ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

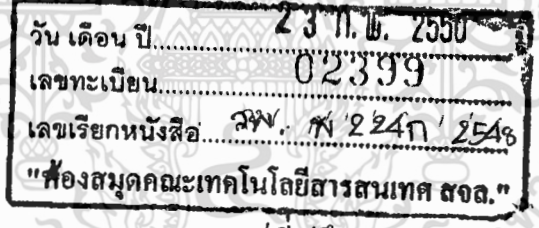
การพัฒนาระบบดาต้าไมน์นิ่งโดยใช้ Decision Tree

Development of Data Mining system using Decision Tree

โดย

พยูน พาณิชย์กุล

รหัสประจำตัว 47066120



อาจารย์ที่ปรึกษา

รศ.ดร.วรพจน์ กรีสุระเดช

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคฤดูร้อน ปีการศึกษา 2548

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การพัฒนาระบบดาต้าไมน์นิ่งโดยใช้ Decision Tree
นักศึกษา	นายพยูน พาณิชย์กุล
อาจารย์ที่ปรึกษา	รศ.ดร.วรพจน์ กรีสุระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2548

บทคัดย่อ

ในปัจจุบันระบบธุรกิจต่างๆ มีการแข่งขันกันสูงมากโดยแต่ละองค์กรต่างก็มีการแย่งชิงความได้เปรียบทางธุรกิจ ข่าวสาร ข้อมูลต่างๆ โดยใครที่มีความสามารถนำเอาข้อมูลเหล่านี้มาใช้งานได้มากที่สุดก็จะถือว่าได้เปรียบคู่แข่งเป็นอย่างมาก ซึ่งในแต่ละองค์กรได้มีการนำเอาข้อมูลที่มีอยู่มาวิเคราะห์เพื่อช่วยให้การตัดสินใจในการดำเนินธุรกิจเพื่อช่วยในการตัดสินใจ ซึ่งมีรูปแบบหนึ่งของกระบวนการทางด้าน Data Mining ที่ช่วยในการตัดสินใจ คือ Decision Tree ซึ่งเป็นรูปแบบที่ง่ายต่อการเข้าใจ เพราะเป็นรูปแบบที่คนส่วนใหญ่เคยสัมผัสมาบ้างแล้ว จึงเป็นรูปแบบที่หน้าสนใจสำหรับใช้ในการเคราะห์ข้อมูลที่มีอยู่อย่างมีประสิทธิภาพและง่ายต่อการเข้าใจ.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Development of Data Mining system using Decision Tree
Student	Mr. Payoon Panichkul
Advisor	Assoc. Prof. Dr. Worapoj Kreesuradej
Level of Study	Master of Science in Information
Major	Information Science
Academic	2005

ABSTRACT

Present, There is high compete in business system. Every company have to fight for business advantage. Data and information can make more advantage and can be the intelligence business. Each company use all information they have to make decision with their business and there are 1 of Data Mining process that can helpful to easy decision is "Decision Tree". Decision Tree has easy to understand because most of business ever seen before. Then, it is interesting to use for more efficiency data analyze and easy to understand anyway.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ในการพัฒนาและศึกษาโครงงานการพัฒนาระบบดาต้าไมน์นิ่งโดยใช้ Decision Tree ได้รับความสนับสนุนเป็นอย่างดีจากหลายฝ่าย ไม่ว่าจะไปในด้านคำแนะนำ คำปรึกษา กำลังใจ ตลอดทั้งค่าใช้จ่ายในการดำเนินงาน เพื่อให้ได้ระบบที่มีประสิทธิภาพและตรงตามวัตถุประสงค์ที่วางไว้ ผู้จัดทำจึงใคร่ขอขอบพระคุณบุคคลดังต่อไปนี้

1. ผศ.ดร. วรพจน์ กรีสุระเดช อาจารย์ที่ปรึกษา ผู้ให้ความรู้ คำแนะนำ ข้อเสนอแนะ ตลอดทั้งแนะนำหนังสือเทคนิควิธีการแก้ไขปัญหาต่างๆ
2. บิดาและมารดา ที่คอยเป็นกำลังใจและกำลังทรัพย์เป็นอย่างดี
3. นาย ฐานันด์ บวรศักดิ์สกุล ผู้ให้คำปรึกษาแก้ข้อสงสัยเกี่ยวกับ Microsoft Visual Studio .NET 2003
4. เพื่อนๆ ทุกคนที่คอยให้คำปรึกษาและคอยเป็นกำลังใจ

นาย พยูน พาณิชย์กุล

11 เมษายน 2549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... I

บทคัดย่อภาษาอังกฤษ..... II

กิตติกรรมประกาศ..... III

สารบัญ..... IV

สารบัญตาราง..... VI

สารบัญภาพ..... VII

บทที่

1. บทนำ

1.1 ความเป็นมาของปัญหา..... 1

1.2 วัตถุประสงค์ของโครงการ..... 1

1.3 ขอบเขตการศึกษา..... 2

1.4 องค์ประกอบของการพัฒนาโครงการ..... 2

1.5 ขั้นตอนและวิธีการดำเนินโครงการ..... 2

1.6 ประโยชน์ที่คาดว่าจะได้รับ..... 3

2. ดาต้าไมน์นิ่ง

2.1 ดาต้าไมน์นิ่ง (Data Mining)..... 4

2.2 ความเป็นมาของดาต้าไมน์นิ่ง..... 4

2.3 กระบวนการของดาต้าไมน์นิ่ง..... 5

2.4 รูปแบบของดาต้าไมน์นิ่ง..... 6

2.5 การแบ่งแยกประเภท..... 9

2.6 ดาต้าไมน์นิ่งช่วยระบบธุรกิจได้อย่างไร..... 10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3. ดิซิชั่นทรี

3.1 ความหมายของดิซิชั่นทรี..... 11

3.2 กระบวนการทำงานของดิซิชั่นทรี..... 11

3.3 ขั้นตอนการสร้างดิซิชั่นทรี..... 12

3.4 อัลกอริทึม ID 3..... 12

4. วิเคราะห์และออกแบบโปรแกรม

4.1 วิเคราะห์และออกแบบโปรแกรม..... 15

4.2 Process Model..... 16

4.2.1 Use Case Diagram..... 16

4.2.2 Sequence Diagram..... 17

4.2.3 Activity Diagram..... 23

4.3 Structure Chart และ Flow – Chart แสดงการทำงาน..... 25

4.4 Data dictionary..... 30

5. การประยุกต์ใช้โปรแกรม

5.1 การสร้างโปรเจ็กต์ใหม่..... 32

5.2 การทดสอบโมเดล..... 51

6. สรุปผลการศึกษาและข้อเสนอแนะ

6.1 สรุปผลการดำเนินงาน..... 60

6.2 ข้อเสนอแนะ..... 60

บรรณานุกรม..... 61

ประวัติผู้เขียน..... 62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่

3.1 แสดงข้อมูลตัวอย่างในการใช้อัลกอริทึม ID3..... 13

4.1 อธิบายรายละเอียดตาราง number_datatree..... 30

4.2 อธิบายรายละเอียดตาราง main_datatree 30

4.3 อธิบายรายละเอียดตาราง exten_datatree 30

4.4 อธิบายรายละเอียดตาราง refer_datatree 31

4.5 อธิบายรายละเอียดตาราง field_datatree 31

4.6 อธิบายรายละเอียดตาราง class_datatree 31

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

หน้า

รูปที่

2.1 การบวนการดาต้าไมน์นิ่ง..... 5

2.2 แสดงลักษณะของโครงสร้างต้นไม้..... 7

2.3 แสดงลักษณะของ Rule Induction..... 7

2.4 ลักษณะของ K-means Clustering 8

2.5 กระบวนการ Classification 9

3.1 การทำงานของดิซิชั่นทรี..... 11

4.1 แสดงโครงสร้าง Array ในการจัดเก็บข้อมูล..... 15

4.2 แสดงUse Case ของระบบ Decision Tree 16

4.3 แสดง Sequence Diagram ของการติดต่อฐานข้อมูล..... 17

4.4 แสดง Sequence Diagram ของการเลือกข้อมูลสำหรับทำไมน์นิ่ง..... 18

4.5 แสดง Sequence Diagram ของการแก้ไขข้อมูล..... 19

4.6 แสดง Sequence Diagram ของการแปลงข้อมูล..... 20

4.7 แสดง Sequence Diagram ของการทดสอบโมเดล..... 22

4.8 แสดง Activity Diagram ของการสร้างโปรเจ็กต์..... 23

4.9 แสดง Activity Diagram ของการทดสอบโมเดล..... 24

4.10 Structure Chart แสดงระบบ Decision Tree 25

4.11 Structure Chart แสดงการสร้างโปรเจ็กต์..... 25

4.12 Structure Chart แสดงการทดสอบโมเดล..... 26

4.13 Structure Chart แสดงการติดต่อฐานข้อมูล..... 26

4.14 Structure Chart แสดงการแก้ไขข้อมูล..... 27

4.15 Structure Chart แสดงการแปลงข้อมูล..... 27

4.16 Structure Chart แสดงการติดต่อฐานข้อมูล..... 28

4.17 Flow – Chart แสดงการทำงานของการสร้าง Decision Tree..... 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ (ต่อ)

หน้า

รูปที่

5.1 หน้าจอเริ่มต้นของระบบ..... 32
5.2 หน้าจอติดต่อฐานข้อมูล Microsoft SQL Server เพื่อทำการสร้างโปรเจ็กต์..... 33
5.3 หน้าจอเลือกตารางข้อมูลที่ต้องการ..... 34
5.4 หน้าจอแสดงชื่อดาต้าเบสและตารางข้อมูลที่ได้ทำการเลือก..... 35
5.5 หน้าจอเลือกตารางข้อมูลสำหรับการทำไมน์นิ่ง..... 35
5.6 หน้าจอเลือกฟิลด์ข้อมูลสำหรับการทำไมน์นิ่ง..... 36
5.7 หน้าจอแสดงชื่อฟิลด์ ชนิด และขนาดของฟิลด์..... 37
5.8 หน้าจอการตรวจสอบข้อมูลในกรณีเป็นตัวอักษร..... 38
5.9 หน้าจอแสดงการแก้ไขข้อมูลที่มีค่าว่างในกรณีที่เป็นตัวอักษร..... 39
5.10 หน้าจอการตรวจสอบข้อมูลในกรณีเป็นตัวเลข..... 40
5.11 หน้าจอแสดงการแก้ไขข้อมูลที่มีค่าว่างในกรณีที่เป็นตัวเลข..... 41
5.12 หน้าจอแสดงการแก้ไขข้อมูลในกรณีกดปุ่ม “อัตโนมัติ”..... 42
5.13 หน้าจอแสดงข้อมูลทั้งหมดหลังจากการแก้ไข..... 43
5.14 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลขโดยจัดเป็น 2 กลุ่ม..... 44
5.15 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลขโดยจัดเป็น 3 กลุ่ม..... 45
5.16 หน้าจอแสดงข้อมูลหลังจากมีการแปลงข้อมูล..... 46
5.17 หน้าจอแสดงการสร้างโครงสร้างต้นไม้..... 47
5.18 หน้าจอสำหรับทำการเลือกจำนวนเรคคอร์ดในการแตกโครงสร้างต้นไม้..... 48
5.19 หน้าจอแสดงการเพิ่มขนาดหน้าจอการแสดงโครงสร้างต้นไม้..... 49
5.20 หน้าจอแสดงการบันทึกโปรเจ็กต์..... 50
5.21 หน้าจอแสดงการเลือกโมเดลมาทดสอบ..... 51
5.22 หน้าจอแสดงการติดต่อฐานข้อมูล Microsoft SQL Server สำหรับทดสอบ โมเดล..... 52
5.23 หน้าจอแสดงกระบวนการแม็พข้อมูล..... 53
5.24 หน้าจอแสดงการลบค่าข้อมูลที่ค่าว่าง..... 54
5.25 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลขโดยแบ่งเป็น 2 กลุ่ม..... 55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ (ต่อ)

หน้า

รูปที่

5.26 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลขโดยแบ่งเป็น 3 กลุ่ม..... 56
5.27 หน้าจอแสดงการตรวจสอบข้อมูลในกรณีใช้งานได้.....57
5.28 หน้าจอแสดงการตรวจสอบข้อมูลในกรณีใช้งานไม่ได้..... 58
5.29 หน้าจอแสดงผลลัพธ์การทดสอบโมเดล..... 59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

ในปัจจุบันองค์กรต่างๆ ได้มีการนำเอาระบบสารสนเทศเข้ามาใช้งานเพื่อเพิ่มประสิทธิภาพในการดำเนินงานขององค์กรเพิ่มมากขึ้น จึงทำให้มีข้อมูลจำนวนมากเกิดขึ้นภายในระบบ โดยข้อมูลที่เกิดขึ้นนั้นได้ถูกจัดเก็บลงฐานข้อมูลที่มีความแตกต่างกันไปแล้วแต่ความต้องการของแต่ละองค์กรซึ่ง Microsoft SQL Server ก็เป็น Relational Database Management Systems (RDBMS) ตัวหนึ่งที่ได้รับความนิยมเป็นอย่างมากในการจัดเก็บและดูแลรักษาข้อมูลขององค์กรต่างๆ ในปัจจุบัน ซึ่งข้อมูลต่างๆ ที่เกิดขึ้นในระบบนั้นต่างก็เป็นข้อมูลที่มีประโยชน์อย่างมากหากแต่ละองค์กรรู้จักวิธีการที่จะจัดการกับข้อมูลที่ถูกต้อง แต่ในความเป็นจริงแล้วในหลายองค์กรยังไม่สามารถที่จะนำข้อมูลที่มีประโยชน์เหล่านี้มาใช้งานได้อย่างคุ้มค่าเท่าที่ควร ดังนั้นจึงได้มีแนวคิดเกี่ยวกับการนำกระบวนการทางด้าน Data Mining ซึ่งเป็นกระบวนการในการเพิ่มประสิทธิภาพของระบบข้อมูลข่าวสารที่มีอยู่ให้มีประโยชน์สูงสุด มาใช้กับข้อมูลต่างๆ ที่เกิดของภายในองค์กร โดยได้นำเอาเทคนิค การจัดกลุ่มของข้อมูล (Classification) แบบ โครงสร้างต้นไม้ (Decision Tree) มาประยุกต์ใช้ในการวิเคราะห์ข้อมูลที่เกิดขึ้นของแต่ละองค์กร โดยจะแสดงผลลัพธ์ที่ได้ออกมาในรูปแบบของโครงสร้างต้นไม้ที่มีความสะดวกและง่ายต่อการเข้าใจมากกว่าการนำเสนอข้อมูลหรือผลลัพธ์ในรูปแบบเก่าๆ ที่ออกมาเป็นในรูปแบบของเอกสารทั่วๆ ไปไม่ได้ผ่านกระบวนการจัดการข้อมูลที่ดี ดังนั้นในแต่ละองค์กรจึงจำเป็นที่จะต้องจัดหา Tools ที่มีความสามารถเกี่ยวกับการจัดการบริหารข้อมูลภายในองค์กรให้มีประสิทธิภาพและเกิดประโยชน์สูงสุด

1.2 วัตถุประสงค์ของโครงการ

1) เพื่อศึกษากระบวนการทางด้าน Data Mining
2) เพื่อเรียนรู้วิธีในการนำหลักการทางด้าน Data Mining โดยการใช้เทคนิคการ Classification แบบ Decision Tree มาใช้ประโยชน์
3) เพื่อเพิ่มประสิทธิภาพของข้อมูลที่มีอยู่ให้เกิดประโยชน์สูงสุด
4) เพื่อสร้าง Software Tools ที่สามารถจัดการข้อมูลที่มีอยู่ให้เกิดประโยชน์สูงสุด
5) เพื่อให้ได้ข้อมูลที่สามารถนำมาช่วยในการตัดสินใจได้อย่างรวดเร็ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตการศึกษา

โครงการนี้เป็นการศึกษาและพัฒนาระบบงานเกี่ยวกับทางด้าน Data Mining โดยจะนำหลักการและเทคนิคในการจัดกลุ่มของข้อมูล โดยใช้รูปแบบของ โครงสร้างต้นไม้ มาใช้ในการบรรยายผลลัพธ์ของข้อมูลที่เกิดขึ้นหลังจากที่ข้อมูลนั้นได้ผ่านกระบวนการทางด้าน Data Mining แล้ว โดยข้อมูลจะถูกจัดเก็บอยู่ใน Relational Database Management คือ Microsoft SQL Server ซึ่งจะศึกษาเกี่ยวกับวิธีใช้หลักการทำงานต่างๆ ของ Microsoft SQL Server ตลอดจนภาษาที่ใช้ในการสื่อสารกับฐานข้อมูลซึ่งก็คือ Structure Query Language หรือเรียกง่ายๆว่าภาษา SQL Statement

โดยขอบเขตของระบบจะเป็นการดำเนินการตามกระบวนการทางด้าน Data Mining โดยจะเริ่มจากการกำจัดข้อมูลที่ไม่มีส่วนเกี่ยวข้องออกไป ต่อมาก็ทำการเลือกข้อมูลให้ตรงกับความต้องการ ผ่านกระบวนทางด้าน Data Mining ให้ได้รูปแบบของข้อมูลที่เราต้องการ

1.4 องค์ประกอบของการพัฒนาโครงการ

1) คอมพิวเตอร์ Fujitsu รุ่น S-6230
2) ระบบปฏิบัติการ Microsoft Windows XP Professional corporate Edition Version 2002 Service Pack 2
3) Microsoft SQL Server
4) Microsoft Visual Studio .NET 2003
5) Microsoft Office system 2003 Professional Edition

1.5 ขั้นตอนและวิธีการดำเนินโครงการ

1) ศึกษากระบวนการทางด้าน Data Mining
2) เลือก Data Mining Model ที่เหมาะสมกับโครงการ
3) ศึกษา Data Mining Model ที่ต้องการคือ Decision Tree และอัลกอริทึมต่างๆ
4) เลือกอัลกอริทึมในการสร้าง Decision Tree และทำการศึกษา
5) ศึกษาการทำงานของ Microsoft SQL Server
6) ศึกษา Structure Query Language
7) ศึกษาการใช้งาน Microsoft Visual Studio .NET 2003
8) ศึกษาการติดต่อฐานข้อมูลโดยวิธีการต่างๆของ Microsoft Visual Studio .NET 2003
9) เลือกวิธีการติดต่อฐานข้อมูลที่เหมาะสม
10) ออกแบบและวิเคราะห์ระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สร้างฐานข้อมูลสำหรับทำการทดสอบระบบ

11) ทำการพัฒนาระบบ

12) ทดสอบและปรับปรุงแก้ไขระบบ

13) จัดทำเอกสารประกอบ

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1) เพิ่มความรวดเร็วในการทำงานเกี่ยวกับการจัดการกับข้อมูลที่มีอยู่

2) เพิ่มประสิทธิภาพการใช้ข้อมูลให้เกิดประโยชน์สูงสุด

3) นำความรู้และเทคนิคที่ได้ศึกษาเกี่ยวกับกระบวนการต่างๆ ไปประยุกต์ใช้ในการวิเคราะห์ข้อมูลเพื่อเพิ่มประสิทธิภาพในการดำเนินงานขององค์กร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ดาต้าไมน์นิ่ง

2.1 ดาต้าไมน์นิ่ง (Data Mining)

ดาต้าไมน์นิ่ง (Data Mining) เป็นกระบวนการเพิ่มประสิทธิภาพของระบบข้อมูลข่าวสาร โดยการดึงเอาข้อมูลที่มีประโยชน์ที่ซ่อนเร้นอยู่ภายในข้อมูลดิบออกมาใช้งานเพื่อเพิ่มประสิทธิภาพของข้อมูลและใช้ประโยชน์จากข้อมูลให้มากที่สุด ไม่ว่าข้อมูลนั้นจะถูกจัดเก็บในรูปแบบใด หรือขนาดของข้อมูลจะมีขนาดเท่าไหร่ก็ตาม เช่นการดึงข้อมูลจากฐานข้อมูล การดึงข้อมูลจากไฟล์ข้อมูล โดยมีระบบ วิศวกรรมซอฟแวร์ซึ่งเป็นระบบที่สามารถจะทำการดึงข้อมูลออกมาแล้วทำการหาผลลัพธ์โดยรวมของข้อมูลทั้งหมด โดยสามารถที่จะทำเป็นตัวรายงานออกเพื่อให้มีความน่าสนใจและเข้าใจ
มากขึ้น และยังได้รูปแบบ (Pattern) ใหม่ๆออกมาจากข้อมูลซึ่งเป็นสิ่งที่เราไม่เคยรู้มาก่อน ซึ่งผลลัพธ์ที่ได้ออกมานั้น เป็นผลลัพธ์ที่ปราศจากการแทรกแซงของความคิดมนุษย์ ซึ่งดาต้าไมน์นิ่งสามารถเรียกอีกอย่างหนึ่งได้ว่า Knowledge Discovery from very large Database (KDD)

โดยดาต้าไมน์นิ่งมีลักษณะที่สำคัญอยู่ 2 อย่างด้วยกันคือ

1) Predictive Methods เป็นวิธีที่ใช้ในกาเตรียมข้อมูลเพื่อใช้ในการทำนายพฤติกรรมต่างๆ ในอนาคตที่อาจจะเกิดขึ้น

2) Descriptive Methods เป็นวิธีที่ใช้สำหรับบรรยายหรืออธิบายรูปแบบต่างของข้อมูล จัดกลุ่มข้อมูล และสร้างความสัมพันธ์

2.2 ความเป็นมาของดาต้าไมน์นิ่ง

ดาต้าไมน์นิ่ง (Data Mining) พัฒนามาจากเทคนิคทางสถิติ ฐานข้อมูล และการเรียนรู้ของเครื่องจักรกล (Machine – Learning) โดยจะเน้นที่ข้อมูลที่มีขนาดใหญ่ เพื่อสร้างตัวแบบสำหรับทำนายพฤติกรรมของลูกค้า ซึ่งในปัจจุบันเทคโนโลยีทำให้กระบวนการดาต้าไมน์นิ่งเป็นไปอย่างอัตโนมัติ มีการรวมเข้ากับดาต้าแวร์เฮาส์ และนำเสนอผลลัพธ์ในหลายๆ ทางที่ผู้ใช้ต้องการได้อย่างสะดวกมากขึ้น โดยทั้งสามวิธีนั้นจะมีแนวทางที่ดาต้าไมน์นิ่งนำหลักการมาใช้แตกต่างกันไป คือ

1) ฐานข้อมูล (Database Technology) การทำงานที่ดาต้าไมน์นิ่งนำมาจากฐานข้อมูลคือการเก็บข้อมูล การคำนวณเกี่ยวกับการทำงานที่ทำซ้ำๆ และให้ข้อมูลเฉพาะบางส่วนที่เก็บไว้เท่านั้น

เอกสารนี้ [illegible] ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) สถิติ (Statistics) คือการรวบรวมข้อมูลที่มีอยู่แล้วนำทฤษฎีทางสถิติมาวิเคราะห์เพื่อที่จะบอกถึงค่าความเป็นไปได้ต่างๆ ที่อาจจะเกิดขึ้น

3) การเรียนรู้ของเครื่องจักรกล (Machine Learning) โดยในการทำนายค่าต่างๆ ของดาต้าไมน์นิ่งได้นำความรู้จักวิทยาศาสตร์ทางคอมพิวเตอร์ที่ได้คิดค้นแนวทางที่ทำให้เครื่องจักรเรียนรู้ข้อมูล เช่น Decision Tree

โดยในปัจจุบันกระบวนการทางดาต้าไมน์นิ่งนั้นจะนำมาใช้สำหรับแก้ปัญหาเกี่ยวการดำเนินธุรกิจที่เกิดขึ้น ซึ่งปัญหาทางธุรกิจที่ได้นำเอาดาต้าไมน์นิ่งมาประยุกต์ใช้ เช่น

1) การเพิ่มขึ้นของคู่แข่งและความเสี่ยงทางธุรกิจ มีความขึ้นโดยเมื่อพิจารณาแล้วพบว่าแนวโน้มของสินค้าจะมีการซื้อขายกันทั่วโลก โดยผ่านทางอินเตอร์เน็ตทำให้ยากที่จะเก็บข้อมูลต่างอีกทั้งแนวโน้มการเปลี่ยนแปลงของลูกค้าก็เป็นไปอย่างรวดเร็วทำให้การดำเนินธุรกิจเกิดความเสี่ยง

