการค้นหากฎความสัมพันธ์ถือว่าเป็นเทคนิคที่สำคัญในสาขาวิขาการทำเหมืองข้อมูล และได้มีการนำเทคนิคการทำเหมืองข้อมูลดังกล่าวไปประยุกต์ใช้งานได้จริงในหลายๆสาขาด้วยกัน ซึ่งโครงงานการพัฒนาระบบนี้ได้เลือกใช้เทคนิคทั้งสองมาเข้ามาใช้ทำงานร่วมกันในการพัฒนาระบบงาน

2.3.1 เทคนิคการจำแนกประเภทข้อมูล (Data Classification)

การจำแนกประเภทข้อมูลเป็นหนึ่งในเทคนิคของการทำเหมืองข้อมูลเป็นเทคนิคที่ทำหน้าที่สืบค้นความรู้เพื่อสรุปหาแบบจำลองหรือโมเดลของฐานข้อมูลนั้น(Quinlan, 1993) เพื่อใช้ในการทำนายข้อมูลใหม่ (Unseen Data) โดยการจำแนกประเภทข้อมูลเป็นเทคนิคแบบ Supervise Learning คือ การที่จะสามารถสร้างแบบจำลองออกมาได้นั้นจะต้องทำการสอนระบบก่อน ซึ่งเราจำเป็นต้องทราบจำนวนคลาสปลายทาง(Class Label) และ จำนวนแอตทริบิวท์(Attribute) ที่แน่นอน และจะต้องทำการแบ่งข้อมูลออกเป็น 2 ส่วน คือ ส่วนที่สำหรับใช้สอนระบบ(Training) และส่วนที่ 2 คือ ส่วนที่ใช้สำหรับทดสอบความแม่นยำของแบบจำลองที่ถูกสร้างออกมา(Testing Data) โดยปกติจะแบ่งสัดส่วนข้อมูลระหว่างข้อมูลสำหรับใช้สอนระบบและข้อมูลสำหรับทดสอบความแม่นยำออกเป็น 80 ต่อ 20 โดยมีขั้นตอนในการสร้างแบบจำลองดังนี้ เริ่มจากนำข้อมูลสอนระบบเข้าสู่กระบวนการสร้างแบบจำลองการจำแนกกลุ่มข้อมูล(Model Constructor) เพื่อให้ได้แบบจำลองออกมา

หลังจากได้แบบจำลองกลุ่มข้อมูลออกมาแล้ว วิธีทดสอบว่าแบบจำลองที่ถูกสร้างขึ้นมามีความแม่นยำมากเพียงพอที่จะนำไปใช้ได้หรือไม่ จะใช้ข้อมูลทดสอบระบบเพื่อทดสอบความแม่นยำของแบบจำลองที่ถูกสร้างขึ้นมา(Model Evaluate) ถ้าแบบจำลองไม่ผ่านเกณฑ์ที่กำหนดไว้ ก็จะต้องกลับไปปรับปรุงในกระบวนการสร้างแบบจำลองเสียก่อน แต่ถ้าแบบจำลองมีความแม่นยำผ่านตามเกณฑ์ที่ได้กำหนดไว้ ก็สามารถนำแบบจำลองที่สร้างมานั้นไปประยุกต์ใช้เพื่อทำนายกลุ่มหรือประเภทข้อมูลใหม่ที่ไม่ทราบกลุ่มหรือประเภทข้อมูลต่อไป (Unknown Class Label) โดยภาพรวมขั้นตอนของเทคนิคการจำแนกประเภทข้อมูล สามารถตัวอย่างได้จากรูปที่ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



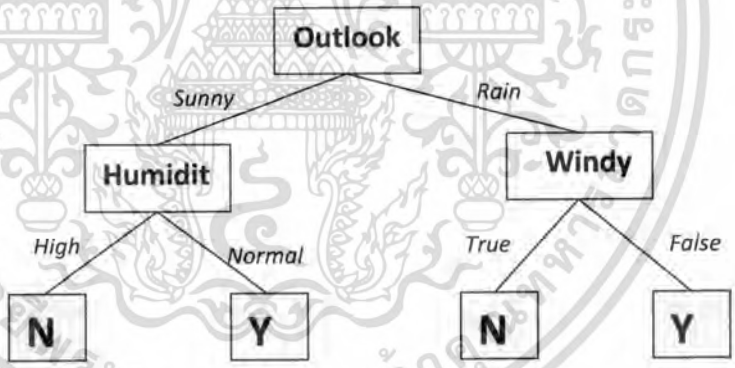
รูปที่ 2.1 ขั้นตอนของเทคนิคการจำแนกประเภทข้อมูล

ตัวอย่างการนำเทคนิคการจำแนกประเภทข้อมูลมาใช้คือ การใช้วิเคราะห์สภาพอากาศเพื่อตัดสินใจว่าในขณะนั้น สมควรที่จะออกไปทำธุระข้างนอกหรือไม่ เริ่มต้นจากการสอนระบบ โดยการนำข้อมูลที่มีอยู่ ซึ่งประกอบด้วย สภาพอากาศ(Outlook) อุณหภูมิ (Temperature) ความชื้น (Humidity) และแรงลม (Windy) พร้อมกับผลการตัดสินใจว่าสมควรออกไปนอกบ้านหรือไม่ (Class) มาสอนให้ระบบเรียนรู้เพื่อการจำแนก ผลลัพธ์ที่ได้จากกระบวนการดังกล่าว ซึ่งแสดงออกมาเป็นรูปแผนผังต้นไม้ เมื่อกรอบสี่เหลี่ยมคือ เงื่อนไขต่างๆของการตัดสินใจ N คือ ไม่สมควร และ Y คือสมควรออกนอกบ้าน จากแผนผังนี้ทำให้สามารถตัดสินใจได้ว่า เมื่อสภาพอากาศ อุณหภูมิ ความชื้นหรือแรงลมในขณะนั้นมีลักษณะหนึ่งๆแล้ว เหมาะสมที่จะออกไปนอกบ้านหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 ตัวอย่างข้อมูลสภาพอากาศ และการตัดสินใจว่าจะออกไปนอกบ้านหรือไม่

Outlook	Temperature	Humidity	Windy	Class
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Rain	Mild	High	False	Y
Rain	Cool	Normal	False	Y
Rain	Cool	Normal	True	N
Sunny	Mild	High	False	N
Sunny	Cool	Normal	False	Y
Rain	Mild	Normal	False	Y
Sunny	Mild	Normal	True	Y
Rain	Mild	High	True	N



รูปที่ 2.2 ตัวอย่างแบบจำลองที่ได้จากเทคนิคการจำแนกประเภทข้อมูล

ในขั้นตอนการสอนระบบเพื่อสร้างแบบจำลองประเภทข้อมูล สามารถสร้างได้จากหลายเทคนิค ซึ่งประกอบด้วย Decision Tree, Bayesian Classification, Neural Network, Genetic Algorithm, Fuzzy Sets และ Association-Based Classification เทคนิค Decision Tree เป็นเทคนิคที่ได้รับความนิยมสูงสุดในการสร้างแบบจำลองการจำแนกประเภทข้อมูล เนื่องจากมีความเร็วใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเรียนรู้เพื่อสร้างแบบจำลองการจำแนกสูงกว่าตัวอื่นๆ โดยที่ให้เปอร์เซ็นต์ความถูกต้องของการจำแนกผลลัพธ์พอๆกัน

2.3.2 เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule Discovery)

การค้นหากฎความสัมพันธ์เป็นอีกหนึ่งในเทคนิคการทำเหมืองข้อมูลที่สำคัญ ซึ่งมีหลักการทำงานคือ การค้นหากฎความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ที่มีอยู่ เพื่อนำกฎที่ได้ไปวิเคราะห์เพื่อช่วยในการตัดสินใจ โดยส่วนใหญ่แล้วจะนำไปใช้ทางด้านธุรกิจ เช่น การวิเคราะห์พฤติกรรมลูกค้าที่ซื้อสินค้าในซุปเปอร์มาเกต โดยดูว่าลูกค้ามักจะซื้ออะไรด้วยกันเพื่อที่จะนำข้อมูลการซื้อสินค้าของลูกค้าเหล่านั้นมาช่วยในการวางแผนทางการตลาด เช่น การวางสินค้าที่มักจะถูกซื้อด้วยกันไว้ใกล้ๆกัน เป็นต้น

โดยอัลกอริทึมหนึ่งที่รู้จักกันดีในการการค้นหากฎความสัมพันธ์คือ อัลกอริทึม Apriori (Agrawal, 1994) ซึ่งมีหลักการคือ จะทำการคำนวณหาความสัมพันธ์ของ itemsets ที่มักจะเกิดขึ้นพร้อมๆกันในฐานข้อมูล โดยความสัมพันธ์ของ itemsets นั้นเรียกว่า กฎความสัมพันธ์ (Association rule) ซึ่งอยู่ในรูปแบบดังต่อไปนี้

{item1, item2} => {item3}[ค่า Support%, ค่า Confidence%]

รูปที่ 2.3 ตัวอย่างกฎความสัมพันธ์ (Agrawal, 1994)

โดยอัลกอริทึม Apriori กฎเกณฑ์ของกฎความสัมพันธ์ทุกกฎที่ได้จะต้องมีการกำหนดค่าสนับสนุนขั้นต่ำ (Minimum support) และค่าค่าความเชื่อมั่นขั้นต่ำ (Minimum confidence) ซึ่งในการกำหนดค่าขั้นต่ำของทั้งสองค่านี้ ขึ้นอยู่กับผู้ใช้งานระบบเป็นผู้กำหนดเอง หรือจะใช้ผู้เชี่ยวชาญ (Expert user) เป็นผู้กำหนดให้ก็ได้ โดยกฎความสัมพันธ์ที่ได้จะต้องมีค่าสนับสนุน (Support) และค่าความเชื่อมั่น (confidence) ไม่น้อยกว่าค่าขั้นต่ำที่ได้กำหนดไว้ข้างต้น

คำศัพท์ของเทคนิคการค้นหากฎความสัมพันธ์ ดังนี้

- ค่าสนับสนุนขั้นต่ำ(Minimum Support) คือ ค่าสนับสนุนต่ำสุดที่ทำให้เกิดกฎเกณฑ์
- ค่าความเชื่อมั่นขั้นต่ำ(Minimum Support) คือค่าความเชื่อมั่นต่ำสุดที่ทำให้เกิดกฎเกณฑ์
- ค่าสนับสนุน(support) คือ เปอร์เซ็นต์ของจำนวน itemsetsทั้งหมดที่เกิดขึ้นในฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

• ค่าความเชื่อมั่น(Confidence) คือ เปอร์เซ็นต์ของจำนวน itemsets ทั้งหมดที่เกิดขึ้นในฐานข้อมูลต่อ จำนวน itemsets ที่เกิดขึ้นทางซ้ายมือ

ตัวอย่างของการใช้เทคนิคการค้นหากฎความสัมพันธ์(Association Rule Discovery) คือ การหากฎความสัมพันธ์ของข้อมูลการซื้อสินค้า ตารางที่ 2.2 แสดงตัวอย่างการซื้อสินค้า เมื่อ Trans-ID คือ รายการซื้อสินค้าแต่ละครั้ง และตัวอักษร A B C D และ E แทนชื่อสินค้าแต่ละชนิด ตารางที่ 2.3 แสดงกฎความสัมพันธ์ที่หาได้จากข้อมูลดังกล่าวตามเทคนิคการค้นหากฎความสัมพันธ์

ตารางที่ 2.2 ตัวอย่างข้อมูลรายการซื้อสินค้า

Trans-ID	Items
1	ACD
2	BCE
3	ABCE
4	BE
5	ABCE

ตารางที่ 2.3 ตัวอย่างกฎความสัมพันธ์ที่หาได้จากเทคนิคการค้นหากฎความสัมพันธ์เมื่อกำหนดค่าสนับสนุนขั้นต่ำ (Support) = 60% และ ค่าความเชื่อมั่นขั้นต่ำ (Confidence) = 75%

Association Rule	Support (%)	Confidence (%)
{BC} -> {E}	60	100
{BE}-> {C}	60	75
{CE}->{B}	60	100
{B}->{CE}	60	75
{C}->{BE}	60	75
{B}->{BC}	60	75

จากตัวอย่างกฎความสัมพันธ์ข้างบน กฎที่ 1 {BC} -> {E} Support 60% Confidence 100% มีความหมายว่า จากจำนวนรายการสินค้าทั้งหมด เมื่อมีการซื้อสินค้า B และ C จะมีการซื้อสินค้า E ด้วยเสมอ โดยปริมาณรายการขายสินค้าที่มีการขาย B C และ E พร้อมกันมีจำนวน 60 เปอร์เซ็นต์จากจำนวนรายการขายสินค้าทั้งหมด และมีความเชื่อมั่น 100 เปอร์เซ็นต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 การจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative classification)

การจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์เป็นเทคนิคที่เกิดจากการรวมกันระหว่าง 2 เทคนิค (Liu, Bing and Hsu,Wynne and Ma,Yiming, 1998) ด้วยกันดังที่ได้กล่าวไว้ตอนต้นแล้ว คือ เทคนิคการจำแนกประเภทข้อมูล (Data Classification) และ เทคนิคการค้นหากฎความสัมพันธ์ (Associative Rule Discovery) โดยเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative Classification) ได้แบ่งการทำงานออกเป็น 2 ส่วนหลักๆ คือ ส่วนที่ใช้ในการสร้างกฎความสัมพันธ์ (Rule Generator Phase) และส่วนที่นำกฎความสัมพันธ์ไปสร้างแบบจำลองเพื่อใช้ในการทำนายข้อมูล (Classification Builder Phase)

ส่วนของการสร้างกฎความสัมพันธ์(Rule generator) จะใช้หลักการเดียวกับเทคนิคการค้นหากฎความสัมพันธ์(Association rule technique)เกือบทั้งหมด คือ หลักการของอัลกอริทึม Apriori แต่สิ่งที่แตกต่างกันคือ กฎที่ถูกสร้างจากกระบวนการสร้างกฎความสัมพันธ์จะต้องเป็นกฎเฉพาะ ที่เรียกว่า กฎความสัมพันธ์แบบมีคลาส หรือ class associative rules(CARs) คือ กฎความสัมพันธ์ที่สับเซตของกฎทางขวามือจะต้องเป็นแอตทริบิวท์ Class เท่านั้น เช่น {A, B, C → **Class**} โดยอัลกอริทึมการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ที่มีอยู่จะถูกดัดแปลงเพื่อค้นหา CARs ทั้งหมดที่ผ่านค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำ

การทำงานส่วนที่สองได้แก่ส่วนของการสร้างแบบจำลองเพื่อการทำนายข้อมูล(Classifier builder) หลังจากได้เซตของกฎทั้งหมดแล้ว (CARs) ก็จะนำกฎความสัมพันธ์ที่ได้จากส่วนของการสร้างกฎนั้นมาใช้เพื่อสร้างแบบจำลองในการทำนายข้อมูลต่อไป โดยอาศัยหลักการเรียงกฎ (Sort CARs) และสร้างแบบแบบจำลองเพื่อใช้ในการทำนายข้อมูลต่อไป โดยในภาพรวมขั้นตอนทั้งหมดของเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ สามารถดูได้จากรูปที่ 2.3

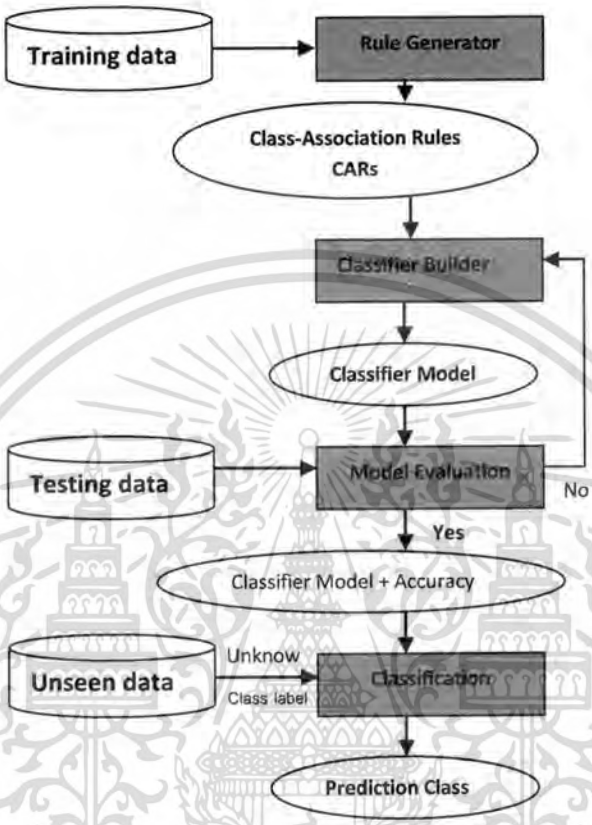
Class Association Rules (CARs) (Liu, Bing and Hsu,Wynne and Ma,Yiming 1998) คือ กฎความสัมพันธ์ที่ได้จากกระบวนการค้นหากฎความสัมพันธ์ แต่เป็นกฎที่บ่งบอกความสัมพันธ์ของข้อมูลที่มีต่อประเภทข้อมูลนั้นๆ (Class) กฎที่ได้จะมีลักษณะดังต่อไปนี้

{item1, item2} → Class ค่าสนับสนุน%, ค่าความเชื่อมั่น %

รูปที่ 2.4 ตัวอย่างรูปแบบกฎความสัมพันธ์แบบมีคลาส (CARs)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปแบบกฎความสัมพันธ์ข้างบน มีความหมายว่า เมื่อมี item 1 และ item 2 อยู่ในชุดข้อมูลพร้อมกัน ชุดข้อมูลนั้นน่าจะจัดอยู่ในประเภทข้อมูล Class โดยมีปริมาณข้อมูลที่สอดคล้องตามกฎความสัมพันธ์ของ ค่าเปอร์เซ็นต์ค่าสนับสนุนและค่าเปอร์เซ็นต์ของค่าความเชื่อมั่น



รูปที่ 2.5 ขั้นตอนของเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์

2.5 CBA อัลกอริทึม (Classification Based on Associations)

CBA อัลกอริทึม (Liu et al., 1998) เป็นอัลกอริทึมต้นแบบของเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative classification) ซึ่งอัลกอริทึม CBA ประกอบด้วย 2 ส่วนคือ ส่วนของการสร้างกฎความสัมพันธ์ (Rule generator) เรียกว่า CBA-RG ดังที่ได้กล่าวไว้ข้างต้นแล้วว่า อัลกอริทึมส่วนแรกนี้จะอ้างอิงกับอัลกอริทึม Apriori และในส่วนของการสร้างแบบจำลองในการทำนาย (Classifier builder) เรียกว่า CBA-CB

ก่อนที่จะทำความเข้าใจอัลกอริทึมในส่วนของ CBA-RG จะต้องทราบเกี่ยวกับแนวคิดพื้นฐานที่ใช้ในอัลกอริทึมในส่วนของ CBA-RG โดยจุดมุ่งหมายในส่วนของ CBA-RG คือ การ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค้นหา ruleitem ทั้งหมดที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำ โดยที่ ruleitem จะอยู่ในรูปแบบของ

<condset, y>

รูปที่ 2.6 ตัวอย่างรูปแบบของ ruleitem

เมื่อ condset เป็นเซตของไอเท็ม และ y คือ คลาสปลายทาง (class label)
โดยที่ ค่า support count of the condset (condsupCound) คือ จำนวนครั้งของทรานแซกชันที่เกิด condset ขึ้นในฐานข้อมูล และค่า support count of the ruleitem (rulesupCount) คือ จำนวนครั้งของทรานแซกชันที่เกิด condset และคลาสปลายทางเป็น y พร้อมกัน แต่ละ ruleitem เขียนแทนด้วยกฎความสัมพันธ์ดังนี้

condset → y, ค่าสนับสนุน%, ค่าความเชื่อมั่น%

รูปที่ 2.7 ตัวอย่างรูปแบบ ruleitem

โดยที่ค่าสนับสนุน(Support) หาได้จากสมการดังนี้

(rulesupCount / |D|) * 100% (2.1)

โดยที่ |D| คือ ขนาดของฐานข้อมูลหรือจำนวนแถวของข้อมูล (Transaction)
และ ค่าความเชื่อมั่น(Confidence) หาได้จากสมการดังนี้

(rulesupCound / condsupCound)* 100% (2.2)

กฎความสัมพันธ์ที่ผ่านค่าสนับสนุนขั้นต่ำ(Minimum support) เรียกว่า Frequent ruleitem ส่วนกฎความสัมพันธ์ที่เหลือ เรียกว่า Infrequent ruleitem

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับกรณีที่ทุกๆ ruleitem ที่มี condset เหมือนกัน จะทำการเลือกกฎความสัมพันธ์ที่มีค่าความเชื่อมั่นสูงที่สุด(highest confidence) เรียกว่า Possible rule(PR) เป็นตัวแทนกลุ่มของ ruleitem ที่เหลือ เช่น

เมื่อ ruleitem1 ประกอบด้วย แอตทริบิวท์ A มีค่า item เป็น 1 และแอตทริบิวท์ B มีค่า item เป็น 1 และ ruleitem2 ประกอบด้วยแอตทริบิวท์ A มีค่า item เป็น1 และแอตทริบิวท์ B มีค่า item เป็น 1 ซึ่ง ruleitem ทั้งคู่มีค่า condset เหมือนกัน แต่มีคลาสปลายทางต่างกันคือ มีคลาสปลายทางเป็น 1 และ 2

ตัวอย่าง ruleitem ที่มี condset เหมือนกัน ดังนี้

1. <{(A,1),(B,1)},(class: 1)>.
2. <{(A,1),(B,1)},(class: 2)>.

