ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

การจัดการวัตถุดิบสำหรับการผลิตหินเจียรโดยใช้ดาตาไมน์นิ่ง

The using of data mining method for manage material of Grinding Wheel Production

โดย

เชษฐพล รัตนานนท์

รหัสประจำตัว 43067003

อาจารย์ที่ปรึกษา

ผศ.ดร. วรพจน์ กรีสุระเดช

H002305

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 2 ปีการศึกษา 2547

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

b 11704 0x
i1 24424 5x

วัน เดือน ปี 21 ก.พ. 2550
เลขทะเบียน 02305
เลขเรียกหนังสือ วพ ช 7221 2547
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การจัดการวัตถุดิบสำหรับการผลิตหินเจียรโดยใช้ดาตาไมน์นิ่ง
นักศึกษา	เชษฐพล รัตนานนท์
อาจารย์ที่ปรึกษา	ผศ.ดร.วรพจน์ กรีสุระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2547

บทคัดย่อ

สำหรับอุตสาหกรรมที่ต้องการขอใช้สิทธินำเข้าวัตถุดิบโดยยกเว้นภาษีจาก BOI จะต้องแจ้งรายการวัตถุดิบที่จะนำเข้า และสูตรหรือสัดส่วนการใช้วัตถุดิบในการผลิตสินค้าชนิดต่าง ๆ ให้กับ BOI เพื่อนำไปใช้ควบคุมการจัดเก็บวัตถุดิบ ซึ่งสูตรที่แจ้งดังกล่าวเหมือนเป็นตัวแทนของสูตรที่ใช้จริง จะต้องครอบคลุม ใช้งานได้นาน ตลอดจนสามารถพยากรณ์ค่าในอนาคตได้ ซึ่งในปัจจุบันหินเจียรที่ผลิตจากโรงงานได้ถูกนำไปใช้ในอุตสาหกรรมต่าง ๆ มากขึ้น ลูกค้าแต่ละรายต้องการหินเจียรในรูปแบบและส่วนผสมต่าง ๆ กัน นั่นคือ ทำให้สูตรการใช้วัตถุดิบเพิ่มขึ้นเป็นจำนวนมาก เมื่อมีการนำเข้าวัตถุดิบสำหรับสูตรที่เพิ่มขึ้น จะไม่สามารถใช้สิทธิยกเว้นภาษีได้ หรือใช้สิทธิจากสูตรที่เคยแจ้ง ก็จะมีความคลาดเคลื่อนได้

ในการพัฒนาระบบงานนี้ จะนำกระบวนการทาง data mining มาใช้ในการวางแผนการใช้สิทธินำเข้าวัตถุดิบ โดยจะนำสูตรการใช้วัตถุดิบที่มีทั้งหมดมาจัดกลุ่มโดยกระบวนการ segmentation ตามอัลกอริทึมของ K-mean เพื่อให้ได้สูตรมาตรฐานกลางที่จะนำไปแจ้งกับ BOI และ classify สูตรที่จะเกิดขึ้นมาใหม่ในอนาคตให้เข้าอยู่ในกลุ่มที่ถูกต้องโดยใช้วิธีเดียวกัน ตลอดจนพยากรณ์ยอดการใช้งานในแต่ละสูตรเพื่อยื่นขอนำเข้าวัตถุดิบได้ใกล้เคียงกับความเป็นจริงมากขึ้นโดยใช้วิธีของ Back propagation

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	The using of data mining method for manage material of Grinding Wheel Production
Student	Mr. Chettapol Rattananonta
Advisor	Asst. Prof. Dr. Worapoj Kreesuradej
Level of Study	Master of Science in Information Technology
Major	Information Science
Academic	2004

Abstract

Every BOI Privilege of Import Raw Material industries must inform about List of Import material and Production Formula (mixed ratio of raw material) of every products to BOI for control stock of import and export . These formulas must long term, and forecast the future value. Currently, Grinding Wheel is used in more distinct industries, each Customer need product in more distinct formula. Therefore, The New formula will not use BOI Privilege or will be wrong if use the Old inform formula.

In this project will be use data mining method for planning about raw material Import. For example, Will use segmentation technical (K-Mean method) for clustering standard groups of exist formulas for inform to BOI and use back propagation method for classify new formula to these groups. Finally wile use linear regression method for forecast the Import value in the future.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ..... II

สารบัญ..... III

บทที่

1. บทนำ..... 1

    1.1. แนวคิดและความเป็นมา..... 1

    1.2. วัตถุประสงค์ของการศึกษา 2

    1.3. ขอบเขตของการศึกษา 2

    1.4. ขั้นตอนและวิธีการดำเนินงาน 2

    1.5. ผลที่คาดว่าจะได้รับ 3

2. ทฤษฎีที่เกี่ยวข้อง 4

    2.1. ความหมายของ Data Mining 4

    2.2. การทำ Data Mining 4

    2.3. เทคนิคการทำ Data Mining 6

    2.4. การจัดกลุ่มข้อมูลโดยใช้วิธี k-mean 8

    2.5. การพยากรณ์โดยใช้โครงข่ายประสาทเทียม 9

    2.6. โครงข่าย Back propagation 11

    2.7. ข้อดีของ Neural Network 14

    2.8. ข้อจำกัดของ Neural Network 14

    2.9. สรุป 14

3. การออกแบบระบบสำหรับการจัดการวัตถุดิบ 15

    3.1. การออกแบบระบบจัดกลุ่มสูตรหินเจียรโดยใช้ Database Segmentation 15

    3.2. การออกแบบระบบพยากรณ์ยอดการผลิตหินเจียรโดยใช้
    โครงข่ายประสาทเทียม 19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4. วิธีการใช้งานระบบ 22
   4.1. เริ่มต้นใช้งานโปรแกรม 22
   4.2. Cleaning Data Process 23
   4.3. Clustering Process 26
   4.4. Prediction Process 28
   4.5. Summary 32
5. สรุป 36
   5.1. สรุปผลการจัดกลุ่ม 36
   5.2. สรุปผลการพยากรณ์ 38

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 แนวคิดและความเป็นมา

ในการใช้สิทธินำเข้าวัตถุดิบเพื่อใช้ผลิตสินค้าสำหรับส่งออกโดยยกเว้นภาษีจากสำนักงานคณะกรรมการส่งเสริมการลงทุน (BOI) ของผู้ขอใช้สิทธิต่าง ๆ ในปัจจุบันนั้น ผู้ขอใช้สิทธิจะต้องยื่นแจ้งรายละเอียดเกี่ยวกับสูตรหรือสัดส่วนการใช้วัตถุดิบในการผลิตสินค้าชนิดต่าง ๆ ให้กับ BOI เพื่อเป็นข้อมูลในการควบคุมการนำเข้า การจัดเก็บ และการส่งออกวัตถุดิบ กล่าวคือ เมื่อผู้ใช้สิทธิต้องการจะนำเข้าวัตถุดิบเพื่อผลิตสินค้าสำหรับส่งออกโดยยกเว้นภาษีทุกครั้ง จะต้องยื่นจำนวนวัตถุดิบโดยอิงตามสูตรที่ยื่นแจ้งไป และเมื่อมีการส่งสินค้าออก ก็จะต้องยื่นตัดบัญชีวัตถุดิบ ตามสูตรที่แจ้งเช่นกัน ดังนั้นในกรณีที่ผู้ขอใช้สิทธิมีสูตรวัตถุดิบที่แน่นอนตายตัว และไม่มากนัก ก็จะไม่มีปัญหาในการยื่นแจ้งสูตร แต่สำหรับโรงงานที่มีสูตรการผลิตที่หลากหลาย และเพิ่มขึ้นได้ตลอดเวลา สูตรที่จะแจ้งดังกล่าวจะต้องครอบคลุม ใช้งานได้นาน ตลอดจนสามารถพยากรณ์ค่าในอนาคตได้อย่างใกล้เคียง ซึ่งในปัจจุบันโรงงานหินเจียรอุตสาหกรรมเป็นไปในลักษณะดังกล่าว ลูกค้าแต่ละรายต้องการหินเจียรในรูปแบบและส่วนผสมต่าง ๆ กัน นั่นคือ ทำให้สูตรการใช้วัตถุดิบเพิ่มขึ้นเป็นจำนวนมาก ถ้าสูตรที่แจ้งไปไม่ครอบคลุมพอ การควบคุมวัตถุดิบก็จะมีความคลาดเคลื่อนได้

ในการผลิตหินเจียรของโรงงาน จะต้องใช้วัตถุดิบแบ่งเป็น 2 ส่วนคือ

1. เม็ดหิน (Grain) เป็นองค์ประกอบหลัก ซึ่งการใช้งานในปัจจุบันแบ่งตามชนิดแร่ธาตุได้ 2 กลุ่ม คือ Aluminium Oxide และ Silicon Carbide ซึ่งมีขนาดแตกต่างกันตั้งแต่ 1 – 6000 mm และในการยื่นจดกับ BOI จะแบ่งเป็น 2 ขนาด คือ ไม่เกิน 240 mm และตั้งแต่ 240 mm ขึ้นไป
2. สารประกอบอื่น ๆ ที่ใช้ในการผสม เช่น กาว แป้ง หรือสารเร่งปฏิกิริยา ฯลฯ ซึ่งเกิดจากการนำสารประมาณ 40 ชนิด มาผสมกันในอัตราส่วนที่แตกต่างกันไป

เมื่อมีการสั่ง Order หินเจียรจากลูกค้า ซึ่งจะระบุรูปแบบ, ขนาด , สี ความคงทน หรือ grain หลักที่ต้องการมาให้ ทางโรงงานจะกำหนดส่วนผสมของวัตถุดิบเพื่อให้ได้ผลิตภัณฑ์ตามที่ต้องการ จากวัตถุดิบทั้งสองส่วนข้างต้น ซึ่งในปัจจุบัน ทางโรงงานมีสูตรการผลิตกว่า 10,000 สูตร ซึ่งรองรับความต้องการของลูกค้าที่แตกต่างกันและจะมีเพิ่มขึ้นเรื่อย ๆ ในแต่ละวันที่รับ order จากลูกค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในปัจจุบัน ทางโรงงานได้เคยยื่นแจ้งสูตรให้กับ BOI เพียง 5 สูตร ซึ่งใช้มาตั้งแต่ตอนที่สูตรผลิตภัณฑ์ยังไม่มาก โดย 5 สูตรนี้จะเป็นการจัดกลุ่มสูตรที่มีอยู่ตามวิธีการผลิต , ชนิดของหิน และขนาดของผลิตภัณฑ์เมื่อผลิตเสร็จ และเมื่อจะมีการนำเข้าวัตถุดิบโดยใช้สิทธิยกเว้นภาษีนำเข้าของ BOI ทางโรงงานจะต้องประมาณยอดนำเข้าวัตถุดิบแต่ละชนิดเพื่อใช้ในการผลิตเพื่อส่งออกในช่วงเวลา 6 เดือน โดยจะต้องอ้างอิงจาก 5 สูตรดังกล่าว ซึ่งถ้ายอดที่ตั้งไว้สูงจากยอดที่ใช้งานจริงมากเกินไปก็อาจจะต้องเสียภาษีส่วนที่ใช้ไม่หมดเมื่อโรงงานมีการเลิกกิจการหรือสิ้นสุดการใช้สิทธิ หรือถ้าประมาณไว้ต่ำจากความเป็นจริง ก็จะต้องเสียเวลาในการยื่นใช้สิทธิเพื่อสั่งนำเข้าเพิ่มหรือเสียเวลาในการรอวัตถุดิบที่อาจจะมาไม่ทันกำหนด ทำให้มีข้อยุ่งยากอื่น ๆ ตามมาภายหลัง

1.2 วัตถุประสงค์ของการศึกษา

ในการพัฒนาระบบงานนี้ จะนำกระบวนการทาง data mining มาใช้ในการวางแผนการใช้สิทธินำเข้าวัตถุดิบ โดยจะนำสูตรการใช้วัตถุดิบที่มีทั้งหมดมาจัดกลุ่มโดยกระบวนการ segmentation เพื่อให้ได้สูตรมาตรฐานกลางสำหรับยื่นแจ้งแก่ BOI และ classify สูตรที่จะเกิดขึ้นมาใหม่ในอนาคตให้เข้าอยู่ในกลุ่มที่ถูกต้อง ตลอดจนพยากรณ์ยอดการใช้งานในแต่ละสูตรเพื่อยื่นขอนำเข้าวัตถุดิบได้ใกล้เคียงกับความเป็นจริงมากขึ้น

1.3 ของเขตของการศึกษา

การศึกษาโครงการพัฒนาระบบงานนี้มีขอบเขตของการศึกษาคือ เป็นการพัฒนาโปรแกรมช่วยตัดสินใจในการวางแผนการนำเข้าวัตถุดิบ โดยโปรแกรมดังกล่าว จะต้องจัดกลุ่มสูตรวัตถุดิบของโรงงานทั้งหมด ให้ได้ไม่เกิน 20 กลุ่ม โดยใช้หลักการของ K-Mean algorithm เพื่อให้ได้สูตรมาตรฐานกลางที่จะนำไปแจ้งกับ BOI และรองรับการ classify สูตรที่จะเกิดขึ้นมาใหม่ในอนาคตให้เข้าอยู่ในกลุ่มที่ถูกต้อง (มีการใช้วัตถุดิบในอัตราส่วนที่ใกล้เคียงที่สุด) ตลอดจนสามารถพยากรณ์ยอดการใช้งานในแต่ละสูตรจากข้อมูลที่ผ่าน ๆ มา โดยใช้วิธีของ back propagation Neural Network

1.4 ขั้นตอนและวิธีการดำเนินงาน

การศึกษาโครงการพัฒนาระบบงานนี้มีขั้นตอนและวิธีการดำเนินงานคือ

1. ศึกษาหลักการและเทคนิคของ Data mining ที่จำเป็นต้องใช้ เช่น K-Mean algorithm , Back propagation Neural Network
2. รวบรวมข้อมูลของวัตถุดิบที่มีอยู่เดิม
3. ออกแบบและพัฒนาระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. ทดสอบการใช้งานของระบบ
5. สรุปผลการศึกษาและการดำเนินงาน

1.5 ผลที่คาดว่าจะได้รับ

การศึกษาโครงการพัฒนาระบบงานนี้มีผลที่คาดว่าจะได้รับคือ

1. เพื่อเข้าใจหลักการและเทคนิคของ Data mining
2. เพื่อนำระบบงานนี้ไปช่วยในการตัดสินใจในการนำเข้าวัตถุดิบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

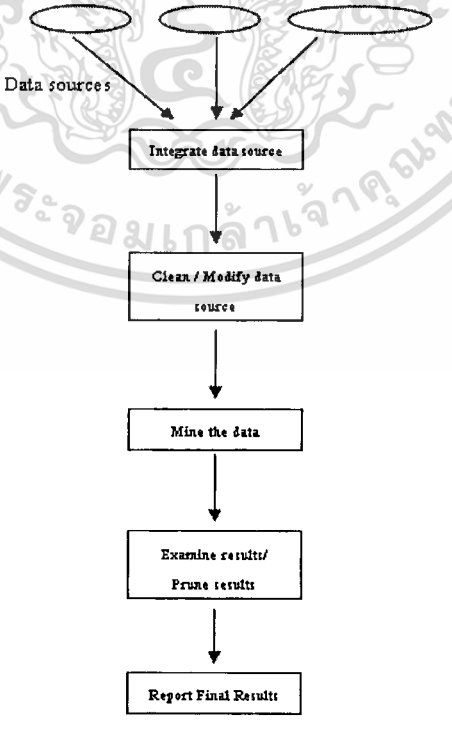
2.1 ความหมายของ Data mining

data mining เป็นกระบวนการที่เป็นการนำเอาข้อมูลที่ซ่อนอยู่ภายใต้ฐานข้อมูลขนาดใหญ่ที่เป็นข้อมูลที่ไม่ทราบมาก่อน มาทำให้เกิดประโยชน์ในการตัดสินใจเลือกแนวทางในการดำเนินธุรกิจ

ข้อมูลที่ได้มาจากการทำ Mining ไม่ได้เกิดจากสมมุติฐาน หรือจากการคาดคะเนจากประสบการณ์ แต่เป็นข้อมูลหรือความสัมพันธ์ที่เกิดขึ้นจริงที่ซ่อนอยู่ภายใต้ข้อมูลที่เรามีอยู่ ดังนั้นการทำ data mining มักจะไม่ใช้การตั้งสมมุติฐานแต่จะเป็นการดูผลลัพธ์ที่ได้จากการทำ mining เลยซึ่งจะแตกต่างจากวิธีการวิเคราะห์ข้อมูลทางสถิติแบบอื่น

2.2 การทำ Data mining

ประกอบด้วยกระบวนการต่างๆ แบ่งออกได้เป็นขั้นตอนใหญ่ๆ ได้ 5 ขั้นตอนดังรูป



รูปที่ 2.1 ขั้นตอนการทำ Data mining

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.1 ขั้นตอนกำหนดวัตถุประสงค์และแหล่งข้อมูลของการทำ Data mining

เป็นขั้นตอนส่วนสำคัญที่จะกำหนดถึง ความต้องการหรือปัญหา ที่ต้องการทราบ ซึ่งมักเป็นความต้องการที่มุ่งเพื่อนำคำตอบที่ได้ มาใช้เพื่อให้เกิดประโยชน์ แต่จะไม่ใช่เกิดจากการตั้งสมมุติฐาน และเป็นการกำหนดถึงแหล่งที่มาของข้อมูลที่จะทำการ mining ซึ่งก็เป็นแหล่งข้อมูลที่คาดว่าจะได้คำตอบจากสิ่งที่ต้องการทราบ

2.2.2 ขั้นตอนการจัดเตรียมข้อมูลที่จะทำการ mining

เป็นขั้นตอนที่ต้องใช้เวลามากที่สุด เนื่องจากต้องมีการพิจารณาข้อมูลในแทบจะทุกเรื่องเช่น ความเกี่ยวข้องของข้อมูลกับวัตถุประสงค์ของการทำ ชนิดของข้อมูล การตรวจสอบข้อมูลว่าเป็นข้อมูลที่เหมาะสมหรือไม่ ทั้งนี้ต้องคำนึงถึงอายุของข้อมูลด้วย โดยอาจต้องมีการกำจัดข้อมูลที่ไม่จำเป็นหรือไม่ถูกต้องออกไป รวมทั้งเป็นการเตรียมข้อมูลให้พร้อมที่จะทำการ mining โดยการปรับเปลี่ยนรูปแบบของข้อมูลเพื่อให้เหมาะสมกับอัลกอริทึมที่จะเลือกใช้ ซึ่งเป็นการจัดเพื่อให้การ mining ทำไปได้อย่างมีประสิทธิภาพ

2.2.3 ขั้นตอนการทำ Data Mining

ขั้นตอนการทำ Data Mining ถือว่าเป็นหัวใจหลักของการทำ Data Mining เพราะการเลือกเอาวิธีการและกระบวนการอัลกอริทึมที่ถูกต้องจะทำให้การทำ Mining ได้ผลอย่างรวดเร็วและถูกต้องตามจุดประสงค์ที่ต้องการ

2.2.4 ขั้นตอนการประเมินผลที่ได้จากการทำ Data Mining

เป็นเสมือนขั้นตอนการอธิบายและประเมินถึงผลที่ได้จากการทำ mining ว่าสามารถนำมาใช้ให้บรรลุถึงจุดประสงค์ที่ต้องการหรือไม่ รวมทั้งเป็นการประเมินถึงความถูกต้องของผลที่ได้จากการทำ ซึ่งก็นับว่าเป็นสิ่งสำคัญอย่างหนึ่งเช่นกันเพราะบางครั้งผลที่ได้จากการทำ mining อาจมีข้อผิดพลาดซึ่งอาจเกิดได้จากหลายสาเหตุที่เราอาจคาดไม่ถึง จึงต้องมีการตรวจสอบผลที่ได้

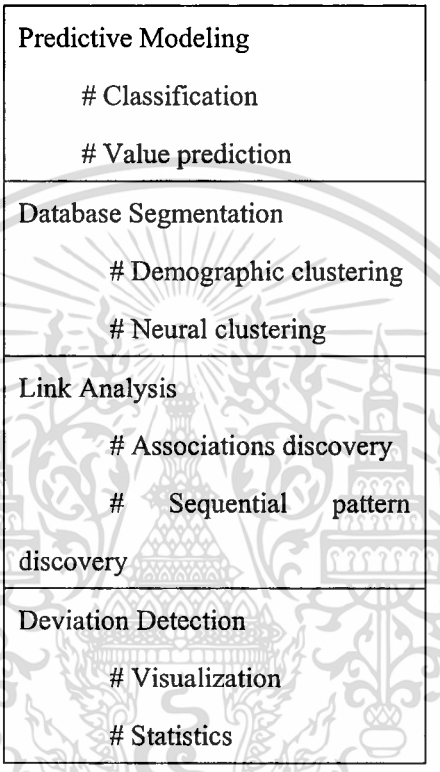
2.2.5 ขั้นตอนนำเสนอความรู้ที่ได้

การนำเสนอความรู้ได้เป็นขั้นตอนสุดท้ายของกระบวนการทั้งหมด เป็นการนำเสนอถึงผลที่ได้จากการทำ Data Mining และนำเสนอถึงวิธีการที่จะนำผลที่ได้นี้ไปใช้ให้เกิดประโยชน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 เทคนิคการทำ Data Mining

