

การพัฒนาโปรแกรมประยุกต์สำหรับธุรกิจสำรองห้องพักโรงแรม.

โดยใช้ Decision Tree

Hotels Reservations Application Development Using Decision Tree

โดย

นายกุลชาติ ตั้งกาญจนภรณ์

รหัส 45066084



\*H002184\*

อาจารย์ที่ปรึกษา

ผศ.ดร. วรพจน์ กรีสระเดช

วัน เดือน ปี..... 0 8 ก.พ. 2550

เลขทะเบียน..... 02184

เลขเรียกหนังสือ..... อพ. ก925ก 2546

"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน  
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 2 ปีการศึกษา 2546

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ชื่อหัวข้อ	การพัฒนาโปรแกรมประยุกต์สำหรับธุรกิจสำรองห้องพักรวมโดยใช้ Decision Tree
นักศึกษา	นายกฤษชาติ ตั้งกาญจนภาสน์
อาจารย์ที่ปรึกษา	ผศ.ดร. วรพงษ์ กรีสระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2546

### บทคัดย่อ

ธุรกิจสำรองห้องพักโรงแรมจากเว็บไซต์เป็นธุรกิจที่นำเอาเทคโนโลยีอินเทอร์เน็ต มาประยุกต์ใช้เพิ่มมูลค่าทางธุรกิจ และนำมาใช้เป็นช่องทางการจัดจำหน่ายได้อย่างมีประสิทธิภาพ เนื่องจากการที่ธุรกิจประเภทนี้มีข้อมูลที่เกี่ยวข้องเป็นจำนวนมาก จึงมีความจำเป็นที่จะต้องค้นหาวิธีการจัดการ กับข้อมูลเหล่านั้น เพื่อค้นหาสารสนเทศมาช่วยสนับสนุนการตัดสินใจต่างๆที่เป็นประโยชน์ต่อธุรกิจ จึงนำหลักการ คาด้าไมนิง ซึ่งเป็นกระบวนการค้นหา วิเคราะห์ข้อมูลที่มีความหมายจากข้อมูลขนาดใหญ่ เพื่อค้นหาข้อสรุปไปสู่ข้อมูลสารสนเทศที่เป็นประโยชน์มากขึ้น

การศึกษานี้เป็นการศึกษาความเป็นไปได้ในการนำเอาเทคโนโลยี คาด้าไมนิง มาใช้เพื่อค้นหาความสัมพันธ์ของข้อมูลต่างๆของ ธุรกิจ เพื่อที่จะนำความรู้ที่ได้ไปเป็นแนวทางในการสนับสนุนการตัดสินใจทางธุรกิจต่อไป

<b>Title</b>	Hotels Reservations Application Development Using Decision Tree
<b>Student</b>	Mr. Kunlachart Tangkanjanapas
<b>Advisor</b>	Asst.Prof. Dr. Worapoj Kreesuradej
<b>Level of Study</b>	Master of Science in Information Technology
<b>Major</b>	Information Science
<b>Academic Year</b>	2003

## ABSTRACT

Online Hotels Reservation Business is one of Business section that uses Internet Technology for business purpose. Because of Mass of data in business, It is necessary to find how to transform data to information. Data Mining is the process of extracting knowledge hidden from large volumes of raw data.

This report study possibility to use Data Mining Technology for extracting knowledge hidden from Online Hotels Reservation Database.

## กิตติกรรมประกาศ

ความสำเร็จของการพัฒนาโครงการพัฒนาระบบงานฉบับนี้ สำเร็จขึ้นได้จากความช่วยเหลือของบุคคลหลายๆ ท่าน ข้าพเจ้ามีความรู้สึกขอบคุณทุกท่านที่มีส่วนช่วยเหลือในด้านต่างๆ ด้วยความจริงใจ หากขาดบุคคลที่จะกล่าวถึงดังต่อไปนี้ ก็จะไม่ส่งผลกระทบต่อความสำเร็จของโครงการพัฒนาระบบงานฉบับนี้

ขอขอบคุณ บิดา มารดาผู้ให้กำเนิดผู้ซึ่งให้ทุกอย่าง

ขอขอบคุณ อาจารย์ ผศ.ดร. วรพจน์ กฤษระเดช อาจารย์ที่ปรึกษา ที่ให้ความรู้ แนะนำ หนังสือและให้คำปรึกษา ตลอดจนชี้แนะแนวทางการแก้ไขปัญหามาโดยตลอด

ขอขอบคุณ คุณวรพล สติรกุล Database Developer บริษัท Anacott Asia Pacific co., ltd. ผู้ให้คำแนะนำชี้แนะในเรื่องการเชื่อมโยงฐานข้อมูล

สุดท้าย ขอขอบคุณ คุณจิตวรี ตรงต่อศักดิ์ และคุณแดง บัวสุก เพื่อนร่วมงานที่คอยช่วยเหลือ และเป็นกำลังใจให้เสมอมา

กุตชาติ ตั้งกาญจนภาสน์

# สารบัญ

หน้า

บทคัดย่อภาษาไทย .....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ .....	IV
สารบัญตาราง.....	V
สารบัญภาพ .....	VII
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมาของปัญหา .....	1
1.2 วัตถุประสงค์ของการพัฒนาระบบ .....	2
1.3 ขอบเขตของการพัฒนาระบบ.....	2
1.4 เทคโนโลยีที่ใช้พัฒนาระบบ .....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ .....	2
1.6 ระยะเวลา .....	3
2. ธุรกิจสำรองห้องพักจากเว็บไซต์และความจำเป็นของคาด้าไมนิ่งต่อธุรกิจ .....	6
2.1 ธุรกิจสำรองห้องพักโรงแรม .....	6
2.2 ความจำเป็นของคาด้าไมนิ่งต่อธุรกิจ .....	7
3. แนวคิดและทฤษฎีที่เกี่ยวข้อง.....	8
3.1 คาด้าไมนิ่ง .....	8
4. เทคนิค Classification และอัลกอริทึม Decision Tree.....	21
4.1 Classification .....	21
4.2 Tree Induction .....	21
4.3 วิธีการ Decision Tree .....	23
5. การออกแบบและพัฒนาระบบงาน .....	31
5.1 องค์ประกอบต่างๆที่ใช้ในการพัฒนา.....	32

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

หน้า

5.3 Decision Tree System Web Site Diagram .....	34
5.4 System Design Framework.....	35
5.5 Activity Diagram Decision Tree System.....	35
5.6 Use Case Diagram Decision Tree System.....	36
5.7 Sequence Diagram Decision Tree System.....	37
5.8 Class Diagram Decision Tree System .....	43
6. การเตรียมข้อมูล.....	52
6.1 แหล่งที่มาของข้อมูล .....	52
6.2 การคัดเลือกข้อมูล .....	58
6.3 การทำความสะอาดข้อมูล.....	59
6.4 การแปลงข้อมูล .....	59
6.5 การนำข้อมูลเข้าตารางข้อมูลสำหรับการเทรนนิ่ง.....	60
7. การเทรนนิ่งและการทำนาย .....	67
7.1 ส่วนงานการเทรนนิ่ง (Training Phase).....	68
7.2 ส่วนงานการทำนาย (Prediction Phase).....	88
8. บทสรุป.....	91
8.1 สรุปหลักการที่ใช้ในระบบ .....	91
8.2 สรุปกระบวนการในการทำงาน .....	91
8.3 สรุปผลการทดสอบ.....	92
8.4 ข้อเสนอแนะ .....	92
บรรณานุกรม .....	93
ประวัติผู้เขียน.....	94

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญตาราง

หน้า

ตารางที่

1.1 ระยะเวลาในการทำงาน.....	3
5.1 รายละเอียดของ Sequence Diagram ของ Use Case Authentication .....	38
5.2 รายละเอียดของ Sequence Diagram ของ Use Case Setup .....	39
5.3 รายละเอียดของ Sequence Diagram ของ Use Case Training .....	40
5.4 รายละเอียดของ Sequence Diagram ของ Predicting.....	42
6.1 รายละเอียดตาราง CSDB.BKINFO.....	54
6.2 รายละเอียดตาราง CSDB.BKSTATUS.....	55
6.3 รายละเอียดตาราง CSDB.BKCLIENT .....	57
6.4 รายละเอียดตาราง CSDB.BKCHARGE .....	57
6.5 รายละเอียดตาราง CSDB.TBL_XML_BOOKING_INFO .....	58
6.6 ตารางที่ใช้ในการเทรนนิ่ง .....	66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญภาพ

หน้า

ภาพที่

3.1	คาค้าไม้หนึ่งในกระบวนการของ Knowledge Discovery .....	9
3.2	เป้าหมาย คาค้าไม้หนึ่ง ในรูปแบบต่างๆ .....	9
3.3	กระบวนการทำ คาค้าไม้หนึ่ง.....	14
4.1	Decision tree for the concept Play Tennis .....	23
4.2	ความสัมพันธ์ของการสุ่มข้อมูล และ ค่า Entropy .....	24
4.3	แนวคิดของอัลกอริทึมID3 .....	26
5.1	โครงสร้างพื้นฐาน โดยรวมของระบบงาน .....	32
5.2	Decision Tree System Web Site Diagram .....	34
5.3	System Design Framework .....	35
5.4	Activity Diagram Decision Tree System.....	36
5.5	Use Case Diagram .....	36
5.6	Sequence Diagram ของ Use Case Authentication .....	37
5.7	Sequence Diagram ของ Use Case Setup .....	39
5.8	Sequence Diagram ของ Use Case Training .....	40
5.9	Sequence Diagram ของ Use Case Predicting.....	42
5.10	User Service Class .....	43
5.11	ucLeftBar Class .....	44
5.12	ucTopBar Class .....	44
5.13	Autorization Class .....	44
5.14	Setup Class .....	45
5.15	Training Config Class .....	46
5.16	ExistingTrees Class .....	46
5.17	displayTreeDetails Class .....	47
5.18	displayTreeView Class.....	47

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาตจากเจ้าของลิขสิทธิ์

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญญภาพ (ต่อ)

ภาพที่	หน้า
5.20 PredictingSamples Class .....	48
5.21 Business Service Class .....	49
5.22 Training Process Class Class .....	49
5.23 cDataMiningServer Class .....	50
5.24 cNode Class .....	50
5.25 colNodes Class .....	51
5.26 cSamples Class .....	51
5.27 cDecisionTrees Class .....	52
5.28 PredictingSamplesProcess Class .....	52
5.29 cDataAccess Object Class .....	53
7.1 หน้าจอเริ่มต้นของระบบ .....	67
7.2 หน้าจอปรับค่าการติดต่อฐานข้อมูล .....	68
7.3 หน้าจอเริ่มต้นของกระบวนการเทรนนิ่ง .....	69
7.4 หน้าจอเลือกตารางที่ต้องการใช้ในการ Training.....	69
7.5 หน้าจอเลือก Target Attribute เพื่อใช้ในการ Training.....	70
7.6 หน้าจอเลือก Descriptor Attributes เพื่อใช้ในการ Training.....	70
7.7 หน้าจอขึ้นย่นค่า Attributes ที่จะใช้ในการ Mining .....	71
7.8 หน้าจอแสดงผลลัพธ์หลังจากเสร็จสิ้นการการสร้าง Model.....	74
7.9 หน้าจอเลือกรายการ Model ต่างๆที่ได้ทดสอบมา .....	74
7.10 หน้าจอแสดงผลลัพธ์ในรูปแบบ Tree View.....	75
7.11 หน้าจอแสดงผลลัพธ์ในรูปแบบ Rules .....	76
7.12 หน้าจอแสดงรายการ Model ที่มีเพื่อเลือกใช้ในการ Predicting .....	88
7.13 หน้าจอเลือกตารางที่ต้องการจะนำมาทำนายผล.....	89
7.14 หน้าจอเลือก Reference Attribute .....	89
7.15 หน้าจอแสดงผลลัพธ์จากการประมวลผลทำนายตาม Model .....	90

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การข่งานนี้เพื่อการศึกษาเท่านั้น เมื่อผู้ยู่ติให้เข้าไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาของปัญหา

ปัจจุบันเทคโนโลยีอินเทอร์เน็ตได้ถูกนำมาใช้เพื่อวัตถุประสงค์ทางธุรกิจกันอย่างแพร่หลาย ธุรกิจสำรองห้องพัก เป็นธุรกิจหนึ่งที่น่าเอาเทคโนโลยีนี้ มาใช้เป็นช่องทางในการเพิ่มมูลค่าทางธุรกิจ โดยนำมาใช้เป็นช่องทางการจัดจำหน่ายได้อย่างมีประสิทธิภาพ เนื่องจากเป็นวิธีทางธุรกิจที่ทำให้ลูกค้า สามารถที่จะรับข้อมูล รายละเอียดต่างๆของสินค้าได้อย่างครบถ้วน ไม่ว่าจะเป็นรายละเอียดห้องพัก ขนาดของห้องพัก ราคา ตลอดจนไปถึงข้อมูลบริการอื่นๆ อีกทั้ง กระบวนการสำรองห้องพักผ่านระบบอินเทอร์เน็ต ทำให้สามารถดำเนินการได้อย่าง รวดเร็วและง่ายดาย ส่งผลให้ธุรกิจ ประเภทนี้ได้รับการ ตอบสนองจากลูกค้า เป็นอย่างดี

จากข้อมูลสถิติจากเว็บไซต์ทางการของการท่องเที่ยวแห่งประเทศไทย พบว่าการเข้ามาท่องเที่ยวของนักท่องเที่ยวในประเทศไทยของปี 2545 มีจำนวน 940,653 คน เพิ่มขึ้นจากปี 2544 ถึง 5.91 เปอร์เซ็นต์ ซึ่งจากจำนวนนักท่องเที่ยวที่เข้ามาดังกล่าวคาดว่าจะมีการเข้าพักห้องพักในโรงแรมต่อปี ซึ่งถ้าคิดประมาณ 70 เปอร์เซ็นต์ ก็จะได้ตัวเลขห้องพักประมาณ 650,000 ห้อง/ปี ซึ่งจะเห็นได้ว่ามีความต้องการต่อความต้องการใช้บริการห้องพักเป็นจำนวนมาก หากแต่ว่าในการให้บริการสำรองห้องพักของโรงแรมในปัจจุบันยังมีข้อจำกัด เช่นต้องค้นหาข้อมูลของโรงแรมจากในหลาย ๆ แหล่งข้อมูล ทำให้ผู้ใช้บริการไม่ได้รับความสะดวกสบาย ธุรกิจสำรองห้องพักในโรงแรมผ่านอินเทอร์เน็ต จึงเป็นอีกทางเลือกหนึ่งของธุรกิจ โดยจะเป็นการนำเอาเทคโนโลยีด้านอินเทอร์เน็ตมาช่วยในการให้บริการสำรองห้องพักโรงแรมอย่างครบวงจร ซึ่งจะเป็นธุรกิจขนาดเล็ก และมีเงินทุนจำกัด โดยจะเป็นตัวกลางระหว่างธุรกิจโรงแรมกับผู้ใช้บริการ เพื่อที่จะทำให้ผู้บริโภคมีความสะดวกสบายในการสำรองห้องพักโรงแรมมากขึ้น อีกทั้งยังสามารถขยายบริการไปยังบริการที่เกี่ยวข้องกับการสำรองห้องพักของโรงแรม เช่น การสำรองแพ็คเกจทัวร์ การสำรองรถเช่า เป็นต้น

เนื่องจากการที่ธุรกิจสำรองห้องพักนี้ มีข้อมูล ที่เกี่ยวข้องเป็นจำนวนมาก ซึ่งหากไม่ได้นำข้อมูลเหล่านั้น มาใช้ให้เป็นประโยชน์ ข้อมูลเหล่านั้นก็สูญเปล่า อีกทั้ง การวิเคราะห์จากฐานข้อมูล เดี่ยวอาจไม่ให้ความรู้ที่เพียงพอ จึงมีความจำเป็นที่จะต้องจัดการกับข้อมูลเหล่านั้น เพื่อค้นหาสารสนเทศมาช่วยสนับสนุนการตัดสินใจแผนงานต่างๆที่เป็นประโยชน์ต่อธุรกิจ ซึ่งหลักการการค้า ไม่นิ่งซึ่งเป็นกระบวนการ ในการสืบค้นหารูปแบบความสัมพันธ์ที่มีประโยชน์ของข้อมูลใน ฐานข้อมูลขนาดใหญ่ จึงถูกนำมาใช้เพื่อค้นหาข้อสรุปไปสู่ข้อมูลที่เป็นประโยชน์ต่อไป การพัฒนา  
ไม่ว่าการณีใดจทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมประยุกต์นี้เป็นการพัฒนาโดยใช้หลักการของ Decision Tree ซึ่งเป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้เพื่อเรียนรู้จากข้อมูลต่างๆของธุรกิจเพื่อค้นหากลุ่มลูกค้าที่มีศักยภาพ และนำความรู้ที่ได้ไปเป็นแนวทางในการสนับสนุนการตัดสินใจทางธุรกิจต่อไป

## 1.2 วัตถุประสงค์ของการพัฒนาระบบ

การพัฒนาโปรแกรมประยุกต์สำหรับธุรกิจสำรองห้องพักโรงแรมโดยใช้ Decision Tree นี้จัดทำขึ้นเพื่อ

1. เพื่อศึกษาทฤษฎีค่าต้นไม้เชิงเทคนิค Decision Tree
2. เพื่อนำมาประยุกต์ใช้ค้นหา พฤติกรรมของข้อมูลต่างๆของธุรกิจสำรองห้องพักโรงแรมผ่านทางเว็บไซต์
3. เพื่อที่จะค้นหาลักษณะของลูกค้าที่มีศักยภาพ (Potential Customer) เพื่อเป็นแนวทางในการสนับสนุนการตัดสินใจทางธุรกิจต่อไป
4. เพื่อที่จะประมวลผลในการวิเคราะห์ข้อมูลลูกค้าใหม่ว่าเป็นลูกค้าที่มีศักยภาพหรือไม่ จากกฎที่ค้นหาได้

## 1.3 ขอบเขตของการพัฒนาระบบ

การพัฒนาโปรแกรมประยุกต์สำหรับธุรกิจสำรองห้องพักโรงแรมโดยใช้ Decision Tree นี้ประกอบด้วยส่วนงานต่างๆดังต่อไปนี้

1. ระบบการวิเคราะห์หากฎ Decision Tree ของข้อมูลลูกค้าที่สำรองห้องพัก
2. ระบบการนำข้อมูลลูกค้ามาประมวลผลเปรียบเทียบกับตามกฎ Decision Tree เพื่อวิเคราะห์หาความเป็นไปได้ที่จะเป็นลูกค้าที่มีศักยภาพ

## 1.4 เทคโนโลยีที่ใช้ในการพัฒนาระบบ

ระบบได้รับการพัฒนาเป็น Web Application บนพื้นฐาน .Net Framework 1.1 ซึ่งทำงานในระบบปฏิบัติการ Microsoft Windows 2000 โดยมีรายละเอียดดังนี้

- การพัฒนา Web Application ใช้โปรแกรม Microsoft Visual Studio .net 2003
- ใช้ Microsoft SQL Server 2000 เป็นฐานข้อมูลหลักในการเก็บข้อมูล
- ผลลัพธ์ของโมเดลเก็บในรูปแบบ XML
- Tree View เพื่อใช้ในการแสดงผลในรูปแบบต้นไม้ใน Browser

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

### 1.5.1 ธุรกิจสำรองห้องพักโรงแรม

- 1.5.1.1 ผู้บริหารสามารถมองเห็นถึงภาพรวมของลักษณะของลูกค้าที่มีศักยภาพได้มากขึ้น เพื่อนำไปเป็นแนวทางในการสนับสนุนการตัดสินใจทางธุรกิจต่อไป
- 1.5.1.2 เจ้าหน้าที่บริการลูกค้าสามารถนำข้อมูลลูกค้ามาประมวลผลเปรียบเทียบตามกฎ เพื่อให้ทราบถึงความเป็นไปได้ที่จะเป็นลูกค้าที่มีศักยภาพ จากนั้นจึงเน้นให้บริการกับลูกค้ากลุ่มที่มีความเป็นไปได้ที่จะเป็นลูกค้าที่มีศักยภาพเป็นพิเศษ เพื่อลดความสูญเสียทางธุรกิจใน โดยเฉพาะในกรณีการสูญเสียลูกค้าที่มีศักยภาพเนื่องจากสาเหตุการบริการล่าช้า
- 1.5.1.3 ได้รับสารสนเทศที่สะดวกรวดเร็ว เนื่องจากพัฒนาเป็น โปรแกรมประยุกต์ทางเว็บไซต์ ผู้ใช้งานไม่จำเป็นต้องลงโปรแกรมเพิ่มเติม อีกทั้งยังสามารถเข้าใช้งานจากที่ใดก็ได้

### 1.5.2 ผู้พัฒนาระบบ

- 1.5.2.1 ได้รับความรู้เกี่ยวกับเทคโนโลยีดาต้า ไมนิ่ง
- 1.5.2.2 ได้นำทฤษฎีเกี่ยวกับการ โปรแกรมเชิงวัตถุและเทคโนโลยี Microsoft .NET มาประยุกต์ใช้ในการออกแบบและพัฒนาระบบงานจริง
- 1.5.2.3 ใช้เป็นแนวทางในการพัฒนาระบบที่เกี่ยวข้องกับเทคโนโลยีดาต้า ไมนิ่งที่มีความซับซ้อนมากขึ้นในอนาคต
- 1.5.2.4 ใช้เป็นแนวทางให้กับนักศึกษา ผู้บริหาร ตลอดจนบุคคลทั่วไป ในการนำเทคโนโลยีดาต้า ไมนิ่งมาใช้ในงาน

## 1.6 ระยะเวลา

### ตารางที่ 1.1 ตารางระยะเวลาในการทำงาน

กิจกรรม	2546		2547	
	พ.ย.	ธ.ค.	ม.ค.	ก.พ.
ศึกษาความเป็นไปได้				
วิเคราะห์ระบบ				
ศึกษาเครื่องมือที่จะพัฒนาระบบ				
รวบรวมข้อมูล วิเคราะห์ออกแบบและจัดทำฐานข้อมูล				
ออกแบบและพัฒนาระบบ				
ทดสอบระบบ				
จัดทำเอกสารประกอบการใช้งาน				

โดยมีกำหนดระยะเวลาในการพัฒนาระบบต้นแบบเริ่ม 3 พฤษภาคม 2546 และสิ้นสุดวันที่ 12 กุมภาพันธ์ 2547 โดยมีรายละเอียดระยะเวลาในการพัฒนาระบบดังแสดงในแผนงานในการพัฒนาระบบ

#### รายละเอียด

1. ศึกษาความเป็นไปได้ ประกอบด้วยขั้นตอนย่อยต่างๆดังต่อไปนี้
  - 1.1 ศึกษาเทคโนโลยีดาต้าไมนิ่ง โดยเน้นศึกษา อัลกอริทึมของเทคนิค Decision Tree
  - 1.2 ศึกษาการลักษณะข้อมูลที่มีอยู่ ของระบบสำรองห้องพักโรงแรมจากเว็บไซต์
2. วิเคราะห์ระบบ
  - 2.1 วิเคราะห์ความต้องการของระบบโดยรวม
  - 2.2 วิเคราะห์ความต้องการสำหรับแต่ละส่วนคือ ส่วนงานการวิเคราะห์หากฎ Decision Tree ของข้อมูลลูกค้าที่สำรองห้องพัก และส่วนงานการนำข้อมูลลูกค้ามาประมวลผลเปรียบเทียบตามกฎ Decision Tree เพื่อวิเคราะห์หาความเป็นไปได้ที่จะเป็นลูกค้าที่มีศักยภาพ
3. ออกแบบระบบ
  - 3.1 ออกแบบสถาปัตยกรรมของระบบ
  - 3.2 ออกแบบโครงสร้างพื้นฐานของระบบ
  - 3.3 ออกแบบพฤติกรรมการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ออกแบบส่วนเก็บข้อมูล

1.5 ออกแบบส่วนติดต่อกับผู้ใช้

4. พัฒนาระบบ

4.1 พัฒนาระบบค้ำไม้หนึ่ง ส่วนงานการวิเคราะห์หากฎ Decision Tree ของข้อมูลลูกค้าที่สำรองห้องพัก

4.2 พัฒนาระบบค้ำไม้หนึ่งส่วนงานการนำข้อมูลลูกค้ามาประมวลผลเปรียบเทียบตามกฎ Decision Tree ที่ได้มาจากการวิเคราะห์

5. ทดสอบระบบ



## บทที่ 2

### ธุรกิจสำรองห้องพักจากเว็บไซต์และความจำเป็นของดาต้าไมนิ่งต่อธุรกิจ

#### 2.1 ธุรกิจสำรองห้องพักโรงแรม

ธุรกิจสำรองห้องพักโรงแรมผ่านทางเว็บไซต์ เริ่มต้นจาก การพัฒนาเว็บไซต์ที่ให้ข้อมูลทั่วไป เกี่ยวกับโรงแรมต่างๆ เช่นสถานที่ตั้ง บริการต่างๆ ตลอดจนสถานที่ท่องเที่ยวใกล้เคียง โดยอาศัยรายได้จากค่าโฆษณาที่ได้จากโรงแรมที่ติดต่อขอลงข้อมูลส่งเสริมการขาย จากนั้นก็พัฒนาระบบสำรองห้องพักของโรงแรมจากเว็บไซต์เพื่อให้บริการครบวงจรในการสำรองห้องพักโรงแรม เมื่อธุรกิจขยายตัวขึ้น มีการติดต่อกับตัวแทนจำหน่ายห้องพักของโรงแรม (Destination Management Company - DMC) แทนที่จะติดต่อกับโรงแรมโดยตรง เนื่องจาก DMC ที่เป็นบริษัทขนาดใหญ่ที่ไปรับเหมาซื้อการสำรองห้องพักโรงแรมจำนวนมาก สามารถต่อรองเพื่อได้มาซึ่งต้นทุนของสินค้าที่ถูกกว่าการไปติดต่อกับโรงแรมโดยตรงในลักษณะสำรองห้องพักรายย่อย ซึ่งการที่บริษัทติดต่อกับธุรกิจกับ บริษัท DMC เหล่านี้ นอกจากจะได้ต้นทุนของการสำรองห้องพักในอัตราที่ต่ำแล้ว ยังเป็นการลดค่าใช้จ่ายที่จำเป็นต้องใช้ในการติดต่อเจรจาต่อรองเพื่อนำห้องพักโรงแรมแต่ละแห่งมาจำหน่ายอีกด้วย เนื่องจากการอาศัย ข้อตกลงทางธุรกิจที่ DMC มีกับโรงแรมต่างๆอยู่แล้วมาใช้

ลักษณะการดำเนินธุรกิจของธุรกิจสำรองห้องพักโรงแรมผ่านทางเว็บไซต์ ในกรณีศึกษานี้เป็นธุรกิจที่พัฒนาเว็บไซต์ รับสำรองห้องพักโรงแรมต่างๆทั่วโลก ให้บริการกับลูกค้าทั่วไป โดยลูกค้าสามารถที่จะรับข้อมูล รายละเอียดต่างๆ ของสินค้าโรงแรม ไม่ว่าจะเป็นรายละเอียดห้องพัก ขนาดของห้องพัก ราคา ตลอดจนไปถึงข้อมูลบริการอื่นๆ อีกทั้ง กระบวนการสำรองห้องพักผ่านระบบอินเทอร์เน็ต สามารถดำเนินการได้อย่างรวดเร็วและง่ายดาย ทำให้ธุรกิจ ประเภทนี้ ได้รับการตอบสนองจากลูกค้า เป็นอย่างดี ทางด้านที่มาของสินค้า จะเป็นการนำเอาสินค้าห้องพักโรงแรมต่างๆ จากตัวแทน DMC ต่างๆ มาขาย ซึ่ง DMC แต่ละบริษัท ก็มีความชำนาญ และความสัมพันธ์ทางธุรกิจกับโรงแรมต่างๆ แตกต่างกันไป ทำให้สินค้าและราคาที่ได้จาก DMC แต่ละรายอาจไม่เท่ากัน DMC บางรายอาจมีข้อเสนอที่ดีกว่าในสินค้าโรงแรมกลุ่มทวีปเอเชีย บางรายอาจมีข้อเสนอที่ดีกว่าในสินค้าโรงแรมกลุ่มทวีปยุโรป เป็นต้น ซึ่งโดยปกติแล้วบริษัทจะขายโรงแรมใดๆจาก DMC รายที่ให้ข้อเสนอกับบริษัทที่ดีที่สุด

## 2.1.1 จุดเด่นของธุรกิจสำรองห้องพักโรงแรม

- 2.1.1.1 มีการลงทุนต่ำเมื่อเทียบกับธุรกิจผลิตสินค้าและบริการทั่วไป
- 2.1.1.2 ใช้อินเทอร์เน็ตในการแสดงข้อมูลของห้องพักและสถานที่ท่องเที่ยวที่อยู่ใกล้เคียง ทำให้ลดเวลาในการค้นหาข้อมูลโรงแรมและสถานที่ท่องเที่ยวของนักท่องเที่ยว
- 2.1.1.3 สามารถยื่นความต้องการสำรองห้องพักโรงแรมผ่านอินเทอร์เน็ตได้ในทันที
- 2.1.1.4 ราคาต่ำกว่าการสำรองห้องพักแบบปกติ เนื่องจากต้นทุนการจัดการต่ำกว่า

## 2.1.2 การวิเคราะห์สภาพแวดล้อมภายนอก

- 2.1.2.1 คู่แข่ง : เว็บไซต์ที่ให้บริการในปัจจุบันอาจจะมีเฉพาะของในแต่ละโรงแรม หรือเว็บไซต์ที่มีการให้บริการในลักษณะเดียวกัน ยังมีขั้นตอนในการใช้บริการที่ซับซ้อนหรือใช้พนักงานทำงาน เช่น ต้องให้นักท่องเที่ยวส่งแฟกซ์ การสำรองห้องพัก เป็นต้น
- 2.1.2.2 นโยบายของรัฐบาล : การส่งเสริมการท่องเที่ยวเป็นนโยบายที่สำคัญของรัฐบาล ซึ่งจะทำให้ธุรกิจประเภทนี้มีโอกาสเติบโตสูง

## 2.2 ความจำเป็นของ ดาต้าไมนิ่ง ต่อธุรกิจ

เนื่องจากธุรกิจสำรองห้องพักนี้มีข้อมูล ที่เกี่ยวข้องเป็นจำนวนมาก ซึ่งหากไม่ได้นำข้อมูลเหล่านั้น มาใช้ให้เป็นประโยชน์ ข้อมูลเหล่านั้นก็สูญเปล่า หรือแม้กระทั่งการนำข้อมูลมาใช้โดยการวิเคราะห์จากฐานข้อมูลเดียวอาจไม่ให้เห็นสารสนเทศที่มากเพียงพอ จึงมีความจำเป็นที่จะต้องจัดการกับข้อมูลเหล่านั้น ดังนั้นเพื่อค้นหาสารสนเทศมาช่วยสนับสนุนการตัดสินใจต่างๆที่เป็นประโยชน์ต่อธุรกิจ การนำหลักการของดาต้าไมนิ่ง ซึ่งเป็นกระบวนการในการสืบค้นหารูปแบบความสัมพันธ์ที่มีประโยชน์ของข้อมูลในฐานข้อมูลขนาดใหญ่ จึงถูกนำมาใช้เพื่อค้นหาข้อสรุปไปสู่อีกรูปแบบที่เป็นประโยชน์ต่อไป การพัฒนาโปรแกรมประยุกต์นี้เป็นการพัฒนาโดยใช้หลักการของ Decision Tree ซึ่งเป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้เพื่อเรียนรู้จากข้อมูลต่างๆของธุรกิจเพื่อค้นหากลุ่มลูกค้าที่มีศักยภาพ และนำความรู้ที่ได้ไปเป็นแนวทางในการสนับสนุนการตัดสินใจทางธุรกิจต่อไป