กำหนดให้

ค่า condsupCount เท่ากับ 3

ค่า rulesupCount ของ ruleitem 1 เท่ากับ 2

ค่า rulesupCount ของ ruleitem 2 เท่ากับ 1

ขนาดของฐานข้อมูล(D) เท่ากับ 10

สามารถคำนวณค่าความเชื่อมั่นได้จากสมการดังนี้

$$(\text{rulesupCount}/\text{condsupCount}) * 100\% \qquad (2.3)$$

สำหรับ ruleitem1 จะได้ค่าความเชื่อมั่นเท่ากับ (2/3) *100% = 66.7%

สำหรับ ruleitem2 จะได้ค่าความเชื่อมั่นเท่ากับ (1/3) * 100% = 33.3%

จากผลลัพธ์ที่ได้แสดงให้เห็นว่า ruleitem 1 มีค่าความเชื่อมั่นสูงกว่า ruleitem 2

ดังนั้น ruleitem1คือ (A,1),(B,1) $\rightarrow$(class,1) [support = 20%, confidence = 66.7%] ถูกเลือกไปสร้างกฎความสัมพันธ์ต่อไป โดยที่กฎความสัมพันธ์นั้นจะต้องมีค่าความเชื่อมั่นมากกว่าค่าความเชื่อมั่นขั้นต่ำ(Minimum confidence) เราจะบอกว่ากฎเหล่านั้นมีความถูกต้อง(Accurate)

ดังนั้นเซตที่เรียกว่า Class-Association Rule (CARs) จะประกอบด้วยกฎที่เรียกว่า Frequent และ Accurate

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.1 อัลกอริทึม CBA-RG

โดยอัลกอริทึม CBA-RG จะมีขั้นตอนการทำงาน 3 ขั้นตอนหลักๆ คือ

1. ทำการค้นหา Frequence ruleitem ทั้งหมด (Frequent ruleitem คือ ruleitems ที่มีค่าสนับสนุนสูงกว่าค่าสนับสนุนขั้นต่ำ) โดยขั้นตอนนี้จะเหมือนกับ Apriori ซึ่งผลลัพธ์ที่ได้นั้นจะเลือกเฉพาะกฎที่เป็น Class Association Rules (CARs) อยู่ในรูปแบบของ

condset → y, ค่าสนับสนุน%, ค่าความเชื่อมั่น%

รูปที่ 2.8 ตัวอย่างรูปแบบกฎความสัมพันธ์แบบ CARs

2. สร้างกฎความสัมพันธ์ประเภท CARs จาก Frequent ruleitems และสำหรับทุกๆ ruleitems ที่มี condset เหมือนกัน จะทำการเลือกเพียงกฎเดียว โดยกฎที่มีค่าความเชื่อมั่นสูงที่สุดจะถูกเลือก

3. ทำการ Prune ruleitems ที่มีค่าความเชื่อมั่นน้อยกว่าค่าความเชื่อมั่นขั้นต่ำ (Minimum confidence)

ตัวแปรที่ต้องนำมาพิจารณา ดังนี้

k-ruleitem คือ ruleitem ที่มีสมาชิก(condset)จำนวน k ตัว

F_k คือ เซตของ frequent k-ruleitems ซึ่งทุกเซตของ ruleitems จะต้องมีค่าสนับสนุนสูงกว่าค่าสนับสนุนขั้นต่ำ แต่ละสมาชิกของเซตอยู่ในรูปแบบของ

<(condset, condsupcount), (y, rulesupcount)>.

รูปที่ 2.9 ตัวอย่างรูปแบบ Frequent ruleitem

และ C_k คือ เซตของ candidate k-ruleitems เป็นเซตที่ถูกเลือกมาจาก F_k

อธิบายหลักการทำงานของอัลกอริทึม CBA-RG

ในการหา Frequent ruleitem จะเริ่มจากการสแกนฐานข้อมูลแล้วนับ item และ class เพื่อกำหนดค่า Frequent ruleitem ซึ่งทุกเซตของ ruleitems จะต้องมีค่าสนับสนุนสูงกว่าค่าสนับสนุนขั้นต่ำ หลังจากทำการ prune ค่าที่ได้จะได้กฎความสัมพันธ์ออกมาเป็น CAR_k

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นเข้าสู่ลูป มีขั้นตอนการทำงานหลักๆ ด้วยกัน 4 ขั้นตอน คือ

1.จะใช้ Frequent ruleitem ก่อนหน้านี้ (F_{k-1}) เพื่อหาค่า candidate ruleitem (C_k) โดยใช้ฟังก์ชัน CandidateGen จะได้ candidate ruleitem ออกมาเป็นในรูปแบบของ <condset, y>

2. ทำการสแกนฐานข้อมูลและเข้าสู่ฟังก์ชัน ruleSubset โดยเอาตัว candidate ruleitem (C_k) และ จำนวนทรานแซกชั่นของ condset(d) เพื่อไปนับค่าของ ruleitem ใน C_k โดยพิจารณาค่าสนับสนุน(Support)โดยค่า d จะได้ผลลัพธ์เป็น C_d จากนั้นทำการนับค่า condsupCount และ rulesupCount ของ C_d

3. จะทำการเลือก Candidate ruleitem ที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำไปเป็นตัวตั้งในการหา Frequent ruleitem ต่อไป และสร้างกฎความสัมพันธ์ประเภท CAR_k โดยใช้ฟังก์ชัน genRules

4. ทำการ prune กับเซตของกฎความสัมพันธ์ที่ได้ (CAR_k) โดยทำการ prune ruleitem ที่มีค่าความเชื่อมั่นน้อยกว่าค่าความเชื่อมั่นขั้นต่ำ โดยจะทำการคัดกฎความสัมพันธ์ที่มีค่าความเชื่อมั่นน้อยกว่าค่าความเชื่อมั่นขั้นต่ำทิ้งไป

ในส่วนสุดท้าย จะได้กฎความสัมพันธ์ที่เหลือทั้งหมดซึ่งถูกนำไปใช้สร้างแบบจำลองในการทำนายต่อไป ผลลัพธ์จากการ prune เซตของกฎความสัมพันธ์ประเภท CARs เรียบร้อยแล้ว คือ prCARs สามารถดูตัวอย่างขั้นตอนการทำงานของ CBA-RG อัลกอริทึมได้จากรูปที่ 2.8

```
1   F_1 = {large 1-ruleitems};
2   CAR_1 = genRules(F_1);
3   prCAR_1 = pruneRules(CAR_1);
4   for (k = 2; F_{k-1} ≠ ∅; k++) do
5       C_k = candidateGen(F_{k-1});
6       for each data case d ∈ D do
7           C_d = ruleSubset(C_k, d);
8           for each candidate c ∈ C_d do
9               c.condsupCount++;
10              if d.class = c.class then c.rulesupCount++;
11          end
12      end
13      F_k = {c ∈ C_k | c.rulesupCount ≥ minsup};
14      CAR_k = genRules(F_k);
15      prCAR_k = pruneRules(CAR_k);
16  end
17  CARs = ∪_k CAR_k;
18  prCARs = ∪_k prCAR_k;
```

รูปที่ 2.10 อัลกอริทึม CBA-RG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
06388

2.5.2 การสร้างแบบจำลองเพื่อการทำนาย (Building a Classifier)

หลังจากได้เซตของกฎความสัมพันธ์ (CARs) ทั้งหมดแล้ว ก็จะถึงส่วนต่อไปนั่นคือ การสร้างแบบจำลองเพื่อการทำนาย โดยใช้อัลกอริทึม CBA-CB โดยในส่วนนี้จะกล่าวถึงหลักการในการเรียงกฎ และขั้นตอนในการสร้างแบบจำลองเพื่อใช้ในการทำนายข้อมูล

หลักการเบื้องต้นที่ใช้ในการเรียงกฎ มีดังต่อไปนี้

กำหนดให้ 2 rules คือ r_i และ r_j เราสามารถกำหนดว่า $r_i \succ r_j$ (r_i มีศักย์สูงกว่า r_j) ได้ก็ต่อเมื่อ

- ค่าความเชื่อมั่น (confidence) ของ r_i มากกว่า r_j หรือ
- ค่าความเชื่อมั่น (confidence) เท่ากัน แต่ค่าสนับสนุน (support) r_i มากกว่า r_j หรือ
- ทั้งค่าความเชื่อมั่น(confidence)และค่าสนับสนุน(support)เท่ากันแต่ r_i ถูกสร้างมาก่อน r_j

ให้ R เป็นเซตของกฎความสัมพันธ์ที่ได้จากขั้นตอนในการสร้างกฎความสัมพันธ์ (CARs หรือ pCARs) และ D เป็นข้อมูลที่เราทราบกลุ่มอยู่แล้ว (Training data) แนวคิด คือ เลือกเซตของกฎที่มีศักย์สูงสุดจาก R ซึ่งอยู่ในข้อมูลที่เราทราบกลุ่มแล้ว (D)

Classifier ที่ได้อยู่ในรูปแบบของ

<r_1, r_2,..... default_class>,

รูปที่ 2.11 ตัวอย่างรูปแบบ Classifier

โดยที่ $r_i \in R$, $r_a \succ r_b$ ถ้า b > a แล้ว default_class เป็น classที่กำหนดให้เป็น default class

ขั้นตอนในการสร้างแบบจำลองเพื่อการทำนาย(Building a Classifier) มี 3 ขั้นตอนหลักๆ คือ

1. เรียงลำดับเซตของกฎความสัมพันธ์ (CARs) ที่เป็นผลลัพธ์จากอัลกอริทึม CBA-RG ทั้งหมด โดยใช้หลักการที่ใช้ในการเรียงกฎ จากการพิจารณาตามศักย์ของกฎ ที่ได้กล่าวมาแล้ว เพื่อทำให้เรามั่นใจได้ว่ากฎที่ถูกเลือกมาสร้างแบบจำลองนั้น เป็นกฎที่มีศักย์สูงที่สุด

2. เลือกกฎความสัมพันธ์เพื่อสร้างแบบจำลอง (Classifier) โดยจะเลือกจาก R ซึ่งกฎความสัมพันธ์ในเซต R นั้นเป็นกฎความสัมพันธ์ที่เรียงลำดับเรียบร้อยแล้ว (sorted sequence) ต่อจากนั้นทำการวนลูปกฎความสัมพันธ์โดยนำกฎความสัมพันธ์ไปตรวจสอบกับตัวข้อมูล (ข้อมูลที่ใช้ในการตรวจสอบเป็นข้อมูลที่เราทราบกลุ่มของข้อมูลอยู่แล้ว เรียกว่า Training data) ถ้ากฎ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความสัมพันธ์สามารถรองรับ (Satisfy) ตัวข้อมูลทางด้านซ้าย (Condset) และค่าทางด้านขวา (Class) แล้วจะนำกฎนั้นไปเก็บไว้ในเซตของกฎความสัมพันธ์ (Set C) ก็จะได้กฎความสัมพันธ์ที่จำแนกข้อมูลอย่างถูกต้องเพื่อใช้ในการทำนายข้อมูล (Correctly classify) เป็นอย่างน้อยจำนวน 1 กฎความสัมพันธ์ที่เกิดขึ้น จากนั้นเลือก Default class เพื่อใช้ในการทำนายข้อมูล ในกรณีที่ไม่มีกฎความสัมพันธ์ไหนรองรับข้อมูลแล้วและนำเซตของกฎความสัมพันธ์นั้นมาคำนวณหาเปอร์เซ็นต์ข้อผิดพลาด(Compute error)

3. นำเซตของกฎความสัมพันธ์ทุกๆเซต(Set C) มาเปรียบเทียบกัน แล้วเลือกเซตของกฎความสัมพันธ์ที่มีเปอร์เซ็นต์ความผิดพลาดน้อยที่สุด (ส่วนเซตของกฎความสัมพันธ์ที่ไม่ได้รับการเลือกจะถูกคัดทิ้งไปเพราะว่าเซตของกฎความสัมพันธ์เหล่านั้นจะสร้างความผิดพลาดในการทำนายได้) เพื่อนำเซตของกฎความสัมพันธ์ไปใช้เป็นแบบจำลองเพื่อการทำนายข้อมูลที่เราไม่ทราบกลุ่ม (Unseen data) ต่อไป สามารถดูขั้นตอนการทำงานของ CBA-CB อัลกอริทึมได้จากรูปที่ 2.10

```
1   R = sort(R);
2   for each rule r ∈ R in sequence do
3       temp = ∅;
4       for each case d ∈ D do
5           if d satisfies the conditions of r then
6               store d.id in temp and mark r if it correctly
                    classifies d;
7       if r is marked then
8           insert r at the end of C;
9           delete all the cases with the ids in temp from D;
10          selecting a default class for the current C;
11          compute the total number of errors of C;
12      end
13  end
14  Find the first rule p in C with the lowest total number
        of errors and drop all the rules after p in C;
15  Add the default class associated with p to end of C,
        and return C (our classifier).
```

รูปที่ 2.12 อัลกอริทึม CBA-CB

2.5.3 ตัวอย่างการทำงานของอัลกอริทึม CBA

ตัวอย่างการทำงานของอัลกอริทึม CBA ตั้งแต่การค้นหากฎความสัมพันธ์ (CBA-RG) และการสร้างแบบจำลองในการทำนายข้อมูล (CBA-CB) โดยในตารางที่ 2.4 คือ ตัวอย่างชุดข้อมูลของคนไข้ ซึ่งประกอบด้วย 6 คอลัมน์ โดยมีคลาสปลายทางเป็น Status หรือ สถานะของคนไข้

โดยที่กำหนดค่าสนับสนุนขั้นต่ำ(Minimum Support) เท่ากับ 30% และค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) เท่ากับ 60%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.4 ตัวอย่างข้อมูลคนไข้

TID	Sex	Cholesterol (250)	Blood Sugar	Vassel color Number	Status
1	Male	High	False	1	Sick
2	Male	Low	True	3	Well
3	Female	Low	False	3	Sick
4	Female	High	True	2	Well
5	Female	High	False	3	Well
6	Male	Low	True	1	Sick
7	Female	Low	False	3	Sick
8	Female	High	True	3	Well
9	Male	Low	False	1	Few
11	Female	Low	True	1	Well

จากตัวอย่างข้อมูลคนไข้ในตารางที่ 2.4 ก็จะถูกนำเข้าสู่กระบวนการในส่วนของการสร้างกฎความสัมพันธ์ โดยขั้นตอนวิธีการสร้างกฎความสัมพันธ์สามารถดูได้จาก ตารางที่ 2.5 ถึง 2.6 และกฎความสัมพันธ์ที่ได้ทั้งหมดที่ผ่านค่าสนับสนุนขั้นต่ำสามารถดูได้จาก ตารางที่ 2.7 ส่วนตารางที่ 2.8 จะแสดงกฎความสัมพันธ์ที่ผ่านค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำ ซึ่งหลังจากนั้นอัลกอริทึม CBA-RG จะทำการจัดเรียงกฎความสัมพันธ์แบบมีคลาสใหม่โดยเรียงลำดับตามค่าความเชื่อมั่น โดยดูจากตารางที่ 2.9

ตารางที่2. 5 อัลกอริทึม CBA-RG ทำการค้นหา Frequent ruleitems รอบแรก

1-ruleitem C1	1-ruleitem C1	Count	%	Large 1-ruleitem L1	Count	%
{Male, Sick}	{Male, Sick}	2	20			
{Male, Well}	{Male, Well}	1	10			
{Female, Sick}	{Female, Sick}	2	20			
{Female, Well}	{Female, Well}	4	40	{Female, Well}	4	40
{Male, Few}	{Male, Few}	1	10			
{High, Sick}	{High, Sick}	1	10			
{Low, Well}	{Low, Well}	2	20			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2. 5 (ต่อ)

{Low, Sick}	{Low, Sick}	3	30	{Low, Sick}	3	30
{High, Well}	{High, Well}	3	30	{High, Well}	3	30
{Low, Few}	{Low, Few}	1	10			
{False, Sick}	{False, Sick}	3	30	{False, Sick}	3	30
{True, Well}	{True, Well}	4	40	{True, Well}	4	40
{True, Sick}	{True, Sick}	1	10			
{False, Well}	{False, Well}	1	10			
{False, Few}	{False, Few}	1	10			
{1, Sick}	{1, Sick}	2	2			
{3, Well}	{3, Well}	3	3	{3, Well}	3	30
{3, Sick}	{3, Sick}	2	2			
{2, Well}	{2, Well}	1	1			
{1, Few}	{1, Few}	1	1			
{1, Well}	{1, Well}	1	1			
a)Generate Phase				b1) Count Phase		

ตารางที่ 2.5 แสดงขั้นตอนของอัลกอริทึม CBA-RG ทำการค้นหา ruleitem ที่มีค่าสนับสนุนมากกว่า ค่าสนับสนุนขั้นต่ำ ซึ่งได้กำหนดไว้คือ 30 เปอร์เซ็นต์ โดย ruleitem ที่ผ่านค่าสนับสนุนขั้นต่ำ จะถูกนำไปใช้ในการหา ruleitem ใน ระดับที่สอง ถัดไป

ตารางที่ 2.6 อัลกอริทึม CBA-RG ทำการค้นหา Frequent ruleitems รอบที่สอง

2-ruleitems C2	2-ruleitems C2	Count	%	Large 2-ruleitems L2	Count	%
{Female,High,Well}	{Female,High, Well}	3	30	{Female,High, Well}	3	30
{Female,True, Well}	{Female, True, Well}	3	30	{Female, True, Well}	3	30
{Female, 3, Well}	{Female, 3, Well}	2	20			
{High, True, Well}	{High, True, Well}	2	20			
{High, 3, Well}	{High, 3, Well}	2	20			
{True,3, Well}	{True,3, Well}	2	20			
{Low, False, Sick}	{Low, False, Sick}	2	20			
a)Generate Phase	b1) Count Phase			b2) Select Phase		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 2.6 ซึ่งเป็นขั้นตอนในการหา ruleitems ในระดับที่สองที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำ จะเห็นได้ว่าในระดับนี้จะเป็นระดับสุดท้าย เนื่องจากไม่สามารถสร้าง ruleitem ในระดับที่สามได้เพราะเหลือ ruleitem ที่ผ่านค่าสนับสนุนในระดับที่สอง เพียง 2 ตัวเท่านั้น

ตารางที่ 2.7 CBA-RG สร้าง Class-Association Rule(CARs) ที่ผ่านค่าสนับสนุน

CARs	ค่าสนับสนุน (%)	ค่าความเชื่อมั่น(%)
{Female}=>{ Well}	40	4/6=66.7
{Low}=>{Sick}	30	3/6=50
{High}=>{Well}	30	3/4=75
{False}=>{Sick}	30	3/5=60
{True}=>{Well}	40	4/5=80
{3}=>{ Well}	30	3/5=60
{Female, High}=> Well}	30	3/3=100
{Female, True}=>{Well}	30	3/3=100

จากตารางที่ 2.7 แสดงกฎความสัมพันธ์แบบมีคลาส(CARs) ในทุกระดับของ ruleitem ที่ผ่านค่าสนับสนุนขั้นต่ำ

ตารางที่ 2.8 CBA-RG สร้าง Class-Association Rules (CARs) ที่ผ่านค่าความเชื่อมั่น

CARs	ค่าสนับสนุน (%)	ค่าความเชื่อมั่น(%)
{Female}=>{ Well}	40	4/6=66.7
{Low}=>{Sick}	30	3/6=50
{High}=>{Well}	30	3/4=75
{False}=>{Sick}	30	3/5=60
{True}=>{Well}	40	4/5=80
{3}=>{ Well}	30	3/5=60
{Female, High}=> Well}	30	3/3=100
{Female, True}=>{Well}	30	3/3=100

ตารางที่ 2.8 แสดงกฎความสัมพันธ์แบบมีคลาส (CARs) ในทุกๆระดับของ ruleitems ที่ผ่านค่าสนับสนุนขั้นต่ำ และค่าความเชื่อมั่นขั้นต่ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากตารางที่ 2.8 จะเป็นการจบในส่วนของการสร้างกฎความสัมพันธ์ (CBA-RG) ซึ่งในขั้นตอนถัดไปจะเป็นขั้นตอนในส่วนการสร้างแบบจำลองในการทำนาย (CBA-CB)

ตารางที่ 2.9 CBA-CB เรียงกฎตามค่าความเชื่อมั่น

Rid	CARs	ค่าสนับสนุน (%)	ค่าความเชื่อมั่น (%)
r1	{Female, High}=> Well}	30	3/3=100
r2	{Female, True}=>{Well}	30	3/3=100
r3	{True}=>{Well}	40	4/5=80
r4	{High}=>{Well}	30	3/4=75
r5	{Female}=>{ Well}	40	4/6=66.7
r6	{False}=>{Sick}	30	3/5=60
r7	{3}=>{ Well}	30	3/5=60