Data Mining มีเทคนิคและอัลกอริทึมที่สามารถนำมาใช้งานอยู่หลายประเภท ขึ้นอยู่กับรูปแบบ Application ที่ต้องการนำมาใช้งาน แต่สามารถ แบ่งออกเป็น รูปแบบต่างๆได้ดังรูป



รูปที่ 2.2 กระบวนการและเทคนิคต่างๆของ Data Mining

2.3.1 Predictive Modeling

เป็นการคาดคะเนทำนายถึงความเป็นไปได้ โดยใช้การสังเกตจากรูปแบบของข้อมูลที่มีอยู่ โดยมีลักษณะเป็นการเรียนรู้จากกลุ่มข้อมูลที่ได้กำหนดไว้ แล้วจึงนำไปวิเคราะห์กลุ่มข้อมูลที่ต้องการ จึงเป็นลักษณะแบบ Supervised learning จึงมีรูปแบบการพัฒนาใน 2 ช่วงคือ

- ช่วงการเรียนรู้ (Training Phase) เป็นการสร้างโมเดลโดยการใช้ข้อมูลในอดีต และมีจำนวนข้อมูลจำนวนมาก
- ช่วงการทดสอบ (Testing Phase) เป็นการตรวจสอบความน่าเชื่อถือและประสิทธิภาพของโมเดลที่สร้างขึ้น จึงเป็นข้อมูลที่มีจำนวนไม่มากนัก

Predictive modeling ยังสามารถแบ่งย่อยได้อีก เป็น 2 เทคนิค คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. Classification ซึ่งเป็นการแบ่งกลุ่มของข้อมูลตามชนิดของกลุ่มข้อมูลที่ควรจะเป็น สามารถแบ่งกลุ่มข้อมูลได้อย่างชัดเจน ซึ่งมีอัลกอริทึมที่นิยมคือ Tree Induction และ Neural Induction
   - Tree Induction คือการนำเอาข้อมูลมาสร้างแบบจำลองในรูปแบบของ Decision Tree ซึ่ง Decision Tree นั้นมีการทำงานแบบ Supervised Learning คือสามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่ได้ไว้ก่อนล่วงหน้าที่เรียกว่า Training Set ได้โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วยรูปของ Tree ซึ่งประกอบไปด้วย root node, node และ leaf node
   - Neural Induction คือเทคนิคที่นำเสนอแบบจำลองที่มีโครงสร้างประกอบด้วย Node และค่าถ่วงน้ำหนักของแต่ละ link ที่เชื่อมระหว่าง Node ซึ่งมีพื้นฐานการทำงานแบบ Neural Network ซึ่งเป็นการสร้างแบบจำลองที่เลียนแบบการทำงานของสมองมนุษย์ มีโครงสร้างเป็นกลุ่มของ Node ที่เชื่อมโยงถึงกัน
2. Value prediction เป็นการทำนายถึง ค่าความต่อเนื่องของข้อมูล เป็นการทำนายค่าที่เป็นตัวเลข โดยมีเทคนิคที่นิยมใช้คือ Linear regression และ Nonlinear regression

2.3.2 Database Segmentation

Segmentation หรือ Clustering เป็นการทำการแบ่งกลุ่มย่อยข้อมูลเพื่อทำการแยกออกให้ทราบว่าข้อมูลชุดนี้มีทั้งหมดกี่กลุ่ม ซึ่งการแบ่งกลุ่มข้อมูลนี้ไม่สามารถกำหนดได้ว่าข้อมูลนี้ควรจะอยู่กลุ่มใด แต่เป็นการกำหนดกลุ่มของข้อมูลจากธรรมชาติของข้อมูลเองไม่ได้ใช้ความรู้สึกหรือประสบการณ์ในการตัดสินใจ แบ่งกลุ่มข้อมูลแต่จะจัดการโดยอัลกอริทึมที่เหมาะสมของแต่ละกลุ่มข้อมูล จึงเป็นการเรียกว่าเป็นรูปแบบของ unsupervised learning ซึ่งก็มีเทคนิคย่อยอีก ซึ่งที่นิยมใช้คือ Demographic clustering และ Neural clustering

2.3.3 Link Analysis

เป็นการศึกษาความสัมพันธ์ของข้อมูลหรือกลุ่มของข้อมูลว่ามีความสัมพันธ์กันในรูปแบบลักษณะใด โดยเรียกความสัมพันธ์นี้ว่าเป็น “ Association” เป็นโมเดลที่นิยมมากในการวิเคราะห์เพื่อหาความสัมพันธ์ระหว่างลูกค้ากับสินค้าหรือบริการ สามารถแบ่งย่อยได้อีก 3 ลักษณะตามการวิเคราะห์ข้อมูลคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. Association discovery เป็นการวิเคราะห์ข้อมูลที่เกิดขึ้นพร้อมกันภายในกลุ่มข้อมูลเดียวกัน เป็นเทคนิคที่นิยมมากชนิดหนึ่ง มักใช้ในการวิเคราะห์ถึงพฤติกรรมการซื้อของผู้บริโภค จึงเป็นเทคนิคที่มีอีกชื่อหนึ่งว่า Market basket analysis
2. Sequential pattern discovery เป็นการศึกษาความสัมพันธ์ระหว่างข้อมูล โดยเทียบข้อมูลกับเวลา ซึ่งเป็นการศึกษาพฤติกรรมในระยะยาว
3. Similar time sequence discovery เป็นการศึกษาพฤติกรรมของข้อมูลที่เกิดขึ้นทั้งหมดหรือเกิดขึ้นในช่วงเวลาเดียวกันเพื่อหาความสัมพันธ์ระหว่างกลุ่มของข้อมูลเหล่านี้

2.3.4 Deviation Detection

เป็นโมเดลที่จะใช้เทคนิคทางสถิติ และการทำให้เห็นภาพ (Visualization) ซึ่งเป็นรูปแบบการสรุปข้อมูลให้ออกมาในรูปแบบการแสดงผลแบบกราฟฟิก เช่นการใช้ graph

2.4 การจัดกลุ่มข้อมูลโดยใช้วิธี k-mean

กำหนดให้มีออบเจ็ค n รายการ และ มีการจัดกลุ่ม k กลุ่ม ในการแบ่งกลุ่มออบเจ็คดังกล่าวเข้าไปไว้ในแต่ละกลุ่ม ซึ่งจะเรียกกลุ่มว่า Cluster โดย Cluster เป็นเหมือนขอบเขตที่ถูกสร้างบนสิ่งที่เรียกว่า similarity function ที่ใช้พิจารณาเช่น ระยะห่างระหว่างออบเจ็ค เป็นต้น โดย ออบเจ็คที่อยู่ใน cluster เดียวกันจะเรียกว่า “similar” ในขณะที่ออบเจ็คต่าง cluster กัน เรียกว่า “dissimilar” ในเทอมของแอตริบิวของออบเจ็ค

ในอัลกอริทึม k-mean จะต้องระบุพารามิเตอร์ k แทนจำนวนกลุ่มสำหรับกำหนดให้ออบเจ็ค n จำนวน โดยค่าของ Cluster similarity วัดจากระยะห่างจากออบเจ็คถึงค่ากลาง (mean) ของกลุ่ม ซึ่งค่ากลางนี้ เรียกว่า center of gravity. ซึ่งมีกระบวนการในการหาดังนี้

ขั้นตอนแรก จะสุ่มเลือกออบเจ็คมา k จำนวน เพื่อทำหน้าที่เป็น mean หรือ center ของกลุ่ม k กลุ่ม หลังจากนั้นสำหรับออบเจ็คที่เหลือแต่ละตัวจะถูกนำมาจัดเข้ากลุ่มที่ใกล้เคียงที่สุด โดยวัดจากระยะห่างระหว่างออบเจ็คนั้นกับ center ของแต่ละกลุ่ม แล้วคำนวณหาค่า center ใหม่ของทุก ๆ กลุ่ม แล้วนำออบเจ็คที่เคยจัดกลุ่มแล้วทั้งหมดมาตรวจสอบกลุ่มใหม่อีกครั้ง ซึ่งกระบวนการดังกล่าวจะวนซ้ำจนกว่าไม่มีออบเจ็คใดเปลี่ยนแปลงกลุ่ม ในการพิจารณาว่าจำนวนกลุ่ม k เหมาะสมหรือไม่ ให้กำหนดค่า k แตกต่างกัน แล้วพิจารณาจากค่า squared-error criterion ซึ่งหาได้จาก

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ E คือค่าผลรวมของ square-error ของทุก ๆ ออบเจ็คในฐานข้อมูล โดยที่ p คือจุดใน space ที่ใช้แทนออบเจ็ค mi คือค่า mean ของ cluster Ci (ซึ่ง p และ mi อาจอยู่ในหลาย dimention)

ในการกำหนดค่า k ใด ๆ เมื่อใช้อัลกอริทึมดังกล่าวแล้วทำให้ค่า squared-error น้อยที่สุด จะสรุปได้ว่า การกำหนดกลุ่มด้วยค่า k ดังกล่าวเหมาะสมที่สุด

2.5 การพยากรณ์โดยใช้โครงข่ายประสาทเทียม

พื้นฐานการทำงานของโครงข่ายประสาทเทียมนั้นจะจำลองแบบจำลองทางคณิตศาสตร์อย่างง่ายมาจากการทำงานของสมอง มาใช้ในการทำงาน โดยเมื่อมีการนำโครงข่ายประสาทเทียมเข้ามาประยุกต์ใช้กับข้อมูล time series นั้น การทำงานจะทำในลักษณะของ non-linear คือทำงานกับข้อมูลซึ่งไม่สามารถแสดงในรูปของ linear ได้ และในการพยากรณ์นั้นจะต้องใช้ปริมาณของข้อมูลที่ทำการสังเกตมามากกว่าวิธีอื่นๆ อย่างวิธีทางสถิติ เป็นต้น แม้จะต้องใช้ข้อมูลจำนวนมาก แต่ส่งผลให้มีข้อดีคือ มีความยืดหยุ่นมากขึ้นในการพยากรณ์ และ สามารถทำงานที่ซับซ้อนให้ง่ายขึ้นได้

ด้านความหมายของคำจำกัดความต่างๆ ที่ใช้ในวิธีโครงข่ายประสาทเทียมนี้จะแตกต่างจากวิธีอื่น ได้แก่ จะใช้ network แทนคำว่า model หรือ ใช้ weight แทน parameter ต่างๆ หรือ ใช้คำว่า training the network แทนคำว่า estimating parameters

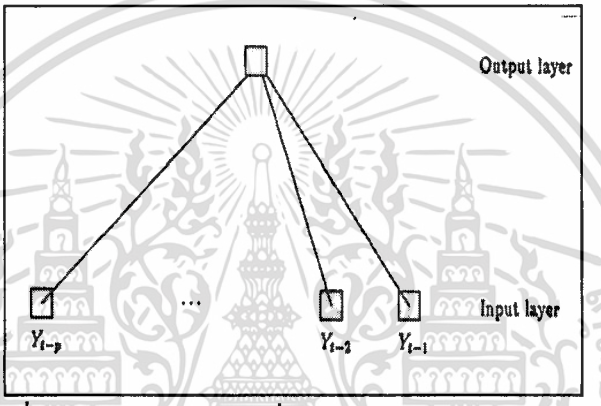
ลักษณะของ neural network นั้นสามารถอธิบายได้ว่า เป็นโครงข่ายของ neuron ที่เรียงตัวอยู่เป็นชั้นๆ ซึ่งที่ชั้นล่างสุดนั้นเป็นส่วนของ input units ส่วนชั้นบนสุดนั้นเป็นส่วนของ output units โดย units แต่ละอันในแต่ละชั้นจะถูกเชื่อมต่อเข้ากับ units ในชั้นที่อยู่เหนือกว่า

องค์ประกอบที่จำเป็นในการสร้าง neural network นั้นประกอบด้วย

1. **network architecture** คือ จำนวนชั้น และ units ในโครงข่าย และวิธีการเชื่อมโยงกันของ units ในแต่ละชั้น
2. **activation function** คือ วิธีที่แต่ละ unit ใช้ในการรวม input เข้าด้วยกันให้ออกมาเป็น output
3. **cost function** คือ วิธีการวัดความถูกต้องของการพยากรณ์ เช่น MSE เป็นต้น
4. **training algorithm** คือ วิธีการฝึกโครงข่ายเพื่อหาค่า parameter ที่ทำให้เกิดค่า cost function ที่น้อยที่สุด

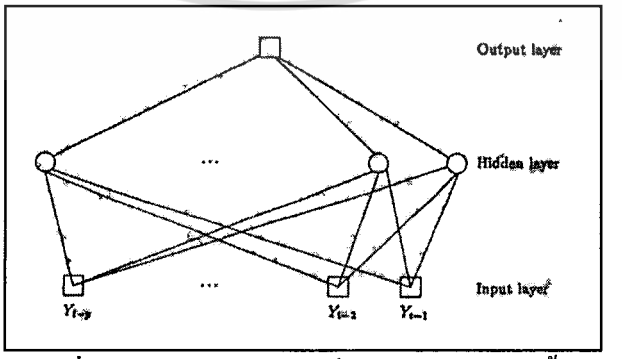
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างของการใช้ neural network นั้นได้แก่ การใช้ neural network อธิบาย AR(p) model (Auto regression) ดังรูปที่ 2.3 แสดงไดอะแกรมของ neural network สำหรับ linear AR(p) model โดยที่ชั้น input จะให้ค่าตัวแปรที่เป็น input ให้กับ network ในตัวอย่าง ใช้ค่าที่สังเกต ทั้งหมด p ค่า โดยชั้น output นั้นสามารถหาผลลัพธ์ได้จาก linear function ของ input ซึ่งแทนโดยเส้นเชื่อมระหว่างสองชั้น โดยค่า weight สำหรับแต่ละ input คือ parameter $\varnothing_1,\ldots,\varnothing_p$ ซึ่งจะถูกเลือกมาโดยใช้ learning algorithm (learning algorithm) ซึ่งเป็นค่าที่ทำให้ค่า cost function น้อยที่สุด



รูปที่ 2.3 Neural network ที่ใช้แสดง linear AR(p) model

ส่วนที่ทำให้ neural network มีประสิทธิภาพมากขึ้น คือ การเพิ่มชั้นกลางระหว่างสองชั้นเข้ามา คือ hidden unit ซึ่งประกอบด้วย non-linear hidden units อยู่ระหว่างชั้น input และ output ตัวอย่างเช่น ในรูปที่ 2.4 แสดง network ที่มี hidden layer 1 ชั้น โดย input จะเชื่อมต่อกับชั้น hidden unit และ hidden unit จะเชื่อมต่อกับ output unit 1 unit



รูปที่ 2.4 Neural network ที่มี hidden layer 1 ชั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าที่ได้ออกมาจาก unit เรียกว่า activation value โดยทั่วไป non-linear activation function จะได้จากการใช้ linear function ของ input แล้วจึงใช้ non-linear squashing function หรือที่เรียกว่า sigmoid function กับค่าที่ได้จาก linear function อีกที เพื่อให้เห็นภาพ ขอยกตัวอย่าง จากรูปที่ 2.4 input ของ hidden unit สามารถคำนวนได้จาก linear function (1)

$$Z = b + \sum_{i=1}^{p} w_i Y_{t-1} \qquad (1)$$

เมื่อ w_i คือ weight ของ unit i

Y_{t-i} คือ input ที่ t-i

จากนั้นค่าที่ได้จะเป็นค่า input สำหรับ non-linear function ในที่นี้คือ sigmoid function

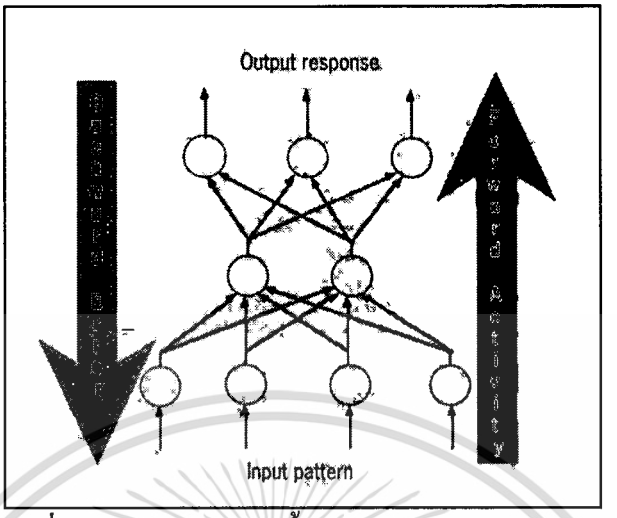
$$S(Z) = \frac{1}{1 + e^{-aZ}} \qquad (2)$$

จากนั้นค่า S(Z) ที่ได้จาก hidden unit แต่ละ unit จะถูกรวมกันด้วย linear function อีกครั้ง จึงได้ผลลัพธ์ออกมาเป็น output ซึ่งคือ ผลการพยากรณ์นั่นเอง

2.6 โครงข่าย Back propagation

โครงข่าย BPN จะทำงานเป็น 2 ขั้นตอน คือ ขั้นแรก เป็นการใส่ input pattern ให้กับ ชั้น input ของโครงข่าย ดังรูปที่ 2.5 จากนั้นผลลัพธ์จะไหลไปตามโครงข่ายจากชั้นล่างขึ้นไปยังชั้นถัดไป จนได้ผลลัพธ์ออกมาที่ชั้น output ส่วนขั้นตอนที่ 2 นั้น เป็นขั้นตอนการเปรียบเทียบ output ที่ได้จากโครงข่าย กับค่า output ที่คาดว่าจะได้จาก input pattern ที่ใส่เข้าไป หากค่าที่ได้ไม่ตรงกันจะหมายถึงมี error เกิดขึ้น ซึ่งค่า error นี้จะถูกส่งย้อยกลับมาในโครงข่าย จากชั้น output กลับมายังชั้น input ซึ่งจะทำให้ค่า weight ของชั้นภายในที่เชื่อมต่อกันนั้นถูกเปลี่ยนค่าไปตามค่า error ที่ถูกส่งกลับมา โดย BPN นั้นใช้หลักการของ minimum-error learning ที่เรียกว่า LMS rule หรือ delta rule มาดัดแปลงให้เหมาะกับโครงข่ายที่มีหลายชั้น จนได้เป็นกฎที่เรียกว่า generalized delta rule สำหรับใช้เป็นกฎในการปรับค่า weight

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 โครงสร้างและขั้นตอนการทำงานของ BPN

โครงข่าย BPN นั้นก็เหมือนกับโครงข่ายประสาทเทียมชนิดอื่น ๆ ที่มีสิ่งที่ต้องพิจารณา คือ การเชื่อมต่อกันของ neurode (unit), transfer function ที่ใช้โดย neurode และกฎการเปลี่ยน weight ตามปกติโครงข่าย BPN จะมีชั้นเชื่อมต่อกันอยู่ 3 ชั้นขึ้นไปโดยเชื่อมต่อแบบทั่วถึงกัน (fully connected) คือ ทุก unit ในแต่ละชั้นจะเชื่อมต่อกับทุก unit ในชั้นถัดไป โดยจะไม่มีการเชื่อมต่อกันภายในชั้นหรือกระโดดข้ามชั้น ดังนั้นชั้น input ก็จะเชื่อมต่อกับ ชั้น hidden ที่อยู่ตรงกลาง และชั้น hidden ก็จะเชื่อมต่อกับชั้น output เสมอ

ส่วน transfer function ที่ unit ใช้ คือ sigmoid function นั่นเอง โดยแต่ละ unit นั้นจะคำนวณค่า output ตามลำดับต่อไปนี้ คือ เริ่มจาก แต่ละ unit คำนวณ input I จากสมการ (3)

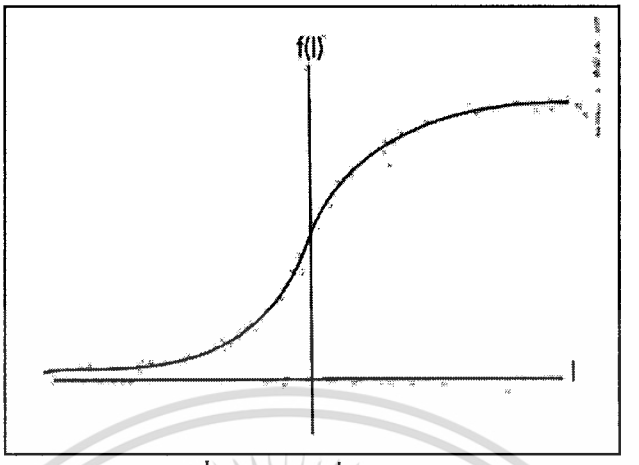
$$I_i = \sum_{j=1}^{n} w_i x_j \qquad (3)$$