## บทที่ 3

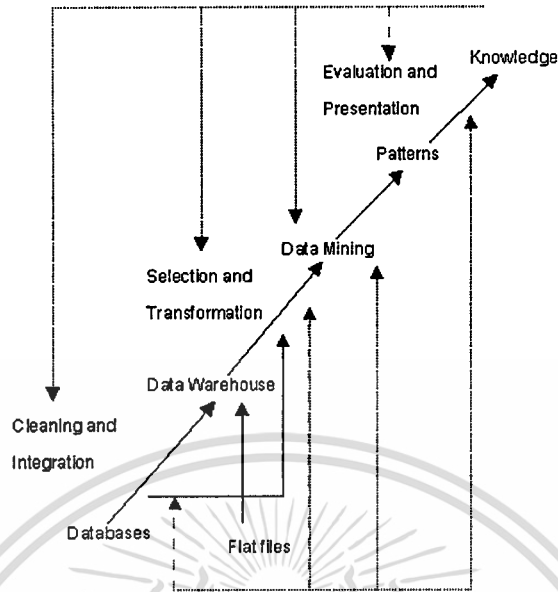
### แนวคิดและทฤษฎีที่เกี่ยวข้อง

#### 3.1 คาด้าไมนิ่ง (Data Mining)

คาด้าไมนิ่ง เป็นเทคนิคที่ใช้ในการค้นหา รูปแบบสารสนเทศ ซ่อนเร้นจาก ข้อมูลขนาดใหญ่ เปรียบเสมือน กระบวนการสนับสนุนการตัดสินใจจากรูปแบบที่แอบแฝงอยู่ในข้อมูล เป็นกระบวนการของการตัดทอนข้อมูลจากฐานข้อมูลขนาดใหญ่ เช่น ความสัมพันธ์ แนวโน้มหรือรูปแบบเฉพาะ(Pattern) ซึ่งสารสนเทศ เหล่านี้มีประโยชน์อย่างมากในการตัดสินใจเชิงธุรกิจ เพื่อให้องค์กรมีประสิทธิภาพในการแข่งขันมากยิ่งขึ้น ในปัจจุบันองค์กรส่วนใหญ่จะเผชิญกับปัญหาของข้อมูลดิบจำนวนมากแต่ข้อมูลที่ประยุกต์ใช้นั้นมีน้อย คาด้าไมนิ่งจึงเป็นสาขาที่คาดว่าจะเป็นที่รู้จักและนำมาใช้ประยุกต์ใช้อย่างแพร่หลายในอนาคต เนื่องจากคาด้าไมนิ่ง สามารถดึงเอาความรู้ ออกจากข้อมูลจำนวนมากที่ถูกเก็บสะสมไว้นั่นเอง

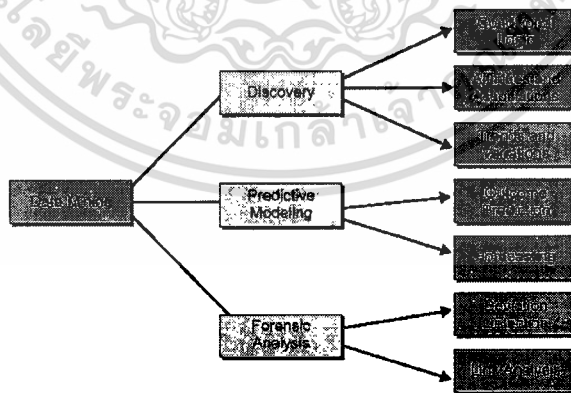
ในโลกของธุรกิจปัจจุบันบริษัทต่างๆจะพยายามหาเทคนิคต่างๆที่สามารถนำความสำเร็จมาสู่บริษัทเช่นธุรกิจขนาดย่อมจะสร้างความสัมพันธ์กับลูกค้า โดยสังเกตจากความต้องการ ความชอบ และความสนใจของลูกค้า และอาจมีการเรียนรู้ได้จากข้อมูลในอดีตว่าจะทำอย่างไรให้การบริการลูกค้ามีประสิทธิภาพดีขึ้นในอนาคต หรือ บริษัทที่เป็นผู้ออกบัตรเครดิตและธนาคารต่างๆ จะมีขบวนการที่ใช้คาด้าไมนิ่งให้เป็นประโยชน์ ในการช่วยตัดสินใจว่าลูกค้ากลุ่มใดเป็นกลุ่มที่ดี , ทำความเข้าใจลูกค้า , ช่วยในการแยกประเภทของลูกค้าและจะทำนายกลุ่มของประชากรที่คาดว่าจะมาเป็นลูกค้าในอนาคต เป็นต้น อย่างไรก็ตามการเรียนรู้นั้นต้องมากกว่าการเก็บสะสมข้อมูลโดยตรงไปตรงมา ซึ่งจะทำให้การทำงานไม่มีประสิทธิภาพ

การนำเทคโนโลยี คาด้าไมนิ่ง มาประยุกต์ใช้กับข้อมูลธุรกิจผู้พัฒนาจำเป็นต้องคำนึงถึงว่าสามารถนำเทคโนโลยีของ คาด้าไมนิ่ง มาประยุกต์ใช้กับธุรกิจได้อย่างไร การนำ คาด้าไมนิ่ง มาประยุกต์ใช้กับงานนั้นมีหลายรูปแบบ เช่น ธุรกิจด้านการสื่อสาร ได้แก่ การค้นหาลักษณะของกลุ่มลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการ หรือในธุรกิจด้านการตลาดแบบ Direct marketing ที่นำ คาด้าไมนิ่ง มาใช้ในการทำนายคุณลักษณะของลูกค้าที่มีแนวโน้มซื้อสินค้าประเภทต่างๆ ซึ่งจะทำให้สามารถวางแผนการทำตลาดของสินค้าไปที่กลุ่มลูกค้าที่ทำรายได้ให้สินค้ามากๆ ได้ดียิ่งขึ้น



รูปที่ 3.1 คาด้าไมนึ่ง ในกระบวนการของ Knowledge Discovery

เป้าหมายของการทำคาด้าไมนึ่ง มีแตกต่างกันไป เช่น Discovery, Predictive Modeling และ Forensic Analysis ซึ่งแต่ละแบบ มีข้อดีข้อด้อยและ ได้มาจากเทคนิคที่แตกต่างกัน จึงมีความจำเป็นที่จะต้องคัดเลือก เทคนิคที่จะนำมาใช้ในการวิเคราะห์ให้เหมาะสมกับข้อมูล



รูปที่ 3.2 เป้าหมาย คาด้าไมนึ่ง ในรูปแบบต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.1.1 ความเป็นมาของดาต้าไมนิ่ง

ดาต้าไมนิ่ง พัฒนามาจากเทคนิคทางสถิติและการเรียนรู้ของเครื่องกล (Machine-Learning) เพื่อสร้างตัวแบบสำหรับทำนายพฤติกรรมของลูกค้า ในปัจจุบันเทคโนโลยีทำให้กระบวนการดาต้าไมนิ่งเป็นไปอย่างอัตโนมัติ มีการรวมเข้ากับดาต้าแวร์เฮาส์ และนำเสนอผลลัพธ์ในหลาย ๆ ทางที่ผู้ใช้ต้องการ

### 3.1.2 ปัญหาของธุรกิจที่ทำให้มีการนำเอาดาต้าไมนิ่งมาประยุกต์ใช้

เมื่อมีการวิจัยและใช้เทคโนโลยีสารสนเทศก้าวหน้าขึ้น จึงทำให้เกิดการพัฒนาเรื่องของการตัดสินใจทางธุรกิจ โดยสิ่งแวดล้อมที่มีผลต่อการเปลี่ยนแปลงของธุรกิจซึ่งมีดังนี้

#### 3.1.2.1 รูปแบบพฤติกรรมของผู้บริโภค (Customer behavior patterns)

พฤติกรรมของผู้บริโภคมีการเปลี่ยนแปลงไป ซึ่งจะเกิดจากการปรับเปลี่ยนตามสภาพแวดล้อมหรือสภาพของเศรษฐกิจ อีกทั้งลูกค้ายังมีความต้องการและมีการรับรู้ข่าวสารได้รวดเร็วขึ้น จึงทำให้ผู้บริโภคซื้อสินค้าได้รวดเร็วตามที่ต้องการ

3.1.2.2 **สถานะตลาดที่อิ่มตัว (Market saturation)** เกิดขึ้นจากปริมาณของการเสนอขายสินค้ามีมากเกินไปกว่าความต้องการของผู้บริโภคเป็นผลให้โอกาสในการขายส่วนแบ่งทางการตลาดมีน้อยลง ทำให้คู่แข่งจำเป็นต้องปรับตัวเองเพื่อสนองตอบความต้องการของลูกค้าให้ได้มากที่สุด

3.1.2.3 **การเพิ่มขึ้นของสินค้า (Increased commoditization)** สินค้าและบริการหลายชนิดมีการเพิ่มที่แตกต่างจากสายของสินค้านั้นๆ เพราะว่ามีตลาดกลุ่มใหม่ๆ เกิดขึ้นตลอดเวลา เนื่องจากว่าพฤติกรรมของผู้บริโภคแต่ละกลุ่มมีความชอบที่แตกต่างกัน ทำให้เกิดพฤติกรรมการบริโภคแบบใหม่ซึ่งหลบซ่อนอยู่

3.1.2.4 **แนวโน้มทางการตลาดภายใต้สถานะแรงกดดัน (Traditional marketing approaches under pressure)** เกิดจากจำนวนข่าวสารที่เพิ่มมากขึ้นแต่การดำเนินการเกี่ยวกับฐานข้อมูล เริ่ม ไม่มีประสิทธิภาพขณะที่ความต้องการของลูกค้าเพิ่มมากขึ้น

3.1.2.5 **เวลา (Time To Market)** ถูกให้ความสำคัญมากขึ้นเพราะว่าคู่แข่งในตลาดมีมากขึ้น องค์กรแต่ละองค์กรจึงพยายามหาทางที่จะทำให้สินค้าไปสู่ผู้บริโภคเร็วที่สุดเท่าที่เป็นไปได้ ซึ่งหากว่าใครดำเนินการก่อนย่อมมีโอกาสมากกว่า

3.1.2.6 **วงจรชีวิตของผลิตภัณฑ์สั้นลง (Shorter product life cycle)** เนื่องจากผลิตภัณฑ์ต่างๆ สามารถหาซื้อจากตลาดได้อย่างรวดเร็วและพฤติกรรมของผู้บริโภคมีการ

เปลี่ยนแปลงอย่างรวดเร็วมาก ทำให้อายุของสินค้ามีช่วงชีวิตที่สั้นลง เป็นผลให้ผู้ใช้ต้องซื้อสินค้าชนิดนั้นๆ บ่อยขึ้นอีกทั้งผู้ขายก็มีเวลาในการทำกำไรน้อยลง

**3.1.2.7 การเพิ่มขึ้นของกลุ่มคู่แข่งและความเสี่ยงทางธุรกิจ (Increased competition and business risks)** มีมากขึ้น โดยเมื่อพิจารณาแล้วพบว่าแนวโน้มของสินค้าจะมีการซื้อขายกันทั่วโลกโดยผ่านทางอินเทอร์เน็ตทำให้ยากที่จะเก็บข้อมูลต่างๆ อีกทั้งแนวโน้มการเปลี่ยนแปลงของลูกค้ายิ่งเป็นไปอย่างรวดเร็วทำให้ธุรกิจเกิดความเสี่ยงอย่างไม่เคยเป็นมาก่อน

### 3.1.3 ความสามารถของ คาด้าไมนิ่ง

คาด้าไมนิ่ง เป็นกระบวนการในการค้นหาแนวโน้ม และรูปแบบของข้อมูลที่ซ่อนอยู่เพื่อสร้างความรู้ใหม่เกี่ยวกับข้อมูลนั้นๆ โดยใช้การวิเคราะห์ทางสถิติ และเทคนิคในการสร้างแบบจำลอง ที่วิธีการธรรมดาไม่สามารถทำได้ คาด้าไมนิ่ง นำเอาวิธีการสร้างแบบจำลอง (Model) มาช่วยในการค้นหารูปแบบและความสัมพันธ์ของข้อมูลแบบจำลอง เป็นเสมือนแผนที่ หรือแบบจำลองของสถานการณ์จริง ซึ่ง แบบจำลอง ที่ดีก็จะมีประโยชน์ในการทำ ความเข้าใจกับธุรกิจ และบอกได้ถึงสิ่งที่ควรปฏิบัติเพื่อทำให้เกิดความสำเร็จในธุรกิจ

เราสามารถจำแนกแบบจำลอง ได้เป็น 2 ประเภท ประเภทแรกคือ Predictive Model ซึ่งจะ ใช้ข้อมูลที่รู้ผลลัพธ์แล้วมาสร้างแบบจำลองที่สามารถนำไปใช้ในการพยากรณ์ค่าของข้อมูลที่แตกต่างออกไป เช่น เราสามารถสร้างแบบจำลองขึ้นโดยใช้ประวัติการชำระหนี้ของลูกค้าหนึ่งเพื่อสามารถระบุลักษณะของลูกค้าหนึ่งที่มีแนวโน้มจะเป็นหนี้ที่มีปัญหา ส่วนแบบจำลองประเภทที่สองคือ Descriptive Model ที่จะอธิบายรูปแบบ (Patterns) ของข้อมูลที่มีอยู่ ซึ่งสามารถนำไปใช้เป็นแนวทางในการตัดสินใจ

ความแตกต่างโดยพื้นฐานของแบบจำลองทั้งสองประเภทคือแบบจำลองประเภท Predictive Model แสดงผลการพยากรณ์อย่างชัดเจน (เช่นลักษณะของหนี้ที่มีปัญหา) ในขณะที่แบบจำลองประเภท Descriptive Model นำมาใช้เพื่อสร้าง Predictive Model อีกทีหนึ่ง หรือแสดงการพยากรณ์แบบไม่ชัดเจน

บริษัทที่เข้าใจถึงธุรกิจ และลูกค้า ก็เข้าใจถึงความสำคัญและมูลค่าของรูปแบบการเกิดข้อมูลที่พนักงานได้สังเกตเห็น ดังนั้นนอกจาก คาด้าไมนิ่งจะเป็นเครื่องมือที่ช่วยยืนยันในสิ่งที่ได้สังเกตเห็นและค้นพบแล้ว ยังสามารถช่วยในการค้นพบ รูปแบบ (Pattern) ของข้อมูลที่ซ่อนอยู่ได้อีก ด้วย ความรู้ใหม่ที่เกิดขึ้นก็สามารถเพิ่มอัตราผลตอบแทนได้อย่างชัดเจนและต่อเนื่อง

อย่างไรก็ตาม คาด้าไมนิง ก็ไม่ใช่เครื่องมือที่จะทำงานกับฐานข้อมูลและคอยสังเกตสิ่งที่เกิดขึ้นและแจ้งผู้ใช้เมื่อมีรูปแบบที่น่าสนใจ คาด้าไมนิง ยังต้องการความรู้ความเข้าใจเกี่ยวกับข้อมูลและธุรกิจ หรือต้องการความเข้าใจในเรื่องวิธีการวิเคราะห์ข้อมูล คาด้าไมนิง จะเป็นเครื่องมือที่จะช่วยให้ให้นักวิเคราะห์ธุรกิจค้นหารูปแบบ (Patterns) และความสัมพันธ์ระหว่างข้อมูล แต่จะไม่บอกถึงมูลค่าของ รูปแบบที่เกิดขึ้นต่อองค์กร

ถึงแม้ว่า เครื่องมือทางคาด้าไมนิงที่ดีจะช่วยให้ผู้ใช้ไม่จำเป็นต้องรู้ถึงเทคนิคทางสถิติ แต่ผู้ใช้ก็จำเป็นต้องเข้าใจในการทำงานของเครื่องมือที่เลือกใช้ และวิธีการที่เลือกใช้ด้วย อาทิเช่น ค่าติดตั้งต่างๆของเครื่องมือจะมีผลต่อความถูกต้อง และความเร็วในการทำงานของแบบจำลอง คาด้าไมนิง ไม่ได้มีขึ้นเพื่อทดแทนทักษะการทำงานของนักวิเคราะห์ธุรกิจ หรือผู้บริหาร แต่ คาด้าไมนิง เป็นเครื่องมือที่มีประสิทธิภาพที่จะช่วยเพิ่มคุณภาพของงานที่ทำอยู่

### 3.1.4 กระบวนการทำงานของคาด้าไมนิง

กระบวนการทำงานของ คาด้าไมนิง เป็น กระบวนการของการสร้างแบบจำลองของกลุ่มข้อมูล เพื่อสร้างความเข้าใจในแนวโน้ม รูปแบบ และความเกี่ยวข้องกันของกลุ่มข้อมูล เพื่อใช้ในวิเคราะห์ข้อมูลในแง่มุมต่างๆ โดยกระบวนการของ คาด้าไมนิง แบ่งเป็น ขั้นตอนต่างๆดังนี้

#### 3.1.4.1 กำหนดวัตถุประสงค์ทางธุรกิจ (Business Objectives Determination)

การกำหนดวัตถุประสงค์ทางธุรกิจและความต้องการทางธุรกิจที่จะเป็นตัวกำหนดทิศทางในการทำ คาด้าไมนิง ซึ่งจะเป็นการวิเคราะห์ทางธุรกิจและวิเคราะห์ข้อมูล เบื้องต้นว่ามีข้อมูลอะไรอยู่และต้องการอะไรจากข้อมูล มีขอบเขตของปัญหาแค่ไหน ซึ่งในขั้นตอนนี้ จะวิเคราะห์หาอัลกอริทึมและฐานข้อมูลเบื้องต้นที่สัมพันธ์กับวัตถุประสงค์ทางธุรกิจได้

#### 3.1.4.2 การเตรียมข้อมูล (Data preparation)

การเตรียมข้อมูลเป็นขั้นตอนที่ใช้เวลานานที่สุดประมาณ 60 % ของการทำคาด้าไมนิง โดยจะแบ่งการทำงานเป็น 3 ขั้นตอนย่อย ดังนี้

### 3.1.4.2.1 การเลือกข้อมูล (Data Selection)

ข้อมูลก่อนทำค้ำไ่มิ่งอาจจะมาจากข้อมูลหลายๆแหล่งรวมกันจึงต้องมีการเลือกข้อมูลที่สำคัญออกมา วัตถุประสงค์ของการเลือกข้อมูลคือ การระบุลักษณะข้อมูล , เลือกข้อมูลที่ต้องการ และนำข้อมูลที่ไม่ต้องการออกไป ซึ่งการเลือกข้อมูลจะขึ้นอยู่กับวัตถุประสงค์ของแต่ละธุรกิจ

### 3.1.4.2.2 การกลั่นกรองข้อมูล (Data preprocessing)

เมื่อเลือกข้อมูลที่ต้องการแล้วจะต้องกลั่นกรองข้อมูลเพื่อให้ข้อมูลที่เหมาะสมกับจะนำไปทำค้ำไ่มิ่ง โดยการตรวจสอบข้อมูลเหล่านั้น ว่า มีความถูกต้องหรือไม่ ซึ่งความผิดพลาดของข้อมูลอาจเกิดขึ้นได้ขณะที่มีการรวบรวมข้อมูลจากฐานข้อมูลหลายๆแหล่ง ซึ่งหากมีข้อมูลขาดหายไปก็จะตัดส่วนนั้นทิ้งไป หรือหาค่าเฉลี่ยเพื่อใส่ข้อมูลลงไปเป็นต้น

### 3.1.4.2.3 การแปลงข้อมูล (Data Transformation)

เป็นการแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับการทำค้ำไ่มิ่ง ซึ่งขั้นตอนนี้เป็นการพิจารณาถึงข้อมูลที่จะนำมาใช้ว่าเป็นข้อมูลประเภทใด เพราะบางอัลกอริทึมที่เราใช้จะรับข้อมูลประเภทหนึ่งแต่ข้อมูลที่นำเข้ามาอาจเป็นอีกประเภทหนึ่งก็ได้ จึงจำเป็นต้องมีการแปลงข้อมูลให้เหมาะสมกับอัลกอริทึมที่ใช้ เช่น Neural network รับข้อมูลที่เป็นตัวเลขแต่หากข้อมูลที่นำเข้ามาเป็นข้อมูลแบบหมวดหมู่ (Categorical) ก็ต้องทำการแปลงให้เป็นตัวเลขเสียก่อน

### 3.1.4.3 การทำค้ำไ่มิ่ง (Data Mining)

เป็นขั้นตอนการประมวลผลข้อมูลตามอัลกอริทึมที่กำหนดไว้ โดยถ้าผลลัพธ์ที่ได้จากขั้นตอนนี้ไม่เป็นไปอย่างที่คาดหวังก็อาจย้อนกลับไปทำขั้นตอนนี้การกลั่นกรองข้อมูลใหม่ได้ ในขั้นตอนนี้จะเกี่ยวข้องกับการเลือก อัลกอริทึมในการทำค้ำไ่มิ่งซึ่งจะต้องพิจารณาลักษณะของปัญหาเป็นหลัก เพราะในแต่ละปัญหาต้องเลือกใช้อัลกอริทึมที่เหมาะสมจึงจะได้ผลการวิเคราะห์ที่ถูกต้อง ซึ่งอาจจะใช้หลายอัลกอริทึมเพื่อเปรียบเทียบผลลัพธ์ได้

### 3.1.4.4 การวิเคราะห์ผลลัพธ์ที่ได้จากการทำค้ำไ่มิ่ง (Analysis of Results)

เมื่อเราได้รูปแบบของความสัมพันธ์หรือโมเดลแล้ว ขั้นตอนนี้จะเป็นการทดสอบโมเดลที่ได้กับชุดข้อมูลอีกชุดหนึ่งที่เรารวบรวมผลลัพธ์อยู่แล้ว ว่าเมื่อทดสอบกับโมเดลแล้ว

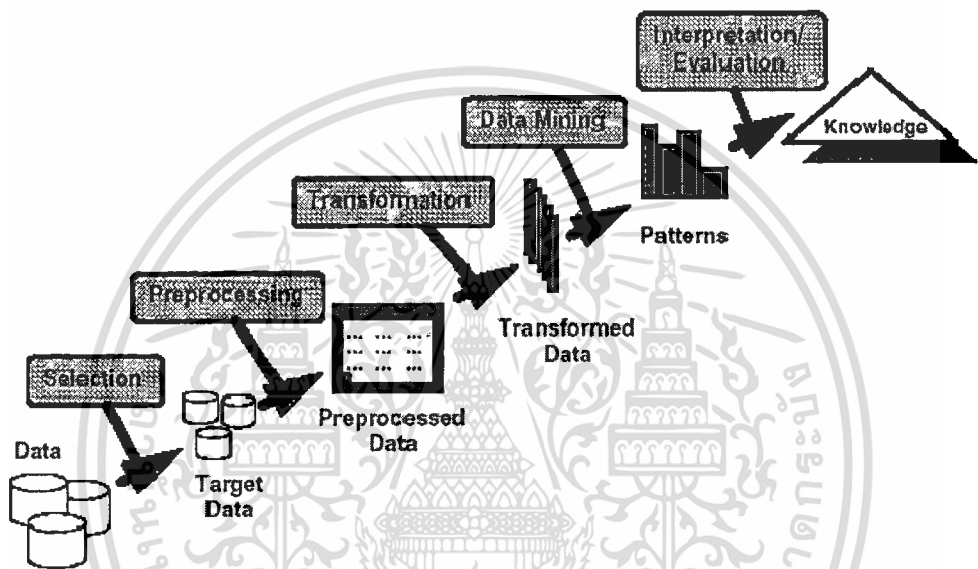
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับงานเพื่อการศึกษาเท่านั้น มิใช่ผูกขาดเนื้อหาเบ็ดเสร็จเช่นงานด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ได้ผลลัพธ์ถูกต้องหรือยอมรับได้หรือไม่ หากยอมรับไม่ได้ก็อาจจะแก้ไขโดยเพิ่มจำนวนข้อมูลให้มากขึ้น หรือเปลี่ยนอัลกอริทึมใหม่

### 3.1.4.5 การปรับความรู้ที่ได้เข้ากับธุรกิจ (Assimilation of Knowledge)

เป็นการรวบรวมความเข้าใจทางธุรกิจที่มีผลจากขั้นตอนการวิเคราะห์ผลลัพธ์เพื่อนำไปใช้ในโอกาสต่อไป



รูปที่ 3.3 กระบวนการทำ ดาต้าไมนิ่ง

### 3.1.5 ชนิดของงาน ดาต้าไมนิ่ง

ปัญหาต่างๆที่นำ ดาต้าไมนิ่ง มาช่วยแก้ไข มีกระบวนการแตกต่างกันออกไป ขึ้นกับลักษณะของปัญหา และวัตถุประสงค์ของการแก้ปัญหา การแก้ปัญหของงานชนิดต่างๆ โดยใช้วิธี ดาต้าไมนิ่งนั้น มีเทคนิคที่จะทำไปใช้ได้อย่างเหมาะสมแตกต่างกันไปในแต่ละงาน

#### 3.1.5.1 Association Rule Discovery

เป็นเทคนิคการค้นหความสัมพันธ์ของข้อมูลจาก ข้อมูลขนาดใหญ่ที่มีอยู่ เพื่อนำไปใช้ในการวิเคราะห์ทำนาย ปรากฏการณ์ต่างๆ หรือมาจากการวิเคราะห์การซื้อสินค้าของลูกค้าเรียกว่า Market Basket Analysis ซึ่งประเมิน จากข้อมูลในตารางที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้ จะเป็นคำตอบของปัญหา ซึ่งการวิเคราะห์แบบนี้เป็นการใช้ กฎความสัมพันธ์ (Association Rule) เพื่อหาความสัมพันธ์ของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้拿去ใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.1.5.2 Classification / Predictive

เป็นกระบวนการสร้าง Model จัดการข้อมูลให้อยู่ในกลุ่มที่กำหนด ตัวอย่างเช่น การแบ่งประเภทลูกค้าว่าเชื่อถือได้หรือไม่ โดยใช้ Model Construction ซึ่งเป็นการสร้าง Model โดยการเรียนรู้จากข้อมูลที่ได้กำหนดกลุ่มไว้เรียบร้อยแล้ว ซึ่ง Model ที่ได้อาจแสดงในรูปของ แบบต้นไม้ (Decision Tree) หรือ แบบนิวรอลเน็ตเวิร์ค (Neural Network)

### 3.1.5.3 Database Clustering /Segmentation

เป็นเทคนิคการลดขนาดของข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกัน ทำให้สามารถค้นหาข้อมูลที่ถูกกลบหายไปได้ เทคนิคนี้มัก ถูกใช้เป็นขั้นตอนเบื้องต้นในการทำดาต้าไมนิ่ง

### 3.1.5.4 Deviation Detection

เป็นกรรมวิธีในการหาค่าที่แตกต่างไปจากค่ามาตรฐาน หรือค่าที่คาดคิดไว้ว่าต่างไปมากน้อยเพียงใด โดยทั่วไปมักใช้วิธีการทางสถิติ หรือการแสดงให้เห็นภาพ สำหรับเทคนิคนี้ใช้ในการตรวจสอบ ลายเซ็นปลอม หรือบัตรเครดิตปลอม เป็นต้น

### 3.1.5.5 Sequential Analysis

ในการวิเคราะห์ลำดับ เพื่อค้นพบรูปแบบของการปรากฏของข้อมูล ซึ่งปรากฏในรายการที่แยกออกมา เช่นถ้าผู้ซื้อของซื้อสินค้า A แล้วเขาจะซื้อสินค้า B ในภายหลัง

## 3.1.6 ประโยชน์ของ ดาต้าไมนิ่ง

3.1.6.1 ผลลัพธ์ที่ได้อยู่ในรูปของตัวแบบที่ง่ายต่อการเข้าใจ ผู้ที่ไม่มีพื้นฐานทางสถิติก็สามารถแปรความจากตัวแบบได้ สามารถนำเสนอสารสนเทศที่ได้ไปใช้ในกระบวนการทางธุรกิจได้

3.1.6.2 สามารถวิเคราะห์ข้อมูลจำนวนมากได้

3.1.6.3 ดาต้าไมนิ่ง จะค้นพบในสิ่งที่เราคาดไม่ถึง เนื่องจากการที่มีตัวแบบและการตรวจสอบที่หลากหลาย จะพบว่าการค้นหาจากตัวแปรที่นำมารวมกัน ทำให้ได้สารสนเทศที่เป็นประโยชน์ต่อธุรกิจ

3.1.6.4 ตัวแปรไม่จำเป็นต้องถูกบันทึก ค่าค่าใดหนึ่ง สามารถรองรับได้ทั้งตัวแปรที่เป็นตัวเลขหรือเป็นกลุ่ม ตัวแปรเหล่านี้จะแสดงให้เห็นในตัวแบบโดยมีรูปแบบเช่นเดียวกับที่เก็บในฐานข้อมูล

3.1.6.5 ตัวแบบมีความถูกต้อง เนื่องจากการทดสอบด้วยเทคนิคทางสถิติ

3.1.6.6 ตัวแบบถูกสร้างได้อย่างรวดเร็ว ตัวแบบต่างๆ จะถูกทดสอบและเลือกตัวแบบที่ดีที่สุดสำหรับผู้ใช้

### 3.1.7 การประยุกต์ใช้ ค่าค่าใดหนึ่ง

จากประโยชน์ที่เห็นได้ชัดของ ค่าค่าใดหนึ่งในการควบคุมค่าใช้จ่าย และเพิ่มรายได้ให้กับองค์กร ทำให้ค่าค่าใดหนึ่ง ได้รับความนิยมนมากขึ้น ตัวอย่างเช่น หลายๆ บริษัทนำ ค่าค่าใดหนึ่งไปใช้ในการบริหารความสัมพันธ์กับลูกค้า โดยการหาคุณลักษณะของลูกค้าที่มีแนวโน้มจะผลจากบริษัทไปหาคู่แข่ง เพื่อที่บริษัทจะได้หาวิธีการที่จะดึงลูกค้ากลับมา เพราะการรักษาฐานลูกค้าเก่าเอาไว้จะประหยัดค่าใช้จ่ายมากกว่าเมื่อเทียบกับการหาลูกค้ารายใหม่

นอกจากนี้ยังมีการนำ ค่าค่าใดหนึ่ง ไปใช้กับฐานข้อมูลทางการตลาดในด้านการหากลุ่มลูกค้าที่น่าจะส่งจดหมายแนะนำสินค้า หรือนำเสนอรายการสินค้า เพื่อจะสามารถลดค่าใช้จ่าย และเพิ่มรายได้ที่จะเกิดจากการขายสินค้านั้นๆ ได้ การโฆษณาประชาสัมพันธ์สินค้ากับกลุ่มลูกค้าเป้าหมายจะได้ผลกลับมาดีกว่าจะเห็นได้ว่า ค่าค่าใดหนึ่ง สามารถสร้างคุณค่าให้กับหลายๆ ภาคธุรกิจ ซึ่งจะสามารถสรุปลักษณะการประยุกต์ใช้ ค่าค่าใดหนึ่ง ได้ดังนี้

#### 3.1.7.1 Direct Marketing

ค่าค่าใดหนึ่งมีบทบาทสำคัญในการทำการตลาดของธุรกิจขายตรงเพราะความสามารถในการทำนายพฤติกรรมลูกค้าเพื่อพิจารณาว่าลูกค้าลักษณะใดน่าจะมีการตอบรับที่ดีในการประชาสัมพันธ์สินค้าด้วยไปรษณีย์ หรือทำนายกลุ่มลูกค้าที่มีแนวโน้มจะซื้อสินค้าชนิดหนึ่ง ซึ่งจะทำให้การประชาสัมพันธ์สินค้าประเภทหนึ่งๆ ต่อกลุ่มลูกค้าเป้าหมายได้ผลมากขึ้น นอกจากนั้นยังช่วยลดค่าใช้จ่ายจากการไม่ต้องส่งใบโฆษณาไปยังลูกค้าทุกราย แต่จะส่งเฉพาะลูกค้าที่มีแนวโน้มจะตอบรับเท่านั้น

