ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

การพัฒนาระบบการแบ่งกลุ่มข้อมูลด้วยเซ็ลฟออร์แกไนซิงแม๊ป

DEVELOPMENT OF DATA CLUSTERING SYSTEM USING SELF-ORGANIZING MAP

H004812

โดย

ธวิทย์ เชิดเกียรติสกุล

TAWIT CHOEDKIATSAKUL

อาจารย์ที่ปรึกษา

รศ.ดร.วรพจน์ กรีสุระเดช

วน.
ว399 ก
2550

เลขหมู่.....
เลขทะเบียน 04812
วัน,เดือน,ปี - 8 ต.ค. 2551

.b 11989117.....
.i.....

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 2 ปีการศึกษา 2550

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

DEVELOPMENT OF DATA CLUSTERING SYSTEM USING SELF-ORGANIZING MAP

TAWIT CHOEDKIATSAKUL

A SYSTEM DEVELOPMENT PROJECT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

2/ 2007

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2008

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การพัฒนาระบบการแบ่งกลุ่มข้อมูลด้วยเซ็ลฟออร์แกไนซิงแม็ป
นักศึกษา	นายธวิทย์ เชิดเกียรติสกุล
รหัสนักศึกษา	49066412
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2550
อาจารย์ที่ปรึกษา	รศ.ดร.วรพจน์ กรีสุระเดช

บทคัดย่อ

ในโลกของการดำเนินธุรกิจในปัจจุบันยากที่จะหลีกเลี่ยงการแข่งขัน ไม่ว่าจะเนินธุรกิจใดๆ ก็จำเป็นเข้าสู่เวทีการแข่งขัน การเก็บรวมข้อมูลข่าวสารรวมไปถึงการนำข้อมูลที่เก็บไว้อยู่มาใช้ให้เกิดประโยชน์จึงเป็นส่วนสำคัญเพื่อนำไปใช้ในการวางแผนกลยุทธ์และสร้างความได้เปรียบเหนือคู่แข่ง ดังนั้นการนำเครื่องมือดาต้าไมนิ่งมาใช้จึงเป็นอีกทางเลือกที่สามารถช่วยปรับปรุงการทำงานให้เกิดประสิทธิภาพมากยิ่งขึ้น

โครงงานนี้จึงได้เสนอกระบวนการทำเหมืองข้อมูล โดยอัลกอริทึมที่นำมาใช้ในการทำเหมืองข้อมูลคือ เซ็ลฟออร์แกไนซิงแม็ป เพื่อแบ่งกลุ่มข้อมูล ซึ่งได้นำระบบการเตรียมข้อมูลและการสำรวจข้อมูลรวมกับการแบ่งกลุ่มข้อมูลเพื่อให้การทำดาต้าไมนิ่งมีความสมบูรณ์มากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Development of Data Clustering System Using Self-Organizing Map
Student	Mr. Tawit Choedkiatsakul
Student ID.	49066412
Degree	Master of Science
Programme	Information Science
Academic Year	2007
Advisor	Assoc. Prof. Dr. Worapoj Kreesuradej

ABSTRACT

In the modern world of the business operation, it is difficult to avoid competitions. Any activity needs to enter into a competition. Collecting information and using information are important parts which are used for the best benefit to plan a strategy and to gain an advantage with a competitor. Data mining is one of the choice that can improve the effectiveness of systematic operation.

This project propose data mining process by use Self-Organizing Map algorithm for clustering data so to combine data preparation and exploration system with data mining.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงงานพัฒนาระบบงานฉบับนี้สำเร็จได้ด้วยความกรุณาของ รศ.ดร.วรพจน์ กรีสุระเดช อาจารย์ที่ปรึกษาโครงงานพัฒนาระบบงานซึ่งได้ให้คำปรึกษา คำแนะนำเทคนิคต่างๆ ที่เป็นประโยชน์ต่อการพัฒนาระบบ และสละเวลาในการตรวจสอบแก้ไขข้อบกพร่องของโครงการนี้ และขอขอบพระคุณ คณาจารย์คณะเทคโนโลยีสารสนเทศ ทุกท่านที่ประสิทธ์ประสาทวิชาความรู้ให้

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาวิทยาการสารสนเทศ ขอบคุณเพื่อนร่วมรุ่น IS21.1 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกคนที่ให้คำแนะนำต่างๆ และคอยให้กำลังใจเสมอมารวมไปถึงผู้มีพระคุณทุกท่านที่มิได้เอ่ยนามไว้นะที่นี้

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณบิดา มารดา และครอบครัวที่ได้ให้กำลังใจ ให้การสนับสนุนในทุกๆ เรื่องที่ทำให้โครงงานฉบับนี้สำเร็จลุล่วงเป็นอย่างดี

ธวิทย์ เชิดเกียรติสกุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....I

บทคัดย่อภาษาอังกฤษ.....II

กิตติกรรมประกาศ.....III

สารบัญ.....IV

สารบัญตาราง.....VI

สารบัญรูป.....VII

บทที่ 1 บทนำ.....1

1.1 ความเป็นมาและความสำคัญของปัญหา.....1

1.2 วัตถุประสงค์และประโยชน์ที่คาดว่าจะได้รับ.....1

1.3 ขอบเขตการวิจัย.....2

1.4 ขั้นตอนการศึกษา.....2

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....3

2.1 การทำเหมืองข้อมูล.....3

2.1.1 ความหมายของการทำเหมืองข้อมูล.....3

2.1.2 ขั้นตอนการทำเหมืองข้อมูล.....4

2.1.3 วิธีการทำเหมืองข้อมูล.....5

2.2 ทฤษฎีโครงข่ายประสาทเทียม.....6

2.2.1 โครงสร้างของโครงข่ายประสาทเทียม.....6

2.2.2 การทำงานของโครงข่ายประสาทเทียม.....8

2.2.3 ลักษณะการเรียนรู้ของโครงข่ายประสาทเทียม.....9

2.3 อัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ป.....9

2.3.1 ลักษณะของอัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ป.....9

2.3.2 โครงสร้างของเซ็ลฟออร์แกไนซิงแม็ป.....10

2.3.3 ขั้นตอนการทำงานของเซ็ลฟออร์แกไนซิงแม็ป.....10

2.3.4 อัลกอริทึมในการพัฒนาระบบการแบ่งกลุ่มข้อมูลด้วยเซ็ลฟออร์แกไนซิงแม็ป.11

บทที่ 3 วิเคราะห์และออกแบบระบบ.....12

3.1 ความต้องการและข้อจำกัดของระบบ.....12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.1.1 ความต้องการของผู้ใช้งานระบบ.....12

3.1.2 ความต้องการเชิงฟังก์ชัน.....12

3.1.3 ความต้องการเชิงประสิทธิภาพ.....12

3.2 การวิเคราะห์ระบบโดยใช้ยูสเคสวิวและแอคทิวิตีไดอะแกรม.....13

3.2.1 ยูสเคสไดอะแกรมของระบบ.....13

3.2.2 อธิบายแอคเตอร์และยูสเคส.....13

3.2.3 รายละเอียดยูสเคส.....15

3.2.4 แอคทิวิตีไดอะแกรม.....21

3.2.5 การออกแบบระบบโดยใช้คลาสไดอะแกรม.....27

3.2.6 การออกแบบระบบโดยใช้ซีเควนไดอะแกรม.....28

บทที่ 4 ออกแบบและพัฒนาระบบ.....30

4.1 เครื่องมือที่ใช้ในการออกแบบและพัฒนาระบบ.....30

4.2 ขั้นตอนการทำงานของระบบ.....31

4.3 การออกแบบหน้าจอการทำงานของระบบ.....32

4.4 การใช้งานโปรแกรม.....33

บทที่ 5 สรุปและข้อเสนอแนะ.....42

5.1 สรุปผลการดำเนินงาน.....42

5.2 สรุปผลการทดลอง.....42

5.3 ปัญหาที่พบในการพัฒนา.....42

5.4 ข้อเสนอแนะ.....43

บรรณานุกรม.....44

ประวัติผู้เขียน.....45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

3.1 รายละเอียดยูสเคส การเลือกข้อมูล.....15

3.2 รายละเอียดยูสเคส การคลีนข้อมูล.....16

3.3 รายละเอียดยูสเคส แปลงข้อมูล.....17

3.4 รายละเอียดยูสเคส สำรวจข้อมูล.....18

3.5 รายละเอียดยูสเคส ออกแบบโครงข่าย.....19

3.6 รายละเอียดยูสเคส การใช้ประโยชน์โครงข่าย.....20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

2.1 กระบวนการทำเหมืองข้อมูล.....4

2.2 โครงสร้างของเซลล์ประสาท.....7

2.3 สถาปัตยกรรมของโครงข่ายประสาทเทียม.....8

2.4 ขั้นตอนการทำงานของโครงข่ายประสาทเทียม.....8

2.5 Two self-organized feature maps 10

2.6 อัลกอริทึมการทำงานของ SOM 11

3.1 ยูสเคสไดอะแกรมระบบแบ่งกลุ่มข้อมูล.....13

3.2 แอคทิวิตีไดอะแกรมการเลือกข้อมูล.....21

3.3 แอคทิวิตีไดอะแกรมการคลีนข้อมูล.....22

3.4 แอคทิวิตีไดอะแกรมการแปลงข้อมูล.....23

3.5 แอคทิวิตีไดอะแกรมการสำรวจข้อมูล.....24

3.6 แอคทิวิตีไดอะแกรมสร้างโครงข่ายใหม่.....25

3.7 แอคทิวิตีไดอะแกรมการใช้งานโครงข่าย.....26

3.8 คลาสไดอะแกรมของระบบ.....27

3.9 ซีเควนไดอะแกรมการสร้างโครงข่ายใหม่.....28

3.10 ซีเควนไดอะแกรม การแบ่งกลุ่มด้วย SOM 29

4.1 ขั้นตอนการทำงานของระบบการแบ่งกลุ่มด้วย SOM 31

4.2 หน้าจอรับข้อมูล.....32

4.3 หน้าจอระบบแบ่งกลุ่มด้วย SOM.....32

4.4 ขั้นตอนการติดต่อฐานข้อมูล.....33

4.5 กล่องข้อความแสดงการผลการเชื่อมต่อฐานข้อมูล.....33

4.6 ขั้นตอนการเลือกข้อมูลจาก 1 ตาราง.....34

4.7 ขั้นตอนการเลือกข้อมูลจากหลายตาราง.....34

4.8 กล่องข้อความแสดงผลการเลือกข้อมูล.....35

4.9 การคลีนข้อมูล.....35

4.10 กล่องข้อความแสดงผลการคลีนข้อมูลเสร็จเรียบร้อย.....35

4.11 ขั้นตอนการแปลงข้อมูล.....36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.12 ขั้นตอนการสำรวจข้อมูล.....37

4.13 กล่องข้อความแสดงผลการสร้างตารางข้อมูลสำหรับการทำไมนิ่ง..... 37

4.14 ขั้นตอนการเลือกแบบจำลองที่ใช้ทำเหมืองข้อมูล.....38

4.15 หน้าจอ Create New Map ของแบบจำลองการแบ่งกลุ่มด้วย SOM.....39

4.16 หน้าจอ Create New Map เลือกแท็บ Advance.....39

4.17 กล่องข้อความแสดงหลังจากทำการบันทึกข้อมูล Model เรียบร้อยแล้ว.....39

4.18 กล่องข้อความแสดงหลังจากโหลด Model เรียบร้อยแล้ว.....40

4.19 หน้าจอ Deployment Map ของแบบจำลองการแบ่งกลุ่มด้วย SOM.....40

4.20 กล่องข้อความแสดงหลังจากทำการบันทึกข้อมูลผลลัพธ์เรียบร้อยแล้ว.....41

4.21 แสดงกราฟค่า mean แต่ละมิติของข้อมูลแต่ละกลุ่ม.....41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเทคโนโลยีสารสนเทศเข้ามามีบทบาทอย่างกว้างขวางในทุกวงการ และส่งผลให้กลายมาเป็นเครื่องมือสำคัญในการทำงานทุกด้าน นับตั้งแต่ด้านเศรษฐกิจ พานิชยกรรม อุตสาหกรรม การวิจัยและการพัฒนาตลอดจนด้านการเมืองและราชการ ทั้งนี้ได้มีการนำเทคโนโลยีสารสนเทศมาประยุกต์ใช้กับข้อมูลที่มีอยู่ภายในองค์กรนั้นๆ เพื่อพัฒนาเป็นระบบการตัดสินใจที่ช่วยวิเคราะห์ข้อมูลต่างๆ เพื่อนำสารสนเทศที่ได้มาใช้ประโยชน์ในการตัดสินใจเพื่อให้บรรลุวัตถุประสงค์ขององค์กร เนื่องจากหากผู้บริหารมีข้อมูลที่มีประสิทธิเพื่อใช้ในการประกอบการตัดสินใจของผู้บริการ ผลลัพธ์ของการตัดสินใจย่อมมีประสิทธิภาพต่อหน่วยงานด้วยเช่นกัน การศึกษาและพัฒนาโครงงานนี้จึงได้นำเทคโนโลยีการทำเหมืองข้อมูลมาใช้ในการวิเคราะห์ข้อมูล เพื่อใช้วิเคราะห์ข้อมูลและมุ่งเน้นอธิบายความสัมพันธ์ของข้อมูลที่ต้องการ เพื่อนำความสัมพันธ์ของข้อมูลมาใช้ให้เกิดประโยชน์มากยิ่งขึ้น ทั้งนี้ได้มีการนำระบบเตรียมข้อมูลและสำรวจข้อมูลมาใช้ประกอบกับการทำเหมืองข้อมูลด้วย

1.2 วัตถุประสงค์และประโยชน์ที่คาดว่าจะได้รับ

โครงงานฉบับนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการนำระบบโครงข่ายประสาทเทียม มาใช้ในการแบ่งกลุ่มข้อมูล ซึ่งข้อมูลที่นำมาใช้จะอยู่ในรูปข้อมูลตัวเลข โดยในการพัฒนาโครงงานมีวัตถุประสงค์ ดังต่อไปนี้

1. ศึกษาเทคโนโลยีการทำเหมืองข้อมูล เพื่อนำมาใช้ออกแบบและพัฒนาระบบ
2. ศึกษาวิธีการและขั้นตอนในการทำเหมืองข้อมูลโดยที่เน้นไปที่อัลกอริทึมเซ็ลฟออร์แกไนซิงแม๊ป
3. สร้างระบบแบ่งกลุ่มด้วยอัลกอริทึมเซ็ลฟออร์แกไนซิงแม๊ป
4. เพิ่มมูลค่าของข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลภายในองค์กร โดยนำมาอธิบายในรูปความสัมพันธ์เพื่อให้เกิดประโยชน์ในการใช้งานสูงสุด
5. เพื่อนำผลลัพธ์จากการทำเหมืองข้อมูลไปใช้สนับสนุนการทำงานขององค์กร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตการวิจัย

การทำงานของระบบการแบ่งกลุ่มด้วยอัลกอริทึมเซ็ลฟออร์แกไนซิงแม๊ปมีขอบเขตการศึกษาและพัฒนาดังนี้

1. นำข้อมูลไปผ่านกระบวนการเตรียมข้อมูลก่อนทำดาต้าไมนิ่ง
2. สร้างอัลกอริทึมเพื่อนำกลุ่มของข้อมูลมาประมวลผล
3. นำผลลัพธ์ของการประมวลผลกลุ่มข้อมูลมาแสดงผล

1.4 ขั้นตอนของการศึกษา

การพัฒนาโครงงานนี้มีการนำข้อมูลจากแหล่งข้อมูลที่น่าเชื่อถือ และต้องมีการจัดเตรียมการศึกษา ออกแบบ และพัฒนา โดยมีขั้นตอนดังนี้

1. ศึกษาทฤษฎีที่เกี่ยวข้องเกี่ยวกับการทำเหมืองข้อมูลเพื่อนำมาประยุกต์ใช้ในการพัฒนาระบบ
2. การจัดการให้ข้อมูลอยู่รูปแบบที่เหมาะสมและมีการจัดกลุ่มเปลี่ยนแปลงค่าเพื่อให้สามารถนำไปวิเคราะห์หาผลลัพธ์ได้ตรงตามวัตถุประสงค์
3. สร้างรูปแบบจำลองเพื่อหาความสัมพันธ์ของข้อมูลที่ผ่านการจัดรูปแบบให้ถูกต้องแล้วโดยใช้อัลกอริทึมเซ็ลฟออร์แกไนซิงแม๊ปในการวิเคราะห์เพื่อหาผลลัพธ์
4. จัดเตรียมรายงานผลความสัมพันธ์ที่ผ่านการประมวลผลจากอัลกอริทึมในรูปแบบรายงานที่ผู้ใช้งานสามารถเข้าใจง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

ในหัวข้อนี้จะกล่าวถึงทฤษฎีพื้นฐานต่างๆ ที่เกี่ยวข้องในการพัฒนาระบบ เนื้อหาในบทนี้จะกล่าวถึง ทฤษฎีดาต้าไมนิ่ง อัลกอริทึมเซ็ลฟออร์แกไนซิงแม๊ปเป็นอัลกอริทึมที่ใช้แบ่งกลุ่มข้อมูลที่ใช้หลักการของระบบโครงข่ายประสาทเทียม โดยมีรายละเอียดที่เกี่ยวข้องกับการพัฒนาระบบงาน ดังนี้

