

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

การพัฒนาระบบการจัดกลุ่มพฤติกรรมการใช้บริการระบบออนไลน์อัตโนมัติ

ATS SERVICE CUSTOMER BEHAVIOR CLUSTERING SYSTEM  
DEVELOPMENT



H005933



ดพ.

๘๓๙๖ ๗

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน

๘๐๐/

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

เลขหมู่.....

05933

คณะเทคโนโลยีสารสนเทศ

เลขทะเบียน.....

วัน,เดือน,ปี ๕ 3 ๓.๗. 2553

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

b. ๑๒๑๗๖๓๓  
i.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ภายในที่เรียนศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ภาคฤดูร้อน ปีการศึกษา 2551

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ATS SERVICE CUSTOMER BEHAVIOR CLUSTERING SYSTEM  
DEVELOPMENT**



**A SYSTEM DEVELOPMENT PROJECT  
OF THE REQUIREMENT FOR THE DEGREE OF  
MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY  
FACULTY OF INFORMATION TECNOLOGY  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ภายในเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
**SUMMER/2008**  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**COPYRIGHT 2009**

**FACULTY OF INFORMATION TECHNOLOGY**

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG** บัณฑิตยสถาน

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การพัฒนาระบบการจัดกลุ่มพฤติกรรมการใช้บริการระบบออนไลน์อัตโนมัติ
นักศึกษา	นายรัช วิวัฒน์วารมย์
รหัสนักศึกษา	50066502
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2551
อาจารย์ที่ปรึกษา	รศ.ดร.วรพจน์ กรีสระเดช

### บทคัดย่อ

โครงการพัฒนาระบบการจัดกลุ่มพฤติกรรมการใช้บริการระบบออนไลน์อัตโนมัติ ของธนาคารพาณิชย์แห่งหนึ่งในประเทศไทย โดยอาศัยเทคนิคการจัดกลุ่มข้อมูลเพื่อวิเคราะห์หากลุ่มลูกค้าที่มีคุณค่า (Valued Customers) เพื่อใช้ในการกำหนดกลยุทธ์ทางการตลาด หรือแผนการส่งเสริมการขายของแต่ละกลุ่มลูกค้าได้อย่างเหมาะสม ในส่วนของการพัฒนาแอปพลิเคชันนั้นได้นำเอาเครื่องมือและเทคโนโลยีต่างๆ ได้แก่ MS Analysis Server 2005, MS Visual Studio 2005 Professional Edition และ MS .Net Framework

<b>Title</b>	ATS Service Customer Behavior Clustering System Development
<b>Student</b>	Mr. Thawat Wiwatthanawarom
<b>Student ID.</b>	50066502
<b>Degree</b>	Master of Science
<b>Programme</b>	Information Science
<b>Academic Year</b>	2008
<b>Advisor</b>	Assoc. Prof. Dr. Worapoj Kreesuradej

### ABSTRACT

ATS Service Customer Behavior Clustering System Development that use the automatic transfer system (ATS) of once commercial bank in Thailand. Technical for Customer Segmentation are grouping data for find valued customers that to make properly the marketing strategies or planning the promotion of costumer's group in the future. To complete this project, we need some tools such as MS Analysis Server 2005, MS Visual Studio 2005 Professional Edition and MS .Net Framework.

# กิตติกรรมประกาศ

โครงการนี้สำเร็จได้อย่างดีด้วยคำแนะนำ และคำปรึกษาจาก รศ.ดร.วรพจน์ กรีสู่ระเดชซึ่งเป็นอาจารย์ที่ปรึกษา และคอยให้คำปรึกษาตลอดทั้งโครงการ

ขอกราบขอบพระคุณคณาจารย์คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า

ขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกเรื่องๆ ทำให้ข้าพเจ้าสามารถทำโครงการนี้สำเร็จลุล่วงด้วยดี

สุดท้ายนี้ข้าพเจ้าหวังเป็นอย่างยิ่งที่จะนำความรู้ที่ได้ไปใช้ประโยชน์ให้มากที่สุด

ธวัช วิวัฒนารามย์



# สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	1
1.3 ขอบเขตการดำเนินการ.....	2
1.4 ขั้นตอนการดำเนินการ.....	2
1.5 เครื่องมือที่ใช้ในการ.....	2
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	4
บทที่ 2 ทฤษฎีพื้นฐานที่เกี่ยวข้องกับคาด้าไมนิ่ง.....	5
2.1 คาด้าไมนิ่ง (Data Mining).....	5
2.1.1 คำจำกัดความของคาด้าไมนิ่ง.....	5
2.1.2 ประวัติการทำเหมืองข้อมูล.....	5
2.1.3 วัตถุประสงค์ในการทำคาด้าไมนิ่ง.....	6
2.1.4 ปัจจัยที่สนับสนุนการนำข้อมูลมาทำคาด้าไมนิ่ง (Data Mining).....	6
2.1.5 ตัวอย่างการทำเหมืองข้อมูล.....	7
2.1.6 ประโยชน์ของการใช้ Data Mining.....	9
2.1.7 งานของคาด้าไมนิ่ง (Data Mining Tasks).....	9
2.2 การจัดกลุ่ม (Clustering) โดยวิธีเคมีนอัลกอริทึม (K-Means Algorithm).....	10
2.2.1 การจัดกลุ่ม (Clustering).....	10
2.2.2 เทคนิคการจัดกลุ่มด้วย K-Means.....	12

# สารบัญ (ต่อ)

หน้า

2.3 ไมโครซอฟต์แวร์คลัสเตอร์ริงอัลกอริทึม (Microsoft Clustering Algorithm).....	15
2.3.1 ไมโครซอฟต์แวร์คลัสเตอร์ริง (Microsoft Clustering).....	15
2.3.2 พารามิเตอร์ในไมโครซอฟต์แวร์คลัสเตอร์ริงอัลกอริทึม.....	17
2.3.3 การวัดประสิทธิภาพการจัดกลุ่ม.....	18
<b>บทที่ 3 การวิเคราะห์และออกแบบระบบงาน.....</b>	<b>19</b>
3.1 ลักษณะการทำงานของระบบงานปัจจุบัน.....	19
3.1.1 การบริการนำเงินเข้าบัญชีโดยระบบอัตโนมัติ (Credit Fund Transfer).....	19
3.1.2 การบริการหักบัญชีโดยระบบอัตโนมัติ (Debit Fund Transfer).....	19
3.2 ปัญหาที่พบในระบบงานปัจจุบัน.....	20
3.3 การศึกษาความต้องการของระบบงานใหม่.....	20
3.4 การเตรียมข้อมูล.....	21
3.5 ยูสเคสไดอะแกรม (Use-case Diagram).....	25
3.6 ลักษณะการทำงานของระบบ (Activity Diagram).....	26
3.7 คลาสไดอะแกรม (Class Diagram).....	29
3.8 ซีควเอนซ์ไดอะแกรม (Sequence Diagram).....	30
<b>บทที่ 4 การพัฒนาระบบการจัดกลุ่มพฤติกรรมการใช้บริการระบบโอนเงินอัตโนมัติ.....</b>	<b>31</b>
4.1 การสร้างแบบจำลองการจัดกลุ่ม.....	31
4.2 การไม่นิ่งข้อมูล.....	33
4.3 การแสดงผลการจัดกลุ่ม.....	33
4.4 การประมาณค่าธรรมเนียมจากแบบจำลอง.....	37
4.5 การวิเคราะห์ผลการจัดกลุ่ม (Clustering).....	39
<b>บทที่ 5 สรุปผลการดำเนินงาน.....</b>	<b>33</b>
5.1 สรุปผลการดำเนินงาน.....	33
5.2 ข้อเสนอแนะ.....	33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา V และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

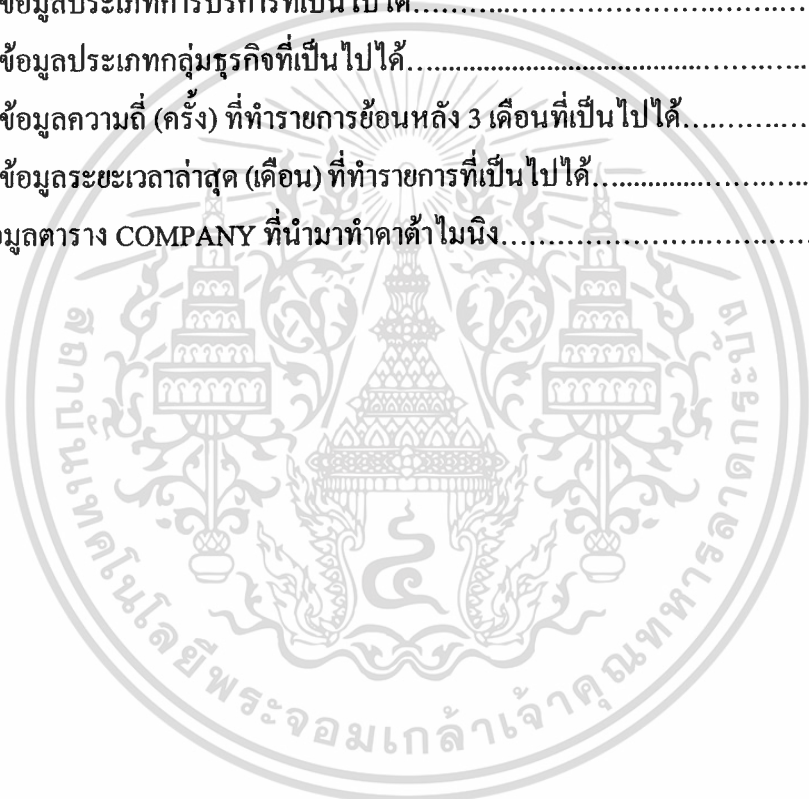
# สารบัญ (ต่อ)

	หน้า
บรรณานุกรม.....	41
ประวัติผู้เขียน.....	42



# สารบัญตาราง

ตารางที่	หน้า
2.1 ประวัติการทำเหมืองข้อมูล.....	6
2.2 พารามิเตอร์ต่างๆที่ใช้ในแบบจำลองระบบ.....	17
3.1 แสดงค่าข้อมูลสถานะการเก็บค่าธรรมเนียมขั้นต่ำที่เป็นไปได้.....	22
3.2 แสดงค่าข้อมูลประเภทการทำรายการที่เป็นไปได้.....	22
3.3 แสดงค่าข้อมูลประเภทการทำรายการที่เป็นไปได้.....	22
3.4 แสดงค่าข้อมูลประเภทการบริการที่เป็นไปได้.....	22
3.5 แสดงค่าข้อมูลประเภทกลุ่มธุรกิจที่เป็นไปได้.....	23
3.6 แสดงค่าข้อมูลความถี่ (ครั้ง) ที่ทำรายการย้อนหลัง 3 เดือนที่เป็นไปได้.....	23
3.7 แสดงค่าข้อมูลระยะเวลาล่าสุด (เดือน) ที่ทำรายการที่เป็นไปได้.....	24
3.8 แสดงข้อมูลตาราง COMPANY ที่นำมาทำคาด้าไมนิ่ง.....	24



# สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างการจัดกลุ่ม (Clustering).....	11
2.2 กำหนดที่ตั้งของตัวแทนกลุ่ม.....	12
2.3 กำหนดขอบเขตเริ่มต้นของกลุ่ม.....	13
2.4 คำนวณหาศูนย์กลางของกลุ่ม และย้ายตำแหน่ง.....	13
2.5 ทำซ้ำทุกกลุ่มที่กำหนดค่าใหม่.....	14
3.1 การบริการนำเงินเข้าบัญชีโดยระบบอัตโนมัติ.....	19
3.2 การบริการหักบัญชีโดยระบบอัตโนมัติ.....	20
3.3 ยูสเคสไดอะแกรมของระบบการจัดกลุ่มพฤติกรรมการใช้บริการ โอนเงินอัตโนมัติ.....	25
3.4 แอกทिवิตีไดอะแกรมการ Create Data Mining Model.....	26
3.5 แอกทिवิตีไดอะแกรมการ Process Data Mining Model.....	27
3.6 แอกทिवิตีไดอะแกรมการ View Data Mining Model.....	28
3.7 แอกทिवิตีไดอะแกรมการ Estimate Fee.....	28
3.8 แสดงคลาสไดอะแกรมของระบบการจัดกลุ่มพฤติกรรมการใช้บริการ โอนเงินอัตโนมัติ.....	29
3.9 ซีควเอนซ์ไดอะแกรมการจัดกลุ่มพฤติกรรมการใช้บริการ โอนเงินอัตโนมัติ.....	30
4.1 การบริการหักบัญชีโดยระบบอัตโนมัติ.....	32
4.2 แบบจำลองการค้าไมนิ่ง ATSPProject .....	32
4.3 หน้าจอการไมนิ่งข้อมูล.....	33
4.4 หน้าจอการแสดงผลรายละเอียด Content ของแบบจำลอง.....	34
4.5 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Diagram.....	34
4.6 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Profile.....	35
4.7 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Characteristics.....	35
4.8 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Discrimination.....	36
4.9 หน้าจอการแสดงผลการจัดกลุ่ม: การ Drill Through.....	36
4.10 หน้าจอการประมาณค่าธรรมเนียมจากแบบจำลอง.....	31
4.11 แสดงผลการจัดกลุ่มพฤติกรรมการใช้บริการระบบ โอนเงินอัตโนมัติ.....	39

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ธุรกิจการเงินการธนาคารเป็นอีกธุรกิจหนึ่งที่มีการแข่งขันกันค่อนข้างสูง ธนาคารพาณิชย์แต่ละแห่งต่างก็พยายามที่สร้างผลิตภัณฑ์ หรือช่องทางในการบริการใหม่ๆ เพื่อตอบสนองต่อความต้องการของลูกค้าให้มากที่สุด การสร้างความแตกต่างของผลิตภัณฑ์ หรือการบริการเป็นอีกวิธีหนึ่งเพื่อสร้างความได้เปรียบคู่แข่ง ซึ่งระบบโอนเงินอัตโนมัติ (ATS : Automatic Transfer System) เป็นอีกช่องทางหนึ่งในการให้บริการที่สามารถตอบสนองต่อความต้องการของลูกค้าด้านความสะดวก และรวดเร็วในการใช้บริการ พร้อมทั้งประหยัดค่าใช้จ่ายในการเดินทาง ซึ่งการบริการดังกล่าวยังไม่ประสบความสำเร็จเท่าที่ควร เนื่องจากลูกค้ายังขาดความไว้วางใจเรื่องความปลอดภัย ความถูกต้องแม่นยำในการใช้บริการ ตลอดจนในระยะแรกที่เปิดให้บริการยังมีข้อจำกัดของประเภทการให้บริการยังไม่ครอบคลุม หรือประเภทที่ให้บริการที่มีอยู่ไม่ตรงต่อความต้องการใช้งานของผู้ใช้บริการ ซึ่งนับเป็นสาเหตุที่สำคัญอย่างหนึ่ง ดังนั้นการนำข้อมูลการใช้งานระบบโอนเงินอัตโนมัติมาศึกษาถึงพฤติกรรมการใช้บริการ โดยการจัดกลุ่มข้อมูล เพื่อใช้เป็นแนวทางในการกำหนดกลยุทธ์ทางการตลาดหรือแผนการส่งเสริมการขายของแต่ละกลุ่มลูกค้าได้อย่างเหมาะสม

### 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

ในการดำเนินโครงการเพื่อพัฒนาระบบการจัดการกลุ่มพฤติกรรมการใช้บริการระบบโอนเงินอัตโนมัติ เกิดจากความต้องการที่จะพัฒนาระบบโอนเงินอัตโนมัติ (ATS : Automatic Transfer System) ให้มีประสิทธิภาพและสอดคล้องกับความต้องการของลูกค้ามากขึ้น ซึ่งมีวัตถุประสงค์ ดังนี้

1. เพื่อศึกษาและประยุกต์ใช้เคมีนอัลกอริทึม (K-Means Algorithm) ในการจัดกลุ่มข้อมูลการให้บริการของลูกค้า
2. เพื่อพัฒนาระบบการจัดการกลุ่มพฤติกรรมการใช้บริการระบบโอนเงินอัตโนมัติ (ATS : Automatic Transfer System) เพื่อช่วยในการวิเคราะห์การจัดกลุ่มลูกค้า
3. เพื่อใช้เป็นแนวทางในการปรับปรุงผลิตภัณฑ์ และช่องทางในการให้บริการ
4. เพื่อใช้เป็นแนวทางในการกำหนดกลยุทธ์ทางการตลาด หรือแผนการส่งเสริมการขาย

ให้แต่ละกลุ่มได้อย่างเหมาะสม

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 1.3 ขอบเขตการดำเนินการ

ในการศึกษานี้จะนำข้อมูลการใช้บริการของธนาคารพาณิชย์แห่งหนึ่งในประเทศไทยช่วงเดือนมกราคม พ.ศ. 2550 ถึงเดือนธันวาคม พ.ศ. 2551 เพื่อใช้เป็นข้อมูลในการพัฒนาระบบ โดยมีขอบเขตในการดำเนินการ ดังนี้

1. การจัดเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์การจัดกลุ่มข้อมูลการใช้บริการของระบบการโอนเงินอัตโนมัติ
2. การพัฒนาโปรแกรมประยุกต์ เพื่อใช้ในการสร้าง และการจัดการ ไม่นิ่ง โมเดล ตลอดจนการแสดงผลการวิเคราะห์

### 1.4 ขั้นตอนการดำเนินการ

ในการพัฒนาระบบ มีขั้นตอนในการดำเนินการ ดังนี้

1. ศึกษาข้อมูลในระบบโอนเงินอัตโนมัติ เพื่อกำหนดขอบเขตข้อมูลที่จะนำไปใช้ในการทำเหมืองข้อมูล
2. วิเคราะห์ปัญหาและความต้องการของผู้ใช้ ระบบโอนเงินอัตโนมัติ
3. การเตรียมข้อมูลที่จะใช้ในการพัฒนา
4. วิเคราะห์และออกแบบ ไม่นิ่ง โมเดล
5. ศึกษาและพัฒนาโปรแกรมประยุกต์ในการจัดการ ไม่นิ่ง โมเดล
6. ทดสอบและปรับปรุงแก้ไขข้อผิดพลาดของ โปรแกรมประยุกต์
7. สรุปผลการศึกษา และการใช้งานระบบที่พัฒนา
8. จัดทำเอกสารประกอบโครงการ

