

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

ระบบวิเคราะห์พฤติกรรมผู้บริโภคเพื่อการตัดสินใจ

CONSUMER BEHAVIOR ANALYSIS SYSTEM  
FOR DECISION MAKING



\*H004855\*



ฉท.  
๕ ๖๖๑ ๖  
๑๕๖๐

เลขหมู่.....

เลขทะเบียน.....04855.....

วัน,เดือน,ปี..... ๘ ๗.ค. ๒๕๕๑.....

๑๑๙๗๗๙๙๑

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการศึกษาระดับปริญญาตรี

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ภาคเรียนที่ ๒ ปีการศึกษา ๒๕๕๐

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**CONSUMER BEHAVIOR ANALYSIS SYSTEM  
FOR DECISION MAKING**



**A SPECIAL STUDY 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**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
2/2007  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**COPYRIGHT 2008**

**FACULTY OF INFORMATION TECHNOLOGY**

เอกสารนี้ King Mongkut's Institute of Technology Ladkrabang การค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

|                  |  |
|------------------|--|
| หัวข้อ           | ระบบวิเคราะห์พฤติกรรมผู้บริโภคเพื่อการตัดสินใจ |
| นักศึกษา         | นางสาวรุจิรา แซ่เตีย                           |
| รหัสนักศึกษา     | 49066613                                       |
| ปริญญา           | วิทยาศาสตรมหาบัณฑิต                            |
| สาขาวิชา         | เทคโนโลยีสารสนเทศ                              |
| แขนงวิชา         | การจัดการเทคโนโลยีสารสนเทศ                     |
| ปีการศึกษา       | 2550   |
| อาจารย์ที่ปรึกษา | ผศ.ดร. ภัทรชัย ทลิตโรจน์วงศ์                   |

### บทคัดย่อ

ในปัจจุบันการค้ามีธุรกิจมีการแข่งขันกันอย่างสูง องค์กรธุรกิจจำเป็นต้องปรับกลยุทธ์ เพื่อเพิ่มศักยภาพให้สามารถแข่งขันกับองค์กรอื่นได้ โดยทิศทางการแข่งขันในปัจจุบันเป็นยุคของ ข้อมูลข่าวสารและมุ่งเน้นถึงความต้องการของลูกค้าเป็นสำคัญ จึงได้มีการนำเทคนิคดาต้าไมน์นิ่ง มาช่วยวิเคราะห์และค้นหาสารสนเทศที่มีประโยชน์จากข้อมูลจำนวนมาก ช่วยวิเคราะห์พฤติกรรม ผู้บริโภค ซึ่งจะช่วยให้องค์กรสามารถสร้างรายได้ เพิ่มส่วนแบ่งการตลาดตลอดจนการรักษาฐาน ลูกค้าขั้นดีให้อยู่กับองค์กรได้ โดยในรายงานฉบับนี้ได้นำเสนอการวิเคราะห์ออกแบบระบบซึ่ง ดำเนินการตามขั้นตอนของเทคนิคดาต้าไมน์นิ่งและใช้ซอฟต์แวร์ Enterprise Miner ของบริษัท SAS เป็นเครื่องมือในการวิเคราะห์ข้อมูลต่างๆ เพื่อจะได้ผลลัพธ์ ซึ่งนำไปสู่การประยุกต์ใช้งานและ สนับสนุนการตัดสินใจขององค์กรสำหรับการวางแผนกลยุทธ์ ในด้านต่างๆ ให้สามารถรองรับ ความต้องการของลูกค้าและสามารถแข่งขันในตลาดได้อย่างมีประสิทธิภาพ

|                      |   |
|----------------------|---|
| <b>Title</b>         | Consumer Behavior Analysis System for Decision Making |
| <b>Student</b>       | Miss. Rujira Saetia                                   |
| <b>Student ID.</b>   | 49066613  |
| <b>Degree</b>        | Master of Science                                     |
| <b>Programme</b>     | Information Technology Management                     |
| <b>Academic Year</b> | 2007  |
| <b>Advisor</b>       | Asst. Prof. Dr. Pattarachai Lalitrojwong              |

## ABSTRACT

Presently, there is high competition in various businesses. The entrepreneurs and business organization need to adjust their strategies in order to increase their potential to compete with other companies. The direction of the competition is the information trend and focus on customer needs. So, Data mining is applied to analyze and searched the useful information from various databases. Most important thing in using data mining is analyzing the consumer behavior. This is a strategy to increase revenue, market share and including maintain customer base. This report mentions system, design decision making process and data mining techniques. The tool employed in this project is Enterprise Miner software of SAS Co.,Ltd. The result of data mining process can be applied to use and support decision making of company strategy. This process will be an idea to fulfill customer needs and compete efficiently in the market.

# กิตติกรรมประกาศ

ความสำเร็จของการพัฒนาโครงการศึกษาระดับพิเศษฉบับนี้ สำเร็จขึ้นได้จากความช่วยเหลือของบุคคลหลายๆ ท่าน ข้าพเจ้าขอขอบคุณทุกท่านที่มีส่วนช่วยเหลือในด้านต่างๆ ด้วยความจริงใจ หากขาดบุคคลที่จะกล่าวถึงต่อไปนี้จะส่งผลกระทบต่อความสำเร็จของโครงการศึกษาระดับพิเศษฉบับนี้ ข้าพเจ้าจึงขอขอบคุณทุกท่านในความช่วยเหลือด้วยความจริงใจของทุกท่าน

ขอขอบพระคุณ ผศ.ดร. ภัทรชัย ลลิตโรจน์วงศ์ อาจารย์ที่ปรึกษาโครงการ ที่ได้ให้คำแนะนำในด้านต่างๆ ทำให้เกิดประโยชน์ต่อการออกแบบระบบเป็นอย่างมาก รวมทั้งเสียสละเวลา ในการช่วยเหลือตลอดระยะเวลาในการศึกษา

ขอขอบพระคุณ คณาจารย์ทุกท่านที่ได้ประสาทวิชาความรู้ซึ่งช่วยทำให้ข้าพเจ้าได้รับความรู้ทั้งหลายทั้งมวลและได้นำความรู้ที่ได้รับไปใช้ประโยชน์ที่ก่อให้เกิดความเจริญก้าวหน้าต่อไป

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดา ตลอดจนญาติพี่น้องในครอบครัวทุกท่านที่ให้การสนับสนุนและเป็นกำลังใจในการศึกษาโดยตลอด

รุจิรา แซ่เตีย

# สารบัญ

|  | หน้า |
|--|------|
| บทคัดย่อภาษาไทย.....                     | I    |
| บทคัดย่อภาษาอังกฤษ.....                  | II   |
| กิตติกรรมประกาศ.....                     | III  |
| สารบัญ.....                              | IV   |
| สารบัญตาราง.....                         | VI   |
| สารบัญรูป.....                           | VII  |
| <br>                                     |      |
| บทที่ 1 บทนำ.....                        | 1    |
| 1.1 ความเป็นมา.....                      | 1    |
| 1.2 วัตถุประสงค์.....                    | 1    |
| 1.3 ขอบเขตการศึกษา.....                  | 2    |
| 1.4 ประโยชน์ที่คาดว่าจะได้รับ.....       | 2    |
| 1.5 ขั้นตอนการศึกษา.....                 | 2    |
| <br>                                     |      |
| บทที่ 2 ทฤษฎีพื้นฐาน.....                | 4    |
| 2.1 การตัดสินใจและกระบวนการตัดสินใจ..... | 4    |
| 2.2 ความหมายและหลักการของค้ำไม้ค้ำ.....  | 6    |
| 2.3 ขั้นตอนการทำค้ำไม้ค้ำ.....           | 9    |
| 2.4 เทคนิคค้ำไม้ค้ำ.....                 | 10   |
| <br>                                     |      |
| บทที่ 3 การเตรียมข้อมูล.....             | 20   |
| 3.1 แหล่งข้อมูล.....                     | 20   |
| 3.2 ระบบปัจจุบันและปัญหาของระบบ.....     | 22   |
| 3.3 การเตรียมข้อมูล.....                 | 23   |

## สารบัญ (ต่อ)

|  | หน้า |
|--|------|
| บทที่ 4 การนำข้อมูลเข้าสู่ฐานข้อมูล .....                                | 35   |
| 4.1 มเดลความสัมพันธ์ของฐานข้อมูลใหม่ .....                               | 35   |
| 4.2 การนำข้อมูลเข้าสู่ฐานข้อมูลใหม่.....                                 | 39   |
| บทที่ 5 ซอฟต์แวร์สำหรับการทำคาด้าไมน์นึ่ง .....                          | 44   |
| 5.1 การเลือกซอฟต์แวร์สำหรับการทำคาด้าไมน์นึ่ง .....                      | 44   |
| 5.2 คุณสมบัติหลักของซอฟต์แวร์ .....                                      | 44   |
| 5.3 ตัวอย่างบริษัทที่พัฒนาผลิตภัณฑ์เกี่ยวกับคาด้าไมน์นึ่ง.....           | 45   |
| บทที่ 6 การประยุกต์ใช้เทคนิคคาด้าไมน์นึ่งด้วยซอฟต์แวร์ SAS .....         | 50   |
| 6.1 วิเคราะห์พฤติกรรมการซื้อขายของลูกค้า.....                            | 50   |
| 6.2 วิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทและวิธีการขนส่งสินค้า..... | 55   |
| 6.3 วิเคราะห์พฤติกรรมของการเลือกขนส่งสินค้าของลูกค้า.....                | 60   |
| 6.4 วิเคราะห์กลุ่มลูกค้าที่สั่งซื้อสินค้า.....                           | 64   |
| 6.5 ทำนายยอดขายในวันถัดไป.....   | 68   |
| บทที่ 7 สรุปผลการศึกษา และข้อเสนอแนะ .....                               | 73   |
| 7.1 สรุปผลการศึกษา.....  | 73   |
| 7.2 ข้อเสนอแนะ .....   | 73   |
| บรรณานุกรม.....  | 75   |
| ภาคผนวก .....  | 76   |
| ประวัติผู้เขียน .....  | 82   |

# สารบัญตาราง

| ตารางที่   | หน้า |
|--|------|
| 3.1 รายละเอียดของแอดทริบิวต์ที่ใช้ในการวิเคราะห์พฤติกรรมการซื้อขายสินค้า .....   | 24   |
| 3.2 ตัวอย่างข้อมูลสำหรับวิเคราะห์พฤติกรรมการซื้อขายสินค้าของลูกค้า.....  | 25   |
| 3.3 รายละเอียดของแอดทริบิวต์ที่ใช้ในการวิเคราะห์สาเหตุของการส่งคืนสินค้า .....   | 26   |
| 3.4 ตัวอย่างข้อมูลสำหรับวิเคราะห์สาเหตุการขอคืนสินค้าแยกตามประเภทของสินค้าและการขนส่ง.....   | 27   |
| 3.5 ตัวอย่างข้อมูลที่ผ่านกระบวนการทำความสะอาดข้อมูล แปลงข้อมูลและรวบรวมข้อมูลสำหรับวิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทของสินค้าและการขนส่ง..... | 27   |
| 3.6 รายละเอียดของแอดทริบิวต์ที่ใช้วิเคราะห์การเลือกขนส่งสินค้า .....   | 28   |
| 3.7 ตัวอย่างข้อมูลสำหรับวิเคราะห์พฤติกรรมเลือกส่งขนส่งสินค้าของลูกค้าตามประเภทของสินค้า.....   | 29   |
| 3.8 ตัวอย่างข้อมูลหลังจากผ่านกระบวนการเตรียมข้อมูลสำหรับวิเคราะห์พฤติกรรมเลือกส่งขนส่งสินค้าของลูกค้าตามประเภทของสินค้า .....                          | 30   |
| 3.9 รายละเอียดของแอดทริบิวต์ที่ใช้วิเคราะห์กลุ่มของลูกค้าที่สั่งซื้อสินค้า.....  | 30   |
| 3.10 ตัวอย่างข้อมูลสำหรับวิเคราะห์กลุ่มของลูกค้าที่สั่งซื้อสินค้า.....   | 31   |
| 3.11 ตัวอย่างข้อมูลหลังจากผ่านกระบวนการเตรียมข้อมูลสำหรับวิเคราะห์กลุ่มของลูกค้าที่สั่งซื้อสินค้า.....   | 32   |
| 3.12 รายละเอียดของแอดทริบิวต์ที่ใช้สำหรับทำนายยอดขายสินค้า .....   | 33   |
| 3.13 ตัวอย่างข้อมูลสำหรับทำนายยอดขายสินค้า .....   | 34   |
| 6.1 ผลลัพธ์จากการรันโปรแกรมด้วยเทคนิคแผนภาพต้นไม้ในรูปแบบของกฎ “ถ้า...แล้ว” สำหรับหัวข้อที่ 3.2 .....  | 58   |
| 6.2 ผลลัพธ์จากการรันโปรแกรมด้วยเทคนิคแผนภาพต้นไม้ในรูปแบบของกฎ “ถ้า...แล้ว” สำหรับหัวข้อที่ 6.3 .....  | 62   |

# สารบัญรูป

| รูปที่ | หน้า  |
|--------|---|
| 2.1    | กระบวนการตัดสินใจและแก้ปัญหา..... 6   |
| 2.2    | สถาปัตยกรรมพื้นฐานของคาด้าไมน์นิ่ง..... 8   |
| 2.3    | ขั้นตอนการทำคาด้าไมน์นิ่ง ..... 9   |
| 2.4    | ประเภทของเทคนิคคาด้าไมน์นิ่งและตัวอย่างการทำงาน ..... 11  |
| 2.5    | ตัวอย่างแผนภาพต้นไม้เพื่อวิเคราะห์โอกาสที่ลูกค้าผู้เช่าบ้านจะซื้อบ้าน..... 13                           |
| 2.6    | ตัวอย่างแผนภาพการแบ่งกลุ่มข้อมูล ..... 14   |
| 2.7    | ตัวอย่างแผนภาพนิเวศน์เน็ตเวิร์กเพื่อวิเคราะห์หาผลลัพธ์..... 15  |
| 2.8    | ผลการขายสินค้าประเภทผ้าอ้อมและเปียร์ ..... 16   |
| 3.1    | ความสัมพันธ์ของตารางข้อมูลของระบบปัจจุบัน ..... 21  |
| 3.2    | ตารางความสัมพันธ์ของข้อมูล ORD_No และ PRD_Name..... 24  |
| 3.3    | ตารางความสัมพันธ์ของข้อมูล CLM_Title CAT_Name และ SPT_Name ..... 27                                     |
| 3.4    | ตารางความสัมพันธ์ของข้อมูล CAT_Name, SPT_Name, SPT_Description,<br>ORD_Weight และ ADD_PostCode ..... 29 |
| 3.5    | ตารางความสัมพันธ์ของข้อมูล CAT_Name, PDT_Price, ORD_Weight,<br>ADD_PostCode และ USR_DateOfBirth..... 32 |
| 4.1    | ความสัมพันธ์ของตารางข้อมูลที่จำเป็นต้องใช้ในการวิเคราะห์ข้อมูล ..... 35                                 |
| 4.2    | การคิวรีข้อมูลจากตารางความสัมพันธ์ตามวัตถุประสงค์ที่ต้องการ ..... 36                                    |
| 4.3    | ตาราง Buy Product จากการคิวรีข้อมูลจากตารางความสัมพันธ์ ..... 36  |
| 4.4    | ตาราง Claim จากการคิวรีข้อมูลจากตารางความสัมพันธ์..... 37   |
| 4.5    | ตาราง Shipping Order จากการคิวรีข้อมูลจากตารางความสัมพันธ์..... 37                                      |
| 4.6    | ตาราง Buy Product จากการคิวรีข้อมูลสำหรับวัตถุประสงค์การวิเคราะห์กลุ่มลูกค้า... 38                      |
| 4.7    | ตาราง Promotion จากการคิวรีข้อมูลจากตารางความสัมพันธ์..... 38   |
| 4.8    | ตารางสำหรับฐานข้อมูลใหม่..... 39  |
| 4.9    | การสร้าง New DTS Package และการออกแบบ DTS Package ..... 40  |
| 4.10   | การสร้างงานสำหรับการโอนข้อมูลโดยเลือกข้อมูลด้านต้นทาง ..... 41  |
| 4.11   | การสร้างตารางปลายทางสำหรับฐานข้อมูลใหม่..... 41   |
| 4.12   | การกำหนดการเทียบเคียงข้อมูลระหว่างต้นทางและปลายทาง..... 42  |
| 4.13   | การกำหนดไฟล์ข้อมูล..... 43  |

# สารบัญรูป (ต่อ)

| รูปที่   | หน้า |
|--|------|
| 4.14 การกำหนดช่วงเวลาในการดึงข้อมูลแบบอัตโนมัติ.....                           | 43   |
| 5.1 เริ่มเปิดโปรแกรมของซอฟต์แวร์ Intelligent Miner.....                        | 45   |
| 5.2 การแสดงผลลัพธ์ในรูปแบบของแผนภาพของซอฟต์แวร์ Intelligent Miner .....        | 45   |
| 5.3 เริ่มเปิดโปรแกรมของซอฟต์แวร์ Enterprise Miner.....                         | 46   |
| 5.4 การนำข้อมูลเข้าสู่โปรแกรมและใช้เทคนิคค้ำไมน์นิ่งของ Enterprise Miner ..... | 46   |
| 5.5 เริ่มเปิดโปรแกรมของซอฟต์แวร์ Clementine .....                              | 47   |
| 5.6 การนำข้อมูลเข้าสู่โปรแกรมของซอฟต์แวร์ Clementine .....                     | 47   |
| 5.7 การแสดงผลลัพธ์ในรูปแบบแผนภาพของซอฟต์แวร์ Clementine .....                  | 48   |
| 5.8 เริ่มเปิดโปรแกรมของซอฟต์แวร์ WEKA .....                                    | 48   |
| 5.9 การเลือกเทคนิคค้ำไมน์นิ่งแบบต่างๆ ของซอฟต์แวร์ WEKA .....                  | 49   |
| 5.10 การแสดงผลลัพธ์ในรูปแบบแผนภาพของซอฟต์แวร์ WEKA .....                       | 49   |
| 6.1 การนำข้อมูลเข้าสู่โปรแกรมและเลือกประเภทของข้อมูล .....                     | 50   |
| 6.2 การสร้างโมเดลด้วยเทคนิคกฎความสัมพันธ์ .....                                | 51   |
| 6.3 การกำหนดค่าพารามิเตอร์ข้อมูลนำเข้าสำหรับเทคนิคกฎความสัมพันธ์ .....         | 51   |
| 6.4 การกำหนดค่าพารามิเตอร์ของ Association Model สำหรับเทคนิคกฎความสัมพันธ์...  | 52   |
| 6.5 ผลลัพธ์ในแท็บ Rules ของเทคนิคกฎความสัมพันธ์.....                           | 52   |
| 6.6 ผลลัพธ์ในแท็บ Rules ของเทคนิคกฎความสัมพันธ์.....                           | 53   |
| 6.7 ผลลัพธ์ในแท็บ Frequencies ของเทคนิคกฎความสัมพันธ์ .....                    | 53   |
| 6.8 การสร้างโมเดลด้วยเทคนิคแผนภาพต้นไม้.....                                   | 55   |
| 6.9 การกำหนดค่าพารามิเตอร์ของ Input Data สำหรับเทคนิคแผนภาพต้นไม้.....         | 55   |
| 6.10 การกำหนดค่าพารามิเตอร์ของ Data Partition สำหรับเทคนิคแผนภาพต้นไม้ .....   | 56   |
| 6.11 การกำหนดค่าพารามิเตอร์ของ Tree Model สำหรับเทคนิคแผนภาพต้นไม้.....        | 56   |
| 6.12 ผลลัพธ์ของเทคนิคแผนภาพต้นไม้สำหรับการส่งคืนสินค้า .....                   | 57   |
| 6.13 ผลลัพธ์ในรูปแบบของแผนภาพของเทคนิคแผนภาพต้นไม้สำหรับการส่งคืนสินค้า...     | 57   |
| 6.14 ผลลัพธ์ของเทคนิคแผนภาพต้นไม้สำหรับการขนส่งสินค้า .....                    | 61   |
| 6.15 ผลลัพธ์ในรูปแบบของแผนภาพของเทคนิคแผนภาพต้นไม้สำหรับการขนส่งสินค้า...      | 61   |
| 6.16 การสร้างโมเดลของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means .....            | 64   |
| 6.17 การกำหนดพารามิเตอร์ของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means .....      | 65   |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

## สารบัญรูป (ต่อ)

| รูปที่   | หน้า |
|--|------|
| 6.18 ผลลัพธ์ในรูปแบบของกราฟของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means เพื่อดูการกระจายของข้อมูล ..... | 65   |
| 6.19 ผลลัพธ์ในรูปแบบแผนภาพวงกลมของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means                             | 66   |
| 6.20 ค่า RMS-SD ในแต่ละกลุ่มของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means .....                          | 66   |
| 6.21 การสร้างโมเดลของเทคนิคพยากรณ์อนุกรมเวลา .....   | 68   |
| 6.22 การสร้างโปรเจ็คใหม่สำหรับเทคนิคพยากรณ์อนุกรมเวลา.....   | 68   |
| 6.23 การเลือกข้อมูลอินพุตเข้าสู่โมเดลพยากรณ์อนุกรมเวลา .....   | 69   |
| 6.24 การกำหนดพารามิเตอร์สำหรับเทคนิคพยากรณ์อนุกรมเวลา.....   | 69   |
| 6.25 การจำลองโมเดลเพื่อเลือกโมเดลที่ดีที่สุด .....   | 70   |
| 6.26 การกระจายข้อมูลของ โมเดล Mean ตามข้อมูลอินพุต .....   | 71   |
| 6.27 การเลือกทำนายค่าตามวันที่ต้องการ .....  | 71   |
| 6.28 ผลลัพธ์จากการทำนายค่า .....   | 72   |

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมา

การดำเนินธุรกิจในยุคปัจจุบันที่เทคโนโลยีมีการเปลี่ยนแปลงอย่างรวดเร็วและตลอดเวลา เช่นเดียวกับ ระบบธุรกิจที่มีการแข่งขันกันค่อนข้างสูง ซึ่งผู้บริโภคเป็นผู้กำหนดทิศทางของสินค้าและบริการ หากองค์กรใดไม่สามารถเข้าถึงพฤติกรรมของผู้บริโภคได้แล้วนั้น อาจทำให้รายได้ขององค์กรลดลงและมีความเสี่ยงที่จะสูญเสียลูกค้าให้กับคู่แข่งที่เพิ่มขึ้น จึงเป็นสิ่งสำคัญที่การที่องค์กรจะอยู่รอดได้ จะต้องมีการใช้ข้อมูลสารสนเทศที่ดี มีประโยชน์ ทันท่วงที หาหนทางที่จะเก็บรวบรวมข้อมูลสารสนเทศที่มีคุณค่าและนำข้อมูลเหล่านั้นมาวิเคราะห์ เพื่อใช้ในการตัดสินใจและวางแผนทางหรือกลยุทธ์ขององค์กรให้ตรงตามความต้องการของผู้บริโภคให้ได้มากที่สุด

โดยแต่ละองค์กรสามารถค้นหาข้อมูลของลูกค้าได้จากแหล่งข้อมูลหลายๆแหล่งแตกต่างกันไป ตัวอย่างเช่น ใบสั่งซื้อสินค้า ข้อมูลประวัติการสั่งซื้อสินค้า ข้อมูลความสนใจหรือเยี่ยมชมสินค้าและบริการจากเว็บไซต์ เป็นต้น ข้อมูลเหล่านี้ล้วนแล้วแต่เป็นแหล่งข้อมูลดิบที่เกิดขึ้นจากการดำเนินงานขององค์กรในแต่ละวัน แต่ละเดือน ซึ่งยากต่อการนำมาใช้ในการวิเคราะห์ ดังนั้น การพัฒนาระบบวิเคราะห์พฤติกรรมผู้บริโภคเพื่อการตัดสินใจจะสามารถช่วยให้องค์กรตัดสินใจวางกลยุทธ์ในด้านต่างๆได้

### 1.2 วัตถุประสงค์

การวิเคราะห์พฤติกรรมผู้บริโภคสำหรับองค์กรมีวัตถุประสงค์ดังต่อไปนี้

1. เพื่อเป็นแนวทางในการพัฒนาระบบวิเคราะห์พฤติกรรมผู้บริโภคเพื่อการตัดสินใจของธุรกิจอีคอมเมิร์ซหรือพาณิชย์อิเล็กทรอนิกส์
2. เพื่อก่อให้เกิดการนำข้อมูลที่มีอยู่มาใช้ได้อย่างมีประสิทธิภาพสูงสุดในเชิงพาณิชย์
3. เพื่อคัดเลือกข้อมูลจากข้อมูลในระดับปฏิบัติการให้อยู่ในรูปของสารสนเทศที่มีความหมายในเชิงวิเคราะห์และนำไปสู่การสนับสนุนการตัดสินใจ
4. เพื่อวิเคราะห์สารสนเทศที่มาจากกรคัดเลือกข้อมูลด้วยเทคนิคค้ำไ่มนึ่งแบบต่างๆ
5. เพื่อช่วยให้ผู้บริหารสามารถนำผลลัพธ์ที่ได้มาช่วยในการวางแผนกลยุทธ์ขององค์กรในต่างๆ อย่างมีประสิทธิภาพ

### 1.3 ขอบเขตการศึกษา

โครงการนี้มีขอบเขตของการศึกษา ดังนี้

1. ศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้องกับกระบวนการตัดสินใจและเทคนิคคาด้าไมน์นิ่ง
2. ศึกษาขั้นตอนในการเลือกซอฟต์แวร์สำหรับเทคนิคคาด้าไมน์นิ่งที่เหมาะสมเพื่อใช้ในการวิเคราะห์ปัญหาและความต้องการของระบบ
3. วิเคราะห์ปัญหาและความต้องการของระบบ
4. เตรียมข้อมูลเพื่อนำไปสู่กระบวนการทำคาด้าไมน์นิ่งอย่างถูกต้องและเหมาะสม
5. วิเคราะห์ข้อมูลด้วยเทคนิคคาด้าไมน์นิ่งแบบต่างๆ ด้วยซอฟต์แวร์
6. นำเสนอผลลัพธ์ที่ได้จากการใช้เทคนิคคาด้าไมน์นิ่งในการวิเคราะห์ข้อมูล

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

เมื่อดำเนินการพัฒนาโครงการเสร็จสิ้น คาดว่าจะได้รับประโยชน์ ดังนี้

1. ธุรกิจอีคอมเมิร์ซหรือพาณิชย์อิเล็กทรอนิกส์ได้แนวทางในการพัฒนาระบบวิเคราะห์พฤติกรรมผู้บริโภคเพื่อการตัดสินใจ
2. สามารถนำข้อมูลที่มีอยู่มาใช้ได้อย่างมีประสิทธิภาพสูงสุดในเชิงพาณิชย์
3. สามารถจัดเตรียมข้อมูลได้อย่างถูกต้องและเหมาะสมกับเทคนิคคาด้าไมน์นิ่ง
4. ผู้บริหารมีข้อมูลหรือได้รับผลลัพธ์ที่สามารถช่วยในการสนับสนุนการตัดสินใจในการวางแผนกลยุทธ์ขององค์กรในต่างๆ อย่างมีประสิทธิภาพ

### 1.5 ขั้นตอนของการศึกษา

โครงการศึกษากรณีพิเศษฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ

บทที่ 1 กล่าวถึงความเป็นมา ความมุ่งหมายและวัตถุประสงค์ ขอบเขตของการศึกษา ประโยชน์ที่คาดว่าจะได้รับ และขั้นตอนการศึกษาของโครงการฯ

บทที่ 2 กล่าวถึงทฤษฎีพื้นฐานที่ใช้ในการศึกษา ซึ่งประกอบด้วยกระบวนการตัดสินใจ ขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคคาด้าไมน์นิ่งและเทคนิคคาด้าไมน์นิ่งแบบต่างๆ

บทที่ 3 กล่าวถึงวิธีการคัดเลือกและการเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคคาด้าไมน์นิ่งจากข้อมูลในระดับปฏิบัติการ

บทที่ 4 กล่าวถึงขั้นตอนการนำข้อมูลเข้าสู่ฐานข้อมูลใหม่ที่ใช้สำหรับการวิเคราะห์ข้อมูลด้วยเทคนิคคาด้าไมน์นิ่ง

บทที่ 5 กล่าวถึงซอฟต์แวร์สำหรับวิเคราะห์ข้อมูลด้วยเทคนิคคาด้าไมน์นิ่ง

บทที่ 6 กล่าวถึงการนำซอฟต์แวร์ดาต้าไมนิ่งมาใช้ในการวิเคราะห์ปัญหาตามความต้องการของระบบ พร้อมทั้งการนำเสนอรูปแบบของผลลัพธ์ที่ได้

บทที่ 7 เป็นบทสรุปผลการศึกษาของโครงการฯและข้อเสนอแนะ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

# ทฤษฎีพื้นฐาน

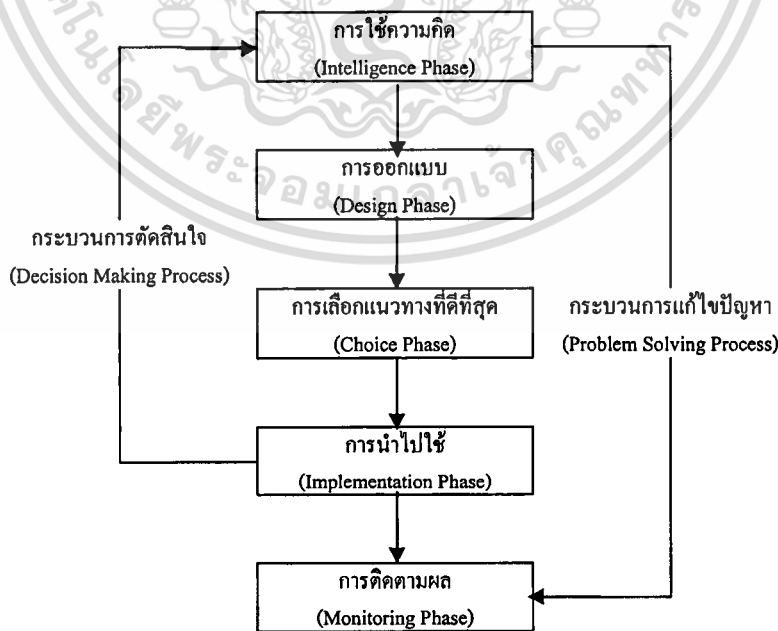
ในหัวข้อนี้จะกล่าวถึงทฤษฎีพื้นฐานต่างๆ ที่เกี่ยวข้องในออกแบบและพัฒนาระบบฯ ซึ่งเนื้อหาในบทนี้จะกล่าวถึงกระบวนการตัดสินใจ และเทคนิคที่นำมาใช้กับระบบการสนับสนุนการตัดสินใจซึ่งมีหลายเทคนิคด้วยกัน โดยในบทนี้จะนำเสนอเทคนิคคาด้าไมน์นึ่ง ซึ่งจะประกอบไปด้วยหลักการ ขั้นตอนการดำเนินการ เทคนิคคาด้าไมน์นึ่งและอัลกอริทึมแบบต่างๆ

### 2.1 การตัดสินใจและกระบวนการตัดสินใจ

องค์กรแต่ละองค์กรจำเป็นต้องตัดสินใจในการวางแผนกลยุทธ์ด้านต่างๆ เพื่อให้เข้าถึงความต้องการของลูกค้าให้ได้มากที่สุด ซึ่งการตัดสินใจนั้นต้องรวดเร็ว ถูกต้องและตรงกับเป้าหมายที่ต้องการ ดังนั้น ก่อนการตัดสินใจจึงควรมีกระบวนการในการคิดและตัดสินใจที่ดี

การตัดสินใจ (Decision Making) เป็นกระบวนการคัดเลือกแนวทางปฏิบัติจากทางเลือกต่างๆ เพื่อให้บรรลุวัตถุประสงค์ที่ต้องการ ซึ่งจัดเป็นส่วนหนึ่งของกระบวนการแก้ไขปัญหา

กระบวนการตัดสินใจ (Decision Making Process) เป็นการกำหนดขั้นตอนในการตัดสินใจแก้ไขปัญหาที่เกิดขึ้นภายในองค์กรอย่างมีหลักเกณฑ์ ด้วยการกำหนดขั้นตอนตั้งแต่ขั้นแรกจนถึงขั้นตอนสุดท้ายเพื่อให้ได้ผลลัพธ์ที่ต้องการ โดยมีขั้นตอนดังรูปที่ 2.1 (กิตติ ภัคดีวิณะกุล. 2550)



รูปที่ 2.1 กระบวนการตัดสินใจและแก้ปัญหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. การใช้ความคิด (Intelligence Phase) เป็นการค้นหาสาเหตุของปัญหา โดยศึกษาถึงต้นเหตุของปัญหา ประเมินผลที่จะเกิดขึ้นหากไม่ทำการแก้ไขปัญหา วิเคราะห์ปัจจัยแวดล้อมของปัญหา เพื่อสร้างแบบจำลองที่ใช้อธิบายลักษณะและสาเหตุของปัญหา ซึ่งผลที่ได้รับจากขั้นตอนนี้เรียกว่า การระบุปัญหา (Decision Statement)

2. การออกแบบ (Design Phase) เป็นการสร้างและวิเคราะห์ทางเลือกในการตัดสินใจ โดยทางเลือกที่สร้างขึ้นมาจะต้องมีความเป็นไปได้ในการแก้ไขปัญหาให้ได้ผลประโยชน์สูงสุด และต้องมีการกำหนดวัตถุประสงค์ของการตัดสินใจ ในขั้นตอนนี้อาจมีการสร้างแบบจำลอง (Model) แผนภาพการตัดสินใจ (Decision Tree) หรือตารางการตัดสินใจ (Decision Table) ก็ได้ทั้งนี้เพื่อใช้ในการพัฒนาทางเลือกในการตัดสินใจ

3. การเลือกแนวทางที่ดีที่สุด (Choice Phase) เป็นการประเมินทางเลือกต่างๆที่ได้จากขั้นตอนการออกแบบและคัดเลือกให้เหลือทางเลือกเดียว ซึ่งผลที่ได้รับจากขั้นตอนนี้คือ ทางเลือกเพื่อนำไปใช้จริงในการแก้ไขปัญหา

4. การนำไปใช้ (Implementation Phase) เป็นการนำทางเลือกในการแก้ไขปัญหามาได้จากขั้นตอนที่ 3 ไปลงมือปฏิบัติจริง ซึ่งอาจจะประสบผลสำเร็จหรือล้มเหลวก็ได้ หากพบว่าล้มเหลวก็จะย้อนกลับไปทบทวนกระบวนการในขั้นตอนก่อนหน้านี้ใหม่ได้

5. การติดตามผล (Monitoring Phase) เป็นขั้นตอนสุดท้ายของกระบวนการตัดสินใจ ในขั้นตอนนี้ผู้ตัดสินใจจะมีการประเมินผลหลังจากนำแนวทางที่ได้เลือกไปใช้ ซึ่งหากไม่เป็นที่น่าพอใจจะต้องพิจารณาถึงสาเหตุว่าเกิดขึ้นจากขั้นตอนใดหรือขาดสารสนเทศส่วนใดไปบ้าง เพื่อนำไปปรับปรุงกระบวนการตัดสินใจแก้ไขปัญหาลงครั้ง

การนำเทคนิคดาต้าไมนิงมาประยุกต์ใช้เพื่อสนับสนุนการตัดสินใจนั้นจะเป็นเทคนิคที่สอดคล้องกับกระบวนการตัดสินใจในเฟสการใช้ความคิด เฟสการออกแบบและเฟสการเลือกแนวทางที่ดีที่สุด จากนั้นเมื่อได้ผลลัพธ์แล้วจึงนำไปสู่เฟสการนำไปใช้และการติดตามผลต่อไป

โดยข้อมูลต่าง ๆ ที่ใช้ในระบอบสนับสนุนการตัดสินใจนั้นต้องทำการเก็บรวบรวมจากแหล่งข้อมูลหลาย ๆ แหล่ง เพื่อให้ระบบได้รับข้อมูลที่ครบถ้วน สมบูรณ์ สามารถนำไปประมวลผลเป็นสารสนเทศที่ดีและสามารถนำไปใช้ประโยชน์ได้อย่างสูงสุด โดยเครื่องมือที่จะมาช่วยวิเคราะห์ข้อมูลดังกล่าวเช่น คลังข้อมูล (Data Warehouse) เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลในหลายมิติ (OLAP) และดาต้าไมนิง (Data Mining) เป็นต้น

## 2.2 ความหมายและหลักการของดาต้าไมน์นิ่ง

ในอดีตการจะค้นหาข้อมูลที่มีประโยชน์จากฐานข้อมูลนั้นเป็นเรื่องยาก ยิ่งถ้าหากเป็นฐานข้อมูลที่มีขนาดใหญ่หลายๆ ก็จะต้องใช้เวลาในการค้นหามากและการที่เราจะค้นหาข้อมูลที่เป็นสารสนเทศที่เราต้องการจากแหล่งข้อมูลดิบที่มีมากมาย เราจำเป็นต้องมีเครื่องมือที่จะช่วยในการค้นหาสารสนเทศเหล่านั้น ซึ่งหนึ่งในเทคนิคแบบต่างๆ ก็คือ เทคนิคดาต้าไมน์นิ่ง (กฤษณะ ไวยมัย. 2549)

ดาต้าไมน์นิ่ง เป็นกระบวนการค้นหา วิเคราะห์สารสนเทศที่สำคัญ โดยจะค้นหารูปแบบหรือกฎเกณฑ์ที่เป็นประโยชน์และน่าสนใจที่ซ่อนอยู่ออกมาจากข้อมูลที่มีจำนวนมากมาย แล้วนำมาตั้งเป็นสมมติฐาน หลังจากนั้น ก็นำข้อมูลที่ต้องการทราบมาทำการทดสอบสมมติฐานที่สร้างไว้ เพื่อสนับสนุนการตัดสินใจในด้านต่างๆ

### 2.2.1 ลักษณะของสารสนเทศที่ได้

1. ต้องเป็นข้อมูลที่ไม่เคยรู้ล่วงหน้ามาก่อน (Unknown) หมายถึงข้อมูลสารสนเทศที่ได้รับนั้นต้องไม่เคยค้นพบมาก่อนหน้าและไม่สามารถคาดเดาได้ว่าผลที่ได้รับจะออกมาเป็นลักษณะใด
2. ต้องเป็นข้อมูลที่มีความถูกต้อง (Valid) หมายถึงสารสนเทศที่ได้รับต้องเป็นสารสนเทศที่มีความถูกต้อง เนื่องจากต้องนำไปใช้ประกอบกับข้อมูลส่วนอื่นๆ ดังนั้นต้องมีความถูกต้องน่าเชื่อถือ
3. สามารถนำไปใช้ประโยชน์ได้ (Actionable) หมายถึงต้องสามารถนำเอาข้อมูลและสารสนเทศที่ค้นพบออกไปใช้ประโยชน์ในด้านอื่นๆ ได้ เช่น นำมาช่วยตัดสินใจในการวางแผนการตลาด เพื่อสร้างความได้เปรียบทางการแข่งขันในเชิงธุรกิจ เป็นต้น