2.1 การทำเหมืองข้อมูล

2.1.1 ความหมายของการทำเหมืองข้อมูล

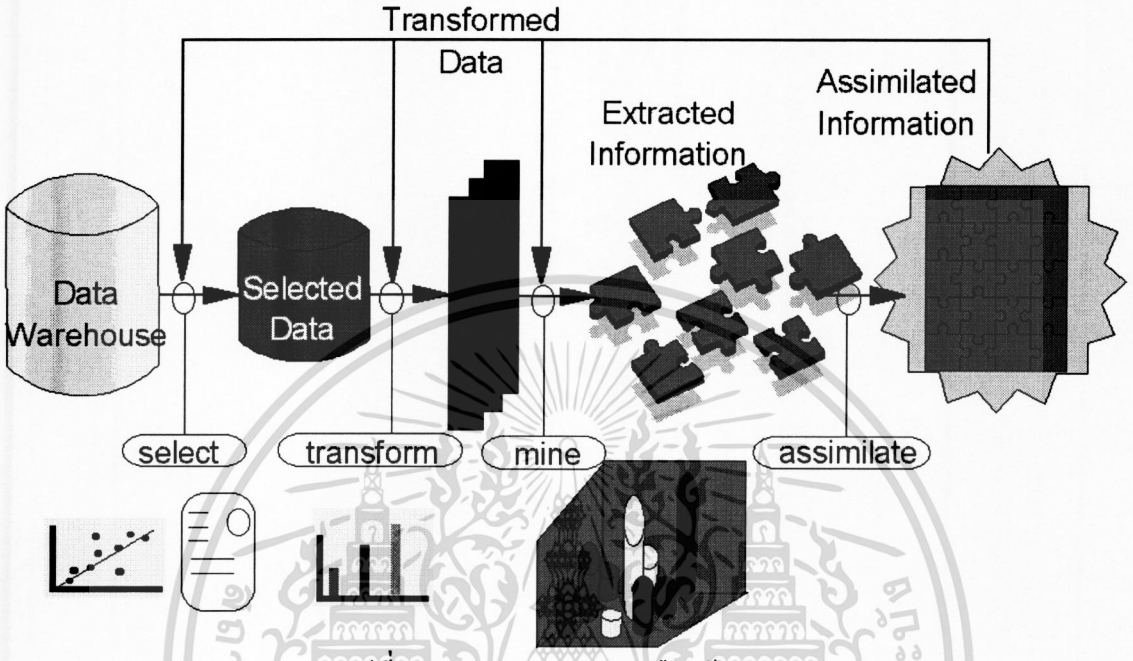
การทำเหมืองข้อมูลเป็นส่วนประกอบหนึ่งที่สำคัญในการทำธุรกิจแบบอัจริยะ (Business Intelligence) มีความแตกต่างจากภาษาคิวรีที่ตอบได้เพียงคำถามแบบง่ายๆ เท่านั้น เช่น ข้อมูลที่อยู่ในแถวมีอะไรบ้าง แต่ไม่สามารถตอบปัญหาที่มีความสัมพันธ์ และมีความซับซ้อนได้ ซึ่งการทำเหมืองข้อมูลมีเป้าหมายที่ชัดเจนเพื่อมุ่งเน้นไปที่การค้นหาแหล่งความรู้ใหม่ที่ถูกซ่อนไว้ในตัวข้อมูล

การทำเหมืองข้อมูลถูกจัดให้ถูกจัดให้เป็นกระบวนการค้นหาความสัมพันธ์ในหลายๆ รูปแบบข้อมูลหรือหาสิ่งที่มีประโยชน์ และไม่เคยเจอมาก่อนจากฐานข้อมูลที่มีตารางขนาดใหญ่ เพื่อเพิ่มมูลค่าให้กับข้อมูลที่หน่วยงานนั้นๆ มีอยู่ โดยที่ข้อมูลนั้นสามารถหาได้จากหลายๆ แหล่ง เช่น ข้อมูลที่ได้จากระดับปฏิบัติการ, จากฐานข้อมูล, คลังข้อมูล มาผ่านกระบวนการจนได้ความรู้แล้วนำความรู้ที่ได้มานั้นไปเป็นส่วนช่วยในการดำเนินงานขององค์กรหรือธุรกิจทางด้านต่างๆ ในหลากหลายสาขา อาทิ

- การนำแหล่งข้อมูลขนาดใหญ่ที่ไม่ได้ใช้งานแล้ว นำกลับมาใช้เพื่อเป็นตัวทำนายอนาคตได้
- การนำผลการทำนาย และการวิเคราะห์มาใช้เป็นกลยุทธ์ในการเพิ่มยอดขายให้กับธุรกิจการค้า
- การทำนายการซื้อสินค้าในช่วงเวลาต่างๆ ช่วยให้สามารถนำมาออกแบบการจัดซื้อผลิตภัณฑ์ในคลังสินค้าได้ถูกต้อง และคุ้มค่ากับต้นทุนที่จ่ายไป
- การวิเคราะห์สินค้า เพื่อให้ได้สินค้าที่ตรงกับความต้องการของลูกค้า รวมไปถึงการกำหนดราคาสินค้าที่เหมาะสม
- การทำนายข้อมูลความสัมพันธ์ กับการซื้อสินค้าแต่ละชนิดของลูกค้าทำให้บริษัทสามารถกำหนดที่วางสินค้าให้เหมาะสม และเป็นการช่วยให้มีการสร้างยอดขายให้เพิ่มมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- อีคอมเมิร์ซ ทำการวิเคราะห์พฤติกรรมการใช้เว็บไซต์ของลูกค้าตามลำดับก่อนหลัง ซึ่งจะได้ทราบถึงสิ่งที่ลูกค้าต้องการเพื่อนำไปปรับปรุงเว็บไซต์ให้เข้ากับความต้องการของลูกค้า



รูปที่ 2.1 กระบวนการทำเหมืองข้อมูล

จากรูปที่ 2.1 แสดงถึงกิจกรรมของแต่ละขั้นในกระบวนการทำเหมืองข้อมูล ซึ่งแต่ละกิจกรรมสามารถทำได้มากกว่า 1 ครั้งหรือ วนกลับมาทำใหม่ได้

2.1.2 ขั้นตอนการทำเหมืองข้อมูล

ในการทำเหมืองข้อมูลมีขั้นตอนต่างๆ ดังต่อไปนี้

1. คัดเลือกข้อมูล (Data Selection)

เป็นการเลือกข้อมูลหรือระบุถึงแหล่งข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล ให้ตรงกับวัตถุประสงค์ทางธุรกิจที่ได้กำหนดไว้

2. ปรับเปลี่ยนรูปแบบของข้อมูล (Data Transformation)

เป็นการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลที่เหมาะสมตรงตามอัลกอริทึมและแบบจำลองที่เลือกใช้ทำเหมืองข้อมูลที่จะนำมาวิเคราะห์

3. การทำเหมืองข้อมูล (Data Mining)

เป็นขั้นตอนการเลือกอัลกอริทึมให้เหมาะสมกับข้อมูลที่เตรียมไว้ และทำการประมวลผลของข้อมูลตามอัลกอริทึมที่เลือก ซึ่งการจะใช้อัลกอริทึมใดนั้นขึ้นอยู่กับความต้องการหรือปัญหาที่เกิดขึ้นเป็นหลัก และในการเลือกใช้อัลกอริทึมที่เหมาะสมจะทำให้ได้ผลลัพธ์เพื่อนำไปใช้ในการวิเคราะห์ที่ถูกต้องมากขึ้น

เอกสารนี้ ... ไม่ว่ากรณีใดๆ ... และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. วิเคราะห์ผลลัพธ์และรวบรวมความรู้ (Assimilate Information)

เป็นการวิเคราะห์และรวบรวมความรู้ที่ได้จากการผ่านกระบวนการทำเหมืองข้อมูลว่ามีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ โดยที่ผลลัพธ์อาจจะแสดงในรูปแบบที่เป็นกราฟหรือตารางข้อมูล ซึ่งจะเป็นข้อมูลความรู้ (Knowledge) เพื่อเป็นส่วนช่วยในการตัดสินใจทางธุรกิจ

2.1.3 วิธีการทำเหมืองข้อมูล

วิธีการทำเหมืองข้อมูลตามมาตรฐานทั่วไปสามารถแบ่งออกเป็น 4 วิธี คือ

1. Predictive Model

เป็นการสร้างแบบจำลองจากข้อมูลเพื่อใช้ในการคาดคะเนหรือทำนายความเป็นไปได้ของข้อมูลที่มีอยู่ เช่น การทำนายอุณหภูมิ, การทำนายหุ้น เป็นต้น โดย Predictive Modeling แบ่งเป็น 2 ลักษณะคือ

- Classification เป็นการจำลองเพื่อการแบ่งกลุ่มของข้อมูลที่ต้องการโดยที่กลุ่มของข้อมูลนั้นจะแบ่งกลุ่มตามชนิดของข้อมูลที่ควรจะเป็น เช่น การแบ่งกลุ่มการพิจารณาการให้วงเงินสินเชื่อของลูกค้า โดยเทคนิคที่นิยมใช้คือ Tree Induction และ Neural Induction
- Value Prediction เป็นการพยากรณ์ค่าที่เป็นไปได้ หรือการกระจายของค่าที่เป็นไปได้ของตัวแปรใดๆ ในกลุ่มข้อมูล การทำนายค่าที่เป็นตัวเลข เช่น การทำนายภาษีที่จะเก็บได้ในแต่ละปี หรือการทำนายราคาหุ้น เป็นต้น โดยเทคนิคที่นิยมใช้คือ Linear regression และ Nonlinear regression

2. Database Segmentation

เป็นการแบ่งกลุ่มของข้อมูลที่ไม่สามารถกำหนดได้ว่าข้อมูลอยู่ในกลุ่มใด โดยเหมาะกับข้อมูลที่มีลักษณะเป็นกลุ่ม โดยเฉพาะจำนวนของกลุ่มน้อย ๆ โดยที่ข้อมูลใดที่มีลักษณะคล้ายกันหรือใกล้เคียงกันก็จะจัดเป็นกลุ่มเดียวกัน เพื่อให้ง่ายต่อการวิเคราห์ เช่น การแบ่งกลุ่มตามรายได้ของประชากร เพื่อนำผลลัพธ์ที่ได้ไปวิเคราะห์และเสนอสินค้าให้ตรงกับกลุ่มที่ได้ทำการแบ่งไว้ กล่าวคือ หากประชากรที่มีรายได้อยู่ในกลุ่มปานกลางก็ควรเสนอสินค้าราคาและคุณภาพที่ไม่สูงและไม่ต่ำเกินไป เป็นต้น ซึ่งเราสามารถเรียกการทำ Database Segmentation อีกอย่างหนึ่งว่า Clustering โดยเทคนิคที่นิยมใช้คือ Demographic Clustering และ Neural Clustering

3. Link Analysis

การหาความสัมพันธ์ของข้อมูลแต่ละเรคคอร์ดหรือกลุ่มเรคคอร์ด ซึ่งการหาความสัมพันธ์ของข้อมูลนี้ส่วนใหญ่จะนำไปใช้ในการวิเคราะห์หาความสัมพันธ์ของสินค้าที่ลูกค้ามักจะซื้อคู่กัน โดย Link Analysis แบ่งเป็น 3 ชนิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Association discovery ใช้วิเคราะห์การซื้อสินค้าภายในรายการเดียวกัน ศึกษาถึงความสัมพันธ์อย่างใกล้ชิดที่ถูก ปิดซ่อนอยู่ของสินค้า ซึ่งสินค้าเหล่านั้นมักมีแนวโน้มที่จะถูกซื้อควบคู่กันไป การวิเคราะห์แบบนี้เรียกว่า "Market Basket Analysis" คือรายการทั้งหมดที่ลูกค้าซื้อต่อครั้งที่ซุปเปอร์มาร์เก็ต Sequential pattern discover ถูกใช้ระบุความเกี่ยวเนื่องกันของพฤติกรรมในระยะยาวว่าพฤติกรรมหรือเหตุการณ์ใดเกิดขึ้นแล้วจะมีพฤติกรรมหรือเหตุการณ์ใดเกิดขึ้นตามมา เช่น ผู้ขายอาจพบว่าลูกค้าที่ซื้อทีวีมีแนวโน้มที่จะซื้อวีดีโอในเวลาต่อมา
- Similar time sequence discovery ใช้ค้นหาความเกี่ยวเนื่องกันระหว่างกลุ่มของข้อมูล 2 กลุ่มซึ่งการขึ้นต่อกัน ทางด้านเวลา โดยมีรูปแบบการเคลื่อนที่เหมือนกัน ผู้ขายสินค้ามักจะใช้เพื่อดูแนวโน้มเพื่อเตรียมสต๊อก เช่นเมื่อไรก็ตามที่ ยอดขาย สินค้าน้ำอัดลมสูงขึ้น ยอดขายมันฝรั่งจะสูงขึ้นตาม

4. Deviation Detection

เป็นเทคนิคที่ใช้ในการหาค่าที่ต่างไปจากที่คาดไว้ว่าค่านั้นต่างไปมากน้อยเพียงใด ซึ่งใช้เทคนิคทางด้านสถิติช่วยในการสรุปข้อมูลออกมาให้อยู่ในรูปแบบของกราฟิกเพื่อแสดงให้ภาพ เพื่อใช้วัดความน่าเชื่อถือของข้อมูลและแสดงให้เห็นภาพ โดยที่เทคนิคนี้สามารถนำไปใช้ตรวจสอบได้หลายด้าน เช่น ตรวจสอบลายเซ็นปลอม, บัตรเครดิตปลอม เป็นต้น

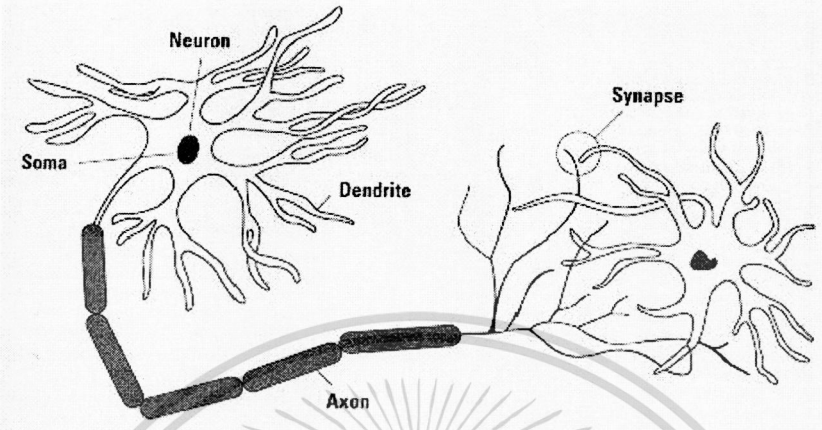
2.2 ทฤษฎีโครงข่ายประสาทเทียม

2.2.1 โครงสร้างของโครงข่ายประสาทเทียม

โครงสร้างของโครงข่ายประสาทเทียม มีลักษณะคล้ายคลึงกับสมองของมนุษย์ซึ่งประกอบด้วยเซลล์ประสาทที่เรียกว่านิวรอล (Neuron) จำนวนมากเชื่อมต่อกัน โดยแต่ละนิวรอลจะมีเส้นเชื่อมต่อกันเพื่อทำการรับส่งข้อมูล ซึ่งแต่ละนิวรอลสามารถมีอินพุตได้หลายอินพุต แต่จะมีเอาท์พุตเพียงเอาท์พุตเดียว เอาท์พุตจะส่งต่อไปยังอินพุตของนิวรอลอื่นๆ เป็นการส่งสัญญาณในระยะเวลาที่รวดเร็วมาก เนื่องจากมีรูปแบบการส่งสัญญาณในลักษณะขนาน (Parallel) กัน ทำให้สามารถทำงานพร้อมๆ กันได้หลายงาน และสำหรับวิธีการแก้ปัญหาของมนุษย์จะใช้ประสบการณ์ที่เกิดขึ้นจากการเรียนรู้ข้อมูลต่างๆ ในอดีตนำมาวิเคราะห์ แต่ในขณะที่สถาปัตยกรรมของคอมพิวเตอร์นั้น จะประกอบด้วยหน่วยประมวลผลเพียงอย่างเดียวเท่านั้น รวมถึงในการทำงานนั้นจะใช้ชุดคำสั่งที่เป็นลำดับ ทำให้การแก้ปัญหาต้องทำตามลำดับที่แน่นอน จึงอาจมีปัญหาที่ไม่สามารถแก้ปัญหาโดยใช้การทำงานในลักษณะนี้ได้ เช่น การจดจำหน้าตาของคน ซึ่งสมองของมนุษย์นั้นสามารถแก้ปัญหาลักษณะเช่นนี้ได้ ถึงแม้เวลาจะผ่านมาสักระยะหนึ่งแล้ว ซึ่งต่างจากคอมพิวเตอร์ ดังนั้นจึงได้มีการพัฒนาคอมพิวเตอร์ให้สามารถประมวลผลได้ใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะที่คล้ายคลึงกับสมองของมนุษย์ หรือที่รียกว่า โครงข่ายประสาทเทียม (Artificial Neural Network)



