

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

การประยุกต์ใช้ดาต้าไมนิ่งในการตรวจสอบการหลีกเลี่ยง  
ภาษีเงินได้นิติบุคคล

Data Mining Application in Corporate Income Tax Avoidance Auditing



วัน เดือน ปี.....	09 ส.ค. 2550
เลขทะเบียน.....	01725
เลขเรียกหนังสือ.....	จท. ข 677ก 2543
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."	

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน  
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ  
ภาคเรียนที่ 1 ปีการศึกษา 2543  
คณะเทคโนโลยีสารสนเทศ

เอกสารนี้เป็นเอกสารที่สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**Title** Data Mining Application in Corporate Income Tax Avoidance Auditing  
**Student** Miss.Busaya Promphibal  
**Advisor** Assoc.Prof.Dr.Wichian Premchaiswadi  
**Level of Study** Master of Science in Information Technology  
**Major** Information Science  
**Academic Year** 2000



## ABSTRACT

The main functions of the Revenue Department are to collect tax and to generate revenue for the whole nation. Most revenue are generate from Corporate Income Tax, in which the Revenue Department has its own database on this kind of taxation. In this regard, there is another utilization an these database. The Revenue Department then employs the data mining technique on these collected data, especially in the audit on tax evasion.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

เอกสารฉบับนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่สำเร็จลงได้ด้วยความกรุณาของ รศ.ดร.วิเชียร เปรมชัยสวัสดิ์ ที่ได้กรุณาเสียสละเวลาอันมีค่ารับเป็นอาจารย์ที่ปรึกษาของผู้เขียน และได้กรุณาให้คำปรึกษาแนะนำอันเป็นประโยชน์มาโดยตลอด รวมทั้งขอระลึกถึงผู้บังคับบัญชา เพื่อนร่วมงานที่สละเวลาให้ข้อมูลและความช่วยเหลือแก่ผู้เขียนเพื่อใช้ในการจัดทำเอกสารฉบับนี้



บุษยา พรหมภินาล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญภาพ.....	VII
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมา.....	1
1.2 Data Process ภาษีเงินได้นิติบุคคล.....	1
1.3 ขั้นตอนในการตรวจสอบภาษีเงินได้นิติบุคคลในปัจจุบัน.....	2
1.4 ปัญหาที่พบในการดำเนินการตรวจสอบภาษีเงินได้นิติบุคคล.....	2
1.5 วัตถุประสงค์ของการพัฒนาระบบงาน.....	3
1.6 หลักการที่เกี่ยวข้องในการพัฒนาระบบงาน.....	3
1.7 องค์ประกอบของการพัฒนาระบบงาน.....	3
2. Data Mining และ Neural Network.....	4
2.1 คำจำกัดความของ Data Mining.....	4
2.2 กระบวนการของ Data Mining.....	4
2.3 โอเปอเรชั่นของ Data Mining.....	5
2.4 โครงข่ายประสาทเทียม(Neural Network).....	12
2.5 การเรียนรู้ของโครงข่ายประสาทเทียม.....	15
2.6 แบบจำลองของ ANN.....	16
3. แนวคิดและการออกแบบ.....	19
3.1 ฐานข้อมูลที่ใช้ในการพัฒนาระบบ.....	19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ

หน้า

3.2 การออกแบบ User Interface .....	19
4. การประยุกต์ใช้คำตำไมนิ่งในการตรวจสอบการหลีกเลี่ยงภาษีเงินได้นิติบุคคล.....	22
4.1 การคัดเลือกข้อมูลสำหรับใช้ในการตรวจสอบ .....	22
4.2 การ Clean ข้อมูล .....	27
4.3 การ Transform ข้อมูล.....	28
4.4 การคัดเลือกอัลกอริทึมในการตรวจสอบการหลีกเลี่ยงภาษีเงินได้นิติบุคคล .....	29
4.5 การวิเคราะห์ผลลัพธ์.....	34
5. บทสรุปและข้อเสนอแนะ.....	43
5.1 บทสรุป.....	43
5.2 ข้อเสนอแนะ.....	43
บรรณานุกรม.....	44
ประวัติผู้เขียน.....	45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

	หน้า
ตารางที่	
2.1 Transaction Database .....	10
2.2 Customer Sequences .....	10
2.3 Sequential patterns with support >4% .....	11
4.1 ฐานข้อมูลที่ใช้ในการตรวจสอบ.....	23



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญภาพ

ภาพที่	หน้า
2.1 กระบวนการของ Data Mining .....	5
2.2 ตัวอย่างของการ Classification โดยใช้ Tree Induction.....	6
2.3 ตัวอย่างของการ Classification โดยใช้ Neural Induction .....	7
2.4 การนำเทคนิค Visualization ไปใช้ .....	11
2.5 Biological Neurons.....	12
2.6 แบบจำลองเซลล์ประสาทของ McColloch-Pitts.....	13
2.7 แบบจำลองเซลล์ประสาทเทียม .....	14
2.8 บล็อกไดอะแกรมของโครงข่าย Feedforward .....	16
2.9 บล็อกไดอะแกรมของโครงข่าย Feedback.....	17
2.10 โครงข่าย Multilayer Perceptron ที่มี 3 ชั้น .....	17
3.1 หน้าจอเมนูหลัก .....	19
3.2 หน้าจอการคัดเลือกข้อมูล .....	20
3.3 หน้าจอ Transform ข้อมูล .....	20
3.4 หน้าจอระบบรักษาความปลอดภัย .....	21
4.1 เมนูหลักของการทำงาน .....	26
4.2 การคัดเลือก Attribute ที่ใช้ในการตรวจสอบการหลักเสียงภาษา .....	27
4.3 หน้าจอการ Transfer ข้อมูล .....	28
4.4 ตัวอย่าง Backpropagation network.....	30
4.5 แสดงการ Train ข้อมูลในรูปแบบกราฟโดยโปรแกรม MATLAB .....	35
4.6 แสดงการ Train ข้อมูลในรูปแบบข้อความโดยโปรแกรม MATLAB .....	35
4.7 แสดงข้อมูล Attribute รายได้จากการขายโดยตรงที่ใช้ในการ Train.....	36
4.8 แสดงข้อมูล Attribute ต้นทุนขายที่ใช้ในการ Train.....	36
4.9 แสดงข้อมูล Attribute กำไร/ขาดทุนขั้นต้นใช้ในการ Train.....	37
4.10 แสดงข้อมูล Attribute รายได้ที่ไม่เกี่ยวกับการขายโดยตรงที่ใช้ในการ Train.....	37

เอกสาร 4.11 แสดงข้อมูล Attribute รายจ่ายในการดำเนินงานที่ใช้ในการ Train ให้มองไปใต้ประโยชน์ (38) การค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญภาพ

ภาพที่	หน้า
4.12 แสดงข้อมูล Attribute กำไร/ขาดทุนสุทธิที่ใช้ในการ Train.....	38
4.13 แสดงข้อมูล Attribute ต้นทุนการผลิตที่ใช้ในการ Train.....	39
4.14 แสดงข้อมูล Attribute รวมสินทรัพย์ที่ใช้ในการ Train.....	39
4.15 แสดงข้อมูล Attribute รวมหนี้สินและทุนที่ใช้ในการ Train.....	40
4.16 แสดงข้อมูล Attribute ยอดรวมรายได้ที่ใช้ในการ Train.....	40
4.17 แสดงข้อมูล Attribute ยอดรวมรายจ่ายที่ใช้ในการ Train .....	41
4.18 แสดงข้อมูล Attribute กำไรขาดทุนสุทธิ(ภ.ง.ด.51)ที่ใช้ในการ Train .....	41
4.19 การแสดงผลลัพธ์ในรูปแบบกราฟโดยโปรแกรม MATLAB .....	42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมา

กรมสรรพากรเป็นหน่วยงานที่ทำหน้าที่ในการจัดเก็บภาษีอากร และเป็นหน่วยงานจัดเก็บรายได้หน่วยงานหนึ่งของประเทศ โดยรายได้ในการจัดเก็บภาษีอากรส่วนใหญ่มาจากภาษีเงินได้นิติบุคคล ซึ่งผู้ที่มีหน้าที่ชำระภาษีเงินได้นิติบุคคลได้แก่บริษัทหรือห้างหุ้นส่วนนิติบุคคล ซึ่งหากผู้ที่มีหน้าที่ชำระภาษีเงินได้นิติบุคคล มีความซื่อตรงในการยื่นแบบแสดงรายการภาษีเงินได้นิติบุคคล บริษัทหรือห้างหุ้นส่วนนิติบุคคล (ภ.ง.ด.50) ตามจริงก็จะไม่เกิดกรณีของการตรวจสอบเงินได้ของนิติบุคคลขึ้น แต่ในความเป็นจริง ผู้ที่มีหน้าที่ชำระภาษีเงินได้นิติบุคคลจำนวนมากพยายามที่จะหลีกเลี่ยงการชำระภาษีเงินได้นิติบุคคล ดังนั้นกรมสรรพากรจึงต้องทำการค้นหาผู้ที่มีหน้าที่ชำระภาษีเงินได้นิติบุคคลที่มีแนวโน้มว่าจะหลีกเลี่ยงภาษีจากข้อมูลของแบบแสดงรายการภาษีเงินได้ ซึ่งในปัจจุบันหลักเกณฑ์และเงื่อนไขที่ใช้อยู่ขาดความรัดกุมจึงทำให้ผู้ที่มีหน้าที่เสียภาษีเงินได้นิติบุคคลบางรายหลีกเลี่ยงจากการตรวจสอบได้ ดังนั้นเพื่อให้เกิดประสิทธิภาพในการตรวจสอบภาษีเงินได้นิติบุคคลจึงนำเทคนิคคาด้า ไม่นิ่งเข้ามาช่วยในการคัดเลือกผู้ที่มีหน้าที่ชำระภาษีเงินได้นิติบุคคลที่มีความเสี่ยงต่อการหลีกเลี่ยงภาษีได้รัดกุมมากยิ่งขึ้น

### 1.2 Data Process ของภาษีเงินได้นิติบุคคล

- 1.2.1 ผู้มีหน้าที่เสียภาษีเงินได้นิติบุคคลยื่นแบบแสดงรายการภาษีเงินได้บริษัทหรือห้างหุ้นส่วนนิติบุคคล(ภ.ง.ด.50) เพื่อชำระภาษีเงินได้ที่สรรพากรอำเภอหรือสรรพากรเขต
- 1.2.2 เจ้าหน้าที่สรรพากรอำเภอหรือสรรพากรเขตทำการรับแบบ ตรวจสอบความถูกต้องของแบบแสดงรายการ(ภ.ง.ด.50)จากนั้นก็ทำการบันทึกข้อมูลเข้าเครื่องออกใบเสร็จ(POS) และคืนใบเสร็จให้กับผู้ที่มีหน้าที่เสียภาษีเงินได้นิติบุคคล
- 1.2.3 เจ้าหน้าที่สรรพากรอำเภอหรือสรรพากรเขตจัดเพิ่มนำส่งแบบแสดงรายการ(ภ.ง.ด.50)ให้กับสำนักงานสรรพากรพื้นที่หรือสำนักงานสรรพากรภาคต้นสังกัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1.2.4 ฝ่ายบริหารของสำนักงานสรรพากรพื้นที่หรือสรรพากรภาค เมื่อได้รับแบบเบียร์ร้อยแล้วก็จัดส่งแบบแสดงรายการ(ภ.ง.ด.50)ดังกล่าวให้กับฝ่ายบริหารงานกรรมวิธีของสรรพากรพื้นที่เพื่อบันทึกแบบแสดงรายการ(ภ.ง.ด.50) ส่วนสรรพากรภาค ฝ่ายบริหารจะจัดส่งให้ฝ่ายประมวลผลของภาคดำเนินการบันทึกแบบแสดงรายการ(ภ.ง.ด.50) และเจ้าหน้าที่ทำการจัดส่งข้อมูลที่ได้บันทึกไว้ให้กับสำนักเทคโนโลยีสารสนเทศ (โดยเจ้าหน้าที่คัดลอกข้อมูลลงแผ่น Diskette แล้วทำการจัดส่งให้กับสำนักเทคโนโลยีสารสนเทศ)
- 1.2.5 เมื่อบันทึกแบบแสดงรายการ(ภ.ง.ด.50)เสร็จฝ่ายกรรมวิธีหรือฝ่ายประมวลผลก็จะจัดส่งแบบแสดงรายการ(ภ.ง.ด.50)ที่บันทึกแล้วคืนให้กับฝ่ายบริหารเพื่อดำเนินการจัดเก็บต่อไป

### 1.3 ขั้นตอนในการตรวจสอบภาษีเงินได้นิติบุคคลในปัจจุบัน

- 1.3.1 การตรวจสอบภาษีเงินได้นิติบุคคลที่ทำการประเมินตนเอง กระทำได้โดยการเปรียบเทียบยอดรายได้ในแบบแสดงรายการ(ภ.ง.ด.50) กับฐานรายได้ที่ผู้เสียภาษีได้แจ้งไว้
- 1.3.2 การตรวจสอบภาษีเงินได้นิติบุคคลที่ทำการตรวจสอบโดยเจ้าหน้าที่ประเมิน กระทำได้ดังนี้
- 1.3.2.1 เจ้าหน้าที่ทำการสุ่มตรวจจากแบบแสดงรายการที่ผู้เสียภาษีได้ยื่นไว้และออกไปสำรวจกิจการของผู้เสียภาษีโดยการสุ่มตรวจ
- 1.3.2.2 กรณีที่มีผู้แจ้งข้อมูลการหลีกเลี่ยงภาษีของผู้เสียภาษีเข้ามาเจ้าหน้าที่ตรวจสอบต้องขอข้อมูลการยื่นแบบแสดงรายการภาษีมาที่สำนักเทคโนโลยีสารสนเทศเพื่อใช้ในการตรวจสอบ
- 1.3.2.3 ทำการวิเคราะห์แบบแสดงรายการของผู้เสียภาษีโดยเปรียบเทียบตามหลักเกณฑ์ การวิเคราะห์จากสภาพธุรกิจ สภาพเศรษฐกิจ และภาพรวมของกิจการในประเภทเดียวกันเพื่อหาความแตกต่าง

### 1.4 ปัญหาในการดำเนินการตรวจสอบภาษีเงินได้นิติบุคคล

- 1.4.1 การจัดเก็บภาษีเงินได้นิติบุคคลโดยวิธีการประเมินตนเอง ผู้เสียภาษีจะเป็นผู้แสดงยอดรายได้ในแบบแสดงรายการภาษี และชำระภาษีตามผลการคำนวณจากฐาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายได้ที่ผู้เสียภาษีแจ้งไว้ ซึ่งยอดรายได้ที่ผู้เสียภาษีได้แจ้งไว้อาจจะแสดงไว้ต่ำกว่าความเป็นจริง

- 1.4.2 การจัดเก็บภาษีเงินได้นิติบุคคล โดยวิธีการตรวจสอบของเจ้าหน้าที่ประเมิน
  - 1.4.2.1 ผู้ที่มีหน้าที่เสียภาษีเงินได้นิติบุคคลมีจำนวนมากแต่จำนวนเจ้าหน้าที่ตรวจสอบมีไม่เพียงพอที่จะทำการตรวจสอบแบบแสดงรายการภาษีหรือออกไปสำรวจ
  - 1.4.2.2 ผู้เสียภาษีอาจจะมีการคิดสินบนเจ้าพนักงานเพื่อไม่ให้ทำการตรวจสอบกิจการของตนทำให้เกิดความไม่เป็นธรรมกับผู้เสียภาษีคนอื่น ๆ
  - 1.4.2.3 เกิดความล่าช้าในการตรวจสอบ

## 1.5 วัตถุประสงค์ของการพัฒนาระบบงาน

- 1.5.1 สามารถเพิ่มประสิทธิภาพในการคัดเลือกรายผู้ที่มีหน้าที่เสียภาษีเงินได้นิติบุคคลที่มีโอกาสหลีกเลี่ยงภาษี
- 1.5.2 สามารถเพิ่มความสะดวกและรวดเร็วในการตรวจสอบภาษีให้กับเจ้าพนักงานได้มากยิ่งขึ้น
- 1.5.3 สามารถนำข้อมูลสารสนเทศที่ได้จากการคัดเลือกรายไปใช้ในการวางแผนการตรวจสอบภาษีเงินได้นิติบุคคลของผู้บริหารได้อย่างมีประสิทธิภาพ
- 1.5.4 สามารถทำการตรวจแบบแสดงรายการภาษีเงินได้นิติบุคคลของผู้เสียภาษีได้ครบทุกรายและทำให้เกิดความเป็นธรรมแก่ผู้เสียภาษี

