ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

การพัฒนาระบบการค้นหากฎความสัมพันธ์แบบอพิออริอัลกอริทึม

DEVELOPMENT OF ASSOCIATION RULE DISCOVERY SYSTEM USING APRIORI ALGORITHM

โดย

นิอร หงส์วัฒนกุล

NION HONGWATTANAKUL

อาจารย์ที่ปรึกษา

รศ.ดร.วรพจน์ กรีสุระเดช

QW
4697ก
2549

H004423

เลขหมู่.....
เลขทะเบียน 04423
วัน,เดือน,ปี - 5 ส.ค. 2551

b.11922314
i.....

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคฤดูร้อน ปีการศึกษา 2549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

DEVELOPMENT OF ASSOCIATION RULE DISCOVERY SYSTEM USING APRIORI ALGORITHM

NION HONGWATTANAKUL

A SYSTEM DEVELOPMENT PROJECT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

SUMMER/2006

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2007

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การพัฒนาระบบการค้นหากฎความสัมพันธ์แบบอพิออริอัลกอริทึม
นักศึกษา	นางสาวนิอร หงส์วัฒนกุล
รหัสนักศึกษา	47066432
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2549
อาจารย์ที่ปรึกษา	รศ.ดร.วรพจน์ กรีสุระเดช

บทคัดย่อ

การพัฒนาระบบค้นหากฎความสัมพันธ์ของข้อมูล โดยใช้การทำเหมืองข้อมูลแบบกฎความสัมพันธ์ เพื่อหาผลลัพธ์ความสัมพันธ์ของข้อมูลที่ผ่านกระบวนการทำเหมืองข้อมูล และนำผลลัพธ์ที่ได้มาทำนายหาสาเหตุปัจจัยหลัก และปัจจัยรองอื่นๆ เพื่อช่วยในการตัดสินใจในการดำเนินธุรกิจ ซึ่งผลลัพธ์ที่ได้จะช่วยให้สามารถอธิบายเหตุผลการตัดสินใจให้กับผู้ที่เกี่ยวข้องให้เข้าใจได้ง่ายขึ้น เนื่องจากการดำเนินงานธุรกิจในปัจจุบันได้มีการแข่งขันกันอย่างสูง โดยเฉพาะการแข่งขันทางด้านข้อมูล ซึ่งหากองค์กรใดมีข้อมูลในการดำเนินงานที่ดีกว่าก็จะได้เปรียบในการดำเนินธุรกิจ

โครงงานนี้จึงได้เสนอกระบวนการทำเหมืองข้อมูล โดยอัลกอริทึมที่นำมาใช้ในการทำเหมืองข้อมูลคือ อพิออริอัลกอริทึม เพื่อหาความสัมพันธ์ของข้อมูล ซึ่งได้นำระบบการเตรียมข้อมูลและการสำรวจข้อมูลรวมกับการหากฎความสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Development of Association Rule Discovery System Using Apriori Algorithm
Student	Miss Nion Hongwattanakul
Student ID.	47066432
Degree	Master of Science
Programme	Information Science
Academic Year	2006
Advisor	Associate Professor Dr. Worapoj Kreesuradej

ABSTRACT

Development of Association Rule Discovery System by use association rule of data mining process for searching relationship of information that are used data mining process and take the results to predict for searching main factor and others factor for making a decision in business process so the results can describe decision reason to stakeholder understand easily. Since the present, business process has highly competition specially competition about information so in case of others organization have good information for business process, also it have advantage over in business process.

This project propose data mining process by use apriori algorithm for searching relationship of information so to combine data preparation and exploration system with association rule discovery system.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงงานฉบับนี้ได้รับความกรุณาจากอาจารย์ที่ปรึกษา รศ.ดร.วรพจน์ กรีสุระเดช ที่ได้ให้คำชี้แนะ และให้คำปรึกษาตลอดจนช่วยแก้ปัญหาด้านต่างๆ ข้าพเจ้าขอขอบพระคุณอย่างสูงในความอนุเคราะห์ของท่านอาจารย์

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาวิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกคนที่ให้คำแนะนำต่างๆ และคอยให้กำลังใจเสมอมา

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณบิดา มารดา และครอบครัวที่ได้ให้กำลังใจ ให้การสนับสนุนในทุกๆ เรื่องที่ทำให้โครงงานฉบับนี้สำเร็จลุล่วงเป็นอย่างดี

นิอร หงส์วัฒนกุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์และประโยชน์ที่คาดว่าจะได้รับ 1

1.3 ขอบเขตของระบบ 1

1.4 ขั้นตอนการดำเนินงาน 2

บทที่2 ทฤษฎีที่เกี่ยวข้อง 3

2.1 ความหมายของการทำเหมืองข้อมูล 3

2.2 ขั้นตอนการทำเหมืองข้อมูล 4

2.3 การทำเหมืองข้อมูล 5

บทที่3 การค้นหากฎความสัมพันธ์ (Association Rule Discovery) 7

3.1 กฎความสัมพันธ์ (Association Rule) 7

3.1.1 การคิดหาค่าของกฎความสัมพันธ์ 7

3.2 อัลกอริทึม 8

บทที่4 วิเคราะห์ และออกแบบระบบ 15

4.1 ความต้องการ และข้อจำกัดของโปรแกรม 15

4.1.1 ความต้องการของผู้ใช้งาน 15

4.1.2 ความต้องการเชิงฟังก์ชัน 15

4.1.3 ความต้องการที่ไม่ใช่เชิงฟังก์ชัน 16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2 การวิเคราะห์ระบบโดยใช้ยูเคสไดอะแกรม และแอคทิวิตีไดอะแกรม 16
4.2.1 ยูสเคสไดอะแกรม 16
4.2.2 รายละเอียดแอคเตอร์ 16
4.2.3 รายละเอียดยูสเคส 17
4.2.4 แอคทิวิตีไดอะแกรม 24
4.3 การออกแบบระบบโดยใช้คลาสไดอะแกรม และซีเควนไดอะแกรม 28

บทที่5 ออกแบบ และพัฒนาระบบ 40
5.1 เครื่องมือที่ใช้ในการออกแบบ และพัฒนาระบบ 40
5.1.1 ฮาร์ดแวร์ 40
5.1.2 ซอฟต์แวร์ 40
5.2 สถาปัตยกรรมระบบ 40
5.3 การออกแบบหน้าจอการรับข้อมูล 42
5.4 การออกแบบหน้าจอรายงานผลลัพธ์ 42
5.5 การใช้งานโปรแกรม 43
5.5.1 ขั้นตอนการติดต่อฐานข้อมูล 43
5.5.2 ขั้นตอนการเลือกข้อมูล 43
5.5.3 ขั้นตอนการคลีนข้อมูล 45
5.5.4 ขั้นตอนการแปลงข้อมูล 46
5.5.5 ขั้นตอนการสร้างกฎความสัมพันธ์ 48
5.5.6 ขั้นตอนการแสดงกฎความสัมพันธ์ผลลัพธ์ 51

บทที่6 สรุปและข้อเสนอแนะ 53
6.1 สรุปผลการพัฒนาระบบ 53
6.2 ปัญหาที่พบในการพัฒนา 53
6.2.1 ปัญหาที่พบ 53
6.2.2 ข้อเสียของการทำเหมืองข้อมูล 53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

6.3 ข้อเสนอแนะ 54
บรรณานุกรม 55
ประวัติผู้เขียน 56

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

3.1 ทรานส์แอ็กชันของข้อมูลสำหรับร้านขายสินค้า 11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
2.1	กระบวนการทำเหมืองข้อมูล	4
3.1	อัลกอริทึมการหา frequent itemset	9
3.2	อัลกอริทึมขั้นตอนการ Join	10
3.3	อัลกอริทึมขั้นตอนการ Prune	10
3.4	Itemset และ การหา frequent itemset โดยกำหนดให้ minimum support = 2	13
4.1	ยูเคสระบบการค้นหาความสัมพันธ์แบบอพิออริอัลกอริทึม	16
4.2	ยูสเคสการเลือกข้อมูล	18
4.3	ยูสเคสการคลีนข้อมูล	19
4.4	ยูสเคสการแปลงข้อมูล	20
4.5	ยูสเคสการสำรวจข้อมูล	21
4.6	ยูสเคสการสร้างกฎความสัมพันธ์ของข้อมูล	22
4.7	ยูสเคสเรียงกฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล	23
4.8	ยูสเคสค้นหากฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล	24
4.9	แอคทิวิตีไดอะแกรมการเตรียมข้อมูล	25
4.10	แอคทีวิตีไดอะแกรมสร้างกฎความสัมพันธ์ของข้อมูล	26
4.11	แอคทิวิตีไดอะแกรมเรียงกฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล	27
4.12	แอคทิวิตีไดอะแกรมค้นหากฎความสัมพันธ์การทำเหมืองข้อมูล	27
4.13	คลาสไดอะแกรมการค้นหากฎความสัมพันธ์ (High Level)	28
4.14	คลาสไดอะแกรมระบบการค้นหากฎความสัมพันธ์	29
4.15	ซีเควนไดอะแกรมการหากฎความสัมพันธ์	30
4.16	คลาสไดอะแกรม Datamining	31
4.17	คลาสไดอะแกรมApriori	32
4.18	คลาสไดอะแกรม Itemset	33
4.19	คลาสไดอะแกรม ItemsetArrayList	34
4.20	คลาสไดอะแกรม Groupinh table	35
4.21	คลาสไดอะแกรม Class Information	36
4.22	คลาสไดอะแกรม View Data	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่		หน้า
4.23	คลาสไดอะแกรม Data	37
4.24	คลาสไดอะแกรม Database	38
4.25	คลาสไดอะแกรม Northwind Data Control	39
4.26	คลาสไดอะแกรม ProgressBar Event Handle	40
5.1	สถาปัตยกรรมระบบ	42
5.2	หน้าจอการรับข้อมูล	43
5.3	หน้าจอแสดงผลลัพธ์การประมวลผล	43
5.4	ขั้นตอนการติดต่อฐานข้อมูล	44
5.5	กล่องข้อความแสดงการผลการเชื่อมต่อฐานข้อมูล	44
5.6	ขั้นตอนการเลือกข้อมูลจาก 1 ตาราง	45
5.7	ขั้นตอนการเลือกข้อมูลจากหลายตาราง	45
5.8	กล่องข้อความแสดงผลการเลือกข้อมูล	46
5.9	การคลีนข้อมูล	46
5.10	กล่องข้อความแสดงผลการคลีนข้อมูล	47
5.11	ขั้นตอนการจัดกลุ่ม	47
5.12	กล่องข้อความแสดงผลการจัดกลุ่ม	48
5.13	สิ้นสุดขั้นตอนการจัดกลุ่มข้อมูล	48
5.14	กล่องข้อความแสดงผลการแปลงข้อมูล	49
5.15	ขั้นตอนการเลือกแบบจำลองที่ใช้ทำเหมืองข้อมูล	49
5.16	การสร้างกฎความสัมพันธ์	50
5.17	การจัดเก็บกฎความสัมพันธ์ผลลัพธ์	51
5.18	การเลือกตำแหน่งและกำหนดชื่อไฟล์การจัดเก็บ	51
5.19	การเรียงกฎความสัมพันธ์ผลลัพธ์	52
5.20	การค้นหากฎความสัมพันธ์ผลลัพธ์	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในสภาวะเศรษฐกิจของปัจจุบันหน่วยงานต่างๆ ได้มีการแข่งขันทางการตลาดกันอย่างสูง จึงทำให้ต้องค้นหาวิธีต่างๆ เพื่อพัฒนาหน่วยงานของตนให้มีประสิทธิภาพมากขึ้น ซึ่งส่วนสำคัญของการพัฒนาเช่น การตัดสินใจของผู้บริหาร โดยอาศัยสภาพแวดล้อมต่างๆ กล่าวคือ เมื่อมีเหตุการณ์ที่ต้องการประเมิน หรือ ตัดสินใจบางอย่าง ก็ควรที่จะสนใจสถานะทางเศรษฐกิจ ณ ขณะนั้นรวมไปถึงสิ่งสำคัญที่สุดด้วย คือ ข้อมูล และ เหตุการณ์ในอดีต ดังนั้นหากการตัดสินดังกล่าวได้ข้อมูลที่ดีและมีประสิทธิภาพจะทำให้ผลลัพธ์ที่เกิดจากการตัดสินใจของผู้บริหารก็มีประสิทธิภาพต่อหน่วยงานด้วยเช่นกัน

การพัฒนา และศึกษาโครงงานนี้จึงได้นำเทคโนโลยีการทำเหมืองข้อมูลมาใช้วิเคราะห์ข้อมูลและมุ่งเน้นอธิบายความสัมพันธ์ของกลุ่มข้อมูลที่ต้องการ เพื่อนำมาหาความสัมพันธ์ของข้อมูลเพื่อทำให้ข้อมูลที่มีอยู่ให้เกิดประโยชน์มากยิ่งขึ้น โดยได้นำระบบการเตรียมข้อมูลและการสำรวจมาทำการรวมเข้ากับการทำเหมืองข้อมูล

1.2 วัตถุประสงค์และประโยชน์ที่คาดว่าจะได้รับ

1. ศึกษาขั้นตอนและวิธีการในการทำเหมืองข้อมูลโดยมุ่งเน้นไปที่อพิออริอัลกอริทึม
2. ศึกษาเทคโนโลยีการทำเหมืองข้อมูล เพื่อนำมาใช้ออกแบบและพัฒนาโปรแกรม
3. เพิ่มมูลค่าของข้อมูลที่ถูกเก็บไว้ โดยนำมาอธิบายในรูปแบบความสัมพันธ์ เพื่อก่อให้เกิดประโยชน์สูงสุด
4. เพื่อนำผลลัพธ์ของการทำเหมืองข้อมูลที่ได้ไปใช้สนับสนุนการทำงานขององค์กร

1.3 ขอบเขตของระบบ

การทำงานของระบบการค้นหากฎความสัมพันธ์แบบอพิออริอัลกอริทึมมีขอบเขตการศึกษาและพัฒนาดังนี้

1. นำข้อมูลไปผ่านกระบวนการแปลงข้อมูล
2. สร้างอัลกอริทึมเพื่อนำกลุ่มของข้อมูลมาประมวลผล
3. นำผลลัพธ์ของการประมวลผลกลุ่มข้อมูลมาแสดงรายงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ขั้นตอนการดำเนินงาน

การพัฒนาโครงงานนี้มีการนำข้อมูลจากแหล่งข้อมูลที่น่าเชื่อถือ และต้องมีการจัดเตรียมการศึกษา ออกแบบ และพัฒนา โดยมีขั้นตอนดังนี้

1. ศึกษาทฤษฎีที่เกี่ยวข้องเกี่ยวกับการทำเหมืองข้อมูลเพื่อนำมาประยุกต์ใช้ในการพัฒนาระบบ
2. การจัดการให้ข้อมูลอยู่รูปแบบที่เหมาะสมและมีการจัดกลุ่มเปลี่ยนแปลงค่าเพื่อให้สามารถนำไปวิเคราะห์หาผลลัพธ์ได้ตรงตามวัตถุประสงค์
3. สร้างรูปแบบจำลองเพื่อหาความสัมพันธ์ของข้อมูลที่ผ่านการจัดรูปแบบให้ถูกต้องแล้ว โดยใช้กฎความสัมพันธ์และใช้อัลกอริทึมอพิออริในการวิเคราะห์เพื่อหาผลลัพธ์
4. ศึกษาข้อมูลที่ได้รับมา และนำมาหาความสัมพันธ์ระหว่างข้อมูลเพื่อนำมาจัดกลุ่มตามรูปแบบจำลองของกฎความสัมพันธ์
5. จัดเตรียมรายงานผลความสัมพันธ์ที่ผ่านการประมวลผลจากอัลกอริทึมในรูปแบบรายงานที่ผู้ใช้งานสามารถเข้าใจได้ง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่2
ทฤษฎีที่เกี่ยวข้อง

การพัฒนาโครงการนี้ เป็นการออกแบบ พัฒนาระบบ โดยอาศัยเทคโนโลยีการทำเหมืองข้อมูลเข้ามาใช้ในการทำนายหาผลลัพธ์ความสัมพันธ์กันของข้อมูลที่รับเข้ามาโดยใช้พื้นฐานอ้างอิงการทำนายข้อมูลจากความสัมพันธ์ของข้อมูลในอดีต ดังนั้นจึงต้องมีการศึกษาการทำเหมืองข้อมูลโดยมีรายละเอียดที่เกี่ยวข้องกับการพัฒนาระบบงาน ดังนี้

2.1 ความหมายของการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นส่วนประกอบหนึ่งที่สำคัญในการทำธุรกิจแบบอัฉริยะ (Business Intelligence) มีความแตกต่างจากภาษาคิวรีที่ตอบได้เพียงคำถามแบบง่ายๆ เท่านั้น เช่น ข้อมูลที่อยู่ในแถวมีอะไรบ้าง แต่ไม่สามารถตอบปัญหาที่มีความสัมพันธ์ และมีความซับซ้อนได้ ซึ่งการทำเหมืองข้อมูลมีเป้าหมายที่ชัดเจนเพื่อมุ่งเน้นไปที่การค้นหาแหล่งความรู้ใหม่ที่ถูกซ่อนไว้ในตัวข้อมูล

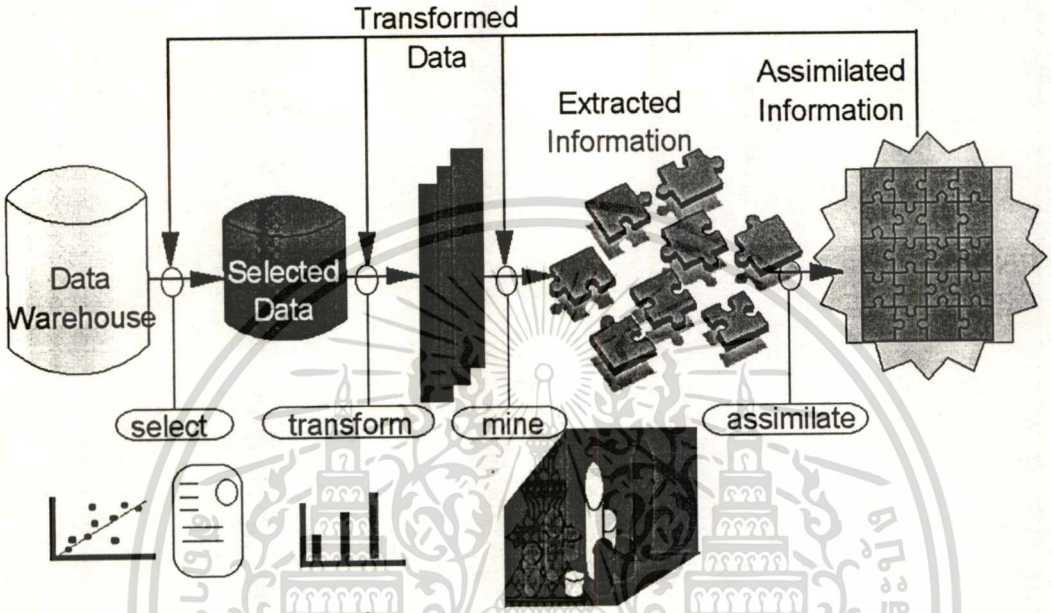
การทำเหมืองข้อมูลถูกจัดให้เป็นกระบวนการค้นหาความสัมพันธ์ในหลายๆ รูปแบบข้อมูลหรือหาสิ่งที่มีประโยชน์ และไม่เคยเจอมาก่อนจากฐานข้อมูลที่มีตารางขนาดใหญ่ เพื่อเพิ่มมูลค่าให้กับข้อมูลที่หน่วยงานนั้นๆ มีอยู่ โดยที่ข้อมูลนั้นสามารถหาได้จากหลายๆ แหล่ง เช่น ข้อมูลที่ได้จากระดับปฏิบัติการ, จากฐานข้อมูล, คลังข้อมูล มาผ่านกระบวนการจนได้ความรู้ แล้วนำความรู้ที่ได้มานั้นไปเป็นส่วนช่วยในการดำเนินงานขององค์กรหรือธุรกิจทางด้านต่างๆ ในหลากหลายสาขาอาทิ

- การนำแหล่งข้อมูลขนาดใหญ่ที่ไม่ได้ใช้งานแล้ว นำกลับมาใช้เพื่อเป็นตัวทำนายอนาคตได้
- การนำผลการทำนาย และการวิเคราะห์มาใช้เป็นกลยุทธในการเพิ่มยอดขายให้กับธุรกิจการค้า
- การทำนายการซื้อสินค้าในช่วงเวลาต่างๆ ช่วยให้สามารถนำมาออกแบบการจัดซื้อผลิตภัณฑ์ในคลังสินค้าได้ถูกต้อง และคุ้มค่ากับต้นทุนที่จ่ายไป
- การวิเคราะห์สินค้า เพื่อให้ได้สินค้าที่ตรงกับความต้องการของลูกค้า รวมไปถึงการกำหนดราคาสินค้าที่เหมาะสม
- การทำนายข้อมูลความสัมพันธ์ กับการซื้อสินค้าแต่ละชนิดของลูกค้าทำให้บริษัทสามารถกำหนดที่วางสินค้าให้เหมาะสม และเป็นการช่วยให้มีการสร้างยอดขายให้เพิ่มมากขึ้น
- อีคอมเมิร์ซ ทำการวิเคราะห์พฤติกรรมการใช้เว็บไซต์ของลูกค้าตามลำดับก่อนหลัง ซึ่งจะได้ทราบถึงสิ่งที่ลูกค้าต้องการเพื่อนำไปปรับปรุงเว็บไซต์ให้เข้ากับความต้องการของลูกค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 ขั้นตอนการทำเหมืองข้อมูล

กระบวนการทำเหมืองข้อมูลบางครั้งถูกเรียกชื่ออีกอย่างหนึ่งซึ่งมีความหมายอย่างเดียวกันว่า Knowledge discovery in database (KDD) โดยมีความหมายว่า การค้นหาความรู้ที่มีประโยชน์ และยังไม่เคยเจอมาก่อนจากฐานข้อมูล



รูปที่2.1 กระบวนการทำเหมืองข้อมูล

จากรูปที่2.1 แสดงถึงกิจกรรมของแต่ละขั้นในกระบวนการทำเหมืองข้อมูล ซึ่งแต่ละกิจกรรมสามารถทำได้มากกว่า 1 ครั้งหรือ วนกลับมาทำใหม่ได้ โดยแต่ละขั้นตอนมีดังนี้

- คัดเลือกข้อมูล (Data Selection)

  เป็นการเลือกข้อมูลหรือระบุถึงแหล่งข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล ให้ตรงกับวัตถุประสงค์ทางธุรกิจที่ได้กำหนดไว้

- ปรับเปลี่ยนรูปแบบของข้อมูล (Data Transformation)

  เป็นการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลที่เหมาะสมตรงตามอัลกอลิทึมและแบบจำลองที่เลือกใช้ทำเหมืองข้อมูลที่จะนำมาวิเคราะห์

- การทำเหมืองข้อมูล (Data Mining)