รูปที่ 2.2 โครงสร้างของเซลล์ประสาท

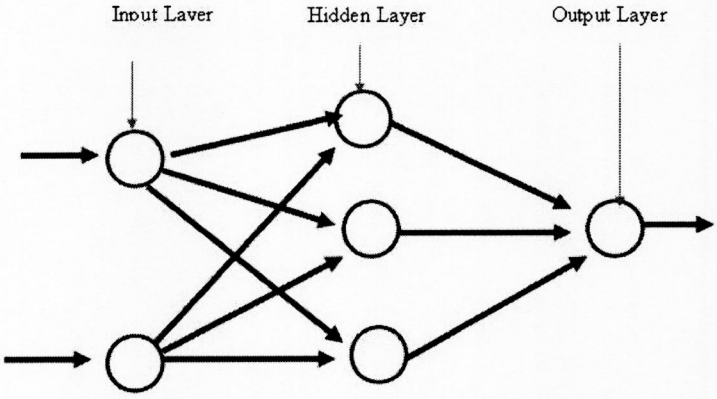
โครงข่ายประสาทเทียม ประกอบด้วยกลุ่มของข้อมูลอินพุต และข้อมูลเอาท์พุต ที่มีการเชื่อมต่อกันนั้น จะมีการให้น้ำหนัก (Weights) เพื่อให้ได้ผลลัพธ์ที่ถูกต้องมากที่สุด ลักษณะการเชื่อมต่อระหว่างหน่วยที่อยู่ต่างชั้น (Layer) เพื่อทำการส่งสัญญาณมี 2 ลักษณะ คือ

- Feedforward : ลักษณะการไหลของข้อมูลของการเชื่อมต่อลักษณะนี้คือ ข้อมูลจะไหลไปตามการเชื่อมต่อในทิศทางเดียว จะไม่มีการไหลย้อนกลับ
- Feedback : ลักษณะการไหลของข้อมูลของการเชื่อมต่อลักษณะนี้คือ ข้อมูลจะไหลย้อนกลับได้ หรืออาจจะมีการไหลของข้อมูลในลักษณะเรียกซ้ำได้

โดยทั่วไปโครงข่ายประสาทเทียมจะแบ่งออกเป็น 3 ชั้น (Layer) ด้วยกันคือ

1. ชั้นอินพุต (Input Layer) ทำหน้าที่รับสัญญาณจากภายนอกผ่านเข้ามาในโครงข่าย แล้วกระจายไปในชั้นถัดไป โดยไม่มีการคำนวณใดๆ
2. ชั้นซ่อน (Hidden Layer) เป็นชั้นที่อยู่ระหว่างกลางระหว่างชั้นอินพุตและชั้นเอาท์พุต ซึ่งในชั้นนี้สามารถมีได้มากกว่า 1 ชั้น
3. ชั้นเอาท์พุต (Output Layer) ทำหน้าที่ส่งผลลัพธ์ที่คำนวณได้ออกจากโครงข่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

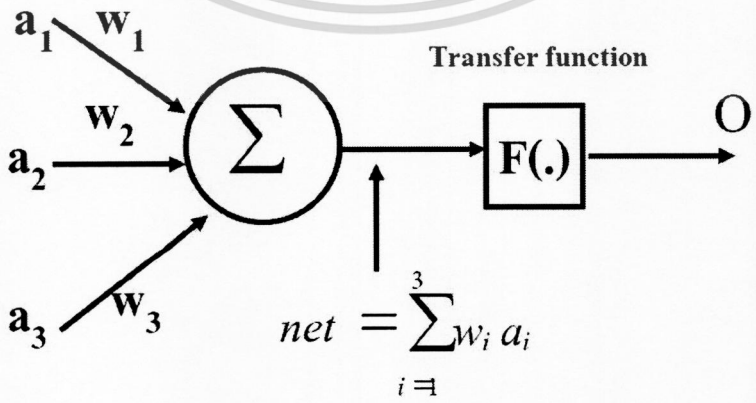


รูปที่ 2.3 สถาปัตยกรรมของโครงข่ายประสาทเทียม

2.2.2 การทำงานของโครงข่ายประสาทเทียม

การทำงานของโครงข่ายประสาทเทียมจะขึ้นอยู่กับค่าน้ำหนัก และฟังก์ชันโยกย้าย (Transfer Function) โดยมีการกำหนดค่าน้ำหนักเริ่มต้นของโครงข่ายประสาทเทียมอยู่ระหว่าง -0.5 ถึง 0.5 และสำหรับฟังก์ชันโยกย้ายหรือฟังก์ชันกระตุ้นนั้น จะถือเป็นฟังก์ชันที่ทำหน้าที่กระตุ้นการทำงานให้เกิดการย้ายค่าอินพุตไปยังเอาท์พุตซึ่งจะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปแล้วฟังก์ชันโยกย้ายที่นิยมใช้ คือ Sigmoid Function ซึ่งมีขั้นตอนการทำงานดังนี้

- ขั้นตอนแรก เมื่อได้อินพุตเข้ามาในโครงข่ายแล้ว ก็จะคูณค่าอินพุตนั้นกับค่าน้ำหนักในแต่ละลิ้งค์
- ขั้นที่สอง ทำการบวกผลลัพธ์ที่ได้จากขั้นตอนแรกเข้าด้วยกัน
- ขั้นที่สาม จะนำผลรวมในขั้นตอนที่สองมาผ่านฟังก์ชันโยกย้าย แล้วจะได้ผลลัพธ์เป็นค่าเอาท์พุตของโครงข่าย



รูปที่ 2.4 ขั้นตอนการทำงานของโครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.3 ลักษณะการเรียนรู้ของโครงข่ายประสาทเทียม

ลักษณะการเรียนรู้ของโครงข่ายประสาทเทียม สามารถแบ่งได้เป็น 2 ประเภท คือ

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้แบบมีผู้สอน คือ จะมีการกำหนดเซตของการฝึกหัดให้กับโครงข่ายประสาทเทียม ซึ่งเซตนี้จะประกอบด้วยอินพุต และเอาท์พุตที่ต้องการ โดยทำการจับคู่สอน (Train Pair) เมื่อทำการป้อนอินพุตให้กับโครงข่ายแล้ว โครงข่ายจะทำการประมวลผลจนได้คำตอบ และค่าถ่วงน้ำหนัก (Weight) ออกมาชุดหนึ่ง สำหรับคำตอบที่ได้จริง (Actual Output) เพื่อนำมาคำนวณค่าผิดพลาด โดยสามารถหาได้ จากการนำคำตอบที่เราต้องการ (Target Output) ลบกับคำตอบที่ได้จริงจากการเรียนรู้ของโครงข่าย ถ้าค่าความผิดพลาดที่ได้จากการลบกันระหว่างค่าที่ได้จริงกับค่าที่เราต้องการแล้วยังมีค่าสูง หรือ ในบางกรณีอาจจะต้องปรับให้มีค่าเป็นศูนย์ จึงต้องการทำปรับค่าถ่วงน้ำหนักไปจนกว่าจะได้ค่าที่ยอมรับได้จึงหยุดการสอนโครงข่าย

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน คือ การเรียนรู้โดยที่ไม่ต้องมีการจับคู่สอน เนื่องจากเมื่อเราใส่ค่าอินพุตเซตเข้าสู่โครงข่ายแล้วมันจะพยายามจัดกลุ่มอินพุตเซตที่มีลักษณะเดียวกันให้เอาท์พุตออกมาจากโครงข่ายที่เดียวกัน คือ มีการปรับตัวเองภายใน (Self-Organization) การเรียนรู้โดยวิธีนี้จะป้อนอินพุตเข้าสู่โครงข่ายภายใน โครงข่ายจะมีเอาท์พุตโหนด อยู่หลายโหนด โดยแต่ละโหนด แทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนอินพุตเข้าสู่โครงข่าย โครงข่ายจะทำการคำนวณค่าความสัมพันธ์ที่มีภายในเซตของอินพุตของโครงข่าย การเรียนรู้ คือการใช้ค่าน้ำหนักเป็นตัวแยกความแตกต่างของอินพุต ทำให้การเรียนรู้ในลักษณะนี้ผู้ใช้งานจะต้องทำการกำหนดเอาท์พุตเอง ส่งผลให้ค่าที่ได้อาจจะไม่ถูกต้อง

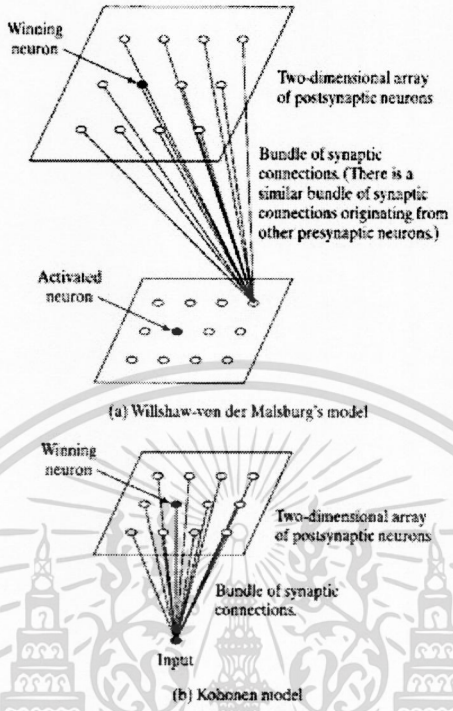
2.3 อัลกอริทึมเซ็ลฟออร์แกไนซิงแม๊ป

2.3.1 ลักษณะของอัลกอริทึมเซ็ลฟออร์แกไนซิงแม๊ป

เซ็ลฟออร์แกไนซิงแม๊ปพัฒนาโดย Kohonen แบบจำลอง SOM เป็นแบบจำลองระบบโครงข่ายแบบไม่มีผู้สอน (Unsupervised Learning) คือ การเรียนรู้ที่ไม่ต้องมีการจับคู่สอน เมื่อใส่ค่าอินพุตเซตเข้าสู่โครงข่ายแล้วมันจะพยายามจัดกลุ่มอินพุตเซตที่มีลักษณะเดียวกันให้เอาท์พุตออกมาจากโครงข่ายที่เดียวกัน คือ การปรับตัวภายใน (Self –Organizing) การเรียนรู้โดยวิธีนี้จะป้อนอินพุตเข้าสู่โครงข่ายภายใน โครงข่ายจะมีเอาท์พุตโหนดอยู่หลายโหนด โดยแต่ละโหนดแทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนอินพุตเข้าสู่โครงข่าย โครงข่ายจะคำนวณหาค่าความสัมพันธ์ที่มีภายในเซตของอินพุตของโครงข่าย การเรียนรู้ คือ การใช้ค่าน้ำหนักเป็นตัวแยกความแตกต่างของอินพุต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 โครงสร้างของเซ็ลฟออร์แกไนซิงแม๊ป



รูปที่ 2.5 Two self-organized feature maps

เซ็ลฟออร์แกไนซิงแม๊ป ประกอบด้วยชั้นของเซลล์ประสาทที่แท้จริงเพียงชั้นเดียวคือ ชั้นฮิด-เดนเลเยอร์ เพราะเป็นชั้นที่นำค่าอินพุตเข้ามาคิดคำนวณ แล้วนำค่าที่ได้ส่งไปชั้นเอาต์พุตต่อไป ลักษณะของ Self-Organizing เป็นการปรับตัวภายใน การเรียนรู้โดยวิธีนี้จะป้อนอินพุตเข้าสู่เครือข่าย ภายในเครือข่ายจะมีเอาท์พุตอยู่หลายโหนด โดยแต่ละโหนดแทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนอินพุตเข้าสู่เครือข่าย เครือข่ายจะทำการคำนวณหาค่าความสัมพันธ์ที่มีภายในเซตของอินพุต จุดหนึ่งที่ไม่เหมือนกับการทำ Clustering ในแบบอื่นๆ ของ SOM ไม่ได้ใช้การเปรียบเทียบระหว่างข้อมูลด้วยกันเองแต่เป็นการเปรียบเทียบข้อมูลกับ neural output

2.3.3 ขั้นตอนการทำงานของเซ็ลฟออร์แกไนซิงแม๊ป

วิธีการคิดของ Kohonen's Self-Organizing Map คือการทำซ้ำของในแต่ละข้อมูลเพื่อที่จะได้ค่าน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการแบ่ง มีวิธีการคิดดังนี้

1. กำหนดค่าเริ่มต้นให้กับ weight vector วิธีการหนึ่งที่ใช้ในการกำหนดค่า weight vector คือการสุ่มจากข้อมูลเข้า
2. ทำการเลือกข้อมูล x เข้าด้วยการสุ่ม
3. เปรียบเทียบข้อมูลเข้า x เพื่อหาค่า winning neural โดยใช้ minimum-distance Euclidean เป็นดังสมการ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\forall i, \| x(t) - m_c(t) \| \leq \| x(t) - m_i(t) \| \qquad (2.1)$$

4. นำข้อมูลของกลุ่มที่มีค่าน้อยที่สุดมาคำนวณหาน้ำหนักใหม่ โดยใช้สมการ 2.2

$$m_i(t+1) = m_i(t) + h_{c(x),i}(x(t) - m_i(t)) \qquad (2.2)$$

โดย Neighborhood Function มักใช้สมการของ Gaussian ซึ่งเป็นดังสมการ 2.3

$$h_{c(x),i} = \alpha(t) \exp\left(- \frac{\| r_i - r_c \|^2}{2\sigma^2(t)} \right) \qquad (2.3)$$

ซึ่ง $0 < \alpha(t) < 1$ เป็นค่าแฟกเตอร์ หรือค่า Learning rate

5. เริ่มทำซ้ำตั้งแต่ข้อสองจนกระทั่งไม่มีความเปลี่ยนแปลงค่า distance ในแต่ละกลุ่ม

2.3.4 อัลกอริทึมในการพัฒนาระบบการแบ่งกลุ่มข้อมูลด้วยเซ็ลฟออร์แกไนซิงแม๊ป

1. Initialize random assign each a unique position in low-dimension grid (map)
2. Initialize α_0 and σ_0 to some values
3. While the number of iterations and user-specified limit **do**
4. Take random point x from the data set
5. Find the winner
$\forall i, \| x(t) - m_c(t) \| \leq \| x(t) - m_i(t) \|$
6. Update the node in map
$m_i(t+1) = m_i(t) + h_{c(x),i}(x(t) - m_i(t))$
7. Reduce the neighborhood radius and learning rate
8. End While

รูปที่ 2.6 อัลกอริทึมการทำงานของ SOM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิเคราะห์และออกแบบระบบ

ขั้นตอนการวิเคราะห์และออกแบบรูปแบบจำลองระบบ เพื่อนำมาพัฒนาโครงงานนี้เป็นการนำเทคโนโลยีการทำเหมืองข้อมูลใช้ในการประมวลผล ดังนั้นการออกแบบจะใช้รูปแบบจำลองจากยูเอ็มแอล เพื่อนำมาอธิบายขั้นตอนและเนื้อหาในการพัฒนาโปรแกรมมีรายละเอียดดังนี้

3.1 ความต้องการและข้อจำกัดของโปรแกรม

โครงการพัฒนาระบบงานนี้สามารถกำหนดความต้องการ และข้อจำกัดของโปรแกรมเพื่อใช้เป็นขอบเขตในการพัฒนาโครงการ ดังนี้

3.1.1 ความต้องการของผู้ใช้งานระบบ

- ระบบต้องสามารถแบ่งกลุ่มข้อมูลและรู้ได้ว่าข้อมูลตกอยู่ในกลุ่มใด
- ระบบต้องแสดงผลลัพธ์ของการทำการแบ่งกลุ่มให้อยู่ในรูปแบบที่เข้าใจง่าย

3.1.2 ความต้องการเชิงฟังก์ชัน (Functional Requirement)

- ระบบต้องสามารถเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมก่อนการทำดาต้าไนนิ่งได้
- ระบบต้องสามารถนำเข้าข้อมูลจากแหล่งเก็บไฟล์ภายในเครื่องไปยัง internal data structure ได้
- ระบบสามารถทำการสุ่มเพื่อกำหนดเป็นค่าเริ่มต้นให้กับ Weight Vector
- ระบบสามารถปรับเปลี่ยนตัวแปรของอัลกอริทึมที่ใช้ในโครงข่าย ได้แก่ (initial radius, map height, map width, initial learning rate, epsilon, number of iterations)
- ระบบสามารถทำการบันทึกโครงข่ายลงในแหล่งเก็บไฟล์ภายในเครื่องได้
- ระบบสามารถเปิดไฟล์โครงข่ายที่มีอยู่แล้วภายในเครื่องขึ้นมาใช้งานได้
- ระบบสามารถแสดงผลลัพธ์ในรูปแบบที่เข้าใจง่ายในลักษณะตาราง
- ระบบสามารถบันทึกผลลัพธ์จากการแบ่งกลุ่มลงในแหล่งเก็บไฟล์ในเครื่องได้

3.1.3 ความต้องการเชิงประสิทธิภาพ (Nonfunctional Requirements)

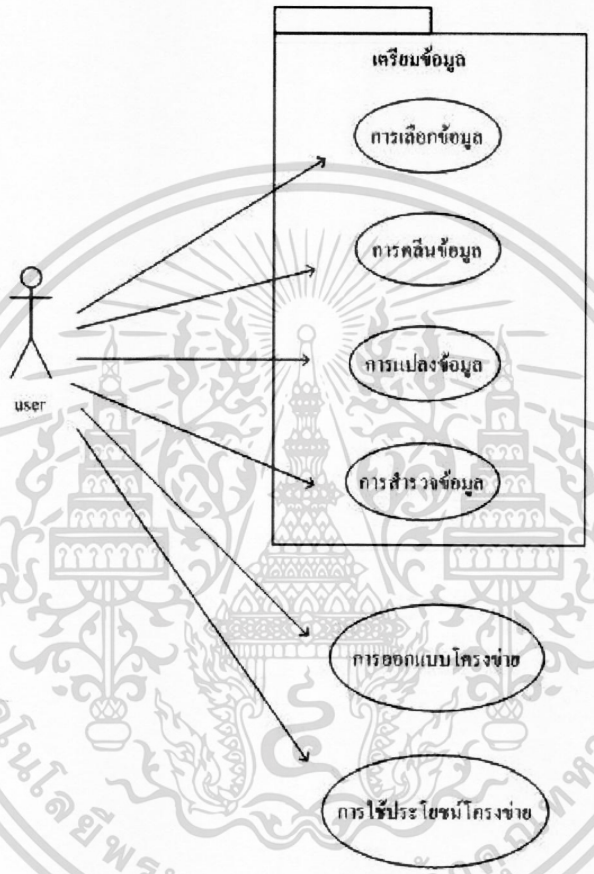
- ระบบต้องมีหน้าจอการใช้งาน (User Interface) ที่ใช้งานง่าย สื่อความหมาย
- ระบบต้องตอบสนองการทำงาน ณ ขั้นตอนต่างๆ ขณะประมวลผลได้
- ระบบบอกสถานะการทำงาน ณ ขั้นตอนต่างๆ ขณะประมวลผลได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การวิเคราะห์ระบบโดยใช้ยูสเคสวิวและแอคทิวิตีไดอะแกรม

