

# การหาความสัมพันธ์ข้อมูลการซื้อขายหลักทรัพย์ โดยใช้ Association Rules

## Using Association Rules to Determine the Relation of Securities trading

โดย

นายมนัส แยมโคกสูง

รหัส 42067083

อาจารย์ที่ปรึกษา

ดร.วรพจน์ กริสุระเดช

วัน เดือน ปี..... 09 ส.ค. 2550  
เลขทะเบียน..... 01754  
เลขเรียกหนังสือ..... กว. ๒164 ก 2543  
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 2 ปีการศึกษา 2543

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง



\*H001754\*

ชื่อหัวข้อ	การหาความสัมพันธ์ข้อมูลการซื้อขายหลักทรัพย์ โดยใช้ Association Rules
นักศึกษา	นายมนัส เข้ม โคนสูง
อาจารย์ที่ปรึกษา	ดร.วรพจน์ กรีสระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2543

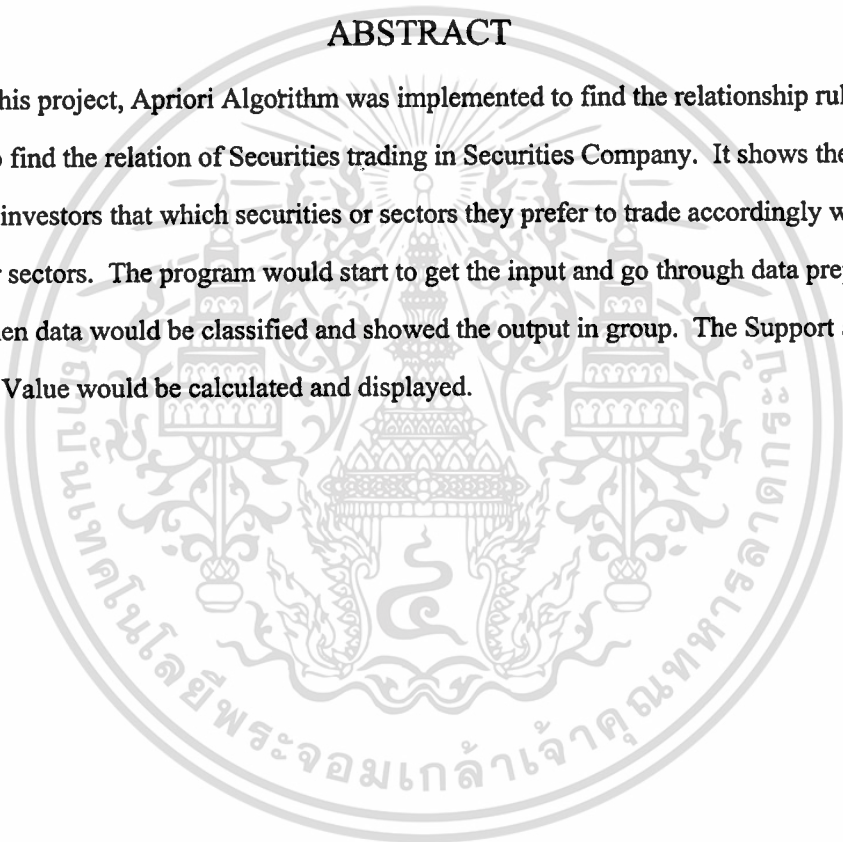
### บทคัดย่อ

ในโปรเจกต์นี้ใช้ Apriori อัลกอริทึมในการค้นหาความสัมพันธ์เพื่อใช้หาความสัมพันธ์ การซื้อขายข้อมูลด้านธุรกิจหลักทรัพย์เพื่อที่จะได้เข้าใจพฤติกรรมการลงทุนของผู้ลงทุนว่าผู้ลงทุน นิยมซื้อขายหลักทรัพย์ตัวใดหรือกลุ่มใดควบคู่กัน โดยโปรแกรมเริ่มการทำงาน โดยการรับข้อมูล เข้ามา จากนั้น โปรแกรมเข้าสู่การเตรียมข้อมูลโดยสามารถคัดเลือกข้อมูลเป็นบางกลุ่มได้หรือแสดงผล ในลักษณะเป็นกลุ่มๆ จากนั้นจะทำการค้นหาค่า Support และ Confidence มาแสดง

**Title** Using Association Rules to determine the Relation of Securities trading  
**Student** Mr.Manat Yamkoksoong  
**Advisor** Dr.Worapoj Keesuradej  
**Level of Study** Master of Science in Information Technology  
**Major** Information Science  
**Academic Year** 2543

## ABSTRACT

In this project, Apriori Algorithm was implemented to find the relationship rule. This rule helps to find the relation of Securities trading in Securities Company. It shows the trading behavior of investors that which securities or sectors they prefer to trade accordingly with which securities or sectors. The program would start to get the input and go through data preparation process. Then data would be classified and showed the output in group. The Support and Confidence Value would be calculated and displayed.



## กิตติกรรมประกาศ

ในการพัฒนาโปรแกรมการหาความสัมพันธ์ข้อมูลการซื้อขายหลักทรัพย์ โดยใช้

Association Rules ต้องอาศัยแหล่งความรู้ต่าง ๆ และคำแนะนำ ความช่วยเหลือ ช่วยดู้น ตลอดจน  
กำลังใจอีกมากมายเกินจะกล่าว แต่ที่ต้องขอบคุณเป็นพิเศษ ก็คือ

1. ท่านบิดามารดา ผู้ซึ่งเป็นทุกอย่างในชีวิต
2. ท่านอาจารย์วรพงษ์ กรีสระเดช ท่านอาจารย์ผู้ซึ่งให้ทั้งโอกาสและคำปรึกษาตลอดจน  
ความ

ช่วยเหลือแก่ข้าพเจ้า

ขอขอบคุณครับ

มนัส แฉ่ม โศกสูง

ผู้จัดทำ

# สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญภาพ.....	VII
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ในการพัฒนาระบบงาน.....	2
1.3 เป้าหมายของการพัฒนาระบบงาน.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 ทฤษฎีที่ใช้ในการพัฒนาระบบงาน.....	2
2. กระบวนการทำงานของ Data Mining.....	3
2.1 Business Objectives Determination.....	3
2.2 Data Preparation.....	3
2.3 การทำ Data Mining.....	4
3. กระบวนการ Association Rule Algorithm.....	6
4. วิธีการ Apriori.....	9
4.1 การค้นพบรายการส่วนใหญ่( Discovery Large Itemsets).....	9
4.2 สัญลักษณ์ที่ใช้.....	9
4.3 Algorithm Apriori.....	10
4.4 การสร้างรายการ Apriori คู่แข่ง.....	11
4.5 การค้นหากฎ.....	11
4.6 ตัวอย่างการหาผลลัพธ์.....	12
5. การออกแบบและพัฒนาระบบงาน.....	15
5.1 เครื่องมือที่ใช้ในการพัฒนาระบบงาน.....	15

## สารบัญ(ต่อ)

	หน้า
5.2 แนวทางการออกแบบระบบ.....	15
5.3 ขั้นตอนทำงานของระบบ.....	17
5.4 สรุปผลการพัฒนาโปรแกรม .....	23
5.5 ข้อเสนอแนะ.....	23
บรรณานุกรม .....	24
ภาคผนวก ก.....	25
ภาคผนวก ข.....	25
ภาคผนวก ค.....	26
ภาคผนวก ง.....	27
ภาคผนวก จ.....	27
ประวัติผู้แต่ง.....	28



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญตาราง

หน้า

บทที่ตารางที่

3.1 แสดงค่าต่าง ๆ ของ Association Rule.....	7
4.1 แสดงสัญลักษณ์ที่ต้องใช้ใน Apriori .....	10



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญภาพ

ภาพที่	หน้า
2.1 แสดงลำดับการทำงานของ Data Mining .....	3
5.1 แสดงขั้นตอนการทำงานของระบบงาน .....	17
5.2 การเปิดใช้งาน Text File .....	17
5.3 แสดงผลการดึงข้อมูลจาก Text File.....	18
5.4 แสดง Dialog เปิดฐานข้อมูล Access.....	18
5.5 แสดง Database Window .....	19
5.6 แสดงการเปิดตารางใน Microsoft Access.....	19
5.7 การใช้คำสั่ง SQL ในการเลือกข้อมูล.....	19
5.8 แสดงการติดต่อผ่านทาง ODBC.....	20
5.9 แสดงหน้าต่างทำการ.....	20
5.10 แสดงข้อมูลที่ทำการเลือกมาแล้ว .....	21
5.12 แสดงข้อมูลที่เลือกแบบ Virtual Group.....	22
5.13 แสดงผลลัพธ์จากการค้นหาความสัมพันธ์.....	22
5.14 การจัดการกลุ่ม.....	23

