

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

การพัฒนาระบบเพื่อวิเคราะห์ความสัมพันธ์ของบริการเสริมในระบบโทรศัพท์เคลื่อนที่

Association of mobile value added service analysis system development

โดย

นางสาวเนตรนฤน อูปกาจิม

รหัส 42067163



H001867

อาจารย์ที่ปรึกษา

ดร. วรพจน์ กรีสระเดช

วัน เดือน ปี.....	19	ธ.ค.	2550
เลขทะเบียน.....	01867		
เลขเรียกหนังสือ.....	วทท 4๖78๕๓ 2544		
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."			

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
ภาคเรียนที่ 1 ปีการศึกษา 2544
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การพัฒนาระบบเพื่อวิเคราะห์ความสัมพันธ์ของบริการเสริมในระบบ โทรศัพท์เคลื่อนที่
นักศึกษา	นางสาว เนตรีนฤณ อุปกาศิม
อาจารย์ที่ปรึกษา	ดร. วรพจน์ กรีสระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
พ.ศ.	2544

บทคัดย่อ

โครงการนี้เป็นการพัฒนาโปรแกรมสำเร็จรูป เพื่อใช้ในการวิเคราะห์ความสัมพันธ์ของบริการเสริมที่มีในระบบโทรศัพท์เคลื่อนที่ โดยใช้ อีพริออริ อัลกอริทึม (Apriori Algorithm) ซึ่งเป็นอัลกอริทึมขั้นพื้นฐานในการหากฎความสัมพันธ์ (Association Rule) ของข้อมูล มาเป็นพื้นฐานในการพัฒนาโปรแกรม ผลการวิเคราะห์ที่ได้จากการใช้งานโปรแกรมจะอยู่ในรูปแบบของกฎความสัมพันธ์ ซึ่งจะมีค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confident) ในแต่ละกฎความสัมพันธ์ เพื่อใช้ในการตัดสินใจหรือคัดเลือกกฎความสัมพันธ์ที่สามารถนำไปใช้ประโยชน์ต่อไปได้ เช่น ในด้านการตลาด สามารถนำกฎความสัมพันธ์ที่ได้ มาจัดทำชุดของบริการเสริม (Package) เพื่อเพิ่มจุดขายให้กับระบบโทรศัพท์เคลื่อนที่

Title	Association of mobile value added service analysis system development
Student	Ms. Neadnarin Uppakasim
Advisor	Dr. Worapoj Kreesuradej
Level of Study	Master of Science in Information Technology
Major	Information Science
Academic Year	2001

ABSTRACT

The purpose of project is to develop an application program to help in analyzing the relationship of features of mobile phone system. Apriori Algorithm, which is a basic algorithm is finding an association rule of data, was used in this project.

The association rule is the result of the application. Support and Confident values of each rule can be used to determine or select a potential rule to gain a benefit to the mobile phone system. For example, in marketing, a package of feature can be launched to the market for gaining a sales promotion to the mobile phone system.

กิตติกรรมประกาศ

ในโครงการฉบับนี้ ที่ผู้จัดทำสามารถพัฒนาจนสำเร็จลุล่วงไปด้วยดีนั้น เนื่องด้วยได้รับคำแนะนำและความช่วยเหลือ รวมถึงกำลังใจที่ดี ทั้งนี้ผู้จัดทำขอกล่าวคำขอบคุณกับบุคคลและกลุ่มบุคคลต่างๆ ดังนี้

1. บิดามารดา ที่เป็นผู้ให้ทุกอย่าง รวมถึงกำลังใจที่ดีเสมอมา
2. ดร.วรพจน์ กริสุระเดช อาจารย์ที่ปรึกษาวิชาโครงการพัฒนาระบบงาน ขอขอบคุณสำหรับคำแนะนำ ข้อเสนอแนะ และการสนับสนุนการทำโครงการฉบับนี้ตั้งแต่เริ่มต้นจนสำเร็จ
3. พี่ๆ เพื่อนๆ น้องๆ ทุกคน ขอขอบคุณในการช่วยเหลือทุกสิ่ง ไม่ว่าจะเป็นด้านฮาร์ดแวร์ ซอฟต์แวร์ ด้านข้อมูล หนังสืออ้างอิง คำแนะนำ ความช่วยเหลือ ในทุกๆ ด้าน และกำลังใจ รวมถึงการดูแลต่างๆ ที่มีให้ตลอดเวลา

ด้วยความขอบคุณเป็นอย่างสูง

เนตรนฤน อุปกาศิม

ผู้จัดทำ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญภาพ	VII
บทที่	
1. บทนำ	1
1.1 ความเป็นมา	1
1.2 วัตถุประสงค์ของการศึกษา	1
1.3 ขอบเขตการศึกษา	2
1.4 ขั้นตอนของการดำเนินการศึกษา	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ	2
2. แนวคิดและทฤษฎีที่เกี่ยวข้อง	3
2.1 ความหมายของดาต้าไมนิ่ง (Data Mining)	3
2.2 ขั้นตอนการวิเคราะห์ข้อมูล	4
2.3 ประเภทของกระบวนการทำดาต้าไมนิ่ง (Data Mining)	8
2.4 กระบวนการของลิงค์อานาไลซิส (Link Analysis)	8
2.5 อีพออริ อัลกอริทึม (Apiori Algorithm)	13
3. วิธีการดำเนินการศึกษาเพื่อทำการวิเคราะห์ความสัมพันธ์ของบริการเสริม ในระบบโทรศัพท์เคลื่อนที่	21
3.1 กำหนดวัตถุประสงค์	21
3.2 คัดเลือกข้อมูล	21
3.3 การแปลงและทำความสะอาดข้อมูล	26
3.4 ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ระบบที่พัฒนาขึ้น	26

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้า
บทที่	
4. สรุปผลการศึกษาและข้อเสนอแนะ	33
4.1 สรุปผลการศึกษา	33
4.2 ข้อเสนอแนะ	33
เอกสารอ้างอิง	35
ภาคผนวก ก.	36
ประวัติผู้เขียน	43



สารบัญตาราง

ตารางที่	หน้า
2.1 หมายเลขโทรศัพท์และบริการเสริมที่เปิดให้บริการ	16
2.2 เขตของ C ₁	17
2.3 เขตของ L ₁	17
2.4 เขตของ C ₂	18
2.5 เขตของ L ₂	18
2.6 เขตของ C ₃	18
2.7 เขตของ L ₃	19
2.8 กฎความสัมพันธ์และค่า Confident ของแต่ละกฎ	19
2.9 กฎความสัมพันธ์ที่มีค่า Confident มากกว่าค่า Minimum Confident	20
3.1 ตารางข้อมูลลูกค้า	22
3.2 ตารางบริการเสริมในระบบ	23
3.3 ตารางการขอเปิดให้บริการเสริม	24
3.4 ตารางประวัติการขอใช้บริการเสริมของลูกค้า	24
3.5 ตารางของบริการเสริมที่สนใจ	25
3.6 ตารางการให้บริการเสริมของแต่ละหมายเลขโทรศัพท์	25
3.7 กฎความสัมพันธ์ที่มีค่า Support และค่า Confident สูง 5 อันดับแรก	31

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

ภาพที่	หน้า	
2.1	แผนภาพแสดงความสัมพันธ์ของดาต้าไมนิ่งกับความรู้แขนงต่างๆ	3
2.2	ขั้นตอนการวิเคราะห์ข้อมูลของ Data Mining	4
2.3	ประเภทของการแปลงข้อมูล	5
2.4	ประเภทของการแสดงผลการวิเคราะห์	6
2.5	แผนภาพแสดงเวลาที่ใช้ในแต่ละขั้นตอนของการวิเคราะห์ข้อมูล	7
2.6	การจัดหมวดสินค้า	10
2.7	การสร้างสินค้าจำลอง	11
2.8	การแก้ปัญหาของสินค้าจำลอง	12
2.9	Main Pseudo Code	14
2.10	Generate Pseudo Code	15
2.11	Generating Rule Pseudo Code	15
3.1	หน้าจอเพื่อติดต่อฐานข้อมูล	27
3.2	หน้าจอการคัดเลือกข้อมูล	27
3.3	หน้าจอกำหนดค่า	28
3.4	ตารางแสดงผลการวิเคราะห์	29
3.5	ผลการวิเคราะห์ข้อมูล	30
ก.1	หน้าจอโปรแกรม Miner Project	36
ก.2	หน้าจอแสดงการติดต่อฐานข้อมูลที่เป็น Microsoft Access	37
ก.3	หน้าจอแสดงการติดต่อฐานข้อมูลผ่าน ODBC	37
ก.4	หน้าจอแจ้งการติดต่อฐานข้อมูลได้สำเร็จ	38
ก.5	หน้าจอคัดเลือกข้อมูล	38
ก.6	เลือกข้อมูลที่ต้องการวิเคราะห์	39
ก.7	หน้าจอแสดงการสร้างกลุ่มข้อมูล	40
ก.8	หน้าจอกำหนดค่าต่างๆ	41
ก.9	หน้าจอแสดงผลการวิเคราะห์	42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมา

ระบบการสื่อสารในปัจจุบัน มีการแข่งขันทางการตลาดกันอย่างมาก โดยเฉพาะอย่างยิ่ง การสื่อสารในรูปแบบของโทรศัพท์เคลื่อนที่ ดังนั้นแต่ละบริษัทที่ให้บริการระบบโทรศัพท์เคลื่อนที่จึงจำเป็นต้องคิดหาวิธีหรือรูปแบบการให้บริการแก่ลูกค้า ที่จะทำให้ลูกค้าที่สนใจจะใช้โทรศัพท์เคลื่อนที่ ตัดสินใจที่จะใช้บริการกับระบบของบริษัท หรือทำให้ลูกค้าที่อยู่ในระบบ มีความซื่อสัตย์ต่อระบบยิ่งขึ้น ฉะนั้นวิธีหรือรูปแบบการให้บริการที่จะมีระบบจะต้องสอดคล้องกับความต้องการของลูกค้าในตลาดโทรศัพท์เคลื่อนที่มากที่สุด จากข้อมูลลูกค้าในระบบที่มีอยู่ในปัจจุบัน ถ้าบริษัทสามารถหารูปแบบการใช้งาน รวมถึงความสนใจของลูกค้าที่แฝงอยู่ในข้อมูลลูกค้าได้ จะทำให้บริษัทสามารถพัฒนาหรือปรับปรุงการให้บริการให้ตรงตามความต้องการของลูกค้ามากที่สุด ซึ่งจะช่วยให้บริษัทมีลูกค้าในระบบเพิ่มขึ้น

จากแนวคิดดังกล่าว ทำให้ต้องมีการวิเคราะห์ข้อมูลลูกค้าที่มีในระบบ เพื่อศึกษาและวิเคราะห์ถึงพฤติกรรมการใช้บริการเสริมของลูกค้า เพื่อหาความสัมพันธ์ของบริการเสริมแต่ละตัวว่า บริการเสริมใดบ้างที่มีการขอใช้บริการควบคู่กันได้ ดังนั้นการทำดาต้าไมนิ่ง (Data Mining) จึงถูกนำมาใช้ประโยชน์เพื่อค้นหารูปแบบการขอเปิดบริการของลูกค้า เพื่อจะนำไปวิเคราะห์และสร้างเป็นชุดของบริการเสริมต่อไป

1.2 วัตถุประสงค์ของการศึกษา

จากที่กล่าวมาแล้วข้างต้นว่า โทรศัพท์เคลื่อนที่ได้เข้ามามีบทบาทในชีวิตประจำวันของบุคคลทั่วไป บริษัทผู้ให้บริการมีการแข่งขันกันมาก และจากความต้องการที่จะจัดชุดของบริการเสริมให้ตรงกับความต้องการของลูกค้า ดังนั้นในการศึกษาร้านี้จึงเน้นในเรื่องของการศึกษาขั้นตอนการทำดาต้าไมนิ่ง (Data Mining) , ลิงค์อานาไลซิส (Link Analysis) ซึ่งเป็นกระบวนการหนึ่งของดาต้าไมนิ่งที่เหมาะสมในการหาความสัมพันธ์ และรูปแบบของการขอเปิดใช้บริการเสริมของลูกค้าในระบบ เพื่อที่จะสามารถประยุกต์วิธีการดังกล่าว เป็นโปรแกรมสำเร็จรูป เพื่อช่วยในการวิเคราะห์ความสัมพันธ์ของข้อมูลการเปิดบริการเสริมของลูกค้า

1.3 ขอบเขตการศึกษา

การศึกษาโครงการนี้กำหนดขอบเขตในการศึกษา เป็นการพัฒนาโปรแกรมสำเร็จรูป โดยใช้ไอพอรื อัลกอริทึม (Apriori Algorithm) เป็นพื้นฐานในการพัฒนา เพื่อให้โปรแกรมสามารถหาความสัมพันธ์ของบริการเสริมของระบบโทรศัพท์เคลื่อนที่ได้ โดยที่ข้อมูลที่ใช้ในการวิเคราะห์เป็นตัวอย่างข้อมูลที่สมมุติขึ้นจำนวนหนึ่งของผู้ใช้บริการโทรศัพท์เคลื่อนที่ในองค์กร เพื่อใช้ในการวิเคราะห์โดยโปรแกรมจะทำการหาค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confident) และคัดเลือกเฉพาะกฎความสัมพันธ์ที่มีค่าทั้ง 2 ค่า มากกว่าค่าที่กำหนดไว้

1.4 ขั้นตอนการดำเนินการศึกษา

1. กำหนดวัตถุประสงค์ของการศึกษา
2. ศึกษาพื้นฐานของการทำค้ำด้าไมนิง (Data Mining) ด้วยกระบวนการของ ลิงค้อนาไลซิด (Link Analysis) โดยใช้เทคนิคการหากฎความสัมพันธ์ (Association Rule) และอัลกอริทึมที่ใช้ในการในการพัฒนาโปรแกรมคือ อไอพอรื อัลกอริทึม (Apriori Algorithm)
3. คัดเลือกข้อมูลที่จะใช้ทำการวิเคราะห์
4. ออกแบบและพัฒนาโปรแกรมเพื่อทำการวิเคราะห์ข้อมูล
5. วิเคราะห์ผลที่ได้จากการวิเคราะห์ เพื่อนำว่าสัมพันธ์ของบริการเสริมมาสร้างเป็นชุดของบริการเสริม
6. สรุปผลการศึกษา