จากนั้น ค่า input ที่ได้จะถูกส่งไปยัง activation function (squashing function) โดยทั่วไปจะอยู่ในรูป

$$f(I) = \frac{1}{1+e^{-I}} \qquad (4)$$

จากรูปที่ 2.6 คือกราฟที่ได้จาก f(I) โดยค่าที่ได้จะมีค่าอยู่ระหว่าง 0 และ 1 และ มีค่า slope เป็นบวกเสมอ โดยค่า output ที่ได้ของแต่ละ unit ส่วนใหญ่เป็น ค่า activation นี้เลย หรืออาจเป็นค่า activation ลบกับค่า threshold ก็ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.6 กราฟที่ได้จาก f(I)

จากนั้น ส่วนที่เหลือที่ต้องทำ คือ การทำ weight change law หรือ learning law โดยกฎที่ใช้คือ generalized delta rule ดังสมการ (4)

$$\Delta W_{ij} = \beta E F(I) \qquad (4)$$

โดย E คือ error ของ unit นั้นๆ

β คือ learning constant ซึ่งมีค่าระหว่าง 0 และ 1

F(I) คือ input ของ unit

ซึ่งการหาค่า error E นั้น แบ่งได้เป็น 2 กรณี

1. สำหรับชั้น output ค่า E หาได้จาก

$$E_j^{output} = Y_j^{desired} - Y_j^{actual} \qquad (5)$$

2. สำหรับชั้น middle ค่า E หาได้จาก

$$E_i^{middle} = \frac{df(I_i^{middle})}{dI} \sum_{j=1}^{n} (w_{ij} E_j^{output}) \qquad (6)$$

จากสมการ (5) และ (6) output และ middle คือชั้นของ unit ที่จะปรับ weight ส่วน unit j คือ unit ในชั้นoutput ส่วน i คือ unit ในชั้น middle ดังนั้น w_{ij} จะหมายถึง การตั้งค่า weight เมื่อมีทิศทางข้อมูลจาก unit i (ในชั้น middle) ไปยัง unit j (ในชั้น output) จะเห็นว่าจะเป็นไปในทิศทางตรงข้ามกับการปรับ weight จากชั้น output มายังชั้น middle ดังนั้นหากต้องการจะปรับ weight จากทิศทางย้อนกลับก็จะต้องทำการสลับค่า subscript ของตัวแปรต่างๆ ก่อนที่จะทำได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7 ข้อดีของ Neural Network

1. สามารถทำงานกับข้อมูลแบบ non-linear ได้
2 .สามารถทำงานกับข้อมูลที่มีความแตกต่างจากกลุ่มของข้อมูลส่วนใหญ่ข้อมูลที่ควรจะเป็น (noisy data) หรือทำงานกับข้อมูลที่มีบางส่วนของข้อมูลขาดหายไป (missing data) ซึ่งอาจเกิดจากข้อมูลไม่ได้ถูกเลือกจากขั้นตอนการเลือกข้อมูล
3. สามารถสร้างความสัมพันธ์ของข้อมูลได้โดยไม่ต้องใช้สมการ
4. สามารถทำงานกับข้อมูลที่มีตัวแปรหรือพารามิเตอร์จำนวนมากๆได้
5. สามารถใช้ในการแก้ปัญหาทั่วไปด้วยการทำนายที่มีความถูกต้อง

2.8 ข้อจำกัดของ Neural Network

1. Neural Network รับข้อมูลได้เฉพาะข้อมูลที่เป็นตัวเลขซึ่งมีค่าระหว่าง 0 ถึง 1 หรือ –1 ถึง 1 เท่านั้นดังนั้น ถ้า input เป็นข้อมูลที่มีค่าไม่อยู่ในช่วงดังกล่าวจะต้องทำการปรับให้อยู่ในช่วงตัวเลขที่กำหนดหรือในกรณีที่เป็นข้อมูลชนิดอื่นซึ่งไม่ใช่ตัวเลขก็ต้องแปลงให้เป็นตัวเลขก่อนเช่น ข้อมูลวันในสัปดาห์อาจแทนด้วยเลข 1,2,3,…,7 เป็นต้น แล้วปรับให้อยู่ในช่วงที่กำหนดอีกที
2. ไม่มีกฎเกณฑ์ในการกำหนดจำนวน hidden layer แต่ถ้าใช้ hidden layer มากกว่า 1 layerจะทำให้ Network ใช้เวลาเรียนรู้นาน ดังนั้นวิธีที่ดีที่สุดคือให้มีจำนวน hidden layer น้อยที่สุดเท่าที่จะทำได้ถ้าเป็น feed forward network ควรใช้ 1 hidden layer
3. การกำหนดจำนวน hidden unit ใน hidden layer ไม่มีกฎเกณฑ์ตายตัวต้องลองผิดลองถูกเพื่อหาจำนวน hidden node ที่ทำให้เกิด error น้อยที่สุด
4. ต้องกำหนดชุดของข้อมูลที่ใช้ในการเรียนรู้ของ neural network ให้ครอบคลุมเพราะถ้า neural network เรียนรู้ไม่ครอบคลุมจะทำให้เกิดความผิดพลาดในการพยากรณ์ได้

2.9 สรุป

จากการนำเสนอ Data Miningในการจัดการกับข้อมูลทำให้เราสามารถวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพ เพื่อนำไปใช้ให้เกิดประโยชน์ทาง ธุรกิจได้อย่างดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การออกแบบระบบสำหรับการจัดการวัตถุดิบ

3.1 การออกแบบระบบ จัดกลุ่มสูตรหินเจียรโดยใช้ Database Segmentation

ในกระบวนการทำงานของการผลิตหินเจียร เมื่อได้รับ Order จากลูกค้า (ในรูปของ Product Lot) ซึ่งจะระบุเป็น รูปร่างหน้าตาของผลิตภัณฑ์ รวมไปถึง ขนาด สี ความแข็งแรง ซึ่งทางโรงงานก็จะนำมาออกแบบ และผสมวัตถุดิบให้ได้ตามต้องการ โดยอัตราส่วนในการผสมอาจจะเป็นสูตรเดิมที่มีอยู่ หรือคิดสูตรใหม่เพื่อความเหมาะสม การทำระบบ Database segmentation เพื่อจัดกลุ่มสูตรที่มีให้เข้ากลุ่ม เพื่อประโยชน์ในการยื่นจดขอใช้สิทธิยกเว้นภาษีนำเข้าวัตถุดิบจาก BOI โดยพิจารณาจากสิ่งต่อไปนี้

3.1.1. Data Cleaning Process ทางโรงงานมีฐานข้อมูลในการที่จะทำ Segmentation ดังนี้

ตารางที่ 3.1 ตารางแสดงส่วนผสมวัตถุดิบ

	Lot 1	Lot 2	Lot 3	Lot 4	Lot 5
CodeNo	1000000011	1000000011	1000000022	1000000022	1000000022
LotNo	339-00038	337-00046	347-00278	349-00416	347-00375
Mix Date	18-Feb-04	18-Sep-03	23-Aug-04	15-Nov-04	17-Sep-04
BOITS2N	32A	32A	K A	K A	K A
BOITS2M	80	80	80	80	80
BOITS2	5600	5600	2800	2800	2800
BOITS3N			TWA	TWA	TWA
BOITS3M	0	0	80	80	80
BOITS3	0	0	1400	1400	1400
BOITS4N			M A	M A	M A
BOITS4M	0	0	80	80	80
BOITS4	0	0	1400	1400	1400
BOIMS1N	8J	8J	8M	8M	8M
BOIMS1	560	560	660	660	660
BOIMS2N					
BOIMS2	0	0	0	0	0
BOIMS3N	NND	NND	DL	DL	DL
BOIMS3	340	340	170	170	170
BOIMS4N	25%L	25%L	EW	EW	EW
BOIMS4	120	120	120	120	120

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตาราง แสดงข้อมูล material ของ Lot การผลิต 5 Lot ซึ่ง 2 Lot แรกจะใช้สูตรการผลิตเดียวกัน คือสูตรหมายเลข 1000000011 ส่วน 3 Lot ที่เหลือใช้สูตรหมายเลข 1000000022

- ฟิล์ดที่ขึ้นต้นด้วย BOITS และ BOITI จะเป็นจำนวนของ Grain หลักที่ใช้ ซึ่งถ้าลงท้ายด้วย N หมายถึงชื่อ Grain ถ้าลงท้ายด้วย M หมายถึงขนาดของเม็ด Grain และฟิล์ดที่เหลือเป็นปริมาณการใช้
- ฟิล์ดที่ขึ้นต้นด้วย BOIMS และ BOIMI จะเป็น material ส่วนประกอบ โดยถ้าลงท้ายด้วย N หมายถึงชื่อ และฟิล์ดที่เหลือเป็นปริมาณการใช้

จากตารางจะเห็นได้ว่าในสูตรเดียวกัน จะใช้วัตถุดิบเหมือนกันเช่นในสูตรหมายเลข 1000000011 จะใช้ Grain ชื่อ 32 A ขนาด 80 mm ในปริมาณ 5600 g เป็นต้น

ตารางที่ 3.2 ตารางตัวอย่างชื่อวัตถุดิบที่ขอยื่นจดกับ BOI

MatNo	MatName
1	ALUMINIUM OXIDE Type I
2	ALUMINIUM OXIDE Type II
3	ALUMINIUM OXIDE Type III
4	SILICON CARBIDE Type I
5	SILICON CARBIDE Type II

ตารางที่ 3.3 ตารางแสดงส่วนผสมของวัตถุดิบที่ใช้ผลิตกับวัตถุดิบที่ขอยื่นจดกับ BOI

mgtName	25%L	32A	8J	8M	EW	KA	MA	NND	TWA
mType	B	G	B	B	B	G	G	B	G
M01	0	1	0	0	0	1	1	0	1
M02	0	0	0	0	0	0	0	0	0
M03	0	0	0	0	0	0	0	0	0
M04	0	0	0	0	0	0	0	0	0
M05	0	0	0	0	0	0	0	0	0
M06	0	0	0.4	0.4	0	0	0	0	0
M07	0	0	0	0	0	0	0	0	0
M08	0	0	0.1	0.4	0	0	0	0	0
M09	0	0	0.3	0	0	0	0	0	0
M10	0	0	0.27	0.27	0	0	0	0	0
M11	0	0	0	0	0	0	0	0	0
M12	0	0	0.03	0.03	0	0	0	0	0
M13	0	0	0	0	0	0	0	0	0
M14	0	0	0	0	0	0	0	0	0
M15	0	0	0	0.008	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

จากตารางที่ 3.3 แสดงส่วนผสมของ Material ที่ยื่นจดกับ BOI ไปเป็นวัตถุดิบที่ใช้ในการผลิต เช่น 8J ที่ใช้สำหรับผลิตหินเจียร เกิดจากนำวัตถุดิบที่ยื่นจดกับ BOI หมายเลข 6,8,10,12 ในอัตราส่วน 40 % , 40 % , 27 % และ 3 % ตามลำดับ

จากตารางดังกล่าวจะต้องมีการจัดเตรียมข้อมูลเพื่อใช้ในการ Segmentation โดยจะต้องทำการแปลงข้อมูลให้อยู่ในรูปของความสัมพันธ์ระหว่าง Lot กับอัตราส่วนวัตถุดิบทั้ง 59 ชนิด ซึ่งได้ดังตาราง

ตารางที่ 3.4 ตารางแสดงความสัมพันธ์ระหว่าง Lot กับอัตราส่วนวัตถุดิบ

LotNo	339-00038	337-00046	347-00278	349-00416	347-00375
ClassID					
M01	100	100	100	100	100
M02	0	0	0	0	0
M03	0	0	0	0	0
M04	0	0	0	0	0
M05	0	0	0	0	0
M06	4	4	4.80000019	4.80000019	4.80000019
M07	0	0	0	0	0
M08	1	1	4.80000019	4.80000019	4.80000019
M09	3	3	0	0	0
M10	2.70000005	2.70000005	3.24000001	3.24000001	3.24000001
M11	0	0	0	0	0
M12	0.30000001	0.30000001	0.36000001	0.36000001	0.36000001
M13	0	0	0	0	0
M14	0	0	0	0	0
M15	0	0	0.096	0.096	0.096
M16	0	0	0.03	0.03	0.03
M17	0	0	0	0	0
M18	0	0	0	0	0
M19	1.81817997	1.81817997	1.5	1.5	1.5
M20	0	0	0	0	0
M21	0.5	0.5	1.5	1.5	1.5
M22	0.54545999	0.54545999	0	0	0
M23	3.63635993	3.63635993	0	0	0
M24	0	0	1.40077996	1.40077996	1.40077996
M25	0	0	0	0	0
M26	0	0	0	0	0

จากตาราง แสดงความสัมพันธ์ระหว่าง Lot กับเปอร์เซ็นต์ของวัตถุดิบที่ยื่นจดบัญชี BOI ซึ่งโดยทั่วไปแล้วผลรวมของ M01 – M05 (ซึ่งเป็น Grain) จะมีค่า 100 % ส่วนตั้งแต่ M06 ขึ้นไปจะเปรียบเทียบเปอร์เซ็นต์กับน้ำหนักของ Grain เช่น จาก Lot ที่ 339-00038 ถ้าใช้ Grain 100 g ก็จะใช้วัตถุดิบหมายเลข 6 เป็นจำนวน 4 g เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกระบวนการดังกล่าว จะได้อัตราส่วนในแต่ละ Lot ซึ่งมีข้อมูลทั้งหมดประมาณ 75,000 รายการ แต่จากตารางตัวอย่างข้างต้น เมื่อนำมาจัดกลุ่มเบื้องต้น จะเหลือประมาณ 5,000 รายการที่แตกต่างกัน ซึ่งจะนำมาเป็นข้อมูลเบื้องต้นในการทำ K-mean ต่อไป

3.1.2 Database Segmentation Process

จากรายการที่แตกต่างกัน (n = 5000) จะนำมาทดสอบหาจำนวนกลุ่มที่เหมาะสมที่สุด โดยเริ่มตั้งแต่ k = 3 ถึง 35 แล้วหาค่า Sum of square-error ในแต่ละค่า k ซึ่งแสดงดังตาราง

ตารางที่ 3.5 ตารางบันทึกผลการทดสอบการทำ Segmentation

TotSegment	NumRec	SumErr_2
3	500	240561.7656
4	500	198737.8594
5	500	180991.2656
6	200	16655.19336
6	300	29151.76953
6	400	47229.42188
6	500	73504.17969
6	600	107826.6094
6	700	137306.6875
6	800	166913.1563

จากตาราง คอลัมน์แรกจะเป็นค่า k จากรูป เป็นการจัดเก็บผลการทดสอบเมื่อค่า k = 3 ถึง 6 และคอลัมน์ที่สองแสดงจำนวนข้อมูลที่จะใช้ทดสอบ และคอลัมน์ที่สามแสดงค่า Sum of Squire Error ในการทดสอบ จากนั้น นำผลที่ได้จากการทดสอบที่ค่า k และ n ต่าง ๆ กัน แล้วเลือกที่เหมาะสมที่สุดเพื่อที่จะนำมาทำการ prediction เพื่อหายอดการใช้ต่อไป

3.1.3 Database Classification Process

จากการทำ Data Segmentation จะมีการบันทึกเก็บ center ของแต่ละกลุ่มไว้ดังตัวอย่าง

ตารางที่ 3.6 ตารางบันทึก Center ของกลุ่มที่ผ่านการ Segmentation

TotSegment	SegmentID	NumRec	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10
6	1	1500	14.535	84.370	0.2728	0.4387	0.3835	2.8789	1.2371	1.5412	2.5207	1.0621
6	2	1500	43.036	7.7119	0.7614	47.033	1.6949	0.2377	0.0432	0.3292	0	0.0288
6	3	1500	97.848	1.0006	0.2165	0.9651	0	0.1532	0	0.1033	0	0
6	4	1500	60.909	38.538	0.5657	0	0	3.0588	1.1567	1.6774	2.6945	1.3113
6	5	1500	0	0	0.2989	99.713	0	2.4845	1.1942	5.1161	0.0766	0.6462
6	6	1500	98.423	1.5773	0	0	0	3.9802	1.5139	2.8518	3.0613	1.3416

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

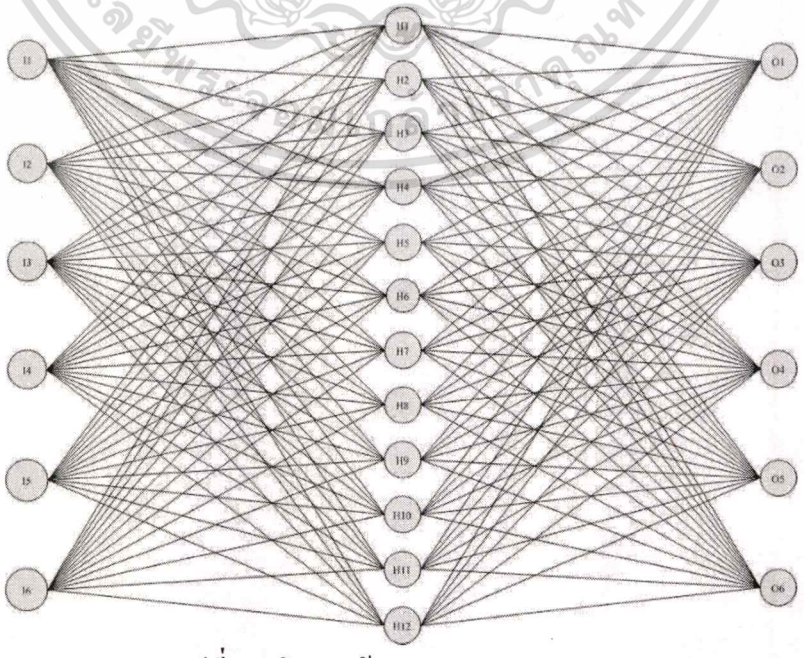
จากตารางเป็นการบันทึกค่า Center ของการจัดกลุ่มทั้งหมด 6 กลุ่ม เมื่อจำนวนข้อมูลเป็น 1500 เรคอร์ด โดยในแต่ละกลุ่มจะมีอัตราส่วนการใช้วัตถุดิบที่แตกต่างกันดังตาราง หลังจากเลือกกลุ่มที่ต้องการแล้ว ก็จะนำค่าในตารางข้างต้น ไป Classify ข้อมูลในแต่ละ Lot โดยดูจาก distance function ของ Lot นั้น ๆ กับ Center ในแต่ละกลุ่มแล้วเข้ากลุ่มที่มีค่า distance function น้อยที่สุด และเมื่อมีการสั่งสินค้า Lot ใหม่เกิดขึ้น ก็จะทำในทำนองเดียวกัน

3.2 การออกแบบระบบพยากรณ์ยอดการผลิตหินเจียรโดยใช้โครงข่ายประสาทเทียม

จากการแบ่งกลุ่มหินเจียรในหัวข้อ 3.1 ทำให้สามารถหาได้ว่า ในแต่ละเดือนมีการผลิตหินเจียรในแต่ละกลุ่มจำนวนเท่าใด และจากข้อมูลการผลิตในปีผ่าน ๆ มา ตั้งแต่ 1995 – 2004 จะนำมาเป็นข้อมูลเบื้องต้นในการพยากรณ์ยอดการผลิต ซึ่งใช้วิธี Backpropagation โดยมีการออกแบบระบบดังนี้

3.2.1. Neural Network design ขั้นตอนนี้ จะเป็นการกำหนดโครงสร้างของ Neural network ซึ่งจะแยกชุดกันในแต่ละกลุ่มโดยมีโครงสร้างที่เหมือนกันคือ

1 ข้อมูลนำเข้า เกิดจากยอดการผลิตใน 6 เดือนย้อนหลัง เดือนที่จะพยากรณ์
2 ชั้นซ่อนเร้น (Hidden Layers) ในโครงข่ายกำหนดไว้ 1 ชั้น โดยมี จำนวน 12 โหนด
3 ข้อมูลออกที่ต้องการเป็นยอดการผลิต 6 เดือนถัดไป ดังนั้นขนาดของเอาท์พุทของโครงข่ายประสาทเทียมจึงมี 6 โหนด



รูปที่ 3.1 โครงสร้างของ Neural network

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2. Data Preparing Process การจัดเตรียม Input และ Output ซึ่งจัดเตรียมดังตาราง