### 2.2.2 ปัจจัยที่ทำให้ดาต้าไมน์นิ่งเป็นที่นิยม

1. จำนวน ขนาดของข้อมูลใหญ่ และการขยายตัวอย่างรวดเร็วจากหลากหลายแหล่ง รวมถึงข้อมูลจากระบบเครือข่ายต่างๆ ไม่ว่าจะเป็นอินเทอร์เน็ต ดาวเทียม เครื่องอ่านบาร์โค้ด เครดิตการ์ดหรืออีคอมเมิร์ซ เป็นต้น
2. ความต้องการนำข้อมูลที่จัดเก็บไปสร้างระบบการสนับสนุนการตัดสินใจเพื่อให้ง่ายต่อการนำข้อมูลมาใช้ในการวิเคราะห์และตัดสินใจ
3. เนื่องจากเทคนิคดาต้าไมน์นิ่งประกอบไปด้วยอัลกอริทึมที่มีความซับซ้อนและความต้องการการคำนวณสูงจำเป็นต้องใช้งานกับระบบคอมพิวเตอร์ที่มีสมรรถนะสูงมาทำงาน และในปัจจุบันระบบคอมพิวเตอร์สมรรถนะสูงมีราคาต่ำลง จึงเป็นการเพิ่มโอกาสและศักยภาพการทำงานให้กับเทคนิคดาต้าไมน์นิ่งมากขึ้น
4. การแข่งขันในด้านอุตสาหกรรมและการค้า ในปัจจุบันมีการแข่งขันกันอย่างสูงทำให้มี

เอกสารที่เก็บสะสมข้อมูลเป็นจำนวนมากแต่ไม่ได้นำมาใช้ให้เกิดประโยชน์ จึงเป็นโอกาสที่ดีในการนำข้อมูลที่ถูกละทิ้ง ไม่ว่าจะเป็นข้อมูลเก่าหรือข้อมูลใหม่ให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จัดการและสืบค้นสารสนเทศที่ซ่อนอยู่ เพื่อนำสารสนเทศที่ได้ไปช่วยสนับสนุนการตัดสินใจในการจัดการระบบงานต่างๆ

### 2.2.3 ประเภทข้อมูลที่สามารถทำดาต้าไมน์นิ่งได้

1. ฐานข้อมูลเชิงสัมพันธ์ (Relational Databases) เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบข้อมูลที่สัมพันธ์กันสามารถแสดงให้เห็นได้ในรูปแบบของตารางข้อมูลโดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของตารางข้อมูลสามารถแสดงได้โดยใช้โมเดลความสัมพันธ์ระหว่างเอนทิตี (Entity-Relationship Model)

2. คลังข้อมูล เป็นการเก็บข้อมูลในลักษณะหลายมิติ หรือ คิวบ์ (Cube) โดยมีการรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบโครงสร้างใหม่ที่สร้างขึ้นเพื่อให้สะดวกต่อการวิเคราะห์ข้อมูลแบบออนไลน์ (On-Line Analytical Processing)

3. ฐานข้อมูลแบบทรานแซกชัน (Transactional Databases) เป็นฐานข้อมูลที่ประกอบไปด้วยข้อมูลที่แต่ละทรานแซกชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลซื้อลูกค้าและรายการสินค้าที่ลูกค้ารายนั้นซื้อ เป็นต้น

4. ฐานข้อมูลขั้นสูง (Advanced Database) เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่นๆ เช่น ข้อมูลเชิงวัตถุ (Object-Oriented) ข้อมูลที่เป็นไฟล์ข้อมูล ข้อมูลมัลติมีเดีย ข้อมูลในรูปแบบของเว็บไซต์ เป็นต้น

### 2.2.4 ลักษณะเฉพาะของข้อมูลที่สามารถทำดาต้าไมน์นิ่ง

1. ข้อมูลขนาดใหญ่เกินกว่าจะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า หรือ โดยการใช้ Database Management System (DBMS) ในการจัดการฐานข้อมูล

2. ข้อมูลที่มาจากหลายแหล่ง โดยอาจรวบรวมมาจากหลายระบบปฏิบัติการหรือหลาย DBMS เช่น Oracle, DB2, MS SQL, MS Access เป็นต้น

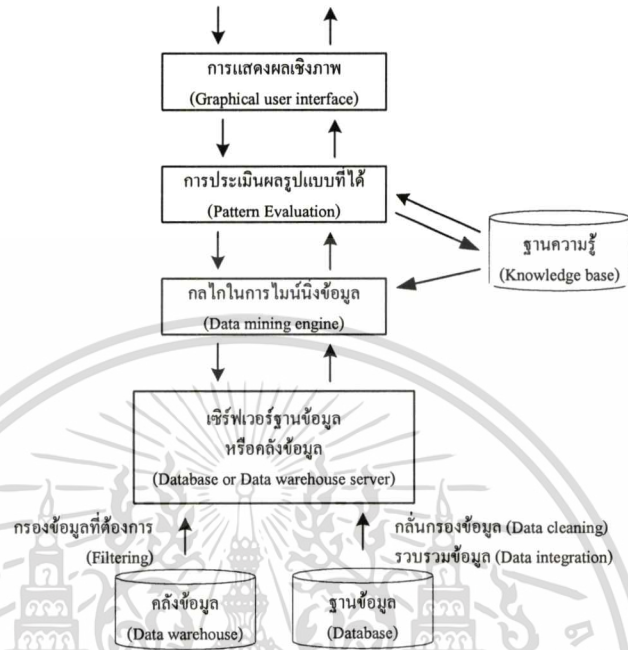
3. ข้อมูลที่ไม่มีการเปลี่ยนแปลง ตลอดช่วงเวลาที่ทำดาต้าไมน์นิ่ง หากข้อมูลที่มีอยู่นั้นเป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาจะต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้นไว้และนำฐานข้อมูลที่บันทึกไว้มาทำดาต้าไมน์นิ่ง แต่เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลา จึงทำให้ผลลัพธ์ที่ได้จากการทำดาต้าไมน์นิ่ง สมเหตุสมผลในช่วงเวลาหนึ่งเท่านั้น ดังนั้น เพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องเหมาะสมอยู่ตลอดเวลาจึงต้องทำดาต้าไมน์นิ่งใหม่ทุกครั้งในช่วงเวลาที่เหมาะสม ซึ่งช่วงเวลาที่เหมาะสมนั้นขึ้นอยู่กับลักษณะของธุรกิจในแต่ละองค์กร

4. ข้อมูลที่มีโครงสร้างซับซ้อน เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย ข้อมูลเหล่านี้สามารถนำมาทำดาต้าไมน์นิ่งได้เช่นกันแต่ต้องใช้เทคนิคการทำดาต้าไมน์นิ่งขั้นสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.2.5 สถาปัตยกรรมพื้นฐานของดาต้าไมนิง

สถาปัตยกรรมพื้นฐานของดาต้าไมนิงมีส่วนประกอบหลักดังรูปที่ 2.2 (กฤษณะ ไวยมัย, 2549)



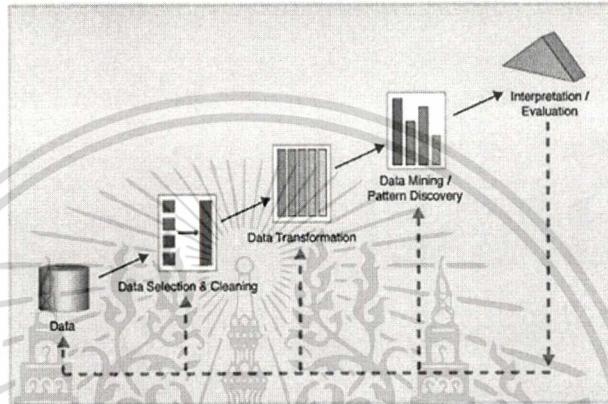
รูปที่ 2.2 สถาปัตยกรรมพื้นฐานของดาต้าไมนิง

1. ฐานข้อมูล คลังข้อมูล และแหล่งเก็บสะสมข้อมูลอื่นๆ หมายถึงฐานข้อมูลตั้งแต่หนึ่งฐานข้อมูลขึ้นไปหรือกลุ่มของฐานข้อมูล คลังข้อมูล สเปรดชีตหรือแหล่งที่เก็บข้อมูลในรูปแบบอื่นๆ การได้มาซึ่งข้อมูลนั้นต้องผ่านกระบวนการเตรียมข้อมูล โดยใช้เทคนิคกลั่นกรองข้อมูล และการรวบรวมข้อมูลก่อน
2. เซิร์ฟเวอร์ฐานข้อมูลหรือคลังข้อมูล จะทำหน้าที่ในการให้บริการข้อมูลที่ต้องการตามที่ได้ร้องขอในกระบวนการทำดาต้าไมนิงนั่นเอง
3. ฐานความรู้ จะเป็นที่เก็บรวบรวมความรู้ หรือรูปแบบกฎเกณฑ์ที่ได้ค้นพบ
4. กลไกในการทำดาต้าไมนิง เป็นกระบวนการในการนำเทคนิคดาต้าไมนิงมาใช้กับข้อมูลอย่างเหมาะสม
5. การประเมินผลรูปแบบที่ได้ เป็นขั้นตอนการแปลความหมายและประเมินผลลัพธ์ที่ได้ นั้นมีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ ซึ่งต้องใช้ทักษะการวิเคราะห์เชิงธุรกิจเข้าช่วย
6. การแสดงผลเชิงภาพ เป็นขั้นตอนการนำผลลัพธ์ที่ได้มาแสดงผลโดยทั่วไปควรมีการแสดงผลในรูปแบบที่สามารถเข้าใจได้โดยง่าย เช่น กราฟ แผนภูมิเชิงมิติ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.3 ขั้นตอนการทำดาต้าไมนนิ่ง

ขั้นตอนในการทำดาต้าไมนนิ่งหรือเรียกอีกอย่างหนึ่งว่า กระบวนการค้นหาความรู้จากฐานข้อมูล (Knowledge Discovery in Database) ซึ่งเป็นกระบวนการสร้างแบบจำลองหรือรูปแบบกฎเกณฑ์จากกลุ่มข้อมูล ทำให้เกิดความเข้าใจในลักษณะรูปแบบความเกี่ยวข้องสัมพันธ์กันของกลุ่มข้อมูลและแนวโน้มเพื่อใช้ในการทำนายข้อมูลนั้นๆ โดยมีขั้นตอนการทำงานที่สำคัญดังรูปที่ 2.3 (กฤษณะ ไวยมัย. 2549)



รูปที่ 2.3 ขั้นตอนการทำดาต้าไมนนิ่ง

### 2.3.1 การกำหนดวัตถุประสงค์ทางธุรกิจ

การกำหนดวัตถุประสงค์ทางธุรกิจ (Business Objective Determination) เป็นกระบวนการขั้นแรกของการทำดาต้าไมนนิ่ง ซึ่งจะนำไปสู่การกำหนดรูปแบบและขอบเขตข้อมูลเบื้องต้น การเลือกเทคนิควิธีการ เครื่องมือที่ใช้ รวมถึงแนวทางที่ชัดเจนในการทำงาน แม้ว่าเราจะยังไม่ทราบผลลัพธ์ที่ได้อย่างชัดเจนก็ตาม โดยการกำหนดวัตถุประสงค์อาจจะมาจากปัญหาที่เกิดขึ้นในองค์กรที่ต้องการแก้ไขก็ได้

### 2.3.2 การเตรียมข้อมูล

การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่สำคัญและใช้เวลานานที่สุดของเวลาทั้งหมด เนื่องจากข้อมูลที่ผิดพลาดจะทำให้ผลลัพธ์ที่ได้ผิดพลาดตามไปด้วย โดยกระบวนการเตรียมข้อมูลสำหรับทำดาต้าไมนนิ่งมีวัตถุประสงค์เพื่อให้ข้อมูลมีความถูกต้องสมบูรณ์และเหมาะสมกับการทำดาต้าไมนนิ่งนั่นเอง การเตรียมข้อมูลสามารถแบ่งออกได้เป็น 3 ขั้นตอนคือ

#### 1. การคัดเลือกข้อมูล

การคัดเลือกข้อมูล (Data Selection) มีจุดประสงค์เพื่อเลือกข้อมูลที่ถูกต้องเหมาะสมมาทำการวิเคราะห์ให้ได้ตรงตามความต้องการ ซึ่งทั่วไปผู้ที่เลือกข้อมูลควรจะเป็นผู้เชี่ยวชาญหรือเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกี่ยวข้องกับข้อมูลในธุรกิจนั้นๆ การเลือกข้อมูลจำเป็นต้องมีความเข้าใจเกี่ยวกับคุณสมบัติของข้อมูล ดังนี้

- ชนิดของข้อมูลที่จัดเก็บ ซึ่งมี 2 ลักษณะ คือ ข้อมูลที่เป็นตัวเลข เช่น อายุ ราคา จำนวนที่ซื้อ เป็นต้น และข้อมูลที่เป็นตัวอักษร เช่น ชื่อ ที่อยู่ เป็นต้น
- ประเภทของข้อมูล (Categorical) คือ Nominal เป็นข้อมูลที่ลำดับของข้อมูลไม่มีผลกับค่า เช่น เพศ (ชาย, หญิง) เป็นต้น และ Ordinal เป็นข้อมูลที่ลำดับของข้อมูลมีผลกับค่า เช่น ลำดับของสินค้า เป็นต้น
- คุณสมบัติของข้อมูล คือคุณสมบัติเชิงปริมาณ (Quantitative Attribute) เป็นข้อมูลที่แสดงให้เห็นค่าตัวเลขที่เป็นปริมาณชัดเจนของตัวแปรนั้น และคุณสมบัติเชิงประเภท (Categorical Attribute) เป็นข้อมูลที่แสดงตามกลุ่มประเภท ลำดับชั้น หรือช่วงปริมาณข้อมูลที่เป็นตัวอักษร

ข้อพิจารณาเบื้องต้นในการคัดเลือกข้อมูล ได้แก่

- เลือกเฉพาะข้อมูลที่สนใจ
- ไม่นำคอลัมน์ที่มีค่าสำหรับทุกแถวเป็นค่าเดียวกันมาใช้
- ตัดคอลัมน์ที่มีค่าไม่ซ้ำกันเลขออก เนื่องจากข้อมูลเหล่านี้ไม่สามารถหาแถวที่มีข้อมูลสัมพันธ์กันได้เลย ยกเว้นบางคอลัมน์ที่ต้องการนำไปใช้จริง แต่จะมีการจัดการกับข้อมูลนี้ในลำดับต่อไป
- คอลัมน์ที่มีค่าส่วนมากเป็นอย่างเดียวกัน จะต้องพิจารณาว่าค่าที่แตกต่างกันนั้นสำคัญหรือไม่ หากข้อมูลส่วนน้อยนั้นไม่สำคัญก็สามารถตัดคอลัมน์นั้นทิ้งไปได้
- คุณสมบัติหรือแอตทริบิวต์บางอย่างที่อยู่ในตารางข้อมูลสามารถใช้สร้างเป็นคุณสมบัติใหม่ได้ โดยอาศัยเงื่อนไขต่าง ๆ มากำหนด

## 2. การทำความสะอาดข้อมูล

วัตถุประสงค์ของการทำความสะอาดข้อมูล (Data Cleaning) คือการทำให้ข้อมูลที่คัดเลือกมานั้นมีความถูกต้องและเหมาะสมที่จะนำไปวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นึ่งมากที่สุด โดยข้อผิดพลาดที่อาจเกิดขึ้นกับข้อมูลมีดังนี้

- ค่าที่ขาดหาย (Missing Values) สามารถเกิดขึ้นได้จากการเก็บข้อมูลทั่วไป โดยอาจมีค่าในบางคอลัมน์ขาดหายไป ซึ่งอาจเกิดจากสาเหตุต่างๆ เช่น ผู้ใช้บันทึกข้อมูลไม่ครบถ้วน หรือเป็นข้อมูลที่สามารรถเว้นว่างได้ เป็นต้น

- ข้อมูลที่มีสิ่งรบกวน (Noisy Data) เป็นความผิดพลาดที่เกิดขึ้นแบบไม่ตั้งใจ หรือแปรปรวนไปจากที่ควรจะเป็นเช่น ข้อมูลอายุเป็น 200 ปี ซึ่งไม่น่าจะเป็นไปได้

- ข้อมูลที่ขัดแย้งกัน (Inconsistent Data) เป็นข้อมูลที่ขัดแย้งกันหรือไม่สอดคล้องตรงกันของข้อมูล เช่น ข้อมูลจากเอกสารทางการขายสินค้ากับข้อมูลที่ป้อนเข้าสู่ระบบคอมพิวเตอร์ซึ่งอาจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกิดจากการป้อนข้อมูลผิดพลาด เป็นต้น

เมื่อมีความผิดพลาดเกิดขึ้นกับข้อมูลจำเป็นต้องปรับปรุงข้อมูลให้มีค่าสอดคล้องกันกับข้อมูลเดิมที่มีอยู่ เช่น ข้อมูลในคอลัมน์ที่ขาดหายไป ควรแทนค่าข้อมูลที่ขาดหายไปที่ละค่าจนครบทุกรายการ ซึ่งการแทนค่าอาจจะแทนค่าด้วย “Unknown” หรือ “0” เหมือนกันทั้งหมด หรือใช้ค่าเฉลี่ยรวมแทนที่ทั้งหมด แต่ทั้งนี้ในทางปฏิบัติจริงวิธีการนี้จะเสียเวลามากและไม่เหมาะกับข้อมูลที่ขาดหายไปเป็นจำนวนมาก

### 3. การรวบรวมข้อมูลและแปลงข้อมูล

เนื่องจากข้อมูลที่น่ามาใช้อาจอยู่ในฐานข้อมูลหลากหลายที่แตกต่างกัน ทำให้ต้องมีขั้นตอนที่ทำการรวบรวมข้อมูล (Data Integration) หรือแปลงข้อมูล (Data Transformation) เพื่อให้ข้อมูลอยู่ในระดับเดียวกัน และอยู่ในรูปแบบที่เหมาะสมกับอัลกอริทึมที่ใช้ในการทำค้ำไมน์นิ่ง เช่น การรวบรวมข้อมูลที่ถูกเก็บอยู่ในลักษณะหลายระดับในหลายๆตาราง จำเป็นต้องพิจารณารวบรวมข้อมูลอยู่ในระดับเดียวกันหรือตารางเดียวกัน การแปลงข้อมูลเป็นการกำหนดรูปแบบของข้อมูล โดยมีเงื่อนไขที่สอดคล้องกับการวิเคราะห์ข้อมูลในเรื่องนั้นๆ เพื่อให้ง่ายต่อการวิเคราะห์ข้อมูล

#### 2.3.3 การวิเคราะห์ข้อมูลด้วยเทคนิคค้ำไมน์นิ่ง

เทคนิคค้ำไมน์นิ่ง เป็นขั้นตอนในการประมวลผลข้อมูลตามวิธีและอัลกอริทึมที่เลือกไว้ให้มีความเหมาะสมกับการใช้งานซึ่งอาจจะต้องใช้วิธีการและเทคนิคต่างๆมารวมกันเพื่อให้ได้ผลลัพธ์ที่ดีซึ่งการดำเนินการที่นิยมใช้โดยทั่วไปมีดังรูปที่ 2.4

|              | Market Management   |                                  | Risk Management |                                 | Fraud Management       |               |
|--------------|---------------------|----------------------------------|-----------------|---------------------------------|------------------------|---------------|
| Applications | ✓                   | Target Marketing                 | ✓               | Forecasting                     | ✓<br>Fraud<br>Detecton |               |
|              | ✓                   | Customer Relationship Management | ✓               | Customer Relation               |                        |               |
|              | ✓                   | Market Basket Analysis           | ✓               | improved Underwriting           |                        |               |
|              | ✓                   | Cross Selling                    | ✓               | Quality Control                 |                        |               |
|              | ✓                   | Market Segmentation              | ✓               | Competitive Analysis            |                        |               |
| Operations   | Predictive Modeling | Database Segmentation            | Link Analysis   |                                 | Deviation Detection    |               |
| Techniques   | ✓                   | Classification                   | ✓               | Demographic Clustering          | ✓                      | Visualization |
|              | ✓                   | Value Prediction                 | ✓               | Neural Clustering               | ✓                      | Statistic     |
|              |                     |                                  | ✓               | Associations Discovery          |                        |               |
|              |                     |                                  | ✓               | Sequential Pattern Discovery    |                        |               |
|              |                     |                                  | ✓               | Similar Time Sequence Discovery |                        |               |

รูปที่ 2.4 ประเภทของเทคนิคค้ำไมน์นิ่งและตัวอย่างการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. Predictive Modeling เป็นการนำข้อมูลมาใช้ในการสร้างโมเดล เพื่อนำไปใช้ในการทำนายค่า แบ่งออกได้เป็น 2 เทคนิคคือ

- Classification เป็นการทำนายข้อมูลในอนาคตว่าข้อมูลที่ต้องการพิจารณาควรจะอยู่ในกลุ่มใด โดยมีการแบ่งประเภทกลุ่มไว้ก่อนแล้ว

- Forecasting เป็นโมเดลที่ใช้ในการทำนายแนวโน้มข้อมูลที่เป็นตัวเลขในอนาคต เช่น การพยากรณ์อากาศ การทำนายราคาหุ้น เป็นต้น

2. Database Segmentation เป็นการแบ่งข้อมูลออกเป็นกลุ่มย่อยๆ โดยที่ข้อมูลภายในแต่ละกลุ่มมีลักษณะที่เหมือนกันหรือใกล้เคียงกัน โดยที่เรายังไม่เคยรู้มาก่อน เช่น ใช้ในการแบ่งกลุ่มของลูกค้าว่ามีจำนวนกี่กลุ่ม

3. Link Analysis เป็นการหาความสัมพันธ์ของข้อมูลในแต่ละเรคอร์ด หรือกลุ่มของเรคอร์ดในฐานข้อมูล เช่น การหาความสัมพันธ์ของสินค้าว่าลูกค้ามักจะทำซื้อสินค้าอะไรควบคู่กันในการซื้อหนึ่งครั้ง (Association Rule) หรือการศึกษาการซื้อสินค้าในระยะยาว (Sequential Pattern Discovery)

4. Deviation Detection เป็นเทคนิคที่ใช้แสดงข้อมูลที่มีลักษณะผิดปกติไปจากข้อมูลทั่วไปแบ่งเป็น 2 ประเภทดังนี้

- Visualization เป็นเทคนิคที่ใช้ในการแสดงข้อมูลในรูปแบบต่างๆ เช่น แผนที่รูปภาพ กราฟสามมิติ ซึ่งมีประสิทธิภาพในการสื่อสารค่อนข้างมาก

- Statistics เป็นการใช่วิธีทางสถิติเข้ามาช่วยตรวจสอบข้อมูล

### 2.3.4. การประเมินรูปแบบและกฎเกณฑ์ที่ได้

การประเมินรูปแบบและกฎเกณฑ์ที่ได้ (Pattern Evaluation) เป็นขั้นตอนการแปลความหมายและประเมินผลลัพธ์ที่ได้ นั่นว่ามีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ ซึ่งต้องใช้ทักษะการวิเคราะห์ธุรกิจเข้าช่วย โดยทั่วไปควรมีการแสดงผลในรูปแบบที่สามารถเข้าใจได้โดยง่าย

### 2.3.5. การนำเสนอความรู้

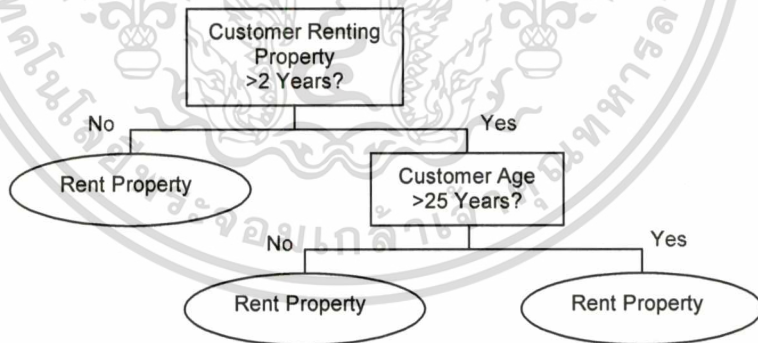
การนำเสนอความรู้ (Knowledge Presentation) เป็นขั้นตอนการนำผลลัพธ์ที่ได้ไปประยุกต์ใช้งานจริงต่อไป โดยภายหลังจากการใช้งานจริงควรมีการติดตามผลอย่างต่อเนื่องด้วย

## 2.4 เทคนิคดาต้าไมน์นิ่ง

การแก้ปัญหาของงานชนิดต่างๆ โดยใช้เทคนิคดาต้าไมน์นิ่งในแต่ละงานนั้น จะใช้เทคนิคดาต้าไมน์นิ่งที่ไม่เหมือนกัน ซึ่งขึ้นอยู่กับความเหมาะสมของข้อมูล และผลลัพธ์ที่คาดว่าจะได้รับ โดยเทคนิคดาต้าไมน์นิ่งนั้นมีมากมาย ซึ่งจะขอยกตัวอย่างดังนี้ (Rob and Coronel. 2000)

### 2.4.1 แผนภาพต้นไม้

แผนภาพต้นไม้ (Decision Tree) เป็นการนำโครงสร้างต้นไม้ มาสร้างแบบจำลองในการตัดสินใจซึ่งเป็นวิธีการแก้ปัญหาทางตรรกศาสตร์มีลักษณะเหมือนแผนภูมิองค์กร ตัวอย่างแผนภาพต้นไม้ดังรูปที่ 2.5 แสดงให้เห็นถึงการวิเคราะห์ห้กลุ่มลูกค้าบ้านเช่า โดยใช้ปัจจัยคือ อายุของลูกค้า และระยะเวลาที่เช่าเพื่อวิเคราะห์ความสนใจที่จะซื้อบ้านเป็นของตนเอง ซึ่งวิธีการทำงานของแผนภาพต้นไม้ จะใช้อัลกอริทึมที่มีหลักการค้นหาแอดทริบิวต์ที่สำคัญที่สุดเป็น Root Node เพื่อแบ่งข้อมูลว่าข้อมูลใดควรอยู่กลุ่มใดจากรูปคือ ระยะเวลาที่เช่าซึ่งทำให้ที่ Root Node เมื่อทดสอบข้อมูล จะมีการแบ่งข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มที่มีระยะเวลาเช่าไม่เกิน 2 ปี (Left Node) และกลุ่มที่มีระยะเวลาเช่ามากกว่า 2 ปี (Right Node) จากนั้น Descendent Node ซึ่งเป็นแอดทริบิวต์ที่เหมาะสมและถูกเลือกในระดับความสำคัญถัดมา ซึ่งจะตัดสินใจต่อไปในระดับนี้ก็คือ อายุของลูกค้า โดยระบบจะทำการทดสอบไปเช่นนี้จนกระทั่งสิ้นสุดแผนภาพต้นไม้ สิ่งที่ต้องคำนึงถึงคือ ลำดับของ Descendent Node ซึ่งจะต้องเรียงลำดับ



รูปที่ 2.5 ตัวอย่างแผนภาพต้นไม้เพื่อวิเคราะห์โอกาสที่ลูกค้าผู้เช่าบ้านจะซื้อบ้าน

#### ข้อดีของแผนภาพต้นไม้

1. วิธีการและหลักการเข้าใจได้ง่าย
2. ในการแบ่งกลุ่มข้อมูลกระทำได้ง่ายและสามารถเลือกแอดทริบิวต์ที่ดีที่สุดในการแบ่งได้

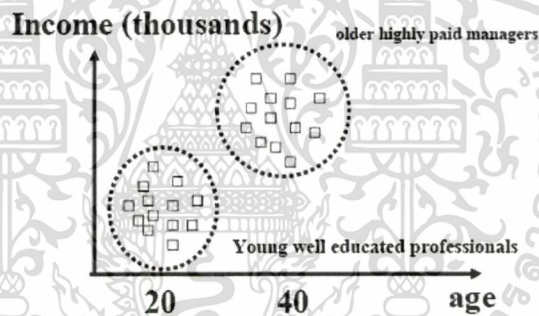
#### ข้อเสียของแผนภาพต้นไม้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ความน่าเชื่อถือจะลดลงเมื่อระดับของแผนภาพต้นไม้มีความซับซ้อนและมีจำนวนโหนดมากขึ้น ซึ่งจะก่อให้เกิดปัญหา Overfilling หรือ Overstraining
2. ข้อมูลที่จัดกลุ่มจะน้อยลงทำให้เกิดปัญหา Fragmentation ตามมา
3. กรณีข้อมูลเป็นค่าต่อเนื่อง เช่น รายได้ อายุ ซึ่งการจัดเก็บค่าของข้อมูลอาจเป็นช่วง เมื่อจัดกลุ่มอาจไม่สามารถทำได้

#### 2.4.2 คลัสเตอร์ลิ่ง

คลัสเตอร์ลิ่ง (Clustering) เป็นวิธีการที่นำเอาข้อมูลมาแบ่งเป็นกลุ่ม ที่มีความสัมพันธ์กัน เพื่อวิเคราะห์หาลักษณะที่เหมือน หรือแตกต่างกันของข้อมูลในแต่ละกลุ่ม โดยไม่รู้ล่วงหน้าว่าจะมีทั้งหมดกี่กลุ่มที่รูปแบบถ้ารายการในข้อมูลมีลักษณะคล้ายคลึงเป็นกลุ่มเดียวกันได้ก็จะรวมเข้าด้วยกัน เพื่อให้ง่ายต่อการวิเคราะห์ เช่น ข้อมูลประวัติลูกค้า หรือกลุ่มเป้าหมายทางการตลาด เป็นต้น อัลกอริทึมที่ใช้ในการทำคลัสเตอร์ลิ่งได้แก่ K-Mean และ Kohonen Feature Maps เป็นต้น โดยผลลัพธ์จากการทำคลัสเตอร์ลิ่งสามารถแสดงได้ดังรูปที่ 2.6



รูปที่ 2.6 ตัวอย่างแผนภาพการแบ่งกลุ่มข้อมูล

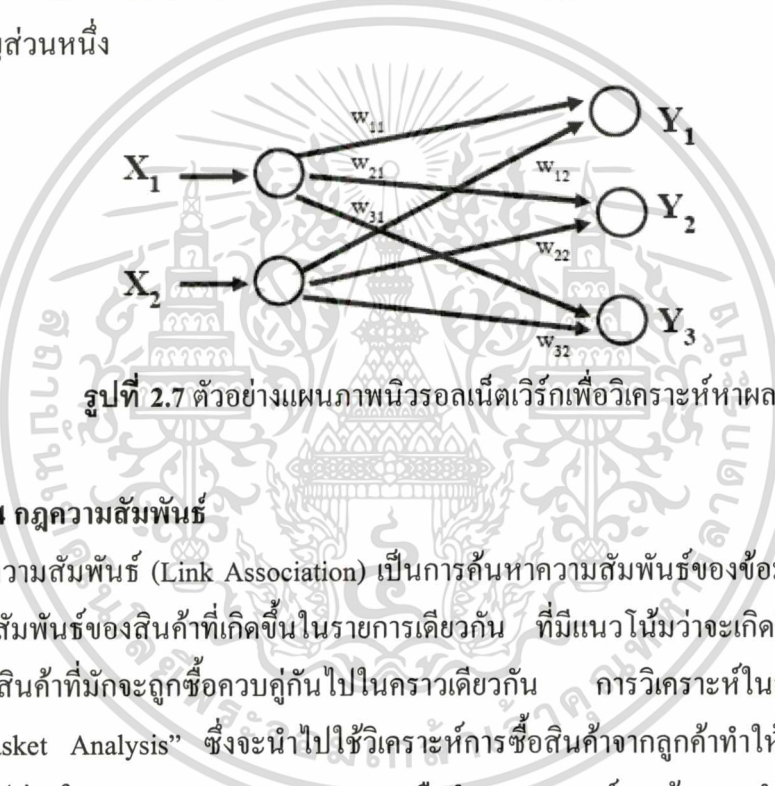
#### 2.4.3 นิวรอลเน็ตเวิร์ก

นิวรอลเน็ตเวิร์ก (Neural Network) เป็นระบบที่มีการประมวลผลข้อมูลโดยรวมคุณสมบัติของไบโอลอจิกคอลล ซึ่งถูกพัฒนาขึ้นโดยโมเดลทางคณิตศาสตร์ของกระบวนการเรียนรู้ของมนุษย์ เป็นการเลียนแบบการทำงานของสมองมนุษย์ และจะเรียนรู้จากชุดข้อมูลความรู้เทรนนิ่งเซต (Training Set) นิวรอลเน็ตเวิร์ก ประกอบด้วยหน่วยความจำจำนวนมากเรียกว่า นิวรอนเซลล์ หรือ โหนด แต่ละนิวรอนต่อกันโดยคอนเนกชันลิงค์ (Connection Link) ที่มีค่าน้ำหนักของมันอยู่ ในแต่ละการเชื่อมต่อ โดยค่าน้ำหนักจะแสดงรายละเอียดที่เน็ตเวิร์กใช้ในการแก้ปัญหา เครือข่ายนิวรอนประกอบด้วยส่วนนำเข้าข้อมูล (Input Layer) ส่วนแสดงผลลัพธ์ (Output Layer) และส่วนประมวลผลซ่อนพราง (Hidden Layer) ระบบเครือข่ายนิวรอนจะถูกป้อนด้วยชุดข้อมูลที่มีการเตรียมไว้ล่วงหน้าซึ่งจะสร้างผลลัพธ์หรือข้อสรุปที่กำหนดโดยกระบวนการนี้เป็นการสอนให้ระบบ

หรือคอมพิวเตอร์เรียนรู้สิ่งที่ถูกต้องจากตัวอย่างที่กำหนดให้ เมื่อคอมพิวเตอร์ได้รับข้อมูลมากขึ้นก็  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลแต่ละชุดจะถูกนำไปเปรียบเทียบกับคุณลักษณะของข้อมูลที่ทราบล่วงหน้าแล้ว ถ้าหากมีความแตกต่างกันส่วนประมวลผลที่ซ่อนพรางไว้ก็จะทำหน้าที่แก้ไขให้ถูกต้อง กระบวนการนี้จะเกิดขึ้นจนกว่าเงื่อนไขบางอย่างได้รับการตอบสนองเรียบร้อยแล้วจึงจะหยุดทำงาน เทคนิคนิวรอลเน็ตเวิร์กนี้นิยมนำมาใช้ในการตรวจสอบพฤติกรรมการณ์ของลูกค้าบัตรเครดิต หรือลูกค้าสินเชื่อประเภทต่างๆ โดยแผนภาพนิวรอลเน็ตเวิร์กแสดงได้ดังรูปที่ 2.7 และแม้ว่าจะสามารถนำไปประยุกต์ใช้กับงานหลายๆ ชนิดได้อย่างมีประสิทธิภาพ แต่นิวรอลเน็ตเวิร์กก็ยังมีข้อเสียอยู่บ้าง ดังนี้

1. เป็นวิธีที่ยากต่อการทำความเข้าใจในโมเดลที่ถูกผลิตออกมา
2. มีคุณสมบัติที่ไวต่อรูปแบบของอินพุท ถ้าเราแทนข้อมูลด้วยรูปแบบที่แตกต่างกันจะทำให้สามารถผลิตผลลัพธ์ที่แตกต่างกันออกมา ดังนั้น การกำหนดค่าเริ่มต้นให้กับข้อมูลจึงเป็นส่วนที่มีความสำคัญส่วนหนึ่ง



รูปที่ 2.7 ตัวอย่างแผนภาพนิวรอลเน็ตเวิร์กเพื่อวิเคราะห์หาผลลัพธ์

### 2.4.4 กฎความสัมพันธ์

กฎความสัมพันธ์ (Link Association) เป็นการค้นหาความสัมพันธ์ของข้อมูล ซึ่งนิยมใช้ในการหาความสัมพันธ์ของสินค้าที่เกิดขึ้นในรายการเดียวกัน ที่มีแนวโน้มว่าจะเกิดขึ้นพร้อมๆ กัน เช่น พิจารณาสินค้าที่มักจะถูกซื้อควบคู่กันไปในคราวเดียวกัน การวิเคราะห์ในลักษณะนี้เรียกว่า “Market Basket Analysis” ซึ่งจะนำไปใช้วิเคราะห์การซื้อสินค้าจากลูกค้าทำให้ผู้ประกอบการสามารถนำไปช่วยในการวางแผนทางการตลาด หรือกำหนดกลยุทธ์ทางการจำหน่ายสินค้าและบริการได้เช่น การจัด โปรโมชัน การวางตำแหน่งของสินค้า เป็นต้น

ผลลัพธ์ที่ได้จากการทำคาน่าไมน์นึ่งด้วยเทคนิคนี้จะได้ออกมาเป็นการหากฎความสัมพันธ์ (Association rules) ซึ่งรูปแบบของกฎความสัมพันธ์ สามารถแสดงได้ดังนี้

If X Then Y

When Condition 1 then Condition 2

$X \Rightarrow Y$

เมื่อ X หรือ Condition 1 คือกฎหรือเหตุการณ์ที่เกิดขึ้นก่อน (Antecedent)

Y หรือ Condition 2 คือ ผลที่ตามมา (Consequent)

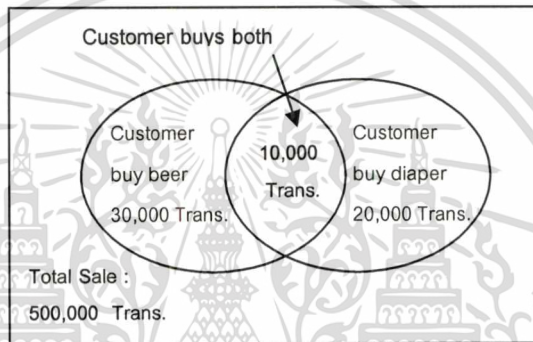
เอกสารนี้เป็นเอกสารที่เกี่ยวกับข้องในการสร้างกฎนี้ได้แก่ศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ค่าความถี่ของเหตุการณ์ X และ Y ที่เกิดขึ้นจากจำนวนเหตุการณ์ทั้งหมด (Support Factor)
- ค่าความน่าเชื่อถือของเหตุการณ์ Y ที่จะนำเกิดขึ้น เทียบจากเหตุการณ์ X และ Y พร้อมกัน (Confidence Factor)

**ตัวอย่างกฎความสัมพันธ์**

Buys (X, "diapers") => buys (X, "beers") [2%, 50%]

แปลความหมายได้ว่า มีลูกค้าที่นิยมซื้อผ้าอ้อมและเบียร์ไปพร้อมกันด้วยความถี่ 2% ของรายการที่ซื้อสินค้าทั้งหมดและมีรายการที่ซื้อผ้าอ้อมแล้วซื้อเบียร์ไปด้วย ค่าความน่าเชื่อถือ 50% ด้วยกัน การขายสินค้าประเภทผ้าอ้อมและเบียร์ไปพร้อมๆกัน แสดงได้ดังรูปที่ 2.8



**รูปที่ 2.8 ผลการขายสินค้าประเภทผ้าอ้อมและเบียร์**

โดยส่วนใหญ่ กฎความสัมพันธ์ที่น่าสนใจ คือกฎที่มีค่าความน่าเชื่อถือที่สูง เนื่องจากมีโอกาที่จะเกิดขึ้นสูงตามด้วยและนอกจากความถี่และค่าความน่าเชื่อถือ ยังมีตัววัดค่าความน่าเชื่อถือของกฎที่สร้างขึ้นเรียกว่า ลิฟต์

ค่าลิฟต์ จะแสดงถึงความสำคัญของความสัมพันธ์หรือเหตุการณ์ว่ามีมากน้อยเพียงใด โดยหาได้จากผลหารของค่าความน่าเชื่อถือกับจำนวนข้อมูลที่มีรายการซื้อสินค้าอย่างที 2 ซึ่งหากค่าลิฟต์มีค่ามากกว่า 1 หมายถึงเหตุการณ์นั้นน่าสนใจ และในกรณีที่ค่าลิฟต์มากหรือน้อยเกินไปอาจพิจารณาได้ว่า กฎนั้นไม่เป็นจริง