1.5 ประโยชน์ที่คาดว่าจะได้รับ

จากการศึกษาทฤษฎีและเทคนิคของการทำค้ำด้าไมนิง (Data Mining) และได้ทำการคัดเลือกที่จะใช้เทคนิคการหาความสัมพันธ์ (Association Rule) เป็นเทคนิคที่เหมาะสมในการพัฒนาระบบเพื่อทำการวิเคราะห์ข้อมูลการขอใช้บริการเสริมในระบบโทรศัพท์เคลื่อนที่ เพื่อหาความสัมพันธ์ของข้อมูล คาดว่าเมื่อนำระบบที่ได้พัฒนาขึ้น มาทำการวิเคราะห์ข้อมูล จะได้ผลลัพธ์ที่เป็นความสัมพันธ์ของตัวบริการเสริมแต่ละตัว ซึ่งจะสามารถนำไปสร้างเป็นชุดของบริการเสริมเพื่อนำเสนอต่อลูกค้าในอนาคตได้ รวมถึงเป็นพื้นฐานในการวิเคราะห์ข้อมูลอื่นๆ ในระบบโทรศัพท์เคลื่อนที่ต่อไป

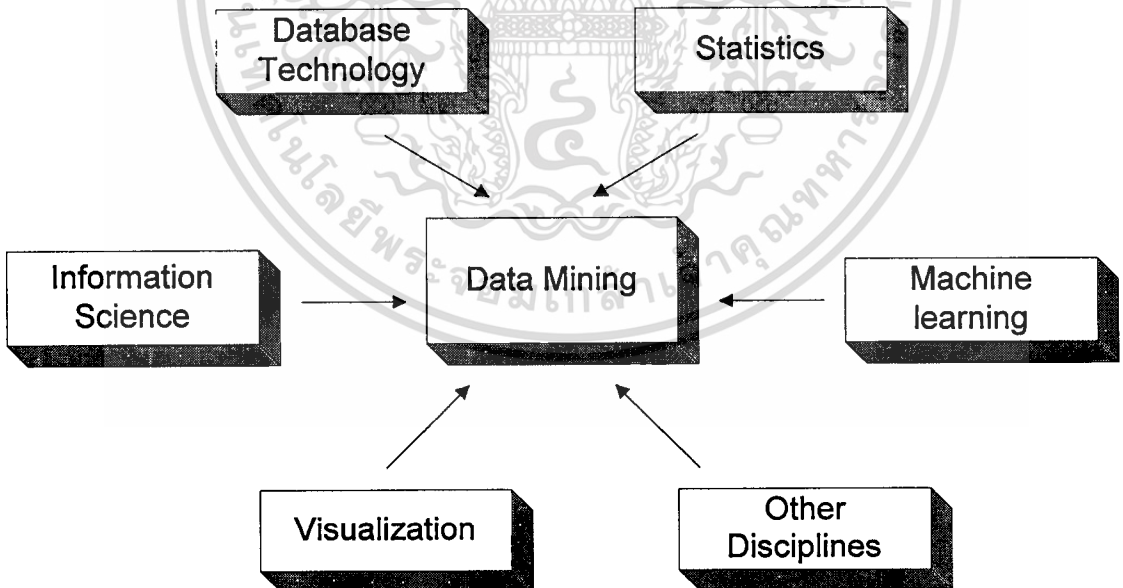
บทที่ 2

แนวคิดและทฤษฎีที่เกี่ยวข้อง

2.1 ความหมายของดาต้าไมนิ่ง (Data Mining)

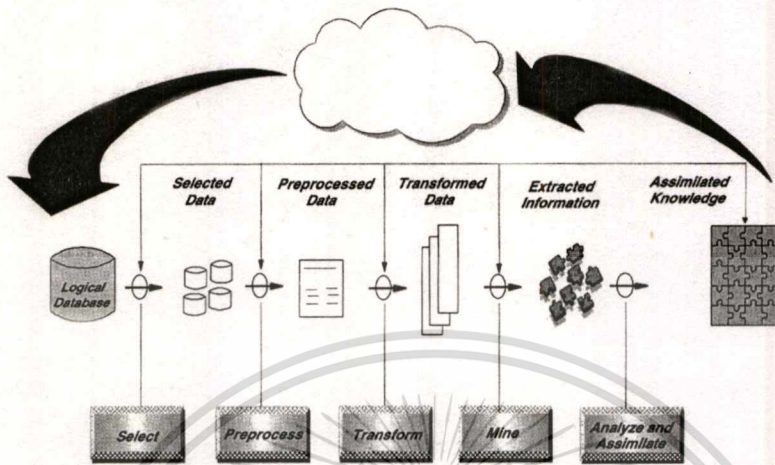
เป็นกระบวนการที่มีจุดมุ่งหมายเพื่อวิเคราะห์ข้อมูลที่มีอยู่ให้เกิดประโยชน์ ไม่ว่าจะเป็นการหาความสัมพันธ์ของข้อมูลหรือ การวิเคราะห์หาแนวโน้มความเป็นไปได้ของข้อมูล เพื่อนำมาใช้ประโยชน์ทางการตลาด รวมไปถึงการวิเคราะห์ถึง สิ่งผิดปกติที่เกิดขึ้นในฐานข้อมูล เพื่อหาจุดบกพร่องในการทำงาน

ดาต้าไมนิ่ง เป็นกระบวนการที่รวมเอาความสัมพันธ์ของการทำงานหรือความรู้หลายแขนงมาประยุกต์เข้าด้วยกัน ไม่ว่าจะเส้นทางด้านระบบฐานข้อมูล สถิติ ระบบสารสนเทศ ขึ้นอยู่กับข้อมูลที่ต้องการวิเคราะห์ ว่าจะมีความสัมพันธ์กับความรู้ด้านใดบ้าง ดังนั้นผู้ที่ทำดาต้าไมนิ่ง ควรมีความรู้พื้นฐานในแขนงต่างที่เกี่ยวข้องกับข้อมูลที่ต้องการทำการวิเคราะห์เพื่อที่จะได้ประโยชน์จากการทำดาต้าไมนิ่ง และจะสามารถทำดาต้าไมนิ่งได้ถูกต้องที่สุด



รูปที่ 2.1 แผนภาพแสดงความสัมพันธ์ของดาต้าไมนิ่งกับความรู้แขนงต่างๆ [3]

2.2 ขั้นตอนการวิเคราะห์ข้อมูล



รูปที่ 2.2 ขั้นตอนการวิเคราะห์ข้อมูลของ Data Mining [2]

2.2.1 การวางวัตถุประสงค์ของการวิเคราะห์ข้อมูล

เป็นส่วนที่สำคัญที่สุดของการวิเคราะห์ เนื่องจากถ้าเราวางวัตถุประสงค์ได้ชัดเจนแล้ว ในขั้นตอนต่อๆ มาจะสามารถทำงานได้อย่างตรงประเด็น ทำให้ไม่เสียเวลาในการที่จะต้องเริ่มต้นใหม่ เนื่องจากการวางวัตถุประสงค์ที่ไม่แน่ชัด ผลที่วิเคราะห์ได้ก็อาจจะคลุมเครือ ไม่สามารถนำมาประยุกต์ใช้ต่อไปได้

2.2.2 การคัดเลือกข้อมูล

ในขั้นตอนนี้เป็นขั้นตอนที่ค่อนข้างใช้เวลามากที่สุดในการวิเคราะห์ข้อมูล เป็นการคัดข้อมูลที่อยู่ในฐานข้อมูลที่สัมพันธ์กับวัตถุประสงค์ในการวิเคราะห์ ซึ่งข้อมูลนี้จะต้องเป็นข้อมูลที่มีความสมบูรณ์ และถูกต้องที่สุด ซึ่งเราสามารถแบ่งประเภทของข้อมูลได้เป็น

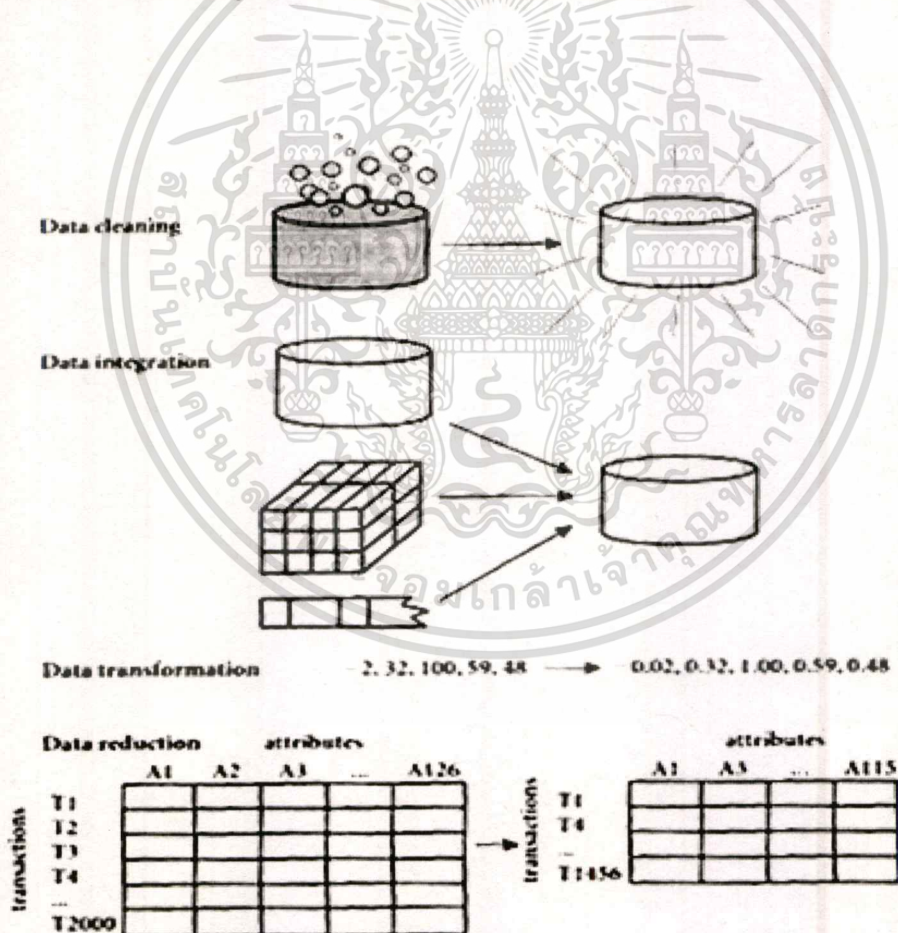
- ข้อมูลตัวเลข (Quantitative) คือ ตัวเลขจำนวนเต็ม และ จำนวนจริง เช่น รหัสพนักงาน, อายุ, ค่าจ้าง เป็นต้น
- ข้อมูลที่ไม่ใช่ตัวเลข (Categorical Data) สามารถแบ่งได้เป็น
 - ข้อมูลที่ไม่มีลำดับความสำคัญ เช่น ชื่อ-นามสกุล, เพศ
 - ข้อมูลที่มีลำดับความสำคัญ เช่น เกรด (A, B, C, D, F)

2.2.3 การแปลงข้อมูล

ในข้อมูลบางชนิดที่จะต้องนำมาประมวลผลหรือเป็นข้อมูลที่อยู่ในรูปแบบที่ไม่เหมาะกับการนำมาวิเคราะห์ สามารถทำการแปลงรูปแบบของข้อมูลให้อยู่ในรูปแบบที่เหมาะสมได้ เช่น ใช้การกำหนดตัวเลข แทนสถานภาพสมรถ โดยกำหนดให้

- 1 = สถานภาพโสด
- 2 = สถานภาพแต่งงาน
- 3 = สถานภาพหย่าร้าง

หรือข้อมูลที่มีค่าที่ผิดไปจากความเป็นจริง เช่น วันที่ ต้องตรวจสอบว่ามีข้อมูลใดที่มีค่าไม่ตรงกับความ เป็นจริงหรือไม่ รวมถึงจัดการข้อมูลบางส่วนที่ขาดหายไป ซึ่งอาจจะกำหนดค่าเฉพาะที่ทำให้ทราบว่าเป็นค่าที่ใช้แทนข้อมูลว่าง หรือจะใช้ค่าเฉลี่ยของข้อมูลทั้งหมดมากำหนดแทนก็ได้



รูปที่ 2.3 ประเภทของการแปลงข้อมูล [3]

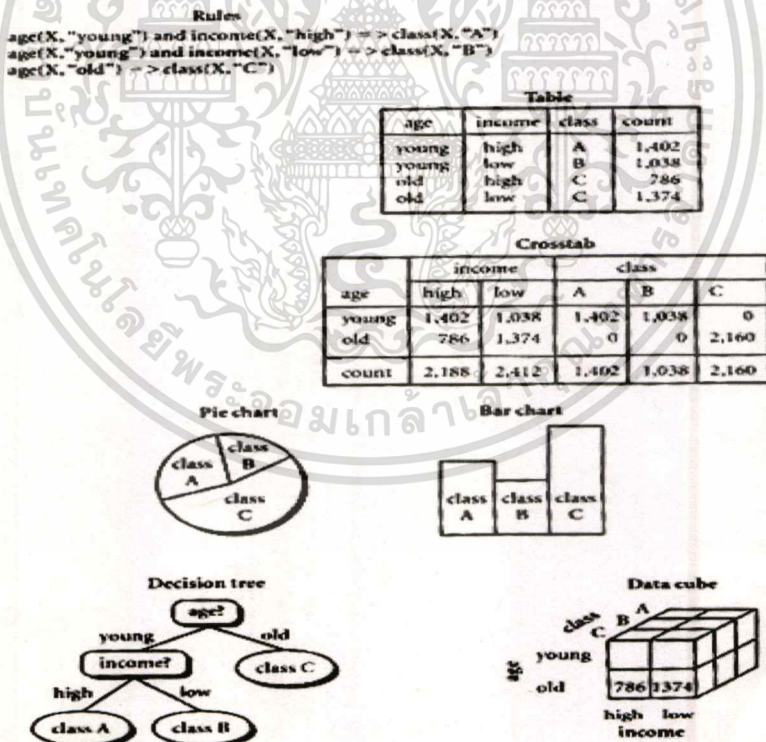
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.4 การทำดาต้าไมนิ่ง (Data Mining)

เป็นการนำเอาเทคนิคและกระบวนการมาที่เหมาะสมของดาต้าไมนิ่ง (Data Mining) ช่วยในการวิเคราะห์ข้อมูล ซึ่งการพิจารณาว่าเทคนิคใดจะเหมาะสม ต้องพิจารณาถึงวัตถุประสงค์ที่ตั้งไว้ ข้อมูลที่มีอยู่ และผลที่ต้องการจะได้รับจากการวิเคราะห์