ตารางที่ 3.7 ตารางแสดงการจัดเตรียมข้อมูลสำหรับ Neural network

ClassID	TimeID	I1	I2	I3	I4	I5	I6	O1	O2	O3	O4	O5	O6
1	14	0.2392	0.3046	0	0.2927	0	0.2773	0.4414	0.2908	0.9606	0.3849	0.6147	0.4496
1	15	0.3046	0	0.2927	0	0.2773	0.4414	0.2908	0.9606	0.3849	0.6147	0.4496	0.9753
1	16	0	0.2927	0	0.2773	0.4414	0.2908	0.9606	0.3849	0.6147	0.4496	0.9753	0.5151
1	17	0.2927	0	0.2773	0.4414	0.2908	0.9606	0.3849	0.6147	0.4496	0.9753	0.5151	0.7926
1	18	0	0.2773	0.4414	0.2908	0.9606	0.3849	0.6147	0.4496	0.9753	0.5151	0.7926	0.3468
1	19	0.2773	0.4414	0.2908	0.9606	0.3849	0.6147	0.4496	0.9753	0.5151	0.7926	0.3468	0.9116
1	20	0.4414	0.2908	0.9606	0.3849	0.6147	0.4496	0.9753	0.5151	0.7926	0.3468	0.9116	0.4353
1	21	0.2908	0.9606	0.3849	0.6147	0.4496	0.9753	0.5151	0.7926	0.3468	0.9116	0.4353	0.2901
1	22	0.9606	0.3849	0.6147	0.4496	0.9753	0.5151	0.7926	0.3468	0.9116	0.4353	0.2901	0.3411
1	23	0.3849	0.6147	0.4496	0.9753	0.5151	0.7926	0.3468	0.9116	0.4353	0.2901	0.3411	0.3117
1	24	0.6147	0.4496	0.9753	0.5151	0.7926	0.3468	0.9116	0.4353	0.2901	0.3411	0.3117	0
1	25	0.4496	0.9753	0.5151	0.7926	0.3468	0.9116	0.4353	0.2901	0.3411	0.3117	0	0.2722
1	26	0.9753	0.5151	0.7926	0.3468	0.9116	0.4353	0.2901	0.3411	0.3117	0	0.2722	0.2708

จากตารางเป็นข้อมูลของกลุ่มที่ 1 โดยกำหนด Time Series เริ่มจากเดือน มค. 1995 เป็น 1 ซึ่งจากตาราง เริ่มข้อมูลในเดือน ก.พ. 1996 (TimeID = 14) ซึ่งในเรคอร์ดแรก I1 – I6 จะเป็นยอดผลิตของเดือน ก.พ. ถึง ก.ค. และ O1 – O6 เป็นยอดผลิตของเดือน ส.ค. 1996 – ม.ค. 1997 ตามลำดับ ซึ่งข้อมูลดังกล่าว ผ่านการ Transform ให้อยู่ในช่วง 0 – 1 การเตรียมข้อมูล Weigth และ Bias ของแต่ละ Node ซึ่งเกิดจากการสุ่มค่าตั้งแต่ -1 ถึง 1 โดยเก็บในรูปตารางดังนี้

ตารางที่ 3.8 ตารางแสดงค่า weight และ bias เริ่มต้นของแต่ละกลุ่ม

bKey	bValue
bH_0	-0.6
bH_1	0.4
bH_10	-0.8
bH_11	0
bH_2	0.2
bH_3	0.8
bH_4	0.8
bH_5	-0.6
bH_6	0.2
bH_7	0.2
bH_8	1
bH_9	-0.4
bOut_0	-0.4
bOut_1	-0.3
bOut_2	0.7
bOut_3	0.7
bOut_4	-0.5
bOut_5	0.6

bKey	bValue
wHO_4_0	-0.9
wHO_4_1	0.5
wHO_4_2	0.5
wHO_4_3	-0.5
wHO_4_4	0.3
wHO_4_5	1
wHO_5_0	0.3
wHO_5_1	0.1
wHO_5_2	0.6
wHO_5_3	-0.9
wHO_5_4	-0.3
wHO_5_5	0.3
wHO_6_0	-0.5
wHO_6_1	-0.1
wHO_6_2	0.5
wHO_6_3	-0.7
wHO_6_4	-0.3
wHO_6_5	-0.8

bKey	bValue
wIH_0_9	-0.1
wIH_1_0	-0.9
wIH_1_1	-0.5
wIH_1_10	0.1
wIH_1_11	0.9
wIH_1_2	-1
wIH_1_3	-1
wIH_1_4	0.2
wIH_1_5	0.2
wIH_1_6	0.5
wIH_1_7	1
wIH_1_8	0.5
wIH_1_9	-0.6
wIH_2_0	0.5
wIH_2_1	0.2
wIH_2_10	-0.1
wIH_2_11	0.6
wIH_2_2	0.6

bKey	bValue
wIH_5_4	-0.6
wIH_5_5	0.2
wIH_5_6	-0.8
wIH_5_7	-0.9
wIH_5_8	0.8
wIH_5_9	-0.7
wIH_6_0	0.3
wIH_6_1	-0.6
wIH_6_10	-0.9
wIH_6_11	0.3
wIH_6_2	-0.1
wIH_6_3	1
wIH_6_4	0.3
wIH_6_5	0.8
wIH_6_6	0.6
wIH_6_7	-0.3
wIH_6_8	-0.9
wIH_6_9	-0.2

bKey	bValue
wHO_2_2	0.9
wHO_2_3	0.9
wHO_2_4	0.2
wHO_2_5	-0.8
wHO_3_0	0.2
wHO_3_1	-0.4
wHO_3_2	0.2
wHO_3_3	-0.5
wHO_3_4	0.9
wHO_3_5	-0.9
wIH_0_16	0.7
wIH_0_17	0.4
wIH_0_18	-0.4
wIH_0_19	-0.3
wIH_0_2	0.9
wIH_0_3	0.9
wIH_0_4	0
wIH_0_5	-0.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3 Data Learning Process

จะทำการ learn ทีละกลุ่มโดยเลือกข้อมูลบางส่วนของแต่ละกลุ่มมาทำการ learning โดยกำหนดจำนวนรอบ และ Check Point ค่า Sum of Squared Error ซึ่งจัดเก็บดังตาราง

ตารางที่ 3.9 ตารางจัดเก็บผลการ Learning

	TotClass	CheckID	Learnf	OutErr_2
▶	15	50000	0.04	5.58340883
	15	100000	0.04	4.68513775
	15	150000	0.04	4.33438063
	15	200000	0.04	4.13301373
	15	250000	0.04	3.93749738
	15	300000	0.04	3.81084824
	15	350000	0.04	3.70808005
	15	400000	0.04	3.64935732
	15	450000	0.04	3.60627151
	15	500000	0.04	3.5723846
	15	550000	0.03	3.52940631
	15	600000	0.03	3.51303983
	15	650000	0.03	3.4987216
	15	700000	0.03	3.4860425
	15	750000	0.03	3.47442651

จากตาราง เป็นการ Learn โดยใช้ learning Rate = 0.03 มีการทำงาน 750,000 รอบ และ Check Point ทุก ๆ 50,000 รอบ ซึ่งเพิ่มจำนวนรอบได้จนกว่าค่า Sum of Squared Error เป็นที่น่าพอใจ

3.2.4 Data Prediction Process

ทำเมื่อมีการ Learning จนได้ Model ของแต่ละกลุ่ม เป็นการพยากรณ์ 6 เดือน โดยแบ่งเป็นการทดสอบกับข้อมูลส่วนที่เหลือที่ไม่ได้ learning ซึ่งมีการเก็บข้อมูลเพื่อนำไปสร้างกราฟดังตาราง

ตารางที่ 3.10 ตัวอย่างตารางจัดเก็บผลการพยากรณ์

	TimeID	RealValue	PredicValue
▶	12	279200	279200
	13	223776	223776
	14	4560300	4560300
	15	4121540	4121540
	16	7439300	7439300
	17	6321700	6321700
	18	7117000	6219306.5
	19	6492100	6264988
	20	5738200	5951525
	21	5949300	6111425.5
	22	14120000	6087863.5
	23	8602600	5973160
	24	8045200	6267812.5
	25	14101800	6176766
	26	9640800	6063183
	27	8455300	6169931
	28	5682900	5938156.5
	29	8590700	5816967
	30	5059200	6741982
	31	11771700	6894911.5
	32	6902500	6531260.5
	33	5630300	6842601.5
	34	7723900	6812823.5
	35	4139000	6343014
	36	3089015	6224553
	37	1693586	6255606
	38	1206500	6194181
	39	4426100	6352290.5
	40	2336700	6170598
	41	3400800	6103410.5
	42	3128800	4136253
	43	4216800	4436734.5
	44	4370300	4586706.5
	45	4946500	4430834
	46	5818400	4657864.5
	47	2974100	4694278
	48	0	4316665
	49	0	4270918.5
	50	181000	4397775

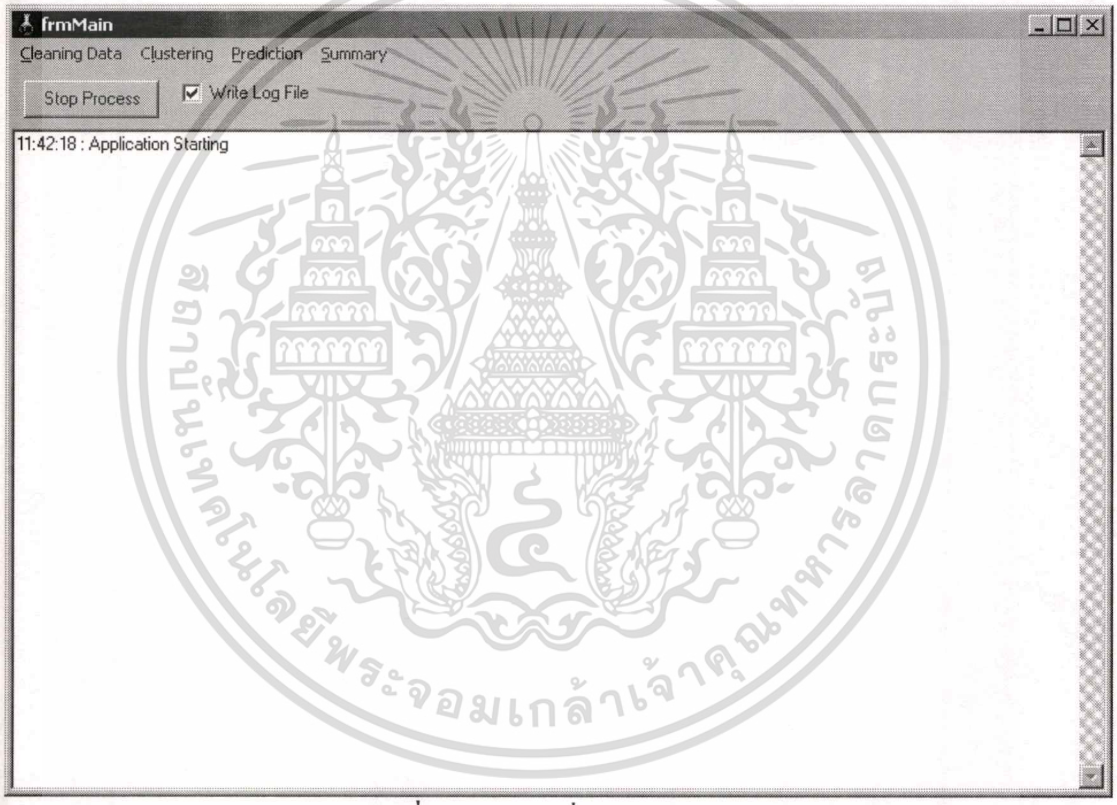
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

วิธีการใช้งานระบบ

4.1 เริ่มต้นใช้งานโปรแกรม

จากการออกแบบในที่กล่าวมาข้างต้น ได้มีการพัฒนาโปรแกรมสำหรับจัดการตามขั้นตอนดังกล่าว ซึ่ง เมื่อเปิดโปรแกรม จะมี Interface ดังรูป



รูปที่ 4.1 หน้าจอเริ่มต้นโปรแกรม

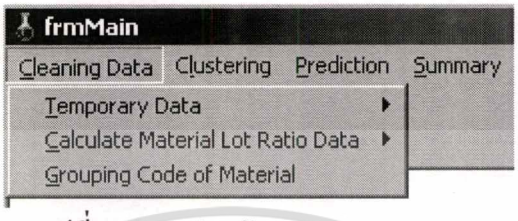
จากรูปประกอบไปด้วย

- เมนูการทำงาน แบ่งเป็น เมนูสำหรับ Cleaning Data , Clustering และ Prediction
- ปุ่ม button ที่ใช้สำหรับ stop การทำงานของ process ที่กระทำอยู่
- Checkbox สำหรับเลือกที่จะ write รายละเอียดการทำงานลง File
- ส่วน monitor สำหรับแสดงผลการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 Cleaning data Process

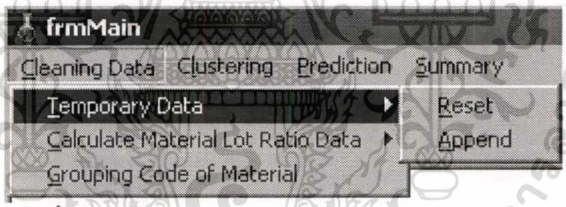
การ Cleaning data คือการจัดเตรียมข้อมูลเพื่อเข้าสู่ระบบ โดยมีกระบวนการต่าง ๆ ดังเมนูต่อไปนี้



รูปที่ 4.2 เมนูสำหรับการ Cleaning data

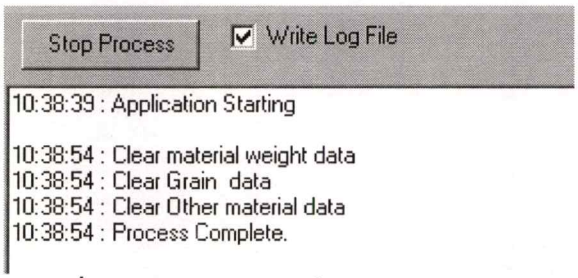
จากรูป แบ่งกระบวนการออกเป็น 3 ขั้นตอนย่อย คือ

- การจัดการ Temporary Data ซึ่งเป็นขั้นตอนในการ load ข้อมูลจาก table ที่เกี่ยวข้องซึ่งอยู่ในรูปของ query ที่ซับซ้อนและใช้เวลาในการ process นาน มาพักไว้ใน Temporary table เพื่อให้การทำงานต่อไปทำได้เร็วขึ้น โดยมีกระบวนการย่อย 2 ขั้นตอนดังรูป



รูปที่ 4.3 เมนูสำหรับการจัดการ Temporary Data

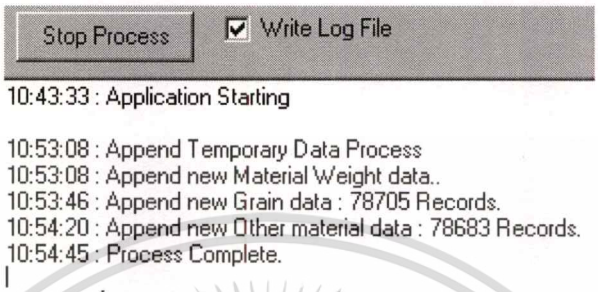
  - Reset คือการล้างข้อมูลทุกอย่างที่อยู่ใน temporary table ทำเมื่อต้องการจะ clear ข้อมูลใหม่ ซึ่งเมื่อเลือกทำงาน จะแสดงผลการทำงานดังรูป



รูปที่ 4.4 ผลการทำงานเมื่อเลือก Reset ข้อมูล

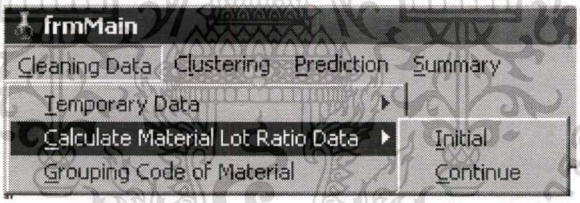
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

○ Append คือการ load ข้อมูลเฉพาะข้อมูลที่ยังไม่มีเข้ามาไว้ใน temporary table ซึ่งเมื่อเลือกทำงานจะแสดงผล ดังรูป



รูปที่ 4.5 ผลการทำงานเมื่อเลือก Append ข้อมูล

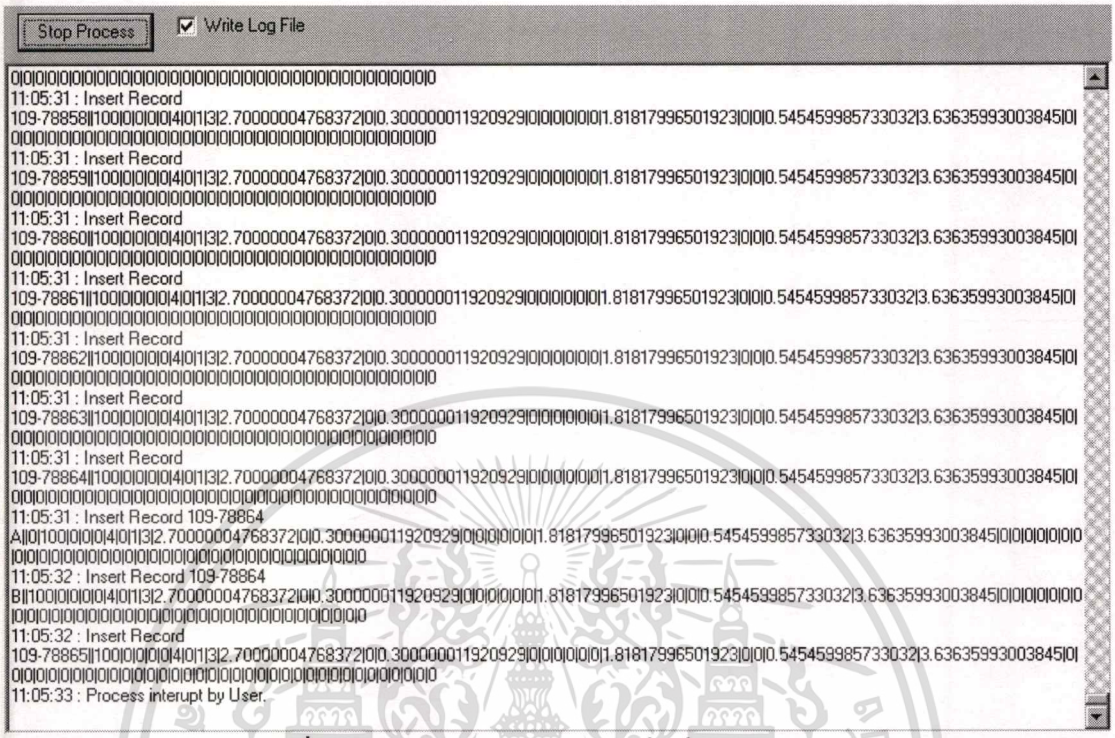
- Calculate Material Lot Ratio Data ใช้สร้างข้อมูลแสดงความสัมพันธ์ระหว่าง Lot ของ Product กับอัตราส่วน material ที่ใช้ผลิตจดกับ BOI ซึ่งมีขั้นตอนย่อย 2 ขั้นตอนดังรูป



รูปที่ 4.6 เมนูสำหรับคำนวณอัตราส่วนการใช้ Material ในแต่ละ Lot

○ Initial คือการเริ่มต้นกระบวนการ Calculate ซึ่งจะ ล้างข้อมูลเดิมที่มีอยู่ และสร้างข้อมูลขึ้นมาใหม่ โดยจะแสดงผลของการ Insert ข้อมูลลง Table ดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

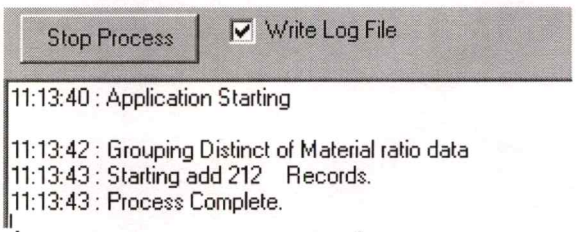


รูปที่ 4.7 แสดงผลของการทำงานเมื่อเริ่ม Insert ข้อมูล

จากรูป เป็นการแสดงผลการคำนวณอัตราส่วนการใช้วัตถุดิบในแต่ละ Lot แล้วเพิ่มรายการลงใน Table ซึ่งเป็นขั้นตอนที่ใช้เวลานาน ผู้ใช้สามารถหยุดการทำงานโดยคลิกที่ Stop Process

- Continue คือกระบวนการ Calculate ข้อมูลต่อจากที่ทำค้างไว้ โดยโปรแกรมจะตรวจสอบ record ที่ยังไม่ได้คำนวณแล้ว Insert ข้อมูลเพิ่ม โดยจะมีการแสดงผลเหมือนกับรูป 4.7

- Grouping Code of Material คือกระบวนการในการจัดกลุ่มของ Lot Ratio Data ที่มีอัตราส่วนการใช้วัตถุดิบเหมือนกัน โดยจะแสดงผลการทำงานดังรูป

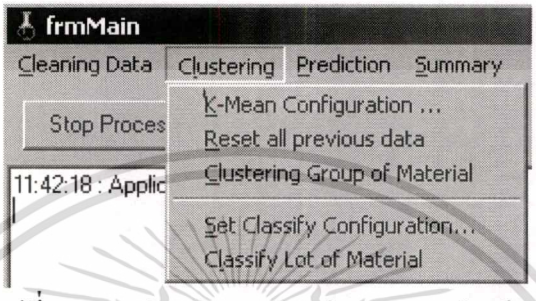


รูปที่ 4.8 แสดงผลของการทำงานเมื่อทำการ group ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 Clustering Process