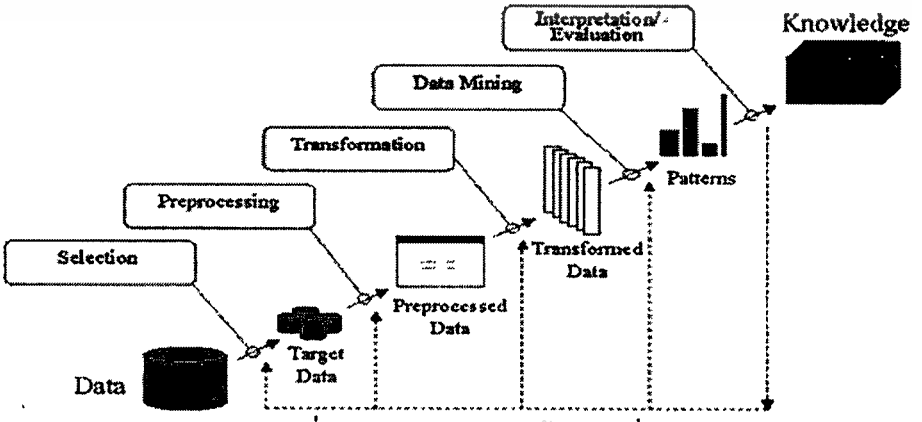
2) การเพิ่มขึ้นของสินค้า โดยสินค้าและบริการหลายชนิดมีการเพิ่มที่แตกต่างจากสายของสินค้าชนิดนั้นๆ เพราะว่ามีตลาดกลุ่มใหม่ๆ เกิดขึ้นตลอดเวลา เนื่องจากตลาดของผู้บริโภคมีความชอบที่แตกต่างกันไป

3) เวลา ถูกให้ความสำคัญมากขึ้นเพราะว่า ขู่แข่งขันในตลาดมีมากขึ้นอย่างรวดเร็ว

4) วงจรชีวิตของผลิตภัณฑ์สั้นลง เนื่องจากผลิตภัณฑ์ต่างๆ สามารถหาซื้อจากตลาดได้อย่างรวดเร็วและพฤติกรรมของผู้บริโภคมีการเปลี่ยนแปลงอย่างรวดเร็วมาก ทำให้อายุของสินค้ามีช่วงชีวิตที่สั้นลง

5) รูปแบบพฤติกรรมของผู้บริโภค พฤติกรรมของผู้บริโภคมีการเปลี่ยนแปลงไป ซึ่งจะเกิดจากการปรับเปลี่ยนตามสภาพแวดล้อมหรือสภาพของเศรษฐกิจ อีกทั้งลูกค้ายังมีความต้องการและมีการรับรู้ข่าวสารได้รวดเร็วขึ้น

2.3 กระบวนการของดาต้าไมน์นิ่ง (Process of Data Mining)



รูปที่ 2.1 การบวนการดาต้าไมน์นิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการ[illegible]ตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระบวนการในการทำดาต้าไมน์นิ่งนั้นจะเริ่มจากการกำหนดวัตถุประสงค์ในการทำ หลังจากนั้นจะนำข้อมูลที่มีอยู่ผ่านกระบวนการต่างๆ เพื่อให้ได้เป็นข้อมูลใหม่ที่เป็นความรู้ขึ้นมาดังรูป โดยกระบวนการในการทำดาต้าไมน์นิ่ง ประกอบด้วย

1) Data cleaning จะทำการลบข้อมูลส่วนที่ไม่เกี่ยวข้องกับความต้องการออกไปก่อนหรือที่เรียกว่า ข้อมูลลบกวน (Noise) เพื่อลดจำนวนข้อมูลที่มีอยู่ให้มีเฉพาะข้อมูลที่มีความจำเป็นและมีส่วนเกี่ยวข้องเท่านั้น

2) Data integration จะเป็นการรวบรวมข้อมูลที่จำเป็นที่มีอยู่ที่ คือหากว่าข้อมูลที่เราต้องการใช้งานนั้นมีอยู่อย่างมากมายและอยู่ในหลายๆ ที่เราก็จำเป็นที่จะต้องรวบรวมข้อมูลเหล่านั้นเพื่อที่สามารถที่จะใช้งานข้อมูลทั้งหมดที่มีอยู่อย่างคุ้มค่าและเกิดประโยชน์ที่สุด ดีกว่าจะตัดทิ้งไป

3) Data selection จะเป็นการเลือกข้อมูลให้ตรงกับความต้องการ ที่มีประโยชน์ต่อการใช้งาน หรือตัดข้อมูลที่ไม่ได้ส่งผลต่อผลลัพธ์ออกไป เพราะถ้ามีข้อมูลที่ไม่จำเป็นมากเกินไปจะใช้เวลาในการทำงานมากเกินความจำเป็น โดยในขั้นตอนนี้จะต้องเข้าใจถึงแหล่งกำเนินข้อมูล ความหมายของข้อมูล ประเภทของข้อมูล ลักษณะของฐานข้อมูลที่ใช้อยู่ ค่าที่เป็นไปได้และลักษณะอื่นๆ

4) Data transformation เป็นการแปลงข้อมูลให้พร้อมสำหรับการทำดาต้าไมน์นิ่ง

5) Data mining เป็นการทำไมน์นิ่งข้อมูลคือเป็นกระบวนการในการประยุกต์ใช้ข้อมูลเพื่อค้นหารูปแบบของข้อมูล (Pattern) เช่น การหากฎการจัดกลุ่มของข้อมูลที่มีอยู่ (Classification rules) การสร้างต้นไม้ของข้อมูล (Trees) หรือการทำ Clustering ฯลฯ

6) Pattern evaluation หลังจากที่ได้ทำการไมน์นิ่งข้อมูลแล้ว จะต้องทำการวิเคราะห์พิจารณารูปแบบที่ได้จากการทำไมน์นิ่ง พิจารณาความเป็นไปได้ของการเกิดรูปแบบดังกล่าว เพื่อสามารถที่จะนำรูปแบบที่ได้นั้นไปใช้ประโยชน์ได้ โดยรูปแบบบางอย่าที่ได้จากการทำดาต้าไมน์นิ่งนั้นก็ไม่สามารถที่จะนำไปใช้ประโยชน์ได้เช่นกัน ซึ่งอาจมาจากปัจจัยต่างๆ ที่เกี่ยวข้อง

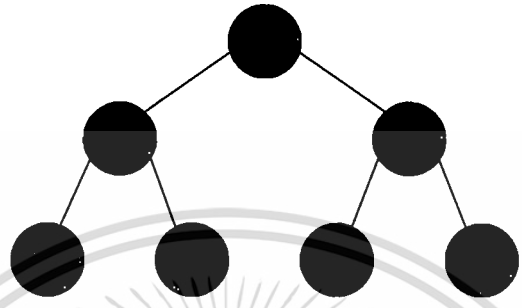
7) Knowledge presentation เมื่อผ่านกระบวนการต่างๆ เรียบร้อยแล้วก็จะได้องค์ความรู้ออกมา ขั้นสุดท้ายเป็นการแสดงความรู้และเทคนิคต่างให้กับผู้ใช้งานได้รับรู้

2.4 รูปแบบของดาต้าไมน์นิ่ง (Model of Data Mining)

1) Decision Tree เป็นลักษณะของโครงสร้างต้นไม้ซึ่งเป็นที่นิยมกันมากเนื่องจากเป็นลักษณะที่คนส่วนใหญ่คุ้นเคยกันดี จึงทำให้เข้าใจได้ง่าย ซึ่งลักษณะของ Decision Tree นั้นจะมีลักษณะคล้ายกับ แผนภูมิองค์กร หรือที่เรียกกันว่า Flow-Chart โดยแต่ละโนดนั้นจะแทนการทดสอบบนหนึ่ง Attribute คือหมายถึงว่า แต่ละโนดนั้นจะแทนคุณสมบัติของการทดสอบนั้นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แล้วแต่ละกิ่งก้านสาขาของ Decision Tree นั้นจะแสดงผลในการทดสอบและลีฟโนด แสดงคลาสที่กำหนดไว้

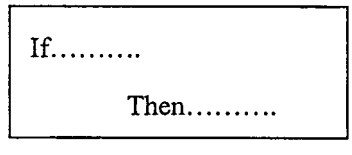


รูปที่ 2.2 แสดงลักษณะของโครงสร้างต้นไม้

2) Neural Networks นิวรอลเน็ตเวิร์ก หรือชื่อเต็มๆ ก็คือArtificial Neural Networks หรือ ANN เป็นเทคโนโลยีที่มีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ Artificial Intelligence:AI เพื่อใช้ในการคำนวณค่าฟังก์ชั่นจากกลุ่มข้อมูล วิธีการของ นิวรอลเน็ตเวิร์ก เป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบ แล้วทำการฝึกให้ระบบได้รู้จักที่จะคิดแก้ปัญหาที่เกิดขึ้นนอกเหนือจากปัญหาที่ได้ทำการทดสอบหรือปัญหาที่ได้ทำการฝึกนั้นได้ ในโครงสร้างของนิวรอลเน็ตเวิร์กจะประกอบด้วยโหนด สำหรับ Input – Output และการประมวลผล กระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ input layer , output layer และ hidden layers การประมวลผลของนิวรอลเน็ตเวิร์กจะอาศัยการส่งการทำงานผ่านโหนดต่าง ๆ ใน layer เหล่านี้

3) Nearest Neighbor Classification หรือเรียกว่า k-nearest neighbor หรือเรียกสั้นๆว่า KNN เป็นอีกหนึ่งวิธีการที่ใช้ในการทำนาย (prediction) โดย KNN classifier algorithm นั้นเป็นวิธีการคิดแยกประเภทโดยต้องการตัวแปรสำหรับเลือกในการตัดสินใจแบ่งแยกประเภท โดยเทคนิคนี้ไม่สามารถให้คำตอบที่ดีในการทำนายได้ในกรณีที่ข้อมูลมีลักษณะที่เหมือนกัน

4) Rule Induction ลักษณะของ Rule Induction นั้นจะเป็นดังรูปที่ 2.3



รูปที่ 2.3 แสดงลักษณะของ Rule Induction

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่าง

If Car = Ford and Age = 30..40

Then Defaults = Yes

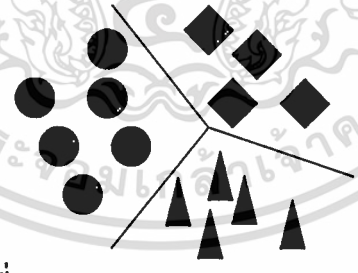
If Age = 25..35 and Prior_purchase = No

Then Defaults = No

จากตัวอย่างการนำเอา Rule Induction มาใช้ก็คือว่า ถ้าเกิดรถยี่ห้อ Ford และ มีอายุระหว่าง 30 – 40 ปี นั้นจะกำหนดค่าสถานะเริ่มต้นให้เป็น Yes แต่ถ้าหากว่ามีอายุระหว่าง 25 – 35 ปีและยังไม่เคยซื้อรถ Ford มาก่อนเลยจะกำหนดให้สถานะเริ่มต้นเป็น Noโดยลักษณะการทำงานของ Rule Induction นั้น

1. ไม่ควรมีลักษณะการทำงานที่ซ้ำซ้อนกัน
2. เริ่มต้นการพิจารณาทีละกฎ เช่น If A Then B คือถ้าใช้เงื่อนไข A ก็ไปทำ B
3. ทำการรวมปัญหาที่มีความเป็นไปได้ที่จะรวมด้วยกัน มารวมเข้าด้วยกัน
4. หากการเปรียบเทียบขนาดเล็กนั้นสามารถนำไปรวมกับการเปรียบเทียบขนาดใหญ่ได้ก็ให้นำไปรวมกัน

5) K-means Clustering จะเป็นการจัดแบ่งกลุ่มของข้อมูลให้ตรงกับความต้องการแบ่งแยกประเภทของข้อมูลที่มีลักษณะเหมือนกันให้มาอยู่ด้วยกัน โดยทำการแบ่งเป็นกลุ่มย่อยๆ



รูปที่ 2.4 ลักษณะของ K-means Clustering

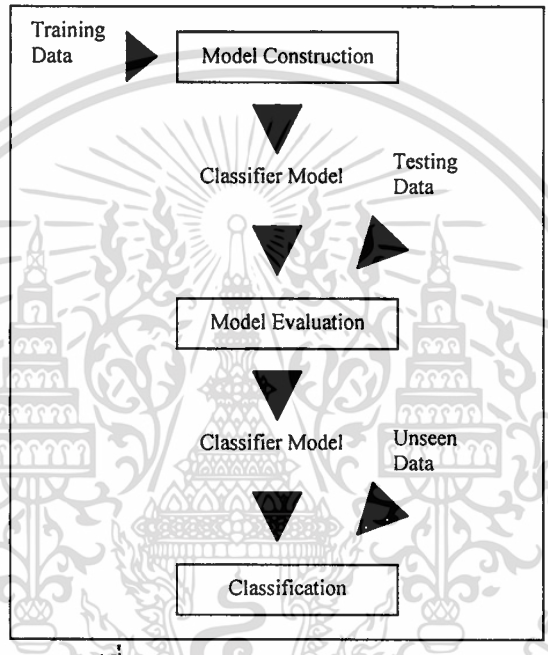
จากรูปจะเห็นได้ว่ามีการจัดแบ่งกลุ่มแยกประเภทออกตามลักษณะ โดยลักษณะที่มีความเหมือนกันก็อยู่ด้วยกัน ลักษณะที่ต่างกันก็อยู่คนละกลุ่มกัน

โดยวิธีการของ K-means Clustering นั้นจะเริ่มจากการให้ ผู้ที่ต้องการที่จะแบ่งกลุ่มนั้นกำหนดจำนวนกลุ่มที่ต้องการก่อนว่าต้องการจะแบ่งออกเป็นกี่กลุ่ม หลังจากนั้นการจะทำการสุ่ม (random) ข้อมูลแล้วทำการจัดแบ่งตามความต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 การแบ่งแยกประเภท (Classification)

Classification เป็นกระบวนการสร้าง model จัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ ตัวอย่างเช่น จัดกลุ่มนักเรียนว่า ดีมาก ดี ปานกลาง ไม่ดี โดยพิจารณาจากประวัติและผลการเรียน หรือแบ่งประเภทของลูกค้าว่าเชื่อถือได้ หรือไม่โดยพิจารณาจากข้อมูลที่มีอยู่ โดยกระบวนการ classification มีขั้นตอนดังรูปที่ 2.5



รูปที่ 2.5 กระบวนการ Classification

- Training Data คือเป็นการนำข้อมูลที่เรามีอยู่หรือข้อมูลตัวอย่าง ที่เรียกว่า Training data set มาทำการจัดกลุ่มข้อมูลโดยผ่านกระบวนการจำพวก classification rules , decision trees หรือ กระบวนการทางคณิตศาสตร์ เช่น classification rule ในการแบ่งลูกค้าของร้านขายสินค้าแห่งหนึ่ง โดยมี classification rules คือต้องการแบ่งลูกค้าออกเป็น 2 กลุ่มคือ กลุ่มลูกค้าที่มีเครดิตดี และกลุ่มลูกค้าที่มีเครดิตที่ไม่ดี

- Testing Data เป็นการนำข้อมูลที่มาทดสอบกับรูปแบบโครงสร้างต่างๆ ที่เราได้ทำการสร้างไว้แล้วโดยตัว ข้อมูลที่จะนำมาทดสอบนั้นส่วนใหญ่จะเป็นข้อมูลที่เราทราบผมอยู่แล้วว่าเมื่อได้ทำการทดลองแล้วจะได้คำตอบเป็นอย่างไรเพราะว่าถือเป็นการตรวจสอบว่า รูปแบบ ที่เราได้ทำไปนั้นถูกต้องหน้าเชื่อถือได้หรือไม่ สามารถที่จะนำไปใช้งานได้หรือยัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Unseen Data จะเป็นข้อมูลที่เราไม่เคยเห็นมาก่อนเลย และเราไม่รู้ว่าเมื่อใช้ข้อมูลตัวนี้แล้ว ผลลัพธ์ที่ได้นั้นจะมีค่าออกมาเป็นอย่างไร หรืออีกนัยหนึ่งก็คือการนำรูปแบบที่ได้นั้นไปใช้งานจริงนั้นเอง

- Model Constructionเป็นขั้นการสร้าง model โดยการเรียนรู้จากข้อมูลที่ได้กำหนดคลาสไว้เรียบร้อยแล้ว ซึ่ง model ที่ได้นั้นอาจแสดงในรูปของ

    1) แบบโครงสร้างต้นไม้ (Decision Tree)

    2) แบบ นิวรอล เน็ตเวิร์ก (Neural Network)

- Model Evaluation (Accuracy)เป็นขั้นการประมาณความถูกต้องโดยอาศัยข้อมูลที่ใช้ทดสอบ (testing data) ซึ่งคลาสที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้จะถูกนำมาเปรียบเทียบกับคลาสที่หามาได้จาก model เพื่อทดสอบความถูกต้อง

- Model Usage (Classification) เป็น Model สำหรับใช้ข้อมูลที่ไม่เคยเห็นมาก่อน (unseen data) โดยจะทำการกำหนดคลาสให้กับ object ใหม่ที่ได้มา หรือ ทำนายค่าออกมาตามที่ต้องการ

2.6 ดาต้าไมน์นิ่งช่วยระบบธุรกิจได้อย่างไร

ในระบบธุรกิจ หากว่าธุรกิจมีขนาดเล็กจนแทบจะรู้จักกับลูกค้าทุกคนได้เป็นอย่างดี มีความสัมพันธ์กับลูกค้าเป็นอย่างดี สามารถที่จะตอบสนองความต้องการของลูกค้าได้อย่างครบถ้วนสมบูรณ์ แต่อย่างไรก็ตาม เมื่อธุรกิจมีขนาดใหญ่โตขึ้น ดังเช่นธุรกิจจำนวนมากในปัจจุบัน สถานการณ์ก็แปรเปลี่ยนไป ดังนั้นแนวความคิดเกี่ยวการรวบรวมข้อมูล เพื่อนำมาวิเคราะห์หาความจริงในเรื่องความต้องการต่างๆ ของลูกค้า ตลอดจนแนวโน้มความเป็นไปได้ต่างๆ ดังที่เรียกกันว่า ดาต้าไมน์นิ่ง (Data Mining) จึงถือว่าเป็นเครื่องมือสำคัญในการดำเนินธุรกิจ เช่น ธุรกิจผับโดยบางผับใหญ่ที่มีสาขาในอังกฤษ ใช้ซอฟต์แวร์ เพื่อประเมินยอดขาย อันเป็นผลจากการลดราคาของเบียร์แต่ละชนิดในช่วง "Happy Hours" ถ้าเบียร์ชนิดใดลดราคาแล้วยอดขายพุ่ง รุ่งขึ้นก็จะลดราคาซ้ำแบบเดิม ถ้ายอดขายไม่เพิ่มก็จะลองลดราคาตัวใหม่ไปเรื่อยๆ Data Mining แบบทันใจเช่นนี้มีส่วนอย่างสำคัญในการวางกลยุทธ์ปรับราคาสินค้า ทั้งนี้ ต้องสอดคล้องกับสต๊อกสำรองของสินค้า และระบบการบริหารที่ทำให้สามารถปรับราคาได้อย่างรวดเร็วด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

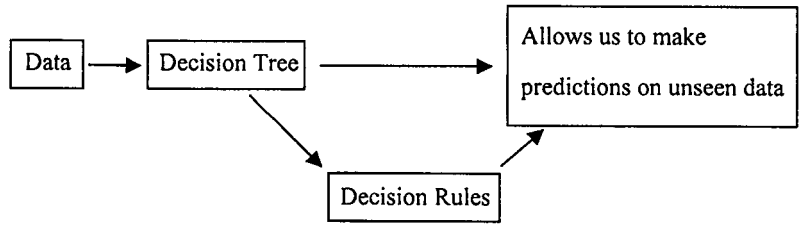
บทที่ 3
ดิซิชั่นทรี

3.1 ความหมายของดิซิชั่นทรี

ดิซิชั่นทรี (Decision Tree) หมายถึง โครงสร้างต้นไม้เพื่อช่วยในการตัดสินใจ โดยในกระบวนการทางด้านดาต้าไมน์นิ่งนั้นได้นำเอาดิซิชั่นทรีมาใช้เป็นตัวช่วยในการทำงานเกี่ยวกับทางด้านการช่วยในการตัดสินใจต่างๆ ไม่ว่าจะเป็นทางด้านระบบธุรกิจต่างๆ เช่น ช่วยตัดสินในเกี่ยวกับการดำเนินกิจการการวางแผนงานของระดับผู้บริหารองค์กรต่างๆ เนื่องจาก Decision Tree เป็นลักษณะที่คนส่วนใหญ่คุ้นเคยกันดีอยู่แล้ว จึงทำให้เข้าใจได้ง่าย โดยลักษณะของ Decision Tree นี้จะมีลักษณะคล้ายกับ แผนภูมิองค์กร หรือที่เรียกกันก็คือ Flow – Chart นั่นเอง

3.2 กระบวนการทำงานของดิซิชั่นทรี

กระบวนการทำงานของดิซิชั่นทรี (Decision tree) นั้นจะเริ่มจากการที่เรามีข้อมูลแล้วนำข้อมูลตัวนั้นแตกออกมาทำเป็นโครงสร้างต้นไม้หรือที่เรียกว่าดิซิชั่นทรีโดยโครงสร้างของต้นไม้ที่ได้นั้นจะเป็นโครงสร้างที่มีกฎต่างๆ เกิดขึ้นตามเป้าหมายของการใช้งานในแต่ละงาน เมื่อเราได้โครงสร้างต้นไม้ขึ้นมาแล้ว เราก็สามารถนำไปใช้งานได้ไม่ว่าจะเป็นจากข้อมูลทั่วๆ ไปโดยนำข้อมูลนั้นมาผ่าน Decision Rules เพื่อให้ได้โครงสร้างต้นไม้ที่ตรงตามความต้องการเพื่อนำไปใช้งาน เช่น ช่วยในการทำนายค่าความต้องการซื้อของของลูกค้าหรือดูโอกาสการเจริญเติบโตของธุรกิจที่มีในหลายๆ ด้านว่าเราควรขยายตัวในด้านไหน



รูปที่ 3.1 การทำงานของดิซิชั่นทรี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 ขั้นตอนการสร้างดิซิชั่นทรี

ในการสร้างดิซิชั่นทรี (Decision Tree) นั้นจะเริ่มจากการที่เรามีข้อมูลอยู่แล้วนำข้อมูลมาทำการแตกออกมาทำเป็นโครงสร้างต้นไม้ โดยโครงสร้างของต้นไม้ที่ได้นั้นจะเป็นโครงสร้างที่มีกฎต่างๆ เกิดขึ้นตามเป้าหมายของการใช้งานในแต่ละงาน โดยเมื่อเราได้โครงสร้างต้นไม้แล้วเราก็สามารถนำโครงสร้างที่ได้นั้นไปใช้งานกับข้อมูลอื่นๆ ได้ โดยนำข้อมูลนั้นผ่าน Decision Rules เพื่อให้ได้โครงสร้างต้นไม้ตามความต้องการ โดยขั้นตอนหลักๆของการสร้างดิซิชั่นทรี คือ

1) ต้องทำการหา Root node ก่อนโดยการหา Attribute ที่มีความน่าเชื่อถือมากที่สุด โดยสามารถหาได้จากอัลกอริทึมต่างๆ เช่น ID3 , C4.5 ฯลฯ

2) หลังจากที่ได้ Root node แล้วก็ทำการแตกค่าความเป็นไปได้ของ Root node นั้นออกมาเป็นแต่ละกิ่งของต้นไม้

3) ทำการแบ่งข้อมูลตามกิ่งของต้นไม้

4) นำข้อมูลแต่ละกลุ่มที่ได้มีการแบ่งกลุ่มแล้วมาทำซ้ำตั้งแต่ขั้นตอนที่หนึ่ง

3.4 อัลกอริทึม ID3

จะเป็นการหาค่าความน่าเชื่อถือของข้อมูลของแต่ละ Attribute ว่า Attribute ไหนมีความน่าเชื่อถือเหมาะที่จำไปใช้

<u>สูตรการคำนวณ Entropy</u>

$$\text{Entropy} = \sum_{c} -\left(\frac{N_{bc}}{N_b}\right)\log_2\left(\frac{N_{bc}}{N_b}\right)$$

<u>สูตรการคำนวณ Average Entropy(x)</u>

$$\text{Average Entropy(x)} = \sum_{b}\left(\frac{N_b}{N_t}\right)\times\left[\sum_{c} -\left(\frac{N_{bc}}{N_b}\right)\log_2\left(\frac{N_{bc}}{N_b}\right)\right]$$

<u>สูตรการคำนวณ Gain(x)</u>

Gain(x) = Entropy – Average Entropy

กำหนดให้

N_b = จำนวนตัวอย่างของกิ่ง b

N_{bc} = จำนวนของตัวอย่างของกิ่ง b ของ class c

N_t = จำนวนตัวอย่างทั้งหมดของกิ่งๆ นั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างการคำนวณโดยใช้อัลกอริทึม ID3

ตารางที่ 3.1 แสดงข้อมูลตัวอย่างในการใช้อัลกอริทึม ID3

Name	Hair	Height	Weight	Lotion	Result
Sarah	blonde	average	light	no	Sunburned
Dana	blonde	tall	average	yes	None
Alex	brown	short	average	yes	None
Annie	blonde	short	average	no	Sunburned
Emily	red	average	heavy	no	Sunburned
Pete	brown	tall	heavy	no	None
John	brown	average	heavy	no	None
Katie	blonde	short	light	yes	None