2.2.5 การวิเคราะห์ผล

ผลที่ได้จากการทำดาต้าไมนิ่ง (Data Mining) เป็นเพียงข้อมูลที่จะช่วยให้มองเห็นรูปแบบของข้อมูลที่มีอยู่ว่าเป็นอย่างไร ดังนั้นต้องวิเคราะห์ว่าผลที่ได้ตรงกับวัตถุประสงค์ที่วางไว้หรือไม่ และได้ประโยชน์มากน้อยเพียงใดและเหมาะสมที่จะนำไปเป็นแนวทางในการเพิ่มผลกำไรในธุรกิจนั้นๆ หรือไม่ ถ้าผลที่ได้ไม่เป็นที่น่าพอใจ ก็สามารถที่จะกลับไปเริ่มต้นใหม่ โดยอาจจะพิจารณาหาว่ามีจุดบกพร่องที่ขั้นตอนใด แล้วจึงย้อนไปแก้ไขที่จุดนั้นได้

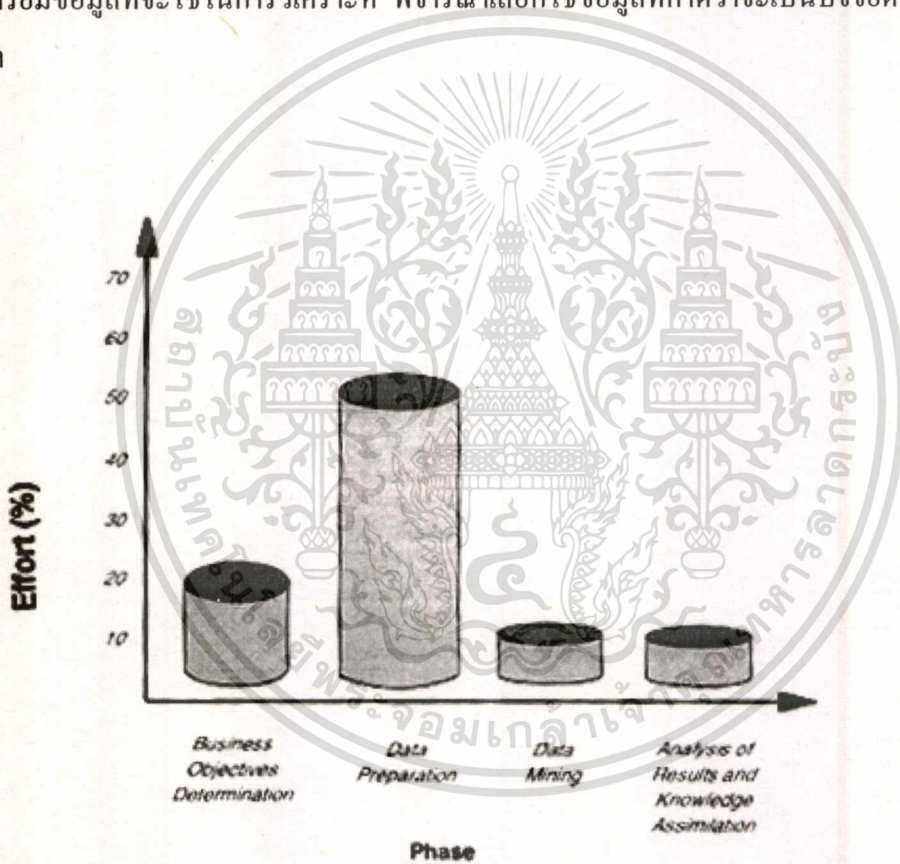


รูปที่ 2.4 ประเภทของการแสดงผลการวิเคราะห์ [3]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการแสดงผลการวิเคราะห์นั้น สามารถทำได้ในหลายลักษณะ ดังแสดงในรูปที่ 2.4 แล้วแต่การนำเสนอของแต่ละการ โปรแกรมหรือกระบวนการเพื่อให้ผลการวิเคราะห์นั้นเข้าใจง่ายที่สุด เช่น แสดงผลในรูปของกฎ, ตาราง, กราฟ เป็นต้น

ทุกขั้นตอนในการวิเคราะห์ข้อมูลที่กล่าวมา จะต้องใช้เวลาในการจัดการเตรียมข้อมูลต่างๆ โดยเราสามารถเปรียบเทียบเวลาที่ใช้ในขั้นตอนต่างๆ ได้ดังรูปที่ 2.5 จะเห็นว่าเวลาส่วนใหญ่ของการวิเคราะห์จะอยู่ในส่วนของการเตรียมข้อมูล ดังนั้น ในการวิเคราะห์ใดๆ ควรให้ความสนใจต่อการเตรียมข้อมูลที่จะใช้ในการวิเคราะห์ พิจารณาเลือกใช้ข้อมูลที่คาดว่าจะป้จจัยต่อรูปแบบของข้อมูล



รูปที่ 2.5 แผนภาพแสดงเวลาที่ใช้ในแต่ละขั้นตอนของการวิเคราะห์ข้อมูล [2]

2.3 ประเภทของกระบวนการทำดาต้าไมนิ่ง (Data Mining)

2.3.1 Classification Model หรือ Prediction model

เป็นการทำนายว่าสิ่งที่เราสนใจจะอยู่ในประเภทไหนหรือกลุ่มไหน โดยประเภทหรือกลุ่มดังกล่าวมีขึ้นล่วงหน้าแล้ว หรือทำนายแนวโน้มของข้อมูลที่จะเกิดขึ้น

2.3.2 Clustering Model

Clustering เป็นวิธีการในการจัดกลุ่มข้อมูลที่มีลักษณะคล้ายกันเข้าไว้ด้วยกัน หรือเรียกอีกอย่างหนึ่งได้ว่าการ Database segmentation โดยที่กลุ่มข้อมูลที่จัดขึ้น ไม่ได้มีการกำหนดไว้ก่อนล่วงหน้า

2.3.3 Link Analysis

เป็นการวิเคราะห์หาความสัมพันธ์ของข้อมูล ว่าสิ่งใดควรไปด้วยกันได้ มักถูกใช้ในด้านของการตลาดเพื่อที่จะวิเคราะห์ว่ามีสินค้าชนิดไหนที่ถูกซื้อควบคู่ไปด้วยกัน ดังนั้นบางครั้งจะเรียกการวิเคราะห์ในลักษณะนี้ว่า "Market Basket Analysis" (MBA)

2.3.4 Deviation Detection

เป็นการวิเคราะห์ข้อมูล เพื่อหาจุดที่แตกต่างที่เกิดขึ้นในตัวข้อมูล โดยอาศัยการแสดงผลในรูปแบบของกราฟ เพื่อที่จะมองเห็นสิ่งผิดปกติที่เกิดขึ้นในข้อมูลที่กำลังทำการวิเคราะห์

ในโครงงานนี้จะนำเสนอในเรื่องของลิงค์อณาไลซิส เพื่อจะศึกษาถึงแนวคิดและขั้นตอนการทำงาน เพื่อที่จะนำไปประยุกต์ใช้ให้ตรงตามวัตถุประสงค์ของการทำโครงงานนี้ ซึ่งจะกล่าวถึงในบทต่อไป

2.4 กระบวนการของลิงค์อณาไลซิส (Link Analysis)

ลิงค์อณาไลซิส (Link Analysis) หรือ Market basket analysis เป็นกระบวนการที่ใช้ในการวิเคราะห์ถึงความสัมพันธ์ของข้อมูลที่มีอยู่ ว่ามีสิ่งไหนสามารถไปกันได้กับสิ่งอีกสิ่งหนึ่ง เช่นในการขายสินค้า สามารถนำหลักการของ ลิงค์อณาไลซิส (Link Analysis) วิเคราะห์หาว่าสินค้าชนิดใดที่ถูกค้ำมักจะซื้อ ไปคู่กันหรือในทางการสื่อสาร โทรคมนาคมจะพิจารณาการใช้โทรศัพท์ของลูกค้ารายหนึ่งๆ ผลที่ได้จากการวิเคราะห์ความสัมพันธ์ จะนำมาประยุกต์เพื่อเพิ่มยอดขายให้มากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นต้น เทคนิคของลิงก์อานาไลซิส (Link Analysis) ที่นิยมนำมาใช้คือเทคนิคของการหาความสัมพันธ์ (Association Rule) เทคนิคนี้เป็นเทคนิคที่ใช้ในการหาความสัมพันธ์ของข้อมูล โดยที่ผลที่ได้จะออกมาในรูปแบบของกฎความสัมพันธ์ ซึ่งสามารถเข้าใจได้ง่าย

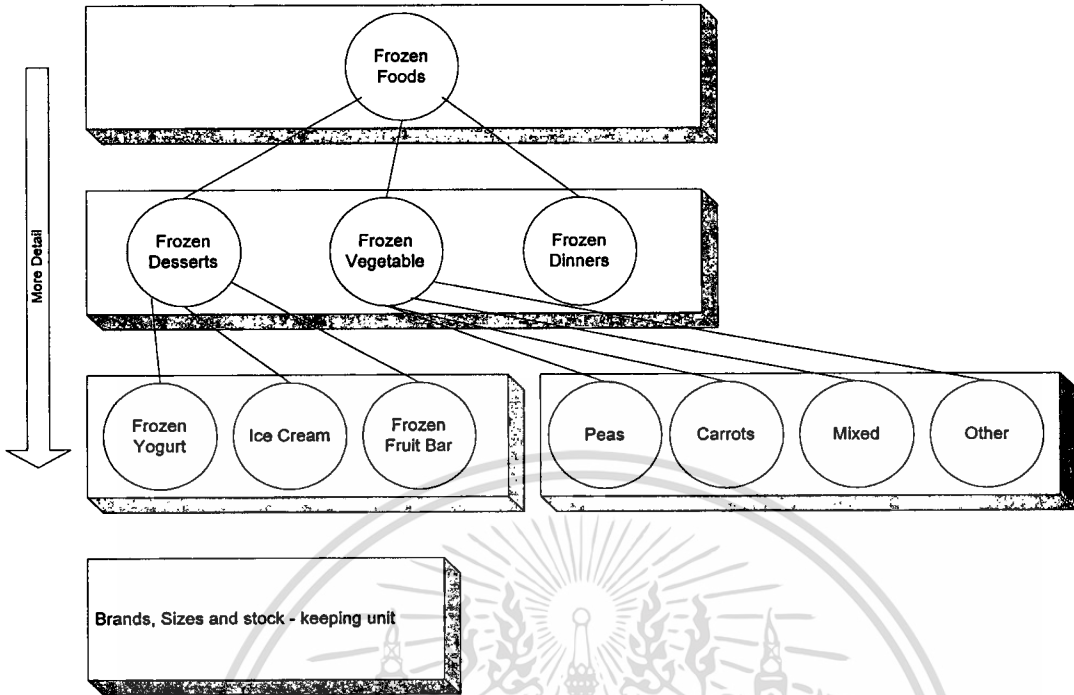
2.4.1 ขั้นตอนการทำ Market Basket Analysis

2.4.1.1 คัดเลือกกลุ่มข้อมูลที่เราสนใจ

ในร้านค้าหนึ่งๆ มีสินค้าอยู่มากมายหลายชนิด ซึ่งในการวิเคราะห์ความสัมพันธ์ จะมองที่ตัวสินค้าแต่ละตัวที่มีอยู่ว่ามีความสัมพันธ์กันอย่างไรทำให้ในการวิเคราะห์ประกอบไปด้วยข้อมูลจำนวนมาก ดังนั้นในการจะเลือกสินค้าชนิดใดมาใช้ในการวิเคราะห์ ควรจะพิจารณาถึงความละเอียดในการเจาะจงตัวสินค้า เช่น ถ้าในการขาย 1 เหตุการณ์ (Transaction) มีการซื้อพิซซ่าเพิ่มชีส ในการนำสินค้าชนิดนี้มาใช้เราอาจใช้ “พิซซ่า” แทนสินค้าตัวนี้ ซึ่งจะรวมไปถึง พิซซ่าประเภทอื่นๆ ด้วย หรือถ้าต้องการความละเอียด ก็จะใช้ “พิซซ่าเพิ่มชีส” เป็นตัวเจาะจงสินค้าชนิดนี้ ซึ่งในการตัดสินใจว่าจะใช้ตัวข้อมูลที่มีความละเอียดมากน้อยเพียงใดขึ้นอยู่กับวัตถุประสงค์ของงานนั้นๆ วิธีที่สามารถช่วยในการกำหนดตัวสินค้านี้

2.4.1.1.1 การแบ่งหมวดสินค้า (Taxonomy)

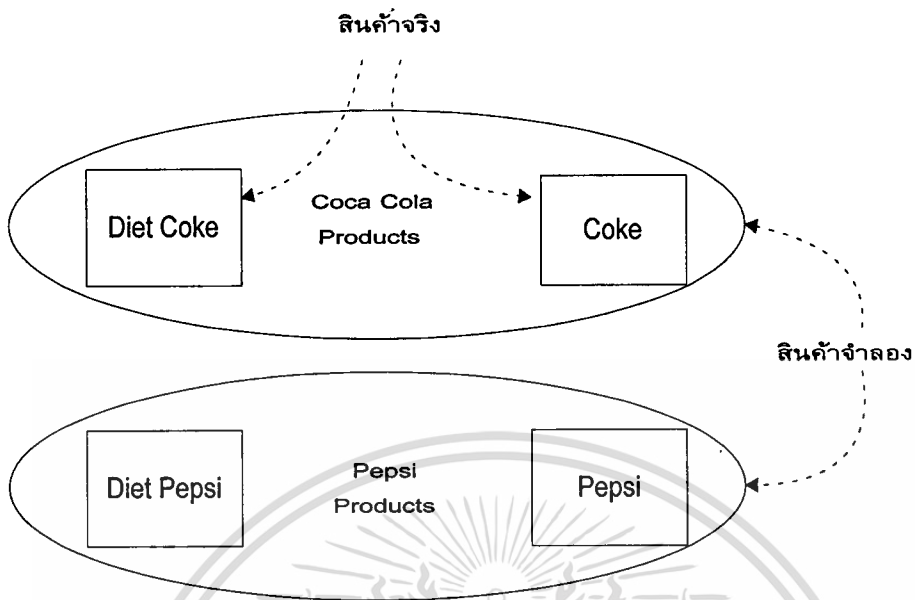
เป็นการแบ่งสินค้าแต่ละชนิดออกเป็นหมวดหมู่โดยแสดงออกมาในรูปแบบของ hierarchy จะแบ่งและสร้างกลุ่มของข้อมูลเป็นระดับ ๆ ระดับด้านล่างจะมีความละเอียดของสินค้ามากที่สุด เราจะใช้ชื่อสินค้าในระดับที่เราต้องการเป็นตัวแทนสินค้าที่อยู่ในระดับที่ต่ำกว่า เพื่อเป็นการลดจำนวนของสินค้าลง ซึ่งวิธีการนี้มีประโยชน์กับร้านที่มีจำนวนเหตุการณ์ (Transaction) มากๆ เช่นใน ซูเปอร์มาเก็ต ในการเลือกว่าจะใช้สินค้าในระดับในการวิเคราะห์ ก็ขึ้นกับความต้องการของงานนั้น ๆ ว่าต้องการความละเอียดของสินค้ามากน้อยเพียงใด เพราะในบางงานสินค้าในระดับล่างสุดก็จะสนับสนุนการสร้างกฎความสัมพันธ์ (Association Rule) ได้ดีกว่าสินค้าในระดับบน