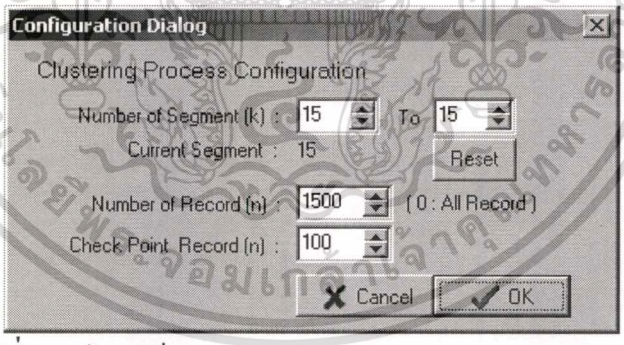
คือกระบวนการตั้งแต่การนำข้อมูลอัตราส่วนวัตถุดิบของ Lot ที่แตกต่างกันมาจัดกลุ่มโดยวิธีการ k-mean อัลกอริทึม ไปจนถึงการจำแนกกลุ่มของ Lot ที่เกิดขึ้นมาภายหลัง ซึ่งมีเมนูดังต่อไปนี้



รูปที่ 4.9 เมนูสำหรับการจัดกลุ่มของ Lot หินเจียร

จากเมนู ประกอบไปด้วย

- K-Mean Configuration เป็นขั้นตอนในการกำหนด Configuration ให้กับ k-mean อัลกอริทึม ซึ่งเมื่อคลิกเข้าไป จะแสดงไดอะล็อก ดังรูป



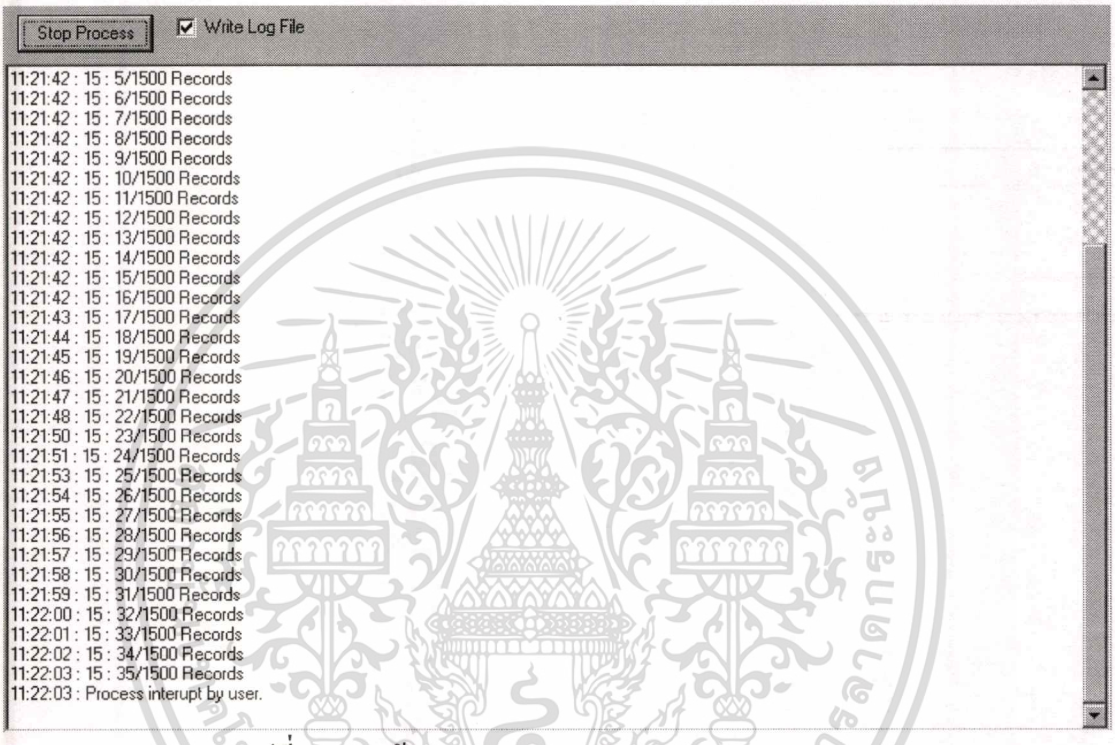
รูปที่ 4.10 ไดอะล็อกแสดงการ configuration k-mean algorithm

จากรูป ผู้ใช้จะต้องกำหนด

- จำนวนของ Segment ที่จะแบ่ง ซึ่งโปรแกรมสามารถ Process ได้ต่อเนื่องตามช่วงที่กำหนดไว้ ซึ่งจากรูป มีการเจาะจงเฉพาะค่า Segment = 15
- จำนวนของ Record ใน Distinct Record Lot Ratio Table จากตัวอย่าง จะ Process ในค่า k แต่ละค่า เป็นจำนวน 1500 เรคอร์ด
- จำนวนเรคอร์ดในการเก็บ Check Point ซึ่งประกอบไปด้วย Sum of Square Error และ Center ของแต่ละกลุ่ม ซึ่งจากตัวอย่าง จะบันทึกเมื่อสิ้นสุดทีละ 100 Record

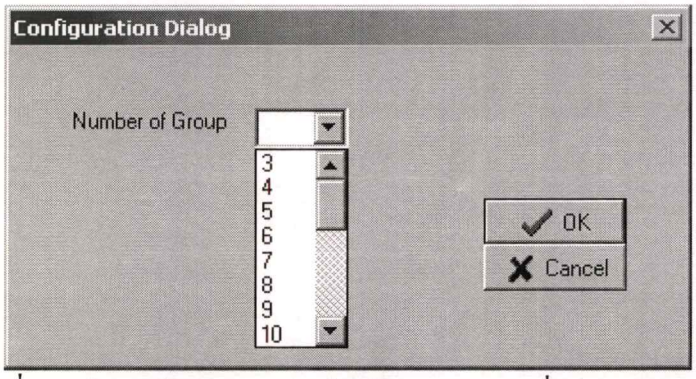
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Reset all previous data คือ การล้างข้อมูลการทำ Segmentation เดิมทั้งหมด
- Clustering Group of Material คือการจัดกลุ่มตามค่าที่กำหนดใน Configuration ซึ่งอาจใช้เวลานานขึ้นอยู่กับปริมาณข้อมูลและจำนวน Cluster โดยแสดงผลการทำงานดังรูป



รูปที่ 4.11 หน้าจอแสดงผลการทำ k-mean segmentation

- Set Classify Configuration คือกระบวนการในการกำหนด configuration เพื่อที่จะ classify Lot ของ Product ที่มีอยู่ ซึ่งเมื่อเลือกเมนูดังกล่าวจะแสดงไดอะล็อก ดังรูป

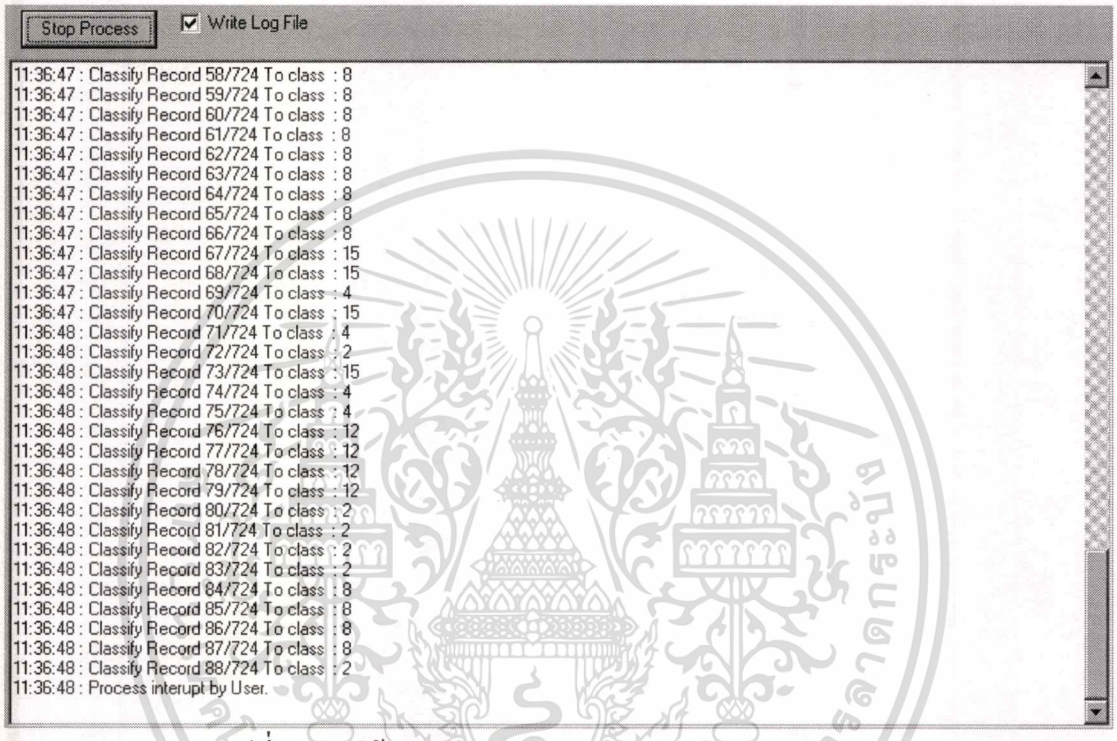


รูปที่ 4.12 ไดอะล็อกแสดงการเลือกจำนวน group ที่จะ classify

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป ผู้ใช้จะต้องเลือกจำนวนของกลุ่มที่จะใช้แบ่ง ซึ่งค่าที่สามารถเลือกได้ จะต้องมีการเก็บ center ของแต่ละกลุ่มจากกระบวนการก่อนหน้ามาก่อนแล้ว

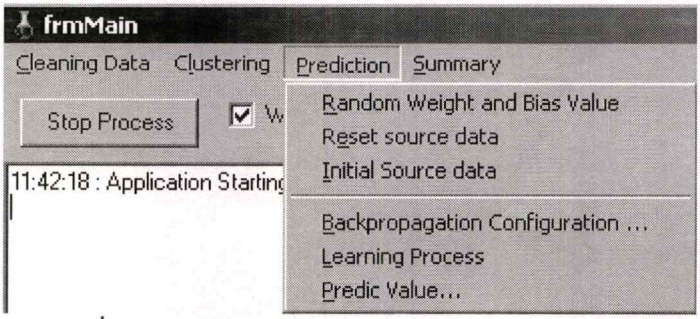
- Classify Lot of Material คือการจัดกลุ่มของ Lot โดยมีการแสดงผลการทำงาน ดังรูป



รูปที่ 4.13 หน้าจอแสดงผลของการ Classify Lot of Material

4.4 Prediction Process

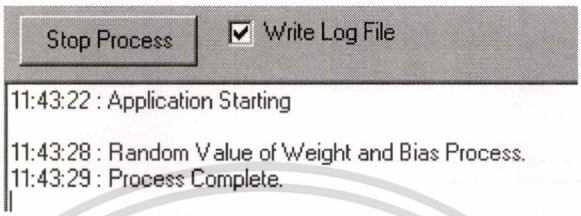
เป็นกระบวนการที่เริ่มตั้งแต่การสร้างโมเดลของการพยากรณ์ยอดการผลิตโดยอิงจากกลุ่มของ Lot of Product ที่ได้จากกระบวนการ Clustering จนถึงการแสดงผลการพยากรณ์ตามโมเดลที่สร้างขึ้น ซึ่งมีเมนูการทำงานดังนี้



รูปที่ 4.14 เมนูการทำงานในกระบวนการ Prediction

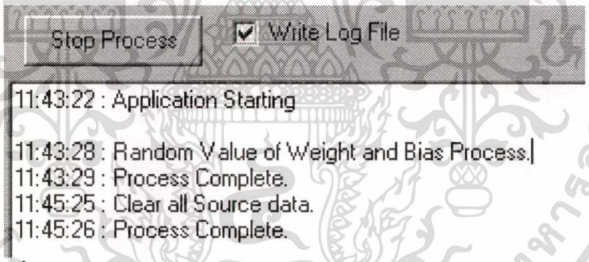
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Random Weight and Bias Value คือการในการสุ่มค่า Weight และ Bias ของ Backpropagation neural network ซึ่งจะใช้เป็นค่าเริ่มต้นของโมเดลที่ยังไม่เคยมีการ Learning โดยแสดงผลการทำงานดังรูป



รูปที่ 4.15 ผลการทำงานเมื่อเลือกเมนู Random Weight and Bias Value

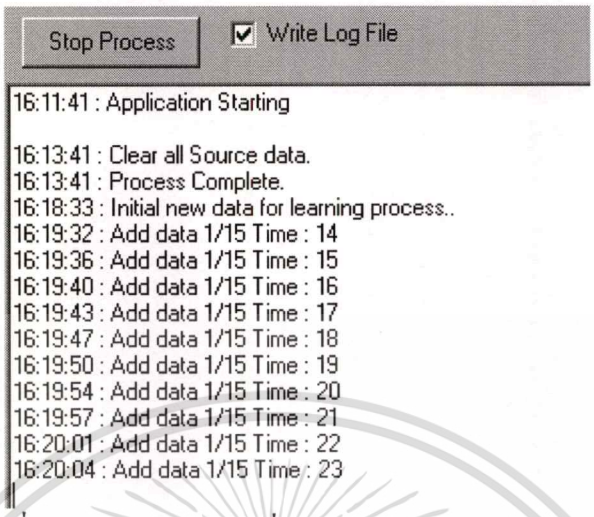
- Reset source data คือกระบวนการล้างข้อมูลเดิมที่จะนำมา leaning ที่มีอยู่เดิม มักจะใช้หลังจากมีการทำ Clustering ข้อมูลขึ้นมาใหม่และกลุ่มของ Lot of Product เปลี่ยนแปลง โดยแสดงผลการทำงาน ดังรูป



รูปที่ 4.16 ผลการทำงานเมื่อเลือกเมนู Reset Source data

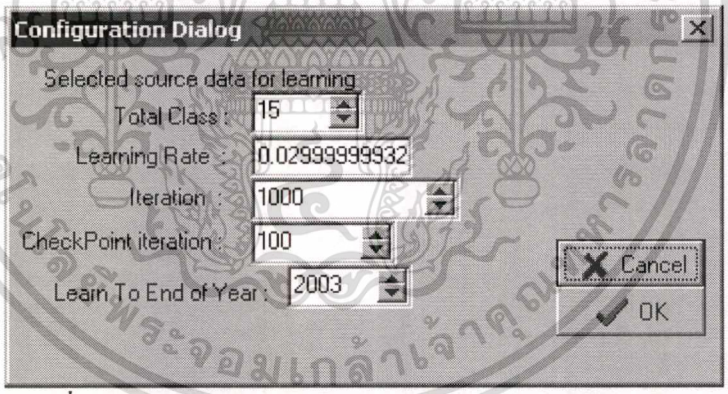
- Initial Source data คือการแปลงข้อมูล Segment ของ Lot of Product ให้อยู่ในรูปของข้อมูลที่จะใช้ในการ learning โดยแสดงผลการทำงานดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.17 ผลการทำงานเมื่อเลือกเมนู Initial Source data

- Backpropagation Configuration คือการจัดเตรียมค่าที่จะใช้ในการ learning ซึ่งเมื่อเลือกเมนูดังกล่าว จะแสดง ไดอะล็อก ดังรูป



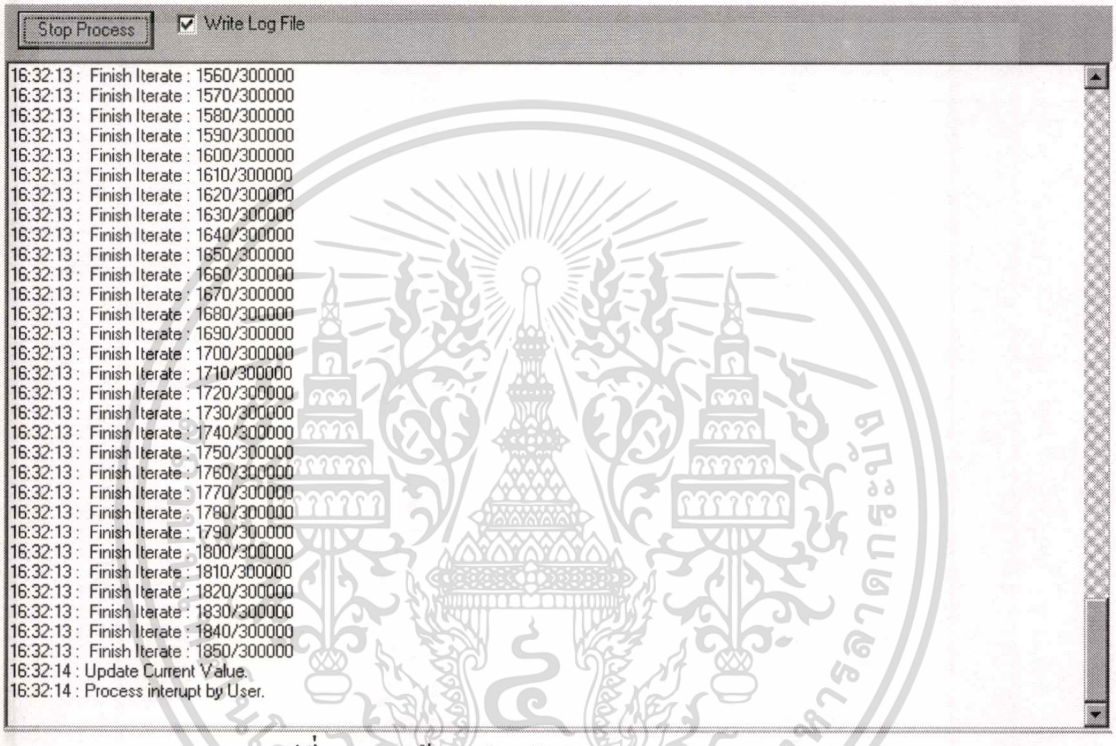
รูปที่ 4.18 ไดอะล็อกแสดงการ configuration neural network

จากรูป ผู้ใช้จะต้อง

- เลือกจำนวน class ที่จะสร้างโมเดลเน็ตเวิร์ค
- กำหนด Learning Rate ซึ่งควรจะมีค่ามากกว่า 0 และน้อยกว่า 1
- เลือกจำนวนรอบที่จะทำการ learning ซึ่งจากรูป เลือก 1000 รอบ โดยในการ learning จะถือเอาข้อมูลใน table ที่จะใช้ learn ที่เกี่ยวข้องเป็นทั้งหมด 1 รอบ
- เลือกรอบที่จะมีการบันทึก Check Point Data ซึ่งจากรูป จะมีการบันทึกเมื่อผ่านไปทีละ 100 รอบ โดยจะบันทึกดังตาราง
- เลือกปีสุดท้ายที่จะทำการ learning โดยจากรูป จะ learning ถึงปี 2003

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

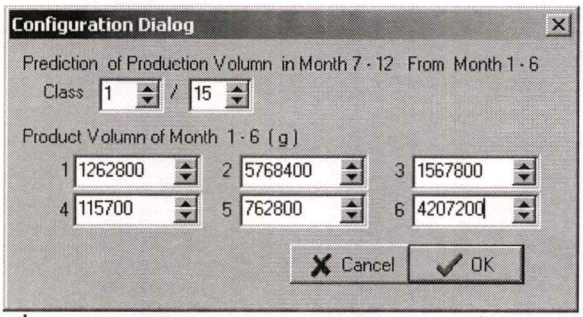
- Learning Process เป็นการ Learning ข้อมูลตามค่าที่กำหนดใน Configuration โดยจะ load ค่า weight และ bias ของโมเดลที่ต้องการจากข้อมูลเดิมที่มีอยู่ ถ้าไม่พบก็จะกำหนดค่าเริ่มต้นให้ตามค่าที่ Random ขึ้นมา โดยจะแสดงผล ดังรูป



รูปที่ 4.19 หน้าจอสำหรับแสดงผลของการ learning

จากรูปแสดงการ learning โดยจำนวนรอบทั้งหมด 300,000 รอบ และทำไปแล้ว 1,850 รอบ

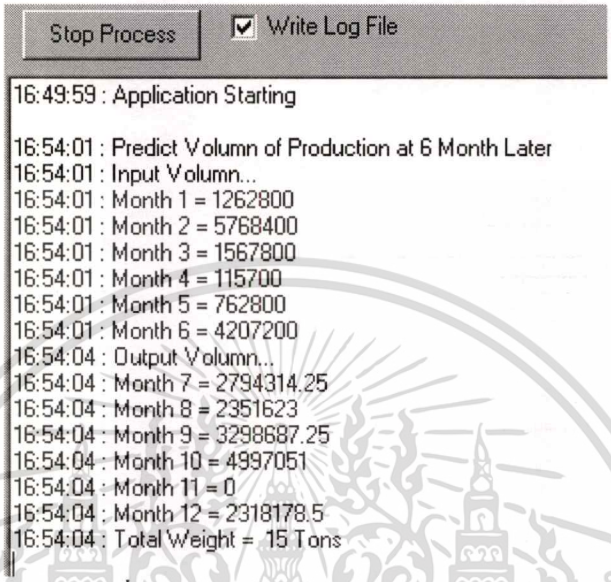
- Predict Value เป็นการนำโมเดลที่ผ่านการ Learning มาพยากรณ์ยอดผลิต 6 เดือน โดยจะแสดงไดอะล็อกเพื่อให้ผู้ใช้เลือกกลุ่ม และข้อมูลการผลิต 6 เดือนก่อนหน้า ดังรูป