ผลที่ได้จากการหากฎความสัมพันธ์จะมีอยู่มากมาย หลากหลายจึงต้องมีการกำจัดหรือตัดกฎบางกฎที่ไม่น่าสนใจออก หรือมีค่าที่น่าสนใจน้อยออกเพื่อเป็นการลดจำนวนกฎที่มีโอกาสที่จะเกิดขึ้นน้อยออกไป โดยการกำหนดค่าความถี่น้อยที่สุด (Minimum Support) และค่าความน่าเชื่อถือน้อยที่สุด (Minimum Confidence) ซึ่งหากค่าความถี่และค่าความน่าเชื่อถือที่ได้จากกฎความสัมพันธ์มีค่าน้อยกว่าที่กำหนดจะไม่นำมาพิจารณา ทำให้เวลาที่ใช้ในการหากฎความสัมพันธ์สั้นลง และในการกำหนดค่าความถี่น้อยที่สุดและค่าความน่าเชื่อถือน้อยที่สุดนี้ ขึ้นอยู่กับสถานการณ์ หรือ

**ข้อดี**

● ผู้ใช้สามารถควบคุมจำนวนผลลัพธ์ได้โดยการระบุค่าความถี่น้อยที่สุดและค่าความน่าเชื่อถือน้อยที่สุด

- สามารถทำงานได้ดีกับข้อมูลขนาดใหญ่
- ในกรณีที่ข้อมูลไม่สมบูรณ์ ก็สามารถทำการไต่เน่กับข้อมูลบางส่วนได้
- ไม่จำเป็นต้องระบุขอบเขตของกลุ่มข้อมูล
- สามารถจัดเก็บข้อมูลที่อยู่ในรูปแบบที่แตกต่างกันได้
- มีการแสดงผลด้วยสัญลักษณ์ ทำให้ง่ายต่อการทำความเข้าใจ

**ข้อเสีย**

- ในการกำหนดค่าความถี่น้อยที่สุดและค่าความน่าเชื่อถือน้อยที่สุด นั้นขึ้นอยู่กับลักษณะของธุรกิจและนักวิเคราะห์ซึ่งหากกำหนดไม่ดีอาจทำให้เกิดความผิดพลาดในการวิเคราะห์ข้อมูลได้
  - กฎที่ได้มานั้นอาจเป็นกฎที่เกิดขึ้นบ่อยๆและรู้อยู่แล้วทำให้ไม่เกิดการนำไปใช้งานได้จริงในทางปฏิบัติ
  - กฎที่ได้มาสามารถบอกได้เพียงแนวโน้มที่จะเกิดขึ้นด้วยกัน ไม่ได้บอกเรื่องของความเป็นเหตุเป็นผลของกฎซึ่งต้องอาศัยประสบการณ์ของนักวิเคราะห์ในการวิเคราะห์กฎต่างๆ

**2.4.5 การพยากรณ์ข้อมูลอนุกรมเชิงเวลา**

การวิเคราะห์อนุกรมเวลาเป็นระเบียบทางสถิติที่สามารถแปลงประสบการณ์ในอดีตไปพยากรณ์เหตุการณ์ในอนาคต

**1. อนุกรมเวลา**

อนุกรมเวลา (Time Series) หมายถึงข้อมูลหรือค่าสังเกตที่เปลี่ยนแปลงไปตามลำดับเวลาที่เกิดขึ้น โดยข้อมูลที่สังเกตได้ จะเก็บรวบรวมในช่วงเวลาใดก็ได้ ซึ่งจะทำให้เห็นรูปแบบของการเปลี่ยนแปลงค่าสังเกตในช่วงเวลาที่ผ่านมาเพื่อใช้ในการพยากรณ์ค่าสังเกตดังกล่าวในอนาคตได้

วิธีการพยากรณ์อนุกรมเวลา (Time Series Forecasting Method) เป็นการวิเคราะห์ลักษณะพฤติกรรมในอดีตของตัวแปรอนุกรมเวลา เพื่อพยากรณ์พฤติกรรมในอนาคตโดยถ้าค้นพบพฤติกรรมที่เป็นระบบบางอย่างในตัวแปรอนุกรมเวลาผู้ตัดสินใจก็จะสามารถสร้างแบบจำลองของพฤติกรรมของตัวแปรตามแล้วนำมาใช้ในการพยากรณ์พฤติกรรมของตัวแปรเหล่านั้นในอนาคตได้

**2. องค์ประกอบของอนุกรมเวลามีอะไรบ้าง**

ในการวิเคราะห์อนุกรมเวลาผู้วิเคราะห์จะแยกองค์ประกอบต่างๆ ที่ประกอบกันขึ้นเป็นอนุกรมเวลาโดยจะมีการเปลี่ยนแปลงไปตามอิทธิพลต่างๆ เช่นการเปลี่ยนแปลงการผลิตเทคโนโลยี สภาพอากาศ เป็นต้น ในการหาคุณลักษณะของอนุกรมเวลานั้นเราสามารถที่ใช้แบบจำลองได้หลายแบบ ซึ่งแบบจำลองที่ใช้โดยนักเศรษฐศาสตร์แบบหนึ่ง คือแบบจำลองแบบคลาสสิก (Classical Model) เป็นแบบจำลองที่อธิบายถึงองค์ประกอบของการแปรผันของอนุกรมเวลา 4 ส่วน ดังนี้

- ค่าแนวโน้ม (Secular trend) แทนด้วย  $T_t$
- การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล (Seasonal Variation) แทนด้วย  $S_t$
- การเปลี่ยนแปลงหรือความผันแปรตามวัฏจักร (Cyclical Variation) แทนด้วย  $C_t$
- การเปลี่ยนแปลงหรือความผันแปรเนื่องจากเหตุการณ์ผิดปกติ (Irregular Variation) แทนด้วย  $I_t$

แทนด้วย  $I_t$

### 3. รูปแบบของอนุกรมเวลา

จากปัจจัยทั้ง 4 ข้างต้น ถ้า  $Y$  แทนข้อมูลอนุกรมเวลาชุดหนึ่ง ๆ เราสามารถกำหนดแบบจำลองได้ 2 แบบ ดังนี้

- แบบจำลองผลบวก (Additive model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลบวกขององค์ประกอบทั้ง 4 อย่าง

$$Y_t = T_t + S_t + C_t + I_t$$

- แบบจำลองผลคูณ (Multiplicative model) ถือว่าข้อมูลในแต่ละอนุกรมเวลาประกอบด้วยผลคูณขององค์ประกอบทั้ง 4 อย่าง

$$Y_t = T_t * S_t * C_t * I_t$$

### 4. การพยากรณ์ (Introduction to Forecasting)

การพยากรณ์ความต้องการทางการตลาด ทำให้ผู้ผลิตสามารถวางแผนการผลิตให้สอดคล้องกับความต้องการนั้นในแต่ละช่วงเวลา อีกทั้งจะช่วยให้วางแผนการจัดซื้อวัตถุดิบต่างๆ วางแผนกำลังคน และ วางแผนปริมาณสินค้าคงคลัง และ วางแผนการซ่อมบำรุงรักษาได้อีกด้วย

ลักษณะของการพยากรณ์ที่ดี

- มีการกำหนดช่วงเวลาที่เหมาะสม เช่น การพยากรณ์ความต้องการในเดือนถัดไป หรือ ไตรมาสถัดไป
- เลือกเทคนิคที่มีความเหมาะสม และ เกิดความคลาดเคลื่อนน้อย เช่น การพยากรณ์ความต้องการสินค้าบางอย่างต้องพยากรณ์ตามฤดูกาล เป็นต้น
- มีการกำหนดหน่วยพยากรณ์ เช่น จำนวนชิ้น จำนวนเงิน เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 5. ขั้นตอนในการพยากรณ์

- กำหนดจุดประสงค์ในการพยากรณ์ เพื่อเลือกเทคนิคที่เหมาะสม
- กำหนดช่วงเวลาที่เหมาะสม ถ้าช่วงเวลาไกลหรือใกล้เกินไป จะเกิดความ

คลาดเคลื่อนมากขึ้น

- เลือกเทคนิคในการพยากรณ์ โดยพิจารณา ข้อมูลที่หาได้ ความง่าย งบประมาณ บุคลากร และ ทรัพยากรที่มีอยู่

- รวบรวมและวิเคราะห์ข้อมูลในอดีต หารูปแบบของข้อมูล เช่น

- ทำการพยากรณ์

- แสดงผลการพยากรณ์ กำหนดค่าความคลาดเคลื่อน (Error,  $e_t$ ) เมื่อเปรียบเทียบกับ ข้อมูลจริงในอดีต

$$\text{ความคลาดเคลื่อน } (e_t) = \text{ค่าจริง } (A_t) - \text{ค่าพยากรณ์ } (F_t)$$

- ปรับปรุงการพยากรณ์โดยอาจจะเก็บข้อมูลเพิ่มเติม หรือ ใช้เทคนิคการพยากรณ์ใหม่

ประเภทของการพยากรณ์ สามารถแบ่งได้เป็น 3 ประเภทดังนี้

- การพยากรณ์ระยะสั้น (Short range forecasting) เช่น การพยากรณ์ช่วงเดือน, หรือ ช่วงไตรมาส เป็นต้น

- การพยากรณ์ระยะปานกลาง (Intermediate range forecasting) เช่น การพยากรณ์ ในช่วง 1 ปี ใช้ในการวางแผนการผลิตหลัก เป็นต้น

- การพยากรณ์ระยะยาว (Long range forecasting) เช่น ถ้ามาการขยายกำลังการผลิตเพิ่มเติม อาจจะมีการพยากรณ์ ช่วงยาวกว่า 1 ปี เป็นต้น

## บทที่ 3

### การคัดเลือกและเตรียมข้อมูล

ในหัวข้อนี้จะกล่าวถึงการวิเคราะห์ปัญหาหรือความต้องการทางธุรกิจเพื่อที่จะนำไปสู่ขั้นตอนการคัดเลือกและการเตรียมข้อมูลตามหลักการและทฤษฎีของค้ำไมน์นิ่งเพื่อให้สามารถนำข้อมูลไปใช้วิเคราะห์ได้อย่างถูกต้อง รวดเร็ว และเหมาะสมกับเทคนิคค้ำไมน์นิ่งแบบต่างๆ

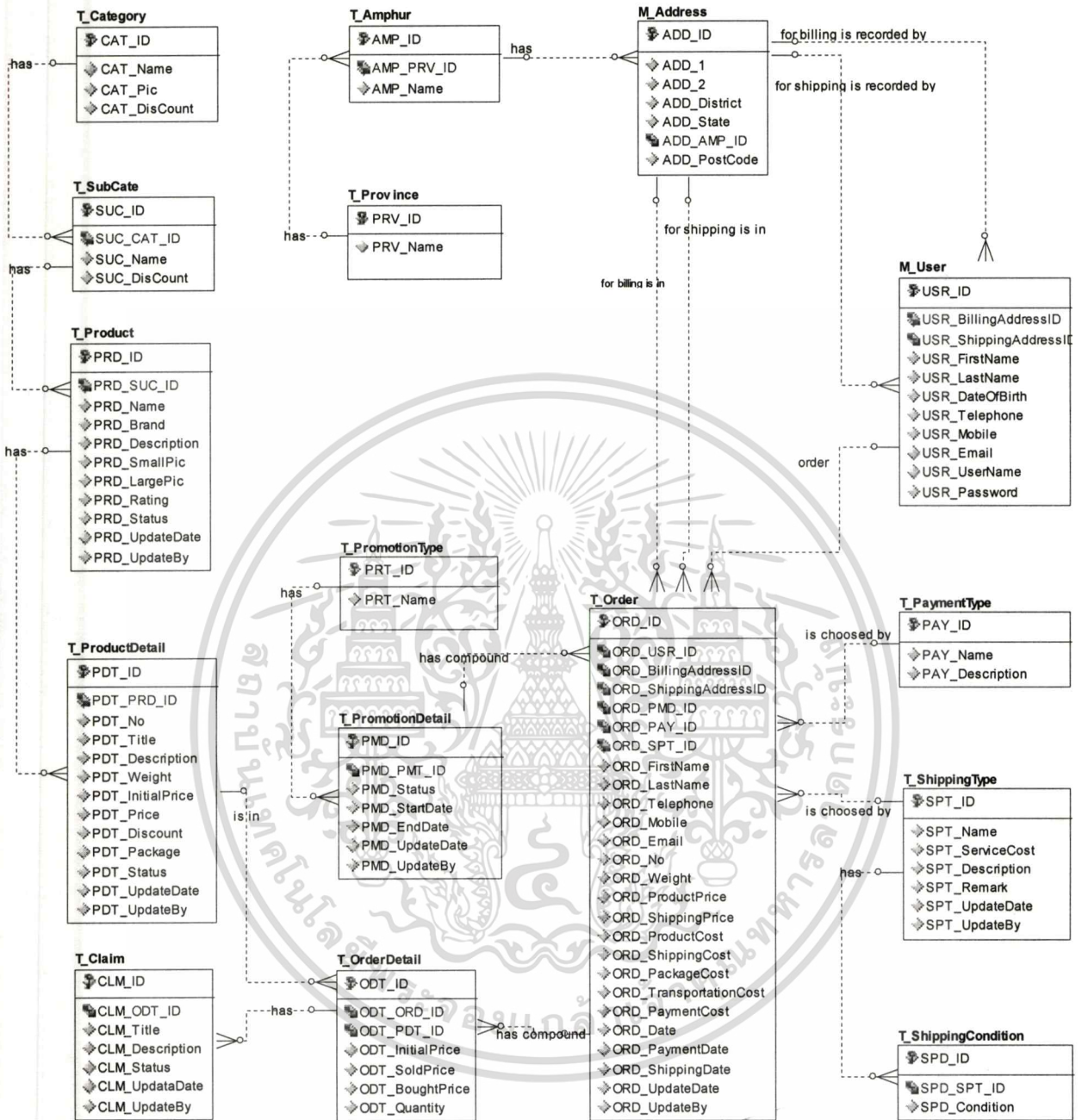
#### 3.1 แหล่งข้อมูล

แหล่งข้อมูลที่นำมาวิเคราะห์มาจากระบบปฏิบัติงานหลัก โดยเก็บข้อมูลลงในฐานข้อมูลเชิงสัมพันธ์ ซึ่งเป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบข้อมูลที่สัมพันธ์กัน สามารถแสดงให้เห็นได้ในรูปของตารางข้อมูล โดยความสัมพันธ์ของตารางข้อมูลใช้แบบจำลองความสัมพันธ์ระหว่างแอตทริบิวต์ ซึ่งข้อมูลที่นำมาพิจารณาจะเป็นข้อมูลเฉพาะที่มีแนวโน้มเกี่ยวข้องกับวัตถุประสงค์ที่ใช้ในการวิเคราะห์ข้อมูลเท่านั้น

โดยในฐานข้อมูลจะมีตารางหลักๆ ที่ใช้เก็บข้อมูลจากระบบปฏิบัติการหลัก โดยจะเริ่มจากการสมัครสมาชิกเพื่อซื้อสินค้า เก็บข้อมูลลงตาราง M\_User และ M\_Address เมื่อลูกค้าทำการเลือกซื้อสินค้าลงตะกร้าสินค้าฐานข้อมูลจะเชื่อมโยงไปที่ตาราง T\_Product และ T\_ProductDetail เพื่อแสดงรายละเอียดของสินค้าที่ต้องการ จากนั้นลูกค้าเลือกวิธีการชำระเงินและการขนส่งสินค้าฐานข้อมูลจะเชื่อมโยงไปที่ตาราง T\_PaymentType และ T\_ShippingType เมื่อลูกค้าตกลงที่จะซื้อสินค้าและชำระเงิน ฐานข้อมูลจะเชื่อมโยงไปที่ตาราง T\_Order และ T\_OrderDetail เพื่อเก็บรายละเอียดจากสั่งซื้อสินค้าของลูกค้า และพร้อมจัดส่งสินค้าให้กับลูกค้า และเมื่อมีการส่งคืนสินค้าฐานข้อมูลจะเก็บข้อมูลลงตาราง T\_Claim เพื่อเก็บข้อมูลสาเหตุการส่งคืนสินค้าคืนต่อไป ดังรูปที่ 3.1

#### 3.2 ระบบปัจจุบันและปัญหาของระบบ

จากเทคนิคค้ำไมน์นิ่งดังที่กล่าวข้างต้น สามารถนำมาประยุกต์ใช้ได้กับองค์กรที่ต้องการวิเคราะห์ข้อมูลที่ซ่อนอยู่ในฐานข้อมูลโดยองค์กรที่นำมาเป็นกรณีศึกษา ได้แก่บริษัท อะควาเรียม จำกัด ซึ่งดำเนินธุรกิจเกี่ยวกับการจัดจำหน่ายต้นไม้ พันธุ์ไม้น้ำ เครื่องกรองน้ำ เคมีภัณฑ์ ยา อาหาร อุปกรณ์การเลี้ยงปลา และตู้ปลา โดยบริษัทมีช่องทางการจำหน่ายผ่านทางอีคอมเมิร์ซหรือพาณิชย์อิเล็กทรอนิกส์เท่านั้น ไม่มีหน้าร้านเพื่อจัดจำหน่ายสินค้าโดยตรง และมีคลังเก็บสินค้าซึ่งตั้งอยู่เขตรังสิต จังหวัดปทุมธานี



รูปที่ 3.1 ความสัมพันธ์ของตารางข้อมูลของระบบปัจจุบัน

### 3.2.1 ระบบปัจจุบันและปัญหาของระบบ

ระบบปฏิบัติการหลักของบริษัทฯ มาจากการขายสินค้าอีคอมเมิร์ซหรือพาณิชย์อิเล็กทรอนิกส์ โดยผ่านเว็บไซต์ของบริษัท และมีการเก็บข้อมูลโดยใช้ฐานข้อมูลเชิงสัมพันธ์ ซึ่งปัจจุบันยังไม่มีระบบสารสนเทศที่ช่วยในการสนับสนุนการตัดสินใจในการวางแผนกลยุทธ์หรือมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ช่วยในการวิเคราะห์ข้อมูล ทำให้ผู้บริหารไม่สามารถวิเคราะห์ข้อมูลได้อย่างสะดวก รวดเร็วและถูกต้องเท่าที่ควร ปัญหาที่ผู้บริหารพบหรือต้องการมีดังนี้

- ไม่สามารถทราบถึงพฤติกรรมการซื้อขายของลูกค้าหรือความนิยมของผู้ซื้อสินค้าแต่ละประเภท
- ไม่สามารถวิเคราะห์ถึงสาเหตุของสินค้าที่มีการส่งคืนสินค้าได้
- ไม่สามารถวิเคราะห์ได้ว่า เมื่อลูกค้าซื้อสินค้าในแต่ละประเภทมัก จะเลือกวิธีการขนส่งสินค้าแบบใด และวิธีการขนส่งสินค้าแบบใดที่มีประสิทธิภาพและทำให้ลูกค้าพึงพอใจที่สุด
- ต้องการทราบถึงกลุ่มของลูกค้าที่ซื้อสินค้าเพื่อช่วยในการจัดโปรโมชั่นและทำ Direct Mail ถึงลูกค้าแต่ละกลุ่ม
- ต้องการคาดคะเนล่วงหน้าถึงยอดขายในเดือนหรือในปีถัดไปเพื่อช่วยในการวางแผนกลยุทธ์ของบริษัท
- แนวทางการขายสินค้าว่าในช่วงเวลาใดที่สินค้าประเภทใดขายดี หรือเป็นที่นิยมต่อลูกค้า และสินค้าประเภทใดที่ไม่ได้รับความนิยมนต่อลูกค้าเท่าที่ควร

### 3.2.2 การกำหนดวัตถุประสงค์ทางธุรกิจ

จากปัญหาที่เกิดขึ้นของบริษัททำให้บริษัทจำเป็นต้องหาหนทางที่จะแก้ไขปัญหาดังกล่าว ซึ่งสามารถแก้ไขปัญหาได้โดยวิธีการของคาด้า ไม่นิ่ง ดังนั้น ขั้นตอนแรกของการทำคาด้า ไม่นิ่งคือการวิเคราะห์ถึงความต้องการของระบบหรือความต้องการของพนักงาน ผู้บริหาร ซึ่งจากการสอบถามผู้บริหารของบริษัท สามารถวิเคราะห์ความต้องการ ได้ดังนี้

- วิเคราะห์พฤติกรรมกรรมการซื้อสินค้าของลูกค้า เช่น ลูกค้าส่วนใหญ่เมื่อซื้อสินค้าอย่างหนึ่งแล้ว มักจะซื้อสินค้าคู่กับสินค้าอะไร
- วิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทของสินค้าและวิธีการขนส่งสินค้า
- วิเคราะห์พฤติกรรมของการเลือกขนส่งสินค้าของลูกค้าตามประเภทของสินค้าที่สั่งซื้อ
- วิเคราะห์กลุ่มลูกค้าในการสั่งซื้อสินค้า
- ทำนายยอดขายสินค้า

### 3.3 การเตรียมข้อมูล

การออกแบบฐานข้อมูลสำหรับจัดเก็บข้อมูลเพื่อความสะดวกและเหมาะสมกับการนำข้อมูลไปใช้ทำคาด้า ไม่นิ่ง โดยหลักการในการออกแบบฐานข้อมูลนั้นเพื่อให้ได้ฐานข้อมูลที่เหมาะสมและครอบคลุมการวิเคราะห์ข้อมูลจะพิจารณาจากการวิเคราะห์ปัญหาในหัวข้อ 3.2.2 การกำหนดวัตถุประสงค์ทางธุรกิจ ซึ่งจะต้องคัดเลือกข้อมูลที่สำคัญ และเกี่ยวข้องกับการวิเคราะห์ปัญหาในแต่ละหัวข้อดังนี้

**3.3.1 ต้องการวิเคราะห์พฤติกรรมการซื้อขายของลูกค้า ได้แก่ “ลูกค้าส่วนใหญ่เมื่อซื้อสินค้าอย่างหนึ่งแล้วมักจะซื้อสินค้าคู่กับสินค้าอะไร”**

- การคัดเลือกข้อมูล ในกระบวนการนี้จะคัดเลือก แอตทริบิวต์จำนวน 2 แอตทริบิวต์ที่เกี่ยวข้องกับการวิเคราะห์พฤติกรรมการซื้อขายของลูกค้ามาใช้งานดังตารางที่ 3.1 โดยทั้ง 2 แอตทริบิวต์จะอยู่บนตาราง โดยแต่ละตารางสามารถเชื่อมความสัมพันธ์ด้วยคีย์ ORD\_ID (รหัสคำสั่งซื้อสินค้า)

- การทำความสะอาดข้อมูล เนื่องจากข้อมูลที่เกี่ยวข้องอยู่ในฐานข้อมูลนั้นอาจไม่ถูกต้องหรือไม่เหมาะสมสำหรับการทำคาน่าไม้นิ่ง ดังนั้น จำเป็นต้องตรวจสอบข้อมูลให้ถูกต้องก่อน โดยเริ่มจากการตรวจสอบข้อมูลที่ขาดหายไป (Missing data) ในแต่ละคอลัมน์ ถ้าพบว่าข้อมูลขาดหายไป ก็จะมีการเติมข้อมูลลงไปให้ครบ โดยใช้ข้อมูลที่ได้กำหนดไว้แล้วตามความเหมาะสมของธุรกิจ จากการตรวจสอบข้อมูลพบว่าแอตทริบิวต์ ORD\_No คือเลขที่ใบสั่งซื้อสินค้าและ แอตทริบิวต์ PRD\_Name คือชื่อสินค้า นั้นไม่มีข้อมูลขาดหายไปเนื่องจากเป็นข้อมูลที่ระบบสร้างขึ้นและไม่อนุญาตให้มีค่าว่างเกิดขึ้น จากนั้นตรวจสอบข้อมูลที่มีสิ่งรบกวน (Noisy data) ซึ่งเป็นความผิดพลาดที่เกิดขึ้นแบบไม่ตั้งใจ โดยเกิดจากการกรอกข้อมูลที่ผิดพลาด ทำให้ค่าไม่เป็นไปตามที่ควรจะเป็น สำหรับทั้ง 2 แอตทริบิวต์นี้จะไม่เกิดขึ้นเนื่องจากระบบสร้างขึ้น

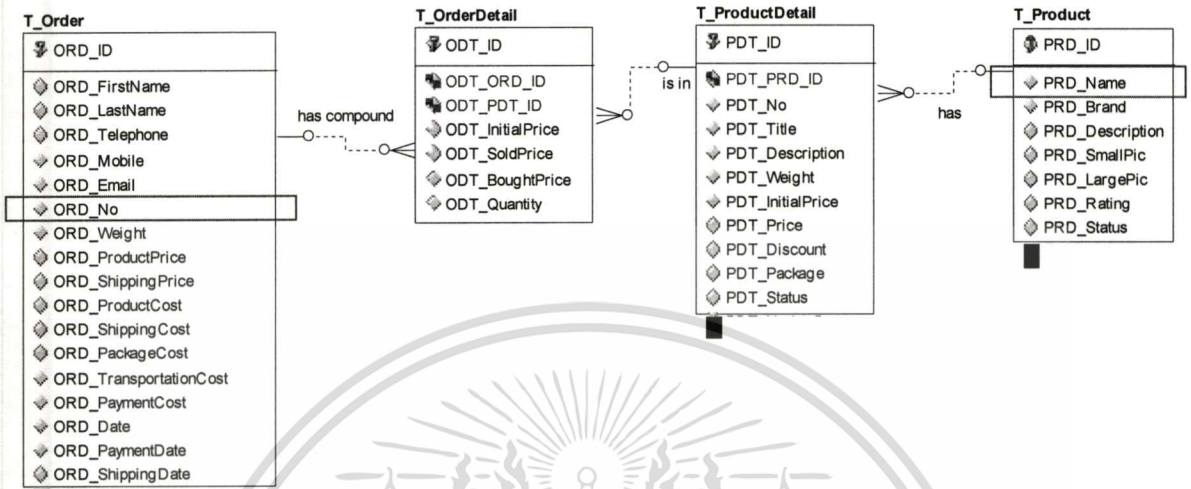
ตารางที่ 3.1 รายละเอียดของแอตทริบิวต์ที่ใช้ในการวิเคราะห์พฤติกรรมการซื้อขาย

|                    |                                |
|--------------------|--------------------------------|
| <b>แอตทริบิวต์</b> | <b>ORD_No</b>                  |
| คำอธิบาย           | เลขที่ใบสั่งซื้อสินค้า         |
| ตาราง              | T_Order                        |
| รูปแบบของข้อมูล    | yyyymmddxxxxx                  |
| ชนิดของข้อมูล      | ข้อมูลที่เป็นตัวเลข (Numeric)  |
| <b>แอตทริบิวต์</b> | <b>PRD_Name</b>                |
| คำอธิบาย           | ชื่อสินค้า                     |
| ตาราง              | T_Product                      |
| รูปแบบของข้อมูล    | xxxxxxxxxxxxx                  |
| ชนิดของข้อมูล      | ข้อมูลที่เป็นอักษร (Character) |

- การแปลงข้อมูล เป็นการทำให้ข้อมูลอยู่ในรูปแบบที่เหมาะสม โดยอาศัยเงื่อนไขเพื่อให้ข้อมูลสอดคล้องกับการวิเคราะห์ โดยทั้ง 2 แอตทริบิวต์นี้ไม่ต้องแปลงข้อมูลเนื่องจากเป็นข้อมูลที่สื่อความหมายได้ชัดเจนและสามารถนำไปวิเคราะห์ได้ทันที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- การรวบรวมข้อมูลจากตารางที่ 3.1 จะเห็นได้ว่าแอตทริบิวต์ที่นำมาใช้นั้นมาจาก 2 ตาราง ซึ่งการนำข้อมูลไปใช้งานจะต้องนำข้อมูลมารวมให้อยู่ในระดับเดียวกันและเป็นตารางเดียวกัน โดยทั้ง 2 แอตทริบิวต์มีการเชื่อมโยงความสัมพันธ์ดังรูปที่ 3.2



รูปที่ 3.2 ตารางความสัมพันธ์ของข้อมูล ORD\_No และ PRD\_Name

จากรูป 3.2 จะได้แอตทริบิวต์ ORD\_No จากตาราง T\_Order เชื่อมโยงข้อมูลไปยังตาราง T\_OrderDetail ด้วยคีย์ ORD\_ID จากนั้นเชื่อมโยงไปยังตาราง T\_ProductDetail ด้วยคีย์ ODT\_ID และเชื่อมโยงไปยังตาราง T\_Product ด้วยคีย์ PDT\_ID ซึ่งจะได้แอตทริบิวต์ PRD\_Name ดังตารางที่ 3.2

ตารางที่ 3.2 ตัวอย่างข้อมูลสำหรับวิเคราะห์พฤติกรรมการซื้อขายของลูกค้า

| ORD_No        | PRD_Name         |
|---------------|------------------|
| 2005120100012 | Air pump         |
| 2005120100013 | Filter           |
| 2005120100014 | Water Treatments |
| 2005120100015 | Filter           |

จากตารางที่ 3.2 เป็นตัวอย่างข้อมูลที่ผ่านกระบวนการเตรียมข้อมูล ทำความสะอาดข้อมูล และแปลงข้อมูลแล้ว สามารถนำข้อมูลนี้ไปใช้ในการวิเคราะห์ได้อย่างถูกต้องและมีประสิทธิภาพ

### 3.3.2 ต้องการวิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทของสินค้าและวิธีการขนส่งสินค้า

- การคัดเลือกข้อมูล ในกระบวนการนี้จะคัดเลือกแอตทริบิวต์จำนวน 3 แอตทริบิวต์ที่เกี่ยวข้องกับการวิเคราะห์ความเสี่ยงและสาเหตุของการส่งคืนสินค้าโดยแยกตามประเภทของสินค้าและการขนส่งสินค้าดังตารางที่ 3.3

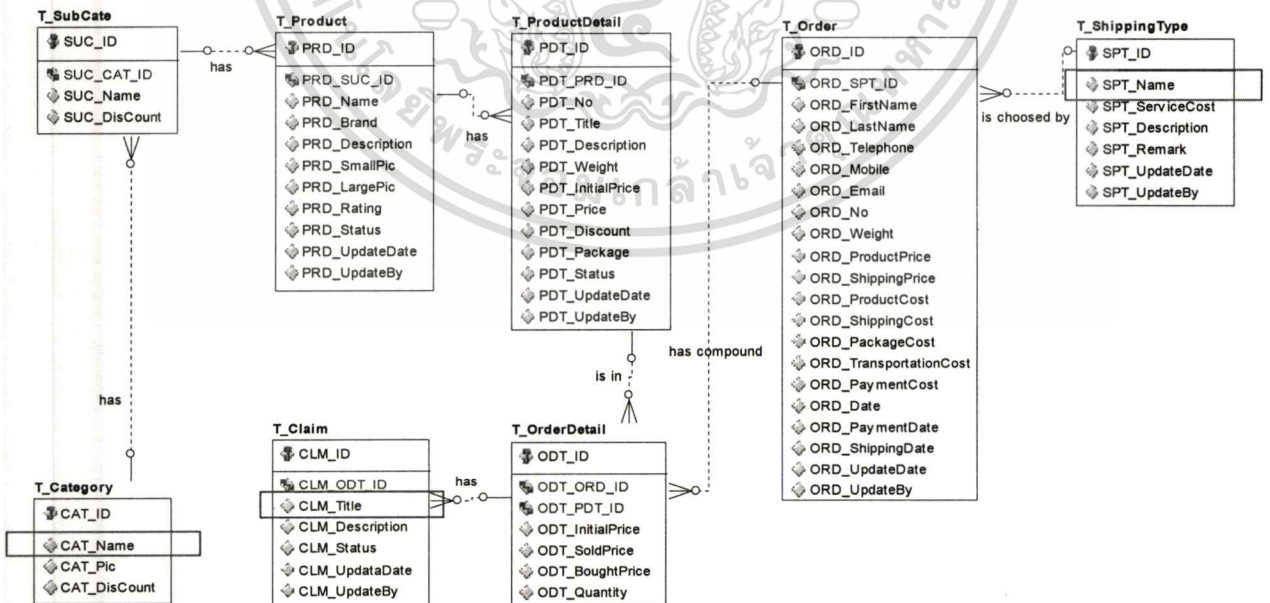
- การทำความสะอาดข้อมูล จากข้อมูลตารางที่ 3.3 สามารถทำความสะอาดข้อมูลได้โดยเริ่มจากการตรวจสอบข้อมูลที่ขาดหายไปในแต่ละคอลัมน์ ถ้าพบว่าข้อมูลขาดหายไปก็จะมีการเติมข้อมูลโดยใช้ข้อมูลที่ได้กำหนดไว้ ได้แก่ ข้อมูล “Unknown” สำหรับแอตทริบิวต์ CLM\_Title SPD\_Name และ CLM\_OTD\_ID ไม่พบข้อมูลที่ผิดพลาดเนื่องจากระบบไม่อนุญาตให้มีค่าว่างเกิดขึ้น จากนั้นตรวจสอบข้อมูลที่มีสิ่งรบกวน ซึ่งเป็นความผิดพลาดที่เกิดขึ้นแบบไม่ตั้งใจ ซึ่งอาจเกิดจากการกรอกข้อมูลที่ผิดพลาด ทำให้ค่าไม่เป็นไปตามที่ควรจะเป็น สำหรับทั้ง 3 แอตทริบิวต์นี้จะไม่เกิดขึ้นเนื่องจากระบบสร้างขึ้น

- การแปลงข้อมูลเป็นการทำให้ข้อมูลอยู่ในรูปแบบที่เหมาะสม ได้แก่ แอตทริบิวต์ CLM\_ODT\_ID คือรหัสสินค้าที่ส่งคืน พบว่าหากนำข้อมูลดังกล่าวไปใช้งานจะทำให้การแปลความหมายของข้อมูลยาก เราจึงจัดกลุ่มของข้อมูลโดยให้เป็นสินค้าในหมวดต่างๆ ได้แก่ ตู้ปลา (Aquarium), อุปกรณ์ตู้ปลา (Equipment), เคมีภัณฑ์ (Chemical), อาหารปลา (Fish Food) และ ต้นไม้ น้ำ (Plant) โดยการเชื่อมต่อข้อมูลไปยัง CAT\_Name ซึ่งเป็นประเภทของสินค้า

- การรวบรวมข้อมูล เป็นการทำให้ข้อมูลที่เกี่ยวข้องอยู่ในลักษณะหลายระดับในหลายๆตารางให้อยู่ในระดับและตารางเดียวกัน เพื่อให้สามารถทำงานได้อย่างถูกต้องและรวดเร็ว โดยแต่ละตารางเชื่อมโยงความสัมพันธ์ดังรูปที่ 3.3 ตาราง T\_Claim จะได้ข้อมูลการส่งคืนสินค้า แอตทริบิวต์ CLM\_Title และแอตทริบิวต์ CLM\_OTD\_ID เชื่อมโยงข้อมูลไปยังตาราง T\_Order เพื่อทราบข้อมูลรหัสส่งคืนสินค้า และตาราง T\_ShippingType เพื่อแปลงข้อมูลเป็นประเภทการขนส่งสินค้า SPT\_Name จากนั้นแปลงข้อมูลแอตทริบิวต์ CLM\_OTD\_ID ให้เป็นข้อมูลที่อ่านแล้วเข้าใจได้มากขึ้น โดยแปลงข้อมูลให้เป็นประเภทของสินค้าคือแอตทริบิวต์ CAT\_Name โดยการเชื่อมโยงข้อมูลตาราง T\_OrderDetail ไปยังตาราง T\_ProductDetail T\_Product T\_SubCate และตาราง T\_Catagory นำข้อมูลทั้งให้อยู่ในตารางเดียวกัน ดังตารางที่ 3.4

ตารางที่ 3.3 รายละเอียดของแอตทริบิวต์ที่ใช้ในการวิเคราะห์สาเหตุของการส่งคืนสินค้า

|   |  |
|---|--|
| <p><b>แอตทริบิวต์</b></p> <p>คำอธิบาย</p> <p>ตาราง</p> <p>รูปแบบของข้อมูล</p> <p>ชนิดของข้อมูล</p>  | <p><b>CLM_ODT_ID</b></p> <p>รหัสสินค้าที่ส่งคืน</p> <p>T_Claim</p> <p>A001 – Z999</p> <p>ข้อมูลที่เป็นอักษร (Character)</p>  |
| <p><b>แอตทริบิวต์</b></p> <p>ตาราง</p> <p>คำอธิบาย</p> <p>รูปแบบของข้อมูล*</p> <p>ชนิดของข้อมูล</p> | <p><b>SPT_Name</b></p> <p>T_ShippingType</p> <p>ประเภทการขนส่งสินค้า</p> <p>Motorcycle, EMS, Parcel, TNT</p> <p>ข้อมูลที่เป็นอักษร (Character)</p>                 |
| <p><b>แอตทริบิวต์</b></p> <p>ตาราง</p> <p>คำอธิบาย</p> <p>รูปแบบของข้อมูล</p> <p>ชนิดของข้อมูล</p>  | <p><b>CLM_Title</b></p> <p>T_Claim</p> <p>ประเภทของสาเหตุการส่งคืนสินค้า</p> <p>Broken, Unavailable, Package, Miss order</p> <p>ข้อมูลที่เป็นอักษร (Character)</p> |



รูปที่ 3.3 ตารางความสัมพันธ์ของข้อมูล CLM\_Title CAT\_Name และ SPT\_Name

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 ตัวอย่างข้อมูลสำหรับวิเคราะห์สาเหตุการขอคืนสินค้าแยกตามประเภทของสินค้าและการขนส่ง

| CLM_ODT_ID | ORD_SPT_ID | CLM_Title   |
|------------|------------|-------------|
| A004       | Motorcycle | Broken      |
| E001       | EMS        | Unavailable |
| P003       | Parcel     | Miss order  |
| C003       | Motorcycle | Broken      |

ตารางที่ 3.5 ตัวอย่างข้อมูลที่ผ่านกระบวนการทำความสะอาดข้อมูล แปลงข้อมูลและรวบรวมข้อมูล สำหรับวิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทของสินค้าและการขนส่ง

| Product Category | Shipping Type | Claim Type  |
|------------------|---------------|-------------|
| Aquarium         | Motorcycle    | Broken      |
| Equipment        | EMS           | Unavailable |
| Plant            | Parcel        | Miss order  |
| Chemical         | Motorcycle    | Broken      |

จากตารางที่ 3.5 เป็นตัวอย่างข้อมูลที่ผ่านกระบวนการเตรียมข้อมูล ทำความสะอาดข้อมูล แปลงข้อมูลและรวบรวมข้อมูลให้อยู่ในระดับและตารางเดียวกันแล้วสามารถนำข้อมูลไปใช้วิเคราะห์ได้อย่างมีประสิทธิภาพ

### 3.3.3 วิเคราะห์พฤติกรรมของการเลือกขนส่งสินค้าของลูกค้าตามประเภทของสินค้าที่สั่งซื้อ

- การคัดเลือกข้อมูล ในกระบวนการนี้จะคัดเลือก แอดทริบิวต์จำนวน 5 แอดทริบิวต์ที่มีแนวโน้มเกี่ยวข้องกับกรณีวิเคราะห์ความเสี่ยงในการขนส่งสินค้าให้กับลูกค้าดังตารางที่ 3.6

- การทำความสะอาดข้อมูล โดยเริ่มจากการตรวจสอบข้อมูลที่ขาดหายไปและสิ่งแปลกปลอมของข้อมูลในแต่ละคอลัมน์ จากระบบพบว่าในคอลัมน์ของแต่ละแอดทริบิวต์นั้นไม่พบข้อมูลที่ขาดหายไปเนื่องจากระบบไม่อนุญาตให้มีค่าว่างเกิดขึ้น และตรวจสอบความผิดปกติของข้อมูลว่ามีค่าที่ไม่เป็นไปได้อหรือไม่ ซึ่งพบว่าค่าทุกค่าของข้อมูลนั้นถูกต้อง เนื่องจากเป็นการป้อนข้อมูลโดยเลือกข้อมูลที่มีอยู่ในระบบเท่านั้น ไม่ได้มีการป้อนข้อมูลเอง