## 1.6 หลักการที่เกี่ยวข้องในการพัฒนาระบบงาน

- 1.6.1 เทคนิคของ Data Mining
- 1.6.2 หลักการของ Neural Network โดยใช้อัลกอริทึม Backpropagation

## 1.7 องค์ประกอบของการพัฒนาระบบงาน

- 1.7.1 คอมพิวเตอร์ที่ประกอบด้วย CPU Celeron 500 MHz , RAM ขนาด 64 MB
- 1.7.2 MATLAB version 5.3
- 1.7.3 Microsoft Visual FoxPro 6.0 และ Microsoft Visual Basic 6.0
- 1.7.4 Windows 98 หรือ Windows NT

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

### Data Mining และ Neural Network

#### 2.1 คำจำกัดความของ Data Mining

ได้มีผู้ให้คำจำกัดความของ Data Mining ไว้หลายความหมาย ยกตัวอย่าง เช่น Data Mining เป็นกระบวนการในการค้นกรองข้อมูลที่ไม่รู้มาก่อนจากฐานข้อมูลที่มีขนาดใหญ่ และข้อมูลที่กลั่นกรองออกมานั้นต้องถูกต้องและนำไปใช้ประโยชน์เพื่อประกอบการตัดสินใจ Data Mining เป็นการค้นพบสารสนเทศใหม่ ๆ จากข้อมูลขนาดใหญ่ที่ได้รวบรวมไว้

ดังนั้น Data Mining คือการค้นหาสารสนเทศที่เป็นประโยชน์จากข้อมูลจำนวนมากที่จัดเก็บไว้

#### 2.2 กระบวนการของ Data Mining

ประกอบด้วย 3 ขั้นตอน คือ

2.2.1 Data Preparation ประกอบด้วย 3 ขั้นตอนย่อย ๆ คือ

2.2.1.1 Selection Data เมื่อกำหนดวัตถุประสงค์เรียบร้อยแล้วก็ต้องทำการคัดเลือกว่า จะใช้ข้อมูลอะไรจากส่วนไหนเพื่อทำให้บรรลุตามวัตถุประสงค์ที่ได้กำหนดไว้

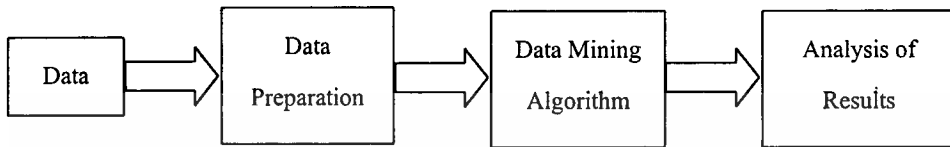
2.2.1.2 Preprocessed Data เมื่อทำการคัดเลือกข้อมูลได้แล้วก็ต้องทำการ Clean Data และตรวจสอบข้อมูลว่าถูกต้องหรือไม่ มีข้อมูลขาดหายไปหรือไม่ ขั้นตอนนี้มีวัตถุประสงค์เพื่อตรวจสอบความผิดพลาดของข้อมูล

2.2.1.3 Transform Data ขั้นตอนนี้จะพิจารณาถึงข้อมูลที่จะนำมาใช้กับอัลกอริทึมที่เราใช้ คือต้องแปลงข้อมูลให้เป็นประเภทเดียวกับอัลกอริทึมที่จะใช้ เช่น ถ้าใช้ Neural network ต้องใช้ข้อมูลที่เป็นตัวเลข

2.2.2 Data Mining Algorithm เป็นการเลือกใช้อัลกอริทึมที่เหมาะสม โดยการเลือกจะขึ้นอยู่กับลักษณะของปัญหา และลักษณะของข้อมูล แล้วนำมาเอาข้อมูลที่ Transform แล้วไปประมวลผลกับอัลกอริทึมนั้น เพื่อให้ได้สารสนเทศออกมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมีการดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.3 Analysis of Results เป็นการวิเคราะห์ผลลัพธ์ที่ได้จากการประมวลผล เพราะว่าผลลัพธ์ที่ได้อาจจะไม่สามารถนำไปใช้ตรง ๆ ได้ ซึ่งอาจจะเป็นการตีความหมายจากผลลัพธ์ที่อยู่ในรูปของตัวเลขก็ได้



รูปที่ 2.1 กระบวนการของ Data Mining

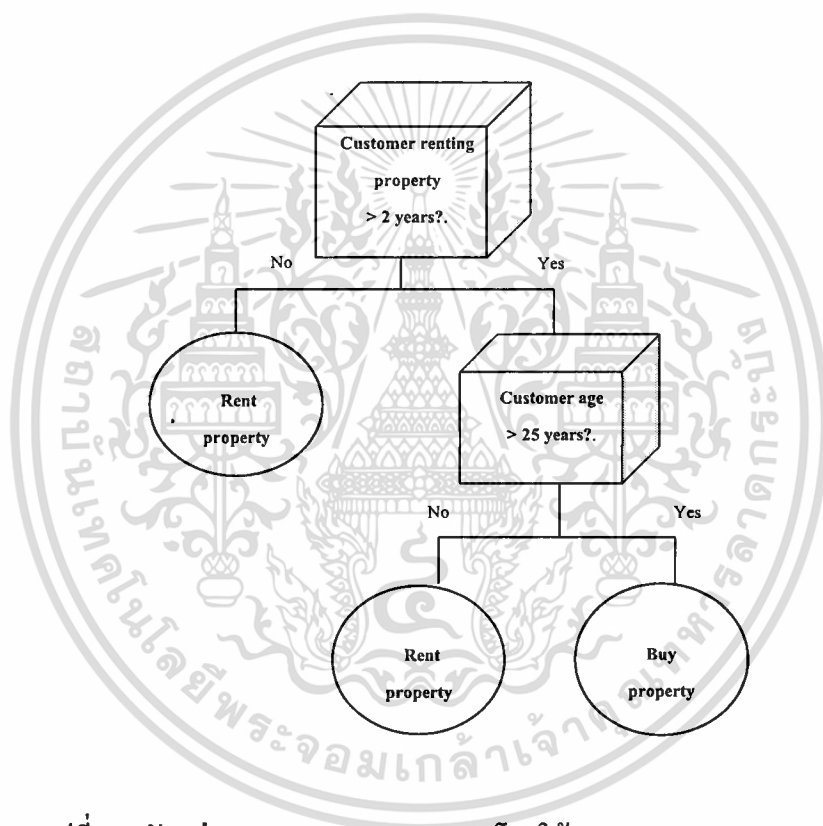
## 2.3 โอเปอเรชั่นของ Data Mining

2.3.1 Predictive modeling แบบจำลองการทำนายจะคล้ายๆ กับการเรียนรู้ประสิทธิภาพของมนุษย์โดยการใช้การสังเกตจากรูปแบบของคุณลักษณะที่สำคัญของเหตุการณ์ที่ประทับใจบางอย่าง แบบจำลองการทำนายนี้สามารถที่จะใช้ในการวิเคราะห์ภายในฐานข้อมูลเพื่อที่จะกำหนดคุณลักษณะที่สำคัญบางอย่างเกี่ยวกับกลุ่มข้อมูลได้ รูปแบบนี้ถูกพัฒนาโดยใช้วิธีการควบคุมการเรียนรู้ (Supervised learning) ซึ่งมี 2 เฟส คือการเรียนรู้ (Training) และการทดสอบ (Testing) การเรียนรู้สร้างโดยใช้ตัวอย่างของประวัติข้อมูลที่มีขนาดใหญ่มาก ซึ่งถูกเรียกว่า กลุ่มการเรียนรู้ (Training Set) ขณะที่การทดสอบ เป็นเรื่องเกี่ยวกับการทดสอบแบบจำลองกับข้อมูลใหม่ที่ไม่เคยเห็นมาก่อน เพื่อที่จะสรุปความถูกต้องและคุณลักษณะทางกายภาพของผลที่ได้ การประยุกต์ใช้แบบจำลองนี้เช่น การจัดการเรื่องของการรักษาลูกค้า การอนุมัติบัตรเครดิต การขายของ และเป้าหมายการตลาด แบบจำลองการทำนายประกอบด้วย 2 เทคนิคคือ การจัดหมวดหมู่ (Classification) และ การทำนายมูลค่า ( Value Predictive)

โดยการจัดหมวดหมู่ประกอบด้วย 2 เทคนิคคือ Tree Induction และ Neural Induction ตัวอย่างของการจำแนกโดยใช้ Tree Induction แสดงในรูปที่ 2.2 ซึ่งในตัวอย่างนี้เราสนใจในการทำนายว่าลูกค้าคนที่เข้าทรัพย์สินในปัจจุบันใครที่สนใจจะซื้อทรัพย์สินนั้นบ้าง แบบจำลองการทำนายนี้ถูกสรุปว่ามีตัวแปรเพียงแค่ 2 ตัวแปรที่น่าสนใจคือระยะเวลาที่ลูกค้าเข้าทรัพย์สิน และอายุของลูกค้า การตัดสินใจแบบต้นไม้เป็นการนำเสนอการวิเคราะห์โดยใช้สัญชาตญาณโดยรูปแบบการทำนายนี้ แสดงให้เห็นว่าลูกค้าที่มีการเข้าทรัพย์สินมากกว่า 2 ปี และอายุมากกว่า 25 ปี ส่วนมากแล้วจะสนใจที่จะซื้อทรัพย์สินนั้น ตัวอย่างของการ Classification โดยการใช้ Neural Induction แสดงให้เห็นในรูปที่ 2.3 โดยใช้ตัวอย่างเดียวกันกับ Tree Induction ในกรณีนี้ การ Classification ข้อมูลถูกทำให้เป็นผลสำเร็จโดยใช้ Neural Network ซึ่ง Neural Network เป็นการ

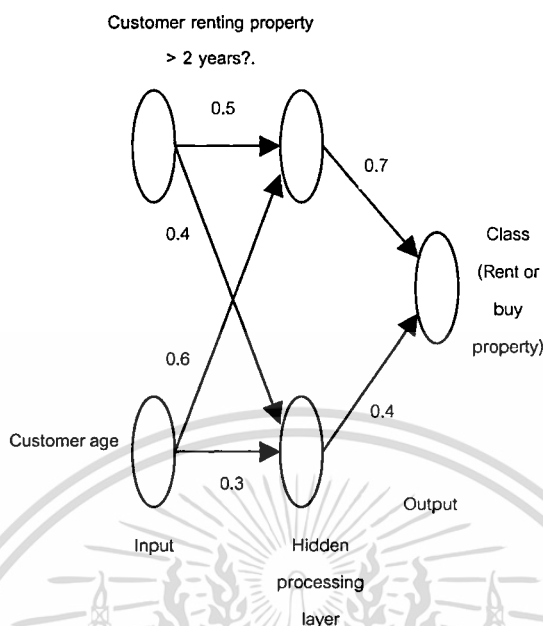
เอกลักรุ่นเป็นเอกสารลิขสิทธิ์สงวนลิขสิทธิ์การเขียนเพื่อการศึกษาเท่านั้น เมื่อผู้ดูแลเนื้อหาใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รวบรวมการเชื่อมต่อ node ของข้อมูลเข้า ข้อมูลออกและกระบวนการในแต่ละ node ระหว่างข้อมูลเข้าและข้อมูลออกที่แสดงให้เห็น อาจจะมีชั้นประมวลผลซ่อนอยู่ แต่ละหน่วยประมวลผลในหนึ่งชั้นถูกเชื่อมต่อไปที่แต่ละหน่วยประมวลผลในชั้นถัดไป โดยการถ่วงค่าน้ำหนัก ซึ่งเป็นการแสดงให้เห็นถึงความสัมพันธ์ที่ชัดเจน เครือข่ายพยายามที่จะเลียนแบบวิธีการทำงานของสมองมนุษย์ในการที่จะยอมรับรูปแบบโดยใช้หลักการทางคณิตศาสตร์ในการรวมตัวแปรที่มีความสัมพันธ์ทั้งหมดด้วยการให้ค่าจุดทศนิยมกับข้อมูล



รูปที่ 2.2 ตัวอย่างของการ Classification โดยใช้ Tree Induction

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



### รูปที่ 2.3 ตัวอย่างของการ Classification โดยใช้ Neural Induction

การทำนายมูลค่า (Value Prediction) การทำนายมูลค่าถูกใช้เพื่อประเมินค่าของตัวเลขที่ต่อเนื่อง ซึ่งมีความเกี่ยวข้องกับเรคคอร์ดในฐานข้อมูล เทคนิคนี้ใช้เป็นเทคนิคที่เกี่ยวกับสถิติของ Linear Regression และ Nonlinear Regression เทคนิคนี้ได้รับการยอมรับว่าเป็นเทคนิคที่ดี ก่อนข้างง่ายต่อการที่จะใช้งานและการทำความเข้าใจ Linear Regression เป็นการพอร์ตข้อมูลให้เป็นเส้นตรง ซึ่งเส้นที่เกิดจากการพอร์ตจะแสดงให้เห็นถึงค่าเฉลี่ยของการสังเกตทั้งหมดได้ดีที่สุด ปัญหาของ Linear Regression คือเทคนิคนี้ทำงานได้ดีกับ Linear Data เท่านั้น และไวต่อการปรากฏของสิ่งภายนอก (นั่นคือ ค่าของข้อมูล) ซึ่งไม่ได้ทำให้ระดับถัวเฉลี่ยเป็นไปตามที่คาดหวังไว้ถึงแม้ว่า Nonlinear Regression จะหลีกเลี่ยงปัญหาหลักๆของ Linear Regression ได้แต่ก็ยังคงไม่มีความยืดหยุ่นเพียงพอที่จะคลุมรูปร่างของกราฟข้อมูลได้ทั้งหมด นี่คือสิ่งที่ทำให้กระบวนการในการวิเคราะห์ทางสถิติและกระบวนการทำดาต้าไมนิ่งแตกต่างกัน วิธีการวัดที่อาศัยหลักทางสถิตินี้ถูกค้นหาเพื่อที่จะสร้างแบบจำลองของ Linear เพื่อที่จะได้พยากรณ์ในรายละเอียดของข้อมูลต่อไปได้ อย่างไรก็ตามข้อมูลส่วนใหญ่ไม่ได้เป็น Linear โดยธรรมชาติดาต้าไมนิ่ง ต้องการกระบวนการทางสถิติเพื่อที่จะปรับให้ ข้อมูลที่เป็น Nonlinearity Outlier และ Non-Numeric ปรับเข้าหากันได้ การประยุกต์ใช้ของการทำนายค่า เช่น การป้องกันการฉ้อโกงบัตรเครดิต หรือการกำหนดรายชื่อของเป้าหมายที่ต้องการส่งจดหมาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 Database Segmentation จุดประสงค์ของการแบ่งส่วนฐานข้อมูลก็เพื่อที่จะแบ่งสรรฐานข้อมูลออกเป็นส่วนๆ หรือการจับกลุ่มของข้อมูลที่มีความคล้ายกัน โดยไม่รู้จำนวนของส่วน หรือ กลุ่มที่ต้องการแบ่ง นั่นคือ เรคคอร์ดที่มีคุณสมบัติร่วมกันจะถูกพิจารณาว่ามีความเหมือนกัน (ภายในส่วนเดียวกันมีความเหมือนกันมาก และ คนละส่วนจะมีความแตกต่างกันมาก) วิธีนี้ใช้การเรียนรู้ที่ไม่อาจควบคุมได้ เพื่อค้นหาความเหมือนกันในส่วนย่อยๆ ของประชากรในฐานข้อมูล ที่มีการตรวจสอบความแม่นยำของโครงสร้างแล้ว การแบ่งส่วนฐานข้อมูลเป็นโอเปอเรชันที่มีความแม่นยำน้อยกว่าโอเปอเรชันอื่น ดังนั้นจึงมีความไวต่อการซ้ำซ้อนและคุณสมบัติที่ไม่ตรงประเด็นน้อย ความไวนี้สามารถที่จะถูกลดลงได้โดยการละเลยกลุ่มย่อยของ แอตทริบิว ซึ่งเป็นรายละเอียดของค่าคงที่แต่ละตัวหรือ โดยการให้น้ำหนักกับปัจจัยแต่ละตัวแปร การแบ่งส่วนฐานข้อมูลประกอบด้วยเทคนิค Demographic Clustering และ Neural Clustering