3.2.1 ยูสเคสไดอะแกรม (Use Case Diagram)

ขอบเขตการพัฒนาระบบการแบ่งกลุ่มข้อมูลโดยใช้อัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ป นำมาอธิบายด้วยยูสเคสไดอะแกรมได้ดังนี้



รูปที่ 3.1 ยูสเคสไดอะแกรมระบบแบ่งกลุ่มข้อมูล

3.2.2 อธิบายแอคเตอร์และยูสเคส

จากรูปที่ 3.1 เป็นการใช้ยูสเคสไดอะแกรมแสดงให้เห็นถึงภาพรวมของการพัฒนาระบบโดยมีแอคเตอร์และยูสเคส ดังนี้

3.2.2.1 แอคเตอร์และบทบาทหน้าที่ของแอคเตอร์ที่ติดต่อกับโปรแกรม มีดังนี้

- ผู้ใช้งาน

เป็นบุคคลที่ป้อนอินพุตเข้าสู่ระบบ และเป็นผู้รับผลลัพธ์การแบ่งกลุ่มที่ได้จากระบบ

- ระบบฐานข้อมูล

เป็นระบบฐานข้อมูลที่ใช้ติดต่อกับฐานข้อมูล เพื่อนำข้อมูลเข้าสู่ระบบและบันทึกเป็น Temp Database สำหรับใช้ระหว่างการทำงานของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1.1 กิจกรรมหลักและรายละเอียดของกิจกรรมที่ทำงานภายในโปรแกรม ดังนี้

1. ระบบเตรียมข้อมูล

กิจกรรมหลัก คือ การติดต่อกับระบบฐานข้อมูล และทำการเตรียมข้อมูลให้เหมาะกับอัลกอริทึมที่ใช้ ซึ่งประกอบไปด้วยฟังก์ชัน การเลือกข้อมูล การคลีนข้อมูล การแปลงข้อมูล และการสำรวจข้อมูล ซึ่งมีรายละเอียดดังนี้

- **การเลือกข้อมูล**

กิจกรรมหลัก คือ การระบุแหล่งของข้อมูลที่ต้องการและทำการดึงเอาข้อมูลออกมาเพื่อแสดง และนำไปใช้สำหรับการวิเคราะห์เบื้องต้นเพื่อเตรียมข้อมูลสำหรับการทำดาต้าไมนิ่งในขั้นตอนต่อไป โดยผู้ใช้จะเลือก
แอตทริบิวต์ที่ต้องการใช้งาน เมื่อเสร็จสิ้นกระบวนการนี้จะได้ Temp Table ของข้อมูลที่เลือกไว้

- **การคลีนข้อมูล**

กิจกรรมหลัก คือ การกรองข้อมูลเพื่อให้ข้อมูลมีความถูกต้อง เลือกแอตทริบิวต์ที่ต้องการคลีนและประเภทของการคลีนข้อมูลที่ต้องการ

- **การแปลงข้อมูล**

กิจกรรมหลัก คือ การปรับขอบเขตของข้อมูลให้อยู่ในรูปแบบและช่วงข้อมูลที่เหมาะสมต่อการนำไปใช้ทำดาต้าไมนิ่ง

- **การสำรวจข้อมูล**

กิจกรรมหลัก คือ เป็นการอธิบายถึงการแสดงผลลัพธ์ข้อมูลที่ผ่านกระบวนการแปลงข้อมูลเรียบร้อยแล้วแสดงรายชื่อแอตทริบิวต์ทั้งหมดที่ได้จากการแปลงข้อมูล ซึ่งเป็นข้อมูลก่อนที่จะนำไปทำดาต้าไมนิ่ง

2. ออกแบบโครงข่าย

กิจกรรมหลัก คือ การกำหนดค่าต่างๆ ให้กับโครงข่าย ได้แก่
ความกว้างของโครงข่าย ความสูงของโครงข่าย อัตราการเรียนรู้ของโครงข่าย รัศมีเริ่มต้นของฟังก์ชั่น รวมทั้งจำนวนรอบสูงสุดของการฝึกสอนโครงข่าย เพื่อใช้ในการประมวลผลให้ได้ผลลัพธ์ที่ต้องการจากนั้นจึงทำการ ฝึกสอนโครงข่ายโดยใช้ข้อมูลที่เลือกไว้ การฝึกสอนโครงข่ายให้เรียนรู้จากข้อมูลการฝึกสอน จนเกิดเป็นโครงข่ายการแบ่งกลุ่มที่ต้องการ โดยใช้การฝึกสอนโครงข่ายด้วยวิธีของอัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ปและบันทึก Map ลงในเครื่อง

3. ใช้ประโยชน์โครงข่าย

กิจกรรมหลัก คือ การแบ่งกลุ่มข้อมูล โดยนำข้อมูลที่ต้องการแบ่งมาทำการแบ่งกลุ่มโดยใช้ Map ของโครงข่ายที่ผ่านการฝึกสอนมาแล้ว และแสดงผลลัพธ์ออกมาให้อยู่ในรูปแบบที่สื่อความหมายต่อผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3 รายละเอียดยูสเคส

3.2.3.1 รายละเอียดยูสเคสระบบเตรียมข้อมูล

- การเลือกข้อมูล

ตารางที่ 3.1 รายละเอียดยูสเคส การเลือกข้อมูล

ยูสเคส เลือกข้อมูล
รายละเอียด เป็นยูสเคสอธิบายถึงขั้นตอนการเลือกข้อมูลที่ต้องและติดต่อฐานข้อมูล โดยจะติดต่อกับ Microsoft SQL Server 2005
แพกเกจ เตรียมข้อมูล
แอคเตอร์ 1. ผู้ใช้งาน 2. ระบบฐานข้อมูล
เงื่อนไขก่อนเข้ายูสเคส -
ลำดับเหตุการณ์หลัก 1. ผู้ใช้งานระบุชื่อเครื่องเซิร์ฟเวอร์ 2. ผู้ใช้งานระบุชื่อฐานข้อมูล 3. ระบบทำการตรวจสอบฐานข้อมูลถูกต้อง 4. ระบบแสดงตารางฐานข้อมูล 5. ผู้ใช้งานเลือกตารางฐานข้อมูล 6. ระบบแสดงแอตทริบิวต์ของตารางฐานข้อมูล 7. ผู้ใช้งานเลือกแอตทริบิวต์ 8. ผู้ใช้งานคลิกปุ่ม Execute
ลำดับเหตุการณ์ย่อย 2a) ระบบทำการตรวจสอบฐานข้อมูลไม่ถูกต้อง : ระบบแสดงข้อความเตือนให้ระบุชื่อเครื่องเซิฟเวอร์และชื่อฐานข้อมูลใหม่
เงื่อนไขก่อนออกยูสเคส 1. ระบบทำการเลือกข้อมูลเรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- การคลีนข้อมูล

ตารางที่ 3.2 รายละเอียดยูสเคส การคลีนข้อมูล

ยูสเคส คลีนข้อมูล
รายละเอียด เป็นยูสเคสอธิบายถึงขั้นตอนการคลีนข้อมูล ซึ่งทำการแก้ไขปัญหาที่พบในข้อมูล เพื่อทำให้ข้อมูลมีคุณภาพก่อนที่จะนำข้อมูลไปประมวลผลในขั้นต่อไป
แพกเกจ เตรียมข้อมูล
แอคเตอร์ 1. ผู้ใช้งาน 2. ระบบฐานข้อมูล
เงื่อนไขก่อนเข้ายูสเคส 1. ต้องติดต่อฐานข้อมูลก่อน
ลำดับเหตุการณ์หลัก 1. ผู้ใช้งานเลือกแอตทริบิวต์ที่ต้องการคลีน 2. ผู้ใช้งานเลือกประเภทการคลีน 3. ระบบทำการคลีนข้อมูลถูกต้อง ลำดับเหตุการณ์ย่อย 3a) ระบบทำการคลีนข้อมูลไม่ถูกต้อง : ระบบแสดงข้อความเตือนระบบไม่สามารถคลีนข้อมูลได้
เงื่อนไขก่อนออกยูสเคส 1. ระบบทำการคลีนข้อมูลเรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

- การแปลงข้อมูล

ตารางที่ 3.3 รายละเอียดยูสเคส แปลงข้อมูล

ยูสเคส แปลงข้อมูล
รายละเอียด เป็นยูสเคสอธิบายถึงขั้นตอนการปรับเปลี่ยนข้อมูลตามประเภทของข้อมูลที่จะนำไปใช้ในการทำเหมืองข้อมูล
แพกเกจ เตรียมข้อมูล
แอคเตอร์ 1. ผู้ใช้งาน 2. ระบบฐานข้อมูล เงื่อนไขก่อนเข้ายูสเคส 1 ต้องผ่านการคลีนข้อมูลก่อน
ลำดับเหตุการณ์หลัก 1. ระบบแสดงแอตทริบิวต์ที่เลือกไว้ 2. ผู้ใช้ทำการแปลงข้อมูล วนซ้ำข้อ 2
ลำดับเหตุการณ์ย่อย 2a) ผู้ใช้ทำการแปลงข้อมูลแบบ Normalize : ผู้ใช้งานเลือกประเภทการ Normalize 2b) ผู้ใช้ทำการแปลงข้อมูลแบบ Construct New Attribute : ผู้ใช้งานระบุค่าและเลือกวิธีการคำนวณแอตทริบิวต์ใหม่ 2c) ผู้ใช้ทำการแปลงข้อมูลแบบ Numeric to Categorical : ผู้ใช้งานระบุค่าที่ต้องการแปลง 2d) ผู้ใช้ทำการแปลงข้อมูลแบบ Categorical to Numeric : ผู้ใช้งานเลือกประเภทการแปลง 2e) ผู้ใช้ทำการแปลงข้อมูลแบบ Grouping : ผู้ใช้งานเลือกข้อมูลและกำหนดชื่อกลุ่มของข้อมูล
เงื่อนไขก่อนออกยูสเคส 1. ระบบทำการแปลงข้อมูลเรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

04812

- การสำรวจข้อมูล

ตารางที่ 3.4 รายละเอียดยูสเคส สำรวจข้อมูล

ยูสเคส สำรวจข้อมูล
รายละเอียด เป็นยูสเคสที่อธิบายถึงการแสดงผลลัพธ์ข้อมูลที่ผ่านกระบวนการแปลงข้อมูล เรียบร้อยแล้ว โดยที่จะแสดงผลลัพธ์ในรูปแบบรายละเอียดและกราฟ
แพกเกจ เตรียมข้อมูล
แอคเตอร์
1. ผู้ใช้งาน 2. ระบบฐานข้อมูล
เงื่อนไขก่อนเข้ายูสเคส 1. ระบบต้องผ่านขั้นตอนการแปลงข้อมูลก่อน ลำดับเหตุการณ์หลัก 1. ผู้ใช้งานคลิกแท็บ Exploration 2. ระบบแสดงรายชื่อแอตทริบิวต์ทั้งหมดที่ได้จากการแปลงข้อมูล 3. ผู้ใช้เลือกแอตทริบิวต์ 4. ระบบแสดงรายละเอียดของแอตทริบิวต์ที่เลือก
ลำดับเหตุการณ์ย่อย -
เงื่อนไขก่อนออกยูสเคส 1. ระบบแสดงรายละเอียดของแอตทริบิวต์เรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3.2 รายละเอียดยูสเคสระบบแบ่งกลุ่มข้อมูล

- ออกแบบโครงข่าย

ตารางที่ 3.5 รายละเอียดยูสเคส ออกแบบโครงข่าย

ยูสเคส ออกแบบโครงข่าย
รายละเอียด เป็นยูสเคสที่อธิบายถึงการกำหนดค่าต่างๆ ให้กับโครงข่าย ได้แก่ ความกว้างของโครงข่าย ความสูงของโครงข่าย อัตราการเรียนรู้ของโครงข่าย รัศมีเริ่มต้นของฟังก์ชั่น รวมทั้งจำนวนรอบสูงสุดของการฝึกสอนโครงข่าย เพื่อใช้ในการประมวลผลให้ได้ผลลัพธ์ที่ต้องการจากนั้นจึงทำการฝึกสอนโครงข่ายโดยใช้ข้อมูลที่เลือกไว้ การฝึกสอนโครงข่ายให้เรียนรู้จากข้อมูลการฝึกสอน จนเกิดเป็นโครงข่ายการแบ่งกลุ่มที่ต้องการ
แอคเตอร์ 1. ผู้ใช้งาน
เงื่อนไขก่อนเข้ายูสเคส 1. ข้อมูลต้องผ่านขั้นตอนการเตรียมข้อมูลก่อน
ลำดับเหตุการณ์หลัก 1. ผู้ใช้งานคลิกแท็บ New Map 2. ผู้ใช้เลือกข้อมูลที่ต้องการใช้ฝึกสอนโครงข่าย 3. ระบบแสดงรายละเอียดทั่วไปของข้อมูล 4. ผู้ใช้กำหนดค่าเริ่มต้นในการสร้างโครงข่าย 5. ระบบแสดง Pattern ที่ได้จากการฝึก 6. ผู้ใช้ทำการบันทึก Model
ลำดับเหตุการณ์ย่อย -
เงื่อนไขก่อนออกยูสเคส 3a) ระบบตรวจสอบว่ามีการนำข้อมูลเข้าแล้ว หากยังไม่มีข้อมูลจะแจ้งเตือนข้อผิดพลาด 4a) ผู้ใช้กำหนดค่าเริ่มต้นไม่ถูกต้อง ระบบจะแจ้งเตือนข้อผิดพลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- การใช้ประโยชน์โครงข่าย

ตารางที่ 3.6 รายละเอียดยูสเคส การใช้ประโยชน์โครงข่าย

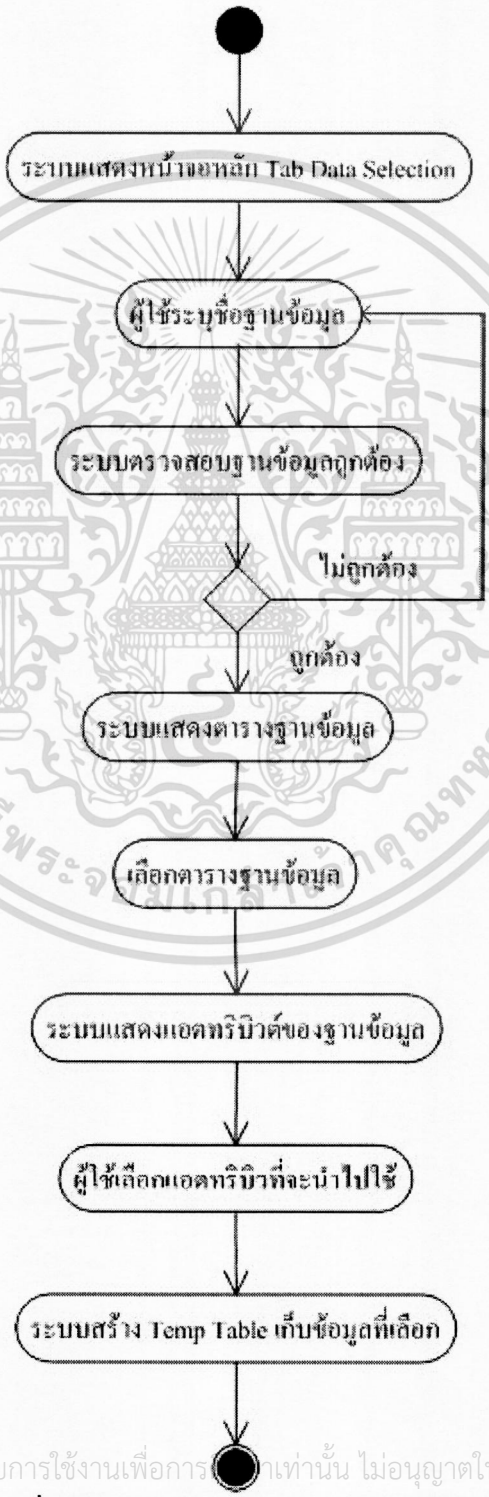
ยูสเคส การใช้ประโยชน์โครงข่าย
รายละเอียด เป็นยูสเคสที่อธิบายถึงการแบ่งกลุ่มข้อมูล โดยนำข้อมูลที่ต้องการแบ่งมาทำการแบ่งกลุ่มโดยใช้ Pattern ของโครงข่ายที่ผ่านการฝึกสอนมาแล้ว และแสดงผลลัพธ์ออกมาให้อยู่ในรูปแบบที่สื่อความหมายต่อผู้ใช้
แอคเตอร์ 1. ผู้ใช้งาน
เงื่อนไขก่อนเข้ายูสเคส 1. ระบบต้องมี Pattern ที่ใช้สำหรับแบ่งกลุ่มข้อมูล
ลำดับเหตุการณ์หลัก 1. ผู้ใช้งานคลิกแท็บ Deployment map 2. ผู้ใช้เลือกข้อมูลที่ต้องการแบ่งกลุ่ม 3. ระบบแสดงรายละเอียดทั่วไปของข้อมูล 4. ผู้ใช้เลือก Pattern ที่ต้องการนำมาใช้ในการแบ่งกลุ่ม 5. ระบบโหลด Model และแสดงรายละเอียด Model ที่เลือก 6. ผู้ใช้กดปุ่มเริ่มต้นการแบ่งกลุ่ม 7. ระบบแสดงผลลัพธ์การแบ่งกลุ่มที่สื่อความหมาย 6. ผู้ใช้ทำการบันทึกผลลัพธ์
ลำดับเหตุการณ์ย่อย -
เงื่อนไขก่อนออกยูสเคส 3a) ระบบตรวจสอบว่ามีการนำข้อมูลเข้าแล้ว หากยังไม่มีข้อมูลจะแจ้งเตือนข้อผิดพลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.4 แอคทิวิตีไดอะแกรม