# บทที่ 1

## บทนำ

การดำเนินธุรกิจในปัจจุบันมีสถานะการแข่งขันที่สูงและรุนแรงมากขึ้น สารสนเทศจึงได้เข้ามามีบทบาทในส่วนช่วยในการดำเนินธุรกิจของบริษัทต่าง ๆ เช่น การประเมินความเสี่ยงของธนาคาร การประเมินสถานการณ์ล่วงหน้าการเก็บภาษีของ กรมสรรพากร การประเมินการต่อประกันภัยของลูกค้าบริษัทประกัน เป็นต้น ซึ่งในช่วงเวลาดังกล่าว วิชาการด้านสารสนเทศก็ได้รับการพัฒนาให้มีศักยภาพเพิ่มมากขึ้น จากเดิมที่เป็นระบบ Transaction Process เมื่อข้อมูลมากขึ้น ก็ได้รับการจัดเก็บเพื่อการเรียกใช้ได้อย่างมีประสิทธิภาพเรียกว่าระบบ Data Warehouse และเมื่อได้ Data Warehouse มา จึงมีแนวคิดที่จะใช้ข้อมูลที่มีดังกล่าวให้เป็นประโยชน์จึงได้รับการพัฒนาเป็นกระบวนการ Data Mining เพื่อที่ช่วยสนับสนุนการตัดสินใจ (Decision Support) ในทางธุรกิจทางหนึ่ง ดังนั้นในบทนี้จะกล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ของการพัฒนาระบบงาน เป้าหมายของการพัฒนาระบบงาน ประโยชน์ที่คาดว่าจะได้รับ ทฤษฎีที่นำมาใช้ในการพัฒนาระบบ และรายละเอียดในบทต่าง ๆ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในการดำเนินธุรกิจหลักทรัพย์นั้น ก็มีการดำเนินการมายาวนานพอสมควร เริ่มจากมีการซื้อขายกันด้วยระบบ Manual จากนั้นได้มีการพัฒนาให้มีการซื้อขายด้วยระบบอิเล็กทรอนิกส์ผ่านทาง Broker ต่าง ๆ มี Tool มากมายได้รับการพัฒนาเพื่อใช้ในการดำเนินธุรกรรมของบริษัท Broker และกับตัวลูกค้าเอง เช่นระบบการทำนายราคาหุ้น ระบบบริหารความเสี่ยงของ Port ลูกค้า ระบบการบริหาร Port ของบริษัท เป็นต้น ซึ่งระบบดังกล่าวเป็นระบบทั่วไปซึ่งเน้นไปที่การนำข้อมูลระดับราคาของหุ้นแต่ละวันและข่าวสารที่มีผลกระทบต่อบริษัทมาประมวลผล แต่ในแง่ของการบริหารองค์กรเอง ในระบบเหล่านี้ไม่มี ในเวลาที่ผ่านมาลูกค้ามักจะได้มาจากฝ่ายการตลาด (Marketing) เมื่อฝ่ายการตลาดลาออกจากองค์กรไป ก็จะทำลูกค้าไปด้วยทำให้ฐานลูกค้าของบริษัทหายไป ซึ่งเป็นที่ตระหนักของผู้บริหารในองค์กรเพื่อที่จะรักษาฐานลูกค้าให้ได้ดีที่สุด โดยการที่จะให้เจ้าหน้าที่ฝ่ายการตลาดสามารถอำนวยความสะดวกแทนกันได้ทำให้ลูกค้าไม่รู้สึกรังเกียจ และไม่เห็นความแตกต่างของการให้บริการ

## 1.2 วัตถุประสงค์ในการพัฒนาระบบงาน

- เพื่อวิเคราะห์ถึงความสัมพันธ์ แนวโน้มการซื้อขายลูกค้าแต่ละบุคคลหรือรายกลุ่ม
- เพื่อให้เกิดความเข้าใจความต้องการของลูกค้าได้รวดเร็วยิ่งขึ้น

## 1.3 เป้าหมายของการพัฒนาระบบงาน

- โปรแกรมสามารถค้นหาความสัมพันธ์ของข้อมูลการซื้อขายพร้อมแสดงผลได้
- สามารถดึงข้อมูลได้หลายแหล่ง
- มีความยืดหยุ่นในการเตรียมข้อมูลเพื่อให้สามารถทดลองจากข้อมูลต่างลักษณะได้

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

ผู้ใช้และองค์กรจะได้ประโยชน์เพื่อในไปปรับเปลี่ยนกลยุทธ์ในการที่จะให้ความสะดวกกับลูกค้า นอกจากนี้แล้วโปรแกรมยังสามารถนำไปประยุกต์ใช้ในงานอื่นที่ลักษณะความต้องการเหมือนกันได้ด้วย

## 1.5 ทฤษฎีที่ใช้ในการพัฒนาระบบงาน

- ขั้นตอนตามขบวนการทำ Data Mining เพื่อให้ได้ลักษณะการทำงานที่ถูกต้อง
- ในการที่จะทำ Association Rule Algorithms โดยนำเอา Apriori Algorithm มาประยุกต์

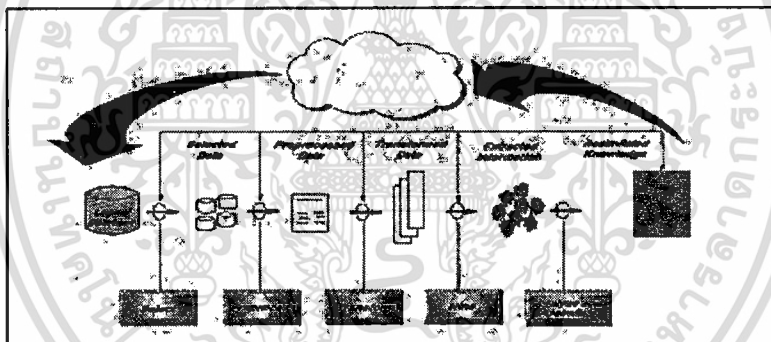
## บทที่ 2

### กระบวนการทำงานของ Data Mining

Data Mining คือกรรมวิธีดึงสิ่งที่ไม่รู้มาก่อน, สิ่งที่เป็นเหตุผล, ข่าวสาร ออกจากฐานข้อมูลขนาดใหญ่เพื่อเป็นข่าวสารที่ใช้ในการตัดสินใจในธุรกิจที่สำคัญ โดยทั่วไป ลำดับการทำงานของ Data Mining สามารถแบ่งได้ดังนี้

#### 2.1 Business Objectives Determination

เป็นการกำหนดปัญหาหรือความต้องการ สิ่งนี้จะเป็นหัวใจของการทำ Data Mining เพราะถ้ากำหนดปัญหาได้ชัดเจนแล้ว กระบวนการ Mining จะง่ายขึ้น



ภาพที่ 2.1 แสดงลำดับการทำงานของ Data Mining

(ที่มา : Discovering Data Mining from concept to implementation)

#### 2.2 Data Preparation

เป็นการเตรียมข้อมูลเป็นช่วงเวลาที่ใช้เวลาค่อนข้างมาก เพราะถ้าข้อมูลที่เตรียมมาไม่เหมาะสม อาจจะต้องมีการเตรียมข้อมูลใหม่อีกครั้ง ซึ่งขั้นตอนสามารถแบ่งเป็น 2 ขั้นตอนย่อยได้ดังนี้

1) **Data Selection** เป็นการเลือกสรรข้อมูลที่ต้องการจะนำมาใช้งาน รายละเอียดของข้อมูลที่นำมาทำ Mining ลักษณะข้อมูลมี 3 ประเภท

1.1 **Categorical** ค่าที่เป็นได้แน่นอนและต่างชนิดกัน ซึ่งมีค่าแบ่งเป็น 2 subtype คือ ordinal และ Nominal

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- **Nominal** เช่น สถานะการแต่งงาน (Marital Status) จะมีค่า Single, Married, divorced เป็นต้น

- **Ordinal** เป็นตัวแปรแบบมีลำดับค่าที่เป็นไปได้ เช่น Customer Credit rating จะมีค่าเป็น good, regular, poor

**1.2 Quantitative** มีความแตกต่างที่เป็นไปได้ 2 subtype คือ Continuous และ Discrete

- **Continuous** ได้แก่ เลขจำนวนจริง(Real Number) เช่น 0.1,0.2,0.3,0.5

- **Discrete** ได้แก่ เลขจำนวนเต็ม (Integer) เช่น 1,2,3,4,5

2) **Data Preprocessing** ทำการเรียนรู้ชนิดของข้อมูลเพื่อเป็นการวิเคราะห์ในการกำหนดชนิดของข้อมูลที่จะมาทำ เพื่อให้มั่นใจว่าข้อมูลที่เลือกมานั้นถูกต้องและเหมาะสม

3) **Mining Data Transformation** เป็นการนำข้อมูลมาแปลงให้อยู่ใน Model ที่จะทำการวิเคราะห์ เพื่อเลือก Algorithms ที่เหมาะสม รูปแบบข้อมูลที่ต้องการ

## 2.3 การทำData Mining

เป็นการนำข้อมูลจากหัวข้อ 2.3 ที่เลือกมารวมกับ Algorithms ตามความเหมาะสม ในขั้นตอนนี้จะเร็วมากและ ทำโดยอัตโนมัติ โดยตัว Algorithms สามารถแบ่งได้ 4 หมวดดังนี้

