ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

การพัฒนาระบบงานเพื่อการพยากรณ์ความต้องการใช้ชุดบริการเสริมพิเศษในธุรกิจวิทยุติดตามตัวโดยการประยุกต์ใช้ดาต้าไมนิ่ง

System Development for Forecasting of Value Added Service Package in Pager Business by applying Data Mining.

โดย

นางสาวกิ่งกาญจน์ เหลืองพุทธรัตน์

รหัส 42067067

H001760

อาจารย์ที่ปรึกษา

ดร. วรพจน์ กรีสุระเดช

วัน เดือน ปี 09 ม.ค. 2550
เลขทะเบียน 01760
เลขเรียกหนังสือ สพ. ก63ก 2543
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 2 ปีการศึกษา 2543

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การพัฒนาระบบงานเพื่อการพยากรณ์ความต้องการใช้ชุดบริการเสริมพิเศษในธุรกิจวิทยุติดตามตัวโดยการประยุกต์ใช้ดาต้าไมนิ่ง
นักศึกษา	นางสาว กิ่งกาญจน์ เหลืองพุทธรัตน์
อาจารย์ที่ปรึกษา	ดร. วรพจน์ กรีสุระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2543

บทคัดย่อ

เนื่องมาจากในธุรกิจวิทยุติดตามตัวมีการแข่งขันกันสูงระหว่างผู้ให้บริการ จึงมักจะมีการจัดรายการส่งเสริมการขายต่างๆ อย่างต่อเนื่อง และการจัดทำชุดบริการเสริมพิเศษเพื่อมอบให้กับผู้บริโภคที่ซื้อเครื่องใหม่ก็เป็นรายการส่งเสริมการขายหนึ่งที่ผู้ให้บริการมักจะจัดให้มีขึ้นอย่างสม่ำเสมอ แต่จะทำอย่างไรจึงจะสามารถทราบความต้องการที่แท้จริงของผู้บริโภค โครงการนี้จะนำเสนอถึงขั้นตอนและวิธีการพัฒนาระบบงานเพื่อทำการพยากรณ์ความต้องการใช้ชุดบริการเสริมพิเศษของผู้บริโภคแต่ละคน โดยจะทำการศึกษาจากข้อมูลพฤติกรรมการใช้บริการของกลุ่มผู้บริโภคที่มีอยู่เดิม แล้วนำความรู้ที่ได้มาทำการพยากรณ์เพื่อจัดหมวดหมู่(Classification) สำหรับผู้บริโภครายใหม่ๆ โดยในโครงการฉบับนี้มุ่งเน้นที่จะศึกษาถึงวิธีการสร้างแบบจำลองพยากรณ์ (Predictive modeling) โดยการประยุกต์ใช้ Neural induction ซึ่งเป็นเทคนิคหนึ่งของการจัดหมวดหมู่ (Classification) ของดาต้าไมนิ่ง ในการแก้ปัญหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	System Development for Predictive of Value Added Service package requirement in Pager Business by Data Mining approach.
Student	Ms. Kingkan Luengputtarat
Advisor	Dr. Worapoj Kreesuradej
Level of Study	Information of Science in Information Technology
Major	Information Science
Academic Year	2000

ABSTRACT

There is the increasingly on competition among Pager service business. The promotion that organization always set for the new customer is the package of "Value added service". The package of "Value added service" for promotion in Pager business is very important, but to hardly to meet customer's need. This project shows the step and implementation for the Data mining to offer suitable "Value added service" package for each customer. To study about the Predictive modeling which intend to predict and analysis the behavior trend of customer service in order to provide good solution "Value added service" package by using Neural Induction which is the "Classification" operation to solve the problem.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงการพัฒนาระบบงานฉบับนี้สำเร็จลุล่วงลงได้ด้วยดี ด้วยคำแนะนำและความช่วยเหลือเป็นอย่างดียิ่งของ ดร. วรพจน์ กรีสุระเดช อาจารย์ที่ปรึกษาวิชาโครงการพัฒนาระบบงาน

ดิฉันขอกราบขอบพระคุณบุพการี รวมทั้งผู้ใกล้ชิดทุกท่าน ซึ่งได้ให้การสนับสนุนตลอดมา และ ขอกราบขอบพระคุณคณาจารย์ทุกท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้แก่ดิฉัน จนกระทั่งโครงงานสำเร็จลงได้ด้วยดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญรูปภาพ	VII
บทที่	
1. บทนำ	1
1.1. ความเป็นมา	1
1.2. วัตถุประสงค์	2
1.3. ขอบเขตของโครงการ	2
1.4. ขั้นตอนการดำเนินงาน	2
1.5. ประโยชน์ที่คาดว่าจะได้รับ	2
2. ทฤษฎีที่เกี่ยวข้อง	4
2.1. ดาต้าไมนิ่ง	4
2.2. ขั้นตอนการทำดาต้าไมนิ่ง (Data Mining Process)	4
2.3. Neural Induction	7
2.4. โครงข่ายประสาทเทียมแบบการแพร่ย้อนกลับ (Back-propagation Algorithm)	12
3. การประยุกต์ใช้ดาต้าไมนิ่งเพื่อการออกแบบและพัฒนาระบบงาน	18
3.1. กำหนดวัตถุประสงค์	18
3.2. จัดเตรียมข้อมูล	18
3.3. การนำข้อมูลมาทำดาต้าไมนิ่ง	21
4. วิเคราะห์ผลการดำเนินการ	31
4.1. ผลที่ได้จากการดำเนินการ	31
4.2. การวัดความถูกต้องของแบบจำลอง	32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้า

4.3. ข้อดีและข้อเสียที่เกิดขึ้นจากการนำเทคนิคโครงข่ายประสาทเทียมมาใช้เพื่อการพยากรณ์ 34
5. สรุปผลการศึกษาและข้อเสนอแนะ 36
5.1. สรุปผลการดำเนินงาน 36
5.2. ข้อเสนอแนะ 36
บรรณานุกรม 38
ประวัติผู้เขียน 39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่

3.1. รายการบริการเสริม 19

4.1. ตัวอย่างชุดข้อมูลทดสอบและผลลัพธ์ที่โครงข่ายคำนวณให้ 32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

หน้า

ภาพที่

2.1. เปอร์เซ็นต์ของเวลาที่ใช้ในแต่ละขั้นตอน 7

2.2. ลักษณะเครือข่ายประสาทสิ่งมีชีวิต 8

2.3. แบบจำลองเซลล์ประสาทเทียม 9

2.4. กราฟของฟังก์ชัน Bipolar 10

2.5. กราฟของฟังก์ชัน Unipolar 11

2.6. โครงข่ายแบบหลายชั้น (Multilayer) 11

2.7. โครงข่ายประสาทเทียมแบบการแพร่ย้อนกลับ (Back-propagation Algorithm) 13

2.8. Back-propagation Training Algorithm 17

3.1. หน้าจอสำหรับการสอนและทดสอบแบบจำลองโครงข่ายประสาทเทียม 23

3.2. หน้าจอสำหรับการพยากรณ์บริการเสริมให้ลูกค้า 24

3.3. หน้าจอสำหรับการตั้งค่าให้กับโครงข่ายประสาทเทียม 25

3.4. หน้าจอสำหรับการกำหนดกลุ่มบริการเสริม 26

3.5. หน้าจอสำหรับการเลือกบริการเสริมในแต่ละกลุ่ม 26

3.6. หน้าจอการกำหนดข้อมูลนำเข้า 27

3.7. หน้าจอแสดงรายการข้อมูลที่เลือก 28

3.8. หน้าจอการแปลงข้อความเป็นตัวเลข 28

3.9. หน้าจอการแปลงวันที่เป็นตัวเลข 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1. ความเป็นมา

เนื่องมาจากความต้องการในการติดต่อสื่อสารของผู้บริโภค ทำให้ภาคธุรกิจสื่อสารได้เข้ามามีบทบาทในชีวิตประจำวันของเรามากขึ้น เช่น ธุรกิจการติดต่อสื่อสารผ่านมือถือ, การสื่อสารโดยใช้จดหมายอิเล็คโทรนิค, อินเทอร์เน็ต หรือการส่งข้อความผ่านวิทยุติดตามตัว(Pager) ในการที่ผู้บริโภคจะเลือกใช้บริการแบบไหนนั้นก็ขึ้นอยู่กับปัจจัยหลายๆอย่าง เช่น ค่าใช้จ่าย, ลักษณะการใช้งาน เป็นต้น

สำหรับการติดต่อสื่อสารด้วยวิทยุติดตามตัวก็เป็นอีกทางเลือกหนึ่งซึ่งสามารถตอบสนองความต้องการบางส่วนของผู้บริโภคได้เป็นอย่างดี เช่น กลุ่มผู้บริโภคที่ต้องการส่งหรือรับข้อความเพียงอย่างเดียว ทั้งยังจะมีค่าใช้จ่ายที่ไม่สูงมากนักเมื่อเทียบกับโทรศัพท์มือถือ และยังมีความคล่องตัวมากกว่าการใช้โทรสารหรือจดหมายอิเล็คโทรนิคอีกด้วย

แต่อย่างไรก็ตามในธุรกิจการสื่อสารด้วยวิทยุติดตามตัวนี้ก็มีการแข่งขันกันสูงทั้งในด้านของการให้บริการ, ราคาเครื่อง และราคาค่าบริการ โดยบริษัทผู้ให้บริการก็มักจะมีการจัดรายการส่งเสริมการขายต่างๆอย่างต่อเนื่อง เช่น การลดราคาเครื่อง การลดราคาค่าบริการ หรือการจัดชุดบริการเสริมพิเศษให้กับลูกค้าที่ซื้อเครื่องใหม่ในช่วงระยะเวลาต่างๆ เพื่อสร้างแรงจูงใจให้ลูกค้าหันมาใช้บริการของบริษัท ทั้งยังเป็นการให้ลูกค้าได้ทดลองใช้บริการเสริมใหม่ๆ โดยได้รับส่วนลดค่าบริการหรือไม่ต้องเสียค่าบริการในช่วงระยะเวลาหนึ่งก่อนตัดสินใจว่าจะเลือกใช้บริการนั้นๆต่อไปหรือไม่ ดังนั้นสิ่งหนึ่งที่ผู้ประกอบการต้องการทราบ ก็คือ ความเหมาะสมในการจัดชุดบริการเสริมให้กับลูกค้า ว่าควรจะจัดบริการเสริมแบบใดให้กับลูกค้าจึงจะตรงกับความต้องการที่แตกต่างกัน

ซึ่งในโครงงานฉบับนี้จะได้นำปัญหาดังกล่าวมาวิเคราะห์ โดยการประยุกต์ใช้ดาต้าไมนิ่ง โดยจะนำเสนอถึงขั้นตอนและวิธีการพัฒนาระบบงานเพื่อทำการพยากรณ์ความต้องการใช้ชุดบริการเสริมพิเศษของผู้บริโภคแต่ละราย โดยจะทำการศึกษาจากข้อมูลพฤติกรรมการใช้บริการของกลุ่มผู้บริโภคที่มีอยู่เดิม แล้วนำความรู้ที่ได้มาทำการพยากรณ์สำหรับผู้บริโภครายใหม่ๆ โดยมุ่งเน้นที่จะศึกษาถึงวิธีการสร้างแบบจำลองพยากรณ์ (Predictive Modeling) โดยการประยุกต์ใช้หลักการโครงข่ายประสาทเทียม (Neural Induction) ซึ่งเป็นเทคนิคหนึ่งของวิธีการจัดหมวดหมู่ (Classification) ในดาต้าไมนิ่งเพื่อการแก้ปัญหา

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2. วัตถุประสงค์

1. เพื่อศึกษาและประยุกต์ใช้ดาต้าไมนิ่ง (Data Mining) ในการทำการพยากรณ์ชุดบริการเสริมให้กับลูกค้าแต่ละราย
2. เพื่อเป็นแนวทางในการนำไปประยุกต์ใช้ในการสนับสนุนการตัดสินใจของผู้บริหารสำหรับวางแผนเพื่อทำรายการส่งเสริมการขาย โดยการจัดชุดบริการเสริมพิเศษที่เหมาะสมกับความต้องการของลูกค้า

1.3. ขอบเขตของโครงการ

1. วิเคราะห์ข้อมูลที่มีอยู่ภายในองค์กร เพื่อนำมาใช้ในการพยากรณ์
2. ออกแบบและพัฒนาระบบงานเพื่อทำการพยากรณ์ความต้องการใช้ชุดบริการเสริมพิเศษของผู้บริโภคแต่ละราย โดยใช้โครงข่ายประสาทเทียม (Neural Network)

1.4. ขั้นตอนการดำเนินงาน

1. ศึกษาและเก็บรวบรวมข้อมูลที่มีอยู่ในองค์กร
2. ศึกษาทฤษฎีดาต้าไมนิ่ง (Data Mining) เพื่อการนำมาประยุกต์ใช้
3. ศึกษาทฤษฎีโครงข่ายประสาทเทียม (Neural Network) เพื่อนำมาสร้างแบบจำลองพยากรณ์ (Predictive Modeling)
4. ออกแบบและพัฒนาระบบงานเพื่อทำการพยากรณ์ความต้องการใช้ชุดบริการเสริมพิเศษของผู้บริโภคแต่ละราย
5. ประเมินผลและวิเคราะห์ผลที่ได้จากแบบจำลองพยากรณ์ (Predictive Modeling)
6. สรุปผลการศึกษาของโครงงาน
7. จัดทำรายงาน และข้อเสนอแนะ