### 1.5 เครื่องมือที่ใช้ในการ

ในการพัฒนาระบบ มีการใช้งานเครื่องมือหรือซอฟต์แวร์ต่างๆ ดังนี้

1. รายละเอียดในด้านฮาร์ดแวร์
  - 1.1 เครื่องคอมพิวเตอร์ที่ใช้เป็น Server มีคุณสมบัติ ดังนี้
    - หน่วยประมวลผล (CPU) ชนิด Core 2 Duo มีความเร็วไม่น้อยกว่า 2.66 GHz
    - หน่วยความจำหลัก (RAM) มีความจุไม่น้อยกว่า 1 GB
    - หน่วยความจำสำรอง (Harddisk) มีความจุไม่น้อยกว่า 320 GB
    - จอภาพ (Monitor) เป็นชนิด LCD ขนาดไม่ต่ำกว่า 17 นิ้ว
    - DVD writer 16X หรือดีกว่า

เอกสารนี้เป็นเอกสารที่สงวนเวลาสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีเหตุผลบางประการที่ และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 1.2 เครื่องคอมพิวเตอร์ที่ใช้เป็น Client มีคุณสมบัติ ดังนี้

- หน่วยประมวลผล (CPU) ชนิด Intel Celeron M ความเร็ว 1.7 GHz
- หน่วยความจำหลัก (RAM) ความจุ 512 MB
- หน่วยความจำสำรอง (Harddisk) ความจุ 80 GB
- จอภาพ (Monitor) เป็นชนิด LCD ขนาด 15"
- DVD writer 16X หรือดีกว่า
- LAN card , scroll mouse , USB 2 port

### 1.3 เครื่องคอมพิวเตอร์ที่ใช้ในการพัฒนาระบบ มีคุณสมบัติ ดังนี้

- หน่วยประมวลผล (CPU) ชนิด Intel Pentium IV ความเร็ว 1.6 GHz
- หน่วยความจำหลัก (RAM) ความจุ 1 GB
- หน่วยความจำสำรอง (Harddisk) ความจุ 100 GB
- จอภาพ (Monitor) เป็นชนิด LCD ขนาด 15"
- DVD writer 16X หรือดีกว่า
- LAN card , scroll mouse , USB 2 port

## 2. รายละเอียดในด้านซอฟต์แวร์

### 2.1 เครื่องคอมพิวเตอร์ที่ใช้เป็น Server มีคุณสมบัติ ดังนี้

- ระบบปฏิบัติการ Microsoft Windows 2003 Server Enterprise Edition
- Relational Database ใช้ Microsoft SQL Server 2005
- Mining Model ใช้ Microsoft SQL Analysis Service 2005

### 2.2 เครื่องคอมพิวเตอร์ที่ใช้เป็น Client มีคุณสมบัติ ดังนี้

- ระบบปฏิบัติการ Microsoft Windows XP Service Pack 2
- ติดตั้ง Microsoft Data Access Component 2.8 เพื่อใช้เชื่อมต่อกับ SQL Server
- ติดตั้ง Microsoft SQL Server 2005 Analysis Services 9.0 OLE DB Provider เพื่อใช้เชื่อมต่อกับ Microsoft SQL Server 2005 Analysis Services

### 2.3 เครื่องคอมพิวเตอร์ที่ใช้ในการพัฒนาระบบ มีคุณสมบัติ ดังนี้

- ระบบปฏิบัติการ Microsoft Windows XP Service Pack 2
- ติดตั้ง Microsoft Data Access Component 2.8 เพื่อใช้เชื่อมต่อกับ SQL Server
- ติดตั้ง Microsoft SQL Server 2005 Analysis Services 9.0 OLE DB Provider เพื่อใช้เชื่อมต่อกับ Microsoft SQL Server 2005 Analysis Services
- ติดตั้ง Visual Studio .Net 2005 ในการพัฒนาโปรแกรมประยุกต์
- ติดตั้ง Microsoft Office 2007 ในการจัดทำเอกสาร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางให้ผู้บริหารในการกำหนดกลยุทธ์การตลาด หรือแผนการส่งเสริมการขายให้ดียิ่งขึ้น
2. เพื่อเป็นแนวทางในการปรับปรุงผลิตภัณฑ์ และช่องทางในการให้บริการ
3. เพื่อความสะดวกรวดเร็ว และลดความผิดพลาดในการนำข้อมูลมาวิเคราะห์
4. โปรแกรมประยุกต์ที่พัฒนาขึ้นช่วยให้สามารถทำการวิเคราะห์ข้อมูลจากสถานะที่อื่นได้ โดยทำการเชื่อมต่อกับ Server



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

# ทฤษฎีพื้นฐานที่เกี่ยวข้องกับดาต้าไมนิ่ง

ในเอกสารบทนี้จะกล่าวถึงทฤษฎีพื้นฐานต่างๆ และเครื่องมือที่ใช้ในการพัฒนาระบบการ จัดกลุ่มพฤติกรรมการใช้บริการระบบ โอนเงินอัตโนมัติ ดังนี้

## 2.1 ดาต้าไมนิ่ง (Data Mining)

### 2.1.1 คำจำกัดความของดาต้าไมนิ่ง

ดาต้าไมนิ่ง (Data Mining) หรือการทำเหมืองข้อมูล คือ ขั้นตอนหนึ่งของกระบวนการ ที่เกี่ยวข้องกับการค้นหารูปแบบ ความสัมพันธ์ของข้อมูลที่ถูกเก็บรวบรวมไว้เป็นจำนวนมากโดย จะใช้เทคนิคทางสถิติหรือเทคนิคทางคณิตศาสตร์ นอกจากนี้ยังมีผู้ให้นิยามอื่นอีกหลายนิยาม เช่น เหมืองข้อมูล คือ มองเป็นกระบวนการทั้งหมดในการทำดาต้าไมนิ่ง (Data Mining) ซึ่งเรียกว่า “Knowledge discovery in database หรือ KDD”

จากนิยามการทำเหมืองข้อมูลจะเห็นว่ามีความหลากหลายแต่ใจความหลัก คือ กระบวนการที่ ช่วยให้สามารถได้สารสนเทศที่ซ่อนอยู่ในข้อมูล และยังเป็นการใช้ข้อมูลอย่างคุ้มค่า แต่การทำ เหมืองข้อมูลเป็นวิธีการที่ค่อนข้างยุ่งยาก ปัจจุบันจึงมีการนำโปรแกรมคอมพิวเตอร์เข้ามาช่วยใน การวิเคราะห์เช่น SPSS , SAS , Oracle ,Microsoft SQL Server เป็นต้น ซึ่งการเหมืองข้อมูลยัง สามารถพัฒนาโปรแกรมเองเพื่อช่วยในการวิเคราะห์เองได้ด้วย

### 2.1.2 ประวัติการทำเหมืองข้อมูล

เทคนิคเหมืองข้อมูลเป็นผลจากการวิจัยและกระบวนการพัฒนาผลผลิตเป็นเวลานาน โดย ในตารางที่ 2.1 แสดงวิวัฒนาการของเหมืองข้อมูล เริ่มระยะแรกในปี ค.ศ. 1960 เริ่มมีการนำ คอมพิวเตอร์เข้ามาช่วยในการดำเนินการทางธุรกิจ ในการวิเคราะห์ข้อมูลระยะแรกเป็นของ เทคนิค Data collection เป็นการใช้ข้อมูลที่เก็บรวบรวมมาวิเคราะห์ซึ่งเป็นการคำนวณข้อมูลอย่าง ง่ายเช่น การหาค่าเฉลี่ย หรือผลรวม สารสนเทศที่สร้างขึ้นมานี้เป็นการตอบคำถามทาง ธุรกิจเช่น รายได้รวม หรือรายได้เฉลี่ยในแต่ละช่วงเวลา ต่อมาในปี ค.ศ. 1980 เป็นระยะเป็นของ เทคนิค Data access มีการใช้เทคโนโลยีฐานข้อมูลมาช่วยในการเก็บข้อมูล มีเริ่มมีการใช้ระบบ การจัดการรายงานเป็นเครื่องมือในการช่วยสร้างสารสนเทศ ปี ค.ศ.1990 เป็นระยะของเทคนิค Data navigation โดยเริ่มมีการใช้ฐานข้อมูลเชิงหลายมิติ, OLAP และคลังข้อมูลเพื่อเข้ามาตอบ คำถามและปัญหาต่างๆ ที่ซับซ้อนขึ้น พร้อมทั้งให้ทันต่อเหตุการณ์ที่เกิดขึ้น เช่น ยอดขายรวมใน แต่ละภูมิภาคแยกตามคาบระยะเวลาที่ต้องการ สิ้นสุดท้ายในปี ค.ศ. 2000 เป็นระยะเริ่มใช้เทคนิค

ดาต้าไมนิง (Data mining) ซึ่งเป็นวิเคราะห์ที่หาสารสนเทศที่ซ่อนอยู่ในข้อมูลโดยอาศัยข้อมูลจำนวนมาก

## ตารางที่ 2.1 ประวัติการทำเหมืองข้อมูล

เทคนิค (ปีที่เริ่ม)	คำถามหรือปัญหาทางธุรกิจ	เทคโนโลยีที่เกิดขึ้น
Data collection(1960s)	ค่าเฉลี่ยของรายได้ห้าเดือนที่ผ่านมา	คอมพิวเตอร์, เทป, Disks
Data Access (1980s)	ยอดขายที่นิวอิงแลนด์เมื่อเดือนเมษายน	ฐานข้อมูลเชิงสัมพันธ์, SQL, ODBC
Data Navigation(1990s)	ยอดขายที่นิวอิงแลนด์เมื่อเดือนเมษายน โดยแยกตามเมืองต่างๆ	OLAP, ฐานข้อมูลเชิงมิติ, คลังข้อมูล
เหมืองข้อมูล(2000)	อะไรน่าจะเกิดขึ้นจากการขายในบอสตันของเดือนหน้า เพราะอะไร	อัลกอริทึมขั้นสูง, ฐานข้อมูลขนาดใหญ่, การประมวลคอมพิวเตอร์หลายขั้นตอน

### 2.1.3 วัตถุประสงค์ในการทำดาต้าไมนิง

1. ข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลหากเก็บไว้เฉย ๆ ก็จะไม่เกิดประโยชน์ดังนั้นจึงต้องมีการสกัดสารสนเทศไปใช้การสกัดสารสนเทศ หมายถึง การคัดเลือกข้อมูลออกมาใช้งานในส่วนที่เราต้องการ
2. ในอดีตเราใช้คนเป็นผู้สืบค้นข้อมูลต่าง ๆ ในฐานข้อมูลซึ่งผู้สืบค้นจะทำการสร้างเงื่อนไขขึ้นมาตามภูมิปัญญาของผู้สืบค้น
3. ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูลเดียวอาจไม่ให้ความรู้เพียงพอและลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูงและมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นต้องรวบรวมฐานข้อมูลหลาย ๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่า “คลังข้อมูล” (Data Warehouse) ดังนั้นเราจึงจำเป็นต้องใช้ดาต้าไมนิง (Data Mining) ในการดึงข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะนำข้อมูลนั้นมาใช้งานให้เกิดประโยชน์สูงสุด

### 2.1.4 ปัจจัยที่สนับสนุนการนำข้อมูลมาทำดาต้าไมนิง (Data Mining)

1. จำนวนและขนาดข้อมูลขนาดใหญ่ถูกผลิตและขยายตัวอย่างรวดเร็ว การสืบค้นความรู้จะมีความหมายก็ต่อเมื่อฐานข้อมูลที่ใช้มีขนาดใหญ่มาก ปัจจุบันมีจำนวนและขนาดข้อมูลขนาดใหญ่ที่ขยายตัวอย่างรวดเร็ว โดยผ่านทางอินเทอร์เน็ต ดาวเทียม และแหล่งผลิตข้อมูลอื่นๆ เช่น เครื่องอ่านบาร์โค้ด, เครดิตการ์ด, อีคอมเมิร์ซ

2. ข้อมูลถูกจัดเก็บเพื่อนำไปสร้างระบบการสนับสนุนการตัดสินใจ (Decision Support System) เพื่อเป็นการง่ายต่อการนำข้อมูลมาใช้ในการวิเคราะห์เพื่อการตัดสินใจ ส่วนมากข้อมูลจะ

ถูกจัดเก็บแยกมาจากระบบปฏิบัติการ (Operational System) โดยจัดอยู่ในรูปของคลังหรือเหมืองข้อมูล (Data Warehouse) ซึ่งเป็นการง่ายต่อการนำเอาไปใช้ในการสืบค้นความรู้

3. ระบบคอมพิวเตอร์สมรรถนะสูงมีราคาต่ำลง เทคนิคดาต้าไมนิง (Data Mining) ประกอบไปด้วยอัลกอริทึม (Algorithm) ที่มีความซับซ้อนและความต้องการการคำนวณสูง จึงจำเป็นต้องใช้งานกับระบบคอมพิวเตอร์สมรรถนะสูง ปัจจุบันระบบคอมพิวเตอร์สมรรถนะสูงมีราคาต่ำลง พร้อมด้วยเริ่มมีเทคโนโลยีที่นำเครื่องไมโครคอมพิวเตอร์จำนวนมากมาเชื่อมต่อกัน โดยเครือข่ายความเร็วสูง (PC Cluster) ทำให้ได้ระบบคอมพิวเตอร์สมรรถนะสูงในราคาต่ำ

4. การแข่งขันอย่างสูงในด้านอุตสาหกรรมและการค้า เนื่องจากปัจจุบันมีการแข่งขันอย่างสูงในด้านอุตสาหกรรมและการค้า มีการผลิตข้อมูลไว้อย่างมากมายแต่ไม่ได้นำมาใช้ให้เกิดประโยชน์ จึงเป็นการจำเป็นอย่างยิ่งที่ต้องควบคุมและสืบค้นความรู้ที่ถูกซ่อนอยู่ในฐานข้อมูลความรู้ที่ได้รับสามารถนำไปวิเคราะห์เพื่อการตัดสินใจในการจัดการในระบบต่างๆ ซึ่งจะเห็นได้ว่าความรู้เหล่านี้ถือว่าเป็นผลิตผลอีกชิ้นหนึ่งเลยทีเดียว

### 2.1.5 ตัวแบบการทำเหมืองข้อมูล

จากที่กล่าวมาแล้วว่าการทำเหมืองข้อมูลเป็นขั้นตอนที่สำคัญของ Knowledge Discovery in Data (KDD) ดังนั้นเราสามารถในตัวแบบในการทำ KDD เข้ามาช่วยซึ่ง KDD คือ กระบวนการในการกำหนดและแสวงหารูปแบบอันเป็นองค์ความรู้ใหม่ที่มีประโยชน์และสามารถเข้าใจได้จากสิ่งที่ซ่อนอยู่ในข้อมูล ซึ่งประกอบด้วยขั้นตอนต่างๆ ดังนี้

1. กำหนดลักษณะของจุดมุ่งหมาย (Goals Identification) การกำหนดวัตถุประสงค์ขอบเขตของปัญหาที่จะทำการศึกษาให้ชัดเจน ซึ่งจะมีผลต่อทุกๆ ขั้นตอนในการทำ KDD

2. การคัดเลือกข้อมูล (Data Selection) การระบุแหล่งข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล รวมถึงการนำข้อมูลออกมาใช้สำหรับการวิเคราะห์เบื้องต้น โดยการเลือกข้อมูลนั้นจะแตกต่างกันไปตามวัตถุประสงค์ของแต่ละธุรกิจที่ได้กำหนดไว้ และตามลักษณะงานที่จะนำมาใช้อีกด้วย ตัวแปรที่ถูกเลือกมาแต่ละตัวนั้นจะต้องถูกทำความเข้าใจว่า มีความหมายว่าอะไร ไม่เพียงแต่คำจำกัดความทางธุรกิจเท่านั้น แต่จะต้องสามารถอธิบายถึงชนิดของข้อมูล, ค่าที่เป็นไปได้, แหล่งกำเนิดของข้อมูล และลักษณะอื่นๆ โดยตัวแปรสามารถแบ่งได้ 2 ประเภทคือ

#### ■ ตัวแปรแบบ Categorical หรือตัวแปรจัดจำแนก

- Nominal Variables กล่าวถึงตัวแปรชนิดนี้ object ที่มันอ้างถึงจะไม่มีลำดับในค่าที่เป็นไปได้ (Possible Value) เช่น สถานะการแต่งงาน (โสด, แต่งงาน, หย่า, ไม่ทราบ), เพศ (ชาย, หญิง), ระดับการศึกษา (ปริญญาโท, ปริญญาตรี, มัธยมปลาย)
- Ordinal Variables เป็นตัวแปรชนิดที่มีลำดับสำหรับค่าที่เป็นไปได้ เช่น ลำดับ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ตัวแปรแบบ Quantitative หรือตัวแปรเชิงปริมาณ
  - Continuous Variables (ตัวแปรต่อเนื่อง) เป็นตัวแปรใดๆที่มีค่าของตัวแปรเป็นสมาชิกของจำนวนจริง เช่น รายได้, ค่าเฉลี่ยจำนวนครั้งที่ซื้อ
  - Discrete Variables (ตัวแปรที่เป็นจำนวนเต็ม) เช่น จำนวนพนักงาน, เวลาปี (เดือน, ไตรมาส)

3. การเตรียมข้อมูล (Data Preprocessing) เป็นการทำความสะอาดข้อมูล (Data Cleaning) เป็นกระบวนการที่ใช้ในเตรียมและตรวจสอบข้อมูลให้ถูกต้อง ก่อนที่จะนำไปใช้งาน

4. การแปลงข้อมูล (Data Transformation) เป็นการแปลงข้อมูลทีเลือกมาให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปใช้วิเคราะห์ตามอัลกอริทึม (Algorithm) และแบบจำลองที่ใช้ในการทำค้ำไมน์นิง (Data Mining) ต่อไป การแปลงข้อมูลที่เราคุ้นเคยกันเช่น