รูปที่ 2.6 การจัดหมวดสินค้า [1]

2.4.1.1.2 การสร้างสินค้าจำลอง (Virtual item)

สินค้าจำลอง (Virtual item) คือสินค้าที่ไม่มีอยู่จริงในร้านค้า เป็นเพียงชื่อเรียกที่เราสร้างขึ้นเพื่อจะเพิ่มประสิทธิภาพในการวิเคราะห์ ทำให้จำนวนสินค้าที่ต้องการวิเคราะห์ลดลง และสามารถลดจำนวนความสัมพันธ์ที่เป็นลักษณะของความสัมพันธ์ตามธรรมชาติ เช่น ขนบปังกับแยม ซึ่งเป็นความสัมพันธ์ที่เกิดขึ้นได้เป็นปกติไม่จำเป็นต้องทำการวิเคราะห์ก็ทราบได้ หรือเป็นการจัดกลุ่มให้กับสินค้าเพื่อ โดยจัดให้อยู่ในระดับที่ต้องการวิเคราะห์ เช่น จัดตามยี่ห้อ, ประเภทของสินค้า เป็นต้น โดยที่สามารถรวมสินค้าจริงที่มีอยู่ในร้านหลายๆ ตัว มารวมเป็นกลุ่มเดียวกันได้ เช่น สินค้าจำลองคือ Calvin Klein แทนสินค้าที่เป็นเสื้อยี่ห้อ Calvin Klein และน้ำหอมยี่ห้อ Calvin Klein ไปด้วยกัน เป็นต้น เราควรที่จะสร้างสินค้าจำลองก็ต่อเมื่อสร้างแล้วสินค้าจำลองนั้นสามารถนำมาประยุกต์ใช้แล้วมีประสิทธิภาพดีกว่าการใช้ตัวสินค้าจริง



รูปที่ 2.7 การสร้างสินค้าจำลอง [1]

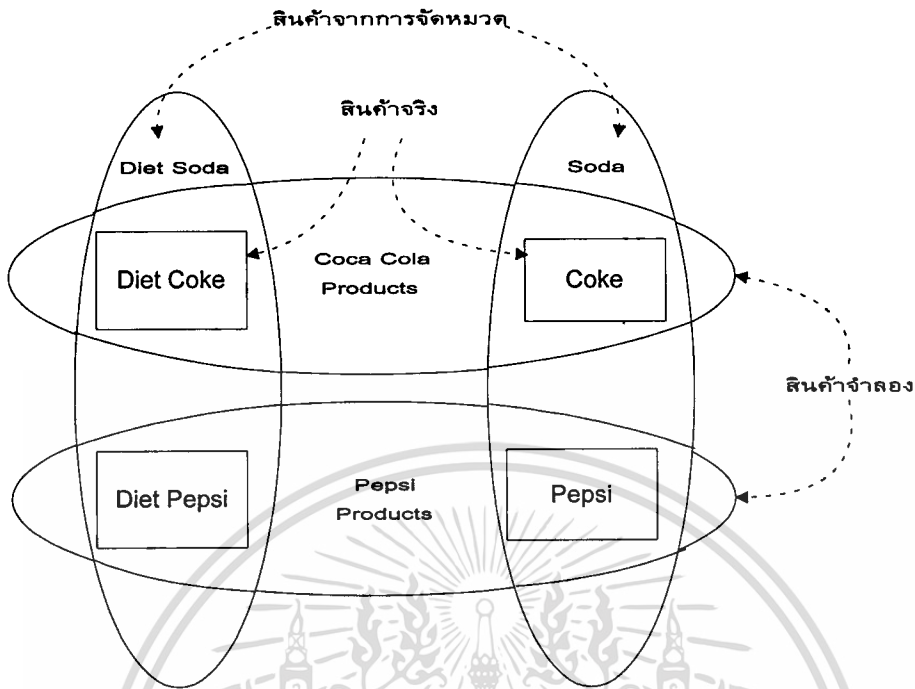
จากรูปที่ 2.7 แสดงการสร้างสินค้าจำลอง โดยจัดเป็นกลุ่มสินค้าตามยี่ห้อของสินค้า ดังนั้นสินค้าที่เป็น Coke จะนำมารวมกัน แล้วสร้างเป็นสินค้าจำลอง จะได้

Coca Cola Product แทนสินค้าที่เป็น Diet Coke และ Coke

เช่นเดียวกับสินค้ายี่ห้อ Pepsi นำมาจัดกลุ่มเข้าด้วยกัน แล้วสร้างเป็นสินค้าจำลอง จะได้

Pepsi Product แทนสินค้าที่เป็น Diet Pepsi และ Pepsi

การสร้างสินค้าจำลองในลักษณะนี้ อาจทำให้เกิดการซ้ำซ้อนของกฎความสัมพันธ์ที่จะเกิดขึ้น เช่นถ้าใน เหตุการณ์ (Transaction) มีการซื้อสินค้าที่เป็น Diet Coke เราก็จะแทนสินค้าตัวนั้นด้วย Coca Cola Product และถ้าใน เหตุการณ์ (Transaction) มีสินค้าที่เป็น Coke เราก็จะแทนด้วย Coca Cola Product เช่นกัน วิธีหลีกเลี่ยงความซ้ำซ้อนนี้สามารถทำได้โดยอาศัยสินค้าในการจัดหมวดหมู่ (Taxonomy) มาซ้อนทับกับสินค้าจำลองที่ตั้งไว้ เพื่อเจาะจงตัวสินค้าได้ดียิ่งขึ้น



รูปที่ 2.8 การแก้ปัญหาของสินค้าจำลอง [1]

จากรูปที่ 2.8 จะเห็นว่าการสร้างสินค้าจำลองโดยจัดกลุ่มสินค้าตามประเภทของสินค้า ด้วย โดยที่

Diet Soda แทนสินค้าที่เป็น Diet Coke และ Diet Pepsi

และ

Soda แทนสินค้าที่เป็น Coke และ Pepsi

ดังนั้นถ้ามีการตั้งกฎว่า

ถ้า “Coca Cola Product” และ “Diet Soda” แล้ว “Pretzels”

จะเท่ากับ

ถ้า “Diet Coke” แล้ว “Pretzels”

2.4.1.2 สร้างกฎหรือหลัก จากข้อมูลที่เลือกมา

กฎความสัมพันธ์ (Association Rule) ประกอบด้วย 2 ส่วนคือ ส่วนของเงื่อนไข และ ส่วนของผลลัพธ์ สามารถเขียนได้ในรูปแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$X \rightarrow Y$$

โดยที่ X เป็นเงื่อนไข และ Y เป็นผลลัพธ์

หรือ

"IF condition1 THEN condition2"

โดยเราจะเรียก X หรือ Condition 1 ว่า Left-Hand-Side (LHS) และเรียก Y หรือ Condition 2 ว่า Right-Hand-Side (RHS) ในการหาความสัมพันธ์จะมีค่าที่ต้องนำมาพิจารณาเพื่อให้ได้ความสัมพันธ์ที่เป็นไปตามวัตถุประสงค์ของการวิเคราะห์ คือ

Support คือ ความน่าจะเป็นที่สินค้าที่เป็น LHS และสินค้าที่เป็น RHS จะเกิดขึ้นในเหตุการณ์เดียวกัน

Confident คือ ความน่าจะเป็นที่จะเกิดสินค้าที่เป็น RHS ขึ้นในเหตุการณ์ที่มีการเกิดสินค้าที่เป็น LHS

Expected Confidence คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจ ซึ่งอาจเป็นส่วนของ LHS หรือ RHS อย่างใดอย่างหนึ่ง เทียบกับจำนวนเหตุการณ์ทั้งหมดภายในชุดข้อมูล

Lift คือ สัดส่วนระหว่างค่า Confidence กับ Expected Confidence ของจำนวนชุดข้อมูลที่เป็น RHS ค่า Lift ที่ได้จะช่วยในการตัดสินใจว่ากฎความสัมพันธ์ที่หาได้นั้น น่าสนใจมากน้อยเพียงใด

2.5 อีพอริ อัลกอริทึม (Apriori Algorithm)

เป็นอัลกอริทึมขั้นพื้นฐานสำหรับวิเคราะห์ความสัมพันธ์ในกระบวนการวิเคราะห์ข้อมูล โดยใช้กระบวนการของการหาความสัมพันธ์ (Association Rule)

2.5.1 หลักการของอีพอริ อัลกอริทึม (Apriori Algorithm)

1. หาความถี่ของการเกิดของเซตของข้อมูล โดยที่ความถี่ที่จะใช้พิจารณา จะต้องมากกว่าหรือเท่ากับค่า Support ที่น้อยที่สุด (Minimum Support) ที่กำหนด และทุกๆ สับเซตในเซตนั้นๆ ก็จะต้องมีค่าความถี่ของตัวเองด้วย

2. นำเซตข้อมูลที่ได้มาสร้างเป็นกฎความสัมพันธ์ และกฎที่ได้จะต้องเป็นไปตามเงื่อนไขของค่า Minimum Support และ ค่า Confident ที่ได้กำหนดไว้

ความหมายของตัวแปรที่ต้องพิจารณามีดังนี้

k-itemset คือ เซตของข้อมูลที่แต่ละเซตประกอบด้วยสมาชิกจำนวน k ตัว

L_k คือ เซตของ k-itemset (ซึ่งทุกเซตมีความถี่ในการเกิดมากกว่าหรือเท่ากับ Minimum Support)

C_{k+1} คือ เซตที่ถูกคัดเลือกแล้วของ k- itemset (ซึ่งได้จาก L_k)

2.5.2 ขั้นตอนการทำงานของเอปอริ อัลกอริทึม (Apriori Algorithm)

ขั้นที่ 1 ขั้นตอนการเชื่อม (Join Step) เป็นขั้นตอนการสร้าง C_k จากการเชื่อมกันของ L_{k-1} กับเซตของตัว L_{k-1} เอง

ขั้นที่ 2 ขั้นตอนการตัดทิ้ง (Prune Step) คือขั้นตอนของการตัดทิ้งสับเซต(k-1) ที่เป็นสมาชิกใน C_k ที่ไม่ได้เป็นสมาชิกของ L_{k-1} ออกจาก C_k

ขั้นที่ 3 ขั้นตอนการสร้างกฎความสัมพันธ์ เมื่อได้เซตของข้อมูลที่มีความสัมพันธ์กันในขั้นที่ 2 แล้ว นำเซตนั้นมาหากฎความสัมพันธ์โดยที่ กฎที่ได้จะต้องมีค่า Confident มากกว่าค่า Minimum Confident

Main Apriori Algorithm

$L_1 = \{ \text{large 1- itemset} \};$

For ($k = 1$; $L_k \neq \emptyset$; $k++$) do begin

$C_{k+1} = \text{Generate candidate}(L_k);$

For all transactions t in database do

Increment the count of all candidates in C_{k+1} that are contained in t ;

$L_{k+1} = \{ \text{candidates in } C_{k+1} \text{ with minimum support} \}$

End

Return $\bigcup_k L_k;$

รูปที่ 2.9 Main Pseudo Code

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Generate candidate function

● Join Step

Insert into C_k Select $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-1}$ Form $L_{k+1} p, L_{k+1} q$ Where $p.item_1 = q.item_1, \dots, p.item_{k-2} = q.item_{k-2},$ $p.item_{k-1} < q.item_{k-1}$

● Prune Step

For all itemset c in C_k doFor all $(k-1)$ -subsets s of c doIf (s is not in L_{k-1}) thenDelete c from C_k ;**รูปที่ 2.10 Generate Pseudo Code**Generating RulesFor all Large itemsets $L_k, k \geq 2$ doCall genrules(L_k, L_k)Procedure genrules(L_k : large k -itemset, A_m : Large m -itemset) $A = \{(m-1)\text{-itemsets } A_{m-1} \mid A_{m-1} \subset A_m\}$ For all $A_{m-1} \in A$ do beginConf = support(L_k) / support(A_{m-1})**รูปที่ 2.11 Generating Rules Pseudo Code**

```

If (Conf  $\geq$  minconf) then begin
    Output the rule  $A_{m-1} \Rightarrow (L_k - A_{m-1})$ , with
    Confident = conf and Support = support( $L_k$ )
    If (m-1 > 1) then
        Call genrules( $L_k, A_{m-1}$ )
    End
End
End

```

รูปที่ 2.11(ต่อ) Generating Rules Pseudo Code

2.5.3 ตัวอย่างของการทำงานของอโพรอริ-อัลกอริทึม (Apriori Algorithm)

จากหลักการของอโพรอริ อัลกอริทึม (Apriori Algorithm) ข้างต้น นำมาประยุกต์ใช้กับข้อมูลทางด้านโทรศัพท์เคลื่อนที่ เช่น ถ้าสนใจข้อมูลการขอใช้บริการเสริมของลูกค้า โดยสุ่มข้อมูลของลูกค้าจำนวน 10.หมายเลขโทรศัพท์ นำมาสร้างเป็นฐานข้อมูลเพื่อการวิเคราะห์

Mobile_no	Feature_code
8010000	3 6 7
8479273	6 7
9046842	1
8010005	6
9060994	3 7
9061885	1 2 3 6 7
8163885	6 7
9063529	3 4 5
8325556	5 6
8222332	5 6 7

ตารางที่ 2.1 หมายเลขโทรศัพท์และบริการเสริม
ที่เปิดให้บริการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

หาเซตของ C_1 โดยใช้จำนวนลูกค้ำที่ใช้บริการเป็นความถี่ของบริการเสริมแต่ละตัว ในการวิเคราะห์นี้จะกำหนดค่า Minimum Support = 2