1. **Database Segmentation** เป็นการแยกข้อมูลที่เหมือนกันใน Database ออกเป็นส่วน ๆ หรือเป็นกลุ่ม ๆ โดยแบ่งตามคุณสมบัติที่เหมือนกัน ซึ่งผลลัพธ์ จะได้เป็นกลุ่มข้อมูล
2. **Predictive Modeling** เป็นขบวนการสร้างการทำนาย ซึ่งแบ่งได้ 2 ลักษณะ

**2.1 Classification** เราจะพิจารณา 2 Data Mining เฉพาะการจัดหมู่(Classification) คือแบบ Tree ซึ่งอาจจะใช้ กระบวนการทางสถิติ เช่น C4.5 Methods และ Neural เพราะ ทั้งคู่ง่ายต่อการเรียนรู้ นอกจากนี้ยังมี rough sets, Database Oriented ฯลฯ การสร้างรูปแบบการจัดหมู่ Classification Model) โดย สร้างจากกลุ่มรายการข้อมูลที่เรียกว่า Training Set โดย Record ใน Training Set ต้องเป็นข้อมูลที่ผู้วิเคราะห์ต้องการกำหนดไว้ก่อน

**Tree Induction** เป็นเทคนิคการสร้างรูปแบบการพยากรณ์ในรูปของ Decision Tree และ Binary Decision Tree มีข้อเสียหลายประการ โดยเฉพาะการทำงานกับค่าที่ต่อเนื่อง เช่นรายได้ หรือ ราคา ซึ่งค่าของข้อมูลอาจเป็นกลุ่มใน Range ดังนั้นการจัดกลุ่มด้วย Decision Tree จะไม่สามารถจัด Patterns ได้ อีกปัญหาหนึ่งคือ การจัดการกับสิ่งที่ไม่มีค่า(Miss Value)ว่าจะจัดการอย่างไร นอกจากนี้ยังทำให้เกิดเหตุ

การที่เรียกว่า Fragmentation คือเกิดเมื่อระดับของ Tree ซับซ้อนมากขึ้น ข้อมูลที่จัดกลุ่มมาก็จะน้อยตามไปด้วยซึ่งจะทำให้ข้อมูลมีขนาดเล็กมาก ขากแก่การจัดความถูกต้อง

Neural เป็น Techniques การจำลอง ของระบบประสาท ของมนุษย์ แบ่งการทำงานเป็น 2 ส่วนคือ

**Training Mode** เป็น Mode การให้ Neural ทำการเรียนรู้ ซึ่งแบ่งเป็น 2 ประเภท คือ Supervised Learning คือ มีการกำหนด Output ที่ต้องการและ Unsupervised Learning คือ ไม่ระบุ Output ที่ต้องการ

**Query (Interrogation) Mode** เป็นการนำ Neural ไปใช้งาน Neural นำเสนอด้วยลักษณะของ Node และสายเชื่อมระหว่าง Node ซึ่งเป็นพื้นฐานของ Neural Network ซึ่งเป็นที่รวมการเชื่อมต่อของ Hidden Node กับส่วนรับข้อมูล(Input)

2.2 Value Prediction ใช้ในการประมาณการ ตัวเลขที่เป็นค่าถัดไป ซึ่งสัมพันธ์กับ

Database Record เช่น ผู้ค้าปลีกรถยนต์ต้องการทราบช่วงระยะเวลาของลูกค้าใหม่ โดย Data Mining จะหาค่าจากประวัติช่วงเวลาของลูกค้า หรือ การประมาณค่าการจัดเก็บรายได้ของกรมสรรพากร เป็นต้น

3. **Link Analysis** มีจุดมุ่งหมายเพื่อหาสิ่งที่เชื่อมโยงระหว่าง Record หรือ กลุ่ม Record ใน Database ซึ่งเราอาจเรียกบ่อย ๆ ว่า “Associations” Classic Application ที่ใช้หลักการเป็นการค้นพบความสัมพันธ์ระหว่างสินค้าหรือบริการ ว่าลูกค้ามักชอบซื้อสินค้าเหล่านี้ด้วยกันหรือลำดับการซื้อผลิตภัณฑ์นี้หลายครั้ง นอกจากนี้ ยังใช้ Link Analysis ในธุรกิจอื่นเช่น ใช้ในการวิเคราะห์ เป้าหมายทางการตลาดและ การเคลื่อนไหวของราคาสินค้า

4. **Deviation Detection** เป็น Operation ใหม่ที่เพิ่งจะเริ่มเป็นที่รู้จัก บางครั้งหลังจากทำการ Data mining แล้ว ผลที่แสดงอยู่ไกลจากการคาดหมายหรือ ไกลจากปกติ deviation detection จะใช้ Statistic และ visualization techniques บอกความสิ่งที่ต่างออกไปของข้อมูล ซึ่งอาจใช้ Computer ช่วยในการ Summarization และ แสดงเป็นผล Graphic ซึ่งเป็นการง่ายต่อการตรวจพบ

### บทที่ 3

#### กระบวนการ Association Rule Algorithms

กระบวนการ Association Rule คือกฎซึ่งค้นหาความสัมพันธ์ที่แฝงอยู่ระหว่างกลุ่มของวัตถุ เป็นต้นว่า “อะไรมีการกระทำด้วยกัน” หรือ “มีอะไรที่ซ่อนอีกสิ่งหนึ่ง” ในฐานข้อมูล โดยในกลุ่มของ Transaction เมื่อแต่ละ Transaction คือกลุ่มรายการ(Items) ซึ่ง Association Rule จะแสดงนิพจน์ในรูปแบบ X, Y เมื่อ X, Y คือกลุ่มรายการ(Items) ซึ่งความหมายในแต่ละกฎคือ Transaction ในฐานข้อมูล ซึ่งมีค่า X แล้วมีค่า Y ด้วย การวิเคราะห์นี้บางครั้งเรียกว่า “Market Basket Analysis”

ผู้ที่ทำธุรกิจหรือผู้วิเคราะห์ สามารถใช้ Market Basket Analysis ในการวางแผน

- การใช้คูปองและการลดราคา คงไม่เป็นการดีแน่ ถ้าจะลดราคาเบียร์และผ้าอ้อมในเวลาเดียวกัน ถ้าเรามีแนวโน้มที่จะซื้อของเหล่านี้ด้วยกัน ควรจะลดราคาสินค้าตัวหนึ่งเพื่อดึงยอดขายสินค้าอีกตัวหนึ่ง
- การวางตำแหน่งสินค้า การวางตำแหน่งสินค้าที่มีความสัมพันธ์กันไว้ใกล้กัน ย่อมทำให้การหยิบสินค้าง่ายขึ้น

รูปแบบของกฎจะอยู่ในลักษณะ “If X Then Y” หรือ “IF Condition1 THEN Condition2” ซึ่งทั้ง X และ Y เกิดขึ้นพร้อมกัน Transaction เดียวกันหรือเรียกได้ว่าเมื่อได้ก็ตามที่เกิดเหตุการณ์ X หรือ Condition1 แล้วจะเกิดเหตุการณ์ Y หรือเหตุการณ์ที่ 2 ขึ้นด้วย โดยเรียก X ว่า “Left-hand side” หรือ “Antecedent” และ Y เรียกว่า “Right-hand side” หรือ “Antecedent” ผลที่ได้จากกฎจะมีค่าสำคัญอยู่ 2 ตัวคือ “Confidence factor” และ “Support factor”

ค่า Confidence และ Support หรือบางผลิตภัณฑ์เรียกว่า “predictability” และ “prevalence” แทน ซึ่งสามารถยกตัวอย่างการหาค่าต่าง ๆ พิจารณาจากตัวอย่างเบียร์และผ้าอ้อม ดังนี้

- รายการทั้งหมด 500,000 รายการ
- เป็นรายการของผ้าอ้อม 20,000 รายการ (เป็น 4% ของรายการทั้งหมด)
- เป็นรายการของเบียร์ 30,000 รายการ (เป็น 6% ของรายการทั้งหมด)
- เป็นรายการที่ทั้งมีผ้าอ้อมและเบียร์ 10,000 รายการ (เป็น 2% ของรายการทั้งหมด)

ค่า Support (prevalence) เป็นจำนวนเหตุการณ์ที่เกิดขึ้นร่วมกัน ซึ่งคิดเป็นเปอร์เซ็นต์ต่อรายการที่ดำเนินงานทั้งหมดในตัวอย่างนี้ คือ 2% (10,000/500,000)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า **Confidence** (predictability) เป็นจำนวนรายการที่เจาะจงที่ขึ้นกับอีกรายการหนึ่งว่าเป็นเท่าไร ตัวอย่างก็คือรายการที่มีทั้งผ้าอ้อมและเบียร์มีทั้งหมด 10,000 รายการเมื่อเทียบกับรายการผ้าอ้อมที่มีรายการ 20,000 รายการแล้วคิดเป็น 50 %

กฎในทางตรงกันข้าม(Reverse Rule) สามารถเขียนได้คือ “เมื่อคนซื้อเบียร์แล้วเขาทั้งหลายจะซื้อผ้าอ้อมด้วย” คิดเป็น  $1/3(10,000/30,000)$  คิดเป็น 33.33 %

\*หมายเหตุ กฎทั้ง 2 ข้อมีค่า Support เหมือนกัน คือ 2 %

ค่า 4% และ 6% ซึ่งเป็นสัดส่วนของผ้าอ้อมและเบียร์ เมื่อเปรียบเทียบกับรายงานทั้งหมด เรียกอีกอย่างว่า Expected confidence