- Data Normalization เป็นการแปลงข้อมูลโดยจะเปลี่ยนค่าข้อมูลให้อยู่ในรูปมาตรฐานอันดับที่กำหนด เนื่องจากบางครั้งอัลกอริทึมที่ใช้ในการทำเหมืองข้อมูลมีค่าอยู่ในช่วงที่กำหนด เช่น การใช้อัลกอริทึม Neural Networks ในการจำแนก ซึ่งค่าสเกลของตัวเลขที่ใช้วัดจะอยู่ในช่วง 0 และ 1 วิธีการ Data normalization มีอยู่หลายวิธี เช่น
  - Decimal Scaling เป็นการหารข้อมูลด้วยตัวเลขโดยส่วนมากตัวเลขที่เรานำมาหารนั้นจะใช้เลขสิบยกกำลัง เช่น ถ้าเรารู้ว่าข้อมูลมากและน้อยสุดอยู่ในช่วง -1000 ถึง 1000 ถ้าเราต้องการให้ข้อมูลอยู่ในช่วง -1 ถึง 1 เราจะนำ 1000 มาหาร
  - Min-Max Normalization เป็นเทคนิคที่ต้องรู้ค่าสูงสุดและค่าต่ำสุดของข้อมูล จะทำให้ข้อมูลที่ได้นั้นอยู่ในช่วง 0 และ 1

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A$$

- การทำให้อยู่ให้รูปคะแนนมาตรฐาน Z (Z-scores) โดยต้องทราบค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (stand\_dev)

$$v' = \frac{v - \text{mean}_A}{\text{stand\_dev}_A}$$

- การเปลี่ยนชนิดของข้อมูล (Data type conversion) เทคนิคเหมืองข้อมูลบางเทคนิคไม่สามารถวิเคราะห์ข้อมูลเป็นข้อมูลเชิงกลุ่มได้ จึงต้องทำการแปลงข้อมูลเชิงกลุ่มให้เป็นตัวเลขก่อนการวิเคราะห์

5. ค้ำไมน์นิง (Data Mining) เป็นการเลือกอัลกอริทึมในการทำเหมืองข้อมูล ซึ่งสามารถ

6. การแปลผลและการประเมินผล (Interpretation and Evaluation) การพิจารณาผลลัพธ์จากการทำเหมืองข้อมูลว่าจะสามารถตอบคำถามหรือตรงตามวัตถุประสงค์หรือไม่ มีการตัดสินใจว่า จะต้องทำในขั้นตอนที่ 5 ซ้ำหรือไม่ และรวมถึงการแปลผลการวิเคราะห์ให้แก่ผู้ใช้ข้อมูลเข้าใจ

7. การนำความรู้ไปใช้ (Knowledge Presentation) เป็นขั้นตอนสุดท้ายที่จะนำผลลัพธ์ของค้ำค่าไมน์นิ่งที่ได้ไปใช้

### 2.1.6 ประโยชน์ของการใช้ Data Mining

ค้ำค่าไมน์นิ่ง (Data Mining) ถูกนำมาใช้สนับสนุนการตัดสินใจ โดยการสร้างมูลค่าเพิ่มให้กับข้อมูลที่มีอยู่ ประโยชน์ที่แต่ละองค์กรได้รับจากการใช้ค้ำค่าไมน์นิ่ง (Data Mining) สรุปได้ดังนี้

1. การเอาชนะคู่แข่งชั้น ลูกค้ำที่ค้ำมีกจะเป็นที่ชื่นชอบของบริษัทคู่แข่งเช่นกัน บริษัทเหล่านั้นจะพยายามแย่ง ส่วน แบ่งจาก Segment ที่สามารถสร้างผลค้ำไรให้กับบริษัทได้ และก็พยายามที่จะแย่งชิงส่วนแบ่งตลาดนั้นมาทำค้ำค่าไมน์นิ่ง (Data Mining) สามารถนำมาใช้ประโยชน์ได้ทั้งการแย่งส่วนแบ่งตลาด และในแง่การป้องกันมิให้เกิดการเสิช ส่วนแบ่งตลาด

2. ทำให้เกิดความรู้ที่สามารถนำมาใช้ หรือประกอบการตัดสินใจได้ เนื่องจากค้ำค่าไมน์นิ่ง (Data Mining) จะใช้เทคนิคที่ซับซ้อน และมีลักษณะเป็น Artificial Intelligence ในการสร้างโมเดลที่อิงกับข้อมูล ซึ่งรวบรวมจากแหล่งต่างๆ เช่น รายงานทางธุรกิจ ข้อมูลประวัติลูกค้ำ และข้อมูลอื่นๆ จากแหล่งภายนอก ความรู้ที่ได้จึงช่วยให้องค์กร สามารถคาดการณ์อนาคต และสามารถเจาะกลุ่มตลาดได้ถูกต้องมากขึ้น

3. ใช้ในการหาข้อผิดพลาดของการปฏิบัติงาน หรือการให้บริการได้ (Fraud Detection)

4. ช่วยประหยัดค้ำใช้จ่าย โดยการทำให้ขั้นตอนการทำงานมีประสิทธิภาพมากขึ้น

5. การกำหนดเป้าหมายกลุ่มลูกค้ำได้อย่างมีประสิทธิภาพมากขึ้น ทำให้ยอดขายเพิ่มขึ้น เพิ่มจำนวนลูกค้ำ และลด โอกาสของความเสิชต่างๆ

### 2.1.7 งานของค้ำค่าไมน์นิ่ง (Data Mining Tasks)

ในทางปฏิบัติจริงค้ำค่าไมน์นิ่งจะประสบความสำเร็จกับงานบางกลุ่มเท่านั้น และต้องอยู่ภายใต้ภาวะที่จำกัดปัญหาให้เหมาะสม การใช้เทคนิคค้ำค่าไมน์นิ่งมักจะเป็นปัญหาที่ต้องใช้เหตุผลในการแก้ เป็นปัญหาที่เกี่ยวข้องกับเศรษฐศาสตร์และการเงิน ซึ่งจะสามารถจัดรูปแบบของธุรกิจให้อยู่ในรูปแบบของงานทั้ง 6 งานได้ ดังนี้

1. การจัดหมวดหมู่ (Classification)

เป็นการจำแนกของวัตถุที่ต้องการ โดยอาศัยลักษณะที่คล้ายคลึงกันหรือแตกต่างกัน เป็นการจัดวัตถุที่เข้ามาใหม่เข้ากลุ่มที่จัดไว้แล้วให้เข้ากลุ่มได้ถูกต้อง โดยเราต้องรู้ว่าวัตถุที่ต้องการจัดเข้ากลุ่มนั้นมีจำนวนกลุ่มที่แน่นอน โดยตัวแปรตามจะเป็นตัวแปรเชิงกลุ่ม การจำแนกกลุ่มจะเน้นการสร้างตัวแบบที่กำหนดว่าวัตถุที่เข้ามาใหม่นั้นจะถูกจัดเข้ากลุ่มใดกลุ่มหนึ่งที่กำหนด เช่น การ

จัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำระดับกลาง และระดับสูง ของความเสี่ยงที่จะได้รับ เป็นต้น

## 2. การประเมินค่า (Estimation)

จุดประสงค์หลักในการประมาณค่านั้นจะเน้นการกำหนดค่าของตัวแปรตามที่เราไม่ทราบค่าจากตัวแปรอิสระต่างๆ ซึ่งตัวแปรตามนั้นจะมีค่าเป็นตัวเลขหรือมีลักษณะเป็นข้อมูลชนิดต่อเนื่อง เช่น การประมาณรายได้ของพนักงานแต่ละคน

## 3. การทำนายล่วงหน้า (Prediction)

การทำนายล่วงหน้ามีลักษณะคล้ายกับการประมาณค่า แต่ตัวแบบในการทำนายจะมุ่งเน้นเป็นการศึกษาพฤติกรรมในอนาคตมากกว่าในปัจจุบัน โดยตัวแปรตามที่จะทำนายนั้นจะเป็นข้อมูลชนิดต่อเนื่องหรือไม่ต่อเนื่องก็ได้ เช่น การทำนายยอดขายของบริษัทในเดือนหน้า

## 4. การค้นหากฎความสัมพันธ์ (Association rule)

เป็นการหาความสัมพันธ์หรือความเกี่ยวเนื่องของข้อมูลโดยอาศัยหลักของกฎ ซึ่งจะอยู่ในรูปแบบ “ถ้า สิ่งที่เกิดขึ้น แล้ว ผลที่จะตามมา”

## 5. การรวมตัว (Clustering)

เป็นการรวมวัตถุที่คล้ายคลึงกัน ไว้ในกลุ่มเดียวกัน โดยจะไม่มีข้อสมมุติเกี่ยวกับจำนวนกลุ่มว่ามีกี่กลุ่ม แตกต่างจากการจำแนกกลุ่มตรงที่การจำแนกกลุ่มเป็นการจัดวัตถุใหม่เข้ากลุ่มใดกลุ่มหนึ่งจากกลุ่มที่มีอยู่ การรวมกลุ่มจะไม่มีตัวแปรตามแต่จะมีแต่ตัวแปรอิสระที่ใช้วัดความคล้ายคลึงหรือใช้ในการคำนวณความคล้ายคลึง เช่น การรวมกลุ่มของลูกค้าที่มีพฤติกรรมการซื้อที่มีลักษณะคล้ายคลึงกัน เทคนิคที่ใช้ เช่น เคมีน (K-Means) เป็นต้น

## 6. การบรรยาย (Description)

การหาคำอธิบายถึงสิ่งที่จะเกิดขึ้น โดยอาศัยข้อมูลจากฐานข้อมูล มักจะเป็นการอธิบายรูปแบบและแนวโน้มที่อยู่ในข้อมูลหรือการสรุปข้อมูล

## 2.2 การจัดกลุ่ม (Clustering) โดยวิธีเคมีนอัลกอริทึม (K-Means Algorithm)

### 2.2.1 การจัดกลุ่ม (Clustering)

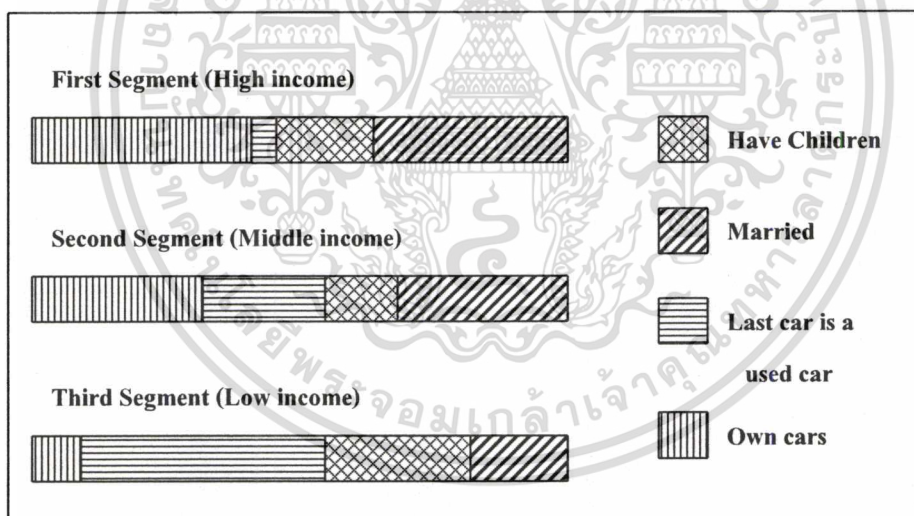
Clustering (Database clustering หรือ Segmentation Cluster) เป็นการแบ่งฐานข้อมูลออกเป็นกลุ่มต่างๆ เป็นเทคนิค ใช้จำแนกหรือแบ่ง เป็นกรณี (Case) เช่น คน, สัตว์, สิ่งของ, หรือองค์กร ฯลฯ หรือแบ่งตัวแปรออกเป็นกลุ่มย่อยๆตั้งแต่ 2 กลุ่มขึ้นไป กรณีที่อยู่กลุ่มเดียวกันจะมีลักษณะที่เหมือนกันหรือคล้ายกัน ส่วนกรณีที่อยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน ดังนั้นการพิจารณาเลือกลักษณะตัวแปรที่จะนำมาใช้แบ่งกลุ่มกรณีจึงมีความสำคัญ นอกจากนั้นกรณีใดกรณีหนึ่งจะอยู่ในกลุ่มหนึ่งเพียงกลุ่มเดียวตัวอย่างเช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ใช้ศึกษาพฤติกรรมกรรมการบริโศขของกลุ่มผู้บริโศขที่อยู่ต่างกลุ่มกัน ทำให้สามารถวางกลยุทธ์ทางการตลาดได้อย่างมีประสิทธิภาพมากขึ้น ตัวแปรที่ใช้ในการแบ่งกลุ่ม เช่น อาชีพ อายุ รายได้
- การแบ่งกลุ่มประเทศด้านสาธารณสุข ซึ่งตัวแปรที่ใช้ เช่น จำนวนแพทย์ จำนวนเภสัชกร จำนวนพยาบาล จำนวนเตียง ประเทศที่มีระบบสาธารณสุขคล้ายกันจะอยู่ด้วยกัน
- บริษัทจำหน่ายรถยนต์ได้แยกกลุ่มลูกค้าออกเป็น 3 กลุ่ม คือ
  1. กลุ่มผู้มีรายได้สูง (>\$80,000)
  2. กลุ่มผู้มีรายได้ปานกลาง (\$25,000 to \$ 80,000)
  3. กลุ่มผู้มีรายได้ต่ำ (less than \$25,000)

และภายในแต่ละกลุ่มยังประกอบด้วย

- Have Children
- Married
- Last car is a used car
- Own cars



รูปที่ 2.1 ตัวอย่างการจัดกลุ่ม (Clustering)

จากข้อมูลข้างต้น ดังแสดงในรูปที่ 2.1 ทำให้ทางบริษัทรู้ว่าเมื่อมีลูกค้าเข้ามาที่บริษัทควร จะเสนอขายรถประเภทใด เช่น ถ้าเป็นกลุ่มผู้มีรายได้สูงควรจะเสนอรถใหม่ เป็นรถครอบครัว ขนาดใหญ่พอสมควร แต่ถ้าเป็นผู้มีรายได้ค่อนข้างต่ำควรเสนอรถมือสอง ขนาดค่อนข้างเล็ก

เทคนิค Clustering แบ่งเป็นหลายประเภทโดยเทคนิค K-Means Clustering เป็นเทคนิค หนึ่งที่ได้รับความนิยม เป็นเทคนิคการจำแนก Case ออกเป็นกลุ่มย่อย โดยจะใช้กับ Case ที่มี เอกสารเป็นเอกสารที่ส่งวนเวียนสำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้า จำนวนมากและต้องกำหนดจำนวนกลุ่ม ตัวแปรที่ใช้จะต้องเป็นตัวแปรเชิงปริมาณคือเป็นสเกล ไม่วากรณีใดๆทั้งสิ้น อีกทั้งห้ามมีเหตุบังเอิญและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

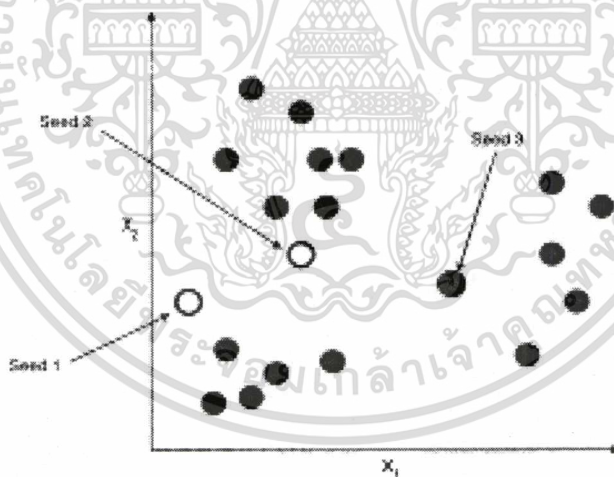
อันตรภาค (Interval Scale) หรือ สเกลอัตราส่วน (Ratio Scale) เท่านั้น ผู้วิเคราะห์จะต้องทำข้อมูลให้เป็นมาตรฐานก่อน

### 2.2.2 เทคนิคการจัดกลุ่มด้วย K-Means

เทคนิคการจัดกลุ่มข้อมูลด้วย K-Means อัลกอริทึมเป็นการแบ่งกลุ่มข้อมูลใดๆ ออกเป็นกลุ่มต่างๆ โดยที่สมาชิกภายในกลุ่มเดียวกันจะมีความคล้ายคลึงกันมากกว่าสมาชิกที่ต่างกลุ่มกัน สมาชิกภายในกลุ่มแต่ละกลุ่มจะมีระยะห่างจากจุดศูนย์กลาง หรือตัวแทนของกลุ่ม (Mean) การทำงานของอัลกอริทึมนี้ขั้นตอนแรกจะทำการสุ่มกำหนดจำนวนกลุ่มเริ่มต้น  $k$  กลุ่ม กำหนดตัวแทนของกลุ่ม พิจารณาความคล้ายคลึงกันของข้อมูลที่เราสนใจ โดยพิจารณาจากระยะทางระหว่างข้อมูลที่เราสนใจกับตัวแทนของกลุ่มนั้น แล้วจึงทำการคำนวณหาค่าเฉลี่ยหรือตัวแทนของกลุ่มใหม่ การทำงานนี้จะทำซ้ำไปเรื่อยๆ จนกระทั่งค่าเฉลี่ยหรือตัวแทนของกลุ่มไม่มีการเปลี่ยนแปลง

#### ■ ตัวอย่างการจัดกลุ่มโดยใช้ K-Means

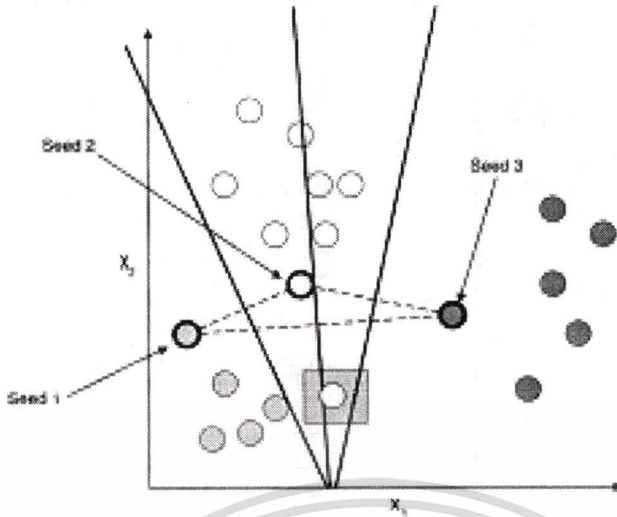
1. สุ่มเลือกตัวแทนกลุ่ม  $k$  ข้อมูล คือกลุ่มเริ่มต้นซึ่งไปยังเมล็ดพันธุ์ที่เป็นตัวแทนของทั้ง 3 กลุ่ม ดังแสดงในรูปที่ 2.2