ขั้นตอนของกิจกรรมที่เกิดขึ้นในการพัฒนาระบบสามารถอธิบายได้ด้วยแอคทิวิตีไดอะแกรมดังรูปต่อไปนี้

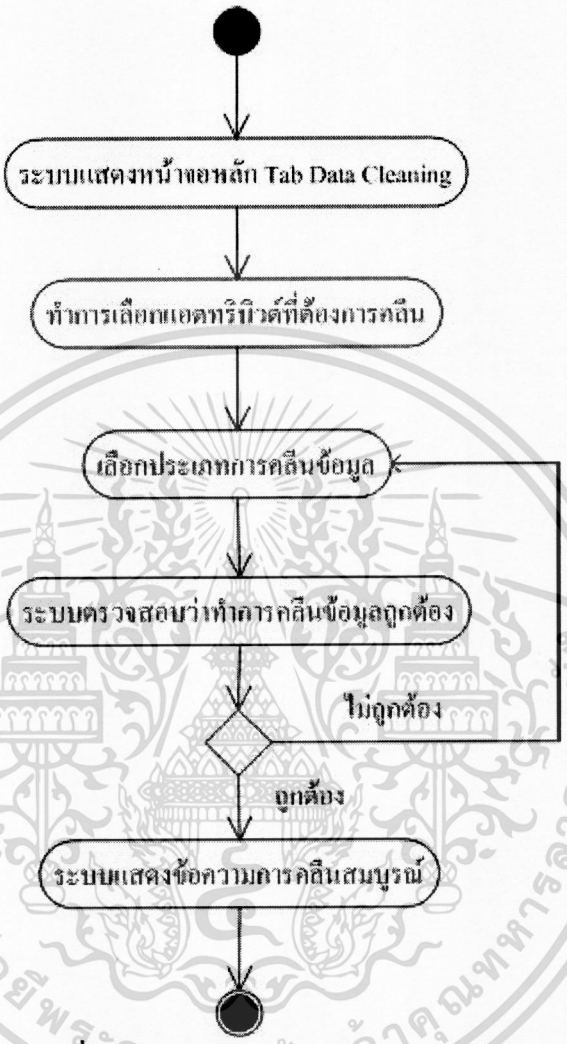
- แอคทิวิตีไดอะแกรมการเลือกข้อมูล แสดงขั้นตอนการเลือกข้อมูลจากแหล่งข้อมูลที่ต้องการ เพื่อนำไปใช้เป็นข้อมูลเข้าในการทำดาต้าไมนิ่ง



รูปที่ 3.2 แอคทิวิตีไดอะแกรมการเลือกข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารทุกครั้งที่มีการนำไปใช้

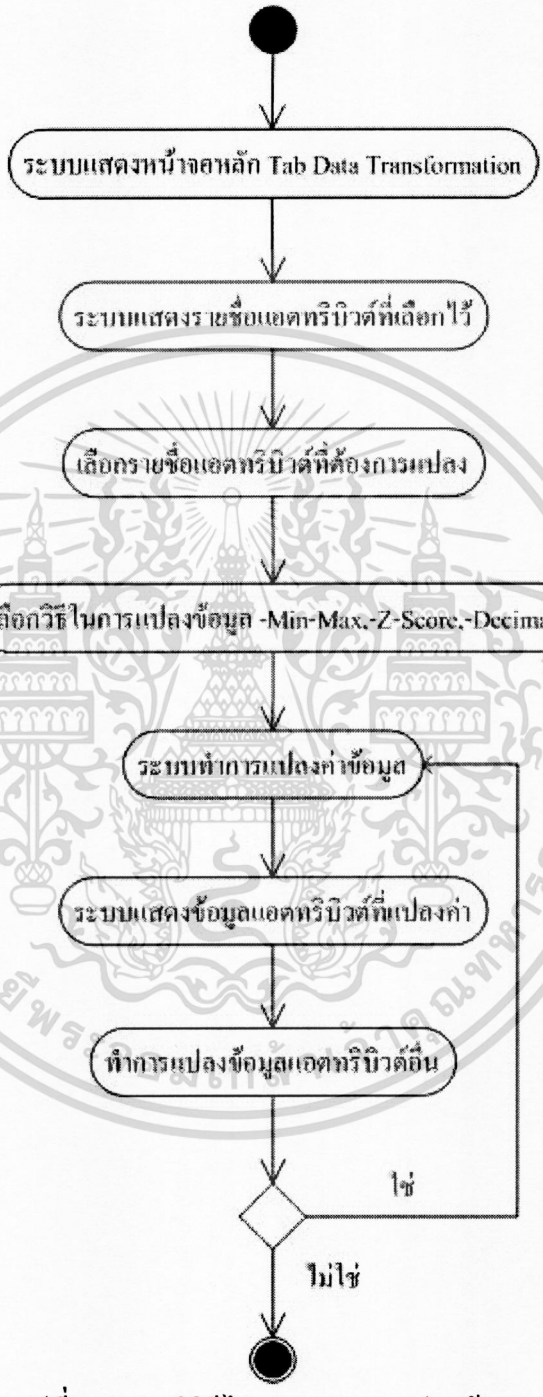
- แอคทิวิตีไดอะแกรมการคลีนข้อมูล แสดงขั้นตอนการคลีนข้อมูลซึ่งเป็นขั้นตอนหนึ่งในการเตรียมข้อมูล



รูปที่ 3.3 แอคทิวิตีไดอะแกรมการคลีนข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- แอคทิวิตีไดอะแกรมการแปลงข้อมูล แสดงขั้นตอนการแปลงข้อมูล เพื่อให้อยู่ในรูปแบบที่เหมาะสมกับอัลกอริทึมที่เลือกใช้ประมวลผลในขั้นตอนการทำดาต้าไมนิ่ง



รูปที่ 3.4 แอคทิวิตีไดอะแกรมการแปลงข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- แอคทิวิตีไดอะแกรมการสำรวจข้อมูล แสดงการสำรวจข้อมูลเพื่อดูแนวโน้มการกระจายตัวของข้อมูลในแต่ละช่วง

รูปที่ 3.5 แอคทิวิตีไดอะแกรมการสำรวจข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- แอคทิวิตีไดอะแกรมการออกแบบโครงข่าย แสดงขั้นตอนการออกแบบโครงข่ายที่ใช้ในการฝึกสอน กิจกรรมหลักเป็นการกำหนดค่าตัวแปรต่างๆ ในการสร้างโครงข่ายที่จะนำไปใช้ในการแบ่งกลุ่มข้อมูล

รูปที่ 3.6 แอคทิวิตีไดอะแกรมสร้างโครงข่ายใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

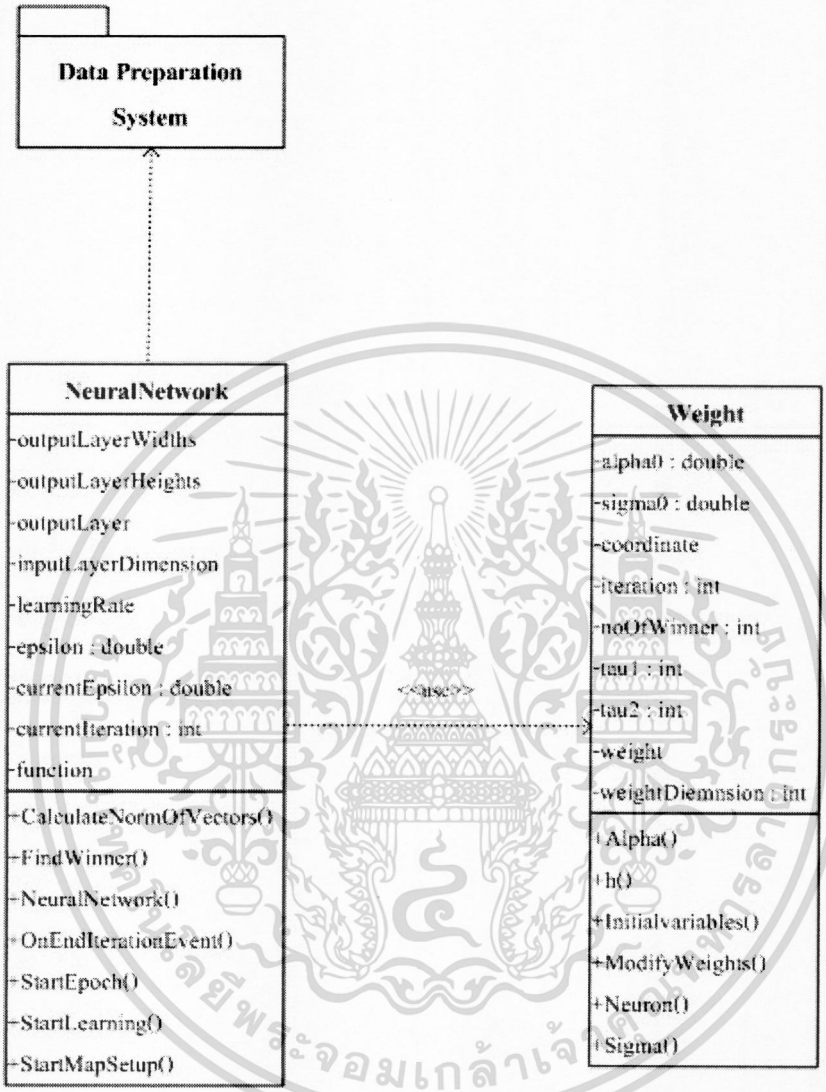
- แอคทิวิตีไดอะแกรมการใช้ประโยชน์โครงข่าย แสดงขั้นตอนในการแบ่งกลุ่มข้อมูลด้วยอัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ป เป็นการใช้ประโยชน์โครงข่ายที่ผ่านการฝึกสอนแล้ว

รูปที่ 3.7 แอคทิวิตีไดอะแกรมการใช้งานโครงข่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.5 การออกแบบระบบโดยใช้คลาสไดอะแกรม

รูปแสดงการทำงานของคลาสต่างๆ ที่มีในระบบ โดยมีรายละเอียดดังรูปต่อไปนี้



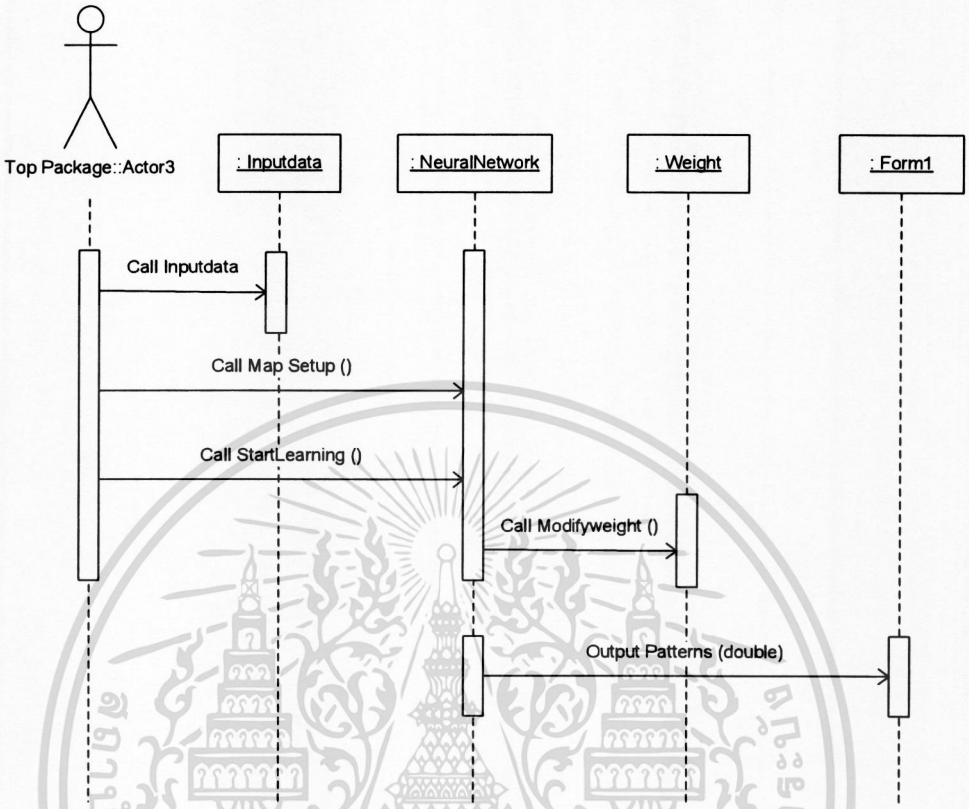
รูปที่ 3.8 คลาสไดอะแกรมของระบบ

คลาสไดอะแกรมของระบบประกอบด้วย

- เพคเกจ Data Preparation System ทำหน้าที่ในการเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมก่อนการทำดาต้าไมนิ่ง
- คลาส Neural Network ทำหน้าที่ในการสร้างโครงข่ายเพื่อใช้ในแบ่งกลุ่มข้อมูล
- คลาส Weight ทำหน้าที่ในการปรับค่า weight ให้กับโครงข่าย

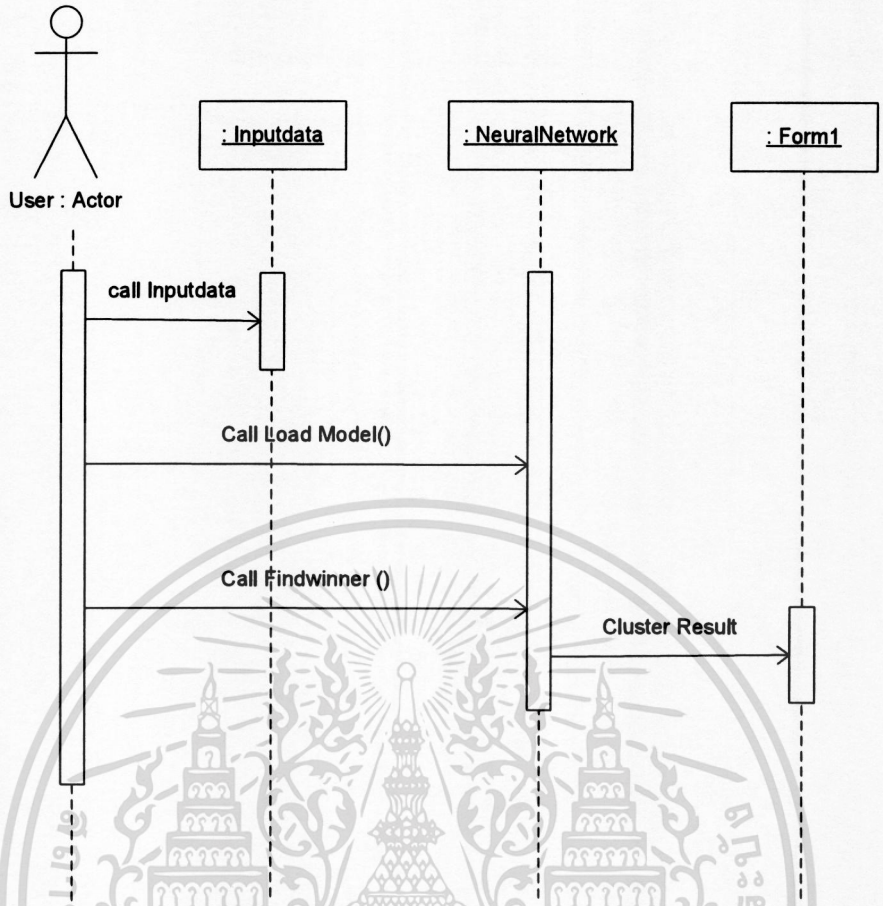
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.6 การออกแบบระบบโดยใช้ซีเควนไดอะแกรม



รูปที่ 3.9 ซีเควนไดอะแกรมการสร้างโครงข่ายใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 ซีเควนไดอะแกรม การแบ่งกลุ่มด้วย SOM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ออกแบบและพัฒนาระบบ

4.1 เครื่องมือที่ใช้ในการออกแบบและพัฒนาระบบ

4.1.1 ฮาร์ดแวร์

การพัฒนาระบบใช้เครื่องคอมพิวเตอร์ที่มีคุณสมบัติดังนี้

- หน่วยประมวลผลกลาง Intel core2duo ความเร็ว 1.8 GHz
- ฮาร์ดดิสก์ความจุ 120 GB
- หน่วยความจำ 2 GB