  เป็นขั้นตอนการเลือกอัลกอริทึมให้เหมาะสมกับข้อมูลที่เตรียมไว้ และทำการประมวลผลของข้อมูลตามอัลกอลิทึมที่เลือก ซึ่งการจะใช้อัลกอริทึมใดนั้นขึ้นอยู่กับความต้องการหรือปัญหาที่เกิดขึ้นเป็นหลัก และในการเลือกใช้อัลกอริทึมที่เหมาะสมจะทำให้ได้ผลลัพธ์เพื่อนำไปใช้ในการวิเคราะห์ที่ถูกต้องมากขึ้น

- วิเคราะห์ผลลัพธ์และรวบรวมความรู้ (Assimilate Information)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นการวิเคราะห์และรวบรวมความรู้ที่ได้จากการผ่านกระบวนการทำเหมืองข้อมูลว่ามีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ โดยที่ผลลัพธ์อาจจะแสดงในรูปแบบที่เป็นกราฟหรือตารางข้อมูล ซึ่งจะเป็นข้อมูลความรู้ (Knowledge) เพื่อเป็นส่วนช่วยในการตัดสินใจทางธุรกิจ

2.3 การทำเหมืองข้อมูล

วิธีการทำเหมืองข้อมูลตามมาตรฐานทั่วไปสามารถแบ่งออกเป็น 4 วิธี คือ

1. Predictive Model

เป็นการสร้างแบบจำลองจากข้อมูลเพื่อใช้ในการคาดคะเนหรือทำนายความเป็นไปได้ของข้อมูลที่มีอยู่ เช่น การทำนายอุณหภูมิ, การทำนายหุ้น เป็นต้นโดย Predictive Modeling แบ่งเป็น 2 ลักษณะคือ

- Classification เป็นการจำลองเพื่อการแบ่งกลุ่มของข้อมูลที่ต้องการโดยที่กลุ่มของข้อมูลนั้นจะแบ่งกลุ่มตามชนิดของข้อมูลที่ควรจะเป็น เช่น การแบ่งกลุ่มการพิจารณาการให้วงเงินสินเชื่อของลูกค้า โดยเทคนิคที่นิยมใช้คือ Tree Induction และ Neural Induction
- Value Prediction เป็นการพยากรณ์ค่าที่เป็นไปได้ หรือการกระจายของค่าที่เป็นไปได้ของตัวแปรใดๆ ในกลุ่มข้อมูล การทำนายค่าที่เป็นตัวเลข เช่น การทำนายภาษีที่จะเก็บได้ในแต่ละปี หรือการทำนายราคาหุ้น เป็นต้น โดยเทคนิกที่นิยมใช้คือ Linear regression และ Nonlinear regression

2. Database Segmentation

เป็นการแบ่งกลุ่มของข้อมูลที่ไม่สามารถกำหนดได้ว่าข้อมูลอยู่ในกลุ่มใด โดยเหมาะกับข้อมูลที่มีลักษณะเป็นกลุ่ม โดยเฉพาะจำนวนของกลุ่มน้อย ๆ โดยที่ข้อมูลใดที่มีลักษณะคล้ายกันหรือใกล้เคียงกันก็จะจัดเป็นกลุ่มเดียวกัน เพื่อให้ง่ายต่อการวิเคราห์ เช่น การแบ่งกลุ่มตามรายได้ของประชากร เพื่อนำผลลัพธ์ที่ได้ไปวิเคราะห์และเสนอสินค้าให้ตรงกับกลุ่มที่ได้ทำการแบ่งไว้ กล่าวคือ หากประชากรที่มีรายได้อยู่ในกลุ่มปานกลางก็ควรเสนอสินค้าราคาและคุณภาพที่ไม่สูงและไม่ต่ำเกินไป เป็นต้น ซึ่งเราสามารถเรียกการทำ Database Segmentation อีกอย่างหนึ่งว่า Clustering โดยเทคนิคที่นิยมใช้คือ Demographic Clustering และ Neural Clustering

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. Link Analysis

การหาความสัมพันธ์ของข้อมูลแต่ละเรคคอร์ดหรือกลุ่มเรคคอร์ด ซึ่งการหาความสัมพันธ์ของข้อมูลนี้ส่วนใหญ่จะนำไปใช้ในการวิเคราะห์หาความสัมพันธ์ของสินค้าที่ลูกค้ามักจะซื้อคู่กัน โดย Link Analysis แบ่งเป็น 3 ชนิด

- Association discovery ใช้วิเคราะห์การซื้อสินค้าภายในรายการเดียวกัน ศึกษาถึงความสัมพันธ์อย่างใกล้ชิดที่ถูก ปิดซ่อนอยู่ของสินค้า ซึ่งสินค้าเหล่านั้นมักมีแนวโน้มที่จะถูกซื้อควบคู่กันไป การวิเคราะห์แบบนี้เรียกว่า "Market Basket Analysis" คือ รายการทั้งหมดที่ลูกค้าซื้อต่อครั้งที่ซุปเปอร์มาร์เก็ต
- Sequential pattern discover ถูกใช้ระบุความเกี่ยวเนื่องกันของพฤติกรรมในระยะยาว ว่าพฤติกรรมหรือเหตุการณ์ใดเกิดขึ้นแล้วจะมีพฤติกรรมหรือเหตุการณ์ใดเกิดขึ้นตามมา เช่น ผู้ขายอาจพบว่าลูกค้าที่ซื้อทีวีมีแนวโน้มที่จะซื้อวีดีโอในเวลาต่อมา
- Similar time sequence discovery ใช้ค้นหาความเกี่ยวเนื่องกันระหว่างกลุ่มของข้อมูล 2 กลุ่ม ซึ่งการขึ้นต่อกัน ทางด้านเวลา โดยมีรูปแบบการเคลื่อนที่เหมือนกัน ผู้ขายสินค้ามักจะใช้เพื่อดูแนวโน้มเพื่อเตรียมสต๊อก เช่นเมื่อไรก็ตามที่ ยอดขาย สินค้าน้ำอัดลมสูงขึ้น ยอดขายมันฝรั่งจะสูงขึ้นตาม

4. Deviation Detection

เป็นเทคนิคที่ใช้ในการหาค่าที่ต่างไปจากที่คาดไว้ว่าค่านั้นต่างไปมากน้อยเพียงใด ซึ่งใช้เทคนิคทางด้านสถิติช่วยในการสรุปข้อมูลออกมาให้อยู่ในรูปแบบของกราฟฟิกเพื่อแสดงให้ภาพ เพื่อใช้วัดความน่าเชื่อถือของข้อมูลและแสดงให้เห็นภาพ โดยที่เทคนิคนี้สามารถนำไปใช้ตรวจสอบได้หลายด้าน เช่น ตรวจสอบลายเซ็นปลอม, บัตรเครดิตปลอม เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การค้นหากฎความสัมพันธ์ (Association Rule Discovery)

3.1 กฎความสัมพันธ์ (Association Rule)

Association Rule เป็นเทคนิคที่ใช้ในการหากฎความสัมพันธ์ที่เกิดขึ้น เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายพฤติกรรมต่างๆ ที่เกิดขึ้น เช่น การหาความสัมพันธ์พฤติกรรมการซื้อสินค้าของลูกค้า โดยรูปแบบความสัมพันธ์จะเป็นความสัมพันธ์ระหว่างไอเทม ต่างๆ โดยสามารถคิดหาค่ากฎความสัมพันธ์ได้ดังนี้

3.1.1 การคิดหาค่าของกฎความสัมพันธ์

การค้นหาความสัมพันธ์ที่สนใจระหว่างไอเทม ในชุดของข้อมูลที่มีอยู่ในฐานข้อมูล ตัวอย่างในการหาความสัมพันธ์แบบนี้ คือ วิเคราะห์ความเป็นไปได้ในการบริโภคสินค้าในตลาด (market basket analysis) ซึ่งเป็นการหาความสัมพันธ์ของสินค้าที่ลูกค้ามักจะซื้อไปด้วยกัน เพื่อนำไปช่วยในการวางแผนและส่งเสริมการ เช่น ถ้าลูกค้าคนไหนซื้อคอมพิวเตอร์ก็จะซื้อซอฟแวร์ไปด้วย ในเวลาเดียวกัน การวางฮาร์ดแวร์ไว้กับซอฟแวร์อาจจะช่วยเพิ่มยอดขายของสินค้าทั้ง 2 ชนิดนั้นได้

กฎในการหาความสัมพันธ์จะเป็นการหาความสัมพันธ์ของรายการสินค้าที่เกิดขึ้นในแต่ละรายการ แล้วนับจำนวนความสัมพันธ์ของสินค้าที่เกิดขึ้น โดยอาศัยการสร้างตาราง Co-occurrence table เพื่อเปรียบเทียบจำนวนของสินค้าที่มีการซื้อไปด้วยกัน

ทั่วไปกฎความสัมพันธ์ของการสั่งซื้อสินค้าจะอยู่ในรูปไวยกรณ์ดังนี้

if A then B

โดยใช้สัญลักษณ์ $A \rightarrow B$ ซึ่งมีตัวแปรที่ใช้ในการวัดค่าข้อมูลอยู่ด้วยกัน 2 ตัวแปร คือ

- Support factor คือ เปอร์เซ็นต์ที่แสดงความสัมพันธ์ระหว่างจำนวนรายการของข้อมูล $A \rightarrow B$ ที่เกิดขึ้นเปรียบเทียบกับจำนวนรายการของข้อมูลทั้งหมด เช่น จากกฎในการหาความสัมพันธ์ คอมพิวเตอร์ → ซอฟต์แวร์ เมื่อลูกค้าซื้อเครื่องคอมพิวเตอร์ก็จะซื้อซอฟต์แวร์คิดเป็น 2% เทียบกับสินค้าที่ขายทั้งหมดในร้าน ซึ่งหาได้จาก

$$\frac{\text{จำนวนที่ลูกค้าซื้อคอมพิวเตอร์และซอฟต์แวร์คู่กัน}}{\text{จำนวนสินค้าที่ขายทั้งหมดในร้าน}}$$

ดังนั้นค่า Support factor สามารถเขียนให้อยู่ในรูปแบบสมการได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Support(A \rightarrow B) = P(A \cup B)$$

- Confidence factor คือ เปอร์เซ็นต์ที่แสดงค่าความสัมพันธ์ที่ซ่อนอยู่ในตัวเองกับสินค้ารายการอื่น เช่น จากกฎในการหาความสัมพันธ์ คอมพิวเตอร์ → ซอฟต์แวร์ เมื่อลูกค้าซื้อเครื่องคอมพิวเตอร์ก็จะซื้อซอฟต์แวร์คิดเป็น 70% ของลูกค้าที่ซื้อคอมพิวเตอร์ทั้งหมด ซึ่งหาได้จาก

$$\frac{\text{จำนวนที่ลูกค้าซื้อคอมพิวเตอร์และซอฟต์แวร์คู่กัน}}{\text{จำนวนสินค้าที่ลูกค้าซื้อคอมพิวเตอร์ทั้งหมด}}$$

ดังนั้นค่า Confidence factor สามารถเขียนให้อยู่ในรูปแบบสมการได้ดังนี้

$$Confidence\ (A \rightarrow B) = P(B \mid A)$$

3.2 อัลกอรึทึม

อพิออริเป็นอัลกอริทึมที่ใช้ในการหากฎความสัมพันธ์ของข้อมูลซึ่งมีขั้นตอนการทำอยู่ 2 ขั้นตอน ดังนี้

ขั้นตอนที่1 หาความถี่ของกลุ่มไอเทม(Frequent itemsets) ทั้งหมดโดยที่ความถี่ของกลุ่มไอเทมจะเป็นไอเทมที่เกิดขึ้นอย่างน้อย x% ซึ่งค่า x นั้นเราเป็นผู้กำหนด ดังนั้นเราจะเรียกค่านี้ว่า ความเป็นไปได้ต่ำสุด (minimum_support)

ในการหาความถี่ของกลุ่มไอเทม นั้นจะมีการทำงานซ้ำไปเรื่อยๆ กล่าวคือความถี่ของกลุ่มไอเทมมีค่าเท่ากับ k จะใช้หา frequent itemset (k+1)-itemsets โดยที่ขั้นแรก เซตของ frequent 1 – itemset แทน L_1 และ L_1 ก็จะถูกใช้ในการหา frequent itemset 2 และเซตของ frequent itemset 2 – itemset ซึ่งจะถูกใช้ในการหา L_3 ไปเรื่อยๆ

อัลกอริทึมการค้นหาความถี่ของกลุ่มไอเทม(Frequent itemsets) มีตัวแปรที่นำมาใช้ ดังนี้

D คือ ฐานข้อมูลที่เก็บแต่ละทรานส์แอ็กชัน

L_k คือ เซตของ frequent k-itemset

K-Itemset คือ จำนวนสมาชิกของเซตของข้อมูลที่มี K ตัว

C_k คือ เซตของ candidate k-itemset

t คือ จำนวนทรานส์แอ็กชันในฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

L_1 = find_frequent_1-itemsets(D);

for ($k = 2$; $L_{k-1} \neq \phi$; k++) {

C_k = **apriori_gen**(L_{k-1}, *min_sup*);

for each transaction $t \in D$ { // scan D for counts

C_t = subset(C_k, t); // get the subsets of t that are candidates

for each candidate $c \in C_t$

c.count ++;

}

$L_k = \{c \in C_k | c.count \geq min_sup\}$

}

return $L = \cup_k L_k$;

Procedure apriori_gen(L_{k-1}: frequent (k - 1)-itemsets; *min_sup*: minimum support threshold)

for each itemset $l_1 \in L_{k-1}$

for each itemset $l_2 \in L_{k-1}$

if ($l_1[1] = l_2[1]) \wedge (l_1[2] = l_2[2]) \wedge \ldots \wedge (l_1[k\text{-}2] = l_2[k\text{-}2]) \wedge (l_1[k\text{-}1] < l_2[k\text{-}1])$ **then**

{

$c = l1 \bowtie l2$; // join step: generate candidates

if has_infrequent_subset(c, L_{k-1}) **then**

delete c; // prune step: remove unfruitful candidate

else add c **to** C_k;

}

return C_k;

Procedure has_infrequent_subset(c : candidate k-itemset; L_{k-1} : frequent (k - 1)-itemsets);

// use prior knowledge

for each (k-1)-subset s **of** c

if $s \notin L_{k-1}$ **then**

return TRUE;

return FALSE;

รูปที่3.1 อัลกอริทึมการหา frequent itemset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการหา frequent itemset ประกอบด้วย 2 ขั้นตอน ดังนี้

- **Join** ในการหา L_k นั้น เป็นการนำ L_{k-1} มาทำการ Join กับตัวมันเอง เพื่อหา Candidate itemset-K (C_k) ซึ่งจะเขียนเป็นสัญลักษณ์คือ $L_{k-1} \bowtie L_{k-1}$ โดยที่ให้ $l_i[j]$ แทนสมาชิกแต่ละตัวของ L_{k-1}, j คือลำดับที่ของแต่ละสมาชิก และ $l_i[j]$ จะ join กันได้นั้น $l_1[j]$ จะต้องเท่ากับ $l_2[j]$ กล่าวคือ สมาชิกตั้งแต่ตัวแรกจนถึงตัวรองสุดท้ายจะต้องเหมือนกันจึงสามารถทำการ join กันได้

Join Step

Insert into C_k

Select $p.item_1, p.item_2, \ldots, p.item_{k-1}, q.item_{k-1}$

From L_{k-1} p, L_{k-1} q

Where $p.item_1 = q.item_1, \ldots, p.item_{k-2} = p.item_{k-2}, p.item_{k-1} < p.item_{k-1}$

รูปที่3.2 อัลกอริทึมขั้นตอนการ Join

- **Prune** Candidate itemset-K (C_k) คือ superset ของ L_k อาจจะมีจำนวนมากเพราะเนื่องมาจากการ join ของ L_{k-1} จึงต้องมีการลดจำนวนของ C_k โดย item ที่ได้จากการ join ของ L_{k-1} กับ L_{k-1} จะเป็น C_k ได้ก็ต่อเมื่อ ทุกๆ subset ของตัวมันเองเป็นสมาชิกของ L_{k-1} หากมี subset ใดไม่เป็นสมาชิกของ L_{k-1} แล้วนั้น itemsets ก็จะถูกลบทิ้งไป

Pruning Step

Forall itemsets $c \in C_k$ **do**

  Forall $(k-1)$ – subset s of c **do**

    If $(s \notin L_{k-1})$ **then** delete c from C_k

รูปที่3.3 อัลกอริทึมขั้นตอนการ Prune

<u>ขั้นตอนที่2</u> สร้างกฎความสัมพันธ์จาก frequent itemsets เมื่อได้ค่า frequent itemsets แล้วก็จะนำมาสร้างเป็นกฎความสัมพันธ์ โดยกฎความสัมพันธ์ได้มาจาก แต่ละ frequent itemset L หา subset ของ frequent itemset ของ L และ ทุกๆ subset s ของ L ยกเว้นเซตว่าง จะได้ผลลัพธ์คือ s $\rightarrow$ (L-s)

สูตรในการหาค่า confidence คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{confidence}(A \rightarrow B) = \frac{\text{support_count}(A \cup B)}{\text{support_count}(A)}$$

โดยที่ support_count(A U B) คือจำนวนของทรานส์แอ็กชันที่มี itemset A U B และ support_count(A) คือจำนวนของทรานส์แอ็กชันที่มี itemset A ซึ่งกฎจะมีความน่าเชื่อถือได้นั้น ผลลัพธ์ที่ได้จากการหาค่าความสัมพันธ์จะต้องมีค่ามากกว่าหรือเท่ากับ minimum confidence

ตัวอย่างที่1 การศึกษาร้านขายสินค้า โดยใช้อัลกอริทึมอพิออริ

ตารางที่3.1 ทรานส์แอ็กชันของข้อมูลสำหรับร้านขายสินค้า

TID	List of item
T001	I1,I2,I5
T002	I2,I4
T003	I2,I3
T004	I1,I2,I4
T005	I1,I3
T006	I2,I3
T007	I1,I3
T008	I1,I2,I3,I5
T009	I1,I2,I3

การหา frequent itemset

1. นับจำนวนของแต่ละไอเทมที่เกิดขึ้น ของแต่ละ ทรานส์แอ็กชันซึ่งจะได้ค่า C_1
2. ทำการเปรียบเทียบจำนวนสินค้าแต่ละไอเทมที่นับได้จากข้อที่ 1 กับ minimum support โดยที่กำหนดให้ค่า minimum support = 2 จะได้ L_1
3. ทำการหา C_2 โดยการ join L_1 กับ L_1 ด้วยกันเอง
4. ทำการนับจำนวนของ itemset ที่มีการ join กันในขั้นที่ 3
5. นำจำนวนที่นับได้มาเปรียบเทียบกับ minimum support ได้เป็น L_2
6. ทำการหา C_3 โดยการ join L_2 กับ L_2
7. นับจำนวนของ itemset ที่ join กันระหว่าง L_2
8. นำจำนวนที่นับได้มาเปรียบเทียบกับ minimum support ได้เป็น L_3

ตัวอย่าง แสดงการหา C_3 โดยการ join L_2 กับ L_2

1) ขั้นตอนการ Join

เอกสารนี้ใช้เอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

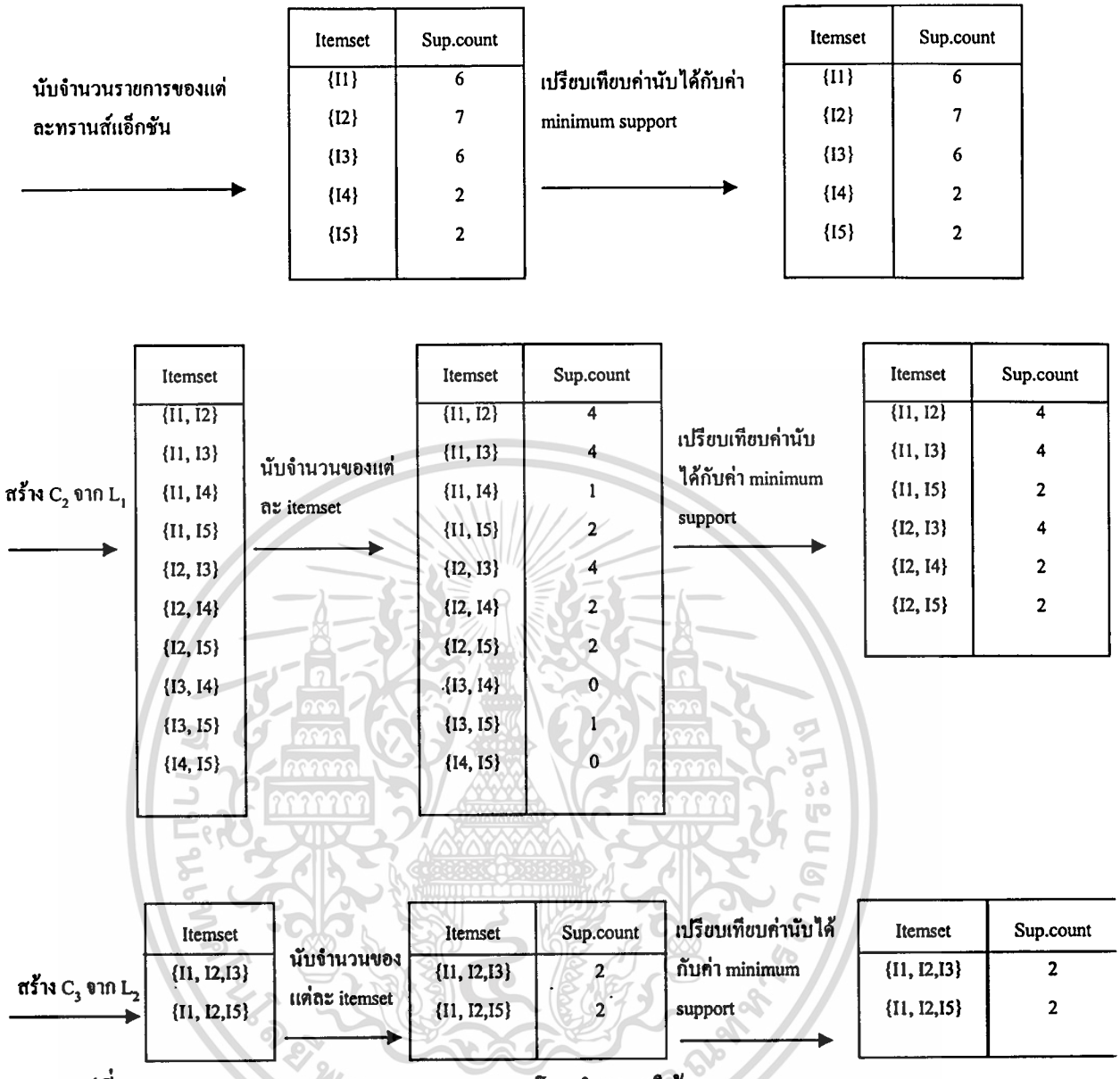
$C_3 = L_2 \bowtie L_2$

{{I1, I2},{I1, I3},{I1, I5},{I2, I3},{I2, I4},{I2, I5}} ⋈ {{I1, I2},{I1, I3},{I1, I5},{I2, I3},{I2, I4}, {I2, I5}} = {{I1,I2,I3},{I1,I2,I5},{I1,I3,I5},{I2,I3,I4},{I2,I3,I5},{I2,I4,I5}}