เซตของลูกค้ำ	Support
{ 1 }	2
{ 2 }	1
{ 3 }	4
{ 4 }	1
{ 5 }	3
{ 6 }	7
{ 7 }	6

ตารางที่ 2.2 เซตของ C_1

เมื่อได้เซตของ C_1 แล้วพิจารณาค่าความถี่ในที่นี้คือค่า Support ทำการตัดเซตที่มีค่า Support ที่มีค่าน้อยกว่าค่า Minimum Support เพื่อจะได้เซต L_1

เซตของลูกค้ำ	Support
{ 1 }	2
{ 3 }	4
{ 5 }	3
{ 6 }	7
{ 7 }	6

ตารางที่ 2.3 เซตของ L_1

จากนั้นทำการวนรอบ เพื่อหาเซตของ C_k และ เซตของ L_k ต่อๆไป จนกระทั่ง ได้เซต ของ L_k ที่ไม่สามารถหา L_{k+1} ได้อีกดังนี้

เซตของสมาชิก	Support
{1, 2}	1
{1, 5}	0
{1, 6}	1
{1, 7}	1
{2, 3}	1
{3, 6}	2
{3, 7}	3
{5, 6}	2
{5, 7}	0
{6, 7}	5

ตารางที่ 2.4 เซตของ C_2

เซตของสมาชิก	Support
{3, 6}	2
{3, 7}	3
{5, 6}	2
{6, 7}	5

ตารางที่ 2.5 เซตของ L_2

เซตของสมาชิก	Support
{3, 6, 7}	2

ตารางที่ 2.6 เซตของ C_3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เซตของสัณฐาน	Support
{ 3, 6, 7 }	2

ตารางที่ 2.7 เซตของ L_3

สุดท้ายเราจะได้ เซตของ $L_3 = \{ 3, 6, 7 \}$ เป็นเซตของบริการเสริม ที่มีความสัมพันธ์กัน ในตัวอย่างนี้ เมื่อได้ผลการวิเคราะห์แล้วขั้นตอนถัดไปจะเป็นการนำเอาผลที่ได้พิจารณาสร้างเป็นกฎความสัมพันธ์ จากอัลกอริทึมข้างต้น ต้องหากฎที่ได้จากทั้ง L_2 และ L_3 ของเซต { 3, 6, 7 } จะทำให้เราได้กฎความสัมพันธ์และค่า Confident ของแต่ละกฎดังแสดงในตารางที่ 2.8

กฎความสัมพันธ์	ค่า Confident
3 \Rightarrow 6	0.50
3 \Rightarrow 7	0.50
6 \Rightarrow 3	0.28
6 \Rightarrow 7	0.71
7 \Rightarrow 3	0.50
7 \Rightarrow 6	0.83
3 \Rightarrow 6,7	0.50
6 \Rightarrow 3,7	0.28
7 \Rightarrow 6,3	0.33
3,6 \Rightarrow 7	1.00
3,7 \Rightarrow 6	0.66
6,7 \Rightarrow 3	0.40

ตารางที่ 2.8 กฎความสัมพันธ์และค่า Confident ของแต่ละกฎ

ถ้าในการหาความสัมพันธ์กำหนดค่า Minimum Confident = 0.5 ดังนั้นกฎที่เราให้ความสนใจจะเหลือเพียงกฎที่มีค่า Confident มากกว่าหรือเท่ากับค่า Minimum Confident เท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กฎความสัมพันธ์	ค่าConfident
3 \Rightarrow 6	0.50
3 \Rightarrow 7	0.50
6 \Rightarrow 7	0.71
7 \Rightarrow 6	0.83
3 \Rightarrow 6,7	0.50
3,6 \Rightarrow 7	1.00
3,7 \Rightarrow 6	0.66

ตารางที่ 2.9 กฎความสัมพันธ์ที่มีค่า Confident มากกว่าค่า Minimum Confident

จากนั้นทำการพิจารณาเลือกกฎที่มีค่า Confident ที่น่าพอใจ ซึ่งจากตารางที่ 2.9 กฎความสัมพันธ์ที่ดีที่สุด คือ

$$3, 6 \Rightarrow 7 \text{ ซึ่งมีค่า Confident} = 1$$

ดังนั้นสามารถกล่าวได้ว่า

“ถ้ามีการใช้บริการที่ 3 และ บริการที่ 6 แล้ว ลูกค้าจะมีการใช้บริการที่ 7 ด้วย”

เมื่อสามารถหากฎความสัมพันธ์แล้ว ก็สามารถนำกฎที่ได้มาประยุกต์ใช้ต่อไป เช่น ทำโฆษณา (Campaign) โดยการจัดชุด (Package) ของบริการเสริม ที่เราพบว่ามีความสัมพันธ์กัน เพื่อดึงดูดความสนใจของลูกค้า เป็นต้น

บทที่ 3

วิธีดำเนินการศึกษาเพื่อทำการวิเคราะห์ความสัมพันธ์ของบริการเสริม ในระบบโทรศัพท์เคลื่อนที่

3.1 กำหนดวัตถุประสงค์

เนื่องจากในระบบของโทรศัพท์เคลื่อนที่ที่มีความต้องการที่จะจัดชุดของบริการเสริมเพื่อเพิ่มจุดสนใจให้กับลูกค้าที่จะทำการขอเปิดบริการ แต่จะต้องจัดชุดให้ได้ใกล้เคียงกับความต้องการของลูกค้ามากที่สุดจึงจะเกิดผลดีกับบริษัท จึงเกิดแนวความคิดที่จะวิเคราะห์ข้อมูลที่มีอยู่ในปัจจุบัน เพื่อให้ได้เห็นถึงรูปแบบการขอเปิดบริการเสริมของลูกค้าโดยทั่วไป ว่าให้ความสนใจกับบริการเสริมประเภทใดบ้าง และบริการเสริมประเภทใด ที่มักจะถูกรขอเปิดบริการด้วยกันเสมอๆ จึงเป็นวัตถุประสงค์ที่ทำให้เกิดมีการพัฒนาระบบที่สามารถจะวิเคราะห์หาความสัมพันธ์ของบริการเสริมเหล่านั้นได้ เพื่อเพิ่มประสิทธิภาพและความสะดวกแก่การวิเคราะห์ข้อมูล

3.2 การคัดเลือกข้อมูล

ในระบบของการสื่อสารในรูปแบบของโทรศัพท์เคลื่อนที่จะมีข้อมูลเกี่ยวกับลูกค้าแต่ละรายเป็นจำนวนมาก เพื่อเก็บข้อมูลในส่วนต่างๆ ที่มีความเกี่ยวข้องกับลูกค้าในแต่ละด้าน เช่น ข้อมูลการจดทะเบียนเพื่อเปิดบริการของลูกค้า ข้อมูลทางด้านการชำระค่าใช้บริการรายเดือน ข้อมูลการขอใช้บริการต่างๆ รวมถึงข้อมูลการขอเปลี่ยนแปลงแก้ไขต่างๆ ของลูกค้า จึงถือได้ว่าเป็นฐานข้อมูลที่มีขนาดใหญ่ ดังนั้นในการที่จะนำข้อมูลมาทำการวิเคราะห์นั้น ผู้ทำการวิเคราะห์จะต้องมีความเข้าใจถึงข้อมูลต่างๆ ในระบบก่อน เพื่อที่จะใช้ในการพิจารณาว่าข้อมูลใดบ้างที่จะมีผลต่อการวิเคราะห์ข้อมูล เนื่องจากส่วนของการคัดเลือกข้อมูลนั้น นับได้ว่ามีผลต่อผลของการวิเคราะห์ข้อมูลเป็นอย่างมาก ถ้าทำการคัดเลือกข้อมูลที่ไม่เหมาะสมไปทำการวิเคราะห์ ผลการวิเคราะห์ที่ได้จะไม่เป็นไปตามวัตถุประสงค์ที่ได้กำหนดไว้ ก็เท่ากับว่าการวิเคราะห์นั้นไม่เป็นผลสำเร็จ จำเป็นต้องมีการเริ่มทำการรวบรวมการวิเคราะห์ใหม่อีกครั้ง

สำหรับในโครงการนี้ ได้คัดเลือกข้อมูลที่มีความสัมพันธ์เกี่ยวข้องกับลูกค้าในส่วนของการขอใช้บริการเสริมต่างๆ ในระบบโทรศัพท์เคลื่อนที่ และได้ทำการจำลองฐานข้อมูลนั้นในรูปแบบฐานข้อมูลของ Microsoft Access ซึ่งข้อมูลที่มีประกอบด้วย

ชื่อข้อมูล	ประเภทของข้อมูล	ความหมาย
Customer_ID	Varchar2(10)	เลขที่ลูกค้าจะได้รับตอนจดทะเบียน
Customer_Type	Varchar2(2)	รหัสสำหรับประเภทลูกค้า
Customer_Name	Varchar2(100)	ชื่อขององค์กรหรือชื่อของบุคคล
Last_Name	Varchar2(100)	นามสกุลของลูกค้า
Mobile_No.	Varchar2(10)	หมายเลขโทรศัพท์เคลื่อนที่ของลูกค้า
Account_Status	Varchar2(2)	สถานะของเลขหมาย
Language	Varchar2(2)	ภาษาที่ใช้กับลูกค้า
Bill_Cycle	Number(2)	รอบบิลสำหรับลูกค้า
Bill_Format	Number(2)	รูปแบบของบิลตามที่ต้องการ
Group_Code	Varchar2(6)	ชื่อพิเศษที่ใช้แสดงถึงกลุ่มของลูกค้า
ID_Card_Type	Varchar2(3)	ประเภทของบัตรแสดงลักษณะเฉพาะของบุคคล <ol style="list-style-type: none"> 1. บัตรประชาชน 2. บัตรข้าราชการ 3. หนังสือเดินทาง 4. หนังสือจดทะเบียนการค้า
ID_Card_No.	Varchar2(20)	เลขที่บัตรใช้แสดงลักษณะเฉพาะของบุคคล
Register_Type	Varchar2(3)	ประเภทของ ID จะแสดงว่าเป็นองค์กรใด
Register_No	Varchar2(20)	หมายเลขของ ID
Segment_Date	Date	วันที่ลูกค้าถูกจำแนกกลุ่ม
Segment_Probable	Number(3)	เปอร์เซ็นต์ที่ลูกค้าจะออกจากกลุ่ม
Bill_Deposit	Number(12)	เงินประกันของลูกค้า
Payment_Method	Varchar2(2)	ประเภทของการชำระเงิน
Payment_Term	Varchar2(1)	ช่วงการชำระ รายสัปดาห์, รายเดือน
Credit_Limit	Number(12)	จำนวนสูงสุดที่ยอมให้ลูกค้าเป็นหนี้ใน 1 รอบบิล (ปัจจุบันยังไม่ได้ใช้)

ตารางที่ 3.1 ตารางข้อมูลลูกค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อข้อมูล	ประเภทของข้อมูล	ความหมาย
Credit_Rate	Number(12)	คะแนนสะสมเครดิตแก่ลูกค้า
Churn_Date	Date	วันที่ลูกค้ายกเลิกการใช้บริการ
Churn_Type	Varchar2(2)	รหัสแสดงประเภทของการยกเลิกบริการ
Churn_Probability	Number(3)	เปอร์เซ็นต์ที่ลูกค้าจะยกเลิกบริการ
Active_Date	Date	วันที่ลูกค้าเข้ามาในระบบ
Bank_Code	Number(3)	รหัสธนาคาร
Credit_Card_Type	Varchar2(2)	ประเภทของบัตรเครดิต

ตารางที่ 3.1 (ต่อ) ตารางข้อมูลลูกค้า

ชื่อข้อมูล	ประเภทของข้อมูล	ความหมาย
Feature_Code	Number(4)	รหัสบริการเสริม
Desc_Thai	Varchar2(50)	คำบรรยายภาษาไทย
Desc_Eng	Varchar2(50)	คำบรรยายภาษาอังกฤษ
Effective_Date	Date	วันที่เริ่มใช้งาน
Expire_Date	Date	วันสุดท้ายที่ใช้งาน

ตารางที่ 3.2 ตารางบริการเสริมในระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อข้อมูล	ประเภทของข้อมูล	ความหมาย
Location_Code	Number(4)	รหัสสาขาที่ทำการขอเปิดบริการ
Order_No	Number(10)	ลำดับการขอทำรายการ
Cust_Group_Acc	Varchar2(10)	เลขที่ลูกค้าจะได้รับตอนจดทะเบียน
Mobile_No	Varchar2(10)	หมายเลขโทรศัพท์เคลื่อนที่ของลูกค้า
Feature_Code	Number(4)	รหัสบริการเสริมที่ขอเปิดบริการ
Tariff_Package_Code	Varchar2(4)	ประเภทของการคิดค่าบริการ
Effective_Date	Date	วันที่เริ่มใช้บริการเสริม
Action_Type	Varchar2(1)	สถานะของบริการเสริม

ตารางที่ 3.3 ตารางการขอเปิดใช้บริการเสริม

ชื่อข้อมูล	ประเภทของข้อมูล	ความหมาย
Customer_ID	Varchar2(10)	เลขที่ลูกค้าจะได้รับตอนจดทะเบียน
Mobile_No	Varchar2(10)	หมายเลขโทรศัพท์เคลื่อนที่ของลูกค้า
Feature_Code	Number(4)	รหัสบริการเสริมที่ขอเปิดบริการ
Seq_No	Number(3)	ลำดับที่ของบริการเสริมที่ขอเปิด
Network_Type	Varchar2(3)	เครือข่ายที่ใช้บริการ
Effective_Date	Date	วันที่เริ่มใช้บริการเสริม
Expire_Date	Date	วันที่สิ้นสุดการใช้บริการเสริม

ตารางที่ 3.4 ตารางประวัติการขอใช้บริการเสริมของลูกค้า

จากข้อมูลดิบข้างต้น นำมาพิจารณาถึงรูปแบบการจัดเก็บข้อมูล ประเภทของข้อมูล รวมถึงค่าที่เป็นไปได้ในแต่ละตัวข้อมูล ว่ามีรูปแบบที่เหมาะสม สามารถนำไปวิเคราะห์ได้สะดวกหรือไม่ สำหรับข้อมูลข้างต้น ที่ต้องการนำว่าหาความสัมพันธ์ของการเปิดใช้บริการเสริม โดยที่บริการเสริมที่โครงการนี้ให้ความสนใจ คือบริการเสริม 7 ประเภทดังต่อไปนี้

ชื่อบริการเสริมภาษาไทย	ชื่อบริการเสริมภาษาอังกฤษ
บริการรับสายเรียกซ้อน	Call Waiting (CW)
บริการระงับการใช้เครื่อง	Call Barring (CB)
บริการรับฝากข้อความ	Voice Mail (VM)
บริการโอนสายอัตโนมัติ	Call Forwarding (CF)
บริการประชุมทางโทรศัพท์	Call Conference(CC)
บริการแสดงหมายเลขเรียกเข้าอัตโนมัติ	Calling Line Identification Presentation (CLI)
บริการรับส่งข้อความอัตโนมัติ	Short Message Service(SMS)