ค่า **Lift** เป็นค่าที่แสดงความน่าเชื่อถือของผลกระทบที่เกิดขึ้น ถ้าเราพบว่า ลูกค้าซื้อผ้าอ้อมและซื้อเบียร์ด้วยมีค่าเป็น 8 % แล้วผลกระทบของการซื้อผ้าอ้อมบนการซื้อเบียร์จะมีความถูกต้องน้อยถ้า Expected confidence เท่ากับ 6 ค่า confidence มีค่า 50 % และค่า Lift มากกว่า 8 แล้วผลระหว่าง ผ้าอ้อมและเบียร์ จะมีความน่าเชื่อถือมาก

	Left-Hand side		Right-hand side	Expected Conf(%)	Confidence (%)	Lift Ratio	Support (%)
1	Diapers	>	Beer	6.00	50.00	8.33	2.00
2	Beer	>	Diapers	4.00	33.33	8.33	2.00
3	Diapers	>	Wipes	2.00	40.00	20.00	1.60
4	Wipes	>	Diapers	4.00	80.00	20.00	1.60
5	Wipes	>	Beer	6.00	2.20	0.37	0.04
6	Beer	>	Wipes	2.00	0.73	0.37	0.04
7	Diapers and Wipes	>	Beer	6.00	2.50	0.42	0.04
8	Diapers and Beer	>	Wipes	2.00	2.00	1.00	0.04
9	Wipers and Beer	>	Diapers	4.00	90.91	22.73	0.04
10	Diapers	>	Wipes and Beer	0.04	1.00	22.73	0.04
11	Wipers	>	Diapers and Beer	0.04	1.00	22.73	0.04
12	Beer	>	Diapers and Wipes	1.60	0.67	0.42	0.04

ตารางที่ 3.1 แสดงค่าต่าง ๆ ของ Association Rule

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า Lift ที่ดีที่สุดจะพบใน กฎข้อที่ 9 ค่า Lift = 22 มีความหมายว่า ลูกค้าน่าจะซื้อน้ำยาทำความสะอาดและเบียร์ 22 ครั้ง มีความเป็นไปได้มากที่จะเขาซื้อผ้าอ้อมด้วยมากกว่าคนที่ไม่ซื้อน้ำยาทำความสะอาดและเบียร์ ถ้าค่า Lift มีค่าเป็นลบหรือมีค่าน้อยกว่า 1 เช่นรายการที่ 5,6,7 และ 12 โดย 7และ 12 มีค่า Lift 0.42 เราสามารถอธิบายได้ว่า ลูกค้าน่าจะทำความสะอาดและผ้าอ้อมมีความเป็นไปได้ น้อยมากที่เขาจะซื้อเบียร์ ถ้าค่า Lift ของกฎและค่า Lift ของ กฎที่ตรงข้าม(Reverse Rule) มีค่า เหมือนกันแล้ว เป็นการยืนยันได้ว่าค่า Lift ถูกคำนวณในลักษณะที่เหมือนกัน สิ่งหนึ่งที่น่าสนใจ มากที่สุดในกฎคือถ้าค่า Confidence มีค่าสูงมาก ๆ หรือต่ำมาก ๆ บางครั้งถือเป็นกฎที่ผิดปกติ เช่น ค่า Confidence เป็น 100 % จากกฎ “เมื่อไรก็ตามที่ลูกค้าซื้ออาหารสัตว์เลี้ยงแล้วเขาจะซื้ออุปกรณ์ ของสัตว์เลี้ยงด้วย” ซึ่งไม่น่าจะเป็นไปได้ ซึ่งอาจจะเป็นไปได้ว่าข้อมูลที่น่ามาอาจจะ เป็นข้อมูลเพียงวัน เดียว ซึ่งเป็นวันที่มีการจัดของก้านัลในวันนั้น หรืออาจจะเมื่อมีการซื้ออาหารสัตว์เลี้ยงก็มีการแถม ของให้ด้วย เป็นต้น

ในกฎข้ออื่น ๆ อาจเป็นของที่ต้องใช้เป็นประจำตัวอย่างเช่นนมหรือไข่ ซึ่งจะทำให้แสดง ให้กฎเป็นจำนวนมาก ซึ่งเป็นไปได้ที่เราจะต้องยกเว้นรายการเหล่านี้ให้ตอนใช้ Association Tool โดยทั่วไปการวิเคราะห์ค่า Lift ที่สูงมาก ๆ หรือ ต่ำมาก ๆ ซึ่งไม่ได้หมายถึงว่าจะมีรายการปรากฏอยู่ ในรายการดำเนินการมาก ๆ กฎที่มีค่า Support ต่ำอาจจะข้อผิดพลาดเล็ก ๆ น้อย ๆ

## บทที่ 4

### วิธีการ Apriori

วิธีการ Apriori เป็นวิธีการสำหรับค้นหาความสัมพันธ์โดยการค้นหาจะค้นหาจากกลุ่มข้อมูลทั้งหมด โดยมีค่าสนับสนุน(Support) มากกว่าค่าสนับสนุนต่ำสุด(Minimum Support) ซึ่งค่าสนับสนุน(Support) คือจำนวนข้อมูลที่มีอยู่ในกลุ่มรายการ(Itemset) เมื่อนำกลุ่มรายการ(Itemset) มารวมกับค่าสนับสนุนสามารถเรียกว่า กลุ่มรายการส่วนใหญ่(Large Itemset) คือกลุ่มข้อมูลที่มีอยู่ที่มีความสัมพันธ์กัน(มีค่าอย่างน้อยเท่ากับค่าสนับสนุนที่น้อยที่สุด)

เมื่อเราได้กลุ่มรายการส่วนใหญ่มาแล้ว จะสามารถนำไปสร้างกฎ ซึ่งแนวคิดโดยทั่วไป ถ้าพูดว่า ถ้า ABCD และ AB คือ ข้อมูลรายการส่วนใหญ่แล้ว เราสามารถพิจารณาได้ว่า ถ้ากฎ  $AB \Rightarrow CD$  โดยคำนวณจากสัดส่วนของ ค่าความเชื่อมั่น(Coherence) = ค่าสนับสนุน (ABCD) / ค่าสนับสนุน (AB) ซึ่งถ้าค่าความเชื่อมั่น  $\geq$  ค่าความเชื่อมั่นต่ำสุด(Minimum Coherence)

#### 4.1 การค้นพบรายการส่วนใหญ่(Discovery Large Itemsets)

การค้นหาข้อมูลขนาดใหญ่จะมีการดำเนินการกับข้อมูลหลายครั้ง โดยเริ่มจากการนับจำนวนรายการ(ค่าสนับสนุน) ซึ่งมีมากกว่าค่าสนับสนุนต่ำสุด ในลำดับต่อมาเราจะเริ่มค้นหาข้อมูลจากรายการครั้งที่แล้ว ซึ่งกลุ่มที่ได้มาใหม่นั้นจะเป็นกลุ่มรายการกลุ่มใหม่ ซึ่งเราเรียกว่า กลุ่มรายการคู่แข่ง (Candidate Itemset) และจะทำการนับค่าสนับสนุนกลุ่มรายการใหม่ ซึ่งเมื่อจบขั้นตอนนั้นเราจะพิจารณา รายการในกลุ่มนี้กับรายการการดำเนินงาน(Transaction) ซึ่งเราจะทำอย่างนี้ไปเรื่อยไปจนกว่าจะหากกลุ่มข้อมูลใหม่ไม่พบ

#### 4.2 สัญลักษณ์ที่ใช้

เราสมมุติว่าในแต่ละรายการ(Transaction) มีการเรียงลำดับตามพจนานุกรมซึ่งเราจะเก็บอยู่ในฐานข้อมูล D โดยเก็บในลักษณะ  $\langle TID, Item \rangle$  คือ รหัสรายการ(TransactionID) กับ รายการข้อมูล เราเรียกจำนวนรายการ(Item)ในกลุ่มรายการ(itemset) ว่าขนาด(size) ซึ่งแทนด้วย  $K(K\text{-itemset})$  คือขนาด  $K$  ข้อมูล เราใช้สัญลักษณ์  $C[1], C[2], \dots, C[K]$  แสดงถึงขนาดของ  $K$  ของกลุ่มรายการ  $C$  โดยที่  $C[1] < C[2] < \dots < C[K]$

k-itemset	An itemset having k items.
$L_k$	Set of large k-itemsets(those with minimum support). Each member of this set has two fields : i) itemset and ii)support count.
$C_k$	Set of candidate k-itemsets (potentially large itemsets). Each member of this set has two fields : i) itemset and ii)support count.