2) ขั้นตอนการ Prune

- subset ของ {I1,I2,I3} = {I1,I2},{I1,I3},{I2,I3} subset ของ {I1,I2,I3}เป็นสมาชิกของ L_2 ดังนั้น {I1,I2,I3} จึงอยู่ใน C_3
- subset ของ {I1,I2,I5} = {I1,I2},{I1,I5},{I2,I5} subset ของ {I2,I3,I5} เป็นสมาชิกของ L_2 ดังนั้น {I1,I2,I5} จึงอยู่ใน C_3
- subset ของ {I1,I3,I5} = {I1,I3},{I1,I5},{I3,I5} เนื่องจาก {I3,I5} ไม่ได้เป็นสมาชิกของ L_2 ดังนั้น {I1,I3,I5} จึงไม่อยู่ใน C_3
- subset ของ {I2,I3,I4} = {I2,I3},{I2,I4},{I3,I4}เนื่องจาก {I3,I4} ไม่ได้เป็นสมาชิกของ L_2 ดังนั้น {I2,I3,I4} จึงไม่อยู่ใน C_3
- subset ของ {I2,I3,I5} = {I2,I3},{I2,I5},{I3,I5}เนื่องจาก {I3,I5} ไม่ได้เป็นสมาชิกของ L_2 ดังนั้น {I2,I3,I5} จึงไม่อยู่ใน C_3
- subset ของ {I2,I4,I5} = {I2,I4},{I2,I5},{I4,I5}เนื่องจาก {I4,I5} ไม่ได้เป็นสมาชิกของ L_2 ดังนั้น {I2,I4,I5} จึงไม่อยู่ใน C_3 ดังนั้นจะได้ C_3 = {{I1, I2,I3},{I1, I2,I5}} หลังจากการ Prune แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Itemset	Sup.count
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2

Itemset	Sup.count
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2

Itemset
{I1, I2}
{I1, I3}
{I1, I4}
{I1, I5}
{I2, I3}
{I2, I4}
{I2, I5}
{I3, I4}
{I3, I5}
{I4, I5}

Itemset	Sup.count
{I1, I2}	4
{I1, I3}	4
{I1, I4}	1
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	2
{I3, I4}	0
{I3, I5}	1
{I4, I5}	0

Itemset	Sup.count
{I1, I2}	4
{I1, I3}	4
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	2

Itemset
{I1, I2,I3}
{I1, I2,I5}

Itemset	Sup.count
{I1, I2,I3}	2
{I1, I2,I5}	2

Itemset	Sup.count
{I1, I2,I3}	2
{I1, I2,I5}	2

รูปที่3.4 Itemset และ การหา frequent itemset โดยกำหนดให้ minimum support = 2

การสร้างกฎความสัมพันธ์จาก frequent itemsets

$$L = \{I1, I2, I5\}$$

สามารถแยกออกเป็น subset ยกเว้นเซตว่างได้ดังนี้ คือ {I1},{I2},{I5},{I1,I2},{I1,I5},{I2,I5}

I1 → I2 & I5 ค่าความเชื่อมั่น = 2/6 = 33%

I2 → I1 & I5 ค่าความเชื่อมั่น = 2/7 = 29%

I5 → I1 & I2 ค่าความเชื่อมั่น = 2/2 = 100%

I1 & I2 → I5 ค่าความเชื่อมั่น = 2/4 = 50%

I1 & I5 → I2 ค่าความเชื่อมั่น = 2/2 = 100%

I2 & I5 → I1 ค่าความเชื่อมั่น = 2/2 = 100%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ้าค่า minimum confidence = 70% กฎที่น่าเชื่อถือคือ กฎข้อที่ 3,5 และ 6 ซึ่งกฎที่ได้นี้สามารถนำไปใช้ในการวางแผน หรือวิเคราะห์เพื่อการดำเนินธุรกิจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่4

วิเคราะห์ และออกแบบระบบ

ขั้นตอนการวิเคราะห์ และออกแบบรูปแบบจำลองระบบ เพื่อนำมาพัฒนาโครงงานนี้เป็นการนำเทคโนโลยีการทำเหมืองข้อมูลใช้ในการประมวลผล ดังนั้นการออกแบบจะใช้รูปแบบจำลองจากยูเอ็มแอล เพื่อนำมาอธิบายขั้นตอนและเนื้อหาในการพัฒนาโปรแกรมมีรายละเอียดดังนี้

4.1 ความต้องการ และข้อจำกัดของโปรแกรม

โครงการพัฒนาระบบงานนี้สามารถกำหนดความต้องการ และข้อจำกัดของโปรแกรมเพื่อใช้เป็นขอบเขตในการพัฒนาโครงการ ดังนี้

4.1.1 ความต้องการของผู้ใช้งาน

จากการสอบถามความต้องการของผู้ใช้งานมีรายละเอียดความต้องการต่างๆ ดังนี้

- ต้องการทราบความสัมพันธ์ของแต่ละข้อมูล
- ต้องสามารถกำหนดมาตรฐานหรือกฎเกณฑ์เพื่อใช้อธิบายความสัมพันธ์ได้
- ต้องสามารถบอกให้ทราบถึงผลลัพธ์หรือความสัมพันธ์อื่นๆ ที่เกิดขึ้นจากกระบวนการหาความสัมพันธ์ได้
- ต้องสามารถบอกความเป็นไปได้ของจำนวนครั้งความสัมพันธ์ที่นำมาวิเคราะห์ได้

4.1.2 ความต้องการเชิงฟังก์ชั่น

การทำงานของโปรแกรมต้องสามารถรองรับการทำงานต่างๆ ดังนี้

- ค้นหาความสัมพันธ์ของข้อมูลโดยใช้อัลกอริทึมอพิออริ
- ระบบต้องสามารถหาความสัมพันธ์ของข้อมูลได้
- ระบบต้องสามารถหาความสัมพันธ์ของข้อมูลออกมาเป็นกฎได้
- ระบบต้องสามารถเรียงลำดับกฎความสัมพันธ์ตามค่า Support factor หรือ Confidence factor ได้
- ระบบต้องสามารถระบุชื่อกลุ่มที่ใช้ในการจัดกลุ่ม เพื่อหาความสัมพันธ์กับข้อมูลอื่นๆด้วยการใช้เทคนิคการจัดกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

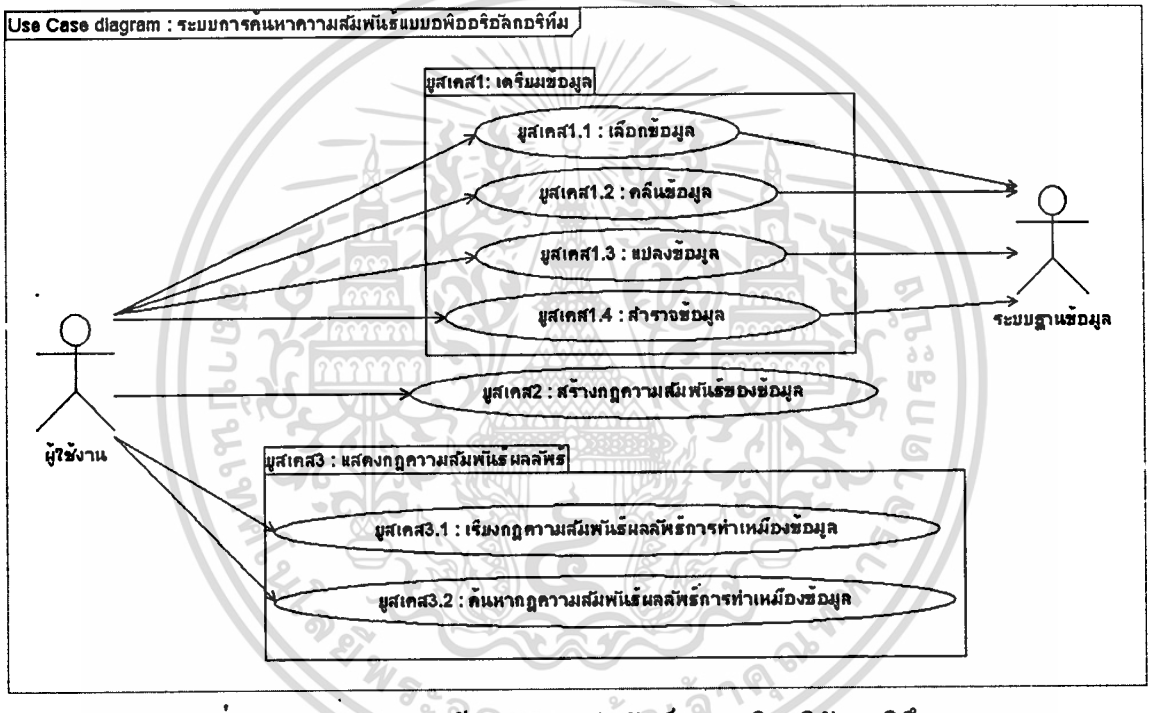
4.1.3 ความต้องการที่ไม่ใช่เชิงฟังก์ชั่น

- ระบบต้องตอบสนองการทำงาน ณ ขั้นตอนต่างๆ ขณะประมวลผลได้
- ระบบบอกสถานะการทำงาน ณ ขั้นตอนต่างๆ ขณะประมวลผลได้

4.2 การวิเคราะห์ระบบโดยใช้ยูเคสไดอะแกรม และแอคทิวิตีไดอะแกรม

4.2.1 ยูสเคสไดอะแกรม (Use Case Diagram)

ขอบเขตการพัฒนาระบบค้นหากฎความสัมพันธ์นำมาอธิบายเป็นรูปแบบจำลองโดยใช้รูปยูสเคสได้ดังนี้



รูปที่4.1 ยูเคสระบบการค้นหาความสัมพันธ์แบบอพิออริอัลกอริทึม

4.2.2 รายละเอียดแอคเตอร์

จากรูปที่3.1 เป็นการใช้ยูสเคสแสดงให้ทราบถึงขอบเขตการพัฒนาโครงงานระบบ โดยมีแอคเตอร์ ซึ่งมีรายละเอียด ดังนี้

1. ผู้ใช้งาน

เป็นบุคคลที่ใส่ค่าอินพุตเข้าสู่ระบบ และเป็นผู้รอผลลัพธ์การประมวลผลเพื่อนำไปทำนายรวมกับความรู้ของตนเอง

2. ระบบฐานข้อมูล

เป็นระบบที่ใช้ในการเก็บข้อมูลต่างๆ เพื่อนำมาใช้ในการทำเหมืองข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

4.2.3 รายละเอียดยูสเคส (Use Case Description)

4.2.3.1 ยูสเคสที่1 : แพคเกจการเตรียมข้อมูล เป็นแพคเกจที่ภายในประกอบด้วย 6 ยูสเคส ที่อธิบายถึงขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับการทำเหมืองข้อมูล และอัลกอริทึมที่ได้เลือกไว้ โดยแพคเกจสำหรับการเตรียมข้อมูลเป็นการนำระบบการเตรียมข้อมูลและการสำรวจข้อมูลของโครงงานพัฒนาระบบการพัฒนาระบบการเตรียมข้อมูลและการสำรวจ สำหรับการทำดาต้าไมนิ่งมารวม เพื่อนำมาใช้พัฒนาในโครงงานนี้ โดยแต่ละยูสเคสมีดังนี้

- ยูสเคสที่1.1 : ยูสเคสเลือกข้อมูล เป็นยูสเคสอธิบายถึงขั้นตอนการเลือกข้อมูลที่ต้องและติดต่อฐานข้อมูล โดยจะติดต่อกับ Microsoft SQL Server 2005 ซึ่งได้อธิบายไว้ในรูปที่ 4.2
- ยูสเคสที่1.2 : ยูสเคสคลีนข้อมูล เป็นยูสเคสอธิบายถึงขั้นตอนการคลีนข้อมูล เพื่อทำการแก้ไขค่าว่างโดยวิธีต่างๆ ซึ่งได้อธิบายดังรูปที่ 4.3
- ยูสเคสที่1.3 : ยูสเคสแปลงข้อมูล เป็นยูสเคสอธิบายถึงขั้นตอนการปรับเปลี่ยนข้อมูลตามประเภทของข้อมูลที่จะนำไปใช้ในการทำเหมืองข้อมูล ซึ่งได้อธิบายดังรูปที่4.4
- ยูสเคสที่1.4 : ยูสเคสสำรวจข้อมูล เป็นยูสเคสที่อธิบายถึงการแสดงผลลัพธ์ข้อมูลที่ผ่านกระบวนการแปลงข้อมูลเรียบร้อยแล้ว โดยที่จะแสดงผลลัพธ์ในรูปแบบรายละเอียดและกราฟ ซึ่งได้อธิบายดังรูปที่ 4.5

4.2.3.2 ยูสเคสที่2 : สร้างกฎความสัมพันธ์ของข้อมูล เป็นยูสเคสที่อธิบายขั้นตอนในการสร้างกฎความสัมพันธ์ของข้อมูล โดยใช้อพิออริอัลกอริทึมในการหากฎความสัมพันธ์ ซึ่งข้อมูลที่จะนำมาหากฎความสัมพันธ์จะต้องผ่านการเตรียมข้อมูลมาก่อน ซึ่งได้อธิบายดังรูปที่ 4.6

4.2.3.3 ยูสเคสที่3 : แพคเกจแสดงกฎความสัมพันธ์ผลลัพธ์ เป็นแพคเกจที่ภายในประกอบด้วย 2 ยูสเคสที่อธิบายถึงการแสดงกฎหลังจากการสร้างกฎความสัมพันธ์ของข้อมูลแล้ว โดยมียูสเคสดังนี้

- ยูสเคสที่3.1 : เรียงกฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล เป็นยูสเคสที่อธิบายขั้นตอนในการเรียงกฎความสัมพันธ์ผลลัพธ์ที่ได้แสดงไว้หลังจากผ่านกระบวนการหากฎความสัมพันธ์ ซึ่งได้อธิบายดังรูปที่ 4.7
- ยูสเคสที่3.2 : ค้นหากฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล เป็นยูสเคสที่อธิบายขั้นตอนในการค้นหากฎความสัมพันธ์ผลลัพธ์ที่ได้แสดงไว้หลังจากผ่านกระบวนการหากฎความสัมพันธ์ ซึ่งได้อธิบายดังรูปที่ 4.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
04423

ยูสเคส1.1 เลือกข้อมูล

รายละเอียด

เป็นยูสเคสอธิบายถึงขั้นตอนการเลือกข้อมูลที่ต้องและติดต่อฐานข้อมูล โดยจะติดต่อกับ Microsoft SQL Server 2005

แพกเกจ เตรียมข้อมูล

แอคเตอร์

1. ผู้ใช้งาน
2. ระบบฐานข้อมูล

เงื่อนไขก่อนเข้ายูสเคส

-

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานระบุชื่อเครื่องเซิฟเวอร์
2. ผู้ใช้งานระบุชื่อฐานข้อมูล
3. ระบบทำการตรวจสอบฐานข้อมูลถูกต้อง
4. ระบบแสดงตารางฐานข้อมูล
5. ผู้ใช้งานเลือกตารางฐานข้อมูล
6. ระบบแสดงแอดทริบิวของตารางฐานข้อมูล
7. ผู้ใช้งานเลือกแอดทริบิว
8. ผู้ใช้งานคลิกปุ่ม Execute

ลำดับเหตุการณ์ย่อย

2a) ระบบทำการตรวจสอบฐานข้อมูลไม่ถูกต้อง : ระบบแสดงข้อความเตือนให้ระบุชื่อเครื่องเซิฟเวอร์และชื่อฐานข้อมูลใหม่

เงื่อนไขก่อนออกยูสเคส

1. ระบบทำการเลือกข้อมูลเรียบร้อย

รูปที่4.2 ยูสเคสการเลือกข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส1.2 คลีนข้อมูล

รายละเอียด

เป็นยูสเคสอธิบายถึงขั้นตอนการคลีนข้อมูล ซึ่งทำการแก้ไขปัญหาที่พบในข้อมูล เพื่อทำให้ข้อมูลมีคุณภาพก่อนที่จะนำข้อมูลไปประมวลผลในขั้นต่อไป

แพกเกจ เตรียมข้อมูล

แอคเตอร์

1. ผู้ใช้งาน
2. ระบบฐานข้อมูล

เงื่อนไขก่อนเข้ายูสเคส

1. ต้องติดต่อฐานข้อมูลก่อน

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานเลือกแอคทริบิวที่ต้องการคลีน
2. ผู้ใช้งานเลือกประเภทการคลีน
3. ระบบทำการคลีนข้อมูลถูกต้อง

ลำดับเหตุการณ์ย่อย

3a) ระบบทำการคลีนข้อมูลไม่ถูกต้อง : ระบบแสดงข้อความเตือนระบบไม่สามารถคลีนข้อมูลได้

เงื่อนไขก่อนออกยูสเคส

1. ระบบทำการคลีนข้อมูลเรียบร้อย

รูปที่4.3 ยูสเคสการคลีนข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส1.3 แปลงข้อมูล

รายละเอียด

เป็นยูสเคสอธิบายถึงขั้นตอนการปรับเปลี่ยนข้อมูลตามประเภทของข้อมูลที่จะนำไปใช้ในการทำเหมืองข้อมูล

แพกเกจ เตรียมข้อมูล

แอคเตอร์

1. ผู้ใช้งาน
2. ระบบฐานข้อมูล

เงื่อนไขก่อนเข้ายูสเคส

1 ต้องผ่านการคลีนข้อมูลก่อน

ลำดับเหตุการณ์หลัก

1. ระบบแสดงแอดทริบิวที่เลือกไว้
2. ผู้ใช้ทำการแปลงข้อมูล

วนซ้ำข้อ 2

ลำดับเหตุการณ์ย่อย

2a) ผู้ใช้ทำการแปลงข้อมูลแบบ Normalize : ผู้ใช้งานเลือกประเภทการ Normalize

2b) ผู้ใช้ทำการแปลงข้อมูลแบบ Construct New Attribute : ผู้ใช้งานระบุค่าและเลือกวิธีการคำนวณแอดทริบิวใหม่

2c) ผู้ใช้ทำการแปลงข้อมูลแบบ Numeric to Categorical : ผู้ใช้งานระบุค่าที่ต้องการแปลง

2d) ผู้ใช้ทำการแปลงข้อมูลแบบ Categorical to Numeric : ผู้ใช้งานเลือกประเภทการแปลง

2e) ผู้ใช้ทำการแปลงข้อมูลแบบ Grouping : ผู้ใช้งานเลือกข้อมูลและกำหนดชื่อกลุ่มของข้อมูล

เงื่อนไขก่อนออกยูสเคส

1. ระบบทำการแปลงข้อมูลเรียบร้อย

รูปที่4.4 ยูสเคสการแปลงข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส1.5 สำรวจข้อมูล

รายละเอียด

เป็นยูสเคสที่อธิบายถึงการแสดงผลลัพธ์ข้อมูลที่ผ่านกระบวนการแปลงข้อมูลเรียบร้อยแล้ว โดยที่จะแสดงผลลัพธ์ในรูปแบบรายละเอียดและกราฟ

แพกเกจ เตรียมข้อมูล

แอคเตอร์

1. ผู้ใช้งาน
2. ระบบฐานข้อมูล

เงื่อนไขก่อนเข้ายูสเคส

1. ระบบต้องผ่านขั้นตอนการแปลงข้อมูลก่อน

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานคลิกแท็บ Exploration
2. ระบบแสดงรายชื่อแอคทริบิวทั้งหมดที่ได้จากการแปลงข้อมูล
3. ผู้ใช้เลือกแอคทริบิว
4. ระบบแสดงรายละเอียดของแอคทริบิวที่เลือก

ลำดับเหตุการณ์ย่อย

-

เงื่อนไขก่อนออกยูสเคส

1. ระบบแสดงรายละเอียดของแอคทริบิวเรียบร้อย

รูปที่4.5 ยูสเคสการสำรวจข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส2 : สร้างกฎความสัมพันธ์ของข้อมูล

รายละเอียด

เป็นยูสเคสที่อธิบายขั้นตอนในการหากฎความสัมพันธ์ของข้อมูลที่ได้มาจากขั้นตอนการเตรียมข้อมูล โดยใช้อพิออริอัลกอริทึม ซึ่งในการหากฎความสัมพันธ์จะใช้ข้อมูลเพียง 2 ฟิลด์ เท่านั้นในการหาความสัมพันธ์

แอคเตอร์

1. ผู้ใช้งาน

เงื่อนไขก่อนเข้ายูสเคส

1. ข้อมูลต้องผ่านขั้นตอนการเตรียมข้อมูลก่อน

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานเลือกประเภทในการทำเหมืองข้อมูล
2. ผู้ใช้งานคลิกปุ่ม Choose mining method
3. ระบบยืนยันการเลือกประเภทการทำเหมืองข้อมูล
4. ผู้ใช้งานเลือกข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล
5. ผู้ใช้งานกำหนดข้อมูล Transaction
6. ผู้ใช้งานกำหนดข้อมูล Item
7. ผู้ใช้งานกำหนดค่า Minimum Support
8. ผู้ใช้งานกำหนดค่า Minimum Confidence
9. ผู้ใช้งานคลิกปุ่ม Generate Rule
10. ระบบแสดงกฎความสัมพันธ์ของข้อมูล

ลำดับเหตุการณ์ย่อย

10a) ระบบทำการสร้างกฎความสัมพันธ์ไม่ถูกต้อง : ระบบแจ้งเตือนข้อผิดพลาดและหยุดการทำงาน

เงื่อนไขก่อนออกยูสเคส

1. ระบบสร้างกฎความสัมพันธ์ของข้อมูลเสร็จเรียบร้อยแล้ว

รูปที่4.6 ยูสเคสการสร้างกฎความสัมพันธ์ของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส3.1 : เรียงกฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล

รายละเอียด

เป็นยูสเคสที่อธิบายขั้นตอนการเรียงกฎความสัมพันธ์ผลลัพธ์เมื่อผ่านกระบวนการทำเหมืองข้อมูล โดยการเรียงผลลัพธ์กฎความสัมพันธ์นั้นผู้ใช้งานสามารถเรียงลำดับกฎตามค่า Support factor , Confidence factor หรือ Analysis

แพคเกจ การแสดงกฎความสัมพันธ์ผลลัพธ์

แอคเตอร์

1. ผู้ใช้งาน

เงื่อนไขก่อนเข้ายูสเคส

1. ผ่านกระบวนการทำเหมืองข้อมูลแล้ว

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานคลิกชื่อคอลัมน์ผลลัพธ์ที่ต้องการเรียง
2. ระบบเรียงกฎความสัมพันธ์

ลำดับเหตุการณ์ย่อย

1a) ผู้ใช้งานเลือกเรียงข้อมูลตาม Analysis : ระบบทำการเรียงกฎตามตัวอักษรของกฎผลลัพธ์

