ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

ระบบวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการระบบโทรศัพท์เคลื่อนที่

Data Mining Application in Mobile Fraud Management

โดย

นภาพร พันธ์โยธาชาติ

รหัส 45066078

H002217

อาจารย์ที่ปรึกษา

ดร. ภัทรชัย ลลิตโรจน์วงศ์

วัน เดือน ปี 08 ก.พ. 2550
เลขทะเบียน 02217
เลขเรียกหนังสือ วพ. น๑๑๙ ร 2547
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

b 11698556
i 1 2547 no

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชาโครงการพัฒนาระบบงาน

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 1 ปีการศึกษา 2547

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	ระบบวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการระบบโทรศัพท์เคลื่อนที่
นักศึกษา	นางสาว นภาพร พันธ์โยธาชาติ
อาจารย์ที่ปรึกษา	ดร. ภัทรชัย ลลิตโรจน์วงศ์
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2547

บทคัดย่อ

ธุรกิจการให้บริการโทรศัพท์เคลื่อนที่ ในปัจจุบันนั้นขยายตัวเพิ่มขึ้นอย่างรวดเร็ว และมีผู้ใช้บริการเป็นจำนวนมาก ปัญหาที่พบตามมาคือ มีผู้ทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่เพิ่มมากขึ้น จนกลายเป็นปัญหาตลอดระยะเวลา 10 ปีที่ผ่านมา การบริหารงานจัดการที่ไม่มีระเบียบรูปแบบของการให้บริการที่หลากหลาย และจำนวนของผู้ให้บริการระบบโทรศัพท์เคลื่อนที่ที่มีเพิ่มขึ้น ทำให้โอกาสในการทุจริตเพิ่มสูงขึ้น โดยเฉลี่ยแล้วธุรกิจให้บริการโทรศัพท์เคลื่อนที่และธุรกิจโทรคมนาคมอื่นๆ จะสูญเสียเงินจำนวน 3% ของรายได้ทั้งหมดต่อปี ให้กับการทุจริตในรูปแบบต่างๆ ซึ่งปัญหาเหล่านี้ถือเป็นเรื่องใหญ่ในการดำเนินธุรกิจ ซึ่งจำเป็นที่จะต้องได้รับการแก้ไข ดังนั้นผู้ให้บริการระบบโทรศัพท์เคลื่อนที่ จึงพยายามหาวิธีการจัดการกับผู้ที่ทำการทุจริต (Fraud Management) เพื่อตรวจสอบพฤติกรรมการทุจริต และหาวิธีป้องกันแก้ไขต่อไป

โครงการนี้เป็นการพัฒนาโปรแกรมสำหรับการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการระบบโทรศัพท์เคลื่อนที่ โดยนำทฤษฎีของคาต้าไมน์นิ่งและหลักการของเครือข่ายประสาทเทียมแบบ Back Propagation มาประยุกต์ใช้ เพื่อนำข้อมูลที่วิเคราะห์ได้มาใช้ในการป้องกัน และตรวจจับลูกค้าที่มีแนวโน้มที่จะหลีกเลี่ยงการชำระค่าบริการ ซึ่งเป็นประโยชน์ต่อองค์กรในการดำเนินการให้บรรลุเป้าหมายได้อย่างมีประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Data Mining Application in Mobile Fraud Management
Student	Miss Napaporn Phanyothachart
Advisor	Dr. Pattarachai Lalitrojwong
Level of Study	Master of Science in Information Technology
Major	Information Science
Academic Year	2004

ABSTRACT

In the present time, mobile telecommunication fraud has become a massive problem. Mobile operators have been defined various types of fraud like SIM card cloning. Fraudsters can use services or network without intention to pay. The telecommunication companies lose a large amount of revenue to fraud as opposed to bad debt. This report presents an analysis in mobile fraud activity using data mining to prevent and identify fraudsters in the future.

This project develops an application for analyzing fraud in mobile telecommunication services using data mining and back propagation neural networks. This application has benefits to prevent and protect the customers who tend to be fraudsters and also very useful for the company to efficiently achieve the goal.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ในการพัฒนาระบบระบบวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการระบบโทรศัพท์เคลื่อนที่นี้ ผู้จัดทำขอขอบพระคุณ ดร. ภัทรชัย ลลิตโรจน์วงศ์ ที่ได้กรุณาให้แนวทางการศึกษาและให้คำปรึกษาตลอดช่วงเวลาการดำเนินงาน จนสามารถพัฒนาโครงการได้สำเร็จลุล่วงด้วยดี

สุดท้ายนี้ขอขอบคุณทุกๆกำลังใจจากบุคคลในครอบครัวและเพื่อนๆ ทุกคนที่เป็นส่วนหนึ่งที่ช่วยให้สามารถแก้ไขปัญหาและผ่านอุปสรรคต่างๆได้เป็นอย่างดี

นภาพร พันธ์โยธาชาติ

4 ตุลาคม 2547

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... I
บทคัดย่อภาษาอังกฤษ..... II
กิตติกรรมประกาศ..... III
สารบัญ..... IV
สารบัญตาราง..... VII
สารบัญรูป..... VIII
บทที่
1. บทนำ..... 1
1.1 ความเป็นมาของปัญหา..... 1
1.2 วัตถุประสงค์..... 1
1.3 ขอบเขตการศึกษา..... 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 2
2. ทฤษฎีที่เกี่ยวข้อง..... 3
2.1 คาต้าไมน์นิ่ง..... 3
2.1.1 ขั้นตอนในการทำคาต้าไมน์นิ่ง..... 3
2.1.2 ลักษณะของข้อมูลที่ใช้ในการทำคาต้าไมน์นิ่ง..... 6
2.1.3 ประเภทของการใช้งานคาต้าไมน์นิ่ง..... 7
2.1.4 การประยุกต์ใช้งานคาต้าไมน์นิ่ง..... 8
3. เทคนิคที่เหมาะสมสำหรับวิเคราะห์การทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่..... 8
3.1 เครือข่ายประสาทเทียม..... 9
2.3.1 หลักการทำงานของเครือข่ายประสาทเทียม..... 9
2.3.2 รูปแบบของการเชื่อมโยงระหว่างโหนด..... 11
2.3.3 กระบวนการเรียนรู้ของเครือข่ายประสาทเทียม..... 12
2.3.4 การคำนวณในเครือข่ายประสาทเทียม..... 13
2.3.5 การคำนวณของโหนดในเครือข่ายประสาทเทียม..... 13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.3.6. การแปลงค่าข้อมูลให้เหมาะสมกับการทำงานของเครือข่ายประสาทเทียม.... 15
2.3.7. เครือข่ายประสาทเทียมแบบ Back Propagation..... 15
2.3.8 ข้อดีของเครือข่ายแบบ Back Propagation..... 19
2.3.9 ข้อเสียของเครือข่ายแบบ Back Propagation..... 20
2.3.10 การกำหนดจำนวนฮิดเดนเลเยอร์..... 20
3. การวิเคราะห์และออกแบบระบบ..... 21
3.1 การวิเคราะห์การทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่..... 21
3.2 วัตถุประสงค์ทางธุรกิจ..... 22
3.3 การเตรียมข้อมูล..... 23
3.4 การออกแบบลักษณะของข้อมูล..... 23
3.5 การออกแบบระบบ..... 26
3.5.1 ส่วนการเลือกอัลกอริทึม..... 27
3.5.2 ส่วนติดต่อฐานข้อมูล..... 27
3.5.3 ส่วนการเตรียมข้อมูล..... 28
3.5.4 ส่วนการสร้างแบบจำลอง..... 28
3.5.5 ส่วนการใช้งานแบบจำลอง..... 34
4. การพัฒนาระบบ..... 35
4.1 ส่วนการเลือกอัลกอริทึม..... 35
4.2 ส่วนติดต่อฐานข้อมูล..... 36
4.3 ส่วนการเตรียมข้อมูล..... 37
4.4 ส่วนการสร้างแบบจำลอง..... 40
4.5 ส่วนการใช้งานแบบจำลอง..... 46
5. การทดลองการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่..... 49
5.1 วัตถุประสงค์ทางธุรกิจ..... 49
5.2 การเตรียมข้อมูล..... 49

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

5.3 การทดลองที่1 51

5.3.1 การฝึกอบรมเครือข่ายประสาทเทียม 51

5.3.2 การทดสอบเครือข่ายประสาทเทียม 54

5.3.3 การวิเคราะห์ข้อมูลจากโปรแกรม 56

5.3.4 สรุปผลการทดลอง 57

5.4 การทดลองที่2 57

5.4.1 การฝึกอบรมเครือข่ายประสาทเทียม 57

5.4.2 การทดสอบเครือข่ายประสาทเทียม 60

5.4.3 การวิเคราะห์ข้อมูลจากโปรแกรม 62

5.4.4 สรุปผลการทดลอง 63

6. สรุปการศึกษาและข้อเสนอแนะ 64

6.1 สรุปผลการศึกษา 64

6.2 ข้อเสนอแนะ 65

บรรณานุกรม 66

ประวัติผู้เขียน 67

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ **หน้า**

3.1 ข้อมูลชุดฝึกอบรมที่ใช้สำหรับการสอนและการทดสอบระบบ.....24
3.2 ข้อมูลที่ใช้การนำไปวิเคราะห์หาผู้ที่ทำการทุจริต.....24
3.3 รายละเอียดของข้อมูลที่ใช้ในการฝึกอบรมเครือข่ายประสาทเทียม.....25
5.1 ผลการวิเคราะห์ข้อมูลจากโปรแกรม (การทดลองที่ 1).....56
5.2 ผลการวิเคราะห์ข้อมูลจากโปรแกรม (การทดลองที่ 2).....62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูปที่

2.1 โครงสร้างของระบบประสาทของมนุษย์ 10
2.2 โครงสร้างพื้นฐานของแบบจำลองเครือข่ายประสาทเทียม 11
2.3 ค่าฟังก์ชั่นกระตุ้นของซิกมอยด์ 14
2.4 รูปแบบการคำนวณเอาต์พุตของแต่ละโหนดที่อยู่ในฮิดเดนเลเยอร์และเอาต์พุตเลเยอร์ 14
3.1 แผนภาพการทำงานของระบบ 26
3.2 แผนภาพแสดงส่วนการเลือกอัลกอริทึม 27
3.3 แผนภาพแสดงส่วนติดต่อฐานข้อมูล 27
3.4 แผนภาพแสดงส่วนการเตรียมข้อมูล 28
3.5 แผนภาพแสดงการสอนให้ระบบเรียนรู้ 32
3.6 แผนภาพแสดงส่วนการทดสอบแบบจำลอง 33
3.7 แผนภาพแสดงส่วนการใช้งานแบบจำลอง 34
4.1 หน้าจอหลักของระบบ 35
4.2 หน้าจอของระบบเมื่อทำการเลือกอัลกอริทึมเรียบร้อยแล้ว 36
4.3 หน้าจอรับข้อมูลในการติดต่อกับฐานข้อมูล 37
4.4 ข้อความเมื่อทำการติดต่อฐานข้อมูลสำเร็จ 37
4.5 หน้าจอการเตรียมข้อมูล 38
4.6 หน้าจอการเลือกไฟล์ที่ต้องการโหลดเข้าในระบบ 38
4.7 ข้อความเมื่อทำการโหลดข้อมูลเสร็จเรียบร้อย 39
4.8 หน้าจอเลือกอัลกอริทึมในการทำความสะอาดข้อมูลที่ขาดหายไป 39
4.9 ข้อความเมื่อทำความสะอาดข้อมูลเรียบร้อย 39
4.10 หน้าจอให้เลือกว่าต้องการจะสร้างแบบจำลองใหม่หรือว่ารันจากแบบจำลองเดิม 40
4.11 หน้าจอแสดงแอตทริบิวต์ของข้อมูล 41
4.12 หน้าจอรับข้อมูลในการสร้างเครือข่ายประสาทเทียม 42
4.13 หน้าจอฝึกอบรมเครือข่ายประสาทเทียม 42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

หน้า

รูปที่

4.14 หน้าจอแสดงผลของการฝึกอบรม....43
4.15 รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียม....43
4.16 รายงานรายละเอียดของผลการฝึกอบรมเครือข่ายประสาทเทียม....44
4.17 หน้าจอแสดงผลการทดสอบเครือข่ายประสาทเทียม....44
4.18 รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียม....45
4.19 รายงานรายละเอียดของผลการทดสอบเครือข่ายประสาทเทียม....45
4.20 หน้าจอในการใช้งานแบบจำลอง....46
4.21 หน้าจอการเลือกไฟล์แบบจำลองที่ต้องการ....47
4.22 หน้าจอการใช้งานแบบจำลองเมื่อเลือกแบบจำลองเรียบร้อยแล้ว....47
4.23 รายงานแสดงผลการพยากรณ์จากแบบจำลอง....48
5.1 การเตรียมข้อมูลในการทำไมน์นิ่ง....50
5.2 การทำความสะอาดข้อมูลโดยการตัดข้อมูลที่ขาดหายไปทิ้ง....50
5.3 การเลือกข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 1)....52
5.4 การกำหนดค่าข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 1)....52
5.5 ผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 1)....53
5.6 รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 1)....53
5.7 ผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 1)....54
5.8 รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 1)....55
5.9 การเลือกข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 2)....58
5.10 การกำหนดค่าข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 2)....58
5.11 ผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 2)....59
5.12 รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 2)....59
5.13 ผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 2)....60
5.14 รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 2)....61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1. ความเป็นมาของปัญหา

ธุรกิจการให้บริการโทรศัพท์เคลื่อนที่ ในปัจจุบันนั้นขยายตัวเพิ่มขึ้นอย่างรวดเร็ว และมีผู้ใช้บริการเป็นจำนวนมาก ปัญหาที่พบตามมาคือ มีผู้ทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่เพิ่มมากขึ้น จนกลายเป็นปัญหาตลอดระยะเวลา 10 ปีที่ผ่านมา การบริหารงานจัดการที่ไม่มีระเบียบรูปแบบของการให้บริการที่หลากหลาย และจำนวนของผู้ให้บริการระบบโทรศัพท์เคลื่อนที่ที่มีเพิ่มขึ้น ทำให้โอกาสในการทุจริตเพิ่มสูงขึ้น โดยเฉลี่ยแล้วธุรกิจให้บริการโทรศัพท์เคลื่อนที่และธุรกิจโทรคมนาคมอื่นๆ จะสูญเสียเงินจำนวน 3% ของรายได้ทั้งหมดต่อปี ให้กับการทุจริตในรูปแบบต่างๆ ซึ่งปัญหาเหล่านี้ถือเป็นเรื่องใหญ่ในการดำเนินธุรกิจ ซึ่งจำเป็นที่จะต้องได้รับการแก้ไข ดังนั้นผู้ให้บริการระบบโทรศัพท์เคลื่อนที่ จึงพยายามหาวิธีการจัดการกับผู้ที่ทำการทุจริต (Fraud Management) เพื่อตรวจสอบพฤติกรรมการทุจริต และหาวิธีป้องกันแก้ไขต่อไป

ดาต้าไมน์นิ่งเป็นวิธีการหนึ่งที่นิยมนำมาให้ในการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ โดยนำมาใช้หาความสัมพันธ์ของข้อมูลต่างๆของผู้ใช้บริการ เพื่อนำข้อมูลมาใช้ในการป้องกัน และตรวจจับลูกค้าที่มีแนวโน้มที่จะหลีกเลี่ยงการชำระค่าบริการ ซึ่งเป็นประโยชน์ต่อองค์กรในการดำเนินการให้บรรลุเป้าหมายได้อย่างมีประสิทธิภาพ

1.2. วัตถุประสงค์

1. เพื่อศึกษาหลักการเบื้องต้นของทฤษฎีดาต้าไมน์นิ่งและการนำมาประยุกต์ใช้ในการพัฒนาระบบงานสำหรับการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์ เคลื่อนที่
2. เพื่อศึกษาทฤษฎี และการประยุกต์ใช้อัลกอริทึมของเครือข่ายประสาทเทียม ในการวิเคราะห์หาความสัมพันธ์ของข้อมูล
3. เพื่อศึกษาวิธีการ และขั้นตอนของการพัฒนาระบบงานที่มีประสิทธิภาพในการสร้างระบบงานสำหรับการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์ เคลื่อนที่
4. เพื่อใช้เป็นพื้นฐานในการศึกษาการประยุกต์ใช้ทฤษฎีดาต้าไมน์นิ่งสำหรับงานในด้านอื่นๆ ได้อย่างมีประสิทธิภาพต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3. ขอบเขตของการศึกษา

โครงการนี้เป็นการศึกษาและพัฒนาระบบที่ใช้ในการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการระบบโทรศัพท์เคลื่อนที่ โดยมีขอบเขตการศึกษาดังต่อไปนี้

1. เป็นการพัฒนาระบบตามหลักการของดาต้าไมน์นิ่งโดยประยุกต์ใช้ทฤษฎีเครือข่ายประสาทเทียม ในการสร้างแบบจำลองของระบบเพื่อใช้วิเคราะห์พฤติกรรมการทุจริตของผู้ใช้บริการระบบโทรศัพท์เคลื่อนที่
2. ข้อมูลที่ใช้จะถูกจำลองขึ้นจากแอตทริบิวต์ที่มีการใช้งานจริง ของบริษัทโทเทิ่ล แอ็คเซ็ส คอมมูนิเคชั่น (DTAC) ซึ่งข้อมูลจะถูกแบ่งออกเป็น 2 ส่วน โดยข้อมูลชุดแรกเป็นข้อมูลชุดฝึกอบรมที่ใช้ในการเรียนรู้และทดสอบความแม่นยำของเครือข่ายประสาทเทียม และข้อมูลชุดที่สองเป็นข้อมูลที่ต้องการวิเคราะห์หาผู้ที่ทำการทุจริต โดยนำเครือข่ายประสาทเทียมเข้ามาพยากรณ์

1.4. ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถเรียนรู้ และเข้าใจทฤษฎีเบื้องต้นของดาต้าไมน์นิ่ง โดยใช้หลักการของทฤษฎีเครือข่ายประสาทเทียม
2. สามารถนำโปรแกรมไปประยุกต์ใช้งานในธุรกิจต่างๆได้จริง
3. เป็นข้อมูลให้ผู้บริหาร ในการวางแผนงานด้านธุรกิจต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1. ดาต้าไมน์นิ่ง

ปัจจุบันระบบสนับสนุนข้อมูลเพื่อใช้ในการตัดสินใจ ได้เข้ามามีอิทธิพลในการดำเนินธุรกิจเป็นอย่างมาก ข้อมูลที่เป็นประโยชน์เพื่อใช้ในการตัดสินใจนั้นจะถูกค้นหามาจากคลังข้อมูล (Data Warehouse) ซึ่งเป็นฐานข้อมูลขนาดใหญ่ที่ประกอบไปด้วยข้อมูลเป็นเทอราไบต์ ทำให้ยากแก่การค้นหาได้อย่างรวดเร็วด้วย DBMS จึงมีการนำเอาดาต้าไมน์นิ่งเข้ามาช่วยในเรื่องของเทคนิคในการจัดการข้อมูล ซึ่งด้วยเทคนิคนี้เราสามารถใช้ค้นหาข้อมูลสำคัญที่ปะปนกับข้อมูลอื่นๆในฐานข้อมูลที่ไม่ใช่แค่การสุ่มหา บางคนเรียกว่า KDD (Knowledge Discovery in Database) หรือการค้นหาข้อมูลด้วยความรู้ โดยกระบวนการทำดาต้าไมน์นิ่งมีดังต่อไปนี้

2.1.1. ขั้นตอนในการทำดาต้าไมน์นิ่ง

ขั้นตอนในการทำดาต้าไมน์นิ่งประกอบไปด้วย 5 ขั้นตอนดังต่อไปนี้ (Cabena *et.al.* 1998)

1. การตั้งวัตถุประสงค์ทางธุรกิจ (Business Objective Determination)

   การตั้งวัตถุประสงค์ทางธุรกิจ เป็นขั้นตอนที่สำคัญที่สุด คือ จำเป็นต้องระบุวัตถุประสงค์ในการทำให้ชัดเจน ถ้าระบุไม่ชัดเจนแล้ว ก็จะไม่สามารถเริ่มทำกระบวนการอื่นๆได้

2. การเตรียมข้อมูล (Data Preparation)

   การเตรียมข้อมูล เป็นการเตรียมข้อมูลเพื่อนำไปใช้ประโยชน์ สามารถแบ่งเป็นขั้นตอนย่อยๆ ได้ดังนี้

   2.1 การทำความสะอาดข้อมูล (Data Cleansing)

   การทำความสะอาดข้อมูลเป็นการทำให้ข้อมูลสมบูรณ์ครบถ้วน โดยทั่วไปในขั้นตอนนี้จะคำนึงถึงข้อมูลใน 2 ลักษณะคือ ข้อมูลที่มีค่าผิดไปจากค่าปกติ (Noisy Data) และข้อมูลที่ขาดหายไป (Missing Value) การทำความสะอาดข้อมูลมีวิธีการและหลักที่นำมาใช้ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ตัดข้อมูลที่ขาดหายไปทิ้ง โดยปกติแล้ววิธีนี้จะถูกใช้เมื่อข้อมูลที่ใช้กำหนดคลาสในการแบ่งกลุ่มขาดหายไป ในกรณีที่มีการใช้เทคนิคการแบ่งกลุ่ม เหมาะกับกรณีที่ข้อมูลนั้นประกอบไปด้วยแอตทริบิวต์ที่มีข้อมูลขาดหายไปจำนวนมาก
- ทำการเติมข้อมูลที่ขาดหายเข้าไปใหม่ด้วยตัวเอง วิธีการนี้จะไม่เหมาะกับข้อมูลขนาดใหญ่ที่มีข้อมูลที่ขาดหายไปจำนวนมาก เพราะต้องใช้เวลานาน
- ใช้ตัวแปรกลางแทนค่าข้อมูลที่ขาดหายไป คือการแทนค่าข้อมูลที่ขาดหายไปทั้งหมดด้วยตัวแปรตัวเดียวกัน เช่นระบุว่าเป็น “Unknown” หรือ -∞ วิธีการนี้เป็นวิธีการที่ง่าย แต่ข้อเสียคือในการทำไมน์นิ่งอาจมีข้อผิดพลาด ที่เกิดจากการนำตัวแปรนี้ไปวิเคราะห์ได้
- แทนข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ย ยกตัวอย่างเช่น ถ้าเงินเดือนโดยเฉลี่ยของลูกค้าของบริษัทแห่งหนึ่งมีค่าเท่ากับ 28,000 บาท ดังนั้นจึงใช้ค่า 28,000 แทนค่าข้อมูลเงินเดือนของลูกค้าที่ขาดหายไป
- แทนข้อมูลที่ขาดหายไปด้วยค่าเฉลี่ยของข้อมูลที่อยู่ภายในคลาสเดียวกัน ยกตัวอย่างเช่น ถ้าข้อมูลของลูกค้าที่อยู่ในกลุ่มเครดิตไม่ดี ให้แทนค่าข้อมูลเงินเดือนของลูกค้าจากค่าเฉลี่ยที่คิดจากกลุ่มลูกค้าที่มีเครดิตไม่ดีเท่านั้น
- แทนค่าข้อมูลที่ขาดหายไปด้วยค่าที่เป็นไปได้มากที่สุด ซึ่งค่านี้สามารถคำนวณได้จากสมการถดถอย หรือโครงสร้างต้นไม้ได้

2.2 การแปลงรูปแบบข้อมูล (Data Transformation)

การแปลงรูปแบบข้อมูลเป็นการปรับข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับโมเดลที่จะใช้ในการทำดาต้าไมน์นิ่ง เช่น การแปลงค่าข้อมูลเป็นช่วงเพื่อให้เหมาะสมกับโมเดลแบบโครงสร้างต้นไม้ หรือการแปลงค่าข้อมูลเป็น 0 กับ 1 เพื่อให้เหมาะกับโมเดลแบบเครือข่ายประสาทเทียมเป็นต้น

3. การทำไมน์นิ่ง (Data Mining)

จากลักษณะงานหลายประเภทในการทำดาต้าไมน์นิ่งนั้นทำให้มีการพัฒนาเทคนิค หรือวิธีการที่เหมาะสมเพื่อใช้กับงานแต่ละประเภท ซึ่งวิธีการที่ต่างกันก็จะมีความเหมาะสมกับวัตถุประสงค์ของผู้ใช้งานในด้านต่างๆกันด้วย โดยแต่ละวิธีก็มีข้อดีและข้อเสียที่แตกต่างกัน นอกจากนี้ในงานแต่ละอย่างนั้นไม่ได้ระบุว่าจะต้องใช้เทคนิคใดเทคนิคหนึ่งโดยเฉพาะเท่านั้น แต่ขึ้นอยู่กับลักษณะของปัญหาที่นำมาวิเคราะห์ ซึ่งในบางครั้งการนำเทคนิคหลายๆอย่างมาใช้เพื่อเปรียบเทียบความเหมาะสมกับแต่ละงาน หรือการนำหลายๆวิธีการมาใช้ร่วมกัน อาจจะมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนช่วยในการค้นหาความหมาย หรือความสัมพันธ์จากข้อมูลออกมาได้อย่างมีประสิทธิภาพมากกว่าการใช้เพียงวิธีการเดียว ซึ่งตัวอย่างของเทคนิคที่สามารถนำมาใช้ในการทำดาต้าไมน์นิ่งได้แก่