รูปที่ 4.20 ไดอะล็อกแสดงการกรอกข้อมูลเพื่อพยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

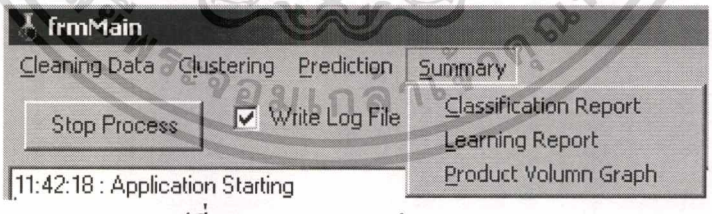
เมื่อคลิกที่ปุ่ม ก็จะนำโมเดลของกลุ่มที่เลือกมาพยากรณ์แล้วแสดงผลดังรูป



รูปที่ 4.21 หน้าจอแสดงผลการพยากรณ์

4.5 Summary

ใช้สำหรับสร้างรายงานสรุปผลการทดสอบการแบ่งกลุ่มและการ learning เพื่อสร้างโมเดลการพยากรณ์ ตลอดจนสร้างกราฟแสดงการพยากรณ์ ซึ่งมีเมนูการทำงานดังรูป



รูปที่ 4.22 แสดงเมนูย่อย Summary

- Classification Report ใช้สร้างรายการผลการจัดกลุ่ม โดยจะแสดงจำนวน lot ในแต่ละกลุ่ม น้ำหนักรวม , เฉลี่ย และการกระจายในแต่ละกลุ่ม โดยจะแสดงรายงานดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Classification Summary Report

Class	Number of Lot	Sum of Weight (kg .)	Average of Weight (kg.)	Standard Var.
1	1,747	145,681.900	83.390	61.644
2	15,538	800,977.548	51.550	59.088
3	705	29,207.610	41.429	35.467
4	11,785	494,304.969	41.944	54.738
5	1,527	89,913.310	58.882	67.169
6	2,101	102,653.742	48.859	54.022
7	6,085	398,979.271	65.568	78.961
8	2,806	160,179.093	57.084	54.102
9	4,244	231,068.170	54.446	49.916
10	384	14,011.260	36.488	23.550
11	2,041	172,288.962	84.414	78.585
12	9,086	402,163.749	44.262	59.277
13	489	33,470.000	68.446	45.325
14	13,852	848,802.690	61.277	59.755
15	6,558	523,502.847	79.827	77.718

รูปที่ 4.23 รายงานแสดงผลการ Classification

- Learning Report ใช้สร้างรายการผลการเรียนรู้ในแต่ละกลุ่ม โดยจะแสดงจำนวนรอบของการ Learn และค่าเปอร์เซนต์ของ Sum of Squire Error โดยคำนวนจากสูตร

  % ของ Sum of Squire Error = Sum of Squire Error x 100 / จำนวนข้อมูลที่นำมา Learn

  โดยจะแสดงรายงานดังรูป

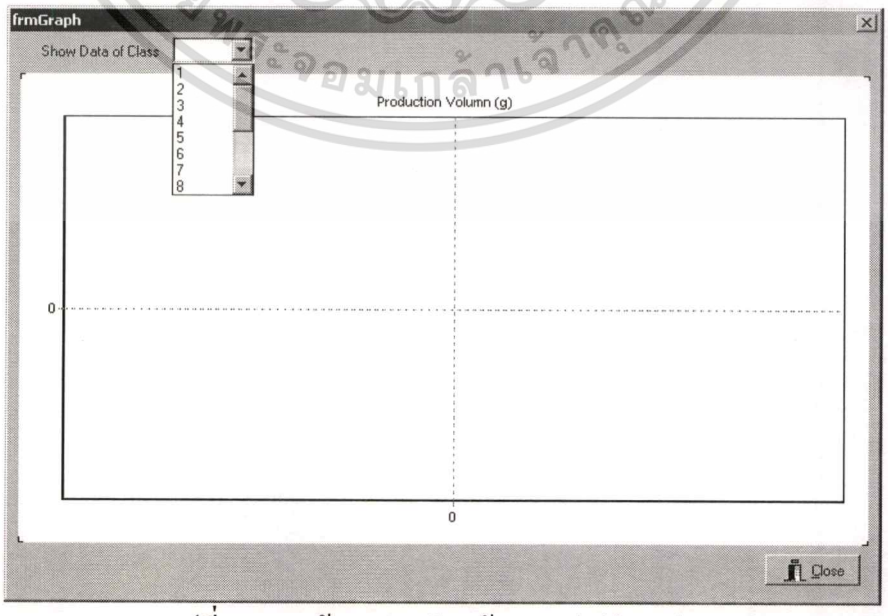
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Backpropagation Learning Report : % Sum of Squire Error

Loop of Learn	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Class 12	Class 13	Class 14	Class 15
50,000	6.27	4.28	8.69	5.35	9.47	8.85	7.63	9.24	7.15	8.33	7.66	7.19	6.09	3.94	6.99
100,000	5.26	3.69	7.03	4.00	7.24	7.17	6.16	7.54	4.97	7.09	6.12	5.12	4.50	3.21	5.76
150,000	4.87	3.35	6.31	3.63	6.69	6.54	5.79	6.93	4.40	6.74	5.57	4.39	4.46	2.85	4.78
200,000	4.64	3.10	6.15	3.49	6.42	6.30	5.65	6.65	4.06	6.52	5.01	4.08	4.45	2.66	4.53
250,000	4.42	2.92	6.02	3.42	6.25	6.17	5.53	6.45	3.84	6.43	4.48	3.95	4.44	2.54	4.35
300,000	4.28	2.76	5.93	3.36	6.12	6.07	5.45	6.40	3.70	6.37	4.14	3.89	4.44	2.44	4.27
350,000	4.17	2.64	5.86	3.31	5.93	6.01	5.40	6.39	3.62	6.32	3.96	3.83	4.44	2.37	4.20
400,000	4.10	2.57	5.78	3.25	5.74	5.96	5.36	6.38	3.56	6.27	3.86	3.78	4.43	2.32	4.14
450,000	4.05	2.51	5.70	3.18	5.61	5.92	5.31	6.37	3.51	6.20	3.78	3.74	4.43	2.28	4.10
500,000	4.01	2.46	5.64	3.11	5.52	5.89	5.27	6.37	3.46	6.14	3.70	3.70	4.43	2.23	4.08
550,000	3.97	2.41	5.60	3.05	5.45	5.85	5.23	6.37	3.42	6.10	3.63	3.67	4.43	2.18	4.07
600,000	3.95	2.37	5.55	3.00	5.40	5.81	5.20	6.36	3.38	6.07	3.58	3.65	4.43	2.13	4.06
650,000	3.93	2.33	5.52	2.97	5.35	5.76	5.17	6.36	3.34	6.05	3.53	3.62	4.43	2.09	4.05
700,000	3.92	2.28	5.49	2.94	5.31	5.73	5.14	6.36	3.32	6.03	3.48	3.60	4.43	2.06	4.05
750,000	3.90	2.24	5.46	2.92	5.28	5.71	5.11	6.35	3.29	6.01	3.43	3.58	4.42	2.04	4.04
800,000	3.89	2.20	5.43	2.89	5.25	5.69	5.07	6.35	3.27	5.99	3.38	3.57	4.42	2.01	4.03
850,000	3.88	2.16	5.39	2.86	5.23	5.68	4.98	6.35	3.25	5.97	3.35	3.56	4.42	1.98	4.02
900,000	3.87	2.12	5.36	2.84	5.20	5.66	4.94	6.34	3.24	5.94	3.33	3.55	4.42	1.95	4.02
950,000	3.87	2.09	5.34	2.82	5.19	5.65	4.92	6.34	3.22	5.90	3.31	3.54	4.42	1.93	4.01
1,000,000	3.86	2.06	5.32	2.81	5.17	5.63	4.90	6.33	3.21	5.87	3.30	3.53	4.42	1.90	4.00

รูปที่ 4.24 รายงานแสดงผลการ Learning

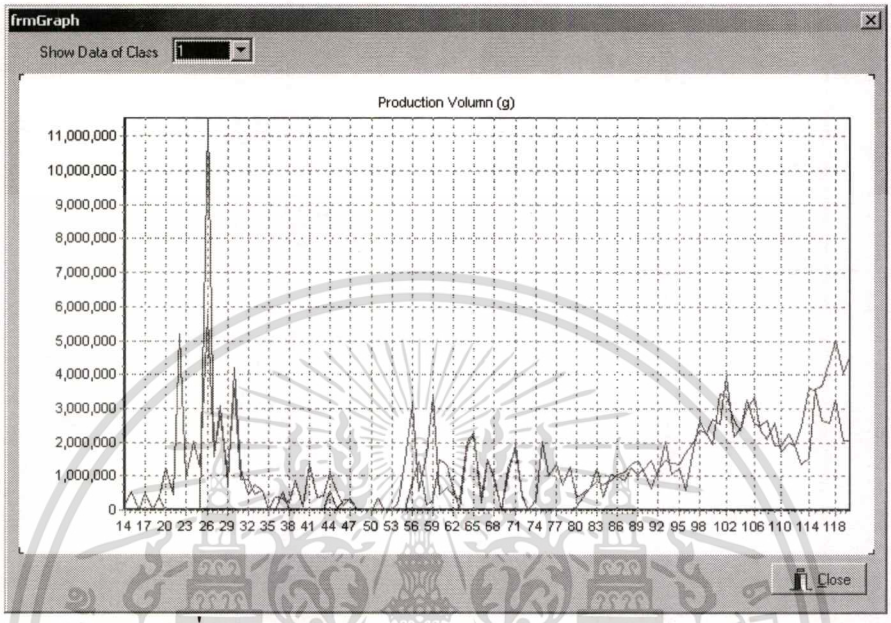
- Product Volumn Graph ใช้สร้างกราฟเปรียบเทียบยอดการผลิตกับยอดพยากรณ์ โดยสร้างจากข้อมูลทั้งหมด (ทั้งที่ผ่านการ learn และที่ใช้ทดสอบ) ซึ่งจะแสดงไดอะล็อก ดังรูป



รูปที่ 4.25 หน้าจอสำหรับสร้างกราฟเปรียบเทียบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป ผู้ใช้จะต้องเลือกกลุ่มที่จะแสดงกราฟ แล้วรอผลการแสดงกราฟ ซึ่งจะแสดงดังรูป



รูปที่ 4.26 แสดงกราฟยอดการผลิตและการพยากรณ์

จากรูป เส้นสีแดงแทนปริมาณการผลิตจริงในกลุ่มที่ 1 ส่วนเส้นสีเขียวคือ ปริมาณการผลิตที่ได้จากการพยากรณ์ตามโมเดลของนิวรอล เน็ตเวิร์คที่ได้จัดเตรียมไว้สำหรับกลุ่มที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุป

5.1 สรุปผลการจัดกลุ่ม

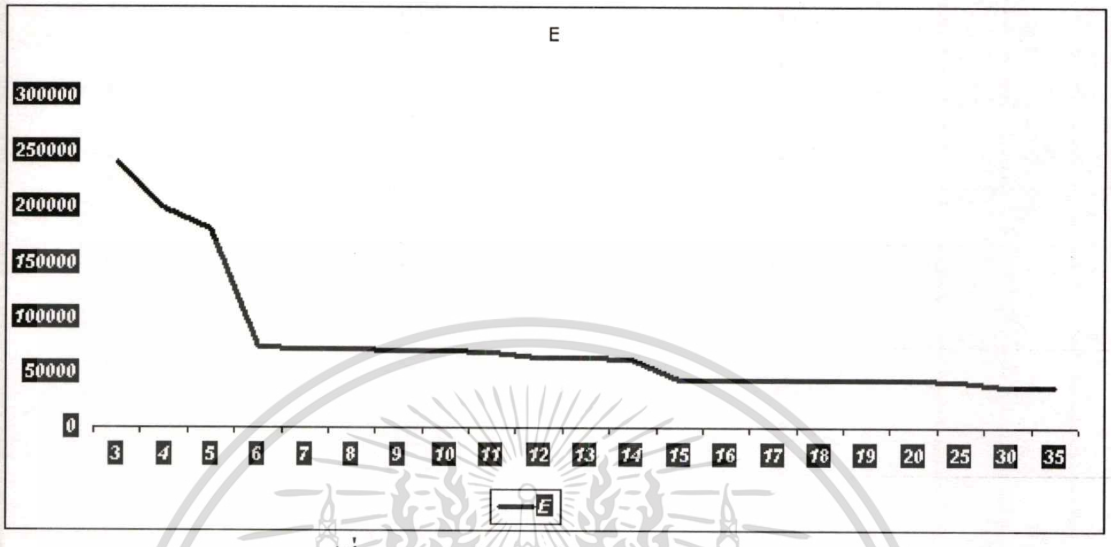
จากการจัดกลุ่มโดยวิธี k-mean โดยใช้ค่า k ต่าง ๆ กัน แสดงค่า sum of square error ดังตาราง

ตารางที่ 5.1 ตารางแสดงค่า Sum of Squire error

TotSegment	NumRec	SumErr_2
3	500	240561.766
4	500	198737.859
5	500	180991.266
6	500	73504.1797
7	500	71564.9609
8	500	70362.1406
9	500	69417.2109
10	500	68874.7656
11	500	68470.1094
12	500	64199.8555
13	500	63821.4688
14	500	63233.7461
15	500	43630.2852
16	500	43577.6367
17	500	43556.1758
18	500	43526.0781
19	500	43395.3633
20	500	43390.3086
25	500	41653.7344
30	500	38435.2891
35	500	37426.5781
0	0	0

จากตาราง สามารถแสดงกราฟได้ดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.1 กราฟแสดงค่า Sum of Squre Error

จากรูป พบว่าเมื่อกำหนดค่า k = 6 และ 15 จะทำให้ค่า sum of square error ลดลงอย่างเห็นได้ชัด จึงเลือกจำนวนกลุ่ม 15 กลุ่ม แล้วตรวจสอบดูการกระจายของ Lot of Product โดยแสดงได้ดังตาราง

ตารางที่ 5.2 ตารางแสดงการกระจายของ Lot ในแต่ละกลุ่ม

ClassID	Count Lot	%	Sum of Weight	%
1	1,747	2.21	145,681,900	3.28
2	15,538	19.68	800,977,548	18.01
3	705	0.89	29,207,610	0.66
4	11,785	14.93	494,304,969	11.11
5	1,527	1.93	89,913,310	2.02
6	2,101	2.66	102,653,742	2.31
7	6,085	7.71	398,979,271	8.97
8	2,806	3.55	160,179,093	3.60
9	4,244	5.38	231,068,170	5.20
10	384	0.49	14,011,260	0.32
11	2,041	2.59	172,288,962	3.87
12	9,086	11.51	402,163,749	9.04
13	489	0.62	33,470,000	0.75
14	13,852	17.55	848,802,690	19.09
15	6,558	8.31	523,502,847	11.77
Total	**78,948**	**100.00**	**4,447,205,121**	**100.00**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 สรุปผลการพยากรณ์

เมื่อกำหนดให้จำนวนของ Cluster เป็น 15 แล้วแยกยอดการผลิตในแต่ละเดือนไปทำ Time series แล้วนำไปผ่านกระบวนการ learning โดยแยกแต่ละโมเดล โดยในการ learning จะทำการวนซ้ำโมเดลละ 1,000,000 รอบ และมีการตรวจสอบค่า Sum of Squire error ของแต่ละโมเดล ทุก ๆ 50,000 รอบ ได้ดังตาราง

ตารางที่ 5.3 ตารางแสดงค่า Sum of Squire error ในแต่ละกลุ่ม

Loop	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
50000	5.583	3.897	7.563	4.92	7.768	8.051	6.944	8.406	6.361	7.084	6.667	6.616	1.705	3.502	6.360
100000	4.685	3.36	6.118	3.676	5.936	6.525	5.609	6.86	4.426	6.03	5.328	4.712	1.261	2.858	5.238
150000	4.334	3.048	5.492	3.340	5.484	5.951	5.267	6.306	3.916	5.729	4.85	4.041	1.249	2.538	4.348
200000	4.133	2.823	5.351	3.211	5.264	5.736	5.144	6.053	3.616	5.542	4.36	3.750	1.246	2.364	4.122
250000	3.937	2.659	5.236	3.142	5.129	5.613	5.032	5.872	3.419	5.463	3.894	3.634	1.245	2.259	3.962
300000	3.811	2.515	5.161	3.095	5.019	5.523	4.963	5.827	3.296	5.417	3.598	3.578	1.243	2.174	3.882
350000	3.708	2.403	5.095	3.047	4.861	5.467	4.915	5.811	3.222	5.376	3.444	3.521	1.242	2.107	3.824
400000	3.649	2.335	5.029	2.992	4.706	5.425	4.874	5.803	3.167	5.331	3.354	3.476	1.242	2.062	3.764
450000	3.606	2.286	4.963	2.928	4.602	5.390	4.833	5.799	3.123	5.274	3.286	3.44	1.241	2.027	3.734
500000	3.572	2.239	4.911	2.860	4.527	5.358	4.795	5.795	3.083	5.220	3.223	3.408	1.241	1.988	3.717
550000	3.529	2.192	4.868	2.805	4.468	5.325	4.762	5.792	3.044	5.186	3.161	3.379	1.240	1.942	3.705
600000	3.513	2.156	4.833	2.763	4.424	5.286	4.731	5.79	3.008	5.161	3.111	3.354	1.24	1.897	3.697
650000	3.499	2.120	4.803	2.732	4.388	5.244	4.701	5.787	2.977	5.142	3.068	3.332	1.239	1.863	3.69
700000	3.486	2.079	4.776	2.706	4.355	5.216	4.676	5.784	2.951	5.125	3.026	3.313	1.239	1.836	3.683
750000	3.474	2.036	4.749	2.682	4.328	5.197	4.653	5.781	2.928	5.109	2.982	3.297	1.239	1.812	3.676
800000	3.465	2.1884	4.723	2.658	4.306	5.18	4.611	5.778	2.909	5.092	2.944	3.284	1.238	1.788	3.669
850000	3.456	1.966	4.691	2.635	4.286	5.166	4.531	5.774	2.894	5.073	2.916	3.273	1.238	1.763	3.662
900000	3.448	1.933	4.667	2.615	4.268	5.152	4.496	5.770	2.88	5.047	2.896	3.265	1.238	1.738	3.655
950000	3.441	1.902	4.646	2.599	4.252	5.139	4.475	5.765	2.866	5.018	2.882	3.257	1.238	1.715	3.648
1000000	3.434	1.874	4.626	2.585	4.238	5.126	4.461	5.76	2.854	4.991	2.871	3.25	1.237	1.695	3.640

จากตาราง พบว่าค่า Sum of Squire Error จะลดลงเรื่อย ๆ เมื่อมีการ learning จำนวนรอบมากขึ้น ซึ่งแสดงว่าข้อมูลที่ได้จากการพยากรณ์ ใกล้เคียงค่าข้อมูลจริงมากขึ้น โดยในแต่ละกลุ่มจะมีค่าแตกต่างกันไป ตามลักษณะของข้อมูล และเมื่อนำผลการ learning มาคำนวณหาเปอร์เซนต์ของ Squire Error ตามสูตร