#### 3.1.7.2 Customer Acquisition

สำหรับการทำการตลาดของบริษัทขนาดเล็ก การดึงลูกค้าเข้ามาก็เป็นวิธีการหนึ่งที่เป็นที่ปฏิบัติกัน โดยค่าค่าใดหนึ่งถือได้ว่าเป็นเครื่องมือสำคัญในการจัดกลุ่มลูกค้าที่เป็นลูกค้าชั้นดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของบริษัท จัดกลุ่มของตลาด และเพิ่มประสิทธิภาพในโครงการการตลาดแบบ Direct marketing กิจกรรมเหล่านี้ในสมัยก่อนต้องอาศัยนักสถิติที่มีความชำนาญ หรือจ้างให้บริษัทที่เชี่ยวชาญเฉพาะทางทำ ซึ่งจะทำให้เกิดความล่าช้าในการเริ่มโครงการใหม่ๆ และไม่ทันต่อสภาพแวดล้อมที่มีการแข่งขันสูง อีกทั้งมีโครงการใหม่ๆ ออกสู่ตลาดได้น้อยกว่าความต้องการของธุรกิจ การนำค้ำไ่มิ่งมาพัฒนาเพื่อเป็นเครื่องมือให้นักการตลาดผลักดันให้เกิดโครงการใหม่ๆที่เหมาะสมกับกลุ่มลูกค้าได้มากขึ้น และยังทำให้การวิเคราะห์ผลตอบรับจากลูกค้าทำได้มีประสิทธิภาพ

### 3.1.7.3 Customer Retention

ในการทำธุรกิจ การรักษาฐานลูกค้าไว้ก็เป็นสิ่งสำคัญยิ่งกว่าการหาลูกค้ารายใหม่ เพราะลูกค้ารายเดิมมีโอกาสที่จะสร้างรายได้ให้กับบริษัทได้มากกว่า ประกอบกับต้นทุนในการหาลูกค้ารายใหม่จะสูงกว่า ตัวอย่างเช่น ในธุรกิจโทรคมนาคมทางบริษัทจะรู้ว่าลูกค้าจะไปจากบริษัทก็ต่อเมื่อลูกค้าโทรศัพท์เข้ามาขอยกเลิกบริการ ซึ่งเมื่อลูกค้าตั้งใจจะไปแล้วทางบริษัทไม่มีโอกาสที่จะชักชวนให้ลูกค้าเปลี่ยนใจ แต่เมื่อมีการนำ ค้ำไ่มิ่ง มาใช้เพื่อสร้าง Model ของรูปแบบการใช้บริการของลูกค้าในอดีตที่ได้ยกเลิกบริการ โดยการศึกษาประวัติการชำระเงิน ลักษณะของลูกค้า และข้อมูลอื่นๆ ซึ่ง model นี้จะสามารถนำมาใช้เพื่อทำนายว่าลูกค้าที่คิดในปัจจุบันจะมีแนวโน้มจะยกเลิกบริการในอนาคตหรือไม่ เมื่อนักการตลาดมีข้อมูลของลูกค้าในลักษณะนี้จะทำให้สามารถวางกลยุทธ์การตลาดเพื่อที่จะรักษาลูกค้าไว้ในช่วงที่เขายังเป็นลูกค้าอยู่ซึ่งจะได้ผลดีกว่าการดึงลูกค้าตอนที่เขายกเลิกบริการไปแล้ว

### 3.1.7.4 Cross Selling

การเพิ่มมูลค่าของลูกค้าก็เป็นงานการตลาดที่สำคัญประการหนึ่ง การเพิ่มมูลค่าของลูกค้าแตกต่างจากการเพิ่มส่วนแบ่งทางการตลาดตรงที่การเพิ่มส่วนแบ่งตลาดได้จากการเพิ่มจำนวนลูกค้า แต่การเพิ่มมูลค่าของลูกค้าคือการเพิ่มปริมาณการซื้อสินค้าของลูกค้าแต่ละราย ซึ่งในอดีตใช้วิธีการเคาะว่าในขณะที่ลูกค้าซื้อสินค้าชนิดหนึ่งๆ ควรจะนำเสนอสินค้าใดเพิ่ม แต่เมื่อมีการนำ ค้ำไ่มิ่ง มาใช้ทำให้นักการตลาดทราบว่าลูกค้ารายใดน่าจะซื้อสินค้าชนิดใหม่ของบริษัทด้วย หรือควรจะนำเสนอสินค้าชนิดใดไปพร้อมๆ กัน

### 3.1.7.5 Trend Analysis

การทำความเข้าใจแนวโน้มของตลาด มีความสำคัญในการลดต้นทุน และการนำสินค้าออกสู่ตลาดในช่วงเวลาที่เหมาะสม สถาบันการเงินต่างๆ ต้องการวิธีการที่จะรู้ถึงรูปแบบของการเปลี่ยนแปลงปริมาณเงินฝากและถอนอย่างรวดเร็ว ส่วนผู้ค้าปลีกก็ต้องการที่จะรู้ว่าสินค้าชนิดใดจะขายได้เมื่อวางขายด้วยกัน

### 3.1.8 การเปรียบเทียบ คาด้าไมนิ่ง กับเทคนิคอื่นๆ

#### 3.1.8.1 เปรียบเทียบ คาด้าไมนิ่ง กับ Data Warehouse

ในกระบวนการจัดทำ คาด้าไมนิ่ง นั้น ไม่จำเป็นต้องเข้าถึงข้อมูลใน Data Warehouse มาก่อน เพราะองค์กรธุรกิจที่มีการแข่งขันทางการตลาดสูงต้องมีความสามารถในการปรับตัวให้ทันต่อการเปลี่ยนแปลงอย่างรวดเร็วของแนวทางการตลาด เพื่อแก้ไขปัญหาดังกล่าวองค์กรจำเป็นต้องเข้าถึงข้อมูลทั้งหมดเพื่อนำมาใช้เป็นปัจจัยในการตัดสินใจ ทำให้ต้องสืบค้นข้อมูลในอดีตเพื่อนำมาวิเคราะห์และคาดเดาแนวโน้มในอนาคต

ข้อมูลจำนวนมากถูกจัดเก็บลงฐานข้อมูลขนาดใหญ่ วิธีที่ดีที่สุดในการเข้าถึงข้อมูลสำคัญที่นำมาใช้ในการตัดสินใจเหล่านี้คือการสร้าง Data Warehouse ขึ้นมา ในองค์กรธุรกิจส่วนใหญ่พบว่า มีฐานข้อมูลขนาดใหญ่สำหรับกระบวนการการทำงานสำหรับ Transaction รายวัน ฐานข้อมูลนี้มีชื่อเรียกว่า Operational Database ซึ่งโดยส่วนใหญ่ Operational Database ไม่ได้สร้างขึ้นเพื่อเก็บ Historical Data แต่สร้างขึ้นเพื่อใช้กับ Application สำหรับ Transactions จึงเหมาะกับงานประเภทรายวันที่ต้องการความรวดเร็วในการทำงาน และสามารถรองรับการทำงานของผู้ใช้จำนวนมากได้

ฐานข้อมูลที่พบในองค์กรอีกลักษณะหนึ่งคือ Data Warehouse ทำหน้าที่เก็บข้อมูลขนาดใหญ่แต่ข้อมูลนั้นสามารถแสดงถึงความหมายของข้อมูลจำนวนพันล้านรายการที่มีอยู่ ถูกออกแบบมาเพื่อสนับสนุนกลยุทธ์ทางการตลาดและการตัดสินใจของผู้บริหาร ข้อมูลที่จำเป็นสำหรับการตัดสินใจจะถูก Extract จาก Operational Database นอกจากนี้ Data Warehouse ขนาดเล็กที่เป็น Local Data Warehouse มีชื่อเรียกว่า Datamart

ในขณะที่ Data Warehouse ทำหน้าที่เก็บเฉพาะข้อมูลที่สำคัญ คาด้าไมนิ่ง กลับเป็นแนวทางที่สนับสนุนการตัดสินใจด้วยการนำข้อมูลจาก Data Warehouse มาสร้างกฎเกณฑ์ (rule) ทำให้เกิดแนวความคิดใหม่ ๆ และเป็นการทำนายอนาคต ตัวอย่างคำถามที่ผู้บริหารต้องการทราบต้องใช้กระบวนการ คาด้าไมนิ่ง เพื่อให้ได้คำตอบที่เหมาะสมที่สุด เช่น ผลิตภัณฑ์หรือบริการใดที่ควรจะผลิตถัดไปเพื่อรองรับความต้องการของลูกค้าก่อให้เกิดกำไรสูงสุด หรือ สาขาใหม่ของบริษัทควรตั้งที่ไหน คำตอบเหล่านี้เกี่ยวข้องกับข้อมูลที่มีอยู่แน่นอน แต่ต้องอาศัยคาด้าไมนิ่ง เพื่อเอ็กสารนเป็นเอ็กสารทงสวนวิสาหรับการใชงานเพื่การศึกษาเท่านั้น เมื่อนูญาตเ็นไปใชประโยชนดานการคาไมวากรณีใดทงทั้งสิ้น อีกทงห้ามมิใหัดดแปลงเนื้อหา และตองอ้างอิงถึงเจาของเอ็กสารทุกครั้งที่มีการนำไปใช้

ค้นหาคำตอบขึ้นมา ซึ่งแอปพลิเคชันของ เทคนิคดาต้าไมนิ่งสร้างขึ้นมาจาก 2 วิธี คือ โดยการใช้ Data Warehouse และ Extract ขึ้นจาก Data Warehouse

ฟังก์ชันหลักของ Data Warehouse คือ การให้ข้อมูลเพื่อการตัดสินใจ ในบางกรณีสามารถใช้ SQL Tool ทำงานก็เพียงพอแล้ว แต่ถ้าต้องการเปรียบเทียบข้อมูลจำนวนมากและไม่ทราบประเภทของข้อมูลที่ต้องการแน่ชัด หรือต้องการข้อมูลที่ไม่สามารถใช้ SQL สืบค้นได้โดยตรง แนวทางแก้ไขปัญหาลักษณะนี้คือ การใช้ เทคนิคดาต้าไมนิ่ง ด้วยการนำเอาข้อมูลจาก Data Warehouse มาวิเคราะห์ เรียนรู้ และเปลี่ยนเป็น Actionable Information โดยทั่วไปมีหลายวิธี

เทคนิคดาต้าไมนิ่ง ถูกนำไปใช้ในหลายลักษณะงานขึ้นอยู่กับความต้องการของผู้ใช้ เมื่อเข้าใจถึงความต้องการของผู้ใช้แล้วจึงสามารถสร้าง Data Warehouse ที่เหมาะสมสำหรับ ดาต้าไมนิ่ง ได้ โดยต้องแบ่งและจัดการทำงานของคอมพิวเตอร์เพื่อทำ ดาต้าไมนิ่ง เนื่องจากการดึงข้อมูลจาก Operational Database มาสร้างนั้นไม่เพียงพอและเป็น ไปไม่ได้ถึงแม้ว่าจะมีแอปพลิเคชัน มีชนิดข้อมูล (Data Type) และแอตทริบิวต์ (Attribute) ที่แตกต่างกัน โดยที่ข้อมูลทั้งหมดของการสร้าง Data Warehouse มาจาก Operational Database จึงอาจกล่าวได้ว่า การสร้าง ดาต้าไมนิ่ง ขึ้นมานั้นก็เพื่อเพิ่มความฉลาด (Intelligence) ให้แก่ Data Warehouse

ดังนั้น ดาต้าไมนิ่ง จึงเป็นกระบวนการการค้นหาความรู้จากฐานข้อมูลที่มีอยู่ ต้องอาศัยข้อมูลจำนวนมากและพิจารณาความสัมพันธ์ระหว่างข้อมูลเหล่านั้น โดยนำแนวความคิดและเทคนิคต่าง ๆ คือ การนำหลักสถิติมาประยุกต์ใช้กับ Artificial Intelligent, Database, Marketing หรือแนวความคิดอื่น เพื่อนำมาสร้างกฎและสมมติฐานเพื่อนำไปวิเคราะห์ให้เกิดประโยชน์ต่อไป ดาต้าไมนิ่ง ถูกนำไปประยุกต์ใช้งานได้หลากหลายยิ่ง เช่น ทางการแพทย์ ห้างสรรพสินค้า

### 3.1.8.2 เปรียบเทียบ KDD และ ดาต้าไมนิ่ง

KDD (Knowledge Discovery in Database) เป็นกระบวนการที่สำคัญของการค้นหาความรู้ (Knowledge) ใหม่ที่เป็นประโยชน์จากข้อมูลที่มีอยู่ ความรู้ที่จะระบุขึ้นมานั้น จะนำมาใช้เพื่ออธิบายข้อมูลเก่าที่มีอยู่ และช่วยในการตัดสินใจ หรือแยกประเภทข้อมูลใหม่ที่เข้ามา รวมทั้งช่วยในการสรุป content ของฐานข้อมูลขนาดใหญ่ เพื่อช่วยให้การตัดสินใจทำได้ง่ายขึ้น

มีความสับสนเกิดขึ้นระหว่างความหมายของ ดาต้าไมนิ่ง กับ KDD ว่าเหมือนกันหรือไม่ ในการประชุมนานาชาติเกี่ยวกับเรื่อง KDD ครั้งแรกที่ Montreal ในปี 1995 มีข้อเสนอว่าในส่วน ของ KDD นั้นใช้ในการอธิบายกระบวนการทั้งหมดของการ Extract ความรู้จากข้อมูล ในที่นี้ ความรู้หมายถึง Pattern และความสัมพันธ์ระหว่าง Data element ส่วน ดาต้าไมนิ่ง นั้น ควรใช้ สำหรับขั้นของการ Discovery ในกระบวนการของ KDD (KDD Process) ด้วยเหตุที่ KDD Process

เกี่ยวข้องกับการวิเคราะห์และตีความ Pattern จึงถูกดึงออกมาได้ด้วย คำค้นหาใด ๆ เพื่อตัดสิน Subset ของ Pattern เป็นองค์ประกอบของความรู้ KDD Process ประกอบไปด้วย 6 ขั้นตอนหลักๆ ดังนี้

**3.1.8.2.1 Data Selection** เป็นขั้นที่ทำการเลือก Operational Data ออกมาจากฐานข้อมูลที่มีขนาดใหญ่และกระจัดกระจาย โดยที่คัดลอกของ Operational Data ถูกแยกเก็บในฐานข้อมูลต่างหาก

**3.1.8.2.2 Cleaning** เป็นขั้นตอนการทำความสะอาดข้อมูลที่ผิดพลาดจาก Operational Data เช่น ข้อมูลไม่สมบูรณ์ มีข้อมูลซ้ำกัน ข้อมูลผิด ขั้นตอนนี้จึงเป็นการแก้ไขข้อมูลให้ถูกต้อง

**3.1.8.2.3 Enrichment** เป็นขั้นตอนที่ทำให้ข้อมูลที่เลือกมา มีความสมบูรณ์มากขึ้น โดยอาจหาข้อมูลมาจากภายนอก ในฟิลด์ (Field) ที่น่าจะเป็นประโยชน์ในการค้นหาความรู้ (knowledge)

**3.1.8.2.4 Coding** ทำการเลือกเรคอร์ดจากข้อมูลในขั้น Enrichment ที่มีความสมบูรณ์ครบถ้วนเพียงพอที่จะนำมาใช้ในการค้นหาความรู้ตามที่ต้องการ เช่น การเลือกเรคอร์ดที่มีข้อมูลอยู่ครบทุกฟิลด์ เป็นต้น

**3.1.8.2.5 คำค้นหาใด ๆ** ถือเป็น Discovery Stage ของ KDD Process ซึ่งใช้ในการค้นหาความรู้ การทำ คำค้นหาใด ๆ นั้นมีหลายเทคนิควิธี แต่ละเทคนิคจะใช้กับจุดประสงค์ที่แตกต่างกัน ตัวอย่างเทคนิคของการ คำค้นหาใด ๆ เช่น Query Tools, Statistical Techniques, Visualization, Online Analytical Processing (OLAP), Case - Based learning ( k - Nearest Neighbor), Neural Networks, Genetic Algorithms, Association Rules, Classification Rules เป็นต้น

**3.1.8.2.6 Reporting** เป็นขั้นตอนของการรายงานผลจากการทำ คำค้นหาใด ๆ ซึ่งโดยทั่วไปสามารถแสดงได้เป็นข้อความหรือเป็นภาพกราฟฟิก

## บทที่ 4

### เทคนิค Classification และอัลกอริทึม Decision Tree

เทคนิค Classification หรือเทคนิคในส่วนของการแบ่งระดับ เป็นการจัดกลุ่มข้อมูลจากหมวดหมู่ของข้อมูลที่กำหนดไว้และใช้ทำนายค่าของข้อมูลที่ต้องการทราบว่าอยู่กลุ่มใด เป็นเทคนิคการทำนายว่าสิ่งที่เราสนใจจะอยู่ในประเภทไหน หรือกลุ่มไหน โดยจะเห็นว่าการแบ่งระดับเป็นการจำแนกคลาสของเรคคอร์ดในข้อมูล ซึ่งคลาสต้องเป็นเซตของความเป็นไปได้ที่มีค่าแน่นอน โดยมีการกำหนดค่าของคลาสไว้ก่อนล่วงหน้าแล้ว แบบจำลองจะถูกสร้างขึ้นจาก Training Set ที่มีและจากค่าใน Attributes ของข้อมูล โดยโมเดลที่ใช้ในการทำนาย ซึ่งใช้ข้อมูลในอดีตเพื่อสร้างโมเดลขึ้นมา

#### 4.1 Classification

เทคนิค Classification อยู่บนพื้นฐานของการเรียนรู้แบบมีเป้าหมาย (Supervised Learning) คือต้องมีการกำหนดรูปแบบของอินพุตและเอาต์พุตมาก่อนจากข้อมูลตัวอย่างที่เรียกว่า Training Set หรือคือกลุ่มของข้อมูลที่จัดแบ่งไว้ส่วนหนึ่งเพื่อนำมาใช้ในการสร้างแบบจำลองนั่นเอง โดยเปรียบเทียบกับการเรียนรู้ของมนุษย์คือ เป็นการเรียนรู้ถึงแบบแผนของข้อมูลจากกลุ่มของข้อมูลที่จัดเตรียมไว้ ซึ่งโมเดลที่ได้มาจากการ Training สามารถแสดงให้เห็นได้ทั้งแบบดิซิชันทรี หรือแบบกฎ IF THEN โดยดิซิชันทรี เป็นการจำแนกค่าของ Attributes เพื่อจัดเรียงเป็นแบบต้นไม้ จากโหนดที่อยู่บนสุดของต้นไม้ (Root Node) ไล่ไปจนถึงโหนดที่อยู่ต่ำสุดของต้นไม้ (Leaf Node) โดยแต่ละโหนดในต้นไม้จะทำการแตกกิ่งออกตามค่าต่างๆที่เป็นไปได้ของ Attribute นั้น เพื่อสร้างลำดับถัดไปของต้นไม้ กระบวนการนี้จะถูกทำซ้ำในต้นไม้ย่อยๆจนกว่าจะถึงโหนดสุดท้ายของต้นไม้ โดย Classification มีเทคนิคที่ใช้ในการทำงานอยู่หลายเทคนิค แต่ในที่นี้จะขอกล่าวถึงรายละเอียดในส่วน of เทคนิค ทรีอินดักชัน (Tree Induction) เท่านั้น

#### 4.2 Tree Induction

เป็นการสร้างแบบจำลองในรูปแบบของ Decision Tree โดยส่วนใหญ่จะอยู่ในรูปแบบของ Binary Tree นั่นคือ เป็น Tree ที่มีทางเลือกเป็น 2 ทางเลือกเพื่อความสะดวกและง่ายในการพิจารณาจุดที่ใช้กำหนดการตัดสินใจเลือกทางเลือกเรียกว่า Nodes โดยมี Root Node เป็นปัจจัยสำคัญที่สุดที่มีผลต่อการตัดสินใจ และจุดสุดท้ายที่ข้อมูลเลือกทางเลือกในการตัดสินใจเรียกว่า Leaves

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.2.1 หลักการทำงานของ Tree Induction

4.2.1.1 กำหนดตัวแปร (Attribute) ที่มีผลต่อการกำหนดกลุ่ม (Classification) ที่สำคัญที่สุดในการทำนายเป็น Root Node

4.2.1.2 เมื่อเลือกตัวแปร (Attribute) ที่สำคัญได้ ก็ทำการแตกกิ่งจาก Root Node จากการใช้ค่าของตัวแปรที่กำหนดนั้น Node ที่ได้จากการแตกกิ่งเรียกว่า Child Node

4.2.1.3 หลังจากนั้น กระบวนการทำงานจะซ้ำเดิมสำหรับแต่ละ Node ที่แตกออกมา

โดยหลักการทำงานพื้นฐานของ Tree Induction มีวิธีการทำงานโดยการกำหนดตัวแปร หรือ Attribute ที่มีค่าความสำคัญสูงสุด ที่มีผลกระทบในการกำหนดการจัดกลุ่มเพื่อนำมาเป็น Root Node จากนั้นทางเลือกลงจาก Root Node จะเป็นการกำหนดเป็น Node ต่อๆ ไป สำหรับค่าที่เป็นไปได้ในการเลือกทางเลือกลงไป กระบวนการทั้งหมดจะถูกกระทำซ้ำๆ บน Training Data ที่คัดเลือกไว้ ปัญหาที่เกิดขึ้น คือจำนวน Node ที่เหมาะสมควรจะเป็นกี่ Nodes และการพิจารณาเลือก Attribute ที่ใช้กำหนดเป็น Root Node ที่ต่างกันจะมีผลกระทบต่อแบบจำลองหรือไม่ อย่างไรก็ตาม Tree Induction เป็นเทคนิคหนึ่งที่มีประสิทธิภาพในแง่ของการประหยัดเวลาที่ใช้ในการ Process และ วิเคราะห์ผลลัพธ์

#### 4.2.2 ข้อดีและข้อเสียของ Tree Induction

##### 4.2.2.1 ข้อดีของ Tree Induction

- มีประสิทธิภาพทางด้านเวลาในการประมวลผล ในช่วงหลังจากการทำแบ่งระดับ (Classification) แล้ว
- การแตกตัวของต้นไม้ (Tree) มีลักษณะที่เข้าใจได้ง่ายเมื่อสร้างเป็นกฎ ถ้าเทียบกับนอร์อล อินดักชัน (neural induction) ที่บางครั้งก็ไม่สามารถรู้ได้เลยว่าให้คำตอบมาได้อย่างไร และสามารถบอกได้ว่าปัจจัยใดที่มีอิทธิพลต่อการทำนายมากที่สุด

##### 4.2.2.2 ข้อเสียของ Tree Induction

- การสร้างต้นไม้ (Tree) เป็นเรื่องยากที่จะการคำนวณและมีความซับซ้อนมาก
- ยิ่งข้อมูลน้อยๆ ความผิดพลาดก็จะสูงขึ้น ซึ่งจะเห็นได้จากว่าเมื่อทำการแตกต้นไม้ (Tree) ไปเรื่อยๆ จำนวนของข้อมูลจะลดลงเรื่อยๆ ซึ่งจำนวนน้อยๆจะเป็นตัวแทนทางสถิติได้ยาก โดยยิ่งทำให้มี Level มากขึ้นความน่าเชื่อถือยิ่งน้อยลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

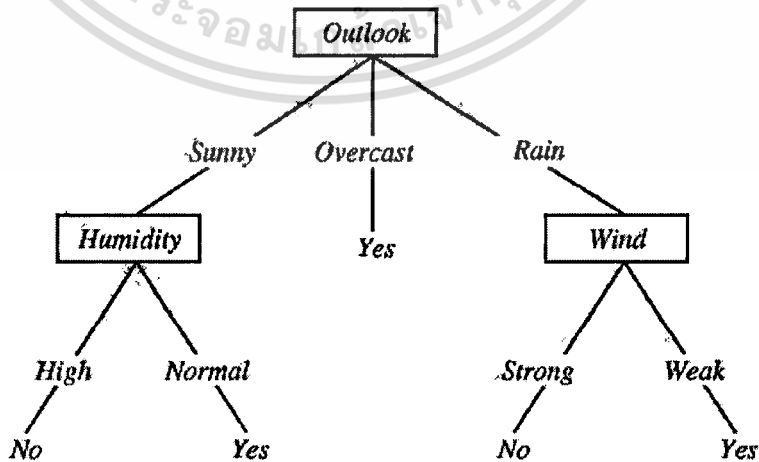
### 4.3 วิธีการ Decision Tree

Decision Tree เป็นโครงสร้างที่สามารถมองได้ในรูปแบบแผนภูมิต้นไม้ โดยแต่ละกิ่งของ โหนดแสดงให้เห็นถึงทางเลือกระหว่างจำนวนของทางเลือก และในส่วนใหญ่ของต้นไม้ หรือ โหนดที่เป็นใบ (Leaf Node) แสดงให้เห็นถึงการจำแนกพวกหรือการตัดสินใจนั่นเอง

Decision Tree เป็นที่นิยมกันมากเนื่องจากเป็นลักษณะที่คนจำนวนมากคุ้นเคย และเข้าใจ ได้ง่าย มีลักษณะเหมือนแผนภูมิองค์กร โดยที่แต่ละ โหนดแสดง Attributes แต่ละกิ่งแสดงผลในการ ทดสอบ และลิฟโหนดแสดงคลาสที่กำหนดไว้

สมมติว่าบริษัทขนาดใหญ่แห่งหนึ่ง ทำธุรกิจอสังหาริมทรัพย์มีสำนักงานสาขาอยู่ประมาณ 50 แห่ง แต่ละสาขามีพนักงานประจำ เป็นผู้จัดการและพนักงานขาย พนักงานเหล่านี้แต่ละคนจะ ดูแลอาคารต่าง ๆ หลายแห่งรวมทั้งลูกค้าจำนวนมาก บริษัทจำเป็นต้องใช้ระบบฐานข้อมูลที่กำหนด ความสัมพันธ์ระหว่างองค์ประกอบเหล่านี้ เมื่อรวบรวมข้อมูลแบ่งเป็นตารางพื้นฐานต่าง ๆ เช่น ข้อมูลสำนักงานสาขา ข้อมูลพนักงาน ข้อมูลทรัพย์สิน และข้อมูลลูกค้า พร้อมทั้งกำหนด ความสัมพันธ์ของข้อมูลเหล่านี้ เช่น ประวัติการเช่าบ้านของลูกค้า รายการให้เช่า รายการขาย สินทรัพย์ เป็นต้น ต่อมาเมื่อมีประชุมกรรมการผู้บริหารของบริษัท ส่วนหนึ่งของรายงานจาก ฐานข้อมูลสรุปว่า

“40% ของลูกค้าที่เช่าบ้านนานกว่าสองปี และมีอายุเกิน 25 ปี จะซื้อบ้านเป็นของตนเอง โดยกรณีเช่นนี้เกิดขึ้นกับ35% ของลูกค้าผู้เช่าบ้านของบริษัท” เป็นต้น



รูปที่ 4.1 ตัวอย่าง: Decision tree for the concept Play Tennis


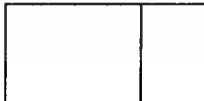

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยอัลกอริทึมที่ใช้ในโครงการนี้ คืออัลกอริทึม ID3 โดย ID3 เป็นอัลกอริทึมพื้นฐานของ Decision Tree ซึ่งเป็นการเรียนรู้แบบมีเป้าหมาย (Supervised Learning) เป็นการสร้างกฎในการแบ่งระดับ (Classification Rules) ในรูปแบบของ Decision Tree ซึ่งจะใช้เซตของการเทรนนิ่ง (Training Set) เป็นอินพุตและทำการสร้าง Decision Tree ขึ้นมา โดยการแบ่งส่วนของ เซตของการเทรนนิ่ง ซึ่ง Attribute จะถูกเลือกใช้เพื่อที่จะเป็นตัวแบ่งเซตของข้อมูลโดยอาศัยหลักของ Entropy และ Information Gain โดย Tree จะถูกสร้างตามค่าย่อยๆของ Attribute นั้น ทำเช่นนี้จนกระทั่งสมาชิกทุกตัวมีคลาสเดียวกัน

### 4.3.1 Entropy

Entropy คือกระบวนการที่จะใช้ค้นหาระดับของความบริสุทธิ์ ที่สร้างโดยตัวพรรณา (Descriptor) เพื่อสะท้อนการกระจายของตัวแบ่งแยก (Classifier) Entropy เป็นแนวคิดที่จะนำมาใช้เพื่อจัดเรียง Descriptors ต่างๆ ตามระดับนัยสำคัญ (Significant) ของ Descriptor หรืออธิบายได้อีกอย่างหนึ่งคือ Entropy จะเป็นตัวที่จะบ่งบอกถึงระดับของความไม่แน่นอนของ ตัว Descriptor นั้นเอง

#### Entropy and Randomness

<p><b>Most Random</b> Entropy = 1.0</p>		<p>A set with equal numbers of each type</p>
<p><b>Somewhat Random</b> 1.0 &gt; Entropy &gt; 0.0</p>		<p>A set of an inproportionate number of one type</p>
<p><b>Not Random</b> Entropy = 0.0</p>		<p>A set of all of the same type</p>

รูปที่ 4.2 ความสัมพันธ์ของการสุ่มข้อมูล และ ค่า Entropy

จากแนวความคิดของการสุ่มข้อมูล ค่า Entropy ถูกนำมาอ้างอิงโดยค่า Entropy จะแสดงถึงความเป็นชุดข้อมูลที่เหมือนกัน จากรูป กลุ่มตัวอย่างที่ประกอบด้วยชุดของ 2 ชนิดที่เป็นไปได้ ชุดแบบสุ่มมากที่สุดคือชุดที่มีความเป็นไปได้ของข้อมูลสองชนิดเท่ากันทั้ง 2 ชนิด คือชุดข้อมูลที่มีค่า Entropy = 1 ในขณะที่ ชุดข้อมูลอีกชุดหนึ่งที่มีความเป็นไปได้ของข้อมูลเพียงชนิดเดียว หรือไม่ว่าสุ่มอย่างไรก็ได้ข้อมูลแบบเดียวกันเสมอชุดข้อมูลชุดนี้มีค่า Entropy = 0 ชุดข้อมูลนี้เรียกอีกอย่างหนึ่งว่า ชุดข้อมูลที่มีโครงสร้าง (Structured Set)

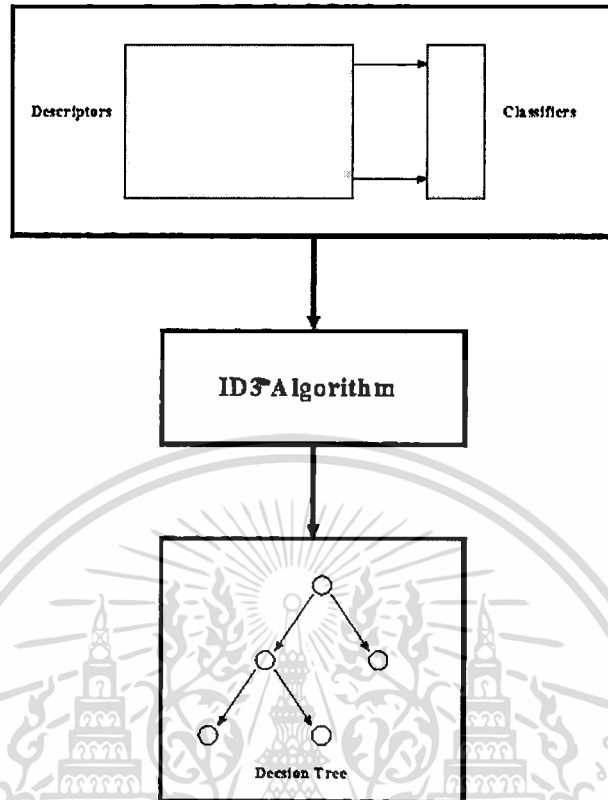
#### 4.3.2 Information Gain Measure

เป็นวิธีที่ใช้ในการวัดการปรับปรุงค่า Entropy หลังการแบ่งกับก่อนการแบ่ง เพื่อคำนวณหา ค่า Information Gain ที่สูงที่สุดไปสร้าง Decision Tree

$$\text{Information Gain} = \text{Entropy ก่อน Split} - \text{Entropy หลัง Split}$$

#### 4.3.3 ID3 Algorithm

อัลกอริทึม ID3 เป็นอัลกอริทึมที่มีเป้าหมายในการสร้าง Decision Tree จากข้อมูลที่ให้คำอธิบายที่มีประสิทธิภาพของการแยกประเภท (Classify) บนพื้นฐานของ Entropy ซึ่งจะทำการสร้าง Decision Tree ที่สามารถนำไปใช้เพื่อทำนายหรือแยกประเภทของข้อมูล โดยอาศัยแนวความคิด ของ Entropy มาช่วยในการพิจารณาพฤติกรรมของ Classifier