ทำการหาค่า Entropy

$$\text{Entropy} = -3/8 * \log_2(3/8) - 5/8 * \log_2(5/8)$$

$$= 0.954 \text{ bits}$$

พิจารณาที่ attribute Hair ซึ่งแบ่งข้อมูลได้เป็น 3 subset

$$\text{Average Entropy}(x) = 4/8 * (-2/4 * \log_2(2/4) - 2/4 * \log_2(2/4))$$
$$+ 3/8 * (-3/3 * \log_2(3/3))$$
$$+ 1/8 * (-1/1 * \log_2(1/1))$$
$$= 0.5 \text{ bits}$$

ดังนั้นค่า Gain ของ attribute Hair คือ

$$\text{Gain}(x) = 0.954 - 0.5 = 0.454 \text{ bits}$$

พิจารณาที่ attribute Height ซึ่งแบ่งข้อมูลได้เป็น 3 subset

$$\text{Average Entropy}(x) = 3/8 * (-2/3 * \log_2(2/3) - 1/3 * \log_2(1/3))$$
$$+ 2/8 * (-2/2 * \log_2(2/2))$$
$$+ 3/8 * (-2/3 * \log_2(2/3) - 1/3 * \log_2(1/3))$$
$$= 0.69 \text{ bits}$$

ดังนั้นค่า Gain ของ attribute Height คือ

$$\text{Gain}(x) = 0.954 - 0.69 = 0.264 \text{ bits}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พิจารณาที่ attribute Weight ซึ่งแบ่งข้อมูลได้เป็น 3 subset

Average Entropy(x) = $2/8 * (-1/2 * \log_2(1/2) - 1/2 * \log_2(1/2))$

$+ 3/8 * (-2/3 * \log_2(2/3) - 1/3 * \log_2(1/3))$

$+ 3/8 * (-2/3 * \log_2(2/3) - 1/3 * \log_2(1/3))$

= 0.94 bits

ดังนั้นค่า Gain ของ attribute Weight คือ

Gain(x) = 0.954 - 0.94 = 0.014 bits

พิจารณาที่ attribute Lotion ซึ่งแบ่งข้อมูลได้เป็น 2 subset

Average Entropy(x) = $5/8 * (-3/5 * \log_2(3/5) - 2/5 * \log_2(2/5))$

$+ 3/8 * (-2/3 * \log_2(2/3))$

= 0.61 bits

ดังนั้นค่า Gain ของ attribute Lotion คือ

Gain(x) = 0.954 - 0.61 = 0.344 bits

ดังนั้น

Attribute	**Gain(x)**
Hair Color	0.454
Height	0.264
Weight	0.014
Lotion	0.344

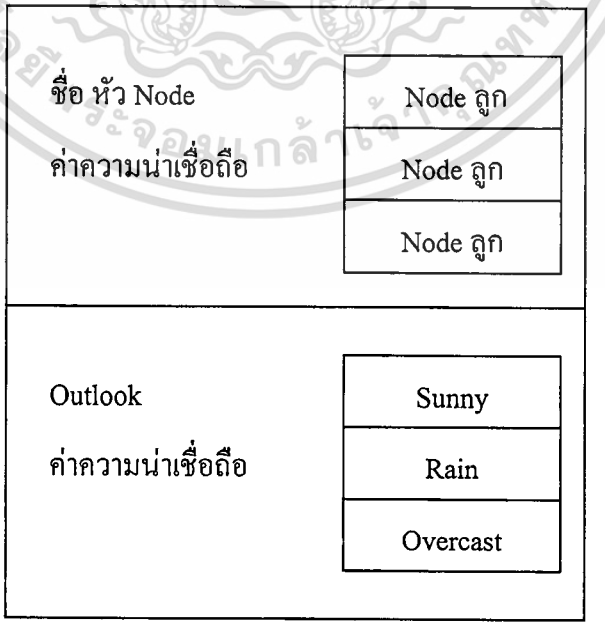
ซึ่งจะเห็นได้ว่า Attribute Hair Color มีค่า Gain(x) มาที่สุดจึงมีค่าความน่าเชื่อถือมากที่สุด จึงเหมาะที่สุดที่จะเป็น Root Node

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
วิเคราะห์และออกแบบโปรแกรม

4.1 วิเคราะห์และออกแบบโปรแกรม

ระบบ Decision Tree เป็นระบบที่จะทำการติดต่อกับ Relational Database Management Systems คือ Microsoft SQL Server เพื่อจะทำการเลือก ฐานข้อมูล ตาราง และฟิลด์ต่างๆ ที่ต้องการเพื่อที่จะนำมาใช้ในการคำนวณ เพื่อทำการสร้างโครงสร้างต้นไม้ออกมาตามอัลกอริทึมที่ได้ทำการเลือกใช้นั้นก็คือ อัลกอริทึม ID3 โดยค่าของข้อมูลต่างๆ ที่ได้ทำการดึงขึ้นมาจากฐานข้อมูลจำเป็นที่จะต้องมีการตรวจสอบข้อมูลว่ามีค่าว่างหรือไม่ และทำการแก้ไข โดยหากว่าข้อมูลนั้นมีชนิดของข้อมูลที่เป็นตัวเลขก็จะเป็นที่จะแปลงข้อมูลเป็นตัวอักษร โดยระบบมีความสามารถในการแปลงข้อมูลออกเป็น 2 ประเภทด้วยกันก็คือ สามารถแปลงได้เป็น 2 กลุ่ม หรือ 3 กลุ่มตามแต่ความต้องการ โดยในการคำนวณนั้นต้องคำนึงถึงการเก็บข้อมูล ที่จำเป็นที่จะต้องนำไปสร้างเป็นโครงสร้างต้นไม้ให้ได้ รวมทั้งคำนึงถึงการจัดเก็บข้อมูลเพื่อสามารถที่จะนำกลับมาใช้ใหม่ได้ จึงมีแนวความคิดในการจัดเก็บข้อมูลในขณะคำนวณนั้น จะจัดเก็บอยู่ในรูปของ array ซึ่งลักษณะของโครงสร้าง array สามารถดูได้จากรูปที่ 4.1



รูปที่ 4.1 แสดงโครงสร้าง Array ในการจัดเก็บข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนในการทดสอบโมเดลนั้นจะต้องการดึงรูปแบบโครงสร้างของโมเดลที่ได้ทำการบันทึกลงไปในฐานข้อมูลกลับขึ้นมาแสดงผล พร้อมทั้งทำการออกแบบการติดต่อกับฐานข้อมูลเพื่อที่จะนำข้อมูลที่ใช้ในการทดสอบมาทำการแม็พกับฐานข้อมูลที่ในการสร้างโมเดล หลังจากนั้นก็ทำการทดสอบและแสดงผลลัพธ์ออกมา

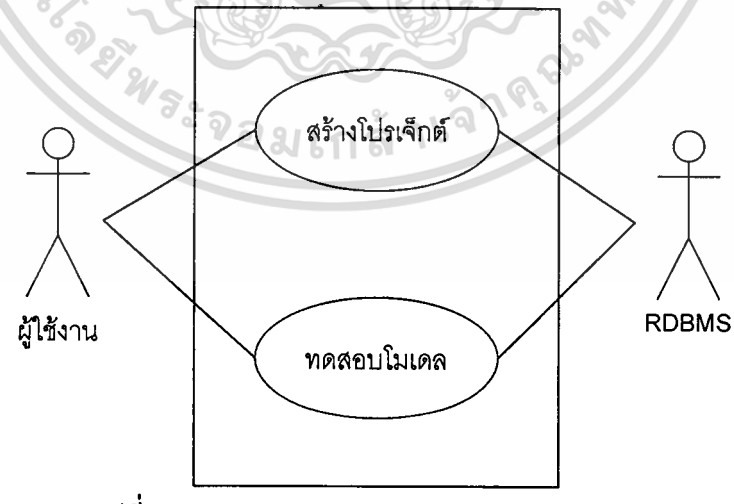
4.2 Process Model

4.2.1 Use Case Diagram

Use Case Diagram แสดงถึงระบบการทำงานว่าทำอะไรได้บ้าง มีการทำงานหลักๆ อะไร ผู้ใช้ใช้อะไรได้บ้าง ซึ่งตัวแปรแกรมได้แบ่งฟังก์ชันงานออกเป็น 2 ส่วนด้วยกัน คือ การสร้างโปรเจ็กต์เพื่อทำการสร้างโมเดลขึ้นมา และการทดสอบโมเดลที่ได้ทำการสร้างขึ้นมา โดย

1) การสร้างโปรเจ็กต์ จะเป็นการสร้างงานใหม่ขึ้นมา โดยจะทำการติดต่อกับ Relational Database Management systems คือ Microsoft SQL Server เพื่อทำการเลือกข้อมูลที่ต้องการสำหรับการสร้างโครงสร้างต้นไม้ขึ้นมาหลังจากนั้นก็จะทำการบันทึกข้อมูลรูปแบบของโมเดลที่ทำการสร้างขึ้นมาลงฐานข้อมูล

2) ทดสอบโมเดล จะเป็นการดึงข้อมูลรูปแบบโมเดลที่ได้ทำการบันทึกลงฐานข้อมูลเอาไว้นั้น มาทำการทดสอบกับฐานข้อมูลอื่น ที่มีลักษณะของมูลความเป็นไปได้ของค่าแต่ละฟิลด์ที่มีลักษณะเหมือนกัน



รูปที่ 4.2 แสดงUse Case ของระบบ Decision Tree

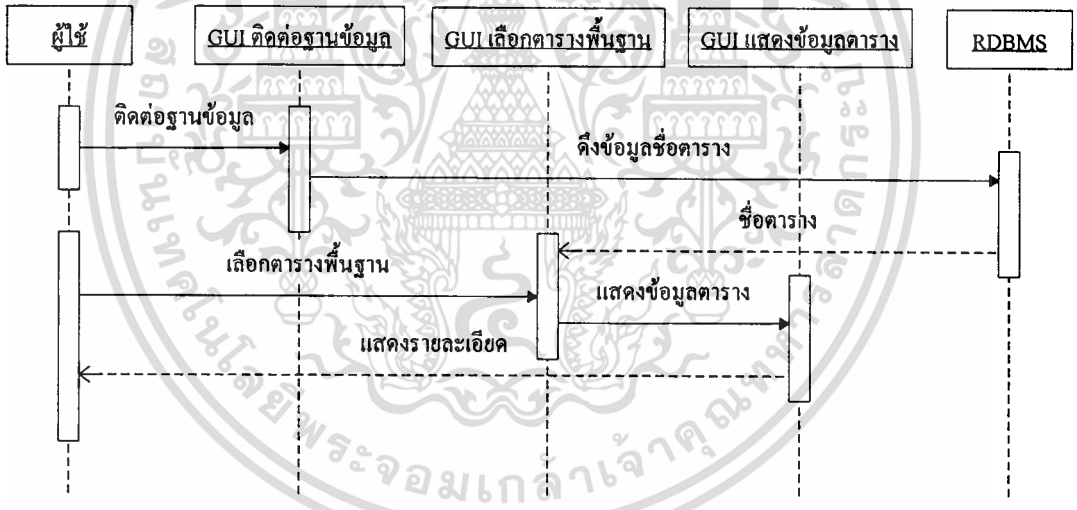
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

4.2.2 Sequence Diagram

1) Sequence Diagram ของการติดต่อฐานข้อมูล

- ผู้ใช้ทำการเรียกใช้ GUIติดต่อฐานข้อมูล ให้ทำการติดต่อกับฐานข้อมูลที่ต้องการ โดยผู้ใช้จะต้องใส่เชื่อเซิฟเวอร์ และ ดาต้าเบสที่ต้องการลงไป
- ต่อมา GUIติดต่อฐานข้อมูล จะทำการติดต่อฐานข้อมูลและดาต้าเบสที่ผู้ใช้ได้ระบุไว้ แล้วทำการการดึงข้อมูลที่ต้องการขึ้นมายัง GUIเลือกตารางพื้นฐานโดนจะแสดงตารางทั้งหมดที่มีอยู่
- ผู้ใช้ทำการเลือกตารางพื้นฐานที่ต้องการจากหน้าจอ GUIเลือกตารางพื้นฐาน ที่ต้องการ
- หลังจากนั้น GUIแสดงข้อมูลตาราง จะทำการแสดงข้อมูลที่ผู้ใช้ได้ทำการเลือก

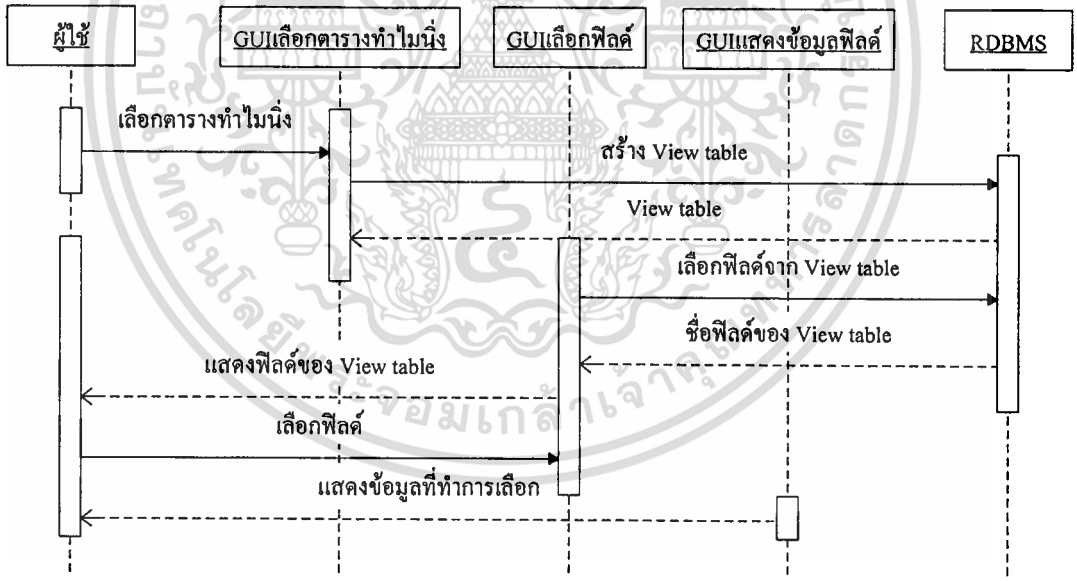


รูปที่ 4.3 แสดง Sequence Diagram ของการติดต่อฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) Sequence Diagram ของการเลือกข้อมูลสำหรับทำไมน์นิ่ง

- ผู้ใช้ทำการเลือกตารางข้อมูลสำหรับทำไมน์นิ่ง
- GUIเลือกตารางทำไมน์นิ่ง ทำการเรียก RDBMS ให้ทำการสร้าง View table ตามที่ผู้ใช้ได้ทำการเลือกเอาไว้
- RDBMS ทำการสร้าง View table แล้วส่งค่ากลับมายัง GUIเลือกตารางทำไมน์นิ่ง
- GUI เลือกฟิลด์ จะทำการดึงชื่อฟิลด์ของ View table จาก RDBMS
- RDBMS จะทำการดึงข้อมูลชื่อฟิลด์ของ View table หลังจากนั้น GUIเลือกฟิลด์ก็จะแสดงชื่อฟิลด์ต่างๆ ให้ผู้ใช้
- ผู้ใช้ทำการเลือกฟิลด์ที่ต้องการ
- GUIแสดงข้อมูลฟิลด์ จะแสดงข้อมูลฟิลด์ที่ผู้ใช้ได้ทำการเลือกไว้ให้กับผู้ใช้งานได้ตรวจสอบว่าถูกต้องหรือไม่

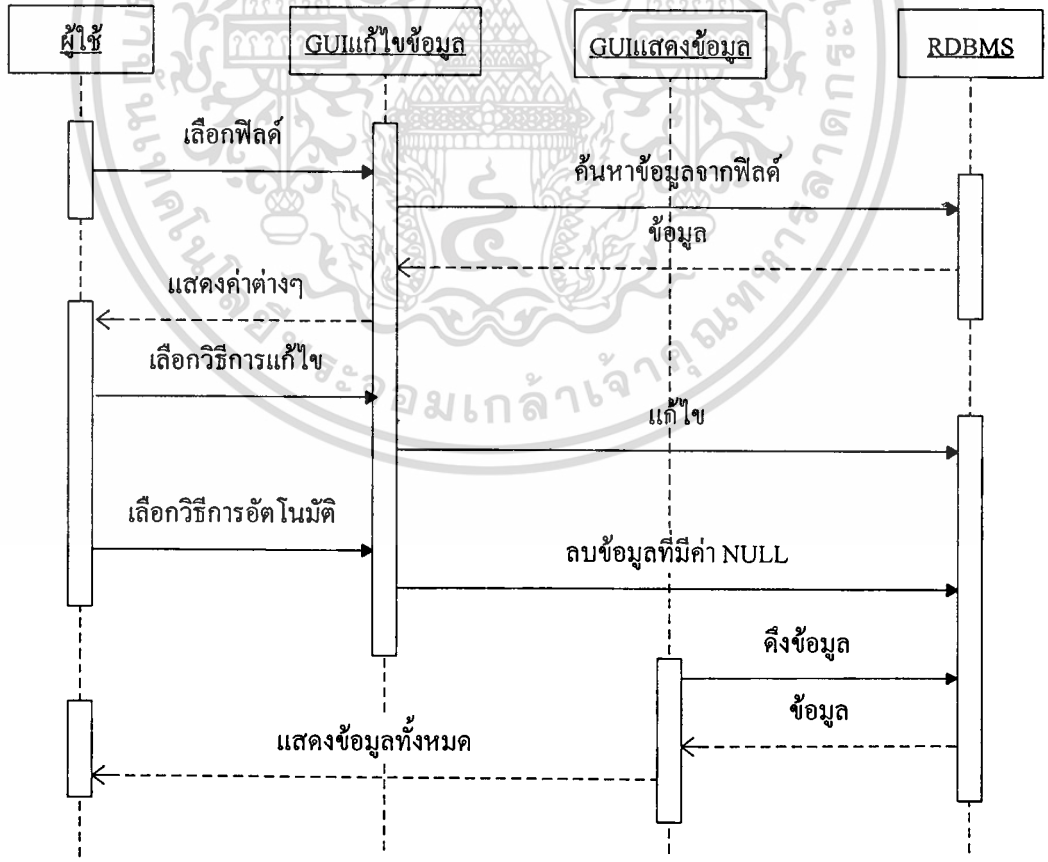


รูปที่ 4.4 แสดง Sequence Diagram ของการเลือกข้อมูลสำหรับทำไมน์นิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) Sequence Diagram ของการแก้ไขข้อมูล

- ผู้ใช้ทำการเลือกฟิลด์แต่ละฟิลด์สำหรับตรวจสอบและทำการแก้ไขข้อมูล
- GUIแก้ไขข้อมูล ทำการเรียกให้ RDBMS ค้นหาข้อมูลตามฟิลด์ที่ผู้ใช้ได้ทำการเลือก
- RDBMS ส่งข้อมูลที่ GUIแก้ไขข้อมูล ต้องการ
- GUIแก้ไขข้อมูล ทำการแสดงผมข้อมูลต่างๆ ที่ได้การดึงมาจากฐานข้อมูลให้ผู้ใช้ได้ทราบถึงข้อมูลต่างๆ พร้อมทั้งแสดงวิธีการที่ผู้ใช้ต้องการที่แก้ไขข้อมูลชุดนี้
- ผู้ใช้ทำการเลือกวิธีการแก้ไขข้อมูลหากข้อมูลมีค่าว่าง
- GUIแก้ไขข้อมูล เรียกไปยัง RDBMS ให้ทำการแก้ไขข้อมูล
- หรือหากว่าผู้ใช้ทำการเลือกให้แก้ไขแบบอัตโนมัติ GUIแก้ไขข้อมูลจะทำการลบข้อมูลที่มีค่า NULL ออก
- GUIแสดงข้อมูลจะทำการดึงข้อมูลทั้งหมดจาก RDBMS แสดงให้กับผู้ใช้งาน

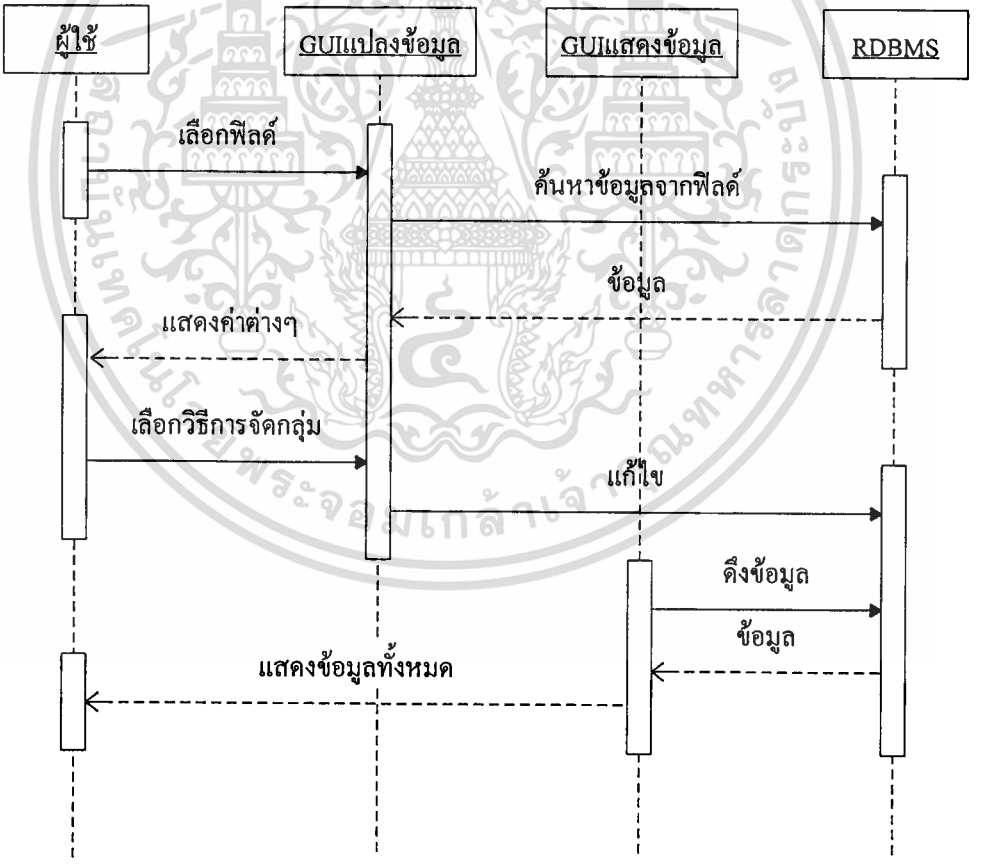


รูปที่ 4.5 แสดง Sequence Diagram ของการแก้ไขข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) Sequence Diagram ของการแปลงข้อมูล

- ผู้ใช้ทำการเลือกฟิลด์ข้อมูลที่ต้องการแปลงข้อมูล
- GUIแปลงข้อมูล ทำการค้นหาข้อมูลจากฟิลด์ที่ผู้ใช้ได้ทำการเลือก
- RDBMS ส่งข้อมูลที่ต้องการกลับมา
- GUIแปลงข้อมูล แสดงค่าต่างๆ ให้ผู้ใช้ทราบ
- ผู้ใช้ทำการเลือกวิธีการจัดกลุ่มว่าต้องการกี่กลุ่มและแต่ละกลุ่มจะแบ่งอย่างไร
- GUIแปลงข้อมูล สั่งให้ RDBMS ทำการแก้ไขข้อมูลตามที่ผู้ใช้เลือกวิธีการแก้ไข
- GUIแสดงข้อมูล ทำการดึงข้อมูลทั้งหมด
- RDBMS ส่งข้อมูลที่ GUIแสดงข้อมูล ต้องการ
- GUIแสดงข้อมูล ทำการแสดงข้อมูลค่าทั้งหมดให้กับผู้ใช้งาน