% of Squire Error = Sum of Squire Error x 100 / จำนวนข้อมูลที่ใช้ Learn ในกลุ่ม

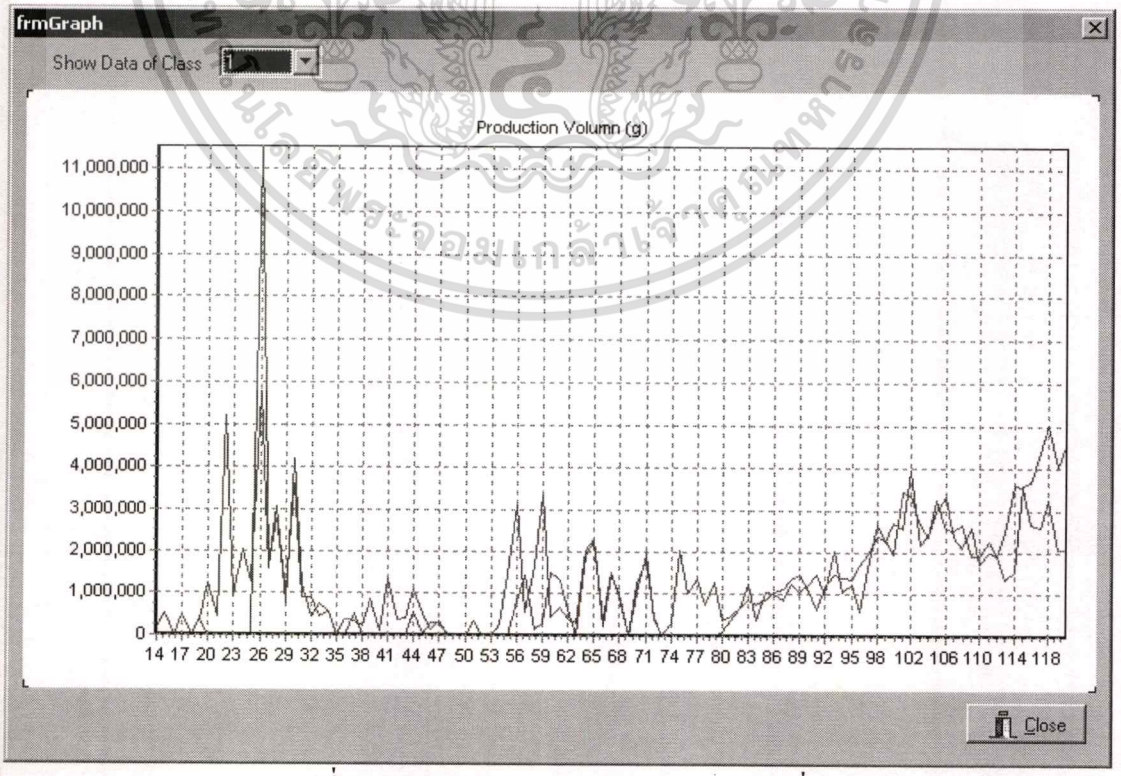
เมื่อ Learn ครบ 1,000,000 รอบ ซึ่งแสดงได้ดังตาราง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.4 ตารางแสดงค่าเปอร์เซนต์เปรียบเทียบ

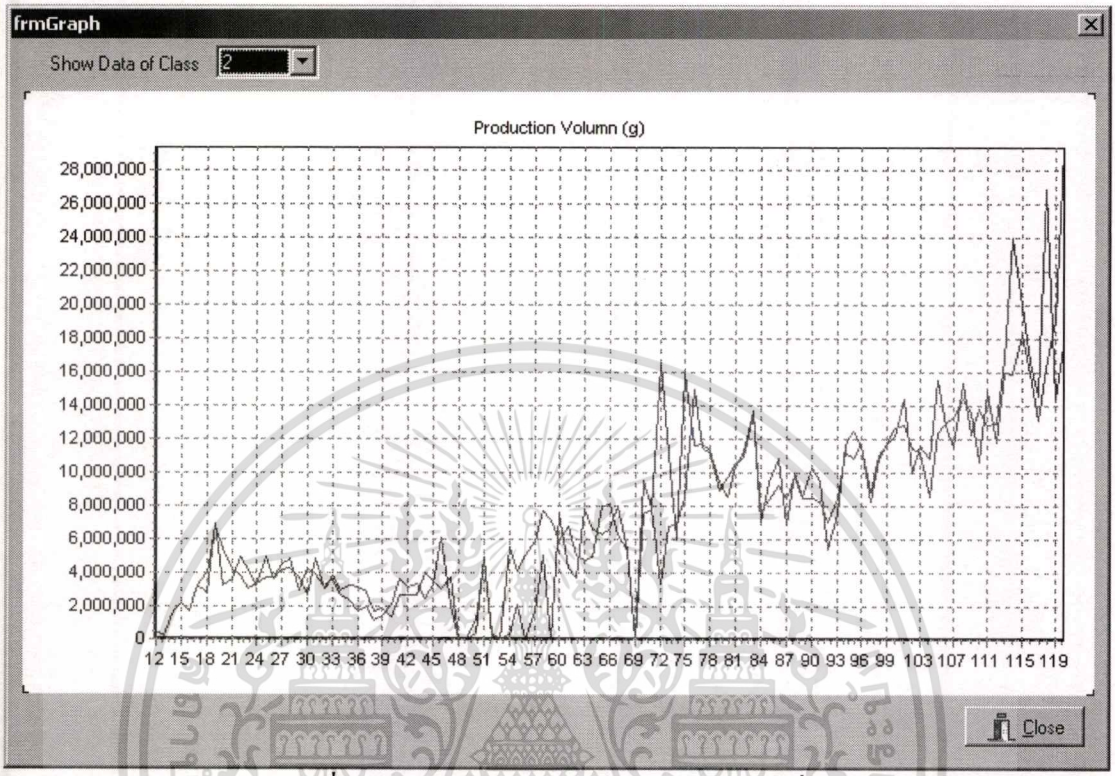
ClassID	Train	Test	% of Test	Err_2	%of Err
1	89	7	7.29	3.43	3.86
2	91	7	7.14	1.87	2.06
3	87	7	7.45	4.63	5.32
4	92	7	7.07	2.59	2.81
5	82	7	7.87	4.24	5.17
6	91	7	7.14	5.13	5.63
7	91	7	7.14	4.46	4.90
8	91	7	7.14	5.76	6.33
9	89	7	7.29	2.85	3.21
10	85	7	7.61	4.99	5.87
11	87	7	7.45	2.87	3.30
12	92	7	7.07	3.25	3.53
13	28	7	20.00	1.24	4.42
14	89	7	7.29	1.70	1.90
15	91	7	7.14	3.64	4.00

และเมื่อแสดงผลของการ learning ในรูปของกราฟเปรียบเทียบของผลการผลิตจริง กับผลที่เกิดจากการพยากรณ์ จะแสดงได้ดังรูป

