

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

การพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes
A Development tool of Data Segmentation by K-Prototypes Algorithm

โดย

กิติภัทร กิติกุล

รหัสประจำตัว 46066722

วัน เดือน ปี..... 21 ก.พ. 2550
เลขทะเบียน..... 02325
เลขเรียกหนังสือ..... อท. ก ๖๙๓ ก 2548
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

อาจารย์ที่ปรึกษา

ผศ.ดร.วรพจน์ กรีสुरเดช

611 707070
11 28 ๖7๒๕๐



H002325

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 1 ปีการศึกษา 2548

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes
นักศึกษา	นายกิติภัทร กิติกุล
อาจารย์ที่ปรึกษา	ผศ.ดร.วรพจน์ กรีสระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2548

บทคัดย่อ

ในปัจจุบันข้อมูลต่างๆทางธุรกิจมีขนาดใหญ่และซับซ้อนขึ้นทำให้ยากที่จะคาดคะเนหรือตัดสินใจ จึงได้เกิดแนวคิดการจัดกลุ่มข้อมูลตามลักษณะหรือคุณสมบัติที่คล้ายคลึงกัน เพื่อให้ได้ข้อมูลใหม่ๆ ช่วยในการพยากรณ์ข้อมูลที่น่าจะเกิดขึ้นในอนาคตอันใกล้ สามารถใช้ช่วยสนับสนุนในการตัดสินใจของผู้บริหารได้ ซึ่งในโครงการพัฒนาระบบนี้ได้นำอัลกอริทึม K-Prototypes ซึ่งสามารถจัดกลุ่มข้อมูลได้ทั้งที่เป็นตัวเลขและไม่ใช่ตัวเลขมาใช้ในการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูล

Title	A Development tool of Data Segmentation by K-Prototypes Algorithm
Student	Mr. Kitipat Kitikul
Advisor	Asst.Prof.Dr.Worapoj Kreesuradej
Level of Study	Master of Science in Information Technology
Major	Information Science
Academic Year	2005

ABSTRACT

At present, business data have large size and more complexity that hard to forecast and make decision. This is the initiative to study for data segmentation that classified by characteristic, or attribute which is resembled. It will be gathering new data as advantageous to forecast data which will be happen in the near future, and will be help to support decision-making for the executive. In this project brings K-Prototypes algorithm which can use with numerical and categorical in data segmentation.

กิตติกรรมประกาศ

ในการทำโครงการเรื่องการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes (A Development tool of Data Segmentation by K-Prototypes Algorithm) เป็นระบบที่มีการทำงานหลายขั้นตอนโดยมีความช่วยเหลือจากบุคคลหลาย ๆ ท่าน ทำให้การพัฒนาระบบนั้นสำเร็จลุล่วงไปได้ด้วยดี ข้าพเจ้าจึงขอขอบพระคุณ มา ณ. ที่นี้ด้วย

ข้าพเจ้าต้องขอขอบพระคุณ ผศ.ดร.วราพงษ์ กรีสระเดช อาจารย์ที่ปรึกษาวิชาโครงการพัฒนาระบบงานที่กรุณาให้คำแนะนำและเป็นที่ปรึกษา อันเป็นประโยชน์ต่อการพัฒนาระบบ รวมทั้งเป็นผู้ตรวจสอบความถูกต้องของโครงการฉบับนี้

ขอขอบพระคุณ คุณพ่อ คุณแม่ และบุคคลในครอบครัว ที่ได้ให้ความสนับสนุนทางด้านกำลังใจในการเรียนจนการทำโครงการพัฒนาระบบนี้สำเร็จด้วยดี รวมทั้งขอขอบคุณพี่ๆและเพื่อนๆ IS 16.1 ทุกคนที่ให้ความช่วยเหลือในด้านต่างๆ ที่เกี่ยวกับโครงการไว้ ณ. ที่นี้



สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมาของโครงการ.....	1
1.2 วัตถุประสงค์ของโครงการ.....	2
1.3 ขอบเขตการศึกษา.....	2
1.4 ผลที่คาดว่าจะได้รับ.....	2
1.5 ผลการดำเนินงาน.....	2
1.6 อุปกรณ์และซอฟต์แวร์ที่คาดว่าจะใช้.....	3
2. ทฤษฎีที่เกี่ยวข้อง.....	5
2.1 ความหมายของค้ำ数据挖掘.....	5
2.2 วิวัฒนาการของเทคโนโลยีฐานข้อมูล.....	6
2.3 ลักษณะเฉพาะของข้อมูลที่สามารถทำค้ำ数据挖掘.....	7
2.4 งานของค้ำ数据挖掘 (Task of Data Mining).....	7
2.5 การเตรียมข้อมูลสำหรับทำค้ำ数据挖掘.....	9
2.5.1 การคัดเลือกข้อมูล (Data Selection).....	9
2.5.2 การแก้ไขข้อมูล (Data Cleaning).....	9
2.5.3 การปรับเปลี่ยนข้อมูล (Data Transformation).....	10
2.6 เนื้อหาและหลักการของอัลกอริทึม K-Prototypes.....	11

สารบัญ (ต่อ)

หน้า

2.6.1 การวิเคราะห์ห้วงของข้อมูลใน K-Prototypes	11
2.6.2 อัลกอริทึม K-Prototypes	12
2.6.3 ตัวอย่างการทำงานของอัลกอริทึม K-Prototypes กับข้อมูล	17
3. การออกแบบระบบงาน	22
3.1 ความต้องการของระบบ (Functional Requirement).....	22
3.2 องค์ประกอบของระบบงาน	23
3.3 ขั้นตอนการทำงานของระบบ	24
3.4 การออกแบบหน้าจการทำงานจากระบบ	27
3.4.1 การติดต่อกับฐานข้อมูลที่ต้องการวิเคราะห์.....	27
3.4.2 การเลือกตารางในฐานข้อมูลที่ต้องการวิเคราะห์	28
3.4.3 การเลือกฟิลด์ในฐานข้อมูลที่ต้องการวิเคราะห์	29
3.4.4 การแก้ไขข้อมูล (Data Cleaning)	29
3.4.5 การปรับเปลี่ยนข้อมูล (Data Transformation)	34
3.4.6 การกำหนดจำนวนกลุ่มที่ต้องการจัดกลุ่ม	34
3.4.7 การแสดงผลการจัดกลุ่ม	35
4. การประยุกต์ใช้โปรแกรม	37
4.1 การกำหนดวัตถุประสงค์	37
4.2 การเตรียมข้อมูลสำหรับทำค้ำไม่นิ่ง	37
4.3 การทำค้ำไม่นิ่งในการจัดกลุ่มข้อมูล	40
4.4 การวิเคราะห์ผลที่ได้จากการทำค้ำไม่นิ่งและการนำความรู้มาใช้	42
5. สรุปผลการศึกษาและข้อเสนอแนะ	44
5.1 สรุปผลการศึกษา	44
5.2 ประโยชน์ที่ได้จากการศึกษาและพัฒนาระบบ	44
5.3 ข้อเสนอแนะ	44

สารบัญ (ต่อ)

หน้า

บรรณานุกรม	46
ภาคผนวก ก คู่มือการใช้งาน โปรแกรม	47
ประวัติผู้เขียน	71



สารบัญตาราง

หน้า

ตารางที่

1.1 แผนการดำเนินงานของการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes.....	3
2.1 แสดงชนิดข้อมูลที่จะนำมาทำไมนิ่ง.....	17
2.2 แสดงข้อมูลที่จะนำมาทำ ไมนิ่ง.....	17
2.3 แสดงข้อมูลในกลุ่มที่ 1.....	18
2.4 แสดงข้อมูลในกลุ่มที่ 2.....	18
2.5 แสดงข้อมูลในกลุ่มที่ 3.....	18
2.6 แสดงจุดศูนย์กลางของข้อมูล 3 กลุ่ม.....	19
2.7 แสดงการเปรียบเทียบข้อมูลกับจุดศูนย์กลางกลุ่มที่ 1.....	19
2.8 แสดงการเปรียบเทียบข้อมูลกับจุดศูนย์กลางกลุ่มที่ 2.....	19
2.9 แสดงการเปรียบเทียบข้อมูลกับจุดศูนย์กลางกลุ่มที่ 3.....	20
2.10 แสดงข้อมูลในกลุ่มที่ 1 หลังจากมีการย้ายกลุ่ม.....	20
2.11 แสดงข้อมูลในกลุ่มที่ 3 หลังจากมีการย้ายกลุ่ม.....	20
2.12 สรุปจุดศูนย์กลางของกลุ่มทั้ง 3 กลุ่มหลังถูก update.....	21
4.1 ข้อมูล customer ที่ทำการเลือกมาจัดกลุ่ม.....	37

สารบัญรูป

หน้า

รูปที่

2.1	วิวัฒนาการเทคโนโลยีฐานข้อมูล	6
2.2	กระบวนการเริ่มต้นจัดกลุ่มข้อมูล	13
2.3	กระบวนการจัดกลุ่มข้อมูลใหม่	15
2.4	กราฟแสดงการทำงานของอัลกอริทึม K-Prototypes	16
3.1	ผังงานกิจกรรมการทำงานหลักของระบบ	26
3.2	หน้าจอสำหรับเลือกข้อมูล	27
3.3	เลือกฐานข้อมูลที่ต้องการนำมาวิเคราะห์	28
3.4	เลือกตารางในฐานข้อมูลที่ต้องการวิเคราะห์	28
3.5	เลือกฟิลด์ในฐานข้อมูลที่ต้องการวิเคราะห์	29
3.6	การแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Numerical ไม่มีค่า null	30
3.7	การแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ไม่มีค่า null	31
3.8	การแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Numerical ที่มีค่า null	32
3.9	การแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ที่มีค่า null	33
3.10	การปรับเปลี่ยนข้อมูล	34
3.11	กำหนดจำนวนกลุ่มที่ใช้ในการจัดกลุ่ม	35
3.12	การแสดงผลการจัดกลุ่ม	36
4.1	หน้าจอเลือกฟิลด์ข้อมูลที่ต้องการจัดกลุ่ม	38
4.2	หน้าจอการแก้ไขข้อมูล	39
4.3	หน้าจอการปรับเปลี่ยนข้อมูล	40
4.4	หน้าจอการกำหนดจำนวนกลุ่มข้อมูล	41
4.5	หน้าจอแสดงผลการจัดกลุ่มข้อมูล	42
ก.1	หน้าจอแรกเมื่อเข้าสู่การติดตั้งโปรแกรม	48
ก.2	หน้าจอที่สองของการติดตั้งโปรแกรม	49

สารบัญรูป (ต่อ)

หน้า

รูปที่

ก.3 หน้าจอแสดงกระบวนการติดตั้งระบบ	49
ก.4 หน้าจอแสดงการติดตั้ง โปรแกรมเสร็จสมบูรณ์	50
ก.5 แสดงไคเรกทอรีที่โปรแกรมได้ถูกติดตั้ง	50
ก.6 การเข้าสู่โปรแกรมโดยเลือกจาก Start Menu	51
ก.7 การเข้าสู่โปรแกรมโดยเลือกจาก Shortcut	51
ก.8 หน้าจอเริ่มต้นของ โปรแกรม	52
ก.9 หน้าจอแสดงผล เมนู File	52
ก.10 หน้าจอเลือกฐานข้อมูล	53
ก.11 หน้าจอเลือกตารางและฟิลด์	54
ก.12 หน้าจอแสดงผลการแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Numerical ที่ไม่มีค่า Null	55
ก.13 หน้าจอแสดงผลการแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ที่ไม่มีค่า Null	56
ก.14 หน้าจอแสดงผลการแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Numerical ที่มีค่า Null	57
ก.15 หน้าจอแสดงผลการแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ที่มีค่า Null	58
ก.16 หน้าจอแสดงผลการปรับเปลี่ยนข้อมูล	59
ก.17 หน้าจอแสดงผลการกำหนดจำนวนกลุ่มข้อมูล	60
ก.18 หน้าจอแสดงผลการจัดกลุ่มข้อมูล	61
ก.19 หน้าจอก่อนทำการดูผลสรุปแบบกราฟ	61
ก.20 หน้าจอแสดงผลสรุปแบบกราฟ	62
ก.21 หน้าจอแสดงผลโดยเลือกเมนู File > Export Text	62
ก.22 หน้าจอแสดงผลโดยเลือกคลิกที่ปุ่ม Export Text	63
ก.23 หน้าจอแสดงผลการบันทึกข้อมูลแบบ Text	63
ก.24 หน้าจอแสดงผลข้อมูลเมื่อบันทึกข้อมูลแบบ Text	64
ก.25 หน้าจอแสดงผลโดยเลือกเมนู File > Export Excel	64

สารบัญรูป (ต่อ)

หน้า

รูปที่

ก.26 หน้าจอแสดงผลโดยเลือกคลิกที่ปุ่ม Export Excel	65
ก.27 หน้าจอแสดงผลการบันทึกข้อมูลแบบ Excel	65
ก.28 หน้าจอแสดงผลข้อมูลเมื่อบันทึกข้อมูลแบบ Excel จาก Worksheet ชื่อ Center	66
ก.29 หน้าจอแสดงผลข้อมูลเมื่อบันทึกข้อมูลแบบ Excel จาก Worksheet ชื่อ Member	66
ก.30 หน้าจอแสดงผลโดยเลือกเมนู File > Export XML	67
ก.31 หน้าจอแสดงผลการบันทึกข้อมูลแบบ XML	67
ก.32 หน้าจอแสดงผลการบันทึกข้อมูลแบบ XML	68
ก.33 หน้าจอแสดงผลข้อมูลเมื่อบันทึกข้อมูลแบบ XML	68
ก.34 หน้าจอการปิดโปรแกรมโดยเลือกเมนู File > Quit	69
ก.35 หน้าจอแสดงการปิดโปรแกรมโดยเลือกที่ปุ่ม Quit	69
ก.36 ข้อความเตือนเมื่อฟิลด์ไม่มีข้อมูลใดๆอยู่	70
ก.37 ข้อความแสดงเตือนเมื่อเลือกเฉพาะฟิลด์ประเภทใดประเภทหนึ่งเท่านั้น	70

บทที่ 1

บทนำ

1.1 ความเป็นมาของโครงการ

ในปัจจุบันนี้เป็นโลกของธุรกิจที่จำเป็นต้องพึ่งพิงข้อมูลข่าวสารที่เป็นประโยชน์ต่อตัวองค์กรเอง ทำให้การเก็บข้อมูลนั้นมีขนาดใหญ่มาก ซึ่งข้อมูลที่เป็นสื่อบันทึกมีราคาถูกลง และความเร็วในการประมวลผลของเครื่องคอมพิวเตอร์มีมากขึ้น จึงเป็นการเปิดโอกาสให้มีการสืบค้นและใช้ประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่เหล่านี้

ซึ่งแนวทางหนึ่งในการพัฒนาคำนี้คือค้ำไม่นิ่ง (Data Mining :Knowledge Discovery in Database : KDD) เป็นการแสวงหาความรู้และรูปแบบพฤติกรรมจากข้อมูลดิบที่สะสมอยู่ในฐานข้อมูล (Unknown Pattern in Database : Hidden Predictive Information) โดยที่ผลที่ได้จากการทำค้ำไม่นิ่งจะไปช่วยในเรื่องระบบสนับสนุนการตัดสินใจของผู้บริหารในแต่ละสาขา โดยเฉพาะอย่างยิ่งงานในด้านธุรกิจ ซึ่งจำเป็นต้องมีการแข่งขันกับคู่แข่งทางด้านการค้า (Competitor) ในตัวของสารสนเทศเองนั้นลักษณะจะคล้ายกับทรัพยากรในเมืองแร่ที่ถูกขุดขึ้นมาเพื่อใช้งาน ทั้งนี้เพราะการดำเนินงานในส่วนของการประมวลผลธุรกรรม (Transaction Processing) แต่เพียงอย่างเดียวนั้นไม่เพียงพอในการนำไปใช้งานเพราะการประมวลผลธุรกรรมเปรียบเสมือนเปลือกนอกของสารสนเทศ ซึ่งไม่ได้ถึงส่วนที่เป็นสาระสำคัญที่อยู่ภายในออกมาใช้ประโยชน์ได้

1.2 วัตถุประสงค์ของโครงการ

การจัดทำโครงการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes ซึ่งมีวัตถุประสงค์ต่างๆดังนี้

1. เพื่อเป็นการได้ปฏิบัติงานจริง หลังจากที่ได้ศึกษาทฤษฎีของการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes
2. เพื่อการนำระบบคอมพิวเตอร์มาช่วยวิเคราะห์ในการจัดกลุ่มข้อมูล
3. เพื่อทราบจำนวนกลุ่มข้อมูลหลังจากที่ได้ทำการแบ่งกลุ่มเรียบร้อยแล้ว

1.3 ขอบเขตการศึกษา

1. ทำการจัดกลุ่มฐานข้อมูลโดยใช้อัลกอริทึม K-Prototypes
2. ข้อมูลที่สามารถทำการวิเคราะห์ในการจัดกลุ่มข้อมูลต้องเป็นฐานข้อมูลของ Microsoft Access 2000 , Access XP หรือ Access 2003 เท่านั้น
3. สามารถส่งออกไฟล์ในรูปแบบของเท็กซ์ไฟล์(TXT) เอกสาร Microsoft Excel (XLS) และเอกสาร eXtensible Markup Language (XML)
4. สามารถแสดงผลสรุปข้อมูลในรูปแบบของกราฟได้

1.4 ผลที่คาดว่าจะได้รับ

1. ได้เครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes
2. สามารถนำแนวทางด้านค่าใดมาประยุกต์กับการดำเนินงาน
3. สามารถแปลงข้อมูลดิบจากฐานข้อมูล กลายเป็นข้อมูลที่มีประโยชน์ได้
4. สามารถจัดกลุ่มเพื่อทำนายขนาดของพฤติกรรมหรือการประเมินค่าที่จะเกิดขึ้นในอนาคตได้
5. มีความสะดวก รวดเร็ว แม่นยำ ในการจัดกลุ่มข้อมูลที่มีขนาดใหญ่จากฐานข้อมูล

1.5 แผนการดำเนินงาน

ในการดำเนินงานของการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes จะต้องมีการศึกษาทฤษฎีและเทคโนโลยีต่างๆที่ใช้ การสำรวจและรวบรวมข้อมูล การออกแบบโปรแกรม เขียนโปรแกรมและทดสอบโปรแกรม รวมถึงการติดตั้งระบบและการจัดทำเอกสารประกอบการใช้งาน โดยแผนการดำเนินงานมีดังนี้

การประมาณการวางแผนสำหรับการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้ อัลกอริทึม K-Prototypes มีระยะเวลาที่ใช้ในการดำเนินงาน ดังแสดงใน ตารางที่ 1.1

ตารางที่ 1.1 แผนการดำเนินงานของการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม

K-Prototypes

ระยะเวลาการดำเนินงาน	พ.ศ. 2548															
	มิถุนายน				กรกฎาคม				สิงหาคม				กันยายน			
	สัปดาห์				สัปดาห์				สัปดาห์				สัปดาห์			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1. การศึกษาทฤษฎีและ เทคโนโลยีที่ใช้	■															
2. การสำรวจและรวบรวม ข้อมูล	■															
3. การออกแบบโปรแกรม	■															
4. การพัฒนา-ทดสอบ																
- การเขียนโปรแกรม																
- การทดสอบย่อย																
- การทดสอบรวม																
5. การทดสอบติดตั้งระบบ																
6. การจัดทำเอกสาร	■															

หมายเหตุ

■ ระยะเวลาในการดำเนินงาน

1.6 อุปกรณ์และซอฟต์แวร์ที่คาดว่าจะใช้

ในการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes ได้นำโปรแกรมหลักที่ใช้ในการเขียนโปรแกรมคือภาษา Microsoft Visual Basic6 โปรแกรม Microsoft Access2003 ซึ่งเป็นระบบฐานข้อมูล โปรแกรม Microsoft Excel ในการออกรายงานโดยสามารถทำกราฟแสดงกลุ่มข้อมูลต่างๆ และมีโปรแกรมเสริมในการช่วยพิมพ์เอกสารคือ Microsoft Word โดยการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes มีเครื่องมือและอุปกรณ์ที่ใช้ในการปฏิบัติงานดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ฮาร์ดแวร์ (Hardware) ที่มีรายละเอียดดังนี้
 - ซีพียู (CPU) : อินเทล เพนเทียม (Intel Pentium 4) 1.50 GHz
 - แรม (RAM) : RDRAM ขนาด 512 MB
 - Hard Disk : ขนาด 40 GB
 - DVD Drive : สำหรับการ ติดตั้งโปรแกรมที่ใช้ในการทำงานในระบบ
 - เมาส์ (Mouse) : PS2
 - เครื่องพิมพ์ (Printer) แบบอิงค์เจ็ท (Inkjet) ในการจัดพิมพ์เอกสารต่างๆ
2. ระบบปฏิบัติการ (Operating System) ที่มีรายละเอียดดังนี้
 - Microsoft Windows XP
3. เครื่องมือในการพัฒนาระบบ (Development Tools) ที่มีรายละเอียดดังนี้
 - ฐานข้อมูล (Database) : Microsoft Access 2003 เป็น โปรแกรมฐานข้อมูลซึ่งสามารถทำงานได้ในระบบ Win32
 - Development Language : Microsoft Visual Basic 6 เป็น โปรแกรมที่ใช้สร้างโปรแกรมประยุกต์สำหรับระบบปฏิบัติการ Windows สามารถทำงานร่วมกับ Microsoft Access และ Microsoft Excel ได้เป็นอย่างดี ซึ่งจะนำมาใช้ในการสร้างแอปพลิเคชันระบบการจัดกลุ่มข้อมูล และสร้างรายงานแสดงผลทางหน้าจอ
 - Software Packages : Installshield X Express เป็น โปรแกรมที่สร้าง Installer สำหรับที่ให้ผู้ผู้ติดตั้งระบบ

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 ความหมายของดาต้าไมนิ่ง

ดาต้าไมนิ่ง (Data Mining) คือการค้นหาความสัมพันธ์และรูปแบบ ซึ่งอยู่ในฐานข้อมูล (Database) แต่ได้ถูกซ่อนไว้ในข้อมูลจำนวนมาก ดาต้าไมนิ่งจะทำการสำรวจและวิเคราะห์อย่างอัตโนมัติหรือกึ่งอัตโนมัติ ในปริมาณข้อมูลจำนวนมากให้อยู่ในรูปแบบที่เต็มไปด้วยความหมายและกฎ (Meaning Patterns And Rules) โดยความสัมพันธ์เหล่านี้แสดงให้เห็นถึงความรู้ต่างๆที่มีประโยชน์ที่เก็บอยู่ในฐานข้อมูล

เทคโนโลยีทางด้านฐานข้อมูลได้ถูกพัฒนาขึ้นอย่างกว้างขวาง จึงเป็นสิ่งที่ไม่ยากนักที่จะเก็บข้อมูลดิบจำนวนมากลงในคอมพิวเตอร์อย่างมีประสิทธิภาพ และยังสามารถดึงข้อมูลกลับออกมาเมื่อใดก็ได้ตามที่ผู้ใช้ต้องการ ในปัจจุบันองค์กรส่วนใหญ่ยังเผชิญกับปัญหาของข้อมูลดิบที่มีจำนวนมากแต่ข้อมูลที่ประยุกต์ใช้ได้นั้นมีน้อย จากปัญหานี้จึงเป็นแรงกระตุ้นให้มีศาสตร์ทางด้านดาต้าไมนิ่ง เนื่องจากดาต้าไมนิ่งสามารถดึงความรู้ออกมาจากฐานข้อมูลจำนวนมากที่ถูกสะสมไว้

ในของธุรกิจบริษัทต่างๆพยายามเสาะหาเทคนิคที่สามารถนำความสำเร็จมาสู่บริษัท เช่น ในธุรกิจขนาดกลางและขนาดย่อมจะสร้างความสัมพันธ์กับลูกค้า โดยสังเกตจากความต้องการ ความชอบ และความสนใจของลูกค้า และอาจจะเรียนรู้ได้จากผลสะท้อนในอดีตว่าจะทำอย่างไรให้การบริการลูกค้านั้นมีประสิทธิภาพมากยิ่งขึ้นไปในอนาคต แต่การเรียนรู้นั้นต้องการมากกว่าการเก็บสะสมข้อมูลจำนวนมากอย่างตรงไปตรงมา

ความเป็นจริงแล้วบริษัทหลายๆแห่งมีการเก็บข้อมูลในฐานข้อมูลจำนวนมากมาขมมหาศาล เพื่อการค้นหาข้อมูลพร้อมทั้งการวิเคราะห์การสังเคราะห์รูปแบบ การออกแบบกฎ การเกิดขึ้นของความคิดใหม่ และการนำสิ่งต่างๆที่เก็บไว้ไปทำนายเกี่ยวกับเหตุการณ์ในอนาคต ซึ่งความสามารถเหล่านี้ถือว่าเป็นข้อได้เปรียบของดาต้าไมนิ่ง ในโครงการขนาดใหญ่สามารถนำดาต้าไมนิ่งไปใช้ใน ระบบ เพื่อจุดประสงค์ในการได้ส่วนแบ่งทางการตลาดเพิ่มมากขึ้น ดาต้าไมนิ่งจะช่วยในการสร้างความเข้าใจในให้กับบริษัทในการตัดสินใจแบ่งกลุ่มของลูกค้าเพื่อออกสินค้า เข้าใจพฤติกรรมลูกค้าที่ซื้อสินค้ากับทางบริษัทและทำนายประชากรที่คาดว่าจะมาเป็นลูกค้าในอนาคตได้

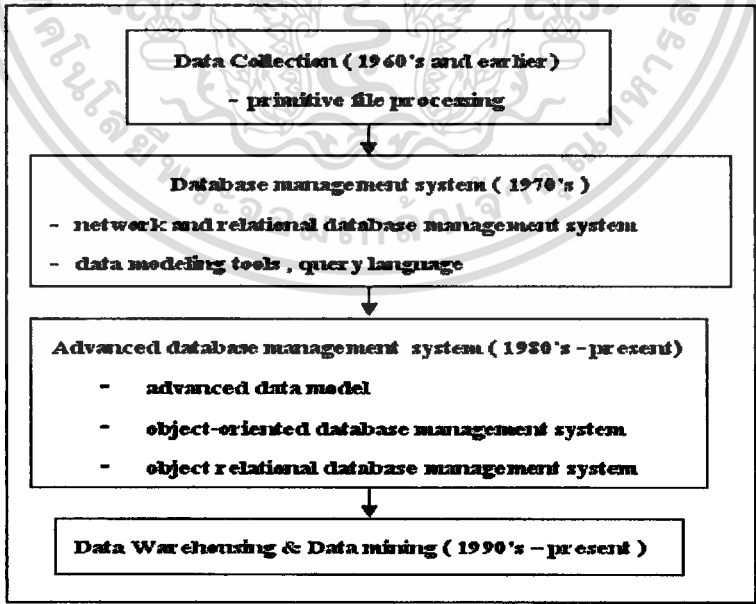
2.2 วิวัฒนาการของเทคโนโลยีฐานข้อมูล

ในปี 1960 เทคโนโลยีฐานข้อมูลได้เริ่มพัฒนามาจาก File processing พื้นฐาน การค้นคว้า และพัฒนาระบบฐานข้อมูลมีมาเรื่อย ๆ

ปี 1970 ได้นำไปสู่การพัฒนากระบวนการเก็บข้อมูลในรูปแบบตาราง (Relational Database System) มีเครื่องมือจัดการโมเดลข้อมูล และมีเทคนิคการใช้อินเด็กซ์และการบริหารข้อมูล นอกจากนี้ผู้ใช้ยังได้รับความสะดวกในการเข้าถึงข้อมูลโดยการใช้ภาษาในการเรียกข้อมูล (Query Language)

ปี 1980 เทคโนโลยีฐานข้อมูลได้เริ่มมีการปรับปรุงและพัฒนาในการหาระบบจัดการที่มี ศักยภาพมากขึ้น ความก้าวหน้าในเทคโนโลยีฮาร์ดแวร์ ใน 30 ปีที่ผ่านมา ได้นำไปสู่การจัดเก็บ ข้อมูลจำนวนมากที่มีความซับซ้อนได้อย่างมีประสิทธิภาพเพิ่มขึ้น