Demographic Clustering เป็นหลักการของการให้คะแนนแบบง่ายๆ ซึ่งถูกเรียกว่า Condorset ด้วยการวัดระยะห่างระหว่างเรคคอร์ด ที่จะใส่เข้าไปและกำหนดให้เรคคอร์ดเหล่านี้เข้ากลุ่มใดกลุ่มหนึ่ง โดยข้อมูล 2 เรคคอร์ดจะถูกเปรียบเทียบกันด้วยค่าในแต่ละฟิลด์ ฟิลด์ที่มีค่าเหมือนกันจะให้คะแนนเป็น +1 ฟิลด์ที่มีค่าต่างกันจะให้คะแนนเป็น -1 เรคคอร์ดที่มีจำนวนฟิลด์เหมือนกันหลายฟิลด์จะจัดให้อยู่ในกลุ่มที่คล้ายกัน ถ้ามีหลายฟิลด์ที่ไม่เหมือนกันจะจัดให้อยู่ในกลุ่มที่แตกต่างกัน สุดท้ายจะมีการรวมคะแนนของเรคคอร์ดที่ได้มา เพื่อนำเรคคอร์ดนั้นไปใส่ใน segment ที่เหมาะสม โดยถ้าคะแนนรวมของเรคคอร์ดนั้นสูงกว่าคะแนนรวมของ segment นั้น ๆ ก็ จะปรับให้ไปอยู่ใน segment อื่น แต่ถ้าคะแนนรวมของเรคคอร์ดต่ำกว่าคะแนนรวมของ segment ก็ ให้แยกไปตั้ง segment ใหม่ ข้อดีของเทคนิคนี้คือสามารถที่จะสร้าง segment ให้เหมาะสมกับข้อมูลได้โดยอัตโนมัติ และสามารถแบ่งส่วนข้อมูลที่มีขนาดใหญ่ๆ ได้อย่างชัดเจนและรวดเร็ว ดังนั้นวิธีนี้จึงเหมาะกับข้อมูลที่มีลักษณะเป็นกลุ่ม โดยเฉพาะกลุ่มข้อมูลที่มีขนาดเล็ก

Neural Clustering เป็นเทคนิคที่เหมาะสมกับข้อมูลเข้าที่เป็นตัวเลขเท่านั้น โดย Neural Network นั้นมีอัลกอริทึมในการทำงานหลายชนิด เช่น อัลกอริทึมของ Kohonen Feature Map ซึ่งการทำงานประกอบด้วยหน่วยประมวลผล 2 ชั้น โดยชั้นของข้อมูลเข้า (Input Layer) จะเชื่อมต่อกับชั้นของผลลัพธ์ (Output) อย่างสมบูรณ์ เมื่อมีรูปแบบของข้อมูลเข้ามาในชั้นของข้อมูลเข้า หน่วยต่าง ๆ ในชั้นของผลลัพธ์ที่ชนะคือหน่วยที่ถูกลงความเห็นว่าคุณคล้ายกับรูปแบบของข้อมูลเข้ามากที่สุด และจะได้รับสิทธิ์ที่จะถ่วงน้ำหนักการเชื่อมต่อ สำหรับการประยุกต์ใช้ Database segmentation เช่น การรักษาลูกค้า ส่วนแบ่งหรือเป้าหมายทางการตลาด

2.3.3 Link Analysis จุดประสงค์ของ Link Analysis ก็เพื่อการค้นหาความสัมพันธ์ระหว่างข้อมูลแต่ละตัวหรือ แต่ละกลุ่มในฐานข้อมูล ซึ่งมีอยู่ 3 วิธีคือ การค้นหาความสัมพันธ์ (Associations Discovery) การค้นหารูปแบบที่เกี่ยวข้องกัน (Sequential Pattern Discovery) และการค้นหาลำดับเหตุการณ์ที่เกิดขึ้นในเวลาใกล้เคียงกัน (Similar Time Sequence Discovery)

การค้นหาความสัมพันธ์ (Associations Discovery) เป็นการค้นหารายการที่มีความสำคัญซึ่งปรากฏร่วมอยู่กับรายการอื่นในเหตุการณ์ที่คล้ายกัน ความคล้ายกันระหว่างรายการนี้ถูกนำเสนอโดยกฎเกณฑ์ของความสัมพันธ์ ตัวอย่างเช่น เมื่อลูกค้าเช่าทรัพย์สินเกินกว่า 2 ปีและมีอายุมากกว่า 25 ปี แล้วพบว่า 40% ในกรณีนี้ลูกค้าจะซื้อทรัพย์สินนั้นเลย ซึ่งความสัมพันธ์นี้เกิดขึ้น 35% ของลูกค้าทั้งหมดที่ทำการเช่าทรัพย์สิน

การค้นหารูปแบบที่เกี่ยวข้องกัน (Sequential Pattern Discovery) เป็นการค้นหารูปแบบระหว่างเหตุการณ์ ซึ่งเป็นการแสดงให้เห็นว่า 1 ชุดของรายการหนึ่งแปรผันตามอีก 1 ชุดของรายการหนึ่งในฐานข้อมูลในช่วงเวลาหนึ่ง ตัวอย่างนี้ถูกใช้เพื่อที่จะทำให้เข้าใจถึงพฤติกรรมของการซื้อของลูกค้าในระยะยาว เช่น ในตารางที่ 2.1 แสดง Transaction Database ในตารางที่ 2.2 แสดง customer Sequence และในตารางที่ 2.3 แสดง Sequential Patterns with support > 40%

การค้นหาลำดับเหตุการณ์ที่เกิดขึ้นในเวลาใกล้เคียงกัน (Similar Time Sequence Discovery) ถูกใช้ในกรณีตัวอย่างของการค้นหาความสัมพันธ์ระหว่างข้อมูล 2 กลุ่มที่ขึ้นอยู่กับเวลา และอยู่บนระดับของความเหมือนกันระหว่างรูปแบบ ซึ่ง เป็นการแสดงให้เห็นลำดับของเวลาของทั้ง 2 กลุ่ม ตัวอย่างเช่น ผู้ชายปลีกต้องการการสั่งซื้อสินค้าและการจัดเก็บสินค้าไว้ในคลังสินค้าที่ดีที่สุด ผู้ชายปลีกก็จะทำการสังเกตการขายสินค้าทุกวัน ทุกสัปดาห์ว่าขายสินค้ากลุ่มไหนไปบ้างในช่วงเวลาที่ใกล้เคียงกัน และก็จะสังเกตได้ถึงจำนวนสินค้าที่นิยมใช้ในช่วงระยะเวลานั้นๆ ซึ่งข้อมูลนี้ผู้ชายปลีกก็จะสามารถใช้ในการทำนายการจัดซื้อสินค้าเพื่อมาเก็บไว้ในคลังสินค้าให้เหมาะสมในฤดูกาล (ช่วงระยะเวลา) เดียวกันของปีถัดไปได้

Customer	Transaction Time	Items Bought
B.Adams	June 21,1994 5:27 pm	Beer
B.Adams	June 22,1994 10:34 am	Brandy
J.Brown	June 20,1994 10:13 am	Juice,Coke
J.Brown	June 20,1994 11:47 am	Beer
J.Brown	June 21,1994 9:22 am	Wine,water,cider
J.Mitchell	June 21,1994 3:19 pm	Beer,Gin,Cider
B.Moore	June 21,1994 2:32 pm	Beer
B.Moore	June 21,1994 6:17 pm	Wine,Cider
B.Moore	June 21,1994 5:03 pm	Brandy
F.Zappa	June 21,1994 11:02 pm	Brandy

ตารางที่ 2.1 แสดง Transaction Database

Customer	Customer Sequences
B.Adams	(Beer)(Brandy)
J.Brown	(Juice,coke)(Beer)(Wine,Water,Cider)
J.Mitchell	(Beer,Gin,Cider)
B.Moore	(Beer)(Wine,Cider)(Brandy)
F.Zappa	(Brandy)

ตารางที่ 2.2 แสดง Customer Sequences

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sequential Patterns with Support > 40%	Supporting Customers
(Beer)(Brandy)	B.Adams, B.Moore
(Beer)(Wine,Cider)	J.Brown, B.Moore

ตารางที่ 2.3 แสดง Sequential Patterns with Support > 40%

การประยุกต์ใช้ Link Analysis เช่น การวิเคราะห์ความสัมพันธ์ของผลิตภัณฑ์ต่างๆ การซื้อขายโดยตรง และ การวิเคราะห์พฤติกรรมการณ์การซื้อสินค้า

2.3.4 Deviation Detection คือ โอเปอเรชั่นใหม่ที่คำนึงถึงผลประโยชน์ ซึ่งหาได้จากเครื่องมือของการทำดาต้าไมนิ่งเป็นสำคัญ อย่างไรก็ตามการค้นหาค่าเบี่ยงเบนเป็นแหล่งของความจริงที่ค้นพบ เพราะว่ามันชี้ให้เห็นถึงสิ่งที่อยู่ภายนอก (Outliers) ซึ่งค่าเบี่ยงเบนนั้นเกิดจากการคาดหมายบางอย่างที่รู้มาก่อนและมาตรฐาน ที่แน่นอน โอเปอเรชั่นนี้สามารถถูกปฏิบัติโดยใช้เทคนิคของ Statistics และ Visualization หรือ โดยใช้ผลิตภัณฑ์ของ ดาต้าไมนิ่ง ตัวอย่างเช่น Linear Regression ทำให้ง่ายขึ้นด้วยการกำหนดสิ่งที่อยู่ภายนอกข้อมูล ขณะที่เทคนิค Visualization จะแสดงให้เห็นถึง ผลสรุปและแสดงให้เห็นด้วยกราฟ ซึ่งทำให้ง่ายที่จะค้นหาการเบี่ยงเบน

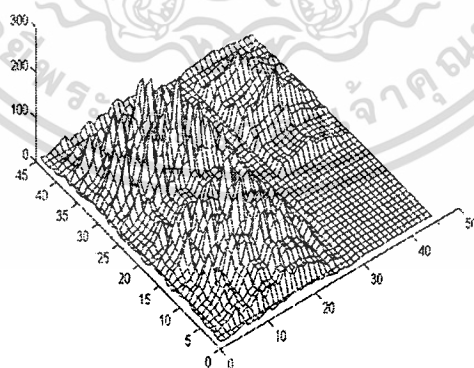


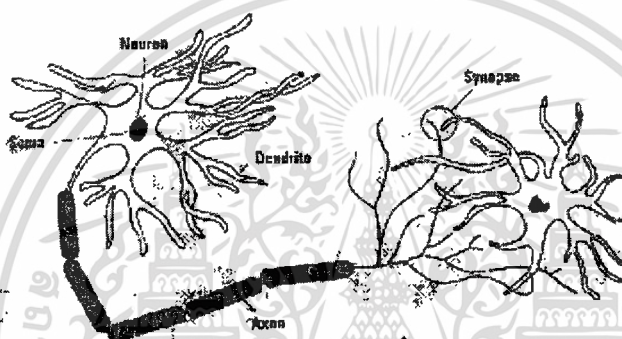
Figure 25.8 An example of visualization of the data shown in Figure 25.7.

## รูปที่ 2.4 การนำเทคนิค Visualization ไปใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในรูป 2.4 แสดงให้เห็นถึงการนำเทคนิค Visualization ไปใช้กับข้อมูลของการสังเกต 200 ครั้งพบว่า เป็น Banknote จริง 100 ครั้ง และเป็น Banknote ปลอม 100 ครั้ง การประยุกต์ใช้ค้นหา ค่าเบี่ยงเบน เช่น การป้องกันการฉ้อโกงในการใช้บัตรเครดิตและการเรียกร้องการประกันภัย การควบคุมคุณภาพ และการตรวจค้นหาจุดบกพร่อง

## 2.4 โครงข่ายประสาทเทียม (Neural Network)



รูปที่ 2.5 Biological Neurons

โครงสร้างทางธรรมชาติของระบบประสาทจะประกอบด้วย Axon ซึ่งเป็นแอกซอนซึ่งแอกซอนนี้จะส่งสัญญาณจาก Neural หนึ่งไป Neural หนึ่ง โดยส่งผ่าน Axon โดยที่ จะมีเซลล์ประสาทที่รับสัญญาณเรียกว่า dendrites เป็นเซลล์ที่ใช้รับสัญญาณเข้ามาให้กับ neural โดย neural เป็นจุดกลางที่ใช้ประมวลผลเมื่อประมวลผลเสร็จ ก็ส่งแอกซอนออกไปให้กับ dendrites ของอีกเซลล์หนึ่ง ช่วงต่อระหว่าง Axon กับ dendrites จะมี Gap เรียกว่า synapse โดย synapse จะทำหน้าที่หลังสารเคมีออกมาเมื่อต้องการส่งสัญญาณ สารนี้ก็จะเข้าไปสู่ dendrites ของอีก neural หนึ่ง

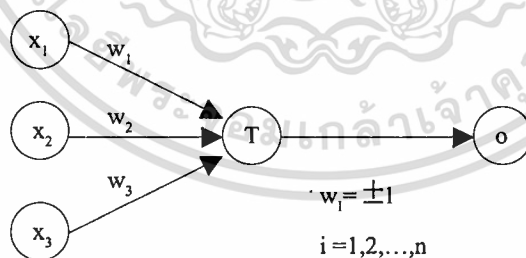
โครงข่ายประสาทเทียมเป็นวิธีหนึ่งซึ่งช่วยให้คอมพิวเตอร์มีความสามารถมากขึ้น โดยเฉพาะการประมวลผลข่าวสารที่มีความยุ่งยากซับซ้อน เช่น การทำให้คอมพิวเตอร์สามารถรู้จำวัตถุ รู้จำตัวอักษร ซึ่งต้องมีการตัดสินใจที่ต้องอาศัยความรู้และประสบการณ์ โดยถ้าใช้เทคนิคทางคณิตศาสตร์ธรรมดาแก้ปัญหาจะมีความซับซ้อนมาก แต่ถ้าใช้ระบบโครงข่ายประสาทเทียมก็จะช่วยลดความยุ่งยากลงได้มาก ซึ่งระบบที่สร้างขึ้นมานี้ถูกเรียกว่าระบบแบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Network System :ANNS) ที่ทำให้คอมพิวเตอร์มีความสามารถใน

การเรียนรู้และสามารถตัดสินใจให้ระบบได้ โดยจะถอดแบบมากจากการทำงานของระบบสมองของมนุษย์

แบบจำลองระบบประสาทที่ใช้ในการประมวลผลโดยเครื่องคอมพิวเตอร์เพื่อนำไปใช้ควบคุม(รักษาสมดุล) ต่าง ๆ นั้นระบบแรกถูกนำเสนอโดย McCulloch และ Pitt ในปีค.ศ.1943 ซึ่งเป็นแบบจำลองของเซลล์ประสาทดังรูปที่ 2.6 อินพุต  $x_i$  (สำหรับ  $i = 1, 2, \dots, n$ ) จะมีค่าเป็น  $\{0, 1\}$  ซึ่งจะขึ้นอยู่กับสัญญาณอินพุตจากเซลล์อื่นในขณะนั้นว่าจะมีหรือไม่มีสัญญาณ ส่วนสัญญาณที่จะส่งต่อไปยังเซลล์ถัดไปซึ่งเป็นเซลล์ผลลัพธ์(จะแทนด้วย  $o$ ) และ Firing Level ของแบบจำลองนี้ถูกกำหนดโดย

$$o^{k+1} = \begin{cases} 1 & \text{if } \sum_{i=1}^n w_i x_i^k \geq T \\ 0 & \text{otherwise} \end{cases}$$

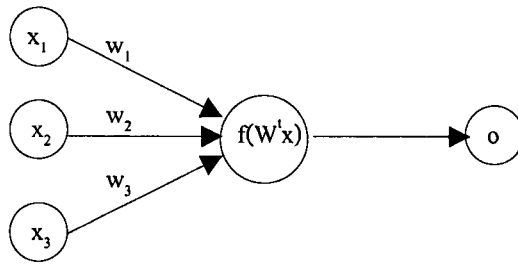
โดยที่  $k=0, 1, 2, \dots$  เป็นช่วงเวลาแบบไม่ต่อเนื่อง  $w_i$  เป็นค่าถ่วงน้ำหนัก ที่เชื่อมต่อกับอินพุตที่  $i$  ซึ่งถ้า  $w_i = +1$  แสดงถึงการกระตุ้นของซินแนปส์และ ถ้า  $w_i = -1$  ซินแนปส์จะมีการยับยั้งการส่งผ่านสัญญาณ และ  $T$  เป็นค่าความต่างศักย์เทรชโฮลด์หรือขีดเริ่มเปลี่ยนซึ่งถ้าค่าผลรวมของผลคูณระหว่างค่าถ่วงน้ำหนักกับสัญญาณอินพุตจะต้องมากกว่า  $T$  จึงจะมีสัญญาณผ่านไปยังเซลล์อื่นได้