ตารางที่ 3.5 ตารางของบริการเสริมที่สนใจ

ดังนั้นสามารถจัดการให้ข้อมูลที่สนใจอยู่ในรูปแบบที่เหมาะสมกับการวิเคราะห์ได้ดังนี้ คือทำการสร้างตารางข้อมูลใหม่ เพื่อเก็บข้อมูลของการเปิดใช้บริการเสริมของแต่ละหมายเลข โทรศัพท์เคลื่อนที่ จะได้ตารางข้อมูลดังนี้

ชื่อข้อมูล	ประเภทของข้อมูล	ความหมาย
Mobile_No	Varchar2(10)	หมายเลขโทรศัพท์เคลื่อนที่ของลูกค้า
CW	Boolean	บริการรับสายเรียกซ้อน
CB	Boolean	บริการระงับการใช้เครื่อง
VM	Boolean	บริการรับฝากข้อความ
CF	Boolean	บริการ โอนสายอัตโนมัติ
CC	Boolean	บริการประชุมทางโทรศัพท์
CLI	Boolean	บริการแสดงหมายเลขเรียกเข้าอัตโนมัติ
SMS	Boolean	บริการรับส่งข้อความอัตโนมัติ

ตารางที่ 3.6 ตารางการใช้บริการเสริมของแต่ละหมายเลขโทรศัพท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การแปลงและทำความสะอาดข้อมูล

ในขั้นตอนการแปลงและการทำความสะอาดข้อมูลเพื่อให้ได้ผลการวิเคราะห์ที่ถูกต้องมากที่สุด จำเป็นจะต้องทำให้ข้อมูลสะอาดและ สะดวกต่อการวิเคราะห์มากที่สุด โดยในส่วนนี้ จะขอกกล่าวถึงการแปลงหรือการจัดการข้อมูลก่อนที่จะนำไปผ่านระบบวิเคราะห์ข้อมูลโดยโปรแกรมที่พัฒนาขึ้น โดยเริ่มตั้งแต่การคัดข้อมูลที่จะนำมาทำการวิเคราะห์ จะเลือกจากฐานข้อมูลที่มีอยู่ ดังตารางที่ 3.1 ถึง ตารางที่ 3.4 โดยมีเงื่อนไขคือ ข้อมูลที่ไม่ระบุรหัสของบริการเสริม (Feature_Code) หรือข้อมูลที่ระบุรหัสบริการเสริมแต่ไม่อยู่ในขอบเขตที่พิจารณา จะถูกคัดทิ้ง

3.4 ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ระบบที่ทำการพัฒนาขึ้น

เมื่อได้ฐานข้อมูลที่จะทำการวิเคราะห์แล้ว ขั้นตอนในการทำการวิเคราะห์โดยใช้โปรแกรมมีขั้นตอนดังนี้

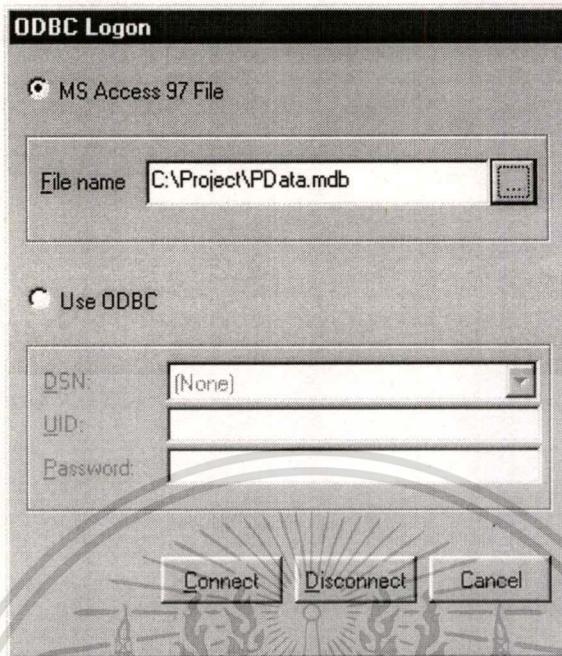
3.4.1 ติดต่อฐานข้อมูล

ในการทำการวิเคราะห์ข้อมูลโดยใช้โปรแกรมที่พัฒนาขึ้น ให้ทำการเลือกฐานข้อมูลที่ต้องการทำการวิเคราะห์ โดยในโครงงานนี้ฐานข้อมูลที่ใช้จะอยู่ในรูปแบบของ Microsoft Access (*.mdb) ซึ่งโปรแกรมนี้สามารถรองรับได้ทั้งการติดต่อฐานข้อมูลของ Microsoft Access โดยตรง และติดต่อฐานข้อมูลผ่าน ODBC

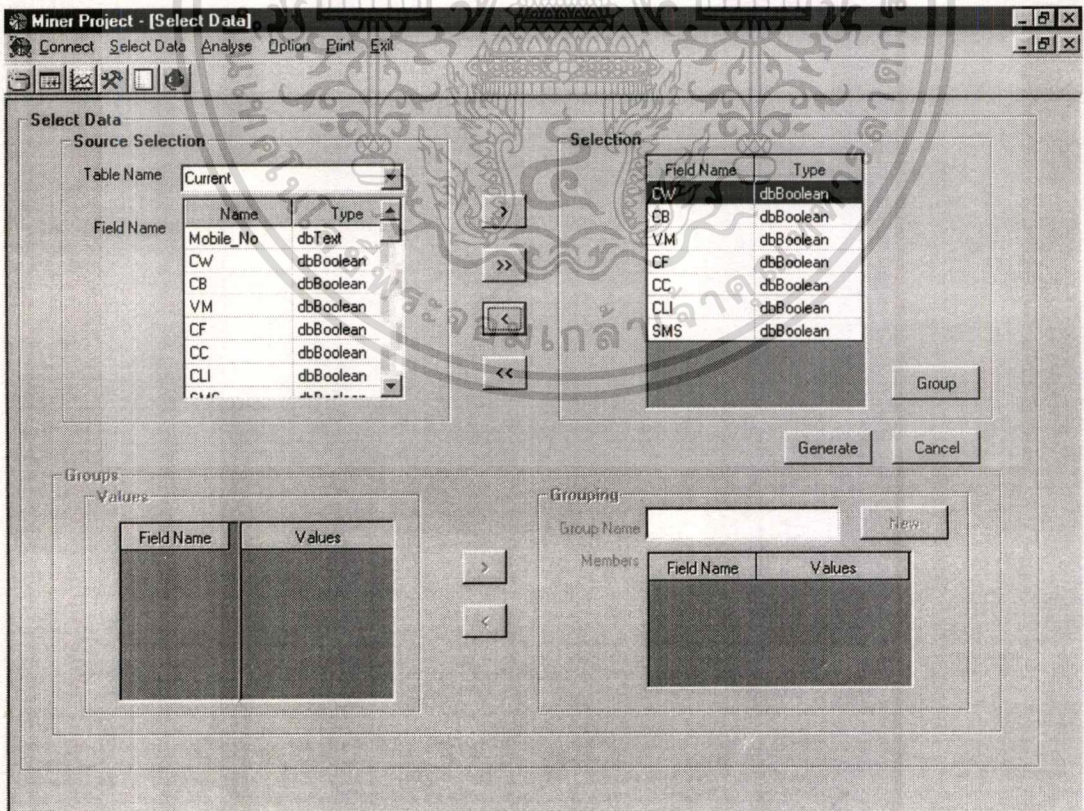
3.4.2 คัดเลือกข้อมูล

เลือกตารางข้อมูลที่ต้องการทำการวิเคราะห์ ซึ่งเป็นตารางที่มีการคัดเลือกข้อมูลและจัดรูปแบบเรียบร้อยแล้ว (ตารางที่ 3.6) ในการคัดเลือกข้อมูลนี้เป็นการเลือกเฉพาะบริการเสริมที่สนใจในการวิเคราะห์ ซึ่งอาจจะเลือกบางบริการเสริมหรือจะเลือกทั้งหมดก็ได้ รวมถึงในขั้นตอนการคัดเลือกนี้สามารถจัดกลุ่มของบริการเสริมได้ ก่อนที่จะทำการวิเคราะห์ ในโครงงานนี้ สนใจบริการเสริมทั้งสิ้น 7 ประเภท ดังที่ได้กล่าวไว้ข้างต้น

* ขั้นตอนการทำงานของโปรแกรมโดยละเอียด จะกล่าวในภาคผนวก
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์หรือมีการสงวนเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 หน้าจอเพื่อติดต่อฐานข้อมูล



รูปที่ 3.2 หน้าจอการคัดเลือกข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3 ผลการวิเคราะห์ข้อมูล

เมื่อทำการคัดเลือกข้อมูลที่จะให้โปรแกรม สามารถที่จะสั่งให้โปรแกรมทำการวิเคราะห์หา กฎความสัมพันธ์ได้ทันที โดยที่ผลของการวิเคราะห์ที่ได้จะใช้ค่า Minimum Support, ค่า Minimum Confident และจำนวนสมาชิกในเซต ที่เป็นค่า Default ของโปรแกรม แต่ถ้าต้องการกำหนดค่า Minimum Support, ค่า Minimum Confident และจำนวนสมาชิกในเซต ตามที่ต้องการ สามารถทำได้โดยกำหนดค่าต่างๆ ในโปรแกรม ดังรูปที่ 3.3

The image shows a dialog box titled "Option" with the following settings:

- Support:**
 - Minimum Support: 0.5 (range 0.0-1.0)
 - Minimum Frequency: 20
- Confident:**
 - Minimum Confident: 0.5 (range 0.0-1.0)
- Item Set:**
 - Maximum Item in set: 3 Item(s)

Buttons: OK, Default

รูปที่ 3.3 หน้าจอกำหนดค่า

ในการแสดงผล โปรแกรมจะแสดงสมาชิกในแต่ละเซตที่สามารถเกิดกฎความสัมพันธ์ได้ และมีค่า Support มากกว่าหรือเท่ากับค่า Minimum Support ที่กำหนด โดยแสดงค่า Support ที่เกิดขึ้นของแต่ละเซตไว้ด้วย และแสดงกฎความสัมพันธ์ที่สามารถเกิดขึ้นของเซตข้อมูลนั้น ซึ่งเป็นกฎความสัมพันธ์ที่มีค่า Confident มากกว่าหรือเท่ากับค่า Minimum Confident ที่กำหนด ซึ่งก็จะแสดงเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า Confident ของแต่ละกฎความสัมพันธ์ไว้เช่นกัน ในการแสดงกฎความสัมพันธ์จะบอกถึงส่วนที่เป็น Left-Hand-Side (LHS) และส่วนที่เป็น Right-Hand-Side (RHS) ของแต่ละกฎที่เกิดขึ้น

Miner Project - [Rule Generate]

Connect Select Data Analyse Option Print Exit

Association

	LHS	RHS	Support	Confident
Set 1	CW = N VM = Y		0.4367	
		VM = Y		0.8901
Set 2	CW = N VM = Y	CW = N	0.5529	
		VM = Y		0.8846
Set 3	CW = N CLI = Y	CLI = Y	0.5367	
		CW = N		0.8626
Set 4	CW = N SMS = Y	SMS = Y	0.5689	
		CW = N		0.7091
Set 5	CW = Y CB = Y	CB = Y	0.6119	
		CW = Y		0.7356
Set 6	CW = Y VM = Y	VM = Y	0.6298	
		VM = Y		0.8421
Set 7	CW = Y CLI = Y	CLI = Y	0.3504	
		CLI = Y		
Set 8	CW = Y SMS = Y	SMS = Y		
		SMS = Y		
Set 9	CB = N VM = Y	VM = Y		
		VM = Y		
Set 10	CB = N CLI = Y	CLI = Y		
		CLI = Y		

Total Rules: 228 Sort OK

รูปที่ 3.4 ตารางแสดงผลการวิเคราะห์

จากผลที่ได้จากการวิเคราะห์ เราจะได้กฎความสัมพันธ์ออกมาจำนวนหนึ่ง ซึ่งในขั้นต่อไปในการทำดาต้าไมนิง คือ นำผลที่ได้ไปทำการวิเคราะห์ เพื่อนำกฎที่มีประโยชน์ ซึ่งตรงกับวัตถุประสงค์ของการวิเคราะห์ มาประยุกต์ใช้ต่อไป

3.4.4 การวิเคราะห์ข้อมูล

ตารางของกฎความสัมพันธ์ที่ได้ เราสามารถนำมาหากฎที่คาดว่าจะมีประโยชน์และตรงตามวัตถุประสงค์ของการวิเคราะห์ที่ได้ตั้งไว้ คือการจัดชุดของบริการเสริมให้ตรงกับความต้องการของลูกค้า ดังนั้นพิจารณาผลการวิเคราะห์

จากข้อมูลตัวอย่างการขอใช้บริการเสริม และกำหนดเงื่อนไขการหาความสัมพันธ์ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ค่า Minimum Support = 0.45
- ค่า Minimum Confident = 0.5
- ค่า Maximum ItemSet = 3

ดังนั้นผลที่ได้จากโปรแกรมหากฎความสัมพันธ์ ดังแสดงในรูปที่ 3.5

	LHS	RHS	Support	Confident
Set 1	SMS = Y CLI = Y		0.6765	
	SMS = Y	CLI = Y		0.9094
	CLI = Y	SMS = Y		0.8366
Set 2	VM = Y CLI = Y		0.6658	
	VM = Y	CLI = Y		0.8430
	CLI = Y	VM = Y		0.8234
Set 3	SMS = Y VM = Y		0.6334	
	SMS = Y	VM = Y		0.8515
	VM = Y	SMS = Y		0.8020
Set 4	CLI = Y SMS = Y VM = Y		0.5876	
	CLI = Y SMS = Y	VM = Y		0.8686
	CLI = Y	VM = Y SMS = Y		0.7267
	SMS = Y	VM = Y CLI = Y		0.7899
	VM = Y	CLI = Y SMS = Y		0.7440
Set 5	SMS = Y CF = N		0.4528	
	SMS = Y	CF = N		0.6087
Set 6	VM = Y SMS = Y CLI = Y		0.5876	
	VM = Y SMS = Y	CLI = Y		0.9277
Set 7	CF = N VM = Y		0.4744	
	CF = N	VM = Y		0.8713
Set 8	VM = Y CLI = Y SMS = Y		0.5876	