รูปที่ 2.2 กำหนดที่ตั้งของตัวแทนกลุ่ม

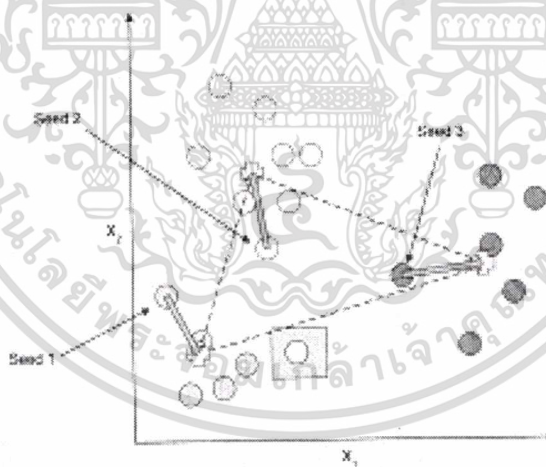
2. กำหนดขอบเขตของเมล็ดพันธุ์ที่คล้ายกันที่สุด โดยพิจารณาทางเรขาคณิตในการแบ่งระยะจากจุดใดๆ เช่น จุด A และ B ระยะห่างที่เท่ากันจาก A และ B จะพบเส้นที่ตั้งฉากจากแกน  $x$  หรือเรียกว่าเส้นแบ่งระหว่างกลุ่ม ดังแสดงในรูปที่ 2.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



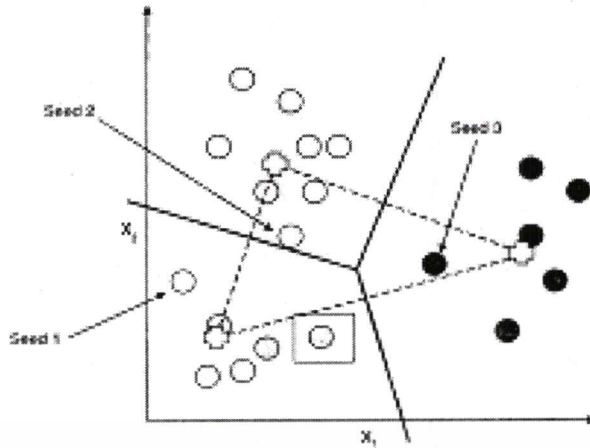
รูปที่ 2.3 กำหนดขอบเขตเริ่มต้นของกลุ่ม

3. เปลี่ยนตำแหน่งเมล็ดพันธุ์ที่เป็นตัวแทนของกลุ่มให้อยู่ในตำแหน่งกึ่งกลางของกลุ่ม ดังแสดงในรูปที่ 2.4



รูปที่ 2.4 คำนวณหาศูนย์กลางของกลุ่ม และย้ายตำแหน่ง

4. ทำการลากเส้นเพื่อกำหนดขอบเขตของเมล็ดพันธุ์ใหม่ แต่ละจุดจะถูกกำหนดกลุ่มที่ใกล้กับศูนย์กลางอีกครั้ง ดังแสดงในรูปที่ 2.5



รูปที่ 2.5 ทำซ้ำทุกกลุ่มที่กำหนดค่าใหม่

■ ขั้นตอนวิธีการจัดกลุ่มโดย K-Means มีดังต่อไปนี้

1. กำหนดค่า  $k$  เป็นค่าเริ่มต้นว่าต้องการแบ่งกลุ่มกี่กลุ่ม และนำข้อมูลใส่ในแต่ละกลุ่ม
2. คำนวณหาตัวแทนของกลุ่มข้อมูลปัจจุบัน (centroids of the clusters)
3. พิจารณาข้อมูลทุกตัวว่าอยู่กลุ่มใดถึงจะเหมาะสม โดยใช้หลักการว่าจุดใกล้กลุ่มใดให้อยู่กลุ่มนั้น ในที่นี้คือมองจุดที่เราสนใจเทียบกับตัวแทนของกลุ่ม โดยใช้สูตร ดังสมการที่ (2.1)

$$d(x_i, m^{(k)}) = \sqrt{(x_i - m^{(k)})^T (x_i - m^{(k)})} \quad (2.1)$$

เมื่อ  $d(x_i, m^{(k)})$  คือระยะทางระหว่างข้อมูลตัวที่  $i$  กับตัวแทนของกลุ่มที่  $K$   
 $x_i$  คือข้อมูลตัวที่  $i$   
 $m^{(k)}$  คือตัวแทนของกลุ่มที่  $K$

4. เมื่อข้อมูลเปลี่ยนกลุ่ม ตัวแทนของกลุ่มจะต้องเปลี่ยนด้วย ทำการคำนวณหาตัวแทนของกลุ่มใหม่ โดยใช้สูตร ดังสมการที่ (2.2)

$$m^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^{(k)} \quad (2.2)$$

เมื่อ  $m^{(k)}$  คือตัวแทนของกลุ่มที่  $K$   
 $n_k$  คือจำนวนข้อมูลทั้งหมดของกลุ่มที่  $K$   
 $x_i^{(k)}$  คือข้อมูลตัวที่  $i$  ในกลุ่มที่  $K$

5. ทำซ้ำข้อ 2 – 5 ไปเรื่อยๆ จนกว่าข้อมูล หรือตัวแทนของกลุ่มไม่มีการเปลี่ยนแปลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ▪ การวัดประสิทธิภาพการจับกลุ่ม

วิธีการวัดประสิทธิภาพโดยทั่วไปแล้วจะใช้หลักของ Square error ( $e_k^2$ )

$$e_k^2 = \sum_{i=1}^{n_k} \left( x_i^{(k)} - m^{(k)} \right)^T \left( x_i^{(k)} - m^{(k)} \right) \quad (2.3)$$

เมื่อ	$e_k^2$	คือค่าการกระจายตัวของข้อมูลกลุ่มที่ K
	$n_k$	คือจำนวนข้อมูลทั้งหมดของกลุ่มที่ K
	$x_i^{(k)}$	คือข้อมูลตัวที่ i ในกลุ่มที่ K
	$m^{(k)}$	คือตัวแทนของกลุ่มที่ K

ค่าการกระจายตัวของข้อมูลภายในกลุ่ม ดังสมการที่ (2.3) ถ้ามีค่าการกระจายมาก ค่า Square error ( $e_k^2$ ) จะมาก ซึ่งแสดงให้เห็นถึงประสิทธิภาพการจับกลุ่มว่ามีมากน้อยเพียงใด หากค่าการกระจายตัวของข้อมูลภายในกลุ่มมีน้อย ย่อมหมายถึงการจับกลุ่มของกลุ่มนั้นดี และใช้แสดงภาพรวมในการแบ่งกลุ่มว่าเป็นการแบ่งกลุ่มที่ดีหรือไม่ โดยใช้สูตร ดังสมการที่ (2.4) ดังนี้

$$E_k^2 = \sum_{i=1}^K e_i^2 \quad (2.4)$$

เมื่อ	$E_k^2$	คือผลรวมของค่าการกระจายตัวของข้อมูล K กลุ่ม
	$e_k^2$	คือค่าการกระจายตัวของข้อมูลกลุ่มที่ i

## 2.3 ไมโครซอฟท์คลัสเตอร์ริงอัลกอริทึม (Microsoft Clustering Algorithm)

### 2.3.1 ไมโครซอฟท์คลัสเตอร์ริง (Microsoft Clustering)

ไมโครซอฟท์คลัสเตอร์ริงเป็นอัลกอริทึมในการค้นหาหากลุ่มตามธรรมชาติภายในข้อมูล ซึ่งยังไม่ทราบกลุ่มที่ชัดเจน หรือเป็นการค้นหาตัวแปรที่ซ่อนอยู่ที่บ่งบอกถึงประเภทข้อมูล โดยส่วนใหญ่นิยมใช้เทคนิคการจับกลุ่มสำหรับงานด้านการตลาด ซึ่งสามารถใช้เทคนิคการจับกลุ่มเพื่อศึกษาข้อมูลเกี่ยวกับกลุ่มลูกค้าเพื่อกำหนดเป้าหมายเฉพาะในแต่ละกลุ่มได้ ความสามารถในการกำหนดและระบุกลุ่มของลูกค้าดังกล่าวจะช่วยให้การดำเนินธุรกิจมีประสิทธิภาพมากยิ่งขึ้น โดยไมโครซอฟท์ได้เสนอวิธีในการจับกลุ่ม 2 วิธี ดังนี้

1. Expectation Maximization (EM)
2. K-means algorithm

## EM Clustering

เป็นอัลกอริทึมที่ใช้วิธีกำหนดกลุ่มเริ่มต้นในแบบจำลองต่างๆจนกว่าจะเหมาะกับข้อมูล และใช้ความน่าจะเป็นในการตัดสินใจว่าจุดของข้อมูลควรอยู่ในกลุ่มใด ในการประมวลผลจะหยุดเมื่อความน่าจะเป็นของแบบจำลองเหมาะสมกับข้อมูล โดยใช้ฟังก์ชันในการตัดสินใจจากความเป็นไปได้ของข้อมูลที่ได้จากแบบจำลอง

EM algorithm เป็นวิธีที่กำหนดให้เป็นมาตรฐาน เนื่องจากมีข้อดีหลายประการเมื่อเปรียบเทียบกับวิธี K-means algorithm ดังนี้

- ต้องมีการสแกนฐานข้อมูลมากที่สุดหนึ่งครั้ง
- สามารถทำงานได้โดยที่หน่วยความจำ (แรม) จำกัด
- สามารถที่จะเลื่อนเคอร์เซอร์ไปข้างหน้าได้เท่านั้น
- เป็นวิธีการสุ่มที่ให้ผลที่ดีกว่า

ไมโครซอฟท์ได้นำเสนอ EM algorithm 2 วิธี ซึ่งเป็นเทคนิคที่มักใช้กับฐานข้อมูลขนาดใหญ่ ดังนี้

1. Scalable EM ซึ่งเป็นวิธีมาตรฐานจะใช้ข้อมูล 50,000 ระเบียบแรกเพื่อเริ่มสแกน เมื่อดำเนินการสำเร็จ แบบจำลองจะใช้ข้อมูลเพียงเท่านั้น กรณีไม่สำเร็จจากการใช้ข้อมูล 50,000 ระเบียบแรก จะทำการเพิ่มข้อมูลอีกทีละ 50,000 ระเบียบ วิธีนี้จะทำการประมวลผลในหน่วยความจำการทำซ้ำระหว่างข้อมูลจะเร็วกว่าวิธี Non-scalable EM

2. Non-scalable EM เป็นวิธีที่จะอ่านข้อมูล Dataset ทั้งหมดโดยไม่คำนึงถึงของขนาด วิธีนี้จะสร้างกลุ่มซึ่งมีความถูกต้องมากกว่าแต่ทำให้ต้องใช้หน่วยความจำจำนวนมาก

## K-Means Clustering

เป็นวิธีการรวมกลุ่มวัตถุที่มีคุณลักษณะที่คล้ายคลึงกัน ให้รวมอยู่ในกลุ่มเดียวกัน ในการวัดความคล้ายคลึงสิ่งสำคัญที่ต้องพิจารณานั้นได้แก่ ชนิดของตัวแปร (ต่อเนื่อง, ไม่ต่อเนื่อง) ซึ่ง K-Means จะใช้ชนิดตัวแปรแบบต่อเนื่องในการจัดกลุ่ม โดยการคำนวณระยะทาง กรณีชนิดของตัวแปรแบบไม่ต่อเนื่อง ไมโครซอฟท์ได้ดัดแปลงให้ K-Means สามารถจัดกลุ่มข้อมูลชนิดตัวแปรแบบไม่ต่อเนื่องได้ โดยระยะห่างของข้อมูลจากกลุ่มจะคำนวณได้จาก  $1 - P(\text{data point, cluster})$

ไมโครซอฟท์ได้นำเสนอ K-Means algorithm 2 วิธี ดังนี้

1. Scalable K-Means ซึ่งเป็นวิธีมาตรฐานจะใช้ข้อมูล 50,000 ระเบียบแรก แล้วจึงอ่านข้อมูลเพิ่มเมื่อต้องการ

2. Non-scalable K-Means เป็นวิธีที่จะอ่านข้อมูล Dataset ทั้งหมดและสร้างเป็น 1 กลุ่มให้ผ่าน

### 2.3.2 พารามิเตอร์ในไมโครซอฟท์คลัสเตอร์ริงอัลกอริทึม

พารามิเตอร์ที่ใช้ในการทำคาค่าไมนิ่ง โดย Microsoft Clustering Algorithm มีพารามิเตอร์ดังแสดงในตารางที่ 2.2 ดังนี้

ตารางที่ 2.2 พารามิเตอร์ต่างๆที่ใช้ในแบบจำลองระบบ

พารามิเตอร์	ค่าที่ระบุ	ค่ามาตรฐาน	ค่าพารามิเตอร์ที่รองรับ
CLUSTERING METHOD	3	1	1,2,3,4
CLUSTER COUNT	0	10	[0,...)
CLUSTER SEED		0	[0,...)
MINIMUM SUPPORT		1	[0,...)
MODELLING CARDINALITY		10	[1,50]
STOPPING TOLERANCE		10	[0,...)
SAMPLE SIZE		50000	0, [100,...)
MAXIMUM INPUT ATTRIBUTES		255	[0,65535]
MAXIMUM STATES		100	0, [2,65535]

CLUSTERING METHOD คือ พารามิเตอร์ที่ใช้ระบุวิธี Clustering ที่จะใช้อัลกอริทึมใด โดยมีให้เลือก ดังนี้

- 1 Scalable EM
- 2 Non-scalable EM
- 3 Scalable K-Means
- 4 Non-scalable K-Means

โดยค่ามาตรฐานที่ระบบกำหนดไว้คือ 1

CLUSTER COUNT คือ พารามิเตอร์ที่กำหนดจำนวนของกลุ่มเริ่มต้นเพื่อจัดกลุ่มข้อมูลที่ใกล้เคียงกัน กรณีอัลกอริทึมไม่สามารถจัดกลุ่มตามจำนวนของกลุ่มเริ่มต้นที่กำหนดไว้ อัลกอริทึมอาจจัดกลุ่มขึ้นมากกว่าหรือน้อยกว่าที่กำหนด การกำหนดค่าพารามิเตอร์ CLUSTER\_COUNT เท่ากับ 0 จะทำให้ช่วยค้นหาจำนวนกลุ่มที่ดีที่สุดที่จะนำมาจัดกลุ่ม โดยค่ามาตรฐานที่ระบบกำหนดไว้คือ 10

CLUSTER SEED คือ พารามิเตอร์ที่ใช้ในการสุ่มเพื่อสร้างกลุ่มเริ่มต้นสำหรับการสร้างแบบจำลองในตอนแรก การเปลี่ยนพารามิเตอร์นี้จะสามารถเปลี่ยนวิธีการเริ่มต้นสร้างกลุ่มได้ โดยค่ามาตรฐานที่ระบบกำหนดไว้คือ 0

MINIMUM SUPPORT คือ พารามิเตอร์ที่ใช้ระบุจำนวนข้อมูลที่ต่ำที่สุดเพื่อใช้สร้างกลุ่มกรณีจำนวนกลุ่มข้อมูลต่ำกว่าจำนวนพารามิเตอร์นี้ กลุ่มข้อมูลที่ได้จะไม่มีความและจะถูกล้าง โดยมีความมาตรฐานที่ระบบกำหนดไว้คือ 1

MODELLING CARDINALITY คือ พารามิเตอร์ที่ระบุจำนวนแบบจำลองตัวอย่างขณะที่มีการสร้างแบบจำลอง ซึ่งการลดค่าพารามิเตอร์นี้จะช่วยปรับปรุงประสิทธิภาพที่เสี่ยงต่อการสูญหายของบางแบบจำลองได้ โดยมีความมาตรฐานที่ระบบกำหนดไว้คือ 10

STOPPING TOLERANCE คือ พารามิเตอร์ที่เป็นตัวบอกที่จุดเสร็จสิ้นในการสร้างแบบจำลอง โดยเมื่อรวมการเปลี่ยนแปลงของกลุ่มน่าจะน้อยกว่าอัตราส่วนของค่าพารามิเตอร์หารด้วยขนาดของแบบจำลอง โดยมีความมาตรฐานที่ระบบกำหนดไว้คือ 10

SAMPLE SIZE คือ พารามิเตอร์ที่ระบุจำนวนข้อมูลที่ใช้อัลกอริทึมใช้ในแต่ละครั้ง กรณีพารามิเตอร์ CLUSTERING\_METHOD กำหนดเป็นแบบที่ปรับขนาดได้ การกำหนดค่าพารามิเตอร์นี้เป็น 0 จะทำให้ข้อมูลทั้งกลุ่มถูกโหลดเข้าสู่หน่วยความจำในครั้งเดียว ซึ่งจะทำให้เกิดปัญหาเรื่องประสิทธิภาพการทำงานและหน่วยความจำได้ โดยมีความมาตรฐานที่ระบบกำหนดไว้คือ 50000

MAXIMUM INPUT ATTRIBUTES คือ พารามิเตอร์ที่ระบุจำนวนสูงสุดของ input attributes ที่อัลกอริทึมสามารถจัดการได้ กรณีกำหนดเป็น 0 หมายถึงไม่มีจำนวน input attributes สูงสุด ซึ่งการเพิ่มจำนวน input attributes จะทำให้ลดประสิทธิภาพ โดยมีความมาตรฐานที่ระบบกำหนดไว้คือ 255

MAXIMUM STATES คือ พารามิเตอร์ที่ระบุจำนวนสูงสุดของ attribute states ที่อัลกอริทึมรองรับ ถ้า attribute มี states มากกว่าค่าสูงสุดอัลกอริทึมจะใช้ states ที่ทำงานอยู่และ ignores states เหลืออยู่ โดยมีความมาตรฐานที่ระบบกำหนดไว้คือ 100