3.1. การวิเคราะห์ข้อมูลในเชิงตัวเลข (Statistical Method)

การวิเคราะห์ข้อมูลในเชิงตัวเลข มีวัตถุประสงค์หลักเพื่อแยกความแตกต่าง หรือหาความสัมพันธ์ของข้อมูลภายในกลุ่มข้อมูล โดยใช้วิธีการคำนวณตามหลักการทางสถิติ เช่น การสร้างและทดสอบสมมติฐาน (Hypothesis Testing) และการวิเคราะห์สมการการถดถอย (Regression Analysis) เพื่อหาความสัมพันธ์ของข้อมูล เป็นต้น ซึ่งสิ่งสำคัญสำหรับวิธีนี้คือ ข้อมูลที่นำมาวิเคราะห์ต้องอยู่ในรูปตัวเลข และมีการแบ่งกลุ่มเพื่อใช้ในการวิเคราะห์เปรียบเทียบกัน

3.2. การเรียนรู้จากประสบการณ์ในอดีต (Case-Based Reasoning)

การเรียนรู้จากประสบการณ์ในอดีต เป็นการใช้ข้อมูลที่เกิดจากประสบการณ์และผลลัพธ์ที่ได้จากในอดีต ซึ่งตามปกติในแต่ละครั้งของการเกิดปัญหาต่างๆนั้น ส่วนมากมักเป็นปัญหาที่เคยเกิดขึ้น และได้รับการแก้ไขสำเร็จผ่านมาแล้วในอดีต ดังนั้นในการพิจารณาปัญหาใหม่ที่เกิดขึ้นนั้น จะมีการนำรายละเอียดของปัญหามาวิเคราะห์ และระบบก็จะค้นหากรณีต่างๆ จากข้อมูลที่มีการเก็บรวบรวมไว้ โดยหาลักษณะที่ใกล้เคียงกันมากที่สุด เมื่อพบกรณีดังกล่าวแล้ว ก็จะนำวิธีการแก้ไขนั้นมาใช้กับปัญหาใหม่ที่เกิดขึ้น และผลลัพธ์ของกรณีใหม่นี้ก็จะถูกเก็บไว้เพื่อใช้อ้างอิงต่อไป

3.3. เครือข่ายประสาทเทียม (Neural Networks)

เครือข่ายประสาทเทียม เป็นวิธีการที่ใช้ระบุรูปแบบของข้อมูล หรือจัดกลุ่มข้อมูลโดยสร้างแบบจำลองที่ประกอบด้วยโหนดที่มีการเชื่อมต่อกันเป็นเครือข่าย มีลักษณะคล้ายกับระบบประสาทของมนุษย์ โดยที่แต่ละโหนดมีการตอบสนองต่อตัวแปร หรือข้อมูลที่เข้ามากระตุ้นระบบ โดยในช่วงเวลาใดๆอาจมีโหนดที่ได้รับการกระตุ้นให้มีการทำงานได้พร้อมกันหลายโหนด ซึ่งลักษณะการทำงานที่ตอบสนองต่อข้อมูลที่เข้าสู่ระบบในแต่ละครั้งนั้นจะเสมือนเป็นการสร้างให้ระบบมีการเรียนรู้ลักษณะต่างๆของข้อมูล และทำการจัดกลุ่มข้อมูลตามลักษณะที่คล้ายกัน

3.4. โครงสร้างต้นไม้ (Decision Trees)

โครงสร้างต้นไม้ เป็นการวิเคราะห์เพื่อค้นหากฎเกณฑ์ หรือความสัมพันธ์ในลักษณะที่แบ่งย่อยลงไปตามโครงสร้างแบบต้นไม้ โดยที่แต่ละโหนดที่ไม่ใช่โหนดสุดท้ายของโครงสร้างจะแทนจุดที่มีการตรวจสอบเงื่อนไข หรือเป็นการตัดสินใจเกี่ยว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กับลักษณะของข้อมูลที่นำมาพิจารณาและจากผลลัพธ์ที่ได้ในแต่ละครั้งก็จะมีการเลือกเส้นทางที่แน่นอน โดยในการจัดแบ่งประเภทของข้อมูลนั้น จะเริ่มพิจารณาจากโหนดที่เป็นโหนดแรก และอ้างอิงลงมาตามโครงสร้างต้นไม้เรื่อยๆ จนกระทั่งถึงโหนดสุดท้ายของโครงสร้าง ซึ่งแสดงว่ากระบวนการตัดสินใจ หรือการวิเคราะห์ข้อมูลนั้นเสร็จสิ้นลงแล้ว และสามารถบอกลักษณะหรือประเภทของข้อมูลนั้นได้ตามโหนดที่ข้อมูลผ่านการตรวจสอบมา

3.5. กฎความสัมพันธ์ (Association Rule)

กฎความสัมพันธ์เป็นวิธีการหาข้อมูลที่มีความสัมพันธ์กัน หรือข้อมูลที่มีลักษณะการเกิดขึ้นร่วมกันภายในกลุ่มข้อมูลที่นำมาพิจารณา แต่จากผลที่ได้นั้นไม่ได้หมายความว่าความสัมพันธ์ของข้อมูลที่พบนั้นจะเป็นสิ่งที่มีความสำคัญเสมอไป ดังนั้นจึงต้องนำผลที่ได้ไปพิจารณาโดยใช้วิธีที่เหมาะสมอีกครั้ง

4. การวิเคราะห์ผล (Analysis of Results)

การวิเคราะห์ผล เป็นการเก็บผลลัพธ์ของดาต้าไมน์นิ่ง และสรุปความหมายของผลลัพธ์ที่ได้ ทำการประเมินคำตอบ และวิเคราะห์ผลว่าได้ตามความต้องการหรือผิดพลาดอะไรบ้าง

5. การนำเสนอความรู้ (Assimilation of Knowledge)

การนำเสนอความรู้ เป็นการนำความรู้ใหม่ๆ ที่ได้มาแสดงผลในรูปแบบที่เข้าใจง่าย เพื่อนำมาใช้ให้เป็นประโยชน์ต่อการดำเนินธุรกิจหรือช่วยในการตัดสินใจต่างๆ

2.1.2. ลักษณะของข้อมูลที่ใช้ในการทำดาต้าไมน์นิ่ง

ลักษณะของข้อมูลที่ใช้ในการทำดาต้าไมน์นิ่งจะมีดังต่อไปนี้

- ข้อมูลขนาดใหญ่ เกินกว่าจะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า หรือโดยการใช้ DBMS ในการจัดการฐานข้อมูล
- ข้อมูลที่มาจากหลายแหล่ง โดยอาจรวบรวมมาจากหลายระบบปฏิบัติการหรือหลาย DBMS เช่น Oracle DB2 และ MS Access เป็นต้น
- ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดช่วงเวลาที่ทำการไมน์นิ่ง ถ้าหากข้อมูลที่มีอยู่นั้นเป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาจะต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้นไว้และนำฐานข้อมูลที่บันทึกไว้มาทำไมน์นิ่ง แต่เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลา จึงทำให้ผลลัพธ์ที่ได้จากการทำไมน์นิ่ง สมเหตุสมผลในช่วงเวลาหนึ่งเท่านั้น ดังนั้นเพื่อให้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ได้ผลลัพธ์ที่มีความถูกต้องเหมาะสมอยู่ตลอดเวลา จึงต้องทำไมน์นิ่งใหม่ทุกครั้งในช่วงเวลาที่เหมาะสม

- ข้อมูลที่มีโครงสร้างซับซ้อน เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย ซึ่งข้อมูลเหล่านี้สามารถทำไมน์นิ่งได้เช่นกัน

2.1.3. ประเภทของการใช้งานดาต้าไมน์นิ่ง

ดาต้าไมน์นิ่งเป็นวิธีการที่ถูกนำมาใช้อย่างกว้างขวางในการวิเคราะห์หาความหมาย หรือลักษณะเด่นที่แฝงอยู่ในข้อมูลจำนวนมากๆ โดยการเลือกเทคนิคหรือแบบจำลองต่างๆ เพื่อนำมาประยุกต์ใช้ จะขึ้นอยู่กับประเภทของงานและวัตถุประสงค์ของผู้ใช้เป็นหลัก โดยทั่วไปประเภทของงานตามลักษณะแบบจำลองที่ใช้ในการทำดาต้าไมน์นิ่งนั้น สามารถแบ่งกลุ่มได้เป็น 2 ประเภทคือ

1. งานด้านการพยากรณ์ (Predictive Data Mining)

งานด้านการพยากรณ์เป็นการคาดคะเนหรือประมาณค่าที่ชัดเจนของข้อมูลที่เกิดขึ้น โดยใช้พื้นฐานจากข้อมูลที่ผ่านมาในอดีต สามารถแบ่งลักษณะงานได้ดังต่อไปนี้

- การแยกประเภท (Classification) เป็นการจัดกลุ่มให้กับแต่ละข้อมูลในฐานข้อมูล โดยมีการระบุค่าหรือลักษณะที่เป็นไปได้ของข้อมูลภายในแต่ละกลุ่ม เช่น การจัดกลุ่มของผู้ป่วยตามผลของการใช้ยารักษา เพื่อระบุรูปแบบการรักษาให้กับผู้ป่วยใหม่ที่เข้ารับการรักษาเป็นต้น
- การพยากรณ์ (Prediction) เป็นการพยากรณ์ค่าที่เป็นไปได้ หรือการกระจายของค่าที่เป็นไปได้ของตัวแปรใดๆ ในกลุ่มข้อมูล

2. งานด้านการแบ่งกลุ่ม (Descriptive Data Mining)

งานด้านการแบ่งกลุ่ม เป็นการทำแบบจำลองเพื่ออธิบายลักษณะบางอย่างของข้อมูลที่มีอยู่ ซึ่งโดยส่วนมากจะเป็นลักษณะการแบ่งกลุ่มให้กับข้อมูล สามารถแบ่งลักษณะของงานได้ดังต่อไปนี้

- การวิเคราะห์ความเชื่อมโยง (Link Analysis) เป็นการหาความสัมพันธ์ของข้อมูลภายในกลุ่มของข้อมูล เพื่อใช้ลักษณะของข้อมูลหนึ่งๆ ในการบอกลักษณะที่จะเกิดขึ้นกับข้อมูลอีกตัวหนึ่ง ซึ่งอาจจะเป็นการหาความสัมพันธ์ของข้อมูลในกลุ่มเดียวกัน เช่น การระบุว่าในกลุ่มของลูกค้าที่ซื้อนมนั้นจะมีลูกค้า 64 % ที่ซื้อขนมปังด้วย หรืออาจจะเป็นการหาความสัมพันธ์ของตัวแปรระหว่างกลุ่มข้อมูลก็ได้ เช่น ในทุกๆครั้งที่ดัชนีของตลาดหุ้นหนึ่งลดลง 5% ดัชนีของตลาดหุ้นอื่นจะเพิ่มขึ้น 13% ภายในช่วง 2-6 เดือนต่อจากนั้นเป็นต้น ซึ่งลักษณะของการหาความสัมพันธ์นั้นอาจแบ่งได้เป็น 3 กลุ่มคือ การหาความสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระหว่างข้อมูล (Association Discovery) การหาความสัมพันธ์ในลักษณะที่เป็นลำดับของข้อมูล (Sequential Pattern Discovery) และการหาความสัมพันธ์ของข้อมูลกับช่วงเวลาใดๆ (Similar Time Sequence Discovery)

- การจัดกลุ่ม (Clustering) เป็นการแบ่งกลุ่มของข้อมูลโดยที่ในแต่ละกลุ่มนั้นประกอบด้วยข้อมูลที่มีลักษณะคล้ายกัน โดยในการแบ่งกลุ่มของข้อมูลที่ดีนั้น จะทำให้ได้กลุ่มที่มีคุณภาพคือมีความแตกต่างกันระหว่างกลุ่มมาก และมีความแตกต่างกันภายในกลุ่มน้อย เช่น การแบ่งกลุ่มของลูกค้าโดยใช้พฤติกรรมการซื้อเป็นเกณฑ์
- การตรวจจับความเบี่ยงเบน (Deviation Detection) เป็นเทคนิคที่ใช้ในการแสดงลักษณะของข้อมูลที่ผิดปกติ หรือผิดไปจากที่คาดไว้ โดยมีการแสดงผลอยู่ในลักษณะที่สามารถทำความเข้าใจและแปลความหมายได้ง่าย เช่น การใช้กราฟ เป็นต้น

2.1.4. การประยุกต์ใช้งานดาต้าไมน์นิ่ง

ดาต้าไมน์นิ่งสามารถนำมาประยุกต์ใช้งานกับธุรกิจต่างๆในปัจจุบัน ดังต่อไปนี้ (สุวิมล คงศักดิ์ตระกูล. 2544)

- ธุรกิจค้าปลีกสามารถใช้ดาต้าไมน์นิ่ง ในการพิจารณาหากลยุทธ์ให้เป็นที่สนใจกับผู้บริโภคในรูปแบบต่างๆ เช่น ที่ว่างในชั้นวางของจะจัดการอย่างไรจึงจะเพิ่มยอดขายได้
- การวิเคราะห์ผลิตภัณฑ์ สามารถเก็บรวบรวมลักษณะและราคาของผลิตภัณฑ์ทั้งหมดเพื่อนำไปสร้างแบบจำลองด้วยเทคนิคดาต้าไมน์นิ่ง และใช้แบบจำลองในการทำนายราคาผลิตภัณฑ์ตัวอื่นๆ
- การวิเคราะห์บัตรเครดิต การทำดาต้าไมน์นิ่งจะช่วยบริษัทบัตรเครดิตตัดสินใจในการที่จะให้บัตรเครดิตกับลูกค้าหรือไม่ สามารถนำไปใช้แบ่งประเภทของลูกค้าว่ามีความเสี่ยงในเรื่องเครดิตต่ำ ปานกลาง หรือสูง และใช้ป้องกันปัญหาเรื่องการทุจริตบัตรเครดิต
- การวิเคราะห์ลูกค้า การทำดาต้าไมน์นิ่งสามารถช่วยแบ่งกลุ่มและวิเคราะห์ลูกค้าเพื่อที่จะผลิตและเสนอสินค้าได้ตรงตามกลุ่มเป้าหมายแต่ละกลุ่ม และช่วยในการพยากรณ์ได้ว่าลูกค้าคนใดจะเลิกใช้บริการจากบริษัทภายใน 6 เดือนข้างหน้า

2.2. เทคนิคที่เหมาะสมสำหรับวิเคราะห์การทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่

ดาต้าไมน์นิ่งเป็นเทคนิคที่มีประโยชน์มากในการจัดการกับการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ โดยจะใช้วิเคราะห์พฤติกรรมของลูกค้าในระยะยาว ซึ่งการทำดาต้าไมน์นิ่งจะไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถค้นหาผู้ทุจริตได้ในทันทีที่กระทำการทุจริต เหมือนกับการตรวจสอบแบบทันที (Real Time) แต่สามารถใช้วิเคราะห์รูปแบบการทุจริตในระยะยาว หรือการทุจริตที่ไม่ได้เกิดขึ้นบ่อยๆได้

การทำไมน์นิ่งสำหรับการวิเคราะห์การทุจริตในการใช้งานโทรศัพท์เคลื่อนที่มีเทคนิคที่เหมาะสมอยู่มากมาย เช่น Deviation Detection และเครือข่ายประสาทเทียม ในโครงการนี้เลือกใช้เทคนิคเครือข่ายประสาทเทียมสำหรับการวิเคราะห์ทุจริตในการใช้งานโทรศัพท์เคลื่อนที่ เนื่องจากมีความยืดหยุ่นในการเรียนรู้ทำให้สามารถเรียนรู้ความสัมพันธ์รูปแบบต่างๆได้จำนวนมาก (Taniguchi. 1998)

2.3. เครือข่ายประสาทเทียม (Han and Kamber. 2001)

เครือข่ายประสาทเทียม เป็นเทคโนโลยีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ ในการคำนวณค่าฟังก์ชันจากกลุ่มข้อมูล วิธีการของเครือข่ายประสาทเทียมเป็นวิธีการที่ให้ระบบเรียนรู้จากตัวอย่างต้นแบบแล้วฝึกอบรม (Training) ให้ระบบได้รู้จักคิดและแก้ปัญหาที่กว้างขึ้นได้ ในโครงสร้างของเครือข่ายประสาทเทียมจะประกอบด้วยโหนดสำหรับอินพุต-เอาต์พุตกระจายอยู่ในโครงสร้างเป็นเลเยอร์ ได้แก่ อินพุตเลเยอร์ เอาต์พุตเลเยอร์ และฮิดเดนเลเยอร์ การประมวลผลของเครือข่ายประสาทเทียม จะอาศัยการส่งการทำงานผ่านโหนดต่างๆในเลเยอร์เหล่านี้ส่งผ่านกันเป็นลำดับจนสุดท้ายได้เป็นเอาต์พุตออกมา

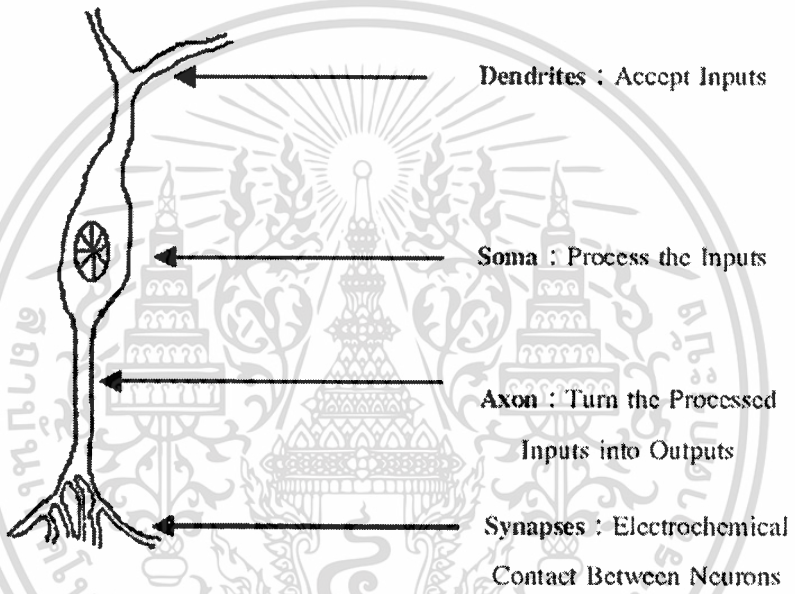
2.3.1. หลักการทำงานของเครือข่ายประสาทเทียม

เครือข่ายประสาทเทียมนั้นเป็นวิธีการที่นำพื้นฐานการคำนวณของแบบจำลองทางคณิตศาสตร์ มาใช้ในการสร้างเครือข่ายให้เกิดการเรียนรู้ โดยจะมีการทำงานในลักษณะเดียวกับระบบประสาทของมนุษย์ ซึ่งประกอบด้วยกลุ่มเซลล์ประสาทหรือนิวรอน ที่ทำหน้าที่ในการจดจำ คิด นำความรู้ และประสบการณ์ที่บันทึกไว้ในสมองมาใช้ตัดสินใจในชีวิตประจำวัน แต่ละนิวรอนอาจจะเชื่อมต่อกับนิวรอนอื่นๆ มากถึง 200,000 นิวรอน โดยนิวรอนของมนุษย์ตามหลักประสาทวิทยา จะแบ่งออกเป็น 4 ส่วน แต่ละส่วนมีฟังก์ชันการทำงานดังนี้

- เดนไดรท์ (Dendrites) คือ ส่วนที่ทำหน้าที่รับข้อมูล
- โซมา (Soma) คือ ส่วนที่ประมวลผลข้อมูลเบื้องต้น
- แอคซอน (Axon) คือ ส่วนที่แปลงข้อมูลที่ได้ประมวลเบื้องต้นเป็นผลลัพธ์ที่ต้องการ
- ไซแนปส์ (Synapses) คือ เส้นประสาทที่ทำหน้าที่เชื่อมต่อเพื่อการสื่อสารกับนิวรอนอื่นในระบบสมอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยการทำงานของระบบประสาทของมนุษย์เริ่มจาก เดนไดรท์จะทำหน้าที่รับข้อมูลจากแหล่งต่างๆ แล้วส่งให้โซมาประมวลผลข้อมูลเบื้องต้น หลังจากนั้นแอคซอนจะแปลงข้อมูลต่างๆ เป็นผลลัพธ์แล้ว ไซแนปส์จะส่งผลลัพธ์ให้นิวรอนอื่น เพื่อช่วยกันสร้างผลลัพธ์สุดท้าย ในความเป็นจริงแล้วระบบการทำงานของระบบประสาทในสมองมนุษย์มีความซับซ้อนเกินกว่าที่มนุษย์จะเข้าใจได้ โครงสร้างของระบบประสาทของมนุษย์สามารถแสดงได้ดังรูปที่ 2.1

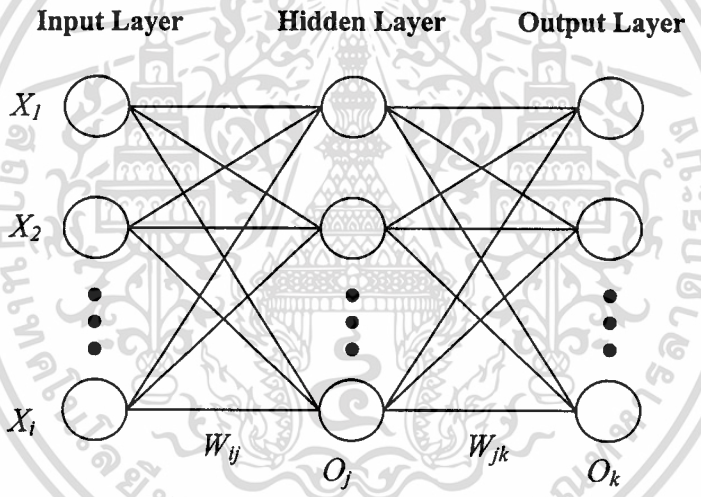


รูปที่ 2.1 โครงสร้างของระบบประสาทของมนุษย์

การทำงานของระบบเครือข่ายประสาทเทียมจะมีการส่งข้อมูลการเรียนรู้ของระบบเชื่อมต่อกันเป็นเครือข่ายจากอินพุตจนถึงเอาต์พุต โดยในแบบจำลองของเครือข่ายประสาทเทียมนั้นจะมีการเชื่อมต่อกันของข้อมูลในลักษณะที่เป็นลำดับชั้น ดังรูปที่ 2.2 จะเห็นได้ว่าโครงสร้างของเครือข่ายประสาทเทียม เป็นลักษณะของเครือข่ายที่มีการเชื่อมต่อกันของกลุ่มอินพุต และกลุ่มของเอาต์พุต โดยผ่านโหนดต่างๆตามลำดับชั้นของการเชื่อมต่อ ซึ่งในการเชื่อมต่อกันระหว่างแต่ละโหนดจากอินพุตจนถึงเอาต์พุตนั้นจะมีค่าถ่วงน้ำหนัก เฉพาะที่แตกต่างกันไป โดยในระหว่างการสอนให้เครือข่ายประสาทเทียมเกิดการเรียนรู้นั้น ค่าถ่วงน้ำหนักเหล่านี้ก็จะถูกปรับเปลี่ยนไปตามค่าที่ได้จากการคำนวณตามอัลกอริทึมที่ใช้จนกระทั่งได้ค่าที่เหมาะสมที่สุด ซึ่งองค์ประกอบที่สำคัญของการพัฒนาแบบจำลองเครือข่ายประสาทเทียมมีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- โครงสร้างของเครือข่ายประสาทเทียม (Architecture) ประกอบด้วยจำนวนเลเยอร์ภายในเครือข่าย คือจำนวนของโหนดในอินพุตเลเยอร์ เอาต์พุตเลเยอร์ รวมถึงจำนวนของโหนดที่อยู่ในฮิดเดนเลเยอร์ ที่เชื่อมต่ออยู่ระหว่างอินพุตเลเยอร์และเอาต์พุตเลเยอร์ด้วย
- ฟังก์ชันกระตุ้น (Activation Function) ใช้ในการคำนวณค่าที่ได้จากแต่ละโหนดภายในเครือข่ายประสาทเทียมจากอินพุตจนถึงเอาต์พุต
- ฟังก์ชันที่ใช้ทดสอบหาค่าความผิดพลาดของผลลัพธ์ที่ได้ (Cost Function) เป็นการเปรียบเทียบค่าผลลัพธ์ที่คำนวณได้จากแบบจำลองกับค่าข้อมูลที่เกิดขึ้น
- อัลกอริทึมที่ใช้สอนให้เครือข่ายประสาทเทียมเกิดการเรียนรู้ (Training Algorithm) โดยมีการกำหนดวิธีการปรับค่าตัวแปรต่างๆภายในเครือข่ายประสาทเทียม เพื่อให้ค่าความผิดพลาดที่คำนวณได้จากฟังก์ชันทดสอบความผิดพลาดมีค่าน้อยที่สุด