ตารางที่ 2.9 แสดงกฎความสัมพันธ์แบบมีคลาส (CARs) ที่ถูกเรียงใหม่ตามค่าความเชื่อมั่นเพื่อจะสามารถนำกฎความสัมพันธ์นี้ไปใช้ในการสร้างแบบจำลองในการทำนายข้อมูล

ตารางที่ 2.10 CBA-CB สร้างแบบจำลองในการทำนาย

แบบจำลองการทำนาย	ความถูกต้อง(%)
R1={r1, r2, r3, r6, Default Class = Sick}	80

จากตารางที่ 2.10 จะเห็นว่าอัลกอริทึม CBA-CB หรือส่วนของการสร้างแบบจำลองเพื่อการทำนายจะมีการสร้างแบบจำลองออกมาเพื่อเป็นแบบจำลองต้นแบบของการทำนายประเภทหรือกลุ่มข้อมูลใหม่ที่ไม่ทราบกลุ่มข้อมูลต่อไป (Unknown Class label)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิเคราะห์และออกแบบระบบ

ขั้นตอนการวิเคราะห์และออกแบบระบบ การพัฒนาโครงงานการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเออัลกอริทึม ใช้เทคโนโลยีการทำเหมืองข้อมูลเข้ามาช่วยในการประมวล ผลและทำการออกแบบระบบโดยใช้ยูเอ็มแอล เพื่ออธิบายขั้นตอนการทำงานและแสดงความสัมพันธ์ของข้อมูลในระบบ โดยสร้างแบบจำลองระบบในลักษณะรูปภาพ เพื่อให้ง่ายต่อการทำความเข้าใจในขั้นตอนการทำงาน การวิเคราะห์และออกแบบระบบมีรายละเอียดดังนี้

3.1 ความต้องการและข้อจำกัดของโปรแกรม

ในการพัฒนาระบบจะต้องมีการกำหนดความต้องการในส่วนของผู้ใช้งานระบบ ความต้องการเชิงฟังก์ชันและความต้องการที่ไม่ใช่เชิงฟังก์ชัน เพื่อสนับสนุนการทำงานของระบบ มีรายละเอียดดังนี้

3.1.1 ความต้องการของผู้ใช้งาน

1. สามารถกำหนดเงื่อนไขหรือกฎเกณฑ์เพื่อสร้างแบบจำลอง

2. สามารถทราบหรือบอกผลลัพธ์ของแบบจำลองต้นแบบที่ได้จากกระบวนการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ได้

3. สามารถบันทึกผลลัพธ์แบบจำลองต้นแบบในรูปแบบฐานข้อมูลได้

4. สามารถเรียกดูแบบจำลองผ่านทางหน้าจอได้

3. ต้องการทราบผลการทำนายหรือจำแนกข้อมูลใหม่ในอนาคตได้

3.1.2 ความต้องการเชิงฟังก์ชัน

1. ระบบสร้างแบบจำลองเพื่อการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเออัลกอริทึมได้

2. ระบบสามารถจัดการข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปใช้ในจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเออัลกอริทึมได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ระบบสามารถค้นหากฎความสัมพันธ์เพื่อใช้ในกระบวนการสร้างแบบจำลองสำหรับจำแนกประเภทข้อมูลใหม่ได้

4. ระบบสามารถแสดงผลลัพธ์แบบจำลองการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเออัลกอริทึมได้

5. ระบบสามารถจัดเรียงลำดับแบบจำลองการจำแนกประเภทข้อมูลตามค่าสนับสนุน หรือค่าความเชื่อมั่นได้

6. สามารถบันทึกแบบจำลองต้นแบบในรูปแบบฐานข้อมูลได้

7. สามารถเรียกดูแบบจำลองต้นแบบได้

8. สามารถทดสอบและประเมินความถูกต้องของแบบจำลองได้

9. ระบบสามารถทำนายหรือจำแนกประเภทข้อมูลใหม่ในอนาคตได้

3.1.3 ความต้องการที่ไม่ใช่เชิงฟังก์ชัน

1. ระบบตอบสนองการทำงานในขั้นตอนต่างๆได้

2. ระบบแสดงสถานการณ์ทำงานหรือผลลัพธ์ในการทำงานในแต่ละขั้นตอนได้

3.2 การวิเคราะห์ระบบโดยใช้ยูสเคสไดอะแกรม และ แอคทิวิตีไดอะแกรม

ภาพรวมกระบวนการทำงานของ ระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ อัลกอริทึม สามารถอธิบายโดยใช้แบบจำลองยูสเคสไดอะแกรม ดังรูปที่ 3.1

3.2.1 รายละเอียดแอ็กเตอร์

จากรูปที่ 4.1 เป็นการใช้ยูสเคสแสดงภาพรวมของระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ อัลกอริทึม โดยมีผู้ที่เกี่ยวข้องกับระบบดังนี้

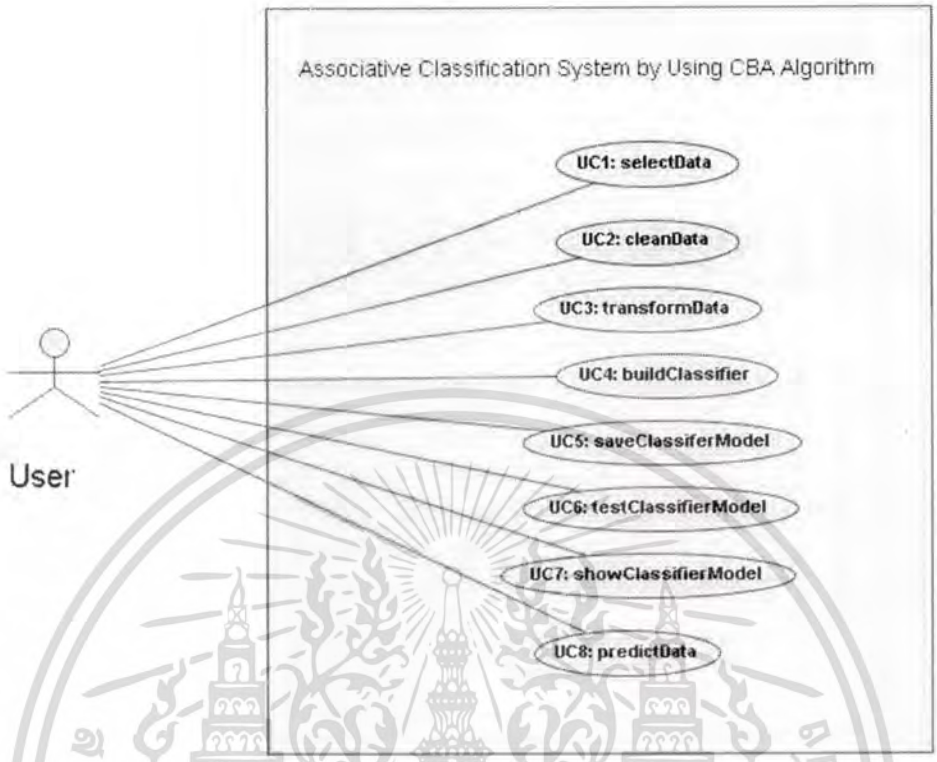
1. ผู้ใช้งาน(User)

เป็นบุคคลที่ทำงานร่วมกับระบบ โดยนำข้อมูลอินพุตเข้าสู่ระบบ เพื่อทำการประมวลผลแล้วได้ผลลัพธ์ออกมา และนำผลลัพธ์ที่ได้ไปใช้ในกระบวนการวิเคราะห์ข้อมูลและช่วยในการตัดสินใจในการทำงาน

2. ระบบฐานข้อมูล(Database System)

เป็นส่วนที่ใช้ในการเก็บข้อมูลต่างๆเพื่อนำไปใช้ในการทำเหมืองข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 ภาพรวมระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ อัลกอริทึม

3.2.2 รายละเอียดยูสเคส

3.2.2.1 UC1: selectData เป็นยูสเคสสำหรับการเลือกข้อมูล อธิบายขั้นตอนการติดต่อกับฐานข้อมูล โดยทำการติดต่อกับ Microsoft SQL Server 2005 และทำการเลือกข้อมูล เพื่อใช้เป็นอินพุตของระบบ

3.2.2.2 UC2: cleanData เป็นยูสเคสสำหรับการคลีนข้อมูล อธิบายขั้นตอนการแก้ไขข้อมูลที่เป็นค่าว่าง (Null Value) เมื่อพบแอตทริบิวท์ที่มีค่าว่าง จะทำการลบแถวของข้อมูลแอตทริบิวท์ที่มีค่าว่าง

3.2.2.3 UC3: transformData เป็นยูสเคสสำหรับขั้นตอนการแปลงข้อมูล อธิบายขั้นตอนการปรับเปลี่ยนรูปแบบข้อมูล เพื่อให้ได้ข้อมูลในรูปแบบที่เหมาะสมสำหรับใช้ในการประมวลผลในขั้นตอนการสร้างแบบจำลองในส่วนของการสร้างกฎความสัมพันธ์ (CBA-RG)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2.4 UC4: buildClassifier เป็นยูสเคสอธิบายขั้นตอนการสร้างแบบจำลองการจำแนกประเภทข้อมูลด้วยอัลกอริทึม CBA โดยข้อมูลที่สามารถนำไปใช้สำหรับการประมวลผลต้องผ่านกระบวนการเลือกข้อมูล คลีนข้อมูล และแปลงข้อมูลแล้ว โดยนำข้อมูลที่ได้มาใช้ในการสร้างแบบจำลอง

3.2.2.5 UC5: saveClassiferModel เป็นยูสเคสอธิบายขั้นตอนการบันทึกผลลัพธ์แบบจำลองการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ที่ได้จากการนำข้อมูลเข้ามาประมวลผล โดยทำการบันทึกข้อมูลลงในรูปแบบฐานข้อมูล

3.2.2.6 UC6: testClassifierModel เป็นยูสเคสอธิบายขั้นตอนการทดสอบความถูกต้องของแบบจำลองระบบ ทำการทดสอบความแม่นยำด้วยข้อมูลทดสอบระบบ (Testing data)

3.2.2.7 UC7: showClassifierModel เป็นยูสเคสอธิบายขั้นตอนการแสดงแบบจำลอง

3.2.2.8 UC8: predictData เป็นยูสเคสอธิบายขั้นตอนการทำนายประเภทกลุ่มข้อมูลให้กับข้อมูลใหม่ที่เข้ามาในระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส UC1: selectData

รายละเอียด

เป็นยูสเคสอธิบายขั้นตอนการติดต่อฐานข้อมูลโดยทำการติดต่อกับ Microsoft SQL Server 2005 และทำการเลือกข้อมูลเพื่อใช้เป็นอินพุตของระบบ

แอคเตอร์

1. ผู้ใช้งาน(User)

เงื่อนไขก่อนเข้ายูสเคส

-

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานระบุชื่อเครื่องเซิร์ฟเวอร์
2. ผู้ใช้งานระบุชื่อฐานข้อมูล
3. ผู้ใช้งานกดปุ่ม Connect
4. ระบบทำการตรวจสอบฐานข้อมูลใน Microsoft SQL Server แล้วพบฐานข้อมูล
5. ระบบแสดงตารางทั้งหมดในฐานข้อมูล
6. ผู้ใช้งานเลือกตารางในฐานข้อมูล
7. ระบบแสดงแอตทริบิวท์ของตารางที่ผู้ใช้งานเลือก
8. ผู้ใช้งานเลือกแอตทริบิวท์ในตาราง
9. ผู้ใช้งานกดปุ่ม Execute
10. แสดงผลลัพธ์การเลือกข้อมูล

ลำดับเหตุการณ์ย่อย

3a) ระบบไม่พบฐานข้อมูล

3a1. ระบบแจ้งข้อความเตือน

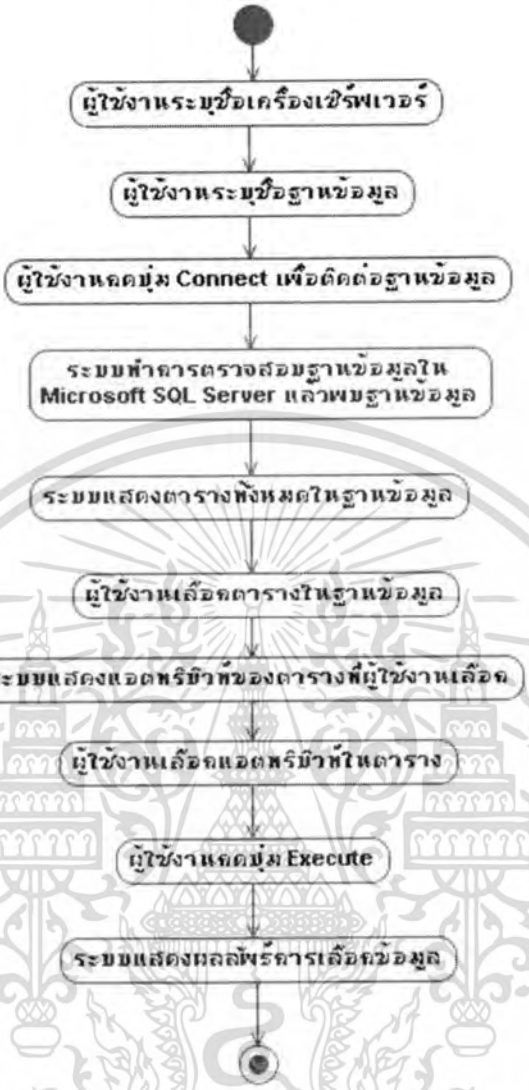
3a2. ผู้ใช้งานระบุชื่อฐานข้อมูลใหม่

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบแสดงการเลือกข้อมูลเรียบร้อยแล้ว

รูปที่ 3.2 คำอธิบายยูสเคส UC1: select Data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 แอคทิวิตีไดอะแกรม selectData

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส UC2: cleanData

รายละเอียด

เป็นยูสเคสอธิบายขั้นตอนการแก้ไขข้อมูลที่เป็นค่าว่าง(Null Value) เพื่อให้ได้ข้อมูลที่เหมาะสมในการนำไปใช้ในการสร้างแบบจำลองเพื่อใช้ในการจำแนกประเภทข้อมูล เมื่อพบแอตทริบิวท์ที่มีค่าว่าง จะทำการลบแถวของข้อมูลแอตทริบิวท์ที่มีค่าว่าง

แอคเตอร์

1. ผู้ใช้งาน(User)

เงื่อนไขก่อนเข้ายูสเคส

1. ผู้ใช้ต้องติดต่อกับฐานข้อมูลและทำการเลือกข้อมูลเรียบร้อยแล้ว

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานกดปุ่ม Auto Cleaning
2. ระบบทำการลบแถวของข้อมูล ที่ข้อมูลในแอตทริบิวท์เป็นค่าว่าง
3. ระบบแจ้งข้อความ Auto Cleaning Complete

ลำดับเหตุการณ์ย่อย

2a) ระบบทำการลบข้อมูลที่ไม่ถูกต้อง : ระบบแสดงข้อความเตือนไม่สามารถคลีนข้อมูลได้

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบทำการคลีนข้อมูลเรียบร้อย

รูปที่ 3.4 คำอธิบายยูสเคส UC2: cleanData

รูปที่ 3.5 แอคทิวิตีไดอะแกรม cleanData

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส UC3: transformData

รายละเอียด

เป็นยูสเคสอธิบายขั้นตอนการปรับเปลี่ยนรูปแบบข้อมูล เพื่อให้ข้อมูลอยู่ในรูปแบบที่เหมาะสมในการประมวลผลเพื่อนำไปใช้ในขั้นตอนการทำเหมืองข้อมูล โดยในการพัฒนาระบบนี้ทำการแปลงข้อมูลโดยการ Discrete ข้อมูลจากข้อมูลแบบค่าต่อเนื่องแปลงไปเป็นข้อมูลแบบเป็นช่วงข้อมูล โดยผู้ใช้งานกำหนดข้อมูลที่ต้องการ Discrete เพื่อให้ได้แบบข้อมูลที่ต้องการ

แอคเตอร์

1. ผู้ใช้งาน(User)

เงื่อนไขก่อนเข้ายูสเคส

1. ข้อมูลต้องผ่านกระบวนการคลีนข้อมูลเรียบร้อยแล้ว

ลำดับเหตุการณ์หลัก

1. ระบบแสดงแอตทริบิวท์ในการแปลงข้อมูล
2. ผู้ใช้งานเลือกแอตทริบิวท์ที่ต้องการเพื่อนำมาแปลงข้อมูลแบบ Discrete
3. ระบบแสดงค่า min, max, distinct Value ของแอตทริบิวท์ที่เลือก
4. ผู้ใช้งานกำหนดจำนวนกลุ่มของช่วงข้อมูลที่ต้องการแบ่ง
5. ผู้ใช้งานกำหนดค่าเริ่มต้นและค่าสิ้นสุดของแต่ละช่วงข้อมูล
6. ผู้ใช้งานกดปุ่ม Add
7. ระบบแสดงแอตทริบิวท์ที่ทำการ Discrete

   วนซ้ำข้อ4
8. ผู้ใช้งานกดปุ่ม Discretize Data
9. ระบบทำการ Discrete ข้อมูลตามที่ผู้ใช้งานกำหนด
10. ระบบแจ้งข้อความ Discretize Data Complete
11. ระบบแสดงแอตทริบิวท์แบบ Discrete

ลำดับเหตุการณ์ย่อย

-

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบทำการแปลงข้อมูลเรียบร้อยแล้ว

รูปที่ 3.6 คำอธิบายยูสเคส UC3: transformData

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.7 แอคทิวิตีไดอะแกรม transformData

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส UC4: buildClassifier

รายละเอียด

เป็นยูสเคสอธิบายขั้นตอนการสร้างแบบจำลองการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ โดยใช้ข้อมูลที่ได้จากการแปลงข้อมูลมาใช้ในการสร้างแบบจำลอง โดยใช้ซีบีเอ อัลกอริทึม โดยพิจารณาจากค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำ

แอคเตอร์

1. ผู้ใช้งาน(User)

เงื่อนไขก่อนเข้ายูสเคส

1. ข้อมูลต้องผ่านกระบวนการแปลงข้อมูลเรียบร้อยแล้ว

ลำดับเหตุการณ์หลัก

1. ระบบแสดงแอตทริบิวท์ของตารางที่แปลงข้อมูลแล้ว
2. ผู้ใช้งานเลือกแอตทริบิวท์ในตาราง
3. ผู้ใช้งานกดปุ่ม OK เพื่อยืนยันการเลือก
4. ระบบแสดงแอตทริบิวท์ที่ผู้ใช้งานเลือก
5. ผู้ใช้งานกำหนดแอตทริบิวท์ค่า Condset
6. ผู้ใช้งานกำหนดคลาสปลายทาง(Class Label)
7. ผู้ใช้งานกำหนดค่าสนับสนุนขั้นต่ำ(Minimum Support)
8. ผู้ใช้งานกำหนดค่าความเชื่อมั่นขั้นต่ำ(Minimum Condidence)
9. ผู้ใช้งานกดปุ่ม Build Classifier Model
10. ระบบทำการประมวลผลเพื่อสร้างแบบจำลอง
11. ระบบแสดงผลแบบจำลองที่ถูกสร้างขึ้น

ลำดับเหตุการณ์ย่อย

8a) ผู้ใช้ไม่ได้กำหนดค่าสนับสนุนขั้นต่ำหรือค่าความเชื่อมั่นขั้นต่ำ

8a1) ระบบแจ้งข้อความเตือน “กรุณาระบุ Minimum Support”

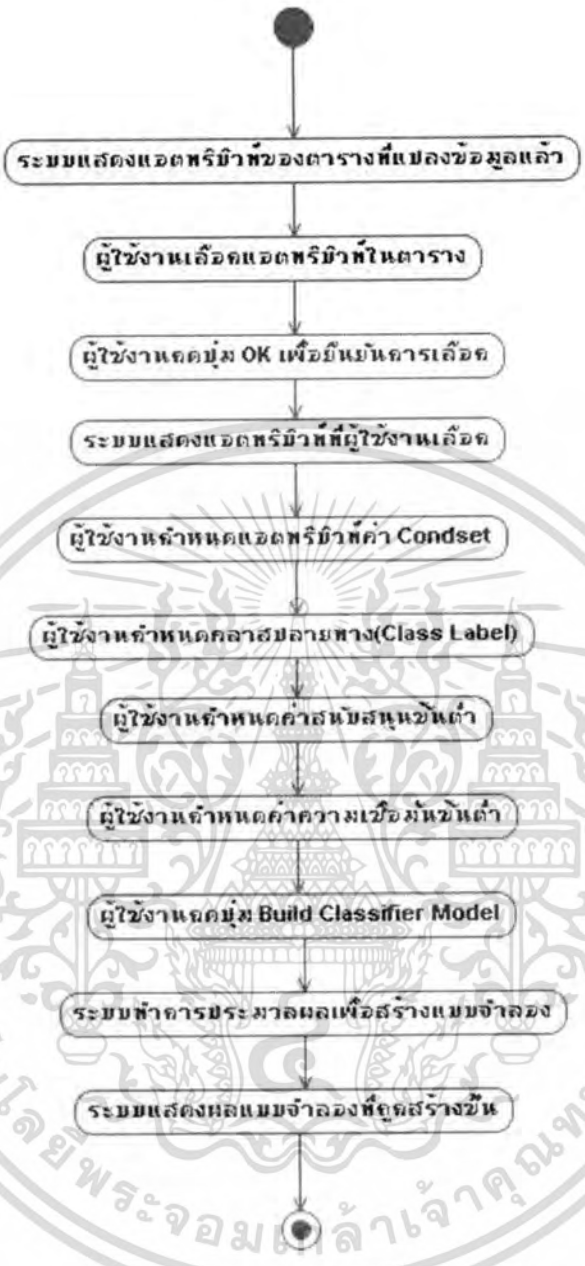
8a2) ระบบแจ้งข้อเตือน “กรุณาระบุ Minimum Confidence”