### 2.3.3 การวัดประสิทธิภาพการจัดกลุ่ม

การวัดประสิทธิภาพการจัดกลุ่มสามารถดูได้จากคอลัมน์ MSOLAP\_NODE\_SCORE จาก Model Content ซึ่งเป็นคะแนนที่ได้จากค่าเฉลี่ยของ case likelihood ในแต่ละ cases ตัวอย่าง Query ในการแสดงค่า MSOLAP\_NODE\_SCORE

```
SELECT TOP 1 MSOLAP_NODE_SCORE FROM CustomerClusters.CONTENT
```

คะแนนที่ได้จะเป็นตัวเลขระหว่าง 0 และ 1 คะแนนที่สูงขึ้นจะดีและใช้วัดคุณภาพของการจัดกลุ่มที่ได้

## บทที่ 3

### การวิเคราะห์และออกแบบระบบงาน

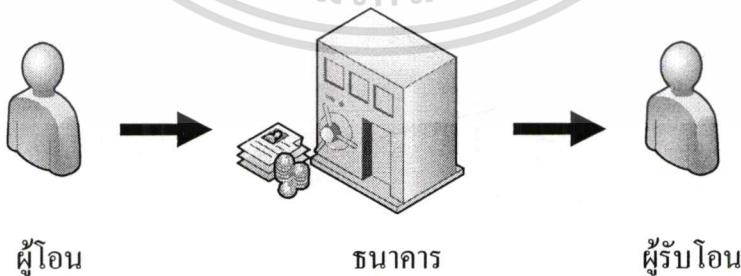
ในบทนี้จะกล่าวถึงขั้นตอนกระบวนการต่างๆ ในการวิเคราะห์และออกแบบระบบงานของโครงการพัฒนาระบบการจัดกลุ่มพฤติกรรมการใช้บริการระบบโอนเงินอัตโนมัติ โดยจะกล่าวถึงลักษณะการทำงานของระบบงานปัจจุบัน, ปัญหาที่พบในระบบงานปัจจุบัน และการศึกษาความต้องการของระบบงานใหม่

#### 3.1 ลักษณะการทำงานของระบบงานปัจจุบัน

ระบบโอนเงินอัตโนมัติ (Automatic Transfer System) หรือ ATS เป็นการโอนเงินรายย่อย ซึ่งมีงวดชำระที่แน่นอนและมีปริมาณรายการจำนวนมากให้แก่ลูกค้าที่มีข้อตกลงล่วงหน้า โดยผู้โอนและผู้รับโอนจะต้องมีบัญชีเงินฝากกับธนาคาร ซึ่งแบ่งเป็น 2 ประเภท คือ บริการนำเงินเข้าบัญชีโดยระบบอัตโนมัติ (Credit Fund Transfer) และบริการหักบัญชีโดยระบบอัตโนมัติ (Debit Fund Transfer)

##### 3.1.1 การบริการนำเงินเข้าบัญชีโดยระบบอัตโนมัติ (Credit Fund Transfer)

เป็นบริการโอนเงินอัตโนมัติเพื่อชำระค่าสินค้าหรือบริการ รวมถึงการจ่ายเงินเดือน โบนัส เงินปันผล ฯลฯ ให้กับผู้รับเงินจำนวนมากรายที่มีบัญชีออมทรัพย์หรือกระแสรายวัน กับสาขาของธนาคารทั่วประเทศ ซึ่งจะตอบสนองความต้องการของลูกค้าในการทำรายการโอนเงินที่มีจำนวนรายการมากๆ ด้วยการส่งคำสั่งครั้งเดียว (ในรูปแบบ One To Many) ดังแสดงในรูปที่ 3.1

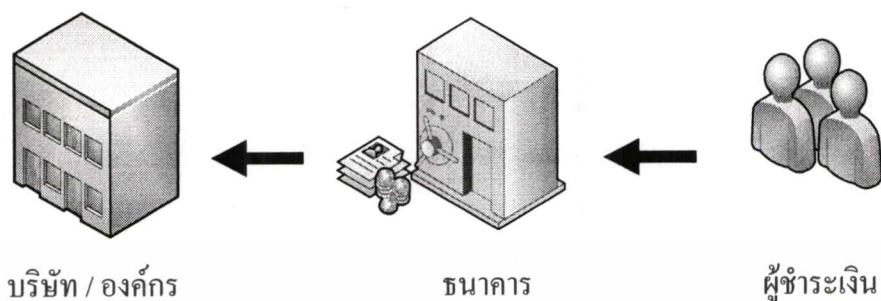


รูปที่ 3.1 การบริการนำเงินเข้าบัญชีโดยระบบอัตโนมัติ

##### 3.1.2 การบริการหักบัญชีโดยระบบอัตโนมัติ (Debit Fund Transfer)

เป็นการบริการรับชำระเงินค่าสินค้าและบริการ ที่ผู้ขายหรือผู้ให้บริการสามารถเรียกเก็บเงินจากผู้ชำระเงินจำนวนมากราย โดยการหักเงินจากบัญชีออมทรัพย์หรือกระแสรายวันของผู้ชำระ

ชำระเงินที่ฝากไว้กับธนาคารทุกสาขาทั่วประเทศซึ่งจะตอบสนองต่อความต้องการลูกค้าในการทำรายการหักบัญชีที่มีจำนวนรายการมากๆ ด้วยการส่งคำสั่งเพียงครั้งเดียว (ในรูปแบบ Many To One) ดังแสดงในรูปที่ 3.2



รูปที่ 3.2 การบริการหักบัญชีโดยระบบอัตโนมัติ

จากการบริการทั้ง 2 ประเภทอาจมองเป็นรูปแบบของการรับชำระหรือจ่ายเงินผ่านธนาคารก็ได้ โดยมีข้อตกลงและการยินยอมจากทั้ง 3 ฝ่ายคือ 1.ลูกค้า 2.ธนาคาร 3.บริษัท / องค์กร ซึ่งธนาคารจะเป็นผู้ดำเนินการ โอนเงินระหว่างบัญชีลูกค้า และบริษัท / องค์กร และได้รับค่าธรรมเนียมเป็นค่าดำเนินการ

### 3.2 ปัญหาที่พบในระบบงานปัจจุบัน

ระบบงานในปัจจุบันพบปัญหาในการทำงานที่ต้องปรับปรุง ดังนี้

1. ประเภทการให้บริการ ในปัจจุบันยังไม่ครอบคลุมต่อความต้องการของผู้ใช้บริการ
2. ธนาคารประสบปัญหาที่ผู้ใช้บริการหันไปใช้บริการ โอนเงินจากธนาคารอื่น ทำให้ธนาคารเสียหายได้
3. ระบบปัจจุบันเป็นระบบที่ใช้ในการทำรายการ โอนเงิน ยังขาดส่วนที่จะช่วยให้ผู้บริหารนำมาใช้เป็นแนวทางในการตัดสินใจ หรือวางแผนทางการตลาด

### 3.3 การศึกษาความต้องการของระบบงานใหม่

จากการศึกษาระบบ สามารถสรุปความต้องการของระบบงานใหม่ได้ ดังนี้

1. ผู้ใช้ระบบงานสามารถสร้างและการจัดการแบบจำลองค่าใดมั่งหนึ่ง เพื่อใช้ในการวิเคราะห์ข้อมูลได้
2. ระบบงานสามารถแสดงผลการจัดกลุ่มพฤติกรรมการใช้บริการระบบโอนเงิน

เอกสออัตโนมัติได้สารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.4 การเตรียมข้อมูล

การวิเคราะห์จะนำข้อมูลการใช้บริการระบบโอนเงินอัตโนมัติ (Automatic Transfer System) หรือ ATS ของธนาคารพาณิชย์แห่งหนึ่งในประเทศไทยช่วงวันที่ 23 มกราคม พ.ศ. 2550 ถึงวันที่ 17 มีนาคม พ.ศ. 2552 จากการศึกษาในฐานข้อมูลของระบบ พบว่ามี 7 ตารางที่นำมาใช้ในการวิเคราะห์ ดังนี้

1. ตาราง ATS\_COMPANY ประกอบด้วย 13 ฟیلด์ ใช้เก็บข้อมูลบริษัท
2. ตาราง ATS\_SENDING\_TYPE ประกอบด้วย 2 ฟیلด์ ใช้เก็บข้อมูลช่องทางการส่งข้อมูล
3. ตาราง ATS\_TRANSFER\_TYPE ประกอบด้วย 2 ฟیلด์ ใช้เก็บข้อมูลประเภทการทำรายการ
4. ตาราง ATS\_SERVICE\_TYPE ประกอบด้วย 2 ฟیلด์ ใช้เก็บข้อมูลประเภทการบริการ
5. ตาราง ATS\_BUSINESS\_GROUP ประกอบด้วย 2 ฟیلด์ ใช้เก็บข้อมูลประเภทกลุ่มธุรกิจ
6. ตาราง ATS\_FEE\_RATE ประกอบด้วย 14 ฟیلด์ ใช้เก็บข้อมูลค่าธรรมเนียม
7. ตาราง ATS\_HEADER ประกอบด้วย 16 ฟیلด์

เมื่อตรวจสอบข้อมูลแล้วจะมีบางฟیلด์ที่ไม่ได้เก็บข้อมูล หรือไม่จำเป็นต่อการวิเคราะห์ข้อมูล ดังนั้นทำการสร้างตาราง COMPANY เพื่อจัดเก็บฟیلด์ข้อมูลที่จะใช้ในการวิเคราะห์ข้อมูลได้ ดังนี้

1. ข้อมูลจากตาราง ATS\_COMPANY ประกอบด้วย 8 ฟیلด์ ดังนี้  
COMP\_CODE, COMP\_NAME, BEGIN\_DATE, COMP\_ACCT, SEND\_TYPE, TRANSFER\_TYPE, SERVICE\_TYPE และ BUSINESS\_GROUP
2. ข้อมูลจากตาราง ATS\_FEE\_RATE 4 ฟیلด์ ดังนี้  
MIN\_RATE, COMP\_FEE, REGION\_FEE และ WAIVE\_MIN\_RATE
3. ข้อมูลจากตาราง ATS\_HEADER 3 ฟیلด์ ดังนี้  
TOTAL\_ITEM, TOTAL\_AMOUNT และ TOTAL\_FEE
4. ข้อมูลจากการสกัดข้อมูล (Extract Data) จากตาราง ATS\_HEADER 2 ฟیلด์ ดังนี้  
TOTAL\_TIME และ NUM\_MONTHS\_LAST\_SEND

ข้อมูลที่ได้จะเป็นข้อมูลสรุปยอดรวมของแต่ละบริษัท โดยจะใช้คำสั่ง SQL ในการทำงาน จะมีการแปลงข้อมูลให้มีความสอดคล้องกับโครงสร้างของตาราง COMPANY ที่ออกแบบไว้ พร้อมทั้งตรวจสอบความถูกต้องและแก้ไขข้อมูลที่จะเก็บลงตารางข้อมูลให้มีความถูกต้อง โดยสามารถอธิบายรายละเอียดค่าข้อมูลที่เป็นไปได้ดังแสดงในตารางที่ 3.1 – 3.7

ตารางที่ 3.1 แสดงค่าข้อมูลสถานะการเก็บค่าธรรมเนียมขั้นต่ำที่เป็นไปได้

ค่าข้อมูลของ WAIVE_MIN_RATE	คำอธิบาย
Y	ไม่เก็บค่าธรรมเนียมขั้นต่ำ
N	เก็บค่าธรรมเนียมขั้นต่ำ

ตารางที่ 3.2 แสดงค่าข้อมูลประเภทการทำรายการที่เป็นไปได้

ค่าข้อมูลของ SEND_TYPE	คำอธิบาย
01	Media
02	Internet
03	Modem
04	FTP Internal

ตารางที่ 3.3 แสดงค่าข้อมูลประเภทการทำรายการที่เป็นไปได้

ค่าข้อมูลของ TRANSFER_TYPE	คำอธิบาย
00	Credit
01	Debit

ตารางที่ 3.4 แสดงค่าข้อมูลประเภทการบริการที่เป็นไปได้

ค่าข้อมูลของ SERVICE_TYPE	คำอธิบาย
01	เงินเดือน ค่าจ้าง บำเหน็จ บำนาญ
02	ค่าสินค้า บริการ
03	ขายหลักทรัพย์
04	เงินกู้
05	ค่าสาธารณูปโภค (ค่าน้ำ, ค่าไฟ เป็นต้น)
06	ค่าเบี้ยประกันภัย ประกันชีวิต
07	ซื้อหลักทรัพย์
08	ค่าผ่อนชำระ
99	อื่นๆ

ตารางที่ 3.5 แสดงค่าข้อมูลประเภทกลุ่มธุรกิจที่เป็นไปได้

ค่าข้อมูลของ BUSINESS_GROUP	คำอธิบาย
01	เกษตรกรรม การล่าสัตว์ และการป่าไม้
02	การประมง
03	การทำเหมืองแร่และเหมืองหิน
04	การผลิต
05	การไฟฟ้า แก๊ส และการประปา
06	การก่อสร้าง
07	การขายส่ง การขายปลีก การซ่อมแซมยานยนต์ จักรยานยนต์ ของใช้ส่วนบุคคลและของใช้ในครัวเรือน
08	โรงแรมและภัตตาคาร
09	การขนส่ง สถานที่เก็บสินค้าและการคมนาคม
10	ตัวกลางทางการเงิน
11	บริการด้านอสังหาริมทรัพย์ การให้เช่าและบริการทางธุรกิจ
12	การบริหารราชการและการป้องกันประเทศ รวมทั้งการประกันสังคมภาคบังคับ
13	การศึกษา
14	การบริการด้านสุขภาพและงานสังคมสงเคราะห์
15	การให้บริการชุมชน สังคม และบริการส่วนบุคคลอื่นๆ
16	ลูกจ้างในครัวเรือนส่วนบุคคล
17	องค์การระหว่างประเทศและองค์การต่างประเทศอื่นๆ และสมาชิก

ตารางที่ 3.6 แสดงค่าข้อมูลความถี่ (ครั้ง) ที่ทำรายการย้อนหลัง 3 เดือนที่เป็นไปได้

ค่าข้อมูลของ TOTAL_TIME	คำอธิบาย
<20	น้อยกว่า 20 ครั้ง
20-60	20 ครั้ง ถึง 60 ครั้ง
>60	มากกว่า 60 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.7 แสดงค่าข้อมูลระยะเวลาล่าสุด (เดือน) ที่ทำรายการที่เป็นไปได้

ค่าข้อมูลของ NUM_MONTHS_LAST_SEND	คำอธิบาย
<1	น้อยกว่า 1 เดือน
1-3	1 เดือน ถึง 3 เดือน
>3	มากกว่า 3 เดือน

ตารางที่ 3.8 แสดงข้อมูลตาราง COMPANY ที่นำมาทำดาต้าไมนิ่ง

ลำดับ	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย
1	COMP_CODE	Character	รหัสบริษัท
2	COMP_NAME	Character	ชื่อบริษัท
3	BEGIN_DATE	Date	วันที่เริ่มใช้สัญญา
4	COMP_ACCT	Character	เลขที่บัญชี
5	SEND_TYPE	Character	ช่องทางการส่งข้อมูล
6	TRANSFER_TYPE	Character	ประเภทการทำรายการ
7	SERVICE_TYPE	Character	ประเภทการบริการ
8	BUSINESS_GROUP	Character	ประเภทกลุ่มธุรกิจ
9	MIN_RATE	Number	ค่าธรรมเนียมขั้นต่ำ
10	COMP_FEE	Number	ค่าธรรมเนียมที่ในเขต
11	REGION_FEE	Number	ค่าธรรมเนียมที่อยู่นอกเขต
12	WAIVE_MIN_RATE	Character	สถานะการเก็บค่าธรรมเนียมขั้นต่ำ
13	TOTAL_TIME	Character	ความถี่ (จำนวนครั้ง) ที่ทำรายการ ย้อนหลัง 3 เดือน
14	NUM_MONTHS_LAST_SEND	Character	ระยะเวลาล่าสุด (เดือน) ที่ทำรายการ
15	TOTAL_ITEM	Number	จำนวนรายการทั้งหมด
16	TOTAL_AMOUNT	Number	จำนวนเงินรวมทั้งหมด
17	TOTAL_FEE	Number	จำนวนเงินค่าธรรมเนียมรวม ทั้งหมด

เมื่อจัดเตรียมข้อมูลเรียบร้อยแล้ว สามารถสรุปฟิลด์ที่จะใช้ในการวิเคราะห์ข้อมูลทั้งหมด

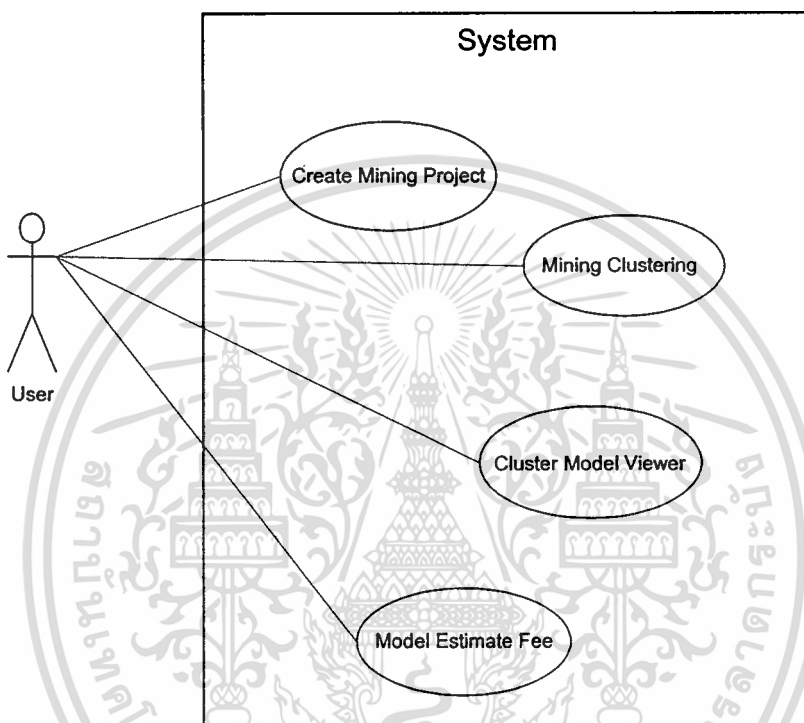
17 ฟิลด์ คือ COMP\_CODE, COMP\_NAME, BEGIN\_DATE, COMP\_ACCT, SEND\_TYPE, TRANSFER\_TYPE, SERVICE\_TYPE, BUSINESS\_GROUP, MIN\_RATE, COMP\_FEE,