รูปที่ 2.2 โครงสร้างพื้นฐานของแบบจำลองเครือข่ายประสาทเทียม

2.3.2. รูปแบบของการเชื่อมโยงระหว่างโหนด

รูปแบบของการเชื่อมโยงระหว่างโหนดแต่ละโหนดภายในเครือข่ายประสาทเทียมนั้น แต่ละโหนดจะเชื่อมต่อกับโหนดอื่นตามระบบการเชื่อมต่อที่สร้างขึ้น เอาต์พุตของโหนดหนึ่งจะเป็นอินพุตของโหนดอื่นๆต่อไป โดยการเชื่อมต่ออาจเป็นแบบเชื่อมทางเดียว (Unidirectional) หรือแบบเชื่อมสองทาง (Two-way connection) ก็ได้ โหนดที่อยู่ในเลเยอร์เดียวกันอาจจะเชื่อมต่อกันหรือไม่ก็ได้ แต่โหนดที่อยู่ในเลเยอร์หนึ่งจะต้องเชื่อมโยงกับอย่างน้อย 1 โหนดในเลเยอร์อื่นๆเสมอ การเชื่อมโยงของแต่ละโหนดระหว่างเลเยอร์สามารถแบ่งเป็นแบบต่างๆ ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Fully Connected Neurons คือ แต่ละโหนดในเลเยอร์ที่ 1 จะเชื่อมต่อกับทุกๆโหนดในเลเยอร์ที่ 2
- Partially Connected Neurons คือ แต่ละโหนดในเลเยอร์ที่ 1 จะเชื่อมต่อกับบางโหนดในเลเยอร์ที่ 2 เท่านั้น
- Feed Forward Neurons คือ โหนดในเลเยอร์ที่ 1 จะส่งเอาต์พุตให้โหนดที่อยู่ในเลเยอร์ที่ 2 แต่จะไม่ได้รับข้อมูลจากโหนดที่อยู่ในเลเยอร์ที่ 2 กลับมา
- Bi-directional คือ เมื่อโหนดในเลเยอร์ที่ 1 ส่งเอาต์พุตให้โหนดที่อยู่ในเลเยอร์ที่ 2 แล้ว โหนดที่อยู่ในเลเยอร์ที่ 2 จะส่งเอาต์พุตกลับมาเป็นอินพุตให้กับโหนดที่อยู่ในเลเยอร์ที่ 1 ด้วย

2.3.3. กระบวนการเรียนรู้ของเครือข่ายประสาทเทียม

เครือข่ายประสาทเทียมเรียนรู้โดยการปรับค่าถ่วงน้ำหนัก (Connection Weight) เพื่อให้ได้ผลลัพธ์ที่มีความคลาดเคลื่อนน้อยที่สุด ความสามารถในการเรียนรู้ของเครือข่ายประสาทเทียมขึ้นอยู่กับโครงสร้างของเครือข่ายประสาทเทียมที่ออกแบบไว้ และวิธีการในการฝึกอบรม ซึ่งสามารถแบ่งวิธีการเรียนรู้ออกได้เป็น 2 แบบคือ

1. Unsupervised Learning

Unsupervised Learning เป็นระบบการเรียนรู้ที่จะปรับตัวเองโดยไม่ต้องการความช่วยเหลือจากภายนอก ไม่ต้องมีตัวอย่างผลลัพธ์เพื่อให้โหนดแต่ละโหนดได้เรียนรู้ ซึ่งกระบวนการเรียนรู้แบบนี้เรียกว่า Learning by Doing

2. Supervised Learning

Supervised Learning เป็นระบบการเรียนรู้ซึ่งใช้ชุดข้อมูลฝึกอบรม โดยโหนดที่อยู่ในฮิดเดนเลเยอร์ จะปรับค่าถ่วงน้ำหนักแบบสุ่ม (Random) หรือหมุนเวียน (Rotation) ตามผลของการคำนวณผลลัพธ์ ระบบการเรียนรู้แบบ Supervised นิยมใช้ในกรณีที่มีหลายเลเยอร์ และมีการเชื่อมโยงแต่ละโหนดแบบ Feed Forward ซึ่งถือเป็นระบบการเรียนรู้ที่มีระบบการส่งข้อมูลเกี่ยวกับความคลาดเคลื่อนในการคำนวณ แล้วจึงนำผลลัพธ์กลับไปปรับค่าถ่วงน้ำหนักในการเชื่อมโยงระหว่างเลเยอร์เพื่อเพิ่มประสิทธิภาพในการคำนวณของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.4. การคำนวณในเครือข่ายประสาทเทียม

ความสามารถในการเรียนรู้ของเครือข่ายประสาทเทียม จะถูกพัฒนาในระหว่างที่มีการเรียนรู้หรือกระบวนการสอน โดยในระหว่างกระบวนการนี้จะมีการคำนวณในลักษณะซ้ำๆ กันกับข้อมูลที่นำมาใช้ในเป็นอินพุตหลายๆชุด และพิจารณาลักษณะของเอาต์พุตที่เกิดขึ้นสำหรับอินพุตแต่ละชุด โดยเครือข่ายจะมีการเปลี่ยนแปลงโดยการปรับค่าของตัวแปรต่างๆ ภายในโครงสร้างตามอัลกอริทึมที่นำมาใช้ในการเรียนรู้จนกระทั่งสามารถสร้างเอาต์พุตได้ตามที่ออกแบบไว้ได้ หรือภายใต้เงื่อนไขที่ยอมรับได้ ซึ่งกลุ่มข้อมูลอินพุตและเอาต์พุตที่นำมาใช้นี้เรียกว่า กลุ่มข้อมูลชุดฝึกอบรม (Training Set)

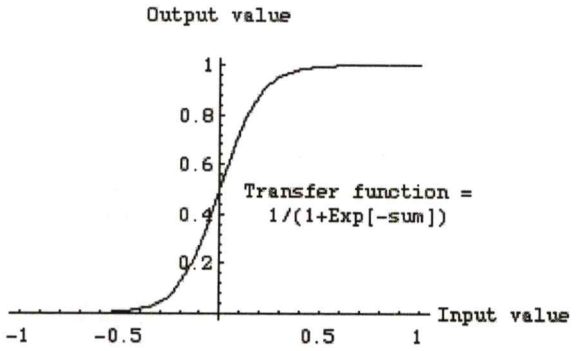
ในการสอนให้เครือข่ายมีการเรียนรู้นั้นจะเกิดประโยชน์ก็ต่อเมื่อข้อมูลที่นำมาใช้สอนต้องมีลักษณะครอบคลุมรูปแบบที่เป็นไปได้ของข้อมูลให้มากที่สุด ดังนั้นจึงมีความสำคัญมากที่ต้องเลือกข้อมูลที่เป็นตัวแทนของทุกรูปแบบของงานที่จะนำแบบจำลองที่ได้มาใช้งาน โดยลักษณะ 2 ประการที่ควรต้องคำนึงถึงในการพิจารณาเลือกข้อมูลที่นำมาใช้ในการสอน คือข้อมูลที่ใช้สอนต้องเป็นส่วนหนึ่งของข้อมูลที่คาดหวังว่าจะนำแบบจำลองไปใช้ในการประมวลผล และข้อมูลที่ใช้สอนต้องถูกเลือกจากช่วงเวลาที่มีการกระจายของข้อมูลครอบคลุมรูปแบบของอินพุตที่เป็นไปได้ทั้งหมด เนื่องจากเครือข่ายประสาทเทียมจะไม่สามารถใช้พยากรณ์ข้อมูลได้อย่างมีประสิทธิภาพ ถ้ามีการสอนด้วยข้อมูลที่มีค่าผิดปกติจำนวนมาก

2.3.5. การคำนวณของโหนดในเครือข่ายประสาทเทียม

ในการคำนวณนั้นสามารถแบ่งประเภทของโหนดตามรูปแบบของฟังก์กระตุ้นที่ใช้ได้เป็น 3 แบบคือ

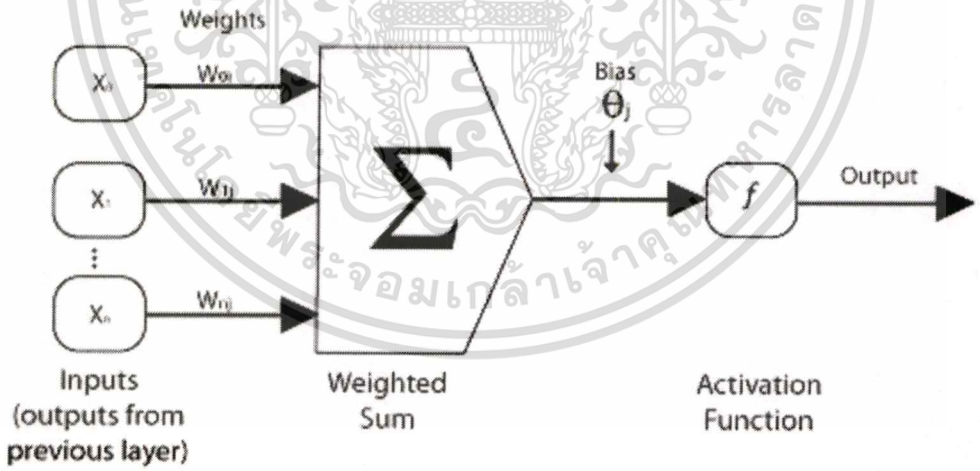
- Threshold Unit เป็นฟังก์ชันกระตุ้นที่จะจำกัดค่าเอาต์พุตไว้เป็น 0 หรือ 1 ตามสถานะก่อนหน้า
- Linear Unit จะใช้ฟังก์ชัน $f(h) = h$
- Non-linear Unit โดยทั่วไปจะใช้การเรียนรู้แบบ Gradient Descent ซึ่งส่วนมากโหนดประเภทนี้จะใช้ฟังก์ชันกระตุ้นของซิกมอยด์ (Sigmoid Function) เช่น $f(h) = \frac{1}{1+e^{-h}}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 ฟังก์ชั่นกระตุ้นของซิกมอยด์

โดยทั่วไป ฟังก์ชันกระตุ้นที่นิยมใช้มากที่สุด คือฟังก์ชั่นของซิกมอยด์ เนื่องจากสามารถหาค่า Differentiate และคำนวณค่า Gradient ได้ง่าย ส่วนฟังก์ชันแบบเชิงเส้นนั้นนิยมใช้ในส่วนที่เป็นโหนดของเอาต์พุตเลเยอร์ของเครือข่ายในการนำไปประยุกต์ใช้กับข้อมูลในลักษณะอนุกรมเวลา รูปแบบการคำนวณเอาต์พุตของแต่ละโหนดที่อยู่ในฮิดเดนเลเยอร์และเอาต์พุตเลเยอร์เป็นดังรูปที่ 2.3



รูปที่ 2.4 รูปแบบการคำนวณเอาต์พุตของแต่ละโหนดที่อยู่ในฮิดเดนเลเยอร์และเอาต์พุตเลเยอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.6. การแปลงค่าข้อมูลให้เหมาะสมกับการทำงานของเครือข่ายประสาทเทียม

ในแบบจำลองที่ใช้หลักการของเครือข่ายประสาทเทียมนั้น กระบวนการแปลงค่าข้อมูลที่ต้องนำมาใช้คือ การนอร์มัลไลซ์ข้อมูล (Normalization) ซึ่งเป็นการลดค่าของข้อมูลให้อยู่ในขอบเขตที่น้อยลง เพื่อให้เหมาะสมกับฟังก์ชันที่ใช้งานในอัลกอริทึมที่นำมาใช้ในการเรียนรู้ของเครือข่ายประสาทเทียม ซึ่งวิธีการนอร์มัลไลซ์ข้อมูลนั้นมีหลายวิธี แต่ที่มีการนำมาใช้กันอย่างกว้างขวางนั้นได้แก่ การแปลงค่าข้อมูลในลักษณะเป็นเชิงเส้น (Min-max Normalization) สามารถคำนวณได้จากสมการดังต่อไปนี้

$$V' = [(\frac{V - \min_A}{\max_A - \min_A})(\max'_A - \min'_A)] + \min'_A$$

โดยค่าต่างๆในสมการแทนได้ดังนี้

- $\min_A$ แทนค่าต่ำสุดของตัวแปร A
- $\max_A$ แทนค่าสูงสุดของตัวแปร A
- V แทนค่าข้อมูลใดๆที่ต้องการแปลงค่า
- V' แทนค่าข้อมูลใหม่ที่ได้จากการคำนวณ
- $\min'_A$ คือค่าต่ำสุดค่าใหม่ของตัวแปร A
- $\max'_A$ คือค่าสูงสุดค่าใหม่ของตัวแปร A

2.3.7. เครือข่ายประสาทเทียมแบบ Back Propagation

ในโครงการนี้จะขอนำเสนอรูปแบบเครือข่ายแบบ Back Propagation ซึ่งมีวิธีการที่เรียนรู้จากตัวอย่างต้นแบบ แล้วฝึกอบรมให้ระบบ ได้รู้จักคิดและแก้ปัญหาที่กว้างขึ้นได้ ในโครงสร้างของเครือข่ายประสาทเทียม จะประกอบด้วยโหนดสำหรับอินพุต-เอาต์พุต กระจายอยู่ในโครงสร้างเป็นเลเยอร์ ได้แก่ อินพุตเลเยอร์ เอาต์พุตเลเยอร์ และฮิดเดนเลเยอร์ การประมวลผลของเครือข่ายประสาทเทียมจะอาศัยการส่งการทำงานผ่านโหนดต่างๆ ในเลเยอร์เหล่านี้

Back Propagation เป็นอัลกอริทึมหนึ่งของเครือข่ายประสาทเทียมซึ่งได้รับความนิยม และมีการนำมาใช้อย่างกว้างขวาง โดยมีลักษณะการเชื่อมต่อแบบ Multilayer Feed-Forward Neural Network และมีการเรียนรู้แบบ Supervised Learning โดยจะเรียนรู้จากการประมวลผลกลุ่มตัวอย่างที่นำมาใช้สำหรับการฝึกอบรมซ้ำกันหลายๆรอบ และนำผลที่ได้จากการพยากรณ์ของเครือข่ายมาเปรียบเทียบกับตัวอย่างข้อมูลจริงที่เตรียมไว้ ซึ่งในการฝึกอบรมแต่ละรอบ ค่าถ่วงน้ำหนักจะถูก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เปลี่ยนแปลง เพื่อให้มีค่าความคลาดเคลื่อนน้อยที่สุด โดยการปรับค่าถ่วงน้ำหนักนั้น จะทำในทิศทางตรงกันข้ามกับการไหลของข้อมูล โดยจะไล่จากเอาต์พุตเลเยอร์ในเลเยอร์ล่างสุดไปจนถึงฮิดเดนเลเยอร์ในเลเยอร์บนสุด

ขั้นตอนการทำงานของ Back Propagation มีดังต่อไปนี้

1. กำหนดค่าถ่วงน้ำหนักเริ่มต้น

การกำหนดค่าเริ่มต้นให้กับค่าถ่วงน้ำหนักจะเริ่มจากการสุ่มตัวเลขจำนวนน้อยๆ เช่นจาก -1.0 ถึง 1.0 หรือ -0.5 ถึง 0.5 โดยแต่ละโหนดก็จะมีค่าเบี่ยงเบนประกอบอยู่ด้วย ซึ่งค่าเบี่ยงเบนก็จะถูกกำหนดค่าเริ่มต้นจากวิธีการสุ่มตัวเลขจำนวนน้อยๆด้วยเช่นกัน

2. การคำนวณหาค่าของอินพุตในทิศทางไปข้างหน้า

ขั้นตอนนี้เป็นการคำนวณค่าอินพุต และค่าเอาต์พุตให้กับแต่ละโหนดในเครือข่าย โดยข้อมูลที่นำมาใช้สอนให้เกิดการเรียนรู้ภายในโครงสร้างนั้นจะถูกป้อนเข้าสู่อินพุตเลเยอร์ของเครือข่าย และมีการคำนวณค่าอินพุตรวมและค่าเอาต์พุตของแต่ละโหนดดังต่อไปนี้

<u>การคำนวณในอินพุตเลเยอร์</u>

สำหรับค่าเอาต์พุตของโหนด j ใดๆในอินพุตเลเยอร์ จะมีค่าเท่ากับค่าอินพุตของโหนดนั้นๆ

$$O_j = I_j$$

ค่าต่างๆในสมการ แทนได้ดังนี้

- O_j = ค่าเอาต์พุตของโหนด j
- I_j = ค่าอินพุตของโหนด j

<u>การคำนวณในฮิดเดนเลเยอร์และเอาต์พุตเลเยอร์</u>

สำหรับค่าอินพุตของโหนด j ใดๆ ที่อยู่ในฮิดเดนเลเยอร์หรือเอาต์พุตเลเยอร์ จะมีค่าเท่ากับการบวกค่าเบี่ยงเบนของโหนด j กับผลรวมทั้งหมดของการคูณค่าเอาต์พุตของทุกๆโหนดที่เชื่อมกับโหนด j กับค่าถ่วงน้ำหนักของการเชื่อมต่อนั้นๆ ดังสมการต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

$$I_j = \sum_{1}^{n} W_{ij} O_i + \theta_j$$

ค่าต่างๆในสมการ แทนได้ดังนี้

- W_{ij} แทนค่าถ่วงน้ำหนักที่เชื่อมต่อระหว่างโหนดที่ i ในเลเยอร์ก่อนหน้ากับโหนดที่ j
- O_i แทนเอาต์พุตที่ได้จากโหนดที่ i ในเลเยอร์ก่อนหน้า
- θ_j แทนค่าเบี่ยงเบนของโหนดที่ j

ส่วนค่าเอาต์พุตสำหรับโหนดใดๆในฮิดเดนเลเยอร์ ได้จากการคำนวณค่าอินพุตรวม สำหรับแต่ละโหนดผ่านฟังก์ชันกระตุ้นที่ใช้ทำการคำนวณผลที่ได้จากโหนดนั้นๆ ซึ่งในอัลกอริทึมนี้จะใช้ฟังก์ชันของซิกมอยด์โดยได้ค่าของเอาต์พุตตามสมการต่อไปนี้

$$O_j = \frac{1}{1 + e^{-I_j}}$$

3. การคำนวณความคลาดเคลื่อนย้อนกลับ

เมื่อได้ค่าของเอาต์พุตครบทุกโหนดแล้ว จึงนำมาคำนวณค่าความผิดพลาดของเอาต์พุตที่ได้ในทิศทางย้อนกลับเพื่อปรับค่าถ่วงน้ำหนักสำหรับการเชื่อมต่อระหว่างโหนดใดๆในเครือข่าย พร้อมทั้งปรับค่าเบี่ยงเบนให้กับแต่ละโหนดด้วย

<u>การคำนวณค่าความคลาดเคลื่อนในเอาต์พุตเลเยอร์</u>

ค่าความคลาดเคลื่อนในแต่ละโหนดของเอาต์พุต สามารถคำนวณได้ดังนี้

$$Err_j = O_j(1 - O_j)(T_j - O_j)$$

ค่าต่างๆในสมการ แทนได้ดังนี้

- O_j แทนเอาต์พุตที่ได้จากโหนดที่ j
- T_j แทนเอาต์พุตที่เป็นจริง โดยได้มาจากการแบ่งกลุ่มในข้อมูลชุดฝึกอบรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การคำนวณความคลาดเคลื่อนในฮิดเดนเลเยอร์

ค่าความคลาดเคลื่อนในฮิดเดนเลเยอร์ที่โหนด j จะนำเอาผลรวมของค่าถ่วงน้ำหนักจากเลเยอร์ที่สูงกว่าที่ติดกัน มาพิจารณา ซึ่งสามารถคำนวณได้ดังนี้

$$Err_j = O_j(1-O_j)\sum Err_k W_{jk}$$

ค่าต่างๆในสมการ แทนได้ดังนี้

- W_{jk} แทนค่าถ่วงน้ำหนักจากดหนด j ไปยังโหนด k ที่อยู่เลเยอร์ที่สูงกว่า
- Err_k แทนค่าความคลาดเคลื่อนที่โหนด k

4. การปรับแต่งค่าถ่วงน้ำหนักและค่าเบี่ยงเบน

ค่าถ่วงน้ำหนักของการเชื่อมต่อระหว่างโหนดใดๆ จะถูกปรับเปลี่ยนด้วยค่าความคลาดเคลื่อนที่เกิดขึ้นต่อเนื่องกันไปตามสมการต่อไปนี้

$$\Delta W_{ij} = (l)Err_{ij}O_i$$
$$\Delta W_{ij} = W_{ij} + \Delta W_{ij}$$

ค่าที่ใช้ในการปรับแต่งค่าเบี่ยงเบนหาได้จากสมการดังต่อไปนี้

$$\Delta\theta_j = (l)Err_j$$
$$\Delta\theta_j = \theta_j + \Delta\theta_j$$

ค่าต่างๆในสมการ แทนได้ดังนี้

- ΔW_{ij} คือค่าการเปลี่ยนแปลงของ W_{ij}
- $\Delta\theta_j$ คือค่าการเปลี่ยนแปลงของ θ_j
- l คืออัตราการเรียนรู้ของเครือข่าย

สำหรับค่า l หมายถึงอัตราการเรียนรู้ของเครือข่าย ซึ่งเป็นค่าที่ใช้ควบคุมความเร็วในการตอบสนองต่อค่าความคลาดเคลื่อนที่เกิดขึ้นในแต่ละครั้ง โดยในกรณีที่ l มีค่ามาก แสดงว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เครือข่ายมีการตอบสนองต่อค่าความคลาดเคลื่อนมาก โดยการปรับค่าถ่วงน้ำหนักให้เปลี่ยนแปลงอย่างชัดเจน แต่ถ้า l มีค่าน้อยๆ แสดงว่ามีการปรับค่าถ่วงน้ำหนักอย่างค่อยเป็นค่อยไป จนกระทั่งได้ค่าที่เหมาะสมที่ทำให้เกิดความผิดพลาดของเอาต์พุตน้อยที่สุด

5. การสิ้นสุดการทำงาน

การฝึกอบรมเครือข่ายจะสิ้นสุดลงเมื่อ

- ค่า ΔW_{ij} ทุกค่า มีค่าน้อยกว่าค่าที่ยอมให้ผิดพลาดได้ (Threshold)
- ค่าอัตราส่วนความคลาดเคลื่อนในการแบ่งกลุ่มมีค่าน้อยกว่าค่าที่ยอมให้ผิดพลาดได้

6. การวัดค่าความเคลื่อน

การเรียนรู้ของเครือข่ายประสาทเทียมนั้นขึ้นอยู่กับค่าความคลาดเคลื่อนที่วัดได้ โดยยึดหลักแนวความคิดที่ว่ายิ่งค่าความคลาดเคลื่อนที่วัดได้ หรือค่าจากฟังก์ชันทดสอบความคลาดเคลื่อนที่คำนวณได้มีค่าน้อยเท่าไรนั่นหมายความว่าค่าถ่วงน้ำหนักต่างๆ ภายในโครงสร้างของเครือข่ายประสาทเทียมยิ่งดีขึ้นหรือมีความเหมาะสมขึ้นนั่นเอง โดยการที่มีฟังก์ชันการวัดค่าความคลาดเคลื่อนนั้นทำให้เราสามารถปรับปรุงค่าถ่วงน้ำหนักได้ โดยส่วนมากในการคำนวณค่าความคลาดเคลื่อนนั้นมักจะใช้ค่า Mean Squared Error (MSE) โดยสามารถคำนวณได้จากสมการต่อไปนี้

$$MSE = \frac{1}{N}\sum (O_j - T_j)^2$$

ค่าต่างๆในสมการแทนได้ดังนี้

- N แทนจำนวนของชุดข้อมูลที่ใช้ในการสอน
- O_j แทนค่าเอาต์พุตของโหนด j ที่ได้จากเครือข่ายประสาทเทียม
- T_j แทนค่าเอาต์พุตจริงของโหนด j จากข้อมูลในอดีต

2.3.8. ข้อดีของเครือข่ายแบบ Back Propagation

- สามารถเรียนรู้ถึงความสัมพันธ์ของรูปแบบได้มาก โดยการเรียนรู้จากข้อมูลชุดฝึกอบรม
- มีความยืดหยุ่นในการเรียนรู้ คือสามารถทำการเลือกจำนวนเลเยอร์ได้
- สามารถนำมาสร้างระบบการตัดสินใจที่ซับซ้อนได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.9. ข้อเสียของเครือข่ายแบบ Back Propagation