- การแปลงข้อมูล เป็นการทำให้ข้อมูลอยู่ในรูปแบบที่เหมาะสมโดยอาศัยเงื่อนไขเพื่อให้ข้อมูลสอดคล้องกับการวิเคราะห์ ได้แก่ แอดทริบิวต์ ADD\_Postcode เปลี่ยนคุณสมบัติให้เป็น โชน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรุงเทพฯ (Bangkok) และต่างจังหวัด (Other province) สำหรับแอตทริบิวต์ ORD\_Weight แบ่งข้อมูลออกเป็น 3 กลุ่มคือ  $\leq 5\text{kg}$  = ขนาดเล็ก (Small),  $>5$  และ  $\leq 10\text{kg}$  = ขนาดกลาง (Medium) และ  $>10\text{ kg}$  = ขนาดใหญ่ (Large)

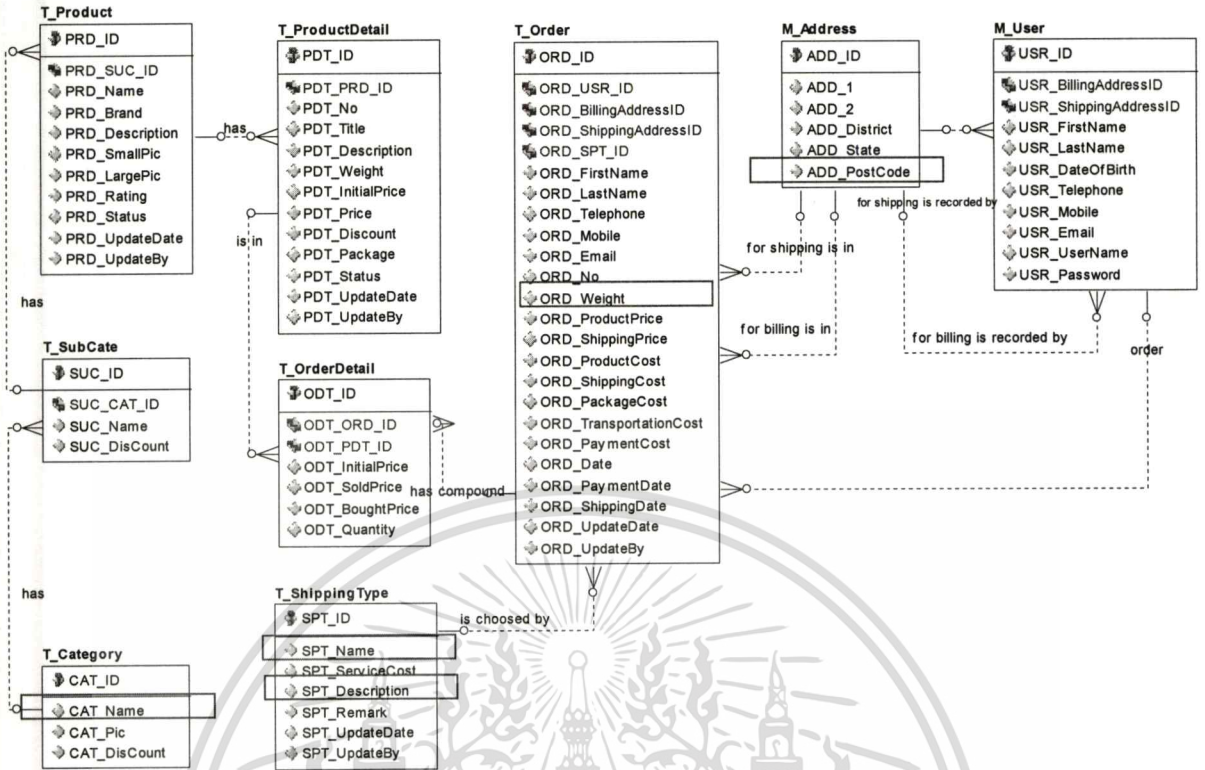
ตารางที่ 3.6 รายละเอียดของแอตทริบิวต์ที่ใช้วิเคราะห์การเลือกขนส่งสินค้า

|                    |   |
|--------------------|---|
| <b>แอตทริบิวต์</b> | <b>CAT_Name*</b>                        |
| <b>แอตทริบิวต์</b> | <b>ADD_PostCode</b>                     |
| ตาราง              | M_Address                               |
| คำอธิบาย           | รหัสไปรษณีย์สำหรับส่งสินค้า             |
| รูปแบบของข้อมูล    | 30000                                   |
| ชนิดของข้อมูล      | ข้อมูลที่เป็นตัวเลข (Numeric)           |
| <b>แอตทริบิวต์</b> | <b>SPT_Description</b>                  |
| ตาราง              | T_ShippingType                          |
| คำอธิบาย           | ระยะเวลาส่งสินค้า                       |
| รูปแบบของข้อมูล    | 1-2 days, 3-5 days, 4-17 days, 1-3 days |
| ชนิดของข้อมูล      | ข้อมูลที่เป็นอักษร (Character)          |
| <b>แอตทริบิวต์</b> | <b>ORD_Weight</b>                       |
| ตาราง              | T_ProductDetail                         |
| คำอธิบาย           | น้ำหนักสินค้า                           |
| รูปแบบของข้อมูล    | 000001-999999 gram                      |
| ชนิดของข้อมูล      | ข้อมูลเป็นตัวเลข (Numeric)              |
| <b>แอตทริบิวต์</b> | <b>SPT_Name**</b>                       |

\*รายละเอียดของแอตทริบิวต์ตามตารางที่ 3.1

\*\*รายละเอียดของแอตทริบิวต์ตามตารางที่ 3.5

• การรวบรวมข้อมูล เริ่มต้นจากการรวมข้อมูลสินค้าในตาราง T\_Order จะได้แอตทริบิวต์ ORD\_Weight ซึ่งจะเชื่อมโยงไปตาราง M\_address จะได้แอตทริบิวต์ ADD\_PostCode และเชื่อมโยงไปยังตาราง T\_ShippingType จะได้แอตทริบิวต์ SPT\_Name และ SPT\_Description ตาราง T\_Category จะได้แอตทริบิวต์ CAT\_Name รวบรวมข้อมูลทุกแอตทริบิวต์ให้อยู่ในตารางเดียวกัน ดังรูปที่ 3.4 และตารางที่ 3.7



รูปที่ 3.4 ตารางความสัมพันธ์ของข้อมูล CAT\_Name, SPT\_Name, SPT\_Description, ORD\_Weight และ ADD\_PostCode

ตารางที่ 3.7 ตัวอย่างข้อมูลสำหรับวิเคราะห์พฤติกรรมการเลือกส่งขนส่งสินค้าของลูกค้าตามประเภทของสินค้า

| CAT_Name  | ADD_PostCode | ORD_Weight | SPT_Name   | SPT_Description |
|-----------|--------------|------------|------------|-----------------|
| Aquarium  | 10320        | 15         | TNT        | 1-3 days        |
| Equipment | 32000        | 8          | EMS        | 3-5 days        |
| Chemical  | 54040        | 3          | EMS        | 3-5 days        |
| Equipment | 10200        | 5          | Motorcycle | 1-2 days        |
| Plant     | 30000        | 2          | EMS        | 3-5 days        |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 3.8 ตัวอย่างข้อมูลหลังจากผ่านกระบวนการเตรียมข้อมูลสำหรับวิเคราะห์พฤติกรรมการเลือก  
ส่งขนส่งสินค้าของลูกค้าตามประเภทของสินค้า**

| CAT_Name  | Destination    | Weight_Type | SPT_Name   | SPT_Description |
|-----------|----------------|-------------|------------|-----------------|
| Aquarium  | Bangkok        | Large       | TNT        | 1-3 days        |
| Equipment | Other province | Medium      | EMS        | 3-5 days        |
| Chemical  | Other province | Small       | EMS        | 3-5 days        |
| Equipment | Bangkok        | Small       | Motorcycle | 1-2 days        |
| Plant     | Other province | Small       | EMS        | 3-5 days        |

จากตารางที่ 3.8 เป็นตัวอย่างข้อมูลที่ผ่านกระบวนการเตรียมข้อมูล ทำความสะอาดข้อมูล  
แปลงข้อมูลและรวบรวมข้อมูลให้อยู่ในระดับและตารางเดียวกันแล้ว

### 3.3.4 วิเคราะห์กลุ่มลูกค้าที่สั่งซื้อสินค้า

- การคัดเลือกข้อมูล ในกระบวนการนี้จะคัดเลือก แอดทริบิวต์จำนวน 5 แอดทริบิวต์ที่มี  
แนวโน้มเกี่ยวข้องกับการวิเคราะห์กลุ่มของลูกค้าในการสั่งซื้อสินค้าดังตารางที่ 3.9

**ตารางที่ 3.9 รายละเอียดของแอดทริบิวต์ที่ใช้วิเคราะห์กลุ่มของลูกค้าที่สั่งซื้อสินค้า**

|                    |                                |
|--------------------|--------------------------------|
| <u>แอดทริบิวต์</u> | CAT_Name*                      |
| <u>แอดทริบิวต์</u> | ADD_PostCode**                 |
| <u>แอดทริบิวต์</u> | ORD_Weight**                   |
| <u>แอดทริบิวต์</u> | USR_DateOfBirth                |
| ตาราง              | M_User                         |
| คำอธิบาย           | วัน/เดือน/ปี เกิดของลูกค้า     |
| รูปแบบของข้อมูล    | 16022005                       |
| ชนิดของข้อมูล      | ข้อมูลเป็นตัวอักษร (Character) |
| <u>แอดทริบิวต์</u> | PDT_Price                      |
| ตาราง              | T_ProductDetail                |
| คำอธิบาย           | ราคาสินค้า                     |
| รูปแบบของข้อมูล    | 0000001-9999999                |
| ชนิดของข้อมูล      | ข้อมูลเป็นตัวเลข (Numeric)     |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

\*รายละเอียดของแอตทริบิวต์ตามตารางที่ 3.1

\*\*รายละเอียดของแอตทริบิวต์ตามตารางที่ 3.6

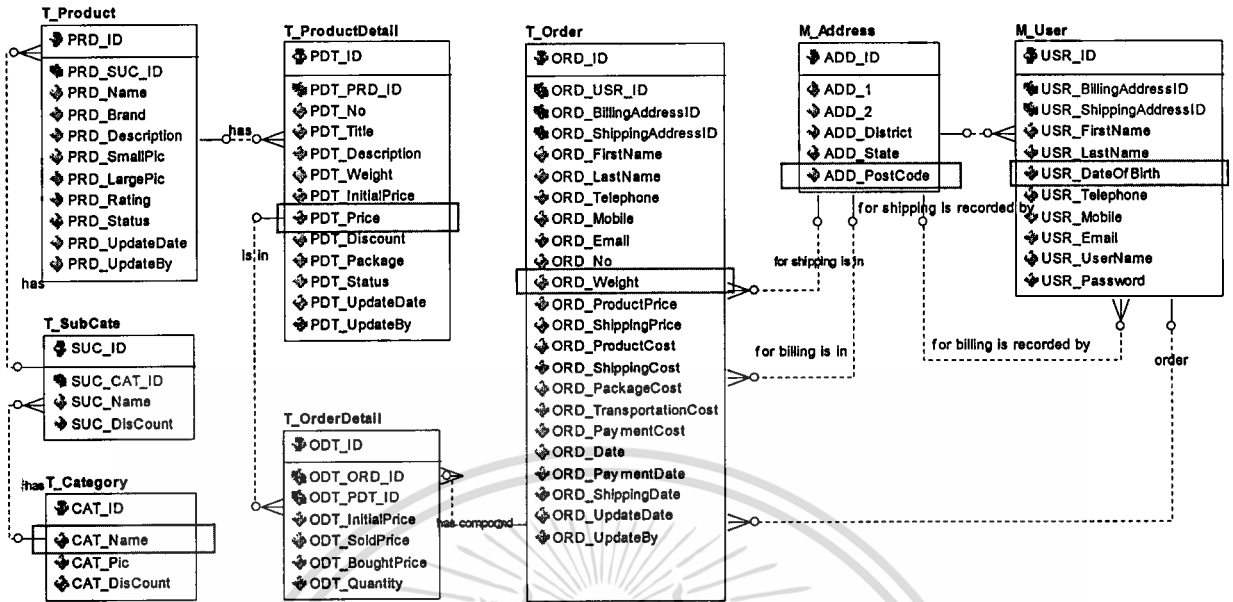
- การทำความสะอาดข้อมูล โดยเริ่มจากการตรวจสอบข้อมูลที่ขาดหายไปในแต่ละคอลัมน์ จากระบบพบว่าในคอลัมน์ของแต่ละแอตทริบิวต์นั้นไม่พบข้อมูลที่ขาดหายไป เนื่องจากระบบไม่อนุญาตให้มีค่าว่างเกิดขึ้น และตรวจสอบความผิดปกติของข้อมูลว่ามีค่าที่ไม่เป็นไปได้หรือไม่ ซึ่งพบว่าค่าทุกค่าของข้อมูลนั้นถูกต้อง เนื่องจากเป็นการป้อนข้อมูลโดยเลือกข้อมูลที่มีอยู่ในระบบเท่านั้น ไม่ได้มีการป้อนข้อมูลเอง

- การแปลงข้อมูล เป็นการทำให้ข้อมูลอยู่ในรูปแบบที่เหมาะสมโดยอาศัยเงื่อนไขเพื่อให้ข้อมูลสอดคล้องกับการวิเคราะห์ ได้แก่ ADD\_PostCode เปลี่ยนคุณสมบัติโดยแบ่งออกเป็น 6 กลุ่ม ได้แก่ ภาคกลาง, ภาคเหนือ, ภาคอีสาน, ภาคใต้, ภาคตะวันออก และกรุงเทพฯ (กรุงเทพฯจะรวมจังหวัดปริมณฑลด้วย) แอตทริบิวต์ ORD\_Weight แบ่งข้อมูลออกเป็น 3 กลุ่มคือ  $\leq 5\text{kg}$  = ขนาดเล็ก (Small),  $>5$  และ  $\leq 10\text{kg}$  = ขนาดกลาง (Medium) และ  $>10\text{ kg}$  = ขนาดใหญ่ (Large) และแอตทริบิวต์ USR\_DateOfBirth เป็นมีการคำนวณอายุของลูกค้า ดังตารางที่ 3.10

ตารางที่ 3.10 ตัวอย่างข้อมูลสำหรับวิเคราะห์กลุ่มของลูกค้าที่สั่งซื้อสินค้า

| CAT_Name  | ADD_PostCode | ORD_Weight | USR_DateOfBirth | PDT_Price |
|-----------|--------------|------------|-----------------|-----------|
| Aquarium  | 10320        | 15         | 18021980        | 5,000     |
| Equipment | 32000        | 8          | 02061978        | 350       |
| Chemical  | 54040        | 3          | 15121985        | 250       |
| Equipment | 10200        | 5          | 22111971        | 1,200     |
| Plant     | 30000        | 2          | 20041980        | 200       |

- การรวบรวมข้อมูล แต่ละตารางมีการเชื่อมโยงความสัมพันธ์กันดังรูปที่ 3.4 โดยเริ่มต้นจากตาราง T\_Order จะได้แอตทริบิวต์ ORD\_Weight ซึ่งจะเชื่อมโยงไปตาราง M\_address จะได้แอตทริบิวต์ ADD\_PostCode และเชื่อมโยงไปยังตาราง M\_User จะได้แอตทริบิวต์ USR\_DateOfBirth ตาราง T\_ProductDetail จะได้แอตทริบิวต์ PDT\_Price และ ตาราง T\_Catagory จะได้แอตทริบิวต์ CAT\_Name รวบรวมข้อมูลทุกแอตทริบิวต์ให้อยู่ในตารางเดียวกัน ดังรูปที่ 3.5



รูปที่ 3.5 ตารางความสัมพันธ์ของข้อมูล CAT\_Name, PDT\_Price, ORD\_Weight, ADD\_PostCode และ USR\_DateOfBirth

ตารางที่ 3.11 ตัวอย่างข้อมูลหลังจากผ่านกระบวนการเตรียมข้อมูลสำหรับวิเคราะห์กลุ่มของลูกค้าที่สั่งซื้อสินค้า

| CAT_Name  | Zone      | ORD_Weight | Age | Price |
|-----------|-----------|------------|-----|-------|
| Aquarium  | Bangkok   | Large      | 28  | 5,000 |
| Equipment | Northeast | Medium     | 30  | 350   |
| Chemical  | North     | Small      | 22  | 250   |
| Equipment | Bangkok   | Small      | 36  | 1,200 |
| Plant     | Northeast | Small      | 27  | 200   |

จากตารางที่ 3.11 เป็นตัวอย่างข้อมูลที่ผ่านกระบวนการเตรียมข้อมูล ทำความสะอาดข้อมูล แปลงข้อมูลและรวบรวมข้อมูลให้อยู่ในระดับและตารางเดียวกันแล้ว

### 3.3.5 ทำนายยอดขายสินค้าในวันถัดไป

- การคัดเลือกข้อมูล ในกระบวนการนี้จะคัดเลือก แอตทริบิวต์จำนวน 2 แอตทริบิวต์ที่มีแนวโน้มเกี่ยวข้องกับการวิเคราะห์กลุ่มของลูกค้าในการสั่งซื้อสินค้าดังตารางที่ 3.12

ตารางที่ 3.12 รายละเอียดของแอตทริบิวต์ที่ใช้สำหรับทำนายยอดขายสินค้า

|                    |                                |
|--------------------|--------------------------------|
| <u>แอตทริบิวต์</u> | ORD_Price*                     |
| <u>แอตทริบิวต์</u> | ORD_Date                       |
| ตาราง              | T_Order                        |
| คำอธิบาย           | วันที่สั่งซื้อสินค้า           |
| รูปแบบของข้อมูล    | dd/mm/yyyy                     |
| ชนิดของข้อมูล      | ข้อมูลเป็นตัวอักษร (Character) |

\*รายละเอียดของแอตทริบิวต์ตามตารางที่ 3.11

- การทำความสะอาดข้อมูล โดยเริ่มจากการตรวจสอบข้อมูลที่ขาดหายไปในแต่ละคอลัมน์ จากระบบพบว่าในคอลัมน์ของแต่ละแอตทริบิวต์นั้นไม่พบข้อมูลที่ขาดหายไป เนื่องจากระบบไม่อนุญาตให้มีค่าว่างเกิดขึ้น และตรวจสอบความผิดปกติของข้อมูลว่ามีค่าที่ไม่เป็นไปได้อหรือไม่ ซึ่งพบว่าค่าทุกค่าของข้อมูลนั้นถูกต้อง เนื่องจากการป้อนข้อมูลโดยเลือกข้อมูลที่มีอยู่ในระบบเท่านั้น ไม่ได้มีการป้อนข้อมูลเอง
- การแปลงข้อมูล เป็นการทำให้ข้อมูลอยู่ในรูปแบบที่เหมาะสม โดยอาศัยเงื่อนไขเพื่อให้ข้อมูลสอดคล้องกับการวิเคราะห์ สำหรับแอตทริบิวต์ ORD\_Date นั้นรูปแบบของข้อมูลสามารถนำไปใช้งานได้ทันที และสำหรับ ORD\_Price จะมีการแปลงข้อมูลโดยการคำนวณยอดขายในแต่ละยอดการสั่งซื้อสินค้าให้เป็นยอดขายใน 1 วัน
- การรวบรวมข้อมูล จากข้อมูลข้างต้นเป็นข้อมูลที่มาจาก 2 ตารางจึงต้องทำการรวบรวมข้อมูลให้อยู่ในตารางเดียวกัน ดังตารางที่ 3.13

**ตารางที่ 3.13** ตัวอย่างข้อมูลสำหรับทำนายยอดการขายสินค้า

| ORD_Date   | PDT_Price |
|------------|-----------|
| 01/08/2005 | 5,000     |
| 02/08/2005 | 3500      |
| 02/08/2005 | 2500      |
| 05/09/2005 | 1,200     |
| 15/10/2005 | 2000      |

จากตารางที่ 3.13 เป็นตัวอย่างข้อมูลที่ผ่านกระบวนการเตรียมข้อมูล ทำความสะอาดข้อมูล แปลงข้อมูลและรวบรวมข้อมูลให้อยู่ในระดับและตารางเดียวกันแล้ว



## บทที่ 4

# การนำข้อมูลเข้าสู่ฐานข้อมูลใหม่

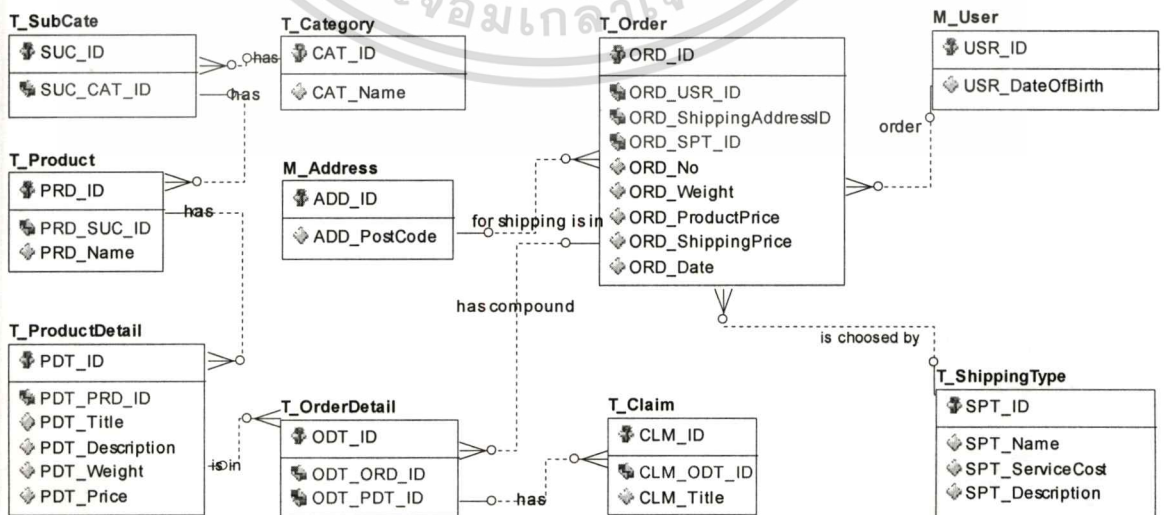
เมื่อทำวิเคราะห์ข้อมูลตามความต้องการหรือปัญหาของระบบ โดยผ่านการคัดเลือกข้อมูล ทำความสะอาดข้อมูล แปลงข้อมูลและรวบรวมข้อมูลแล้ว ขั้นตอนต่อไปคือการออกแบบฐานข้อมูลใหม่เพื่อให้สอดคล้องกับความต้องการของระบบ สามารถนำข้อมูลไปวิเคราะห์โดยใช้ซอฟต์แวร์ใดก็ได้ที่ทันที่ และการใช้เทคนิคใดก็ตามนี้จำเป็นจะต้องมีการออกแบบฐานข้อมูลใหม่เนื่องจากในการวิเคราะห์ข้อมูลจำเป็นต้องมีการดึงข้อมูลไปใช้ซึ่งจะต้องไม่กระทบกับระบบปฏิบัติการหลักของบริษัท

### 4.1 โมเดลความสัมพันธ์ของฐานข้อมูลใหม่

ในการวิเคราะห์และออกแบบฐานข้อมูลใหม่ ปัจจัยหลักคือข้อมูลที่ต้องการใช้ในกระบวนการวิเคราะห์พฤติกรรมลูกค้าดังที่กล่าวไว้ในบทที่ 3 และนอกจากนั้นในการออกแบบฐานข้อมูลใหม่จำเป็นต้องออกแบบฐานข้อมูลให้สามารถครอบคลุมถึงความต้องการในอนาคตได้ในกรณีที่ต้องการวิเคราะห์ข้อมูลในเรื่องอื่นๆเพิ่มเติม

#### 4.1.1 วิเคราะห์ข้อมูลสำหรับสร้างฐานข้อมูลใหม่

จากการคัดเลือกข้อมูลในบทที่ 3 นั้นจะเห็นได้ว่าการคัดเลือกข้อมูลใดไปใช้นั้นขึ้นอยู่กับวัตถุประสงค์ของระบบ ซึ่งแต่ละวัตถุประสงค์ก็มีการใช้ข้อมูลแตกต่างกัน โดยสามารถสรุปความสัมพันธ์ของข้อมูลทั้งหมดที่จำเป็นต้องใช้ได้ดังรูปที่ 4.1



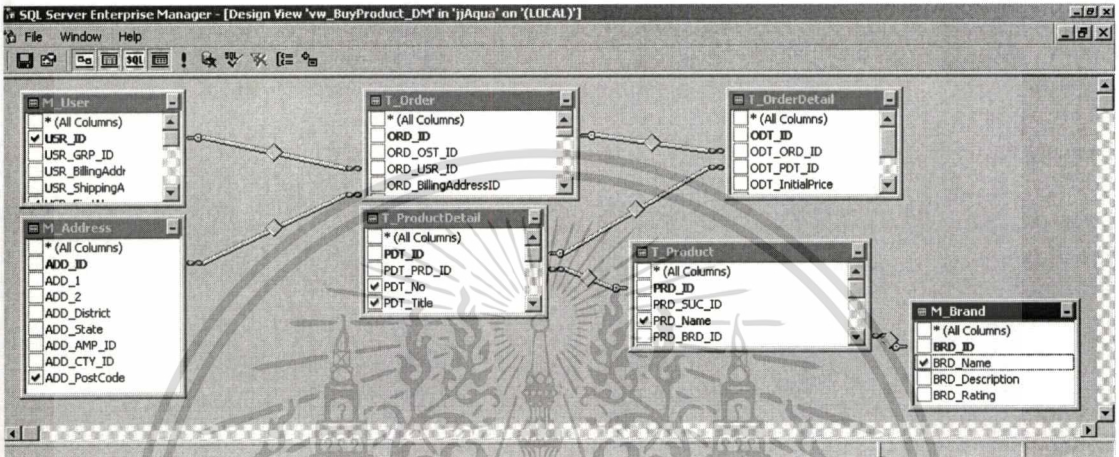
รูปที่ 4.1 ความสัมพันธ์ของตารางข้อมูลที่จำเป็นต้องใช้ในการวิเคราะห์ข้อมูล

เอกสารนี้เป็นเอกสารทรัพย์สินทางปัญญาของบริษัทฯ เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.1.2 การคิวรีข้อมูลจากฐานข้อมูล

การคิวรีข้อมูลจากฐานข้อมูลเพื่อให้ได้ข้อมูลที่สอดคล้องกับวัตถุประสงค์ โดยจะมีการคิวรีข้อมูลอื่นที่คาดว่าจะมีประโยชน์ได้ในอนาคตมาด้วย วิธีการคิวรีข้อมูลทำได้โดยการเลือกตารางซึ่งมีเส้นความสัมพันธ์เชื่อมโยงแต่ละตารางเข้าด้วยกัน จากนั้นเลือกข้อมูลที่ต้องการนำไปใช้ ดังรูปที่

4.2



รูปที่ 4.2 การคิวรีข้อมูลจากตารางความสัมพันธ์ตามวัตถุประสงค์ที่ต้องการ

โดยในแต่ละวัตถุประสงค์ ก็จะมีการเลือกข้อมูลต่างๆ ที่จำเป็นแตกต่างกันดังนี้

- วิเคราะห์พฤติกรรมการซื้อสินค้าของลูกค้า เช่น ลูกค้าส่วนใหญ่เมื่อซื้อสินค้าอย่างหนึ่งแล้ว มักจะซื้อสินค้าคู่กับสินค้าอะไร ได้คิวรีข้อมูลตามที่ต้องการคือ ORD\_No และ PRD\_Name สำหรับแอตทริบิวต์อื่นๆ เป็นการคิวรีข้อมูลสำหรับวัตถุประสงค์อื่นและรองรับวัตถุประสงค์ที่อาจเปลี่ยนแปลงได้ในอนาคต โดยกำหนดชื่อตารางว่า **Buy Product** ดังรูปที่ 4.3

The screenshot shows the design of the 'BuyProduct' table in SQL Server Enterprise Manager. The table has the following columns:

| Column Name       | Data Type | Length | Allow Nulls |
|-------------------|-----------|--------|-------------|
| USR_ID            | int       | 4      |             |
| Age               | int       | 4      | ✓           |
| ORD_No            | nvarchar  | 50     | ✓           |
| ORD_Weight        | decimal   | 9      | ✓           |
| ORD_ProductPrice  | money     | 8      | ✓           |
| ORD_ShippingPrice | money     | 8      | ✓           |
| BRD_Name          | nvarchar  | 100    | ✓           |
| PRD_Name          | nvarchar  | 50     | ✓           |
| PDT_No            | nvarchar  | 50     | ✓           |
| PDT_Title         | varchar   | 50     | ✓           |
| PDT_Weight        | decimal   | 9      | ✓           |
| PDT_InitialPrice  | money     | 8      | ✓           |
| PDT_Price         | money     | 8      | ✓           |
| [Zone]            | nvarchar  | 5      | ✓           |
| CAT_Name          | nvarchar  | 80     | ✓           |

รูปที่ 4.3 ตาราง Buy Product จากการคิวรีข้อมูลจากตารางความสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- วิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทของสินค้าและวิธีการขนส่งสินค้า ได้  
คิวรีข้อมูลตามที่ต้องการคือ SPT\_Name, Claim\_Type และ CAT\_Name สำหรับแอดทริบิวต์อื่นๆ  
เป็นการคิวรีข้อมูลสำหรับวัตถุประสงค์อื่นและรองรับวัตถุประสงค์ที่อาจเปลี่ยนแปลงได้ในอนาคต  
โดยกำหนดชื่อตารางว่า **Claim** ดังรูปที่ 4.4

| Column Name | Data Type | Length | Allow Nulls |
|-------------|-----------|--------|-------------|
| ORD_ID      | int       | 4      |             |
| ORD_No      | nvarchar  | 50     | ✓           |
| ORD_Weight  | decimal   | 9      | ✓           |
| ORD_Date    | datetime  | 8      | ✓           |
| PDT_No      | nvarchar  | 50     | ✓           |
| PDT_Title   | nvarchar  | 50     | ✓           |
| SPT_Name    | nvarchar  | 50     | ✓           |
| Claim_Type  | nvarchar  | 50     | ✓           |
| CAT_Name    | nvarchar  | 80     | ✓           |

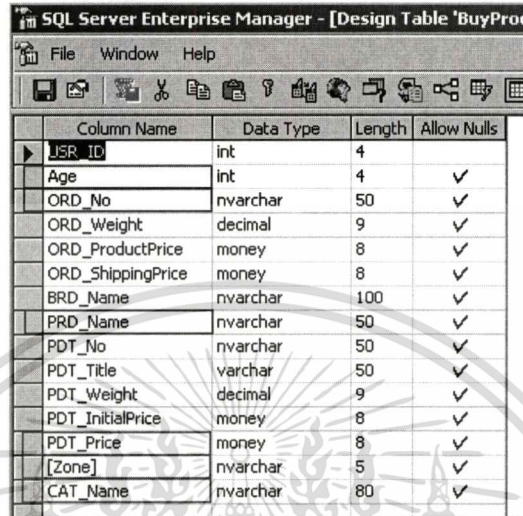
รูปที่ 4.4 ตาราง Claim จากการคิวรีข้อมูลจากตารางความสัมพันธ์

- วิเคราะห์พฤติกรรมของการเลือกขนส่งสินค้าของลูกค้าตามประเภทของสินค้าที่สั่งซื้อ  
ได้คิวรีข้อมูลตามที่ต้องการคือ Weight\_Type, SPT\_Description, Destination และ CAT\_Name  
สำหรับการคิวรีข้อมูลสำหรับวัตถุประสงค์อื่นและรองรับวัตถุประสงค์ที่อาจ  
เปลี่ยนแปลงได้ในอนาคต โดยกำหนดชื่อตารางว่า **Shipping Order** ดังรูปที่ 4.5

| Column Name       | Data Type | Length | Allow Nulls |
|-------------------|-----------|--------|-------------|
| ORD_ID            | int       | 4      |             |
| ORD_No            | nvarchar  | 50     | ✓           |
| Weight_Type       | decimal   | 9      | ✓           |
| ORD_ProductPrice  | money     | 8      | ✓           |
| ORD_ShippingPrice | money     | 8      | ✓           |
| SPT_Name          | nvarchar  | 50     | ✓           |
| SPT_ServiceCost   | money     | 8      | ✓           |
| SPT_Description   | nvarchar  | 100    | ✓           |
| BRD_Name          | nvarchar  | 100    | ✓           |
| PRD_Name          | nvarchar  | 50     | ✓           |
| PDT_No            | nvarchar  | 50     | ✓           |
| PDT_Title         | nvarchar  | 50     | ✓           |
| PDT_Weight        | decimal   | 9      | ✓           |
| PDT_InitialPrice  | money     | 8      | ✓           |
| PDT_Price         | money     | 8      | ✓           |
| Destination       | nvarchar  | 50     | ✓           |
| CAT_Name          | nvarchar  | 80     | ✓           |

รูปที่ 4.5 ตาราง Shipping Order จากการคิวรีข้อมูลจากตารางความสัมพันธ์

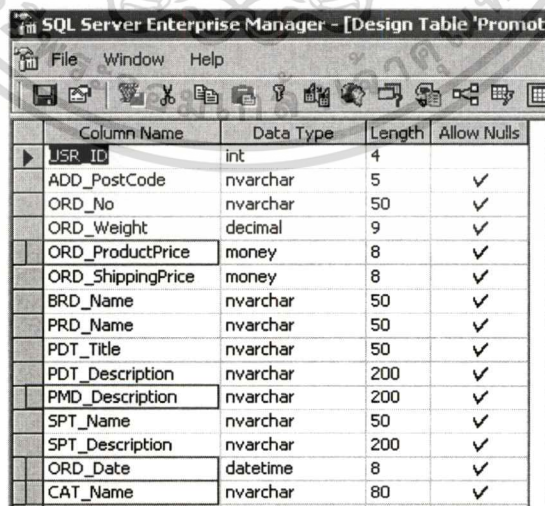
- วิเคราะห์กลุ่มลูกค้าในการสั่งซื้อสินค้า ได้คิวรีข้อมูลตามที่ต้องการคือ Age, PDT\_Price, Zone และ CAT\_Name ซึ่งเป็นการคิวรีข้อมูลรวมกับตาราง Buy Product ดังรูปที่ 4.6



| Column Name       | Data Type | Length | Allow Nulls |
|-------------------|-----------|--------|-------------|
| USR_ID            | int       | 4      |             |
| Age               | int       | 4      | ✓           |
| ORD_No            | nvarchar  | 50     | ✓           |
| ORD_Weight        | decimal   | 9      | ✓           |
| ORD_ProductPrice  | money     | 8      | ✓           |
| ORD_ShippingPrice | money     | 8      | ✓           |
| BRD_Name          | nvarchar  | 100    | ✓           |
| PRD_Name          | nvarchar  | 50     | ✓           |
| PDT_No            | nvarchar  | 50     | ✓           |
| PDT_Title         | varchar   | 50     | ✓           |
| PDT_Weight        | decimal   | 9      | ✓           |
| PDT_InitialPrice  | money     | 8      | ✓           |
| PDT_Price         | money     | 8      | ✓           |
| [Zone]            | nvarchar  | 5      | ✓           |
| CAT_Name          | nvarchar  | 80     | ✓           |

รูปที่ 4.6 ตาราง Buy Product จากการคิวรีข้อมูลสำหรับวัตถุประสงค์การวิเคราะห์กลุ่มลูกค้า

- ทำนายยอดขายในวันถัดไป ได้คิวรีข้อมูลตามที่ต้องการคือ ORD\_ProductPrice, PMD\_Description, ORD\_Date และ CAT\_Name สำหรับแอดทริบิวต์อื่นๆ เป็นการคิวรีข้อมูลสำหรับวัตถุประสงค์อื่นและรองรับวัตถุประสงค์ที่อาจเปลี่ยนแปลงได้ในอนาคต โดยกำหนดชื่อตารางว่า Promotion ดังรูปที่ 4.7



| Column Name       | Data Type | Length | Allow Nulls |
|-------------------|-----------|--------|-------------|
| USR_ID            | int       | 4      |             |
| ADD_PostCode      | nvarchar  | 5      | ✓           |
| ORD_No            | nvarchar  | 50     | ✓           |
| ORD_Weight        | decimal   | 9      | ✓           |
| ORD_ProductPrice  | money     | 8      | ✓           |
| ORD_ShippingPrice | money     | 8      | ✓           |
| BRD_Name          | nvarchar  | 50     | ✓           |
| PRD_Name          | nvarchar  | 50     | ✓           |
| PDT_Title         | nvarchar  | 50     | ✓           |
| PDT_Description   | nvarchar  | 200    | ✓           |
| PMD_Description   | nvarchar  | 200    | ✓           |
| SPT_Name          | nvarchar  | 50     | ✓           |
| SPT_Description   | nvarchar  | 200    | ✓           |
| ORD_Date          | datetime  | 8      | ✓           |
| CAT_Name          | nvarchar  | 80     | ✓           |

รูปที่ 4.7 ตาราง Promotion จากการคิวรีข้อมูลจากตารางความสัมพันธ์

### 4.1.3 ตารางสำหรับฐานข้อมูลใหม่

เมื่อทำการคิวรีข้อมูลตามที่ต้องการแล้วขั้นต่อไปเป็นการสร้างตารางสำหรับฐานข้อมูลใหม่ ซึ่งจะเป็นฐานข้อมูลที่จัดเก็บข้อมูลสำหรับจะนำไปใช้ในการทำคาน่าไมน์นิ่ง ดังรูปที่ 4.8