ตารางที่ 4.1 แสดงสัญลักษณ์ที่ต้องใช้ใน Apriori

### 4.3 Algorithm Apriori

สำหรับการทำงานของวิธีการ เริ่มต้นจากการนับรายการ เพื่อพิจารณาจำนวนข้อมูลเพื่อ กำหนดเป็นกลุ่มใหญ่ลำดับที่ 1 (large 1-itemsets) ในลำดับต่อมาจะทำการผ่านข้อมูลขนาด K ค่า ซึ่ง แบ่งเป็น 2 ส่วน โดยส่วนแรกคือกลุ่มข้อมูลรายการใหญ่  $L_{k-1}$  ซึ่งจะนำไปสร้างรายการคู่แข่ง (Candidate Itemset)  $C_k$  โดยฟังก์ชัน Apriori-Gen ที่จะกล่าวในลำดับถัดไปและค่าในส่วนที่สองคือ ค่าที่นับจากฐานข้อมูลเป็นค่าการสนับสนุน(support)

- 1)  $L_1 = \{\text{large 1-itemsets}\};$
- 2) **For** ( $k = 2; L_{k-1} \neq \emptyset; k++$ ) **do begin**
- 3)  $C_k = \text{apriori-get}(L_{k-1});$  // New candidate
- 4) **forall** transaction  $t \in D$  **do begin**
- 5)  $C_t = \text{subset}(C_k, t);$  //Candidates contained in t
- 6) **forall** candidates  $c \in C_t$  **do**
- 7)  $c.\text{count}++;$
- 8) **end**
- 9)  $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$
- 10) **end**
- 11)  $\text{Answer} = \bigcup_k L_k$

#### 4.4 การสร้างรายการ Apriori คู่แข่ง

ฟังก์ชัน apriori-gen จะทำรายการ  $L_{k-1}$ , ซึ่งฟังก์ชันจะส่งค่ากลับเป็นค่า Superset ของรายการ  $K$  ทั้งหมด โดยในลำดับแรกจะทำการเชื่อม (Join) รายการ  $L_{k-1}$  กับ รายการ  $L_{k-1}$  โดยในลำดับ

- 1) insert into  $C_k$
- 2) select  $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-1}$
- 3) from  $L_{k-1} p, L_{k-1} q$
- 4) where  $p.item_1 = p.item_2 = \dots = p.item_{k-2} = q.item_{k-2}, p.item_{k-1} < q.item_{k-1}$ ;

จากนั้นจะมาถึงการ prune ซึ่งเป็นการลบทุก itemsets  $c \in C_k$  บาง (k-1)-subset ของ  $C$  ที่ไม่มีใน  $L_{k-1}$  ซึ่งมีวิธีการ ดังนี้:

- 1) forall itemsets  $c \in C_k$  do
- 2) forall (k-1)-subsets  $s$  of  $c$  do
- 3) if ( $s \notin L_{k-1}$ ) then
- 4) delete  $c$  from  $C_k$ ;

ตัวอย่าง ให้  $L_3$  คือ  $\{\{1\ 2\ 3\}, \{1\ 2\ 4\}, \{1\ 3\ 4\}, \{1\ 3\ 5\}, \{2\ 3\ 4\}\}$  หลังจากผ่านขั้นตอนการ Join จะได้  $C_4$  คือ  $\{\{1\ 2\ 3\ 4\}, \{1\ 3\ 4\ 5\}\}$  ขบวนการ prune จะทำการลบ itemset  $\{1\ 3\ 4\ 5\}$  เพราะ itemset  $\{1\ 4\ 5\}$  ไม่มีค่าอยู่ใน  $L_3$

#### 4.5 การค้นหากฎ (Rule)

การค้นหาความสัมพันธ์มีความสัมพันธ์ มีหลักการว่า การสร้างกฎสำหรับกลุ่มข้อมูล  $I$  เราจะค้นหา Subset ของ  $I$  ที่มีค่าไม่ต่ำทั้งหมด ซึ่งทุก ๆ Subset  $A$ . ผลลัพธ์ของกฎ  $a$  จะอยู่ในรูปแบบ  $a \Rightarrow (I-a)$  ถ้าสัดส่วนระหว่างค่าสนับสนุน  $I$  ต่อ ค่าสนับสนุน  $A$  ซึ่งค่าน้อยที่สุดเท่ากับค่าความเชื่อมั่นต่ำสุด (Minimum Confidence) จากนั้นเราจะพิจารณา Subset ของ  $I$  ทั้งหมดเพื่อสร้างกฎร่วมหลายตัวในภายหลัง โดยวิธีการดังนี้

**Procedure getrules( $I_k$ : large k-itemset,  $a_m$ : large m-itemset)**

- 1)  $A = \{(m-1)\text{-itemsets } a_{m-1} \mid a_{m-1} \subset a_m\}$ ;
- 2) Forall  $a_{m-1} \in A$  do begin
- 3)  $Conf = \text{support}(I_k) / \text{support}(a_{m-1})$
- 4) if ( $conf \geq \text{miniconf}$ ) then begin
- 5) output the rule  $a_{m-1} \Rightarrow (I_k - a_{m-1})$ , with confidence = conf and support = support( $I_k$ );
- 6) if ( $m-1 > 1$ ) then

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 7) **call** getrules( $l_k, a_{m-1}$ ); // to generate rules with subsets of  $a_{m-1}$  as the antecedents
- 8) **end**
- 9) **end**

#### 4.6 ตัวอย่างการหาผลลัพธ์

กำหนดให้ ค่าสนับสนุน(Minimum Support) = 2

1. Data Input คือ

TID	ITEMS
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

2. Large 1-itemsets ดังนี้ โดย ค่า Support เป็นการนับ ค่าจากข้อมูลData Input

ITEMSET	SUPPORT
{1}	2
{2}	3
{3}	3
{5}	3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. หลังจากผ่าน Function Apriori-gen (L1) และ จะได้ Candidate 2-itemsets ดังนี้

ITEMSET
{1 2}
{1 3}
{1 5}
{2 3}
{2 5}
{3 5}

4. ทำการ หา Large 2-itemsets ดังนี้ โดย ค่า Support เป็นการนับ ค่าจาก Data Input โดยค่าที่ได้จะต้องมีค่า  $\text{Count}(\text{Support}) \geq 2$  ซึ่งจะได้

ITEMSET	SUPPORT
{1 3}	2
{2 3}	2
{2 5}	3
{3 5}	2

\* ค่า {1 2} ได้ถูกกระบวนการ Prune ออก เนื่องจากมี set อยู่จริง แต่ ค่า Support มีเพียง 1 เท่านั้น

5. หลังจากผ่าน Function Apriori-gen (L2) และ จะได้ Candidate 3-itemsets ดังนี้

ITEMSET
{2 3 5}

6. ทำการ หา Large 3-itemsets ดังนี้ โดย ค่า Support เป็นการนับ ค่าจาก Data Input

ITEMSET	SUPPORT
{2 3 5}	2



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### การออกแบบและพัฒนาระบบงาน

ในส่วนที่ผ่านมาได้พูดถึง กระบวนการทาง Data Mining กระบวนการ Association Rule และ Apriori Algorithm มาแล้ว ในบทนี้จะพูดถึงการออกแบบและพัฒนาระบบงานการหาความสัมพันธ์ข้อมูลการซื้อขายหลักทรัพย์ โดยใช้ Association Rule ในส่วนต่าง ๆ ดังต่อไปนี้

#### 5.1 เครื่องมือที่ใช้ในการพัฒนาระบบงาน

- Microsoft Word และ Microsoft Excel ใช้ในการผลิตเอกสารประกอบโครงการ
- Visio และ Paint สำหรับการทำไคอะแกรม (Flow Chart) และ เดิมแต่งรูป
- Microsoft Visual Basic 6.0 และ Service Pack 4 สำหรับการพัฒนาโปรแกรมในส่วน User Interface, การติดกับฐานข้อมูล และ การทำ Data Preparation
- Microsoft Visual C++ 6.0 สำหรับพัฒนาส่วนของการค้นหาความสัมพันธ์ และ การสร้างกฎ

◆ ทั้ง Visual Basic และ Visual C++ ต้องเป็นชุด Microsoft Visual Studio

#### 5.2 แนวทางการออกแบบ

สำหรับแนวทางการออกแบบ ได้มุ่งหวังต้องการให้ระบบงานมีความซับซ้อนที่น้อย มีความยืดหยุ่นในการนำข้อมูลเข้า การจัดกลุ่มข้อมูล ซึ่งได้เน้นกระบวนการในส่วน Data Preparation เพราะ กระบวนการ Mining ที่จะได้ผลต้องมีการเตรียมข้อมูลที่ดี ดังจะได้นำเสนอจากนี้

##### 5.2.1 การนำข้อมูลเข้า

โปรแกรมสามารถ รับข้อมูลเข้าได้ 3 ทางซึ่งทำให้ผู้ใช้มีความยืดหยุ่นต่อการใช้งานมากข โดยโปรแกรมจะรับข้อมูล 3 ลักษณะดังนี้

1 Text File เป็น File ข้อมูลที่สร้างจากโปรแกรม Text Editor ทั่วไป หรือ NotePad ซึ่งจะต้องแยกแต่ละ Column ด้วยช่องว่าง และข้อมูลจะต้องเป็นลักษณะแต่ละ Item ใน Item Set ต้องแยกบรรทัดกัน ดังตัวอย่าง