- ต้องใช้เวลาในการเรียนรู้นานเมื่อเครือข่ายมีขนาดใหญ่
- มีโอกาสเกิดความล้มเหลวจากการเรียนรู้ของเครือข่ายประสาทเทียม ซึ่งแม้จะมีการเพิ่มข้อมูลเข้าไปก็ไม่สามารถปรับเครือข่ายประสาทเทียมได้ แต่สามารถแก้ไขได้โดยการเพิ่มจำนวนเลเยอร์ หรือทำการเปลี่ยนชุดของข้อมูลที่นำมาใช้ในการฝึกอบรม
- ต้องใช้เวลาในการประมวลผลเพื่อออกแบบให้เหมาะสม
- ค่าถ่วงน้ำหนักยากต่อการตีความ

2.3.10 การกำหนดจำนวนฮิดเดนเลเยอร์

การกำหนดจำนวนฮิดเดนเลเยอร์ของเครือข่ายประสาทเทียมนั้น ยังไม่มีวิธีการใดสามารถกำหนดค่าที่แน่นอนได้ แต่มีทฤษฎีกล่าวไว้ว่า ฮิดเดนเลเยอร์เพียงเลเยอร์เดียวก็เพียงพอแล้วสำหรับการนำไปใช้กับปัญหาที่ไม่เป็นเชิงเส้นใดๆ แต่ในบางกรณีการเพิ่มจำนวนฮิดเดนเลเยอร์สามารถช่วยให้การเรียนรู้ง่ายขึ้น รวดเร็วขึ้น และอีกปัญหาหนึ่งคือจำนวนโหนดในฮิดเดนเลเยอร์ควรจะมีจำนวนเท่าใด ซึ่งเช่นเดียวกันคือ ยังไม่มีกฎเกณฑ์ตายตัวใดๆกำหนดได้อย่างแน่นอนได้ เนื่องจากจำนวนโหนดในฮิดเดนเลเยอร์นี้ขึ้นอยู่กับความซับซ้อนของปัญหา มีเทคนิคหลายอย่างถูกนำเสนอออกมาเพื่อใช้หาจำนวนโหนดที่เหมาะสมในฮิดเดนเลเยอร์ ซึ่งวิธีที่นับว่าแพร่หลายพอสมควรคือใช้โหนดให้มีจำนวนมากไว้ก่อนแล้วค่อยๆลดจำนวนลงมา จนได้จำนวนที่น้อยที่สุดที่เครือข่ายประสาทเทียมยังคงเรียนรู้ได้ เทคนิคดังกล่าวเรียกว่าพรุนนิ่ง (Pruning)

ในการเลือกโครงสร้างของเครือข่ายประสาทเทียมสามารถพิจารณาจากแนวทางต่างๆได้ 3 แบบคือ

1. Ad Hoc Approach เป็นการออกแบบโครงสร้างโดยอาศัยพื้นฐานจากประสบการณ์การทดลองที่มีมาในครั้งก่อนๆ
2. Dynamic Approach เป็นขั้นตอนการเลือกแบบอัตโนมัติ ซึ่งมีแนวคิดว่าให้มีการปรับเพิ่มหรือลดจำนวนตัวแปรต่างๆในเครือข่ายประสาทเทียมลงในระหว่างกระบวนการสอน โดยพิจารณาจากค่าความผิดพลาดของระบบ
3. Distribution Approach เป็นแนวคิดที่พยายามค้นหาหรือตั้งสมมติฐานของ Distribution Function ในการประมาณและหาวิธีการทางสถิติเพื่อวัดประสิทธิภาพของเครือข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การวิเคราะห์และออกแบบระบบ

การวิเคราะห์และออกแบบระบบงานสำหรับการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ จะนำทฤษฎีของดาต้าไมน์นิ่งและหลักการของเครือข่ายประสาทเทียมมาประยุกต์ใช้ โดยนำดาต้าไมนิ่งมาใช้ในลักษณะของงานด้านพยากรณ์ข้อมูล ซึ่งใช้เทคนิคของเครือข่ายประสาทเทียมที่มีโครงสร้างแบบ Multilayer Feed-Forward และมีอัลกอริทึมในการเรียนรู้แบบ Back Propagation

3.1 การวิเคราะห์การทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่

การทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ มีได้หลายหลายรูปแบบ สามารถแบ่งได้เป็น 2 ประเภทดังนี้ (Chorleywood. 2002)

1. การคัดลอกหมายเลขโทรศัพท์เคลื่อนที่ (Mobile Phone Cloning)

การคัดลอกหมายเลขโทรศัพท์เคลื่อนที่ เป็นการทุจริตที่สามารถทำได้ง่ายมากในระบบเครือข่ายแบบแอนะล็อก แต่สามารถทำได้ยากในระบบ GSM เนื่องจากเครื่องโทรศัพท์ในระบบแอนะล็อกจะมีหมายเลขที่ใช้ระบุอยู่ 2 หมายเลขคือ Mobile Identification Number (MIN) ใช้สำหรับระบุหมายเลขของโทรศัพท์เคลื่อนที่ และ Electronic Serial Number (ESN) ใช้สำหรับระบุหมายเลขของผู้ให้บริการโทรศัพท์เคลื่อนที่ ซึ่งผู้ที่จะทำการทุจริตจะลักลอบนำหมายเลขเหล่านี้ไปใช้งาน โดยการลักลอบดักจับทางอากาศ แต่ในระบบเครื่องข่ายแบบดิจิทัล เช่นระบบ GSM จะทำการเข้ารหัสหมายเลขเหล่านี้ก่อนที่จะทำการส่งออกไป จึงทำให้ปลอดภัยกว่าระบบแอนะล็อก

การตรวจสอบลักษณะการทุจริตโดยการคัดลอกหมายเลขโทรศัพท์เคลื่อนที่สามารถทำได้หลายรูปแบบ ดังนี้

- การตรวจสอบการใช้งานที่มีระยะเวลายาวนานผิดปกติ ผู้ที่ทำการทุจริตส่วนมากจะมีการใช้งานที่ยาวนานผิดปกติ เช่น ในระยะเวลา 1 วันมีการใช้งานโทรศัพท์ติดต่อกันยาวนานถึง 20 ชม.
- การพิจารณาสถานที่ที่ใช้โทรศัพท์ที่มีความขัดแย้งกัน เช่น การโทรจากสถานที่สองแห่งที่ห่างไกลกันมาก ภายในระยะเวลาที่ใกล้เคียงกันเกินกว่าที่จะเดินทางจริงๆได้ เช่น ในเวลา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

12:00 มีการใช้โทรศัพท์จากจังหวัดเชียงใหม่ และเวลา 12:05 มีการใช้โทรศัพท์จากจังหวัดสงขลา ซึ่งในความเป็นจริง ไม่สามารถเดินทางจากจังหวัดเชียงใหม่ไปยังจังหวัดสงขลาภายในระยะเวลา 5 นาทีได้

- การใช้งานหมายเลขเดียวกันมากกว่าหนึ่งครั้งในเวลาเดียวกัน

2. การสมัครใช้บริการโดยตั้งใจที่จะทุจริต (Subscription Fraud)

การสมัครใช้บริการโดยตั้งใจที่จะทุจริต เป็นการทุจริตที่สามารถป้องกันได้ยากที่สุด เพราะการทุจริตประเภทนี้ไม่มีความเกี่ยวข้องใดๆ กับเทคโนโลยีหรือวิทยาการสมัยใหม่ โดยวิธีทำการทุจริตคือผู้ทุจริตจะเข้ามาสมัครเป็นลูกค้าเพื่อใช้บริการโทรศัพท์เคลื่อนที่ โดยที่ตั้งใจจะไม่ชำระค่าบริการอยู่แล้ว

ข้อมูลที่ใช้สำหรับวิเคราะห์การทุจริตมีดังนี้

- ความสัมพันธ์ระหว่างจำนวนเงินที่จ่ายกับการใช้งาน
- รูปแบบของเลขหมายที่ทำการโทร ในขณะที่ผู้ทำการทุจริตจะเปลี่ยนเลขหมายที่โทรออกตลอดเวลา แต่เลขหมายปลายทางมักจะเหมือนกันอยู่เสมอ และการโทรไปยังหมายเลขที่เรียงกัน ยังสามารถนำไปพิจารณาถึงการพยายามที่จะบุกรุก (Hack) ไปยังตู้สาขา (PABX) อีกด้วย
- รูปแบบการใช้บริการโทรศัพท์ที่แตกต่างจากลูกค้าทั่วไป เช่น การโทรศัพท์ไปต่างประเทศเป็นระยะเวลานานๆ
- สถานที่ที่ใช้โทรศัพท์ ใช้วิเคราะห์ในกรณีที่สถานที่ที่ใช้โทรศัพท์มีความขัดแย้งกันที่เกิดจากการคัดลอกหมายเลขโทรศัพท์เคลื่อนที่

3.2. วัตถุประสงค์ทางธุรกิจ

วัตถุประสงค์ในการทำระบบวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่คือ เพื่อทำการวิเคราะห์หากลุ่มลูกค้าที่มีโอกาสทำการทุจริตในการใช้บริการ เพื่อทำการป้องกันไม่ให้ลูกค้าเหล่านี้เข้ามาใช้บริการอีกต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3. การเตรียมข้อมูล

สามารถแบ่งประเภทของข้อมูลที่จะนำมาใช้ในการวิเคราะห์พฤติกรรมการทุจริตได้ดังนี้

- ข้อมูลของลูกค้า (Customer) เป็นรายละเอียดของลูกค้า เช่น ระยะเวลาในการเป็นลูกค้านับตั้งแต่เปิดบริการระบบโทรศัพท์ที่ใช้งาน เป็นต้น
- ข้อมูลการชำระเงิน (Payment) เช่น ข้อมูลการชำระเงินในแต่ละรายการ เช่นชำระเงินตรงตามกำหนดเวลาหรือไม่ เป็นต้น
- ข้อมูลจำนวนการใช้บริการ (Billing) เช่น จำนวนเงินเฉลี่ยที่ใช้บริการต่อเดือน เป็นต้น

โดยแอตทริบิวต์ที่เหมาะสมในการทำดาต้าไมน์นิ่งสำหรับการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ มีดังต่อไปนี้

- Age คือ อายุของลูกค้า
- Subscription Age คือ ระยะเวลาในการเป็นลูกค้านับตั้งแต่เปิดบริการ
- Salary คือ เงินเดือนของลูกค้า
- Credit Limit คือ วงเงินในการใช้บริการ
- Number of Change Address คือจำนวนครั้งที่ทำการเปลี่ยนที่อยู่
- Number of In Due Payment คือจำนวนครั้งที่ชำระเงินตรงตามกำหนดเวลา
- Number of Overdue Payment คือ จำนวนครั้งที่ชำระเงินเกินกำหนดเวลา
- Average Amount คือ จำนวนเงินเฉลี่ยที่ใช้บริการต่อเดือน
- Fraud คือ สถานะของลูกค้าว่าเป็นลูกค้าที่ทำการทุจริตหรือไม่

3.4. การออกแบบลักษณะของข้อมูล

ข้อมูลที่ใช้ในระบบจะแบ่งออกเป็น 2 ส่วนคือ ส่วนข้อมูลชุดฝึกอบรมที่ใช้สำหรับการสอนและ การทดสอบระบบ และข้อมูลที่ต้องการจะนำไปวิเคราะห์หาผู้ที่ทำการทุจริต ซึ่งข้อมูลชุดฝึกอบรมจะถูกเตรียมมาจากข้อมูลของลูกค้าในอดีตที่เก็บไว้ในระบบคลังข้อมูล ส่วนข้อมูลที่จะนำไปวิเคราะห์หาผู้ที่ทำการทุจริตจะถูกเลือกข้อมูลมาเก็บไว้ตามความต้องการของผู้ใช้งาน ซึ่งจากการรวบรวมความต้องการของผู้ใช้งานมา พบว่าข้อมูลที่ต้องการนำไปวิเคราะห์คือลูกค้าที่เปิดบริการมาไม่ต่ำกว่า 3 เดือนและมีการเปลี่ยนที่อยู่อย่างน้อย 1 ครั้ง ดังนั้นจึงทำการเลือกข้อมูลเหล่านี้จากระบบคลังข้อมูล โดยมีการแปลงค่าของข้อมูลเพื่อให้เหมาะสมกับการนำไปวิเคราะห์ และเก็บลงฐานข้อมูล โดยแต่ละตารางโดยมีโครงสร้างดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ข้อมูลชุดฝึกอบรมที่ใช้สำหรับการสอนและการทดสอบระบบ

TRAIN_ANLS	
COLUMN NAME	**DATA TYPE**
subr_numb	INTEGER
age	INTEGER
subr_age	INTEGER
salary	INTEGER
crlm_amnt	INTEGER
chng_addr	INTEGER
numb_in_due	INTEGER
numb_over_due	INTEGER
avrg_usag	MONEY(12,2)
Fraud	INTEGER

ตารางที่ 3.2 ข้อมูลที่ใช้การนำไปวิเคราะห์หาผู้ที่ทำการทุจริต

FRAUD_ANLS	
COLUMN NAME	**DATA TYPE**
subr_numb	INTEGER
age	INTEGER
subr_age	INTEGER
salary	INTEGER
crlm_amnt	INTEGER
chng_addr	INTEGER
numb_in_due	INTEGER
numb_over_due	INTEGER
avrg_usag	MONEY(12,2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 รายละเอียดของข้อมูลที่ใช้ในการฝึกอบรมเครือข่ายประสาทเทียม

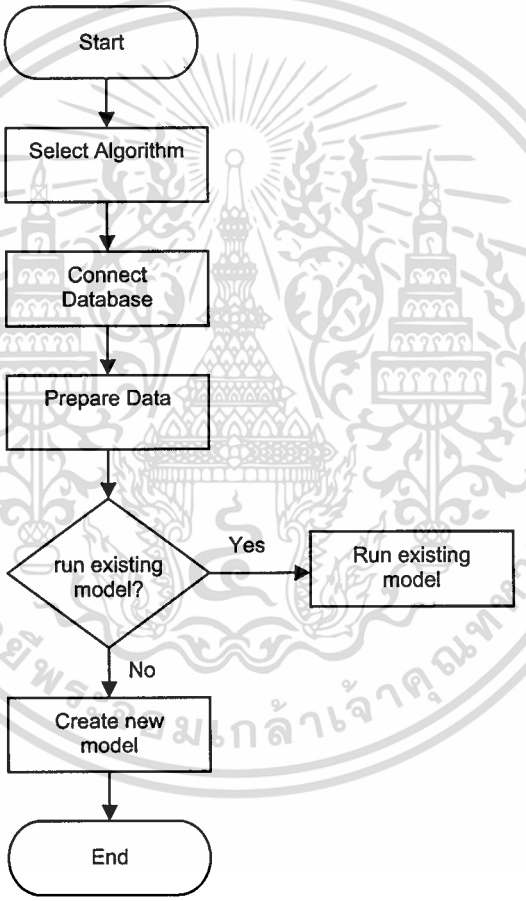
ชื่อ Field	ชื่อภาษาอังกฤษ	ชื่อภาษาไทย	ตัวอย่างข้อมูล
subr_numb	subscriber number	หมายเลขเบอร์โทรศัพท์ของลูกค้า	6611111111
age	subscriber's age	อายุของลูกค้า (วัน)	1136
subr_age	subscription's age	ระยะเวลาในการเป็นลูกค้านับตั้งแต่เปิดบริการ(วัน)	890
salary	Salary	เงินเดือน	25000
crlm_amnt	Credit limit amount	วงเงินในการใช้บริการ	4000
chng_addr	Number of change address	จำนวนครั้งในการเปลี่ยนที่อยู่	0
numb_in_due	Number of in due payment	จำนวนครั้งในการชำระเงินตรงเวลา	20
numb_over_due	Number of over due payment	จำนวนครั้งในการชำระเงินเกินกำหนดระยะเวลา	4
avrg_usag	Average usage	จำนวนเงินเฉลี่ยที่ใช้บริการต่อเดือน	1200
fraud	fraud	สถานะของลูกค้าที่ทำการทุจริต ถ้าข้อมูลเป็น 0 แสดงถึงสถานะของลูกค้าที่ไม่ทุจริต ถ้าข้อมูลเป็น 1 แสดงถึงสถานะของลูกค้าที่ทำการทุจริต	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5. การออกแบบระบบ

ระบบจะแบ่งการทำงานออกเป็น 5 ส่วนดังนี้

1. ส่วนการเลือกอัลกอริทึม
2. ส่วนการติดต่อกับฐานข้อมูล
3. ส่วนการเตรียมข้อมูล
4. ส่วนการสร้างแบบจำลอง
5. ส่วนการใช้งานแบบจำลอง

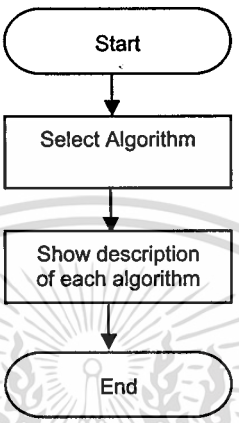


รูปที่ 3.1 แผนภาพแสดงการทำงานของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.1. ส่วนการเลือกอัลกอริทึม

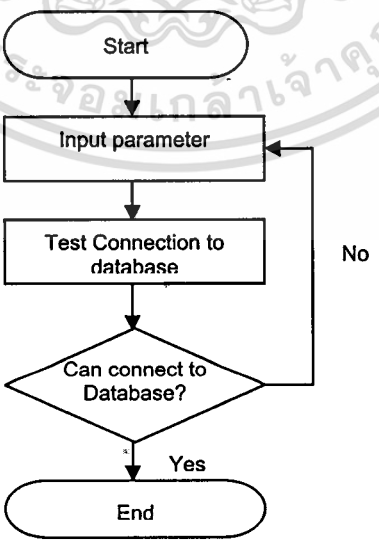
ส่วนการเลือกอัลกอริทึม เป็นการให้ผู้ใช้งานเลือกอัลกอริทึมของเครือข่ายประสาทเทียมที่ต้องการใช้งาน



รูปที่ 3.2 แผนภาพแสดงส่วนการเลือกอัลกอริทึม

3.5.2. ส่วนติดต่อฐานข้อมูล

ส่วนติดต่อฐานข้อมูล จะรับข้อมูลของชื่อผู้ใช้งาน รหัสผ่าน ชื่อฐานข้อมูล และชื่อเครื่องเซิร์ฟเวอร์เพื่อทำการติดต่อกับฐานข้อมูล

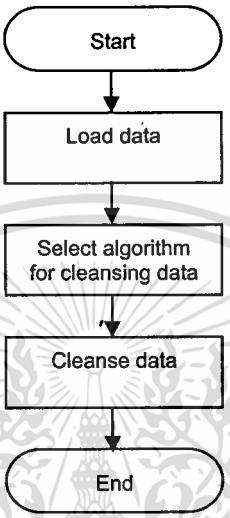


รูปที่ 3.3 แผนภาพแสดงส่วนติดต่อฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.3. ส่วนการเตรียมข้อมูล

ส่วนการเตรียมข้อมูลเป็นการเตรียมข้อมูลให้สมบูรณ์ครบถ้วน โดยให้ผู้ใช้งานอินพุตข้อมูลเข้ามาในระบบ และเลือกอัลกอริทึมที่เหมาะสมในการจัดการกับข้อมูลที่ขาดหายไป



รูปที่ 3.4 แผนภาพแสดงส่วนการเตรียมข้อมูล

3.5.4. ส่วนการสร้างแบบจำลอง

ในการสร้างแบบจำลองของเครือข่ายประสาทเทียม จะใช้อัลกอริทึมแบบ Back Propagation ซึ่งผู้ใช้งานสามารถกำหนดค่าตัวแปรต่างๆได้ เพื่อช่วยให้ระบบมีการเรียนรู้ได้อย่างมีประสิทธิภาพมากขึ้น สำหรับการสร้างแบบจำลองของระบบนั้น ประกอบไปด้วยขั้นตอนดังต่อไปนี้

3.5.4.1. การกำหนดตัวแปรต่างๆให้ระบบ

ในขั้นตอนแรก ผู้ใช้จะต้องกำหนดค่าตัวแปรต่างๆ ที่จำเป็นในการสร้างแบบจำลอง เช่น จำนวนอินพุตเลเยอร์ จำนวนเอาต์พุตเลเยอร์ จำนวนฮิดเดนเลเยอร์ จำนวนรอบของการเรียนรู้ ค่าความคลาดเคลื่อนที่ยอมรับได้ จำนวนข้อมูลที่ใช้ในการสอน และจำนวนข้อมูลที่ใช้ในการทดสอบระบบ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนในการกำหนดตัวแปรให้ระบบ

1. ระบบอ่านค่าแอตทริบิวต์ทั้งหมดมาจากฐานข้อมูล
2. ผู้ใช้ทำการเลือกว่าจะให้แอตทริบิวต์ใดบ้างเป็นอินพุตโหนด และเอาต์พุตโหนด
3. ระบบจะกำหนดให้จำนวนแอตทริบิวต์ที่ผู้ใช้เลือกเป็นอินพุตโหนดเป็นจำนวนอินพุตเลเยอร์ และจำนวนแอตทริบิวต์ที่ผู้ใช้เลือกเป็นเอาต์พุตโหนดเป็นจำนวนเอาต์พุตเลเยอร์
4. ผู้ใช้กำหนดจำนวนอินพุตเลเยอร์ จำนวนเอาต์พุตเลเยอร์ จำนวนฮิดเดนเลเยอร์ จำนวนรอบของการเรียนรู้ ค่าความคลาดเคลื่อนที่ยอมรับได้ จำนวนข้อมูลที่ใช้ในการสอน และจำนวนข้อมูลที่ใช้ในการทดสอบระบบ

3.5.4.2. การสอนให้ระบบเรียนรู้

เมื่อกำหนดตัวแปรเรียบร้อยแล้ว จึงการสอนให้ระบบเกิดการเรียนรู้ โดยการนำข้อมูดชุดฝึกอบรม มาทำตามอัลกอริทึมแบบ Back Propagation เพื่อปรับตัวแปรภายในเครือข่ายจนกระทั่งได้แบบจำลองที่เหมาะสม โดยระบบจะใช้ค่าจำนวนรอบของการเรียนรู้ และค่าความผิดพลาดที่ยอมรับได้ที่ผู้ใช้กำหนดไว้มาเป็นเงื่อนไขในการจบการเรียนรู้ของแบบจำลอง

ขั้นตอนในการสอนให้ระบบเรียนรู้

1. ทำการแปลงค่าข้อมูลให้เหมาะสมกับอัลกอริทึมที่ใช้สร้างแบบจำลอง
2. ทำการสร้างแบบจำลองตามข้อมูลที่ผู้ใช้กำหนด
3. อ่านข้อมูลชุดฝึกอบรมจากฐานข้อมูล
4. ทำการสอนระบบตามอัลกอริทึม Back Propagation

อัลกอริทึมที่ใช้ในการสอนให้ระบบเรียนรู้

Algorithm: Data Transform

Method:

1) Read data V from database.
2) Find 'V' for or all V

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$V' = [(\frac{V - \min_A}{\max_A - \min_A})(\max'_A - \min'_A)] + \min'_A$$

Algorithm: Back propagation

Method:

1) Initialize all weights and biases in network.
2) While terminating condition is not satisfied {
3) For each training sample X in samples {
4) //Propagate the inputs forward:
5) For each hidden or output layer unit j {
6) $I_j = \sum_1^n W_{ij} O_i + \theta_j$; //compute the net input of unit j
7) $O_j = \frac{1}{1+e^{-I_j}}$; //compute the output of each unit j
8) For each unit j in the output layer
9) $Err_j = O_j(1 - O_j)(T_j - O_j)$;// compute the error
10) For each unit j in the hidden layers, form the last to the first hidden layer
11) $Err_j = O_j(1-O_j)\sum Err_k W_{jk}$; //compute the error
12) For each weight W_{ij} in network {
13) $\Delta W_{ij} = (l) Err_{ij} O_i$;//weight increment
14) $\Delta W_{ij} = W_{ij} + \Delta W_{ij}$; }//weight update
15) For each bias θ_j in network {
16) $\Delta \theta_j = (l) Err_j$; //bias increment
17) $\Delta \theta_j = \theta_j + \Delta \theta_j$ } //bias update
18) }}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