1b) ผู้ใช้งานเลือกเรียงข้อมูลตาม Support factor : ระบบทำการเรียงกฎตามค่า Support factor

1c) ผู้ใช้งานเลือกเรียงข้อมูลตาม Confidence factor : ระบบทำการเรียงกฎตามค่า Confidence factor

เงื่อนไขก่อนออกยูสเคส

1. ระบบเรียงกฎความสัมพันธ์เรียบร้อย

รูปที่4.7 ยูสเคสเรียงกฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยูสเคส3.2 : ค้นหากฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล

รายละเอียด

เป็นยูสเคสที่อธิบายขั้นตอนค้นหากฎความสัมพันธ์ผลลัพธ์เมื่อผ่านกระบวนการทำเหมืองข้อมูล โดยการค้นหาผลลัพธ์กฎความสัมพันธ์นั้นผู้ใช้งานสามารถระบุกฎที่ต้องการได้

แพคเกจ แสดงกฎความสัมพันธ์ผลลัพธ์

แอคเตอร์

1. ผู้ใช้งาน

เงื่อนไขก่อนเข้ายูสเคส

1. ผ่านกระบวนการทำเหมืองข้อมูลแล้ว

ลำดับเหตุการณ์หลัก

1. ผู้ใช้งานระบุไอเทมความสัมพันธ์ถูกต้อง
2. ผู้ใช้งานคลิกปุ่ม Search Analysis
3. ระบบแสดงกฎความสัมพันธ์ที่ค้นหา

ลำดับเหตุการณ์ย่อย

1a) ผู้ใช้งานระบุไอเทมความสัมพันธ์ไม่ถูกต้อง : ระบบแสดงข้อความเตือน

เงื่อนไขก่อนออกยูสเคส

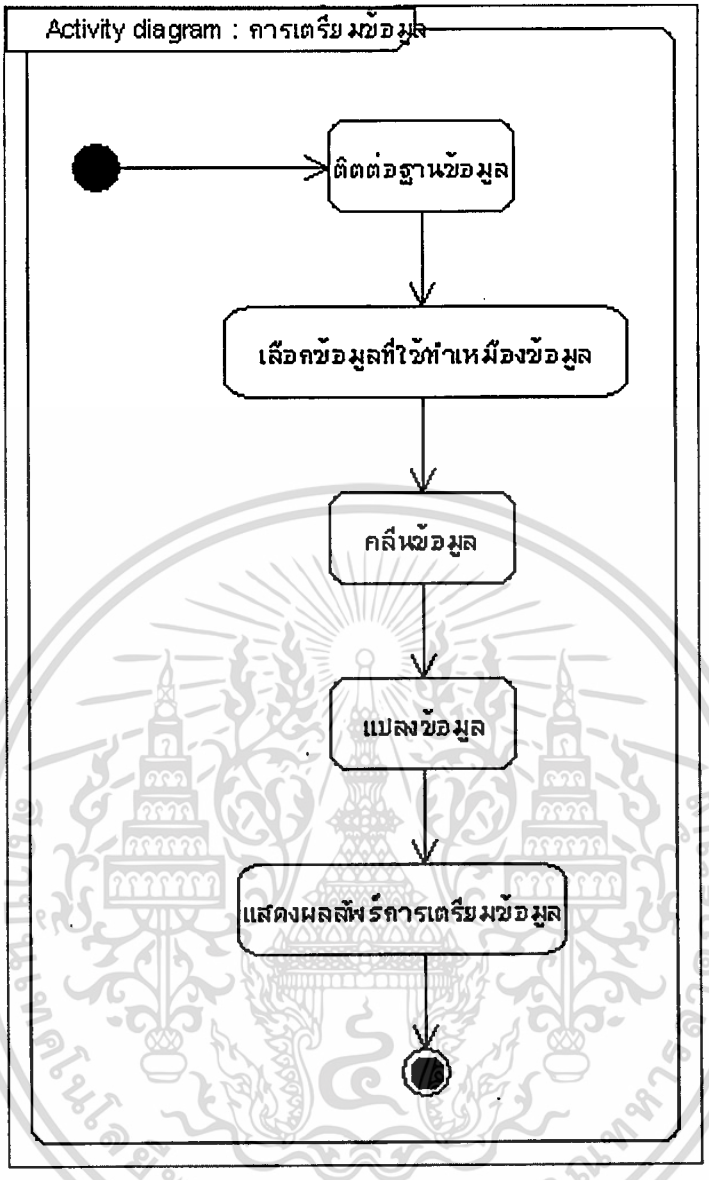
1. ระบบแสดงกฎความสัมพันธ์ที่ค้นหา

รูปที่4.8 ยูสเคสค้นหากฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล

4.2.4 แอคทิวิตีไดอะแกรม (Activity Diagram)

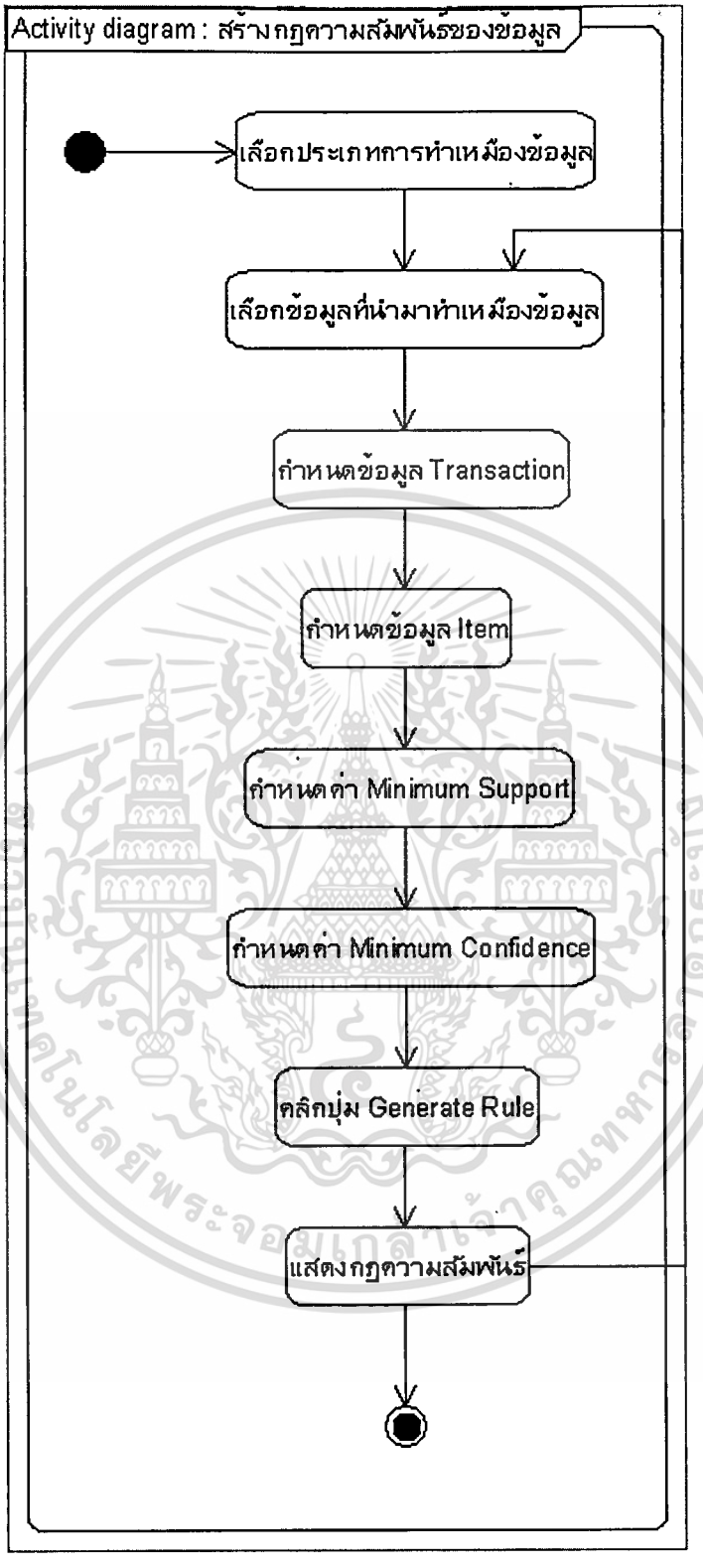
ขั้นตอนของกิจกรรมที่เกิดขึ้นในการพัฒนาระบบสามารถอธิบายได้ด้วยแอคทิวิตีไดอะแกรมดังรูปต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



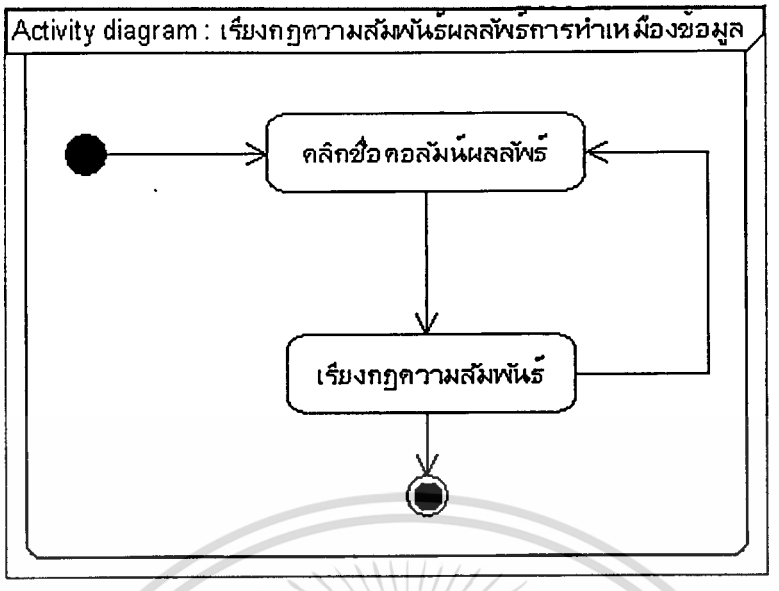
รูปที่4.9 แอคทิวิตีไดอะแกรมการเตรียมข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

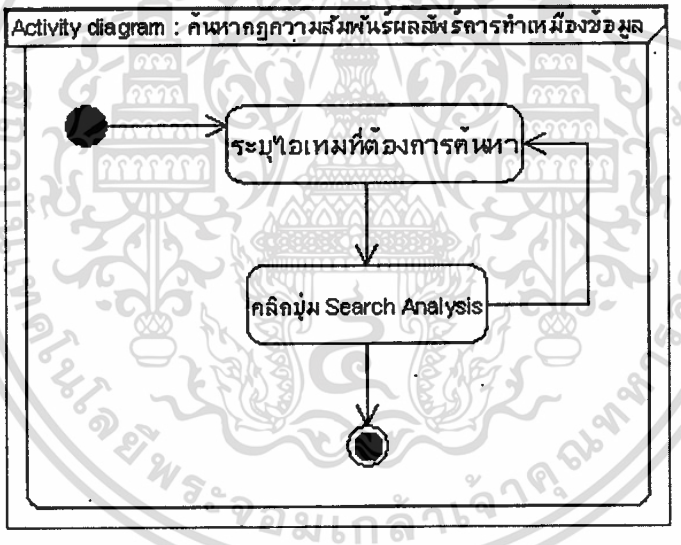


รูปที่4.10 แอคทีวิตีไดอะแกรมสร้างกฎความสัมพันธ์ของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่4.11 แอคทิวิตีไดอะแกรมเรียงกฎความสัมพันธ์ผลลัพธ์การทำเหมืองข้อมูล

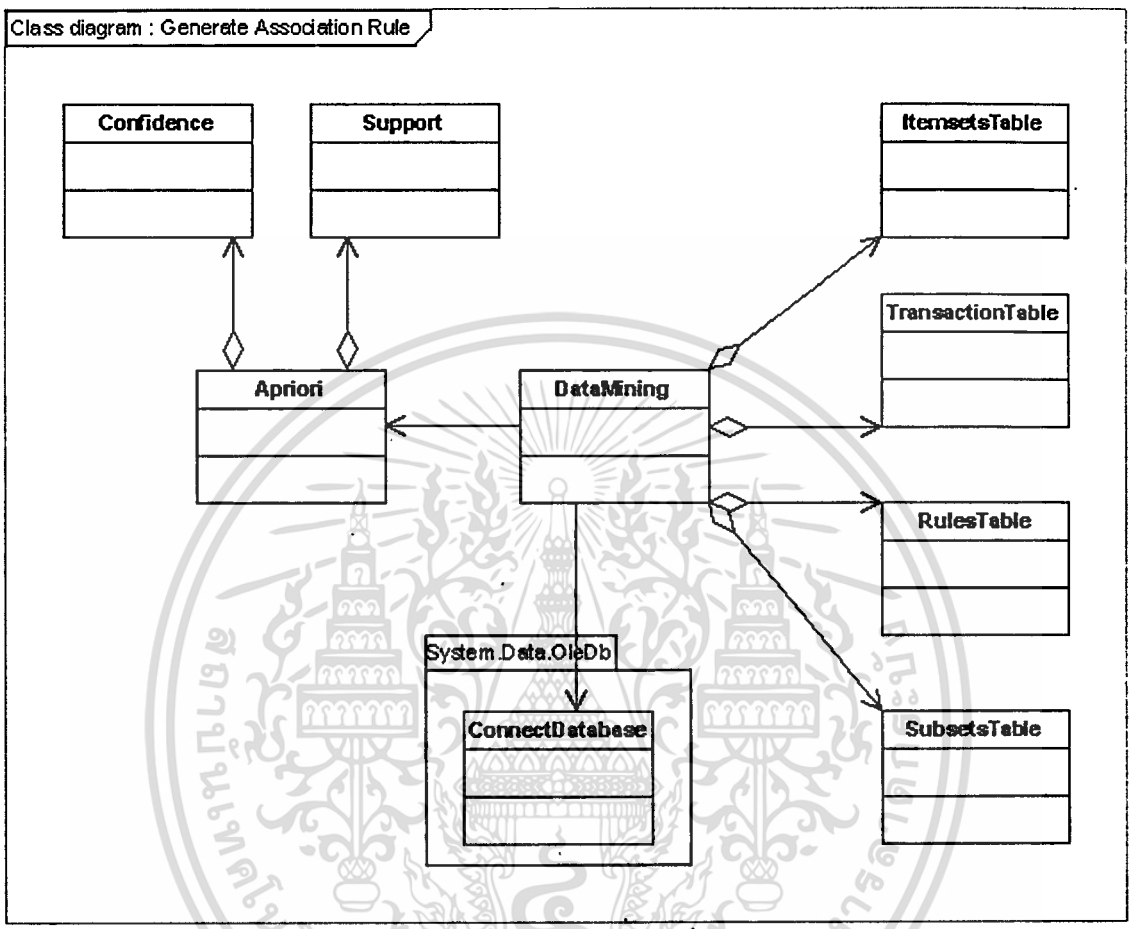


รูปที่4.12 แอคทิวิตีไดอะแกรมค้นหากฎความสัมพันธ์การทำเหมืองข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

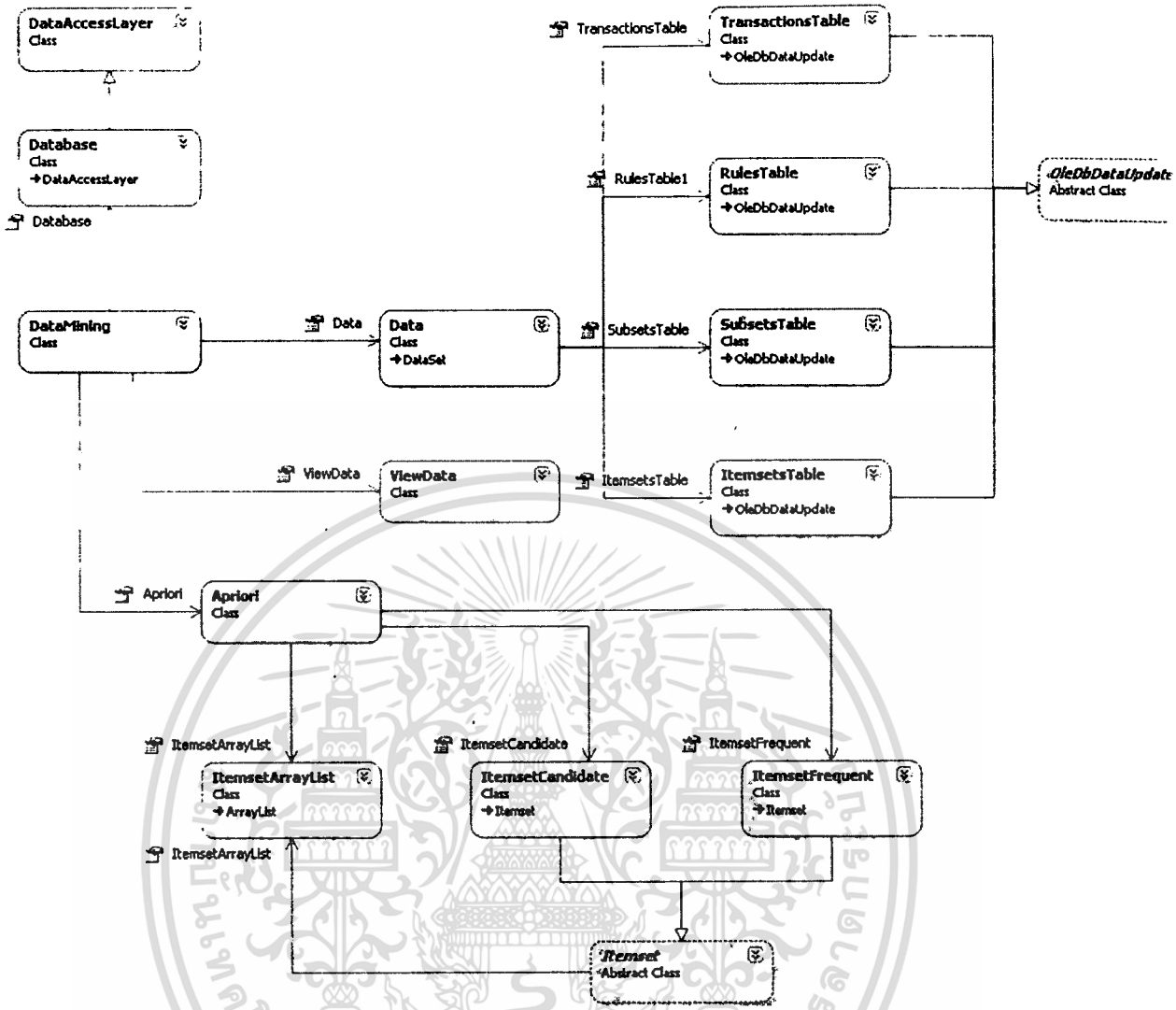
4.3 การออกแบบระบบโดยใช้คลาสไดอะแกรม และซีเควนไดอะแกรม

ภาพแสดงการทำงานของคลาสต่างๆ ที่มีในระบบ โดยมีรายละเอียดดังรูปต่อไปนี้



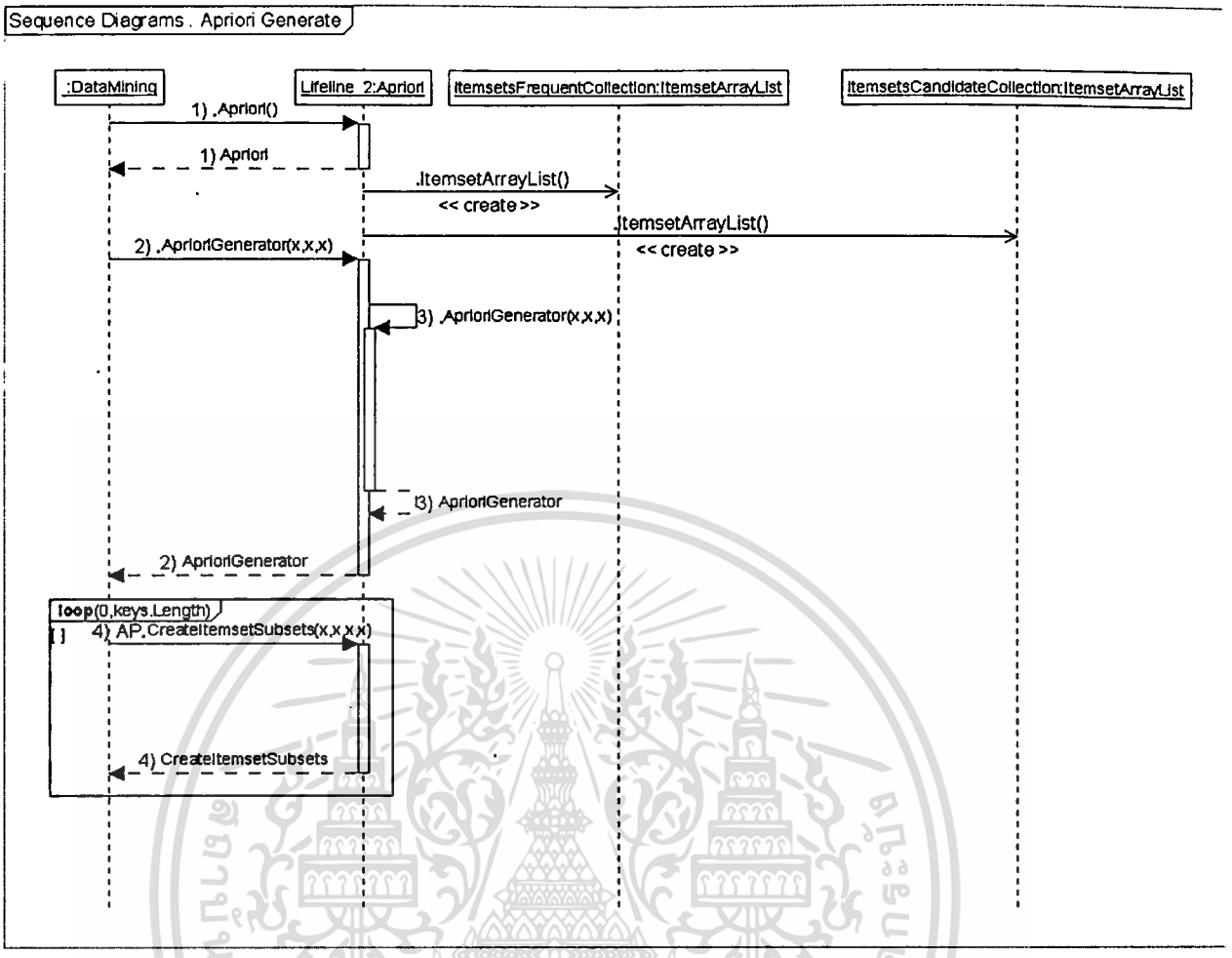
รูปที่4.13 คลาสไดอะแกรมการค้นหากฎความสัมพันธ์ (High Level)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่4.14 คลาสไดอะแกรมระบบการค้นหากฎความสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