Total Rules: 20 Sort OK

รูปที่ 3.5 ผลการวิเคราะห์ข้อมูล

ผลที่ได้มีกฎความสัมพันธ์ทั้งสิ้น 20 กฎ ซึ่งแต่ละกฎมีค่า Support และค่า Confident มากกว่าค่าต่ำสุดที่กำหนด จากนั้นเราจะพิจารณาเฉพาะกฎที่ตรงกับวัตถุประสงค์ในการวิเคราะห์ที่ได้กำหนดไว้ตั้งแต่ต้น คือกฎที่มีการขอเปิดบริการเสริมเหล่านั้นด้วยกัน และคัดเฉพาะที่มีค่า Support และค่า Confident สูงๆ 4 อันดับ ซึ่งแสดงว่าเป็นกฎที่มีการเกิดขึ้นได้บ่อย กฎที่น่าสนใจจะแสดงในตารางที่

3.7

LHS	RHS	Support	Confident
SMS = Y	CLI = Y	0.6765	0.9094
VM = Y	CLI = Y	0.6658	0.8430
SMS = Y	VM = Y	0.6334	0.8515
CLI = Y	VM = Y SMS = Y	0.5876	0.8686

ตารางที่ 3.7 กฎความสัมพันธ์ที่มีค่า Support และค่า Confident สูง 4 อันดับแรก

จากตารางที่ 3.7 ทำให้เราทราบถึงกฎที่น่าสนใจคือ

- กฎที่ 1 ถ้ามีการเปิดบริการ Short Message Service แล้วจะมีการเปิดบริการแสดงหมายเลขเรียกเข้าอัตโนมัติ ด้วยค่า Support = 0.6765 และค่า Confident = 0.9094
- กฎที่ 2 ถ้ามีการเปิดบริการ Voice Mail แล้วจะเปิดบริการ แสดงหมายเลขเรียกเข้าอัตโนมัติ ด้วยค่า Support = 0.6658 และค่า Confident = 0.8430
- กฎที่ 3 ถ้ามีการเปิดบริการ Short Message Service แล้ว จะเปิดบริการ Voice Mail ด้วยค่า Support = 0.6334 และค่า Confident = 0.8515
- กฎที่ 4 ถ้ามีการเปิดบริการแสดงหมายเลขเรียกเข้าอัตโนมัติ แล้วจะมีการเปิดบริการ Short Message Service และ Voice Mail ด้วยค่า Support = 0.5876 และค่า Confident = 0.9277

พิจารณากฎทั้ง 4 เราสามารถที่จะนำทุกกฎมาทำเป็นชุดของบริการเสริม หรือจะคัดเฉพาะกฎที่เราให้ความสนใจมากที่สุดก็ได้ จากกฎข้างต้น ถ้านำทุกกฎมาจัดเป็นชุดของบริการเสริม เราจะสามารถจัดได้ทั้งสิ้น 4 ชุด ดังนั้นในการที่จะจัดชุดของบริการเสริมจะได้ชุดของบริการเสริมดังนี้

บริการเสริมชุดที่ 1 ประกอบด้วย

- บริการแสดงหมายเลขเรียกเข้าอัตโนมัติ
- บริการรับส่งข้อความ

บริการเสริมชุดที่ 2 ประกอบด้วย

- บริการแสดงหมายเลขเรียกเข้าอัตโนมัติ
- บริการรับฝากข้อความอัตโนมัติ

บริการเสริมชุดที่ 3 ประกอบด้วย

- บริการรับฝากข้อความอัตโนมัติ
- บริการรับส่งข้อความ

บริการเสริมชุดที่ 4 ประกอบด้วย

- บริการรับฝากข้อความอัตโนมัติ
- บริการรับส่งข้อความ
- บริการแสดงหมายเลขเรียกเข้าอัตโนมัติ



บทที่ 4

สรุปผลการศึกษาและข้อเสนอแนะ

4.1 สรุปผลการศึกษา

โครงการที่ได้ศึกษานี้ ถูกจัดทำขึ้นเพื่อให้ผู้จัดทำได้ศึกษาถึงหลักการทำงานของการทำดาต้าไมนิ่ง โดยเลือกที่จะใช้การทำดาต้าไมนิ่งในรูปแบบของลิ่งค่อนาไลซิซ ซึ่งสามารถนำมาประยุกต์ใช้ในธุรกิจที่มีอยู่จริงได้ เพื่อให้ได้ผลของการวิเคราะห์ที่มีประโยชน์ต่อองค์กรที่เกี่ยวข้อง โดยการประยุกต์ใช้เทคนิคของการหากฎความสัมพันธ์ ซึ่งเป็นเทคนิคหนึ่งของลิ่งค่อนาไลซิซ เป็นการประยุกต์ขั้นพื้นฐาน โดยที่ผู้จัดทำนำเทคนิคและอัลกอริทึมของการหากฎความสัมพันธ์มาประยุกต์เพื่อสร้างโปรแกรมที่สามารถวิเคราะห์หากฎความสัมพันธ์ของบริการเสริมที่มีในระบบโทรศัพท์เคลื่อนที่ในปัจจุบัน โดยโปรแกรมสามารถทำการวิเคราะห์ข้อมูลและมีความยืดหยุ่นได้ในระดับหนึ่งเท่านั้น แต่ถือได้ว่าเป็นระดับพื้นฐานที่ยังสามารถนำไปพัฒนาให้เป็น โปรแกรมในการวิเคราะห์ข้อมูลที่ซับซ้อนมากขึ้นได้

4.2 ข้อเสนอแนะ

1. อีโพอริ อัลกอริทึม ที่นำมาใช้ในการพัฒนาโปรแกรม ถือเป็นอัลกอริทึมขั้นพื้นฐานในการหากฎความสัมพันธ์ของข้อมูล ดังนั้นประสิทธิภาพในการทำงานของโปรแกรมจะอยู่ในระดับหนึ่งเท่านั้น ถ้าในการพัฒนาโปรแกรมที่จะนำมาใช้วิเคราะห์ข้อมูลที่มีอยู่จริงในระบบโทรศัพท์เคลื่อนที่ ควรที่จะใช้อัลกอริทึมที่สามารถทำการวิเคราะห์ข้อมูลที่ซับซ้อนมากขึ้น และเป็นข้อมูลที่มีปริมาณมากๆ หรือมีขั้นตอนการทำงานที่กระชับขึ้น เพื่อให้ผลการวิเคราะห์ที่ได้มีความถูกต้องที่สุด และมีประโยชน์มากที่สุด
2. ในการวิเคราะห์ความสัมพันธ์ในโครงการที่ได้จัดทำขึ้น จะเป็นการวิเคราะห์ว่าบริการเสริมแต่ละตัวมีความสัมพันธ์กันอย่างไรเท่านั้น โดยนำข้อมูลจากการเปิดใช้บริการเสริมของลูกค้าแต่ละรายเท่านั้น ดังนั้นในการวิเคราะห์เพื่อที่จะได้ผลการวิเคราะห์ที่ถูกต้องมากขึ้น ควรนำปัจจัยอื่นมาร่วมในการวิเคราะห์ด้วย เช่น พื้นที่ใช้งานของแต่ละหมายเลขโทรศัพท์ อายุ และปัจจัยอื่นๆ ที่อาจจะมีผลต่อการขอเปิดบริการเสริมแต่ละประเภทตัว

3. จากที่ได้ทำการศึกษากระบวนการต่างๆ ในการหาคุณค่าความสัมพันธ์ เห็นว่าเป็นกระบวนการที่น่าสนใจ ที่สามารถนำมาประยุกต์ใช้ได้กับในหลายๆ ธุรกิจ ซึ่งจะทำให้เกิดประโยชน์กับธุรกิจต่างๆ ได้ ผู้จัดหวังว่าในการทำโครงการนี้ขึ้นจะสามารถทำให้ผู้ที่สนใจในการวิเคราะห์ข้อมูลที่ตนเองหรือองค์กรมีอยู่ ให้กลายเป็นข้อมูลที่มีประโยชน์มากขึ้น ได้รับแนวทางที่จะนำไปประยุกต์ใช้ต่อให้เข้ากับข้อมูลที่มีได้

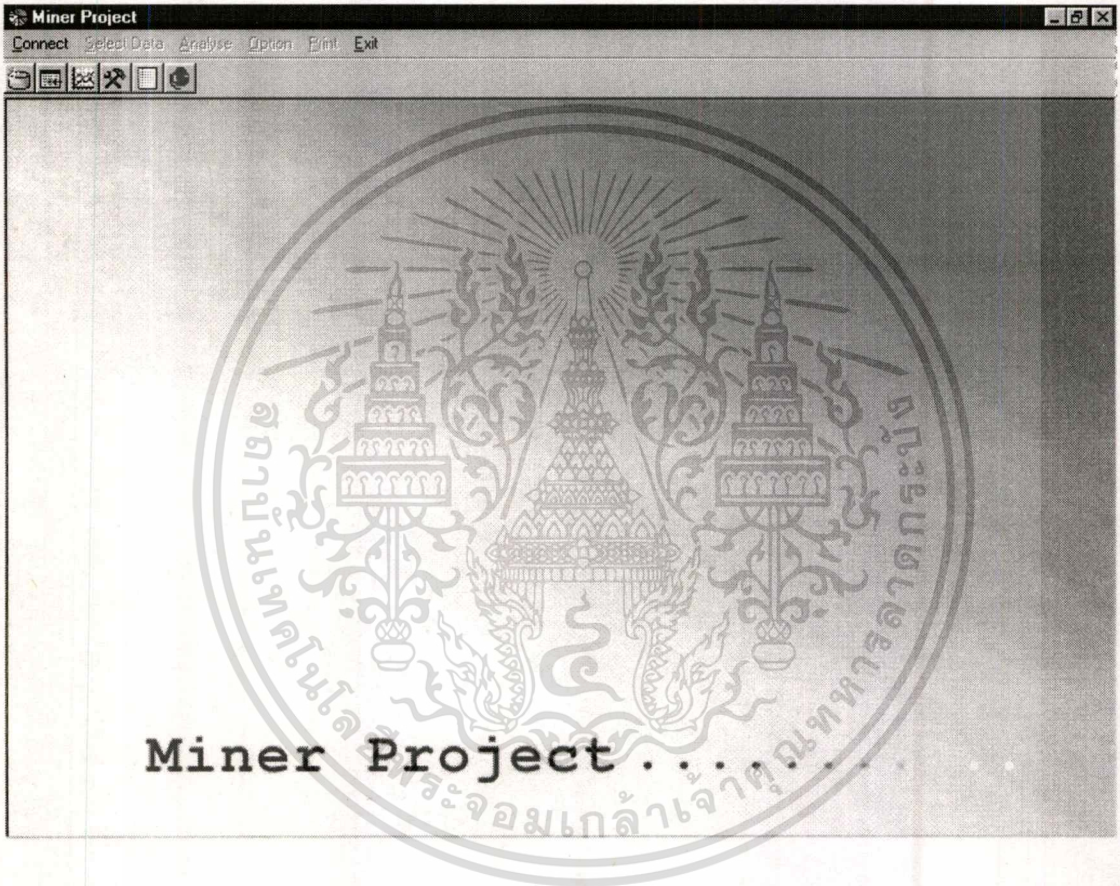


เอกสารอ้างอิง

- [1] Michael, J.A. Berry and Gordon ,S. Linoff , 1997. **Data Mining Techniques For Marketing, Sales, and Customer Support** . John Wiley & Sons.
- [2] Cabena, Hadjinian, Stadler, Verhees, Zanasi. 1998. **Discovering data mining : from concept to implementation**. Prentice Hall.
- [3] Jiawei Han, Micheline Kamber, 2001. **Data Mining : Concepts and Techniques**. Morgan Kaufann.
- [4] Rakesh Agrawal and Ramakrishnam Srikant. 1994. **Fast Algorithms for Mining Association Rules**. IBM Almaden Research Center.
- [5] Grant Bugher. **Market Basket Analysis technique : While paper**. [Online] Available: <http://www.megaputer.com/html/mba.html>.
- [6] Two Crows Corporation. 1999. **Introduction to Data Mining and Knowledge Discovery**. [Online] Available : <http://www.twocrows.com> .

ภาคผนวก ก.

ก. คู่มือการใช้งานโปรแกรม Miner Project



รูปที่ ก.1 หน้าจอโปรแกรม Miner Project

1. การติดต่อฐานข้อมูล

เลือกเมนู Connect โปรแกรมจะแสดงหน้าจอเพื่อทำการติดต่อฐานข้อมูล ซึ่งสามารถติดต่อข้อมูลได้ 2 ลักษณะ คือ

- ติดต่อฐานข้อมูลที่เป็น Microsoft Access

เลือก Microsoft Access แล้วระบุฐานข้อมูลที่ต้องการติดต่อในช่อง File Name จะได้ผลดังรูปที่ ก.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ODBC Logon

MS Access 97 File

File name: C:\Project\pdata.mdb

Use ODBC

DSN: (None)

UID:

Password:

Connect Disconnect Cancel

รูปที่ ก.2 หน้าจอแสดงการติดต่อฐานข้อมูลที่เป็น Microsoft Access

- ติดต่อฐานข้อมูลผ่าน ODBC
เลือก Use ODBC แล้วทำการระบุ DSN ที่ต้องการ รวมถึง User Name และ Password ของผู้ใช้ ดังแสดงในรูปที่ ก.3

ODBC Logon

MS Access 97 File

File name:

Use ODBC

DSN: Project

UID: User

Password: xxxxxx

Connect Disconnect Cancel

รูปที่ ก.3 หน้าจอแสดงการติดต่อฐานข้อมูลผ่าน ODBC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