4.1.2 ซอร์ฟแวร์

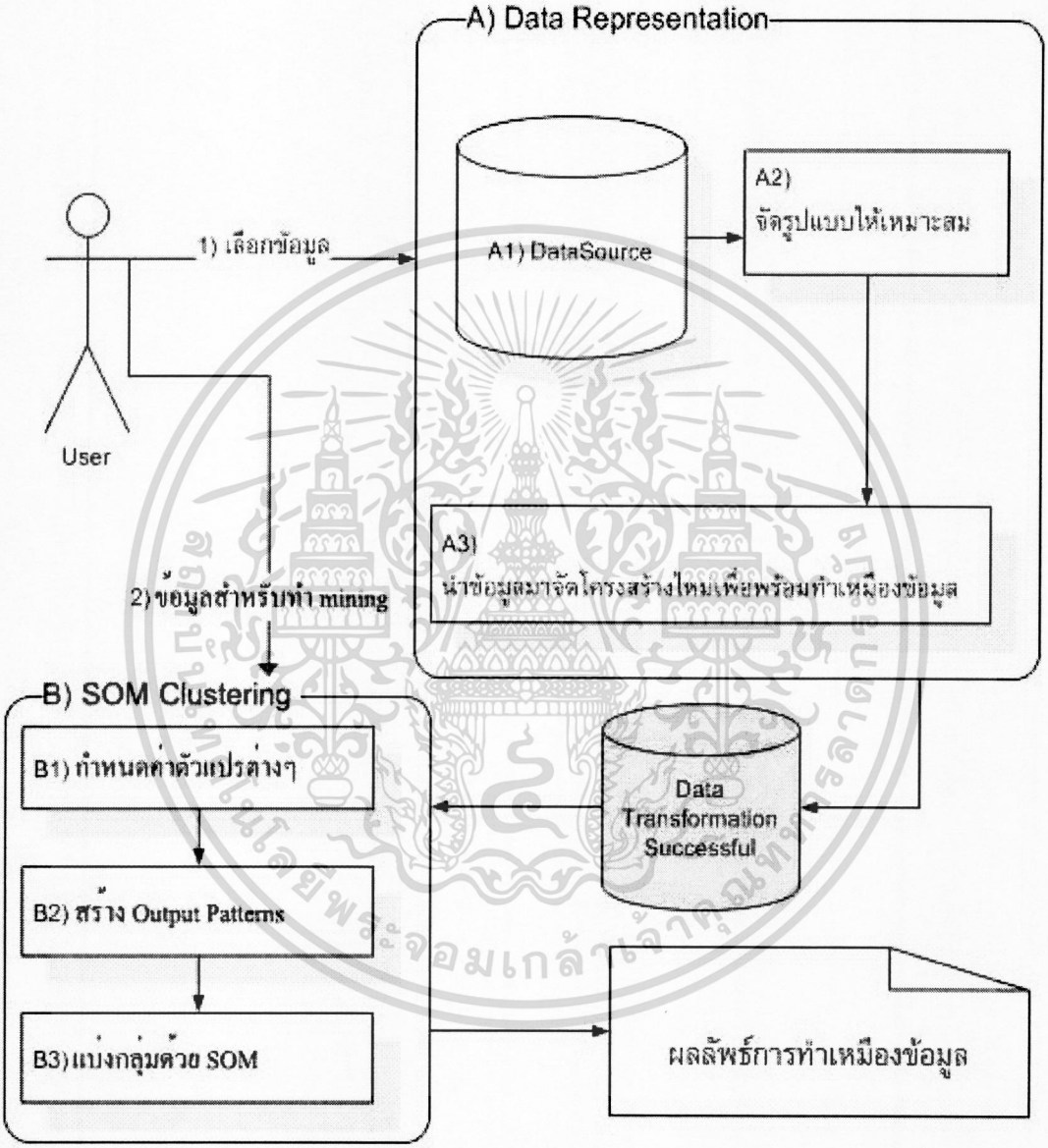
เครื่องมือและโปรแกรมที่ใช้ในการพัฒนามีรายละเอียดดังนี้

- ระบบปฏิบัติการ คือ Windows XP SP2
- ภาษาที่ใช้ในการพัฒนา คือ C# (.Net Framework 2.0)
- เครื่องมือที่ใช้ในการพัฒนาคือ Microsoft SQL Server 2005 Developer Edition
- การสร้างไฟล์จากส่วนการวิเคราะห์และออกแบบ คือ Microsoft Visio 2003

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ขั้นตอนการทำงานของระบบ

ในการออกแบบขั้นตอนการทำงานของระบบได้แบ่งออกเป็น 2 ส่วนใหญ่ๆ คือ Data Representation และ SOM Clustering โดยทั้ง 2 ส่วนมีความต่อเนื่องกัน ดังรูป



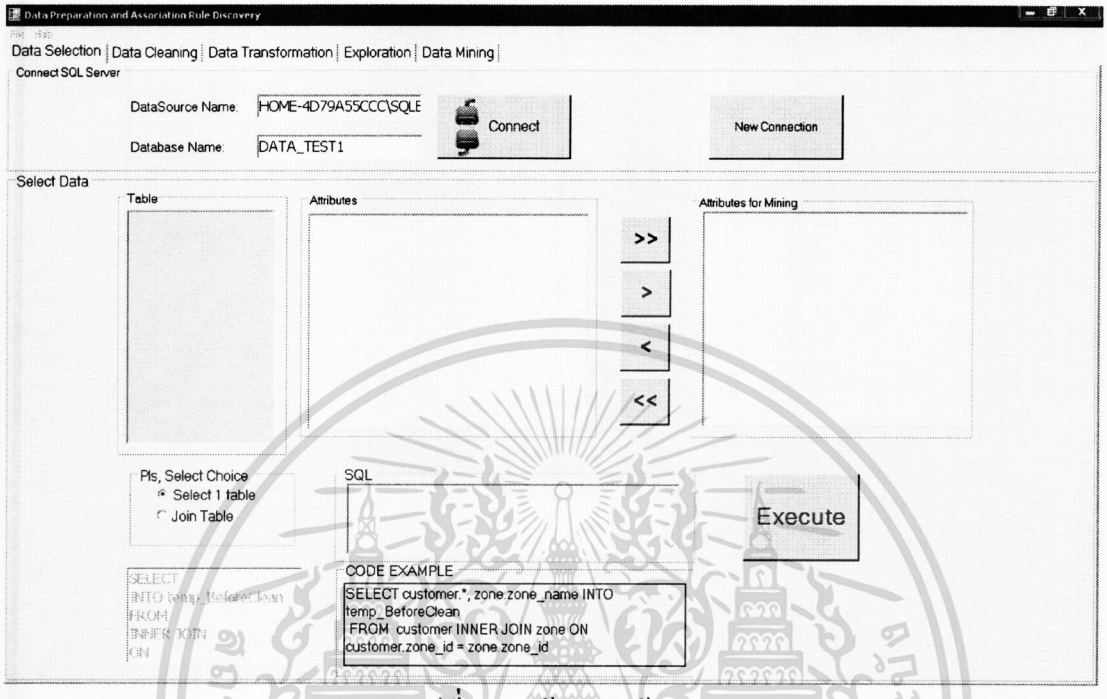
รูปที่ 4.1 ขั้นตอนการทำงานของระบบการแบ่งกลุ่มด้วย SOM

จากรูปที่ 4.1 เป็นการออกแบบสถาปัตยกรรมของระบบ โดยผู้ใช้งานทำการเลือกข้อมูลจากฐานข้อมูลเพื่อนำข้อมูลไปจัดรูปแบบ และจัดโครงสร้างใหม่ให้เหมาะสม เพื่อพร้อมใช้ในการทำเหมืองข้อมูล ซึ่งผู้ใช้งานจะสั่งให้กระบวนการประมวลผลการทำเหมืองข้อมูลเริ่มทำงาน โดยใช้ข้อมูลที่จัดเตรียมไว้แล้ว เมื่อระบบทำงานเสร็จสิ้นก็จะได้ผลลัพธ์การทำเหมืองข้อมูลออกมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

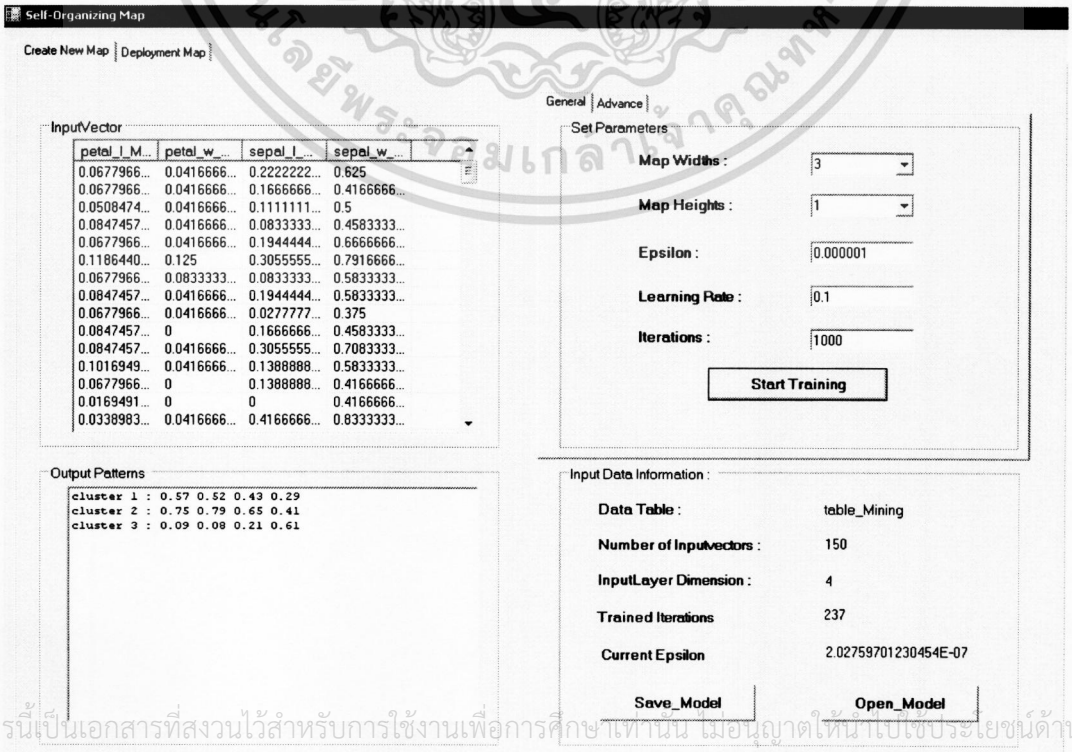
4.3 การออกแบบหน้าจอการทำงานของระบบ

4.3.1 หน้าจอรับข้อมูล



รูปที่ 4.2 หน้าจอรับข้อมูล

4.3.2 หน้าจอระบบแบ่งกลุ่มด้วย SOM



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

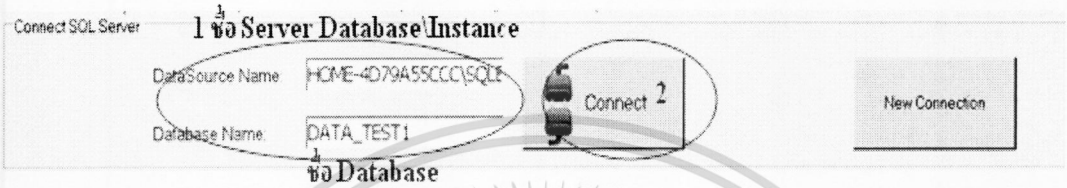
รูปที่ 4.3 หน้าจอระบบแบ่งกลุ่มด้วย SOM

4.4 การใช้งานโปรแกรม

การทำงานของระบบจะเป็นไปตามกระบวนการของการทำเหมืองข้อมูล ดังนี้

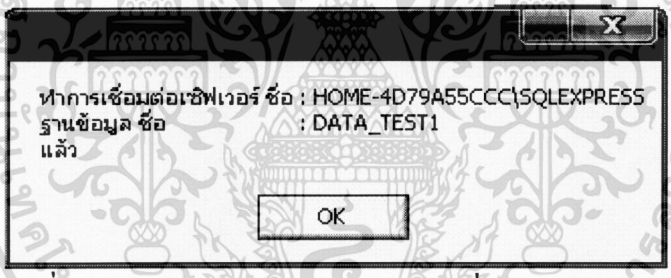
4.4.1 ขั้นตอนการติดต่อฐานข้อมูล

1. ผู้ใช้งานกำหนด ชื่อเซิร์ฟเวอร์ และชื่อฐานข้อมูล
2. คลิกปุ่ม Connect เพื่อเชื่อมต่อกับฐานข้อมูล



รูปที่ 4.4 ขั้นตอนการติดต่อฐานข้อมูล

3. ระบบแสดงกล่องข้อความให้ทราบว่าได้ทำการเชื่อมต่อกับฐานข้อมูลเรียบร้อย



รูปที่ 4.5 กล่องข้อความแสดงการผลการเชื่อมต่อฐานข้อมูล

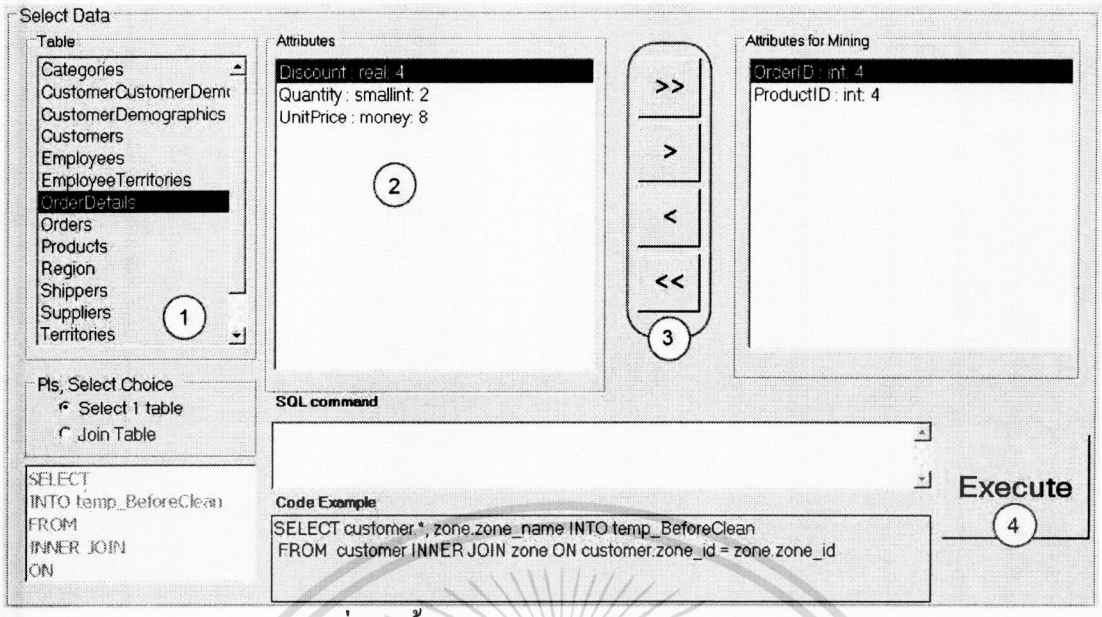
4.4.2 ขั้นตอนการเลือกข้อมูล

ในการขั้นตอนการเลือกข้อมูลสามารถเลือกข้อมูลได้ 2 กรณีคือ กรณีที่ข้อมูลมาจาก 1 ตารางและกรณีที่ข้อมูลมาจากมากกว่า 1 ตารางในฐานข้อมูลเดียวกัน โดยทั้ง 2 กรณีมีวิธีเลือกดังนี้

กรณีที่1 เลือกข้อมูลจาก 1 ตาราง

1. ผู้ใช้งานคลิกเลือกตาราง
2. ผู้ใช้งานเลือกชื่อแอตทริบิวที่จะใช้ในการ
3. ผู้ใช้งานคลิกปุ่มลูกศรเพื่อเลือกชื่อแอตทริบิวที่ได้เลือกไว้
4. คลิกปุ่ม Execute เพื่อสิ้นสุดขั้นตอนการเลือกข้อมูล

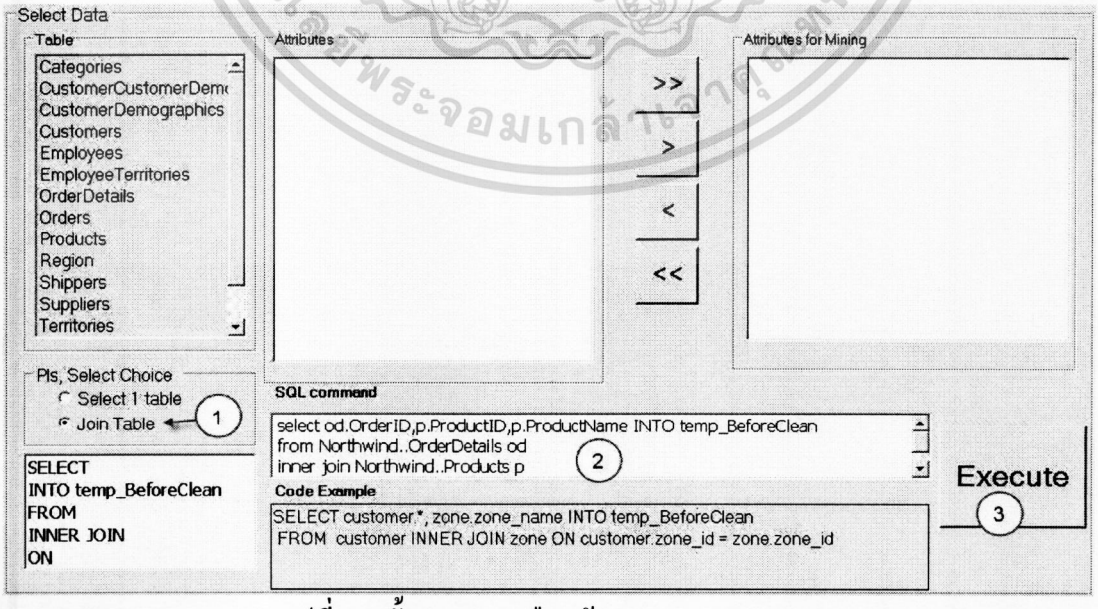
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 ขั้นตอนการเลือกข้อมูลจาก 1 ตาราง