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบสร้างแบบจำลองเรียบร้อย

รูปที่ 3.8 คำอธิบายยูสเคส UC4: buildClassifier

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 แอคทิวิตีไดอะแกรม buildClassifier

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส UC5: saveClassiferModel

รายละเอียด

ยูสเคสอธิบายขั้นตอนการบันทึกผลลัพธ์แบบจำลองการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ที่ได้จากการนำข้อมูล(training data)เข้ามาประมวลผล โดยทำการบันทึกข้อมูลแบบจำลองลงในรูปแบบฐานข้อมูล

แอคเตอร์

1. ผู้ใช้งาน(User)
2. ระบบ(System)

เงื่อนไขก่อนเข้ายูสเคส

ระบบเรียนรู้และแสดงผลลัพธ์แบบจำลองผ่านทางหน้าจอระบบ

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานกดปุ่ม Save Classifier Model
2. ระบบแสดงหน้าจอเพื่อบันทึกชื่อ Classifier Model
3. ผู้ใช้งานพิมพ์ชื่อ Classifier Model
4. ผู้ใช้งานกดปุ่ม Save เพื่อบันทึกแบบจำลอง
5. ระบบทำการบันทึกแบบจำลอง
6. ระบบแสดงข้อความ "ระบบบันทึกชื่อ Classifier Model เรียบร้อยแล้ว"

ลำดับเหตุการณ์ย่อย

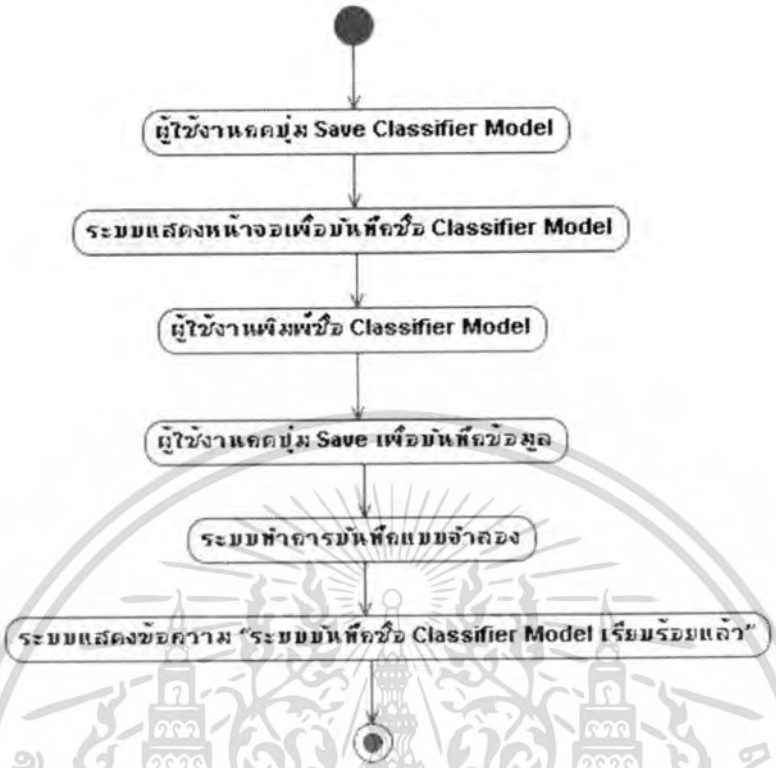
3a) ผู้ใช้ไม่ได้พิมพ์ชื่อ Classifier Model : ระบบแจ้งเตือน "กรุณาระบุชื่อ Classifier Model"

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบบันทึกแบบจำลองในรูปแบบฐานข้อมูลเรียบร้อยแล้ว

รูปที่ 3.10 คำอธิบายยูสเคส UC5: saveClassiferModel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.11 แอคทิวิตีไดอะแกรม saveClassiferModel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส UC6: testClassifierModel

รายละเอียด

ยูสเคสอธิบายขั้นตอนการทดสอบความแม่นยำของแบบจำลองจากข้อมูลทดสอบระบบ โดยพิจารณาค่าความแม่นยำจากค่าเปอร์เซ็นต์ความถูกต้องจากการทดสอบ

แอคเตอร์

1. ผู้ใช้งาน(User)

เงื่อนไขก่อนเข้ายูสเคส

1. ระบบสร้างแบบจำลองเรียบร้อย

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานกดปุ่ม Test Model
2. ผู้ใช้กำหนดชื่อเครื่องเซิร์ฟเวอร์ และชื่อฐานข้อมูล
3. ผู้ใช้กดปุ่ม Connect
4. ระบบแสดงตารางทั้งหมดที่มีในฐานข้อมูล
5. ผู้ใช้งานเลือกตารางในฐานข้อมูลเพื่อนำมาเป็นข้อมูลทดสอบ
6. ระบบแสดงแอตทริบิวท์ที่มีในตาราง
7. ผู้ใช้เลือกแอตทริบิวท์สำหรับการทำเหมืองข้อมูล
8. ผู้ใช้งานกดปุ่ม Execute เพื่อประมวลผล
9. ระบบแสดงข้อมูลแบบจำลองต้นแบบที่มีในฐานข้อมูล
10. ผู้ใช้งานเลือกชื่อแบบจำลอง
11. ระบบแสดงแบบจำลองตามชื่อที่เลือก
12. ผู้ใช้งานพิจารณาแบบจำลองจากค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำ
13. ระบบทำการทดสอบความแม่นยำของแบบจำลองด้วยข้อมูลทดสอบระบบ
14. แสดงผลลัพธ์การทดสอบความแม่นยำของแบบจำลองผ่านทางหน้าจอระบบ

ลำดับเหตุการณ์ย่อย

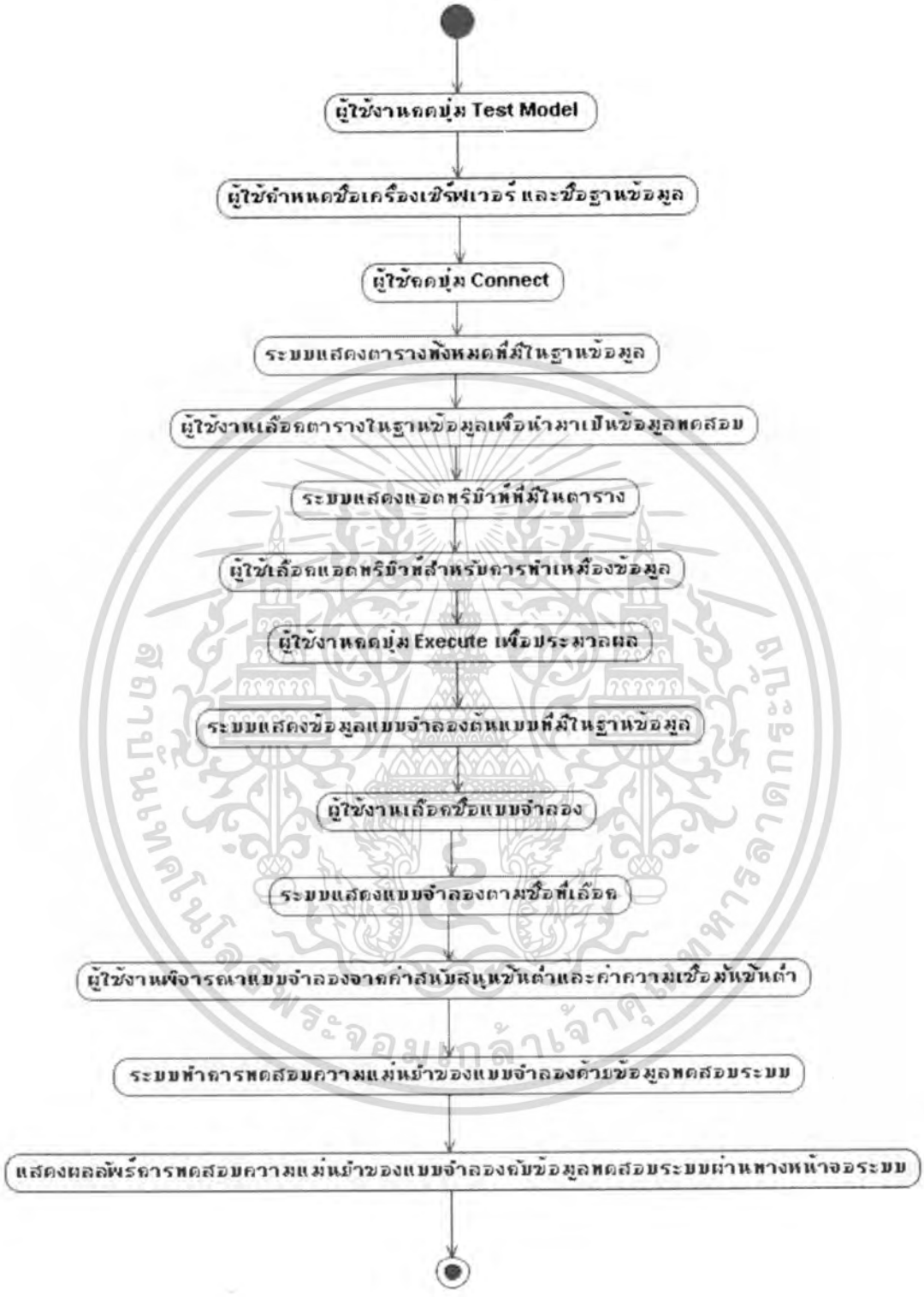
-

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบแสดงผลความแม่นยำของแบบจำลองกับข้อมูลทดสอบระบบผ่านทางหน้าจอ

รูปที่ 3.12 คำอธิบายยูสเคส UC6: testClassifierModel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.13 แอคทิวิตีไดอะแกรม testClassifierModel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส UC7: showClassifierModel

รายละเอียด

ยูสเคสอธิบายขั้นตอนการเรียกดูรายละเอียดแบบจำลอง(Classifier Model)การจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ โดยพิจารณาแบบจำลองจากค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำ ที่มีค่าเปอร์เซ็นต์ความถูกต้องสูงสุดจากการประเมินแบบจำลองและจากการทดสอบแบบจำลอง

แอคเตอร์

1. ระบบ
2. ผู้ใช้งาน(User)

เงื่อนไขก่อนเข้ายูสเคส

1. ระบบสร้างและจัดเก็บข้อมูลแบบจำลองไว้ในรูปแบบฐานข้อมูลแล้ว

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานเลือกเมนู Load Classifier Model
2. ระบบแสดงหน้าจอรายละเอียด Classifier Model
3. ผู้ใช้งานคลิกเลือกแบบจำลองจากชื่อ Classifier Model
4. ระบบแสดงแบบจำลองตามชื่อ Classifier Model ที่ถูกเลือก
5. ผู้ใช้งานคลิกเลือกแบบจำลองจากค่าสนับสนุนและค่าความเชื่อมั่นขั้นต่ำที่มีเปอร์เซ็นต์ความถูกต้องจากการประเมินและการข้อมูลทดสอบระบบที่มีค่าสูงสุด
6. ระบบแสดงผลรายละเอียดแบบจำลองที่ถูกเลือกผ่านทางหน้าจอระบบ

ลำดับเหตุการณ์ย่อย

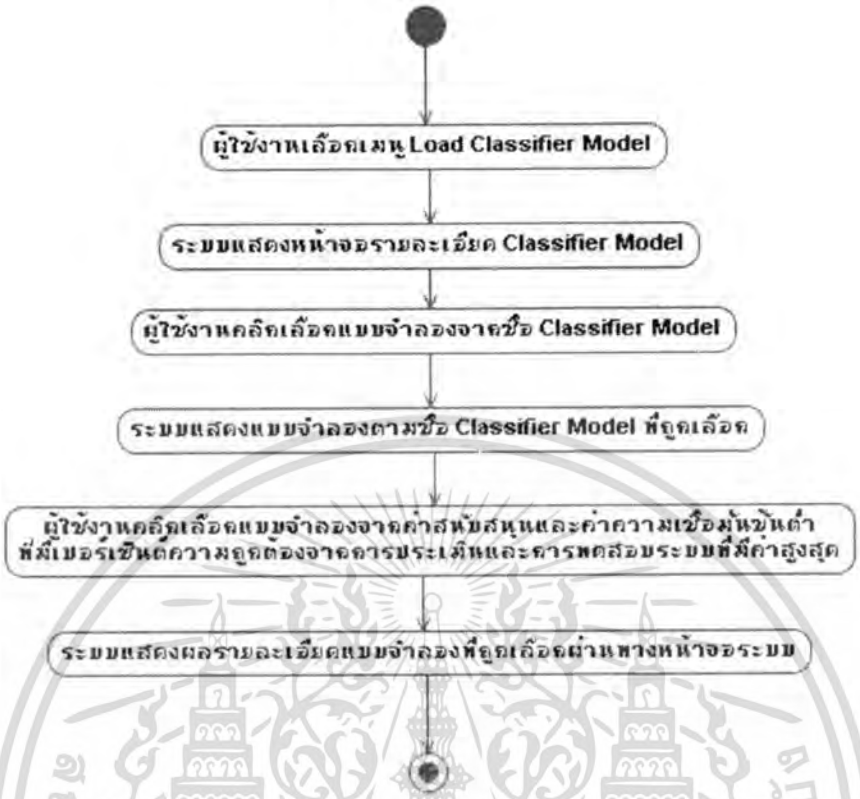
-

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบแสดงผลรายละเอียดแบบจำลองที่ถูกเลือกผ่านทางหน้าจอเรียบร้อยแล้ว

รูปที่ 3.14 คำอธิบายยูสเคส UC7: showClassifierModel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.15 แอคทิวิตีไดอะแกรม showClassifierModel

ยูสเคส UC8: predictData

รายละเอียด

ยูสเคสอธิบายขั้นตอนการทำนายประเภทกลุ่มข้อมูลของข้อมูลใหม่(Unseen data)ที่เข้ามาโดยใช้แบบจำลองที่ผู้ใช้พิจารณาเพื่อทำการทำนายประเภทกลุ่มข้อมูล

แอคเตอร์

1. ผู้ใช้งาน(User)

เงื่อนไขก่อนเข้ายูสเคส

1. ระบบได้ทำการจัดเก็บแบบจำลองไว้ในฐานข้อมูลแล้ว

ลำดับเหตุการณ์หลัก

1. ผู้ใช้กำหนดชื่อเครื่องเซิร์ฟเวอร์ และชื่อฐานข้อมูลเพื่อเลือกข้อมูลใหม่สำหรับทำนาย
2. ผู้ใช้กดปุ่ม Connect

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ระบบแสดงตารางทั้งหมดในฐานข้อมูล
4. ผู้ใช้งานเลือกตารางในฐานข้อมูลเพื่อนำข้อมูลมาใช้ในการทำนาย
5. ผู้ใช้งานเลือกแอตทริบิวท์ในตาราง
6. ผู้ใช้งานกดปุ่ม Execute
7. ระบบแสดงแบบจำลองสำหรับการทำนายข้อมูล
8. ผู้ใช้งานคลิกเลือก ชื่อแบบจำลองเพื่อนำมาเป็นแบบจำลองสำหรับการทำนาย
9. ระบบแสดงแบบจำลองตามชื่อที่ผู้ใช้งานเลือก
10. ผู้ใช้งานเลือกแบบจำลองพิจารณาจากเปอร์เซ็นต์ความถูกต้องจากการประเมินของแบบจำลองและจากข้อมูลการทดสอบระบบ
11. ระบบแสดงรายละเอียดแบบจำลองที่ผู้ใช้พิจารณาเลือกผ่านทางหน้าจอ
12. ผู้ใช้กดปุ่ม OK เพื่อยืนยันการเลือกแบบจำลอง
13. ระบบแสดงหน้าจอสำหรับการทำนายข้อมูล
14. ระบบแสดงแอตทริบิวท์สำหรับเลือกเพื่อทำนายข้อมูล
15. ผู้ใช้งานกำหนดค่าแอตทริบิวท์ที่เป็นข้อมูลใหม่ตามต้องการ
16. ผู้ใช้งานกดปุ่ม Add เพื่อเพิ่มข้อมูล

    วนซ้ำข้อ 15
17. ระบบแสดงรายการข้อมูลใหม่ที่เลือกสำหรับใช้ในการจำแนกประเภทกลุ่มข้อมูล
18. ผู้ใช้งานกดปุ่ม Start Predict ทำนายข้อมูล
19. ระบบแสดงผลลัพธ์การทำนายประเภทกลุ่มข้อมูลให้กับข้อมูลใหม่

ลำดับเหตุการณ์ย่อย

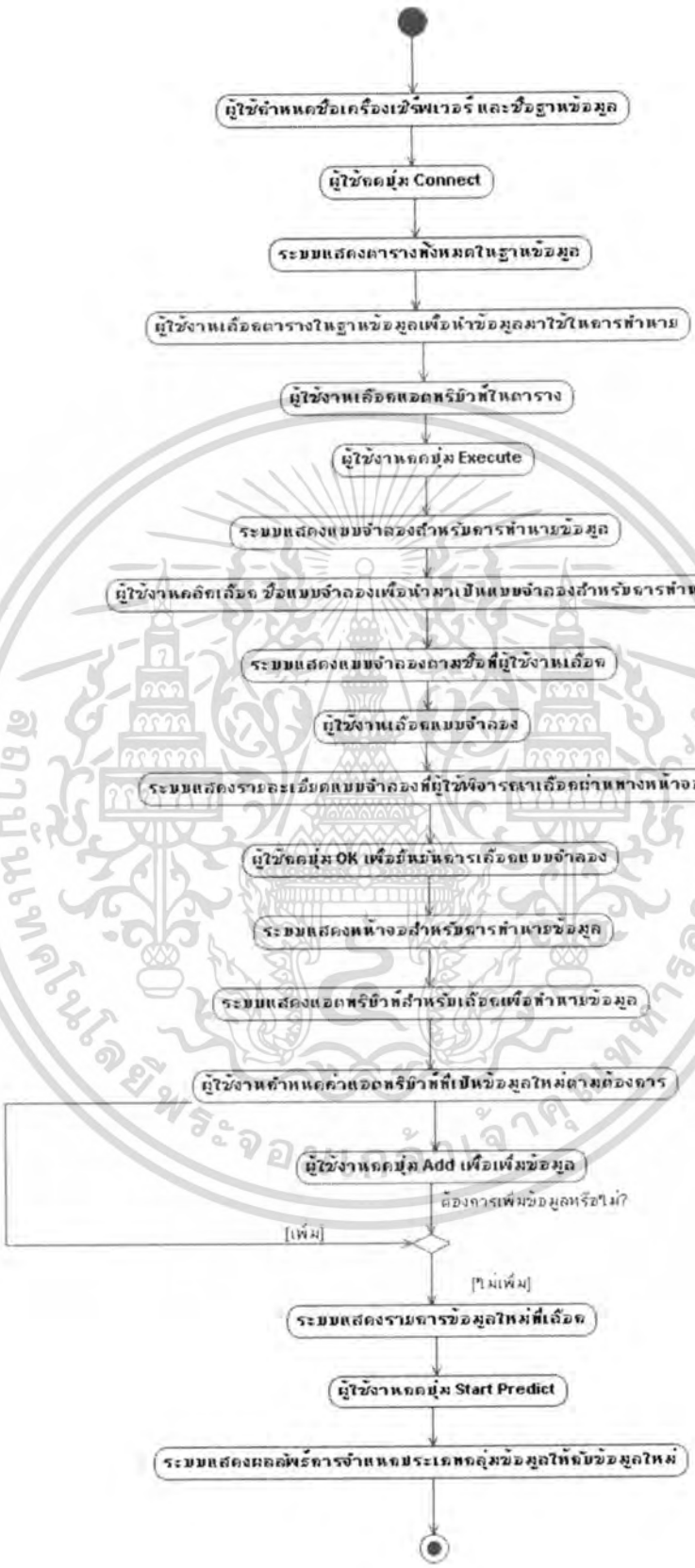
16a) ผู้ใช้งานไม่ได้เลือกแอตทริบิวท์เพื่อการทำนาย : ระบบแสดงข้อความ "กรุณาเลือกรายการก่อน"