1.5. ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นแนวทางในการนำดาต้าไมนิ่งมาประยุกต์ใช้กับข้อมูลทางธุรกิจ
2. เข้าใจถึงหลักการและขั้นตอนการทำดาต้าไมนิ่ง
3. เข้าใจหลักการทำงานของโครงข่ายประสาทเทียม (Neural Network) แบบการแพร่ย้อนกลับ (Back-propagation Algorithm)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. เป็นข้อมูลสำหรับผู้บริหาร เพื่อช่วยในการตัดสินใจและวางแผนการส่งเสริมการขาย โดยการจัดชุดบริการเสริมพิเศษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1. ดาต้าไมนิ่ง

ในภาคธุรกิจปัจจุบันมีการแข่งขันสูง จึงมีความพยายามที่จะคิดค้นและพัฒนาเทคนิคต่างๆ ขึ้น เพื่อให้สามารถแข่งขันได้ในตลาด และก็เป็นที่ยอมรับกันโดยทั่วไปว่า ข้อมูล คือหัวใจสำคัญในการทำธุรกิจ การที่เรารู้ข้อมูลมากและสามารถนำไปใช้ได้ถูกต้อง ก็จะสร้างโอกาสให้กับธุรกิจมากขึ้น และดาต้าไมนิ่ง (Data Mining) ก็เป็นเทคนิคหนึ่งที่ถูกนำมาใช้

ดาต้าไมนิ่ง เป็นการรวบรวมเทคนิคจากงานต่างๆ เช่น การเรียนรู้ของเครื่องจักร (Machine Learning), การจดจำรูปแบบ (Pattern Recognition), หลักสถิติ (Statistics) และฐานข้อมูล (Database) เป็นต้น เพื่อนำมาใช้ในการค้นหาความสัมพันธ์ของข้อมูลและข่าวสาร (Information) ที่มีประโยชน์ที่แฝงอยู่ในฐานข้อมูลขนาดใหญ่ โดยข้อมูลที่ได้มาต้องสามารถนำไปใช้ประโยชน์ได้จริง

2.2. ขั้นตอนการทำดาต้าไมนิ่ง (Data Mining Process)

ในการทำดาต้าไมนิ่งจะมีขั้นตอนการทำงาน ดังนี้

2.2.1. กำหนดวัตถุประสงค์ (Business Objectives Determination)

เป็นการกำหนดปัญหาและวัตถุประสงค์ขององค์กรให้ชัดเจน เพื่อนำมาวิเคราะห์ว่า สามารถนำดาต้าไมนิ่งไปช่วยทำอะไรได้บ้าง

2.2.2. จัดเตรียมข้อมูล (Data Preparation)

เป็นขั้นตอนที่สำคัญและใช้เวลานานที่สุดประมาณ 60 % ของเวลาทั้งหมด เนื่องจากข้อมูลที่จะนำมาใช้ต้องคัดเลือกจากข้อมูลที่มีความเหมาะสมและอยู่ในประเด็นที่เราต้องการ ซึ่งมีขั้นตอนในการทำงาน ดังนี้

2.2.2.1. คัดเลือกข้อมูล (Data Selection) คือการคัดเลือกข้อมูลสำหรับดาต้าไมนิ่งจากข้อมูลทั้งหมดขององค์กร โดยต้องคำนึงถึงวัตถุประสงค์ในการนำข้อมูลมาใช้งาน นอกจากนี้ต้องทำความเข้าใจกับข้อมูลและประเภทข้อมูลที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะนำมาใช้ด้วย และสิ่งสำคัญอีกสิ่งที่ต้องพิจารณาในการคัดเลือกข้อมูล คือ อายุการใช้งานของข้อมูลที่เลือก เช่น การนำประเภทงานของลูกค้าเข้ามาพิจารณา โอกาสที่ลูกค้าจะเปลี่ยนงานในแต่ละปีจะมีสูง เพราะฉะนั้นต้องทำการตรวจสอบให้แน่ชัดว่าข้อมูลนั้นถูกต้องหรือไม่ โดยประเภทข้อมูลแบ่งเป็น

- ข้อมูลตัวเลข (Quantitative) ซึ่งได้แก่ ค่าจำนวนเต็ม (Discrete) เช่น จำนวนพนักงาน และ เลขจำนวนจริง (Continuous) เช่น รายได้, ค่าเฉลี่ยต่างๆ เป็นต้น
- ข้อมูลที่ไม่ใช่ตัวเลข (Categorical data) ซึ่งได้แก่ ข้อมูลที่ลำดับไม่มีความสำคัญ (Nominal categorical) เช่น เพศ(ชาย,หญิง) และ ข้อมูลที่ลำดับมีความสำคัญ (Ordinal categorical) เช่น เกรด(A,B,C,D,F) ถ้าแปลงให้อยู่ในรูปตัวเลขต้องให้ได้ความหมายเดิม

2.2.2.2. การตรวจสอบข้อมูล (Data Preprocess) ทำการตรวจสอบข้อมูลที่คัดเลือกมาว่าเป็นข้อมูลที่เหมาะสมหรือไม่ โดยใช้หลักการทางสถิติ เช่น ข้อมูลที่เป็น Categorical การวัดการกกระจายของข้อมูลจะทำให้เข้าใจข้อมูลที่มีอยู่ได้ดียิ่งขึ้น วิธีการที่ง่ายที่สุด คือ การนำข้อมูลนั้นมาสร้างเป็นกราฟ ซึ่งช่วยให้เห็นความโน้มเอียงของข้อมูลและข้อมูลที่ผิดปกติได้ ส่วนข้อมูลที่เป็นตัวเลข การวิเคราะห์ข้อมูลทำได้โดยการหาค่าสูงสุด (Max) , ค่าต่ำสุด (Min) , ค่าเฉลี่ย (Mean) ,ค่าที่ปรากฏบ่อย (Mode) , ค่ากลาง (Median) ซึ่งสิ่งที่จะปรากฏให้เห็นในขั้นตอนนี้ คือ

- Noisy data คือ ค่าของข้อมูลที่ผิดไปจากที่ควรจะเป็น อาจเกิดจากการป้อนข้อมูลผิด เช่น ป้อนอายุเป็น 650 ปี หรือป้อนรายได้เป็นค่าติดลบ เป็นต้น
- Missing data มีข้อมูลบางส่วนหายไป อาจเกิดจากความผิดพลาดในการป้อนข้อมูลหรือไม่มีข้อมูลส่วนนี้ในขณะที่รับข้อมูล แก้ไขได้โดยการตัดข้อมูลนั้นทิ้งทั้งรายการ หรือบันทึกส่วนที่หาไปด้วยค่าเฉลี่ย (Mean) หรือค่าที่ปรากฏบ่อย (Mode) สำหรับข้อมูลที่ไม่ใช่ตัวเลขอาจบันทึกด้วยค่าที่ปรากฏบ่อย (Mode) หรือบันทึกเป็น 'UNKNOWN'

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2.3. การปรับเปลี่ยนรูปแบบของข้อมูล (Data Transformation) เพื่อให้เหมาะสมกับอัลกอริทึ่ม เช่น การแปลงข้อมูลตัวเลขให้เป็นช่วงเพื่อใช้กับ Decision Tree หรือ การปรับอัตราส่วนตัวเลขให้อยู่ในช่วง 0-1 เพื่อใช้กับอัลกอริทึ่ม Neural Network

2.2.3. ดาต้าไมนิ่ง (Data Mining)

เลือกวิธีการและอัลกอลิทึ่มที่เหมาะสมเพื่อใช้กับข้อมูลที่ได้เตรียมไว้แล้ว ซึ่งมีอยู่ 4 ชนิด ด้วยกัน คือ

2.2.3.1. Data Segmentation เป็นกระบวนการการแบ่งฐานข้อมูล (Database) ออกเป็นกลุ่มเพื่อให้ง่ายต่อการวิเคราะห์ เช่น การแบ่งลูกค้าออกตามอายุ,เพศ, รายได้ เป็นต้น

2.2.3.2. Predictive Modeling เป็นการสร้างแบบจำลองพยากรณ์ แบ่งออกเป็น

- Classification เป็นการทำนายกลุ่มของรายการจากข้อมูลนำเข้า เช่น การทำนายว่าเป็นลูกค้าที่อยู่ในกลุ่มที่ควรส่งจดหมายแนะนำสินค้าและบริการใหม่ไปให้หรือไม่ โดยดูจากประวัติและพฤติกรรมการบริโภค เป็นต้น
- Value Prediction การทำนายค่าที่เป็นตัวเลข เช่น การทำนายภาษีที่จะเก็บได้ในปีหน้า หรือการทำนายราคาหุ้น เป็นต้น

2.2.3.3. Link Analysis ศึกษาว่าข้อมูลแต่ละรายการมีความสัมพันธ์กันหรือไม่อย่างไร เช่น Market basket เก็บข้อมูลการซื้อสินค้าแต่ละครั้งของลูกค้าเพื่อศึกษาแนวโน้มการซื้อและสินค้าที่ซื้อว่า เพื่อนำมาทำรายการส่งเสริมการขาย และการจัดชั้นวางสินค้าให้เหมาะสม

2.2.3.4. Deviation Detection การตรวจจับสิ่งผิดปกติต่างๆ เช่น การจับการโกง เป็นต้น

2.2.4. วิเคราะห์ผลลัพธ์ที่ได้ (Analysis of Result)

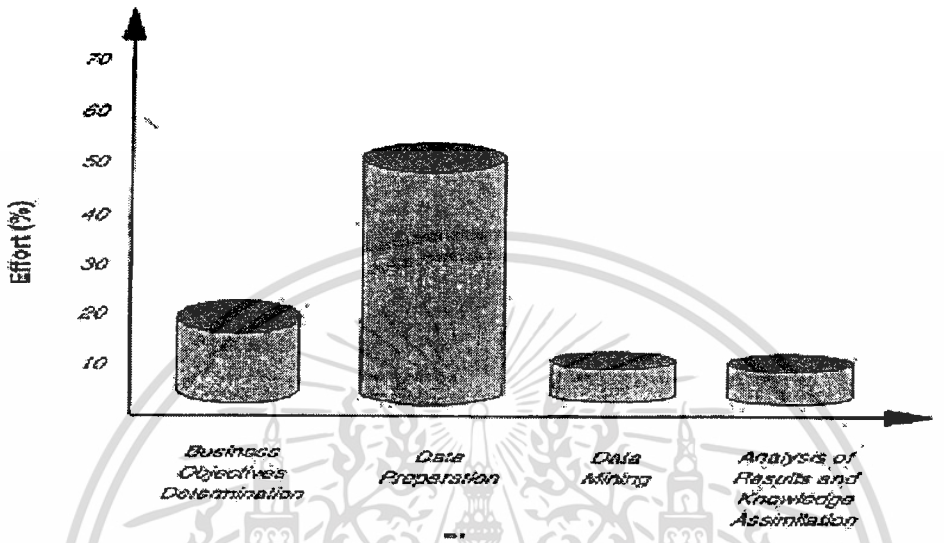
การแปลความหมายและประเมินผลลัพธ์ที่ได้จากขั้นตอนการทำดาต้าไมนิ่ง

2.2.6. ปรับความรู้ที่ได้เข้ากับธุรกิจ (Assimilation of Knowledge)

นำความรู้ที่ได้มาปรับให้เข้ากับธุรกิจ เพื่อนำมาใช้ให้เป็นประโยชน์กับองค์กร

เอกสารนี้เป็นเอกสาร... ...ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งในแต่ละขั้นตอนจะมีสัดส่วนของเปอร์เซ็นต์เวลาที่ใช้ ดังรูปที่ 2.1



รูปที่ 2.1 เปอร์เซ็นต์ของเวลาที่ใช้ในแต่ละขั้นตอน (Cabena et. al. 1997 : 43)

สำหรับในโครงงานที่นำจะขอนำเสนอวิธีการสร้างแบบจำลองพยากรณ์ (Predictive Modeling) โดยใช้วิธีการโครงข่ายประสาทเทียม (Neural Induction) เพื่อการจัดหมวดหมู่ (Classification) ดังได้กล่าวไว้ในบทที่ 1