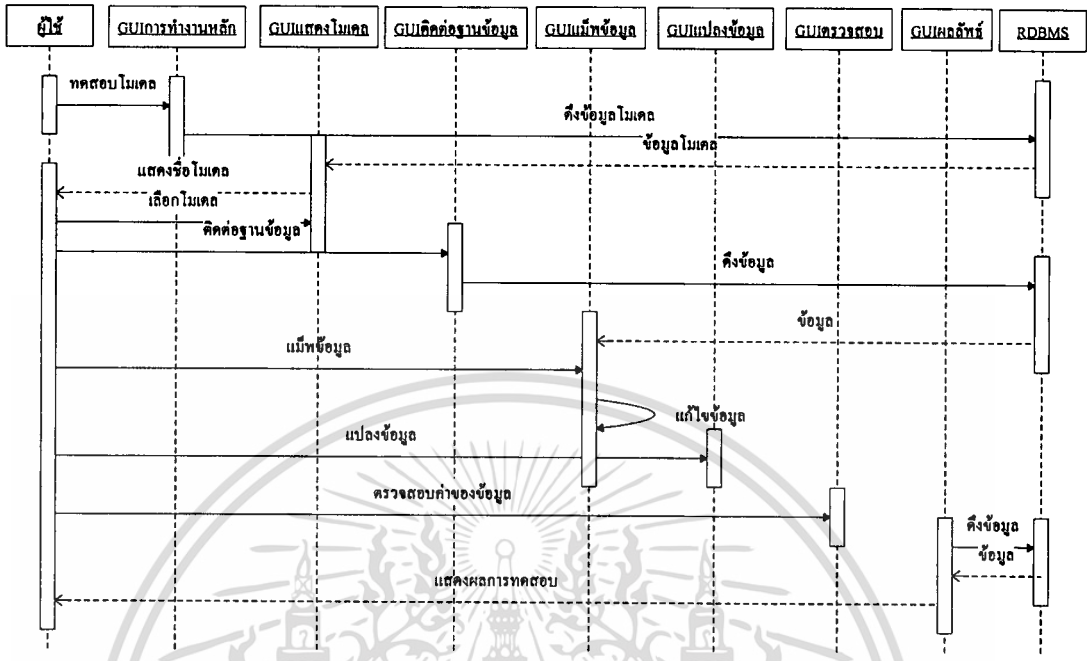
รูปที่ 4.6 แสดง Sequence Diagram ของการแปลงข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) Sequence Diagram ของการทดสอบโมเดล

- ผู้ใช้ทำการทดสอบโมเดล
- GUIแสดงโมเดล ทำการดึงข้อมูลชื่อโมเดลต่างๆ ขึ้นมาแสดงให้ผู้ใช้ดู
- ผู้ใช้ทำการเลือกโมเดลที่ต้องการจะทำการทดสอบ
- ผู้ใช้ทำการติดต่อฐานข้อมูลที่ต้องการจะนำมาทดสอบกับโมเดลที่ได้ทำการเลือกเอาไว้
- GUIติดต่อฐานข้อมูลจะทำการติดต่อฐานข้อมูลแล้วดึงข้อมูลมาแสดงยังGUIแม็พข้อมูล โดยให้ผู้ใช้งานทำการแม็พข้อมูลที่ต้องการ แล้ว GUIแม็พข้อมูลจะทำการแก้ไขข้อมูลคือจะลบข้อมูลที่มีค่า NULL ออกไป
- หากข้อมูลมีค่าเป็นตัวเลขผู้ใช้ก็จะต้องทำการแปลงข้อมูลในหน้า GUIแปลงข้อมูล
- โดยในการแปลงข้อมูลนั้นผู้ใช้ก็สามารถทำการแปลงข้อมูลได้คือสามารถเลือกการจัดกลุ่มว่าต้องการที่จะจัดกลุ่มกี่กลุ่มและแต่ละกลุ่มต้องการที่จะแบ่งค่าอย่างไร
- GUIตรวจสอบ จะทำการแสดงข้อมูลทั้งหมดออกมาแล้วให้ผู้ใช้ทำการตรวจสอบข้อมูลว่าสามารถที่จะทำการทดสอบโมเดลได้หรือไม่ โดยในการทดสอบนั้นจะดูว่าข้อมูลที่ได้ทำการแม็พกันมานั้นค่าความเป็นไปได้ของข้อมูลในแต่ละฟิลด์นั้นเหมือนกันหรือไม่ หากเหมือนกันก็สามารถที่จะทำการทดสอบได้
- GUIผลลัพธ์ จะทำการดึงข้อมูลที่ได้ทำการแม็พไว้แล้วทำการตรวจสอบกับโมเดลที่เราได้ทำการเลือกไว้ตั้งแต่ต้นว่าผลลัพธ์จะออกมาเป็นอย่างไร แล้วแสดงผลลัพธ์ที่ได้ให้กับผู้ใช้รับทราบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

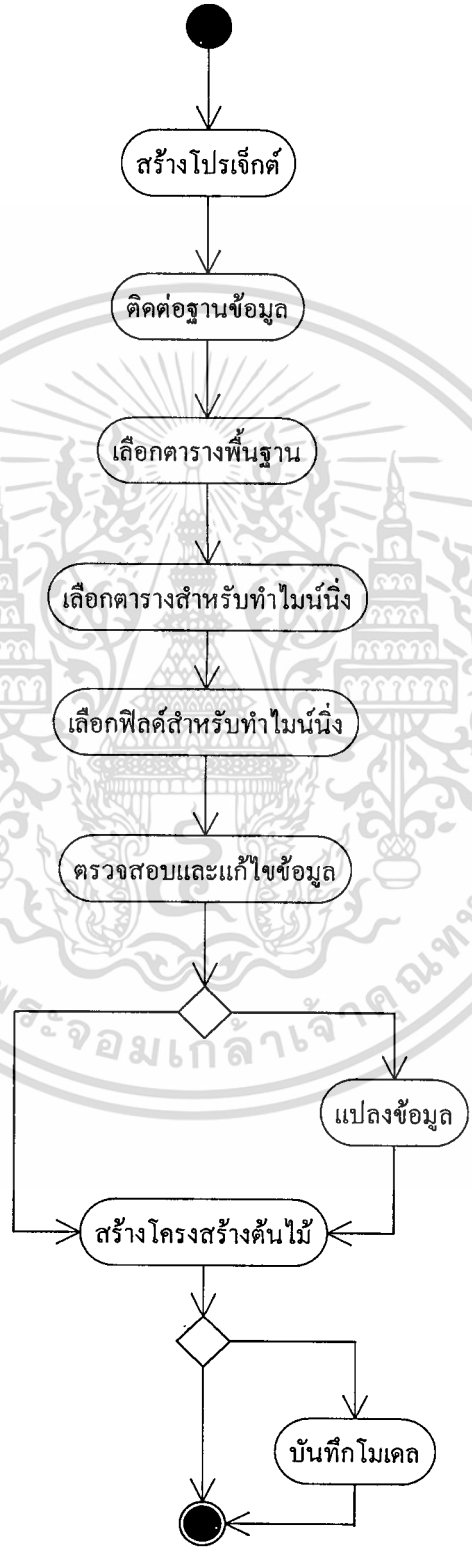


รูปที่ 4.7 แสดง Sequence Diagram ของการทดสอบโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 Activity Diagram

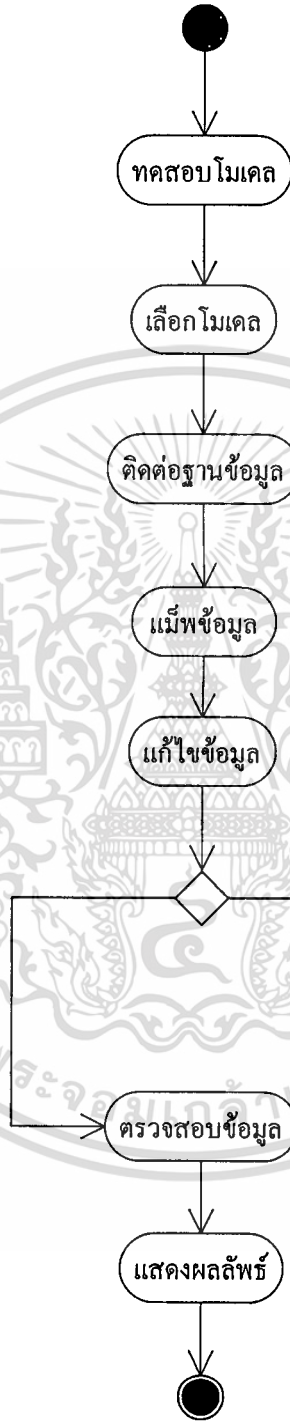
1) Activity Diagram ของการสร้างโปรเจ็กต์



รูปที่ 4.8 แสดง Activity Diagram ของการสร้างโปรเจ็กต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

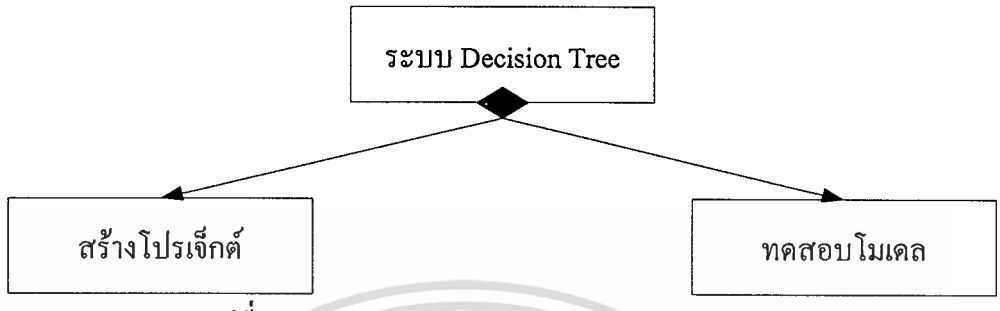
2) Activity Diagram ของการทดสอบโมเดล



รูปที่ 4.9 แสดง Activity Diagram ของการทดสอบโมเดล

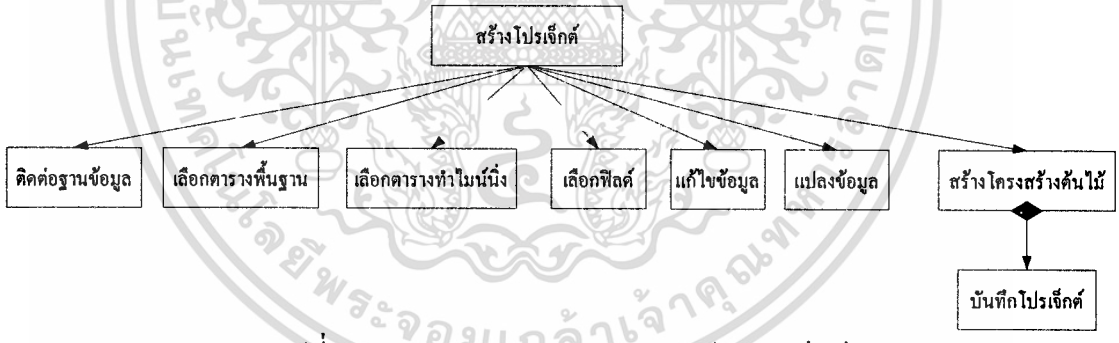
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 Structure Chart และ Flow – Chart แสดงการทำงาน



รูปที่ 4.10 Structure Chart แสดงระบบ Decision Tree

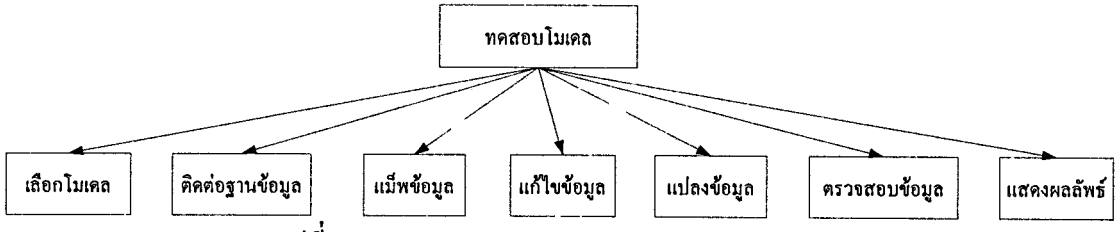
ระบบ Decision Tree จะประกอบไปด้วย 2 ส่วนหลักๆ ด้วยกัน คือ การสร้างโปรเจ็กต์และการทดสอบโมเดล โดย ในการสร้างโปรเจ็กต์นั้นจะเป็นการสร้างโมเดลของ Decision Tree ขึ้นมาตามความต้องการของผู้ใช้งานและสามารถที่จะบันทึกโมเดลนั้นได้ ส่วนการทดสอบโมเดลนั้นจะเป็นการนำโมเดล Decision Tree ที่ได้ทำการสร้างและบันทึกเอาไว้มาทำการทดสอบกับข้อมูลอื่นๆ



รูปที่ 4.11 Structure Chart แสดงการสร้างโปรเจ็กต์

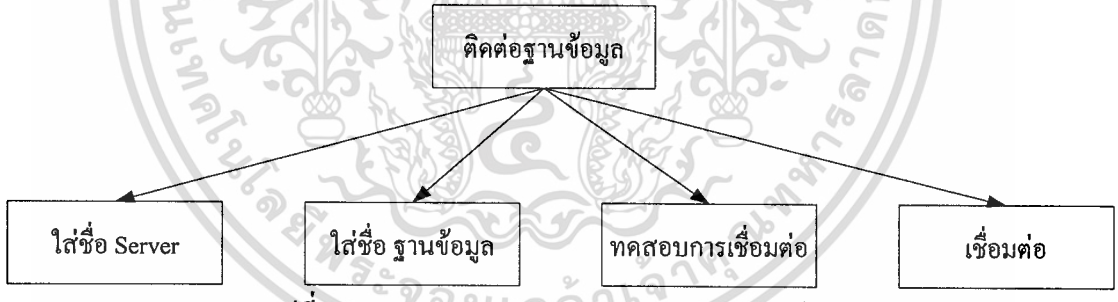
ในการสร้างโปรเจ็กต์นั้นจะเริ่มจากให้ผู้ใช้งานทำการติดต่อฐานข้อมูลที่ต้องการจะนำมาสร้างโมเดล Decision Tree หลักจากนั้นให้ผู้ใช้งานทำการเลือกตารางพื้นฐาน ในที่นี้ก็คือจะเป็นการทดสอบว่าผู้ใช้งานนั้นสามารถที่จะทำการติดต่อกับตารางที่ต้องการได้หรือไม่ ต่อมาผู้ใช้ต้องทำการเลือกตารางข้อมูลสำหรับทำไมน์นิ่ง ในที่นี้ก็คือเลือกตารางสำหรับทำ Decision Tree ต่อมาก็ทำการเลือกฟิลด์จากตารางที่ได้ทำการเลือกไว้ ต่อมาก็ต้องทำการแก้ไขข้อมูลในกรณีที่ข้อมูลมีค่าว่าง ต่อมาหากมีฟิลด์ข้อมูลมีค่าเป็นตัวเลขก็ต้องทำการแปลงข้อมูลจากตัวเลขเป็นตัวอักษร หลังจากนั้นก็ทำการสร้างโครงสร้างต้นไม้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 Structure Chart แสดงการทดสอบโมเดล

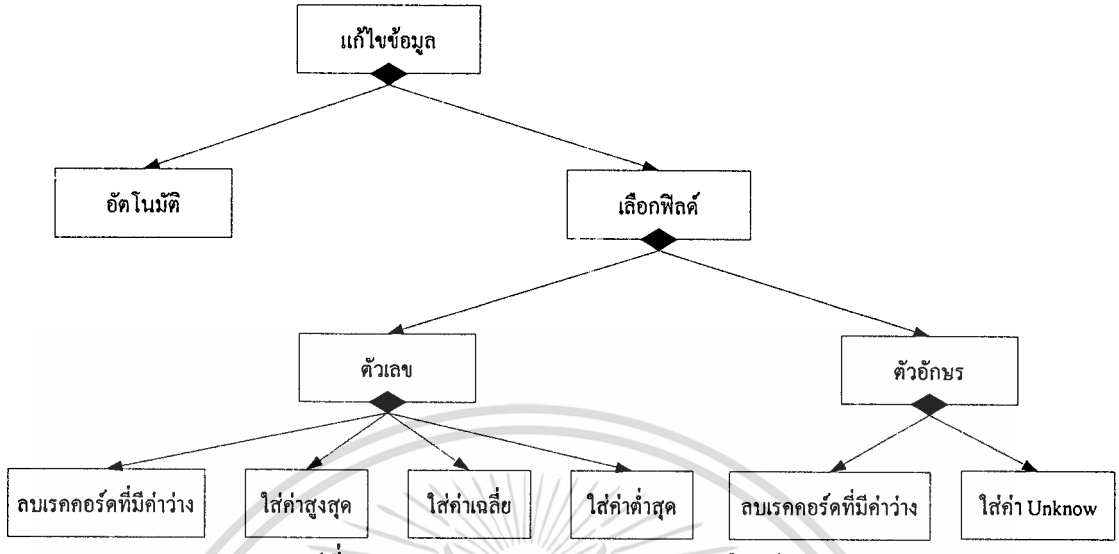
การทดสอบโมเดล ให้ผู้ใช้ทำการเลือกโมเดลที่ต้องการจะทำการทดสอบ หลังจากนั้นให้ผู้ใช้ทำการติดต่อฐานข้อมูลที่ผู้ใช้ต้องการจะนำมาทดสอบกับโมเดลที่ได้ทำการเลือกไว้ ทำการแม็พข้อมูลระหว่าง ข้อมูลของโมเดลและข้อมูลของฐานข้อมูลที่จะนำมาทดสอบ ต่อมาก็ต้องทำการแก้ไขข้อมูลที่ค่าว่าง โดยตัวโปรแกรมจะทำการลบ record ที่มีค่าว่างทิ้งไปทั้งหมด หลังจากนั้นก็ทำการแปลงข้อมูลโดยการแปลงข้อมูลจะทำการแปลงข้อมูลเหมือนกับการแปลงข้อมูลของฟิลด์ที่ได้ทำการแม็พกับตัวโมเดล ต่อมาระบบก็จะทำการตรวจสอบข้อมูลโดยดูว่าค่าความเป็นไปได้ของข้อมูลแต่ละฟิลด์ของข้อมูลที่นำมาทดสอบนั้นเป็น subset ของข้อมูลโมเดลหรือไม่ ถ้าใช้งานได้ก็จะแสดงผลลัพธ์ออกมา



รูปที่ 4.13 Structure Chart แสดงการติดต่อฐานข้อมูล

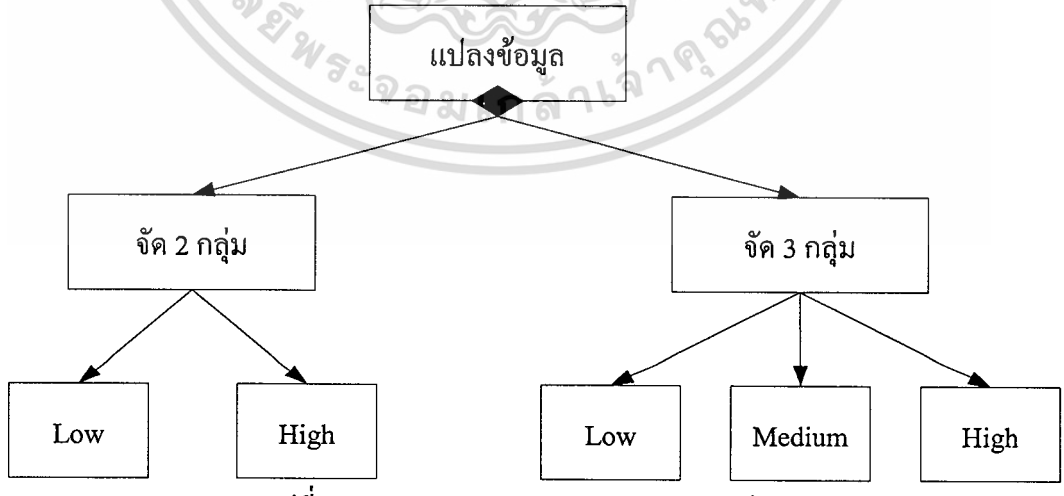
ในการติดต่อฐานข้อมูลให้ผู้ใช้ทำการใส่ชื่อ Server และใส่ชื่อ Database ที่ต้องการจะติดต่อ โดยผู้ใช้งานสามารถที่จะทำการทดสอบการเชื่อมต่อฐานข้อมูลได้ แล้วก็ทำการเชื่อมต่อจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.14 Structure Chart แสดงการแก้ไขข้อมูล

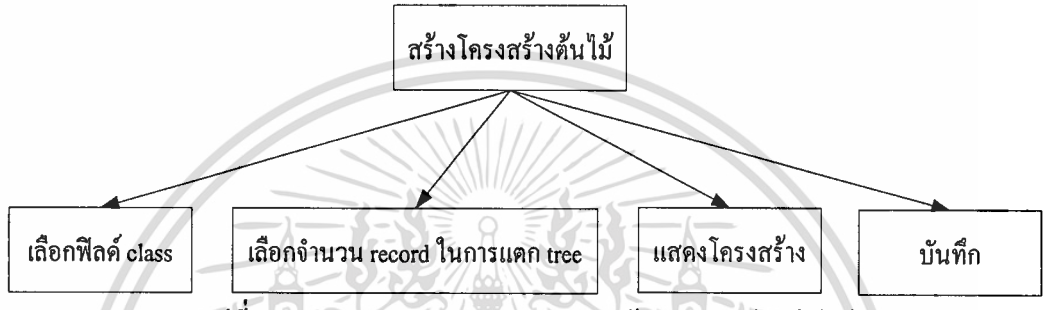
การแก้ไขข้อมูลแบ่งเป็น 2 ส่วนหลักๆ ด้วยกันคือ อัตโนมัติ คือหากผู้ใช้งานเลือกการทำงานแบบอัตโนมัติระบบจะทำการลบข้อมูลที่มีค่าว่างออกไปทั้งหมด และหารผู้ใช้เลือกการแก้ไขข้อมูลแบบทีละฟิลด์ ให้ผู้ใช้งานทำการเลือกฟิลด์ที่ต้องการจะแก้ไข ในกรณีที่ข้อมูลเป็นตัวเลข ให้ผู้ใช้งานทำการเลือกวิธีการแกไขข้อมูลได้ 4 วิธีด้วยกันคือ ลบเรคคอร์ดที่มีค่าว่าง ใส่ค่าสูงสุด ใส่ค่าเฉลี่ย หรือเลือกที่จะใส่ค่าต่ำสุด ส่วนข้อมูลที่มีลักษณะเป็นตัวอักษร ผู้ใช้งานสามารถเลือกวิธีแก้ไขข้อมูลได้ 2 วิธีด้วยกันคือ ลบเรคคอร์ดที่มีค่าว่าง หรือ ใส่ค่า Unknow เข้าไปแทนที่ค่าว่าง



รูปที่ 4.15 Structure Chart แสดงการแปลงข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

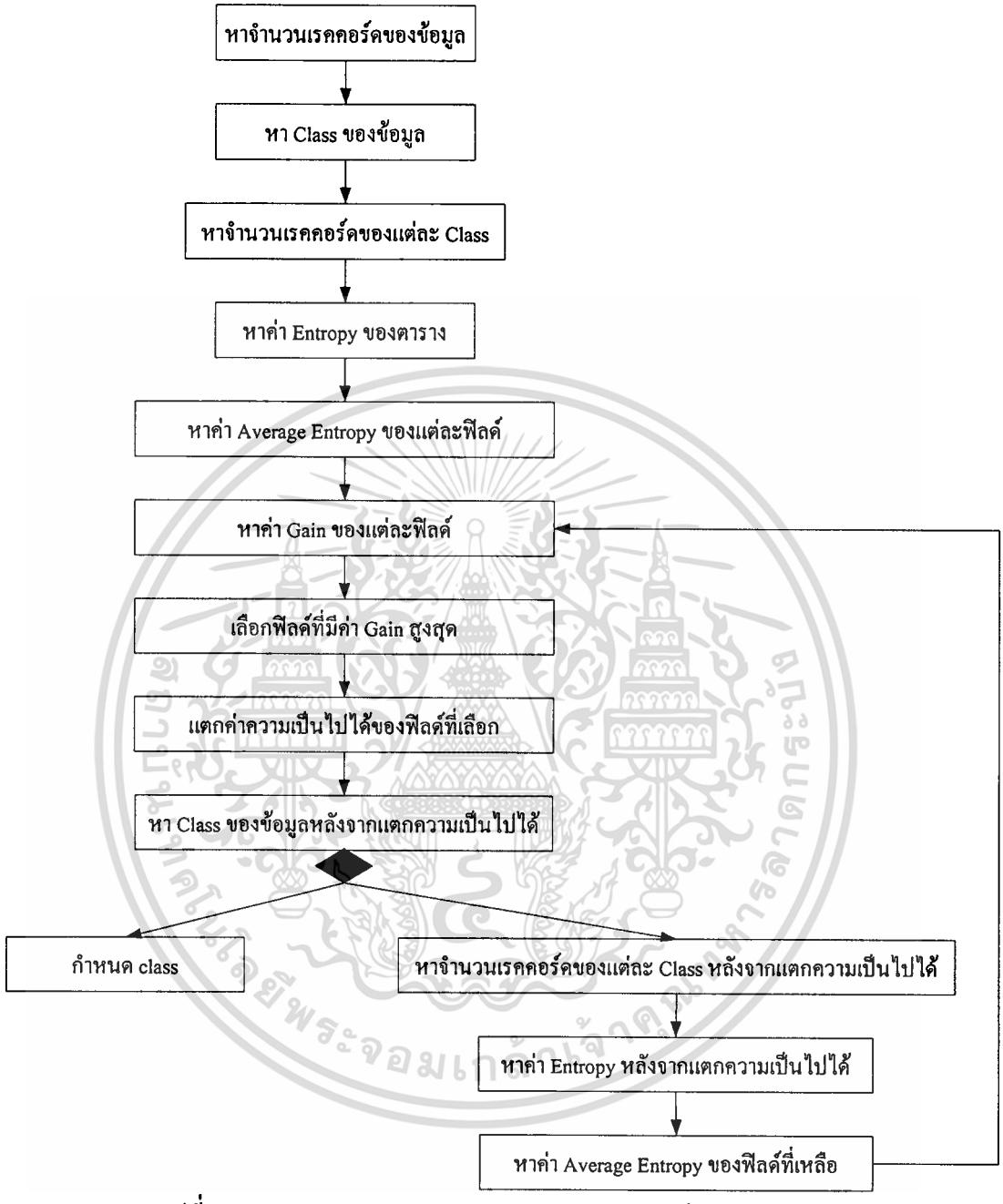
ในการแปลงข้อมูลผู้ใช้งานสามารถเลือกวิธีการแปลงข้อมูลได้ 2 แบบด้วยกันคือ จัดข้อมูลเป็น 2 กลุ่ม หรือทำการจัดข้อมูลเป็น 3 กลุ่ม โดยในการจัดข้อมูลเป็น 2 กลุ่มนั้นให้ผู้ใช้งานทำการเลือกข้อมูลกลุ่มที่ 1 โดยจะแทนค่าข้อมูลด้วยคำว่า Low ส่วนในกลุ่มที่ 2 จะแทนด้วยคำว่า High ส่วนในการจัดกลุ่มข้อมูลเป็น 3 กลุ่มนั้น กลุ่มที่ 1 จะแทนด้วยคำว่า Low ส่วนกลุ่มที่ 2 จะแทนด้วยคำว่า Medium ส่วนกลุ่มที่ 3 จะแทนด้วยคำว่า High