เงื่อนไขก่อนออกจากยูสเคส

1. ระบบแสดงผลลัพธ์การทำนายประเภทข้อมูลให้กับข้อมูลใหม่เรียบร้อย

รูปที่ 3.16 คำอธิบายยูสเคส UC8: predictData

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

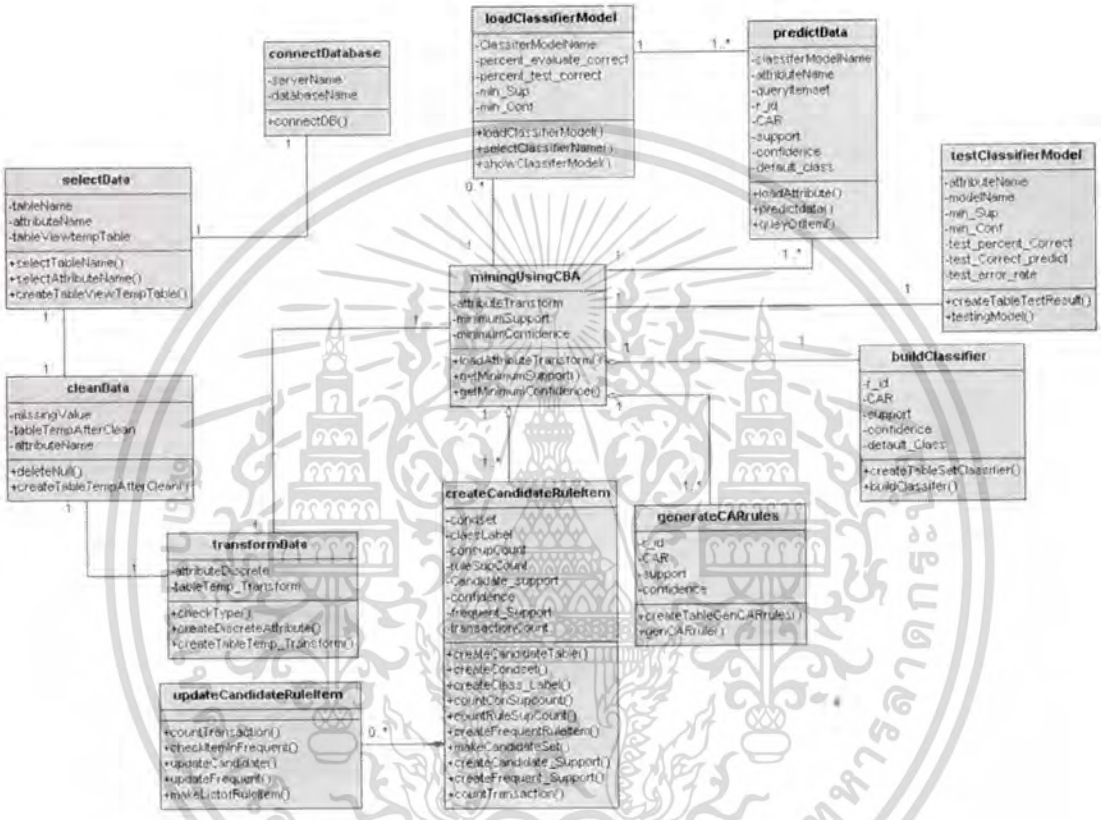


รูปที่ 3.17 แอคทิวิตีไดอะแกรม predictData

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การออกแบบระบบโดยใช้คลาสไดอะแกรม และ ซีเควนไดอะแกรม

การออกแบบคลาสและความสัมพันธ์ระหว่างคลาส ในกระบวนการพัฒนาระบบ สามารถอธิบายโดยใช้แบบจำลองคลาสไดอะแกรมและซีเควนไดอะแกรม ดังนี้



รูปที่ 3.18 คลาสไดอะแกรมการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์

จากรูปที่ 3.18 สามารถอธิบายการทำงานของ คลาสไดอะแกรมการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ ได้ดังนี้

- คลาส user คือ คลาสของผู้ใช้งานระบบ
- คลาส connectDatabase มีหน้าที่เชื่อมต่อฐานข้อมูล มีความสัมพันธ์กับคลาส selectData
- คลาส selectData มีหน้าที่เลือกข้อมูล มีความสัมพันธ์กับคลาส connectDatabase โดยเมื่อต้องการ selectData จะต้องทำการติดต่อกับคลาส connectDatabase ก่อน และมีความสัมพันธ์กับคลาส cleanData คือเลือกข้อมูลเพื่อเตรียมเข้าสู่ขั้นตอนการคลีนข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

• คลาส cleanData มีหน้าที่คลีนข้อมูลที่เป็นค่าว่าง มีความสัมพันธ์กับคลาส selectData เมื่อคลาส cleanData ต้องการทำงานในฟังก์ชันนี้จะต้องติดต่อกับคลาส selectData เพื่อเลือกข้อมูลสำหรับคลีนก่อน และมีความสัมพันธ์กับ คลาส transform เพื่อเตรียมข้อมูลก่อนเข้าสู่ขั้นตอนการแปลงข้อมูล

• คลาส transformData มีหน้าที่แปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับขั้นตอนต่อไปคือขั้นตอนการทำเหมืองข้อมูล มีความสัมพันธ์กับคลาส miningUsingCBA

• คลาส miningUsingCBA มีหน้าที่แสดงแอตทริบิวท์จากคลาส transformData เข้ามา และรับค่า minimumSupport และค่า minimumConfidence เพื่อใช้ในการสร้างแบบจำลอง

• คลาส createCandidataRuleItem มีหน้าที่สร้าง candidateRuleItem เพื่อใช้สร้างกฎความสัมพันธ์ มีความสัมพันธ์กับคลาส updateCandidateRuleItem

• คลาส updateCandidateRuleItem มีหน้าที่ updateFrequentRuleItem และ updateCandidateruleItem จากคลาส createCandidateRuleItem

• คลาส genarateCARrules มีหน้าที่สร้างกฎความสัมพันธ์แบบ CARs เพื่อใช้ในการสร้างแบบจำลอง

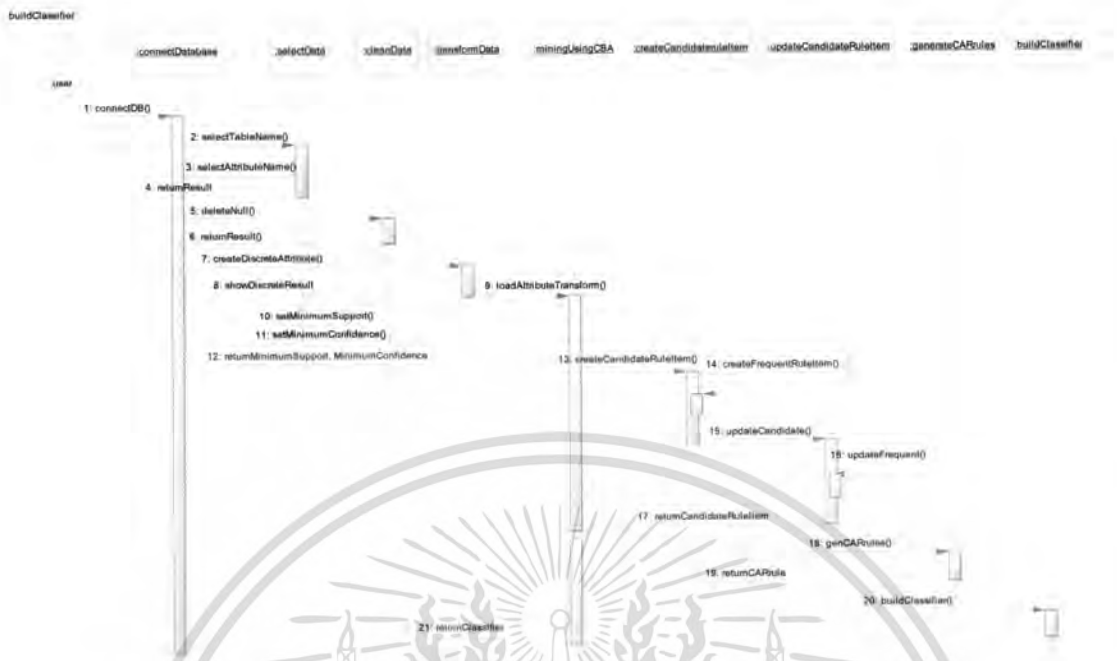
• คลาส buildClassifier มีหน้าที่ สร้างแบบจำลองจำแนกประเภทข้อมูล

• คลาส testClassifierModel มีหน้าที่ทดสอบความแม่นยำของแบบจำลอง ที่สร้างได้จากคลาส miningUsingCBA

• คลาส predictData มีหน้าที่จำแนกประเภทข้อมูลให้กับข้อมูลใหม่ มีความสัมพันธ์กับคลาส loadClassifierModel เพื่อเรียกดูข้อมูลรายละเอียด ClassifierModel

• คลาส loadClassifierModel มีหน้าที่เรียกดูและแสดง ClassifierModel มีความสัมพันธ์กับคลาส predictData และ คลาส miningUsingCBA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.19 ซีเควนไดอะแกรมการสร้างแบบจำลองจำแนกประเภทข้อมูล

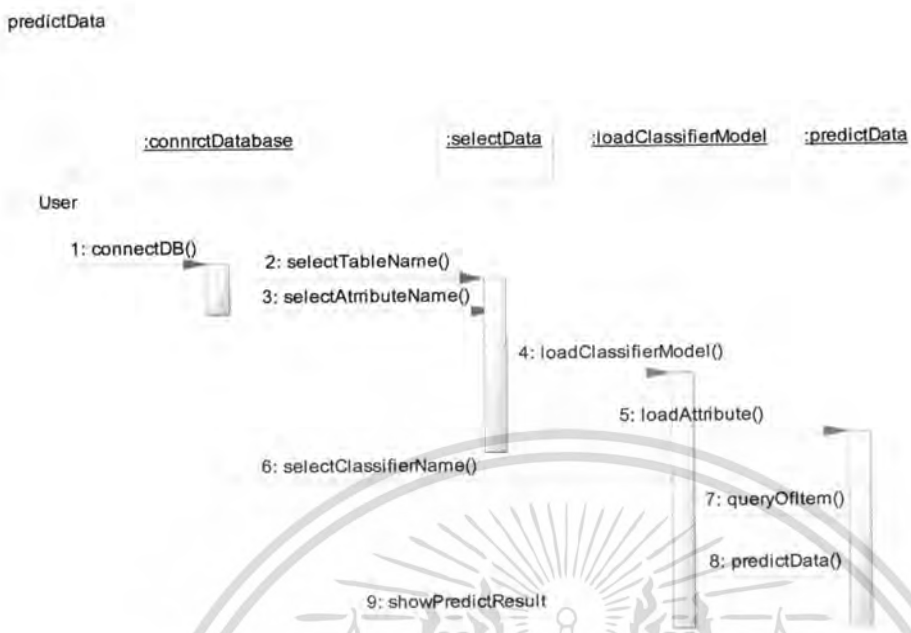
จากรูปที่ 3.19 ซีเควนไดอะแกรมการสร้างแบบจำลองจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ สามารถอธิบายการทำงานได้ดังนี้

1. ผู้ใช้งานติดต่อคลาส connectDatabase เรียกใช้เมทธอด connectDB ทำการเชื่อมต่อฐานข้อมูล
2. คลาส selectData เรียกใช้เมทธอด selectTableName เพื่อเลือกตาราง และเลือกเมทธอด selectAttributeName เพื่อเลือกแอตทริบิวท์ให้กับคลาส selectData
3. คลาส selectData ทำการ return ค่าแอตทริบิวท์ที่ผู้ใช้งานเลือก
4. คลาส cleanData เรียกใช้เมทธอด deleteNull เพื่อลบข้อมูลที่เป็นค่าว่างออกทั้งเรคคอร์ด
5. คลาส cleanData ทำการ return ค่าแอตทริบิวท์ที่ทำการคลีนข้อมูลเรียบร้อยแล้ว
6. คลาส TransformData เรียกใช้เมทธอด createDiscreteAttribute เพื่อสร้างแอตทริบิวท์ที่ผ่านการแปลงข้อมูล และ ทำการ showDiscreteAttribute เพื่อแสดงค่าแอตทริบิวท์ที่ทำการแปลงแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. คลาส miningUsingCBA เรียกใช้เมทธอด loadAttributeTransform เพื่อแสดงข้อมูลแอตทริบิวท์ที่ผ่านการแปลงข้อมูลแล้วสำหรับขั้นตอนการทำเหมืองข้อมูล
8. ผู้ใช้กำหนดค่าสนับสนุนขั้นต่ำให้กับคลาส miningUsingCBA
9. ผู้ใช้กำหนดค่าความเชื่อมั่นขั้นต่ำให้กับคลาส miningUsingCBA
10. คลาส miningUsingCBA ทำการแสดงค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำไปสู่ผู้ใช้งาน
11. คลาส createCandidateruleItem เรียกใช้เมทธอด createCandidateRuleItem เพื่อสร้างข้อมูล candidateRuleItem และเรียกใช้เมทธอด createFrequentRuleItem เพื่อสร้างข้อมูล frequentRuleItem
12. คลาส updateCandidateRuleItem เรียกใช้เมทธอด updateCandidate เพื่อปรับปรุงข้อมูล candidateRuleItem และเรียกใช้เมทธอด updateFrequent เพื่อปรับปรุงข้อมูล FrequentRuleItem
13. คลาส UpdateCandidateRuleItem ทำการ Return ค่า updateCandidateRuleItem
14. คลาส generateCARrules เรียกใช้เมทธอด genCARrules เพื่อสร้างข้อมูล generateCARrule
15. คลาส buildClassifier เรียกใช้เมทธอด buildClassifier เพื่อสร้างข้อมูล ClassifierModel
16. คลาส buildClassifier ทำการส่งผลลัพธ์ returnClassifier ไปยังผู้ใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.20 ซีเควนไดอะแกรมการทำนายประเภทข้อมูล

จากรูปที่ 3.20 ซีเควนไดอะแกรมการทำนายประเภทข้อมูล สามารถอธิบายการทำงานได้ดังนี้

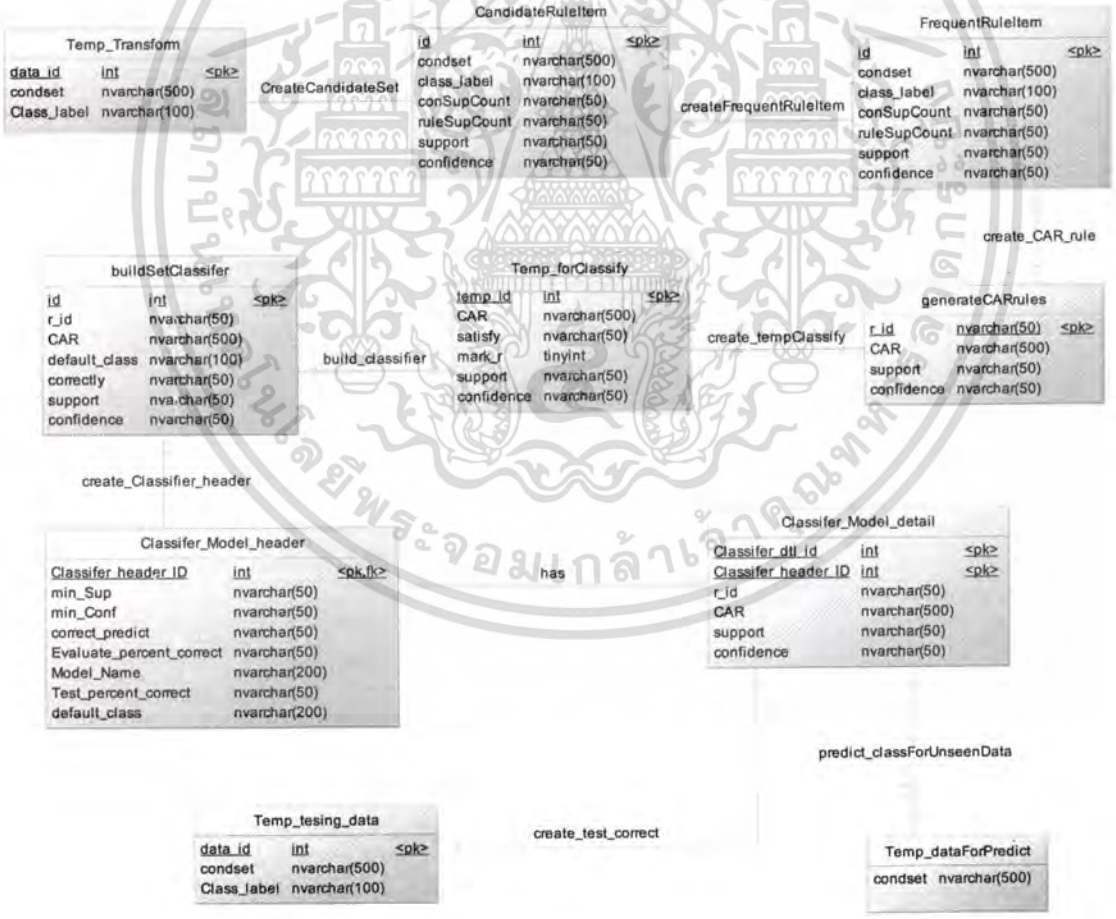
1. ผู้ใช้งานทำการเชื่อมต่อฐานข้อมูลกับคลาส ConnectDatabase
2. คลาส SelectData เรียกใช้เมทธอด selectTableName เพื่อเลือกตารางในฐานข้อมูล และ เรียกใช้เมทธอด selectAttributeName เพื่อเลือกแอตทริบิวท์
3. คลาส loadClassifierModel เรียกใช้เมทธอด loadclassifierModel เพื่อดึงข้อมูล ClassifierModel มาแสดงข้อมูล
4. คลาส predictData เรียกใช้เมทธอด loadAttribute เพื่อดึงข้อมูลที่ผู้ใช้เลือกสำหรับการทำนายประเภทข้อมูล
5. ผู้ใช้งานทำการเลือก ClassifierModel ตามชื่อของ classifier

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. คลาส predictData เรียกใช้เมทธอด queryOfItem เพื่อเลือกข้อมูลสำหรับจำแนกประเภทข้อมูลแบบ 1 เรคคอร์ด และ เรียกใช้เมทธอด predictData เพื่อจำแนกประเภทข้อมูลให้กับข้อมูลใหม่ที่ไม่ทราบคลาสปลายทาง
7. คลาส predictData ทำการแสดงค่าผลการจำแนกประเภทข้อมูล

3.4 การออกแบบฐานข้อมูล

การออกแบบฐานข้อมูลของระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเอ อัลกอริทึม สามารถอธิบายเอนทิตีของระบบและความสัมพันธ์ระหว่างเอนทิตีโดยใช้อีอาร์ไดอะแกรม แสดงดังรูปที่ 3.21 ดังนี้



รูปที่ 3.21 อีอาร์ไดอะแกรมการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในรูปที่ 3.21 สามารถอธิบายเอนทิตีของระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบซีบีเอ ได้ดังนี้

1. Temp_Transform คือ ตารางจัดเก็บข้อมูลที่นำไปใช้ในการทำเหมืองข้อมูล โดยข้อมูลที่จัดเก็บผ่านขั้นตอนการเตรียมข้อมูลเพื่อปรับปรุงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมในกระบวนสร้างแบบจำลองจากการค้นหากฎความสัมพันธ์

2. CandidateRuleItem คือ ตารางจัดเก็บข้อมูล CandidateRuleItem เก็บข้อมูลที่ใช้ในการค้นหากฎความสัมพันธ์

3. FrequentRuleItem คือ ตารางจัดเก็บข้อมูล FrequentRuleItem เก็บความสัมพันธ์ของข้อมูลที่พิจารณาผ่านค่า Minimum Support

4. generateCARrule คือ ตารางจัดเก็บข้อมูลกฎความสัมพันธ์แบบมีคลาส(Class Association Rules) ที่ได้จากกระบวนการค้นหากฎความสัมพันธ์เพื่อนำไปใช้ในขั้นตอนการสร้างแบบจำลอง

5. Temp_forClassify คือ ตารางจัดเก็บข้อมูล satisfy และตรวจสอบกฎความสัมพันธ์กับข้อมูลทรานแซคชั่นที่ผ่านค่า correct satisfy

6. buildSetClassifier คือ ตารางเก็บข้อมูลกลุ่มของ Classifier ที่ผ่านค่า correct satisfy เพื่อใช้ในการคำนวณหาค่าความถูกต้องของแบบจำลอง

7. Classifer_Model_header คือตารางจัดเก็บการบันทึกผลลัพธ์จากการพิจารณาค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำของแต่ละแบบจำลอง บันทึกชื่อของแบบจำลอง เปอร์เซ็นต์ความถูกต้องของแบบจำลองจากการประเมิน และจากการทดสอบความแม่นยำของแบบจำลอง

8. Classifer_Model_detail คือตารางจัดเก็บรายละเอียดข้อมูลของแบบจำลองจากแต่ละแบบจำลองที่พิจารณาจากค่าความเชื่อมั่นและค่าสนับสนับสนุนขั้นต่ำ