2.3. Neural Induction

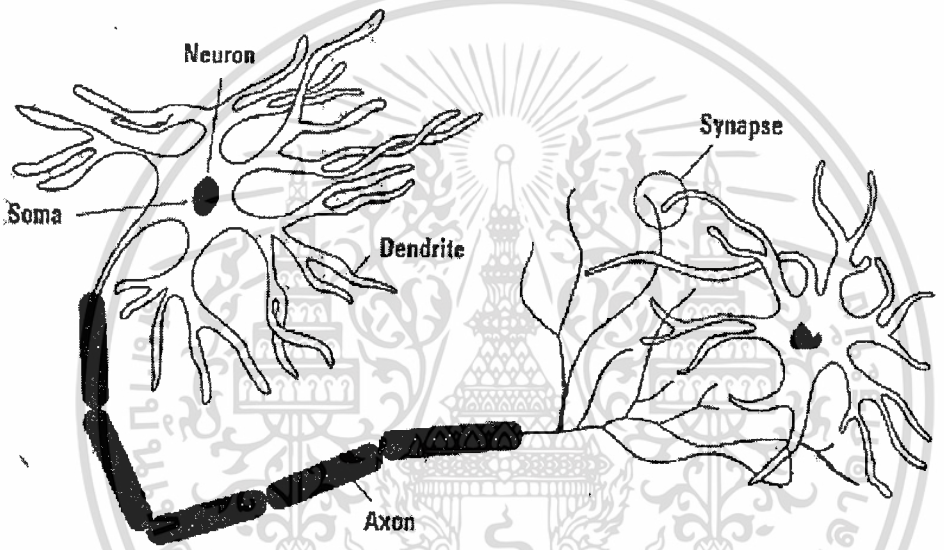
เป็นเทคนิคที่นำเสนอแบบจำลองที่มีโครงสร้างประกอบไปด้วย Node และค่าถ่วงน้ำหนักของแต่ละ Link ที่เชื่อมโยงระหว่าง Node ซึ่งมีพื้นฐานการทำงานแบบ Neural Network

Neural Network เป็นแบบจำลองที่สร้างขึ้นเพื่อเลียนแบบการทำงานของเครือข่ายประสาทภายในสมองของมนุษย์ ซึ่งประกอบไปด้วยเซลล์ประสาทจำนวนมาก (Nerve cell) ประมาณ 10^{11} เซลล์ ซึ่งเซลล์เหล่านี้ทำหน้าที่เป็นหน่วยประมวลผล และมีบริเวณที่มีการส่งสัญญาณติดต่อกันระหว่างเซลล์ที่เรียกว่า ไซแนปส์ (Synaps) ประมาณ 10^{14} จุด รูปแบบการส่งสัญญาณทำในลักษณะขนานกัน ทำให้เกิดการเรียนรู้โดยการนำข้อมูลในอดีตมาวิเคราะห์และใช้แก้ปัญหาในปัจจุบัน ลักษณะเซลล์ประสาททางชีวภาพจะประกอบไปด้วย 3 ส่วนหลัก คือ เดนไดรท์ (Dendrites), โซมา (Soma) และ แอคซอน (Axon)

โดยเดนไดรท์เป็นส่วนที่ยื่นออกไปจากเซลล์ประสาท มีลักษณะคล้ายกิ่งก้านทำหน้าที่รับสัญญาณจากเซลล์ประสาทผ่านแอคซอนซึ่งเป็นเส้นไฟเบอร์ขนาดยาวทำหน้าที่เหมือนสายส่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัญญาณ นำสัญญาณจากเซลล์ประสาทมาแปลงเป็นสัญญาณกระตุ้นเพื่อส่งต่อ โดยบริเวณที่แอคซอนและเดนไดรท์ติดต่อสื่อสารกัน เรียกว่า ไซแนปส์ ซึ่งสัญญาณที่มาถึงบริเวณไซแนปส์จะเป็นสัญญาณไฟฟ้าและการรับสัญญาณของเดนไดรท์ก็เกิดจากปฏิกริยาทางเคมีที่เกิดขึ้นในบริเวณนี้ หลังจากนั้นสัญญาณจะถูกส่งไปยังโซมาหรือตัวเซลล์ประสาทเพื่อรวบรวมสัญญาณแล้วส่งไปยังเซลล์ประสาทอื่นต่อๆไป

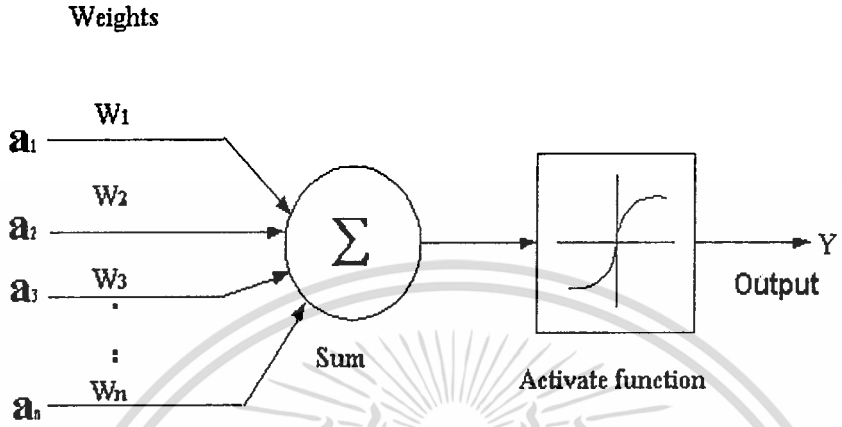


รูปที่ 2.2 ลักษณะเครือข่ายประสาทสิ่งมีชีวิต (Zirilli. 1997 : 4)

ซึ่งในแบบจำลองโครงข่ายประสาทเทียมที่สร้างขึ้นเพื่อเลียนแบบการทำงานของสมองมนุษย์นี้ ก็จะประกอบด้วย 3 ส่วนหลัก เช่นเดียวกัน ดังแสดงในรูปที่ 2.3 คือ

- เส้นที่เชื่อมโยงระหว่างชั้นเพื่อรับค่าอินพุท เปรียบเสมือนเดนไดรท์ และบริเวณที่มีการรับข้อมูลมีการใส่ค่าถ่วงน้ำหนัก (Weight) ให้กับค่าอินพุทที่รับเข้ามา ก็เปรียบเสมือนปฏิกริยาเคมีที่เกิดขึ้นบริเวณไซแนปส์
- Summation ที่ทำหน้าที่รวบรวมข้อมูลนำเข้าแล้วคำนวณออกมาเป็นค่าค่าหนึ่ง เรียกว่าเป็นค่า *net* เปรียบเสมือนการรวมสัญญาณของโซมาหรือตัวเซลล์ประสาท
- ส่วนการแปลงค่าเพื่อส่งให้เซลล์อื่นๆ โดยการผ่านฟังก์ชันกระตุ้น (Activate function) จะเปรียบเสมือนแอคซอนที่ทำหน้าที่เหมือนสายส่งสัญญาณที่ทำการนำสัญญาณจากเซลล์ประสาทมาแปลงเป็นเป็นสัญญาณกระตุ้นเพื่อส่งต่อไปยังโหนดอื่นต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 แบบจำลองเซลล์ประสาทเทียม

สำหรับการคำนวณหาค่า *net* มีสูตรดังนี้

$$net = w_1 a_1 + w_2 a_2 + w_3 a_3 + \ldots + w_n a_n$$

หรือ

$$net = \sum_{i=1}^{n} w_i a_i$$

โดยที่ *net* คือ ค่าผลลัพธ์จาก Summation

w_i คือ ค่าถ่วงน้ำหนักลำดับที่ i

a_i คือ ค่าอินพุทลำดับที่ i

สำหรับฟังก์ชันกระตุ้น (Activate Function) จะมีอยู่ด้วยกันหลายรูปแบบ ในที่นี้จะขอยกตัวอย่างเฉพาะซิกมอยด์ฟังก์ชัน (Sigmoid Function) เท่านั้น แบ่งได้เป็น 2 ประเภท แยกตามลักษณะผลลัพธ์ที่คำนวณได้

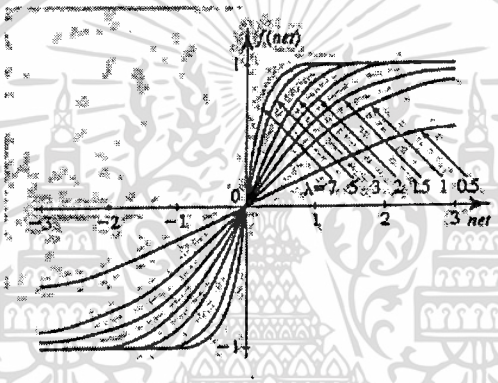
- Bipolar Binary Activation Function ค่า *net* ที่ผ่านฟังก์ชันจะถูกแปลงให้อยู่ในช่วง -1 ถึง 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$f(net) = \frac{2}{1+\exp(-\lambda net)} - 1$$

$$f\ (\ net\ \ )\ =\ \begin{cases} +\ 1\ ,\ net & >\ 0 \\ -\ 1\ ,\ net & <\ 0 \end{cases}$$

โดยที่ λ เป็นค่าคงที่ แปรผันตรงกับความชันของกราฟ



รูปที่ 2.4 กราฟของ Sigmoid Bipolar Binary Activation Function (Zurada. 1992 : 34)

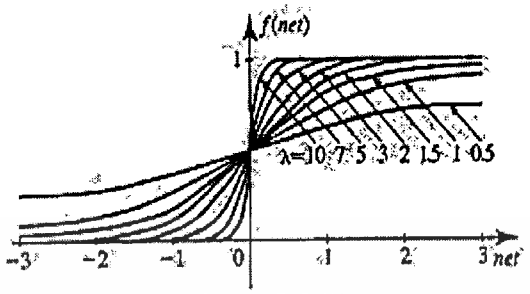
- Unipolar Binary Activation Function ค่า *net* ที่ผ่านฟังก์ชันจะถูกแปลงให้อยู่ในช่วง 0 ถึง 1

$$f(net) = \frac{1}{1+\exp(-\lambda net)}$$

$$f(net) = \begin{cases} 0, net > 0 \\ 1, net < 0 \end{cases}$$

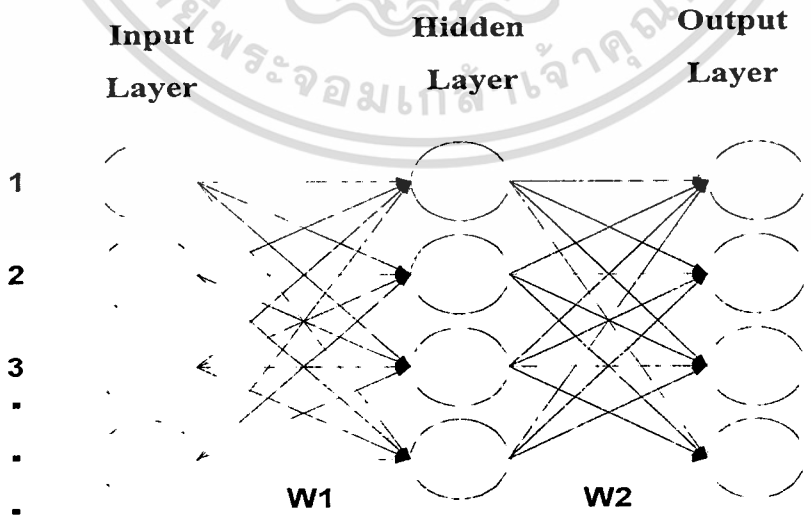
โดยที่ λ เป็นค่าคงที่ แปรผันตรงกับความชันของกราฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 Sigmoid Unipolar Binary Activation Function (Zurada. 1992 : 34)

จากที่กล่าวข้างต้นเป็นการคำนวณภายใน 1 โหนด ของเซลล์ประสาทเทียมเท่านั้น ซึ่งในแต่ละโหนดจะเป็น 1 ยูนิตของ Input, Process และ Output และภายในโครงข่ายประสาทเทียมจะประกอบไปด้วยโหนดจำนวนมากที่เชื่อมต่อกัน โดยการเชื่อมต่อจะแบ่งเป็นชั้น (Layer) มีอย่างน้อย 2 ชั้น คือ ชั้นอินพุท (Input layer) และ ชั้นเอาท์พุท (Output layer) เรียกว่า โครงข่ายแบบชั้นเดียว (Single layer) แต่ถ้ามีจำนวนชั้นมากขึ้น คือ เพิ่มชั้นซ่อน (Hidden layer) ซึ่งมีได้มากกว่า 1 ชั้น จะเรียกว่า โครงข่ายแบบหลายชั้น (Multilayer) จะมีประสิทธิภาพมากกว่าสามารถแก้ปัญหาข้อมูลนำเข้าที่ซับซ้อนขึ้นกว่าโครงข่ายแบบชั้นเดียว



รูปที่ 2.6 โครงข่ายแบบหลายชั้น (Multilayer)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

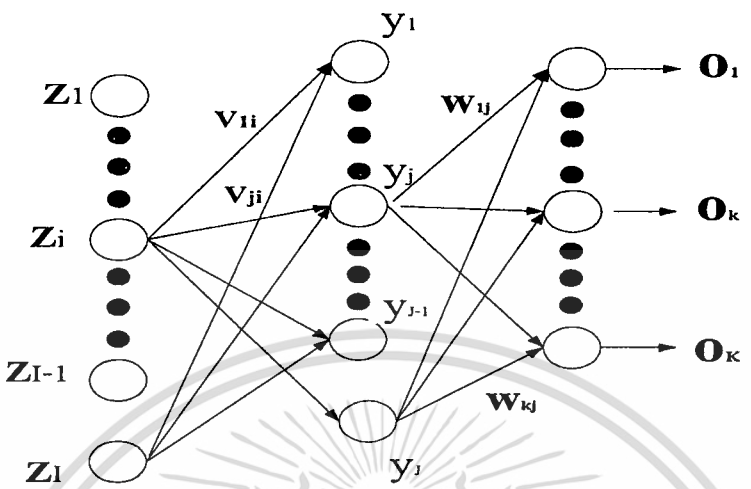
ซึ่งในการนำโครงข่ายประสาทเทียม (Neural Network) มาใช้ในการจัดหมวดหมู่ (Classification) เพื่อพยากรณ์ชุดบริการเสริมให้กับลูกค้านั้น ได้เลือกอัลกอลิทึ่มการแพร่ย้อนกลับ (Back-propagation Algorithm) ซึ่งมีโครงข่ายหลายชั้น มาใช้ในโครงงาน