ID001 ADVANC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ID002 IEC

ID002 CEI

ID002 COCO

ID003 BC

ID003 EFS

ID004 ASIA

**\* โปรแกรมสามารถ รับ Column Text File ได้ทั้งหมด 10 Column**

2 ฐานข้อมูล Microsoft Access เนื่องจาก เครื่องมือที่ใช้พัฒนาสามารถเข้าถึงไฟล์ฐานข้อมูลของโปรแกรม Microsoft Access ได้โดยตรงและ โปรแกรม Microsoft เองนั้นกำลังได้รับความนิยม

3 ODBC เนื่องจาก ODBC เป็นมาตรฐานหลักที่ใช้ในการเชื่อมต่อกับฐานข้อมูลชนิดต่างๆ ซึ่งทำให้ง่ายต่อการดึงข้อมูลมาใช้งาน สำหรับการติดตั้ง ODBC สามารถดูได้จากภาคผนวก

### 5.2.2 การดึงข้อมูล

1. Text File โปรแกรมจะสามารถดึง File ข้อมูลมาประมวลผลได้ครั้งละ 1 File เท่านั้น
2. ฐานข้อมูล Microsoft Access และ ODBC ผู้ใช้สามารถเลือกข้อมูลได้ จาก ตาราง, View หรือ เลือกข้อมูลตามต้องการได้จาก คำสั่ง SQL

### 5.2.3 การเลือกข้อมูล

1. ผู้ใช้สามารถเลือก Column ที่จะนำมาเป็น Transaction ID และ Data Item ได้ (การหาความสัมพันธ์โดยใช้ Association Rule จะใช้ข้อมูลเพียง 2 Column เท่านั้น)
2. ผู้ใช้สามารถที่จะกรองข้อมูลให้เหลือเฉพาะกลุ่มที่ต้องการใช้งานได้ โดยผู้ใช้งานต้องทำการสร้างกลุ่มไว้ล่วงหน้า
3. ผู้ใช้ดูความสัมพันธ์ของรายการแบบเป็นกลุ่ม(Virtual Group)ได้ เช่น กลุ่มธนาคาร ก็ประกอบด้วย ธนาคาร ทุกรายการรวมเป็น 1 รายการ หรือกลุ่ม Entertainment & Recreation ก็ประกอบด้วย BEC Word, CVD, Grammy ฯลฯ เป็นต้น โดยหุ้นหรือรายการใด ที่ไม่เข้ากลุ่ม ก็จะถือเป็นกลุ่มอื่น ๆ

### 5.2.4 การหาความสัมพันธ์

ก่อนจะหาความสัมพันธ์ โปรแกรมสามารถให้ผู้ใช้งานระบุค่า Minimum Support และค่า Minimum Confidence โดยถ้าไม่ระบุ โปรแกรมจะตั้งค่าเริ่มต้น Minimum Support ไว้ที่ 10% และค่า Minimum Confidence เท่ากับ 80%

### 5.3 ขั้นตอนการทำงานของระบบงาน

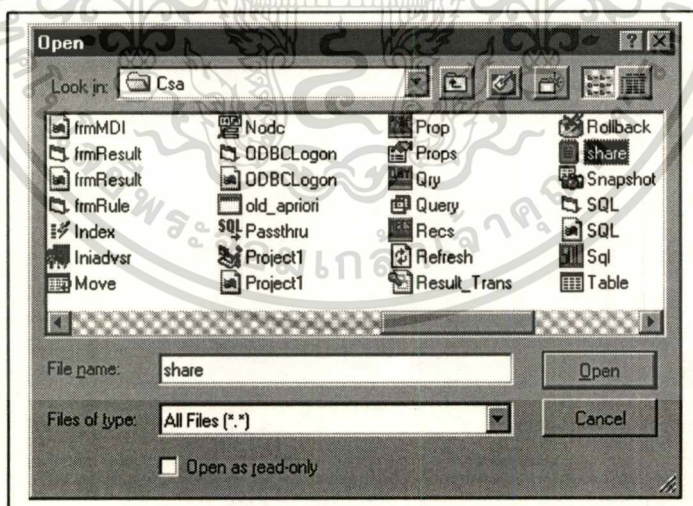
การทำงานของโปรแกรมระบบงานได้เป็นไปตามกระบวนการ Mining ดังนี้



รูปที่ 5.1 แสดงขั้นตอนการทำงานของระบบงาน

#### 5.3.1 ขั้นตอนดึงข้อมูล

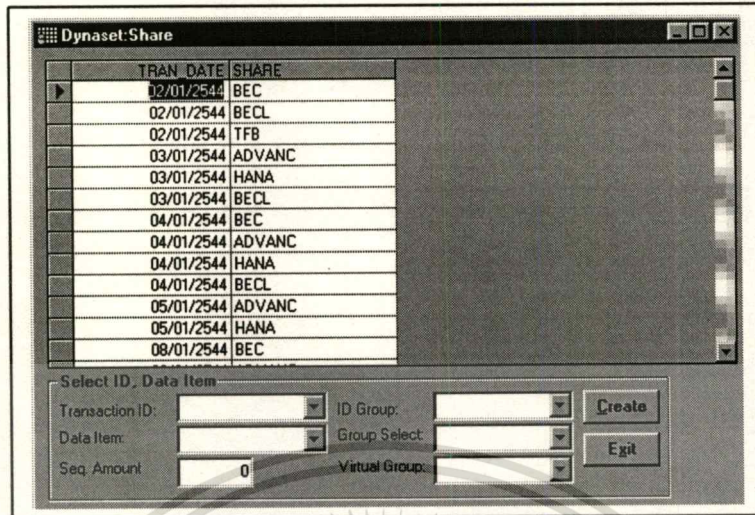
- 1) Text File สามารถเรียกใช้จาก Menu File->Open->Text Files หรือเรียกใช้จาก Icon  โปรแกรมจะแสดง Dialog เพื่อให้ผู้ใช้เลือกชื่อเพิ่มข้อมูลที่ต้องการเลือก



รูป 5.2 การเปิดใช้งาน Text File

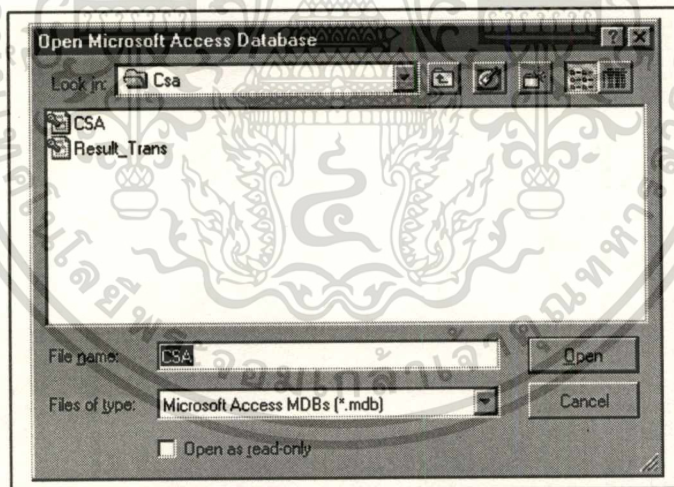
เมื่อ Click ปุ่ม Open โปรแกรมจะทำการ Import ข้อมูลแล้วนำมาแสดงในพื้นที่ทำการดึงรูป ซึ่งโปรแกรมสามารถรับข้อมูลได้สูงสุด 10 Column

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



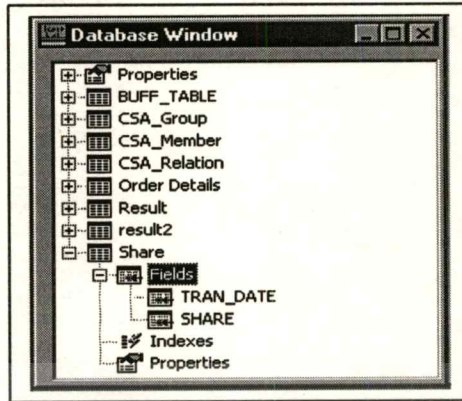
รูปที่ 5.3 แสดงผลการดึงข้อมูลจาก Text File

- 2) ฐานข้อมูล Microsoft Access สามารถเรียกใช้จาก File -> Open -> Microsoft Access หรือ เรียกใช้จาก Icon  โปรแกรมจะแสดง Dialog ให้เลือก File ที่ต้องการ เปิด



รูปที่ 5.4 แสดง Dialog เปิดฐานข้อมูล Access

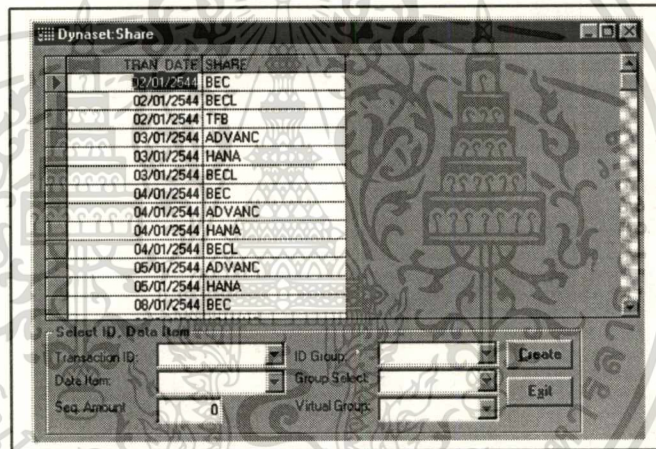
หลังจากที่เลือก File แล้ว Click ปุ่ม Open โปรแกรมจะแสดง Database Window ซึ่งแสดง Table และ View ที่อยู่ Access File นั้น ซึ่งผู้ใช้สามารถที่จะ Click ที่ชื่อ Table เพื่อดูรายชื่อ Field และ ค่า Property ต่าง ๆ ของตารางได้ด้วย