รูปที่ 4.3 แสดงถึงแนวคิดของอัลกอริทึม ID3

ID3 เป็นอัลกอริทึมพื้นฐานโดยที่ ชุดตัวอย่าง (Examples) คือเซตของข้อมูลที่ใช้ในการเรียนรู้ (Training Example) แอททริบิวเป้าหมาย (Target Attribute) คือ Attribute ที่นำค่าไปใช้ในการทำนายผลในต้นไม้ และ แอททริบิว (Attributes) คือ Attribute อื่นๆที่ใช้ในการสร้างโหนดในต้นไม้ และไม่ใช่ Target Attribute ซึ่งลักษณะของอัลกอริทึมมีดังนี้

### ID3 (Examples, Target\_Attribute, Attributes)

- Create a root node for the tree
- IF all examples are positive, Return the single-node tree Root, with label = +
- If all examples are negative, Return the single-node tree Root, with label = -
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples
- Otherwise Begin
  - $A \leftarrow$  The Attribute that best classifies examples
  - Decision Tree attribute for Root  $\leftarrow A$
  - For each positive value,  $v_i$ , of  $A$ ,
    - Add a new tree branch below Root, corresponding to the test  $A = v_i$
    - Let  $\text{Examples}(v_i)$ , be the subset of examples that have the value  $v_i$  for  $A$
    - If  $\text{Examples}(v_i)$  is empty
      - Then below this new branch add a leaf node with label = most common target value in the examples
      - Else below this new branch add the subtree  $\text{ID3}(\text{Examples}(v_i), \text{Target\_Attribute}, \text{Attributes} - \{A\})$
- End
- Return Root

ซึ่ง ID3 เป็นอัลกอริทึมที่จะสร้าง Decision Tree โดยจะมีการนำตัวอย่างเป็นข้อมูลที่จะเข้าไปทดสอบเพื่อหา Attribute มาสร้างเป็น โหนดใน Decision Tree โดยการทำงานของอัลกอริทึม ID3 มีขั้นตอนต่างๆดังนี้

1. ทำการทดสอบว่าตัวอย่างทุกตัวที่รับเข้ามามีค่าของ Target Attribute เป็นบวกทั้งหมดหรือไม่ ถ้าใช่ก็จะคืน โหนดที่มีค่าเป็นบวกให้ แล้วจบการทำงาน แต่ถ้าไม่ใช่ก็ทำต่อไป
2. ทำการทดสอบว่าตัวอย่างทุกตัวที่รับเข้ามามีค่าของ แอททริบิวเป้าหมาย (Target Attribute) เป็นลบทั้งหมดหรือไม่ ถ้าใช่ก็จะคืน โหนดที่มีค่าเป็นลบให้ แล้วจบการทำงาน แต่ถ้าไม่ใช่ก็ทำต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ทำการทดสอบว่าแอททริบิวต์อื่นๆที่ไม่ใช่ แอททริบิวต์เป้าหมาย (Target Attribute) มีค่าหรือ ไม่ ถ้าไม่มีค่าก็จบการทำงาน แต่ถ้ายังมีค่าอยู่ก็ทำกระบวนการดังนี้

1. หาค่า Information Gain ของทุกๆ แอททริบิวต์ ที่ไม่ใช่แอททริบิวต์เป้าหมาย ซึ่ง Attribute ใดมีค่า Gain สูงสุดจะถูกกำหนดให้เป็น โหนดบนสุด (Root Node) ตามสมการดังนี้

$$\text{Gain}(S, A) = \text{Entropy}(S) - S \left( \frac{|S_v|}{|S|} * \text{Entropy}(S_v) \right) \text{ -----(1)}$$

Gain(S,A) คือค่า Information Gain ของ A

S คือค่าแต่ละค่าที่เป็นไปได้ ของ Attribute A

S<sub>v</sub> คือตัวอย่างทั้งหมดของค่าที่เป็นไปได้หนึ่งๆของ Attribute A ที่มีค่า v

|S<sub>v</sub>| จำนวน Elements ใน S<sub>v</sub>

|S| จำนวน Elements ใน S

Entropy(S) คือ Entropy ของการรวบรวม S แบบเคม

Entropy(S<sub>v</sub>) คือ ค่าที่ถูกคาดหว้งของ Entropy หลังจาก S ถูกแบ่งด้วยการใช้ Attribute A

จากสมการ (1) สามารถหาค่า Entropy ได้จาก สมการ(2)

$$\text{Entropy}(S) = -p(+) \log_2 p(+) - p(-) \log_2 p(-) \text{ -----(2)}$$

โดย

p(+) คืออัตราส่วนของตัวอย่างที่เป็นบวกใน S

p(-) คืออัตราส่วนของตัวอย่างที่เป็นลบใน S

ถ้าค่า Target Attribute สามารถมีได้ c ค่าแล้ว Entropy ของ S จะเป็นดังนี้

$$\text{Entropy}(S) = \sum -p(I) \log_2 p(I) \text{ ----- (3)}$$

โดยที่ p(I) คือความสัมพันธ์ของ S ที่คลาส I จะเห็นได้ว่าถ้า Target Attribute สามารถรับค่าได้ c ค่าแล้ว Entropy ก็จะทำกับ log<sub>2</sub>c

2. กำหนดให้ A เท่ากับ Attribute ที่มีค่า Information Gain สูงสุดแล้วให้ A เป็น โหนดบนสุด (Root Node)

3. ทำการรวมรูป ค่าต่างๆทั้งหมดที่เป็นไปได้ของ A โดยให้ v แทนค่าของ A ตัวที่ 1 ถึง I ตามขั้นตอนด้านล่างนี้ จนกว่าจะหมดแล้วทำการเลือกชุดตัวอย่าง ของ Attribute A ที่มีค่าเท่ากับ v<sub>i</sub> มา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ทำตั้งแต่ ข้อ 1 ขึ้นจนกว่าจะสร้างต้นไม้เสร็จ โดยทำการส่งค่าพารามิเตอร์ต่างๆเข้าไป ดังนี้
1. ค่าตัวอย่างทั้งหมดที่เลือกมา
  2. ค่า Target Attribute
  3. ค่าของ Attribute ที่ลบเอา Attribute A ออก

สรุปขั้นตอนการทำงานโดย เริ่มจากให้อัลกอริทึมทำงานบนเซตของเรคคอร์ดที่ใช้ในการเทรนนิ่งซึ่งในที่นี้คือ S แล้วทำการตรวจสอบเงื่อนไขว่าถ้าเซตของเรคคอร์ดทั้งหมด (All Instances) ใน S เป็น คลาส C ก็ทำการสร้าง โหนด C แล้วหยุด แต่ถ้ามีคลาสอื่นปนอยู่ด้วยก็ต้องทำการเลือก Attribute A และสร้าง โหนดการตัดสินใจ (Decision Node) ขึ้นมา แล้วทำการแบ่งส่วนเรคคอร์ดที่ใช้ในการเทรนนิ่งใน S ตามค่า V ของ Attribute A จากนั้นก็ทำซ้ำไปซ้ำมาในแต่ละเซตย่อยของ  $S_v$  ต่อไป

ในการเทรนนิ่ง จะทำการแตกกิ่งไปเรื่อยๆ จนถึงข้อมูลตัวสุดท้าย โดยที่ต้นไม้ จะทำการแตกกิ่งไปเรื่อยๆ แต่ในความเป็นจริงแล้วไม่ได้ง่ายอย่างที่คิดไว้ เพราะจะต้องพบกับอุปสรรคมากมายในระหว่างการสร้างต้นไม้ โดยเฉพาะอย่างยิ่งเมื่อต้องการให้ความถูกต้องสูงขึ้นก็ต้องใช้ข้อมูลจำนวนมากขึ้นในการเทรนนิ่ง แต่ถ้าเป็นกรณีของการทดสอบข้อมูล ความถูกต้องจะเพิ่มขึ้นถึงจุดๆหนึ่งเท่านั้นแล้วถ้ามีการแตกกิ่งออกไปเรื่อยๆจะทำให้ความถูกต้องลดลง ซึ่งจุดที่ความถูกต้องถึงจุดสูงสุดของการทดสอบข้อมูล (Test Data) นี้เรียกว่า จุด โอเวอร์ฟิต (Over fit) จึงเป็นผลให้เกิดการ Pruning Tree (การแต่งกิ่ง หรือการตัดกิ่ง) และสาเหตุที่ทำให้เกิด Over fit อีกอย่างหนึ่งคือ เมื่อข้อมูลที่ใช้มีสิ่งรบกวน (noise) อยู่ หรือเมื่อข้อมูลที่ใช้ในการ Training มีจำนวนน้อยเกินกว่าที่จะสร้าง Tree ได้ ซึ่งสิ่งเหล่านี้ทำให้ อัลกอริทึมผลัดคิด Tree ที่เกิดการโอเวอร์ฟิตได้ ซึ่ง ณ จุดที่เกิดการ โอเวอร์ฟิต ของข้อมูลนั้นจะมีผลทำให้ความถูกต้องของโมเดลที่ได้ลดลง ดังนั้นทางที่ดีจึงควรใช้ข้อมูลจำนวนมากในการทำการเทรนนิ่ง เพื่อที่จะทำให้ได้โมเดลที่ครอบคลุมมากที่สุด และใช้ข้อมูลจำนวนน้อยๆในการทดสอบ

#### 4.3.4 การหลีกเลี่ยงการเกิด โอเวอร์ฟิต (Overfit) ในข้อมูล

การหลีกเลี่ยงการเกิด โอเวอร์ฟิต (Overfit) สามารถทำได้ 2 แบบคือ

**แบบที่ 1** หยุดต้นไม้ (Tree) ก่อนที่ ต้นไม้จะแตกไปเรื่อยๆจนเกิด โอเวอร์ฟิต (Overfit) คือ ในระหว่างการแตกต้นไม้ ก็ทำการทดสอบด้วยว่ามันเกิด โอเวอร์ฟิต (Overfit) หรือไม่ไปด้วย ถ้าผลที่ได้จากการทดสอบทำให้ความถูกต้องแม่นยำลดลงก็หยุดทันที

**แบบที่ 2** วิธีการ โปสพรันทรี (Post-Prune Tree) คือทำการแตกต้นไม้ (Tree) จนหมดก่อน แล้วค่อยมาตัดแต่งกิ่งต้นไม้ ในภายหลัง

จากทั้งหมดที่กล่าวมาจึงพอสรุปได้ว่า ID3 อยู่ในกลุ่มของอัลกอริทึมการเรียนรู้ Decision Tree โดยใช้ทฤษฎีข่าวสาร (information theory) เพื่อตัดสินใจ ซึ่งเชริแอททริบิวโดยการรวบรวม เรคคอร์ดเพื่อแตกข้อมูลต่อไป แอททริบิวที่ถูกเลือกในวิธีนี้ถูกทำซ้ำไปซ้ำมาจนกระทั่ง Decision Tree สมบูรณ์ นั่นคือการจำแนกทุกๆอินพุตที่เข้ามา ถ้าข้อมูลมีสิ่งรบกวนจะมีผลทำให้เรคคอร์ด อาจจะถูกจำแนกผิดพลาด อาจจะเป็นไปได้ที่จะทำการแต่ง Decision Tree เพื่อที่จะลดความผิดพลาดในการจำแนกข้อมูลที่มีสิ่งรบกวน อยู่ได้ จากที่กล่าวมาทั้งหมดเป็นเพียงทฤษฎีของอัลกอริทึมนี้ แต่บางท่านอาจจะยังไม่เห็นถึงประโยชน์เท่าไรนั้น แต่ในบทถัดไปจะเป็นการนำทฤษฎีที่ได้ศึกษามาไปใช้ประโยชน์ได้กับงานจริง

## บทที่ 5

### การออกแบบและพัฒนาระบบงาน

ในการพัฒนาโปรแกรมประยุกต์สำหรับธุรกิจสำรองห้องพักโรงแรมโดยใช้ Decision Tree ในวิชาโครงการศึกษาระดับปริญญาตรีพิเศษนี้ จะใช้การพัฒนาระบบในลักษณะของ Web Application จึงได้มีการพัฒนาส่วนติดต่อผู้ใช้ทั้งหมดด้วยเทคโนโลยี ASP.NET ทำงานบน Internet Information Services (IIS) Manager เพื่อให้สามารถทำงานผ่าน Web Browser ได้ ทั้งนี้สาเหตุที่เลือกพัฒนาระบบ ด้วยเทคโนโลยี ASP.NET ซึ่งเป็นเทคโนโลยีที่สนับสนุนการพัฒนาระบบงานบนอินเทอร์เน็ต ประกอบกับเทคโนโลยีนี้มีความเข้ากันได้กับส่วนประกอบอื่น ๆ ที่ใช้ในการพัฒนาระบบงานบนอินเทอร์เน็ต อีกทั้งมีสาเหตุเนื่องมาจาก มี Base Classes หรือ Class library พื้นฐานต่างๆเตรียมไว้ให้ใช้งานอย่างมากมาย ซึ่งผู้พัฒนาสามารถนำเอา Classes ต่างๆมาใช้งานได้ตามความเหมาะสม ซึ่งช่วยทำให้ลดระยะเวลาในการพัฒนาได้เป็นอย่างมาก การ ASP.NET เป็นเทคโนโลยีที่กำลังได้รับความนิยมในวงกว้างในการใช้งาน และเป็นภาษาที่ใช้ในการพัฒนาที่มีผู้ที่สามารถพัฒนาได้เป็นจำนวนมาก เพื่อไม่ให้เกิดปัญหาหากมีความจำเป็นต้องปรับปรุง แก้ไขเพิ่มเติม ระบบงานในภายหลัง อีกทั้งยังเป็นภาษาที่มีการพัฒนาอย่างต่อเนื่อง ตอบสนองได้กับการพัฒนาระบบตามวัตถุประสงค์ของโครงการศึกษาระดับปริญญาตรีพิเศษของผู้จัดทำ นอกจากนี้ยังสามารถรองรับการทำงานของระบบงาน ซึ่งจะมีผู้ใช้งานระบบจำนวนมาก ได้เป็นอย่างดี

#### 5.1 องค์ประกอบต่างๆที่ใช้ในการพัฒนามีรายละเอียดดังนี้

- Microsoft Visio ใช้เป็นเครื่องมือในการวิเคราะห์และออกแบบระบบ
- การพัฒนา Web Application ใช้โปรแกรม Microsoft Visual Studio .net 2003
- ใช้ Microsoft SQL Server 2000 เป็นฐานข้อมูลหลักในการเก็บข้อมูล
- ผลลัพธ์ของโมเดลเก็บในไฟล์รูปแบบ XML
- Tree View Component ในใช้ในการแสดงผลรูปแบบต้นไม้บน Browser

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนของการพัฒนาโปรแกรมประยุกต์สำหรับธุรกิจสำรองห้องพักโรงแรมโดยใช้ Decision Tree นี้ ได้แบ่งระบบงานออกเป็น 2 ส่วน หลักๆ คือ

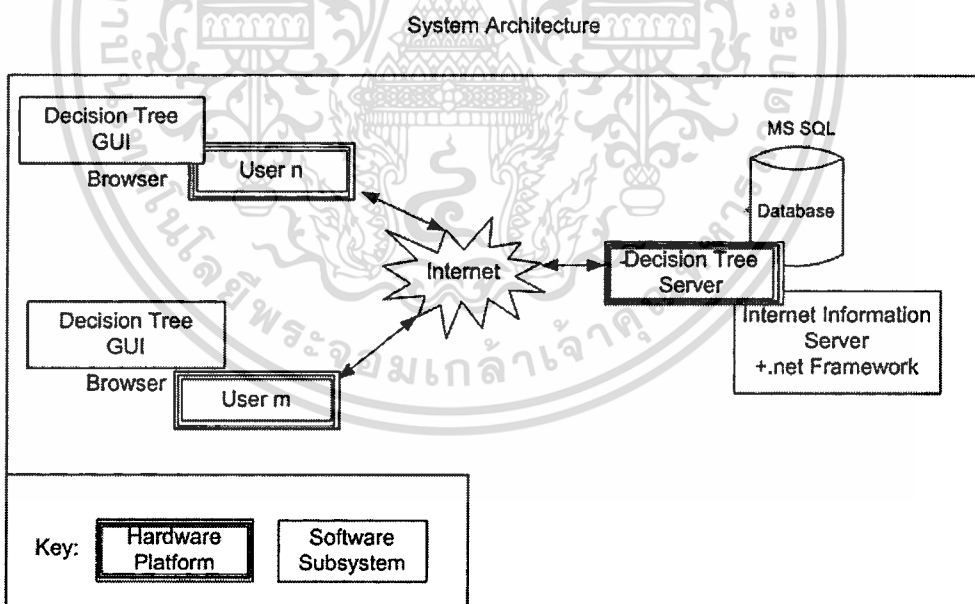
- ส่วนงานการเทรนนิ่ง (Training Phase)

การเทรนนิ่ง เป็นกระบวนการนำข้อมูลที่ได้จากการเตรียมข้อมูลเข้าสู่การประมวลผลด้วย อัลกอริทึมของคาค่าไมนิ่ง สำหรับระบบนี้จะใช้อัลกอริทึมในการทำงานคือ ID3

- ส่วนงานการทำนาย (Predicting Phase)

ส่วนงานการทำนายเป็นกระบวนการนำเอาโมเดลที่ได้มาใช้ในการทำนายกับข้อมูลอีกชุดหนึ่ง เพื่อวิเคราะห์หาความเป็นไปได้หรือผลลัพธ์ของแอททริบิวเป้าหมาย

### 5.2 System Architecture



รูปที่ 5.1 System Architecture

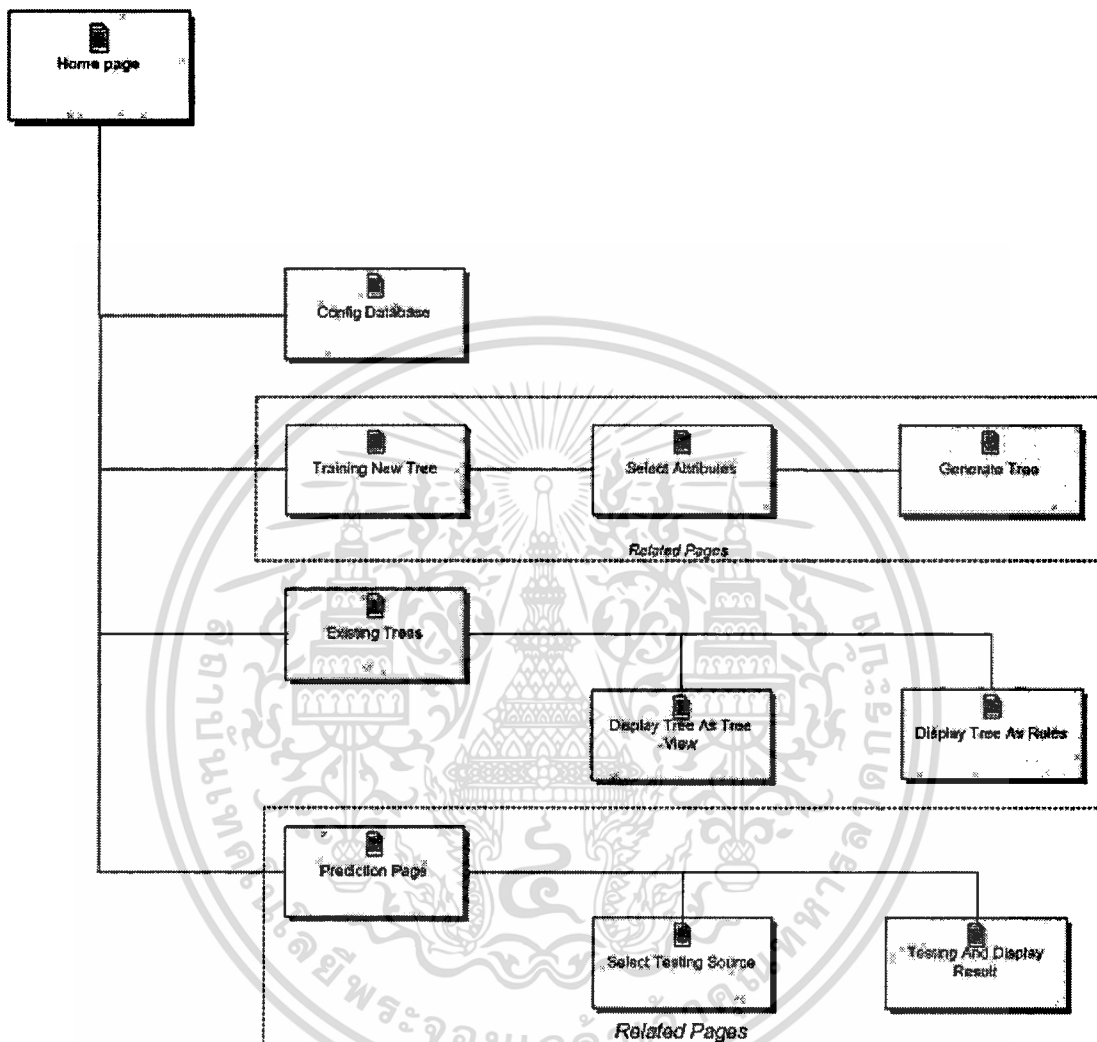
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## โครงสร้างพื้นฐานโดยรวมของระบบงานประกอบด้วย

- ผู้ให้บริการ
  - Hardware
    - เครื่อง Server 600MHZ Pentium III Class ขึ้นไป
    - Memory 512 MB
  - Software
    - Web Server IIS 5.0 ขึ้นไป
    - .NET Framework 1.1
    - Microsoft SQL Server Database or Microsoft Access Database
  - Network
    - สายสื่อสารที่เชื่อมต่อเข้าอินเทอร์เน็ต เช่น สายโทรศัพท์, ISDN, ADSL ฯลฯ
    - อุปกรณ์เชื่อมต่อสายสื่อสารกับอุปกรณ์คอมพิวเตอร์ เช่น Modem ฯลฯ
    - ชั่วโมงการใช้งานอินเทอร์เน็ตจากผู้ให้บริการอินเทอร์เน็ต
- ผู้ใช้บริการ
  - Hardware
    - เครื่อง PC Computer 500 MHZ Pentium III Class ขึ้นไป
  - Software
    - Microsoft Windows 95, 98, XP, 2000
    - Internet Explorer ตั้งแต่ Version 5.0
  - Network
    - สายสื่อสารที่เชื่อมต่อเข้าอินเทอร์เน็ต เช่น สายโทรศัพท์, ISDN, ADSL ฯลฯ
    - อุปกรณ์เชื่อมต่อสายสื่อสารกับอุปกรณ์คอมพิวเตอร์ เช่น Modem ฯลฯ
    - ชั่วโมงการใช้งานอินเทอร์เน็ตจากผู้ให้บริการอินเทอร์เน็ต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 5.3 Decision Tree System Web Site Diagram



รูปที่ 5.2 Decision Tree System Web Site Diagram

1. หน้า Homepage หลัก

เป็นหน้า Webpage หลัก เพื่อให้ผู้ใช้งานทราบถึงรายละเอียดของระบบ

2. หน้า Database Configuration

เป็นหน้า Webpage ที่ให้ผู้ใช้งานปรับแต่ง Database Configuration

3. หน้า Training New Tree

เป็นหน้า Webpage ที่ให้ผู้ใช้งานทำการสร้าง Decision Tree โดยจะทำการเลือก Table

และ Attributes ต่างๆ จากนั้นจึงทำการ Generate Decision Tree Model เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นับญาติให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4. หน้า Existing Tree

เป็นหน้า Webpage ที่ให้ผู้ใช้งานสามารถเลือกดู Tree ที่ได้สร้างไว้เรียบร้อยแล้ว โดยสามารถแสดงผลได้ในรูปแบบ Tree View และ Rules

#### 5. หน้า Predicting Samples

เป็นหน้า Webpage ที่แสดงผู้ใช้งานทำการทดสอบชุดข้อมูลกับ Model ที่มี เพื่อวิเคราะห์หาผลลัพธ์ และความเป็นไปได้ ของ Target Attribute ของชุดข้อมูลใหม่

### 5.4 System Design Framework



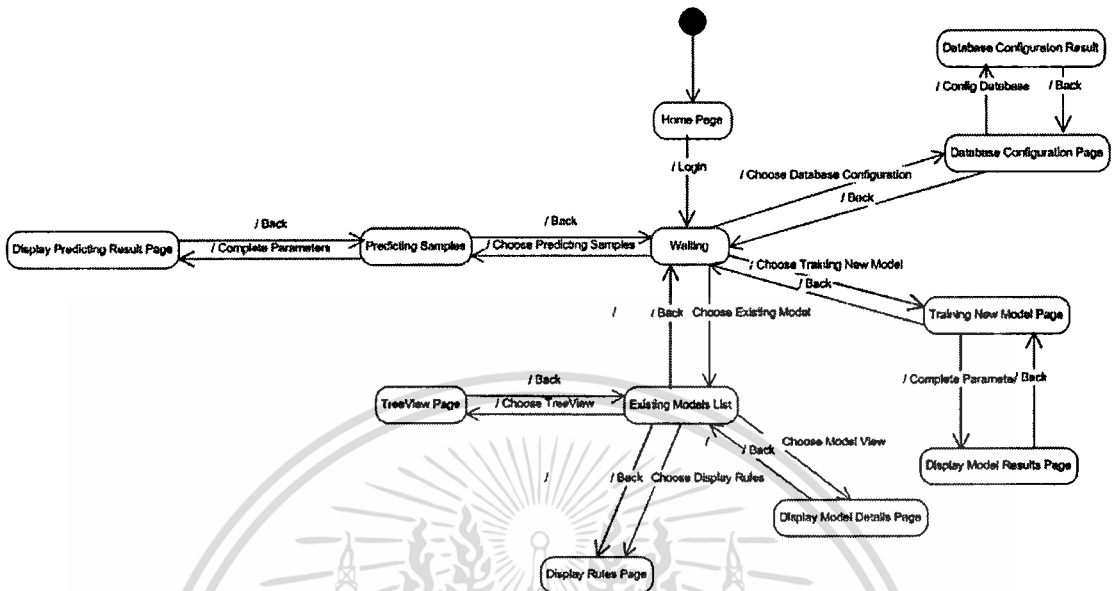
รูปที่ 5.3 System Design Framework

ในการออกแบบระบบ ใช้การออกแบบในรูปแบบของ Model View ,Control Pattern ซึ่งเป็นโครงสร้างที่ใช้ในการพัฒนา Web Application โดยมีส่วนประกอบดังนี้

- ชุดการแสดงผล (User Service / Presentation Tier View) เป็นส่วนที่เกี่ยวข้องกับการติดต่อกับผู้ใช้งานระบบโดยตรง ซึ่งได้แก่ ส่วนติดต่อผู้ใช้ (User Interface), ส่วนของการแสดงผลลัพธ์ (Output) และส่วนของการนำเข้า (Input)
- ชุดการประมวลผล (Business Service / Business Logic Tier) เป็นส่วนของระบบที่เกี่ยวข้องกับการทำงานที่เกิดขึ้นจริง ๆ เช่น ส่วนของการคำนวณ, กลไกการจัดตั้งข้อมูลที่ได้รับมาเพื่อนำมาใช้งาน, กลไกการบันทึกข้อมูล เป็นต้น
- ชุดข้อมูล (Data Service / Data Tier) เป็นส่วนการจำลอง Data Item ต่าง ๆ ที่ถูกจัดเก็บอยู่ในสื่อบันทึกข้อมูลในรูปแบบของฐานข้อมูล หรือ File

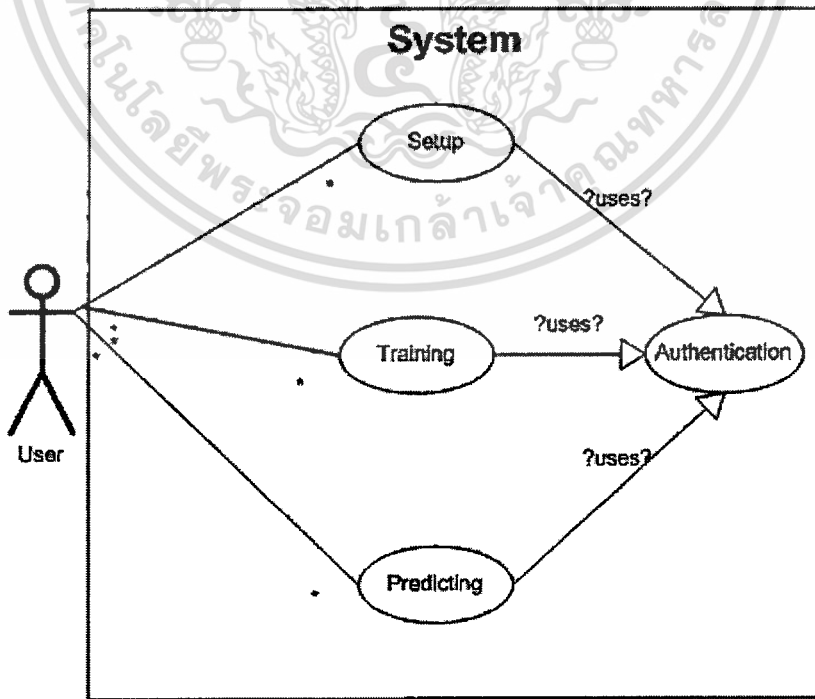
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.5 Activity Diagram Decision Tree System



รูปที่ 5.4 Activity Diagram Decision Tree System

5.6 Use Case Diagram Decision Tree System



รูปที่ 5.5 Use Case Diagram Decision Tree System

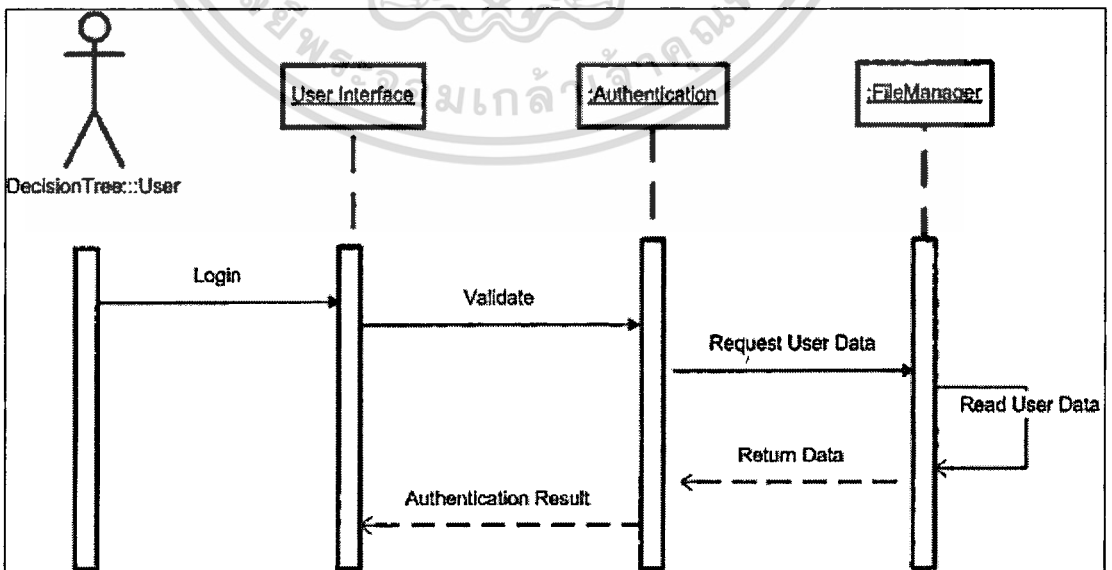
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบงานประกอบด้วย ยูสเคส ต่าง ๆ ดังนี้