รูปที่ 2.6 แบบจำลองเซลล์ประสาทของ McCulloch-Pitts

โครงข่ายอีกแบบหนึ่งซึ่งคล้ายกับแบบจำลองของ McCulloch-Pitts แต่แตกต่างกันตรงที่ค่าของตัวแปรต่าง ๆ ที่ใช้ในแบบจำลอง ANNs เป็นเลขจำนวนจริงและมีค่าถ่วงน้ำหนักจะได้รับการเรียนรู้ของระบบซึ่งแบบจำลองนี้แสดงในรูปที่ 2.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 แบบจำลองเซลล์ประสาทเทียม

จากรูปที่ 2.7 แสดงโครงข่ายการเชื่อมต่อของแบบจำลองเซลล์ประสาทที่สามารถสอนให้โครงข่ายตัดสินใจได้ โดยมี  $x_i$  เป็นสัญญาณอินพุต และ  $w$  เป็นค่าถ่วงน้ำหนักที่ได้จากการสอนโครงข่าย และแต่ละโหนดในโครงข่ายจะใช้แทนเซลล์ประสาทแต่ละเซลล์ ซึ่งบางครั้งจะเรียกว่าหน่วยประมวลผลพื้นฐาน (Process Element Unit) และมีซีแนปส์ ซึ่งจะเชื่อมต่อโหนดเพื่อใช้ในการส่งสัญญาณการกระตุ้นหรือยับยั้งสัญญาณจะขึ้นกับค่าถ่วงน้ำหนัก  $w_i$  และสำหรับสัญญาณเอาต์พุตสามารถจะคำนวณได้ดังนี้

$$o = f(W'x)$$

โดยที่  $W$  เป็นเวกเตอร์ของค่าถ่วงน้ำหนักซึ่งสามารถกำหนดได้ดังนี้

$$W = [w_1, w_2, \dots, w_n]^t$$

และ  $X$  เป็นเวกเตอร์อินพุต  $X = [x_1, x_2, \dots, x_n]^t$  เมื่อ  $t$  เป็นตัวดำเนินการทรานสโพสค์ของเมตริก ฟังก์ชันกำหนดสัญญาณเอาต์พุตในสมการที่ต่อไปนี้จะถูกเรียกว่าฟังก์ชันการเร่งเร้าหรือแอกติเวชันฟังก์ชัน (Activation Function) และกำหนดให้

$$net = W^t X = \sum_{i=1}^n w_i x_i$$

ฟังก์ชันการเร่งเร้ามีคุณสมบัติคล้ายกับกราฟของศักย์ไฟฟ้าขณะทำงาน มีด้วยกันสองชนิดคือ ชนิดที่เป็นเชิงเส้นและชนิดที่ไม่เป็นเชิงเส้น การที่จะกำหนดว่าฟังก์ชันการเร่งเร้าใดมีคุณสมบัติที่สุดคงเป็นไปไม่ได้ เนื่องจาก ANNs ไม่มีโครงข่ายที่แน่นอน นั่นคือโครงข่ายหนึ่งจะใช้แก้ปัญหาใดปัญหาหนึ่งเท่านั้น ดังนั้นการเลือกฟังก์ชันการเร่งเร้าไฉจะต้องพิจารณาให้เหมาะสมกับปัญหานั้น ๆ ตัวอย่างแอกติเวชันฟังก์ชันแบบต่อเนื่องที่นิยมใช้กันมากคือฟังก์ชัน Sigmoid

$$f(net) = 1 / (1 + e^{-net})$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.5 การเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้ของโครงข่าย ANNS จะมีประสิทธิภาพเพียงใดนั้นขึ้นอยู่กับค่าถ่วงน้ำหนักของโครงข่าย ซึ่งการฝึกสอน(Training)โครงข่าย ก็คือการหาค่าถ่วงน้ำหนักที่เหมาะสมให้แก่โครงข่ายนั้น ๆ วิธีการสอน ANNS มีอยู่ 2 แบบด้วยกันคือ

2.5.1 การเรียนรู้แบบชี้แนะ(Supervised Learning) การสอน โดยวิธีนี้จะกำหนดเซตของการสอนให้กับโครงข่าย ซึ่งเซตนี้ประกอบด้วยอินพุตและเอาต์พุตที่ต้องการ(Output Desired) เมื่อป้อนอินพุตให้กับโครงข่าย โครงข่ายก็จะมีการประมวลผลจนได้คำตอบและค่าถ่วงน้ำหนักออกมาชุดหนึ่ง สำหรับคำตอบที่ได้จากโครงข่ายจะถูกนำมาคำนวณค่าความผิดพลาดโดยวัดเป็นระยะทางว่ามีความห่างจากคำตอบที่ต้องการของอินพุตในชุดเดียวกันมากน้อยเพียงใด ถ้ายังมีความผิดพลาดสูงอยู่ก็จะมีการปรับค่าถ่วงน้ำหนัก และทำการสอนต่อไปจนกว่าค่าความผิดพลาดระหว่างคำตอบโครงข่ายกับเอาต์พุตที่ต้องการมีค่าน้อยพอที่จะยอมรับได้จึงจะหยุดการสอน และค่าถ่วงน้ำหนักที่ได้จะเป็นเหมือนฟังก์ชันที่ใช้ในการแปลงข้อมูล

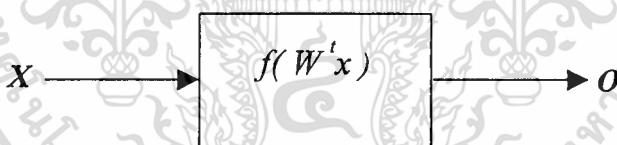
2.5.2 การเรียนรู้แบบไม่มีการชี้แนะ(Unsupervised Learning) การสอนโดยวิธีนี้จะป้อนอินพุตเข้าสู่โครงข่าย และภายในโครงข่ายจะมีเอาต์พุตโหนดอยู่หลายโหนดด้วยกัน โดยแต่ละโหนดจะแทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนอินพุตเข้าสู่โครงข่าย โครงข่ายจะคำนวณค่าความสัมพันธ์ที่มีอยู่ภายในเซตของอินพุต โดยอาศัยค่าถ่วงน้ำหนักเป็นตัวแยกความแตกต่างของอินพุตไปเก็บไว้ในโหนดเอาต์พุตของ โครงข่าย การสอนโดยวิธีนี้ จะไม่สามารถระบุได้ว่าเอาต์พุตโหนดใดเป็นของข้อมูลกลุ่มไหน ผู้ใช้จะต้องกำหนดเองซึ่งแตกต่างจากการสอนแบบชี้แนะที่โครงข่ายสามารถระบุกลุ่มเอาต์พุตได้อย่างแน่นอน

การสอนโครงข่ายเซลล์ประสาทเป็นการหาฟังก์ชันการแปลง และฟังก์ชันการแปลงที่ได้จะมีคุณสมบัติไม่เป็นเชิงเส้น ซึ่งฟังก์ชันการแปลงของ ANN ในที่นี้คือเซตของค่าถ่วงน้ำหนักของโครงข่าย ดังนั้นฟังก์ชันการแปลงจะมีศักยภาพมากน้อยเพียงใดนั้นจะขึ้นอยู่กับค่าถ่วงน้ำหนักของโครงข่ายนั้น ๆ ว่ามีเสถียรภาพมากน้อยเพียงใด และ ค่าถ่วงน้ำหนักคำนวณได้จากการสอนโครงข่าย ซึ่งการสอนโครงข่ายมีหลายแบบด้วยกัน เช่น กฎการสอนของ Hebb กฎการสอนแบบ Perceptron ของ Rosenblatt กฎการสอนแบบเดลต้า กฎของ Widrow-Take-All และกฎการสอนแบบ Outsta ของ Grossberg

## 2.6 แบบจำลองของ ANN

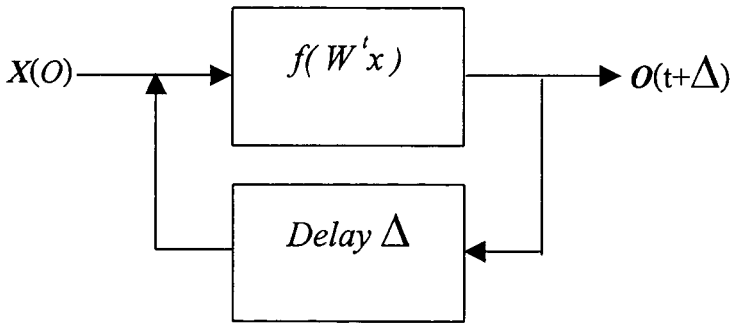
โครงข่ายเซลล์ประสาทที่เป็นแบบจำลองและเป็นโครงข่ายของเซลล์ประสาทจริงของมนุษย์ จะมีการเชื่อมต่อกันของโหนดในลักษณะของโครงข่ายอย่างแน่นอนหา เพื่อให้โครงข่ายสามารถเรียนรู้และสามารถจดจำสิ่งที่เรียนรู้มาแล้วได้ ซึ่งการเชื่อมโยงของโครงข่ายจะมีอยู่สองลักษณะด้วยกันคือ

2.6.1 โครงข่ายที่ส่งสัญญาณไปข้างหน้า(Feedforward Networks) โครงข่ายชนิดนี้จะประกอบด้วยชั้นต่าง ๆ ของโครงข่ายโดยชั้นแรกจะเป็นอินพุตและชั้นสุดท้ายเป็นชั้นของเอาต์พุต ส่วนระหว่างชั้นอินพุตกับเอาต์พุตอาจจะมีหรือไม่มีชั้นที่แทรกอยู่ภายในก็ได้ ซึ่งจะขึ้นอยู่กับอัลกอริทึมที่ใช้ในการสอนโครงข่าย เช่นถ้าเป็นโครงข่าย Perceptron แบบหลายชั้น(Multilayer Perceptron) ก็จะมีชั้นที่อยู่ระหว่างอินพุตกับเอาต์พุตอีกซึ่งอาจมีมากกว่าหนึ่งชั้นก็ได้ ส่วนโครงข่าย Self-Organizing Map ของ Kohonen จะมีเพียงชั้นของอินพุตกับเอาต์พุตเท่านั้น การเชื่อมต่อระหว่างโครงข่ายแบบ Feedforward จะมีค่าถ่วงน้ำหนักเป็นตัวเชื่อม และสัญญาณ อินพุตที่เข้ามาจะถูกส่งไปตามทิศทางของลูกศรจนถึงชั้นเอาต์พุต โดยไม่มีการป้อนกลับ ดังรูปที่ 2.8



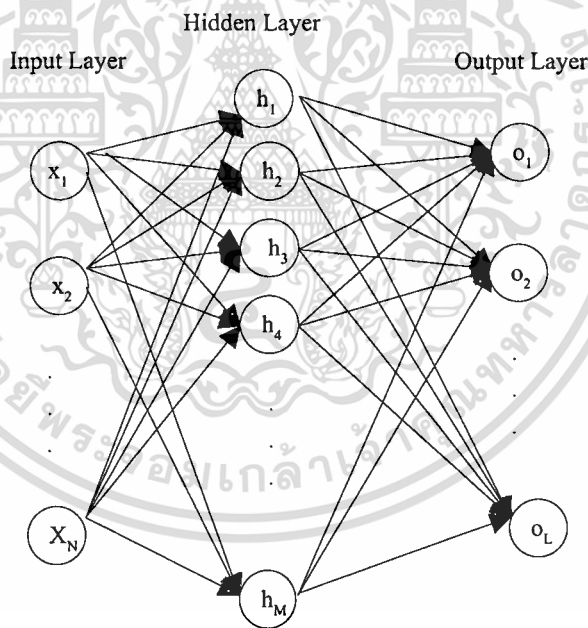
รูปที่ 2.8 บล็อกไดอะแกรมของโครงข่าย Feedforward

2.6.2 โครงข่ายที่มีการป้อนกลับ(Feedback Networks) ในส่วนแรกของโครงข่ายนี้จะเป็นโครงข่าย Feedforward เหมือนกับแบบแรก และส่วนที่เพิ่มเข้ามาคือส่วนของการป้อนกลับซึ่งมีการหน่วงเวลาไปจากเวลาเดิมเท่ากับ  $\Delta$  ดังรูปที่ 2.9



รูปที่ 2.9 บล็อกไดอะแกรมของโครงข่าย Feedback

2.6.3 การแพร่กระจายกลับ(Backpropagation) การแพร่กระจายกลับหรือแบคโพรพาแกชัน (Backpropagation) เป็นขั้นตอนที่ใช้สอนโครงข่ายแบบ Multilayer Perceptron ซึ่งเป็นแบบจำลองโครงข่ายเซลล์ประสาทที่มีการเชื่อมโยงกันเป็นโครงข่ายแบบเป็นชั้น ๆ ดังรูปที่ 2.10



รูปที่ 2.10 โครงข่าย Multilayer Perceptron ที่มี 3 ชั้น

โครงข่ายนี้มีการเชื่อมโยงกัน 3 ชั้น ซึ่งประกอบด้วยชั้นของอินพุตซึ่งมีเซลล์ประสาทอยู่ N โหนด ถัดมาเป็นชั้นของฮิดเดนหรือชั้นภายใน (Hidden Layer) ซึ่งประกอบด้วยโหนดต่าง ๆ จำนวน M โหนด และสุดท้ายคือ ชั้นของเอาต์พุตซึ่งมีโหนดต่าง ๆ อยู่ L โหนด โครงข่ายแบบ Multilayer Perceptron ในรูปที่ 2.10 แต่ละโหนดในชั้นเดียวกันจะไม่มี การเชื่อมโยงกัน การเชื่อมโยงกันจะมีเฉพาะระหว่างชั้นเท่านั้น และการเชื่อมโยงนี้จะต่อถึงกันทุกโหนด โครงข่ายแบบนี้อาจจะมีความซับซ้อนมากขึ้นได้ด้วยการเพิ่มจำนวนโหนดในแต่ละชั้น การเชื่อมโยงกันจะทำได้โดยการคำนวณค่าความผิดพลาดของโหนดในแต่ละชั้น และนำค่าความผิดพลาดนี้ไปใช้ในการปรับปรุงน้ำหนักการเชื่อมโยงกันของโหนดในชั้นก่อนหน้า การปรับปรุงน้ำหนักการเชื่อมโยงกันนี้จะทำซ้ำไปเรื่อย ๆ จนกว่าโครงข่ายจะทำงานได้ตามที่ต้องการ

Multilayer Perceptron ไม่จำเป็นต้องมีสามชั้นเหมือนในรูปที่ 2.10 อาจจะมีจำนวนชั้นมากกว่านั้นก็ ได้ ซึ่งอาจมีสี่ชั้นโดยการเพิ่มชั้นฮิดเดนเข้าไปอีกหนึ่งชั้น หรือถ้าต้องการจำนวนชั้นมากกว่านี้ก็ สามารถทำได้ ชั้นฮิดเดนเป็นตัวเพิ่มความสามารถให้แก่โครงข่ายเซลล์ประสาท โครงข่ายเซลล์ ประสาทแบบ Multilayer Perceptron ถ้าไม่มีชั้นฮิดเดนก็จะกลายเป็นโครงข่ายแบบ Perceptron การที่จะกำหนดชั้นของฮิดเดนว่าใน โครงข่ายหนึ่ง ๆ ควรมีฮิดเดนกี่ชั้นและแต่ละชั้นประกอบด้วยกี่ โหนดนั้นไม่มีกฎเกณฑ์ที่แน่นอน ดังนั้นการกำหนดจำนวนชั้นและจำนวนโหนดของแต่ละชั้นของ ฮิดเดนสามารถทำได้โดยการทดลอง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

### แนวคิดและการออกแบบ

#### 3.1 ฐานข้อมูลที่ใช้ในการพัฒนาระบบ

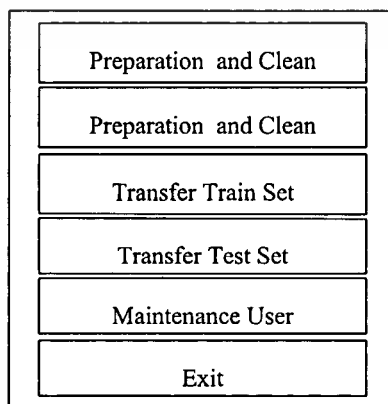
ประกอบด้วยฐานข้อมูลจำนวน 4 ฐานข้อมูลคือ

- 3.1.1 ฐานข้อมูลของภาษีเงินได้นิติบุคคล ซึ่งเป็นฐานข้อมูลที่จัดเก็บรายละเอียดตามแบบแสดงรายการภาษีเงินได้นิติบุคคล(ภ.ง.ด.50)
- 3.1.2 ฐานข้อมูลของภาษีเงินได้นิติบุคคลครึ่งปี ซึ่งเป็นฐานข้อมูลที่จัดเก็บรายละเอียดตามแบบแสดงรายการภาษีเงินได้นิติบุคคลครึ่งปี(ภ.ง.ด.51)
- 3.1.3 ฐานข้อมูลการตรวจสอบภาษี ซึ่งเป็นฐานข้อมูลที่จัดเก็บผลการตรวจสอบที่ได้จากการการตรวจสอบจริง
- 3.1.4 ฐานข้อมูลประเภทกิจการ ซึ่งเป็นฐานข้อมูลที่จัดเก็บรายละเอียดของประเภทกิจการต่างตามที่กรมสรรพากรได้จัดประเภทไว้