รูปที่ 5.5 แสดง Database Window

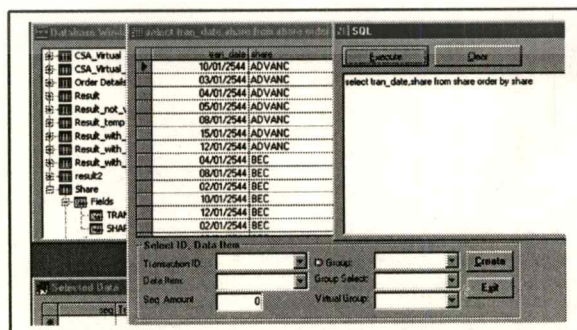
ถ้าต้องการจะเรียกใช้ข้อมูล Table หรือ View ใดก็ Double Click ที่ Table นั้น ซึ่งจะได้ผล

ดังนี้



รูปที่ 5.6 แสดงรูปการเปิดตารางใน Microsoft Access

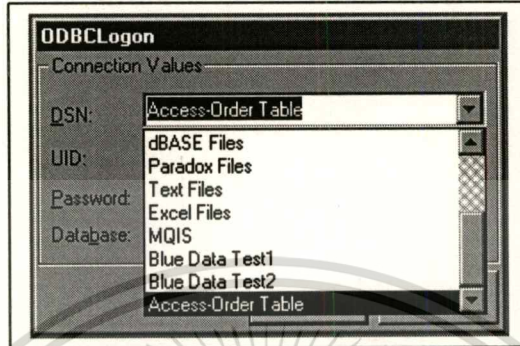
นอกจากนี้ผู้ใช้ยังสามารถใช้คำสั่ง จากหน้าจอ SQL เพื่อใช้คำสั่ง SQL ใน SQL Window เพื่อทำการคัดข้อมูลได้อีกด้วย



รูปที่ 5.7 การใช้คำสั่ง SQL ในการเลือกข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 3) การเปิดด้วย ODBC จะต้องมีการกำหนดตัว Drive ไว้ก่อน โดยสามารถเรียกใช้ได้จาก File-> Open -> ODBC หรือ Click จาก Icon 



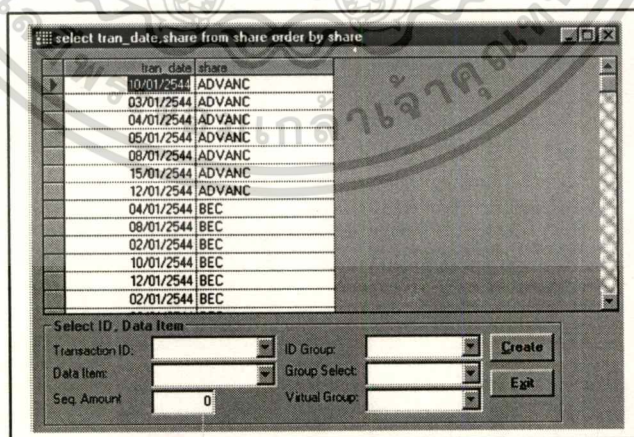
รูปที่ 5.8 แสดงการติดต่อผ่านทาง ODBC

ผู้ใช้ DSN ที่ระบุใน Control Panel นอกจากนี้ ถ้า ODBC ตัวนั้น ต้องการ รายละเอียดอื่นๆแล้ว ผู้ใช้ก็ต้องป้อนให้ด้วย เช่น User ID, Password, Database เป็นต้น ซึ่งเมื่อเลือกถูกต้องแล้วกดปุ่ม OK โปรแกรมจะทำการดึง Table ที่มีตาม ODBC ระบุออกมาแสดงในจอ Database Window เหมือนกับการดึงข้อมูลจาก Database Microsoft Access

### 5.3.2 การเลือกข้อมูล

เมื่อทำการดึงข้อมูลมาสู่หน้าต่างการทำงานแล้ว โปรแกรมสามารถให้ผู้ใช้เลือกข้อมูลได้ดัง

นี้



รูปที่ 5.9 แสดงหน้าต่างทำการ

ในหน้าต่างทำการนี้ ผู้ใช้สามารถ เลือก Column Transaction ID จาก Combo box

Transaction ID และ เลือก Data Item จาก Data Item Combo box ส่วน Combo box ID Group นั้นใช้เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการเลือก Group Sequence Number เข้าด้วยกัน จากเดิม เมื่อ Transaction ID มีค่าเปลี่ยน ตัว Sequence จะเปลี่ยนด้วย แต่ ถ้ามีการ Set ค่า Group Sequence Number เป็น 2 ก็ จะทำการ Group ทุก 2 Transaction ID เข้าด้วยกัน

seq	Transaction_ID	data_item	group_code
1	02/01/2544	BEC	
1	02/01/2544	BECL	
1	02/01/2544	TFB	
1	03/01/2544	ADVANC	
1	03/01/2544	BECL	
1	03/01/2544	HANA	
2	03/01/2544	ADVANC	
2	03/01/2544	BECL	
2	03/01/2544	HANA	
2	03/08/2544	CCC	
3	03/08/2544	CCC	
3	04/01/2544	ADVANC	
3	04/01/2544	BEC	
3	04/01/2544	BECL	

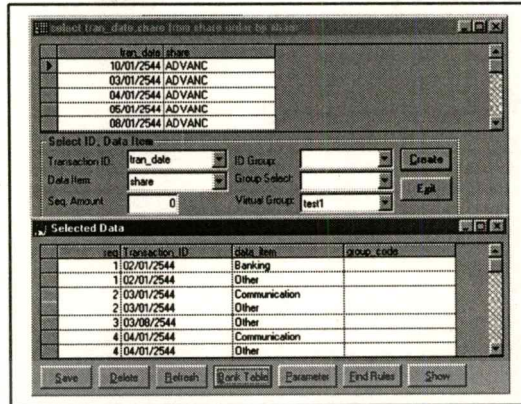
### รูปที่ 5.10 แสดงข้อมูลที่ทำการเลือกมาแล้ว

กรณีที่ต้องการจำนวน Sequence ที่แน่นอนก็สามารถระบุได้ในช่อง Seq. Amount ในส่วน Combo box Group Select เป็นการเลือก Group หรือ Visual Group โดยเลือก Group เมื่อต้องการเลือกข้อมูลเฉพาะ กลุ่มนั้น และเลือก Visual Group เมื่อต้องการหา ความสัมพันธ์ของ Group ของ Data Item แทน ซึ่งเมื่อทำการประมวลผลแล้วข้อมูลจะปรากฏใน Window Selected Data

seq	Transaction_ID	data_item	group_code
1	03/01/2544	ADVANC	Communication
2	04/01/2544	ADVANC	Communication
3	05/01/2544	ADVANC	Communication
4	08/01/2544	ADVANC	Communication
5	10/01/2544	ADVANC	Communication
6	12/01/2544	ADVANC	Communication
7	15/01/2544	ADVANC	Communication

### รูปที่ 5.11 แสดงข้อมูลที่ทำการเลือกเฉพาะกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

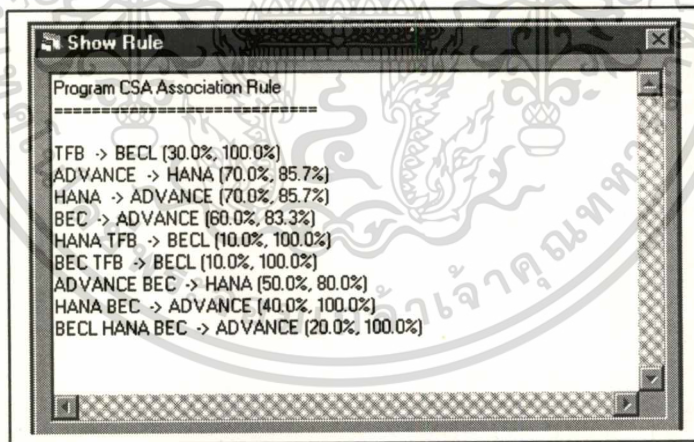


รูปที่ 5.12 แสดงข้อมูลที่ทำให้การเลือกแบบ Virtual Group

เมื่อได้ข้อมูลแล้วผู้ใช้งานสามารถตรวจสอบเช็คได้ว่าเป็นไปตามเงื่อนไขที่ต้องการหรือไม่ถ้าตรงความต้องการแล้วก็สามารถดำเนินการขั้นต่อไปได้

### 5.3.3 การค้นหากฎ

การค้นหาจะเป็นการนำเอาข้อมูลที่เราได้ทำการเตรียมข้อมูล มา Run ผ่าน Apriori Algorithm แล้วจะแสดงผลออกมาดังนี้