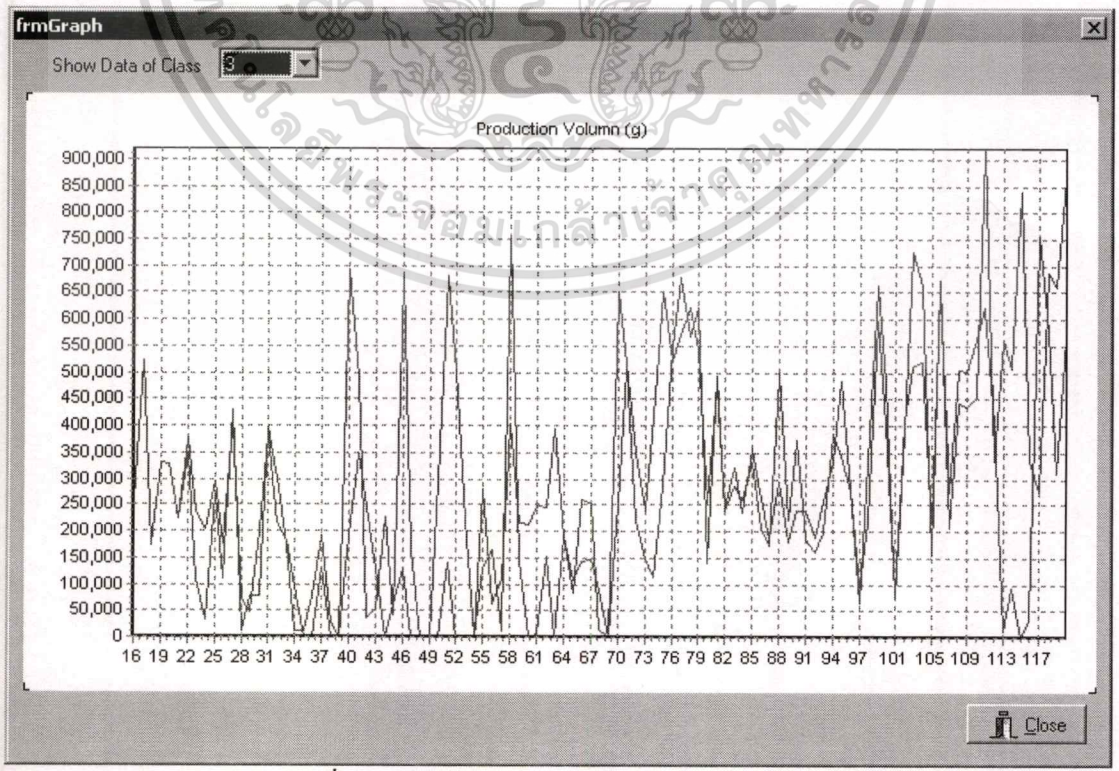


รูปที่ 5.2 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

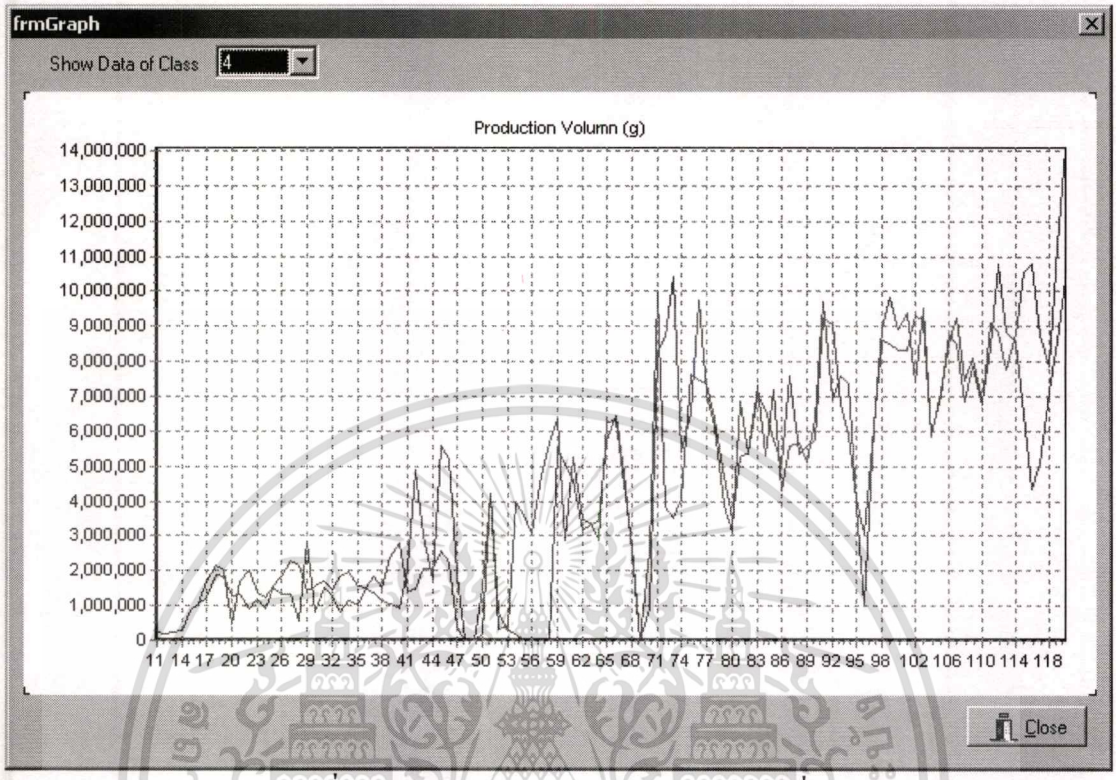


รูปที่ 5.3 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 2

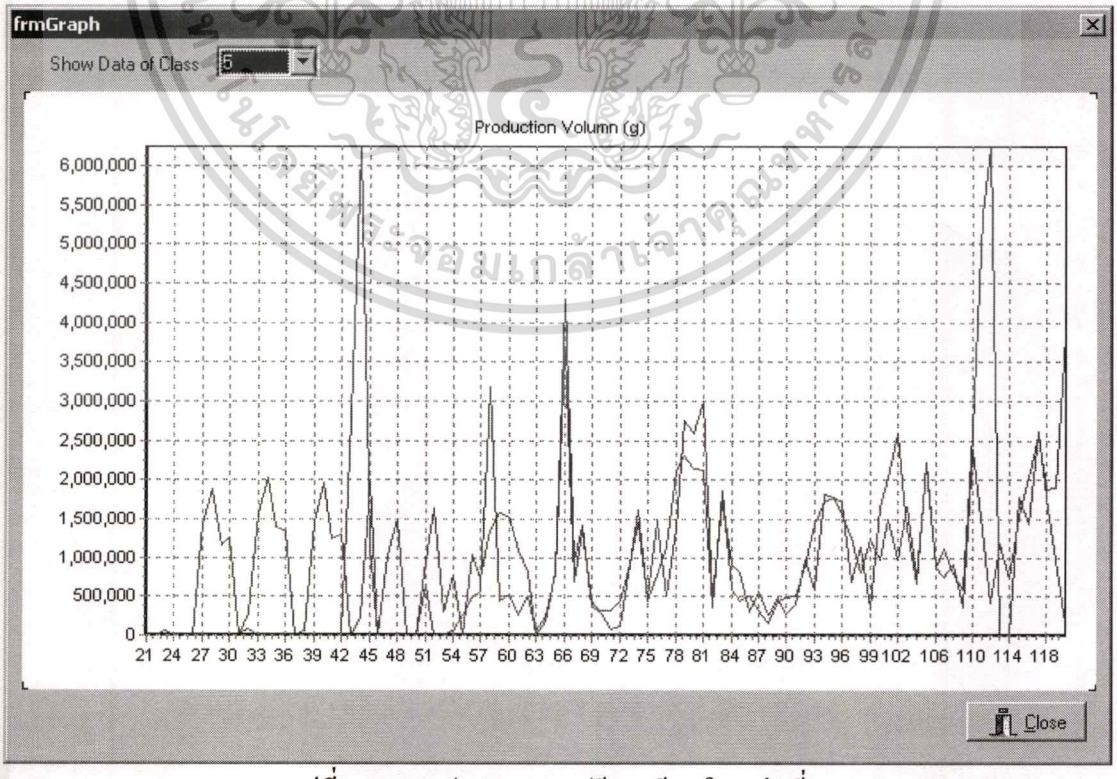


รูปที่ 5.4 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

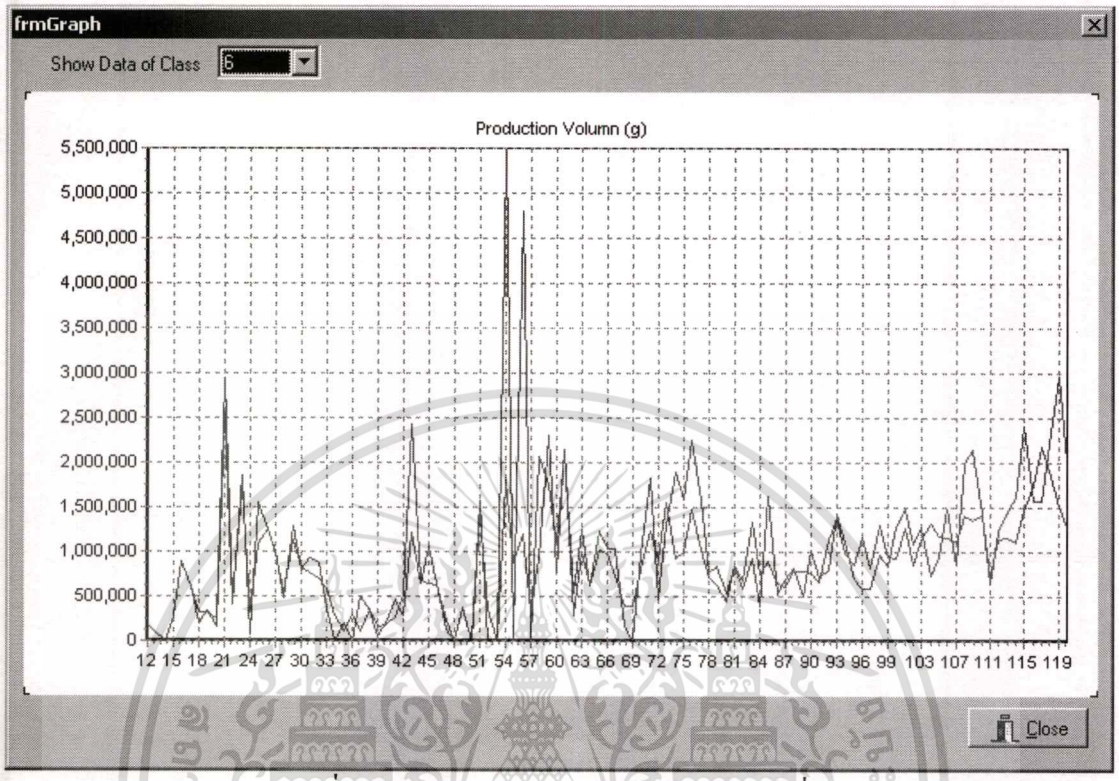


รูปที่ 5.5 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 4

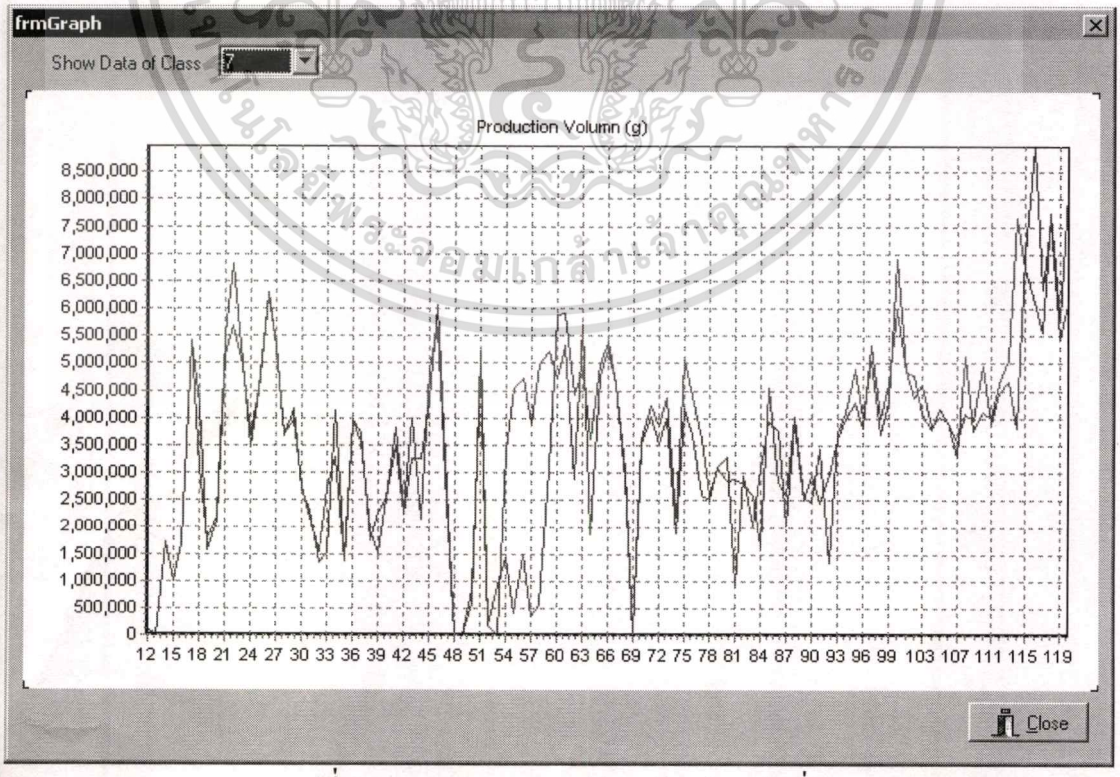


รูปที่ 5.6 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

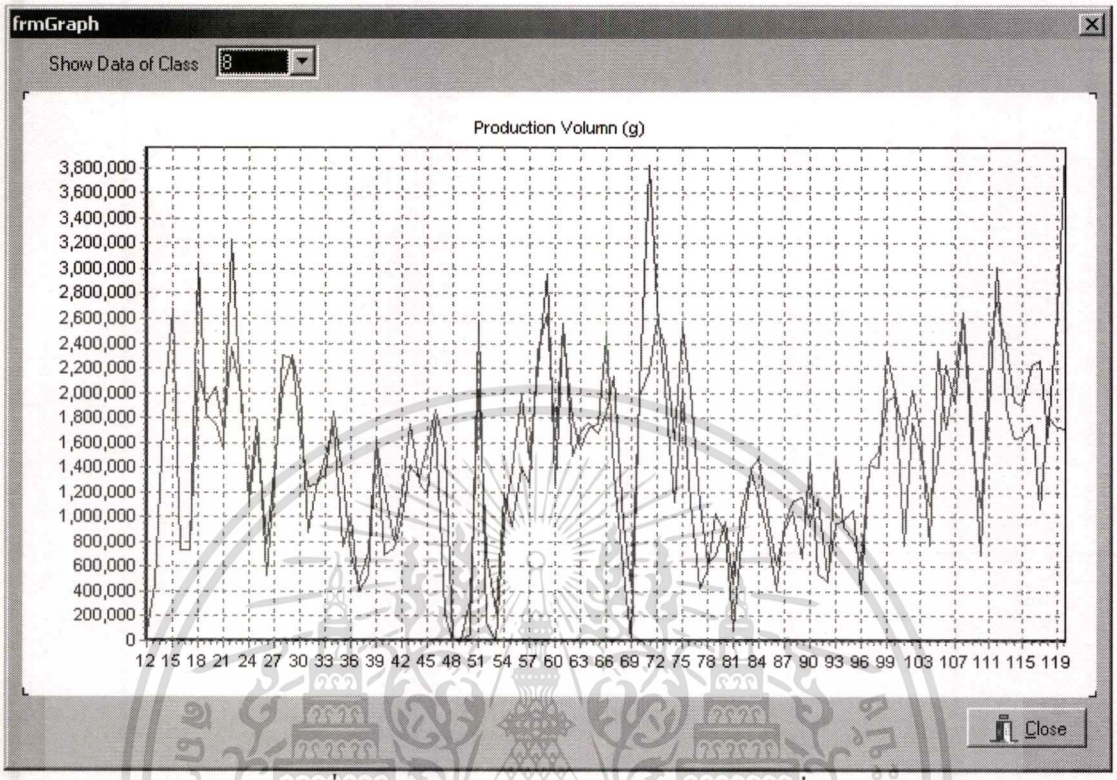


รูปที่ 5.7 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 6

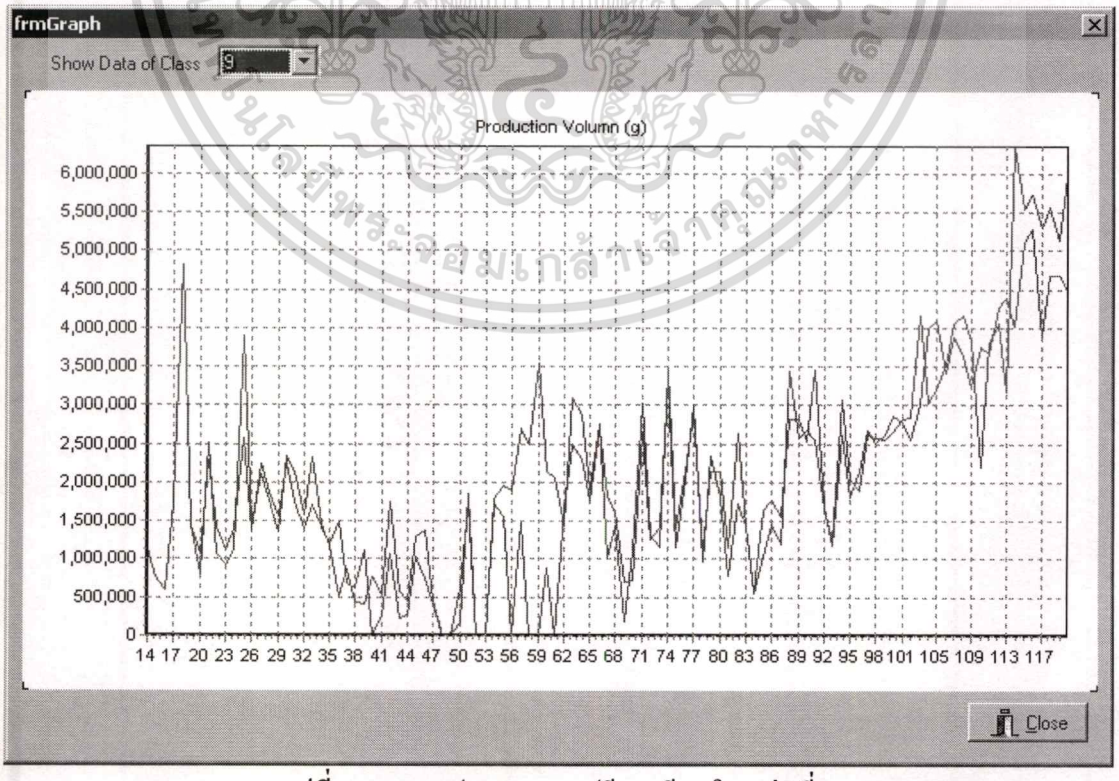


รูปที่ 5.8 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

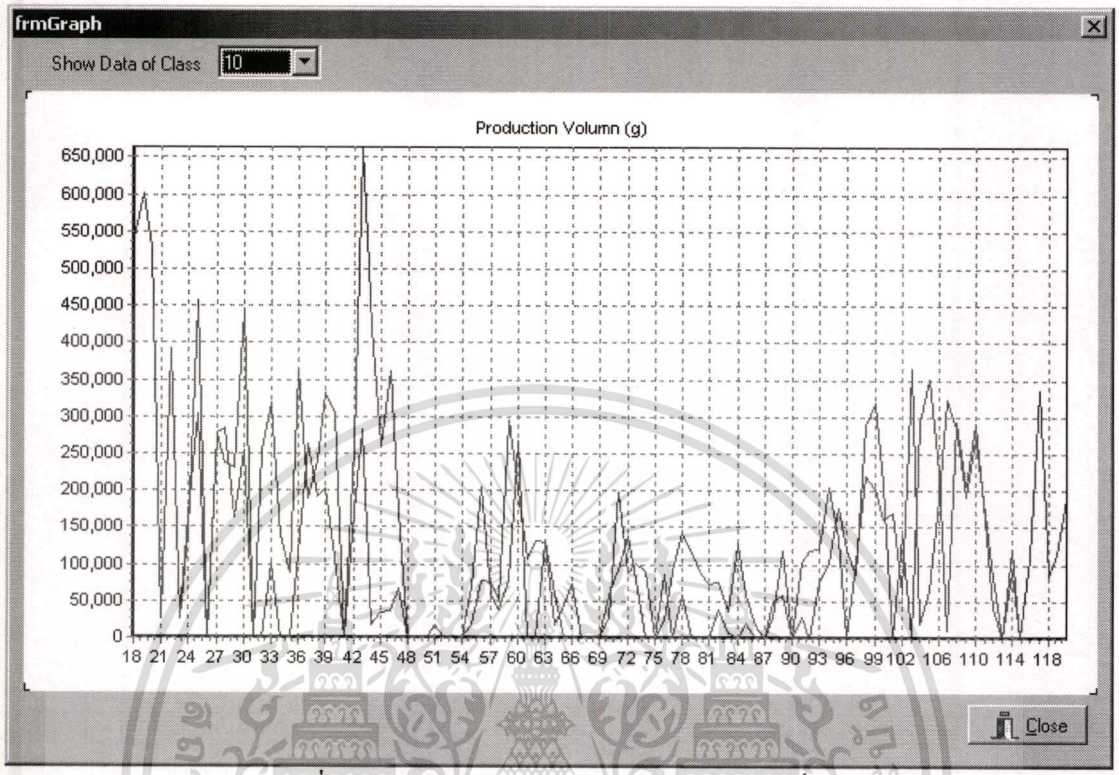


รูปที่ 5.9 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 8

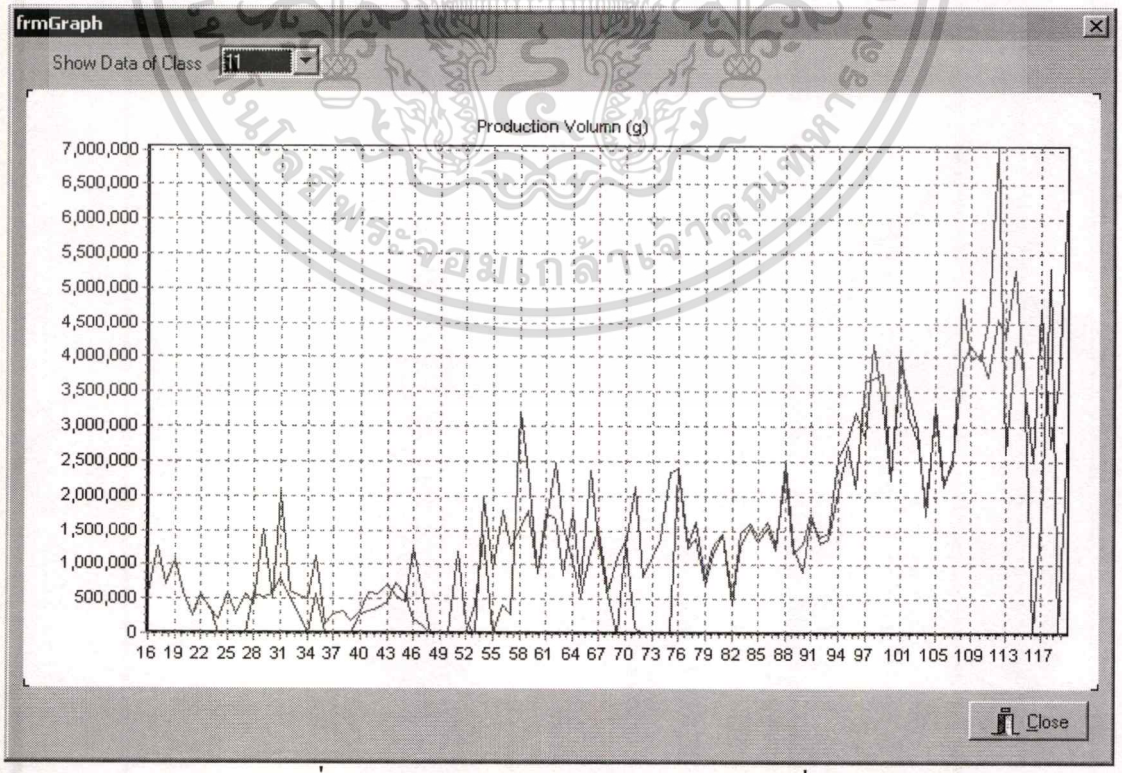


รูปที่ 5.10 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

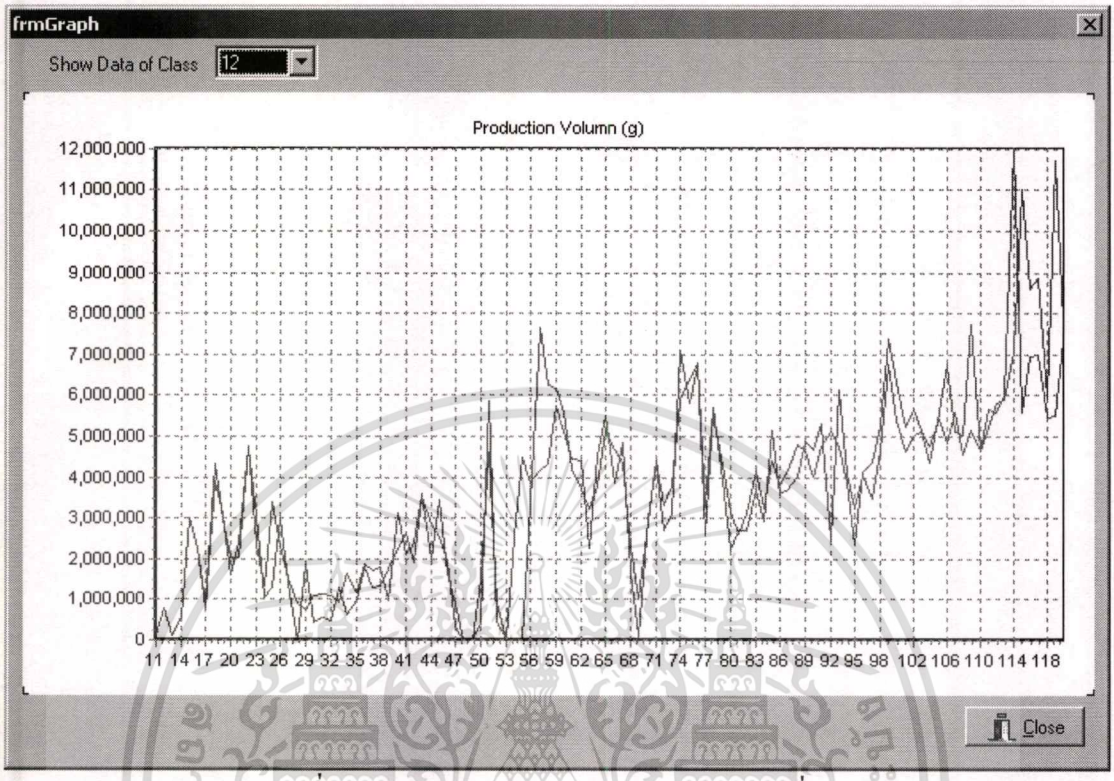


รูปที่ 5.11 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 10

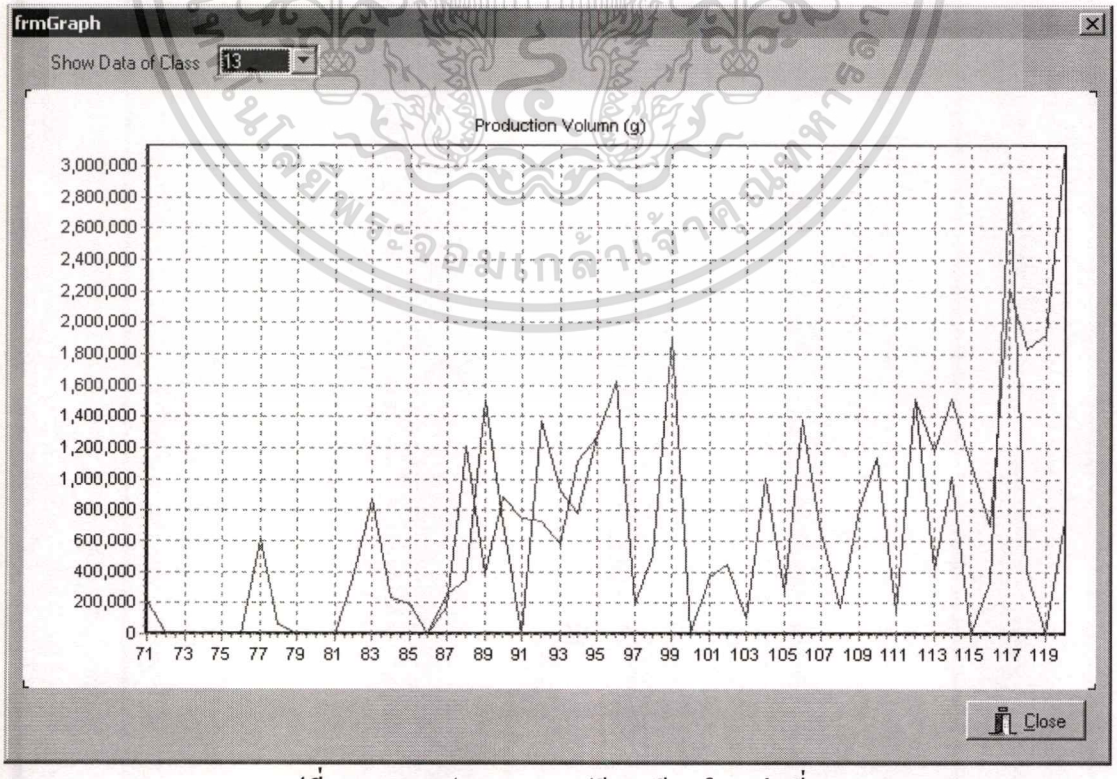


รูปที่ 5.12 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

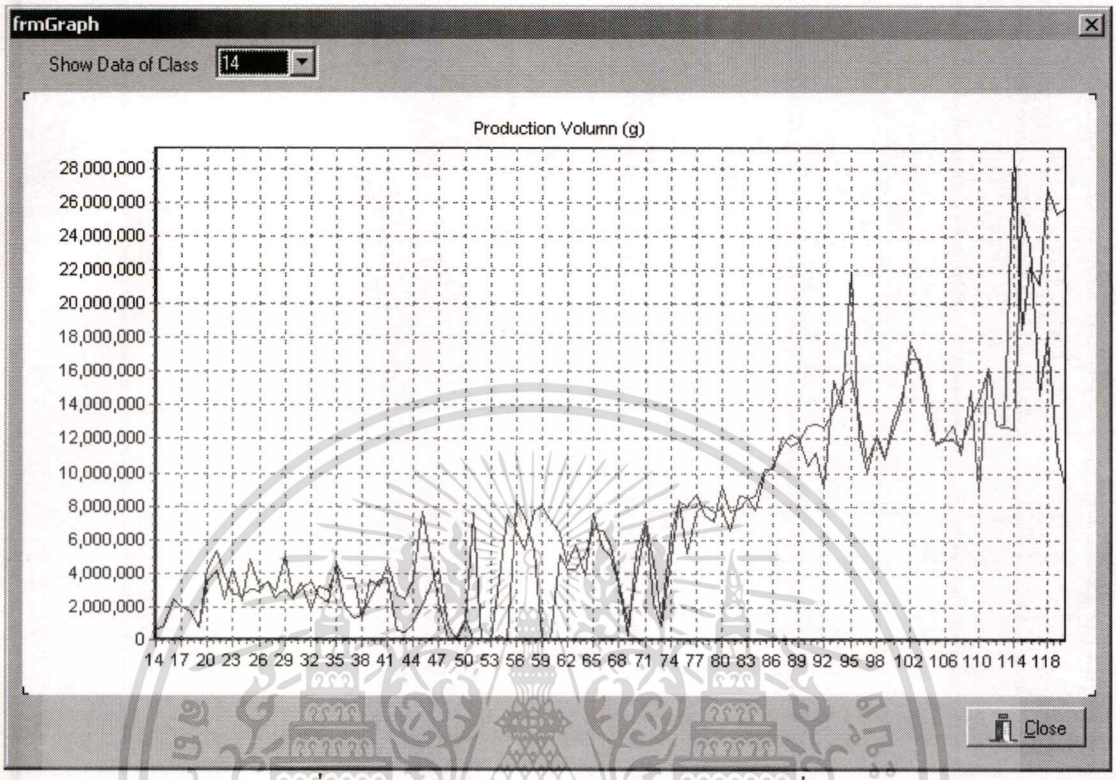


รูปที่ 5.13 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 12

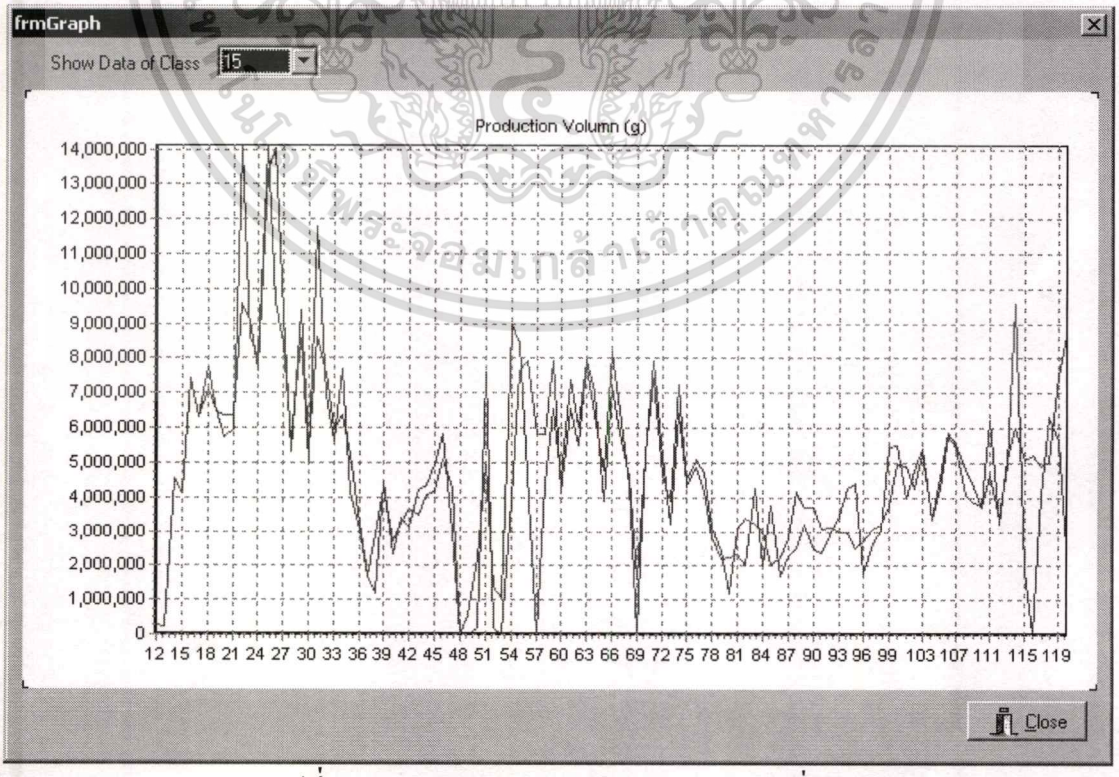


รูปที่ 5.14 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.15 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 14



รูปที่ 5.16 กราฟแสดงการเปรียบเทียบในกลุ่มที่ 15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Jiawei Han ,and Micheline Kamber.2001. **Datamining: Concepts and Techniques.** New York : Morgan Kaufmann Publishers.

Maureen , Caudill and Charless , Butler. 1992. **Understanding Neural Networks.** London : The MIT Press.

Peter ,Cabena. 1997 . **Discovering data mining : from concept to implementation.** New Jersey: Prentice Hall PTR.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้