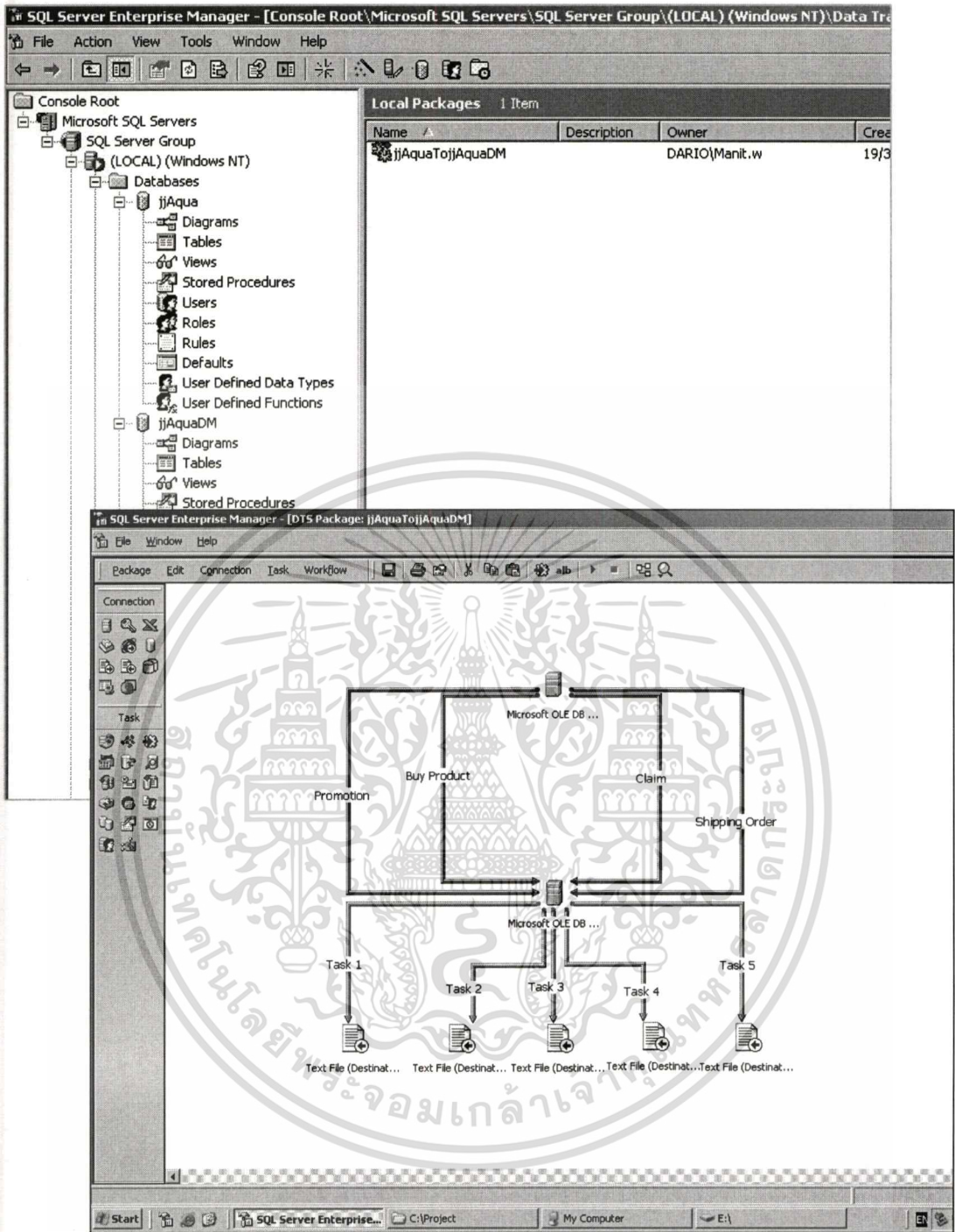
| BuyProduct   | Claim   | ShippingOrder   | Promotion   |
|--|---|---|---|
| <ul style="list-style-type: none"> <li>◆ USR_ID</li> <li>◆ Age</li> <li>◆ ORD_No</li> <li>◆ ORD_Weight</li> <li>◆ ORD_ProductPrice</li> <li>◆ ORD_ShippingPrice</li> <li>◆ BRD_Name</li> <li>◆ PRD_Name</li> <li>◆ PDT_No</li> <li>◆ PDT_Title</li> <li>◆ PDT_Weight</li> <li>◆ PDT_InitialPrice</li> <li>◆ PDT_Price</li> <li>◆ Zone</li> <li>◆ CAT_Name</li> </ul> | <ul style="list-style-type: none"> <li>◆ ORD_ID</li> <li>◆ ORD_No</li> <li>◆ ORD_Weight</li> <li>◆ ORD_Date</li> <li>◆ PDT_No</li> <li>◆ PDT_Title</li> <li>◆ SPT_Name</li> <li>◆ Claim_Type</li> <li>◆ CAT_Name</li> </ul> | <ul style="list-style-type: none"> <li>◆ ORD_ID</li> <li>◆ ORD_No</li> <li>◆ Weight_Type</li> <li>◆ ORD_ProductPrice</li> <li>◆ ORD_ShippingPrice</li> <li>◆ SPT_Name</li> <li>◆ SPT_ServiceCost</li> <li>◆ SPT_Description</li> <li>◆ BRD_Name</li> <li>◆ PRD_Name</li> <li>◆ PDT_No</li> <li>◆ PDT_Title</li> <li>◆ PDT_Weight</li> <li>◆ PDT_InitialPrice</li> <li>◆ PDT_Price</li> <li>◆ Destination</li> <li>◆ CAT_Name</li> </ul> | <ul style="list-style-type: none"> <li>◆ USR_ID</li> <li>◆ ADD_PostCode</li> <li>◆ ORD_No</li> <li>◆ ORD_Weight</li> <li>◆ ORD_ProductPrice</li> <li>◆ ORD_ShippingPrice</li> <li>◆ BRD_Name</li> <li>◆ PRD_Name</li> <li>◆ PDT_Title</li> <li>◆ PDT_Description</li> <li>◆ PMD_Description</li> <li>◆ SPT_Name</li> <li>◆ SPT_Description</li> <li>◆ ORD_Date</li> <li>◆ CAT_Name</li> </ul> |

รูปที่ 4.8 ตารางสำหรับฐานข้อมูลใหม่

## 4.2 การนำข้อมูลเข้าสู่ฐานข้อมูลใหม่

หลังจากที่ได้ออกแบบความสัมพันธ์ของข้อมูลสำหรับฐานข้อมูลใหม่แล้วขั้นต่อไปจะเป็นการนำข้อมูลจากระบบปฏิบัติการเดิมเข้าสู่ฐานข้อมูลใหม่ โดยการสร้างแพ็คเกจของ DTS (Data Transformation Service) โดยมีขั้นตอนดังนี้

1. ติดต่อ SQL Server เลือกโฟลเดอร์ไปยัง Data Transformation Service > New Package
2. ที่ DTS Designer ให้ทำการคลิกไอคอน SQL Server เลือกเมนู Connection > Microsoft OLE DB Provider for SQL Server เพื่อสร้างการเชื่อมต่อไปยังฐานข้อมูล
3. สร้างการเชื่อมต่อไปยังฐานข้อมูลปลายทางหรือฐานข้อมูลใหม่ ทำเช่นเดียวกันกับขั้นตอนที่ 2 ดังรูปที่ 4.9

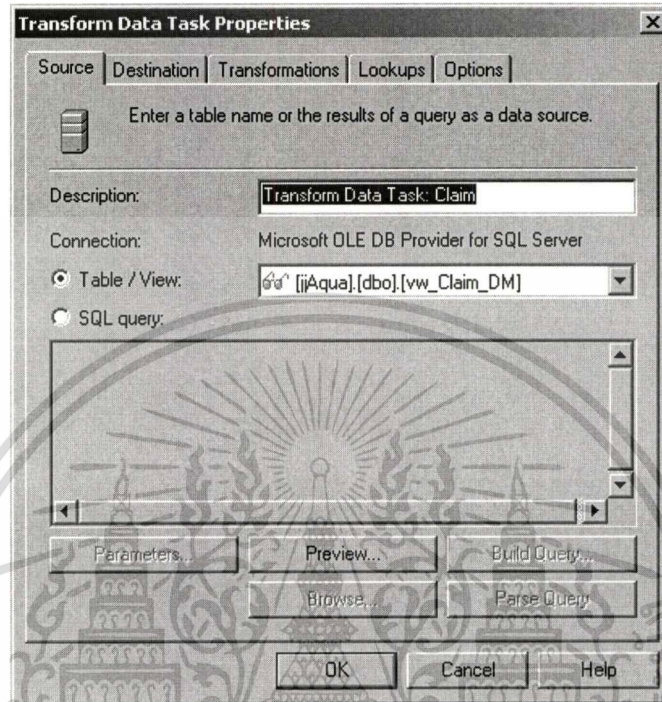


รูปที่ 4.9 การสร้าง New DTS Package และการออกแบบ DTS Package

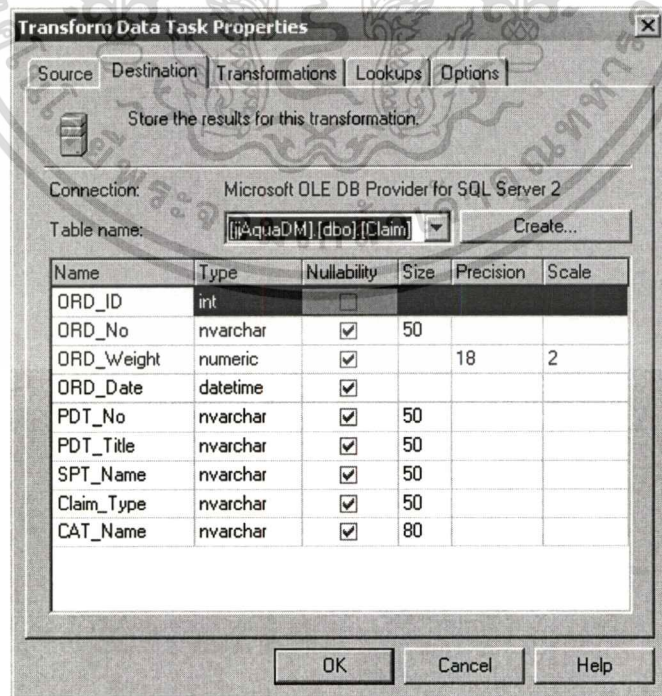
4. สร้างงานสำหรับการโอนข้อมูลระหว่างเซิร์ฟเวอร์ต้นทางและปลายทาง โดยเลือกที่เมนู Transform Data > Task > Transform Data Task จะ ได้เส้นเชื่อมโยงโดยลูกศรชี้ไปยังปลายทาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. ดับเบิลคลิกที่เส้นเชื่อมโยงเพื่อกำหนดการทำงานของงานการโอนย้ายข้อมูล โดยเริ่มต้นจากการระบุข้อมูลต้นทาง คลิกปุ่ม Create เพื่อสร้างตารางปลายทางตามที่ต้องการ ดังรูปที่ 4.10-4.11



รูปที่ 4.10 การสร้างงานสำหรับการโอนข้อมูลโดยเลือกข้อมูลด้านต้นทาง

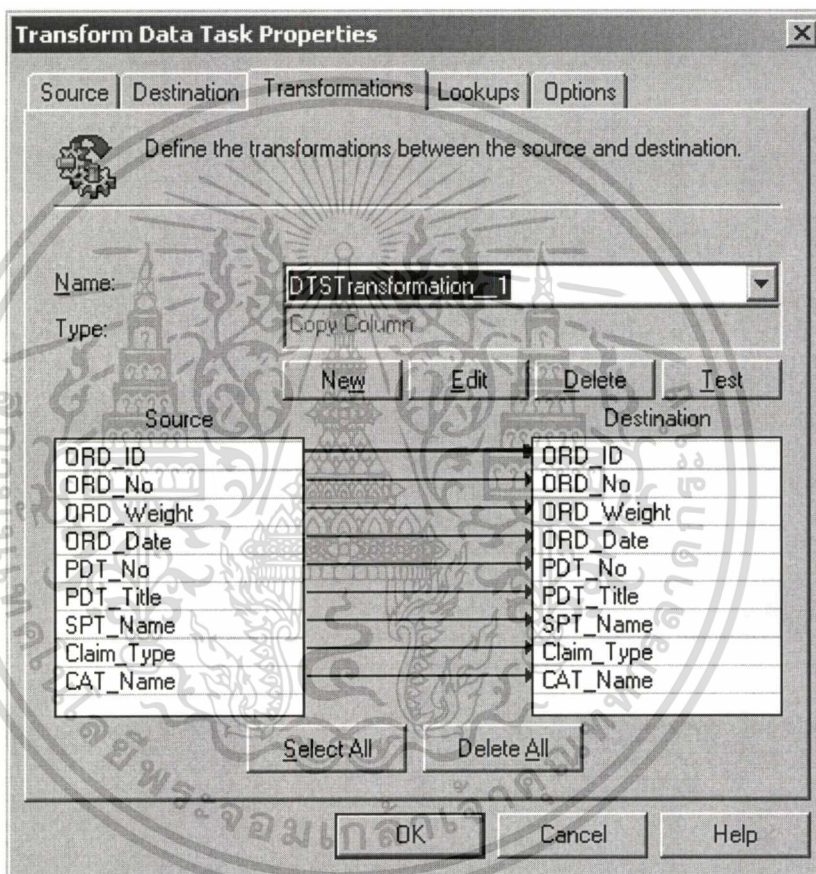


รูปที่ 4.11 การสร้างตารางปลายทางสำหรับฐานข้อมูลใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

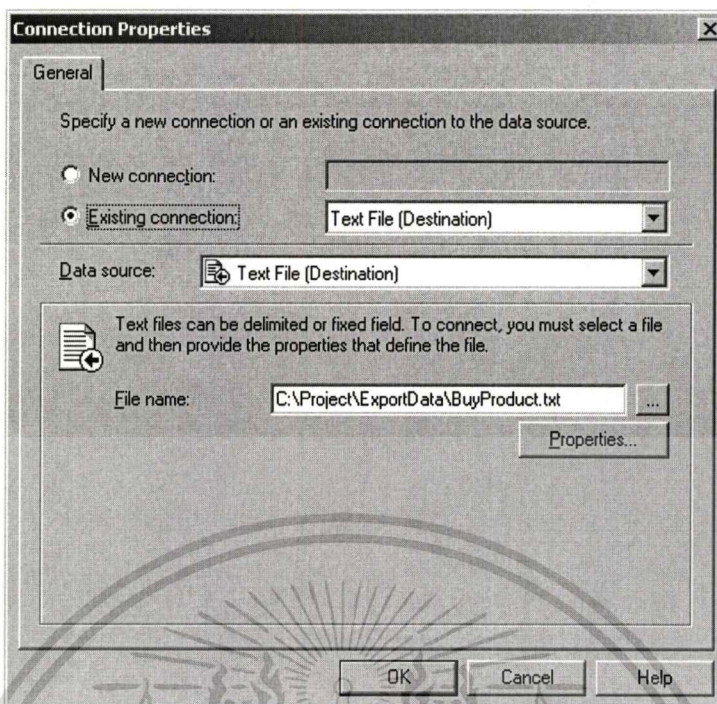
6. ที่แท็บ Transformation เป็นการกำหนดการเทียบเคียงกันระหว่างคอลัมน์ของตารางต้นทางและปลายทาง โดยระบบจะพยายามจับคู่ให้ระหว่างคอลัมน์ที่ชื่อหรือประเภทของข้อมูลเหมือนกันหรือคล้ายกัน

7. กำหนดวิธีการโอนข้อมูล สามารถเลือกได้ 3 แบบคือ ActiveX Script เป็นการโอนข้อมูลของคอลัมน์ โดยการนำภาษาสคริปต์มาใช้ในการควบคุมข้อมูล Copy Column เป็นการโอนข้อมูลของคอลัมน์ระหว่างกันโดยตรง และสำหรับรายการอื่นๆ เป็นการโอนข้อมูลโดยใช้ฟังก์ชันพื้นฐานทั่วไปที่มักใช้บ่อย ดังรูปที่ 4.12



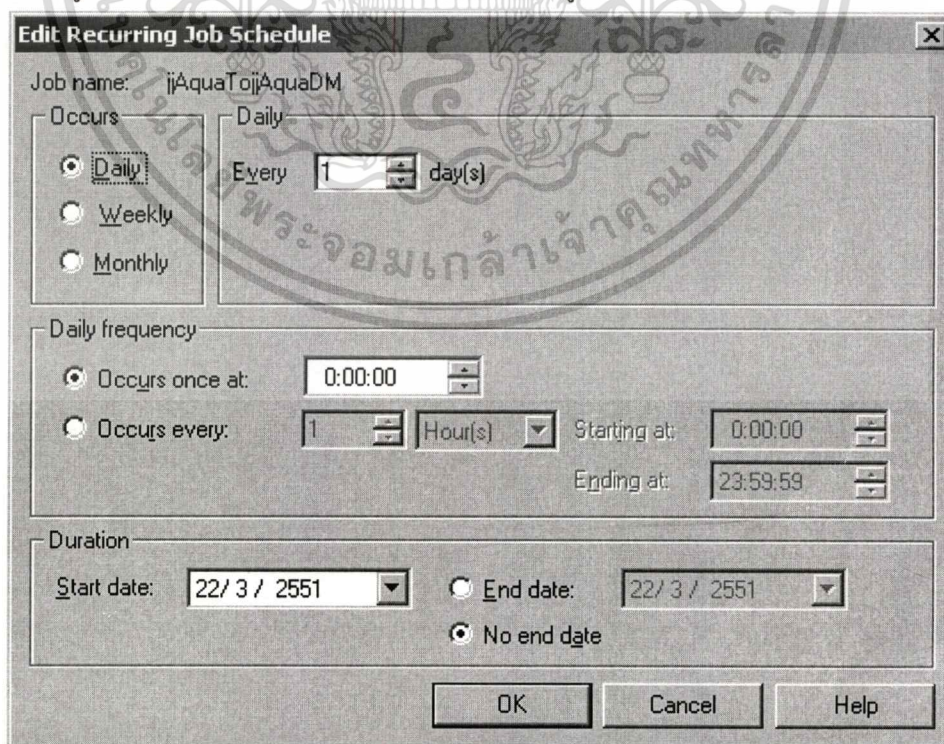
รูปที่ 4.12 การกำหนดการเทียบเคียงข้อมูลระหว่างต้นทางและปลายทาง

8. สร้างงานสำหรับสร้างไฟล์ข้อมูลที่เป็นแบบแท็บไฟล์ เพื่อนำข้อมูลที่ได้ไปใช้เป็นอินพุตสำหรับโปรแกรม SAS 9.1 โดยมีการกำหนดคุณสมบัติและเงื่อนไขของไฟล์ดังกล่าวได้ดังรูปที่ 4.13



รูปที่ 4.13 การกำหนดไฟล์ข้อมูล

9. กำหนดช่วงเวลาในการอัปเดตข้อมูลเมื่อมีการเปลี่ยนแปลงข้อมูลซึ่งอาจจะกำหนดให้มีการดึงข้อมูลจากฐานข้อมูลเก่าสู่ฐานข้อมูลใหม่และสร้างไฟล์ข้อมูลไว้เป็นอาทิตย์หรือเดือนละ 1 ครั้ง ก็ได้ขึ้นอยู่กับความต้องการของนักวิเคราะห์ระบบ ดังรูปที่ 4.14



รูปที่ 4.14 การกำหนดช่วงเวลาในการดึงข้อมูลแบบอัตโนมัติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

# ซอฟต์แวร์สำหรับการทำดาต้าไมน์นิ่ง

ในบทนี้จะกล่าวถึงซอฟต์แวร์สำหรับการทำดาต้าไมน์นิ่งที่จะนำมาประยุกต์ใช้กับระบบ เพื่อให้ได้ผลลัพธ์ตามวัตถุประสงค์ที่ต้องการ

### 5.1 การเลือกซอฟต์แวร์สำหรับการทำดาต้าไมน์นิ่ง

ปัจจุบันดาต้าไมน์นิ่งได้รับความนิยมใช้กันในองค์กรต่างๆ มากขึ้น เนื่องจากปัจจัยหลายประการ เช่น ปริมาณข้อมูลที่เพิ่มขึ้นของแต่ละองค์กร การทำการวิเคราะห์ข้อมูลขนาดใหญ่เพื่อสร้างคุณค่าให้กับข้อมูลทำได้ยากแต่เป็นสิ่งจำเป็น ความสามารถของซีพียูที่เพิ่มขึ้นในขณะที่ราคาลดลง การพัฒนาผลิตภัณฑ์ที่เกี่ยวข้องกับดาต้าไมน์นิ่งที่ทำงานบนเครื่องคอมพิวเตอร์ส่วนบุคคลมีมากขึ้น เป็นต้น ทำให้องค์กรต่างๆ เริ่มหันมาใช้เครื่องมือดาต้าไมน์นิ่งมากขึ้น

### 5.2 คุณสมบัติหลักของซอฟต์แวร์

ปัจจุบันมีซอฟต์แวร์หลากหลายจากบริษัทต่างๆ ดังนั้นการเลือกซอฟต์แวร์จำเป็นต้องเลือกให้เหมาะสม ตรงตามความต้องการมากที่สุด โดยซอฟต์แวร์จะต้องมีคุณสมบัติหลักดังนี้

1. มีเครื่องมือช่วยในการเตรียมข้อมูล ซึ่งรวมถึงการแปลงข้อมูลจากแหล่งข้อมูลหลายๆ แหล่ง หลายรูปแบบให้สามารถนำมาใช้กับดาต้าไมน์นิ่งได้อย่างถูกต้อง
2. เปิดโอกาสให้เลือกใช้เทคนิคการวิเคราะห์ต่างๆ ของดาต้าไมน์นิ่งได้หลายเทคนิค
3. ให้ผู้ใช้เลือกซื้อซอฟต์แวร์ได้ตามขนาดของข้อมูลและประสิทธิภาพตามต้องการได้
4. มีการแสดงผลของข้อมูลที่ชัดเจน เข้าใจง่ายและน่าสนใจ

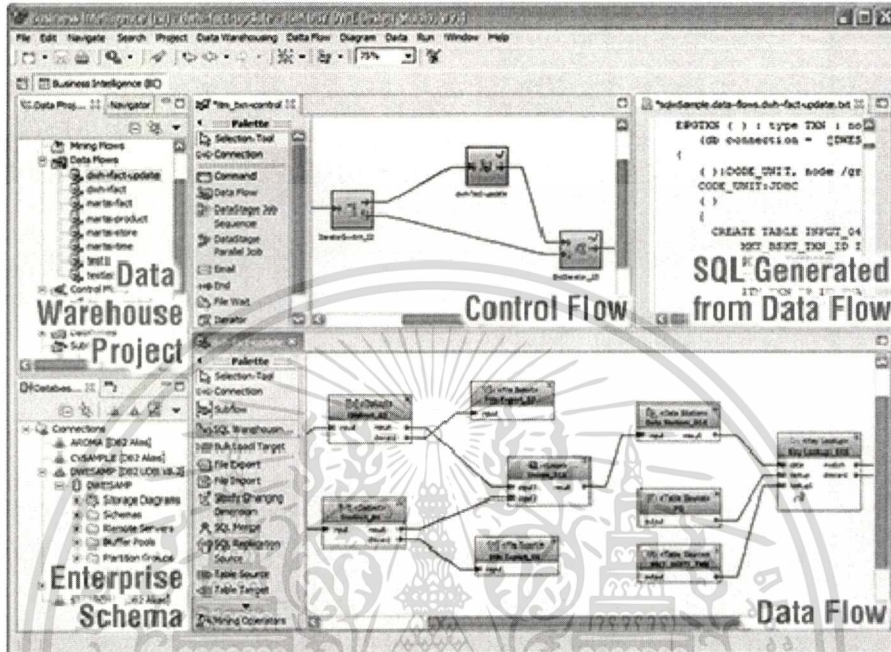
อย่างไรก็ตามซอฟต์แวร์สำหรับการทำดาต้าไมน์นิ่งเป็นเพียงเครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลเท่านั้น ไม่สามารถเข้าใจธุรกิจหรือเข้าใจข้อมูลได้ดีเท่าคน ดังนั้นผู้ใช้ซอฟต์แวร์สำหรับการทำดาต้าไมน์นิ่งจึงจำเป็นต้องมีความรู้ความเข้าใจในข้อมูลธุรกิจ เครื่องมือที่นำมาใช้งาน และอัลกอริทึมของเทคนิคดาต้าไมน์นิ่งแบบต่างๆ เป็นอย่างดี

และเทคนิคดาต้าไมน์นิ่งเป็นเพียงเครื่องมือที่ช่วยค้นหาคำตอบ รูปแบบ ความสัมพันธ์ของข้อมูลและแสดงผลลัพธ์ในรูปแบบต่างๆ ซึ่งเป็นเครื่องมือที่ช่วยนักวิเคราะห์หรือผู้บริหารในการตัดสินใจหรือสนับสนุนการตัดสินใจ ทำให้การวิเคราะห์หรือตัดสินใจกับผลลัพธ์ที่ได้ในแต่ละครั้งนั้น ไม่เหมือนกันขึ้นอยู่กับทักษะ ประสบการณ์ และความเชี่ยวชาญของนักวิเคราะห์หรือผู้บริหารเอง

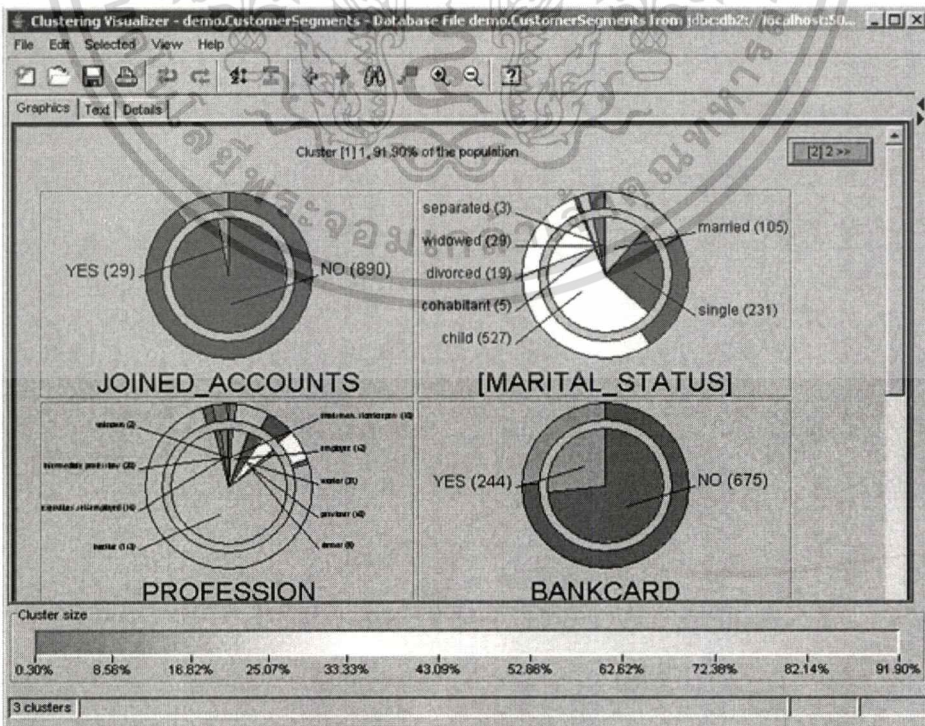
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 5.3 ตัวอย่างบริษัทที่พัฒนาผลิตภัณฑ์เกี่ยวกับดาต้าไมนนิ่ง

1. **Intelligent Miner** โดยบริษัทไอบีเอ็ม ใช้เทคนิคการแยกประเภท (Classification) กฎความเกี่ยวข้องกัน (Association Rule) และแบบจำลองการทำนาย (Predictive Model) ใช้กับระบบปฏิบัติการ UNIX (AIX) บน DB2 ตัวอย่างหน้าจอการใช้งาน โปรแกรมดังรูปที่ 5.1 และ 5.2



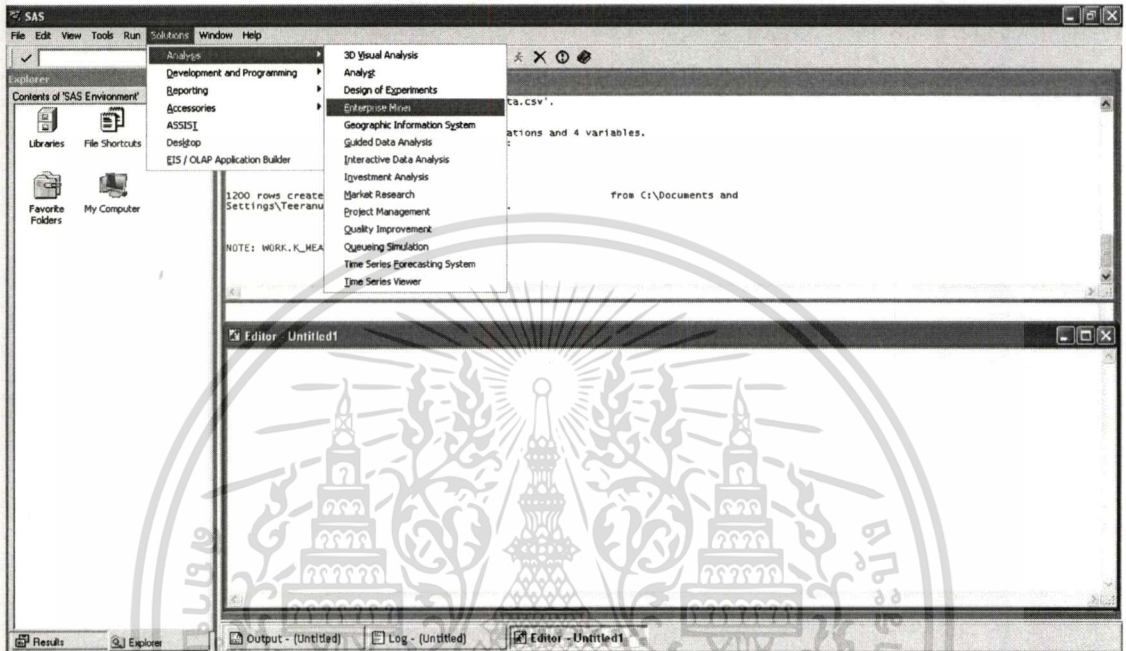
รูปที่ 5.1 หน้าจอเริ่มเปิดโปรแกรมของซอฟต์แวร์ Intelligent Miner



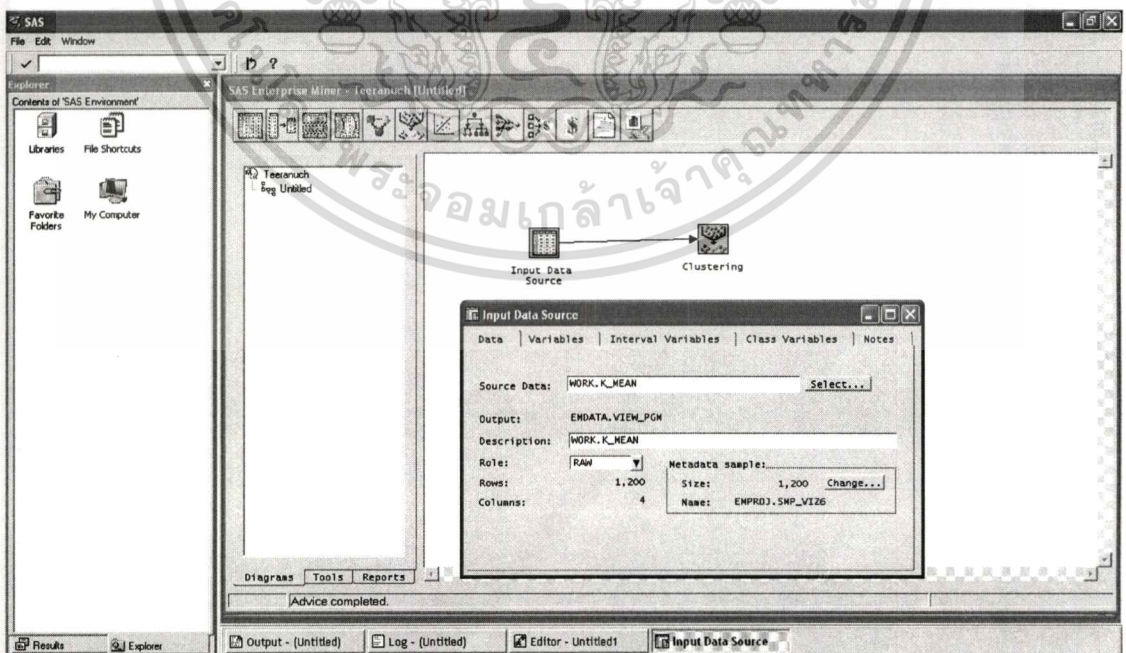
รูปที่ 5.2 หน้าจอการแสดงผลลัพธ์ในรูปแบบของแผนภาพของซอฟต์แวร์ Intelligent Miner

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. **Enterprise Miner** โดยบริษัท SAS ใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) กฎความเกี่ยวข้องกัน (Association Rule) เครือข่ายประสาทเทียม (Neural Network) และการจัดกลุ่ม (Clustering) ใช้กับระบบปฏิบัติการ UNIX (Solaris) หรือ Windows NT ติดต่อกับฐานข้อมูลผ่าน ODBC และสามารถสนับสนุน Oracle ตัวอย่างหน้าจอการใช้งานโปรแกรมดังรูปที่ 5.3 และ 5.4



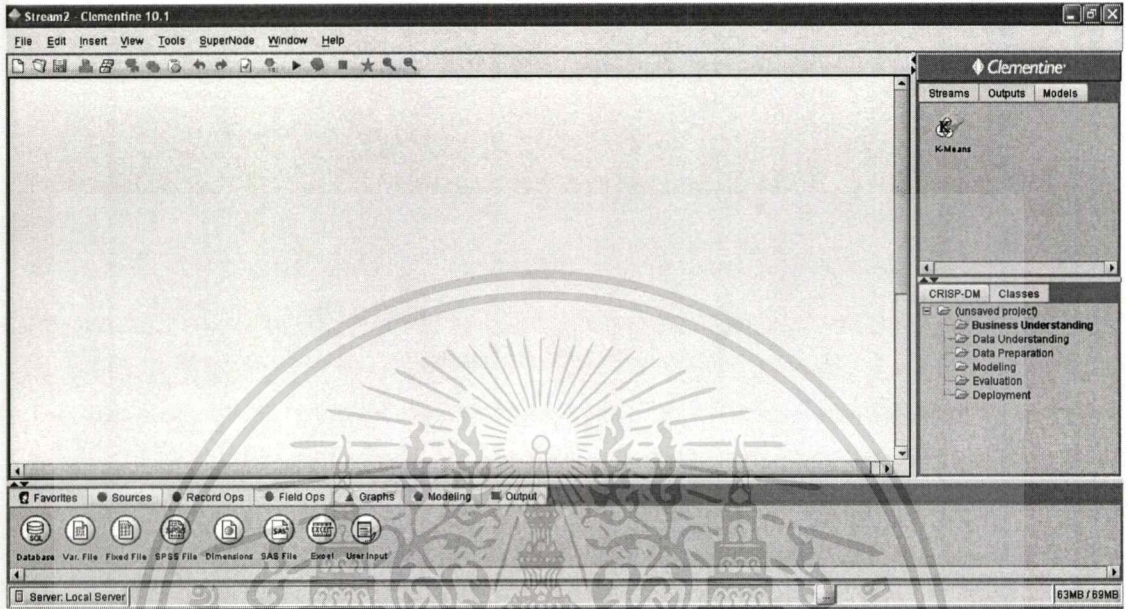
รูปที่ 5.3 หน้าจอเริ่มเปิดโปรแกรมของซอฟต์แวร์ Enterprise Miner



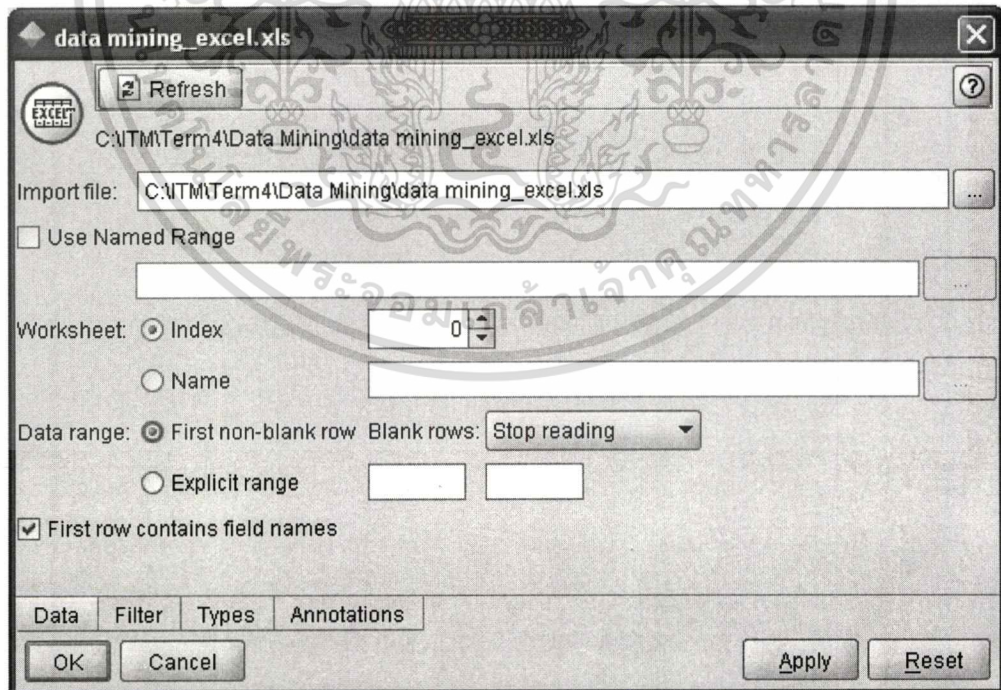
รูปที่ 5.4 หน้าจอการนำข้อมูลเข้าสู่โปรแกรมและใช้เทคนิคดาต้าไมน์นิ่งของ Enterprise Miner

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. **Clementine** โดยบริษัท SPSS ใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) กฎความเกี่ยวข้องกัน (Association Rule) เครือข่ายประสาทเทียม (Neural Network) และการจัดกลุ่ม (Clustering) ใช้กับระบบปฏิบัติการ Windows, Mac และ Linux ตัวอย่างหน้าจอการใช้งานโปรแกรมดังรูปที่ 5.5 – 5.7