กรณีที่2 เลือกข้อมูลจากหลายตาราง

1. ผู้ใช้งานเลือก Join Table ในช่อง Pls, Select Choice
2. ผู้ใช้งานพิมพ์คำสั่งของ SQL ลงในช่อง SQL Command เพื่อเลือกข้อมูลที่ต้องการ โดยผู้ใช้ระบบต้องทราบความสัมพันธ์ของข้อมูลในแต่ละตาราง และสามารถใช้คำสั่ง SQL พื้นฐานในการเชื่อมความสัมพันธ์เหล่านั้น เพื่อเลือกข้อมูลในการทำเหมืองข้อมูล
3. ผู้ใช้งานคลิกปุ่ม Execute เพื่อสิ้นสุดขั้นตอนการเลือกข้อมูล



รูปที่ 4.7 ขั้นตอนการเลือกข้อมูลจากหลายตาราง

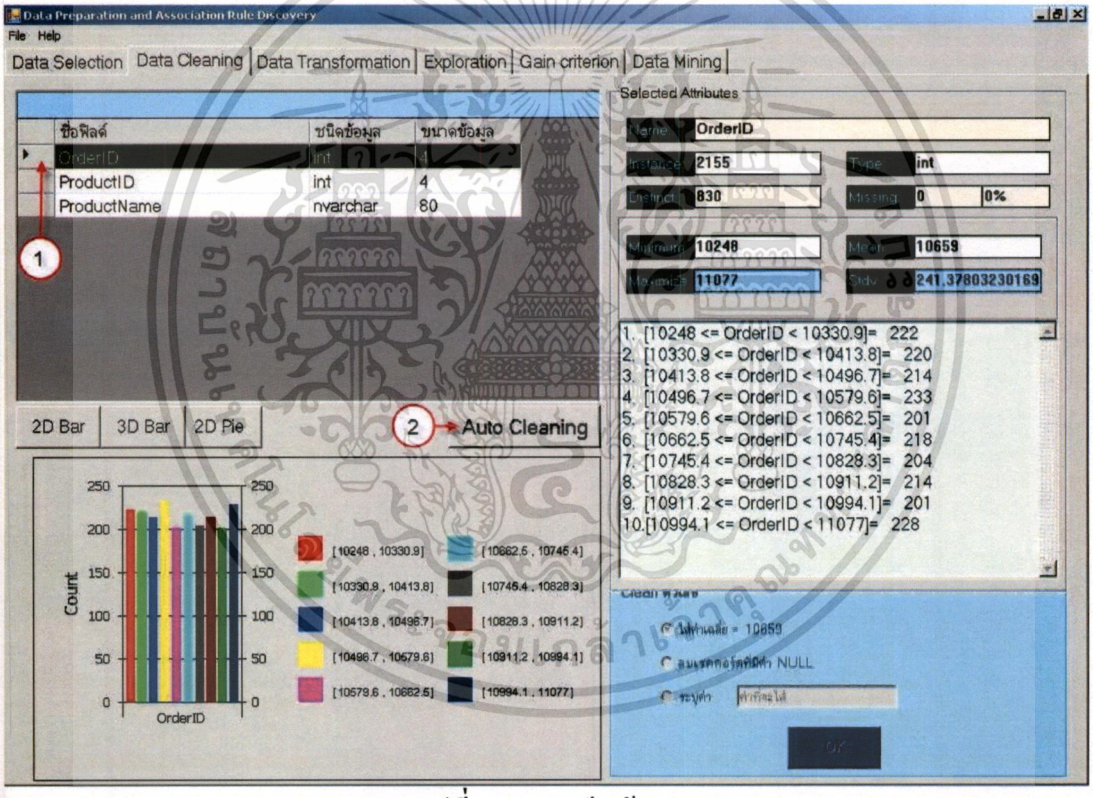
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ระบบแสดงกล่องข้อความบอกให้ผู้ใช้งานทราบว่าได้เสร็จสิ้นขั้นตอนการเลือกข้อมูลแล้ว

รูปที่ 4.8 กล่องข้อความแสดงผลการเลือกข้อมูล

4.4.3 ขั้นตอนการคลีนข้อมูล

1. ผู้ใช้งานคลิกเลือกแอตทริบิวต์ โดยระบบจะแสดงรายละเอียดของแอตทริบิวต์และแสดงกราฟของจำนวนข้อมูลในแอตทริบิวต์ที่ผู้ใช้เลือก
2. ผู้ใช้งานคลิกปุ่ม Auto Cleaning เพื่อทำการคลีนข้อมูลที่ได้เลือกไว้



รูปที่ 4.9 การคลีนข้อมูล

3. ระบบแสดงข้อความให้ทราบว่าทำการคลีนข้อมูลเรียบร้อย

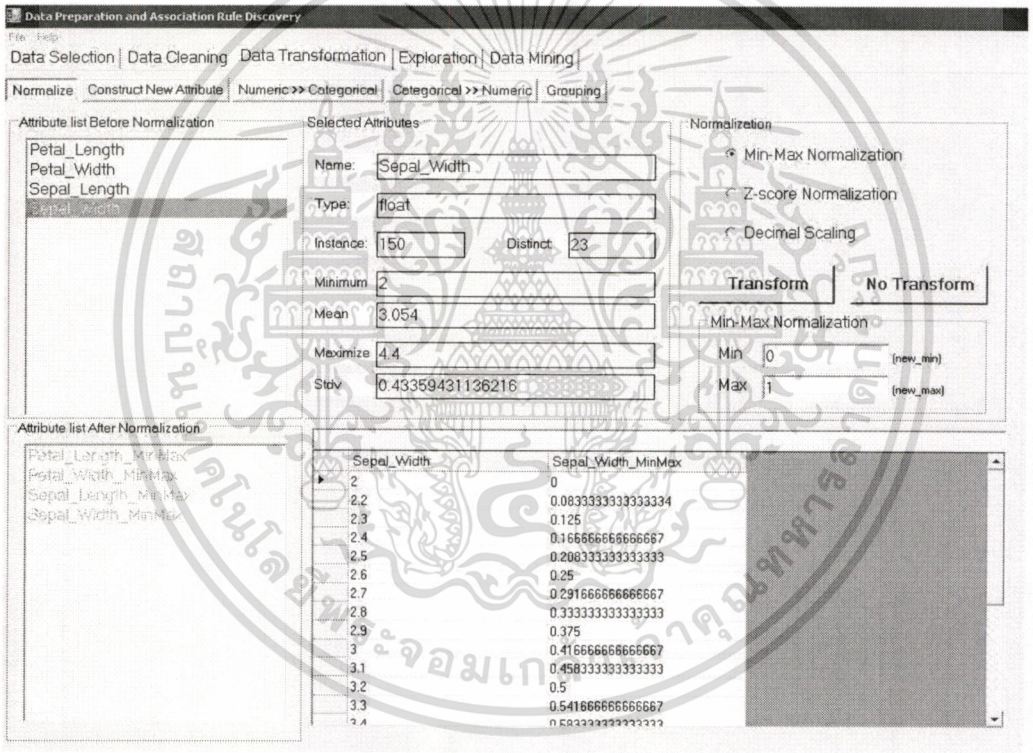
รูปที่ 4.10 กล่องข้อความแสดงผลการคลีนข้อมูลเสร็จเรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.4 ขั้นตอนการแปลงข้อมูล

ในขั้นตอนการแปลงข้อมูลสำหรับการทำเหมืองข้อมูลโดยใช้แบบจำลองของ SOM ในที่นี้จะกล่าวถึงกระบวนแปลงข้อมูล โดยมีขั้นตอนดังนี้

1. ผู้ใช้งานคลิกปุ่ม Normalize เพื่อเลือกประเภทการแปลงข้อมูล
2. ผู้ใช้งานเลือกแอตทริบิวต์ที่ต้องการจะนำมาแปลงข้อมูล
3. ผู้ใช้งานเลือกวิธีการแปลงข้อมูลในที่นี้เลือกเป็น Min-Max Normalization
4. ผู้ใช้งานคลิกปุ่ม Transformation เพื่อทำการแปลงข้อมูล
5. ผู้ใช้เลือกแอตทริบิวต์ที่ต้องการและทำซ้ำตั้งแต่ขั้นตอนที่ 2 จนกว่าจะได้ข้อมูลตามที่ต้องการครบ

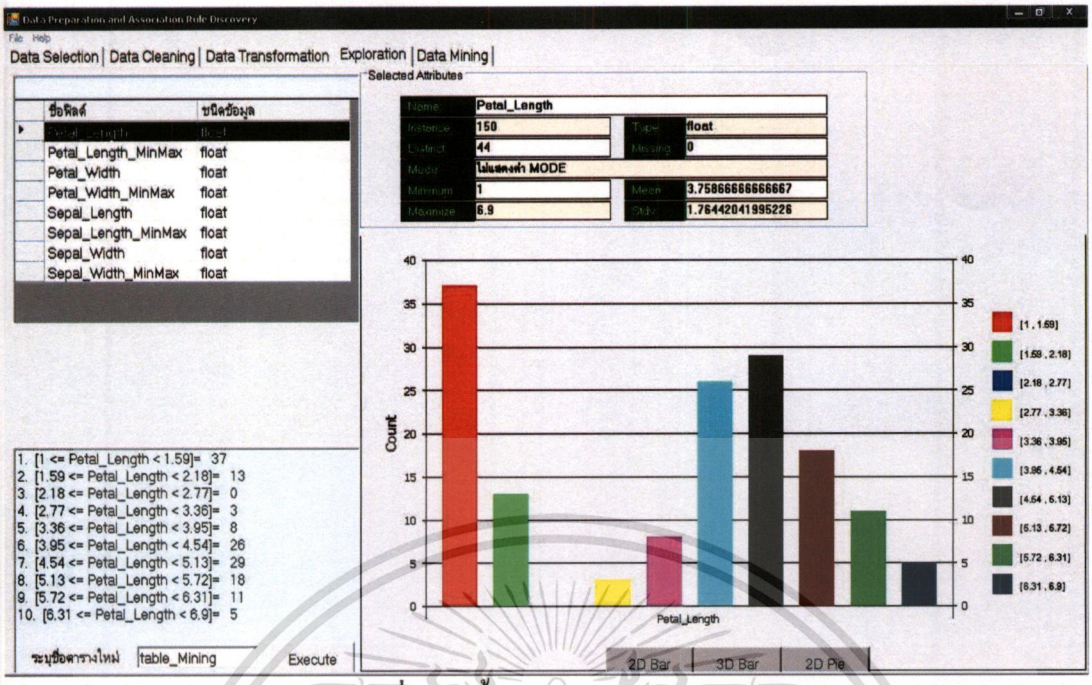


รูปที่ 4.11 ขั้นตอนการแปลงข้อมูล

4.4.5 ขั้นตอนการสำรวจข้อมูล

1. ผู้ใช้เลือกรายชื่อแอตทริบิวต์ที่ต้องการดูคุณสมบัติต่างๆ เช่น การกระจายตัวของข้อมูล ค่า Min-Max
2. ผู้ใช้ตั้งชื่อตารางข้อมูลชั่วคราวไว้ใช้ในกระบวนการทำดาต้าไมนิ่งในที่นี้ตั้งชื่อว่า table_Mining
3. คลิกปุ่ม Execute เพื่อทำการสร้างตารางข้อมูลสำหรับทำดาต้าไมนิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 ขั้นตอนการสำรวจข้อมูล

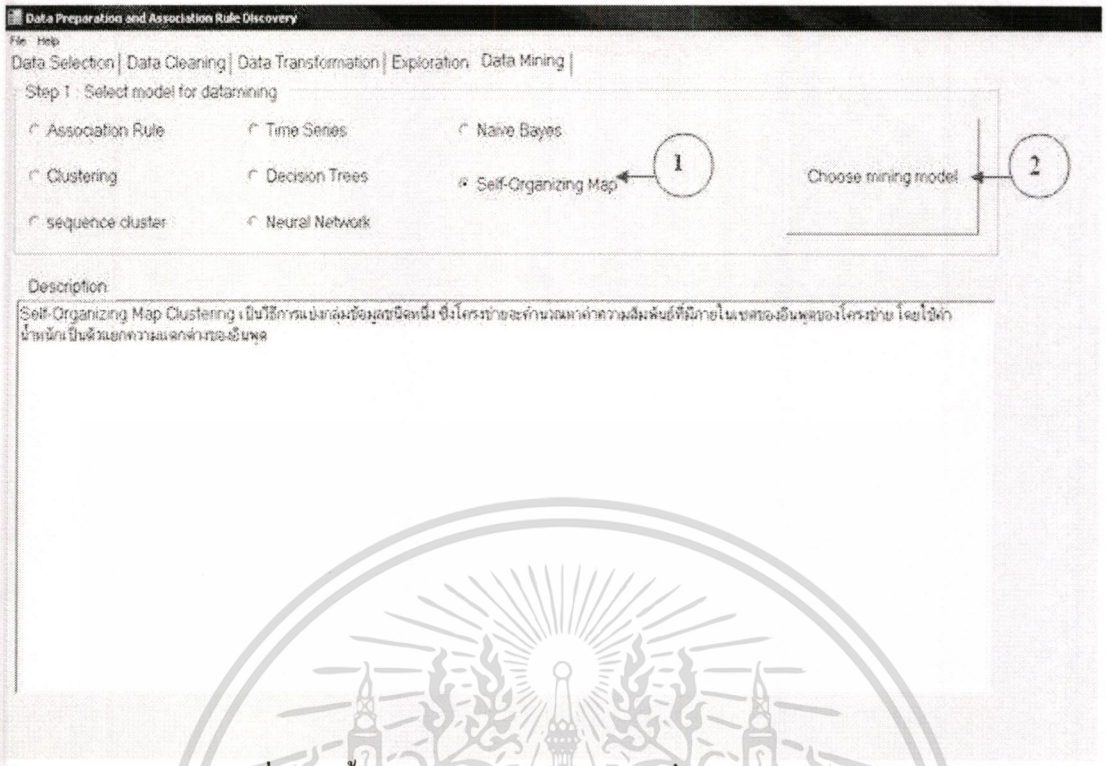
9. ระบบแสดงกล่องแสดงข้อความให้ผู้ใช้ทราบว่าทำสร้างตารางข้อมูลสำหรับการทำดาต้าไมนิ่งเรียบร้อย

รูปที่ 4.13 กล่องข้อความแสดงผลการสร้างตารางข้อมูลสำหรับการทำไมนิ่ง

4.4.6 ขั้นตอนการทำดาต้าไมนิ่ง

1. หลังจากจบขั้นตอนการแปลงข้อมูลแล้วผู้ใช้งานทำการคลิกเลือกที่แท็บ Data Mining เพื่อเลือกแบบจำลองในการทำเหมืองข้อมูล ซึ่งในที่นี้จะเลือกเซ็ลฟออร์แกไนซิงแม็ป
2. ผู้ใช้งานคลิกปุ่ม Choose mining model เพื่อยืนยันการเลือกแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