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษานี้เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

REGION\_FEE, WAIVE\_MIN\_RATE, TOTAL\_TIME, NUM\_MONTHS\_LAST\_SEND, TOTAL\_ITEM, TOTAL\_AMOUNT และ TOTAL\_FEE ดังแสดงในตารางที่ 3.8

### 3.5 ยูสเคสไดอะแกรม (Use-case Diagram)

เมื่อทำการวิเคราะห์และออกแบบระบบงานสามารถเขียนเป็นยูสเคสไดอะแกรมได้ ดังนี้



รูปที่ 3.3 ยูสเคสไดอะแกรมของระบบการจัดการกลุ่มพฤติกรรมการใช้บริการ โอนเงินอัตโนมัติ

1. แอกเตอร์ (Actor) แอกเตอร์เป็นการแสดงถึงบุคคลที่เกี่ยวข้องกับระบบ ดังนี้
  - 1.1 User เป็นผู้ดูแลและจัดการระบบ
2. ยูสเคส (Use case) ยูสเคส จะอธิบายถึงหน้าที่และการทำงานของระบบทั้งหมด ว่ามีรายละเอียดการทำงานอย่างไร

2.1 Create Mining Project เป็นส่วนที่สร้างแบบจำลองค้ำไมนิ่ง โดยทำการระบุชื่อ Server และรายละเอียดในการสร้างแบบจำลอง

2.2 Mining Clustering เป็นส่วนที่ประมวลผลแบบจำลองที่สร้างขึ้น โดยสามารถปรับเปลี่ยนพารามิเตอร์ที่ใช้ในการประมวลผลได้

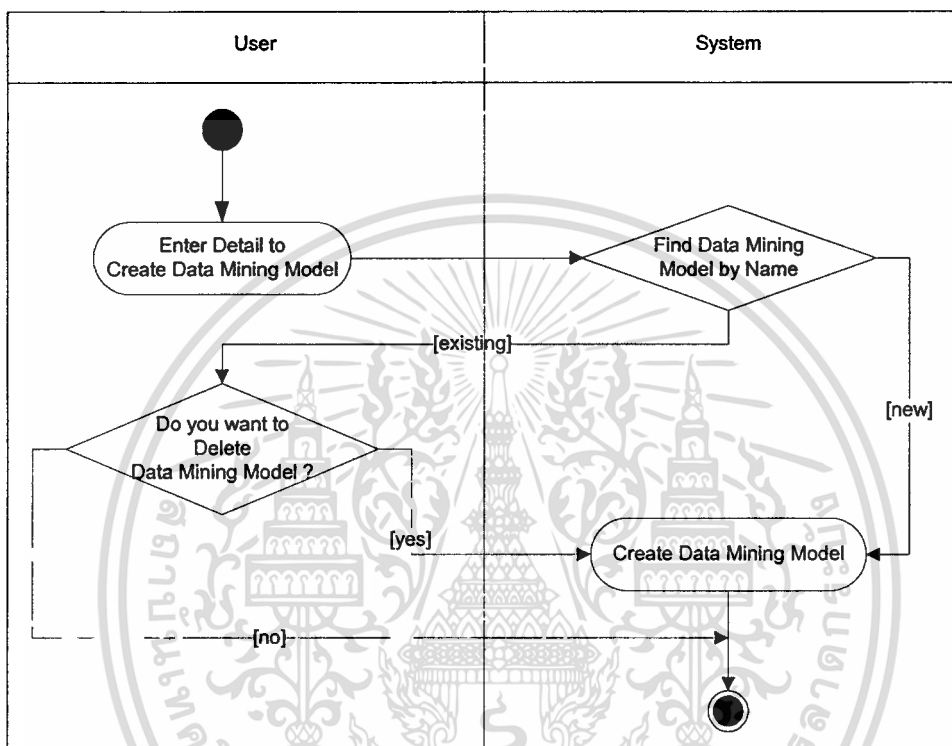
2.3 Cluster Model Viewer เป็นส่วนที่ใช้แสดงผลการจัดกลุ่ม

2.4 Model Estimate Fee เป็นส่วนที่ใช้ประมาณค่าธรรมเนียมเบื้องต้นในการเสนอต่อ

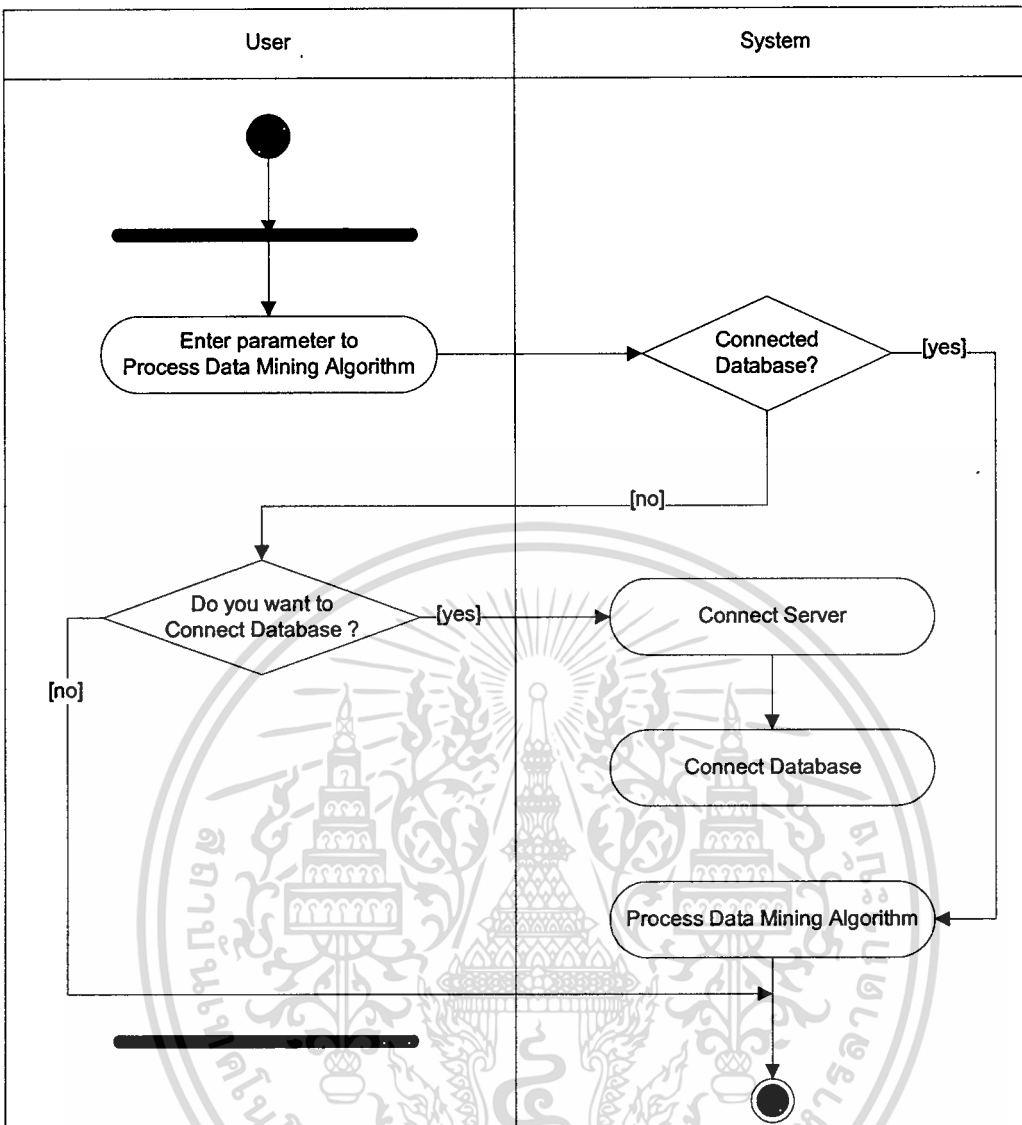
เอกสารนี้เป็นเอกสารลิขสิทธิ์สงวนลิขสิทธิ์ของสถาบันวิจัยและพัฒนาปัญญาประดิษฐ์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
ไม่ว่ากรณีใดๆ ผู้ถูกค้า อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.6 ลักษณะการทำงานของระบบ (Activity Diagram)

ระบบการจัดกลุ่มเป็นแอปพลิเคชันที่เชื่อมต่อกับ Analysis Server เพื่อใช้ในการประมวลผลแบบจำลอง โดยขั้นตอนการทำงานนั้นสามารถอธิบายด้วยแผนภาพแอกทิวิตีไดอะแกรมได้ ดังแสดงในรูปที่ 3.4

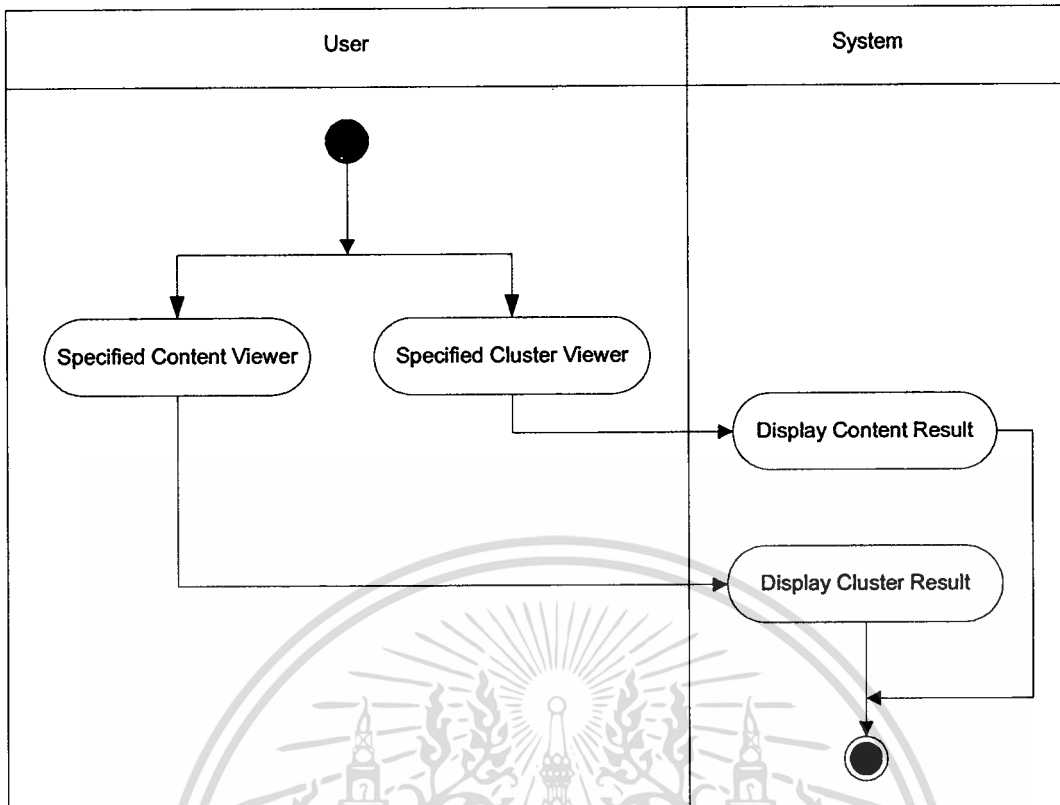


รูปที่ 3.4 แอกทิวิตีไดอะแกรมการ Create Data Mining Model

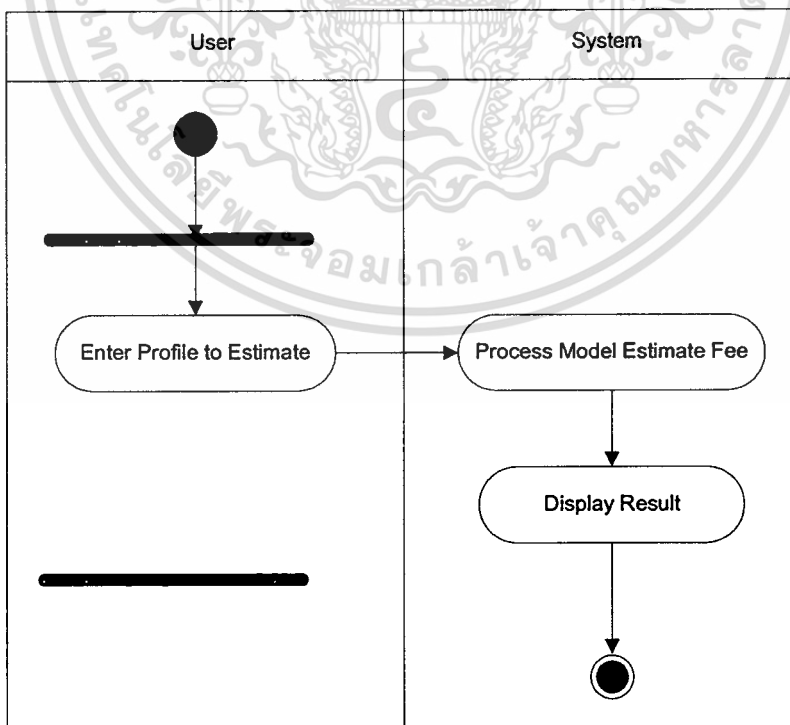


รูปที่ 3.5 แยกทิวทัศน์ไดอะแกรมการ Process Data Mining Model

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 แยกทิวทัศน์โคอะแกรมการ View Data Mining Model

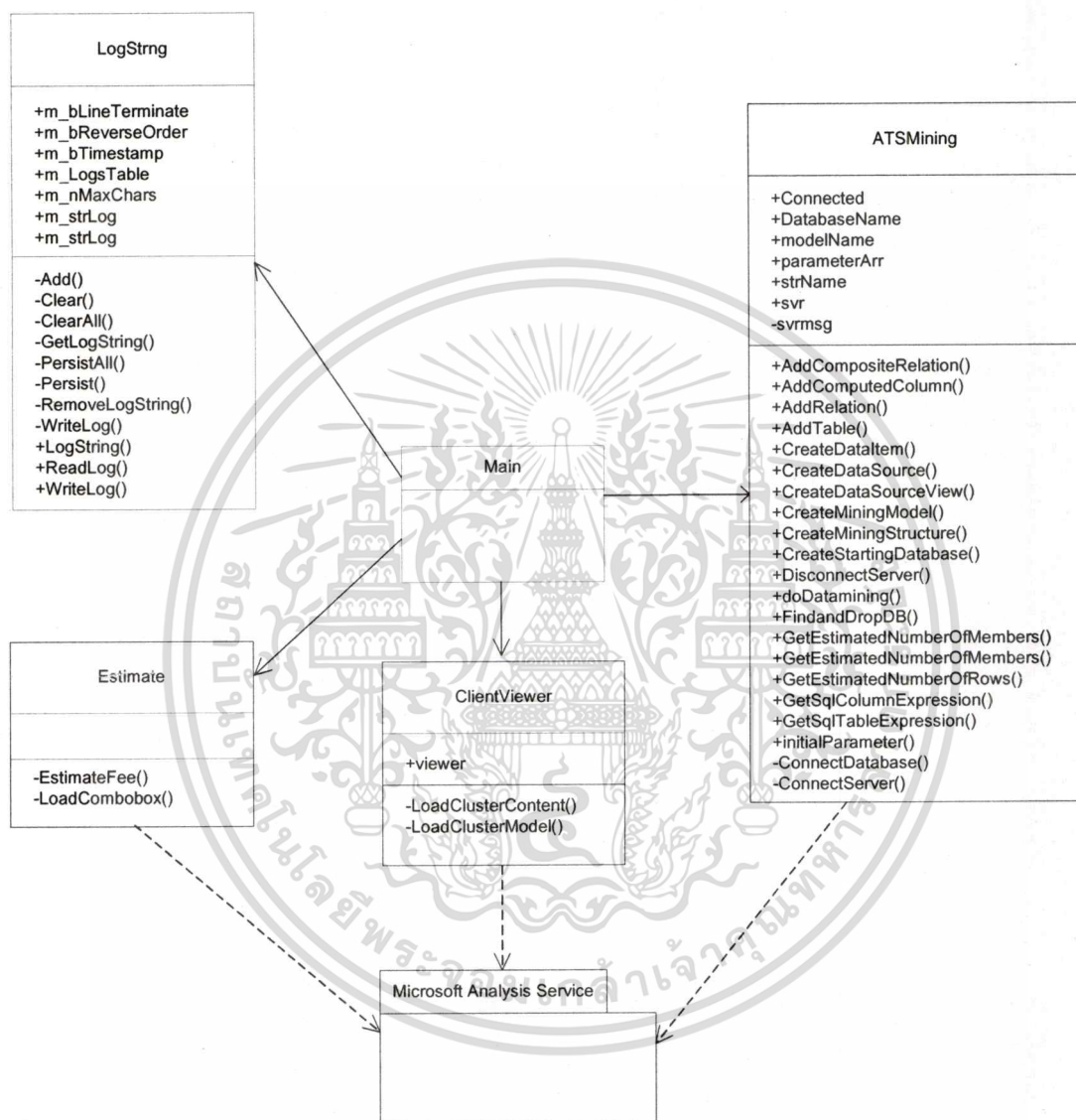


รูปที่ 3.7 แยกทิวทัศน์โคอะแกรมการ Estimate Fee

เอกสารนี้เป็นเอกสารที่สงวนไว้เพื่อการใช้งานเฉพาะที่ขอเท่านั้น มิได้อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.7 คลาสไดอะแกรม (Class Diagram)

จากข้อมูลที่เราได้ทำการวิเคราะห์และออกแบบระบบ เราสามารถแบ่งส่วนย่อยตามหน้าที่การทำงาน นำมาสร้างเป็นคลาสไดอะแกรมได้ ดังแสดงในรูปที่ 3.8 ดังนี้