- Authentication ตรวจสอบสิทธิของผู้ใช้งาน มีรายละเอียดดังนี้
  - ผู้ใช้งานกรอกข้อมูลรหัสผู้ใช้และรหัสผ่าน เพื่อทำการตรวจสอบสิทธิ
- Setup ปรับแต่งค่าการติดต่อฐานข้อมูล มีรายละเอียดดังนี้
  - ผู้ใช้งานสามารถเพิ่มหรือลบค่า Database Configuration และสามารถเลือก Activate Configuration ที่ต้องการใช้
- Training ทำการ Train Decision Tree มีรายละเอียดดังนี้
  - ผู้ใช้งานสามารถเลือก Table, Target Attribute และ Descriptor Attributes เพื่อทำการ Training เพื่อหา Decision Tree Model แล้วจึงนำผลลัพธ์ เก็บไว้ในรูปแบบ XML เพื่อจะแปลงกลับมาเป็น Tree View หรือ Model Rules ต่อไป
- Predicting ทำการเปรียบเทียบเพื่อทำนายค่าของ Target Attribute มีรายละเอียดดังนี้
  - ผู้ใช้งานสามารถเลือก แล้ว Existing Model หรือ โมเดลที่ได้ทำการสร้างไว้เพื่อที่จะใช้ในการทำนายและทำการเลือก Table ของกลุ่มข้อมูลชุดใหม่ ที่ยังไม่ทราบผลลัพธ์ จะมีการเช็คค่าชุดข้อมูลใหม่นั้นมีค่า Attributes ต่างๆ ที่จำเป็นต้องใช้ในการทดสอบตาม Model ที่เลือกหรือไม่ จากนั้นจึงทำการทดสอบ ว่ากลุ่มข้อมูลเหล่านั้น มีความน่าจะเป็นที่จะมีค่า Target Attribute อย่างไร

5.7 Sequence Diagram Decision Tree System

5.7.1 Authentication



รูปที่ 5.6 Sequence Diagram ของ Use Case Authentication

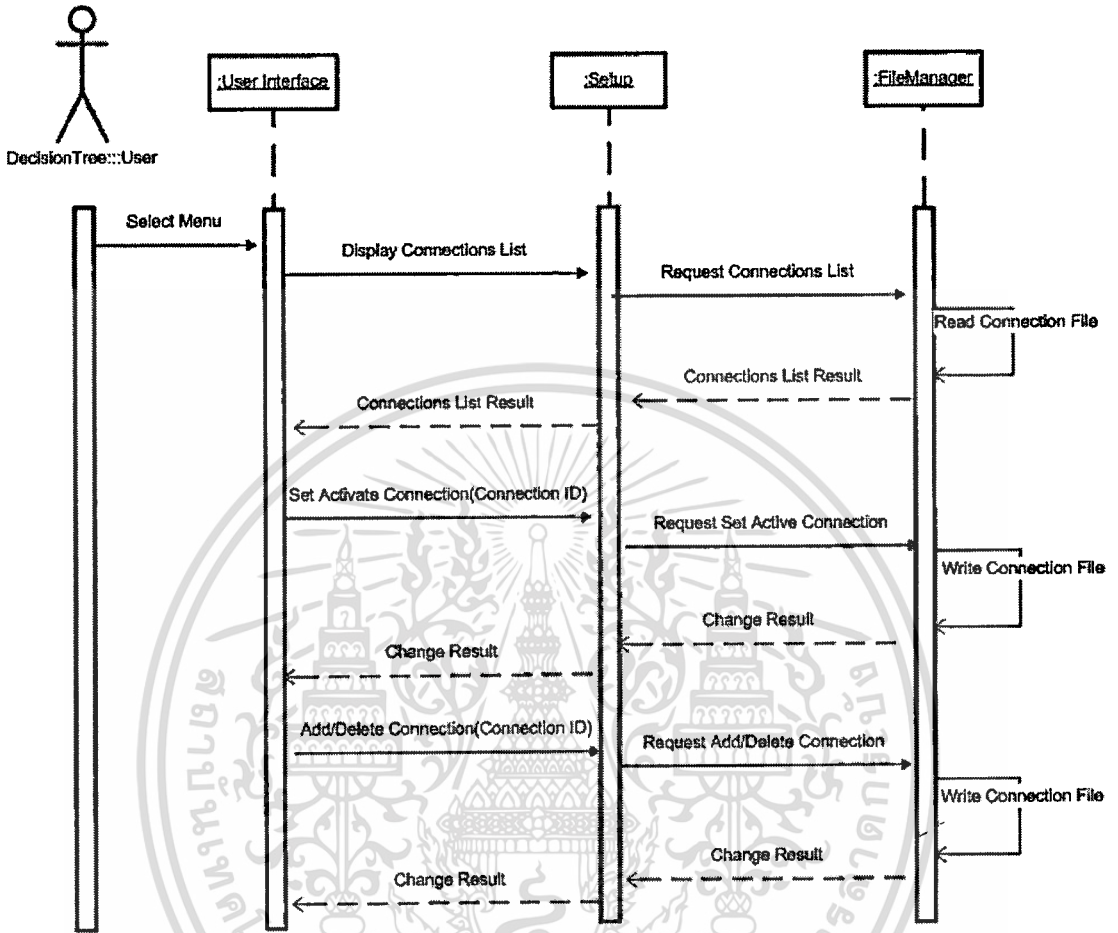
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เพื่อการศึกษาค้นคว้าเท่านั้น เมื่อผู้ผู้ใดเห็นปะโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 รายละเอียดของ Sequence Diagram ของ Use Case Authentication

Use Case	Authentication
วัตถุประสงค์	เพื่อให้ผู้ใช้งานทำการตรวจสอบสิทธิเพื่อเข้าสู่ระบบ
เงื่อนไขเมื่อเริ่มต้น	-
เมื่อทำงานสำเร็จ	ผู้ใช้งานจะสามารถทำการเข้าใช้ระบบได้
เมื่อทำงานไม่สำเร็จ	ผู้ใช้งานจะไม่สามารถเข้าใช้ระบบได้ และ ระบบจะแจ้งให้ผู้ใช้งานทำการเข้าใช้ใหม่อีกครั้ง
Actor ที่เกี่ยวข้อง	User
สิ่งกระตุ้นการทำงาน	ผู้ใช้งานทำการ Login
Input	Username และ Password
Output	แจ้งผลว่าสามารถเข้าใช้ระบบได้หรือไม่
รายละเอียด	<ol style="list-style-type: none"> <li>1. ผู้ใช้งานทำการ login เข้าสู่ระบบ โดยใส่ User name และ Password</li> <li>2. Validate จะทำการส่ง User name และ Password ไปค้นหาข้อมูลของผู้ใช้งาน</li> <li>3. หากไม่พบ ข้อมูล User name และ Password หรือข้อมูลที่ค้นหามาได้ไม่ถูกต้องกับค่าผู้ใช้งานส่งเข้ามา จะทำการแจ้งให้ผู้ใช้งานทราบถึงปัญหาที่พบ</li> </ol>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.7.2 Setup



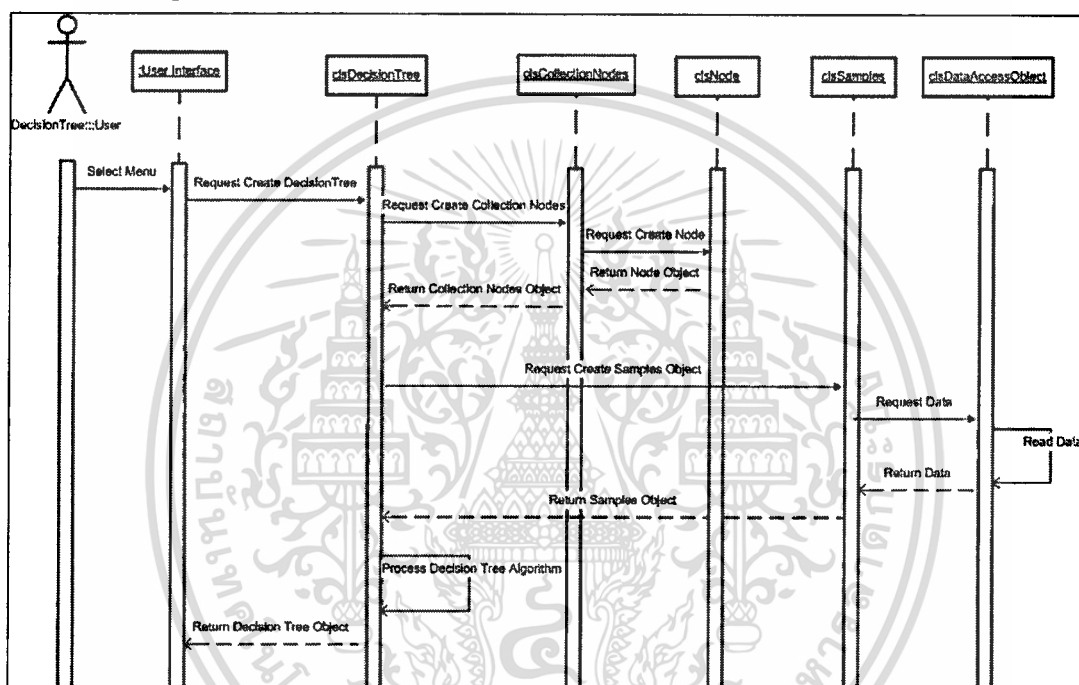
รูปที่ 5.7 Sequence Diagram ของ Use Case Setup

ตารางที่ 5.2 รายละเอียดของ Sequence Diagram ของ Use Case Setup

Use Case	Setup
วัตถุประสงค์	เพื่อให้ผู้ใช้งานสามารถทำการปรับแต่งค่าการติดต่อฐานข้อมูล
เงื่อนไขเมื่อเริ่มต้น	ผู้ใช้ต้องผ่านการ Authentication
เมื่อทำงานสำเร็จ	ผู้ใช้งานจะสามารถทำการปรับแต่งค่าการติดต่อฐานข้อมูลได้
เมื่อทำงานไม่สำเร็จ	ผู้ใช้งานจะไม่สามารถทำการปรับแต่งค่าการติดต่อฐานข้อมูลได้
Actor ที่เกี่ยวข้อง	User
สิ่งกระตุ้นการทำงาน	ผู้ใช้งานทำการเลือกปรับแต่งค่าการติดต่อฐานข้อมูล

Input	ค่า Configuration
Output	แจ้งผลว่าสามารถบันทึกการปรับแต่งค่าการติดต่อฐานข้อมูลได้หรือไม่
รายละเอียด	1. ผู้ใช้งานเลือก Add หรือ ลบ ข้อมูลการติดต่อฐานข้อมูล 2. ผู้ใช้งานกรอกข้อมูลการตั้งค่าต่างๆ แล้วจึงกด Save

### 5.7.3 Training



รูปที่ 5.8 Sequence Diagram ของ Use Case Training

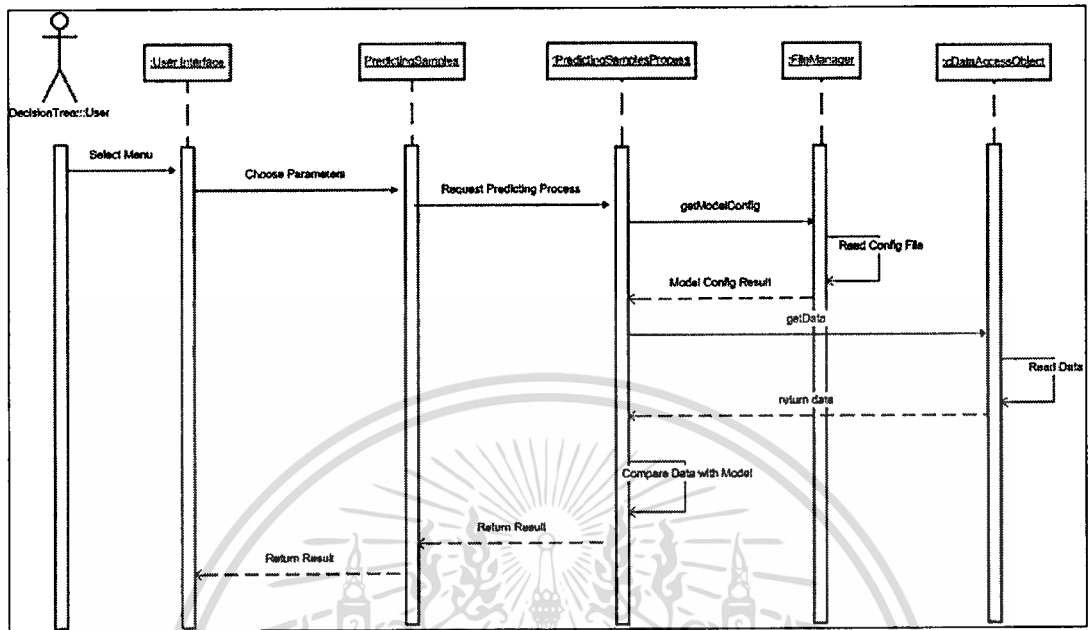
ตารางที่ 5.3 รายละเอียดของ Sequence Diagram ของ Use Case Training

Use Case	Training
วัตถุประสงค์	เพื่อทำการ Traing ข้อมูลตาม อัลกอริทึม ID3
เงื่อนไขเมื่อเริ่มต้น	ผู้ใช้งานต้องผ่านการ Authentication
เมื่อทำงานสำเร็จ	ทำการสร้าง Decision Tree Model บันทึกเก็บไว้ในรูปแบบ XML
เมื่อทำงานไม่สำเร็จ	ระบบจะแจ้งให้ผู้ใช้งานทำการแก้ไขถึงปัญหาที่เกิดขึ้น
Actor ที่เกี่ยวข้อง	User
สิ่งกระตุ้นการทำงาน	ผู้ใช้งานทำการเลือกข้อมูล Attributes ต่างๆ และสั่งให้ทำการ Training

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Input	ชื่อตาราง Target Attribute และ Descriptor Attributes ที่ถูกเลือกมาเพื่อทำการ Training
Output	แจ้งผลว่าสามารถทำการ Training Model สำเร็จหรือไม่ สามารถดูได้อย่างไร
รายละเอียด	<ul style="list-style-type: none"> <li>• User Interface เรียกใช้ Object Decision Tree เพื่อทำการสร้าง Decision Tree ตามอัลกอริทึม โดยได้ผลลัพธ์กลับมาเป็น Object ประเภท Collection Nodes แล้วจะนำไปเก็บไว้ในรูปแบบ XML เพื่อนำ Model ที่ได้ไปใช้ต่อไป</li> <li>• Object Decision Tree เรียกใช้ Object Collection Nodes เพื่อสร้างกลุ่มของ Nodes</li> <li>• Object Collection Nodes เรียกใช้ Object Node เพื่อสร้าง Object ประเภท Node</li> <li>• Object Decision Tree เรียกใช้ Object Samples เพื่อสร้าง Object ประเภท Samples ในการเก็บชุดข้อมูลต่างๆที่ใช้ในระหว่างการ Process ตามอัลกอริทึม</li> <li>• Object Samples เรียกใช้ Object DataAccessObject เพื่อใช้ในการติดต่อกับ Database</li> <li>• จากนั้น Object Decision Tree จะทำการ ดึงข้อมูลจากฐานข้อมูลตามเงื่อนไขของ ตาราง Target Attribute และ Descriptor Attributes ที่ผู้ใช้งานเลือกมา แล้วจึงนำข้อมูลไปทำงานตามอัลกอริทึม</li> </ul>

### 5.7.4 Predicting



รูปที่ 5.9 Sequence Diagram ของ Use Case Predicting

ตารางที่ 5.4 รายละเอียดของ Sequence Diagram ของ Predicting

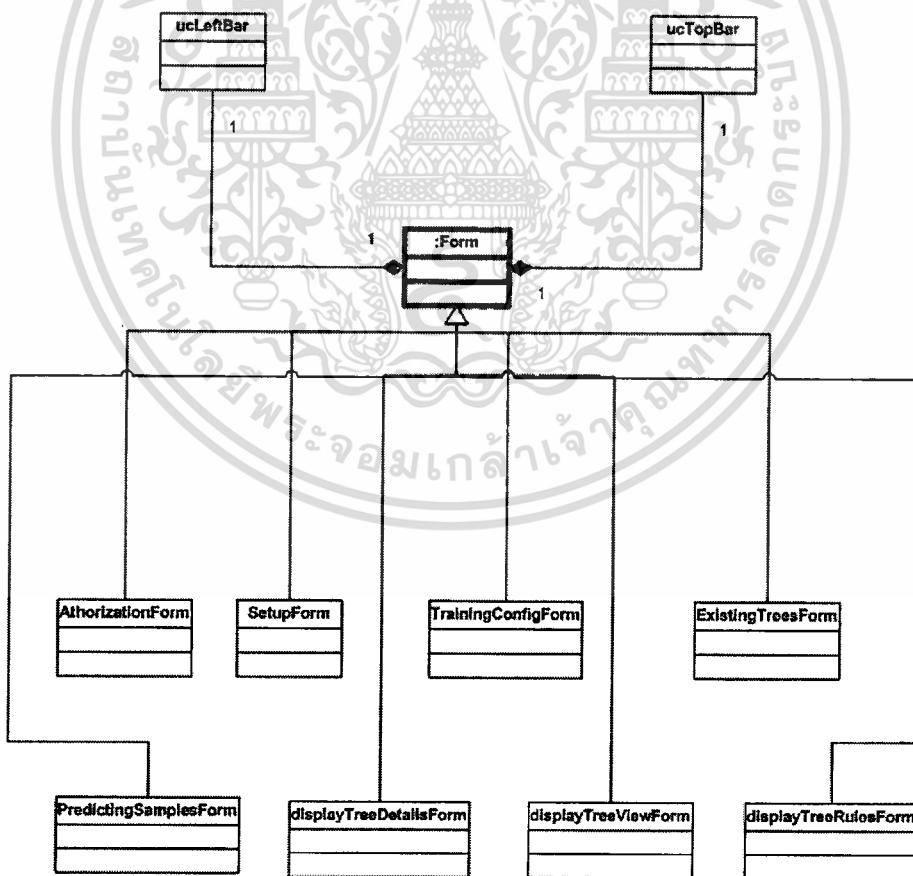
Use Case	Predicting
วัตถุประสงค์	เพื่อประมวลผลผลลัพธ์ของการประเมินของกลุ่มตัวอย่างจาก Model ที่มี
เงื่อนไขเมื่อเริ่มต้น	ผู้ใช้ต้องผ่านการ Authentication
เมื่อทำงานสำเร็จ	ผู้ใช้งานจะได้รับผลลัพธ์การประเมิน
เมื่อทำงานไม่สำเร็จ	ระบบจะแจ้งให้ผู้ใช้ทำการแก้ไขถึงปัญหาที่เกิดขึ้น
Actor ที่เกี่ยวข้อง	User
สิ่งกระตุ้นการทำงาน	ผู้ใช้ทำการเลือก Model ที่มีและตารางข้อมูลที่ต้องการนำมาประมวลผลและสั่งให้ทำการ Predicting
Input	Model ที่ต้องการใช้ ชื่อตาราง
Output	แจ้งผลการประมวลผล Predicting

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<p>รายละเอียด</p>	<ol style="list-style-type: none"> <li>1. ผู้ใช้งานทำการเลือก Model ที่ต้องการใช้ในการประมวลผล จากนั้นจึงเลือกตารางข้อมูลที่ต้องการใช้ในการ Predicting</li> <li>2. ระบบจะทำการตรวจสอบว่าตารางที่ผู้ใช้งานเลือกมามี Attributes ต่างๆ ครบถ้วนตามที่ Model หากไม่ครบถ้วนจะทำการยกเลิกการประมวลผลและจะแจ้งให้ผู้ใช้งานทราบถึงปัญหาที่พบ</li> <li>3. หากมี Attributes ครบถ้วนตามที่ Model ต้องการก็จะทำการประมวลผลเปรียบเทียบตามกฎต่างๆของ Model</li> <li>4. หลังจากประมวลผลเสร็จเรียบร้อยแล้ว ก็จะแสดงผลลัพธ์ให้ผู้ใช้งาน</li> </ol>
-------------------	--

### 5.8 Class Diagram Decision Tree System

#### User Service Class



รูปที่ 5.10 User Service Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## User Service Class ประกอบไปด้วย Class ต่าง ๆ ดังนี้

- ucLeftBar Class เป็นส่วนที่เก็บส่วน Left ของ Page ซึ่งจะแสดงอยู่ในทุก ๆ Page

ucLeftBar
-designerPlaceholderDeclaration : Object
-InitializeComponent()
-Page_Init(in sender : Object, in e : EventArgs)
-Page_Load(in sender : Object, in e : EventArgs)

รูปที่ 5.11 ucLeftBar Class

- ucTopBar Class เป็นส่วนที่เก็บส่วน Top ของ Page ซึ่งจะแสดงอยู่ในทุก ๆ Page

ucTopBar
-designerPlaceholderDeclaration : Object
-InitializeComponent()
-Page_Init(in sender : Object, in e : EventArgs)
-Page_Load(in sender : Object, in e : EventArgs)

รูปที่ 5.12 ucTopBar Class

- Authorization เป็นส่วนติดต่อผู้ใช้ในส่วนของการ Authorizatioin

Authorization
#txtUserName : TextBox
#txtPassword : TextBox
#lblResult : Label
-designerPlaceholderDeclaration : Object
-InitializeComponent()
-Page_Init(in sender : Object, in e : EventArgs)
-Page_Load(in sender : Object, in e : EventArgs)
-btnLogin_Click(in sender : Object, in e : EventArgs)

รูปที่ 5.13 Authorization Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Setup เป็นส่วนติดต่อผู้ใช้ในส่วนของการ Setup

Setup
<pre> #dgConfig : DataGrid #panelAddNew : Panel #txtNewName : TextBox #ddlProvider : DropDownList #txtDataSource : TextBox #txtUserName : TextBox #txtPassword : TextBox #lnkAddNew : LinkButton #lnkSave : LinkButton #txtCatalog : TextBox #lblText : Label -designerPlaceholderDeclaration : Object -InitializeComponent() -Page_Init(in sender : Object, in e : EventArgs) -Page_Load(in sender : Object, in e : EventArgs) -blnAdd_Click(in sender : Object, in e : EventArgs) +dgConfig_EditCommand(in s : Object, in e : DataGridCommandEventArgs) +dgConfig_Cancel(in Source : Object, in E : DataGridCommandEventArgs) +dgConfig_UpdateCommand(in s : Object, in e : DataGridCommandEventArgs) +BindData() +dgConfig_DeleteCommand(in s : Object, in e : DataGridCommandEventArgs) +dgConfig_ActivateCommand(in s : Object, in e : DataGridCommandEventArgs) +ActivateId(in intActivate : Integer) -lnkAddNew_Click(in sender : Object, in e : EventArgs) -lnkSave_Click(in sender : Object, in e : EventArgs) </pre>

รูปที่ 5.14 Setup Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- TrainingConfig เป็นส่วนติดต่อผู้ใช้ในส่วนของการ Training Config

```

TrainingConfig
#dgRadio : DataGrid
#dgTable : DataGrid
#lnkListTable : LinkButton
#lnkListColumnSchema : LinkButton
#lnkChooseTargetAtt : LinkButton
#lnkChooseDescriptorAtt : LinkButton
#lnkConfirm : LinkButton
#dgDescriptorAtt : DataGrid
#lnkReset : LinkButton
#lit : Literal
-designerPlaceholderDeclaration : Object
-InitializeComponent()
-Page_Init(in sender : Object, in e : EventArgs)
-Page_Load(in sender : Object, in e : EventArgs)
+ResetParams()
+dplStatus()
+DetermineSelection(in item : DataGridItem, inout count : Integer) : String
+DetermineRadioSelection(in item : DataGridItem, inout count : Integer) : String
+DetermineRadioTableSelection(in item : DataGridItem, inout count : Integer) : String
-btnListTable_Click(in sender : Object, in e : EventArgs)
-lnkListTable_Click(in sender : Object, in e : EventArgs)
-lnkListColumnSchema_Click(in sender : Object, in e : EventArgs)
-btnGetColumn_Click(in sender : Object, in e : EventArgs)
-lnkChooseTargetAtt_Click(in sender : Object, in e : EventArgs)
-lnkChooseDescriptorAtt_Click(in sender : Object, in e : EventArgs)
-lnkConfirm_Click(in sender : Object, in e : EventArgs)
-lnkReset_Click(in sender : Object, in e : EventArgs)
    
```

รูปที่ 5.15 Training Config Class

- ExistingTrees เป็นส่วนติดต่อผู้ใช้ในส่วนของการแสดง Existing Tree

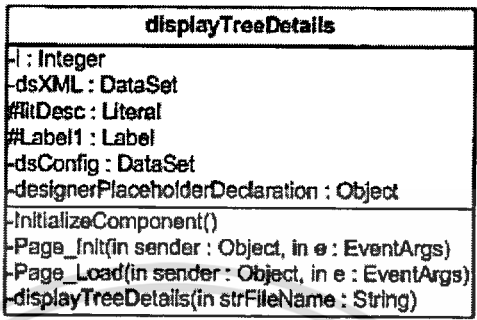
```

ExistingTrees
#dgExistingTrees : DataGrid
-designerPlaceholderDeclaration : Object
-InitializeComponent()
-Page_Init(in sender : Object, in e : EventArgs)
-Page_Load(in sender : Object, in e : EventArgs)
-loadExistingTrees()
    
```

รูปที่ 5.16 ExistingTrees Class

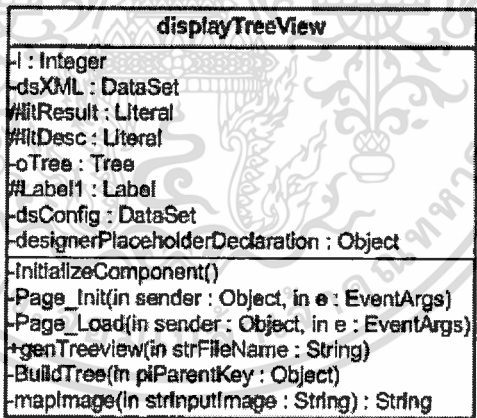
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- displayTreeDetails เป็นส่วนติดต่อผู้ใช้ในส่วนของการแสดง Tree Details



รูปที่ 5.17 displayTreeDetails Class

- displayTreeView เป็นส่วนติดต่อผู้ใช้ในส่วนของการแสดง Tree View



รูปที่ 5.18 displayTreeView Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- displayTreeRules เป็นส่วนติดต่อผู้ใช้ในส่วนของการแสดง Tree Rules

displayTreeRules
#HitDesc : Literal #Label1 : Label -dsTreeRules : DataSet -dsConfig : DataSet #HitText : Literal -designerPlaceholderDeclaration : Object
-InitializeComponent() -Page_Init(in sender : Object, in e : EventArgs) -Page_Load(in sender : Object, in e : EventArgs) +loadTree(in strFileName : String)

รูปที่ 5.19 displayTreeRules Class

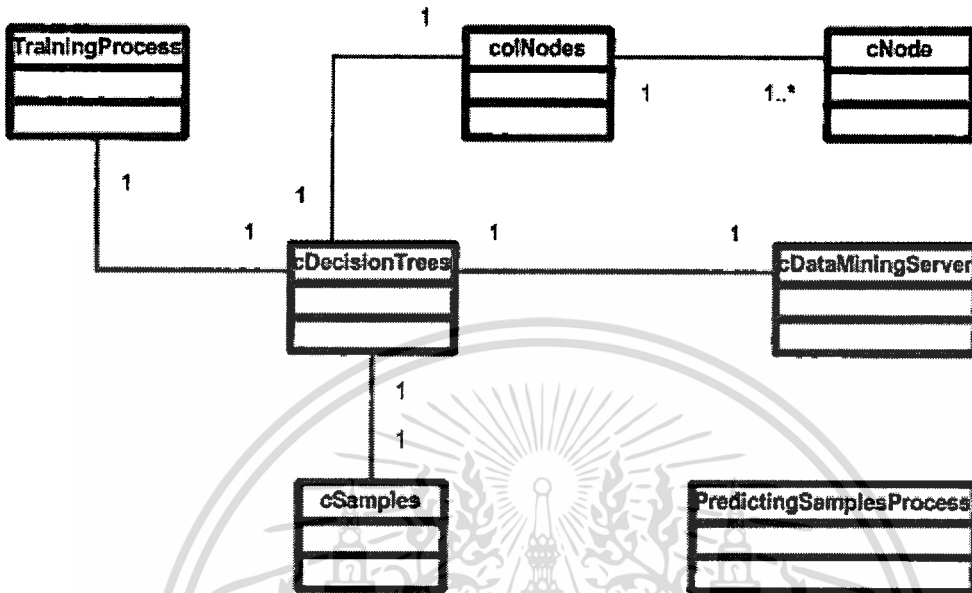
- PredictingSampls เป็นส่วนติดต่อผู้ใช้ในส่วนของการ Predicting Samples

PredictingSampls
#dgExistingTrees : DataGrid -designerPlaceholderDeclaration : Object
-InitializeComponent() -Page_Init(in sender : Object, in e : EventArgs) -Page_Load(in sender : Object, in e : EventArgs) -loadExistingTrees()

รูปที่ 5.20 PredictingSampls Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

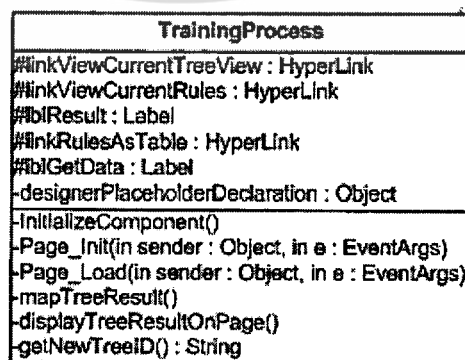
## Business Service Class



รูปที่ 5.21 Business Service Class

Business Service Class ประกอบไปด้วย Class ต่างๆดังนี้

- Training Process Class เป็นส่วนที่ใช้ในการ Training Decision Tree โดยรับค่าจาก Training Process Config แล้วเรียกใช้ Class ต่างๆ ที่จำเป็นแล้วทำการประมวลผลตาม อัลกอริทึม



รูปที่ 5.22 Training Process Class

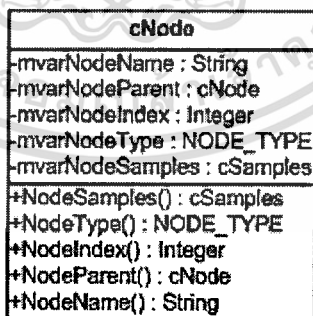
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- cDataMiningServer Class ใช้สำหรับเก็บค่าตัวแปรต่างๆ ไม่ว่าจะเป็น Connection String, ค่าตารางที่เลือก, ค่า Attributes ต่างๆที่เลือกใช้



รูปที่ 5.23 cDataMiningServer Class

- cNode Class ใช้สำหรับเก็บค่าตัวแปรต่างๆ ในใช้เป็น Node ของต้นไม้



รูปที่ 5.24 cNode Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- colNodes Class ใช้สำหรับเก็บค่าตัวแปรประเภท Node โดยเก็บเป็น Collection เพื่อใช้ในการสร้าง Decision Tree

colNodes
-mCol : Collection
-mvarNodeMaximum : Integer = 100
+Add(in clsNode : cNode, inout sKey ; String = "") : cNode
+Item(in vntIndexKey : Object) : cNode
+Count() : Integer
+GetEnumerator() : IEnumerator
+Remove(inout vntIndexKey : Object)
-Class_Initialize_Renamed()
+New()
-Class_Terminate_Renamed()
#Finalize()

รูปที่ 5.25 colNodes Class

- cSamples Class ใช้สำหรับเก็บพิกัดของข้อมูลต่างๆในการทำงาน

cSamples
mvarClass : String
mvarDataSet : Recordset
mvarAttributes : Recordset
mvarRecordCount : Integer
mvarAttributesCount : Integer
mvarUniqueClassValues : Boolean
mvarSamplesDataTable : DataTable
mvarSelectedAttributes : DataTable
+UniqueClassValues() : Boolean
+AttributesCount() : Integer
+RecordCount() : Integer
+Attributes() : Recordset
+SelectedAttributes() : DataTable
+DataSet() : Recordset
+SamplesDataTable() : DataTable
+Class_Renamed() : String
+Filter_Renamed(in rstSource : Recordset, in varFilter : Object) : Recordset
+FirstFilter(in rstSource : Recordset, in varFilter : Object) : Recordset
+Clone(in rstSource : Recordset) : Recordset
+Info(in ClassName : String, in ClassValues : Recordset) : Double
+Partition_Renamed(in Attribute_Name : String, in Attribute_Value : String, in Sample : cSamples) : cSamples
+SameClass(in ClassName : String, in ClassValues : Recordset) : Boolean
+MajorityClass(in ClassName : String, in ClassValues : Recordset) : String
+Maximum_Gain(in ClassName : String, in ClassValues : Recordset, in mvarUniqueAttributeValues : Recordset) : Recordset
+GetUniqueValues(in mvarUniqueAttributeValues : Recordset, in AttributeName : String) : Recordset