9. Temp_tesing_data คือตารางจัดเก็บข้อมูลสำหรับใช้ทดสอบระบบ(Testing Data)

10. Temp_dataForPredict คือ ตารางจัดเก็บข้อมูลใหม่สำหรับใช้ทำนายประเภทกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การออกแบบและพัฒนาระบบ

4.1 เครื่องมือที่ใช้ในการออกแบบและพัฒนาระบบ

4.1.1 ฮาร์ดแวร์

การพัฒนาระบบงานใช้เครื่องคอมพิวเตอร์ที่ทีคุณสมบัติดังนี้

- หน่วยประมวลผลกลาง Pentium M ความเร็ว 1.73 GHz
- ฮาร์ดดิสก์ความจุขนาด 80 GB
- หน่วยความจำ 512 MB

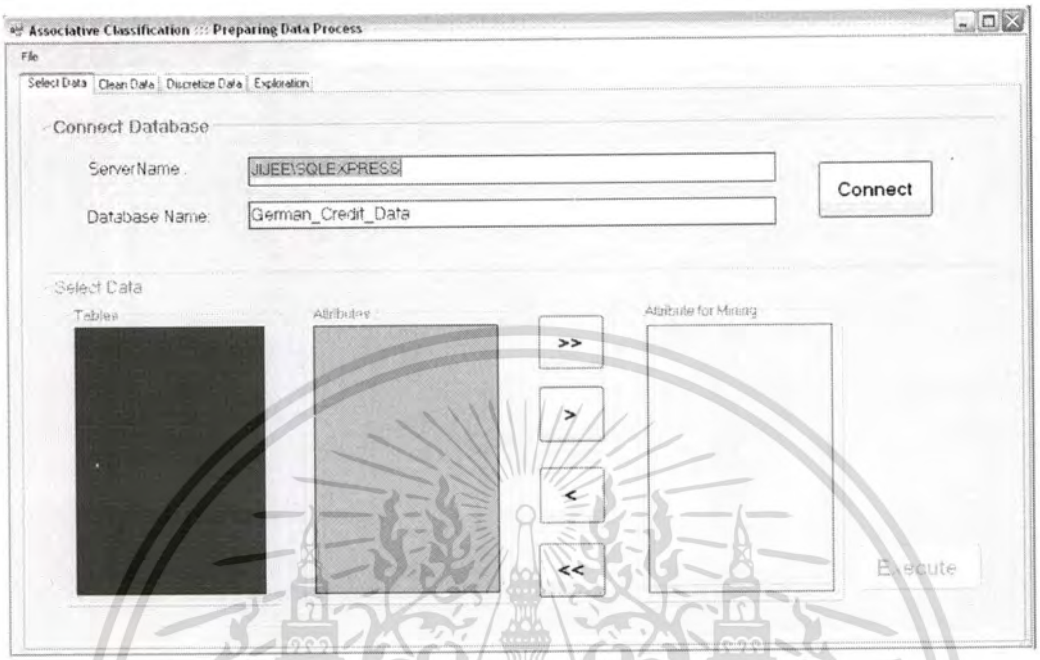
4.1.2 ซอฟต์แวร์

เครื่องมือ และโปรแกรมที่ใช้ในการพัฒนามีรายละเอียดดังนี้

- ระบบปฏิบัติการ คือ Windows XP SP3
- ภาษาที่ใช้ในการพัฒนา คือ VB.NET (Framework 2.0)
- เครื่องมือที่ใช้ในการพัฒนา(IDE) คือ Visual Studio .NET 2005
- ระบบฐานข้อมูล คือ Microsoft SQL Server 2005

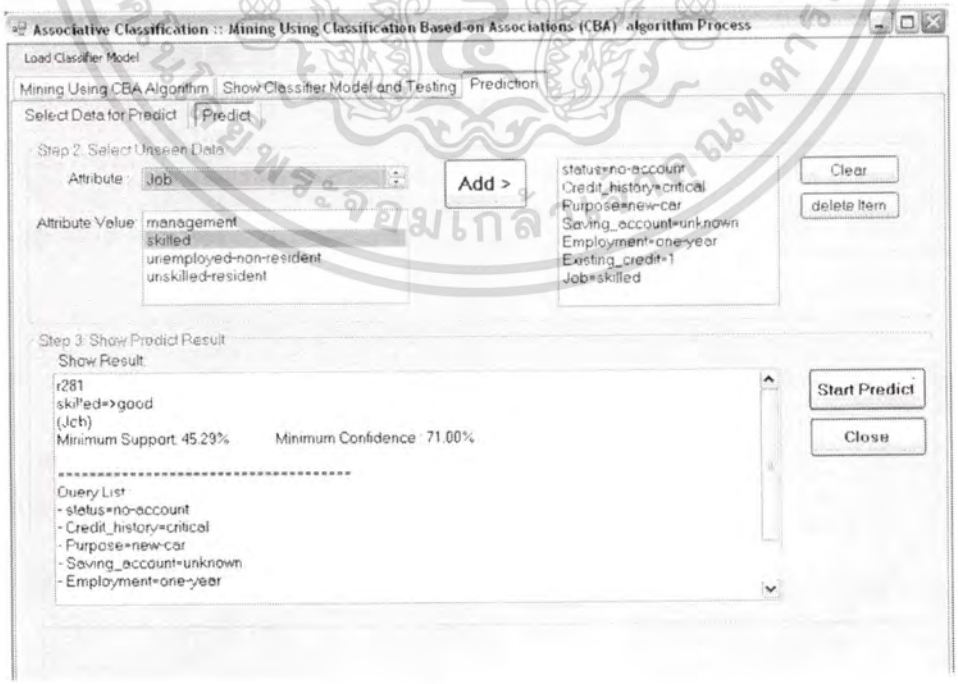
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 การออกแบบหน้าจอการรับข้อมูล



รูปที่ 4.1 หน้าจอการรับข้อมูล

4.3 การออกแบบหน้าจอรายงานผลลัพธ์



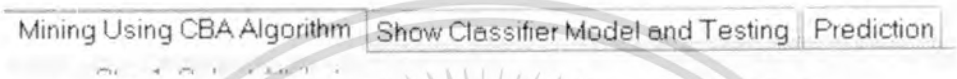
รูปที่ 4.2 หน้าจอแสดงผลลัพธ์การประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 การใช้งานโปรแกรม

การทำงานของระบบจะเป็นไปตามกระบวนการของการทำเหมืองข้อมูล ดังนี้

รูปที่ 4.3 เมนูการทำงานของระบบในส่วนขั้นตอนของการเตรียมข้อมูล



รูปที่ 4.4 เมนูการทำงานของระบบในส่วนขั้นตอนการทำเหมืองข้อมูล

4.4.1 ขั้นตอนการติดต่อฐานข้อมูล

รูปที่ 4.5 ขั้นตอนการติดต่อฐานข้อมูล

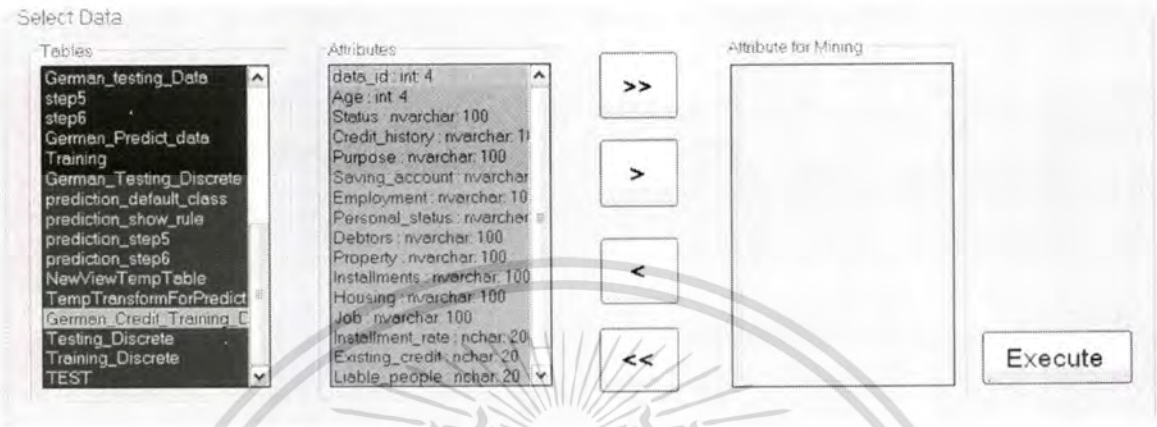
1. ผู้ใช้งานกำหนด ชื่อเซิร์ฟเวอร์ และชื่อฐานข้อมูล
2. กดปุ่ม Connect เพื่อเชื่อมต่อฐานข้อมูล
3. ระบบแสดงข้อความเมื่อระบบเชื่อมต่อกับฐานข้อมูลเรียบร้อยแล้ว

รูปที่ 4.6 กล่องข้อความเมื่อระบบเชื่อมต่อกับฐานข้อมูลเรียบร้อยแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

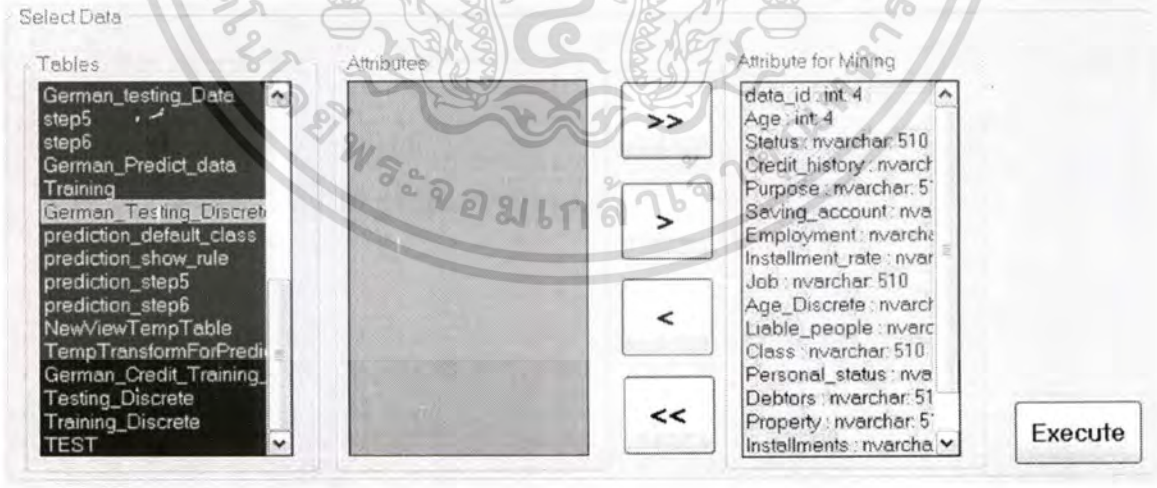
4.4.2 ขั้นตอนการเลือกข้อมูล

ในขั้นตอนการเลือกข้อมูล มีขั้นตอนการทำงานดังนี้



รูปที่ 4.7 การเลือกข้อมูลจากตารางและแอตทริบิวท์

1. ผู้ใช้งานคลิกเลือกตาราง และจะแสดงแอตทริบิวท์ของตารางที่เลือกในช่องแอตทริบิวท์
2. ผู้ใช้งานคลิกเลือกชื่อแอตทริบิวท์ที่จะใช้ในการทำเหมืองข้อมูล

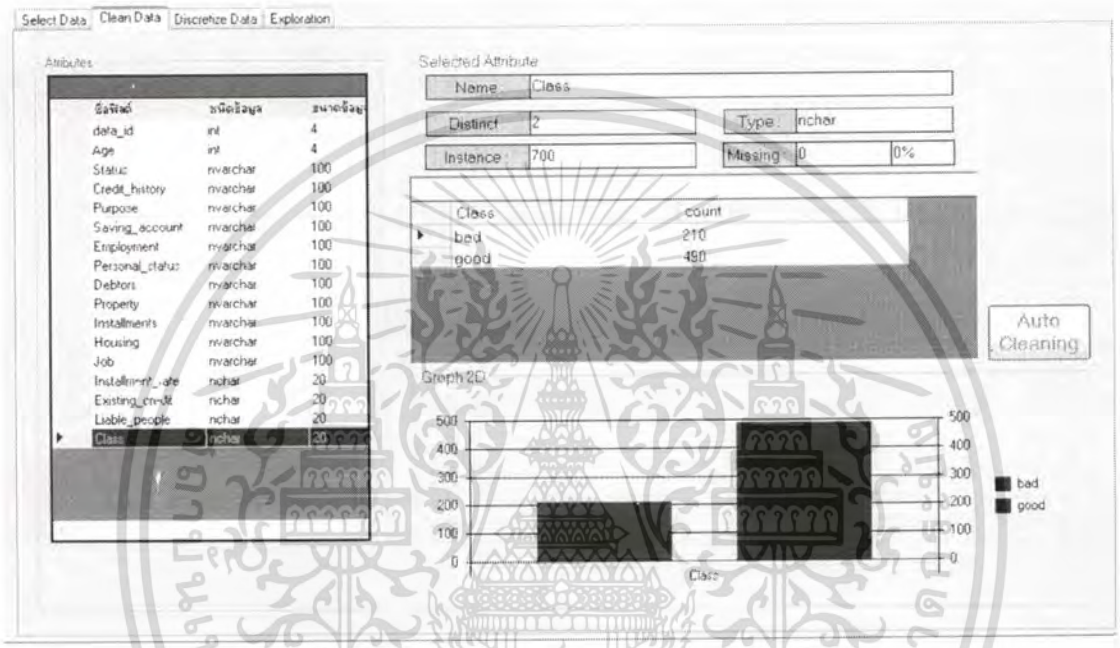


รูปที่ 4.8 การเลือกข้อมูลแอตทริบิวท์สำหรับการทำเหมืองข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ผู้ใช้งานกดปุ่ม [>>] เพื่อเลือกชื่อแอตทริบิวท์ที่ได้เลือกไว้ แอตทริบิวท์ที่ถูกเลือกจะแสดงในช่อง [>] Attribute for Mining
4. กดปุ่ม Execute เพื่อสิ้นสุดขั้นตอนการเลือกข้อมูล

4.4.3 ขั้นตอนการคลีนข้อมูล



รูปที่ 4.9 การคลีนข้อมูล

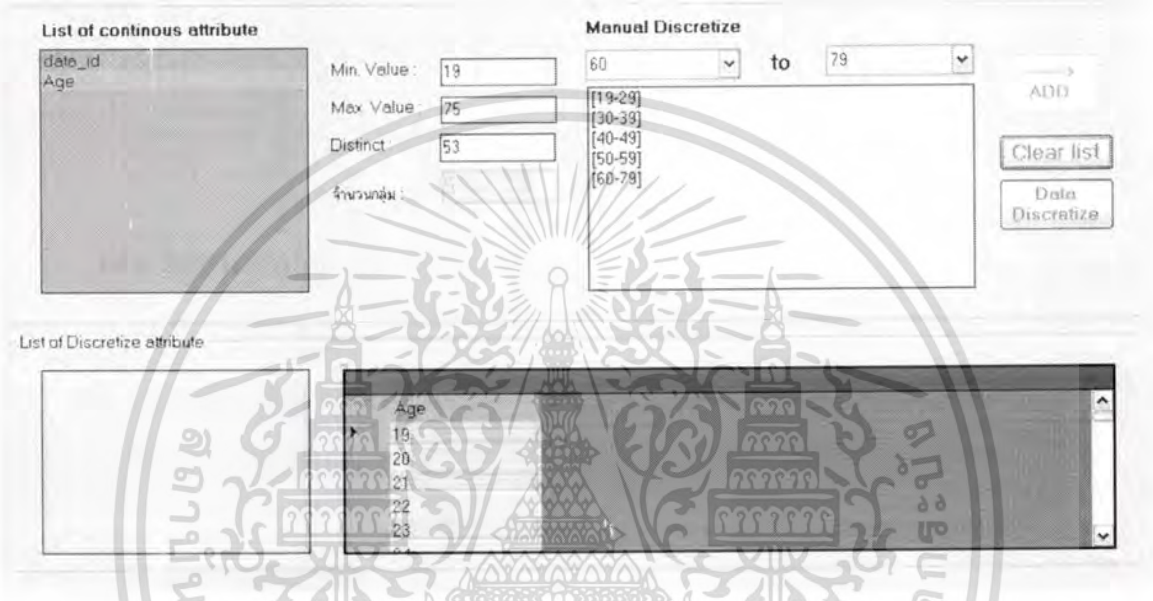
1. ผู้ใช้งานเลือกแอตทริบิวท์ โดยระบบจะแสดงรายละเอียดของแอตทริบิวท์และแสดงกราฟของจำนวนข้อมูลในแอตทริบิวท์ที่ผู้ใช้เลือก
2. ผู้ใช้งานกดปุ่ม [Auto Cleaning] เพื่อทำการคลีนข้อมูลแอตทริบิวท์ที่ได้เลือกไว้
3. ระบบแสดงข้อกล่องข้อความเมื่อทำขั้นตอนเลือกข้อมูลเรียบร้อย

รูปที่ 4.10 กล่องข้อความเมื่อทำขั้นตอนการคลีนข้อมูลเรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.4 ขั้นตอนการแปลงข้อมูล

ในการขั้นตอนการแปลงข้อมูลสำหรับการทำเหมืองข้อมูลแบบจำลองการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ แบบซีบีเอ อัลกอริทึม ในที่นี้จะกล่าวถึงกระบวนการแปลงเป็นข้อมูลแบบ Discrete โดยมีขั้นตอนดังนี้



รูปที่ 4.11 ขั้นตอนการ Discrete ข้อมูล

1.ผู้ใช้งานเลือกแอตทริบิวท์ที่ต้องการนำมาปรับเปลี่ยนเป็นรูปแบบ Discrete

2. ผู้ใช้งานกำหนดจำนวนกลุ่มช่วงข้อมูลที่ต้องการแบ่ง

3. ผู้ใช้งานกำหนดค่าเริ่มต้น

4. และผู้ใช้กำหนดค่าสิ้นสุดของแต่ละช่วงข้อมูล

5. ผู้ใช้งานกดปุ่มAdd เพื่อเพิ่มข้อมูลลงรายการ

6. ผู้ใช้งานกดปุ่ม Discretize Data เมื่อทำการกำหนดช่วงข้อมูลครบตามจำนวนกลุ่มที่ได้กำหนดไว้แล้ว เพื่อแปลงข้อมูลเป็นข้อมูลแบบ discrete

6. ระบบแสดงกล่องข้อความ Discretize Data Complete เมื่อทำขั้นตอนการ Discrete ข้อมูลเรียบร้อยแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

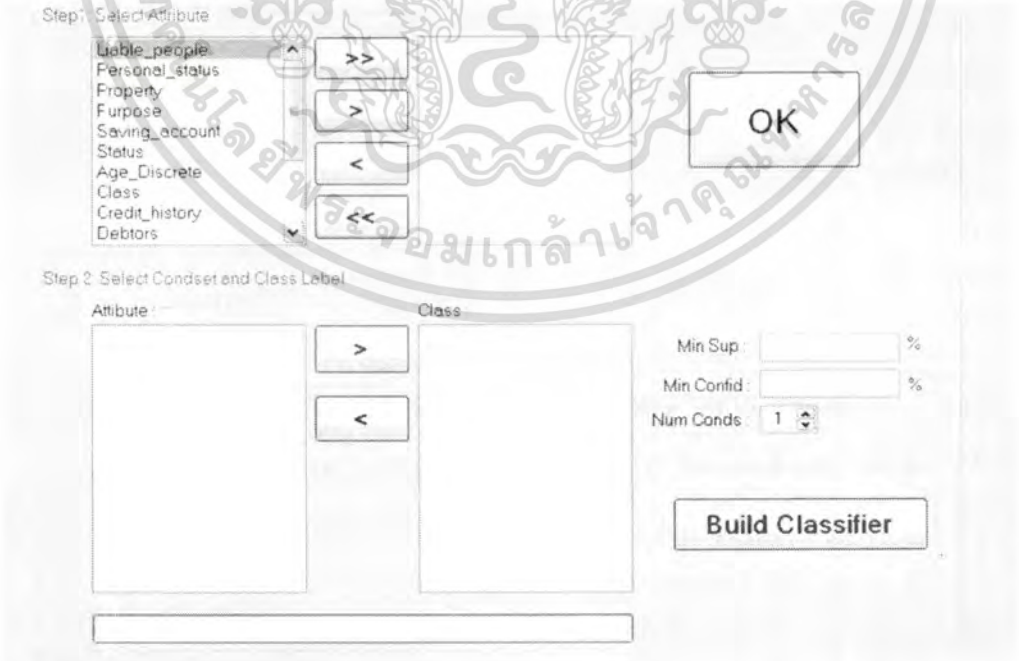
รูปที่ 4.12 กล่องข้อความเมื่อทำขั้นตอนการแปลงข้อมูลเรียบร้อยแล้ว