รูปที่ 4.16 Structure Chart แสดงการสร้างโครงสร้างต้นไม้

ในการสร้างโครงสร้างต้นไม้ผู้ใช้งานจะต้องทำการเลือก ฟิลด์ที่ใช้เป็น class สำหรับสร้าง Decision Tree แล้วทำการเลือกจำนวนเรคคอร์ดในการแตกโครงสร้างต้นไม้ คือเป็นการเลือกจำนวนเรคคอร์ดในการแตกกิ่งของต้นไม้เพื่อลดการพรุนนิ่งของข้อมูล ระบบจะทำการแสดงโครงสร้างต้นไม้ให้ผู้ใช้งานโดยผู้ใช้งานสามารถจะทำการบันทึกโปรเจ็กต์ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.17 Flow – Chart แสดงการทำงานของการสร้าง Decision Tree

ในการทำงานในการสร้าง Decision Tree โดยใช้อัลกอริทึม ID3 เริ่มจากการหาจำนวนเรคคอร์ดของข้อมูล หา Class ของข้อมูลว่ามี Class อะไรบ้าง หลังจากนั้นหาจำนวนเรคคอร์ดของข้อมูลแต่ละ Class หาค่า Entropy ของทั้งตาราง หาค่า Average Entropy ของแต่ละฟิลด์ของตาราง หาค่า Gain ของแต่ละฟิลด์ ทำการเลือกฟิลด์ที่มีค่า Gain สูงสุด แตกค่าความเป็นไปได้ของฟิลด์ที่ได้ทำการเลือก หา Class ของข้อมูลหลังจากแตกความเป็นไปได้ หาก Class ของข้อมูลที่ได้ทำการแตกความเป็นไปได้เหลือเพียงแค่ Class เดียวหรือ ฟิลด์ของข้อมูลไม่เหลือสำหรับแตกกิ่งของต้นไม้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แล้วให้ทำการกำหนด Class ของข้อมูลลงไป แต่หากยังมี Class มากกว่า 1 Class และยังมีฟิลด์ข้อมูลเหลืออยู่ให้ทำการหาจำนวนเรคคอร์ดของแต่ละ Class หลังจาที่ได้ทำการแตกความเป็นไปได้ของข้อมูล หาค่า Entropy หาค่า Average Entropy ของฟิลด์ที่เหลือแล้ววนกลับไปทำตามขั้นตอนเดิมจนกว่าจะสร้าง Decision Tree สำเร็จ

4.4 Data dictionary

ตารางที่ 4.1 อธิบายรายละเอียดตาราง number_datatree

Field Name	Description	Type	Key	Null
Time_id	ครั้งที่บันทึกโปรเจ็กต์	Int(4)	PK	No
Namerule	ชื่อโปรเจ็กต์	Varchar(50)		No
Persen	เปอร์เซ็นความผิดพลาด	Varchar(50)		No
detailtree	เก็บรายละเอียดการแปลงข้อมูล	Text(16)		No

ตารางที่ 4.2 อธิบายรายละเอียดตาราง main_datatree

Field Name	Description	Type	Key	Null
Time_number	ครั้งที่บันทึกโปรเจ็กต์	Int(4)	PK	No
Name_field	ค่าแม่ของ array	Varchar(50)		No
Key_reference	Key ไว้ติดต่อกับ number_datatree	Varchar(50)	PK	No
Index	เก็บลำดับ	Int(4)		No

ตารางที่ 4.3 อธิบายรายละเอียดตาราง exten_datatree

Field Name	Description	Type	Key	Null
Time_number_ext	ครั้งที่บันทึกโปรเจ็กต์	Int(4)		No
Key_reference_ext	Key ไว้ติดต่อกับ main_datatree	Varchar(50)		No
Field_ext	ค่าของ array ลูก	Varchar(50)		No
Belive_ext	ค่าความน่าเชื่อถือ	Varchar(50)		No
indexx	เก็บลำดับ	Int(4)		No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 อธิบายรายละเอียดตาราง refer_datatree

Field Name	Description	Type	Key	Null
Time_number_ref	ครั้งที่บันทึกโปรเจ็กต์	Int(4)		No
Name_field_ref	ชื่อฟิลด์ที่ใช้ในโปรเจ็กต์	Varchar(50)		No
Value_ref	ค่าความเป็นไปได้ของฟิลด์	Varchar(50)		No

ตารางที่ 4.5 อธิบายรายละเอียดตาราง field_datatree

Field Name	Description	Type	Key	Null
Time_number_field	ครั้งที่บันทึกโปรเจ็กต์	Int(4)		No
Name_possible	ชื่อฟิลด์	Varchar(50)		No
Insert_value	ค่าที่ใช้นำหน้าค่าในฟิลด์	Varchar(50)		No
Num_division	จำนวนการแบ่งกลุ่ม	Int(4)		No
Val1	ค่ากลุ่มที่ 1	Float (8)		No
Val2	ค่ากลุ่มที่ 2	Float(8)		No

ตารางที่ 4.6 อธิบายรายละเอียดตาราง class_datatree

Field Name	Description	Type	Key	Null
Time_number_class	ครั้งที่บันทึกโปรเจ็กต์	Int(4)		No
Name_class	ค่าของ class แต่ละโปรเจ็กต์	Varchar(50)		No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
การประยุกต์ใช้โปรแกรม

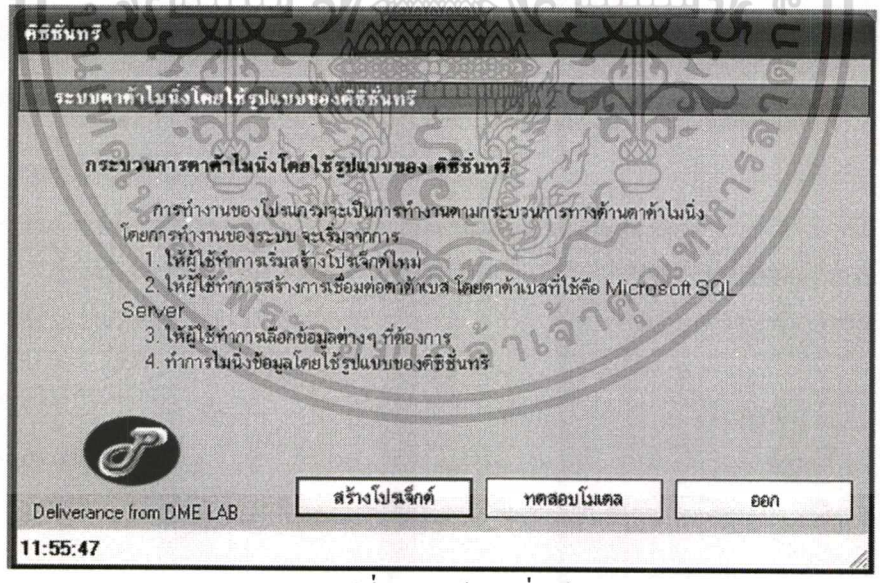
5.1 การสร้างโปรเจ็กต์ใหม่

5.1.1 เมื่อทำการเปิดโปรแกรมดิซิชั่นทรี จะมีหน้าต่างขึ้นมาดังรูปที่ 5.1 โดยในหน้าต่างนี้จะมีการทำงานอยู่ด้วยกัน 3 ปุ่มด้วยกัน คือ

ปุ่ม สร้างโปรเจ็กต์ คือ จะเป็นการสร้างโปรเจ็กต์ใหม่ขึ้นมา หากต้องการที่จะทำการสร้างโปรเจ็กต์ใหม่ขึ้นมาเพื่อทำการสร้างโครงสร้างต้นไม้ให้ทำการกดปุ่ม สร้างโปรเจ็กต์

ปุ่ม ทดสอบโมเดล คือ หากต้องการที่จะทดสอบโมเดลที่ได้จากการสร้างโปรเจ็กต์ให้กดปุ่ม ทดสอบโมเดล

ปุ่ม ออก คือ ต้องการที่จะออกจากโปรแกรม

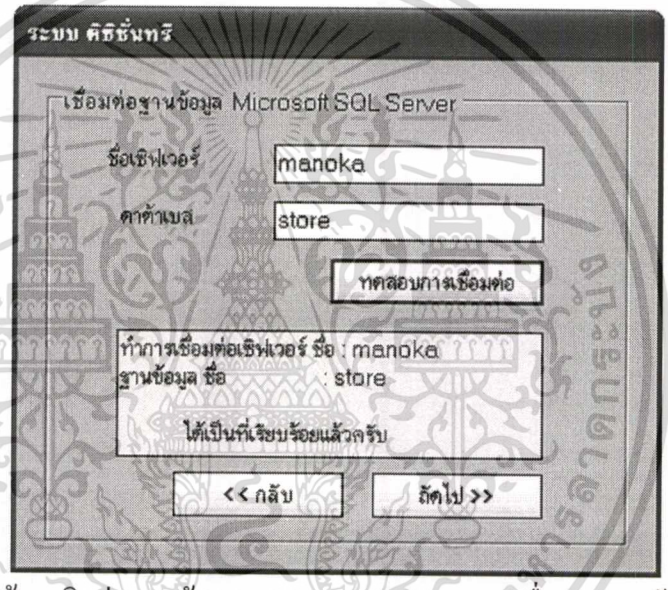


รูปที่ 5.1 หน้าจอเริ่มต้นของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.2 หลังจากกดปุ่ม สร้างโปรเจ็กต์ จะมีหน้าต่างขึ้นดังรูปที่ 5.2 เพื่อทำการติดต่อกับฐานข้อมูลที่ต้องการ โดยฐานข้อมูลที่จะทำการติดต่อด้วยคือ Microsoft SQL Server โดยผู้ใช้ต้องทำการกรอกข้อมูลคือ ชื่อเซิฟเวอร์ และ ชื่อดาต้าเบส หลังจากนั้นผู้ใช้สามารถกดปุ่ม ทดสอบการเชื่อมต่อหรือไม่ก็ได้ แล้วทำการกดปุ่ม ถัดไป >>

โดยหากต้องการที่ทดสอบการติดต่อฐานข้อมูลให้ กดปุ่ม ทดสอบการเชื่อมต่อ หรือไม่ผู้ใช้ก็สามารถที่กดปุ่ม ถัดไป >> โดยระบบจะทำการตรวจสอบการเชื่อมต่อฐานข้อมูลให้อย่างอัตโนมัติ และหากท่านต้องการย้อนกลับไปหน้าก่อนหน้าให้กดปุ่ม << กลับ



รูปที่ 5.2 หน้าจอติดต่อฐานข้อมูล Microsoft SQL Server เพื่อทำการสร้างโปรเจ็กต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.3 หน้าต่างสำหรับเลือกตารางพื้นฐานสำหรับใช้ในการสร้างโครงสร้างต้นไม้ โดยให้ทางการเลือกชื่อตารางข้อมูลจาก Listbox ด้านซ้าย ไปยังด้านขวา โดย Listbox ด้านขวาคือส่วนที่เราต้องการใช้งาน หลังจากนั้นให้กดปุ่ม ถัดไป >> เพื่อทำงานต่อ โดยมีปุ่มการทำงานหลัก คือ

ปุ่ม > ทำการย้ายข้อมูลจาก Listbox ด้านซ้ายไปยังด้านขวาทีละตัวตามแล้วแต่ที่ผู้ใช้งานได้ทำการเลือกไว้

ปุ่ม < ทำการย้ายข้อมูลจาก Listbox ด้านขวาไปยังด้านซ้ายทีละตัวตามแล้วแต่ที่ผู้ใช้งานได้ทำการเลือกไว้

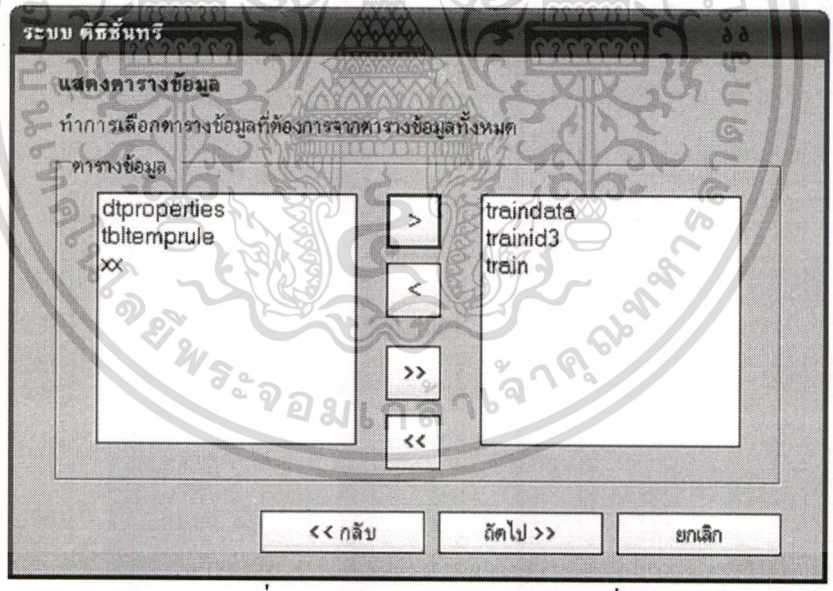
ปุ่ม >> ทำการย้ายข้อมูลจาก Listbox ด้านซ้ายไปยังด้านขวาทั้งหมด

ปุ่ม << ทำการย้ายข้อมูลจาก Listbox ด้านขวาไปยังด้านซ้ายทั้งหมด

ปุ่ม << กลับ คือจะกลับไปทำงานก่อนหน้าใหม่

ปุ่ม ถัดไป >> คือจะทำงานต่อไปในขั้นตอนถัดไป

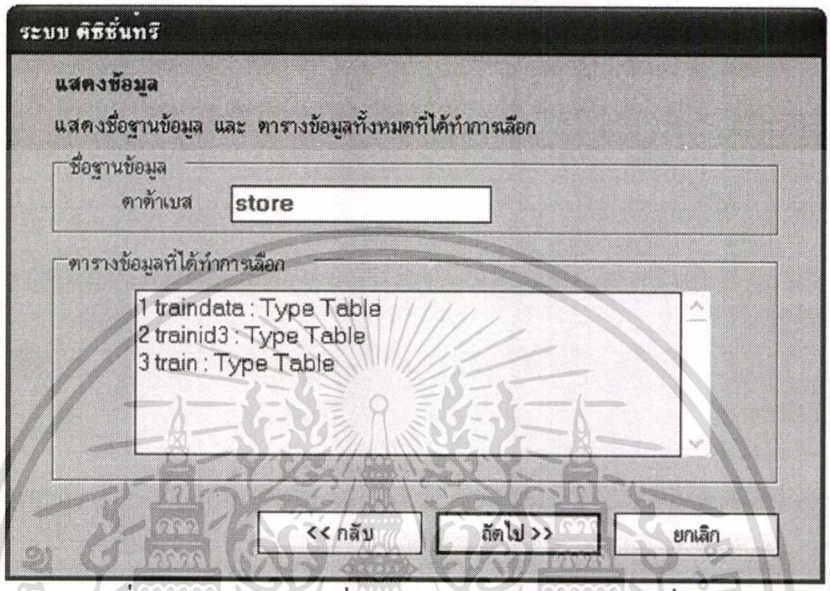
ปุ่ม ยกเลิก หากต้องการยกเลิกการทำงานแล้วกลับไปหน้าหลัก



รูปที่ 5.3 หน้าจอเลือกตารางข้อมูลที่ต้องการ

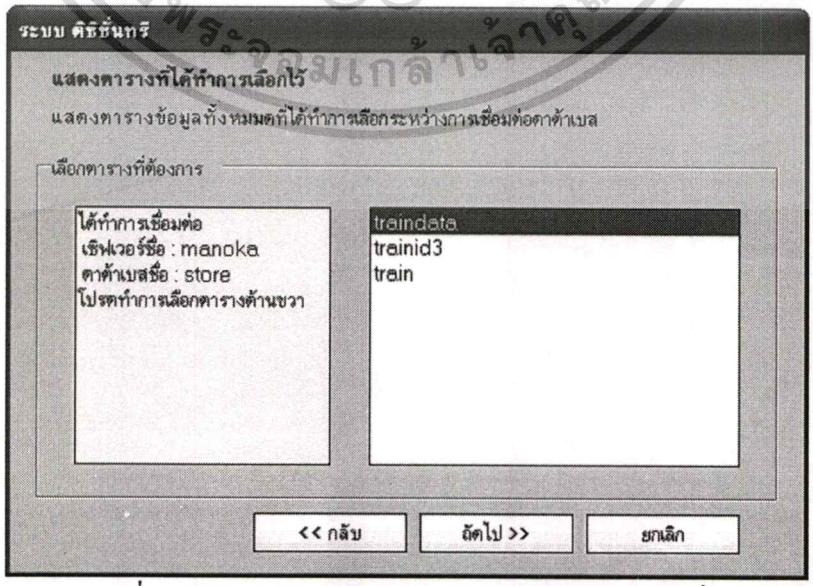
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.4 จะเป็นหน้าต่างแสดงชื่อดาต้าเบส และ ชื่อตารางต่างๆ ที่ผู้ใช้ได้ทำการเลือกไว้ก่อนหน้า ให้ผู้ใช้ทำการตรวจสอบอีกครั้ง หลังจากนั้นให้กดปุ่ม ถัดไป >> เพื่อทำงานขั้นตอนถัดไป



รูปที่ 5.4 หน้าจอแสดงชื่อดาต้าเบสและตารางข้อมูลที่ได้ทำการเลือก

5.1.5 รูปที่ 5.5 เป็นหน้าต่างสำหรับให้ผู้ใช้ทำการเลือกตารางที่ต้องการสำหรับทำการไมน์นิ่งข้อมูล คือเลือกตารางสำหรับที่จะนำไปสร้างเป็นโครงสร้างต้นไม้ โดยผู้ใช้ต้องทำการเลือกตารางที่ต้องการหลังจากนั้นให้กดปุ่ม ถัดไป >> หรือไม่ต้องการจะทำงานต่อให้กดปุ่ม ยกเลิก



รูปที่ 5.5 หน้าจอเลือกตารางข้อมูลสำหรับการทำไมน์นิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.6 ให้ผู้ใช้งานทำการเลือกฟิลด์ข้อมูลจากตารางที่ได้ทำการเลือกไว้ก่อนหน้าโดยให้ทำการเลือกฟิลด์ข้อมูลสำหรับทำการไมน์นิ่ง อย่างน้อย 2 ฟิลด์ขึ้นไป หลังจากนั้นให้กดปุ่ม ถัดไป >> โดยในหน้าต่างนี้จะมี ปุ่มการทำงานหลักๆ คือ

ปุ่ม > ทำการย้ายข้อมูลจาก Listbox ด้านซ้ายไปยังด้านขวาทีละตัวตามแล้วแต่ที่ผู้ใช้งานได้ทำการเลือกไว้

ปุ่ม < ทำการย้ายข้อมูลจาก Listbox ด้านขวาไปยังด้านซ้ายทีละตัวตามแล้วแต่ที่ผู้ใช้งานได้ทำการเลือกไว้

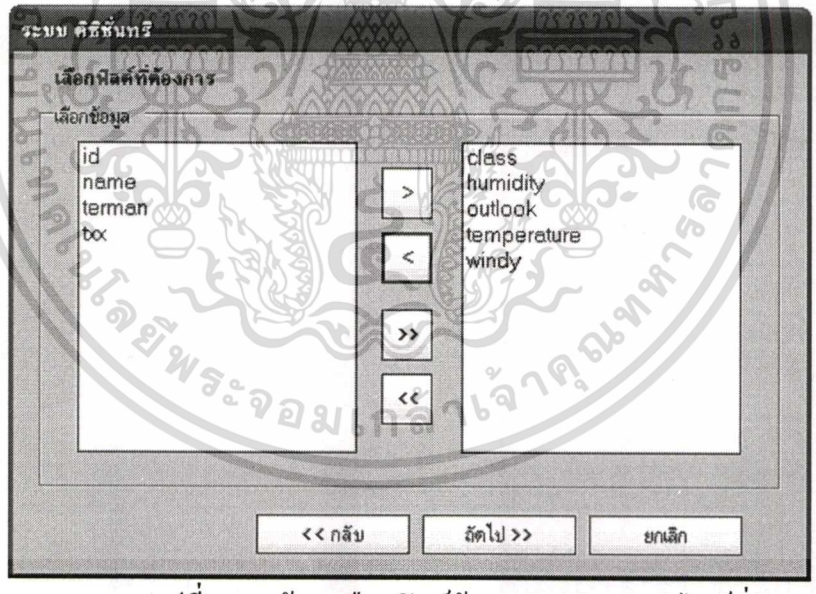
ปุ่ม >> ทำการย้ายข้อมูลจาก Listbox ด้านซ้ายไปยังด้านขวาทั้งหมด

ปุ่ม << ทำการย้ายข้อมูลจาก Listbox ด้านขวาไปยังด้านซ้ายทั้งหมด

ปุ่ม << กลับ คือจะกลับไปทำงานก่อนหน้าใหม่

ปุ่ม ถัดไป >> คือจะทำงานต่อไปในขั้นตอนถัดไป

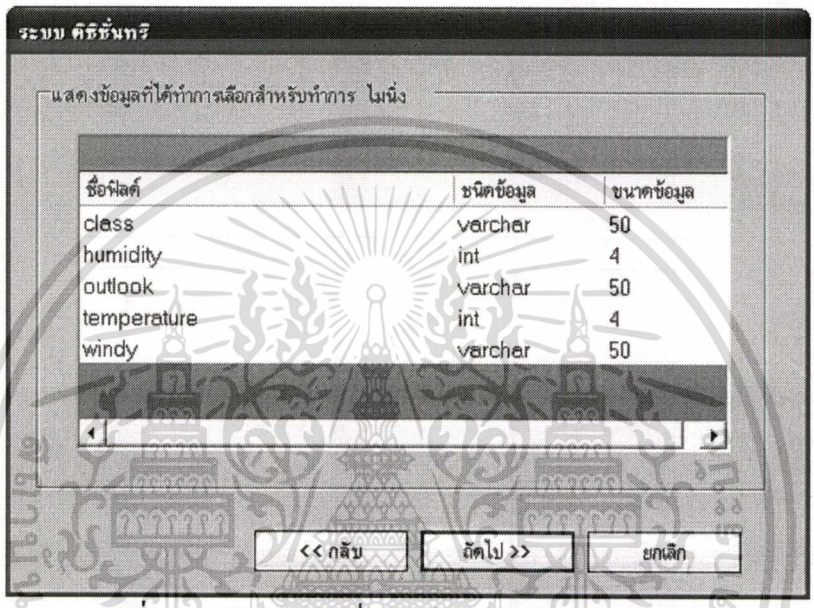
ปุ่ม ยกเลิก หากต้องการยกเลิกการทำงานแล้วกลับไปหน้าหลัก