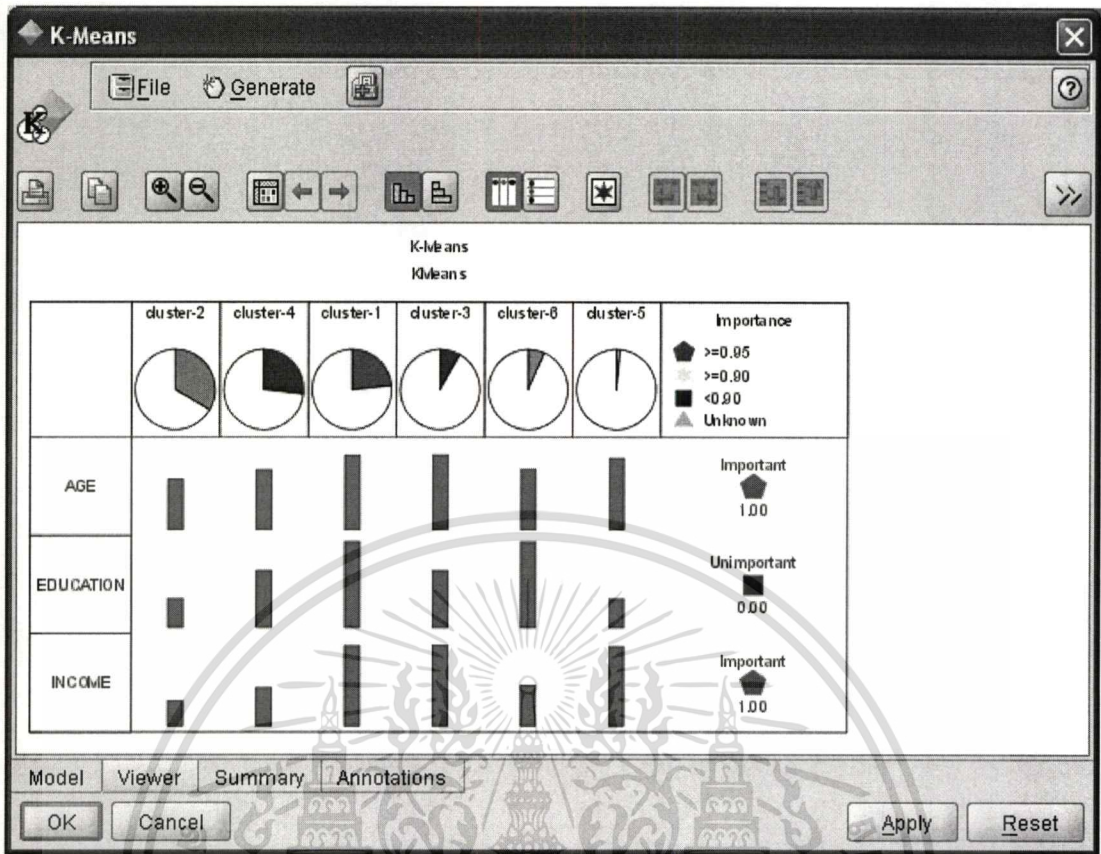


รูปที่ 5.5 หน้าจอเริ่มเปิดโปรแกรมของซอฟต์แวร์ Clementine



รูปที่ 5.6 หน้าจอการนำข้อมูลเข้าสู่โปรแกรมของซอฟต์แวร์ Clementine

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

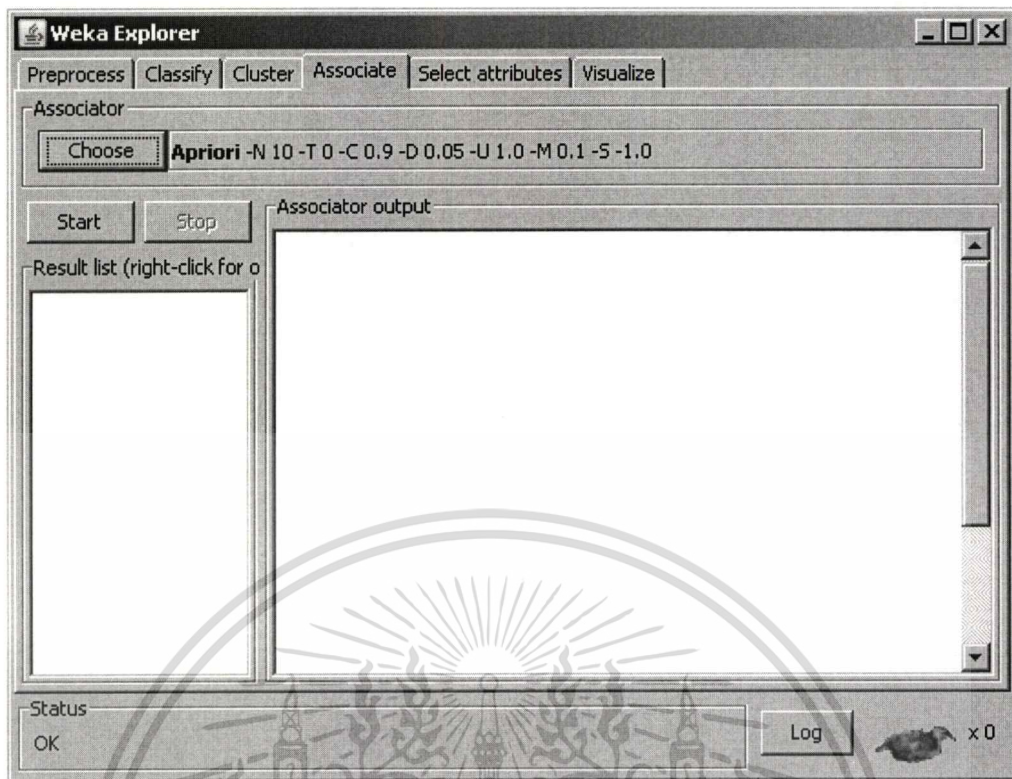


รูปที่ 5.7 หน้าจอการแสดงผลลัพธ์ในรูปแบบแผนภาพของซอฟต์แวร์ Clementine

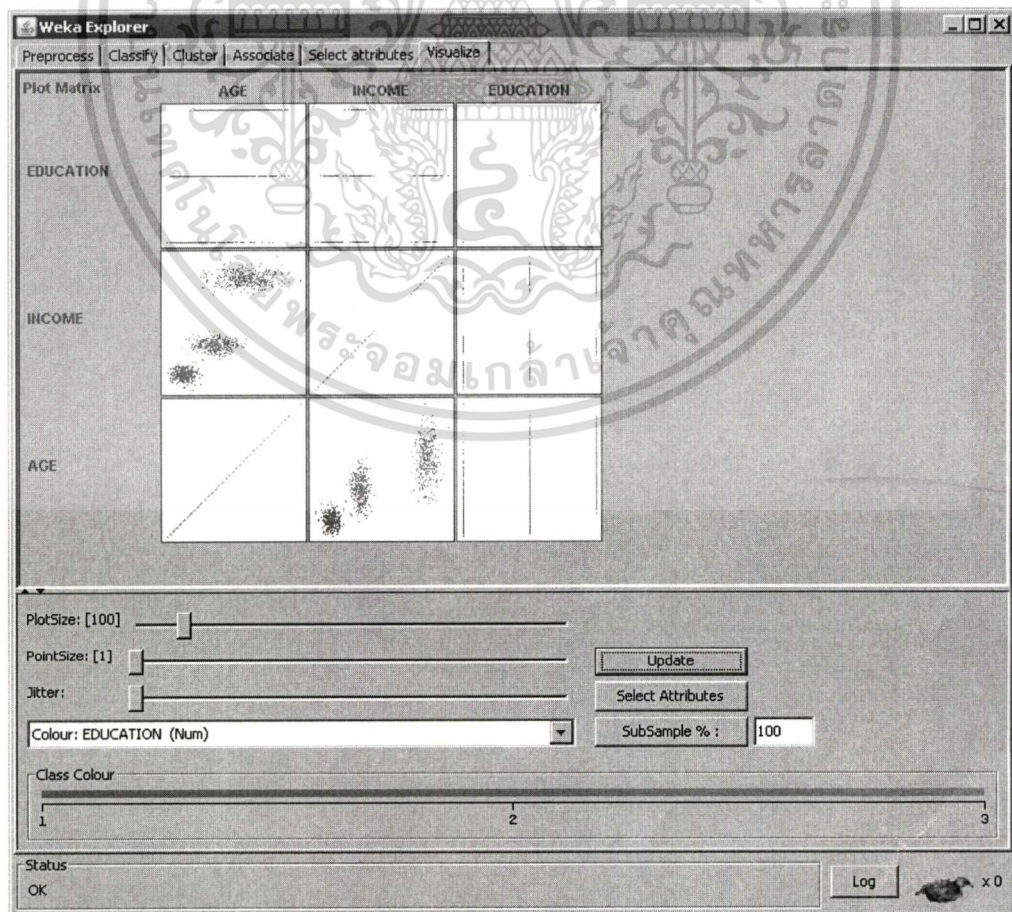
4. **WEKA** เป็นซอฟต์แวร์แบบ Open Source ซึ่งมีการพัฒนาอย่างต่อเนื่อง โดยใช้เทคนิคกฎความเกี่ยวข้องกัน (Association Rule) และการจัดกลุ่ม (Clustering) ใช้กับระบบปฏิบัติการ Windows NT ติดต่อกับฐานข้อมูลผ่าน ODBC ตัวอย่างหน้าจอการใช้งาน โปรแกรมดังรูปที่ 5.8 – 5.10



เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์รูปที่ 5.8 หน้าจอเริ่มเปิดโปรแกรมของซอฟต์แวร์ WEKA ใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.9 หน้าจอการเลือกเทคนิคดาต้าไมน์นิ่งแบบต่างๆ ของซอฟต์แวร์ WEKA



รูปที่ 5.10 หน้าจอการแสดงผลพีชในแบบแผนภาพของซอฟต์แวร์ WEKA

เอกสารนี้เป็นเอกสารที่เผยแพร่ภายใต้ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง การนำเอกสารนี้ไปใช้ในการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 6

# การประยุกต์ใช้เทคนิคดาต้าไมน์นิ่งด้วยซอฟต์แวร์ SAS

ในหัวข้อนี้จะกล่าวถึงการนำเทคนิคดาต้าไมน์นิ่งมาทำการวิเคราะห์ข้อมูลที่ได้จัดเตรียมไว้แล้วในบทที่ 3 และ 4 ซึ่งจะสอดคล้องกับความต้องการทางธุรกิจ โดยในแต่ละความต้องการทางธุรกิจหรือปัญหาทางธุรกิจนั้นจะใช้เทคนิคดาต้าไมน์นิ่งที่แตกต่างกันดังนั้นจำเป็นจะต้องเลือกเทคนิคดาต้าไมน์นิ่งให้เหมาะสมกับความต้องการนั้นๆ

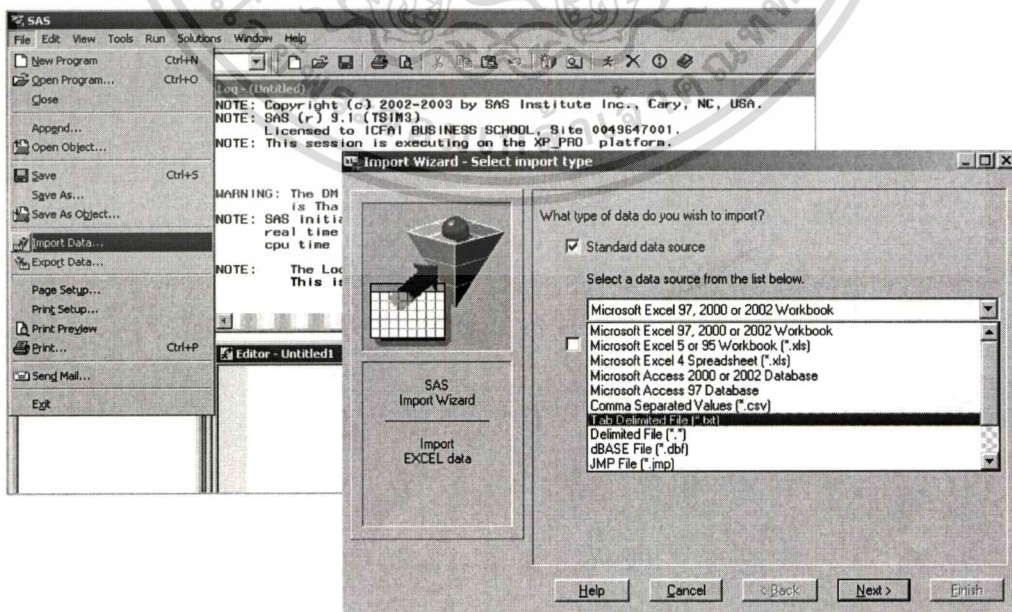
### 6.1 วิเคราะห์พฤติกรรมการณ์ซื้อสินค้าของลูกค้า ได้แก่ “ลูกค้าส่วนใหญ่เมื่อซื้อสินค้าอย่างหนึ่งแล้วมักจะซื้อสินค้าคู่กับสินค้านั้นอะไร”

ในหัวข้อนี้จะวิเคราะห์พฤติกรรมการณ์ของลูกค้าด้วยเทคนิคการค้นหากฎความสัมพันธ์ ซึ่งเป็นเทคนิคที่นิยมใช้วิเคราะห์หาความสัมพันธ์ของข้อมูลที่เกิดขึ้นในรายการเดียวกัน โดยผลลัพธ์ที่ได้จากการทำดาต้าไมน์นิ่งด้วยเทคนิคนี้จะได้ออกมาเป็นการหากฎความสัมพันธ์ของข้อมูล

ซึ่งปัจจัยที่เกี่ยวข้องในการพิจารณาความสำคัญของกฎหรือผลลัพธ์ที่ได้คือ ค่าความถี่ของข้อมูล ค่าความน่าเชื่อถือของข้อมูล และค่าลิฟท์

#### 6.1.1 ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ซอฟต์แวร์ SAS

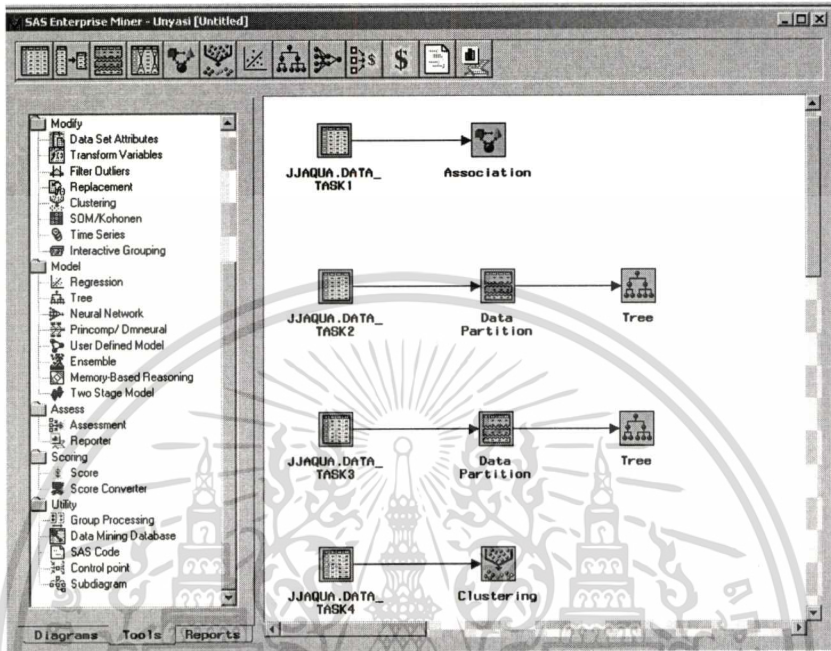
1. นำข้อมูลจากฐานข้อมูลที่ได้ออกแบบไว้เข้าสู่โปรแกรม โดยเลือกเมนู File > Import Data > เลือกประเภทของข้อมูลจากแหล่งข้อมูล ซึ่งจะเห็นได้ว่าสามารถเลือกแหล่งข้อมูลได้หลากหลายประเภท



รูปที่ 6.1 การนำข้อมูลเข้าสู่โปรแกรมและเลือกประเภทของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เข้าสู่เมนูการวิเคราะห์ข้อมูล เลือกเมนู Solution > Analyze > Enterprise Miner
3. การสร้างโมเดล โดยเลือก ไอคอน ข้อมูลนำเข้า และ ไอคอน Association มาวางไว้ที่หน้าจอ จากนั้นทำการสร้างเส้นเชื่อมจาก ข้อมูลนำเข้า ไปสู่ Association



รูปที่ 6.2 การสร้าง โมเดลด้วยเทคนิคกฎความสัมพันธ์

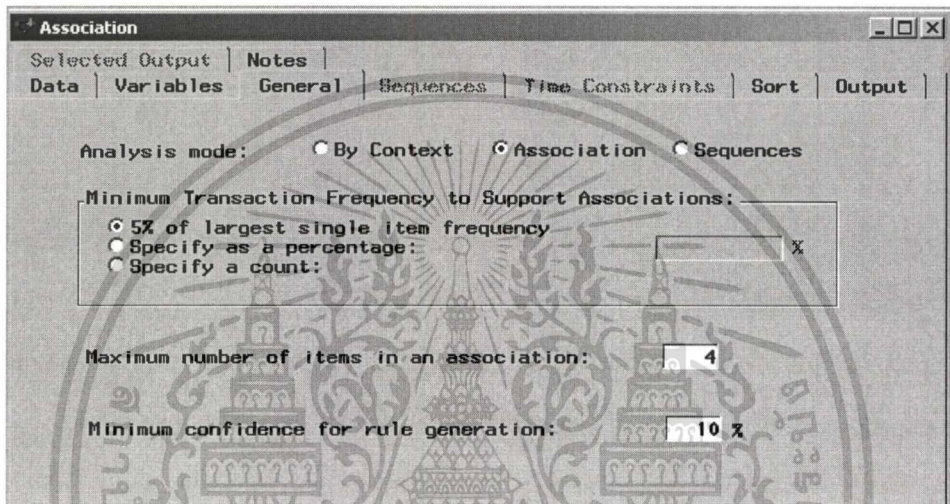
4. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับ ข้อมูลนำเข้า ในแท็บ Data จะทำการเลือกข้อมูลที่ได้เตรียมไว้แล้ว และในแท็บ Variables แต่ละเทคนิคจะกำหนดค่าไม่เหมือนกันโดยเทคนิคกฎความสัมพันธ์นี้จะต้องการกำหนดค่าของข้อมูล 2 ค่าด้วยกันคือ id และ target โดย id คือ ข้อมูลทรานแซกชัน และ target คือ ข้อมูลที่ต้องการให้หากฎความสัมพันธ์

The screenshot shows the 'Input Data Source' dialog box in SAS Enterprise Miner. It has tabs for 'Data', 'Variables', 'Interval Variables', 'Class Variables', and 'Notes'. The 'Data' tab is active, showing 'Source Data' as 'JJAQUA.DATA\_TASK1' and 'Output' as 'EMDATA.VIEM\_XGB'. The 'Variables' tab is also visible, showing a table with columns for Name, Model Role, Measurement, Type, and F. The table contains two rows: 'CUSTOMER\_ID' with Model Role 'id' and Type 'num', and 'PRODUCT' with Model Role 'target' and Type 'char'. A dropdown menu is open over the 'id' role, showing options like 'input', 'target', 'rejected', 'id', 'group', 'cost', 'freq', 'brial', 'predict', 'sequence', 'crossid', and 'timeid'. The 'id' option is currently selected.

เอกสารนี้เป็นรูปที่ 6.3 การกำหนดค่าพารามิเตอร์ของข้อมูลนำเข้าสำหรับเทคนิคกฎความสัมพันธ์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

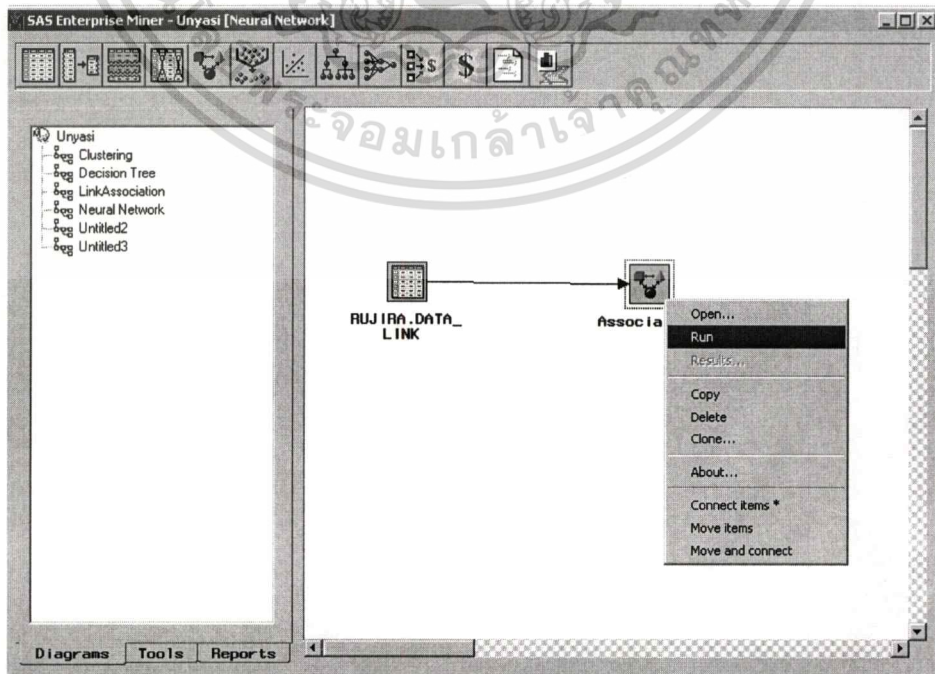
5. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับ Association Model โดยจะมีการกำหนดค่าความถี่น้อยที่สุดและค่าความน่าเชื่อถือน้อยที่สุด เพื่อตัดกฎความสัมพันธ์ที่ไม่น่าสนใจออกและลดเวลาในการรันโมเดลอีกด้วย ในการกำหนดค่าพารามิเตอร์อาจมีการกำหนดใหม่ได้หลังจากได้วิเคราะห์ผลลัพธ์ที่ออกมาแล้ว ยังไม่เป็นที่น่าพอใจหรือได้กฎความสัมพันธ์ที่ไม่น่าสนใจโดยได้กำหนดค่าพารามิเตอร์ต่างๆ ดังนี้

- ความถี่น้อยที่สุด (Minimum Transaction Frequency to Support Associations) เท่ากับ 5%
- ค่าความน่าเชื่อถือน้อยที่สุด (Minimum confidence for rule generation) เท่ากับ 10%



รูปที่ 6.4 การกำหนดค่าพารามิเตอร์ของ Association Model สำหรับเทคนิคกฎความสัมพันธ์

## 6. ทำการรันโปรแกรม



เอกสารนี้เป็นเอกสารที่สงวนรูปที่ 6.5 ผลลัพธ์ในแท็บ Rules ของเทคนิคกฎความสัมพันธ์ใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 7. ผลลัพธ์จากขั้นตอนที่ 6

| Rules | Frequencies | Code | Log        | Notes         |                   |   |
|-------|-------------|------|------------|---------------|-------------------|---|
|       | Relations   | Lift | Support(%) | Confidence(%) | Transaction Count | Rule                                      |
| 1     | 2           | 5.16 | 3.07       | 35.85         | 19.00             | Water pump ==> Plant                      |
| 2     | 2           | 6.08 | 3.07       | 42.22         | 19.00             | UV Sterilizer ==> Cutting                 |
| 3     | 2           | 5.16 | 3.07       | 44.19         | 19.00             | Plant ==> Water pump                      |
| 4     | 2           | 6.08 | 3.07       | 44.19         | 19.00             | Cutting ==> UV Sterilizer                 |
| 5     | 2           | 4.76 | 2.91       | 40.00         | 18.00             | Heaters ==> Temper Glass Aquarium         |
| 6     | 2           | 4.76 | 2.91       | 34.62         | 18.00             | Temper Glass Aquarium ==> Heaters         |
| 7     | 2           | 4.66 | 2.91       | 34.62         | 18.00             | Temper Glass Aquarium ==> Water Test Kits |
| 8     | 2           | 4.66 | 2.91       | 39.13         | 18.00             | Water Test Kits ==> Temper Glass Aquarium |
| 9     | 2           | 3.53 | 2.58       | 29.63         | 16.00             | Thermometers ==> Dissolved Oxygen         |
| 10    | 2           | 3.97 | 2.58       | 27.59         | 16.00             | Fish food ==> Plant                       |
| 11    | 2           | 3.53 | 2.58       | 30.77         | 16.00             | Dissolved Oxygen ==> Thermometers         |
| 12    | 2           | 3.97 | 2.58       | 37.21         | 16.00             | Plant ==> Fish food                       |
| 13    | 2           | 4.80 | 2.42       | 33.33         | 15.00             | UV Sterilizer ==> Lighting                |
| 14    | 2           | 4.80 | 2.42       | 34.88         | 15.00             | Lighting ==> UV Sterilizer                |
| 15    | 2           | 2.77 | 2.26       | 24.14         | 14.00             | Fish food ==> Thermometers                |
| 16    | 2           | 2.82 | 2.26       | 24.14         | 14.00             | Fish food ==> Water pump                  |
| 17    | 2           | 2.77 | 2.26       | 25.93         | 14.00             | Thermometers ==> Fish food                |
| 18    | 2           | 3.80 | 2.26       | 26.42         | 14.00             | Water pump ==> Lighting                   |
| 19    | 2           | 3.80 | 2.26       | 32.56         | 14.00             | Lighting ==> Water pump                   |
| 20    | 2           | 2.82 | 2.26       | 26.42         | 14.00             | Water pump ==> Fish food                  |
| 21    | 2           | 3.59 | 1.94       | 26.67         | 12.00             | Heaters ==> Water Test Kits               |
| 22    | 2           | 3.59 | 1.94       | 26.09         | 12.00             | Water Test Kits ==> Heaters               |
| 23    | 2           | 3.05 | 1.78       | 25.58         | 11.00             | CO2-Tube ==> Dissolved Oxygen             |
| 24    | 2           | 3.05 | 1.78       | 21.15         | 11.00             | Dissolved Oxygen ==> CO2-Tube             |
| 25    | 2           | 3.35 | 1.62       | 23.26         | 10.00             | Cutting ==> Lighting                      |
| 26    | 2           | 3.20 | 1.62       | 23.81         | 10.00             | Biological filter ==> Water Test Kits     |
| 27    | 2           | 7.11 | 1.62       | 17.24         | 10.00             | Fish food ==> Pond Liner                  |
| 28    | 2           | 7.11 | 1.62       | 66.67         | 10.00             | Pond Liner ==> Fish food                  |
| 29    | 2           | 3.35 | 1.62       | 23.26         | 10.00             | Lighting ==> Cutting                      |
| 30    | 2           | 3.35 | 1.62       | 23.26         | 10.00             | Plant ==> Lighting                        |
| 31    | 2           | 2.87 | 1.62       | 20.83         | 10.00             | External filter ==> Heaters               |
| 32    | 2           | 3.35 | 1.62       | 23.26         | 10.00             | Lighting ==> Plant                        |
| 33    | 2           | 3.20 | 1.62       | 21.74         | 10.00             | Water Test Kits ==> Biological filter     |

รูปที่ 6.6 ผลลัพธ์ในแท็บ Rules ของเทคนิคกฎความสัมพันธ์

| Rules | Frequencies | Code                           | Log | Notes |
|-------|-------------|--------------------------------|-----|-------|
|       | Count       | Item                           |     |       |
| 1     | 58          | Fish food                      |     |       |
| 2     | 54          | Thermometers                   |     |       |
| 3     | 53          | Water pump                     |     |       |
| 4     | 52          | Temper Glass Aquarium          |     |       |
| 5     | 52          | Dissolved Oxygen               |     |       |
| 6     | 48          | External filter                |     |       |
| 7     | 46          | Water Test Kits                |     |       |
| 8     | 45          | UV Sterilizer                  |     |       |
| 9     | 45          | Heaters                        |     |       |
| 10    | 43          | Plant                          |     |       |
| 11    | 43          | Lighting                       |     |       |
| 12    | 43          | Cutting                        |     |       |
| 13    | 43          | CO2-Tube                       |     |       |
| 14    | 42          | Biological filter              |     |       |
| 15    | 15          | Pond Liner                     |     |       |
| 16    | 15          | Manifolds                      |     |       |
| 17    | 15          | Garden pond                    |     |       |
| 18    | 14          | Conductivity/TDS Meters        |     |       |
| 19    | 13          | Photometers (Ion Specific)     |     |       |
| 20    | 13          | Multi-Parameter Instruments    |     |       |
| 21    | 12          | Aquarium include Filter System |     |       |

รูปที่ 6.7 ผลลัพธ์ในแท็บ Frequencies ของเทคนิคกฎความสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 6.1.2 วิเคราะห์ผลลัพธ์พร้อมทั้งนำเสนอรูปแบบผลลัพธ์ที่ได้

จากการรันโปรแกรมด้วยโมเดลกฎความสัมพันธ์พบว่า โปรแกรมสร้างกฎความสัมพันธ์ให้หลากหลายกฎ มากกว่า 1,000 กฎด้วยกัน ทำให้ยากในการพิจารณาหากฎที่น่าสนใจ ดังนั้นจะต้องมีการพิจารณาค่าพารามิเตอร์ต่างๆ ประกอบการวิเคราะห์ โดยกฎความสัมพันธ์ที่น่าสนใจนั้นจะดูจากค่าความถี่ ค่าความน่าเชื่อถือที่มาก และค่าลิฟท์ ยิ่งมากยิ่งหมายความว่ากฎความสัมพันธ์ที่น่าสนใจ ซึ่งจะเห็นได้ว่าเมื่อกำหนดค่าความถี่มากกว่าหรือเท่ากับ 2% และค่าความน่าเชื่อถือมากกว่าหรือเท่ากับ 30% จำนวนกฎลดลงเหลือ 12 กฎความสัมพันธ์ ซึ่งทำให้สามารถวิเคราะห์กฎความสัมพันธ์ได้ง่ายขึ้น จากนั้นพิจารณากฎความสัมพันธ์ที่น่าสนใจและเป็นประโยชน์ต่อบริษัท จากรูปที่ 6.6 สามารถแสดงกฎความสัมพันธ์ที่น่าสนใจได้ดังนี้

|    |                     |                       |        |          |             |
|----|---------------------|-----------------------|--------|----------|-------------|
| 1. | Cutting =>          | UV Sterilizer         | [3.07% | 44.19%], | Lift = 6.08 |
| 2. | Heater =>           | Temper Glass Aquarium | [2.91% | 40.00%], | Lift = 4.76 |
| 3. | Water Test Kits =>  | Temper Glass Aquarium | [2.91% | 39.13%], | Lift = 4.66 |
| 4. | Dissolved Oxygen => | Thermometers          | [2.58% | 30.77%], | Lift = 3.53 |

จากกฎความสัมพันธ์ที่ได้พบว่ามีกฎความสัมพันธ์ที่น่าสนใจ 4 กฎและมีสินค้าที่ลูกค้านิยมซื้อควบคู่กันจากกฎความสัมพันธ์ที่ได้ 7 รายการสามารถนำไปช่วยในการสนับสนุนการตัดสินใจในการวางแผนกลยุทธ์ด้านการตลาด การขายได้

ตัวอย่างกลยุทธ์ทางการตลาด ได้แก่ การจัด โปรโมชันประจำเดือน ซึ่งสินค้าที่มักซื้อคู่กันจะไม่จัดโปรโมชันคู่กันอีกเนื่องจากเป็นสินค้าที่ลูกค้ามักจะซื้ออยู่แล้วแต่อาจจะเพิ่มราคาสินค้าอย่างใดอย่างหนึ่งเพื่อเพิ่มยอดขาย ลดราคาสินค้าอย่างใดอย่างหนึ่งเพื่อเพิ่มปริมาณการสั่งซื้อหรือลดราคาของสินค้าที่เกี่ยวข้องกับสินค้าทั้ง 7 รายการนี้เพื่อเป็นการเพิ่มยอดขายแทน เป็นต้น

จากผลลัพธ์ดังรูปที่ 6.7 ซึ่งบอกจำนวนปริมาณการสั่งซื้อสินค้าแต่ละประเภท ซึ่งสามารถนำไปช่วยสนับสนุนการตัดสินใจในเรื่องการจัดเก็บคลังสินค้าได้เช่น อาหารปลาหมักยอดการสินค้ามากและเป็นสินค้าที่ลูกค้านิยมซื้อคู่กับต้น ไม้ดังนั้น จำเป็นต้องมีการจัดซื้อสินค้าทั้ง 2 ชนิดนี้ให้เพียงพอกับความต้องการของลูกค้า หรืออาจนำไปเป็นข้อมูลในการต่อรองราคากับผู้ผลิตในการสั่งซื้อสินค้าปริมาณมาก และสำหรับสินค้าบางประเภทซึ่งพบว่าการสั่งซื้อสินค้าปริมาณน้อยและไม่บ่อยจึงไม่ควรที่จะสั่งซื้อสินค้าชนิดนี้ไว้ปริมาณมากหรือหากมีลูกค้าสั่งซื้อจึงค่อยสั่งซื้อสินค้ากับผู้ผลิตเป็นครั้งๆ ไปเพื่อเป็นการลดต้นทุนในการสั่งซื้อสินค้าจากผู้ผลิตและลดต้นทุนในการจัดเก็บสินค้าอีกด้วย

## 6.2 วิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทและ วิธีการขนส่งสินค้า

ในหัวข้อนี้จะวิเคราะห์สาเหตุการส่งคืนสินค้าแยกตามประเภทและวิธีการขนส่งสินค้าด้วยเทคนิคแผนภาพต้นไม้ซึ่งเป็นเทคนิคที่นิยมใช้ในการหาคุณลักษณะและทำนายสิ่งที่จะเกิดขึ้นในอนาคต โดยได้เน้นถึงปัจจัยที่ส่งผลให้เกิดการส่งคืนสินค้า โดยผลลัพธ์ที่ได้จะสามารถนำไปสนับสนุนการตัดสินใจในการเรื่องการให้บริการให้ลูกค้าได้รับความพึงพอใจมากที่สุด

### 6.2.1 ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ซอฟต์แวร์ SAS

- นำข้อมูลจากฐานข้อมูลที่ได้ออกแบบไว้เข้าสู่โปรแกรม โดยเลือกเมนู File > Import Data > เลือกประเภทของข้อมูลจากแหล่งข้อมูล ซึ่งโปรแกรมสามารถรองรับประเภทของแหล่งข้อมูลได้หลากหลายประเภท
- เข้าสู่เมนูการวิเคราะห์ข้อมูล เลือกเมนู Solution > Analyze > Enterprise Miner
- การสร้างโมเดล โดยเลือก ไอคอน ข้อมูลนำเข้า ไอคอน Data Partition และ ไอคอน Tree มาวางไว้ที่หน้าจอ จากนั้นทำการสร้างเส้นเชื่อมจาก ข้อมูลนำเข้า ไปสู่ Tree



รูปที่ 6.8 การสร้างโมเดลด้วยเทคนิคแผนภาพต้นไม้

- กำหนดค่าพารามิเตอร์ต่างๆ สำหรับ ข้อมูลนำเข้า ในแท็บ Data จะทำการเลือกข้อมูลที่ได้เตรียมไว้แล้ว และในแท็บ Variables แต่ละเทคนิคจะกำหนดค่าไม่เหมือนกันโดยเทคนิคแผนภาพต้นไม้จะสนใจข้อมูลที่เป็น Target เป็นหลักจะต้องทำการกำหนดค่าของข้อมูลประเภทการส่งคืนเป็น Target

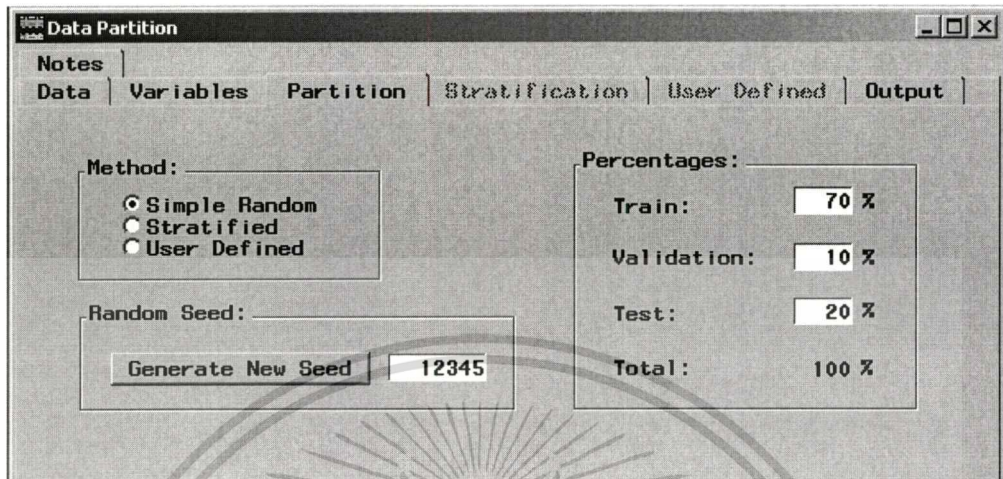
| Input Data Source |           |                    |             |                 |        |          |
|-------------------|-----------|--------------------|-------------|-----------------|--------|----------|
| Data              | Variables | Interval Variables |             | Class Variables |        | Notes    |
|                   | Name      | Model Role         | Measurement | Type            | Format | Informat |
|                   | CAT_NAME  | input              | nominal     | char            | \$9.   | \$9.     |
|                   | SPT_TYPE  | target             | nominal     | char            | \$10.  | \$10.    |
|                   | CLM_TITLE | input              | nominal     | char            | \$11.  | \$11.    |

รูปที่ 6.9 การกำหนดค่าพารามิเตอร์ของ ข้อมูลนำเข้า สำหรับเทคนิคแผนภาพต้นไม้

- กำหนดค่าพารามิเตอร์ต่างๆ สำหรับ Data Partition ในส่วนนี้จะเป็เครื่องมือที่ช่วยในการแบ่งข้อมูลออกเป็นส่วย่อย ซึ่งการใช้เทคนิคนี้จะต้องแบ่งข้อมูลออกเป็น 3 ส่วนคือ Training, Validation และ Test ซึ่งเริ่มต้นจะมีการนำข้อมูล Training มาสร้างโมเดลหลังจากนั้นจะใช้ข้อมูล Validation และ Test ในการทดสอบโมเดลว่ามีความถูกต้องมากเพียงใด โดยโปรแกรมสามารถให้

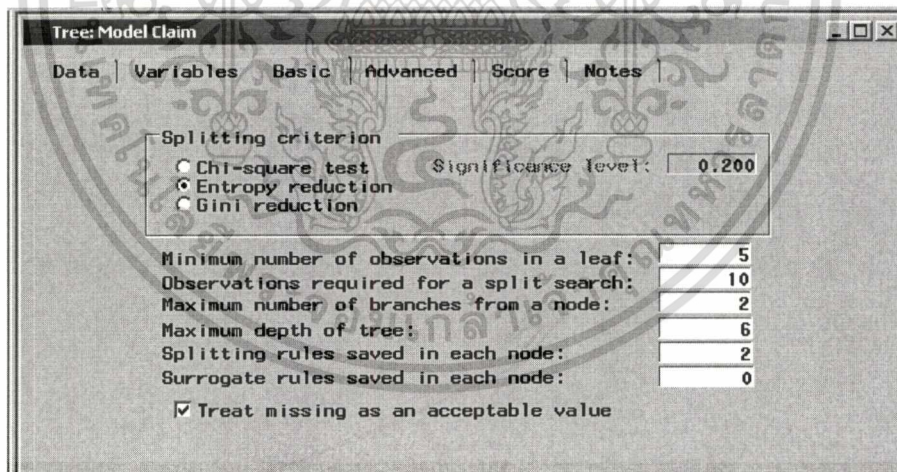
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดเปอร์เซ็นต์ของข้อมูลทั้ง 3 ส่วนนี้ได้อย่างอิสระซึ่งโดยทั่วไปมีการแบ่งข้อมูลออกเป็น 70%-10%-20%



รูปที่ 6.10 การกำหนดค่าพารามิเตอร์ของ Data Partition สำหรับเทคนิคแผนภาพต้นไม้

6. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับโมเดลแผนภาพต้นไม้ โดยกำหนดที่แท็บ Basic เป็นการเลือกเงื่อนไขและอัลกอริทึมที่ใช้ในการสร้างโมเดล โดยในที่นี้จะเลือก Entropy reduction



รูปที่ 6.11 การกำหนดค่าพารามิเตอร์ของ Tree Model สำหรับเทคนิคแผนภาพต้นไม้

7. ทำการรันโปรแกรม
8. ผลลัพธ์จากขั้นตอนที่ 7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ตารางที่ 6.1 ผลลัพธ์จากการรัน โปรแกรมด้วยเทคนิคแผนภาพต้นไม้ในรูปแบบของกฎ “ถ้า...แล้ว”  
สำหรับหัวข้อที่ 6.2

|   |  |
|---|--|
| IF CAT_Name EQUALS AQUARIUM<br>THEN<br>NODE : 2<br>N : 13<br>TNT : 38.5%<br>EMS : 38.5%<br>MOTORCYC: 23.1%<br>PARCEL : 0.0%   | IF CAT_Name EQUALS CHEMICAL<br>THEN<br>NODE : 4<br>N : 17<br>TNT : 0.0%<br>EMS : 70.6%<br>MOTORCYC: 29.4%<br>PARCEL : 0.0%   |
| IF CAT_Name EQUALS PLANT<br>THEN<br>NODE : 5<br>N : 7<br>TNT : 0.0%<br>EMS : 0.0%<br>MOTORCYC: 100.0%<br>PARCEL : 0.0%  | IF CLM_Title EQUALS MISS ORDER<br>AND CAT_Name IS ONE OF: EQUIPMENT FISH FOOD<br>THEN<br>NODE : 6<br>N : 20<br>TNT : 20.0%<br>EMS : 50.0%<br>MOTORCYC: 20.0%<br>PARCEL : 10.0% |
| IF CLM_Title EQUALS BROKEN<br>AND CAT_Name IS ONE OF: EQUIPMENT FISH FOOD<br>THEN<br>NODE : 7<br>N : 22<br>TNT : 0.0%<br>EMS : 31.8%<br>MOTORCYC: 40.9%<br>PARCEL : 27.3% | IF CLM_Title EQUALS UNAVAILABLE<br>AND CAT_Name IS ONE OF: EQUIPMENT FISH FOOD<br>THEN<br>NODE : 8<br>N : 16<br>TNT : 0.0%<br>EMS : 50.0%<br>MOTORCYC: 50.0%<br>PARCEL : 0.0%  |
| IF CLM_Title EQUALS PACKAGE<br>AND CAT_Name IS ONE OF: EQUIPMENT FISH FOOD<br>THEN<br>NODE : 9<br>N : 5<br>TNT : 0.0%<br>EMS : 100.0%<br>MOTORCYC: 0.0%<br>PARCEL : 0.0%  |  |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 6.2.2 วิเคราะห์ผลลัพธ์พร้อมทั้งนำเสนอรูปแบบผลลัพธ์ที่ได้