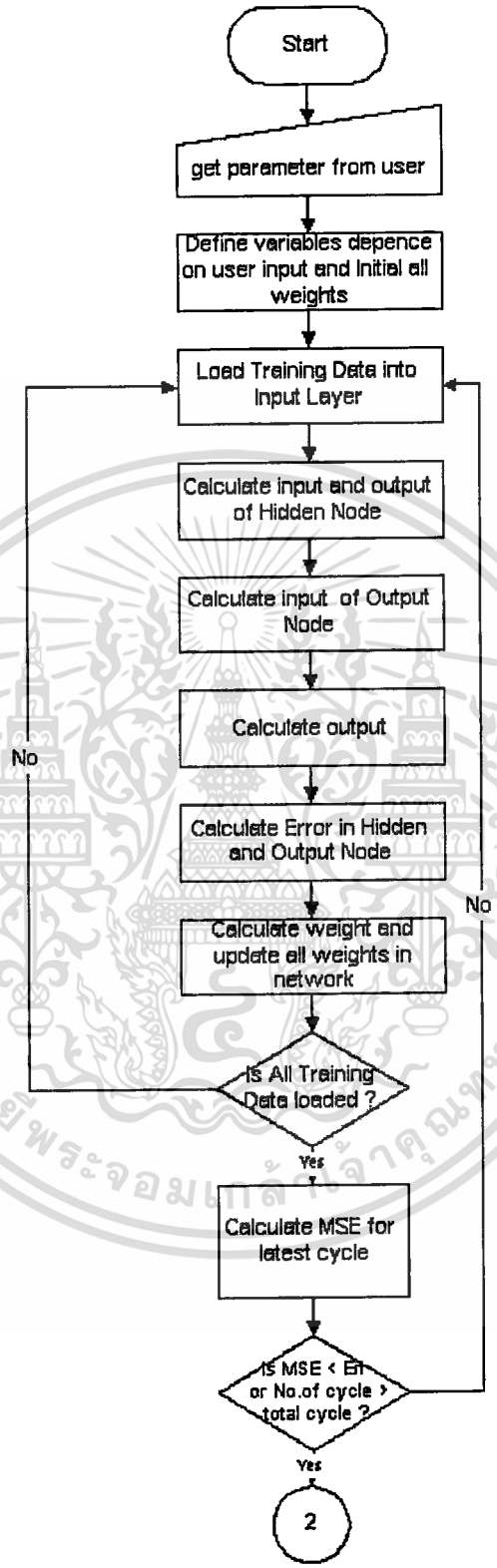
3.5.4.3 การทดสอบแบบจำลองที่ได้จากการสอน

การทดสอบแบบจำลองที่ได้จากการสอน เป็นการประเมินแบบจำลองโดยใช้ข้อมูลชุดสอบ มาทำการคำนวณหาเอาต์พุตของเครือข่ายเพื่อเปรียบเทียบกับค่าจริงที่เกิดขึ้น แล้วหาค่าความคลาดเคลื่อนเพื่อประเมินประสิทธิภาพของแบบจำลองที่ได้ ซึ่งถ้าทดสอบแล้วประสิทธิภาพของระบบเป็นที่น่าพอใจ ผู้ใช้สามารถบันทึกแบบจำลองนี้เก็บไว้ เพื่อนำไปใช้ในการพยากรณ์ได้ต่อไป

<u>ขั้นตอนในการทดสอบแบบจำลองที่ได้จากการสอน</u>

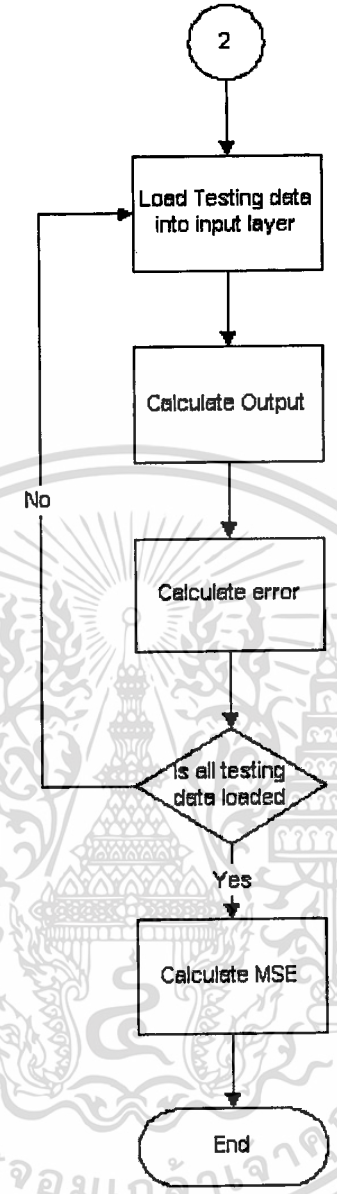
1. ระบบจะอ่านข้อมูลชุดฝึกอบรมเพื่อนำไปทดสอบ โดยจำนวนข้อมูลที่ทำการทดสอบจะถูกกำหนดโดยผู้ใช้
2. นำข้อมูลที่อ่านได้ไปทดสอบกับแบบจำลอง และแจ้งผลการทดสอบให้ผู้ใช้งานรับทราบ
3. ถ้าผู้ใช้งานพอใจผลการทดสอบสามารถบันทึกแบบจำลองเพื่อนำไปใช้งานต่อไปได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 แผนภาพแสดงการสอนให้ระบบเรียนรู้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 แผนภาพแสดงส่วนการทดสอบแบบจำลอง

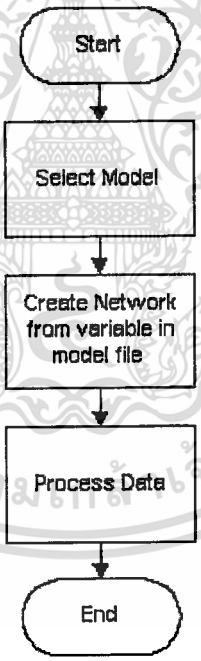
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5.5. ส่วนการใช้งานแบบจำลอง

การใช้งานแบบจำลอง เป็นการเลือกแบบจำลองที่สร้างไว้แล้วมาใช้งาน ซึ่งในการใช้งานแบบจำลอง จะประกอบไปด้วยขั้นตอนดังต่อไปนี้

<u>ขั้นตอนในการใช้งานแบบจำลอง</u>

1. ผู้ใช้สามารถเลือกแบบจำลองของระบบที่เก็บไว้ หลังจากการฝึกสอนจนเป็นที่น่าพอใจแล้ว ขึ้นมาทำการใช้งานได้ตามความเหมาะสมกับข้อมูลที่จะนำมาใช้ในการพยากรณ์
2. ระบบอ่านค่าจากไฟล์ข้อมูลที่ผู้ใช้เลือก และทำการสร้างแบบจำลองขึ้นมาใหม่โดยใช้ค่าต่างๆที่เก็บไว้ในไฟล์
3. อ่านข้อมูลจากฐานข้อมูล
4. นำข้อมูลมาคำนวณหาค่าเอาต์พุต



รูปที่ 3.7 แผนภาพแสดงส่วนการใช้งานแบบจำลอง

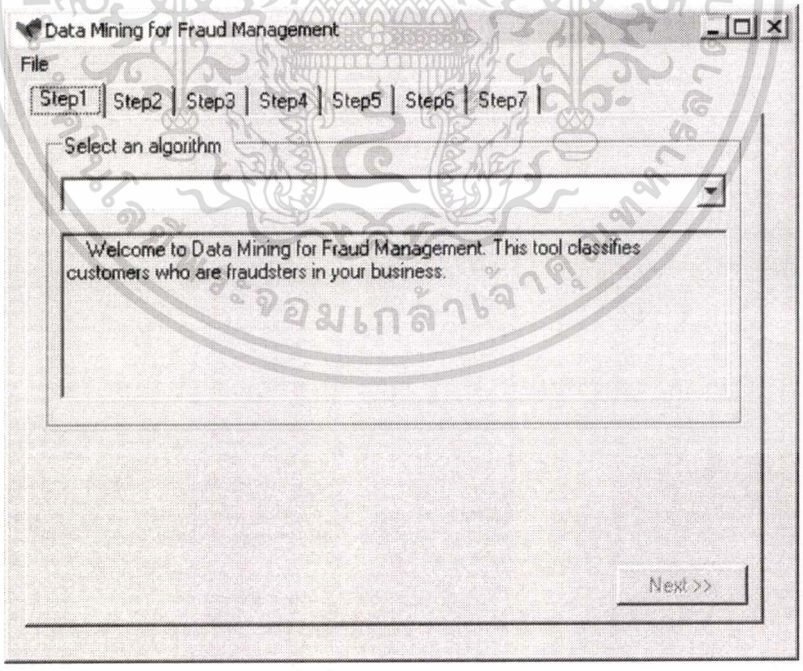
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การพัฒนาระบบ

ในบทนี้จะกล่าวถึงรายละเอียดของการใช้งานระบบงานที่ทำการพัฒนาขึ้น โดยการพัฒนาระบบจะใช้ Visual Basic .Net และ Microsoft SQL Server Version 2000 มาใช้ในการพัฒนาโปรแกรม โดยขั้นตอนการใช้งานระบบวิเคราะห์พฤติกรรมการทุจริตในการใช้งานบริการโทรศัพท์เคลื่อนที่ มีดังต่อไปนี้

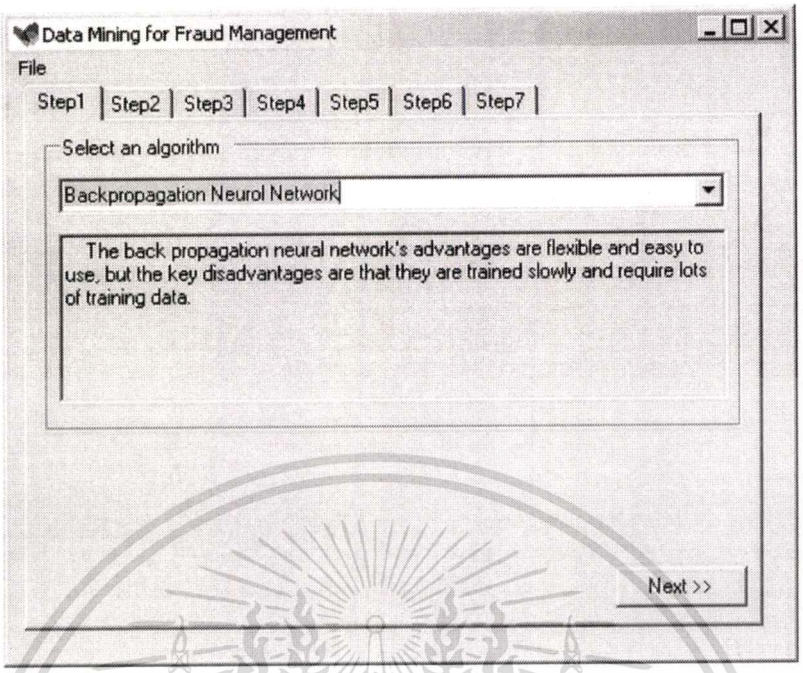
4.1. ส่วนการเลือกอัลกอริทึม

ส่วนการเลือกอัลกอริทึม เป็นหน้าจอแรกของระบบการวิเคราะห์พฤติกรรมการทุจริตของผู้ใช้บริการโทรศัพท์เคลื่อนที่ โดยระบบจะให้เลือกว่าจะใช้อัลกอริทึมใดในการวิเคราะห์ และเมื่อเลือกอัลกอริทึมเรียบร้อยแล้ว จะมีคำอธิบายข้อดีและข้อเสียของอัลกอริทึมนั้นด้วย



รูปที่ 4.1 หน้าจอหลักของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



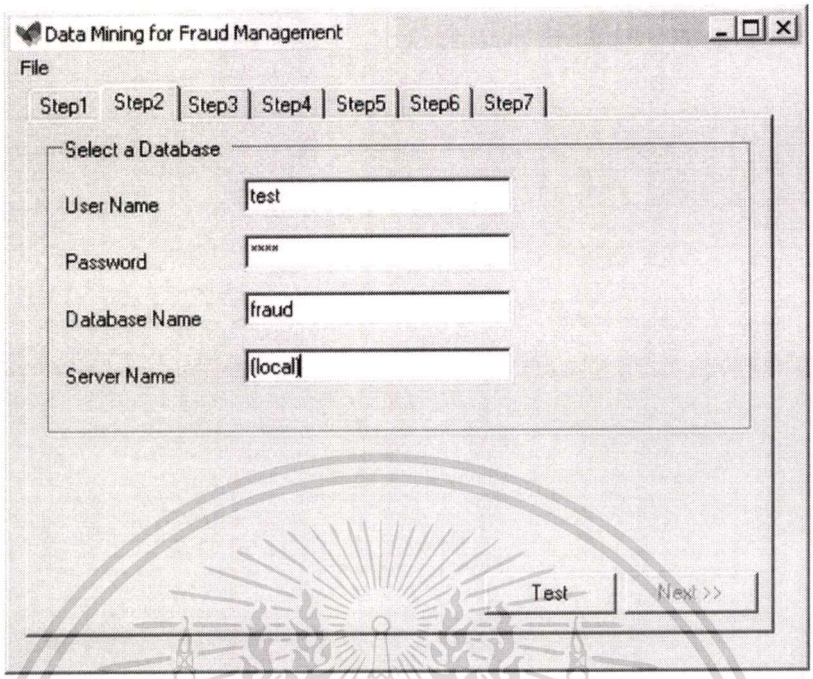
รูปที่ 4.2 หน้าจอเมื่อทำการเลือกอัลกอริทึมเรียบร้อย

4.2. ส่วนติดต่อฐานข้อมูล

เมื่อเลือกอัลกอริทึมเสร็จเรียบร้อยแล้ว ระบบจะทำการติดต่อกับฐานข้อมูลโดยมีขั้นตอนในการทำงานดังต่อไปนี้

1) ใส่ชื่อผู้ใช้ รหัสผ่าน ชื่อฐานข้อมูล และชื่อเครื่องเซิร์ฟเวอร์ที่เก็บฐานข้อมูลที่ต้องการจะติดต่อ
2) ทำการทดสอบว่าระบบสามารถติดต่อฐานข้อมูลได้หรือไม่ ถ้าได้จึงจะสามารถเริ่มใช้งานระบบได้ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 หน้าจอรับข้อมูลในการติดต่อกับฐานข้อมูล

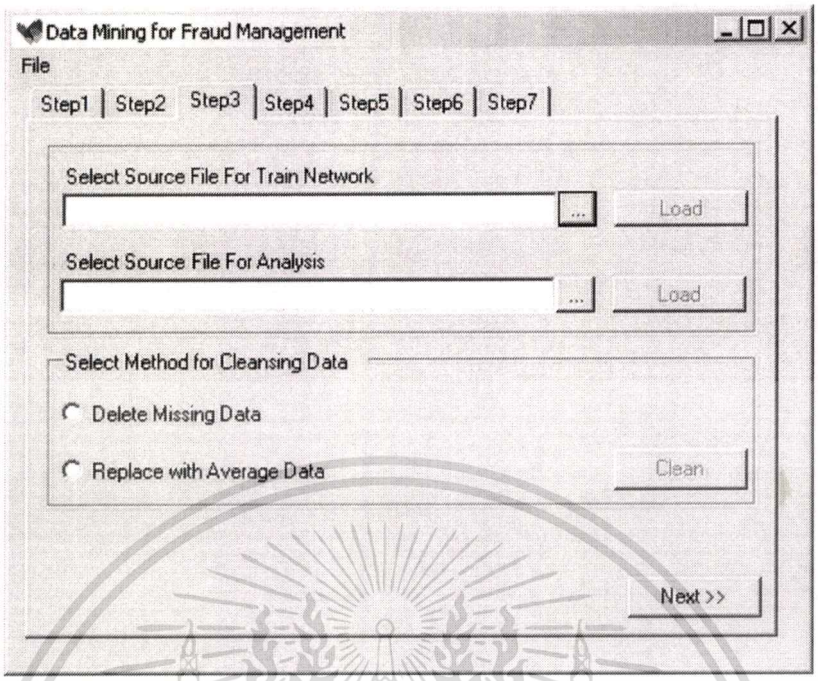
รูปที่ 4.4 ข้อความเมื่อทำการติดต่อฐานข้อมูลสำเร็จ

4.3. ส่วนการเตรียมข้อมูล

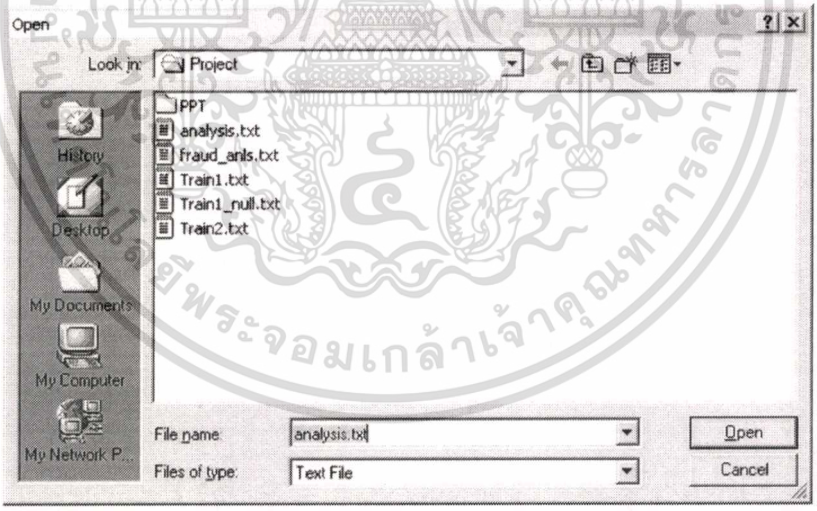
ส่วนการเตรียมข้อมูล มีขั้นตอนการทำงานดังนี้

1. โหลดข้อมูลเข้าไปในระบบ
2. เลือกอัลกอริทึมที่เหมาะสมในการจัดการกับข้อมูลที่ขาดหายไป
3. เมื่อเตรียมข้อมูลเรียบร้อยแล้ว จะมีหน้าจอให้เลือกว่าผู้ใช้ต้องการจะสร้างแบบจำลองใหม่หรือเลือกจากแบบจำลองเดิมที่มีอยู่แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

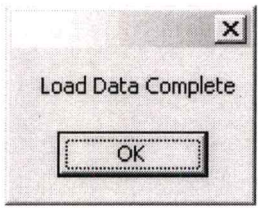


รูปที่ 4.5 หน้าจอการเตรียมข้อมูล

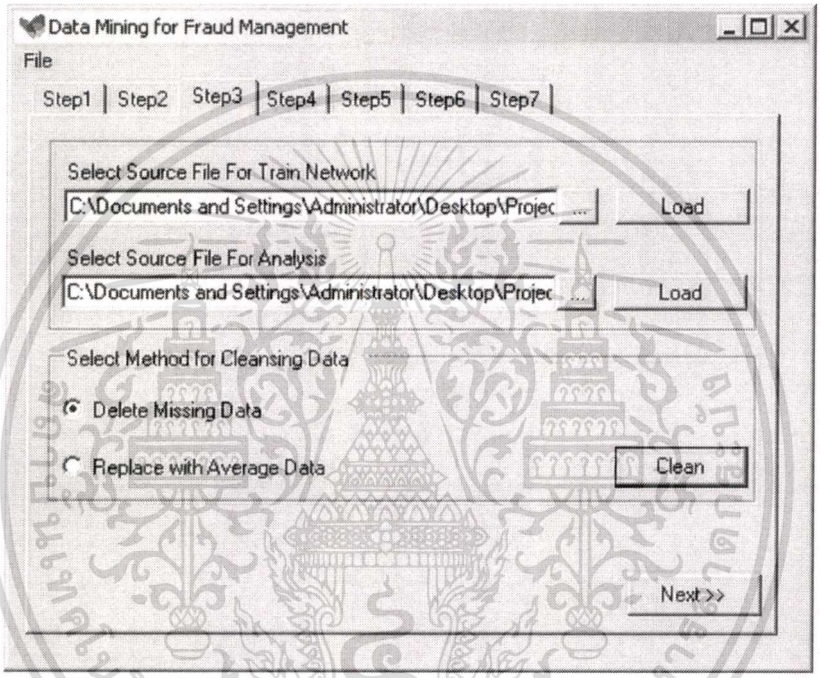


รูปที่ 4.6 หน้าจอการเลือกไฟล์ที่ต้องการโหลดเข้าในระบบ

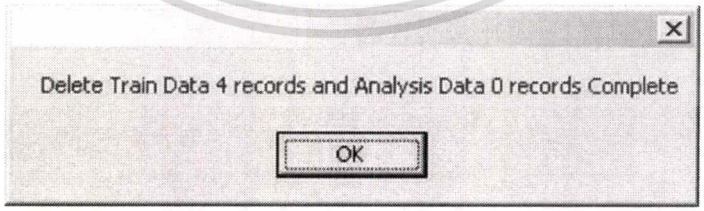
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.7 ข้อความเมื่อทำการโหลดข้อมูลเสร็จเรียบร้อย

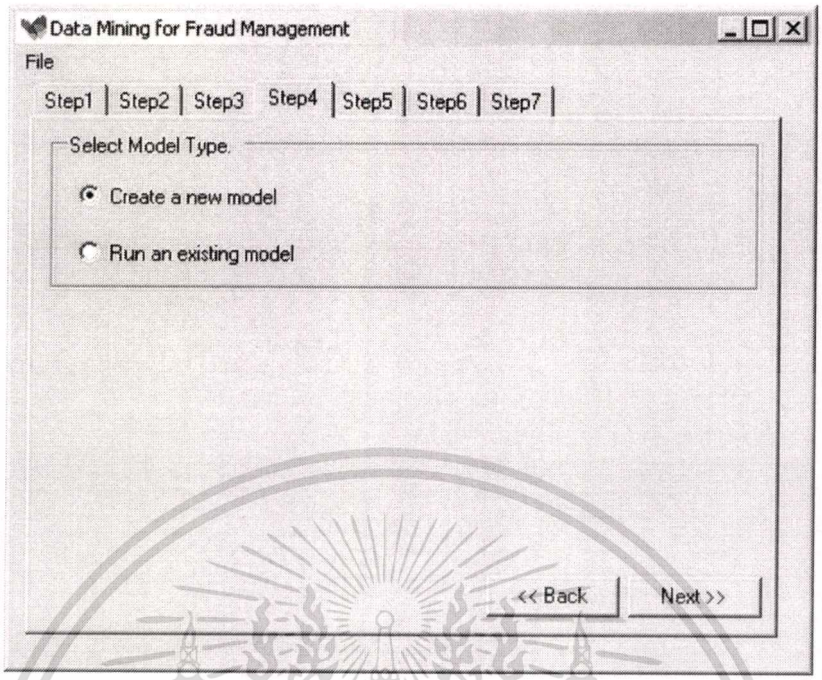


รูปที่ 4.8 หน้าจอเลือกอัลกอริทึมในการทำความสะอาดข้อมูลที่ขาดหายไป



รูปที่ 4.9 ข้อความเมื่อทำความสะอาดข้อมูลเรียบร้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 หน้าจอให้เลือกว่าต้องการจะสร้างแบบจำลองใหม่หรือว่ารันจากแบบจำลองเดิม

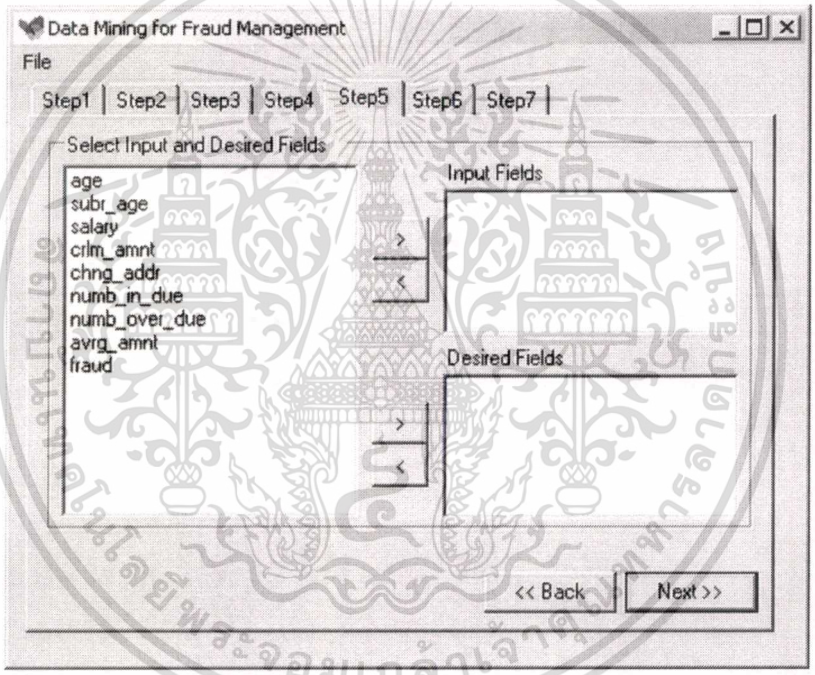
4.4 ส่วนการสร้างแบบจำลอง

ขั้นตอนในการสร้างแบบจำลองระบบ มีดังต่อไปนี้

1. ผู้ใช้กำหนดค่าตัวแปรที่เป็นอินพุตฟิลด์ (Input Field) และฟิลด์ที่ต้องการพยากรณ์ (Desired Field)
2. จากนั้นระบบจะทำการกำหนดให้จำนวนอินพุตยูนิต มีค่าเท่ากับจำนวนฟิลด์อินพุตที่ผู้ใช้ใส่เข้ามา และกำหนดให้จำนวนเอาท์พุตยูนิต มีค่าเท่ากับจำนวนฟิลด์ที่ต้องการพยากรณ์ ที่ผู้ใช้กำหนด
3. ผู้ใช้กำหนดค่าฮิดเดนยูนิต จำนวนรอบสูงสุดที่จะให้เครือข่ายเรียนรู้ และค่าความคลาดเคลื่อนที่ยอมรับได้
4. ผู้ใช้ทำการเลือกจำนวนของข้อมูลที่ใช้ฝึกอบรม และจำนวนข้อมูลที่ใช้ทดสอบ โดยกำหนดเป็นร้อยละของจำนวนข้อมูลชุดฝึกอบรมทั้งหมด
5. เมื่อกำหนดค่าต่างๆ เรียบร้อยแล้วให้กดปุ่ม Train เพื่อทำการฝึกอบรมเครือข่าย
6. เมื่อฝึกอบรมเสร็จเรียบร้อยแล้ว ระบบจะแจ้งผลของการฝึกอบรมให้ทราบ โดยผู้ใช้สามารถกดปุ่ม Report เพื่อให้ระบบแสดงผลการฝึกอบรมในรูปแบบของรายงานและกราฟ