2.4. โครงข่ายประสาทเทียมแบบการแพร่ย้อนกลับ (Back propagation Algorithm)

อัลกอลิทึ่มการแพร่ย้อนกลับ (Back Propagation) เป็นอัลกอริทึ่มหนึ่งของโครงข่ายประสาทเทียม (Neural Network) และเป็นอัลกอริทึ่มประเภท *Supervise Learning* คือ การเรียนรู้แบบที่ต้องมีการชี้นำ โดยต้องส่งข้อมูลเป็นคู่ของข้อมูลนำเข้า (Input) และข้อมูลผลลัพธ์ (Desired Output) เข้าไปในเครือข่าย โดยในเครือข่ายของ Network นั้นจะเป็นแบบโครงข่ายหลายชั้น คือ ประกอบด้วย ชั้นอินพุท (Input layer), ชั้นซ่อน (Hidden layer) และ ชั้นเอาท์พุท (Output layer) มีขั้นตอนการทำงาน 3 ขั้นตอน ดังนี้

1. ข้อมูลนำเข้าจะถูกส่งเข้าไปในชั้นอินพุท ซึ่งมีหน้าที่เพียงแค่รับข้อมูลแล้วแพร่ไปสู่ทุกๆโหนดในชั้นถัดไปโดยยังไม่มีการคำนวณเกิดขึ้น เมื่อข้อมูลไปถึงที่ชั้นซ่อน จึงจะมีการคำนวณค่า จากนั้นก็จะส่งข้อมูลที่ผ่านการคำนวณค่าและถูกแปลงโดยผ่านฟังก์ชันกระตุ้น ต่อไปยังให้กับชั้นซ่อนถัดๆไป จนกระทั่งถึงชั้นเอาท์พุท จึงจะได้ค่าผลลัพธ์ที่โครงข่ายคำนวณให้
2. เนื่องจากอัลกอลิทึ่มการแพร่ย้อนกลับเป็นอัลกอริทึ่มประเภท *Supervise Learning* จึงต้องมีการนำค่าเป้าหมายหรือค่าผลลัพธ์ที่ต้องการ (Desired output) เข้ามาเป็นส่วนหนึ่งของข้อมูลที่ใช้สำหรับฝึกสอนโครงข่ายให้เรียนรู้ โดยการนำผลลัพธ์ที่คำนวณได้จากชั้นเอาท์พุทไปเปรียบเทียบกับค่าเป้าหมาย แล้วคำนวณหาค่าความผิดพลาด (Error signal) แล้วแพร่ค่าความผิดพลาดกลับเข้าไปในโครงข่าย
3. โครงข่ายจะทำการปรับค่าถ่วงน้ำหนัก (weight) โดยเรียนรู้จากค่าความผิดพลาดที่เกิดขึ้น จากนั้นเริ่มต้นคำนวณใหม่ตั้งแต่ชั้นอินพุท วนไปเรื่อยๆ จนกระทั่งผลลัพธ์ที่คำนวณได้มีค่าความผิดพลาดเป็นที่ยอมรับได้จึงจะหยุดสอนเครือข่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 โครงข่ายประสาทเทียมแบบการแพร่ย้อนกลับ (Back propagation Algorithm)

ในการสร้างแบบจำลอง แบบ Multilayer Feedforward Network ของอัลกอริทึ่มการแพร่ย้อนกลับ (Back propagation algorithm) ประกอบด้วยขั้นตอน ดังนี้

1. จับคู่ข้อมูลนำเข้าและค่าเป้าหมายหรือผลลัพธ์ที่ต้องการ จำนวน p จากข้อมูล

$$(\vec{z}_1, \vec{d}_1), (\vec{z}_2, \vec{d}_2), ..., (\vec{z}_p, \vec{d}_p)$$

โดย z คือ ข้อมูลนำเข้า (Input)
d คือ ค่าเป้าหมายหรือผลลัพธ์ที่ต้องการ (Desired Output)

2. กำหนดค่าถ่วงน้ำหนักให้กับแต่ละเส้นของทุกๆชั้น โดยการสุ่มตัวเลขที่มีค่าเข้าใกล้ศูนย์
3. เลือกข้อมูลนำเข้าและค่าเป้าหมายออกมาที่ละคู่ แล้วคำนวณค่าผลลัพธ์จาก f(net) ในแต่ละ Node แล้วเคลื่อนไปข้างหน้า จนถึง Output Layer
4. คำนวณค่า Error ที่เกิดขึ้น

$$E \leftarrow E + \frac{1}{2}(d_k - o_k)^2$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดย d_k คือ ค่าเป้าหมายหรือผลลัพธ์ที่ต้องการตัวที่ k

o_k คือ ค่าผลลัพธ์ที่โครงข่ายคำนวณได้ ตัวที่ k

5. ถ้าผลลัพธ์ที่คำนวณได้ต่างจากค่าเป้าหมาย ต้องคำนวณค่าความผิดพลาด (Error signal) ที่เกิดขึ้นทั้งในชั้นเอาท์พุทและชั้นซ่อน แล้วย้อนกลับเข้าไปในโครงข่าย เพื่อสอนให้โครงข่ายเรียนรู้ข้อผิดพลาดที่เกิดขึ้น ซึ่งสมการที่ใช้จะแตกต่างกัน ถ้าใช้ฟังก์ชันกระตุ้นต่างกัน โดย

- กรณีใช้ Bipolar binary activation function

$$\delta_{ok} = \frac{1}{2}\left[(d_k - o_k)(1 - o_k^2)\right], \rightarrow k = 1,2,3,..,K$$

โดย δ_{ok} คือ ค่าความผิดพลาดในชั้นเอาท์พุท ตัวที่ k

d_k คือ ค่าเป้าหมายหรือผลลัพธ์ที่ต้องการตัวที่ k

o_k คือ ค่าผลลัพธ์ที่โครงข่ายคำนวณได้ ตัวที่ k

$$\delta_{yi} = \frac{1}{2}(1 - y_j^2)\sum_{k=1}^{K} \delta_{ok} w_{kj}, \rightarrow \quad j = 1,2,3,..., \ J$$

โดย δ_{yj} คือค่าความผิดพลาดของโหนดในชั้นซ่อน ตัวที่ k

y_j คือ ค่าผลลัพธ์ของโหนดในชั้นซ่อน

W_{kj} คือ ค่าถ่วงน้ำหนักระหว่างชั้นเอาท์พุทตัวที่ k และชั้นซ่อนตัวที่ j

- Unipolar binary activation function

$$\delta_{ok} = \left[(d_k - o_k)(1 - o_k)o_k\right], \rightarrow k = 1,2,3,..., K$$

โดย δ_{ok} คือ ค่าความผิดพลาด (Error Signal) ของชั้นเอาท์พุทตัวที่ k

d_k คือ ค่าเป้าหมายหรือผลลัพธ์ที่ต้องการตัวที่ k

o_k คือ ค่าผลลัพธ์ที่โครงข่ายคำนวณได้ ตัวที่ k

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\delta_{yi} = y_i(1-y_j)\sum_{k=1}^{K}\delta_{ok}w_{kj}, \rightarrow \ j = 1,2,3,\ldots, J$$

โดย δ_{yj} คือค่าความผิดพลาดของโหนดในชั้นซ่อน ตัวที่ k

y_j คือ ค่าผลลัพธ์ของโหนดในชั้นซ่อน

W_{kj} คือ ค่าถ่วงน้ำหนักระหว่างชั้นเอาท์พุทตัวที่ k และชั้นซ่อน ตัวที่ j

6. ปรับค่าถ่วงน้ำหนักที่ชั้นเอาท์พุทให้สอดคล้องกับ Error

$$w_{kj} \leftarrow w_{kj} + \eta\delta_{ok}y_j, \rightarrow \ k = 1,2,3,\ldots, K$$
$$j = 1,2,3,\ldots,J$$

โดย W_{kj} คือ ค่าถ่วงน้ำหนักระหว่างชั้นเอาท์พุทตัวที่ k และชั้นซ่อนตัวที่ j

η คือ ค่าอัตราการเรียนรู้ (Leaning rate) มีค่าอยู่ระหว่าง 0 – 1 โดยปกติจะมีค่าเท่ากับ 0.1 หรือ 0.01

δ_{ok} คือ ค่าความผิดพลาด (Error Signal) ของชั้นเอาท์พุทตัวที่ k

y_j คือ ค่าผลลัพธ์ของโหนดในชั้นซ่อนตัวที่ j

7. ปรับค่าถ่วงน้ำหนักที่ชั้นซ่อน

$$v_{ji} \leftarrow v_{ji} + \eta\delta_{yi}z_i, \rightarrow \ j = 1,2,3,\ldots, J$$
$$i = 1,2,3,\ldots,I$$

โดย V_{ji} คือ ค่าถ่วงน้ำหนักระหว่างชั้นอินพุทตัวที่ j และชั้นซ่อนตัวที่ i

η คือ ค่าอัตราการเรียนรู้ (Leaning rate) มีค่าอยู่ระหว่าง 0 – 1 โดยปกติจะมีค่าเท่ากับ 0.1 หรือ 0.01

δ_{yj} คือ ค่าความผิดพลาด (Error Signal) ของชั้นซ่อนตัวที่ j

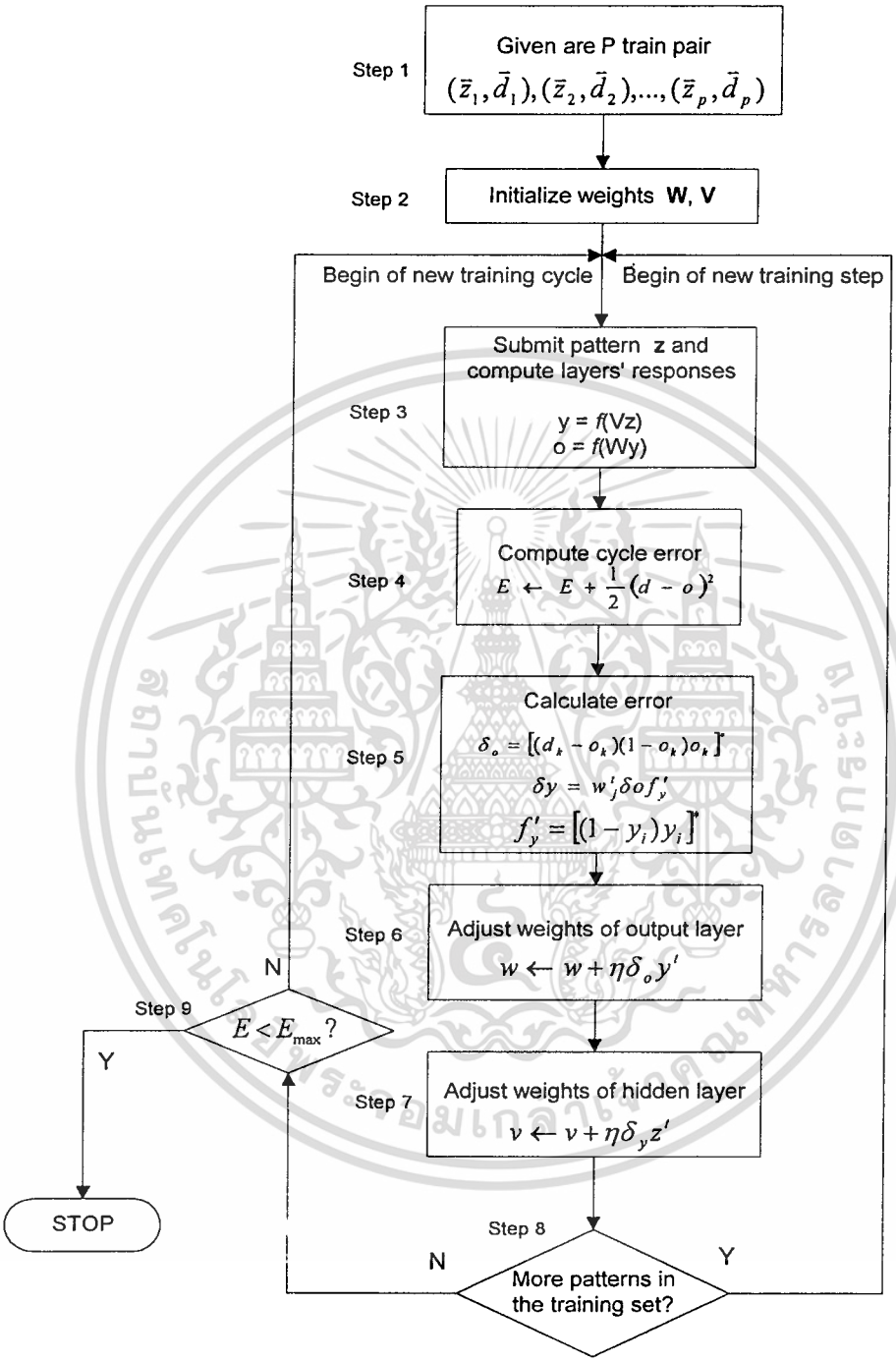
Z_i คือ ข้อมูลนำเข้า (Input) ตัวที่ i

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8. กลับไปทำขั้นตอนที่ 3 ถ้ายังมีข้อมูลที่ใช้ในการสอนอยู่ (จำนวนคู่ของข้อมูลนำเข้า (input) และ เอาท์พุทที่ต้องการ (desire Output) < P คู่) ถ้าไม่มีไปทำขั้นตอนต่อไป
9. ตรวจสอบค่า Error ในการสอนที่เกิดขึ้น ถ้ามีค่าน้อยกว่าที่กำหนดก็จะหยุดการสอน แต่ถ้าไม่ใช่กลับไปทำขั้นตอนที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.