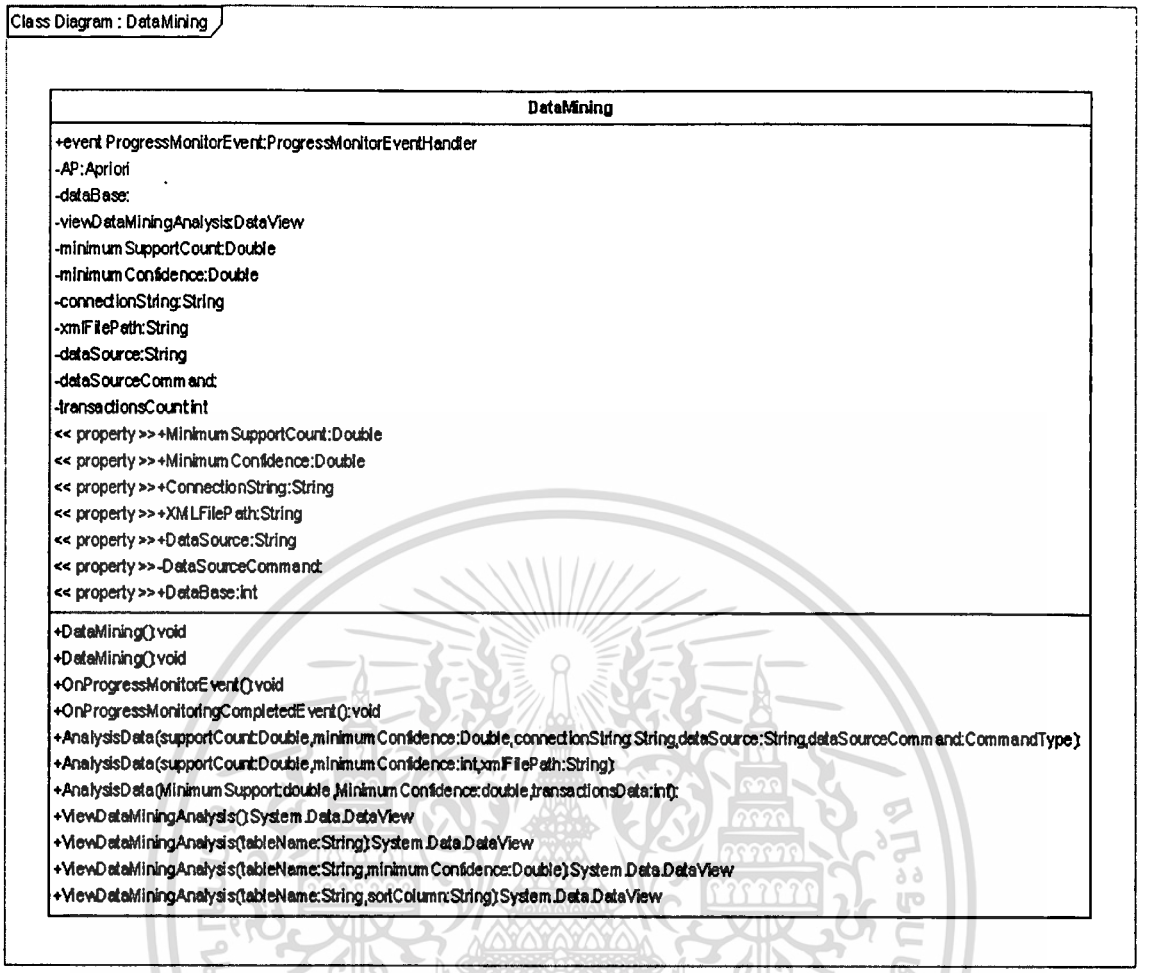


รูปที่4.15 ซีเควนไดอะแกรมการหากฎความสัมพันธ์

จากรูปที่4.15 ซีเควนไดอะแกรมการหากฎความสัมพันธ์ สามารถอธิบายการทำงานได้ดังนี้

1. คลาส DataMining ทำการเรียกไปที่ Constructor Apriori เพื่อสร้างอ๊อบเจ็ค Apriori ซึ่งมีการสร้าง ItemArrayList ที่ใช้สำหรับเก็บค่า Frequent Itemset และ Candidate Itemset
2. คลาส Apriori ทำการเรียกไปที่ Constructor ItemsetArrayList เพื่อสร้างอ๊อบเจ็ค itemsetsFrequentCollection และอ๊อบเจ็ค itemsetsCandidateCollection
3. คลาส DataMining ทำการเรียกใช้งาน Method AprioriGenerator เพื่อทำการสร้าง Candidate Itemset
4. คลาส DataMining ทำการเรียกใช้งาน Method CreateItemsetSubsets ของอ๊อบเจ็ค Apriori เพื่อสร้าง Subset และสร้างเป็นกฎความสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

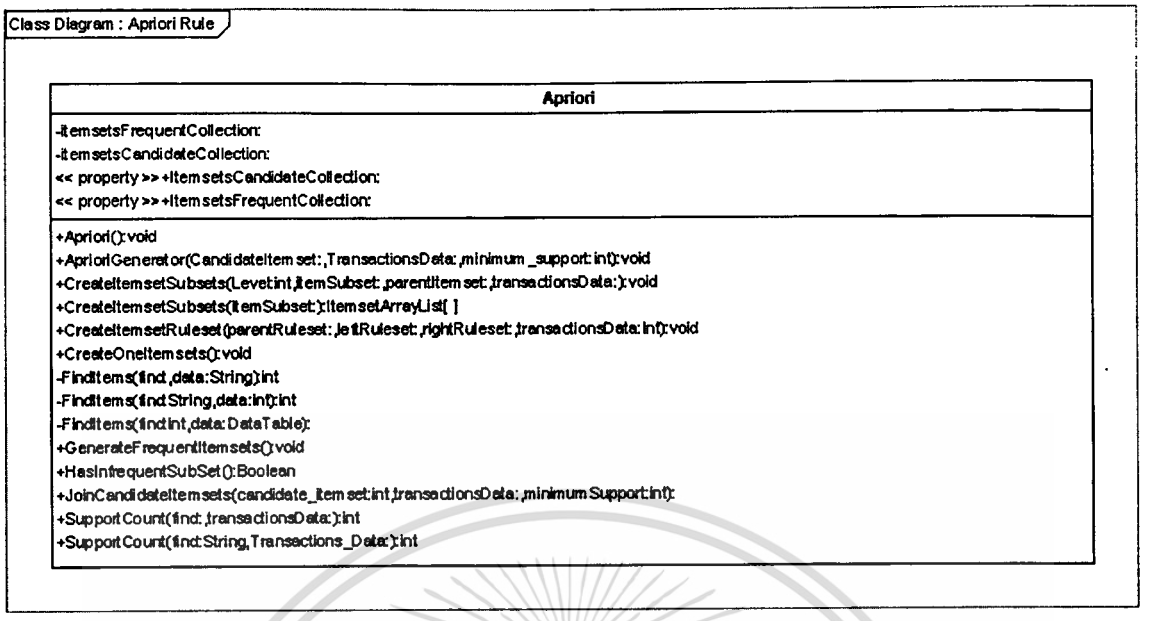


รูปที่4.16 คลาสไดอะแกรม DataMining

จากรูปที่4.16 คลาสไดอะแกรม DataMining มีหน้าที่คือ

1. คำนวณ และกำหนดค่า Support Count
2. คำนวณ และกำหนดค่า Minimum Confidence
3. ส่งค่า Candidate Itemset ไปให้ อ๊อบเจ็ค Apriori ประมวลผล
4. แสดงขั้นตอนการประมวลผลว่าอยู่ขั้นตอนไหนแล้วจาก Event ของ Progress Bar

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

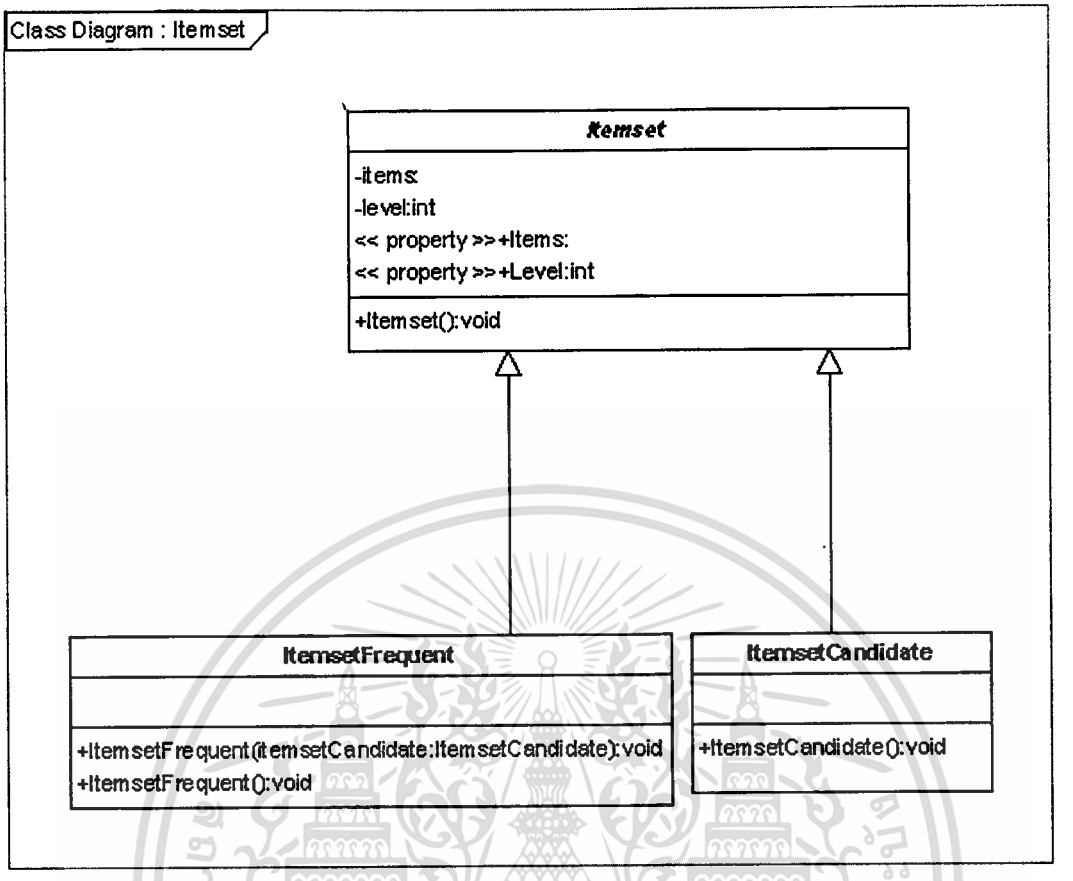


รูปที่4.17 คลาสไดอะแกรม Apriori

จากรูปที่4.17 คลาสไดอะแกรมอพิออริมีหน้าที่คือ

1. นับจำนวนของไอเทม
2. หาค่า Support Count ของไอเทมทรานส์แอ็กชัน
3. สร้างค่า frequent itemset
4. สร้าง itemset โดยทำการ Join candidate itemset กับตัวมันเอง
5. สร้างกฎความสัมพันธ์ระหว่างไอเทม
6. มีการวนซ้ำของการสร้าง subset ทั้งหมดที่อยู่ใน itemset และทำการลงฐานข้อมูล
7. สร้าง subset ของ itemset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

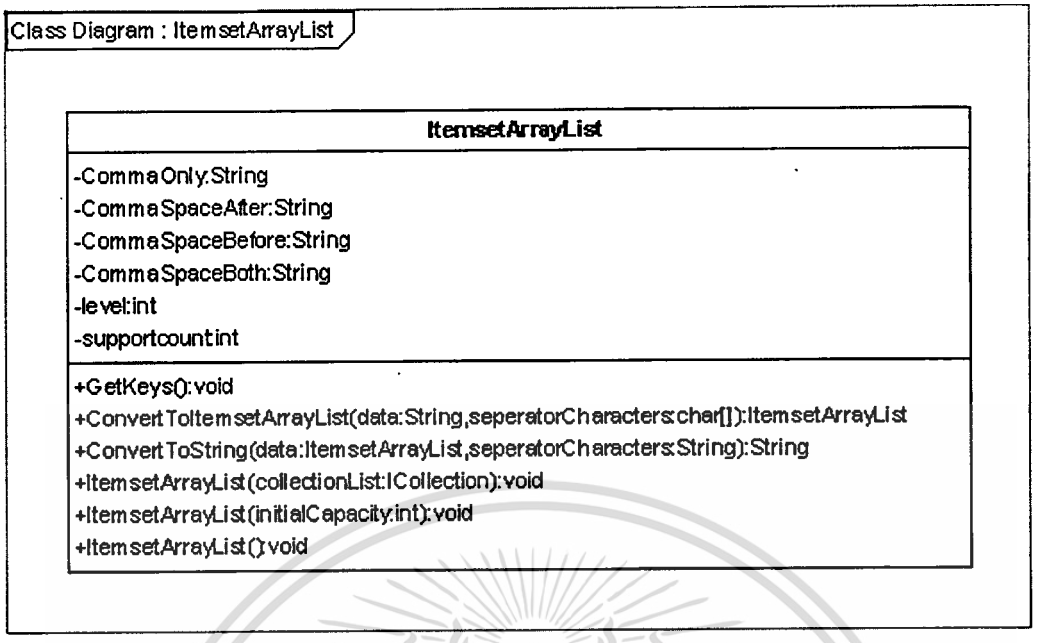


รูปที่4.18 คลาสไดอะแกรม Itemset

จากรูปที่4.18 คลาสไดอะแกรม ItemSet มีหน้าที่คือ

1. เก็บข้อมูลของ Frequent Itemset
2. เก็บข้อมูลของ Candidate Itemset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

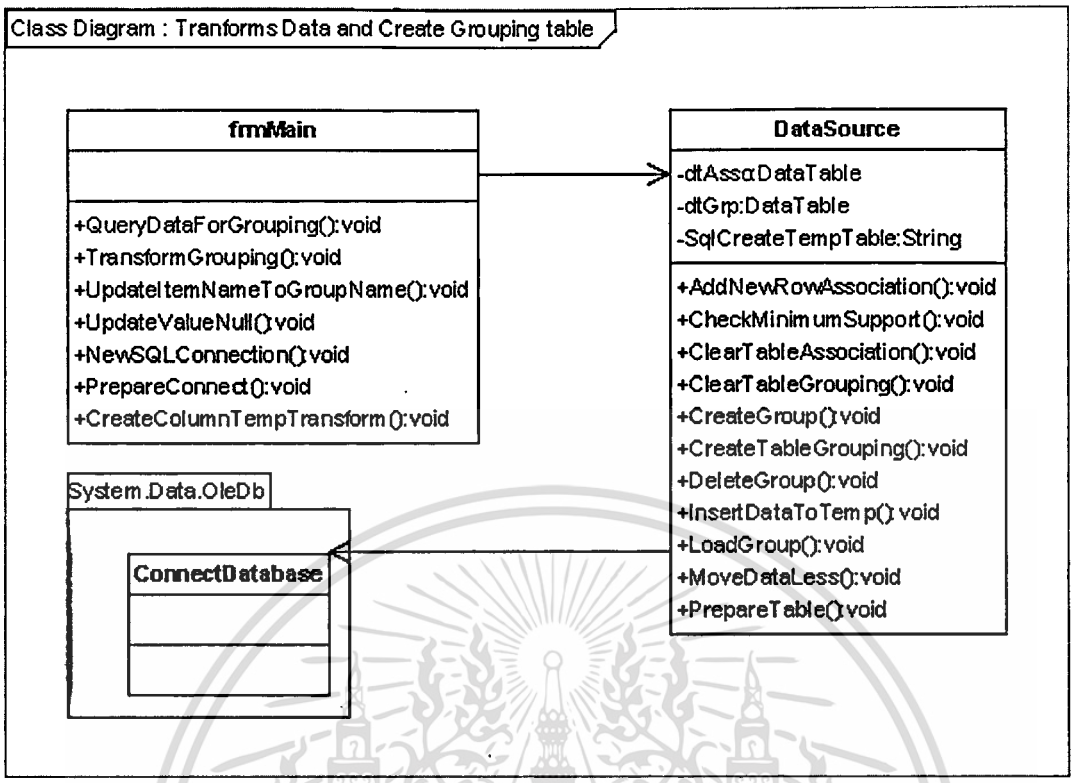


รูปที่4.19 คลาสไดอะแกรม ItemsetArrayList

จากรูปที่4.19 คลาสไดอะแกรม ItemsetArrayList เป็นคลาสที่สืบทอดคุณสมบัติมาจากคลาส ArrayList และมีหน้าที่คือ

1. สร้างและกำหนดค่าเริ่มต้นให้อินสแตนซ์ของคลาส ItemArrayList
2. แปลงค่าตัวอักษรให้อยู่ในรูปแบบของคลาส ItemArrayList
3. แปลงข้อมูลของอ๊อบเจ็ค ItemArrayList ให้อยู่ในรูปแบบของตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

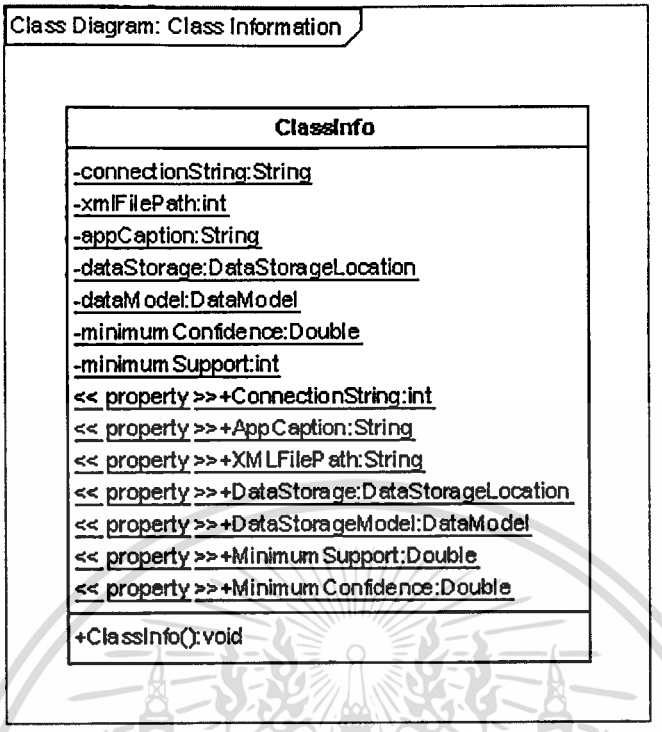


รูปที่4.20 คลาสไดอะแกรม Grouping table

จากรูปที่4.20 คลาสไดอะแกรม Grouping table มีหน้าที่คือ

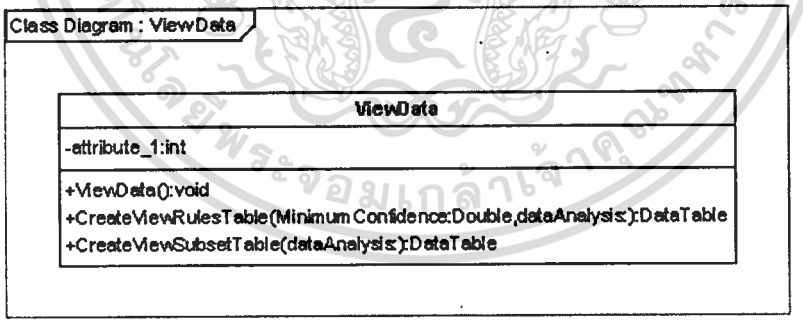
1. สร้างตารางสำหรับเก็บข้อมูลจากกลุ่มข้อมูลที่ได้เลือกไว้
2. ตรวจสอบชื่อข้อมูลที่ทำการจัดกลุ่มและเปลี่ยนแปลงให้เป็นชื่อกลุ่มที่กำหนด
3. นำข้อมูลที่เปลี่ยนเป็นชื่อกลุ่มที่กำหนดแล้วเก็บลงฐานข้อมูล
4. ลบ แก้ไขและแสดงข้อมูลที่ทำการจัดกลุ่มแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่4.21 คลาสไดอะแกรม Class Information

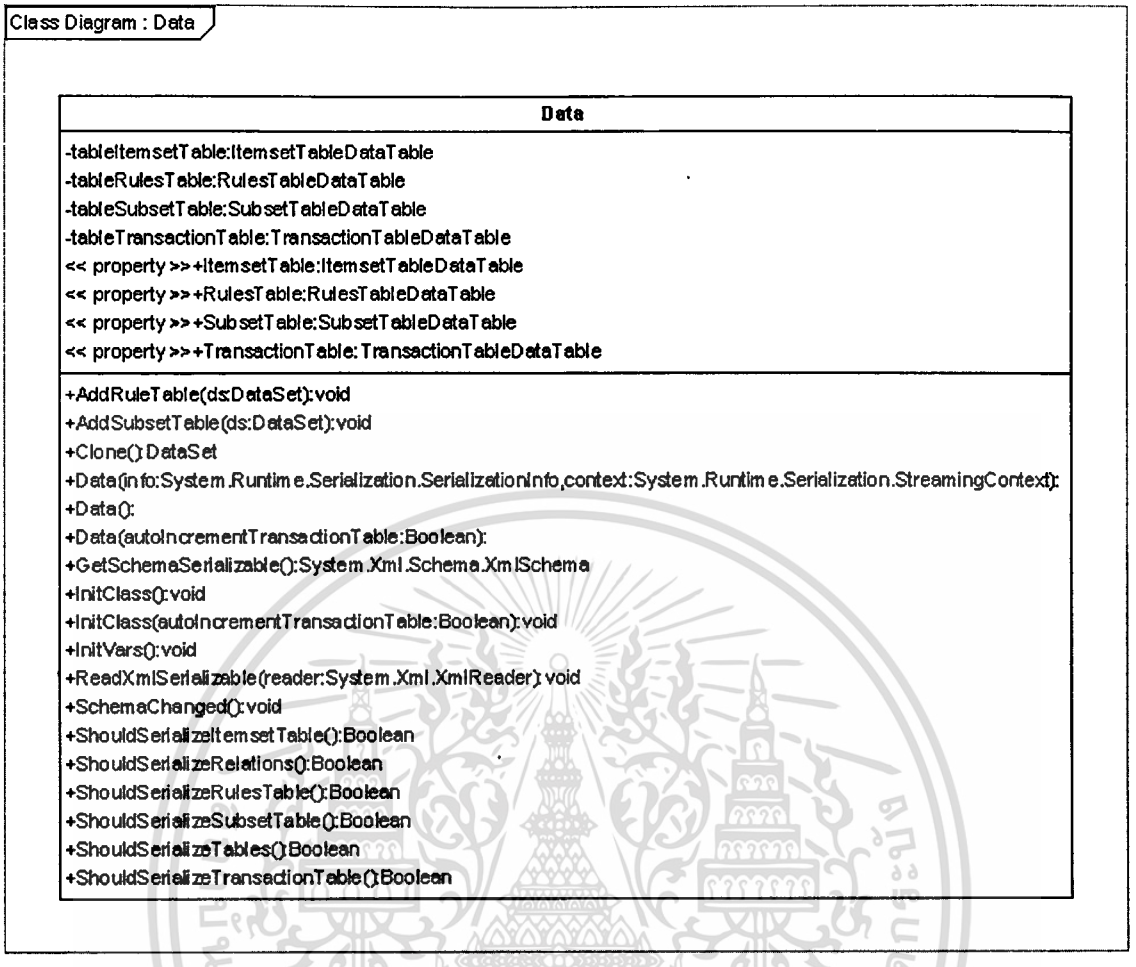
จากรูปที่4.21 คลาสไดอะแกรม Class Information มีหน้าที่คือ เก็บข้อมูลที่ทุกคลาสใช้งานร่วมกัน เช่น Connection String สำหรับติดต่อระบบฐานข้อมูล