รูปที่ 5.13 แสดงผลลัพธ์จากการค้นหาความสัมพันธ์

จากรูปที่ 5.10 ผลที่แสดงออกมากจะเป็นการแสดงในรูปแบบ

Condition 1 -> Condition 2 (Support percent, Confidence percent) หรือ  
Left-hand side -> Right-hand side (Support percent, Confidence percent)

จากตัวอย่างจะเห็นว่า เมื่อซื้อหุ้น TFB และจะซื้อหุ้น BECL ด้วยคิดเป็นค่า support ได้ 30 %  
และเมื่อซื้อหุ้น TFB แล้วมีโอกาสซื้อหุ้น BECL คิดเป็น 100 %

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 5.3.4 การจัดกลุ่ม Data Item

เป็นการจัดกลุ่มเพื่อนำมาทำการ Select Data ซึ่งสามารถเลือกได้จาก Maintenance -> Group ซึ่งจะแบ่งเป็น Tab 3 Tab คือ Group, Member, Relationship, Virtual Group, Virtual Relation ดังรูป



รูปที่ 5.14 การจัดการกลุ่ม

### 5.4 สรุปผลการพัฒนาโปรแกรม

จากพัฒนาโปรแกรมได้ทำให้ผู้พัฒนาได้เข้าใจถึงการในเอากระบวนการ Data Mining มาใช้ประโยชน์ ผลจากทดลองนั้นไปผลอยู่ในระดับหนึ่ง ซึ่งการทำงานของ Data Mining นั้นขึ้นอยู่กับความพร้อมข้อมูลที่ดี ซึ่งถ้าขั้นตอนนี้ได้กระทำไม่ดีแล้ว ผลการทดสอบก็จะไม่ได้ผล ซึ่งโปรแกรมก็พยายามยืดหยุ่นให้ผู้ใช้สามารถจัดการกับข้อมูลได้ง่ายขึ้น

### 5.5 ข้อเสนอแนะ

การนำเอากระบวนการ Data Mining ที่นำมาประยุกต์กับการซื้อขายหลักทรัพย์นี้ เป็นเพียงแนวทางหนึ่งของการประยุกต์ใช้งานเท่านั้น อีกทั้งยังขาดในแง่ความคล่องตัวในการใช้งานอีกหลายอย่าง ซึ่งการดูแลพฤติกรรม ที่ดูจาก Association Rule อย่างเดียวก็น่าจะยังไม่เพียงพอ ต้องดูปัจจัยอื่นประกอบด้วย จึงจะทำให้การดำเนินงานมีความเที่ยงตรงขึ้น นอกจากนี้แล้วผู้พัฒนาหวังว่าการนำการประยุกต์มาใช้ในงานครั้งนี้ จะทำให้ผู้ที่สนใจ เกิดความคิดในการนำกระบวนการ Data Mining ไปประยุกต์ใช้งานอื่นต่อไป

## บรรณานุกรม

Estelle Brand and Rob Gerritsen. 1998. **Association and Sequencing**. [Online]

Available: <http://www.dbmsmag.com/9807m03.html>

Giwebb&Associates. 1999, 2000. **Using Magnum Opus with basket data**. [Online]

Available: <http://www.rulequest.com/MOUsingBasket.html>

Karuna Pande Joshi. 1997. **Analysis of Data Mining Algorithms**. [Online] Available:

[http://userpages.umbc.edu/~kjoshi1/data-mine/proj\\_rpt.htm](http://userpages.umbc.edu/~kjoshi1/data-mine/proj_rpt.htm)

Rakesh Agrawal and Ramakrishnan Srikant. **Fast Algorithms for Mining Association Rules**.

[Online] Available: [http://www.cs.berkeley.edu/~1gribble/summaries/database/fast\\_mining.html](http://www.cs.berkeley.edu/~1gribble/summaries/database/fast_mining.html)

Stonebraker and Hellerstein, eds. 1998. **Data Mining**. [Online] Available:

<http://redbook.cs.berkeley.edu/lec29.html>

## ภาคผนวก

### ก. ความต้องการระบบ

- 1) Microsoft windows 98 Mode การแสดงผล 800 \* 600
- 2) Visual Basic Run Time และ Service Pack 4

### ข. รายละเอียด File CSA.INI

CSA.INI เป็น แฟ้มข้อมูลที่โปรแกรมเรียกใช้เพื่อที่จะให้ทราบถึงที่อยู่และชื่อแฟ้มข้อมูลที่ใช้เก็บ Data ที่ผ่านการทำ Data Selection ในขั้นตอน Data Preparation และ แฟ้มข้อมูลต้นทางเพื่อที่จะส่งให้ส่วนการค้นหาคความสัมพันธ์จนถึงแฟ้มข้อมูลผลลัพธ์ โดยแฟ้มข้อมูล CSA.INI จะถูกติดตั้งไว้ใน Sub Directory “C:\windows” โดยมีรายละเอียด ดังนี้

- “DB” เป็นคำที่ใช้บ่งบอกว่าเราเก็บฐานข้อมูลไว้ที่ไหน สามารถกำหนดการใช้ดังนี้  
**DB=C:\PROJECT\CSA\CSA.MDB**  
เป็นการบอกโปรแกรมว่าเพิ่มฐานเก็บผลลัพธ์ ชื่อ CSA.MDB เก็บอยู่ที่ Sub Directory C:\PROJECT\CSA
- “INPUT” เป็นชื่อ Text File ที่ส่งข้อมูล Input เพื่อส่งเข้าโปรแกรมค้นหาคความสัมพันธ์ สามารถกำหนดการใช้ดังนี้  
**INPUT=C:\PROJECT\CSA\CSA\_IN.TAB**
- “OUTPUT” เป็นชื่อ Text File ที่ส่งข้อมูล Output จาก โปรแกรมค้นหาคความสัมพันธ์ สามารถกำหนดการใช้งานได้ดังนี้  
**OUTPUT=C:\PROJECT\CSA\CSA\_OUT.RUL**
- “PROCESS” ชื่อโปรแกรมที่ใช้สำหรับค้นหาคความสัมพันธ์ มีการกำหนดการใช้งานดังนี้  
**PROCESS=C:\PROJECT\CSA\APRIOR1.EXE**

### ค. โครงสร้างตารางที่ใช้ในระบบ

ในระบบงานได้มีการใช้งานตารางภายในระบบเอง ซึ่งเก็บอยู่ใน ฐานข้อมูล Access ชื่อ CSA.MDB ดังมีรายละเอียดดังนี้

#### ● ตาราง Result

Seq	Number	ลำดับ Transaction.
Transaction_ID	Text (50)	Original Transaction ID.
Data_Item	Text (50)	Data Item.

#### ● ตาราง CSA\_Group

Group_Code	Text (50)	รหัสกลุ่มหุ้น
Group_Detail	Text (50)	ชื่อกลุ่ม

#### ● ตาราง CSA\_Member

Member_Symbol	Text (50)	รหัสสมาชิก
Member_Detail	Text (50)	รายละเอียดกลุ่มหุ้น

#### ● ตาราง CSA\_Relation

Group_Code	Text (50)	รหัสกลุ่มหุ้น
Member_Symbol	Text (50)	รหัสสมาชิก

#### ● ตาราง CSA\_Virtual

Virtual_Code	Text (50)	รหัสกลุ่ม Virtual
Virutal_Desc	Text (50)	ชื่อกลุ่ม

#### ● ตาราง CSA\_Virtual\_Rel

Virtual_Code	Text (50)	รหัสกลุ่ม Virtual
Group_Code	Text (50)	รหัสกลุ่มหุ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### **๓. Library Reference**

Visual Basic for Applications (MSVBVM60.DLL)

Visual Basic for runtime objects and procedures (MSVBVM60.DLL\3)

Visual Basic objects and procedures (VB6.OLB)

OLE Automation (StdOle2.Tlb)

Help System Services 1.0 Type Library

Microsoft DAO 3.51 Object Library

Microsoft Visual Basic 6.0 Extensibility (VBExt.OLB)

Microsoft ActiveX Data Objects 2.5 Library (MSADO15.DLL)

Microsoft Data Binding Collection VB6.0(SP4) (MSBIND.DLL)

### **๔. Visual basic Components**

Microsoft ADO Data control 6.0 (SP4)(OLEDB)

Microsoft Common Dialog Control 6.0 (SP3)

Microsoft Data Bound Grid Control 5.0 (SP3)

Microsoft Data Bound List Control 6.0

Microsoft Data Control 6.0 (SP4) (OLEDB)

Microsoft Rich Textbox Control 6.0 (SP4)

Microsoft Tabbed Dialog Control 6.0 (SP4)

Microsoft Windows Common Control 6.0 (SP4)

## ประวัติผู้เขียน

ชื่อผู้เขียน

นายมนัส เข้ม โศกสูง

วันเดือนปีเกิด

28 สิงหาคม 2513

สถานที่เกิด

จังหวัด นครราชสีมา

วุฒิการศึกษาระดับปริญญาตรี

วท.บ.(วิทยาการคอมพิวเตอร์)

สถานที่สำเร็จการศึกษา

คณะวิทยาศาสตร์ วิทยาลัยครูจันทระเกษม

ปีที่สำเร็จการศึกษา

ปีการศึกษา 2536

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้