รูปที่ 5.26 cSamples Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- cDecisionTrees Class ใช้สำหรับการสร้าง Decision Tree ตามอัลกอริทึม ID3

cDecisionTrees
+NODE_LIMIT : Integer = 100
+oDataSet : DataSet = New System.Data.DataSet
+oConn : OleDbConnection = New System.Data.OleDb.OleDbConnection
-Attributes_Selected() : String
-Class_Selected : String
-Table_Selected : String
-mvarUniqueAttributeValues : Recordset
-mvarDataSet : Recordset
-mvarMajorityClass : String
-mvarDataSet_Second : Recordset
-mvarUniqueClassValues : Recordset
+UniqueClassValues() : Recordset
+DataSet_Second() : Recordset
+MajorityClass() : String
+UniqueAttributeValues() : Recordset
+DataSet() : Recordset
+CopySamples(in Sample : cSamples) : cSamples
+GetUniqueValues(in AttributeName : String) : Recordset
+FilterDataTableByAttribute(in InputDataTable : DataTable, in AttributeName : String) : DataTable
+DecisionTreeMine(inout Samples : cSamples, in NodeX : cNode) : colNodes
+Entropy(inout Values() : Integer) : Double
+LogN(inout x : Double, inout BaseN : Double) : Double
-Class_Initialize_Renamed()
+New()
+CreateDecisionTree(in ConnectionString : String) : colNodes
+CreateUniqueAttributeValues()
+CreateNode(in NodeText : String, in Parent : cNode, in NodeType : NODE_TYPE) : cNode

รูปที่ 5.27 cDecisionTrees Class

- PredictingSamplesProcess Class ใช้สำหรับ ประมวลผลชุดข้อมูลใหม่เปรียบเทียบกับ Model ที่ถูกเลือกขึ้นมา เพื่อค้นหาความน่าจะเป็นของ

PredictingSamplesProcess
-dsTreeRules : DataSet
#panelChooseTable : Panel
#lblSelect : Label
#litError : Literal
#litResult : Literal
-dsConfig : DataSet
#dgTable : DataGrid
#lnkValidateAttributes : LinkButton
#litDesc : Literal
-designerPlaceholderDeclaration : Object
-InitializeComponent()
-Page_Init(in sender : Object, in e : EventArgs)
-Page_Load(in sender : Object, in e : EventArgs)
-loadTableSchema()
-lnkValidateAttributes_Click(in sender : Object, in e : EventArgs)

รูปที่ 5.28 cDataMiningServer Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Service Class ประกอบไปด้วย Class ต่างๆดังนี้

Class cDataAccess Object Class ใช้สำหรับติดต่อกับฐานข้อมูล ดึงข้อมูล ตลอดจนบันทึกเก็บข้อมูลพักไว้ใน เพื่อทำการประมวลผลต่อไป

```

cDataAccessObject
+ConnectionObject : Connection
+ConnectionString : String
+adoPrimaryRS : Recordset
+DoingRequery : Boolean
+mvwFieldNames() : Object
- ConnectionUsed : String
- Class Selected : String
- Table Selected : String
- Attributes Selected() : String
+MoveComplete()
+Class_Initialize_Renamed()
+New()
+ExecuteSQL(Inout QueryString : String, Inout Connect : String, Inout ReturnRecordset : Recordset = Nothing, Inout ReturnArray : Object = Nothing) : Boolean
+WriteToXML(Inout rstSchema : Recordset, Inout strPath : String) : Boolean
+Class_GetDataMember(Inout DataMember : String, Inout Data : Object)
+adoPrimaryRS_MoveComplete(In adReason : EventReasonEnum, In pError : Error, Inout adStatus : EventStatusEnum, In pRecordset : Recordset)
+adoPrimaryRS_WillChangeRecord(In adReason : EventReasonEnum, In cRecords : Integer, Inout adStatus : EventStatusEnum, In pRecordset : Recordset)
+EditingRecord() : Boolean
+AbsolutePosition() : Integer
+FieldNames() : Object
+AddNew()
+Delete()
+Requery()
+Update()
+Cancel()
+MoveFirst()
+MoveLast()
+MoveNext()
+MovePrevious()
+GetDatabaseTables_SQL(In strConn : String) : DataTable
+GetDatabaseColumns_SQL(In strConn : String, In strTable : String) : DataTable
+ConnectionUsed() : String
+Class_Selected() : String
+Table_Selected() : String
+Attributes_Selected() : String

```

รูปที่ 5.29 cDataAccessObject Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 6

### การเตรียมข้อมูล

เพื่อเป็นการเพิ่มประสิทธิภาพและประสิทธิผลของการดำเนินธุรกิจสำรองห้องพักจากเว็บไซต์ จึงได้มีการศึกษาแนวทางที่จะสามารถค้นหาลูกค้าที่มีศักยภาพของธุรกิจ เพื่อมุ่งเน้นการใช้ทรัพยากรของบริษัทที่มีอยู่จำกัด ไปยังลูกค้ากลุ่มที่มีศักยภาพ ในการนำทฤษฎีเกี่ยวกับค้ำไ่มิ่งมาประยุกต์ใช้โดยมีวัตถุประสงค์เพื่อหารูปแบบของการตัดสินใจสำหรับการค้นหาและคัดเลือกลูกค้าที่มีศักยภาพ แต่เนื่องจากข้อมูลในฐานข้อมูลเดิม ไม่สามารถนำมาดำเนินการได้ทันที จึงต้องมีกระบวนการต่างๆสำหรับจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมก่อนนำมาใช้งาน ซึ่งจะได้กล่าวดังต่อไปนี้

#### 6.1 แหล่งที่มาของข้อมูล

ข้อมูลที่ต้องนำมาใช้ในการดำเนินการ ได้มาจากฐานข้อมูลลูกค้า ซึ่งเก็บข้อมูลไว้โดยใช้ฐานข้อมูล Microsoft SQL Server 2000 ดังมีรายละเอียดตารางข้อมูลดังนี้

ตารางที่ 6.1 CSDB.BKINFO เป็นตารางที่ใช้เก็บข้อมูลทั่วไปของ Booking

PHYSICAL NAME	DATATYPE	SIZE	DESCRIPTION
REFNO	varchar	20	หมายเลขอ้างอิงของ Booking
HOTEL_CODE	varchar	10	รหัสโรงแรม
HOTEL	varchar	100	ชื่อโรงแรม
AREA	varchar	50	บริเวณที่ตั้งของ โรงแรม
CITY	varchar	50	เมืองที่โรงแรมตั้งอยู่
COUNTRY	varchar	50	ประเทศที่โรงแรมตั้งอยู่
DMCCODE	varchar	15	รหัส ตัวแทนจำหน่าย
DMCNAME	varchar	50	ชื่อตัวแทนจำหน่าย
INDATE	datetime	8	วันที่เข้าพัก
OUTDATE	datetime	8	วันที่ออก
NORM	int	4	จำนวนห้องที่สำรอง
RMDESC	varchar	150	คำอธิบายห้องพัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาดูเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

RMOCC	smallint	2	รหัสประเภทห้องพัก
ALTHOTEL1_CODE	varchar	10	รหัสโรงแรมอื่นๆที่เลือก
ALTHOTEL1	varchar	100	ชื่อโรงแรมอื่นๆที่เลือก
NOADULT	smallint	2	จำนวนผู้ใหญ่ที่เข้าพัก
NOCHILD	smallint	2	จำนวนเด็กที่เข้าพัก
AGECHILD	varchar	80	อายุของเด็ก
SMOKING	varchar	5	ความต้องการห้องที่สูบบุหรี่ได้
FLOOR	varchar	5	ความต้องการชั้นเฉพาะเจาะจง
BEDTYPE	varchar	15	ประเภทเตียง
GUESTS	varchar	255	ชื่อแขกที่เข้าพัก
RMTYPE	varchar	150	ประเภทห้องพัก

ตารางที่ 6.2 CSDB.BKSTATUS เป็นตารางที่ใช้เก็บข้อมูลสถานะต่างๆ Booking

PHYSICAL NAME	DATATYPE	SIZE	DESCRIPTION
REFNO	varchar	20	หมายเลขอ้างอิงของ Booking
BKTYPE	varchar	30	ประเภทของ Booking
STATUS	varchar	30	สถานะของ Booking
MADEBY	varchar	30	ผู้ทำการบันทึกข้อมูล Booking
MADEON	datetime	8	วันที่ทำการบันทึก
AMEND	smallint	2	มีการแก้ไข Booking
MODIFYBY	varchar	30	แก้ไข Booking โดย
MODIFYON	datetime	8	วันที่ทำการแก้ไข
CXL	varchar	1	ค่าการยกเลิก
CXLREASON	varchar	150	เหตุผลที่ยกเลิก
CXLBY	varchar	30	ยกเลิกโดย
CXLON	datetime	8	วันที่ทำการยกเลิก
CONFBYDMC	varchar	1	ยืนยันโดย DMC แล้วหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

CONFBY	varchar	30	ยืนยัน โดยผู้ใด
CONFON	datetime	8	วันที่ทำการยืนยัน
REJBYDMC	varchar	1	ถูกปฏิเสธ โดย DMC หรือไม่
REJREASON	varchar	150	เหตุผลที่ปฏิเสธ
REJBY	varchar	30	ผู้ที่ปฏิเสธ
REJON	datetime	8	วันที่ทำการปฏิเสธ
DMCDUE	datetime	8	วันกำหนดสุดท้ายของ DMC ในการยืนยัน
GETCC	varchar	1	ได้รับCredit Card แล้วหรือไม่
GETCCON	datetime	8	วันที่ได้รับCredit Card
GETCCBY	varchar	30	ได้รับCredit Card โดย
AUTHORIZEON	datetime	8	วันที่ตรวจสอบ Credit Card
AUTHORIZEBY	varchar	30	ตรวจสอบ Credit Card โดย
CHARGECC	varchar	1	ทำการการเก็บเงินชาร์ต Credit Card
CHARGECCON	datetime	8	วันที่ชาร์ต Credit Card
CHARGECCBY	varchar	30	ชาร์ต Credit Card โดย
DMCSTATUS	varchar	30	สถานะของ DMC
URGENT	varchar	1	Booking เร่งด่วนหรือไม่
DMCREVIEW	varchar	1	DMC ตรวจสอบแล้วหรือไม่
FULLYBOOK	varchar	1	ห้องพักเต็ม
LMB	varchar	1	Booking เร่งด่วนเป็นพิเศษหรือไม่
CXLCCON	datetime	8	ถูกยกเลิกเมื่อไร
CXLCCBY	varchar	30	ถูกยกเลิกโดย
PAYMENTTYPE	varchar	25	รูปแบบการชำระเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 6.3 CSDB.BKCLIENT เป็นตารางที่ใช้เก็บข้อมูลลูกค้าแต่ละ Booking

NAME	DATATYPE	SIZE	DESCRIPTION
REFNO	varchar	20	หมายเลขอ้างอิงของ Booking
CUSTID	varchar	10	รหัสลูกค้า
LASTNAME	varchar	30	ชื่อสกุลของลูกค้า
FIRSTNAME	varchar	30	ชื่อของลูกค้า
SALUTE	varchar	25	คำนำหน้าชื่อของลูกค้า
EMAIL	varchar	60	อีเมลล์
TELAR	varchar	5	รหัสพื้นที่โทรศัพท์
TELNO	varchar	20	เบอร์โทรศัพท์
NATIONALITY	varchar	50	เชื้อชาติ
RESIDENCY	varchar	50	ประเทศที่พักอาศัยอยู่

ตารางที่ 6.4 CSDB.BKCHARGE เป็นตารางที่ใช้เก็บข้อมูลเกี่ยวกับการค่าใช้จ่ายของ Booking

NAME	DATATYPE	SIZE	DESCRIPTION
REFNO	varchar	20	หมายเลขอ้างอิงของ Booking
DISCOUNT	smallmoney	4	ส่วนลด
CUSTTOTAL	money	8	ราคารวมที่เก็บจากลูกค้า
RATEID	int	4	รหัส Rate

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 6.5 CSDB.TBL\_XML\_BOOKING\_INFO เป็นตารางที่ใช้เก็บข้อมูลการเชื่อมโยงของ Booking ที่ใช้การส่ง XML Document ติดต่อกับ Supplier เพื่อดำเนินธุรกรรมการสำรองห้องพัก

PHYSICAL NAME	DATATYPE	SIZE	DESCRIPTION
PH_REFNO	varchar	10	หมายเลขอ้างอิงของ Booking
AGENT_REFNO	varchar	10	หมายเลขอ้างอิงของตัวแทนขาย
DMC_REFNO	varchar	10	หมายเลขอ้างอิงของ DMC
DMCCODE	varchar	10	รหัส DMC
COUNTRY_CODE	varchar	10	รหัสประเทศ
DEPARTURE_DATE	datetime	8	วันเข้าพัก

## 6.2 การคัดเลือกข้อมูล

การคัดเลือก Attributes ต่างๆเพื่อนำมาใช้วิเคราะห์เป็นงานส่วนที่มีความสำคัญมาก ทั้งนี้หากเลือก Attribute ที่ไม่มีความสำคัญหรือไม่เหมาะสมก็จะมีผลต่อโมเดลที่จะได้ ดังนั้นในการคัดเลือกจึงจำเป็นต้องใช้ความระมัดระวังอย่างมาก เพื่อที่จะได้ Attribute ที่มีแนวโน้มว่าจะเกี่ยวข้องกับผลลัพธ์เป้าหมายที่ต้องการ เพื่อที่จะนำ Attributes เหล่านั้นมาใช้ในการเทรนนิ่งในขั้นตอนต่อไป โดยได้ทำการคัดเลือก Attributes จากตารางต่างๆมีรายละเอียดดังต่อไปนี้

6.1.1 จากตาราง BKINFO มี Attributes ที่เลือกมาดังนี้ Refno, Brefreq, Smoking, Floor, Rmtype, Indate, Outdate, Country

6.1.2 จากตาราง BKSTATUS มี Attributes ที่เลือกดังนี้ Amend, Chargecc , Urgent , Dmcstatus , Madeon, Confon

6.1.3 จากตาราง BKCLIENT มี Attributes ที่เลือกมาดังนี้ Salute, Nationality, Resident

6.1.4 จากตาราง BKCHARGE มี Attributes ที่เลือกมาดังนี้ Custtotal

6.1.5 จากตาราง TBL\_XML\_BOOKING\_INFO มี Attribute ที่เลือกมาคือ ph\_refno

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากข้อมูลการสำรองห้องพักแบ่งเป็นสามประเภทใหญ่ๆคือ Normal Booking, Instant Booking และ XML Booking แต่ไม่ได้มีการแยกประเภทของบ็อกกิ้งอย่างชัดเจนใน Database จึงจำเป็นต้องใช้ S Q L C o m m a n d ช่วยในการแยกประเภทของข้อมูลสำรองห้องพัก โดยมีหลักการแยกดังนี้

ข้อมูลการสำรองห้องพักประเภท Instant จะมีค่าในคอลัมน์ STATUS ในตาราง BKSTATUS เป็น instant จึงใช้ คำสั่ง SQL ดังนี้

```
select * from csdb.csdb_user.bkstatus where status = 'instant'
```

สำหรับข้อมูลการสำรองห้องพักประเภท XML จะมี Attribute REFNO อ้างอิงไปยังตาราง csdb.dbo.tbl\_xml\_booking\_info จึงใช้ คำสั่ง SQL ดังนี้

Booking อื่นนอกเหนือจากสองประเภทที่กล่าวมาถือได้ว่าเป็น Normal Booking จึงนำเอาข้อมูล Booking มากรองส่วนที่ไม่ใช่ instant และ XML ออกไป โดยใช้คำสั่ง SQL เงื่อนไข

```
select * from csdb.csdb_user.bkinfo where Where (Status <> 'Instant') and (refno not in (select ph_refno from csdb.dbo.tbl_xml_booking_info))
```

### 6.3 การทำความสะอาดข้อมูล

เป็นกระบวนการที่อยู่ในส่วนการเตรียมข้อมูล โดยจะทำการทำความสะอาดข้อมูลหรือตัดข้อมูลที่ไม่ว่างต้องออกด้วย เพื่อให้ได้เฉพาะข้อมูลที่ดี ที่จะนำเข้าไปใช้ในการเทรนนิ่ง โดยการดำเนินการตัดข้อมูลที่ไม่ว่างต้องทิ้งมีเงื่อนไขดังนี้

- Attribute ที่มีค่าเท่ากับ 0 หรือ Null
- Attribute ที่มีค่าไม่ตรงกับความต้องการ เช่น N/A

### 6.4 การแปลงข้อมูล

ซึ่งจากการวิเคราะห์การคัดเลือกข้อมูล เห็นว่าข้อมูลที่เก็บไว้ในฐานข้อมูลบางส่วน ต้องการการประมวลผลบางอย่างเพื่อที่จะได้ข้อมูลที่นัยสำคัญและนำมาใช้ในการวิเคราะห์ได้ จึงได้ใช้คำสั่ง SQL เพื่อให้ได้ผลลัพธ์ตามต้องการ ได้แก่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำนวนวันระหว่างวันที่ลูกค้ายื่นใช้บริการสำรองห้องพักกับวันที่เข้าพัก ใช้คำสั่ง SQL

```
Select datediff (d,madeon, indate) from csdb_user.bkinfo A inner join
csdb_user.bkstatus B on (A.refno = B.refno)
```

ข้อมูลใน Attribute ที่เป็นจำนวนเงิน ซึ่งมีค่าต่อเนื่องกันมาก จึงนำมาแปลงให้อยู่ในรูปของ
กลุ่มข้อมูลโดย Custtotal ได้นำมาจัดกลุ่มดังนี้

กลุ่มที่ 1 เป็นกลุ่มข้อมูลสำรองห้องพักที่มีราคารวม < 200

กลุ่มที่ 2 เป็นกลุ่มข้อมูลสำรองห้องพักที่มีราคารวมสูงกว่า 200 เหรียญ (USD)

โดยใช้คำสั่ง SQL เพื่อจัดข้อมูลการสำรองห้องพักเป็นกลุ่มต่างๆ โดยคำสั่ง SQL

```
TOTALPRICE = Case
when (D.Custtotal <= 200) Then 'LOWPRICE'
Else 'HIGHPRICE'
End ,
```

### 6.5 การนำข้อมูลเข้าตารางข้อมูลสำหรับการเทรนนิ่ง

โดยหลังจากทำการคัดเลือก Attribute ที่ต้องการได้แล้ว จึงใช้คำสั่ง SQL เพื่อนำข้อมูลจาก
ตารางต่างๆมารวมกันเป็นตารางเดียว โดยมีคำสั่ง SQL ดังนี้

```
Select Top 5000
A.REFNO, B.CHARGECC, A.Rmocc AS ROOMOCC, B.URGENT,
A.Noadult AS PERSON, A.SMOKING, A.BEDTYPE, A.FLOOR,
A.Norm AS NO_OF_ROOM,
SEX = Case
when (C.SALUTE='MR.') Then 'M'
when (C.SALUTE='MISS.') or (C.SALUTE='MRS.') or
```

```

(C.SALUTE='MS.') Then 'F'
    Else 'M'
    End
, F.CONTINENT_NAME AS HOTEL_CONTINENT,
CUST_CONTINENT= Case
    when (H.CONTINENT_NAME='Africa') Then 'Africa'
    when (H.CONTINENT_NAME='Asia') Then 'Asia'
    when (H.CONTINENT_NAME='Central America And Caribbean') Then
'America'
    when (H.CONTINENT_NAME='Europe') Then 'Europe'
    when (H.CONTINENT_NAME='Middle East') Then 'Middle East'
    when (H.CONTINENT_NAME='North America') Then 'America'
    when (H.CONTINENT_NAME='Pacific Ocean And Australia') Then
'Pacific'
    when (H.CONTINENT_NAME='South America') Then 'America'
    Else 'Asia'
    End
,
NODAYBOOKTOCHECKIN_TYPE = Case
    when (IsNull(Datediff(D, B.Madeon, A.Indate),0) <= 15) Then 'SHORT'
    when (IsNull(Datediff(D, B.Madeon, A.Indate),0) > 15) and
(Isnull(Datediff(D, B.Madeon, A.Indate),0) <= 40)Then 'MEDIUM'
    Else 'LONG'
    End
,
TOTALPRICE = Case
    when (D.Custtotal <= 200) Then 'LOWPRICE'
    Else 'HIGHPRICE'
    End
,
BOOKING_TYPE = 'INSTANT'
From    dbo.Bkinfo as A

```

```

Inner Join dbo.Bkstatus as B On (A.REFNO = B.REFNO)

```

```

Inner Join dbo.Bkclient as C On (A.REFNO = C.REFNO)
Inner Join      dbo.Bkcharge as D On (A.REFNO = D.REFNO)
Inner Join dbo.TBL_COUNTRIES AS E On
(A.COUNTRY=E.COUNTRY_NAME)
Inner Join dbo.TBL_CONTINENTS AS F On (E.CONTINENT_ID =
F.CONTINENT_ID)
Inner Join dbo.TBL_COUNTRIES AS G On
(C.NATIONALITY=G.COUNTRY_NAME)
Inner Join dbo.TBL_CONTINENTS AS H On (G.CONTINENT_ID =
H.CONTINENT_ID)
Where (B.Fullybook <> 'Y') and (A.REFNO in (select ph_REFNO from
dbo.tbl_xml_booking_info))
Union
Select Top 5000
A.REFNO, B.CHARGECC, A.Rmocc AS ROOMOCC, B.URGENT,
A.Noadult AS PERSON, A.SMOKING, A.BEDTYPE, A.FLOOR,
A.Norm AS NO_OF_ROOM,
SEX = Case
when (C.SALUTE='MR.') Then 'M'
when (C.SALUTE='MISS.') or (C.SALUTE='MRS.') or
(C.SALUTE='MS.') Then 'F'
Else 'M'
End ,
F.CONTINENT_NAME AS HOTEL_CONTINENT,
CUST_CONTINENT= Case
when (H.CONTINENT_NAME='Africa') Then 'Africa'
when (H.CONTINENT_NAME='Asia') Then 'Asia'
when (H.CONTINENT_NAME='Central America And Caribbean') Then
'America'
when (H.CONTINENT_NAME='Europe') Then 'Europe'

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับอาจารย์และบุคลากรศึกษาเท่านั้น ไม่อนุญาตให้ทำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

when (H.CONTINENT_NAME='Middle East') Then 'Middle East'
when (H.CONTINENT_NAME='North America') Then 'America'
when (H.CONTINENT_NAME='Pacific Ocean And Australia') Then
'Pacific'
when (H.CONTINENT_NAME='South America') Then 'America'
Else 'Asia'
End ,
NODAYBOOKTOCHECKIN_TYPE = Case
when (IsNull(Datediff(D, B.Madeon, A.Indate),0) <= 15) Then 'SHORT'
when (IsNull(Datediff(D, B.Madeon, A.Indate),0) > 15) and
(Isnull(Datediff(D, B.Madeon, A.Indate),0) <= 40)Then 'MEDIUM'
Else 'LONG'
End ,
TOTALPRICE = Case
when (D.Custtotal <= 200) Then 'LOWPRICE'
Else 'HIGHPRICE'
End ,
BOOKING_TYPE = 'INSTANT'
From dbo.Bkinfo as A
Inner Join dbo.Bkstatus as B On (A.REFNO = B.REFNO)
Inner Join dbo.Bkelient as C On (A.REFNO = C.REFNO)
Inner Join dbo.Bkcharge as D On (A.REFNO = D.REFNO)
Inner Join dbo.TBL_COUNTRIES AS E On
(A.COUNTRY=E.COUNTRY_NAME)
Inner Join dbo.TBL_CONTINENTS AS F On (E.CONTINENT_ID =
F.CONTINENT_ID)
Inner Join dbo.TBL_COUNTRIES AS G On
(C.NATIONALITY=G.COUNTRY_NAME)
Inner Join dbo.TBL_CONTINENTS AS H On (G.CONTINENT_ID =
H.CONTINENT_ID)

```

Where (B.Fullybook <> 'Y') and (B.Status = 'Instant')

Union

Select Top 5000

A.REFNO, B.CHARGECC, A.Rmocc AS ROOMOCC, B.URGENT,  
A.Noadult AS PERSON, A.SMOKING, A.BEDTYPE, A.FLOOR,  
A.Norm AS NO\_OF\_ROOM,

SEX = Case

when (C.SALUTE='MR.') Then 'M'

when (C.SALUTE='MISS.') or (C.SALUTE='MRS.') or

(C.SALUTE='MS.') Then 'F'

Else 'M'

End ,

F.CONTINENT\_NAME AS HOTEL\_CONTINENT,

CUST\_CONTINENT= Case

when (H.CONTINENT\_NAME='Africa') Then 'Africa'

when (H.CONTINENT\_NAME='Asia') Then 'Asia'

when (H.CONTINENT\_NAME='Central America And Caribbean') Then  
'America'

when (H.CONTINENT\_NAME='Europe') Then 'Europe'

when (H.CONTINENT\_NAME='Middle East') Then 'Middle East'

when (H.CONTINENT\_NAME='North America') Then 'America'

when (H.CONTINENT\_NAME='Pacific Ocean And Australia') Then  
'Pacific'

when (H.CONTINENT\_NAME='South America') Then 'America'

Else 'Asia'

End ,

NODAYBOOKTOCHECKIN\_TYPE = Case

when (IsNull(Datediff(D, B.Madeon, A.Indate),0) <= 15) Then 'SHORT'