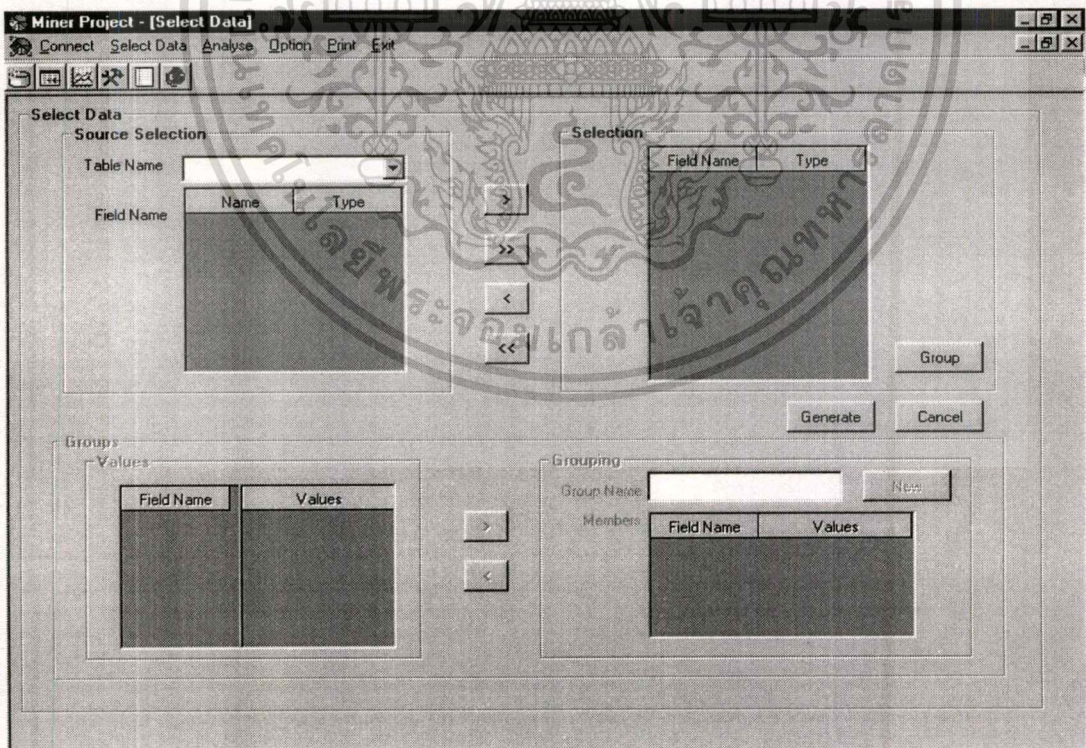
เมื่อกดปุ่ม Connect เพื่อยืนยันกันติดต่อฐานข้อมูล โปรแกรมจะทำการติดต่อฐานข้อมูลที่ระบุไว้ เมื่อสามารถติดต่อได้สำเร็จจะมีข้อความแจ้งว่าได้ติดต่อฐานข้อมูลได้สำเร็จ



รูปที่ ก.4 หน้าจอแจ้งการติดต่อฐานข้อมูลได้สำเร็จ

2. การคัดเลือกข้อมูล

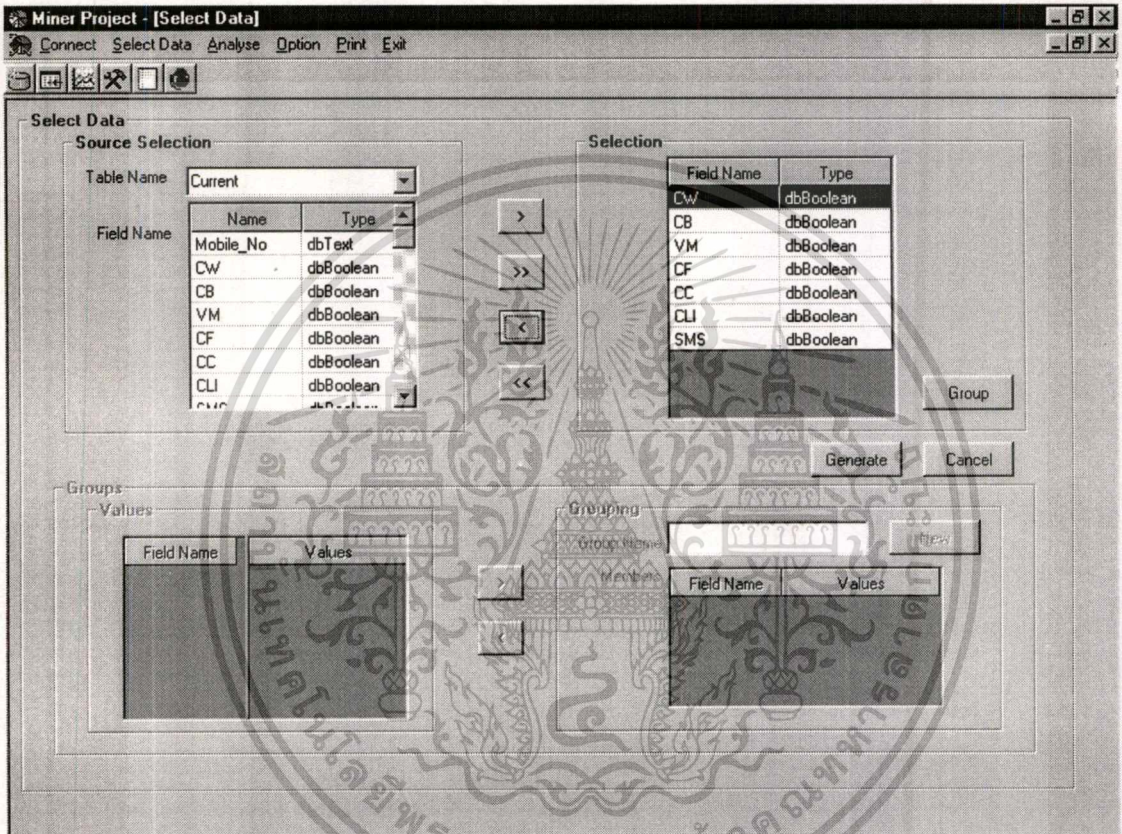
หลังจากที่มีการติดต่อฐานข้อมูลแล้ว ผู้ใช้ต้องทำการเลือกข้อมูลที่ต้องการทำการวิเคราะห์ โดยเลือกที่เมนู Select Data จะปรากฏหน้าจอ ดังรูปที่ ก.5



รูปที่ ก.5 หน้าจอคัดเลือกข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้ใช้จะต้องระบุตารางข้อมูลที่ต้องการ ในช่องของ Table Name โปรแกรมจะทำการแสดงชื่อข้อมูล (Field) ในตารางที่ระบุให้ ผู้ใช้สามารถเลือกเฉพาะชื่อข้อมูลที่ต้องการวิเคราะห์บางรายการ หรือสามารถเลือกทั้งหมดได้ โดยใช้การดับเบิลคลิก หรือ ใช้ Select Button (>) และ Select all Button (>>)

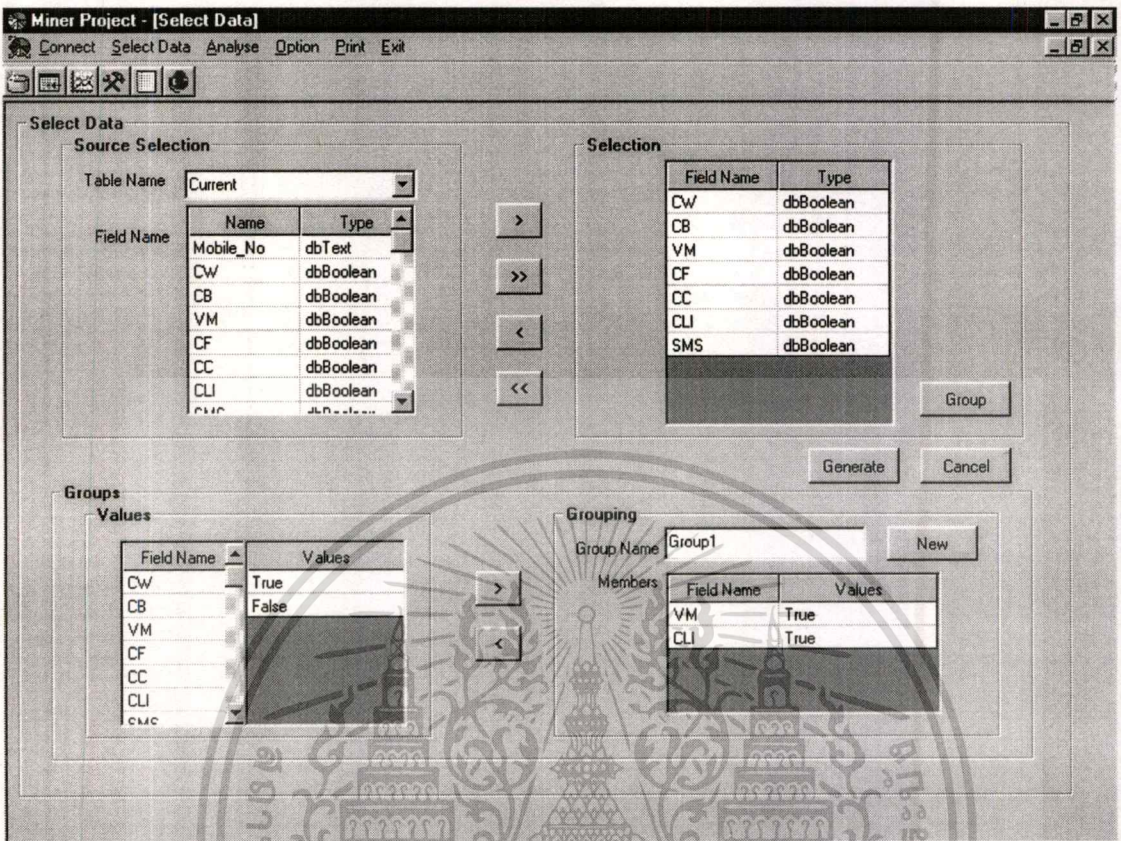


รูปที่ ก.6 เลือกข้อมูลที่ต้องการวิเคราะห์

ในการคัดเลือกข้อมูล ผู้ใช้สามารถจะจัดกลุ่มของข้อมูลที่ได้ถูกคัดเลือกมาแล้วได้ เพื่อให้มีการวิเคราะห์มีประสิทธิภาพมากขึ้น โดยกดปุ่ม Group แล้วทำการจัดกลุ่มข้อมูลที่ต้องการในหน้าจอส่วน Groups ซึ่งในส่วนที่เป็น Values จะเป็นการแสดงว่าข้อมูล(Field) ที่เลือก มีค่าข้อมูลใดบ้าง

ทำการเลือกข้อมูล (Field) และค่าของข้อมูล แล้วกดปุ่ม > เพื่อเป็นการเลือกสมาชิกของกลุ่ม เมื่อได้สมาชิกที่ต้องการ ให้ระบุชื่อของกลุ่ม จากนั้นกดปุ่ม New เพื่อเป็นการยืนยันการสร้างกลุ่มข้อมูล ชื่อของกลุ่มจะไปแสดงอยู่ในส่วนของ Selection

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.7 หน้าจอแสดงการสร้างกลุ่มข้อมูล

กดปุ่ม Generate เพื่อยืนยันการคัดเลือกข้อมูลทั้งหมด หรือปุ่ม Cancel เพื่อยกเลิกการคัดเลือกข้อมูล

3. การกำหนดเงื่อนไข

ในการทำการวิเคราะห์ข้อมูล ผู้ใช้สามารถกำหนดค่าที่ใช้ในการตัดสินใจได้ โดยเลือกที่เมนู Option จะปรากฏหน้าจอ เพื่อให้ทำการกำหนดค่าต่างๆ คือ ค่า Minimum Support ค่า Minimum Confident และค่า Maximum in Itemset สำหรับค่า Minimum Support ผู้ใช้สามารถ เลือกได้ว่า ต้องการให้ค่าที่ใช้ในการวิเคราะห์เป็นค่าความน่าจะเป็นของการเกิดข้อมูล หรือเป็นจำนวนข้อมูลที่เกิดขึ้นจริง ซึ่งจะส่งผลต่อการแสดงค่าในหน้าจอการวิเคราะห์ข้อมูลด้วย

Option

Support

Minimum Support 0.5 (0.0 - 1.0)

Minimum Frequency 20

Confident

Minimum Confident 0.5 (0.0 - 1.0)

Item Set

Maximum Item in set 3 Item(s)

OK Default

รูปที่ ก.8 หน้าจอกำหนดค่าต่างๆ

4. การวิเคราะห์ข้อมูล

เมื่อคัดเลือกข้อมูลและกำหนดค่าต่างๆ แล้ว ผู้ใช้สามารถทำการวิเคราะห์ข้อมูล โดยเลือกเมนู Analyse โปรแกรมจะทำการวิเคราะห์หากกลุ่มของข้อมูลที่มีความสัมพันธ์กัน และกฎความสัมพันธ์ที่สามารถเกิดขึ้นได้ พร้อมด้วยค่า Support และค่า Confident

Miner Project - [Rule Generate]

Connect Select Data Analyse Option Print Exit

Association

	LHS	RHS	Support	Confident
Set 1	CB = Y CLI = Y		0.4582	
	CB = Y	CLI = Y		0.7762
	CLI = Y	CB = Y		0.5667
Set 2	VM = Y CF = N		0.4744	
	VM = Y	CF = N		0.6007
	CF = N	VM = Y		0.8713
Set 3	VM = Y CLI = Y		0.6658	
	VM = Y	CLI = Y		0.8430
	CLI = Y	VM = Y		0.8234
Set 4	VM = Y SMS = Y		0.6334	
	VM = Y	SMS = Y		0.8020
	SMS = Y	VM = Y		0.8515
Set 5	CF = N CLI = Y		0.4690	
	CF = N	CLI = Y		0.8613
	CLI = Y	CF = N		0.5800
Set 6	CF = N SMS = Y		0.4528	
	CF = N	SMS = Y		0.8316
	SMS = Y	CF = N		0.6087
Set 7	CLI = Y SMS = Y		0.6765	
	CLI = Y	SMS = Y		0.8366
	SMS = Y	CLI = Y		0.9094
Total Rules	20		Sort	OK

รูปที่ ก.9 หน้าจอแสดงผลการวิเคราะห์

ถ้าต้องการดูผลการวิเคราะห์เรียงตามค่า Support จากมากไปน้อย ให้กดปุ่ม Sort โปรแกรมจะทำการจัดเรียงผลการวิเคราะห์ให้ใหม่

5. การพิมพ์ผลการวิเคราะห์

เลือกเมนู Print → Print Preview เพื่อทำการดูรูปแบบผลการวิเคราะห์ก่อนพิมพ์ เมื่อต้องการพิมพ์ผลการวิเคราะห์ออกทางเครื่องพิมพ์ให้เลือกเมนู Print → Print

6. การออกจากโปรแกรม

เลือกเมนู Exit เพื่อออกจากโปรแกรม

ประวัติผู้เขียน

นางสาว เนตรนฤณ อุปกาซิม เกิดวันที่ 31 ตุลาคม พ.ศ 2519 สำเร็จการศึกษา วิทยาศาสตร์บัณฑิต จากภาควิชาคณิตศาสตร์ประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ปีการศึกษา 2540 ปัจจุบันทำงานในตำแหน่งโปรแกรมเมอร์ ให้กับบริษัท การบินไทย จำกัด มหาชน



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้