ในปัจจุบันฐานข้อมูลทั้ง 4 ฐานข้อมูลได้ถูกออกแบบและใช้งานอยู่แล้วจึงไม่จำเป็นต้องทำการออกแบบใหม่อีก ซึ่งการนำไปประยุกต์ใช้ในการตรวจสอบการหลีกเลี่ยงภาษีเงินได้นิติบุคคลก็เพียงแค่คัดเลือกข้อมูลบางส่วนไปใช้เท่านั้น

#### 3.2 การออกแบบ User Interface

##### 3.2.1 หน้าจอเมนูหลัก



รูปที่ 3.1 หน้าจอเมนูหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2.2 หน้าจอการคัดเลือกข้อมูล( Preparation Train Set และ Preparation Test Set )

Isiccode:

Field Name:

Field Name :

>>

<<

ALL

OK

Cancel

รูปที่ 3.2 หน้าจอการคัดเลือกข้อมูล

### 3.2.3 หน้าจอการ Transform ข้อมูล

**Transfer Data**

**Preview Data:**

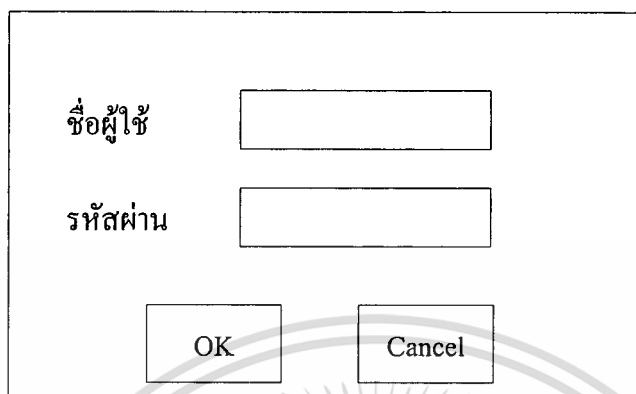
Aaaa	bbbbbb	ccccc	ddddddd	eeeee	ffffff	gggggg
1.						
2.						
3.						
.						
.						
.						

Transfer      Cancel

รูปที่ 3.3 หน้าจอการ Transform ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เฉพาะในเพื่อการวิจัยเท่านั้น มิใช่ผู้จัดทำให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2.4 หน้าจอรระบบรักษาความปลอดภัย



ชื่อผู้ใช้

รหัสผ่าน

OK Cancel

รูปที่ 3.4 หน้าจอรระบบรักษาความปลอดภัย



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 4

### การประยุกต์ใช้ดาต้าไมนิ่งในการตรวจสอบการหลีกเลี่ยงภาษีเงินได้นิติบุคคล

#### 4.1 การคัดเลือกข้อมูลสำหรับใช้ในการตรวจสอบ

ฐานข้อมูลที่ใช้ในการคัดเลือก Attribute เพื่อคัดเลือกรายสำหรับตรวจสอบนั้น มี 4 ฐานข้อมูล คือ

- 4.1.1 ฐานข้อมูลของภาษีเงินได้นิติบุคคล(ภ.ง.ด.50) จำนวน 14 Attribute ได้แก่ เลขประจำตัวผู้เสียภาษี, รายได้จากการขายหรือการประกอบการโดยตรง, ต้นทุนขาย, รหัสกำไร/ขาดทุนขั้นต้น, กำไร/ขาดทุนขั้นต้น, รายได้ที่ไม่เกี่ยวกับการขายหรือการประกอบการโดยตรง, รายจ่ายในการดำเนินงาน, รหัสกำไร/ขาดทุนสุทธิ, กำไร/ขาดทุนสุทธิ, ต้นทุนการผลิต, รวมสินทรัพย์, รวมหนี้สินและทุน, ยอดรวมรายได้, ยอดรวมรายจ่าย
- 4.1.2 ฐานข้อมูลของภาษีเงินได้นิติบุคคลครึ่งปี (ภ.ง.ด.51) จำนวน 2 Attribute ได้แก่ รหัสกำไร/ขาดทุนสุทธิ(ภ.ง.ด.51), กำไร/ขาดทุนสุทธิ(ภ.ง.ด.51)
- 4.1.3 ฐานข้อมูลผลการตรวจสอบภาษี จำนวน 1 Attribute ได้แก่ ผลการตรวจสอบภาษี
- 4.1.4 ฐานข้อมูลประเภทกิจการ จำนวน 1 Attribute ได้แก่ รหัสประเภทกิจการ

ลำดับ	Attribute	ประเภทข้อมูล	ขนาดข้อมูล
1.	เลขประจำตัวผู้เสียภาษี	Character	10
2.	รายได้จากการขายหรือการประกอบการโดยตรง	Numeric	16
3.	ต้นทุนขาย	Numeric	16
4.	รหัสกำไร/ขาดทุนขั้นต้น	Character	1
5.	กำไร/ขาดทุนขั้นต้น	Numeric	16
6.	รายได้ที่ไม่เกี่ยวกับการขายหรือการประกอบการ โดยตรง	Numeric	16
7.	รายจ่ายในการดำเนินงาน	Numeric	16
8.	รหัสกำไร/ขาดทุนสุทธิ	Character	1

ลำดับ	Attribute	ประเภทข้อมูล	ขนาดข้อมูล
9.	กำไร/ขาดทุนสุทธิ	Numeric	16
10.	ต้นทุนการผลิต	Numeric	16
11.	รวมสินทรัพย์	Numeric	16
12.	รวมหนี้สินและทุน	Numeric	16
13.	ยอดรวมรายได้	Numeric	16
14.	ยอดรวมรายจ่าย	Numeric	16
15.	รหัสกำไร/ขาดทุนสุทธิ(ภ.ง.ค.51)	Character	1
16.	กำไร/ขาดทุนสุทธิ(ภ.ง.ค.51)	Numeric	16
17.	ผลการตรวจสอบภาษี	Numeric	1
18.	รหัสประเภทกิจการ	Numeric	6

#### ตารางที่ 4.1 ฐานข้อมูลที่ใช้ในการตรวจสอบ

จากเมนูดังแสดงในรูปที่ 4.1 เป็นการจัดเตรียมข้อมูลเบื้องต้นเพื่อใช้ในการคัดเลือกข้อมูลต่อไป โดยการเตรียมข้อมูลและการ Clean ข้อมูล test set (Preparation and Clean (Test Set)) ใช้รูปแบบคำสั่งดังต่อไปนี้

```

select dist a.fctin,a.fcisiccd1,a.fni04ct49+a.fni04ct50+a.fni04ct51 as
income_di,a.fni04ct52 as cost1,a.fci04ct53 as gross_c,a.fni04ct54 as gross_pl,
a.fni04ct57+a.fni04ct58+a.fni04ct59+a.fni04ct61 AS income_indi,a.fni04ct80 as sale_ad,
a.fci04ct81 as net_c,a.fni04ct82 as net_pl,a.fni07ct137 as assets,a.fni05ct108 as
cost2,a.fni07ct162 as lia_cap,
a.fni04ct49+a.fni04ct50+a.fni04ct51+a.fni04ct57+a.fni04ct58+a.fni04ct59+a.fni04ct61
as T_income,a.fni04ct52+a.fni05ct108+a.fni04ct80 as T_cost,b.c5121 as
net51_c,b.c5122 as net_pl51
from database\cit5040 a, database\cit5140 b
Where a.fccdboi='0' ;
and a.fctin=b.ctin ;
and a.fctin!='0000000000';

```

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

and !empty(a.fctin) ;
and !empty(a.fci04ct53) ;
and !empty(a.fci04ct81) ;
and !empty(b.c5121) ;
and val(a.fci04ct53)>=0 ;
and val(a.fci04ct53)<=2 ;
and a.fni04ct49>=0 ;
and a.fni04ct50>=0 ;
and a.fni04ct51>=0 ;
and a.fni04ct52>=0 ;
and a.fni04ct54>=0;
and a.fni04ct57>=0;
and a.fni04ct58>=0;
and a.fni04ct59>=0;
and a.fni04ct61>=0;
and a.fni04ct80>=0;
and val(a.fci04ct81)>=0 ;
and val(a.fci04ct81)<=2 ;
and a.fni04ct82>=0;
and a.fni07ct137>=0;
and a.fni07ct162>=0;
and a.fni05ct108>=0;
and val(b.c5121)>=0 ;
and val(b.c5121)<=2 ;
and b.c5122>=0;
into table database\cit1

```

การคัดเตรียมข้อมูลและการ Clean ข้อมูล train set (Preparation and Clean (Train Set)) ใช้  
รูปแบบคำสั่งดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

select dist a.fctin,a.fcisiccd1,a.fni04ct49+a.fni04ct50+a.fni04ct51 as
income_di,a.fni04ct52 as cost1,a.fci04ct53 as gross_c,a.fni04ct54 as
gross_pl,a.fni04ct57+a.fni04ct58+a.fni04ct59+a.fni04ct61 AS income_indi,a.fni04ct80
as sale_ad,a.fci04ct81 as net_c,a.fni04ct82 as net_pl,a.fni07ct137 as assets,a.fni05ct108
as cost2,a.fni07ct162 as
lia_cap,a.fni04ct49+a.fni04ct50+a.fni04ct51+a.fni04ct57+a.fni04ct58+a.fni04ct59+a.fni
04ct61 as T_income,a.fni04ct52+a.fni05ct108+a.fni04ct80 as T_cost,b.c5121 as
net51_c,b.c5122 as net_pl51,c.result
from c:\project\database\cit5040 a, c:\project\database\cit5140 b
,c:\project\database\result c
Where a.fcdboi='0' ;
and a.fctin=b.ctin ;
and a.fctin=c.fctin ;
and a.fctin!='0000000000' ;
and !empty(a.fctin) ;
and !empty(a.fci04ct53) ;
and !empty(a.fci04ct81) ;
and !empty(b.c5121) ;
and !empty(c.fctin) ;
and val(a.fci04ct53)>=0 ;
and val(a.fci04ct53)<=2 ;
and a.fni04ct49>=0 ;
and a.fni04ct50>=0 ;
and a.fni04ct51>=0 ;
and a.fni04ct52>=0 ;
and a.fni04ct54>=0;
and a.fni04ct57>=0;
and a.fni04ct58>=0;
and a.fni04ct59>=0;
and a.fni04ct61>=0;

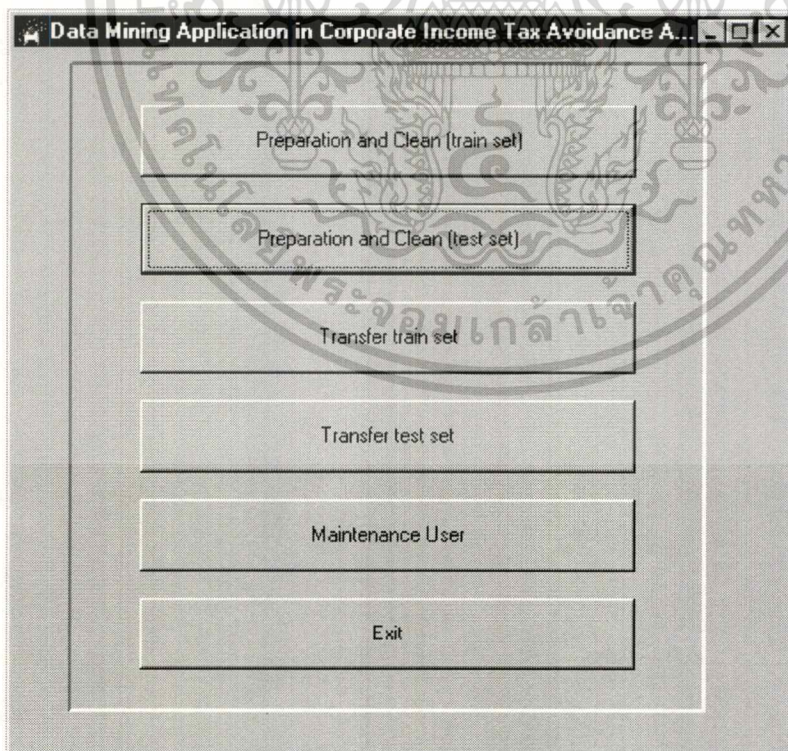
```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

and a.fni04ct80>=0;
and val(a.fci04ct81)>=0 ;
and val(a.fci04ct81)<=2 ;
and a.fni04ct82>=0;
and a.fni07ct137>=0;
and a.fni07ct162>=0;
and a.fni05ct108>=0;
and val(b.c5121)>=0 ;
and val(b.c5121)<=2 ;
and b.c5122>=0;
and a.fctin=c.fctin ;
and a.fcisiccd1=c.fcisiccd1 ;
into table c:\project\database\cit