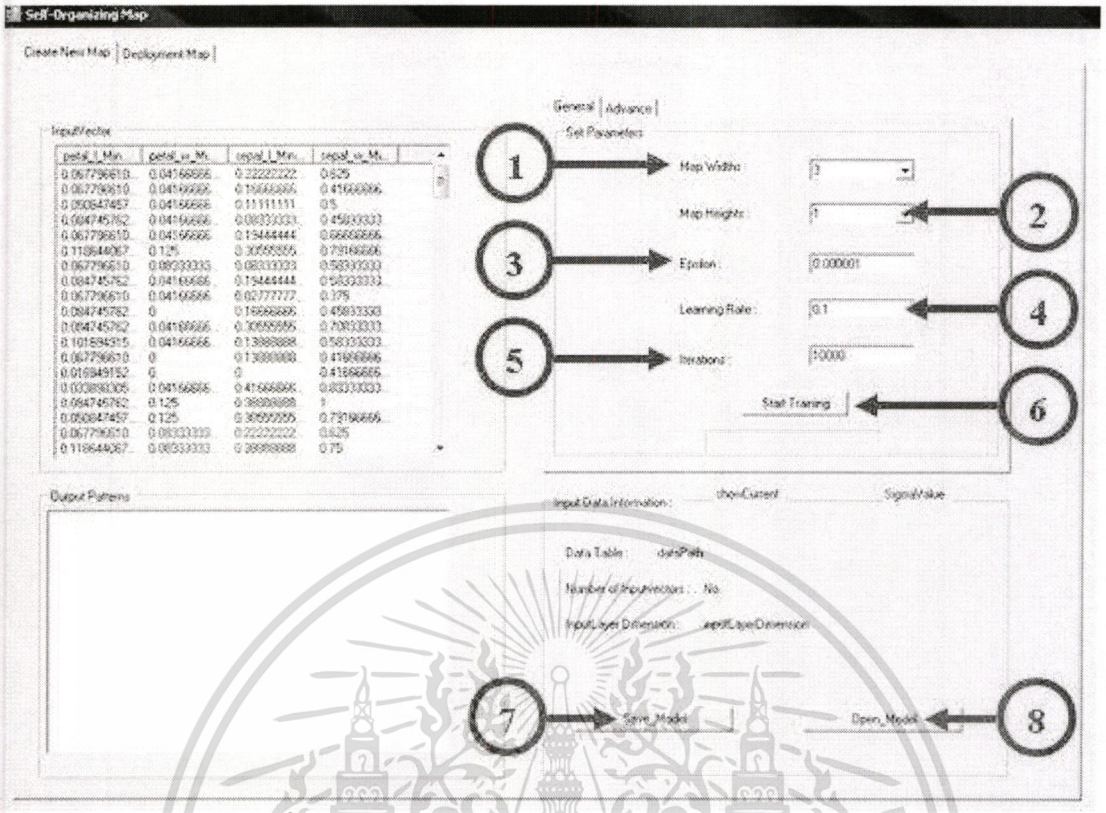


รูปที่ 4.14 ขั้นตอนการเลือกแบบจำลองที่ใช้ทำเหมืองข้อมูล

4.4.6.1 ขั้นตอนการสร้างโครงข่ายใหม่แบ่งออกเป็นสองแท็บย่อย

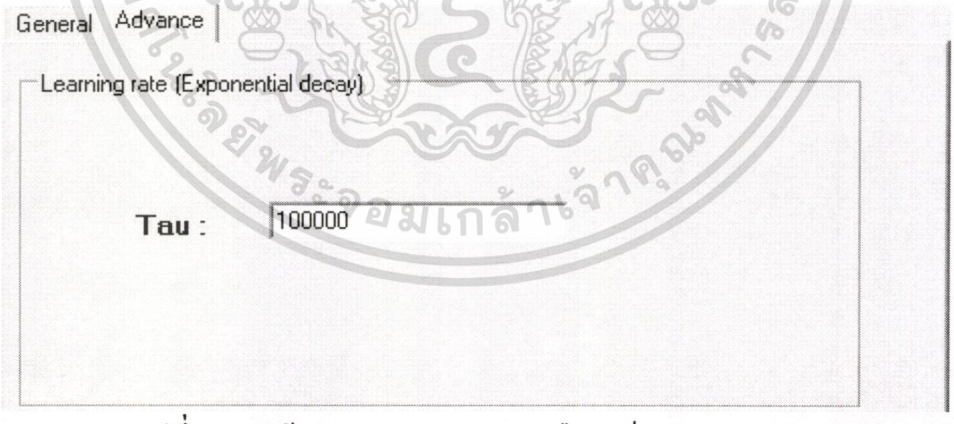
1. แท็บ General ไว้กำหนดค่าตัวแปรต่างๆ เพื่อเริ่มต้นสร้าง map
   1. ผู้ใช้กำหนดค่า Map Widths
   2. ผู้ใช้กำหนดค่า Map Heights
   3. ผู้ใช้กำหนดค่า Epsilon
   4. ผู้ใช้กำหนดค่า Learning Rate
   5. ผู้ใช้กำหนดค่า Iterations
   6. คลิกปุ่ม Start Training เพื่อเริ่มการฝึกสอนโครงข่าย
   7. เมื่อเสร็จสิ้นการฝึกสอนผู้ใช้คลิกปุ่ม Save Model เพื่อบันทึกข้อมูลของแบบจำลองให้อยู่ในรูปไฟล์ XML ซึ่งเก็บ ค่าตัวแปรต่างๆ ที่กำหนดตอนเริ่มต้น และค่า Pattern เก็บไว้
   8. เมื่อต้องการเปิดแบบจำลองที่เคยฝึกสอนขึ้นมาใช้ผู้ใช้คลิกปุ่ม Open Model เพื่อโหลดข้อมูล ค่าตัวแปรต่างๆ และ Pattern ขึ้นมาใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

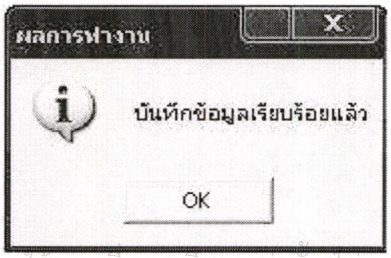


รูปที่ 4.15 หน้าจอ Create New Map เลือกแท็บ General

2. แท็บ Advance ไว้กำหนดค่า Tau ซึ่งเป็นค่าที่ใช้กำหนดการลดของค่า Learning rate

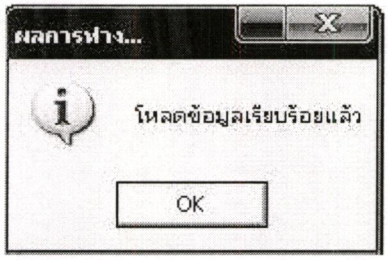


รูปที่ 4.16 หน้าจอ Create New Map เลือกแท็บ Advance



รูปที่ 4.17 กล่องข้อความแสดงหลังจากทำการบันทึกข้อมูล Model เรียบร้อยแล้ว

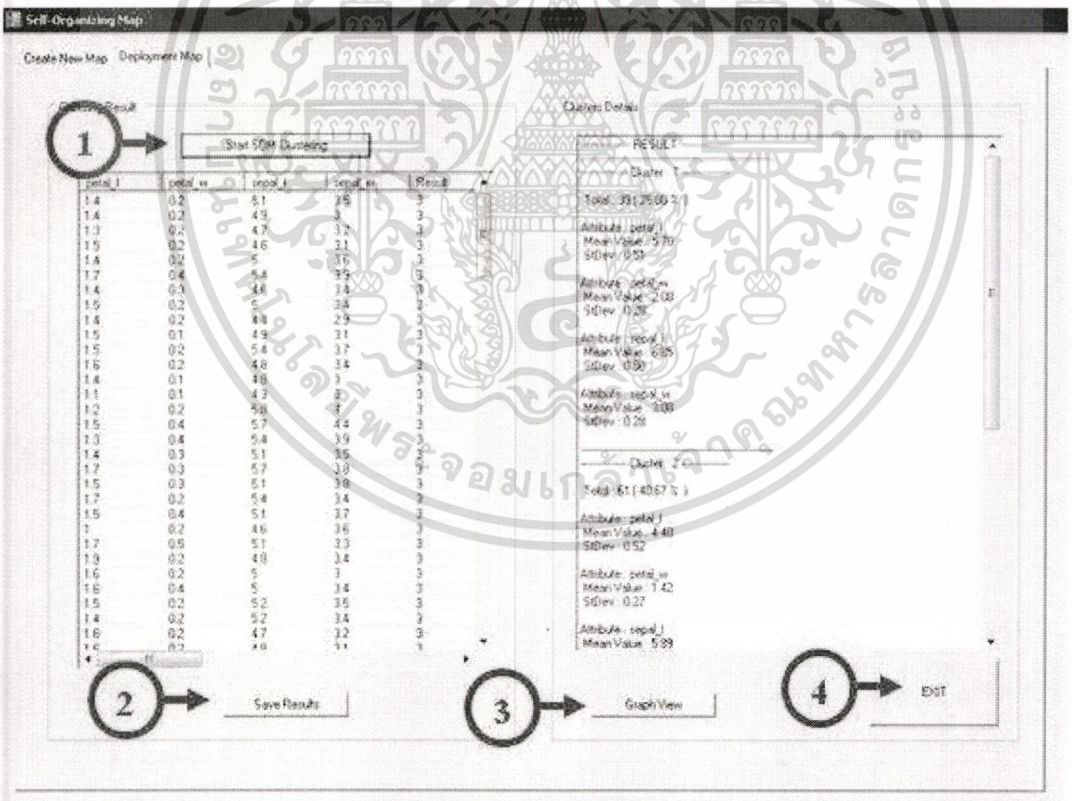
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.18 กล่องข้อความแสดงหลังจากโหลด Model เรียบร้อยแล้ว

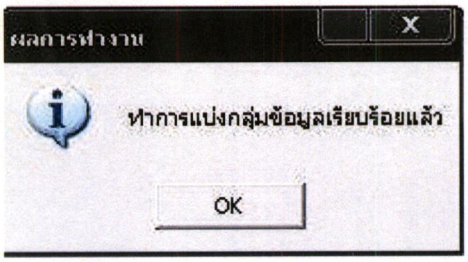
4.4.6.2 ขั้นตอนการแบ่งกลุ่มด้วย SOM

1. เมื่อผู้ใช้คลิกปุ่ม Start SOM Clustering เพื่อทำการประมวลผลโครงข่ายและแสดงผลบนหน้าจอ
2. ผู้ใช้สามารถคลิกปุ่ม Save Results เลือกบันทึกผลลัพธ์เก็บไว้รูปแบบไฟล์ XLS
3. ผู้ใช้สามารถคลิกปุ่ม Graph View เพื่อดูกราฟค่า means แต่ละกลุ่มแสดงตามมิติของข้อมูล
4. คลิกปุ่ม EXIT เมื่อต้องการออกจากหน้าจอการแบ่งกลุ่มด้วย SOM

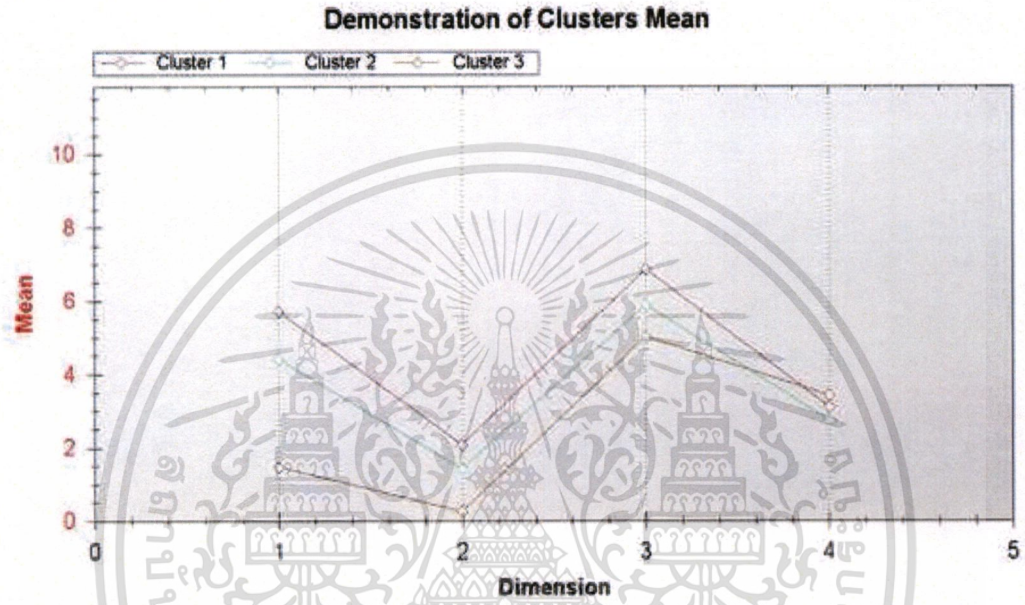


รูปที่ 4.19 หน้าจอ Deployment Map ของแบบจำลองการแบ่งกลุ่มด้วย SOM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.20 กล่องข้อความแสดงหลังจากทำการบันทึกข้อมูลผลลัพธ์เรียบร้อยแล้ว



รูปที่ 4.21 แสดงกราฟค่า mean แต่ละมิติของข้อมูลแต่ละกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุป และข้อเสนอแนะ

ผลที่ได้จากการทำโครงงานพัฒนาระบบงานในหัวข้อเรื่อง "การพัฒนาระบบแบ่งกลุ่มด้วย SOM" สามารถสรุปผลได้ดังนี้

5.1 สรุปผลการดำเนินงาน

จากการศึกษาในโครงการนี้ สามารถสรุปผลการดำเนินงานในการออกแบบโครงข่ายประสาทเทียมได้ดังนี้

1. ในการนำเข้าข้อมูลมาใช้ในการเรียนรู้และทดสอบนั้น ต้องมีลักษณะของข้อมูลที่อยู่ในช่วงเดียวกัน หรือค่าของข้อมูลมีการกระจายตัวอย่างเหมาะสม จะทำให้ระบบมีการเรียนรู้ที่ดีขึ้น
2. การเรียนรู้ในแต่ละครั้งของโครงข่าย ผลการทดลองจะให้ค่าที่แตกต่างกัน เนื่องจากผลลัพธ์ที่ได้จากการทำงานจะขึ้นอยู่กับ หลายปัจจัย เช่น จำนวนข้อมูลที่ใช้ในการฝึกสอน การกำหนดอัตราการเรียนรู้ การกำหนดขนาดของโครงข่าย การกำหนดค่าตัวแปรที่เป็นเงื่อนไข ดังนั้นในการฝึกสอนและการกำหนดค่าต่างๆ ให้กับระบบต้องอาศัยความชำนาญและประสบการณ์ของผู้ใช้เข้ามาเกี่ยวข้องด้วย เพื่อให้เกิดความผิดพลาดน้อยที่สุด

5.2 สรุปผลการทดลอง

1. ค่าน้ำหนักเริ่มต้น ที่ได้จากการสุ่มมีผลต่อการเริ่มต้นการทำงานของระบบ รวมไปถึงผลลัพธ์ที่จะได้รับจากการคำนวณ
2. การกำหนดค่าตัวแปรต่างๆ ในตอนเริ่มต้นการสร้างโครงข่ายมีผลต่อผลลัพธ์ที่ได้รับจากการคำนวณโครงข่าย

5.3 ปัญหาที่พบในการพัฒนา

5.3.1 ปัญหาที่พบ

เนื่องจากระบบที่ทำการพัฒนาได้นำระบบการเตรียมข้อมูลและสำรวจมารวมไว้ด้วยกัน จึงต้องอาศัยเวลาในการทำความเข้าใจ และสอบถามขั้นตอนการใช้งานของระบบดังกล่าว จึงจะสามารถสรุปและนำมาพัฒนาระบบต่อไปได้ ทำให้ต้องใช้เวลาค่อนข้างมากในการพัฒนาระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3.2 ข้อเสียของการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นเพียงเครื่องมือที่ใช้สำหรับวิเคราะห์เท่านั้น โดยที่ไม่สามารถเข้าใจถึงข้อมูลได้ดีเท่ากับมนุษย์ ดังนั้นผู้ใช้เครื่องมือการทำเหมืองข้อมูล รวมไปถึงอัลกอริทึมที่เลือกจึงต้องมีความรู้ความเข้าใจเป็นอย่างดี อย่างไรก็ตามการทำเหมืองข้อมูลจะเป็นเพียงเครื่องมือที่ช่วยประกอบการตัดสินใจ ซึ่งไม่สามารถช่วยระบุค่าของข้อมูล หรือค่าที่แสดงความสัมพันธ์ของรูปแบบและความสัมพันธ์ของข้อมูลนั้นจริง จะต้องให้ผู้ใช้งานทำการตัดสินใจอีกครั้งหนึ่ง

5.4 ข้อเสนอแนะ

ระบบที่พัฒนาขึ้นสามารถเชื่อมต่อกับฐานข้อมูลในระบบ Microsoft SQL Server เท่านั้น หากผู้ใช้ต้องการทำงานกับโปรแกรมฐานข้อมูลระบบอื่น จะต้องทำการนำข้อมูลเข้า โดยผ่าน DTS Import/Export Wizard ที่ Microsoft SQL Server รองรับ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

ศุภชัย สมพานิช. 2549. **Database Programming ด้วย VB 2005 & VC# 2005 ฉบับสมบูรณ์.** นนทบุรี : ไอดีซีฯ.

สมพร จิวรสกุล. 2545. **SQL Server2000 ฉบับสมบูรณ์**. Infopress Developer Book.

สัจจะ จรัสรุ่งรวีวร. 2550. **คู่มือ Visual C# 2005 ฉบับสมบูรณ์.** นนทบุรี : ไอดีซีฯ.

อาทิตยา เชื้อจันอัด. 2549. "การพัฒนาระบบการเตรียมข้อมูลและการสำรวจ สำหรับการทำดาต้าไมนิ่ง"โครงการพัฒนาระบบงานวิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

Anil K. Jan, Richard C. Dubes.1988. **Algorithms for Clustering Data**, Prentice Hall,

J. Schatzman. 2003. "Using Self-Organizing Maps to Visualize Clusters and Trends in Multidimensional Datasets", Department of Computing Data Mining Group,

Jiawei Han and Micheline Kamber. 2001. **Data Mining: Concept and Techniques**. San Francisco. Morgan Kaufman Publishers.

Microsoft Corporation. 2005. **Microsoft SQL Server 2005**. [Online]. Available: http://www.microsoft.com

T. Kohonen. 1998. "The Self-Organizing map" Helsinki University of Technology Nueral Networks Research Center, P.O.Box 2200, FIN-02015, Finland.

T. Kohonen. 1985. **Median strings.** ***Pattern Recognition Letters***, 3:309.313.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายธวิทย์ เชิดเกียรติสกุล
วัน เดือน ปีเกิด	26 มิถุนายน 1983
สถานที่เกิด	ชุมพร
วุฒิการศึกษาระดับปริญญาตรี	บริหารธุรกิจบัณฑิต
สถานที่สำเร็จการศึกษา	คณะวิทยาการจัดการ มหาวิทยาลัยสงขลานครินทร์
ปีที่สำเร็จการศึกษา	2549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้