ปี 1990 – ปัจจุบัน สามารถจัดเก็บข้อมูลได้ในหลายรูปแบบแตกต่างกันทั้งระบบปฏิบัติการ หรือการจัดเก็บฐานข้อมูล ซึ่งการนำข้อมูลทั้งหมดมารวมและจัดเก็บไว้ในรูปแบบเดียวกันเรียกว่า Data Warehouse เพื่อความสะดวกในการจัดการ ซึ่งเทคโนโลยี Data Warehouse รวมไปถึง Data Cleaning, Data Integration และ On-Line Analytical Processing (OLAP) เป็นเทคนิคการวิเคราะห์ ข้อมูลในหลาย ๆ มิติ นั้นเกิดขึ้นมาตามลำดับ (สุวิมล คงศักดิ์ตระกูล และสุวิสาข์ โกพล. 2544 : 2-3)



รูปที่ 2.1 วิวัฒนาการเทคโนโลยีฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 ลักษณะเฉพาะของข้อมูลที่สามารถทำค้ำไ่มนึ่ง

ข้อมูลขนาดใหญ่ เกินกว่าจะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่า หรือ โดยการใช้ Database Management System (DBMS) ในการจัดการฐานข้อมูล

ข้อมูลที่มาจกหลายแหล่ง โดยอาจรวบรวมมาจากหลายระบบปฏิบัติการหรือหลาย DBMS เช่น Oracle , DB2 , MS SQL , MS Access เป็นต้น

ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดช่วงเวลาที่ทำการไ่มนึ่ง หากข้อมูลที่มีอยู่นั้นเป็นข้อมูล ที่เปลี่ยนแปลงตลอดเวลาคงต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้นไว้และนำฐานข้อมูลที่ บันทึกไว้มาทำไ่มนึ่ง แต่เนื่องจากข้อมูลนั้นมีการเปลี่ยนแปลงอยู่ตลอดเวลา จึงทำให้ผลลัพธ์ที่ได้ จากการทำไ่มนึ่ง สมเหตุสมผลในช่วงเวลาหนึ่งเท่านั้น ดังนั้นเพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องจึง ต้องทำไ่มนึ่งใหม่ทุกครั้งในการทำไ่มนึ่ง (สุวิมล คงศักดิ์ตระกูล และสุวิสาข์ โกพล. 2544 : 5)

2.4 งานของค้ำไ่มนึ่ง (Task of Data Mining)

ในทางปฏิบัติจริงค้ำไ่มนึ่งจะประสบความสำเร็จกับงานบางกลุ่มเท่านั้น และต้องอยู่ ภายใต้อาไรที่จัด ปัญหาเหมาะสมกับการใช้เทคนิคค้ำไ่มนึ่งจะเป็นปัญหาที่ค้ำไ่มนึ่งให้เหตุผลในการ แก้ ปัญหาที่เกี่ยวข้องกับเศรษฐศาสตร์และการเงิน ซึ่งจะสามารถจัดรูปแบบของธุรกิจให้อยู่ใน รูปแบบของงานค้ำไ่มนึ่ง

1. การจัดกลุ่มข้อมูล (Clustering)

การจัดกลุ่มข้อมูล คือการรวบรวมข้อมูลที่มีความคล้ายคลึงกันไว้ในกลุ่มเดียวกันและ ข้อมูลที่ไม่คล้ายคลึงกันจะอยู่ในกลุ่มอื่น ซึ่งข้อมูลแต่ละตัวจะต้องอยู่ในกลุ่มใดกลุ่มหนึ่งเพียงกลุ่ม เดียว ซึ่งข้อมูลแต่ละกลุ่มที่ได้นั้นจะเรียกว่า “Cluster” นั่นคือลักษณะการแบ่งแยกข้อมูลบนพื้นฐาน ของความคล้ายคลึงกันในตัวอง

วัตถุประสงค์ของการจัดกลุ่มข้อมูล เพื่อจัดกลุ่มข้อมูลซึ่งเป็นประโยชน์ในงานด้าน ต่างๆ เช่น การตลาด การแพทย์ การปกครอง ฯลฯ (กัลยา วาณิชย์บัญชา. 2546 : 125) ดังตัวอย่าง ต่อไปนี้

- ใช้ศึกษาพฤติกรรมผู้บริโภคของกลุ่มผู้บริโภคที่อยู่ต่างกลุ่มกัน ซึ่งจะทำได้ สามารถวางกลยุทธ์ทางการตลาดได้อย่างมีประสิทธิภาพมากขึ้น การที่จะสามารถแยกกลุ่มผู้บริโภค ออกเป็นกลุ่มย่อยได้ จะต้องพิจารณาถึงตัวแปรที่ใช้ในการจัดกลุ่มผู้บริโภค ที่จะทำให้ผู้ที่อยู่ต่าง กลุ่มกันมีพฤติกรรมผู้บริโภคที่แตกต่างกัน ตัวแปรดังกล่าวอาจจะประกอบด้วย อาชีพ อายุ รายได้ เป็นต้น

- ใช้ในการวางแผนเพื่อการทดสอบการตลาด เช่น อาจจะมีการจัดกลุ่มพื้นที่ หรือ จังหวัด โดยรวมพื้นที่ หรือจังหวัดที่คล้ายกันไว้ด้วยกันเพื่อจะได้กำหนดกลยุทธ์ทางการตลาดที่แตกต่างกันสำหรับพื้นที่ที่อยู่ต่างกลุ่มกัน สำหรับตัวแปรที่ควรนำมาพิจารณาในการจัดกลุ่มอาจจะเป็นจำนวนประชากร รายได้เฉลี่ย อาชีพของคนในพื้นที่ พฤติกรรม ทศนคติของคนในพื้นที่ เป็นต้น

- การเปรียบเทียบรถยนต์ยี่ห้อต่างๆ โดยที่ 1 case คือ รถยนต์ 1 ยี่ห้อ ซึ่งพิจารณาจากตัวแปร เช่น ความถี่ในการซ่อม ลูกสูบ ระบบเบรก ค่าใช้จ่ายต่อกิโลเมตร ราคา เป็นต้น ซึ่งแบ่งประเภทของการจัดกลุ่มได้ดังนี้

1) Partitioning Methods

ซึ่งการทำงานจะแบ่งจำนวนข้อมูล (จำนวนเรคอร์ด) n ออกเป็น cluster ต่างๆ จำนวน k cluster โดยให้ k น้อยกว่าหรือเท่ากับ n และในแต่ละกลุ่มต้องมีอย่างน้อย 1 เรคอร์ด และในทางกลับกันในแต่ละเรคอร์ด ต้องมีอย่างน้อย 1 กลุ่ม

การทำ Clustering ด้วยวิธีนี้จะต้องกำหนดค่า k หรือจำนวน cluster ก่อน ว่าต้องการจำนวนกลุ่มเพียงใด จากนั้น algorithm จะสุ่มสร้าง cluster เริ่มต้นขึ้นมา โดยการวนลูปย้ายเรคอร์ดที่มีคุณสมบัติใกล้เคียงกันไปยังกลุ่มต่างๆ จนครบทั้งทุกเรคอร์ด

วิธีนี้เหมาะสมกับฐานข้อมูลขนาดเล็กและฐานข้อมูลขนาดกลางซึ่งไม่เหมาะสมกับฐานข้อมูลที่มีขนาดใหญ่

2) Hierarchical Methods

เป็นการจัดกลุ่มแบบลำดับขั้น โดยแบ่งเป็นสองประเภท ได้แก่

2.1) Agglomerative หรือแบบ Bottom-Up เป็นการรวบรวมข้อมูลจากระดับล่างทีละเรคอร์ด ขึ้นมาจนถึงระดับบน จนสุดท้ายได้เป็นกลุ่มข้อมูลกลุ่มใหญ่

2.2) Divisive หรือแบบ Top-Down ทำงานแบบบนลงล่างซึ่งตรงข้ามกับแบบ Agglomerative

3) Density-based Methods

วิธีนี้จะทำการจัดกลุ่มโดยพิจารณาความหนาแน่นของข้อมูลแทนการใช้ระยะห่างของแต่ละเรคอร์ด

4) Grid-based Methods

วิธีนี้ทำการ Quantize ข้อมูลไปเรื่อยๆจนได้ Cells จำนวนหนึ่งเพื่อใช้ในการสร้างโครงสร้าง Grid ข้อดีคือทำงานได้เร็วและไม่ขึ้นกับจำนวนเรคอร์ดในฐานข้อมูล

5) Model-based Methods

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีนี้จะทำการจัดกลุ่มข้อมูลโดยเทียบกับ โมเดลที่สร้างขึ้นมาซึ่งมี 2 Approach หลักๆ คือ Statistical approach และ Neural Network approach

2. การทำนายค่า (Prediction) และการประเมินค่า (Estimation)

งานการประเมินค่าจะเกี่ยวข้องกับผลลัพธ์แบบต่อเนื่อง (Continuous) เพื่อต้องการจะทำนายหรือประมาณค่าสำหรับตัวแปร ที่เราไม่สามารถทราบค่าที่แน่นอนจากตัวแปรต่างๆ ที่เราทราบจากฐานข้อมูลที่มีนั่นเอง สำหรับการทำนายค่าจะเป็นลักษณะในการทำนายไปยังอนาคต

3. การวิเคราะห์ความสัมพันธ์ (Association Analysis)

การวิเคราะห์ความสัมพันธ์ของแต่ละสิ่งเพื่อหากฎความสัมพันธ์หรือความเกี่ยวข้องต่อกัน หรือมีความน่าจะเป็นที่จะเกิดขึ้นพร้อมกัน

ซึ่งไม่มีเทคนิคใดเลยที่จะสามารถแก้ปัญหของคาด้าไมนิ่งได้ทุกปัญหา ดังนั้นเลือกเทคนิคของคาด้าไมนิ่งให้เหมาะกับงานแต่ละอย่างจะนำไปสู่วิธีการแก้ปัญหของคาด้าไมนิ่งได้ดีที่สุด

2.5 การเตรียมข้อมูลสำหรับทำคาด้าไมนิ่ง

การเตรียมข้อมูลมีความสำคัญมากสำหรับในการทำคาด้าไมนิ่ง เพื่อการทำนาย โดยเริ่มต้นนั้นข้อมูลที่เพื่อใช้ในการจัดกลุ่มข้อมูลเป็นข้อมูลที่ยังไม่สมบูรณ์ที่จะสามารถนำไปใช้ผ่านกระบวนการทำคาด้าไมนิ่งได้ จึงต้องมีการจัดการข้อมูลทำการปรับเปลี่ยนหรือแก้ไขข้อมูลบางส่วนเพื่อให้ข้อมูลมีความเหมาะสมและถูกต้อง ซึ่งในการเตรียมข้อมูลเบื้องต้น มีวิธีการ 3 ขั้นตอนดังนี้

2.5.1 การคัดเลือกข้อมูล (Data Selection)

การคัดเลือกข้อมูลจะคัดเลือกสามารถนำมาใช้ประโยชน์ต่อการวิเคราะห์ข้อมูลได้ หากข้อมูลมีจำนวนไม่มาก อาจทำให้ข้อมูลไม่เพียงพอในการนำมาวิเคราะห์ อาจทำให้มีข้อมูลผิดพลาดได้ในการทดสอบได้

2.5.2 การแก้ไขข้อมูล (Data Cleaning)

การแก้ไขข้อมูลจะเป็นการทำให้ข้อมูลที่ต้องการทำการวิเคราะห์มีความสมบูรณ์และถูกต้อง (Integrity) โดยมีขั้นตอนและวิธีในการทำดังนี้

1) เลือกเฉพาะคอลัมน์ที่สำคัญที่คาดว่าจะสามารถนำมาใช้ประโยชน์ได้ และเป็นคอลัมน์ที่มีข้อมูลค่อนข้างครบถ้วน (กฤษณะ ไวยมัย และคณะ. 2547 ก) ส่วนในบางคอลัมน์ที่มีความสำคัญแต่มีข้อมูลน้อยมากนั้นจะไม่นำมาพิจารณา

2) สำหรับคอลัมน์ที่มีค่าสำหรับทุกแถวเป็นค่าเดียวกัน จะเป็นข้อมูลที่ไม่สามารถแยกความแตกต่างของแต่ละแถวได้เลย ดังนั้นในการทำคาด้าไมนิ่งจะไม่สามารถใช้ประโยชน์จากคอลัมน์นี้ ดังนั้นจึงไม่นำคอลัมน์นี้มาพิจารณา (กฤษณะ ไวยมัย และคณะ. 2547 ข)

3) คอลัมน์ที่มีค่าที่ไม่ซ้ำกันเลย ได้แก่ หมายเลขโทรศัพท์ เป็นต้น ข้อมูลเหล่านี้ไม่สามารถหาแถวที่มีข้อมูลสัมพันธ์กันได้เลย การทำคาด้าไมนิ่งจึงไม่สามารถนำข้อมูลเหล่านี้มาใช้ประโยชน์ได้ ดังนั้นในการทำคาด้าไมนิ่ง ควรกำจัดคอลัมน์ที่มีข้อมูลไม่ซ้ำกันเลยออก (กฤษณะ ไวยมัย และคณะ. 2547 ค)

4) การแก้ไขค่าว่างของข้อมูล (Missing Data) ซึ่งสามารถแก้ไขได้หลายวิธี เช่น แก้ไขโดยกำจัดข้อมูลที่ในแถวเป็นค่าว่าง (NULL) ออกไป (กฤษณะ ไวยมัย และคณะ. 2547 ง) หรือ อีกวิธีคือ ทำการเติมค่าที่ขาดหายไป ในกรณีที่เป็นค่าที่ไม่ใช่ตัวเลข (Categorical) อาจแทนข้อมูลที่ขาดหายไปด้วยคำฐานนิยมหรือค่าที่สร้างขึ้นใหม่ เช่น “unknown” เป็นต้น ส่วนกรณีที่เป็นตัวเลข (Numerical) มักจะแทนค่าด้วยค่าที่ใกล้เคียง เช่น ค่าเฉลี่ย เป็นต้น

5) การแก้ไขข้อมูลที่มีความคลาดเคลื่อน (Noisy data) ความคลาดเคลื่อนนี้อาจเกิดขึ้นได้จากหลายสาเหตุ เช่น เกิดจากการเก็บข้อมูลผิดพลาดของมนุษย์ เช่น การเก็บอายุพนักงาน 30 ปี แต่ใส่ข้อมูลผิดเป็น 300 ปี หรือการบันทึกอัตราเงินเดือนเป็นจำนวนลบ เป็นต้น จะทำให้ข้อมูลนั้นมีความคลาดเคลื่อนออกไป โดยค่านี้จะเรียกว่า “Outliers”

6) การจัดกลุ่มข้อมูลเพื่อลดการกระจาย (Binning Data) เช่นกรณีเกรดของนักศึกษา เนื่องมาจากข้อมูลของนักศึกษามีจำนวนไม่มาก แต่เกรดในแต่ละวิชาที่สามารถมีได้นั้นมีจำนวนมากถึง 10 ตัวด้วยกันคือ {A, B+, B, C+, C, D+, D, F, W, I} ดังนั้นเพื่อลดการกระจายของข้อมูลเกรดของนักศึกษาที่มีมากเมื่อเทียบกับจำนวนนักศึกษา จึงได้จัดกลุ่มเกรดของนักศึกษาเป็น 3 กลุ่ม ดังนี้ คือ เกรด {A, B+, B} เป็น High, เกรด {C+, C} เป็น Medium และ เกรด {D+, D, F, W, I} เป็น Low (กฤษณะ ไวยมัย และคณะ. 2547 จ)

2.5.3 การปรับเปลี่ยนข้อมูล (Data Transformation)

การปรับเปลี่ยนข้อมูลของคาด้าไมนิ่งมีวัตถุประสงค์คือ ทำให้ไม่ningมีประสิทธิภาพมากขึ้นและทำให้สอดคล้องกับโมเดลที่จะนำมาใช้ เนื่องจากข้อมูลที่จะนำมาใช้ทำคาด้าไมนิ่ง ในบางครั้งอยู่ในรูปแบบที่ไม่เหมาะสมกับอัลกอริทึมที่เลือกใช้โดยวิธีการปรับเปลี่ยนข้อมูลจะมีหลายวิธีขึ้นอยู่กับปัญหาของข้อมูล ซึ่งเทคนิคและวิธี เช่น

วิธี Normalization เป็นวิธีไม่ให้ข้อมูลแอตทริบิวต์คนละแอตทริบิวต์ มีช่วงค่าที่ต่างกัน อาจทำให้ผลของการทำ ไมนิ่งคลาดเคลื่อน โดยแปลงข้อมูลให้อยู่ในช่วงหนึ่งๆ เช่น Min-Max Normalization มีสูตรในการคำนวณ ดังแสดงในสมการที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A \quad (1)$$

V' = ค่าข้อมูลใหม่ที่ได้หลังจากการแปลง

V = ข้อมูลเดิมที่จะนำมาทำการแปลง

\min_A = ค่าต่ำสุดของข้อมูลเดิมในแอตทริบิวต์ A

\max_A = ค่าสูงสุดของข้อมูลเดิมในแอตทริบิวต์ A

new_min_A = ค่าต่ำสุดของข้อมูลใหม่ที่ต้องการทำการแปลงข้อมูลแอตทริบิวต์ A

new_max_A = ค่าสูงสุดของข้อมูลใหม่ที่ต้องการทำการแปลงข้อมูลแอตทริบิวต์ A

2.6 เนื้อหาและหลักการของอัลกอริทึม K-Prototypes

เป็นการจัดกลุ่มข้อมูลโดยใช้วิธี Partitioning เนื่องจาก อัลกอริทึม K-Prototypes มักนิยมใช้กับข้อมูลในฐานข้อมูลจริงเพราะในฐานข้อมูลนั้นส่วนใหญ่จะมีทั้งข้อมูลที่เป็นแบบตัวเลข (Numerical) และ ไม่ใช่ตัวเลข (Categorical) ซึ่งหากเป็น K-means จะเหมาะกับข้อมูลที่เป็นตัวเลข เพราะจะมีการแทนค่าเฉลี่ย (means) ในกลุ่มข้อมูล หากเป็น K-modes นั้นจะเหมาะกับ ข้อมูลที่ไม่ใช่ตัวเลข เนื่องจากจะมีการแทนค่าเฉลี่ยด้วยฐานนิยม (modes) และดังที่กล่าวมาแล้ว K-Prototypes จึงเป็นการผสมผสานคุณลักษณะของ K-means และ K-modes เข้าด้วยกัน

2.6.1 การวัดระยะห่างของข้อมูลใน K-Prototypes

การจัดกลุ่มข้อมูลนี้จะแบ่ง X ออกเป็น k กลุ่ม โดยที่ k เป็นจำนวนเต็มบวก วิธีการในการแบ่งกลุ่มใช้วิธีการวัดความเหมือนกันของข้อมูล (Cost Function) ดังแสดงในสมการที่ 2

$$d(X_i, Q_l) = \sum_{j=1}^{m_r} (x_{ij}^r - q_{lj}^r)^2 + \gamma_l \sum_{j=1}^{m_c} \delta(x_{ij}^c, q_{lj}^c) \quad (2)$$

- ให้ข้อมูล $X = \{X_1, X_2, \dots, X_n\}$ แทนเซตข้อมูล n ตัว
- $X_i = \{X_{i1}, X_{i2}, \dots, X_{im}\}$ แทนข้อมูลที่จะนำมาจัดกลุ่มซึ่งเป็นแอตทริบิวต์ของ X
- $Q_l = [q_{l1}, q_{l2}, \dots, q_{lm}]$ เป็นจุดศูนย์กลางของกลุ่มข้อมูล 1
- d เป็นหน่วยที่วัดความคล้ายคลึงกันของข้อมูล
- $\delta(p, q) = 0$ แล้ว $p = q$ หมายถึงกรณีที่ข้อมูล categorical ถ้าเหมือนกันจะกำหนดให้ค่าเท่ากับ 0

- $\delta(p,q) = 1$ แล้ว $p \neq q$ หมายถึงกรณีที่ข้อมูล categorical ถ้าไม่เหมือนกันจะกำหนดให้ค่าเท่ากับ 1
- x_{ij}^n เป็นค่า numeric attributes ของข้อมูล i และ q_{ij}^n เป็นจุดศูนย์กลางของข้อมูล i
- x_{ij}^c เป็นค่า categorical attribute ของข้อมูล i และ q_{ij}^c เป็นจุดศูนย์กลางของกลุ่มข้อมูล i
- m_n เป็นจำนวน attribute ของ numeric
- m_c เป็นจำนวน attribute ของ categorical
- γ_i เป็นค่า weight ของ categorical attribute ซึ่งแล้วแต่ว่าจะให้ความสำคัญกับค่านี้เท่าไร ตามความต้องการของผู้ใช้

2.6.2 อัลกอริทึม K-Prototypes

สำหรับขั้นตอนในการทำงานของ K-Prototypes มีดังนี้

1. เลือก k เริ่มต้น เป็นจำนวนกลุ่มข้อมูลที่ต้องการจัดกลุ่มของข้อมูล X
2. นำข้อมูลแต่ละตัวที่อยู่ใน X มาเปรียบเทียบกับจุดศูนย์กลางของแต่ละกลุ่มข้อมูล หากอยู่ใกล้กับจุดศูนย์กลางของกลุ่มข้อมูลใดมากที่สุด ก็นำข้อมูลนั้นเข้าไปในกลุ่มได้เลย และทำกับสมาชิกทุกๆตัวจนครบ N ตัว ซึ่งหาได้จากสมการที่ 2 และทำการคำนวณค่าจุดศูนย์กลางของแต่ละกลุ่มในกรณีที่มีการย้ายกลุ่มของข้อมูลแต่ละตัว
3. หลังจากข้อมูลทั้งหมดถูกจัดอยู่ในกลุ่มแล้ว ให้เอาข้อมูลแต่ละกลุ่มมาเทียบกับจุดศูนย์กลางของกลุ่มข้อมูลอีกครั้ง ถ้าพบว่าข้อมูลอยู่ใกล้กับจุดศูนย์กลางของกลุ่มข้อมูลในกลุ่มอื่นมากกว่า ให้ทำการย้ายข้อมูลไปยังกลุ่มข้อมูลนั้น และทำการปรับเปลี่ยนจุดศูนย์กลางของกลุ่มข้อมูลใหม่อีกครั้ง

4. ทำซ้ำกับข้อ 3 จนกระทั่งข้อมูลใน X ทุกตัวไม่เปลี่ยนกลุ่ม

ซึ่งจะเห็นได้ว่าอัลกอริทึม K-Prototypes มีการทำงานหลักๆอยู่ 3 กระบวนการ คือ การเลือกกำหนดค่าจำนวนกลุ่มข้อมูลที่ต้องการจะจัดกลุ่ม จัดกลุ่มข้อมูลเริ่มต้นและจัดกลุ่มข้อมูลใหม่ ซึ่งกระบวนการแรกจะใช้วิธีการสุ่มค่า k ขึ้นมา ส่วนกระบวนการที่สองเริ่มกำหนดจุดศูนย์กลางของกลุ่มข้อมูลจากนั้นจะทำการจัดข้อมูลให้อยู่ในกลุ่มข้อมูล และทำการปรับเปลี่ยนค่าจุดศูนย์กลางกลุ่มข้อมูลหลังจากการจัดกลุ่มข้อมูล ส่วนกระบวนการที่สาม คล้ายกับกระบวนการที่สอง แต่แตกต่างกันที่หลังจากการจัดกลุ่มข้อมูลใหม่ จะต้องทำการปรับเปลี่ยนจุดศูนย์กลางกลุ่มข้อมูลของกลุ่มข้อมูลก่อนย้ายและหลังย้าย (Huang, 1997 : 6)

ในรูปที่ 2.2 และรูปที่ 2.3 เป็นกระบวนการบางส่วนที่อยู่ใน K-Prototypes ซึ่งประกอบด้วยตัวแปรและฟังก์ชันดังต่อไปนี้ (Huang, 1997 : 7)

$X[i]$ = แทนข้อมูล X ตัวที่ i

$X[i,j]$ = แทนค่าของแอตทริบิวต์ j สำหรับข้อมูล X ตัวที่ i

$O_prototypes[]$ = เก็บ numeric attribute ของจุดศูนย์กลาง cluster

$C_prototypes[]$ = เก็บ categorical attribute ของจุดศูนย์กลาง cluster

$O_prototypes[i,j]$ = ข้อมูลแบบ numeric ที่เป็นสมาชิกของ cluster i

$C_prototypes[i,j]$ = ข้อมูลแบบ categorical ที่เป็นสมาชิกของ cluster i

$Distance()$ = เป็นฟังก์ชันในการหาค่า square Euclidean distance จากสมการที่ 2

$Sigma()$ = เป็นฟังก์ชันในการหาค่า $\delta()$

$Clustership[]$ = เก็บข้อมูลที่อยู่ใน cluster

$ClusterCount[]$ = เก็บจำนวนข้อมูลที่อยู่ใน cluster

$SumInCluster[]$ = ผลรวมค่า numeric ของข้อมูลใน cluster และใช้ในการปรับเปลี่ยนค่า numeric attribute สำหรับ cluster

$FrequencyInCluster[]$ = เก็บความถี่ของค่าที่เปลี่ยนแปลงของ categorical attribute ใน cluster

$HighestFreq()$ = เป็นฟังก์ชันในการปรับเปลี่ยน categorical attribute ของ cluster

```

FOR i = 1 TO NumberOfObjects
Mindistance= Distance(X[i],O_prototypes[1])+gamma*Sigma(X[i],C_prototypes[1])
FOR j = 1 TO NumberOfClusters
distance= Distance(X[i],O_prototypes[j])+ gamma * Sigma(X[i],C_prototypes[j])
IF (distance < Mindistance)
Mindistance=distance
cluster=j
ENDIF
ENDFOR
Clustership[i]=cluster

```

รูปที่ 2.2 กระบวนการเริ่มต้นจัดกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

ClusterCount[cluster] + 1
FOR j=1 TO NumberOfNumericAttributes
SumInCluster[cluster,j] + X[i,j]
O_prototypes[cluster,j]=SumInCluster[cluster,j]/ClusterCount[cluster]
ENDFOR
FOR j=1 TO NumberOfCategoricAttributes
FrequencyInCluster[cluster,j,X[i,j]] + 1
C_prototypes[cluster,j]=HighestFreq(FrequencyInCluster,cluster,j)
ENDFOR
ENDFOR

```

รูปที่ 2.2 (ต่อ) กระบวนการเริ่มต้นจัดกลุ่มข้อมูล

กระบวนการเริ่มต้นจัดกลุ่มข้อมูลมีดังนี้

1. กำหนดค่าต่ำสุดของ cost function
2. คำนวณค่า cost function ของข้อมูล i ที่อยู่ใน cluster j
3. เปรียบเทียบค่า cost function ที่คำนวณได้ใหม่กับค่า cost function ที่ต่ำสุด โดยที่ถ้าค่า cost function ที่คำนวณได้ใหม่มีค่าน้อยกว่าค่าต่ำสุดของ cost function ให้ค่าต่ำสุดมีค่าเท่ากับค่า cost function ที่คำนวณได้ใหม่ และเก็บค่า cluster ที่ค่าต่ำสุดนั้นอยู่
4. กลับไปทำข้อ 2 ซ้ำ จนกระทั่งทำครบทุก cluster
5. เก็บ cluster ที่ข้อมูลนั้นอยู่ที่มีค่า cost function ต่ำสุด
6. เก็บค่าจำนวนข้อมูลใน cluster เพิ่มขึ้นหนึ่ง
7. วนรูปเพื่อหาผลรวมของค่า numeric และหาค่าต่ำสุดของค่า total cost ของ numeric attribute
8. วนรูปเพื่อหาความถี่ที่เปลี่ยนแปลง และหาค่าต่ำสุดของ categorical attribute
9. กลับไปทำข้อ 1 ซ้ำจนกระทั่งทำครบทุกข้อมูล