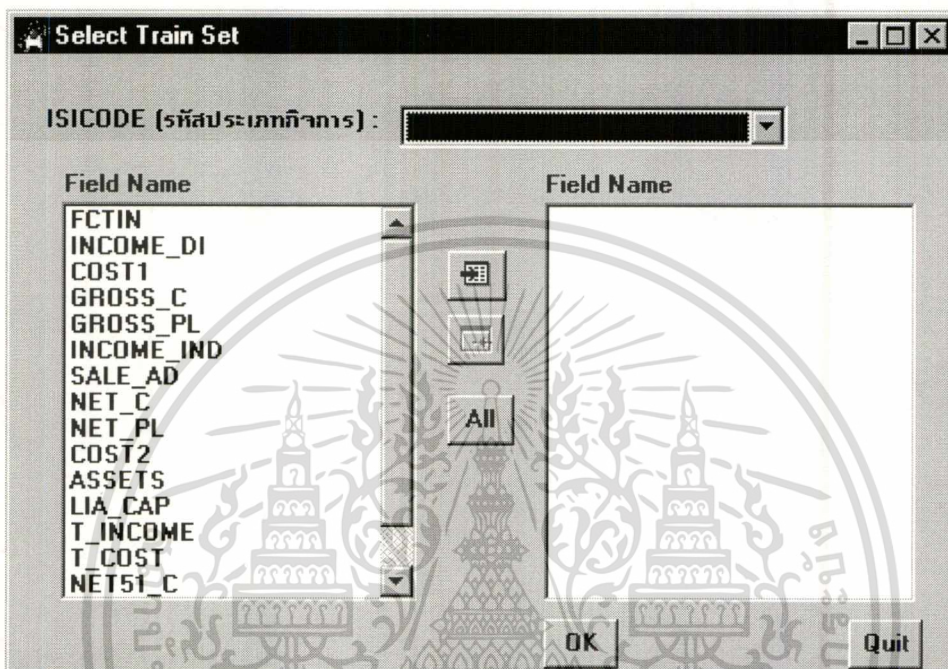
```



#### รูปที่ 4.1 เมนูหลักของการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อได้ Attribute ครบถ้วนแล้วก็ทำการคัดเลือก Attribute ที่จำเป็นต้องใช้ในการตรวจสอบ การหลีกเลี่ยงภาษีอีกครั้งหนึ่งเนื่องจากแต่ละประเภทกิจการจะใช้ Attribute ในการตรวจสอบที่ไม่ เหมือนกัน



รูปที่ 4.2 การคัดเลือก attribute ที่ใช้ในการตรวจสอบการหลีกเลี่ยงภาษี

#### 4.2 การ Clean ข้อมูล

การ Clean ข้อมูลจะกระทำไปพร้อมๆกับขั้นตอนการคัดเลือกข้อมูล โดยเงื่อนไขที่ใช้ในการ Clean ข้อมูลมีดังนี้

- 4.2.1 ทุก Attribute ต้องไม่เป็นค่าว่าง และไม่ซ้ำกัน
- 4.2.2 สำหรับ Attribute ที่เป็นตัวเลขยกเว้นรหัสประเภทกิจการ จะต้องมีค่ามากกว่าหรือเท่ากับ 0 (ต้องไม่เป็นค่าที่ติดลบ)
- 4.2.3 รหัสประจำตัวผู้เสียภาษีต้องไม่มีค่าเป็น '000000000'
- 4.2.4 รหัสกำไรขาดทุน/ขั้นต้น, รหัสกำไร/ขาดทุนสุทธิ, รหัสกำไร/ขาดทุนสุทธิ (ภ.ง.ด.51) ต้องมีค่าเป็น 0 หรือ 1 หรือ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.3 การ Transform ข้อมูล

เมื่อเลือกประเภทกิจการและเลือก Attribute ที่จะเป็นต้องใช้ในแต่ละประเภทกิจการแล้วจะต้องทำการ Transfer ข้อมูลให้มีค่าระหว่าง 0 ถึง 1 โดยมีวิธีการ Transform ข้อมูล ดังนี้

4.3.1 Attribute ที่เป็นตัวเลขจะต้องทำการหาค่า Max และ ค่าMin ใน Attribute นั้น ๆ แล้ว ใช้เงื่อนไขในการแปลงดังนี้

4.3.1.1 ค่าที่เท่ากับ Max ให้แปลงเป็น 1

4.3.1.2 ค่าที่เท่ากับ Min ให้แปลงเป็น 0

4.3.1.3 ค่าที่ไม่ใช่ค่า Max หรือ Min ให้หารด้วยค่า Max

4.3.2 Attribute รหัสกำไรขาดทุน/ขั้นต้น, รหัสกำไร/ขาดทุนสุทธิ, รหัสกำไร/ขาดทุนสุทธิ(ภ.ง.ด.51) ใช้เงื่อนไขในการแปลงดังนี้

4.3.2.1 ถ้ามีค่าเป็น 0 ให้แปลงเป็น 0

4.3.2.2 ถ้ามีค่าเป็น 1 ให้แปลงเป็น 0.5

4.3.2.3 ถ้ามีค่าเป็น 2 ให้แปลงเป็น 1

4.3.3 Attribute ผลการตรวจสอบภาษี(ในส่วนของ Preparation Train Set) ไม่ต้องแปลง เนื่องจากมีค่าเป็น 1 หรือ 0 อยู่แล้ว

FCTIN	INCOME DI	GROSS C	GROSS PL	INCOME IN
▶ 3011000144	0.0		0	
3011000816	0.0		0	5661542.7
3011000847	0.0		0	622620
3011001875	0.0		0	
3011002209	2871443	1	2871443	
3011002698	0.0		0	1197.9
3011005460	0.0		0	
3011007402	0.0		0	1686.9
3011010310	0.0		0	10000
3011011287	478194	1	478194	1060.1
3011012598	700000	1	40218.5	
3011012871	0.0		0	7503304.7
3011012888	1866813.76	1	1866813.76	
3011013069	0.0		0	
3011013274	0.0		0	
3011014882	0.0		0	10000
3011015957	0.0		0	
3011017104	0.0		0	
3011017944	360000	1	360000	5975.9
3011019078	0.0		0	
3011020203	0.0		0	103874.5
3011020869	0.0		0	
3011020982	0.0		0	
3011021019	0.0		0	
3011021644	0.0		0	
3011022528	0.0		0	

Max	Expr1000	Expr1001	Expr1002	Expr1003	Expr1004
▶	3861009083	2825311057.14	2	1552391556.38	839414910.48

Min	Expr1000	Expr1001	Expr1002	Expr1003	Expr1004
▶	3011000144	0.0		0	0.1

รูปที่ 4.3 หน้าจอการ Transform ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.4 การคัดเลือกอัลกอริทึมในการตรวจสอบการหลีกเลี่ยงภาษีเงินได้นิติบุคคล

##### 4.4.1 อัลกอริทึมที่เลือกใช้คือ Backpropagation Algorithm ซึ่งมีขั้นตอนดังต่อไปนี้

###### 4.4.1.1 กำหนดค่า Weight เริ่มต้น

4.4.1.2 ตรวจสอบ Stopping Condition ว่าเป็นจริงหรือเท็จถ้าเป็นจริงให้จบการทำงาน(ขั้นตอนที่4.4.1.10) ถ้าเป็นเท็จทำขั้นตอนต่อไป

4.4.1.3 รับค่า Training Pair (Training Pattern) แล้วส่ง input signal ไปยังทุก ๆ unit ใน layer ถัดไป(Hidden Layer) .....(propagate the input forward)

4.4.1.4 แต่ละ Hidden Unit รับค่า input signal แล้วรวมค่า Weight และใช้ activation function เพื่อคำนวณ output signal แล้วส่ง output signal นั้นไปยังทุก ๆ unit ใน layer ถัดไป(Output Layer) โดย activation function มีดังต่อไปนี้ .....(propagate the input forward)

- Logistic function

$$f(\text{net}) = 1 / (1 + e^{-\text{net}})$$

- Hyperbolic tangent function

$$\begin{aligned} f(\text{net}) &= \tanh(\text{net} / 2) \\ &= (1 - e^{-\text{net}}) / (1 + e^{-\text{net}}) \\ &= (2 / (1 + e^{-\text{net}})) - 1 \\ &= (e^{\text{net}} - e^{-\text{net}}) / (e^{\text{net}} + e^{-\text{net}}) \end{aligned}$$

- Identity function

$$f(\text{net}) = \text{net}$$

โดยที่ net เป็น input weight

4.4.1.5 แต่ละ Output Unit รวมค่า weight และใช้ activation function เพื่อคำนวณ output signal .....(propagate the input forward)

4.4.1.6 แต่ละ Output Unit รับ Target Pattern มีลักษณะเช่นเดียวกับ Training Pattern คำนวณ error information และคำนวณค่า Weight และค่า bias แล้วส่ง error information ไปยัง unit ใน layer ก่อนหน้านั้น(Hidden Layer) .....( propagate the sensitivities backward)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.1.7 แต่ละ Hidden Unit รวมค่าที่อยู่ในเทอมของ error information และใช้ derivative ของ activation function คำนวณ error information และคำนวณค่า Weight และค่า bias โดย derivative ของ activation function มีดังต่อไปนี้ ..... (propagate the sensitivities backward)

- ถ้า  $f(\text{net}) = 1 / (1 + e^{-\text{net}})$

$$f'(\text{net}) = f(\text{net})[1-f(\text{net})]$$

- ถ้า  $f(\text{net}) = (2 / (1 + e^{-\text{net}})) - 1$

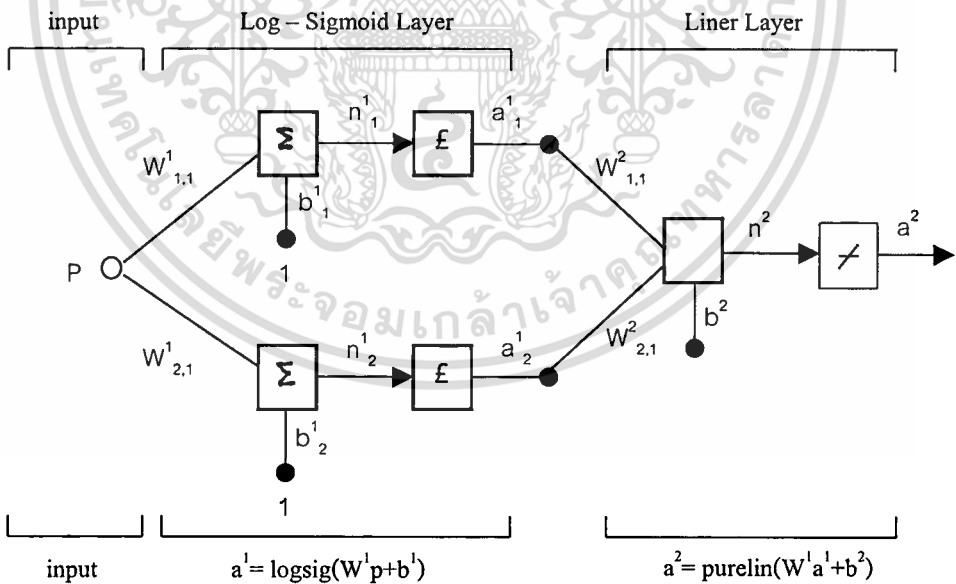
$$f'(\text{net}) = 0.5[1+f(\text{net})][1-f(\text{net})]$$

4.4.1.8 แต่ละ Output Unit และ Hidden Unit ปรับค่า Weight และค่า bias

4.4.1.9 ตรวจสอบ Stopping Condition (ขั้นตอนที่ 4.4.1.2)

4.4.1.10 จบการทำงาน

4.4.2 ตัวอย่างการทำงานของ backpropagation algorithm



รูปที่ 4.4 ตัวอย่าง Backpropagation network

สูตรที่ใช้ในการหาค่าต่าง ๆ มีดังต่อไปนี้

- training pair =  $\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_q, t_q\}$ ;  $q=1, 2, 3, \dots$

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ โดยที่ p = input, t = target  
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$- a^0 = p \text{ สำหรับ layer ที่ } 1 \quad \dots\dots(4.1)$$

$$- a^{m+1} = f^{m+1}(W^{m+1}a^m + b^{m+1}); m = 0, 1, 2, \dots, M-1 \quad \dots\dots(4.2)$$

โดยที่  $M =$  จำนวนของ layer ใน network

$W =$  weight

$a =$  activation function(logsin และ purelin)

$b =$  bias

$$- a = a^M \text{ สำหรับ layer สุดท้าย} \quad \dots\dots(4.3)$$

$$- S^M = -2F^M(n^M)(t-a) \quad \dots\dots(4.4)$$

โดยที่  $S =$  sensitivity backward

$F =$  derivative ของ activation function

$$- S^m = F^m(n^m)(W^{m+1})^T S^{m+1} \quad \dots\dots(4.5)$$

โดยที่  $m = M-1, \dots, 2, 1$

$$- W^m(k+1) = W^m(k) - \alpha S^m (a^{m-1})^T \quad \dots\dots(4.6)$$

โดยที่  $\alpha =$  learning rate

$$- b^m(k+1) = b^m(k) - \alpha S^m \quad \dots\dots(4.7)$$

$$\text{โจทย์กำหนดให้ } W^1(0) = \begin{bmatrix} -0.27 \\ -0.41 \end{bmatrix}, b^1(0) = \begin{bmatrix} -0.48 \\ -0.13 \end{bmatrix}, W^2(0) = [0.09 \quad -0.17],$$

$$b^2(0) = [0.48], p = 1, \alpha = 0.1 \text{ และ } t = \left\{ 1 + \sin\left(\frac{\pi}{4} p\right) \right\} \text{ โดยที่ } -2 \leq p \leq 2$$

$$\text{จาก } a^0 = p = 1$$

$$\text{output ของ layer ที่ } 1 \text{ คือ } a^1 = f^1(W^1 a^0 + b^1)$$

$$= \text{logsig}\left(\begin{bmatrix} -0.27 \\ -0.41 \end{bmatrix} [1] + \begin{bmatrix} -0.48 \\ -0.13 \end{bmatrix}\right)$$

$$= \text{logsig}\left(\begin{bmatrix} -0.75 \\ -0.54 \end{bmatrix}\right)$$

$$= \begin{bmatrix} \frac{1}{1 + e^{-0.75}} \\ \frac{1}{1 + e^{-0.54}} \end{bmatrix}$$

$$= \begin{bmatrix} 0.321 \\ 0.368 \end{bmatrix}$$

$$\text{output ของ layer ที่ } 1 \text{ คือ } a^2 = f^2(W^2 a^1 + b^2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น  $= \text{purelin}\left([0.09 \quad -0.17] \begin{bmatrix} 0.321 \\ 0.368 \end{bmatrix} + [0.48]\right)$  ใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$= [0.446]$$

ค่า error คือ

$$e = t - a$$

$$= \left\{ 1 + \sin\left(\frac{\pi}{4} p\right) \right\} - a^2$$

$$= \left\{ 1 + \sin\left(\frac{\pi}{4} 1\right) \right\} - 0.446$$

$$= 1.261$$

derivative ของ activation function ของ layer ที่ 1 คือ  $f^1(n) = \frac{d}{dn} \left( \frac{1}{1+e^{-n}} \right)$

$$= \frac{e^{-n}}{(1+e^{-n})^2}$$

$$= \left( 1 - \frac{1}{1+e^{-n}} \right) \left( \frac{1}{1+e^{-n}} \right)$$

$$= (1 - a^1)(a^1)$$

derivative ของ activation function ของ layer ที่ 2 คือ  $f^2(n) = \frac{d}{dn}(n) = 1$

sensitivity backward ของ layer ที่ 2 คือ  $S^2 = -2F^2(n^2)(t-a)$

$$= -2[f^2(n^2)](1.261)$$

$$= -2[1](1.261) = -2.522$$

sensitivity backward ของ layer ที่ 1 คือ  $S^1 = F^1(n^1)(W^2)^T S^2$

$$= \begin{bmatrix} (1-a_1^1)(a_1^1) & 0 \\ 0 & (1-a_2^1)(a_2^1) \end{bmatrix} \begin{bmatrix} 0.09 \\ -0.17 \end{bmatrix} [-2.522]$$

$$= \begin{bmatrix} (1-0.321)(0.321) & 0 \\ 0 & (1-0.368)(0.368) \end{bmatrix} \begin{bmatrix} 0.09 \\ -0.17 \end{bmatrix} [-2.522]$$

$$= \begin{bmatrix} 0.218 & 0 \\ 0 & 0.233 \end{bmatrix} \begin{bmatrix} -0.227 \\ 0.429 \end{bmatrix}$$

$$= \begin{bmatrix} -0.0495 \\ 0.0997 \end{bmatrix}$$

weight ที่ปรับ ใน layer ที่ 2 คือ  $W^2(1) = W^2(0) - \alpha S^2(a^1)^T$

$$= [0.09 \quad -0.17] - 0.1[-2.522][0.321 \quad 0.368]$$

$$= [0.171 \quad -0.0772]$$

bias ที่ปรับ ใน layer ที่ 2 คือ  $b^2(1) = b^2(0) - \alpha S^2$

$$= [0.48] - 0.1[-2.522] = [0.732]$$

weight ที่ปรับ ใน layer ที่ 1 คือ  $W^1(1) = W^1(0) - \alpha S^1(a^0)^T$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษา =  $\begin{bmatrix} -0.27 \\ -0.41 \end{bmatrix} - 0.1 \begin{bmatrix} -0.0495 \\ 0.0997 \end{bmatrix} [1]$  ซึ่งประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$= \begin{bmatrix} -0.265 \\ -0.420 \end{bmatrix}$$

bias ที่ปรับ ใน layer ที่ 1 คือ  $b^1(1) = b^1(0) - \alpha S^1$

$$= \begin{bmatrix} -0.48 \\ -0.13 \end{bmatrix} - 0.1 \begin{bmatrix} -0.0495 \\ 0.0997 \end{bmatrix} = \begin{bmatrix} -0.475 \\ -0.140 \end{bmatrix}$$

สำหรับโปรแกรมซึ่งเขียนด้วย MATLAB มีดังต่อไปนี้