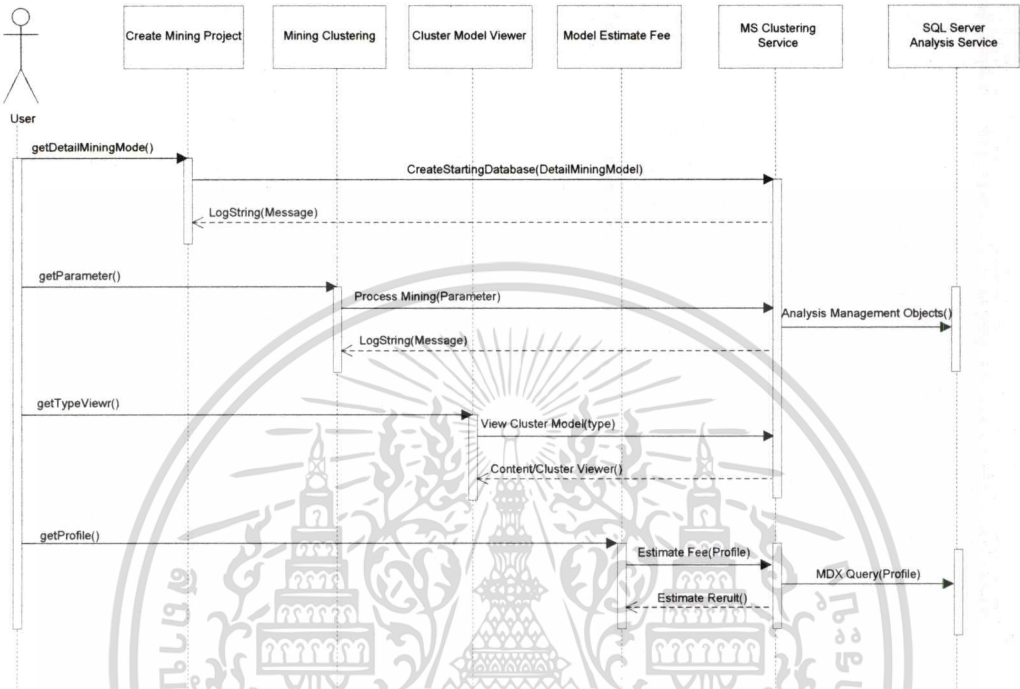


รูปที่ 3.8 แสดงคลาสไดอะแกรมของระบบการจัดการกลุ่มพฤติกรรมบริการใช้บริการ โอนเงินอัตโนมัติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.8 ซีเควนซ์ไดอะแกรม (Sequence Diagram)

แสดงแบบจำลองกิจกรรมเพื่อใช้เข้าใจการทำงานของระบบโดยง่าย ดังแสดงในรูปที่ 3.9  
ดังนี้



รูปที่ 3.9 ซีเควนซ์ไดอะแกรมการจัดการจัดกลุ่มพฤติกรรมบริการออนไลน์อัตโนมัติ

## บทที่ 4

# การพัฒนาระบบการจัดกลุ่มพฤติกรรมการใช้บริการ ระบบโอนเงินอัตโนมัติ

ในบทนี้จะกล่าวถึงส่วนประกอบหลักของโปรแกรมที่พัฒนาขึ้น โดยประกอบด้วยส่วน 4 ส่วนหลัก ได้แก่ การสร้างแบบจำลองการจัดกลุ่ม การ Mining ข้อมูล การแสดงผลการจัดกลุ่ม และ ส่วนการพยากรณ์จากแบบจำลอง

### 4.1 การสร้างแบบจำลองการจัดกลุ่ม

ฟังก์ชันการทำงานในส่วนนี้ประกอบด้วยการทำงานหลัก ดังนี้

#### 1. Connect Analysis Server

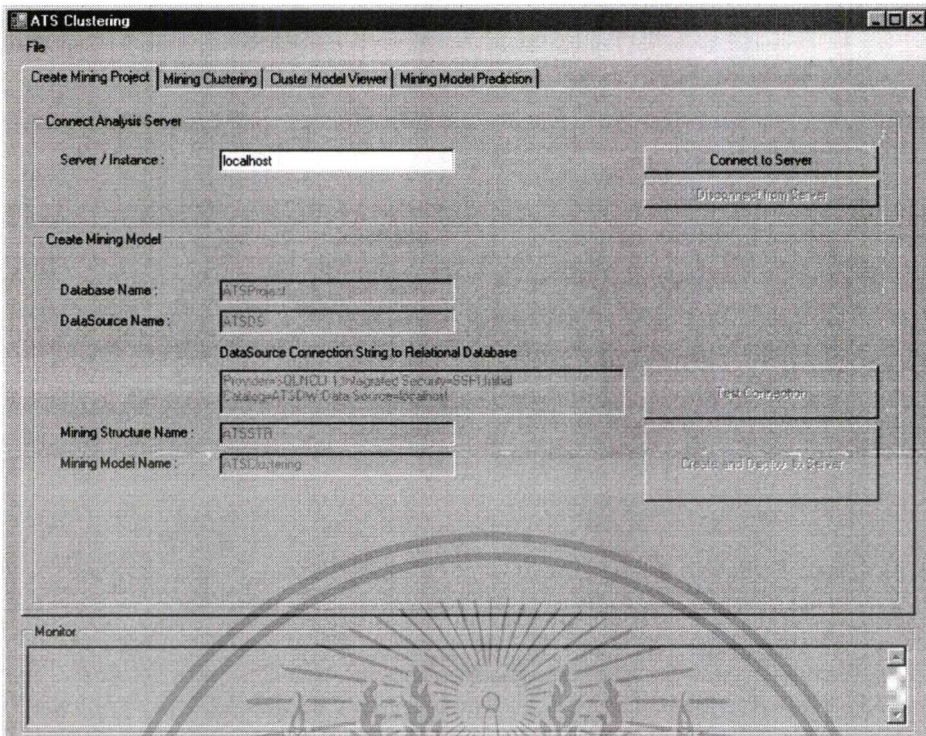
เป็นส่วนที่ทำหน้าที่ในการติดต่อกับ Analysis Server เพื่อเริ่มใช้ระบบงาน โดยระบุชื่อ Server และคลิกปุ่ม Connect to Server และกรณีต้องการยกเลิกการติดต่อกับ Server ทำการคลิกที่ปุ่ม Disconnect from Server ดังแสดงในรูปที่ 4.1

#### 2. Create Mining Model

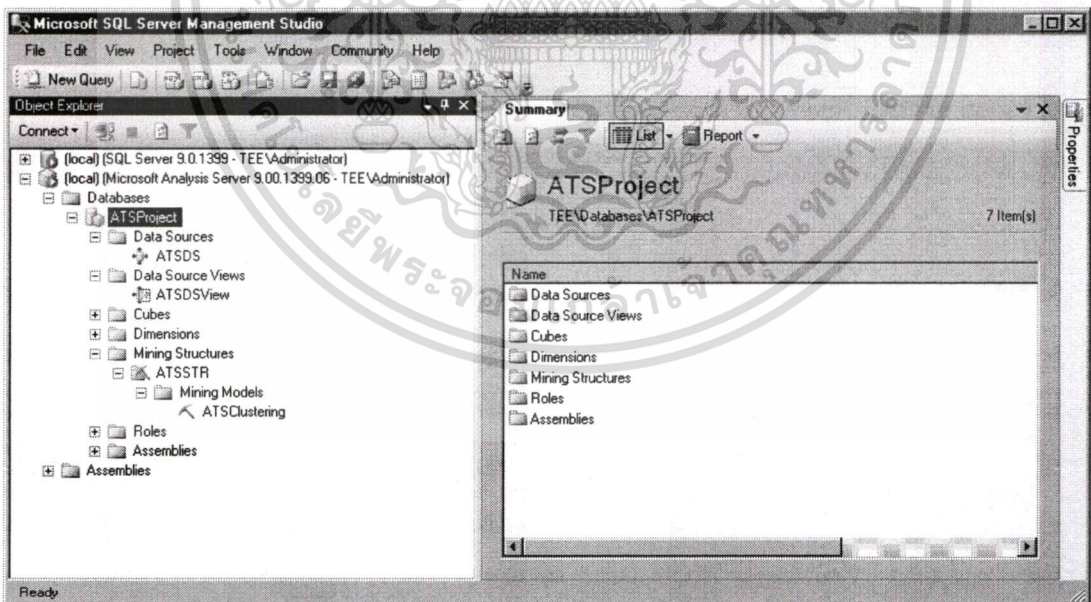
เป็นส่วนที่สร้างแบบจำลองค้ำไม้ Mining โดยระบุข้อมูลดังนี้

- Database Name กำหนดชื่อฐานข้อมูล
- DataSource Name กำหนดชื่อแหล่งที่มาของข้อมูล และชุดคำสั่งในการติดต่อแหล่งที่มาของข้อมูล (DataSource Connect String to Relation Database) โดยสามารถคลิกปุ่ม Test Connection เพื่อทดสอบการติดต่อแหล่งที่มาของข้อมูลที่ระบุได้
- Mining Structure Name กำหนดชื่อโครงสร้างค้ำไม้ Mining
- Mining Model Name กำหนดชื่อแบบจำลองค้ำไม้ Mining

เมื่อระบุข้อมูลในการสร้างแบบจำลองค้ำไม้ Mining สร็จแล้ว คลิกปุ่ม Create and Deploy to Server เพื่อสร้างแบบจำลองค้ำไม้ Mining และติดตั้งแบบจำลองบน Server ดังแสดงในรูปที่ 4.2



รูปที่ 4.1 หน้าจอการสร้างแบบจำลอง



รูปที่ 4.2 แบบจำลองดาต้าไมนิ่ง ATSPROJ

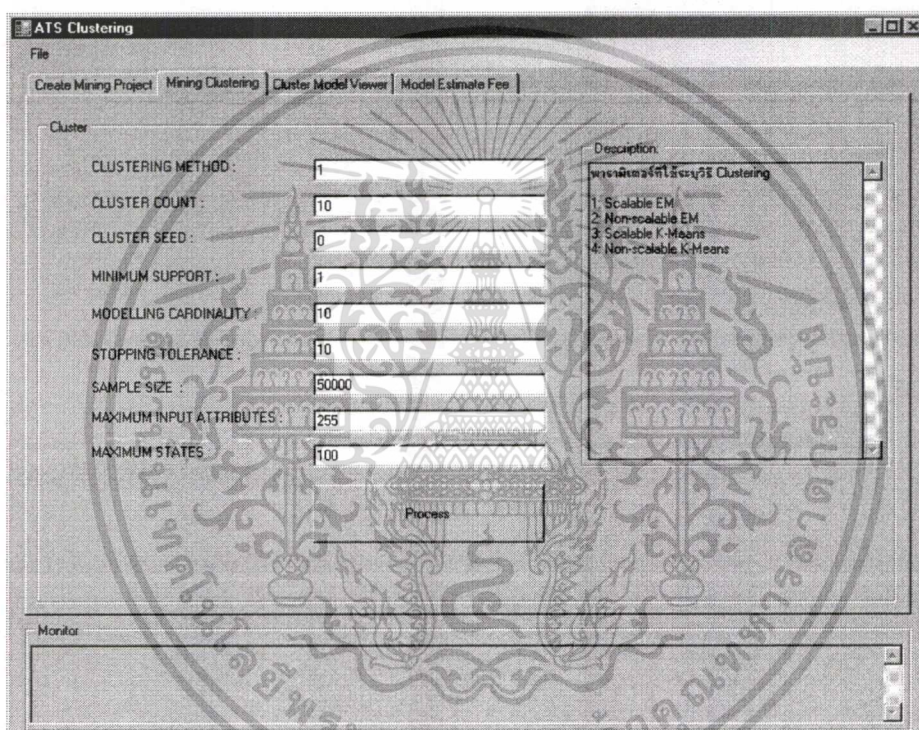
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.2 การไมนิ่งข้อมูล

ฟังก์ชันการทำงานในส่วนนี้เป็นการประมวลผลแบบจำลองที่ได้สร้างขึ้น ประกอบด้วยการทำงานหลัก ดังนี้

1. การกำหนดค่าพารามิเตอร์ที่ในการทำดาต้าไมนิ่ง ซึ่งบริษัทไมโครซอฟท์กำหนดค่าพารามิเตอร์ที่ใช้กับ Microsoft Clustering Algorithm เมื่อคลิกช่องแต่ละช่องพารามิเตอร์ระบบแสดงคำรายละเอียดอย่างย่อของแต่ละพารามิเตอร์ ดังแสดงในรูปที่ 4.3

2. คลิกปุ่ม Process เพื่อประมวลผล



รูปที่ 4.3 หน้าจอการไมนิ่งข้อมูล

## 4.3 การแสดงผลการจัดกลุ่ม

ฟังก์ชันการทำงานในส่วนนี้เป็นการแสดงผลการจัดกลุ่มจากการไมนิ่งข้อมูล ประกอบด้วยการทำงานหลัก ดังนี้

1. การแสดงผลการจัดกลุ่ม

คลิกปุ่ม Content Viewer เพื่อแสดงรายละเอียด Content ของแบบจำลอง ดังแสดงในรูปที่

4.4 หรือคลิกปุ่ม Cluster Viewer เพื่อแสดงผลการจัดกลุ่ม ระบบแสดงผลการจัดกลุ่มในส่วน

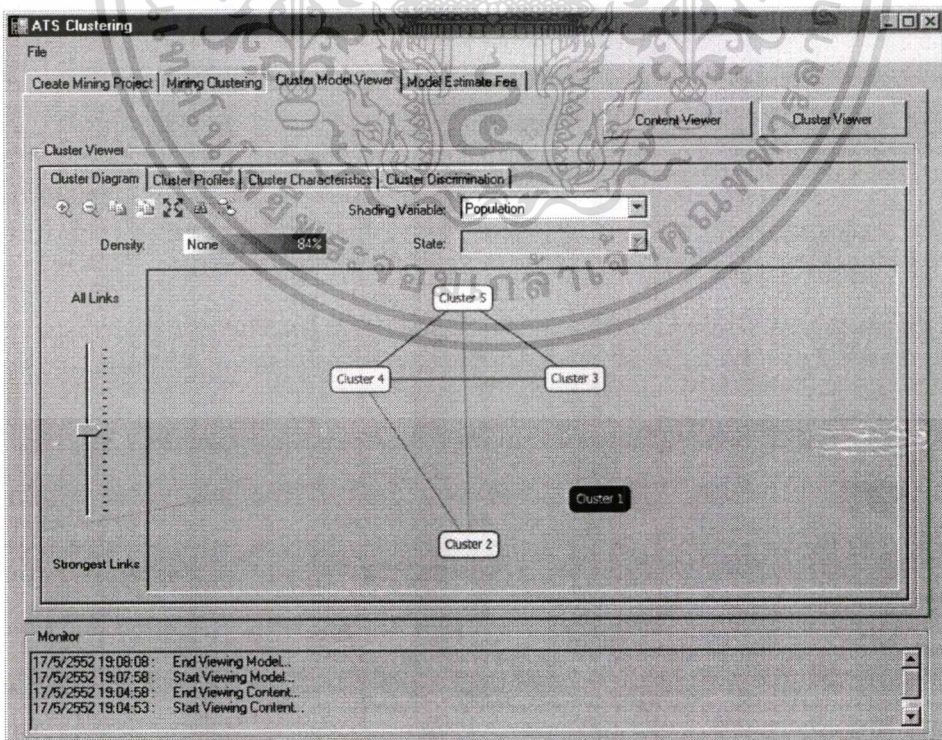
เอกส Cluster Viewer ประกอบด้วย 4 แท็บข้อมูล ดังแสดงในรูปที่ 4.5 ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

NODE_CAPTION	NODE_DESCRIPTION	NODE_PROBABIL	NODE_SUPPORT	MSOLAP_NODE_SCOR
Cluster Model	All	1	160	0.98280874588790967
Cluster 1	SEND_TYPE=03_MIN_...	0.84375	135	
Cluster 2	696.2155190435 <=TO...	0.05625	9	
Cluster 3	2412.2073214947 <=TO...	0.05	8	
Cluster 4	5434.6142302997 <=TO...	0.0375	6	
Cluster 5	36455 <=TOTAL_FEE <...	0.0125	2	

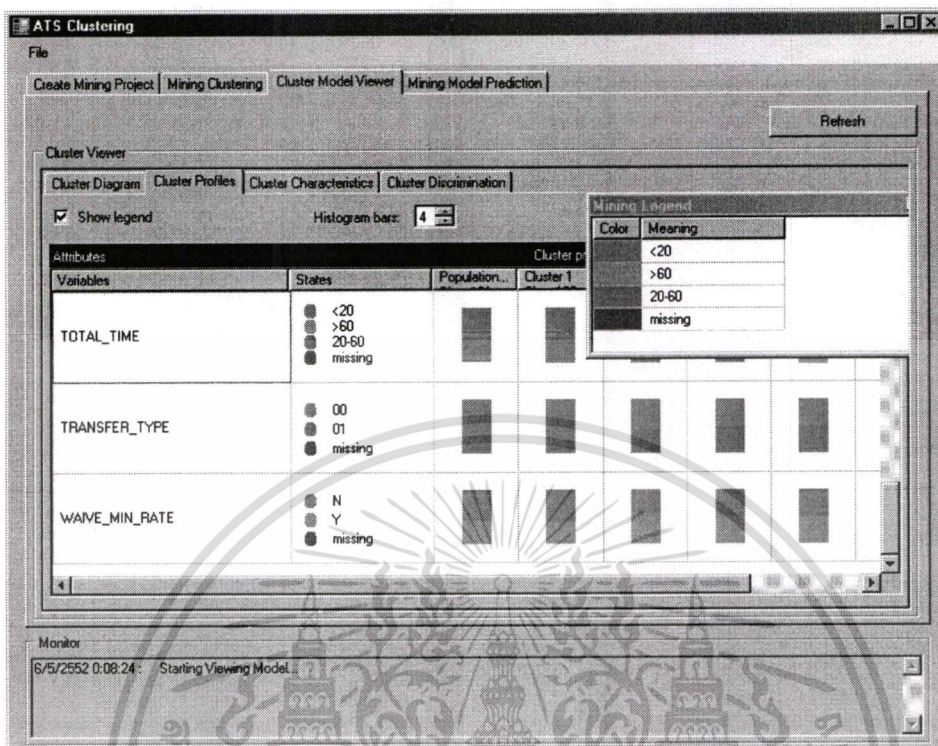
Monitor  
 17/5/2552 19:04:58: End Viewing Content...  
 17/5/2552 19:04:53: Start Viewing Content...