รูปที่4.22 คลาสไดอะแกรม View Data

จากรูปที่4.22 คลาสไดอะแกรม View Data มีหน้าที่คือ นำข้อมูลผลลัพธ์ที่ได้จากการทำ Mining มาสร้างเป็น DataTable เพื่อใช้สำหรับแสดงผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

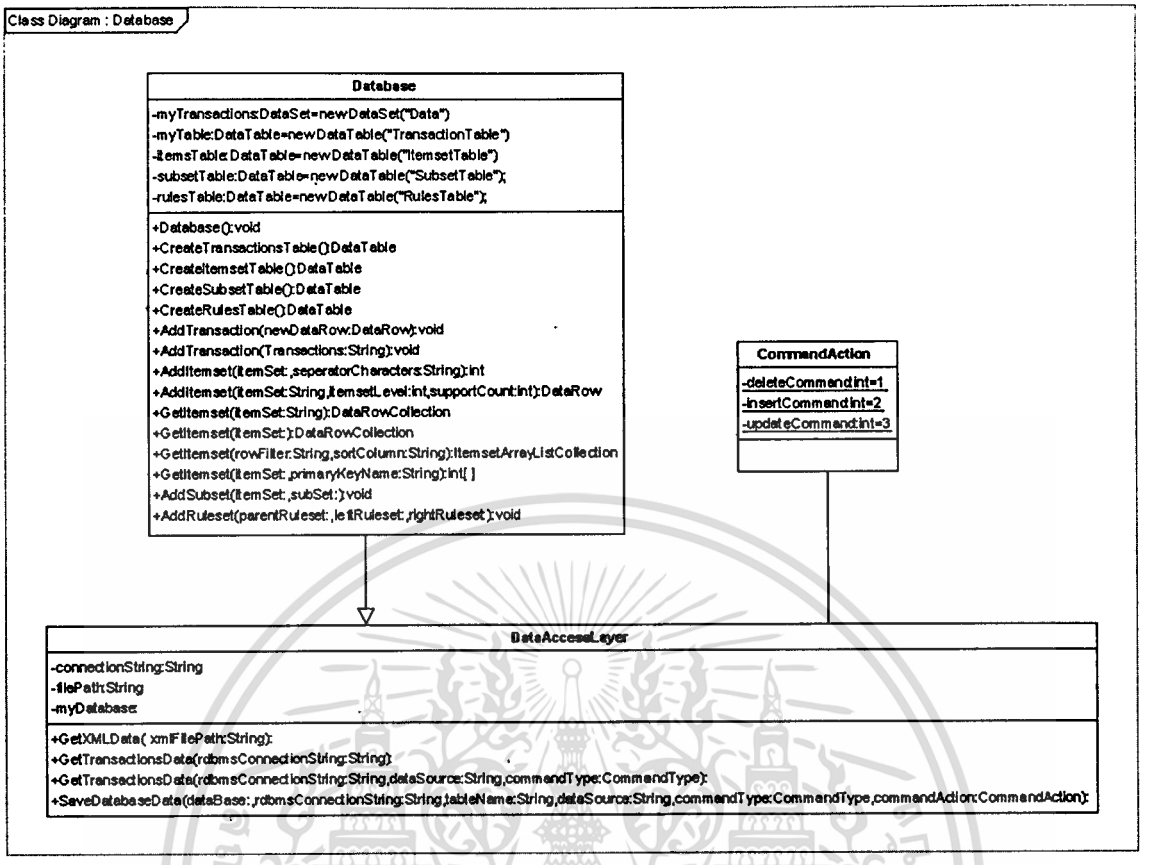


รูปที่4.23 คลาสไดอะแกรม Data

จากรูปที่4.23 คลาสไดอะแกรม Data มีหน้าที่คือ

1. กำหนดโครงสร้าง และค่าเริ่มต้นให้กับอ๊อบเจ็คของคลาส TransactionTableDataTable
2. กำหนดโครงสร้าง และค่าเริ่มต้นให้กับอ๊อบเจ็คของคลาส ItemsetTableDataTable
3. กำหนดโครงสร้าง และค่าเริ่มต้นให้กับอ๊อบเจ็คของคลาส SubsetTableDataTable
4. กำหนดโครงสร้าง และค่าเริ่มต้นให้กับอ๊อบเจ็คของคลาส RulesTableDataTable

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่4.24 คลาสไดอะแกรม Database

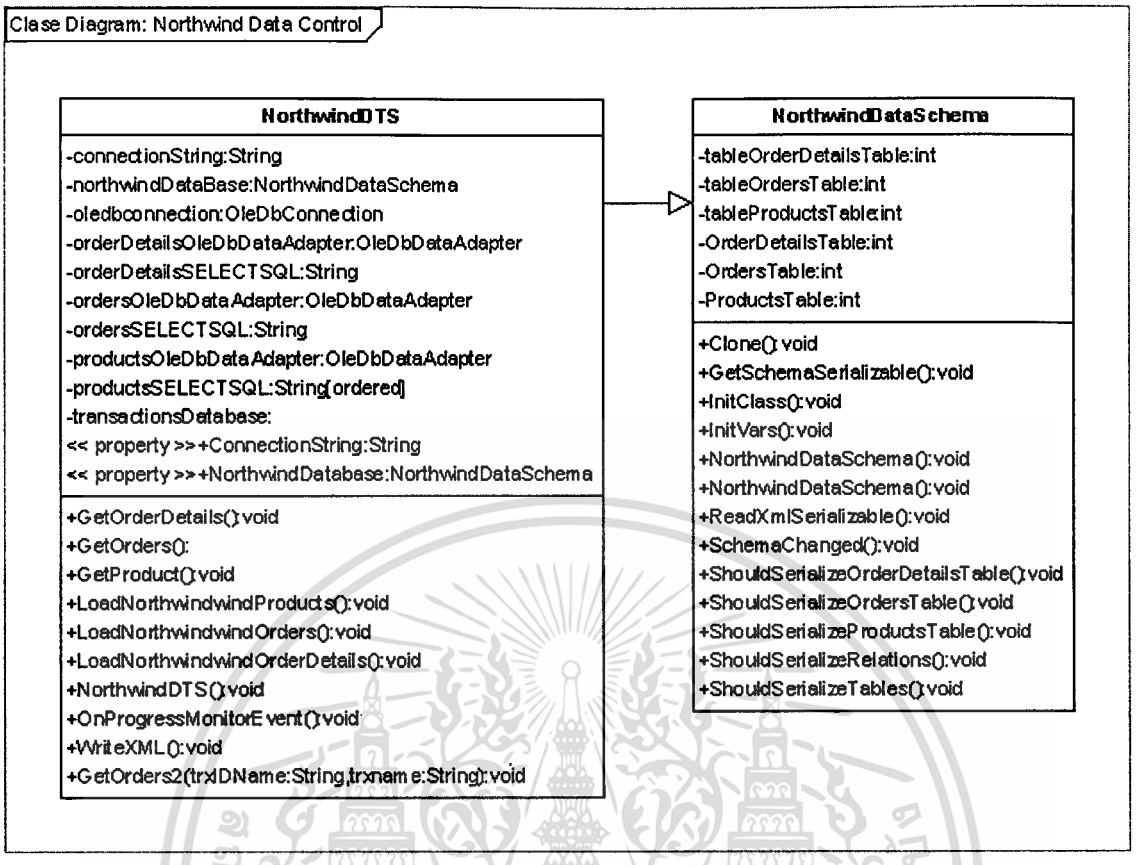
จากรูปที่4.24 คลาสไดอะแกรม Database มีหน้าที่คือ

1. สร้างตาราง Transaction สำหรับเก็บข้อมูล Transaction ต่างๆ โดยกำหนดให้รายการในแถวใช้เครื่องหมาย "," การแยก
2. สร้างตาราง ItemSet เพื่อกำหนดเป็นที่ใช้เก็บค่า Level และ Support Count
3. สร้างตาราง Subset
4. สร้างตาราง Rules เพื่อเก็บค่าต่างๆ พร้อมทั้งคำนวณค่า Confidence และค่า Support

จากรูปที่4.18 คลาสไดอะแกรม DataAccessLayer มีหน้าที่คือ

1. ติดต่อข้อมูลจากไฟล์เอกสาร XML
2. ติดต่อข้อมูลจากระบบฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่4.25 คลาสไดอะแกรม Northwind Data Control

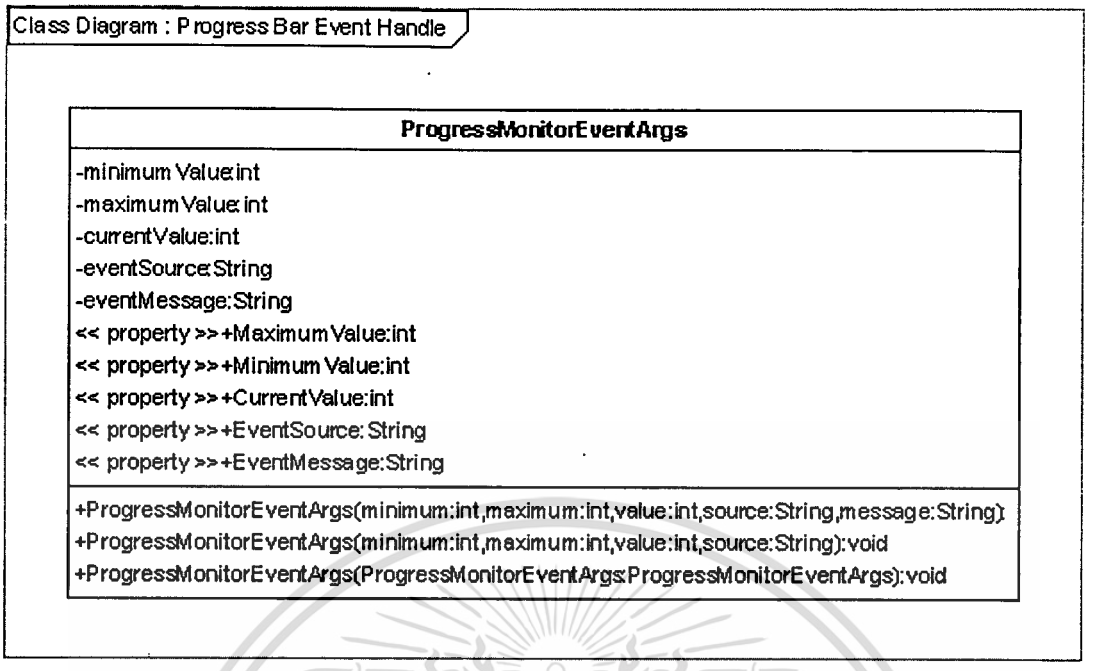
จากรูปที่4.25 คลาสไดอะแกรม Northwind Data Control มีหน้าที่คือ

1. ติดต่อ และดึงข้อมูลจากตาราง Product Order และ OrderDetail มาเก็บไว้

2. สร้างตารางข้อมูลเพื่อเก็บลงหน่วยความจำสำหรับนำไปใช้สร้างตาราง Transaction สำหรับประมวลผล

3. ตรวจสอบ และเปรียบเทียบโครงสร้างข้อมูลเพื่อใช้สำหรับการทำ Mining

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่4.26 คลาสไดอะแกรม ProgressBar Event Handle

จากรูปที่4.26 คลาสไดอะแกรม ProgressBar Event Handle มีหน้าที่คือ

1. เชื่อมโยงเหตุการณ์เข้ากับ Progress Bar
2. สั่งงาน Progress Bar ให้มีการแก้ไขสถานะค่า Minimum และ Maximum

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่5

ออกแบบ และพัฒนาระบบ

5.1 เครื่องมือที่ใช้ในการออกแบบ และพัฒนาระบบ

5.1.1 ฮาร์ดแวร์

การพัฒนาระบบงานใช้เครื่องคอมพิวเตอร์ที่มีคุณสมบัติดังนี้

- หน่วยประมวลผลกลาง Pentium M ความเร็ว 1.5 GHz
- ฮาร์ดดิสก์ความจุขนาด 40 GB
- หน่วยความจำ 512 MB

5.1.2 ซอฟต์แวร์

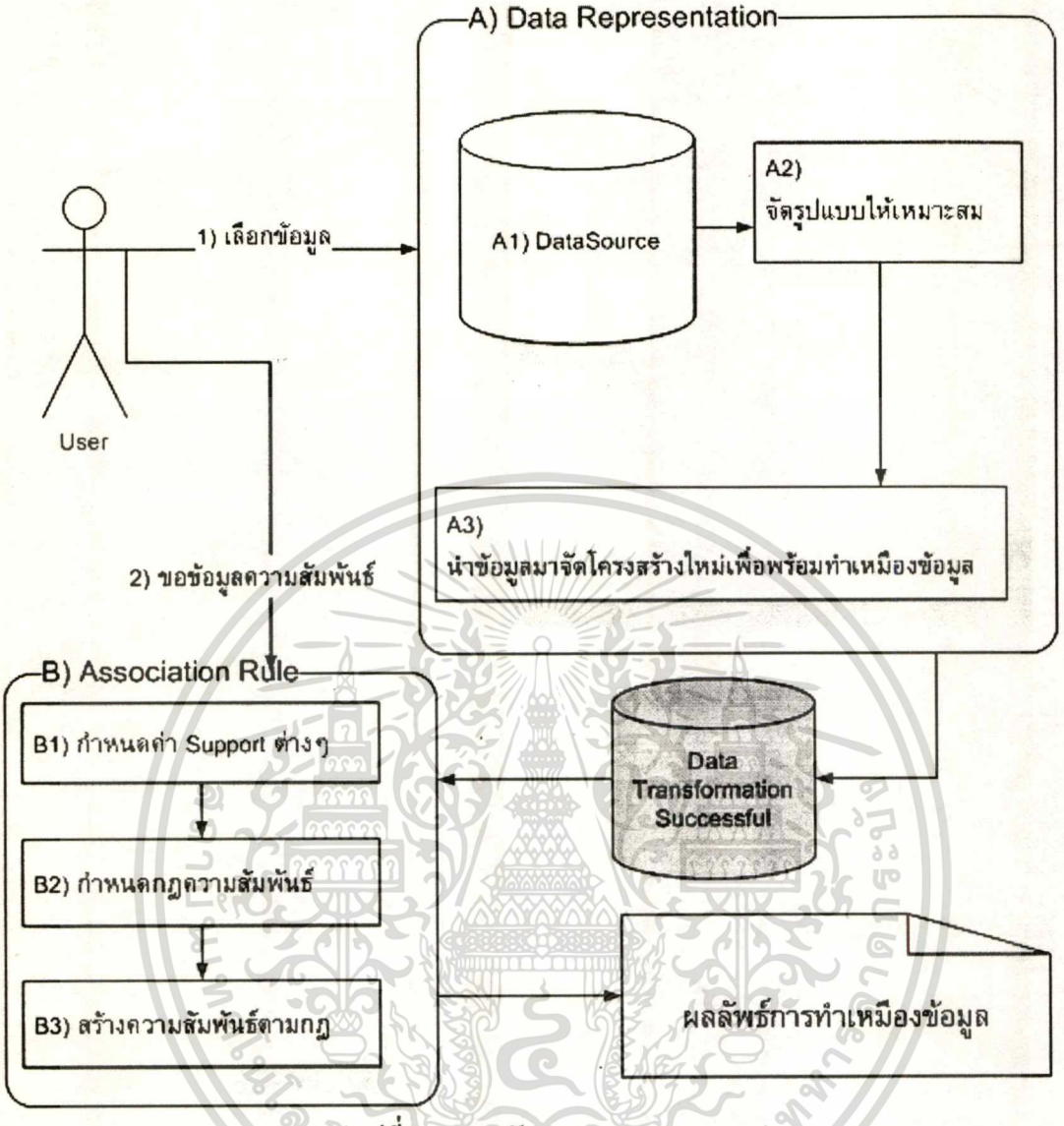
เครื่องมือ และโปรแกรมที่ใช้ในการพัฒนามีรายละเอียดดังนี้

- ระบบปฏิบัติการ คือ Windows XP SP2
- ภาษาที่ใช้ในการพัฒนา คือ C#.NET และ VB.NET (Framework 2.0)
- เครื่องมือที่ใช้ในการพัฒนา(IDE) คือ Visual Studio .NET 2005
- ระบบฐานข้อมูล คือ Microsoft SQL Server 2005 Developer Edition
- การสร้างไฟล์จากส่วนการวิเคราะห์ และออกแบบ คือ Poseidon for UML Professional v4.12

5.2 สถาปัตยกรรมระบบ

ในการออกแบบสถาปัตยกรรมของระบบได้แบ่งออกเป็น 2 ส่วนใหญ่ๆ คือ Data Representation และ Association Rule โดยทั้ง 2 ส่วนมีความต่อเนื่องกัน ดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

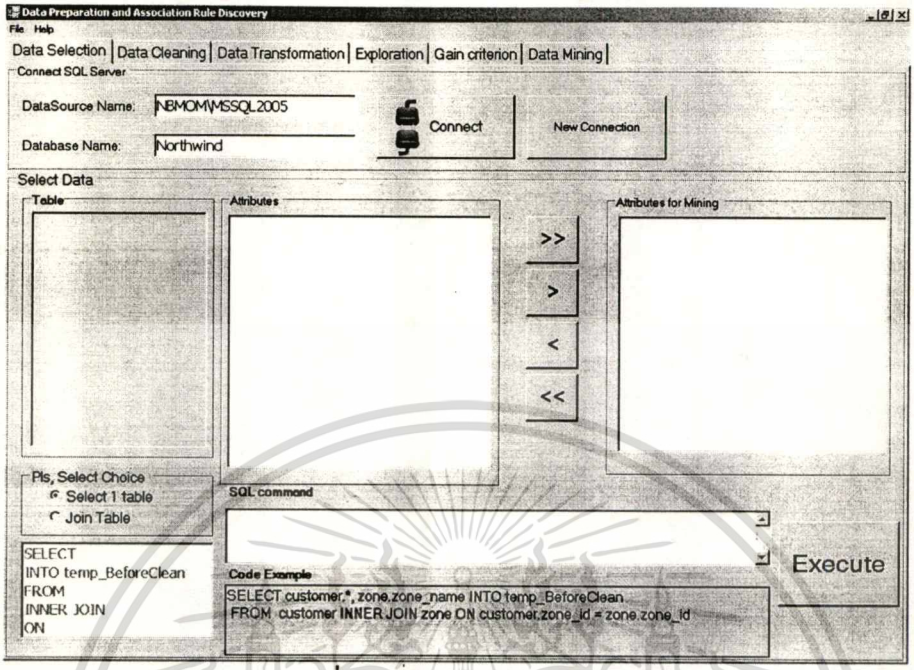


รูปที่ 5.1 สถาปัตยกรรมระบบ

จากรูปที่5.1 เป็นการออกแบบสถาปัตยกรรมของระบบ โดยผู้ใช้งานทำการเลือกข้อมูลจากฐานข้อมูลเพื่อนำข้อมูลไปจัดรูปแบบ และจัดโครงสร้างใหม่ให้เหมาะสม เพื่อพร้อมใช้ในการทำเหมืองข้อมูล ซึ่งผู้ใช้งานจะสั่งให้กระบวนการประมวลผลการทำเหมืองข้อมูลเริ่มทำงาน โดยใช้ข้อมูลที่จัดเตรียมไว้แล้ว เมื่อระบบทำงานเสร็จสิ้นก็จะได้ผลลัพธ์การทำเหมืองข้อมูลออกมา

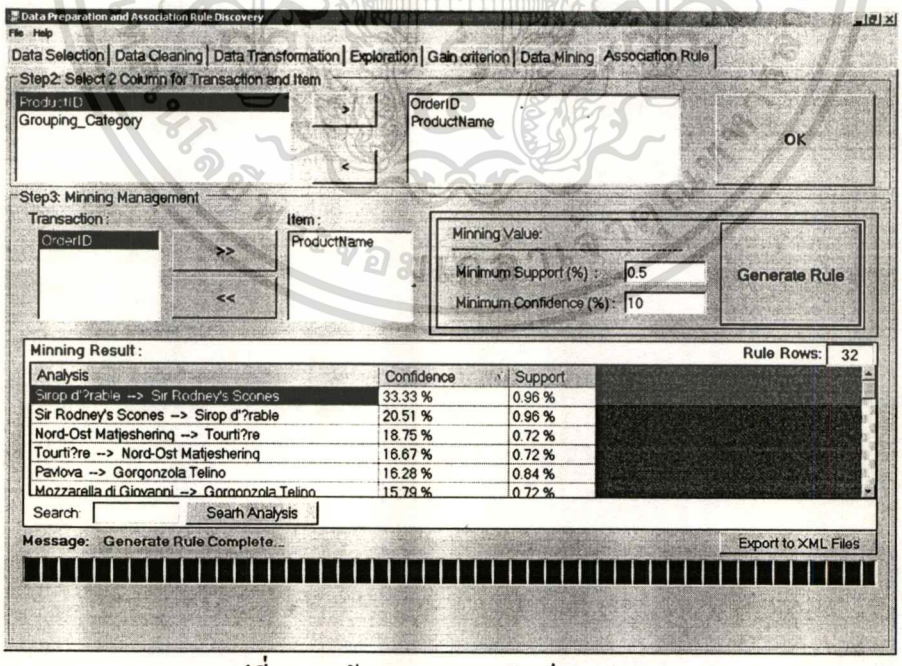
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 การออกแบบหน้าจอการรับข้อมูล



รูปที่5.2 หน้าจอการรับข้อมูล

5.4 การออกแบบหน้าจอรายงานผลลัพธ์



รูปที่5.3 หน้าจอแสดงผลลัพธ์การประมวลผล

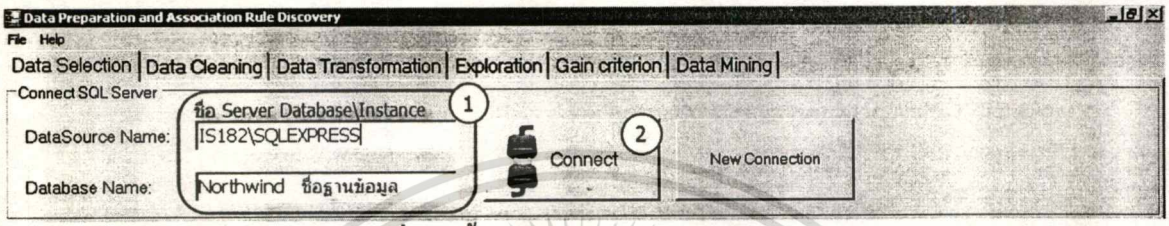
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.5 การใช้งานโปรแกรม