*If *f(net)* given by *Bipolar binary activation function* is used in Step 2, then in Step 4 use

$$\delta_o = \frac{1}{2}[(d_k - o_k)(1 - o_k^2)]$$

$$f_y' = \frac{1}{2}[(1 - y_i^2)]$$

รูปที่ 2.8 Back-propagation Training Algorithm

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การประยุกต์ใช้ดาต้าไมนิ่งเพื่อการออกแบบและพัฒนาระบบงาน

ในการนำดาต้าไมนิ่งมาประยุกต์ใช้ในโครงงาน มีขั้นตอนดังต่อไปนี้

3.1. กำหนดวัตถุประสงค์

การนำดาต้าไมนิ่งมาประยุกต์ใช้ก็เพื่อพัฒนาระบบงานที่สามารถทำการพยากรณ์กลุ่มบริการเสริมที่เหมาะสมให้กับลูกค้า เนื่องมาจากความต้องการขององค์กรในการที่หาวิธีที่จะจัดชุดบริการเสริมที่เหมาะสมให้กับลูกค้าแต่ละราย

3.2. จัดเตรียมข้อมูล

ในการจัดเตรียมข้อมูลประกอบด้วย 3 ขั้นตอน คือ

3.2.1. คัดเลือกข้อมูล

ทำการคัดเลือกเฉพาะข้อมูลลูกค้าที่เปิดใช้บริการในปี พ.ศ. 2543 โดยนำข้อมูลจากฐานข้อมูลระบบการคิดค่าบริการ (Billing System) ทั้งหมดตั้งแต่เปิดบริษัท และคัดเลือกเฉพาะคุณลักษณะที่เกี่ยวข้องกับข้อมูลลูกค้า คือ

- รหัสลูกค้า
- หมายเลขเพจเจอร์
- รหัสรุ่นของเครื่องที่ใช้
- วันที่เปิดใช้บริการเสริม
- วัน/เดือน/ปี เกิดของลูกค้า
- เพศ
- การศึกษา
- รายได้
- ที่อยู่ ประกอบด้วยรหัสจังหวัด, รหัสอำเภอ และรหัสตำบล
- รหัสบริการที่ลูกค้าแต่ละรายใช้ โดยบริการเสริมของบริษัทมีทั้งสิ้น 13 รายการ ดังตารางที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัส	รายละเอียด
491809	บริการเชื่อมโยงมือถือ
491820	บริการ VOICE MAIL
491821	บริการ MEET ME
491822	บริการดูดวงประจำวัน
491823	บริการฝากข้อความ
491824	บริการข่าว 100 ช่อง
491826	บริการโชว์เบอร์โทรศัพท์เรียกเข้า
491827	บริการ INFO ON DEMND
491828	บริการกดหมายเลขวิทยุติดตามตัวผ่านโทรศัพท์
491829	บริการข่าวสาร
491830	บริการตู้เพลงผ่านวิทยุติดตามตัว
491831	บริการเลขาส่วนตัว
491833	บริการ E-mail Alert

ตารางที่ 3.1 รายการบริการเสริม

3.2.2. ตรวจสอบข้อมูล

จากข้อมูลลูกค้าที่ได้คัดเลือกมามีลูกค้าเปิดใช้บริการในช่วงปี 2543 ประมาณ 40,000 ราย และลูกค้าแต่ละรายสามารถมีหมายเลขเพจเจอร์ได้มากกว่า 1 หมายเลข ได้ทำการพิจารณาแต่ละคุณลักษณะที่เกี่ยวข้องกับข้อมูลลูกค้า ดังนี้

- การศึกษา หลังจากตรวจสอบข้อมูลแล้ว มีการป้อนข้อมูลแค่ 30% ของข้อมูลทั้งหมด และ 50 % ของข้อมูลที่ป้อนมีค่าเท่ากับ 1 คือไม่ระบุการศึกษา จึงได้ตัดข้อมูลส่วนนี้ทิ้งไม่นำมาพิจารณา
- รายได้ ข้อมูลรายได้มีค่าตั้งแต่ 5 – 250,000 บาท รายการที่มีข้อมูลมีแค่ 6% ของข้อมูลทั้งหมด จึงได้ตัดข้อมูลส่วนนี้ทิ้งไม่นำมาพิจารณาเช่นกัน
- อายุ พิจารณาเฉพาะอายุในช่วง 11 – 70 ปี นอกนั้นตัดทิ้งทั้งรายการ เนื่องจากข้อมูลอายุที่ไม่อยู่ในช่วงนี้มีสัดส่วนที่น้อยเมื่อเทียบกับข้อมูลทั้งหมด และยังมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โอกาสที่จะเป็นข้อมูลที่ผิดพลาดสูงด้วย เช่น เป็นค่าติดลบ หรืออายุที่มากกว่า 100 ปี

- สำหรับข้อมูลในฟิลด์อื่นๆ ที่มีข้อมูลไม่ครบ ก็ได้ทำการตัดรายการนั้นๆ ทิ้งไปทั้งรายการ ซึ่งทำให้เหลือข้อมูลที่ครบถ้วนสมบูรณ์ 24,945 รายการ
- รหัสบริการที่ลูกค้าใช้ ใน 1 หมายเลขเพจเจอร์ จะมีบริการเสริมได้มากกว่า 1 รายการ จึงได้ทำการจัดบริการเสริมเป็นชุดจากข้อมูลที่ลูกค้าใช้จริง จากข้อมูลก่อนตัดรายการที่ไม่สมบูรณ์ออก ซึ่งมีทั้งสิ้นประมาณ 40,000 รายการ ได้จำนวนชุดที่แตกต่างกัน 30 ชุด แต่หลังจากตัดข้อมูลที่ไม่สมบูรณ์ออกแล้วเหลือจำนวนชุดบริการเสริม 25 ชุด แต่ชุดมีลูกค้าใช้บริการในจำนวนที่แตกต่างกันไป สำหรับชุดที่มีลูกค้าใช้เป็นจำนวนมากกว่าชุดอื่นๆ มีเพียง 3 ชุดเท่านั้น คือ

  ชุดที่ 1 ประกอบด้วย บริการข่าว 100 ช่อง ลูกค้าใช้บริการ จำนวนทั้งสิ้น 7,425 รายการ

  ชุดที่ 2 ประกอบด้วย บริการข่าว 100 ช่องและบริการโชว์เบอร์โทรศัพท์เรียกเข้า จำนวนทั้งสิ้น 13,851 รายการ

  ชุดที่ 3 ประกอบด้วย บริการข่าว, บริการโชว์เบอร์โทรศัพท์เรียกเข้า และบริการ E-mail Alert จำนวนทั้งสิ้น 780 รายการ

  ส่วนข้อมูลบริการเสริมชุดอื่นๆ ซึ่งมีลูกค้าใช้บริการกระจัดกระจายในจำนวนไม่มากนัก จะไม่นำมาพิจารณา

3.2.3. ปรับเปลี่ยนรูปแบบของข้อมูล

พิจารณาแต่ละรายการซึ่งก็คือ 1 หมายเลยเพจเจอร์แล้วทำการปรับเปลี่ยนค่า ดังนี้

- ข้อมูลรหัสรุ่นของเครื่องที่ใช้ แต่ละรหัสจะแยกตามยี่ห้อ, รุ่น และสี ได้ทำการแปลงค่าจากรหัสรุ่นที่มีอยู่ทั้งหมดเป็นตัวเลขตามลำดับ เช่น ‘C01’ แปลงเป็น 1, ‘C02’ แปลงเป็น 2 เป็นต้น มีทั้งสิ้น 37 รุ่น
- วันที่เปิดใช้บริการเสริม นำเฉพาะค่าเดือนที่เปิดใช้บริการมาใช้
- วัน/เดือน/ปี เกิดของลูกค้า นำมาคำนวณอายุลูกค้า
- เพศ จากข้อมูล ‘M’ หมายถึง เพศชาย และ ‘F’ หมายถึง เพศหญิง ทำการแปลงค่า ‘M’ เป็น 1 และ ‘F’ เป็น 2
- ที่อยู่ ประกอบด้วยรหัสจังหวัด, รหัสอำเภอ และรหัสตำบล อย่างละ 2 หลัก นำทั้ง 3 รหัสมารวมเป็นรหัสเดียวกันเป็นจำนวน 6 หลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับข้อมูลนำเข้าทั้ง 5 คอลัมน์นั้น เมื่อได้ทำการปรับเปลี่ยนเป็นค่าตัวเลขแล้ว จึงทำการปรับลดสัดส่วนของข้อมูลให้อยู่ในช่วง 0 ถึง 1 เนื่องจากข้อจำกัดของฟังก์ชันกระตุ้นซิกมอยด์ฟังก์ชัน โดยการนำค่าสูงสุดในแต่ละฟิลด์มาหารทั้งคอลัมน์ ส่วนค่าเป้าหมาย คือชุดบริการเสริมทั้งสามชุด ทำการแปลงค่าชุดบริการเสริมโดยการแยกคอลัมน์เป้าหมายเป็น 3 คอลัมน์ แล้วพิจารณาแต่ละรายการโดยรายการใดที่มีค่าชุดบริการเสริมเป็นชุดที่ 1 ค่าเป้าหมายคอลัมน์แรกมีค่าเป็น 1 ส่วนคอลัมน์อื่นๆมีค่าเป็น 0 เช่นเดียวกับรายการที่มีค่าชุดบริการเสริมเป็นชุดที่ 2 และ 3 ค่าเป้าหมายในคอลัมน์ที่ 2 และ 3 มีค่าเป็น 1 ส่วนคอลัมน์อื่นมีค่าเป็น 0 ตามลำดับ

จากนั้นสุ่มเลือกรายการมา 1050 รายการ แบ่งเป็นข้อมูลสำหรับการสอน 750 รายการ และข้อมูลสำหรับการทดสอบ 300 รายการ

3.3. การนำข้อมูลมาทำดาต้าไมนิ่ง

เมื่อได้ทำการจัดเตรียมข้อมูลเรียบร้อยแล้ว ก็เป็นขั้นตอนการทำดาต้าไมนิ่ง ซึ่งวิธีการที่ใช้สำหรับการพยากรณ์ชุดบริการเสริม คือ การสร้างแบบจำลองพยากรณ์ (Predictive Modeling) โดยเป็นการพยากรณ์กลุ่มของข้อมูล (Classification) โดยใช้อัลกอลิทึ่มการแพร่ย้อนกลับ (Back-propagation Algorithm) ซึ่งเป็นอัลกอลิทึ่มหนึ่งสำหรับสร้างแบบจำลองพยากรณ์โดยใช้โครงข่ายประสาทเทียม (Neural Network) และใช้ฟังก์ชันกระตุ้นซิกมอยด์ฟังก์ชั่น (Sigmoid Function) แบบ Unipolar Binary Activation Function ซึ่งค่าที่ผ่านฟังก์ชันจะถูกแปลงให้อยู่ในในช่วง 0 - 1 แบ่งการทำงานในขั้นตอนนี้ออกเป็น 2 ส่วน คือ

3.3.1. การพัฒนาระบบงานเพื่อสร้างแบบจำลองพยากรณ์

สำหรับเครื่องมือที่ใช้ในการพัฒนาระบบงานเพื่อสร้างแบบจำลองพยากรณ์ในโครงงานฉบับนี้ ได้เลือกใช้โปรแกรมวิชวลเบสิค เวอร์ชัน 6 (Visual Basic Version 6) ในการเขียนโปรแกรม โดยระบบงานที่ได้พัฒนาแบ่งออกเป็น 4 ส่วน คือ

3.3.1.1. ส่วนการสร้างแบบจำลองพยากรณ์

3.3.1.2. ส่วนการพยากรณ์บริการเสริมให้ลูกค้า (Prediction)

3.3.1.3. การตั้งค่าให้กับโครงข่ายประสาทเทียม (Neural Network setup)

3.3.1.4. การตั้งค่าสำหรับเลือกข้อมูลนำเข้าและค่าเป้าหมายหรือผลลัพธ์ที่ต้องการ (Input and output setup)