```

moves=0
FOR i = 1 TO NumberOfObjects
...
(To find the cluster whose prototype is the nearest to object i. Same as Figure 2.2)
...
IF (Clustership[i] < cluster)
moves+1
oldcluster=Clustership[i]
ClusterCount[cluster] + 1
ClusterCount[oldcluster] - 1
FOR j=1 TO NumberOfNumericAttributes
SumInCluster[cluster,j] + X[i,j]
SumInCluster[oldcluster,j] - X[i,j]
O_prototypes[cluster,j]=SumInCluster[cluster,j]/ClusterCount[cluster]
O_prototypes[oldcluster,j]= SumInCluster[oldcluster,j]/ClusterCount[oldcluster]
ENDFOR
FOR j=1 TO NumberOfCategoricAttributes
FrequencyInCluster[cluster,j,X[i,j]] + 1
FrequencyInCluster[oldcluster,j,X[i,j]] - 1
C_prototypes[cluster,j]=HighestFreq(cluster,j)
C_prototypes[oldcluster,j]=HighestFreq(oldcluster,j)
ENDFOR
ENDIF
ENDFOR

```

รูปที่ 2.3 กระบวนการจัดกลุ่มข้อมูลใหม่

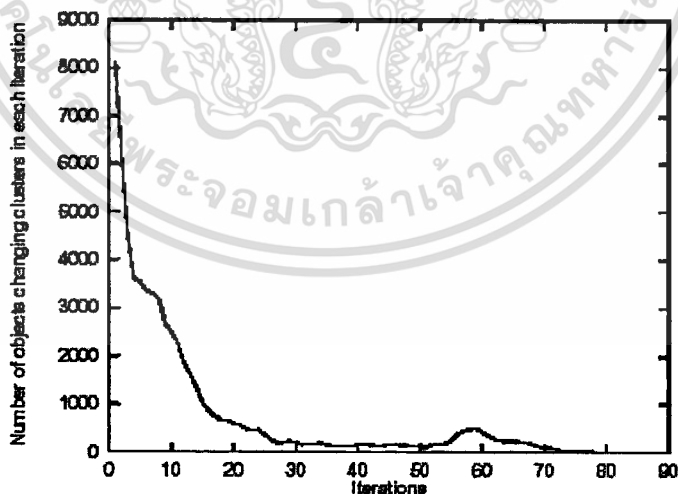
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระบวนการจัดกลุ่มข้อมูลใหม่ มีดังนี้

1. กำหนดค่า moves เริ่มต้นให้เท่ากับศูนย์
2. จัดข้อมูลให้อยู่ใน cluster (ตามกระบวนการดังแสดงในรูปที่ 2.2)
3. เปรียบเทียบ Clustership[i] ว่าเท่ากับ cluster หรือไม่ ถ้าไม่เท่ากัน ให้ทำดังนี้
 - ให้เพิ่มค่า moves ขึ้น 1
 - เพิ่มจำนวนข้อมูลใน cluster ใหม่ขึ้น 1
 - ลดจำนวนข้อมูลใน cluster เดิมลง 1
 - วนรูปเพื่อหาผลรวมของค่า numeric และหาค่าต่ำสุดของ total cost ค่า numeric attribute ทั้งหมดของทั้ง 2 cluster (cluster ใหม่และ cluster เดิม)
 - วนรูปเพื่อหา ความถี่ของค่าที่เปลี่ยนแปลง และหาค่าต่ำสุดของ categorical attribute ของทั้ง 2 cluster (cluster ใหม่และ cluster เดิม)

4. กลับไปทำข้อ 2 ซ้ำจนกระทั่งครบข้อมูลทุกตัว

อัลกอริทึมจะทำซ้ำจนกว่าไม่มีการย้ายกลุ่ม โดยค่า $move = 0$ ซึ่งจะชี้ได้ว่าอัลกอริทึมนั้นเสถียรแล้ว ซึ่งกราฟนี้ได้มาจากการทดสอบกับชุดข้อมูล 75,808 เรคอร์ด ซึ่งมี 20 แอททริบิวต์ จำนวนในการแบ่งกลุ่มข้อมูลแบ่งออกเป็น 64 กลุ่ม ดังแสดงในรูปที่ 2.4



รูปที่ 2.4 กราฟแสดงการทำงานของอัลกอริทึม K-Prototypes

ซึ่ง K-Prototypes จะมีการเปลี่ยนแปลงกลุ่มข้อมูลอย่างรวดเร็วในช่วงเริ่มต้น และจากนั้นความเร็วจะลดลงก่อนที่จะไม่มีเปลี่ยนแปลงข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การคำนวณค่า Cost ของอัลกอริทึมนี้คือ $O((t+1)kn)$ โดยที่ n เป็นข้อมูลที่นำมาใช้ในการแบ่งกลุ่ม ค่า k คือจำนวนกลุ่มที่ต้องการที่จะทำการจัดกลุ่ม และค่า t เป็นจำนวนรอบในกระบวนการจัดกลุ่มใหม่ ซึ่งปกติแล้วค่า k ต้องน้อยกว่าค่า n มาก และค่า t จะไม่เกิน 100 ซึ่งอัลกอริทึมนี้จะมีประสิทธิภาพกับการจัดกลุ่มข้อมูลที่มีขนาดใหญ่ (Huang, 1997 : 8)

2.6.3 ตัวอย่างการทำงานของอัลกอริทึม K-Prototypes กับข้อมูล

ตัวอย่างการทำงานของอัลกอริทึม K-prototypes กับข้อมูลจะแสดงตัวอย่างขั้นตอนการทำงานและการคำนวณของอัลกอริทึม โดยกำหนดให้ข้อมูลนี้ผ่านขั้นตอนการเตรียมข้อมูลมาแล้ว ซึ่งประกอบด้วย 3 แอตทริบิวต์ ดังนี้

- Age (อายุ)
- Gender (เพศ) ประกอบด้วย ชาย และ หญิง
- Color (สีที่ชอบ) ประกอบด้วย Black , Blue , Orange , White

โดยข้อมูลแต่ละแอตทริบิวต์มีชนิดของข้อมูลดังนี้

ตารางที่ 2.1 แสดงชนิดข้อมูลที่จะนำมาทำไบนิ่ง

Attribute	Data Type
Age	Numerical
Gender	Categorical
Color	Categorical

ตารางที่ 2.2 แสดงข้อมูลที่จะนำมาทำไบนิ่ง

Age	Gender	Color
26	Female	Black
27	Female	Blue
20	Male	Blue
23	Male	Orange
15	Female	White
18	Female	Blue
25	Male	Blue
19	Female	Orange

ตารางที่ 2.2 (ต่อ) แสดงข้อมูลที่จะนำมาทำไบนารี

30	Female	White
27	Male	Black

จากอัลกอริทึม สามารถแสดงขั้นตอนการทำงานต่างๆของ K-Prototypes Algorithm ได้ตั้ง
ขั้นตอนต่างๆต่อไปนี้

ขั้นตอนที่ 1 จะทำการกำหนดจำนวนกลุ่มข้อมูล และหาจุดศูนย์กลางของข้อมูล ซึ่งใน
ตัวอย่างจะทำการกำหนด 3 กลุ่ม ดังต่อไปนี้

ตารางที่ 2.3 แสดงข้อมูลในกลุ่มที่ 1

Age	Gender	Color
26	Female	Black
18	Female	Blue
25	Male	Blue

ตารางที่ 2.4 แสดงข้อมูลในกลุ่มที่ 2

Age	Gender	Color
20	Male	Blue
23	Male	Orange
19	Female	Orange

ตารางที่ 2.5 แสดงข้อมูลในกลุ่มที่ 3

Age	Gender	Color
27	Female	Blue
15	Female	White
30	Female	White
27	Male	Black

ซึ่งจะกำหนดจุดศูนย์กลางของแต่ละกลุ่มดังต่อไปนี้

ตารางที่ 2.6 แสดงจุดศูนย์กลางของข้อมูล 3 กลุ่ม

จุดศูนย์กลางของกลุ่มที่ 1	Age = 23	Gender = Female	Color = Blue
จุดศูนย์กลางของกลุ่มที่ 2	Age = 20.66667	Gender = Male	Color = Orange
จุดศูนย์กลางของกลุ่มที่ 3	Age = 24.75	Gender = Female	Color = White

ขั้นตอนที่ 2 จะนำเอาข้อมูลแต่ละตัวมาหาค่า Distance ดังสมการที่ (2) กับจุดศูนย์กลางกลุ่มต่างๆ โดยสมมุติค่า weight ของข้อมูล categorical ดังนี้

- ให้ค่า weight ของ Gender = 0.4
- ให้ค่า weight ของ Color = 0.5

เริ่มพิจารณาทีละกลุ่ม

กลุ่มที่ 1 หาค่า distance ของข้อมูล

ตารางที่ 2.7 แสดงการเปรียบเทียบข้อมูลกับจุดศูนย์กลางกลุ่มที่ 1

ข้อมูล	จุดศูนย์กลางกลุ่มที่ 1
26	23
Female	Female
Black	Blue

$$\begin{aligned} \text{Distance} &= (26-23)^2 + 0.4(0) + 0.5(1) \\ &= 9.5 \end{aligned}$$

กลุ่มที่ 2 หาค่า distance ของข้อมูล

ตารางที่ 2.8 แสดงการเปรียบเทียบข้อมูลกับจุดศูนย์กลางกลุ่มที่ 2

ข้อมูล	จุดศูนย์กลางกลุ่มที่ 2
26	20.66667
Female	Male
Black	Orange

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\begin{aligned}\text{Distance} &= (26-20.66667)^2 + 0.4(1) + 0.5(1) \\ &= 29.34441\end{aligned}$$

กลุ่มที่ 3 หาค่า distance ของข้อมูล

ตารางที่ 2.9 แสดงการเปรียบเทียบข้อมูลกับจุดศูนย์กลางกลุ่มที่ 3

ข้อมูล	จุดศูนย์กลางกลุ่มที่ 3
26	24.75
Female	Female
Black	White

$$\begin{aligned}\text{Distance} &= (26-24.75)^2 + 0.4(0) + 0.5(1) \\ &= 2.0625\end{aligned}$$

เมื่อกำหนดหาค่า distance กับจุดศูนย์กลางของทั้ง 3 กลุ่ม พบว่า ค่า distance ของกลุ่มที่ 3 มีค่าต่ำที่สุด ดังนั้นกำหนดให้ข้อมูล Age = 26 Gender = Female Color = Black อยู่กลุ่มที่ 3

ตารางที่ 2.10 แสดงข้อมูลในกลุ่มที่ 1 หลังจากมีการย้ายกลุ่ม

Age	Gender	Color
18	Female	Blue
25	Male	Blue

ตารางที่ 2.11 แสดงข้อมูลในกลุ่มที่ 3 หลังจากมีการย้ายกลุ่ม

Age	Gender	Color
27	Female	Blue
15	Female	White
30	Female	White
27	Male	Black
26	Female	Black

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 3 จะทำการ update จุดศูนย์กลางของกลุ่มที่ 1 และกลุ่มที่ 3 โดยข้อมูลที่เป็น Numerical นั้นจะคำนวณโดยการหาค่าเฉลี่ย (mean) และส่วนข้อมูลที่เป็น Categorical นั้นจะคำนวณได้จากการหาฐานนิยม (mode) ซึ่งทำการคำนวณดังนี้

Update กลุ่มที่ 1

$$\text{Age} = (18+25)/2 = 21.5$$

$$\text{Gender} = (\text{Female}, \text{Male})$$

$$= \text{Female}$$

$$\text{Color} = (\text{Blue}, \text{Blue})$$

$$= \text{Blue}$$

Update กลุ่มที่ 3

$$\text{Age} = (27+15+30+27+26)/5 = 25$$

$$\text{Gender} = (\text{Female}, \text{Female}, \text{Female}, \text{Male}, \text{Female})$$

$$= \text{Female}$$

$$\text{Color} = (\text{Blue}, \text{White}, \text{White}, \text{Black}, \text{Black})$$

$$= \text{White}$$

ดังนั้นจะได้จุดศูนย์กลางของกลุ่มทั้ง 3 กลุ่ม ดังนี้

ตารางที่ 2.12 สรุปจุดศูนย์กลางของกลุ่มทั้ง 3 กลุ่มหลังถูก update

จุดศูนย์กลางของกลุ่มที่ 1	Age = 21.5	Gender = Female	Color = Blue
จุดศูนย์กลางของกลุ่มที่ 2	Age = 20.66667	Gender = Male	Color = Orange
จุดศูนย์กลางของกลุ่มที่ 3	Age = 25	Gender = Female	Color = White

และจะต้องทำขั้นตอนที่ 2 และ 3 กับทุกข้อมูล ดังนั้นจุดศูนย์กลางจะต้อง update ไปเรื่อยๆ จนกระทั่งข้อมูลทั้งหมดกำหนดไปในกลุ่มต่างๆ หลังจากนั้นให้ทำการทดสอบค่า distance กับจุดศูนย์กลางอีกครั้ง ถ้าพบข้อมูลนั้นใกล้กับกลุ่มอื่น ๆ มากกว่าก็ให้ย้ายข้อมูลไปยังกลุ่มนั้นแล้วทำการ update จุดศูนย์กลางใหม่อีกครั้ง ซึ่งให้ทำงานกระทั่งไม่มีข้อมูลใดๆ เปลี่ยนกลุ่มอีก

บทที่ 3

การออกแบบระบบงาน

3.1 ความต้องการของระบบ (Functional Requirement)

1. ฟังก์ชันการเลือกข้อมูล (Data Selection)

เป็นฟังก์ชันของการเลือกข้อมูลซึ่งเป็นขั้นตอนแรกในเตรียมข้อมูลสำหรับการทำคาค่าไมนิ่ง ซึ่งผู้ใช้สามารถเลือกฐานข้อมูลของ Microsoft Access เลือกตาราง และเลือกฟิลด์ ซึ่งจะเลือกทำได้กับฐานข้อมูลและตารางเดียวเท่านั้น ส่วนฟิลด์สามารถเลือกตามแต่ผู้ใช้ต้องการ

2. ฟังก์ชันการแก้ไขข้อมูล (Data Cleaning)

เป็นฟังก์ชันในการแก้ไขข้อมูลเป็นขั้นตอนที่สองสำหรับการเตรียมข้อมูลสำหรับการทำคาค่าไมนิ่ง ซึ่งจะให้ผู้ใช้ทำการแก้ไขค่าว่างต่างๆ เพื่อที่จะทำให้ผลลัพธ์ของการทำคาค่าไมนิ่งนั้นมีประสิทธิภาพ

3. ฟังก์ชันการปรับเปลี่ยนข้อมูล (Data Transformation)

เป็นฟังก์ชันในการปรับเปลี่ยนข้อมูลซึ่งเป็นขั้นตอนสุดท้ายสำหรับการเตรียมข้อมูลสำหรับการทำคาค่าไมนิ่ง จะเป็นการปรับเปลี่ยนข้อมูลของ Numerical ให้อยู่ในช่วงๆหนึ่ง โดยผู้ใช้จะทำการกำหนดเองว่าต้องการปรับให้ข้อมูล Numerical อยู่ในช่วงใด เพื่อให้ผลลัพธ์ข้อมูลระหว่างแอตทริบิวต์ต่างๆ หลังจากประมวลผลคาค่าไมนิ่งแล้วค่าจะไม่ต่างกันเกินไป

4. ฟังก์ชันการกำหนดจำนวนกลุ่ม

เป็นฟังก์ชันซึ่งเป็นขั้นตอนในการทำคาค่าไมนิ่งของการจัดกลุ่มโดยใช้อัลกอริทึม K-Prototypes โดยผู้ใช้สามารถกำหนดจำนวนกลุ่มเพื่อให้ระบบแสดงผลลัพธ์ โดยจำนวนกลุ่มต้องไม่มากกว่าจำนวนข้อมูล ซึ่งระบบจะทำการตรวจสอบให้โดยอัตโนมัติ

5. ฟังก์ชันการแสดงผลการจัดกลุ่ม

ซึ่งระบบจะแสดงผลการจัดกลุ่มทางหน้าจอ โดยแบ่งเป็นสองส่วนคือ จุดศูนย์กลางของกลุ่มข้อมูล และ สมาชิกต่างๆที่อยู่ในกลุ่มข้อมูลนั้นๆ

6. ฟังก์ชันการแสดงผลการจัดกลุ่มในรูปของกราฟ

จะแสดงผลในรูปของกราฟเฉพาะข้อมูลที่เป็น Numerical ของกลุ่มข้อมูล ทางหน้าจอ

7. ฟังก์ชันการบันทึกข้อมูล

สามารถบันทึกข้อมูลในการจัดกลุ่มได้ทั้ง รูปแบบของ เท็กซ์ไฟล์ (TXT) รูปแบบของ เอกสาร Microsoft Excel (XLS) และ เอกสาร XML

3.2 องค์ประกอบของระบบงาน

ระบบการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes มีการทำงานแบบ Standalone ซึ่งผู้ใช้งานมีความสะดวกและมีความคล่องตัวเนื่องจากสามารถเรียกใช้งานได้ โดยผ่านเครื่องคอมพิวเตอร์ของผู้ใช้ได้เอง โดยไม่ต้องอาศัยการทำงานผ่านเครือข่าย

องค์ประกอบของระบบงานจะประกอบด้วย 3 ส่วน ดังนี้

1. ส่วนนำข้อมูลเข้า

ส่วนข้อมูลนำเข้าจะเป็นขั้นตอนในการนำข้อมูลเข้ามามาเพื่อใช้ในการจัดกลุ่ม ซึ่งอยู่ในขั้นตอนการประมวลผล โดยข้อมูลนำเข้ามีดังนี้

- (1) ไฟล์ฐานข้อมูลซึ่งสำหรับโครงการพัฒนาระบบนี้ได้สนับสนุนไฟล์ฐานข้อมูลเฉพาะ Microsoft Access (MDB) เท่านั้น
- (2) ตารางข้อมูล โดยทำการเลือกจากฐานข้อมูลในข้อที่ 1
- (3) ฟิลด์ข้อมูล โดยทำการเลือกจากตารางข้อมูลในข้อที่ 2
- (4) จำนวนกลุ่มที่ต้องการจัดแบ่ง

2. ส่วนวิเคราะห์และประมวลผล

ในส่วนวิเคราะห์และประมวลผลจะเป็นการประมวลผลจากข้อมูลที่นำเข้ามาจากส่วนข้อมูลนำเข้า โดยการประมวลผลจะแบ่งเป็นสองส่วนคือ การเตรียมข้อมูล และการทำเหมืองหน้าที่ในการนำเข้าสู่ฐานข้อมูล

- ส่วนการเตรียมข้อมูล เป็นขั้นตอนในการเตรียมข้อมูลเพื่อจะนำไปใช้ต่อในกระบวนการเหมือง โดยมีขั้นตอนดังนี้
 - (1) เลือกฟิวด์ข้อมูลที่ต้องการนำมาจัดกลุ่ม
 - (2) ทำการแก้ไขข้อมูล (Data Cleaning) โดยจะตรวจสอบค่าที่หายไป (Missing Value) ถ้าตรวจพบก็จะมีการแก้ไข โดยผู้ใช้