รูปที่ 5.6 หน้าจอเลือกฟิลด์ข้อมูลสำหรับการทำไมน์นิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.7 รูปที่ 5.7 จะทำการแสดงข้อมูลชื่อฟิลด์ ชนิดของข้อมูล และขนาดของข้อมูล ในแต่ละฟิลด์ที่ผู้ใช้ได้ทำการเลือกไว้ก่อนหน้า หลังจากผู้ใช้ทำการตรวจสอบข้อมูลเป็นที่เรียบร้อยแล้ว ให้ผู้ใช้ทำการกดปุ่ม ถัดไป >> เพื่อทำงานขั้นต่อไป หรือหากไม่ต้องการทำงานต่อผู้ใช้ก็สามารถที่จะกดปุ่ม ยกเลิก เพื่อกลับไปหน้าเมนูหลักของโปรแกรม

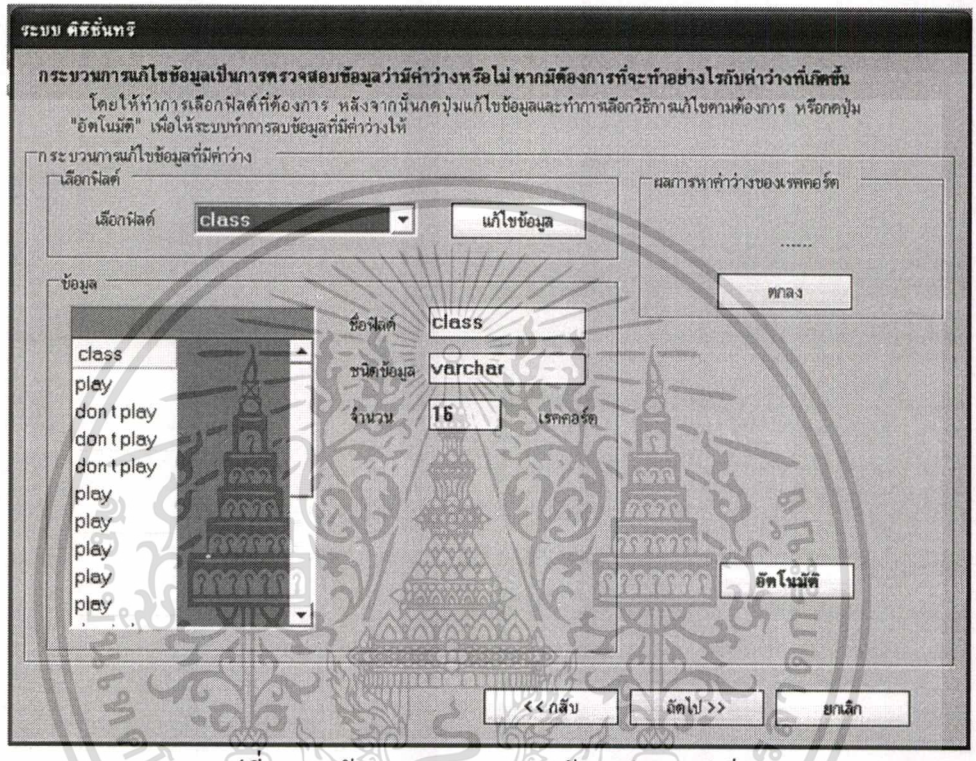


รูปที่ 5.7 หน้าจอแสดงชื่อฟิลด์ ชนิด และขนาดของฟิลด์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.8 รูปที่ 5.8 แสดงหน้าของการตรวจสอบข้อมูลที่มีค่าว่างหรือไม่ โดยการทำงานคือ

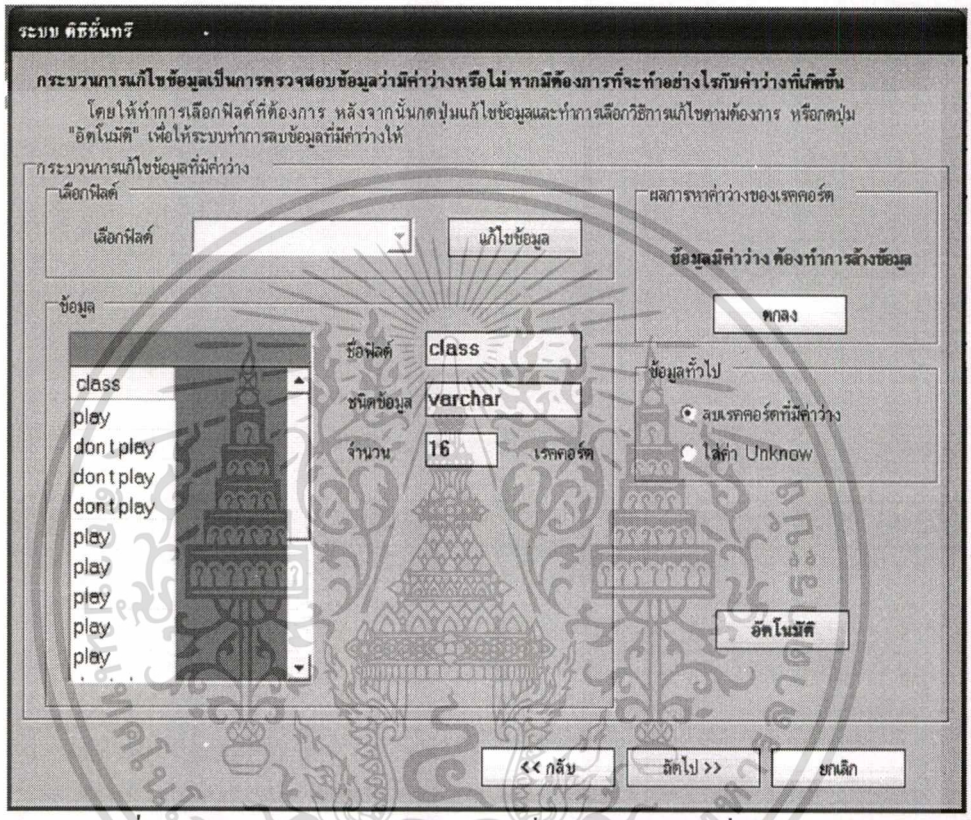
ให้ผู้ใช้ทำการเลือกฟิลด์ที่ต้องการตรวจสอบ โดยตัวโปรแกรมจะแสดงข้อมูลต่างๆ ดังรูป ให้ผู้ใช้กดปุ่ม แก้ไขข้อมูลเพื่อทำการแก้ไขข้อมูล



รูปที่ 5.8 หน้าจอการตรวจสอบข้อมูลในกรณีเป็นตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

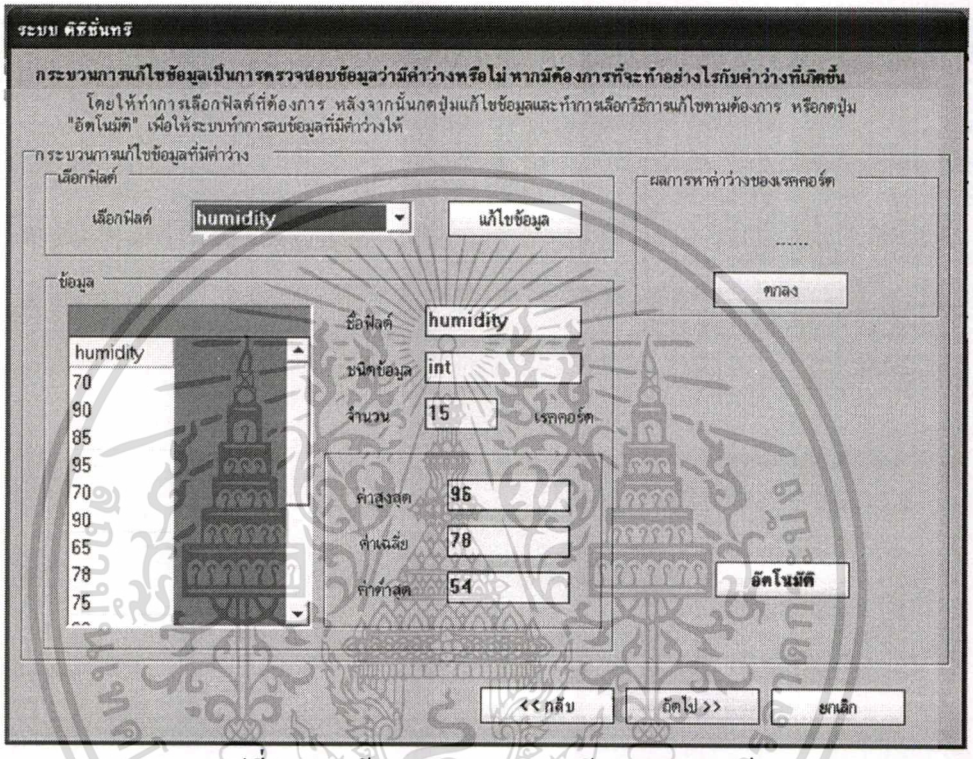
5.1.9 เมื่อผู้ใช้ทำการกดปุ่ม แก้ไขข้อมูล หากข้อมูลมีค่าว่างเกิดขึ้น ด้านขวามือ โปรแกรมก็จะแสดงวิธีการแก้ไขข้อมูลที่ผู้ใช้สามารถทำได้ โดยให้ผู้ใช้ทำการแก้ไขข้อมูลโดยเลือกวิธีการที่ต้องการแล้วกดปุ่ม ตกลง เพื่อทำการแก้ไขข้อมูล



รูปที่ 5.9 หน้าจอแสดงการแก้ไขข้อมูลที่มีค่าว่างในกรณีที่เป็นตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

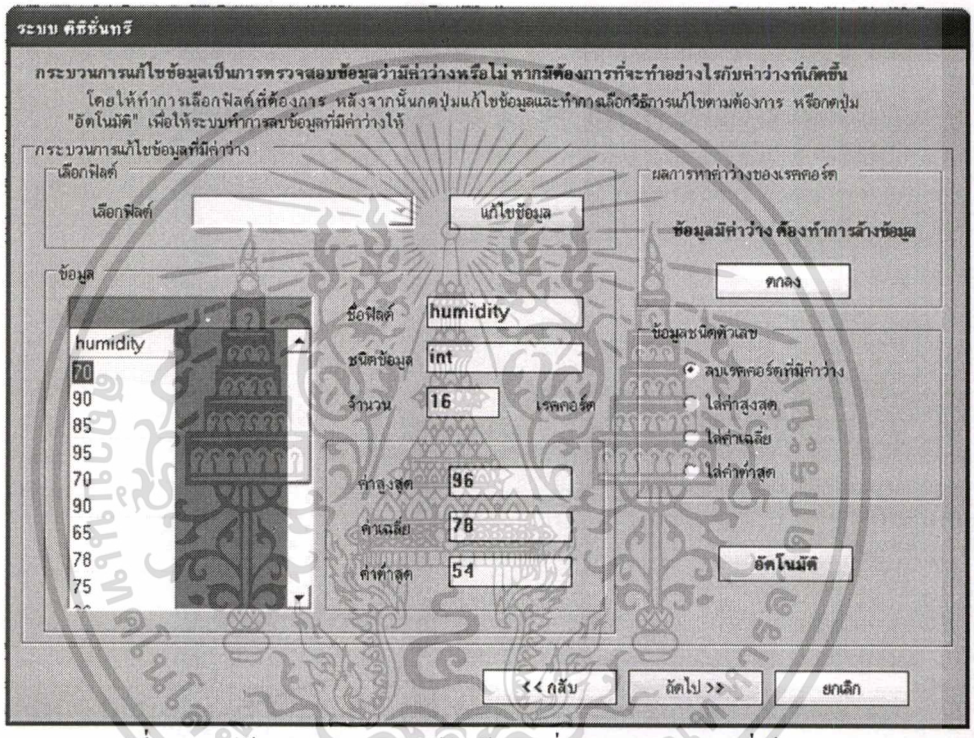
5.1.10 ในกรณีที่ข้อมูลเป็นตัวเลขเมื่อผู้ใช้ทำการเลือกฟิลด์ที่ต้องการจะตรวจสอบ โปรแกรมก็จะแสดงข้อมูล ค่าสูงสุด ค่าต่ำสุด และค่าเฉลี่ย เพิ่มเติมขึ้นมาจากข้อมูลที่เป็นตัวอักษร หลังจากนั้นให้ผู้ใช้กดปุ่ม แก้ไขข้อมูล



รูปที่ 5.10 หน้าจอการตรวจสอบข้อมูลในกรณีเป็นตัวเลข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

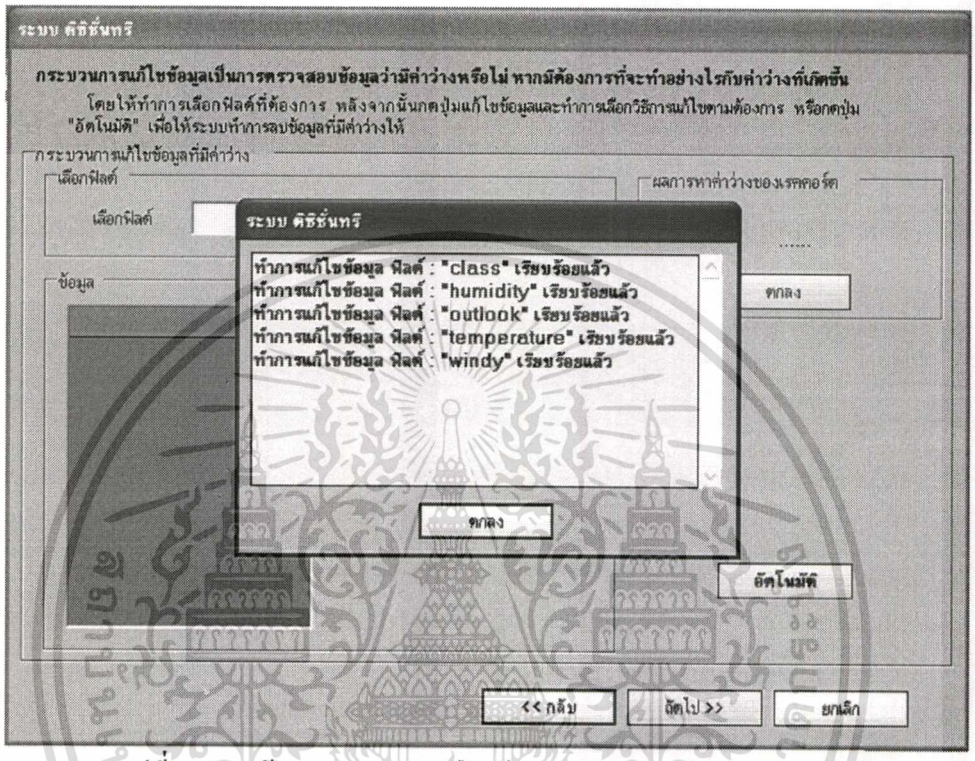
5.1.11 ในการแก้ไขข้อมูลในกรณีที่เป็นตัวเลข ด้านขวามือจะแสดงวิธีการในการแก้ไขข้อมูลโดยผู้ใช้สามารถทำการแก้ไขข้อมูลชนิดที่เป็นตัวเลขได้ 4 วิธีด้วยกันคือ ลบเรคคอร์ดที่มีค่าว่าง ใส่ค่าสูงสุด ใส่ค่าเฉลี่ย และใส่ค่าต่ำสุด ให้ผู้ใช้งานทำการเลือกวิธีการแก้ไขข้อมูล และกดปุ่มตกลง เมื่อผู้ให้งานทำการแก้ไขข้อมูลครบทุกฟิลด์แล้วให้กดปุ่ม ถัดไป >> เพื่อทำงานขั้นตอนต่อไป



รูปที่ 5.11 หน้าจอแสดงการแก้ไขข้อมูลที่มีค่าว่างในกรณีที่เป็นตัวเลข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

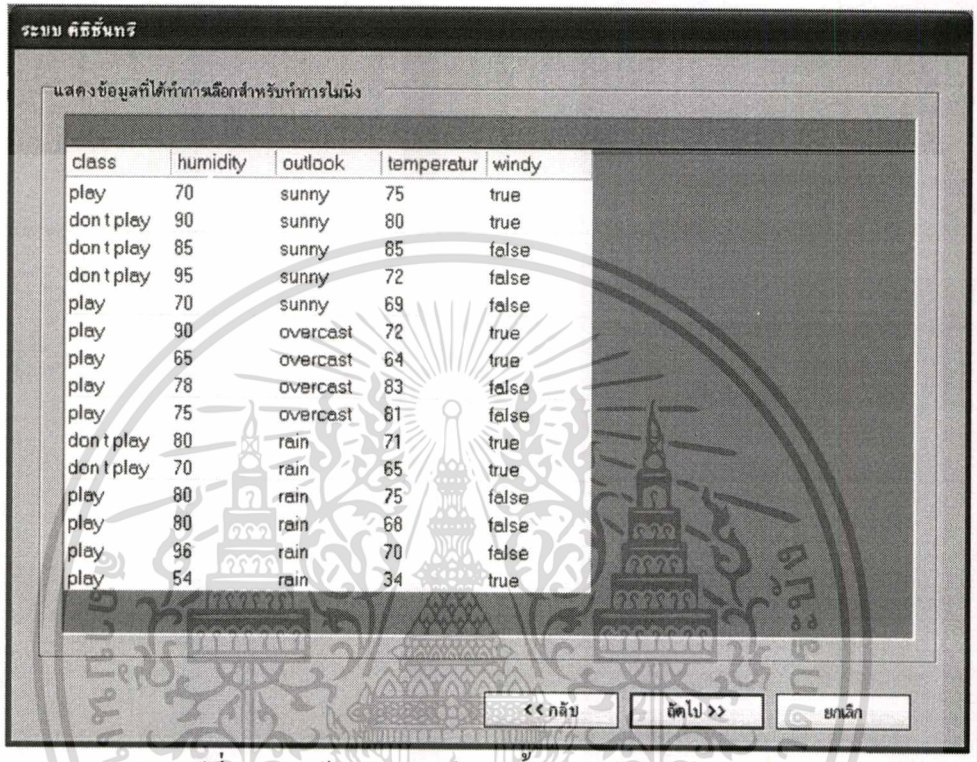
5.1.12 ในกรณีของการแก้ไขข้อมูลผู้ใช้สามารถกดปุ่ม อัตโนมัติ เพื่อทำการลบเรคคอร์ดที่มีค่าว่างทั้งหมดของตารางได้เลย



รูปที่ 5.12 หน้าจอแสดงการแก้ไขข้อมูลในกรณีกดปุ่ม “อัตโนมัติ”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

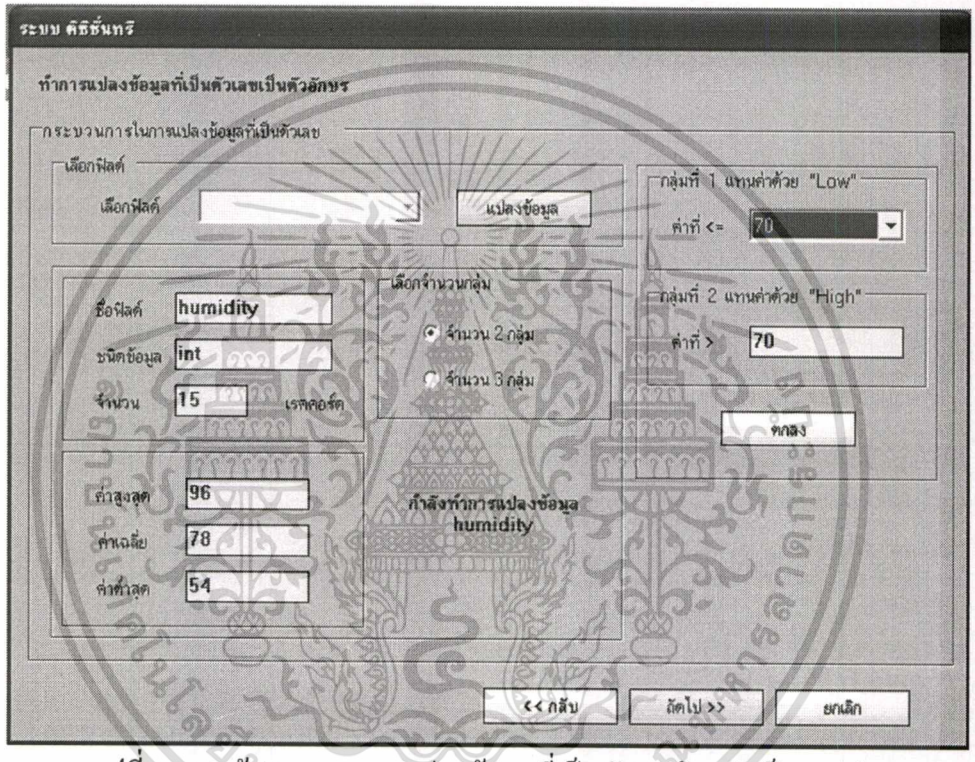
5.1.13 โปรแกรมจะแสดงข้อมูลทั้งหมดที่ได้ทำการแก้ไขแล้ว ต่อมาให้ผู้ใช้งานทำการกดปุ่ม ถัดไป >> เพื่อทำงานขั้นตอนต่อไป



รูปที่ 5.13 หน้าจอแสดงข้อมูลทั้งหมดหลังจากการแก้ไข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

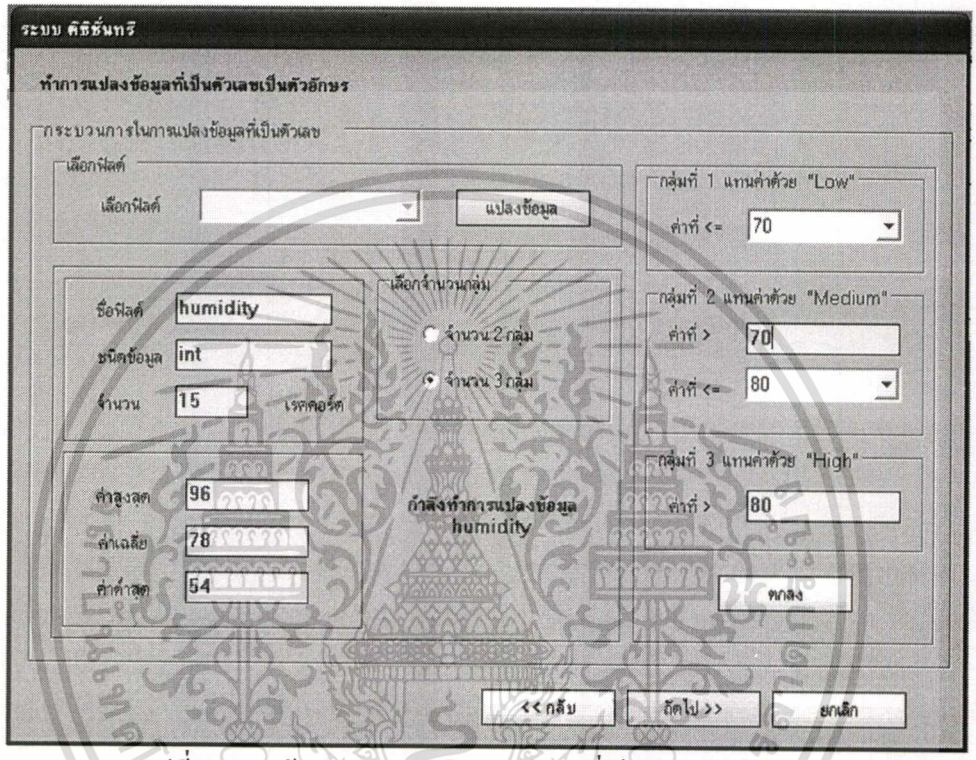
5.1.14 หากฟิลด์ของข้อมูลที่ได้การเลือกไว้มี ชนิดเป็นตัวเลขโปรแกรมจะให้ผู้ใช้งานทำการแปลงข้อมูลที่เป็นตัวเลขนั้นเป็นตัวอักษร โดยให้ผู้ใช้งานทำการเลือกฟิลด์ที่ต้องการ แล้วกดปุ่มแปลงข้อมูล โปรแกรมก็จะแสดงส่วนของการเลือกจำนวนกลุ่ม โดยผู้ใช้สามารถที่จะเลือกจำนวนกลุ่มได้ 2 แบบด้วยกัน คือ แบบ 2 กลุ่ม และแบบ 3 กลุ่ม โดยในรูป 5.14 แสดงการแปลงข้อมูลแบบ 2 กลุ่ม ต่อมาให้ผู้ใช้งานทำการเลือกข้อมูลแต่ละกลุ่ม แล้วกดปุ่ม ตกลง



รูปที่ 5.14 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลข โดยจัดเป็น 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