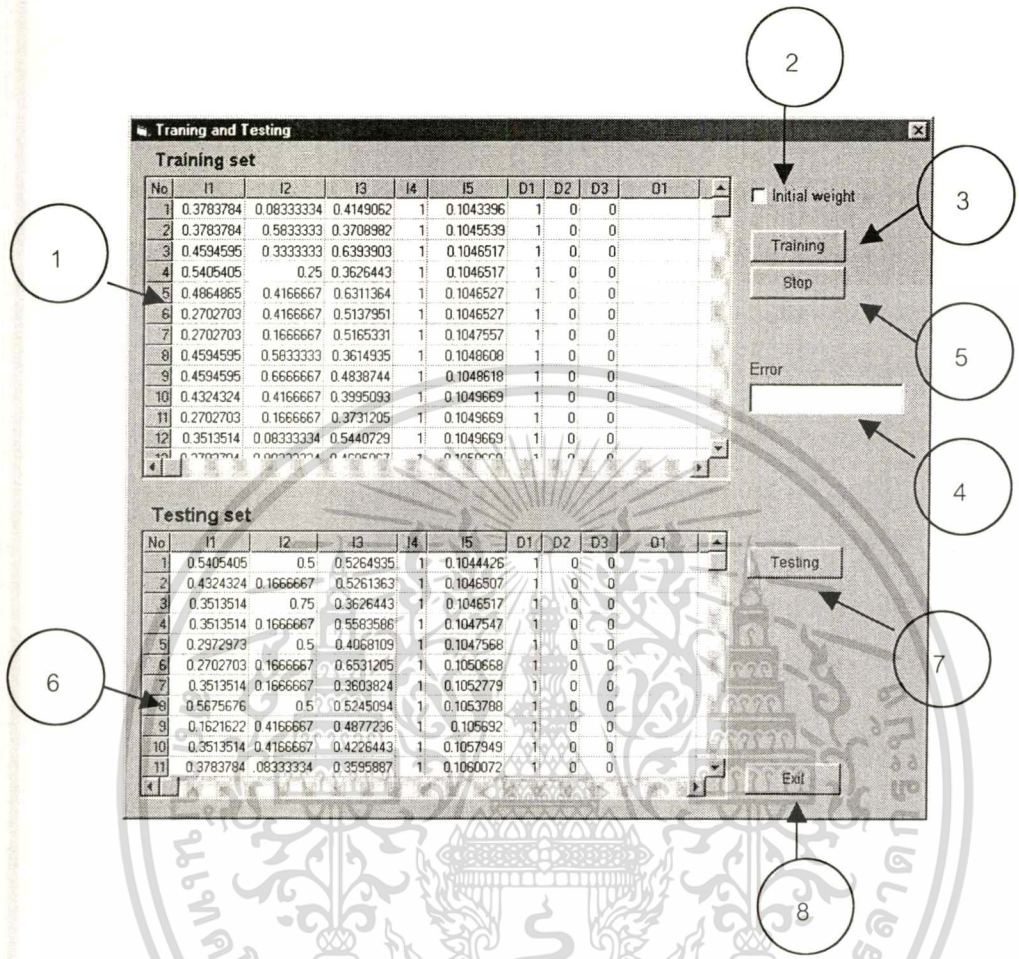
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.1.1. ส่วนการสร้างแบบจำลองพยากรณ์

สำหรับการสร้างแบบจำลองพยากรณ์โดยการประยุกต์ใช้โครงข่ายประสาทเทียม(Neural Network) นี้ จะใช้อัลกอลิทึ่มแบบการแพร่ย้อนกลับ (Back-propagation Algorithm) โดยในหน้าจอสำหรับสร้างแบบจำลองพยากรณ์จะมีฟังก์ชันการทำงาน 2 ส่วน คือ การสอนโครงข่ายให้เรียนรู้จากข้อมูลชุดฝึกสอน (Training set) และการทดสอบความถูกต้องในการทำนายผลลัพธ์ของโครงข่ายด้วยข้อมูลชุดทดสอบ (Testing set) โดยโปรแกรมจะทำการอ่านข้อมูลทั้ง 2 ชุด จากไฟล์ข้อความ (Text file) ซึ่งได้มีการจัดเตรียมไว้ก่อน ซึ่งรายละเอียดหน้าจอ มีดังนี้

1. ตารางข้อมูลฝึกสอนโครงข่าย (Training set) ซึ่งจะประกอบไปด้วยข้อมูลนำเข้า(Input), และค่าเป้าหมายหรือผลลัพธ์ที่ต้องการ (Desired output) ซึ่งจากหน้าจอในตารางของชุดข้อมูล ค่าข้อมูลนำเข้าจะอยู่ในคอลัมน์ I1 – I5 และค่าเป้าหมาย จะอยู่ในคอลัมน์ D1 – D3 ส่วนคอลัมน์ O1 – O3 จะเป็นค่าผลลัพธ์ที่โครงข่ายคำนวณให้
2. การกำหนดค่าถ่วงน้ำหนักเบื้องต้น (Initial weight) เลือกเมื่อต้องการให้โปรแกรมทำการสุ่มค่าถ่วงน้ำหนักใหม่
3. ปุ่มสำหรับสอนโครงข่าย (Training button) โปรแกรมจะทำการอ่านชุดข้อมูลฝึกสอนโครงข่ายที่ละรายการ แล้วคำนวณค่าต่างตามอัลกอลิทึ่ม จากนั้นจะแสดงผลลัทธ์ในตารางของชุดข้อมูลฝึกสอนต่อไป
4. ค่าความผิดพลาด (Error) แสดงค่าผลรวมของความผิดพลาดในแค่ละรอบของการฝึกสอน
5. ปุ่มสำหรับหยุดการสอน (Stop button) กดปุ่มนี้เมื่อต้องการหยุดสอน โปรแกรมจะทำการเก็บค่าถ่วงน้ำหนักไว้ในเท็กซ์ไฟล์ เพื่อนำไปใช้ในการทำนายต่อไป
6. ตารางข้อมูลทดสอบโครงข่าย (Testing set) เช่นเดียวกับตารางข้อมูลฝึกสอน คือ จะแสดงค่าข้อมูลนำเข้า(Input), และค่าเป้าหมายหรือผลลัพธ์ที่ต้องการ ส่วนค่าผลลัพธ์ที่โครงข่ายคำนวณจะแสดงเมื่อทำการคำนวณเสร็จแล้ว
7. ปุ่มสำหรับทดสอบค่าโครงข่าย โปรแกรมจะทำการอ่านข้อมูลชุดทดสอบที่ละรายการ แล้วคำนวณค่าผลลัพธ์เพื่อนำไปแสดงในตารางต่อไป
8. ปุ่มออกจากโปรแกรม (Exit)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 หน้าจอสำหรับการสอนและทดสอบแบบจำลองโครงข่ายประสาทเทียม

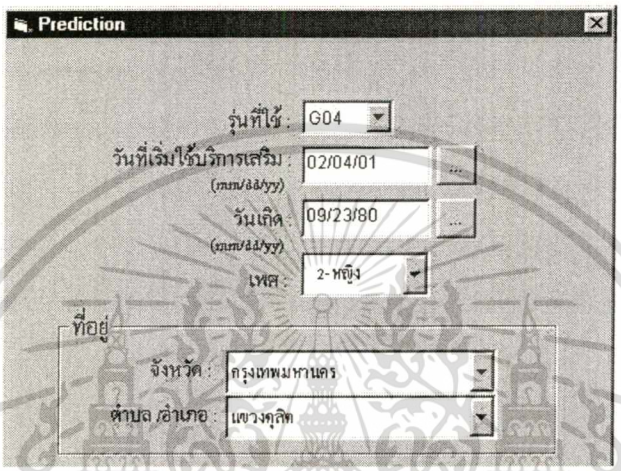
3.3.1.2. ส่วนการพยากรณ์บริการเสริมให้ลูกค้า (Prediction)

หน้าจอนี้ใช้สำหรับทำนายกลุ่มบริการให้ลูกค้า โดยใช้แบบจำลองโครงข่ายประสาทเทียมที่ได้ทำการฝึกสอนและทดสอบโครงข่ายจนได้ค่าความผิดพลาดอยู่ในช่วงที่ยอมรับได้แล้ว ข้อมูลที่ป้อนต้องเป็นข้อมูลประเภทเดียวกับที่ใช้ในการฝึกสอนและทดสอบโครงข่าย คือ

- ข้อมูลรหัสรุ่นของเครื่องที่ใช้
- วันที่เปิดใช้บริการเสริม
- วัน/เดือน/ปี เกิดของลูกค้า
- เพศ
- ที่อยู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมจะทำการอ่านข้อมูลแล้วจะทำการแปลงให้อยู่ในรูปแบบที่เหมาะสมกับอัลกอลิทึ่ม จากนั้นจึงจะพยากรณ์กลุ่มบริการเสริมให้กับลูกค้ารายนี้ต่อไป

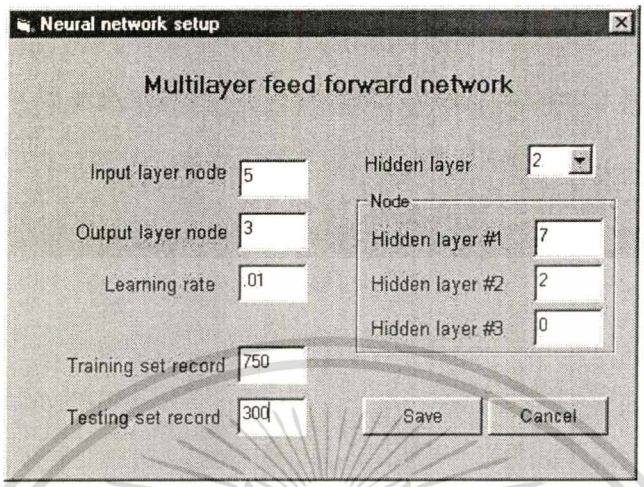


รูปที่ 3.2 หน้าจอสำหรับการพยากรณ์บริการเสริมให้ลูกค้า

3.3.1.3. การตั้งค่าให้กับโครงข่ายประสาทเทียม (Neural Network setup) เป็นการตั้งค่าที่จำเป็นสำหรับโครงข่าย ดังนี้

- จำนวนฟิลด์ข้อมูลนำเข้า (Input layer node)
- จำนวนกลุ่มเป้าหมายหรือกลุ่มผลลัพธ์ที่ต้องการ (Output layer node)
- จำนวนชั้นของชั้นซ่อน (Hidden layer) มีได้ไม่เกิน 3 ชั้น
- จำนวนโหนดของชั้นซ่อนชั้นที่ 1,2 และ 3 ตามลำดับ
- อัตราการเรียนรู้ (Learning rate) มักกำหนดเป็นค่าน้อยๆ เช่น 0.1, 0.01 เป็นต้น
- จำนวนรายการข้อมูลชุดฝึกสอน (Training set record)
- จำนวนรายการข้อมูลชุดทดสอบ (Testing set record)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



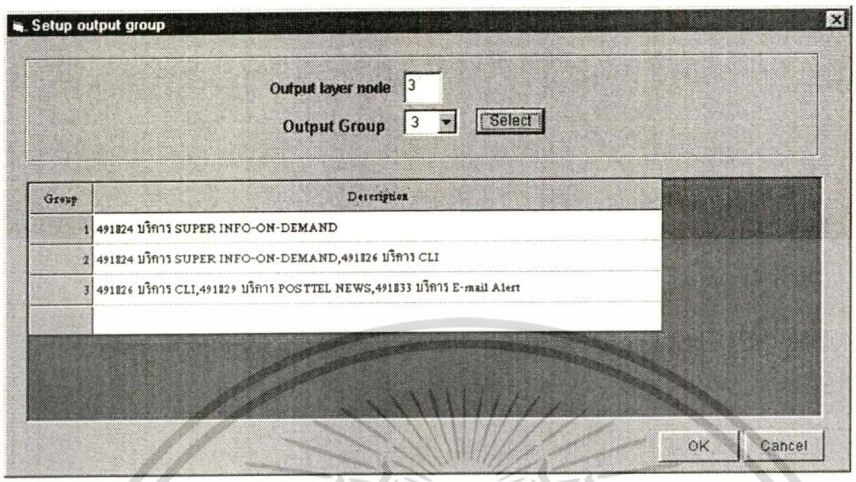
รูปที่ 3.3 หน้าจอสำหรับการตั้งค่าให้กับโครงข่ายประสาทเทียม

3.3.1.4. การตั้งค่าสำหรับเลือกข้อมูลนำเข้าและค่าเป้าหมายหรือผลลัพธ์ (Input and output setup)