การทำงานของระบบจะเป็นไปตามกระบวนการของการทำเหมืองข้อมูล ดังนี้

5.5.1 ขั้นตอนการติดต่อฐานข้อมูล

1. ผู้ใช้งานกำหนด ชื่อเซิฟเวอร์ และชื่อฐานข้อมูล
2. คลิกปุ่ม Connect เพื่อเชื่อมต่อกับฐานข้อมูล



รูปที่5.4 ขั้นตอนการติดต่อฐานข้อมูล

3. ระบบแสดงกล่องข้อความให้ทราบว่าได้ทำการเชื่อมต่อกับฐานข้อมูลเรียบร้อย

รูปที่5.5 กล่องข้อความแสดงการผลการเชื่อมต่อฐานข้อมูล

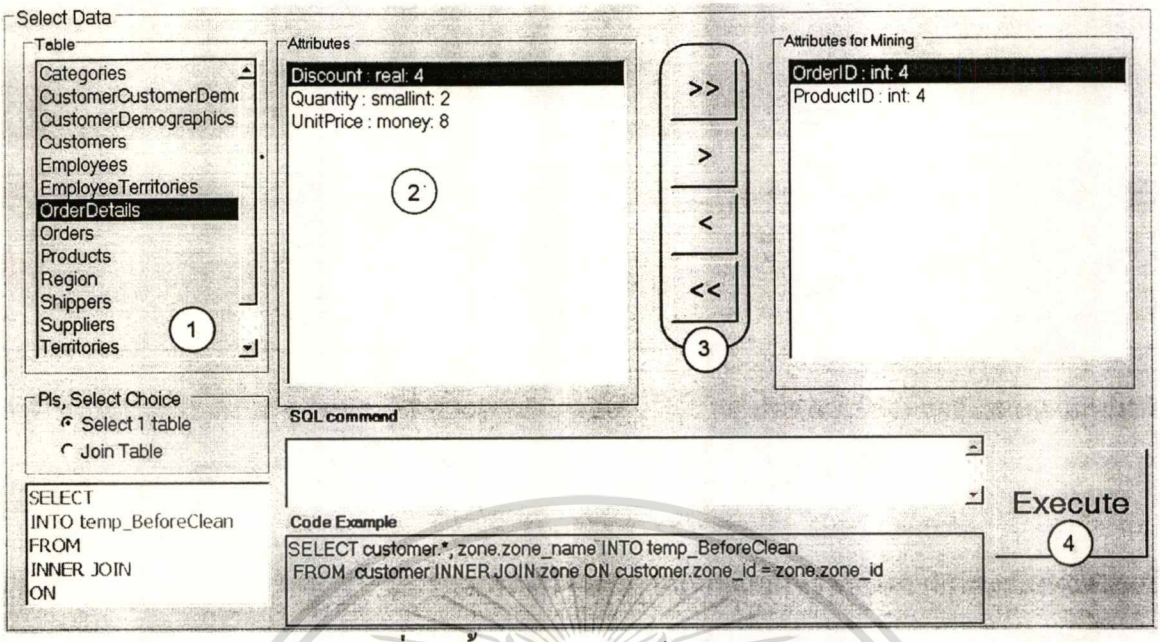
5.5.2 ขั้นตอนการเลือกข้อมูล

ในการขั้นตอนการเลือกข้อมูลสามารถเลือกข้อมูลได้ 2 กรณีคือ กรณีที่ข้อมูลมาจาก 1 ตารางและกรณีที่ข้อมูลมาจากมากกว่า 1 ตารางในฐานข้อมูลเดียวกัน โดยทั้ง 2 กรณีมีวิธีเลือกดังนี้

กรณีที่1 เลือกข้อมูลจาก 1 ตาราง

1. ผู้ใช้งานคลิกเลือกตาราง
2. ผู้ใช้งานเลือกชื่อแอตทริบิวที่จะใช้ในการ
3. ผู้ใช้งานคลิกปุ่มลูกศรเพื่อเลือกชื่อแอตทริบิวที่ได้เลือกไว้
4. คลิกปุ่ม Execute เพื่อสิ้นสุดขั้นตอนการเลือกข้อมูล

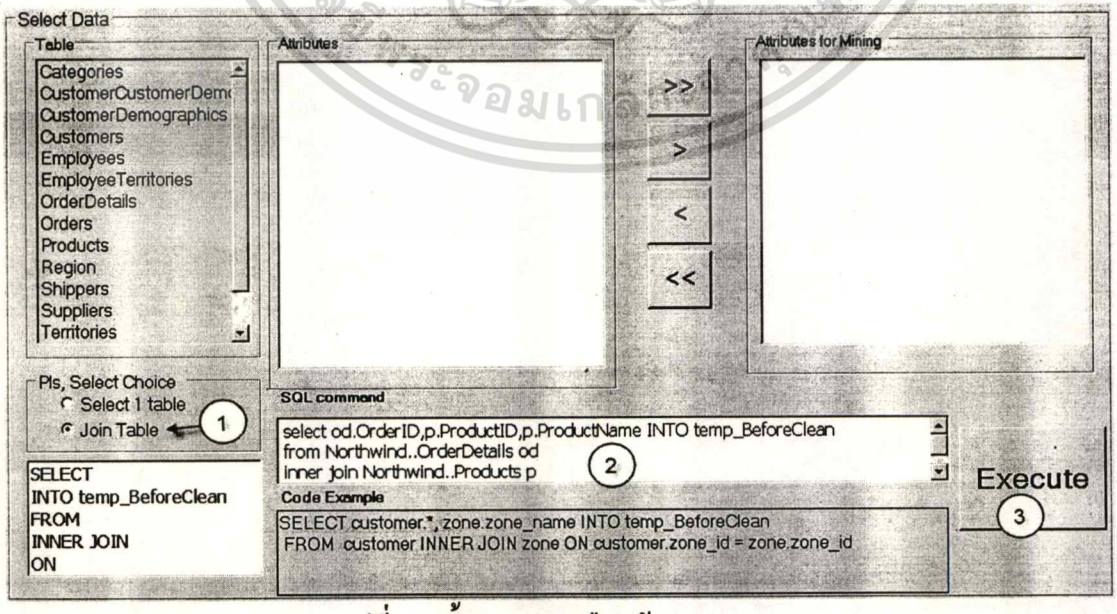
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่5.6 ขั้นตอนการเลือกข้อมูลจาก 1 ตาราง

กรณีที่2 เลือกข้อมูลจากหลายตาราง

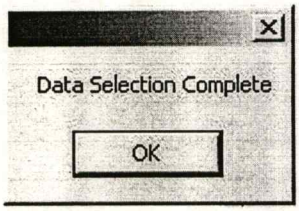
1. ผู้ใช้งานเลือก Join Table ในช่อง Pls, Select Choice
2. ผู้ใช้งานพิมพ์คำสั่งของ SQL ลงในช่อง SQL Command เพื่อเลือกข้อมูลที่ต้องการ โดยผู้ใช้ระบบต้องทราบความสัมพันธ์ของข้อมูลในแต่ละตาราง และสามารถใช้คำสั่ง SQL พื้นฐานในการเชื่อมความสัมพันธ์เหล่านั้น เพื่อเลือกข้อมูลในการทำเหมืองข้อมูล
3. ผู้ใช้งานคลิกปุ่ม Execute เพื่อสิ้นสุดขั้นตอนการเลือกข้อมูล



รูปที่5.7 ขั้นตอนการเลือกข้อมูลจากหลายตาราง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

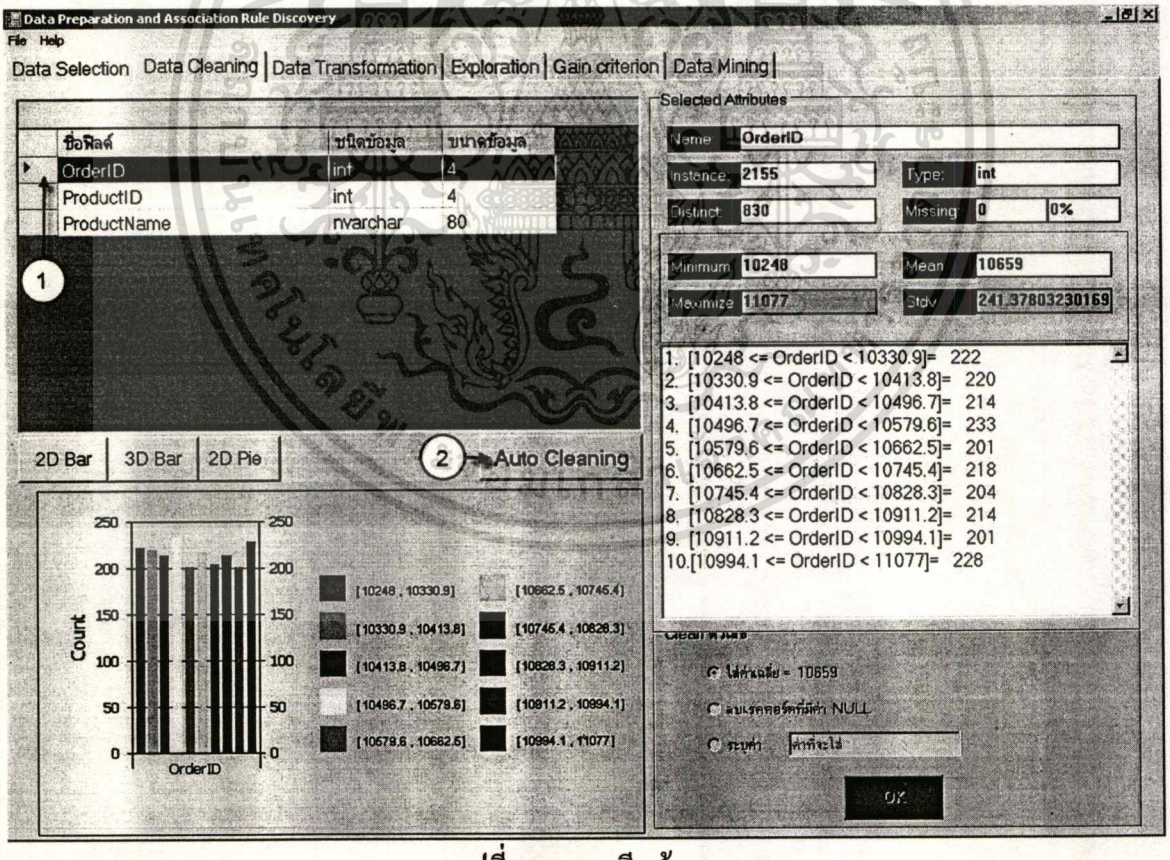
4. ระบบแสดงกล่องข้อความบอกให้ผู้ใช้งานทราบว่าได้เสร็จสิ้นขั้นตอนการเลือกข้อมูลแล้ว



รูปที่5.8 กล่องข้อความแสดงผลการเลือกข้อมูล

5.5.3 ขั้นตอนการคลีนข้อมูล

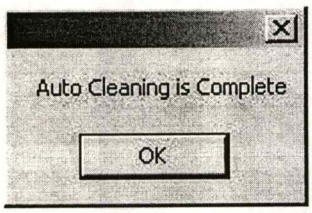
1. ผู้ใช้งานคลิกเลือกแอตทริบิว โดยระบบจะแสดงรายละเอียดของแอตทริบิวและแสดงกราฟของจำนวนข้อมูลในแอตทริบิวที่ผู้ใช้เลือก
2. ผู้ใช้งานคลิกปุ่ม Auto Cleaning เพื่อทำการคลีนข้อมูลที่ได้เลือกไว้



รูปที่5.9 การคลีนข้อมูล

3. ระบบแสดงข้อความให้ทราบว่าทำการคลีนข้อมูลเรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

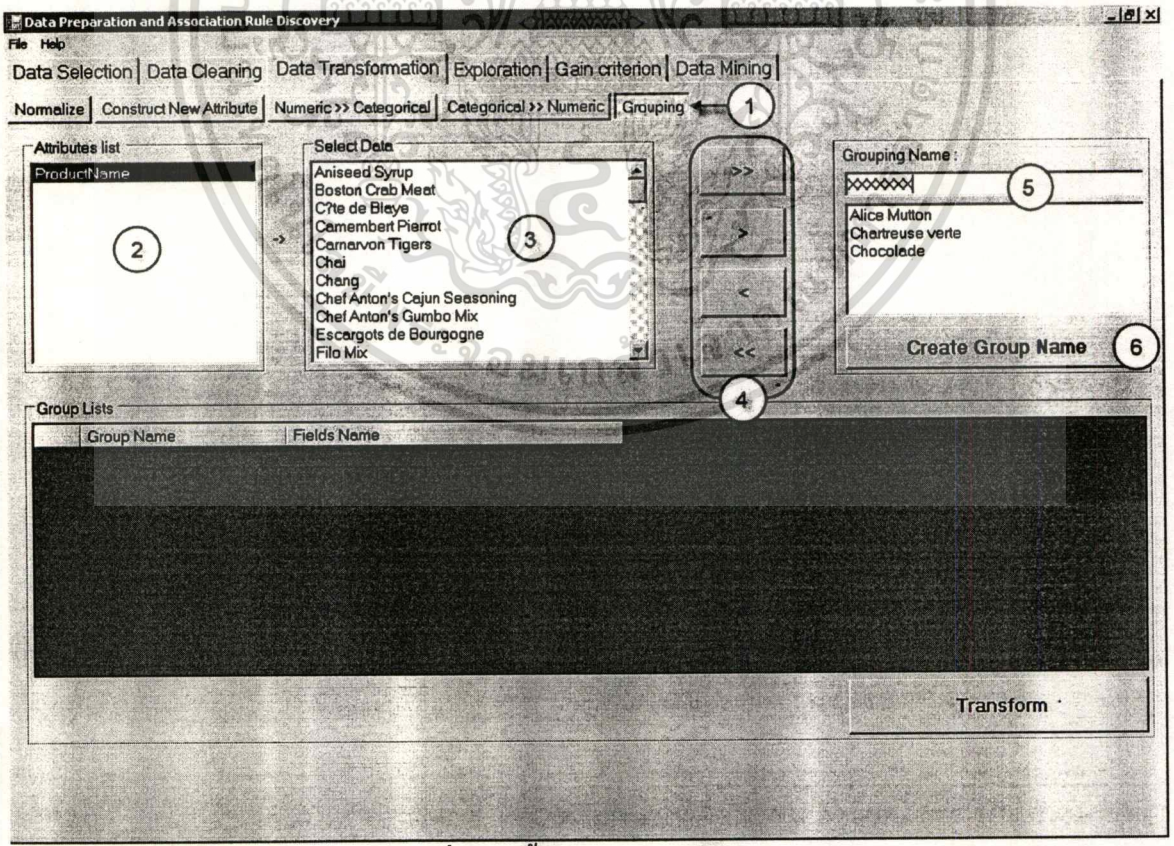


รูปที่5.10 กล่องข้อความแสดงผลการคลีนข้อมูล

5.5.4 ขั้นตอนการแปลงข้อมูล

ในขั้นตอนการแปลงข้อมูลสำหรับการทำเหมืองข้อมูลแบบค้นหากฎความสัมพันธ์ ในที่นี้จะกล่าวถึงกระบวนการจัดกลุ่มข้อมูล โดยมีขั้นตอนดังนี้

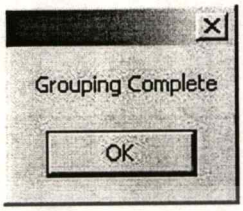
1. ผู้ใช้งานคลิกปุ่ม Grouping เพื่อเลือกประเภทการแปลงข้อมูล
2. ผู้ใช้งานเลือกแอตทริบิวที่ต้องการจะนำมาจัดกลุ่ม
3. ผู้ใช้งานเลือกข้อมูลที่จะนำมาจัดกลุ่ม
4. ผู้ใช้งานคลิกปุ่มลูกศรเพื่อเลือกข้อมูลที่ได้เลือกไว้
5. ผู้ใช้กำหนดชื่อกลุ่มข้อมูล
6. คลิกปุ่ม Create Group Name เพื่อจัดกลุ่มข้อมูล



รูปที่5.11 ขั้นตอนการจัดกลุ่ม

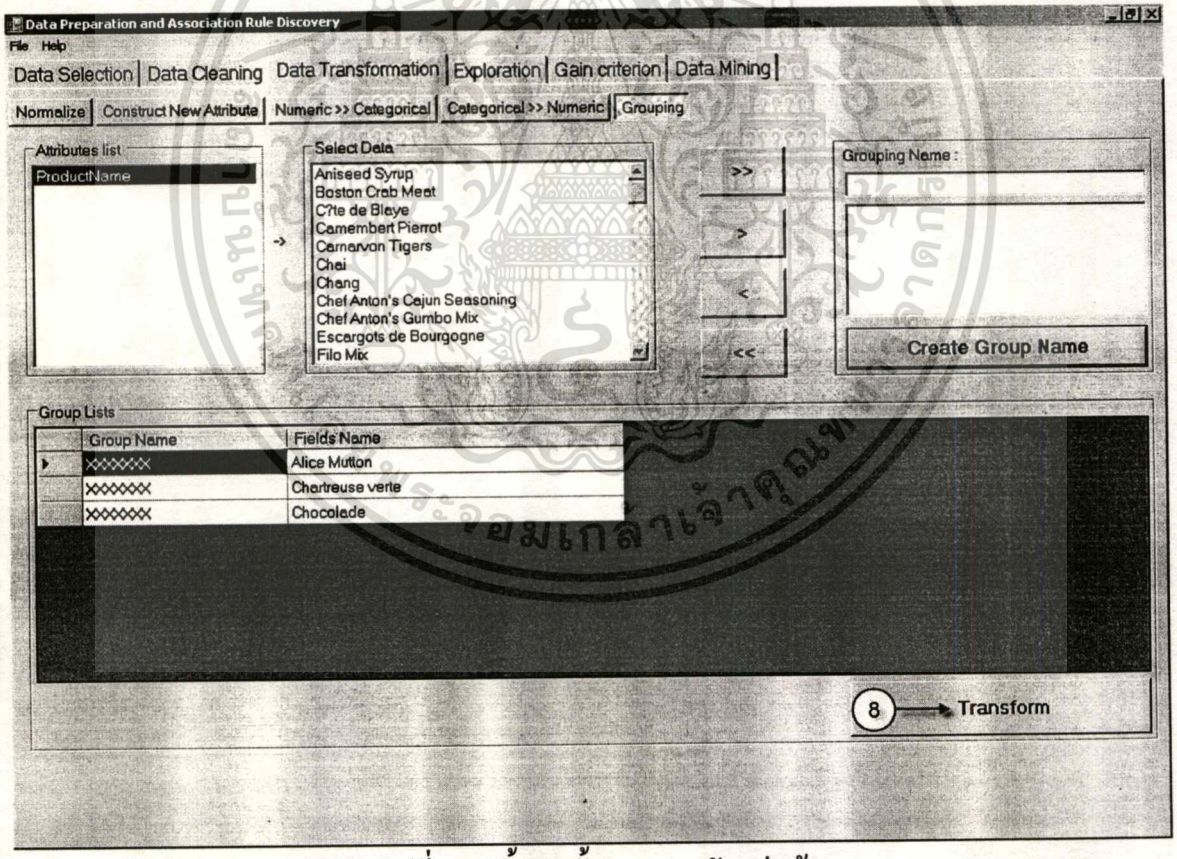
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. ระบบแสดงกล่องข้อความให้ผู้ใช้ทราบว่าได้จัดกลุ่มข้อมูลเรียบร้อยแล้ว ซึ่งในการจัดกลุ่มข้อมูลสามารถจัดกลุ่มข้อมูลได้มากกว่า 1 กลุ่ม โดยผู้ใช้สามารถกลับไปเลือกข้อมูลที่จะจัดกลุ่มได้



รูปที่5.12 กล่องข้อความแสดงผลการจัดกลุ่ม

8. เมื่อผู้ใช้งานทำการจัดกลุ่มเรียบร้อยแล้ว ทำการคลิกปุ่ม Transform เพื่อสิ้นสุดขั้นตอนการแปลงข้อมูล



รูปที่5.13 สิ้นสุดขั้นตอนการจัดกลุ่มข้อมูล

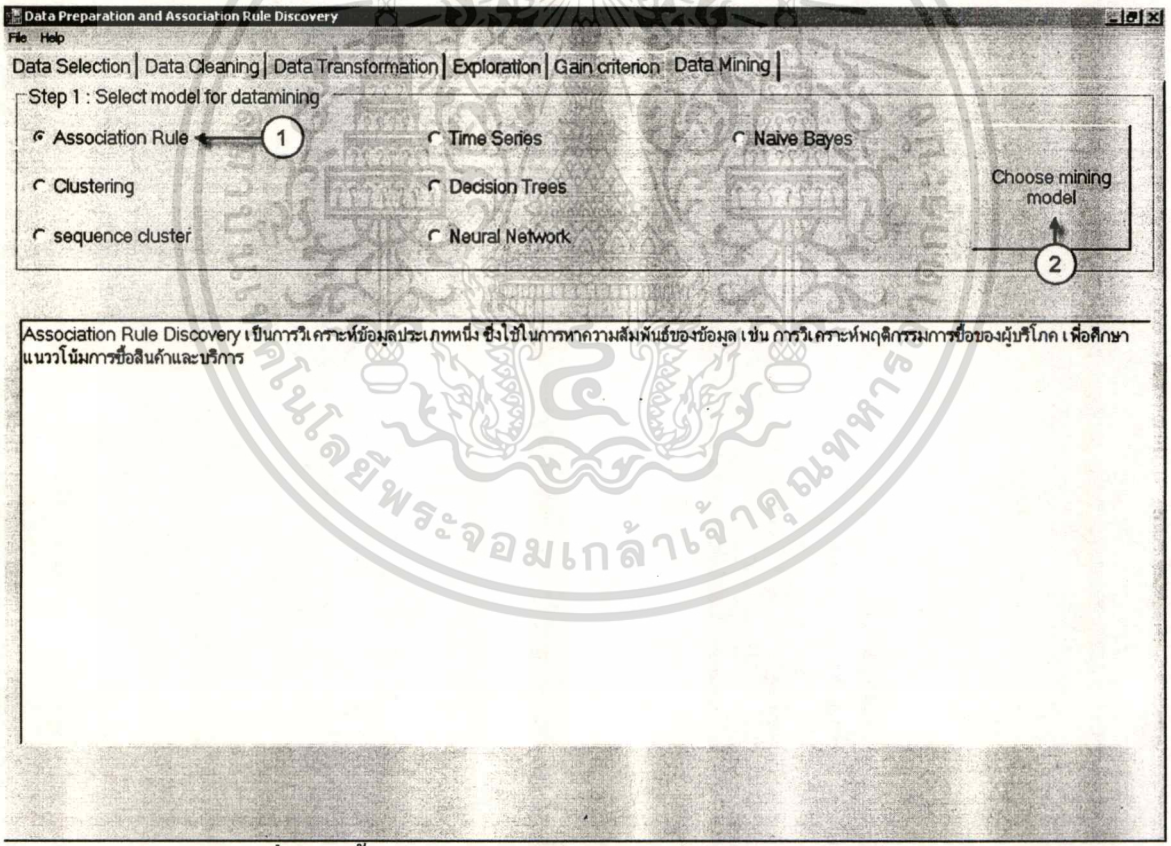
9. ระบบแสดงกล่องแสดงข้อความให้ผู้ใช้ทราบว่าทำการแปลงข้อมูลในรูปแบบการจัดกลุ่มเรียบร้อยแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่5.14 กล่องข้อความแสดงผลการแปลงข้อมูล