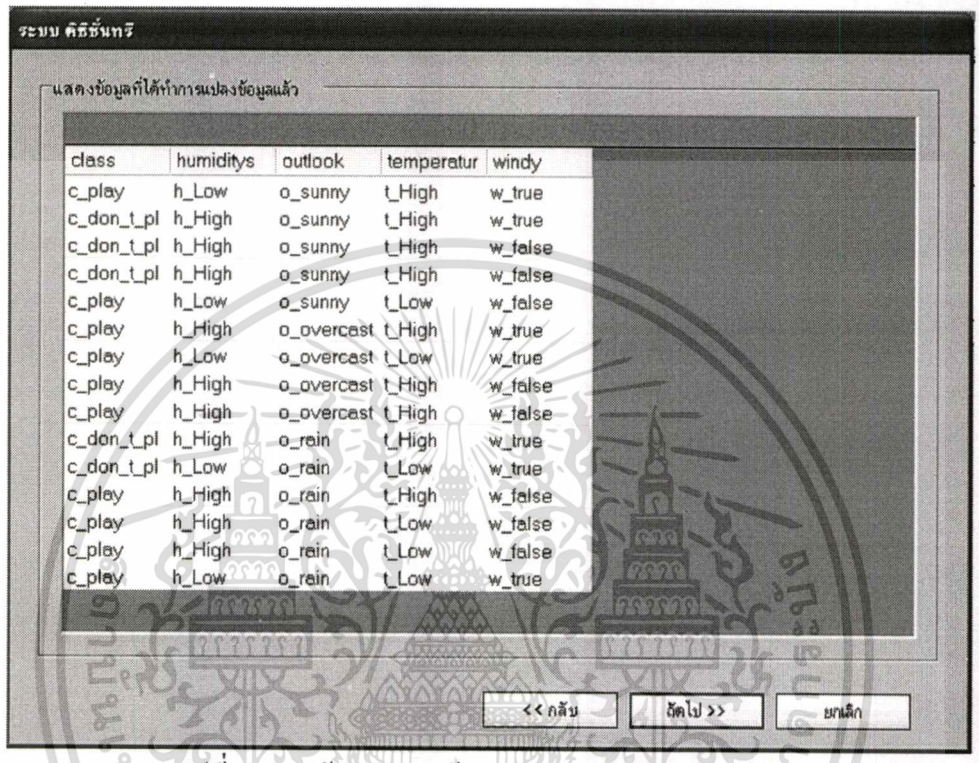
5.1.15 รูปที่ 5.15 แสดงลักษณะการแปลงข้อมูลในกรณีที่ผู้ใช้ต่อการแปลงเป็น 3 กลุ่ม โดยให้ผู้ใช้งานทำการเลือกข้อมูลแต่ละกลุ่มแล้วกดปุ่ม ตกลง หลังจากนั้นเมื่อผู้ใช้งานทำการแปลงข้อมูลที่มีชนิดเป็นตัวเลขครบหมดทุกฟิลด์แล้วให้กดปุ่ม ถัดไป >> เพื่อทำงานขั้นตอนต่อไป



รูปที่ 5.15 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลขโดยจัดเป็น 3 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

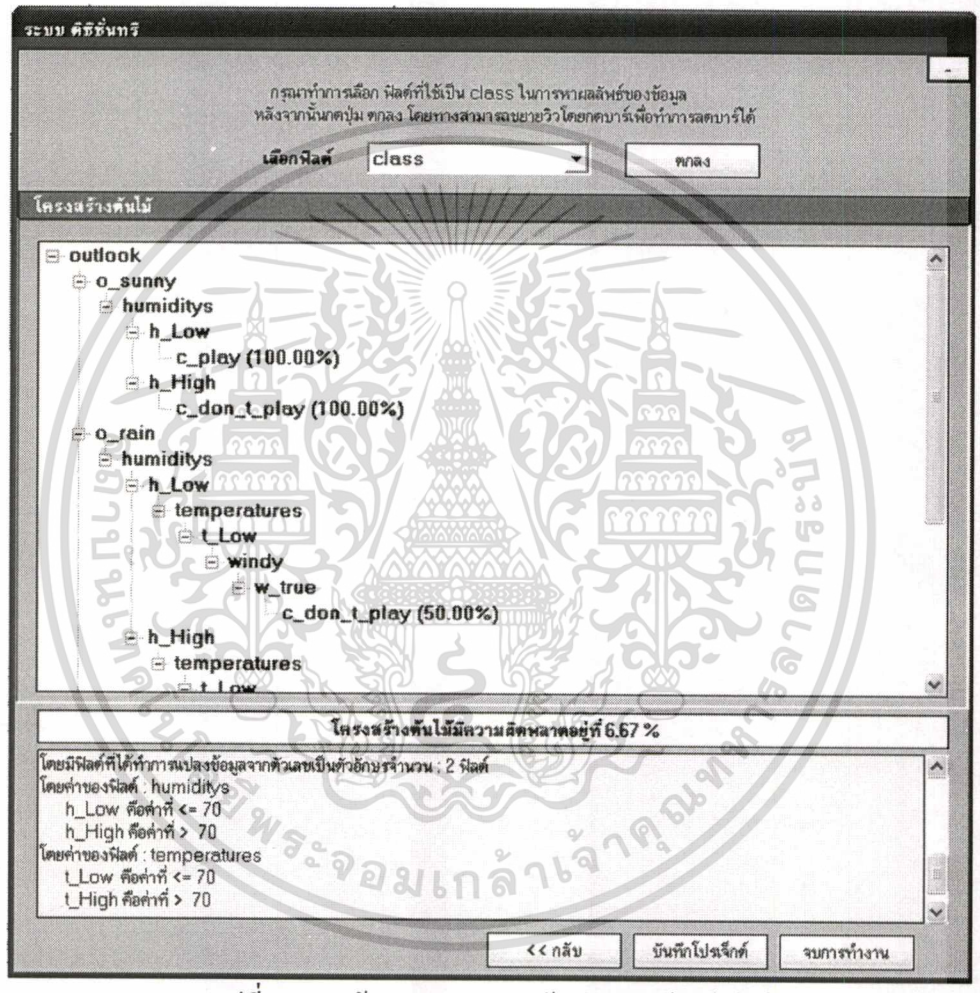
5.1.16 รูปที่ 5.16 แสดงข้อมูลที่ได้ทำการแปลงและแก้ไขเรียบร้อยแล้วให้ผู้ใช้งานกดปุ่ม ถัดไป >> เพื่อทำงานขั้นตอนต่อไป



รูปที่ 5.16 หน้าจอแสดงข้อมูลหลังจากมีการแปลงข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

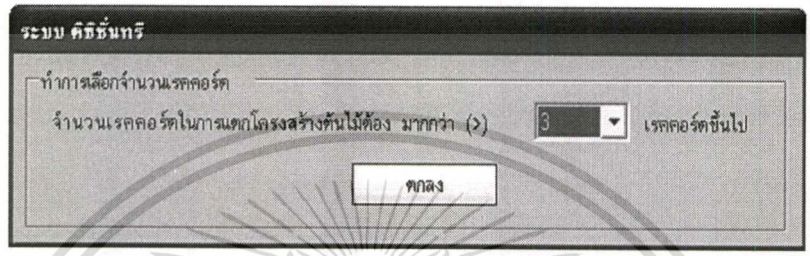
5.1.17 จะเป็นการสร้างโครงสร้างต้น โดยให้ผู้ใช้งานทำการเลือก ชื่อฟิลด์ที่ใช้สำหรับเป็น Class ที่ต้องการหาค่า ตรง เลือกฟิลด์จากหน้าต่าง โดยจากรูปจะเป็นการเลือกฟิลด์ชื่อ Class เป็นตัว Class ในการหาโครงสร้างของต้นไม้ หลังจากนั้นให้กดปุ่ม ตกลง โปรแกรมก็จะแสดงโครงสร้างต้นไม้ให้กับผู้ใช้งาน โดยจะบอกถึงความผิดพลาดของโครงสร้างต้นไม้และข้อมูลต่างๆ



รูปที่ 5.17 หน้าจอแสดงการสร้างโครงสร้างต้นไม้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

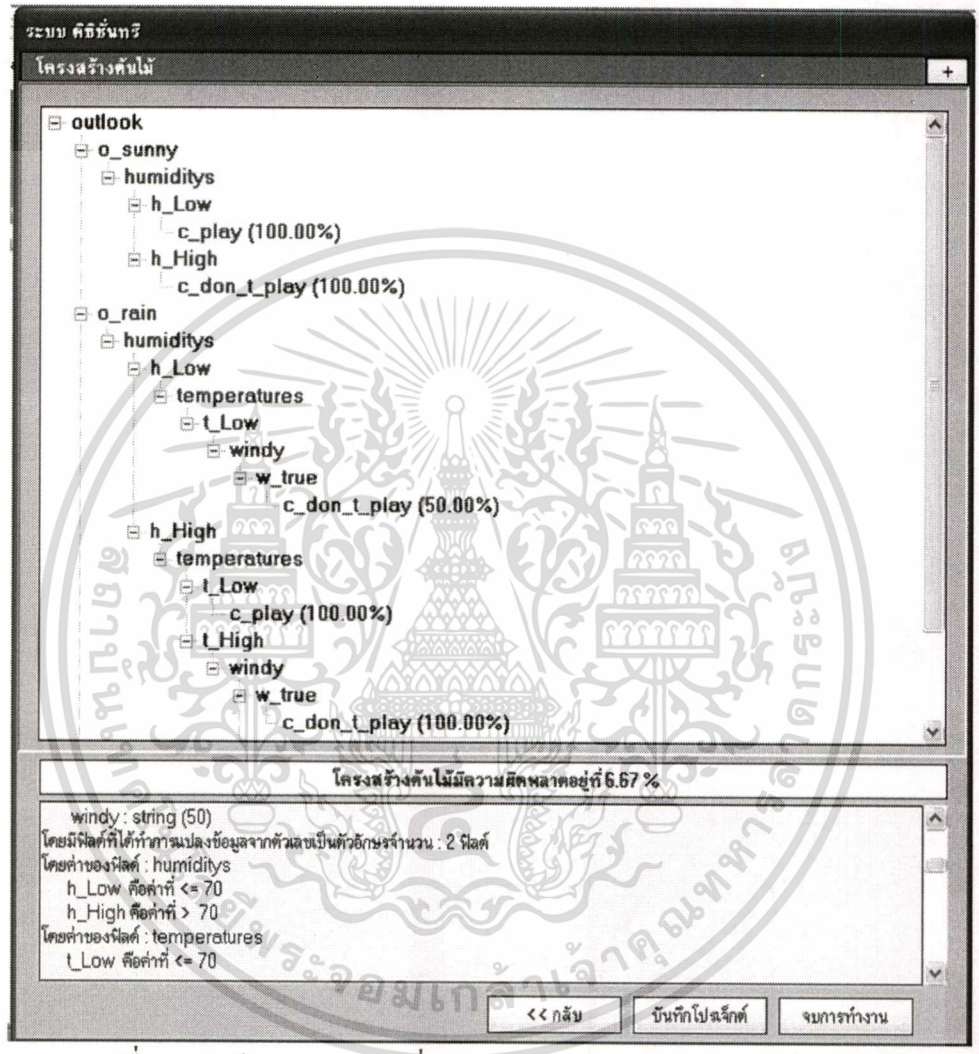
5.1.18 หลังจากนั้นจะแสดงหน้าต่างให้ผู้ใช้ทำการเลือกจำนวนเรคคอร์ดในการแตกโครงสร้างของต้นไม้ เช่น ถ้าผู้ใช้เลือก 3 หมายถึงว่าต้องมีจำนวนเรคคอร์ดมากกว่า 3 เรคคอร์ดถึงจะทำการแตกต้นไม้ในระดับต่อไปอีกถ้าไม่ถึงก็จะไม่ทำการแตกอีก หลังจากนั้นก็จะทำการแสดงโครงสร้างต้นไม้ที่เกิดขึ้น



รูปที่ 5.18 หน้าจอสำหรับทำการเลือกจำนวนเรคคอร์ดในการแตกโครงสร้างต้นไม้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

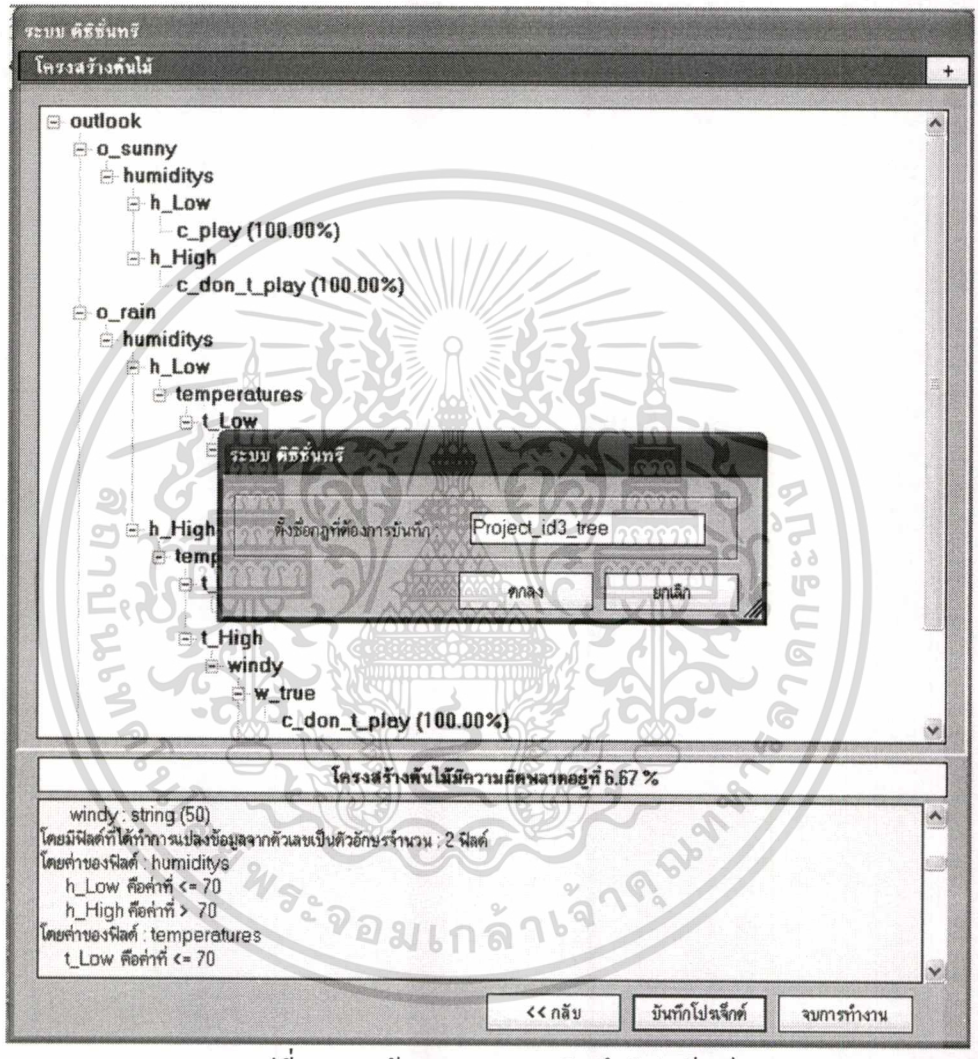
5.1.19 ผู้ใช้งานสามารถที่จะขยายหน้าต่างของการแสดงโครงต้นไม้ได้โดยกดปุ่ม – หรื อ+ ทางมุ่งบนขวามือ ตามรูป



รูปที่ 5.19 หน้าจอแสดงการเพิ่มขนาดหน้าจอการแสดงโครงสร้างต้นไม้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.1.20 หากผู้ใช้งานต้องการบันทึกรูปแบบโมเดลที่ได้ให้กดปุ่ม บันทึกโปรเจ็กต์ หลังกจากนั้นก็ทำการใส่ชื่อ โปรเจ็กต์แล้วกด ตกลง เพื่อทำการบันทึกข้อมูล หรือหากไม่ต้องการบันทึกข้อมูลให้กดปุ่ม จบการทำงาน

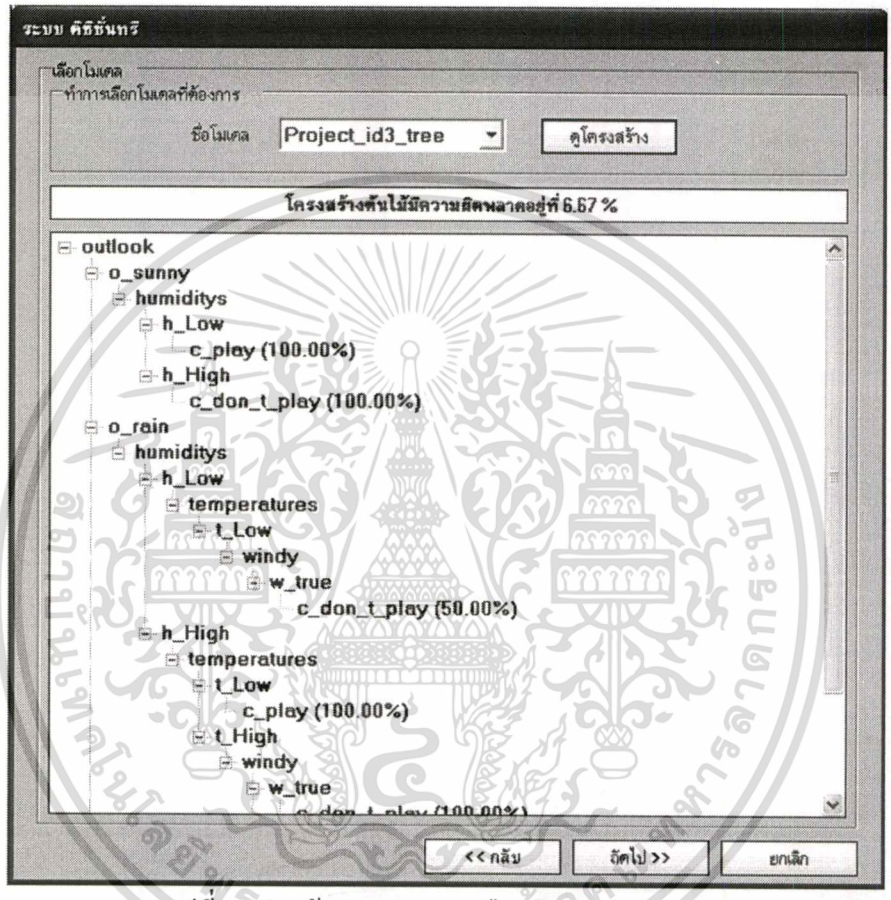


รูปที่ 5.20 หน้าจอแสดงการบันทึกโปรเจ็กต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 การทดสอบโมเดล

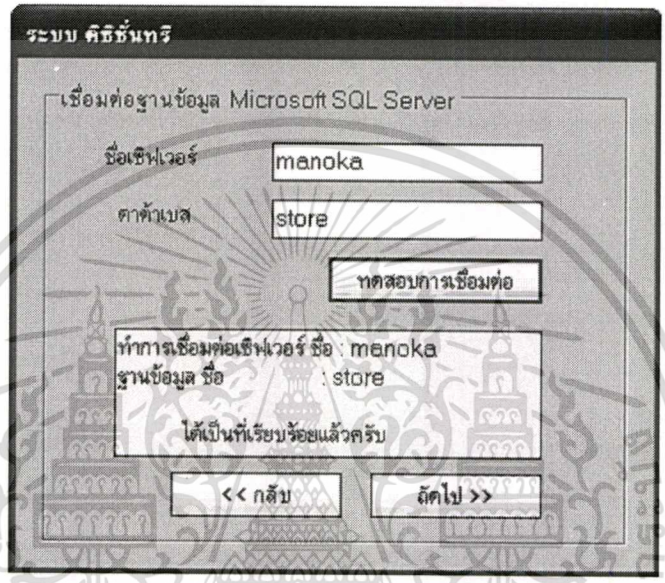
5.2.1 ให้ผู้ใช้งานทำการเลือกชื่อโมเดลที่ต้องการจะทดสอบแล้ว กดปุ่ม ดูโครงสร้าง เมื่อทำการเลือกโมเดลที่ต้องการได้แล้วให้กดปุ่ม ถัดไป >> เพื่อทำงานขั้นตอนต่อไป



รูปที่ 5.21 หน้าจอแสดงการเลือกโมเดลมาทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

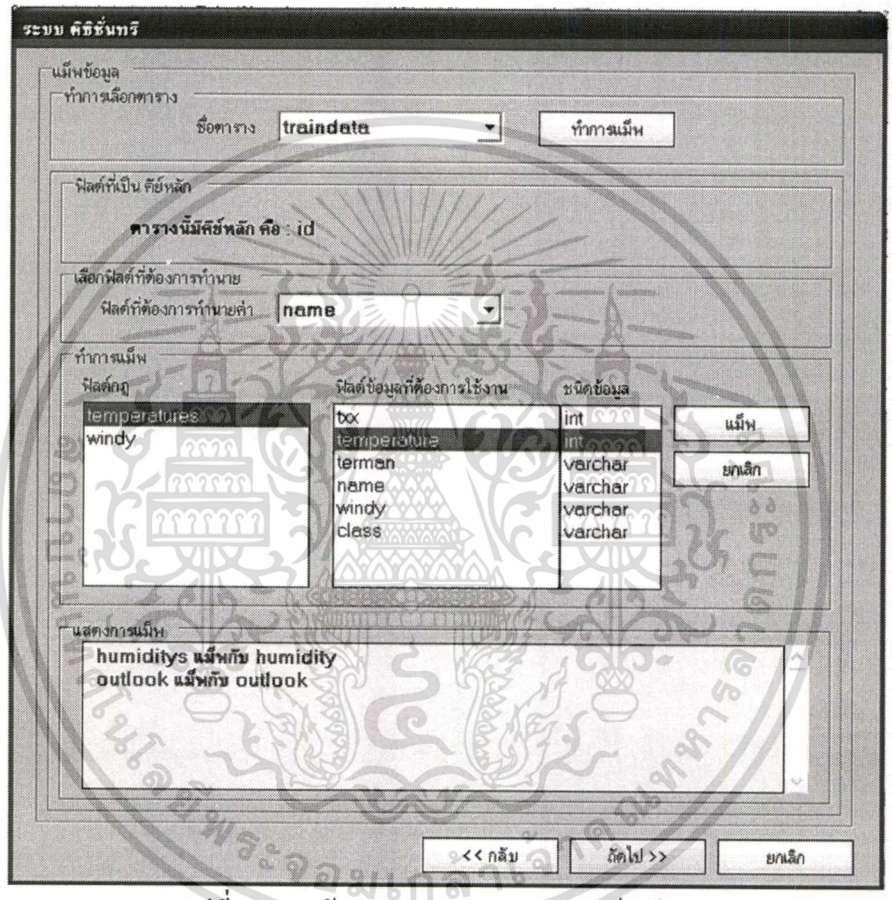
5.2.2 ทำการติดต่อกับฐานข้อมูลที่ต้องการ โดยฐานข้อมูลที่จะทำการติดต่อด้วยคือ Microsoft SQL Server โดยผู้ใช้ต้องทำการกรอกข้อมูลคือ ชื่อเซิฟเวอร์ และ ชื่อดาต้าเบส หลังจากนั้นผู้ใช้สามารถกดปุ่ม ทดสอบการเชื่อมต่อหรือไม่ก็ได้ แล้วทำการกดปุ่ม ถัดไป >>



รูปที่ 5.22 หน้าจอแสดงการติดต่อฐานข้อมูล Microsoft SQL Server สำหรับทดสอบโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

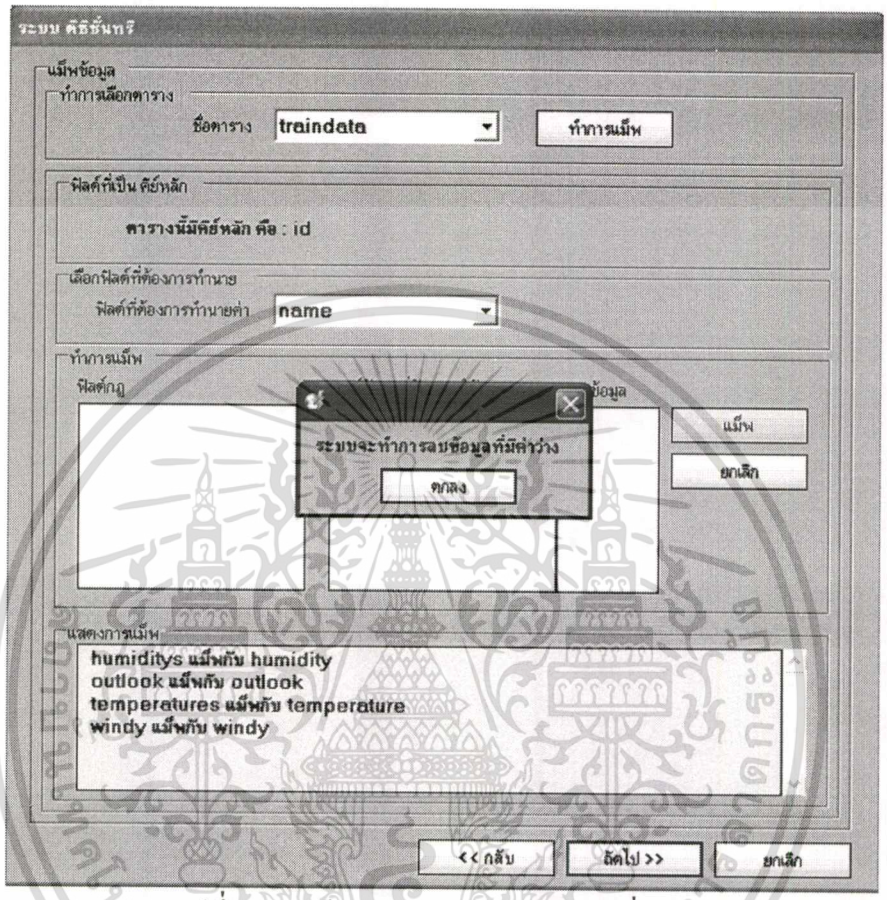
5.2.3 ทำการแม็พข้อมูล โดยให้ทำการเลือก ชื่อตารางที่ต้องการแม็พกับข้อมูลของโมเดลที่ได้ทำการเลือก หลังกจากนั้นให้กดปุ่ม ทำการแม็พ ต่อมาให้เลือกฟิลด์ที่ต้องการทำนายค่า แล้วทำการแม็พฟิลด์ของโมเดลกับฟิลด์ของข้อมูลที่ต้องการทดสอบ แล้วกดปุ่ม แม็พ เพื่อทำการแม็พครบทุกฟิลด์ให้กดปุ่ม ถัดไป >>