when (IsNull(Datediff(D, B.Madeon, A.Indate),0) > 15) and

(IsNull(Datediff(D, B.Madeon, A.Indate),0) <= 40)Then 'MEDIUM'

```

Else 'LONG'
End ,
TOTALPRICE = Case
when (D.Custtotal <= 200) Then 'LOWPRICE'
Else 'HIGHPRICE'
End ,
BOOKING_TYPE = 'NORMAL'

From dbo.Bkinfo as A
Inner Join dbo.Bkstatus as B On (A.REFNO = B.REFNO)
Inner Join dbo.Bkclient as C On (A.REFNO = C.REFNO)
Inner Join dbo.Bkcharge as D On (A.REFNO = D.REFNO)
Inner Join dbo.TBL_COUNTRIES AS E On
(A.COUNTRY=E.COUNTRY_NAME)
Inner Join dbo.TBL_CONTINENTS AS F On (E.CONTINENT_ID =
F.CONTINENT_ID)
Inner Join dbo.TBL_COUNTRIES AS G On
(C.NATIONALITY=G.COUNTRY_NAME)
Inner Join dbo.TBL_CONTINENTS AS H On (G.CONTINENT_ID =
H.CONTINENT_ID)
Where (B.Fullybook <> 'Y') and (B.Status <> 'Instant') and (A.REFNO not in
(select ph_REFNO from dbo.tbl_xml_booking_info))

Order By A.REFNO Desc

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากคำสั่ง SQL จะเชื่อมโยงตารางต่างๆด้วยหมายเลขอ้างอิง Booking (Ref no.) โดยจะคัดเลือกมาเฉพาะตามเงื่อนไขที่ตั้งไว้ ทำให้ได้ตารางใหม่ตามโครงสร้างที่วางไว้

ตารางที่ 6.6 ตารางที่ใช้ในการเทรนนิ่ง

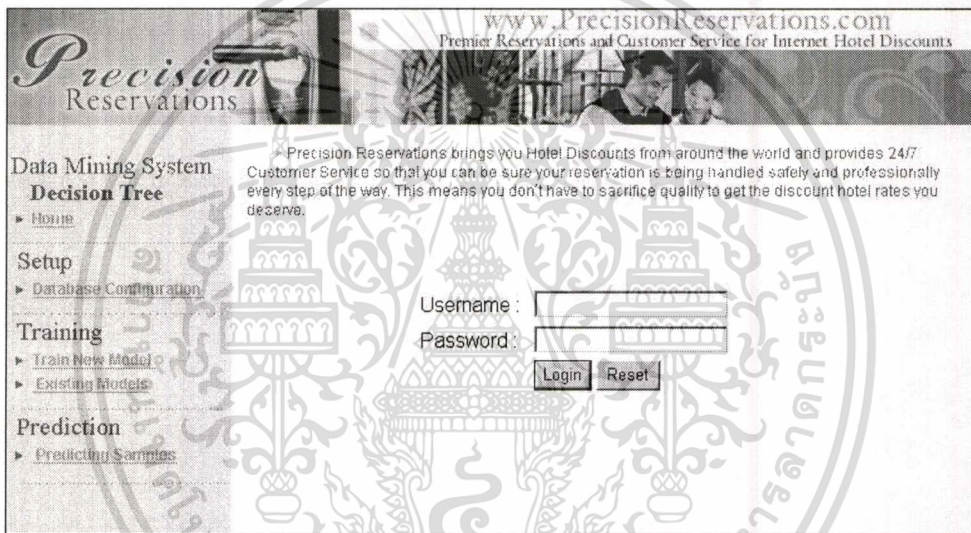
PHYSICAL NAME	DATATYPE	SIZE	DESCRIPTION
REFNO	varchar	20	หมายเลขอ้างอิง Booking
CHARGECC	varchar	1	การชาร์ต Credit Card
URGENT	varchar	1	Booking เร่งด่วน
SALUTE	varchar	25	เพศ
NODAYBOOKTOCHECKIN	varchar	4	ระยะห่างวันสำรองกับวันเข้าพัก
CUST_CONTINENT	varchar	4	ทวีปที่อยู่อาศัยของลูกค้า
CUSTTOTAL	varchar	8	ราคารวม
BOOKINGTYPE	varchar	7	ประเภทของ Booking

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 7

### การเทรนนิ่งและการทำนาย

จากบทที่ผ่านมาเป็นการนำเสนอกระบวนการในการจัดเตรียมข้อมูลเพื่อให้ได้ข้อมูลที่มีคุณภาพ ซึ่งเมื่อจัดเตรียมข้อมูลเรียบร้อยแล้ว จึงนำข้อมูลที่เตรียมไว้ นั้นมาทำการเทรนนิ่งเพื่อสร้างโมเดลสำหรับการค้นหาลูกค้าที่มีศักยภาพของธุรกิจสำรองห้องพักโรงแรม โดยในบทนี้จะกล่าวถึงกระบวนการทำงานเทรนนิ่งและนำเอาผลที่ได้จากการเทรนนิ่งนั้น ไปใช้ในทำการทำนายต่อไป

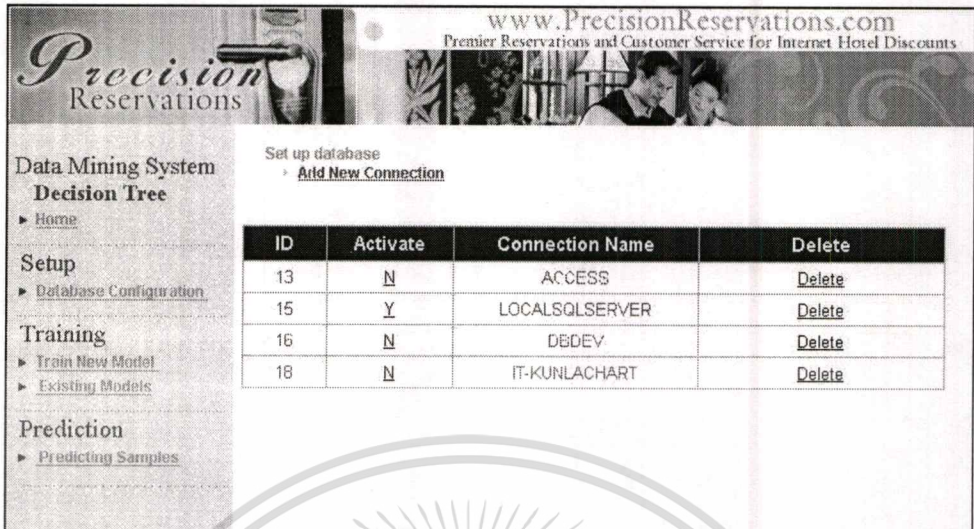


รูปที่ 7.1 หน้าจอเริ่มต้นของระบบ

หน้าจอแรกของระบบ Decision Tree สำหรับธุรกิจสำรองห้องพักโรงแรมจะมีเมนูต่างๆ ตามหมวดหมู่ ดังนี้

- Home : เมื่อดำเนินการไปยังหน้าจอต้อนรับของระบบ รวมถึงการ หน้าจอการ Login
- Database Configuration : เมื่อดำเนินการไปยังหน้าจอปรับแต่งค่าต่างๆของการติดต่อกับฐานข้อมูล
- Training New Model : เมื่อดำเนินการไปยังส่วนงานการเทรนนิ่งข้อมูล
- Existing Models : เมื่อดำเนินการไปยังหน้าจอลิส Model ที่ได้สร้างขึ้นมา
- Predicting Samples : เมื่อดำเนินการไปยังหน้าจอประมวลผลเปรียบเทียบเพื่อทำนายความเป็นไปได้ของกลุ่มตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

**Precision Reservations**

Data Mining System  
**Decision Tree**

- ▶ Home
- ▶ Setup
  - ▶ Database Configuration
- ▶ Training
  - ▶ Train New Model
  - ▶ Existing Models
- ▶ Prediction
  - ▶ Predicting Samples

Set up database  
▶ [Add New Connection](#)

ID	Activate	Connection Name	Delete
13	N	ACCESS	<a href="#">Delete</a>
15	Y	LOCALSQLSERVER	<a href="#">Delete</a>
16	N	DEDEV	<a href="#">Delete</a>
18	N	IT-KUNLACHART	<a href="#">Delete</a>

รูปที่ 7.2 หน้าจอปรับค่าการติดต่อฐานข้อมูล

หน้าจอปรับค่าการติดต่อฐานข้อมูล โดยจะสามารถทำการเพิ่ม ลบ และทำการเซตค่าให้ใช้งานของ Connection ได้

## 7.1 ส่วนงานการเทรนนิ่ง (Training Phase)

การเทรนนิ่ง เป็นกระบวนการนำข้อมูลที่ได้จากการเตรียมข้อมูลเข้าสู่การประมวลผลด้วยอัลกอริทึมของคาค่าไมนิ่ง สำหรับระบบนี้จะใช้อัลกอริทึมในการทำงานคือ ID3 โดยมีรายละเอียดดังนี้

### 7.1.1 กระบวนการเทรนนิ่ง

ก่อนจะทำการเทรนนิ่ง จะมีขั้นตอนในเลือกตารางเพื่อที่จะเลือก Target Attribute และ Descriptor Attributes เพื่อใช้ในการเทรนนิ่ง จากนั้นจึงทำการเทรนนิ่งข้อมูลตามอัลกอริทึม

www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

**Precision Reservations**

Data Mining System  
**Decision Tree**

- ▶ [Home](#)

Setup

- ▶ [Database Configuration](#)

Training

- ▶ [Train New Model](#)
- ▶ [Existing Models](#)

- ▶ [Reset](#)
- ▶ [Choose Table](#)
- ▶ [Choose Target Attribute](#)
- ▶ [Choose Descriptor Attributes](#)
- ▶ [Confirm Attributes for Mining Process](#)

รูปที่ 7.3 หน้าจอเริ่มต้นของกระบวนการเทรนนิ่ง

ขั้นตอนแรก เลือกคลิก List Table Schema เพื่อแสดงรายการตารางในฐานข้อมูล

www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

**Precision Reservations**

Data Mining System  
**Decision Tree**

- ▶ [Home](#)

Setup

- ▶ [Database Configuration](#)

Training

- ▶ [Train New Model](#)
- ▶ [Existing Models](#)

Prediction

- ▶ [Predicting Samples](#)

- ▶ [Reset](#)
- ▶ [Choose Table](#)
- ▶ [Choose Target Attribute](#)
- ▶ [Choose Descriptor Attributes](#)
- ▶ [Confirm Attributes for Mining Process](#)

There are 4 steps.  
In first step, please click List Table Schema Link to choose Table that you desire to use.

TABLE	TABLE_NAME	DATE_CREATED
<input type="radio"/>	TESTING	2/11/2004 8:17:38 AM
<input checked="" type="radio"/>	TRAININGDATA	2/12/2004 12:02:48 AM

รูปที่ 7.4 หน้าจอเลือกตารางที่ต้องการใช้ในการ Training

จากนั้นเลือกตารางที่ต้องการใช้ แล้วคลิก List Column Schema เพื่อแสดงรายการ Attributes ต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

**Precision Reservations**

Data Mining System  
**Decision Tree**

- ▶ Home

Setup

- ▶ Database Configuration

Training

- ▶ Train New Model
- ▶ Existing Models

Prediction

- ▶ Predicting Samples

- ▶ Reset
- ▶ Choose Table
- ▶ Choose Target Attribute
- ▶ Choose Descriptor Attributes
- ▶ Confirm Attributes for Mining Process

You have selected Table : TRAININGDATA

Next Step, Please select Target Attribute.

TARGET	COLUMN_NAME
<input checked="" type="checkbox"/>	REFNO
<input type="checkbox"/>	CHARGECC
<input type="checkbox"/>	ROOMOCC
<input type="checkbox"/>	URGENT

รูปที่ 7.5 หน้าจอเลือก Target Attribute เพื่อใช้ในการ Training

จากนั้นเลือก Attribute ที่ต้องการให้เป็น Target Attribute แล้วจึงกด Choose Target Attribute เพื่อทำการเลือก Descriptor Attributes ต่อไป

www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

**Precision Reservations**

Data Mining System  
**Decision Tree**

- ▶ Home

Setup

- ▶ Database Configuration

Training

- ▶ Train New Model
- ▶ Existing Models

Prediction

- ▶ Predicting Samples

- ▶ Reset
- ▶ Choose Table
- ▶ Choose Target Attribute
- ▶ Choose Descriptor Attributes
- ▶ Confirm Attributes for Mining Process

You have selected Table : TRAININGDATA  
Selected Target Attribute : REFNO

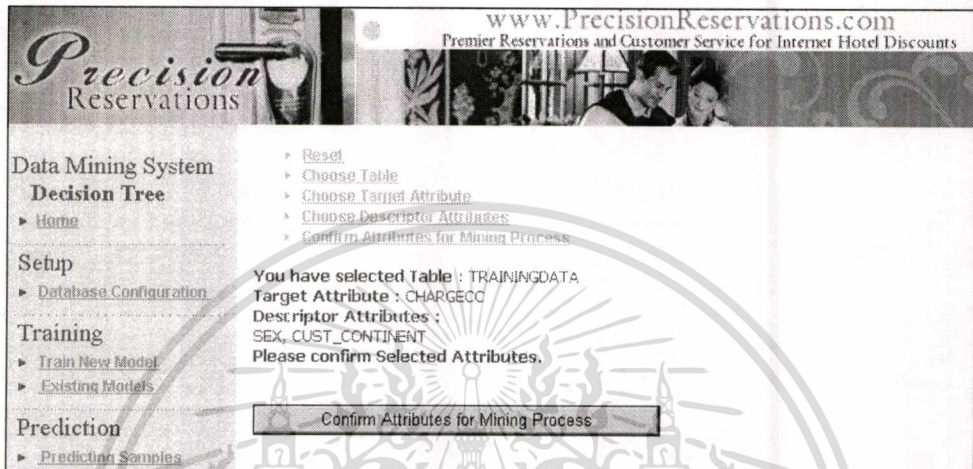
Next Step, Please select Descriptor Attributes.

	COLUMN_NAME
<input type="checkbox"/>	REFNO
<input type="checkbox"/>	CHARGECC
<input checked="" type="checkbox"/>	ROOMOCC
<input type="checkbox"/>	URGENT
<input checked="" type="checkbox"/>	PERSON
<input type="checkbox"/>	SMOYING

รูปที่ 7.6 หน้าจอเลือก Descriptor Attributes เพื่อใช้ในการ Training

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากเลือก Attributes ที่ต้องการให้เป็น Descriptor Attributes แล้วจึงกด Choose Descriptor Attributes เพื่อเข้าสู่หน้าจอสรุป Attributes ต่างๆที่ได้ทำการเลือกมาก่อนที่จะทำการ Process ตามอัลกอริทึม



รูปที่ 7.7 หน้าจอยืนยันค่า Attributes ที่จะใช้ในการ Mining

หน้าจอสรุปรายละเอียดที่ได้เลือกไป จากนั้นกด Confirm เพื่อยืนยันการประมวลผล จึงทำการ Training ข้อมูล เมื่อโปรแกรมประมวลผลเสร็จ โปรแกรมจะทำการเก็บค่า Configuration ต่างๆ ของโมเดลและผลลัพธ์ที่ได้จากการ Training เป็นไฟล์รูปแบบ XML ใน Folder Tree Results ตามรูปแบบ ปีเดือนวันตามด้วย Version

ตัวอย่างการเก็บค่า Configuration ต่างๆ ของโมเดลในรูปแบบ XML

```
<?xml version="1.0" standalone="yes"?><PARAMS>
  <PARAM>
    <NAME>CONNECTION</NAME>
    <VALUE>Provider=sqloledb;Data Source=localhost;Initial Catalog=DTDB;User
    Id=DTDB_user;Password=chart;</VALUE>
  </PARAM>
  <PARAM>
    <NAME>TABLE</NAME>
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    <VALUE>TRAININGDATA</VALUE>
  </PARAM>
  <PARAM>
    <NAME>TARGET_ATTRIBUTE</NAME>
    <VALUE>CHARGECC</VALUE>
  </PARAM>
  <PARAM>
    <NAME>ATTRIBUTE</NAME>
    <VALUE>SEX</VALUE>
  </PARAM>
  <PARAM>
    <NAME>ATTRIBUTE</NAME>
    <VALUE>CUST_CONTINENT</VALUE>
  </PARAM>
  <PARAM>
    <NAME>ATTRIBUTE</NAME>
    <VALUE>TOTALPRICE</VALUE>
  </PARAM>
</PARAMS>

```

ตัวอย่างผลลัพธ์ของโมเดลที่เก็บในรูปแบบ XML

```

<?xml version="1.0" standalone="yes"?>
<TreeNodes>
  <TreeNode>
    <NodeIndex>1</NodeIndex>
    <NodeParent>0</NodeParent>
    <NodeName>ROOT</NodeName>
    <NodeType>adRootNode</NodeType>

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

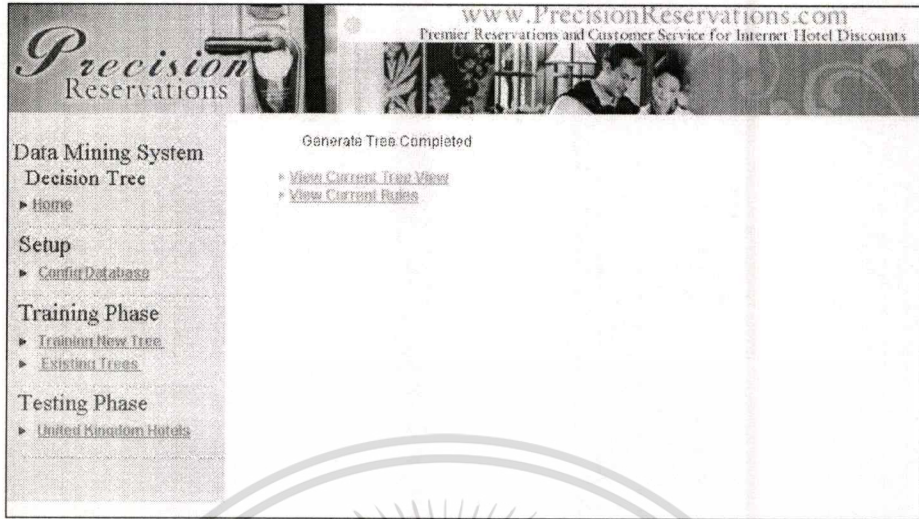
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

</TreeNode>
<TreeNode>
  <NodeIndex>2</NodeIndex>
  <NodeParent>1</NodeParent>
  <NodeName>CUST_CONTINENT_NAME</NodeName>
  <NodeType>adAttributeNode</NodeType>
</TreeNode>
<TreeNode>
  <NodeIndex>3</NodeIndex>
  <NodeParent>2</NodeParent>
  <NodeName>AFRICA</NodeName>
  <NodeType>adValueNode</NodeType>
</TreeNode>
<TreeNode>
  <NodeIndex>4</NodeIndex>
  <NodeParent>3</NodeParent>
  <NodeName>DMC_CONTINENT_NAME</NodeName>
  <NodeType>adAttributeNode</NodeType>
</TreeNode>
</TreeNodes>

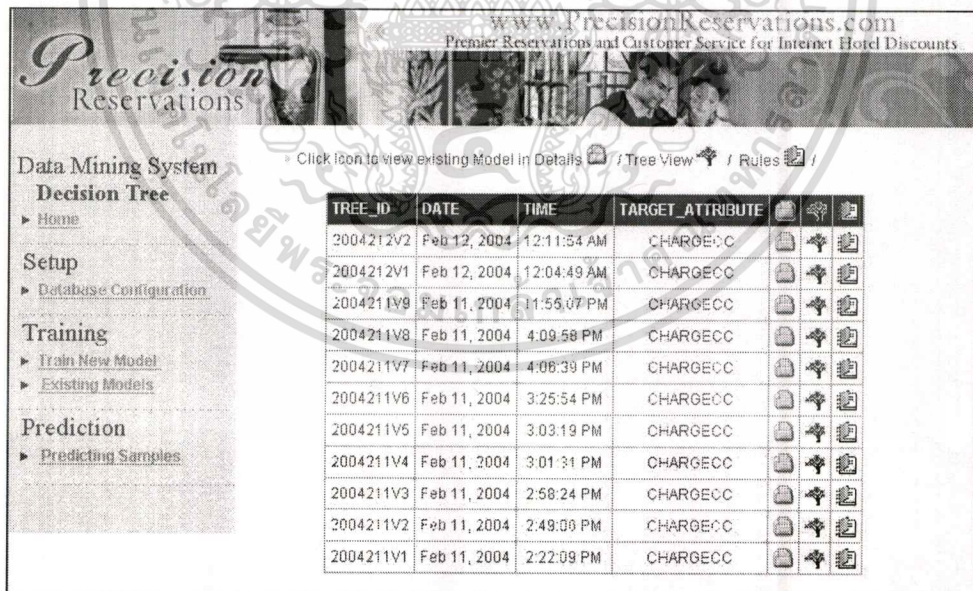
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 7.8 หน้าจอแสดงผลลัพธ์หลังจากเสร็จสิ้นการการสร้าง Model

หลังจากประมวลผลเรียบร้อยแล้ว หน้าจอจะแสดงลิงค์สำหรับการเปิดดูผลลัพธ์ของการ Training โดยจะแบ่งเป็นสองส่วนคือ Tree View กับ Rules List



รูปที่ 7.9 หน้าจอเลือกรายการ Model ต่างๆที่ได้ทดสอบมา

หน้าจอแสดงรายการการทดสอบต่างๆที่ผ่านมาซึ่งสามารถดูรายละเอียดของโมเดลในทั้งในรูปแบบ Tree View หรือ Rules

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

► Tree Details

CREATED : Feb 12, 2004 12:04:49 AM  
 TARGET\_ATTRIBUTE : CHARGECC  
 TARGET\_ATTRIBUTE\_VALUE : N,Y  
 ATTRIBUTES : SEX,CUST\_CONTINENT,TOTALPRICE

Decision Tree Results

Potential Customers Rules

ROOT

TOTALPRICE

LOWPRICE (59.81%)

CUST\_CONTINENT

ASIA (25.93%)

SEX

M (69.64%)

N 52.05% / Y 47.95%

F (30.36%)

N 60.00% / Y 40.00%

PACIFIC (10.19%)

SEX

M (68.85%)

Y 60.54% / N 39.46%

F (31.15%)

Y 59.40% / N 40.60%

AMERICA (8.42%)

SEX

M (71.67%)

Y 54.15% / N 45.85%

F (28.33%)

Y 64.00% / N 36.00%

EUROPE (52.17%)

SEX

M (75.49%)

Y 59.78% / N 40.22%

F (24.51%)

Y 56.34% / N 43.66%

MIDDLE EAST (2.00%)

SEX

M (85.71%)

N 52.78% / Y 47.22%

F (14.29%)

N 58.33% / Y 41.67%

AFRICA (1.29%)

SEX

M (81.48%)

N 63.64% / Y 36.36%

F (18.52%)

Y 50.00% / N 50.00%

รูปที่ 7.10 หน้าจอแสดงผลลัพธ์ในรูปแบบ Tree View

หน้าจอแสดงผลลัพธ์ในรูปแบบ Tree View โดยจะรับ Input Tree ID มาแล้วจึงไปดึงข้อมูล

จากไฟล์ XML ผลลัพธ์ที่เก็บไว้ก่อนหน้า เพื่อนำมาแสดงผลในรูปแบบต้นไม้ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

*Precision*  
Reservations

**Data Mining System**  
**Decision Tree**

▶ Home

**Setup**

▶ Database Configuration

**Training**

▶ Train New Model

▶ Existing Models

**Prediction**

▶ Predicting Samples

▶ Tree Details

CREATED : Feb 12, 2004 12:11:54 AM  
TARGET\_ATTRIBUTE : CHARGECC  
TARGET\_ATTRIBUTE\_VALUE : N,Y  
ATTRIBUTES : SEX,CUST\_CONTINENT,TOTALPRICE,BOOKING\_TYPE

1) IF TOTALPRICE = 'LOWPRICE' AND CUST\_CONTINENT = 'ASIA' AND BOOKING\_TYPE = 'INSTANT' AND SEX = 'M' THEN RESULT IS Y  
\* Total Samples Matched:278 / Correct:151 / Incorrect:128 / Ratio:54.12

2) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'ASIA' AND SEX = 'F' AND BOOKING\_TYPE = 'INSTANT' THEN RESULT IS N  
\* Total Samples Matched:90 / Correct:56 / Incorrect:34 / Ratio:62.22

3) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'ASIA' AND SEX = 'F' AND BOOKING\_TYPE = 'NORMAL' THEN RESULT IS N  
\* Total Samples Matched:70 / Correct:40 / Incorrect:30 / Ratio:57.14

4) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'PACIFIC' AND SEX = 'M' AND BOOKING\_TYPE = 'INSTANT' THEN RESULT IS N  
\* Total Samples Matched:81 / Correct:44 / Incorrect:37 / Ratio:54.32

5) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'PACIFIC' AND SEX = 'M' AND BOOKING\_TYPE = 'NORMAL' THEN RESULT IS N  
\* Total Samples Matched:79 / Correct:43 / Incorrect:36 / Ratio:54.43

6) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'PACIFIC' AND SEX = 'F' AND BOOKING\_TYPE = 'INSTANT' THEN RESULT IS N  
\* Total Samples Matched:55 / Correct:32 / Incorrect:23 / Ratio:58.18

7) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'PACIFIC' AND SEX = 'F' AND BOOKING\_TYPE = 'NORMAL' THEN RESULT IS N  
\* Total Samples Matched:21 / Correct:12 / Incorrect:9 / Ratio:57.14

8) IF TOTALPRICE = 'LOWPRICE' AND CUST\_CONTINENT = 'ASIA' AND BOOKING\_TYPE = 'INSTANT' AND SEX = 'F' THEN RESULT IS N  
\* Total Samples Matched:110 / Correct:62 / Incorrect:48 / Ratio:56.36

9) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'AMERICA' AND BOOKING\_TYPE = 'INSTANT' AND SEX = 'M' THEN RESULT IS N  
\* Total Samples Matched:82 / Correct:47 / Incorrect:35 / Ratio:57.32

### รูปที่ 7.11 หน้าจอแสดงผลลัพธ์ในรูปแบบ Rules

หน้าจอแสดงผลลัพธ์ในรูปแบบ Rules โดยจะรับ Input Tree ID มาแล้วจึงไปดึงข้อมูลจากไฟล์ XML ผลลัพธ์ที่เก็บไว้ก่อนหน้านี้ เพื่อนำมาแสดงผลในรูปแบบ Rules โดยในหน้าจอนี้ จะแสดงผลลัพธ์ตามเงื่อนไขต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 7.1.2 การวิเคราะห์ผลของ Model ที่ได้จากการเทรนนิ่ง

หลังจากทำการเทรนนิ่งข้อมูลจะได้ผลลัพธ์มาในรูปแบบกฎ IF-THEN ซึ่งเมื่อพิจารณา Decision Tree Model ที่ได้มา สามารถวิเคราะห์ความหมายของแต่ละกิ่งของต้นไม้ได้ โดยการทดสอบจะใช้ CHARGECC หรือการเรียกเก็บเงินโดยบัตรเครดิตเป็น Target Attribute และได้แบ่งการทดสอบออกเป็นกรณีต่างๆที่ใช้ Descriptor Attributes ที่แตกต่างกันเพื่อวิเคราะห์หาความสัมพันธ์ของ Target Attribute และ Descriptor Attributes ในมุมมองต่างกั้ดังนี้

#### 7.1.2.1 ทดสอบโดยใช้ เพศ (SEX)และราคารวม(TOTALPRICE)เป็น Descriptor Attributes ผลลัพธ์ที่ได้คือ

- 1) IF TOTALPRICE = 'HIGHPRICE' AND SEX = 'M' THEN RESULT IS Y (57.83%)
- 2) IF TOTALPRICE = 'HIGHPRICE' AND SEX = 'F' THEN RESULT IS N (50.29%)
- 3) IF TOTALPRICE = 'LOWPRICE' AND SEX = 'M' THEN RESULT IS Y (70.49%)
- 4) IF TOTALPRICE = 'LOWPRICE' AND SEX = 'F' THEN RESULT IS Y (62.18%)

แสดงให้เห็นถึงว่าราคาและเพศมีผลต่อการซื้อของลูกค้า จากผลลัพธ์ที่ได้พบว่า Booking ราคาสูง อัตราการตัดบัตรเครดิตจะน้อยกว่า ซึ่งแสดงให้เห็นว่าอีกว่า ลูกค้าที่เป็นหญิงให้ความสำคัญกับเรื่องของราคามากกว่าลูกค้าที่เป็นผู้ชาย

#### 7.1.2.2 ทดสอบโดยใช้ เพศ (SEX)และช่วงห่างของวันเริ่ม Booking กับวันที่จะเข้าพักโรงแรม (CHECKINPERIOD)เป็น Descriptor Attributes ผลลัพธ์ที่ได้คือ

- 1) IF CHECKINPERIOD = 'SHORT' AND SEX = 'M' THEN RESULT IS Y (79.43%)
- 2) IF CHECKINPERIOD = 'SHORT' AND SEX = 'F' THEN RESULT IS Y (73.12%)
- 3) IF CHECKINPERIOD = 'LONG' AND SEX = 'M' THEN RESULT IS N (55.30%)
- 4) IF CHECKINPERIOD = 'LONG' AND SEX = 'F' THEN RESULT IS N (60.83%)
- 5) IF CHECKINPERIOD = 'MEDIUM' AND SEX = 'M' THEN RESULT IS Y (64.64%)
- 6) IF CHECKINPERIOD = 'MEDIUM' AND SEX = 'F' THEN RESULT IS Y (59.56%)

แสดงให้เห็นถึงว่าช่วงห่างของวันเริ่ม Booking กับวันที่จะเข้าพักโรงแรม มีผลต่อการซื้อของลูกค้า ซึ่งเห็นได้ว่า Bookingที่ระยะเวลายาวนาน อัตราการยกเลิกจะสูงกว่า Booking ที่ระยะเวลาดสั้น ในส่วนปัจจัยเรื่องเพศ จะเห็นได้ว่า ระยะเวลาในการ Booking จะมีผลต่อการยกเลิกของผู้หญิงมากกว่าผู้ชาย

**7.1.2.3 ทดสอบโดยใช้ ประเภทของ Booking (BOOKING\_TYPE)และเพศ (SEX) เป็น Descriptor Attributes ผลลัพธ์ที่ได้คือ**

- 1) IF BOOKING\_TYPE = 'NORMAL' AND SEX = 'M' THEN RESULT IS N (61.61%)
- 2) IF BOOKING\_TYPE = 'NORMAL' AND SEX = 'F' THEN RESULT IS N (63.62%)
- 3) IF BOOKING\_TYPE = 'INSTANT' AND SEX = 'M' THEN RESULT IS Y (50.14%)
- 4) IF BOOKING\_TYPE = 'INSTANT' AND SEX = 'F' THEN RESULT IS N (53.58%)

แสดงให้เห็นถึงว่าประเภทของ Booking มีผลต่อการตัดสินใจซื้อของลูกค้าผู้ชายมากกว่าผู้หญิง ซึ่งจากกฎแสดงให้เห็นว่าผู้ชายจะมีอัตราการยืนยันการซื้อใน Booking ประเภทที่มีสำรองที่นั่งแน่นอน (INSTANT BOOKING) มากกว่าผู้หญิง หรืออีกนัยหนึ่งก็คือ หากลูกค้าทราบว่าได้ห้องพักแน่นอนแล้ว ลูกค้าผู้ชายมีโอกาสที่จะตัดสินใจซื้อเลยได้มากกว่าลูกค้าผู้หญิงนั่นเอง

**7.1.2.4 ทดสอบโดยใช้ ใช้ ทวีปที่อยู่อาศัยของลูกค้า (CUST\_CONTINENT)และราคารวม (TOTALPRICE) เป็น Descriptor Attribute ผลลัพธ์ที่ได้คือ**

- 1) IF CUST\_CONTINENT = 'PACIFIC' AND TOTALPRICE = 'LOWPRICE' THEN RESULT IS Y (62.67%)
- 2) IF CUST\_CONTINENT = 'PACIFIC' AND TOTALPRICE = 'HIGHPRICE' THEN RESULT IS N (56.49%)
- 3) IF CUST\_CONTINENT = 'AMERICA' AND TOTALPRICE = 'LOWPRICE' THEN RESULT IS Y (54.39%)
- 4) IF CUST\_CONTINENT = 'AMERICA' AND TOTALPRICE = 'HIGHPRICE' THEN RESULT IS N (57.62%)
- 5) IF CUST\_CONTINENT = 'EUROPE' AND TOTALPRICE = 'LOWPRICE' THEN RESULT IS Y (58.68%)
- 6) IF CUST\_CONTINENT = 'EUROPE' AND TOTALPRICE = 'HIGHPRICE' THEN RESULT IS N (52.36%)
- 7) IF CUST\_CONTINENT = 'MIDDLE EAST' AND TOTALPRICE = 'LOWPRICE' THEN RESULT IS N (52.14%)
- 8) IF CUST\_CONTINENT = 'MIDDLE EAST' AND TOTALPRICE = 'HIGHPRICE' THEN RESULT IS N (72.17%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9) IF CUST\_CONTINENT = 'AFRICA' AND TOTALPRICE = 'LOWPRICE' THEN  
RESULT IS N (64.84%)

10) IF CUST\_CONTINENT = 'AFRICA' AND TOTALPRICE = 'HIGHPRICE' THEN  
RESULT IS N (78.38%)

11) IF CUST\_CONTINENT = 'ASIA' AND TOTALPRICE = 'LOWPRICE' THEN  
RESULT IS N (55.30%)

12) IF CUST\_CONTINENT = 'ASIA' AND TOTALPRICE = 'HIGHPRICE' THEN  
RESULT IS N (67.07%)

แสดงให้เห็นถึงว่าลูกค้าหลักของธุรกิจ คือลูกค้าที่อยู่ในทวีป PACIFIC , AMERICA และยุโรป โดยราคามีผลต่อการตัดสินใจซื้อของลูกค้าอย่างมาก

**7.1.2.5 ทดสอบโดยใช้ ทวีปที่อยู่อาศัยของลูกค้า(CUST\_CONTINENT) และประเภทของBooking (BOOKING\_TYPE)เป็น Descriptor Attribute ผลลัพธ์ที่ได้คือ**

1) IF CUST\_CONTINENT = 'PACIFIC' AND BOOKING\_TYPE = 'INSTANT' THEN  
RESULT IS Y (52.00%)

2) IF CUST\_CONTINENT = 'PACIFIC' AND BOOKING\_TYPE = 'NORMAL' THEN  
RESULT IS N (55.79%)

3) IF CUST\_CONTINENT = 'AMERICA' AND BOOKING\_TYPE = 'INSTANT' THEN  
RESULT IS Y (51.39%)

4) IF CUST\_CONTINENT = 'AMERICA' AND BOOKING\_TYPE = 'NORMAL' THEN  
RESULT IS N (60.93%)

5) IF CUST\_CONTINENT = 'EUROPE' AND BOOKING\_TYPE = 'INSTANT' THEN  
RESULT IS Y (51.14%)

6) IF CUST\_CONTINENT = 'EUROPE' AND BOOKING\_TYPE = 'NORMAL' THEN  
RESULT IS N (56.00%)

7) IF CUST\_CONTINENT = 'MIDDLE EAST' AND BOOKING\_TYPE = 'INSTANT'  
THEN RESULT IS N (56.14%)

8) IF CUST\_CONTINENT = 'MIDDLE EAST' AND BOOKING\_TYPE = 'NORMAL'  
THEN RESULT IS N (79.15%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 9) IF CUST\_CONTINENT = 'AFRICA' AND BOOKING\_TYPE = 'INSTANT' THEN  
RESULT IS N (74.19%)
- 10) IF CUST\_CONTINENT = 'AFRICA' AND BOOKING\_TYPE = 'NORMAL' THEN  
RESULT IS N (77.23%)
- 11) IF CUST\_CONTINENT = 'ASIA' AND BOOKING\_TYPE = 'INSTANT' THEN  
RESULT IS N (55.24%)
- 12) IF CUST\_CONTINENT = 'ASIA' AND BOOKING\_TYPE = 'NORMAL' THEN  
RESULT IS N (73.02%)

แสดงให้เห็นถึงว่าใช้ ทวีปที่อยู่อาศัยของลูกค้ากับประเภทของBooking มีผลต่อการซื้อของลูกค้าชัดเจนเป็นบางกรณี กรณีนี้ขอเน้นที่กลุ่มลูกค้าหลักของธุรกิจที่ได้มาจากผลลัพธ์ข้อ 7.1.2.4 นั่นคือลูกค้าที่อยู่ในทวีป PACIFIC, AMERICA และลูกค้าในทวีป EUROPE จากกฎที่ได้ สังเกตว่าลูกค้าที่อยู่ในทวีปอเมริกา มีอัตราการยืนยันการซื้อใน Booking ประเภท มีสินค้าขึ้นชั้น (INSTANT) มากกว่า Booking ปกติ ในขณะที่ลูกค้าที่อยู่ในทวีปยุโรป มีอัตราการยืนยันการซื้อใน Booking ประเภท มีสินค้า ไม่ต่างกันมากนัก แสดงให้เห็นถึงว่า ลูกค้าที่อยู่ในทวีปอเมริกาให้ความสำคัญกับความมั่นใจในการได้รับสินค้าอย่างแน่นอน มากกว่าลูกค้าในทวีปยุโรป

**7.1.2.6 ทดสอบโดยใช้ช่วงห่างของวันเริ่ม Booking กับวันที่จะเข้าพักโรงแรม (CHECKINPERIOD) และทวีปที่อยู่อาศัยของลูกค้า(CUST\_CONTINENT)เป็น Descriptor Attribute ผลลัพธ์ที่ได้คือ**