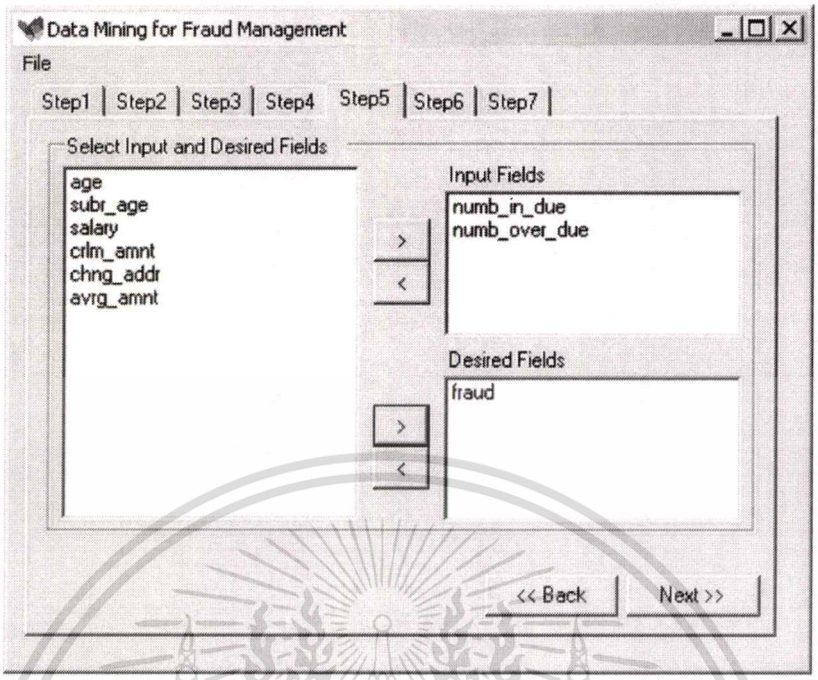
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. เมื่อฝึกอบรมเครือข่ายจนเป็นที่พอใจแล้ว ให้ทำการทดสอบเครือข่ายโดยการกดปุ่ม Test
8. เมื่อระบบทำการทดสอบเสร็จเรียบร้อยแล้ว ระบบจะแจ้งผลของการทดสอบให้ทราบ โดยผู้ใช้สามารถกดปุ่ม Report เพื่อให้ระบบแสดงผลการทดสอบ ในรูปแบบของรายงานและกราฟ
9. ผู้ใช้สามารถบันทึกเครือข่ายที่ทำการฝึกอบรมจนเป็นที่พอใจแล้ว ได้ในรูปแบบของ *.mdl เพื่อให้สามารถนำเครือข่ายมาใช้ในการใช้งานได้ในภายหลัง โดยการกดปุ่ม Save

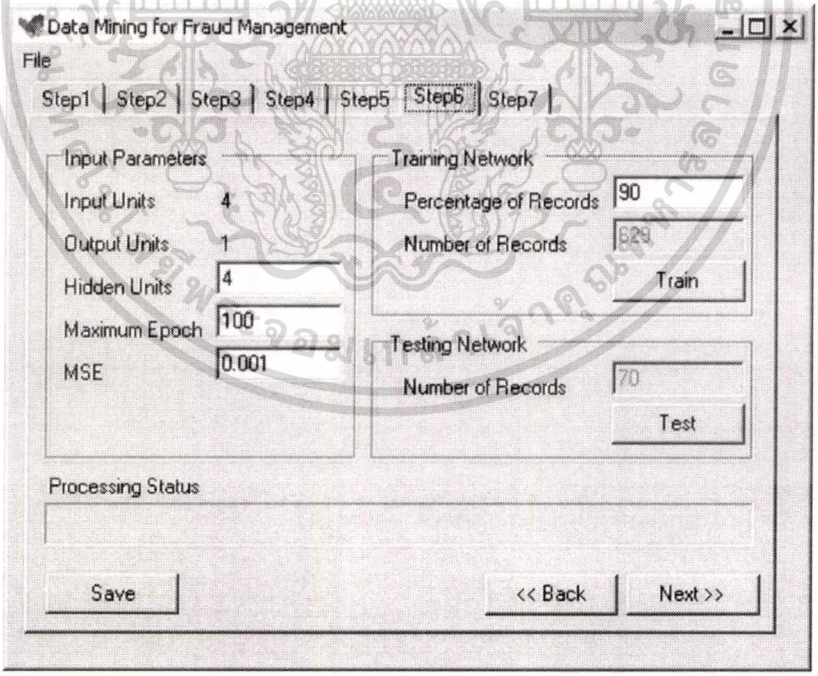


รูปที่ 4.11 หน้าจอแสดงแอตทริบิวต์ของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

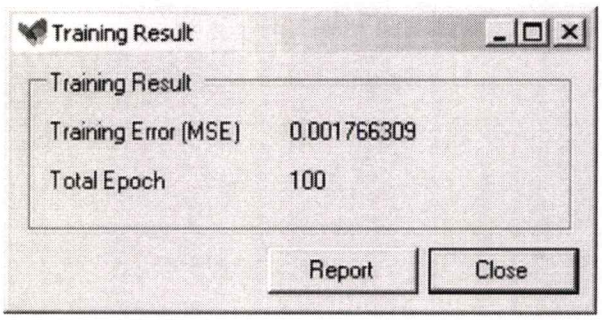


รูปที่ 4.12 หน้าจอรับข้อมูลในการสร้างเครือข่ายประสาทเทียม

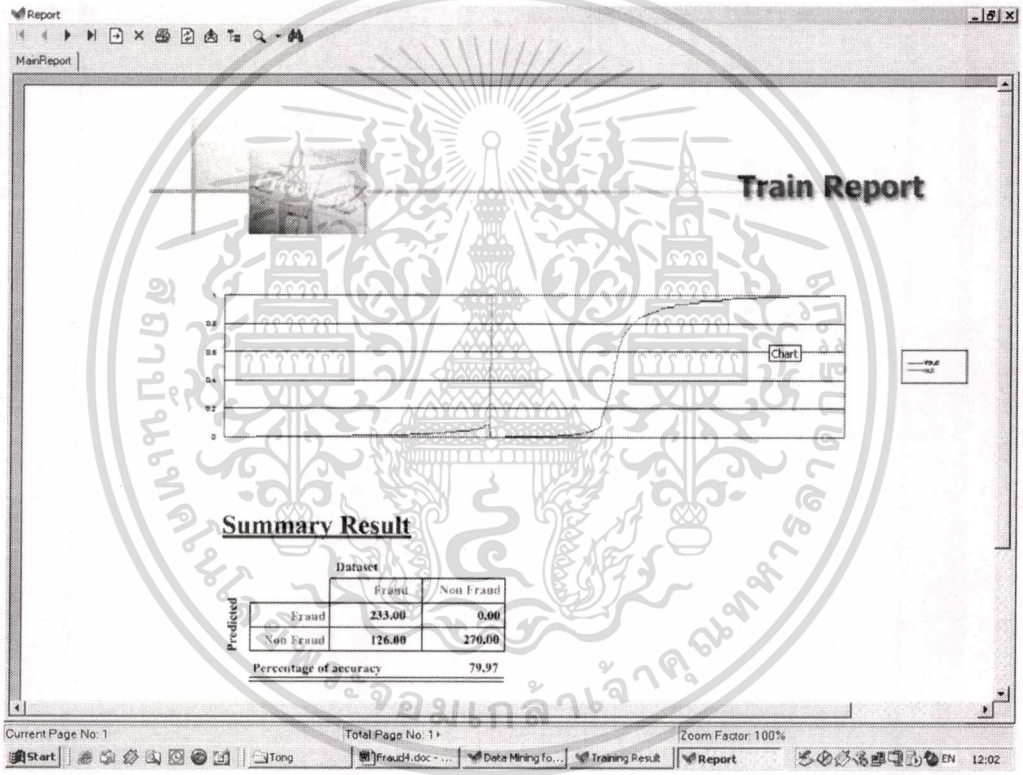


รูปที่ 4.13 หน้าจอฝึกอบรมเครือข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

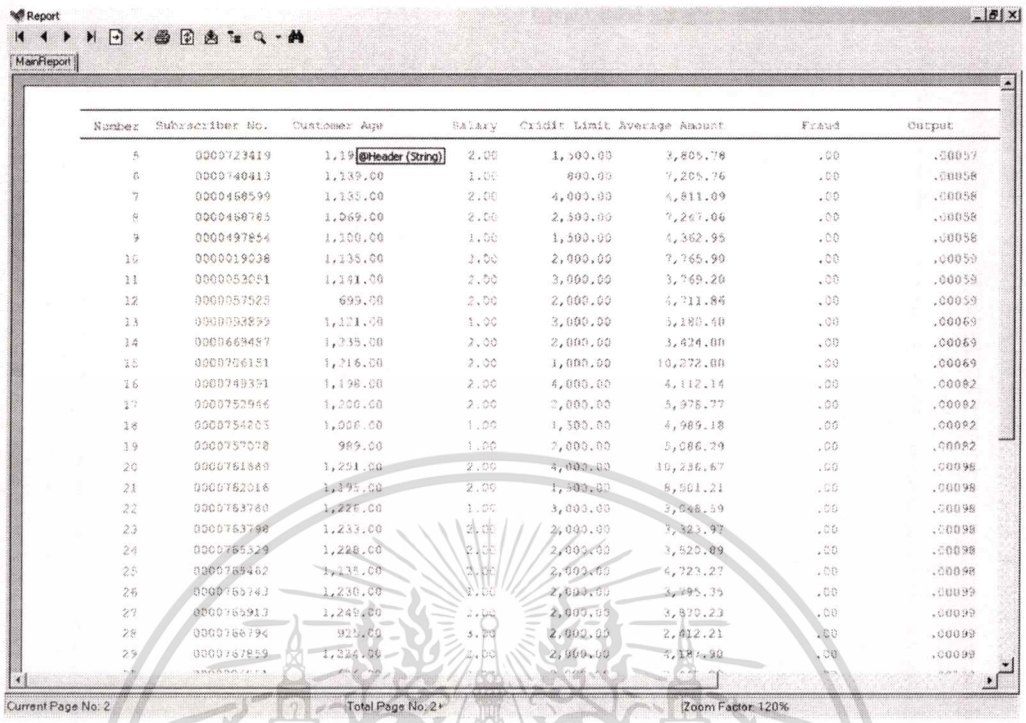


รูปที่ 4.14 หน้าจอแสดงผลการฝึกอบรม

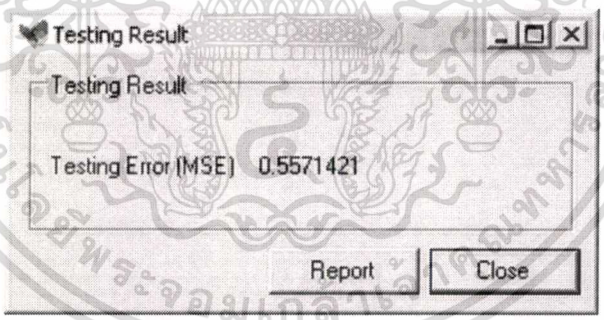


รูปที่ 4.15 รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

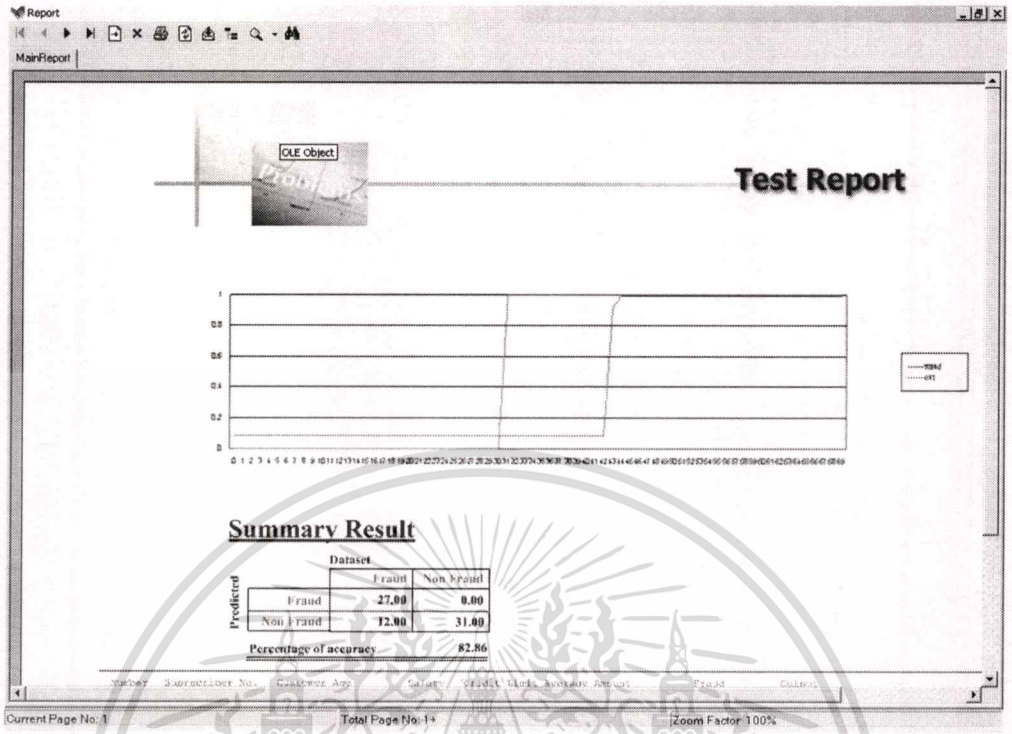


รูปที่ 4.16 รายงานรายละเอียดของผลการฝึกอบรมเครือข่ายประสาทเทียม

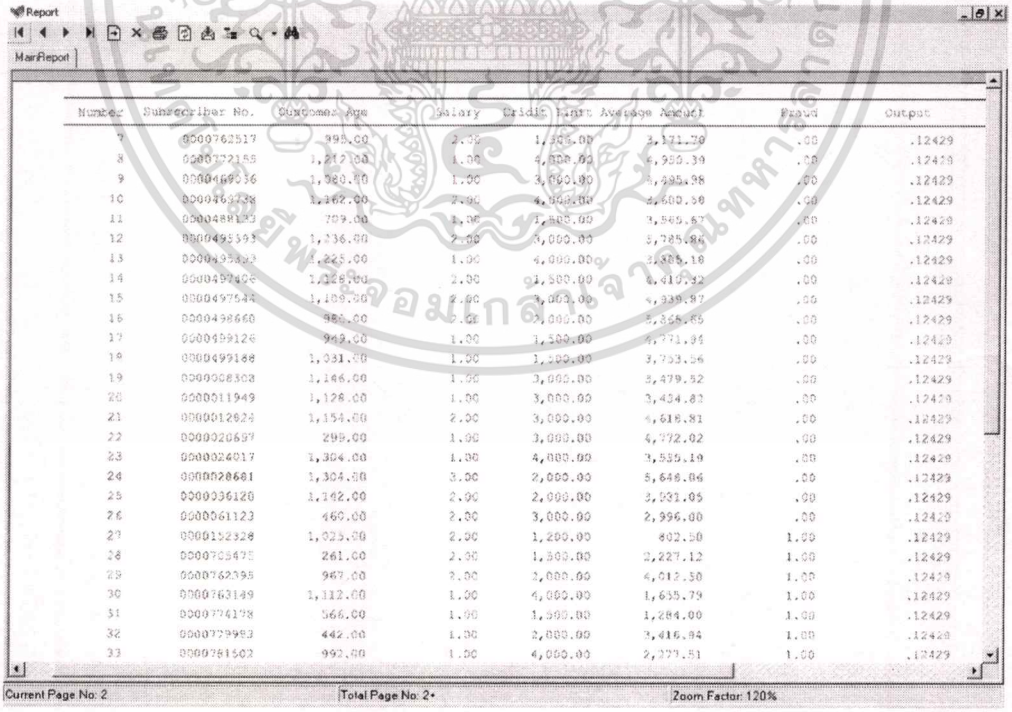


รูปที่ 4.17 หน้าจอแสดงผลการทดสอบเครือข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.18 รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียม



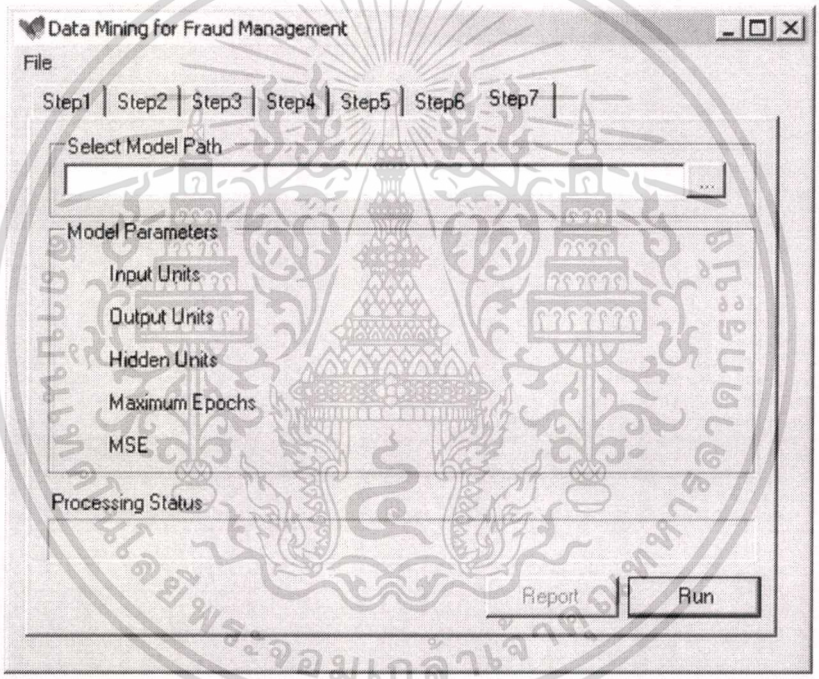
รูปที่ 4.19 รายงานรายละเอียดของผลการทดสอบเครือข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5. ส่วนการใช้งานแบบจำลอง

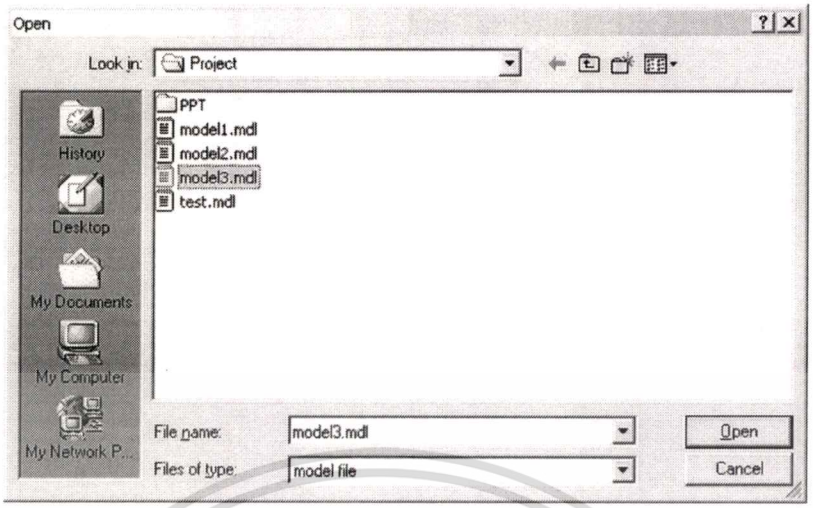
ในการใช้งานแบบจำลอง มีขั้นตอนดังต่อไปนี้

1. เลือกแบบจำลองที่ต้องการจะนำไปใช้งาน
2. เมื่อเลือกแบบจำลองเสร็จเรียบร้อยแล้ว ระบบจะทำการแสดงรายละเอียดของแบบจำลองนั้น
3. ผู้ใช้กดปุ่ม Run เพื่อทำการคำนวณผลลัพธ์จากระบบ
4. ผู้ใช้สามารถกดปุ่ม Report เพื่อให้ระบบแสดงผลลัพธ์จากระบบในรูปแบบของรายงานและกราฟ

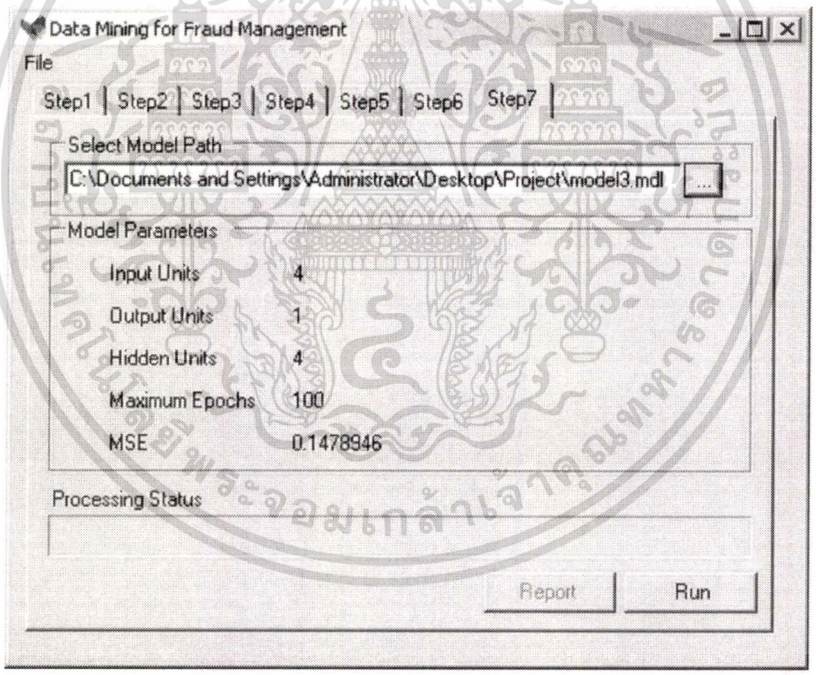


รูปที่ 4.20 หน้าจอในการใช้งานแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

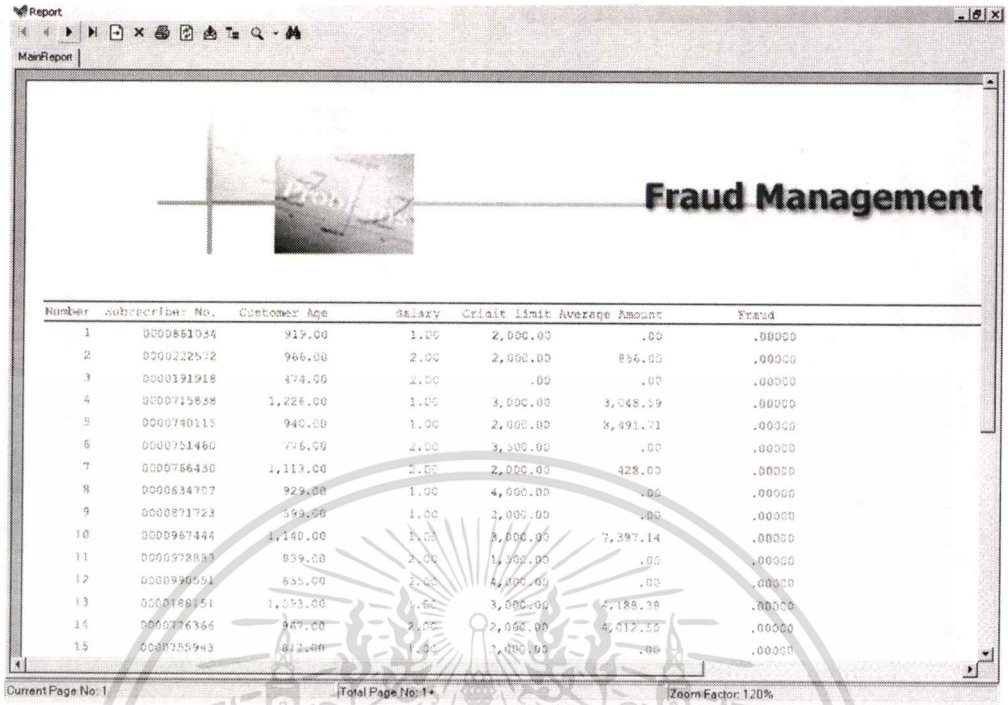


รูปที่ 4.21 หน้าจอการเลือกไฟล์แบบจำลองที่ต้องการ



รูปที่ 4.22 หน้าจอการใช้งานแบบจำลองเมื่อเลือกแบบจำลองเรียบร้อยแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.23 รายงานแสดงผลการพยากรณ์จากแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
การทดลองการวิเคราะห์พฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่