รูปที่ 5.23 หน้าจอแสดงกระบวนการแม็พข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

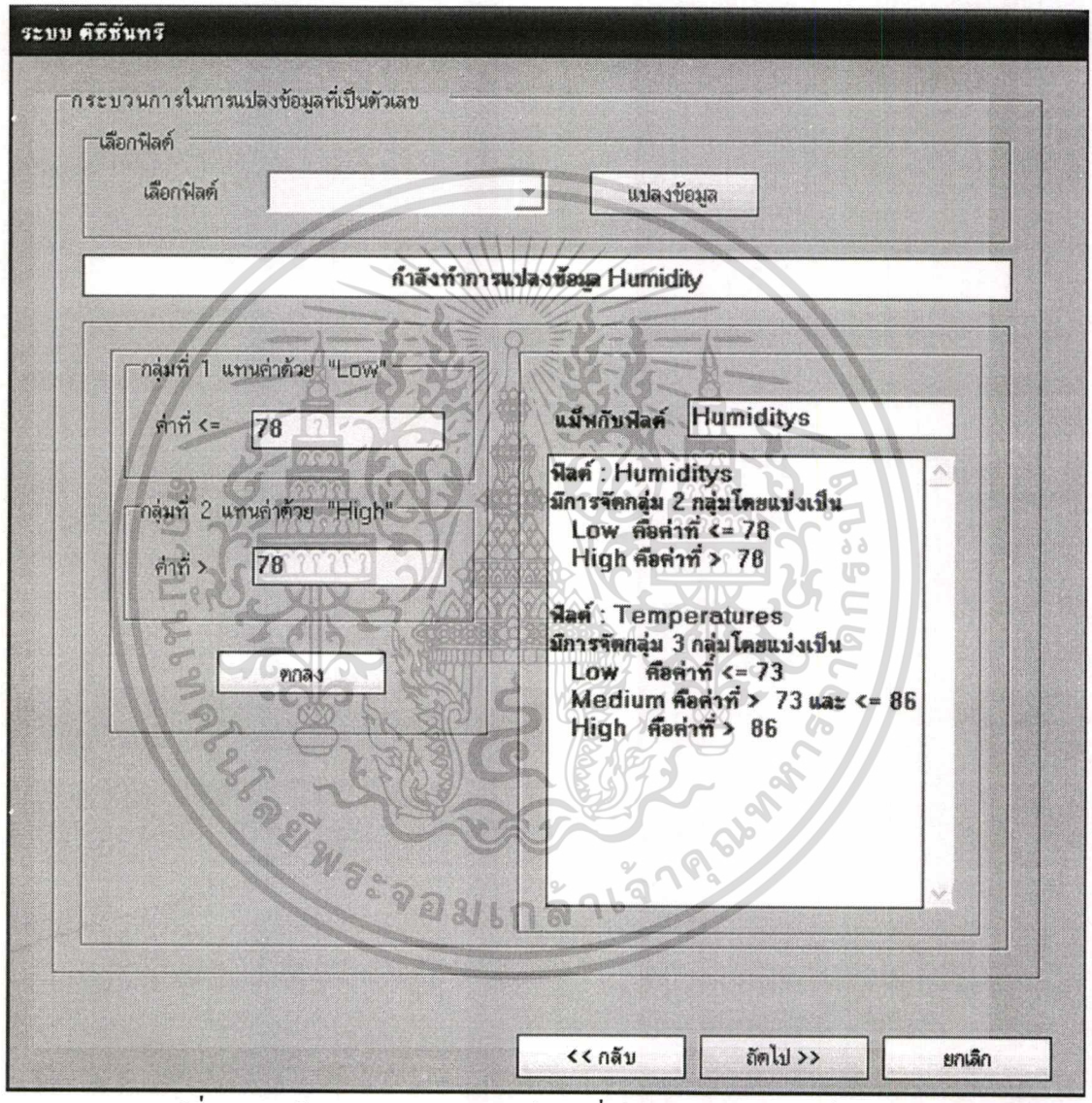
5.2.4 เมื่อกดปุ่ม ถัดไป >> โปรแกรมจะทำการลบข้อมูลที่มีว่างให้โดยอัตโนมัติให้กดปุ่มตกลง



รูปที่ 5.24 หน้าจอแสดงการลบค่าข้อมูลที่ค่าว่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

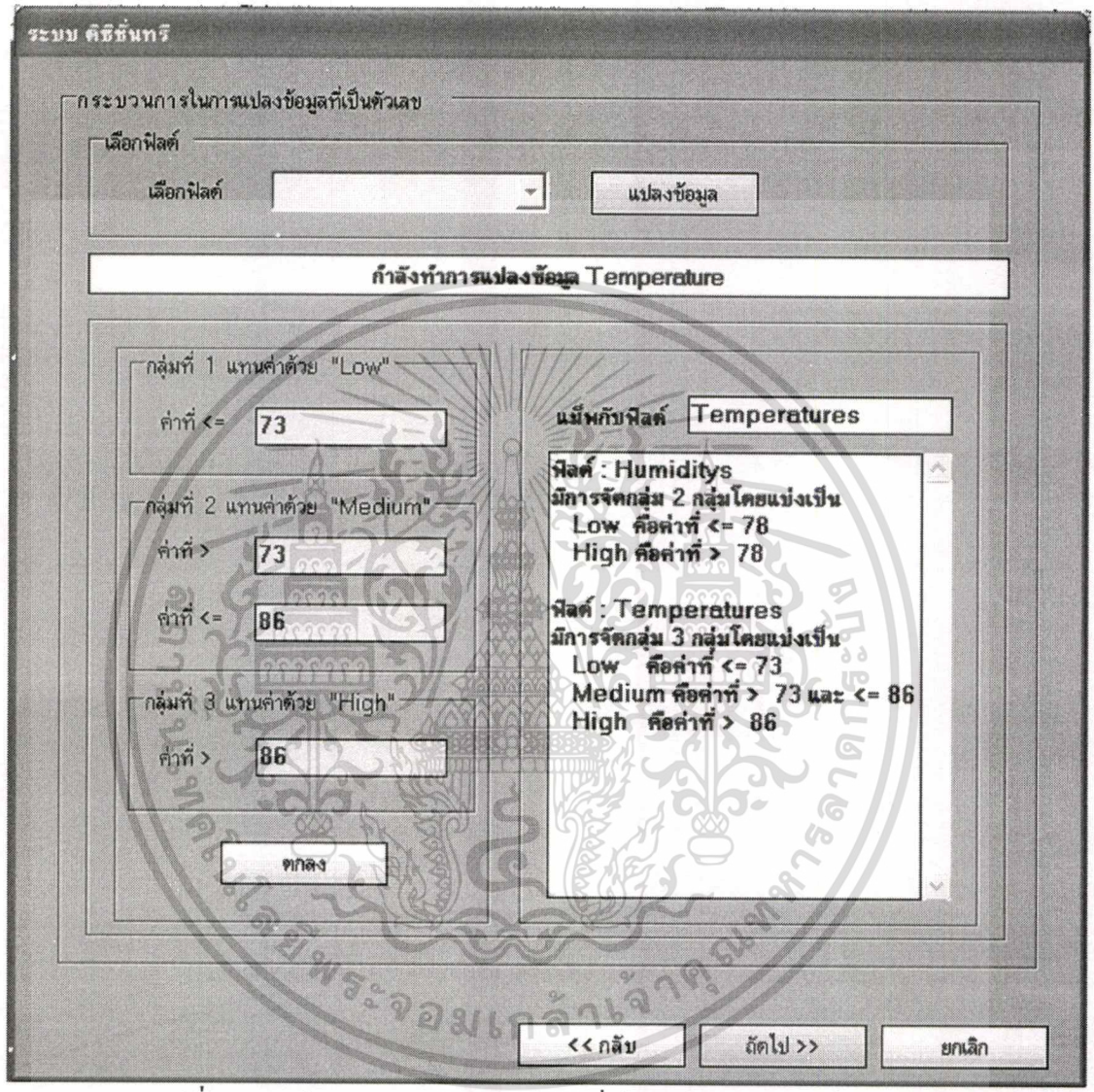
5.2.5 หากข้อมูลใหม่มีข้อมูลที่มีชนิดเป็นตัวเลขให้ทำการแปลงข้อมูลให้เป็นตัวอักษร โดยเลือกฟิลด์ แล้วกดปุ่ม แปลงข้อมูล หลังจากนั้นระบบจะทำการดึงข้อมูลเกี่ยวกับการแปลงข้อมูลของฟิลด์ที่ได้ทำการแม็พกันออกมา แล้วกด ตกลง เมื่อทำการแปลงครบทุกฟิลด์แล้วให้กดปุ่ม ถัดไป >>



รูปที่ 5.25 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลขโดยแบ่งเป็น 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

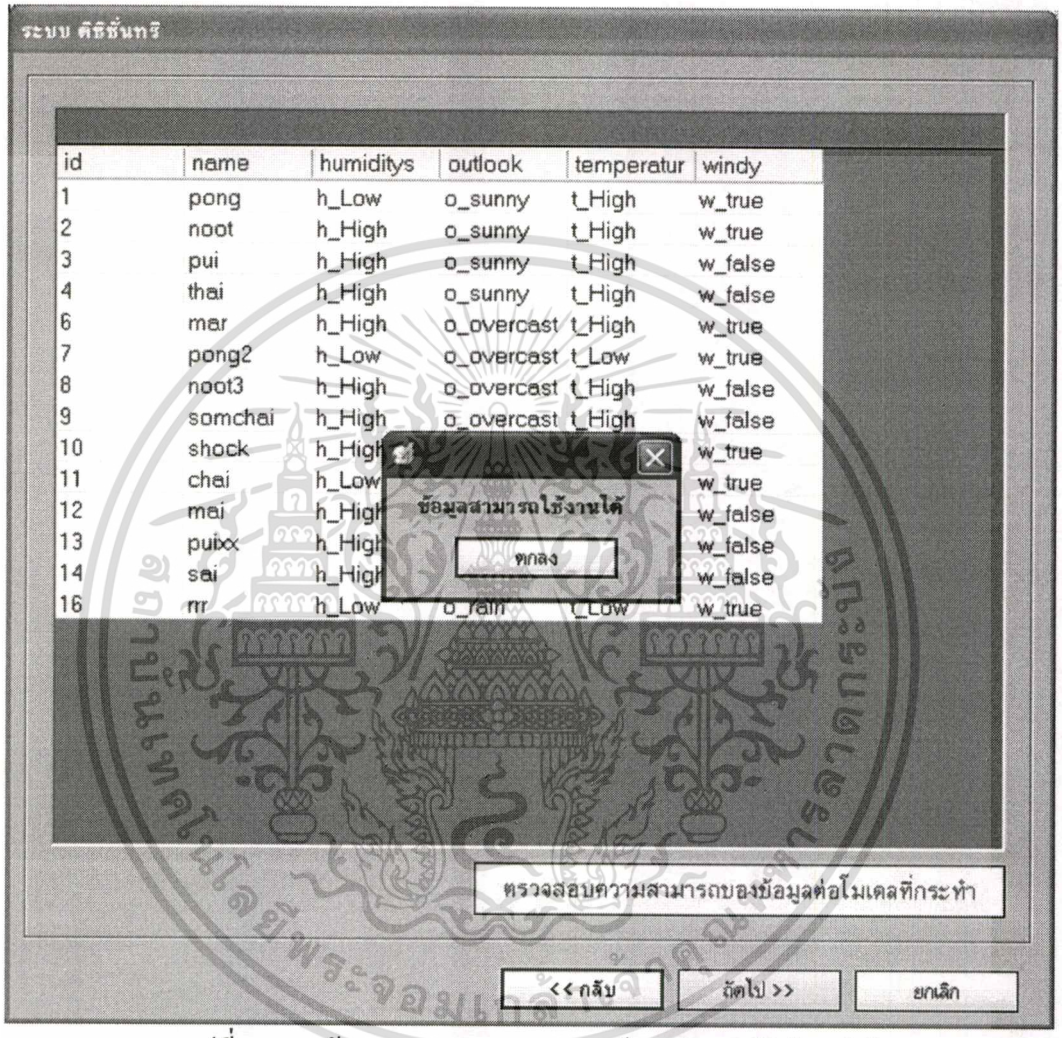
5.2.6 แสดงการแปลงข้อมูลในกรณีที่ผู้ใช้ต้องการจัดกลุ่มเป็น 3 กลุ่ม



รูปที่ 5.26 หน้าจอแสดงการแปลงข้อมูลที่เป็นตัวเลขโดยแบ่งเป็น 3 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

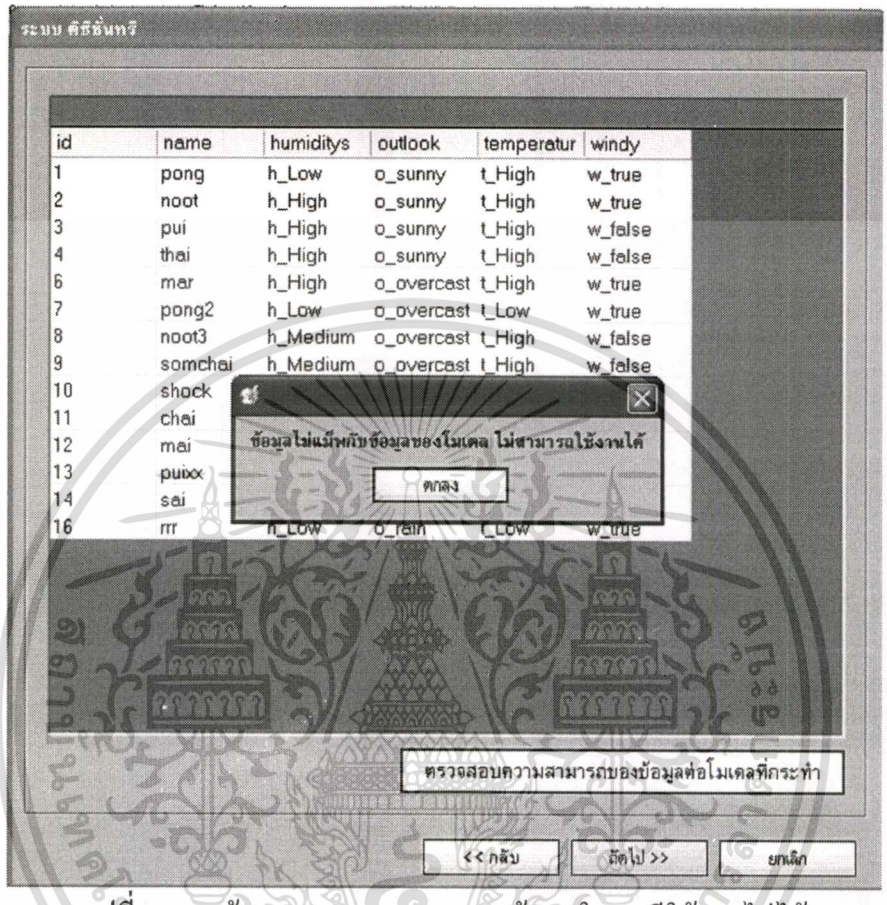
5.2.7 ทำการทดสอบข้อมูลทั้งหมดว่าสามารถที่จะใช้งานกับโมเดลนี้ได้หรือไม่โดยให้ผู้ใช้งานทำการกดปุ่ม ตรวจสอบความสามารถของข้อมูลต่อโมเดลที่กระทำ



รูปที่ 5.27 หน้าจอแสดงการตรวจสอบข้อมูลในกรณีใช้งานได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

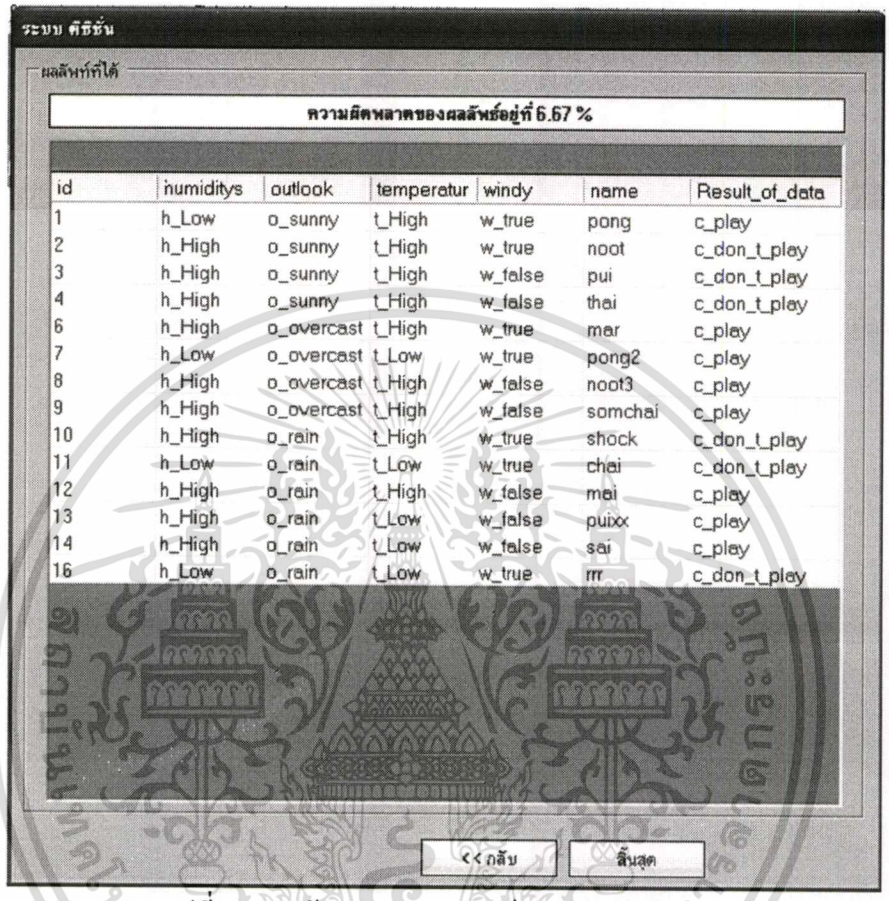
5.2.8 หากไม่สามารถใช้งานได้โปรแกรมก็จะทำการแจ้งให้ผู้ใช้ทราบให้กดปุ่ม ตกลง



รูปที่ 5.28 หน้าจอแสดงการตรวจสอบข้อมูลในกรณีใช้งานไม่ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.9 แสดงผลลัพธ์ของการทดสอบ โดยค่าของผลลัพธ์จะอยู่ในฟิลด์ที่มีชื่อว่า Result_of_data



รูปที่ 5.29 หน้าจอแสดงผลลัพธ์การทดสอบโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปผลการศึกษา และ ข้อเสนอแนะ

โครงการพัฒนาระบบ Decision Trees นี้จัดทำขึ้นเพื่อให้สามารถนำข้อมูลที่มีอยู่นำมาใช้ให้มีประสิทธิภาพเพิ่มมากขึ้น และช่วยเพิ่มประโยชน์ให้กับข้อมูลที่มีอยู่ โดยการนำข้อมูลทั้งหมดที่มีอยู่ มาผ่านกระบวนการทางด้านดาต้าไมน์นิ่งโดยใช้รูปแบบของ Decision Tree ผ่านอัลกอริทึม ID3 แล้วทำการสร้างโครงสร้างต้นไม้ขึ้นมาเพื่อใช้ประโยชน์ต่อไป

6.1 สรุปผลการดำเนินงาน

ในการพัฒนาระบบดาต้าไมน์นิ่งแบบ Decision Tree ปัจจัยที่มีความสำคัญก็คือ ข้อมูลที่นำมาใช้ในการสร้าง Decision Tree นั้นข้อมูลจะต้องมีความสัมพันธ์กันโดยตัวโปรแกรมไม่สามารถที่จะทราบได้ว่าข้อมูลมีความสัมพันธ์กันจริงหรือไม่ โดยระบบ Decision Tree ที่ได้ทำการพัฒนาขึ้นมา สามารถที่จะทำการติดต่อกับ Relational Database Management Systems คือ Microsoft SQL Server โดยผู้ใช้สามารถที่จะระบุถึง ชื่อเซิฟเวอร์ และ ดาต้าเบส ที่ต้องการที่จะติดต่อได้ โดยผู้ใช้ต้องการเลือกชื่อตาราง ได้เพียงตารางเดียวสำหรับการสร้างโครงสร้างต้นไม้ในแต่ละครั้ง และผู้ใช้สามารถที่จะทำการเลือกฟิลด์ที่ต้องการได้ หาย ชนิดของฟิลด์ที่ต้องการเป็นค่าที่เป็นตัวเลข ผู้ใช้สามารถที่จะทำการแปลงเป็นตัวอักษรได้ โดยเมื่อผู้ใช้ทำการสร้างโครงสร้าง ต้นไม้ออกมาแล้ว สามารถที่จะทำการบันทึกข้อมูลของโมเดลต้นไม้ที่ได้ทำการสร้างขึ้นเพื่อที่จะสามารถนำมาทดสอบอีกครั้งได้

6.2 ข้อเสนอแนะ

โปรแกรมที่ได้ทำการพัฒนาขึ้นนั้นในการสร้างโครงสร้างต้นไม้ในแต่ละครั้งผู้ใช้สามารถที่จะทำการเลือกตารางข้อมูลได้เพียงตารางเดียวต่อการสร้างโครงสร้างต้นไม้หนึ่งครั้งดังนั้นควรที่จะสามารถใช้ฟิลด์จากหลายๆ ตารางได้ในการสร้างโครงสร้างต้นไม้ในแต่ละครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กิตติ ภักดีวัฒนะกุล , กิตติพงษ์ กลมกล่อม. 2521. **UML วิเคราะห์และออกแบบระบบเชิงวัตถุ** กรุงเทพ: เคทีพี คอมพ์ แอนด์ คอนซัลท์.

กฤษณะ ไวยมัย และคณะ. 2547. **Data Mining การเตรียมข้อมูลสำหรับดาต้าไมนิง (2).** [Online]. Available: http://micro.se-ed.com/content/mc189/mainframe.asp?tar=mc189_91.asp

สมพร จิวรสกุล. 2545. **คู่มือการติดตั้งและใช้งาน Microsoft SQL Server 2000.** นนทบุรี: อินโฟเพรส.

ศุภชัย สมพานิช. 2546. **สร้างระบบงานฐานข้อมูลด้วย Visual Basic .NET.** กรุงเทพฯ: DevBook.

Ching-Yi Wu, Suan Zhu, Rohit Thadani, Kevin Kirkpatrick and Christopher Swope. 2005. **DATA MINING.** [Online]. Available: http://userfs.cec.wustl.edu/~cse530/2004 /DataMIning.ppt

Decision Tree & C4.5 Algorithm [Online]. Available : http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/c4.5/tutorial.html

SQL Server 2005 Hands-On Labs [Online]. Available : http://blogs.developpeur.org/themit/ archive/2005/04/19/6185.aspx

US-CERT [Online]. Available : http://www.kb.cert.org

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ : นายพยูน พาณิย์กุล

วันเดือนปีเกิด : 16 พฤศจิกายน 2525

สถานที่เกิด : จังหวัด ตรัง

ประวัติการศึกษา :

มัธยมต้น : โรงเรียนสภาราชินีจังหวัดตรัง

มัธยมปลาย : โรงเรียนสภาราชินีจังหวัดตรัง

ปริญญาตรี : มหาวิทยาลัยกรุงเทพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้