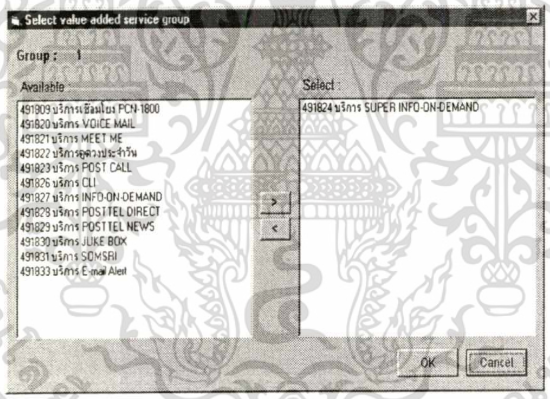
ในกรณีที่ต้องการสร้างแบบจำลองการพยากรณ์บริการเสริมจากฐานข้อมูลจะสามารถทำได้โดยเลือกการตั้งค่าข้อมูลนำเข้าและกลุ่มเป้าหมาย ซึ่งจะสามารถเลือกฟิลด์ที่ต้องการใช้เป็นข้อมูลนำเข้า, จัดกลุ่มบริการเสริมซึ่งเป็นค่าเป้าหมาย และสามารถปรับเปลี่ยนข้อมูลได้ตามต้องการ โดยมีขั้นตอนการทำงานดังนี้

1. กำหนดจำนวนกลุ่มของบริการเสริมซึ่งจะนำไปใช้เป็นค่าเป้าหมายสำหรับโครงข่ายประสาทเทียม โดยจะสามารถกำหนดจำนวนชุดบริการเสริมและแสดงรายละเอียดบริการเสริมในแต่ละชุดได้จากหน้าจอในรูปที่ 3.4 และกำหนดบริการเสริมสำหรับแต่ละชุดหน้าจอ ดังรูป 3.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



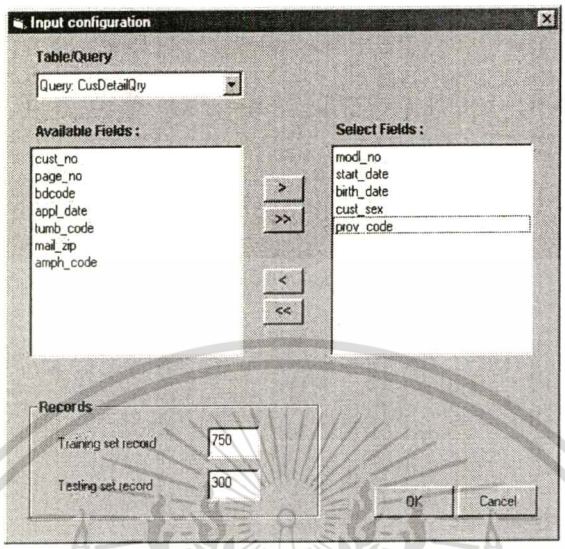
รูปที่ 3.4 หน้าจอสำหรับการกำหนดกลุ่มบริการเสริม



รูปที่ 3.5 หน้าจอสำหรับการเลือกบริการเสริมในแต่ละกลุ่ม

2. เลือก Table หรือ Query ที่ต้องการกำหนดให้เป็นข้อมูลนำเข้าจากฐานข้อมูล จากนั้นทำการเลือกฟิลด์ที่ต้องการ กำหนดจำนวนรายการข้อมูลที่ต้องการให้เป็นข้อมูลชุดฝึกสอนและชุดทดสอบ

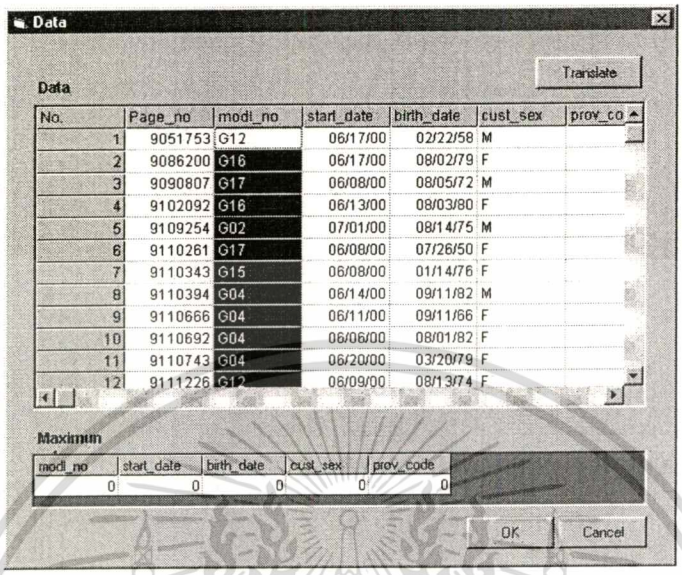
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 หน้าจอการกำหนดข้อมูลนำเข้า

3. เมื่อทำการจัดกลุ่มบริการเสริมแล้ว โปรแกรมจะทำการดึงข้อมูลจากฐานข้อมูลมาแสดง ดังรูปที่ 3.7 ซึ่งในขั้นตอนต่อไปคือ ต้องแปลงข้อมูลอยู่ให้ในรูปแบบที่อัลกอลิทึ่มต้องการ ซึ่งมีอยู่ 2 ขั้นตอน คือ
   - แปลงค่าข้อมูลนำเข้าให้เป็นตัวเลขทั้งหมด ทำได้โดยเลือกฟิลด์ที่ต้องการจากตารางข้อมูล จากนั้นกดปุ่มแปลงข้อมูล (Translate button) ในกรณีข้อมูลในฟิลด์นั้นมีชนิดเป็นข้อความ จะปรากฏหน้าจอที่ 3.8 และถ้าเป็นวันที่ก็จะปรากฏหน้าจอที่ 3.9
   - แปลงข้อมูลให้อยู่ในช่วง 0 – 1 โดยนำค่าสูงสุดในแต่ละฟิลด์หรือค่าที่ผู้ใช้กำหนดมาหารทั้งคอลัมน์

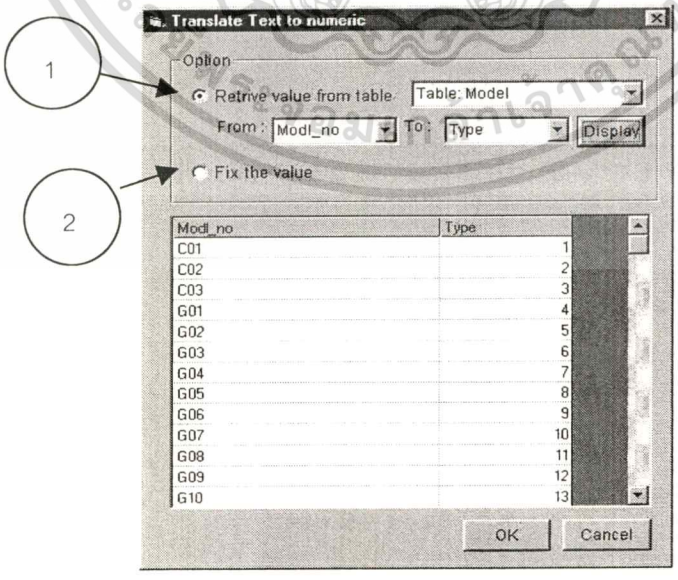
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 หน้าจอแสดงรายการข้อมูลที่เลือก

สำหรับการแปลงข้อมูลชนิดข้อความให้เป็นตัวเลขสามารถทำได้ 2 ทางเลือก คือ

1. โดยใช้ข้อมูลจากฐานข้อมูลโดยกำหนดชื่อตารางแล้วเลือกชื่อฟิลด์ต้นแบบและฟิลด์ค่าที่เป็นตัวเลขตามต้องการ
2. ป้อนข้อมูลจากหน้าจอเอง

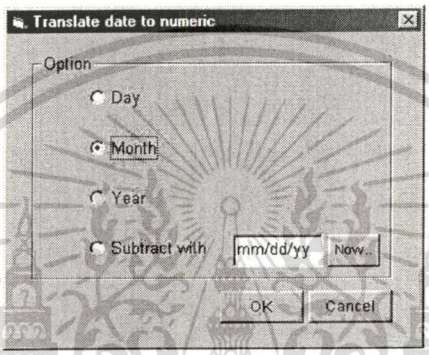


รูปที่ 3.8 หน้าจอการแปลงข้อความเป็นตัวเลข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับการข้อมูลวันที่ให้เป็นตัวเลข มี 4 ฟังก์ชัน ที่สามารถทำได้ คือ คือ

- เลือกใช้เฉพาะวันที่
- เลือกใช้เฉพาะเดือน
- เลือกใช้เฉพาะปี
- คำนวณระยะเวลาจนถึงวันที่ที่ต้องการ โดยคำนวณออกมาเป็นปี



รูปที่ 3.9 หน้าจอการแปลงวันที่เป็นตัวเลข

3.3.1.1. เมื่อเลือกข้อมูลและทำการแปลงข้อมูลแล้ว โปรแกรมจะทำการบันทึกข้อมูลลงแฟ้มข้อมูลแบบข้อความ (Text file) โดยแยกออกเป็น 2 แฟ้ม คือ แฟ้มข้อมูลสำหรับฝึกสอนโครงข่าย ชื่อ *Train.txt* และแฟ้มข้อมูลสำหรับทดสอบโครงข่าย ชื่อ *Test.txt*

3.3.2. การฝึกสอนและทดลองแบบจำลองพยากรณ์

นำข้อมูลที่ได้คัดเลือกในขั้นตอน 3.2 มาใช้ในการฝึกสอนและทดสอบโครงข่าย โดยกำหนดจำนวนโหนดต่างๆ ดังนี้

- จำนวนโหนดข้อมูลนำเข้า (Input layer) เท่ากับ 5 โหนด ซึ่งก็คือ
  - ข้อมูลรหัสรุ่นของเครื่องที่ใช้
  - เดือนที่เปิดใช้บริการเสริม
  - อายุลูกค้า
  - เพศ
  - ที่อยู่ ประกอบด้วยรหัสจังหวัด, รหัสอำเภอ และรหัสตำบล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- จำนวนชั้นซ่อน (Hidden Layer) 2 ชั้น โดยชั้นที่ 1 มีจำนวนโหนดเท่ากับ 4 โหนด และชั้นที่ 2 มีจำนวนโหนดเท่ากับ 5 โหนด
- จำนวนโหนดค่าเป้าหมาย (Output layer) เท่ากับ 3 โหนด คือ ชุดบริการเสริมชุดที่ 1, 2 และ 3 ตามลำดับ

ทำการฝึกสอนโครงข่ายด้วยข้อมูลชุดฝึกสอน จนกระทั่งได้ค่าความผิดพลาดมีค่าคงที่เท่ากับ 2.5629 จึงได้หยุดทำการสอนโครงข่าย จากนั้นทำการทดสอบโครงข่ายโดยใช้ข้อมูลทดสอบต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
วิเคราะห์ผลการดำเนินการ

4.1. ผลที่ได้จากการดำเนินการ

ผลจากการทดสอบโครงข่ายโดยใช้ข้อมูลชุดทดสอบจำนวน 300 รายการ ประกอบด้วยรายการที่เป็นชุดบริการเสริมชุดที่ 1, 2 และ 3 อย่างละ 100 รายการ โดยมีตัวอย่างของชุดข้อมูลที่ใช้ทดสอบและผลลัพธ์ที่โครงข่ายคำนวณดังแสดงในตาราง 4.1 ซึ่งมีรายละเอียดแต่ละคอลัมน์ ดังนี้

- I1 – I5 คือค่าข้อมูลนำเข้า
- D1 – D3 คือค่าเป้าหมาย โดย D1 มีค่าเป็น 1 กรณีเป็นบริการเสริมชุดที่ 1
  D2 มีค่าเป็น 1 กรณีเป็นบริการเสริมชุดที่ 2
  D3 มีค่าเป็น 1 กรณีเป็นบริการเสริมชุดที่ 3
- O1 – O3 คือผลลัพธ์ที่โครงข่ายคำนวณให้

ปรากฏว่าผลลัพธ์ที่โครงข่ายทำนายค่าชุดบริการเสริมมาให้นั้น มีค่าใกล้เคียงกับค่าที่เป็นผลลัพธ์ที่แท้จริงมาก โดยพิจารณาที่คอลัมน์ O1 – O2 และตัดสินใจว่ารายการใดเป็นบริการเสริมชุดใดจากคอลัมน์ที่มีค่ามากที่สุด มีรายการที่ถูกต้องทั้งสิ้น 293 รายการ แบ่งเป็น

- รายการที่เป็นบริการเสริมชุดที่ 1 จำนวน 94 รายการ
- รายการที่เป็นบริการเสริมชุดที่ 2 จำนวน 98 รายการ
- รายการที่เป็นบริการเสริมชุดที่ 3 จำนวน 100 รายการ

และรายการที่โครงข่ายทำนายค่าเป้าหมายผิดจำนวนทั้งสิ้น 7 รายการ โดย

- รายการที่เป็นชุดบริการเสริมชุดที่ 1 ทำนายเป็น บริการเสริมชุดที่ 2 จำนวน 6 รายการ
- รายการที่เป็นชุดบริการเสริมชุดที่ 2 ทำนายเป็น บริการเสริมชุดที่ 3 จำนวน 2 รายการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