7. เมื่อผู้ใช้งานทำการ Disctete ข้อมูลเรียบร้อยแล้ว จะแสดงข้อมูลชนิด Discrete

รูปที่ 4.13 แสดงคอลัมน์ที่แปลงข้อมูลแล้วเป็นข้อมูลประเภท Discrete

4.4.5 ขั้นตอนการทำเหมืองข้อมูลแบบซีบีเออัลกอริทึม สร้างแบบจำลอง

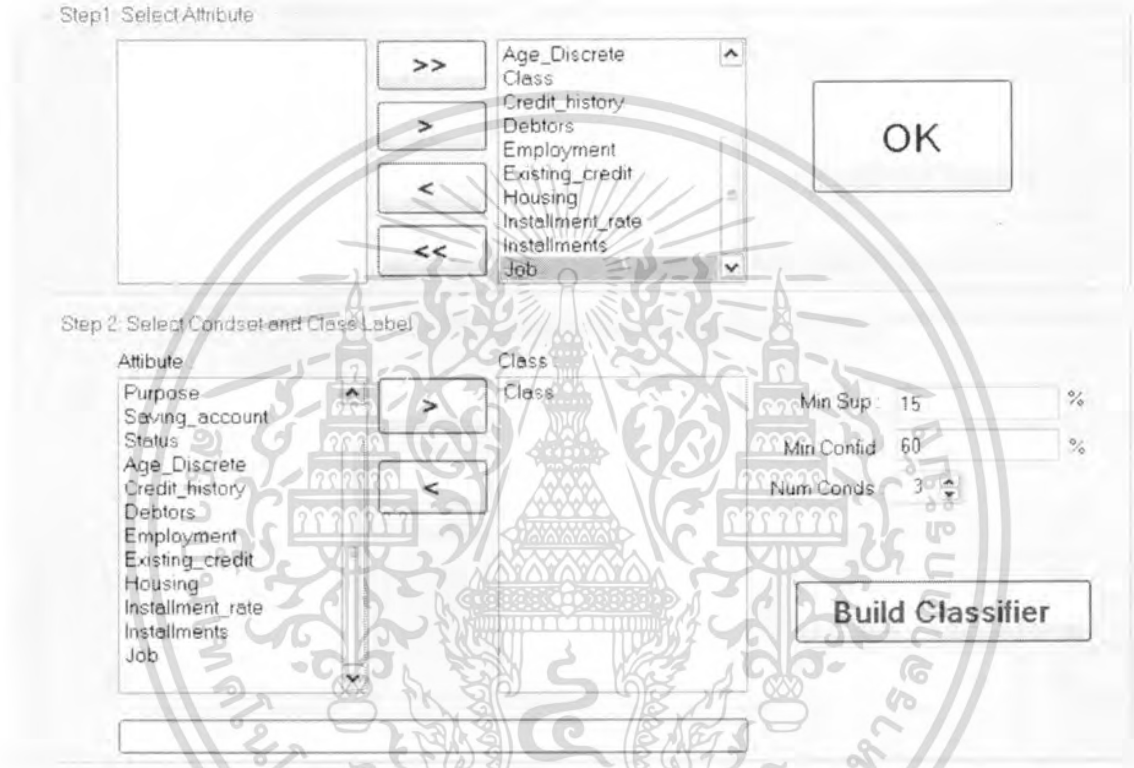
หลังจากจบขั้นตอนการแปลงข้อมูลแล้ว เริ่มต้นทำการสร้างแบบจำลองเพื่อใช้สำหรับการทำนายประเภทกลุ่มข้อมูล โดยใช้ซีบีเอ อัลกอริทึม มีขั้นตอนการทำงานดังนี้



รูปที่ 4.14 การขั้นตอนการสร้างแบบจำลองแบบซีบีเอ อัลกอริทึม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ผู้ใช้งานเลือกข้อมูลที่จะนำมาใช้ในการสร้างแบบจำลอง โดยเลือกข้อมูลที่ต้องการ
2. กดปุ่ม [>>] เพื่อเลือกข้อมูลที่จะใช้สำหรับการทำเหมืองข้อมูล
   [>]
3. แสดงข้อมูลที่เลือกสำหรับขั้นตอนการทำเหมืองข้อมูล

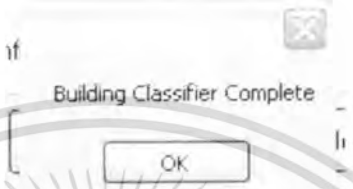


รูปที่ 4.15 ขั้นตอนการสร้างแบบจำลองแบบซีบีเอ อัลกอริทึม เลือกข้อมูลสำหรับทำเหมืองข้อมูล

4. ผู้ใช้งานคลิกปุ่ม OK เพื่อยืนยันการเลือกข้อมูล
5. ผู้ใช้งานกำหนดข้อมูล Condset และข้อมูล Class ถ้าต้องการให้ข้อมูลใดเป็น Class ให้ทำการคลิกเลือกข้อมูลแล้วกดปุ่มลูกศรไปทางขวา โดยสามารถเลือกได้เพียง 1 แอตทริบิวท์เท่านั้น และข้อมูลจะถูกแสดงอยู่ในช่อง Class ส่วนข้อมูลที่เหลือก็จะเป็นข้อมูลแบบ condset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. ผู้ใช้งานกำหนดค่า Minimum Support และ Minimum Confidence โดยทั้งค่า Minimum Support และ Minimum Confidence จะต้องกำหนดเป็นค่าเปอร์เซ็นต์
7. ผู้ใช้งานคลิกปุ่ม Build Classifier เพื่อสร้างแบบจำลอง
8. ระบบแสดงกล่องข้อความว่า Build Classifier Complete



รูปที่ 4.16 แสดงข้อความในกล่องข้อความสร้างแบบจำลองเรียบร้อย

4.4.6 ขั้นตอนการแสดงแบบจำลองและการประเมินความแม่นยำของแบบจำลอง

หลังได้ผ่านกระบวนการสร้างแบบจำลองแล้วจะได้แบบจำลองออกมา พร้อมด้วยเปอร์เซ็นต์แสดงความถูกต้องของแบบจำลอง ดังนี้

รูปที่ 4.17 แสดงผลลัพธ์แบบจำลองที่ถูกสร้างขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.7 ขั้นตอนการเรียงลำดับแบบจำลองโดยเรียงลำดับตามค่าความเชื่อมั่นและค่าสนับสนุน

1. ผู้ใช้งานกดปุ่ม Sorting by Conf แล้วระบบแสดงผลการเรียงลำดับแบบจำลองตามค่าความเชื่อมั่นสูงสุด

2. เมื่อผู้ใช้งานกดปุ่ม Sorting by Sup แล้วระบบแสดงผลการเรียงลำดับแบบจำลองตามค่าสนับสนุนสูงสุด

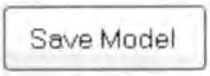
รูปที่ 4.18 แสดงผลแบบจำลองที่ถูกจัดเรียงตามค่าความเชื่อมั่น

รูปที่ 4.19 แสดงผลแบบจำลองที่ถูกจัดเรียงตามค่าสนับสนุน

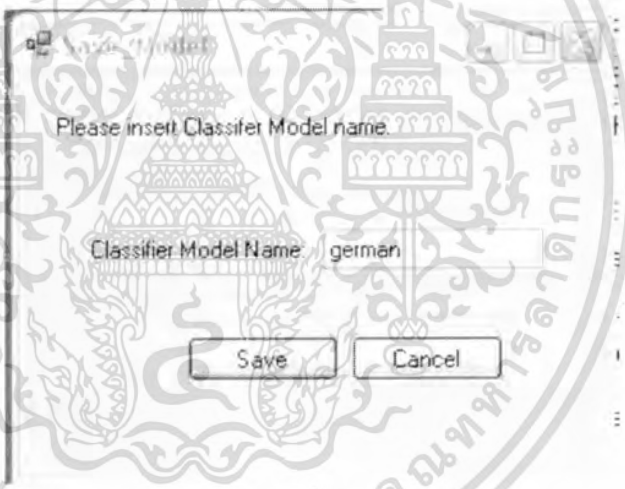
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.8 ขั้นตอนการบันทึกแบบจำลอง

หลังจากได้สร้างแบบจำลองแสดงออกมาแล้วขั้นตอนต่อไป ผู้ใช้งานจะต้องทำการบันทึกแบบจำลองลงในรูปแบบฐานข้อมูลเพื่อผู้ใช้งานคนครั้งต่อไปสามารถเรียกดูแบบจำลองได้

1. ผู้ใช้งานกดปุ่ม 

เพื่อบันทึกแบบจำลอง
2. แสดงหน้าจอเพื่อให้ผู้ใช้งานพิมพ์ชื่อ Classifier Model Name เพื่อทำการบันทึก
3. ผู้ใช้งานกดปุ่ม Save
4. ระบบแสดงข้อความ บันทึกชื่อ Classifier Model เรียบร้อยแล้ว



รูปที่ 4.20 ฟอร์มการบันทึกชื่อ Classifier Model Name

รูปที่ 4.21 แสดงข้อความยืนยันว่าบันทึกแบบจำลองเรียบร้อย

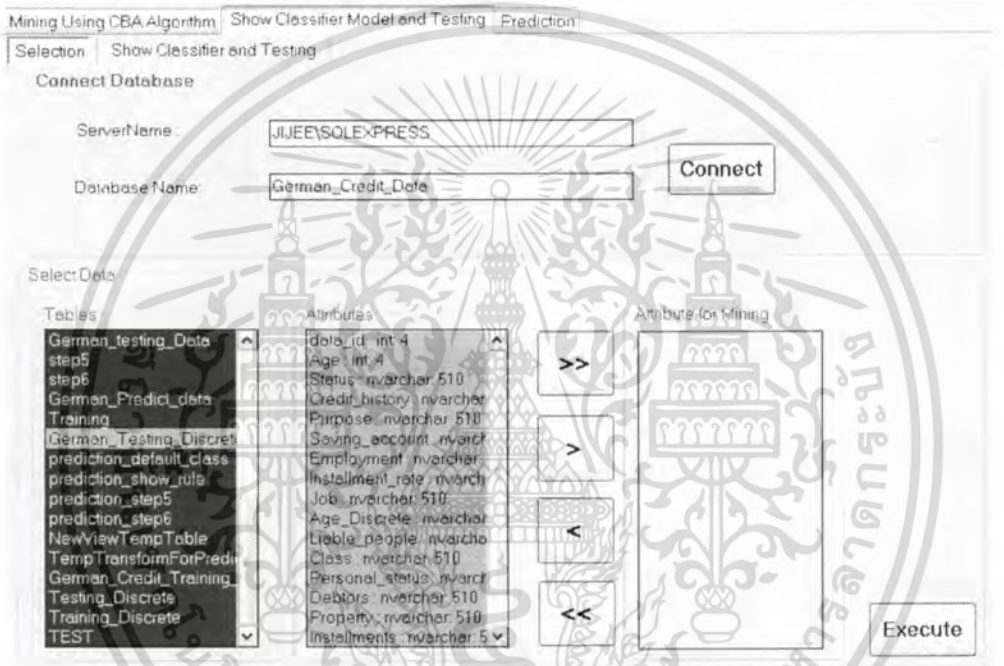
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.9 ขั้นตอนการทดสอบความแม่นยำของแบบจำลอง

เป็นขั้นตอนการทดสอบความแม่นยำของแบบจำลองด้วยข้อมูลทดสอบระบบ มีขั้นตอนการทำงานดังนี้

1. ผู้ใช้กดปุ่ม

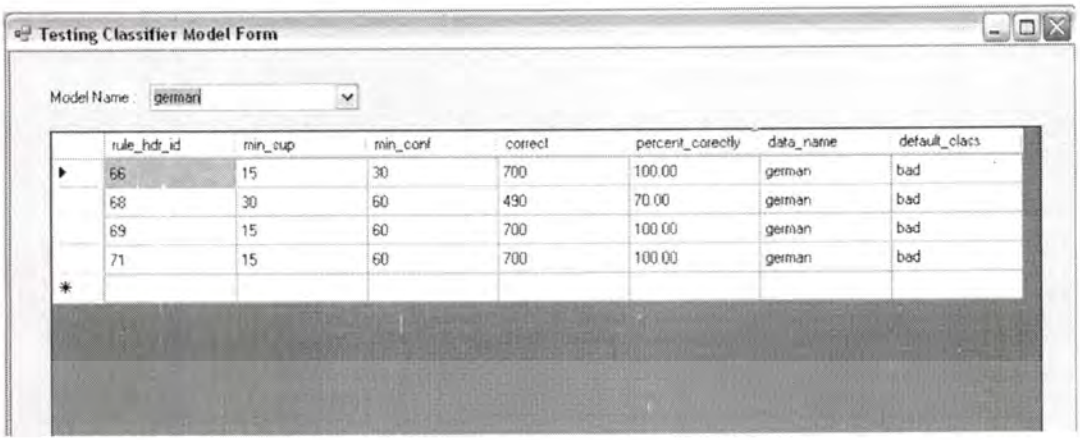
2. เข้าสู่หน้าจอการเลือกข้อมูลสำหรับการทดสอบระบบ



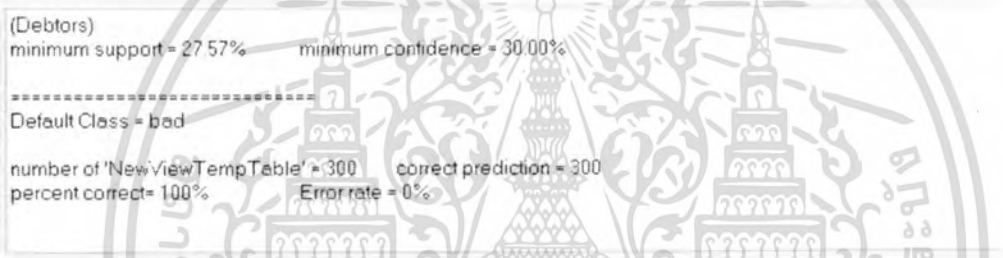
รูปที่ 4.22 หน้าจอการเลือกข้อมูลสำหรับการทดสอบระบบ

3. เลือกตารางและแอตทริบิวท์สำหรับข้อมูลทดสอบระบบ
4. แสดงแบบจำลองและพิจารณาเลือกจากค่าสนับสนุนและค่าความเชื่อมั่นต่ำ ที่มีเปอร์เซนต์ความถูกต้องจากการประเมินสูงสุด
5. แสดงผลลัพธ์การทดสอบแบบจำลองจากข้อมูลทดสอบระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.23 แสดงแบบจำลองโดยพิจารณาจากค่าสนับสนุนและค่าความเชื่อมั่นต่ำสุด



รูปที่ 4.24 แสดงผลลัพธ์การทดสอบแบบจำลองจากข้อมูลทดสอบระบบ

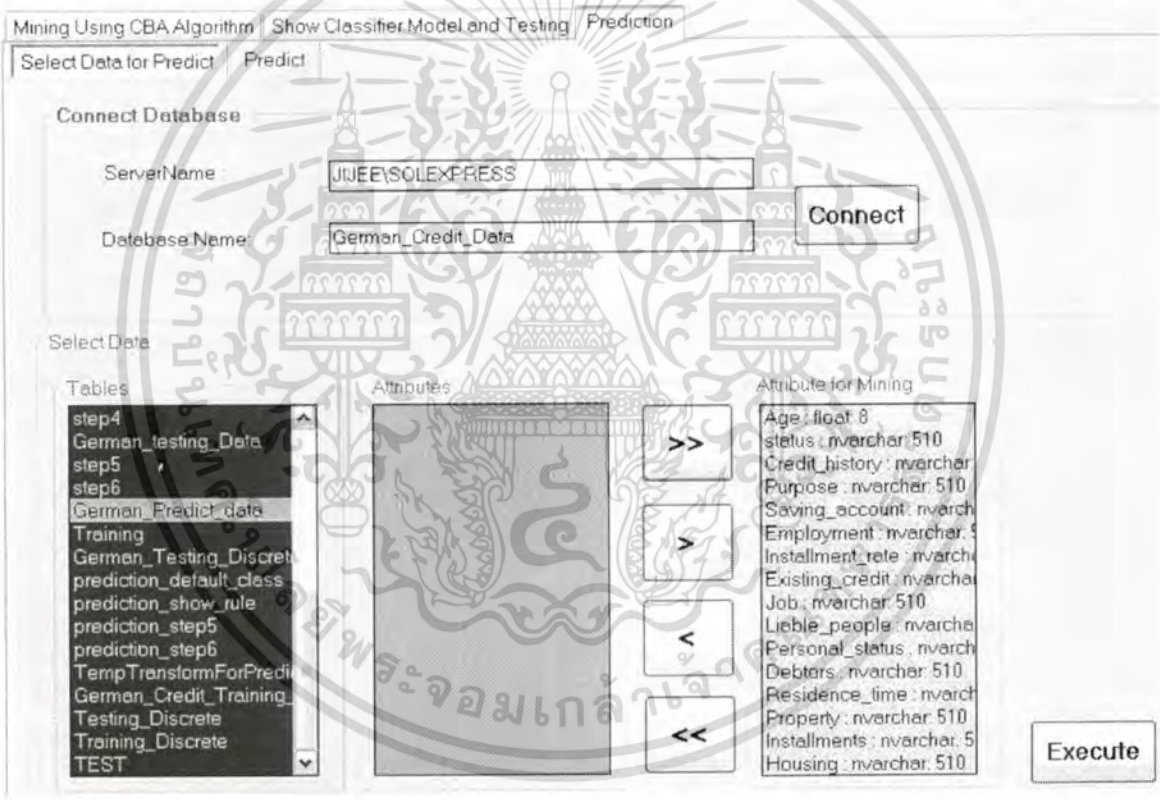
รูปที่ 4.25 เปรียบเทียบการประเมินความถูกต้องของแบบจำลองและการทดสอบความแม่นยำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.10 ขั้นตอนการทำนายกลุ่มข้อมูลจากแบบจำลองแบบ ซีบีเอ อัลกอริทึม

ขั้นตอนการทำนายกลุ่มข้อมูลใหม่ จะสามารถกระทำได้ก็ต่อเมื่อจะต้องข้อมูลใหม่เข้าสู่ระบบและมีแบบจำลองของระบบแล้ว ระบบสามารถทำนายกลุ่มข้อมูลตามขั้นตอน ดังนี้

1. ผู้ใช้งานเลือกเมนู PredictData
2. เข้าสู่หน้าจอการเลือกข้อมูลจากฐานข้อมูลสำหรับใช้ในการทำนาย
3. ทำการติดต่อเครื่องเซิร์ฟเวอร์และชื่อฐานข้อมูลเพื่อดึงข้อมูล
4. ทำการเลือกข้อมูลสำหรับการทำนาย

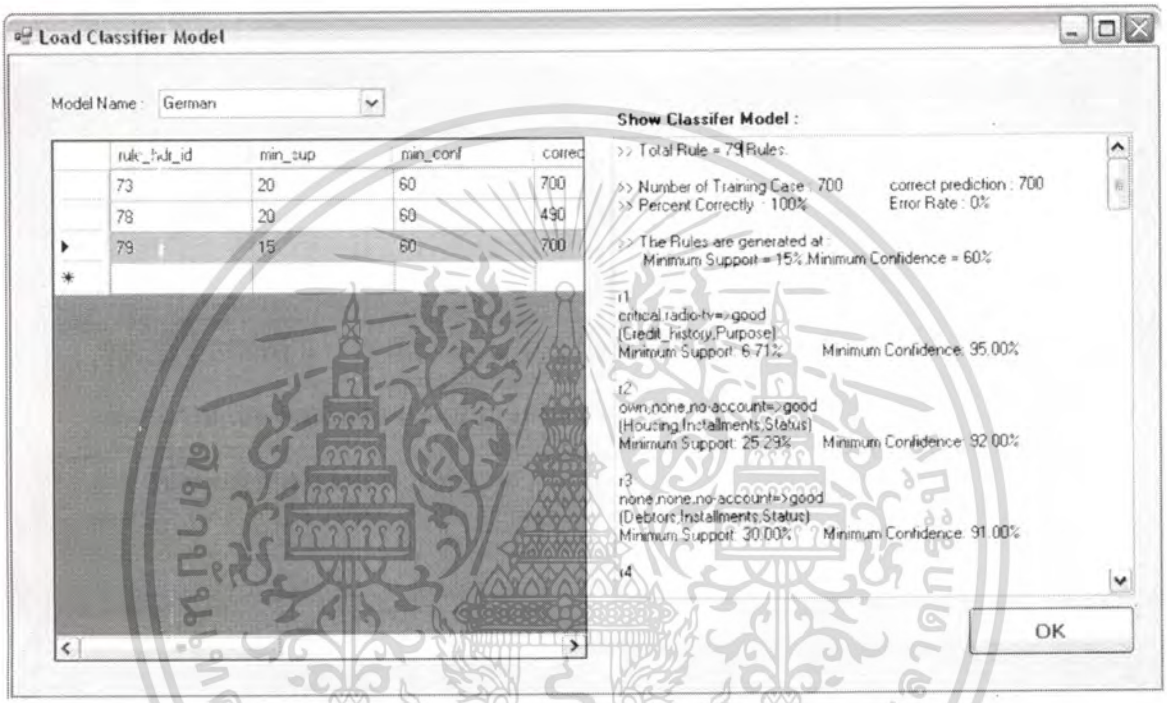


รูปที่ 4.26 ขั้นตอนการเลือกข้อมูลจากฐานข้อมูลสำหรับการทำนายกลุ่มข้อมูล

5. แสดงฟอร์ม ClassifierModel พิจารณาเลือก Classifier Model จากค่าเปอร์เซนต์ความถูกต้องจากการประเมินแบบจำลองและจากการทดสอบความแม่นยำที่กำหนด
6. แสดงหน้าจอการ PredictData สำหรับการทำนายข้อมูล
7. เลือกข้อมูลจากแอตทริบิวท์เพื่อเพิ่มลงในรายการ โดยกดปุ่ม Add >