5.1. วัตถุประสงค์ทางธุรกิจ

เพื่อทำการแบ่งกลุ่มผู้ที่มีพฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ โดยมีวัตถุประสงค์ในการทำการทดลอง 2 ข้อ ดังนี้

1. เพื่อทำการแบ่งกลุ่มผู้ที่มีพฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ โดยการคัดลอกหมายเลขโทรศัพท์เคลื่อนที่กับผู้ใช้บริการปกติ โดยพิจารณาจาก ระยะเวลาในการเป็นลูกค้านับตั้งแต่เปิดบริการ เงินเดือนของลูกค้า วงเงินในการใช้บริการ จำนวนครั้งที่ทำการเปลี่ยนที่อยู่ และจำนวนเงินเฉลี่ยที่ใช้บริการ
2. เพื่อทำการแบ่งกลุ่มผู้ที่มีพฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ โดยการการสมัครใช้บริการโดยตั้งใจที่จะทุจริต กับผู้ใช้บริการปกติ โดยพิจารณาจากจำนวนครั้งที่ชำระเงินตรงตามกำหนดเวลา และ จำนวนครั้งที่ชำระเงินเกินกำหนดเวลา

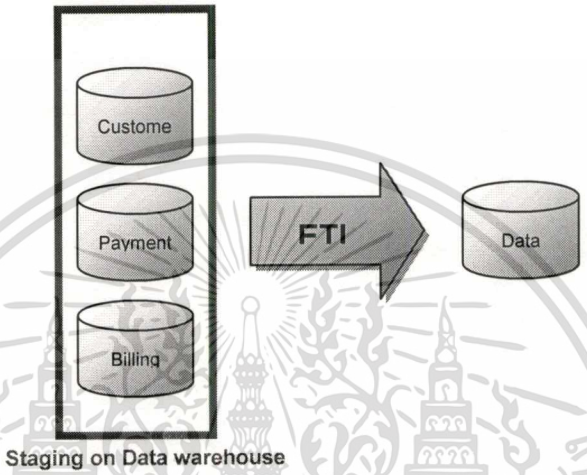
5.2. การเตรียมข้อมูล

ข้อมูลจะถูกเตรียมจาก ที่พักข้อมูลบนระบบคลังข้อมูล (Staging) และผ่านกระบวนการ ETL ในการแปลงข้อมูลให้อยู่ในรูปแบบ Text File เพื่อนำไปใช้ในการทำไมน์นิ่งต่อไป โดยข้อมูลจะถูกแบ่งออกเป็น 2 ชุด คือ ข้อมูลชุดฝึกอบรมและข้อมูลที่ต้องการจะนำไปวิเคราะห์หาผู้ที่ทำการทุจริต ซึ่งข้อมูลชุดฝึกอบรมจะถูกเตรียมมาจากข้อมูลของลูกค้าในอดีตที่เก็บไว้ในระบบคลังข้อมูล ส่วนข้อมูลที่ต้องการจะนำไปวิเคราะห์หาผู้ที่ทำการทุจริตจะถูกเลือกข้อมูลมาเก็บไว้ตามความต้องการของผู้ใช้งาน ซึ่งจากการรวบรวมความต้องการของผู้ใช้งานมา พบว่าข้อมูลที่ต้องการนำไปวิเคราะห์คือลูกค้าที่เปิดบริการมาไม่ต่ำกว่า 3 เดือนและมีการเปลี่ยนที่อยู่อย่างน้อย 1 ครั้ง โดยมีขั้นตอนการเตรียมข้อมูลดังนี้

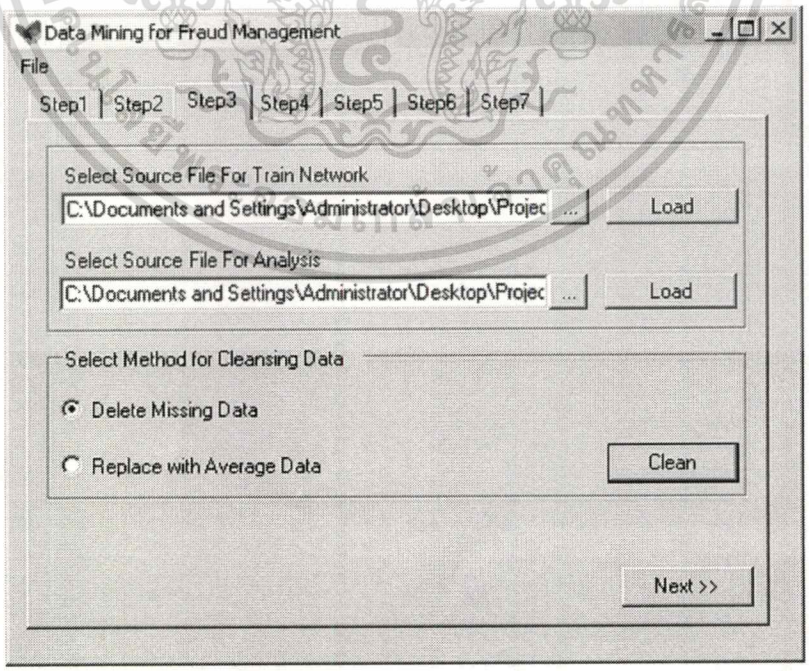
1. เตรียมข้อมูลจากระบบข้อมูลลูกค้า (Customer) โดยเลือกแอตทริบิวต์ของลูกค้าคือ อายุของลูกค้า, ระยะเวลาในการเป็นลูกค้านับตั้งแต่เปิดบริการ เงินเดือนของลูกค้า วงเงินในการใช้บริการ จำนวนครั้งที่ทำการเปลี่ยนที่อยู่ และสถานะของการทุจริตของลูกค้า
2. เตรียมข้อมูลจากระบบข้อมูลการชำระเงิน (Payment) โดยเลือกแอตทริบิวต์คือ จำนวนครั้งที่ชำระเงินตรงตามกำหนดเวลา และ จำนวนครั้งที่ชำระเงินเกินกำหนดเวลา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เตรียมข้อมูลจากระบบข้อมูลการใช้บริการ (Billing) โดยเลือกแอตทริบิวต์คือจำนวนเงินเฉลี่ยที่ใช้บริการต่อเดือน
4. เมื่อเตรียมข้อมูลเรียบร้อยแล้ว จะต้องทำความสะอาดข้อมูล โดยวิธีที่เลือกใช้ในการทำการทดลองครั้งนี้คือการตัดข้อมูลที่ขาดหายไปทิ้ง



รูปที่ 5.1 การเตรียมข้อมูลในการทำไมน์นิ่ง



รูปที่ 5.2 การทำความสะอาดข้อมูลโดยการตัดข้อมูลที่ขาดหายไปทิ้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3. การทดลองที่ 1

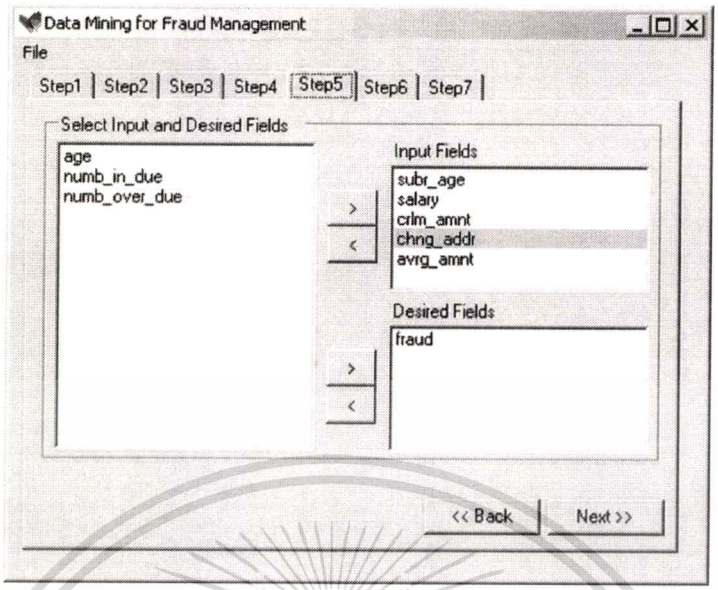
การทดลองที่ 1 เป็นการทำไมน์นิ่งเพื่อวิเคราะห์ข้อมูลตามวัตถุประสงค์แรกคือ เพื่อทำการแบ่งกลุ่มผู้ที่มีพฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ โดยการคัดลอกหมายเลขโทรศัพท์เคลื่อนที่กับผู้ใช้บริการปกติ โดยพิจารณาจาก ระยะเวลาในการเป็นลูกค้านับตั้งแต่เปิดบริการ เงินเดือนของลูกค้า วงเงินในการใช้บริการ จำนวนครั้งที่ทำการเปลี่ยนที่อยู่ และจำนวนเงินเฉลี่ยที่ใช้บริการ

5.3.1. การฝึกอบรมเครือข่ายประสาทเทียม

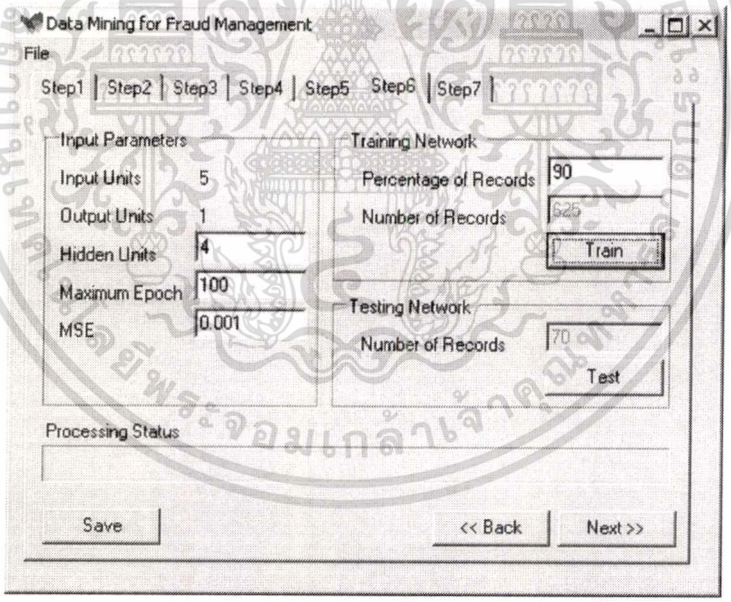
การฝึกอบรมเครือข่ายประสาทเทียม มีขั้นตอนดังนี้คือ

1. เลือกแอตทริบิวต์ของลูกค้าตามที่ตั้งวัตถุประสงค์ไว้คือ ระยะเวลาในการเป็นลูกค้านับตั้งแต่เปิดบริการ เงินเดือนของลูกค้า วงเงินในการใช้บริการ จำนวนครั้งที่ทำการเปลี่ยนที่อยู่ เข้ามาในระบบ ใส่ในช่องอินพุตฟิลด์
2. เลือกแอตทริบิวต์ที่ต้องการวิเคราะห์คือ สถานะของการทุจริตของลูกค้า ใส่ในช่องฟิลด์ที่ต้องการพยากรณ์
3. กำหนดค่าฮิดเดนยูนิตเป็น 4
4. กำหนดจำนวนรอบสูงสุดที่จะให้เครือข่ายเรียนรู้เป็น 100 รอบ
5. กำหนดค่าความคลาดเคลื่อนที่ยอมรับได้ เป็น 0.001
6. กำหนดสัดส่วนของข้อมูลสำหรับฝึกอบรมให้เป็น 90% ระบบจะกำหนดให้สัดส่วนของข้อมูลที่ต้องการจะทดสอบเป็น 10% โดยอัตโนมัติ
7. เมื่อกำหนดค่าเรียบร้อยแล้วจึงทำการฝึกอบรมเครือข่ายประสาทเทียม โดยค่าความคลาดเคลื่อนจากการฝึกอบรมครั้งนี้คือ 0.0759228

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

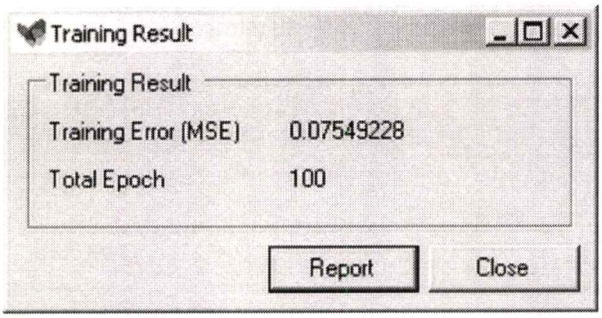


รูปที่ 5.3 การเลือกข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 1)

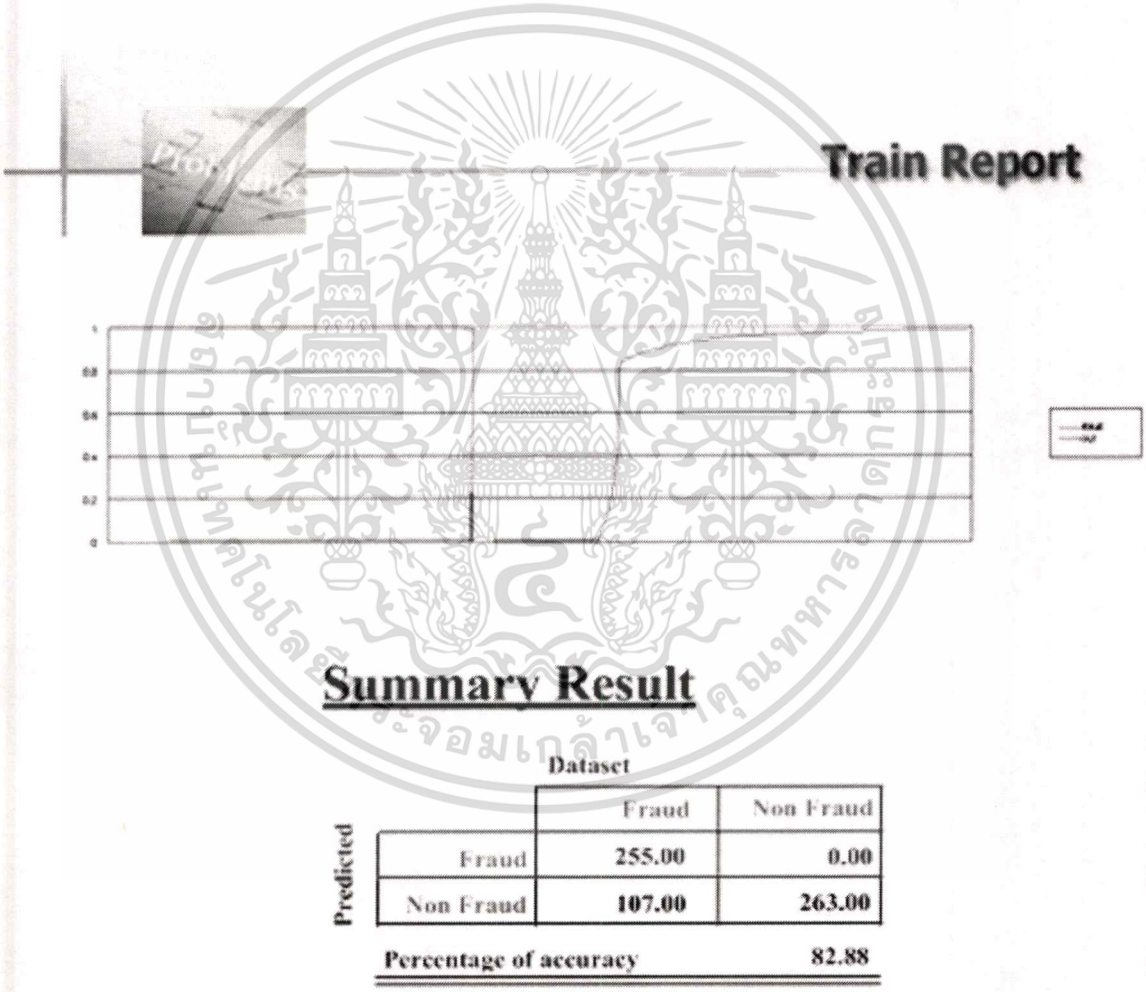


รูปที่ 5.4 การกำหนดค่าข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.5 ผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 1)



รูปที่ 5.6 รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 1)

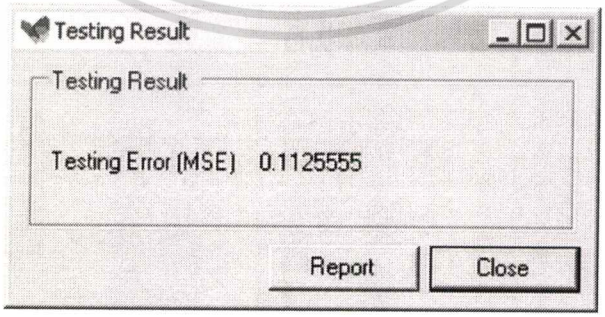
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียมจะมี 2 รูปแบบคือ

1. รูปแบบกราฟ โดยค่าที่เป็นผลลัพธ์ที่ถูกต้องจะแสดงในเส้นสีน้ำเงิน ส่วนค่าที่แบบจำลองพยากรณ์ได้จะแสดงในเส้นสีแดง
2. รูปแบบตารางสรุปผล จะแสดงค่าความถูกต้องของการพยากรณ์ โดยจะแสดงค่าที่แบบจำลองพยากรณ์ได้เทียบกับค่าผลลัพธ์ที่ถูกต้องจากข้อมูลชุดฝึกอบรม และสรุปค่าเปอร์เซ็นต์ความถูกต้องของแบบจำลอง ซึ่งจากการทดลองนี้สามารถสรุปผลได้ดังนี้
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 255 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มผู้ที่ทำการทุจริตจำนวน 107 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มใช้งานปกติจำนวน 0 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 263 คน
   - เปอร์เซ็นต์ความถูกต้องของการพยากรณ์ผลจากการทดสอบเครือข่ายประสาทเทียมนี้คือ 82.88%

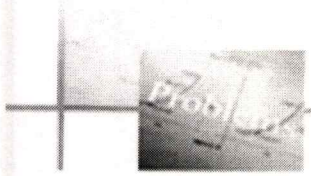
5.3.2. การทดสอบเครือข่ายประสาทเทียม

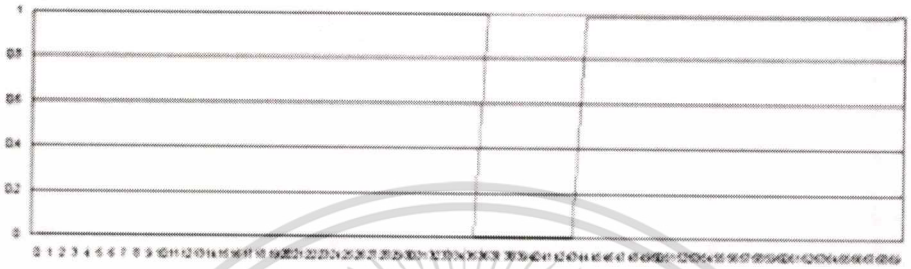
เมื่อทำการฝึกอบรมเครือข่ายประสาทเทียมเรียบร้อยแล้ว จึงทำการทดสอบเครือข่ายประสาทเทียม โดยการทดสอบนี้มีค่าความคลาดเคลื่อนเท่ากับ 0.1125555



รูปที่ 5.7 ผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





Summary Result

Predicted \ Dataset	Fraud	Non Fraud
Fraud	26.00	0.00
Non Fraud	8.00	36.00
Percentage of accuracy		88.57

รูปที่ 5.8 รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 1)

รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียมจะมี 2 รูปแบบคือ

1. รูปแบบกราฟ โดยค่าที่เป็นผลลัพธ์ที่ถูกต้องจะแสดงในเส้นสีน้ำเงิน ส่วนค่าที่แบบจำลองพยากรณ์ได้จะแสดงในเส้นสีแดง
2. รูปแบบตารางสรุปผล จะแสดงค่าความถูกต้องของการพยากรณ์ โดยจะแสดงค่าที่แบบจำลองพยากรณ์ได้เทียบกับค่าผลลัพธ์ที่ถูกต้องจากข้อมูลชุดฝึกอบรม และสรุปค่าอัตราส่วนความถูกต้องของแบบจำลอง ซึ่งจากการทดลองนี้สามารถสรุปผลได้ดังนี้
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 26 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มผู้ที่ทำการทุจริตจำนวน 8 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มใช้งานปกติจำนวน 0 คน
- ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 36 คน
- เปอร์เซ็นต์ความถูกต้องของการพยากรณ์ผลจากการทดสอบเครือข่ายประสาทเทียมนี้คือ 88.57%

5.3.3. การวิเคราะห์ข้อมูลจากโปรแกรม

หลังจากที่ทำการฝึกอบรมเครือข่ายจนเป็นที่พอใจแล้ว จึงนำข้อมูลของลูกค้าที่ต้องการนำมาแบ่งกลุ่ม มารันผ่านเครือข่าย สามารถแบ่งกลุ่มลูกค้าแต่ละคนตัวอย่างของผลการรันเป็นดังนี้

ตารางที่ 5.1 ผลการวิเคราะห์ข้อมูลจากโปรแกรม (การทดลองที่ 1)

Number	Subscriber No.	Customer Age	Salary	Credit Limit	Change Address	Average Amount	Fraud
1	0000000001	1,093.00	1	3,000.00	0	4,188.38	0.00791
2	0000000002	1,207.00	1	1,500.00	0	2,744.11	0.00791
3	0000000003	1,233.00	1	4,000.00	2	7,403.13	0.00791
4	0000000004	1,144.00	2	1,500.00	1	1,369.92	0.00791
5	0000000005	1,178.00	1	4,000.00	0	4,611.27	0.00791
6	0000000006	1,051.00	1	3,000.00	0	2,093.44	0.00791
7	0000000316	504	1	2,000.00	0	0	0.98979
8	0000000317	934	1	4,000.00	3	0	0.98979
9	0000000318	840	3	2,000.00	1	253.65	0.98979
10	0000000319	933	2	2,000.00	2	0	0.98979
11	0000000320	1,055.00	2	4,000.00	4	0	0.98979

สมมติฐานของการแบ่งกลุ่มในการทดลองครั้งนี้คือถือความน่าจะเป็นที่จะมีพฤติกรรมการทุจริตน้อยกว่า 0.5 ถือว่าเป็นลูกค้าปกติ แต่ถ้ามีค่ามากกว่า 0.5 ถือว่าลูกค้าที่มีพฤติกรรมการทุจริต สามารถอ่านค่าของข้อมูลได้ดังนี้

- พิจารณาข้อมูลลูกค้าหมายเลขโทรศัพท์ 0000000001 ที่มีค่าอายุของลูกค้าเป็น 1093 เงินเดือนเป็น 1 จำนวนวงเงินเครดิตเป็น 3000 จำนวนการเปลี่ยนที่อยู่เป็น 0 และจำนวนเงินเฉลี่ยที่ใช้บริการเป็น 4188.38 มีความเป็นไปได้ที่จะมีพฤติกรรมการทุจริตเป็น 0.00791 ซึ่งถือว่าอยู่ในกลุ่มลูกค้าปกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- พิจารณาข้อมูลลูกค้าหมายเลขโทรศัพท์ 0000000319 ที่มีค่าอายุของลูกค้าเป็น 993 เงินเดือนเป็น 2 จำนวนวงเงินเครดิตเป็น 2000 จำนวนการเปลี่ยนที่อยู่เป็น 2 และจำนวนเงินเฉลี่ยที่ใช้บริการเป็น 0 มีความเป็นไปได้ที่จะมีพฤติกรรมการทุจริตเป็น 0.98979 ซึ่งถือว่าอยู่ในกลุ่มลูกค้าที่มีพฤติกรรมการทุจริต

5.3.4. สรุปผลการทดลอง

จากการทดลองสามารถสรุปได้ดังนี้ จากข้อมูลที่ต้องการมาวิเคราะห์ทั้งหมด 500 คน สามารถแบ่งกลุ่มออกเป็น ผู้ที่มีพฤติกรรมการทุจริตในการใช้โทรศัพท์เคลื่อนที่ 290 คน ผู้ใช้บริการปกติ 210 คนโดยมีค่าความคลาดเคลื่อน 0.112555