5.5.5 ขั้นตอนการสร้างกฎความสัมพันธ์

1. หลังจากจบขั้นตอนการแปลงข้อมูลแล้วผู้ใช้งานทำการคลิกเลือกที่แท็บ Data Mining เพื่อเลือกแบบจำลองในการทำเหมืองข้อมูล ซึ่งในที่นี้จะเลือกแบบ Association Rule ในการทำ
2. ผู้ใช้งานคลิกปุ่ม Choose mining model เพื่อยืนยันการเลือกแบบจำลอง



รูปที่5.15 ขั้นตอนการเลือกแบบจำลองที่ใช้ทำเหมืองข้อมูล

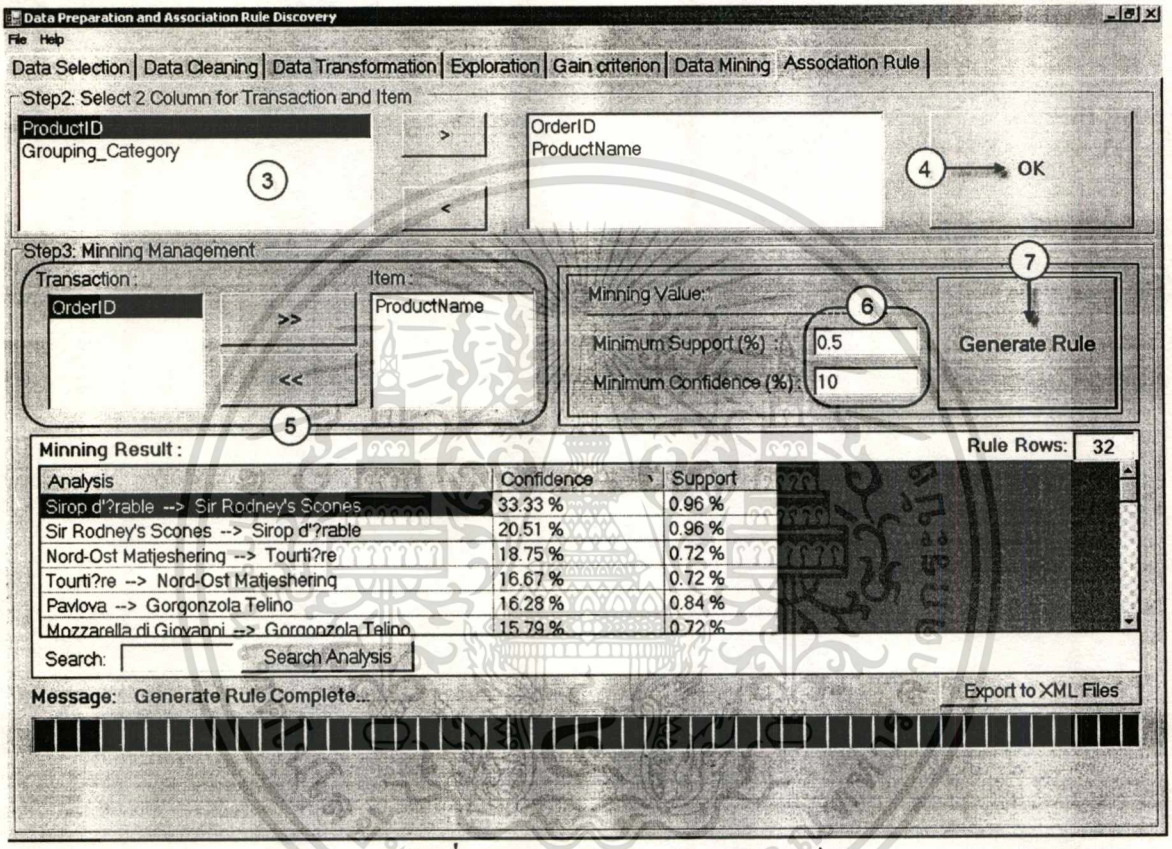
3. ผู้ใช้งานเลือกข้อมูลที่จะนำมาใช้ในหากฎความสัมพันธ์ ซึ่งในการหากฎความสัมพันธ์จะเลือกใช้ข้อมูลเพียง 2 ฟิลด์ เท่านั้น
4. ผู้ใช้งานคลิกปุ่ม OK เพื่อยืนยันการเลือกข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. ผู้ใช้งานกำหนดข้อมูล Transaction และข้อมูล Item โดยหากต้องการให้ข้อมูลใดเป็น Item ให้ทำการคลิกเลือกที่ข้อมูลนั้นแล้วคลิกที่ลูกศรไปทางขวา

6. ผู้ใช้งานกำหนดค่า Minimum Support และMinimum Confidence โดยทั้งค่า Minimum Support และ Minimum Confidence จะต้องกำหนดเป็นเปอร์เซ็น

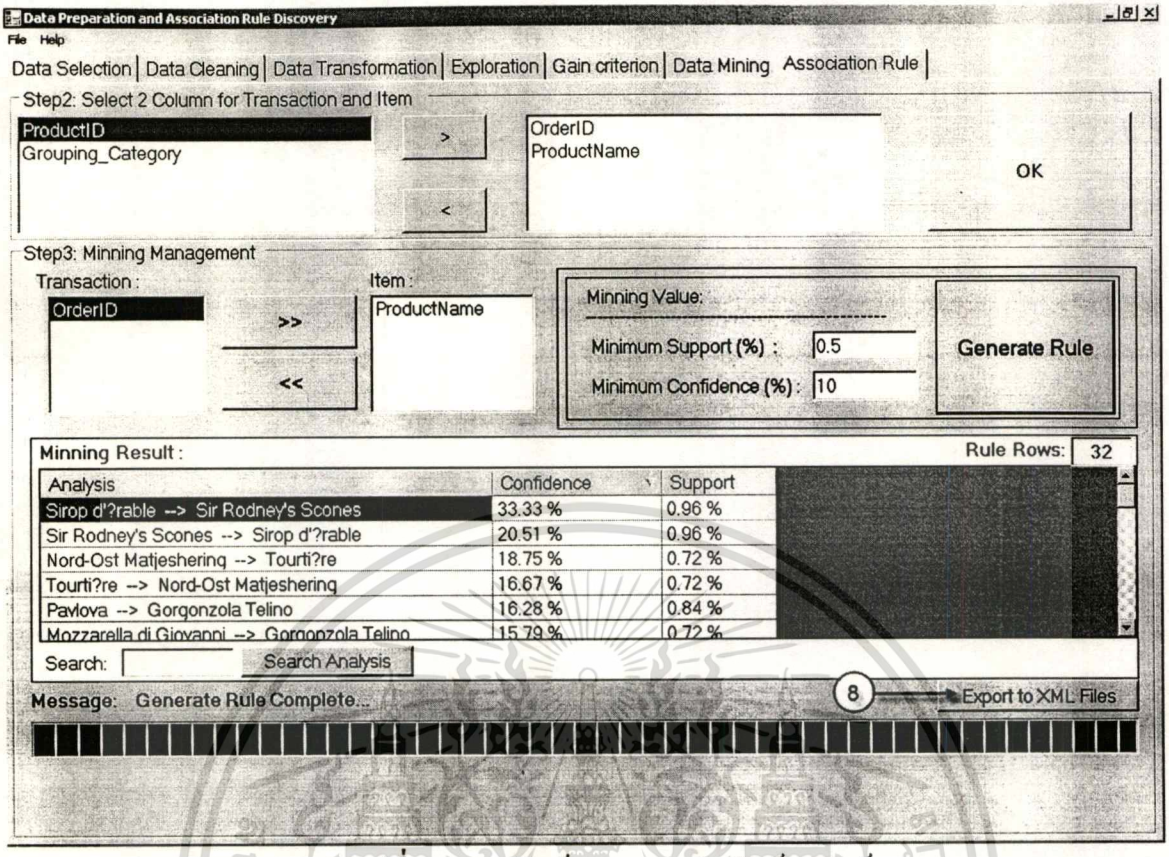
7. ผู้ใช้งานคลิกปุ่ม Generate Rule เพื่อทำการสร้างกฎความสัมพันธ์



รูปที่5.16 การสร้างกฎความสัมพันธ์

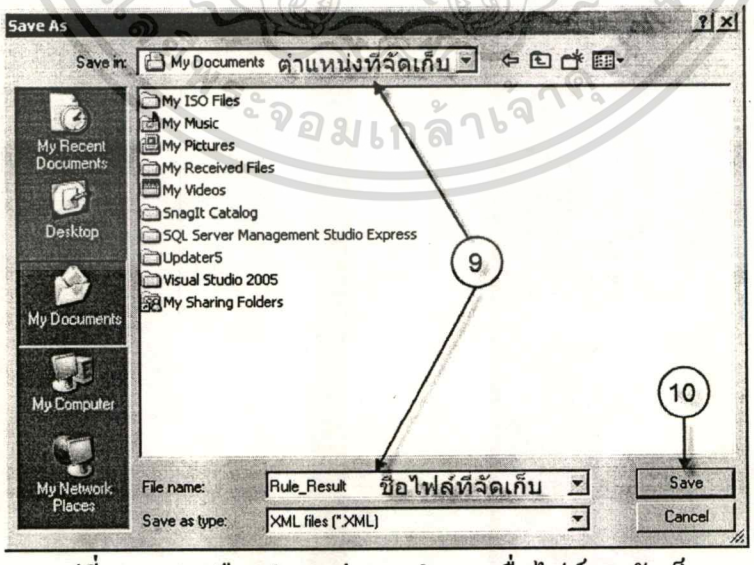
8. ผู้ใช้งานสามารถจัดเก็บกฎความสัมพันธ์ผลลัพธ์เป็นแบบ XML File ได้ โดยคลิกปุ่ม Export to XML Files

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่5.17 การจัดเก็บกฎความสัมพันธ์ผลลัพธ์

9. ผู้ใช้งานเลือกตำแหน่งที่จะจัดเก็บ และกำหนดชื่อไฟล์
10. ผู้ใช้งานคลิกปุ่ม Save เพื่อยืนยันการจัดเก็บไฟล์



รูปที่5.18 การเลือกตำแหน่งและกำหนดชื่อไฟล์การจัดเก็บ

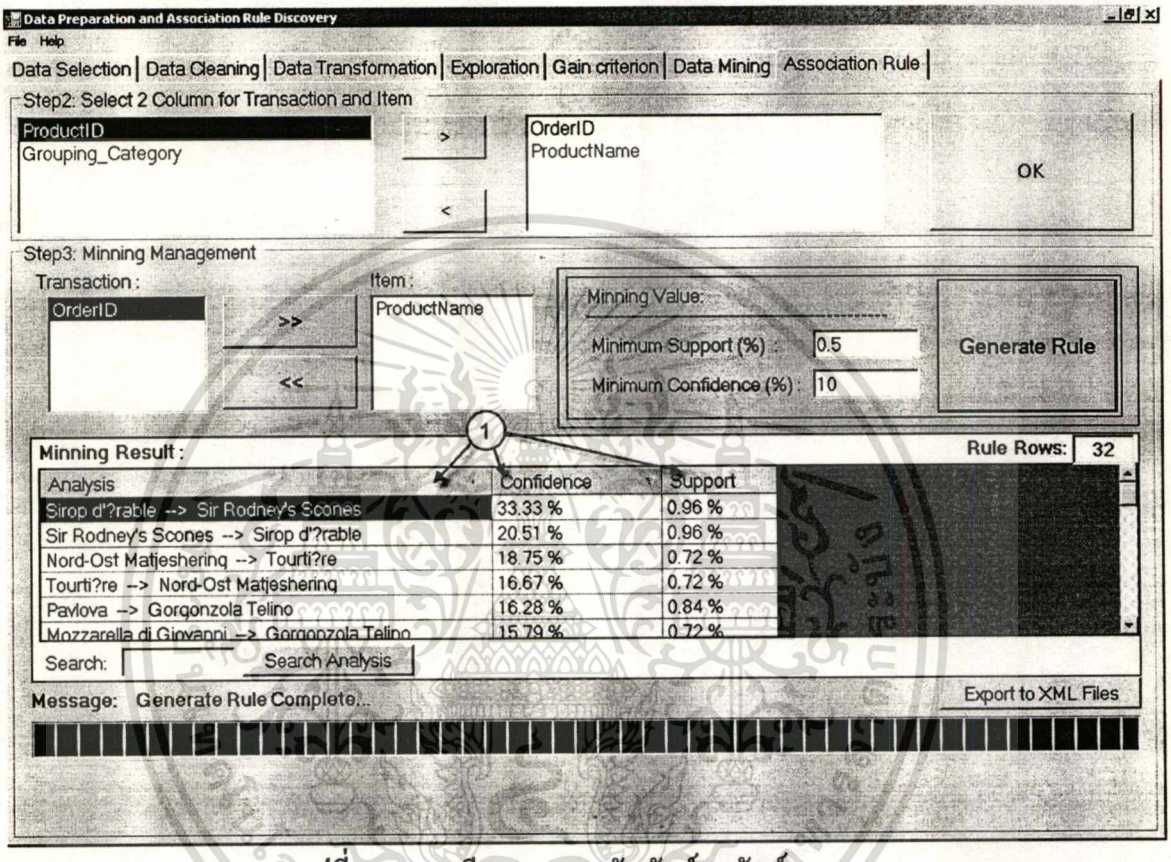
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.5.6 ขั้นตอนการแสดงกฎความสัมพันธ์ผลลัพธ์

ในการแสดงกฎความสัมพันธ์สามารถแบ่งออกเป็น 2 กรณี

กรณีที่1 การเรียงกฎความสัมพันธ์ผลลัพธ์ มีขั้นตอนดังนี้

1. ผู้ใช้งานทำการคลิกที่หัวคอลัมน์ที่ต้องการเรียง

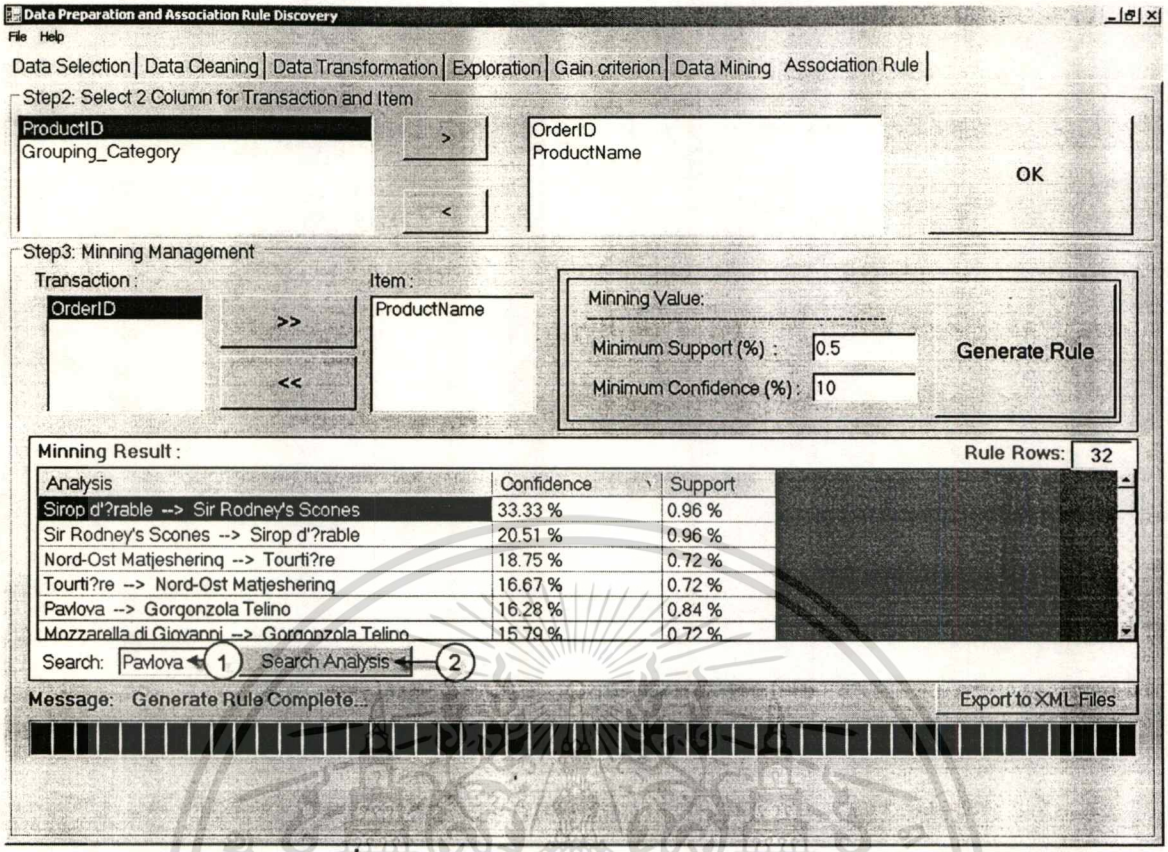


รูปที่5.19 การเรียงกฎความสัมพันธ์ผลลัพธ์

กรณีที่2 การค้นหากฎความสัมพันธ์ผลลัพธ์ มีขั้นตอนดังนี้

1. ผู้ใช้งานระบุกฎความสัมพันธ์ที่ต้องการค้นหา
2. ผู้ใช้งานคลิกปุ่ม Search Analysis เพื่อทำการค้นหากฎความสัมพันธ์ผลลัพธ์ที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่5.20 การค้นหากฎความสัมพันธ์ผลลัพธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่6
สรุป และข้อเสนอแนะ

ผลที่ได้จากการทำโครงงานพัฒนาระบบงานในหัวข้อเรื่อง "การพัฒนาระบบการค้นหากฎความสัมพันธ์แบบอพิออริอัลกอริทึม" สามารถสรุปผลได้ดังนี้

6.1 สรุปผลการพัฒนาระบบ

โครงงานนี้ออกแบบ และพัฒนาอยู่บนพื้นฐานการทำเหมืองข้อมูล ด้วยการใช้ความสัมพันธ์ของข้อมูลอดีต ซึ่งมีการนำอัลกอริทึมอพริออริมาใช้ในการประมวลผลหาความสัมพันธ์ดังกล่าว จากการพัฒนาระบบจะเน้นผลลัพธ์ที่ถูกต้องจากการหาความสัมพันธ์ โดยความสัมพันธ์ที่ได้เป็นการอาศัยข้อมูลในอดีต จึงสามารถนำมาวิเคราะห์ และประมวลผลได้

6.2 ปัญหาที่พบในการพัฒนา

จากการวิเคราะห์ ออกแบบ และพัฒนา พบว่ามีขั้นตอนการทำงานที่มีรายละเอียดเกี่ยวกับข้อมูลหลายอย่าง จึงทำให้ขั้นตอนการวิเคราะห์ข้อมูลเบื้องต้นจำเป็นต้องใช้เวลานานกว่างานอื่นๆ โดยมีรายละเอียดดังต่อนี้

6.2.1 ปัญหาที่พบ

เนื่องจากระบบที่ทำการพัฒนาได้นำระบบการเตรียมข้อมูลและสำรวจมารวมไว้ด้วยกัน จึงต้องอาศัยเวลาในการทำความเข้าใจ และสอบถามขั้นตอนการใช้งานของระบบดังกล่าว จึงจะสามารถสรุปและนำมาพัฒนาระบบต่อไปได้ ทำให้ต้องใช้เวลาค่อนข้างมากในการพัฒนาระบบ

6.2.2 ข้อเสียของการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นเพียงเครื่องมือที่ใช้สำหรับวิเคราะห์เท่านั้น โดยที่ไม่สามารถเข้าใจถึงข้อมูลได้ดีเท่ากับมนุษย์ ดังนั้นผู้ใช้เครื่องมือการทำเหมืองข้อมูล รวมไปถึงอัลกอริทึมที่เลือกจึงต้องมีความรู้ความเข้าใจเป็นอย่างดี อย่างไรก็ตามการทำเหมืองข้อมูลจะเป็นเพียงแค่การทำนายเท่านั้น ซึ่งไม่สามารถช่วยระบุค่าของข้อมูล หรือค่าที่แสดงความสัมพันธ์ของรูปแบบและความสัมพันธ์ของข้อมูลนั้นจริง จะต้องให้ผู้ใช้งานทำการตัดสินใจอีกครั้งหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.3 ข้อเสนอแนะ

จากการศึกษาและพัฒนาระบบจะเห็นว่าลักษณะการประมวลผลของของอพิออริอัลกอริทึมจะเหมาะสมกับจำนวนข้อมูลที่มีปริมาณไม่มาก เนื่องจากมีการนับจำนวนข้อมูลในทุกๆ รอบ ดังนั้นหากองค์กรใดที่มีขนาดใหญ่และต้องการหาความสัมพันธ์ของข้อมูลที่มีปริมาณมากๆ อาจจะต้องหาอัลกอริทึมที่มีประสิทธิภาพในการประมวลผลมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

อาทิตยา เชื้อจันอัด. 2549. "การพัฒนาระบบการเตรียมข้อมูลและการสำรวจ สำหรับการทำดาต้าไมนิ่ง"โครงการพัฒนาระบบงานวิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศบัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

Dr.Dobb's Journal. 1996. "Algorithm Alley." [Online]. Available : http://www.ddj.com/ftp/ 1996 Berry, M. J.A. and Linoff, G., Data Mining Techniques: for marketing, sales, and customer support, Wiley Computer Publishing, 1997.

Jagoda Crawford, Frank Crawford "Data Mining in a Scientific Environment." [Online]. Available: http://www.csu.edu.au/special/auugwww96/proceedings/crawford/ crawford.html. 1995.

Jiawei Han and Micheline Kamber. 2001. **Data Mining:Concept and Techniques**. San Francisco. Morgan Kaufman Publishers.

Megaputer. 2005. "What is datamining?". [Online]. Available: http://www.megaputer.com/ dm/dm101.php3.

Microsoft Corporation. 2005. "Microsoft SQL Server 2005." [Online]. Available: http://www.microsoft.com

Pablo Hadjinian, Peter Cabena, et.al. 1997. **Discovering Data Mining from concept to implementation**. Newjersy. Prentice Hall PTR.

Peter Bajcsy. 2005. "Knowledge Discovery in Databases."[Online]. Available: http://alg.ncsa.uiuc.edu/tools/docs/d2k/manual/dataMining.html.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ – นามสกุล	นางสาวนิอร หงส์วัฒนกุล
วัน เดือน ปีเกิด	7 ธันวาคม 2523 ที่ราชบุรี
ที่อยู่	33 หมู่ 4 ต.สวนกล้วย อ.บ้านโป่ง จ.ราชบุรี 70110
ประวัติการศึกษา	2546 วิทยาศาสตร์บัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยกรุงเทพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้