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

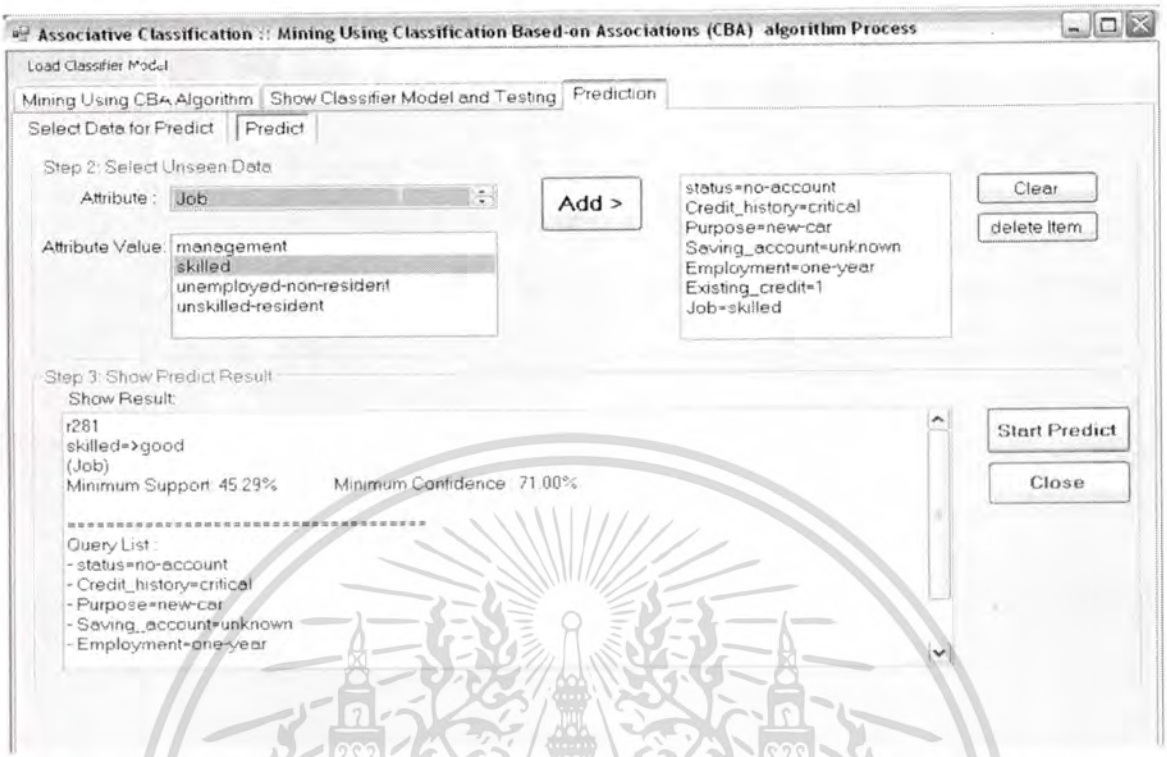
8. แสดงผลรายการข้อมูลที่เลือกแบบ 1 เรคคอร์ด
9. ผู้ใช้งานกดปุ่ม Start Predict เพื่อทำนายกลุ่มข้อมูลให้กับข้อมูลใหม่
10. ระบบแสดงผลลัพธ์จำแนกกลุ่มข้อมูลของข้อมูลใหม่ โดยพิจารณาจากกฎที่มีค่าความเชื่อมั่นสูงสุด ดังแสดงรูปที่ 4.29



รูปที่ 4.27 หน้าจอ ClassifierModel แสดงผลแบบจำลองที่เลือก

รูปที่ 4.28 หน้าจอการทำนายข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



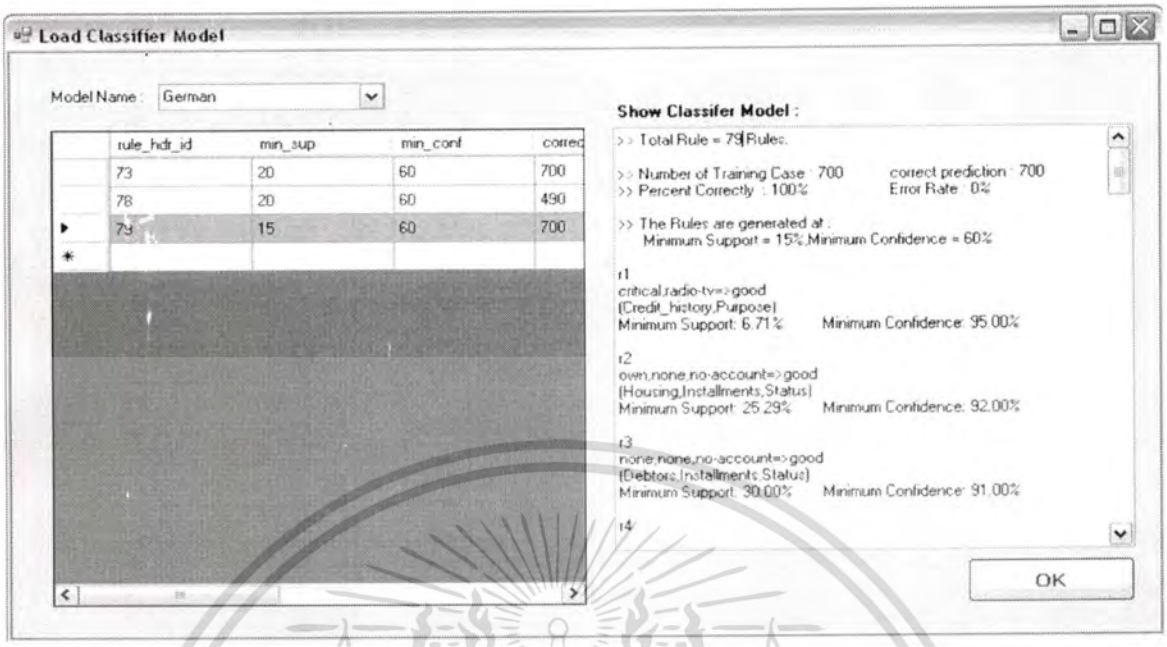
รูปที่ 4.29 แสดงผลลัพธ์การทำนายข้อมูล

4.4.10 ขั้นตอนการเรียกดูแบบจำลอง

ขั้นตอนการเรียกดูแบบจำลองสามารถกระทำได้ก็ต่อเมื่อมีการสร้างและบันทึกแบบจำลองลงในรูปแบบฐานข้อมูลเรียบร้อยแล้ว โดยมีขั้นตอนการทำงานดังนี้

1. ผู้ใช้งานเลือกเมนู Load Classifier Model เพื่อเรียกดูแบบจำลอง
2. แสดงแบบจำลองและแสดงรายละเอียดแบบจำลองที่เลือก ดังรูปที่ 4.30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.30 แสดงแบบจำลองโดยพิจารณาค่าจากความถูกต้องของแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การพัฒนา ระบบการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์แบบ ซีบีเอ อัลกอริทึม จัดทำขึ้นเพื่อนำเสนอให้เห็นถึงประโยชน์ของการนำทฤษฎีการทำเหมืองข้อมูลมาใช้เพื่อเพิ่มประสิทธิภาพในส่วนของการจำแนกประเภทข้อมูลให้มีความแม่นยำในการทำนายมากกว่าเทคนิคการจำแนกประเภทข้อมูลแบบทั่วไป โดยได้นำเทคนิคการค้นหากฎความสัมพันธ์เข้ามาประยุกต์ใช้ในส่วนของการสร้างกฎความสัมพันธ์ แต่ส่วนนี้แตกต่างจากการสร้างกฎความสัมพันธ์แบบทั่วไปๆ คือเป็นกฎความสัมพันธ์แบบมีคลาสปลายทางติดมาด้วย และสามารถสร้างแบบจำลองได้จากกฎความสัมพันธ์แบบมีคลาสปลายทางที่ค้นหาได้มาจากขั้นตอนที่ผ่านมา ซึ่งแบ่งส่วนการทำงานหลักๆออกเป็น 2 ส่วน คือ ส่วนแรก ส่วนของการค้นหากฎความสัมพันธ์ (Rule Generator) และ ส่วนที่สองคือส่วนของการสร้างแบบจำลอง (Building a Classifier) ด้วยกระบวนการดังกล่าวทำให้เทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์มีประสิทธิภาพและมีความแม่นยำมากยิ่งขึ้น และสามารถนำแบบจำลองที่ได้ไปประยุกต์ใช้ในการทำงานได้

5.1 สรุปผลการพัฒนาระบบ

จากการพัฒนาระบบทำให้ผู้พัฒนาเข้าใจถึงหลักการทำงานของการทำเหมืองข้อมูล เพื่อสามารถนำไปใช้ประโยชน์ในการทำนาย ผลการทดลองนั้นมีความถูกต้องในระดับหนึ่ง ซึ่งการทำงานของการทำเหมืองข้อมูลต้องอาศัยการเตรียมข้อมูลที่ดี ซึ่งถ้าในขั้นตอนดังกล่าวกระทำไม่ดีแล้ว จะส่งผลกระทบต่อการทดสอบให้ได้ผลในระดับที่ไม่ดีเท่าทีควร ทั้งนี้ผลลัพธ์ที่ได้ออกมาจะเกิดประโยชน์ได้ต้องอาศัยกระบวนการต่างๆตามที่ได้กล่าวมาแล้วข้างต้น ถ้าการทำงานขั้นตอนใดขั้นตอนหนึ่งผิดพลาดไปก็จะส่งผลให้ได้ผลลัพธ์ที่ผิดพลาดตามไปด้วย

5.2 ปัญหาที่พบในการพัฒนา

จากการศึกษา วิเคราะห์ ออกแบบ และพัฒนาระบบ พบว่าโครงงานพัฒนาระบบนี้มีปัญหาเรื่องขั้นตอนการศึกษาการทำงานของอัลกอริทึม การเตรียมข้อมูล ทำให้ต้องใช้เวลาค่อนข้างนาน โดยมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1 ปัญหาที่พบ

1. ปัญหาในการใช้เวลาในการศึกษาการทำงานของเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์(Associative Classification) ค่อนข้างนาน เนื่องจากเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์เป็นวิธีการสร้างแบบจำลองแบบใหม่ ซึ่งเป็นการนำเทคนิคการค้นหากฎความสัมพันธ์(Association Rule Discovery) เข้ามาประยุกต์ใช้ทำงานร่วมกับงานการจำแนกประเภทข้อมูล คือ นำเทคนิคการค้นหากฎความสัมพันธ์เข้ามาประยุกต์ทำงานในส่วนของขั้นตอนการสร้างกฎความสัมพันธ์ แต่เป็นการสร้างกฎความสัมพันธ์แบบมีคลาส(CARs) ซึ่งแตกต่างจากการสร้างแบบจำลองในงานจำแนกประเภทข้อมูลทั่วไป ดังนั้นการค้นคว้าตัวอย่างงานวิจัยของเทคนิคนี้จึงมีไม่แพร่หลายมากนัก ทำให้ผู้พัฒนาระบบต้องใช้เวลาในการศึกษาทำความเข้าใจถึงขั้นตอนการทำงานของอัลกอริทึมค่อนข้างนาน

2. ปัญหาการเตรียมข้อมูล เนื่องจากในกระบวนการทำเหมืองข้อมูล มีหน้าที่คือ การสืบค้นความรู้ ความสัมพันธ์ และรูปแบบทั้งหมดที่ถูกซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ออกมา เพื่อช่วยในการวิเคราะห์ข้อมูล และช่วยในการตัดสินใจในการทำงาน ดังนั้นการที่จะสามารถเตรียมข้อมูลให้มีความถูกต้องตามรูปแบบที่กำหนดไว้ มีรูปแบบข้อมูลที่ถูกต้องไม่มีความผิดพลาดนั้น เป็นเรื่องที่ค่อนข้างยากเพราะว่าข้อมูลสำหรับการทำเหมืองข้อมูลเป็นข้อมูลจำนวนมาก ดังนั้นโครงงานการพัฒนาระบบนี้ได้ใช้ชุดข้อมูลทดลองสำหรับการทำเหมืองข้อมูลสำหรับศึกษาการทำงานของอัลกอริทึม ซีบีเอ ทางด้านการทำเหมืองข้อมูลเท่านั้น ซึ่งอาจส่งผลให้ข้อมูลไม่สามารถเป็นตัวแทนทางสถิติได้ทั้งหมด

5.2.2 ข้อเสียของการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นเทคนิคหนึ่งที่ช่วยให้การทำงานรวดเร็วและมีประสิทธิภาพมากขึ้น การที่จะสามารถสร้างแบบจำลองเพื่อช่วยในการตัดสินในการทำงานนั้น จำเป็นที่ผู้ใช้งานจะต้องมีความรู้ทั้งในเรื่องของตัวข้อมูลนั้นๆเป็นอย่างดี รวมไปถึงเรื่องของการกำหนดกฎเกณฑ์การทำงานต่างๆ การเลือกอัลกอริทึมที่เหมาะสมกับวัตถุประสงค์หรือปัญหาที่ต้องการใช้เทคนิคการทำเหมืองข้อมูลนำมาแก้ปัญหาได้เป็นอย่างดี จึงจะช่วยให้การนำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้กับการทำงานมีประสิทธิภาพมากขึ้น เมื่อได้แบบจำลองออกมาแล้วจึงจำเป็นจะต้องใช้ผู้ที่มีความชำนาญงานตัดสินใจอีกครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการทำเหมืองข้อมูลไม่มีเทคนิคใดที่จะสามารถแก้ปัญหาของการทำเหมืองข้อมูลได้ทุกปัญหา ความหลากหลายของเทคนิคจึงเป็นสิ่งที่จำเป็นในการนำไปสู่วิธีการแก้ปัญหาของการทำเหมืองข้อมูลได้ดีที่สุด ทั้งนี้ขึ้นอยู่กับความหลากหลายของเทคนิคในการนำไปประยุกต์ใช้ และขึ้นอยู่กับการกำหนดวัตถุประสงค์หรือช่วยแก้ปัญหาในเรื่องใดเป็นสำคัญ และเทคนิคการทำเหมืองข้อมูลเป็นเพียงเครื่องมือช่วยผู้ใช้ในการวิเคราะห์และช่วยในการตัดสินใจให้การทำงานเกิดความรวดเร็วและมีประสิทธิภาพมากขึ้นเท่านั้น การวิเคราะห์และการตัดสินใจจะเกิดประสิทธิภาพสูงสุดต้องมาจากการตัดสินใจของผู้ใช้งานที่เป็นมนุษย์ซึ่งมีความรู้ความเข้าใจเรื่อง

5.3 ปัญหา ข้อจำกัด และข้อเสนอแนะ

เทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์เป็นวิธีการสร้างแบบจำลองแบบใหม่ซึ่งมีความแม่นยำในการทำนายข้อมูลสูงกว่าเทคนิคการจำแนกประเภทข้อมูลทั่วไป และสามารถทำนายคลาสได้อย่างรวดเร็ว เพราะมีการสร้างแบบจำลองของกฎความสัมพันธ์แบบมีคลาสเอาไว้ก่อน ถึงแม้ข้อมูลที่นำมาใช้จะมีความซับซ้อนและมีความหลากหลายของคลาสก็ตาม แต่ก็สามารถที่จะให้เหตุผลในการทำนายข้อมูลได้โดยดูจากกฎความสัมพันธ์ที่ใช้ในการทำนาย ดังนั้นผู้ใช้งานสามารถทำความเข้าใจกับผลการทำนายได้ง่าย แต่เนื่องจากการสร้างแบบจำลองแบบมีคลาสเอาไว้ก่อน เมื่อมีการเพิ่มของข้อมูล จึงต้องทำการปรับปรุงในส่วนของการค้นหากฎความสัมพันธ์ และแบบจำลองใหม่ทุกครั้ง ดังนั้นจึงไม่เหมาะกับข้อมูลที่มีการเปลี่ยนแปลงปริมาณมาก(Incremental Data) และเนื่องจากเทคนิคการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ ในส่วนของการสร้างแบบจำลอง (Classifier a Building) จะมีอัลกอริทึมการสร้างแบบจำลองด้วยกัน 2 รูปแบบคือ อัลกอริทึมแบบ M1 ซึ่งเป็นอัลกอริทึมที่เหมาะกับข้อมูลขนาดเล็ก ประสิทธิภาพการประมวลผลค่อนข้างช้า ถ้ามีจำนวนข้อมูลปริมาณมากอาจจะต้องใช้อัลกอริทึมแบบ M2 ซึ่งมีประสิทธิภาพในการทำงานดีกว่าแบบ M1

ส่วนของโปรแกรมระบบยังมีข้อจำกัดในส่วนของการเตรียมข้อมูลคือ ส่วนของฟังก์ชันการคลีนข้อมูล (Clean Data) เนื่องจากผู้พัฒนามีระยะเวลาจำกัดในการพัฒนาระบบทำให้ยังขาดความครบถ้วนในรายละเอียดของบางฟังก์ชันคือ ฟังก์ชันการคลีนข้อมูลซึ่งมีวิธีการคลีนข้อมูลแค่เพียงวิธีการเดียวคือ การลบเรคคอร์ดที่มีค่าว่าง (Null) ทิ้ง ซึ่งอาจส่งผลให้ข้อมูลเกิดความผิดพลาดได้ ควรเพิ่มวิธีการในฟังก์ชันการคลีนข้อมูล ได้แก่ วิธีการเติมค่าที่ขาดหายไป เช่น เติมค่าฐานนิยม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Mode) ลงในข้อมูลที่หายไป เติม unknown ลงในฐานข้อมูลที่ขาดหายไป วิธีการกำหนดค่าที่เกินขอบเขต เพื่อลดความคลาดเคลื่อน และวิธีการทำให้ข้อมูลตรงกัน เป็นต้น อีกทั้งในส่วนของฟังก์ชันการแปลงข้อมูล(Transform Data) วิธีการทำงานในระบบทำได้เพียงข้อมูลที่เป็นแบบ continous ผู้พัฒนาเห็นว่าควรเพิ่มประเภทข้อมูลสำหรับขั้นตอนนี้ คือ ควรพัฒนาระบบให้สามารถแปลงข้อมูลได้ทุกประเภทข้อมูล ทั้งนี้การแปลงข้อมูลต้องรองรับและขึ้นอยู่กับวัตถุประสงค์การนำข้อมูลนั้นไปใช้งานด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

ดวงกมล มหาวีระ. 2546. "ระบบการพัฒนาระบบการค้นหากฎความสัมพันธ์โดยใช้ Apriori Algorithm." โครงการพัฒนาระบบงานวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

ธีระวัฒน์ พงษ์ศิริปรีดา. 2544. "การจัดสรรกฎหมายที่เหมาะสมให้กับคดีความโดยอัตโนมัติ." วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, มหาวิทยาลัยเกษตรศาสตร์.

สุรสิทธิ์ คิวประสพศักดิ์. 2546. **อินไซท์ Visual Basic .NET ฉบับสมบูรณ์.** กรุงเทพฯ: โปรวิชัน.

อาทิตยา เชื้อจันอัด. 2549. "การพัฒนาระบบการเตรียมข้อมูลและการสำรวจ สำหรับการทำดาต้าไมนิ่ง." โครงการพัฒนาระบบงานวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

Agrawal, Rakesh and Srikant, Ramakrishnan. 1994. **"Fast Algorithm for Mining Association Rules in Large Databases."** [Online] Available: http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf

Liu, Bing and Hsu,Wynne and Ma,Yiming. 1998. **"Integrating Classification and Association Rule Mining."** [Online] Available: http://www.aaai.org/Papers/KDD/1998/KDD98-012.pdf

Quinlan, JR. 1993. **C4.5: Programs for Machine Learning.** Morgan Kaufmann.

Wasilewska, Anita. **Data Mining Concepts and Techniques.** [Online] Available: http://www.cs.sunysb.edu~cse634presentationsCSE643Association%20Web%20Mining%Group%206.ppt

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน	นางสาว วัลภรณ์ เอตะนาม
วันเดือนปีเกิด	7 พฤศจิกายน 2525
ที่อยู่	176 ม.5 ต.บางมะเดื่อ ถ.สุราษฎร์-ตะกั่วป่า
	อ.พุนพิน จ.สุราษฎร์ธานี 84130
วุฒิการศึกษาระดับปริญญาตรี	สาขา สารสนเทศศึกษา
	สำนักวิชาสารสนเทศศาสตร์
	มหาวิทยาลัยวลัยลักษณ์
ปีที่สำเร็จการศึกษา	2548

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้