 - (3) ทำการปรับเปลี่ยนข้อมูล (Data Transformation) ซึ่งในกระบวนการนี้จะทำหรือไม่ขึ้นอยู่กับผู้ใช้งานที่ต้องการ
- ส่วนการทำเหมือง ซึ่งจะเป็นการจัดกลุ่มโดยใช้ K-Prototypes Algorithm ในการวัดความเหมือนกันของข้อมูล โดยวัดระยะห่างจากจุดศูนย์กลาง

3. ส่วนแสดงรายงาน

ทำหน้าที่ในการสรุปผลลัพธ์การจัดกลุ่มซึ่งจะแสดงทั้งจุดศูนย์กลางของกลุ่มข้อมูลต่างๆ และแสดงสมาชิกที่อยู่ในกลุ่มข้อมูลนั้นๆ ทางหน้าจอของโปรแกรม นอกจากนี้ยังบันทึกเป็นรายงานในการสรุปด้วยเอกสารในรูปแบบต่างๆ ได้แก่

(1) เอกสารที่เป็นเท็กซ์ไฟล์ (TXT) ซึ่งเมื่อบันทึกไฟล์ออกมาจะมีข้อมูลในการจัดกลุ่มข้อมูลซึ่งประกอบไปด้วย จุดศูนย์กลางของกลุ่มต่างๆ และส่วนของข้อมูลเรคอร์ดต่างๆ ที่ถูกจัดเข้าเป็นสมาชิกของกลุ่ม

(2) เอกสารของ Microsoft Excel (XLS) ซึ่งเมื่อบันทึกไฟล์ออกมาจะมีข้อมูลในการจัดกลุ่มข้อมูลซึ่งประกอบไปด้วย จุดศูนย์กลางของกลุ่มต่างๆ อยู่ใน Sheet ที่ชื่อ Member และ ส่วนของข้อมูลเรคอร์ดต่างๆ ที่ถูกจัดเข้าเป็นสมาชิกของกลุ่ม จะอยู่ใน Sheet ที่ชื่อ Center

(3) เอกสาร XML ซึ่งเมื่อบันทึกไฟล์ออกมาจะมีข้อมูลในการจัดกลุ่มข้อมูลซึ่งประกอบไปด้วย Schema ของข้อมูลจุดศูนย์กลางของกลุ่มต่างๆ และแสดงข้อมูลจุดศูนย์กลางกลุ่มต่างๆ

(4) แสดงผลลัพธ์ในรูปของกราฟ เมื่อคุณผลลัพธ์ในรูปแบบของกราฟจะแสดงกราฟที่ข้อมูลจุดศูนย์กลางของกลุ่มต่างๆ และสามารถสั่งพิมพ์กราฟนี้ออกมาเป็นเอกสารได้

3.3 ขั้นตอนการทำงานของระบบ

ในการทำงานของระบบมีขั้นตอนดังต่อไปนี้

1. ผู้ใช้เลือกไฟล์ฐานข้อมูล จากนั้นเลือกตาราง และฟิลด์ที่ต้องการนำมาจัดกลุ่มข้อมูล
2. ผู้ใช้ทำการแก้ไขข้อมูลที่มีค่าว่างที่ขาดหายไป (Data Cleaning) กับทุกๆ ฟิลด์ โดยข้อมูลมีกรณีต่างๆ ดังนี้
 - 2.1 หากค่าว่างนั้นเป็น Numerical ให้ใส่ค่าเฉลี่ย หรือลบเรคอร์ดที่มีค่าว่าง
 - 2.2 หากค่าว่างนั้นเป็น Categorical ให้ใส่ค่าฐานนิยม หรือค่าใดๆตามที่กรอกในช่องรับข้อมูล หรือลบเรคอร์ดที่มีค่าว่าง
 - 2.3 หากไม่พบค่าว่าง ผู้ใช้ไม่ต้องแก้ไขข้อมูล
3. ผู้ใช้ทำการปรับเปลี่ยนข้อมูล (Data Transformation) กับทุกๆ ฟิลด์ โดยมีกรณีต่างๆ ดังนี้
 - 3.1 สำหรับข้อมูลที่เป็น Numerical หากผู้ใช้ต้องการปรับเปลี่ยนข้อมูล ให้เลือกกรณีที่ต้องการปรับเปลี่ยนข้อมูล

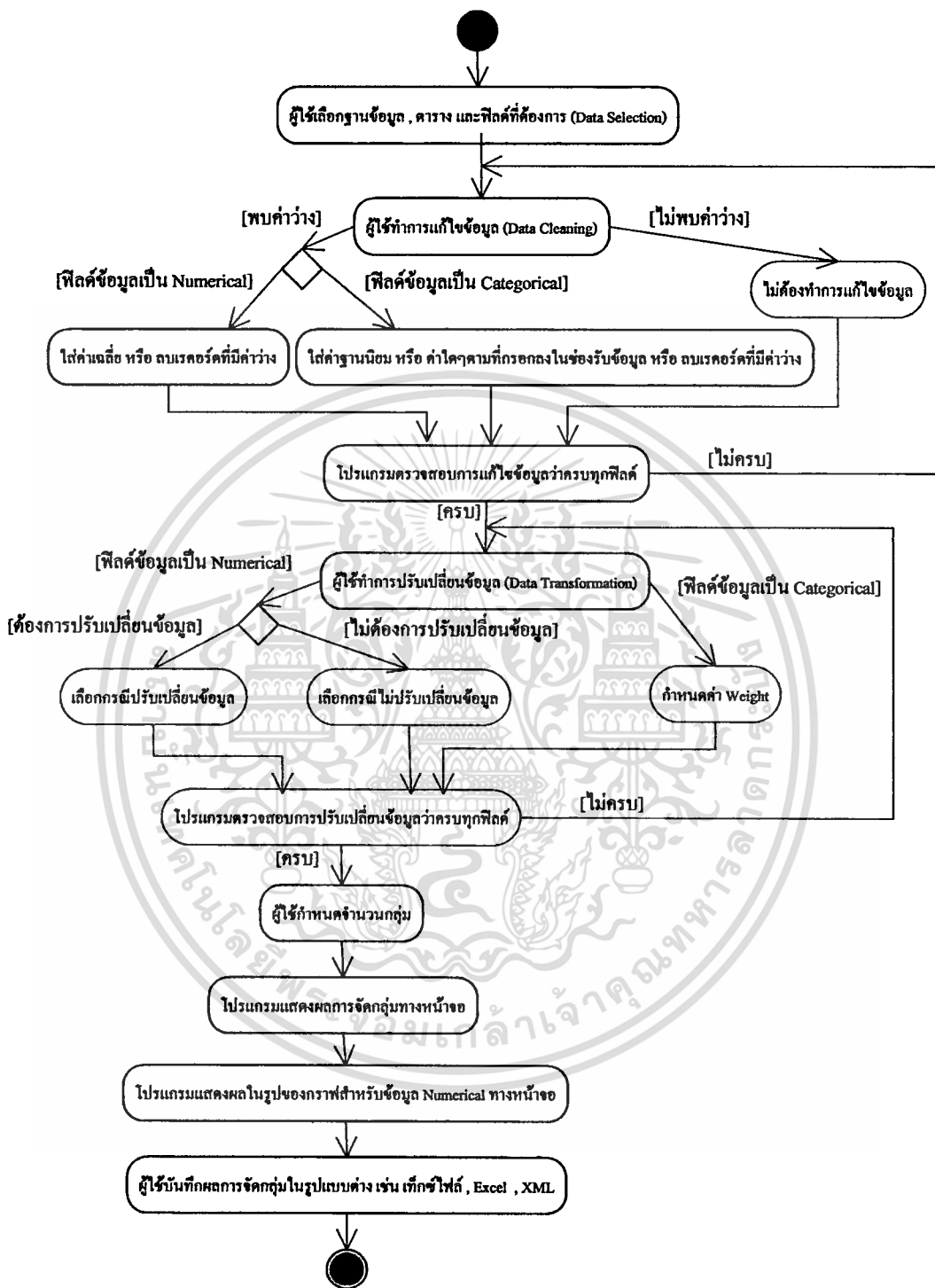
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 สำหรับข้อมูลที่เป็น Numerical หากผู้ใช้ไม่ต้องการปรับเปลี่ยนข้อมูล ให้เลือกกรณีไม่ต้องการปรับเปลี่ยนข้อมูล

3.3 สำหรับข้อมูลที่เป็น Categorical ให้ผู้ใช้กำหนดค่า Weight ให้กับฟิลด์ต่างๆ เพื่อจะนำไปใช้ในกระบวนการของโมเดล

4. ผู้ใช้กำหนดจำนวนกลุ่มที่ต้องการนำมาจัดกลุ่มข้อมูล
5. โปรแกรมจะทำการประมวลผล และทำการจัดกลุ่มแสดงทางหน้าจอ
6. โปรแกรมแสดงผลในรูปของกราฟสำหรับข้อมูลที่เป็น Numerical ทางหน้าจอ
7. ผู้ใช้สามารถเลือกการบันทึกผลการจัดกลุ่มในรูปแบบต่างๆ เช่น เท็กซ์ไฟล์ (TXT) เอกสารของ Microsoft Excel (XLS) และ เอกสาร XML โดยแสดงด้วยผังงานกิจกรรมการทำงานหลักของระบบดังแสดงในรูปที่ 3.1



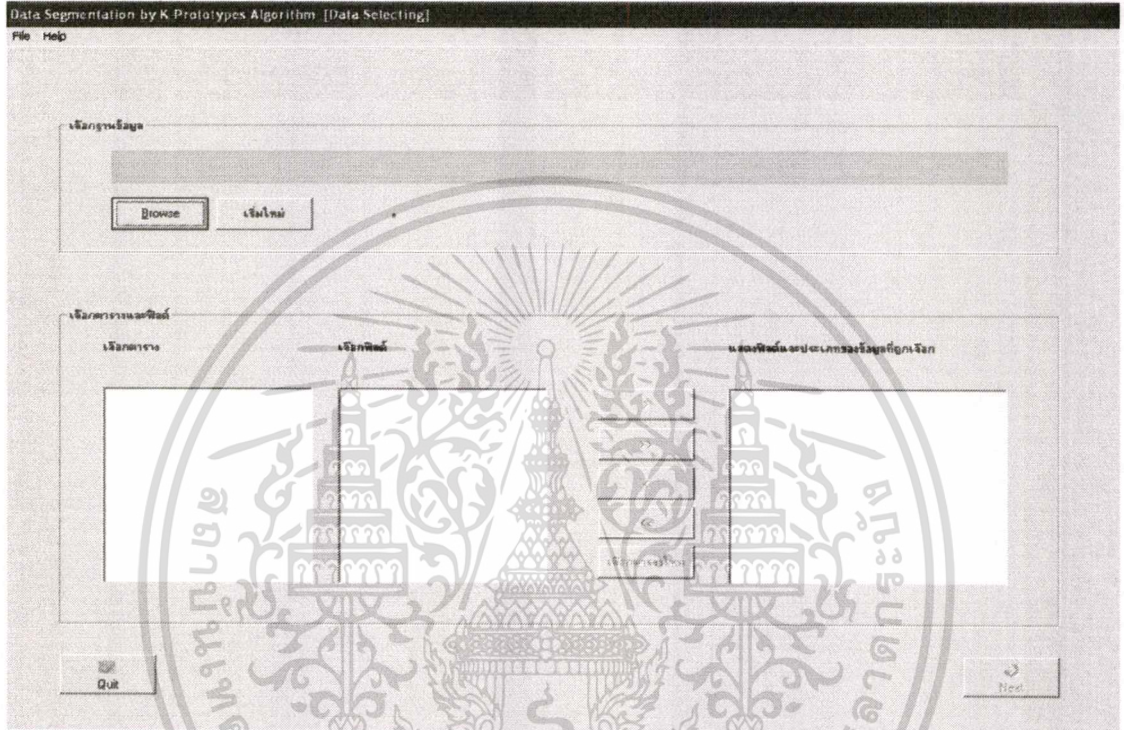


รูปที่ 3.1 ฟังงานกิจกรรมการทำงานหลักของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การออกแบบหน้าจอการทำงานของระบบ

เมื่อเข้าสู่โปรแกรมจะปรากฏหน้าจอสำหรับเลือกข้อมูล (Data Selection) ดังรูปที่ 3.2



รูปที่ 3.2 หน้าจอสำหรับเลือกข้อมูล

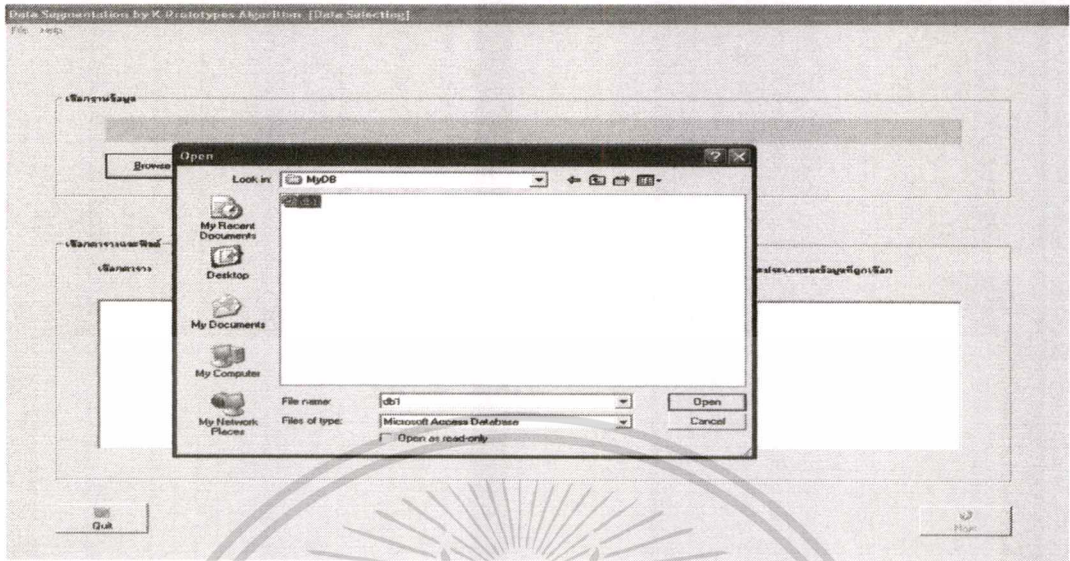
ซึ่งหน้าจอนี้คือหน้าจอสำหรับการเลือกข้อมูล ประกอบด้วย 4 ส่วนคือ

- เลือกฐานข้อมูล - เป็นส่วนที่ให้ผู้ใช้งานทำการเลือกฐานข้อมูลขึ้นมา
- เลือกตาราง - เป็นส่วนที่ให้ผู้ใช้งานทำการเลือกตารางจาก listbox
- เลือกฟิลด์ - เป็นส่วนที่ให้ผู้ใช้งานทำการเลือกฟิลด์จาก listbox
- แสดงฟิลด์และประเภทของข้อมูลที่ถูกลเลือก - เป็นส่วนแสดงฟิลด์และประเภทของข้อมูลที่ถูกเลือกโดยผู้ใช้งาน

3.4.1 การติดต่อกับฐานข้อมูลที่ต้องการวิเคราะห์

กดปุ่ม Browse ในส่วนเลือกฐานข้อมูล เพื่อเลือกไฟล์ฐานข้อมูลนามสกุล .mdb ที่ต้องการนำมาให้ระบบวิเคราะห์ในการจัดกลุ่มข้อมูลต่อไป ดังแสดงในรูปที่ 3.3

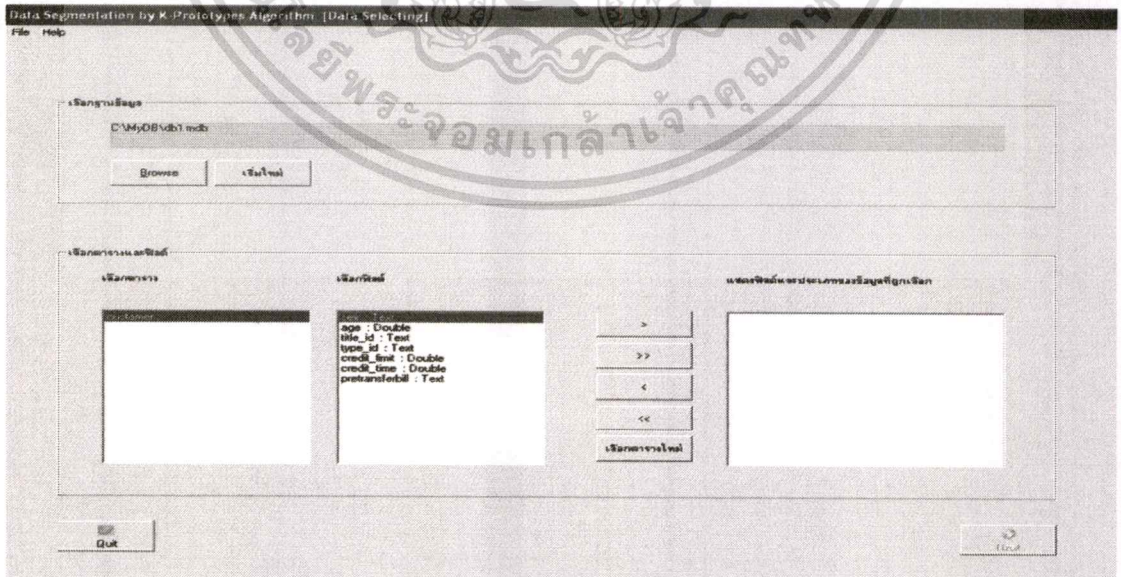
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 เลือกฐานข้อมูลที่ต้องการนำมาวิเคราะห์

3.4.2 การเลือกตารางในฐานข้อมูลที่ต้องการวิเคราะห์

หลังจากเมื่อเลือกฐานข้อมูลเรียบร้อยแล้ว ระบบจะแสดงตารางทั้งหมดของฐานข้อมูลนั้นๆที่ถูกเลือกมาให้ผู้ใช้ทำการเลือกตารางที่ต้องการนำมาวิเคราะห์ ดังแสดงในรูปที่ 3.4

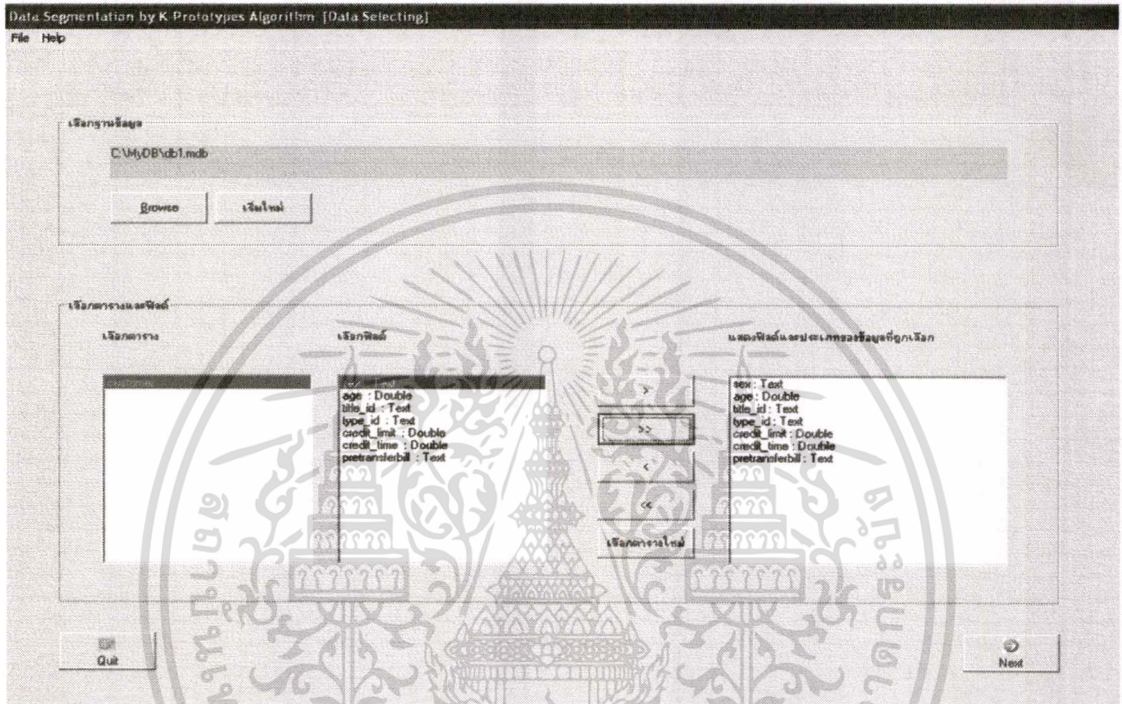


รูปที่ 3.4 เลือกตารางในฐานข้อมูลที่ต้องการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3 การเลือกฟิลด์ในฐานข้อมูลที่ต้องการวิเคราะห์

เมื่อทำการเลือกฟิลด์ที่ต้องการแล้ว ระบบจะแสดงฟิลด์ที่ถูกเลือก รวมถึงประเภทข้อมูล (Data type) ให้ผู้ใช้ทราบ ดังแสดงในรูปที่ 3.5



รูปที่ 3.5 เลือกฟิลด์ในฐานข้อมูลที่ต้องการวิเคราะห์

หลังจากนั้นสามารถคลิกปุ่ม Next เพื่อทำงานในขั้นตอนต่อไปได้

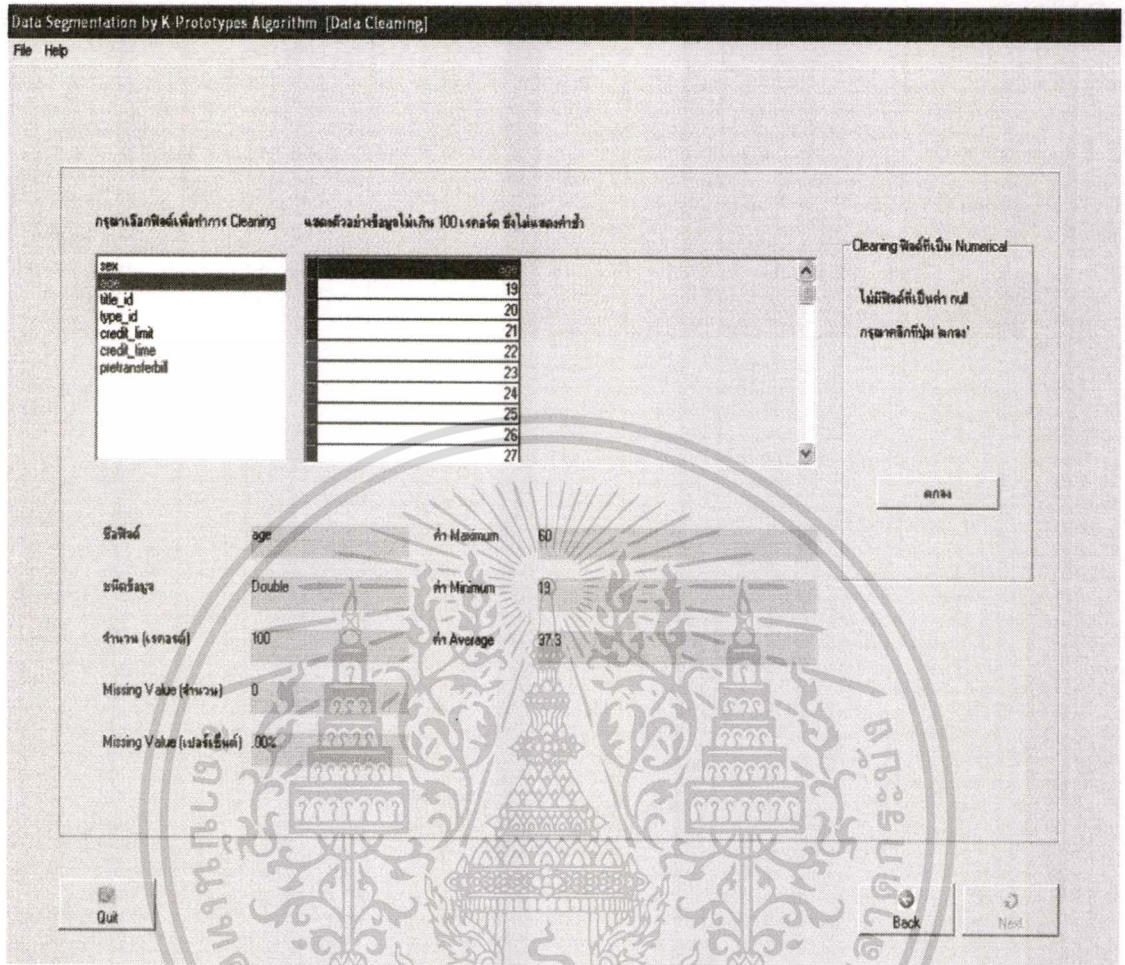
3.4.4 การแก้ไขข้อมูล (Data Cleaning)

หลังจากที่ทำการเลือกข้อมูล ทั้งฐานข้อมูล ตาราง และฟิลด์เรียบร้อยแล้ว ขั้นตอนต่อไป จะเป็นการแก้ไขข้อมูล เพื่อจัดการกับเรคคอร์ดต่างๆ ที่มีค่า null และระบบจะแสดงรายละเอียดของฟิลด์นั้นๆ ให้ผู้ใช้ทราบ

ซึ่งการทำแก้ไขข้อมูลในระบบมี 4 กรณี ดังนี้

กรณีที่ 1 ฟิลด์ที่เป็น Numerical ไม่มีค่า null

สำหรับฟิลด์ Numerical ที่ไม่มี record ใดๆ เป็นค่า null ระบบจะแสดงดังรูปที่ 3.6



รูปที่ 3.6 การแก้ไขข้อมูลสำหรับพิลด์ที่เป็น Numerical ไม่มีค่า null

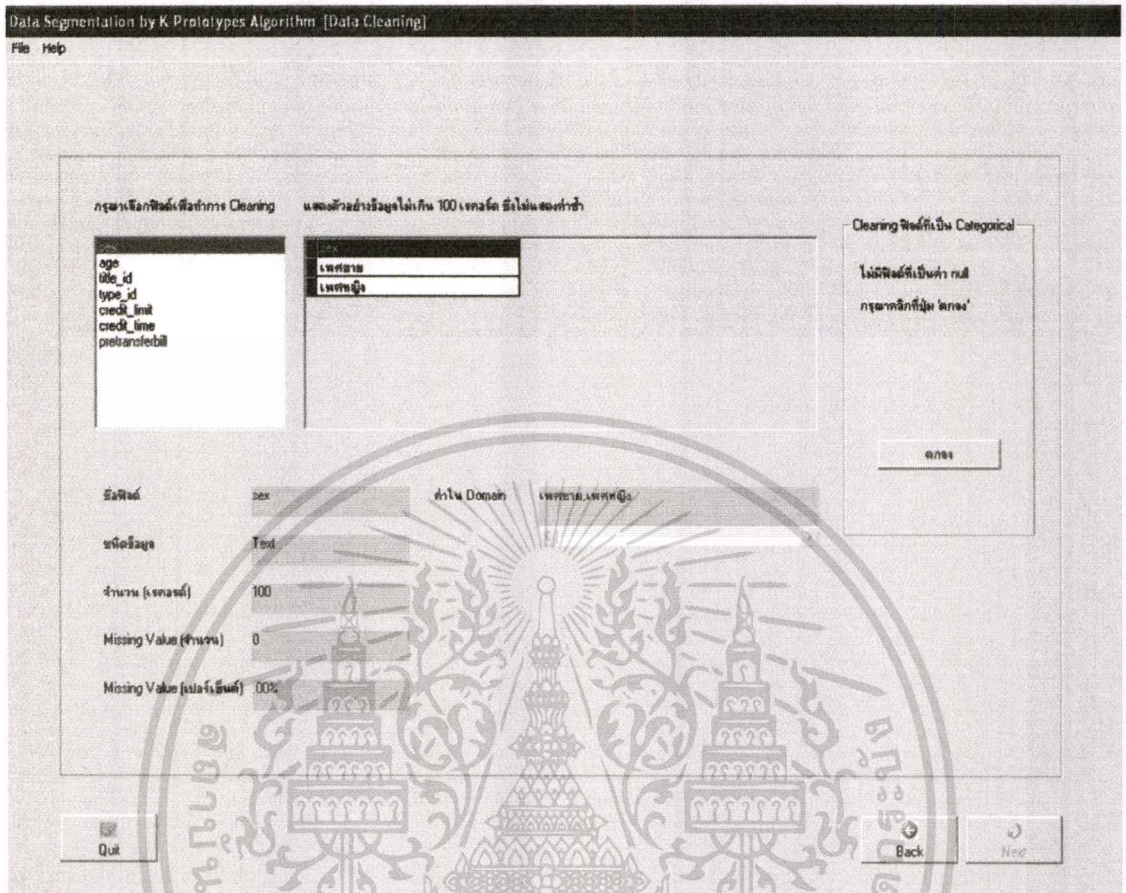
จะมีข้อมูลแสดงรายละเอียดต่างๆ ในหน้าจอดังนี้

- ส่วนแสดงตัวอย่างข้อมูลในตาราง
- ชื่อพิลด์ - เป็นส่วนที่แสดงชื่อพิลด์
- ชนิดข้อมูล - เป็นส่วนที่แสดงชนิดข้อมูล
- จำนวนเรคอร์ด - เป็นส่วนที่แสดงจำนวนเรคอร์ดขึ้นมาทั้งหมด
- จำนวนค่าว่าง , จำนวนเปอร์เซ็นต์ค่าว่าง
- ค่า Minimum , ค่า Maximum , ค่า Average

กรณีที่ 2 พิลด์ที่เป็น Categorical ไม่มีค่า null

สำหรับพิลด์ Categorical ที่ไม่มี record ใดๆเป็นค่า null ระบบจะแสดงดังรูปที่ 3.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 การแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ไม่มีค่า null

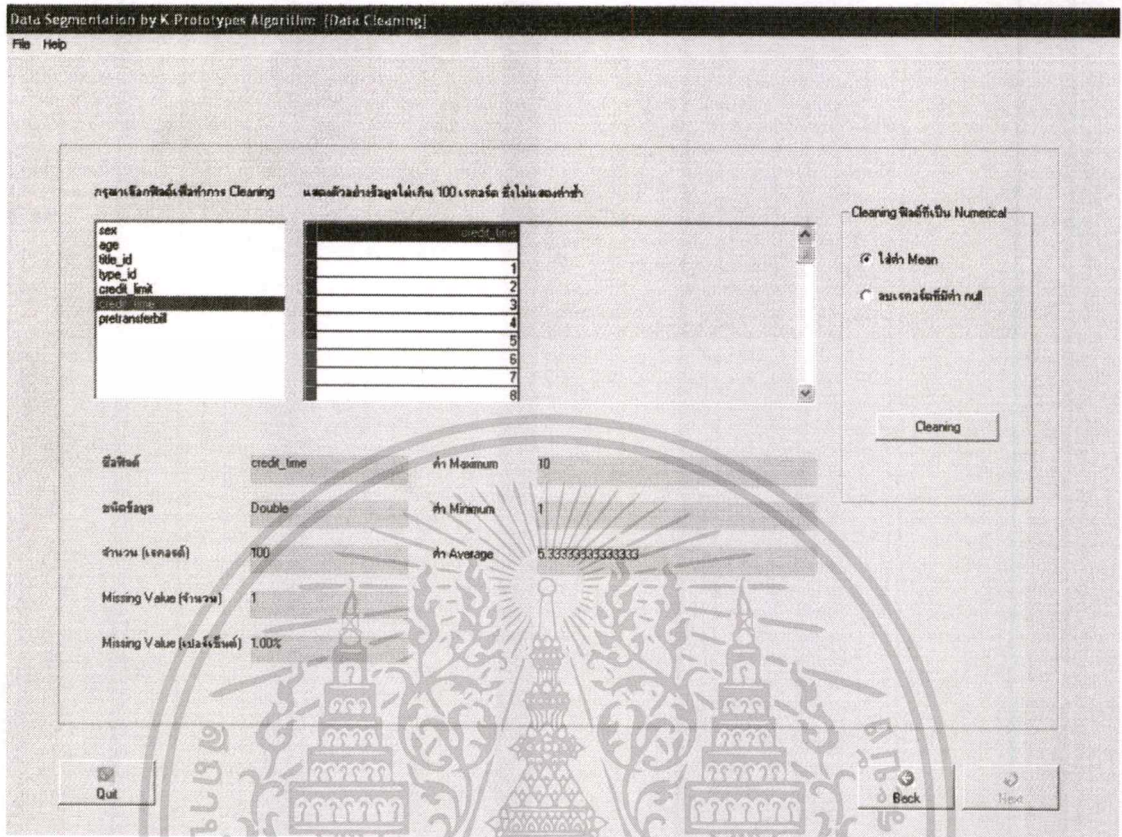
จะมีข้อมูลแสดงรายละเอียดต่างๆในหน้าจอดังนี้

- ส่วนแสดงตัวอย่างข้อมูลในตาราง
- ชื่อฟิลด์ - เป็นส่วนที่แสดงชื่อฟิลด์
- ชนิดข้อมูล - เป็นส่วนที่แสดงชนิดข้อมูล
- จำนวนเรคอร์ด - เป็นส่วนที่แสดงจำนวนเรคอร์ดขึ้นมาทั้งหมด
- จำนวนค่าว่าง , จำนวนเปอร์เซ็นต์ค่าว่าง
- ค่าใน domain - เป็นส่วนที่แสดงค่าในโดเมนของฟิลด์นี้ทั้งหมด

กรณีที่ 3 ฟิลด์ที่เป็น Numerical ที่มีค่า null

สำหรับฟิลด์ Numerical ที่มี record ใดๆเป็นค่า null ระบบจะแสดงดังรูปที่ 3.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.8 การแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Numerical ที่มีค่า null

จะมีข้อมูลแสดงรายละเอียดต่างๆในหน้าจอดังนี้

- ส่วนแสดงตัวอย่างข้อมูลในตาราง
- ชื่อฟิลด์ - เป็นส่วนที่แสดงชื่อฟิลด์
- ชนิดข้อมูล - เป็นส่วนที่แสดงชนิดข้อมูล
- จำนวนเรคอร์ด - เป็นส่วนที่แสดงจำนวนเรคอร์ดขึ้นมาทั้งหมด
- จำนวนค่าว่าง , จำนวนเปอร์เซ็นต์ค่าว่าง
- ค่า Minimum , ค่า Maximum , ค่า Average

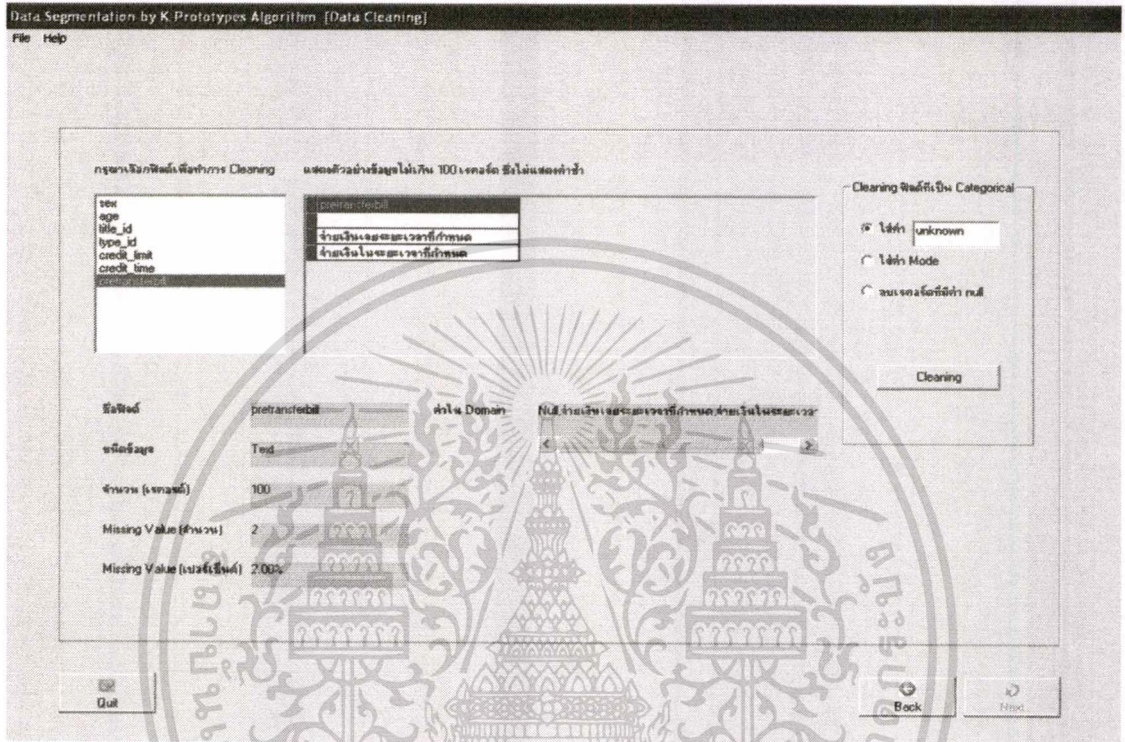
และมีตัวเลือกในการแก้ไขข้อมูลดังนี้

- ใส่ค่า Mean (ค่าเฉลี่ย) โดยจะใส่แทนข้อมูลที่เป็นค่า null
- ลบเรคอร์ดที่มีค่า null

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีที่ 4 ฟิลด์ที่เป็น Categorical ที่มีค่า null

สำหรับฟิลด์ Categorical ที่มี record ใดๆ เป็นค่า null ระบบจะแสดงดังรูปที่ 3.9



รูปที่ 3.9 การแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ที่มีค่า null

จะมีข้อมูลแสดงรายละเอียดต่างๆ ในหน้าจอ ดังนี้

- ส่วนแสดงตัวอย่างข้อมูลในตาราง
- ชื่อฟิลด์ - เป็นส่วนที่แสดงชื่อฟิลด์
- ชนิดข้อมูล - เป็นส่วนที่แสดงชนิดข้อมูล
- จำนวนเรคอร์ด - เป็นส่วนที่แสดงจำนวนเรคอร์ดขึ้นมาทั้งหมด
- จำนวนค่าว่าง , จำนวนเปอร์เซ็นต์ค่าว่าง
- ค่าใน domain - เป็นส่วนที่แสดงค่าใน โดเมนของฟิลด์นี้ทั้งหมด

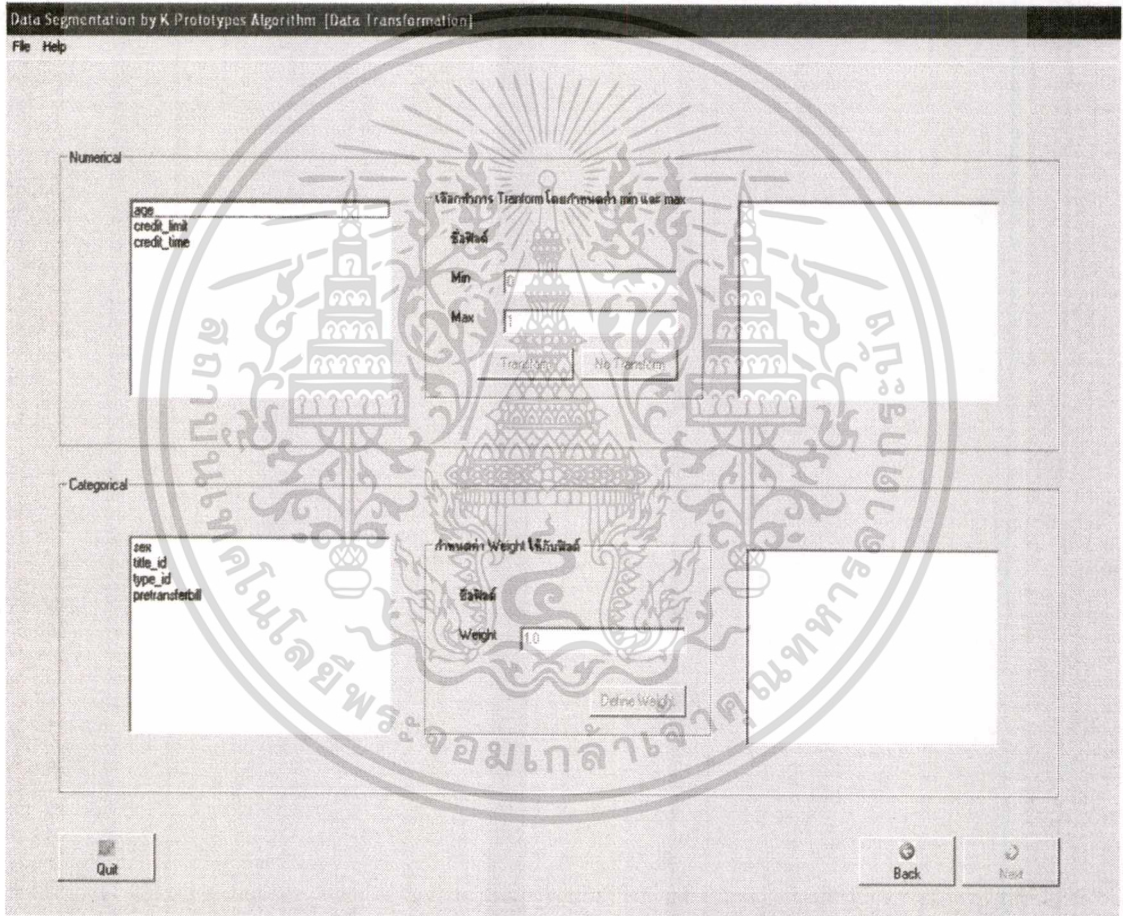
และมีตัวเลือกในการแก้ไขข้อมูลดังนี้

- ใส่ค่าโดยที่ผู้ใช้กำหนดเอง
- ใส่ค่า Mode (ฐานนิยม) โดยจะใส่แทนข้อมูลที่เป็นค่า null
- ลบเรคอร์ดที่มีค่า null

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.5 การปรับเปลี่ยนข้อมูล (Data Transformation)

ขั้นตอนการปรับเปลี่ยนข้อมูล ผู้ใช้สามารถทำการปรับเปลี่ยนข้อมูลของค่าที่เป็น Numerical ให้อยู่ในช่วงๆหนึ่งได้ โดยระบุค่า Min และ Max โดยโปรแกรมจะกำหนดให้ค่า Min น้อยกว่าค่า Max เสมอ หรือในกรณีหากไม่ต้องการปรับเปลี่ยนในกรณีที่ผู้ใช้ต้องการคงค่าเดิมไว้ก็สามารถเลือกได้ สำหรับค่าที่เป็น Categorical ผู้ใช้จำเป็นต้องกำหนดค่า Weight เพื่อนำไปใช้ในขั้นตอนต่อไป ดังแสดงในรูปที่ 3.10



รูปที่ 3.10 การปรับเปลี่ยนข้อมูล

3.4.6 การกำหนดจำนวนกลุ่มที่ต้องการจัดกลุ่ม

ขั้นตอนในการกำหนดจำนวนกลุ่มให้ข้อมูลที่ต้องการจัดกลุ่ม ให้ผู้ใช้ใส่จำนวนที่ต้องการจัดกลุ่มลงในช่อง textbox จากนั้นให้คลิกที่ปุ่มตกลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Segmentation by K-Prototypes Algorithm [Data Mining]

File Help

sex	age	type_id	type_id	credit_line	credit_line	probability
เพศหญิง	0.17073171	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	1	0.77777779	ชำระเงินโดยระยะเวลา
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555558	0.65656563	ชำระเงินโดยระยะเวลา
เพศหญิง	0.14634146	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555558	0.65656563	ชำระเงินโดยระยะเวลา
เพศหญิง	4.6790495E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.44444445		ชำระเงินโดยระยะเวลา
เพศหญิง	0.17073171	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.65656568	0.11111111	ชำระเงินโดยระยะเวลา
เพศหญิง	0	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.77777779		ชำระเงินโดยระยะเวลา
เพศหญิง	0.2439044	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.22222222	0.11111111	ชำระเงินโดยระยะเวลา
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.33333334	0.33333334	ชำระเงินโดยระยะเวลา
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.88888889		ชำระเงินโดยระยะเวลา
เพศหญิง	9.7560972E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.44444445	0.33333334	ชำระเงินโดยระยะเวลา
เพศหญิง	0.19512194	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.22222222	0.22222222	ชำระเงินโดยระยะเวลา
เพศหญิง	0.12195122	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.44444445	0.11111111	ชำระเงินโดยระยะเวลา
เพศหญิง	0.14634146	พนักงานทั่วไป	ชำระเงินเป็นเงินสด		1	ชำระเงินโดยระยะเวลา
เพศหญิง	0.17073171	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.77777779	0.44444445	ชำระเงินโดยระยะเวลา
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.65656568		ชำระเงินโดยระยะเวลา
เพศหญิง	0.19512194	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.22222222	0.11111111	ชำระเงินโดยระยะเวลา