```
load cit.txt
load cit2.txt
s=size(cit);
ss=size(cit2);
p=transpose(cit);
P=p(1:s(2)-1,1:s(1));
Q=transpose(cit2);
Q=Q(1:ss(2),1:ss(1));
T=p(s(2),1:s(1));
net=newff(minmax(P),[11 1],{'logsig','purelin'});
net.trainParam.epochs=1000;
net.trainParam.lr=0.01;
net=train(net,P,T);
Y=sim(net,Q);
X=transpose(Y);
m=mean(X);
M=ones(ss(1),ss(2));
M=M*m;
figure(2);
plot(X);
hold;
plot(M,'m');
grid;
disp(X);
disp(m);
%h=hist(X);
%figure(3);
```

เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

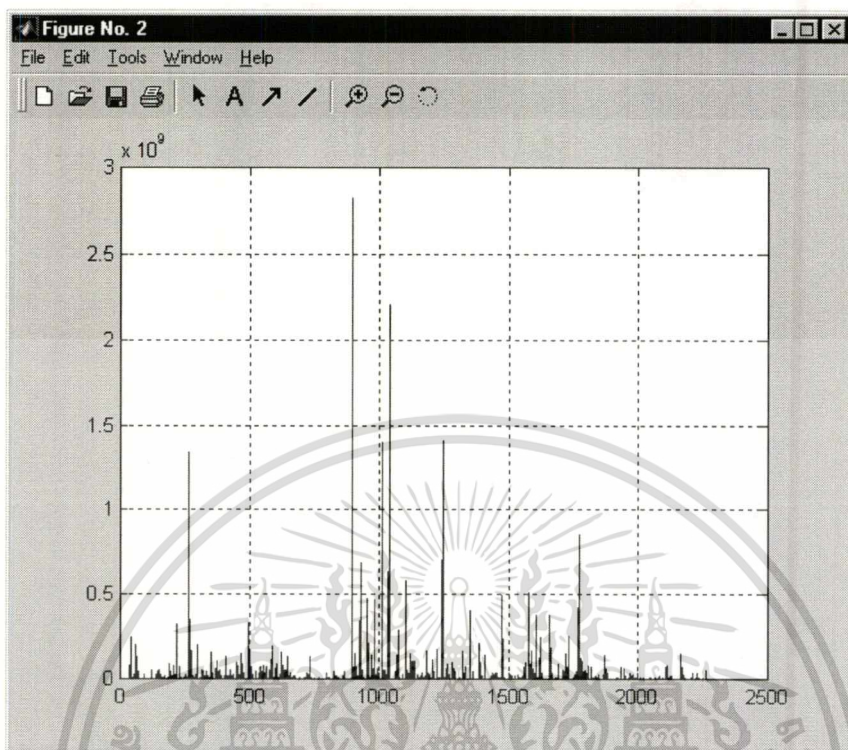
```
%plot(h);
data=cit2;
data(:,ss(2)+1)=X;
save Output.txt data -ascii -tabs;
```

#### 4.5 การวิเคราะห์ผลลัพธ์

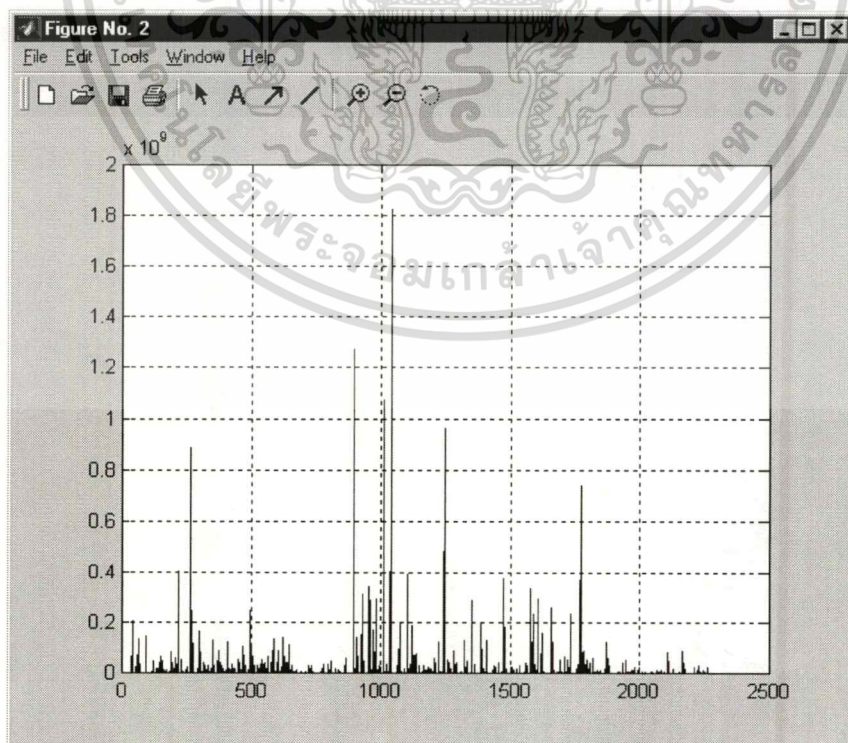
ผลลัพธ์ที่ได้จากการRun โปรแกรมจะอยู่ในคอลัมน์สุดท้ายของไฟล์ชื่อ output.txt ซึ่งสามารถเปิดอ่าน และสามารถสร้างกราฟแสดงผลได้จากโปรแกรม Microsoft Excel อีกทั้งยังสามารถอ่านผลลัพธ์จากกราฟขณะที่ Run โปรแกรม (โปรแกรม MATLAB) โดยผลลัพธ์ที่ได้จะเป็นตัวเลขซึ่งมีความหมายดังนี้

- 4.5.1 ถ้าผลลัพธ์มีค่าเท่ากับ 0 หมายถึง ผู้เสียหายในบรรทัดนั้นไม่เสียหาย
- 4.5.2 ถ้าผลลัพธ์มีค่ามากกว่า 0 แต่น้อยกว่า 0.5 หมายถึง ผู้เสียหายในบรรทัดนั้นน่าจะ  
ไม่เสียหาย
- 4.5.3 ถ้าผลลัพธ์มีค่ามากกว่า 0.5 แต่น้อยกว่า 1 หมายถึง ผู้เสียหายในบรรทัดนั้นน่าจะ  
เสียหาย
- 4.5.4 ถ้าผลลัพธ์มีค่าเท่ากับ 1 หมายถึง ผู้เสียหายในบรรทัดนั้นเสียหาย

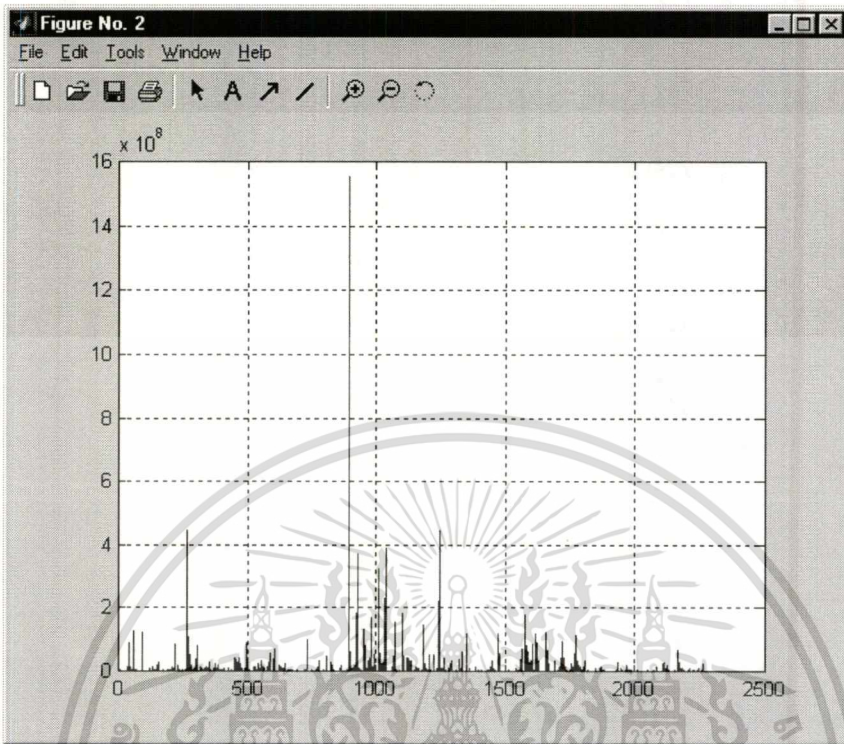




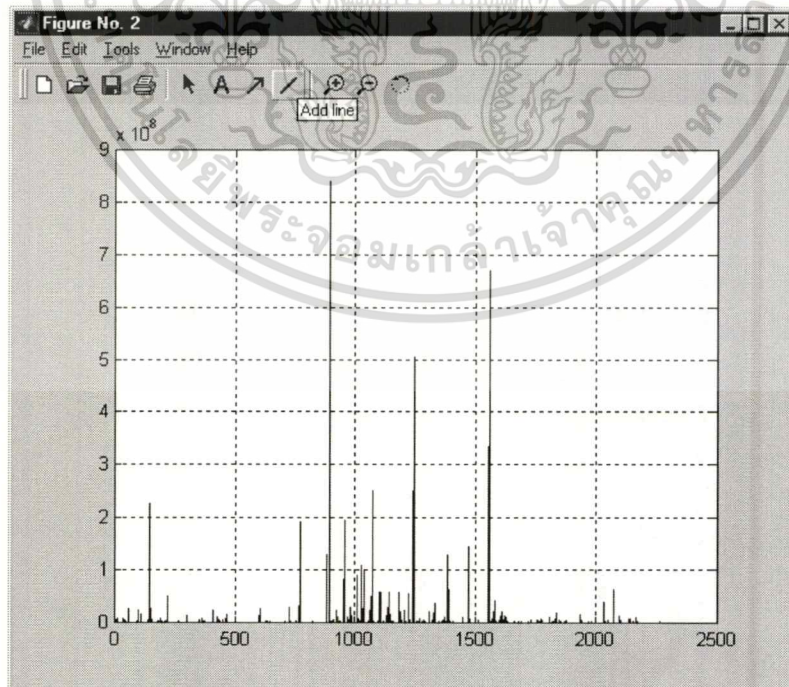
รูปที่ 4.7 แสดงข้อมูล Attribute รายได้จากการขายหรือการประกอบการโดยตรงที่ใช้ในการ Train



เอกสารนี้เป็นเอกสารที่ **รูปที่ 4.8 แสดงข้อมูล Attribute ต้นทุนขายที่ใช้ในการ Train** ไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

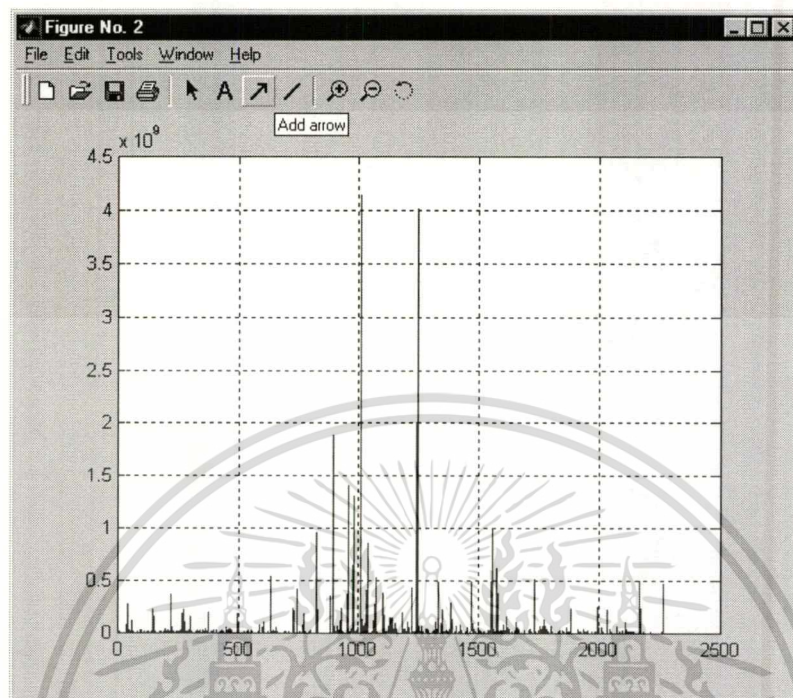


รูปที่ 4.9 แสดงข้อมูล Attribute กำไร/ขาดทุนขั้นต้นที่ใช้ในการ Train

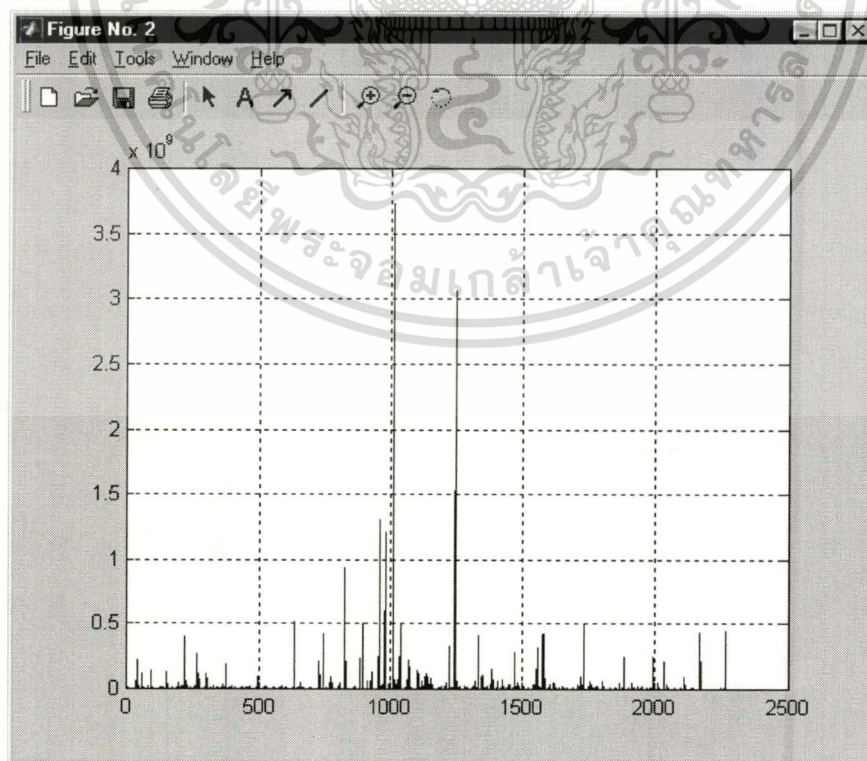


รูปที่ 4.10 แสดงข้อมูล Attribute รายได้ที่ไม่เกี่ยวกับการขายโดยตรงที่ใช้ในการ Train

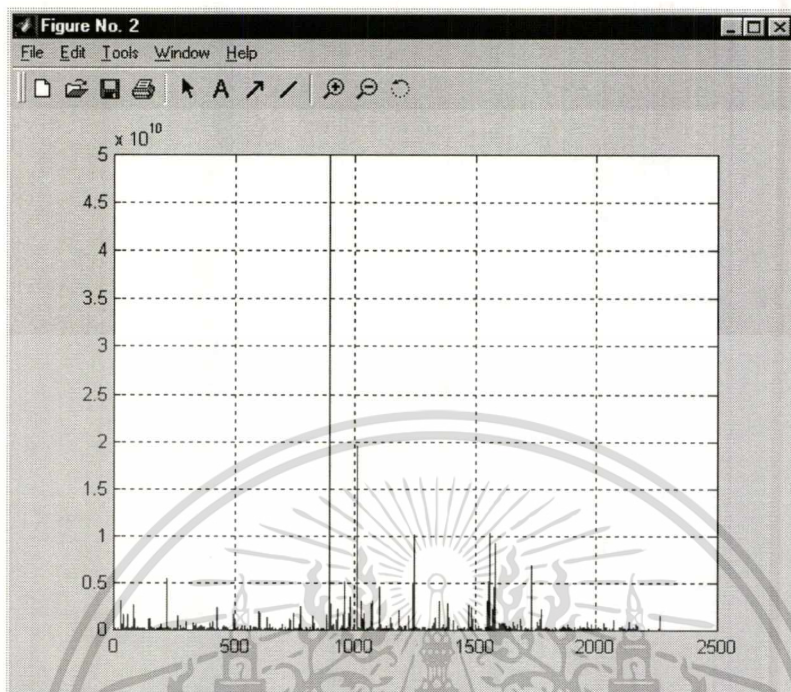
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



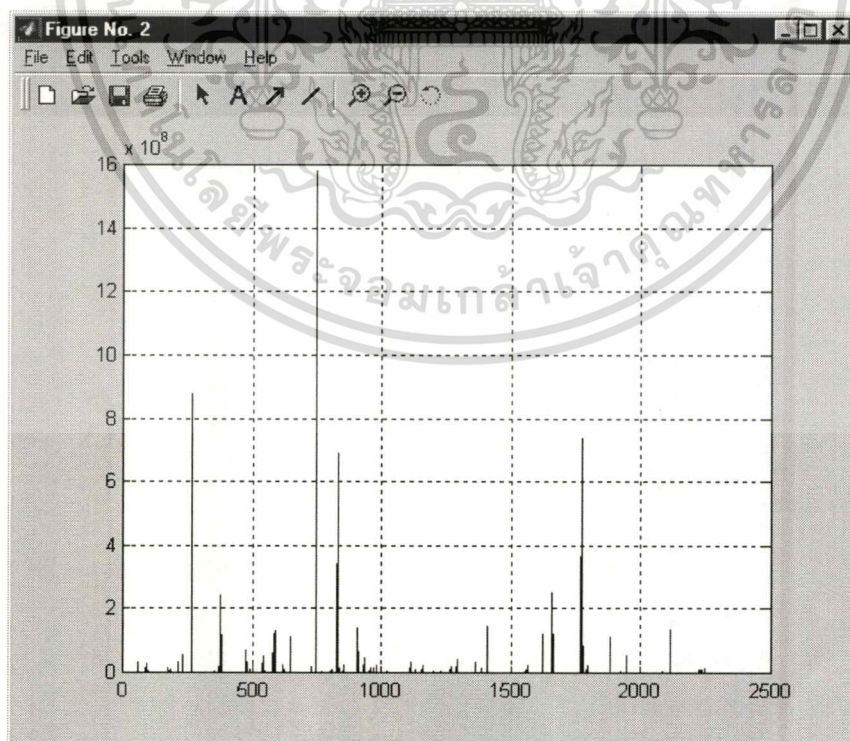
รูปที่ 4.11 แสดงข้อมูล Attribute รายจ่ายในการดำเนินงานที่ใช้ในการ Train



เอกสารนี้เป็นเอกสารรูปที่ 4.12 แสดงข้อมูล Attribute กำไร/ขาดทุนสุทธิที่ใช้ในการ Train ใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

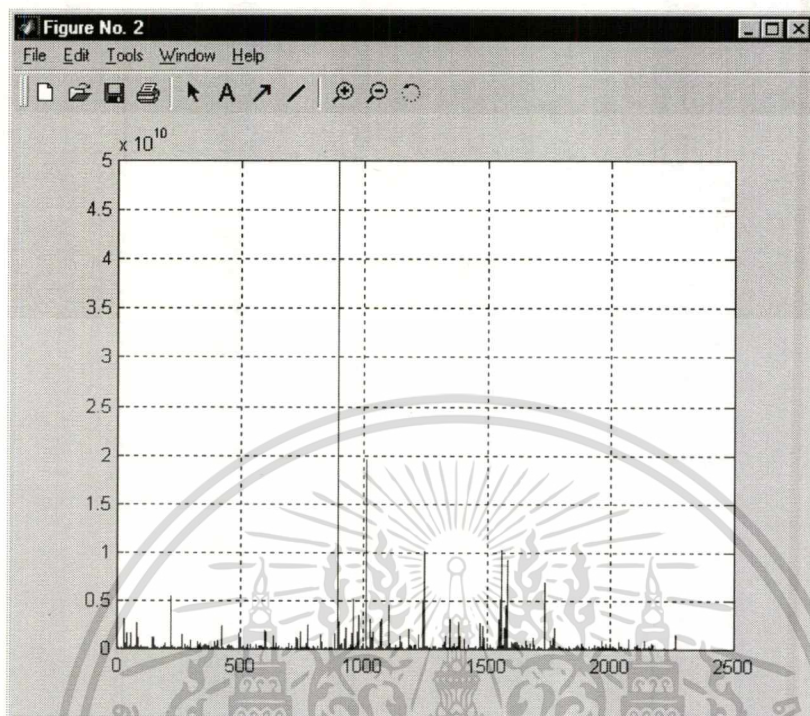


รูปที่ 4.13 แสดงข้อมูล Attribute ต้นทุนการผลิตที่ใช้ในการ Train

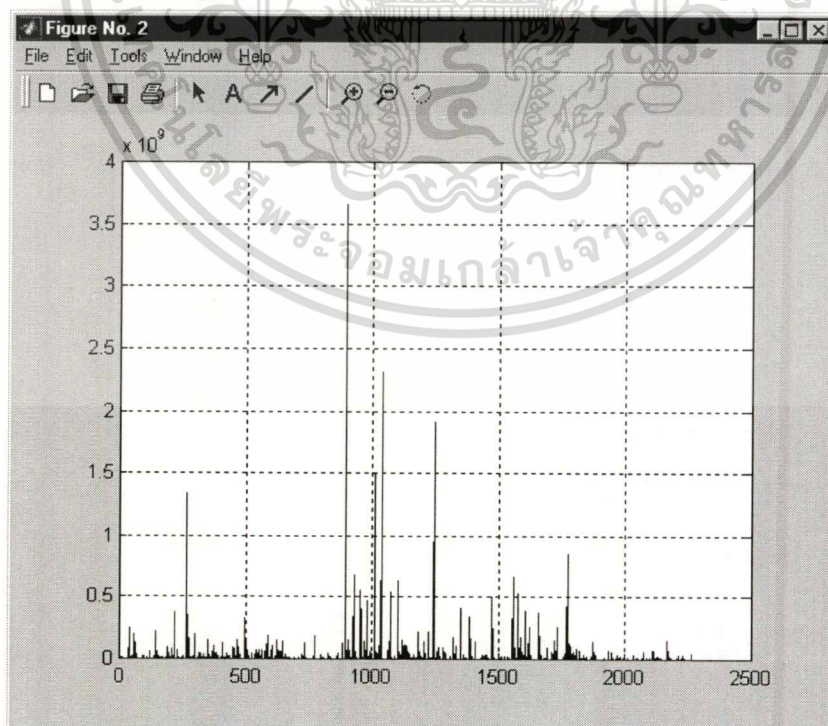


รูปที่ 4.14 แสดงข้อมูล Attribute รวมสินทรัพย์ที่ใช้ในการ Train

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

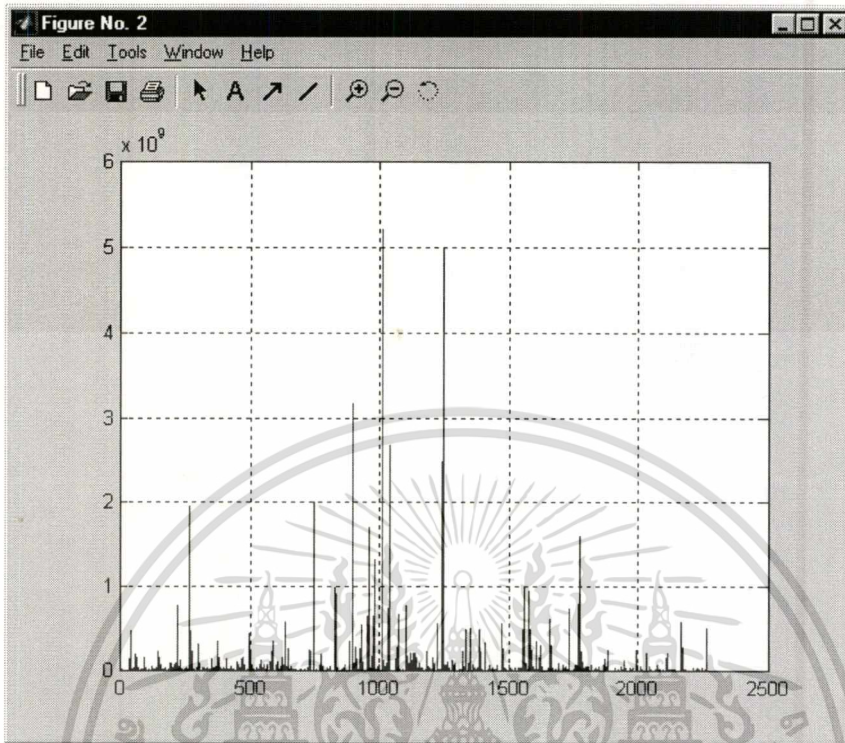


รูปที่ 4.15 แสดงข้อมูล Attribute รวมหนี้สินและทุนที่ใช้ในการ Train

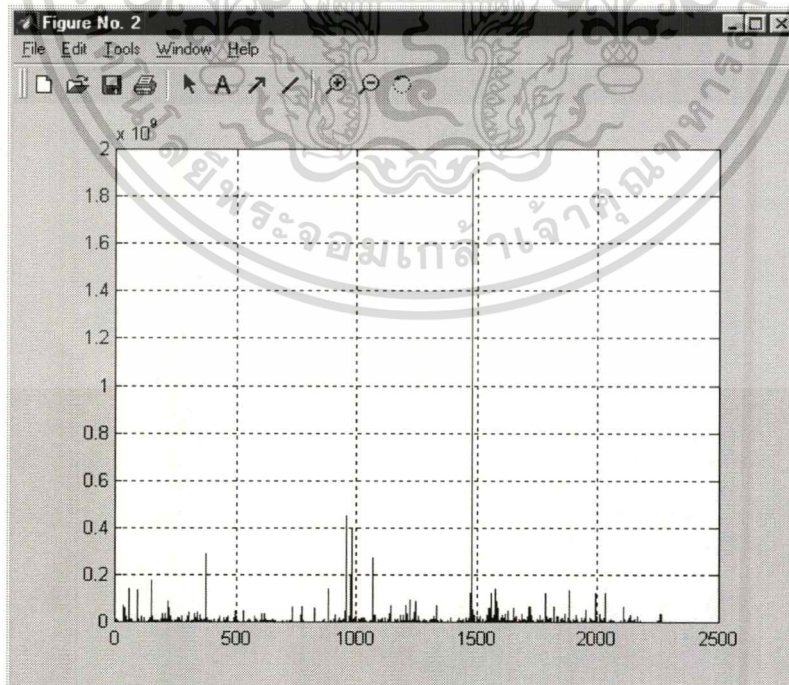


รูปที่ 4.16 แสดงข้อมูล Attribute ยอดรวมรายได้ที่ใช้ในการ Train

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการศึกษาเท่านั้น ไม่อนุญาตให้ไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

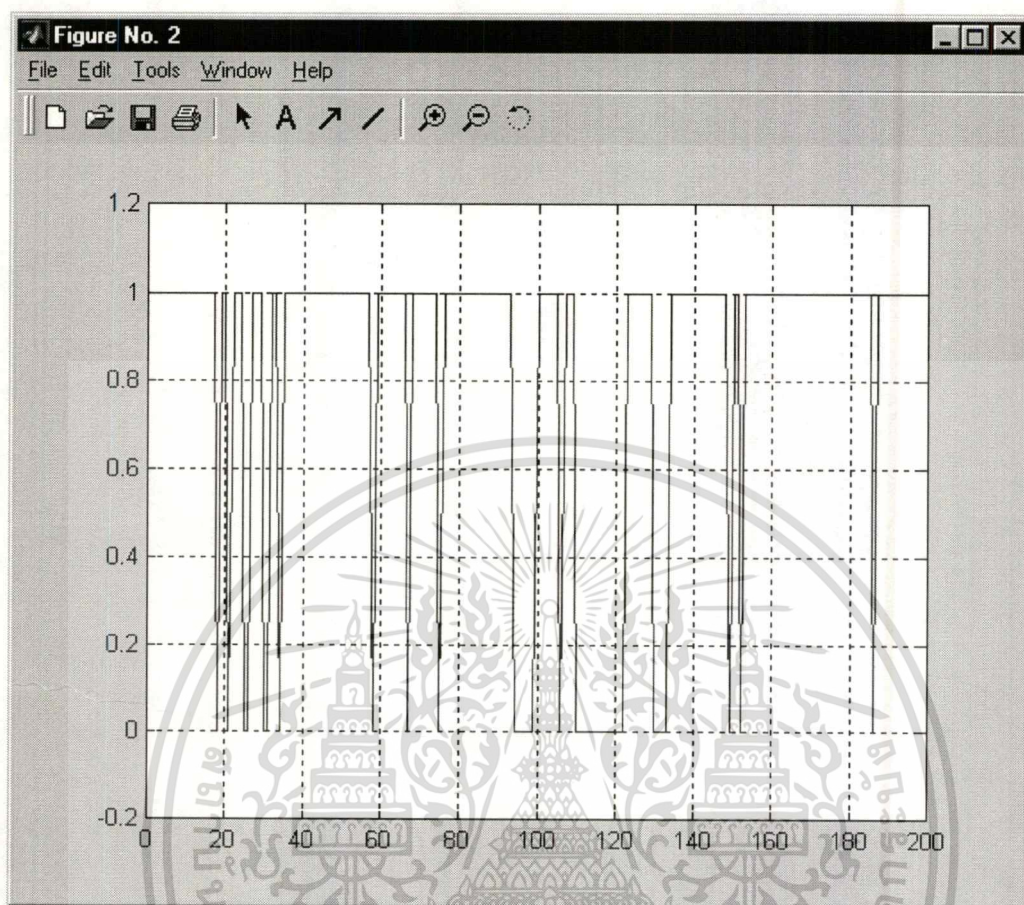


รูปที่ 4.17 แสดงข้อมูล Attribute ยอดรวมรายจ่ายที่ใช้ในการ Train