I1	I2	I3	I4	I5	D1	D2	D3	O1	O2	O3
0.3783784	0.1666667	0.4126443	1	0.8747523	1	0	0	0.999989	0.000004	0.000000
0.3783784	0.08333334	0.3505808	1	0.8749593	1	0	0	0.091683	0.934195	0.000000
0.3243243	0.08333334	0.4732792	1	0.9365548	1	0	0	0.999997	0.000001	0.000000
0.3243243	0.1666667	0.5928427	1	0.9373746	1	0	0	0.999998	0.000001	0.000000
0.4594595	0.25	0.4240332	1	0.9373746	1	0	0	0.999994	0.000002	0.000000
0.4594595	0.4166667	0.5986363	1	0.9987496	1	0	0	0.999998	0.000001	0.000000
0.05405406	0.4166667	0.3781205	1	0.9997909	1	0	0	0.999095	0.000403	0.000000
0.3513514	0.1666667	0.2776442	0.5	0.1044437	0	1	0	0.000032	0.999672	0.000000
0.3783784	0.4166667	0.3438346	1	0.1046517	0	1	0	0.006298	0.993922	0.000000
0.2702703	0.5	0.3204616	1	0.1046527	0	1	0	0.000000	0.998749	0.000218
0.2972973	0.1666667	0.3109378	1	0.1049669	0	1	0	0.000000	0.583959	0.405039
0.5675676	0.5	0.2319696	0.5	0.1051708	0	1	0	0.000000	0.008754	0.987375
0.5405405	0.5	0.3214537	1	0.1053799	0	1	0	0.000000	0.997708	0.000342
0.3513514	0.25	0.3428029	1	0.105692	0	1	0	0.003669	0.996263	0.000000
0.5135135	0.4166667	0.225939	0.5	0.7699997	0	0	1	0.000000	0.001887	0.999125
0.2702703	0.5833333	0.2245898	0.5	0.8331431	0	0	1	0.000000	0.000643	0.999745
0.02702703	0.5	0.2213755	0.5	0.8331442	0	0	1	0.000000	0.000042	0.999988
0.2972973	0.25	0.2198676	0.5	0.8343925	0	0	1	0.000000	0.000018	0.999995
0.4864865	0.4166667	0.2279628	0.5	0.8753765	0	0	1	0.000000	0.004080	0.998283
0.4054054	0.25	0.2245501	0.5	0.9372727	0	0	1	0.000000	0.000187	0.999945
0.5405405	0.5	0.2212168	0.5	0.9373777	0	0	1	0.000000	0.000067	0.999974

ตารางที่ 4.1 ตัวอย่างชุดข้อมูลทดสอบและผลลัพธ์ที่โครงข่ายคำนวณให้

4.2. การวัดความถูกต้องของแบบจำลอง

จากผลการทดสอบข้างต้น สามารถวัดความถูกต้องของแบบจำลองการจัดหมวดหมู่สำหรับการทำนายข้อมูลในแต่ละกลุ่ม โดยมีสูตรการคำนวณค่าความถูกต้องในการทำนาย ดังนี้

$$\Pr ecisian_i = \frac{t_rec_i}{rec_i}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดย $Precisian_i$ คือ ค่าความถูกต้องของกลุ่มที่ i

t_rec_i คือ จำนวนรายการที่ทายถูกว่าเป็นกลุ่ม i

rec_i คือ จำนวนรายการทั้งหมดของกลุ่ม i ที่ใช้ทดสอบ

- รายการที่เป็นบริการเสริมชุดที่ 1 จำนวน 94 รายการ

$$G1 = \frac{94}{100}$$

$$= 0.94$$

โดย G1 ค่าความถูกต้องในการทำนายว่าเป็นค่าชุดบริการเสริมชุดที่ 1

- รายการที่เป็นบริการเสริมชุดที่ 2 จำนวน 98 รายการ

$$G2 = \frac{98}{100}$$

$$= 0.98$$

โดย G2 ค่าความถูกต้องในการทำนายว่าเป็นค่าชุดบริการเสริมชุดที่ 2

- รายการที่เป็นบริการเสริมชุดที่ 3 จำนวน 100 รายการ

$$G3 = \frac{100}{100}$$

$$= 1$$

โดย G3 ค่าความถูกต้องในการทำนายว่าเป็นค่าชุดบริการเสริมชุดที่ 3

สำหรับค่าความถูกต้องโดยรวมของแบบจำลองในการทำนายกลุ่มหาได้จาก

$$Accuracy = \sum_{i=1}^{n} \left(\Pr ecisian_i \frac{rec_i}{total_rec} \right)$$

โดย Accuracy คือ ค่าความถูกต้องในการทำนายค่าของแบบจำลอง

$Precisian_i$ คือ ค่าความถูกต้องของกลุ่มที่ i

rec_i คือ จำนวนรายการที่ทายถูกว่าเป็นกลุ่ม i

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

total_rec คือ จำนวนรายการทั้งหมดที่ใช้ทดสอบ

$$Accuracy = 0.94\left(\frac{100}{300}\right) + 0.98\left(\frac{100}{300}\right) + 1\left(\frac{100}{300}\right)$$

$$Accuracy\ rate = 97.33\ \%$$

สรุปรายการที่โครงข่ายสามารถทำนายผลได้ถูกต้องคิดเป็น 97.33 เปอร์เซ็นต์ของข้อมูลทั้งหมด เมื่อพิจารณาแล้วเห็นว่าแบบจำลองนี้สามารถพยากรณ์ชุดบริการเสริมให้กับลูกค้าได้ใกล้เคียงกับชุดบริการเสริมที่ลูกค้าใช้อยู่จริงมาก จึงได้นำแบบจำลองนี้ไปใช้ในการพยากรณ์สำหรับลูกค้าแต่ละราย โดยต้องทำการป้อนข้อมูลนำเข้าชนิดเดียวกับข้อมูลนำเข้าในแบบจำลอง ซึ่งประกอบด้วย รุ่นเพจเจอร์ที่ใช้, วันที่เริ่มใช้บริการเสริม, วันเดือนปีเกิด และที่อยู่ เพื่อใช้ในการพยากรณ์ต่อไป ดังได้กล่าวมาแล้ว ในบทที่ 3 หัวข้อที่ 3.3.1.2. การพยากรณ์บริการเสริมให้ลูกค้า (Prediction)

4.3. ข้อดีและข้อเสียที่เกิดขึ้นจากการนำเทคนิคโครงข่ายประสาทเทียมมาใช้เพื่อการพยากรณ์

การนำเทคนิคโครงข่ายประสาทเทียม (Neural Network) มาใช้เพื่อการพยากรณ์ชุดบริการเสริมให้กับลูกค้านั้น เป็นวิธีหนึ่งซึ่งบริษัทสามารถนำไปใช้ในการประเมินความต้องการของลูกค้าเพื่อเป็นแนวทางในการจัดทำรายการส่งเสริมการขายได้ ซึ่งจากการทำงานจริงพบข้อดีและข้อเสียของโครงข่ายประสาทเทียม ดังนี้

ข้อดี

1. มีความสามารถในการเรียนรู้ สามารถพยากรณ์ข้อมูลได้อย่างถูกต้อง
2. มีความยืดหยุ่นในการใช้งาน สามารถปรับแต่งจำนวนชั้นซ่อน, จำนวนโหนดในชั้นซ่อน และค่าอัตราการเรียนรู้ เพื่อให้เหมาะสมกับรูปแบบของปัญหา

ข้อเสีย

1. ต้องใช้เวลานานในการฝึกสอนโครงข่ายให้เรียนรู้
2. เมื่อทำการฝึกสอนโครงข่ายไประยะหนึ่ง มีโอกาสที่จะเกิดความถูกต้องกับข้อมูลฝึกสอน (Overfitting) เท่านั้น จึงต้องมีการทดสอบเครือข่ายกับชุดข้อมูลทดสอบเพื่อดูค่าความผิดพลาดที่ได้ด้วย ถ้าค่าความผิดพลาดที่ได้จากชุดทดสอบมีค่ามากกว่าค่าความผิดพลาดที่ได้จากข้อมูลชุดสอนก็ต้องหยุดการสอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ในการทำงานของ Neural Network เปรียบเสมือนกับกล่องดำ คือ ไม่มีคำอธิบายถึงผลลัพธ์ที่ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปและข้อเสนอแนะ

5.1. สรุปผลการดำเนินงาน

สำหรับโครงงานพัฒนาระบบฉบับนี้มีวัตถุประสงค์หลักเพื่อที่จะนำเสนอและประยุกต์ใช้ดาต้าไมนิ่งในธุรกิจ ซึ่งดาต้าไมนิ่งนั้นเป็นกระบวนการที่ใช้เพื่อค้นหาข้อมูลที่มีประโยชน์ออกจากฐานข้อมูล เพื่อนำมาช่วยในการตัดสินใจ ซึ่งวิธีการแก้ปัญหาด้วยดาต้าไมนิ่งนั้นมีอยู่ด้วยกันหลายรูปแบบขึ้นอยู่กับวัตถุประสงค์ของการทำงาน โดยในโครงงานนี้ได้เสนอเทคนิคการสร้างแบบจำลองพยากรณ์ (Predictive Modeling) เพื่อทำนายชุดบริการเสริมให้กับลูกค้ารายใหม่จากข้อมูลลูกค้าของบริษัทที่มีอยู่ในระบบการคิดค่าบริการ (Billing System) โดยใช้อัลกอลิทึ่มการแพร่ย้อนกลับ (Back-propagation Algorithm) ซึ่งเป็นอัลกอลิทึ่มหนึ่งของโครงข่ายสมองเทียม (Neural Network) ที่มีการใช้งานกันอย่างกว้างขวาง อันเนื่องมาจากความมีประสิทธิภาพในการแก้ปัญหา, มีความยืดหยุ่นสูง และเข้าใจง่าย ซึ่งจากผลการศึกษาพบว่า โครงข่ายสามารถทำนายค่าผลลัพธ์ชุดบริการเสริมได้อย่างถูกต้องใกล้เคียงกับผลลัพธ์ที่แท้จริงมาก คือ รายการที่โครงข่ายสามารถทำนายได้ถูกต้องคิดเป็น 97.33 เปอร์เซ็นต์ ของข้อมูลที่ใช้ทดสอบ

5.2. ข้อเสนอแนะ

1. สำหรับการพยากรณ์ชุดบริการเสริมในโครงงานฉบับนี้เป็นการนำข้อมูลที่มีอยู่ในระบบมาฝึกสอนให้โครงข่ายประสาทเทียมได้เรียนรู้และทำการพยากรณ์นั้นจะสามารถทำนายได้เฉพาะชุดบริการเสริมที่บริษัทเคยจัดมาแล้วเท่านั้น เนื่องจากคุณสมบัติของโครงข่ายคือการเรียนรู้ข้อมูลในอดีตเพื่อนำมาใช้ในการแก้ปัญหาในปัจจุบัน ดังนั้นสำหรับบริการเสริมที่มีขึ้นใหม่ หรือชุดบริการเสริมที่บริษัทจะจัดขึ้นใหม่นั้น โครงข่ายจะไม่สามารถทำนายได้ ดังนั้นบริษัทอาจจะต้องทำการเก็บข้อมูลในส่วนนี้ไปสักระยะหนึ่งก่อนที่จะดึงข้อมูลมาใช้เพื่อฝึกสอนโครงข่ายต่อไป
2. อัลกอลิทึ่มแบบการแพร่ย้อนกลับในโครงงานฉบับนี้สามารถนำไปประยุกต์ใช้กับข้อมูลอื่นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. จากที่กล่าวมาแล้วข้างต้นว่าโครงงานมีวัตถุประสงค์หลักเพื่อที่จะนำเสนอและประยุกต์ใช้ดาต้าไมนิ่งในธุรกิจ โดยทำการสร้างแบบจำลองเพื่อทำนายชุดบริการเสริมให้กับลูกค้า ดังนั้นโครงงานนี้สามารถใช้เป็นแนวทางในการประยุกต์ใช้ดาต้าไมนิ่งกับธุรกิจอื่นๆได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Berry , J.M. and Linoff ,G. 1997. **Data Mining Techniques**. Canada : John Wiley & Sons.

Cabena , P. et. al. 1997. **Discovering Data Mining**. New Jersey : Prentice Hall.

Logical design consulting inc.,**THINKS and ThinksPro:Personal and Professional Neural Network Software for Windows** , Website available:<URL:www.sigma-research.com>.

Mitchell, T.M. 1997. **Machine leaning**. McGraw-Hill.

Skapura, D.M. 1996. **Building Neural Networks**. New York : Addison-Wesley.

Weiss, S.M. and Indurkhya, N. 1998. **Predictive data mining.** , San Francisco : Morgan Kaufmann.

Zirilli, S. Joseph. 1997. **Financial Prediction using Neural Networks**. London : International Thomson Computer Press.

Zurada, M. Jacek. 1992. **Artificial Neural Systems**. New York : West Publishing.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

นางสาว กิ่งกาญจน์ เหลืองพุทธรัตน์ เกิดวันที่ 23 สิงหาคม พ.ศ. 2512 สำเร็จการศึกษาปริญญาบริหารธุรกิจบัณฑิต เกียรตินิยมอันดับหนึ่ง จากสาขาคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจ มหาวิทยาลัยรังสิต ในปีการศึกษา 2534 และศึกษาต่อในระดับปริญญามหาบัณฑิต สาขาวิทยาการสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2542 ปัจจุบันทำงานในตำแหน่ง Senior Software Engineer ให้กับบริษัท Atos Origin (Thailand) Ltd.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้