เพศหญิง	0.12195122	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.11111111	0.88888889	ชำระเงินโดยระยะเวลา
เพศหญิง	0.29268292	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.77777779	0.11111111	ชำระเงินโดยระยะเวลา
เพศหญิง	7.3170729E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555558	0.88888889	ชำระเงินโดยระยะเวลา
เพศหญิง	7.3170729E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555558	0.65656563	ชำระเงินโดยระยะเวลา
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.33333334	0.77777779	ชำระเงินโดยระยะเวลา
เพศชาย	0.2682927	ข้าราชการทั่วไป	ชำระเงินผ่านบัตรเครดิต	0.11111111		ชำระเงินโดยระยะเวลา
เพศชาย	0.19512194	ข้าราชการทั่วไป	ชำระเงินผ่านบัตรเครดิต	1	0.88888889	ชำระเงินโดยระยะเวลา
เพศชาย	0.14634146	ข้าราชการทั่วไป	ชำระเงินผ่านบัตรเครดิต	0.88888889	0.65656563	ชำระเงินโดยระยะเวลา
เพศชาย	0.14634146	ข้าราชการทั่วไป	ชำระเงินผ่านบัตรเครดิต	0.77777779	0.44444445	ชำระเงินโดยระยะเวลา

กำหนดจำนวนกลุ่มที่ใช้ในการจัดกลุ่ม

Quit Back 7100

รูปที่ 3.11 กำหนดจำนวนกลุ่มที่ใช้ในการจัดกลุ่ม

3.4.7 การแสดงผลการจัดกลุ่ม

หลังจากที่ระบบได้ทำการวิเคราะห์และจัดกลุ่มข้อมูลให้เรียบร้อยแล้ว ก็จะมีการแสดงผล (Output) ซึ่งประกอบไปด้วย 2 ส่วนคือ

- Cluster Center จะแสดงจุดศูนย์กลางของแต่ละกลุ่ม และบอกจำนวนสมาชิกที่อยู่ในกลุ่มนั้นๆ
- Cluster Membership จะแสดงข้อมูลแต่ละเรคอร์ดที่อยู่ในกลุ่มใดๆ และบอกถึงระยะห่างระหว่างสมาชิกกับจุดศูนย์กลางของกลุ่มนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Segmentation by K Prototypes Algorithm [Output]

File Help

จำนวนข้อมูลทั้งหมด 100 เกรด ค่าจำนวนกลุ่ม 3 กลุ่ม เวลาในการประมวลผล : 10.963750 วินาที

Cluster Center

age	credit limit	credit time	sex	file id	type id	pretransferbill
0.42461197247959401	0.72727273878726095	0.31818182630972403	เพศชาย	งอมผู้จัดการ คู่สาย พิเศษ	ชำระเงินผ่านบัตรเครดิต	ชำระเงินในระยะเวลาที่กำหนด
0.53139888420400201	0.44179895005765402	0.60317461476439505	เพศชาย	ผู้บริหาร ใช้งานการสมัคร	ชำระเงินผ่านบัตรเครดิต	ชำระเงินในระยะเวลาที่กำหนด
0.36043360539608499	0.53256260280264703	0.438271610066295	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด

Cluster Membership

age	credit limit	credit time	sex	file id	type id	pretransferbill
0.17073170840740204	1	0.77777779102325439	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.21951219439506531	0.55955598204650879	0.66666666853488153	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.14634145796296981	0.55955598204650879	0.8888888958116272	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
8780495987663269E-2	0.4444444477598136	1	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.17073170840740204	0.66666666853488159	0.111111119389534	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0	0.77777779102325439	1	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.24390244483947754	0.222222238779068	0.111111119389534	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.12195122241973877	0.3333333432674408	0.3333333432674408	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.21951219439506531	0.8888888958116272	0	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
7960971975326538E-2	0.4444444477598136	0.3333333432674408	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.19512194395065308	0.222222238779068	0.222222238779068	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.12195122241973877	0.4444444477598136	0.111111119389534	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.14634145796296981	1	1	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.17073170840740204	0.77777779102325439	0.4444444477598136	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.21951219439506531	0.66666666853488159	1	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด
0.19512194395065308	0.222222238779068	0.111111119389534	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินในระยะเวลาที่กำหนด

Back Show Graph Export Text Export Excel Export XML Quit

รูปที่ 3.12 การแสดงผลการจัดกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การประยุกต์ใช้โปรแกรม

4.1 การกำหนดวัตถุประสงค์

สำหรับบริษัทต่าง ๆ นั้นมักจะมีการให้สินเชื่อกับลูกค้า และหลังจากให้สินเชื่อไปแล้ว บางครั้งผู้บริหารอาจมีนโยบายการปรับเปลี่ยนรายละเอียดการให้สินเชื่อ เช่น เพิ่มช่วงเวลาการชำระเงินของลูกค้า หรือเพิ่มวงเงินในการให้สินเชื่อ ซึ่งปัญหาที่มักจะเกิดขึ้นในการปรับเปลี่ยนสินเชื่อลูกค้าคือ ข้อมูลในการนำมาพิจารณานั้นมีจำนวนมาก ทำให้ยากต่อการวิเคราะห์ข้อมูลทั้งหมดให้เสร็จภายในระยะเวลาที่รวดเร็ว ดังนั้นจึงมีความคิดที่จะนำเอาค่าไมนิ่งมาประยุกต์ใช้กับการแก้ปัญหาในครั้งนี้ โดยวัตถุประสงค์เพื่อจัดกลุ่มลูกค้าออกเป็นประเภทต่างๆ ซึ่งการจัดกลุ่มข้อมูลออกเป็น 3 กลุ่มลูกค้า จำนวน 100 เรคอร์ด แบ่งโดยแยกเป็นอายุ อาชีพ และการชำระเงินตรง กำหนดหรือไม่ จากนั้นนำข้อมูลที่ได้ออกมาวิเคราะห์ดูว่าควรปรับเปลี่ยนสินเชื่อลูกค้าอย่างไร เช่น ควรเพิ่มลดวงเงินหรือระยะเวลาการให้สินเชื่อของลูกค้า เพื่อสนับสนุนการตัดสินใจของผู้บริหาร

4.2 การเตรียมข้อมูลสำหรับทำค่าไมนิ่ง

1. การคัดเลือกข้อมูล (Data Selection)

ข้อมูลที่นำมาใช้เป็นข้อมูลจากฐานข้อมูลของบริษัทแห่งหนึ่ง โดยทำการจัดเก็บข้อมูลไว้ในฐานข้อมูล Microsoft Access 2000 ที่ชื่อ db1 โดยข้อมูลที่จะนำมาจัดกลุ่มนี้มาจากตาราง customer ซึ่งจะทำารคัดเลือกฟิลด์ดังนี้

ตารางที่ 4.1 ข้อมูล customer ที่ทำการเลือกมาจัดกลุ่ม

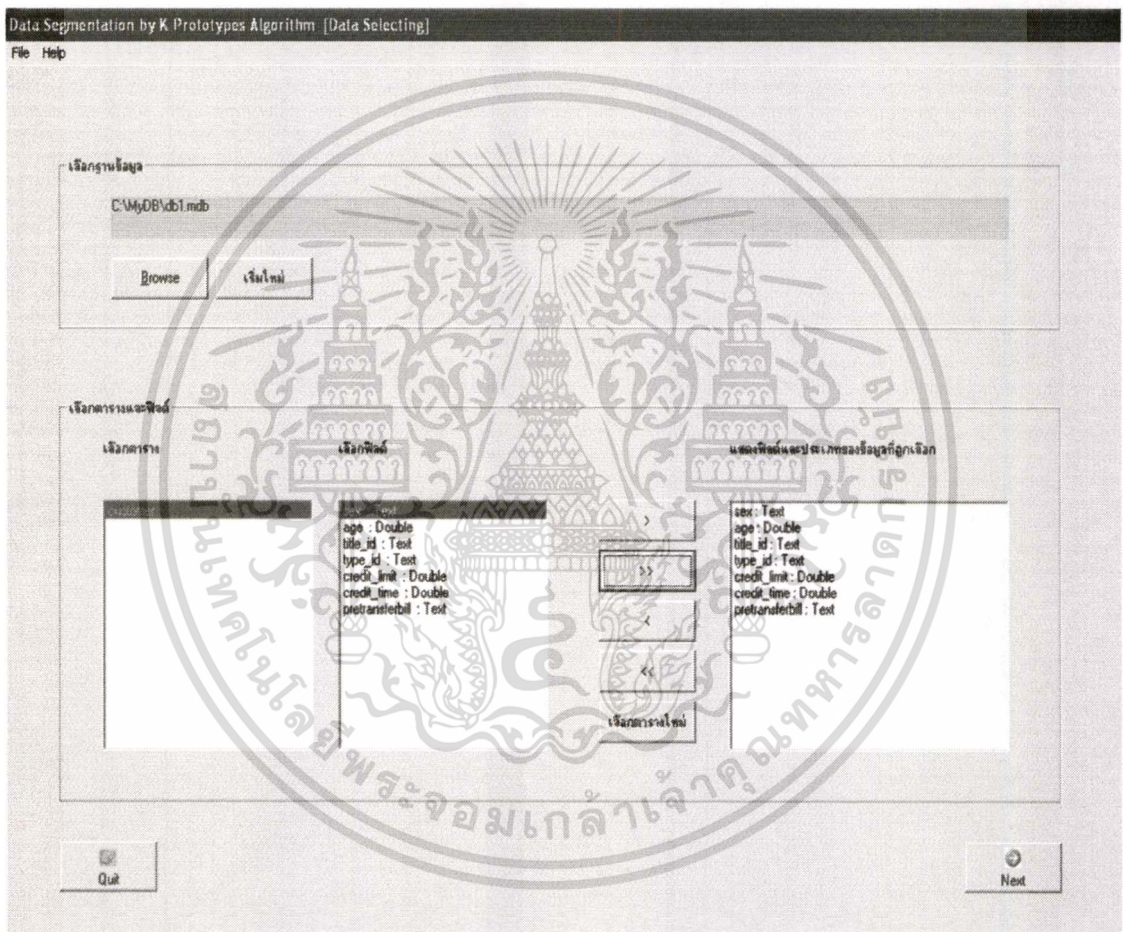
ชื่อฟิลด์ข้อมูล	ประเภทของข้อมูล
เพศ (sex)	Text (Categorical)
อายุ (age)	Double (Numerical)
อาชีพ (title_id)	Text (Categorical)
วิธีการชำระเงิน (type_id)	Text (Categorical)
วงเงินในการให้สินเชื่อ (credit_limit)	Double (Numerical)
ระยะเวลาการให้สินเชื่อ (credit_time)	Double (Numerical)

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อใช้ในการศึกษาเท่านั้น ไม่สามารถนำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 (ต่อ) ข้อมูล customer ที่ทำการเลือกมาจัดกลุ่ม

รายละเอียดการชำระเงินตรงกำหนด (pretransferbill)	Text (Categorical)
--	--------------------

ซึ่งหน้าจอการคัดเลือกข้อมูลของ โปรแกรมจะแสดงดังรูปที่ 4.1

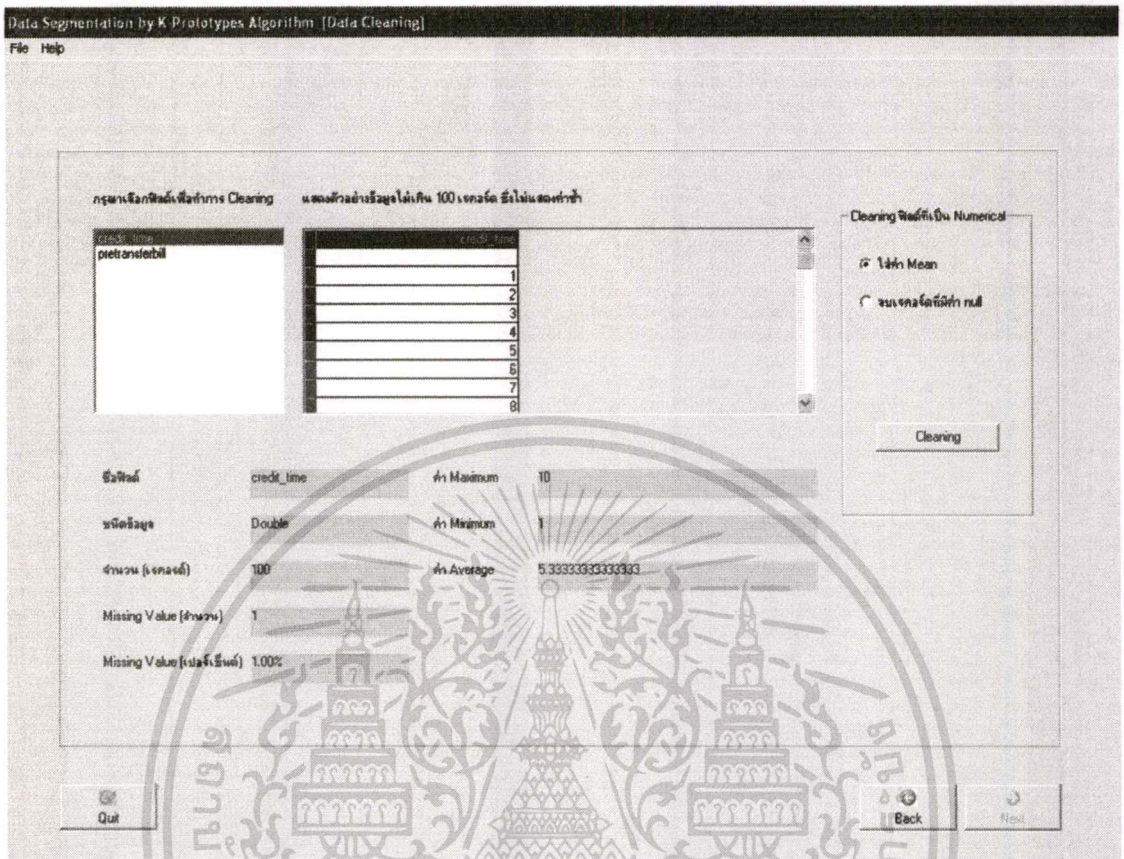


รูปที่ 4.1 หน้าจอเลือกฟิลด์ข้อมูลที่ต้องการจัดกลุ่ม

2. การแก้ไขข้อมูลที่มีค่าว่าง (Data Cleaning)

ในขั้นตอนนี้จะพบว่าฟิลด์ที่มีค่าว่าง คือฟิลด์ credit_time และฟิลด์ pretransferbill โดยการแก้ไขข้อมูลสำหรับฟิลด์ credit_time จะทำการแก้ไขโดยการเติมค่า Mean ส่วนฟิลด์ pretransferbill จะแก้ไขโดยการใส่ค่า unknown ส่วนฟิลด์อื่นๆ ไม่พบค่าว่างจึงไม่จำเป็นต้องทำการแก้ไขข้อมูล ดังแสดงในรูปที่ 4.2

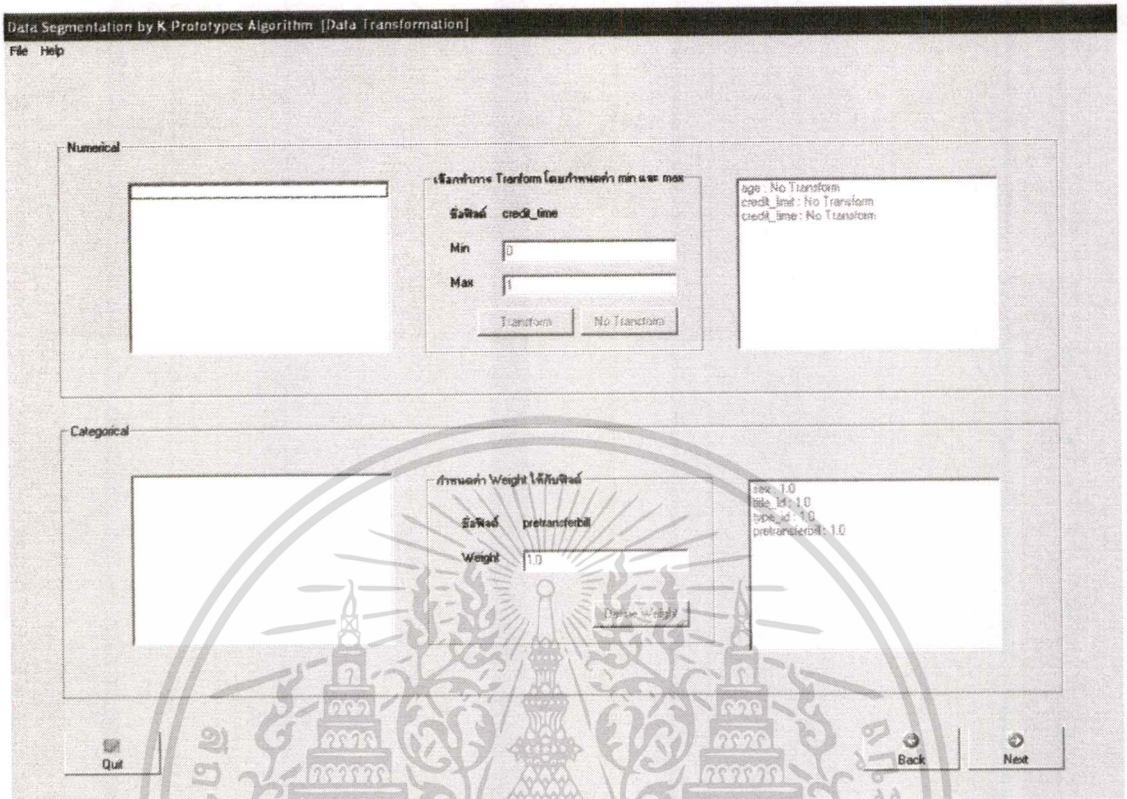
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 หน้าจอการแก้ไขข้อมูล

3. การปรับเปลี่ยนข้อมูล (Data Transformation)

ในขั้นตอนนี้จะไม่ทำการ transform ฟิวด์ข้อมูลที่เป็น numerical ซึ่งได้แก่ age, credit_limit, credit_time เนื่องจากการนำเสนอข้อมูลต้องการแสดงผลข้อมูล numerical ที่ไม่ถูกปรับเปลี่ยนค่าให้อยู่ในช่วงต่างๆ ส่วนฟิวด์ข้อมูลที่เป็น categorical จะกำหนดค่า weight ให้เป็น 1.0 ทั้งหมด ดังแสดงในรูปที่ 4.3



รูปที่ 4.3 หน้าจอการปรับเปลี่ยนข้อมูล

4.3 การทำคาด้าไมนิ่งในการจัดกลุ่มข้อมูล

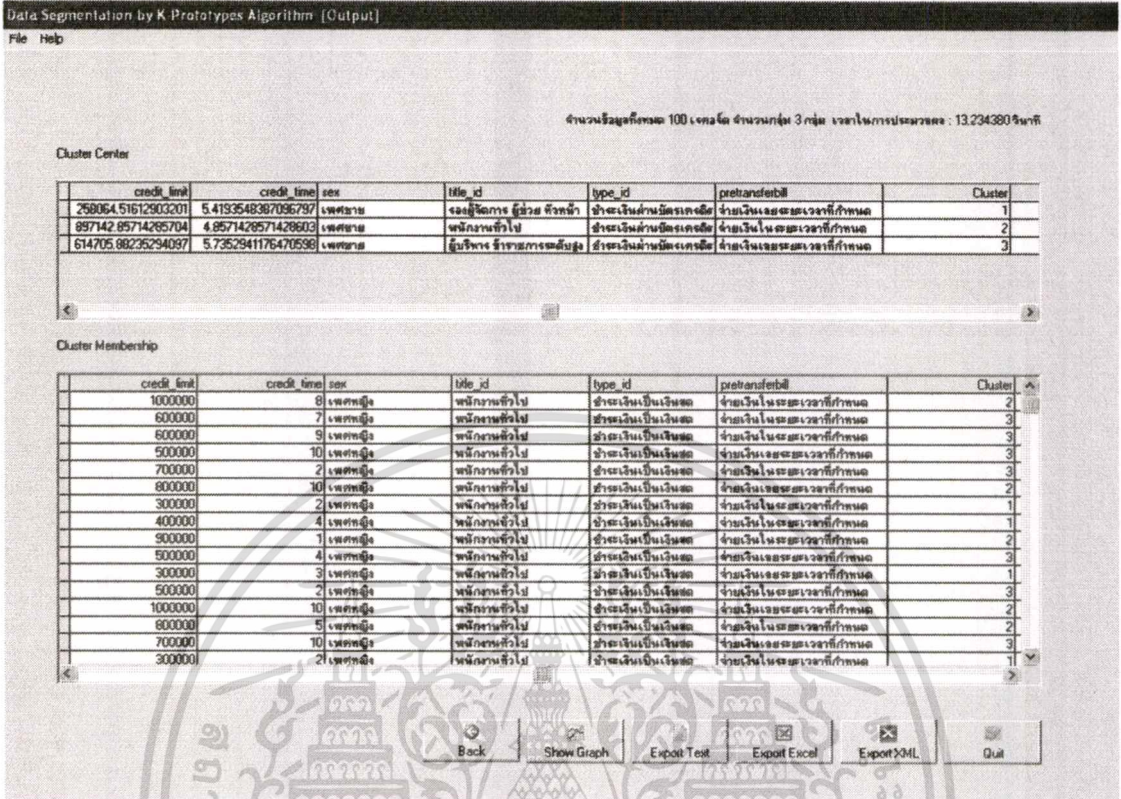
ขั้นตอนการทำคาด้าไมนิ่งเป็นการนำข้อมูลที่ผ่านการเตรียมข้อมูลทั้ง 3 ขั้นตอนที่ได้กล่าวมาแล้วในหัวข้อที่ 4.2 ซึ่งขั้นตอนนี้จะเป็นการกำหนดจำนวนกลุ่มข้อมูลที่ต้องการแบ่งออกเป็น 3 กลุ่ม โดยนำมาจากวัตถุประสงค์ ดังแสดงในรูปที่ 4.4

sex	age	title_id	type_id	credit limit	credit time	pretransferbill
เพศหญิง	26	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	1000000	8	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	28	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	7	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	26	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	9	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	21	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	500000	10	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	26	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	700000	2	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	19	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	10	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	29	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	300000	2	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	24	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	400000	4	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	28	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	900000	1	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	23	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	500000	4	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	27	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	300000	3	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	24	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	500000	2	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	25	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	1000000	10	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	26	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	5	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	28	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	700000	10	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	27	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	300000	2	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	24	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	200000	9	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	31	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	2	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	22	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	9	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	22	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	7	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศหญิง	22	พนักงานทั่วไป	ชำระหนี้เงินเงินสด	600000	8	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศชาย	30	ข้าราชการทั่วไป ชั้นที่ ๓	ชำระหนี้ผ่านบัตรเครดิต	200000	10	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศชาย	27	ข้าราชการทั่วไป ชั้นที่ ๓	ชำระหนี้ผ่านบัตรเครดิต	1000000	9	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศชาย	25	ข้าราชการทั่วไป ชั้นที่ ๓	ชำระหนี้ผ่านบัตรเครดิต	900000	7	ชำระหนี้เงินระยะเวลาที่กำหนด