จากการรัน โปรแกรมด้วยโมเดลแผนภาพต้นไม้พบว่าโปรแกรมได้ทำการสร้างแบบจำลอง และเสนอผลลัพธ์ออกมา 6 เงื่อนไขซึ่งดังนี้

1. ถ้าสินค้าประเภทตู้ปลา สาเหตุการคืนสินค้าทั้งหมดพบว่าเป็นการส่งสินค้าด้วย TNT และ Motorcycle เท่าๆ กัน
2. ถ้าสินค้าประเภทอุปกรณ์ตู้ คืนสินค้าด้วยสาเหตุสินค้าไม่ตรงตามความต้องการของลูกค้า พบว่ามีแนวโน้มในการส่งสินค้าแบบ TNT มากกว่าที่สุดและรองมาคือ แบบพัสดุไปรษณีย์
3. ถ้าสินค้าประเภทอุปกรณ์ตู้ คืนสินค้าด้วยสาเหตุสินค้าสินค้าแตกเสียหายไม่สามารถใช้งานได้ พบว่ามีเป็นการส่งสินค้าด้วยรถมอเตอร์ไซด์
4. ถ้าสินค้าประเภทอุปกรณ์ตู้ คืนสินค้าด้วยสาเหตุสินค้าสินค้าไม่สามารถใช้งานได้ พบว่ามีเป็นการส่งสินค้าแบบ EMS
5. ถ้าสินค้าประเภทอุปกรณ์ตู้ คืนสินค้าด้วยสาเหตุสินค้าสินค้าบรรจุหีบห่อไม่อยู่ในสภาพปกติ พบว่ามีเป็นการส่งสินค้าแบบ EMS
6. ถ้าสินค้าประเภทเคมีภัณฑ์ ที่มีการส่งคืนสินค้า พบว่าเป็นการส่งสินค้าแบบ EMS
7. ถ้าสินค้าประเภทต้นไม้ ที่มีการส่งคืนสินค้า พบว่าเป็นการส่งสินค้าด้วยรถมอเตอร์ไซด์ ซึ่งจากเงื่อนไขที่ได้ทั้ง 7 เงื่อนไขนี้สามารถนำไปช่วยสนับสนุนการวางแผนทางด้าน CRM โดยพิจารณาจากความเสียหายที่เกิดขึ้นซึ่งส่งผลให้เกิดความไม่พึงพอใจต่อลูกค้าได้ โดยพบว่าสาเหตุการส่งคืนสินค้านั้นมักเกิดจากการเลือกส่งสินค้าแบบ EMS และด้วยรถมอเตอร์ไซด์ซึ่งผู้เกี่ยวข้องจำเป็นต้องเข้าไปดูแลหรือปรับเปลี่ยนวิธีการขนส่งให้ดีขึ้นไม่ว่าจะเป็นการบรรจุหีบห่อ การเคลื่อนย้ายสินค้า และอีกสาเหตุที่ทำให้มีการส่งคืนสินค้านั้นคือการส่งสินค้าไม่ตรงตามความต้องการของลูกค้า ซึ่งถือว่าเป็นเรื่องสำคัญมาก ดังนั้นผู้ที่เกี่ยวข้องควรมีการพิจารณาถึงขั้นตอนการสั่งซื้อสินค้าและจัดส่งสินค้าให้มีความถูกต้องมากขึ้น และผู้บริหารอาจมีการพิจารณาการขนส่งสินค้ากับบริษัทขนส่งรายอื่นๆ เพื่อให้เกิดทางเลือกที่มากขึ้นและมีการพัฒนาระบบการขนส่งสินค้าที่ดีขึ้นได้

## 6.2.3 ข้อเสนอแนะจากการวิเคราะห์สาเหตุการขอคืนสินค้าของลูกค้า

เนื่องด้วยมีปัจจัยอื่นๆ ที่เกี่ยวข้องกับสาเหตุการขอคืนสินค้าของลูกค้าเช่น การปฏิบัติงานของพนักงานแต่ละฝ่ายที่เกี่ยวข้อง ซึ่งเป็นปัจจัยหลักในการขอคืนสินค้าจากสาเหตุสินค้าไม่ตรงตามความต้องการ หรือการคัดเลือกบริษัทผู้ผลิตที่ได้มีคุณภาพได้มาตรฐานซึ่งส่งผลต่อสาเหตุการขอคืนสินค้าเนื่องจากสินค้าไม่สามารถใช้งานได้ เป็นต้น โดยปัจจัยเหล่านี้อาจมีความสำคัญมากกว่าการพิจารณาประเภทการขนส่งสินค้าแต่ไม่มีการเก็บข้อมูลหรือไม่สามารถเก็บข้อมูลลงฐานข้อมูลได้ทำให้ข้อมูลในฐานข้อมูลที่นำมาวิเคราะห์สาเหตุนั้นไม่เพียงพอ ดังนั้นควรมีการพิจารณาเก็บข้อมูลปัจจัยอื่นๆ ที่เกี่ยวข้องเพิ่มมากขึ้น

ใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 6.3 วิเคราะห์พฤติกรรมของการเลือกขนส่งสินค้าของลูกค้าตามประเภทของสินค้าที่สั่งซื้อ

ในหัวข้อนี้จะวิเคราะห์พฤติกรรมการเลือกขนส่งสินค้าของลูกค้าตามประเภทของสินค้าที่สั่งซื้อ ซึ่งในการเลือกวิธีการขนส่งสินค้านี้ลูกค้าจะเป็นผู้เลือกและจ่ายค่าใช้จ่ายเอง ซึ่งการเลือกของลูกค้าอาจจะเลือกมาจากความคุ้นเคยหรือเคยได้รับรู้ถึงบริการและราคาค่าบริการเป็นหลัก โดยผลลัพธ์ที่ได้จะสามารถนำไปสนับสนุนการตัดสินใจในการเรื่องการให้บริการลูกค้าให้ได้รับความพึงพอใจมากที่สุด การวิเคราะห์พฤติกรรมของลูกค้านี้สามารถแสดงให้เห็นถึงความนิยมของลูกค้าและยังสามารถนำไปพิจารณาการพัฒนาระบบให้สามารถคัดเลือกหรือแนะนำวิธีการขนส่งให้เหมาะสมกับสินค้าของลูกค้าได้

#### 6.3.1 ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ซอฟต์แวร์ SAS

1. ข้อมูลจากฐานข้อมูลที่ได้ออกแบบไว้เข้าสู่โปรแกรม โดยเลือกเมนู File > Import Data > เลือกประเภทของข้อมูลจากแหล่งข้อมูล ซึ่งโปรแกรมสามารถรองรับประเภทของแหล่งข้อมูลได้หลากหลายประเภท

2. เข้าสู่เมนูการวิเคราะห์ข้อมูล เลือกเมนู Solution > Analyze > Enterprise Miner

3. การสร้างโมเดล โดยเลือก ไอคอน ข้อมูลนำเข้า ไอคอน Data Partition และ ไอคอน Tree มาวางไว้ที่หน้าจอ จากนั้นทำการสร้างเส้นเชื่อมจาก ข้อมูลนำเข้า ไปสู่ Tree

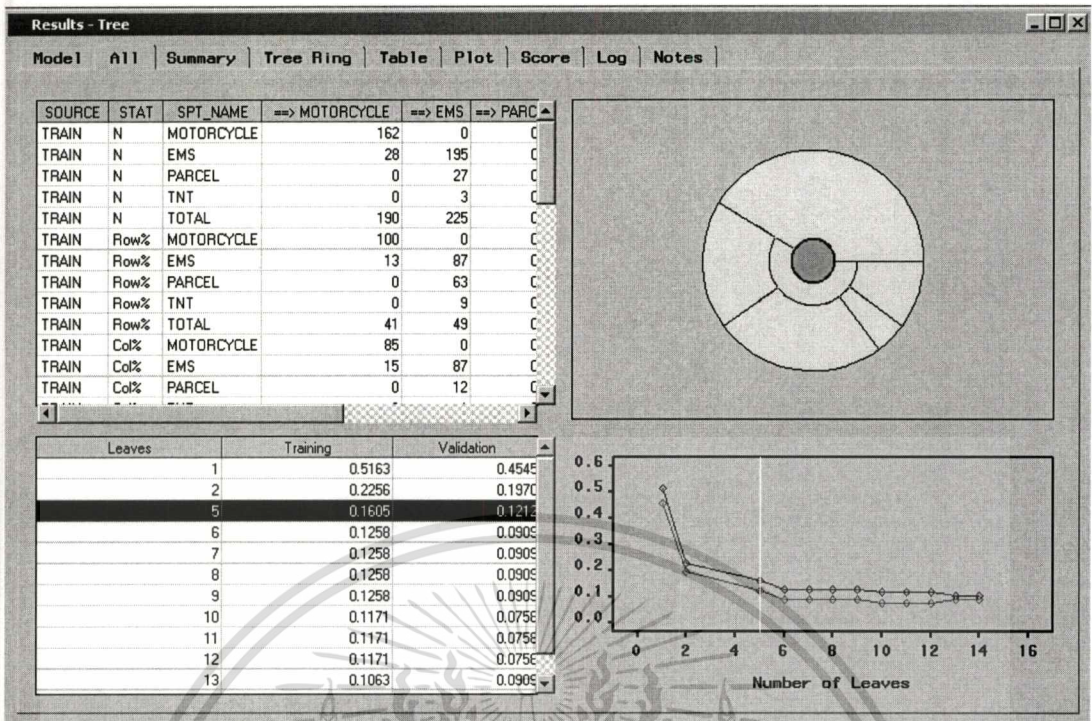
4. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับ ข้อมูลนำเข้า ในแท็บ Data จะทำการเลือกข้อมูลที่ได้เตรียมไว้แล้ว และในแท็บ Variables แต่ละเทคนิคจะกำหนดค่าไม่เหมือนกันโดยเทคนิคแผนภาพต้นไม้จะสนใจข้อมูลที่เป็น Target เป็นหลักจะต้องทำการกำหนดค่าของข้อมูลประเภทการส่งสินค้าเป็น Target

5. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับ Data Partition โดยแบ่งข้อมูลออกเป็น 70%-10%-20% ดังในหัวข้อ 6.2

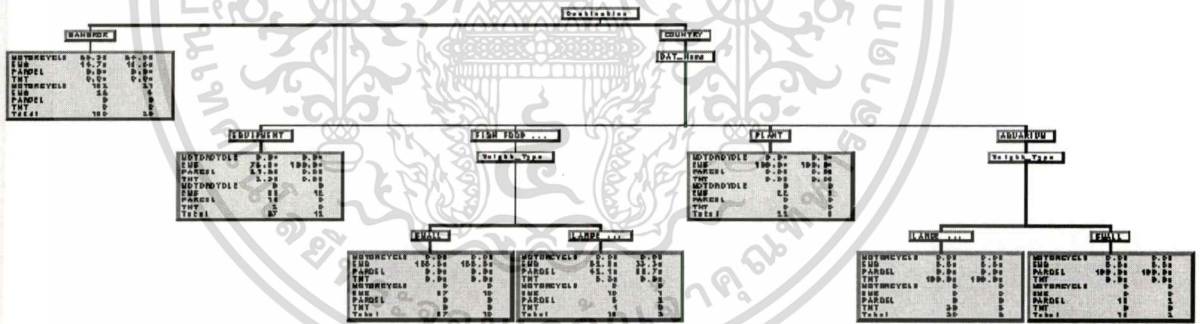
6. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับโมเดลแผนภาพต้นไม้ โดยกำหนดเหมือนกับหัวข้อ 6.2

7. ทำการรันโปรแกรม

8. ผลลัพธ์จากขั้นตอนที่ 7



รูปที่ 6.14 ผลลัพธ์ของเทคนิคแผนภาพต้นไม้สำหรับการขนส่งสินค้า



รูปที่ 6.15 ผลลัพธ์ในรูปแบบของแผนภาพของเทคนิคแผนภาพต้นไม้สำหรับการขนส่งสินค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 6.2 ผลลัพธ์จากการรันโปรแกรมด้วยเทคนิคแผนภาพต้นไม้ในรูปแบบของกฎ “ถ้า...แล้ว”  
สำหรับหัวข้อที่ 6.3

|   |  |
|---|--|
| <p>IF Destination EQUALS BANGKOK<br/>THEN<br/>NODE : 2<br/>N : 190<br/>MOTORCYC: 85.3%<br/>EMS : 14.7%<br/>PARCEL : 0.0%<br/>TNT : 0.0%</p>   | <p>IF CAT_Name EQUALS EQUIPMENT<br/>AND Destination EQUALS COUNTRY<br/>THEN<br/>NODE : 8<br/>N : 87<br/>MOTORCYC: 0.0%<br/>EMS : 75.9%<br/>PARCEL : 21.8%<br/>TNT : 2.3%</p>   |
| <p>IF CAT_Name EQUALS PLANT<br/>AND Destination EQUALS COUNTRY<br/>THEN<br/>NODE : 10<br/>N : 22<br/>MOTORCYC: 0.0%<br/>EMS : 100.0%<br/>PARCEL : 0.0%<br/>TNT : 0.0%</p>   | <p>IF Weight_Type EQUALS SMALL<br/>AND CAT_Name IS ONE OF: FISH FOOD CHEMICAL<br/>AND Destination EQUALS COUNTRY<br/>THEN<br/>NODE : 17<br/>N : 97<br/>MOTORCYC: 0.0%<br/>EMS : 100.0%<br/>PARCEL : 0.0%<br/>TNT : 0.0%</p>            |
| <p>IF Weight_Type EQUALS SMALL<br/>AND CAT_Name IS ONE OF: FISH FOOD CHEMICAL<br/>AND Destination EQUALS COUNTRY<br/>THEN<br/>NODE : 17<br/>N : 97<br/>MOTORCYC: 0.0%<br/>EMS : 100.0%<br/>PARCEL : 0.0%<br/>TNT : 0.0%</p> | <p>IF Weight_Type IS ONE OF: LARGE MEDIUM<br/>AND CAT_Name IS ONE OF: FISH FOOD CHEMICAL<br/>AND Destination EQUALS COUNTRY<br/>THEN<br/>NODE : 18<br/>N : 19<br/>MOTORCYC: 0.0%<br/>EMS : 52.6%<br/>PARCEL : 42.1%<br/>TNT : 5.3%</p> |
| <p>IF Weight_Type IS ONE OF: LARGE MEDIUM<br/>AND CAT_Name EQUALS AQUARIUM<br/>AND Destination EQUALS COUNTRY<br/>THEN<br/>NODE : 19<br/>N : 30<br/>MOTORCYC: 0.0%<br/>EMS : 0.0%<br/>PARCEL : 0.0%<br/>TNT : 100.0%</p>    | <p>IF Weight_Type EQUALS SMALL<br/>AND CAT_Name EQUALS AQUARIUM<br/>AND Destination EQUALS COUNTRY<br/>THEN<br/>NODE : 20<br/>N : 16<br/>MOTORCYC: 0.0%<br/>EMS : 0.0%<br/>PARCEL : 100.0%<br/>TNT : 0.0%</p>                          |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 6.3.2 วิเคราะห์ผลลัพธ์พร้อมทั้งนำเสนอรูปแบบผลลัพธ์ที่ได้

จากการรันโปรแกรมด้วยโมเดลแผนภาพต้นไม้พบว่าโปรแกรมได้ทำการสร้างแบบจำลองและเสนอผลลัพธ์ออกมา 8 เงื่อนไขซึ่งดังนี้

1. ลูกค้าที่อยู่กรุงเทพฯส่วนมากหรือเกือบทั้งหมดมีการเลือกส่งสินค้าด้วยรถมอเตอร์ไซค์
2. ลูกค้าที่อยู่ต่างจังหวัด และซื้อสินค้าประเภทอุปกรณ์ตู้ ส่วนมากมีการเลือกส่งสินค้าแบบ EMS
3. ลูกค้าที่อยู่ต่างจังหวัด และซื้อสินค้าประเภทต้นไม้ น้ำ ทั้งหมดมีการเลือกส่งสินค้าแบบ EMS
4. ลูกค้าที่อยู่ต่างจังหวัด และซื้อสินค้าประเภทอุปกรณ์ตู้ มีการเลือกส่งสินค้าแบบ EMS
5. ลูกค้าที่อยู่ต่างจังหวัด และซื้อสินค้าประเภทอาหารปลาด้วยบรรจุภัณฑ์ น้ำหนักน้อย ทั้งหมด มีการเลือกส่งสินค้าแบบ EMS
6. ลูกค้าที่อยู่ต่างจังหวัด และซื้อสินค้าประเภทอาหารปลาด้วยบรรจุภัณฑ์ น้ำหนักปานกลางและน้ำหนักมากมีแนวโน้มในการเลือกส่งสินค้าแบบ EMS และพัสดุไปรษณีย์
7. ลูกค้าที่อยู่ต่างจังหวัด และซื้อสินค้าประเภทตู้ปลา น้ำหนักปานกลางและน้ำหนักมาก ทั้งหมด มีการเลือกส่งสินค้าแบบ TNT
8. ลูกค้าที่อยู่ต่างจังหวัด และซื้อสินค้าประเภทตู้ปลา น้ำหนักเบา มีการเลือกส่งสินค้าแบบพัสดุไปรษณีย์

ซึ่งจากเงื่อนไขที่ได้ทั้ง 8 เงื่อนไขนี้สามารถนำไปช่วยสนับสนุนการวางแผนทางด้าน CRM โดยพบว่าในกรณีที่ลูกค้าในกรุงเทพฯมักจะสั่งซื้อสินค้าที่มีน้ำหนักเบาและเลือกส่งสินค้าด้วยรถมอเตอร์ไซค์ สำหรับลูกค้าต่างจังหวัดที่สั่งซื้อสินค้าที่มีน้ำหนักเบาจะเลือกส่งสินค้าแบบพัสดุไปรษณีย์ และสินค้าน้ำหนักปานกลางถึงน้ำหนักมากมักจะเลือกส่งสินค้าแบบ TNT จากข้อมูลที่ได้สามารถนำไปช่วยวิเคราะห์และพัฒนาระบบปฏิบัติการให้สามารถเลือกหรือแนะนำประเภทการส่งสินค้าให้กับลูกค้าได้ หรืออาจช่วยในการคาดการณ์การเลือกส่งสินค้าเช่น ถ้าเป็นลูกค้าต่างจังหวัดมักจะเลือกขนส่งแบบ EMS เสมอ เป็นต้น และหากผู้บริหารต้องการสร้างความพึงพอใจให้กับลูกค้าและสร้างพันธมิตรทางการค้ากับบริษัทขนส่ง ก็สามารถนำผลลัพธ์ที่ได้นี้ไปช่วยสนับสนุนการตัดสินใจในการเจรจากับบริษัทอื่นๆได้

#### 6.4 วิเคราะห์กลุ่มลูกค้าที่สั่งซื้อสินค้า

ในหัวข้อนี้จะวิเคราะห์กลุ่มลูกค้าที่สั่งซื้อสินค้าโดยการนำปัจจัยต่างๆที่เกี่ยวข้องกับการสั่งซื้อสินค้าของลูกค้ามาวิเคราะห์พฤติกรรมและแบ่งกลุ่ม ซึ่งผลลัพธ์ที่ได้จากการแบ่งกลุ่มนั้นจะสามารถนำไปพิจารณาการจัดโปรโมชั่นให้ตรงกับกลุ่มเป้าหมายหรือสามารถนำไปทำ Direct mail ถึงลูกค้าเฉพาะกลุ่มได้ โดยได้นำเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means มาช่วยในการแบ่งกลุ่มข้อมูล

##### 6.4.1 ขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ซอฟต์แวร์ SAS

1. นำข้อมูลจากฐานข้อมูลที่ได้ออกแบบไว้เข้าสู่โปรแกรม โดยเลือกเมนู File > Import Data > เลือกประเภทของข้อมูลจากแหล่งข้อมูล ซึ่งโปรแกรมสามารถรองรับประเภทของแหล่งข้อมูลได้หลากหลายประเภท
2. เข้าสู่เมนูการวิเคราะห์ข้อมูล เลือกเมนู Solution > Analyze > Enterprise Miner
3. การสร้างโมเดล โดยเลือก ไอคอน ข้อมูลนำเข้า และไอคอน Clustering มาวางไว้ที่หน้าจอ จากนั้นทำการสร้างเส้นเชื่อมจาก ข้อมูลนำเข้า ไปสู่ Clustering



รูปที่ 6.16 การสร้างโมเดลของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means

4. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับข้อมูลนำเข้าในแท็บ Data จะทำการเลือกข้อมูลที่ได้เตรียมไว้แล้ว และในแท็บ Variables แต่ละเทคนิคจะกำหนดค่าไม่เหมือนกัน โดยเทคนิคการแบ่งกลุ่มโดยใช้อัลกอริทึม K-Means นี้จะสนใจข้อมูลทั้งหมดที่นำมาเป็นอินพุต จึงทำการกำหนดข้อมูลทั้งหมดเป็นอินพุต และในการแบ่งกลุ่มค่าความเป็นไปได้ของข้อมูลที่เป็นอินพุตนั้นจะมีผลต่อการแบ่งกลุ่มเป็นอย่างมาก เช่น ข้อมูลเพศจะมีค่าความเป็นไปได้เพียง 3 ค่า คือ Male, Female และ Unknown แต่สำหรับข้อมูลอายุจะมีค่าความเป็นไปได้จำนวนมากเนื่องจากเป็นค่าตัวเลข ดังนั้น จะต้องมีการทำนอร์มอลไลเซชันกับข้อมูลที่มีค่าความเป็นไปได้มากโดยเลือก Standardization > Range ซึ่งจะทำการแปลงข้อมูลที่มีค่าต่อเนื่องให้เป็นข้อมูลแบบเป็นช่วงๆ

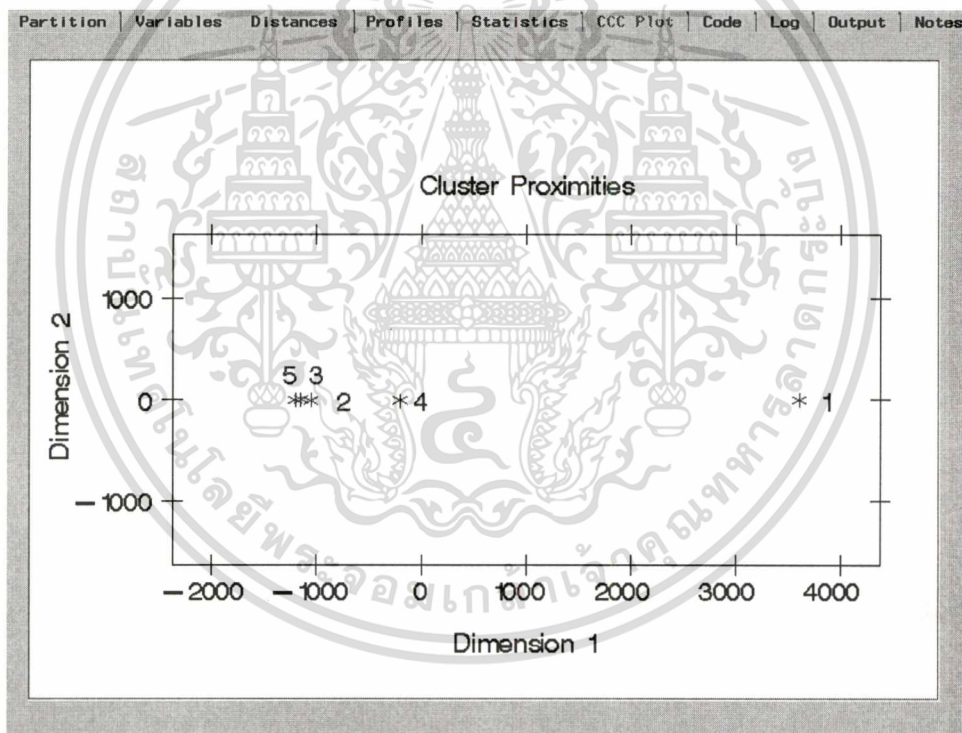
5. กำหนดค่าพารามิเตอร์ต่างๆ สำหรับโมเดล Clustering โดยทั่วไปเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means นี้จำเป็นต้องกำหนดกลุ่มของข้อมูลเริ่มต้นก่อน ในแท็บ Clusters > Number of Clusters > User specify แต่สำหรับ โปรแกรม Enterprise Miner นี้มีฟังก์ชันพิเศษคือสามารถเลือกให้โปรแกรมทำการแบ่งกลุ่มอัตโนมัติได้ ซึ่งหากผลลัพธ์ที่ได้ไม่เป็นที่น่าพอใจก็สามารถกำหนดกลุ่มตัวเอง

| Data  | Variables | Clusters                                | Seeds | Missing Values                                      | Output | Notes |
|---|-----------|---|-------|---|--------|-------|
| Segment Identifier:                           |           |   |       |   |        |       |
| Variable name:                                |           | <input type="text" value="_SEGMNT_"/>   |       |   |        |       |
| Variable label:                               |           | <input type="text" value="Cluster ID"/> |       |   |        |       |
| Role:   |           | <input type="text" value="group"/>      |       |   |        |       |
| Number of Clusters:                           |           |   |       |   |        |       |
| <input checked="" type="radio"/> User specify |           | <input type="text" value="5"/>          |       | <input type="text" value="Selection Criterion..."/> |        |       |
| <input type="radio"/> Automatic               |           |   |       |   |        |       |

รูปที่ 6.17 การกำหนดพารามิเตอร์ของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means

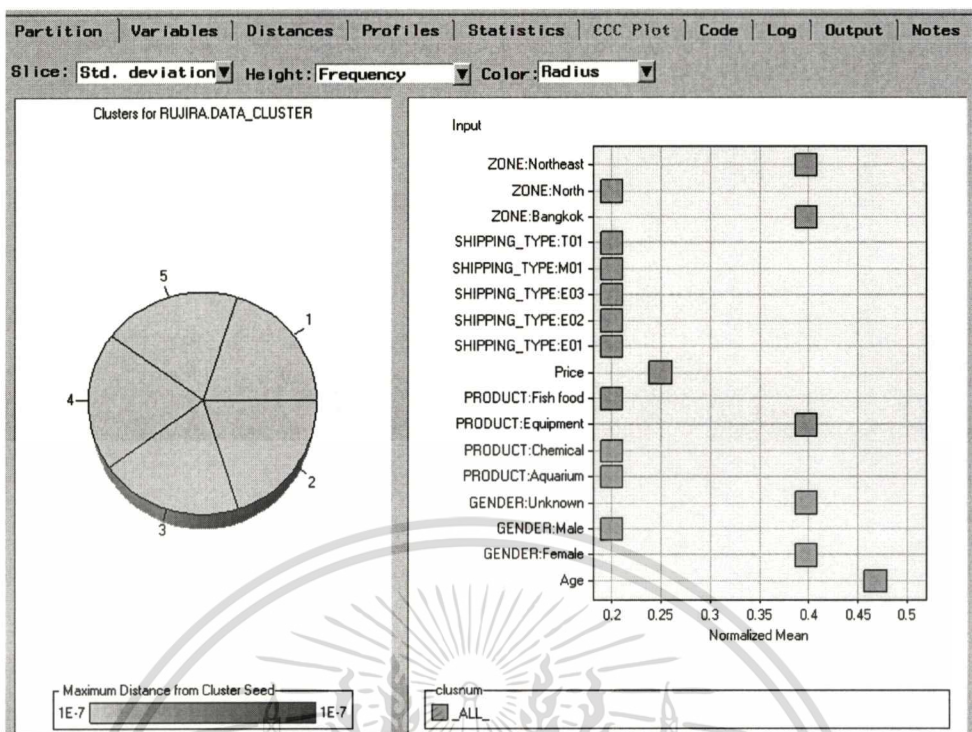
6. ทำการรันโปรแกรม

7. ผลลัพธ์จากขั้นตอนที่ 6

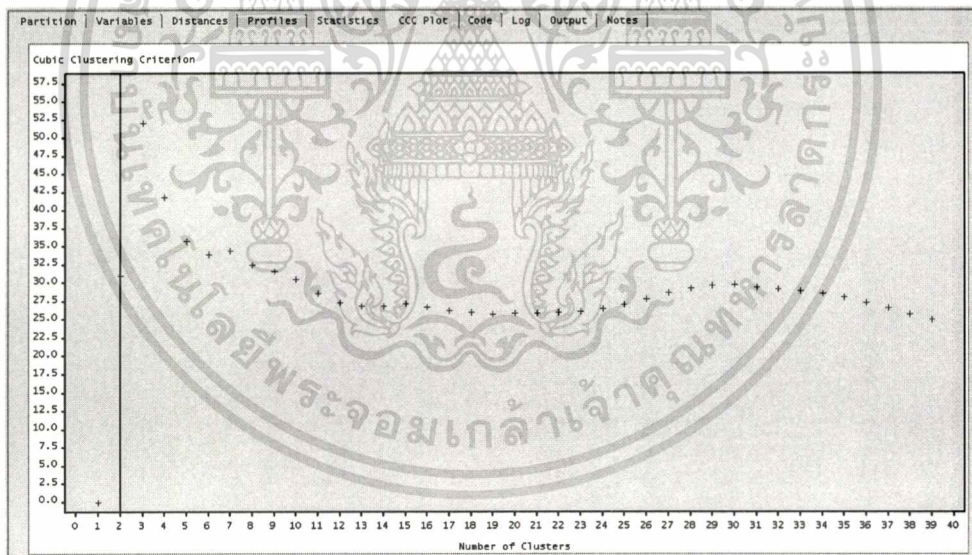


รูปที่ 6.18 ผลลัพธ์ในรูปแบบของกราฟของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means เพื่อดูการกระจายของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 6.19 ผลลัพธ์ในรูปแบบแผนภาพวงกลมของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means



รูปที่ 6.20 แผนภาพค่า RMS-SD ในแต่ละกลุ่มของเทคนิคการแบ่งกลุ่มด้วยอัลกอริทึม K-Means

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 6.4.2 วิเคราะห์ผลลัพธ์พร้อมทั้งนำเสนอรูปแบบผลลัพธ์ที่ได้

ในการวิเคราะห์การแบ่งกลุ่มของข้อมูลนั้นสิ่งที่จำเป็นในการพิจารณาเลือกกว่าแบ่งกลุ่มเท่าไรจึงเหมาะสมนั้นขึ้นอยู่กับค่า  $E^2$  คือผลรวมของค่าความผิดพลาดของข้อมูลในแต่ละกลุ่ม โดยทั่วไปจะพิจารณาค่าอย่างน้อยยิ่งดี ซึ่งโปรแกรมนี้จะแสดงด้วยค่า RMS-SD ยิ่งมีการแบ่งกลุ่มมากเท่าไรห้ค่า RMS-SD ยิ่งมีค่าน้อยลงแต่ไม่ได้หมายความว่าต้องมีการแบ่งกลุ่มมากๆ เนื่องจากการแบ่งกลุ่มมากๆ นั้นอาจไม่มีผลต่อธุรกิจเลยดังนั้นหลักการเลือกจะเลือกจำนวนกลุ่มที่มีค่าน้อยและไม่เปลี่ยนแปลงมาก ซึ่งจากรูปที่ 6.21 พบว่ากราฟจะลดลงเรื่อยๆ จนถึงระดับหนึ่งก็จะเริ่มคงที่

จากผลลัพธ์ในการรัน โปรแกรมพบว่าสามารถแบ่งกลุ่มข้อมูลของลูกค้าได้ทั้งหมด 5 กลุ่ม ดังนี้

1. กลุ่มลูกค้าที่มีการซื้อสินค้ามากกว่า 3,000 บาทขึ้นไป ซึ่งเป็นลูกค้าที่อยู่ต่างจังหวัด
2. กลุ่มลูกค้าเพศชายที่มีการซื้อสินค้าตั้งแต่ 1,000 - 3,000 บาทขึ้นไป ซึ่งเป็นลูกค้าที่อยู่ต่างจังหวัด
3. กลุ่มลูกค้าเพศหญิงที่มีการซื้อสินค้าไม่เกิน 1,000 บาทและเป็นลูกค้าที่อยู่ต่างจังหวัด
4. กลุ่มลูกค้าเพศชายที่มีการซื้อสินค้าไม่เกิน 550 บาทและเป็นลูกค้าที่อยู่ในกรุงเทพฯ และอายุระหว่าง 17 – 25 ปี
5. กลุ่มลูกค้าเพศชายที่มีอายุมากกว่า 25 ปี ซึ่งเป็นลูกค้าที่อยู่ต่างจังหวัดและมีการซื้อสินค้าไม่เกิน 1,000 บาท

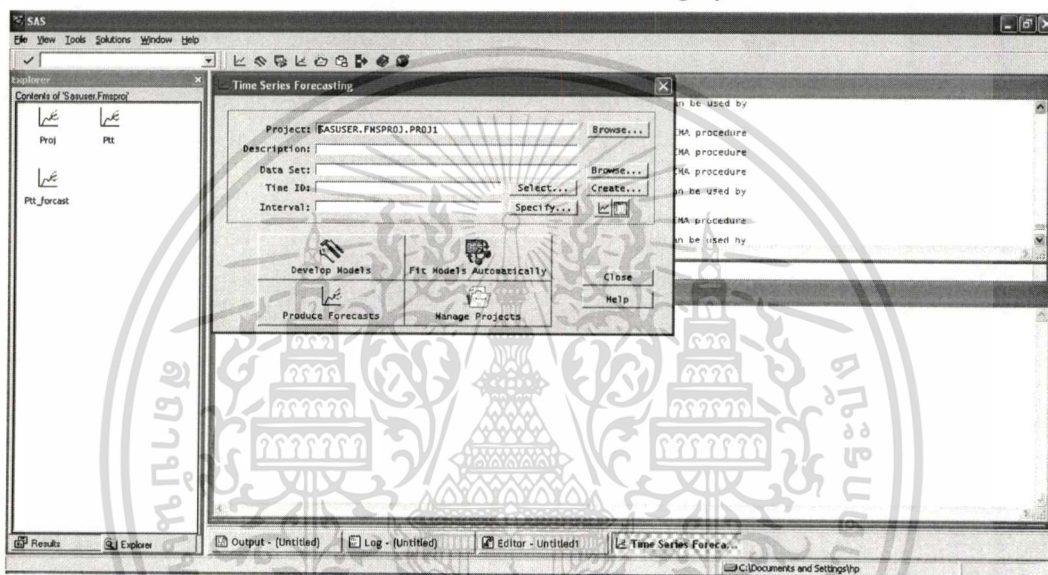
ลักษณะกลุ่มของลูกค้าที่แบ่งได้นั้นสามารถนำไปช่วยสนับสนุนการตัดสินใจเรื่องการตลาดและ CRM ได้เช่นมีการเสนอสินค้าใหม่หรือโปรโมชันของสินค้าที่มีราคาไม่เกิน 1,000 บาทให้กับลูกค้าที่อยู่ในกลุ่ม 3, 4 และ 5 และยังพบว่าลูกค้าส่วนใหญ่ที่อยู่ในกรุงเทพฯ มักจะสั่งซื้อสินค้าราคาไม่แพงหรือไม่เกิน 550 บาท ซึ่งอาจเนื่องมาจากหากซื้อสินค้าราคาไม่แพงหรือจำนวนน้อยอาจจะคุ้มค่าที่จะออกไปซื้อสินค้านอกบ้าน และหากต้องการซื้อสินค้าที่มีราคาแพงหรือปริมาณมากจะไม่เลือกซื้อสินค้าทางอินเทอร์เน็ตเนื่องจากอยู่ใกล้แหล่งซื้อสินค้าดังนั้นอาจมีการเสนอโปรโมชันของสินค้าที่แตกต่างจากสินค้าที่มีอยู่ในท้องตลาดหรือมีบริการจัดส่งสินค้าฟรีสำหรับลูกค้าที่สั่งซื้อสินค้านี้ราคาสูงกว่า 1,000 บาทและมีปริมาณมากเพื่อเป็นการกระตุ้นยอดขายและเพิ่มความเชื่อมั่นในการสั่งซื้อสินค้าผ่านอินเทอร์เน็ต เป็นต้น

## 6.5 ทำนายยอดขายสินค้าในวันถัดไป

ในหัวข้อนี้จะทำนายยอดขายสินค้าในแต่ละวัน โดยการนำข้อมูลยอดขายสินค้าย้อนหลังมาเป็นปัจจัยในการทำนาย ซึ่งผลลัพธ์ที่ได้จากการทำนายนั้นสามารถนำไปพิจารณาการวางแผนกลยุทธ์ในการขายสินค้า หรือวางเป้าหมายยอดขายในแต่ละวัน เดือนหรือปียอดขายประจำปีได้ โดยได้นำเทคนิคพยากรณ์อนุกรมเวลา (Time Series Forecasting) มาช่วยในการทำนายยอดขาย

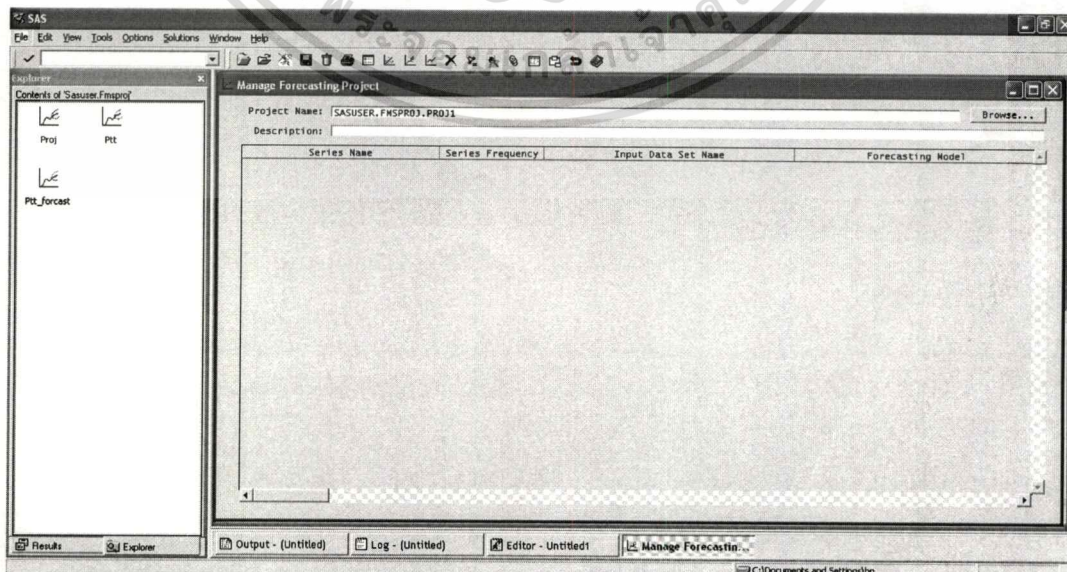
### 6.5.1 ขั้นตอนการทำนายโดยใช้เทคนิค Time Series Forecasting ของซอฟต์แวร์ SAS