- 1) IF CHECKINPERIOD = 'MEDIUM' AND CUST\_CONTINENT = 'AMERICA' THEN  
RESULT IS Y (63.55%)
- 2) IF CHECKINPERIOD = 'MEDIUM' AND CUST\_CONTINENT = 'EUROPE' THEN  
RESULT IS Y (69.28%)
- 3) IF CHECKINPERIOD = 'MEDIUM' AND CUST\_CONTINENT = 'MIDDLE EAST'  
THEN RESULT IS N (55.00%)
- 4) IF CHECKINPERIOD = 'MEDIUM' AND CUST\_CONTINENT = 'AFRICA' THEN  
RESULT IS Y (50.00%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 5) IF CHECKINPERIOD = 'SHORT' AND CUST\_CONTINENT = 'ASIA' THEN  
RESULT IS Y (67.34%)
- 6) IF CHECKINPERIOD = 'SHORT' AND CUST\_CONTINENT = 'PACIFIC' THEN  
RESULT IS Y (78.19%)
- 7) IF CHECKINPERIOD = 'SHORT' AND CUST\_CONTINENT = 'AMERICA' THEN  
RESULT IS Y (81.82%)
- 8) IF CHECKINPERIOD = 'SHORT' AND CUST\_CONTINENT = 'EUROPE' THEN  
RESULT IS Y (84.11%)
- 9) IF CHECKINPERIOD = 'SHORT' AND CUST\_CONTINENT = 'MIDDLE EAST'  
THEN RESULT IS Y (77.97%)
- 10) IF CHECKINPERIOD = 'SHORT' AND CUST\_CONTINENT = 'AFRICA' THEN  
RESULT IS Y (68.00%)
- 11) IF CHECKINPERIOD = 'LONG' AND CUST\_CONTINENT = 'ASIA' THEN  
RESULT IS N (64.32%)
- 12) IF CHECKINPERIOD = 'LONG' AND CUST\_CONTINENT = 'PACIFIC' THEN  
RESULT IS N (58.33%)
- 13) IF CHECKINPERIOD = 'LONG' AND CUST\_CONTINENT = 'AMERICA' THEN  
RESULT IS N (65.22%)
- 14) IF CHECKINPERIOD = 'LONG' AND CUST\_CONTINENT = 'EUROPE' THEN  
RESULT IS N (53.93%)
- 15) IF CHECKINPERIOD = 'LONG' AND CUST\_CONTINENT = 'MIDDLE EAST'  
THEN RESULT IS N (61.90%)
- 16) IF CHECKINPERIOD = 'LONG' AND CUST\_CONTINENT = 'AFRICA' THEN  
RESULT IS N (65.00%)
- 17) IF CHECKINPERIOD = 'MEDIUM' AND CUST\_CONTINENT = 'ASIA' THEN  
RESULT IS Y (52.25%)
- 18) IF CHECKINPERIOD = 'MEDIUM' AND CUST\_CONTINENT = 'PACIFIC' THEN  
RESULT IS Y (64.24%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แสดงให้เห็นถึงว่าช่วงห่างของวันเริ่ม Booking กับวันที่จะเข้าพักโรงแรม มีผลต่อการซื้อของลูกค้า ซึ่งเห็นได้ว่า Booking ที่ระยะเวลายาวนาน อัตราการยกเลิกจะสูงกว่า Booking ที่ระยะเวลานั้น ในส่วนปัจจัยทวีปที่พักอาศัยของลูกค้ามีผลที่ไม่แตกต่างกันมากนัก

**7.1.2.7 ทดสอบโดยใช้ ทวีปที่อยู่อาศัยของลูกค้า(CUST\_CONTINENT)และทวีปของโรงแรมที่สำรองห้องพัก (HOTEL\_CONTINENT) เป็น Descriptor Attributes ผลลัพธ์ที่ได้คือ**

- 1) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'EUROPE' THEN RESULT IS N (55.81%)
- 2) IF CUST\_CONTINENT = 'AFRICA' AND HOTEL\_CONTINENT = 'EUROPE' THEN RESULT IS N (80.00%)
- 3) IF CUST\_CONTINENT = 'AFRICA' AND HOTEL\_CONTINENT = 'CENTRAL AMERICA AND CARIBBEAN' THEN RESULT IS N (100.00%)
- 4) IF CUST\_CONTINENT = 'AFRICA' AND HOTEL\_CONTINENT = 'MIDDLE EAST' THEN RESULT IS N (75.00%)
- 5) IF CUST\_CONTINENT = 'AFRICA' AND HOTEL\_CONTINENT = 'SOUTH AMERICA' THEN RESULT IS Y (66.67%)
- 6) IF CUST\_CONTINENT = 'AFRICA' AND HOTEL\_CONTINENT = 'AFRICA' THEN RESULT IS N (75.00%)
- 7) IF CUST\_CONTINENT = 'AFRICA' AND HOTEL\_CONTINENT = 'PACIFIC OCEAN AND AUSTRALIA' THEN RESULT IS N (100.00%)
- 8) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'CENTRAL AMERICA AND CARIBBEAN' THEN RESULT IS Y (100.00%)
- 9) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'MIDDLE EAST' THEN RESULT IS N (57.14%)
- 10) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'SOUTH AMERICA' THEN RESULT IS N (71.43%)
- 11) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'AFRICA' THEN RESULT IS Y (54.55%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 12) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'PACIFIC OCEAN AND AUSTRALIA' THEN RESULT IS Y (50.00%)
- 13) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'ASIA' THEN RESULT IS Y (55.88%)
- 14) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'NORTH AMERICA' THEN RESULT IS Y (71.43%)
- 15) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'EUROPE' THEN RESULT IS Y (68.75%)
- 16) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'CENTRAL AMERICA AND CARIBBEAN' THEN RESULT IS N (100.00%)
- 17) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'MIDDLE EAST' THEN RESULT IS Y (50.00%)
- 18) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'SOUTH AMERICA' THEN RESULT IS Y (70.00%)
- 19) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'AFRICA' THEN RESULT IS Y (57.14%)
- 20) IF CUST\_CONTINENT = 'PACIFIC' AND HOTEL\_CONTINENT = 'PACIFIC OCEAN AND AUSTRALIA' THEN RESULT IS N (52.73%)
- 21) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'ASIA' THEN RESULT IS N (50.06%)
- 22) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'NORTH AMERICA' THEN RESULT IS N (57.14%)
- 23) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'EUROPE' THEN RESULT IS Y (58.06%)
- 24) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'CENTRAL AMERICA AND CARIBBEAN' THEN RESULT IS Y (100.00%)
- 25) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'MIDDLE EAST' THEN RESULT IS N (58.33%)
- 26) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'SOUTH AMERICA' THEN RESULT IS N (56.52%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 27) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'AFRICA'  
THEN RESULT IS N (60.00%)
- 28) IF CUST\_CONTINENT = 'AMERICA' AND HOTEL\_CONTINENT = 'PACIFIC  
OCEAN AND AUSTRALIA' THEN RESULT IS N (76.92%)
- 29) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'ASIA' THEN  
RESULT IS N (59.81%)
- 30) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'ASIA' THEN  
RESULT IS Y (54.66%)
- 31) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'NORTH  
AMERICA' THEN RESULT IS Y (51.67%)
- 32) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'EUROPE'  
THEN RESULT IS Y (52.10%)
- 33) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'CENTRAL  
AMERICA AND CARIBBEAN' THEN RESULT IS Y (57.89%)
- 34) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'MIDDLE  
EAST' THEN RESULT IS Y (70.27%)
- 35) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'SOUTH  
AMERICA' THEN RESULT IS Y (55.75%)
- 36) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'AFRICA'  
THEN RESULT IS N (53.85%)
- 37) IF CUST\_CONTINENT = 'EUROPE' AND HOTEL\_CONTINENT = 'PACIFIC  
OCEAN AND AUSTRALIA' THEN RESULT IS N (65.85%)
- 38) IF CUST\_CONTINENT = 'MIDDLE EAST' AND HOTEL\_CONTINENT = 'ASIA'  
THEN RESULT IS N (61.06%)
- 39) IF CUST\_CONTINENT = 'ASIA' AND HOTEL\_CONTINENT = 'NORTH  
AMERICA' THEN RESULT IS N (57.89%)
- 40) IF CUST\_CONTINENT = 'MIDDLE EAST' AND HOTEL\_CONTINENT = 'NORTH  
AMERICA' THEN RESULT IS Y (100.00%)
- 41) IF CUST\_CONTINENT = 'MIDDLE EAST' AND HOTEL\_CONTINENT =  
'EUROPE' THEN RESULT IS N (66.67%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สังเกตได้ว่าโรงแรมส่วนใหญ่ที่ได้รับการยืนยันการจองคือโรงแรมในทวีปเอเชีย ยุโรป และลูกค้าเป็นลูกค้าในกลุ่มทวีปอเมริกา ยุโรป และ แปซิฟิก แต่หากสังเกตโดยละเอียดจะพบว่าลูกค้าที่อยู่ในทวีปเอเชีย ซึ่งจากผลลัพธ์ก่อนหน้านี้ ไม่ใช่กลุ่มลูกค้าที่มีศักยภาพ แต่จากการทดสอบนี้พบว่าลูกค้ากลุ่มนี้มีการยืนยันการจองโรงแรมที่อยู่ในทวีป แปซิฟิก อเมริกาใต้ แอฟริกา และออสเตรเลีย ทำให้เห็นถึงข้อมูลที่ซ่อนอยู่ ซึ่งการมองจากภาพรวมอาจไม่สามารถค้นหาถึงข้อมูลนี้ได้

7.1.2.8 ทดสอบโดยใช้ ทวีปที่อยู่อาศัยของลูกค้า (CUST\_CONTINENT) , ราคาของ Booking (TOTALPRICE) และ เพศ (SEX) เป็น Descriptor Attributes ผลลัพธ์ที่ได้คือ

- 1) IF HOTEL\_CONTINENT = 'PACIFIC' AND SEX = 'F' AND TOTALPRICE = 'LOWPRICE' THEN RESULT IS N (70.27%)
- 2) IF HOTEL\_CONTINENT = 'AMERICA' AND SEX = 'F' AND TOTALPRICE = 'LOWPRICE' THEN RESULT IS N (76.92%)
- 3) IF HOTEL\_CONTINENT = 'MIDDLE EAST' AND SEX = 'M' AND TOTALPRICE = 'LOWPRICE' THEN RESULT IS Y (66.67%)
- 4) IF HOTEL\_CONTINENT = 'AFRICA' AND TOTALPRICE = 'LOWPRICE' AND SEX = 'M' THEN RESULT IS N (66.67%)
- 5) IF HOTEL\_CONTINENT = 'ASIA' AND TOTALPRICE = 'LOWPRICE' AND SEX = 'M' THEN RESULT IS Y (73.53%)
- 6) IF HOTEL\_CONTINENT = 'AFRICA' AND TOTALPRICE = 'HIGHPRICE' AND SEX = 'F' THEN RESULT IS N (75.00%)

แสดงให้เห็นถึงว่าลูกค้าผู้ชายที่อยู่ในทวีปเอเชียและตะวันออกกลางที่สำรองห้องพักที่มีราคาต่ำ มีโอกาสเป็นลูกค้าที่มีศักยภาพ

**7.1.2.9 ทดสอบโดยใช้ ทวีปที่อยู่อาศัยของลูกค้า (CUST\_CONTINENT) , ประเภทของ Booking (BOOKING\_TYPE) และ เพศ (SEX) เป็น Descriptor Attributes ผลลัพธ์ที่ได้ คือ**

- 1) IF CUST\_CONTINENT = 'PACIFIC' AND SEX = 'M' AND BOOKING\_TYPE = 'INSTANT' THEN RESULT IS Y (51.23%)
- 2) IF CUST\_CONTINENT = 'PACIFIC' AND SEX = 'M' AND BOOKING\_TYPE = 'NORMAL' THEN RESULT IS Y (61.14%)
- 3) IF CUST\_CONTINENT = 'PACIFIC' AND SEX = 'F' AND BOOKING\_TYPE = 'INSTANT' THEN RESULT IS Y (53.61%)
- 4) IF CUST\_CONTINENT = 'PACIFIC' AND SEX = 'F' AND BOOKING\_TYPE = 'NORMAL' THEN RESULT IS N (50.42%)
- 5) IF CUST\_CONTINENT = 'AMERICA' AND SEX = 'M' AND BOOKING\_TYPE = 'INSTANT' THEN RESULT IS Y (52.81%)
- 6) IF CUST\_CONTINENT = 'AMERICA' AND SEX = 'M' AND BOOKING\_TYPE = 'NORMAL' THEN RESULT IS N (53.10%)
- 7) IF CUST\_CONTINENT = 'AMERICA' AND SEX = 'F' AND BOOKING\_TYPE = 'INSTANT' THEN RESULT IS N (52.05%)
- 8) IF CUST\_CONTINENT = 'AMERICA' AND SEX = 'F' AND BOOKING\_TYPE = 'NORMAL' THEN RESULT IS Y (52.45%)
- 9) IF CUST\_CONTINENT = 'EUROPE' AND BOOKING\_TYPE = 'INSTANT' AND SEX = 'M' THEN RESULT IS Y (52.28%)
- 10) IF CUST\_CONTINENT = 'EUROPE' AND BOOKING\_TYPE = 'INSTANT' AND SEX = 'F' THEN RESULT IS N (52.31%)
- 11) IF CUST\_CONTINENT = 'EUROPE' AND BOOKING\_TYPE = 'NORMAL' AND SEX = 'M' THEN RESULT IS Y (55.49%)
- 12) IF CUST\_CONTINENT = 'EUROPE' AND BOOKING\_TYPE = 'NORMAL' AND SEX = 'F' THEN RESULT IS Y (55.33%)

แสดงให้เห็นถึงว่าลูกค้าผู้ชายที่อยู่ในทวีปแปซิฟิก เพศชายที่สำรองห้องพักประเภทธรรมดามีโอกาสเป็นลูกค้าที่มีศักยภาพ ในขณะที่ลูกค้าที่อยู่ในทวีปอเมริกาและทวีปยุโรป เพศชายสนใจห้องพักแบบที่มีสินค้า (INSTANT) มากกว่า อย่างไรก็ตามก็ดีจากกฎที่ได้พบว่า เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลูกค้าผู้หญิงที่อยู่ในทวีปอเมริกามีอัตราการยืนยันการซื้อสินค้าประเภทธรรมดามากกว่าประเภทมีสินค้า แสดงให้เห็นว่าลูกค้าผู้หญิงที่อยู่ในทวีปอเมริกาไม่ได้ให้ความสำคัญกับห้องพักแบบที่มีสินค้ามากที่สุด ซึ่งอาจมีปัจจัยอื่นลูกค้ากลุ่มนี้ให้ความสำคัญมากกว่าเช่น ราคา

#### 7.1.2.10 ใช้ ทวีปที่อยู่อาศัยของลูกค้า (CUST\_CONTINENT) , ราคา (TOTALPRICE) และ เพศ (SEX) เป็น Descriptor Attribute ผลลัพธ์ที่ได้คือ

- 1) IF TOTALPRICE = 'LOWPRICE' AND CUST\_CONTINENT = 'AMERICA' AND SEX = 'M' THEN RESULT IS Y (57.82%)
- 2) IF TOTALPRICE = 'LOWPRICE' AND CUST\_CONTINENT = 'AMERICA' AND SEX = 'F' THEN RESULT IS Y (55.56%)
- 3) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'AMERICA' AND SEX = 'M' THEN RESULT IS N (52.35%)
- 4) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'AMERICA' AND SEX = 'F' THEN RESULT IS N (67.21%)
- 5) IF TOTALPRICE = 'LOWPRICE' AND CUST\_CONTINENT = 'EUROPE' AND SEX = 'M' THEN RESULT IS Y (60.60%)
- 6) IF TOTALPRICE = 'LOWPRICE' AND CUST\_CONTINENT = 'EUROPE' AND SEX = 'F' THEN RESULT IS N (51.33%)
- 7) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'EUROPE' AND SEX = 'M' THEN RESULT IS Y (50.98%)
- 8) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'EUROPE' AND SEX = 'F' THEN RESULT IS N (61.45%)
- 9) IF TOTALPRICE = 'LOWPRICE' AND CUST\_CONTINENT = 'PACIFIC' AND SEX = 'M' THEN RESULT IS Y (61.20%)
- 10) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'PACIFIC' AND SEX = 'M' THEN RESULT IS N (56.22%)
- 11) IF TOTALPRICE = 'HIGHPRICE' AND CUST\_CONTINENT = 'PACIFIC' AND SEX = 'F' THEN RESULT IS N (56.32%)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แสดงให้เห็นถึงว่าลูกค้าส่วนใหญ่ให้ความสำคัญเรื่องของเราอย่างมาก ยกเว้นกรณีลูกค้าผู้ชายที่อาศัยอยู่ในทวีปยุโรป ซึ่งมีการยืนยันการสั่งซื้อที่สูงกว่ากลุ่มทวีปอื่น ซึ่งนอกจากนั้นยังแสดงให้เห็นถึงอัตราการยืนยันของลูกค้าในทุกทวีปที่เป็นผู้ชายมากกว่าลูกค้าผู้หญิงอีกด้วย

## 7.2 ส่วนงานการทำนาย (Prediction Phase)

ส่วนงานการทำนายเป็นกระบวนการนำเอาโมเดลที่ได้มาใช้ในการทำนายกับข้อมูลอีกชุดหนึ่ง เพื่อวิเคราะห์หาความเป็นไปได้หรือผลลัพธ์ของเอททริบิวเป้าหมาย

The screenshot shows the 'Precision Reservations' Data Mining System interface. The main content area displays a table of existing models under the 'Decision Tree' category. The table has columns for TREE\_ID, DATE, TIME, and TARGET\_ATTRIBUTE. Each row represents a trained model with a unique ID, training date and time, and a target attribute of 'CHARGECC'. There are also navigation icons for each model, such as 'View', 'Delete', and 'Train Again'.

TREE_ID	DATE	TIME	TARGET_ATTRIBUTE
2004317V4B	Mar 17, 2004	3:33:18 PM	CHARGECC
2004317V45	Mar 17, 2004	3:29:10 PM	CHARGECC
2004317V44	Mar 17, 2004	3:24:02 PM	CHARGECC
2004317V27	Mar 17, 2004	2:17:03 PM	CHARGECC
2004317V26	Mar 17, 2004	2:15:51 PM	CHARGECC
2004317V25	Mar 17, 2004	2:14:32 PM	CHARGECC
2004317V24	Mar 17, 2004	1:58:46 PM	CHARGECC

รูปที่ 7.12 หน้าจอแสดงรายการ Model ที่มีเพื่อเลือกใช้ในการ Predicting

หน้าจอแสดงรายการการทดสอบต่างๆที่ผ่านสามารถเลือกดูรายละเอียด Model ก่อนได้ หลังจากนั้นจึงทำการเลือกเพื่อ Model เพื่อที่จะนำไปประมวลผลเปรียบเทียบกับเพื่อทำนายค่าของข้อมูลต่อไป ซึ่งหากต้องการลบ Model ที่ได้ทำการสร้างไว้แล้ว ก็สามารถเลือกลบได้จากหน้านี้

www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

**Precision Reservations**

Data Mining System  
**Decision Tree**

- ▶ Home
- ▶ Setup
  - ▶ Database Configuration
- ▶ Training
  - ▶ Train New Model
  - ▶ Existing Models
- ▶ Prediction
  - ▶ Predicting Samples

▶ Predicting Samples Process

Please Select Table that you want to Process.

TABLE	TABLE_NAME	DATE_CREATED
<input checked="" type="radio"/>	TESTING	2/11/2004 6:17:38 AM
<input type="radio"/>	TRAININGDATA	2/12/2004 12:02:48 AM

After you choose Table, please  
Click [Validate Attributes and Process Prediction](#)

รูปที่ 7.13 หน้าจอเลือกตารางที่ต้องการจะนำมาทำนายผล

จากนั้นจึงเลือกตารางที่ต้องการใช้ในการทำนายผล

www.PrecisionReservations.com  
Premier Reservations and Customer Service for Internet Hotel Discounts

**Precision Reservations**

Data Mining System  
**Decision Tree**

- ▶ Home
- ▶ Setup
  - ▶ Database Configuration
- ▶ Training
  - ▶ Train New Model
  - ▶ Existing Models
- ▶ Prediction
  - ▶ Predicting Samples

Please select a reference attribute for your convenient in relating to the results.  
Choose Attribute and Process Prediction

TARGET	COLUMN_NAME	IS_NULLABLE	MAX_LENGTH
<input checked="" type="radio"/>	REFNO	False	20
<input type="radio"/>	CHARGECC	True	1
<input type="radio"/>	ROOMOCC	True	
<input type="radio"/>	URGENT	True	1
<input type="radio"/>	PERSON	True	
<input type="radio"/>	SMOKING	True	5
<input type="radio"/>	BEDTYPE	True	15
<input type="radio"/>	FLOOR	True	5
<input type="radio"/>	NO_OF_ROOM	True	
<input type="radio"/>	SEX	False	1
<input type="radio"/>	HOTEL_CONTINENT	True	30
<input type="radio"/>	CUST_CONTINENT	False	11

รูปที่ 7.14 หน้าจอเลือก Reference Attribute

หน้าจอแสดงลิส Attribute เพื่อให้ User ทำการเลือก Attribute ที่ต้องการในการอ้างอิง เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

You have selected REFNO as your reference attribute.

REF_ATT	CUST_CONTINENT	TOTALPRICE	MATCH_RULE	PREDICTION	CONFIDENCE
1027685	Europe	HIGHPRICE	6	Y	46.48
1027856	Pacific	HIGHPRICE	2	Y	44.49
1015023	Pacific	LOWPRICE	1	Y	60.19
1038523	Asia	LOWPRICE	11	Y	45.54
1108564	Europe	HIGHPRICE	6	Y	46.48
1032965	America	LOWPRICE	3	Y	56.94
1049044	Europe	HIGHPRICE	6	Y	46.48
1052138	Europe	HIGHPRICE	6	Y	46.48
1107267	Europe	LOWPRICE	5	Y	58.94
1074251	Europe	HIGHPRICE	6	Y	46.48
1081343	Europe	LOWPRICE	5	Y	58.94
1039309	Europe	LOWPRICE	5	Y	58.94

รูปที่ 7.15 หน้าจอแสดงผลลัพธ์จากการประมวลผลทำนายตาม Model

หลังจากเลือก Model และ Attributes ระบบก็จะทำการประมวลผลเปรียบเทียบค่าของกลุ่มตัวอย่างกับ Model ที่เลือกใช้ เพื่อค้นหาว่าแต่ละชุดของกลุ่มตัวอย่าง มีความน่าจะเป็นตาม Rules ของ Model อย่างไร แล้วจึงแสดงผลออกมาเป็นคำตอบว่ามีความน่าจะเป็นของ Target Attribute อย่างไร และคิดเป็นสัดส่วนเท่าไรอย่างไร

## บทที่ 8

### บทสรุป

โครงการพัฒนาโปรแกรมประยุกต์สำหรับธุรกิจสำรองห้องพักรวมโดยใช้ Decision Tree นี้เป็น โครงการที่จัดทำขึ้นมาเพื่อนำเสนอให้เห็นถึงประโยชน์ของการนำเอาทฤษฎีค้ำไม้หนึ่ง มาใช้เพิ่มประสิทธิภาพในการค้นหาลูกค้ำที่มีศักยภาพ ผลลัพธ์ที่ได้ทำให้สามารถมองเห็นภาพรวมของลูกค้ำที่มีศักยภาพ อีกทั้งการนำเอาโมเดลที่ได้มาทำการทดสอบกับลูกค้ำใหม่ที่ยังไม่สิ้นสุดการดำเนินงานทางธุรกิจเพื่อคาดคะเนหาคำตอบว่าลูกค้ำดังกล่าวมีความเป็นไปได้ที่จะเป็นลูกค้ำที่มีศักยภาพมากน้อยเพียงใด ทำให้ธุรกิจสามารถมุ่งเน้น ไปให้บริการกับลูกค้ำกลุ่มที่มีความเป็นไปได้ที่จะเป็นลูกค้ำที่มีศักยภาพได้

#### 8.1 สรุปหลักการที่ใช้ในระบบ

เทคนิคที่ใช้ในการค้นหาลูกค้ำที่มีศักยภาพ ได้นำเอาเทคนิคของการ Classification ในส่วนของ Decision Tree มาใช้ในการพัฒนาระบบ ซึ่งเป็นเทคนิคของการสร้างรูปแบบการทำนายโดยอาศัยประสบการณ์การเรียนรู้หรือความสนใจในสิ่งบางอย่างที่จะเป็นหลักสำคัญในการตัดสินใจ โดยอัลกอริทึมที่เลือกใช้คือ อัลกอริทึม ID3 ซึ่งเป็นกระบวนการการเรียนรู้แบบมีรูปแบบของอินพุตและเอาต์พุต ไว้ก่อนแล้วจากข้อมูลสำรองห้องพักรวมที่มีอยู่เดิม จากนั้นจึงนำเอาข้อมูลที่ได้มาเลือกหา Attributes ที่มีความสำคัญเพียงพอที่คาดว่าจะมีความสัมพันธ์ต่อ Attribute เป้าหมาย มาทำการเทรนนิ่ง โดยจะนำเอา Attributes ต่างๆที่เลือกมา มาดำเนินการค้นหาว่า Attribute ไหนมีความสำคัญมากที่สุด โดยการใช้วิธีการของ Information Gain เพื่อเลือก Attribute ที่มีค่ามากที่สุดเป็น Attribute เริ่มต้นสำหรับการทดสอบเพื่อสร้างโมเดล Decision Tree หลังจากได้โมเดลแล้ว จึงนำเอาโมเดลที่ได้ไปทำการทดสอบกับข้อมูลอื่นๆที่ไม่ใช่ข้อมูลที่ใช้ในการสร้าง โมเดล เพื่อทำการทดสอบว่าโมเดลที่ได้มีความถูกต้องหรือไม่ด้วย

#### 8.2 สรุปกระบวนการในการทำงาน

โครงการนี้เริ่มต้นจากการค้นหา Attribute ที่มีความสำคัญเพียงพอที่จะระบุได้ว่าลูกค้ำคนใดมีแนวโน้มที่จะเป็นลูกค้ำที่มีศักยภาพ โดยจากการวิเคราะห์ Attributes ที่เกี่ยวข้อง ทำให้การคัดเลือก Attributes ที่คาดว่าจะจะเป็นปัจจัยที่มีความสำคัญขึ้นมา จากนั้นจึงนำข้อมูลมาทำความสะอาด พร้อมทั้งแปลงข้อมูลให้อยู่ในรูปแบบที่แบ่งกลุ่มไว้ แล้วจึงนำเอาข้อมูลเหล่านั้นไปทำการ

เทรนนิ่ง เพื่อให้ได้ผลลัพธ์ออกมาเป็น โมเดล โดยจะเก็บผลลัพธ์เป็นไฟล์ใน รูปแบบ XML เพื่อความสะดวกในการแปลงกลับเป็นรูปแบบต่างๆ โดยส่วนงานแสดงผลจะแบ่งเป็นสองส่วนคือส่วนที่แสดงผลเป็นรูปแบบ Tree View และ ส่วนแสดงผลในรูปแบบกฎ IF-THEN นอกจากนี้ส่วนงานการทำนายสามารถนำเอาข้อมูลลูกค้าใหม่ที่ยังไม่สิ้นสุดการดำเนินงานทางธุรกิจเพื่อหาคำตอบว่าลูกค้าดังกล่าวมีความเป็นไปได้ที่จะเป็นลูกค้าที่มีศักยภาพหรือไม่

### 8.3 สรุปผลการทดสอบ

เมื่อนำเอาผลที่ได้จากการเทรนนิ่งไปทำการทดสอบกับข้อมูลการสำรองห้องพักโรงแรมจำนวน 10,000 ชุดข้อมูล พบว่ามีปัจจัยต่างๆมีความสัมพันธ์เกี่ยวข้องกับความสำเร็จในการให้บริการการสำรองห้องพัก ไม่ว่าจะเป็นราคา เพศ ราคา ตลอดจนถึง ประเภทของ booking ซึ่งจากการทดสอบหลายๆเงื่อนไข ทำให้พบความสัมพันธ์หลากหลายรูปแบบเช่น ลูกค้าที่เป็นหญิงให้ความสำคัญกับเรื่องของราคามากกว่าลูกค้าที่เป็นผู้ชาย ลูกค้าที่เป็นผู้ชายให้ความสำคัญต่อประเภทของการ Booking แบบมีสำรองสินค้ามากกว่าผู้หญิง ตลอดจนการค้นพบว่าลูกบางกลุ่มที่เดิมเข้าใจว่าไม่ใช่ลูกค้าที่มีศักยภาพ แต่แท้จริงแล้วมีความเป็นไปได้ที่จะประสบความสำเร็จในการให้บริการการสำรองห้องพัก โดยธุรกิจสามารถนำความรู้ที่ได้มาประยุกต์ใช้ในการเสนอสินค้าต่อลูกค้าให้ตรงเป้าหมายมากยิ่งขึ้น ตลอดจนธุรกิจจะได้ทุ่มทรัพยากรและการบริการที่มีอยู่อย่างจำกัด ไปยังลูกค้ากลุ่มที่มีศักยภาพ อันจะนำไปซึ่งการเพิ่มขึ้นของประสิทธิภาพและประสิทธิผลของการดำเนินงานธุรกิจต่อมา

### 8.4 ข้อเสนอแนะ

เนื่องจากแต่เดิมในการสำรองห้องพักเน้นความสะดวกรวดเร็ว เพื่อให้ลูกค้าใช้เวลาน้อยที่สุดในการสำรองห้องพัก จึงมีการเก็บข้อมูลลูกค้าเท่าที่จำเป็น ทำให้ข้อมูลลูกค้าที่มีอยู่ไม่เพียงพอต่อการวิเคราะห์อย่างลึกซึ้ง ซึ่งหากมีการเก็บข้อมูล ประวัติลูกค้าที่ละเอียดมากยิ่งขึ้น ไม่ว่าจะเป็นการเก็บข้อมูลในเรื่องอายุ อาชีพ ลักษณะงานอดิเรก กีฬา สถานที่ท่องเที่ยวที่ชอบเป็นต้น จะทำให้สามารถค้นหาลักษณะของลูกค้าที่มีศักยภาพได้อย่างมีประสิทธิภาพมากขึ้นอย่างแน่นอน

## บรรณานุกรม

สุรพล ศรีบุญทรง. 2541. “Data Mining เครื่องมือการตลาดยุคดิจิทัล.” **IT SOFT** มกราคม 2541 : 116-123

Kurt Thearling. **An Overview of Data Mining Techniques**. [Online]. Available:

<http://www.thearling.com/text/dmtechniques/dmtechniques.htm>

Statsoft. **Classification Trees**. [Online]. Available:

<http://www.statsoftinc.com/textbook/stclatre.html>

Insightful Corporation, 2002. **Decision Trees for Predictive Modeling**. [Online]. Available:

<http://www.datamining.com>

Helge Grenager Solheim. 1996. **ID3** [Online]. Available : [http://www.recursive-partitioning.com/Classification\\_Trees/more36.html](http://www.recursive-partitioning.com/Classification_Trees/more36.html).

Edward S. Blurock. **The ID3 Algorithm**. [Online]. Available: <http://www.risc.unilinz.ac.at/people/blurock/ANALYSIS/manual/document/node26.html>

Information Discovery. **A Characterization of Data Mining Technologies and Process 2002 December 15** [Online]. Available: <http://www.datamining.com>

Peter Cabena. 1998. **Discovering Data Mining: From Concept to Implementation**. New Jersey. Prentice Hall PTR.

Salford System. **An Overview of Decision Tree**. [Online]. Available : <http://www.salford-systems.com/whitepaper.html>

## ประวัติผู้เขียน

ชื่อผู้เขียน	นายกุลชาติ ตั้งกาญจนภาสน์
วันเดือนปีเกิด	3 สิงหาคม พ.ศ. 2520
สถานที่เกิด	เชียงใหม่
อีเมล	kunlachart@yahoo.com

### ประวัติการศึกษา

- สำเร็จการศึกษาระดับประถม โรงเรียนมงฟอร์ตวิทยาลัย เชียงใหม่
- เข้ารับการศึกษาระดับมัธยม โรงเรียนมงฟอร์ตวิทยาลัย เชียงใหม่
- สำเร็จการศึกษามัธยมปลาย จากศูนย์การศึกษานอกโรงเรียน เชียงใหม่
- สำเร็จการศึกษาระดับปริญญาตรี คณะเศรษฐศาสตร์ สาขาเศรษฐศาสตร์ มหาวิทยาลัยเชียงใหม่
- กำลังศึกษาระดับปริญญาโท คณะเทคโนโลยีสารสนเทศ สาขาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีเจ้าคุณทหารลาดกระบัง

### ประวัติการทำงาน

- บริษัทนัมซี่เส็งลิซซิ่งจำกัด (ปี พ.ศ. 2541)  
ตำแหน่ง เจ้าหน้าที่สินเชื่อและเร่งรัดหนี้สิน
- ITAG Consulting co., Ltd. (ปี พ.ศ. 2542 - 2543)  
ตำแหน่ง Web Designer
- บริษัทสามารถคอร์ปอเรชั่นจำกัด (มหาชน) (ปี พ.ศ. 2543 - 2546)  
ตำแหน่ง Web Programmer
- Anacott Asia-Pacific co., Ltd. (ปี พ.ศ. 2546 – ปัจจุบัน)  
ตำแหน่ง Web Team Leader

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้