เพศชาย	25	ข้าราชการทั่วไป ชั้นที่ ๓	ชำระหนี้ผ่านบัตรเครดิต	800000	5	ชำระหนี้เงินระยะเวลาที่กำหนด

กำหนดจำนวนกลุ่มที่ใช้ในการจัดกลุ่ม

รูปที่ 4.4 หน้าจอการกำหนดจำนวนกลุ่มข้อมูล

หลังจากที่ใส่จำนวนกลุ่มที่ต้องการทำการจัดกลุ่มแล้ว โปรแกรมจะทำการจัดกลุ่มข้อมูล โดยผ่านอัลกอริทึม K-Prototypes โดยหน้าจอ จะแบ่งเป็น 2 ส่วน โดยส่วนแรกจะแสดงจุดศูนย์กลางของกลุ่มข้อมูลที่ได้ทำการจัดกลุ่มและจำนวนข้อมูลที่อยู่ในกลุ่ม ส่วนที่สองจะแสดงข้อมูลทั้งหมดว่าอยู่ในกลุ่มใด โดยทั้งสองส่วนนี้จะบอกถึงค่า distance ซึ่งเป็นระยะห่างจากจุดศูนย์กลางของกลุ่มข้อมูล ดังแสดงในรูปที่ 4.5



รูปที่ 4.5 หน้าจอแสดงผลการจัดกลุ่มข้อมูล

4.4 การวิเคราะห์ผลที่ได้จากการทำค้ำไม้หนึ่งและการนำความรู้มาใช้

หลังจากทำการ ไม้หนึ่งเสร็จแล้วจะได้ผลลัพธ์ ซึ่งจะต้องนำมาวิเคราะห์ต่อ โดยต้องตรงกับวัตถุประสงค์ที่ตั้งไว้ คือ ข้อมูลอายุ (age) ข้อมูลเพศ (sex) ข้อมูลอาชีพ (title_id) ข้อมูลวิธีการชำระเงิน (type_id) ข้อมูลวงเงินการให้สินเชื่อ (credit_limit) ข้อมูลระยะเวลาการให้สินเชื่อ (credit_time) และ ข้อมูลรายละเอียดการชำระเงินตรงกำหนด (pretransferbill) จำนวน 100 เรคอร์ด โดยแบ่งเป็น 3 กลุ่ม ผลลัพธ์พบว่า

กลุ่มที่ 1 มีจำนวนข้อมูลทั้งหมดในกลุ่มนี้เท่ากับ 31 เรคอร์ด โดยลักษณะทั่วไปของกลุ่มนี้คือ อายุ 36.8387096774194 ปี เป็น เพศชาย อาชีพ รองผู้จัดการ ผู้ช่วย หัวหน้า โดยชำระเงินผ่านบัตรเครดิต มีวงเงินการให้สินเชื่อ 258064.516129032 บาท ระยะเวลาการให้สินเชื่อ 5.41935483870968 หน่วย โดยกลุ่มนี้จ่ายเงินโดยระยะเวลาที่กำหนด

กลุ่มที่ 2 มีจำนวนข้อมูลทั้งหมดในกลุ่มนี้เท่ากับ 35 เรคอร์ด โดยลักษณะทั่วไปของกลุ่มนี้คือ อายุ 36.8387096774194 ปี เป็น เพศชาย อาชีพ พนักงานทั่วไป โดยชำระเงินผ่านบัตรเครดิต มี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วงเงินการให้สินเชื่อ 897142.857142857 บาท ระยะเวลาการให้สินเชื่อ 4.85714285714286หน่วย โดยกลุ่มนี้จ่ายเงินในระยะเวลาที่กำหนด

กลุ่มที่ 3 มีจำนวนข้อมูลทั้งหมดในกลุ่มนี้เท่ากับ 34 เรคอร์ด โดยลักษณะทั่วไปของกลุ่มนี้ คือ อายุ 38.5294117647059 ปี เป็น เพศชาย อาชีพ ผู้บริหาร ข้าราชการระดับสูง โดยชำระเงินผ่าน บัตรเครดิต มีวงเงินการให้สินเชื่อ 614705.882352941 บาท ระยะเวลาการให้สินเชื่อ 5.73529411764706 หน่วย โดยกลุ่มนี้จ่ายเงินเลยระยะเวลาที่กำหนด

ผลจากการจัดกลุ่มทั้งสามกลุ่มดังที่กล่าวมาแล้ว จะพบว่ากลุ่มที่ 1 และ กลุ่มที่ 3 จ่ายเงินเลย ระยะเวลาที่กำหนด ซึ่งจะมีวงเงินการให้สินเชื่อน้อยกว่ากลุ่มที่จ่ายเงินในระยะเวลาที่กำหนด แต่ทั้งสองกลุ่มนี้มีระยะเวลาการให้สินเชื่อกว่ากลุ่มที่จ่ายเงินในระยะเวลาที่กำหนดเช่นกัน ส่วนกลุ่มที่ 2 ถือว่าเป็นลูกค้าชั้นดีซึ่งมีอาชีพเป็นพนักงานทั่วไป โดยกลุ่มนี้จ่ายเงินในระยะเวลาที่กำหนดและเป็นกลุ่มที่มีสมาชิกในกลุ่มมากที่สุด ได้วงเงินมากที่สุดแต่กลับ ได้ระยะเวลาการชำระเงินน้อยที่สุด ทั้งนี้ผู้บริหารอาจต้องปรับลดวงเงินของกลุ่มที่ 1 และกลุ่มที่ 3 ที่จ่ายเงินเลยระยะเวลาที่กำหนด และ อาจเพิ่มระยะเวลาการชำระเงินสำหรับกลุ่มที่ 2 ซึ่งทั้งนี้อาจต้องใช้กระบวนการวิเคราะห์เชิงธุรกิจ เข้ามาช่วยตัดสินใจต่อไป

บทที่ 5

สรุปผลการศึกษาและข้อเสนอแนะ

5.1 สรุปผลการศึกษา

จากการที่ได้ศึกษาทฤษฎีของคาค้าไมนิ่ง ทำให้ได้เรียนรู้ว่าคาค้าไมนิ่งเป็นกระบวนการที่ค้นหาข้อมูลที่เป็นประโยชน์จากภายในฐานข้อมูลที่มีอยู่ ทำให้ได้รับสารสนเทศที่เป็นประโยชน์ และสามารถนำสารสนเทศนั้นไปช่วยสนับสนุนการตัดสินใจและการประยุกต์นำไปใช้งานกับธุรกิจต่างๆได้ ซึ่งกระบวนการดังกล่าวจะเริ่มตั้งแต่การกำหนดวัตถุประสงค์ของการทำคาค้าไมนิ่ง จากนั้นก็มีขั้นตอนการเตรียมข้อมูลมาวิเคราะห์ ซึ่งจะประกอบไปด้วยการคัดเลือกข้อมูล การทำความสะอาดข้อมูล และการแปลงข้อมูลให้เหมาะสม หลังจากนั้นก็จะทำคาค้าไมนิ่งเมื่อข้อมูลผ่านการทำไมนิ่ง ก็จะได้ผลลัพธ์ที่เกิดประโยชน์ในทางธุรกิจ

เมื่อได้ศึกษาหลักการดังกล่าวแล้ว จึงได้มีแนวทางในการพัฒนาเครื่องมือในการจัดกลุ่มข้อมูล โดยใช้อัลกอริทึม K-Prototypes ซึ่งเป็นคาค้าไมนิ่งในการจัดกลุ่มข้อมูลประเภทหนึ่งที่สามารถรองรับข้อมูลประเภท Categorical และ Numerical ได้เป็นอย่างดี สามารถทำงานกับฐานข้อมูล Microsoft Access ได้อย่างกว้างขวางเนื่องจากพัฒนาโปรแกรมไม่ได้เจาะจงกับฐานข้อมูลตัวใดตัวหนึ่งเท่านั้น ทำให้นำไปงานได้จริง

5.2 ประโยชน์ที่ได้จากการศึกษาและพัฒนาระบบ

1. ทำให้ทราบถึงปัญหาและเข้าใจการพัฒนาระบบอย่างเป็นขั้นตอน
2. สามารถนำทฤษฎีที่ได้ศึกษามาปฏิบัติงานจริงให้เป็นรูปธรรม

5.3 ข้อเสนอแนะ

1. ระบบที่พัฒนาขึ้นสามารถใช้ได้กับฐานข้อมูล Microsoft Access เท่านั้น ไม่สามารถทำการวิเคราะห์กับฐานข้อมูลอื่นๆนอกเหนือจากนี้ได้
2. ระบบที่พัฒนาขึ้นอาจใช้เวลาในการประมวลผลเพื่อแสดงผลลัพธ์ต่างกัน โดยขึ้นอยู่กับความเร็วของหน่วยประมวลผลกลาง ขนาดของหน่วยความจำ จำนวนเรคอร์ดและจำนวนแอททริบิวต์ที่นำมาประมวลผล และจำนวนครั้งของการย้ายกลุ่ม เป็นต้น

3. การวัดระยะห่างของอัลกอริทึม K-Prototypes ไม่สามารถบอกรายละเอียดที่เจาะลึกได้ว่าข้อมูลในฟิลด์เดียวกันนั้นแต่ละตัวห่างกันเท่าไร เช่น เมื่อวัดระยะห่างสำหรับข้อมูลที่เป็น Categorical ในกรณีฟิลด์อาชีพเมื่อเปรียบเทียบข้อมูลกับจุดศูนย์กลางของกลุ่มข้อมูล โดยผู้บริหารระดับสูงเปรียบเทียบกับผู้บริหารระดับกลาง จะมีระยะห่างเท่ากับ ผู้บริหารระดับสูงเปรียบเทียบกับพนักงานระดับล่าง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กัลยา วานิชย์บัญชา. 2546. การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows. กรุงเทพฯ:

ธรรมสาร

กฤษณะ ไวยมัย และคณะ. 2547. **Data Mining การเตรียมข้อมูลสำหรับดาต้าไมนิ่ง (2)**. [Online].

Available: http://micro.se-ed.com/content/mc189/mainframe.asp?tar=mc189_91.asp

สุพร เวทีวิทยา. 2546. ระบบการจัดกลุ่มข้อมูลลูกค้า. สัมนา 2 ภาคเรียนที่ 1 คณะเทคโนโลยี

สารสนเทศ สาขาวิชาวิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร

ลาดกระบัง.

สุวิมล คงศักดิ์ตระกูล และสุวิสาข์ โกพล. 2544. **Data Mining คืออะไร**. [Online]. Available:

<http://project.cs.kku.ac.th/2544/seminar/day2/413356-4and413357-5/datamining.doc>

Huang Zhexue. 1997. **Clustering Large Data Sets with Mixed Numeric and Categorical**

Values. [Online]. Available: <http://www.act.cmis.csiro.au/gjw/papers/apkdd.pdf>

ภาคผนวก



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.1 การติดตั้งโปรแกรม

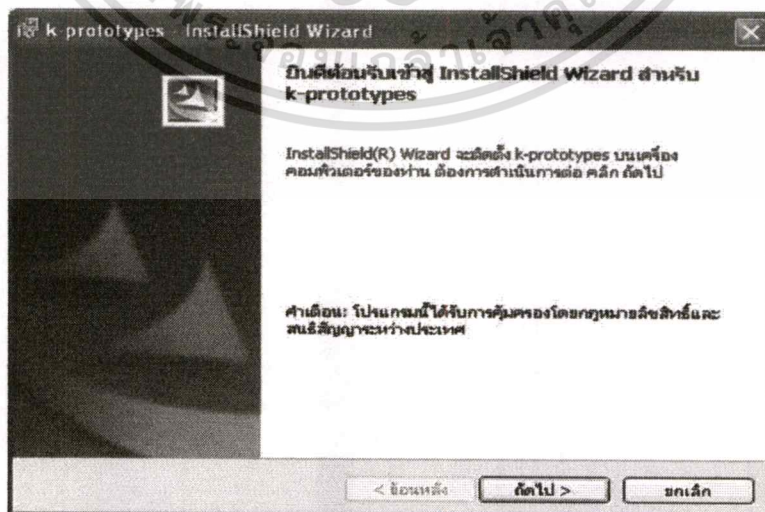
สำหรับการติดตั้ง โปรแกรมเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes มีความต้องการของระบบขั้นต่ำที่มีรายละเอียดดังนี้

- ระบบปฏิบัติการ Windows XP
- เครื่องคอมพิวเตอร์ที่ใช้หน่วยประมวลผลระดับ Pentium2 400 MHz เป็นอย่างน้อย แต่ถ้าต้องการรองรับฐานข้อมูลขนาดใหญ่จำเป็นต้องใช้เครื่องที่มีหน่วยประมวลผลมากกว่านี้
- หน่วยความจำอย่างน้อย 256 MB หากต้องการประสิทธิภาพมากกว่านี้ จำเป็นต้องใช้หน่วยความจำมากขึ้น
- พื้นที่ในการติดตั้ง 24.23 MB หรือมากกว่า ขอแนะนำให้มีพื้นที่ว่างอย่างน้อย 50 MB เพื่อเพียงพอสำหรับการประมวลผลของโปรแกรม
- จอแสดงผลขนาด 1024 * 768
- ติดตั้ง Microsoft Access 2003 และ Microsoft Excel 2003 เรียบร้อยแล้ว
- หากต้องการได้ประสิทธิภาพเพิ่มขึ้นอาจติดตั้งเครื่องพิมพ์และติดตั้ง โปรแกรม Acrobat Professional สำหรับการพิมพ์ผลลัพธ์

สำหรับการติดตั้ง โปรแกรมให้ใส่แผ่น CD ติดตั้ง โปรแกรม

1. เริ่มการติดตั้ง โปรแกรมโดยทำการดับเบิลคลิกที่ไฟล์  ที่อยู่ในไดรฟ์

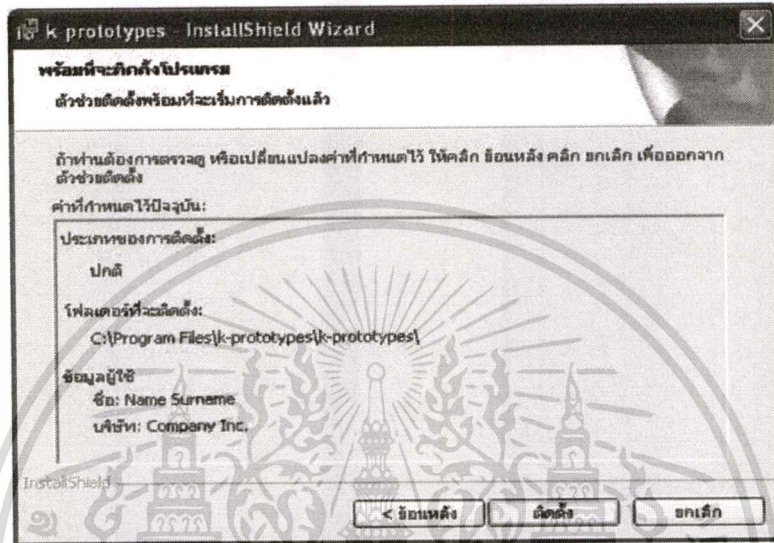
CD-ROM เพื่อใช้ในการติดตั้ง โปรแกรม หลังจากนั้นหน้าจอจะปรากฏดังรูปที่ ก.1



รูปที่ ก.1 หน้าจอแรกเมื่อเข้าสู่การติดตั้ง โปรแกรม

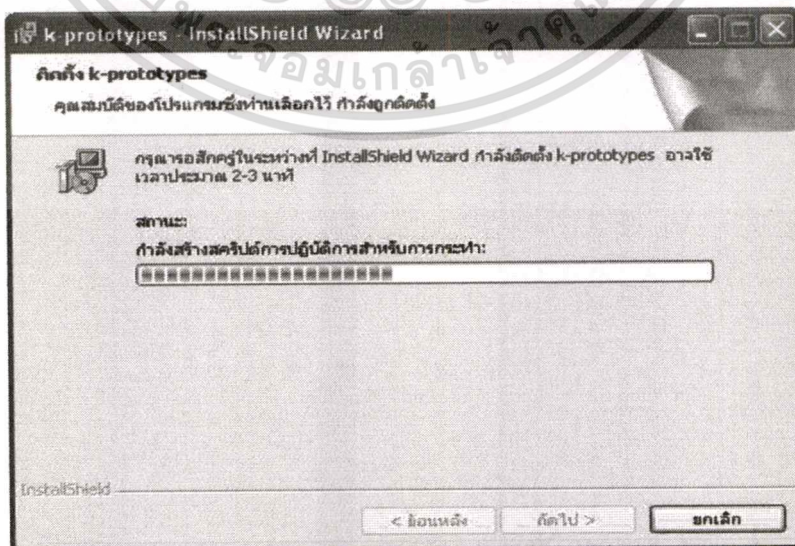
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. หลังจากคลิกที่ปุ่ม **ถัดไป >** จะเข้าสู่หน้าจอการติดตั้งโปรแกรมในหน้าที่สอง ดังแสดงในรูปที่ ก.2



รูปที่ ก.2 หน้าจอที่สองของการติดตั้งโปรแกรม

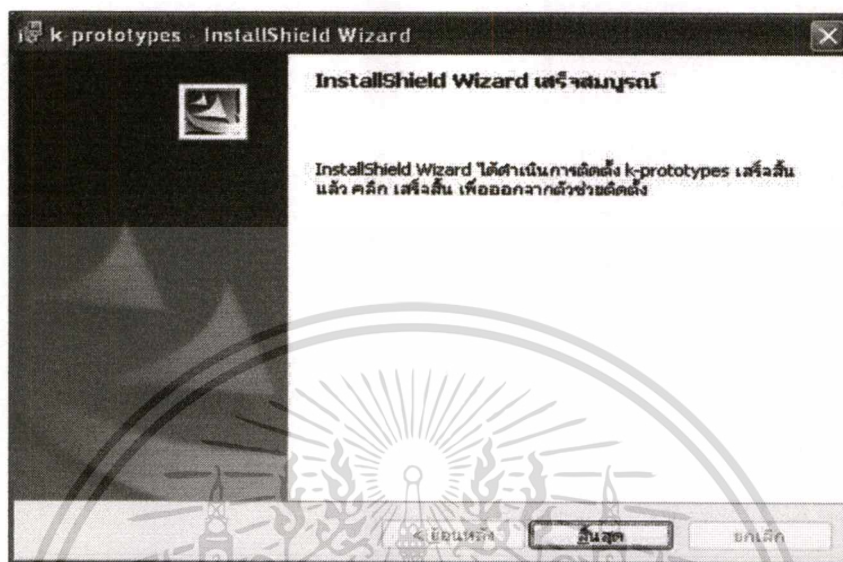
3. เมื่อคลิกปุ่ม **ติดตั้ง** ระบบจะทำการติดตั้ง ดังแสดงในรูปที่ ก.3



รูปที่ ก.3 หน้าจอแสดงกระบวนการติดตั้งระบบ

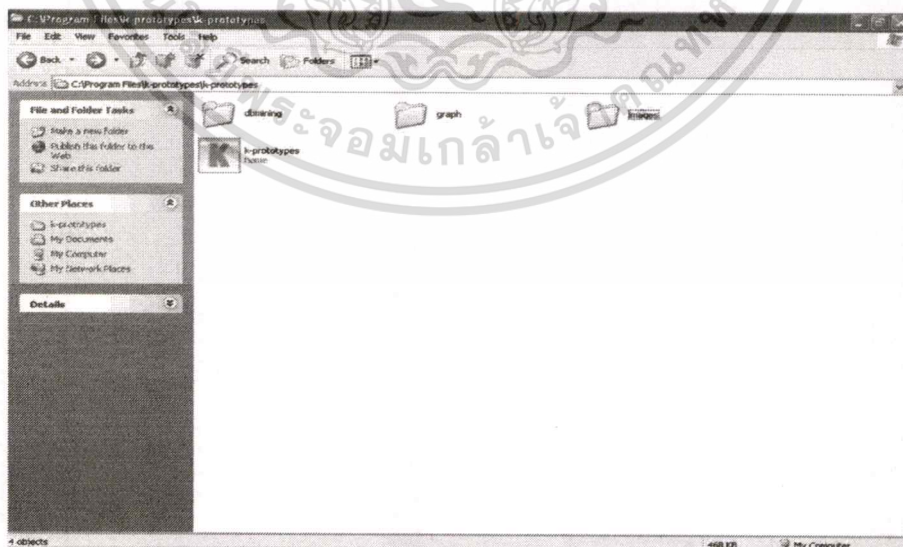
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการรู้เชิงงานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. เมื่อการติดตั้งโปรแกรมเสร็จสมบูรณ์ ให้คลิกที่ปุ่ม **สิ้นสุด**



รูปที่ ก.4 หน้าจอแสดงการติดตั้ง โปรแกรมเสร็จสมบูรณ์

โปรแกรมจะถูกติดตั้งไว้ในไดเรกทอรี C:\Program Files\k-prototypes\k-prototypes ดังแสดงในรูปที่ ก.5

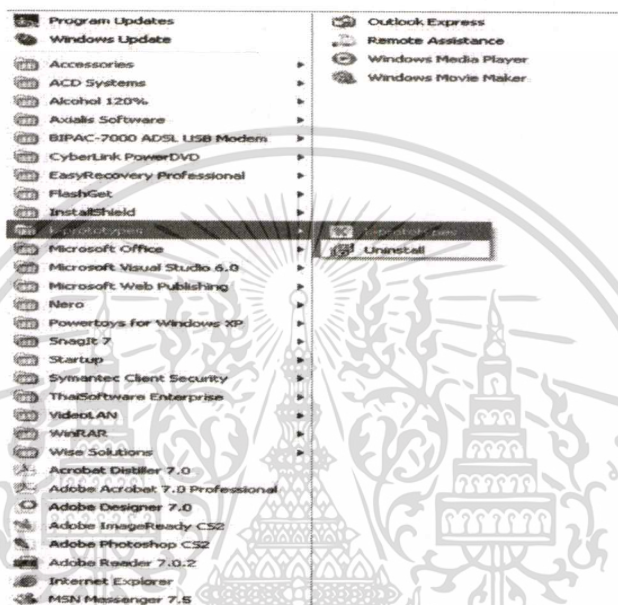


รูปที่ ก.5 แสดงไดเรกทอรีที่โปรแกรมได้ถูกติดตั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

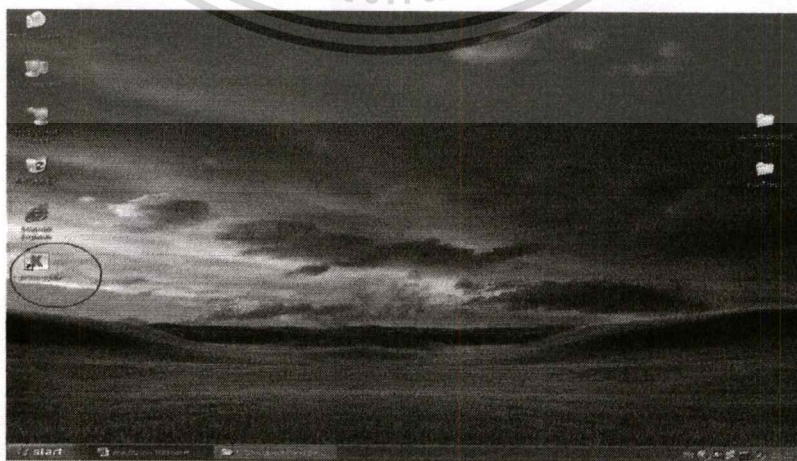
ก.2 การเข้าสู่โปรแกรม

การเข้าสู่โปรแกรม Data Segmentation by K-prototypes Algorithm สามารถทำได้โดยวิธีที่ 1. คลิกเมาส์ที่ Start Menu เข้าไปยัง All programs จะพบกับโฟลเดอร์ K-prototypes หลังจากนั้นให้เลือก Shortcut ที่ชื่อ K-prototypes ดังแสดงในรูปที่ ก.6



รูปที่ ก.6 การเข้าสู่โปรแกรม โดยเลือกจาก Start Menu

วิธีที่ 2 ดับเบิลคลิกที่ Shortcut ที่ชื่อ k-prototypes จากหน้า Desktop ดังแสดงในรูปที่ ก.7

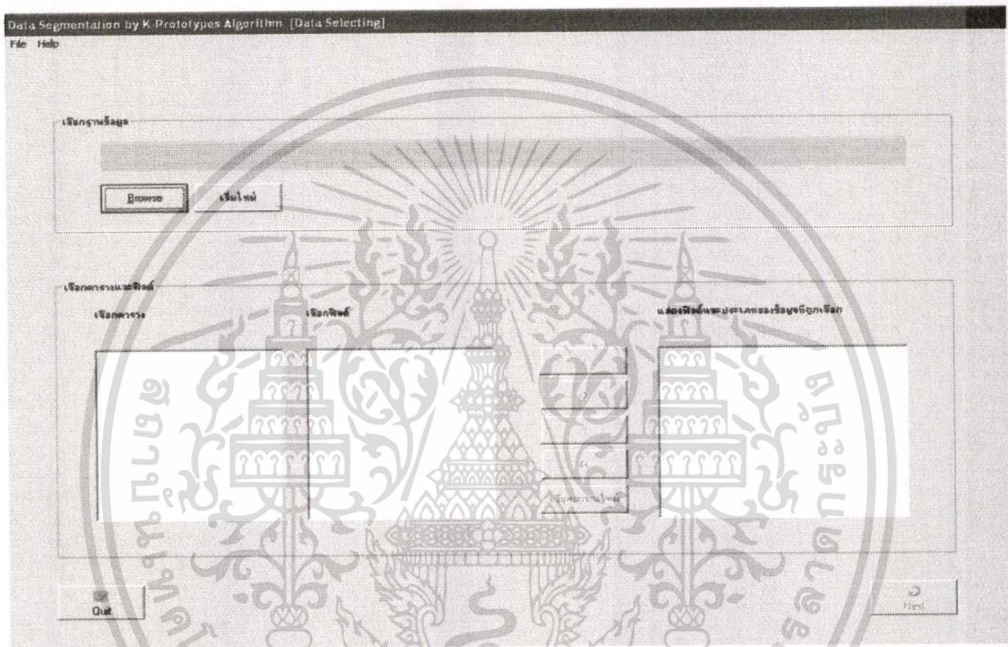


เอกสารนี้เป็นเอกสารที่สงวนไว้รูปที่ ก.7 การเข้าสู่โปรแกรม โดยเลือกจาก Shortcut เพื่อให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีที่ 3 สามารถเข้าถึงโดยระบุไปยัง path ของโปรแกรมดังนี้ C:\Program Files\k-prototypes\k-prototypes\k-prototypes.exe

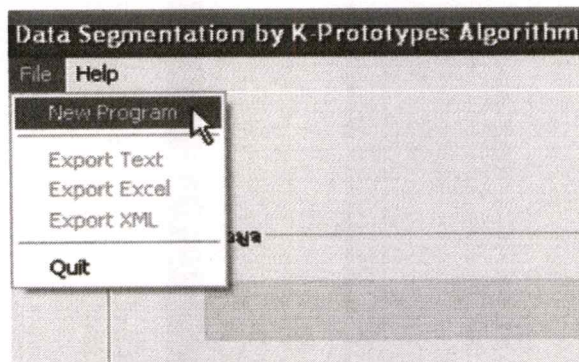
ก.3 การใช้งานโปรแกรม

เมื่อเข้าสู่โปรแกรมแล้วจะพบกับหน้าจอเริ่มต้นของ โปรแกรมดังแสดงรูปที่ ก.8



รูปที่ ก.8 หน้าจอเริ่มต้นของ โปรแกรม

เลือกที่เมนู File จะปรากฏเมนูย่อยดังแสดงในรูปที่ ก.9

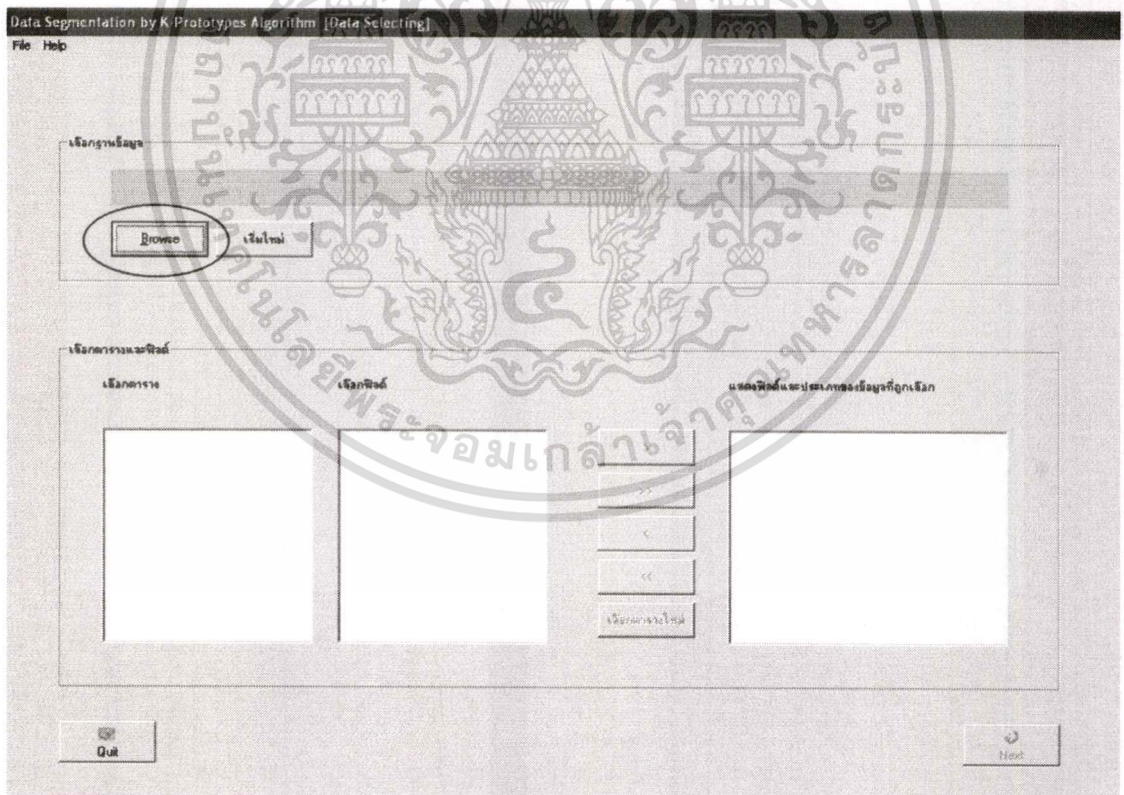


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้รูปที่ ก.9 หน้าจอแสดงผล เมนู File หน้าไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- New Program – ทำหน้าที่ในการเริ่มโปรแกรมใหม่ (Restart)
- Export Text – ทำหน้าที่ในการบันทึกผลการจัดกลุ่มข้อมูลออกมาเป็นไฟล์ประเภท Text
- Export Excel – ทำหน้าที่ในการบันทึกผลการจัดกลุ่มข้อมูลออกมาเป็นไฟล์ประเภท Excel
- Export XML – ทำหน้าที่ในการบันทึกผลการจัดกลุ่มข้อมูลออกมาเป็นไฟล์ประเภท XML
- Exit – ทำการออกจากโปรแกรม

ก.4 การเลือกฐานข้อมูล (Data Selection)

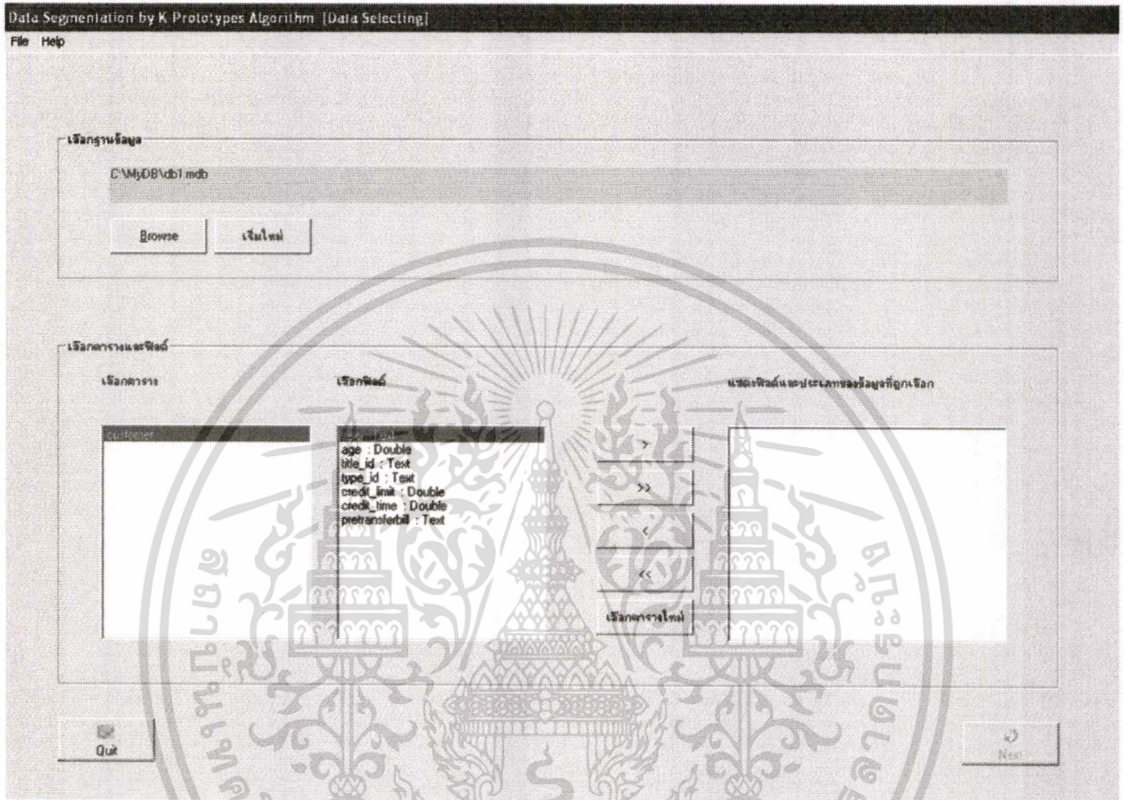
ขั้นที่ 1 คลิกที่ปุ่ม Browse เพื่อทำการเลือกฐานข้อมูล ดังแสดงในรูปที่ ก.10




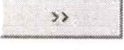
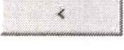
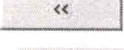
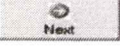
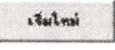
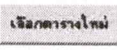
รูปที่ ก.10 หน้าจอเลือกฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นที่ 2 หลังจากเลือกฐานข้อมูลเรียบร้อยแล้ว จะปรากฏหน้าจอให้เลือกตารางและฟิลด์ ดังรูปที่ ก.11

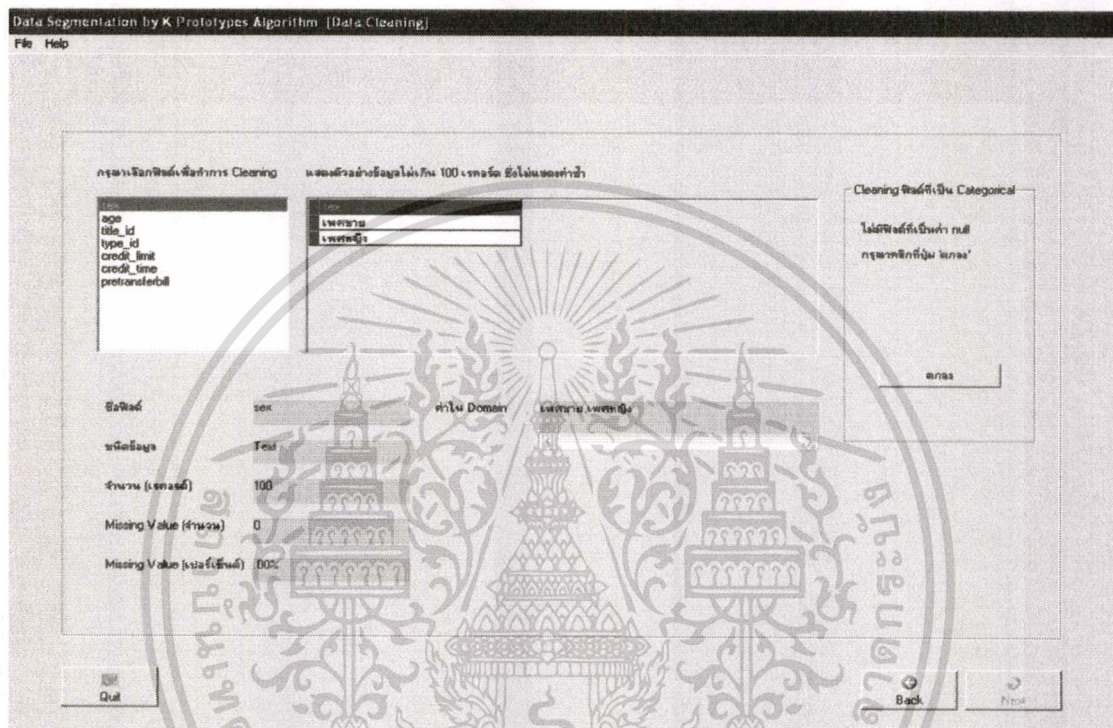