รูปที่ 4.18 แสดงข้อมูล Attribute กำไร/ขาดทุนสุทธิ(ภ.ง.ด.51)ที่ใช้ในการ Train

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.19 การแสดงผลพัลส์ในรูปของกราฟโดยโปรแกรม MATLAB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### บทสรุปและข้อเสนอแนะ

#### 5.1 สรุปผลการพัฒนาระบบ

จากการพัฒนาระบบครั้งนี้ การนำเทคนิคของคาค่าไมนิ่งมาใช้ในด้านการตรวจสอบของกรมสรรพากรสามารถกระทำได้อย่างดี เนื่องจาก กรมสรรพากรมีการจัดเก็บข้อมูลต่าง ๆ ในรูปแบบฐานข้อมูลอยู่แล้วจึงเป็นการง่ายที่จะนำเทคนิคคาค่าไมนิ่งมาใช้ในการวิเคราะห์ข้อมูล อีกทั้งอุปกรณ์ต่าง ๆ ไม่ว่าจะเป็นฮาร์ดแวร์และซอฟต์แวร์ที่มีอยู่ก็สามารถสนับสนุนเทคนิคดังกล่าวด้วย

#### 5.2 ข้อเสนอแนะ

จากผลของการพัฒนาระบบ สามารถยอมรับได้ว่าสามารถใช้งานได้ดีในระดับหนึ่ง เนื่องจากการพัฒนาระบบเป็นการทดลองกับข้อมูลจำนวนไม่มากนัก(ข้อมูลที่ใช้ Train จำนวน 2,262 เรคคอร์ด ส่วนข้อมูลที่ใช้ในการ Test จำนวน 200 เรคคอร์ด ) การกำหนดจำนวนโหนด อัตราการเรียนรู้ และจำนวนรอบในการเรียนรู้ของโครงข่ายประสาทเทียมก็เกิดจากการทดลองปรับเปลี่ยนโดยผู้เขียนเอง ซึ่งเมื่อนำไปใช้งานจริงอาจจะพบกับปัญหาของจำนวนของข้อมูลซึ่งมีเป็นจำนวนมาก การกำหนดจำนวนโหนด อัตราการเรียนรู้ และจำนวนรอบในการเรียนรู้ของโครงข่ายประสาทเทียมก็ต้องทำการทดลองและปรับเปลี่ยนให้เหมาะสมกับจำนวนข้อมูลนั้น ๆ ด้วย อีกทั้งอาจจะต้องพบกับปัญหาการเรียนรู้ของโครงข่ายที่ต้องใช้เวลาในการเรียนรู้เพิ่มมากขึ้นด้วย

## บรรณานุกรม

ฝ่ายวิเคราะห์ระบบและเอกสารภาษี กองนโยบาย กรมสรรพากร. 2533. **ภาษีเงินได้นิติบุคคล.**  
กรุงเทพฯ : ไอเดียสแควร์.

สาธิต อินทจักร์ และ กิตติ ไพฑูรย์วัฒนกิจ. 2540. “นิรอลเน็ตเวิร์ก.” สาร NECTEC. 26-42

Alex Berson and Stephen J. Smith. 1997. **Data Warehouse, Data Mining, & OLAP.** McGraw Hill.

Dan W. Patterson. 1996. **Artificial Neural Network.** Singapore : Prentice Hall.

Matin T. Hagan and Howard B. Demuth University of Idaho. 1996. **Neural Network.** PWS Publishing Company.

Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees, Alessandro Zanasi. 1998.  
**Discovering Data Mining: From Concept to Implementation.** Nerjersey : Prentice Hall.

Ruby L. Kennedy, Yuchun Lee, Benjamin Van Roy, Christopher D. Reed and Richard P. Loppmann. 1995-1997. **Solving Data Mining Problem through Patten recognition.** Nerjersey : Prentice Hall.

Sholum M. Weiss and Nith Indurkhya. 1998. **Predictive Data Mining.** San Francisco : Mogan Kaufmann.

“Neural net.” 1999. [Online]. Available: [http://www.shef.ac.uk/psychology/notes/l2/section3\\_1-.html](http://www.shef.ac.uk/psychology/notes/l2/section3_1-.html).

“Software Computing 2.” 1999. [Online]. Available: <http://www2.psy.uq.edu.au/cogPsychlic-210/backprop/sld016.html>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ประวัติผู้เขียน

ชื่อผู้เขียน	นางสาวนุชชา พรหมภิบาล
วุฒิการศึกษาระดับปริญญาตรี	วิทยาศาสตร์บัณฑิต
สถานที่สำเร็จการศึกษา	สถาบันราชภัฏอุตรดิตถ์
ปีที่สำเร็จการศึกษา	ปีการศึกษา 2540
อาชีพปัจจุบัน	รับราชการตำแหน่ง นักวิชาการคอมพิวเตอร์ 4 สำนักเทคโนโลยีสารสนเทศ กรมสรรพากร



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้