รูปที่ 4.4 หน้าจอการแสดงผลรายละเอียด Content ของแบบจำลอง

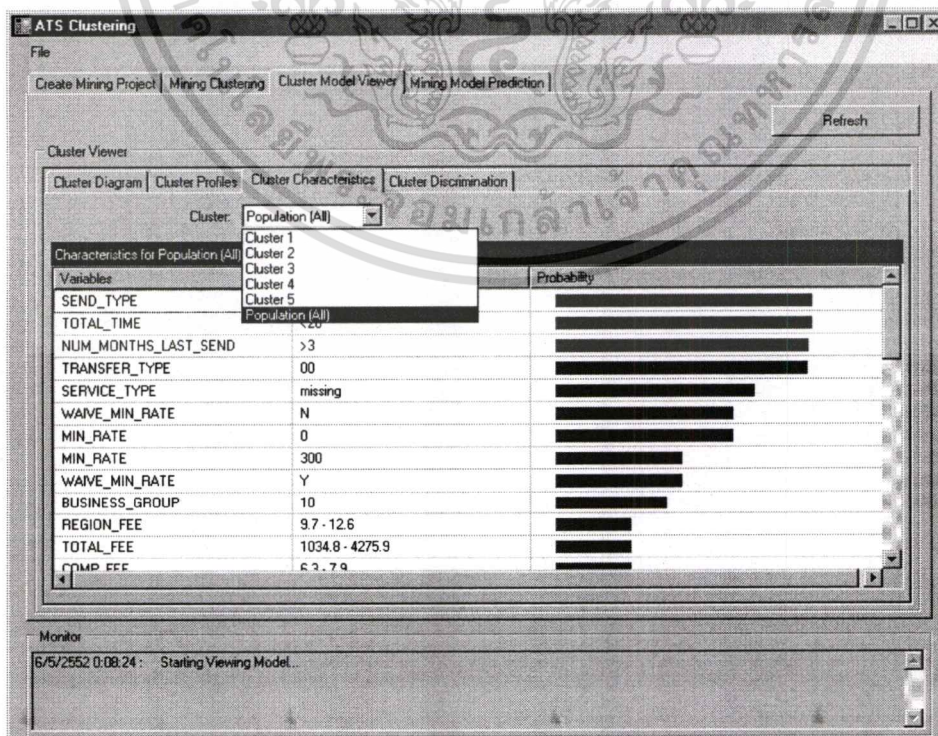


รูปที่ 4.5 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Diagram

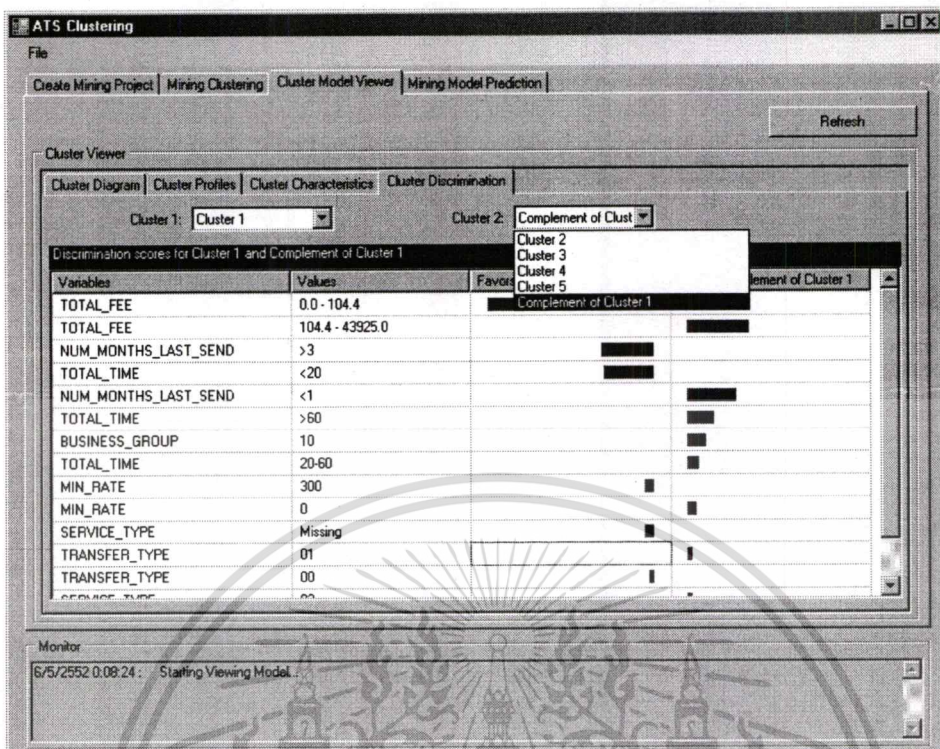
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Profile



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 รูปที่ 4.7 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Characteristics



รูปที่ 4.8 หน้าจอการแสดงผลการจัดกลุ่ม: แท็บ Cluster Discrimination

Drill Through

Cases Classified to:  
Cluster 1

GR...	MIN_RATE	COMP_FEE	REGION_FEE	WAVE_MIN_R...	TOTAL_TIME	NUM_MONTHS...
0	10	15	15	N	<20	>3
0	10	15	15	N	<20	>3
0	10	15	15	Y	<20	>3
0	10	15	15	N	<20	>3
0	10	15	15	Y	<20	>3
0	10	15	15	N	<20	>3
0	10	15	15	Y	<20	>3
0	10	15	15	N	<20	>3
3กค	5	5	5	N	<20	>3

Query execution completed with 136 rows fetched

รูปที่ 4.9 หน้าจอการแสดงผลการจัดกลุ่ม: การ Drill Through

### Cluster Diagram

แท็บแสดงแผนภาพ Cluster แสดงกลุ่มข้อมูลจากการทำค้ำไม่นิ่ง Cluster ที่มีสีเข้ม แสดงถึงมีจำนวนสมาชิกมากที่สุด และแต่ละ Cluster มีเส้นเชื่อมโยงไป Cluster อื่นโดยสีของเส้น เชื่อมโยงบ่งบอกถึงความคล้ายคลึงกันของกลุ่มข้อมูล กรณีเป็นสีสว่างหรือไม่มีสีเป็นกลุ่มที่มี

ไม่ว่าการณ์ใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมาชิกไม่คล้ายคลึงกันมาก หากสีเข้มมากแสดงถึงสามารถชกภายในกลุ่มมีความคล้ายคลึงกันมาก ดังแสดงในรูปที่ 4.5

เมื่อต้องการดูรายละเอียดข้อมูลของสมาชิกในแต่ละกลุ่มสามารถคลิกเมา์ที่ขวาเลือกเมนู ลัด Drill Through ระบบจะแสดงข้อมูลที่ถูกจัดไว้กลุ่ม ดังแสดงในรูปที่ 4.9

#### Cluster Profile

แท็บรายละเอียดของแต่ละ attribute ของแต่ละกลุ่ม แสดงถึงการกระจายของ attribute สถิติการกระจาย จำนวนประชากร ซึ่งแสดงในรูปแบบของ color bar หรือ diamond chart ดังแสดงในรูปที่ 4.6

#### Cluster Characteristics

แท็บแสดงลักษณะข้อมูลของแต่ละกลุ่ม โดยเลือกกลุ่มที่ต้องการจาก Cluster list ระบบจะแสดง attribute ที่ประกอบภายในกลุ่มที่เลือกและแสดงค่าข้อมูลและค่าความน่าจะเป็นสำหรับ attribute นั้นๆ ดังแสดงในรูปที่ 4.7

#### Cluster Discrimination

แท็บแสดงการเปรียบเทียบลักษณะข้อมูลระหว่างสองกลุ่ม เช่น Cluster 1 และ Cluster 2 โดยดู attribute ที่ต้องการเปรียบเทียบกัน ดังแสดงในรูปที่ 4.8

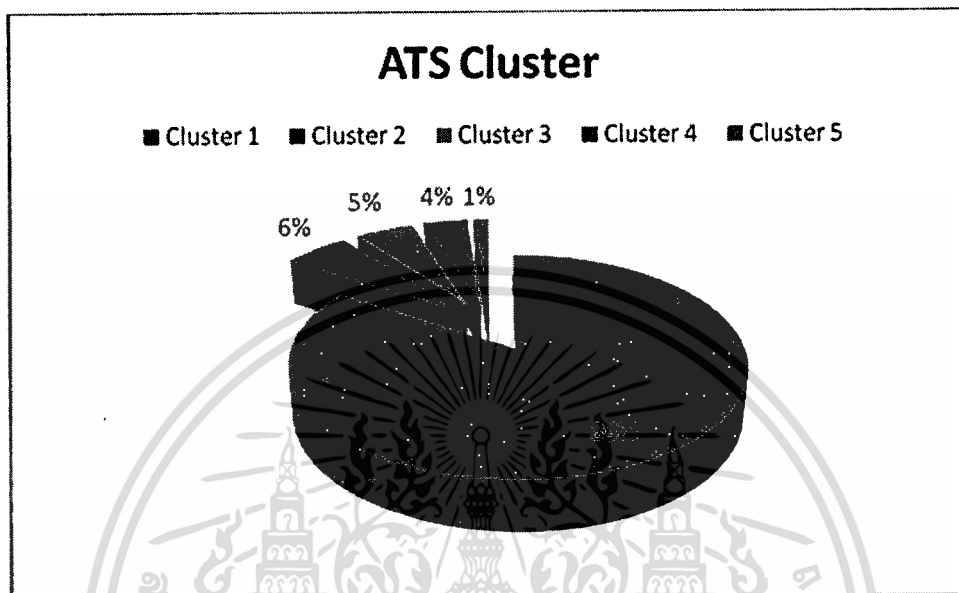
### 4.4 การประมาณค่าธรรมเนียมจากแบบจำลอง

ฟังก์ชันการทำงานในส่วนนี้เป็นการนำข้อมูล ช่องทางการส่งข้อมูล (Sending Type), ประเภทการทำรายการ (Transfer Type), ประเภทการบริการ (Service Type) และประเภทกลุ่มธุรกิจ (Business Group) เป็น input โดยจำลองเหตุการณ์เมื่อบริษัทที่เป็นลูกค้ามีลักษณะข้อมูลดังกล่าว แล้วประมาณว่าควรกำหนดอัตราค่าธรรมเนียมเท่าใด เพื่อใช้เป็นข้อมูลเบื้องต้นในการบริหารงานต่อไป ประกอบด้วยการทำงานหลัก ดังนี้



## 4.5 การวิเคราะห์ผลการจัดกลุ่ม (Clustering)

ผลจากการจัดกลุ่มพฤติกรรมกรรมการใช้บริการระบบโอนเงินอัตโนมัติ สามารถจัดกลุ่มได้ทั้งหมด 5 กลุ่ม ดังแสดงในรูปที่ 4.10



รูปที่ 4.11 แสดงผลการจัดกลุ่มพฤติกรรมกรรมการใช้บริการระบบโอนเงินอัตโนมัติ

**Cluster 1 :** เป็นกลุ่มลูกค้าที่ขาดการติดต่อกับธนาคารนานเกินกว่า 3 เดือน มีจำนวนการทำรายการน้อยกว่า 20 ครั้งภายใน 3 เดือนที่ผ่านมา ซึ่งเป็นกลุ่มลูกค้าที่มีการใช้บริการจ่ายเงินเดือน จ่ายค่าจ้าง ให้แก่พนักงาน

**กลยุทธ์ :** หาสาเหตุ และกำหนดกลยุทธ์ที่เหมาะสม

**Cluster 2 :** เป็นกลุ่มลูกค้าที่ลักษณะธุรกิจเป็นวิศวกรทางการเงิน มีการติดต่อกับธนาคารเมื่อไม่เกิน 1 เดือนที่ผ่านมา มีการทำรายการ 20 – 60 ครั้งภายใน 3 เดือนที่ผ่านมา

**กลยุทธ์ :** กำหนดแผนการส่งเสริมการขายเพื่อกระตุ้นยอดการใช้บริการให้มากขึ้น

**Cluster 3, 4, 5 :** เป็นกลุ่มลูกค้าที่ลักษณะธุรกิจเป็นวิศวกรทางการเงิน มีการติดต่อกับธนาคาร เมื่อไม่เกิน 1 เดือนที่ผ่านมา มีการทำรายการมากกว่า 60 ครั้งภายใน 3 เดือนที่ผ่านมา โดยทั้ง 3 Cluster ต่างกันที่แต่ละ Cluster จ่ายค่าธรรมเนียมแก่ธนาคารไม่เท่ากัน

**กลยุทธ์ :** กำหนดแผนการส่งเสริมการขายเพื่อลูกค้าเกิดความพึงพอใจ และภักดีต่อธนาคาร

## บทที่ 5

### สรุปผลการดำเนินงาน

ในการศึกษาเทคนิคเหมืองข้อมูลสำหรับการวิเคราะห์การจัดกลุ่มพฤติกรรมการใช้บริการระบบโอนเงินอัตโนมัติ (ATS : Automatic Transfer System) เป็นโครงการที่จัดทำขึ้นเพื่อประยุกต์ใช้เคมีอัลกอริทึม (K-Means Algorithm) ในการจัดกลุ่มข้อมูล โดยใช้โปรแกรม Microsoft SQL Server 2005 ซึ่งเป็นโปรแกรมจัดการฐานข้อมูลในการจัดเก็บข้อมูลระบบโอนเงินอัตโนมัติ (ATS : Automatic Transfer System) และ Microsoft SQL Server 2005 Analysis Services (SSAS) ช่วยในส่วนการวิเคราะห์ข้อมูล เพื่อช่วยเป็นแนวทางในการกำหนดกลยุทธ์ทางการตลาดหรือแผนการส่งเสริมการขายของแต่ละกลุ่มลูกค้าได้อย่างเหมาะสม

#### 5.1 สรุปผลการดำเนินงาน

โครงการพัฒนาระบบการจัดกลุ่มพฤติกรรมการใช้บริการระบบโอนเงินอัตโนมัติ (ATS : Automatic Transfer System) แบ่งการทำงานหลักเป็น 4 ส่วน ได้แก่

1. ส่วนการสร้างแบบจำลองการจัดกลุ่ม ซึ่งสามารถแก้ไข ปรับเปลี่ยน ได้สะดวก
2. ส่วนการไมนิ่งข้อมูลโดยสามารถปรับเปลี่ยนค่าพารามิเตอร์ได้ และสามารถประมวลผลเพื่อไมนิ่งข้อมูลได้หลายรอบ
3. ส่วนการแสดงผลการจัดกลุ่ม
4. ส่วนการพยากรณ์จากแบบจำลอง

จากการศึกษา สามารถประยุกต์ใช้ระบบงานที่พัฒนาขึ้นเพื่อรวมกลุ่มลูกค้าเพื่อช่วยให้สามารถดำเนินกลยุทธ์ทางการตลาด การทำการส่งเสริมการขายได้อย่างเหมาะสม ตลอดจนปรับปรุงผลิตภัณฑ์ และช่องทางในการให้บริการ ได้ถูกกลุ่มเป้าหมายมากขึ้น

#### 5.2 ข้อเสนอแนะ

ข้อเสนอแนะเพื่อการวิเคราะห์ในครั้งต่อไป

1. นำเทคนิคอื่นๆ ไปประยุกต์ใช้ในการทำเหมืองข้อมูล เช่น การหากฎความสัมพันธ์โดยการทำ Association Rule หรือวิธีอื่นๆ ในการทำเหมืองข้อมูล เพื่อให้ได้สารสนเทศใหม่ๆ

2. เนื่องจากระบบโอนเงินอัตโนมัติ (ATS : Automatic Transfer System) เป็นระบบที่ทางธนาคารเริ่มใช้ได้ไม่นาน ดังนั้นข้อมูลที่นำมาใช้ในการวิเคราะห์จึงมีปริมาณไม่มาก ซึ่งอาจทำให้ผลการวิเคราะห์ไม่สมบูรณ์เท่าที่ควร

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของมหาวิทยาลัยราชภัฏวชิราวุฒวิทยาลัย เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บรรณานุกรม

- นุสรรา จิรเจริญจิตต์. 2549. “การพัฒนากระบวนการวิเคราะห์กลุ่มลูกค้าเป้าหมายที่สร้างความคุ้มค่าให้กับธุรกิจการประกันภัย.” วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง.
- สุชาติ แก้วยศ. 2550. “การประยุกต์ใช้การค้นหากฎความสัมพันธ์กับระบบห้องสมุด.” วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง.
- สุธี พงศาสกุลชัย และทศชนก งามอินทร์. 2549. คัมภีร์ Visual C# 2005. กรุงเทพฯ : เลทีพี.
- Jiawei Han and Micheline Kamber. 2006. **Data Mining: Concepts and Techniques**. 2nd ed. Morgan Kaufmann Publisher.
- Lynn Langit. 2007. **Foundations of SQL Server 2005 Business Intelligence**. Press.
- Microsoft. 2009. **Microsoft Clustering Algorithm**. [Online].  
Available :<http://msdn.microsoft.com/en-us/library/ms174879.aspx>
- Microsoft. 2009. **SQL Server TechCenter**. [Online].  
Available : [http://technet.microsoft.com/en-us/library/ms174879\(SQL.90\).aspx](http://technet.microsoft.com/en-us/library/ms174879(SQL.90).aspx)
- ZhaoHui Tang and Jamie MacLennan. 2005. **Data Mining with SQL Server 2005**. Indiana : Wiley Publishing, Inc.

## ประวัติผู้เขียน

ชื่อผู้เขียน	นาย ธวัช วิวัฒนารามย์
สถานที่เกิด	สมุทรสาคร
การศึกษา	ระดับปริญญาตรี วท.บ วิทยาศาสตร์บัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยรามคำแหง
ประสบการณ์ทำงาน	Programmer บริษัท SOAT Solution Co. Ltd (2545-2547) TQA ธนาคารชนชาติ (2548 - ปัจจุบัน)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้