รูปที่ ก.11 หน้าจอเลือกตารางและฟิลด์

- 1) เลือกตารางที่ต้องการ
- 2) เลือกฟิลด์ข้อมูลจาก listbox จากหัวข้อเลือกฟิลด์ แล้วคลิกที่ปุ่มดังนี้
 -  เพื่อทำการเลือกฟิลด์ข้อมูลที่ต้องการ
 -  เพื่อทำการเลือกฟิลด์ข้อมูลทั้งหมด
 -  เพื่อทำการย้ายฟิลด์ข้อมูลกลับ
 -  เพื่อทำการย้ายฟิลด์ข้อมูลกลับทั้งหมด
- 3) คลิกที่ปุ่ม  เพื่อทำขั้นตอนถัดไป แต่ถ้าต้องการเลือกฐานข้อมูลใหม่ให้คลิกที่ปุ่ม  และหากต้องการเลือกตารางใหม่ให้คลิกที่ปุ่ม 

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ให้คลิกที่ปุ่ม  หากฟิลด์นี้เป็นฟิลด์สุดท้ายในการแก้ไขข้อมูล ให้คลิกปุ่ม  เพื่อทำขั้นตอนถัดไป หากไม่ใช่ฟิลด์สุดท้ายให้ดำเนินการกับฟิลด์อื่นๆแล้วแต่กรณี กรณีที่ 2 ฟิลด์ที่เป็น Categorical ที่ไม่มีค่า Null



รูปที่ ก.13 หน้าจอแสดงผลการแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ที่ไม่มีค่า Null

สำหรับฟิลด์ที่เป็น Categorical ที่ไม่มีค่า Null จะมีข้อมูลแสดงรายละเอียดต่างๆในหน้าจอดังนี้

- ส่วนแสดงตัวอย่างข้อมูลในตาราง
- ชื่อฟิลด์ - เป็นส่วนที่แสดงชื่อฟิลด์
- ชนิดข้อมูล - เป็นส่วนที่แสดงชนิดข้อมูล
- จำนวนเรคอร์ด - เป็นส่วนที่แสดงจำนวนเรคอร์ดขึ้นมาทั้งหมด
- จำนวนค่าว่าง , จำนวนเปอร์เซ็นต์ค่าว่าง
- ค่าใน domain - เป็นส่วนที่แสดงค่าในโดเมนของฟิลด์นี้ทั้งหมด

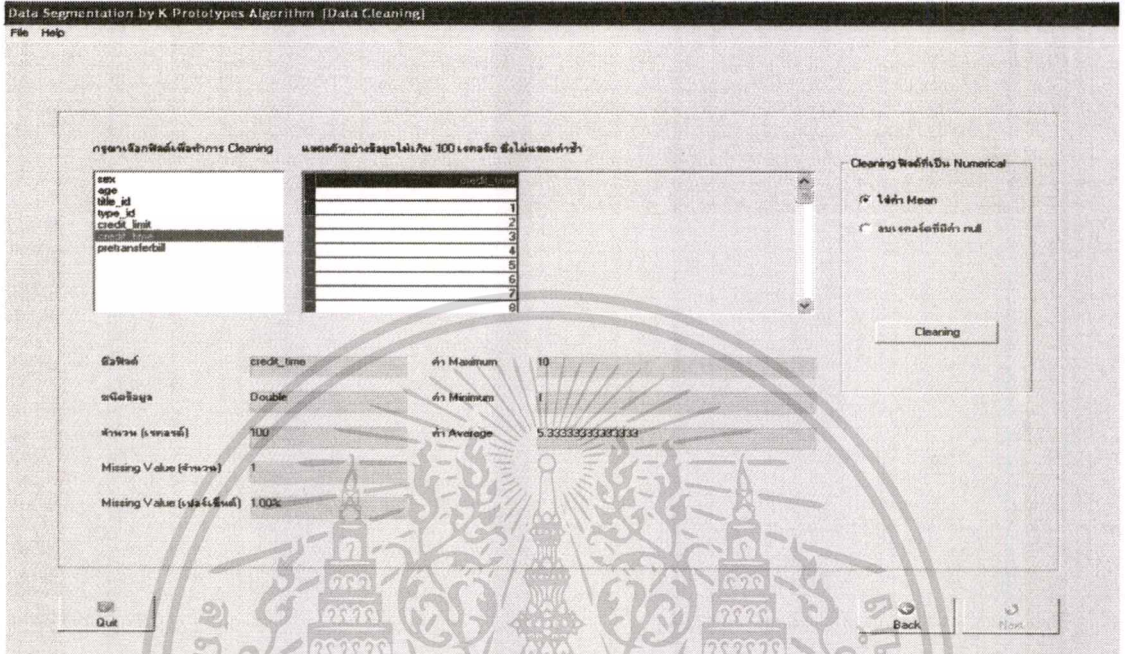
ให้คลิกที่ปุ่ม  หากฟิลด์นี้เป็นฟิลด์สุดท้ายในการแก้ไขข้อมูล ให้คลิกปุ่ม

 เพื่อทำขั้นตอนถัดไป หากไม่ใช่ฟิลด์สุดท้ายให้ดำเนินการกับฟิลด์อื่นๆแล้วแต่กรณี

เอกสารนี้เป็นเอกสารทสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีที่ 3 ฟิลด์ที่เป็น Numerical ที่มีค่า Null



รูปที่ ก.14 หน้าจอแสดงผลการแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Numerical ที่มีค่า Null

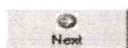
สำหรับฟิลด์ที่เป็น Numerical ที่มีค่า Null จะมีข้อมูลแสดงรายละเอียดต่างๆในหน้าจอดังนี้

- ส่วนแสดงตัวอย่างข้อมูลในตาราง
- ชื่อฟิลด์ - เป็นส่วนที่แสดงชื่อฟิลด์
- ชนิดข้อมูล - เป็นส่วนที่แสดงชนิดข้อมูล
- จำนวนเรคอร์ด - เป็นส่วนที่แสดงจำนวนเรคอร์ดขึ้นมาทั้งหมด
- จำนวนค่าว่าง , จำนวนเปอร์เซ็นต์ค่าว่าง
- ค่า Minimum , ค่า Maximum , ค่า Average

และมีตัวเลือกในการแก้ไขข้อมูลดังนี้

- ใส่ค่า Mean (ค่าเฉลี่ย) โดยจะใส่แทนข้อมูลที่เป็นค่า null
- ลบเรคอร์ดที่มีค่า null

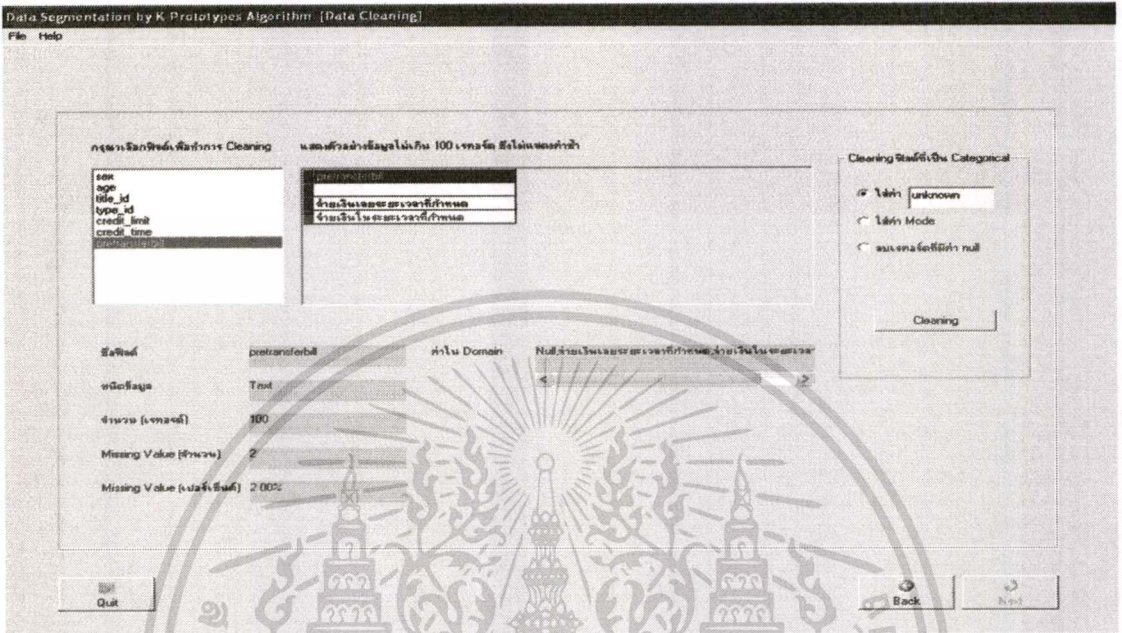
ให้คลิกที่ปุ่ม **Cleaning** หากฟิลด์นี้เป็นฟิลด์สุดท้ายในการแก้ไขข้อมูล ให้คลิกปุ่ม



เพื่อทำขั้นตอนถัดไป หากไม่ใช่ฟิลด์สุดท้ายให้ดำเนินการกับฟิลด์อื่นๆแล้วแต่กรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีศึกษาที่ 4 ฟิลด์ที่เป็น Categorical ที่มีค่า Null



รูปที่ ก.15 หน้าจอแสดงผลการแก้ไขข้อมูลสำหรับฟิลด์ที่เป็น Categorical ที่มีค่า Null

สำหรับฟิลด์ที่เป็น Categorical ที่มีค่า Null จะมีข้อมูลแสดงรายละเอียดต่างๆในหน้าจอดังนี้

- ส่วนแสดงตัวอย่างข้อมูลในตาราง
- ชื่อฟิลด์ - เป็นส่วนที่แสดงชื่อฟิลด์
- ชนิดข้อมูล - เป็นส่วนที่แสดงชนิดข้อมูล
- จำนวนเรคอร์ด - เป็นส่วนที่แสดงจำนวนเรคอร์ดขึ้นมาทั้งหมด
- จำนวนค่าว่าง , จำนวนเปอร์เซ็นต์ค่าว่าง
- ค่าใน domain - เป็นส่วนที่แสดงค่าใน โดเมนของฟิลด์นี้ทั้งหมด

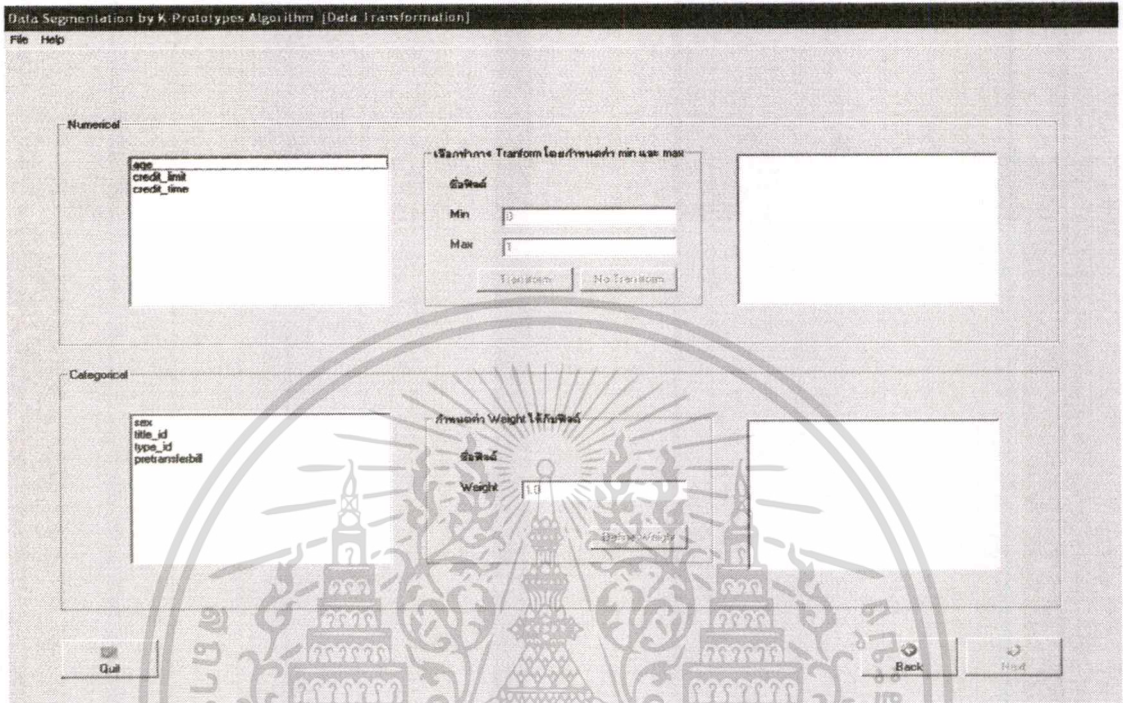
และมีตัวเลือกในการแก้ไขข้อมูลดังนี้

- ใส่ค่าโดยที่ผู้ใช้งานกำหนดเอง
- ใส่ค่า Mode(ฐานนิยม)โดยจะใส่แทนข้อมูลที่เป็นค่า null
- ลบเรคอร์ดที่มีค่า null

ให้คลิกที่ปุ่ม หากฟิลด์นี้เป็นฟิลด์สุดท้ายในการแก้ไขข้อมูล ให้คลิกปุ่ม

เอกสารนี้ Next ออก เพื่อทำขั้นตอนถัดไป หากไม่ใช่ฟิลด์สุดท้ายให้ดำเนินการกับฟิลด์อื่นๆแล้วแต่กรณี การดำเนินการนี้ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.6 การปรับเปลี่ยนข้อมูล (Data Transformation)



รูปที่ ก.16 หน้าจอแสดงผลการปรับเปลี่ยนข้อมูล

ขั้นที่ 1 ข้อมูลประเภท Numerical ให้ผู้ใช้เลือกฟิลด์ที่เป็น Numerical ใน listbox หลังจากนั้นจะมี 2 กรณีให้ผู้ใช้ได้เลือก ได้แก่

กรณีที่ 1 กำหนดให้มีการปรับเปลี่ยนข้อมูลของฟิลด์ที่เป็น Numerical มีค่าอยู่ในช่วงๆหนึ่ง กำหนดช่วงในการปรับเปลี่ยนข้อมูลในช่อง Min ใน textbox ซึ่งเป็นค่า Numerical ที่เป็นค่าน้อยที่สุดและช่อง Max ใน textbox ซึ่งเป็นค่า Numerical ที่เป็นค่าที่สูงที่สุดในฟิลด์นั้น จากนั้นให้คลิกที่ปุ่ม **Transform**

กรณีที่ 2 ไม่ต้องการให้มีการปรับเปลี่ยนข้อมูล

เมื่อไม่ต้องการปรับเปลี่ยนข้อมูล เพื่อให้ข้อมูลที่เป็น Numerical มีค่าเดียวกับค่าที่อยู่ในฐานข้อมูลตั้งแต่เริ่มต้น สามารถทำได้โดยคลิกที่ปุ่ม **No Transform**

ขั้นที่ 2 ข้อมูลประเภท Categorical ให้ผู้ใช้เลือกฟิลด์ที่เป็น Categorical ใน listbox จากนั้น ให้ใส่ค่าที่ต้องการกำหนด Weight ให้ข้อมูลใน textbox จากนั้นคลิกที่ปุ่ม **Define Weight**

ขั้นที่ 3 เมื่อทำครบทุกฟิลด์แล้วให้คลิกที่ปุ่ม **Next** เพื่อทำในขั้นตอนถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.7 การกำหนดจำนวนกลุ่มที่ต้องการจัดกลุ่ม

ขั้นตอนในการกำหนดจำนวนกลุ่มให้ข้อมูลที่ต้องการจัดกลุ่ม ให้ผู้ใช้ใส่จำนวนที่ต้องการลงในช่อง textbox จากนั้นให้คลิกที่ปุ่ม **ตกลง** ดังแสดงในรูปที่ ก.17

sex	age	title_id	type_id	credit_limit	credit_time	credit_card_status
เพศหญิง	0.17073171	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	1	0.77777779	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555556	0.66666667	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.14634146	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555556	0.99999999	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	4.8780486E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.44444445	1	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.17073171	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555556	0.11111111	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.77777779	1	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.24390244	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.22222222	0.11111111	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.12195122	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.33333334	0.33333334	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.98989899	0	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	9.7560972E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.44444445	0.33333334	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.19512194	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.22222222	0.22222222	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.12195122	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.44444445	0.11111111	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.14634146	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	1	1	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.17073171	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.77777779	0.44444445	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.66666667	0	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.19512194	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.22222222	0.11111111	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.12195122	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.11111111	0.98989899	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.25266262	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.77777779	0.11111111	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	7.3170725E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555556	0.88888889	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	7.3170725E-2	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.55555556	0.66666667	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศหญิง	0.21951219	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	0.33333334	0.77777779	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศชาย	0.26823077	ข้าราชการทั่วไป ชั้นสูง	ชำระเงินผ่านบัตรเครดิต	0.11111111	1	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศชาย	0.19512194	ข้าราชการทั่วไป ชั้นสูง	ชำระเงินผ่านบัตรเครดิต	1	0.98989899	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศชาย	0.14634146	ข้าราชการทั่วไป ชั้นสูง	ชำระเงินผ่านบัตรเครดิต	0.98989899	0.55555556	ชำระเงินโดยบัตรเครดิต/กำหนด
เพศชาย	0.14634146	ข้าราชการทั่วไป ชั้นสูง	ชำระเงินผ่านบัตรเครดิต	0.77777779	0.44444445	ชำระเงินโดยบัตรเครดิต/กำหนด

รูปที่ ก.17 หน้าจอแสดงผลการกำหนดจำนวนกลุ่มข้อมูล

หลังจากกำหนดจำนวนกลุ่มข้อมูลแล้วให้คลิกที่ปุ่ม **Next** เพื่อให้ระบบทำการประมวลผลและแสดงผลลัพธ์

ก.6 การแสดงผลการจัดกลุ่มข้อมูล

หลังจากที่ระบบได้ทำการวิเคราะห์และจัดกลุ่มข้อมูลให้เรียบร้อยแล้ว ก็จะมีการแสดงผล (Output) ซึ่งประกอบไปด้วย 2 ส่วนคือ

- Cluster Center จะแสดงจุดศูนย์กลางของแต่ละกลุ่ม และบอกจำนวนสมาชิกที่อยู่ในกลุ่มนั้นๆ
- Cluster Membership จะแสดงข้อมูลแต่ละเรคอร์ดที่อยู่ในกลุ่มใดๆ และบอกถึงระยะห่างระหว่างสมาชิกกับจุดศูนย์กลางของกลุ่มนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ระหว่างมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Segmentation by K-Prototypes Algorithm [Output]

File: Hep

จำนวนข้อมูลทั้งหมด: 100 เกรดดี จำนวนกลุ่ม: 3 กลุ่ม เวลาในการประมวลผล: 10.921880 วินาที

Cluster Center

age	credit_limit	credit_limit	sex	title	type	postcard_send
0.4246119724795401	0.7272727272727273	0.37619182630972403	เพศชาย	นางสาว/นาง	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.5312268942640021	0.44173695005765403	0.6321746147642925	เพศชาย	ผู้บริหาร	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.3844330263900499	0.932954029024703	0.438271610065296	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด

Cluster Membership

age	credit_limit	credit_limit	sex	title	type	postcard_send
0.17023170940740204	1	0.7777779102225439	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.2106121943950631	0.95959595020490079	0.55666666053405199	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.14634146736726981	0.95959595020490079	0.88888888055116272	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด

Buttons: Back, Show Graph, Export Text, Export Excel, Export HTML, Quit

รูปที่ ก.18 หน้าจอแสดงผลการจัดกลุ่มข้อมูล

ก.7 การแสดงผลการจัดกลุ่มข้อมูลในรูปแบบกราฟ

เมื่อผู้ใช้ดูผลการจัดกลุ่มข้อมูลเรียบร้อยแล้ว สามารถดูข้อมูลสรุปในรูปแบบกราฟได้ ซึ่งข้อมูลที่แสดงจะเป็นข้อมูลจุดศูนย์กลางของฟิลด์ที่เป็น Numerical เท่านั้น

Data Segmentation by K-Prototypes Algorithm [Output]

File: Hep

จำนวนข้อมูลทั้งหมด: 100 เกรดดี จำนวนกลุ่ม: 3 กลุ่ม เวลาในการประมวลผล: 10.921880 วินาที

Cluster Center

age	credit_limit	credit_limit	sex	title	type	postcard_send