#### 1. เข้า โปรแกรม SAS เลือก Time Series Forecasting System



รูปที่ 6.21 การสร้างโมเดลของเทคนิคพยากรณ์อนุกรมเวลา

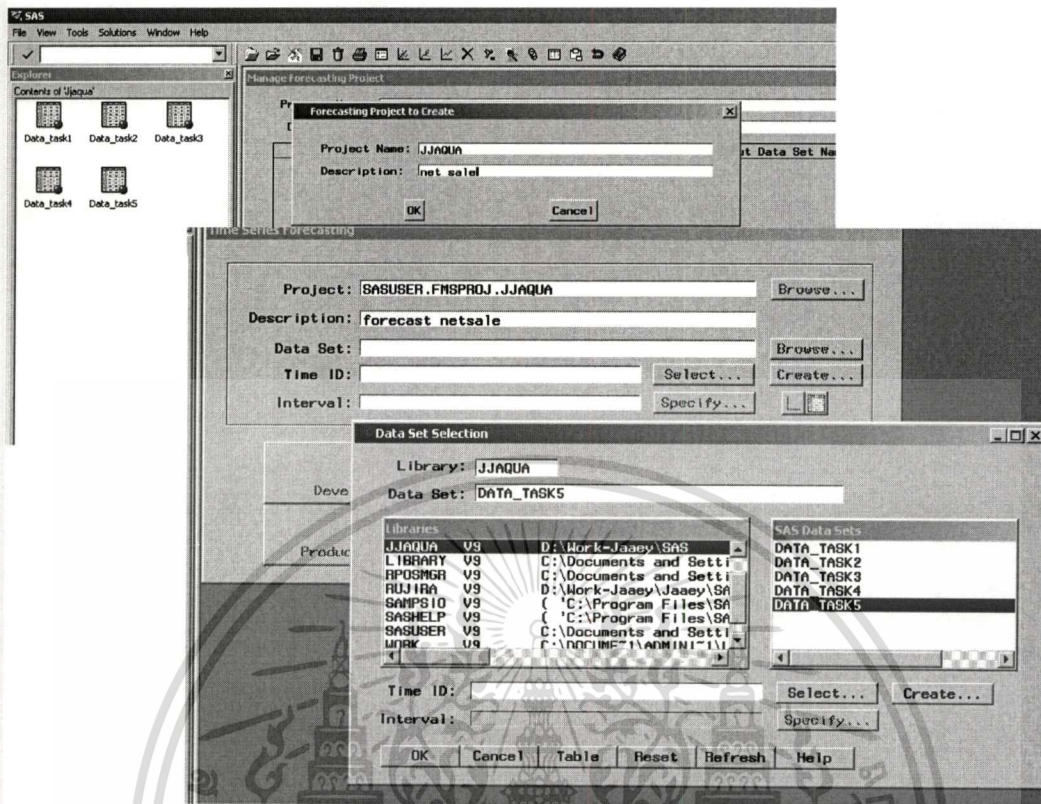
#### 2. เลือก Manage Forecasting Project



รูปที่ 6.22 การสร้างโปรเจกต์ใหม่สำหรับเทคนิคพยากรณ์อนุกรมเวลา

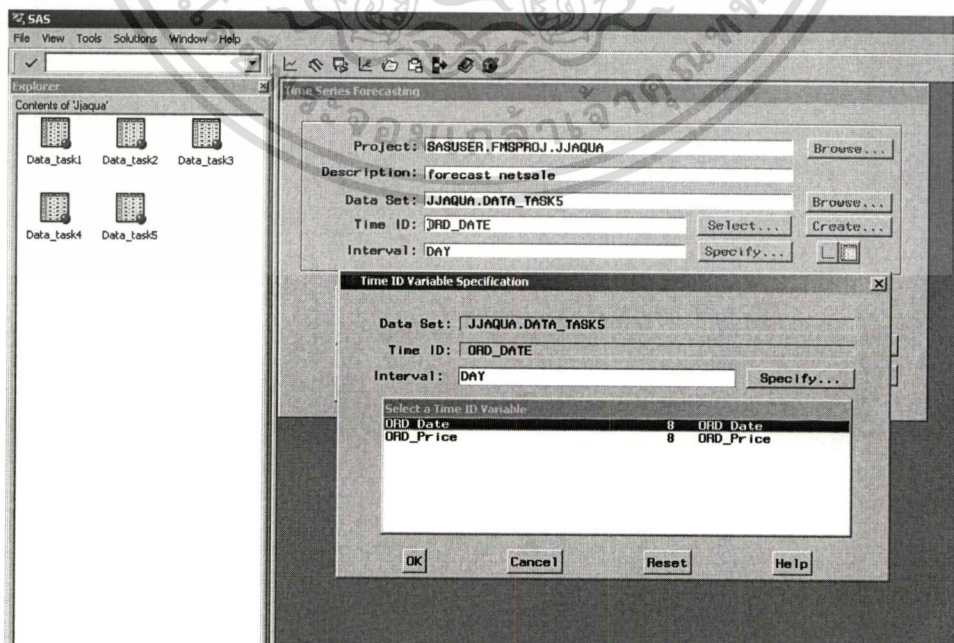
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษเท่านั้น เมื่อผู้ผู้เห็นใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3. สร้าง Project โดยเริ่มจากตั้งชื่อ Project เลือกอินพุทของข้อมูล



รูปที่ 6.23 การเลือกข้อมูลอินพุทเข้าสู่โมเดลพยากรณ์อนุกรมเวลา

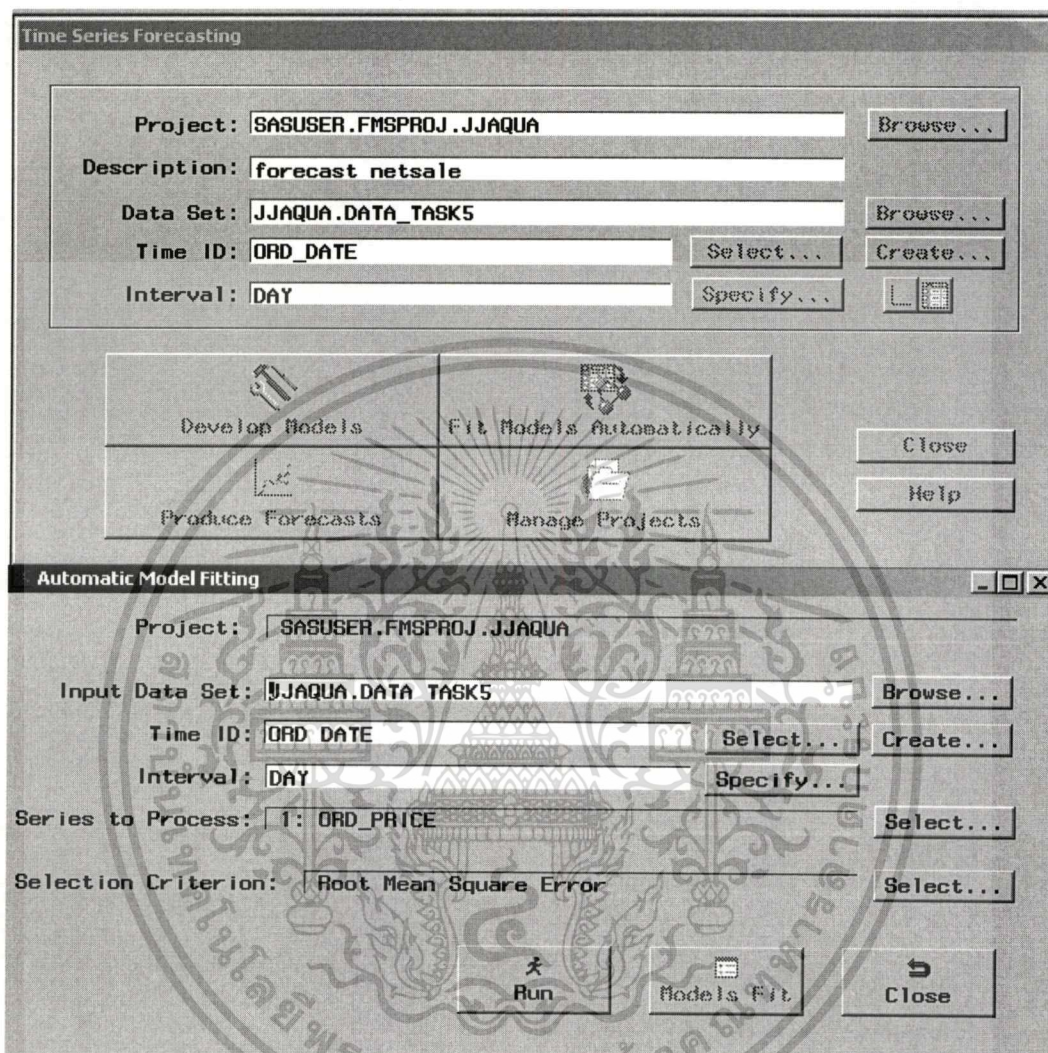
### 4. กำหนดพารามิเตอร์ในการจำลองโมเดล โดยเลือก Time ID เป็น ORD\_Date และ เลือก Interval เป็น Day



รูปที่ 6.24 การกำหนดพารามิเตอร์สำหรับเทคนิคพยากรณ์อนุกรมเวลา

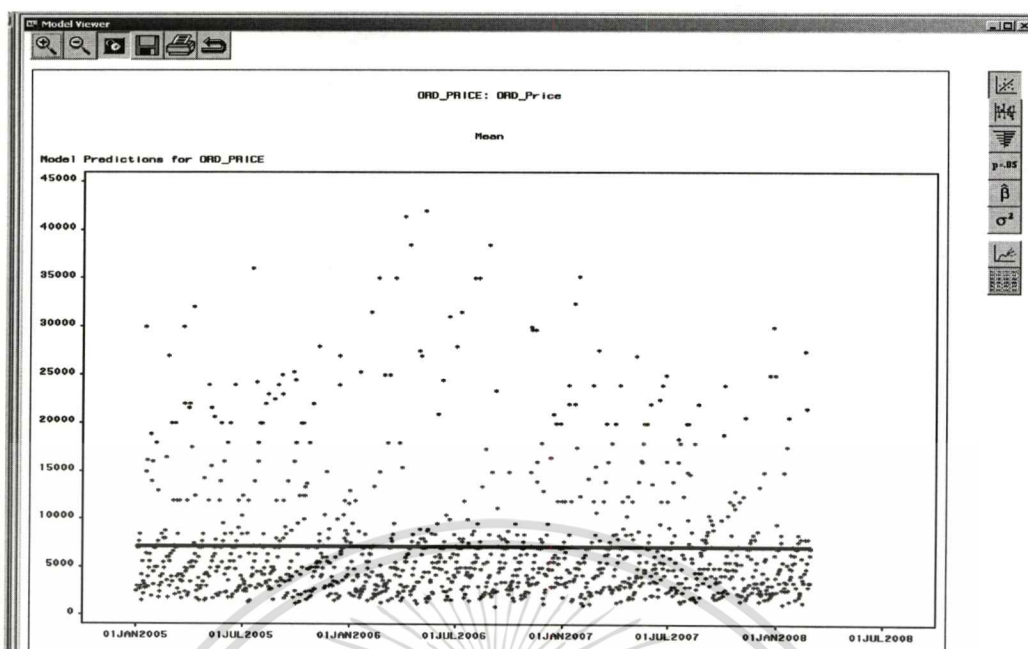
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. เลือกโมเดลสำหรับการทำนายโดยให้โปรแกรมทำการเลือกโมเดลที่ดีที่สุด โดยเลือกที่ Automatic Model Fitting



รูปที่ 6.25 การจำลอง โมเดลเพื่อเลือกโมเดลที่ดีที่สุด

6. โปรแกรมเลือกโมเดลที่ดีที่สุดให้อัตโนมัติ โดยเลือกโมเดล  $> \text{Mean}$
- ข้อมูลที่นำมาทำนายนั้นสามารถแสดงกราฟตาม โมเดล Mean ได้ดังรูป



รูปที่ 6.26 การกระจายข้อมูลของโมเดล Mean ตามข้อมูลอินพุต

- ขั้นตอนการทำนาย โดยเลือก Produce Forecast เลือกทำนายยอดขายสินค้าในวันที่ 1 มีนาคม 2008 เนื่องจากข้อมูลที่นำมาใช้นั้นมีข้อมูลล่าสุดคือวันที่ 29 กุมภาพันธ์ 2008

รูปที่ 6.27 การเลือกทำนายค่าตามวันที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 8. ผลลัพธ์ที่ได้จากการทำนาย

| VIEWTABLE: WORK.FORECAST |           |      |       |     |           |
|--------------------------|-----------|------|-------|-----|-----------|
|                          | ORD_Date  | YEAR | MONTH | DAY | ORD_PRICE |
| 1138                     | 12FEB2008 | 2008 | 2     | 12  | 4200      |
| 1139                     | 13FEB2008 | 2008 | 2     | 13  | 8000      |
| 1140                     | 14FEB2008 | 2008 | 2     | 14  | 6000      |
| 1141                     | 15FEB2008 | 2008 | 2     | 15  | 1400      |
| 1142                     | 16FEB2008 | 2008 | 2     | 16  | 3780      |
| 1143                     | 17FEB2008 | 2008 | 2     | 17  | 3500      |
| 1144                     | 18FEB2008 | 2008 | 2     | 18  | 27500     |
| 1145                     | 19FEB2008 | 2008 | 2     | 19  | 4500      |
| 1146                     | 20FEB2008 | 2008 | 2     | 20  | 8000      |
| 1147                     | 21FEB2008 | 2008 | 2     | 21  | 21600     |
| 1148                     | 22FEB2008 | 2008 | 2     | 22  | 4900      |
| 1149                     | 23FEB2008 | 2008 | 2     | 23  | 5600      |
| 1150                     | 24FEB2008 | 2008 | 2     | 24  | 3000      |
| 1151                     | 25FEB2008 | 2008 | 2     | 25  | 8000      |
| 1152                     | 26FEB2008 | 2008 | 2     | 26  | 3500      |
| 1153                     | 27FEB2008 | 2008 | 2     | 27  | 3500      |
| 1154                     | 28FEB2008 | 2008 | 2     | 28  | 6300      |
| 1155                     | 29FEB2008 | 2008 | 2     | 29  | 7000      |
| 1156                     | 01MAR2008 | 2008 | 3     | 1   | 7137      |

รูปที่ 6.28 ผลลัพธ์จากการทำนายค่า

## 6.5.2 วิเคราะห์ผลลัพธ์พร้อมทั้งนำเสนอรูปแบบผลลัพธ์ที่ได้

ในการทำนายยอดขายนั้นสิ่งที่จำเป็นในการทำนายได้แก่ข้อมูลอินพุตย้อนหลังที่นำมาใช้ ประกอบกับข้อมูลวันที่ที่ต้องมีความถูกต้องและต่อเนื่อง ซึ่งข้อมูลวันที่ที่นำมาใช้นั้นจะต้องเป็นข้อมูลที่ต่อเนื่องและเรียงลำดับเรียบร้อยแล้ว

จากผลลัพธ์ในการรันโปรแกรมพบว่าการทำนายยอดขายในวันที่ 1 มีนาคม 51 มีค่าเท่ากับ 7,137 บาท เมื่อนำไปเปรียบเทียบกับยอดขายจริงพบว่า ยอดขายมีค่าเท่ากับ 9,580 บาท คิดเป็นเปอร์เซ็นต์ความผิดพลาดอยู่ที่ 25.51%

จากโมเดลการทำนายยอดขายได้ทำนายยอดขายในแต่ละวันเพิ่มเติมเพื่อตรวจสอบดูค่าความผิดพลาด โดยเริ่มตั้งแต่วันที่ 1 มีนาคม 51 ถึงวันที่ 15 มีนาคม 51 และนำผลลัพธ์ที่ได้จากการทำนายมาเปรียบเทียบกับข้อมูลจริงพบว่าเปอร์เซ็นต์ความผิดพลาดเฉลี่ยอยู่ที่ 18.57% ทั้งนี้ค่าความผิดพลาดที่เกิดขึ้นเนื่องมาจากข้อมูลย้อนหลังที่ใช้ในการทำนายนั้นเป็นข้อมูลยอดขายแต่ละวันเพียงอย่างเดียว ไม่มีการนำปัจจัยอื่นๆ มาช่วยในการวิเคราะห์ด้วย ซึ่งการทำนายยอดขายในแต่ละวันนั้นมีปัจจัยที่ไม่แน่นอนและยังไม่มีเก็บข้อมูลในฐานะข้อมูล ดังนั้นหากองค์กรต้องการนำโมเดลการทำนายยอดขายนี้ไปใช้งานจริง จำเป็นจะต้องเผื่อเปอร์เซ็นต์ความผิดพลาดเข้าไป หรือมีการเก็บข้อมูลอื่นๆ ที่เป็นปัจจัยในการขายสินค้าในแต่ละวันเพิ่มมากขึ้นเช่น อัตราแลกเปลี่ยนเงินตราต่างประเทศ สภาวะราคาน้ำมัน เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 7

# สรุปผลการศึกษา และข้อเสนอแนะ

### 7.1 สรุปผลการศึกษา

โครงการศึกษานี้ได้ดำเนินการศึกษาการพัฒนาระบบวิเคราะห์พฤติกรรมผู้บริโภคสำหรับธุรกิจอีคอมเมิร์ซหรือพาณิชย์อิเล็กทรอนิกส์ด้วยการออกแบบระบบฐานข้อมูลใหม่ให้สอดคล้องกับแต่ละปัญหา สามารถนำข้อมูลไปใช้วิเคราะห์พฤติกรรมผู้บริโภคได้อย่างเหมาะสมและไม่ส่งผลกระทบต่อระบบปฏิบัติการหลักขององค์กร โจทะนั้นได้มีการนำเทคนิคดาต้าไมน์นิ่งแบบต่างๆมาช่วยในการวิเคราะห์พฤติกรรมผู้บริโภคในแต่ละเรื่องแต่ละปัญหาเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด โดยในแต่ละปัญหานั้นได้มีการคัดเลือกข้อมูล เตรียมข้อมูลและใช้เทคนิคดาต้าไมน์นิ่งที่แตกต่างกันมาทำการวิเคราะห์ข้อมูล

การพัฒนาระบบจะใช้ SQL Server เป็นระบบจัดการฐานข้อมูล และซอฟต์แวร์ Enterprise Miner ของบริษัท SAS ในการวิเคราะห์พฤติกรรมผู้บริโภคด้วยเทคนิคดาต้าไมน์นิ่ง ซึ่งการทำงานหลักของระบบจัดการฐานข้อมูลนั้นจะทำการโอนย้ายข้อมูลจากระบบปฏิบัติการหลักสู่ฐานข้อมูลใหม่ แปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับการวิเคราะห์ข้อมูลและมีการกำหนดช่วงเวลาให้ในการโอนย้ายข้อมูลอัตโนมัติ และมีการตรวจสอบความถูกต้องของข้อมูลก่อนที่จะนำไปใช้งาน จากนั้นมีการ Query ข้อมูลเพื่อให้ได้ข้อมูลตามต้องการและสามารถนำไปใช้กับซอฟต์แวร์ดาต้าไมน์นิ่งได้ถูกต้องตามเงื่อนไขและอัลกอริทึมของเทคนิคดาต้าไมน์นิ่ง

ผลลัพธ์จากการออกแบบระบบและใช้เทคนิคดาต้าไมน์นิ่งช่วยวิเคราะห์ข้อมูลนั้น จะได้สารสนเทศที่ดี มีประโยชน์ มีความถูกต้องและน่าเชื่อถือ เป็นข้อมูลช่วยสนับสนุนการตัดสินใจของผู้บริหารในการวางกลยุทธ์ในด้านต่างๆ ได้ไม่ว่าจะเป็นด้านการขาย การตลาด การบริการ CRM และการบริหารคลังสินค้า เป็นต้น และผลลัพธ์ที่ได้นี้อยู่ในรูปแบบที่เข้าใจง่าย ไม่ซับซ้อนและสามารถดูข้อมูลได้หลายมิติซึ่งทำให้การวิเคราะห์ข้อมูลมีประสิทธิภาพมากยิ่งขึ้น

### 7.2 ข้อเสนอแนะ

ในการออกแบบและพัฒนาระบบงานนี้เป็นแนวทางในการวิเคราะห์พฤติกรรมผู้บริโภคสำหรับธุรกิจอีคอมเมิร์ซหรือพาณิชย์อิเล็กทรอนิกส์อื่นๆ ใช้ได้กับการทำงานจริง และผลลัพธ์ที่ได้ นั้นจะยังมีความน่าเชื่อถือมากขึ้นเมื่อองค์กรมีการจัดเก็บข้อมูลต่างๆที่เกี่ยวข้องได้อย่างถูกต้องและครอบคลุม ดังนั้นการพัฒนาระบบให้มีประสิทธิภาพเพิ่มมากขึ้นเพื่อให้ตรงกับความต้องการในอนาคตที่อาจมีการเปลี่ยนแปลงได้โดยมีข้อเสนอแนะเป็นดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. การที่จะทำให้ระบบนี้สามารถวิเคราะห์พฤติกรรมผู้บริโภคได้ความถูกต้อง หลากหลาย และน่าเชื่อถือมากขึ้น ต้องมีการจัดเก็บข้อมูลที่ดีและครอบคลุม ควรจะมีการจัดเก็บข้อมูลที่เกี่ยวข้องกับลูกค้าให้มากขึ้นเช่น อาชีพ รายได้ บัตรเครดิต เป็นต้น
2. ให้มีการนำผลลัพธ์ที่ได้ไปใช้งานจริงเพื่อเป็นการประเมินความถูกต้องที่ได้และนำมาปรับปรุงกระบวนการและพัฒนาระบบต่อไป
3. นำซอฟต์แวร์ Enterprise Miner มาช่วยในการวิเคราะห์พฤติกรรมผู้บริโภคด้วยเทคนิค คาด้าไมน์นิ่งแบบอื่นๆเพิ่มเติม และในกรณีที่มีการปรับเปลี่ยนระบบจัดการฐานข้อมูล สามารถนำซอฟต์แวร์ Enterprise Miner มาช่วยในการจัดการข้อมูลแทนได้โดยไม่ต้องใช้ SQL Server



## บรรณานุกรม

กิตติ ภัคดีวัฒนะกุล. 2550. ระบบสนับสนุนการตัดสินใจและระบบผู้เชี่ยวชาญ. พิมพ์ครั้งที่ 2.  
กรุงเทพฯ : เททีพีคอมพิวเตอร์แอนด์คอนซัลท์.

กิตติพงษ์ กลมกล่อม. 2546. การออกแบบและพัฒนาคลังข้อมูล. กรุงเทพฯ : เททีพีคอมพิวเตอร์แอนด์คอนซัลท์.

กฤษณะ ไวยมัย. 2549. คลังข้อมูลและการทำเหมืองข้อมูล. เอกสารคำสอน วิชา 214554. ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

สมพร จิวรสกุล. 2545. คู่มือการติดตั้งและใช้งาน Microsoft SQL Server 2000. พิมพ์ครั้งที่ 1.  
กรุงเทพฯ : เอช เอ็น กรู๊ป.

Han, Jiawei and Kamber, Micheline. 2001. **Data Mining: Concepts and Techniques**. San Francisco, CA: Morgan Kaufmann Publishers.

Rob, Peter and Coronel, Carlos. 2000. **Database Systems: Design, Implementation & Management**. 7<sup>th</sup> Edition. Boston, MA: Thomson Course Technology.

## ภาคผนวก

### พจนานุกรมข้อมูลของระบบฐานข้อมูลเดิม

ตารางที่ 1 : พจนานุกรมของตาราง T\_Order

| ชื่อตาราง              | T_Order                        |               |      |                 |
|------------------------|--------------------------------|---------------|------|-----------------|
| ชื่อแอตทริบิวต์        | คำอธิบาย                       | ชนิดข้อมูล    | คีย์ | ตารางที่อ้างอิง |
| ORD_ID                 | รหัสการสั่งซื้อสินค้า          | INT(4)        | PK   |                 |
| ORD_USR_ID             | รหัสลูกค้าที่สั่งซื้อสินค้า    | INT(4)        | FK   | M_User          |
| ORD_BillingAddressID   | รหัสที่อยู่สำหรับออกใบแจ้งหนี้ | INT(4)        | FK   | M_Address       |
| ORD_ShippingAddressID  | รหัสที่อยู่สำหรับส่งสินค้า     | INT(4)        | FK   | M_Address       |
| ORD_SPT_ID             | รหัสประเภทการส่งสินค้า         | INT(4)        | FK   | T_ShippingType  |
| ORD_FirstName          | ชื่อ                           | NVARCHAR(50)  |      |                 |
| ORD_LastName           | นามสกุล                        | NVARCHAR(50)  |      |                 |
| ORD_Telephone          | หมายเลขโทรศัพท์                | NVARCHAR(20)  |      |                 |
| ORD_Mobile             | หมายเลขมือถือ                  | NVARCHAR(20)  |      |                 |
| ORD_Email              | อีเมล                          | NVARCHAR(50)  |      |                 |
| ORD_No                 | เลขที่ใบสั่งซื้อ               | NVARCHAR(50)  |      |                 |
| ORD_Weight             | น้ำหนัก                        | DECIMAL(18,2) |      |                 |
| ORD_ProductPrice       | ราคาสินค้าเก็บจากลูกค้า        | MONEY(19,4)   |      |                 |
| ORD_ShippingPrice      | ราคาขนส่งเก็บจากลูกค้า         | MONEY(19,4)   |      |                 |
| ORD_ProductCost        | ราคาค่าต้นทุน                  | MONEY(19,4)   |      |                 |
| ORD_ShippingCost       | ราคาค่าขนส่ง                   | MONEY(19,4)   |      |                 |
| ORD_PackageCost        | ราคาค่าบรรจุหีบห่อ             | MONEY(19,4)   |      |                 |
| ORD_TransportationCost | ราคาค่าขนส่ง                   | MONEY(19,4)   |      |                 |
| ORD_PaymentCost        | ราคาค่าจ้างพนักงานขนส่ง        | MONEY(19,4)   |      |                 |
| ORD_Date               | วันที่สั่งซื้อสินค้า           | DATETIME      |      |                 |
| ORD_PaymentDate        | วันที่ชำระเงิน                 | DATETIME      |      |                 |
| ORD_ShippingDate       | วันที่ส่งสินค้า                | DATETIME      |      |                 |
| ORD_UpdateDate         | วันที่ปรับปรุงข้อมูล           | DATETIME      |      |                 |
| ORD_UpdateBy           | ชื่อผู้ปรับปรุงข้อมูล          | NVARCHAR(50)  |      |                 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2 : พจนานุกรมของตาราง M\_User

| ชื่อตาราง             | M_User                  |              |      |                 |
|-----------------------|-------------------------|--------------|------|-----------------|
| ชื่อแอตทริบิวต์       | คำอธิบาย                | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิง |
| USR_ID                | รหัสลูกค้า              | INT(4)       | PK   |                 |
| USR_BillingAddressID  | ที่อยู่สำหรับออกใบเสร็จ | INT(4)       | FK   | M_Address       |
| USR_ShippingAddressID | ที่อยู่สำหรับส่งของ     | INT(4)       | FK   | M_Address       |
| USR_Title             | คำนำหน้า                | CHAR(10)     |      |                 |
| USR_FirstName         | ชื่อ                    | NVARCHAR(50) |      |                 |
| USR_LastName          | นามสกุล                 | NVARCHAR(50) |      |                 |
| USR_DateOfBirth       | วันเดือนปีเกิด          | DATETIME     |      |                 |
| USR_Telephone         | หมายเลขโทรศัพท์         | NVARCHAR(20) |      |                 |
| USR_Mobile            | หมายเลขมือถือ           | NVARCHAR(20) |      |                 |
| USR_Email             | อีเมล                   | NVARCHAR(50) |      |                 |
| USR_UserName          | ชื่อผู้ใช้              | NVARCHAR(20) |      |                 |
| USR_Password          | รหัสผ่าน                | NVARCHAR(20) |      |                 |

ตารางที่ 3 : พจนานุกรมของตาราง T\_OrderDetail

| ชื่อตาราง        | T_OrderDetail        |             |      |                 |
|------------------|----------------------|-------------|------|-----------------|
| ชื่อแอตทริบิวต์  | คำอธิบาย             | ชนิดข้อมูล  | คีย์ | ตารางที่อ้างอิง |
| ODT_ID           | รหัสรายการใบสั่งซื้อ | INT(4)      | PK   |                 |
| ODT_ORD_ID       | รหัสใบสั่งซื้อ       | INT(4)      | FK   | T_Order         |
| ODT_PDT_ID       | รหัสสินค้า           | INT(4)      | FK   | T_ProductDetail |
| ODT_InitialPrice | ราคาเริ่มต้น         | MONEY(19,4) |      |                 |
| ODT_SoldPrice    | ราคาขาย              | MONEY(19,4) |      |                 |
| ODT_BoughtPrice  | ราคาต้นทุน           | MONEY(19,4) |      |                 |
| ODT_Quantity     | ปริมาณที่สั่งซื้อ    | INT(4)      |      |                 |

ตารางที่ 4 : พจนานุกรมของตาราง T\_Category

| ชื่อตาราง       | T_OrderDetail    |              |      |                 |
|-----------------|------------------|--------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย         | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิง |
| CAT_ID          | รหัสประเภทสินค้า | INT(4)       | PK   |                 |
| CAT_Name        | ประเทศสินค้า     | NVARCHAR(80) |      |                 |
| CAT_Pic         | รูปของสินค้า     | IMAGE        |      |                 |
| CAT_Discount    | ส่วนลด           | INT(4)       |      |                 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5 : พจนานุกรมของตาราง T\_SubCate

| ชื่อตาราง       | T_OrderDetail          |              |      |                 |
|-----------------|------------------------|--------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย               | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิง |
| SUB_ID          | เลขที่ประเภทสินค้าย่อย | INT(4)       | PK   |                 |
| SUB_CAT_ID      | รหัสประเทศสินค้าย่อย   | NVARCHAR(50) | FK   | T_Category      |
| SUB_Name        | รายการประเภทสินค้าย่อย | INT(4)       |      |                 |
| SUB_Discount    | ส่วนลด                 | INT(4)       |      |                 |

ตารางที่ 6 : พจนานุกรมของตาราง T\_ProductDetail

| ชื่อตาราง        | T_ProductDetail       |               |      |                 |
|------------------|-----------------------|---------------|------|-----------------|
| ชื่อแอตทริบิวต์  | คำอธิบาย              | ชนิดข้อมูล    | คีย์ | ตารางที่อ้างอิง |
| PDT_ID           | รหัสรายการสินค้า      | INT(4)        | PK   |                 |
| PDT_PRD_ID       | รหัสสินค้า            | INT(4)        | FK   | T_Product       |
| PDT_Title        | ประเภทสินค้า          | NVARCHAR(50)  |      |                 |
| PDT_Description  | รายละเอียดสินค้า      | NVARCHAR(200) |      |                 |
| PDT_Weight       | น้ำหนัก               | DECIMAL(18,2) |      |                 |
| PDT_InitialPrice | ราคาเริ่มต้น          | MONEY(19,4)   |      |                 |
| PDT_Price        | ราคาขาย               | MONEY(19,4)   |      |                 |
| PDT_Discount     | ส่วนลด                | INT(4)        |      |                 |
| PDT_Package      | ประเภทการบรรจุหีบห่อ  | NVARCHAR(50)  |      |                 |
| PDT_Status       | สถานะสินค้า           | BIT           |      |                 |
| PDT_UpdateDate   | วันที่ปรับปรุงข้อมูล  | DATETIME      |      |                 |
| PDT_UpdateBy     | ชื่อผู้ปรับปรุงข้อมูล | NVARCHAR(50)  |      |                 |

ตารางที่ 7 : พจนานุกรมของตาราง T\_Product

| ชื่อตาราง       | DEPARTMENT               |               |      |                 |
|-----------------|--------------------------|---------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย                 | ชนิดข้อมูล    | คีย์ | ตารางที่อ้างอิง |
| PRD_ID          | รหัสสินค้า               | INT(4)        | PK   |                 |
| PRD_Name        | ชื่อสินค้า               | NVARCHAR(50)  |      |                 |
| PRD_Brand       | ยี่ห้อสินค้า             | NVARCHAR(50)  |      |                 |
| PRD_Description | รายละเอียดสินค้า         | NVARCHAR(50)  |      |                 |
| PRD_SmallPic    | รูปภาพแสดงสินค้าขนาดเล็ก | NVARCHAR(50)  |      |                 |
| PRD_LargePic    | รูปภาพแสดงสินค้าขนาดใหญ่ | NVARCHAR(50)  |      |                 |
| PRD_Rating      | ระดับความนิยม            | DECIMAL(18,2) |      |                 |
| PRD_Status      | สถานะสินค้า              | BIT           |      |                 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

|                |                       |              |  |  |
|----------------|-----------------------|--------------|--|--|
| PRD_UpdateDate | วันที่ปรับปรุงข้อมูล  | DATETIME     |  |  |
| PRD_UpdateBy   | ชื่อผู้ปรับปรุงข้อมูล | NVARCHAR(50) |  |  |

ตารางที่ 8 : พจนานุกรมของตาราง T\_ShippingType

| ชื่อตาราง       | T_ShippingType         |              |      |                 |
|-----------------|------------------------|--------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย               | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิง |
| SPT_ID          | รหัสการส่งสินค้า       | INT(4)       | PK   |                 |
| SPT_SPT_Name    | ประเภทการส่งสินค้า     | NVARCHAR(50) |      |                 |
| SPT_ServiceCost | ค่าบริการ              | MONEY(19,4)  |      |                 |
| SPT_Description | รายละเอียดการส่งสินค้า | NVARCHAR(50) |      |                 |
| SPT_UpdateDate  | วันที่ปรับปรุงข้อมูล   | DATETIME     |      |                 |
| SPT_UpdateBy    | ชื่อผู้ปรับปรุงข้อมูล  | NVARCHAR(50) |      |                 |

ตารางที่ 9 : พจนานุกรมของตาราง T\_ShippingCondition

| ชื่อตาราง       | T_ShippingCondition        |               |      |                 |
|-----------------|----------------------------|---------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย                   | ชนิดข้อมูล    | คีย์ | ตารางที่อ้างอิง |
| SPD_ID          | เลขที่เงื่อนไขการส่งสินค้า | INT(4)        | PK   |                 |
| SPD_SPT_ID      | รหัสเงื่อนไขการส่งสินค้า   | INT(4)        | FK   | T_ShippingType  |
| SPD_Condition   | เงื่อนไขการส่งสินค้า       | NVARCHAR(512) |      |                 |

ตารางที่ 10 : พจนานุกรมของตาราง T\_Claim

| ชื่อตาราง       | T_Claim                    |               |      |                 |
|-----------------|----------------------------|---------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย                   | ชนิดข้อมูล    | คีย์ | ตารางที่อ้างอิง |
| CLM_ID          | รหัสการส่งเคลม             | INT(4)        | PK   |                 |
| CLM_ODT_ID      | รหัสส่งสินค้า              | INT(4)        | FK   | T_OrderDetail   |
| CLM_Title       | ประเภทการเคลมสินค้า        | NVARCHAR(50)  |      |                 |
| CLM_Description | รายละเอียดการส่งเคลมสินค้า | NVARCHAR(100) |      |                 |
| CLM_Status      | สถานะการส่งเคลมสินค้า      | INT(4)        |      |                 |
| CLM_UpdateDate  | วันที่ปรับปรุงข้อมูล       | DATETIME      |      |                 |
| CLM_UpdateBy    | ชื่อผู้ปรับปรุงข้อมูล      | NVARCHAR(50)  |      |                 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 11: พจนานุกรมของตาราง M\_Address

| ชื่อตาราง       | M_Address    |               |      |                    |
|-----------------|--------------|---------------|------|--------------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย     | ชนิดข้อมูล    | คีย์ | ตารางที่อ้างอิงถึง |
| ADD_ID          | รหัส ITEM    | INT(4)        | PK   |                    |
| ADD_1           | ที่อยู่ที่ 1 | NVARCHAR(100) |      |                    |
| ADD_2           | ที่อยู่ที่ 2 | NVARCHAR(100) |      |                    |
| ADD_District    | อำเภอ        | NVARCHAR(50)  |      |                    |
| ADD_State       | จังหวัด      | NVARCHAR(50)  |      |                    |
| ADD_AMP_ID      | รหัสอำเภอ    | INT(4)        | FK   | T_Amphur           |
| ADD_PostCode    | รหัสไปรษณีย์ | NVARCHAR(5)   |      |                    |

ตารางที่ 12 : พจนานุกรมของตาราง T\_Amphur

| ชื่อตาราง       | T_Amphur    |              |      |                    |
|-----------------|-------------|--------------|------|--------------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย    | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิงถึง |
| AMP_ID          | รหัสอำเภอ   | INT(4)       | PK   |                    |
| AMP_PRV_ID      | รหัสจังหวัด | INT(4)       | FK   | T_Province         |
| AMP_Name        | ชื่ออำเภอ   | NVARCHAR(50) |      |                    |

ตารางที่ 13 : พจนานุกรมของตาราง M\_Province

| ชื่อตาราง       | LEAVETYPE   |              |      |                    |
|-----------------|-------------|--------------|------|--------------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย    | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิงถึง |
| PRV_ID          | รหัสจังหวัด | INT(4)       | PK   |                    |
| PRV_Name        | ชื่อจังหวัด | NVARCHAR(50) |      |                    |

ตารางที่ 14 : พจนานุกรมของตาราง T\_PaymentType

| ชื่อตาราง       | T_OrderDetail         |               |      |                    |
|-----------------|-----------------------|---------------|------|--------------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย              | ชนิดข้อมูล    | คีย์ | ตารางที่อ้างอิงถึง |
| PAY_ID          | รหัสประเภทการชำระเงิน | INT(4)        | PK   |                    |
| PAY_Name        | ประเภทการชำระเงิน     | NVARCHAR(50)  |      |                    |
| PAY_Description | รายละเอียดการชำระเงิน | NVARCHAR(512) |      |                    |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 15: พจนานุกรมของตาราง T\_PromotionDetail

| ชื่อตาราง       | M_Address                      |              |      |                 |
|-----------------|--------------------------------|--------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย                       | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิง |
| PMD_ID          | เลขที่รายการส่งเสริมการขาย     | INT(4)       | PK   |                 |
| PMD_PMT_ID      | รหัสประเภทรายงานส่งเสริมการขาย | INT(4)       | FK   | T_PromotionType |
| PMD_Status      | สถานะปัจจุบัน                  | NVARCHAR(50) |      |                 |
| PMD_StartDate   | วันเริ่มต้น                    | DATETIME     |      |                 |
| PMD_EndDate     | วันสิ้นสุด                     | DATETIME     |      |                 |
| PMD_UpdateDate  | วันที่ปรับปรุงข้อมูล           | DATETIME     | FK   | T_Amphur        |
| PMD_UpdateBy    | ชื่อผู้ปรับปรุงข้อมูล          | NVARCHAR(50) |      |                 |

ตารางที่ 16 : พจนานุกรมของตาราง T\_PromotionType

| ชื่อตาราง       | LEAVETYPE                      |              |      |                 |
|-----------------|--------------------------------|--------------|------|-----------------|
| ชื่อแอตทริบิวต์ | คำอธิบาย                       | ชนิดข้อมูล   | คีย์ | ตารางที่อ้างอิง |
| PRT_ID          | รหัสประเภทรายงานส่งเสริมการขาย | INT(4)       | PK   |                 |
| PRT_Name        | รายงานส่งเสริมการขาย           | NVARCHAR(50) |      |                 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ประวัติผู้เขียน

|                            |  |
|----------------------------|--|
| ชื่อผู้เขียน               | นางสาวรุจิรา แซ่เตีย                             |
| วันเกิด                    | 20 เมษายน 2523                                   |
| สถานที่เกิด                | นครราชสีมา                                       |
| วุฒิการศึกษาระดับปริญญาตรี | วิศวกรรมศาสตรบัณฑิต                              |
| สถานที่สำเร็จการศึกษา      | สาขาวิศวกรรมไฟฟ้า<br>มหาวิทยาลัยเทคโนโลยีสุรนารี |
| ปีที่สำเร็จการศึกษา        | 2543   |



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้