5.4. การทดลองที่ 2

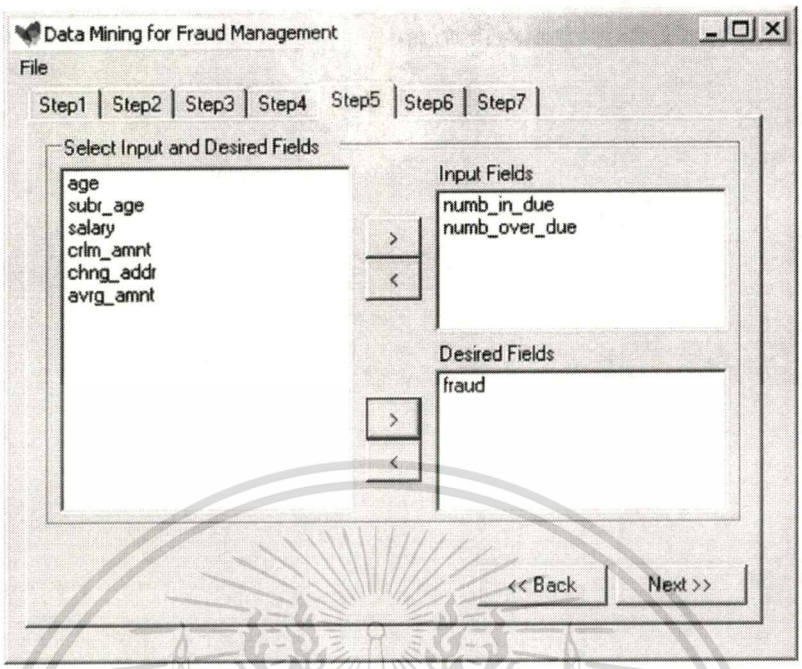
การทดลองที่ 2 เป็นการทำไมน์นิ่งตามวัตถุประสงค์ที่ 2 คือ เพื่อทำการแบ่งกลุ่มผู้ที่มีพฤติกรรมการทุจริตในการใช้บริการโทรศัพท์เคลื่อนที่ โดยการการสมัครใช้บริการโดยตั้งใจที่จะทุจริต กับผู้ใช้บริการปกติ โดยพิจารณาจากจำนวนครั้งที่ชำระเงินตรงตามกำหนดเวลา และจำนวนครั้งที่ชำระเงินเกินกำหนดเวลา

5.4.1. การฝึกอบรมเครือข่ายประสาทเทียม

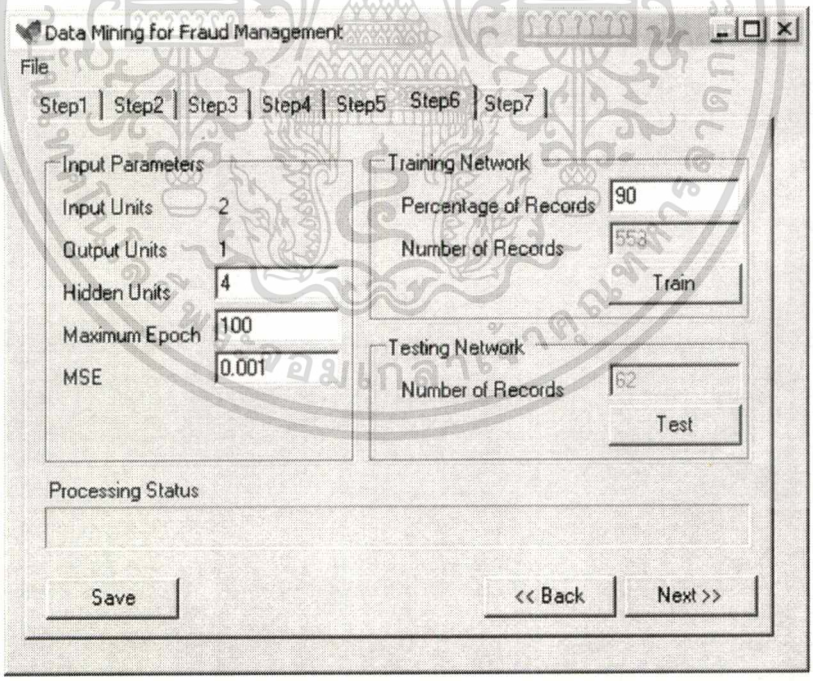
การฝึกอบรมเครือข่ายประสาทเทียม มีขั้นตอนดังนี้คือ

1. เลือกแอตทริบิวต์ของลูกค้าตามที่ตั้งวัตถุประสงค์ไว้คือ จำนวนครั้งที่ชำระเงินตรงตามกำหนดเวลา และ จำนวนครั้งที่ชำระเงินเกินกำหนดเวลา
2. เลือกแอตทริบิวต์ที่ต้องการวิเคราะห์คือ สถานะของการทุจริตของลูกค้า ใส่ในช่องฟิลด์ที่ต้องการพยากรณ์
3. กำหนดค่าฮิดเดนยูนิตเป็น 4
4. กำหนดจำนวนรอบสูงสุดที่จะให้เครือข่ายเรียนรู้เป็น 100 รอบ
5. กำหนดค่าความคลาดเคลื่อนที่ยอมรับได้ เป็น 0.001
6. กำหนดสัดส่วนของข้อมูลสำหรับฝึกอบรมให้เป็น 90% ระบบจะกำหนดให้สัดส่วนของข้อมูลที่ต้องการจะทดสอบเป็น 10% โดยอัตโนมัติ
7. เมื่อกำหนดค่าเรียบร้อยแล้วจึงทำการฝึกอบรมเครือข่ายประสาทเทียม โดยค่าความคลาดเคลื่อนจากการฝึกอบรมครั้งนี้คือ 0.0759228

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

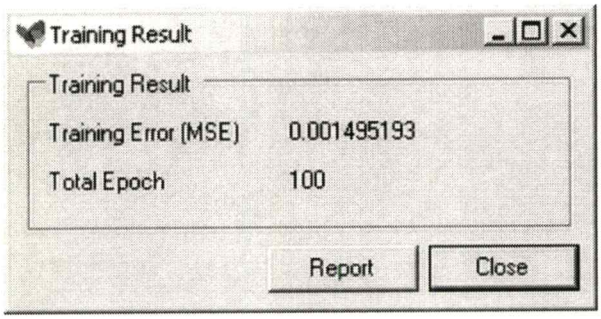


รูปที่ 5.9 การเลือกข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 2)

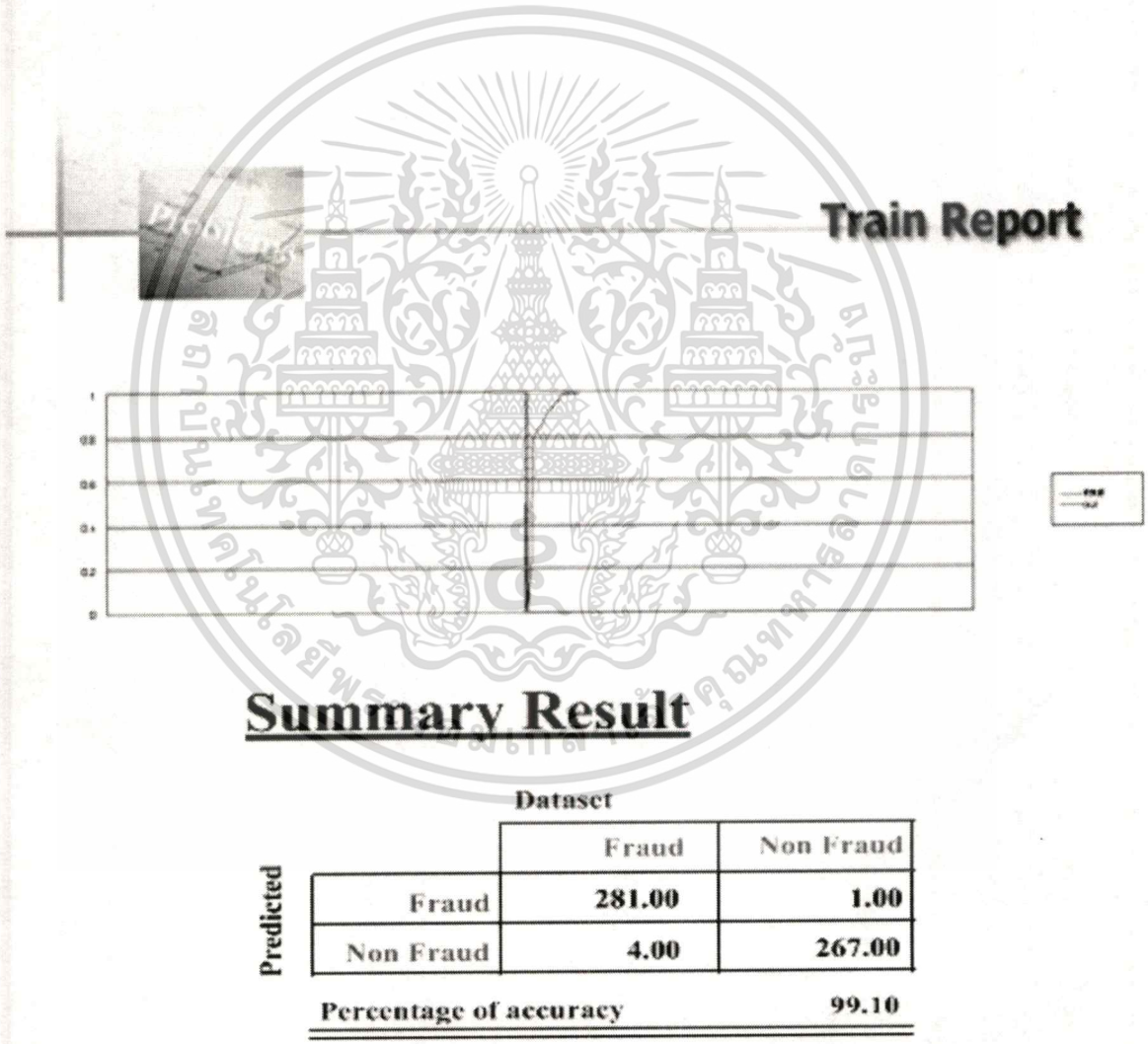


รูปที่ 5.10 การกำหนดค่าข้อมูลในการสร้างเครือข่ายประสาทเทียม (การทดลองที่ 2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.11 ผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 2)



รูปที่ 5.12 รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียม (การทดลองที่ 2)

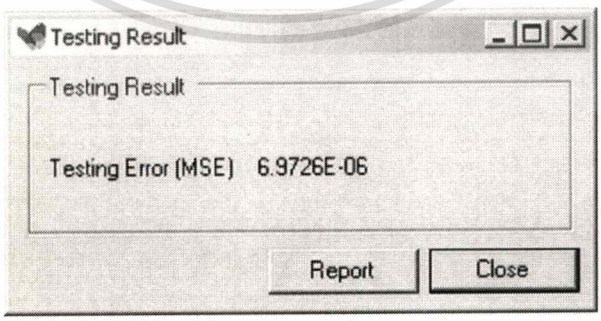
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายงานสรุปผลการฝึกอบรมเครือข่ายประสาทเทียมจะมี 2 รูปแบบคือ

1. รูปแบบกราฟ โดยค่าที่เป็นผลลัพธ์ที่ถูกต้องจะแสดงในเส้นสีน้ำเงิน ส่วนค่าที่แบบจำลองพยากรณ์ได้จะแสดงในเส้นสีแดง
2. รูปแบบตารางสรุปผล จะแสดงค่าความถูกต้องของการพยากรณ์ โดยจะแสดงค่าที่แบบจำลองพยากรณ์ได้เทียบกับค่าผลลัพธ์ที่ถูกต้องจากข้อมูลชุดฝึกอบรม และสรุปค่าเปอร์เซ็นต์ความถูกต้องของแบบจำลอง ซึ่งจากการทดลองนี้สามารถสรุปผลได้ดังนี้
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 281 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มผู้ที่ทำการทุจริตจำนวน 4 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มใช้งานปกติจำนวน 1 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 267 คน
   - เปอร์เซ็นต์ความถูกต้องของการพยากรณ์ผลจากการทดสอบเครือข่ายประสาทเทียมนี้คือ 99.10%

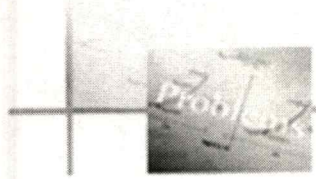
5.4.2. การทดสอบเครือข่ายประสาทเทียม

เมื่อทำการฝึกอบรมเครือข่ายประสาทเทียมเรียบร้อยแล้ว จึงทำการทดสอบเครือข่ายประสาทเทียม โดยการทดสอบนี้มีค่าความคลาดเคลื่อนเท่ากับ 0.0000069

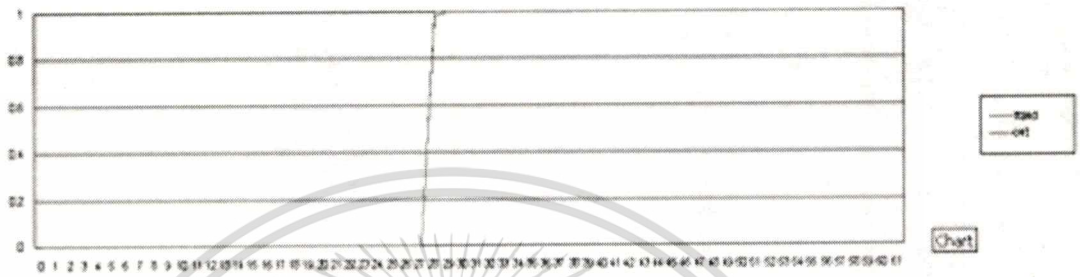


รูปที่ 5.13 ผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Test Report



Summary Result

Dataset

Predicted	Fraud	Non Fraud
Fraud	34.00	0.00
Non Fraud	0.00	28.00
Percentage of accuracy		100.00

รูปที่ 5.14 รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียม (การทดลองที่ 2)

รายงานสรุปผลการทดสอบเครือข่ายประสาทเทียมจะมี 2 รูปแบบคือ

1. รูปแบบกราฟ โดยค่าที่เป็นผลลัพธ์ที่ถูกต้องจะแสดงในเส้นสีน้ำเงิน ส่วนค่าที่แบบจำลองพยากรณ์ได้จะแสดงในเส้นสีแดง
2. รูปแบบตารางสรุปผล จะแสดงค่าความถูกต้องของการพยากรณ์ โดยจะแสดงค่าที่แบบจำลองพยากรณ์ได้เทียบกับค่าผลลัพธ์ที่ถูกต้องจากข้อมูลชุดฝึกอบรม และสรุปค่าอัตราส่วนความถูกต้องของแบบจำลอง ซึ่งจากการทดลองนี้สามารถสรุปผลได้ดังนี้
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 34 คน
   - ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มผู้ที่ทำการทุจริตจำนวน 0 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ที่ทำการทุจริต แต่ผลลัพธ์จากข้อมูลชุดฝึกอบรมจัดให้อยู่ในกลุ่มใช้งานปกติจำนวน 0 คน
- ค่าที่แบบจำลองพยากรณ์ให้อยู่ในกลุ่มผู้ใช้งานปกติ ซึ่งตรงกับผลลัพธ์จากข้อมูลชุดฝึกอบรมจำนวน 28 คน
- เปอร์เซ็นต์ความถูกต้องของการพยากรณ์ผลจากการทดสอบเครือข่ายประสาทเทียมนี้คือ 100%

5.4.3. การวิเคราะห์ข้อมูลจากโปรแกรม

หลังจากที่ทำการฝึกอบรมเครือข่ายจนเป็นที่พอใจแล้ว จึงนำข้อมูลของลูกค้าที่ต้องการนำมาแบ่งกลุ่ม มารันผ่านเครือข่าย สามารถแบ่งกลุ่มลูกค้าแต่ละคนตัวอย่างของผลการรันเป็นดังนี้

ตารางที่ 5.2 ผลการวิเคราะห์ข้อมูลจากโปรแกรม (การทดลองที่ 2)

Number	Subscriber No.	Number of In Due	Number of Over due	Fraud
1	1000000001	30	5	0.00001
2	1000000002	34	6	0.00001
3	1000000003	35	6	0.00001
4	1000000004	35	6	0.00001
5	1000000005	17	3	0.00001
6	1000000006	30	5	0.00001
7	1000000356	21	11	0.99969
8	1000000357	19	10	0.99969
9	1000000358	5	3	0.9997
10	1000000359	5	3	0.9997
11	1000000370	14	8	0.9998

สมมติฐานของการแบ่งกลุ่มในการทดลองครั้งนี้คือถือความน่าจะเป็นที่จะมีพฤติกรรมการทุจริตน้อยกว่า 0.5 ถือว่าเป็นลูกค้าปกติ แต่ถ้ามีค่ามากกว่า 0.5 ถือว่าลูกค้าที่มีพฤติกรรมการทุจริต สามารถอ่านค่าของข้อมูลได้ดังนี้

- พิจารณาข้อมูลลูกค้าหมายเลขโทรศัพท์ 1000000001 ที่มีจำนวนครั้งที่ชำระเงินตรงตามกำหนดเวลาเป็น 30 และ จำนวนครั้งที่ชำระเงินเกินกำหนดเวลาเป็น 5 มีความเป็นไปได้ที่จะมีพฤติกรรมการทุจริตเป็น 0.00001 ซึ่งถือว่าอยู่ในกลุ่มลูกค้าใช้บริการปกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- พิจารณาข้อมูลลูกค้าหมายเลขโทรศัพท์ 1000000359 ที่มีจำนวนครั้งที่ชำระเงินตรงตามกำหนดเวลาเป็น 5 และ จำนวนครั้งที่ชำระเงินเกินกำหนดเวลาเป็น 3 มีความเป็นไปได้ที่จะมีพฤติกรรมการทุจริตเป็น 0.9997 ซึ่งถือว่าอยู่ในกลุ่มลูกค้าที่มีพฤติกรรมทุจริต

5.4.4. สรุปผลการทดลอง

จากการทดลองสามารถสรุปได้ดังนี้ จากข้อมูลที่ต้องการมาวิเคราะห์ทั้งหมด 500 คน สามารถแบ่งกลุ่มออกเป็น ผู้ที่มีพฤติกรรมการทุจริตในการใช้โทรศัพท์เคลื่อนที่ 206 คน ผู้ใช้บริการปกติ 294 คนโดยมีค่าความคลาดเคลื่อน 0.0000069

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
บทสรุป

ในบทนี้จะกล่าวสรุปถึงผลการศึกษาและทำการประเมินระบบที่ได้ทำการพัฒนาทั้งในข้อดีและข้อเสียต่างๆ รวมถึงข้อเสนอแนะเพื่อเป็นแนวทางในการพัฒนาระบบ ให้ทำงานได้อย่างมีประสิทธิภาพมากขึ้น

6.1 สรุปผลการศึกษา

ระบบวิเคราะห์พฤติกรรมการทุจริตของผู้ใช้บริการโทรศัพท์เคลื่อนที่นั้น สร้างขึ้นมาเพื่อศึกษาหลักการทำงานของดาต้าไมน์นิ่งเพื่อนำไปประยุกต์ใช้กับธุรกิจ โดยในระบบงานนี้ได้ใช้เทคนิคของการสร้างแบบจำลองแบบพยากรณ์ (Predictive Modeling) และใช้วิธีการของเครือข่ายประสาทเทียมในการสร้างแบบจำลองเพื่อใช้ในการวิเคราะห์พฤติกรรมการทุจริตของผู้ใช้บริการโทรศัพท์เคลื่อนที่

ระบบงานที่พัฒนาขึ้นนี้ได้เลือกใช้ เครือข่ายประสาทเทียมที่มีโครงสร้างแบบ Multilayer Feed-Forward และมีอัลกอริทึมในการเรียนรู้แบบ Back Propagation ซึ่งจากการทดสอบระบบพบว่า ผลลัพธ์ที่ได้จากการพยากรณ์ขึ้นอยู่กับตัวแปรที่กำหนดให้ในการสร้างแบบจำลองระบบ เช่น จำนวนอินพุตโหนด และจำนวนฮิดเดนโหนด นอกจากนั้น ยังขึ้นอยู่กับจำนวนข้อมูลที่นำมาฝึกอบรมและทดสอบระบบด้วย ดังนั้น การกำหนดตัวแปรที่เหมาะสมจะทำให้ผลลัพธ์ที่ได้มีความคลาดเคลื่อนน้อยลง มีความแม่นยำมากขึ้น การกำหนดตัวแปรให้เหมาะสมนั้นไม่สามารถคำนวณเองได้ แต่จะเกิดจากการประสบการณ์ในการทดสอบหลายๆครั้ง จนได้ค่าที่เหมาะสม

จากการทดลองในบทที่5 พบว่า ผลลัพธ์ของระบบสามารถใช้หาความสัมพันธ์ของข้อมูลต่างๆของผู้ใช้บริการ เพื่อนำข้อมูลมาใช้ในการป้องกัน และตรวจจับลูกค้าที่มีแนวโน้มที่จะหลีกเลี่ยงการชำระค่าบริการได้ ซึ่งสามารถนำข้อมูลเหล่านี้ไปใช้เป็นประโยชน์ต่อองค์กรในการดำเนินการให้บรรลุเป้าหมายได้อย่างมีประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.2 ข้อจำกัดและข้อเสนอแนะ

โปรแกรมนี้ยังมีข้อจำกัดอยู่หลายประการที่ควรจะต้องปรับปรุงแก้ไข เพื่อให้โปรแกรมมีความยืดหยุ่นเหมาะสมกับวัตถุประสงค์ขององค์กรมากที่สุด สิ่งที่ควรจะแก้ไขสำหรับผู้ที่จะพัฒนาต่อไปมีรายละเอียดดังนี้

1. ควรเพิ่มอัลกอริทึมในการวิเคราะห์ให้มากขึ้น เนื่องจากแต่ละอัลกอริทึมจะมีข้อดีข้อเสียที่แตกต่างกัน ดังนั้นควรจะให้ผู้ใช้งานสามารถเลือกอัลกอริทึมที่เหมาะสมตามความต้องการของผู้ใช้งานที่สุด และนอกจากนี้ ในงานแต่ละอย่างนั้นไม่ได้ระบุว่าจะต้องใช้เทคนิคใดเทคนิคหนึ่งโดยเฉพาะเท่านั้น แต่ขึ้นอยู่กับลักษณะของปัญหาที่นำมาวิเคราะห์ ซึ่งในบางครั้งการนำเทคนิคหลายๆอย่างมาใช้เพื่อเปรียบเทียบความเหมาะสมกับแต่ละงาน หรือการนำหลายๆวิธีการมาใช้ร่วมกัน อาจจะมีส่วนช่วยในการค้นหาความหมาย หรือความสัมพันธ์จากข้อมูลออกมาได้อย่างมีประสิทธิภาพมากกว่าการใช้เพียงวิธีการเดียว
2. ระบบยังใช้เวลาค่อนข้างมากในการทำไมน์นิ่งข้อมูลขนาดใหญ่ ซึ่งในการใช้งานจริงข้อมูลอาจจะมีปริมาณมาก ทำให้ไม่สะดวกในการใช้งาน จึงควรมีการปรับแต่งให้ใช้เวลาให้ลดลง
3. ควรจะแสดงรูปภาพและคำอธิบายประกอบกับการทำงานของเครือข่ายประสาทเทียมสำหรับผู้ที่ต้องการเริ่มศึกษาระบบ หรือผู้ที่ต้องการนำระบบไปพัฒนาต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

สุวิมล คงศักดิ์ตระกูล. 2544. **สัมมนาเรื่อง Data Mining.** [Online]. Available: http://project.cs.kku.ac.th/2544/seminar/day2/413356-4and413357-5/apply.html

Cabena, P. et al. 1998. **Discovering Data Mining from Concept to Implementation**. New Jersey: Prentice Hall.

Chorleywood Consulting. 2002. **Application in Data Warehousing**. London: Chorleywood.

Han , J. and Kamber, M. 2001. **Data Mining Concepts and Techniques.** San Francisco: Morgan Kaufmann.

Taniguchi, M. et al. 1998. **Fraud Detection in Communications Networks Using Neural and Probablistic Methods.** [Online]. Available: http://www.cis.hut.fi/jhollmen/Publications/icassp98.pdf

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน	นางสาว นภาพร พันธ์โยธาชาติ
วันเดือนปีเกิด	14 ธันวาคม พ.ศ. 2522
สถานที่เกิด	จ. กรุงเทพ
วุฒิการศึกษาระดับปริญญาตรี	วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์
สถานที่สำเร็จการศึกษา	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีที่สำเร็จการศึกษา	ปีการศึกษา 2542
ประวัติการทำงาน	2542 – ปัจจุบัน เจ้าหน้าที่วิเคราะห์ระบบ บริษัท โทเทิ่ล แอ็คเซ็ส คอมมูนิเคชั่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้