0.4246119724795401	0.7272727272727273	0.37619182630972403	เพศชาย	นางสาว/นาง	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.5312268942640021	0.44173695005765403	0.6321746147642925	เพศชาย	ผู้บริหาร	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.3844330263900499	0.932954029024703	0.438271610065296	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด


Cluster Membership

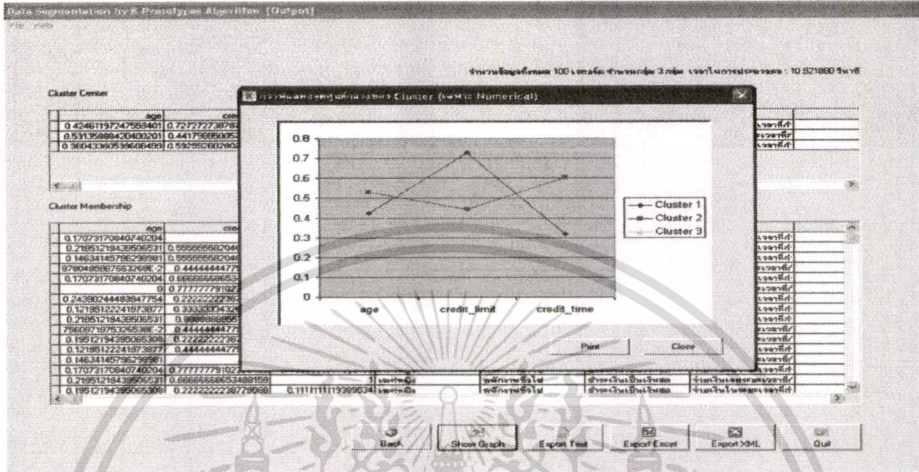
age	credit_limit	credit_limit	sex	title	type	postcard_send
0.17023170940740204	1	0.7777779102225439	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.2106121943950631	0.95959595020490079	0.55666666053405199	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด
0.14634146736726981	0.95959595020490079	0.88888888055116272	เพศหญิง	พนักงานทั่วไป	ชำระเงินเป็นเงินสด	ชำระเงินเป็นเงินสด

Buttons: Back, Show Graph, Export Text, Export Excel, Export HTML, Quit



รูปที่ ก.19 หน้าจอก่อนทำการดูผลสรุปแบบกราฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีกรนำมาใช้

ให้คลิกที่ปุ่ม  หลังจากนั้นจะมีหน้าต่างแสดงกราฟแสดงจุดศูนย์กลางของกลุ่มต่างๆ (Cluster) สำหรับข้อมูลที่เป็น Numerical ขึ้นมา ดังแสดงในรูปที่ ก.20



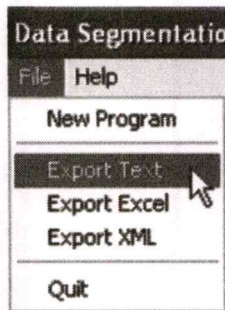
รูปที่ ก.20 หน้าจอแสดงผลรูปแบบกราฟ

- กรณีที่ต้องการพิมพ์กราฟ ให้คลิกที่ปุ่ม 
- กรณีที่ต้องการปิดหน้าต่างที่แสดงกราฟนี้ให้คลิกที่ปุ่ม 

ก.8 การบันทึกผลการจัดกลุ่มข้อมูลออกมาเป็นไฟล์ประเภท Text

1. เมื่อจัดกลุ่มข้อมูลจนเสร็จสมบูรณ์แล้ว เราสามารถที่จะบันทึกข้อมูลเพื่อสามารถเก็บไว้ดูผลลัพธ์ที่ทำการจัดกลุ่มข้อมูลได้ในรูปแบบของ Text File โดยสามารถทำได้ 2 วิธี คือ

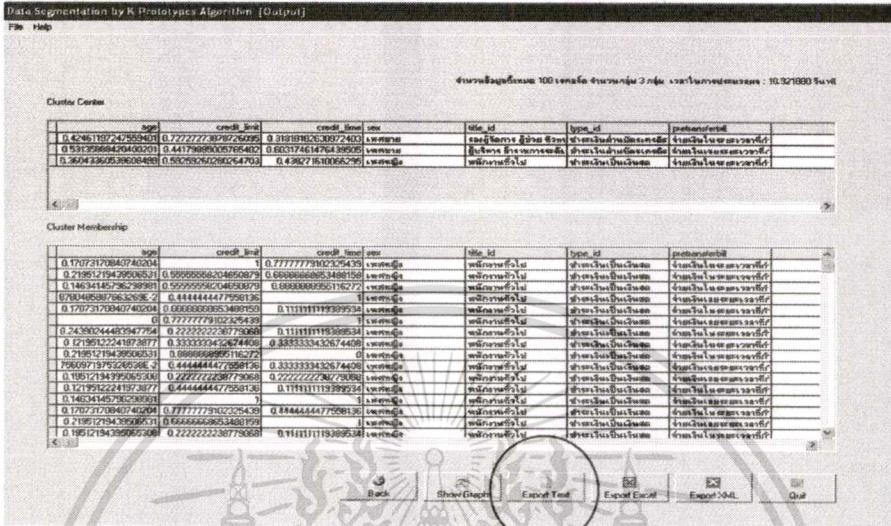
วิธีที่ 1 เลือกจากเมนู File > Export Text ดังแสดงในรูปที่ ก.21



รูปที่ ก.21 หน้าจอแสดงผลโดยเลือกเมนู File > Export Text

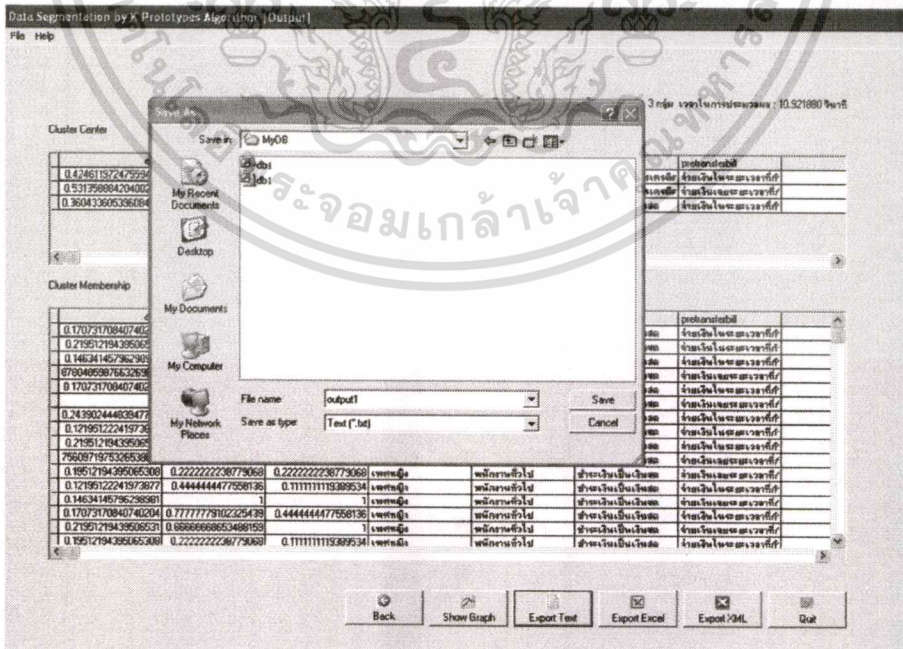
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีที่ 2 คลิกที่ปุ่ม Export Text จากหน้าจอ ดังแสดงในรูปที่ ก.22



รูปที่ ก.22 หน้าจอแสดงผลโดยเลือกคลิกที่ปุ่ม Export Text

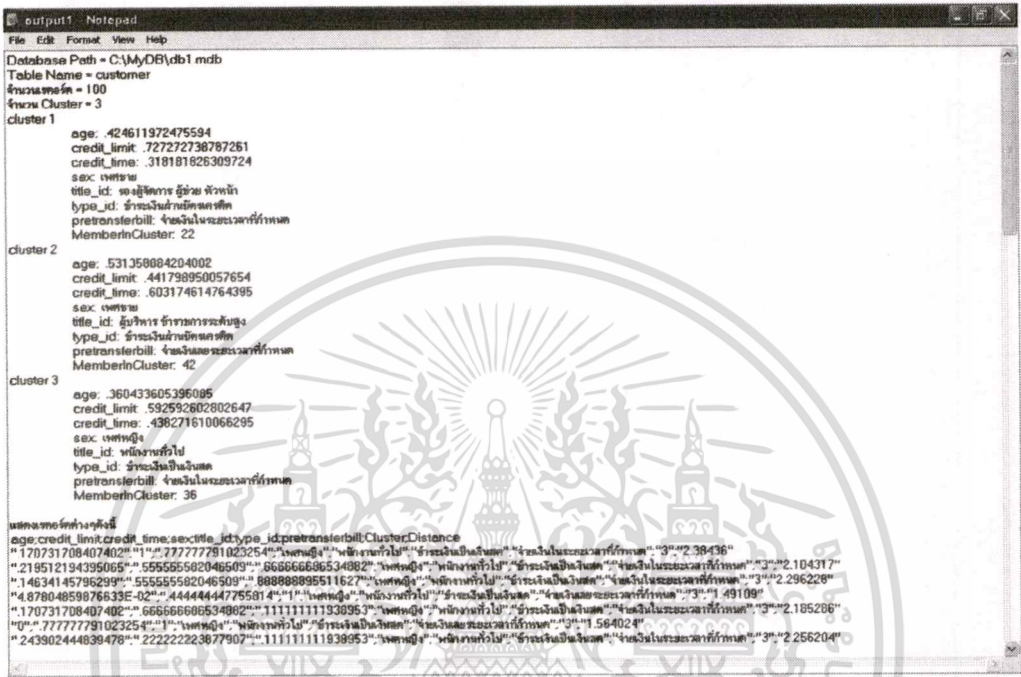
2. ระบุตำแหน่งที่ต้องการในการบันทึกข้อมูล ดังแสดงในรูปที่ ก.23



รูปที่ ก.23 หน้าจอแสดงผลการบันทึกข้อมูลแบบ Text

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เมื่อทำการบันทึกข้อมูลแล้ว เมื่อเข้าไปจะปรากฏผลการจัดกลุ่มข้อมูลในรูปของ Text File ดังแสดงในรูปที่ ก.24

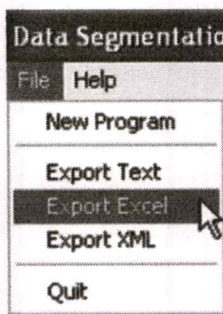


รูปที่ ก.24 หน้าจอแสดงผลข้อมูลเมื่อบันทึกข้อมูลแบบ Text

ก.9 การบันทึกผลการจัดกลุ่มข้อมูลออกมาเป็นไฟล์ประเภท Excel

1. เมื่อจัดกลุ่มข้อมูลจนเสร็จสมบูรณ์แล้ว เราสามารถที่จะบันทึกข้อมูลเพื่อสามารถเก็บไว้ดูผลลัพธ์ที่ทำการจัดกลุ่มข้อมูลได้ในรูปแบบของ Excel โดยสามารถทำได้ 2 วิธี คือ

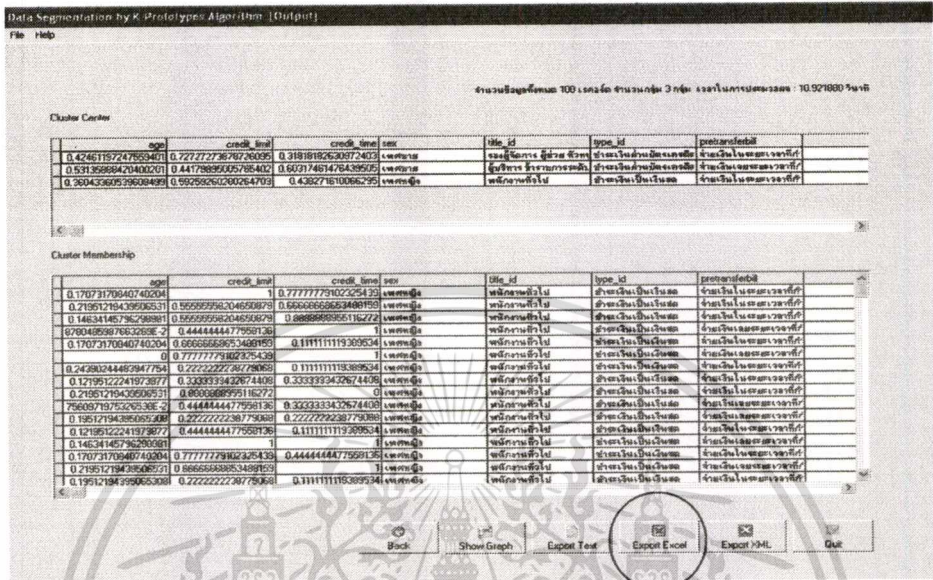
วิธีที่ 1 เลือกจากเมนู File > Export Excel ดังแสดงในรูปที่ ก.25



รูปที่ ก.25 หน้าจอแสดงผลโดยเลือกเมนู File > Export Excel

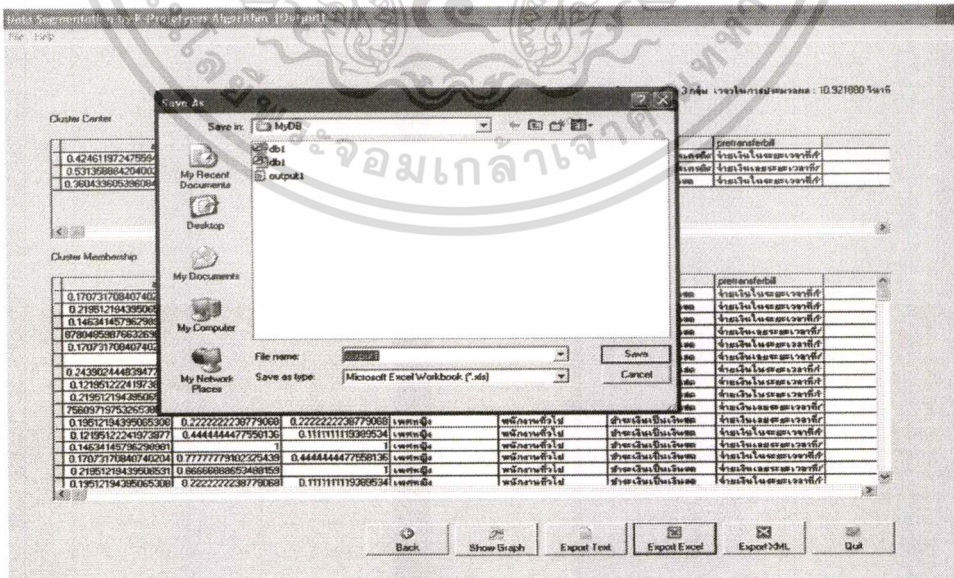
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่ไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีที่ 2 คลิกที่ปุ่ม จากหน้าจอ ดังแสดงในรูปที่ ก.26



รูปที่ ก.26 หน้าจอแสดงผล โดยเลือกคลิกที่ปุ่ม Export Excel

2. ระบุตำแหน่งที่ต้องการในการบันทึกข้อมูล ดังแสดงในรูปที่ ก.27



รูปที่ ก.27 หน้าจอแสดงผลการบันทึกข้อมูลแบบ Excel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิได้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เมื่อทำการบันทึกข้อมูลแล้ว เมื่อเข้าไปจะปรากฏผลจัดกลุ่มข้อมูลในรูปของ Excel ซึ่ง เป็นแบ่ง 2 Worksheet ใน Worksheet ชื่อ Center จะแสดงจุดศูนย์กลางข้อมูลและข้อมูลสรุปต่าง ๆ ดังแสดงในรูปที่ ก.28

Database Path	C:\MyDB\br1.mdb	
Table Name	customer	
จำนวนสมาชิก	100	
จำนวน Cluster	3	
Cluster 1		
age	credit_limit	0.424611972
credit_limit	credit_time	0.727272738
sex	credit_time	0.918181826
title_id	sex	เพศชาย
type_id	title_id	ข้าราชการ (พิเศษ หรือ ทั่วไป)
praftransfirdall	type_id	ข้าราชการกรมวิเทศคดี
MembersInCluster	praftransfirdall	ข้าราชการกรมวิเทศคดี
	MembersInCluster	22
Cluster 2		
age	credit_limit	0.531366884
credit_limit	credit_time	0.441738995
sex	credit_time	0.883174515
title_id	sex	เพศชาย
type_id	title_id	ข้าราชการ (พิเศษ หรือ ทั่วไป)
praftransfirdall	type_id	ข้าราชการกรมวิเทศคดี
MembersInCluster	praftransfirdall	ข้าราชการกรมวิเทศคดี
	MembersInCluster	42
Cluster 3		
age	credit_limit	0.360433605
credit_limit	credit_time	0.592592603
sex	credit_time	0.6927161
title_id	sex	เพศชาย
type_id	title_id	ข้าราชการทั่วไป
praftransfirdall	type_id	ข้าราชการกรมวิเทศคดี
MembersInCluster	praftransfirdall	ข้าราชการกรมวิเทศคดี
	MembersInCluster	36

รูปที่ ก.28 หน้าจอแสดงผลข้อมูลเมื่อบันทึกข้อมูลแบบ Excel จาก Worksheet ชื่อ Center

ส่วน Worksheet ชื่อ Member จะแสดงพิกัด และกลุ่มที่พิกัดนั้นอยู่ รวมถึงแสดงระยะห่าง (Distance) ดังแสดงในรูปที่ ก.29

age	credit_limit	credit_time	sex	title_id	type_id	praftransfirdall	Cluster	Distance
0.170731700	1	0.777777791	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.39436
0.219612194	0.555555582	0.666666687	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.104317
0.146341458	0.555555582	0.988888896	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.296228
0.048780486	0.444444448	1	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	1.49103
0.170731700	0.656565697	0.111111112	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.185236
0.077777791	1	0.777777791	1	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	3	1.564024
0.243902445	0.222222224	0.111111112	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.298204
0.121961222	0.333333343	0.333333343	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.162589
0.219612194	0.888888896	0	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.340429
0.097560972	0.444444448	0.333333343	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	1.140508
0.196121944	0.222222224	0.222222224	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	1.19053
0.121961222	0.444444448	0.111111112	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.217001
0.146341458	1	1	1	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	3	1.801793
0.170731700	0.777777791	0.444444448	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.118529
0.219612194	0.656565697	0.777777791	1	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	3	1.383639
0.196121944	0.222222224	0.111111112	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.277694
0.121961222	0.111111112	0.888888896	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	1.514841
0.292682916	0.777777791	0.111111112	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.170229
0.073170729	0.555555582	0.888888896	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	2.344525
0.073170729	0.555555582	0.666666687	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	3	1.190204
0.219612194	0.333333343	0.777777791	1	เพศหญิง	พนักงานทั่วไป	ข้าราชการกรมวิเทศคดี	3	1.221601
0.268292686	0.111111112	0.111111112	1	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	2	1.348529
0.196121944	1	0.888888896	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	2	1.415563
0.146341458	0.888888896	0.666666687	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	1	1.226113
0.146341458	0.777777791	0.444444448	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	1	1.162005
0.097560972	0.777777791	0.666666687	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	1	1.212853
0.121961222	0.111111112	0.555555582	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	2	1.334208
0.219612194	0.888888896	0.555555582	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	1	2.02653
0.024390245	0.888888896	0.888888896	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	1	1.31817
0.196121944	0.888888896	0.444444448	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	1	1.145014
0.219612194	0.656565697	0.333333343	เพศชาย	ข้าราชการทั่วไป	ข้าราชการกรมวิเทศคดี	ข้าราชการกรมวิเทศคดี	1	1.07662

เอกสารนี้รูปที่ ก.29 หน้าจอแสดงผลข้อมูลเมื่อบันทึกข้อมูลแบบ Excel จาก Worksheet ชื่อ Member ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมีได้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.10 การบันทึกผลการจัดกลุ่มข้อมูลออกมาเป็นไฟล์ประเภท XML

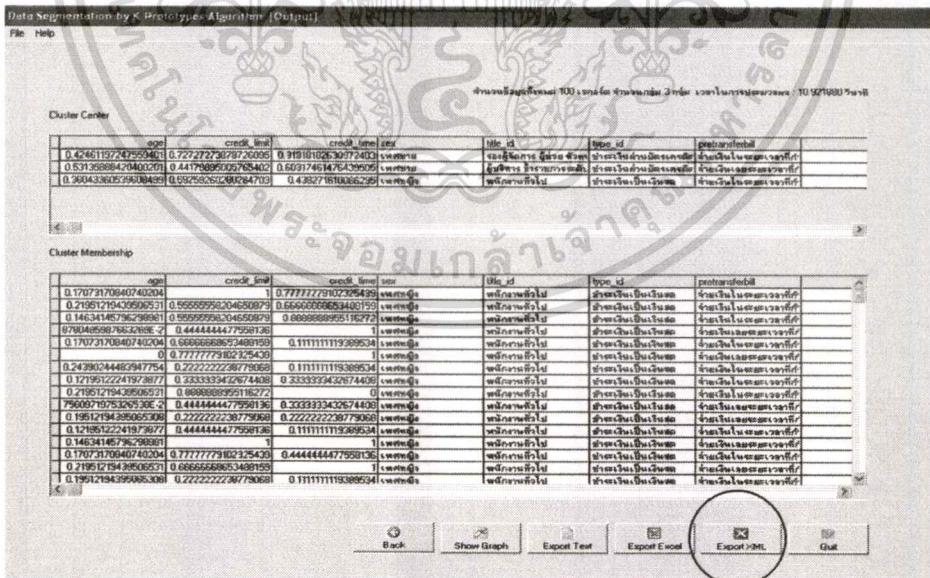
1. เมื่อจัดกลุ่มข้อมูลจนเสร็จสมบูรณ์แล้ว เราสามารถที่จะบันทึกข้อมูลเพื่อสามารถเก็บไว้ดูผลลัพธ์ที่ทำการจัดกลุ่มข้อมูลได้ในรูปแบบของ XML โดยสามารถทำได้ 2 วิธี คือ

วิธีที่ 1 เลือกจากเมนู File > Export XML ดังแสดงในรูปที่ ก.30



รูปที่ ก.30 หน้าจอแสดงผล โดยเลือกเมนู File > Export XML

วิธีที่ 2 คลิกที่ปุ่ม  จากหน้าจอ ดังแสดงในรูปที่ ก.31



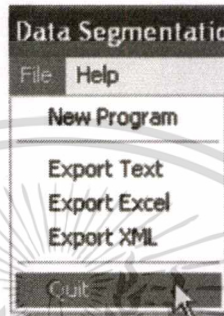
รูปที่ ก.31 หน้าจอแสดงผลการบันทึกข้อมูลแบบ XML

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.11 การปิดโปรแกรม

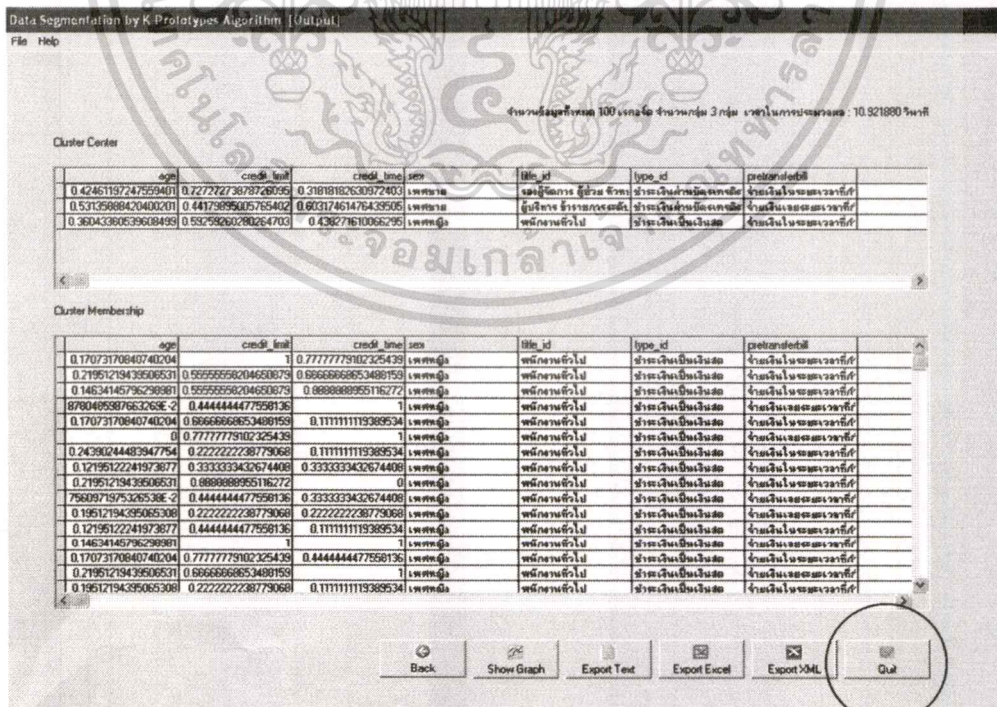
การปิดโปรแกรมเครื่องมือในการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-Prototypes สามารถเลือกวิธีในการปิดโปรแกรมดังกล่าวได้ 2 วิธีดังนี้

วิธีที่ 1 เลือกจากเมนู File > Quit ดังแสดงในรูปที่ ก.34



รูปที่ ก.34 หน้าจอการปิด โปรแกรมโดยเลือกเมนู File > Quit

วิธีที่ 2 คลิกที่ปุ่ม Quit จากหน้าจอทุกๆหน้าที่มีปุ่มดังกล่าวปรากฏ



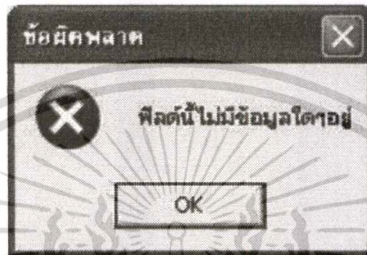
รูปที่ ก.35 หน้าจอแสดงการปิดโปรแกรมโดยเลือกที่ปุ่ม Quit

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.12 Message Box ต่างๆที่มีในโปรแกรมดังนี้

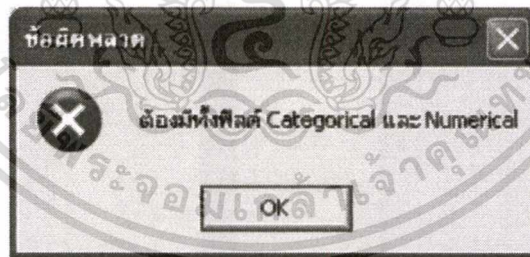
หน้าจอของการแก้ไขข้อมูล (Data Cleaning)

- ข้อความแสดงเตือนเมื่อฟิลด์ที่ผู้ใช้เลือกนั้นไม่มีข้อมูลใดๆอยู่เลย หลังจากนั้นโปรแกรมจะย้อนกลับไปหน้าจอแรกเพื่อทำการเลือกฐานข้อมูลใหม่ ดังแสดงในรูปที่ ก.36



รูปที่ ก.36 ข้อความเตือนเมื่อฟิลด์ไม่มีข้อมูลใดๆอยู่

- ข้อความแสดงเตือนเมื่อผู้ใช้เลือกเฉพาะฟิลด์ที่เป็น Numerical หรือ Categorical เพียงอย่างเดียวเท่านั้น ดังแสดงในรูปที่ ก.37



รูปที่ ก.37 ข้อความแสดงเตือนเมื่อเลือกเฉพาะฟิลด์ประเภทใดประเภทหนึ่งเท่านั้น

ประวัติผู้เขียน

ชื่อผู้เขียน นายกิติภัทร กิติกุล
วันเดือนปีเกิด 10 มกราคม 2523
สถานที่เกิด กรุงเทพมหานคร
ปริญญาตรี มหาวิทยาลัยศรีปทุม
คณะสารสนเทศศาสตร์
สาขาคอมพิวเตอร์ธุรกิจ
ปีที่สำเร็จการศึกษา 2546



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้